

Deep learning approach for forensic facial reconstruction depends on unidentified skull

Doaa M. Mohammed, Mostafa Elgendy, Mohamed Taha

Department of Computer Science, Faculty of Computers and Artificial Intelligence, Benha University, Benha, Egypt

Article Info

Article history:

Received Oct 11, 2023

Revised Feb 24, 2024

Accepted Mar 21, 2024

Keywords:

Facial approximation
Facial reconstruction
Forensic deep learning
Unidentified skulls
Victim's face
Victim's identification

ABSTRACT

Facial reconstruction, or facial approximation, is an essential problem in a criminal investigation involving reconstructing a victim's face from his skull to determine the victim's identification at a crime scene. Facial approximation plays a crucial part when there is a lack of clues with investigators. Investigators utilize facial approximation to guess the victims' identities. This research attempted to use computer-aided face reconstruction rather than traditional approaches. Traditional methods of face reconstruction include the use of clay or gypsum. Traditional procedures necessitate forensic professionals to rebuild the victim's face. This research uses the convolution neural network skull part with sift (CNNSPS) model is employed to reconstruct facial features from a skull image utilizing public datasets CelebAMask-HQ and MUG500+. The proposed algorithm was tested on unidentified skull databases, and celebrity faces were used. The genuine datasets are not available, which is the key issue in this research.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Doaa M. Mohammed

Department of Computer Science, College of Computers and Artificial Intelligence, Benha University

Fareed Nada Street, Banha, Al-Qalyubia Governorate 13511, Egypt

Email: doaa.ibrahem@fci.bu.edu.eg

1. INTRODUCTION

The traditional approaches for human identification are ineffective in some forensic investigations. The police may have just a few clues to identify the individual, DNA, fingerprints, and dental records are the most common identification techniques [1], [2]. The identifying techniques need a known individual with whom to compare evidence. It is challenging to match data with records from an entire community if there are no suspects for identification. In other circumstances, authorities may use a different technique. Forensic facial reconstruction technology is used to aid in identifying human remain [3]–[5].

Forensic face reconstruction or forensic face approximation is a technique used to reconstruct a person's facial features from bone remains. Physical data is used to reconstruct faces from given images or videos. Facial reconstruction is significant in a variety of fields. Facial reconstruction is used in forensics, archaeology, and anthropology to create a sculpted model of an unknown face [6]. Unidentified human remains can be identified through facial reconstruction due to the lack of clues and the global spread of murder cases. The police use face reconstruction to identify human remains at crime scenes [7].

According to world population review, the comprehensive rate of crime is calculated by the overall crimes reported across various categories divided by the overall population, then multiplied by 100,000. There is a distinct association between crime and age, as the majority of offenses, severe ones, are perpetrated by individuals in the age range of 20 to 30 years old. Venezuela holds the top spot for the highest index of crime globally at 83.76, then Papua New Guinea at 80.79 on the crime index, and South Africa ranks as the third-highest crime rate worldwide [8].

According to the report, homicide rates can be influenced by various socioeconomic factors. These reasons include socioeconomic inequality, unemployment, political instability, gun possession, and in particular, gangs, organized crime, and the drug trade. Indeed, "Approximately 65,000 killings per year were associated with organized gangs, and crime between 2000 and 2017, constituting approximately 19% of all recorded homicides worldwide in 2017." As shown in Figures 1 and 2 [8], [9].

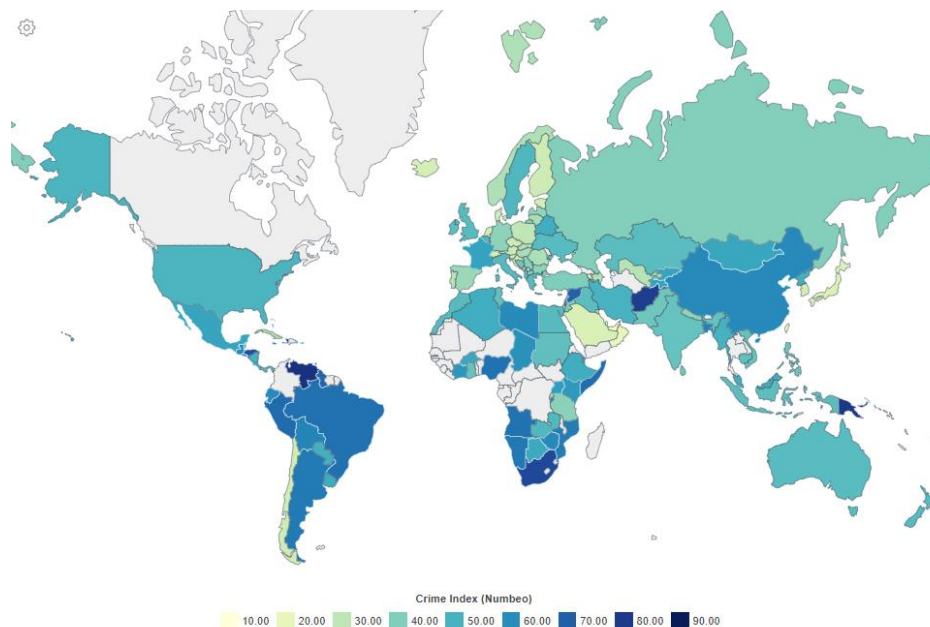


Figure 1. Statistics of crime rate in 2024 [8]

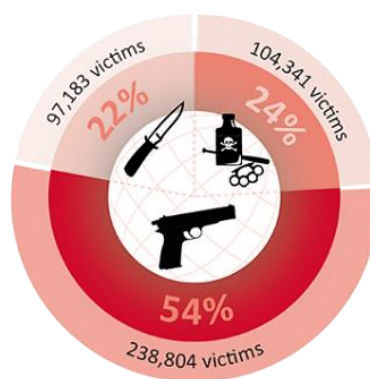


Figure 2. Statistics on the number of victims in 2019 [9]

According to world crime rate and statistics 2000-2024, the increase in 2020 was 5.61 per 100,000 population, equivalent to a rise of 0.74% in annual change from 2019. The statistics in 2019 were 5.56 per 100,000 population, equal to a 3.65% decline in yearly change from 2018. The statistics that existed in 2018 were 5.77 per 100,000 population, equivalent to a 2.24% decline from 2017. The statistics that existed in 2017 were 5.91 per 100,000 population, equal to a 0.69% decline from 2016, as shown in Figure 3 [10].

