

Article

Approach for Tattoo Detection and Identification Based on YOLOv5 and Similarity Distance

Gabija Pocevičė¹, Pavel Stefanovič^{1,*}, Simona Ramanauskaitė² and Ernest Pavlov³

¹ Department of Information Systems, Vilnius Gediminas Technical University, Saulėtekio al. 11, LT-10223 Vilnius, Lithuania

² Department of Information Technology, Vilnius Gediminas Technical University, Saulėtekio al. 11, LT-10223 Vilnius, Lithuania; simona.ramanauskaitė@vilniustech.lt

³ Department of Electronic Systems, Vilnius Gediminas Technical University, Saulėtekio al. 11, LT-10223 Vilnius, Lithuania

* Correspondence: pavel.stefanovic@vilniustech.lt

Abstract: The large number of images in the different areas and the possibilities of technologies lead to various solutions in automatization using image data. In this paper, tattoo detection and identification were analyzed. The combination of YOLOv5 object detection methods and similarity measures was investigated. During the experimental research, various parameters have been investigated to determine the best combination of parameters for tattoo detection. In this case, the influence of data augmentation parameters, the size of the YOLOv5 models (n, s, m, l, x), and the three main hyperparameters of YOLOv5 were analyzed. Also, the efficiency of the most popular similarity distances cosine and Euclidean was analyzed in the tattoo identification process with the purpose of matching the detected tattoo with the person's tattoo in the database. Experiments have been performed using the deMSI dataset, where images were manually labeled to be suitable for use by the YOLOv5 algorithm. To validate the results obtained, the newly collected tattoo dataset was used. The results have shown that the highest average accuracy of all tattoo detection experiments has been obtained using the YOLOv5l model, where mAP@0.5:0.95 is equal to 0.60, and mAP@0.5 is equal to 0.79. The accuracy for tattoo identification reaches 0.98, and the F-score is up to 0.52 when the highest cosine similarity tattoo is associated. Meanwhile, to ensure that no suspects will be missed, the cosine similarity threshold value of 0.15 should be applied. Then, photos with higher similarity scores should be analyzed only. This would lead to a 1.0 recall and would reduce the manual tattoo comparison by 20%.

Keywords: YOLOv5; tattoo detection; data augmentation; hyperparameters; similarity distance; ResNet50



Citation: Pocevičė, G.; Stefanovič, P.; Ramanauskaitė, S.; Pavlov, E. Approach for Tattoo Detection and Identification Based on YOLOv5 and Similarity Distance. *Appl. Sci.* **2024**, *14*, 5576. <https://doi.org/10.3390/app14135576>

Academic Editors: Douglas O'Shaughnessy and Roberto Romaniello

Received: 20 May 2024

Revised: 11 June 2024

Accepted: 26 June 2024

Published: 26 June 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The rapid growth of digital images that could be retrieved from different sources has created a strong demand for efficient object localization and detection methods in various fields, such as medicine, military, manufacturing, law enforcement, and forensics. There is much research in the scientific literature that focuses on the detection of different diseases by analyzing images using various object detection or localization algorithms. For example, the research by Jucevičius et al. [1] and Verbukaitė et al. [2] focuses on the analysis of the image of medicine in which the images of glaucoma and prostate cancer images have been analyzed. In the research of Gupta et al. [3], the image data obtained from an uncrewed aerial vehicle were analyzed. The main aim of the research was to detect military vehicles. The Industry 4.0 context has led to many intelligences or automated solutions based on object detection as well. The research by Usamentiaga et al. [4] used deep learning algorithms to detect product defects in manufacturing. The other research by Li et al. [5] analyzed metal surface defects using object detection methods, where the

results are useful and can be applied to improve manufacturing lines. Object detection and localization tasks are very popular and could be applied in different solutions, for example, for the detection of construction details [6,7] or as a component of travel direction recommendation systems [8].

In law enforcement and forensics, different types of hard biometric data can be used to identify people in the image, such as the iris pattern, facial features, or fingerprint. Despite significant advances in hard biometric identification, a single biometric characteristic cannot guarantee the desired identification accuracy. Characteristics of people who are less unique compared to traditional biometric data are called soft biometric data. As one of the soft biometric features, tattoos are valuable in helping people identify associations, groups, members, gangs, criminals, or victims. Tattoos are considered soft biometrics because, over time, the tattoo on the human body may change, compared to hard biometric characteristics such as fingerprints or iris [9]. However, the accuracy of automatic tattoo identification and detection is challenged by a wide range of artistic compositions, colors, shapes, textures, image conditions, and quality [10]. Therefore, it is more difficult to choose the right model to solve this problem. The concept of tattoo detection and identification using deep learning involves the use of machine learning techniques, specifically deep learning, to identify and locate tattoos on the human body, as well as object detection methods. This is a relatively new area of research, as tattoos have traditionally been difficult to analyze and classify using automated techniques due to their complex and highly variable visual appearance. In the context of tattoo detection and recognition, deep learning techniques can be used to analyze images of the human body and identify the presence and location of tattoos. This can be useful in a variety of applications, including law enforcement, forensic analysis, and medical research. In general, the use of deep learning for tattoo detection and recognition represents a significant advance in the field of machine learning and has the potential to revolutionize the way tattoos are analyzed and understood.

The main aim of this paper is to detect tattoos on a person's body and then link them with the data available in the database to identify to whom it belongs. An experimental investigation has been performed to find out the influence of various hyperparameters of YOLOv5, data augmentation, and similarity distances on tattoo detection and identification. The main contributions of the paper are as follows:

- (1) A total of 135 models have been trained to detect which model of YOLOv5 (n, s, m, l, x) allows obtaining the highest results in tattoo detection. The different combinations of hyperparameters, such as learning rate, momentum, and decay weight, were investigated;
- (2) The influence of the data enhancement parameters on the final results of tattoo detection has been investigated. There is a lack of this kind of research, especially in the context of the tattoo dataset;
- (3) The efficiency of the YOLOv5 algorithm and similarity distances combination have been experimentally investigated to detect tattoos on the person's body and link them to the database of tattoos.

The results of this research may be useful for law enforcement and the field of forensics, as well as for other researchers who focus on object detection tasks. During the research, a large number of parameter combinations were used and five different size YOLOv5 models (n, s, m, l, x) were thoroughly investigated.

The remainder of the paper is organized as follows. In Section 2, related works are reviewed. In Section 3, the background of the experimental investigation is presented. The YOLOv5 algorithm was introduced for tattoo detection. Also, similarity distances have been described that have been used for the person identification process. Section 4 describes the main steps of the tattoo detection and identification process. The experimental investigation of the data augmentation and selection of hyperparameters for YOLOv5 was presented. The limitation of the research performed was discussed in Section 5. Section 6 concludes the paper.

