Style transfer effectiveness for forensic sketch and photo matching

Vasily Rodin
Samara National Research University
Samara, Russia
varodg2000@mail.ru

Aleksey Maksimov
Samara National Research University
Samara, Russia
aleksei.maksimov.ssau@gmail.com

***Abstract*—In this paper, the problem of comparing portrait photographic images and forensic sketches is considered. The paper analyzes the feasibility and potential advantage of applying style transfer methods to solve this problem. A method for comparing photographic and synthetic images is proposed. It consists of the feature extraction from a pair of images, the subsequent element-by-element difference between feature vectors, and further classification. We also consider a modification of the proposed method, which uses the style transfer from a sketch to a photographic image. Experimental research of the two mentioned methods is carried out on a test set of image and sketch pairs. Its results show the advantage of the modified method over the initial one.**

Keywords—style transfer, face recognition, forensic sketch, photo-sketch recognition, machine learning

I. Introduction

The task of automatic comparison of a forensic sketch of a suspect with a photograph from a police database or surveillance video stream is essential for modern-day criminology. However, despite the rapid development of image processing and pattern recognition methods, relatively few works are dedicated to the problem of analyzing synthetic and photographic images. Moreover, a significant part of the work devoted to this topic is focused more on the task of generating a sketch based on a photograph, rather than on the task of forensic sketch face recognition [1-4].

Among the well-known solutions to the discussed forensic sketch recognition problem, it is worth highlighting works based on the usage of HoG [5, 6] and SIFT [7] key-point descriptors. In these works, features in the photographs and sketches are independently extracted and concatenated, after which the distance between the obtained feature vectors is calculated. It is also worth mentioning a well-known solution based on the use of generative networks [8], where the latent feature space is taken advantage of to match the photo to its sketch. The possibility of applying style transfer methods to solve the problem of comparing photographic and synthetic portraits is understudied. The aim of the paper is to explore this possibility.

II. Considered methods

A. Sketch and Photo Matching

We propose a comparison of a portrait photograph and a forensic sketch to be carried out as follows. The input of the method is a pair of images - a photographic and a synthetic one. The output of the proposed method is the binary classification result - whether both images belong to the same person or not.

The initial method considered method is assumed to be as follows. First, features are extracted from the images, after which the forensic sketch face embedding is subtracted elementwise from the photo face embedding. Next, dimensionality reduction is performed. In our research, we use principal component analysis (PCA) [9]. In our work, the best classification metrics were obtained while maintaining the 16 most significant components of the original feature vectors, so we suggest using that number for the current experimental setup. The final step is the binary classification of the resulting feature vector. For classification in this research, quadratic discriminant analysis (QDA) [10] was used. Fig. 1(a) demonstrates the scheme of the proposed method.

A modified version of this method supposes style transfer from a forensic sketch to a photographic image before further feature extraction. Figure 1(b) presents the modified version of the method.

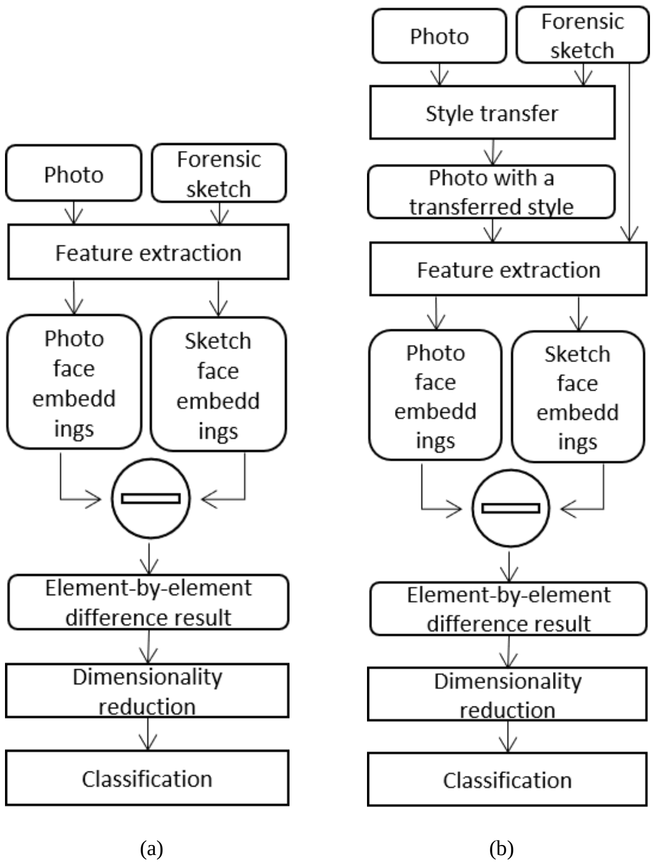


Fig. 1. The examined sketch-photo matching methods - the original (a) and the modified one (b).

