Subclass Contrastive Loss for Injured Face Recognition

Puspita Majumdar, Saheb Chhabra, Richa Singh, Mayank Vatsa
IIIT-Delhi, India
{pushpitam, sahebc, rsingh, mayank}@iiitd.ac.in

Abstract

Deaths and injuries are common in road accidents, violence, and natural disaster. In such cases, one of the main tasks of responders is to retrieve the identity of the victims to reunite families and ensure proper identification of deceased/injured individuals. Apart from this, identification of unidentified dead bodies due to violence and accidents is crucial for the police investigation. In the absence of identification cards, current practices for this task include DNA profiling and dental profiling. Face is one of the most commonly used and widely accepted biometric modalities for recognition. However, face recognition is challenging in the presence of facial injuries such as swelling, bruises, blood clots, laceration, and avulsion which affect the features used in recognition. In this paper, for the first time, we address the problem of injured face recognition and propose a novel Subclass Contrastive Loss (SCL) for this task. A novel database, termed as Injured Face (IF) database, is also created to instigate research in this direction. Experimental analysis shows that the proposed loss function surpasses existing algorithm for injured face recognition.

1. Introduction

Injuries on different body parts including face, head, limbs, and neck are common in road accidents, violence, and natural calamities. Among these, face is one of the most affected regions of the human body. As per the report of World Health Organization (WHO) [1], every year, 1.25 million people are killed and 50 million are injured in road accidents worldwide. According to Jordan and Calhoun [12], 50% to 70% of people surviving traffic accidents have facial injuries. In the majority of such cases, facial regions are partially or completely affected, and identification of victims in such scenarios becomes difficult. The problem becomes more challenging if the victim is in an unconscious state without any identity proofs. Apart from this, many times, natural calamities and violence result in unidentified dead bodies with facial injuries. Image of unidentified bodies with facial injuries, generally, are not accepted by print or electronic media for publication as they may cause anxiety and discomfort to the viewers [5]. In such situations, retrieving the identity of the deceased person becomes an arduous task.

Forensic experts and odontologists use a variety of techniques to ascertain the identity of injured victims and unidentified dead bodies such as DNA profiling and dental profiling [8,17,20]. However, these techniques are generally time-consuming, costly, and challenging to scale-up. In contrast, it is easy to obtain the face images from the victim, which can be matched with the reference images in

![Figure 1. Illustrating the impact of injured faces on existing face recognition models. (a) Sample images showing the change in appearance due to injuries (images are taken from Internet). (b) Performance of existing face recognition models for identifying injured faces at rank-1 (experimental details are available in Section 5.1).](image-url)
large (national level) databases. With this motivation, we propose to use automated face recognition for retrieving the identity of unidentified victims and present a solution for injured face recognition. Figure 1(a) shows sample images with injured faces. During injuries, facial features are affected due to swelling, bruises, mutilation, and laceration which in turn changes the facial structure. As shown in Figure 1(b), existing face recognition models trained on non-injured/normal faces do not perform well for injured face recognition. For matching injured faces with non-injured faces, a novel loss function, termed as Subclass Contrastive Loss (SCL), is proposed. The proposed loss function minimizes the distance between non-injured and injured sub-classes of the same subject while maximizes the distance from other subjects. Additionally, a first of its kind Injured Face (IF) database is created to evaluate the performance of the proposed loss function. Experiments on the IF database and comparison with existing algorithms showcase the efficacy of the proposed SCL.

2. Related Work

In the literature, facial injuries have been studied for various purposes such as identification of dead bodies, distinguishing injuries of domestic violence from accidents, and for studying the psychological effect of facial trauma. Facial features are heavily affected during injuries. This makes the identification of the victim a difficult task. Different techniques such as face reconstruction or sketch of the victim drawn by sketch artists are used to recover the identities of the victims. Singh et al. [19] evaluated the incidence of maxillofacial injury and concluded that road traffic accidents are the major cause of facial injury. It is found that facial injury is more frequent in male as compared to female. Black et al. [4] studied sports injuries and found that more than 41% of sports injuries require medical treatment. They found that in high-risk sports, eye protection, mouth guards, helmets, and face guard plays a key role in reducing facial injuries. However, these safety measures are not followed in most of the high-risk sports. Canzi et al. [6] proposed a measuring scale of the facial trauma that summarizes the surgical duration and care required for the treatment. This score can be used in trauma center decision making.