Previous research used deep learning techniques and computer vision to reconstruct the face of a human from a skull, such as such as object detection, recognition, image segmentation, multi-modal image translation, and 3D reconstruction [11]–[18]. This research proposes a convolution neural network skull part with sift (CNNSPS) model to reconstruct human faces from human remains of unidentified skulls. The same person's skull and skin face datasets have an obstacle due to datasets being unavailable. To overcome this issue, we used two public datasets MUG500+ [19] and CelebAMaskHQ [20]. The proposed algorithm applies some preprocessing on the skull dataset to enhance the skull image. Then, the face dataset and skull dataset are

segmented into parts to be ready for the CNNSPS model. This model is used to map each part of the skull to be equivalent to the face part. The remaining sections of the research paper are organized in the following manner: section 2 related work, section 3 methodology, section 4 experimental and result, and section 5 conclusion.

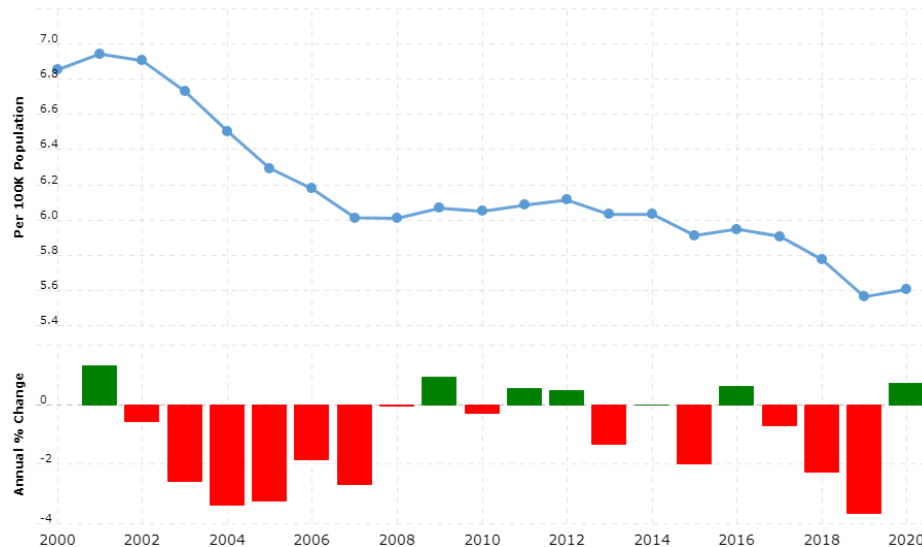


Figure 3. World crime rate and statistics 2000-2023 [10]

2. RELATED WORK

Short overview of some previous studies of facial approximation using some techniques in computer vision and deep learning to reconstruct the face. Knyaz *et al.* [12] presented deep learning techniques showcase cutting-edge outcomes from contemporary deep convolutional neural network (CNN) models across diverse computer vision tasks including object detection and recognition, image segmentation, multi-modal image translation, and 3D reconstruction. The generative adversarial approach in deep learning enables exceptional performance in tasks such as image-to-image translation, image segmentation, and the 3D reconstruction structures from a single photo. This paper presented the skull2face model using generator G and discriminator D of the gan model. The accuracy of the facial approximation was assessed by measuring the surface-to-surface deviation between the reconstructed and the faces real. The results indicate an accuracy of under 2.5 mm for approximately 60% of the reconstructed facial surface [12].

Kniaz *et al.* [13] proposed a deep learning framework is utilized to create and showcase realistic textures for the 3D reconstruction of faces belonging to prehistoric humans. Generative adversarial network (GAN) is a new neural network type. GANs are comprised of two deep CNN: discriminator and generator. A CycleGAN is used to overcome the difficulty of Gan's image-to-image translation tasks. A new model of StyleGAN is implemented to enhance the perceptual realism and quality of the reconstructed image to improve the performance. The issues present in the initial StyleGAN model were partially resolved in the subsequent StyleGANv2 model. For training the developed skull2photo framework, a distinct CelebAMask-HQ and crania-to-facial dataset (C2F) were specifically generated. The outcomes of the experiments reveal that the enhanced generator G1 contributes to an 11% improvement in the quality of 3D reconstruction [13].

Bushra and Maheswari [14] presented an innovative artificial intelligence method to digitally identify a criminal through facial recognition digitally, through forensic sketch is transformed into a real photo using a deep convolutional generative adversarial network (DCGAN). Advanced GAN involves image processing through a CNN known as the DCGAN. Extended difference-of-gaussians (XDoG) algorithm segregates sketches from the clipped real pictures. The datasets employed include the celebrity face attributes (CelebA). The cycle-generative adversarial network (CycleGAN) is trained on a real facial recognition technology (FERET) dataset, consisting of 1,194 sketch-photo pairs of hand-drawn sketch images. Conditional CycleGAN is associated with CycleGAN, achieving an accuracy of 61.34% for CycleGAN and 65.53% for ccGAN [14].

