

Article

Research on Mobile Phone Backplane Defect Segmentation Based on MDAF-UNet

Hao Chen ^{1,2,*} and Byung-Won Min ²¹ School of Computer and Information Engineering, Nantong Institute of Technology, Nantong 226002, China² Division of Information and Communication Convergence Engineering, Mokwon University, Daejeon 35349, Republic of Korea

* Correspondence: chen hao@ntit.edu.cn

Abstract: Mobile phone backplanes are an important part of mobile phones, and are often affected by a wide range of factors during the manufacturing process, resulting in defects of various scales and similar backgrounds. Therefore, accurately identifying these defects is crucial for improving mobile phone quality. To address this challenge, this paper proposes a multi-scale and dynamic attention fusion UNet (MDAF-UNet) model. The model innovatively combines normal convolution with dilated convolution. This allows the model to capture subtle features of defects and to perceive a larger range of feature variations. Moreover, an improved attention mechanism is introduced in this paper. It fuses channel attention and spatial attention, and dynamically adjusts the feature fusion strategy with learnable weights. This allows the model to increase the attention of important features and improve the effectiveness of feature representation. Experimental results on a publicly available dataset show that the MDAF-UNet model achieves 66.9% Mean Intersection over Union (MIoU), outperforming other state-of-the-art models. This result provides an effective solution to the mobile phone backplane defect segmentation problem.

Keywords: mobile phone backplanes; MDAF-UNet; dilated convolution; attention mechanism



Citation: Chen, H.; Min, B.-W. Research on Mobile Phone Backplane Defect Segmentation Based on MDAF-UNet. *Electronics* **2024**, *13*, 1385. <https://doi.org/10.3390/electronics13071385>

Academic Editor: Mehrdad Saif

Received: 19 March 2024

Revised: 28 March 2024

Accepted: 4 April 2024

Published: 5 April 2024



Copyright: © 2024 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

Product quality assurance is a critical component of industrial quality control. Efficient and accurate defect detection is essential for product quality control [1,2]. In mobile phone manufacturing, the appearance and quality standards of mobile phone backplanes are extremely high. Once a tiny defect is found, the product may be judged as substandard. During quality control, in addition to identifying the defects, the shape and location of the defects need to be accurately labeled. This can provide detailed data support for subsequent quality control and assessment [3]. As a result, how to accurately detect defects in mobile phone backplanes is a technical issue that must be addressed as soon as possible.

The field of industrial quality control relies heavily on traditional image processing techniques for defect segmentation. These methods include the threshold method [4], edge detection technique [5], and the watershed algorithm [6]. These methods are effective under specific conditions. However, the defects of mobile phone backplanes have variable sizes and similar backgrounds, which pose great challenges for defect segmentation. Traditional segmentation methods are ineffective in dealing with these complex situations, and it is difficult to meet the requirements of high-precision quality inspection.

With the development of deep learning technology, Convolutional Neural Networks (CNNs) have shown great potential in the field of image processing [7,8]. UNet [9] is designed based on CNNs, further enhancing the capability of CNNs in extracting features from images. UNet effectively captures the contextual information of the image through the unique symmetric structure and jump connections, which significantly improves the accuracy of segmentation. However, considering the diversity and complexity of mobile phone backplane defects, it is still challenging to use UNet directly for mobile phone

backplane defect segmentation. Many researchers have improved UNet to adapt to different application scenarios. This includes the introduction of dilated convolution [10] and attention mechanism [11,12]. In CNNs, smaller convolutional kernels excel at capturing local details of an image, while larger convolutional kernels are adept at extracting global features. Dilated convolution is a special type of convolution that effectively expands the receptive field of the convolutional kernel to help the model capture a broader range of feature information. By integrating various sizes and types of convolutional kernels, CNNs can extract multi-scale features spanning from local to global, thereby achieving a more comprehensive understanding of the image. Moreover, the attention mechanism enables the model to prioritize critical regions in the image while disregarding irrelevant areas during feature extraction. Therefore, by leveraging multi-scale information and the attention mechanism, it provides an effective solution for detecting defects in mobile phone backplanes.

In summary, considering the scale diversity of mobile phone backplane defects, the model needs to flexibly cope with the feature information of different scales. Meanwhile, in terms of the subtle differences and background similarity of mobile phone backplane defects, the model also needs to have the ability to capture these subtle difference variations. To address these challenges, this paper presents two key improvements based on UNet: the first is to strengthen the multi-scale feature extraction capability to enhance the recognition of defects at different scales; the second is to introduce the attention mechanism to enhance the model's differentiation of subtle differences. Therefore, this paper proposes a multi-scale and dynamic attention fusion UNet (MDAF-UNet) model. The main contributions of this paper are as follows:

- (1) An innovative multi-scale fusion technique is proposed, which can effectively recognize defects at different scales. The technique utilizes normal and dilated convolutions to enable the model to capture not only the subtle features of defects, but also a wider range of feature variations.
- (2) An improved attention mechanism is proposed. A fusion module is introduced after channel attention and spatial attention to generate the final attention feature map. Additionally, by introducing learnable weights, the model combines the original features with the fused attention features dynamically, further enhancing the feature representation.
- (3) The proposed model has been validated on a public dataset, and the results show that the improved module enhances UNet's segmentation effect. Compared with the existing algorithms, the MDAF-UNet model demonstrates significant advantages in performance.

The rest of the paper is organized as follows: Section 2 describes related work; Section 3 describes the structure and improved modules of MDAF-UNet; the experimental setup and evaluation indicators are described in Section 4; Section 5 concludes with a detailed description of the experiments and results; Section 6 provides a summary.

2. Related Work

Due to the complexity and variability of object surfaces, defect segmentation poses significant challenges. In the realm of industrial quality inspection, traditional image processing techniques are frequently employed for such tasks. Cao et al. [13] employed a gradient threshold segmentation method, which effectively mitigates segmentation errors caused by uneven illumination and substantially enhances segmentation accuracy on complex object surfaces. Yang et al. [14] combined a supervised multi-threshold segmentation model with the Canny edge detector to effectively recognize similar features on the surface of the target object that are otherwise difficult to differentiate. Meiju et al. [15] proposed a two-dimensional Otsu segmentation algorithm for small defects in mobile phone screens, which achieves accurate segmentation of target and background. While these methods have achieved significant results in specific scenarios, more refined detection strategies

are still required to address the complex situation of mobile phone backplane defects with varying scales and similar backgrounds.

