

# Deep Learning Architectures for Tattoo Detection and De-identification

Tomislav Hrkać\*, Karla Brkić\*, Slobodan Ribarić\* and Darijan Marčetić\*

\*University of Zagreb, Faculty of Electrical Engineering and Computing, Zagreb, Croatia

Email: {tomislav.hrkac, karla.brkic, slobodan.ribaric, darijan.marcetic}@fer.hr

**Abstract**—The widespread use of video recording devices to obtain recordings of people in various scenarios makes the problem of privacy protection increasingly important. Consequently, there is an increased interest in developing methods for de-identification, i.e. removing personally identifying features from publicly available or stored data. Most of related work focuses on de-identifying hard biometric identifiers such as faces. We address the problem of detection and de-identification of soft biometric identifiers – tattoos. We use a deep convolutional neural network to discriminate between tattoo and non-tattoo image patches, group the patches into blobs, and propose the de-identifying method based on replacing the color of pixels inside the tattoo blob area with a values obtained by interpolation of the surrounding skin color. Experimental evaluation on the contributed dataset indicates the proposed method can be useful in a soft biometric de-identification scenario.

## I. INTRODUCTION

The problem of privacy invasion has become increasingly important in recent decades, with the widespread use of video recording devices to obtain recordings of people in various scenarios. In order to reduce privacy risks, the protection of personal data is nowadays strictly regulated by law in many jurisdictions, requiring the stored data to be de-identified (e.g. the Data Protection Directive of the European Union).

In case of images, de-identification of personal data entails obfuscating or removing personally identifying features of the filmed individuals, usually in a reversible fashion so that law enforcement can access them if necessary. One of the most widespread de-identification techniques, used in commercial systems such as Google Street View, involves detecting and blurring the faces of recorded individuals. This approach, however, does not take into account soft biometric and non-biometric features like tattoos, clothing, hair color and similar, that can be used as cues to identify the person [19]. Motivated by the need for soft and non-biometric feature de-identification, we propose a method for detecting and de-identifying tattooed skin regions. We use a deep learning approach and explore several neural network models, training them to act as patch classifiers, labeling each patch of an input image as either belonging to a tattoo or not. After detection, the tattoo area is de-identified by replacing it with the color of the surrounding skin.

## II. RELATED WORK

There have been relatively few works addressing the problem of tattoo detection in unconstrained images. Most research in de-identification deals with hard biometric features, with

an emphasis on the face [8], while a much smaller volume is devoted to soft and non-biometric features [19]. Research on tattoo detection is typically not motivated by de-identification, but by forensic applications, where the goal is to build content-based image retrieval system for tattoos that would help law enforcement. One such system is proposed by Jain et al. [11], where a cropped tattoo is segmented, represented using color, shape and texture features and matched to the database. Han and Jain [9] extend the idea of such a system by enabling sketch-to-image-matching, where the input image is a sketch rather than a photo, while the database contains real tattoo images. They use SIFT descriptors to model shape and appearance, and matching is performed using a local feature-based sparse representation classification scheme. Kim et al. [12] combine local shape context, SIFT descriptors and global tattoo shape for tattoo image retrieval, achieving robustness to partial shape distortions and invariance to translation, scale and rotation.

Heflin et al. [10] consider detecting scars, marks and tattoos “in the wild”, i.e. in unconstrained images that can contain a tattoo of an arbitrary scale anywhere. In their method, tattoo candidate regions are detected using graph-based visual saliency, and GrabCut [20], image filtering and the quasi-connected components technique [4] are employed to obtain the final estimate of the tattoo location.

Marčetić et al. [25] propose an experimental system for tattoo detection and de-identification. To detect the tattoos, the uncovered body parts are first detected based on the skin color model and filtered based on geometrical constraints. Then, the regions of interest are localized based on holes and cutouts in the skin regions. Finally, to confirm the presence of the tattoo, SIFT features are extracted from regions of interest and compared to the SIFT features stored in the tattoo database. The tattoos are then de-identified by replacing the tattoo regions with the patches obtained from the surrounding skin area.

