IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received March 2, 2020, accepted March 20, 2020, date of publication March 24, 2020, date of current version April 7, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2982970

Face Segmentation: A Journey From Classical to Deep Learning Paradigm, Approaches, Trends, and Directions

KHALIL KHAN[1], REHAN ULLAH KHAN[2], KASHIF AHMAD[3], FARMAN ALI[4], AND KYUNG-SUP KWAK[5]

[1]Department of Electrical Engineering, University of Azad Jammu and Kashmir, Muzaffarabad, Pakistan
[2]Department of Information Technology, College of Computer, Qassim University, Al-Mulida, Saudi Arabia
[3]Hamad Bin Khalifa University, Doha, Qatar
[4]Department of Software, Sejong University, Seoul, South Korea
[5]Department of Information and Communication Engineering, Inha University, Incheon, South Korea

Corresponding authors: Khalil Khan (khalil.khan@ajku.edu.pk) and Kyung-Sup Kwak (kskwak@inha.ac.kr)

This work was supported by the National Research Foundation of Korea-Grant through the Korean Government (Ministry of Science) under Grant ICT-NRF-2020R1A2B5B02002478.

ABSTRACT Face segmentation represents an active area of research within the bio-metric community in particular and the computer vision community in general. Over the last two decades, methods for face segmentation have received increasing attention due to their diverse applications in several human-face image analysis tasks. Although many algorithms have been developed to address the problem, face segmentation is still a challenge not being completely solved, particularly for images taken in wild, unconstrained conditions. In this paper, we present a comprehensive review of face segmentation, focusing on methods for both the constrained and unconstrained environmental conditions. The article illustrates the advantages and disadvantages of previously proposed methods in state-of-the-art (SOA). The approaches presented comprise the seminal works on face segmentation and culminating in SOA approaches of the deep learning architecture. An extensive comparison of the previous approaches is intuitively presented, with a discussion of the potential directions for future research on the topic. We believe this comprehensive review and recap will contribute to a number of application domains, and will augment the knowledge of the research community.

INDEX TERMS Face segmentation, face image analysis, deep learning, machine learning.

I. INTRODUCTION

Image segmentation is of immense importance to mid-level computer vision tasks, that target the jointly grouping of image regions into coherent parts of objects. From an implementation point of view, it is the primary task in computer vision, which allows the computer to understand and see the image contents, classify a region or pixel in the image, and divide the image into different parts according to visual understanding. In this regard, extensive research work already exists on image segmentation, mainly reported in the Pascal Visuall Object Classes challenge [1].

Face segmentation is a basic task in face image analysis. In face segmentation, a computer-based algorithm segments a face image according to the different regions in the face. Semantic face segmentation allows the computer to understand the face image contents at the pixel level. As such, for semantic face segmentation, a number of complex features are also employed.

Face segmentation has received immense attention in computer vision, especially during the last two decades. For us in this article, two reasons account for this. First, many face analysis tasks can benefit from precise face segmentation, such as facial expression recognition [2], [3], head pose estimation [4], facial landmark detection [5], [6], sentiment analysis [7], etc. Secondly, although some low level of success has been achieved in the last 10 years, face segmentation is still an open challenge, particularly with images taken in unconstrained environments. For the above two reasons, face segmentation remains an open problem, and

The associate editor coordinating the review of this manuscript and approving it for publication was Ikramullah Lali.

many new SOA algorithms have been proposed from time to time.

Face region segmentation is generally regarded as the entry and starting points for most of the face image analysis tasks. Face segmentation is regarded as an essential and intermediary step for subsequent human face image analysis. This includes applications from the fields of bio-metric based identification and recognition, human indexing, and robotic control, all the way to medical and mental-state understanding. Face segmentation also plays a crucial role in the development of various intelligent environments.

II. FACE SEGMENTATION APPLICATIONS

Several computer vision applications expressively and exhaustively rely on a robust face segmentation output. As such, the following are (but not limited to) the typical face image analysis tasks which strongly rely on accurate face segmentation:

- **Preserving and completion of facial identity:** Due to the ill-posed nature of face images, face completion is quite challenging. Face completion refers to the task of filling those regions missed for one reason or other. These missed regions are filled with realistic synthesized contents. In classical methods, local information is first searched in order to find some existing patterns from the face image [8]–[12]. The patterns are then pasted into the targeted holes. These classical methods rely on low-level features; therefore, they fail when certain patches are not present within the target image. Therefore, face segmentation is one of the best approaches for face identity persevering and face completion. A recent method that uses face parsing for face completion is reported in [13]. As such, the last mentioned approach tackles the two tasks of face parsing and face completion using augmentation in a single framework.
- **Face de-blurring:** With the development and innovations in face parsing approaches, the methods for face image de-blurring have taken new directions. The face de-blurring problem is addressed by exploiting semantic cues between different face regions. As the human face shares various semantic components (eyes, nose, mouth, and chin), semantic cues provide sufficient information for image restoration. Face image de-blurring recovers a comparably high-quality image from a blurred input image. Conventionally, the blurring process involves convolution of a latent clear image with a blur filter. The process thus formulates this de-blurring problem on a maximum a posteriori framework. The most significant approaches to face segmentation for addressing face de-blurring are addressed in [14], and [15].
- **Facial landmark extraction:** Facial landmarks play an essential role in human face-image analysis. Typically, facial landmarks include the important face regions such as the nose, mouth, eyebrows, and eyes. This is a set of high-level features that can easily be differentiated with the naked eye. Typically, facial landmarks can be detected with a traditional machine learning paradigm. It involves the training of machine learning models on facial features using a comprehensive data set. However, these methods have shown lower performance in unconstrained circumstances, for example, in overlapped faces, and wild, low and non optimized lighting conditions. To overcome this problem, facial landmark extraction through semantic face segmentation has been proven an optimal way. Some recent methods that use face parsing for landmark detection are discussed in [5] and [16].
- **Face swapping:** Transferring one face from a source image onto a face appearing in a target image to generate un-edited and realistic looking results is generally termed as the face-swapping. Face-swapping has a number of applications. It is used for preserving privacy [17], data augmentation methods [18], and digital forensics [19]. It is particularly helpful in applications where training data is very scarce, for example, emotion recognition. With the recent developments in face parsing, face swapping has attracted considerable attention, as proved in [18], [20]. To swap a face while not considering their surrounding context, a per-pixel based label information is needed. This information is thus optimally provided by the semantic face segmentation.
- **Face beautification** Makeup makes people more attractive and beautiful. In the market, there are several commercial makeup systems for faces. For example, a virtual hair style (http://www.hairstyles.knowage.info) provides a virtual approach to try different hair styles. Some cosmetic elements such as lipsticks, eyeliners, etc. are used for beautification in the makeover, such as the Taaz (http://www.taaz.com). However, all such applications strongly rely on pre-determined cosmetics, which normally do not satisfy users' needs. Semantic face segmentation provides an easy way to design a real application system that automatically recommends suitable makeup for people. Some recent approaches which use face segmentation for face beautification can be explored in [21]–[23].
- **Head pose estimation:** Head pose estimation predicts the orientation of head from a face image. More specifically, the output of such a system consists of the pitch, yaw, and optionally the roll angles (3D space). Arguably, a powerful relationship exists between different face parts and its corresponding pose. Some excellent methods for estimating head pose using the face segmentation can be explored in [2], [4], [24], [25].
- **Facial expression, age, race, and gender recognition:** Psychology literature claims that close interaction exists between face parts and several other hidden variables in faces [26], [27]. These hidden variables include facial expressions, age, gender, and race information, etc. Face segmentation contributes different face parts information, which helps the human visual system in the

identification of these variables. The literature reported some excellent methods which segment a face image into various semantic parts such as nose, eyes, mouth, hair, skin, and back. These dense semantic class information are then used for modeling a framework for the above-mentioned face analysis tasks [2], [4], [25], [28].

- **Portrait segmentation:** Portrait segmentation is an art from photography and painting. Artists seek to make portraiture stand out and dominant from its surroundings, for example, by making it sharper and brighter. The modern digital world can process a portrait image through photography. Some methods which use face segmentation algorithms in portraits creation are discussed in [29]–[31].
- **Face recognition:** Face recognition was challenging area of computer vision ten years back. Due to recent developments in computer vision methods, face recognition is now a mature area of research with many optimized algorithms. In SOA, face recognition is explored through face segmentation as well. In these approaches, a face image is first segmented into prominent and non prominent regions, and then subsequently performing the process of face recognition [32], [33].

III. MOTIVATION

The face segmentation contributes to a large number of applications with further immense potential for computer vision paradigm; however, face segmentation is far from being solved, especially in the wild and unconstrained conditions and still presents many open challenges. There are also success stories as well, and some convincing work has been reported for face segmentation, especially in the controlled environmental conditions. However, under unconstrained scenarios, the problem of face segmentation is still open for research. Several environmental factors contribute to the robust face segmentation and affecting the performance of an efficient face segmentation system. These include but not limited to; occlusions, changes in illumination conditions, noise in various forms, changes in facial expressions and head poses, etc. Moreover, the number of available datasets for face segmentation is minimal. There are only three major datasets with some convincing data, including HELEN [34], FASSEG [35], and LFW [39]. A subset of LFW is used in literature for face segmentation with name LFW-PL. HELEN is the only dataset with sufficient data and a large number of classes.

In most of the computer vision tasks that are based on the model learning paradigm, the availability of data for model learning is a crucial and essential requirement for the success of the concerned task. The unavailability of a large face dataset is one major bottleneck towards the development of a mature and optimal face segmentation system. Also, over the last decade, some methods have been developed for face segmentation; however, researchers still need efforts towards the development of an optimal and accurate face segmentation framework. Such factors, variables, and issues in SOA motivate us to study the concerned area with a keen interest and analyze the developments, approaches, new applications, and directions in the face segmentation domain. Moreover, the shift from classical to deep learned approaches also motivates us for a very detailed and up-to-dated review, which can help several researchers and contribute to several applications and domains.

