Artificial Intelligence Applied to Facial Image Analysis and Feature Measurement

Thesis submitted in accordance with the requirements of
the University of Liverpool for the degree of Doctor in Philosophy by

Theiab Alzahrani

June 2022

Contents

Illustrations

List of Figures

List of Tables

Notations

This thesis contains the following notations and abbreviations:

ACC: Accuracy
AI: Artificial Intelligence
AR: Augmented Reality
CNNs: Convolutional Neural Networks
CUDA: Compute Unified Device Architecture
DC: Dice Coefficient
FCNs: Fully Convolutional Networks
FP: False Positive
FN: False Negative
F-Heart: Face shape-Heart
F-Oval: Face shape-Oval
F-Rectangle: Face shape-Rectangle
F-Round: Face shape-Round
F-Square: Face shape-Square
Gender-F: Gender-Female
Gender-M: Gender-Male
GUI: Graphical User Interface
JC: Jaccard Score
LRN: Local Response Normalisation
MCC: Matthews Correlation Coefficient
MLP: Multilayer Perceptron
Mask R-CNN: Mask Regional Convolutional Neural Networks
Position-D: Eye position-Downturned
Position-S: Eye position-Straight
Position-U: Eye position-Upturned
PRE: Precision
RBNs: Radial Basis Networks
ResNet101: Residual Network

RNNs: Recurrent Neural Networks
RPN: Region Proposal Network
Shape-A: Eye shape-Almond
Shape-R: Eye shape-Round
Setting-AVG: Eye setting-Average
Setting-C: Eye setting-Close
Setting-W: Eye setting-Wide
SP: Specificity
SN: Sensitivity
TN: True Negative
TP: True Positive
VR: Virtual Reality

Abstract

Beauty has always played an essential part in society, influencing both everyday human interactions and more significant aspects such as mate selection. The continued and expanding use of beauty products by women and, increasingly, men worldwide has prompted and motivated several companies to develop platforms that effectively integrate into the beauty and cosmetics sector. They attempt to improve the customer experience by combining data with personalisation. Global cosmetics spending is worth billions of dollars, and most of it is wasted on unsuitable or incompatible products. This enables artificial intelligence to alter the rules using computer vision and deep learning approaches, allowing customers to be completely satisfied. With the advanced feature extraction in deep learning, especially convolutional neural networks, automatic facial feature analysis from images for the sake of beauty and beautification has become an emerging subject of study. Scholars studying facial aesthetics have recently made breakthroughs in the areas of facial shape beautification and beauty prediction. In the cosmetics sector, a new line of recommendation system research has arisen. Users benefit from recommendation systems since these systems help them narrow down their options.

This thesis has laid the groundwork for a recommendation system related to beautification purposes through hairstyle and eyelashes leveraging artificial intelligence techniques. One of the most potent descriptors for attribution of personality is facial attributes. Various types of facial attributes are extracted in this thesis, including geometrical, automatic and hand-crafted features. The extracted attributes provide rich information for the recommendation system to produce the final outcome. The coexistence of external effects on the faces, like makeup or retouching, could disguise facial features. This might result in degradation in the performance of facial feature extraction and subsequently in the recommendation system. Thus, three methods are further developed to detect the faces wearing the makeup before passing the images into the recommendation system. This would help to provide more reliable and accurate feature extraction and suggest more suitable recommendation results. This thesis also presents a method for segmenting the facial region with the goal of extending the developed recommendation system by incorporating a synthesised hairstyle virtually on the facial region, thereby harnessing the recommended hairstyle generated by our developed system. Hence, the work presented in this thesis shows the benefits of implementing computational intelligence methods in the beauty and cosmetics sector. It also demonstrates that computational intelligence techniques have redefined the notion of beauty and how the consumer communicates with these emerging intelligent facilities that bring solutions to our fingertips.

Acknowledgements

First of all, I would like to express my gratitude to my supervisor, Dr. Waleed Al-Nuaimy, for his guidance, input and support throughout this research. The time and advice he has given me have been invaluable. Without his valuable support, my thesis would not have been accomplished. Second I would like to thank my teammates, especially Dr. Baidaa Al-Bander, for their comprehensive and precise feedback and suggestions while reviewing my thesis and published work. I would also thank the annual Progress Assessment Panel (PAP) committee members for their great advice provided during my PhD.

I would also like to express my gratitude for all academic and technical support received from the people in the Dept. of Electrical Engineering and Electronics. I gratefully acknowledge the Kingdom of Saudi Arabia government for the financial support of my study. Finally, my deep and sincere gratitude to my parents, wife, daughters, and son for their continuous and unparalleled love, help and support.

Chapter 1

Introduction

1.1 Background

”Beauty lies in the eye of the beholder” is a statement defined by Greece philosopher Plato. On an individual basis, different people may have different judgments regarding the beauty and attractiveness of particular faces. However, there are noticeable trends in facial beauty preferences at the population level. Beautification resulting from facial cosmetics, plastic surgery, or facial retouching has also a significant impact on facial beauty and has become dominant in modern society [1]. Human beings have been attracted by beauty since the dawn of humanity. The beauty of the human face has motivated and inspired numerous poets, artists and philosophers. The definition of beauty and cosmetics standards, including smooth hair, clear skin, and beautiful eyes, has evolved to encompass hair colour and extensions, facial contouring, eyebrow enhancement, and eyelashes extension. There is now a product for every curve on the face that could beautify the look and appearance.

Several psychological studies have found that facial beauty and attractiveness are universal concepts that surpass cultural boundaries, with a high cross-cultural agreement in facial attractiveness evaluations among assessors from various socio-economic classes, ethnicities, gender, and ages [2]. From facial symmetry [3], averageness[4], sexual dimorphism [5], to other personality attributes [6], psychologists have explored

a wide range of characteristics that have could formulate the beauty notion.

Over time, the definition of beauty has evolved. Increased cosmetics and beauty needs have resulted from growing global economies, higher incomes, and lifestyle changes. The worldwide beauty sector is now worth $532 billions of dollars. Nowadays, the United States leads the world in beauty, with a 20% market share, followed by China (13%) and Japan (11%). While estimates vary, most experts believe that growth will continue at a 5% to 7% growth rate, reaching approximately 800 billion dollars by 2025. Even if the global economy suffers a setback between now and then, the beauty industry is more likely than other industries to hold its own [7].

In addition to psychological studies of beauty, computer scientists have studied facial beauty based on computerised methods. Many efficient models for beautification have been developed by researchers [8–11]. As Zhang et al comment [12], "Computer-based facial beauty study aims to discover the rules of facial beauty from data using techniques such as image processing, statistical analysis, and machine learning, and develop application systems based on the discovered rules." They also claim that "The history of facial beauty study can be divided into several stages, from the ideal ratios proposed by ancient artists to the measurement-based rules used in aesthetic plastic surgery and from the hypotheses proposed by modern psychologists to data-driven facial beauty modelling emerging in the computer science society".

Facial attribute descriptors are one of the most significant features used in pattern recognition and computer vision research. The researchers in these areas have spent efforts on the extraction of facial attributes from images and adopted them in variety of computer vision applications, such as object detection and classification [13], face recognition, verification, identification [14–16], and facial image searching [17]. It has also been proven that gender identification can be boosted by making use of the interdependencies between age, gender, and other facial characteristics [18].

Artificial intelligence (AI) is a crucial driver of the Fourth Industrial Revolution, and it is transforming how people think, work, and communicate with data and technology.

AI has the potential to revolutionise the development of novel formulations. Data has long been utilised to improve products and formulate better solutions. In the same way, data analysis will result in enhanced cosmetics producing more effective and lasting longer data-driven formulas. AI adds the power of customisation, augmented decision making, process automation, and digital supply chain networks to the consumer goods business, which have previously demanded human intelligence. With the advancement of AI, a growing number of companies are embracing product customisation and providing customers with knowledge-based beauty practices [19].

Users can receive personalised service help through recommendation algorithms, which learn their previous behaviours and individual attributes to predict their preferences or recommend the most suitable selection for specific products. In the development of recommendation systems, artificial intelligence (AI), particularly machine and, more specifically, deep learning approaches, has been effectively utilised to boost prediction accuracy and tackle data sparsity [20]. Virtual try-on technology is now widely used across online retail and social media platforms. Consumers can virtually test on cosmetics via augmented reality before purchasing them, which is especially useful for cosmetics [21, 22].

This chapter aims to deliver an overview of the research, including the problem description and motivation for the study, as well as the objectives and goals. Eventually, the novelties and outcomes produced from the thesis are highlighted. The thesis's contents is also outlined to provide a clear road map to follow while reading the thesis.

1.2 Problem Statement and Motivation

The usage of cosmetics and beauty solutions can be traced back to around 4,000 BC. Laurentini and Bottino [23] have commented that "what produces the human perception of beauty is a long-standing problem in human sciences and, more recently, in medical areas such as plastic surgery and orthodontics". As Liu [24] has observed, based on

the work of Little et al. [25] facial beauty imposes a significant and varied impact on social behaviours such as mate selection and social exchange. Lui [24] also showed, based on the work of Riggio and Woll [26], "attractive people have more dates than less attractive people". Another observation reported by Lui [24], based on the work of [27], showed that males and females who are more attractive are happier with their dates. Final observation given by [24], according to [28], revealed that attractive people have a better chance of being hired than less attractive ones.

Beauty and technology have been separate throughout history, but they have finally blended into one lane. Technology rapidly alters consumer-brand relationships, resulting in a massive revolution in the beauty and personal care market. With so much information and services being presented to the consumer, tailored recommendations are a great addition. The individualised guidance might make products exploration more enjoyable. AI, particularly machine learning and computer vision techniques, can significantly assist in making recommendations that benefit the user. They have the power and authority to determine customers' daily demands based on their age, skin, hair, face and body shape, and other personal characteristics.

Facial attributes like eye features, including shape, size, position, setting, face shape and contour, determine what makeup technique, eyelashes extension and haircut styles should be applied. Therefore, it is important for beautician to know clients' face and eye features before stepping into the application of makeup and eyelashes or the styling of hair. Face and eye features are essential for beauty expert because different types of face shape and eye features are critical information to decide what kind of eye shadows, eyeliners, eyelashes extension, haircut style and colour of cosmetics are best suited to a particular individual.

Thus, automation of facial attribute analysis tasks, based on developing computer-based models for cosmetic purposes, would help to easing people's life and reducing the time and efforts spent by beauty experts. Furthermore, virtual consultation and recommendation systems based on facial and eye attributes [10, 29–34] have secured a

foothold in the market and individuals are opening up to the likelihood of substituting a visit to a physical facility with an online option or using a specific-task software. The utilisation of the innovation of virtual recommendation systems has numerous advantages, including accessibility, enhanced privacy and communication, cost-saving, and comfort. Hence, our work's motivation is to automate the identification of eye attributes and face shape and subsequently produce a recommendation to the user for appropriate eyelashes and a suitable hairstyle.

1.3 Aims and Objectives

This research project aims to develop computer-based facial analysis systems targeting the beauty and cosmetic industry. The targets of this thesis will be the development and design of intelligent computerised methods to automatically identify of facial features that could advance the automation of facial attributes analysis tasks, help to easing people's life and alleviate the need for additional time and efforts made by experts for beautification and aesthetic purposes. To develop these systems, various computer vision and artificial intelligence (i.e.machine and deep learning) strategies are exploited.

In general, this study aims to accomplish the following goals:

1. **Integrated Multi-Model Face Shape and Eye Attribute Identification:** This research direction targets to develop integrated models for face shape and eye attribute identification for hairstyle and eyelashes recommendation.

2. **Facial Makeup Detection:** This research direction focuses on developing a method that helps identify the presence of makeup in unconstrained face images. The detected landmarks' location could be negatively affected and changed by the applied facial cosmetics, which leads to incorrect facial feature detection. Thus, makeup detection is a prerequisite process in the hairstyle and eyelashes recommendation system, which helps to ensure that the processed image is clear of makeup.

3. **Face Contour Localisation for Virtual Hairstyle Makeover Purpose:** The developed hairstyle and eyelashes recommendation system could be extended by synthesis of the hairstyle and virtual application of the recommended hairstyle. This research direction aims to develop an efficient face boundary detection method. The automatically extracted face contour would help try on the haircut virtually on the localised boundary of the face.

1.4 Novelty and Contribution

This section presents the significant novelty and contributions of the work presented in this thesis:

- A new framework merging two types of virtual recommendation, including hairstyle and eyelashes has been introduced. The overall framework is novel and promising.
- A method is able to extract many complex facial features has been proposed, including eye attributes and face shape, accomplishing the extraction of features and detection simultaneously.
- A user-friendly interface system evaluated on a dataset provided with diverse lighting, age, and ethnicity has been presented.
- In the face shape classification method, a method based on merging hand-crafted features with deep learning features has been developed.
- A new geometrical measurement method to determine eye features, including eye shape, position, and setting, based on the coordinates of twelve detected eye landmarks has been presented.

- Makeup detection schemes that help detect facial images covered by makeup using labelled and unlabelled data have been introduced. This has been accomplished by investigating the impact of three different learning strategies to automatically detect the presence of makeup. A thorough analysis and comprehensive study involving six challenging datasets has been carried out. This enables evaluation with facial images collected from various sources, demonstrating the robustness and relatability of the approaches.

- A deep learning-based technique to localise the face region in unconstrained images has been developed. The developed segmentation method is a robust and reliable approach that could handle all face variations and conditions.

1.5 Hardware and Software

Several computer vision libraries and deep learning frameworks have been used to set up and design the developed systems in this thesis. These libraries allow for a high-level implementation without being concerned with low-level implementation. Theano [35], Caffe [36], Tensorflow [37], and Torch [38] are widely used open-source deep learning frameworks implemented using various programming languages such as C++, Python and Java. These frameworks eliminate time-consuming implementation from scratch and support parallel computing in graphical processing units (GPU) for efficient convolutional neural networks implementation and training. NVIDIA[1] released a programming model for their GPUs, named compute unified device architecture (CUDA), enabling for an intensive parallel computing demanded by deep learning algorithms. Experiments conducted in this thesis were performed on an HP Z440 workstation that has RAM of 16GB, E5 3.50GHz processor (Xeon model), and a GPU of 12GB (GTX TITAN X model) running Linux Mint. Tensorflow [37], Keras [2], Scikit-Learn [3], OpenCV

[1] `https://www.nvidia.com/en-me/geforce/`
[2] `https://keras.io/`
[3] `https://scikit-learn.org/stable/`

[4], NumPy [5], SciPy [6], Matplotlib [7], Dlib [8], PyQt [9] are the python libraries used to implement deep learning models, image processing techniques, visualisation, GUI designing, and landmark detection methods in this thesis. To develop our models, many off-the-shelf convolutional neural networks are adopted and adapted to establish the backbone of our models in the thesis, including AlexNet [39], Inception3 [40], VGG16 [41], and ResNet101 [42].

1.6 Evaluation Criteria

This section presents elaboration of the criteria taken into consideration in this thesis to evaluate the success of our developed models. As we deal with classification and segmentation problems, the most popular measurements typically used for classification and segmentation evaluation have been established. The performance of each developed model has been evaluated using one or more of the below evaluation metrics. To formulate the evaluation criteria, four variables including, True Positive (tp), True Negative (tn), False Positive (fp), and False Negative (fn) should be introduced. These four parameters constitute the confusion matrix. A confusion matrix, shown in Figure 1.1, is a method for summarising the performance of a classification and segmentation algorithm. True Positive (tp) are image examples identified by a machine learning model as positive, and the real label was positive too. Likewise, true Negative (tn) are image examples identified by a machine learning model as negative, and the real label was negative too. False Positive (fp) are image examples predicted by a model as positive, and the actual class was negative. False Negative (fn) are image examples predicted by a model as negative, and the actual class was positive. The evaluation criteria are defined as follows:

[4] https://opencv.org/
[5] https://numpy.org/
[6] https://scipy.org/
[7] https://matplotlib.org/
[8] https://pypi.org/project/dlib/
[9] https://matplotlib.org/

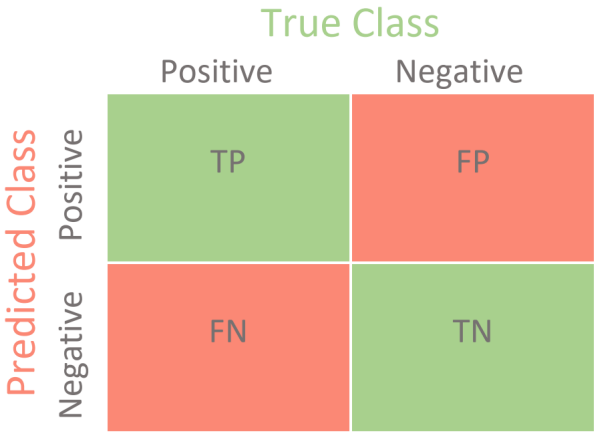


FIGURE 1.1: Confusion matrix.

- Accuracy: This measurement assesses how well the predicted value matches the actual value. The following formula could be used to define it:

$$Accuracy\ (Acc) = \frac{tp + tn}{tp + tn + fp + fn} \tag{1.1}$$

- Precision: The precision evaluates the classifier's ability to exclude irrelevant data. This metric's formula could be given as follows:

$$Precision\ (Pre) = \frac{tp}{tp + fp} \tag{1.2}$$

- Sensitivity: The sensitivity score counts how many relevant samples were accurately identified. It can be expressed in the following way:

$$Sensitivity\ (Recall)\ (Sn) = \frac{tp}{tp + fn} \tag{1.3}$$

- Specificity: It refers to the classifier's ability to recognise the actual negative rate. The following equation can be used to determine the specificity formula:

$$Specificity\ (Sp) = \frac{tn}{tn + fp} \tag{1.4}$$

- F1-Score: This score is calculated as a weighted mean of sensitivity (recall) and accuracy, with both recall and precision contributing equally to the F1-score. The F1-score is described as follows:

$$F_1\ Score = \frac{2(Precision \times Recall)}{Precision + Recall} \tag{1.5}$$

- Matthews Correlation Coefficient (MCC): For real and estimated binary classifications, MCC produces a value in the range of $[-1, +1]$. An outcome of $+1$ typically refers to an accurate prediction, 0 represents a random estimation, while -1 represents total (non-correlation) disagreement between the real values and predictions [43]. MCC is defined as follows:

$$MCC = \frac{t_p \times t_n - f_p \times f_n}{\sqrt{(t_p + f_p)\,(t_p + f_n)\,(t_n + f_p)\,(t_n + f_n)}} \tag{1.6}$$

- Dice Coefficient (DC): DC is two multiplied by the area of overlapping between the predicted segmentation obtained from a model and the ground truth annotated manually divided by the sum of total number of pixels in both images.

$$DC = \frac{(2 \times tp)}{(2 \times tp) + fp + fn} \tag{1.7}$$

- Jaccard Score (JC): JC also known as Intersection Over Union (IoU), the area of overlapping between the estimated segmentation mask and the segmentation mask provided manually divided by the measured area of union between the estimated and real segmentation mask [44, 45]. This score can be defined as follows:

$$Jaccard\ Score\ (JC) = \frac{tp}{tp + fp + fn} \tag{1.8}$$

1.7 Published Work

This research work has resulted in the following papers published in numerous refereed publications.

- Theiab Alzahrani, and Waleed Al-Nuaimy. ”Face segmentation based object localisation with deep learning from unconstrained images.” In 10th International Conference on Pattern Recognition Systems (ICPRS-2019), pp. 47-51. IET, 2019. **Author Contributions:** T.A. conceived the idea, conducted the experiments, wrote the manuscript. W.A.-N. conceived the idea of research work, revised the manuscript, and ran the project.

- Theiab Alzahrani, Waleed Al-Nuaimy, and Baidaa Al-Bander. ”Integrated multi-model face shape and eye attributes identification for hair style and eyelashes recommendation.” Computation 9, no. 5 (2021). **Author Contributions:** T.A. designed and performed the experiments. W.A.-N. conceived the idea of research work and ran the project. B.A.-B. contributed to the data analysis and paper writing up.

- Theiab Alzahrani, Waleed Al-Nuaimy, and Baidaa Al-Bander. ”Hybrid feature learning and engineering based approach for face shape classification.” In 2019 International Conference on Intelligent Systems and Advanced Computing Sciences (ISACS), pp. 1-4. IEEE, 2019. **Author Contributions:** This publication is part of the journal work ”Integrated multi-model face shape and eye attributes identification for hair style and eyelashes recommendation”. The authors contributed in a similar way to the journal work.

- Theiab Alzahrani, Baidaa Al-Bander, and Waleed Al-Nuaimy. ”Deep learning models for automatic makeup detection.” AI 2, no. 4 (2021). **Author Contributions:** T.A. conceived and designed the experiments; T.A. and B.A.-B. performed the experiments; T.A., B.A.-B. and W.A.-N. analysed the data and wrote the paper; W.A.-N. supervised and ran the project.

1.8 Thesis Structure

The following is a list of the remaining chapters in this thesis:

Chapter 2 - Literature Survey

This chapter presents an overview of facial beauty and attractiveness by highlighting the impact of beauty in psychological studies, aesthetic plastic surgery, and computer science. Moreover, beautification methods and their influence on facial beauty are described. Further, a review of artificial intelligence, deep learning and facial image analysis is provided, and several CNN models are exhibited. The role of AI in the beauty and cosmetics industry is also presented. Finally, existing intelligent virtual try-on and recommendation systems for beauty are described.

Chapter 3 - Face Shape and Eye Attribute Identification

This chapter describes the development of system with a friendly graphical user interface that can analyse eye and face attributes from an input image. Based on the detected eye and face features, the system suggests a suitable recommendation of eyelashes and hairstyles for both men and women.

Chapter 4 - Makeup Detection

This chapter introduces makeup detection schemes that help detect facial images covered by makeup using labelled and unlabelled data. A comprehensive comparative study investigating the impact of different learning strategies to detect makeup automatically has been conducted. Evaluation of the developed deep learning models is carried out using six challenging datasets. The obtained results reflect the efficiency of combining different learning strategies by harnessing labelled and unlabelled data.

Chapter 5 - Face Segmentation

This chapter presents a method to (segment) localise the face boundary to help apply a haircut virtually on the localised boundary of the face. Initially, existing face segmentation methods are presented. Following that, the recommended solution for overcoming the earlier methods' shortcomings is described. The performance of the developed

system is discussed, as well as the experimental findings, proving the viability of work.

Chapter 6 - Conclusions and Future Work

Here, the findings of this thesis is highlighted and discussed. The potential future study possibilities based on the findings are also explored in this chapter.

Chapter 2

Literature Review

2.1 Overview of Facial Beauty and Attractiveness

2.1.1 Beauty and Attractiveness: Early Exploration

Human physical beauty and attractiveness have been studied for over 4000 years. Starting from ancient Egypt, artists established specific guidelines to create beautiful and harmonious paintings. These adopted guidelines were based on measuring a series of human body ratios. Sculptors followed the Egyptian guidelines in ancient Greece and Rome. According to [12], Polykleitos (450-420 BC), for example, stated that "the length of the head is 1/8 of the body height, and the length of the face is 1/10 of the body height" [46]. Neoclassical criteria were founded by Renaissance artists who devised more detailed ratio rules. The neoclassical criteria were used to define and establish the proportions between various areas of the head and face. Beautiful looks were also thought to be based on the golden ratio (ϕ = 1.618). This ratio can be described and exemplified by Mona Lisa, the iconic painting, as illustrated in Figure 2.1. In another place on the planet, ancient China, additional ratio-based rules were established, which are known as the vertical thirds and horizontal fifths, as depicted in Figure 2.2. Faces that followed the established rules were thought to be beautiful and attractive [12].

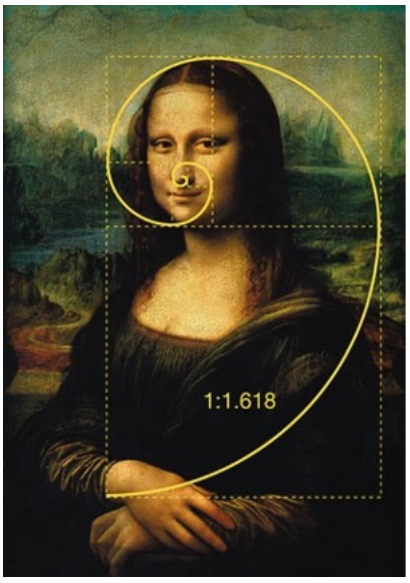
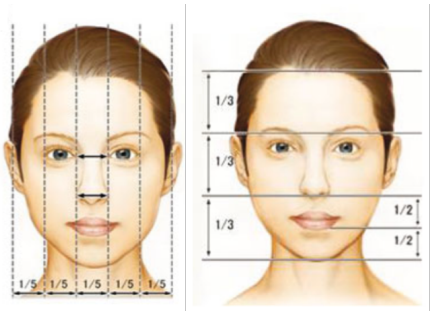


FIGURE 2.1: Mona Lisa and golden ratio (by Leonardo Da Vinci).



(a) (b)

FIGURE 2.2: Explanations of horizontal fifth vertical thirds rule. a) Horizontal fifths, b) Vertical thirds [12].

Early investigations into facial beauty targeted mathematical approaches based on numerical ratio descriptions. In modern arts and cosmetic surgery, some of these rules are still used. Those rules, however, have considerable limits due to time and measurement procedures. For instance, the precision of rules is low, and the majority of ratios are expressed as simple integers or fractions, such as 1, 1/3, or 1/5, which are sometimes based on subjective judgments and opinions [12]. Although these ratios can be utilised as a guide throughout the design process, they are sometimes inconsistent compared to the real measurements of faces [47]. Furthermore, the rules are not all-inclusive. Only two types of ratio measurements, including the horizontal and vertical, were used to characterise the size and position of facial features. This results in omitting crucial

aspects of face beauty represented by the texture of the skin and the shape of facial structure [12].

In the 1960s, as per psychological studies related to facial beauty, two significant findings revolutionised the concept of beauty analysis studies [12]. In this decade, Martin discovered that persons of different backgrounds might exhibit similar perceptions of face appearance [48]. According to Walster et al. [49], facial attractiveness had a considerable impact on dating behaviour. Following that, many researchers joined the face attractiveness research, resulting in thousands of publications. To study the perception of beauty, the psychologist concentrated on historical and social considerations. Exploring the social factors, the researchers considered context (for instance, hetero- or homo-sexual) [50], self-assessment [51]), age (for instance, young adults) [52], inability to recognise the facial preference (for instance, prosopagnosia condition), consequences of beauty (for instance, partnership/marriage) [49], and income [53]. Historical conditions takes into consideration the population and individual preferences for beauty and attractiveness vary through time [54].

In aesthetic plastic surgery, the attributes of beautiful faces were also explored. Apart from the traditional ratio rules used to assess facial beauty, anthropometry, defined as studying measurements and proportions of the human body, is adopted to make these rules more precise and objective compared to the conventional rule measurements [12]. Choe et al. [55] analysed the face of 72 Korean American females and discovered that the faces do not follow the neoclassical norms. Their findings also revealed that the golden ratio does not fit the normal distribution of the established range of ratios. In their assessment study, the authors of [56] created a synthesised face accompanied by the golden ratios; however, the findings revealed a poor prediction performance in terms of beauty level score. Although the conventional ratio-based approaches were shown to be inaccurate in the previous works, the concept of utilising them to explain the geometrical structure of faces has been acknowledged and accepted. Moreover, conventional

methods based on measuring the ratios were employed to aid the diagnostic and planning of plastic surgery. Nonetheless, with so many ratio rules, it is difficult to determine which individual rules are more significant. In practice, a surgeon's experience is more significant than ratios, which are merely used as a guide [12, 57].

2.1.2 Exploration of Facial Beauty Related Features using Computerised Techniques

One of the goals of studying the human perception of beauty is to improve the cosmetics and beauty industry. The majority of studies conducted target natural facial features by developing innovative products and technologies that should be safe to use. On the other hand, in clinical fields, especially, orthodontics and plastic surgery, understanding what causes individuals to perceive facial beauty is a well-established issue [23]. The facial beauty based on computerised approaches is a new field of study. Several thousand papers on this topic have been produced in the last several decades. According to [12], facial beauty analysis can broadly categorised into four phases.

The psychology and aesthetic surgery domains dominated the published works on facial attractiveness studies throughout the first stage (1990–2000). To create stimulus images or calculate ratio values, computer software was used. To achieve the goal of beauty analysis, several challenges were undergone during this phase. Thus, two significant problems were posed but not solved, including finding and setup relative and comprehensive facial attributes for studying the perception of beauty and evaluating the discriminative ability of these attributes. Meanwhile, the internet resources and information processing techniques conducted by advanced computers were rapidly evolving. Many successful methods in computer vision and machine learning were suggested and implemented, laying the groundwork for image-based beauty research [12].

In the second phase, between 2001 and 2007, the researchers concentrated on predicting the beauty of faces using facial ratio features. This phase was a period of transition. On one side, the ratio features were chosen using traditional criteria. On the

other hand, a new concept was introduced represented by facial beauty analysis based on data-driven and machine learning approaches. To conduct data-driven based methods, facial image data collection and manually scoring the beauty level of these images by domain experts are typically prerequisite tasks [58–62]. Second, based on the facial images, the appropriate ratio features were generated [58, 62]. Third, supervised learning approaches such as support vector regression (SVR) and linear regression models [63] were adopted to design a computational model, targeting scoring the facial beauty level [58, 59, 64, 65]. Finally, a correlation metric between the predicted beauty ratings from the model and the manual beauty level scored by an expert (ground truth) was used to assess the model's performance. The limitations of this phase can be seen as follows: (i) the data sets utilised to develop models were small-scale, and (ii) to construct facial beauty prediction models, only attributes related to facial ratios were extracted and used, resulting in a lack of essential attributes such as the anatomical structure of the face and the conditions of the skin. As a result, the models created at this level had restricted capabilities [12].

The facial beauty analysis approaches developed earlier were further improved in the third phase (2008–2009). Several efficient image processing algorithms were used to enhance the automation of beauty analysis [12]. Face detection is achieved with the Viola-Jones [66] technique, facial landmark localisation was handled with active shape models (ASMs) [67], and facial deformation was performed with the Multilevel Free-Form Deformation (MFFD) [68] methodology. Data-driven models were utilised to assist facial beautification, which is one of the facial beauty model applications. The tests were still based on small databases at this time, and the research was still limited to geometric aspects [12].

The fourth phase is from 2010 till now. In this phase, the size of the face image data sets has grown dramatically, and the visual data come from a variety of sources, including the web [69] and videos [8]. Furthermore, local descriptors that function well in facial recognition, such as Local Binary Patterns (LBP) [70] and the Gabor filter

[71], were employed to develop facial beauty models in addition to geometric features. Moreover, cutting-edge machine and deep learning algorithms are used to analyse and model facial beauty, e.g., CNNs [69], cost-sensitive regression [72], self-learning [73], and facial feature extraction based augmented reality (AR) [74]. It can be noticed that the approaches for facial beauty assessments are becoming more diverse as the related research areas develop. Facial beauty analysis and beautification, on the other hand, is still in its infancy when compared to face recognition and expression recognition.

2.2 Beautification Strategies

Beautification is the process of enhancing a person's appearance. Beautification aims to identify, locate and analyse the significant key-points on the face to enhance natural beauty. Beautification, which is induced by plastic surgery and cosmetics, and by means like eyelash extensions and hairstyles, can significantly modify facial appearance. This section sheds light on some facial beautification strategies presented as follows:

2.2.1 Influence of Eyelash Extensions on Facial Beauty

Human eyelashes are one of the facial traits that contribute to the overall appearance of a person's face. Eyelashes have been utilised to enhance and improve facial appearance, particularly the beauty of the eyes, in numerous cultures, including ancient Middle Eastern and North African civilisations. Currently, similar habits are reflected by women's use of mascara and curlers, as having long eyelashes is considered attractive. Eyelashes and brows are significant anatomical structures in the face [75]. Eyelash loss or shortening can indicate dermatological or nutritional problems, as well as endocrine or systemic conditions [76, 77]. Pazhoohi et al. [78] examined the differences between male and female faces. They found that while eyelashes with an optimal ratio of eyelash length to eye width are deemed more beautiful in general, however, this preference is influenced by cultural traditions as well.

According to the findings published by [79], females with longer eyelashes, but not males, are perceived to be more beautiful; perceptions of femininity and health also significantly boost with increasing the length of eyelashes; and older women can take advantage the most from enhanced eyelashes, though longer eyelashes do not affect the perception of age. Figures 2.3 and 2.4 depict the impact of eyelash length on the perceived attractiveness and beauty. Researchers have also discovered that darkening eyelashes, possibly enhanced by mascara, serve to emphasise the sclera (the white area of the eyes) where the brightness of which could refer to youth and health [80].

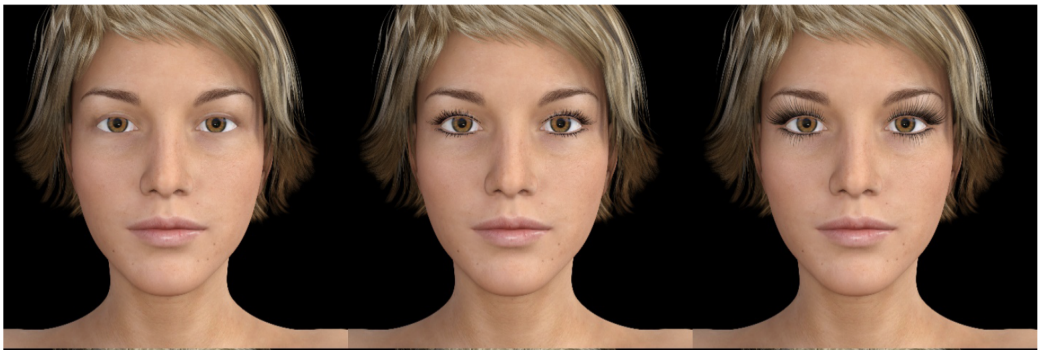


FIGURE 2.3: Image examples show the impact of eyelashes length. The ratio of eyelash length to the width of eye are 0, 0.24, and 0.5, respectively, from the left to right [78].

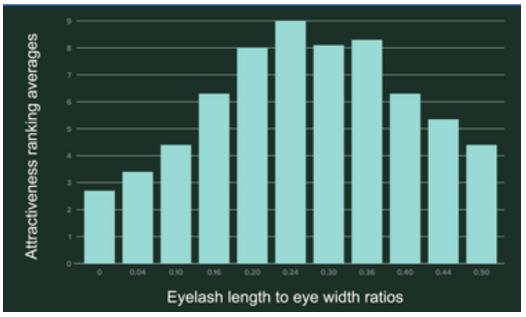


FIGURE 2.4: Eyelash length and attractiveness of female [78].

Artificial eyelash extensions are synthetic lashes which are attached one at a time to natural lashes. Eyelash extensions may be applied on natural lashes using two distinct ways, giving more options for individuals to get the desired appearance. One technique is known as ”Classic.” This is the traditional method of applying for extensions. The classic procedure involves attaching one synthetic lash to one natural lash one at a time.

Because only one extension is placed to each natural lash, it is feasible to utilise various thicknesses, some of which are thicker than the normal lash, to generate variable depths of colour in the eyelashes. The second approach is also known as Volume lashes, Russian lashes, or nD lashes, where n = 2,..., 9. The synthetic eyelash extensions used in this procedure are substantially finer than those used in the Classic method. The process spreads the lashes out before being attached to one particular natural lash [81].

Facial symmetry is a well-studied aspect of facial attractiveness. Symmetry may serve as a signal of genetic integrity and phenotypic in evolutionary terms, and it is valued during mate selection in many cultures. According to the symmetry theory, the more symmetric a face is, the more attractive it is. Symmetry has long been a factor in calculating and modelling facial beauty. The use of 2D and 3D photographs to quantify face asymmetry has been examined. As shown in Figure 2.5, "horizontal distances from a vertical reference line, vertical distances from a horizontal reference line, and the horizontal direction between centres of bilateral points are commonly employed to assess face asymmetry in 2D images" [12, 82]. Synthetic eyelash extensions can help to improve and balance face symmetry while also widening the eyes. Eyelash extensions can also give diverse effects to the eyes, such as the impression of make-up without the use of make-up. The appropriate shape of eyelashes can help open up the eye area and improve the symmetry of the face. The diameter, size, shape, colour, and quality of eyelash extensions vary greatly. They are mainly made of synthetic material and meant to seem like human hair [83].

2.2.2 Impact of Hairstyles on Facial Beauty

Hair is a significant trait that can enhance or detract from attractiveness and beauty. Individual attractiveness can be determined by hair's colour, length, style, and condition. Although the impacts of hair of head on aesthetic judgments have occasionally been studied from an evolutionary point of view, facial characteristics rated beautiful in societies have been addressed using the theory of sexual selection. Hair, on the other hand,

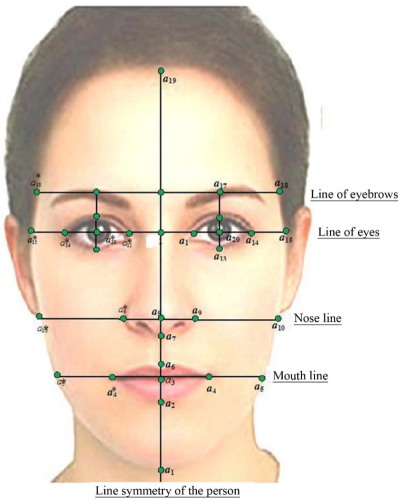


FIGURE 2.5: Identification of bilateral points, and vertical and horizontal lines [84].

plays an important part in the mate selection process [85]. Whenever meeting someone for the first time, one of the primary thoughts that comes to mind is their hair. Blondes were judged as more emotional, attractive, and feminine in an experiment in which the same females were presented as brunettes or blondes, whereas brunettes were assessed as smarter [86]. Blondness, according to the authors of this study, is considered a sign of neoteny, femininity and youthfulness of the face. Mesko and Bereczkei [87] observed, based on a study conducted by the authors of [88], older females keep their hair shorter than younger females and that the condition of the hair is typically associated with female's health. Another observation reported by [87], according to findings of Grammer and colleagues [89], revealed that males may tend to prefer long hair to females. Sexual pheromones generated in apocrine glands could be widely spread on a long hair due to the wide surface granted by the longer hair to host these pheromones, according to Grammer et al. [89].

In many conducted research studies, men with scalp hair were assessed as more attractive, healthy, and energetic than balding men. Male baldness was associated with lower ratings of socially desirable features [87, 90]. On the other hand, a receding hairline may indicate maturity and social control, and baldness is regularly correlated to a higher estimation of age and smartness. Likewise, Mesko and Bereczkei [87] investigated the impact of six hairstyles on female facial beauty and attractiveness, including

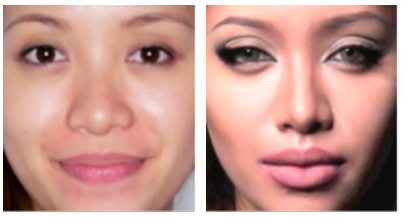
dishevelled, knot, unkempt, short, medium-length, and long, compared to basic faces (do not have visible head hair). Hair can frame faces, highlighting specific features or balancing them off. A well-styled haircut can either enhance or detract from a good first impression. Rather than choosing the latest haircut trend, a haircut that appropriately fits the face shape should be opted for. The right hairstyle and haircut for face shape will effectively frame and balance it while highlighting the most remarkable features for a beautiful and balanced look [91, 92]. Figure 2.6 depicts the impact of different hairstyles and haircuts on Angelina Jolie's face[1].

FIGURE 2.6: The impact of different hairstyles and haircuts on Angelina Jolie face.

2.2.3 Effect of Cosmetics on Facial Beauty

Cosmetics are also used to improve or change the overall appearance by using makeup, cleansers, and lotions. Since Ancient Egypt [93], facial cosmetics have been a vital aspect of daily life in various civilisations, changing people's perceptions of their faces. Figure 2.7 illustrates how cosmetics could alter the perception of beauty, resulting in an improved beauty level [94].

[1] https://www.makeupandbeautyblog.com/hair/angelina-jolie-her-best-hair/



(a) (b)

FIGURE 2.7: Example shows how cosmetics could alter the perception of beauty, resulting in an improved beauty level [94]. Originally was taken from [95].

