A Tattoo Location Problem Approach Using Mask R-CNN Network

Rodrigo Tchalski da Silva
Graduate Program on Electrical Eng. and Informatics
Federal University of Technology - Paraná (UTFPR)
Curitiba, Brazil
rodrigo.tds@gmail.com

Heitor Silvério Lopes
Graduate Program on Electrical Eng. and Informatics
Federal University of Technology - Paraná (UTFPR)
Curitiba, Brazil
hslopes@utfpr.edu.br

***Abstract*—Tattoos are still poorly explored as a biometric factor for human identification, especially in law enforcement, where they can play an important role in identifying criminals, victims, or other persons of interest. Tattoos are classified as soft biometrics as they are not permanent and can change over time, unlike hard biometric traits (fingerprint, iris, DNA, etc.). In this way, the main objective of this work is to apply an approach to the Mask R-CNN network with a tattoo dataset and to fine-tune the set of parameters that presented the best results in training the network. In tattoo location, the results reached an mAP of 0.893, which shows that the Mask R-CNN network has great adaptability to the tattoo environment, in addition to performing a qualitative analysis that helped to understand how the characteristics of images and annotations influence the results. We presented two new datasets for tattoo location, composed of 5,754 new annotated images. Future work will include improving the quality and volume of the databases, conducting a more in-depth study on the fine-tuning of network parameters, and developing models for other problems that make up the tattoo recognition roadmap.**

***Index Terms*—Tattoo location, Mask R-CNN, Computer vision, Tattoo dataset**

I. Introduction

Tattoos constitute a widespread cultural phenomenon and are recognized both as a means of individual expression and as a legitimate form of artistic representation. Their historical presence extends over five millennia, during which they have served various social, cultural, and personal functions [1]. Tattoos can convey a diverse array of information about the individuals who carry them, including aspects of personal history, psychological profile, group identity, and social affiliations [2]. Due to their often unique and highly individualized characteristics, tattoos can carry information that extends beyond their original motivational context. Consequently, they can be classified as biometric markers, offering potential utility in forensic and identity verification applications.

The term biometrics is described as the science of recognizing an individual based on their physical or behavioral characteristics [3]. It is also described as the automated use of physiological or behavioral characteristics to determine or verify an individual's identity [4]. Currently, biometrics are useful for many applications, such as cell phone unblocking, bank transaction authentication, access control, and especially public security-related applications, including people monitoring, criminal identification, identity certification, and border control [5].

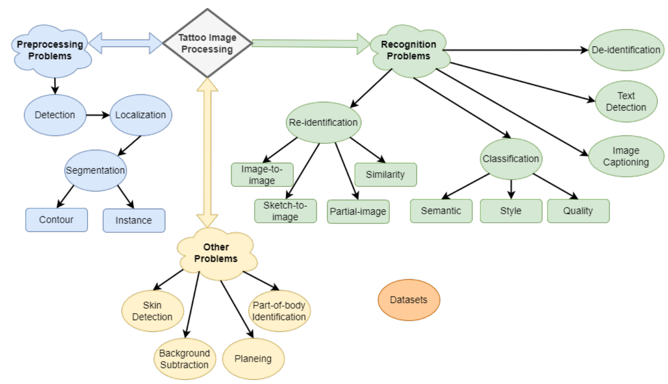

Fig. 1. The tattoo recognition roadmap.

Tattoo recognition encompasses a broad spectrum of challenges that are generally categorized into two primary domains: preprocessing and recognition (Fig. 1). The challenges of pre-processing include the detection, localization, and segmentation of tattoos within visual data. Recognition-related challenges involve re-identification, classification, deidentification, text detection, and image caption. Although numerous methodologies have been proposed to address these tasks, the field remains an active area of research. To date, no approach has yielded results sufficiently robust or consistent to be regarded as a comprehensive solution to these problems.

Among the problems related to the tattoo roadmap, tattoo location is of special importance, as it aims to locate where the tattoos are in an image, insert a bounding box around de tattoo area, and remove the important information (the tattoo) from the rest of the image. In other words, tattoo location methods are responsible for removing the image background, leaving only the tattoo areas to be used in the next steps of the tattoo recognition roadmap (Fig. 1).

