electronics

MDPI

Article

Facial Wrinkle Detection with Multiscale Spatial Feature Fusion Based on Image Enhancement and ASFF-SEUnet

Jiang Chen [1], Mingfang He [1,*] and Weiwei Cai [2]

[1] School of Computer and Information Engineering, Central South University of Forestry and Technology, Changsha 410004, China; 20211100371@csuft.edu.cn
[2] School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi 214122, China; vivitsai@ieee.org
* Correspondence: t20162306@csuft.edu.cn

Abstract: Wrinkles, crucial for age estimation and skin quality assessment, present challenges due to their uneven distribution, varying scale, and sensitivity to factors like lighting. To overcome these challenges, this study presents facial wrinkle detection with multiscale spatial feature fusion based on image enhancement and an adaptively spatial feature fusion squeeze-and-excitation Unet network (ASFF-SEUnet) model. Firstly, in order to improve wrinkle features and address the issue of uneven illumination in wrinkle images, an innovative image enhancement algorithm named Coiflet wavelet transform Donoho threshold and improved Retinex (CT-DIR) is proposed. Secondly, the ASFF-SEUnet model is designed to enhance the accuracy of full-face wrinkle detection across all age groups under the influence of lighting factors. It replaces the encoder part of the Unet network with EfficientNet, enabling the simultaneous adjustment of depth, width, and resolution for improved wrinkle feature extraction. The squeeze-and-excitation (SE) attention mechanism is introduced to grasp the correlation and importance among features, thereby enhancing the extraction of local wrinkle details. Finally, the adaptively spatial feature fusion (ASFF) module is incorporated to adaptively fuse multiscale features, capturing facial wrinkle information comprehensively. Experimentally, the method excels in detecting facial wrinkles amid complex backgrounds, robustly supporting facial skin quality diagnosis and age assessment.

Keywords: wrinkle detection; multiscale; image enhancement; Unet; feature fusion

check for updates

Citation: Chen, J.; He, M.; Cai, W. Facial Wrinkle Detection with Multiscale Spatial Feature Fusion Based on Image Enhancement and ASFF-SEUnet. *Electronics* **2023**, *12*, 4897. https://doi.org/10.3390/electronics12244897

Academic Editor: Hyunjin Park

Received: 9 November 2023
Revised: 29 November 2023
Accepted: 30 November 2023
Published: 5 December 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Wrinkles are facial structures that appear as a person ages. As an individual ages, the collagen and dermal tissues within the facial skin undergo changes, leading to the development of wrinkles. Wrinkle formation is affected by numerous factors such as overexposure to sunlight and an unhealthy lifestyle [1]. Wrinkles play an important role in face-based analysis as a hallmark feature of the normal aging process [2,3]. In addition, wrinkles are quantifiable and are a key indicator for assessing the quality of facial skin [4].

The methods of wrinkle detection can be categorized into two groups: traditional methods and advanced artificial intelligence methods. Traditional wrinkle detection methods rely on filters, edge detectors, enhancement techniques, and other methods to detect wrinkles. The filter-based wrinkle detection method produces an image with the needed filtering effect by first applying the filter to the original image and then configuring it. Gabor and Hessian Hybrid Filter (HFF)-based techniques are widely used for wrinkle detection. Cula et al. [4] proposed an automatic facial wrinkle detection method based on the Gabor filter. The model has strong robustness and is utilized to make decisions by taking into account the length and depth information of wrinkles as well as using an adaptive threshold. However, this has certain limits, as it can only detect wrinkles on the forehead. Batool and Chellappa [5] proposed a deterministic facial wrinkle detection

method combining the Gabor method and a texture orientation field. A Gabor filter bank is used to highlight curvilinear discontinuities in the skin texture on the basis of wrinkles, and then the texture orientation field is fused to localize the curvilinear shape of the wrinkles. Unfortunately, it is prone to false alarms when confronted with shallower wrinkles and does not work well with uneven image illumination. Ng et al. [6] proposed an automatic wrinkle detection Hessian Hybrid Filter based on directional gradient and ridge Gaussian kernels. It utilizes the multiscale second-order local structure of the image for wrinkle detection, and experiments prove that this method outperforms the Gabor method [4]. However, it is ineffective in detecting fine wrinkles. Since the shape of wrinkles changes with age, different edge information is generated. Filter-based methods focus on the frequency domain analysis of wrinkle images; nevertheless, they have limitations when it comes to wrinkles that are growing or changing and are unable to directly extract the edge information of wrinkles. To address this challenge, researchers have developed an edge-detector-based wrinkle detection method and an enhancement-technique-based method. The edge-detector-based method locates the edge portion by detecting discontinuous regions in the depth of the image. An example is the transient wrinkle detection algorithm proposed by Xie et al. [7]. This method uses a canny edge detector for wrinkle edge pair matching, followed by wrinkle structure localization using Active Appearance Modeling (AAM), and, finally, SVM for wrinkle classification. The researchers claim that the performance of the proposed algorithm is comparable to the HHF [6] and Gabor feature [8]-based wrinkle detectors. Enhancement techniques are applied to analyze images using second-order derivatives and Gaussian scale space. Elbashir and Yap [9] proposed an automatic facial wrinkle detection method that utilizes Gaussian scale space and second-order intensity derivatives for wrinkle detection. The performance of the face wrinkle detection algorithm was evaluated on 45 FERET images and 25 Sudan dataset images. A comparison between Gabor and HHF filters was made. The results showed that HHF works best for wrinkle detection on younger faces and the enhancement method works best for wrinkle detection on older faces.

Following an analysis of the aforementioned literature, wrinkle features are usually employed for wrinkle detection using conventional wrinkle detection methods, which treat them as texture or curve objects. Nonetheless, due to the varying scale sizes of wrinkles, the morphology is complex and variable. There are two primary challenges that limit the effectiveness of these methods when it comes to extracting wrinkle features. On the one hand, fine wrinkles under multiscale problems are difficult to detect. Traditional methods work better in the detection of coarse wrinkles on the face but are ineffective in the detection of fine wrinkles under multiscale problems. Complex background textures and noise may interfere with wrinkle extraction, resulting in fine wrinkles being masked or not accurately detected. This has a negative impact on the accuracy of facial wrinkle analysis and assessment. On the other hand, it is difficult to break through the accuracy of full-face wrinkle detection. Based on the traditional methods, the forehead region often has a superior detection effect, while other facial regions have inferior detection effects that are easily impacted by uneven illumination. The wrinkle detection accuracy varies greatly across different regions of the face, making it difficult for traditional methods to account for this variation. In addition, variations in lighting conditions can lead to changes in the appearance of wrinkles, further increasing the challenges of full-face wrinkle detection. Therefore, the further research and development of more comprehensive and effective wrinkle detection methods are needed to address the challenges posed by different scales, morphologies, and illumination conditions.

Nowadays, deep learning technology continues to develop, and it has very obvious advantages in feature extraction and detection in image processing [10]. One of them, the convolutional neural network, can automatically extract data features by introducing operations such as local connections and weight sharing, and has been widely used for age estimation and skin condition analysis [11–15]. Alarifi et al. [16] used a convolutional neural network (CNN) to classify three different types of facial skin patches, namely, normal skin,

spots, and wrinkles. They experimented by tuning the hyperparameters of GoogleNet and demonstrated that their model achieved 85.2% and 89.9% in terms of f1-score and accuracy, respectively. Sabina et al. [17] proposed a nasolabial wrinkle segmentation based on nested convolutional neural networks. The method is a depth encoder–decoder-type network structure suitable for nasolabial wrinkle data extraction. For wrinkles in the nasolabial region, the method achieved an accuracy of 98.9%, which is a significant improvement, but the effect of the direct detection of wrinkles on the entire face might be diminished. Deepa et al. [18] used image processing and deep convolutional neural networks (CNNs) to detect wrinkles on the skin of a human face. The algorithm recognizes regions of interest (ROIs) containing skin wrinkles and facial features, achieving 96% wrinkle detection. Chang et al. [19] used ResNet with different depths to detect facial spots and wrinkles. First, the facial images were divided into three polygonal regions, the forehead, eyes, and cheeks, and then 200 facial wrinkle images were collected. After data enhancement, more than 129,600 image samples were obtained. ResNet-based training was performed on both local and full-face images with an accuracy of 80% to 90%. This may be due to a strong bias caused by the overexpansion of the samples. Therefore, there is a need to design detection models with a moderate sample size and high accuracy.

Wrinkle detection holds significant application value in the field of facial image processing. Traditional wrinkle detection techniques mainly utilize shallow features such as textures and curves. However, with the development of deep learning technology, it has become possible to automatically extract deeper features. This not only enhances wrinkle detection accuracy, but also offers researchers the opportunity to analyze a broader range of facial features. While deep learning techniques demonstrate great potential in wrinkle detection, their direct application still encounters several challenges. First, wrinkles are complex facial features that vary significantly in scale, shape, depth, and direction, and may change as an individual ages. This complexity poses challenges for the accurate detection of various types and degrees of wrinkles. Deep learning methods need to have sufficient flexibility and generalization capabilities to learn and extract discriminative features from different types of wrinkles. Second, wrinkles have a complex set of image characteristics, such as appearing as curves rather than discrete spots, and are less likely to be concentrated in localized areas or intersect with each other. This requires the deep learning model to effectively capture these continuous curve features and have the ability to accurately identify wrinkles of different shapes and directions. In addition, there are some complex factors that will affect the wrinkle detection results. For example, some shallow wrinkles are more similar to the surrounding skin, and moles and dark spots are prone to appear on the faces of individuals with lighter skin. This difficulty hinders the detector from accurate identification and may interfere with the wrinkle detection results. Therefore, deep learning methods need to be robust to these similar features and be able to distinguish them from real wrinkles. Finally, the collection of publicly available wrinkle images often lacks uniform restrictions, resulting in the problem of uneven illumination, which, in turn, affects the accuracy of the detection results.

