

## Article

# A Decision Support System for Face Sketch Synthesis Using Deep Learning and Artificial Intelligence

Irfan Azhar <sup>1</sup>, Muhammad Sharif <sup>1,\*</sup>, Mudassar Raza <sup>1</sup>, Muhammad Attique Khan <sup>2</sup> and Hwan-Seung Yong <sup>3</sup>

<sup>1</sup> Department of Computer Science, COMSATS University Islamabad, Wah Campus, Wah Cantt 47040, Pakistan; irfanazhar681@gmail.com (I.A.); mudassarr@gmail.com (M.R.)

<sup>2</sup> Department of Computer Science, HITEC University Taxila, Taxila 47080, Pakistan; attique.khan@hitecuni.edu.pk

<sup>3</sup> Department of Computer Science & Engineering, Ewha Womans University, Seoul 03760, Korea; hsyong@ewha.ac.kr

\* Correspondence: sharif@ciitwah.edu.pk

**Abstract:** The recent development in the area of IoT technologies is likely to be implemented extensively in the next decade. There is a great increase in the crime rate, and the handling officers are responsible for dealing with a broad range of cyber and Internet issues during investigation. IoT technologies are helpful in the identification of suspects, and few technologies are available that use IoT and deep learning together for face sketch synthesis. Convolutional neural networks (CNNs) and other constructs of deep learning have become major tools in recent approaches. A new-found architecture of the neural network is anticipated in this work. It is called Spiral-Net, which is a modified version of U-Net to perform face sketch synthesis (the phase is known as the compiler network C here). Spiral-Net performs in combination with a pre-trained Vgg-19 network called the feature extractor F. It first identifies the top  $n$  matches from viewed sketches to a given photo. F is again used to formulate a feature map based on the cosine distance of a candidate sketch formed by C from the top  $n$  matches. A customized CNN configuration (called the discriminator D) then computes loss functions based on differences between the candidate sketch and the feature. Values of these loss functions alternately update C and F. The ensemble of these nets is trained and tested on selected datasets, including CUFS, CUFSE, and a part of the IIT photo-sketch dataset. Results of this modified U-Net are acquired by the legacy NLDA (1998) scheme of face recognition and its newer version, OpenBR (2013), which demonstrate an improvement of 5% compared with the current state of the art in its relevant domain.

**Keywords:** smart cities; sketch synthesis; convolutional neural network; Vgg-19 net; U-Net; Spiral-Net; face recognition; NLDA; OpenBR



**Citation:** Azhar, I.; Sharif, M.; Raza, M.; Khan, M.A.; Yong, H.-S. A Decision Support System for Face Sketch Synthesis Using Deep Learning and Artificial Intelligence. *Sensors* **2021**, *21*, 8178. <https://doi.org/10.3390/s21248178>

**Academic Editors:** Robertas Damaševičius, M. Poongodi, Hafiz Tayyab Rauf and Hasan Ali Khattak

Received: 3 November 2021

Accepted: 30 November 2021

Published: 8 December 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The Internet of Things (IoT) [1] has been playing a key role in the smart city sector, for example, in the security of smart homes, where you, using your smartphone, can decide who can enter your home [2]. Through IoT technology, it is easy to monitor your home at any time from anywhere, and this process helps to develop efficient, safer smart cities [3]. The IoT technology integration with IT devices helps to ease the investigation process, especially in the identification of people [4,5]. A very few studies are available on how IoT and information technology (IT) techniques work together [6]. The major applications where these technologies work together are biometric [7], video surveillance [8], Internet of Vehicles [9], and biomedical [10,11].

The formulation of face sketches based on learning from the reference photos and their corresponding forensic sketches has been an active field since the last two decades [12,13]. It helps the law enforcement agencies in the search, isolation, and identification of suspects by enabling them to match sketches against possible candidates from the mug-shot

library [14–16] and/or photo dataset of the target population [17,18]. Forensic or artist sketches are also used in animated movies and/or during the development of CGI-based segments [19]. Presently, many persons like to use a sketch in place of a personal picture as an avatar or a profile image. Therefore, a ready-made scheme to furnish a sketch from a personal picture, without involving a skilled sketch artist, would come in handy [20]. Since 2004, exemplar-based techniques incorporating patch-matching algorithms have been most popular. Photos and corresponding sketches were identically divided into a mosaic of overlapping patches. For each patch of the photo, its nearest patch in all training sketches according to a given property, for example, the Markov random field (MRF) [21], the Markov weight field (MWF), or spatial sketch denoising (SSD), was searched for and marked. This principle was applied successively to all photos and sketches in the training set. Hence, a dictionary was developed. For each test photo patch, a suitable patch was first searched for in the photo and its corresponding patch in the dictionary was selected as part of the resulting sketch [22]. On completion of this search, a resulting sketch was formulated. In previous research, much effort has been devoted to reducing the time spent on and resource overheads of these methods to effectively produce a sketch. Those algorithms did not focus on capturing the subtle non-linearity between the original photo and the forensic sketch. Their results were, however, only reliable for a dataset of subjects devoid of the diversity of ethnicity; age; facial hair; and external elements, such as earrings, glasses, and hairpins. While those methods could replicate major features of the test photo, they did not reproduce minor details, such as contours of the cheekbones, edges of mustaches/beards/hairstyles, or clear outlines of eyeglasses. Lately, neural networks and other tools of deep learning have been employed to learn about the correspondence between photo–sketch pairs, and they try to reproduce intricate features of the photo in the resulting sketch. These methods also have their small inadequacies. Simple CNN-based methods produce sketches that lack sharpness and focus [23,24]. On the contrary, GAN-based methods do produce clear sketches but they are incomplete concerning the outline of the test subject’s photo. This paper includes the following:

- A novel/modified structure of a residual network with skip connections forming a spiral-like shape to act as a compiler entity in the proposed face sketch synthesis phase. The overall scheme is motivated by [25], and a similar approach is presented in [26].
- A pre-trained Vgg-19 network is used to help accomplish the exemplar-based technique of selecting the best possible candidate from the viewed sketches during the training process. This part relies upon the distribution of the input photo into a mosaic of overlapping patches and identical division of the sketches in the reference set.
- The patches are selected by the minimal cosine distance, and a candidate feature map of the sketch is formulated.
- The feature sketch and the raw sketch by the compiler network are then compared through a customized convolutional neural network applying the MSE loss function to render a perceptual loss that monitors the training of the compiler network.
- The adversary loss function is also used to give sharpness to the resulting sketches.

The rest of the paper is arranged in this sequence: Section 2 covers the previous and current works related to the proposed model. Section 3 describes the composition detail of the suggested network. Section 4 provides implementation details and discusses the evaluation and analysis of results. Section 5 gives the conclusion.

## 2. Related Work

The Internet of Things (IoT) and machine learning have shown improved performance in many applications, such as facial recognition, biometrics, and surveillance [27,28]. Recently, the blockchain-based multi-IoT method was presented by Jeong et al. [29]. The presented method works in two layers (layer and layer) with the help of the blockchain technology. Through these layers, information is sent to and received from local IoT groups in more secure ways. Another multi-IoT method was presented by [30] for anomaly detection. They introduced forward and inverse problems to investigate the dependency of the