Women use cosmetics for several reasons, including concern about their facial appearance, compliance to societal norms, and public self-consciousness [96, 97], as well as to look more attractive and assertive to others [96]. Cosmetics are helpful in improving society's views that the wearer may want to change, with people showing up to be healthier [98], demonstrating greater competence, likeability, and trustworthiness [99], and trying to appear more prestigious and influential [100]. Cosmetics also affect the behaviour of others, particularly men, who might tip waitresses wearing cosmetics more generously and more frequently [101]. The boost in attractiveness that cosmetics give faces, which is now a well-known outcome [99], is believed to be the cause of cosmetics' effect on social perceptions. Cosmetics have been shown to modify sex-typical colouration in faces, such as facial contrast [102], and improve the homogeneity of facial skin [103].

The research on how cosmetics impact attractiveness is growing, and many trials employ professionally applied cosmetics to investigate this effect. Professionally applied cosmetics tend to reflect a higher amount of variation in beauty judgments than self-applied cosmetics, according to the authors of [104]. This might indicate that the influence of cosmetics has been exaggerated in the existing literature, with cosmetics improving beauty beyond what can be achieved by daily means. They've also shown that cosmetics have varying effects on various women: more attractive women, especially supermodels, get less of a boost in attractiveness from cosmetics than less attractive women.

2.3 Artificial Intelligence, Deep Learning and Facial Image Analysis

2.3.1 Subsets of Artificial Intelligence

Artificial intelligence (AI) is a method to teach intelligent thinking to a computer, system, or robot. This involves developing algorithms to categorise, analyse, and enable prediction from data. The importance of AI has increased the use of machine learning and deep learning in identifying complex correlations. One area of artificial intelligence that utilises systems to explore data and extract meaningful information used for learning is machine learning (ML) [105]. For training process (model fitting), observations (also termed examples or samples) are needed to determine if there are any implicit patterns hidden within the data. The found patterns are functions or decision boundaries [106, 107].

Machine learning models are typically categorised into supervised or unsupervised models. Supervised models require a data with some samples and classes of those samples, these models can predict future events by learning from the labelled dataset. Unsupervised models require a dataset with some observations without the need for classes of the observations. Unsupervised learning examines how models use unlabelled data to formulate a function that describes a hidden structure in data. Semi-supervised learning lays between supervised and unsupervised learning strategies and uses both labelled and unlabelled data for training [107, 108].

Machine learning is now a widely used technology with many algorithms for implementing intelligent systems. A machine learning technique which simulates the transmission of human brain signals is the artificial neural network (ANN). These networks are comprised of several neurons that resemble neurons in the biological nervous system. These neurons receive data as input and use this to carry out simple calculations. The outputs are then passed to other neurons in a linear or feedback looped pattern [109].

A single and multiple-layer perceptron (or neuron) network is a model of artificial neural networks (ANNs). A perceptron network which has one, or more than one, hidden layer is known as a multilayer perceptron network (MLP). An MLP is also known as a type of feedforward networks, composed of a single one dimension input layer, followed by hidden layers, and a top single output layer. The inputs are accepted by the input layer, then manipulated by the hidden layers, and the result is produced by the output layer. Artificial neural networks (ANNs), including MLP, can learn any non-linear function. Therefore, these networks are usually termed universal function approximators. The mechanism of ANNs' work is to learn weights that map an input to an output. The activation function is one of the main reasons behind universal approximation. Activation functions add non-linear properties to the network, helping it to define the relationship between input and output [110].

Other ANN models include recurrent neural networks (RNNs) and radial basis networks (RBNs). The architecture of RNNs consists of hidden states with recurrent connections. RNNs maintain the sequential information provided in an input data, which is typically ensured by the looping constraint. While making predictions, RNNs capture the sequence of information existing in the data, i.e., dependency among the data. RNNs also share the parameters across different time steps, usually known as parameter sharing. As a result, there are fewer parameters in RNNs to learn and the computing cost is low [111].

Deep learning (DL) is an emerging research trend in ML; the term originates from novel methods for generating features automatically from input data. In DL, the issue of vanishing gradient [112], typically occurred in the top deep layers, has been also overcome. In these layers, the gradients are of insufficient size to produce a training signal and remain rooted in apparent local minima. This can be resolved by developing deep learning architectures with multiple layers which can be effectively trained [113]. Figure 2.8 displays a Venn diagram explaining artificial intelligence as an umbrella terminology, while machine learning is one of its several branches and deep learning as

advanced models of machine learning that do not require extracting the features manually.

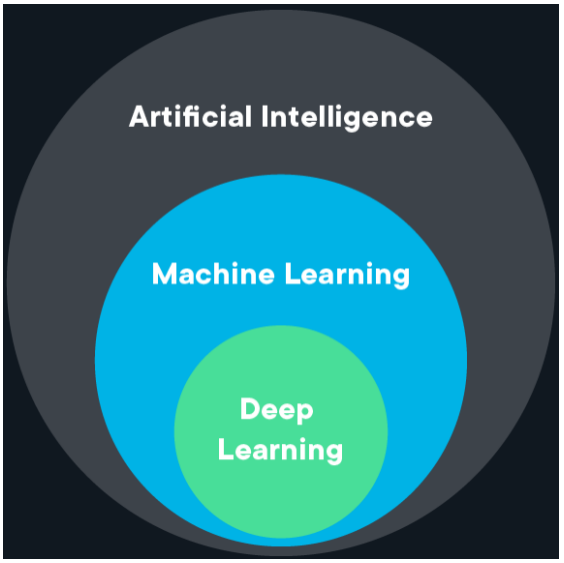


FIGURE 2.8: A Venn diagram depicts artificial intelligence as an umbrella terminology, while machine learning is one of its several branches and deep learning as advanced models of machine learning that do not require extracting the features manually. From [114].

Using layers of neural networks, deep learning interprets higher-level information from raw input data. For example, in image recognition application domains, the earliest layers recognise basic image features such as edges or brightness, the top layers recognise shape details, and a final layer determines the image being presented. A specific type of feedforward multilayer perceptron neural networks, known as deep convolutional neural networks (CNNs), were developed in 1990s, which are much simpler to train with a supervised learning scheme [115]. CNNs are also better at generalising than classic neural networks, which accounts for their popularity in computer vision and speech recognition fields [116].

In addition to supervised learning represented by CNNs, several unsupervised deep learning models have been established, including deep Boltzmann machines (DBMs) [117], auto-encoders [118], deep belief networks (DBNs) [119], long short-term memory (LSTM) [120], and generative adversarial networks (GAN) [121]; these do not require ground truth labels. New computational intelligence algorithms are primarily

reliant upon graphics processing units (GPUs) to efficiently perform sophisticated computations in parallel. GPUs are now preferred over CPUs as the platform for training large and complicated deep learning systems, which has accelerated the success of deep learning network models [122].

CNN models are now widely used as feature learning algorithms. The convolutional neural layers can detect discriminative features from the images and form hierarchies of complex non-linear features whilst penetrating deeper through the convolutional network. CNNs have a multi-layer data representation architecture, in which the earliest layers extract the general features from images, including lines, blobs, and edges. In contrast, the top layers acquire the high-level and complex features containing more information which can then be discriminated among the categories [123], as shown in Figure 2.9.

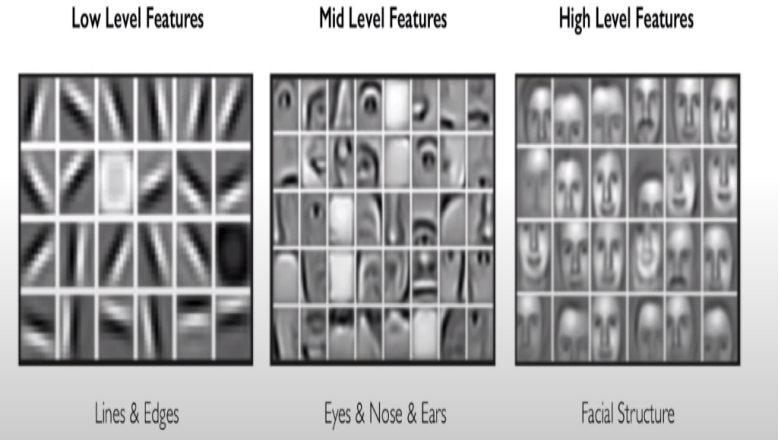


FIGURE 2.9: A feature learning approach is used to produce features [124].

2.3.2 Feature Engineering and Learning

For a while in conventional machine learning models, image features have been manually extracted by using several feature-engineering procedures. Feature engineering is the extraction of meaningful patterns from datasets, which helps machine learning techniques to discernment between binary or multiple classes. Therefore, to make good

predictions, feature engineering is an essential skill [125]. Conventional image prediction, classification and segmentation methods use hand-crafted features, including local binary patterns, edges, colour, histograms oriented gradients, texture, and shape. These hand-crafted features are then fed to machine learning regressors and classifiers, including Multi-class Support Vector Machines (SVM) [126], k-Nearest Neighbours (k-NN) algorithm [127], and Artificial Neural Network (ANN). The classifier then analyses the extracted features by conducting prediction or classification tasks. The current available guidelines for feature extraction from data captured under different environment and domains cannot meet the requirements of different domains. For example, features used for a certain dataset are not always usable for other datasets. Thus, finding new algorithms that can automatically learn features is now essential for developing more accurate machine learning analytics.

To overcome any issues with feature engineering techniques, it has been proposed that feature learning from data based on deep methods be developed. Feature learning is a collection of approaches for automatically discovering the representations needed for feature recognition, including detection, classification, and regression tasks. It can be considered as automated establishment, engineering and extraction of features by intelligent algorithms. The extracted features impose a significant part in computer vision applications, where the kind and meaning of the features determine the accuracy of the applied algorithms [124, 128]. Figure 2.10 states the scheme of conventional machine learning model concept vs. the concept of feature learning algorithm models.

2.3.3 General Components of Deep Convolutional Neural Network Architectures

Because they learn hierarchical features from image data without using hand-crafted features, many convolutional neural networks (CNNs) architecture models achieve superior performance, such as AlexNet [39], VGG [41], GoogLeNet (Inceptionv1) [130], Inception3 [40], InceptionResNetV2 [131], ResNet [42], DenseNet [132], Xception

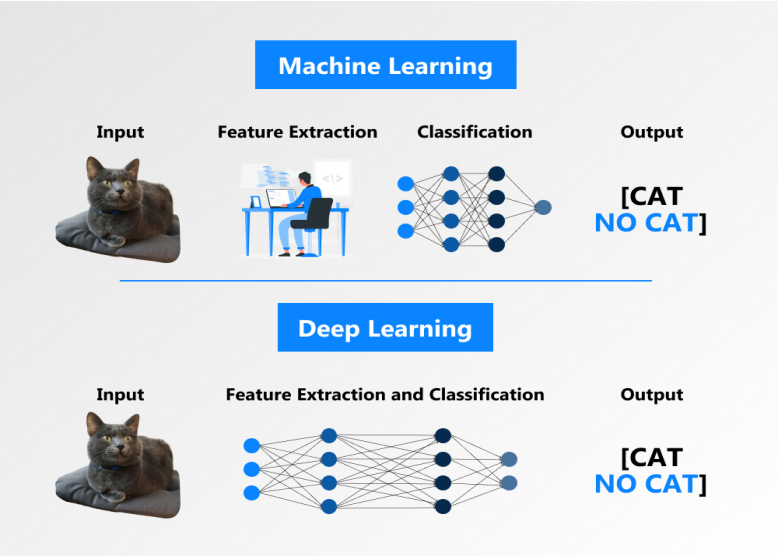


FIGURE 2.10: Machine learning versus deep learning [129].

[133], MobileNet [134], NASNetMobile and NASNetLarge [135], ShuffleNet [136], Darknet-19 [137], Darknet-53 [138], EfficientNetB0 [139] and SqueezeNet [140]. Figure 2.11 illustrates an indication of these networks' relative speeds applied for a classification task. The relative speeds of widely used CNN network models for the classification task are depicted in Figure 2.11. A good network should be fast and accurate. By using a powerful an NVIDIA Tesla P100 GPU and a batch size of 128 samples, the plot shows classification accuracy versus prediction time. The size of each network model architecture on the disc is proportional to each size of blue marker. Here we describe the main structure of the deep CNN model [141]:

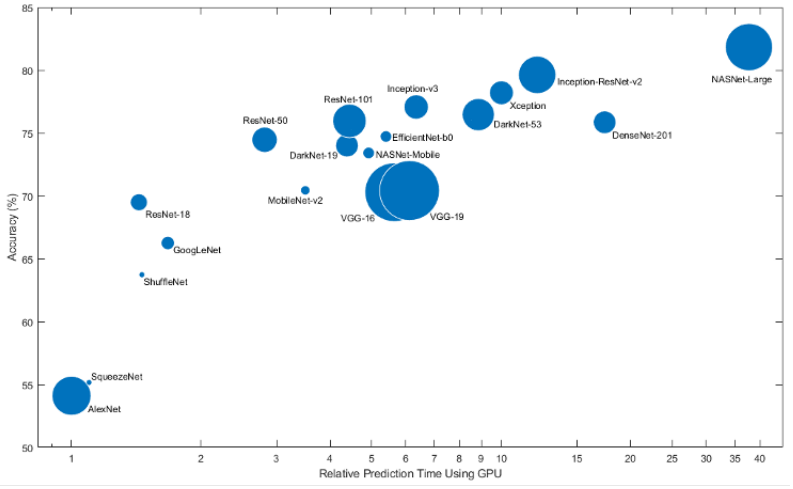


FIGURE 2.11: Various CNN models [141].

1. **Convolutional Layer**: it is a significant component of a CNN model. It's built of many stacked convolutional filters (so-named kernels). The resulted feature map shown in Figure 2.12 is generated by convolving the the width and height of of the input image, which is expressed as N-dimensional metrics, with these kernel filters. For instance, given the receptive field or filter size F, an input size of $N_i \times N_i \times D_i$, provided with stride parameter S and the zero padding of P_i, $N_o \times N_o \times D_o$ can be calculated by a formula to determine the neurons number in the output layer [142]:

$$N_o = \frac{(N_i - F + 2P)}{S + 1}; D_o = K \tag{2.1}$$

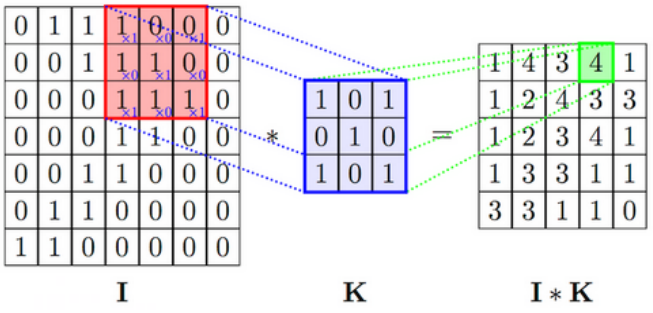


FIGURE 2.12: Explanation of convolution operation applied on a sub-image [143].

The stride parameter S has to be selected so that N_o maintains an integer value. The convolution operation is applied on an image by offsetting filters k over the input image X_i [128]. The formula of convolution operation can be described as follows:

$$X_o = F(\sum_i X_i * k_i + b) \tag{2.2}$$

where b represents the bias of model and F represents any activation function with non-linearity property.

2. **Pooling Layer**: the pooling layer's primary function is to subsample the feature maps with $P \times P$ windows shifted over the feature maps. The value of hyper-parameter P could be defined experimentally. The convolutional functions are

used to build these maps. The size of resulted feature maps is typically reduced by the pooling layer, leading to building a contracted feature map. Thus, the pooling layer helps maintain features, similar to a feature selection phase in machine learning [144]. Both the stride and the pooling kernel are size-assigned before the pooling procedure [145]. There are several forms of pooling strategies that may be used in different pooling layers, including Average and Max pooling, as shown in Figure 2.13.

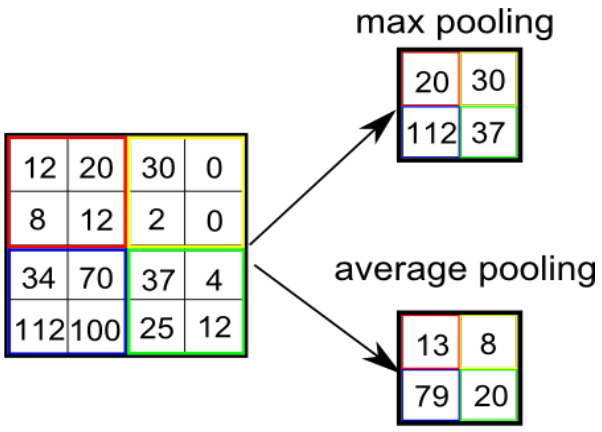


FIGURE 2.13: Average and Max pooling layers. P window size = 2×2.

3. **Fully Connected Layer**: fully connected layers are the final layers of deep neural network architecture which is a feedforward neural networks. The resulted pooling or convolutional layer output is flattened before being input into the fully connected layer. These fully connected layers conduct the same mathematical operations as Artificial Neural Networks on the flattened vectors. The purpose of this layer is to categorise the features maps into several classes after feature extraction. A traditional classifier like support vector machines (SVM) could be used instead of fully connected layers. However, these layers are typically added to make the model end-to-end learnable. Determining the size (number of nodes) in this layer is typically treated as an experimentally chosen hyperparameter [146].

4. **Softmax Layer**: the Softmax activation function is typically attached to the top of the CNN model to compute the likelihood that an input sample belongs to one of several classes. The Softmax function generalises the Logistic Function, which guarantees that all of our predictions sum up to one. In a multi-class problem, Softmax assigns a decimal probability to each class. The sum of those decimal probabilities must equal one. This added restriction allows learning to converge faster than it would without. This function can be represented as follows [147]:

$$\rho(z)_i = \frac{\exp^{z_i}}{\sum_{l=1}^{l} \exp^{z_l}} \tag{2.3}$$

where i = 1, 2, ..., l indexes the output units, l is a number of classes and z are the values represented by a vector to achieve the mapping from an input to output [147]. Auxiliary softmax is a special type of softmax added to the lower levels of the Inception3 network [40], that improves training by mitigating the vanishing gradients issue and speed-up convergence of network training.

5. **Activation Functions**: the weighted sum of the input features is mapped into a non-linear output using activation functions. The activation functions are also known as transfer functions. The selection of the appropriate activation function type has a noticeable influence on the network's capabilities and performance. However, it is not necessary to use the same activation function in all layers of the model. In convolutional neural networks, the Rectified Linear Unit (ReLU) activation function is the most often employed activation function. It is simple to compute and does not saturate or produce the vanishing gradient problem [148]. The mathematical representation of the ReLU and the smoothed ReLU (so-called Softplus) is expressed in the following equations:

$$\theta : x \rightarrow max(0, x) \tag{2.4}$$

$$f(x) = \ln(1 + \exp^{x}) \tag{2.5}$$

$Tanh$ and $sigmoid$ are other types of non-linear activation functions that could be used rather than ReLU. However, they might cause vanishing gradient problems or achieve a lower performance [149].

6. **Normalisation Layer**: during the learning phase, the distribution of input data changes due to the update of parameters, resulting in an internal covariate shift issue that delays the convergence of the network's learning. Normalisation becomes a vital step to guarantee that the model continues to train on the identical input distributions where input data is whitened and decorrelated and speed up network convergence. The normalisation operation is typically conducted before passing the feature maps to the activation function. The normalisation is a standardisation process applied to a batch of data during network learning, resulting in what is called batch normalisation (BN) [150]. Let x represents the layer needs to be normalised, d refers the dimensions of layer x, $x = (x_1, ...x_d)$. Then, the k^{th} dimension can be normalised as follows [128, 150]:

$$\hat{x}^k = \frac{x^k - E[x^k]}{\sqrt{Var[x^k]}} \tag{2.6}$$

Then, $\hat{x}$ is transformed to scale and shift the normalised value as follows:

$$y^k = \gamma^k \hat{x} + \beta^k \tag{2.7}$$

Where the parameters γ and β require tuning during the learning process. Integrating the batch normalisation layer (BN) to the network adapts the formula of the convolution process along with the activation function introduced earlier in Equation(2.2) as follows:

$$z = f(BN(W_x)) \tag{2.8}$$

7. **Regularisation Techniques and Dropout Layer**: the objective of regularisation is to eliminate or decrease the issue of overfitting in convolutional neural networks allowing for better generalisation ability. Several regularisation strategies are developed to tackle the overfitting issue in the CNNs, including norm regularisation ($L2$ type), norm regularisation ($L1$ type) and a dropout scheme. The $L2$ norm (Ridge regression), defined by Equation (2.9), is the most common regularisation method performed by explicitly penalising the parameters of the network model, given the squared magnitude of weights in a pre-defined objective function (least square error cost function in Equation (2.9)). $\frac{1}{2}\lambda w^2$, the penalty term, is introduced to the loss formula, where λ parameter is defined as the degree of regularisation [128].

$$w^* = argmin_w \sum_j \left(t(x_j) - \sum_i w_i h_i(x_j) \right)^2 + \frac{\lambda}{2} \sum_{i=1}^{k} w_i^2 \tag{2.9}$$

To achieve the $L1$ norm regularisation type (Lasso regression), the penalty term $\lambda|w|$ is typically integrated to the cost function for each weight, as defined in the following formula:

$$w^* = argmin_w \sum_j \left(t(x_j) - \sum_i w_i h_i(x_j) \right)^2 + \lambda \sum_{i=1}^{k} |w_i| \tag{2.10}$$

The third regularisation technique is dropout [151]. This layer is an effective approach that introduces noise to the CNNs in a stochastic way to avoid or prevent overfitting. The dropout layer mitigates the overfitting by assigning a certain probability for dropping out the output of some neurons in the hidden layers of a network model. This can be achieved by multiplying hidden units by a variable called Bernoulli distributed random variable. Bernoulli distributed random variable has the value represented by $\frac{1}{p}$ with probability p; otherwise, $p = 1$ refers

to no dropout. In the training stage, a deactivated unit is not participate in forwarding or backpropagation. Each unit is re-activated during the testing phase by multiplying it by one minus the masking probability p [128].

8. **Data Augmentation**: The overfitting problem could be further minimised by using a data augmentation approach that artificially increases data size. During training, the image is randomly augmented at every epoch. Epoch can be defined as a single pass of training data through the CNN model. Random rotation, random vertical and horizontal flips, random zooming, random translations, and random shearing are some examples of data augmentation techniques. The application of various augmentation methods typically enhances network performance [152].

9. **Weight Initialisation**: after completing constructing the architecture of the CNN and prior to learning commencing, the network's weights need to be initialised. Weight initialisation significantly influences a network's convergence rate and performance. Many CNNs' weight initialisation algorithms are based on small and centred around zero initialisation values. One of the most common applied weight initialisation schemes is introduced in Equation (2.11):

$$w \sim \alpha.\upsilon[-\delta, \delta] + \beta.\eta(0, \delta) + \gamma \quad with \quad \alpha, \beta, \gamma \geq 0 \tag{2.11}$$

"Where the term $\eta(0, \delta)$ represents the normal distribution with mean zero and variance of δ and the term $\upsilon[-\delta, \delta]$ represents the uniform distribution" [128]. Several methods have been developed in the literature to define the value of δ, along with the values of α and β parameters, as follows:

- Xavier/Glorot (uniform) [153]:

$$\delta = \sqrt{\frac{6}{n_{in} + n_{out}}} \tag{2.12}$$

where n_{in} = the number of neurons in the previous layer, n_{out} = the number of neurons in the next layer, $\gamma = 0$, $\beta = 0$, and α = 1.

- Xavier/Glorot (normal) [153]:

$$\delta = \sqrt{\frac{2}{n_{in} + n_{out}}} \tag{2.13}$$

where $\alpha = 0$, $\beta = 1$, and $\gamma = 0$.

- He [154]:

$$\delta = \sqrt{\frac{2}{n_{in}}} \tag{2.14}$$

where $\alpha = 0$, $\beta = 1$, and $\gamma = 0$.

10. **Loss Function**: the loss function or objective function is the function that must be minimised and optimised via the network learning phase. Because of an image's true and predicted labels, the loss function guides the learning process by determining the identification error. Loss functions have multiple types used in various machine learning problems and models, including cross-entropy, $L1$, $L2$, Hinge, etc. The cross-entropy loss is typically used for classification tasks, while for regression tasks, L1 and mean squared error (MSE) are usually utilised. The cross-entropy cost function is formulated by this Equation [128]:

$$H_{\hat{p}}(p) = -p \log(\hat{p}) - (1 - p) \log(1 - \hat{p}) \tag{2.15}$$

Where $\hat{p}$ refer to the true label and p represents the predicted value from the CNN model.

11. **Learning/Training and Optimisation**: the CNNs are usually learned/trained by optimising a loss/objective function. The weight parameters in the objective function in a model are typically updated to reach convergence. On a batch of samples, the optimisation process, minimisation task, is typically conducted using a

specific optimisation scheme. Using a gradient-descent approach paired with a back-propagation algorithm, training CNNs is basically equivalent to learning a standard neural network (ANN). However, several particular operations in CNNs, including kernel size, weight sharing, and pooling, necessitate light alterations of back-propagation gradient computation [155]. Convolutions replace matrix multiplications in CNN back-propagation. In addition, the pooling layer calculates the error that occurred by a single winning outcome value, depending on either Max or Average pooling schemes. As a result, in the CNN models, backward propagations vary based on the type of network layer that is propagating. One widely applied optimisation algorithm in the CNN models is Stochastic Gradient Descent (SGD). The SGD updates the weights of the model following this formula:

$$W_{ji}^{(t+1)} \leftarrow W_{ji}^{(t)} + \triangle W_{ji}^{(t+1)} \quad with \quad W_{ji}^{(t+1)} = -\eta \frac{\partial E_B}{\partial W_{ji}} \tag{2.16}$$

Where $\eta \in (0, 1)$ is a parameter that defines the learning rate. Based on SGD, several optimisation algorithms were developed including, Momentum [156], AdaDelta [157], AdaGrad [158], Adam [159], and RMSProp [160]. For example, the Momentum scheme used in this thesis can be represented as follows:

$$W_{ji}^{(t+1)} \leftarrow W_{ji}^{(t)} + \triangle W_{ji}^{(t+1)} \quad with \quad W_{ji}^{(t+1)} = -\eta \frac{\partial E_B}{\partial W_{ji}} + \alpha \triangle W_{ji}^{(t)}, \alpha \in [0, 1] \tag{2.17}$$

12. **Transfer Learning**: is the mapping of knowledge from a certain learned task to a new similar, or different vision task. CNNs require a massive amount of data to be trained; however, the dataset supplied is sometimes insufficient to train the whole network. While data augmentation can artificially double the size of a dataset, transfer learning can help when there are not enough labelled data. Transfer learning may be accomplished by adopting and tuning a pretrained CNN model learned previously on a large amount of data samples. The fine-tuning of the pretrained

CNN model is conducted by deleting the top fully connected layers and retraining the CNN using the trained weights rather than learning from scratch. The transfer learning may also be accomplished by exploiting the pretrained CNN's convolutional layers as a fixed feature-representation extractor. These extracted features are then passed into a linear classifier model to enable its training [161].

2.3.4 Facial Image Analysis

Deep convolutional neural networks (CNNs) have been widely proven to be powerful paradigms and perform excellent performance in several computer vision applications, including video and image processing [39, 41], object recognition [162], facial landmark detection [163, 164], image segmentation [45, 165] and many other applications [166, 167]. Mimic to human vision, computer vision methods process, analyse, recognise, and enhance images and videos. The computer vision methods enable the computers to interpret and perceive what it observes and act on it accordingly [168, 169]. The automatic analysis of images to extract semantic content and flexible features has become increasingly important in recent decades due to the ever-increasing amount of digital images available on the Internet or saved on personal computers. Faces, in particular, carry a lot of useful information compared to other objects or visual elements in images. Specific visual characteristics must be detected, and the facial expressions should be recognised, for example, so that the device can communicate more intuitively and effectively, or facial images from the camera can be encoded and transmitted. As a result, for many applications that involve visual content extraction, automatic facial image analysis is essential. The main goal of facial image analysis is to find an appropriate feature representation, targeting to extract representative attributes, such as gender, identity, position, and expressions in an image [170, 171].

2.4 Artificial Intelligence in the Beauty and Cosmetics Industry

The beauty industry sector has continuously expected to perform well over time (as shown in Figure 2.14) and is significantly different from other sectors. The beauty sector can engage people of all races and ages, from beauty enthusiasts to new advertisers of beauty products. Recent technological developments have greatly facilitated the engagement of new customers by brands. All of the top companies have taken the leading methods of Artificial Intelligence (AI), Augmented Reality (AR) and Virtual Reality (VR) to be industry leaders, such as L'Oreal, Unilever, and Shiseido. With this need for beauty technology, there has been an increase in the companies' challenges and concerns, which require providing suitable solutions [172].

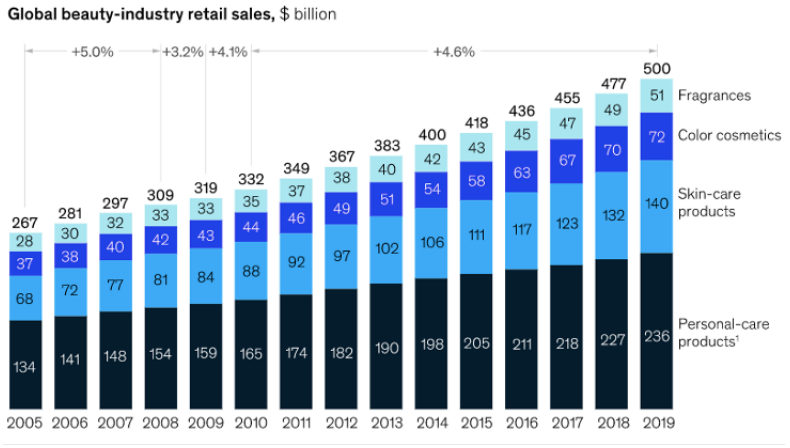


FIGURE 2.14: The global beauty industry market has been consistently resilient [172].

Data and AI can provide the customer with new types of personalisation and customisation through beauty technology [173, 174]. Companies use beauty technology to enhance their correlation with consumers. The idea targets the end-user to feel that the company understands the consumers' needs by providing the best product. Data show that there have been consistent demands for online applications and new beauty innovations, which means that beauty technology can be further explored. COVID-19 crisis is

also expected to accelerate trends that were already impacting the market, such as expanding the use of e-commerce. The worldwide market for cosmetics was estimated at $532,43 billion in 2017 and is projected to grow into $805,61 billion in 2023 [172]. The pandemic brought beauty brands to make a strong connection with AI. The beauty industry will change with the emerging computer vision and deep learning techniques, mainly through the quarantine [175, 176].

Computer vision, which is the branch of artificial intelligence that handles the theoretical development of models that have the ability to discover and analyse meaningful information from image and video data, is the core of future beauty technology [23, 177]. In the beauty sector, computer vision is intended to aid in recognising facial features, the analysis of the data obtained from customers, and the generation of a prediction or decision about their appearance. The beauty and cosmetics industries are experimenting with artificial intelligence to investigate and explore the existing possibilities [178, 179]. AI and deep learning are transforming the cosmetics industry in profound ways [180]. These technologies influence the industry in various ways, from real-time recommendations [181] to product design [182], to virtual assistants (chatbots) and virtual mirrors [183]. All of these solutions aid in the collection of critical consumer data that brands can utilise to understand consumers' demands better and predict market trends. Customer feedback could also be analysed using artificial intelligence systems. The artificial intelligence algorithms can analyse the comments, ratings, and reviews provided by customers on online platforms (e.g. websites or social media) based on incorporating computer vision and natural language processing (NLP) [184]. NLP, machine vision, and deep learning are often used to analyse feedback given by clients in written questionnaires and feedback forms collected from consumers at retail outlets. The final target of analysis of such data is to develop items that customers will be most likely prefer to purchase[185, 186].

Deep learning is also used by Beauty.ai [2] to identify the most beautiful people in

[2] `http://beauty.ai/`

the world. Their system looks at many facial features such as skin colour, gender, wrinkles, face symmetry, age group, and ethnicity to decide the worldwide winners. Yahoo! Research (formerly Yahoo! Labs) [187] has built a deep learning model to classify users' portraits based on multiple facial features. However, Beauty.ai and Yahoo! Research developed systems for assessing the beauty of men and women, i.e. producing scores to the level of beauty. In contrast, the methods developed in this thesis present a recommendation system for beauty improvement. Furthermore, their method is commercial closed source software. They have not described the algorithms developed to achieve their purpose.

2.5 Intelligent Recommendation and Virtual Try-On Systems for Beauty and Cosmetics

Customers in today's world spend a lot of time on their smartphones and computers exploring the products. Beauty technology is the ideal approach for the beauty and cosmetics sector to keep competitive in a saturated globalised economy because such large numbers of people can access the internet. It permits people to try new beauty products that they might otherwise underestimate. Recommender systems have altered interactions with a variety of services. Rather than delivering static data, they provide an interactive experience, the ability to leave feedback, and the ability to personalise the provided options. A recommender system provides tailored informative outcomes for each user independently considering the behaviour of all users towards a particular service [20].

For companies attempting to integrate deep learning and computer vision with beauty, the current step is to build a personalised way to identify the optimal style. Both start-ups and industry leaders provide machine-based recommendations on finding the appropriate personal style and looking more beautiful in the eyes of others, which has

been statistically validated. For example, Sephora's ColorIQ tool [3] employs global tests and over 1,000 foundation combinations to help shoppers choose their exact match. To identify the skin conditions of women, the app records 27 colour-corrected photos, eight light setups, and Ultraviolet light. Mira [4] also leverages computer vision to handle the difficulty of locating influencers, photos, and videos that are relevant to a particular eye shape and appearance. To better select the appropriate products, their method tries to identify an intuitive visual similarity between the eyes of customers and a classifier that extracts human-labelled characteristics given a set of photos of faces.

Another set of advancements in machine learning and artificial intelligence for beauty applications is developing innovative skin care products. These products are personalised according to pre-defined criteria provided by the customer, including medical history, skin type, and age. Proven[5] develops customised skincare products based on the world's "biggest beauty database." It employs machine learning to discover correlations between various product types, their ingredients, and customers' review ratings and then provides consumer recommendations of product ingredients. Curology[6] also leverages machine learning strategies to assess the skin type and medical history of their customers. Following that, customers are associated with a medical specialist who provides special formulae to address the suers' specific skincare needs. Similarly, Function Of Beauty (FoB) [7] applies machine learning to develop personalised shampoo and conditioners with various component combinations considering hair texture, hair type, hair desired outcomes, and other requests.

Some companies even go so far as to design applications to evaluate skin needs and provide customised products. For example, Olay launched the Skin Advisor app [8], which uses a deep learning algorithm to analyse skin and propose beauty products based

[3] https://www.sephora.com/shop/foundation-makeup
[4] https://angel.co/company/mirabeauty
[5] https://www.provenskincare.com/
[6] https://curology.com/
[7] https://www.functionofbeauty.com/
[8] https://www.olayskinadvisor.ca/

on selfies. Atolla Skin Lab [9] considered a similar concept. Recently, the company has relied on a highly specialised database constructed by specialists and a machine learning model to generate combinations of ingredients considering multiple skin features such as hydration of skin, the content of oil, Ultraviolet light damage, age, and skin issues and desired outcome. Atolla is developing a smartphone app that uses computer vision to monitor findings, enhance the developed algorithm, and enable the company to make necessary improvements.

Furthermore, AI algorithms, especially machine and deep learning paradigms, appear to be the foundation of specific innovative ideas conducted by university and research teams who often being ahead of the market. Liu et al. [188] proposed a beauty E-Expert to automatically recommend the suitable makeup and synthesize it. Similarly, Si Liu et al. [189] presented a makeup transfer model, aiming to recommend the best makeup style and synthesize the cosmetics automatically [190]. Kane [191] also built various recommendation systems to help to discover new products and content, using deep learning and computer vision. Salsky [192] explored the advantages of personalisation for both consumer and the brand, what variables lead to a successful implementation, customer behaviour, and the problems related with it. Adebo [193] developed a skincare recommendation algorithm based on the information provided in the users' reviews. The proposed methodology aims to address the problem of rating prediction by combining several review aspects.

Augmented Reality (AR) can be used to help integrate virtual try-on and makeover systems with recommendation systems, allowing users to interact with services. The technology that realistically virtualises and overly an object onto a target individual is known as virtual try-on. Virtual try-on (makeover) methods have recently gained popularity. Integrating virtual try-on with recommendation systems would make it easier for customers to make purchasing selections, resulting in increased sales. Deep convolution neural networks (CNNs) and computer vision (image analysis) are noticeably adopted

[9] `https://atolla.com/`

in most state-of-the-art virtual try-on systems [21, 22, 194]. Many mobile applications have been built recently for hairstyle makeovers and try-on, such as ModiFace [10] and Hairstyle Magic Mirror [11].

2.6 Summary

An overview of facial beauty and attractiveness has been reported in this chapter by highlighting the beauty impact in psychological studies, aesthetic plastic surgery, and computer science (given in§ 2.1). Moreover, beautification methods and their influence on facial beauty were described in § 2.2. Further, § 2.3 provided a review of artificial intelligence, deep learning, components of CNNs and facial image analysis and exhibited several CNN models. The role of AI in the beauty and cosmetics industry was presented in § 2.4. Finally, existing intelligent virtual try-on and recommendation systems for beauty are described in § 2.5.

From the research work and industry applications presented in this chapter, it can be clearly noticed that computer vision and deep learning have demonstrated superior performance in many real-world applications related to the beauty and cosmetics sectors. Thus, we have been motivated to harness the CNNs with facial image analysis for developing a recommendation system (presented in Chapter 3). We also consider many challenging aspects that promote the recommendation system, including the presence of facial makeup and its effect on the developed method (presented in Chapter 4) as well as segmenting the face region that helps in developing virtual try-on hairstyle (shown in Chapter 5). The next chapter will present the framework developed by merging two types of virtual recommendations, including hairstyle and eyelashes, for beautification purposes.

[10] `https://play.google.com/store/apps/details?id=com.modiface.haircolorstudio.free`

[11] `https://play.google.com/store/apps/details?id=air.MagicMirrorFree`

Chapter 3

Integrated Multi-Model Face Shape and Eye Attribute Identification for Hairstyle and Eyelash Recommendation

In this chapter, multiple models have been developed and integrated to identify the face shape and eye attributes for hairstyle and eyelashes recommendation purposes. In § 3.1, an introduction of the eye attributes and face shapes and their definitions are given. In § 3.2 presents an overview of the existing work related to eye and face attributes and recommendation systems. The data preparation and the complete proposed framework integrating three models are described and explained in § 3.3. Next, the experimental results of the proposed system are revealed and discussed in § 3.4. Finally, this work is concluded in § 3.5.

3.1 Introduction

In daily life communication, our faces impose a significant role. Pursuing facial beauty is one of the normal human habits. Due to the increasing demand for aesthetic surgeries over the recent decades, the need for understanding and interpreting beauty has become crucial. Several studies proved that the level of facial beauty has a great effect on some social outcomes, for instance, hiring decisions and mate choices. [195, 196]. Thus, the cosmetic industry has produced various products that target to enhance different parts of the human body, including hair, skin, eye, eyebrow, and lips. Not surprisingly, research topics based on the face features have a long track record in psychology and many other scientific fields [197]. During the last decades, computer vision systems have played a major role in obtaining an image from a camera to process and analyse it in a manner similar to natural human vision system [198].

Computer vision algorithms have recently attracted increasing attention and been considered one of hottest topics due to its significant role in healthcare, industrial and commercial applications [198–202]. Facial image processing and analysis are essential techniques that help extract information from images of human faces. The extracted information such as locations of facial features such as eyes, nose, eyebrows, mouth, and lips, can play a major role in several fields, such as medical purposes, security purposes, cosmetic industry, social media applications, and recognition [199]. Several techniques have been developed to localise these parts and extract them for analysis [203]. The landmarks used in computational face analysis often resemble the anatomical soft tissue landmarks that are used by physicians [204]. Recently, advanced technologies such as artificial intelligence and machine/deep learning algorithms have helped the beauty industry in several ways, from providing statistical bases for attractiveness and helping people alert their looks to developing products which would tackle specific needs of customers [205, 206]. Furthermore, cloud computing facilities and data centre services have gained a lot of attention due to their significant role for customers' access to such products by building web-based and mobile applications [207].