B. Face Embeddibgs Extraction

In this paper, the extraction of features (face embeddings) is based on the ResNet50 model [11] with the usage of the ArcFace loss function [12]. The function is considered to be a SOTA loss function for face recognition tasks according to its creators. Mathematically, it can be represented in the following form:

$$L = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{s\left(\cos\left(\theta_{y_i}+m\right)\right)}}{e^{s\left(\cos\left(\theta_{y_i}+m\right)\right)}+\sum_{j=1,\, j\neq y_i}^{n} e^{s\cos\theta_j}}, \qquad (1)$$

where L - a loss function value,

N - a number of deep feature samples,

y_i - a class, which i^th deep feature sample belongs to,

s - a scale factor,

m - an additive angular margin penalty,

θ_j - an angle between j^th weight and i^th deep feature sample.

We use a pre-trained on the MS-Celeb-1M dataset [13] model. The specified dataset contains 8.2 million images of 100,000 identities. ResNet50 is a convolutional neural network of 50 layers. There are 48 convolutional and two pooling ones. The advantage of this architecture is the application of so-called shortcut connections. They allow skipping one or more layers during the network training process and perform ID matching. Shortcut connections improve the efficiency of training networks with a large number of layers. Fig. 2. shows the architecture of the network.

C. Style Transfer

Style transfer is a digital image manipulation technique that allows recomposing the content of an image in the style of another one. The style transfer algorithm transforms the input image (commonly called "content image") in accordance with the color content and texture obtained from the second image (commonly called "style image"). This paper explores the possibility of using this technology to improve the quality of comparison of natural photographic images and forensic sketches.

In this research, the network [14] was used for style transfer. This architecture can be considered "classic" for the image style transfer application field. It consists of 3 sub-networks: style prediction network (based on Inception-v3 [15] architecture), style transfer network (based on [16] style transfer architecture), and content and loss calculation network (based on VGG [17] architecture). Fig. 3 displays the workflow of the neural network ensemble used in this research for the style transfer stage of the proposed modified method.

III. USED DATA

For the experimental research of our work, we use Tufts Face Database images [18]. This dataset contains 112 portrait photographic images and corresponding forensic sketches.

For this work, a balanced set of pairs [photo embedding, composite portrait embedding] is assembled from the mentioned dataset, containing 112 correct pairs, where both embeddings belong to the same person, and 112 incorrect ones, where embeddings belong to different people. The selected QDA classifier is trained on 70% of this dataset, and 30% is used as a test set.

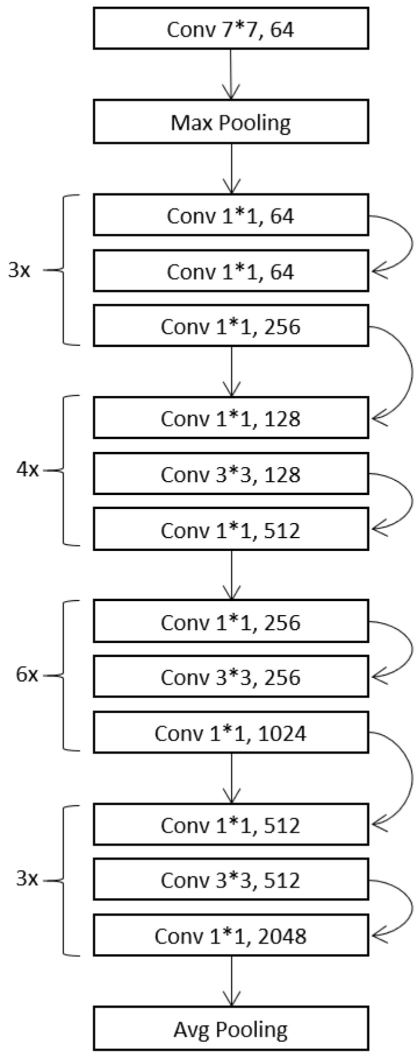


Fig. 2. Architecture of the ResNet50 [11] network, which is used in this research.