2. Related Works

The literature analysis performed has shown that one of the first tattoo identification forensics was the keyword-based matching method. Law enforcement authorities followed the ANSI/NIST-ITL 1-2011 standard, which defines eight major classes (human, animal, plant, flag, object, abstract, symbol, and other) and a total of 70 subclasses (including male face, cat, narcotics, American flag, fire, figure, national symbols, and wording) to categorize tattoos [11] to assign a single keyword to the tattoo image in the database. However, as Jain et al. [12] explain in their paper, in practice, searching for tattoo images based on keywords has several limitations: (1) the ANSI/NIST classes define a limited vocabulary that is not sufficient to describe different tattoo patterns; (2) several keywords may be required to accurately describe the tattoo image; (3) human annotation is subjective, meaning that different people can give quite different labels to the same tattoo. These deficiencies in the keyword-based tattoo image search system have led to the development of a Content-Based Image Recovery System (CBIR) to improve the efficiency and accuracy of tattoo search. To overcome the limitations of keyword-based tattoo matching, Jain et al. [13] proposed the CBIR called Tattoo-ID to match tattoos using the image-to-image method. Tattoo-ID extracts key points from tattoo images with scale-invariant feature transform (SIFT) (Lowe) and uses an unsupervised ensemble ranking algorithm to measure visual similarities between two tattoos [14].

A brief review of the literature on tattoo detection and identification is presented in Table 1. Research, where the main aim was tattoo detection, was usually motivated by forensic applications aimed at building tattoo-content-based image search systems to help law enforcement. Therefore, Han and Jain [15] proposed a system in which a cropped tattoo is segmented, represented by color, shape, and texture characteristics, and matched to the database. Duangphasuk and Kurutach [16] proposed an approach to the detection and segmentation of tattoo skin using image-negative methods in pre-processing to improve the retrieval and matching of tattoo images. The first step in this process was skin detection. The authors used various skin patches to perform the tasks of separating human skin color using the HSV model (hue, saturation, and value (or lightning)) model. In the second step, the negative image method was used to detect clear graphical images of the tattoo. In the third step, they extract the tattoo segment from the skin area of the negative image, and, as a result, negative images of the tattoos are obtained and can be used for further identification.

Table 1. Summary of tattoo detection and identification on related studies.

Authors	Object Detection/Localization Methods	Identification Methods	Dataset Name	Number of Tattoo Images	Non-Tattoo Images	Results
Sun et al. [10]	Faster R-CNN	n/a	Tatt-C: tattoo vs. non-tattoo; Flickr: tattoo vs. non-tattoo	1349 5740	1000 4260	Tatt-C (tattoo vs. non-tattoo): 98.25%; Tatt-C (localization): 45%@0.1FPPI; Flickr (tattoo vs. non-tattoo): 80.66%.
Silva and Lopes [9]	A deep learning model based on transfer learning	n/a	TattDetectB; TattDetectF	1000 1000	1000 1000	96.82% acc.@dense neural network as a classifier, with 10-fold cross-validation.
Han and Jain [15]	Pre-cropped tattoos	SIFT features; Sparse representation classification	MSU Sketch Tattoo	100	101,000	48% acc.@rank-100
Duangphasuk and Kurutach [16]	Image negative	n/a	Royal Thai Police	n/a	n/a	98.3% acc.@SIFT with image negative; 80.1% acc.@SIFT with original RGB image.
Jain et al. [12,13]	Gradient thresholding	Color histogram and correlogram; Shape moments; Edge direction coherence; Fusion of per-feature similarities	Tattoos from the web	2157	43,140	46% prec.@60% recall
Wilber et al. [17]	Pre-cropped tattoos	Exemplary code using HoG features; Random Forest classifier	238 tattoos from 5 classes	238	n/a	63.8% avg. acc. for 5 classes
Di and Patel [18]	AlexNet and SVM (tattoo vs. non-tattoo)	Siamese network with triplet or contrastive loss	Tatt-C: tattoo vs. non-tattoo; Mixed media	1349 181	1000 55	Tattoo vs. non-tattoo: 99.83% Mixed media: 56.9% acc.@rank-10
Manger [19]	n/a	SIFT features; Bag-of-words, hamming embedding, and weak geometry consistency	German police	417	327,049	78% acc.@rank-1
Han et al. [20]	Faster R-CNN	n/a	Tatt-C; NTU_Flickr; WebTattoo;	1349 5740 300,000	1000 260 n/a	87.10% recall (WebTattoo) 61.70% recall (Tatt-C)

Table 1. Cont.

Authors	Object Detection/Localization Methods	Identification Methods	Dataset Name	Number of Tattoo Images	Non-Tattoo Images	Results
Qingyong Xu et al. [21]	CNN	n/a	Tatt-C NTU_Flickr	1349 5740	1000 4260	98.80% acc.
Xu and Kong [22]	Decision tree	Shape Matching algorithm	Unidentified	547	n/a	52.38% acc.
Kim et al. [23]	GraphCut	n/a	Tatt-C (Detection); Evil (Detection)	6308 1105	n/a	Tatt-C: 70.5% acc.@41%recall Evil: 69.9% acc.@67.0%recall
Heflin et al. [24]	Automatic GraphCut and quasi-connected components	LBP-like features, SVM	Tattoo classification	50	500	85% acc.@10% FAR, on average, for 15 classes.
Allen et al. [25]	Segmentation algorithm	n/a	GANGINK tattoo database	256	n/a	~90% acc.

The Bag-of-Words (BoW) model, which uses SIFT functions, was probably the most popular in the early CBIR system for tattoo search [19]. In addition to SIFT features, local binary patterns (LBP) and histograms of oriented gradients (HoG) features were also used in the research by Wilber et al. [17] and Heflin et al. [24] with support vector machine (SVM) and random forest classifiers for tattoo classification. Although these CBIR systems have been reported to provide quite high accuracy on various benchmarks, they require careful manipulation of characteristic descriptors, vocabulary sizes, and indexing algorithms. The success of deep learning has led to the point where CBIR's methods are shifting from handcrafted features and models to deep learning methods. The AlexNet method has been successfully used for tattoo vs. non-tattoo classification in the work of Di and Patel [18].