3.2 Related Work

In the literature, there have been many facial attribute analysis methods presented to recognise whether a specific facial attribute is present in a given image. The main aim of developing attribute analysis methods was to design a connection between feature representations required by real-world computer vision tasks and human-understandable visual descriptions [208, 209]. Deep learning-based facial attribute analysis approaches can generally be categorised into two categories: holistic techniques which exploit the relationships among attributes to get more discriminative cues (i.e. discriminative features that can be well-classified by classifier) [210–212] and part-based approaches that initially locate the positions where attributes appear. Following the acquisition of location clues, features corresponding to certain attribute regions are retrieved and input into attribute classifiers for prediction [213, 214]. Unlike the existing facial attribute analysis methods, which focus on recognising whether a specific facial attribute is present in a given face image or not, our proposed method suggests that all concerned attributes are present but in more than one label (For example, eye setting is an eye class which is available in three labels including, wide, close, proportional)

Furthermore, many automated face shape classification systems were presented in the literature. Many of these published face classification methods consider extracting the face features manually then passing them to three classifiers for classification, including linear discriminant analysis (LDA), artificial neural networks (ANN), and support vector machines (SVM) [215, 216], k-nearest neighbours [217], and probabilistic neural networks [218]. Furthermore, Bansode et al. [219] proposed a face shape identification method based on three criteria which are region similarity, correlation coefficient and fractal dimensions. Recently, Pasupa et al. [33] presented a hybrid approach combining VGG convolutional neural network (CNN) with SVM for face shape classification. Moreover, Emmanuel [220] adopted pretrained Inception CNN for classifying face shape using features extracted automatically by CNN. The work presented by

researchers showed progress in face shape interpretation; however, existing face classification systems require more effort to achieve better performance. The aforementioned methods perform well only on images taken from subjects looking straight towards the camera and their body in a controlled position and acquired under a clear light setting.

Recently, many approaches have been developed for building fashion recommender systems [221, 222]. Conducting a deep search in the literature seeking the existing recommendation systems leads to find out two hairstyle recommendation systems [31, 33]. The system developed by Liang and Wang [31] considers many attributes such as age, gender, skin colour, occupation and customer rating for recommending hairstyle. However, the recommended haircut style might not fit the beauty experts' recommendations based on face shape attributes. Furthermore, the hairstyle recommender system presented by [33] is not general and focuses only on women.

To the best of the authors' knowledge, this proposed eyelashes and hairstyle recommendation system is the first study conducted to automatically make a recommendation of a suitable eyelash type based on computer vision techniques. This research work includes the development of system with friendly graphical user interface that can analyse eye and face attributes from an input image. Based on the detected eye and face features, the system suggests a suitable recommendation of eyelashes (for women only) and hairstyles for both men and women. The proposed framework integrates three main models: face shape classification model, gender prediction model for predicting the gender of user, and eye attribute identification model to make a decision on the input image. Machine and deep learning approaches with various facial image analysis strategies including facial landmark detection, facial alignment, and geometry measurement schemes are designed to establish and realise the developed system.

3.3 Development of Hairstyle and Eyelash Recommendation System

3.3.1 Data Preparation: Facial and Eye Attributes

A publicly available dataset, MUCT [223], is used to conduct our experiments and carry out the evaluation on the developed face shape and eye attribute identification system. The MUCT database consists of 3755 faces captured from 276 subjects with 76 manual landmarks. Facial landmarks are used to localise and describe essential facial features including, eyes, mouth, nose, eyebrows, and jawline, as shown in Figure 3.1. The database was created to provide more diversity of lighting, age and gender with an obstacle (some subjects wear glasses) than other available landmarked 2D face databases. The resolution of images is 480×640. Five webcams were used to photograph each subject. The five webcams are located in five different positions but not on the left of the subject. Each subject was photographed with two or three lighting setting producing ten or fifteen different lighting setups in all five webcams. In order to achieve diversity without too many images, not every subject was photographed with every lighting setup.

The MUCT dataset does not provide face shape and eye attribute labels. Therefore, we recruited beauty experts[1] to provide the ground truth by labelling the images in the dataset. To formulate our research problem, we initially set up the attributes required to be identified. Face shape and eye specifications/attributes required to be detected in this research work are defined as follow:

1. Eye Shape: Two attributes (Almond or Round),

2. Eye Setting: Three attributes (Wide, Close, or Proportional),

3. Eye Position/Pitch: Three attributes (Upturned, Downturned, or Straight),

[1]Experts from Ultimate Group have labelled the data. To label our data, they used the same guideline usually used to measure their customers' eye attributes and face shape before applying for makeup, hairdressing, eyelashes extension. The website of this group: `https://theultimate.group/the-ultimate-team/`

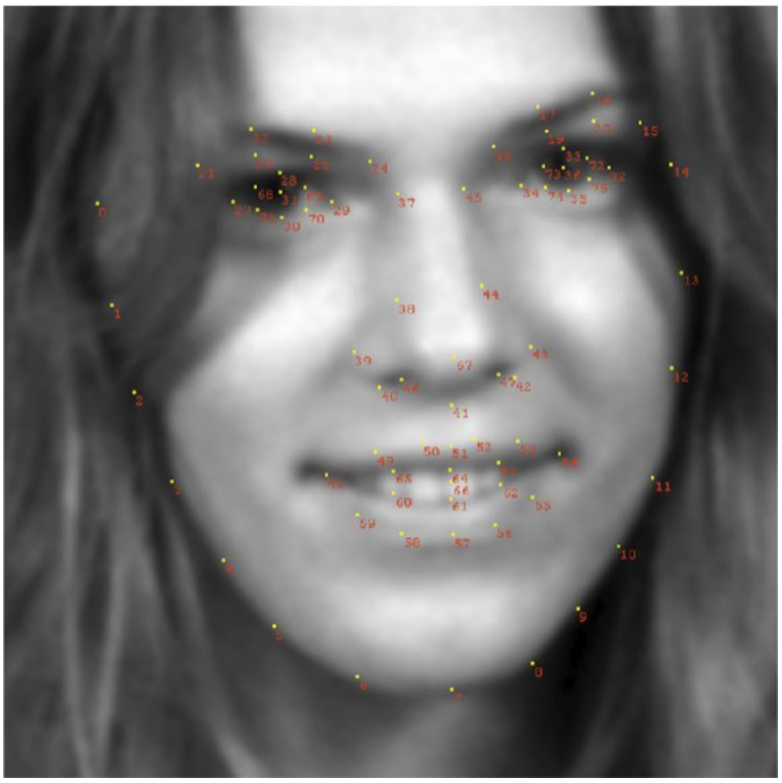


FIGURE 3.1: Seventy six facial landmarks displayed on image from MUCT data.

4. Face Shape: Five attributes (Round, Oval, Square, Rectangle (Oblong), Heart Shape).

Identifying these attributes would help automatically report eyelashes and haircut style recommendations that meet the eye and face attributes. Illustrations of these attributes are shown in Figures 3.2 and 3.3.

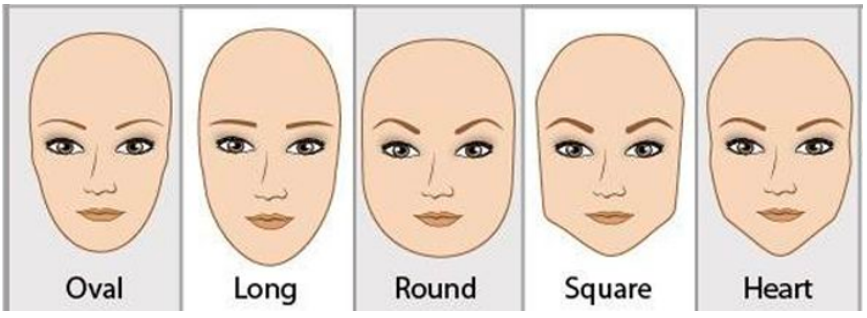
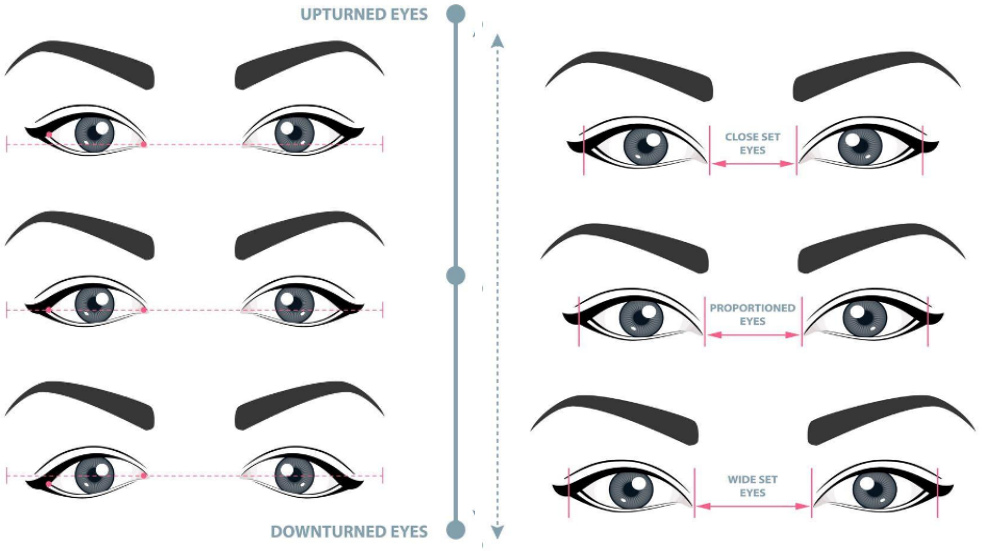


FIGURE 3.2: Categories of face shape [224].

The process of determining the facial shape and eye attributes manually by beauty experts can be carried out in several stages, such as outlining the face, measuring the width and length of the face, jaw, forehead, cheekbones, the gap between eyes, and



(a) Eye position (b) Eye setting

(c) Eye shape: almond (d) Eye shape: round

FIGURE 3.3: Eye attributes. Adapted from [225, 226]

outer and inner corner of the eye. According to the beauty experts, the face shape can be recognised as five categories, namely: oval, square, round, oblong, and heart shapes [227]. The guidelines provided by beauty experts Derrick and Brooke [227, 228] have been followed to label the face shape of 276 subjects into five classes. Furthermore, to provide the ground truth of eye attributes, the detailed guidelines and described criteria presented in [229] has been followed by a beauty expert. The distribution of data labelled manually is depicted by exploratory data analysis in Figures 3.4 and 3.5 describing the distribution of attributes over 276 subjects and per one camera, respectively. Figure 3.4 shows that the number of subjects with Upturned eye position is higher than subjects with Downturned and Straight eye positions. However, the number of images with a label of Straight position shown in Figure 3.5 is higher than Upturned and Downturned labels. This is due to the fact that subjects were photographed with either two or three lighting settings, producing either ten or fifteen different lighting setups and subsequently generating ten or fifteen images per subject. Our study does not examine

the accuracy of face shape class prediction under different camera positions. Thus, the distribution of face shape classes is not explored here.

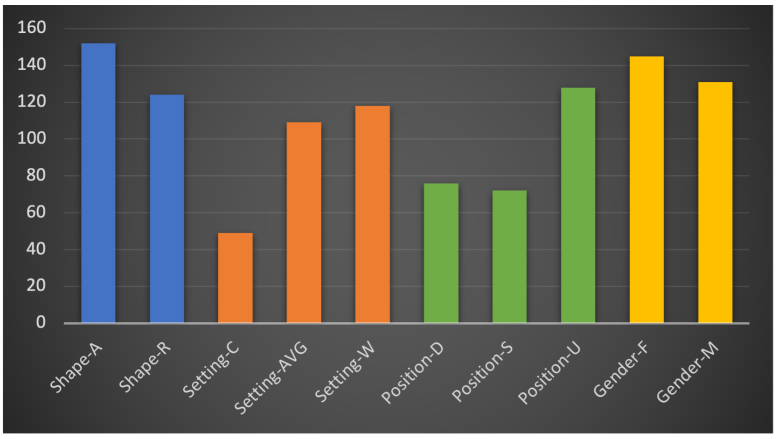


FIGURE 3.4: Distribution of attributes over 276 subject. Shape-A: Eye shape-Almond, Shape-R: Eye shape-Round, Setting-C: Eye setting-Close, Setting-AVG: Eye setting-Average, Setting-W: Eye setting-Wide, Position-D: Eye position-Downturned, Position-S: Eye position-Straight, Position-U: Eye position-Upturned, Gender-F: Gender-Female, Gender-M: Gender-Male

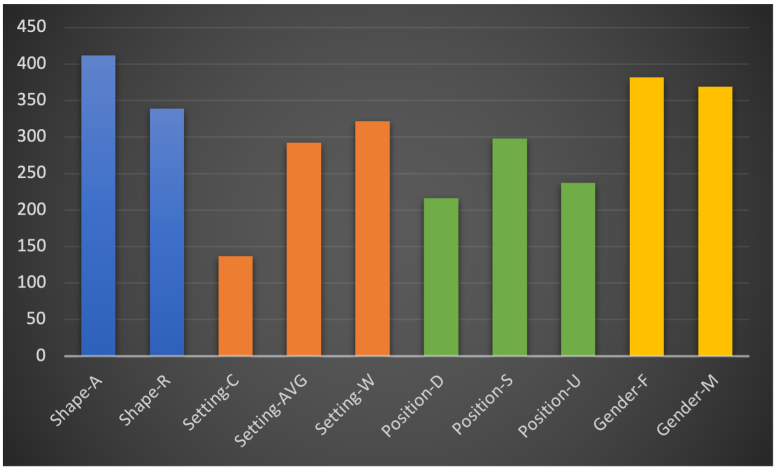


FIGURE 3.5: Distribution of attributes over 751 image (per one camera). Shape-A: Eye shape-Almond, Shape-R: Eye shape-Round, Setting-C: Eye setting-Close, Setting-AVG: Eye setting-Average, Setting-W: Eye setting-Wide, Position-D: Eye position-Downturned, Position-S: Eye position-Straight, Position-U: Eye position-Upturned, Gender-F: Gender-Female, Gender-M: Gender-Male

3.3.2 Developed Multi-Model System

To develop a personalised and automatic approach to finding the appropriate haircut and/or eyelash style for men and women, we develop a multi-model system comprising

three separate models, as shown in Figure 3.6. Each individual model is dedicated to achieving a specific task, including face shape identification, eye attribute identification, and gender identification. The diagram shows that the eye features extracted from an input image by the developed eye feature identification model are merged with identified face shape class. The gender of the individual in the input image is detected by the gender identification model. Thus, the face and eye attributes are combined with identified gender information. If the gender in the input image is a male, then a hairstyle recommendation for men is given; otherwise, both hairstyle and eyelashes recommendations are reported. The description of the three models are explained as follows:

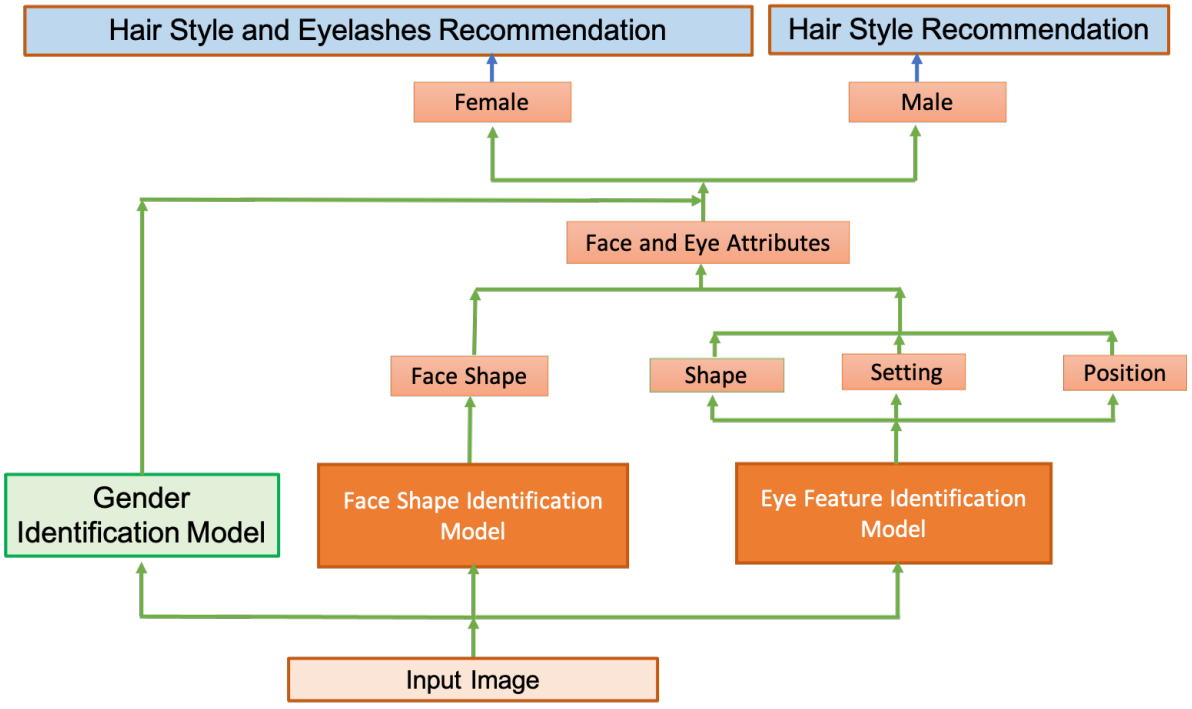


FIGURE 3.6: Block diagram of the proposed method including three models: face shape identification model, eye feature identification model, gender identification model.

Face Shape Identification Model

To identify the face shape, we developed a model [230] that was designed by merging hand-crafted features with automatically learned features. MUCT data has been randomly split into 3240 images for training the developed model, and 500 images were

retained for testing. The developed model (shown in Figure 3.7) achieves the face classification as follows: (1) detect the face region and crop it using a model trained on the histogram of oriented gradients (HOG) features with Support Vector Machine (SVM) model as classifier [231], (2) the detected face is aligned using the detected face landmarks (68 landmarks) by the ensemble of regression tree method (ERT) [232], and (3) Finally, the aligned images are used for training and evaluating Inception V3 convolutional neural network [40] along with hand-engineered HOG features and landmarks to classify the face into one of five classes. Figure 3.7 shows the block diagram of the proposed system, including three main phases of the face shape classification model. These phases are explained as follows:

1. **Face Detection and Cropping**: In this stage of the face shape classification framework, the face is detected and cropped from an image as shown in Figure 3.8. The face detection is carried out using a widely applied face detection model trained on the histogram of oriented gradients (HOG) features with a linear classifier; Support Vector Machine (SVM) model. HOG based method [231] is typically more accurate than Haar cascades [233] with fewer false positives and requires fewer parameters to tune at test time. The dataset used for training consists of 2825 images which are obtained from the LFW dataset [234].

   HOG algorithm iterates on every pixel of an input image; in each pixel, it checks all the pixels around it and figures out how dark the current pixel is compared with the adjacent pixels that are surrounding it. It then measures in which direction the pixel is getting darker, and this direction can be called gradient. This procedure is repeated on every single pixel. The image is then broken down into 16×16 squares, where the number of gradients point in each major direction is counted. That square is typically replaced with the direction which was the strongest. Hence, the image is converted into a representation that captures the basic structure of a face. These features represented by face representations are adopted to learn a linear SVM classifier, targeting to identify the region of

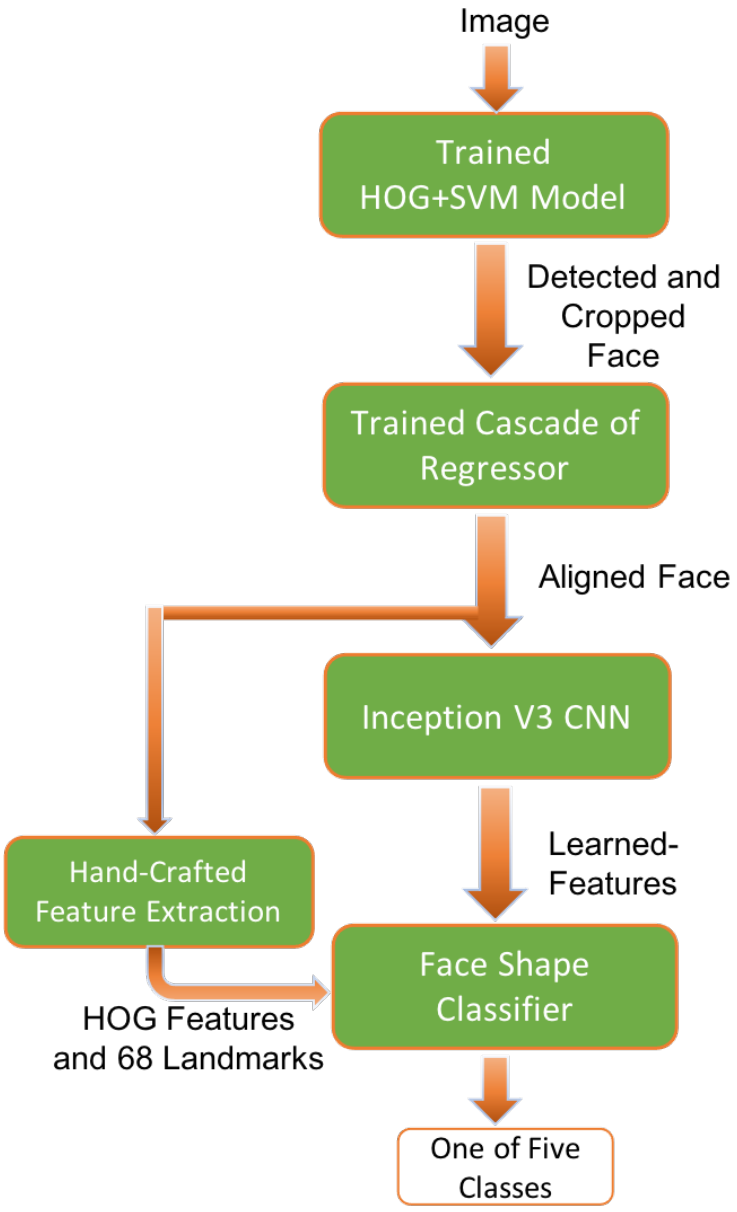
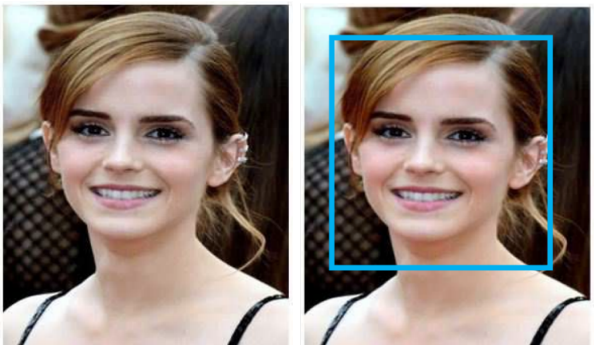


FIGURE 3.7: The developed face shape classification system.

faces in a test image. According to the study conducted by the authors of [235], linear SVMs trained on HOG features demonstrated a superior classification performance in many computer vision tasks.

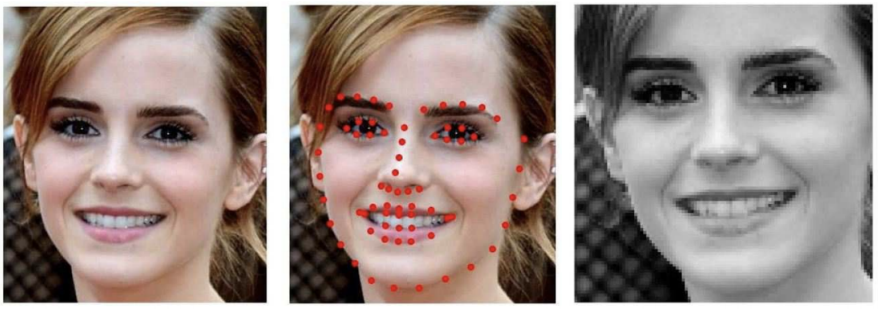
2. **Landmark Detection and Face Alignment**: In this stage, the cropped face is aligned by firstly detecting the face landmarks (68 landmarks) by the ensemble of regression tree method (ERT) [232] and then aligning the face using the detected landmarks as shown in Figure 3.9. As the facial alignment can be seen as a form of



(a) (b)

FIGURE 3.8: (a) Original image (b) Face region localised to be cropped.

data normalisation, faces should be centred and scaled over an entire dataset such that their sizes are almost comparable. The eyes should also lie on a horizontal line. However, the resulted aligned faces across the dataset were not optimal due to the trade-off among those three alignment objectives as shown in Figure 3.9.



(a) (b) (c)

FIGURE 3.9: Emma Watson's face is correctly identified as an oval face shape. (a) Cropped face (b) Detected 68 landmarks (c) Aligned face.

The face alignment is similar to feature vectors normalisation done via scaling to the unit norm or zero centrings prior to learning a machine learning algorithm. A feature vector is an n-dimensional vector of numerical characteristics representing the samples under investigation in machine learning. Because numerical representations of data enable easier processing and statistical analysis, machine learning models accept the numerical representation as input. The normalisation of feature vectors is a scaling method in which data are rescaled to be limited to a certain range. During the learning phase, the normalisation prevents variables

with large numerical values from dominating cost functions. It is common to align the faces in the dataset before training a face recogniser.

The face landmark detection algorithm, subsequently face alignment, adopted in this work is an implementation of the ensemble of regression trees (ERT) presented in 2014 by Kazemi and Sullivan [232]. Our work aims to exploit a landmark detection technique that can achieve feasible application in a real-time environment by reading images from a real camera. Thus, the ERT technique, which utilises simple and fast features (pixel intensities differences) to estimate the landmark positions directly, is adopted. This method guarantees the balance between the model speed and prediction accuracy. The cascade of regressors is adapted to refine these estimated positions with an iterative process. At each iteration, the regressors produce a new estimate from the previous one, trying to reduce the alignment error of the estimated points. To obtain a normalised translation, rotation, and scale representation of the face, the alignment procedure relies on the facial landmarks, particularly the eye regions.

The detected facial landmarks coordinates (68 landmarks) help to align the faces. The alignment helps to make all faces in the dataset centred in the image, scaled such that the size of the faces is almost identical, and rotated such that the eyes located on a horizontal line (the face is rotated such that the eyes are located on a common axis with a constant y-value). In image coordinate systems, (r, c) represents the pixel value. Where x represents the column axis values while y refers to row axis values. Ensemble of regression trees (ERT) [232] for landmark detection is pre-trained on the 300W dataset [236] which contains more than 4000 images in the wild, i.e. captured under an uncontrolled and unconstrained environment. The MUCT dataset is provided with the ground truth of 76 landmarks which are manually annotated. However, the developers of the Dlib library[2] implemented the facial landmarks prediction scheme based on the ensemble of regression trees

[2] `https://pypi.org/project/dlib/`

approach (ERT) [237]. This ERT model is trained on data with 68 landmarks. The detection of the landmarks in our developed framework is accomplished based on the pre-trained ERT model by exploiting a transfer learning scheme. Thus, the remaining eight landmarks have not been used.

3. **Classification by Inception CNN**: In the final stage, the aligned images are used for training and evaluating the model. The training images are fed to the Inception V3 convolutional neural network [40] along with HOG features and landmarks to classify face shape into five classes as shown in Figure 3.10.

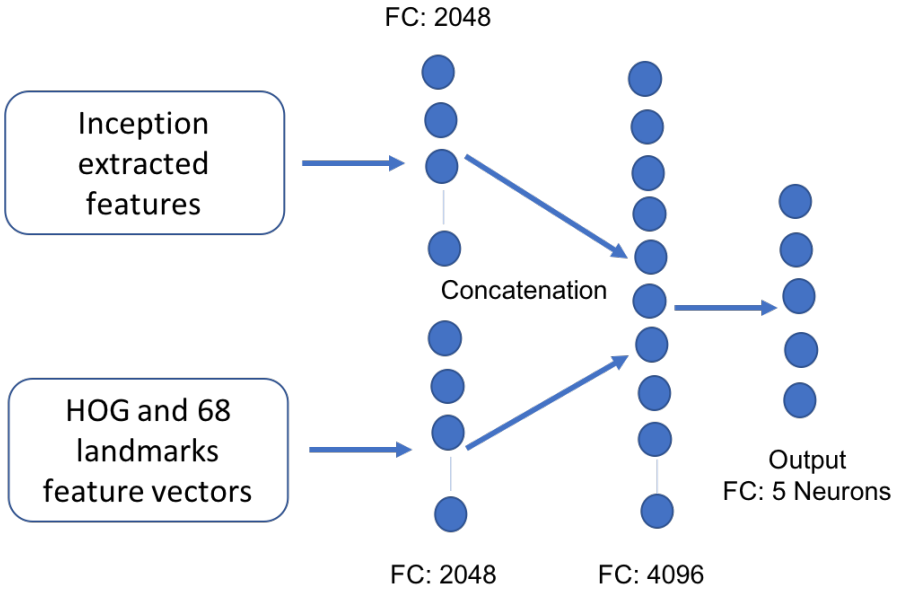


FIGURE 3.10: Block diagram shows merging Inception with hand-crafted features in the face shape classifier.

Inception v3 CNN architecture [40], shown in Figure 3.11, is adopted and adapted to identify the shape of the face. We used the weights of the pre-trained Inception v3 model [238] and fine-tuned it. Thus, the scheme of transfer learning is adopted rather than training net from scratch. The original architecture of the Inception v3 model consists of 1000 output neurons on the top dense layer. This top layer of 1000 neurons is removed and changed with dense layer of 5 neurons (number of face shape classes). Weights of the pre-trained Inception layer are then tuned and the new classifier with five neurons is trained from scratch. Initially, 3255 images

are randomly chosen for training, and the rest 500 images were used for independent testing. The pre-trained Inception network is fine-tuned for 200 epochs on images resized into 299×299 pixels. To learn the network, the SGD with the momentum algorithm was adopted for optimisation with a learning rate of 0.016 and momentum parameter 0.95.

The HOG and 68 face landmarks - extracted from the same training images used to train Inception CNN - are passed into a fully connected layer (FC) that comprises 2048 neurons. HOG features and 68 landmarks are used as hand-crafted features, as shown in Figure 3.7. In the literature, hand-crafted, hand-engineered, and hand-designed terminologies are all used interchangeably to describe the manually extracted features. Integrating these features with automatically learned features from the Inception3 model helps to improve the performance of face class detection. The output of the FC layer in Inception, which has 2048 nodes, are concatenated with the FC layer of hand-crafted features producing a layer of 4096 neurons. This fully connected layer is followed by a second fully connected layer of five neurons (five face shape labels) and a softmax layer for classification purposes.

Eye Attribute Identification

The geometry of face/eye measurements based on the coordinates of detected eye landmarks are determined to develop this model. Thus, the eye specifications (shape, setting, position) are obtained. First, the face region is detected using the same face detection method [231] described in Face Detection and Cropping above. Following that, the landmarks of the eye region are localised using the ERT technique [232]. Here, we customised the landmark predictor to return just the locations of the eyes using the Dlib [3] library with Python. Following the customised selection of landmarks conducted in

[3] `https://pypi.org/project/dlib/`

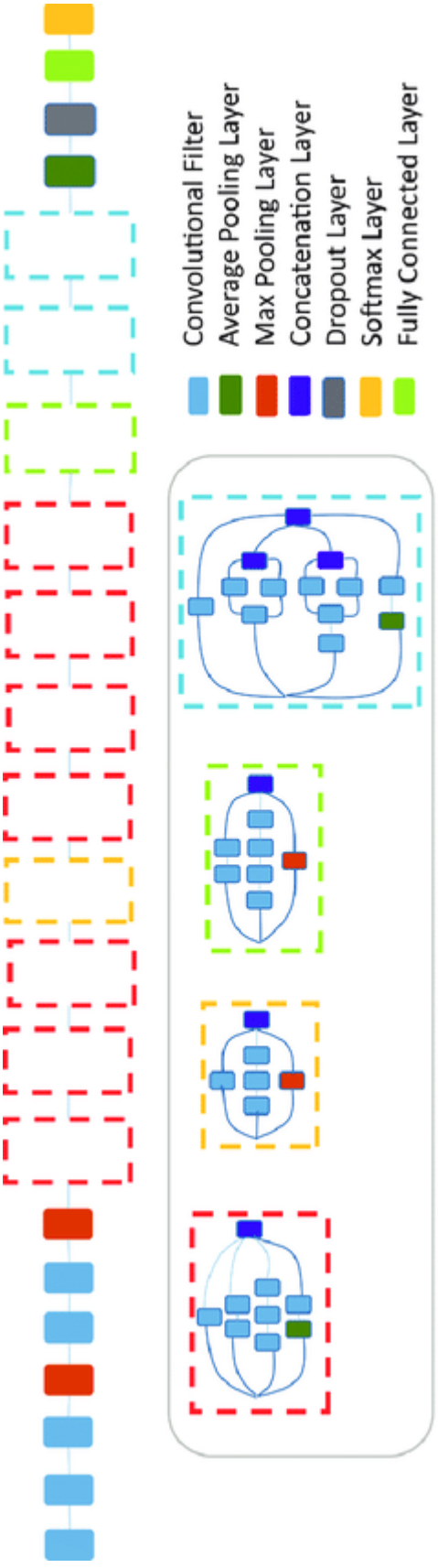


FIGURE 3.11: Inception v3 CNN architecture [40].

[237], only six landmarks per eye -12 in two eyes- are taken into consideration for predicting eye attributes, as shown in Figure 3.12. Training the facial landmark detection model on only six landmarks could help to retain the model size and prediction speed minimal.

Given 12 eye landmark points $P_n = P_0, P_1, P_2, ..., P_n$ where $n = 0, 1, 2, ..., 11$, the eye specifications can be calculated using geometrical eye measurements as follows:

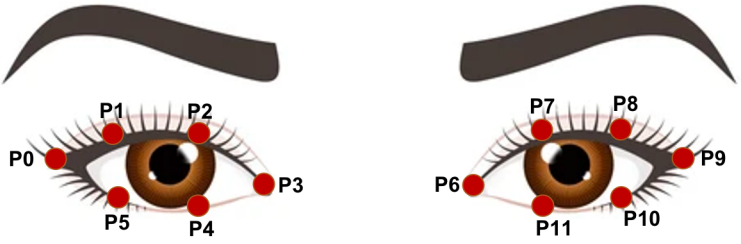


FIGURE 3.12: The six landmarks taken into consideration to predict the eye attributes using geometrical measurements.

- **Eye Setting:** The setting of eye Se can be determined by calculating the ratio of the distance between two eyes L_1 to the eye width W_{eye}. The width of the eye W_{eye} can be found by measuring the distance from the start and the end of an individual eye. The distance between two facial landmark points d can be obtained using the Euclidean distance.

$$d(X, Y) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{3.1}$$

$$\begin{aligned} W_{eye1} &= d(P_0, P_3) \\ W_{eye2} &= d(P_6, P_9) \end{aligned} \tag{3.2}$$

$$L_2 = \frac{W_{eye1} + W_{eye2}}{2} \tag{3.3}$$

$$L_1 = d(P_3, P_6) \tag{3.4}$$

$$Se = \frac{L_1}{L_2} \tag{3.5}$$

The threshold values for category classification are empirically chosen and estimated as follow:

$$Class\ of\ eye\ setting = \begin{cases} Close, & \text{if } Se < 1.57 \\ Proportional, & \text{if } 1.57 \leq Se \leq 1.6 \\ Wide, & \text{Otherwise} \end{cases} \tag{3.6}$$

From Equation(3.6), the threshold value used to determine (Close) label is calculated by finding the median of subjects' eye setting measurements. Thus, the eye setting value of a subject that is less than the median is labelled as (Close). Likewise, the middle number between the median and the highest number (the maximum) is harnessed to differentiate between (Proportional) and (Wide) labels. For example, the eye setting value of a subject that is greater than the median and less than the middle number is labelled as (Proportional); otherwise, it is labelled as (Wide).

- **Eye Shape:** The criterion for discriminating round eye from almond eyes is that, for round eyes, the white sclera area below the iris is visible [229] as illustrated in Figure 3.13. Our proposed criteria exploits eye-aspect-ratio rule to determine the eye shape as follows:

FIGURE 3.13: Round vs. almond shape.

$$Sh = \frac{d(P_1, P_5) + d(P_2, P_4)}{2d(P_0, P_3)} \tag{3.7}$$

The threshold values for category classification are empirically chosen and estimated as follow:

$$Class\ of\ eye\ shape = \begin{cases} Almond, & \text{if } Sh \leq 0.28 \\ Round, & \text{Otherwise} \end{cases} \tag{3.8}$$

From Equation(3.8), the threshold value used to differentiate between (Almond) and (Round) labels is calculated by finding the median of subjects' eye shape measurements. Thus, the eye shape value that is less than the median is labelled as (Almond); otherwise, it is labelled as (Round).

- **Eye Position:** Eye position categorised into up-turned, down-turned and straight can be determined using geometrical calculations by measuring the slope angle (degree). The slope angle is defined as the angle measured between a horizontal line and a slanted line (the line drawn between landmark points (P_3, P_6) and (P_0, P_3) for eye_1, (P_6, P_9) for eye_2). The equation of eye position identification can be defined as follows:

$$Po = \frac{\theta_{eye1} + \theta_{eye2}}{2} \tag{3.9}$$

where θ represents the angle between the line drawn between landmark points (P_3, P_6) and (P_0, P_3) for eye_1, (P_6, P_9) for eye_2 and each point is represented as (x, y) .

$$\theta_{eye} = tan^{-1} \frac{m_1 - m_2}{1 + m_1 m_2} \tag{3.10}$$

where m_1 and m_2 are the slope of each line formulated to calculate the angle θ which is defined as follows:

$$m = \frac{y_2 - y_1}{x_2 - x_1} \tag{3.11}$$

The threshold values for category classification are empirically chosen and estimated as follow:

$$Class\ of\ eye\ position = \begin{cases} Up - turned, & \text{if } Po > 6.9 \\ Straight, & \text{if } 0 \leq Po \leq 6.9 \\ Down - turned, & \text{Otherwise} \end{cases} \tag{3.12}$$

From Equation(3.12), the threshold value used to determine (Up-turned) label is calculated by finding the median of subjects' eye position measurements. Thus, the eye position value that is greater than the median is labelled as (Up-turned). Likewise, the value between zero and the median is labelled as (Straight). Any value of less than zero (negative) is identified as (Down-turned).

Gender Identification

Gender identification is a pre-requisite step in our developed framework. If the gender in the input image is male, then the recommendation engine reports only the suitable hairstyle for men; otherwise, the system recommends hairstyle and eyelash extension for women. The deep learning neural network approach developed in [166] has been adopted to achieve this task. "The network comprises of only three convolutional layers and two fully-connected layers with a small number of neurons. The three subsequent convolutional layers are then defined as follows" [166]: i) "96 filters of size $3 \times 7 \times 7$ pixels are applied to the input in the first convolutional layer, followed by a rectified linear operator (ReLU), a max pooling layer taking the maximal value of 3×3 regions

with two-pixel strides and a local response normalisation layer", ii) "the $96 \times 28 \times 28$ output of the previous layer is then processed by the second convolutional layer, containing 256 filters of size $96 \times 5 \times 5$ pixels, iii) "the third and last convolutional layer operates on the $256 \times 14 \times 14$ blob by applying a set of 384 filters of size $256 \times 3 \times 3$ pixels, followed by ReLU and a max pooling layer" [166]. The block diagram of the CNN architecture adopted for gender identification is shown in Figure 3.14.

Dataset of the Adience Benchmark [166] is used to train the gender identification model. The Adience Benchmark is a collection of unfiltered face images collected from Flickr. It contains 26,580 images of 2,284 unique subjects that are unconstrained, blur, with occlusions, have variety of poses, and different expressions. Scaling has been applied on dataset to size of 256×256 pixels and then cropping to size of 227×227. All MUCT images composed of 52.5% females were used for testing. The training of model has been conducted from scratch. The weights have been initialised using a zero mean Gaussian along with standard deviation of 0.01. Another two layers of dropout with a dropping probability of 0.5 were used in the network architecture to mitigate the effect of overfitting [166]. Data augmentation by randomly mirroring the images in each forward-backwards training pass was also applied. The learning process was conducted using stochastic gradient descent, set with a mini-batch of size 50 images. The initial learning rate is 10^{-3}, reduced to 10^{-4} after 10K iterations.