Shui *et al.* [15] presented manual 2D and 3D approaches for forensic purposes and archaeology. The result of the computerized code of federal regulations (CFR) is contingent on the excellence of the skull digitalized, with the acquisition of a digitized representation achieved through the utilization of computed

tomography (CT) scanning and laser scanning techniques. In a previous study the geometric deviation of models obtained through laser scanning and CT scanning techniques was compared, indicating that both approaches can generate a digital skull of high quality. Classifiers or discriminant equations are provided for sex prediction, in existing research has adopted this to realize sex classification. Support vector machines (SVM) are employed to build a classifier for predicting sex based on PC scores and centroid size. The research enlisted 140 living individuals, comprising 70 females and 70 males of Han Chinese nationality to establish the dataset of the face and skull dataset [15].

Tan *et al.* [16] proposed the recognition of human remains through craniofacial superimposition (CS) as a notable focal point within the field of forensic sciences. The research introduced a (3D-3D) superimposition technique for the model of a 3D skull and 3D face to overcome the challenges mentioned earlier. The analytical curvature B-spline (AC B-spline) begins by reconstructing a 3D face model from a provided 2D face image, by using the mean simplified generic elastic model. The face model is aligned with a 3D skull on the jawline through the application of the AC B-spline. The z-coordinates of the 2D face image are obtained by reconstructing it into a 3D face model using a mean-simplified generic elastic model (MS-GEM). MS-GEM is an enhanced version of S-GEM, amalgamating the face models created through S-GEM to augment facial features more effectively. the AC B-spline achieves an accuracy is 0.803 and the PCA-TR method exhibits an accuracy of 83.5% [16].

Afzal *et al.* [17] proposed a method for reconstructing 3D faces. It takes an input of a single 2D image and yields 3D images reconstructed as the output. The brute force matching (BFM) involves three primary stages: extracting features, determining depth, and generating a 3D image from the processed image through the BFM. Subsequently, scale-invariant feature transform (SIFT) was employed in the reconstruction of 3D faces [17].

Navic *et al.* [18] proposed reproducing a virtual face, with facial soft tissue thickness identified as a key criterion for ensuring reliability and accuracy in 3D computerized facial reconstruction. the primary tools in 3D modeling software and computer animation blender, autodesk, and maya were employed to reconstruct the 3D virtual face. Facial reconstruction in 3D computerization typically relies on a semi-automated method, such as 3D modeling softwares, such as Blender, Zbrush, Freeform, and 3DStudio Max. This approach still incorporates conventional manual techniques and requires strong understanding of anthropology and anatomy, or an automated approach used to reconstruction involves advanced computer-based techniques and craniofacial data to generate facial reconstruction [18].

3. METHOD

Face reconstruction is used to identify the individual from their skull. Face reconstruction is used in forensics, archaeology, and anthropology. We propose a technique of a CNNSPS model. The main procedures in this research will be explained, such as preprocessing, cropping photographs, and the model layers employed, followed by the output image. The main procedures are shown in Figure 4.

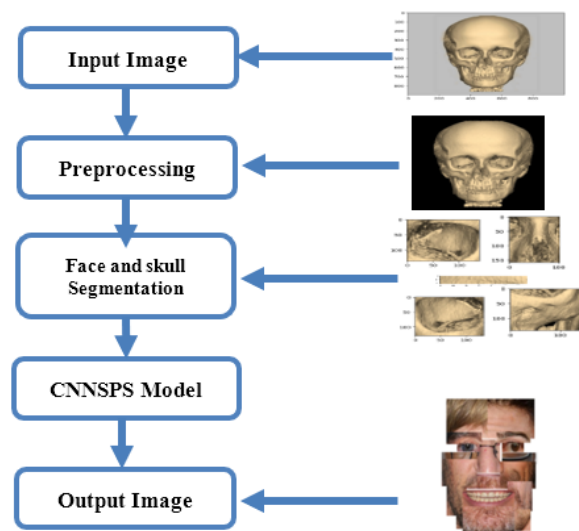


Figure 4. The proposed CNNSPS model

3.1. Preprocessing

This step converts the input image to grayscale and applies some necessary image enhancement, such as noise removal using `fastNlMeansDenoisingColored` in OpenCV. `fastNlMeansDenoisingColored` uses the non-local means denoising algorithm. It compares pixel neighborhoods in the input image to a template neighborhood, identifying and reducing grayscale and color noise. The segmentation technique isolates the skull region from the background using the GrabCut algorithm that applies to iteratively update the mask based on color and spatial information. Feature extraction is used to identify landmarks on the skull bone after detecting the skull region using the SIFT.

The prominent landmarks detected on the skull image. The MUG500+ dataset [19] for the skull is used in the preprocessing step. are: the anterior nasal spine (AntNS), Point A(A), left supra-orbital incisure (LsupO), left inferior lateral orbital rim (LORbi), left gonion (LGoni), left porion (LPori), right supra-orbital incisure (RsupO), right inferior lateral orbital rim (RORbi), right gonion (RGoni), right porion (RPori), and Menton (Mento) as shown in Figure 5.

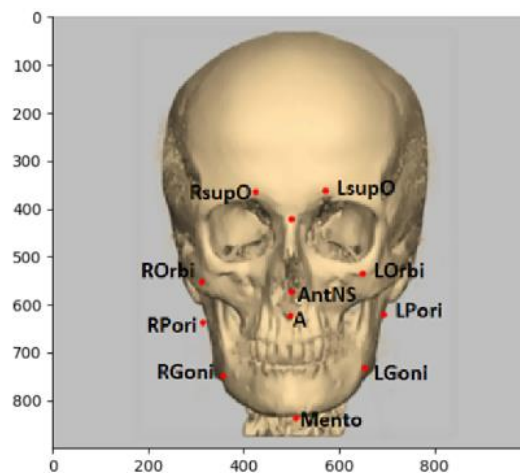


Figure 5. The prominent landmarks on the skull image

3.2. Face and skull segmentation

This stage involves segmenting the skull image into parts like the right and left eyes, right and left cheek, and brow. These segmented parts are then fed into the CNNSPS model for further processing. Detect the face landmarks of the CelebAMask-Hq dataset in [20] consists of 30,000 face images of famous people to detect the 68 points of landmarks, as shown in Figure 6, that are used to crop the face into 10 parts, including right and left eyes, right and left cheek, and brow, as shown in Figures 7(a) and 7(b). Then, training model to get the most part matches with the part of the face with the skull.

