UNSUPERVISED DOMAIN ADAPTATION FOR DISGUISED FACE RECOGNITION

Fangyu Wu[1,2], Shiyang Yan[3], Jeremy S. Smith[2], Wenjin Lu[1], Bailing Zhang[4]

[1] Department of Computer Science and Software Engineering,
Xi'an Jiaotong-liverpool University, SuZhou, JiangSu Province, China,
[2] Department of Electrical Engineering and Electronic, University of Liverpool, Liverpool, L69 3BX, UK,
[3] The Institute of Electronics, Communications and Information Technology,
Queen's University Belfast, NI Science Park, Queen's Road, Queen's Island Belfast, BT3 9DT
[4] School of Computer and Data Engineering,
Ningbo Institute of Technology, Zhejiang University, Ningbo, Zhejiang Province, China
(fangyu.wu, wenjin.lu)@xjtlu.edu.cn, shiyang.yan@qub.ac.uk
J.S.Smith@liverpool.ac.uk,bai_ling_zhang@hotmail.com

ABSTRACT

Disguised face recognition is an extremely challenging task due to the numerous variations that can be introduced with different disguises. Most existing disguised face recognition approaches follow a supervised learning framework. However, due to the domain shift problem, the Convolutional Neural Networks (CNN) model trained on one dataset often fail to generalize well to another dataset. In our attempt, we formulate the DFR as an unsupervised learning problem and propose a unified deep architecture Unsupervised Domain Adaptation Model (UDAM) with three merits. Firstly, UDAM is a unified deep architecture, containing a Domain Style Adaptation subNet (DSN) and an Attention Learning subNet (ALN), which are jointly learned from end-to-end. Secondly, DSN is a well-design generative adversarial network which simultaneously translate the labeled image from source to target domain in an unsupervised manner and maintain the ID label after translation. Thirdly, ALN is a Convolutional Neural Network (CNN) for disguised face recognition with our proposed attention transfer strategy. Extensive experiments over the benchmarks Simple and Complex Face Disguise Dataset and the IIIT-Delhi Disguise Version 1 Face Database have demonstrated that the proposed method yield consistent and competitive performance for disguised face recognition.

***Index Terms*—** Unsupervised Domain Adaptation, Disguised Face Recognition, Generative Adversarial Learning, Attention Transfer

1. INTRODUCTION

Within the past decades, face recognition (FR) has received a tremendous amount of attention owing to its wide range of potential applications, e.g., identity authentication, public security and surveillance. Many innovative and novel methods have been put forward for the tasks of visual face recognition and verification. Meanwhile, great challenges have been confronted by current FR systems, particularly when the accuracy significantly decreases while recognizing the same subjects with disguised appearances, such as wearing a wig or eyeglasses, changing hairstyle and so on [1].

Disguise usually involves intentional and unintentional changes on a face through which one can either impersonate or confuse someone's identity. Fig.1 clearly shows two examples of face obfuscation, in which the appearance of a subject can be varied by using different disguise accessories. To make automatic face recognition secure and usable, it is necessary to address the disguise problem. Current research in disguised face recognition (DFR) typically is based on a single-domain setting [2] [3]. Specifically, an algorithm first learns a Convolutional Neural Networks (CNN) model from the training data, and then applies it to the test data. When the training data and testing data shares the same distribution, the learnt CNN model generally works well, since in this case the training error is an optimal estimate of the test error.

However, in real world applications, there is a need for transferring the learned knowledge from a source domain with abundant labeled data to a target domain where data is unlabeled or sparsely labeled. When CNN models trained on one domain and used on another domain with different distributions, the performance drops dramatically due to the domain bias [4]. To this end, we propose to solve the disguise face recognition task using domain adaptation [5] [6], which attempts to transfer the rich knowledge from the source domain, which is fully annotated, to another, different but related, domain to obtain a better CNN model.

