International Journal of Advanced Technology and Engineering Exploration, Vol 10(106)
ISSN (Print): 2394-5443 ISSN (Online): 2394-7454
http://dx.doi.org/10.19101/IJATEE.2023.10101696

Research Article

Intelligent face sketch recognition system using shearlet transform and convolutional neural network model

Chaymae Ziani* **and Abdelalim Sadiq**
LARI laboratory, Department of computer sciences, Faculty of Sciences, Ibn Tofail University, Kenitra, Morocco

Received: 02-June-2023; Revised: 21-September-2023; Accepted: 23-September-2023

Abstract
Face sketch recognition is a crucial field with applications in identifying suspects and criminals based on verbal descriptions (face sketches) provided by eyewitnesses. Although deep convolutional neural networks (DCNNs) have significantly advanced face recognition from photos, recognizing faces from sketches remains challenging due to texture differences and limited training samples. To overcome these challenges, an innovative methodology that integrates the shearlet transform as a pre-processing layer within the DCNN was proposed. This combination aims to establish a robust learning foundation for identifying individuals from face photos using their corresponding face sketches. Experimental evaluations showcase the effectiveness of our approach, achieving a remarkably high recognition rate. The incorporation of the shearlet transform enhances the DCNN's capability to handle texture disparities between face photos and sketches, resulting in improved performance. Our research marks the first instance of combining DCNN with the shearlet transform for face sketch recognition. Our approach proves highly effective in addressing sketch recognition challenges, as evidenced by an impressively low error rate of only 0.7%. This leads to minimized false positives, a crucial factor in law enforcement applications. A flawless recall score and an F1-score of 100% demonstrate exceptional performance in correctly identifying matches. This advancement carries promising implications for sensitive applications, such as recognizing suspects and criminals based on eyewitness descriptions, ultimately enhancing overall security and law enforcement efforts.

Keywords
Face sketch, CNN, Shearlet transform, Face recognition.

1.Introduction
The recognizing faces from sketches is a critical task with numerous real-world applications, such as identifying suspects and criminals based on verbal descriptions provided by eyewitnesses through sketch drawings. However, compared to recognizing faces from photos, facial sketch recognition presents significant challenges. The primary difficulty arises from the substantial differences in texture between facial sketches and their corresponding face photos [1–4]. Additionally, facial sketches can contain inaccuracies, complicating the recognition process.

The existing literature on facial sketch recognition has explored various techniques, including the use of deep convolutional neural networks (DCNNs) for face recognition.

While DCNNs have achieved remarkable success in recognizing faces from photos [5, 6], their performance in facial sketch recognition has been limited. The major constraints include the texture disparity between sketches and face photos and the limited availability of learning samples. These constraints have hindered the effectiveness of DCNNs in achieving a high recognition rate for facial sketches [7].

The limitations in existing methods and the critical importance of facial sketch recognition in applications like law enforcement and security motivated us to seek an effective solution to enhance the recognition rate. The goal was to address the challenges posed by texture differences and limited learning samples, ultimately enabling accurate identification of individuals from facial sketches.

In this paper, we aim to propose a novel methodology that leverages the shearlet transform as a pre-layer

*Author for correspondence

within the DCNN to improve facial sketch recognition. The shearlet transform has proven effective in generating coefficients representing essential features of images [7]. By incorporating these coefficients into the DCNN, we intend to mitigate the impact of texture disparities and the limitation of input sample size, thereby enhancing the recognition rate for facial sketches.

The primary contribution of this research lies in the integration of the shearlet transform as a pre-layer in the DCNN architecture specifically designed for facial sketch recognition. To the best of our knowledge, this is the first instance where the shearlet transform has been utilized in combination with DCNNs for this purpose. Our approach aims to overcome the challenges faced by previous methods, leading to a substantial improvement in the recognition rate for facial sketches.

The remainder of this paper is organized as follows: Section 2 provides a comprehensive review of the related literature on facial sketch recognition and shearlet transform. In section 3, we present the proposed methodology, detailing the integration of the shearlet pre-layer with the DCNN, experimental setup, dataset, and evaluation metrics used to assess the performance of our approach. The results are presented in section 4, followed by the discussion of findings and implications in section 5. Finally, section 6 concludes the paper, summarizing the contributions and highlighting potential avenues for future research.

2.Literature review

Before delving into this section, it is essential to recall that face detection is a fundamental step in the recognition pipeline, as it involves locating and extracting facial regions from an image or sketch [8]. Also, the use of convolutional neural networks (CNNs) [9–11] in image processing has led to significant improvement and rapid advancement in applications related to this field [12, 13, 9].

This literature review emphasizes recent facial sketch recognition approaches directly relevant to this field of study.

Bahrum et al. [14] in perform a systematic literature review on face sketch recognition, analyzing 35 papers from 2011 to 2021. It highlights the emergence of deep learning techniques for faster and more accurate transformations compared to traditional methods. The generative adversarial networks (GANs) are recognized as effective solutions for various facial recognition challenges. The study underscores GANs' significance and acknowledges limitations tied to data scope, accuracy factors, and the need for large datasets. In [15] the paper presents a deep-learning framework to match face sketches with photos, tackling modality differences through an intermediary latent space and collaborative synthesis. The suggested method successfully bridges these modality gaps, resulting in enhanced performance in face sketch-photo matching. Nonetheless, its effectiveness relies on data availability, while the intricate bidirectional synthesis and StyleGAN-like architecture could hinder practical implementation under resource limitations.

Radman et al. [16] propose DResNet, a method for face sketch synthesis, addressing issues of ethnicity and photo variations. DResNet combines deep residual and feedforward networks for diverse and accurate sketches, outperforming existing methods. chinese university of hong kong (CUHK) face sketch database evaluations demonstrate its effectiveness. However, as with any deep learning approach, potential limitations might include the need for a significant amount of training data, computational resources, and potential challenges in generalization to extremely diverse scenarios not covered by the training dataset.

Yan et al. [17] presents an identity-sensitive generative adversarial network (IsGAN) for face photo-sketch synthesis. IsGAN addresses the challenge of preserving identifiable details in the synthesis process. It incorporates identity information through adversarial learning and introduces an identity recognition loss to maintain detailed identifiable information. The model outperforms state-of-the-art methods in both qualitative and quantitative evaluations on CUHK face sketch (CUFS) and CUHK face sketch feret (CUFSF) datasets. The advantages include improved preservation of identity-sensitive features and high-fidelity synthesis.

Bhoir et al. [18] present a standalone tool for creating and identifying face sketches based on eyewitness descriptions, aiming to improve on existing systems limitations. It employs deep learning and cloud infrastructure for real-time matching with a criminal database. Experimental results show the tool's efficiency and accuracy, achieving over 90%

similarity and 94.6% accuracy in face sketch recognition.

Alhashash et al. [19] introduce the smart switching slime mould algorithm (2SMA), a novel optimization method for fine-tuning pre-trained deep learning models to enhance face sketch recognition accuracy. The performance of 2SMA was evaluated on Multi Modal Verification for Teleservices and Security (XM2VTS), CUFSF, and CEC's 2010 benchmark, consistently demonstrating superior results. For instance, XM2VTS achieved an impressive 98.81% recognition rate in the multiple-model approach. However, a potential limitation is its specialized tuning to specific datasets and deep-face models, potentially affecting its performance on diverse datasets or models with distinct characteristics.

