Computers in Biology and Medicine 174 (2024) 108431

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

A novel approach to the detection of facial wrinkles: Database, detection algorithm, and evaluation metrics

Zijia Liu [a,1], Quan Qi [b,1], Sijia Wang [b,*], Guangtao Zhai [a]

[a] Institute of Image Communication and Network Engineering, Shanghai Jiao Tong University, 800 Dongchuan RD, Shanghai, 200240, China
[b] Shanghai Institute of Nutrition and Health, Chinese Academy of Sciences, 320 Yueyang RD, Shanghai, 200031, China

ARTICLE INFO

Keywords:
Wrinkle detection
Wrinkle assessment metric
Skeleton similarity
Wrinkle database

ABSTRACT

Skin wrinkles result from intrinsic aging processes and extrinsic influences, including prolonged exposure to ultraviolet radiation and tobacco smoking. Hence, the identification of wrinkles holds significant importance in skin aging and medical aesthetic investigation. Nevertheless, current methods lack the comprehensiveness to identify facial wrinkles, particularly those that may appear insignificant. Furthermore, the current assessment techniques neglect to consider the blurred boundary of wrinkles and cannot differentiate images with varying resolutions. This research introduces a novel wrinkle detection algorithm and a distance-based loss function to identify full-face wrinkles. Furthermore, we develop a wrinkle detection evaluation metric that assesses outcomes based on curve, location, and gradient similarity. We collected and annotated a dataset for wrinkle detection consisting of 1021 images of Chinese faces. The dataset will be made publicly available to further promote wrinkle detection research. The research demonstrates a substantial enhancement in detecting subtle wrinkles through implementing the proposed method. Furthermore, the suggested evaluation procedure effectively considers the indistinct boundaries of wrinkles and is applicable to images with various resolutions.

1. Introduction

Wrinkles are structural changes due to a decrease in the elasticity of the facial skin. The factors contributing to this shift can be attributed to intrinsic elements associated with the normal aging process and extrinsic factors, including prolonged exposure to ultraviolet radiation [1] and tobacco smoking [2]. Hence, wrinkles hold significant relevance within dermatology [3–5] and medical aesthetics [6–8]. Nevertheless, automated detection of facial wrinkles is still a challenging task. The paradigm of wrinkle detection technologies and how to objectively and properly assess the quality of wrinkle detection are still open questions.

Wrinkles are considered line segments arranged randomly in space and modeled as different patterns in different studies. In studies of skin aging, wrinkles are considered a textural feature [9]. In the detection task, wrinkles are treated as an edge with fuzzy boundaries [10,11]. Traditionally, in prior studies, the conventional method involves utilizing first- or second-order derivatives for fundamental edge detection, such as Gabor filters [12] or Hessian filters [10,13,14]. Due to these filters cannot perceive high-level semantic information in the image, they will detect all edges without discrimination. This causes other edges in the image to be mistaken for wrinkles. Thus, these methods are limited to particular facial areas, such as the forehead. Deep learning has gained significant popularity in recent years for various computer vision applications because of its capacity to extract complex semantic information from images at several levels automatically [15–20]. Various semantic segmentation models, including Unet [21], Unet++ [22–24], and GCN [25,26] techniques, are employed to identify wrinkles. While these methods do provide flexibility in application scenarios, the recognition process may overlook some insignificant wrinkles due to wrinkles' narrow morphology.

Therefore, this paper presents a novel method for wrinkle detection. The proposed method, similar to recent studies, employs a neural network comprised of an encoder and a decoder. However, unlike other methods that commonly use level-by-level upsampling in semantic segmentation tasks, the decoder in the proposed method utilizes multilevel fusion, which is widely used in edge detection tasks [27–29]. This allows the decoder to preserve multilevel semantic information during decoding better, thus providing a better perception of insignificant wrinkles. Additionally, the paper proposes a distance-based loss function based on the representation of wrinkles in an image. Previous studies have shown that wrinkles exhibit a ridge-valley-ridge pattern [10], where valleys are darker in the image and ridges are brighter and closer to the color of the skin, resulting in

* Corresponding author.
E-mail addresses: zj.liu@sjtu.edu.cn (Z. Liu), qiquanpicb@gmail.com (Q. Qi), wangsijia@picb.ac.cn (S. Wang), zhaiguangtao@sjtu.edu.cn (G. Zhai).
[1] Those authors contribute equally to this work as co-first author.

https://doi.org/10.1016/j.compbiomed.2024.108431
Received 29 February 2024; Received in revised form 19 March 2024; Accepted 7 April 2024
Available online 9 April 2024

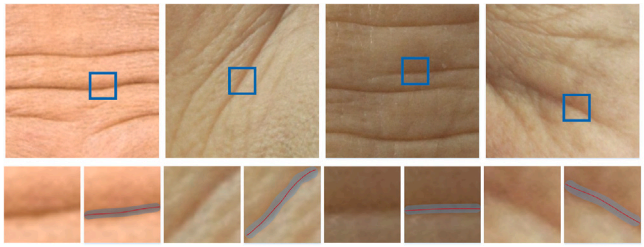

Fig. 1. Wrinkle patterns under different shooting conditions. The red line in the third column of the figure indicates the skeleton line, and the thick gray line indicates the uncertainty region.

blurred borders. The dice loss function, commonly applied in semantic segmentation tasks, is often criticized for its limited capacity to capture edge details [30]. This deficiency is compounded by the presence of blurred borders, which further exacerbates the problem. Therefore, the proposed method generates a distance map as a mask based on the dice loss function centered on the skeleton lines (valleys) to compute a weighted loss based on the distance to enhance the perception of edges.

In addition to the detection method, the evaluation of detection performance is also an issue that needs to be investigated. In this study, we consider the fuzzy boundary of wrinkles as a representation of uncertainty [31]. As shown in Fig. 1, the skeleton line to be detected can be considered as a certainty region. The region between the skeleton line and facial skin, the fuzzy boundary, can be regarded as an uncertainty region. In a previous study, Batool et al. evaluated wrinkles using detection rate, false alarm rate, and miss rate [32]. This method sets up a dilation region based on the certainty region and assumes that all predictions within the dilation region are correct, which is not an entirely rigorous approach. The Jaccard similarity index (JSI) [33] is currently the most popular tool for evaluating wrinkles. Similar to the Dice score [34] used for semantic segmentation, the JSI assessment method dilates the skeleton lines of the wrinkles before calculating the overlap ratio [13,35]. The JSI considers the uncertainty region to some extent but still has flaws. First, the weighting of the certainty region should differ from that of the uncertainty region in the assessment. Dilating the skeleton line is equivalent to changing the uncertainty region into a certainty region. Second, the scale of dilation could not be determined. The number of pixels occupied by wrinkles is different for images with different resolutions. Therefore, using the same dilation scale for images of different resolutions is unreasonable.