Sun et al. [10] also focused on the tasks of tattoo image detection and localization. The authors developed TATT-RBDL, a tattoo detector that can classify images with one or more tattoos. Then, the region-based deep learning method (Faster R-CNN) was applied to the domain-specific data, and a tattoo detector was trained using two datasets, one with tattoo images and one with non-tattoo images. Xu and Kong [22] presented another decision tree-based approach. It achieved only 53.38% accuracy in its own dataset, resulting in a less expressive result. Recently, Han et al. [20] have also presented a detection model using faster R-CNN. This model classified detection problems as examples of image recovery systems in which learning and detection were performed simultaneously. Another basic tattoo detection method [23] uses a GraphCut-based method, with an accuracy of 70.5%. Silva and Lopes [9] presented a deep learning model based on transfer learning for tattoo detection problems.

Due to the wide diversity of tattoo types and the lack of image capture standards, the datasets may be quite different. Therefore, it is difficult or even unreasonable to compare the results. Since the datasets are real-world samples, it is especially important for the efficiency of machine learning methods that the datasets that are used reflect the same diversity. In addition, for multiclass datasets, class samples are another prominent issue, as classifiers tend to be strongly biased. Unfortunately, many real-world datasets do not follow this principle. For example, the Tatt-C dataset presented in the previous table, which is widely used in the literature, consists of images of faces in non-tattoo classes. This may prejudice a classifier trained with this dataset, which may acquire a false concept that images without tattoos are face-type images. Nevertheless, this database was the most used database for tattoo studies to date [9]. The first results on the Tatt-C dataset were published in response to the challenges issued by NIST [26]. According to the report by Ngan et al. [26], four institutions participated in this challenge, with MorphoTrek as the best performance, with 96.3% accuracy. Unfortunately, the algorithms developed by the participants in the NIST challenge have not been published. This was criticized by Qingyong Xu et al. [21] in their work, and it emerged that it was impossible to perform external validation tests.

Although the concept of tattoo detection and identification is theoretically uncomplicated, the process is not simple and depends on various factors. There are no defined standards of what tattoos are in terms of shape, color, size, proportion of individuals, and their location on the body. Additionally, a single image may have several tattoos. Furthermore, the background of an image can introduce significant noise into the detection process because its complexity can be confused with the tattoo itself. Furthermore, it is difficult to compare different studies due to differences in the test procedures, metrics, and datasets used. There are relatively few publications dealing with tattoo detection and identification problems using deep learning. It should also be noted that most previous tattoo studies were based on the NIST Tatt-C dataset, which was discontinued over time and is no longer available for download and use. The lack of standardized datasets for the detection of tattoos is one of the problems in this field of research. Methods such as Faster R-CNN, RetinaNet, YOLO, and SSD, coupled with feature extraction models such as VGG, ResNet, Siamese Networks, and Triplet Networks, collectively contribute to the intricate landscape of tattoo detection and identification [27]. Furthermore, a diverse set

of similarity measures, including Euclidean distance, cosine similarity, and others, form a versatile toolkit for evaluating similarities across various data types.

3. Background of the Experimental Investigation

The review of related works has shown (Table 1) that various deep learning-based algorithms can be used for tattoo detection or localization tasks. In our research, the focus is on a real-time object detection task, so CNN and R-CNN methods are not entirely appropriate. When comparing real-time object detector algorithms YOLO, SSD, and RetinaNet, it should be noted that SSD has low accuracy compared to other alternatives. Furthermore, RetinaNet exhibits better accuracy compared to YOLO or SSD, but lower efficiency for real-time object detection due to its high computational cost. Although YOLOv5 continues to be a widely acclaimed real-time object detection algorithm, boasting improved accuracy over its predecessors and retaining the ability to identify even diminutive objects. So, based on related works, the YOLOv5 object detection algorithm was used in the experimental investigation for tattoo detection. The pre-trained models of YOLOv5 were used as a base [28] (Table 2).

Table 2. The specification of the pre-trained YOLOv5 models used [28].

Models	Image Size (Pixels)	mAP ^{val} (50–95)	mAP ^{val} (50)	Speed (ms) CPU bl	Speed (ms) V100 bl	Speed (ms) V100 b32	Params (M)	FLOPs @640 (B)
YOLOv5n	640	28.0	45.7	45	6.3	0.6	1.9	4.5
YOLOv5s	640	37.4	56.8	98	6.4	0.9	7.2	16.5
YOLOv5m	640	45.4	64.1	224	8.2	1.7	21.2	49.0
YOLOv5l	640	49.0	67.3	430	10.1	2.7	46.5	109.1
YOLOv5x	640	50.7	68.9	766	12.1	4.8	86.7	205.7

During the experimental investigation, all models were trained in an environment with the following specifications: Intel(R) Xeon(R) CPU E5-2630 v4 @ 2.20GHz (20 Threads, 10 Cores). The environment had a Linux operating system with 32 GB DDR4 RAM and GPU, called Tesla P100 PCIe 12GB.

After the tattoo has been detected, the person's identification in the database based on similarity distance has been analyzed. In this investigation, two similarity distances were analyzed: cosine and Euclidean distances. Suppose that we have two images: the tattoo detected in image $X = (x_1, x_2, \dots, x_N)$, and the tattoo from the database $Y = (y_1, y_2, \dots, y_N)$, where N is the dimensionality of the vector corresponding to each datum. In this case, the Euclidean similarity distance can be calculated (1).

$$d(X, Y) = \sqrt{\sum_{k=1}^N (x_k - y_k)^2} \quad (1)$$

The Euclidean distance is the distance between two data points in Euclidean space. In the context of data analysis, it is often used to find the dissimilarity or similarity between data. Smaller values of the Euclidean distance indicate greater similarity. The cosine similarity distance (2) is useful in the analysis of high-dimensional data, such as in the detection of pairwise similarities in images. The value of cosine similarity indicates the cosine of the angle between two vectors in a multidimensional space.

$$d(X, Y) = 1 - \frac{\sum_{k=1}^N x_k y_k}{\sqrt{\sum_{k=1}^N x_k^2} \times \sqrt{\sum_{k=1}^N y_k^2}} \quad (2)$$

4. The Experimental Investigation

The idea of automated person identification based on their tattoo is divided into five steps:

1. Preparation of the tattoo dataset;
2. The experiments performed on the influence of data augmentation parameters;
3. The experiments performed on the influence of hyperparameters of YOLOv5;
4. Preparation of the identification dataset;
5. Estimation of similarity threshold.

All five steps and the principles of it are presented in Figure 1 and are detailed in further sections.