In order to tackle the above challenges, the problems of uneven image illumination, the complex morphology of wrinkles with diverse scale sizes, and the small size of the wrinkle image dataset are addressed. Facial wrinkle detection with multiscale spatial feature fusion based on image enhancement and ASFF-SEUnet is proposed. First, a wrinkle image enhancement method, CT-DIR, is proposed to enhance the features of wrinkles under uneven illumination environments. Then, the ASFF-SEUnet neural network is proposed to detect the wrinkle images after the above processing. In addition, a high-quality dataset of facial wrinkle images is provided. The experimental results show that the method can achieve higher detection accuracy. The contributions made by this paper are as follows:

1. An image enhancement method combined with deep learning for full-face wrinkle detection is proposed, which can realize the high-precision recognition of wrinkle detection images in complex environments.

2. A novel ASFF-SEUnet model is proposed. The model takes Unet as the main body and replaces the encoder part of it with EfficientNet, increasing the network depth, width, and output resolution of the encoder part as a way to improve the feature extraction capability of the network. Subsequently, the SE attention mechanism is implemented to direct the network feature extraction to prioritize wrinkle features. Finally, the ASFF structure is added to realize the adaptive fusion of multiscale features, which solves the problem of wrinkles with different scale sizes and complex morphology.
3. The method achieves high detection accuracy compared to traditional detection methods and deep neural networks without image enhancement.

The remaining sections are organized as follows. Section 2 describes the details of the specific method and the material collection and processing flow, Section 3 verifies the validity and superiority of the proposed method, and Section 4 presents the conclusion.

2. Materials and Methods

2.1. Wrinkle Detection Method Based on Multiscale Spatial Feature Fusion

Wrinkle image processing is a critical and challenging task. It takes competent professionals to capture high-quality wrinkle images, but achieving this comes at a great cost and demands an extensive amount of time. At the same time, external factors such as remote transmission may lead to image quality degradation, requiring denoising and enhancement. Traditional methods are limited in global information processing and are susceptible to light inhomogeneity. Lv et al. [20] proposed a wavelet transform Donoho threshold and improved Retinex (WT-DIR) image enhancement algorithm. It performs well in terms of image improvement, reducing noise in the image, enhancing detail information, and resolving uneven illumination. However, such advantages do not necessarily guarantee that it can solve all problems, such as the facial wrinkle detection mentioned in this paper. For this reason, we designed a CT-DIR algorithm for the problem of uneven illumination and complex wrinkle morphology in wrinkle images. The sym wavelet basis used in the WT-DIR algorithm is replaced with the Coiflet wavelet basis. The stronger edge detection and detail enhancement performance of the Coiflet wavelet basis in image processing is employed to achieve the more accurate capture of edge features in wrinkled images and the enhancement of image details.

Deep neural networks are widely recognized as the best algorithms for extracting high-level semantic features. They display excellent performance in image segmentation detection tasks by extracting more abstract features from images layer by layer [21–23]. Some typical deep neural networks used for image segmentation detection include PspNet [24] and DeeplabV3+ [25]. The network structure of these models improves performance by increasing the depth, consequently requiring large-scale wrinkled image data for training. Nevertheless, it is impractical to acquire a large number of high-quality images of wrinkles. To address the aforementioned issues and improve the accuracy of full-face wrinkle detection, detection models with a moderate sample size and high detection accuracy need to be developed. In this paper, we propose a new ASFF-SEUnet model based on the Unet backbone architecture. Unet [26] is a deep neural network mainly used for image segmentation tasks. Its unique U-shaped structure contains encoder and decoder parts, which effectively integrate feature information from different layers through jump connections. This enables Unet to accurately capture details in an image and excel in areas such as medical image segmentation. Its straightforward yet effective design makes it a popular option for image segmentation tasks. The improvements to ASFF-SEUnet are as follows:

1. An EfficientNet is utilized as an encoder to achieve simultaneous adjustment of the width, depth, and resolution of the network and to improve the feature extraction capability of the network.
2. The SE attention mechanism is implemented to increase the network's performance on wrinkle feature extraction and reduce the effect of redundant features.
3. An ASFF structure is organically integrated to achieve the adaptive fusion of multiscale wrinkle features, which boosts the detection of multiscale wrinkles.

The flowchart of facial wrinkle detection with multiscale spatial feature fusion based on image enhancement and ASFF-SEUnet is shown in Figure 1.

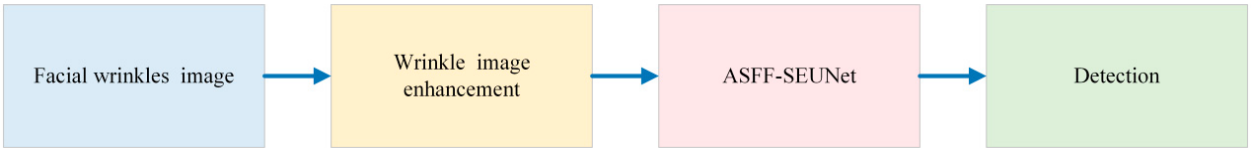


Figure 1. Flowchart of facial wrinkle detection.

First, the facial wrinkle image dataset is obtained (Section 2.2). Second, all facial wrinkle image data are enhanced to obtain enhanced wrinkle feature images (Sections 2.3 and 2.4). Third, the expanded dataset is used to train ASFF-SEUnet (Sections 2.5 and 2.6). Finally, the trained model is applied for facial wrinkle detection (Section 3).

2.2. Graphic Gathering

The importance of the dataset at every stage of wrinkle detection cannot be overstated, from the initial stage of image processing to the final stage of the performance evaluation of the detection algorithm. The wrinkle image dataset for this study was obtained from the high-definition face datasets Flickr Faces High Quality (FFHQ) [27] and UTKFace [28]. As seen in Figure 2, the initial dataset we created contains 1100 images of face wrinkles, including some of the original photographs and labels. In this dataset, most of these images are facial wrinkle images with complex background conditions. Different ages, different skin types, different lighting, different picture resolutions, different genders, etc., are included. The wrinkle features in the pictures also vary in size and have complex and changeable shapes. For example, in terms of age, young people, middle-aged people, and older people are all included. The skin types also include normal skin and skin with interference information such as spots and moles. The lighting also includes overexposed and underlit images. These pictures vary due to the different shooting conditions, and the size was unified to 512 × 512. Although this may result in the loss of some details, this was carried out to meet computational resource and processing efficiency requirements.

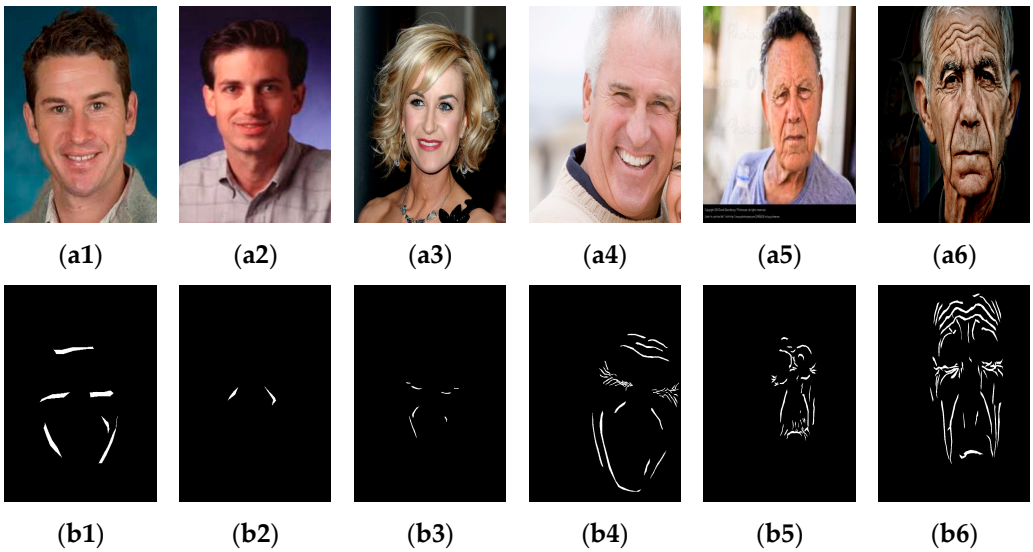


Figure 2. Wrinkle image and labels: (**a1**–**a6**) are the wrinkle images; (**b1**–**b6**) represent the wrinkle labels.