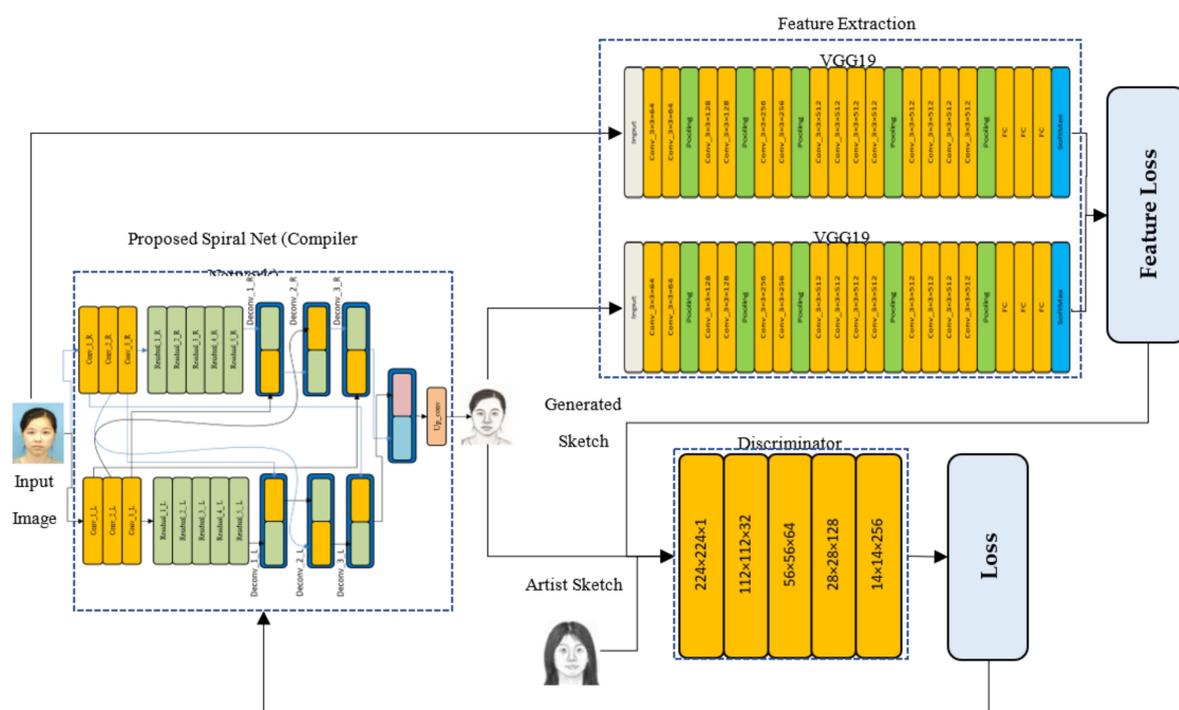
inter-node distance and the size of the IoT network. A new paradigm, named social IoT, was presented by Luigi et al. [31] for the identification of useful guidelines for institution and social management. Khammas et al. [32] presented a cognitive IoT approach to human activity diagnosis. In cognitive computing, the cognitive IoT is the next step to improving the accuracy and reliability of the system. An IoT-based biometric security system was presented by Bobby et al. [11]. In this system, the IoT allows the multiple sensors and scanners to interact with human beings.

The recent developments in the CNN for scene recognition [33], object recognition [34], and action recognition [35] have produced an impressive performance [36]. Tang and Wang [37] introduced in their seminal work a new art of formulating human face sketches based on Eigen transformation. The work is based on pairs of photos and their corresponding viewed sketches. They developed a correlation between input photos and training photos in the Eigenspace. Then, using this correlation, they proposed to construct a sketch from the Eigenspace of the training sketches. Liu et al. [38] proposed the non-linear model of sketch formulation based on locally linear embedding (LLE). In this model, the input photo is divided into overlapping patches. Then, each patch is reshaped by a linear combination of training patches. The same relationship of photo patches was used to formulate respective patches of the resulting sketch. Tang and Wang [39] used Markov random fields (MRF) in the selection of neighboring patches and to improve their relationship. Zhou et al. [40] proposed a model of sketch generation that further builds upon the MRF model. They added weights to linear combinations of best possible candidate patches, and it was called the Markov weight field (MWF). Song et al. [17] presented a model based on spatial sketch denoising (SSD). Gao et al. [41] proposed an adaptive scheme based on the practical benefits of sparse representation theory, and it was called the SNS-SRE method, which relates to sparse neighbor selection and sparse-representation-based enhancement. Wang et al. [42] formulated a solution of neighbor selection by building up a dictionary based on a random sampling of the training photos and sketches. This model was called random sampling and locality constraint (RSLCR). Akram et al. [43] carried out a comparative study of all basic methodologies of the exemplar-based approach as well as two newer methods of sketch synthesis, called FCN [44] and GAN [45], which are based on the convolutional neural network and generative adversarial networks, respectively. The last two works may be included among the pioneer efforts of “learning-based” algorithms of sketch synthesis. Zhang et al. [46] introduced a model to address the problems of texture loss of the FCN setup. Their scheme consisted of two-branched FCN. One computed a content image, and the second branch calculated the texture of the synthesized sketch. This model also inherited the inadequacy of distorted sketches since the two-branched network could not present a well-unified output. Wang et al. [47] proposed a model to generate sketches from training photos and photos from the training sketches by employing a multiscale generative adversarial network. Wang et al. [48] proposed a model of anchored neighborhood index (ANI) that incorporated correlation of photo patches as well as sketch patches during sketch formulation. Moreover, similar to RSLCR, this algorithm also benefited from the development of an off-line dictionary to reduce computational overheads during the testing phase. Jiao et al. [49] presented a deep learning method based on a small CNN and a multilayer perceptron. This work was successful in imparting continuous and faithful facial contours of the input photo to its resulting sketch. Zhang et al. [50] proposed a model based on adversarial neural networks that learned in photo and sketch domains with help of intermediate entities called latent variables. Synthesized sketches of this model bear improvement against blurs and shape deformations. Zhang et al. [51] proposed a model called dual transfer face sketch-photo synthesis (FSPS). It is based on CNN and GAN and realizes inter-domain and intra-domain information transfer to formulate a sketch from the training pairs of photo-viewed sketches. Lin et al. [52] and Fang et al. [53] presented individual works based on neural networks for face-sketch formulation involving the identity of each subject photo. Yu et al. [54] proposed a model to synthesize sketches from photos by GAN that is assisted by composition information of the input photos.

Their work removed blurs and spurious artifacts from the result sketches. Similarly, Lin et al. [55] presented a model to synthesize de-blurred sketches by deep CNN focusing on the estimation of motion blur. Zhu et al. [56] presented a model involving three GANs, in which each network gains knowledge of the photo–sketch pairs and imparts the learned characteristics to resulting sketches directly by a teacher GAN or by the comparison of the two student GANs. Radman et al. [57] proposed a sketch synthesis scheme based on the bidirectional long-short term memory (BL-STM) recurrent neural network.

### 3. Materials and Methods

The proposed framework comprises two neural nets. The first part is a compiler network  $C$ , which is based upon a residual network of two branches, and the skip connections are made in a spiral fashion. It is derived from [58], which was employed for neural-style transfer. For an input photo  $p$ , this part generates a raw sketch named  $s$ . The second part of the scheme is a feature-extractor called  $F$ , based on a pre-trained Vgg-19 network [59]. These net and associated components formulate another intermediate entity, called feature-sketch  $f$ . This composition is shown in Figure 1. The last step of the setup is a customized convolutional neural network, called discriminator  $D$ , to undertake a comparison between raw sketch  $s$  and feature sketch  $f$ . Their difference, combined with other loss functions, is then used to modify the weights of  $C$  and  $D$  networks iteratively during the training process. At end of the training, the  $C$  network is solely used to synthesize automated sketches from the test photos.



**Figure 1.** Schematic diagram of the proposed method.

*Phase-1. Treatment of Images:* Photos/sketches of CUHK and AR datasets are already aligned, and they are of size  $250 \times 200$  pixels. Therefore, they do not need any pre-processing. Photos and viewed sketches of XM2VTS and CUFSF datasets were not aligned. The following operations are executed upon the photos/sketches:

- Sixty-eight face landmarks on the image are detected by the dlib1\* library.
- The image is rescaled in a manner that the two eyes are located at (75; 125) and (125; 125), respectively.
- The resulting image is cropped to a size of  $250 \times 200$ .

*Phase-2. Development of Feature Dictionary:* Patch matching is a time-consuming process. In addition, as already shown by the exemplar-based approaches, the computation of features for patches is resource intensive when conducted at run-time. Therefore, a dictionary of features of patches for all the images, including photos and viewed sketches in the reference set, is pre-computed and stored as a reference bank. Moreover, the entire length of reference sketches is not searched for a possible match. Instead, initially  $n$  top suitable candidate sketches to each input photo are selected at run-time based on their cosine distance at ReLU-5-1 features of the Vgg-19 net. Patch matching is then restricted within these  $n$  reference photos ( $n = 5$  was used in all training runs of all iterations).

### 3.1. Compiler Network C

This network is composed of two identical strains, and each strain is composed of three stages. The first part consists of convolutional layers; it has residual blocks in the middle section and up-sampling layers in the end part. The structure is shown in Figure 2. It is a modified form of U-Net proposed by [58] for image style transfer and super-resolution. To introduce diversity and depth in the network, in a novel fashion, the skip connections in this model are added to an alternate strain instead of the original line. Therefore, each stage of the network on the left side is connected to the corresponding stage on the right side and vice versa. The resulting shape looks similar to a spiral and, therefore, this construct is called Spiral-Net. Skip connections are added in this manner to (a) increase the width of each layer of the net, (b) augment feature matrices at different layers with new feature values from the other strain, and (c) populate feature matrices at different layers such that any half of the matrix vanishing due to ReLU and pooling operations may be repopulated with feature values. The last objective breaks any build-up of monotonous behavior due to ReLU and pooling operations. The compiler network C is a decisive module of this framework, and it plays major role during the implementation and operation phases. During the training phase, the training photo images are fed to this network and a pseudo sketch is formulated at its end. This sketch is further compared by the remaining parts of the overall scheme. Similarly, during the testing phase, a test photo is input to this network and its output is a synthesized sketch.

