ISSN 2007-9737

Proposal of a Methodology to Identify Unidentified Decedent through Artificial Intelligence Techniques

Ramon A. Briseño[1], Edgar G. Cossio[2,*], Blanca I. Félix[3], Ana C. Ayala[4], Fausto A. Bernal[5], Benjamín I. Hernandez[6], Briseyda P. Martinez[4]

[1] Universidad de Guadalajara,
Centro Universitario de Ciencias Económico Administrativas,
Doctorado en Tecnologías de Información,
Mexico

[2] Instituto de Información Estadística y Geográfica de Jalisco,
Mexico

[3] Instituto Tecnológico Superior de Cajeme,
Mexico

[4] Instituto Tecnológico de Estudios Superiores de Zamora,
Mexico

[5] Universidad Autónoma de Occidente,
Mexico

[6] Instituto Tecnológico de la Piedad,
Mexico

alejandro.bmartinez@alumnos.udg.mx, edgar.cossio@iieg.gob.mx,
18130134@e-itesca.edu.mx, {aayala067, bmartinez077}@accitesz.com,
{z.fausto.a.b.b, ivan1516hc}@gmail.com

Abstract. The use of artificial intelligence techniques for the identification and recognition of people is increasing. This article proposes a novel methodology based on artificial neural networks which is based on the classification of tattoos for their analysis and subsequent identification. The techniques used were carefully validated and tested to seek the highest precision in the result. The dataset used was compiled with sample tattoos for training. The model was loaded into a web model that is currently being tested based on Django and python. A dataset and python scripts were built which allowed its implementation.

Keywords. Unknown bodies, artificial intelligence, identification, tattoos, neural network.

1 Introduction

The existence of a human rights crisis in Mexico is a reality. Day by day, people in their daily lives are exposed to crimes such: as murder, torture, forced disappearance, forced displacement, and executions [1]. This is generating unidentified missing persons. In the state of Jalisco, in 2019, 2,100 people were reported missing [2]. By October that same year, 4,060 unidentified decedents (UD) were reported too [3].

The figures of UD exceed the capacities of forensic institutes; therefore, developing strategies that optimize the early identification of bodies is required [4]. Many times, the features of the bodies can change over time; for example, decomposing bodies of white people change the color of their skin and seem to match individuals of other races. The difficulties of recognizing them increase as time passes since they enter into different changes or cadaveric processes [5, 6].

Identifying and registering dead bodies is an interdisciplinary and complex activity that must be consistent, simple, efficient, and economical. There is no universal identification method of bodies because, in each case, the form of identification that best suits the situation presented must be chosen [4].

However, one of the primary identification methods is comparing the information from the UD (physical characteristics, clothing, artifacts, tattoos, among others) with the information on missing persons [5].

The art of tattooing as a decoration and expression technique has gained importance among many population segments, regardless of age or social class. Figure 1 shows the percentage of male and female bodies with at least one tattoo by age group.

Of the unidentified bodies in Jalisco, 45% had at least one tattoo [4]. It can be identified in Figure 1 that tattoos have a growing trend in the bodies of young UDs. Therefore, creating a tattoo catalog would significantly reduce the time it takes to identify a person.

This work intends to detail a methodology based on artificial vision to compare the information of the missing persons and the UD. The methodology implements a model of convolutional neural networks to recognize and classify tattoos through images and thereby identify missing persons.

This paper is articulated as follows: section two describes the societal problem that represents the identification of unidentified decedents; section three explains the proposed solution to the problem statement using artificial vision technologies; section four explores and discusses the results and, finally, section five concludes and mentions the following steps on this research.

2 Problem Statement

Due to the increase in unidentified and unclaimed corpses that have arrived at the Prosecutor's Office facilities since 2006, Forensic Medical Services (SEMEFO as per the acronym in Spanish) presents a saturation of space in its facilities in Jalisco [7].

In forensic identification, there are several methods for the morphological description of corpses, one of which is to identify individual features of each person, such as moles, fractures, and scars, among others. From a forensic point of view, tattoos are interesting because they are "intentional scars" and, at the same time, individualizing features [8].

Forensic identification through tattoos is a method of skin identification currently used by SEMEFO. Said method is effective, efficient, and cheap compared to biological tests [4].

There is a physical catalog, similar to a photo album, where a person searches for a UD among each of the photos until a match is found, the problem is that the search time increases significantly when thousands of bodies have one or more tattoos.