Recently, attention transfer has been proposed and successfully adopted in several domain adaptation tasks [7] [8], which attempts to transfer attention knowledge from a powerful deeper network that is trained with sufficient training sam-

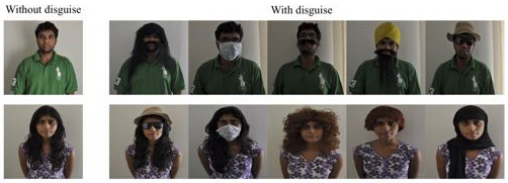

Fig. 1. Two samples of images with different disguise accessories.

ples to a shallower network that can be trained with limited training data with the goal of improving the performance of the latter. However, it is still challenging to train such a high-quality cross-domain model for the DFR due to the large domain shift in the images. To deal with the large domain shift between source domain and target domain for the DFR, we can adopt the data in source domain to synthesize disguised face images as similar as the data in target domain by using generative adversarial networks (GAN) model, which has been proven to generate impressively realistic faces through a two-player game between a generator and a discriminator. For the GAN model, there are many promising image-to-image translation developments [9] [10], but they do not necessarily preserve the identity label of an image. Although the generated image may "look" like that it comes from the auxiliary domain, the underlying identity may be lost after image-image translation. Consequently, the desired model for our task is that it can generate disguised face images which should simultaneously preserve the identity label in source domain and transform helpful content information in target domain.

Inspired by the above discussions, we propose a novel Unsupervised Domain Adaptation Model (UDAM), which jointly transfer the rich knowledge from the source domain and discriminative representation end-to-end that mutually boost each other to achieve the disguised face recognition of target domain. In particular, UDAM includes a Domain Style Adaptation subNet (DSN) and a Attention Learning subNet (ALN) to learn the representations. The DSN introduces unsupervised cross-domain adversarial training and a "learning to learn" strategy with the Siamese discriminator to achieve stronger generalizability and high-fidelity, underlying identity preserving face generation. In this setting, the model can satisfy the specific requirement of retaining identity information after image-image translation in the disguised face recognition, and we are able to create a dataset which has the similar style of the target domain in an unsupervised manner. ALN is a Convolutional Neural Network (CNN) for disguised face recognition with our proposed attention transfer strategy. The CNN model is trained by taking advantage of the sufficient labeled generated images, unlike previous approaches that distill knowledge through class probabilities [11], we propose to learn class-specific energy functions on spatial attention map, which is helpful to obtain an effective CNN model that less affected from the domain shift.

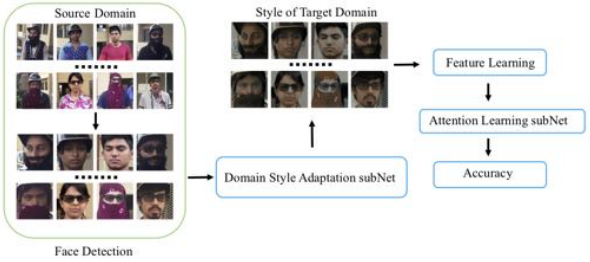

Fig. 2. Unsupervised Domain Adaptation Model (UDAM) for disguised face recognition. First, we predict face and landmark location by MTCNN [12], then the Domain Style Adaptation subNet translates the style of the labeled images from a source dataset to the style of the target dataset. Finally, we train the CNN model with the translated images and use Attention Learning subNet to obtain the disguised face recognition.

Our contributions can be summarized as follows:

- We present a deep architecture unifying image-image translation and disguised face recognition in a mutual boosting way, which inherits the merits of existing domain bias disguised face recognition methods. The proposed model achieves consistent improvement on both controlled and in-the-wild datasets.

- The local and global structural consistency of the style-translated disguised face images has been effectively enforced through pixel cycle-consistency and discriminative loss. Besides, the class-discriminative spatial attention maps from the CNN model trained by source domain are leveraged to boost the performance of disguised face recognition in target domain.

2. PROPOSED METHOD

2.1. Problem Definition

Suppose a labeled dataset A, is used to train a CNN model M_c of disguised face recognition. If the trained M_c is directly applied to a target unlabeled dataset B collected from an entirely different domain with a different set of identities/classes, the model tends to have poor performance, due to the significant differences between A and B. Therefore, we attempt to learn an optimal CNN model for B using knowledge transferred from A.

2.2. Unsupervised Domain Adaptation Model (UDAM)

As shown in Fig.2, the proposed Unsupervised Domain Adaptation Model (UDAM) consists of a Domain Style Adaptation subNet (DSN) and an Attention Learning subNet (ALN)

that jointly generate the domain-aware data and learn the disguised face representation end-to-end. We now present each component in detail.