Based on the above observations, this study proposes an assessment method that utilizes skeleton lines. The method is based on the concept introduced by JSI, in which a dilation region is established. The prediction performance is assessed by calculating the similarity between the annotated and predicted skeleton lines within the region. The similarity metrics used are curve similarity, location similarity, and gradient similarity. The proposed method assigns a weight to the predicted skeleton line within the uncertainty region rather than classifying it as either correct or incorrect. Additionally, the method determines the dilation scale based on images of various resolutions.

The main contribution of this work is summarized as follows:

1. A network and a loss function have been proposed for the purpose of wrinkle detection, with a focus on enhancing the detection of insignificant wrinkles.
2. A novel evaluation method for wrinkle detection has been proposed. This approach is capable of discerning between certainty and uncertainty regions, as well as offering distinct schemes for images with varying resolutions.
3. We collected and annotated the first publicly available wrinkle detection dataset.

The remainder of this paper is organized as follows. In Section 2, we introduce and analyze existing wrinkle detection methods and commonly used evaluation metrics. Section 3 describes the dataset used in this study. The proposed detection methods and assessment metrics are described in Section 4 and Section 5, respectively. In Section 6, experiments are organized to demonstrate the performance of the detection methods and to analyze the rationality of the proposed assessment metrics. Finally, the conclusions are presented in Section 7.

2. Related works

2.1. Wrinkle detection method

In the initial stages of exploration, conventional edge detection methods, such as Gabor filters [6,12], Markov point processes [32,36,37], and Hessian Hybrid filters (HHF) [10,14,38], were utilized to identify wrinkles. Ng et al. applied the Frangi filter to the gradient map in the y-direction and subsequently set a threshold to extract wrinkles [13]. Similarly, Yap et al. modified HHF and devised a wrinkle annotator [14]. However, the selection of parameters poses a challenge for traditional methods, such as the choice of Gaussian kernel scales in HHF techniques. In a study by Yap et al. two Gaussian kernel scales were chosen for two datasets with varying resolutions. On the other hand, the absence of semantic information often results in traditional methods for detecting all edges in an image without discrimination. Consequently, early research efforts have mainly concentrated on specific regions of the face, such as the detection of transverse forehead lines.

Recently, the advancement of deep learning in computer vision has revolutionized the conventional approach to wrinkle detection. The acquisition of semantic information in images by neural networks significantly reduces the perception of non-wrinkled edges [39]. During the iterative training process, models can selectively filter the edge information they wish to detect. Therefore, several semantic segmentation models, such as Unet [21], Unet++ [22], and GCN [25], have been employed to identify wrinkles. Li et al. for instance, employed GCN to detect facial nasolabial folds using the entire face as input [26]. Likewise, Sabina et al. utilized a modified Unet network to accomplish the same objective [40]. Furthermore, convolutional neural networks have also demonstrated strong performance in several full-face wrinkle detection tasks [11,23].

Despite the enhanced flexibility in applications, existing methods are still inadequate for detecting insignificant wrinkles. This is because wrinkles, particularly insignificant wrinkles, have an elongated shape and are more challenging to recognize compared to typical semantic segmentation targets. In addition, the Dice loss function used in the segmentation task is deficient in the perception of edges. Hence, this study introduces a novel network architecture along with a newly formulated loss function for wrinkle detection.

2.2. Wrinkle assessment metrics

Evaluation criteria for wrinkle detection should consider smaller certainty regions while also considering uncertainty areas. Wrinkles are typically thin, making traditional area detection evaluation methods, such as IoU, unsuitable. Such methods are unable to assess predicted wrinkles within the region of uncertainty. In a previous study, Batool et al. first proposed an evaluation method for wrinkle detection [32]. They set a dilated region around the detected wrinkle, and the part of the annotated wrinkle that falls within the area is correctly detected. The ratio of all correctly predicted annotated wrinkles to the total annotated wrinkles was defined as the detection ratio (r_{detect}). Meanwhile, the $1 - r_{detect}$ was defined as the missing ratio (r_{miss}). Then they dilate the annotated wrinkle in the same way, and the part of the detected wrinkle that does not fall within this region is considered a false detection. The false alarm ratio (r_{false}) is the ratio of all false

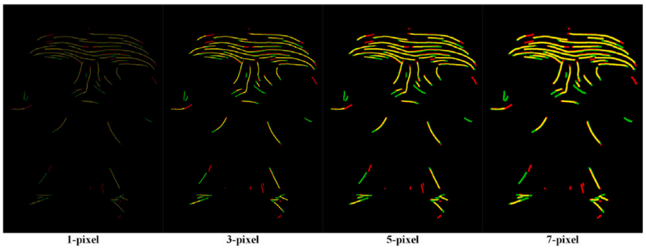

Fig. 2. Four degrees of dilation were performed on the wrinkled skeleton line. The red and green lines represent the annotated and predicted wrinkles, respectively. The yellow line indicates the overlapping part.

detections to the background of the whole image. These three metrics have also been used in some later studies [37,41]. Following this, Ng et al. adopted a comparable but more concise technique for their evaluation [13]. They executed simultaneous dilation of the predicted skeleton lines and the annotated skeleton lines at the same scale, and subsequently, they evaluated the overlap between these two regions using JSI. JSI is currently the most widely evaluated method for wrinkle detection tasks [35].

Although both assessment methods are widely used, their rationalization still has limitations. Firstly, the dilation region considers the uncertainty region of wrinkles, but it is not rigorous to consider the predicted skeleton lines within this region as correct. Instead, it is more appropriate to assign these skeleton lines a reasonable weight. Secondly, determining the dilation scale is another issue that needs to be addressed. As demonstrated in Fig. 2, the performance of different dilation scales varies, and increasing the dilation results in higher overlap and JSI. Therefore, finding a reasonable threshold is a challenge that needs to be overcome. This investigation introduces a novel evaluation approach that employs skeleton lines to assess wrinkles. The method evaluates the similarity between the predicted and annotated skeleton lines within a reasonable dilation range. The approach is inspired by [42,43] and aims to provide a more accurate assessment of wrinkles.

3. Wrinkle detection dataset

Research on wrinkle detection has been carried out for decades, but there is currently no dataset dedicated to wrinkle detection. Commonly used datasets include Bosphorus [44], FG-Net [45], FERET [46], MORPH [47], and PAL [48]. These datasets are intended to be used in other tasks such as age estimation and face recognition. Researchers have typically selected a small subset of these datasets in past studies and annotated them themselves. Currently, there are no publicly available wrinkle annotation datasets available.

We collect various high-resolution face images from different sources and annotate entire face regions. The informed consent was obtained from the volunteers. This research program was conducted with the approval of the Ethics Committee of Fudan University (Ethics Research Approval No. 85), Shanghai, China and followed the principles of the Declaration of Helsinki. In addition, to protect privacy, we covered the volunteers' eyes, noses and mouths. The covered area is determined by connecting landmarks point by point, and the detection of landmarks uses the API of Face++.

Totally 1021 images of frontal faces were collected in several cities across China, which were captured with different digital cameras in various stable indoor environments. The age distribution of the volunteers ranged from 25 to 80 years old. Resolution and lighting conditions are not uniform between images due to different shooting locations. Resolution ranges from 1000 × 700 to 2640 × 1700.