In recent years, researchers have begun to explore the use of CNN-based network models for defect segmentation to improve segmentation accuracy [16–18]. Song et al. [19] developed a U-Net-based surface defect detection technique for identifying surface flaws commonly found in industrial production products, offering robust support for enhancing the quality of industrial production. Jiang et al. [20] further applied UNet to mobile phone backplane defect detection and achieved better results. The primary function of dilated convolution is to expand the receptive field in order to better capture contextual information in the image. This is important when dealing with mobile phone backplane defects, which often contain different scales. Mao et al. [21] introduced dilated convolution and jump connection to effectively extract the feature information of mobile phone screen defects, thus realizing efficient classification. Pan et al. [22] introduced dilated convolution in UNet and verified its superiority in a real mobile phone surface defects task. Considering the subtle differences in the defects of the mobile phone's backplane and their similarity to the background, the model must have the ability to accurately capture these nuances. In this regard, the attention mechanism plays a key role, enabling the model to focus on both local and global information, automatically highlighting critical regions where defects exist [23,24]. Guo et al. [25] introduced the spatial attention mechanism into U-Net, significantly improving the accuracy of the segmentation task. Lu et al. [26] effectively addressed the challenge of diverse surface defects across various products by integrating multiscale features and an attention mechanism. Zhu et al. [27] designed an attention mechanism that combines multi-frequency information and local cross-channel interaction to better represent and emphasize defect features. Therefore, the comprehensive utilization of multi-scale information features and attention mechanisms can effectively enhance the defect segmentation accuracy of mobile phone backplanes.

In summary, the challenges of scale diversity and background similarity exist in the defect detection task of mobile phone backplanes. By improving the model's multi-scale feature extraction capability and ability to differentiate subtle differences, this paper provides an effective solution to the defect detection problem in mobile phone backplane manufacturing.

3. Methodology

3.1. Model Structure

UNet is a widely used image segmentation model consisting primarily of an encoder and a decoder. The encoder gathers contextual information from the image, while the decoder focuses on accurate localization. To effectively enhance the segmentation accuracy of UNet for defects on mobile phone backplanes, this paper introduces a multi-scale fusion module and an attention mechanism module based on UNet. The multi-scale fusion module enhances the network's ability to capture features at different scales, while the improved attention mechanism module, incorporating dynamic weights for the dynamic adjustment of attention features, enhances the focus on defective regions.

Figure 1 depicts the structure of the MDAF-UNet proposed in this paper. Blue arrows represent the use of 3×3 convolutions followed by Rectified Linear Unit (ReLU) activation function to extract image features. Yellow arrows denote the process of multi-scale feature fusion, which combines feature maps from different scales for comprehensive information capture. Red arrows signify 2×2 max pooling, utilized to reduce the spatial dimensions of feature maps. Grey arrows illustrate the copy and crop operation, utilized to achieve fusion of the corresponding feature layers. Green arrows indicate 2×2 deconvolutions, employed to enlarge the feature maps' dimensions. Orange arrows indicate that an attentional mechanism is employed, allowing the model to focus on key regions of the image. Light blue arrows indicate linear interpolation up-sampling. On this basis, firstly, five initial effective feature layers are obtained using VGG16 as the backbone feature extraction network. After each effective feature layer, a multi-scale fusion module is introduced. Subsequently, feature fusion is performed on the five feature layers. The feature fusion is

performed by up-sampling and stacking the feature layers. In this process, an improved attention mechanism is introduced after each feature fusion. Finally, the prediction results are derived based on the input image features.

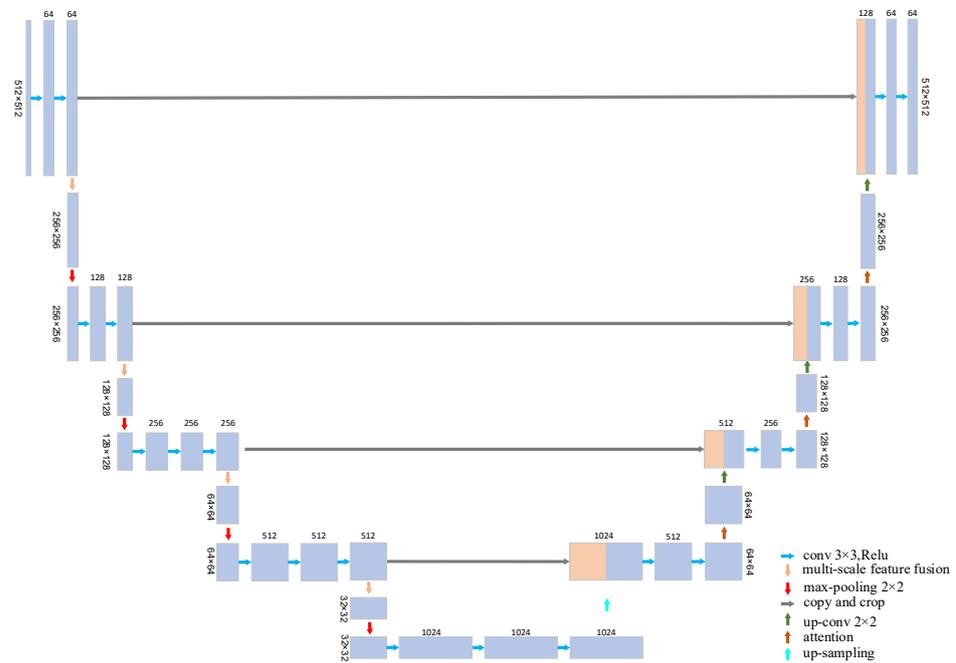


Figure 1. MDAF-UNet structure.