2.2.1. Domain Style Adaptation subNet (DSN)

We first introduce a mapping function G from source domain A to target domain B and train it to produce images that fool an adversarial discriminator D_B. Conversely, the adversarial discriminator attempts to classify the real target data from the source generated data. This corresponds to the loss function:

$$\begin{aligned}\mathcal{L}_{B_{adv}}(G, D_B, P_x, P_y) =& E_{y\sim p_y}[(D_B(y) - 1)^2] + \\ & E_{x\sim p_x}[(D_B(G(x))^2],\end{aligned} \tag{1}$$

where p_x and p_y denote the sample distributions in the source and target domain, respectively. However, with large enough capacity, a network can map the face images in the source domain to any random permutation of images in the target domain. As a result, it is undesirable in the DFR task, where we have to ensure the quality of the generated faces. Thus, we introduce another mapping F from target to source and train it according to the same GAN loss, i.e.,

$$\begin{aligned}\mathcal{L}_{A_{adv}}(F, D_A, P_y, P_x) =& E_{x\sim p_x}[(D_A(x) - 1)^2] + \\ & E_{y\sim p_y}[(D_A(F(y))^2],\end{aligned} \tag{2}$$

We then introduce a cycle-consistency loss [10] to recover the original image after a cycle of translation and reverse translation, thereby enforcing cycle-consistency and preserving local structural information of the face images in source domain. The cycle-consistent loss can be expressed as:

$$\begin{aligned}\mathcal{L}_{cyc}(G, F) =& E_{x\sim p_x}[||F(G(x)) - x||_1] + \\ & E_{y\sim p_y}[||G(F(y)) - y||_1],\end{aligned} \tag{3}$$

To encourage the domain style adaptation to preserve the identity information for each translated image, inspired by [13], we add the contrastive loss [14] in the cycle-consistency loss function to learn a latent space that constrains the learning of the mapping function. We use the contrastive loss [14] to train the Siamese network as follows:

$$\mathcal{L}_{con}(l, i_1, i_2) = (1 - l)\left\{max(0, m - d)\right\}^2 + ld^2, \tag{4}$$

where i_1 and i_2 are a pair of input vectors, which are selected in an unsupervised manner. d denotes the Euclidean distance between normalized embeddings of two input vectors, and l represents the binary label of the pair. If i_1 and i_2 are positive image pair, l equals one. On the contrary, if i_1 and i_2 are negative image pair, l equals zero. $m \in [0, 2]$ represents the margin that defines the separability in the embedding space. The loss of the negative training pair is not back-propagated in the system when m equals zero. Both positive and negative sample pairs are considered if m is larger than zero. A larger m means that the loss of negative training samples has a higher weight in back propagation.

Based on the prior knowledge that the set of ID information is different in the source and target domains, there are two types of negative training pairs designed for generators G and F: 1) $G(i_A)$ and i_B, 2) $F(i_B)$ and i_A. Thus, a translated image should be of different ID information from any target image. Accordingly, the two dissimilar images are pushed away by the network. Taken together, the final Domain Style Adaptation subNet objective can be written as in equation (5) by considering Eqs (1), (2), (3), and (4):

$$\mathcal{L}_{sum} = \mathcal{L}_{B_{adv}} + \mathcal{L}_{A_{adv}} + \mathcal{L}_{cyc} + \mathcal{L}_{con} \tag{5}$$

2.2.2. Attention Learning subNet (ALN)

Baseline Deep DFR Model. Given that the style-translated dataset consisting of the translated images and their associated labels, the ResNet-50 [15] model is slightly improved and used in our experiments as the base network. It is pre-trained on the ImageNet [16] dataset, and fine-tuned on the translated images to classify the training identities. We discard the last 1000-dimensional classification layer and add two fully connected (FC) layers. Besides, to reduce the possibility of overfitting, a dropout layer [17] has been inserted before the final convolutional layer. The last fully-connected layer is modified to have N neurons to predict the N-classes, where N is the number of the classes in the training set.

Attention transfer Learning. Once we obtain the CNN model for the style-translated dataset, we can further address the domain shift problem by using spatial attention map to exploit features from the convolutional layer. Class information and more general convolutional feature are incorporated through attention map, hence more transitions can be made across domains. Let $n \in (1, 2, ..., N)$ be the n-th pre-defined class of the real images in the target domain, where N is the number of classes. For a particular example x with single ground-truth label y, the last convolutional layer of the trained CNN model will produce K feature maps A^k. The image x is first forwardly propagated through the trained CNN model, then we adopt the Grad-CAM [18] to generate the spatial attention map $\mathcal{L}(x, y_n)$ by a weighted combination of the convolutional feature maps,