IV. CONTRIBUTIONS

Among the contributions, from this paper, we present a comprehensive review of face segmentation, focusing on methods for both constrained and unconstrained environmental conditions. We present the advantages and disadvantages of SOA methods by initially focusing on the seminal approaches for face segmentation, culminating in SOA approaches based on the deep learning architecture. A comparison of the previous approaches thus leads us in the potential directions for future research on the topic. We believe such a comprehensive review and recap will contribute to a number of application domains, and will augment the knowledge of the research community.

To the best of our knowledge, our proposed article in this direction is the first effort that combines the literature on face segmentation over the last two decades. We particularly focus on SOA face segmentation systems that have been developed over the last 10 years, with a focus on the shift in SOA towards the new paradigm of the deep learning architecture.

The rest of the paper is organized as follows: Section V presents a description of different datasets that are available for face segmentation. Section VI discusses the face segmentation approaches reported so far. Section VII presents a detailed comparison between the approaches reported to date using the available datasets. Finally, we conclude the article with a useful discussion and some future directions in Section VIII.

V. EXISTING DATASETS

The performance of the face segmentation model was evaluated with publicly available datasets. Face parsing is a relatively new and less explored research area. There is very little available data related to the face parsing task. In this section, we discuss the available face segmentation datasets, that can be used as benchmarks. A list of all face parsing databases and details is presented in Table 1.

A. HELEN

HELEN consists of face images that are collected from Flicker, having much diversity compared to other SOA datasets. A face detection algorithm was run to extract face parts information. The images were searched in Fliker with keywords such as "family," "outdoor", "boy", etc. All the images are high-quality, as low-quality images were filtered out after collection. These images were manually annotated through Amazon Mechanical Truck, which helps somehow to locate face parts such as eyes, nose, mouth, etc easily. This manual annotation requires an unusual amount of review,

TABLE 1. Major face segmentation datasets.

Database	Year	No. of face classes	Face classes	Total no. of images	From the wild conditions
FASSEG [35]	2019	6	skin, nose, hair, mouth, eyes, hair	420	no
Figaro1k [36]	2018	1	hair	1,050	yes
Multi-Face [37]	2017	3	skin, back, and hair	9,645	yes
CelebA [38]	2015	1	hair	250,000	yes
LFW-PL [39]	2013	3	skin, back, and hair	2,927	yes
HELEN [34]	2012	11	back, nose, skin, left brow, right brow, left eye, right eye, upper lip, lower lip, inner mouth, and hair	2,330	yes

FIGURE 1. Sample images from the HELEN [34] dataset.

as post-processing of the data is also needed immensely to ensure the best quality results. Still, this annotation is not accurate as only one person is involved in all the labeling process. Some images from the HELEN database are shown in Figure 1.

To the best of our knowledge, HELEN is the largest database so far addressing maximum face classes. The database consists of 2,330 face images which are annotated with 11 labels. These labels include back, skin, nose, left brow, right brow, left eye, right eye, upper lip, lower lip, inner mouth, and hair. The original database is divided into 2,000 training, 230 validation, and 100 testing images.

Some face parts of HELEN are inaccurately annotated; for example, the skin class is not accurately differentiated from hair and nose mostly. These images were re-annotated by Lin *et al.* [42], which they named the HELEN* database. Only training and validation sets were re-annotated; the testing set was kept as before. Figure 2 shows HELEN images and the corresponding HELEN* face images.

B. FASSEG

FASSEG contains around 500 face images, which were manually annotated into six face classes including, nose, hair, mouth, skin, eyes, and back. The database further consists of four subsets of face images. Three subsets are frontal images, namely frontal01, 02, and 03, whereas the fourth subset, namely Multipose01 consists of face images in multiple poses from $-90°$ to $+90°$.

The first subset (Frontal01) contains 70 RGB images and corresponding ground truth masks. Original RGB images are taken from two other databases, one from the Massachusetts Institute of Technology Center for Biological and Computational Learning [43] and the other from Faculdade de Engenharia Industrial (FEI) in Sao Paulo, Brazil [44]. Images were arranged in two separate folders, one for training and other for testing. The second version (Frontal02) contains the same set of images, but ground truth data was created with extreme care. Figure 3 shows some images from FASSEG V-0 and V-1.

FASSEG V-2 contains 150 face images taken from another database Siblings [45]. Images in this version are high-resolution frontal images that are captured in different orientations, under changing illumination conditions, and with various facial expressions. Figure 4 shows some images from FASSEG V-2.

The last version contains multi-pose 294 face images, which were also manually annotated with image editing software. These images are of different poses between $-90°$ to $+90°$. The difference between the two poses is 30°.

FIGURE 2. Comparison of the HELEN and HELEN* databases. First row shows HELEN and second row corresponding HELEN* images.

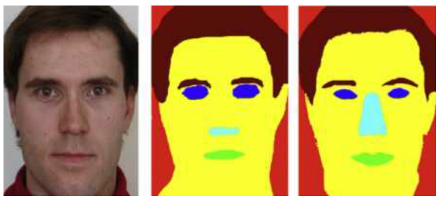

FIGURE 3. Face images from FASSEG database. From left hand side: original image, ground truth (V-0) and ground truth (V-1) respectively.

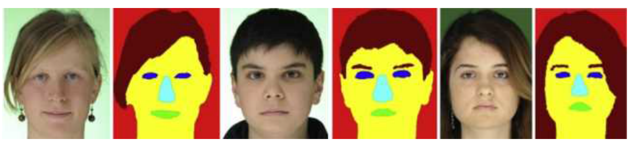

FIGURE 4. Sample original and corresponding ground truth images from FASSEG V-2 database.

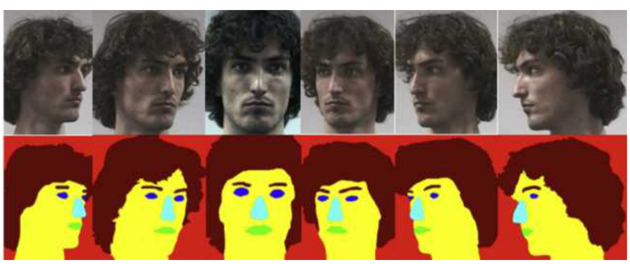

FIGURE 5. Sample face images from FASSEG V-3 database, first row shows original and second row shows the ground truth images.

These images were originally taken from the Pointing'04 [46] database. Figure 5 shows some images from this version of the database.

C. LFW

Labeled faces in the wild (LFW), with parts labeled (PL) database which is also known as (LFW-PL) does not provide class labels for the entire face image. LFW-PL contains 2,927 face images that were manually labeled into three classes: skin, back, and hair. The images were collected from the internet, and face detection was performed by Jones and Viola [47]. All images are collected in unconstrained conditions. LFW is divided into 1500 training, 520 validation, and 927 testing images.

D. MULTI-FACE

Multi-face database was also collected in unconstrained conditions. Multi-Face contains multiple faces in a single image. It is a larger database of 9, 645 face images. Pixel-wise labeling was performed for three classes: skin, back, and hair. The original database images are divided into 9,045 training, 200 validation, and 200 testing images. The image size was kept sufficiently large (512×512) to keep different face contents in good form.

E. Figaro1K

Figaro1k contains 1050 images. All images were annotated for hair class only. Each image is labeled with one of seven hairstyles, including wavy, straight, curly, kinky, braids, short, and dreadlocks. As the database was specifically developed for hair segmentation, the face is often not visible in the image. In some cases, multiple samples are captured for a single candidate. For this reason this database can be used only for hair segmentation.

F. CelebA

CelebA is a large database of more than 200,000 images. The number of participants in the database is also sufficiently large (10,000). These images were captured in unconstrained conditions. Several mis labelling of the hair class can be seen in the CelebA images. Only four hair classes, including black, brown, grey, and blond, are defined for hair tone.

As can be seen, the number of publicly available datasets for face segmentation is very limited. The last two datasets

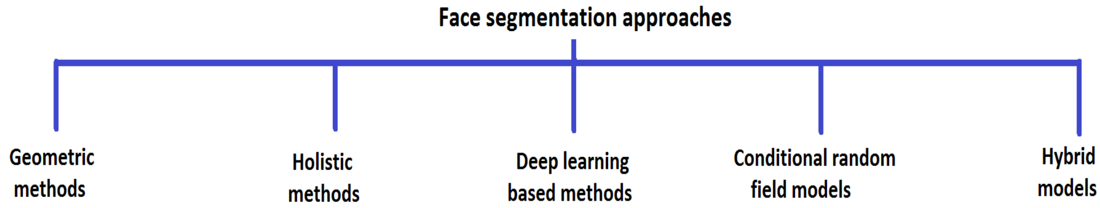

FIGURE 6. Taxanomy of face segmentation used in this paper.

provide class labels for hair only. However, some authors used limited sets of images from Figaro1k and CelebA for face segmentation [48]. Along with these standard datasets, some limited images of other databases are also used for face segmentation, [49]–[54]. However, images in these papers are mostly not available publicly, and details about the images are not provided in some cases. For more details, references in Table 3, 4, and 5 can be explored further.

VI. FACE SEGMENTATION APPROACHES

In this section, we discuss various methods used to address face parsing. We do not claim a generic taxonomy for face segmentation; instead, we organize each face segmentation system based on the fundamental method that underlines implementation. We discuss sufficient references where these proposed algorithms were used previously. We present a detailed discussion regarding the advantages and disadvantages of each algorithm. Figure 6 shows a taxonomy of the approaches discuss in this paper.

A. GEOMETRIC METHODS

Geometric methods are based on first localizing some facial points in a face image and then performing face parsing. These methods strongly rely on active shape modeling(ASM) [55], which is used for statistical shape modeling of various parts of the face. These methods also rely on discriminating clues among different face parts. ASM-based methods are similar to the way the human brain identifies various face parts.

In geometric methods, face parts are first localized and then pre-processed. After pre-processing, landmarks information is extracted, and affine transformation applied, which makes sure that pupils in the image are not disturbed with respect to location after alignment. As a result, facial constituents are of similar sizes in different face images.

