 DOI: 10.3217/978-3-85125-528-7-07

Tattoo Detection for Soft Biometric De-Identification Based on Convolutional Neural Networks*

Tomislav Hrkać[1], Karla Brkić[1], and Zoran Kalafatić[1]

[1] Faculty of Electrical Engineering and Computing
University of Zagreb, Croatia
{*tomislav.hrkac, karla.brkic, zoran.kalafatic*}*@fer.hr*

Abstract

Nowadays, video surveillance is ubiquitous, posing a potential privacy risk to law-abiding individuals. Consequently, there is an increased interest in developing methods for de-identification, i.e. removing personally identifying features from publicly available or stored data. While most of related work focuses on de-identifying hard biometric identifiers such as faces, we address the problem of de-identification of soft biometric identifiers – tattoos. We propose a method for tattoo detection in unconstrained images, intended to serve as a first step for soft biometric de-identification. The method, based on a deep convolutional neural network, discriminates between tattoo and non-tattoo image patches, and it can be used to produce a mask of tattoo candidate regions. We contribute a dataset of manually labeled tattoos. Experimental evaluation on the contributed dataset indicates competitive performance of our method and proves its usefulness in a de-identification scenario.

1. Introduction

In the last decade, video surveillance has spread to almost all aspects of daily life. Storing the recorded surveillance data in its unprocessed form poses a privacy risk to law-abiding individuals, as their whereabouts and activities can be exposed without their consent. Privacy concerns are aggravated by the development of various video retrieval techniques [17, 26, 16] that enable searching for content in large volumes of video data, as well as by the development of techniques for person re-identification across different video sequences [1, 8]. In order to minimize privacy risks, many jurisdictions implement strict regulations for the protection of personal data (see e.g. the Data Protection Directive of the European Union[1]). For video sequences, protection of personal data entails obfuscating or removing personally identifying features of the recorded individuals, usually in a reversible fashion so that law enforcement can access them if necessary.

The process of removing personally identifying features from data is called de-identification. One of the most commonly used de-identification techniques, used in commercial systems such as Google Street View, involves detecting and blurring the faces of recorded individuals. However, this approach ignores soft biometric and non-biometric features like clothing, hair color, birthmarks or tattoos, that can be used as cues to identify the person [6, 20]. In this paper, we propose a method for detecting

*This work has been supported by the Croatian Science Foundation, within the project "De-identification Methods for Soft and Non-Biometric Identifiers" (DeMSI, UIP-11-2013-1544). This support is gratefully acknowledged.

[1] http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:31995L0046

tattooed skin regions that can be used in an advanced de-identification pipeline to obfuscate or remove tattoos. We train a convolutional neural network that acts as a patch classifier, labeling each patch of an input image as either belonging to a tattoo or not.

2. Related work

Current research in de-identification is mainly concerned with de-identifying hard biometric features, especially the face [9]. Considerably less volume of research is devoted to soft and non-biometric features [20]. Tattoo detection is typically studied not in the context of de-identification, but in forensic applications. There the goal is to build a content-based image retrieval system for tattoos that would help law enforcement in finding suspects and other persons of interest, e.g. persons associated with a particular gang etc. [6, 12, 10]. For instance, Jain et al. [12] propose a content-based image retrieval system intended to be used by law enforcement agencies. The query image is a cropped tattoo, which is then segmented, represented using color, shape and texture features and matched to the database. Han and Jain [10] take the concept further by proposing a content-based image retrieval system for sketch-to-image-matching, where a sketch of the tattoo is matched to real tattoo images. Their system uses SIFT descriptors to model shape and appearance patterns in both the sketch and the image, and matches the descriptors using a local feature-based sparse representation classification scheme. Kim et al. [13] propose combining local shape context, SIFT descriptors and global tattoo shape for tattoo image retrieval. Their descriptor is robust to partial shape distortions and invariant to translation, scale and rotation.

The methods used in content-based image retrieval systems often assume that tattoo images are cropped, which limits their potential use in other scenarios. Heflin et al. [11] consider detecting scars, marks and tattoos "in the wild", i.e. in uncropped images, where a tattoo can appear anywhere in the image (or not appear at all) and be of arbitrary size. They propose a method for tattoo detection where tattoo candidate regions are detected using graph-based visual saliency. Further processing of the candidate regions utilizes the GrabCut algorithm [21], image filtering and the quasi-connected components technique [4] to obtain the final estimate of the tattoo location.