Rule-Based Hairstyle and Eyelash Recommendation System

The right shape of artificial eyelash extensions can be used as a way to correct and balance facial symmetry and open the eyes wider. Eyelash Extensions can produce an illusion of various eye effects giving an appearance of makeup on the eye without applying it. It also can create an appearance of a lifted eye, or even a more feline look being produced [239]. For the best hairstyle, instead of choosing the latest fashion, it should be attempted instead to pick a style that fits face shape. The correct haircut that

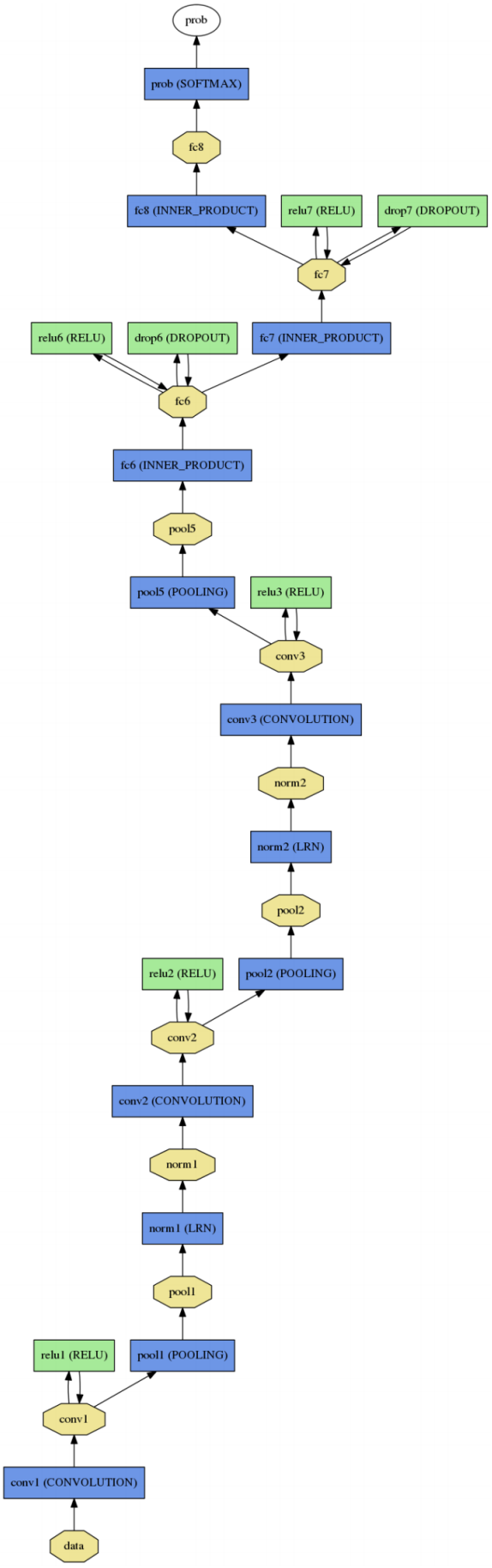


FIGURE 3.14: CNN architecture adopted for gender identification [166].

fits the face shape will frame and balance it expertly while showing ”the best features for a flattering and complementary appearance” [91, 92].

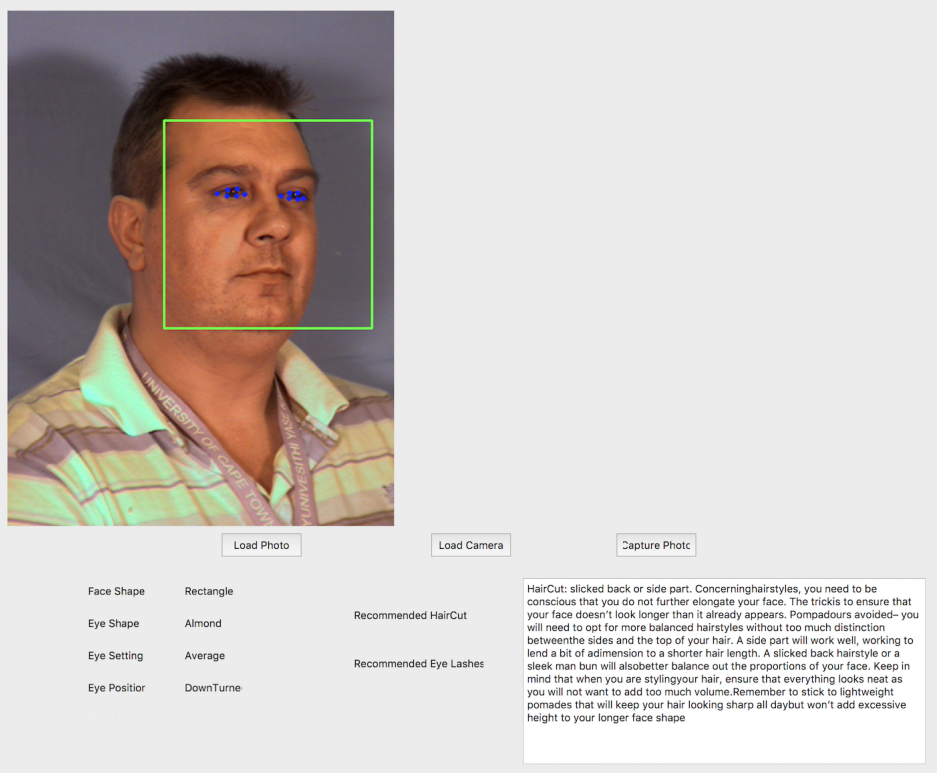
The guidelines from specialised websites [91, 92, 239] are followed to implement the rules of the recommendation engine for hairstyle and eyelashes in our proposed system. The rule-based inference engine system focuses on knowledge representation and applies rules to obtain new information by existing knowledge. When the data (eye attributes, face shape, gender) matches rule conditions, the inference engine can interpret the facts in the knowledge base by applying logical rules and deduce the suitable recommendation using if-then operations. For instance,

IF (Gender is Male) **AND** (Face Shape is Round) **THEN** Haircut Recommendation is Side part, French crop, pompadour.

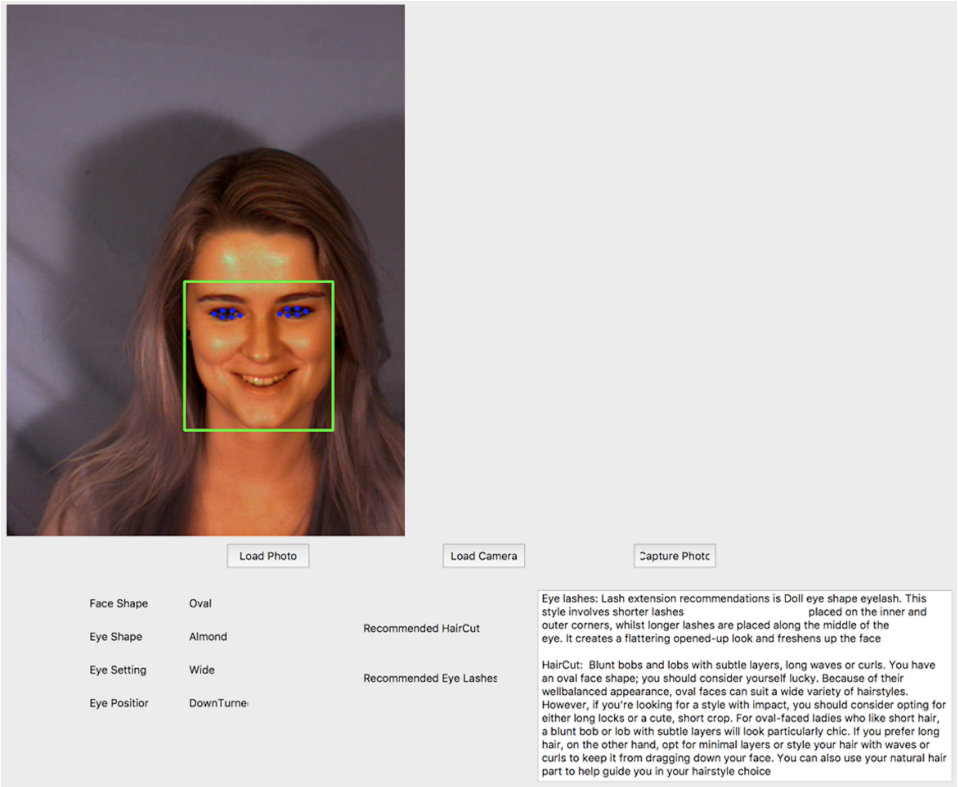
Appendix A describes all the recommendations used in implementing the rule-based inference engine system for hairstyle and eyelashes. Figure 3.15 presents the graphical user interface (GUI) of the developed system showing an input image (man and woman) with the automatically detected landmarks, automatically extracted facial/eye attributes and the recommended haircut and eyelash extension recommendation system. More Examples of the outcome of the recommendation system are shown in Appendix A.

3.4 Results and Discussion

A new approach based on merging three models in one framework has been proposed for the simultaneous detection of the face shape, eye attributes, and gender from web-cam images. The designed framework achieves the detection by extracting hand-crafted features, learning features automatically, and exploiting the face geometry measurements. For assessing the performance of the presented method, several measurements, including accuracy (Acc.), sensitivity (Sn.), specificity (Sp.), Matthews Correlation Coefficient (MCC), precision (Pr), and F_1 score, have been calculated using TP: True Positive, TN: True Negative, FP: False Positive, FN: False Negative.



(a)



(b)

FIGURE 3.15: (a) GUI displays facial and eye attributes predication and recommendation showing man example (b) GUI displays facial and eye attributes predication and recommendation showing woman example. Recommendation text displayed on this GUI is taken from [91, 92]

Unlike binary class confusion matrix described in Section 1.6, there are five face shape classes producing a multi-class task. To calculate the TP, TN, FP, and FN in a multi-class task, the problem has been dealt with as multiple binary classification tasks as multiple binary classification tasks, as shown in Figure 3.16.

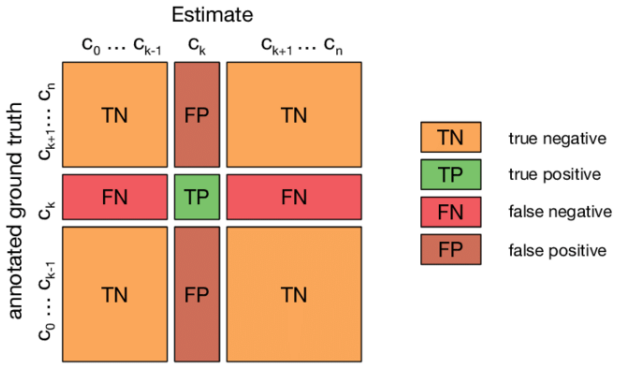


FIGURE 3.16: Confusion matrix of multi-class classification.

Suppose that C_1, C_2, C_3, C_4, and C_5 represent the five face shape classes. TP of C_1 is all C_1 images that are classified as C_1, TN of C_1 is all non C_1 images that are not classified as C_1, FP of C_1 is all non C_1 images that are classified as C_1, and FN of C_1 is all C_1 images that are not classified as C_1. The same procedure is repeated to find these four terms for C_2, C_3, C_4, and C_5. The evaluation metrics described above are computed individually per class than the average performance of all five classes is calculated.

Initially, images belonging to one subject (shown in Figure 3.17) were excluded from data because of failure detecting the eye landmarks of this subject due to extreme eye closing and presence of obstacle (eyeglasses). Thus, the number of subjects that will be considered for evaluation is 275.

Table 3.1 depicts the confusion matrix of face shape classes, reporting an identification accuracy of 85.6% on 500 test images (100 image per each class). We carefully selected only 500 images for testing phase targeting to provide images in the test set with a minimum overlapping with the images of subjects in the training set. This encourages a reliable evaluation scheme. To compare our face shape classification model with the existing face shape classification methods, Table 3.2 presents a comparison in terms of

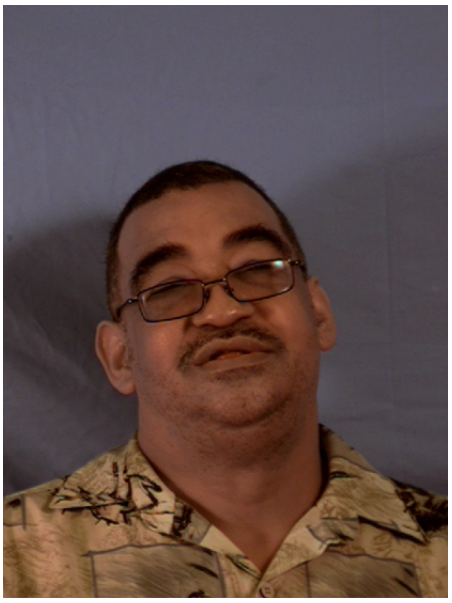


FIGURE 3.17: Subject has been excluded because of failure detecting the eye landmarks of due to extreme eye closing and presence of obstacle (eyeglasses).

classification accuracy. It can be noticed that our developed face shape classification system [230] outperforms the other methods in the literature.

TABLE 3.1: Confusion matrix representing the five classes of face shape.

		Predicted				
Desired	Oval	**59**	11	16	9	5
	Square	0	**99**	0	1	0
	Rectangle	2	0	**92**	1	5
	Round	1	1	4	**91**	3
	bHeart	6	0	7	0	**87**

The authors of [220], who proposed a system based on passing the images directly into the off-the-shelf Inception CNN for feature extraction and classification, reported an accuracy of 84.4%. However, the authors in their study tested their method on the whole data, including data used to train the model, resulting in overfitting. Furthermore, testing on the data used for training is not an indicator of system performance and typically is considered a wrong evaluation scheme. Contrary, in our evaluation scheme, the data is initially split into a training set for training and a test set for testing. The other methods described in Table 3.2 have been evaluated on private datasets.

TABLE 3.2: The performance of proposed face shape classification methodology comparing to the existing methods in literature.

Method	Accuracy
Inception V3 [220]	84.4%
Region Similarity, Correlation and Fractal Dimensions [219]	80%
Active Appearance Model (AAM), segmentation, and SVM [215]	72%
Hybrid approach VGG and SVM [33]	70.3%
3D face data and SVM [216]	73.68%
Geometric features [217]	80%
Probability Neural Network and Invariant Moments [218]	80%
Our model	**85.6%**

For the gender identification, the gender detection model has achieved prediction performance of 0.9018, 0.8923, 0.9103, 0.8992, and 0.8958 in terms of accuracy, sensitivity, specificity, precision, and f_1 score measures; respectively, on 275 subjects. Likewise, the gender detection model achieved 0.9064, 0.8989, 0.9136, 0.9088, and 0.9038; respectively, using the same metrics in terms of 748 image per camera. Table 3.3 presents the confusion matrix obtained from gender identification system. The first part of Table 3.3 shows the gender prediction in terms of the number of subjects, whereas the second part explores the prediction in terms of images per camera. For gender classification, performance evaluation is reported in terms of subject and only one camera position as the detected gender class is the same in all camera positions. For instance, if the gender is identified correctly as male in camera A, it is also correctly detected as male in B, C, D, and E cameras. If the detected gender is misclassified in Camera A, it is also misclassified by the remaining cameras. In terms of comparison with the existing methods, authors in [166] achieved an accuracy of 0.868 on the Adience dataset [166]. On the same dataset, Duan et al. [240] reported accuracy of 0.778 using convolutional neural network (CNN) and extreme learning machine (ELM) method. On Pascal VOC [241] dataset, De et al. [242] obtained accuracy of 0.8739 based on a deep learning approach.

TABLE 3.3: Confusion matrix representing the gender identification. The first part of matrix represents the identification in terms of number of subjects; whereas the second part explores the prediction in terms of images per one camera. F: Female, M: Male, Des.: Desired.

		Predicted			
Des.	F	**132**	13	**349**	33
	M	14	**116**	37	**329**

There are five webcams used to photograph each subject. In the following experiments, we name the webcams with five letters (A-E) including, Cam A, Cam B, Cam C, Cam D, and Cam E. Symbols X and Y (used in Cam X vs. Cam Y) refer to any individual webcam from the five cameras. Our developed system has provided promising results on eye attributes identification. Tables 3.4, 3.5, and 3.6, respectively, reports eye setting attributes detection performance using various evaluation metrics, states the obtained results from the proposed system by comparing the eye setting attribute prediction performance in one camera with ground truth using confusion matrices, and compares the prediction performance in each camera with another camera using confusion matrices.

TABLE 3.4: Performance evaluation of proposed system for eye setting attributes detection. Acc.: accuracy, Sn.: Sensitivity, Sp.: Specificity, Pr.: Precision, F_1: F_1 score, MCC: Matthews Correlation Coefficient, GT: Ground Truth, Cam: Camera.

Webcam Setting	Acc.	Sn.	Sp.	Pr.	F_1	MCC
Cam A vs. GT	0.7086	0.6900	0.8503	0.7236	0.7002	0.5563
Cam B vs. GT	0.6885	0.6694	0.8425	0.7083	0.6767	0.5305
Cam C vs. GT	0.7273	0.7129	0.8625	0.7540	0.7166	0.5958
Cam D vs. GT	0.6845	0.6655	0.8389	0.7024	0.6753	0.5218
Cam E vs. GT	0.7019	0.6849	0.8453	0.7150	0.6939	0.5459
Average (all Cams)	**0.7021**	**0.6845**	**0.8479**	**0.7207**	**0.6926**	**0.5501**
Cam A vs. B	0.6939	0.7009	0.8446	0.6871	0.6927	0.5381
Cam A vs. C	0.6979	0.7142	0.8473	0.6932	0.7003	0.5500
Cam A vs. D	0.7019	0.7033	0.8477	0.6974	0.7001	0.5479
Cam A vs. E	0.7166	0.7192	0.8559	0.7195	0.7187	0.5740
Cam B vs. C	0.6885	0.6892	0.8438	0.6843	0.6862	0.5300
Cam B vs. D	0.7112	0.7037	0.8538	0.7104	0.7066	0.5611
Cam B vs. E	0.6791	0.6761	0.8376	0.6909	0.6819	0.5200
Cam C vs. D	0.6751	0.6691	0.8356	0.6811	0.6732	0.5112
Cam C vs. E	0.6872	0.6840	0.8418	0.7038	0.6910	0.5349
Cam D vs. E	0.6832	0.6845	0.8387	0.6914	0.6871	0.5255

TABLE 3.5: Confusion matrices depict eye setting attributes prediction in each webcam versus the ground truth.

Predicted		Desired		
		Close	Average	Wide
Cam A	Close	111	72	1
	Average	23	162	61
	Wide	3	58	257
Cam B	Close	112	94	3
	Average	19	152	65
	Wide	6	46	251
Cam C	Close	122	95	7
	Average	13	168	58
	Wide	2	29	254
Cam D	Close	110	82	2
	Average	21	152	67
	Wide	6	58	250
Cam E	Close	108	75	3
	Average	25	171	70
	Wide	4	46	246
All Cameras	Close	563	418	16
	Average	101	805	321
	Wide	21	237	1258

Likewise, Tables 3.7, 3.8, and 3.9, respectively, reports eye position classes detection performance using various evaluation metrics, states the obtained results from the proposed system by comparing the eye position classes prediction performance in one camera with ground truth using confusion matrices, and compares the prediction performance in each camera with another camera using confusion matrices, .

Similar to the evaluation strategies mentioned earlier for eye setting and position attributes, Tables 3.10, 3.11, and 3.12, present eye shape labels prediction performance.

From the reported results of eye attributes presented in the above Tables, it can be noticed that the model achieves the highest performance on eye shape detection, reporting accuracy of 0.7535, sensitivity of 0.7168, specificity of 0.7861, precision of 0.7464, F1 score of 0.7311 and MCC of 0.5043 whilst lowest results were obtained from eye position class identification giving 0.6869, 0.7075, 0.8378, 0.6917, 0.6958, and 0.5370 using the same evaluation metrics, respectively. Likewise, the model reveals the results of 0.7021, 0.6845, 0.8479, 0.7207, 0.6926, and 0.5501 on eye setting detection in terms

TABLE 3.6: Confusion matrices depict eye setting attributes prediction in each individual camera versus another camera.

Desired Cam X VS. Cam Y		Predicted Cam B		
		Close	Average	Wide
Cam A	Close	154	29	1
	Average	53	128	65
	Wide	2	79	237
		Predicted Cam C		
Cam A	Close	165	19	0
	Average	50	134	62
	Wide	9	86	223
		Predicted Cam D		
Cam A	Close	147	37	0
	Average	44	133	69
	Wide	3	70	245
		Predicted Cam E		
Cam A	Close	145	38	1
	Average	38	152	56
	Wide	3	76	239
		Predicted Cam C		
Cam B	Close	169	40	0
	Average	47	125	64
	Wide	8	74	221
		Predicted Cam D		
Cam B	Close	156	49	4
	Average	38	132	66
	Wide	0	59	244
		Predicted Cam E		
Cam B	Close	153	54	2
	Average	31	133	72
	Wide	2	79	222
		Predicted Cam D		
Cam C	Close	160	57	7
	Average	33	122	84
	Wide	1	61	223
		Predicted Cam E		
Cam C	Close	161	54	9
	Average	25	140	74
	Wide	0	72	213
		Predicted Cam E		
Cam D	Close	148	44	2
	Average	35	137	68
	Wide	3	85	226

TABLE 3.7: Performance evaluation of proposed system for eye position attributes detection. Acc.: accuracy, Sn.: Sensitivity, Sp.: Specificity, Pr.: Precision, F_1: F_1 score, MCC: Matthews Correlation Coefficient, GT: Ground Truth.

Webcam Setting	Acc.	Sn.	Sp.	Pr.	F_1	MCC
Cam A vs. GT	0.7005	0.7172	0.8453	0.7072	0.7093	0.5574
Cam B vs. GT	0.6778	0.7025	0.8327	0.6808	0.6867	0.5235
Cam C vs . GT	0.6631	0.6809	0.8261	0.6704	0.6724	0.5015
Cam D vs. GT	0.6898	0.7089	0.8395	0.6957	0.6986	0.5412
Cam E vs. GT	0.7032	0.7281	0.8456	0.7044	0.7121	0.5614
Average (all Cams)	**0.6869**	**0.7075**	**0.8378**	**0.6917**	**0.6958**	**0.5370**
Cam A vs. B	0.6123	0.6169	0.7961	0.6308	0.6230	0.4196
Cam A vs. C	0.6203	0.6312	0.8020	0.6323	0.6318	0.4338
Cam A vs. D	0.5802	0.5881	0.7804	0.5950	0.5913	0.3716
Cam A vs. E	0.6297	0.6363	0.8053	0.6474	0.6412	0.4469
Cam B vs. C	0.5668	0.5829	0.7716	0.5707	0.5760	0.3489
Cam B vs. D	0.5655	0.5780	0.7693	0.5715	0.5746	0.3443
Cam B vs. E	0.6003	0.6099	0.7869	0.6074	0.6086	0.3956
Cam C vs. D	0.5548	0.5638	0.7670	0.5698	0.5666	0.3335
Cam C vs. E	0.5922	0.5993	0.7859	0.6090	0.6035	0.3896
Cam D vs. E	0.5508	0.5612	0.7621	0.5651	0.5630	0.3250

TABLE 3.8: Confusion matrices depict eye position attributes prediction in each webcam versus the ground truth.

Predicted		Desired		
		Downturned	Straight	Upturned
Cam A	Downturned	161	78	1
	Straight	52	192	65
	Upturned	0	28	171
Cam B	Downturned	156	70	0
	Straight	56	198	84
	Upturned	1	30	153
Cam C	Downturned	155	85	2
	Straight	56	180	74
	Upturned	2	33	161
Cam D	Downturned	165	68	0
	Straight	48	195	81
	Upturned	0	35	156
Cam E	Downturned	149	75	0
	Straight	64	205	65
	Upturned	0	18	172
All Cameras	Downturned	786	376	3
	Straight	276	970	369
	Upturned	3	144	813

TABLE 3.9: Confusion matrices depict eye position attributes prediction in each individual camera versus another camera.

Desired Cam X VS. Cam Y		Predicted Cam B		
		Downturned	Straight	Upturned
Cam A	Downturned	152	87	1
	Straight	74	179	56
	Upturned	0	72	127
		Predicted Cam C		
Cam A	Downturned	158	80	2
	Straight	83	169	57
	Upturned	1	61	137
		Predicted Cam D		
Cam A	Downturned	152	87	1
	Straight	81	160	68
	Upturned	0	77	122
		Predicted Cam E		
Cam A	Downturned	152	88	0
	Straight	72	183	54
	Upturned	0	63	136
		Predicted Cam C		
Cam B	Downturned	148	76	2
	Straight	92	164	82
	Upturned	2	70	112
		Predicted Cam D		
Cam B	Downturned	145	81	0
	Straight	87	169	82
	Upturned	1	74	109
		Predicted Cam E		
Cam B	Downturned	146	80	0
	Straight	77	187	74
	Upturned	1	67	116
		Predicted Cam D		
Cam C	Downturned	145	96	1
	Straight	86	152	72
	Upturned	2	76	118
		Predicted Cam E		
Cam C	Downturned	146	94	2
	Straight	77	171	62
	Upturned	1	69	126
		Predicted Cam E		
Cam D	Downturned	135	98	0
	Straight	89	161	74
	Upturned	0	75	116

TABLE 3.10: Performance evaluation of proposed system for eye shape attributes detection. Acc.: accuracy, Sn.: Sensitivity, Sp.: Specificity, Pr.: Precision, F_1: F_1 score, MCC: Matthews Correlation Coefficient, GT: Ground Truth.

Webcam Setting	Acc.	Sn.	Sp.	Pr.	F_1	MCC
Cam A vs. GT	0.7540	0.7111	0.7938	0.7619	0.7356	0.5072
Cam B vs. GT	0.7580	0.7171	0.7954	0.7619	0.7388	0.5146
Cam C vs. GT	0.7447	0.7066	0.7783	0.7381	0.7220	0.4865
Cam D vs. GT	0.7540	0.7159	0.7879	0.7500	0.7326	0.5055
Cam E vs. GT	0.7567	0.7333	0.7751	0.7202	0.7267	0.5075
Average (all Cams)	**0.7535**	**0.7168**	**0.7861**	**0.7464**	**0.7311**	**0.5043**
Cam A vs. B	0.9452	0.9389	0.9510	0.9468	0.9428	0.8902
Cam A vs. C	0.9452	0.9306	0.9588	0.9544	0.9423	0.8904
Cam A vs. D	0.9572	0.9444	0.9691	0.9659	0.9551	0.9145
Cam A vs. E	0.9332	0.8889	0.9742	0.9697	0.9275	0.8685
Cam B vs. C	0.9278	0.9160	0.9386	0.9316	0.9237	0.8553
Cam B vs. D	0.9452	0.9356	0.9540	0.9489	0.9422	0.8902
Cam B vs. E	0.9291	0.8880	0.9668	0.9606	0.9229	0.8598
Cam C vs. D	0.9372	0.9345	0.9395	0.9318	0.9331	0.8739
Cam C vs. E	0.9265	0.8917	0.9572	0.9485	0.9192	0.8532
Cam D vs. E	0.9439	0.9091	0.9747	0.9697	0.9384	0.8885

TABLE 3.11: Confusion matrices depict eye shape attributes prediction in each webcam versus the ground truth.

Predicted		Desired	
		Round	Almond
Cam A	Round	256	104
	Almond	80	308
Cam B	Round	256	101
	Almond	80	311
Cam C	Round	248	103
	Almond	88	309
Cam D	Round	252	100
	Almond	84	312
Cam E	Round	242	88
	Almond	94	324
All Cameras	Round	1254	496
	Almond	426	1564

TABLE 3.12: Confusion matrices depict eye shape attributes prediction in each individual camera versus another camera.

Desired Cam X VS. Cam Y		Predicted Cam B	
		Round	Almond
Cam A	Round	338	22
	Almond	19	369
		Predicted Cam C	
Cam A	Round	335	25
	Almond	16	372
		Predicted Cam D	
Cam A	Round	340	20
	Almond	12	376
		Predicted Cam E	
Cam A	Round	320	40
	Almond	10	378
		Predicted Cam C	
Cam B	Round	327	30
	Almond	24	367
		Predicted Cam D	
Cam B	Round	334	23
	Almond	18	373
		Predicted Cam E	
Cam B	Round	317	40
	Almond	13	378
		Predicted Cam D	
Cam C	Round	328	23
	Almond	24	373
		Predicted Cam E	
Cam C	Round	313	38
	Almond	17	380
		Predicted Cam E	
Cam D	Round	320	32
	Almond	10	386

of the same aforementioned metrics. Although the system shows many misclassified prediction of eye attributes, however, it could be considered as tolerated prediction cases (skew is small). By ”tolerated prediction”, we mean that most misclassified cases are resulted from predicting the class into average labels, which is clear in all the reported confusion matrices. For example, in Table 3.8, the major predictions of images in Cam E were misclassified as straight rather than upturned label giving: downturned label: 149, straight: 64, upturned: 0. The Straight label is a middle label between extreme and

opposite labels, i.e., downturned and upturned. Thus, the resulting misclassification is usually mild, into the adjacent category, rather than extreme class. This proves that the predicted measurements of attributes are closer to the ground truth value. This also can be clearly noticed in the reported confusion matrix of the eye setting shown in Table 3.5.

To study the correlation of prediction among different webcams, all the evaluation metrics mentioned earlier, as well as confusion matrices, have been utilised to depict the variance and closeness of prediction. A pair of webcams has been compared by considering one of the webcams as the actual class, and the other is the predicted. For example, A represents the desired class in Cam A vs. B, while B is the predicted. This evaluation is significant to measure the correlation (agreement in the model prediction between different webcams). Suppose the prediction agreement between two webcams is high. In that case, it indicates that the prediction model is consistent in the detection, regardless of the prediction compared to the desired class (the ground truth). If the correlation between a pair of webcams is low, this is an indicator that the model prediction is not consistent under different webcam positions. For instance, for eye setting attribute prediction, the lower part of Table 3.4 presents the correlation between cameras (such as Cam A vs. Cam B) in terms of evaluation metrics. Whilst, Table 3.6 reports the confusion matrices explaining the correlation of prediction among different webcams, which shows good prediction agreement among images captured from different cameras' positions. Matthews Correlation Coefficient (MCC) is a correlation measurement that returns a value between +1 and -1, where -1 value refers to a total disagreement between two assessments and +1 represents the perfect detection (agreement). The range of obtained MCC values (between two webcams) was (0.3250 - 0.9145), which indicates the strong correlation of prediction among various webcam positions.

In comparison with the existing methods considering eye attributes detection, Borza et al. [243] and Aldibaja [244] presented eye shape estimation method. Yet, they have

not reported the accuracy of the shape detection and instead reported eye region segmentation performance. To the best of our knowledge, our proposed method for eye attributes detection seeking the suitable eyelashes recommendation is the first work carried out automatically based on computer vision approaches. Thus, we could not find more related works in the literature to conduct further comparison with the results obtained from our framework. To further investigate our model's generalisation performance, we validated the developed framework on external data represented by sample images of celebrities, as shown in Table 3.13.

TABLE 3.13: Performance evaluation of proposed system on external image data. Red cells refer to incorrect detection whilst last column shows the true label. No extreme misclassification in our model predictions.

Celebrity ID	Face Shape	Eye Shape	Eye Setting	Eye Position	Correction
Olivia Wilde	Square	Round	Wide	Upturned	Almond
Rihanna	Oval	Almond	Wide	Straight	Upturned
Cara Delevingne	Oval	Almond	Wide	Straight	Upturned
George Clooney	Oval	Almond	Wide	Downturned	Round
Adele	Round	Almond	Wide	Downturned	Straight
Scarlett Johansson	Round	Almond	Wide	Straight	Heart
Johnny Depp	Rectangle	Almond	Average	Downturned	Square
Chloe Grace Moretz	Heart	Almond	Wide	Upturned	Average
Camilla Belle	Round	Almond	Wide	Downturned	Average
Will Smith	Oval	Almond	Wide	Straight	-
Hilary Swank	Rectangle	Almond	Average	Straight	Wide
Sarah Jessica Parker	Rectangle	Almond	Wide	Straight	-
Emma Watson	Oval	Round	Average	Downturned	-
Selena Gomez	Round	Round	Wide	Straight	-
Naomi Campbell	Round	Round	Average	Upturned	Heart
Meryl Streep	Rectangle	Round	Average	Straight	Close

On the external dataset, our model achieves an accuracy of 0.8125, 0.8775, 0.75, and 0.8125 for face shape identification, eye shape detection, eye setting prediction, and eye position estimation, respectively. We have demonstrated that the system generalises well on another data and performs well under variant conditions such as camera position, obstacle (wearing glasses), light condition, gender and age. On the other hand, it should be considered that the proposed framework brings up some limitations, including: (1) the developed system has not been evaluated rigorously to compare the

results obtained for diverse ethnic and racial groups, (2) the presented research work has not considered eye situation characteristics analysis such as mono-lid, hooded, crease, deep-set, and prominent. However, these limitations are due to a lack of labelled data captured from variant races and ethnicities. These challenges can be overcome when larger annotated data becomes available. Moreover, the threshold parameter values used for converting scalar value into a nominal label of eye attribute are selected empirically. For future studies, studying the possibility of using automatic parameter tuning strategies (heuristic mechanisms) for determining and setting optimum threshold values could be investigated, (3) unlike the detailed analysis of eye attributes identification carried out in this chapter, face shape classification has not been evaluated from the camera's position perspective in this work. An exhaustive study for face shape classification from different webcam positions, similar to the study conducted on eye attributes identification, is needed to be carried out.

3.5 Conclusions

In this chapter, a decision support system for appropriate eyelashes and hairstyle recommendation has been developed. The developed framework has been evaluated on dataset provided with diversity of lighting, age, and ethnicity. The developed system, based on integrating three models, has proven efficient performance in face shape and eye attributes identification. Face and eye attributes measurement tasks are usually carried out by an expert person manually before applying the treatment of individual eyelashes extension and hairstyle. Measuring face and eye characteristics costs time and efforts. The proposed system could alleviate the need for additional time and efforts made by expert before each treatment. Furthermore, face shape identification automatically is a challenging duty due to the complexity of face and the possible variations in rotation, size, illumination, hairstyle, age and expressions. Moreover, the existence of face occlusion from hats and glasses also adds difficulties to the classification process. Our

model trained on diverse images for automatic face shape classification would be viable in many other decision support systems such as eyeglasses style recommendation systems.

Chapter 4

Facial Makeup Detection

The detected facial landmarks' locations in the recommendation system presented in Chapter 3 could be negatively affected and changed by the applied facial cosmetics, which leads to incorrect facial features detection. Thus, for developing a robust and more reliable recommendation system, the makeup detection process could be incorporated into the recommendation system as a prerequisite step to help ensure that the processed image is clear of makeup. In § 4.1, an introduction about facial makeup and its impact has been provided. § 4.2, the problems that can be tackled with the presence of facial makeup. The related work of makeup detection is presented in § 4.3. In § 4.4, the material and proposed methods are described and explained. The results of the proposed systems are revealed and discussed in § 4.5, and finally, the work has been concluded in § 4.6.

4.1 Introduction

Facial makeup has a long track record. It is an instance of a cosmetic modification that may modify the face's perceived appearance. Ageing, natural biological alterations, and plastic surgery are other forms of alterations. In general, surgical modifications are expensive and permanent. Non-permanent cosmetic improvements, such as makeup,

on the other hand, are considered quick, cost-effective and acceptable in different cultures; similarly, they can alter appearance considerably. As the authors of [245] stated, makeup alterations are aimed at (a) altering the perceived facial form by emphasising contouring methods; (b) altering the shape of the perceived nose and its size by applying contouring tools; (c) enhancing or decreasing the perceived mouth size; (d) altering appearance and contrast of the mouth by adding colour; (e) altering the perceived eyebrow shape, colour and location; (f) altering the perceived form, contrast and size of the eyes; (g) hiding dark pigments in the face, especially, located under the eyes; and (h) altering the texture as well as the colour of the perceived skin [245]. Besides the changes and effects described above, cosmetics and makeup might be utilised to effectively disguise and conceal tattoos, birth moles, wrinkles, and scars [246].

A developing cosmetics industry aims to enhance facial attractiveness, while consider maintaining good health. Cosmetics encompasses a wide range of methods, categories, products and have become socially acceptable in all aspects of our lives. The use of makeup, on the other hand, presents a significant obstacle to biometric systems. Facial makeup has the ability to change and conceal one's natural appearance, making certain identification and verification functions more difficult [247, 248]. An experimental study conducted by [249] revealed that the application of facial makeup leads to changes in the skin texture, smoothness, and colour tone. Many other studies showed that facial makeup carries an inherent ambiguity due to artificial colours, shading, contouring, and varying skin tones. The issue becomes more complicated as makeup changes the symmetry of certain facial features like eyes and lips, affecting the distinctive character of faces [1, 250–255].

With advancements in machine learning and computer vision techniques, deep neural network models are becoming very successful and widespread. Considering that deep learning architectures have been successfully used in various fields, including facial image analysis [230, 256, 257], it could even further be exploited to detect the faces disguised by makeup to overcome the flaws in many facial-related analysis methods.

Those models have the ability to extract the features directly from images without the need for human interaction by training on large datasets using various types of learning schemes. A trained human expert or a physical experiment are required to label the data for a learning problem. Therefore, the costs associated with the labelling process will make providing a massive and fully labelled training sets a difficult task, whereas collection of unlabelled data is relatively inexpensive. In machine learning, semi-supervised learning defines a type of algorithms that attempt to learn from both unlabelled and labelled examples. Many semi-supervised learning with deep neural networks were designed based on generative models such as denoising auto-encoders [258], stacked convolutional auto-encoders [259], variational auto-encoders [260] and generative adversarial networks [261]. Pseudo-labelling of self-learning is another powerful method for semi-supervised learning that has shown success in the context of deep learning models [262]. Semi-supervised learning with self-learning works by assigning approximate classes on unlabelled data by making predictions from a model trained only on labelled data.

4.2 Problem of Makeup Presence

The detection of eye attributes in the presented recommendation system depends on the geometry measurements of the landmarks. However, the detected landmarks' location could be negatively affected and changed by the applied facial cosmetics, including makeup and retouching [253]. Experimental study conducted by [249] revealed that the application of facial makeup leads to changes on the skin texture, smoothness, and colour tone. Many other studies showed that facial makeup carries an inherent ambiguity due to artificial colours, shading, contouring, and varying skin tones. The situation becomes more complicated and confound as cosmetics modify the bilateral size of particular facial features such as lips and eyes, making it difficult to recognise faces [1, 250–255]. Authors of [263] studied the effect of accurate detection of facial key-point landmarks on the performance of recognition systems. Moreover, the effects

of various levels of cosmetics application on evaluations of face beauty were investigated by the authors of [264]. Facial makeup recommendation system, presented by the authors of [32], was developed for selecting the most appropriate makeup style to fit a specific occasion. They study the relation between the makeup style and facial attributes.

Authors of [265] also investigated the problem and revealed that the presence of makeup could affect the detected landmarks' locations as shown in Figure 4.1. Therefore, motivated by the above reported observations, we introduce a makeup detection scheme that helps detect the facial images covered by makeup and reject it to reduce the false detection of facial and eye attributes.

FIGURE 4.1: Landmark detection in images without and with makeup [265]. The presence of makeup could affect the detected landmarks' locations.

4.3 Related Work

The beautification effects caused by cosmetics have been studied in the research literature. In comparison to work on makeup recommendations, research involving an image of an already made-up face appears to be extremely rare. The first study that objectively identified the effect of facial makeup on a face recognition system was performed by Dantcheva et al. [245]. They used three techniques related to face recognition which are Gabor wavelets, local binary pattern, and the commercial Verilook Face Toolkit, to test the accuracy of recognition before and after makeup using two databases: the YouTube

MakeUp (YMU) database and the Virtual MakeUp (VMU) database. The presence of makeup in facial images is also identified in [266] using a feature vector that includes shape, texture, and colour information. Another research, [249], tackled the facial verification issue by extracting features from both a makeup-wearing and a makeup-free face, then matching the two faces using correlation mapping. In [267], an approach for makeup face verification was presented by measuring correlations between face images in a meta subspace[1] where canonical correlation analysis (CCA) and support vector machine (SVM) were used for learning and verification, respectively.

The interrelationship between, on the one hand, the subjective interpretation of female facial beauty and, on the other hand, selected objective parameters including facial features, photo-quality and non-permanent facial features were examined and analysed by the authors of [269]. A supervised deep Boltzmann machine was also proposed by the authors of [270] to solve the problem of classifying face images as original or retouched. The authors of [271] used Gabor filtering and histogram of oriented gradients (HOG) methods for feature extraction from VMU and YMU datasets. These features were combined to form the final feature vectors, which were then reduced using the fisher linear discriminant analysis and classified. The authors of [250] presented a locality constrained low-rank dictionary learning algorithm to determine and locate the usage of cosmetics.