Before image annotation, we first invite dermatologists to grade wrinkles to screen out wrinkles that can significantly reflect skin quality for annotation. Afterward, we asked three coders to independently annotate the skeleton lines of wrinkles based on the grading criteria. The annotation standard is that each coder draws a single-pixel line that they think is the darkest color. All annotation work is done on the iPad. Dermatologists select the final annotation results from the annotation results of the three coders.

4. Wrinkle detection method

4.1. Network structure

In this section, we present a new approach for detecting full-face wrinkles. The architecture of our method is depicted in Fig. 3. It consists of an encoder–decoder structure, with the encoder comprising four stages that extract four levels of semantic features. Between each pair of stages, a downsampling operation is performed, which involves convolution with a kernel size of 3 and a stride size of 2. This reduces feature map size by half, and the network depth is increased accordingly. The depths of the network for the four stages are 32, 64, 160, and 256. The decoding process for each stage includes an attention structure and a convolution operation. Note that the attention structure used here is the same as in stages. Specifically, stages 2–4 are upsampled to their original size using a bilinear interpolation operation, and four outputs are concatenated to produce the final predicted mask through a convolution operation.

The encoder employs a consistent structure for every stage. The structure is inspired by [49] and comprises attention and feedforward modules in sequence. The attention module contains two convolutions with an attention structure in between. The attention structure comprises three branches, each utilizing a convolution with a large-scale kernel to extend the long-range receptive field. The feedforward module implements a bottleneck structure consisting of expansion followed by compression, with an expansion coefficient of 4. The repetitions for the four stages are 3, 3, 5, and 2, respectively. The output of the network is represented as a map that shows the distribution of probabilities. The skeleton lines are obtained by extracting them from a binary image generated with a threshold of 0.5, which is the outcome of the wrinkle detection process. The extraction algorithm for skeleton lines is borrowed from [50].

4.2. Loss function

In previous wrinkle detection studies, the dice loss function is the most commonly employed:

$$L_{dice} = 1 - \frac{2\sum_{i=1}^{N} g_i p_i}{\sum_{i=1}^{N} g_i + \sum_{i=1}^{N} p_i}, \tag{1}$$

where N denotes all pixels. g and p denote the ground truth and the predicted mask, respectively.

The dice loss function aims to maximize the overlap between the positive samples from two distributions. Nevertheless, it disregards the predicted wrinkles within the uncertainty region, as illustrated in Fig. 4. The green line, representing the predicted wrinkles, has minor intersections with the red line, meaning the annotated wrinkles. Predictions made within the uncertainty region are considered incorrect, resulting in a substantial loss value. This evaluation technique overlooks the fuzzy boundaries and is not reasonable enough.

This study proposes a distance-based loss function, followed by [30, 51]. Employing the annotated skeleton line as the central line, a distance map is generated by evaluating the radial distance from the surrounding pixels to the center, as depicted in Fig. 4. The pixel values in the map progressively decrease with increasing distance from the skeleton line. The formula for the loss function is presented below:

$$L_{dist} = -\frac{2\sum_{i=1}^{N} g_i^{dilate} p_i d_i}{\sum_{i=1}^{N} g_i^{dilate} + \sum_{i=1}^{N} p_i}, \tag{2}$$

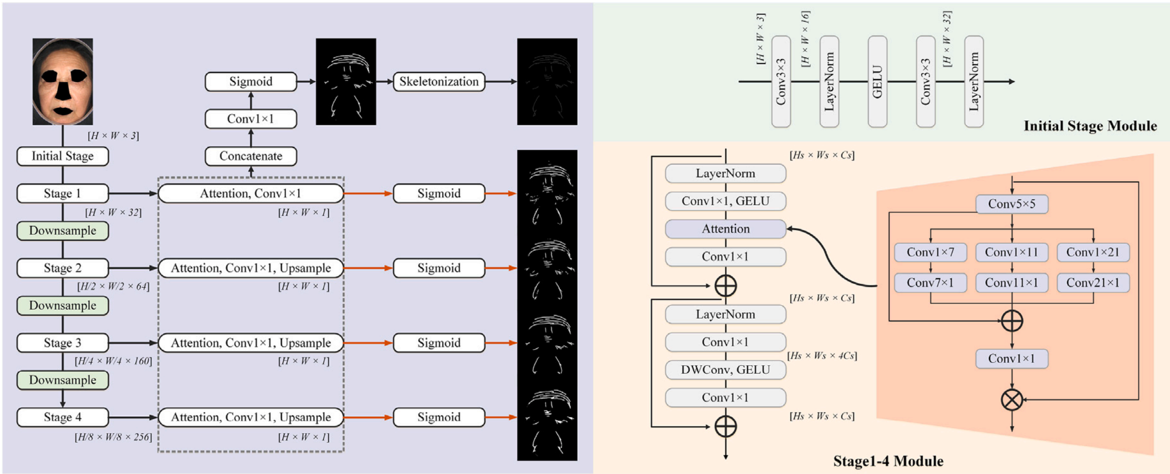

Fig. 3. A general overview of the proposed wrinkle detection method. The flow of this task is demonstrated in the left section, with the right upper part detailing the initiation stage's specific operations and the lower section outlining the subsequent four stages' specific operations.

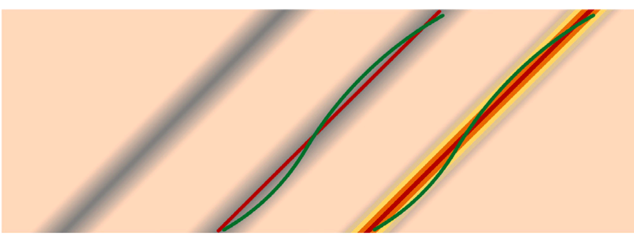

Fig. 4. The left line signifies a simulated wrinkle. The center red line represents annotated wrinkles, while the green line depicts predicted wrinkles. The right-hand side exhibits a gradual thick line, which serves as the distance map.

where g^{dilate} denotes the dilated ground truth, while d denotes the distance map. In this study, the radius of dilation is set to 3. Losses are calculated for each of the 4 stage maps as well as the fused map, and the overall loss is the aggregate of these 5 losses.

5. Skeleton line-based wrinkle assessment metric

The assessment metrics designed in this study aim to objectively evaluate the differences in wrinkle morphology, location, and edge gradients between annotations and predictions through the skeleton line. The overview is shown in Fig. 5.

First, we divide the annotated skeleton lines A^{sl} into single line segments according to the nodes at the intersections between wrinkles. Thereafter, these single line segments are subdivided sequentially such that the longest does not exceed MaxLength and those shorter than MinLength are discarded, as follows:

$$A^{sl} = \sum_{i \in N} A_i^{sl}, \tag{3}$$

where $A_i^{sl} \in [MinLength, MaxLength]$. This study set MinLength and MaxLength to 5 and 20, respectively.