3.2. Multi-Scale Fusion Module

Features at different scales are essential for accurate segmentation of defects in mobile phone backplanes. Small scale features provide detailed information about the image, while large scale features offer global information. The size and shape of mobile phone backplane defects vary greatly at various scales. The introduction of a multi-scale fusion module can capture and adapt to defects at different scales and enhance the feature representation capability of the model. Normal convolution primarily captures localized features, while dilated convolution enlarges the receptive field without increasing parameters, capturing a broader range of contextual information. By combining normal and dilated convolution, the model can perceive a wider range of feature changes while preserving detailed information.

In this paper, the multi-scale feature fusion module is added after the UNet backbone feature extraction network. This approach further enhances the feature representation. Additionally, the backbone feature extraction network helps to extract high-level features by gradually reducing the spatial size of the feature map. The introduction of the multi-scale fusion module enables the fusion of features at different scales without significant loss of spatial information.

Figure 2 illustrates the multi-scale fusion structure. Suppose the input feature map is defined as $F_{in} \in R^{C \times W \times H}$, where C denotes the number of channels, W denotes the width of the feature map, and H denotes the height of the feature map. The output after normal and dilated convolution [10] can be expressed as follows:

$$F' = Conv(F_{in}, \omega, 3), \tag{1}$$

$$F'' = Conv(F_{in}, \omega, 5), \tag{2}$$

$$F''' = Conv(F_{in}, \omega, 3, d), \tag{3}$$

where F' denotes the output feature map after convolution kernel (size 3×3), F'' denotes the output feature map after convolution kernel (size 5×5), F''' denotes the output feature map after dilated convolution, ω denotes the weight of the convolution kernel, and d

denotes the coefficient of the dilated convolution, and here d is equal to 2. The fused features of the three feature maps can be represented as follows:

$$F_m = \text{Concat}(F', F'', F''', \text{dim} = 1), \tag{4}$$

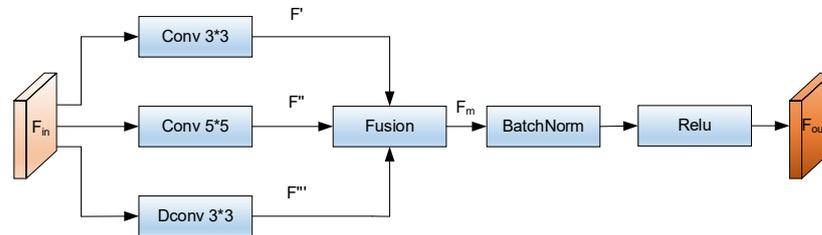


Figure 2. Multi-scale fusion structure.

Next, each channel of F_m is normalized using batch normalization, and then the feature map is scaled and offset. Finally, the multi-scale fused feature map is obtained after *ReLU* activation function. The feature map can be represented as follows:

$$F_{out} = \text{Relu}(\alpha_c(\frac{F_s - \mu_c}{\sqrt{\epsilon_c^2 + \sigma}}) + \beta_c), \tag{5}$$

where μ_c and ϵ_c^2 represent the mean and variance of the channel in the current batch, respectively. α_c and β_c are learnable parameters, independent for each channel, and σ is a smaller constant that prevents the denominator from being zero.

3.3. Attention Mechanism Module

The introduction of the attention mechanism enables the model to automatically focus on the critical areas where the defects are located. This means that the introduction of the attention mechanism can more accurately recognize the defective regions, thus improving the recognition accuracy. However, in mobile phone backplane defect detection, since different defects are difficult to distinguish and have similar backgrounds, it requires finer attention tuning to capture these nuances. The attention to features can be effectively tuned by combining the outputs of channel attention and spatial attention. More detailed feature enhancement is then achieved through additional convolutional processing. As a result, after channel and spatial attention, a fusion module is introduced, which combines the outputs of the two types of attention to produce the final attention feature map. At the decoder stage, an improved attention mechanism is introduced in this paper. Up-sampling is required in the decoder stage to recover the image’s detailed information. The attention mechanism helps the model to process the detailed information meticulously, which is essential for accurate segmentation of similar defects. Additionally, the improved attention mechanism helps the model to better determine which features are useful and which should be suppressed when fusing different levels of features during the up-sampling process. This helps generate more accurate segmentation results.

The structure of the improved attention mechanism is shown in Figure 3, where *Maxpool* denotes max pooling, and *Avgpool* denotes average pooling. Assuming the input feature is defined as f_{in} , after the Channel Attention Mechanism (CAM) [24] can be represented as follows:

$$f_c = f_{in} \cdot \sigma(C_{avg}(f_{in}) + C_{max}(f_{in})), \tag{6}$$

where σ denotes the sigmoid function, C_{avg} represents the result of the feature map after average pooling processed by the fully connected layer, while C_{max} represents the result of the feature map after max pooling processed by the fully connected layer. The feature map obtained after Spatial Attention Mechanism (SAM) can be represented as follows:

$$f_s = f_{in} \cdot \sigma(\text{Conv}(\text{Cat}(\text{Mean}(f_{in}), \text{Max}(f_{in})))), \tag{7}$$

where *Cat* denotes the splicing along the channel direction, *Mean* denotes the average value of the input feature map, and *Max* denotes the maximum value of the input feature map. Following attention fusion, the feature map can be represented as follows:

$$f'_{cs} = \sigma(\text{Conv}(\text{Relu}(\text{Conv}(\text{Cat}(f_c + f_s))))), \tag{8}$$

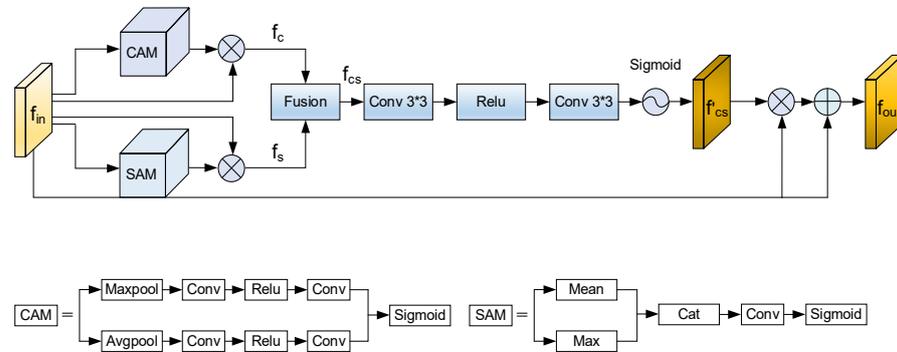


Figure 3. Attention mechanism structure.

