ISSN No:-2456-2165

Cross-Domain Face Recognition for Criminal Identification

Adhert Johnson
Student, Computer science and Engineering Department
Sahrdaya College of Engineering Technology
Thrissur, India

Adithyan Ranjan K C
Student, Computer science and Engineering Department
Sahrdaya College of Engineering Technology
Thrissur, India

Christy Vargheese
Student, Computer science and Engineering Department
Sahrdaya College of Engineering Technology
Thrissur, India

Akhil A S
Student, Computer science and Engineering Department
Sahrdaya College of Engineering Technology
Thrissur, India

Princy T D
Asst. Professor, Computer science and Engineering Department
Sahrdaya College of Engineering Technology
Thrissur, India

Abstract:- The problem of cross-domain face recognition seeks to identify facial images obtained from different domains, and it is gaining popularity due to its numerous applications in law enforcement identification and camera surveillance Existing algorithms typically fail to fully exploit semantic information for identifying cross-domain faces, which could be a strong clue for recognition. In this paper, we present an efficient algorithm for cross-domain face recognition that makes use of semantic information in conjunction with deep convolutional neural networks (CNN). We start with a soft face parsing algorithm that measures the boundaries of facial components as probabilistic values. For cross domain face recognition, we propose a hierarchical soft semantic representation framework. CNN-derived deep features are computed and combined. Which could fully exploit the same semantic clue across cross-domain faces. We present extensive experiments to show that the proposed soft semantic representation algorithm outperforms state-of-the-art methods.

I. INTRODUCTION

FACE images are frequently captured from various domains in various tasks. In an enforcement application, one common scenario is to match face photos with sketches generated from supported eyewitness descriptions of the suspect. Forensic artists are frequently invited to create a hand-drawn sketch based on the description, or software is frequently used to generate a composite sketch for the purpose of identifying the suspect. Face photos and sketches have a large domain difference because photos are captured in a real-life environment, whereas both hand-drawn and software-generated sketches are created manually with artefact [1].Another cross-domain face recognition scenario within the camera surveillance application is to match visual spectra images (VIS) with near-infrared images (NIR), because near-infrared images are resistant to changing lighting conditions. Given that the VIS-NIR faces are captured by different sensors, there is also a significant domain difference between them. Cross-domain face recognition, also referred to as heterogeneous face recognition [4] seeks to match face images across domains. The training subset usually consists of a variety of face image pairs from two different domains, while the remaining face image pairs are used for testing. Due to the various domain discrepancies, cross domain face recognition may be a difficult task. In recent years, a number of cross-domain face recognition algorithms have been developed. Early cross-domain face recognition methods typically used hand-crafted architectures to either transform face images into the same domain using face synthesis techniques [5] or directly match cross-domain faces using a standard subspace [6] or hand-crafted feature descriptors. Hand-crafted approaches, on the other hand, are frequently computationally expensive, hence their performance is usually limited. With the rapid advancement of deep convolutional neural networks (CNN) in classification tasks, a growing number of deep learning-based cross-domain face recognition approaches have recently been developed. To eliminate the domain discrepancy, an end-to-end deep network was used for cross-domain face synthesis [8]. Cross-domain face recognition is also influenced by restricted Boltzmann Machines (RBMs) [9], CNN [10], and generative adversarial networks (GAN) [11].We offer an efficient soft semantic representation based framework (SoftSR) in this research to utilise semantic information for crossdomain face recognition. We offer a soft face parsing technique that performs a soft classification on the pixels around the facial component borders, taking into account the disadvantages of existing approaches that use hard pixel classification based semantic maps. We use the first face image as a guide to create a soft semantic map, in which each pixel can belong to a semantic facial component in part. For cross-domain face recognition, we also build a hierarchical soft semantic representation system..Both the soft semantic component level and, as a result, the face contour level are used to extract deep CNN features.