In [20], the paper introduces a combined approach for face caricature synthesis and recognition using a semantic neural model. This integration addresses two tasks typically treated separately. The study assumes faces in frontal pose, with standard lighting, neutral expression, and no obstructions. Successful synthesis of caricature/image photos was achieved, with an accuracy rate of 64.0%. However, a limitation lies in the method's reliance on specific ideal conditions for input faces, potentially limiting its effectiveness in real-world scenarios where these criteria may not always be met.

An intra-domain enhancement (IDE) method are represented in [21], which targeting both modality gap and quality issues within the same domain for face photo-sketch synthesis. Extensive experiments on public face sketch databases affirm the IDE-based approach's superiority over current state-of-the-art methods, particularly in detail preservation. However, the method's specialized focus on enhancing sketches from low-quality photos might lead to suboptimal performance with higher-resolution or well-detailed images.

Zhong et al. [22] introduces the unsupervised self-attention lightweight (USAL) photo-to-sketch synthesis with feature maps method, utilizing a self-attention module in a generative adversarial network for realistic skin texture and eye-focused synthetic sketch creation. USAL surpasses fixed-architecture models in face sketch-to-photo synthesis, particularly excelling in CUFS database analysis. However, its effectiveness depends on data availability and quality, potentially limiting performance in biased or insufficiently diverse datasets. Additionally, resource-intensive components may pose practicality challenges in resource-constrained environments.

Wan et al. [23] introduces a novel face sketch recognition approach utilizing transfer learning, which leverages knowledge from a related task to enhance performance in face sketch recognition. The authors devise a three-channel CNN architecture, incorporating triplet loss to discern features and decrease intra-class variations. The method successfully aligns facial features from digital photos and corresponding sketches, effectively distinguishing between different identities. However, the scarcity of training data remains a primary challenge in adopting deep learning for face sketch recognition, potentially limiting its effectiveness in scenarios with limited data.

In [24] introduce a modified version of the Viola-Jones algorithm for detecting faces in the video frames. The proposed salp-cat optimization algorithm is utilized to select relevant features for further processing. The selected features are combined with scale-invariant feature transform (SIFT) features, likely to enhance the recognition process. Euclidean distances are computed between feature vectors derived from the sketch and each detected face in the video. The proposed method's performance is assessed using the chokepoint dataset. It achieves an impressive precision of 89.02%, recall of 91.25%, and F-measure of 90.13%. But, the method's performance is demonstrated using the chokepoint dataset, which may not fully represent the diverse conditions that can be encountered in real-world surveillance scenarios.

The paper proposed in [25] uses a deep convolutional generative adversarial network (DCGAN) to convert pencil sketches into high-quality real images. The prepared sketch serves as the training input for the model. The proposed method achieves an average structural similarity index (SSIM) of 0.587, surpassing existing methods with an average SSIM of 0.554. The success of the proposed method is attributed to a unique architecture and modified parameters. The applicability and performance of this model in different contexts may require further exploration and adjustment.

Jacquet et al. [26] conduct a literature review to establish a methodological workflow for developing a Bayesian-based score-based likelihood-ratio computation model in forensic face recognition. It explores different approaches for modeling

variability distributions based on available data and case specificity. The assessment of automatic forensic face recognition systems showcases their versatile applications in intelligence, investigation, and evaluation contexts. The paper outlines a comprehensive global workflow for computing a score-based likelihood ratio (SLR). However, the specific advantages and drawbacks of specific versus generic SLR computation approaches require further investigation.

In [27] and [28] attribute-driven methods are introduced by (Kazemi et al.) and (Iranmanesh et al.). These methods leverage facial attributes to enhance recognition accuracy by incorporating loss functions centered on attributes and joint identity loss functions. While promising, their effectiveness heavily relies on accurate attribute estimation, which can be challenging in practical scenarios. Liu et al. [29] proposed the coupled attribute-guided triplet loss (CAGTL) that addresses issues caused by poorly estimated attributes in an end-to-end network, showing improved performance in heterogeneous face recognition. However, depending solely on attributes for face recognition may limit their effectiveness in dealing with variations in sketches.

In contrast, Liu et al. [30] presented an iterative local re-ranking approach with attribute-guided synthesis, leveraging GANs to enhance matching by improving the quality of generated sketches. Although this approach demonstrates enhanced recognition accuracy, GAN-based methods can be sensitive to the quality of input sketches, potentially affecting the reliability of generated samples.

Another line of research explores local descriptor-based methods for facial sketch recognition. Peng et al. [31] introduced deep local descriptor for cross-modality face (DLFace) recognition, a local descriptor approach based on metric deep learning, which captures fine-grained facial features and shows competitive results in cross-modality face recognition tasks. However, the performance of DLFace may be influenced by the choice of local descriptors and their adaptability to diverse sketch representations.

Fan et al. [32] proposed a triplet network for face sketch recognition, incorporating an image-space attention model to extract features from corresponding locations in photos and sketches. This attention mechanism enhances recognition accuracy by identifying informative regions in images, reducing cross-modality differences. While promising, the effectiveness of the attention model may depend on the quality and alignment of corresponding regions between sketches and photos. Overall, recent research in facial sketch recognition encompasses a range of approaches, each offering unique advantages and limitations. By exploring different methodologies, researchers contribute to advancing the field and addressing challenges in recognizing faces from sketches.

Upon reviewing the literature mentioned, along with other works in the field, a noticeable gap emerges concerning the limited training sample size constraint for CNN. The scarcity of sketch data poses a significant challenge in applying deep learning techniques to facial sketch recognition [1]. Surprisingly, despite the success of CNNs in various domains, their utilization as a primary solution for facial sketch recognition remains relatively unexplored. Consequently, our experimental study aims to emphasize the feasibility and effectiveness of our proposed solution, primarily relying on a CNN. Additionally, we will conduct a comparative analysis between the proposed approach and a CNN to showcase the considerable added value offered by our innovative solution. Through this investigation, we aim to contribute to the field by addressing the limitations of existing methods and demonstrating the potential of CNN-based solutions in facial sketch recognition.

3.Methods

3.1Shearlet transform

Shearlets were introduced to analyze and sparsely represent functions **f** in $L^2(R^2)$. They serve as a multidimensional extension of the wavelet transform, offering efficient encoding of anisotropic structures and the representation of multidimensional data. Therefore, shearlets have demonstrated their superiority over wavelets in achieving an optimal sparse representation for images containing edges. They have been successfully employed in various image processing tasks, including face detection, feature extraction, and denoising [33]. Expressly, the continuous shearlet transform is defined as (Equation 1 and Equation 2):

$$f \longmapsto SH_{\psi} f(i,j,k) = \langle f, \psi_{i,j,k,} \rangle , f \in L^2(R^2) , (i,j,k) \in \mathbb{R}_{>0} \times \mathbb{R} \times \mathbb{R}^2 \quad (1)$$

Where

$$\psi_{i,j,k(x)} = \mid detM_{i,j} \mid^{-\frac{1}{2}} \psi(M_{i,j}^{-1}(x-1)) \quad (2)$$

The shearlets transform function incorporates three variables: i, representing the scale parameter that determines the resolution level; j, denoting the shear parameter indicating the directionality and k, representing the translation parameter that specifies the position.
The system utilizes a parameter ψ known as wavelets with composite dilation (Equation 3).

$$M_{i,j} = \begin{pmatrix} i & j \\ 0 & i \end{pmatrix}. \quad (3)$$

The matrix $M_{i,j}$ is factorized as B_jA_i. In this context, $A_i = \begin{pmatrix} i & 0 \\ 0 & \sqrt{i} \end{pmatrix}$ denotes a parabolic scaling matrix, while $B_j = \begin{pmatrix} 1 & j \\ 0 & 1 \end{pmatrix}$ is a shearing matrix.