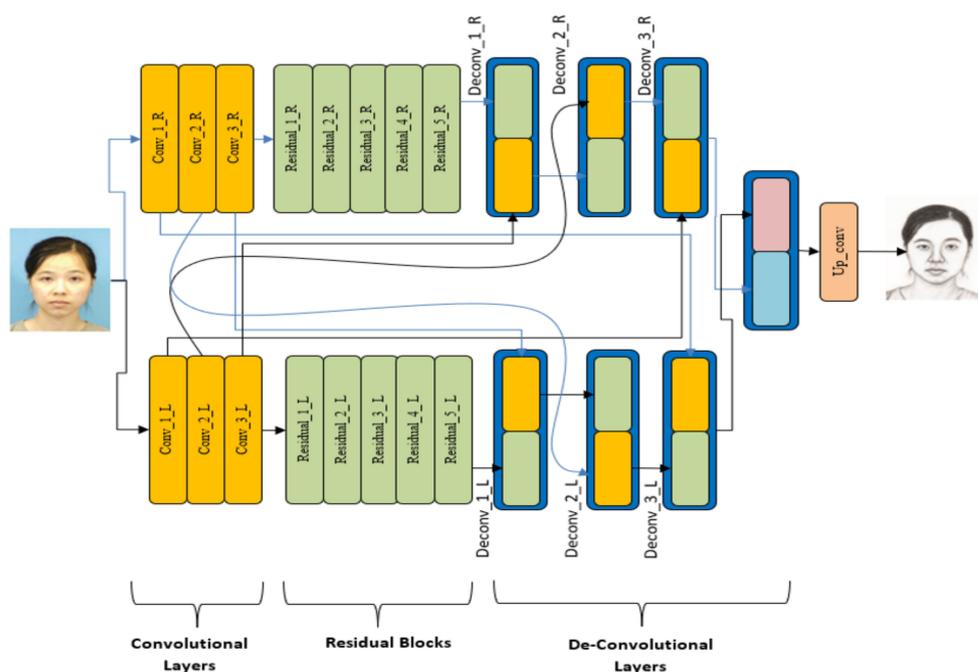


Figure 2. Architecture of Spiral-Net as a sketch compilation network.

### 3.2. Feature Extractor F

A pre-trained model of Vgg-19 is used to extract features of the top  $n$  candidates of viewed sketches from the reference dataset for each train photo, where  $n$  can be set to any value, preferably between 5 and 10. Then input photos and the sketches are divided into identical maps/matrices of overlapping patches. An exemplar approach of the Markov random fields from [60] is preferred here, and it dictates that for each patch of the input photo, any of the candidate patches from the five sketches are selected based upon the shortest distance. This procedure is repeated from the first to the last patch of the input photo. Hence, F shapes up corresponding patches in a proper sequence to yield a feature map that is a representation of the intermediate sketch and is not exactly an image. It is used for comparison with the output of the compiler C through the discriminator D. The loss functions based on these comparisons are used to alternately update C and D networks.

Consider the given dataset as a universal set  $\mathcal{R}$  composed of  $p$  photos and  $s$  sketches, where  $\mathcal{R} = \{(p_i^{\mathcal{R}}, s_i^{\mathcal{R}})\}_{i=1}^N$ . Here,  $N$  is the total number of photo-sketch pairs in the dataset. F aims at formulating a feature map  $\theta^l(p)$  for the input photo  $p$ .  $\theta^l(p)$  is used to augment the synthesis of the sketch  $\hat{s}$ . The MRF principle of [39] is applied to compose a local patch representation of  $p$ . It consists of the following stages:

- To begin with,  $p$  is input to the pre-trained Vgg-19 net.
- The feature map  $\theta^l(p)$  is extracted at the  $l$ -th layer, where  $l = (1, 2, 3, 4, 5)$ , corresponding to  $(relu1 - 1, relu2 - 1, relu3 - 1, relu4 - 1, relu5 - 1)$  of F.
- A dictionary/look-up repository of reference representations is built for the entire dataset in the form of  $\{\theta^l(p_i^{\mathcal{R}})\}_{i=1}^N$  and  $\{\theta^l(s_i^{\mathcal{R}})\}_{i=1}^N$ .
- Let us assume an  $r \times r$  patch centered at point  $j$  of  $\theta^l(p)$  as  $T = \Omega_j(\theta^l(p))$ . Let us also assume corresponding patches  $P = \Omega_j(\theta^l(p_i^{\mathcal{R}}))$  and  $S = \Omega_j(\theta^l(s_i^{\mathcal{R}}))$  from the entire dataset.
- For every patch  $T_j$ , where  $j = \{1, 2, 3, \dots, u\}$  and  $u$  is explained by the relation  $u = (H^l - r) * (W^l - r)$ , where  $H^l$  and  $W^l$  are the height and the width of the map  $\theta^l(p)$ , respectively, we find its closest patch  $P_{j'} = \Omega_{j'}(\theta^l(p_{i'}^{\mathcal{R}}))$  from the look-up repository or dictionary based on the cosine distance.
- The cosine distance is defined with the help of Equation (1).

$$(i', j') = \frac{T_j * P_{j'}}{|T_j|_2 * |P_{j'}|_2} \quad (1)$$

$$(i', j') = \operatorname{argmax}_{\substack{i^*=1 \sim N \\ j^*=1 \sim m}} \frac{[\Omega_j(\theta^l(p)) \cdot \Omega_{j^*}(\theta^l(p_{i^*}^{\mathcal{R}}))]}{\|\Omega_j(\theta^l(p))\|_2 \cdot \|\Omega_{j^*}(\theta^l(p_{i^*}^{\mathcal{R}}))\|_2} \quad (2)$$

- Photos and sketches are aligned in the reference set. We index directly the corresponding feature patches  $M_{j'} = \Omega_{j'}(\theta^l(s_{i'}^{\mathcal{R}}))$  for identified patches  $P_{j'} = \Omega_{j'}(\theta^l(p_{i'}^{\mathcal{R}}))$  by Equation (2).
- Successively,  $M_{j'} = \Omega_{j'}(\theta^l(s_{i'}^{\mathcal{R}}))$  is used in place of every  $T_j = \Omega_j(\theta^l(p_j))$  to formulate a complete feature representation or the feature sketch at given layer  $l$ . Therefore,  $F = \Omega_{j'}\{\theta^l(s)\}_{j=1}^u$ .

### 3.3. Discriminator D

It is a basic convolutional network composed of six layers. Outputs of C and F networks are input to this net. This error, in addition to the other factors discussed later, is used to train the C network.

### 3.4. Loss Function

**Feature Loss:** The difference between the raw sketch  $s$  and the feature map  $f$  is expressed by a feature loss.

$$F_p = \sum_{l=3}^5 \sum_{j=1}^m \left[ \Omega_j(\theta^l(\hat{s})) - \Omega_j'(\theta^l(p)) \right]_2^2 \quad (3)$$

where  $l = 3, 4, 5$  refers to layers relu3-1, relu4-1, and relu5-1, respectively. High-level features after relu3 1 are better representations of textures and more robust against appearance changes and geometric transforms [60]. Features of the initial stages, such as relu1-1 and relu2-1, do not contribute to sketch textures well. Features extracted at a higher stage of the network, e.g., relu5-1, can better preserve textures. As a trade-off,  $r = 3, 4, 5$  is set to improve the performance of the setup and to decrease the computational overhead cost of patch matching procedures.

**GAN Loss:** The least-squares loss was employed when training the neural networks of the proposed setup. It is called LSGAN according to [61]. Equations (4) and (5) give the mathematical relationship of loss parameters/terms.

$$E_{GAN_D} = \frac{1}{2} \mathbb{Q}_{s \sim F_{sketch(s)}} \left[ (D(s) - 1)^2 \right] + \frac{1}{2} \mathbb{Q}_{p \sim F_{photo(p)}} \left[ (D(G(p)))^2 \right] \quad (4)$$

$$E_{GAN_G} = \frac{1}{2} \mathbb{Q}_{p \sim F_{photo(p)}} \left[ (D(G(p)) - 1)^2 \right] \quad (5)$$

**Total Variation Loss:** Sketches generated by a CNN network, used here as the discriminator  $D$ , may be noisy; and they may also contain unwanted artifacts. Therefore, according to previous studies [58,60,62], the total variation loss term was used. It was included to offset the possibility of noise and to improve the quality of the sketch. Its relationship is given by Equation (6).

$$E_{tv}(\hat{s}) = \sum_{x,y} \left( (\hat{s}_{x+1,y} - \hat{s}_{x,y})^2 + (\hat{s}_{x,y+1} - \hat{s}_{x,y})^2 \right) \quad (6)$$

Here,  $\hat{s}_{x,y}$  denotes the intensity value at  $(x, y)$  of the synthesized sketch  $\hat{s}$ .

$$E_G = \delta_p F_p + \delta_{adv} E_{GAN_G} + \delta_{tv} E_{tv} \quad (7)$$

$$E_D = E_{GAN_D} \quad (8)$$

## 4. Results

In this section, a detailed account of the implementation scheme is given. Moreover, it mentions the quality parameters used during this project and, finally, it elaborates upon the evaluation of the performance of the proposed and reference methods.