A method for face segmentation through facial landmarks was introduced by Segundo *et al.* [49]. The framework was developed for automatic process- embedding face recognition with depth information. The segmentation method combined edge detection, shape analysis, and clustering for extracting face regions, whereas landmarks detection combined surface curvature information to find the eyes and nose. The influence of face segmentation on face recognition was also addressed in this work.

Hernandez *et al.* [6] proposed a facial expression recognition system using face parsing. The proposed algorithm first identified regions of interest. These regions included eyes, mouth, forehead. A face image was segmented into two regions, forehead and mouth. In the next phase, each region was segmented into blocks, where each block was characterized with 54 Gabor functions. In the next phase, dimensionality reduction was performed with principle component analysis (PCA). The final feature vector was then given to the support vector machine (SVM) for training.

Guo and Qi [56] addressed face parsing by applying a low-rank matrix decomposition method to face images. In the proposed work, the features parsing problem was formulated as sparse noise detection while recovering a low-rank matrix from the face image. To enhance the feature-parsing, a linear type of transformation matrix of linear type was learned, which further boosted the discriminant feature extraction phase. The algorithm was also extended to facial landmark extraction with derived parsing maps. The method was evaluated on some comprehensive SOA datasets.

A semantic face parsing method proposed in [57] is guided by specific pose information that is encoded in a set of keypoints or landmarks. The framework combined face parsing and facial landmarks estimation in a single model. The proposed framework was based on deep convolutional neural networks (CNNs). The model was evaluated on standard datasets LFW and HELEN, reporting better results compared to other SOA processes.

Luu *et al.* [5] combined ASM and the GrowCut algorithm to develop a face parsing system more robust to certain variations. The variations included multiple facial expression, lighting conditions, and some other environmental factors. Similarly, an automatic method for three-dimensional hair modeling was introduced by Chai *et al.* [58]. Hair modeling was performed from portrait images without user interaction. Several hair geometries were generated through this method, which also estimated hair growth direction along with hair segmentation. Some novel applications of the proposed method were also introduced, such as hairstyle space navigation and hair-aware image retrieval.

Merits and Drawbacks: These methods have some advantages over the other reported methods. For example, geometric methods are robust to pose, facial expressions, translation and rotation variations, and lighting conditions. Very less information is needed for the implementation of these methods; hence the computational cost is very less. ASM-based face parsing methods also face some severe drawbacks. These methods ignore skin and other parts facial texture information that are important cues for different face parts segmentation. As a result the relevant information is lost during feature extraction stage. Occlusions and far-field imagery conditions also significantly affect the performance of these methods. In short, the landmarks localization is itself very tricky and challenging.

B. HOLISTIC METHODS

In holistic methods features are extracted from statistical information by considering the image as one-dimensional vector. These methods always assume a specific relationship between 2D face image properties. Statistical learning methods are used to build a classifier through a large number of images.

Yacoob and Davis [59] built a model for hair segmentation, adopting a region growing strategy. The method faced difficulty when a significant change in hair color occurs from one region to another. The technique was facing specifically difficulty with dark color. Along with skin detection, lips and eye extraction were also explored. Hajiarbabi and Agha [60] introduced a method that performs face detection along with different face parts segmentation. The neural network was used as a classification tool in this work.

In the method proposed by Warrel and Prince [61], face parsing was addressed from an image labeling perspective. First, a per pixel unary classifier was learned, and then dense labeling of facial images for four face parts was estimated. The estimated face parts included mouth, hair, mustache, and hat. The proposed method considered large scale variations both in shape and appearance characteristics in the wild face images. The authors used the Adaboost classifier for classification.

A face segmentation algorithm based on learning vectore quantization was proposed in [62]. Neighboring neurons learned to identify adjacent segments of the input space. The authors tested the proposed framework under different illumination conditions, claiming robustness to illumination changes as compared to other SOA methods.

The problem of multi clas face parsing was addressed by Khan *et al.* [63]. The authors extended the labeled set into six face classes. The proposed framework was evaluated with only 70 images; 20 for training and 50 for testing phase. Three kinds of features, including shape, color, and position information were extracted to train a random forest classifier. The computational cost of the method was high as the algorithm was not properly optimized taking into account the speed factor. However, the authors tried different combinations of features and obtained pixel labeling accuracy of 93%, which is sufficiently high.

Seak *et al.* [64] proposed a face segmentation model utilizing a saliency map incorporating both top-down and bottom-up saliency methods. The top-down approach used skin color data, which was obtained from a training set to bias the skin saliency map. The bottom-up method utilized both color and intensity information maps from the testing images. The saliency map was computed from normalized extracted feature maps and center sound difference. A moving square window was used to find a point with comparatively high energy in the saliency map. The extracted point was marked as a facial region. The method was tested both with simple and complex background scenarios.

Scheffler and Obodez [51] introduced a model using four face classes; hair, back, skin, and clothing. For each label, a spatial prior was also learned. The local label consistency was encoded with a markov random field. The proposed model was evaluated on a limited set of images that were collected under constrained imaging conditions.

Subasic *et al.* [52] proposed a face segmentation algorithm, particularly for electronic identity document recognition. The proposed model segmented a face image into five regions: hair, skin, back, shoulders, and padding frame. The proposed algorithm consisted of two steps: over segmenting and face labeling. Image was segmented into homogeneous regions firstly, and then labeling of the regions was performed. For segmentation, the mean shift segmentation strategy was adapted, whereas for labeling Adaboost classifier was used.

A model robust to hair shape was introduced by Wang *et al.* [65]. The proposed model, which is exemplar-based modeling, has a particular focus on hair segmentation. The same authors improved the earlier work [66]. The output of the segmentation parts were regularized in the latter approach through labeling strategy. The later model was a statistical model, where each face part was utilized, and the co-occurrence probabilities between face parts were estimated. Another exemplar-based face parsing method was proposed by Smith *et al.* [67]. The proposed work was inspired from a general scene segmentation scenario. The method assumed a dataset of exemplar human face images, where each face image was associated with a manually labeled segmentation map. When a test image was given to the framework, a subset of exemplar face images from the dataset was selected. A nonrigid wrap was computed for each image to align the database image with the test image. Finally, labels from exemplar images were propagated to the testing images in a pixel-wise manner.

Hair is an essential element that characterizes people's appearance easily. A method proposed by Svanera *et al.* [68] can detect hair from the remaining face parts. The authors also proposed a novel multi-class image database in the paper. The image patches were classified into hair and non-hair parts. Two kinds of features were extracted from square patches, including histogram of oriented gradients (HOG) [69] and linear ternary patterns [70]. A segmentation accuracy of 85% was obtained with the proposed method. A binary classification method using neural networks was proposed by Guo and Arabi [71]. The proposed model was applied to hair and non-hair patches only. The proposed method used a pre-trained heuristic classifier for this task. The classifier segmented the image data into three clusters, including high-confidence negative, high-confidence positive, and lastly, low confidence set. The high confidence set initially trained a neural network, which further classified the low confidence set.

Another work in [36] tackled the problem of hair analysis. The authors performed three tasks, i.e., hair detection, segmentation, and style classification. First, hair probability maps were built by classifying patches through extracting features through CNNs. For classification, the authors used Random Forest. A database for various hairstyles and colors was also proposed, named Figaro1k. The hairstyle was classified into seven classes including, wavy, kinky, curly, braids, straight, short, and dreadlocks.

Two methods closely related to face parsing are proposed by Shen *et al.* [14], [15]. These methods adapted knowledge of one domain and achieved practical results for the respective applications. Another method introduced in [72] addressed face parsing and face beautification in a single framework. The first function made makeup recommendation for a person. The visually similar face was found from a database of images in the first stage. The second function transfered a reference face makeup to the desired makeup face. Five criteria were fixed for makeup transfer from the reference model to the desired model.

Merits and Drawbacks These methods have some advantages over other competitive methods. For example, these are simple methods with a straightforward implementation strategy. These methods do not need any negative training data in the training phase. Similarly, templates for the training phase can be added at any time for expansion, which allows the architecture to adapt to changes in conditions if needed. These methods are suited both for high and low-resolution images. Literature also reports some serious concerns regarding these methods. For example, holistic approaches assume that the system already detects the head part. The localization error which occurs degrade the system's accuracy. A large amount of training data is needed which makes the system computationally expensive. Similarly, No mechanism exists for handling occlusion problems. And lastly, the pairwise similarity is another issue face by these methods.

C. CONDITIONAL RANDOM FIELDS

Conditional random fields (CRFs) are frequently used for image labeling, and adjacent regions are well modeled with these methods. However, CRFs have limitations when dealing with complex scenarios. Complementary to CRFs, restricted Boltzmann machine (RBMs) are used to model global shapes that are produced by segmentation models. In some recent methods, CRFs are also combined with deep learning architectures, producing promising results. In this subsection, various methods where CRFs, CRFs combined RBMs, and CRFs combined with deep learning architectures are discussed.

An algorithm for face parsing through CRFs was proposed by Khan *et al.* [48]. In the proposed model, each node corresponded to superpixel, whereas the neighboring superpixels are connected through edges to the nodes. The label set was extended to six classes, including mouth, nose, hair, skin, back, and eyes. The proposed model was evaluated on three datasets FASSEG, Figaro, and LFW.

A method proposed in [73] combined the strength of both CRFs and RBMs in a single framework. The combined model predicted labels for three classes, including skin, hair, and background. The model was evaluated with challenging LFW database. The results obtained with the combined framework were much better with CRFs alone.

Face labeling through CRFs having unary and pairwise classifier was introduced by Liu *et al.* [38]. The proposed model was a multi-objective learning method that optimizes a deep CNNs with two different distinct non-structured loss functions. The network was regularized with a non-parametric prior channel in addition to the colored image, achieving better performance as compared to previous results. The algorithm was evaluated with two challenging datasets; LFW and HELEN.

A CRFs over a four connected graph, including CNNs, recurrent neural networks (RNNs) were trained, and then estimation was done via an adversarial process in work proposed by Guclu *et al.* [74]. The model learned both unary and pairwise potentials. The multi-scale contexts were aggregated while controlling higher-order inconsistencies.