By discretizing the scaling, shear, and translation parameters and sampling the continuous shearlet transform accordingly, it becomes feasible to acquire a frame or even a Parseval Frame for $L^2(R)$. The discrete shearlets are obtained by sampling three parameters (Equation 4):

$$i_m = 2^m (m \in \mathbb{z}), \quad j_{m,l} = li_m^{\frac{1}{2}} = l2^{\frac{m}{2}} (l \in \mathbb{z})$$
$$\text{and} \quad k_{m,l,n} = D_{a_m,s_{m,l}} (n \in \mathbb{z}^2) \quad (4)$$

For $\psi \in L^2(R^2)$, the discrete shearlet transform (DST) of $f \in L^2(R^2)$ is the mapping defined by (Equation 5):

$$f \mapsto SH_\psi f(m,l,n) = \langle f, \psi_{m,l,n} \rangle,$$
$$f \in L^2(R^2), (m,l,n) \in \mathbb{z} \times \mathbb{z} \times \mathbb{z}^2 \quad (5)$$

Where Equation 6 is the discrete shearlets function

$$\psi_{m,l,n,}(x) = 2^{\frac{3m}{2}} \psi(B_l A^m x - n),$$
$$j \geq 0, -2^j \leq k \leq 2^j - 1, m \in \mathbb{z}^2 \quad (6)$$

The discretization of $M_{i,j}$ can be represented as $M_{m,l} = B_l A^m$, resulting in the mapping of the function f to the coefficients $SH_\psi f(m,l,n)$ associated with the scale index m, the orientation index l, and the position index n through the operator SH_ψ.

3.2Proposed approach

The main driving factors behind our proposed approach are two significant constraints:

- The disparity in textures between the sketch and the corresponding photo.
- The limited input sample during the CNN learning phase (consisting of a face sketch and a face photo for each individual).

Our approach addresses the texture difference through two levels of intervention:

- The first level involves converting the face photo into a sketch using established techniques in the field of image processing.
- The second level entails generating coefficients for both the face sketch and the corresponding photo using the DST. This step allows us to extract essential points that facilitate better comparability, ultimately minimizing the texture differences indirectly.

To address the issue of the low input sample for the CNN, we will enrich it by incorporating the various coefficients generated for each face sketch and photo. By doing so, we increase the size of the input sample, which will have a positive impact on the recognition rate of our solution. This enrichment allows us to provide the CNN with a more diverse and comprehensive set of data, enabling it to learn and generalize better for improved recognition performance.

The proposed approach is based on a succession of processing layers as shown by the block diagram in *Figure 1*. Each layer performs a very specific task with the final objective being to increase the rate of facial recognition from the sketch of a person. The processing layers are organized as shown in *Figures 1* and *2*, in the following order:

1. Face detection in sketch and photo
2. The conversion of face photos into sketches
3. Generating coefficients using DST
4. Feature extraction and classification through CNN.

3.2.1Face detection

In this step, we employ the Haar-Cascades method [34] to extract the face from both the sketch and the corresponding photo. The primary goal of this method is to remove any non-essential details, focusing solely on isolating the facial region of interest. Haar-Cascades are utilized as a technique to effectively detect and extract faces, helping to streamline the subsequent extraction and recognition steps.

3.2.2Converting face photos into sketch

The subsequent step involves converting face photos into sketches, aiming to minimize the texture difference between the photo and the corresponding sketch. To accomplish this, a range of established techniques from the field of image processing are employed. Specifically, the conversion technique utilized is based on the widely-used OPENCV graphics library [35]. This involves employing a combination of functions, including CvtColor(), BitWise_not(), GaussianBlur(), and Divide(), to

achieve the desired conversion effect. These functions collectively contribute to the process of transforming the face photos into sketch-like representations, reducing the texture disparity between the original photo and the resulting sketch.

3.2.3Generation of coefficients

This step is considered highly significant within the entire processing workflow due to the two guaranteed benefits it provides:

Firstly, this step allows us to augment the size of the CNN's input sample, thereby ensuring a more effective learning process. By applying the DST, shearlet coefficients are generated for each sketch and its corresponding face photo, accounting for different orientations and translations during the decomposition. This process results in the extraction of a substantial amount of essential data (coefficients), effectively expanding the input sample.

Secondly, this step plays a crucial role in minimizing the texture disparity between the sketch and the corresponding face photo. The generated coefficients exhibit closer similarity in terms of texture, enabling improved comparability between the two. As a result, the difference in texture is significantly reduced, reinforcing the overall quality of the recognition process.

Overall, this step holds utmost importance by not only expanding the input sample for better learning but also by effectively addressing the texture differences between the sketch and its corresponding face photo through the generation of closely-related coefficients.

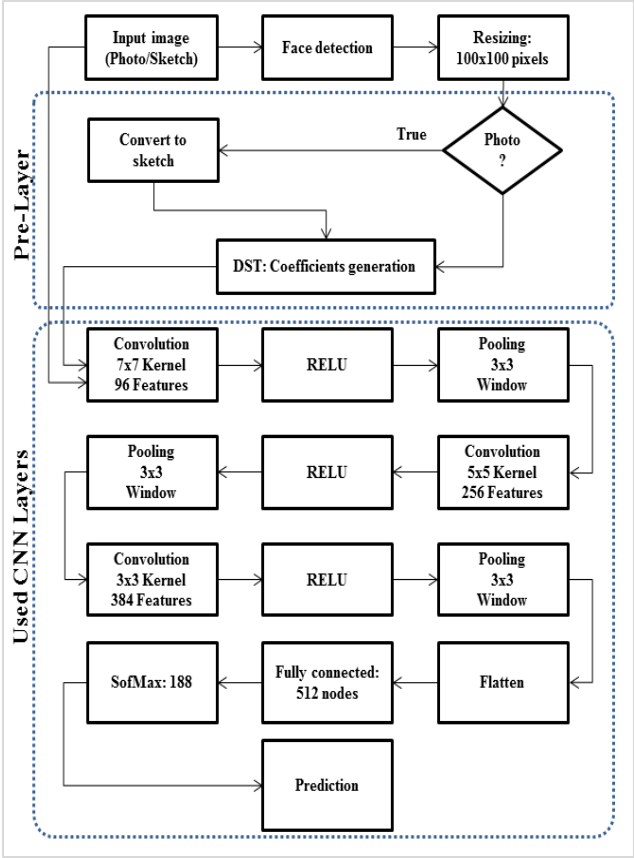

Figure 1 Architectural block diagram of the proposed approach (Learning phase)

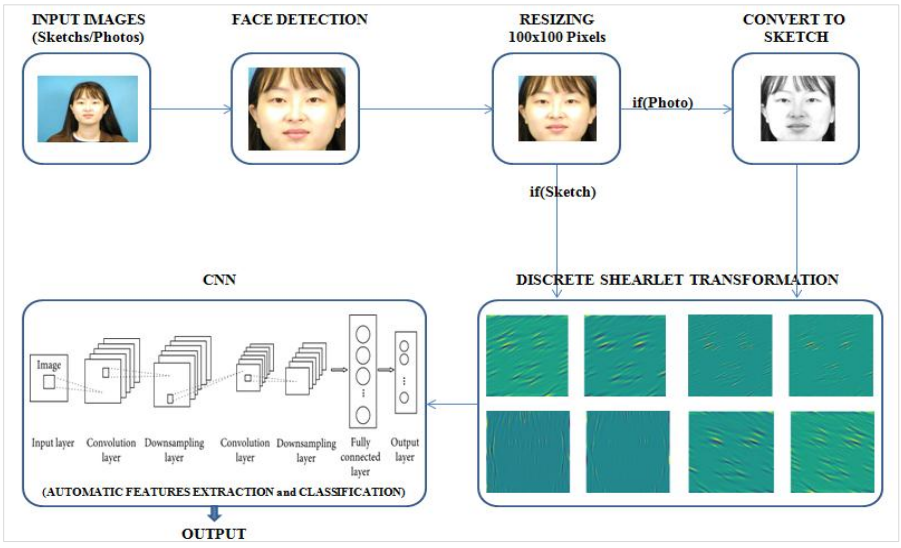

Figure 2 Example of the processing steps (learning phase)

3.2.4Extraction and classification through the CNN

Upon reaching this stage, the coefficients generated by the DST will be integrated into the CNN to commence the learning and testing phases. The CNN employed consists of a series of layers that adjust their parameters during the learning phase to guide the network towards the desired output. The specific layers constituting our CNN are represented in *Figure 1*. Within the various layers of the CNN, crucial information will be extracted from each coefficient and classified based on the training data.