To further improve the performance of the wrinkle detection model, we employ data augmentation techniques to broaden the diversity of the training data during model training. The data enhancement methods used are horizontal flipping, vertical flipping,

horizontal–vertical flipping, and random rotation. The enhancement methods are shown in Table 1. After data enhancement, the size of the sample data is increased fivefold. The number and proportion of enhanced wrinkle images are shown in Table 2. By using data improvement approaches, the model will be able to more effectively adjust to the various kinds of wrinkles and facial features that people have. By expanding the dataset, the performance and generalization of the model can be improved to detect various wrinkle situations more accurately.

Table 1. Data enhancement methods.

Methods	Specific Operations	Methods
Flipping	Horizontal flip, vertical flip, horizontal–vertical flip	Flipping
Rotation	Random rotation at a certain angle with the image center as the origin (0°, 90°, 180°, 270°)	Rotation

Table 2. Number and proportion of wrinkle images after enhancement.

Methods	Number	Proportion
Young people	1175	21.4%
Middle-aged people	3470	63.1%
Older people	855	15.5%

2.3. CT-DIR Wrinkle Image Enhancement Algorithm

Acquired wrinkle images are frequently accompanied by noise interference and low contrast problems. Meanwhile, the regions where wrinkles are present usually exhibit discontinuities in the grayscale, similar to edge features. Therefore, data preprocessing is essential for wrinkle image processing as it maintains image quality while enhancing wrinkle detail information. In order to further improve the image enhancement effect, this paper proposes a new WT-DIR-based algorithm, CT-DIR, which uses Coiflet wavelet bases to replace the Sym wavelet bases employed in the WT-DIR algorithm. Coiflet wavelets are a superior option for image processing because of the following four advantages:

1. Capable of handling sharp changes: Coiflet wavelets are irregular wavelets that can effectively capture and represent sharp changes in an image and are particularly suitable for denoising images with sharp edges or details.
2. An improved quantity of vanishing moments: Coiflet wavelets have a higher number of vanishing moments compared to Sym wavelets, providing a more accurate representation of the signal and better preservation of important features of the image during denoising.
3. Tight support: Coiflet wavelets have tight support, which means that they are localized in both the time and frequency domains. This property allows for efficient and localized denoising, as wavelet coefficients outside the support region can be safely discarded without affecting important information in the image.
4. Resolution flexibility: Coiflet wavelets provide balanced resolution at any time and frequency, which can provide a good trade-off between time and frequency localization, allowing for more accurate analysis of images at different scales.

Our proposed wrinkle image enhancement algorithm, CT-DIR, effectively enhances wrinkle features and solves the problem of uneven illumination. First, by converting the wrinkle image from the RGB space to the HSV space, paying special attention to the wavelet transform of the value component V, we are able to deal with the problem of uneven illumination in a targeted manner. Second, in the wavelet transform, the Coiflet wavelet base is introduced to further improve the processing capability. It copes with problems such as noise interference, weak contrast, and grayscale discontinuity in wrinkle images, and significantly enhances the wrinkle features in wrinkle images. The algorithm flow is shown in Figure 3.

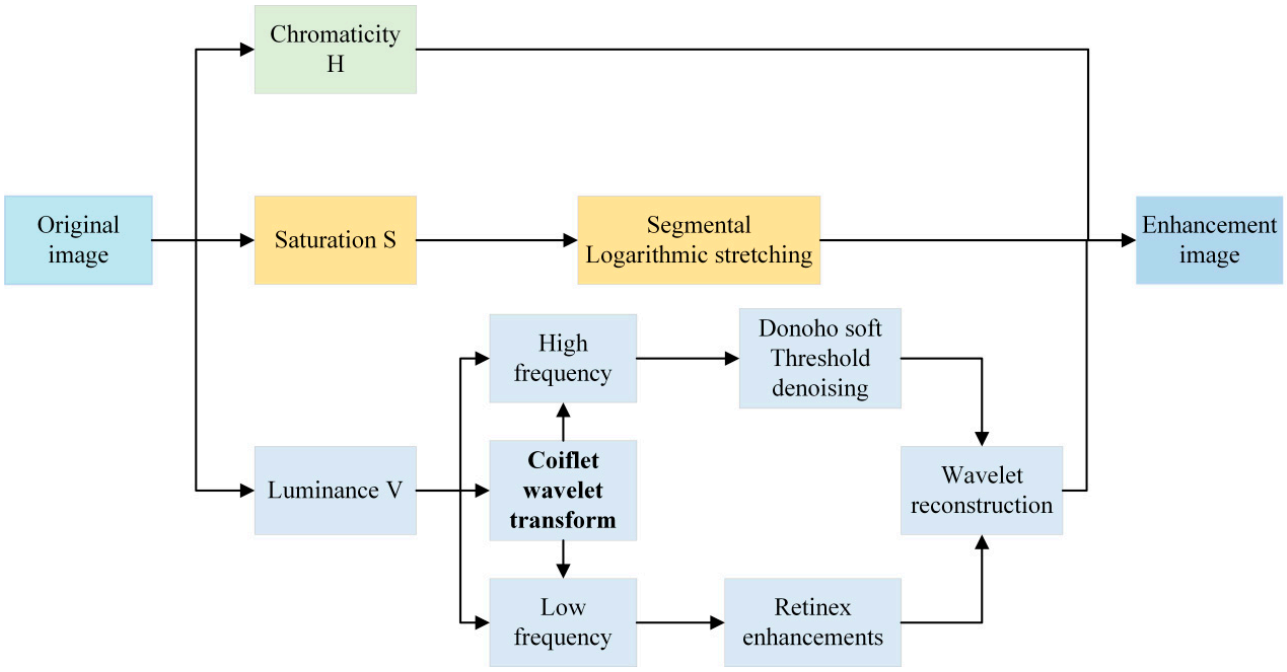


Figure 3. CT-WIR algorithm diagram.

2.4. CT-DIR Wrinkle Image Enhancement Algorithm Theory

- Step 1: Color space conversion

First, the CT-DIR algorithm converts the input facial wrinkle image from RGB color space to HSV color space, which contains three components: brightness (V), hue (H), and saturation (S). The purpose of this step is to separate the color information and brightness information of the image to better process the brightness component, which contains important information about wrinkles. The conversion formula from *RGB* to *HSV* is as follows:

$$H = \begin{cases} 60 * ((G-B)/(V-min(R,G,B))) & if\ V = R \\ 120 + 60 * ((B-R)/(V-min(R,G,B))) & if\ V = G \\ 240 + 60 * ((R-G)/(V-min(R,G,B))) & if\ V = B \end{cases} \quad (1)$$

$$S = \begin{cases} \frac{V-min(R,G,B)}{V} & if\ V \neq 0 \\ 0 & else \end{cases} \quad (2)$$

$$V = max(R,G,B) \quad (3)$$

where V is the brightness, H is the hue, and S is the saturation.

- Step 2: Coiflet wavelet transform

In the HSV space, the CT-DIR algorithm performs Coiflet wavelet transformation on the brightness component V. To better process the features and texture information at various scales, this transformation breaks the image down into several frequency sub-bands. The Coiflet wavelet is the wavelet base of choice because it has an irregular shape that better captures sharp changes and sharp edges in images. Let the brightness component V in formula (3) be expressed as x, and use the characteristics of Coiflet wavelet such as high vanishing moment, compact support, and flexible resolution to transform it in the wavelet domain. The inner product of the continuous signal in the image $x(t)$ and the wavelet basis $\psi_{a,b}(t)$ is as follows:

$$C = < x(t), \psi_{a,b}(t) > = \int_R x(t)\psi_{a,b}(t)dt \quad (4)$$

The inner product C in the Equation (4) of the wavelet basis is used to measure the similarity between the image and a specific wavelet basis, providing information about the structure and features of the image. In wavelet transformation, the computation of discrete inner products is used to determine the projection of the signal or image onto the wavelet basis functions. Through the calculation of discrete inner products, we can achieve discrete wavelet transformation, decomposing the signal or image into wavelet components of different scales and frequencies. This facilitates the more effective handling of signal and image features in a digital environment. Therefore, we can discretize the inner product C.

Let $a = 2^{-j}$ and $b = 2^{-k}$ for j, k $\in$ Z; the discretized form of C can be obtained as follows:

$$D =< x(t), \Psi_{j,k}(t) > \tag{5}$$

where $\Psi_{j,k}(t)$ is the discretized form of $\Psi_{a,b}(t)$.

By employing wavelet bases and scaling functions, we can construct the frequency domain representations of low-pass and high-pass filters. The construction of low-pass and high-pass filters for Coiflet wavelets is as follows. Assume that $\varphi(t)$ and $\Psi'(t)$ are the orthonormal basis of the scale space V_1 and the wavelet basis space W_1, respectively. Then, both $\varphi(t)$ and $\Psi'(t)$ can be linearly expressed using a basis $\{\varphi(2t-k)\}k \in Z$ of the Euclidean space V_0.

$$\begin{cases} \varphi(t) = \sum\limits_{k\in Z} h_k \varphi(2t-k) \\ \Psi'(t) = \sum\limits_{k\in Z} g_k \varphi(2t-k) \end{cases} \tag{6}$$

where h_k, g_k are the weights of the linear combination.