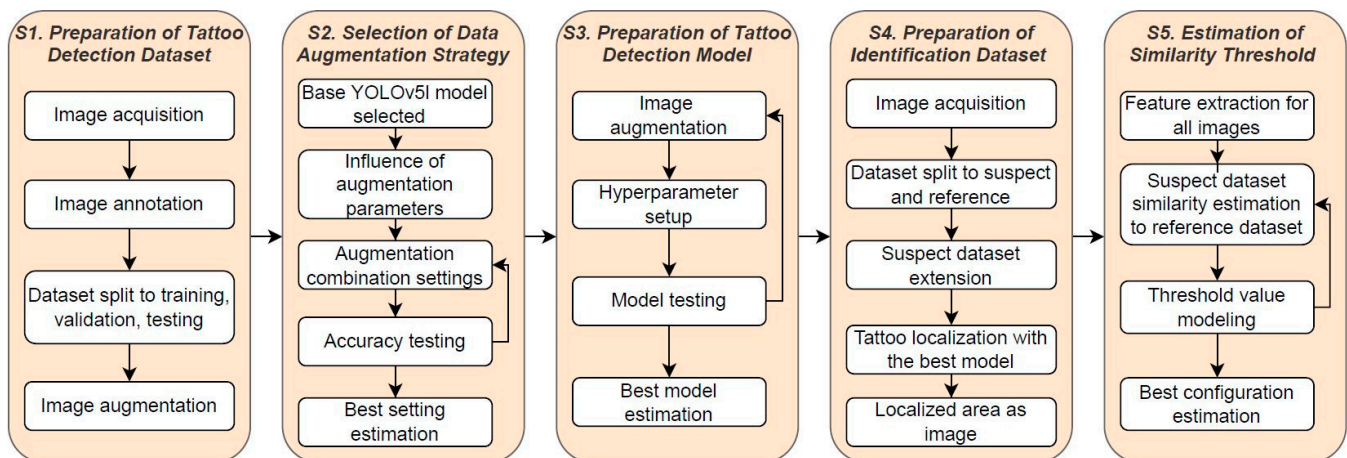


Figure 1. Main principle of automated suspect identification model development based on tattoo photos.

4.1. Preparation of Tattoo Dataset

The analysis of related works has shown that the availability of publicly available tattoo datasets is not very high. In some of the studies, several authors have created their datasets, but the datasets are not publicly available. Additionally, as mentioned, the Tatt-C dataset was frequently used in related studies, but this dataset has been discontinued and can no longer be downloaded. Other equally popular datasets, such as the NTU Tattoo dataset and WebTattoo, require a special agreement to be used for research purposes, which has not been obtained. Therefore, in this paper, only one publicly available deMSI (Hrkać et al., 2016 [27]) dataset was chosen for model training. The sample of the dataset is presented in Figure 2.

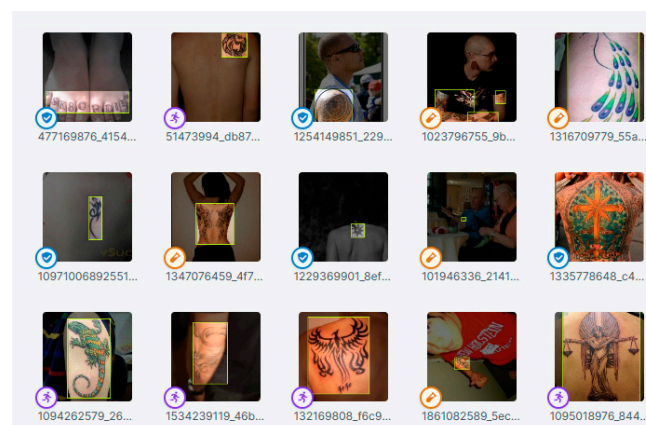


Figure 2. Sample of the deMSI tattoo dataset used in the research.

To facilitate the development and testing of their proposed method, the authors have assembled their dataset by collecting and manually labeling 450 tattoo images from the ImageNet database. Each of the collected images contains one or more tattoos. The authors of this dataset used the ConvNet model in their research and, therefore, annotated each tattoo using a series of connected line segments. However, in the case of this study, such an annotation will not be suitable for our research because the chosen object detection model for tattoo detection is YOLOv5. The chosen model accepts bounding box annotations for each object, where each object in an image is surrounded by a rectangular box that can be described by the coordinates of its top-left corner and its width and height. Therefore, the images in the dataset have been manually annotated. In this study, a dataset was created that contained a total of 1000 images.

4.2. The Experiments Performed on the Influence of Data Augmentation Parameters

As mentioned, the results of tattoo detection can depend on various parameters. First of all, the influence of data augmentation parameters has been analyzed using the pre-trained YOLOv5l model, as according to the related work analysis performed, for such types of tasks, it is the most suitable model. The results of the experiment are presented in the table below (Table 3).

Table 3. Augmentation experiment results.

Augmentation	Augmentation Specifications	Precision	Recall	mAP@ 0.5	mAP@0.5:0.95
No augmentation applied		0.735	0.733	0.739	0.454
Flip	Horizontal and vertical	0.819	0.748	0.799	0.582
90° Rotate	Clockwise, counterclockwise, and upside down	0.834	0.726	0.801	0.608
Crop	0% minimum zoom and 30% maximum zoom	0.884	0.735	0.82	0.585
Rotation	Between -15° and $+15^\circ$	0.861	0.741	0.788	0.51
Shear	$\pm 15^\circ$ horizontal and $\pm 15^\circ$ vertical	0.811	0.733	0.794	0.548
Grayscale	Applied to 25% of images	0.788	0.674	0.735	0.46
Hue	Between -25° and $+25^\circ$	0.835	0.8	0.794	0.568
Saturation	Between -30% and $+30\%$	0.818	0.756	0.804	0.583
Brightness	Between -25% and $+25\%$	0.849	0.689	0.77	0.543
Exposure	Between -15% and $+15\%$	0.816	0.723	0.786	0.551
Blur	Up to 2px	0.814	0.756	0.8	0.566
Noise	Up to 1.49% of pixels	0.78	0.763	0.78	0.572

All augmentation variations were trained on the deMSI dataset with 300 epochs. On the basis of the provided experimental results, some considerations were made. Crop augmentation allowed us to obtain the highest mAP@0.5 (0.82) and good precision and recall scores. It would be an effective option in this study. When talking about balanced precision and recall, hue augmentation showed a good balance between precision and recall, with a decent mAP@0.50 score (0.794). So, this augmentation should also be considered. In addition, computational efficiency must be considered: flip, 90-degree rotation, and blur. These enhancements showed a good balance between precision and recall with moderate mAP@0.5 scores. They may be computationally efficient. Also, it should be mentioned that it is a good practice to avoid low-performing augmentations; in this case, it would be grayscale and rotation. These augmentations had lower mAP@0.5 scores. Depending on priorities, it might be good to consider excluding them from the final augmentation strategy.