After segmentation is complete, region dilation is performed on the A_i^{sl} to divide a tolerable error region for the predicted wrinkle. Since wrinkles have uncertainty areas, even if the predicted skeleton line is offset compared to the annotated skeleton line, it cannot be negated entirely. Of course, it is equally unreasonable to consider wrinkles in the dilated area to be correct predictions. Each pixel on A_i^{sl} is extended outward with a radius r as follows:

$$A_i^{sld} = dilate(A_i^{sl}, r) \tag{4}$$

This denote that the A_i^{sl} is expanded from 1 single pixel to a width of $2r + 1$ pixels. All predicted skeletal line segments P_i^{sl} within the A_i^{sld} are queried and eligible to participate in the evaluation. Next, we will go over the specific evaluation methods for each segment.

5.1. Curve similarity

Curve similarity is used to evaluate the morphology of A_i^{sl} and P_i^{sl}, which can also be considered wrinkles' direction. Specifically, A_i^{sl} and P_i^{sl} can be regarded as a scatter arrangement of pixels within A_i^{sld}, and we use two cubic curves to fit A_i^{sl} and P_i^{sl}, respectively.

$$\begin{cases} A_i^{slc} = \mathcal{F}^3(a1, b1, c1, d1) \\ P_i^{slc} = \mathcal{F}^3(a2, b2, c2, d2), \end{cases} \tag{5}$$

where a, b, c, and d represent the fitting parameters. In the cubic equation $\mathcal{F}^3$, a, b, and c are parameters that determine the direction of the curve, so we measure the difference between the two curves by the vector $\vec{V}_a = [a_1, b_1, c_1]$ and the vector $\vec{V}_p = [a_2, b_2, c_2]$.

$$cs_i = |\frac{\vec{V}_a \cdot \vec{V}_p}{|\vec{V}_a| \cdot |\vec{V}_p|}|, \tag{6}$$

where $cs_i \in [0, 1]$ characterizes the curve similarity, and cs_i equals 1 when the two curves overlap. Fig. 6 shows the fit of the two wrinkles.

5.2. Location similarity

The location similarity assesses the positional relationship between the two skeletal lines A_i^{sl} and P_i^{sl}. First, similar to curve similarity, we also use a function to fit the pixel points, as shown in Fig. 7. The difference is that the location similarity uses a linear function because fitting pixel points using curves is more about depicting the skeleton line orientation, which is difficult to reflect the overall position. Hence, a linear function is preferable. Following fitting, the area enclosed by the two fitted lines is used to measure the distance between A_i^{sl} and P_i^{sl}. We can describe the location similarity as follows.

$$ls_i = 1 - \frac{2 \times AP_i^{sl-area}}{sum(A_i^{sld})}, \tag{7}$$

where $AP_i^{sl-area}$ represents the area enclosed by the two fitted lines. Note that there is a possibility that A_i^{sl} and P_i^{sl} will cross to enclose two areas. In this case, $AP_i^{sl-area}$ takes the absolute value of the difference between the two areas. Because the area resulting from the crossover is not positively correlated with the distance, the slope also plays a role in the size of the area. For example, when the midpoints of two lines intersect in an X-shape, the size of the area is only related to the slope and not the location. Accordingly, the ls_i of these two lines is equal to 1.

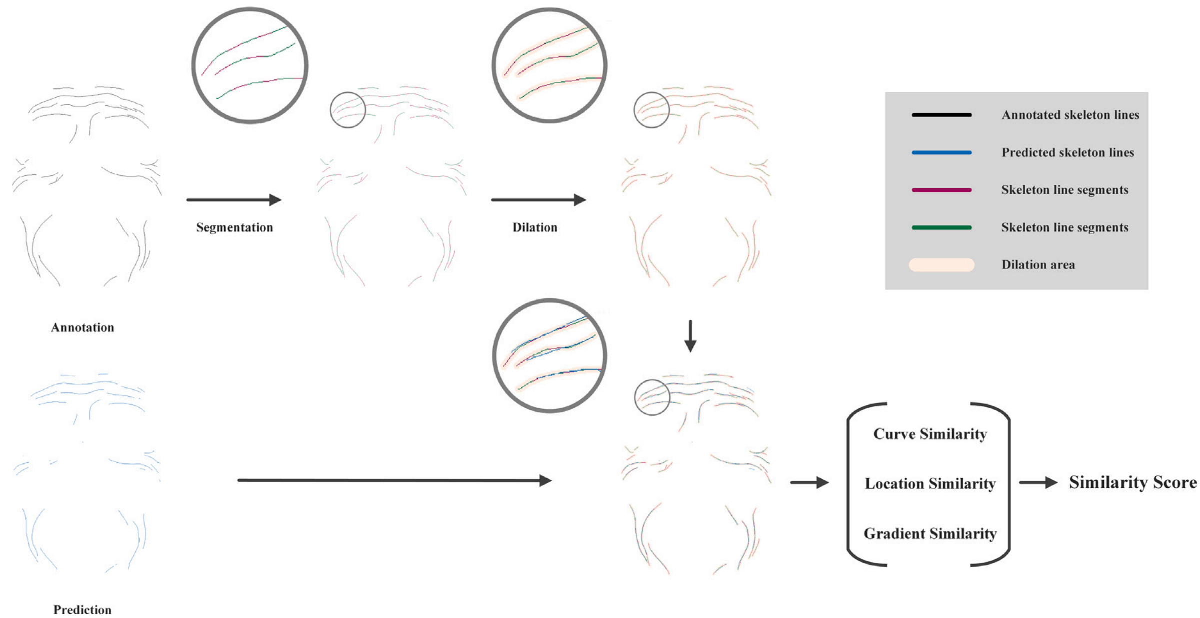

Fig. 5. A general overview of the proposed wrinkle detection assessment method. Following the segmentation and dilation of annotated wrinkles, the predicted wrinkle segments located within the dilated region were assessed.

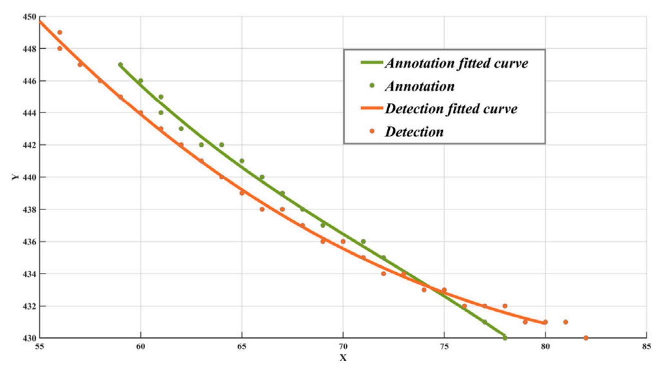

Fig. 6. The green dots represent pixels that annotate wrinkles, and the red dots represent pixels that predict wrinkles. The two curves are obtained via pixel fitting.