From this, the low-pass and high-pass filters of the Coiflet wavelet are as follows:

$$\begin{cases} h(\omega) = \frac{1}{2} \sum\limits_{k\in Z} h_k exp(-qk\omega) \\ g(\omega) = \frac{1}{2} \sum\limits_{k\in Z} g_k exp(-qk\omega) \end{cases} \tag{7}$$

where $h(\omega)$ and $\varphi(t)$ are the corresponding low-pass filters, $g(\omega)$ and $\varphi(t)$ are the corresponding high-pass filters, and $exp(-qk\omega)$ is the Laplace transform factor. The discrete sequence form of $\{h(\omega),g(\omega)\}$ in the time domain can be expressed as $\{h_k, g_k\}$.

- Step 3: Low-frequency coefficient enhancement

After the wavelet transform, the CT-DIR algorithm uses the improved Retinex image enhancement algorithm to process the obtained low-frequency coefficients. The Retinex algorithm is a method widely used in image enhancement to improve the quality of images by adjusting the brightness and contrast of the image. In this step, the improved Retinex algorithm helps maintain the overall brightness characteristics of the image, ensuring that brightness adjustments do not erase wrinkle detail information. In the improved Retinex algorithm, the low-frequency part of V in the wavelet domain is regarded as the product of the illumination component $L(x, y)$ and the reflection component $R(x, y)$. The formula is as follows:

$$I(x, y) = L(x, y) \times R(x, y) \tag{8}$$

where (x, y) represents the two-dimensional coordinates of the low-frequency component of the wrinkle image after wavelet transformation. The reflection component $R(x, y)$ is the clear image that needs to be solved, and it is necessary to transform the estimate of $L(x, y)$ into the logarithmic domain.

$$log\widetilde{\mathrm{R}}(x, y) = logI(x, y) - log(x, y) \tag{9}$$

$\widetilde{\mathrm{R}}(\mathrm{x}, \mathrm{y})$ and $\widetilde{\mathrm{L}}(\mathrm{x}, \mathrm{y})$ represent the estimated values of the low-frequency reflection and illumination components of the wrinkle image. Equation (9) shows that color constancy can be achieved when the low-frequency image of the wrinkle is not affected by ambient

light. The low-frequency illumination component of the wrinkle image can be estimated using the center-around idea. The formula is as follows:

$$\widetilde{L}(x,y) = I(x,y) * F(x,y) \tag{10}$$

where $*$ represents the convolution operation and F represents the center-around function.

Typically, center-around functions exhibit strong dynamic compression capabilities, such as Gaussian functions, which are used to simulate slow signal transformations in wrinkle images.

- Step 4: High-Frequency Coefficient Denoising

High-frequency coefficients contain details and texture information about the image, but usually also include some noise. In order to reduce the effect of this noise, the CT-DIR algorithm applies the Donoho thresholding method to the high-frequency coefficients for denoising. The Donoho thresholding method is a threshold-based image denoising technique. By setting high-frequency coefficients below a certain threshold to zero, the noise is reduced while important image details are retained. The formula for the Donoho soft threshold is as follows:

$$\delta = \sigma\sqrt{2\ln N} \tag{11}$$

where σ denotes the standard deviation of the noise, N denotes the length of the signal, and δ denotes the desired threshold.

After wavelet transform, the V component f (x,y) of the wrinkle image is decomposed into multiple different scales, and the nonlinear threshold of the j-th scale is defined as follows:

$$\delta = \sigma_j\sqrt{[2log(j+1)]/j} \quad j = 1, 2, 3, \cdots, N \tag{12}$$

where σ_j represents the standard deviation of the noise in the wrinkle high-frequency image on the j-th scale. Considering the multiscale characteristics of signal and noise, we select appropriate thresholds to compress the wavelet coefficients of different scales and obtain denoised wrinkle images through wavelet inverse transformation. This method efficiently eliminates noise while maintaining essential features, making it advantageous for the further analysis of wrinkle images.

- Step 5: Logarithmic transformation

Utilizing logarithmic transformation aims to enhance the contrast of the facial wrinkle image while preserving or appropriately reducing high-saturation areas and improving low-saturation regions to achieve an overall enhancement in image clarity and saturation. The formula for performing logarithmic transformation to enhance the saturation component is shown in (13):

$$S(x,y) = \begin{cases} w_1 \times log[1 + S(x,y)] & 0 < S(x,y) \leqslant 0.25 \\ w_2 \times log[1 + S(x,y)] & 0.25 < S(x,y) \leqslant 0.5 \\ w_3 \times log[1 + S(x,y)] & 0.5 < S(x,y) \leqslant 0.75 \\ log[1 + S(x,y)] & 0.75 < S(x,y) \leqslant 1 \end{cases} \tag{13}$$

where w_1, w_2, and w_3 represent the coefficients, and $S(x,y)$ represents the saturation component.

- Step 6: Reconstruct the image

After the denoising of the high-frequency coefficient and the enhancement of the low-frequency component, the enhanced brightness component (V) is reconstructed using the CT-DIR algorithm. This enhanced brightness component is then merged with the hue component (H) and saturation component (S) of the original image to generate the final enhanced image.

The facial wrinkle feature enhancement results of this article's CT-DIR algorithm are shown in Figure 4. This method provides a more powerful tool for wrinkle image analysis, diagnosis, and research.

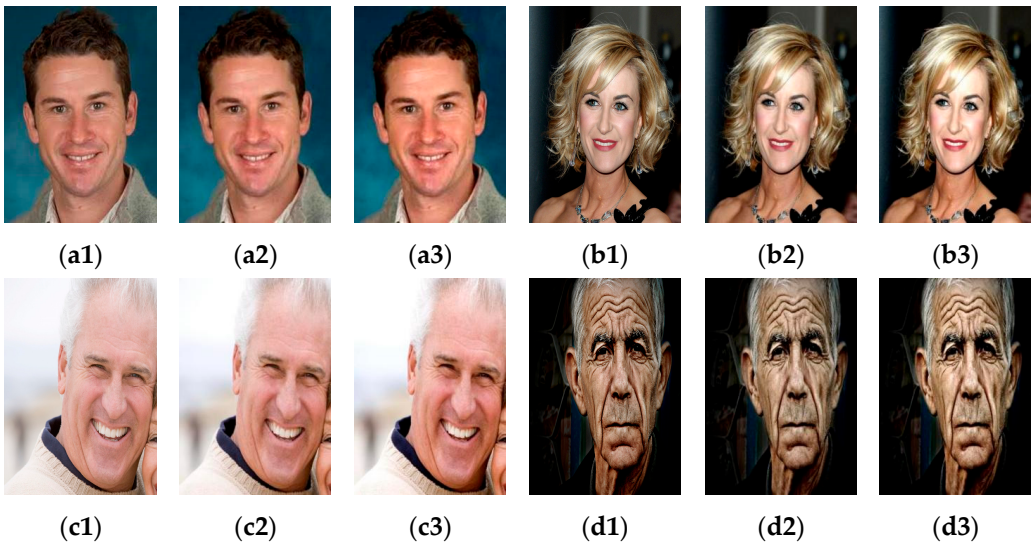


Figure 4. Original image and image enhancement results of WT-DIR and CT-DIR algorithms: (**a1**–**d1**) are original wrinkle images of different ages and genders, (**a2**–**d2**) are the enhanced results of the WT-DIR algorithm, and (**a3**–**d3**) are the enhanced results of the CT-DIR algorithm.

2.5. ASFF-SEUnet

The ASFF-SEUnet network model was proposed by us to address the issue of wrinkles not being evenly distributed throughout the face, having varying sizes, and having intricate and variable shapes. The encoder part in the Unet backbone is replaced with EfficientNet. The output of the input layer of Unet is used as the input of EfficientNet, and the output of four layers is obtained from EfficientNet as the input of the skip connection part. This improvement can simultaneously adjust the network depth, width, and output resolution of the encoder part to improve the wrinkle feature extraction capability of the network. The SE (squeeze-and-excitation) attention mechanism is introduced to enable the model to automatically learn the correlation and importance between features, increase the weight of wrinkle features in the network, and suppress noise and unimportant features. The ASFF (adaptively spatial feature fusion) module is added to adaptively fuse facial wrinkle features at different scales, which strengthens the model's ability to extract facial wrinkle features at different scales, thereby capturing the feature information of facial wrinkles more comprehensively.

The Unet network serves as the foundation for the ASFF-SEUnet network structure. The encoder part is replaced with EfficientNet, and then the SE attention mechanism is introduced and the ASFF module added into the middle skip connection part. The network structure diagram of ASFF-SEUnet is shown in Figure 5. It includes the encoder part, the SE attention mechanism part, the ASFF module part, and the decoder part composed of EfficientNet. The network parameters of ASFF-SEUnet are shown in Table 3.