Also, during the identification of a person, the information collected post-mortem and ante-mortem must be as accurate and objective as possible.

Currently, SEMEFO tries to filter the searches by adding descriptions of the tattoos of UD, however, tattoos are art pieces open to personal interpretation, which creates a bias between viewers and the original tattoo [4].

Therefore, it is imperative to optimize the identification process of a UD, which is why this article proposes a methodology with artificial vision

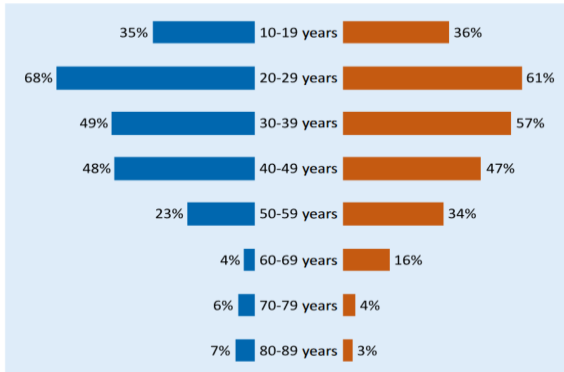

Fig. 1. Percentage of bodies with tattoos. The percentage of men (right) and women (left), grouped by age, had at least one tattoo—source: The use of tattoos to identify unknown bodies experiences from Jalisco, Mexico

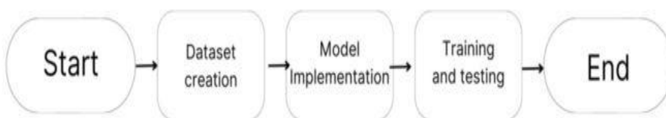

Fig. 2. Diagram of the methodology process

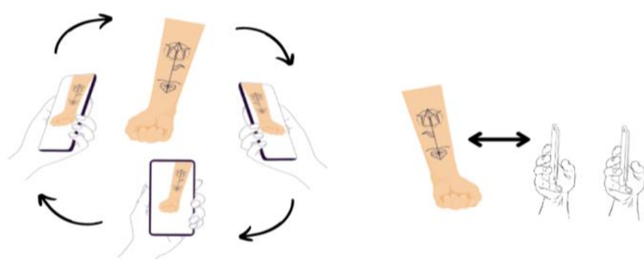

Fig. 3. Movements to capture a video

techniques to reduce the identification time of a UD.

3 Proposal for the Identification of a Body through Tattoos

The identification of people through tattoos can improve with an artificial neural network (ANN). ANNs are computational models based on biological neural networks made up of nodes or neurons, with inputs and outputs simultaneously connected to receiving information [9]. Among them is an image classification field because its main objective is to find behavior patterns from the input data [10].

This paper presents a methodology that implements a convolutional neural network model to identify UD by comparing tattoo images of missing persons with the UD tattoo image database. The methodology consists of three fundamental steps: the elaboration of the dataset (section 3.1), the implementation of the CNN model (section 3.2), and the training and testing of

Fig. 4. Tattoos contemplated for this investigation

Fig 5. Example of the original image, image without background, and image with the random background

the model (section 3.3). Figure 2 shows the steps of the proposed methodology.

A. Dataset

When the body of a UD arrives at the facilities of a forensic medical services unit, its particular features are identified, including tattoos. In order to use the tattoo classification model through images, the proposal is that a 30-second video is taken of each UD tattoo. Later, with a Python script, decompose the video into each frame. Each frame is saved as a different image of the tattoo.

In addition, for the images generated to be different, the videos are taken by moving the camera at different angles and approaching and retracting it from the tattoo. Figure 3 shows the camera's movements while taking a video of a tattoo.

The dataset is made up of 10 different tattoos voluntarily donated to research. Figure 4 shows the tattoos contemplated in this work.

For this research, a smartphone was used to capture the videos. For each 30-second video, 600 to 900 images were generated. The number of frames per video depends on the camera resolution of the device used to take the video.