Designing hand-crafted features that differentiate tattoos from background is very hard, given the degree of variability between tattoo designs and the fact that tattoos are often purposefully designed to resemble many real world objects [18]. Recently, convolutional neural networks (CNNs) were shown to be successful in automatically learning good features for many challenging classification tasks [13], [14]. Also, they have been successfully applied to problems of scene labelling [6] and semantic segmentation [17]. Building on this success,

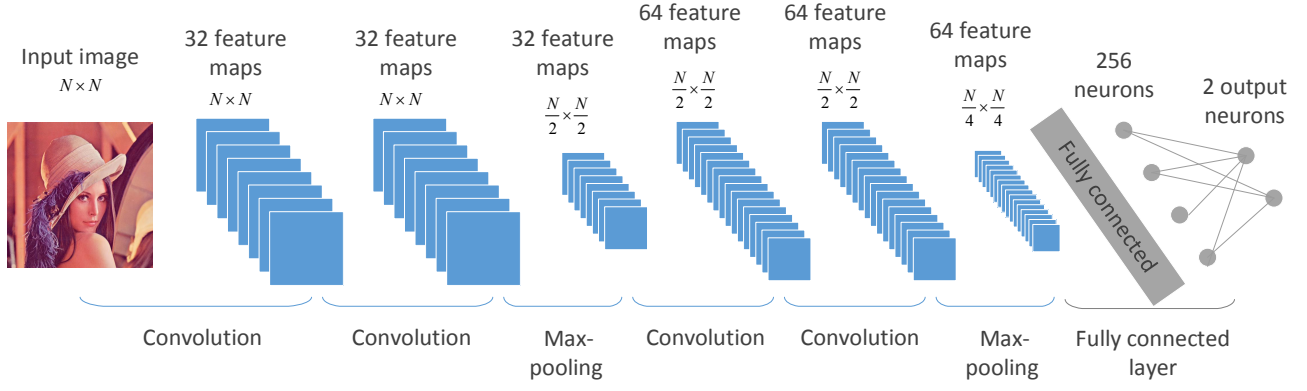


Fig. 1. The architecture of the ConvNet model inspired by the VGGNet that achieved the best results in tattoo patch labeling

in this work we use a deep convolutional neural network as a patch-based tattoo detector.

### III. DETECTING AND DE-IDENTIFYING THE TATTOO AREA

#### A. Tattoo detection

Our proposed method for tattoo detection uses a convolutional neural network for image patch labeling. By traversing the image using a small sliding window of the size  $N \times N$ , we obtain classifications of each window patch as either belonging to a tattoo or not. The final output of our method is a set of masked image regions that are candidate tattoo locations.

Since there is no exact rule on how to design a deep neural network for a particular problem, we consider several deep learning architectures that have recently proved successful in other classification tasks. In particular, we consider the following architectures:

- (i) An architecture consisting only of multiple fully connected layers with no convolutional layers at all, similar to one proposed by Ciresan et al. [28];
- (ii) An architecture inspired by the AlexNet of Krizhevsky et al. [13], the first part of which consists of several convolutional layers, with a max-pooling layer after each of them, followed by several fully connected layers;
- (iii) An architecture consisting of several pairs of convolutional layers of very regular structure, with max-pooling layers between each pair, followed by several fully connected layers (Fig. 1), inspired by the recently proposed VGGNet by Simonyan and Zisserman [23].

The input to the network is an  $N \times N$  color image patch (we assume the RGB color model). The patch 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.

#### B. Tattoo de-identification

The general idea of the proposed process of tattoo de-identification is as follows. First, we locate the tattoo(s) in the image (Fig. 2 (a)) by applying the trained convolutional neural network described in the previous subsection on different positions in the image in a sliding window manner. To increase

the speed of this step, we slide the window with the stride  $k$  (in the experiments, we set  $k = 8$ ). At each position we classify the center of the patch as either belonging to the tattoo area or not, according to the output of the network. If the result is affirmative, we label the pixels in the surrounding local square (of the same size as the stride) as belonging to the tattooed area. To remove the noise, we apply morphological opening, thus removing some of the false positives, followed by morphological closing, thus filling the small gaps inside the tattoo area. The described steps result in the binary mask roughly corresponding to the tattooed area (Fig. 2 (b)).