Iris recognition and dental profiling are some of the techniques for identifying dead bodies. In 2016, Trokielewicz et al. [20] performed postmortem iris recognition. They eval-
Let $\phi_W$ be a Convolutional Neural Network (CNN) with weights $W$ which outputs the feature representation of the input. The convolutional network $\phi_W$ is trained in the verification mode. For this purpose, two sets are created, namely genuine and imposter. The genuine set consists of two pairs of samples of the same subject. The first pair is created by taking a sample from the non-injured subclass and another sample from the injured subclass while the second pair is created by taking two different samples from the injured subclass. On the other hand, the imposter set consists of two pairs of samples of different subjects. The first pair contains samples from the non-injured and injured subclass while the second pair contains samples from the injured subclass. Mathematically, a genuine set and an imposter set are represented as:

$$ G = [\{x^N_{i,p}, x^I_{i,q}\}, \{x^I_{i,q}, x^I_{i,r}\}] $$

(2)

$$ M = [\{x^N_{i,p}, x^I_{j,q}\}, \{x^I_{j,q}, x^I_{j,r}\}] $$

(3)

where, $x^N_{i,p}$ is the $p^{th}$ sample of subclass $X^N_i$ of subject $i$, $x^I_{i,q}$ and $x^I_{i,r}$ are $q^{th}$ and $r^{th}$ samples of subclass $X^I_i$ of subject $i$, and $x^I_{j,q}$ is the $q^{th}$ sample of subclass $X^I_j$ of subject $j$.

The aim is to minimize the intra-class distance between the embeddings of two subclasses of the same subject and to maximize the inter-class distance from others. In order to achieve this, the network is trained by minimizing the following Subclass Contrastive Loss (SCL) function:

$$ \minimize \sum_{G,M} L_{G,M} $$

(4)

where, $L_{G,M}$ represents the loss for a genuine and imposter set. It is mathematically represented as:

$$ L_{G,M} = (1 - Y) L_{G}^{intra} + Y L_{M}^{inter} $$

(5)

where, $Y$ is a binary label with $Y = 0$ for the genuine set and $Y = 1$ for the imposter set. $L_{G}^{intra}$ and $L_{M}^{inter}$ represent the intra-class loss and inter-class loss, which is a distance function of genuine set and imposter set, respectively. For injured face recognition, there are two confounding factors while minimizing the intra-class distance and maximizing the inter-class distance: (i) variations due to the injured and non-injured images and (ii) variations among the injured images (i.e., different types of facial injuries). In order to address the first confounding factor, the distance between the injured and non-injured images of the same subject is reduced, and it is increased for different subjects. Similarly, the second factor is addressed by reducing the distance between the injured images of the same subject and by increasing the distance from others. For any genuine set, loss $L_{G}^{intra}$ is computed as:

$$ L_{G}^{intra} = D(\phi_W(x^N_{i,p}), \phi_W(x^I_{i,q})) $$

$$ + D(\phi_W(x^I_{i,q}), \phi_W(x^I_{i,r})) $$

(6)
For an imposter set, loss $L_{inter}^{M}$ is computed as:

$$L_{inter}^{M} = \max(0, \alpha_1 - D(\phi_{w}(x_{i,p}^N), \phi_{w}(x_{j,q}^I))) + \max(0, \alpha_2 - D(\phi_{w}(x_{j,q}^I), \phi_{w}(x_{i,r}^J)))$$ (7)

where, $D(\cdot)$ is any distance metric. $\alpha_1$ and $\alpha_2$ are the margins with value greater than 0. $\alpha_1$ controls the separation among the non-injured and injured images of different subjects while $\alpha_2$ controls the separation among the injured images of different subjects. In this research, squared Euclidean distance is used as $D(\cdot)$ and Equations 6 and 7 are updated as:

$$L_{intra}^{G} = \|\phi_{w}(x_{i,p}^N) - \phi_{w}(x_{i,q}^I)\|_2^2 + \|\phi_{w}(x_{i,q}^I) - \phi_{w}(x_{i,r}^J)\|_2^2$$ (8)

$$L_{inter}^{M} = \max(0, \alpha_1 - \|\phi_{w}(x_{i,p}^N) - \phi_{w}(x_{j,q}^I)\|_2^2) + \max(0, \alpha_2 - \|\phi_{w}(x_{j,q}^I) - \phi_{w}(x_{i,r}^J)\|_2^2)$$ (9)

where, $\|\cdot\|_2^2$ is the squared Euclidean distance.

**Face Recognition:** For recognition of faces during testing, features corresponding to the non-injured gallery images and injured probe images are extracted using the trained model $\phi_{w}$. Extracted features of the probe images are matched with the gallery images using Euclidean distance for identification.