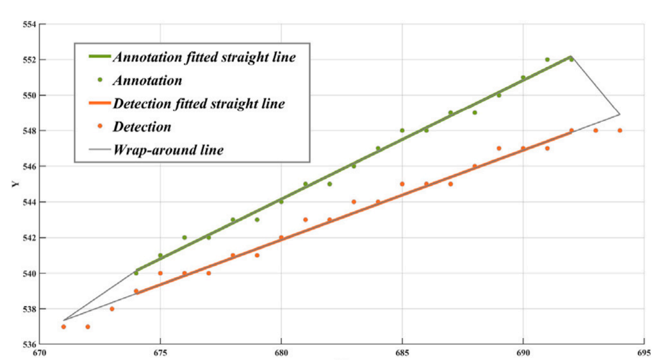

Fig. 7. The green and red dots indicate annotated and predicted pixels, respectively. Two lines are obtained by fitting the pixels using two linear functions.

In addition, $sum(A_i^{sld})$ denotes the total number of pixels of A_i^{sld}, which can also be interpreted as the total area of A_i^{sld}. Since the maximum possible value of the area enclosed by A_i^{sl} and P_i^{sl} is half of $sum(A_i^{sld})$, the numerator is set to be two times of $AP_i^{sl-area}$.

5.3. Gradient similarity

Gradient similarity is used to measure the difference when P_i^{sl} is in a certainty region, uncertainty region, and outside the wrinkles. This is the most significant difference between our and previous evaluation methods, which confounded the expanded certainty area with the uncertainty area. Given an image denoted as $\boldsymbol{I} \in \mathbb{R}^{h\times w}$, the certainty region of wrinkles presents a lower gray value on $\boldsymbol{I}$, while the skin outside the wrinkles has a higher gray value. The uncertainty region is in the middle. Based on this observation, we define the gradient similarity as:

$$gs_i = 1 - 2\times\frac{exp(t)-1}{exp(t)+1}, \tag{8}$$

where t denotes the absolute value of the difference between the average grayscale values of A_i^{sl} and P_i^{sl} at corresponding positions on $\boldsymbol{I}$. Specifically, we query the corresponding pixels of A_i^{sl} and P_i^{sl} on $\boldsymbol{I}$, calculate the average gray value of these two groups of pixels separately, and then make the difference.

$$t = Abs(sum(\boldsymbol{I}(A_i^{sl}))/n - sum(\boldsymbol{I}(P_i^{sl}))/n) \tag{9}$$

Since there is a significant grayscale difference between the uncertainty region and beyond the wrinkles, gs_i can reflect whether the position of P_i^{sl} is in the uncertainty region.

Based on the above, we give the wrinkle similarity scores as follows:

$$wss_i = w_1\cdot cs_i + w_2\cdot ls_i + w_3\cdot gs_i, \tag{10}$$

where w_1, w_2, and w_3 denote the weights of the three similarity metrics, respectively, and $w_1+w_2+w_3 = 1$. In this study, we set all three weights to 1/3. Since all 3 similarity scores are between 0 and 1, $wws_i \in [0,1]$.

After calculating the segmentation score, we give the global score of the whole face, which uses a pixel-level strategy. Specifically, wss_i computes the score of P_i^{sl}. We believe all the pixels contained in P_i^{sl} to be scored as wss_i. From this, we can describe the total score $N\cdot wss_i$ of all pixels in the detected line segments, where N denotes the total number of pixels in P_i^{sl}. The final predicted scores of the whole face are represented as follows:

$$WSS = \frac{\sum_{P_i^{sl}\in P^{sl}} N_i\cdot wss_i}{\sum_{A_i^{sl}\in A^{sl}} N_i}, \tag{11}$$

Fig. 8. Evaluation of the Proposed Method's Detection Performance: The original image is presented in the first column, followed by the four stages of the map in columns 2 to 5. The fused map is depicted in column 6, and the skeleton line extracted from the fused map is shown in column 7. The final column is the annotations.

Table 1
Objective evaluation results of the proposed and compared methods. Bold is the best performance.

	Dilation 3					Dilation 5					Dilation 7				
	HHF	UNet	UNet++	GCN	Proposed	HHF	UNet	UNet++	GCN	Proposed	HHF	UNet	UNet++	GCN	Proposed
ACC	.888	**.993**	**.993**	.992	.992	.851	.988	.987	.987	**.989**	.824	.984	.983	.982	**.988**
F-measure	.036	.178	.109	.068	**.436**	.053	.226	.143	.095	**.561**	.064	.251	.163	.112	**.615**
Dice	.036	.173	.084	.065	**.434**	.053	.219	.110	.091	**.560**	.064	.244	.126	.107	**.613**
JSI	.019	.099	.046	.035	**.283**	.028	.130	.062	.051	**.395**	.033	.147	.072	.060	**.449**
WDice	.063	.166	.079	.057	**.523**	.111	.238	.120	.099	**.712**	.145	.267	.137	.121	**.779**
CDice	.066	.172	.081	.059	**.548**	.118	.245	.121	.103	**.740**	.154	.271	.137	.124	**.802**
LDice	.055	.137	.064	.046	**.439**	.094	.202	.099	.080	**.618**	.123	.235	.119	.101	**.696**
GDice	.068	.189	.093	.067	**.574**	.121	.266	.137	.114	**.768**	.157	.294	.155	.137	**.829**

where WSS denotes the pixel-level mean score of the predicted skeleton lines in the extended region A^{sld}. In other words, WSS is the composite score that can consider both the certainty and uncertainty regions.

5.4. Wrinkle assessment metric

Although WSS can evaluate two close wrinkles, A^{sl} and P^{sl}, it cannot describe over-detection. Because some of the predicted wrinkles are not inside A^{sld}. To solve this problem, we borrow the traditional region evaluation method. Given that the total pixels of the image is $N(\boldsymbol{I})$, the pixels occupied by A^{sld} are $N(A^{sld})$. The traditional evaluation metrics true positive (TP), false negative (FN), true negative (TN), and false positive (FP) can be re-expressed in the following form.

$$\begin{aligned} WTP &= WSS \cdot N(A^{sld}) \\ WFN &= (1 - WSS) \cdot N(A^{sld}) \\ WTN &= N((1 - A^{sld}) \cap (1 - P^{sl})) \\ WFP &= N((1 - A^{sld}) \cap P^{sl}), \end{aligned} \tag{12}$$

where $1-A^{sld}$ denotes the background region outside of A^{sld}, and $1-P^{sl}$ is the same. From this, we depict the final evaluation metrics as follows:

$$WDice = \frac{2 \times WTP}{2 \times WTP + WFP + WFN} \tag{13}$$

6. Experience and analysis

6.1. Experimental details

The experiments were conducted on the dataset we provided. In the wrinkle detection algorithm, we randomly divide all 1021 images into a training set and a test set according to the ratio of 7 to 3, i.e., 714 images in the training set and 307 in the test set. To resist overfitting, we use data augmentation strategies during training, including random rotation $(-\pi, \pi)$ and random flip (horizontal and vertical). Meanwhile, due to the increase in memory cost of the graphics card caused by the large original resolution, we used a resize operation to downsample the images to 512 × 768. Before training, the images were normalized by subtracting the mean and dividing by the standard deviation.