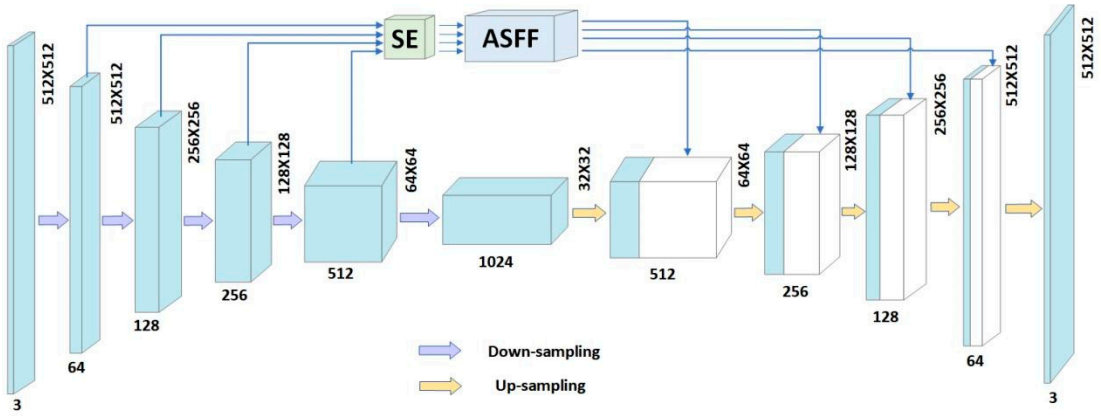


Figure 5. ASFF-SEUNet network structure diagram.

Table 3. ASFF-SEUnet network parameter table.

Layer		Parameter	Follow-Up Action
Input		512 × 512 × 3	
Encoder EfficientNet	Down-sampling 1	Depthwise filter (3 × 3), 1 stride (MBConv1)	Feat1 (SE + ASFF)
	Down-sampling 2	Depthwise filter (3 × 3), 2 strides (MBConv2) Depthwise filter (3 × 3), 2 strides (MBConv3)	Feat2 (SE + ASFF)
	Down-sampling 3	Depthwise filter (5 × 5), 2 strides (MBConv4) Depthwise filter (5 × 5), 2 strides (MBConv5)	Feat3 (SE + ASFF)
	Down-sampling 4	Depthwise filter (3 × 3), 2 strides (MBConv6) Depthwise filter (3 × 3), 2 strides (MBConv7) Depthwise filter (3 × 3), 2 strides (MBConv8) Depthwise filter (5 × 5), 1 stride (MBConv9) Depthwise filter (5 × 5), 1 stride (MBConv10) Depthwise filter (5 × 5), 1 stride (MBConv11)	Feat4 (SE + ASFF)
	Down-sampling 5	Depthwise filter (5 × 5), 2 strides (MBConv12) Depthwise filter (5 × 5), 2 strides (MBConv13) Depthwise filter (5 × 5), 2 strides (MBConv14) Depthwise filter (5 × 5), 2 strides (MBConv15) Depthwise filter (5 × 5), 1 stride (MBConv16)	Feat5
Decoder	Up-sampling 1	Deconv filter (2 × 2), 2 strides (Deconv1) Concat (feat1 (SE+ASFF), Deconv1) Convolution filter (3 × 3), 1 stride (conv1) Convolution filter (3 × 3) 1 strides (conv2)	
	Up-sampling 2	Deconv filter (2 × 2), 2 strides (Deconv2) Concat (feat2 (SE + ASFF), Deconv2) Convolution filter (3 × 3), 1 stride (conv3) Convolution filter (3 × 3) 1 stride (conv4)	
	Up-sampling 3	Deconv filter (2 × 2), 2 strides (Deconv3) Concat (feat3 (SE + ASFF) Convolution filter (3 × 3), 1 stride (conv5) Convolution filter (3 × 3), 1 stride (conv6)	
	Up-sampling 4	Deconv filter (2 × 2), 2 strides (Deconv4) Concat (feat4 (SE + ASFF), Deconv4) Convolution filter (3 × 3), 1 stride (conv7) Convolution filter (3 × 3), 1 stride (conv8)	
Output		512 × 512 × 3	Softmax

2.6. ASFF-SEUnet theory

2.6.1. EfficientNet

The characteristics of wrinkles are multifaceted and changing, exhibiting notable variations across various age groups. To address this problem, considering the encoder feature extraction capability step of the original Unet network, this article replaces the encoder part of the Unet backbone network with EfficientNet. EfficientNet has the following advantages in wrinkle detection. First, it improves feature extraction capabilities, effectively captures skin texture and subtle changes through a deeper and wider structure, and improves detection accuracy. Second, EfficientNet reduces the model complexity and achieves higher efficiency and practicality by automatically adjusting the network width, depth, and resolution.

The structure of the baseline EfficientNet-B0 in the EfficientNet series network [29] is shown in Table 4. The structure includes an initial feature extraction layer (Stem), multiple repeated feature extraction modules (Blocks), and an output layer (Head). The Stem layer is used for preliminary feature extraction and down-sampling of the input image. The Blocks layer is composed of multiple repeated basic blocks. Each basic block contains operations such as depth-wise separable convolution, batch normalization, and activation functions to extract the high-level features of the image. Finally, the Head layer maps the extracted features to the final output category or regression result.

Table 4. EfficientNet-B0 parameter table.

Structure	Stage	Operator	Resolution	Channels	Layers
Stem	1	Conv3 × 3	224 × 224	32	1
Blocks	2	MBconv1, k3 × 3	112 × 112	16	1
	3	MBconv6, k3 × 3	112 × 112	24	2
	4	MBconv6, k3 × 3	56 × 56	40	2
	5	MBconv6, k3 × 3	28 × 28	80	3
	6	MBconv6, k3 × 3	14 × 14	112	3
	7	MBconv6, k3 × 3	14 × 14	192	4
	8	MBconv6, k3 × 3	7 × 7	320	1
Head	9	Conv1 × 1&Pooling&FC	7 × 7	1280	1

2.6.2. SE Attention Mechanism

Facial wrinkle images are complex and changeable, with interference information such as spots and moles. The network is easily affected by this irrelevant feature information when extracting wrinkle features. Therefore, we introduce the SE attention mechanism, namely, the squeeze-and-excitation network (SENet) [30]. This method enables the model to automatically learn the correlation and importance between features, increase the weight of wrinkle features in the network, and suppress noise and unimportant features. The network structure of SENet is shown in Figure 6.

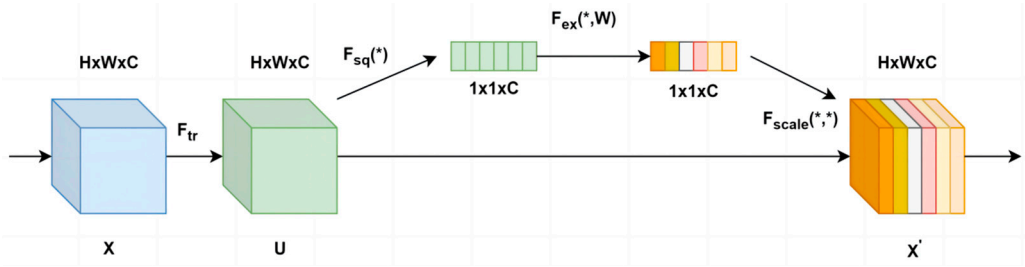


Figure 6. SENet structure diagram.

The feature map X is the input, U is the output after the convolution structure F_{tr}, * represents the independent variable, F_{sq} represents the Squeeze operation, F_{ex} represents the Excitation operation, and F_{scale} represents the element-level multiplication operation of the channel attention vector and the original feature map.

First is the Squeeze operation: the input feature map has dimensions of H × W × C. After global average pooling, a global descriptor is obtained with dimensions of 1 × 1 × C. This process compresses the channel dimension of the feature map but preserves the information between channels. This can be expressed as follows:

$$F_{sq}(U) = Global\,AveragePooling\ (U) \tag{14}$$

Next, the Excitation operation models the relationship between channels by introducing a multi-layer perceptron (MLP). MLP includes two fully connected (FC) layers, where the first fully connected layer is used to reduce the dimensionality, and the second fully connected layer is used to learn the weights between channels. MLP can produce a channel attention vector that illustrates the significance of each channel by utilizing sigmoid functions and activation functions (such as ReLU).

$$S = \sigma(W_2\delta(W_1 z)) \tag{15}$$

where W_1 and W_2 are the weight matrices of the FC layer, δ is the activation function ReLU, σ is the Sigmoid activation function, and S is the excitation value of each channel.

Finally, the channel attention vector is element-wise multiplied with the original feature map. Attention weights are applied to the original feature maps to enhance useful features and suppress useless features. This allows SENet to adaptively adjust the weights of various channels, enhancing the model's capacity to capture important features. The formula is as follows:

$$X' = S * X \tag{16}$$

where X is the original feature map, and X' is the feature map reweighted according to the channel excitation values.

2.6.3. ASFF Module

The network is unable to effectively use multiscale wrinkle features due to the variation in the scales of wrinkles in facial wrinkle images. Therefore, this article adds the ASFF (adaptively spatial feature fusion) module [31] to adaptively fuse the facial wrinkle features of different scales. The model's ability to extract facial wrinkle features at different scales is strengthened, thereby capturing the characteristic information of facial wrinkles more comprehensively. The structure of ASFF is shown in Figure 7.