### 4. Injured Face Database

To the best of our knowledge, there is no publicly available database for injured face recognition. Therefore, we have prepared the Injured Face (IF) database by collecting images from online resources. As part of research contributions, the database URLs (i.e., to these images) will be released to the research community. The details of the creation of the IF database and annotation of injured regions are discussed below:

#### 4.1. Creation of Injured Face Database

Injured Face (IF) database is created by collecting injured and non-injured/normal face images of different subjects from the Internet. Generally, non-injured/normal face images are used for enrollment. To recognize a subject with facial injuries, face recognition algorithms are required to match the injured face image with the enrolled non-injured images. Therefore, two types of images per subject are collected to create the database: i) faces with real injuries and ii) normal or non-injured face images. The database includes injured face images from sports, accidents, and violence cases. The database contains a total of 898 images corresponding to 100 subjects with 346 injured and 552 non-injured face images. Figure 3 shows some sample images of the IF database.

#### 4.2. Annotation of Injured Facial Regions

During facial injuries, different regions of a face are affected. To study the injuries present in different facial regions, injured faces of the IF database are annotated. For this purpose, seven different injured regions of the face are annotated. The regions correspond to forehead, left eye, right eye, left cheek, right cheek, nose, and mouth. Different images of each subject have injuries in one or multiple regions of the face. Table 1 summarizes the number of images with injuries in different facial regions.

#### 5. Experiments and Results

To evaluate the performance of existing deep face recognition models and the proposed algorithm with Subclass Contrastive Loss (SCL) on the IF database, multiple experiments are performed. The first experiment is performed for evaluating the performance of four existing pre-trained face recognition models namely, VGG-Face [16], OpenFace [3], VGGFace2 [7], and LightCNN [22] on the IF database. The second experiment is performed to evaluate the performance of the proposed algorithm with SCL.
on the IF database and compared with Contrastive Loss (CL) [11] and Triplet Loss (TL) [18]. Further, an extended gallery experiment is performed to evaluate the robustness of the proposed algorithm by adding images from the SC-face database [10] to the gallery of the testing set of the IF database. Table 2 presents the details of the experiments. Protocol to perform the experiments and implementation details are discussed below:

**Protocol:** Experiments are performed by dividing the IF database into training and testing sets with non-overlapping 70% subjects in the training set and 30% subjects in the testing set. Five times repeated random sub-sampling is performed for training and testing partitioning. Further, the training and testing sets are divided into gallery and probe, where the gallery contains non-injured/normal images and the probe contains injured images of each subject. Gallery and probe of the training set contain multiple images per subject. In the testing set, the gallery contains a single image per subject while probe contains multiple images per subject. In order to compare the proposed loss function with siamese network trained with Contrastive Loss (CL), pairs are generated using non-injured and injured face images.

**Implementation Details:** The proposed algorithm is implemented in Pytorch. All the experiments are performed on NVIDIA Tesla P100 server with 96GB RAM and 16GB GPU memory. LightCNN-9 network is used as the base network for the proposed algorithm. Initial two layers of the network are frozen, and the rest of the layers are trained by minimizing the proposed Subclass Contrastive Loss (SCL). Adam optimizer is used with a learning rate of 0.0000003. The network is trained for 30 epochs with a batch size of 50. During the experiments, $\alpha_1$ and $\alpha_2$ are set to 2 and 3.1 respectively. LightCNN-9 network with same experimental settings and a margin of 2 is used to train the siamese network with Contrastive Loss (CL) for comparison. LightCNN-9 is also used as the base network for triplet training and the margin is set to 0.4 during the experiments.

Table 3 shows the mean classification accuracy with standard deviation at rank 1, 5, and 10. Figure 4 shows the Cumulative Match Characteristic (CMC) curve. It is observed that existing face recognition models do not perform well for the task of injured face image recognition. Among the pre-trained models, LightCNN-9 performs best with an average classification accuracy of 25.43% at rank 1.

### 5.1. Injured Face Recognition using Pre-trained Models

In order to evaluate the performance of pre-trained models on the IF database, four existing pre-trained deep face recognition models, namely, VGG-Face, OpenFace, VGGFace2, and LightCNN-9 are used. Features are extracted using the pre-trained models from the gallery and probe of the testing set. Extracted features are used to match the probe images with the gallery images using Euclidean distance. Table 5 shows the mean classification accuracy with standard deviation of five times random subsampling based cross-validation at rank 1, rank 5, and rank 10. Figure 4 shows the Cumulative Match Characteristic (CMC) curve. It is observed that existing face recognition models do not perform well for the task of injured face image recognition. Among the pre-trained models, LightCNN-9 performs best with an average classification accuracy of 25.43% at rank 1.