$$\mathcal{L}(x, y_n) = ReLU(\sum_x \alpha_k^{y_n} A^k) \tag{6}$$

The importance of the k-th feature map for the prediction class y_n will be captured by the weight $\alpha_k^{y_n}$ through calculating the back propagating gradients to the convolutional feature map A_k. For the spatial attention map of each image, an energy function has been defined as $\frac{E(\mathcal{L}(x,y_n)}{\sum_{n=1}^{N} E(\mathcal{L}(x,y_n))}$, which is the largest when $y = y_n$, and smaller otherwise. We define E based on a simple yet effective observation: Assuming that the CNN model has been pre-trained on the style-translated source domain to predict certain identity, given an image and its spatial attention map corresponding to an identity, if the

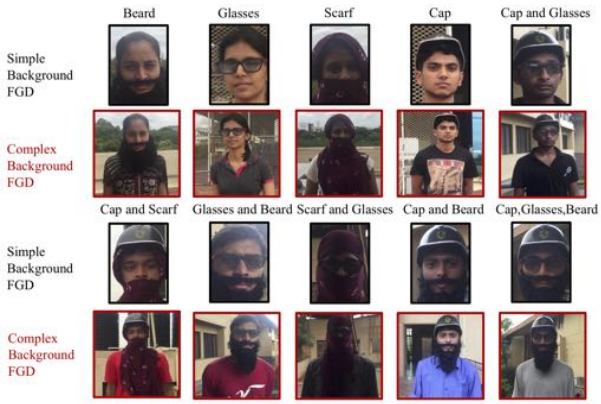

Fig. 3. The illustration shows samples images with different disguises from both the Simple and Complex face disguise (FG) datasets.

Fig. 4. Sample images from the IIIT-Delhi Disguise Version 1 Face Database (ID V1 Database).

facial attribute of the identity exists in the certain region, the attention map will generate the higher activations in the corresponding region. Therefore, a sliding window with size of 4×4 and step size of 1 will be applied over $\mathcal{L}(x, y_n)$. Then we calculate the sum of the value of $\mathcal{L}(x, y_n)$ within each sliding window as the local activation. We use the energy E to express the maximum of all local activations. For the target domain with N classes, we calculate the output score over each label as the mean energy across all local activations,

$$score(x, y_n) = \frac{1}{N} \sum_{C} E(\mathcal{L}(x, y_n)), \tag{7}$$

where C denotes the number of local activations. We infer the one with highest score as the predicted label,

$$y_p = \underset{y_n}{argmax} \ score(x, y_n) \tag{8}$$

3. EXPERIMENT

3.1. Datasets

The Simple and Complex Face Disguise Dataset [3] contain 2000 images of 25 people with 10 different disguises varied each with (i) Simple and (ii) Complex backgrounds that contain people with 8 different background illuminations in the wild. The dataset is split into three fixed parts: 1000 training images, 500 validation images and 500 test images. The example images from each dataset are shown in Fig. 3. We can observe that the samples from the complex background dataset have a relatively complex background compared to the simple dataset.

The IIIT-Delhi Disguise Version 1 Face Database (ID V1 Database) [1] contains 681 visible spectrum images of 75 participants with disguise variations. The dataset is randomly divided into a training set with 35 subjects and a testing set with the remaining 40 subjects. All the face images are almost taken under constant illumination with neutral expression and frontal pose. The sample images from the database are shown in Fig.4.

3.2. Implementation Details

Domain Style Adaptation model. We used Tensorflow [19] to train Domain Style Adaptation subNet using the training images of the dataset. Before the training process, we apply the MTCNN [12] to perform face detection for datasets and reduce the negative affect of the background. With an initial learning rate of 0.0002, and model stops training after 7 epochs. During the testing procedure, we employ the Generator G for Simple and Complex FGD $\rightarrow$ ID V1 Database translation and the Generative F for ID V1 Database $\rightarrow$ Simple and Complex FGD translation. The translated images are used to fine-tune the CNN model.

Feature learning. Specifically, ResNet-50 [15] pre-trained on ImageNet [16] is used for fine-tuning on the translated images. We modify the output of the last fully-connected layer to 25 and 35 for Simple and Complex FGD and ID V1 Database, respectively. A mini-batch SGD is used to train the CNN model on a GTX 1080 GPU. The initial learning rate is set to 0.001, and decays to 0.0001 after 10 epochs. The trained CNN is then used to generate spatial attention maps for test images in target domain. We set the size of the attention map for ResNet50 is 7×7.