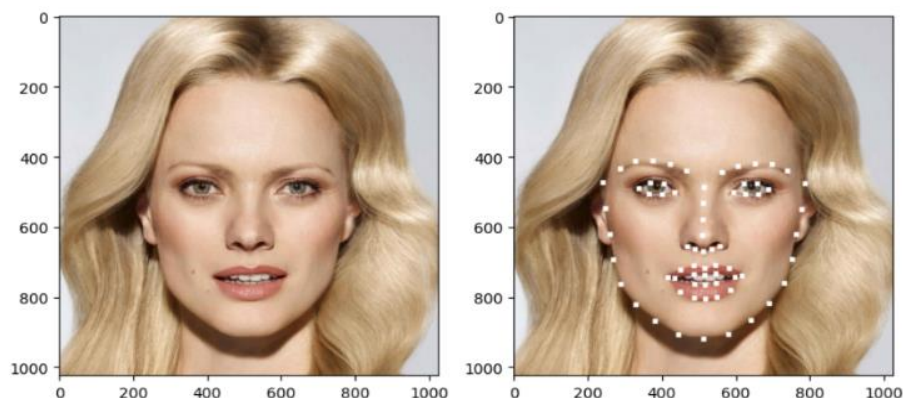


Figure 6. Landmarks of the face

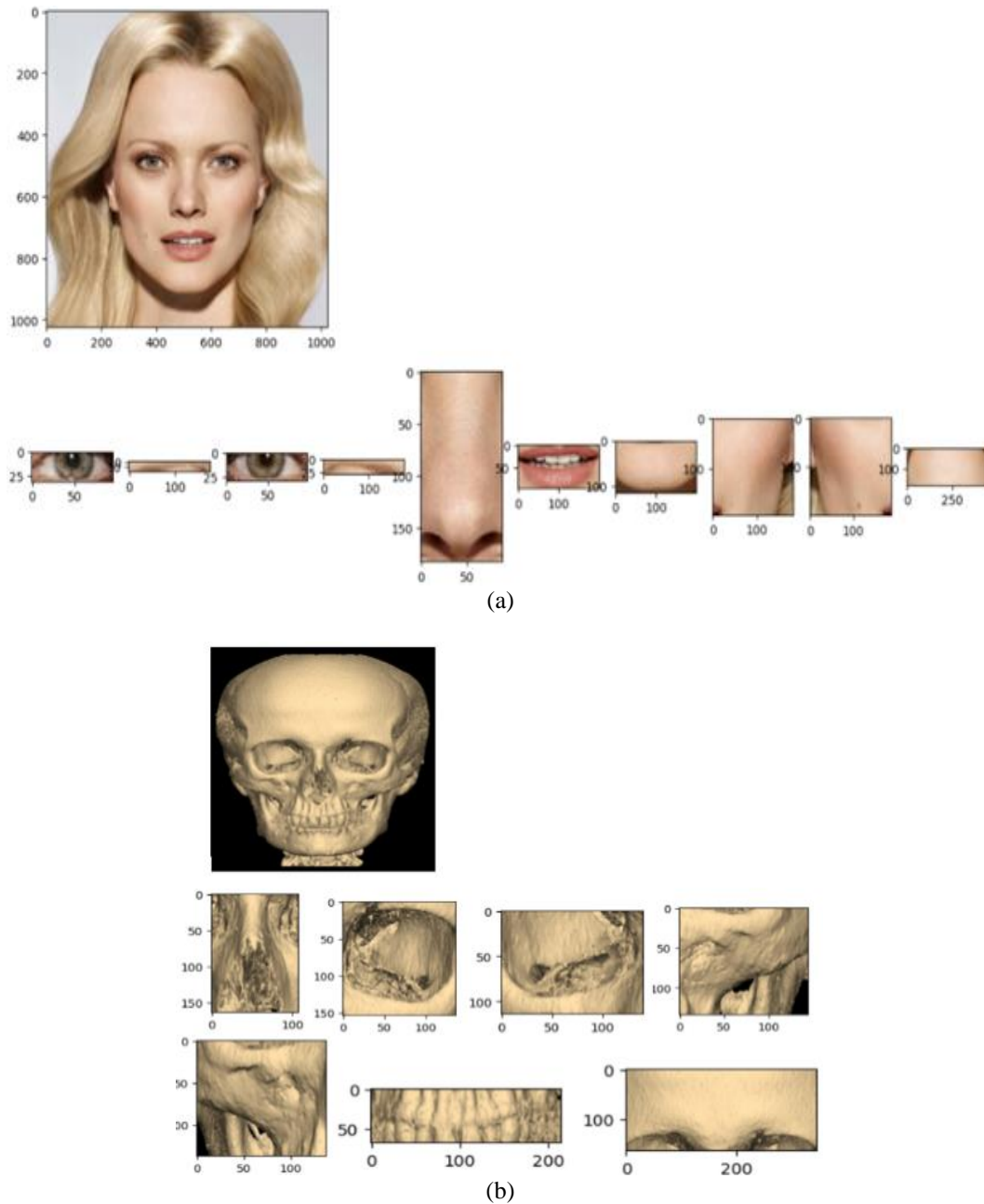


Figure 7. The segmentation of the skulls and faces (a) the segmented parts of the face and (b) the segmented parts of the skull

3.3. Convolution neural network skull part with sift model

Use the output of the segmentation of the face and skull as input to the model. Detect the primary skull landmarks of the image. CNN (also ConvNet) is a deep learning-based neural network with an inherent machine learning technique. CNN processes data, such as images, with a grid pattern. The model architecture comprises five convolutional layers with a kernel size of 3×3 , succeeded by batch normalization, rectified linear unit (ReLU) activation, and max pooling operations. The first convolutional layer comprises three input channels (which represent RGB colors) and 16 output channels. The subsequent layers follow a similar pattern, with 32, 64, 128, and 224 output channels. Batch normalization is used after each convolutional layer to increase convergence and training stability. ReLU activation functions add nonlinearity to the model, helping it detect complicated data patterns. Max-pooling operations minimize spatial dimensions, allowing the model to focus on essential features while removing duplicate data as shown in Figure 8.