Wilber et al. [25] propose a mid-level image representation called Exemplar Codes and apply it to the problem of tattoo classification. Exemplar codes are feature vectors that consist of normalized outputs of simple linear classifiers. Each linear classifier measures the similarity between the input image and an exemplar, i.e. a training image that best captures some property of the tattoo. Decision score outputs from individual linear classifiers are used to estimate probabilities using extreme value theory [23], thus forming exemplar code feature vectors. A random forest classifier is trained on exemplar codes, enabling multi-class tattoo recognition.

Because of great variability of tattoo designs, individual skin color and lighting conditions in real-world tattoo images, as well as the fact that the tattoos resemble many different real world objects, it is very difficult to devise good hand-crafted features suited for differentiating between tattoos and background [19]. In recent times, however, convolutional neural networks (CNNs) were shown to be able to automatically learn good features for many classification tasks [15]. We therefore propose to apply a deep convolutional neural network to the difficult problem of tattoo detection. In seminal work by Krizhevsky et al. [14], convolutional neural networks were proven to be extremely successful on the ImageNet dataset. According to LeCun et al. [15], this success can be attributed to several factors: efficient use of GPUs for network training, use of rectified linear units, use of dropout regularization and augmenting the training set with deformations of the existing images. Convolutional

networks have already been successfully applied to the problem of scene labelling [7] and semantic segmentation [18].

In contrast to related work, in this paper we take a bottom-up approach. Our CNN-based model operates at the level of small image patches and enables classifying each patch as either belonging to a tattoo or not. Our approach can be used on arbitrary images to obtain a low-level estimate of candidate tattooed regions.

We propose this approach with our target application of de-identification in mind. In a de-identification pipeline, the detected candidate tattoo regions can be removed or averaged to remove personally identifying information. We place much greater importance on correctly detecting all tattooed regions than on eliminating false positive detections, as false positives can be eliminated in subsequent stages, e.g. by combining our method with a person detector (e.g. [5]).

3. Our method

Our proposed method for tattoo detection is based on image patch labeling using a convolutional neural network. We do not detect a tattoo as a global entity. Rather, we use the sliding window approach and at each window position we extract a patch of the size $N \times N$. The patch is then classified as either tattoo or background. The output of our method consists of masked image regions that are tattoo candidates.

Convolutional neural networks typically consist of several convolutional layers, followed by one or more fully connected layers. Convolutional layers are in charge of learning good features and they are characterized by (i) local receptive fields (i.e. the neuron in the convolutional layer is not connected to the outputs of all the neurons from the previous layer, but only to the ones in its local neighborhood), and (ii) shared weights, reflecting the intuition that the features are computed in the same way at different image locations. After the convolutional layers, the so-called pooling layers are typically inserted in order to reduce the dimensionality of feature space for subsequent steps. Fully connected layers perform the task of classification and contain the majority of learned weights.

The architecture of our network is broadly inspired by the successful VGGNet model, proposed in 2014 by Simonyan and Zisserman [24]. The VGGNet is characterized by a very homogeneous architecture that only performs 3×3 convolutions and 2×2 pooling from the beginning to the end. However, our model modifies it to accommodate smaller input images and smaller number of output classes. The simplified network, with fewer and smaller layers is faster to train and it proved adequate for our purposes. The proposed network architecture is shown in Fig. 1.

The input to the network is an $N \times N$ color image (we assumed the RGB color model). The image has to be classified either as belonging to the tattoo or not, depending on whether its center lies inside the polygon that demarcates the tattoo.