3.3. Experiment results and Evaluation

To help analyze our model and show the benefit of each module, we design several unsupervised comparison methods as follows:

Setting-1: Source domain to target domain (S2T). This baseline uses the disguised face images in source domain to fine-tune the pre-trained CNN model and then tests it on target domain.

Setting-2: S2T_ DSN(without contrastive loss). We first train the DSN (without contrastive loss) using the source domain, and the generated disguised face images are used to train the CNN model.

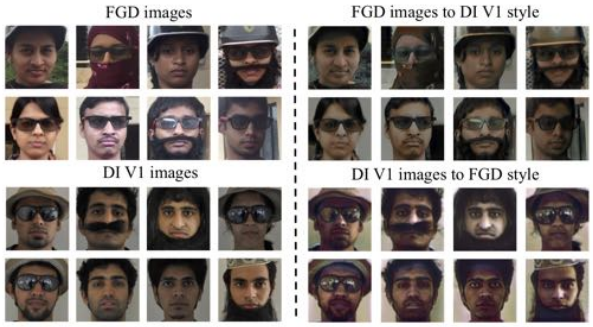

Fig. 5. Upper right: FGD images which are translated to ID V1 style; Lower right: ID V1 images translated to FGD style.

Setting-3: S2T_ DSN. This baseline preserves the identity information for each translated image by adding contrastive loss to setting 2.
Setting-4: S2T_ UDAM(DSN&ALN). Proposed unsupervised domain adaptation method in this paper.

3.3.1. Evaluation on the Simple and Complex Face Disguise Dataset

We first evaluated our method on the Simple and Complex Face Disguise Dataset, which is a disguised face dataset in the wild with varied disguises, covering different backgrounds and under varied illuminations. We translated the image style of ID V1 Database (source domain) to Simple and Complex Face Disguise Dataset (target domain) and then use the translated images to train the disguised face recognition model. Finally, we evaluated the methods on the test set of Simple and Complex Face Disguise Dataset.
Results. Table 1. shows the detailed comparison results between our methods and three aforementioned baseline methods. The proposed method outperforms all the corresponding baselines with 8% to 12.6% improvement and 7.2% and 14.7% on the DFR accuracy for simple and complex version, respectively. We attribute this to the image generator and attention learning strategy in our method. Based on the results in Table 1, it is clear that S2T_ DSN(without contrastive loss) can achieve better performance with the S2T baseline, demonstrating its efficacy to transfer style across domains. With the help of contrastive loss, we preserve the identity information during the image translation process leading to 3% and 5.1% improvement over the Setting-2 for simple and complex version, respectively. Examples of translated images by DSN are shown in Fig. 5.
Comparison with state-of-the-art. Since all of the previous approaches are not unsupervised learning setting, we compared our method with the state-of-the-art supervised learning methods including DFI [3] and ITE [1] in Table 4. For complex FGD, we arrive at an accuracy = 66.1%, which is +3.5% higher than the best results in [3]. Compared with the second best method, ITE [1], our unsupervised domain adaptation method is +1.7% and +12.7% higher in accuracy for Simple and Complex FGD, respectively. The comparisons indicate the competitiveness of the proposed method on the simple and complex FG dataset.

Table 1. Face disguise classification accuracy (%) of our four unsupervised comparative settings on the Simple and Complex Face Disguise Dataset.

Method	Simple FGD	Complex FGD
S2T	54.6%	51.4%
S2T_ DSN (without $\mathcal{L}_{cyc}$)	56.2%	53.8%
S2T_ DSN	59.2%	58.9%
S2T_ UDAM (DSN&ALN)	**67.2**%	**66.1**%

Table 2. Comparison with state-of-the-art methods on Simple and Complex Disguised Face Dataset.

Method	Simple FGD	Complex FGD
DFI [3]	**78.4**%	62.6%
ITE [1]	65.2%	53.4%
S2T_ UDAM (DSN&ALN)	67.5%	**66.1**%

3.3.2. Evaluation on the IIIT-Delhi Disguise Version 1 face database (ID V1 Database).

To further test the effectiveness of our method, we treated the Simple and Complex Face Disguise Dataset and ID V1 Database as source domain and target domain, respectively.
Results. In Table 3, we show the face recognition performance comparison of our method with some baselines. There are several findings from the results. Firstly, the recognition accuracy shown in the last column of this table indicates that the proposed model drastically improve the performance, and the degree of improvement varies between 6% and 15.5%. This well verifies the proposed method is effective when the data in the target domain is limited and unlabeled, which is the general scenario for unsupervised domain adaptation problems. Moreover, the joint learning scheme of domain style

Table 3. Face disguise classification accuracy (%) on the IIIT-Delhi Disguise Version 1 Face Database (ID V1 Database).