To achieve even higher results, it was decided to try to combine multiple augmentations (Table 4).

Table 4. Result of the combined augmentation experiment.

Augmentation	Augmentation Specifications	Precision	Recall	mAP@0.5	mAP@0.5:0.95
Flip	Horizontal and vertical				
Crop	0% minimum zoom and 30% maximum zoom	0.881	0.71	0.793	0.584
Hue	Between -25° and $+25^\circ$				
Hue	Between -25° and $+25^\circ$				
Saturation	Between -30% and $+30\%$	0.796	0.733	0.763	0.533
Brightness	Between -25% and $+25\%$				
Flip	Horizontal and vertical				
Crop	0% minimum zoom and 30% maximum zoom	0.827	0.726	0.81	0.532
Rotation	Between -15° and $+15^\circ$				
90° Rotate	Clockwise, counterclockwise, and upside down				
Shear	$\pm 15^\circ$ horizontal and $\pm 15^\circ$ vertical	0.878	0.744	0.829	0.593
Blur	Up to 2px				
Crop	0% minimum zoom and 30% maximum zoom				
Exposure	Between -15% and $+15\%$	0.8	0.756	0.784	0.586
Noise	Up to 1.49% of pixels				

Taking into account the metrics mAP@0.5 and mAP@0.5:0.95 metrics, group 4 (90° rotation, shear, and blur) appears to have the best overall performance, with the highest values of mAP@0.5 and mAP@0.5:0.95 values. Therefore, group 4 was chosen as the best-performing augmentation strategy. Based on the results of the experiments provided, the following pre-processing and augmentation steps were applied in this study.

- **Resize.** The images were resized to a uniform dimension of 320×320 pixels, which is a common choice for the YOLOv5 models. This step not only standardizes the input size but also enhances training efficiency;
- **90° Rotate.** The images were rotated 90 degrees during the augmentation process. This rotation can help the model become more robust to object orientations in the training data. It introduces variations in the orientation of objects, making the model more versatile;
- **Shear.** Shearing involves shifting one part of an image in a certain direction, creating a “tilted” effect. In this study, shear was applied horizontally and vertically in a range of $\pm 15^\circ$. Shearing introduces distortions that can improve the model’s ability to recognize objects from different perspectives;
- **Blur.** A blur filter was applied to the images, with a maximum blur of up to 2 pixels. Blur helps simulate real-world conditions where images may not be perfectly sharp. It can prevent the model from relying too heavily on fine details and encourage a more generalized understanding of the objects in the images.

Collectively, these pre-processing and augmentation techniques aim to increase the robustness and ability of the model to handle tattoo detection in a wide range of real-world scenarios.

4.3. The Experiments Performed on the Influence of Hyperparameters of YOLOv5

To find out which size of the YOLOv5 model (n, s, m, l, x) and which hyperparameters allow to obtain the highest tattoo detection results, an additional experiment has been performed using the data augmentation options from previous experiment results. Related works have shown that, in various investigations, usually only three hyperparameters are changed to improve the results: learning rate, momentum, and weight decay. Therefore, in this investigation, the combination of five YOLOv5 models and three hyperparameters has been analyzed. The hyperparameters changed in this way:

- learning rate: 0.01; 0.001; 0.0001;
- momentum: 0.9; 0.935; 0.95;
- weight decay: 0.0001; 0.0005; 0.0007.

In this way, a total of 135 models were trained and tested (26 models for each size of the YOLOv5). The other parameters have been chosen considering the primary research performed and have not been changed during the training of all 135 models. This ensured the same condition during the experimental investigations in this research. The fixed parameters of YOLOv5 are as follows:

- image size: 320×320 ;
- batch size: 32;
- number of epochs: 300;
- optimizer: SGD.

In Table 5, the average results for each size of the YOLOv5 model are presented. As we can see (Table 5), the highest averaged precision results are obtained by YOLOv5m (0.87). Slightly small results are obtained by YOLOv5l (0.86), YOLOv5s (0.85), and YOLOv5x (0.81). The smallest averaged precision is obtained by YOLOv5n, which is equal to 0.67. All estimation measures are significantly smaller using YOLOv5n compared to other sizes of the models. The highest average recall value is obtained by the YOLOv5l model (0.70). In the case of mAP values, slightly better results are also obtained using the YOLOv5l model, where mAP@0.5 is equal to 0.79 and mAP@0.5:0.95 is equal to 0.60. In Table 6, the standard deviation for each size of the YOLOv5 model is presented. As we can see, the deviation is not high, so it means that there is no very high influence on which hyperparameters will be used to train the tattoo detection model.

Table 5. The average results of 26 model estimations for each size of the YOLOv5 model.

YOLOv5 Model	YOLOv5n	YOLOv5s	YOLOv5m	YOLOv5l	YOLOv5x
Precision	0.67	0.84	0.87	0.86	0.81
Recall	0.46	0.67	0.68	0.70	0.67
mAP@0.5	0.52	0.75	0.78	0.79	0.76
mAP@0.5:0.95	0.31	0.51	0.58	0.60	0.59

Table 6. The standard deviation results of 26 model estimations for each size of the YOLOv5 model.

YOLOv5 Model	YOLOv5n	YOLOv5s	YOLOv5m	YOLOv5l	YOLOv5x
Precision	0.16	0.06	0.04	0.06	0.07
Recall	0.17	0.02	0.03	0.05	0.03
mAP@0.5	0.19	0.02	0.02	0.03	0.02
mAP@0.5:0.95	0.13	0.02	0.03	0.02	0.01

Summarizing the results of the experimental investigation on the influence of hyperparameters, the YOLOv5l model has been chosen as the basis for tattoo detection. The

highest values of mAP@0.5 (0.82) and mAP0.5:0.95 (0.63) have been obtained using such hyperparameters of YOLOv5l: the learning rate is set at 0.001; momentum is set at 0.95; the weight decay is set at 0.0001. In this case, the precision is equal to 0.87, and the recall is 0.75. In addition, in Figure 3, the graphs of precision, recall, mAP@0.5, and map@0.5:0.95 are presented.