The de-identification of the tattoo area is performed by replacing the colors of the pixels belonging to the tattooed area with the color of the surrounding skin. The skin at different sides of the tattoo can be of different colors, due to various factors such as shadows, lighting conditions, the skin condition, etc. Therefore, simply replacing the whole tattoo area with the same color would not be appropriate, as it would not result in natural looking image. Instead, we calculate the new color of each pixel of the de-identified area by interpolating its value based on all the surrounding skin pixels. To find all surrounding skin pixels, we morphologically dilate the previously found binary mask corresponding to the tattooed area (Fig. 2 (c)) and find the contour of the resulting blob (Fig. 2 (d)). Since the tattoo can be at the edge of the skin (i.e. bordering with the background and not only with the skin), some pixels of the contour may not belong to the skin. We therefore classify each pixel as skin or as non-skin, based on the skin color model, as proposed by Jones and Regh [26]. Contour pixels classified as non-skin are removed from further consideration (Fig. 2 (e)). Each pixel inside the blob corresponding to the tattoo is then replaced with the color obtained by interpolating the color of the contour pixels classified as belonging to the skin, using inverse distance weighing interpolation. This means that the new pixel value is calculated as a weighted sum of skin contour values, where each contour point contributes with a weight inversely proportional to its distance to the pixel being replaced.

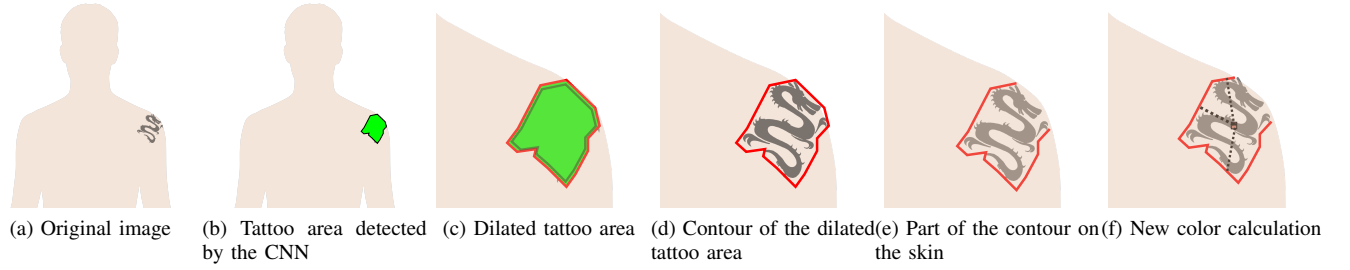


Fig. 2. The process of tattoo detection and de-identification

#### IV. EXPERIMENTS

The experiments were carried out on a publicly available dataset consisting of manually annotated tattoo images from the ImageNet database [21]. The dataset was assembled in the course of our earlier work on tattoo detection [27]

##### A. The employed dataset

Each of the images in the dataset contains one or more tattoos, and each tattoo is annotated using connected line segments. A few examples are shown Fig. 3. It can be seen that the outline of each tattoo is captured quite precisely, given the inherent limitations of a line-based approximation.



Fig. 3. Examples of annotated tattoo images.

##### B. Training the network

The training set for the convolutional neural network was assembled by randomly sampling patches from each tattoo image in the dataset and categorizing them as either positives or negatives depending on their position w.r.t. the annotation. Example patches are shown in Fig. 4.

The network was trained by optimizing the mean squared error loss function. We used stochastic gradient descent with momentum set to 0.9 and the mini-batch of 32. The learning rate was set to 0.1. We performed the training for 40 epochs at most, with early stopping if the validation loss has not improved for 3 epochs. The duration of the training depended on the size of the patches, ranging from 10 minutes for smallest patches to more than 17 hours for the largest.

We implemented the described network architectures in Python, using Theano [2], [3] and Keras [5] libraries.

TABLE I

EVALUATION OF THE NETWORK PERFORMANCE ON DIFFERENT PATCH SIZES

Patch size	$8 \times 8$	$16 \times 16$	$24 \times 24$	$32 \times 32$	$40 \times 40$
False negatives	691	641	887	360	404
False positives	595	459	258	760	673
Accuracy	0.7953	0.8207	0.8134	0.8174	0.8244

##### C. Detection performance evaluation

In total, we extracted 36820 patch images, out of which 18400 positive and 18420 negative examples. The patches were divided into training set (containing 27616 examples: 13800 positive and 13815 negative), validation set (3068 examples: 1533 positive and 1535 negative) and testing set (6136 examples: 3077 positive and 3070 negative). Patches belonging to the same image were all added to the same set.