At this stage, the CNN acts as a powerful tool for feature extraction and pattern recognition. It leverages the input coefficients obtained through the DST to discern important discriminative features, ultimately enabling accurate classification based on the learned representations. The layered architecture of the CNN facilitates the hierarchical extraction of essential data from the coefficients, allowing for effective discrimination and recognition of faces.

Once the CNN is trained, the next step is the test which proceeds as shown in *Figure 3.*

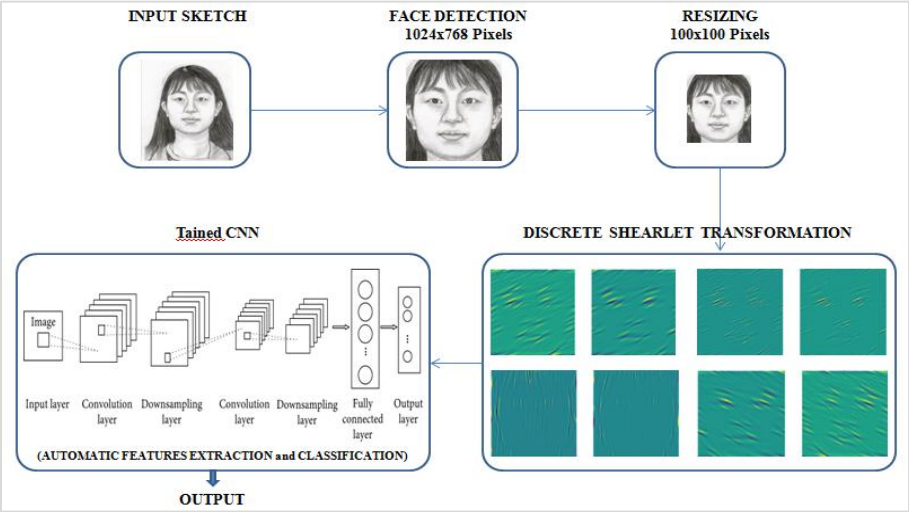

Figure 3 The proposed approach in testing phase

In summary, the integration of the DST-generated coefficients into the CNN marks a crucial phase in the process, where the network learns to extract and classify the essential information required for accurate face recognition.

3.3Discussion CNN architecture and enhancement

The CNN architecture used in our approach is designed to incorporate the shearlet transform pre-layer, which sets it apart from traditional CNNs, as shown in *Figure 1*. Let's discuss the architecture and the reasons why it is enhanced by our approach:

3.3.1Architecture of the CNN Model

The CNN architecture comprises several layers, each serving a specific purpose in feature extraction and classification. The typical layers in a CNN include:

Convolutional layers: These layers use filters (kernels) to perform convolutions over the input image, extracting features from local receptive fields. Convolutional layers are capable of learning meaningful patterns, such as edges, textures, and shapes.

Activation layers: Activation functions introduce non-linearity to the model, allowing it to learn complex relationships between features. Rectified linear unit (ReLU) is a commonly used activation function in CNNs.

Pooling layers: Pooling layers reduce the spatial dimensions of the feature maps, thereby decreasing the computational burden. Max-pooling is a popular pooling technique that retains the most important information in each pooling region.

Fully connected layers: These layers take the output from the previous layers and perform classification tasks, mapping the extracted features to specific classes.

3.3.2Enhancement by our approach:

The enhancement in our approach comes from the integration of the shearlet transform pre-layer with the CNN. This pre-layer serves as a means to incorporate essential features from the input face photo into the CNN, which significantly improves the recognition performance for facial sketch

recognition. Here's why our approach is enhanced by this integration:

Capturing salient features: The shearlet transform is a multi-scale, multi-directional transform that excels at capturing salient points and essential features in an image. By using the shearlet coefficients as the pre-layer input to the CNN, our approach provides the model with vital information that is often lost in traditional image processing techniques.

Addressing texture differences: One of the key challenges in face sketch recognition is the texture difference between face photos and sketches. Traditional CNNs may struggle to handle these differences, leading to reduced recognition accuracy. However, by leveraging the shearlet transform, our approach overcomes this limitation and enhances the CNN's ability to recognize facial sketches accurately.

Mitigating limited training samples: Face sketch recognition typically suffers from a small training sample size, making it challenging for traditional CNNs to generalize effectively. The shearlet pre-layer enriches the input information, compensating for the limited training samples and enabling the CNN to learn more robust representations.

Sparse representation: The shearlet transform provides a sparse representation of images, which is beneficial in handling noise and errors often present in facial sketches. This property helps in making the CNN more robust to noise and improves the overall recognition performance.

The noise removal: The DST is a powerful tool for noise reduction due to its ability to sparsely represent images by capturing essential information in the form of coefficients. By decomposing the face sketches and face photos into different scales, orientations, and translations, the DST can efficiently identify and separate noise from the relevant features.

In conclusion, our enhanced CNN architecture, which incorporates the shearlet transform pre-layer, overcomes the challenges of texture differences and limited training samples, resulting in a significantly improved recognition rate for facial sketch recognition. The integration of shearlet coefficients provides crucial information to the CNN, enhancing its ability to extract discriminative features and improving the accuracy and robustness of the entire face recognition system.

3.4Experience

The proposed approach has been implemented on the jupyter notebook platform using the Python programming language. Python, along with its frameworks and libraries, is renowned for its robustness in the realm of object recognition algorithms from photos. Notably, the keras library, which is employed for the CNN component, exemplifies the capabilities of Python in this domain.

The simulations were conducted on a machine equipped with a 4 GB RAM and a 1.7 GHz i3 processor. The relatively modest complexity of our proposed solution ensures that it can be executed on this machine without encountering any difficulties. The computational requirements of the approach are well-suited for the available hardware, allowing for efficient and seamless execution.

3.4.1CUHK database

The dataset used in the research is the "CUHK FACE SKETCH STUDENT" database [36], which is a widely recognized and established dataset in the field of face sketch recognition. The dataset comprises face sketches along with their corresponding face photos (as shown in *Figure 4*), making it suitable for training and evaluating face recognition systems.

Dataset description:

- Number of Samples: The dataset contains a total of 88 face sketches, each paired with its corresponding face photo, for the training phase. Additionally, there are 100 face sketches with their corresponding face photos for the testing phase.
- Image Size: The original images in the dataset have a pixel size of 1024×768. However, for the purposes of the research, the images were resized to a standardized size of 100×100 pixels to facilitate easier handling and analysis.
- Diversity: The dataset offers a diverse set of face sketches and face photos, covering a range of variations in terms of pose, expression, illumination, and background. This diversity helps in training a robust face recognition model that can generalize well to different scenarios.
- Quality: The face sketches in the dataset are created with high quality, ensuring that they accurately represent the facial features of the corresponding individuals. Similarly, the face photos are of sufficient quality to enable effective recognition.
- Annotations: Each face sketch is paired with its corresponding face photo, providing ground truth annotations for the training and evaluation of the model. The annotations enable supervised learning, where the model learns from labeled examples during the training process.