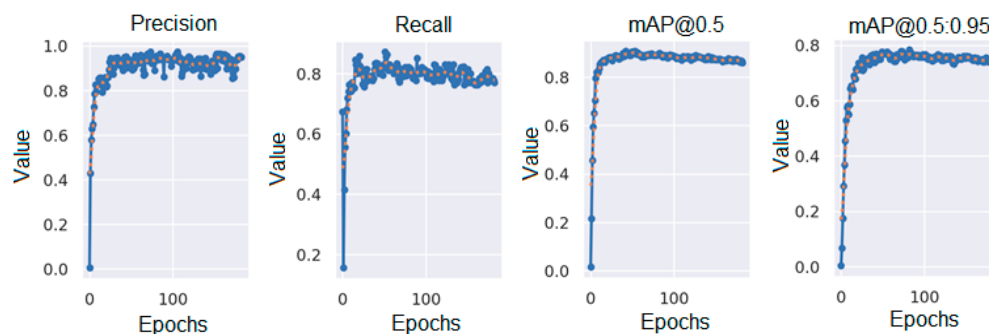


Figure 3. Evaluation of YOLOv5l with the highest mAP@0.5:0.95.

4.4. Preparation of Identification Dataset

To correspond to real conditions, an additional dataset was constructed for identity estimation. Twelve persons were asked to provide at least 6 photos of each of their tattoos. The tattoos had to be taken under good conditions, when the full tattoo is visible, as well as in lower quality when only part of the tattoo is visible, it is partially covered, etc. An ID was assigned to each tattoo. The best photo of the tattoo was selected as a reference model, while other photos of the tattoo were added to the suspect dataset. Additionally, 10 random unused photos from the deMSI dataset were added to the reference dataset and 44 to the suspect dataset. This allows us to reflect on situations when no reference tattoo exists for the analyzed one. There were 43 reference photos and 209 suspect photos.

Using the best model for tattoo detection, both datasets were processed to obtain only the cropped version of the localized tattoo in each photo. After the detection, the reference dataset contained 49 photos, while the suspect dataset contained 245 photos (167 of the tattoos in the reference dataset and 78 not listed in the reference dataset). The increase was affected by the fact that in some photos multiple areas were detected. Sometimes, one tattoo was divided into several parts. In other cases, non-tattoo areas were localized as tattoos.

The suspect dataset was left as it was because this part will have to be performed completely automatically. Meanwhile, the reference dataset was manually revised to leave only photos of good quality. During the revision, redundant or not full photos were eliminated, leaving only 39 reference tattoos, 1 photo for each of the tattoos.

4.5. Estimation of the Similarity Threshold

For further analysis, each photo was resized to a dimension of 224×224 px and pre-processed for ResNet50 suitable feature extraction. Each photo is represented as a (7, 7, 2048) dimension output, where, after flattening, it contains 100,352 values for comparison. This vector was used to estimate the similarity between each suspect and each reference photo. For similarity estimation two most often used similarity methods were used: cosine and Euclidean similarity. Usually, F-score and accuracy metrics are used to define the best model. Using cosine similarity, the threshold value should be 0.45–0.5 to achieve a 48% F-score and 99% accuracy (Figure 4, left chart). For Euclidean distance, the F-score and accuracy optimal threshold value would be 525 and would allow us to achieve a 0.46% F-score and 99% accuracy (Figure 4, right chart).

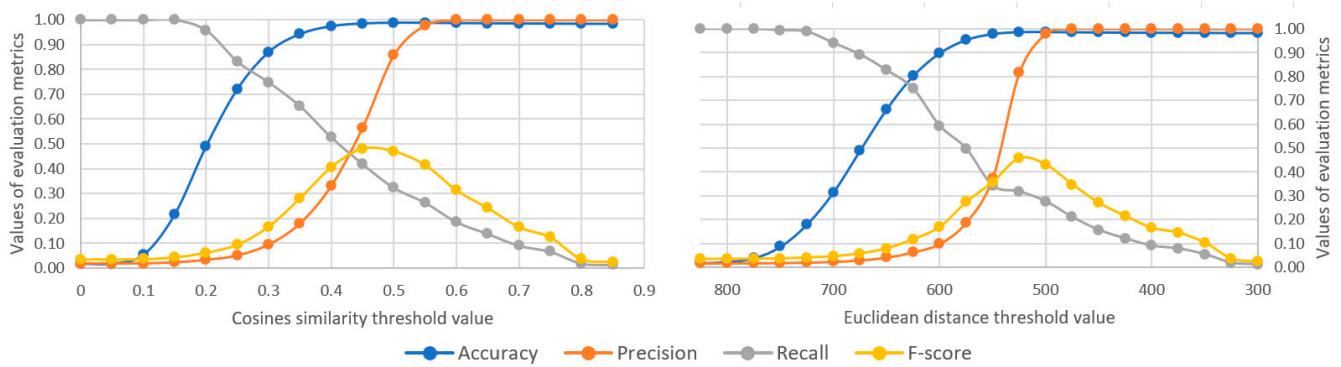


Figure 4. Tattoo identification accuracy metrics modeling results based on selected threshold values for cosines similarity and Euclidean distance.

However, for suspect linking to reference photo task, accuracy and F-score are not the best measurements as the automation should work as decision support and workload reduction, but not a human replacement solution. The final results will have to be verified by a person in any case to ensure that no false results are provided. Therefore, False-positive (FP) and False-negative (FN) values are important. Table 7 presents the threshold values under which tattoo identification achieves 100% recall or precision. Accordingly, it provides numbers that indicate how much workload could be reduced in the case of suspect photos and comparisons. Suspect photos define the suspect photos, which do not require any revision, as all similarity scores to reference tattoos are below or above the threshold value and, therefore, will for sure not be linked to any of the reference photos. Meanwhile, comparison reduction indicates that some of the threshold values do not meet the interval; therefore, the suspect photo has to be compared not to all but just to some reference tattoos. The result indicates that cosine similarity is better for recall assurance situations as it allows a reduction of 2% of photos and 20% of comparisons. Meanwhile, if the task is oriented toward precision assurance, it is better to use the Euclidean distance. The difference to cosine similarity is not very high, but under the same 100% precision, it provides a lower False negative, higher True-positive, and the same True-negative rate.

Table 7. The summary of experimental results using cosine and Euclidean distances.