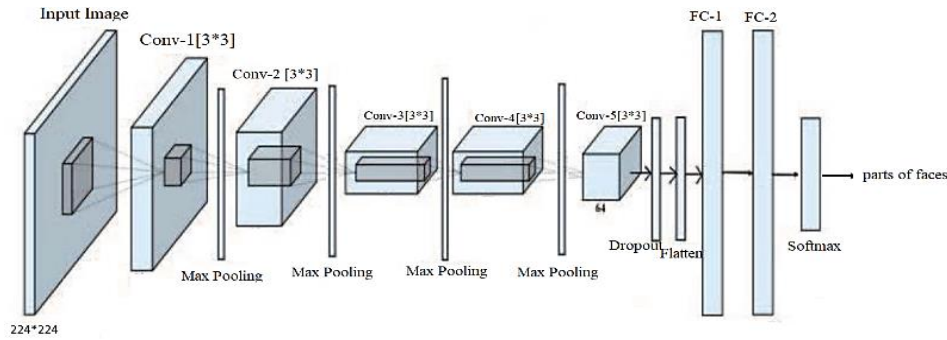


Figure 8. CNNSPS architecture model

3.3.1. Convolution layer

A convolution layer is a key component of the CNN and is responsible for extracting features. It is used to extract essential information. The convolutional layer computes the convolutional operation of the input pictures using filters. The kernel filters have the same dimension as the input image but have lower constant parameters [16]. The convolution layer is calculated in (1) [21]:

$$w_{out} = \left(\frac{w_{in} - ks + 2 * pd}{st} \right) + 1 \quad (1)$$

where w_{in} is an input size, ks is kernel size, pd is padding, and st is stride.

3.3.2. Pooling layer

A pooling layer conducts a standard downsampling operation, decreasing the feature maps' spatial dimensionality. This is done to introduce translation invariance, making the network less sensitive to slight shifts and distortions and decreasing the number of learnable parameters. Max pooling is the most common procedure, taking patches from the input feature maps. It outputs the most significant value in each patch and discards all other values [16]. The equation for max pooling is calculated in (2) [22] and (3) [23]:

$$f_{max}(X) = \max_i x_i \quad (2)$$

J related filters are used for the composition of m th max pooling band $pm = [p1, m, \dots, pj, m, \dots, pj, m] \in RJ$:

$$pj, m = \max(h_j(m-1)N+r) \quad (3)$$

Where $N \in \{1, \dots, R\}$ is termed as a pooling shift that allows enabling overlap within concerning pooling regions when $N < R$.

3.3.3. Fully connected layer

The output feature maps from the last convolution or pooling layer are usually flattened. The fully connected layer is converted into a one-dimensional (1D) array or vector of values. Then, it is linked to one or more fully connected layers, often called dense layers. In these dense layers, every input is connected to every output through adjustable weights. After using CNN, the SIFT algorithm detects distinctive key points or features in an image [24].

3.3.4. Scale invariant feature transform

The SIFT is utilized to extract features that reduces visual information to a set of points that can be used to find similar patterns in other images. This approach is commonly used in computer vision applications like picture matching and object detection. SIFT has many advantages, such as locality, distinctiveness, quantity, efficiency, and extensibility [25]. The steps of SIFT execution are shown in Figure 9.



Figure 9. SIFT algorithm steps

3.3.5. Building the scale-space

The 'Gaussian Blur' approach distinguishes unique spots while disregarding noise. The output image of scale-space is a function $L(x,y,\sigma)$ [16]. The blurred image is calculated in (4) [21], [25]:

$$L(x,y,\sigma) = I(x,y) * G(x,y,\sigma) \quad (4)$$

Gaussian blur operator is calculated from (5) [21], [26]:

$$G(x,y,\sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \quad (5)$$

Where x and y are the location coordinates, σ is the "scale" parameter, and G is the Gaussian blur operator. Then, the difference of Gaussians (DoG) is used to locate relevant key points in an image [21], [27].

$$D(x,y,\sigma) = -L(x,y,\sigma) + L(x,y,k\sigma) \quad (6)$$

Where σ , $k\sigma$ are the scales of images.

3.3.6. Key point localization

Key point localization is used to find the key points in the image. The low-contrast features and those placed on the edges are removed. The technique determines the precise positions of the key points by interpolating neighbor data points for each key point. It can be achieved by employing the Taylor expansion of the DoG image [28].

3.3.7. Orientation assignment

Orientation assignment is used to ensure image rotation invariance. Each key point is assigned an orientation. Depending on the size, a neighborhood is drawn around the keypoint location, and the gradient magnitude and direction are determined there [22], calculated in (7) and (8) [29].

$$Magnitude = \sqrt{G_a^2 + G_b^2} \quad (7)$$

$$Orientation = \arctan(G_b/G_a) \quad (8)$$

Where G_a and G_b represent the gradient in the a and b direction respectively for the image.

3.3.8. Keypoint descriptor and keypoint matching

First, to build a keypoint descriptor, compute the gradient magnitude and direction for each image sample point in an area surrounding the keypoint position. These are weighted by a Gaussian window, represented by the overlaid circle, as shown in Figure 10 [30]. Key point matching is derived from the key point descriptor.