The network was trained using varying patch sizes ( $8 \times 8$ ,  $16 \times 16$ ,  $24 \times 24$ ,  $32 \times 32$  and  $40 \times 40$ ) to determine the optimal patch size. While the larger patches are expected to provide more information about context, the network that utilizes them is slower to train and to test.

Out of the three tested architectures (see section III.A), the architectures (i) and (ii) achieved lower performances, with accuracies ranging from 61.9% to 77.3%. The best results were achieved with the architecture (iii) (shown in Fig. 1.), which achieved accuracy of cca 80% – 83%. The results obtained by this third architecture for different patch sizes are summarized in Table I. As we can see, the results improve in terms of accuracy with the increase in image patch size, however, the difference in accuracy is not very pronounced; i.e. we can say that results for all the patch sizes are similar.

##### D. Blob and contour detection

The results of the blob and contour detection as a preparation for de-identification were evaluated qualitatively. Several successful (first three columns) and unsuccessful (last column) results can be seen in Fig. 5.

As we can see, most of the tattoo area was successfully found by the trained CNN, but some gaps remained inside the tattoo area, and some false positives were found outside of it, both of which were successfully eliminated by applying mathematical morphology. The results obtained in such a way can be used as described in section III.B. to de-identify the tattooed image region.

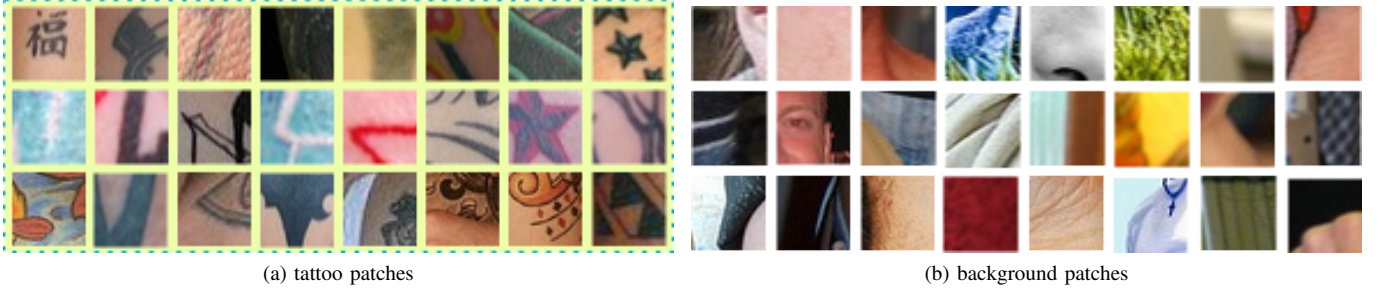


Fig. 4. Example extracted patches from our dataset (patch size  $32 \times 32$ ).

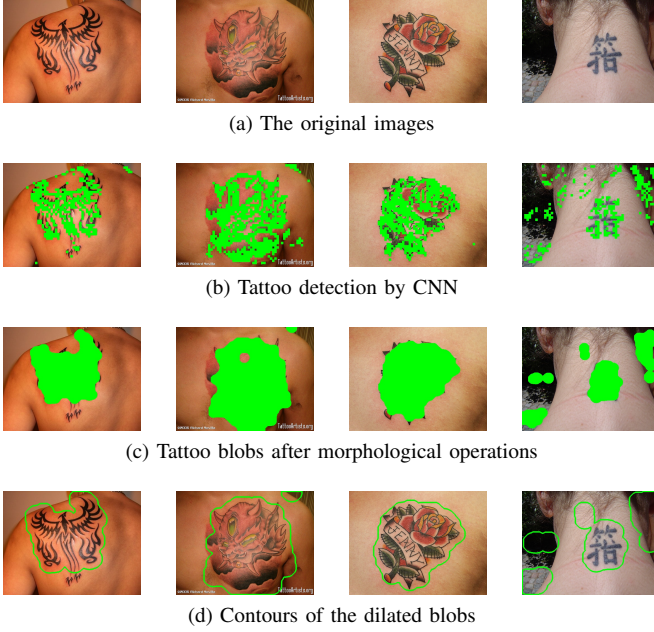


Fig. 5. Results of the tattoo localization.