Task Type	Cosine Distance			Euclidean Distance		
	Threshold Values	Workload Reduction		Threshold Values	Workload Reduction	
		Suspect Photos	Comparisons		Suspect Photos	Comparisons
Guarantee that no suspects will be missed that should be linked to reference tattoos (no False-negatives, recall 100%)	[0.00; 0.15]	2%	20%	[775; 900]	0%	2%
Guarantee that all links to reference tattoos are accurate (no False-positives, precision 100%)	[0.60; 1.00]	12%	0%	[300; 475]	14%	0%

The zero reduction in 100% precision-oriented task for comparisons indicates that the similarity of the suspect photo was the highest among all reference tattoos; therefore, only the True-positive values were left as candidates. If the models were adjusted to take not the threshold value with multiple candidates but to link each photo with the highest similarity/lowest distance reference tattoo, the false positive ratio would increase automatically as not all suspect photos have a reference. Meanwhile, for the cosine similarity case with the highest F-score, the F-score increases from 48% to 52%, while under those conditions, the precision decreases to 42% from 56%, the recall increases from 42% to 65%, and accuracy remains 98%. For Euclidean distance, the accuracy decreases from 99% to

98%, precision from 82% to 41%, recall increases from 32% to 60%, while the F-score remains 46% (Figure 5).

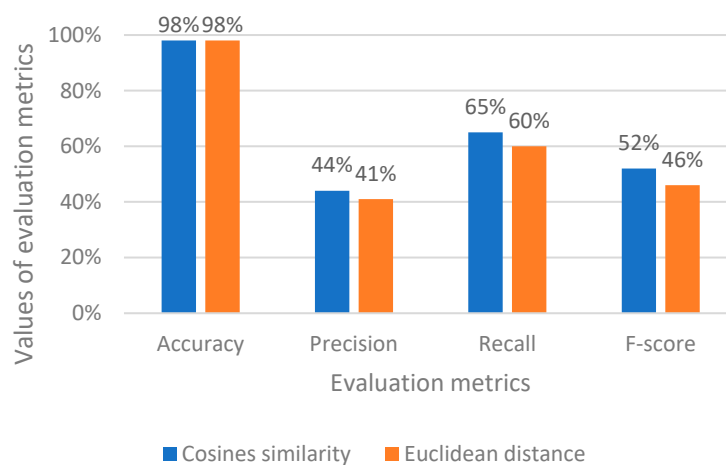


Figure 5. Tattoo identification accuracy scores for cosine similarity and Euclidean distance when only the most similar reference tattoo is taken as the model prediction.

5. Discussion

In this paper, a solution for automated tattoo detection and identification was implemented. The experimental investigation carried out in this research has focused on different types (n, s, n, l, x) of YOLOv5 models, YOLOv5 hyperparameters, data augmentation parameters, and similarity distances used in the identification stage. The newest versions of YOLO have not been analyzed because these algorithms have not been officially released and could be found in the public repository without any scientific investigation. Therefore, in this research, the most popular YOLOv5 version has been used today. Taking into account the research performed by other authors, some parameters have not been investigated due to the cost of time for each model training. A total of 135 models have been trained to find the influence of three main hyperparameters on the results of tattoo detection. It is necessary to admit that a complete investigation using more combinations has not been performed. Even with these research limitations, the results of this research have shown that for such a type of object detection task, the most suitable models are YOLOv5l.

The chosen similarity distances used in the identification stage are the most used distances in various clustering algorithms, similarity detection tasks, recommendation systems, etc. Related works have shown that similarity distances, such as Jaccard, Spearman correlation, Manhattan, and others, can also be used for identification, but primary research has shown that in our case, it was not suitable.

6. Conclusions

In this paper, a solution for automated tattoo detection and identification was implemented. This task is multi-stage as it requires both tattoo detection, its bounds estimation, as well as comparison to the reference tattoo. Such a solution, when the second stage is implemented based on similarity rather than the trained model, has the advantage of the easy extension of the reference dataset; there is no need to retrain the model for added reference tattoos.

After investigation of photo augmentation and the impact of the YOLOv5 hyperparameter on tattoo detection, the highest values of mAP@0.5 (0.82) and mAP0.5:0.95 (0.63) have been obtained. Those were obtained using the YOLOv5l model, with the learning rate set at 0.001, momentum set at 0.95, and the weight decay set at 0.0001, while the photos were augmented using 90° rotation, shear, and blur options. This model was not only able to achieve the best detection but led to the highest recall. This is important as it is

better to have a bigger set of instances for the next stage rather than miss some tattoos or its elements.

In the similarity estimation between tattoos, the highest accuracy in linking the tattoo photo with one reference tattoo with the highest similarity score reached 98%, while the F-score is up to 52%. This would not be an acceptable accuracy for criminal identification or similar tasks. However, the similarity score can be used for the reduction of manual work revising the possible candidates. By applying cosine similarity, all cases where similarity is less than the threshold value of 0.15 can be ignored. This would decrease the workload by 20% while no False-negative cases would be skipped.

The results obtained during the experimental investigation have shown that tattoo detection and identification tasks require larger models than YOLOv5n. Additionally, the learning rate, momentum, and decay weight parameters have not influenced the results too much. Considering the possible implementation of the models obtained in the real environment, such as real-time detection systems, the most suitable are YOLOv5m and YOLOv5l. The training time of these models is lower compared to that of the YOLOv5x; therefore, it would be easy to retrain the models using more tattoo images and improve the quality of tattoo detection. In the future, the newest versions of YOLO could be trained and tested under the same conditions to see how it influences the results of tattoo detection and identification.