### 4.1. Datasets

Initially, two public datasets, namely CUFSS and CUFSSF [63], were employed. Then, the implementation was repeated with the augmentation of these two datasets by part of another set, called DIIT [64]. The details of repeated implementation are provided in Section 4.8 and onward. The composition and training–testing split of these datasets is given in Table 1. CUFSSF is more challenging since its photos were captured under different lighting conditions and its viewed sketches show deformations in shape versus the original photos to mimic inherent properties of forensic sketches.

**Table 1.** Details of initial datasets.

Dataset		Total Pairs	Train	Test
CUFS	CUHK [37]	188	88	100
	AR [65]	123	80	43
	XM2VTS [66]	295	100	195
CUFSF		1194	250	944
Total Pairs		1800	518	1282

#### 4.2. Performance Measures

This section describes those parameters that were selected to gauge the performance of existing and proposed methodologies.

**Structure Similarity Index:** The SSIM [67] gives a measure of visual similarity between two images. It is included here due to its prevalent use in state of the art, but we did not rely upon it as the decisive factor. The mathematical relationship of the SSIM is reproduced here, as Equation (9), from [67]. The value of the SSIM varies between  $-1$  (for totally different inputs) and  $+1$  (for completely identical inputs). Generally, an average value of SSIM scores for respective techniques over a specific dataset is computed to enable their direct comparison with each other.

$$SSIM(P, Q) = \frac{(2h_p h_Q + K_1)(2Z_{PQ} + K_2)}{(h_p^2 + h_Q^2 + K_1)(Z_P^2 + Z_Q^2 + K_2)} \quad (9)$$

**Feature Similarity Index:** The FSIM [68] is a measure of perceptual similarity between two images. It is based upon phase congruence and gradient computations and their comparison in respect of the given images. The FSIM is considered here as a reliable measure of similarity between synthesized sketches and their viewed sketch counterparts. The Feature Similarity Index (FSIM) [68] is a quality metric for two images based on their respective frequency dynamics, called phase congruency (PC), which is then scaled by the gradient magnitude (GM) of light variations of sharp edges at the feature boundaries. It is based on the premise that the human vision system (HVS) is more susceptible to frequency variations (PC) of low-level features in the given image. PC is, however, contrast invariant, whereas information of color or contrast affects the HVS perception of image quality. Therefore, the image gradient magnitude (GM) is employed as the second feature in the FSIM. Inherently, the FSIM is largely invariant to magnitude diversity.

PC and GM play complementary roles in characterizing the image's local quality. PC is a dimensionless parameter defining a local structure. The GM is computed by any of the convolutional masks, such as Sobel, Prewitt, or any other gradient operator. The SSIM compares two images based on their luminance components only, while the FSIM considers the chromatic information in addition to the luminance of colored images.

The FSIM is computed by the following relations according to [66]:  $p(x)$  and  $q(x)$  are two images.  $PC_p$  and  $PC_q$  are their phase congruency maps, and  $G_p$  and  $G_q$  are their gradient magnitudes, respectively.  $Sim_{PC}$  is the similarity between these two images at point  $x$ , given by Equation (10) here.  $Sim_G$ , as mentioned in Equation (11), is their similarity based on the GM only, and  $Sim_L$  is their combined similarity at the point of consideration.  $Sim_L$  is measured by the relation given in Equation (12).

$$Sim_{PC}(x) = \frac{2PC_p(x).PC_q(x) + C_1}{PC_p^2(x) + PC_q^2(x) + C_1} \quad (10)$$

$C_1$  is a constant to ensure the stability of Equation (10).

$$Sim_G(x) = \frac{2G_p(x).G_q(x) + C_2}{G_p^2(x) + G_q^2(x) + C_2} \quad (11)$$

$C_2$  is a constant to ensure the stability of Equation (11).

$$Sim_L(x) = [Sim_{PC}(x)]^\alpha \cdot [Sim_G(x)]^\beta \quad (12)$$

The values of  $\alpha$  and  $\beta$  are adjusted according to the importance of PC and GM contributions. Having determined the SimL at a given point  $x$ , the FSIM is computed for the overall domain of  $p(x)$  and  $q(x)$  images.

$$FSIM = \frac{\sum_{x \in \Omega} S_L(x) \cdot Sim_{PC_m}(x)}{\sum_{x \in \Omega} Sim_{PC_m}(x)} \quad (13)$$

where  $Sim_{PC_m}(x) = \max\{PC_p(x), PC_q(x)\}$  is the maximum value in Equation (13).

#### 4.3. Face Recognition

Face recognition is an important step in the existing state of the art to either determine or validate the efficacy of a proposed methodology of face sketch synthesis. Null-Space Linear Discriminant Analysis (NLDA) was employed to compute the quality of synthesized sketches for face recognition. Training and testing split of the total images to train and run the NLDA scheme is given in Tables 2 and 3. Identical parameters were used during the application of the NLDA process to all sketch methodologies under test. In the repeated implementation OpenBR methodology [69] of face, recognition was additionally employed to ascertain the efficacy of proposed and existing schemes of face sketch synthesis.

**Table 2.** Parameters for processing.

S No	Item	CUFS	CUFSF
1	Hardware	Core i-7 <sup>®</sup> , 7th Gen, NVIDIA 1060 (6GB) GPU	
2	OS	Ubuntu Linux	
3	Environment	PyCharm (CE), Torch 1.4.0	
4	Moderating Weights	$\delta_p$	1
		$\delta_{adv}$	$10^3$
		$\delta_{tv}$	$10^{-5}$
5	Learning Weights	$10^{-3}$ to $10^{-5}$ reducing by a factor of $10^{-1}$	
6	Batch Sizes	4 to 2 for different iterations	
7	Processing Time	See respective tables	

**Table 3.** Distribution of synthesized sketches by the NLDA procedure of face recognition.

Dataset	Total Pairs	Train	Test
CUFS	338	150	188
CUFSF	944	300	644

#### 4.4. Hardware and Software Setup

The compiler C and the discriminator D were updated alternately at every iteration. Neural networks were trained in two parts. In the first run of the setup, the CUFS reference style was used, and in its second part, the system was trained with the CUFSF reference style. In each case, however, the training photo–sketch pairs from both datasets were used. The different parameters and the associated information of training processes are given in Table 2.

#### 4.5. Evaluation of Performance on Public Benchmarks

During the evaluation, we used photos from the CUFS dataset only to test the setup trained in the CUFS reference style. Similarly, photo–sketch pairs of the CUFSF dataset were used to test the proposed model trained in the CUFSF style. To determine the effectiveness of this model, results were compared with nine techniques of face sketch synthesis. They are MRF [39], MWF [40], SSD [17], LLE [38], FCN [44], GAN [45], RSLCR [42], Face2Sketch [25] (which contained a U-Net called SNET by its authors), and BiL-STM [57]. Synthesized sketches of the first seven techniques are available at [70]. We implemented the eighth method, Face2Sketch, ourselves in the PyCharm/UBUNTU environment assisted by NVIDIA GPU, mentioned in Table 2. The sketches were synthesized according to the training/testing parameters specified by its original work. Then SSIM, FSIM, and face recognition scores were computed by using these results of the eight techniques and reference sketches in MATLAB/Windows environment. Moreover, training and testing splits were fixed and identical for all the methods during computation of face recognition scores by the NLDA procedure. This detail is given in Tables 4 and 5.

