




Article

Real-Time ConvNext-Based U-Net with Feature Infusion for Egg Microcrack Detection

Chenbo Shi , Yuejia Li, Xin Jiang, Wenxin Sun, Changsheng Zhu , Yuanzheng Mo , Shaojia Yan and Chun Zhang *

College of Intelligent Equipment, Shandong University of Science and Technology, Tai'an 271019, China; skd996523@sdust.edu.cn (C.S.); 202283230037@sdust.edu.cn (Y.L.); jx1093560806@163.com (X.J.); sunwenx2001@163.com (W.S.); cs.zhu@sdust.edu.cn (C.Z.); 202283230023@sdust.edu.cn (Y.M.); 202383230010@sdust.edu.cn (S.Y.)

* Correspondence: zhangchun@sdust.edu.cn

Abstract: Real-time automatic detection of microcracks in eggs is crucial for ensuring egg quality and safety, yet rapid detection of micron-scale cracks remains challenging. This study introduces a real-time ConvNext-Based U-Net model with Feature Infusion (CBU-FI Net) for egg microcrack detection. Leveraging edge features and spatial continuity of cracks, we incorporate an edge feature infusion module in the encoder and design a multi-scale feature aggregation strategy in the decoder to enhance the extraction of both local details and global semantic information. By introducing large convolution kernels and depth-wise separable convolution from ConvNext, the model significantly reduces network parameters compared to the original U-Net. Additionally, a composite loss function is devised to address class imbalance issues. Experimental results on a dataset comprising over 3400 graded egg microcrack image patches demonstrate that CBU-FI Net achieves a reduction in parameters to one-third the amount in the original U-Net, with an inference speed of 21 ms per image (1 million pixels). The model achieves a Crack-IoU of 65.51% for microcracks smaller than 20 μm and a Crack-IoU and MIoU of 60.76% and 80.22%, respectively, for even smaller cracks (less than 5 μm), achieving high-precision, real-time detection of egg microcracks. Furthermore, on the publicly benchmarked CrackSeg9k dataset, CBU-FI Net achieves an inference speed of 4 ms for 400×400 resolution images, with an MIoU of 81.38%, proving the proposed method's robustness and generalization capability across various cracks and complex backgrounds.

Keywords: poultry eggs; microcrack detection; lightweight model; semantic segmentation; deep learning



Citation: Shi, C.; Li, Y.; Jiang, X.; Sun, W.; Zhu, C.; Mo, Y.; Yan, S.; Zhang, C. Real-Time ConvNext-Based U-Net with Feature Infusion for Egg Microcrack Detection. *Agriculture* **2024**, *14*, 1655. <https://doi.org/10.3390/agriculture14091655>

Academic Editor: Massimo Cecchini

Received: 27 July 2024

Revised: 16 September 2024

Accepted: 19 September 2024

Published: 22 September 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

In 2023, China's poultry egg production and consumption reached approximately 34.5 million tons. Poultry eggs are rich in nutrients and have a significant positive impact on human health and development. The poultry egg industry plays a crucial role in driving national economic growth, promoting agricultural development, and ensuring food safety. However, during production, storage, and distribution, poultry eggs are prone to defects such as cracks, damage, and contamination [1], which typically require manual inspection and sorting. Microcracks, in particular, severely affect egg quality but are often overlooked during inspection [2]. Eggshell damage is a critical issue in egg production and a major source of economic loss for the industry [3]. As a result, researchers both domestically and internationally are increasingly focused on developing high-speed, high-precision poultry egg crack detection technologies. Current research primarily centers on acoustic, optical, and electrical analysis, as well as deep learning methods, achieving notable advancements in these fields.

The acoustic-based eggshell crack detection method leverages the physical properties of sound to analyze the vibration signals generated by striking an eggshell, offering a

novel research approach. Eggshell cracks can affect the structural strength and damping coefficient of eggs, which are reflected in the frequency and intensity of vibration signals. Chia-Chun Lai et al. [4] used cross-correlation analysis and Bayesian classification to classify and detect acoustic response signals generated by eggshell cracks, achieving a detection rate of 97%. Li Sun et al. [5] employed acoustic resonance analysis to detect eggshell cracks by impacting eggshells and collecting their response signals. They calculated Pearson correlation coefficients (PCCs) as feature parameters to compare the signal similarity between intact and cracked eggs at various sampling points. Based on these feature parameters, they established a general linear discriminant function to distinguish between intact and cracked eggs. In their experiment, a mixed sample of 100 chicken eggs and 100 duck eggs was used, and the crack detection rate reached 95.5%. However, these methods are susceptible to environmental conditions, egg shapes, crack locations and impact angles, leading to insufficient detection stability. Moreover, excessive impact force may further damage the eggshell structure. Additionally, since microcracks smaller than 5 mm have minimal impacts on the structural rigidity and resonance characteristics of eggs, acoustic techniques achieve lower accuracy in detecting microcracks [6].