Author Contributions: Conceptualization, G.P., P.S. and S.R.; methodology, P.S. and S.R.; validation, G.P., E.P. and S.R.; formal analysis, G.P. and P.S.; data curation, G.P., E.P. and P.S.; writing—original draft preparation, G.P., E.P., P.S. and S.R.; writing—review and editing P.S. and S.R., visualization, G.P. and P.S.; supervision, P.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Dataset available on request from the authors.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Jucevičius, J.; Treigys, P.; Bernatavičienė, J.; Briedienė, R.; Naruševičiūtė, I.; Trakymas, M. Investigation of MRI prostate localization using different MRI modality scans. In Proceedings of the 2020 IEEE 8th Workshop on Advances in Information, Electronic and Electrical Engineering (AIEEE), Vilnius, Lithuania, 22–24 April 2021; pp. 1–5.
2. Virbukaitė, S.; Bernatavičienė, J. Deep learning methods for glaucoma identification using digital fundus images. *Balt. J. Mod. Comput.* **2020**, *8*, 520–530. [[CrossRef](#)]
3. Gupta, P.; Pareek, B.; Singal, G.; Rao, D.V. Edge device based military vehicle detection and classification from UAV. *Multimed. Tools Appl.* **2022**, *81*, 19813–19834. [[CrossRef](#)]
4. Usamentiaga, R.; Lema, D.G.; Pedrayes, O.D.; Garcia, D.F. Automated surface defect detection in metals: A comparative review of object detection and semantic segmentation using deep learning. *IEEE Trans. Ind. Appl.* **2022**, *58*, 4203–4213. [[CrossRef](#)]
5. Li, W.; Zhang, H.; Wang, G.; Xiong, G.; Zhao, M.; Li, G.; Li, R. Deep learning based online metallic surface defect detection method for wire and arc additive manufacturing. *Robot. Comput. Integr. Manuf.* **2023**, *80*, 102470. [[CrossRef](#)]
6. Kvietkauskas, T.; Stefanovič, P. Influence of Training Parameters on Real-Time Similar Object Detection Using YOLOv5s. *Appl. Sci.* **2023**, *13*, 3761. [[CrossRef](#)]
7. Kvietkauskas, T.; Pavlov, E.; Stefanovič, P.; Pliuskuvienė, B. The Efficiency of YOLOv5 Models in the Detection of Similar Construction Details. *Appl. Sci.* **2024**, *14*, 3946. [[CrossRef](#)]
8. Stefanovič, P.; Ramanauskaitė, S. Travel Direction Recommendation Model Based on Photos of User Social Network Profile. *IEEE Access* **2023**, *11*, 28252–28262. [[CrossRef](#)]
9. Silva, R.T.; Lopes, H.S. A Transfer Learning Approach for the Tattoo Detection Problem. In Proceedings of the XV Brazilian Congress of Computational Intelligence (SBIC), Online, 3–6 October 2021; pp. 1–8.
10. Sun, Z.H.; Baumes, J.; Tunison, P.; Turek, M.; Hoogs, A. Tattoo detection and localization using region-based deep learning. In Proceedings of the 2016 23rd International Conference on Pattern Recognition (ICPR), Cancun, Mexico, 4–8 December 2016; pp. 3055–3060.
11. Mangold, K.C. *Data Format for the Interchange of Fingerprint, Facial and Other Biometric Information*. ANSI/NIST-ITL 1-2011 NIST Special Publication 500-290, 3rd ed.; CreateSpace Independent Publishing Platform: Scotts Valley, CA, USA, 2016.

12. Jain, A.; Lee, J.-E.; Jin, R.; Tong, W. Image Retrieval in Forensics: Tattoo Image Database Application. *IEEE Multimed.* **2012**, *19*, 40–49.
13. Jain, A.; Lee, J.-E.; Jin, R. Unsupervised Ensemble Ranking: Application to Large-Scale Image Retrieval. In Proceedings of the 2010 20th International Conference on Pattern Recognition, Istanbul, Turkey, 23–26 August 2010; pp. 3902–3906.
14. Lowe, D.G. Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [[CrossRef](#)]
15. Han, H.; Jain, A.K. Tattoo based identification: Sketch to image matching. In Proceedings of the 2013 International Conference on Biometrics (ICB), Madrid, Spain, 4–7 June 2013; pp. 1–8.
16. Duangphasuk, P.; Kurutach, W. Tattoo skin detection and segmentation using image negative method. In Proceedings of the 2013 13th International Symposium on Communications and Information Technologies (ISCIT), Samui Island, Thailand, 4–6 September 2013; pp. 354–359.
17. Wilber, M.J.; Rudd, E.; Heflin, B.; Lui, Y.M.; Boulton, T.E. Exemplar codes for facial attributes and tattoo recognition. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision, Steamboat Springs, CO, USA, 24–26 March 2014; pp. 205–212.
18. Di, X.; Patel, V.M. Deep Tattoo Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Las Vegas, NV, USA, 26 June–1 July 2016; pp. 119–126.
19. Manger, D. Large-Scale Tattoo Image Retrieval. In Proceedings of the 2012 Ninth Conference on Computer and Robot Vision, Toronto, ON, Canada, 28–30 May 2012; pp. 454–459.
20. Han, H.; Li, J.; Jain, A.K.; Shan, S.; Chen, X. Tattoo Image Search at Scale: Joint Detection and Compact Representation Learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *41*, 2333–2348. [[CrossRef](#)] [[PubMed](#)]
21. Xu, Q.; Ghosh, S.; Xu, X.; Huang, Y.; Kong, A.W.K. Tattoo detection based on CNN and remarks on the NIST database. In Proceedings of the 2016 International Conference on Biometrics (ICB), Halmstad, Sweden, 13–16 June 2016; pp. 1–7.
22. Xu, X.; Kong, A.W.K. A geometric-based tattoo retrieval system. In Proceedings of the 2016 23rd International Conference on Pattern Recognition (ICPR), Cancun, Mexico, 4–8 December 2016; pp. 3019–3024.
23. Kim, J.; Li, H.; Yue, J.; Ribera, J.; Delp, E.J.; Huffman, L. Automatic and manual tattoo localization. In Proceedings of the 2016 IEEE Symposium on Technologies for Homeland Security (HST), Waltham, MA, USA, 10–11 May 2016; pp. 1–6.
24. Heflin, B.; Scheirer, W.; Boulton, T.E. Detecting and classifying scars, marks, and tattoos found in the wild. In Proceedings of the 2012 IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS), Arlington, VA, USA, 23–27 September 2012; pp. 31–38.
25. Allen, J.D.; Zhao, N.; Yuan, J.; Liu, X. Unsupervised tattoo segmentation combining bottom-up and top-down cues. In Proceedings of the SPIE Defense, Security, and Sensing 2011, Orlando, FL, USA, 29 April 2011.
26. Ngan, M.; Quinn, G.W.; Grother, P. Tattoo Recognition Technology—Challenge (Tatt-C) Outcomes and Recommendations. 2016. Available online: <https://www.nist.gov/programs-projects/tattoo-recognition-technology-challenge-tatt-c> (accessed on 11 June 2024).
27. Hrkač, T.; Brkić, K.; Kalafatić, Z. Tattoo Detection for Soft Biometric Deidentification Based on Convolutional Neural Networks. In Proceedings of the OAGM-ARW Joint Workshop, Wels, Austria, 11–13 May 2016; pp. 131–138.
28. Jocher, G.; Chaurasia, A.; Stoken, A.; Borovec, J.; Kwon, Y.; Michael, K.; Fang, J.; Zeng, Y.; Wong, C.; Montes, D.; et al. *ultralytics/yolov5: v7.0—YOLOv5 SOTA Realtime Instance Segmentation; Version 7.0; Zenodo: Genève, Switzerland, 2022.* [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.