**Table 4.** Comparison of SSIM and FSIM Values for CUFS.

Type	MRF [10]	MWF [11]	LLE [9]	SSD [4]	FCN [15]	GAN [16]	RSLCR [13]	Face2Sketch [6]	BiL-STM [28]	Proposed Spiral-Net
Proc Time (msec/photo)	Not presented by the original works									7.57
SSIM	51.31	53.92	52.58	54.19	52.13	49.38	55.71	54.41	55.19	54.42
FSIM	70.46	71.45	70.32	69.59	69.36	71.54	69.66	72.59	67.77	72.50

**Table 5.** Comparison of face recognition scores for CUFS.

Type	MRF [10]	MWF [11]	LLE [9]	SSD [4]	FCN [15]	GAN [16]	RSLCR [13]	Face2Sketch [6]	BiL-STM [28]	Proposed Spiral-Net
NLDA Score (Equal/Best)	87.34	92.10	90.61	90.61	96.99	93.48	98.38	97.82	94.87	97.04/97.23
No. of Features (Equal/Best)	138	148	144	144	137	139	142	95	-	95/148

#### 4.6. Results of CUFS Dataset

Table 6 shows that the SSIM values of SSD, Face2Sketch, RSLCR, and Spiral-Net are in the same range. Other methods scored less. The SSIM is a too generic a quality parameter to ascertain the visual similarity of images [47,71,72]. It was included in our work for comparison with the results of the previous works. Additionally, the feature similarity measure was computed for these sketch generation methods. Table 6 indicates that the FSIM metrics achieved by Face2Sketch and Spiral-Net are almost identical to each other and slightly higher than the other algorithms. Their difference from other methods' FSIM score is 1–3% higher. In general, all these methods performed fairly similarly in terms of the CUFS dataset, where the viewed sketches lack any difference from the original photos and any variation in light intensity. Computations of the CUFS dataset were included to maintain a harmony of comparison with the previous works.

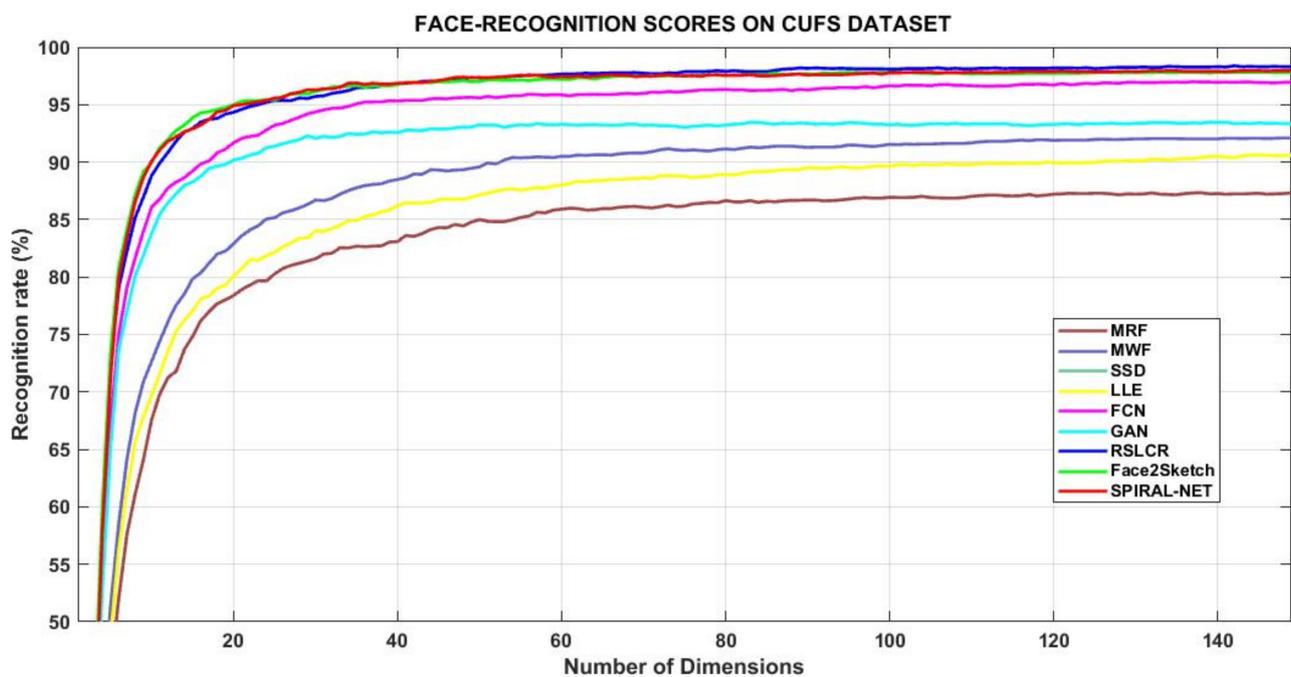
**Table 6.** Comparison of SSIM and FSIM Values for CUFSE.

Type	MRF [10]	MWF [11]	LLE [9]	SSD [4]	FCN [15]	GAN [16]	RSLCR [13]	Face2Sketch [6]	BiL-STM [28]	Proposed Spiral-Net
Proc Time (msec/photo)	Not presented by the original works							4.37	-	7.89
SSIM	35.36	40.83	39.66	41.88	34.39	34.81	42.69	38.97	44.56	38.32
FSIM	66.06	66.76	66.89	64.81	62.91	67.05	63.16	66.87	68.04	68.10

Table 7 records face recognition scores of these methodologies with help of the NLDA procedure, constituted of 142 features/dimensions of the images. Its graphical presentation is in Figure 3. RSLCR, Face2Sketch, and Spiral-Net performed superior to other methods. It is also evident that sketches synthesized by Face2Sketch and Spiral-Net contain more subtle information of the subject persons as compared to other methods since the former two algorithms attain 97% accuracy at 95 dimensions versus the 98% score of RSLCR at 142 dimensions. This improvement in the result also means lesser time complexity of the two methods to reach a rank-1 recognition level.

**Table 7.** Comparison of Face Recognition Scores for CUFSE.

Type	MRF [10]	MWF [11]	LLE [9]	SSD [4]	FCN [15]	GAN [16]	RSLCR [13]	Face2Sketch [6]	BiL-STM [28]	Proposed Spiral-Net
NLDA Score (Equal/Best)	46.03	74.15	70.92	61.76	70.14	71.48	73.05/75.94	73.05	71.35	73.14/78.42
No. of Features (Equal/Best)	223	293	266	274	226	164	102/296	217	-	44/184

**Figure 3.** Comparative view of NLDA scores by different techniques on CUFSE dataset.

#### 4.7. Results of CUFSF Dataset

SSIM, FSIM, and NLDA scores were computed for all eight methodologies, keeping reference parameters identical and intact for all. These values of BL-STM [34] were copied from the original paper. Table 8 records SSIM and FSIM scores of these algorithms for the CUFSF dataset. This dataset contains a diversity of age and ethnicity. Moreover, the viewed sketches were drawn with slight intentional deformations from the photos to render them similar to the properties of forensic sketches. It was observed that SSIM values did not convey any decisive information about the efficacy of the methodologies. RSLCR scored the highest in comparison to other algorithms. The FSIM was considered to be more robust a quality measure. Some of the exemplar-based methods, such as MRF, MWF, and LLE, achieved a 66% score, at par with the Face2Sketch method, which is based on a learning algorithm. The GAN method scored 67%, and it is also based on the neural network. It is seen that the proposed method of Spiral-Net achieved the highest value, of 68%, indicating that sketches synthesized by these methods contain more information of edges, contours, and shapes according to the original photo-sketch pairs.

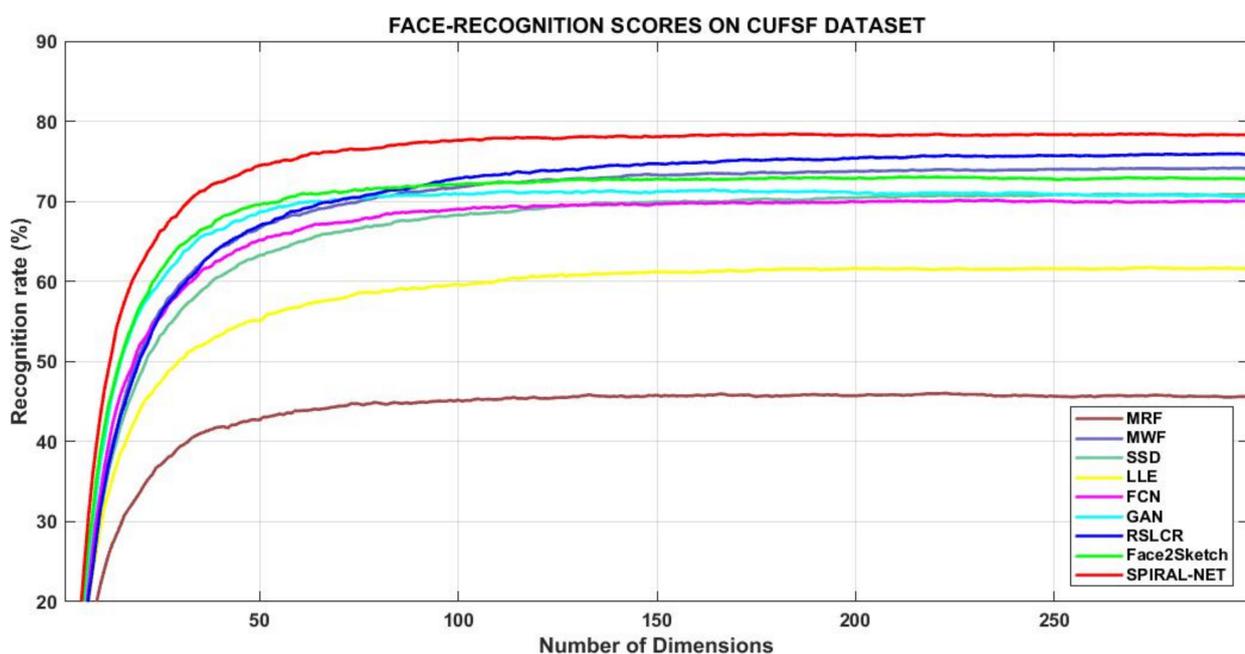
**Table 8.** Details of augmented datasets.