A fully convolutional networks (FCNs) combined with CRFs architecture was used by Zhou *et al.* [75]. The proposed FCNs integrated three sub-networks, including unary, pairwise, and continuous CRFs into a single framework. Low-level details and high-level semantic information were extracted through convolutional and deconvolution structure. The pairwise network learned the semantic edge contexts. Based on the super-pixel pooling layer and continuous CRFs, the pairwise and unary networks were combined via a unique CRFs. The proposed framework was evaluated on HELN and LFW datasets.

A sparse FCNs for face parsing was proposed by Dong *et al.* [76]. As compared to other methods, FCNs have shown strong capabilities in learning representation, specifically for semantic segmentation, which was fully utilized by the authors. As FCNs mostly suffered from redundancy problems, which was solved by the authors through specific strategy. A group Lasso and intra group lasso regularization were used to sparsify the convolutional layers of the fully convolutinal networks. The regularization framework given the algorithm the capability of better feature selection and higher sparsity. The framework was also integrated with CRFs, which refined the output labels of the sparse FCNs. Segmentation accuracy was increased by 11% through sparse convolutional networks.

Merits and Drawbacks: CRFs are probabilistic frameworks used for various face parts labeling. In face parts labeling through CRFs we already define some conditional probability distribution over some specific label sequences. For example, it is very unlikely that nose region will be near the background or hair. The background label will be always close to either skin or hair class. This conditional nature results in the relaxation of the independence assumption that is required by the other competitive methods. Secondly, CRFs based methods also avoid the label bias problem which is faced by other modeling methods. Despite of the recent developments of face segmentation models, its modeling is still an open challenge. A unified shape model is still lacking that can capture all information of face parts in various rotations. It is also time consuming to generate a face segmentation model under various facial shape variations, as more complex model require more training data. For example, if there is occlusion problem, complete face segmentation is infeasible.

D. HYBRID MODELS

Hybrid methods are also known as multi-tasks methods. These methods were first introduced by Caruana [77]. As face segmentation is closely related to several other face image analysis tasks, including head pose estimation, facial expression recognition, gender recognition etc., some recent methods [76], [78], [79] proved that better performance can be obtained with a multi-tasks framework instead of single isolated task model. In this Subsection, hybrid models targeting such multi-tasks methods are discussed.

An interesting hybrid model addressing different facial attributes was introduced by Huang *et al.* [33]. A segmentation model was trained through a limited set of images. The proposed multi-tasks model was used for face recognition in the unconstrained conditions. The framework was also extended to head pose estimation with three simple posses, including frontal, left, and right profile images.

A face parsing model was introduced by Khan *et al.* [63]. The proposed model segmented an image into six semantic classes. The same work was extended by the authors to multi-tasks frameworks in some other papers [2]–[4], [24], [28]. The work proposed in [2], [3] was addressing three different tasks, including facial expression recognition, gender recognition, and head pose estimation. Authors of these papers performed extensive experiments and used different combinations of facial features. Similarly, [4], [24] included head pose estimation along with face segmentation. Some more advanced form of the hybrid frameworks were proposed in [25], where face segmentation, head pose estimation, facial expression and gender classification were included in a single unified frameworks.

An end-to-end face parsing model was presented in [80]. Wei *et al.* proposed a method that automatically regulated receptive fields and obtained better performance on face segmentation. The model generated facial labels in a multi-task scheme. Some improvement and advancements in the lastly mentioned method were proposed in [81] which predicted landmarks information along with segmentation results.

Ghiasi *et al.* [82] designed a model working both for face mask estimation and landmarks localization. This work used a deformable parts model, particularly for occluded faces. Another idea was presented in [83] based on graph cut refinement. The lastly mentioned two methods focused on differentiating face and non-face pixels.

Another hybrid model was proposed by Liu *et al.* [38]. In the first stage, face parsing mask was predicted and then semantic contours for facial parts are estimated. For different hair style classification, a mixture model was introduced by Lee *et al.* [84]. The model also learned color distributions for skin, and back.

Merits and Drawbacks: Many existing algorithms focus on a single face image analysis task. These methods do not consider interaction among some other latent factors within a face image. However, it has been proven that several face image analysis tasks have interactions with each other through latent factors. For example face segmentation is closely related to head pose estimation [24], facial expression and gender recognition [2], [28] etc. Such multi-task learning not only reduces algorithm design computational costs but hidden variables within a face image also help each other through knowledge transfer. Hybrid models jointly train multiple networks that consider the interaction between the target face segmentation task and some other secondary tasks such as head pose estimation, facial expression recognition etc. Such methods requires labeled data from all face image related tasks and in such cases the training phase becomes quite cumbersome, as more and more tasks are involved. This process not only increases the computational cost of the framework, bult also reduces the overall performance of the proposed framework.

E. DEEP LEARNING-BASED METHODS

Deep learning-based methods have shown excellent performance in different visual recognition tasks in recent days. Deep learning-based techniques improve the more complex scenarios in computer vision specifically. Several limitations in the machine learning methods are mitigated with the transition of these methods to the newly introduced deep learning architectures. In this subsection, various face segmentation methods developed through deep learning are discussed.

Two different convolution sampling paths were introduced by Zho [75]. The two sampling paths were named as top-down and bottom-up methods. To each convolutional integral, shared weight was also added. According to the authors, a shared weight improved the accuracy of the face parsing network. The proposed model has fast, as the number of parameters was less and the calculation speed rapid. The model was evaluated on two challenging datasets, HELEN and LFW. The work proposed by Wei *et al.* [85] targeted face parsing network for real-time interface speed. The structure of the traditional FCNs was revisited, and improvements were made to introduce a unique face parsing method. Normalized receptive field was added to the structure to make the system computationally better for real-time applications.

Another function named as statistical contextual loss was also added to the structure of the network. The statistical contextual function regularized features in the training phase. The network was further accelerated by a semi-supervised distillation scheme that transfered the learned knowledge to the network. The performance of the proposed system was evaluated on HELEN and LFW datasets, achieving better results as compared to previous results.

Khalil *et al.* [40] proposed a face parsing module using CNNs and a deep learning architecture. The performance of the proposed framework was evaluated with three datasets HELEN, LFW, and FASSEG. Lin *et al.* [42] proposed a face parsing method inspired by the physiological vision system of human. This was also a convolutional neural network which was addressing the dilemma between a limited sized region of interest and area of peripheral information. The HELEN dataset was also re-labeled as most of the regions in the HELEN images are mislabeled with near-by regions. The authors also used LFW along with HELEN for experiments.

Yan *et al.* [86] proposed a face parsing platform for mobile applications. The proposed model was evaluated on both iPhone and Android systems. The designed method is based on fully convolutional network which works on live face parsing. The proposed CNNs based model was implemented on the iPhone having the Apples's CoreML framework.

A novel face parsing method through a fully convolutional encoder/decoder network was presented in [87]. It is an end-to-end network that was optimized by minimizing the two-loss functions: the negative-log-likelihood and the L1 loss. The network accuracy was also improved with dilated convolution, transfer learning, and skip layer. To further enhance the network performance, maximum connected region extraction was also added to the output of the network. The proposed framework was evaluated with the HELEN database.

Borza *et al.* [88] proposed a method for hair analysis in images collected in the wild. Authors applied an FCNs in their method. A face image was segmented into three dense classes, including hair, face, and back. The proposed framework also provided information if the person has baldness or not. The hair tone was also predicted through a color recognition module. Color features at the super-pixel level were extracted, and then the random forest classifier was used for training. The authors of the paper also contributed a database of 3,500 images. Images from the CelebA database were re-labeled as the labeling process in CelebA is not accurate.

A hierarchical representation of CNNs and RNNs was combined in [38], [89]. The RNNs part enabled an excellent interface over a global space with the help of semantic edges, which were generated by a local CNNs model. The suggested framework was fast, as the CNNs architecture is shallow, and the RNNs have very few parameters due to their spatial nature. The model was applied to both coarse and fine-grained parsing tasks. A two-stage strategy was adopted for fine-grained parsing, first identifying the regions of interest and secondly segmenting the detailed components. The method achieved SOA performance on challenging datasets.

Primarily motivated by semantic-scene understanding, another face parsing method using FCNs was introduced by Vijay *et al.* [90]. The framework consisted of an encoder and a decoder followed by a classification layer. The encoder architecture was topologically same to the thirteen convolutional layers in the VGG16 [91] framework. The proposed network was named SegNet by the authors. To perform non-linear up sampling, the decoder used pooling indices in the max-pooling stage. The proposed face segmentation framework was tested against SOA methods, claiming better results as compared to previous results.

Saito *et al.* [92] proposed a face parsing method considering in the wild environment. According to the authors, the performance of the face parsing frameworks decreased when exposed to datasets collected in the un-controlled conditions. The proposed algorithm also considered conditions such as occlusion, accessories, visual artifacts, and other environmental factors. Non-face regions were initially masked out in the proposed method. The background concept of the algorithm was based on re-purposing CNNs, which are originally designed for general semantic image segmentation. The performance of the framework was improved through specific strategies for data augmentation and designing better complementary characteristics.

A face attribute classification method based on an attribute aware correlation map and CNNs was proposed in [93]. The correlation information between attribute label and pixel location information was provided by attribute aware correlation map. The correlation map of a specific attribute provided sufficient information regarding various regions where relevant segment features were extracted. Different relevant face parts regions were discovered through correlation maps of the particular attributes. Columns of CNNs were trained through face parts information. The proposed framework was evaluated with a subset of LFW dataset.

Merits and Drawbacks: The performance of traditional machine learning methods was impressive with images collected in controlled laboratory conditions. However, when these traditional machine learning methods were exposed to images collected in the wild, their performance decreased significantly. Unlike traditional machine learning methods, deep learning base methods learn a higher level of abstraction from input face image data. The need for feature engineering is much reduced with these methods. Consequently, these methods outperform previous results collected with traditional machine learning methods. Along with advantages, these methods are also facing some serious drawbacks. Deep learning is very complicated procedure which requires different choices by the practitioner. For example, setting the transfer and activation functions of the training algorithm and son on. Researchers mostly rely on a trail and error method to know about more proper model. As a result, deep learning based methods tend to take more time as compared to traditional machine learning methods. In nutshell, deep learning based methods are the definitive methods for solving face parsing problem, but still their use has been sporadic till date. Since these methods are comparatively recent, still the need of establishing their usefulness for face segmentation task is needed. Training a deep learning based face parsing model with many hidden layers and flexible filters is more effective way to learn high and deep level features. However, this process may under-perform if insufficient training data is available.