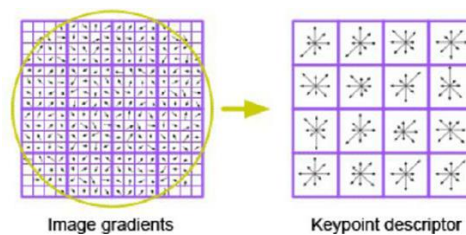


Figure 10. Example of the SIFT descriptor [26]

3.4. Skull-to-face mapping

Skull-to-face mapping proposes a matched image that entails superimposing a photograph of a person's face onto a 2D model of their skull. This procedure enables forensic anthropologists, forensic artists, and plastic surgeons to match the skull's features directly to those of the individual while they are alive. The first step is using images in the CelebAmaskHq available dataset. The second step is using the skull images in the MUG500+ available dataset. Third, match the photos to the 2D model of the skull. This entails resizing and aligning the photograph to match the size and position of the skull. The fourth step is identifying major anatomical landmarks in photographs and the skull. These landmarks may include the eyes, nose, mouth, chin, and other locations along

the face's contours. The final step is to project the features from the image onto the appropriate regions on the skull (the image on the left depicts the skull, while the image on the right illustrates the face) as shown in Figure 11.

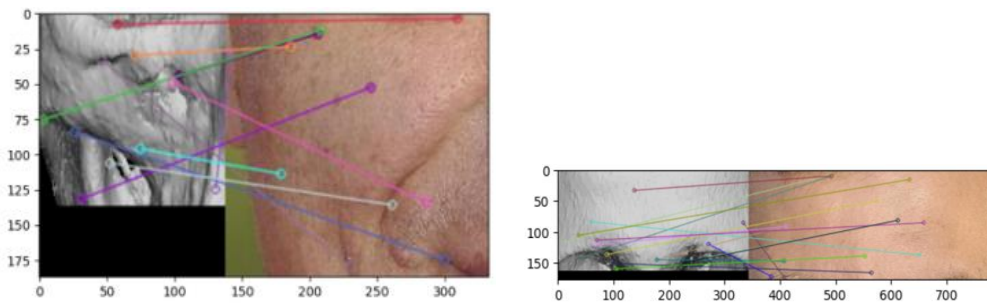


Figure 11. Skull-to-face mapping

4. RESULTS AND DISCUSSION

We train the model on Google Colab. The system's computational backbone is a Google Colab TPU runtime. Featuring an Intel Xeon CPU, 13 GB RAM, and a powerful cloud TPU. Its software stack includes Python 3.10. We used OpenCV for advanced computer vision tasks and leveraged the capabilities of PyTorch for cutting-edge deep learning applications.

4.1. Datasets

The proposed research was tested on the MUG500+ [31] and CelebAMask-HQ datasets. MUG500+ is a collection of 500 healthy skulls, that are segmented from high-resolution head CT images. CelebAMask-HQ has around 30,000 images of celebrity faces. To build an automated diagnosis tool, a huge dataset encompassing over 300,000 head CT images was gathered retrospectively from around 20 centers in India between January 1, 2011 and June 1, 2017. It includes the training and validation subsets (Qure25k and CQ500 datasets, respectively), the original clinical radiology report, and the consensus of three independent radiologists, which are regarded as the gold standard for the Qure25k and CQ500 datasets. This dataset is meant for the development and assessment of automated diagnostic systems, not for face approximation learning [12]. Unfortunately, the publicly accessible CT dataset is not suitable for the job of skull-based face approximation.

The MUG500+ [19] and CelebAMask-HQ [20] datasets were chosen due to the datasets are available to download as public datasets. It is difficult to create datasets of 3D human skulls and 3D human faces since a forensic expert is required to build 3D face datasets and 3D skull datasets. Previous research created 3-D skull and face datasets using 3D -modeling software such as 3DStudio Max, Freeform, Blender, Zbrush15, clay, or gypsum sculpture to create datasets and requires a good knowledge of anatomy, specifically facial anatomy. In this research, we attempted to use real data from unknown human skulls and public 2D faces. However, we ran into an obstacle due to 3D skulls and 3D faces not being available. The reconstructed faces were created by segmenting the skull image into ten parts and matching them to parts of the face, as shown in Figure 12(a). We updated the reconstructed face by clustering the CelebAMaskHQ into male and female. The reconstructed output face segmented the skull image into 18 parts and used a matching model with the same faces as the reconstructed face, as shown in Figure 12(b).

4.2. Experimental details

The input image size is $224 \times 224 \times 3$ to the CNNSPS model, which employs convolution 2D with many input channels, many output channels with a kernel size of 3×3 that moves one pixel at a time in both the horizontal and vertical directions by 1 stride and 1 pixel added to each side of the input image by padding. The model indicates that the batch normalization layer will normalize the number of output channels from the previous convolutional layer. The model uses batch normalization to improve neural network training stability and performance. It normalizes a given layer's activations by subtracting the mean and dividing by the standard deviation determined across a mini-batch of samples. This helps to reduce internal covariate shifts and can speed up network convergence during training. Max pooling is a down-sampling procedure used to minimize the spatial dimensions of the input while maintaining the most important characteristics. It splits the input into non-overlapping rectangular regions and outputs the greatest value from each. The max pooling layer uses a 2×2 kernel and a stride of 2. The model uses a linear transformation to the input data with parameters $224 \times 7 \times 7$.