The final enhanced feature map obtained by adaptive feature fusion computation is represented as follows:

$$f_{out} = \alpha \cdot f_{in} + \beta \cdot (f_{in} \cdot f'_{cs}), \tag{9}$$

where α and β denote the learnable weights after Softmax function. The learnable weights allow us to dynamically adjust the method of combining the original features with the fused attention features.

4. Experimental Setup and Evaluation Indicators

4.1. Dataset

This paper selects a defective segmentation dataset for industrial quality control of mobile phone backplanes. Blemish, corner wear, and crack are the three types of defective targets covered in the dataset, in addition to backgrounds. The dataset contains 864 high-definition images, which provides rich data resources for the quality inspection of mobile phone backplanes. The dataset is scientifically divided in the experiment. The training set for model training is made up of 777 images, while the validation set for model validation and evaluation is made up of 87 images. The dataset’s partial samples are shown in Figure 4. Additionally, this paper adopts the method of 10-fold cross-validation [28] to further enhance the reliability of the model’s results.

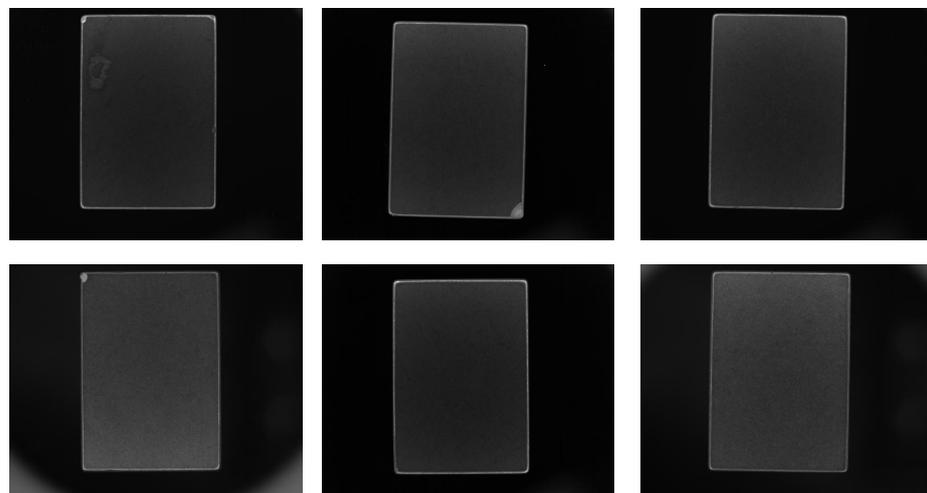


Figure 4. Partial samples.

4.2. Experiment Details

Experiment hardware environment: CPU is Intel Core i7 (Intel Corporation, Santa Clara, CA, USA); GPU is NVIDIA GeForce RTX 3050 Ti (with 12 GB of video memory) (NVIDIA Corporation, Santa Clara, CA, USA). Experiment software environment: the operating system is Windows 10 (Microsoft, Redmond, WA, USA); deep learning framework is PyTorch; the programming language is Python 3.8; CUDA version is 11.6.

This paper's parameters are as follows: the image input size is 512×512 ; the optimizer is Adam; the momentum is 0.9; the initial learning rate is 0.0001; the learning rate decreasing mode is Cosine Annealing; the epoch is 400; the batch size is 8.

4.3. Evaluation Indicators

In this paper, the following three key metrics are chosen to comprehensively evaluate the model's performance in the mobile phone backplane defect segmentation task: precision, recall, and Mean Intersection Over Union (MIoU). These metrics help understand the model's ability to recognize defects from various perspectives and ensure the model's validity and reliability in practical applications. It is assumed that True Positives (TP) denotes the number of correctly predicted positive samples; True Negatives (TN) denotes the number of correctly predicted negative samples; False Positives (FP) denotes the number of negative samples incorrectly predicted as positive samples; and False Negatives (FN) denotes the positive samples incorrectly predicted as negative samples number of positive samples that are incorrectly predicted as negative samples. The formula for calculating precision is as follows:

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (10)$$

Recall is a measure of the model's ability to identify all positive samples, and it focuses on how well the model covers the positive samples. The formula for calculating recall is as follows:

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (11)$$

Intersection Over Union (IoU) is a key metric for evaluating model performance in segmentation tasks, measuring how much the predicted segmentation region overlaps with the actual segmentation region. The formula for calculating Intersection Over Union (IoU) is as follows:

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}}, \quad (12)$$

MIoU is averaged over all IoU categories to assess the model's overall performance in a multi-category segmentation task. The formula for calculating MIoU is as follows:

$$\text{MIoU} = \frac{1}{N} \sum_{i=1}^N \text{IoU}_i, \quad (13)$$

where N represents the number of categories. MDAF-UNet's performance on the task of mobile phone backplane defect segmentation can be comprehensively evaluated using these metrics.

5. Experiments and Results

5.1. Backbone Network Comparison Experiments

Table 1 illustrates the experimental results of VGG16 and ResNet50 as UNet backbone networks, respectively. When used as the backbone network, it has been discovered that VGG16 produces better overall results. Specifically, the MIoU of VGG16 as the backbone network is 4.8% higher than that of ResNet50, indicating that it has a higher degree of overlap between recognized defective regions and real defective regions. Its precision is higher by 8.8%, indicating that VGG16 incorrectly labels fewer regions recognized as

defective. In terms of recall, VGG16 is higher by 4.6% compared to ResNet50, indicating its ability to identify all real defective regions more comprehensively.

Table 1. Backbone network comparison results.