They are then integrated to improve cross-domain face identification performance. The equivalent semantic information across different domains is completely utilised in this manner. The following is a list of the article's contributions :

1) For cross-domain face recognition, we offer a soft face parsing technique that takes advantage of semantic information. The pixels in our soft semantic map, which can deal with erroneous segmentation around semantic face borders, are partially assigned to the component as probabilistic values.
2) To improve the performance of cross-domain face recognition, we develop a hierarchical soft semantic representation framework that is coupled with deep convolutional neural networks.
3) We run extensive experiments on multiple cross-domain face datasets to test the proposed technique and show that it outperforms state-of-the-art approaches.

II. RELATED WORK

Here we look at some of the related works in the field of Face Recognition.

A. Hand -crafted-method

Face identification across domains began with an eigen transformation-based method that converted face pictures into sketches, and then a typical face recognition algorithm was used to match synthesised sketches with sketches in the target gallery. For cross-domain face alteration, Wang et al. combined sparse feature selection with support vector regression. A number of Markov network-based approaches [5] have been presented in recent years to remodel cross-domain faces into a common domain for recognition. Cross-domain face identification started with an eigen transformation-based approach that converted face pictures into the sketch domain. Matching synthesised sketches with sketches in the target gallery was done using a typical face recognition algorithm. For cross-domain face morphing, Wang et al. [25] combined sparse feature selection with support vector regression. A variety of Markov network-based strategies [5] have been proposed to be revised in recent years.

B. Methods based on deep learning

The intrinsic distribution of raw pixels is frequently investigated directly as deep learning progresses. To produce high-quality synthesised faces, an end-to-end cross-domain face synthesis approach [8] was developed. RBMs, CNNs, and GANs were also used to influence cross-domain face recognition [9].Galea and Farrugia recommended that 3D morphable models be fitted to face pictures and sketches so that fresh images may be synthesised for training data augmentation. Later, it was developed to a super-deep CNN with morphing faces and transfer learning [1] for matching software-generated sketches to images. Mutual component analysis was incorporated into CNN by treating it as a separate fully-connected layer capable of extracting domain-independent characteristics for recognition. To deal with thermal to visible face recognition and cross-resolution face recognition [49], Nasrabadi et al. developed attribute-guided algorithms. The variability between ID and spot photographs, the bisample issue (where only two samples are given for each topic), and thus the vast scale classes data are all obstacles in recognising photographs from ID and spot photo domains. They proposed a classification-verification-classification pipeline to transfer models in a progressive manner. Deng et al. suggested a two-branch Residual networks architecture for cross-domain face recognition, featuring a residual compensation module and modality discrepancy loss, and they produced promising results on both sketch-photo matching and NIR-VIS matching tasks. Yu et al. recently proposed using a generative adversarial network to turn a facial sketch into a photo with a loss at the feature level. By incorporating image generation for matching NIR and VIS faces, a variety of ways were presented. However, The semantic clue was ignored by the majority of existing cross-domain face recognition algorithms. According to our observations, the semantic information of facial components is generally similar across domains (with the exception of caricature recognition, which involves shape exaggeration). We will develop cross-domain face correlations and propose a soft semantic representation strategy that is robust across picture domains based on this fact. Unlike the component-based method (CBR), which relied on difficult component segmentation and hand-crafted characteristics, machine learning methods provide the necessary tools. Model validation is carried out by 5-fold cross-validation, which divides the dataset into ve subsets and uses one of the subsets as the testset and the other four as the training set.

III. PROPOSED SYSTEM

We offer our proposed soft semantic representation for cross-domain face recognition in this part. We define X = X1, X1, X M as a collection of M face photos in two domains, and Y = Y1, Y1, YM as a collection of M face photos in two domains. There are two phases to our process. To begin, the photos in both domains are subjected to a soft semantic face parsing algorithm, which provides a soft semantic parsing map for each face picture. Second, using the soft face semantic map as a guide, we create a hierarchical framework for extracting deep CNN features from both the soft semantic and contour levels. For cross-domain face recognition difficulties, using a step-by-step procedure is a common solution. Face photos, for example, are processed first by a face completion method before being put into a CNN for cross-domain face matching. To obtain a new image pattern, first learn a discriminative face representation, and then project the filtered images into one common subspace for identification. Similarly, to tackle cross-domain face recognition, we provide a two-phase methodology. We'll go over each phase in detail, and for the sake of brevity, we'll use face sketch-photo identification as an example.