Fu and Wang [272] have recently created a system capable of identifying, analysing and digitally removing makeup from a facial image. For makeup detection using a supervised deep learning approach, authors of [273] developed a system based on a pre-trained AlexNet [39] for makeup detection. Furthermore, by introducing a bi-level adversarial network (BLAN), Yi Li et al. [255] presented learning from generation method, targeting makeup-invariant face verification. The vulnerability of many face recognition systems has been recently assessed to facial makeup presentation in [274]. Furthermore, the same research team presented a facial retouch detection method based

[1]Meta Subspace Optimisation (MSO) enables the subspace matrix to be computed at each optimisation cycle [268].

on analysis of photo response non-uniformity (PRNU) in [275] and conducted a review and benchmarking for makeup presentation attacks methods in [276].

The scenario of makeup detection can be leveraged in many real-life applications. For instance, the presence of makeup can significantly affect the results of beauty and attractiveness prediction systems [248, 269, 270, 272]. The development of makeup detection methods could grant advantages to beauty prediction systems to refine their outcomes by utilising prior knowledge about the makeup presence in a given facial image. Furthermore, detecting the makeup-wearing can benefit automated age estimation systems [277, 278] and facial recognition models from the perspective of security [245, 266, 274, 275].

Motivated by the above-reported observations and issues, we introduce makeup detection schemes that help detect the facial images covered by makeup using labelled and unlabelled data. The contribution of our work can be shown to be two-fold: (i) conducting a comparative study by investigating the impact of different learning strategies to automatically detect the presence of makeup, and (ii) carrying out a thorough analysis and comprehensive study involving six challenging datasets. This enables evaluation with facial images collected from various sources, demonstrating the robustness and relatability of the methods.

4.4 Materials and Method

4.4.1 Materials

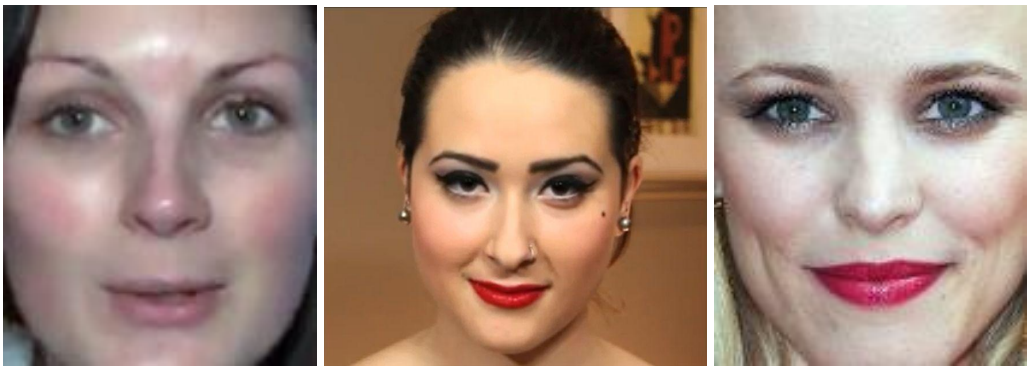
Six publicly available datasets are used in this study including Kaggle [279], YMU [245], VMU [266], MIW [280], MIFS [281], FAM [267]. Kaggle dataset contains facial images of subjects obtained from the Internet comprising 1506 images of people wearing makeup and not wearing makeup. YouTube Makeup (YMU) is a set of 151 facial images collected from YouTube makeup tutorials, focusing on Caucasian females. The images were taken before and after the subjects had their makeup applied [245, 282].

There are four shots per subject: two before makeup application and two after makeup application, totalling 604 images [282]. Virtual Makeup (VMU) is a set of facial images of Caucasian female subjects from the FRGC repository [283]. To simulate the application of makeup, this dataset has been synthetically modified. This change was made by the authors of [266] using the publicly available software Taaz [284]. The total number of images in the dataset is 204, with one before-makeup shot and three after-makeup shots per subject [282]. Makeup in the Wild (MIW) is unconstrained face images, captured under uncontrolled environment, of 125 subjects with and without makeup obtained from the Internet. There are (1-2) images per subject: one before-makeup and one after makeup producing 154 images (77 with makeup and 77 without makeup). Makeup Induced Face Spoofing[2] (MIFS) is a dataset of 624 images collected from 107 subjects (4 images) and 107 target subjects (2 images). "Images of a subject before makeup, images of the same subject after makeup with the intention of spoofing, and images of the target subject who is being spoofed" are three image sets included in this dataset. Finally, FAce Makeup (FAM) dataset contains 519 subjects. Each subject has two face images, one is with makeup, and the other is not, producing 1038 images. All subjects involved in this chapter gave their consent through a data collection process led by the authors, who generated the data and made it publicly accessible for scientific research purposes. Figure 4.2 shows some image samples used in this study.

4.4.2 Proposed Method

This work proposes and designs an automated facial makeup detection system leveraging three types of learning schemes. The deep learning models are evaluated and tested on multiple public datasets, and their performances are compared to ground truth annotations. The classifier in the models deals with makeup identification in an image as a two-class detection task; with makeup and without makeup classes. Using labelled and unlabelled data, we harness semi-supervised strategies and transfer learning schemes

[2]Spoofing by makeup: when a person uses makeup to replicate the appearance of another person [281].



(a) (b) (c)

(d) (e)

(f)

FIGURE 4.2: Samples of images from datasets used in the study. (a) YMU; (b) MIFS; (c) MIW; (d) FAM; (e) VMU; (f) Kaggle.

to implement the proposed systems based on convolutional neural networks (CNNs) as follows:

1. **Supervised: Transfer Learning of Pre-trained VGG16 CNN** pre-trained VGG16 network [41] is fine-tuned on labelled data to extract the facial features and produce a makeup classifier (absence or presence of makeup).

2. **Semi-Supervised: CNN with Self-Learning** the fine-tuned VGG16 network resulted from previous stage can be combined with a self-learning algorithm developed in [262] which is trained on unlabelled data to produce a semi-supervised learning scheme.

3. **Semi-supervised: CNN with Convolutional Auto-encoder (CAE)** in this model, CAE [259] is used to extract the salient visual features in an unsupervised learning manner using the unlabelled makeup data. The trained CAE is then used for initialising the weights of supervised CNN. Whereas the weights of fully connected layers are trained using the labelled data.

Supervised Learning with Pre-trained CNN

Transfer learning aims to use knowledge from the source task to increase learning in the target task. There are three indicators in which the transfer of knowledge might help boosting learning. The first is the initial performance achieved in the target task utilising only transferred knowledge before any further learning. The second factor is the time it takes to learn the target task thoroughly using transferred knowledge versus learning it from scratch. The third factor is the difference between the final performance level achieved in the target task after applying transfer learning and fine-tuning and the final level without transfer. Negative transfer occurs when a method of transfer degrades performance. Positive transfer between adequately related tasks while avoiding negative transfer between tasks that are less correlated is one of the most complex issues in designing transfer strategies. When transfer learning occurs from one task to another,

transferring the properties of one task onto the attributes of the other is often required to establish correspondences [285].

Transfer learning strategies harnessed to boost the evaluation performance in many computer vision tasks can be categorised into three categories,

1. Re-training the entire pre-trained CNN architecture on the target dataset, yet avoiding the random weight initialisation by starting from the pre-trained weight values.

2. Training the new classifier (fully connected layers) related to the target task and freezing the other layers (convolutional and all other layers). In this transfer learning strategy, the pre-trained CNN works as a feature extractor where the weights of the CNN layers, except fully connected layers, are retained without change.

3. Training some of the convolutional layers, especially the top layers of CNN, and the classifier (fully connected layers). The original weights are exploited as a starting point for learning.

To conduct knowledge transfer in our comparative study, we leveraged pre-trained convolutional neural network VGG16 [41]. VGG16 is a remarkable structure that can be applied for a wide range of computer vision tasks. We re-trained the pre-trained VGG16, which is earlier trained on a large-scale hierarchical image database (Imagenet) [286], using our datasets. Figure 4.3 depicts the transfer learning using pre-trained VGG16 model for makeup detection.

Semi-Supervised Learning with Self-Learning Pseudo-Labelling

Unlike the supervised learning scheme that requires completely labelled data, semi-supervised learning combines supervised and unsupervised learning techniques. As the name indicates, semi-supervised learning uses both a labelled set of training data and an unlabelled collection of training data. The self-training scheme introduced in [262]

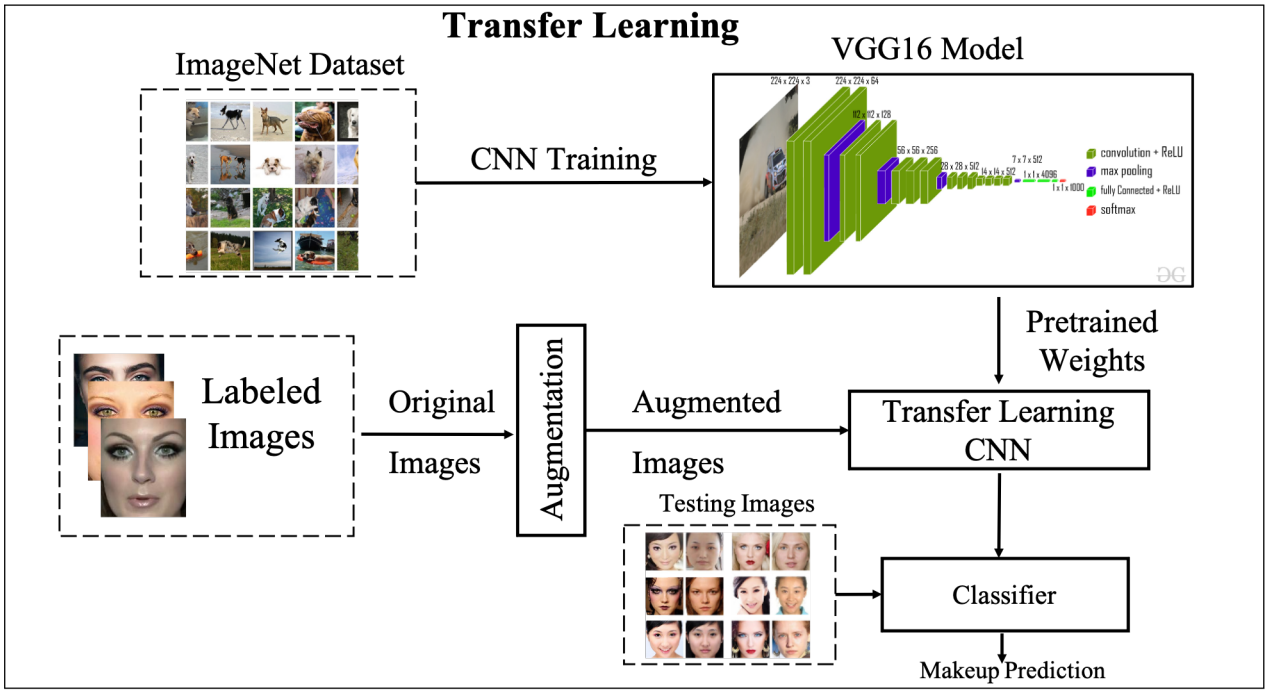


FIGURE 4.3: Supervised learning scheme with a transfer learning using pre-trained VGG16 model for makeup detection.

works by training the classification models (CNNs) on labelled instances and employing the trained classifier to identify class labels of the unlabelled examples. The predicted class labels with the highest probability of being correct are adopted as pseudo-labels. The examples with pseudo-labels are then used to train the classifier. Once the classifier is trained, it can be used to test unseen instances. The description of pseudo-labelling with self-learning merged with classifier is presented in Figure 4.4.

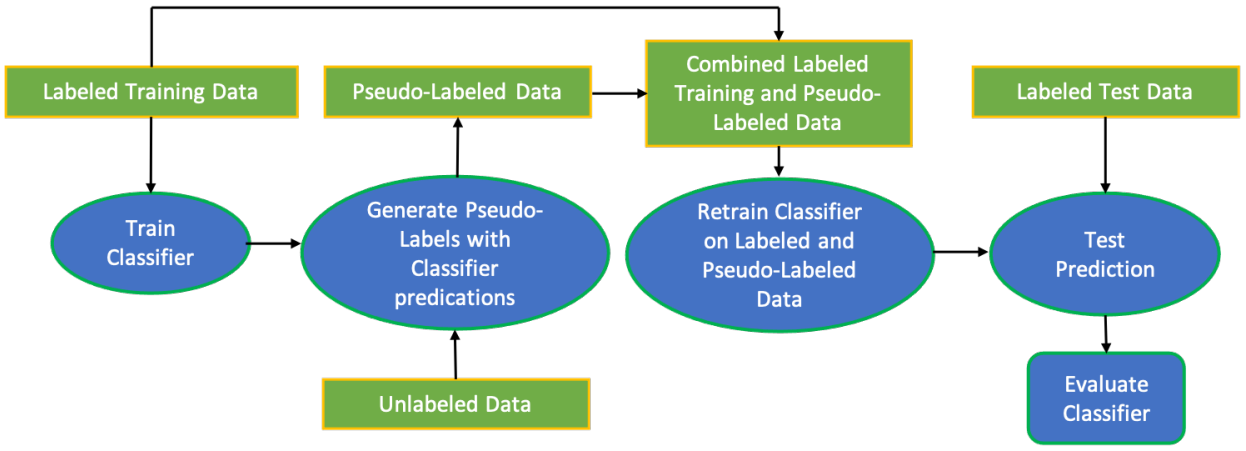


FIGURE 4.4: Self-learning with pseudo-labels scheme.

On a conceptual level, this scheme integrates the pre-trained VGG16 model with the self-learning method to double the size of the training dataset by choosing the most reliable examples from the unlabelled image data. The steps of learning process which are repeated until the convergence using makeup labelled images l and an unlabelled images u could be described as follows:

1. The pre-trained VGG16 is used to predict the labels of the unlabelled images u.
2. The prediction scores with highest confidence rates ($l^{'}$ pseudo-labels) obtained from applying VGG16 model are selected and combined with label images l.
3. The combined pseudo-label images and labelled images are then used to train the classifier. The loss function can be represented as $Loss_{total} = Loss_{labelled} + Loss_{unlabelled}$.

The unlabelled image samples u that are left unclassified after the algorithm's convergence are omitted. Figure 4.5 illustrates the semi-supervised learning scheme with pseudo-labels and self-learning for makeup detection.

Semi-Supervised Learning with Convolutional Auto-Encoder

Autoencoders [287] are neural networks widely used to extract features (feature learning) and dimensionality reduction. Several auto-encoders can be stacked to form a deep hierarchy constituting a convolutional auto-encoder (CAE). In a greedy, layer-wise manner, unsupervised training can be achieved in CAE. Back-propagation can then be used to fine-tune the weights. The top-level activations can be used as feature vectors for classifiers [259]. With multiple layers of non-linear activation functions, Convolutional Auto-Encoders (CAE) can learn non-linear mappings. In CAE, it is not necessary to have dense layers. The CAE can deal with and learn from data captured from different sources, including video, image, and time series, using its convolutional layers. Providing these data, the CAE has the capability to produce a feature representation in each layer. It also can employ a transfer learning scheme to improve the encoder

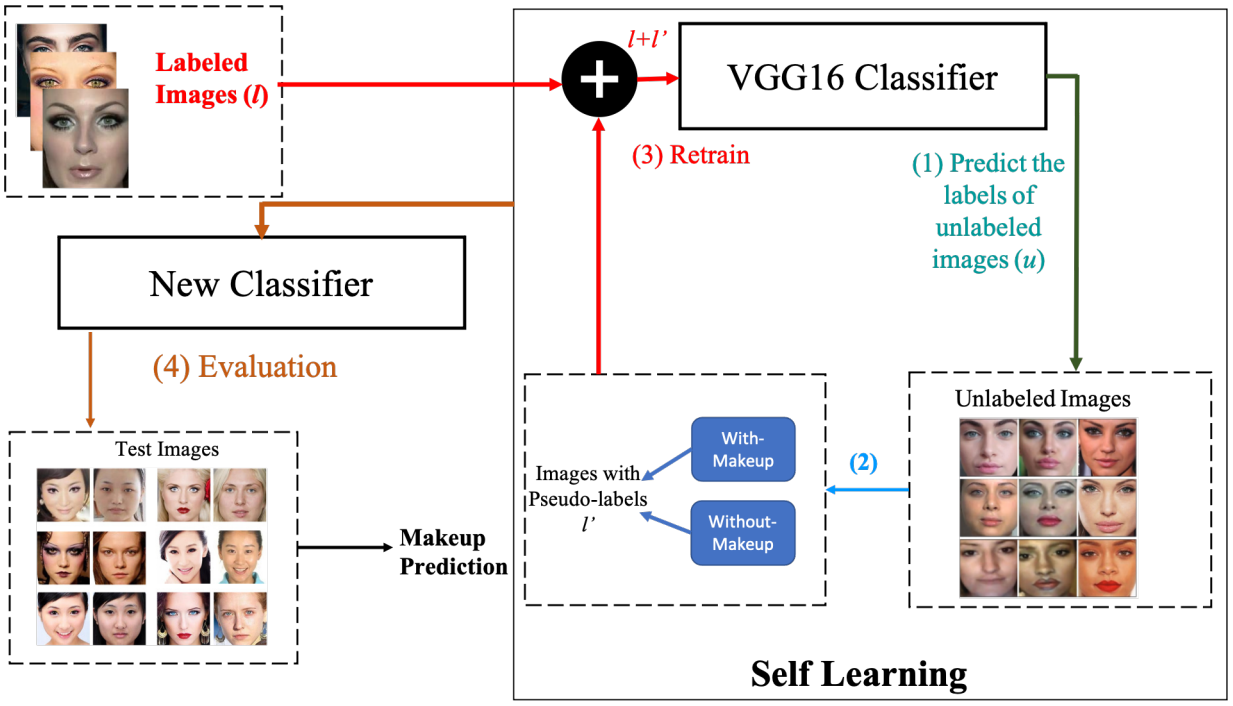


FIGURE 4.5: Semi-supervised learning scheme with pseudo-labels and self-learning for makeup detection.

by using pre-trained layers from another model [288]. Our implemented convolutional auto-encoder (CAE) contains the layers described in Table 4.1. This Table displays the name of layers, output feature space shape, and the number of parameters. The total number of parameters is the sum of all weights in the convolutional layer. The shape of output refers to the dimensions of each image after passing through convolutional filters, producing the size of feature space. Our convolutional auto-encoder (CAE) contains an encoder path, latent features, decoder path. The encoder part comprises eight convolution layers, each followed by batch normalisation, max-pooling, and drop out layers. The decoder part consists of eight up convolution layers, each followed by batch normalisation, upsampling, and drop out layers.

TABLE 4.1: The architecture of the implemented convolutional auto-encoder.

Layer(type)	**Output Shape**	**#Param**
input-1(InputLayer)	[(128, 128, 3)]	0
conv-1(Conv2D)	(128, 128, 32)	896
batchnormalisation	(128, 128, 32)	128

conv-1-2 (Conv2D)	(128, 128, 32)	9248
batchnormalisation-1	(128, 128, 32)	128
pool-1(MaxPooling2D)	(64, 64, 32)	0
dropout(Dropout)	(64, 64, 32)	0
conv-2(Conv2D)	(64, 64, 64)	18496
batchnormalisation-2	(64, 64, 64)	256
conv-2-2(Conv2D)	(64, 64, 64)	36928
batchnormalisation-3	(64, 64, 64)	256
pool-2(MaxPooling2D)	(32, 32, 64)	0
dropout1(Dropout)	(32, 32, 64)	0
conv-3(Conv2D)	(32, 32, 128)	73856
batchnormalisation-4	(32, 32, 128)	512
conv-3-2(Conv2D)	(32, 32, 128)	147584
batchnormalisation-5	(32, 32, 128)	512
pool-3(MaxPooling2D)	(16, 16, 128)	0
dropout-2(Dropout)	(16, 16, 128)	0
conv-4(Conv2D)	(16, 16, 256)	295168
batchnormalisation-6	(16, 16, 256)	1024
conv-4-2(Conv2D)	(16, 16, 256)	590080
batchnormalisation-7	(16, 16, 256)	1024
pool-4(MaxPooling2D)	(8, 8, 256)	0
dropout-3(Dropout)	(8, 8, 256)	0
flatten(Flatten)	(, 16384)	0
latent-feats(Dense)	(, 1024)	16778240
reshape(Reshape)	(2, 2, 256)	0
upsample-4(UpSampling2D)	(4, 4, 256)	0
upconv-4(Conv2D)	(4, 4, 256)	590080
batchnormalisation-8	(4, 4, 256)	1024
upconv-4-2(Conv2D)	(4, 4, 256)	590080
batchnormalisation-9	(4, 4, 256)	1024
upsample-3(UpSampling2D)	(8, 8, 256)	0
dropout-4(Dropout)	(8, 8, 256)	0
upconv-3(Conv2D)	(8, 8, 128)	295040

batchnormalisation-10	(8, 8, 128)	512
upconv-3-2(Conv2D)	(8, 8, 128)	147584
batchnormalisation-11	(8, 8, 128)	512
upsample-2(UpSampling2D)	(16, 16, 128)	0
dropout-5(Dropout)	(16, 16, 128)	0
upconv-2(Conv2D)	(16, 16, 64)	73792
batchnormalisation-12	(16, 16, 64)	256
upconv-2-2(Conv2D)	(16, 16, 64)	36928
batchnormalisation-13	(16, 16, 64)	256
upsample-1(UpSampling2D)	(32, 32, 64)	0
dropout-6(Dropout)	(32, 32, 64)	0
upconv-1(Conv2D)	(32, 32, 32)	18464
batchnormalisation-14	(32, 32, 32)	128
upconv-1-2(Conv2D)	(32, 32, 32)	9248
batchnormalisation-15	(32, 32, 32)	128
upconv-final(Conv2D)	(32, 32, 3)	867

The basic mathematical principle of standard fully connected auto-encoder (AE) requires mapping the unlabelled input image x into a latent representation h in the hidden layers using the following formula:

$$h = s(Wx + b) \tag{4.1}$$

where s is a non-linear activation function, W is the weights of the network, and b is bias. The auto-encoder then tries to decode and reconstruct the unlabelled input image $\hat{x}$ by:

$$\hat{x} = f(W^{'}h + b^{'}) \tag{4.2}$$

The auto-encoder is learned using a back-propagation algorithm to minimise the reconstruction error. The extended version of the conventional fully connected auto-encoder (AE) is convolutional auto-encoders (CAE) which is a combination between fully connected auto-encoder and convolutional neural networks (CNNs). To integrate CNNs with auto-encoders, a corresponding deconvolutional layer must be built for each convolutional layer. Furthermore, because max-pooling layers cause information loss, an un-pooling layer should resemble the original values. Deconvolution layers are equivalent to the convolutional layers but transposed. Unlike AE, the weights in CAE are shared among all locations in the input image, leading to preserve spatial locality. The latent representation of the $k-th$ hidden layer is represented by:

$$h^k = s(x * W^k + b^k) \tag{4.3}$$

The reconstructed image is defined by the formula:

$$\hat{x} = f(h * W^{'} + b^{'}) \tag{4.4}$$

$W^{'}$ identifies the transpose operation of the encoder layer weights, and $*$ refer to the convolution operation. The objective function targeted to be minimised is defined as MSE cost function to measure the error between the original and constructed images. Figure 4.6 explains the semi-supervised learning scheme with a convolutional auto-encoder for makeup detection.

4.5 Experimental Results and Discussion

To conduct the experiments, 2642 labelled images captured from 1060 subject, collected from Kaggle, YMU, VMU, MIW, MIFS, and FAM datasets, are used to train and evaluate the models. Manual investigation and cleaning are carried out on data. Unclear and low-quality images are removed to reduce their negative impact on supervised

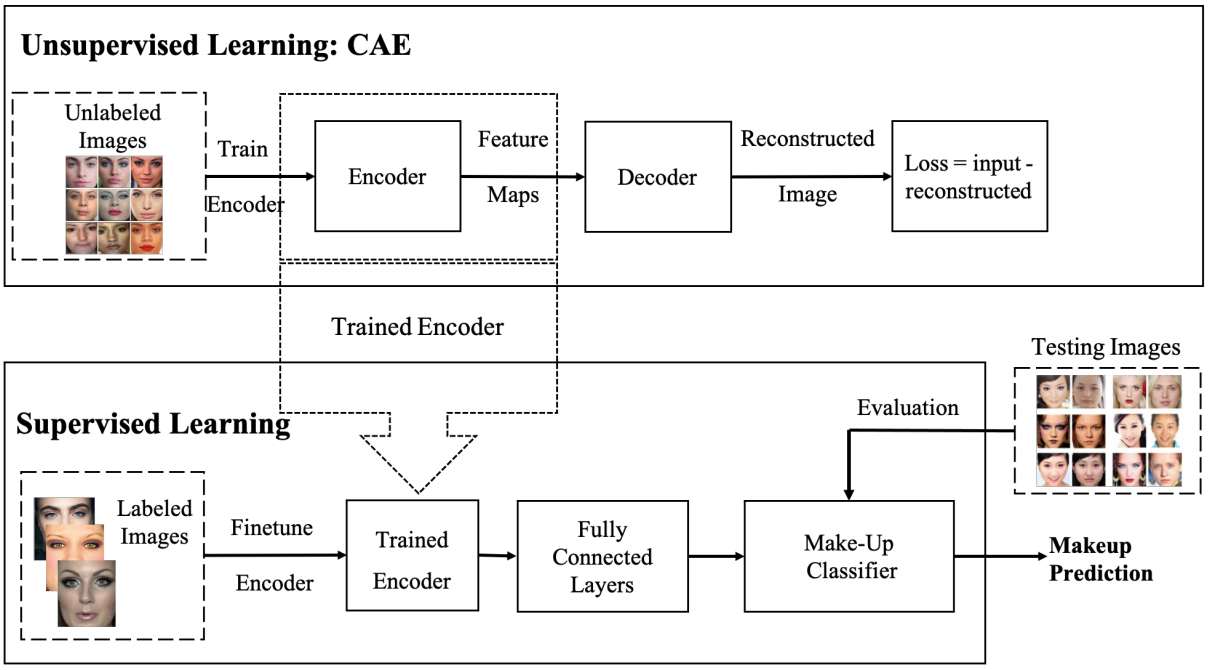


FIGURE 4.6: Semi-supervised learning scheme with a convolutional auto-encoder for makeup detection.

models' learning. After reduction, 2479 images are retained, comprising of 1265 image with makeup and 1214 without makeup. To train the unsupervised learning models, 446 unlabelled images are selected from Kaggle dataset [279]. The first step of our system is to detect and locate the face region and crop it. The model trained on the histogram of oriented gradients (HOG) features with support vector machine (SVM) classifier [231] is used to detect the face region. All the detected and cropped faces are then resized into $128 \times 128 \times 3$ and normalised.

For supervised learning, we used pre-trained convolutional neural network VGG16 [41] for feature extraction and fine-tuned it on our datasets. It typically comprises 16 layers including 13 convolution layers of size 3×3 which are followed by five max-pooling layers of size 2×2. After each convolution layer, a rectified linear unit (ReLU) activation function is appended. The probabilities for each category are then generated using softmax layer that follows three dense layers. To apply the fine-tuning using our data, the dense and softmax layers are replaced with new layers. The new layers consist of two fully connected layers with 4096 neurons and a softmax layer. The Stochastic Gradient Descent (SGD) with learning rate of 0.0001 and momentum of 0.9 was used

as an optimisation algorithm and binary cross-entropy was used as loss function. Four types of data augmentation were applied during the learning stage to double the size of data artificially. These data augmentation types include horizontal flipping, random zooming, and horizontal and vertical shift. The network was trained for 100 epoch with batch size of 32. The learning curve of VGG16 network during training stage is shown in Fig. 4.7a.

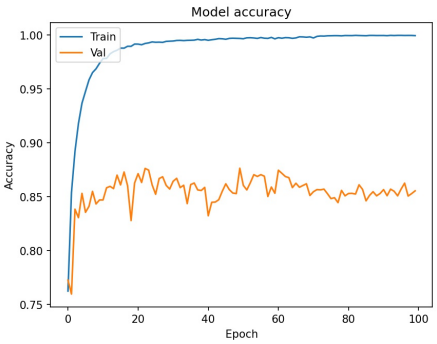
In the second model, the VGG16 trained in the first model is merged with an unsupervised learning scheme via self-learning producing a semi-supervised learning approach. This aims to increase the size of training data by applying self-learning targeting towards producing pseudo-labels. The expansion of training data would help to improve the robustness of the classifier. With samples from the unlabelled dataset that have high confidence when categorised with the VGG16 CNN model, the self-learning approach [262] is exploited to extend the size of the initial training data. Iteratively adding pseudo-labels (obtained from the self-learning model) to the training dataset is achieved by updating the classifier for each iteration and labelling the rest unlabelled instances. Thus, the appending of pseudo-labels to the training data keeps on until the convergence. The unlabelled examples that remain unclassified at the end of the training stage are discarded due to their unreliability. Finally, the testing images are used to validate the resulted classifier. The same number of epochs, optimiser and learning rate used during VGG16 learning are used here. The learning curve of semi-supervised during the training stage is shown in Fig. 4.7b.

In the third model, unsupervised training using unlabelled data is applied by the convolutional auto-encoder (CAE) model before training the classifier on labelled data, producing a semi-supervised learning scheme. Basically, the auto-encoder tries to reconstruct the original input image using a back-propagation algorithm targeting to reduce the reconstruction error. The goal of using CAE is to find the representative features from the unlabelled data. After training the CAE as an unsupervised method, we

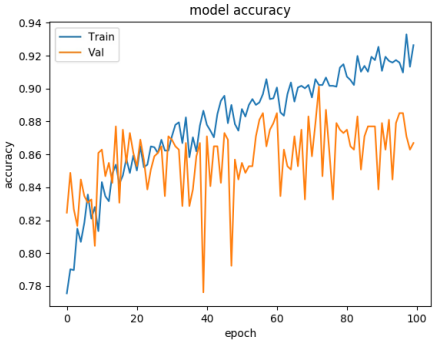
remove the decoder components and use the encoder part of CAE for initialising a supervised CNN. On top of these layers, we add two fully connected layers of size 512 and 256, followed by a classification layer (softmax output unit of size two since there are two classes; makeup and without makeup). The weights of the encoder part trained on unlabelled data are used for initialising the CNN training, whereas the weights of fully connected layers are trained from scratch using labelled data. It has been shown that the pre-training of the network allows for higher generalisation performance than when starting from a random weight initialisation. The auto-encoder was trained for 50 epoch with a batch size of 32. Adam optimiser [159] with binary cross-entropy was adopted for model optimisation with a learning rate of 0.00015. Fig. 4.7c states the learning curve in term of loss for CAE training. To produce the makeup classifier, the encoder part with connected layers are trained on labelled data using an SGD optimiser with a learning rate of 0.001 and momentum of 0.9 for 100 epoch. Fig. 4.7d and Fig. 4.7e show the system performance during the training in terms of error loss and accuracy, respectively.

The proposed makeup detection algorithm's performance is trained and tested using a 5-fold cross-validation scheme. In terms of subjects, there is no overlap between the training and testing sets. The five datasets are separated in an equal percentage through the five folds. Four folds are used for learning the makeup detection model, and the remaining fold is considered for model testing. This process is repeated five times to compute the average over the folds in the 5-fold cross-validation scheme. The performance of the three models are evaluated using two metrics; accuracy and Area Under Curve of Receiver Operating Characteristic AUROC (measurement of the ability of a classifier to distinguish between the positive and negative classes (separability)) as reported in Table 4.2 and displayed in Fig. 4.8.

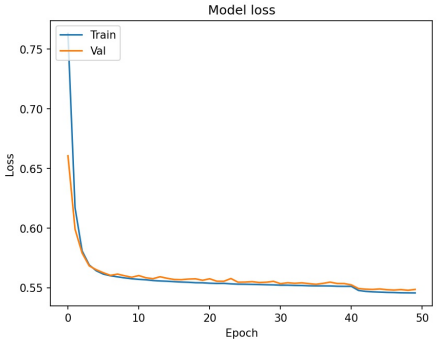
The Receiver Operating Characteristic curve (ROC curve) is an effective measurement when predicting the likelihood of a classifier. It examines the False Positive Rate



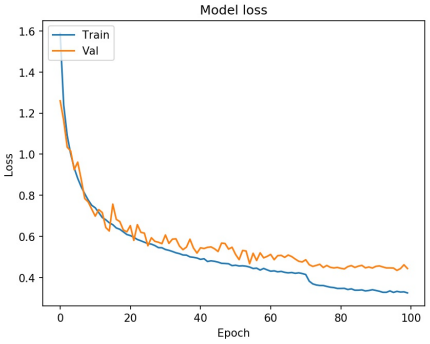
(a) VGG16 (accuracy)



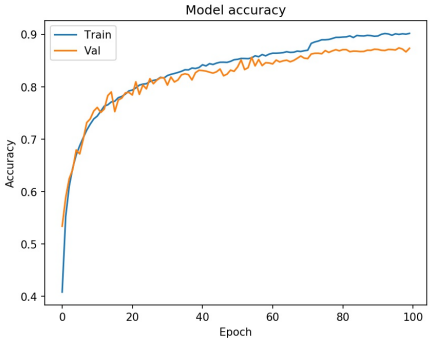
(b) CNN with Self-learning (accuracy)



(c) Convolutional auto-encoder (loss)



(d) Autoencoder-Classifier (loss)



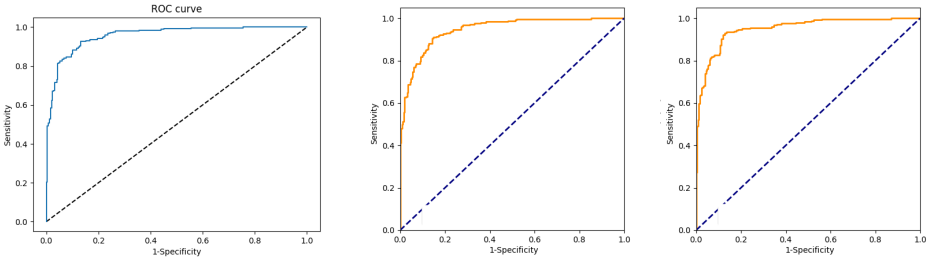
(e) Autoencoder-Classifier (accuracy)

FIGURE 4.7: Learning curves of the implemented models (loss and accuracy curves).

TABLE 4.2: Performance of makeup detection in three models.

Method	**Accuracy**	**AUROC**
VGG16 CNN	86.69%	92.30%
CNN with Self-Learning	87.40%	94.69%
Autoencoder-Classifier	88.33%	95.15%

FPR (1-specificity) versus the True Positive Rate TPR (sensitivity) for various threshold values ranging between 0.0 and 1.0. The performance of a classifier assessed by a ROC curve seeks to develop a decent model that optimises the trade-off between FPR and TPR. The size of the area under the curve is what matters (AUROC) in the ROC curve. The optimal curve is where the classifier can discriminate between negative and positive outcomes with a 100% area under the curve, which is extremely difficult. The AUROC of an excellent model is around 100%, indicating that it has a solid separability performance. A weak model has an AUROC approaching 0%, indicating the lowest separability performance. The predicted FPR and TPR and the curve's shape offer a lot of information about the Roc curve. Lower values on the x-axis of the ROC show lower false positives and greater true negatives. Similarly, higher values on the y-axis of the Roc curve imply more true positives and fewer false negatives [289]. The models under study attained AUROC of 92.30%, 94.69%, and 95.15% in VGG16CNN, CNN with Self-Learning, and Autoencoder-Classifier, respectively. The obtained results reveal high detection outcomes in all classification models represented by curves that bow up to the top left of the plot. This also concludes that the classifier can detect more numbers of true positives and true negatives than false negatives and false positives.



(a) VGG16 (b) CNN with Self-learning (c) Autoencoder-Classifier

FIGURE 4.8: AUROC curves of the developed models.

We conducted a thorough investigation using five publicly available labelled datasets. This helps us compare our methods to earlier work by evaluating image datasets collected from various sources, demonstrating the reliability of our methods. When it comes to comparing our suggested methods to other approaches in the literature, we

carry out a direct comparison with work presented in [273] who tested their system on the same datasets we used in our experiments and used a similar data setup. Their method achieved an accuracy of 79% using a deep supervised learning approach which is apparently lower than our obtained performance. Wang et al. [250] have evaluated their methods on YMU, MIW, and VMU, achieving an accuracy of 91.59%, 91.41%, and 93.75%, respectively. Moreover, Chen et al. [266] reported accuracy of 89.94% on the YMU dataset. However, both research works used features computed manually from the dataset, which are unnecessary to represent other unseen datasets.

Furthermore, Wang et al. [250] pre-processed and aligned all images in the different datasets using facial landmarks. This makes the performance of their method is highly affected and biased by the alignment process. Whereas our method requires no pre-processing step except for image normalisation. We also do not know which subjects were picked up and allocated for training and testing their models, making the comparison not exact. The authors of [245], [267], [271], [281], and [274] are reported face verification and vulnerability assessment performance but not facial makeup detection performance. In our work, we did not conduct face recognition experiments as our methodology's objective is only to detect the presence of the facial makeup. Moreover, our method outperformed the makeup detection method presented in [249] that achieved detection accuracy of 55.5%, 52%, 72.5%, and 52.5% using colour, smoothness, texture, and highlight features, respectively. It is worth mentioning that the AUROC has not been reported in the existing methods.

To sum up, the proposed semi-supervised models could address the lack of labelled image data using CAE and a self-training scheme, whereby only a subset of images is annotated. These approaches mitigate the burden of obtaining large hand-labelled images, which can be costly or impractical

4.6 Conclusions

In this work, we conducted an investigation relating to makeup detection using multiple learning schemes. Three learning mechanisms, including supervised, unsupervised, and semi-supervised with transfer learning approaches, have been exploited to identify the presence of makeup in the input facial images as a binary classification task. The obtained experimental results demonstrated that our methods efficiently identify the makeup and generalise well on unseen data compared to the existing makeup detection methods. It also reveals that unlabelled data can yield changes in learning accuracy when used in combination with a limited amount of labelled data. Thus, semi-supervised learning can be of great practical benefit in such cases, which improves the robustness of the developed models. Considering these findings, in the future, we propose to extend our makeup detection system by developing a makeup removal algorithm that depends on the makeup detection performance. The makeup removal system could recover the bare face, which helps improving facial recognition frameworks and other facial-related systems.

Chapter 5

Face Contour Localisation for Virtual Hairstyle Makeover

The hairstyle and eyelashes recommendation system presented in Chapter 3 could be extended by synthesising the hairstyle and applying the recommended hairstyle virtually. Before applying the hairstyle virtually, the prerequisite step is to recognise the face boundary region from the hair and extract the face contour accurately. Thus, this chapter presents a method to (segment) localise the face boundary to help apply the haircut virtually on the localised boundary of the face. In § 5.1, an introduction of face segmentation and its applications have been presented. In § 5.2, a literature survey related to face segmentation has been explored. The materials, methodology, and results are presented in in § 5.3 and § 5.4, respectively. Finally, the work of this chapter is concluded in § 5.5.

5.1 Introduction

Segmentation of objects plays an important role in many real-life aspects and enables various applications in fields such as education, design, try-on, and e-shopping. The

traditional process of buying a new product from online stores (e-shopping) has a significant shortcoming. Customers cannot see how they will look until the sale is ended, and their final appearance might be different from the in-store trial. This can lead to buyer's disappointment and remorse. Integrating object segmentation with the augmented reality technologies (AR) enables the integration of virtual objects into real-world video, allowing computer-generated objects to be embedded into an input image or video as if they were part of the scene. These technologies improve people's perceptions of virtual try-on systems. Virtual try-on systems allow users to pre-visualise things and have a more credible try-on experience from their own homes. Virtual try-on systems have been designed and developed for objects such as clothes [290] and eyeglasses [291].

Face segmentation is localising and detecting the boundary of faces from different perspectives in a given image. Face segmentation is widely used in many vital applications such as security-related systems, entertainment, computer interaction, beautification and makeup industry, try-on and medical applications by providing essential information obtained from faces [292–294]. With the advancement of face segmentation and parsing of its anatomical structure methods and with the help of AR and deep learning, applying virtual makeup has become feasible [22]. On a similar topic, virtual try-on (makeover) of hairstyle and haircut targets to trial hairstyle before using it in the hair salon. This can be accomplished by blending the synthesis hairstyle with an input image. Many mobile applications have been built recently for hairstyle makeovers and try-on, such as ModiFace [1] and Hairstyle Magic Mirror [2]. Hairstyle makeovers (try-on) are designed for persons who have difficulty communicating their intended haircuts to their hairdressers or wish to see how they look with various hairstyles. The hairstyle makeovers are likely to benefit hairdressers who want to assist their customers in finding the perfect hairstyle or show them different hairstyles without actually cutting their hair.