The network consists of eight layers (not counting the input layer, i.e. the image itself). The first two layers are convolutional layers with 32 feature maps with 3×3 filters and ReLU activation units. The third layer is a max-pooling layer that reduces the feature map dimensions by 2×2. The fourth and the fifth layers are again convolutional layers with ReLU activation units, but with 64 feature maps (again with 3×3 filters). The sixth layer is another max-pooling layer, once more reducing the input dimension by 2×2. The seventh layer is a fully connected layer consisting of 256 neurons. The final,

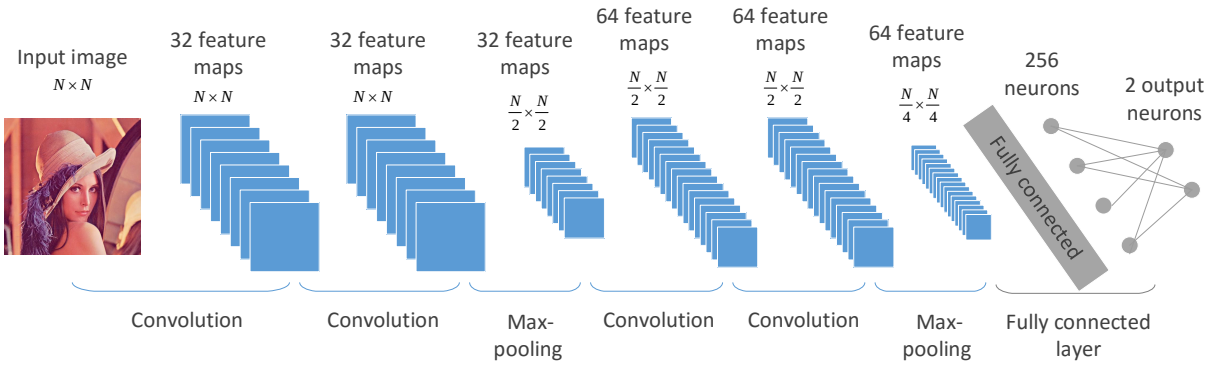


Figure 1: The architecture of the proposed ConvNet model.

Figure 2: Examples of annotated tattoo images.

eighth layer consists of two neurons with the Softmax activation function, corresponding to the two output classes. Dropout, with the dropout ratio set to 0.5, is applied to the fully connected layer.

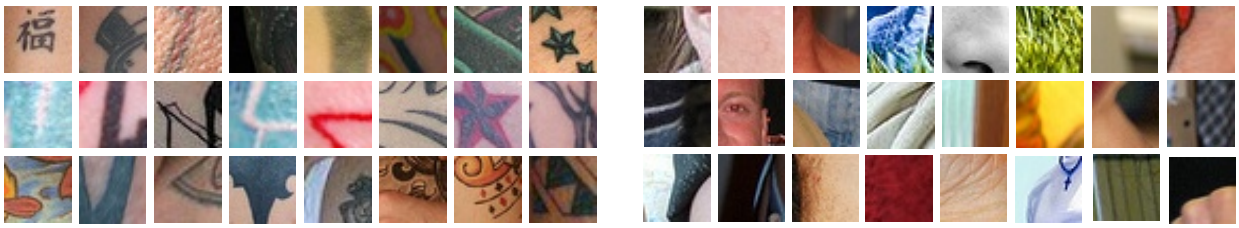
We implemented the described network in Python, using Theano [2, 3] and Keras [2] libraries.

4. Experiments

Given the relatively modest volume of work on tattoo detection, there are no readily available tattoo detection datasets. Recently, a dataset called Tatt-C has been published [19], but it cannot be freely downloaded. Hence, to facilitate the development and testing of our method we have assembled our own dataset[3] by collecting and manually labeling 890 tattoo images from the ImageNet database [22].

Each of the collected images contains one or more tattoos. We annotated each tattoo using a series of connected line segments. Example annotated images from our dataset are shown in Fig. 2. We attempted to closely capture the outline of each tattoo, which can be a challenging task, as tattoos can have highly irregular edges.

[2]https://github.com/fchollet/keras, accessed March 2016.
[3]The dataset is available at http://www.fer.unizg.hr/demsi/databases_and_code/tattoo_dataset.



(a) tattoo patches (b) background patches

Figure 3: Example extracted patches from our dataset (patch size 32×32).

4.1. Training the network

For training, we constructed a training set by randomly sampling a number of image patches of predefined size from each annotated tattoo image in our dataset. This procedure was done both for positive and negative samples, i.e. for patches that do and do not contain tattoos. Examples of extracted patches can be seen in Fig. 3.

The training of the network was carried out by optimizing the mean squared error loss function, using stochastic gradient descent with momentum. We used the mini-batch of 32 samples and the momentum was set to 0.9. The learning rate was set to 0.1. The training was performed for maximaly 40 epochs, with early stopping based on validation loss. The duration of the training varied greatly with the size of the patches, in our case from 10 minutes for smallest patches to 13 hours for the largest.