ISSN No:-2456-2165

A. Semantic Face Parsing with Soft Semantics

Semantic face parsing map as a cross-domain clue for our recognition objective, inspired by earlier publications [12] on applying semantic face parsing to face-related tasks. We use established ways to divide facial features into four categories: brows, eyes, nose, and mouth. We first recognise facial landmarks in a face picture (which could be a photo or a sketch image) using convolutional experts restricted local models. It is possible to employ dense facial landmarks to guide the creation of a semantic parsing map. The semantic face parsing map can be generated in a straightforward method by linking surrounding landmarks around each facial component. Each polygon's pixels are given to the appropriate facial component category. The semantic region for the left eyebrow, for example, is defined as the pixels with associated landmarks around the left eyebrow. However, because there are some long-distance surrounding landmarks, this direct connection operation may create a piecewise linear effect. To address this problem, we use cubic spline interpolation to smooth the component's boundary. Finally, we might create an initial semantic parsing map for each face image, with pixels within facial component regions marked as 1 and the rest as 0.

However, the initial semantic parsing map suffers from the same flaw as previous works in that it performs a hard pixel classification around component borders. The pixels on the inside of the boundary are labelled 1 while those on the outside are labelled 0. However, the component's border may be erroneous, which is overlooked in current methodologies. For example, the nose bridge's boundary may be ambiguous, and the above hard pixel classification may provide misleading information for our next phase. As a result, in this paper, we describe a soft semantic face parsing system in which pixels near facial component boundaries are classified as component pixels. We employ each row/column in the initial semantic parsing map as the input signal S, and every row/column in the input face image as the guidance signal P in our soft semantic face parsing procedure. As a result, the sides of the input face image may be used to guide the semantic parsing map smoothing. This vector shows the likelihood that a pixel corresponds to at least one face component. To get the soft semantic level shape feature, we simply concatenate the R dimensional vectors in the least pixels together, which is designated as follows: FS S = f (Xs m, R) (4), where FS S denotes the soft shape feature that can be retrieved easily using the soft semantic parsing map Xs m as input.When there is a transparent border within the input face image, such as an eye corner, the result is an outsized value and no smoothing is performed. Xs m = SSFP(Xm , Xinim , σr, σs) (3)The soft semantic face parsing operation is denoted by SSFP. Here, the parameters s and r are used. The parameters have an impact on the performance of our method. s is frequently thought of as a spatial parameter introduced in equation. It regulates the recursive edge-preserving filter's feedback coefficient. r is commonly thought of as a range parameter that affects the smoothing operation's range area. The smoothing phenomenon at the face component boundaries becomes more noticeable as s and r grow. When the range parameter r is set to a large value, the parsing map becomes excessively smoothed, and useful face component boundaries become missing.

When the spatial parameter s is extremely big, however, the parsing map will not be over smoothed. As a result, the default parameter setting for r must be a suitable value, whereas the default parameter setting for s can simply be an outsized value. Because the proposed parsing method's computation time is exceedingly long,

When the spatial parameter s is extremely big, however, the parsing map will not be over smoothed. As a result, the default parameter setting for r must be a suitable value, whereas the default parameter setting for s can simply be an outsized value. Because the suggested parsing method has a very short processing time, which will be explained in the complexity analysis chapter, tweaking the two parameters isn't very time demanding. For a variety of cross-domain face recognition challenges,