VII. COMPARATIVE ANALYSIS AND DISCUSSION

We perform a detailed comparison of the existing SOA methods on all available face segmentation datasets. Along with HELEN, FASSEG, and LFW; Multi-Face is another database that provides face labels for three parts, however it has been less used in the literature to date. Details about all datasets are summarized in Table 1. Reported results for all the datasets used between 2009-2020 are summarized in Table 3, 4, and 5. Some concluding remarks that emerged from the propsed methods and reported results are summarized in the following paragraphs:

- The number of datasets introduced for face segmentation is limited. The images ranges from flat background such as FASSEG to more complex scenario, for example LFW. Although the number of images in LFW are comparatively less, however, images are collected in the unconstrained conditions. As compared to FASSEG and LFW, HELEN is a better choice as the number of images are also large and sufficient face classes are included.
- Since only three major datasets exist for complete face parsing, most of the methods follow some standard image setting. LFW has 2,927 images that are collected from the internet, all having unconstrained imaging conditions. The size of these images is 250 × 250 pixels. The LFW-PL database is divided into three parts with 1500 training, 520 validation, and 927 testing images. HELEN is a comparatively larger dataset with 2,330 face images. Class labels are provided for 11 different parts. The images size is 400 × 400 pixels. A standard image setting followe are such that 2,000 images are used for training, 230 validation, and remaining 100 for testing purposes.
- Most of the papers are evaluated with frontal images, as HELEN and LFW are available with frontal images only. Only FASSEG is available in both frontal and profile cases. As quality of the images in FASSEG is comparatively better, the reported results for profile face images are also satisfactory. For more detailed results on FASSEG, these papers [2], [24], [25] can be explored.
- Ground truth data for all the three datasets HELEN, FASSEG, and LFW are created manually. All masks have been produced through a commercial image editing software. In this kind of labeling, no automatic segmentation tool is used. Such kind of labeling is highly depended on the subjective perception of a single person involved in labeling. Hence, it is not possible to provide an accurate label to each and every pixel in a face image. Differentiation and labeling of certain boundary regions are particularly challenging in some cases; for example, the nose region from the skin region can not be differentiated easily. Ground truth data for FASSEG is more accurate as compared to HELEN and LFW, as the dataset is available in three different versions. The authors of the FASSEG tried their best to create precise ground truth, by bringing improvement in each version. However, the number of classes in FASSEG is 6, which is less as compared to HELEN having 11 classes. The LFW ground truth ranks third as far as creating ground truth is concerned. As images in LFW are collected in the wild conditions and quality of the images is also very poor, hence different regions are mixed with each other. The number of class labels in LFW is also only three.
- Two kinds of evaluation metrics are used for face parsing: F-measure and pixel labeling accuracy (PLA). For more details, results are also reported in the form of confusion matrices by some papers [4], [25], [63], [98]. Results reported for FASSEG are in PLA, whereas remaining datasets are evaluated with F-Measure. Confusion matrix presents a clear picture of the results, as it is clear which classes are mis-classified with others. In Table 3 I-mouth represent F-measure/PLA for inner mouth, U/L-lip for upper/lower lip and the overall F-measure/PLA represents a union of all face components including brow, eyes, nose, and mouth labels.
- Face segmentation is an active research topic in computer vision. Table 2 presents a summary of the research conducted between 2009 to 2020 on the face segmentation. Table 3, 4, and 5 present more clear picture as detailed results for each algorithm are reported. The F-Measure and PLA values are reported directly from the original papers. As can be noticed from Table 3, 4, and 5 F-measure and PLA values are improving day by day on the standard datasets with deep learning.
- Some authors claim that a detailed look at the PLA and F-measure values reveals that SOA performance on both deep learning and traditional machine learning is almost same on face segmentation task. For example [2], [24], [25] show that traditional machine learning methods perform better than the newly introduced deep learning methods. From the previously mentioned paper [2], [24], [25] it can be noticed that classical methods where handcrafted features are utilized perform better than deep learning based methods. Through this comparison, the authors are not justifying that deep learning based methods perform poorly as compared to traditional machine learning methods, rather they argue that much better understanding of the deep learning methods and its implementation to face segmentation task is needed. For example in [2], [24], [25], a possible reason for poor performance reported by the authors is

TABLE 2. **Head pose estimation, year wise development.**

Year	Reported paper	Method used	Databases used
2020	Khalil et al. [40]	Deep learning	LFW, FASSEG and HELEN
2019	Liu et al. [41]	Deep learning	LFW and HELEN
	Lin et al. [42]	Deep learning	HELEN and LFW
	Wei et al. [80]	Deep learning	LFW and HELEN
	Benini et al. [2]	Appearnce based	FASSEG
	Ma et al. [13]	Deep learning	HELEN and CelebA
2018	Qu et al. [87]	Deep learning	HELEN
	Nirkin et al. [94]	Deep learning	Caltech
	Yan et al. [86]	Deep learning	HELEN
	Dong et al. [76]	CRFs	LFW-PL
	Borza et al. [88]	Deep learning	CelebA and Figaro1k
	Umar et al. [36]	Appearance based	Figaro1k
2017	Guclu et al. [74]	CRFs	LFW and HELEN
	Nirkin et al. [20]	Deep learning	LFW
	Liu et al. [38]	Deep learning	HELEN and LFW
	Zhou et al. [75], [89]	CRFs	HELEN and LFW
	Li et al. [95]	Deep learning	HELEN and CelebA
	Khan et al. [48]	CRFs	FASSEG, Figaro1k, LFW
2016	Jackson et al. [81]	Deep learning	HELEN
	Saito et al. [83]	Deep learning	Caltech
	Svanera et al. [68]	Appearance based	Figaro1k
	Hernandez et al. [96]	Appearance based	KDEF
	Jackson et al. [81]	Deep learning	HELEN
2015	Kang et al. [93]	Deep learning	LFW
	Ghiasi et al. [82]	Appearance based	Caltech
	Liu et al. [38]	CRFs	HELEN and LFW
	Liang et al. [54]	Deep learning	ATR
	Guo et al. [56], [97]	Geometric methods	LFW
	Hajiarbabi et al. [60]	Appearance based methods	LFW
	Hernandez et al. [6]	Appearance based	KDEF
2014	Ding et al. [53]	Appearance based	Caltech
2013	Smith et al. [38]	Appearance based	HELEN and LFW
	Kae et al. [73]	CRFs	LFW
2012	Luo et al. [18]	Deep learning	LFW
	Luu et al. [5]	Geometric method	Multi-PIE
	Wang et al. [66]	Appearance based method	LFW
	Subasic et al. [52]	Appearance based method	FERET and CALTECH
2011	Poppe et al. [50]	Appearance based	Dutch social network database
	Scheffler et al. [51]	Appearance based	Compaq skin database
2010	Wang et al. [65]	Appearance based	LFW
	Segundo et al. [49]	Geometric based	BU
2009	Warrell et al. [61]	Appearance based	LFW

TABLE 3. **Face segmentation results in the form of F1-measure and comparison of SOA methods for HELEN dataset for 11 different classes including skin, nose, upper lip, inner mouth, lower lip, eye, eyebrows, mouth, and hair.**

Database	Method	year	face	nose	u-lip	in-mouth	L-lip	eye	brows	mouth	hair	overall
HELEN	Khalil et al. [40]	2020	96.2	97.2	–	–	–	78.6	83.2	88.5	98.4	95.2
	Liu et al. [41]	2019	95.1	94.7	80.2	86.6	87.9	87.8	81.9	93.8	–	91.4
	Lin et al. [42]	2019	95.3	95.6	80.8	86.7	89.7	85.9	86.7	95.2	88.7	93.10
	Wei et al. [85]	2019	95.59	95.19	80.02	86.72	86.40	89.03	82.61	93.58	–	91.57
	Guclu et al. [74]	2017	94.36	94.04	79.66	85.50	86.23	88.73	82.26	92.82	–	90.99
	Zhou et al. [89]	2017	91.1	90.6	71.7	79.9	81.7	82.8	75.70	–	–	
	Wei et al. [80]	2017	91.48	93.65	–	–	–	84.66	78.61	91.48	–	90.21
	Liu et al [37]	2017	92.1	93.65	74.3	79.2	81.7	86.9	77.0	89.1	–	88.60
	zhou [89]	2017	–	95.0	75.4	83.6	80.9	87.4	81.3	92.6	–	87.30
	Jackson et al. [81]	2017	81.84	67.11	79.78	83.69	80.9	87.4	81.3	92.6	–	87.30
	Liu et al. [38]	2017	91.0	90.9	62.3	80.8	69.4	76.8	71.3	84.1	–	84.70
	Smith [67]	2017	88.2	92.2	65.1	71.3	70.0	78.5	72.3	85.7	–	80.40
	Qu et al. [87]	2017	93.0	92.2	73.2	71.9	80.8	81.1	75.0	88.8	–	87.20
	zhou [75]	2017	–	91.8	92.8	78.0	80.4	78.0	83.8	77.2	–	90.50

limited data scenario, which is a major weakness faced by deep learning based methods. However, we argue that deep learning-based methods show better results on many challenging datasets. We noticed that deep learning based methods (particularly CNNs) combined with CRFs perform much better as compared to CNNs alone. Although CNNs is an excellent tool for face segmentation, however, CNNs based methods face some

TABLE 4. Face segmentation results reported with pixel labeling accuracy for various methods for 6 different face classes including nose, hair, mouth, skin, eyes, and back.