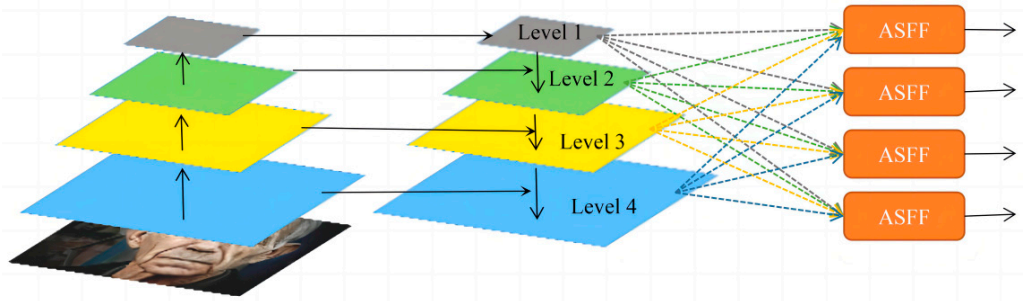


Figure 7. ASFF structure diagram.

Let F_1, F_2, ..., F_n represent the input feature maps and n represent the number of feature maps. For each feature map F_i, ASFF introduces a weight W_i to represent its importance. The weight W_i is obtained by global average pooling of the feature map F_i to obtain the

feature vector x_i, and is processed by a fully connected layer and an activation function (such as sigmoid). Specifically, it can be expressed using the following formula:

$$x_i = GlobalAveragePooling\ (F_i) \tag{17}$$

$$W_i = \sigma\ (W_1 * x_i + b_1) \tag{18}$$

where GlobalAveragePooling represents the global average pooling operation, W_1 and b_1 are the weights and biases of the fully connected layer, and σ represents the activation function.

To ensure that the sum of all weights is 1, ASFF normalizes the weight *Wi*. Specifically, it can be expressed using the following formula:

$$\alpha_i = \frac{W_i}{\sum W_i} \tag{19}$$

where α_i represents the normalized weight.

Finally, through weighted fusion by channel, the fused feature map F is obtained. It can be expressed using the following formula:

$$F = \sum(\alpha_1 * F_i) \tag{20}$$

3. Application and Results Analysis

3.1. Evaluation Metrics

To comprehensively evaluate the proposed model, the evaluation metrics adopted include pixel accuracy (PA), mean intersection over union (MIoU), precision, recall, and F1-Score. Pixel accuracy refers to the ratio of correctly labeled pixels to total pixels. MIoU represents the ratio of the intersection and union of two sets (real values and predicted values). Precision is the ratio of actual wrinkle pixels to predicted wrinkle pixels. Recall is the ratio of correctly predicted wrinkle pixels to actual wrinkle pixels. Moreover, F1-Score takes precision and recall into account and can be viewed as a weighted average of precision and recall. PA is an intuitive metric used to assess the overall pixel level accuracy of a model. MIoU, commonly employed in image segmentation tasks, considers the ratio of intersection to union between predicted and ground truth results, providing a more detailed assessment of the model's segmentation performance across different classes. Precision measures the model's ability to correctly classify positive instances, while Recall assesses the model's capability to detect actual positive instances. F1-Score is a composite metric that takes into account both precision and recall, particularly useful in scenarios with imbalanced class distributions, helping strike a balance between precision and recall. The expressions of the above evaluation indicators are as follows:

$$\mathrm{PA} = \frac{\sum_{i=0}^{k} p_{ij}}{\sum_{i=0}^{k} \sum_{j=0}^{k} p_{ij}} \tag{21}$$

$$\mathrm{MIoU} = \frac{1}{k+1} \sum_{i=0}^{k} \frac{p_{ij}}{\sum_{j=0}^{k} p_{ij} + \sum_{j=0}^{k} p_{ji} - p_{ii}} \tag{22}$$

$$\mathrm{Precision} = \frac{TP}{TP + FP} \tag{23}$$

$$\mathrm{Recall} = \frac{TP}{TP + FN} \tag{24}$$

$$\mathrm{F1-Score} = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{25}$$

where the detection target is k class, P_{ii} represents the number of pixels that divide the i-th class into the i-th class (the number of pixels that are correctly classified), P_{ij} represents

the number of pixels that divide the *i*-th class into the *j*-th class (the number of all pixels), the true positive value (*TP*) refers to the number of pixels correctly identified as wrinkles, and the false positive value (*FP*) represents the number of pixels incorrectly identified as wrinkles. The false negative value (*FN*) represents the number of pixels incorrectly identified as non-wrinkles, and the true negative value (*TN*) represents the number of pixels correctly identified as non-wrinkles.

3.2. Experimental Setup

Hardware environment: Processor: Intel Core i9-9980XE, memory 500GB; graphics card: NVIDIA GeForce RTX 2080 TI; system memory: 64G. Software environment: CUDA 11.1, Python 3.8, TensorFlow-GPU 2.5; operating system: Windows 10. The unified input size of the images is 512 × 512. The enhanced dataset is divided into a training set and a test set according to the ratio of 8:2.

Considering the performance and training effect of the hardware device, the adam optimization algorithm method was employed for the network training in this study. The TensorFlow-GPU 2.5 framework was used for training in the Windows 10 software environment of Python 3.8. The training and test size of each batch was set to 4, which meant the batch size was 2, the momentum parameter was 0.9, the maximum learning rate init_lr = 1×10^{-4}, the minimum learning rate init_lr * 0.01, the learning descent method was cos, and the epoch was 300.

3.3. Ablation Experiment

This article's model presents EfficientNet, the SE attention mechanism, and the ASFF module, mostly using Unet as its backbone. Therefore, the ablation experiment compares and verifies the proposed method with Unet, Unet + EfficientNet, and Unet + EfficientNet + SE. The test results for various indicators are shown in Table 5. The experimental results show that after replacing the encoder part in Unet with EfficientNet, the wrinkle feature extraction capability of the network was enhanced. When compared to the original Unet, the precision and F1-Score improved significantly, by 2% and 1.01%, respectively. Comparing Unet + EfficientNet and Unet + EfficientNet + SE, the PA value of the Unet + EfficientNet + SE model that introduced the SE attention mechanism increased to 99.4%, and the MIoU also increased by 3.31%. This shows that after introducing the SE attention mechanism, the important feature extraction ability of the network was enhanced, and useful wrinkle information could be extracted from complex facial features. Further comparing the proposed ASFF-SEUnet method with Unet + EfficientNet + SE, the proposed approach achieved improvements of 0.2% in PA, 0.35% in MIoU, 2.37% in precision, and 0.86% in F1-Score. This proves that the addition of the ASFF module improved the multiscale wrinkle detection capability. The detection effect for fine wrinkles was also enhanced, in addition to the high detection effect for coarse wrinkles. These experiments reveal that each component is important in wrinkle detection. With a PA value of 99.6%, the proposed ASFF-SEUnet performs exceptionally well in wrinkle detection and is appropriate for multiscale facial wrinkle detection.

Table 5. Ablation experiment table.

Network	PA	MIoU	Precision	Recall	F1-Score
Unet	98.44%	52.29%	62.77%	75.8%	68.67%
Unet + EfficientNet	98.60%	53.11%	64.77%	74.6%	69.68%
Unet + EfficientNet + SE	99.4%	56.42%	67.27%	77.78%	71.56%
ASFF-SEUnet	99.6%	56.77%	69.64%	75.44%	72.42%

The investigation was carried out to obtain a better grasp of the rationale for the proposed method's superiority. The results of the full-face facial wrinkle detection for people of different ages and genders were visually compared, as shown in Figure 8. It can be seen from the figure that Unet cannot detect multiscale wrinkles very well, and there are some

missed detections. When comparing Figure 8(a1–f1,a2–f2) of Unet and Unet + EfficientNet, it can be seen that after using EfficientNet to replace the encoder in Unet, the detection effect is improved, more wrinkles are detected, and fewer are missed. This is because EfficientNet simultaneously adjusts the depth, width, and resolution of the network compared to the original encoder module. The wrinkle feature extraction capability of the network is improved, which makes up for the insufficient feature extraction capability of the original encoder, making the network more comprehensive in detecting wrinkles. Further, comparing Figure 8(a2–f2,a3–f3) of Unet + EfficientNet and Unet + EfficientNet + SE reveals that when the SE attention mechanism is introduced, more accurate wrinkle detection is achieved. It is worth noting that some misdetected areas were excluded after the SE attention mechanism was introduced. This demonstrates how the network can automatically learn the correlation and significance of features once the SE attention mechanism is included, giving wrinkle features more weight in the network and reducing noise and irrelevant information. In the visual comparison between Unet + EfficientNet + SE and ASFF-SEUnet, it can be seen from Figure 8(a3–f3,a4–f4) that small wrinkles that were extremely difficult to distinguish were detected. This shows that after adding the ASFF module, multiscale factors are taken into account, which strengthens the detection performance of local details and improves the local wrinkle detection effect. Additionally, there is reasonable multiscale feature fusion while retaining the global detection effect, the edge details of thick wrinkles are detected more finely, and the wrinkle skeleton can be completely and accurately detected.