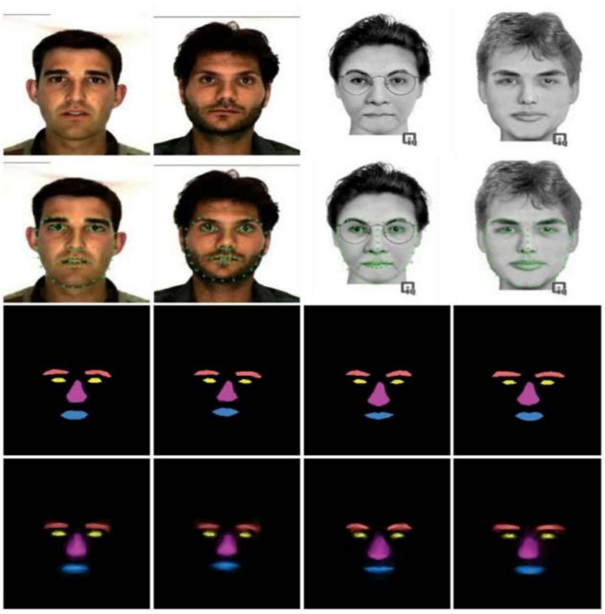

Fig. 1

B. Cross-Domain Face Recognition Using Soft Semantic Representation

To tackle the domain discrepancy in cross-domain face recognition, we provide a hierarchical system providing soft semantic face parsing. Our technique integrates both soft semantic level and contour level deep information, as illustrated in Fig. 1. The soft semantic level features are independent of the domain discrepancy because they may capture full information from both soft semantic component forms and component contents. The former is frequently regarded as a high-level property of the face that is extracted by extracting semantic characteristics at each pixel, whilst the latter can reflect the appearance of every semantic component. We also use contour level characteristics from the holistic face to provide a semantic description of the entire face, which are complementary, and the merger of these hierarchical deep features could enable efficient and robust cross-domain face recognition performance. Below, we go over the soft semantic representation in detail.

1) Soft Semantic Level Feature: High-level semantic features have been shown to be successful in face labelling, face deblurring, and face synthesis, among other applications. We extract the soft semantic clue at each pixel

ISSN No:-2456-2165

position to describe whether the pixel partially belongs to a facial component category or not, because the semantic information of facial components is usually consistent across faces from different domains. At each pixel, we create an R-dimensional probabilistic vector, where R is the number of component categories. This vector shows the likelihood that a pixel corresponds to at least one face component. To acquire the soft semantic level shape feature, we simply concatenate the R dimensional vectors in the least pixels together.

2) Contour Level Feature: We will capture extensive information from facial contour, which is robust to changes in stance and expression, in addition to the soft semantic level feature, which concentrates on local components. The key points in the initial semantic parsing map Xini m are used to extract deep features with the help of the image patch level network described above. The contour level feature, where FHC denotes the original hard semantic level contour feature, concatenates the extracted deep features. The contour level feature can supplement the soft semantic level feature by providing additional information.

3) Soft Semantic Representation Fusion: Each face component category's soft semantic level features FS S and FSC, as well as the contour level feature FHC, are retrieved first. We combine the soft semantic level shape feature with the content feature retrieved at critical places inside this component region to create one soft semantic level representation for each semantic facial component [FS S; FSC]. We'll get six soft semantic level representations if there are six component categories: left eyebrow, left eye, right eyebrow, right eye, nose, and mouth. The contour level representation is formed by concatenating the deep features extracted at the contour landmarks. The root normalises the elements of the six soft semantic level representations and, as a result, the contour level representations, resulting in our soft semantic representation. The set of emotive semantic representations is used for cross-domain face recognition. For starters, we create a classifier for each soft semantic level and contour level representation. Second, the classifiers' scores are combined using a simple averaging technique. Because the dimension of FSC and FHC is usually quite large, we use the principal component analysis and Fisherface algorithms to map them into a discriminant subspace. Further complex classifiers and fusion approaches could be used to improve the performance even more, and this is something that will be looked into in the future. In the experimental section below, we'll go over the implementation details of our framework.