Possibly because of this complexity, to date, only a few studies have been published addressing this issue, even though it has great practical applicability. In addition, the studies published have been based on datasets that are frequently unavailable.

TABLE I
PAPERS APPROACHING THE TATTOO LOCATION PROBLEM.

Reference	Method	Best Result	Dataset Size
[8]	Decision rules in RGB color space + morphological operations	4.9% False Positive Rate	204
[9]	not shown	97% acc.	16,716
[7]	Center-surround feature location	66.20% acc.@Tatt-C	4,308
[7]	Graph-cut	70.46% acc. @Tatt-C	4,308
[6]	Faster R-CNN	98.25% acc.	23,802
[10]	Faster R-CNN	Tatt-C: 61:7% recall@0.1FPPI WebTattoo: 87:1% recall@0.1FPPI	8,026

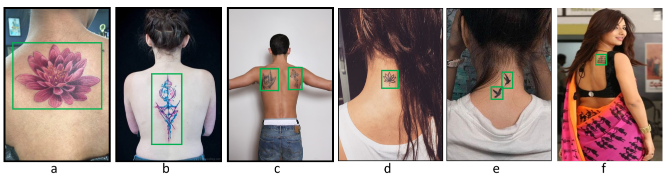

Fig. 2. Examples of tattoo location.

Thus, the main objective of this study is to present a model based on Mask R-CNN (Sec. IV-A) network applied to the tattoo location problem, which has practical applications both in the field of public security, to help identify people, and to help locate tattoos for victim recognition.. Throughout the work, the need for a high-quality dataset for training the models emerged. Therefore, two new, high-quality and annotated databases were created, consisting of 5,754 tattoo instances, called TattLocA and TattLocB. We also used specific data augmentation procedures to improve the robustness of the trained classifier. The datasets are available on the 'link to be published'.

II. Literature Review

The tattoo location problem consists of returning, given an image with one or more tattoos, a group of coordinates containing a position (top and left coordinates) and a size (width and height) representing the bounding box for each tattoo found.

In other words, the challenge is to locate where one or more tattoos are found in the original image and to enclose them in rectangular bounding boxes as shown in Fig. 2.

Usually, the bounding box created around a tattoo is defined by a set of coordinates (x, y, w, h), such that x and y represent the initial coordinates of the bounding box, and w and h represent the width and height of the frame, respectively [6]. The objective is to find, for each disjoint tattoo in the image, a frame with minimal dimensions that is capable of enclosing the entire tattoo [7].

Table I shows a summary of the papers found in the literature, including methods and results for each study. Although the first publications did not present numerical results, the most recent studies made important contributions to the problem of tattoo location using machine learning methods.

The first study was presented by [8], who proposed a two-step approach to locate tattoos on an image. First, the skin is detected by applying decision rules in the RGB color space, followed by geometric restrictions to eliminate skin-like color regions that do not belong to body parts. Next, potential tattoo regions are located in the cropped regions with a different skin color, obtained by the morphological closure operation.

The NIST [11] then proposed a group of challenges in tattoo recognition, including the tattoo location problem. They also presented the Tatt-C dataset, which was widely used in tattoo recognition research from then on.

In [9], where the results for the NIST challenges were presented, the best accuracy of 97% was achieved by the MorphoTrek company. However, algorithms, data used, and how the accuracy was measured were not made available for further validation tests.

Later, [7] presented a specific study on tattoo location using two specific methods. The first is a center-surround feature location method. The second method used a graph-cut segmentation based on the edge of the image, a skin-color model, and a visual bump map. To evaluate the results, the images were previously manually cropped. The second method achieved better accuracy than the first one, reaching 70.46% and 69.91% with the Tatt-C and Evil Tattoo datasets, respectively.

[6] customized a Faster R-CNN, by fine-tuning the VGG_CNN_M_1024 network. The model was trained with Tatt-C and PASCAL Visual Object Classes (VOC) datasets with manually annotated bounding boxes around the tattoos. A validation step was performed using the NTU dataset with 10,000 images. The best accuracy was 98.25%, which was marginally higher (1.25%) than the results presented by [9], which was used as a basis for comparison.