[1] `https://play.google.com/store/apps/details?id=com.modiface.haircolorstudio.free`

[2] `https://play.google.com/store/apps/details?id=air.MagicMirrorFree`

Localising the face region by finding the face segmentation mask [295] is the prerequisite step for any virtual try-on related to the head and face area, including hairstyle and haircut makeovers. For instance, the authors in [296] split the face area into the skin and non-skin areas (hair), targeting to finding the face contour and applying the makeup on the skin area. They then adopted Gaussian Mixture Model (GMM) to achieve the refinement on the skin areas. Moreover, for automated makeup transfer purposes, Lu et al. [297] used an aligned face mesh as an attribute to segment the facial image into two classes, hair or face. They first projected the face geometry onto the image plane and constructed face and hair masks. They then trained a GMM to differentiate the hair pixels located inside the facial mask.

5.2 Related Work

In the literature, several facial segmentation methods have been introduced with a noticeable performance improvement. Ghiasi et al. [298] presented a method for face segmentation by alternating between segmentation and landmark localisation using deformable part models. They reported overlapping[3] of 83.5% on the Caltech Occluded Faces in the Wild (COFW) dataset [300]. Using the regional predictive power (RPP) estimation method, Yang et al. [301] achieved overlapping results of 72.4% on the COFW dataset.

The theoretical development and applications of deep CNN models on a wide range of computer vision tasks, such as image registration, segmentation, and classification, have demonstrated great success [302]. Considering the image segmentation accuracy, Fully Convolutional Network (FCN) [303] has surpassed and dominated other segmentation methods. Saito et al. [304] proposed a real-time 3-D face segmentation system based on FCN reporting overlapping of 83.9% on the COFW dataset. Nirkin et al. [305] used FCN for face segmentation achieving an overlap of 83.7% on the same

[3]Overlapping: is the area of overlap (intersection) between the predicted segmentation outcome and the ground truth provided by manual annotation divided by the area of union between the predicted segmentation and the ground truth [299].

dataset. However, FCN introduced limitations that could hinder its applicability. These limitations could be noticed by the expensive pixel-level annulations required for the model training. Another challenge experienced in the FCN is learning from data with imbalanced class distribution, which leads to an overfitting issue towards the dominated classes. The pixel-level labelling is necessary for all segmentation tasks. Segmentation works by classifying the pixels in the image by recognising the pixels belonging to a certain object. Imbalanced data, in segmentation tasks, include images that have tiny or small objects. In this case, the number of pixels belonging to the object (foreground) is much less the pixels that represent the background. This imbalance leads the network to learn well from the majority labels, which is the background pixels, but not the minority (the foreground pixels).

Facial detection and segmentation are challenging tasks for several reasons. First, under different facial expressions and head poses, facial appearance changes significantly across subjects. Second, the appearance of the faces on the facial images would be affected by environmental conditions such as illumination. Third, incomplete facial appearance information can result from facial occlusion by other objects or self-occlusion due to extreme head poses.

Despite the significant developments of face detection and segmentation algorithms, the earlier works focus on less challenging facial images without the aforementioned facial variations. Later, the developed face detection and segmentation algorithms aimed to handle several variations within images collected in controlled conditions such that facial poses and facial expressions can only be in certain cases and categories. Generally, there is still a lack of a robust and reliable approach that can handle all face variations and conditions.

In this work, we focus on the challenging unconstrained "in-the-wild" conditions in which facial images can undergo arbitrary head poses, variations of facial expressions, occlusions and illumination. We present a feature learning approach to segment faces from unconstrained images. The developed facial segmentation model combines the

functionality of both Faster R-CNN [306], originally developed for localisation, and FCN [303], originally developed for segmentation, in end-to-end learnable model. This integration aims to mitigate the need for a large amount of pixel-level annotations by exploiting the region of interest detected by the object localiser. The network performs object localisation first, which is then used as a cue to guide the learning process of the segmentation network.

5.3 Materials and Methods

5.3.1 Materials

A public dataset [307, 308] was exploited for evaluating the presented method. It consists of 3500 unconstrained (in-the-wild) face images, along with their corresponding manual annotation of the face boundaries. The dataset is split into 2800 training images and 700 test images. The resolution of facial image is 178×218 with large diversities in the faces. The images are a subset from the CelebFaces Attributes dataset, which is a large-scale dataset comprising of 202,599 images [308].

5.3.2 Methods

We adapt Regional Convolutional Neural Network (Mask R-CNN) ([309] to achieve the face segmentation from the unconstrained images. Mask R-CNN [309], originally designed for instance segmentation tasks, integrates the segmentation and localisation in one end-to-end learnable model. This has been adapted here to accomplish the face region segmentation. The proposed face segmentation method shown in Figure 5.1 comprises four major parts, which can be described as follows:

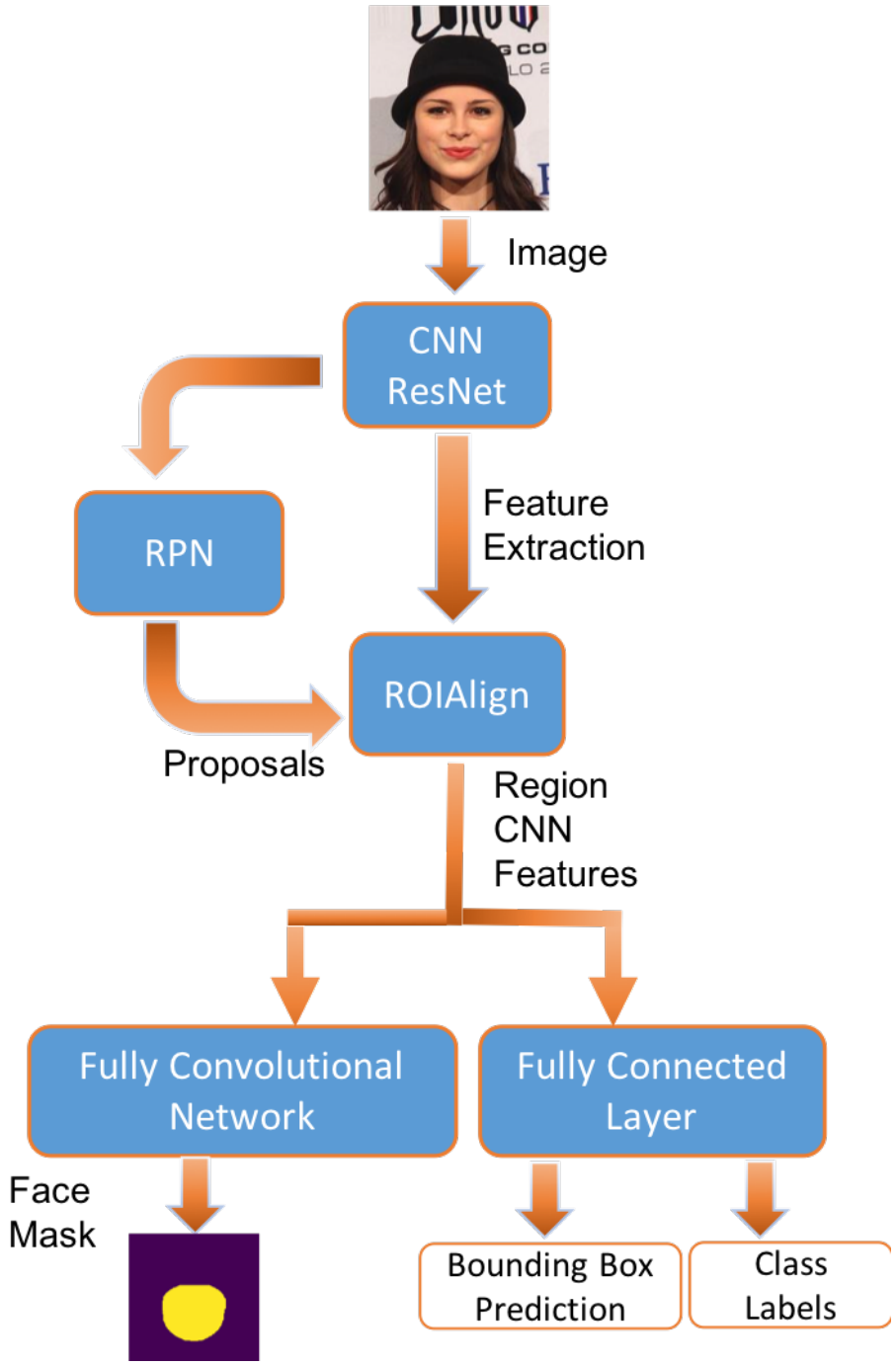


FIGURE 5.1: A block diagram of the proposed face segmentation system. RPN: Region Proposal Network.

Feature Extraction by Resnet101 Model

Our method's initial phase is the feature extraction of images. A conventional CNN with stacked convolutional and pooling layers creates the feature extraction module. This module is adopted as the segmentation task's backbone feature extractor. The backbone network then receives the resized face images along with their masks of size $224 \times 224 \times 3$. Faster R-CNN is implemented with Resnet34, Resnet50 and Resnet101. We use the Resnet101 model [42] as a backbone because it is the deepest architecture. Transfer learning is used to initialise the top fifty layers of the pre-trained Resnet101 model with pre-trained weights rather than training the model from scratch. Faster R-CNN uses the resulting feature map as an input.

The earlier CNN architectures, which precede Resnet architecture, focused on increasing the number of layers in the designed model targeting to achieve better performance. However, as network depth grows, accuracy becomes saturated and then rapidly degrades. ResNet, developed by researchers from Microsoft, has addressed this issue by proposing residual connections (skip connections) in the implemented deep architectures. Thus, there are two primary key reasons to incorporate skip connections which are preventing the vanishing gradients or alleviating the degradation resulting from the accuracy saturation problem. The block diagram of the residual block is shown in Figure 5.2.

As the CNN model grows more complex and deeper, the layers' ability to transport information from shallow levels becomes increasingly challenging, and the information gets lost. The authors of ResNet investigated this issue which causes degrading the performance of CNN models. To overcome the problem, deeper levels must directly transfer information from shallow layers, i.e. identity mapping. Thus, gradients may simply flow from layer to layer by using a skip connections scheme. This aids in the training of deep networks since even the lowest layers receive activations from the upper layers [310].

There are two types of residual connections. First, when both the input and output

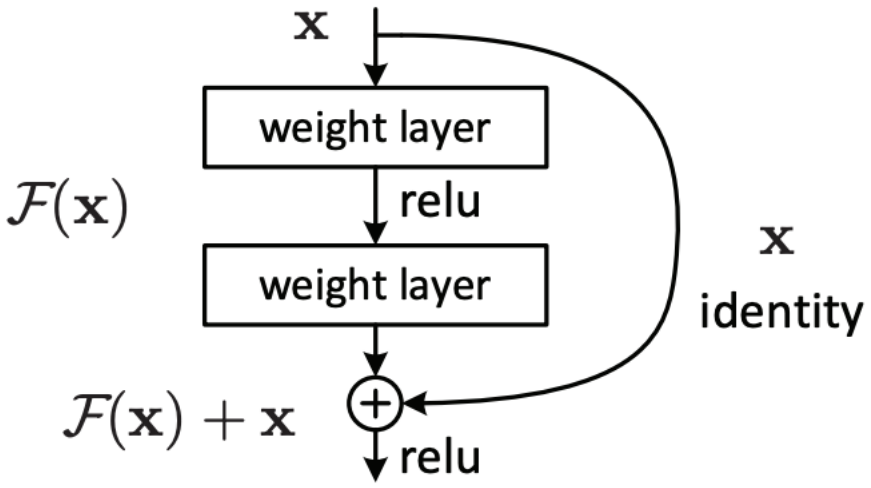


FIGURE 5.2: Block-diagram shows residual connections of Resnet model.

have the same dimensions, the identity shortcut values of (x) could be immediately used as follows:

$$Y = F(x, W_i) + x \tag{5.1}$$

Second, when the dimensions are different, the shortcut still applies identity mapping, with zeros padded with the increased dimension. The complete architecture of Resnet101 is depicted in Figure 5.3.

Localisation by Region Proposal Network

The facial region localisation has been conducted by using the deep Faster R-CNN model [306]. Figure 5.4 shows that Mask and Faster R-CNN show the most frequent used algorithms over time[4]. Faster R-CNN is developed for object detection and localisation. Mask R-CNN is an extension of Faster R-CNN used originally for segmentation. To produce bounding boxes, Faster R-CNN is used to generate and predict numerous regions of interest (ROIs), which are proposals encompassing possible face areas. Region Proposal Network (RPN) is used to locate potential facial regions in Faster R-CNN.

[4]`https://paperswithcode.com/method/fast-r-cnn`

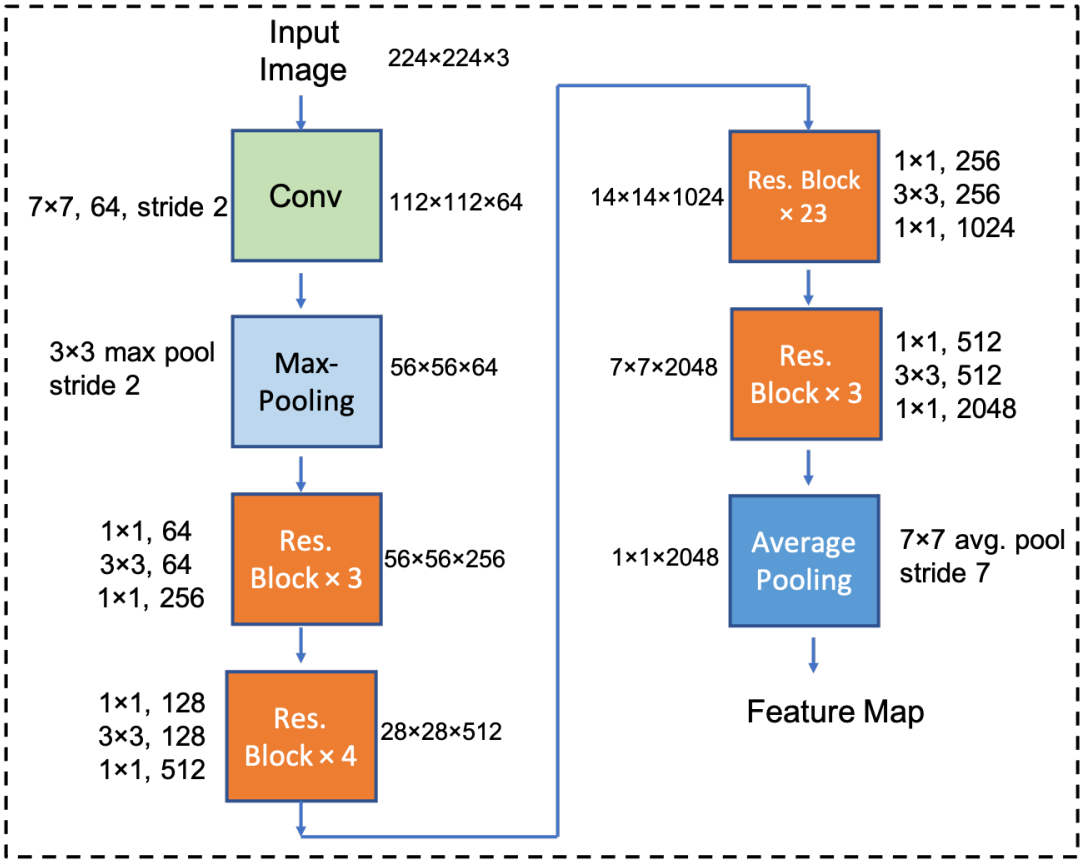


FIGURE 5.3: Resnet101 model architecture.

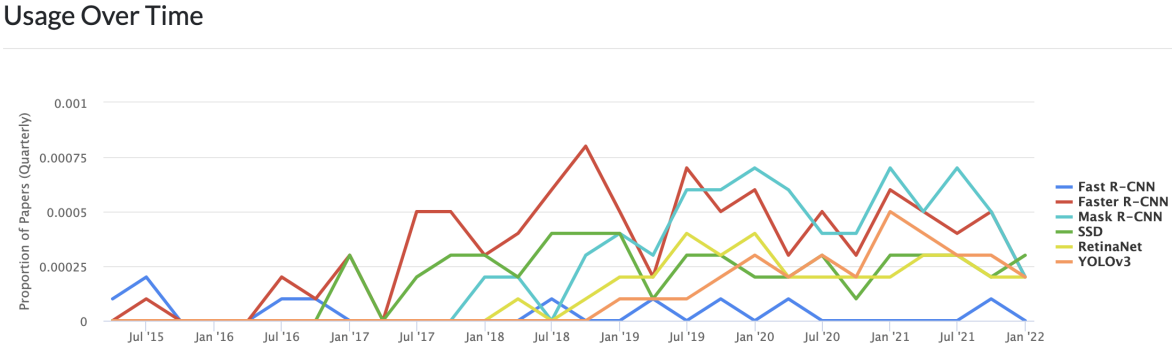


FIGURE 5.4: Usage over time of object detection and localisation algorithms. Mask and Faster R-CNN show the most frequent used algorithms over time.

A fully convolutional network that predicts object boundaries and objectness score values is known as a Region Proposal Network (RPN) [311]. The RPN is a compact and light neural network that examines an input image in a sliding window style to find regions with objects. Anchors are boxes distributed by the RPN throughout an image's feature map to bind the features to their raw image position. For each anchor, the RPN produces two outputs: anchor class and bounding box refinement. The anchor class considers either the foreground (FG) or background (BG) classes. The presence of an object in that box is indicated by the FG class. The foreground anchor, also known as the positive anchor, is appropriately centred above the object thanks to bounding box refinement. Consequently, the RPN evaluates and estimates the change in width and height so that the anchor box can be fine-tuned to fit the object better. Based on RPN predictions, the top anchors most likely to retain objects are selected, and their size location are adjusted. When there are several anchors that are too close together and overlapped, the one with the highest foreground result is kept, while the others are removed using the Non-max Suppression scheme [312].

Non-max Suppression scheme [312] can be accomplished by computing the Intersection over Union (IoU) with the ground truth boxes using the formula: $IoU = Area\ of\ the\ intersection\ /\ Area\ of\ the\ union$. If the IoU is greater than or equal to 0.7, the region is then considered a region of interest. Otherwise, the particular region is neglected. Figure 5.5 illustrates the idea of generating region proposal/anchor boxes that have various scales and aspect ratios.

The RPN examines the backbone feature maps produced by ResNet101 to find suitable region proposals/ anchors. For each anchor, RPN generates two outputs: the class of anchor which is foreground or background and refinement of bounding box (i.e. improve the placement of anchor box). The top nominated anchors of bounding boxes are then chosen since they are most likely to include face objects. An ROI classifier and bounding box regressor assess the ROI class, size, and position of the bounding boxes to check for foreground areas (face) and fine-tune the bounding box's location and size.

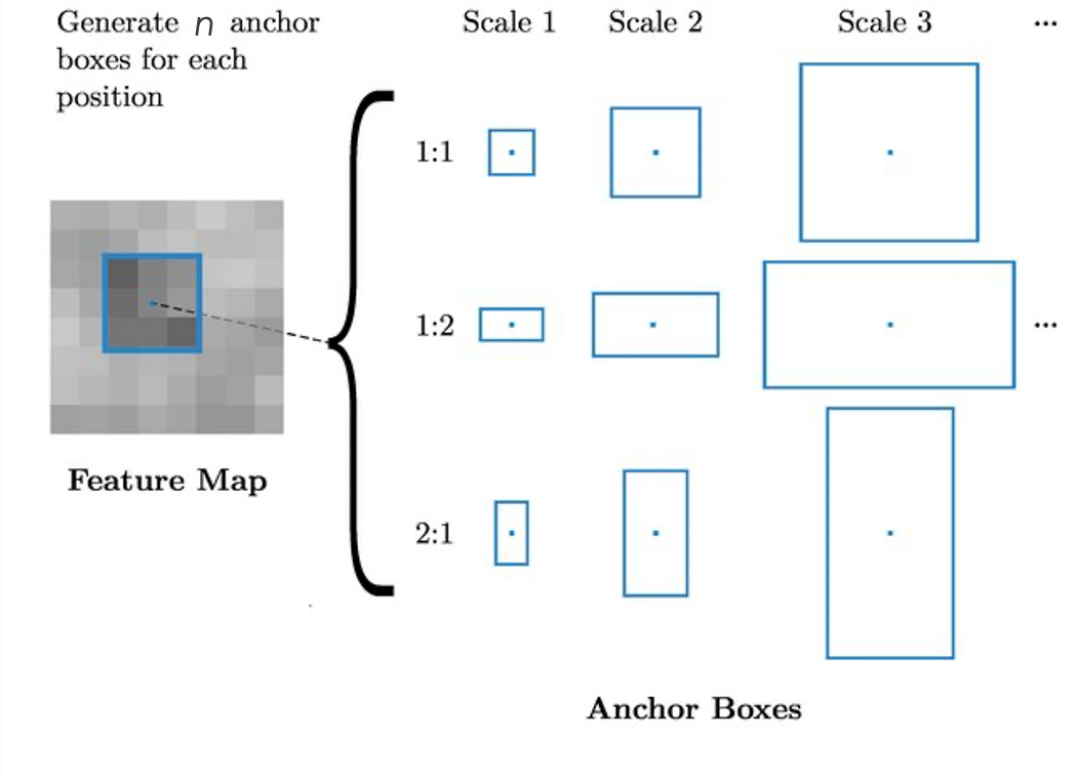


FIGURE 5.5: Anchor Boxes at a particular position in the feature map. The generated anchor boxes have various scales and aspect ratios.

ROI Alignment by ROIAlign

Because the candidate bounding boxes (ROIs) created by the RPN have varying sizes, the dimensions of the ROIs are rectified by using ROIAlign to have identical dimensions. The ROIAlign approach samples the feature map at various positions before applying a bilinear interpolation scheme to construct feature maps that have a fixed size [313], as shown in Figure 5.6. These feature maps are reshaped to be passed through a dense layer (1024 nodes) and then passed into a classification model to make a decision whether the ROI is the positive or negative region. Eventually, the facial positive regions generated by the classifier model is fed to the next stage.

For example, let's have 5×5 feature map is represented with dotted lines in Figure 5.6. Given a rectangle proposal, which is separated into four bins, is placed on this feature map at a certain point. It is not feasible to directly use max-pooling to each

bin due to the region proposal misalignment on the top of feature maps. Thus, in each bin, four points are sampled and bilinear interpolation is used to determine values at those locations, resulting in four nearest point values on the top of feature map. Consequently, in each cell, the largest value among the four is picked. In this scenario, the four cells in this example generated four values. As a result, this could generate the 2×2 feature map that corresponds to the region proposal [309].

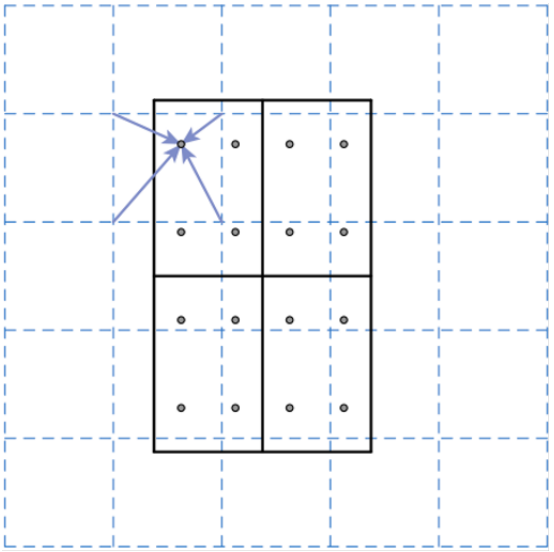


FIGURE 5.6: ROIAlign Strategy [309].

Segmentation Mask Generation

The facial region assigned a positive label by the ROI classifier model becomes the input to the mask branch in the segmentation model, Mask R-CNN, producing the segmentation mask. The FCN in the Mask R-CNN represents the branch that produces the segmentation masks [303]. Thus, it creates masks on the resulted localised facial ROI, resulting in a facial region segmentation. FCN is an end-to-end learning based on pixel-level prediction schemes. The FCN technique replaces the initial fully connected layer in a CNN network with a 1×1 convolutional layer to create a fully convolutional network. FCN can produce and generate a prediction result for each pixel on an image using this technique and then output a prediction mask based on the input image size.

In our study, the positive face ROI is passed through three 3×3 convolutional layers, each followed by batch normalisation and ReLU activation function. The extracted features are then passed through the deconvolution operations with the three deconvolutional layers. This is then subjected to a 1×1 convolutional layer producing the mask of the facial region. From the generated mask, the boundary of facial area can be obtained.

5.4 Results and Discussion

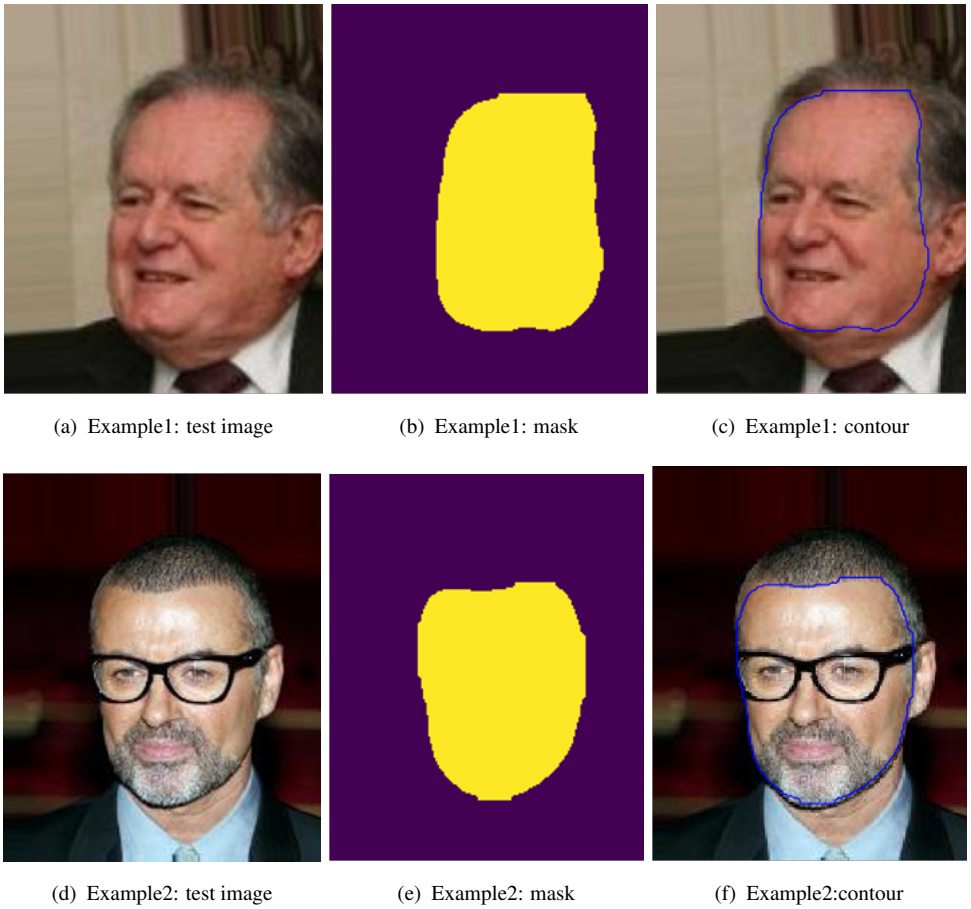
The facial segmentation model is assessed using Jaccard (overlapping), Dice coefficient (F-Measurement), specificity, accuracy, and sensitivity. To set up the experiments, Adam optimiser was used to turn and update the model's weights for 75 epochs. The optimiser has an adjusted learning rate of (10^{-4} - 10^{-6}) with a momentum parameter of 0.9. As a result of the diversity of faces in images, several online image augmentation strategies are adopted during model learning. These strategies are defined by applying the following: i) image scaling with a range of (0.5, 2.0), ii) image rotation with a range of (-90, 90) and (-15, 15), and iii) image cropping randomly to $64 \times 64 \times 3$. The cross-entropy cost function for multi-task class classification is used as a loss function to update the weights of the network. Under the condition of integrating the object localisation and segmentation in the same model, the cross-entropy is adapted to include segmentation, classification, and bounding box localisation losses. This loss function can be defined as follows:

$$L = L_{cls} + L_{bbox} + L_{mask} \tag{5.2}$$

where L_{cls} represents the class loss, L_{bbox} refers to the bounding box loss of the object localiser represented by Faster R-CNN, and L_{mask} represents the mask loss from the fully convolutional network.

The facial segmentation model was assessed on 700 images, producing a dice coefficient of 95.67%, Jaccard score (overlapping) of 91.89%, the accuracy of 98.18%,

sensitivity of 96.14%, and specificity of 98.76%. The obtained results outperform existing automated face segmentation methods, surpassing the state-of-the-art methods such as, Yang et al. [301], Ghiasi et al. [298], and Nirkin et al. [305] who reported overlapping (Jaccard score) of 72.4%, 83.5% and 83.7% on COFW dataset [300], respectively. Our model was not tested on the COFW dataset because the ground truth masks of this dataset are not publicly available. The proposed method is very fast in terms of computational time, taking approximately 0.3s to predict the face mask and draw the contour on a test image. Figure 5.7 presents image examples for mask generation and contour drawing.



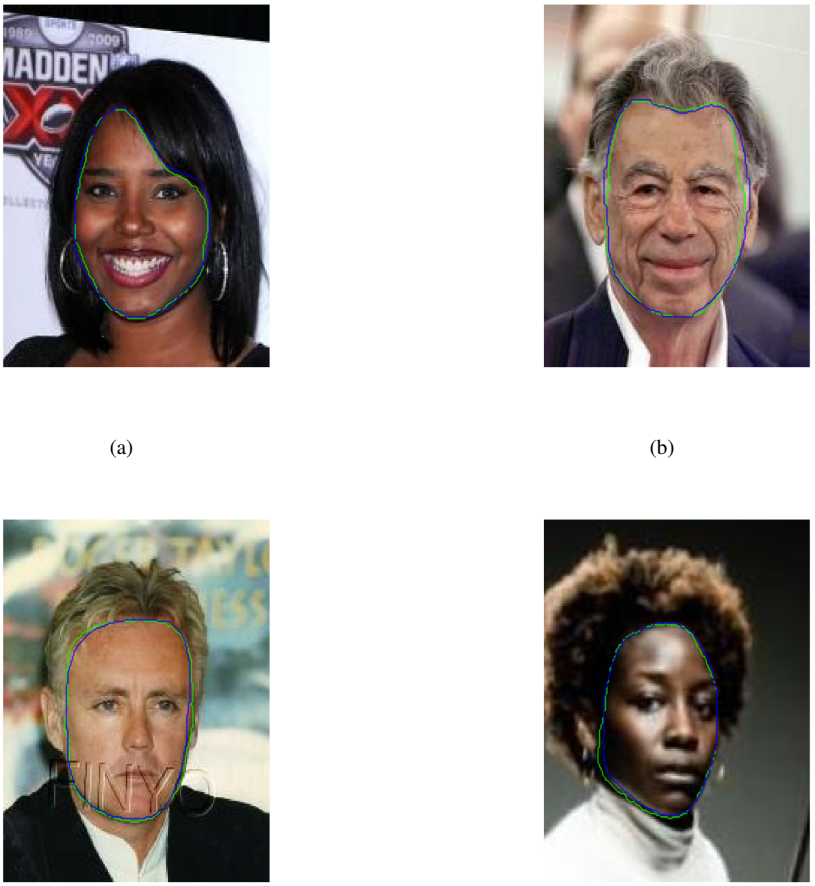
(a) Example1: test image (b) Example1: mask (c) Example1: contour

(d) Example2: test image (e) Example2: mask (f) Example2:contour

FIGURE 5.7: Two image examples show mask generation and contour drawing.

For subjective assessment, Figures 5.8 and 5.9 depict image examples with the localised contour. Figure 5.8 shows correct segmented images came from testing the

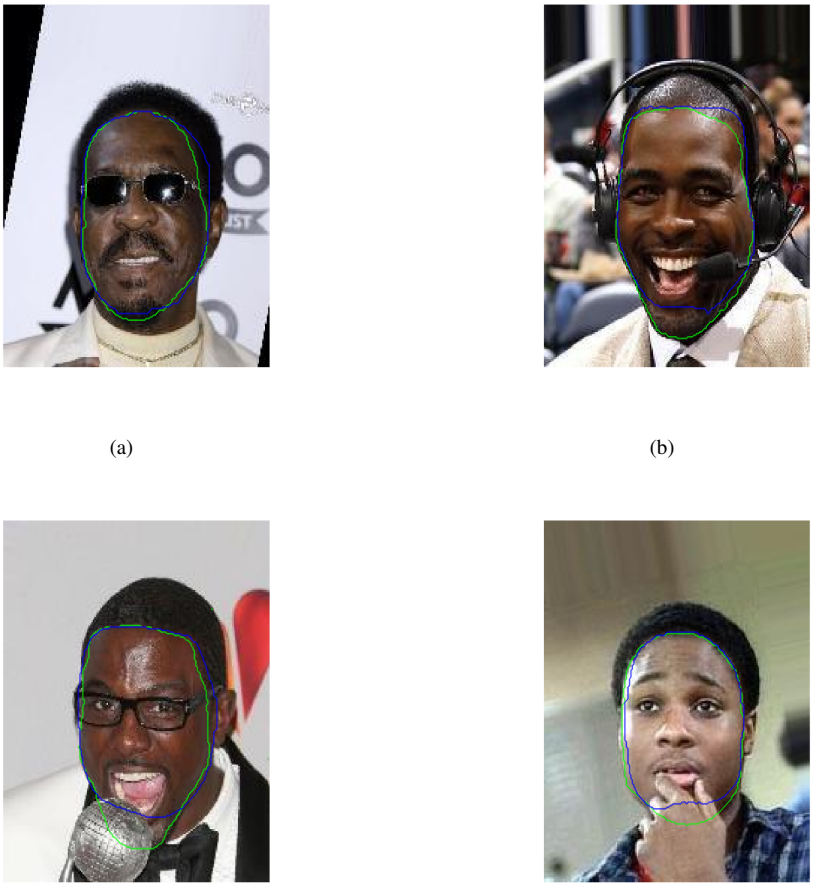
leaned facial segmentation model on test images. Contrary, Figure 5.9 presents examples of images where face segmentation contour is detected and localised with misclassification on some face boundary points. On the other hand, the displayed images show that the presence of obstacles, such as hair covering part of the face, could degrade the performance of the proposed face segmentation model. Further improvement to the presented model is required to overcome this issue by integrating a model that helps to identify the obstacles from the targeted object.



(a) (b)

(c) (d)

FIGURE 5.8: Results of our model (blue contour) are compared with the ground truth annotations (green contour) demonstrating the effectiveness of our model.



(a) (b)

(c) (d)

FIGURE 5.9: Image examples show the test images with some misclassified boundary points.

5.5 Conclusions

In this chapter, a new facial segmentation model from unconstrained image data has been presented. Merging Faster R-CNN with FCN has proven to be efficient and fast in face boundary detection. Incorporating object localisation with segmentation demonstrated its superiority over existing methods in extracting the contour face object in wild facial images. The proposed system has been evaluated on a fairly large and independent dataset that included large diversity of faces. The obtained results demonstrated that the proposed deep learning method is promising in segmenting face regions efficiently and accurately. The proposed object localisation-segmentation framework is general and

could be handily adapted for various image segmentation applications. This system can be further extended by providing a simulation of 3D hairstyles and harnessing augmented reality for designing virtual hairstyle makeovers. This could be accomplished by designing a mobile application that could help users find the hairstyle and haircut that perfectly fit them.

Chapter 6

Conclusions and Future Work

Throughout the work presented in this thesis, efficient computer-based facial analysis and recommendation systems targeting beautification and cosmetic purposes have been developed based on computational intelligence methodologies. Our developed methods studied and suggested recommendations for beautification based on the eye of the machine and deep learning algorithm. In this chapter, we wrap up our project and offer some suggestions for future study based on the findings of this thesis.

6.1 Detailed Conclusions

Several automated analysis approaches have been designed and developed to examine and explore various facial image analysis tasks including, feature extraction, segmentation, detection, and classification. The establishment of numerous deep learning methodologies based on Convolutional Neural Networks (CNNs) alongside handcrafted features, landmark detection, and image processing techniques was investigated to accomplish the facial image analysis tasks.

In the first focus of this thesis, a decision support program (recommendation system) has been developed to constitute an aid system that automatically analyses eye

and face features based on the image taken from a user. The system suggested a suitable recommendation of eyelashes type and hairstyle based on the automatic reported users' eye and face features. To achieve the aim, a multi-model system comprising three separate models; each model targeted a different task, including face shape classification, eye attribute identification and gender detection mode, has been developed. Face shape classification system has been designed based on developing a hybrid framework of hand-crafted and learned features. Eye attributes have been identified by exploiting geometrical eye measurements using the detected eye landmarks. Gender identification system has been realised and designed by implementing a deep learning-based approach. The outputs of three developed models are merged to design a recommendation (decision-support) system for haircut and eyelash extension recommendation. The developed system has been evaluated on publicly available data labelled by a beauty expert. The obtained detection results demonstrated that the proposed method effectively identifies face shape and eye attributes. Identifying human face shape and eye attributes is the most vital process before applying the right hairstyle and eyelashes extension. The developed system could help workers in beauty centres reduce their workload by automating their manual work and producing objective results, particularly with a large dataset.

Makeup could disguise the extracted facial features, which results in degradation in the performance of many facial-related analysis systems, including face recognition, facial landmark characterisation, aesthetic quantification, automated age estimation methods, and facial feature-driven recommendation systems. Thus, facial makeup is likely to affect directly several real-life applications such as cosmetology and virtual cosmetics recommendation systems, security and access control, and social interaction. In the second focus of this thesis, we designed automated facial makeup detection systems leveraging multiple learning schemes from a single unconstrained photograph. We have investigated and studied the efficacy of deep learning models for makeup detection incorporating the use of a transfer learning strategy with semi-supervised learning using

labelled and unlabelled data. First, during the supervised learning, the VGG16 convolution neural network, which is pre-trained on a large dataset, is fine-tuned on makeup labelled data. Secondly, during semi-supervised learning, two unsupervised learning methods, which are self-learning and auto-encoder, are trained on unlabelled data and then incorporated with supervised learning. Comprehensive experiments and comparative analysis have been conducted on 2479 labelled images and 446 unlabelled images collected from six challenging makeup datasets. The obtained results revealed that auto-encoder merged with supervised learning gives the best makeup detection performance. The promising results obtained from conducted experiments revealed and reflected the efficiency of combining different learning strategies for makeup detection by harnessing labelled and unlabelled data. It would also be advantageous to the beauty industry to develop such computational intelligence methods.

In the third focus of the thesis, we laid the groundwork to extend the hairstyle and eyelashes recommendation system. The extension could be accomplished by synthesising the hairstyle and applying the recommended hairstyle virtually. Before applying the hairstyle virtually, the prerequisite step is to recognise the face boundary region from the hair and extract the face contour accurately. Thus, this research theme targeted to introduce a deep feature learning approach to segment face boundaries from unconstrained facial images. The presented deep learning model defined the detection of the face boundary formula as combined object localisation and segmentation task. Integrating two computer vision tasks in one model, represented by localisation and segmentation, assists in improving the segmentation performance accomplished by the fully convolutional network. The weights of the proposed model are updated on 2800 in-the-wild face images and tested on 700 images. The findings demonstrated that the presented model outperforms most state-of-the-art methods in automatic face detection and segmentation on a challenging dataset. The findings have proven the presented face segmentation framework is robust, accurate and reliable for face boundary detection.

To summarise, as people adopt a more holistic approach to beauty that includes

health and wellness, the beauty and cosmetics business responds by incorporating technology into its products. Artificial intelligence would be a real game-changer in the beauty business, revolutionising the products personalisation experience, with the goal of supporting decisions and guidance using knowledge and data-intensive computer-based remedies.

6.2 Future Work

In this section, some of the challenges that still need to be overcome are described in the following list as potential future research directions:

- The developed GUI desktop software, presented in Chapter 3 can be extended to build a cloud-based decision support system using a mobile application. The extended system targets users' facial and eye analysis via cloud server programs based on the images captured and uploaded from mobile phones to the server.