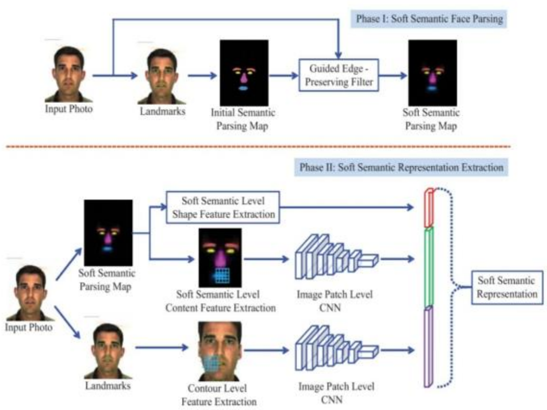

Fig.2

IV. EXPERIMENTS AND RESULTS

The experimental performance of our proposed method on numerous cross-domain face recognition datasets, as well as comparisons to state-of-the-art algorithms, are described in this section. Experiments are performed using cross-domain face databases that are publically available, such as the photo domain with sketch domain, UoM-SGFS [1] and a gathered forensic sketch database, the live photo domain with the ID photo domain, and the VIS domain with the NIR domain. The e-PRIP database has 123 photo-sketch pairs, with the photographs coming from the AR database and the sketches being created using composite sketch software. The database is divided into two subsets at random, each with 48 pairs. The remaining is for testing and classifier training. The PRIP-VSGC database has the same 123 pictures as the e-PRIP database, but the sketches were made with a different composite sketch programme. For evaluation, the same methodology used in the e-PRIP database is used. The photographs for the 600 subjects in the UoM-SGFS database were acquired from the FERET database. A composite artist creates one sketch for each face pic.

One sketch is prepared for each face photo using the EFIT-V composite sketch software, and another sketch is created by further altering the formal sketch with the Corel PaintShop Pro-X7 software. As a result, EFIT-first V's composite sketches are collected as subset A, and the further altered sketches are gathered as subset B after a short time. This article follows the standard evaluation process outlined in. The forensic sketch database contains 168 forensic sketch-photo pairs gathered from real-life law enforcement situations. Because the mug shot images are acquired in varied environments and the forensic sketches are made according to eyewitness descriptions, this is a difficult face sketch-photo recognition database. We employ split protocols, which use 106 image pairings for training and the rest for testing. .We follow the standard protocol to gauge our proposed soft semantic representation on matching faces between live photo domain and ID photo domain. The CASIA NIR-VIS 2.0 database contain A total of 10,000 additional photographs that have been added to the gallery to make it larger. There are 256 high-resolution live photographs in the NJU-ID database, along with accompanying low-resolution ID photographs. To test our proposed soft semantic representation on matching faces between the live photo domain and the ID photo domain, we use the standard protocol. There are 725 subjects in the CASIA NIR-VIS 2.0 database, divided into two views. View 1 is used for parameter adjustment, whereas view 2 is utilised to evaluate performance. There are several facial photos from both the VIS and NIR domains for each person. In this study, we use a comparable experimental technique given by databases, such as the UoM-SGFS database and hence the CASIA NIR-VIS 2.0 database, if they are accessible. Otherwise, we use a similar partitioning protocol as described in the respective studies. We use 48 subjects for training and 75 subjects for testing in the e-PRIP database and the PRIP-VSGC database, according to the partition protocol. For the forensic sketch database, we use an experimental protocol in which 106 subjects are chosen at random for training and the remaining 53 for testing. We use a precise 10-fold cross validation technique for the NJU-ID

ISSN No:-2456-2165

database, dividing the database into 10 folds at random.

A. Implementation Specifications

On the e-PRIP database, we explain the implementation details and parameter settings of our proposed algorithm. All of the face photos used in this text are first aligned and cropped to 200 by 250 pixels. The parameters r and s are set to 0.4 and 40, respectively, during the soft semantic face parsing phase. In the following section, we'll look at the impact of those two variables. The dimensions of picture patches utilised for extracting CNN features in soft semantic representation extraction are around 32 x 32. The important spots are sampled with a density of 10 10 grid during the soft content feature extraction step. The impact of those two parameters during the soft semantic feature extraction step will also be examined further down. To examine facial character around the keypoint, we merely require four image patches surrounding each keypoint. The Brown dataset is used to train the image patch level CNN. After that, we fine-tune the cross-domain databases separately. The hard samples approach is used, as well as data augmentation with random rotation. We choose to utilise a conventional CNN rather than a well-designed complicated module to showcase the contribution of this text because the key contribution is to employ the proposed soft semantic representation technique for cross-domain face recognition.