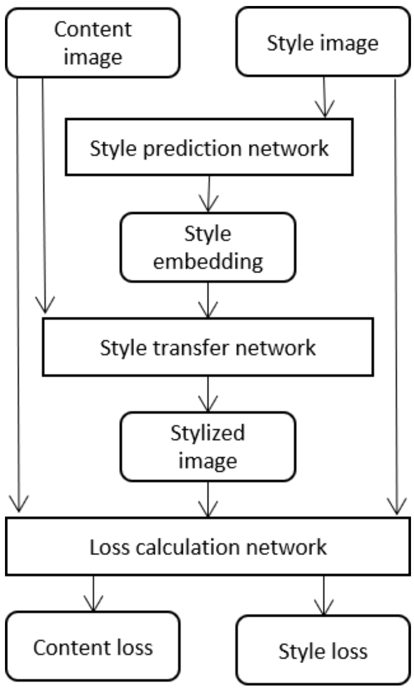


Fig. 3. Workflow of the image style transfer method [14] used in the research

Fig. 4. shows several examples of the dataset images and style transfer results.

(a) (b) (c)

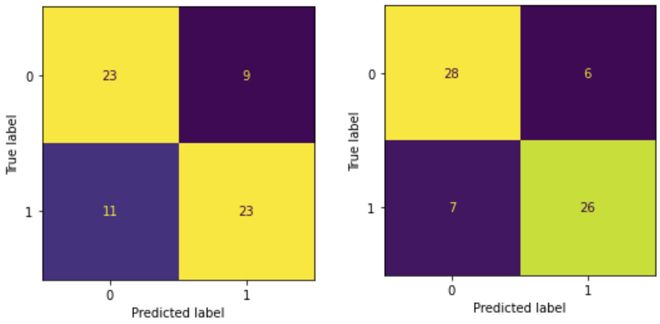
Fig. 4. Examples of images used in research - a photographic portrait (a), a synthetic forensic sketch(b), and a style transfer result (c).

IV. Experimental Research Results

During the experimental research, the values of the precision, recall, accuracy, and f-score metrics [19] were obtained for derived feature vectors classifier. Metric were calculated for both with and without style transfer cases. As can be seen from the results in Table 1, the use of style transfer significantly improves the quality of classification - about 10-12%. Fig. 5 also shows the confusion matrices for the test set.

TABLE I. Classification Metrics Results

Considered method	Precision	Recall	Accuracy	F-score
Without style transfer (pic. 1(a))	0.72	0.67	0.71	0.69
With style transfer (pic. 1(b))	0.82	0.79	0.81	0.80



(a) (b)

Fig. 5. Confusion matrices for the method withot style transfer (a) and for the method with style transfer (b).

V. Discussion and Future Work

The paper considers the problem of comparing photographic and synthetic portrait images. Within the framework of the proposed method, the advantage of applying style transfer to improve the quality of image pair classification is demonstrated. The improvement in classification metrics was about 10-12%.

However, as can be seen from the results shown in Table 1, the overall quality of the proposed method remains below the level that would allow it to be used in practice for the tasks of matching a sketch with a portrait image for forensic purposes. However, the aim of this work is to find out the potential performance of style transfer in this application area. The results show the potential of proposed method, as well as the need for further improvement.

In future research, the authors plan to improve the stages of the proposed method. It is planned to conduct a comparative study of existing face embedding extraction models in order to select the suitable one for the developed method. Is also planned to conduct a comparative study of loss functions to select the appropriate one. ArcFace was used in this work, but it is possible that using CosFace [20] or SphereFace [21] may give better results for such a specific application area as forensic sketch face recognition. It is also planned to compare the quality of different classification methods for the last stage of the proposed method, as well as the impact of different style transfer architectures on the quality of the method.

After researching all mentioned before technologies, the proposed method can be compared with modern solutions, for example, with works [6, 22, 23].

References

[1] Sannidhan, M.S. Evaluating the performance of face sketch generation using generative adversarial network / M.S. Sannidhan, G.A. Prabhu, D.E. Robbins, C. Shasky // Pattern Recognition Letters. – 2019. – Vol.128. – P.452-458. doi: https://doi.org/10.1016/j.patrec.2019.10.010

[2] Chen, C. Face sketch synthesis with style transfer using pyramid column feature / C. Chen, X. Tan, K.Y.K. Wong //2018 IEEE Winter Conference on Applications of Computer Vision. – 2018. – P. 485-493.

[3] Pallavi, S. A novel approach for generating composite sketches from mugshot photographs / S. Pallavi, M.S. Sannidhan, K.B. Sudeepa, A. Bhandary //2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI). – 2018. – P. 460-465.

[4] Zhou, H. Markov weight fields for face sketch synthesis / H. Zhou, Z. Kuang, K.Y.K. Wong // 2012 IEEE Conference on Computer Vision and Pattern Recognition. – 2012. – P. 1091-1097.