In order to increase the number of different images for each tattoo in the data set, two tasks were carried out: 1) the background was removed from each image, leaving a .png image, and 2) a

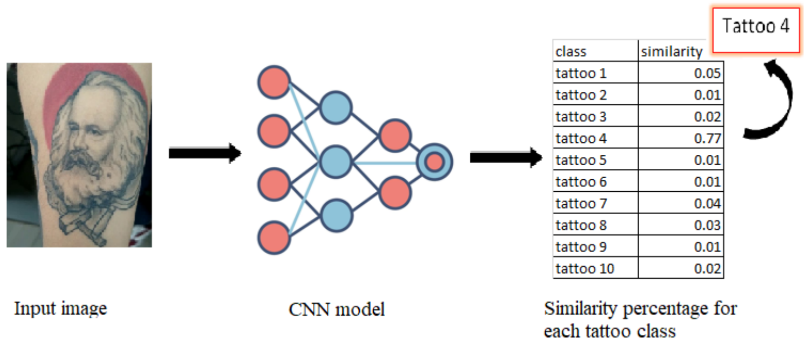

Fig. 6. Classification process

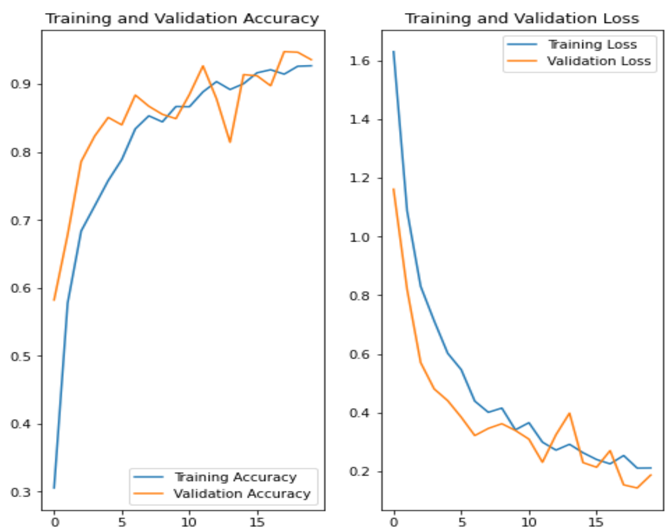

Fig. 7. Precision metrics for model training with 20 epochs

random background was added to the images without a background, from a landscape database.

A photo editing software script was used to remove and assign random backgrounds in bulk. Therefore, the data set contains 600 to 900 original images without background and random backgrounds for each of the ten collected tattoos. Figure 5 shows an original tattoo image, an image with no background, and an image with a random background.

B. Convolutional RNA Model

The model uses a Convolutional Neural Network (CNN) to identify UD through tattoo images for image classification. The model used is the MobileNet V2 created by Google ® and previously trained with 1001 classification classes [11].

Due to the use of a pre-trained image classification model, the hidden layers already have preloaded weights in charge of detecting lines, axes, curves, and positions, which improves the classification in the new training of the model (a process called transfer learning).

To use MobileNet V2 in classifying ten tattoos, a Dense layer for ten classes with Softmax activation was added to the model. By freezing the weights of the MobileNet V2 model, said layer allows us to train the model so that it assigns us a percentage of similarity corresponding to each tattoo in the database. Training that also uses the power of the original Mobile Net V2 model.

To identify a match between a missing person's tattoo and a UD person's tattoo, the model receives an input image and compares it to the classes (UD tattoos) that trained the model. Finally, the model returns an array with the similarity weights for each class. The most prominent similarity percentage represents the tattoo class with the most characteristics in common with the input image. Figure 6, shows the classification process.

The model inputs for training and classification require images with RGB values and a size of 224 x 224 pixels. Since this image size is uncommon in the UD and missing person databases, the images are resized at runtime.

Simultaneously, in each training, the model executes a data augmentation; for each image in the dataset, it creates variations of the same with transformations such as rotation, zoom, translation, and resizing in height and width.

The model executes python scripts to train the model and classify an image using the Keras and Tensorflow libraries.

C. Training and Testing

In training, the data set consisted of two parts: training and validation. The training had 80% of the data set, while the validation had 20%. After multiple trials and errors to find the correct number of training times, the result is that from 20 epochs, the model no longer improves its accuracy.

Therefore, with a training of 20 epochs, the model obtains a 96% accuracy. Figure 7 shows the accuracy of the algorithm over the training periods.

When performing the generalization tests, it was possible to observe that the algorithm classifies well where the conditions of the images were very similar to the training conditions of the algorithm (background colors, contrast, lighting, brightness, zoom, and resolution). Finally, in the cases where the classification failed, the percentage of similarity was found within the five highest.