Model	Module		MIoU (%)	Precision (%)	Recall (%)
	VGG16	ResNet50			
UNet	✓		63.7	78.2	70.0
		✓	58.9	69.4	65.4

Figure 5 displays the IoU of the three targets in different backbone networks. The results show that the choice of backbone network has a significant impact on the IoU performance in the segmentation task. When VGG16 is selected as the backbone network, the IoU of all three targets are higher than the corresponding IoU when ResNet50 is used as the backbone network.

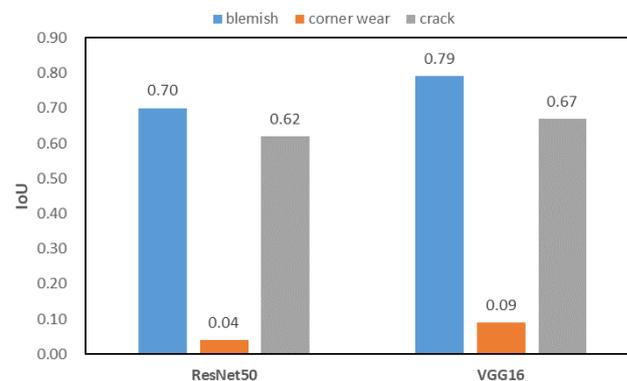


Figure 5. IoU in different backbone networks.

Figure 6 depicts the segmentation results of the two backbone networks, VGG16 and ResNet50, in real-world applications. However, ResNet50 exhibits limitations when dealing with two types of complex defects: cracks, and corner wear. Specifically, it incorrectly segments some cracked regions into corner wear, leading to misclassification of defect types. Furthermore, there are instances where cracks are not detected, significantly reducing the segmentation's completeness and accuracy. In comparison, VGG16 not only accurately recognizes and segments the crack region, but also avoids misclassification and missed detection. It is worth mentioning that both VGG16 and ResNet50 perform relatively well in blemish segmentation. This is due to the more obvious features of blemish, which allow both models to handle it better.

5.2. Ablation Experiments

Ablation experiments are carried out in this paper to investigate the effect of adding multi-scale fusion and attention mechanisms to UNet on image segmentation performance. Table 2 summarizes the results of the ablation experiments. After either multi-scale fusion or the attention mechanism is introduced, the model's performance improves to some extent. Specifically, the MIoU of the model is improved by 2.4% after the introduction of multi-scale fusion, and by 1.8% after the introduction of the attention mechanism. Furthermore, when multi-scale fusion and the attention mechanism are introduced, the model's MIoU increases by 3.2%. Additionally, precision and recall have significantly improved. These results suggest that combining multi-scale fusion and the attention mechanism in UNet can further improve the image segmentation effect.

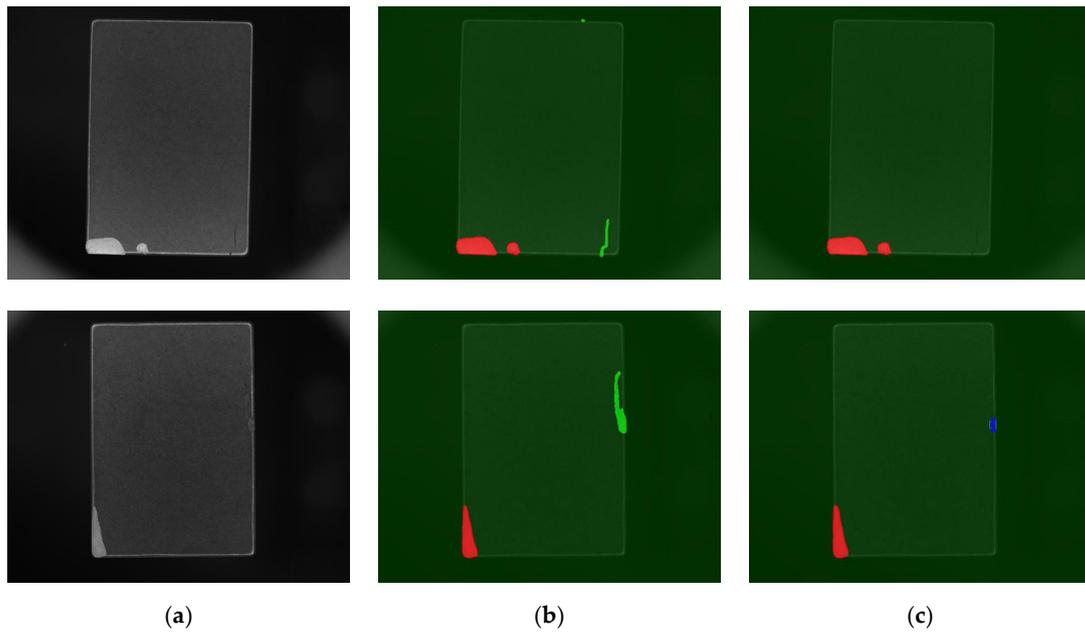


Figure 6. Segmentation results of VGG16 and ResNet50. (a) Original figure; (b) VGG16; (c) ResNet50.

Table 2. Results of ablation experiments.

Model	Module		MIoU (%)	Precision (%)	Recall (%)
	Multi-Scale	Attention			
UNet			63.7	78.2	70.0
	✓		66.1	77.9	73.4
		✓	65.5	76.8	74.0
	✓	✓	66.9	78.6	74.3

Figure 7 demonstrates the IoU of different targets in the ablation experiments. The results show that the IoU of each type of target has slightly improved since the introduction of multi-scale fusion or attention mechanism.

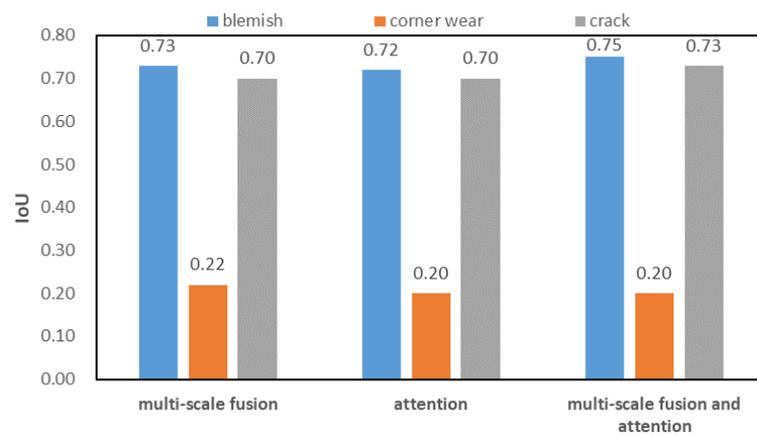


Figure 7. IoU in ablation experiments.