Method	ID VI Database
NoImage+ResNet	41.3%
S2T	29.7%
S2T_ DSN (without $\mathcal{L}_{cyc}$)	35.8%
S2T_ DSN	39.2%
S2T_ UDAM (DSN&ALN)	**45.2**%

adaptation and attention transfer learning also helps, since the two sub-nets leverage each other during end-to-end training to achieve a final win-win outcome.

Comparison with state-of-the-art. We can not find existing methods that conduct experiments on this dataset under the same conditions with us. Thus we directly create baseline NoImage+Resnet, where we directly use the training set of ID V1 to fine-tune a Resnet-50 model. Table 4 shows our methods can achieve better recognition accuracy of 45.2%.

4. CONCLUSION

In this paper, we proposed a novel Unsupervised Domain Adaptation Model (UDAM) to address the challenging face recognition with domain bias. UDAM unifies a Domain Style Adaptation subNet (DSN) and a Attention Learning subNet (ALN) for disguised face recognition in an end-to-end deep architecture. The DSN introduces unsupervised cross-domain adversarial training to provide style-translated images for effective attention transfer learning from ALN. Besides, the underlying (latent) ID information for the disguised face images also has been preserved after image-image translation. We conducted experiments on the Simple and Complex FGD and ID V1 Database, and shown the efficacy of the proposed method to adapt the domain shift problem, especially when the images in the target domain is unlabeled.

5. REFERENCES

[1] Tejas Indulal Dhamecha, Richa Singh, Mayank Vatsa, and Ajay Kumar, "Recognizing disguised faces: Human and machine evaluation," *Plos One*, vol. 9, no. 7, pp. e99212, 2014.

[2] T. I Dhamecha, A Nigam, R Singh, and M Vatsa, "Disguise detection and face recognition in visible and thermal spectrums," in *International Conference on Biometrics*, 2013, pp. 1–8.

[3] Amarjot Singh, Devendra Patil, G Meghana Reddy, and Sn Omkar, "Disguised face identification (dfi) with facial keypoints using spatial fusion convolutional network," in *ICCV Workshop*, 2017, pp. 1648–1655.

[4] A. Torralba and A. A. Efros, "Unbiased look at dataset bias," in *CVPR*, 2011, pp. 1521–1528.

[5] John Blitzer, Ryan McDonald, and Fernando Pereira, "Domain adaptation with structural correspondence learning," *Emnlp*, pp. 120–128, 2006.

[6] Diego Uribe, "Domain adaptation in sentiment classification," in *Ninth International Conference on Machine Learning and Applications*, 2011, pp. 857–860.

[7] Sergey Zagoruyko and Nikos Komodakis, "Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer," in *ICLR*, 2017.

[8] Junnan Li, Yongkang Wong, Qi Zhao, and Mohan S Kankanhalli, "Attention transfer from web images for video recognition," in *Proceedings of the 2017 ACM on Multimedia Conference*. ACM, 2017, pp. 1–9.

[9] Ming-Yu Liu and Oncel Tuzel, "Coupled generative adversarial networks," in *Advances in neural information processing systems*, 2016, pp. 469–477.

[10] Jun Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," pp. 2242–2251, 2017.

[11] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean, "Distilling the knowledge in a neural network," *arXiv preprint arXiv:1503.02531*, 2015.

[12] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1499–1503, 2016.

[13] Weijian Deng, Liang Zheng, Guoliang Kang, Yi Yang, Qixiang Ye, and Jianbin Jiao, "Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person reidentification," in *CVPR*, 2018, vol. 1, p. 6.

[14] R. Hadsell, S. Chopra, and Y. Lecun, "Dimensionality reduction by learning an invariant mapping," in *CVPR*, 2006, pp. 1735–1742.

[15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[16] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, and Michael Bernstein, "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.

[17] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[18] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, Dhruv Batra, et al., "Grad-cam: Visual explanations from deep networks via gradient-based localization.," in *ICCV*, 2017, pp. 618–626.

[19] Martn Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, and Michael Isard, "Tensorflow: a system for large-scale machine learning," 2016.