4 Results and Discussion

The model correctly classifies tattoo images in similar conditions to those used in the training process. On the other hand, due to the limitations in the correct classification of images when the input differs significantly from the images used in the training process, its usefulness is reconsidered since it is unlikely that the input images to classify are similar to the images captured from a UD.

However, the model assigns each input image a percentage of similarity for each class in the data set, with which we could generate a top of classes with the highest match.

Although generating a top of matches does not identify in the first instance who the UD is, this could help reduce the search space and shorten the identification times. In this scenario, the most relevant would not be a correct classification by the model but a correct approximation to the classes with more similarity to the input image.

The limitations presented when classifying images very different from those of the data set show that the model continues to present a certain measure of overfitting.

5 Limitations

The main limitation is the number of images obtained for each tattoo. Typically, a classification system contains around 1000 different images per class. In the case of UD tattoos, it is impossible to get that many different images.

Even though video decomposition into images generates between 600 to 1000 images per tattoo, these are very similar. To eliminate classification problems such as overfitting caused by the small amount of information, an increase in data was generated in creating the dataset and in each runtime training.

Despite the data augmentation techniques, the model showed limitations when classifying tattoos, where the input image is very different in terms of background colors, contrast, lighting, brightness, and zoom.

6 Conclusion and Future Work

The artificial neural network was successfully created and implemented and allowed the identification of people through tattoos. A dataset was built through which the neural network was trained and tested, resulting in an accuracy of 96%.

In future work, the goal is to expand the tattoo database by taking actual data of UD from forensic institutes as a source. This is in order to evaluate the performance of the model by increasing the number of tattoos in a production scenario.

In addition, it is intended to implement a semantic segmentation technique before classification. This is with the aim of reducing the overfitting caused by the little information collected from each UD tattoo.

The possibility of extending the recognition of people through biometric data and dental records is also being sought.

References

1. **Franco-Migues, D. (2019).** Tecnologías de esperanza: Apropiaciones tecnopolíticas para la búsqueda de personas desaparecidas en México: El caso de las rastreadoras del fuerte. Comunicación y Sociedad, Vol. 16. DOI:10.32 870/cys.v2019i0.7280.
2. **Martinez, G. (2020).** Jalisco en primer lugar de desapariciones. Tráfico ZMG.
3. **Franco-Migues, D. (2019).** Gobierno de Jalisco lanza sitio de personas fallecidas no identificadas con sólo una quinta parte de los cuerpos que resguardan. ZonaDocs.
4. **Birngruber, C. G., Martinez-Peña, E. G., Corrales-Blanco, L., Holz, F. (2020).** The use of tattoos to identify unknown bodies. Rechtsmedizin, Vol. 30, No. 4, pp. 219–224. DOI: 10.1007/s00194-020-00396-y.
5. **Kumar, A., Harish, D., Singh, A., Kumar, G. A. (2014).** Unknown dead bodies: Problems and solutions. Journal of Indian Academy of Forensic Medicine, Vol. 36, No. 1, pp. 76–80.
6. **Cárdenas-Velarde, J. M., López-Sullaez, L., C. (2019).** Caracterización de cadáveres NN, que fueron ingresados a la morgue judicial, durante la gestión 2015: Protocolo de Identificación de Cadáveres NN. Repositorio Institucional Universidad Mayor de San Andrés, Tesis de Maestría.
7. **Fortuna, M., Corrales, L., Robinson, A., Farias, R. E., Marquez-Grant, N. (2022).** Cuerpos no identificados en el contexto mexicano. Forensic Anthropology, Vol. 5, No. 3. DOI:10.5744/fa.2022.4004b.
8. **Zimmermann, R. (2022).** Quemaduras químicas y carbonización: efectos sobre tatuajes y su incidencia en la identificación de cadáveres. Repositorio Institucional, Universidad Fasta.
9. **Montaño-Moreno, J. J. (2002).** Redes neuronales artificiales aplicadas al análisis de datos. Doctorado Psicología, Facultad de Psicología, Tesis doctoral de la Facultad de Psicología, Universitat de Les Illes Balears, Palma de Mallorca, España.
10. **Domínguez-Gómez, C. (2018).** FacesDetector: Aplicación práctica de machine learning sobre imágenes para un contexto de seguridad (Bachelor's thesis).

Article received on 01/11/2023; accepted on 17/03/2024.
**Corresponding author is Edgar G. Cossio.*