This paper employs UNet and MDAF-UNet for the segmentation of actual mobile phone backplane defects. Figure 8 illustrates the actual segmentation results. Particularly, both models exhibit similar accuracy in segmentation of blemish defects. However, MDAF-UNet performs more accurately in the segmentation task of crack defects. This suggests that MDAF-UNet effectively captures the detailed information of crack defects, and thus

the introduction of the attention mechanism and the multi-scale fusion module enhances the model's ability to perceive defects and segmentation accuracy.

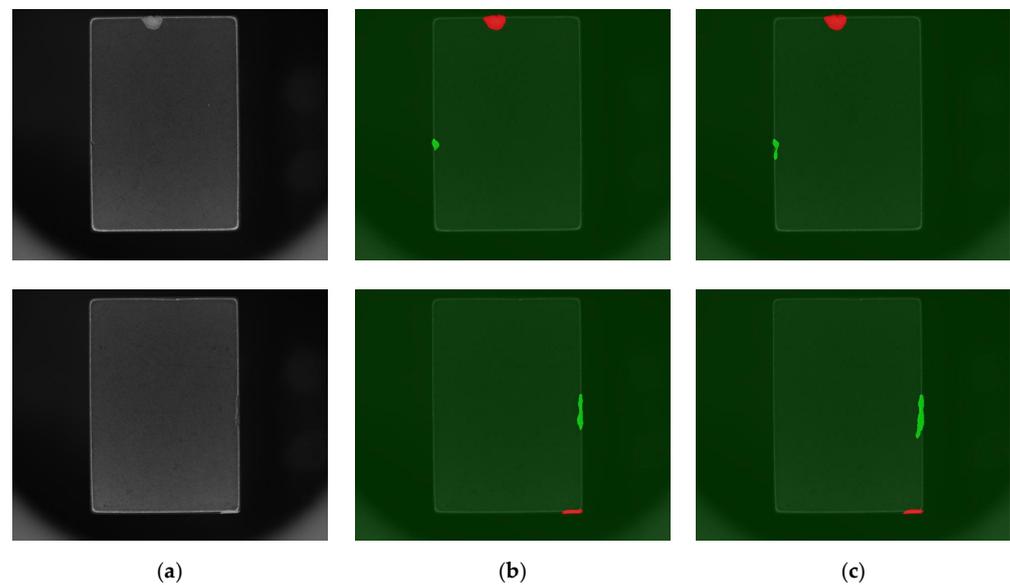


Figure 8. Segmentation results of UNet and MDAF-UNet. (a) Original figure; (b) UNet; (c) MDAF-UNet.

5.3. Comparison Experiments

In this paper, PSPNet [29] and UNet-CBAM are selected for experimental comparison with MDAF-UNet. The comparison experiments' results are shown in Table 3, indicating that MDAF-UNet achieves the best segmentation performance. The MIoU is improved by 7.5% relative to PSPNet and 2.2% relative to UNet-CBAM. Figure 9 illustrates the segmentation results for various targets including blemish, corner wear, and crack. Specifically, MDAF-UNet shows the best performance in the segmentation tasks of blemish defects and crack defects, with IoU of 75% and 73%, respectively, surpassing those of PSPNet and UNet-CBAM. These results indicate that MDAF-UNet achieves better performance in the defect segmentation task of the mobile phone backplane.

Table 3. Comparison of ablation experiments.

Model	MIoU (%)	Precision (%)	Recall (%)
PSPNet [29]	59.4	76.9	72.2
UNet-CBAM	64.7	75.0	72.7
MDAF-UNet	66.9	78.6	74.3

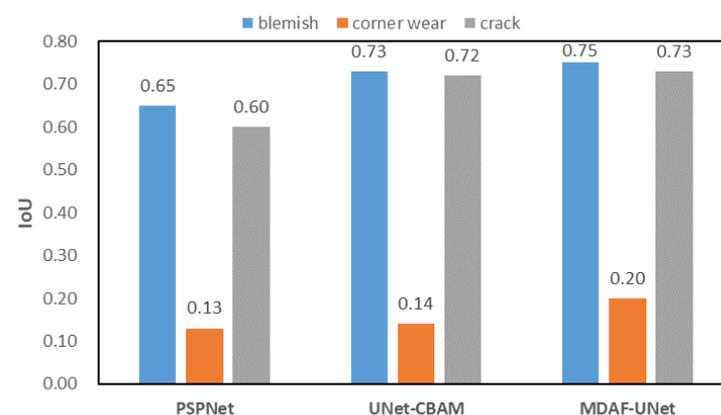


Figure 9. IoU in comparison experiments.

The actual segmentation results of the compared models are shown in Figure 10. For the defect detection task, MDAF-UNet performs the best in terms of segmentation. MDAF-UNet is able to effectively fuse multi-scale features, which enables the model to more comprehensively perceive the characteristics of defects at different scales. This enables MDAF-UNet to segment the blemish region more accurately and avoids mis-segmentation. In contrast, PSPNet made some errors in blemish segmentation and was unable to effectively capture the subtle features of blemish, resulting in some blemish being incorrectly categorized as corner wear. Although the UNet-CBAM model performed well in general, there were some instances where it was unable to handle crack segmentation completely correctly. This is because, while CBAM can improve the model's focus on important features, it is still insufficient for capturing subtle.