	Dataset	Total Pairs	Train	Test
VSC	CUHK [37]	188	88	100
	AR [65]	123	80	43
	XM2VTS [66]	295	100	195
	IIIT-D	234	94	140
	Total Pairs	840	362	478
VSF	CUFSF	1194	250	944
	IIIT-D	234	94	140
	Total Pairs	1428	344	1084

The NLDA procedure was conducted using up to 300 features/dimensions as a validation step of face recognition in respect of all eight methods. Table 9 highlights those scores, and it is also shown graphically by Figure 4. Of the exemplar-based methods, MWF and RSLCR gained high scores, with 74.15% and 75.94% at 293 and 296 dimensions, respectively. Spiral-Net gained a competitive score of 73.14% at 44 dimensions, and it is equal to the Face2Sketch method, which scored equally at 217 dimensions. Therefore, Spiral-Net synthesizes sketches with enhanced features for a dataset that is considered challenging in the state of art. The best score of Spiral-Net is 78.4% at 184 features, and it further establishes the fact that the proposed method can imitate and “learn” subtle properties of the drawing style of the artist during this method’s training phase with photo-viewed-sketch pairs. It achieved 3–7% improvement over competitive methods from the exemplar-based domain (MWF, RSLCR) and the learning domain (GAN, Face2Sketch). It is seen that layers of the compiler C network from the first stage to the later stages were connected in a novel manner as alternate connections. This feature reduced the possibility of the development of monotony of values at subsequent stages since dissimilar layers were connected to each other progressively. As a result, the values in the matrices of layers bear significance, containing information of high-level features of the input photo or a sketch. This, in turn, preserves subtle information of each image throughout the progress of the network. Therefore, as a performance measure, sketches synthesized by Spiral-Net match better with the test photos at lesser dimensions by the NLDA scheme of face recognition as compared to sketches by other techniques.

**Table 9.** Comparative values of performance for augmented datasets using SNET and proposed Spiral-Net.

Type	VSC-SNET	VSC-Spiral-Net	VSF-SNET	VSF-Spiral-Net
Proc Time (msec/photo)	4.3033	8.5619	4.3113	8.1858
SSIM	38.18	<b>46.81</b>	40.33	40.51
FSIM	67.65	<b>68.34</b>	70.25	70.13
NLDA Score (1998) (%)	67.82	<b>69.61</b>	65.99	65.44
OpenBR_FR Score (2013) (%)	66	<b>71.3</b>	30.7	30.4

**Figure 4.** Comparative view of NLDA scores by different techniques on CUFSF dataset.

#### 4.8. Augmented Dataset and New Implementation

We introduced a new dataset from DIIT [64] and added its 234 photo–sketch pairs to the CUFS and CUFSF datasets. This exercise aimed to test our reference and modified schemes on hybrid datasets to verify their accuracy and to check their comparative performance. Detail is given in Table 8.

*Preprocessing of Augmented Datasets. Phase-1. Treatment of Images.* Pre-processing steps of alignment and rescaling of the images were conducted according to Section 4.2, discussed above. *Phase-2. Development of Feature Dictionary.* The initial run was conducted for each scheme of SNET and Spiral-Net to compute feature files for both photo sets and their corresponding sketch sets at layers relu3-1, relu4-1, and relu5-1. The pre-computed files provided by [25] were not useful since they did not cover an additional part of the dataset introduced by this work. NOTE: The remaining parts of the implementation were conducted similar to Sections 4.2–4.4, as discussed above.

#### 4.9. Evaluation of Augmented Datasets

The following text discusses the analysis of the results from experiments conducted on the augmented dataset.

- It is important to note that we cannot compare newer results with any previous work since our modified or augmented dataset is put to use for the first time.
- The setup was implemented for two schemes, namely Face2Sketch (containing SNET as its component) and Spiral-Net. Therefore, the results may be compared between these two techniques.
- The second and third columns of Table 9 relate to these results. The second column gives values of the SNET technique, and the third column depicts result values for the Spiral-Net technique. It is seen that values of the SSIM and the FSIM for Spiral-Net are superior to those of SNET, which means that the proposed setup imparts more accuracy of features to the formulated sketches. Similarly, the face recognition values by NLDA and OpenBR methods for Spiral-Net are better than those for SNET by almost 2% and 5%, respectively. However, this improvement is achieved at the cost of processing time per photo since Spiral-Net contains almost double the layers of SNET (see Table 9).
- It is also observed from columns fourth and fifth, related to the VSF data component employed by SNET and Spiral-Net, respectively, that there is no marked difference of values between the two techniques. It indicates that CUFSF is inherently a challenging dataset since it copies the characteristics of real-life forensic sketches. Therefore, more research effort is required to fine-tune proposed and other new techniques to improve upon results of a singular CUFSF dataset or any combination of sets involving CUFSF.

## 5. Conclusions

In this work, a novel architecture of U-Net comprising two strains instead of one for the forward pass was proposed. Moreover, the skip connections were made cross-wise between the two strains to reduce the possibility of any monotonous build-up of feature values due to ReLU and pooling operations. Experimental results in comparison to exemplar-based and learning-based schemes indicated that the proposed setup enhances the performance benchmark of sketch synthesis by around 5%. Moreover, a newer approach of augmented datasets comprising conventional sets from CUFS/CUFSF and a part of the DIIT photo-sketch dataset was also applied. Then, it was demonstrated that our modified Spiral-Net achieves a superior performance by 5% compared to its original framework of U-Net. In the future, the authors plan to conduct further experimentation to improve the discriminator D neural network of this framework so as to further refine the loss functions of the technique. Moreover, the currently used feature extractor may be replaced with the neural architecture proposed by Li et al. [73,74].

**Author Contributions:** Conceptualization, I.A. and M.S.; methodology, I.A., M.S. and M.R.; software, I.A.; validation, M.S., M.R. and M.A.K.; formal analysis, M.R.; investigation, M.S. and M.A.K.; resources, H.-S.Y.; data curation, M.S. and M.R.; writing—original draft preparation, I.A. and M.S.; writing—review and editing, M.A.K. and H.-S.Y.; visualization, M.R. and M.A.K.; supervision, M.S.; project administration, H.-S.Y. and M.A.K.; funding acquisition, H.-S.Y. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** This study was partially supported by Ewha Womans University.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Atzori, L.; Iera, A.; Morabito, G. The internet of things: A survey. *Comput. Netw.* **2010**, *54*, 2787–2805. [[CrossRef](#)]
2. Yang, S.; Wen, Y.; He, L.; Zhou, M.C.; Abusorrah, A. Sparse Individual Low-rank Component Representation for Face Recognition in IoT-based System. *IEEE Internet Things J.* **2021**. [[CrossRef](#)]