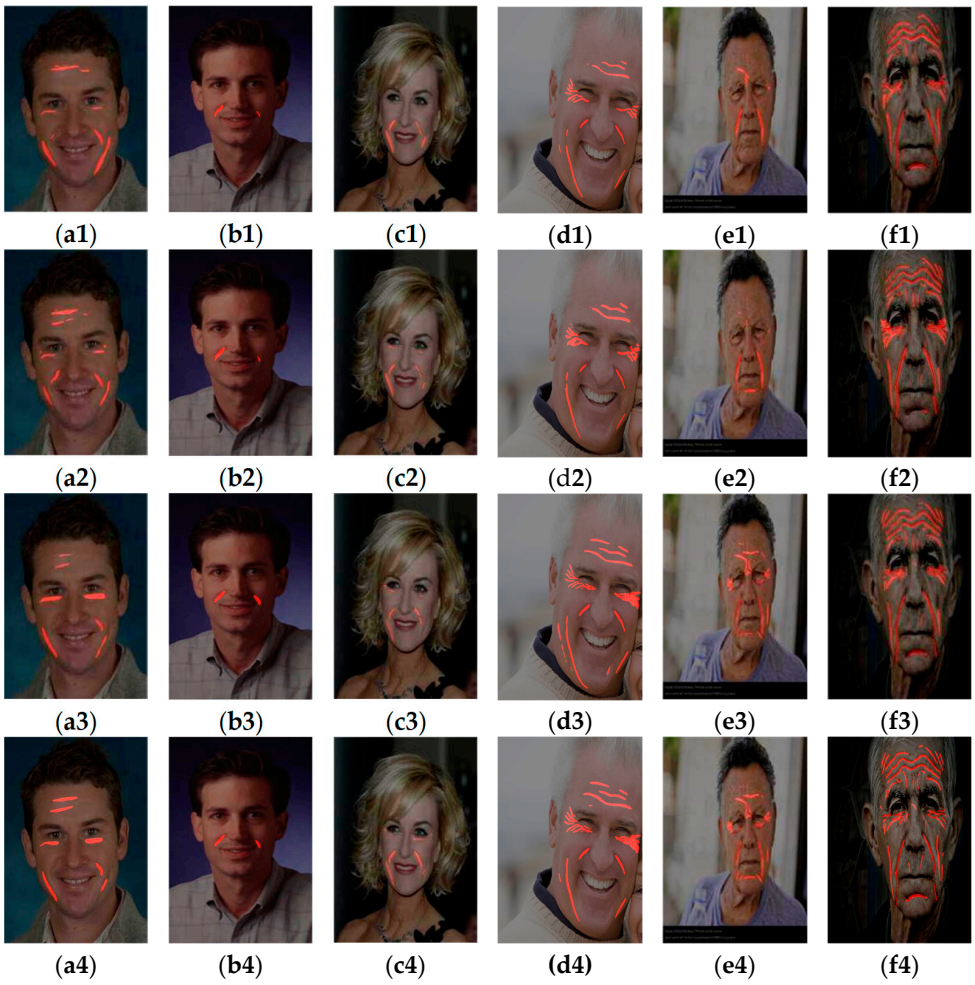


Figure 8. Ablation experiment results: (**a1**–**f1**) show Unet wrinkle detection results; (**a2**–**f2**) show Unet + EfficientNet wrinkle detection results; (**a3**–**f3**) show Unet + EfficientNet + SE wrinkle detection results; and (**a4**–**f4**) show ASFF-SENet Wrinkle detection results.

3.4. Comparative Experiment

In order to more comprehensively evaluate the performance of the ASFF-SENet model, FCN-4s [32], PspNet, Deeplabv3+, BC-DUnet [33], and HC-Unet++ [34] were used to conduct comparative experiments. The experimental results are shown in Table 6. Compared to the nested convolutional neural network (NCNN) presented in [17], ASFF-SEUnet demonstrates notable enhancements with a 0.7% improvement in PA and a 2.42% increase in MioU. Evidently, ASFF-SEUnet excels in wrinkle detection accuracy, exhibiting superior performance in detecting wrinkles across the entire face. In comparison with the three baseline models FCN-4s, PspNet, and Deeplabv3+, ASFF-SEUnet's wrinkle feature extraction capability is more powerful, and the wrinkle detection accuracy has also been greatly improved. ASFF-SEUNet improves PA by 1.1% compared to FCN-4s. This may be due to the jump connection structure in the backbone of Unet-type networks. It can fuse the semantic information of different depths in the network and improve the accuracy of positioning wrinkle pixels, thus improving the accuracy of ASFF-SENet. ASFF-SENet has significantly improved in terms of precision and recall when compared to PspNet and DeeplabV3+. This is because the correlation and importance between features are considered, the weight of important wrinkle features is increased, and the interference of non-wrinkle information is reduced. In comparison with the two advanced models BC-Dunet and HC-Unet++, ASFF-SEUnet has better multiscale wrinkle extraction capabilities, and the network structure is also more suitable for processing wrinkle images. ASFF-SEUNet has a 2.32% improvement in MioU over BC-Dunet. This is because the network targets the multiscale characteristics of wrinkles and achieves adaptive multiscale feature fusion. The model's ability to extract facial wrinkle features at different scales is strengthened, thereby capturing the characteristic information of facial wrinkles more comprehensively. The F1-Score value of ASFF-SEUNet is similar to HC-Unet++, but 0.31% ahead of HC-Unet++. This is because the proposed ASFF-SEUnet has a reasonable structural design for the different sizes and complex shapes of wrinkles in wrinkle images. Based on improving wrinkle feature extraction capabilities, it also pays more attention to wrinkle features and introduces multiscale fusion processing. The analysis shows that ASFF-SEUNet is in a leading position among the six networks and can effectively improve wrinkle detection capabilities. It addresses the issues of wide variations in wrinkle scales, which result in unsatisfactory detection ability, and the uneven distribution of wrinkles around the face, which makes comprehensive detection challenging.

Table 6. Comparative experiment table.

Network	PA	MioU	Precision	Recall	F1-Score
NCNN	98.9%	54.36%	63.32%	73.52%	69.49%
FCN-4s	98.5%	37.5%	43%	74.54%	54.54%
PspNet	99.5%	48.35%	66.13%	64.25%	65.18%
DeeplabV3+	99.2%	50.23%	68.12%	68.2%	68.1%
BC-Dunet	99.2%	54.45%	68.33%	64.22%	68.3%
HC-Unet++	99.3%	55.42%	68.76%	74.55%	71.85%
ASFF-SEUnet	99.6%	56.77%	69.64%	75.44%	72.42%

We find that the combination of image enhancement and ASFF-SEUNet outperforms other networks. The possible reasons for this are that (1) compared with the proposed method, the FCN-4s model does not take into account the fusion of different-depth semantic information. PspNet and DeeplabV3+ ignore the correlation and importance between features. BC-Dunet and HC-Unet++ find it difficult to deal with the multiscale problem of wrinkles; (2) ASFF-SEUnet uses EfficientNet in the encoder part, which can adjust the depth, width, and resolution of the encoder part at the same time, improving feature extraction capabilities. Subsequently, the SE attention mechanism is used to facilitate the automatic learning of feature association and importance by the model. In addition, the ASFF module is added to adaptively fuse facial wrinkle features at different scales,

strengthening the model's ability to extract facial wrinkle features at different scales, thereby more comprehensively capturing the characteristic information of facial wrinkles; (3) our proposed ASFF-SEUnet model is combined with the designed wrinkle image enhancement algorithm CT-DIR.

We visualized the proposed ASFF-SENet and five other networks in order to delve deeper into the causes behind the proposed method's superior performance. Figure 9 shows the wrinkle detection effects of six networks. The figure illustrates that the detection effect of FCN-4s is not good, especially for wrinkle detection in the elderly. The wrinkle detection effects of PspNet and DeeplabV3+ have been improved compared to FCN-4s, and most obvious wrinkles can be detected. BC-Dunet and HC-Unet exhibit better detection results than the first three methods. For wrinkle detection, the proposed ASFF-SEUnet outperforms the other five networks. It is proven that ASFF-SEUnet is an effective wrinkle detection method with broad application prospects, for example, in the fields of age assessment and skin quality diagnosis. On the one hand, during the age assessment process, there are differences in facial wrinkle characteristics between people of different ages. More accurate facial wrinkle feature detection can improve the accuracy of age assessment and reduce misjudgments of age. On the other hand, wrinkles are a key indicator in skin quality diagnosis and have an important impact on such diagnoses. Improving the accuracy of facial wrinkle detection can help skin experts or related personnel diagnose skin quality.



Figure 9. Comparative method wrinkle detection effect diagram: (**a1**,**a2**) show the wrinkle detection effect diagram of FCN-4s; (**b1**,**b2**) show the wrinkle detection effect diagram of PspNet; (**c1**,**c2**) show the results for DeeplabV3+; (**d1**,**d2**) show the wrinkle detection renderings of BC-DUnet; (**e1**,**e2**) show the wrinkle detection renderings of HC-Unet++; and (**f1**,**f2**) are the wrinkle detection renderings of ASFF-SEUnet.