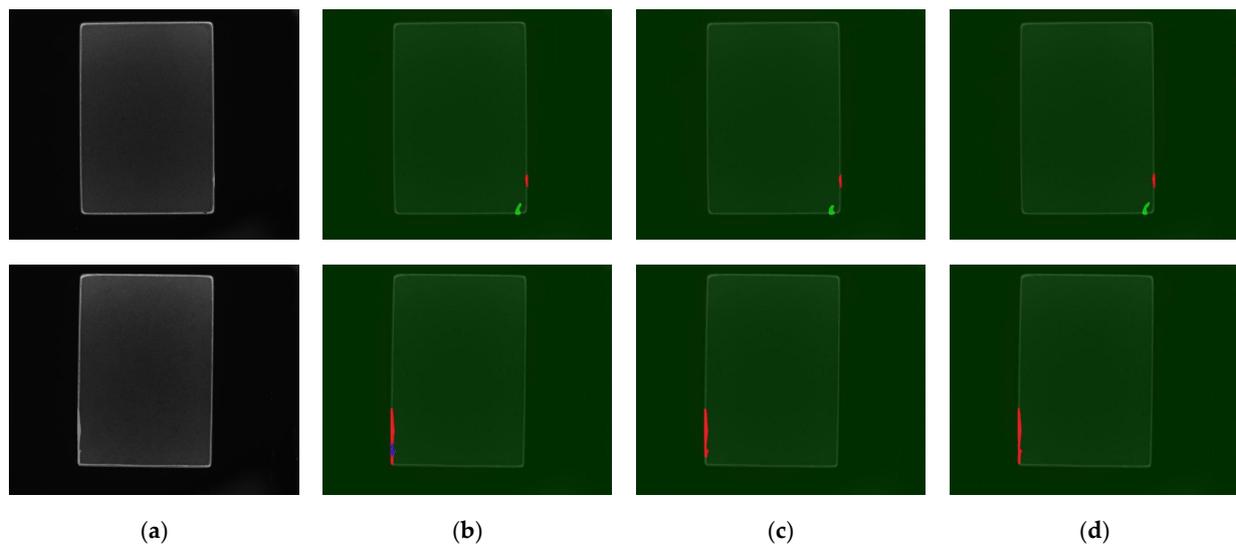


Figure 10. Segmentation results for comparison models. (a) Original figure; (b) PSPNet; (c) UNet-CBAM; (d) MDAF-UNet.

5.4. Discussion

VGG16 increases network depth by stacking multiple smaller convolutional kernels to capture more detailed information. This design allows VGG16 to extract image features in much greater detail, especially when dealing with images with multi-scale defects, such as mobile phone backplanes. Secondly, ResNet50 addresses the issue of gradient vanishing by introducing residual connections. However, in some cases, an excessive number of residual connections may cause the network to become overly complex, hindering the learning of effective feature representations. In the mobile phone backplane defect segmentation task, ResNet50 fails to adequately learn the features specific to defects such as cracks and corner abrasion, leading to mis-classification and missed detection.

UNet is a classical semantic segmentation network widely used in the medical and industrial sectors. However, when dealing with complex scenes or fine features, UNet may not perform well. The results of the ablation experiments demonstrate that the introduction of multi-scale fusion and attention mechanisms has a positive impact on model performance. Firstly, the introduction of the multi-scale fusion module helps the model to capture image features more comprehensively at different scales. Defects on the backplane of a mobile phone often have different scales and shapes, so the multi-scale fusion module can help the model better segment these defects. Secondly, the introduction of the attention mechanism allows the model to focus on critical regions. In the mobile phone backplane defect segmentation task, the attention mechanism helps the model concentrate its attention on regions where defects may exist, reducing the processing of irrelevant information and improving the segmentation accuracy.

PSPNet adopts a global pooling strategy to capture global contextual information, but it may not efficiently handle fine blemish features, resulting in some blemishes being incorrectly segmented as corner wear. UNet-CBAM introduces CBAM to improve the critical target region attention when segmenting mobile phone backplane defects. Its performance is superior to that of UNet. However, due to the model's inability to perceive subtle features, some corner wear and cracks are not properly segmented. The enhancement of the attention mechanism in MDAF-UNet further optimizes the model's attention to critical regions compared to CBAM. Through the fusion of channel attention and spatial attention, the model can locate the defective regions more accurately. Furthermore, using learnable weights to dynamically adjust the way original features are combined with fused attention features improves feature representation. Moreover, the multi-scale fusion module enables the model to comprehensively utilize feature information at different scales, capturing the detailed features more effectively.

In summary, MDAF-UNet can segment defects on mobile phone backplanes more accurately by integrating multi-scale information and optimizing the attention mechanism. However, although MDAF-UNet shows significant advantages in segmenting defects on mobile phone backplanes, it still has some limitations. The model's validation is limited to a single dataset and lacks validation on multiple datasets, which may result in an insufficient assessment of its generalization ability.

6. Conclusions

Aiming to address the challenges of scale variation and background similarity in mobile phone backplane defect segmentation, an MDAF-UNet model is proposed in this paper. The model utilizes a combination of multi-scale fusion and attention mechanism, demonstrating significant advantages in dealing with defect scale variations and background interference. By integrating normal convolution and dilated convolution, defects at different scales can be effectively captured. Meanwhile, channel attention and spatial attention are introduced, and the fusion of attention features is dynamically adjusted by learnable weights. The experimental results validate the superiority of the MDAF-UNet model on the mobile phone backplane defect dataset, achieving a segmentation effectiveness of 66.9% in terms of MIoU.

In the future, this paper aims to further enhance and validate the MDAF-UNet model to ensure its effectiveness in practical applications through a wider range of datasets and a more rigorous experimental design.

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

Funding: This research received no external funding.

Data Availability Statement: The data that support the findings of this study are available in [AI Studio] at <https://aistudio.baidu.com/datasetdetail/179222> (accessed on 1 March 2024).

Acknowledgments: The authors are thankful to all the personnel who either provided technical support or helped with data collection. We also acknowledge all the reviewers for their useful comments and suggestions.