In this experiment, all models are trained for 300 epochs. The learning rate is set to 1e−4 and decays by half every 100 epochs. Adam [52] is used as the optimizer. The batch size is set to 2. The model framework is designed and trained using Pytorch [53]. An NVIDIA RTX 4090 graphics card was used to perform the arithmetic required for the model development.

6.2. Experiment results

In this section, we first evaluate and analyze the wrinkle detection results. Subsequently, we experimentally illustrate the rationality of the proposed assessment method.

6.2.1. Wrinkle detection results

The results of the proposed method's segmentation performance are depicted in Fig. 8. It is evident that the shallow network stage incorrectly detects some non-wrinkled edges, such as hairs. As the network deepens, the high-level semantic features improves the recognition of wrinkles with greater accuracy. The third and fourth stages exhibit sufficient quality to disregard non-wrinkled edges. As the semantic features progressively enhance, the model demonstrates a decreased responsiveness to low-level semantic information, such as edges. Consequently, the fourth stage presents with indistinct boundaries. The multi-stage fusion approach resolves this issue by preserving both rich edge information and high-level semantic information simultaneously.

We evaluated the proposed method by comparing it with the current state-of-the-art wrinkle detection techniques. These techniques encompass the HHF, GCN, UNet, and UNet++ methods. Apart from HHF, which employs traditional techniques, the other three methods apply

Fig. 9. Comparison with previous research methods. The first column is the original image. The following five columns show the detection performance of HHF, UNet, UNet++, GCN, and the proposed method, respectively. The last column is the annotation.

the same training parameter settings as the proposed method. The performances of the methods are shown in Fig. 9. HHF indiscriminately identifies all edges in the images, leading to a significant amount of irrelevant information in its detection results. In contrast, the other three methods perform adequately in detecting more apparent wrinkles, such as forehead wrinkles. However, they fail to detect full-face wrinkles comprehensively; thus, they may fail to detect less obvious wrinkles, such as eye wrinkles.

To objectively evaluate the detection results of wrinkles, we use both the traditional evaluation method and the evaluation method proposed in this paper. The metrics used in the evaluation include the following:

$$\begin{aligned}
Recall &= \frac{TP}{TP+FN} \\
Precision &= \frac{TP}{TP+FP} \\
Accuracy &= \frac{TP+TN}{TP+TN+FP+FN} \\
F-measure &= \frac{2Recall \times Precision}{Recall + Precision} \\
Dice &= \frac{2\times TP}{FP+FN+2\times TP} \\
JSI &= \frac{TP}{TP+FN+FP},
\end{aligned} \tag{14}$$

where TP, FP, TN, and FN are true positive, false positive, true negative, and false negative, respectively. In addition, we take the proposed curve similarity, location similarity, and gradient similarity separately to evaluate the results, denoted as $CDice$, $LDice$, and $GDice$. by setting w_1, w_2, and w_3, respectively, to 1. Since wrinkles have uncertainty regions, we extended the certainty region, i.e., the skeleton line, by three scales in the evaluation. The skeleton line is expanded outward by 1, 2, and 3 pixels, and the expanded wrinkle coarseness is 3-pixel, 5-pixel, and 7-pixel, respectively.

Table 1 showcases the results, which indicate that the three semantic segmentation algorithms exhibit superior performance compared to traditional methods. Among these algorithms, UNet outperforms UNet++ and GCN. The proposed method demonstrates the best performance.

6.2.2. Assessment methods

As described in the previous sections, evaluating the dilated wrinkle skeleton line by JSI would, on the one hand, treat the dilated region as a certainty region and, on the other hand, would fail to determine the extent of dilation. As shown in Fig. 10, we demonstrate the changes in various evaluation metrics during the dilation of wrinkles up to 15 pixels. All metrics are evaluated on the predicted results of the proposed detection method. In addition, considering that multiple resolutions are covered in the dataset and span a wide range, we divided the data into 3 groups, which were evaluated separately: low-resolution, medium-resolution, and high-resolution. This was done because there is a difference in the number of pixels occupied by wrinkles on images with different resolutions. The higher the image's resolution, the more pixels the wrinkles occupy and the larger the area to be dilated. In this experiment, images with resolutions less than or equal to 1000 × 700 in the data are defined as low resolution. Images greater than 1000 × 700

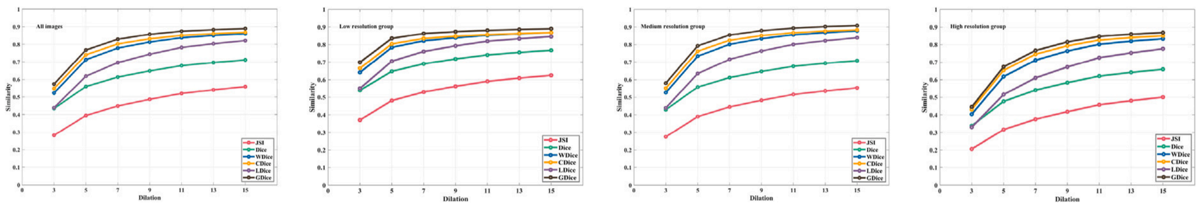

Fig. 10. Growth trend of each assessment metric as dilation increases.

Table 2
The data were divided into three resolution groups—the growth of each evaluation metric as the dilation increases. In the WDice vs. Dice comparison, red font indicates faster growth, whereas green font indicates slower growth.

	Low-resolution						Medium-resolution						High-resolution					
	3-5	5-7	7-9	9-11	11-13	13-15	3-5	5-7	7-9	9-11	11-13	13-15	3-5	5-7	7-9	9-11	11-13	13-15
Dice	.108	.043	.027	.023	.015	.012	.129	.054	.035	.029	.017	.015	.138	.064	.043	.037	.021	.018
JSI	.110	.048	.033	.028	.019	.016	.114	.055	.038	.033	.020	.018	.111	.059	.042	.039	.023	.020
WDice	.143	.037	.018	.014	.009	.007	.208	.068	.032	.022	.013	.010	.214	.093	.052	.038	.018	.014

and less than or equal to 2000 × 1400 are medium resolution. Images larger than 2000 × 1400 are defined as high resolution.

As can be seen from Fig. 10, both JSI and Dice exhibit a nearly smooth upward trend regardless of the resolution. This phenomenon arises because JSI and Dice treat the dilation region as a certainty region. This also leads to the inability to determine the extent of the dilation. On the other hand, the proposed evaluation method has a significant rising saturation phenomenon on all three groups. And there is a significant difference in the location of saturation, i.e., the group with higher resolution will saturate at the location with greater dilation.

The saturation phenomenon arises due to the limitations of the three similarity assessment metrics, as they do not grow indefinitely with dilation. The metrics for CDice and GDice do not grow further because an increase in dilation does not modify the curvature characteristics or the positional relationship between the predicted and annotated wrinkles. For GDice, once the dilation reaches the non-wrinkled area, the wrinkles mistakenly identified as existing in the non-wrinkled region receive a minuscule gradient similarity score, which leads to the cessation of further growth in GDice.