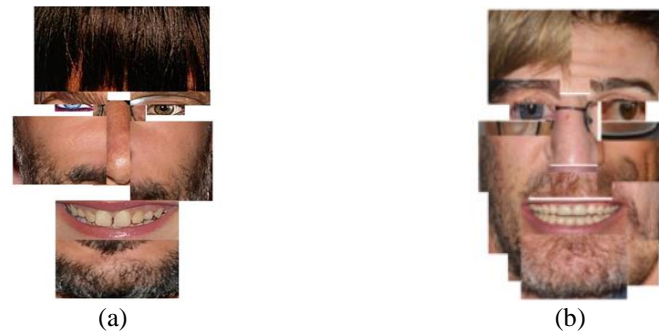


Figure 12. Face reconstruction (a) the reconstructed face of 10 parts and (b) the reconstructed face of 18 parts

5. CONCLUSION

Anthropology, archaeology, and forensics have issues with face approximation based on the skull-computer-aided facial reconstruction techniques based on digitized skull images of unknown people. Computer-aided forensics predicted the victim's nearest face by not employing standard procedures. This research uses the CNNSPS model, based on breaking unidentified skulls into parts and dividing known human images into parts to predict faces. This research makes use of publicly available datasets such as MUG500+ and CelebAMask-HQ. This research illustrates the anticipated outcome of unidentified skulls using a dataset of celebrity persons. Our fundamental issue is that the collection of known genuine skulls and known people does not exist. Thus, we rely on publicly available datasets. The challenge facing us in this research is that the real dataset is not available, and it is difficult to have it.




REFERENCES

- [1] E. A. A. El-Din, "Artificial intelligence in forensic science: invasion or revolution?," *Egyptian Society of Clinical Toxicology Journal*, vol. 10, no. 2, pp. 20–32, 2022, doi: 10.21608/escfj.2022.158178.1012.
- [2] C. Wilkinson, "Computerized forensic facial reconstruction: a review of current systems," *Forensic Science, Medicine, and Pathology*, vol. 1, no. 3, pp. 173–177, 2005, doi: 10.1385/FSMP:1:3:173.
- [3] S. N. Byers and C. A. Juarez, *Introduction to forensic anthropology: Sixth edition*. Routledge, 2023, doi: 10.4324/9781003283935.
- [4] C. N. Stephan, "Facial approximation-from facial reconstruction synonym to face prediction paradigm," *Journal of Forensic Sciences*, vol. 60, no. 3, pp. 566–571, 2015, doi: 10.1111/1556-4029.12732.
- [5] S. Gupta, V. Gupta, H. Vij, R. Vij, and N. Tyagi, "Forensic facial reconstruction: the final frontier," *Journal of Clinical and Diagnostic Research*, vol. 9, no. 9, pp. 26–28, 2015, doi: 10.7860/JCDR/2015/14621.6568.
- [6] M. D. Buhan and C. Nardoni, "A facial reconstruction method based on new mesh deformation techniques," *Forensic Sciences Research*, vol. 3, no. 3, pp. 256–273, 2018, doi: 10.1080/20961790.2018.1469185.
- [7] P. Mesejo, R. Martos, O. Ibáñez, J. Novo, and M. Ortega, "A survey on artificial intelligence techniques for biomedical image analysis in skeleton-based forensic human identification," *Applied Sciences*, vol. 10, no. 14, 2020, doi: 10.3390/app10144703.
- [8] "Crime rate by country 2024," *World Population Review*, 2024. Accessed: Aug. 20, 2023. [Online]. Available: <https://worldpopulationreview.com/country-rankings/crime-rate-by-country>
- [9] UNODC, "Global study on homicide: United Nations office on drugs and crime," *United Nations publication, Sales No. 14.IV.1*, 2019. Accessed: Aug. 20, 2023. [Online]. Available: <https://www.unodc.org/unodc/en/data-and-analysis/global-study-on-homicide-2019.html>
- [10] "World Crime Rate & Statistics 2000-2023," *Macro Trends*, 2023. Accessed: Aug. 25, 2023. [Online]. Available: <https://www.macrotrends.net/countries/WLD/world/crime-rate-statistics>
- [11] H. Q. Nguyen, T. N. Nguyen, V. D. Tran, and T. T. Dao, "A deep learning approach for predicting subject-specific human skull shape from head toward a decision support system for home-based facial rehabilitation," *IRBM*, vol. 44, no. 1, 2023, doi: 10.1016/j.irbm.2022.05.005.
- [12] V. A. Knyaz, V. V. Kniaz, M. M. Novikov, and R. M. Galeev, "Machine learning for approximating unknown face," *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, vol. 43, no. B2, pp. 857–862, 2020, doi: 10.5194/isprs-archives-XLIII-B2-2020-857-2020.
- [13] V. Kniaz, V. Knyaz, and V. Mizginov, "Synthesis and visualization of photorealistic textures for 3d face reconstruction of prehistoric human," *CEUR Workshop Proceedings*, vol. 2744, pp. 1–9, 2020, doi: 10.51130/graphicon-2020-2-4-15.
- [14] S. N. Bushra and K. U. Maheswari, "Crime investigation using DCGAN by forensic sketch-to-face transformation (STF)-a review," in *5th International Conference on Computing Methodologies and Communication, ICCMC 2021*, IEEE, Apr. 2021, pp. 1343–1348, doi: 10.1109/ICCMC51019.2021.9418417.
- [15] W. Shui *et al.*, "A computerized craniofacial reconstruction method for an unidentified skull based on statistical shape models," *Multimedia Tools and Applications*, vol. 79, no. 35–36, pp. 25589–25611, 2020, doi: 10.1007/s11042-020-09189-7.
- [16] J. S. Tan, I. Y. Liao, I. Venkat, B. Belaton, and P. T. Jayaprakash, "Computer-aided superimposition via reconstructing and matching 3D faces to 3D skulls for forensic craniofacial identifications," *Visual Computer*, vol. 36, no. 9, pp. 1739–1753, 2020, doi: 10.1007/s00371-019-01767-7.
- [17] H. M. R. Afzal, S. Luo, M. K. Afzal, G. Chaudhary, M. Khari, and S. A. P. Kumar, "3D face reconstruction from single 2D image using distinctive features," *IEEE Access*, vol. 8, pp. 180681–180689, 2020, doi: 10.1109/ACCESS.2020.3028106.
- [18] P. Navic, P. Palee, S. Prapayasatok, S. Prasitwattanaseree, A. Sinthubua, and P. Mahakkanukrauh, "The development and testing of Thai facial soft tissue thickness data in three-dimensional computerized forensic facial reconstruction," *Medicine, Science and*