### 5.2. Injured Face Recognition using Subclass Contrastive Loss

This experiment is performed to evaluate the performance of the proposed algorithm with SCL, and the results are compared with CL (Contrastive Loss) and TL (Triplet Loss). Among the baseline models, LightCNN-9 performs the best. Therefore, it is used as the base network to train the proposed SCL, CL, and TL. Table 4 shows the mean classification accuracy with standard deviation at rank 1, 5,
and 10 obtained from these algorithms. It is observed that SCL outperforms both CL and TL. For instance, SCL yields 8.03% and 6.44% higher rank-1 accuracies than CL and TL, respectively. The low performance of CL and TL indicates the limitation of existing networks in handling large intra-class variations among the injured faces of the same subject and small inter-class distance among different subjects. This limitation is overcome by the proposed algorithm with SCL as these variations are explicitly encoded in the proposed loss function. Table 5 shows the comparison of the mean inter-class distance of gallery image of each subject with probe images of other subjects using pre-trained LCNN-9, CL, TL, and SCL. It is observed that training with SCL increases the inter-class separation, which in turn results in the improved performance. Figure 5 shows some samples misclassified by the CL based approach but they are correctly classified by the proposed SCL based model.

Verification performance is also evaluated on the IF database using CL, TL, and SCL. For this purpose, 50 genuine and 50 imposter pairs are generated. It is observed that the proposed SCL performs better than CL and TL for the task of verification. For instance, at 0.1 False Accept Rate (FAR), the Genuine Accept Rate (GAR) is 74.00%, 76.00%, and 78.00% corresponding to CL, TL, and SCL. Figure 6 shows the genuine and imposter score distribution using pre-trained LightCNN-9 and SCL, respectively. It is observed that the proposed SCL reduces the overlap between the two distributions, which in turn results in the improved performance.

5.3. Extended Gallery Experiment

The proposed algorithm with SCL is also evaluated for extended gallery experiment. In face recognition, extended gallery experiment is performed to emulate the real-world scenario of matching the probe images with a large gallery [14]. Therefore, an additional 100 subjects with a single image per subject from the SCface dataset are added to the gallery of the testing set to evaluate the performance of the proposed SCL. The SCface dataset contains 4160 static images of 130 subjects in the visible and infrared spectrum.

Verification performance is also evaluated on the IF database using CL, TL, and SCL. For this purpose, 50 genuine and 50 imposter pairs are generated. It is observed that the proposed SCL performs better than CL and TL for the task of verification. For instance, at 0.1 False Accept Rate (FAR), the Genuine Accept Rate (GAR) is 74.00%, 76.00%, and 78.00% corresponding to CL, TL, and SCL. Figure 6 shows the genuine and imposter score distribution using pre-trained LightCNN-9 and SCL, respectively. It is observed that the proposed SCL reduces the overlap between the two distributions, which in turn results in the improved performance.

Table 6. Results of the extended gallery experiment. ‘With EG’ stands for with extended gallery experiment and ‘Without EG’ stands for without extended gallery experiment.

|         | Rank 1     | Rank 5     | Rank 10   |
|---------|------------|------------|-----------|
| With EG | 36.13 ± 3.8| 65.91 ± 6.0| 76.62 ± 4.5|
| Without EG | 36.70 ± 3.6| 65.91 ± 6.0| 82.4 ± 6.5|

For the experimental purpose, images of the visible spectrum are used. The model trained on the training set of the IF database is used for evaluation. Table 6 shows the comparison of the classification accuracy obtained with and without an extended gallery. It is observed that the proposed algorithm with SCL performs equally well in extended gallery experiment.

6. Conclusion

Facial injuries are common in road accidents, violence, and sports. Facial injuries change the physical structure of the face and damage soft tissues. Several cases of unidentified bodies with facial injuries and unconscious victims of road accidents require timely identification. Therefore, in this research, the problem of injured face recognition is studied. For this purpose, a novel Subclass Contrastive Loss (SCL) is proposed for recognition of injured faces. Additionally, first of its kind Injured Face (IF) database is created and baseline performance of existing deep face recognition models is evaluated on the IF database. Multiple experiments showcase the efficacy of the proposed SCL algorithm for injured face recognition.

Acknowledgements

This research is partly supported by the Infosys Center of Artificial Intelligence, IIIT Delhi, India. M. Vatsa is also partially supported by the Department of Science and Technology, Government of India through Swarnajayanti Fellowship. The authors acknowledge Maneet Singh and Rohit Keshari for their constructive and useful feedback.