Another work by [10] also used the Faster R-CNN model and achieved 99% precision when the method was applied to the Tatt-C dataset, in the same joint training process presented in their tattoo detection model.

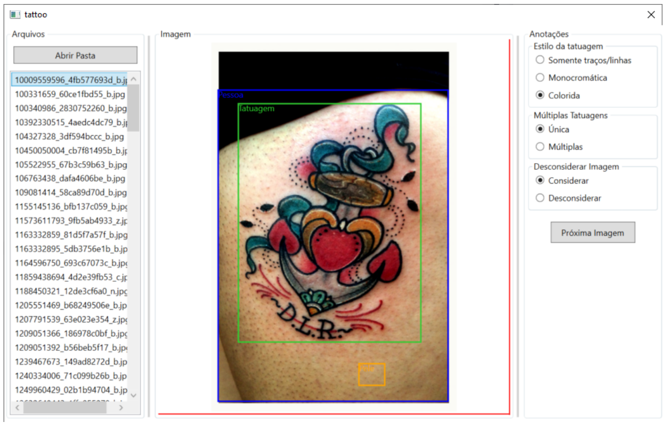

Fig. 3. Tattoo annotation system.

III. DATASETS

The datasets used for tattoo location methods require special attention, as each image must be individually annotated with a specific bounding box information, that is, the initial position (x, y) of the bounding box and its width and height (w, h).

For this study, two datasets were created, one complementary to the other, containing 1,683 and 4,071 images, respectively, called TattLocA and TattLocB. Data were obtained from the Bing and Flickr websites by web scraping, executed in Python. To obtain the data, several terms were used in order to obtain the largest possible variety of images in terms of size, style, and colors.

Each image was manually annotated using a specially developed system for this purpose, as shown in Figure 3. Annotations include the following information:

- Tattoo bounding box: coordinates and tattoo size.
- Person area: approximated area used by the person in the image, used to calculate the proportions of the person and the tattoo in the image.
- Tattoo style: indicates whether the tattoo style is color, monochromatic, or outline.
- Multiple tattoos: indicates the existence of one or more tattoos in the image.
- Consider the image: indicates whether the image should be used or discarded. In general, very confusing or irregular images were not taken into account.

Subsequently, the tattoos were classified in terms of style, color, and proportion in the image (Figure 4). Color tattoos are those that use more than one color to be filled in. Outline tattoos are those unfilled, drawn only with lines. Monochromatic tattoos are those that are filled with one single color (usually black). There are also the text tattoos, represented by those with only text, in general, a particular case of outline tattoos.

In terms of size, tattoos were classified as small, medium, and large. Small are those that occupy less than 1/3 of the total image area, medium are those that occupy more than 1/3 and less than 2/3 of the total image area, and large are those that occupy more than 2/3 of the total image area.

Fig. 4. Sample of different tattoo styles and proportion. Image "a" represents a color tattoo, image "b" represents an outline tattoo, image "c" represents a monochromatic tattoo, image "d" represents a large size tattoo, image "e" represents a medium size tattoo, image "f" represents a small size tattoo and image "g" represents a textual tattoo.

According to what was previously defined, the specifications of the TattLocA and TattLocB datasets are shown in Table II and Table III, respectively.

The goal was to create datasets that considered not only the number of images but also more robust in terms of size, color, and style.

IV. EXPERIMENTS AND RESULTS

The experiments in tattoo location were carried out using scripts in the Python language, running in a Linux Ubuntu environment with Intel(R) Core(TM) i7-9700K CPU @ 3.60GHz processor, 64 GB of memory and NVIDIA TITAN Xp GPU with 12Gb of memory.

The process was based on two main experiments:

1) Evaluation of Mask R-CNN with an initial dataset to verify the adherence of the network to the tattoo location problem.
2) Application of Mask R-CNN to a large dataset to evaluate the convergence capability and fine-tune the model to improve results.