Dataset usage:
The dataset is divided into two main subsets: a training set and a testing set. The training set consists of 88 face sketches with their corresponding face photos, while the testing set contains 100 face sketches along with their corresponding face photos.

Future dataset considerations:
While the "CUHK FACE SKETCH STUDENT" database is a valuable resource for face sketch recognition research, future studies may consider incorporating additional datasets to further enhance the model's performance and generalization ability. Utilizing multiple datasets with different characteristics, such as varying ethnicities, ages, and image resolutions, can help improve the model's robustness and applicability in real-world scenarios.

In summary, the "CUHK FACE SKETCH STUDENT" dataset plays a crucial role in the research, providing a comprehensive collection of face sketches and corresponding face photos for training and evaluation. The dataset's diversity and quality enable the development of an effective face recognition model, and the annotations facilitate supervised learning for accurate recognition.

3.4.2Shearlet transform parameters

The initial key component in our proposed solution is the DST layer, which follows the preprocessing layer consisting of face detection and the conversion of face photos into sketches.
As previously mentioned, the purpose of this layer is twofold: to enhance the input sample of the CNN and to minimize the texture disparity between the sketch and the corresponding face photo. The specific configuration employed for the DST is as follows, and more details are given in *Table 1:*

- Four levels of decomposition
- A range of SHEARs from $2^0$to 2^3
- Scales ranging from 0 to 3
- Four directions (4, 8, 16, 16)

Table 1 DST parameters

Decomposition levels	Number of directions	Size of shear filters
Level 1	4	32×32
Level 2	8	32×32
Level 3	16	16×16
Level 4	16	16×16

Figure 4 photos and its corresponding sketch from CUHK database

The selection of specific values for shearlet parameters in our proposed approach is based on a combination of prior research, experimental analysis, and empirical observations. The chosen values aim to strike a balance between the effectiveness of the shearlet transform in capturing essential facial features and the computational complexity of the overall approach. Let's discuss the reasons behind the selection of these specific values:
Number of decomposition levels: The choice of four decomposition levels is a common practice in the shearlet transform, widely used in various image processing tasks. Four levels provide a good balance between capturing fine details and avoiding excessive computation. Higher levels could lead to more precise representations but may also increase computational complexity significantly, making the approach less efficient.
Shear parameters: We use a range of shear values from 2^0 to 2^3. These values allow the shearlet transform to capture information at different orientations, enabling it to extract anisotropic features effectively. This range of shear parameters has shown to be effective in previous shearlet-based approaches for image processing and feature extraction.
Scale parameters: The scales ranging from 0 to 3 cover a reasonable range of resolutions in the shearlet transform. Lower scales capture global image features, while higher scales capture finer details. With scales up to 3, we can adequately represent the essential facial features without sacrificing too much computational efficiency.

Directions: We employ four directions (4, 8, 16, 16) for the shearlet transform. These directions facilitate the extraction of directional information from the input images, which is important in distinguishing facial characteristics and structures. A moderate number of directions ensures that the shearlet transform can capture essential information without overwhelming the computation.

Overall, these specific values for the shearlet parameters are considered standard in the shearlet transform and have been used effectively in various image processing and recognition tasks. They strike a good balance between representation capabilities and computational efficiency, making them suitable for our face sketch recognition application.

It is important to note that the choice of shearlet parameters can also be influenced by the characteristics of the dataset being used and the specific requirements of the recognition task. Fine-tuning these parameters based on the specific dataset and application scenario can further enhance the performance of the shearlet transform in our proposed approach. Additionally, conducting sensitivity analyses on these parameters may provide insights into their impact on recognition accuracy and help optimize the overall performance of the system.

3.4.3CNN parameters

The CNN architecture, as depicted in *Figure 5*, consists of three convolutional layers. Each convolutional layer is subsequently followed by a max pooling operation to downsample the feature maps. Following the convolutional layers, a Flatten layer is employed to reshape the feature maps into a 1-dimensional vector, facilitating the connection with the subsequent layer. This subsequent layer is a fully-connected layer comprising 512 neurons, which aims to capture higher-level representations of the input data. Finally, the network concludes with a SoftMax layer, which enables the mapping of the learned features to the corresponding class probabilities.

To optimize the performance of the CNN, we have chosen specific sizes for the convolutional layers: 7×7, 5×5, and 3×3. This choice is made to reduce the computational complexity of the network, ultimately minimizing the overall execution time. Considering that the size of the input face sketch and face photo is 100×100 pixels, there is no need to choose a higher neuron count for the fully-connected layer. Raising the number of neurons beyond 512 would not significantly improve the recognition rate. Instead, it would only increase the execution time without providing substantial benefits.

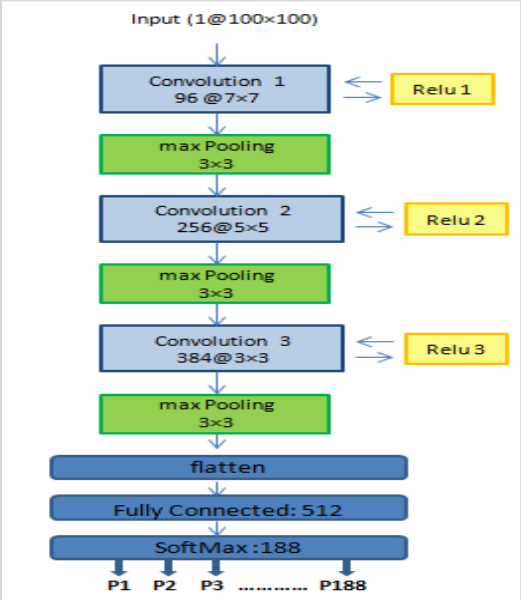

Figure 5 Architecture of the CNN used

It is essential to strike a balance between the network's capacity to capture relevant features and the computational resources required for training and inference. In our case, we have determined that the chosen configuration with the specified convolutional layer sizes and a fully-connected layer of 512 neurons is sufficient to achieve satisfactory recognition performance while maintaining computational efficiency.

The specific values of the CNN parameters were chosen based on a combination of empirical experimentation and standard practices in the field of deep learning.

The goal was to design a CNN architecture that is capable of effectively learning and extracting discriminative features from the input data while maintaining computational efficiency. Here is the rationale behind the chosen values for each layer:

Convolution Layer (96@7×7): The choice of 96 filters with a size of 7×7 was inspired by the architecture of the popular AlexNet model, which showed good performance in image classification tasks. The larger filter size allows the network to capture larger spatial patterns and features, which is beneficial for face recognition tasks where facial features can vary in size and complexity.

Max pooling layer (3×3): Pooling layers with a pool size of 3×3 was used to downsample the feature maps

and reduce the spatial dimensions. This helps in making the network computationally efficient and reduces the risk of overfitting. A pool size of 3x3 is a common choice in many CNN architectures.

Convolution layer (256@5×5): For the second convolutional layer, 256 filters with a size of 5×5 were used to continue capturing higher-level features from the feature maps generated by the first convolutional layer. The choice of 256 filters is a standard practice in deep learning architectures.

Max pooling layer (3×3): Another max pooling layer with a pool size of 3×3 follows the second convolutional layer to further downsample the feature maps and improve the network's ability to learn invariant features.