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

References

1. Yang, J.; Li, S.; Wang, Z.; Dong, H.; Wang, J.; Tang, S. Using deep learning to detect defects in manufacturing: A comprehensive survey and current challenges. *Materials* **2020**, *13*, 5755. [CrossRef] [PubMed]
2. Chen, Y.; Ding, Y.; Zhao, F.; Zhang, E.; Wu, Z.; Shao, L. Surface defect detection methods for industrial products: A review. *Appl. Sci.* **2021**, *11*, 7657. [CrossRef]

3. Tabernik, D.; Šela, S.; Skvarč, J.; Skočaj, D. Segmentation-based deep-learning approach for surface-defect detection. *J. Intell. Manuf.* **2020**, *31*, 759–776. [[CrossRef](#)]
4. Liang, Y.; Xu, K.; Zhou, P. Mask gradient response-based threshold segmentation for surface defect detection of milled aluminum ingot. *Sensors* **2020**, *20*, 4519. [[CrossRef](#)] [[PubMed](#)]
5. Ni, X.; Liu, H.; Wang, C.; Liu, J. Detection for rail surface defects via partitioned edge feature. *IEEE Trans. Intell. Transp. Syst.* **2021**, *23*, 5806–5822. [[CrossRef](#)]
6. Guo, Q.; Wang, Y.; Yang, S.; Xiang, Z. A method of blasted rock image segmentation based on improved watershed algorithm. *Sci. Rep.* **2022**, *12*, 7143. [[CrossRef](#)] [[PubMed](#)]
7. Qin, J.; Pan, W.; Xiang, X.; Tan, Y.; Hou, G. A biological image classification method based on improved CNN. *Ecol. Informa.* **2020**, *58*, 101093. [[CrossRef](#)]
8. Sultana, F.; Sufian, A.; Dutta, P. Evolution of image segmentation using deep convolutional neural network: A survey. *Knowl.-Based Syst.* **2020**, *201*, 106062.
9. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241.
10. Sun, X.; Li, J.; Ma, J.; Xu, H.; Chen, B.; Zhang, Y.; Feng, T. Segmentation of overlapping chromosome images using U-Net with improved dilated convolutions. *J. Intell. Fuzzy Syst.* **2021**, *40*, 5653–5668.
11. Wu, H.; Zhao, Z.; Wang, Z. META-Unet: Multi-scale efficient transformer attention Unet for fast and high-accuracy polyp segmentation. *IEEE Trans. Autom. Sci. Eng.* **2023**. [[CrossRef](#)]
12. Maji, D.; Sigidar, P.; Singh, M. Attention Res-UNet with Guided Decoder for semantic segmentation of brain tumors. *Biomed. Signal Process. Control* **2022**, *71*, 103077. [[CrossRef](#)]
13. Cao, G.; Ruan, S.; Peng, Y.; Huang, S.; Kwok, N. Large-complex-surface defect detection by hybrid gradient threshold segmentation and image registration. *IEEE Access* **2018**, *6*, 36235–36246. [[CrossRef](#)]
14. Yang, Y.; Zhao, X.; Huang, M.; Wang, X.; Zhu, Q. Multispectral image based germination detection of potato by using supervised multiple threshold segmentation model and Canny edge detector. *Comput. Electron. Agric.* **2021**, *182*, 106041. [[CrossRef](#)]
15. Meiju, L.; Rui, Z.; Xifeng, G.; Junrui, Z. Application of improved Otsu threshold segmentation algorithm in mobile phone screen defect detection. In Proceedings of the 2020 Chinese Control and Decision Conference CCDC, Hefei, China, 22–24 August 2020; pp. 4919–4924.
16. Xu, Y.; Li, D.; Xie, Q.; Wu, Q.; Wang, J. Automatic defect detection and segmentation of tunnel surface using modified Mask R-CNN. *Measurement* **2021**, *178*, 109316. [[CrossRef](#)]
17. Cao, J.; Yang, G.; Yang, X. A pixel-level segmentation convolutional neural network based on deep feature fusion for surface defect detection. *IEEE Trans. Instrum. Meas.* **2020**, *70*, 1–12. [[CrossRef](#)]
18. Li, Y.; Li, J. An end-to-end defect detection method for mobile phone light guide plate via multitask learning. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–13. [[CrossRef](#)]
19. Song, Y.; Xia, W.; Li, Y.; Li, H.; Yuan, M.; Zhang, Q. AnomalySeg: Deep Learning-Based Fast Anomaly Segmentation Approach for Surface Defect Detection. *Electronics* **2024**, *13*, 284. [[CrossRef](#)]
20. Jiang, J.; Cao, P.; Lu, Z.; Lou, W.; Yang, Y. Surface defect detection for mobile phone back glass based on symmetric convolutional neural network deep learning. *Appl. Sci.* **2020**, *10*, 3621. [[CrossRef](#)]
21. Mao, J.; Xu, G.; He, L.; Luo, J. Attention-relation network for mobile phone screen defect classification via a few samples. *Digit. Commun. Netw.* **2023**. In Press. [[CrossRef](#)]
22. Pan, J.; Zeng, D.; Tan, Q.; Wu, Z.; Ren, Z. EU-Net: A novel semantic segmentation architecture for surface defect detection of mobile phone screens. *IET Image Process.* **2022**, *16*, 2568–2576. [[CrossRef](#)]
23. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7132–7141.
24. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
25. Guo, C.; Szemenyei, M.; Yi, Y.; Wang, W.; Chen, B.; Fan, C. SA-UNet: Spatial attention U-Net for retinal vessel segmentation. In Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 10–15 January 2021.
26. Lu, P.; Jing, J.; Huang, Y. MRD-net: An effective CNN-based segmentation network for surface defect detection. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 1–12. [[CrossRef](#)]
27. Zhu, Y.; Ding, R.; Huang, W.; Wei, P.; Yang, G.; Wang, Y. HMFCA-Net: Hierarchical multi-frequency based Channel attention net for mobile phone surface defect detection. *Pattern Recognit. Lett.* **2022**, *153*, 118–125. [[CrossRef](#)]

28. Moreno-Torres, J.G.; Sáez, J.A.; Herrera, F. Study on the impact of partition-induced dataset shift on k -fold cross-validation. *IEEE Trans. Neural Netw. Learn. Syst.* **2012**, *23*, 1304–1312. [[CrossRef](#)] [[PubMed](#)]
29. 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.

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.