A. The Mask R-CNN Method

Currently, deep learning are the most used approaches for this problem, with the state-of-the-art in this area based on the evolution of R-CNN networks, there are R-CNN [12], Fast R-CNN [13], Faster R-CNN + RPN [14] and Mask R-CNN [15], and YOLO [16] networks.

[15] presented the Mask R-CNN, "a conceptually simple, flexible, and general framework for object instance segmentation". This network detects objects in an image while simultaneously generating a high-quality segmentation mask for each instance. In addition, the network "extends Faster R-CNN by adding a branch to predict an object mask in parallel with the existing branch to recognize the boundary box" (Fig. 5).

Therefore, the method used for tattoo location consists in using Mask R-CNN with its original weights and then fine-tuning the network hyperparameters to localize tattoos in a given image (Fig. 6).

B. Evaluation of the Mask R-CNN for Tattoo Location

The first experiment aimed to answer the question "Is the Mask R-CNN network suitable for tattoo location?". In other words, this experiment was performed to evaluate whether, even using a small dataset, Mask R-CNN can achieve promising results for the tattoo location problem.

TABLE II
DATASET FOR TATTOO LOCATION - TATTLOCA.

	Color			Monochromatic			Outline		
	Train	**Validation**	**Total**	**Train**	**Validation**	**Total**	**Train**	**Validation**	**Total**
Small	201	50	251	202	50	252	120	30	150
Medium	352	88	440	224	56	280	51	13	64
Large	123	31	154	62	16	78	11	3	14
Total	**676**	**169**	**845**	**488**	**122**	**610**	**182**	**46**	**228**

TABLE III
DATASET FOR TATTOO LOCATION - TATTLOCB.

	Color			Monochromatic			Outline		
	Train	**Validation**	**Total**	**Train**	**Validation**	**Total**	**Train**	**Validation**	**Total**
Small	490	122	612	487	122	609	458	115	573
Medium	650	162	812	446	112	558	308	77	385
Large	222	55	277	100	25	125	96	24	120
Total	**862**	**216**	**1,078**	**1,034**	**258**	**1,292**	**1,361**	**340**	**1,701**

For this evaluation, we applied the model presented in Section IV-A to the TattLocA dataset, divided into 1,346 images for training and 337 images for validation. The model was trained with 100 epochs and kept the hyperparameters fixed with the default values. As a result, Figure 7 shows the evolution of the loss during training.

The objective of this first experiment was to verify if the Mask R-CNN network would reach an acceptable initial convergence, providing some perspective of its robustness and adaptability to tattoo data.

Observing Figure 7, although the loss curve for the validation dataset shows an average stationarity and did not converge with the training curve, its average value of 13.44% can be considered a good starting point. Therefore, even though the loss curve does not show real convergence, its initial level did not lead to an overfitting nor was it extremely high, which was expected to classify it for the next experiments.

Finally, with this initial result it is possible to continue with the next experiments, aiming at achieving even better results for tattoo location. In other words, it is possible to consider that this initial experiment showed that Mask R-CNN has great potential to be used in the tattoo location problem.

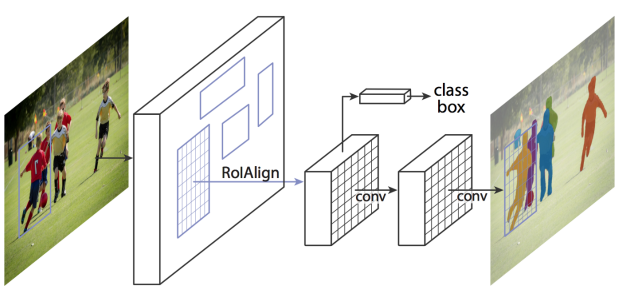

Fig. 5. Mask R-CNN framework.

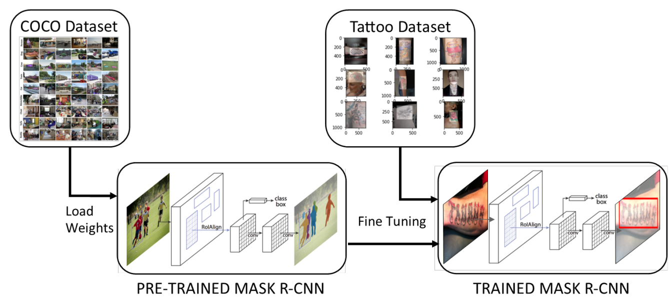

Fig. 6. Mask R-CNN Architecture for Tattoo Location.