Our results show that the features extracted from the picture patch level CNN can perform well on cross-domain face recognition tasks. In reality, within the task of image matching, the image patch level CNN must be present to indicate significant potentiality and effectiveness. The domain gap in our challenge can also be closed using picture patch level CNN as part of our network training technique, which includes fine-tuning on cross-domain databases using hard samples and data augmentation with random rotation.

1) *Soft Semantic Parsing Parameters:*

There are two parameters r and s that need to be determined in our proposed soft semantic face parsing algorithm. In the graphic, we show a visual comparison of the impact of those two characteristics. If the borders inside the input face image are obvious, our suggested soft semantic parsing algorithm can successfully smooth the unclear facial component borders while keeping the strong edge structure, as shown in the figure. Furthermore, we begin by setting s to 40. When r is small, no smoothing around the facial component boundaries is done. The soft semantic parsing map gets highly smooth when r is large. Then, as seen in the second row of the picture, we set r to 0.4. When s is small, there is no smoothing process. The findings are not over smoothed even when s is very large. As a result, during this article, we set r and s to 0.4 and 40, respectively, to provide the simplest performance.

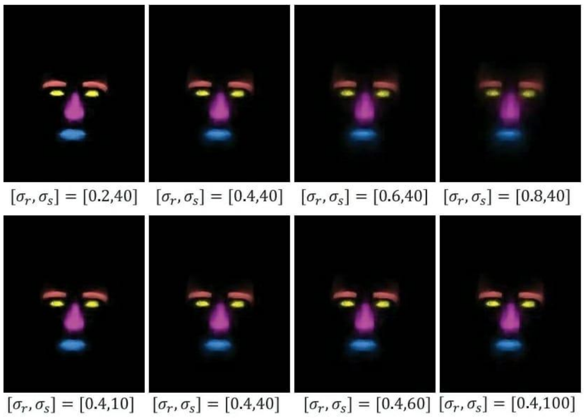

Fig.3

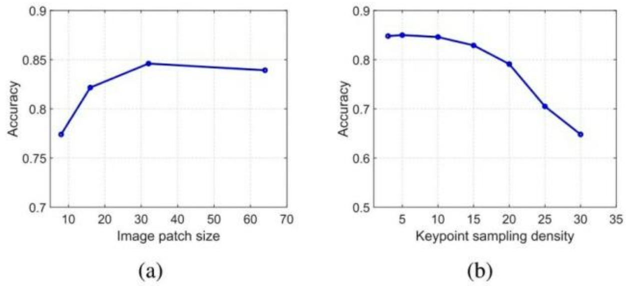

Fig.4

2) The Soft Semantic Representation Extraction Parameters' Influence:

We need to extract CNN features of image patches sampled from cross-domain face images while using the suggested soft semantic face parsing algorithm for cross-domain face recognition. Because we use image patch level CNN for feature extraction, it's plausible to attribute the performance to the image patch size choice. We test our proposed method for feature extraction using different picture patch sizes. It is difficult for CNN to extract discriminative information from small picture patches when the image patch size is tiny. As a result, if we choose an image patch size of 8/8 or 16/16, the performance will be bad. With the increase in image patch size, more relevant and discriminative information, as well as semantic clues, are taken into account, which improves cross-domain face recognition.

The computation cost for image patch level CNN, on the other hand, increases as the image patch size grows. The picture patch size is set to 32 32 as the default setting in this article. The keypoint sampling density, which regulates the density of sampling key points within face components to extract picture patches for recognition,. We densely sample key spots inside the region of face components in the soft content feature extraction phase and generate picture patches around these key locations for deep feature extraction. The keypoint sampling approach at various densities. When the density is low, the image patches will be sampled using a dense grid, which will increase the calculation cost. The results are adequate. The effects of such factors are similar for various cross-domain face recognition tasks. We discovered that using the default parameters on different datasets yields promising results, demonstrating the method's generalisation capabilities.