3. Chauhan, D.; Kumar, A.; Bedi, P.; Athavale, V.A.; Veeraiah, D.; Pratap, B.R. An effective face recognition system based on Cloud based IoT with a deep learning model. *Microprocess. Microsyst.* **2021**, *81*, 103726. [[CrossRef](#)]
4. Kanwal, S.; Iqbal, Z.; Al-Turjman, F.; Irtaza, A.; Khan, M.A. Multiphase fault tolerance genetic algorithm for vm and task scheduling in datacenter. *Inf. Process. Manag.* **2021**, *58*, 102676. [[CrossRef](#)]
5. Sujitha, B.; Parvathy, V.S.; Lydia, E.L.; Rani, P.; Polkowski, Z.; Shankar, K. Optimal deep learning based image compression technique for data transmission on industrial Internet of things applications. *Trans. Emerg. Telecommun. Technol.* **2020**, *32*, e3976. [[CrossRef](#)]
6. Goyal, P.; Sahoo, A.K.; Sharma, T.K.; Singh, P.K. Internet of Things: Applications, security and privacy: A survey. *Mater. Today Proc.* **2021**, *34*, 752–759. [[CrossRef](#)]
7. Akhtar, Z.; Lee, J.W.; Khan, M.A.; Sharif, M.; Khan, S.A.; Riaz, N. Optical character recognition (OCR) using partial least square (PLS) based feature reduction: An application to artificial intelligence for biometric identification. *J. Enterp. Inf. Manag.* **2020**. [[CrossRef](#)]
8. Khan, M.A.; Javed, K.; Khan, S.A.; Saba, T.; Habib, U.; Khan, J.A.; Abbasi, A.A. Human action recognition using fusion of multiview and deep features: An application to video surveillance. *Multimed. Tools Appl.* **2020**, 1–27. [[CrossRef](#)]
9. Sharif, A.; Li, J.P.; Saleem, M.A.; Manogran, G.; Kadry, S.; Basit, A.; Khan, M.A. A dynamic clustering technique based on deep reinforcement learning for Internet of vehicles. *J. Intell. Manuf.* **2021**, *32*, 757–768. [[CrossRef](#)]
10. Khan, M.A.; Zhang, Y.-D.; Alhusseni, M.; Kadry, S.; Wang, S.-H.; Saba, T.; Iqbal, T. A Fused Heterogeneous Deep Neural Network and Robust Feature Selection Framework for Human Actions Recognition. *Arab. J. Sci. Eng.* **2021**, 1–16. [[CrossRef](#)]
11. Khan, M.A.; Muhammad, K.; Sharif, M.; Akram, T.; de Albuquerque, V.H.C. Multi-Class Skin Lesion Detection and Classification via Teledermatology. *IEEE J. Biomed. Heal. Inform.* **2021**, 1. [[CrossRef](#)] [[PubMed](#)]
12. Geremek, M.; Szklanny, K. Deep Learning-Based Analysis of Face Images as a Screening Tool for Genetic Syndromes. *Sensors* **2021**, *21*, 6595. [[CrossRef](#)] [[PubMed](#)]
13. Kim, D.; Ihm, S.-Y.; Son, Y. Two-Level Blockchain System for Digital Crime Evidence Management. *Sensors* **2021**, *21*, 3051. [[CrossRef](#)] [[PubMed](#)]
14. Klare, B.F.; Li, Z.; Jain, A.K. Matching Forensic Sketches to Mug Shot Photos. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *33*, 639–646. [[CrossRef](#)] [[PubMed](#)]
15. Klum, S.J.; Han, H.; Klare, B.F.; Jain, A.K. The FaceSketchID System: Matching Facial Composites to Mugshots. *IEEE Trans. Inf. Forensics Secur.* **2014**, *9*, 2248–2263. [[CrossRef](#)]
16. Galea, C.; Farrugia, R. Forensic Face Photo-Sketch Recognition Using a Deep Learning-Based Architecture. *IEEE Signal Process. Lett.* **2017**, *24*, 1586–1590. [[CrossRef](#)]
17. Song, Y.; Bao, L.; Yang, Q.; Yang, M.-H. Real-Time Exemplar-Based Face Sketch Synthesis. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; pp. 800–813.
18. Klare, B.F.; Jain, A.K. Heterogeneous Face Recognition Using Kernel Prototype Similarities. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *35*, 1410–1422. [[CrossRef](#)] [[PubMed](#)]
19. Negka, L.; Spathoulas, G. Towards Secure, Decentralised, and Privacy Friendly Forensic Analysis of Vehicular Data. *Sensors* **2021**, *21*, 6981. [[CrossRef](#)] [[PubMed](#)]
20. Abayomi-Alli, O.O.; Damaševičius, R.; Maskeliūnas, R.; Misra, S. Few-shot learning with a novel Voronoi tessellation-based image augmentation method for facial palsy detection. *Electronics* **2021**, *10*, 978. [[CrossRef](#)]
21. Liu, P.; Li, X.; Wang, Y.; Fu, Z. Multiple Object Tracking for Dense Pedestrians by Markov Random Field Model with Improvement on Potentials. *Sensors* **2020**, *20*, 628. [[CrossRef](#)]
22. Wei, W.; Ho, E.S.; McCay, K.D.; Damaševičius, R.; Maskeliūnas, R.; Esposito, A. Assessing facial symmetry and attractiveness using augmented reality. *Pattern Anal. Appl.* **2021**, 1–17. [[CrossRef](#)]
23. Ioannou, K.; Myronidis, D. Automatic Detection of Photovoltaic Farms Using Satellite Imagery and Convolutional Neural Networks. *Sustainability* **2021**, *13*, 5323. [[CrossRef](#)]
24. Ranjan, N.; Bhandari, S.; Khan, P.; Hong, Y.-S.; Kim, H. Large-Scale Road Network Congestion Pattern Analysis and Prediction Using Deep Convolutional Autoencoder. *Sustainability* **2021**, *13*, 5108. [[CrossRef](#)]
25. Chen, C.; Liu, W.; Tan, X.; Wong, K.-Y.K. Semi-supervised Learning for Face Sketch Synthesis in the Wild. In Proceedings of the Asian Conference on Computer Vision, Perth, Australia, 2–6 December 2018; pp. 216–231.
26. Chen, C.; Tan, X.; Wong, K.-Y.K. Face Sketch Synthesis with Style Transfer Using Pyramid Column Feature. In Proceedings of the 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV, USA, 12–15 March 2018.
27. Sultan, S.; Javaid, Q.; Malik, A.J.; Al-Turjman, F.; Attique, M. Collaborative-trust approach toward malicious node detection in vehicular ad hoc networks. *Environ. Dev. Sustain.* **2021**, 1–19. [[CrossRef](#)]
28. Khan, M.A.; Kadry, S.; Parwekar, P.; Damaševičius, R.; Mehmood, A.; Khan, J.A.; Naqvi, S.R. Human gait analysis for osteoarthritis prediction: A framework of deep learning and kernel extreme learning machine. *Complex Intell. Syst.* **2021**, 1–19. [[CrossRef](#)]
29. Jeong, Y.-S.; Kim, Y.-T.; Park, G.-C. Blockchain-based multi-IoT verification model for overlay cloud environments. *J. Digit. Converg.* **2021**, *19*, 151–157.
30. Cauteruccio, F.; Cinelli, L.; Corradini, E.; Terracina, G.; Ursino, D.; Virgili, L.; Savaglio, C.; Liotta, A.; Fortino, G. A framework for anomaly detection and classification in Multiple IoT scenarios. *Futur. Gener. Comput. Syst.* **2021**, *114*, 322–335. [[CrossRef](#)]