Convolution layer (384@3×3): The third convolutional layer employs 384 filters with a size of 3×3 to continue extracting higher-level representations. The choice of 384 filters aligns with the approach used in the famous VGGNet architecture.

Max pooling layer (3×3): The final max pooling layer with a pool size of 3×3 further reduces the spatial dimensions and prepares the data for the subsequent fully connected layers.

Flatten layer: After the pooling layers, a flatten layer is used to reshape the 3D feature maps into a 1D vector, making it suitable for the fully connected layers.

Fully connected layer (512): The fully connected layer with 512 neurons allows the network to learn higher-level abstractions from the flattened feature vector. The choice of 512 neurons is a common practice in many CNN architectures.

SoftMax layer (188): The final layer is the SoftMax layer with 188 output neurons, each representing a different class for face recognition. The number of output neurons corresponds to the number of classes in the dataset.

As for the **hyperparameters** (learning rate, batch size, number of epochs, etc.), they were determined through a process of hyperparameter tuning and validation on the validation set. The learning rate was chosen to allow the model to converge to a good solution during training without oscillating or overshooting. The batch size was set based on the available memory of the hardware used for training. The number of epochs was chosen to train the model sufficiently without overfitting to the training data.

It's important to note that the specific values chosen for the CNN parameters were not exhaustive and were based on the constraints of the available hardware and computational resources. Further hyperparameter tuning and optimization can be performed to find even better values for improved performance. Additionally, the choice of these specific values was validated through experimentation and analysis of the recognition results, demonstrating the effectiveness of the chosen CNN architecture in combination with the shearlet transform preprocessing step.

3.4.4Features selection

In the proposed approach for face sketch recognition, the main feature selection technique used is the DST, which generates a set of coefficients that represent essential features of the input images, both face sketches and face photos. Each coefficient corresponds to specific views or angles of the image and captures salient points within the image. After that we select two coefficients of the first decomposition, to incorporating into the CNN.

3.4.5Train-test split selection and variations

The train-test split selection: the dataset is divided into two subsets: the training set and the testing set. The training set uses 88 face sketches with corresponding face photos for training the CNN model. The testing set uses 100 face sketches with their corresponding face photos for testing.

Variations in the train-test split: To ensure the validity and reliability of the results, we have use variations in the train-test split in terms of *"data augmentation"*. We have used variations of the original images, by using DST to generate coefficients, to effectively increasing the size of the training set.

4.Results

Figure 6 illustrates the results of the simulation series conducted on the proposed approach, providing a direct comparison with the CNN model. This comparison is crucial for clarification. As previously explained, we introduced an extra pre-processing layer to the CNN, referred to as the DST layer. This additional layer serves two primary purposes: enhancing the CNN's input data and reducing the texture differences between the facial sketch and its corresponding face photo.

By incorporating the DST layer into the architectures of other existing or future face sketch recognition solutions that rely solely on the CNN, the potential to enhance the recognition rate is evident. This rationale underscores the reason for directly comparing the proposed solution with the conventional CNN. This comparative analysis serves to underscore the value added by our approach and evaluate its efficacy in

enhancing recognition performance when contrasted with traditional CNN-based methodologies.

In *Figure 6*, three curves illustrate the recognition rates of the simulated solutions. The green curve represents our proposed approach, while the blue curve pertains to scenario 1, where the face sketch and its associated face photo were directly input without any preprocessing. The red curve corresponds to scenario 2, depicting the results of the CNN with a preprocessing layer, involving the conversion of face photos into face sketches prior to CNN input to slightly alleviate texture discrepancies.

From the obtained results, it is evident that the proposed approach outperforms the conventional CNN. Notably, the simulations of our proposed approach achieved remarkable recognition rates within twenty epochs. Conversely, for scenarios 1 and 2 (blue and red curves), one hundred epochs were utilized without achieving a satisfactory recognition rate.

We have also evaluated the error rate, F1-score, accuracy, precision, and recall of the face sketch recognition system. This rigorous evaluation is crucial as it provides objective measures for assessing the system's performance in real-world scenarios, particularly in law enforcement applications. Given the obtained average values of the error rate, accuracy, precision, recall, and F1-score metrics as shown in *Figure 7*, it is evident that the proposed approach for face sketch recognition demonstrates exceptional performance. The error rate stands at an impressively low 0.7%, indicating the system's remarkable accuracy in matching face sketches to face images. A perfect precision of 100% reveals that whenever the system identifies a match, it does so correctly, minimizing false positives, which is especially critical in law enforcement scenarios. Additionally, the recall score of 100% ensures that the system successfully identifies all actual positive cases, ensuring that no relevant matches are missed. This outstanding balance between precision and recall is further confirmed by the F1-score of 100%, showcasing a robust overall performance. These results affirm the system's efficacy, making it a compelling choice for real-world applications, such as law enforcement, where precise suspect identification is imperative. The comprehensive evaluation conducted across different total used sketches adds further credibility to the approach's reliability, solidifying its potential as a powerful tool in face sketch recognition tasks.

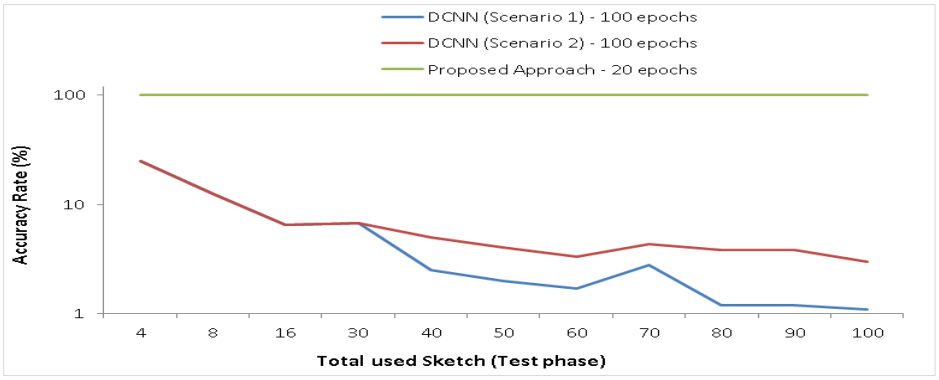

Figure 6 The accuracy rate obtained displayed in logarithmic scale

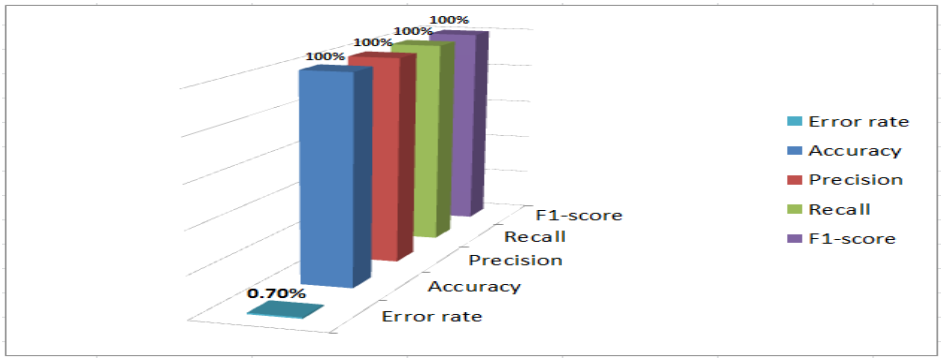

Figure 7 The obtained Error rate, Accuracy, Precision, Recall, and F1-score

5.Discussion

In scenario 1, where the CNN directly encounters the face sketch without any pre-processing, the low recognition rate can be attributed to a combination of factors. First, the CNN struggles to adaptively adjust its weight parameters ("w") to accommodate the inherent texture differences between the face sketch and the corresponding face photo. This results in suboptimal feature extraction and contributes to the reduced recognition performance. Secondly, the shortage of input samples available in face sketch databases poses a significant challenge. Limited data can lead to overfitting, hampering the CNN's ability to generalize effectively across diverse sketches and photos.