V. CONCLUSION

Cross-domain face recognition may be a difficult task due to the wide domain discrepancy. We provide a fully unique soft semantic representation for cross domain face recognition in this article. The first step is to create a soft face parsing method that can produce a soft semantic parsing map using the input face picture to avoid erroneous segmentation around component boundaries in the first hard parsing map. After that, a hierarchical soft semantic representation extraction framework is described, with sentimental semantic representation consisting of soft shape features, soft content features, and hard semantic level contour features. The proposed strategy achieves the simplest performance on numerous cross-domain face datasets by merging the collection of emotive semantic representations. The proposed soft semantic parsing method will be applied to additional fields in the future. We devote a separate subsection to the suggested soft semantic face parsing, which may alternatively be thought of as a pre-processing phase for face recognition. Uncertain visual boundaries around facial components are frequently softened, resulting in incorrect segmentation. This is beneficial for improved cross-domain face representation. We believe that the features of smoothing incorrect segmentation near ambiguous boundaries while retaining edges at clear boundaries of a face picture are frequently integrated into other related tasks such as face deblurring, face denoising, and face synthesis.

REFERENCES

[1]. C. GALEA and R. A. FARRUGIA, "Matching software-generated sketches to face photographs with a very deep CNN, morphed faces, and transfer learning," IEEE Trans. Inf. Forensics Security, vol. 13, no. 6, pp. 1421–1431, Jun. 2018

[2]. Y. SHI and A. K. JAIN, "DocFace: Matching ID document photos to selfies," in Proc. IEEE 9th Int. Conf. Biometrics Theory, Appl. Syst., Oct. 2018, pp. 56–67

[3]. H. ZHANG, H. HAN, J. CUI, S. SHAN, and X. CHEN, "RGB-D face recognition via deep complementary and common feature learning," in Proc. 13th IEEE Int. Conf. Autom. Face Gesture Recognit., May 2018, pp. 1–8

[4]. X. WANG and X. TANG, "Face photo-sketch synthesis and recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, no. 11, pp. 1955–1967, Nov. 2009.

[5]. ABDULLAH ALSHANQITI AND ABDALLAH NAMOUN, Predicting Student Performance and Its Influential Factors Using Hybrid Regression and Multi-Label Classification Digital Object Identifier November19,2020.Digital Object Identifier 10.1109/ACCESS.2020.3036572 Madinah 42351, Saudi Arabia

[6]. RAMIN GHORBANI AND ROUZBEH GHOUS,[1] Comparing Different Resampling Methods in Predicting Students' Performance Using Machine Learning Techniques Digital Object Identifier 10.1109/ACCESS.2020 April 22, 2020. Digital Object Identifier 10.1109/ACCESS.2020.2986809 Tehran 16846-13114, Iran

[7]. HANAN ABDULLAH MENGASH, Using Data Mining Techniques to Predict Student Performance to Support Decision Making in University Admission Systems March 30, 2020. Digital Object Identifier 10.1109/ACCESS.2020.2981905 Riyadh 11351, Saudi Arabia

[8]. AGORITSA POLYZOU AND GEORGE KARYPIS, Feature Extraction for Next-Term Prediction of Poor Student Performance IEEE TRANSACTIONS ON LEARNING TECHNOLOGIES, VOL. 12, NO. 2, APRIL-JUNE 2019

[9]. SANA BHUTTO,DR. QASIM ALI ARAIN,MALEEHA ANWAR, Predicting Students' Academic Performance Through

[10]. Supervised Machine Learning 2020 International Conference on Information Science and Communication Technology

[11]. Y.-H. HU, C.-L. LO, AND S.-P. SHIH, ''Developing early warning systems to predict students' online learning performance,'' Comput. Hum. Behav., vol. 36, pp. 469–478, Jul. 2014.

[12]. M.ZIĘBA, S. K. TOMCZAK, AND J. M. TOMCZAK, ''Ensemble boosted trees with synthetic features generation in application to bankruptcy prediction,'' Expert Syst. Appl., vol. 58, pp. 93–101, Oct. 2016.

[13]. C. PENG, X. GAO, N. WANG, and J. LI, "Graphical representation for heterogeneous face recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 2, pp. 301–312, Feb. 2017

[14]. Y. SONG ET AL., "Joint face hallucination and deblurring via structure generation and detail enhancement," Int. J. Comput. Vis., vol. 127, nos. 6–7, pp. 785–800, Jun. 2019