Database	Method	year	face	nose	eye	mouth	back	hair	overall
FASSEG	Khalil et al. [40]	2020	96.58	87.80	84.30	89.30	94.54	98.20	95.12
	Khan et al. [28]	2019	98.65	75.83	84.2	77.23	95.5	96.65	–
	Khan et al. [48]	2017	97.53	59.23	84.56	74.53	98.65	97.65	92.47
	Khan et al. [4]	2017	93.22	58.75	58.72	83.25	80.23	94.31	–
	Khan et al. [63]	2015	93.39	29.83	82.2	70.23	92.5	95.14	92.95
	Khan et al. [98]	2018	95.5	69.2	67.6	86.33	87.9	97.6	92.95

TABLE 5. Face segmentation results reproted with F1-Measure for SOA methods for LFW dataset. The number of face classes are 3 including skin, back and hair.

Database	Method	year	face	back	hair	overall
LFW	Khalil et al. [40]	2020	96.80	97.20	94.20	93.25
	Lin et al. [42]	2019	95.77	98.26	88.31	96.71
	Wei [85]	2019	95.81	98.18	87.93	96.51
	Zhou et al. [89]	2016	94.37	97.55	83.43	95.46
	Liu et al. [37]	2017	94.10	96.46	85.16	95.28
	Vijay et al. [90]	2017	93.15	84.18	95.25	93.56
	Chen et al. [99]	2016	91.17	78.85	94.95	92.49
	Liu [38]	2015	93.93	97.10	80.70	95.12
	Zheng et al. [100]	2015	92.79	82.75	96.35	94.12
	Kae [73]	2013	95.81	–	–	94.95

technical hurdles when applied to pixel-wise face parsing alone. These reasons are listed bellow;

– Firstly, very diverse, deeply contextual and mutual relationship among various face components for face segmentation should be addressed when providing pixel-wise labels to different face parts. For this reason, combination of CRFs and deep learning is a best choice.
– Secondly, to highlight and recognize the smaller/minor labels (e.g. eyes, eyebrows, nose, and mouth), the estimated label maps are necessarily to be detailed preserved. However, previous works of face segmentation with CNNs only predict the pixel labels of low resolution pixels. Their prediction is also very coarse, and not optimal for the fine grained segmentation.
– Third, the CRFs consider the face segmentation specific context constraints in a much better way, such as the smoothness of local super-pixels and uniqueness for a specific semantic region. The CNNs or other deep learning methods alone does not consider this factor effectively. All pixels within a super-pixel or nearby super-pixels should have more possibilities to be given the same semantic class label. The high probability of estimation for a class label from nearby super-pixels help the label inference through leveraging the specific location priors. Furthermore, to retain a specific region integrity, all pixels with the same region (e.g. skin) are predicted with same class label.
– Lastly, although the combined deep learning and CRFs is a very powerful tool for face segmentation, however the training and inference computational cost of such structured prediction is quite high. It must be noted that the computational cost increases due to deep learning and not CRFs.

- The performance of the traditional machine learning methods in controlled laboratory conditions is better as compared to SOA. However, when these methods are exposed to the datasets collected in the un constrained conditions, we noticed drop in performance. However, deep learning methods perform better with images collected in the un constrained conditions. For example LFW-PL is a dataset collected in the wild conditions. The performance of the traditional machine learning methods is comparatively poor with LFW-PL, however better results are obtained with deep learning methods. We argue again here, although face segmentation is not a fully explored area for deep learning architectures, still better results are obtained with the challenging datasets. A mix response exists of the traditional machine learning methods towards solving face segmentation problem. Our view regarding this, hybrid and CRFs based models produce better results comparatively. For more details, references in the Table 2 can be explored.
- Face segmentation is an active area in computer vision. Tremendous progress has been reported in the last five years. From the results reported it is clear that the F-measure and PLA values are increased day by day. We present summary of the paper published in the last ten in years in Table 2. Noting the trends of most of the other developments in computer vision, which are moving rapidly towards deep learning methods, for face segmentation the speed is not satisfactory. Given the major difficulty of the training stage in deep

learning methods, particularly face segmentation, knowledge transfer [101], [102] is the best option to be explored. In knowledge transferring, benefits from the already trained models are taken. A comparatively less investigated domain in case of knowledge transfer is the option of heterogeneous domain strategy adaption. Considering the deep learning methods for face segmentation task, the keywords are LSTMs, 3D convolution, temporal pooling and optical flow frames. The last and an important point to remember is, for performance improvement of the face parsing systems, carefully and better managed engineering methods are needed. For example, data augmentation [65] is a possible option to be adapted.

VIII. SUMMARY AND CONCLUDING REMARKS

Face segmentation is an essential intermediary step for many face analysis applications. Face segmentation provides rich and sufficient information for several mid-level vision tasks. Face parsing is particularly challenging when images are collected in the wild conditions. However, extensive research work on face segmentation particularly in the last five years resulted several achievements. Due to many applications, we argue that face segmentation in the current stage is beyond the grasp the face parsing systems, therefore, we call researcher working on face segmentation to improve the existing algorithms described in Section VI, particularly exploring the newly introduced deep learning based methods for face segmentation.

One of the main problems face segmentation task is facing is the un-availability of public datasets. We expect to see contribution to the task in the form of large scale and challenging datasets from the research community working on the topic. We also expect to see some excellent evaluations of the recently introduced deep learning methods on the challenging datasets, particularly, collected in the wild conditions in the form of future work. If an efficient face segmentation system is developed, the new face parsing system will have a profound effects on the large scale applications of face segmentation.

We presented a survey of the face segmentation method, also including details about the available face datasets. We investigated different aspects of the existing solutions for the face segmentation problem. We revived the existing methods starting from simple handcrafted representation and then moving to the recently introduced deep learning models. We provided a comparative analysis of the results obtained with SOA so far for face segmentation. Finally, we identified different open problems in face segmentation and presented an outlook into the future of face segmentation task.

REFERENCES

[1] M. Everingham and L. V. Gool, "Christopher KI Williams, John Winn, and Andrew Zisserman, "The Pascal visual object classes (VOC) challenge," *Int. J. Comput. Vis.*, vol. 88, no. 2, p. 303–338, 2010.

[2] S. Benini, K. Khan, R. Leonardi, M. Mauro, and P. Migliorati, "Face analysis through semantic face segmentation," *Signal Process., Image Commun.*, vol. 74, pp. 21–31, May 2019.

[3] K. Khan, M. Mauro, P. Migliorati, and R. Leonardi, "Gender and expression analysis based on semantic face segmentation," in *Image Analysis and Processing—ICIAP*, Catania, Italy, Sep. 2017, Art. no. 11e15.

[4] K. Khan, M. Mauro, P. Migliorati, and R. Leonardi, "Head pose estimation through multi-class face segmentation," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2017, pp. 253–258.

[5] K. Luu, T. H. N. Le, K. Seshadri, and M. Savvides, "Facecut—A robust approach for facial feature segmentation," in *Proc. 19th IEEE Int. Conf. Image Process.*, Sep. 2012, pp. 1841–1844.

[6] A. Hernandez-Matamoros, A. Bonarini, E. Escamilla-Hernandez, M. Nakano-Miyatake, and H. Perez-Meana, "A facial expression recognition with automatic segmentation of face regions," in *Int. Conf. Intell. Softw. Methodol., Tools, Techn.* Cham, Switzerland: Springer, 2015, pp. 529–540.

[7] S. Zohaib, K. Ahmad, N. Conci, and A. Al-Fuqaha, "Sentiment analysis from images of natural disasters," 2019, *arXiv:1910.04416*. [Online]. Available: http://arxiv.org/abs/1910.04416

[8] C. Ballester, M. Bertalmio, V. Caselles, G. Sapiro, and J. Verdera, "Filling-in by joint interpolation of vector fields and gray levels," *IEEE Trans. Image Process.*, vol. 10, no. 8, pp. 1200–1211, Aug. 2001.

[9] C. Barnes, E. Shechtman, A. Finkelstein, and D. B. Goldman, "PatchMatch: A randomized correspondence algorithm for structural image editing," in *ACM Trans. Graph.*, vol. 28, no. 3, 2009, Art. no. 24.

[10] C. Barnes, E. Shechtman, D. B. Goldman, and A. Finkelstein, "The generalized PatchMatch correspondence algorithm," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2010, pp. 29–43.

[11] J. Jia and C.-K. Tang, "Image repairing: Robust image synthesis by adaptive ND tensor voting," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2003, p. 1.

[12] D. Kundu and B. L. Evans, "Full-reference visual quality assessment for synthetic images: A subjective study," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 2374–2378.

[13] R. Ma, H. Hu, W. Wang, J. Xu, and Z. Li, "Photorealistic face completion with semantic parsing and face identity-preserving features," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 15, no. 1, pp. 1–18, Feb. 2019.

[14] X. Shen, A. Hertzmann, J. Jia, S. Paris, B. Price, E. Shechtman, and I. Sachs, "Automatic portrait segmentation for image stylization," *Comput. Graph. Forum*, vol. 35, no. 2, pp. 93–102, May 2016.

[15] X. Shen, X. Tao, H. Gao, C. Zhou, and J. Jia, "Deep automatic portrait matting," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 92–107.

[16] H. Kim, J. Park, H. Kim, and E. Hwang, "Facial landmark extraction scheme based on semantic segmentation," in *Proc. Int. Conf. Platform Technol. Service (PlatCon)*, Jan. 2018, pp. 1–6.

[17] D. Bitouk, N. Kumar, S. Dhillon, P. Belhumeur, and S. K. Nayar, "Face swapping: Automatically replacing faces in photographs," *ACM Trans. Graph.*, vol. 27, no. 3, 2008, Art. no. 39.

[18] P. Luo, X. Wang, and X. Tang, "Hierarchical face parsing via deep learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 2480–2487.

[19] M. A. Oikawa, Z. Dias, A. de R. Rocha, and S. Goldenstein, "Manifold learning and spectral clustering for image phylogeny forests," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 1, pp. 5–18, Jan. 2016.

[20] Y. Nirkin, I. Masi, A. Tran Tuan, T. Hassner, and G. Medioni, "On face segmentation, face swapping, and face perception," in *Proc. 13th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, May 2018, pp. 98–105.