4.2. Performance evaluation

The set of all extracted patches totalled 22700 images (11359 positive and 11341 negative samples). This set was divided into sets for training (containing 15134 samples, out of which 7573 positive and 7560 negative), and validation and testing (both of the same size of 3783 samples, out of which 1893 positive and 1890 negative). We have ensured that all patches extracted from the same image end up in only one of the sets (either training, validation or testing), in order to avoid mixing training and testing data.

We trained and evaluated the network for different patch sizes (8×8, 12×12, 16×16, 24×24, 32×32 and 48×48) to determine the optimal patch size. The larger patches presumably provide more information about context, but the network that utilizes them is slower to train and test.

The test set was used for evaluation. The results are summarized in Table 1. The accuracy was calculated as a total number of misclassifications (false positives and false negatives) divided by the test set size. As we can see, the results improve in terms of accuracy with the increase in image patch size, up to the largest considered size (48×48) that gives slightly worse results than most of the smaller patch sizes. The difference in accuracy is not very pronounced; i.e. we can say that results for all the patch sizes are similar. The other thing that can be noticed is that the improvement in performance with the increase in patch size comes mainly from reducing the number of false positives, while at the same time the number of false negatives rises.

We have done a preliminary qualitative evaluation of the performance of the network in a sliding window setting. Some results are shown in Fig. 4. These examples are relatively simple, with homo-

Patch size	8×8	12×12	16×16	24×24	32×32	48×48
False negatives	152 (8.03%)	229 (12.10%)	187 (9.88%)	213 (11.25%)	248 (13.10%)	290 (15.32%)
False positives	593 (31.37%)	418 (22.12%)	444 (23.49%)	436 (23.07%)	337 (17.83%)	408 (21.59%)
Accuracy	0.8031	0.8290	0.8332	0.8283	0.8454	0.8155

Table 1: Evaluation of the network performance on different patch sizes

geneous skin around the tattoo and simple background. We see that many tattoo patches are correctly detected, but there are also some misclassifications. In more difficult examples with more background containing textured objects, the number of false positives rises. In the context of de-identification, this problem could be addressed by combining this detector with other stages of a de-identification pipeline, e.g. by eliminating detections outside of candidate person locations.

(a) the original images

(b) labeled tattoo patches

Figure 4: The output of the network on full images.

5. Conclusion and outlook

We addressed the challenging problem of tattoo detection for soft biometric de-identification. Instead of hand-crafting image features, we applied deep learning. We trained and evaluated a deep convolutional neural network using the dataset of positive and negative patches generated from a subset of ImageNet tattoo images annotated by hand. Our findings indicate that using a convolutional neural network to classify small image patches can be a reliable way to detect candidate tattoo regions in an image. Patch sizes should be kept small, up to 32×32 patches, in order to obtain best accuracy, good foreground-background segmentation and minimize false negatives.

In our future work, we plan to combine this method with other stages of a de-identification pipeline in order to solve the problem of false positives. As our qualitative analysis shows that the majority of false positives are in the surroundings rather than on the person, one possibility is to run the method only on the outputs of a person detector. We also plan to quantitatively evaluate the performance of our network on full tattoo images (as opposed to patches), and investigate whether this performance could be improved by merging the detections into blobs.

References

[1] D. Baltieri, R. Vezzani, and R. Cucchiara. Mapping appearance descriptors on 3d body models for people re-identification. *International Journal of Computer Vision*, 111(3):345–364, 2014.

[2] F. Bastien, P. Lamblin, R. Pascanu, J. Bergstra, I. J. Goodfellow, A. Bergeron, N. Bouchard, and Y. Bengio. Theano: new features and speed improvements. In *Deep Learning and Unsupervised Feature Learning NIPS 2012 Workshop*, 2012.

[3] J. Bergstra, O. Breuleux, F. Bastien, P. Lamblin, R. Pascanu, G. Desjardins, J. Turian, D. Warde-Farley, and Y. Bengio. Theano: a CPU and GPU math expression compiler. In *Proceedings of the Python for Scientific Computing Conference (SciPy)*, June 2010.