An example of such de-identification is shown in Fig. 6. For the most part, the tattoo is removed, although some problem remain due to the fact that some of the bordering parts of the tattoo were not detected.



Fig. 6. Sample de-identification result.

## V. CONCLUSION AND OUTLOOK

We proposed a method for finding and de-identifying tattooed skin regions. We used a deep convolutional neural network to label image patches as either belonging to a tattoo or not. The tattoo regions found in such a way were de-

identified by replacing their color with the value obtained by interpolation from the surrounding skin.

Our findings indicate that the proposed approach can be used to detect and de-identify candidate tattoo regions in an image. The most critical part of the process is tattoo detection. We estimate that the deep learning approach with convolutional layers has good potential to learn to detect tattooed areas; however, the problem of false positives and false negatives is still visible in the experiments we conducted. The model often either learns to discriminate between skin color and all the rest, resulting in a number of false positives in the background, or learns to count as tattoo patches only the patches bordering with skin color, resulting in false negatives inside the homogenous tattoo areas. We speculate that the problem could be addressed by learning the model with larger patches (e.g.  $64 \times 64$ ,  $128 \times 128$ , or even  $256 \times 256$ ) and significantly enlarging the training set size by adding small modifications (translations, rotations, noise, etc) to the existing dataset, as suggested in many works on deep neural networks. We are planning to investigate these possibilities in the future.

Another possible improvement would consist in combining this method with other stages of a de-identification pipeline, i.e. pedestrian detection and segmentation, 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. Finally, to improve the naturalness of the de-identified regions, the texture of the surrounding skin could also be taken into account along the color.

## ACKNOWLEDGMENT

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), and by the COST Action IC1206 De-Identification for Privacy Protection in Multimedia Content. This support is gratefully acknowledged.

The authors would like to thank prof. Z. Kalafatić for his help on skin detection and interpolation, as well as to UNIZG-FER students D. Bratulić, N. Mrzljak and J. Silović who helped to collect and annotate the dataset and conducted preliminary experiments.



## 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. 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. Oral Presentation.
- [4] T. E. Boulton, 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] F. Chollet. Keras. <https://github.com/fchollet/keras>, 2015.
- [6] 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.
- [7] 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.
- [8] R. Gross, L. Sweeney, J. F. Cohn, F. De la Torre, and then Baker. *Protecting Privacy in Video Surveillance*, chapter Face De-identification, pages 129–146. Springer Publishing Company, Incorporated, 2009.
- [9] 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.
- [10] B. Heflin, W. Scheirer, and T. E. Boulton. 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.
- [11] 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.
- [12] 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.
- [13] 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.
- [14] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553):436–444, 05 2015.
- [15] 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.
- [16] 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.
- [17] 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.
- [18] 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.
- [19] 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.
- [20] C. Rother, V. Kolmogorov, and A. Blake. "grabcut": Interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph.*, 23(3):309–314, Aug. 2004.
- [21] 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.
- [22] W. Scheirer, A. Rocha, R. Micheals, and T. Boulton. *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.
- [23] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2014.
- [24] M. J. Wilber, E. Rudd, B. Heflin, Y.-M. Lui, and T. E. Boulton. Exemplar codes for facial attributes and tattoo recognition. In *Applications of Computer Vision (WACV), 2014 IEEE Winter Conference on*, pages 205–212, March 2014.
- [25] D. Marcetic, S. Ribaric, V. Struc and N. Pavesic. An Experimental Tattoo De-identification System for Privacy Protection in Still Images. In *Special Session on Biometrics, Forensics, De-identification and Privacy Protection, MIPRO 2014*, pages 69–74, 2014.
- [26] M. J. Jones and J. M. Rehg. Statistical color models with application to skin detection. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 123–129, 1999.
- [27] T. Hrkac and K. Brkic. Tattoo Detection for Soft Biometric De-Identification Based on Convolutional Neural Networks. In *OAGM 2016*, to appear.
- [28] D. C. Ciresan, U. Meier, L. M. Gambardella, J. Schmidhuber. Deep Big Simple Neural Nets Excel on Handwritten Digit Recognition. In *Neural Computation*, vol. 22, n0 12, pages 3207–3220, 2010.
- [29] 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.