- The basic GUI designed for our recommendation system requires further improvements. Careful placement and shape selection of input controls, including buttons and text fields, can aid scanning and readability. Furthermore, wisely choosing colour, light, contrast, and texture may all be used to assist in bringing attention to essential items in the window. These improvements have the potential to improve and speed up the interaction with the recommendation system.

- The detected facial and eye attributes extracted in Chapter 3 could be harnessed to develop a makeup recommendation system.

- In Chapter 3, another important aspect could be considered as a potential future work which is studying the computational complexity of the three models to achieve the best and fastest convergence in the developed framework.

- Conducting an exhaustive analysis study for the face shape classification model, by assessing the model detection performance under the five webcam positions, needs to be considered in Chapter 3.

- Following up the findings of Chapter 4, a makeup removal system that removes the makeup from images, rather than omitting them before passing to the recommendation system, could be developed. The research covering makeup removal is still immature. The research works presented in the literature show some progress, for instance, [250, 254, 255, 314–318]. However, all these methods suggest transferring and adding the makeup to bare facial images and then removing it. The research work problem dealt with here is instead seeking to remove pre-existing makeup on a facial image. Generative adversarial networks (GANs) [319, 320] could be adapted for facial makeup removal from images via generating synthesis without-makeup images. The generator's input of GANs could be the images with makeup, while the ground truth is images without makeup, and the generated output images are without makeup facial images.

- This face segmentation method presented in Chapter 5 could be further extended by providing a simulation of 3D hairstyles and harnessing augmented reality (AR) techniques for designing virtual hairstyle makeovers (try-on). This could be accomplished by designing a mobile application to help users find the hairstyle and haircut that perfectly fit them. While developing virtual try-on systems might be feasible, parametrising a rendering engine to synthesising realistic representations of a particular hairstyle/haircut is a time-consuming process requiring specialised computer graphics knowledge. Thus, this challenge could be considered as potential future research work.

- The presence of obstacles, such as hair covering part of the face, could degrade the performance of the proposed face segmentation model presented Chapter 5. Further improvement to the presented model is required to overcome this issue by integrating a model that helps to identify the obstacles from the targeted object.

6.3 Overall Conclusion

Overall, this work has successfully blended skills and experience from several disciplines, such as beauty and aesthetics, artificial intelligence, computer vision and image processing, to produce facial image analysis systems. Our successful results would undoubtedly support the efficacy of various artificial intelligence techniques to compute facial attributes, analyse images automatically and make a suitable recommendation. The developed AI-powered recommendation engine has great potential for practical deployment, targeting less subjective assessment and speeding up the consultation process for facial attributes. Hopefully, this would significantly reduce the number of required consultation visits and customers' long waits, hence enhancing customer satisfaction and saving expenses at beauty salons. The deep learning-based methodologies provided here will also be valuable tools for solving other demanding real-world situations.

Appendix A

Recommendation Texts Displayed on GUI

The guidelines from specialised websites [91, 92, 239] are followed to implement the rules of the recommendation engine for hairstyle and eyelashes in our proposed system. The recommendations are quoted as follows:

A.1 Eyelash Extension Recommendation

If eye shape is almond, display this recommendation message:

"Lash extension recommendations is Doll eye shape eyelash. This style involves shorter lashes placed on the inner and outer corners, whilst longer lashes are placed along the middle of the eye. It creates a flattering opened-up look and freshens up the face"

If eye shape is round, display this recommendation message:

"Lash extension recommendations is Cat eye shape eyelash. This style features shorter lengths gradually lengthening to long lashes on the corner of the eyes. It balances out the eye shape and creates a sultry look"

A.2 Women's Hairstyle Recommendation

- Round: "Recommendations: Long layered cuts, choppy pixie cuts, short side bangs. Round face shapes typically feature a similar length and width as well as prominent, rounded cheeks. Therefore, the most flattering hairstyles for ladies with this face shape are those that add definition and shape. In particular, long layered cuts and choppy pixie cuts tend to look best. If opting for a layered cut, choose long, staggered layers that start around the jawline. You may also want to add a short side fringe that finishes at the eye to help lengthen your face. If you're wondering what to avoid, steer clear of bobs and single-length cuts along with short layers, which can balloon around your face, and full curls, which can also enhance the appearance of roundness"

- Oval: "Recommendations: Blunt bobs and lobs with subtle layers, long waves or curls. You have an oval face shape; you should consider yourself lucky. Because of their well-balanced appearance, oval faces can suit a wide variety of hairstyles. However, if you're looking for a style with impact, you should consider opting for either long locks or a cute, short crop. For oval-faced ladies who like short hair, a blunt bob or lob with subtle layers will look particularly chic. If you prefer long hair, on the other hand, opt for minimal layers or style your hair with waves or curls to keep it from dragging down your face. You can also use your natural hair part to help guide you in your hairstyle choice"

- Square: "Recommendations: Side-parted styles, long and airy layers, short layered bobs, side-swept bangs. A square-shaped face features a broad forehead, wide cheekbones, and a strong jawline. For a square face, selecting a cut that softens these features is essential for a flattering appearance. For a look that complements your bone structure, consider a side-parted style, which will offset the squareness. Long and airy layers can also be flattering and will help to disguise

the sharp angles of your face. If you like a cropped length, consider a short, layered bob. But, remember to add side-swept bangs, which will draw attention to your cheekbones instead of your jawline"

- Rectangle: "Recommendations: Layered cuts, waves or curls, soft and romantic chignons, rounded fringes or curtain bangs. You should work to soften the appearance of your sharp jaw and forehead without further elongating the face. A soft layered cut, for example, can enhance cheekbones while disguising the corners of the face. However, you should be careful to avoid overly long styles, which can further elongate your appearance. If you do opt for a long length, try styling your hair with a blowout, waves or curls, which will add horizontal volume and soften the sharp angles of your face. As for updos, choose soft and romantic chignons over high buns, which will add length, and sleek styles, which will emphasize your strong jaw. When it comes to bangs, preference soft, rounded fringes and curtain bangs over square styles and blunt-cut options"

- Heart-Shaped: "Recommendations: Long side-swept cuts, waves or curls starting below the ear, side-parted pixies, bobs and lobs. With a heart-shaped face, your bone structure features a broad forehead and cheekbones with a narrow jawline and chin. As such, you should aim to balance your face shape with a cut that decreases your brow width and increases the width of the lower half of your face. For example, a long side-swept cut will disguise part of your forehead while drawing eyes down to the bottom of your face. You can also pair this cut with waves or curls starting below the ear to even out your angled jaw. Alternatively, a side-parted pixie cut with textured ends can also appear gorgeous on a heart face shape, as can a bob or lob, which will ensure fullness around your jaw"

A.3 Men's Hairstyle Recommendation

- Round: "Recommendations: Side part, French crop, pompadour. Your hairstyle needs to create angles, given round faces naturally, lack shape and dimension. To elongate your face, try out a hairstyle with longer proportions up top and shorter sides. A pompadour will do just that, working to create more dimension and giving your baby face a more pronounced look. Note that a side part will also create more angles and give you a more mature look"

- Oval: "Recommendations: Pompadour, side part or quiff. You're in luck. Oval face shapes are proportioned well and can handle a broad range of different hairstyles and lengths. Given most hairstyles suit this face shape, it is all about your personal preference with how you style your hair. A conservative quiff works well, but there is also a lot of room to experiment with longer lengths and modernised cuts. Your face shape can handle a voluminous pompadour or even longer, fuller rocker-esque shoulder-length hair. An oval face shape also works well with a shaved head"

- Square: "Recommendations: Undercut, quiff or side part. You're working with a strong jawline, which should be emphasized. It is recommended opting for an undercut, which will make the most of your distinctive features. As this face shape is versatile with different types of hair, don't be afraid to experiment with other lengths and styles. It is recommended a lot of volume when styling your hair, so focus on fullness and add as much product as you desire – this face shape can handle a significant amount of hair. A quiff works well, whether done more classically or with a modern edge. You can also try out deeper side parts which will enhance your features"

- Rectangle (Oblong): "Recommendations: slicked back or side part. Concerning hairstyles, you need to be conscious that you do not further elongate your face. The trick is to ensure that your face doesn't look longer than it already appears.

Pompadours avoided – you will need to opt for more balanced hairstyles without too much distinction between the sides and the top of your hair. A side part will work well, working to lend a bit of a dimension to a shorter hair length. A slicked back hairstyle or a sleek man bun will also better balance out the proportions of your face. Keep in mind that when you are styling your hair, ensure that everything looks neat as you will not want to add too much volume. Remember to stick to lightweight pomades that will keep your hair looking sharp all day but won't add excessive height to your longer face shape"

- Heart-Shaped: "Recommendations: fringe or quiff. When choosing a hairstyle for your heart-shaped face, you will want to balance out the top part of your face with the bottom, so keep your hair proportionate to your features. You should focus on trying to ensure that your chin does not appear too narrow. A textured fringe or dimensional quiff will do the job. you should use a lightweight product to ensure that you are not creating disproportionate angles"

A.4 Image Examples of Recommendation

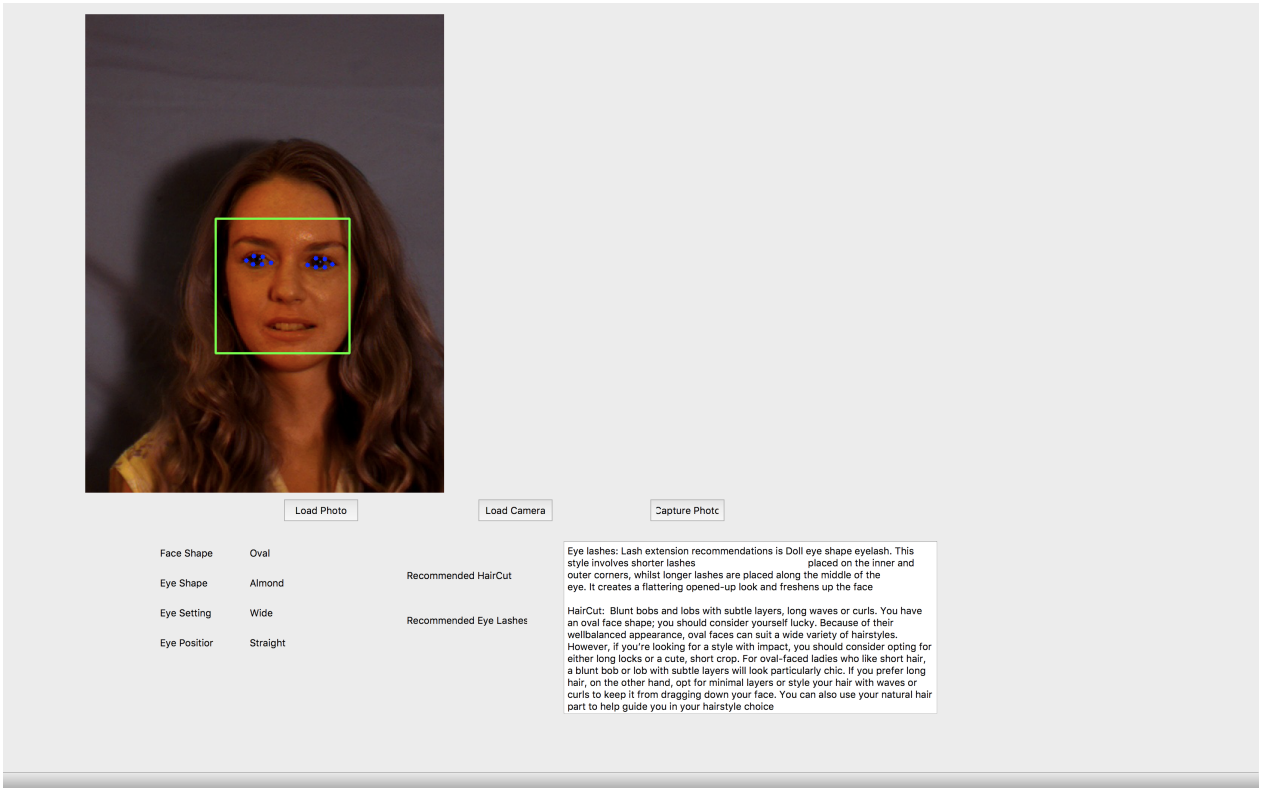


FIGURE A.1: GUI displays facial and eye attributes predication and recommendation: Example1. Recommendation text displayed on this GUI is taken from [91, 92]

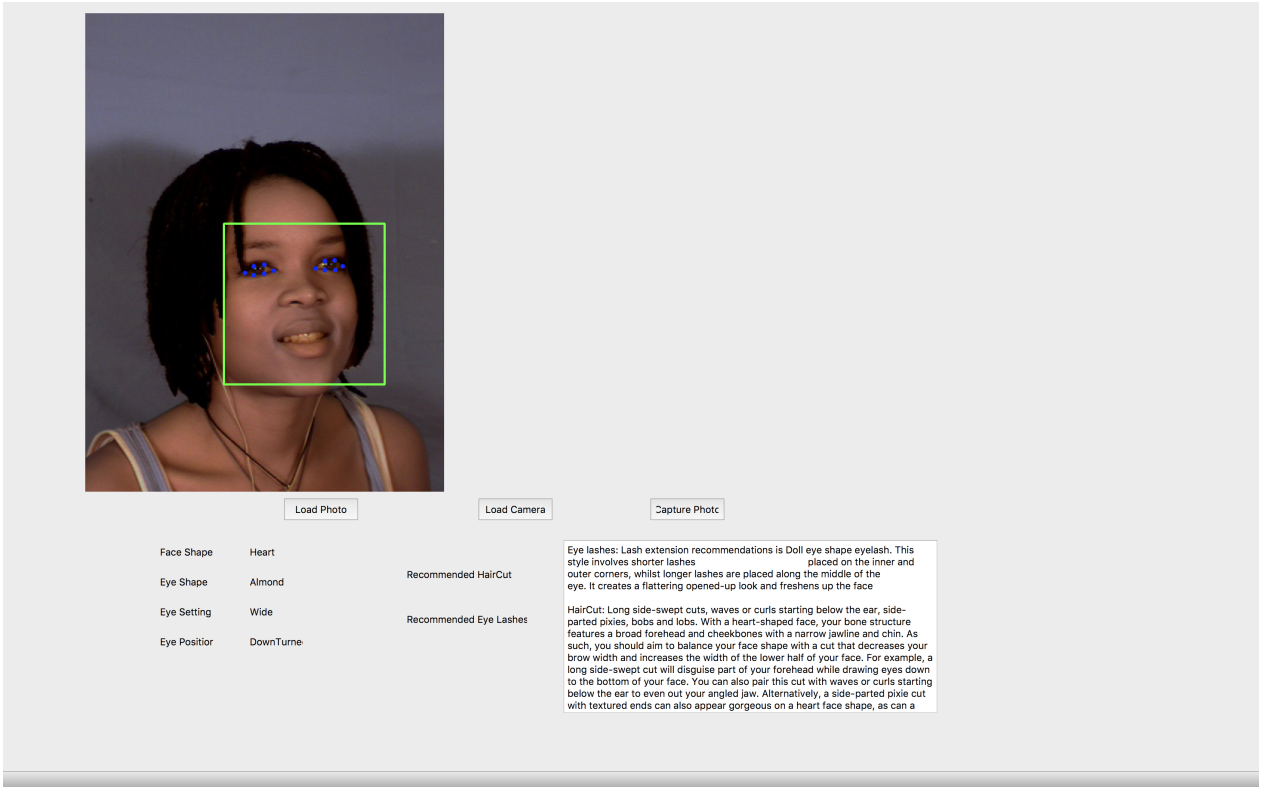


FIGURE A.2: GUI displays facial and eye attributes predication and recommendation: Example2. Recommendation text displayed on this GUI is taken from [91, 92]

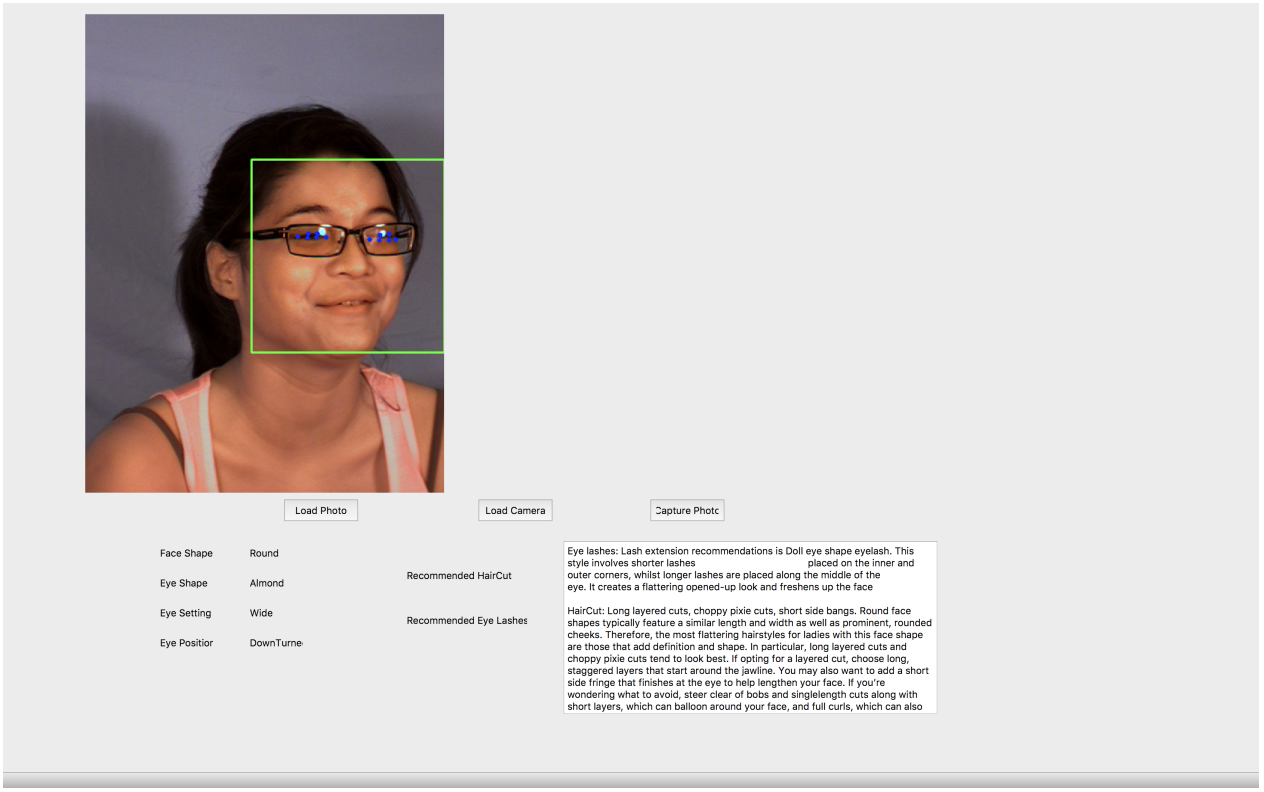


FIGURE A.3: GUI displays facial and eye attributes predication and recommendation: Example3. Recommendation text displayed on this GUI is taken from [91, 92]

FIGURE A.4: GUI displays facial and eye attributes predication and recommendation: Example4. Recommendation text displayed on this GUI is taken from [91, 92]

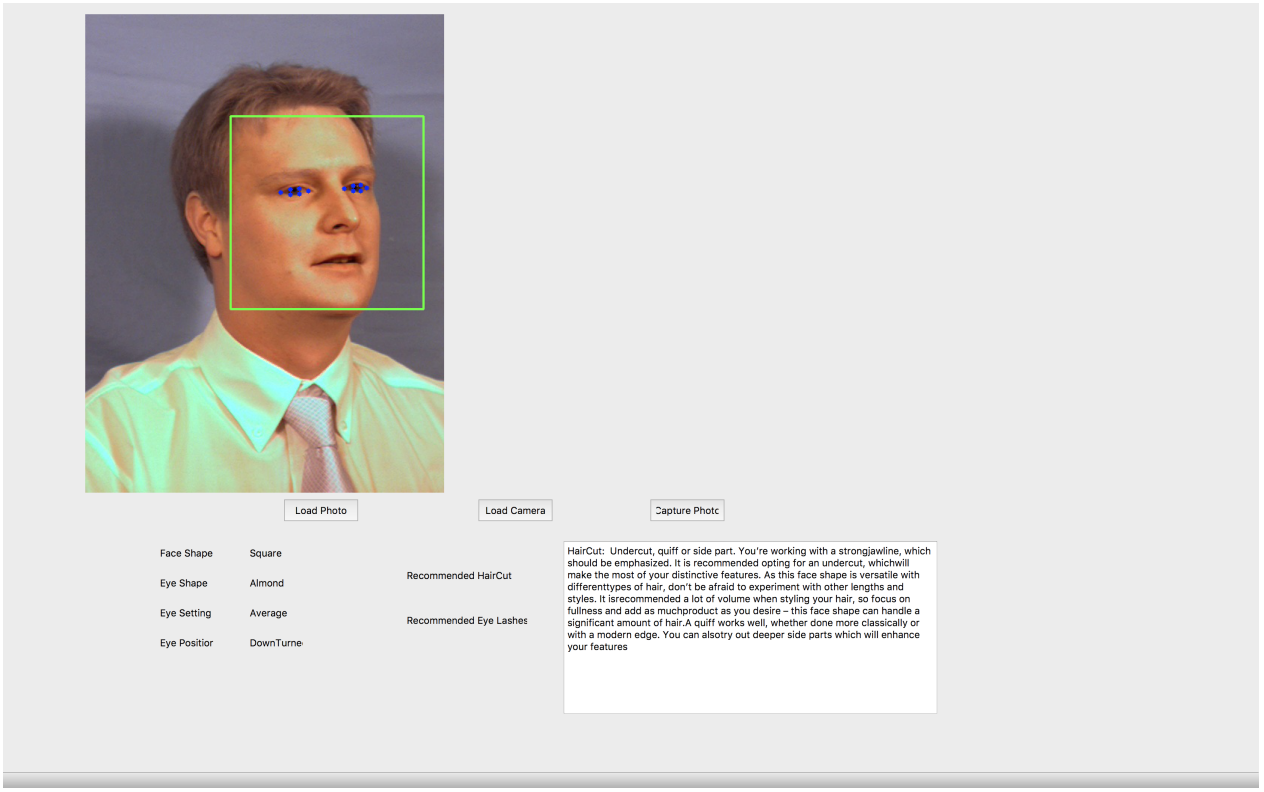


FIGURE A.5: GUI displays facial and eye attributes predication and recommendation: Example5. Recommendation text displayed on this GUI is taken from [91, 92]

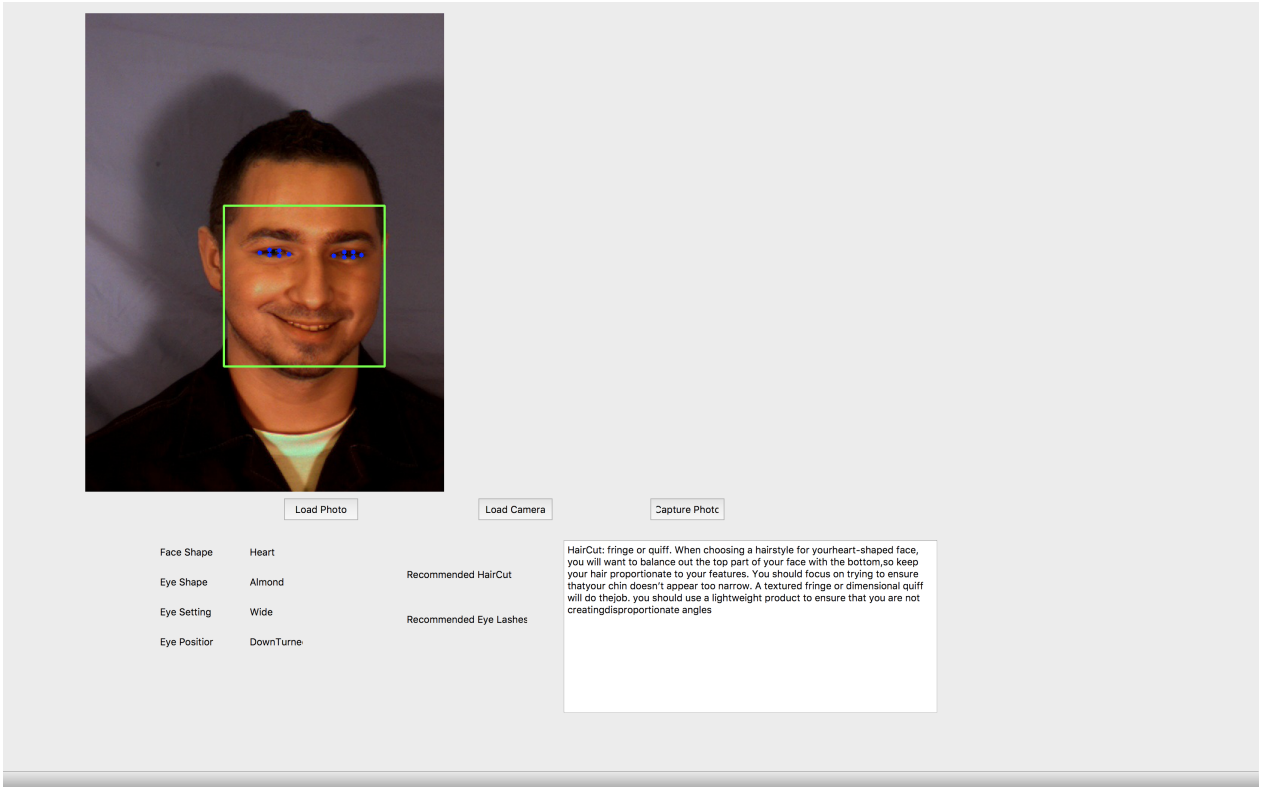


FIGURE A.6: GUI displays facial and eye attributes predication and recommendation: Example6. Recommendation text displayed on this GUI is taken from [91, 92]

Bibliography

[1] C. Rathgeb, A. Dantcheva, and C. Busch, "Impact and detection of facial beautification in face recognition: An overview," *IEEE Access*, vol. 7, pp. 152667–152678, 2019.

[2] D. Jones, *Physical attractiveness and the theory of sexual selection: Results from five populations*, vol. 90. University of Michigan Museum, 1996.

[3] J. E. Scheib, S. W. Gangestad, and R. Thornhill, "Facial attractiveness, symmetry and cues of good genes," *Proceedings of the Royal Society of London. Series B: Biological Sciences*, vol. 266, no. 1431, pp. 1913–1917, 1999.

[4] G. Rhodes and T. Tremewan, "Averageness, exaggeration, and facial attractiveness," *Psychological Science*, vol. 7, no. 2, pp. 105–110, 1996.

[5] G. Rhodes, J. Chan, L. A. Zebrowitz, and L. W. Simmons, "Does sexual dimorphism in human faces signal health?," *Proceedings of the Royal Society of London. Series B: Biological Sciences*, vol. 270, no. 1, pp. S93–S95, 2003.

[6] A. C. Little, D. M. Burt, and D. I. Perrett, "What is good is beautiful: Face preference reflects desired personality," *Personality and Individual Differences*, vol. 41, no. 6, pp. 1107–1118, 2006.

[7] P. Danziger, "6 trends shaping the future of the usd 532b beauty business." `https://www.forbes.com/sites/pamdanziger/2019/09/01/6-trends-shaping-the-future-of-the-532b-beauty-business/?sh=28fb988588d7`. Accessed: 2021-08-5.

[8] T. Leyvand, D. Cohen-Or, G. Dror, and D. Lischinski, "Data-driven enhancement of facial attractiveness," in *ACM SIGGRAPH 2008 Papers*, pp. 1–9, Association for Computing Machinery, New York, United States, 2008.

[9] S. Melacci, L. Sarti, M. Maggini, and M. Gori, "A template-based approach to automatic face enhancement," *Pattern Analysis and Applications*, vol. 13, no. 3, pp. 289–300, 2010.

[10] K. Scherbaum, T. Ritschel, M. Hullin, T. Thormählen, V. Blanz, and H.-P. Seidel, "Computer-suggested facial makeup," in *Computer Graphics Forum*, no. 2, pp. 485–492, Wiley Online Library, 2011.

[11] Q. Liao, X. Jin, and W. Zeng, "Enhancing the symmetry and proportion of 3d face geometry," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 10, pp. 1704–1716, 2012.

[12] D. Zhang, F. Chen, Y. Xu, *et al.*, *Computer models for facial beauty analysis*. Springer, 2016.

[13] X. Zeng, W. Ouyang, B. Yang, J. Yan, and X. Wang, "Gated bi-directional CNN for object detection," in *European Conference on Computer Vision*, pp. 354–369, Springer, 2016.

[14] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 815–823, 2015.

[15] F. Song, X. Tan, and S. Chen, "Exploiting relationship between attributes for improved face verification," *Computer Vision and Image Understanding*, vol. 122, pp. 143–154, 2014.

[16] O. K. Manyam, N. Kumar, P. Belhumeur, and D. Kriegman, "Two faces are better than one: Face recognition in group photographs," in *2011 International Joint Conference on Biometrics (IJCB)*, pp. 1–8, IEEE, 2011.

[17] Y.-H. Lei, Y.-Y. Chen, L. Iida, B.-C. Chen, H.-H. Su, and W. H. Hsu, "Photo search by face positions and facial attributes on touch devices," in *Proceedings of the 19th ACM International Conference on Multimedia*, pp. 651–654, 2011.

[18] J. Bekios-Calfa, J. M. Buenaposada, and L. Baumela, "Robust gender recognition by exploiting facial attributes dependencies," *Pattern Recognition Letters*, vol. 36, pp. 228–234, 2014.

[19] D. Mhlanga, "Artificial intelligence in the industry 4.0, and its impact on poverty, innovation, infrastructure development, and the sustainable development goals: Lessons from emerging economies?," *Sustainability*, vol. 13, no. 11, p. 5788, 2021.

[20] Q. Zhang, J. Lu, and Y. Jin, "Artificial intelligence in recommender systems," *Complex & Intelligent Systems*, vol. 7, no. 1, pp. 439–457, 2021.

[21] X. Han, Z. Wu, Z. Wu, R. Yu, and L. S. Davis, "Viton: An image-based virtual try-on network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7543–7552, 2018.

[22] R. Kips, R. Jiang, S. Ba, E. Phung, P. Aarabi, P. Gori, M. Perrot, and I. Bloch, "Deep graphics encoder for real-time video makeup synthesis from example," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3889–3893, 2021.

[23] A. Laurentini and A. Bottino, "Computer analysis of face beauty: A survey," *Computer Vision and Image Understanding*, vol. 125, pp. 184–199, 2014.

[24] X. Liu, "Facial attributes analysis and applications," Master's thesis, West Virginia University, USA, 2018.

[25] A. C. Little, B. C. Jones, and L. M. DeBruine, "Facial attractiveness: evolutionary based research," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 366, no. 1571, pp. 1638–1659, 2011.

[26] R. E. Riggio and S. B. Woll, "The role of nonverbal cues and physical attractiveness in the selection of dating partners," *Journal of Social and Personal Relationships*, vol. 1, no. 3, pp. 347–357, 1984.

[27] E. Berscheid, K. Dion, E. Walster, and G. W. Walster, "Physical attractiveness and dating choice: A test of the matching hypothesis," *Journal of Experimental Social Psychology*, vol. 7, no. 2, pp. 173–189, 1971.

[28] T. F. Cash and R. N. Kilcullen, "The aye of the beholder: susceptibility to sexism and beautyism in the evaluation of managerial applicants," *Journal of Applied Social Psychology*, vol. 15, no. 4, pp. 591–605, 1985.

[29] Y.-F. Wang, Y.-L. Chuang, M.-H. Hsu, and H.-C. Keh, "A personalized recommender system for the cosmetic business," *Expert Systems with Applications*, vol. 26, no. 3, pp. 427–434, 2004.

[30] K.-Y. Chung, "Effect of facial makeup style recommendation on visual sensibility," *Multimedia Tools and Applications*, vol. 71, no. 2, pp. 843–853, 2014.

[31] Z. Liang and F. Wang, "Research on the personalized recommendation algorithm for hairdressers," *Open Journal of Modelling and Simulation*, vol. 4, no. 3, pp. 102–108, 2016.

[32] T. Alashkar, S. Jiang, and Y. Fu, "Rule-based facial makeup recommendation system," in *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, pp. 325–330, IEEE, 2017.

[33] K. Pasupa, W. Sunhem, and C. K. Loo, "A hybrid approach to building face shape classifier for hairstyle recommender system," *Expert Systems with Applications*, vol. 120, pp. 14–32, 2019.

[34] Z. Zhao, L. Zhou, and T. Zhang, "Intelligent recommendation system for eyeglass design," in *International Conference on Applied Human Factors and Ergonomics*, pp. 402–411, Springer, 2019.

[35] F. Bastien, P. Lamblin, R. Pascanu, J. Bergstra, I. Goodfellow, A. Bergeron, N. Bouchard, D. Warde-Farley, and Y. Bengio, "Theano: new features and speed improvements," *arXiv preprint arXiv:1211.5590*, 2012.

[36] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *Proceedings of the 22nd ACM International Conference on Multimedia*, pp. 675–678, ACM, 2014.

[37] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, *et al.*, "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," *arXiv preprint arXiv:1603.04467*, 2016.

[38] R. Collobert, K. Kavukcuoglu, and C. Farabet, "Torch7: A matlab-like environment for machine learning," in *BigLearn, NIPS Workshop*, no. EPFL-CONF-192376, 2011.

[39] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097–1105, 2012.

[40] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2818–2826, 2016.

[41] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[42] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, 2016.

[43] P. Karatzas, Y. Kiouvrekis, P. Stefaneas, and H. Sarimveis, "An image based approach for predicting the effects of endocrine disrupting chemicals on human health using deep learning," *medRxiv*, 2020.

[44] R. Hong, W.-H. Cheng, T. Yamasaki, M. Wang, and C.-W. Ngo, *Advances in Multimedia Information Processing–PCM 2018: 19th Pacific-Rim Conference on Multimedia, Hefei, China, September 21-22, 2018, Proceedings, Part III*, vol. 11166. Springer, 2018.

[45] S. Minaee, Y. Y. Boykov, F. Porikli, A. J. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

[46] B. Atiyeh and S. Hayek, "Numeric expression of aesthetics and beauty," *Aesthetic Plastic Surgery*, vol. 32, no. 2, pp. 209–216, 2008.

[47] F. Chen and D. Zhang, "Evaluation of the putative ratio rules for facial beauty indexing," in *2014 International Conference on Medical Biometrics*, pp. 181–188, IEEE, 2014.

[48] J. G. Martin, "Racial ethnocentrism and judgment of beauty," *The Journal of Social Psychology*, vol. 63, no. 1, pp. 59–63, 1964.

[49] E. Walster, V. Aronson, D. Abrahams, and L. Rottman, "Importance of physical attractiveness in dating behavior.," *Journal of Personality and Social Psychology*, vol. 4, no. 5, p. 508, 1966.

[50] A. Feingold, "Matching for attractiveness in romantic partners and same-sex friends: A meta-analysis and theoretical critique.," *Psychological Bulletin*, vol. 104, no. 2, p. 226, 1988.

[51] A. Feingold, "Good-looking people are not what we think.," *Psychological Bulletin*, vol. 111, no. 2, p. 304, 1992.

[52] D. I. Perrett, K. A. May, and S. Yoshikawa, "Facial shape and judgements of female attractiveness," *Nature*, vol. 368, no. 6468, pp. 239–242, 1994.

[53] E. Sala, M. Terraneo, M. Lucchini, and G. Knies, "Exploring the impact of male and female facial attractiveness on occupational prestige," *Research in Social Stratification and Mobility*, vol. 31, pp. 69–81, 2013.

[54] H. Huh, D. Yi, and H. Zhu, "Facial width-to-height ratio and celebrity endorsements," *Personality and Individual Differences*, vol. 68, pp. 43–47, 2014.

[55] K. S. Choe, A. P. Sclafani, J. A. Litner, G.-P. Yu, and T. Romo III, "The korean american woman's face," *Archives of Facial Plastic Surgery*, 2004.

[56] J. Fan, K. Chau, X. Wan, L. Zhai, and E. Lau, "Prediction of facial attractiveness from facial proportions," *Pattern Recognition*, vol. 45, no. 6, pp. 2326–2334, 2012.

[57] P. W. Hashim, J. K. Nia, M. Taliercio, and G. Goldenberg, "Ideals of facial beauty.," *Cutis*, vol. 100, no. 4, pp. 222–224, 2017.

[58] K. Schmid, D. Marx, and A. Samal, "Computation of a face attractiveness index based on neoclassical canons, symmetry, and golden ratios," *Pattern Recognition*, vol. 41, no. 8, pp. 2710–2717, 2008.

[59] Y. Eisenthal, G. Dror, and E. Ruppin, "Facial attractiveness: Beauty and the machine," *Neural Computation*, vol. 18, no. 1, pp. 119–142, 2006.

[60] J. Whitehill and J. R. Movellan, "Personalized facial attractiveness prediction," in *2008 8th IEEE International Conference on Automatic Face & Gesture Recognition*, pp. 1–7, IEEE, 2008.

[61] A. Kagian, G. Dror, T. Leyvand, D. Cohen-Or, and E. Ruppin, "A humanlike predictor of facial attractiveness," in *Advances in Neural Information Processing Systems*, pp. 649–656, 2007.

[62] H. Gunes and M. Piccardi, "Assessing facial beauty through proportion analysis by image processing and supervised learning," *International Journal of Human-computer Studies*, vol. 64, no. 12, pp. 1184–1199, 2006.

[63] D. Zhang, Q. Zhao, and F. Chen, "Quantitative analysis of human facial beauty using geometric features," *Pattern Recognition*, vol. 44, no. 4, pp. 940–950, 2011.

[64] R. Duda, P. Hart, and D. Stork, "Pattern classification, ed 2. hoboken," 2000.

[65] V. Vapnik, *The nature of statistical learning theory*. Springer Science & Business Media, 2013.

[66] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol. 1, pp. I–I, IEEE, 2001.

[67] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models-their training and application," *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38–59, 1995.

[68] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[69] D. Gray, K. Yu, W. Xu, and Y. Gong, "Predicting facial beauty without landmarks," in *European Conference on Computer Vision*, pp. 434–447, Springer, 2010.

[70] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, 1996.

[71] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition," *IEEE Transactions on Image Processing*, vol. 11, no. 4, pp. 467–476, 2002.

[72] H. Yan, "Cost-sensitive ordinal regression for fully automatic facial beauty assessment," *Neurocomputing*, vol. 129, pp. 334–342, 2014.

[73] J. Gan, L. Li, Y. Zhai, and Y. Liu, "Deep self-taught learning for facial beauty prediction," *Neurocomputing*, vol. 144, pp. 295–303, 2014.

[74] W. Wei, E. S. Ho, K. D. McCay, R. Damaševičius, R. Maskeliūnas, and A. Esposito, "Assessing facial symmetry and attractiveness using augmented reality," *Pattern Analysis and Applications*, pp. 1–17, 2021.

[75] B. C. Patel, W. Abdul, and Z. P. Joos, "Anatomy, head and neck, eyelash," *StatPearls [Internet]*, 2019.

[76] J. Khong, R. Casson, S. Huilgol, and D. Selva, "Madarosis," *Survey of Ophthalmology*, vol. 51, no. 6, pp. 550–560, 2006.

[77] A. Kumar and K. Karthikeyan, "Madarosis: a marker of many maladies," *International Journal of Trichology*, vol. 4, no. 1, p. 3, 2012.

[78] F. Pazhoohi and A. Kingstone, "The effect of eyelash length on attractiveness: A previously uninvestigated indicator of beauty.," *Evolutionary Behavioural Sciences*, 2020.

[79] A. Adam, "Beauty is in the eye of the beautiful: Enhanced eyelashes increase perceived health and attractiveness.," *Evolutionary Behavioural Sciences*, 2020.

[80] A. Fetters, "Angelina jolie makeup tutorial." `https://www.thecut.com/2018/04/the-psychology-behind-why-we-like-long-dark-eyelashes.html`. Accessed: 2021-06-14.

[81] "Eyelash extensions." http://www.camillalashes.com/eyelash-extensions/. Accessed: 2021-06-14.