[4] T. E. Boult, R. J. Micheals, X. Gao, and M. Eckmann. Into the woods: Visual surveillance of non-cooperative and camouflaged targets in complex outdoor settings. In *Proceedings of the IEEE*, pages 1382–1402, 2001.

[5] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *In CVPR*, pages 886–893, 2005.

[6] J. eun Lee, A. K. Jain, and R. Jin. Scars, marks and tattoos (smt): Soft biometric for suspect and victim identification. In *In Proc. Biometric Symposium, Biometric Consortium Conference*, pages 1–8, 2008.

[7] C. Farabet, C. Couprie, L. Najman, and Y. LeCun. Learning hierarchical features for scene labeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1915–1929, Aug 2013.

[8] J. Garcia, N. Martinel, A. Gardel, I. Bravo, G. L. Foresti, and C. Micheloni. Modeling feature distances by orientation driven classifiers for person re-identification. *Journal of Visual Communication and Image Representation*, 38:115 – 129, 2016.

[9] R. Gross, L. Sweeney, J. F. Cohn, F. De la Torre, and S. Baker. *Protecting Privacy in Video Surveillance*, chapter Face De-identification, pages 129–146. Springer Publishing Company, Incorporated, 2009.

[10] H. Han and A. K. Jain. Tattoo based identification: Sketch to image matching. In *Biometrics (ICB), 2013 International Conference on*, pages 1–8, June 2013.

[11] B. Heflin, W. Scheirer, and T. E. Boult. Detecting and classifying scars, marks, and tattoos found in the wild. In *Biometrics: Theory, Applications and Systems (BTAS), 2012 IEEE Fifth International Conference on*, pages 31–38, Sept 2012.

[12] A. K. Jain, J.-E. Lee, and R. Jin. *Advances in Multimedia Information Processing – PCM 2007: 8th Pacific Rim Conference on Multimedia, Hong Kong, China, December 11-14, 2007. Proceedings*, chapter Tattoo-ID: Automatic Tattoo Image Retrieval for Suspect and Victim Identification, pages 256–265. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007.

[13] J. Kim, A. Parra, J. Yue, H. Li, and E. J. Delp. Robust local and global shape context for tattoo image matching. In *Image Processing (ICIP), 2015 IEEE International Conference on*, pages 2194–2198, Sept 2015.

[14] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012.

[15] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553):436–444, 05 2015.

[16] Y. Li, R. Wang, Z. Huang, S. Shan, and X. Chen. Face video retrieval with image query via hashing across euclidean space and riemannian manifold. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

[17] D. Lin, S. Fidler, C. Kong, and R. Urtasun. Visual semantic search: Retrieving videos via complex textual queries. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 2657–2664, June 2014.

[18] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on*, pages 3431–3440, June 2015.

[19] M. Ngan and P. Grother. Tattoo recognition technology - challenge (tatt-c): an open tattoo database for developing tattoo recognition research. In *Identity, Security and Behavior Analysis (ISBA), 2015 IEEE International Conference on*, pages 1–6, March 2015.

[20] D. Reid, S. Samangooei, C. Chen, M. Nixon, and A. Ross. Soft biometrics for surveillance: an overview. In *Machine Learning: Theory and Applications*, 31, pages 327–352. Elsevier, 2013.

[21] C. Rother, V. Kolmogorov, and A. Blake. "GrabCut": Interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph.*, 23(3):309–314, Aug. 2004.

[22] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252, 2015.

[23] W. Scheirer, A. Rocha, R. Micheals, and T. Boult. *Computer Vision – ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part III*, chapter Robust Fusion: Extreme Value Theory for Recognition Score Normalization, pages 481–495. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.

[24] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2014.

[25] M. J. Wilber, E. Rudd, B. Heflin, Y.-M. Lui, and T. E. Boult. Exemplar codes for facial attributes and tattoo recognition. In *Applications of Computer Vision (WACV), 2014 IEEE Winter Conference on*, pages 205–212, March 2014.

[26] G. Ye, W. Liao, J. Dong, D. Zeng, and H. Zhong. *MultiMedia Modeling: 21st International Conference, MMM 2015, Sydney, NSW, Australia, January 5-7, 2015, Proceedings, Part II*, chapter A Surveillance Video Index and Browsing System Based on Object Flags and Video Synopsis, pages 311–314. Springer International Publishing, Cham, 2015.