- the Law*, vol. 62, no. 2, pp. 113–123, Nov. 2022, doi: 10.1177/00258024211057689.
- [19] J. Li *et al.*, “MUG500+ repository,” *Figshare*, 2021, doi: 10.6084/m9.figshare.9616319.
- [20] K. C. Surendra, “CelebAMask-HQ,” *Kaggle*, 2020. Accessed: Dec. 06, 2023. [Online]. Available: <https://www.kaggle.com/datasets/surendrakc/celebamaskhq>
- [21] Y. Liu *et al.*, “Improved feature point pair purification algorithm based on SIFT during endoscope image stitching,” *Frontiers in Neurorobotics*, vol. 16, 2022, doi: 10.3389/fnbot.2022.840594.
- [22] A. Zafar *et al.*, “A comparison of pooling methods for convolutional neural networks,” *Applied Sciences*, vol. 12, no. 17, 2022, doi: 10.3390/app12178643.
- [23] H. Gholamalinezhad and H. Khosravi, “Pooling methods in deep neural networks, A review,” *arXiv-Computer Science*, pp. 1-16, Sep. 2020, doi: 10.48550/arXiv.2009.07485.
- [24] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, “Convolutional neural networks: an overview and application in radiology,” *Insights into Imaging*, vol. 9, no. 4, pp. 611–629, 2018, doi: 10.1007/s13244-018-0639-9.
- [25] A. Dudhal, H. Mathkar, A. Jain, O. Kadam, and M. Shirole, “Hybrid SIFT feature extraction approach for indian sign language recognition system based on CNN,” in *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMVC-CVB)*, 2019, pp. 727–738, doi: 10.1007/978-3-030-00665-5_72.
- [26] M. F. M. Shaharom and K. N. Tahar, “Multispectral image matching using SIFT and SURF algorithm: A review,” *International Journal of Geoinformatics*, vol. 19, no. 1, pp. 13–21, 2023, doi: 10.52939/ijg.v19i1.2495.
- [27] K. Z. Liu, P. J. Lee, G. C. Xu, and B. H. Chang, “SIFT-enhanced CNN based objects recognition for satellite image,” in *2020 IEEE International Conference on Consumer Electronics - Taiwan, ICCE-Taiwan 2020*, IEEE, Sep. 2020, doi: 10.1109/ICCE-Taiwan49838.2020.9258037.
- [28] Z. H. -Nejad, H. Agahi, and A. Mahmoodzadeh, “Image matching based on the adaptive redundant keypoint elimination method in the SIFT algorithm,” *Pattern Analysis and Applications*, vol. 24, no. 2, pp. 669–683, Nov. 2021, doi: 10.1007/s10044-020-00938-w.
- [29] Y. Xiang, F. Wang, and H. You, “OS-SIFT: A robust SIFT-like algorithm for high-resolution optical-to-SAR image registration in Suburban areas,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 6, pp. 3078–3090, 2018, doi: 10.1109/TGRS.2018.2790483.
- [30] E. Adel, M. Elmogy, and H. Elbakry, “Image stitching based on feature extraction techniques: A survey,” *International Journal of Computer Applications*, vol. 99, no. 6, pp. 1–8, 2014, doi: 10.5120/17374-7818.
- [31] J. Li *et al.*, “MUG500+: database of 500 high-resolution healthy human skulls and 29 craniotomy skulls and implants,” *Data in Brief*, vol. 39, 2021, doi: 10.1016/j.dib.2021.107524.

BIOGRAPHIES OF AUTHORS






Doaa M. Mohammed    received her B.Sc. from the Department of Computer Science, Faculty of Computers and Artificial Intelligence, Benha University, Egypt, 2020. She has been working as a demonstrator in the Faculty of Computers and Artificial Intelligence since 2020, at Benha University, Egypt. Her research interests include digital forensics, machine learning, and computer vision. She can be contacted at email: doaa.ibrahem@fci.bu.edu.eg.



Mostafa Elgendy    received his M.Sc. degree from the Department of Computer Science, Faculty of Computers and Artificial Intelligence, Benha University, Egypt, 2015. He received his Ph.D. degree from the Department of Electrical Engineering and Information Systems, University of Pannonia, Veszprem, Hungary, in 2021. He worked as a demonstrator and assistant lecturer in the Faculty of Computers and Informatics, Benha University, Egypt, from May 2009 to 2021. Now, he is an Assistant professor in the Department of Computer Science, Faculty of Computers and Artificial Intelligence, Benha University, Egypt. His research interests include cloud computing, assistive technology, and machine learning. He can be contacted at email: mostafa.elgendy@fci.bu.edu.eg.



Mohamed Taha    is an Assistant Professor at Benha University, Faculty of Computers and Artificial Intelligence, Department of Computer Science, Egypt. He received his M.Sc. degree and his Ph.D. in computer science at Ain Shams University, Egypt, in February 2009 and July 2015. His research interests concern computer vision (object tracking-video surveillance systems), digital forensics (image forgery detection–document forgery detection-fake currency detection), image processing (OCR), computer networks (routing protocols-security), augmented reality, cloud computing, and data mining (association rules mining-knowledge discovery). Taha has contributed more than 20+ technical papers to international journals and conferences. He can be contacted at email: mohamed.taha@fci.bu.edu.eg.