[5] Narkar, N. Forensic Sketch Recognition using Image Processing / N. Narkar, L. Gasper, A. Menezes, M. Patil, //Advancements in Image Processing and Pattern Recognition. – 2021. – Vol. 4(1). – P.1–8.

[6] Dalal, S. Feature-based Sketch-Photo Matching for Face Recognition / S. Dalal, V.P. Vishwakarma, S. Kumar // Procedia Computer Science. – 2020. – Vol.167. – P.562– 570.

[7] Klare, B. Sketch to Photo Matching: A Feature-based Approach / B. Klare, A. K. Jain // Proceedings of SPIE - The International Society for Optical Engineering. – 2022. – Vol.7667. – №766702.

[8] Bae, S. Exploiting an Intermediate Latent Space between Photo and Sketch for Face Photo-Sketch Recognition / S. Bae, N.U. Din, H. Park, J. Yi // Sensors. – 2022. – Vol.22(19) . – №7299. https://doi.org/10.3390/s22197299

[9] Jolliffe, I.T. Principal component analysis: a review and recent developments / I.T. Jolliffe, J. Cadima // Phil. Trans. R. Soc. – 2016. – A.3742015020220150202.

[10] Cao, W. Quadratic Discriminant Analysis Revisited [Electronic resource] // CUNY Academic Works. – 2015. – https://academicworks.cuny.edu/gc_etds/536

[11] He, K. Deep Residual Learning for Image Recognition / K. He, X. Zhang, S. Ren, J. Sun // 2016 IEEE Conference on Computer Vision and Pattern Recognition. –2016. – P. 770-778.

[12] Deng, J. ArcFace: Additive Angular Margin Loss for Deep Face Recognition / J. Deng, J. Guo, J. Yang, N. Xue, I. Kotsia, S. Zafeiriou // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2022. – Vol.44(10). – P. 5962-5979. doi: 10.1109/TPAMI.2021.3087709.

[13] Guo, Y. MS-Celeb-1M: A Dataset and Benchmark for Large-Scale Face Recognition. / Y. Guo, L. Zhang, X. Hu, X. He, J. Gao //. ECCV 2016. Lecture Notes in Computer Science. – 2016. – Vol. 9907. doi: https://doi.org/10.1007/978-3-319-46487-9_6

[14] Ghiasi, G. Exploring the structure of a real-time, arbitrary neural artistic stylization network/ G. Ghiasi, H. Lee, M. Kudlur, V. Dumoulin, J. Shlens // Proceedings of the British Machine Vision Conference (BMVC). – 2017.

[15] Szegedy, C. Rethinking the inception architecture for computer vision / C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna // IEEE Computer Vision and Pattern Recognition. – 2016. – P. 2818-2826.

[16] Dumoulin, V. A learned representation for artistic style / V. Dumoulin, J. Shlens, M. Kudlur // International Conference of Learned Representations. – 2016. – arXiv preprint arXiv:1610.07629. – 2016.

[17] Gatys, L.A. A Neural Algorithm of Artistic Style / L.A. Gatys, A.S. Ecker, M. Bethge // Journal of Vision. – 2016. – Vol.16. – №326.

[18] Panetta, K. A comprehensive database for benchmarking imaging systems / K. Panetta, Q. Wan, S. Agaian, S. Rajeev, S. Kamath, R. Rajendran, S. Rao // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2018. – Vol.42(3). – P. 509-520, doi: 10.1109/TPAMI.2018.2884458.

[19] Christopher M. Bishop, Pattern Recognition and Machine Learning (Information Science and Statistics). Springer-Verlag, Berlin, Heidelberg, 2006.

[20] Wang, H. CosFace: Large Margin Cosine Loss for Deep Face Recognition / H. Wang, Y. Wang, Z., Zhou, X. Ji, Z. Li, D. Gong, J. Zhou, W. Liu// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. – 2018. – P. 5265-5274.

[21] Liu, W. SphereFace: Deep Hypersphere Embedding for Face Recognition / W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, L. Song // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) . – 2017. – P. 6738-6746.

[22] Galea, C. Forensic Face Photo-Sketch Recognition using a Deep Learning-based Architecture / C. Galea, R. A. Farrugia// EEE Signal Processing Letters. – 2017. – Vol.24(11). – P. 1586-1590. doi: 10.1109/LSP.2017.2749266

[23] Bae, S. Face Photo-Sketch Recognition Using Bidirectional Collaborative Synthesis Network / S. Bae, N.U. Din, H. Park, J. Yi // 2022 16th International Conference on Ubiquitous Information Management and Communication (IMCOM). –2022. – P. 1-8.