[21] S. Liu, X. Ou, R. Qian, W. Wang, and X. Cao, "Makeup like a superstar: Deep localized makeup transfer network," 2016, *arXiv:1604.07102*. [Online]. Available: http://arxiv.org/abs/1604.07102

[22] D. Guo and T. Sim, "Digital face makeup by example," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 73–79.

[23] K. Scherbaum, T. Ritschel, M. Hullin, T. Thormählen, V. Blanz, and H. P. Seidel, "Computer-suggested facial makeup," *Comput. Graph. Forum*, vol. 30, no. 2, pp. 485–492, 2011.

[24] K. Khan, N. Ahmad, F. Khan, and I. Syed, "A framework for head pose estimation and face segmentation through conditional random fields," *Signal, Image Video Process.*, vol. 14, no. 1, pp. 159–166, Feb. 2020.

[25] K. Khan, M. Attique, I. Syed, G. Sarwar, M. A. Irfan, and R. U. Khan, "A unified framework for head pose, age and gender classification through End-to-End face segmentation," *Entropy*, vol. 21, no. 7, p. 647, 2019.

[26] G. Davies, H. Ellis, and J. Shepherd, *Perceiving and Remembering Faces*. New York, NY, USA: Academic, 1981.

[27] P. Sinha, B. Balas, Y. Ostrovsky, and R. Russell, "Face recognition by humans: Nineteen results all computer vision researchers should know about," *Proc. IEEE*, vol. 94, no. 11, pp. 1948–1962, Nov. 2006.

[28] K. Khan, M. Attique, I. Syed, and A. Gul, "Automatic gender classification through face segmentation," *Symmetry*, vol. 11, no. 6, p. 770, 2019.

[29] X. Shen, A. Hertzmann, J. Jia, S. Paris, B. Price, E. Shechtman, and I. Sachs, "Automatic portrait segmentation for image stylization," *Comput. Graph. Forum*, vol. 35, no. 2, pp. 93–102, May 2016.

[30] X. Shen, X. Tao, H. Gao, C. Zhou, and J. Jia, "Deep automatic portrait matting," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 92–107.

[31] H. Averbuch-Elor, D. Cohen-Or, J. Kopf, and M. F. Cohen, "Bringing portraits to life," *ACM Trans. Graph.*, vol. 36, no. 6, pp. 1–13, Nov. 2017.

[32] A. Sepas-Moghaddam, M. S. Moin, and H. R. Kanan, "A new preselection method for face recognition in JPEG domain based on face segmentation," in *Proc. IEEE Int. Conf. Signal Image Process. Appl. (ICSIPA)*, Nov. 2011, pp. 132–137.

[33] G. B. Huang, M. Narayana, and E. Learned-Miller, "Towards unconstrained face recognition," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2008, pp. 1–8.

[34] B. Smith, L. Zhang, J. Brandt, Z. Lin, and J. Yang, "Exemplar-based face parsing," in *Proc. CVPR*, 2013, pp. 3484–3491.

[35] S. Benini, K. Khan, R. Leonardi, M. Mauro, and P. Migliorati, "FASSEG: A FAce semantic SEGmentation repository for face image analysis," *Data Brief*, vol. 24, Jun. 2019, Art. no. 103881.

[36] U. R. Muhammad, M. Svanera, R. Leonardi, and S. Benini, "Hair detection, segmentation, and hairstyle classification in the wild," *Image Vis. Comput.*, vol. 71, pp. 25–37, Mar. 2018.

[37] S. Liu, J. Shi, J. Liang, and M.-H. Yang, "Face parsing via recurrent propagation," 2017, *arXiv:1708.01936*. [Online]. Available: http://arxiv.org/abs/1708.01936

[38] S. Liu, J. Yang, C. Huang, and M.-H. Yang, "Multi-objective convolutional learning for face labeling," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3451–3459.

[39] A. Kae, K. Sohn, H. Lee, and E. Learned-Miller, "Augmenting CRFs with Boltzmann machine shape priors for image labeling," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 2019–2026.

[40] K. Khan, M. Attique, R. U. Khan, I. Syed, and T.-S. Chung, "A multitask framework for facial attributes classification through end-to-end face parsing and deep convolutional neural networks," *Sensors*, vol. 20, no. 2, p. 328, 2020.

[41] Y. Liu, H. Shi, Y. Si, H. Shen, X. Wang, and T. Mei, "A high-efficiency framework for constructing large-scale face parsing benchmark," 2019, *arXiv:1905.04830*. [Online]. Available: http://arxiv.org/abs/1905.04830

[42] J. Lin, H. Yang, D. Chen, M. Zeng, F. Wen, and L. Yuan, "Face parsing with RoI tanh-warping," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 5654–5663.

[43] MIT Center for Biological and Computational Learning (CBCL). *Mit-CBCL Database*. [Online]. Available: http://cbcl.mit.edu/software-datasets/ FaceData2.html

[44] Centro Universitario da FEI. *Fei Database*. [Online]. Available: http://www.fei.edu.br/ecet/facedatabase.html

[45] T. F. Vieira, A. Bottino, A. Laurentini, M. De Simone, *Detecting siblings in image pairs*, *Vis. Comput.*, vol. 30, no. 12, pp. 1333–1345, 2014.

[46] N. Gourier, D. Hall, and J. L. Crowley, "Estimating face orientation from robust detection of salient facial structures," in *Proc. FG Net Workshop Vis. Observ. Deictic Gestures (FGnet)*, Cambridge, U.K., 2004, p. 1e9.

[47] M. Jones and P. Viola, "Fast multi-view face detection," Mitsubishi Electr. Res. Lab., Cambridge, MA, USA, Tech. Rep. TR-20003-96, 2003, no. 14, p. 2.

[48] K. Khan, N. Ahmad, K. Ullah, and I. Din, "Multiclass semantic segmentation of faces using CRFs," *TURKISH J. Electr. Eng. Comput. Sci.*, vol. 25, no. 4, pp. 3164–3174, 2017.

[49] M. P. Segundo, L. Silva, O. R. P. Bellon, and C. C. Queirolo, "Automatic face segmentation and facial landmark detection in range images," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 40, no. 5, pp. 1319–1330, Oct. 2010.

[50] R. Poppe, "Scalable face labeling in online social networks," in *Proc. Face Gesture*, Mar. 2011, pp. 566–571.

[51] Scheffler, Carl, and Jean-Marc Odobez, "Joint adaptive colour modelling and skin, hair and clothing segmentation using coherent probabilistic index maps," in *Proc. Brit. Mach. Vis. Assoc.-Brit. Mach. Vis. Conf.*, 2011, pp. 1–26.

[52] M. Subašić, S. Lončarić, and A. Heđi, "Segmentation and labeling of face images for electronic documents," *Expert Syst. Appl.*, vol. 39, no. 5, pp. 5134–5143, Apr. 2012.

[53] X. Ding, Y. Luo, L. Sun, and F. Chen, "Color balloon snakes for face segmentation," *Optik*, vol. 125, no. 11, pp. 2538–2542, Jun. 2014.

[54] X. Liang, S. Liu, X. Shen, J. Yang, L. Liu, J. Dong, L. Lin, and S. Yan, "Deep human parsing with active template regression," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 12, pp. 2402–2414, Dec. 2015.

[55] F. Fleuret and D. Geman, "Fast face detection with precise pose estimation," in *Proc. 16th Object Recognit. Interact. Service Robots*, vol. 1, Aug. 2002, pp. 235–238.

[56] R. Guo and H. Qi, "Facial feature parsing and landmark detection via low-rank matrix decomposition," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 3773–3777.

[57] A. S. Jackson and M. V. G. Tzimiropoulos, "A CNN cascade for landmark guided semantic part segmentation," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 143–155.

[58] M. Chai, T. Shao, H. Wu, Y. Weng, and K. Zhou, "Autohair: Fully automatic hair modeling from a single image," *ACM Trans. Graph.*, vol. 35, no. 4, 2016.

[59] Y. Yacoob and L. S. Davis, "Detection and analysis of hair," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 7, pp. 1164–1169, Jul. 2006.

[60] M. Hajiarbabi and A. Agah, "Techniques for skin, face, eye and lip detection using skin segmentation in color images," *Int. J. Comput. Vis. Image Process.*, vol. 5, no. 2, pp. 35–57, Jul. 2015.

[61] J. Warrell and S. J. D. Prince, "Labelfaces: Parsing facial features by multiclass labeling with an epitome prior," in *Proc. 16th IEEE Int. Conf. Image Process. (ICIP)*, Nov. 2009, pp. 2481–2484.

[62] E. Cuevas, D. Zaldivar, M. Perez, and E. N. Sanchez, "LVQ neural networks applied to face segmentation," *Intell. Autom. Soft Comput.*, vol. 15, no. 3, pp. 439–450, Jan. 2009.

[63] K. Khan, M. Mauro, and R. Leonardi, "Multi-class semantic segmentation of faces," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 827–831.

[64] Z.-Q. Seak, L.-M. Ang, and K.-P. Seng, "Face segmentation using combined bottom-up and top-down saliency maps," in *Proc. 3rd Int. Conf. Comput. Sci. Inf. Technol.*, vol. 5, Jul. 2010, pp. 477–480.

[65] N. Wang, H. Ai, and S. Lao, "A compositional exemplar-based model for hair segmentation," in *Proc. Asian Conf. Comput. Vis.* Berlin, Germany: Springer, 2010, pp. 171–184.

[66] N. Wang, H. Ai, and F. Tang, "What are good parts for hair shape modeling?" in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 662–669.

[67] B. M. Smith, L. Zhang, J. Brandt, Z. Lin, and J. Yang, "Exemplar-based face parsing," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 3484–3491.

[68] M. Svanera, U. R. Muhammad, R. Leonardi, and S. Benini, "Figaro, hair detection and segmentation in the wild," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 933–937.

[69] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2005, pp. 886–893.

[70] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE Trans. Image Process.*, vol. 19, no. 6, p. 1635–1650, Jun. 2010.