Shi Chenbo et al. [7] proposed a non-destructive detection method for eggshell cracks based on a model of the electrical characteristics of eggs. This method involves establishing an electric field on the eggshell surface and analyzing the subtle current changes caused by cracks in the eggshell. The detection hardware system is designed based on electrical characteristic analysis. Additionally, Shi Chenbo et al. [8] developed a wavelet-scattering convolutional network algorithm utilizing 1D-CNN, LSTM, BiLSTM, and GRU networks for the non-destructive detection of microcracks in eggshells from electrical signals. While this method achieves high accuracy, it has limitations in detection coverage, failing to comprehensively identify all microcracks. Additionally, air humidity can affect conductivity, introducing instability to this method. The detection performance of this method may also be suboptimal for hidden cracks that have not fully penetrated the eggshell or for dry, fine cracks.

Compared to acoustic and electrically based methods, optically based egg crack detection methods offer the advantages of being non-destructive and stable. This technique captures images of the eggshell surface using image acquisition devices and processes and analyzes the image data. Therefore, machine vision-based eggshell crack detection methods are widely used in practical applications and have developed relatively mature technologies [9]. Guanjun Bao et al. [10] addressed irregular dark spots and invisible microcracks on the eggshell surface by using a negative LOG (Laplacian of Gaussian) operator to enhance cracks. They then employed a hysteresis threshold algorithm to eliminate irrelevant dark spots in the binary image, ensuring crack continuity. Finally, they used an improved Local Fitting Image (LFI) indicator to distinguish between cracks and false markings, achieving a crack detection rate of 92.5%. Kunshan Yao et al. [11] proposed methods such as crack enhancement and dual-threshold segmentation to extract the geometric features of cracks. Using the XGBoost classification model, they classified cracked eggs and achieved a recognition accuracy of 93.33%.

Traditional machine vision methods can effectively detect eggshell cracks; however, they often suffer from algorithmic complexity and slow processing speeds. With the recent advancements in deep learning in the field of computer vision, deep learning-based approaches have demonstrated superior performance in crack detection. Currently, the primary research achievements in eggshell crack detection focus on image classification tasks. Bhavya Botta et al. [12] created an eggshell crack dataset containing 468 images, with crack widths ranging from 20 to 40 μm and an average length of 30 mm. By employing convolutional neural networks (CNNs), they achieved a detection accuracy of 95.38%. Wenquan Tang [13] proposed the MobileNetV3_egg model, based on MobileNetV3_large, for real-time detection of damaged preserved eggs, achieving an accuracy of 96.3% and detecting 300 images in 4.267 s. Amin Nasiri et al. [14] collected images of intact, broken, and blood-stained eggs and used a CNN classification algorithm based on the VGG16

architecture, achieving an accuracy of 94.85%. With the introduction of segmentation networks such as U-Net, DeepLab-V3, and PSPNet, improved semantic segmentation models have also been applied to detection tasks. Xiuying Xu et al. [15] developed a semantic segmentation model based on an improved U-Net for the segmentation of corn stalks in fields with complex backgrounds, achieving an accuracy of 93.87%, outperforming U-Net, SegNet, and ResNet models. Chengqi Liu et al. [16] proposed a DeepLab V3+-based deep learning method for pig image segmentation in small sample sizes. They optimized the fusion of high- and low-frequency features using a recursive cascading approach to extract latent semantic information, resulting in a single-label model MIoU of 76.31%. However, these models are complex, with a large number of parameters and high training costs; therefore, they may not be suitable for the detection of narrow microcracks with few pixels in the images.

In the context of microcrack detection in poultry eggs for industrial applications, it is essential to address the model's scale and efficiency. Given the constraints of computational power, a pressing research issue is how to reduce the parameters of semantic segmentation models to make them more lightweight and accelerate inference without compromising detection accuracy. In 2022, Zhuang Liu et al. [17] redesigned a purely convolutional neural network model based on the standard ResNet and named it ConvNeXt. This model outperformed the Swin Transformer in the field of computer vision, with lower computational costs. Zhimeng Han et al. [18] proposed an effective model combining U-Net and ConvNeXt for medical image segmentation, achieving leading results with fewer parameters. Inspired by ConvNeXt, our model introduces improvements such as large convolution kernels and depthwise separable convolutions to reduce the number of model parameters.