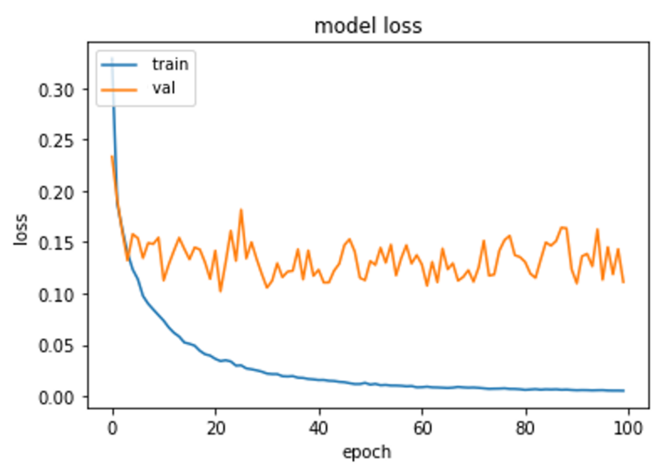

Fig. 7. Loss to tattoo dataset with default Mask R-CNN parameters.

C. Evaluation of Mask R-CNN and Fine-Tuning

Given that the initial experiment showed promising results, the following experiment aimed to answer the question: “Does increasing the training dataset and fine-tuning the hyper-parameters of the model lead to improve the results in terms of accuracy and generalization to the tattoo location problem?”

In this group of experiments, the TattLocB dataset was used. This dataset is composed of 4,071 images organized in terms of tattoo size, fill color, and style.

The main difficulty in using Mask R-CNN network for tattoo location was the identification and adjustment of its more than 40 hyperparameters used both in location and segmentation tasks, and determine which of them have the most positive impact on training the model, including, for instance, learning rate, learning momentum, and weight decay, and parameters related to training ROI per image, ROI positive ratio, and vectors of RPN anchor scales, backbone strides, among others. Since it is not possible to previously know which set of parameters is dependent or independent, a series of tests was randomly performed until a set of satisfactory hyperparameters was established.

After these experiments, the best set of hyperparameters found is shown in Table IV. Using such parameters, we obtained a loss average of 0.1135, as shown in Figure 8. Furthermore, the model reached an mAP (mean Average Precision) of 0.893 with a dVAP (standard deviation Average Precision) of 0.266 in the training dataset and an mAP of 0.761 with a dVAP of 0.223 in the validation dataset (Table V).

TABLE IV
HYPER-PARAMETERS TESTED FOR MASK R-CNN TATTOO LOCATION.

Hyper-Parameter	Value
Epochs	50
Validation Steps	407
Backbone Strides	4, 8, 16, 32, 64, 96, 128
RPN Anchor Scale	32, 64, 128, 256, 512, 768, 1024
Weight Decay	0.5
Train BN	True
ROI Positive Ration	0.7
Train ROIs per Image	512
Detection NMS Threshold	0.7
Image Min Dim	128
Max GT Instances	120
Detection Max Intances	120

TABLE V
FINAL RESULTS FOR MASK R-CNN TATTOO LOCATION.

Evaluation	Train dataset	Validation dataset
Minimum loss	0.034	0.092
Average loss	0.067	0.113
mAP	0.893	0.761
dVAP	0.266	0.223

Due to the number of hyperparameters and the high computational cost of training the model, an automated factorial

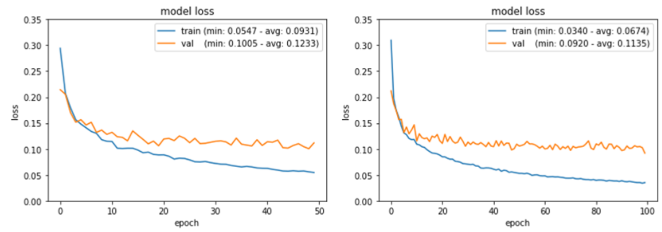

Fig. 8. Loss to tattoo dataset with fine tuning Mask R-CNN.