Moving on to scenario 2, despite employing a conversion technique to transform facial photos into composite portraits and minimize texture disparities, the results remain unsatisfactory. This outcome can be attributed to the underlying issue of the CNN's input sample size problem remaining unaddressed. While the conversion process aims to alleviate texture differences, it fails to tackle the core challenge of insufficient input data, which continues to hinder the CNN's capacity to learn effectively and discern variations in sketches.

Now, in our proposed solution, which introduces the crucial DST layer into the CNN architecture, the DST layer serves as a pivotal enhancement for multiple reasons. First and foremost, the DST layer effectively bridges the gap in texture between face sketches and corresponding face photos. By employing the shearlet transform, it captures essential high-frequency details that are crucial for accurate recognition. This addresses the inherent texture disparity challenge and ensures that both sketches and photos share closely aligned coefficients, making them amenable to direct comparison.

Furthermore, the integration of DST significantly augments the CNN's input sample size. The generated coefficients from the DST process contribute rich and diverse information, which translates into a more comprehensive dataset for the CNN's learning phase. This not only mitigates overfitting but also enhances the CNN's ability to extract discriminative features, ultimately leading to improved recognition accuracy.

It's important to note that the improvement achieved through the proposed approach is substantial. With just twenty epochs, the proposed solution outperforms the traditional CNN, which required a hundred epochs in scenarios 1 and 2 without achieving a satisfactory recognition rate. This efficiency is a direct result of the DST layer's capability to align input data, thereby accelerating the CNN's convergence rate and enhancing its learning process.

The proposed approach offers a significant advancement in facial sketch recognition. The DST layer effectively addresses the challenge of texture disparity while simultaneously enriching the CNN's input data. This holistic improvement leads to improved recognition performance, shortened training times, and increased adaptability to variations in face sketches and photos. The direct comparison of the proposed solution with the CNN in *Figure 6* underscores the added value of our approach and underscores its potential to revolutionize the field of facial sketch recognition.

Limitations:

While the proposed approach demonstrates promising advancements in the realm of facial sketch recognition, it's essential to acknowledge its inherent limitations. One notable constraint is the reliance on the quality of the initial face sketches. The effectiveness of the DST layer and the subsequent CNN learning heavily depends on the accuracy and quality of the input sketches. In cases where the sketches are distorted, incomplete, or of low resolution, the DST layer may struggle to extract meaningful coefficients, thereby affecting the overall recognition performance. Additionally, the proposed approach's success is closely linked to the quality and adaptability of the shearlet transform itself. Variations in transform parameters or its sensitivity to noise could potentially introduce artifacts or inaccuracies in the generated coefficients, leading to suboptimal recognition results. Furthermore, while the DST layer addresses texture disparities, it may not completely eliminate variations stemming from other factors such as lighting conditions, pose changes, and artistic styles in sketches. Lastly, although the integration of DST enhances the CNN's input sample size, the approach may still face challenges in scenarios with extremely limited sketch data availability. Despite these limitations, the proposed approach represents a substantial leap forward in the domain of facial sketch recognition and lays the groundwork for further research and innovation to overcome these challenges. A complete list of abbreviations is shown in *Appendix I.*

6.Conclusion and future work

The importance of face sketch recognition and its applications, particularly in fields such as criminology, along with the limited contributions in this area, served as the motivation behind our proposal for a solution that guarantees a high recognition rate. Initially, it was essential to clearly define the limitations and constraints associated with using CNN in this context. This step allowed us to pinpoint two key constraints that significantly affect the CNN's performance. Building on this comprehension, a promising and effective solution that incorporates a CNN with an additional pre-processing layer based on DST was introduced. The results achieved undeniably demonstrate the effectiveness of our proposed solution. In the future, additional studies will be needed to explore different approaches and conduct comparative studies and investigations in this area.

Acknowledgment

None.

Conflicts of interest

The authors have no conflicts of interest to declare.

Author's contribution statement

Chaymae Ziani: Conceptualization, investigation, data curation, writing – original draft, writing – review and editing. **Abdelalim Sadiq:** Study conception, supervision, investigation on challenges and draft manuscript preparation.

References

[1] Liu L, Shen F, Shen Y, Liu X, Shao L. Deep sketch hashing: fast free-hand sketch-based image retrieval. In proceedings of the conference on computer vision and pattern recognition 2017 (pp. 2862-71). IEEE.

[2] Cao B, Wang N, Li J, Hu Q, Gao X. Face photo-sketch synthesis via full-scale identity supervision. Pattern Recognition. 2022; 124:108446.

[3] Shan C, Gong S, Mcowan PW. Facial expression recognition based on local binary patterns: a comprehensive study. Image and Vision Computing. 2009; 27(6):803-16.

[4] Yu Q, Liu F, Song YZ, Xiang T, Hospedales TM, Loy CC. Sketch me that shoe. In proceedings of the IEEE conference on computer vision and pattern recognition 2016 (pp. 799-807).

[5] Ziani C, Sadiq A. SH-CNN: shearlet convolutional neural network for gender classification. Advances in Science, Technology and Engineering Systems Journal. 2020; 5(20):1328-34.

[6] Ziani C, Sadiq A. Smart approach based on CNN and shearlet transform for age prediction. In proceedings of seventh international congress on information and communication technology: ICICT, London, 2022 (pp. 143-52). Singapore: Springer Nature Singapore.

[7] Kutyniok G, Lim WQ. Image separation using wavelets and shearlets. In curves and surfaces: 7th international conference, Avignon, France, 2012 (pp. 416-30). Springer Berlin Heidelberg.

[8] Mutneja V, Singh S. HAAR-features training parameters analysis in boosting based machine learning for improved face detection. International Journal of Advanced Technology and Engineering Exploration. 2021; 8(80):919-31.

[9] Ashhar SM, Mokri SS, Abd RAA, Huddin AB, Zulkarnain N, Azmi NA, et al. Comparison of deep learning convolutional neural network (CNN) architectures for CT lung cancer classification. International Journal of Advanced Technology and Engineering Exploration. 2021; 8(74):126-34.

[10] Salunke D, Mane D, Joshi R, Peddi P. Customized convolutional neural network to detect dental caries from radiovisiography (RVG) images. International Journal of Advanced Technology and Engineering Exploration. 2022; 9(91):827-38.

[11] Senthil T, Rajan C, Deepika J. An efficient CNN model with squirrel optimizer for handwritten digit recognition. International Journal of Advanced Technology and Engineering Exploration. 2021; 8(78):545-59.

[12] Patel LK, Patel MI. Feature based image registration using CNN features for satellite images having varying illumination level. International Journal of Advanced Technology and Engineering Exploration. 2023; 10(101):440-57.

[13] Chopade PB, Prabhakar N. Human emotion recognition based on block patterns of image and wavelet transform. International Journal of Advanced Technology and Engineering Exploration. 2021; 8(83):1394-409.

[14] Bahrum NN, Setumin S, Abdullah MF, Maruzuki MI, Che AAI. A systematic review of face sketch recognition system. Journal of Electrical and Electronic Systems Research (JEESR). 2023; 22:1-10.