References

[1] WHO statistics. [http://tinyurl.com/y6dz83t4](http://tinyurl.com/y6dz83t4)
[2] A. Abaza, A. Ross, and H. Ammar. Retrieving dental radiographs for post-mortem identification. In IEEE International Conference on Image Processing, pages 2537–2540, 2009.

[3] B. Amos, B. Ludwiczuk, J. Harkes, P. Pillai, K. Elgazzar, and M. Satyanarayanan. Openface: Face recognition with deep neural networks. In IEEE Winter Conference on Applications of Computer Vision, 2016.

[4] A. M. Black, D. A. Patton, P. H. Eliason, and C. A. Emery. Prevention of sport-related facial injuries. Clinics in Sports Medicine, 36(2):257–278, 2017.

[5] P. Bodkha and B. Yadav. A role of digital imaging in identification of unidentified bodies. Journal of Indian Academy of Forensic Medicine, 34(4):0971–0973, 2012.

[6] G. Canzi, E. De Ponti, G. Novelli, F. Mazzoleni, O. Chiara, A. Bozzetti, and D. Sozzi. The cfi score: Validation of a new comprehensive severity scoring system for facial injuries. Journal of Cranio-Maxillofacial Surgery, 47(3):377–382, 2019.

[7] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman. Vggface2: A dataset for recognising faces across pose and age. In IEEE International Conference on Automatic Face & Gesture Recognition, pages 67–74, 2018.

[8] T. Clayton, J. Whitaker, and C. Maguire. Identification of bodies from the scene of a mass disaster using dna amplification of short tandem repeat (str) loci. Forensic Science International, 76(1):7–15, 1995.

[9] T. A. Coulibaly, R. Béogo, I. Traoré, H. M. Kohoun, and B. V. IlÉ. Inter personal violence-related facial injuries: a 10-year survey. Journal of Oral Medicine and Oral Surgery, 24(1):2–5, 2018.

[10] M. Grgic, K. Delac, and S. Grgic. Scface–surveillance cameras face database. Multimedia tools and applications, 51(3):863–879, 2011.

[11] R. Hadsell, S. Chopra, and Y. LeCun. Dimensionality reduction by learning an invariant mapping. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, volume 2, pages 1735–1742, 2006.

[12] J. Jordan and K. Calhoun. Management of soft tissue trauma and auricular trauma. Bailey BJ, Johnson JT; Newlands SD. Head & Neck Surgery: Otolaryngology. Hagerstwon, MD: Lippincott Williams & Wilkins, pages 935–36, 2006.

[13] P. Majumdar, S. Chhabra, R. Singh, and M. Vatsa. On detecting domestic abuse via faces. In IEEE Conference on Computer Vision and Pattern Recognition Workshops, pages 2173–2179, 2018.

[14] S. Nagpal, M. Singh, R. Singh, M. Vatsa, A. Noore, and A. Majumdar. Face sketch matching via coupled deep transform learning. In Proceedings of the IEEE International Conference on Computer Vision, pages 5419–5428, 2017.

[15] D. E. Nassar, A. Abaza, X. Li, and H. Ammar. Automatic construction of dental charts for postmortem identification. IEEE Transactions on Information Forensics and Security, 3(2):234–246, 2008.

[16] O. M. Parkhi, A. Vedaldi, A. Zisserman, et al. Deep face recognition. In British Machine Vision Conference, volume 1, page 6, 2015.

[17] K. Quinet, S. Nunn, and A. Ballew. Who are the unclaimed dead? Journal of Forensic Sciences, 61:S131–S139, 2016.

[18] F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: A unified embedding for face recognition and clustering. In IEEE Conference on Computer Vision and Pattern Recognition, pages 815–823, 2015.

[19] V. Singh, L. Malkunje, S. Mohammad, N. Singh, S. Dhasmana, and S. K. Das. The maxillofacial injuries: A study. National Journal of Maxillofacial Surgery, 3(2):166, 2012.

[20] M. Trokielewicz, A. Czajka, and P. Maciejewicz. Post-mortem human iris recognition. In IEEE International Conference on Biometrics, pages 1–6, 2016.

[21] M. Trokielewicz, A. Czajka, and P. Maciejewicz. Iris recognition after death. IEEE Transactions on Information Forensics and Security, 14(6):1501–1514, 2019.

[22] X. Wu, R. He, Z. Sun, and T. Tan. A light cnn for deep face representation with noisy labels. IEEE Transactions on Information Forensics and Security, 13(11):2884–2896, 2018.