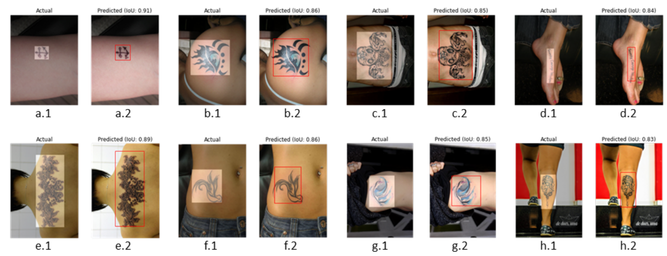

Fig. 9. Sample of good location bounding boxes.

experiment was not possible. However, the results obtained showed that Mask R-CNN is quite adequate for the tattoo location problem.

Although these results cannot be directly compared with the previously presented results in the literature (see Table I) because they used different datasets and quality metrics, the results presented here are very promising, considering both the model and the dataset.

In other words, the process of increasing the datasets, diversifying images in terms of sizes and styles, and performing an adequate fine-tuning in the model parameters brought more comprehensive results with better assertiveness and generalization for the tattoo location problem as a whole.

D. Qualitative Analysis

Finally, a qualitative analysis was carried out with the objective of finding more evidence on the characteristics of the images that were easily located and those that were more difficult to locate. Therefore, it aims to answer the question “Is it possible to identify characteristics in the tattoos that were the best and worst located?”.

Figure 9 presents a collection of images that obtained high IoU (Intersection over Union) values using the model validation dataset. In these images, it was possible to observe that the tattoos had a prominent position in the image and were clearly located within the person’s skin region. In addition, the tattoo images were not too large in terms of the total area of the image and did not have much pollution around them. These characteristics may have favored better results for tattoo location and point to a well-behaved image model for the tattoo location scenario. Regarding tattoo style, color, and size, none of them presented a disadvantage in this scenario, as long as they were well placed and with less noise around them.

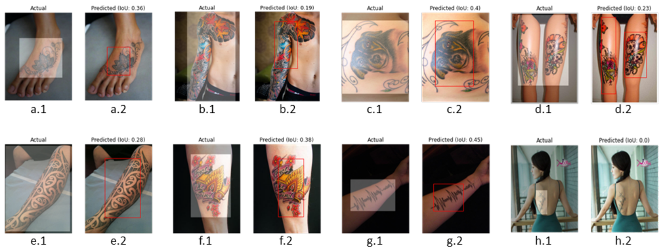

Fig. 10. Sample of bad location bounding boxes.

On the other hand, Figure 10 presents some examples of tattoos that obtained low IoU values for the bounding box generated by the trained model. These images, in general, consisted of tattoos that occupied a large region of the image (see "b.1", for instance), spread across the contour region of the skin with the background of the image ("a.1", "b.1", "e.1", "f.1"), were in a diagonal position in relation to the image orientation ("a.1", "e.1", "g.1"), had poor lighting ("g.1"), had less prominent information around the main image, as in example "c.2", in which the center of the tattoo is well-defined, but there is a continuation of detail around the entire image.

In addition, image "d.1" in Figure 10 presents an example of a wrong annotation in the original dataset: there were two tattoos, but the annotation was made with only one bounding box around both. In this case, only one of the tattoos was found and the IoU calculation was impaired. The tattoo exemplified in "h.1", presents an example where the tattoo was not located. Probably, this is because it is an example where the colors of the tattoo were confused with the colors of the clothing, and, in addition, it is a quite small tattoo, and the network was unable to locate it despite the correct annotation.

Although the model did not perform well for images with specific features, it performed well in the general scope, and these unsuccessful examples can help to infer which characteristics can be considered for further refinements of the model.

V. Conclusions

The present study aimed to develop computer vision-based methods for the tattoo location problem.