[15] Bae S, Din NU, Park H, Yi J. Face photo-sketch recognition using bidirectional collaborative synthesis network. In 16th international conference on ubiquitous information management and communication 2022 (pp. 1-8). IEEE.

[16] Radman A, Sallam A, Suandi SA. Deep residual network for face sketch synthesis. Expert Systems with Applications. 2022; 190:115980.

[17] Yan L, Zheng W, Gou C, Wang FY. IsGAN: Identity-sensitive generative adversarial network for face photo-sketch synthesis. Pattern Recognition. 2021; 119:108077.

[18] Bhoir M, Gosavi C, Gade P, Alte B. A decision-making tool for creating and identifying face sketches. In ITM web of conferences 2022 (pp. 1-6). EDP Sciences.

[19] Alhashash KM, Samma H, Suandi SA. Fine-tuning of pre-trained deep face sketch models using smart

switching slime mold algorithm. Applied Sciences. 2023; 13(8):1-36.
[20] Navuluri C, Jukanti S, Allapuram RR. Semantic neural model approach for face recognition from sketch. arXiv preprint arXiv:2305.01058. 2023.
[21] Peng C, Zhang C, Liu D, Wang N, Gao X. Face photo–sketch synthesis via intra-domain enhancement. Knowledge-Based Systems. 2023; 259:110026.
[22] Zhong K, Chen Z, Liu C, Wu QJ, Duan S. Unsupervised self-attention lightweight photo-to-sketch synthesis with feature maps. Journal of Visual Communication and Image Representation. 2023; 90:103747.
[23] Wan W, Gao Y, Lee HJ. Transfer deep feature learning for face sketch recognition. Neural Computing and Applications. 2019; 31:9175-84.
[24] Lakshmi N, Arakeri MP. A novel sketch based face recognition in unconstrained video for criminal investigation. International Journal of Electrical & Computer Engineering (2088-8708). 2023; 13(2):1499-1509.
[25] Devakumar S, Sarath G. Forensic sketch to real image using DCGAN. Procedia Computer Science. 2023; 218:1612-20.
[26] Jacquet M, Champod C. Automated face recognition in forensic science: review and perspectives. Forensic Science International. 2020; 307:110124.
[27] Kazemi H, Soleymani S, Dabouei A, Iranmanesh M, Nasrabadi NM. Attribute-centered loss for soft-biometrics guided face sketch-photo recognition. In proceedings of the conference on computer vision and pattern recognition workshops 2018 (pp. 499-507). IEEE.
[28] Iranmanesh SM, Kazemi H, Soleymani S, Dabouei A, Nasrabadi NM. Deep sketch-photo face recognition assisted by facial attributes. In 9th international conference on biometrics theory, applications and systems 2018 (pp. 1-10). IEEE.
[29] Liu D, Gao X, Wang N, Li J, Peng C. Coupled attribute learning for heterogeneous face recognition. IEEE Transactions on Neural Networks and Learning Systems. 2020; 31(11):4699-712.
[30] Liu D, Gao X, Wang N, Peng C, Li J. Iterative local re-ranking with attribute guided synthesis for face sketch recognition. Pattern Recognition. 2021; 109:107579.
[31] Peng C, Wang N, Li J, Gao X. DLFace: deep local descriptor for cross-modality face recognition. Pattern Recognition. 2019; 90:161-71.
[32] Fan L, Sun X, Rosin PL. Attention-modulated triplet network for face sketch recognition. IEEE Access. 2021; 9:12914-21.
[33] Easley GR, Labate D, Patel VM. Directional multiscale processing of images using wavelets with composite dilations. Journal of Mathematical Imaging and Vision. 2014; 48:13-34.
[34] Viola P, Jones M. Rapid object detection using a boosted cascade of simple features. In proceedings of the computer society conference on computer vision and pattern recognition. CVPR 2001. IEEE.
[35] Bradski G, Kaehler A. Learning OpenCV: computer vision with the OpenCV library. O'Reilly Media, Inc.; 2008.
[36] Wang X, Tang X. Face photo-sketch synthesis and recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2008; 31(11):1955-67.

Chaymae Ziani is a Ph.D. student at Ibn Tofail University, Faculty of Sciences, Kénitra, Morocco. She possesses a degree in Mathematics and Computer Science, followed by a Master's degree in Computer Engineering. Her specialization lies in the fields of Artificial Intelligence, Image Processing, Computer Vision, and IoT. With a strong passion for research, she actively engages in presenting and publishing numerous research papers within her areas of expertise. Throughout her academic journey, Chaymae has consistently demonstrated a profound dedication to advancing knowledge and understanding in the technology field. Her research primarily focuses on harnessing the potential of Artificial Intelligence, particularly in the realms of image processing and computer vision. Her goal is to develop innovative solutions and algorithms capable of addressing real-world challenges and enhancing various applications within these domains. With a meticulous attention to detail and an unwavering commitment to excellence, Chaymae has made meaningful contributions to the academic community by disseminating her findings through presentations at conferences and symposiums. Furthermore, her research work has garnered recognition and acknowledgment through publications in reputable journals and conference proceedings. Looking towards the future, Chaymae aspires to make substantial contributions to the fields of technology and AI. She aims to leverage her expertise and research findings to tackle pressing challenges and create positive societal impact. Through her work, she seeks to bridge the gap between academia and industry, advocating for the adoption of advanced technologies to solve real-world problems.
Email: chaymae.ziani@uit.ac.ma

Dr. Abdelalim Sadiq is a Full Professor of Computer Science at the Faculty of Sciences, Ibn Tofail University, situated in Kenitra, Morocco. He holds the prestigious position of head of the SIM TEAM at the MISC Laboratory and serves as the coordinator of the Master's program in Software Engineering for Cloud Computing. He earned his Ph.D. in Computer Engineering from the Higher National School of Computer Science and System Analysis (ENSIAS) at the University of Rabat, Morocco, in 2007. With a rich academic background, he has solidified his reputation as a highly accomplished researcher and educator. Dr. Abdelalim Sadiq's research interests span various domains within computer science, with a specific

focus on computer vision, pattern recognition, face recognition, face expression analysis, action recognition, sentiment analysis, and opinion mining. He has made significant contributions to these fields through his extensive research efforts and scholarly publications. As a respected academic and researcher, he actively engages in conferences, workshops, and seminars to disseminate his knowledge and findings to the scientific community. His expertise and contributions have garnered recognition and respect within both national and international research circles. In addition to his research and academic commitments, Dr. Abdelalim Sadiq is dedicated to mentoring and guiding students in their academic pursuits.
Email: a.sadiq@uit.ac.ma

Appendix I

S. No.	Abbreviation	Description
1	2SMA	Smart Switching Slime Mould Algorithm
2	CAGTL	Coupled Attribute-Guided Triplet Loss
3	CNN	Convolutional Neural Networks
4	CUHK	Chinese University of Hong Kong
5	CUFS	CUHK Face Sketch
6	CUFSF	CUHK Face Sketch Feret
7	DCGAN	Deep Convolutional Generative Adversarial Network
8	DCNN	Deep Convolutional Neural Networks
9	DST	Discrete Shearlet Transform
10	GANs	Generative Adversarial Networks
11	IDE	Intra-Domain Enhancement
12	IsGAN	Identity-sensitive Generative Adversarial Network
13	ReLU	Rectified Linear Unit
14	SIFT	Scale-Invariant Feature Transform
15	SLR	Score-based Likelihood Ratio
16	SSIM	Structural SIMilarity Index
17	USAL	Unsupervised Self-Attention Lightweight
18	XM2VTS	Multi Modal Verification for Teleservices and Security