[82] G. Rhodes, F. Proffitt, J. M. Grady, and A. Sumich, "Facial symmetry and the perception of beauty," *Psychonomic Bulletin & Review*, vol. 5, no. 4, pp. 659–669, 1998.

[83] C. Sullivan, "Eyelash extensions 101: Everything you need to know about lash extensions." https://www.today.com/style/eyelash-extensions-101-everything-you-need-know-about-lash-exte Accessed: 2021-06-14.

[84] T. Kazimov and S. Mahmudova, "About a method of recognition of race and ethnicity of individuals based on portrait photographs," *Intelligent Control and Automation*, vol. 2014, 2014.

[85] P. Kingsley, *Hair: An Owner's Handbook*. White Lion Publishing, 1995.

[86] M. R. Cunningham, P. B. Druen, and A. P. Barbee, "Angels, mentors, and friends: Trade-offs among evolutionary, social, and individual variables in physical appearance," *Evolutionary Social Psychology*, vol. 25, pp. 109–140, 1997.

[87] N. Mesko and T. Bereczkei, "Hairstyle as an adaptive means of displaying phenotypic quality," *Human Nature*, vol. 15, no. 3, pp. 251–270, 2004.

[88] V. B. Hinsz, D. C. Matz, and R. A. Patience, "Does women's hair signal reproductive potential?," *Journal of Experimental Social Psychology*, vol. 37, no. 2, pp. 166–172, 2001.

[89] K. Grammer, B. Fink, A. Juette, G. Ronzal, and R. Thornhill, "Female faces and bodies: N-dimensional feature space and attractiveness.," *Advances in Visual Cognition*, 2002.

[90] F. Muscarella and M. R. Cunningham, "The evolutionary significance and social perception of male pattern baldness and facial hair," *Ethology and Sociobiology*, vol. 17, no. 2, pp. 99–117, 1996.

[91] T. Brewer, "The most flattering haircuts for your face shape." `https://www.thetrendspotter.net/haircuts-for-face-shape/`. Accessed: 2020-03-01.

[92] J. Mannah, "The best men's hairstyles for your face shape." `https://www.thetrendspotter.net/find-the-perfect-hairstyle-haircut-for-your-face-shape/`. Accessed: 2020-03-01.

[93] M. Angeloglou, *A History of Make-up*. Macmillan, 1970.

[94] M.-L. Eckert, N. Kose, and J.-L. Dugelay, "Facial cosmetics database and impact analysis on automatic face recognition," in *2013 IEEE 15th International Workshop on Multimedia Signal Processing (MMSP)*, pp. 434–439, IEEE, 2013.

[95] "Angelina jolie makeup tutorial." `https://www.youtube.com/watch?v=rNQ2oDEyVNw`. Accessed: 2021-06-14.

[96] R. Korichi and D. Pelle-de Queral, "Why women use makeup: Implication of psychological," *Journal of Cosmetic Science*, vol. 59, pp. 127–137, 2008.

[97] J. Robertson, G. Fieldman, and T. Hussey, "" who wears cosmetics?" individual differences and their relationship with cosmetic usage.," *Individual Differences Research*, vol. 6, no. 1, 2008.

[98] R. Nash, G. Fieldman, T. Hussey, J.-L. Lévêque, and P. Pineau, "Cosmetics: They influence more than caucasian female facial attractiveness," *Journal of Applied Social Psychology*, vol. 36, no. 2, pp. 493–504, 2006.

[99] N. L. Etcoff, S. Stock, L. E. Haley, S. A. Vickery, and D. M. House, "Cosmetics as a feature of the extended human phenotype: Modulation of the perception of biologically important facial signals," *PloS One*, vol. 6, no. 10, p. e25656, 2011.

[100] V. R. Mileva, A. L. Jones, R. Russell, and A. C. Little, "Sex differences in the perceived dominance and prestige of women with and without cosmetics," *Perception*, vol. 45, no. 10, pp. 1166–1183, 2016.

[101] N. Guéguen and C. Jacob, "Enhanced female attractiveness with use of cosmetics and male tipping behaviour in restaurants.," *Journal of Cosmetic Science*, vol. 62, no. 3, pp. 283–290, 2011.

[102] A. L. Jones, R. Russell, and R. Ward, "Cosmetics alter biologically-based factors of beauty: evidence from facial contrast," *Evolutionary Psychology*, vol. 13, no. 1, p. 147470491501300113, 2015.

[103] B. Fink, K. Grammer, and P. J. Matts, "Visible skin color distribution plays a role in the perception of age, attractiveness, and health in female faces," *Evolution and Human Behaviour*, vol. 27, no. 6, pp. 433–442, 2006.

[104] A. L. Jones and R. S. Kramer, "Facial cosmetics and attractiveness: Comparing the effect sizes of professionally-applied cosmetics and identity," *PloS One*, vol. 11, no. 10, p. e0164218, 2016.

[105] W. Pei, *Models for supervised learning in sequence data*. PhD thesis, Delft University of Technology, 2018.

[106] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.

[107] S. Loukas, "What is machine learning: Supervised, unsupervised, semi-supervised and reinforcement learning methods." `https://towardsdatascience.com/`

what-is-machine-learning-a-short-note-on-supervised-unsupervise
Accessed: 2022-1-13.

[108] J. C. Ang, A. Mirzal, H. Haron, and H. N. A. Hamed, "Supervised, unsupervised, and semi-supervised feature selection: a review on gene selection," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 13, no. 5, pp. 971–989, 2015.

[109] G. Bonaccorso, *Machine learning algorithms*. Packt Publishing Ltd, 2017.

[110] R. Reed and R. J. MarksII, *Neural smithing: supervised learning in feedforward artificial neural networks*. Mit Press, 1999.

[111] T. Mikolov, M. Karafiát, L. Burget, J. Cernockỳ, and S. Khudanpur, "Recurrent neural network based language model.," in *Interspeech*, vol. 2, pp. 1045–1048, Makuhari, 2010.

[112] Y. Bengio *et al.*, "Learning deep architectures for AI," *Foundations and Trends® in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009.

[113] L. Deng, D. Yu, *et al.*, "Deep learning: methods and applications," *Foundations and Trends® in Signal Processing*, vol. 7, no. 3–4, pp. 197–387, 2014.

[114] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

[115] Y. LeCun, Y. Bengio, *et al.*, "Convolutional networks for images, speech, and time series," *The handbook of Brain Theory and Neural Networks*, vol. 3361, no. 10, p. 1995, 1995.

[116] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*. No. 2, MIT Press Cambridge, 2016.

[117] R. Salakhutdinov and G. Hinton, "Deep boltzmann machines," in *Artificial Intelligence and Statistics*, pp. 448–455, PMLR, 2009.

[118] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[119] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.

[120] H. Sak, A. Senior, and F. Beaufays, "Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition," *arXiv preprint arXiv:1402.1128*, 2014.

[121] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, 2014.

[122] S. Mittal, "A survey on optimized implementation of deep learning models on the nvidia jetson platform," *Journal of Systems Architecture*, vol. 97, pp. 428–442, 2019.

[123] A. Garifullin, "deep learning for retinal image segementation," Master's thesis, Lappeenranta University of Technology, 2017.

[124] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.

[125] I. Guyon, S. Gunn, M. Nikravesh, and L. A. Zadeh, *Feature extraction: foundations and applications*, vol. 207. Springer, 2008.

[126] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intelligent Systems and Their Applications*, vol. 13, no. 4, pp. 18–28, 1998.

[127] E. Fix and J. L. Hodges, "Discriminatory analysis. nonparametric discrimination: Consistency properties," *International Statistical Review/Revue Internationale de Statistique*, vol. 57, no. 3, pp. 238–247, 1989.

[128] B. Al-Bander, *Retinal Image Analysis Based on Deep Learning*. PhD thesis, Dept. of Electrical Engineering and Electronics, The University of Liverpool (United Kingdom), 2018.

[129] A. Borodinets, "Deep learning vs. machine learning: What is the difference." `https://www.qulix.com/about/deep-learning-vs-machine-learning/`, 2020. Accessed: 2022-01-14.

[130] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–9, 2015.

[131] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, no. 1, 2017.

[132] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4700–4708, 2017.

[133] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1251–1258, 2017.

[134] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[135] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8697–8710, 2018.

[136] X. Zhang, X. Zhou, M. Lin, and J. Sun, "Shufflenet: An extremely efficient convolutional neural network for mobile devices," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6848–6856, 2018.

[137] J. Redmon and A. Farhadi, "Yolo9000: better, faster, stronger," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7263–7271, 2017.

[138] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.

[139] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International Conference on Machine Learning*, pp. 6105–6114, PMLR, 2019.

[140] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "Squeezenet: Alexnet-level accuracy with 50x fewer parameters and¡ 0.5 mb model size," *arXiv preprint arXiv:1602.07360*, 2016.

[141] "Pretrained deep neural networks." `https://uk.mathworks.com/help/deeplearning/ug/pretrained-convolutional-neural-networks.html#mw_6ab30f15-4136-46b0-b5db-4f9b3b1cfcd2`. Accessed: 2021-06-21.

[142] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[143] P. Velickovic, "Deep learning for complete beginners: convolutional neural networks with keras." `https://cambridgespark.com/content/tutorials/convolutional-neural-networks-with-keras/index.html`. Accessed: 2022-01-13.

[144] B. Al-Bander, W. Al-Nuaimy, B. M. Williams, and Y. Zheng, "Multiscale sequential convolutional neural networks for simultaneous detection of fovea and optic disc," *Biomedical Signal Processing and Control*, vol. 40, pp. 91–101, 2018.

[145] F. J. Huang, Y.-L. Boureau, Y. LeCun, *et al.*, "Unsupervised learning of invariant feature hierarchies with applications to object recognition," in *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pp. 1–8, IEEE, 2007.

[146] Stanford, "Cs231n: Convolutional neural networks for visual recognition." `https://cs231n.github.io/convolutional-networks/`. Accessed: 2022-05-02.

[147] J. Brownlee, "Multi-class neural networks: Softmax." `https://developers.google.com/machine-learning/crash-course/multi-class-neural-networks/softmax`. Accessed: 2022-05-01.

[148] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pp. 315–323, 2011.

[149] J. Brownlee, "How to choose an activation function for deep learning." `https://machinelearningmastery.com/choose-an-activation-function-for-deep-learning/`. Accessed: 2022-05-01.

[150] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International Conference on Machine Learning*, pp. 448–456, 2015.

[151] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting.," *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[152] J. Wang and L. Perez, "The effectiveness of data augmentation in image classification using deep learning," 2017.

[153] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pp. 249–256, 2010.

[154] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1026–1034, 2015.

[155] G. Gwardys, "Convolutional neural networks backpropagation: from intuition to derivation." `https://grzegorzgwardys.wordpress.com/2016/04/22/8/`, 2016. Accessed: 2022-01-11.

[156] D. E. Rumelhart, G. E. Hinton, R. J. Williams, *et al.*, "Learning representations by back-propagating errors," *Cognitive Modeling*, vol. 5, no. 3, p. 1, 1988.

[157] M. D. Zeiler, "Adadelta: an adaptive learning rate method," *arXiv preprint arXiv:1212.5701*, 2012.

[158] J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization," *Journal of Machine Learning Research*, vol. 12, no. Jul, pp. 2121–2159, 2011.

[159] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[160] G. Hinton, N. Srivastava, and K. Swersky, "Lecture 6a overview of mini–batch gradient descent," *Coursera Lecture slides https://class. coursera. org/neuralnets-2012-001/lecture, Online*, 2012.

[161] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?," in *Advances in Neural Information Processing Systems*, pp. 3320–3328, 2014.

[162] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 580–587, 2014.

[163] Z. Zhang, P. Luo, C. C. Loy, and X. Tang, "Facial landmark detection by deep multi-task learning," in *European Conference on Computer Vision*, pp. 94–108, Springer, 2014.

[164] Y. Wu, T. Hassner, K. Kim, G. Medioni, and P. Natarajan, "Facial landmark detection with tweaked convolutional neural networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 12, pp. 3067–3074, 2017.

[165] S. Ghosh, N. Das, I. Das, and U. Maulik, "Understanding deep learning techniques for image segmentation," *ACM Computing Surveys (CSUR)*, vol. 52, no. 4, pp. 1–35, 2019.

[166] G. Levi and T. Hassner, "Age and gender classification using convolutional neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 34–42, 2015.

[167] Q. Dou, H. Chen, L. Yu, L. Zhao, J. Qin, D. Wang, V. C. Mok, L. Shi, and P.-A. Heng, "Automatic detection of cerebral microbleeds from MR images via 3d convolutional neural networks," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1182–1195, 2016.

[168] K. Khan, R. U. Khan, K. Ahmad, F. Ali, and K.-S. Kwak, "Face segmentation: A journey from classical to deep learning paradigm, approaches, trends, and directions," *IEEE Access*, vol. 8, pp. 58683–58699, 2020.

[169] J. O. Kephart and D. M. Chess, "The vision of autonomic computing," *Computer*, vol. 36, no. 1, pp. 41–50, 2003.

[170] S. Duffner, *Face image analysis with convolutional neural networks*. PhD thesis, University of Freiburg, 2008.

[171] S. Aslan, *Multimodal Video-Based Personality Recognition Using Long Short-Term Memory and Convolutional Neural Networks*. PhD thesis, Bilkent Universitesi (Turkey), 2019.

[172] "How artificial intelligence (AI) is transforming beauty and cosmetics retail." https://vue.ai/blog/ai-in-retail/artificial-intelligence-for-beauty-and-cosmetics-retail-industr Accessed: 2021-06-19.

[173] "Transforming the beauty industry with data and analytics." https://clarkstonconsulting.com/insights/beauty-industry-transformation/. Accessed: 2022-03-5.

[174] "Ai is driving efficiencies in the beauty industry — and how brands can benefit." https://www.cbinsights.com/research/artificial-intelligence-applications-in-beauty/. Accessed: 2022-03-5.

[175] M. Priya, "Artificial intelligence and the new age beauty and cosmetics industry." https://www.prescouter.com/2019/06/artificial-intelligence-and-the-new-age-beauty-and-cosmetics-in Accessed: 2020-04-01.

[176] M. T. Chowdhury, A. Sarkar, P. K. Saha, and R. H. Anik, "Enhancing supply resilience in the covid-19 pandemic: a case study on beauty and personal care retailers," *Modern Supply Chain Research and Applications*, 2020.

[177] D. Forsyth and J. Ponce, *Computer vision: A modern approach.* Prentice hall, 2011.

[178] E. L. Spratt and A. Elgammal, "Computational beauty: Aesthetic judgment at the intersection of art and science," in *European Conference on Computer Vision*, pp. 35–53, Springer, 2014.

[179] N. A. Andrade, G. C. Rainatto, D. G. E. Paschoal, F. R. da Silva, and G. Renovato, "Computational vision and business intelligence in the beauty segment-an analysis through instagram," *Journal of Marketing Management*, vol. 7, no. 2, pp. 11–17, 2019.

[180] S.-W. Kim and S.-W. Lee, "Deep learning approach for cosmetic product detection and classification," 2021.

[181] J. Y. Liu, "A survey of deep learning approaches for recommendation systems," in *Journal of Physics: Conference Series*, vol. 1087, p. 062022, IOP Publishing, 2018.

[182] L. Zhou, "Product advertising recommendation in e-commerce based on deep learning and distributed expression," *Electronic Commerce Research*, vol. 20, no. 2, pp. 321–342, 2020.

[183] S. P. R. Karri and B. S. Kumar, "Deep learning techniques for implementation of chatbots," in *2020 International Conference on Computer Communication and Informatics (ICCCI)*, pp. 1–5, IEEE, 2020.

[184] S. Ramaswamy and N. DeClerck, "Customer perception analysis using deep learning and nlp," *Procedia Computer Science*, vol. 140, pp. 170–178, 2018.

[185] A. de Jesus, "Artificial intelligence for beauty and cosmetics – current applications." https://emerj.com/ai-sector-overviews/artificial-intelligence-for-beauty-and-cosmetics-current-applic Accessed: 2021-06-19.

[186] V. Margherita, "How is deep learning transforming the beauty industry?." https://blog.hslu.ch/majorobm/2020/05/19/deep-learning-transforming-beauty-industry-mva/. Accessed: 2021-06-19.

[187] E. Siegel, *Predictive analytics: The power to predict who will click, buy, lie, or die*, vol. 10. Wiley Hoboken, 2013.

[188] L. Liu, J. Xing, S. Liu, H. Xu, X. Zhou, and S. Yan, "Wow! you are so beautiful today!," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 11, no. 1s, pp. 1–22, 2014.

[189] S. Liu, X. Ou, R. Qian, W. Wang, and X. Cao, "Makeup like a superstar: Deep localized makeup transfer network," *arXiv preprint arXiv:1604.07102*, 2016.

[190] X. Wang, K. Wang, and S. Lian, "A survey on face data augmentation for the training of deep neural networks," *Neural Computing and Applications*, vol. 32, pp. 15503–15531, 2020.

[191] F. Kane, *Building Recommender Systems with Machine Learning and AI: Help people discover new products and content with deep learning, neural networks, and machine learning recommendations.* Independently Published, 2018.

[192] E. Salsky, "Personalization in beauty tech using ai and ar: investigating consumer behaviour and benefits of personalization in beauty," Master's thesis, Tampere University of Applied Sciences, 2020.

[193] A. Adebo, *A Natural Language Processing Approach to a Skincare Recommendation Engine*. PhD thesis, Dublin, National College of Ireland, 2020.

[194] G. Yildirim, N. Jetchev, R. Vollgraf, and U. Bergmann, "Generating high-resolution fashion model images wearing custom outfits," in *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, pp. 700–712, 2019.

[195] C. Shahani-Denning, "Physical attractiveness bias in hiring: What is beautiful is good," *Hofstra Horizon*, pp. 14–17, 2003.

[196] S. Kertechian, "The impact of beauty during job applications," *Journal of Human Resources Management Research*, vol. 2016, pp. a1–7, 2016.

[197] A. K. Jain and S. Z. Li, *Handbook of face recognition*, vol. 1. Springer, 2011.

[198] M. S. Nixon, *Feature Extraction & Image Processing for Computer Vision*. Academic Press, 2012.

[199] C. Garcia, J. Ostermann, and T. Cootes, "Facial image processing," *EURASIP Journal on Image and Video Processing*, vol. 207, pp. 1–33, 2008.

[200] M. Zivkovic, N. Bacanin, K. Venkatachalam, A. Nayyar, A. Djordjevic, I. Strumberger, and F. Al-Turjman, "Covid-19 cases prediction by using hybrid machine learning and beetle antennae search approach," *Sustainable Cities and Society*, vol. 66, p. 102669, 2021.

[201] S. Vandenhende, S. Georgoulis, M. Proesmans, D. Dai, and L. Van Gool, "Revisiting multi-task learning in the deep learning era," *arXiv preprint arXiv:2004.13379*, 2020.

[202] G. H. De Rosa, J. P. Papa, and X.-S. Yang, "Handling dropout probability estimation in convolution neural networks using meta-heuristics," *Soft Computing*, vol. 22, no. 18, pp. 6147–6156, 2018.

[203] Y. Wu and Q. Ji, "Facial landmark detection: A literature survey," *International Journal of Computer Vision*, vol. 127, no. 2, pp. 115–142, 2019.

[204] M. Koestinger, P. Wohlhart, P. M. Roth, and H. Bischof, "Annotated facial landmarks in the wild: A large-scale, real-world database for facial landmark localization," in *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, pp. 2144–2151, IEEE, 2011.

[205] "How ai and ar are revolutionising the beauty industry." `https://www.theindustry.fashion/`

the-role-of-ai-and-ar-in-the-beauty-industry/. Accessed: 2022-03-5.

[206] C. Edwards, "Ai in the beauty industry: The tech making the future beautiful." https://www.verdict.co.uk/ai-in-the-beauty-industry/. Accessed: 2020-04-01.

[207] S. K. Panda and P. K. Jana, "Load balanced task scheduling for cloud computing: A probabilistic approach," *Knowledge and Information Systems*, vol. 61, no. 3, pp. 1607–1631, 2019.

[208] X. Zheng, Y. Guo, H. Huang, Y. Li, and R. He, "A survey of deep facial attribute analysis," *International Journal of Computer Vision*, pp. 1–33, 2020.

[209] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3730–3738, 2015.

[210] E. M. Rudd, M. Günther, and T. E. Boult, "Moon: A mixed objective optimization network for the recognition of facial attributes," in *European Conference on Computer Vision*, pp. 19–35, Springer, 2016.

[211] E. M. Hand and R. Chellappa, "Attributes for improved attributes: A multi-task network utilizing implicit and explicit relationships for facial attribute classification," in *proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, pp. 4068–4074, 2017.

[212] J. Cao, Y. Li, and Z. Zhang, "Partially shared multi-task convolutional neural network with local constraint for face attribute learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4290–4299, 2018.

[213] M. M. Kalayeh, B. Gong, and M. Shah, "Improving facial attribute prediction using semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6942–6950, 2017.

[214] U. Mahbub, S. Sarkar, and R. Chellappa, "Segment-based methods for facial attribute detection from partial faces," *IEEE Transactions on Affective Computing*, 2018.

[215] W. Sunhem and K. Pasupa, "An approach to face shape classification for hairstyle recommendation," in *2016 Eighth International Conference on Advanced Computational Intelligence (ICACI)*, pp. 390–394, IEEE, 2016.

[216] P. Sarakon, T. Charoenpong, and S. Charoensiriwath, "Face shape classification from 3D human data by using SVM," in *The 7th 2014 Biomedical Engineering International Conference*, pp. 1–5, IEEE, 2014.

[217] A. Zafar and T. Popa, "Face and eye-ware classification using geometric features for a data-driven eye-ware recommendation system," in *Graphics Interface Conference*, pp. 183–188, ACM Library, 2016.

[218] R. F. Rahmat, M. D. Syahputra, U. Andayani, and T. Z. Lini, "Probabilistic neural network and invariant moments for men face shape classification," in *IOP Conference Series: Materials Science and Engineering*, pp. 12–19, IOP Publishing, 2018.

[219] N. Bansode and P. Sinha, "Face shape classification based on region similarity, correlation and fractal dimensions," *International Journal of Computer Science Issues (IJCSI)*, vol. 13, no. 1, p. 24, 2016.

[220] E. D. Tio., "Face shape classification using inception v3." `https://github.com/adonistio/inception-face-shape-classifier`, 2017. Accessed: 2019-06-01.

[221] W.-C. Kang, C. Fang, Z. Wang, and J. McAuley, "Visually-aware fashion recommendation and design with generative image models," in *2017 IEEE International Conference on Data Mining (ICDM)*, pp. 207–216, IEEE, 2017.

[222] T. Lodkaew, W. Supsohmboon, K. Pasupa, and C. K. Loo, "Fashion finder: A system for locating online stores on instagram from product images," in *2018 10th International Conference on Information Technology and Electrical Engineering (ICITEE)*, pp. 500–505, IEEE, 2018.

[223] S. Milborrow, J. Morkel, and F. Nicolls, "The muct landmarked face database," *Pattern Recognition Association of South Africa*, vol. 201, pp. 1–6, 2010. http://www.milbo.org/muct.

[224] Presleehairstyle, "Wedding veils: Denver wedding hairstylist." https://www.presleehairstyle.com/blog/2017/5/31/wedding-veils-denver-wedding-hairstylist-braiding-specialist-an Accessed: 2020-02-01.

[225] Aumedo, "Make up eyes: make-up tips for every eye shape." https://www.aumedo.de/make-up-tipps-fuer-jede-augenform/. Accessed: 2020-02-01.

[226] A. Moreno, "How to pick the best lashes for your eye shape." https://www.iconalashes.com/blogs/news/how-to-pick-the-best-lashes-for-your-eye-shape. Accessed: 2020-05-01.

[227] J. Derrick, "Is your face round, square, long, heart or oval shaped?." https://www.byrdie.com/is-your-face-round-square-long-heart-or-oval-shaped-345761. Accessed: 2019-06-01.

[228] B. Shunatona, "A guide to determining your face shape (if that's something you care about)." https://www.cosmopolitan.com/style-beauty/beauty/a22064408/face-shape-guide/. Accessed: 2020-02-01.

[229] WikiHow, "How to determine eye shape." https://www.wikihow.com/Determine-Eye-Shape#aiinfo. Accessed: 2020-02-01.

[230] T. Alzahrani, W. Al-Nuaimy, and B. Al-Bander, "Hybrid feature learning and engineering based approach for face shape classification," in *2019 International Conference on Intelligent Systems and Advanced Computing Sciences (ISACS)*, pp. 1–4, IEEE, 2019.

[231] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *International Conference on Computer Vision & Pattern Recognition (CVPR'05)*, vol. 1, pp. 886–893, IEEE Computer Society, 2005.

[232] V. Kazemi and J. Sullivan, "One millisecond face alignment with an ensemble of regression trees," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1867–1874, 2014.

[233] P. I. Wilson and J. Fernandez, "Facial feature detection using haar classifiers," *Journal of Computing Sciences in Colleges*, vol. 21, no. 4, pp. 127–133, 2006.

[234] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," in *Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition*, 2008.

[235] H. Bristow and S. Lucey, "Why do linear svms trained on hog features perform so well?," *arXiv preprint arXiv:1406.2419*, 2014.

[236] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, "300 faces in-the-wild challenge: The first facial landmark localization challenge," in *Proceedings*

of the IEEE International Conference on Computer Vision Workshops, pp. 397–403, 2013.

[237] A. Rosebrock, "Training a custom dlib shape predictor." https://pyimagesearch.com/2019/12/16/training-a-custom-dlib-shape-predictor/, 2019. Accessed: 2022-04-011.

[238] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, *et al.*, "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.

[239] C. Creasey, "Eye shapes and eyelash extensions–what works best for you?." https://www.sydneyeyelashextensions.com/eye-shapes-and-eyelash-extensions-what-works-best-for-you/. Accessed: 2020-03-01.

[240] M. Duan, K. Li, C. Yang, and K. Li, "A hybrid deep learning cnn–elm for age and gender classification," *Neurocomputing*, vol. 275, pp. 448–461, 2018.

[241] M. Everingham, S. A. Eslami, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes challenge: A retrospective," *International Journal of Computer Vision*, vol. 111, no. 1, pp. 98–136, 2015.

[242] L. F. de Araujo Zeni and C. R. Jung, "Real-time gender detection in the wild using deep neural networks," in *2018 31st Conference on Graphics, Patterns and Images (SIBGRAPI)*, pp. 118–125, IEEE, 2018.

[243] D. Borza, A. S. Darabant, and R. Danescu, "Real-time detection and measurement of eye features from color images," *Sensors*, vol. 16, no. 7, p. 1105, 2016.

[244] M. A. J. ALDIBAJA, *Eye shape detection methods based on eye structure modeling and texture analysis for interface systems*. PhD dissertation, Toyohashi University of Technology, 2015.

[245] A. Dantcheva, C. Chen, and A. Ross, "Can facial cosmetics affect the matching accuracy of face recognition systems?," in *2012 IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pp. 391–398, IEEE, 2012.

[246] N. Guéguen, "The effects of women's cosmetics on men's courtship behaviour," *North American Journal of Psychology*, vol. 10, no. 1, pp. 221–228, 2008.

[247] A. Dantcheva and J. Dugelay, "Female facial aesthetics based on soft biometrics and photo-quality," in *Proceedings of ICME*, vol. 2, 2011.

[248] X. Liu, T. Li, H. Peng, I. Chuoying Ouyang, T. Kim, and R. Wang, "Understanding beauty via deep facial features," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 246–256, 2019.

[249] G. Guo, L. Wen, and S. Yan, "Face authentication with makeup changes," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 24, no. 5, pp. 814–825, 2013.

[250] S. Wang and Y. Fu, "Face behind makeup," in *Proceedings of the AAAI Conference on Artificial Intelligence*, no. 1, 2016.

[251] C. Rathgeb, C.-I. Satnoianu, N. Haryanto, K. Bernardo, and C. Busch, "Differential detection of facial retouching: A multi-biometric approach," *IEEE Access*, vol. 8, pp. 106373–106385, 2020.

[252] M. Sajid, N. Ali, S. H. Dar, N. Iqbal Ratyal, A. R. Butt, B. Zafar, T. Shafique, M. J. A. Baig, I. Riaz, and S. Baig, "Data augmentation-assisted makeup-invariant face recognition," *Mathematical Problems in Engineering*, vol. 2018, 2018.

[253] X. Dong, Y. Yan, W. Ouyang, and Y. Yang, "Style aggregated network for facial landmark detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 379–388, 2018.

[254] Y. Li, L. Song, X. Wu, R. He, and T. Tan, "Anti-makeup: Learning a bi-level adversarial network for makeup-invariant face verification," in *Proceedings of the AAAI Conference on Artificial Intelligence*, no. 1, 2018.

[255] Y. Li, L. Song, X. Wu, R. He, and T. Tan, "Learning a bi-level adversarial network with global and local perception for makeup-invariant face verification," *Pattern Recognition*, vol. 90, pp. 99–108, 2019.

[256] T. Alzahrani and W. Al-Nuaimy, "Face segmentation based object localisation with deep learning from unconstrained images," in *In 10th International Conference on Pattern Recognition Systems (ICPRS)*, pp. 47–51, IET, 2019.

[257] T. Alzahrani, W. Al-Nuaimy, and B. Al-Bander, "Integrated multi-model face shape and eye attributes identification for hair style and eyelashes recommendation," *Computation*, vol. 9, no. 5, 2021.

[258] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proceedings of the 25th International Conference on Machine Learning*, pp. 1096–1103, 2008.

[259] J. Masci, U. Meier, D. Cireşan, and J. Schmidhuber, "Stacked convolutional auto-encoders for hierarchical feature extraction," in *International Conference on Artificial Neural Networks*, pp. 52–59, Springer, 2011.

[260] D. P. Kingma, S. Mohamed, D. Jimenez Rezende, and M. Welling, "Semi-supervised learning with deep generative models," *Advances in Neural Information Processing Systems*, vol. 27, pp. 3581–3589, 2014.

[261] A. Odena, "Semi-supervised learning with generative adversarial networks," *arXiv preprint arXiv:1606.01583*, 2016.

[262] D.-H. Lee, "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks," in *Workshop on Challenges in Representation Learning, ICML*, no. 2, 2013.

[263] A. Juhong and C. Pintavirooj, "Face recognition based on facial landmark detection," in *2017 10th Biomedical Engineering International Conference (BMEiCON)*, pp. 1–4, IEEE, 2017.

[264] E. R. Aguinaldo and J. J. Peissig, "Who's behind the makeup? the effects of varying levels of cosmetics application on perceptions of facial attractiveness, competence, and sociosexuality," *Frontiers in Psychology*, vol. 12, 2021.

[265] S. Varshovi, *Facial makeup detection using HSV color space and texture analysis*. PhD thesis, Concordia University, 2012.

[266] C. Chen, A. Dantcheva, and A. Ross, "Automatic facial makeup detection with application in face recognition," in *2013 International Conference on Biometrics (ICB)*, pp. 1–8, IEEE, 2013.

[267] J. Hu, Y. Ge, J. Lu, and X. Feng, "Makeup-robust face verification," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 2342–2346, IEEE, 2013.

[268] Y. Choukroun and M. Katz, "Meta subspace optimization," *arXiv preprint arXiv:2110.14920*, 2021.

[269] A. Dantcheva and J.-L. Dugelay, "Assessment of female facial beauty based on anthropometric, non-permanent and acquisition characteristics," *Multimedia Tools and Applications*, vol. 74, no. 24, pp. 11331–11355, 2015.

[270] A. Bharati, R. Singh, M. Vatsa, and K. W. Bowyer, "Detecting facial retouching using supervised deep learning," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 9, pp. 1903–1913, 2016.

[271] I. A. Kamil and A. S. Are, "Makeup-invariant face identification and verification using fisher linear discriminant analysis-based gabor filter bank and histogram of oriented gradients," *International Journal of Signal and Imaging Systems Engineering*, vol. 10, no. 5, pp. 257–270, 2017.

[272] Y. Fu and W. Shuyang, "System for beauty, cosmetic, and fashion analysis." `https://patents.google.com/patent/US20170076474A1/en`, July 2 2019. US Patent 10,339,685.

[273] M. Li, Y. Li, and Y. He, "Makeup removal system with deep learning," tech. rep., Technical Report, CS230 Deep Learning, Stanford University, 2018.

[274] C. Rathgeb, P. Drozdowski, D. Fischer, and C. Busch, "Vulnerability assessment and detection of makeup presentation attacks," in *2020 8th International Workshop on Biometrics and Forensics (IWBF)*, pp. 1–6, IEEE, 2020.

[275] C. Rathgeb, A. Botaljov, F. Stockhardt, S. Isadskiy, L. Debiasi, A. Uhl, and C. Busch, "PRNU-based detection of facial retouching," *IET Biometrics*, vol. 9, no. 4, pp. 154–164, 2020.

[276] C. Rathgeb, P. Drozdowski, and C. Busch, "Makeup presentation attacks: Review and detection performance benchmark," *IEEE Access*, 2020.

[277] Y. Fu, G. Guo, and T. S. Huang, "Age synthesis and estimation via faces: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 11, pp. 1955–1976, 2010.

[278] K. Kotwal, Z. Mostaani, and S. Marcel, "Detection of age-induced makeup attacks on face recognition systems using multi-layer deep features," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 2, no. 1, pp. 15–25, 2019.

[279] Kaggle, "Makeup or no makeup." https://www.kaggle.com/petersunga/make-up-vs-no-make-up, 2018. Accessed: 2020-11-01.

[280] C. Chen, A. Dantcheva, and A. Ross, "An ensemble of patch-based subspaces for makeup-robust face recognition," *Information Fusion*, vol. 32, pp. 80–92, 2016.

[281] C. Chen, A. Dantcheva, T. Swearingen, and A. Ross, "Spoofing faces using makeup: An investigative study," in *2017 IEEE International Conference on Identity, Security and Behavior Analysis (ISBA)*, pp. 1–8, IEEE, 2017.

[282] "Makeup datasets." http://www.antitza.com/makeup-datasets.html. Accessed on: 2022-06-01.

[283] NIST, "Face recognition grand challenge (FRGC)." http://www.nist.gov/itl/iad/ig/frgc.cfm, 2018. Accessed: 2020-11-01.

[284] Taaz, "Taaz makeover technology." http://www.taaz.com/. Accessed: 2020-11-01.

[285] L. Torrey and J. Shavlik, "Transfer learning," in *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques*, pp. 242–264, IGI Global, 2010.

[286] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255, IEEE, 2009.

[287] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, "Greedy layer-wise training of deep networks," in *Advances in Neural Information Processing Systems*, pp. 153–160, 2007.

[288] S. Deb, "How to perform data compression using autoencoders?." https://medium.com/edureka/autoencoders-tutorial-cfdcebdefe37, 2018. Accessed: 2022-04-11.

[289] V. Dey, "Understanding the auc-roc curve in machine learning classification." https://analyticsindiamag.com/understanding-the-auc-roc-curve-in-machine-learning-classificat 2021. Accessed: 2022-04-11.

[290] M. Cao, Y. Li, Z. Pan, J. Csete, S. Sun, J. Li, and Y. Liu, "Educational virtual-wear trial: More than a virtual try-on experience," *IEEE Computer Graphics and Applications*, vol. 35, no. 6, pp. 83–89, 2015.

[291] Q. Zhang, Y. Guo, P.-Y. Laffont, T. Martin, and M. Gross, "A virtual try-on system for prescription eyeglasses," *IEEE Computer Graphics and Applications*, vol. 37, no. 4, pp. 84–93, 2017.

[292] G. Chen, Y. Shao, C. Tang, Z. Jin, and J. Zhang, "Deep transformation learning for face recognition in the unconstrained scene," *Machine Vision and Applications*, vol. 29, no. 3, pp. 513–523, 2018.

[293] W. Deng, J. Hu, and J. Guo, "Face recognition via collaborative representation: Its discriminant nature and superposed representation," *IEEE Transactions on Pattern Analysis and Machine intelligence*, vol. 40, no. 10, pp. 2513–2521, 2018.

[294] J. Zhao, Y. Cheng, Y. Xu, L. Xiong, J. Li, F. Zhao, K. Jayashree, S. Pranata, S. Shen, J. Xing, *et al.*, "Towards pose invariant face recognition in the wild," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2207–2216, 2018.

[295] T. Issenhuth, J. Mary, and C. Calauzenes, "Do not mask what you do not need to mask: a parser-free virtual try-on," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XX 16*, pp. 619–635, Springer, 2020.

[296] L. Xu, Y. Du, and Y. Zhang, "An automatic framework for example-based virtual makeup," in *2013 IEEE International Conference on Image Processing*, pp. 3206–3210, IEEE, 2013.

[297] J. Lu, K. Sunkavalli, N. Carr, S. Hadap, and D. Forsyth, "A visual representation for editing face images," *arXiv preprint arXiv:1612.00522*, 2016.

[298] G. Ghiasi, C. C. Fowlkes, and C. Irvine, "Using segmentation to predict the absence of occluded parts.," in *British Machine Vision Conference (BMVC)*, pp. 22–1, 2015.

[299] S. Austnes, "Optimization of object detection in newborn resuscitation videos," Master's thesis, University of Stavanger, Norway, 2019.

[300] X. P. Burgos-Artizzu, P. Perona, and P. Dollár, "Robust face landmark estimation under occlusion," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1513–1520, 2013.

[301] H. Yang, X. He, X. Jia, and I. Patras, "Robust face alignment under occlusion via regional predictive power estimation," *IEEE Transactions on Image Processing*, vol. 24, no. 8, pp. 2393–2403, 2015.

[302] S. Liu, J. Yang, C. Huang, and M.-H. Yang, "Multi-objective convolutional learning for face labeling," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3451–3459, 2015.

[303] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431–3440, 2015.

[304] S. Saito, T. Li, and H. Li, "Real-time facial segmentation and performance capture from RGB input," in *European Conference on Computer Vision*, pp. 244–261, Springer, 2016.

[305] Y. Nirkin, I. Masi, A. T. Tuan, T. Hassner, and G. Medioni, "On face segmentation, face swapping, and face perception," in *Automatic Face & Gesture Recognition (FG 2018), 2018 13th IEEE International Conference on*, pp. 98–105, IEEE, 2018.

[306] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 6, pp. 1137–1149, 2017.

[307] D. Borza, T. Ileni, and A. Darabant, "A deep learning approach to hair segmentation and color extraction from facial images," in *International Conference on Advanced Concepts for Intelligent Vision Systems*, pp. 438–449, Springer, 2018.

[308] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proceedings of International Conference on Computer Vision (ICCV)*, 2015.

[309] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Computer Vision (ICCV), 2017 IEEE International Conference on*, pp. 2980–2988, IEEE, 2017.

[310] I. Rajagopal, "Intuition behind residual neural networks." `https://towardsdatascience.com/intuition-behind-residual-neural-networks-fa5d2996b2c7`, 2020. Accessed: 2022-04-11.

[311] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in Neural Information Processing Systems*, vol. 28, 2015.

[312] D. E. King, "Max-margin object detection," *arXiv preprint arXiv:1502.00046*, 2015.

[313] M. Jaderberg, K. Simonyan, A. Zisserman, *et al.*, "Spatial transformer networks," *Advances in Neural Information Processing Systems*, vol. 28, pp. 2017–2025, 2015.

[314] C. Cao, F. Lu, C. Li, S. Lin, and X. Shen, "Makeup removal via bidirectional tunable de-makeup network," *IEEE Transactions on Multimedia*, vol. 21, no. 11, pp. 2750–2761, 2019.

[315] Q. Gu, G. Wang, M. T. Chiu, Y.-W. Tai, and C.-K. Tang, "Ladn: Local adversarial disentangling network for facial makeup and de-makeup," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 10481–10490, 2019.

[316] H. Chang, J. Lu, F. Yu, and A. Finkelstein, "Pairedcyclegan: Asymmetric style transfer for applying and removing makeup," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pp. 40–48, 2018.

[317] D. Horita and K. Aizawa, "SLGAN: Style-and latent-guided generative adversarial network for desirable makeup transfer and removal," *arXiv preprint arXiv:2009.07557*, 2020.

[318] Y. Li, H. Huang, J. Yu, R. He, and T. Tan, "Semantic-aware makeup cleanser," in *2019 IEEE 10th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pp. 1–7, IEEE, 2019.

[319] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *arXiv preprint arXiv:1406.2661*, 2014.

[320] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, "Improved training of wasserstein gans," *arXiv preprint arXiv:1704.00028*, 2017.