According to the above, it is possible to determine how many pixels need to be dilated by determining the location of the saturation point. Table 2 shows more precisely the growth of the metrics as the dilation increases. Since Dice treats the dilation region as a certainty region like JSI, and since our proposed metric references the calculation of Dice (equivalent to Dice that is weighted for 3 regions), we take the amount of growth of Dice with dilation as a reference. In the low-resolution group, WDice grows faster than Dice when the dilation width is less than 5, while WDice appears to increase saturated when the dilation width is greater than 5. Therefore, it is reasonable to conclude that 5 pixels is the best dilation width for low-resolution images. Similarly, the medium-resolution group is 7 pixels, while the high-resolution group is 11 pixels.

7. Conclusion

This study introduces a pioneering technique for detecting wrinkles and a corresponding assessment metric. The proposed method utilizes a multi-stage fused edge detection network and a distance-based loss function, which offers a heightened perception of insignificant wrinkles and, thus, a more comprehensive detection of facial wrinkles. The suggested assessment method employs curve, location, and gradient similarity to evaluate the predicted wrinkles. This method distinguishes between certainty and uncertainty regions, making the assessment more rational. Furthermore, the method provides a reasonable dilation range based on the location of the saturation point, which is applicable to images with different resolutions.

The advancement of imaging equipment presents the potential for increasingly higher-resolution photographs. Wrinkles will increasingly transition from an edge pattern to a regional pattern, representing a promising future research direction.

CRediT authorship contribution statement

Zijia Liu: Writing – original draft, Visualization, Methodology. **Quan Qi:** Validation, Formal analysis, Data curation. **Sijia Wang:** Project administration, Funding acquisition, Conceptualization. **Guangtao Zhai:** Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 92249302, 62225112, 61831015, 32325013), and the CAS Project for Young Scientists in Basic Research, China (No. YSBR-077), and the National Key R and D Program of China (No. 2021YFE0206700), and the Shanghai Municipal Science and Technology Major Project, China (No. 2021SHZDZX0102). and the "Strategic Priority Research Program" of the Chinese Academy of Sciences (No. XDB38020400 to SW), and the Shanghai Science and Technology-Commission Excellent Academic Leaders Program, China (22XD1424700), and the Human Phenome Data Center of Fudan University, China. Thanks to Yinqi Zhang for providing suggestions for this paper.

References

[1] J.H. Chung, J.Y. Seo, M.K. Lee, H.C. Eun, J.H. Lee, S. Kang, G.J. Fisher, J.J. Voorhees, Ultraviolet modulation of human macrophage metalloelastase in human skin in vivo, J. Invest. Dermatol. 119 (2) (2002) 507–512.
[2] D.P. Kadunce, R. Burr, R. Gress, R. Kanner, J.L. Lyon, J.J. Zone, Cigarette smoking: risk factor for premature facial wrinkling, Ann. Intern. Med. 114 (10) (1991) 840–844.
[3] D.J. Day, C.M. Littler, R.W. Swift, S. Gottlieb, The wrinkle severity rating scale: a validation study, Am. J. Clin. Dermatol. 5 (2004) 49–52.
[4] J. Contet-Audonneau, C. Jeanmaire, G. Pauly, A histological study of human wrinkle structures: comparison between sun-exposed areas of the face, with or without wrinkles, and sun-protected areas, Br. J. Dermatol. 140 (6) (1999) 1038–1047.

[5] Y. Chen, W. Liu, L. Zhang, M. Yan, Y. Zeng, Hybrid facial image feature extraction and recognition for non-invasive chronic fatigue syndrome diagnosis, Comput. Biol. Med. 64 (2015) 30–39.

[6] N. Batool, R. Chellappa, Detection and inpainting of facial wrinkles using texture orientation fields and Markov random field modeling, IEEE Trans. Image Process. 23 (9) (2014) 3773–3788.

[7] N. Magnenat-Thalmann, P. Kalra, J.L. Leveque, R. Bazin, D. Batisse, B. Querleux, A computational skin model: fold and wrinkle formation, IEEE Trans. Inform. Technol. Biomed. 6 (4) (2002) 317–323.

[8] X. Ling, M.I. Osotsi, W. Zhang, Y. Wu, Q. Jin, D. Zhang, Bioinspired materials: From distinct dimensional architecture to thermal regulation properties, J. Bionic Eng. 20 (3) (2023) 873–899.

[9] Y. Fu, G. Guo, T.S. Huang, Age synthesis and estimation via faces: A survey, IEEE Trans. Pattern Anal. Mach. Intell. 32 (11) (2010) 1955–1976.

[10] C.-C. Ng, M.H. Yap, N. Costen, B. Li, Wrinkle detection using hessian line tracking, IEEE Access 3 (2015) 1079–1088.

[11] S. Kim, H. Yoon, J. Lee, S. Yoo, Facial wrinkle segmentation using weighted deep supervision and semi-automatic labeling, Artif. Intell. Med. 145 (2023) 102679.

[12] G.O. Cula, P. Bargo, A. Nkengne, N. Kollias, Assessing facial wrinkles: automatic detection and quantification, Skin Res. Technol. 19 (1) (2013) e243–e251.

[13] C.-C. Ng, M.H. Yap, N. Costen, B. Li, Automatic wrinkle detection using hybrid hessian filter, in: Asian Conference on Computer Vision, Springer, 2014, pp. 609–622.

[14] M.H. Yap, J. Alarifi, C.-C. Ng, N. Batool, K. Walker, Automated facial wrinkles annotator, in: European Conference on Computer Vision, Springer, 2018, pp. 676–680.

[15] W. Gómez-Flores, W.C. de Albuquerque Pereira, A comparative study of pre-trained convolutional neural networks for semantic segmentation of breast tumors in ultrasound, Comput. Biol. Med. 126 (2020) 104036.

[16] H. Jiang, Z. Diao, T. Shi, Y. Zhou, F. Wang, W. Hu, X. Zhu, S. Luo, G. Tong, Y.-D. Yao, A review of deep learning-based multiple-lesion recognition from medical images: classification, detection and segmentation, Comput. Biol. Med. (2023) 106726.

[17] T. Lan, Z. Li, J. Chen, FusionSegNet: Fusing global foot features and local wound features to diagnose diabetic foot, Comput. Biol. Med. 152 (2023) 106456.

[18] S. Yasin, A. Othmani, I. Raza, S.A. Hussain, Machine learning based approaches for clinical and non-clinical depression recognition and depression relapse prediction using audiovisual and EEG modalities: A comprehensive review, Comput. Biol. Med. 159 (2023) 106741.

[19] S. Ahmed, T.-M. Groenli, A. Lakhan, Y. Chen, G. Liang, A reinforcement federated learning based strategy for urinary disease dataset processing, Comput. Biol. Med. 163 (2023) 107210.