31. Atzori, L.; Iera, A.; Morabito, G.; Nitti, M. The Social Internet of Things (SIoT)—When social networks meet the Internet of Things: Concept, architecture and network characterization. *Comput. Networks* **2012**, *56*, 3594–3608. [[CrossRef](#)]
32. Jabar, M.K.; Al-Qurabat, A.K.M. Human Activity Diagnosis System Based on the Internet of Things. *J. Phys. Conf. Ser.* **2021**, *1897*, 022079. [[CrossRef](#)]
33. Ansari, G.J.; Shah, J.H.; Khan, M.A.; Sharif, M.; Tariq, U.; Akram, T. A Non-Blind Deconvolution Semi Pipelined Approach to Understand Text in Blurry Natural Images for Edge Intelligence. *Inf. Process. Manag.* **2021**, *58*, 102675. [[CrossRef](#)]
34. Hussain, N.; Khan, M.A.; Kadry, S.; Tariq, U.; Mostafa, R.R.; Choi, J.-I.; Nam, Y. Intelligent Deep Learning and Improved Whale Optimization Algorithm Based Framework for Object Recognition. *Hum. Cent. Comput. Inf. Sci.* **2021**, *11*, 34.
35. Kiran, S.; Khan, M.A.; Javed, M.Y.; Alhaisoni, M.; Tariq, U.; Nam, Y.; Damaševičius, R.; Sharif, M. Multi-Layered Deep Learning Features Fusion for Human Action Recognition. *Comput. Mater. Contin.* **2021**, *69*, 4061–4075. [[CrossRef](#)]
36. Masood, H.; Zafar, A.; Ali, M.U.; Khan, M.A.; Ahmed, S.; Tariq, U.; Kang, B.-G.; Nam, Y. Recognition and Tracking of Objects in a Clustered Remote Scene Environment. *Comput. Mater. Contin.* **2022**, *70*, 1699–1719. [[CrossRef](#)]
37. Xiaoou, T.; Xiaogang, W. Face sketch synthesis and recognition. In Proceedings of the Ninth IEEE International Conference on Computer Vision, Nice, France, 13–16 October 2003; Volume 1, pp. 687–694.
38. Qingshan, L.; Xiaoou, T.; Hongliang, J.; Hanqing, L.; Songde, M. A nonlinear approach for face sketch synthesis and recognition. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–25 June 2005; Volume 1, pp. 1005–1010.
39. Wang, X.; Tang, X. Face Photo-Sketch Synthesis and Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *31*, 1955–1967. [[CrossRef](#)] [[PubMed](#)]
40. Zhou, H.; Kuang, Z.; Wong, K.-Y.K. Markov Weight Fields for face sketch synthesis. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 1091–1097.
41. Gao, X.; Wang, N.; Tao, D.; Li, X. Face Sketch–Photo Synthesis and Retrieval Using Sparse Representation. *IEEE Trans. Circuits Syst. Video Technol.* **2012**, *22*, 1213–1226. [[CrossRef](#)]
42. Wang, N.; Gao, X.; Li, J. Random sampling for fast face sketch synthesis. *Pattern Recognit.* **2018**, *76*, 215–227. [[CrossRef](#)]
43. Akram, A.; Wang, N.; Li, J.; Gao, X. A Comparative Study on Face Sketch Synthesis. *IEEE Access* **2018**, *6*, 37084–37093. [[CrossRef](#)]
44. Zhang, L.; Lin, L.; Wu, X.; Ding, S.; Zhang, L. End-to-End Photo-Sketch Generation via Fully Convolutional Representation Learning. In Proceedings of the 5th ACM on International Conference on Multimedia Retrieval, Shanghai, China, 23–26 June 2015.
45. Isola, P.; Zhu, J.-Y.; Zhou, T.; Efros, A.A. Image-to-Image Translation with Conditional Adversarial Networks. *arXiv* **2017**, arXiv:1611.07004.
46. Zhang, D.; Lin, L.; Chen, T.; Wu, X.; Tan, W.; Izquierdo, E. Content-Adaptive Sketch Portrait Generation by Decompositional Representation Learning. *IEEE Trans. Image Process.* **2016**, *26*, 328–339. [[CrossRef](#)]
47. Wang, L.; Sindagi, V.; Patel, V. High-Quality Facial Photo-Sketch Synthesis Using Multi-Adversarial Networks. In Proceedings of the 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), Xi'an, China, 15–19 May 2018.
48. Wang, N.; Gao, X.; Sun, L.; Li, J. Anchored Neighborhood Index for Face Sketch Synthesis. *IEEE Trans. Circuits Syst. Video Technol.* **2017**, *28*, 2154–2163. [[CrossRef](#)]
49. Jiao, L.; Zhang, S.; Li, L.; Liu, F.; Ma, W. A modified convolutional neural network for face sketch synthesis. *Pattern Recognit.* **2018**, *76*, 125–136. [[CrossRef](#)]
50. Zhang, S.; Ji, R.; Hu, J.; Lu, X.; Li, X. Face Sketch Synthesis by Multidomain Adversarial Learning. *IEEE Trans. Neural Networks Learn. Syst.* **2019**, *30*, 1419–1428. [[CrossRef](#)] [[PubMed](#)]
51. Zhang, M.; Wang, R.; Gao, X.; Li, J.; Tao, D. Dual-Transfer Face Sketch–Photo Synthesis. *IEEE Trans. Image Process.* **2018**, *28*, 642–657. [[CrossRef](#)]
52. Lin, Y.; Ling, S.; Fu, K.; Cheng, P. An Identity-Preserved Model for Face Sketch-Photo Synthesis. *IEEE Signal Process. Lett.* **2020**, *27*, 1095–1099. [[CrossRef](#)]
53. Fang, Y.; Deng, W.; Du, J.; Hu, J. Identity-aware CycleGAN for face photo-sketch synthesis and recognition. *Pattern Recognit.* **2020**, *102*, 107249. [[CrossRef](#)]
54. Xie, F.; Yang, J.; Liu, J.; Jiang, Z.; Zheng, Y.; Wang, Y. Skin lesion segmentation using high-resolution convolutional neural network. *Comput. Methods Programs Biomed.* **2020**, *186*, 105241. [[CrossRef](#)] [[PubMed](#)]
55. Lin, S.; Zhang, J.; Pan, J.; Liu, Y.; Wang, Y.; Chen, J.; Ren, J. Learning to Deblur Face Images via Sketch Synthesis. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 11523–11530.
56. Zhu, M.; Li, J.; Wang, N.; Gao, X. Knowledge Distillation for Face Photo-Sketch Synthesis. *IEEE Trans. Neural Networks Learn. Syst.* **2020**, 1–14. [[CrossRef](#)] [[PubMed](#)]
57. Radman, A.; Suandi, S.A. BiLSTM regression model for face sketch synthesis using sequential patterns. *Neural Comput. Appl.* **2021**, *33*, 12689–12702. [[CrossRef](#)]
58. Johnson, J.; Alahi, A.; Fei-Fei, L. Perceptual Losses for Real-Time Style Transfer and Super-Resolution. *arXiv* **2016**, arXiv:1603.08155.
59. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2015**, arXiv:1409.1556.
60. Li, C.; Wand, M. Combining Markov Random Fields and Convolutional Neural Networks for Image Synthesis. *arXiv* **2016**, arXiv:1601.04589.

61. Mao, X.; Li, Q.; Xie, H.; Lau, R.Y.K.; Wang, Z.; Smolley, S.P. Least Squares Generative Adversarial Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *41*, 2947–2960. [[CrossRef](#)]
62. Kaur, P.; Zhang, H.; Dana, K. Photo-Realistic Facial Texture Transfer. In Proceedings of the 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa Village, HI, USA, 7–11 January 2019; pp. 2097–2105.
63. Zhang, W.; Wang, X.; Tang, X. Coupled information-theoretic encoding for face photo-sketch recognition. In Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition, Washington, DC, USA, 20–25 June 2011.
64. Bhatt, H.S.; Bharadwaj, S.; Singh, R.; Vatsa, M. Memetic approach for matching sketches with digital face images. *IEEE Trans. Inf. Forensics Secur.* **2012**, *7*, 1522–1535. [[CrossRef](#)]
65. Martínez, A.; Benavente, R. The AR face database. *Comput. Vis. Cent.* **2007**, *3*, 5.
66. Messer, K.; Matas, J.; Kittler, J.; Luettin, J.; Maitre, G. XM2VTSDB: The extended M2VTS database. In Proceedings of the Second International Conference on Audio and Video-Based Biometric Person Authentication, Washington, DC, USA, 22–24 March 1999; pp. 965–966.
67. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [[CrossRef](#)]
68. Zhang, L.; Zhang, L.; Mou, Z.; Zhang, D. FSIM: A Feature Similarity Index for Image Quality Assessment. *IEEE Trans. Image Process.* **2011**, *20*, 2378–2386. [[CrossRef](#)]
69. Klontz, J.C.; Klare, B.F.; Klum, S.; Jain, A.K.; Burge, M.J. Open source biometric recognition. In Proceedings of the 2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS), Arlington, VA, USA, 29 September–2 October 2013; pp. 1–8.
70. Rigel, D.S.; Carucci, J.A. Malignant melanoma: Prevention, early detection, and treatment in the 21st century. *CA Cancer J. Clin.* **2000**, *50*, 215–236. [[CrossRef](#)] [[PubMed](#)]
71. Wang, N.; Zha, W.; Li, J.; Gao, X. Back projection: An effective postprocessing method for GAN-based face sketch synthesis. *Pattern Recognit. Lett.* **2018**, *107*, 59–65. [[CrossRef](#)]
72. Ledig, C.; Theis, L.; Huszar, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; et al. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.
73. Li, Z.; Zhou, A. Self-Selection Salient Region-Based Scene Recognition Using Slight-Weight Convolutional Neural Network. *J. Intell. Robot. Syst.* **2021**, *102*, 1–16. [[CrossRef](#)]
74. Li, Z.; Zhou, A.; Shen, Y. An End-to-End Trainable Multi-Column CNN for Scene Recognition in Extremely Changing Environment. *Sensors* **2020**, *20*, 1556. [[CrossRef](#)]