For this, we used the Mask R-CNN network with a fine-tunning in their hiperparameters. The mAP values achieved were 0.893 and 0.761 for the training and validation bases, which is quite significant for this problem.

Two datasets were created for this study, TattLocA and TattLocB, so the latter is much larger than the former. The experiments show the importance of the size and diversity of the dataset used to train the model, since better results were obtained with TattLocB.

Similarly to the tattoo detection problem, the lack of publicly available datasets precluded comparisons with other published results. In addition, published works in this area used different quality metrics as used in this work.

The main difficulty encountered in using the Mask R-CNN network was the performance of fine-tuning, given the large amount of existing hyperparameters, and given the high computational cost taken for each test. Therefore, no factorial test was performed.

In future work, the creation of larger and more diversified datasets could contribute to the development of the model robustness, as well as to a study with data augmentation. A more in-depth study of the hyperparameters of the network would also be helpful to achieve even better results. The use of automated methods for hyperparameters tuning can also improve the results.

Acknowledgment

The authors thank NVIDIA Corp. for the donation of the Titan-Xp GPUs used in the experiments.

References

[1] A. K. Jain, J.-E. Lee, and R. Jin, "Tattoo-ID: Automatic tattoo image retrieval for suspect and victim identification," in *Advances in Multimedia Information Processing*. Heidelberg: Springer, 2007, pp. 256–265.

[2] M. M. Rohith, W. R. Belcher, J. Roy, S. O. Abraham, P. Chakraborty, N. J. Nandaniya, and A. Johnson, "Tattoo in forensic science: An indian perspective," *Journal of Forensic and Legal Medicine*, vol. 74, p. 102022, 2020.

[3] A. Jain, A. Ross, and S. Pankanti, "Biometrics: A tool for information security," *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 2, pp. 125–143, 2006.

[4] A. Sharma, "Biometric system - a review," *International Journal of Computer Science and Information Technologies*, vol. 6, no. 5, pp. 4616–4619, 2015.

[5] R. D. Labati, A. Genovese, E. Muñoz, V. Piuri, F. Scotti, and G. Sforza, "Biometric recognition in automated border control," *ACM Computing Surveys*, vol. 49, no. 2, pp. 1–39, 2016.

[6] Z. H. Sun, J. Baumes, P. Tunison, M. Turek, and A. Hoogs, "Tattoo detection and localization using region-based deep learning," in *Proceedings of the 23rd International Conference on Pattern Recognition (ICPR)*. Piscataway, NJ, USA: IEEE Press, 2016, pp. 3055–3060.

[7] J. Kim, H. Li, J. Yue, J. Ribera, E. J. Delp, and L. Huffman, "Automatic and manual tattoo localization," in *Proc. IEEE Symposium on Technologies for Homeland Security (HST)*, 2016, pp. 1–6.

[8] D. Marcetić, S. Ribarić, V. Struć, and N. Pavesić, "An experimental tattoo de-identification system for privacy protection in still images," in *Proc. 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*. IEEE, 2014, pp. 1288–1293.

[9] M. Ngan, G. Quinn, and P. Grother, "Tattoo recognition technology - challenge (Tatt-C) - outcomes and recommendations," Gaithersburg, MD, USA, Technical Report NISTIR 8078, 2016.

[10] H. Han, J. Li, A. K. Jain, S. Shan, and X. Chen, "Tattoo image search at scale: Joint detection and compact representation learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 10, pp. 2333–2348, 2019.

[11] M. Ngan and P. Grother, "Tattoo recognition technology - challenge (Tatt-C): an open tattoo database for developing tattoo recognition research," in *Proceedings of the IEEE International Conference on Identity, Security and Behavior Analysis (ISBA 2015)*, 2015, pp. 1–6.

[12] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, Jun. 2014.

[13] R. Girshick, "Fast r-CNN," in *2015 IEEE International Conference on Computer Vision (ICCV)*. IEEE, Dec. 2015.

[14] 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 and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[15] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask r-CNN," in *2017 IEEE International Conference on Computer Vision (ICCV)*. IEEE, Oct. 2017.
[16] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," 2015.