[20] Z. Mu, L. Feng, Y. Shang, Q. Liu, L. Hu, F. Zhou, X. Fu, Algorithm analysis of face recognition robot based on deep learning, Int. J. Pattern Recognit. Artif. Intell. 37 (05) (2023) 2356004.

[21] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III. Vol. 18, Springer, 2015, pp. 234–241.

[22] Z. Zhou, M.M. Rahman Siddiquee, N. Tajbakhsh, J. Liang, Unet++: A nested u-net architecture for medical image segmentation, in: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings. Vol. 4, Springer, 2018, pp. 3–11.

[23] Q. Zheng, A. Purwar, H. Zhao, G.L. Lim, L. Li, D. Behera, Q. Wang, M. Tan, R. Cai, J. Werner, et al., Automatic facial skin feature detection for everyone, 2022, arXiv preprint arXiv:2203.16056.

[24] M. Sanchez, G. Triginer, C. Ballester, L. Raad, E. Ramon, Photorealistic facial wrinkles removal, in: Asian Conference on Computer Vision, Springer, 2022, pp. 117–133.

[25] C. Peng, X. Zhang, G. Yu, G. Luo, J. Sun, Large kernel matters–improve semantic segmentation by global convolutional network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4353–4361.

[26] H. Li, K. Wu, H. Cheng, C. Gu, X. Guan, Nasolabial folds extraction based on neural network for the quantitative analysis of facial paralysis, in: 2018 2nd International Conference on Imaging, Signal Processing and Communication, ICISPC, IEEE, 2018, pp. 54–58.

[27] S. Xie, Z. Tu, Holistically-nested edge detection, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 1395–1403.

[28] Z. Su, W. Liu, Z. Yu, D. Hu, Q. Liao, Q. Tian, M. Pietikäinen, L. Liu, Pixel difference networks for efficient edge detection, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 5117–5127.

[29] X. Qin, Z. Zhang, C. Huang, M. Dehghan, O.R. Zaiane, M. Jagersand, U2-Net: Going deeper with nested U-structure for salient object detection, Pattern Recogn. 106 (2020) 107404.

[30] F. Caliva, C. Iriondo, A.M. Martinez, S. Majumdar, V. Pedoia, Distance map loss penalty term for semantic segmentation, 2019, arXiv preprint arXiv:1908.03679.

[31] M. Abdar, F. Pourpanah, S. Hussain, D. Rezazadegan, L. Liu, M. Ghavamzadeh, P. Fieguth, X. Cao, A. Khosravi, U.R. Acharya, et al., A review of uncertainty quantification in deep learning: Techniques, applications and challenges, Inf. Fusion 76 (2021) 243–297.

[32] N. Batool, R. Chellappa, Modeling and detection of wrinkles in aging human faces using marked point processes, in: European Conference on Computer Vision, Springer, 2012, pp. 178–188.

[33] P. Jaccard, Distribution de la flore alpine dans le bassin des Dranses et dans quelques régions voisines, Bull. Soc. Vaudoise Sci. Nat. 37 (1901) 241–272.

[34] L.R. Dice, Measures of the amount of ecologic association between species, Ecology 26 (3) (1945) 297–302.

[35] M.H. Yap, N. Batool, C.-C. Ng, M. Rogers, K. Walker, A survey on facial wrinkles detection and inpainting: Datasets, methods, and challenges, IEEE Trans. Emerg. Top. Comput. Intell. (2021).

[36] N. Batool, R. Chellappa, A Markov point process model for wrinkles in human faces, in: 2012 19th IEEE International Conference on Image Processing, IEEE, 2012, pp. 1809–1812.

[37] S.-G. Jeong, Y. Tarabalka, J. Zerubia, Marked point process model for facial wrinkle detection, in: 2014 IEEE International Conference on Image Processing, ICIP, IEEE, 2014, pp. 1391–1394.

[38] R.M. Elbashir, M. Hoon Yap, Evaluation of automatic facial wrinkle detection algorithms, J. Imag. 6 (4) (2020) 17.

[39] H. Musafer, E. Tokgöz, A facial wrinkle detection by using deep learning with an efficient optimizer, in: Cosmetic and Reconstructive Facial Plastic Surgery: A Review of Medical and Biomedical Engineering and Science Concepts, Springer, 2023, pp. 381–388.

[40] U. Sabina, T.K. Whangbo, Nasolabial wrinkle segmentation based on nested convolutional neural network, in: 2021 International Conference on Information and Communication Technology Convergence, ICTC, IEEE, 2021, pp. 483–485.

[41] N. Batool, R. Chellappa, Fast detection of facial wrinkles based on Gabor features using image morphology and geometric constraints, Pattern Recognit. 48 (3) (2015) 642–658.

[42] Z. Yan, X. Yang, K.-T. Cheng, A skeletal similarity metric for quality evaluation of retinal vessel segmentation, IEEE Trans. Med. Imaging 37 (4) (2017) 1045–1057.

[43] M.E. Gegúndez-Arias, A. Aquino, J.M. Bravo, D. Marín, A function for quality evaluation of retinal vessel segmentations, IEEE Trans. Med. Imaging 31 (2) (2011) 231–239.

[44] A. Savran, B. Sankur, M.T. Bilge, Comparative evaluation of 3D vs. 2D modality for automatic detection of facial action units, Pattern Recogn. 45 (2) (2012) 767–782.

[45] Y. Fu, T.M. Hospedales, T. Xiang, J. Xiong, S. Gong, Y. Wang, Y. Yao, Robust subjective visual property prediction from crowdsourced pairwise labels, in: IEEE TPAMI, 2016.

[46] P.J. Phillips, H. Moon, S.A. Rizvi, P.J. Rauss, The FERET evaluation methodology for face-recognition algorithms, IEEE Trans. Pattern Anal. Mach. Intell. 22 (10) (2000) 1090–1104.

[47] K. Ricanek, T. Tesafaye, Morph: A longitudinal image database of normal adult age-progression, in: 7th International Conference on Automatic Face and Gesture Recognition, FGR06, IEEE, 2006, pp. 341–345.

[48] M. Minear, D.C. Park, A lifespan database of adult facial stimuli, Behav. Res. Methods, Instrum., Comput. 36 (4) (2004) 630–633.

[49] M.-H. Guo, C.-Z. Lu, Q. Hou, Z. Liu, M.-M. Cheng, S.-M. Hu, Segnext: Rethinking convolutional attention design for semantic segmentation, Adv. Neural Inf. Process. Syst. 35 (2022) 1140–1156.

[50] T.Y. Zhang, C.Y. Suen, A fast parallel algorithm for thinning digital patterns, Commun. ACM 27 (3) (1984) 236–239.

[51] J. Ma, J. Chen, M. Ng, R. Huang, Y. Li, C. Li, X. Yang, A.L. Martel, Loss odyssey in medical image segmentation, Med. Image Anal. 71 (2021) 102035.

[52] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, 2014, arXiv preprint arXiv:1412.6980.

[53] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, et al., Pytorch: An imperative style, high-performance deep learning library, in: Advances in Neural Information Processing Systems, vol. 32, 2019.