[71] W. Guo and P. Aarabi, "Hair segmentation using heuristically-trained neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 1, pp. 25–36, Jan. 2018.

[72] S. Liu, X. Ou, R. Qian, W. Wang, and X. Cao, "Makeup like a superstar: Deep localized makeup transfer network," 2016, *arXiv:1604.07102*. [Online]. Available: http://arxiv.org/abs/1604.07102

[73] A. Kae, K. Sohn, H. Lee, and E. Learned-Miller, "Augmenting CRFs with Boltzmann machine shape priors for image labeling," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 2019–2026.

[74] U. Güçlü, Y. Güçlütürk, M. Madadi, S. Escalera, X. Baró, J. González, R. van Lier, and M. A. J. van Gerven, "End-to-end semantic face segmentation with conditional random fields as convolutional, recurrent and adversarial networks," 2017, *arXiv:1703.03305*. [Online]. Available: http://arxiv.org/abs/1703.03305

[75] Y. Zhou, ‘‘Top-down sampling convolution network for face segmentation,’’ in *Proc. 3rd IEEE Int. Conf. Comput. Commun. (ICCC)*, Dec. 2017, pp. 1893–1897.

[76] M. Dong, S. Wen, Z. Zeng, Z. Yan, and T. Huang, ‘‘Sparse fully convolutional network for face labeling,’’ *Neurocomputing*, vol. 331, pp. 465–472, Feb. 2019.

[77] R. Caruana, ‘‘Multitask learning,’’ in *Learning to Learning*. Springer, 1998, pp. 95–133.

[78] R. Ranjan, V. M. Patel, and R. Chellappa, ‘‘HyperFace: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition,’’ *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 1, pp. 121–135, Jan. 2019.

[79] X. Zhu, Z. Lei, X. Liu, H. Shi, and S. Z. Li, ‘‘Face alignment across large poses: A 3D solution,’’ in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, p. 146–155.

[80] Z. Wei, Y. Sun, J. Wang, H. Lai, and S. Liu, ‘‘Learning adaptive receptive fields for deep image parsing network,’’ in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2434–2442.

[81] A. S. Jackson, M. Valstar, and G. Tzimiropoulos, ‘‘A CNN cascade for landmark guided semantic part segmentation,’’ in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 143–155.

[82] G. Ghiasi and C. Fowlkes, ‘‘Using segmentation to predict the absence of occluded parts,’’ in *Proc. Brit. Mach. Vis. Conf.*, 2015, pp. 1–22.

[83] S. Saito, T. Li, and H. Li, ‘‘Real-time facial segmentation and performance capture from RGB input,’’ in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 244–261.

[84] K.-C. Lee, D. Anguelov, B. Sumengen, and S. B. Gokturk, ‘‘Markov random field models for hair and face segmentation,’’ in *Proc. 8th IEEE Int. Conf. Autom. Face Gesture Recognit.*, Sep. 2008, pp. 1–6.

[85] Z. Wei, S. Liu, Y. Sun, and H. Ling, ‘‘Accurate facial image parsing at real-time speed,’’ *IEEE Trans. Image Process.*, vol. 28, no. 9, pp. 4659–4670, Sep. 2019.

[86] Y. Yan, B. Bout, A. Berthelier, X. Naturel, and T. Chateau, ‘‘Face parsing for mobile AR applications,’’ in *Proc. IEEE Int. Symp. Mixed Augmented Reality Adjunct (ISMAR-Adjunct)*, Oct. 2018, pp. 407–408.

[87] X. Qu, C. Peng, T. Wei, P. Du, and C. Chen, ‘‘A novel face attribute segmentation algorithm,’’ in *Proc. 11th Int. Symp. Comput. Intell. Design (ISCID)*, Dec. 2018, pp. 132–135.

[88] D. Borza, T. Ileni, and A. Darabant, ‘‘A deep learning approach to hair segmentation and color extraction from facial images,’’ in *Proc. Int. Conf. Adv. Concepts Intell. Vis. Syst.* Cham, Switzerland: Springer, 2018, pp. 438–449.

[89] L. Zhou, Z. Liu, and X. He, ‘‘Face parsing via a fully-convolutional continuous CRF neural network,’’ 2017, *arXiv:1708.03736*. [Online]. Available: http://arxiv.org/abs/1708.03736

[90] V. Badrinarayanan, A. Kendall, and R. Cipolla, ‘‘SegNet: A deep convolutional encoder-decoder architecture for image segmentation,’’ *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[91] K. Simonyan and A. Zisserman, ‘‘Very deep convolutional networks for large-scale image recognition,’’ 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556

[92] S. Saito, T. Li, and H. Li, ‘‘Real-time facial segmentation and performance capture from RGB input,’’ in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 244–261.

[93] S. Kang, D. Lee, and C. D. Yoo, ‘‘Face attribute classification using attribute-aware correlation map and gated convolutional neural networks,’’ in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 4922–4926.

[94] Y. Nirkin, I. Masi, A. Tran Tuan, T. Hassner, and G. Medioni, ‘‘On face segmentation, face swapping, and face perception,’’ in *Proc. 13th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, May 2018, pp. 98–105.

[95] Y. Li, S. Liu, J. Yang, and M.-H. Yang, ‘‘Generative face completion,’’ in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3911–3919.

[96] A. Hernandez-Matamoros, A. Bonarini, E. Escamilla-Hernandez, M. Nakano-Miyatake, and H. Perez-Meana, ‘‘Facial expression recognition with automatic segmentation of face regions using a fuzzy based classification approach,’’ *Knowl.-Based Syst.*, vol. 110, pp. 1–14, Oct. 2016.

[97] R. Guo and H. Qi, ‘‘Facial feature parsing and landmark detection via low-rank matrix decomposition,’’ in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 3773–3777.

[98] K. Khan, I. Syed, M. S. Khan, M. E. Mazhar, I. Uddin, and N. Ahmad, ‘‘FPL—An end-to-end face parts labeling framework,’’ in *Proc. 24th Int. Conf. Autom. Comput. (ICAC)*, Sep. 2018, pp. 1–6.

[99] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, ‘‘DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs,’’ *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.

[100] S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, and P. H. S. Torr, ‘‘Conditional random fields as recurrent neural networks,’’ in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1529–1537.

[101] Y.-H.-H. Tsai, Y.-R. Yeh, and Y.-C.-F. Wang, ‘‘Learning cross-domain landmarks for heterogeneous domain adaptation,’’ in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 5081–5090.

[102] J. Hoffman, E. Rodner, J. Donahue, B. Kulis, and K. Saenko, ‘‘Asymmetric and category invariant feature transformations for domain adaptation,’’ *Int. J. Comput. Vis.*, vol. 109, nos. 1–2, pp. 28–41, Aug. 2014.

KHALIL KHAN received the B.S. degree in electrical engineering and the M.S. degree in computer engineering from the University of Engineering and Technology, Peshawar, Pakistan, in 2007 and 2012, respectively, and the Ph.D. degree from the Signals and Communication Laboratory, University of Brescia, Italy, in 2016. He is currently working as an Assistant Professor with the Department of Electrical Engineering, University of Azad Jammu and Kashmir, Muzaffarabad, Pakistan. His research interests include a wide range of topics within digital image processing, machine learning, and computer vision.

REHAN ULLAH KHAN received the B.Sc. and M.Sc. degrees in information systems from the University of Engineering and Technology, Peshawar, in 2004 and 2006, respectively, and the Ph.D. degree from the Vienna University of Technology, Austria, in 2011. He is currently an Assistant Professor and the Director of Research and Development at CoC, Qassim University, Saudi Arabia. His current research interests include pattern analysis and its applications in multiple disciplines.

KASHIF AHMAD received the bachelor's and master's degrees from the University of Engineering and Technology, Peshawar, Pakistan, in 2010 and 2013, respectively, and the Ph.D. degree from the University of Trento, Italy, in 2017. He was with the Multimedia Laboratory, DISI, University of Trento. He worked as a Postdoctoral Researcher with ADAPT, Trinity College, Dublin, Ireland. He is currently working as a Research Fellow with the Division of Computer Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar. He authored and coauthored more than 30 journal and conference publications. His research interests include multimedia analysis, computer vision, machine learning, and signal processing.

FARMAN ALI received the B.S. degree in computer science from the University of Peshawar, Pakistan, in 2011, the M.S. degree in computer science engineering from Gyeongsang National University, South Korea, in 2015, and the Ph.D. degree in information and communication engineering from Inha University, South Korea, in 2018. He was a Postdoctoral Fellow with the UWB Wireless Communications Research Center, from September 2018 to August 2019. He is currently an Assistant Professor with the Department of Software, Sejong University. His current research interests include data mining, deep learning, sentiment analysis, recommendation systems, healthcare monitoring systems, ontology, information extraction, information retrieval, and fuzzy logic.

KYUNG-SUP KWAK received the Ph.D. degree from the University of California at San Diego, in 1988. From 1988 to 1989, he was with Hughes Network Systems, San Diego, CA, USA. From 1989 to 1990, he was with the IBM Network Analysis Center, Research Triangle Park, NC, USA. Since then, he has been with the School of Information and Communication Engineering, Inha University, South Korea, as a Professor, where he was the Dean of the Graduate School of Information Technology and Telecommunications, from 2001 to 2002. He has been the Director of the UWB Wireless Communications Research Center (formerly Key National IT Research Center), South Korea, since 2003. In 2006, he served as the President of the Korean Institute of Communication Sciences, and in 2009, the President of Korea Institute of Intelligent Transport Systems. In 2008, he had been selected for an Inha Fellow Professor and currently for an Inha Hanlim Fellow Professor. He has published more than 200 peer-reviewed journal articles and served as the TPC/Track Chair/Organizing Chair for several IEEE related conferences. His research interests include wireless communications, UWB systems, sensor networks, WBAN, and nano communications. He was a recipient of the number of awards, including the Engineering College Achievement Award from Inha University, the LG Paper Award, the Motorola Paper Award, the Haedong Prize of Research, and various government awards from the Ministry of ICT, the President, and the Prime Minister of Korea, for his excellent research performances.