4. Conclusions

As computer vision technology has advanced, experts and academics from both domestic and international universities have studied image analysis techniques in extensive detail. In recent years, facial defect detection based on deep learning has been increasingly widely used in the field of facial wrinkle detection. To improve the accuracy and effectiveness of wrinkle detection, a facial wrinkle detection with multiscale spatial feature fusion based on image enhancement and ASFF-SEUnet is proposed. To minimize the impact of lighting conditions and improve wrinkle details, this method first conducts image enhancement, and then the ASFF-SEUnet model is used for detection, and the detection accuracy has been shown to be as high as 99.6%. Through comparative experimental analysis, the ASFF-SEUnet model is superior to the other three baseline models and two advanced models in terms of precision, recall rate, and F1-Score. It improves the detection effect of

local fine wrinkles and improves the edge detail detection ability of overall coarse wrinkles on the whole face, solving the problem of difficulty in detecting multiscale wrinkle features. The proposed model has made significant progress in facial wrinkle detection, but it still has some potential limitations. Firstly, the model may not encompass the full diversity of facial wrinkle images during the training stage, and in practical applications, it may not generalize to all populations. Secondly, the computational efficiency of the model for practical applications, especially its feasibility on mobile devices or real-time systems, needs evaluation. Going forward, we plan to expand the dataset by collecting more diverse facial wrinkle images to enhance the model's generalization performance. Subsequently, we will deploy the model on mobile devices or real-time systems to verify its computational efficiency. In practical applications, we anticipate extending this model to fields such as medical care, beauty, and age recognition. With the support of cloud architecture, our method can be combined with online detection equipment to provide essential support for intelligent medical decision making.

Author Contributions: Conceptualization, J.C. and M.H.; methodology, M.H. and W.C.; software, J.C. and W.C.; validation, M.H. and W.C.; formal analysis, J.C. and M.H.; investigation, M.H. and W.C.; resources, M.H. and W.C.; writing—original draft preparation, J.C. and M.H.; writing—review and editing, M.H. and W.C.; supervision, M.H. and W.C.; project administration, M.H. and W.C.; funding acquisition, M.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Science Foundation of Hunan Province, (Grant No. 2021JJ41087), and the National Natural Science Foundation of China (Grant No. 61703441).

Data Availability Statement: The data presented in this study are openly available in FFHQ face databased at https://doi.org/10.48550/arXiv.1812.04948 and UTKFace face databased at https://doi.org/10.48550/arXiv.1702.08423, [27,28].

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

References

1. Fu, Y.; Guo, G.; Huang, T.S. Age synthesis and estimation via faces: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 1955–1976. [PubMed]
2. Luu, K.; Dai Bui, T.; Suen, C.Y.; Ricanek, K. Combined local and holistic facial features for age-determination. In Proceedings of the IEEE 2010 11th International Conference on Control Automation Robotics & Vision, Singapore, 7–10 December 2010; pp. 900–904.
3. Ng, C.C.; Yap, M.H.; Cheng, Y.T.; Hsu, G.S. Hybrid ageing patterns for face age estimation. *Image Vis. Comput.* **2018**, *69*, 92–102. [CrossRef]
4. Cula, G.O.; Bargo, P.R.; Nkengne, A.; Kollias, N. Assessing facial wrinkles: Automatic detection and quantification. *Skin Res. Technol.* **2013**, *19*, 243–251. [CrossRef] [PubMed]
5. Batool, N.; Chellappa, R. Detection and inpainting of facial wrinkles using texture orientation fields and Markov random field modeling. *IEEE Trans. Image Proc.* **2014**, *23*, 3773–3788. [CrossRef] [PubMed]
6. Ng, C.C.; Yap, M.H.; Costen, N.; Li, B. Automatic wrinkle detection using hybrid hessian filter. In Proceedings of the 12th Asian Conference on Computer Vision (ACCV 2014), Singapore, 1–5 November 2014; Revised Selected Papers, Part III 12. Springer International Publishing: Berlin/Heidelberg, Germany, 2015; pp. 609–622.
7. Xie, W.; Shen, L.; Jiang, J. A novel transient wrinkle detection algorithm and its application for expression synthesis. *IEEE Trans. Multimed.* **2016**, *19*, 279–292. [CrossRef]
8. Batool, N.; Chellappa, R. Fast detection of facial wrinkles based on Gabor features using image morphology and geometric constraints. *Pattern Recognit.* **2015**, *48*, 642–658. [CrossRef]
9. Elbashir, R.M.; Hoon Yap, M. Evaluation of automatic facial wrinkle detection algorithms. *J. Imaging* **2020**, *6*, 17. [CrossRef] [PubMed]
10. Molinara, M.; Cancelliere, R.; Di Tinno, A.; Ferrigno, L.; Shuba, M.; Kuzhir, P.; Maffucci, A.; Micheli, L. A Deep Learning Approach to Organic Pollutants Classification Using Voltammetry. *Sensors* **2022**, *22*, 8032. [CrossRef] [PubMed]
11. Jeyakumar, J.P.; Jude, A.; Priya Henry, A.G.; Hemanth, J. Comparative Analysis of Melanoma Classification Using Deep Learning Techniques on Dermoscopy Images. *Electronics* **2022**, *11*, 2918. [CrossRef]
12. Pintelas, E.; Livieris, I.E. XSC—An eXplainable Image Segmentation and Classification Framework: A Case Study on Skin Cancer. *Electronics* **2023**, *12*, 3551. [CrossRef]
13. Qin, H.; Deng, Z.; Shu, L.; Yin, Y.; Li, J.; Zhou, L.; Zeng, H.; Liang, Q. Portable Skin Lesion Segmentation System with Accurate Lesion Localization Based on Weakly Supervised Learning. *Electronics* **2023**, *12*, 3732. [CrossRef]

14. Wei, M.; Wu, Q.; Ji, H.; Wang, J.; Lyu, T.; Liu, J.; Zhao, L. A Skin Disease Classification Model Based on DenseNet and ConvNeXt Fusion. *Electronics* **2023**, *12*, 438. [CrossRef]
15. Chen, S.; Zhang, C.; Dong, M.; Le, J.; Rao, M. Using ranking-CNN for age estimation. In Proceedings of the IEEE 2017 Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 5183–5192.
16. Alarifi, J.S.; Goyal, M.; Davison, A.K.; Dancey, D.; Khan, R.; Yap, M.H. Facial skin classification using convolutional neural networks. In Proceedings of the 14th International Conference of Image Analysis and Recognition (ICIAR 2017), Montreal, QC, Canada, 5–7 July 2017; Springer International Publishing: Berlin/Heidelberg, Germany, 2017; pp. 479–485.
17. Sabina, U.; Whangbo, T.K. Nasolabial Wrinkle Segmentation Based on Nested Convolutional Neural Network. In Proceedings of the IEEE 2021 International Conference on Information and Communication Technology Convergence (ICTC), Jeju Island, Republic of Korea, 20–22 October 2021; pp. 483–485.
18. Deepa, H.; Gowrishankar, S.; Veena, A. A Deep Learning-Based Detection of Wrinkles on Skin. In Proceedings of the Computational Vision and Bio-Inspired Computing (ICCVBIC 2021), Online, 25–26 November 2021; Springer: Singapore, 2022; pp. 25–37.
19. Chang, T.R.; Tsai, M.Y. Classifying conditions of speckle and wrinkle on the human face: A deep learning approach. *Electronics* **2022**, *11*, 3623. [CrossRef]
20. Lv, M.; Zhou, G.; He, M.; Chen, A.; Zhang, W.; Hu, Y. Maize leaf disease identification based on feature enhancement and DMS-robust alexnet. *IEEE Access* **2020**, *8*, 57952–57966. [CrossRef]
21. Hevia-Montiel, N.; Haro, P.; Guillermo-Cordero, L. and Perez-Gonzalez, J. Deep Learning–Based Segmentation of Trypanosoma cruzi Nests in Histopathological Images. *Electronics* **2023**, *12*, 4144. [CrossRef]
22. You, Z.; Yu, H.; Xiao, Z.; Peng, T.; Wei, Y. CAS-UNet: A Retinal Segmentation Method Based on Attention. *Electronics* **2023**, *12*, 3359. [CrossRef]
23. Mehta, D.; Skliar, A.; Ben Yahia, H.; Borse, S.; Porikli, F.; Habibian, A.; Blankevoort, T. Simple and Efficient Architectures for Semantic Segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 2628–2636.
24. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid scene parsing network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2881–2890.
25. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 801–818.
26. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the 18th Medical Image Computing and Computer-Assisted Intervention International Conference (MICCAI 2015), Munich, Germany, 5–9 October 2015; Part III 18. Springer International Publishing: Berlin/Heidelberg, Germany, 2015; pp. 234–241.
27. Karras, T.; Laine, S.; Aila, T. A style-based generator architecture for generative adversarial networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2019; pp. 4401–4410.
28. Zhang, Z.; Song, Y.; Qi, H. Age progression/regression by conditional adversarial autoencoder. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 5810–5818.
29. Tan, M.; Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In Proceedings of the 2019 International Conference on Machine Learning (PMLR), Long Beach, CA, USA, 9–15 June 2019; pp. 6105–6114.
30. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE 2018 Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
31. Liu, S.; Huang, D.; Wang, Y. Learning spatial fusion for single-shot object detection. *arXiv* **2019**, arXiv:1911.09516.
32. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
33. Liu, T.; Zhang, L.; Zhou, G.; Cai, W.; Cai, C.; Li, L. BC-DUnet-based segmentation of fine cracks in bridges under a complex background. *PLoS ONE* **2022**, *17*, e0265258. [CrossRef] [PubMed]
34. Cao, H.; Gao, Y.; Cai, W.; Xu, Z.; Li, L. Segmentation Detection Method for Complex Road Cracks Collected by UAV Based on HC-Unet++. *Drones* **2023**, *7*, 189. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.