In microcrack detection, challenges arise due to the narrow width of cracks and the similarity in contrast between the crack ends and the background texture. This paper addresses these challenges through the following two approaches: by enhancing detailed features and fully aggregating feature information for segmentation. In segmentation tasks, cracks can be divided into edge and internal regions [19]. In the field of object detection, some researchers [20] have demonstrated that utilizing and integrating edge features can improve the model's perception accuracy. Crack edge features help the model distinguish and locate crack defects. In digital image processing, edge detection effectively extracts object edge information by identifying pixels with significant brightness changes. Common detection algorithms include the Sobel, Prewitt, and Canny algorithms. Rasha Alshawi et al. [21] proposed using Sobel and Canny filters for edge detection to extract engineering features and manually integrate them into the network model to guide segmentation. This feature enhancement method provides a robust solution for segmentation tasks on limited and skewed datasets. Inspired by data augmentation, this paper's model incorporates a strategy of injecting image edge features into the encoder design to expand the detail feature space, thereby enhancing the model's ability to detect microcracks.

Additionally, Hengshuang Zhao et al. [22] proposed the Pyramid Scene Parsing Network (PSPNet) for pixel-level prediction tasks, providing an excellent framework by aggregating context from different regions. Building on this, Guosheng Lin et al. [23] proposed a multi-stage refinement network (RefineNet) that effectively integrates missing information from downsampling, producing high-resolution predictive images. It is evident that multi-scale feature fusion is beneficial for improving segmentation accuracy. Therefore, our model aggregates multi-scale features in the decoder to obtain fused features. Ablation experiments demonstrate that this approach significantly enhances accuracy compared to single upsampling of high-resolution feature maps for prediction.

In microcrack image segmentation, addressing class imbalance is a critical challenge due to the uneven distribution of foreground and background in the images. The positive-negative cross-entropy loss function is commonly used for most semantic segmentation tasks. However, when foreground pixels are significantly fewer than background

pixels, the background elements dominate the loss function calculation. This dominance causes the model to become overly biased toward the background, adversely affecting training and prediction outcomes. The Dice coefficient-based loss function proposed by Fausto Milletari et al. [24] has been widely adopted to mitigate the negative impact of imbalanced foreground and background areas on model performance. The Dice loss function places more emphasis on capturing the foreground regions, ensuring a lower false-negative rate. However, it suffers from the issue of loss saturation and is typically used in conjunction with the cross-entropy loss function. Michael Yeung et al. [25] proposed a unified focal loss that combines cross-entropy loss and Dice loss to address class imbalance. Experimental results indicate that this loss function exhibits robust performance in handling class imbalance.

In summary, this paper proposes a lightweight and highly real-time algorithm for the detection of microcracks in poultry eggs. The main innovations of this study are summarized as follows:

- This paper proposes a Real-time ConvNext-Based U-Net architecture with Feature Infusion for egg microcrack detection (CBU-FI Net). This architecture leverages the strengths of both U-Net and ConvNeXt, using ConvNeXt's fundamental modules as the backbone network. This approach significantly reduces the model's parameter count and computational complexity.
- To address the challenges in microcrack detection tasks, this paper introduces a feature infusion module within the encoder and employs multi-scale feature aggregation in the decoder for segmentation. By incorporating edge information, this strategy expands the spatial representation of crack features and enhances the extraction of both local details and global semantic information. This approach significantly improves microcrack segmentation accuracy, even with limited training data.
- To tackle the challenge of positive and negative sample imbalance in microcrack images for practical industrial applications, this paper introduces a hybrid loss function that combines cross-entropy loss and Dice loss. This approach significantly improves segmentation performance on microcrack images.

2. Materials and Methods

2.1. Detection Framework

This paper proposes CBU-FI Net, a real-time U-Net architecture enhanced with feature infusion, to meet the requirements for accurate detection and real-time performance in online poultry egg inspection. Figure 1 shows the overall network framework. With a parameter count of approximately 28 million, U-Net is considered a relatively lightweight segmentation model. In this study, U-Net is selected as the baseline model, then improved upon. Given that microcracks are characterized by small detection targets and low contrast against background textures, the primary focus of these improvements is on enhancing the network's ability to detect limited microcrack information and achieving efficient and rapid segmentation tasks with minimal parameters and computational load.

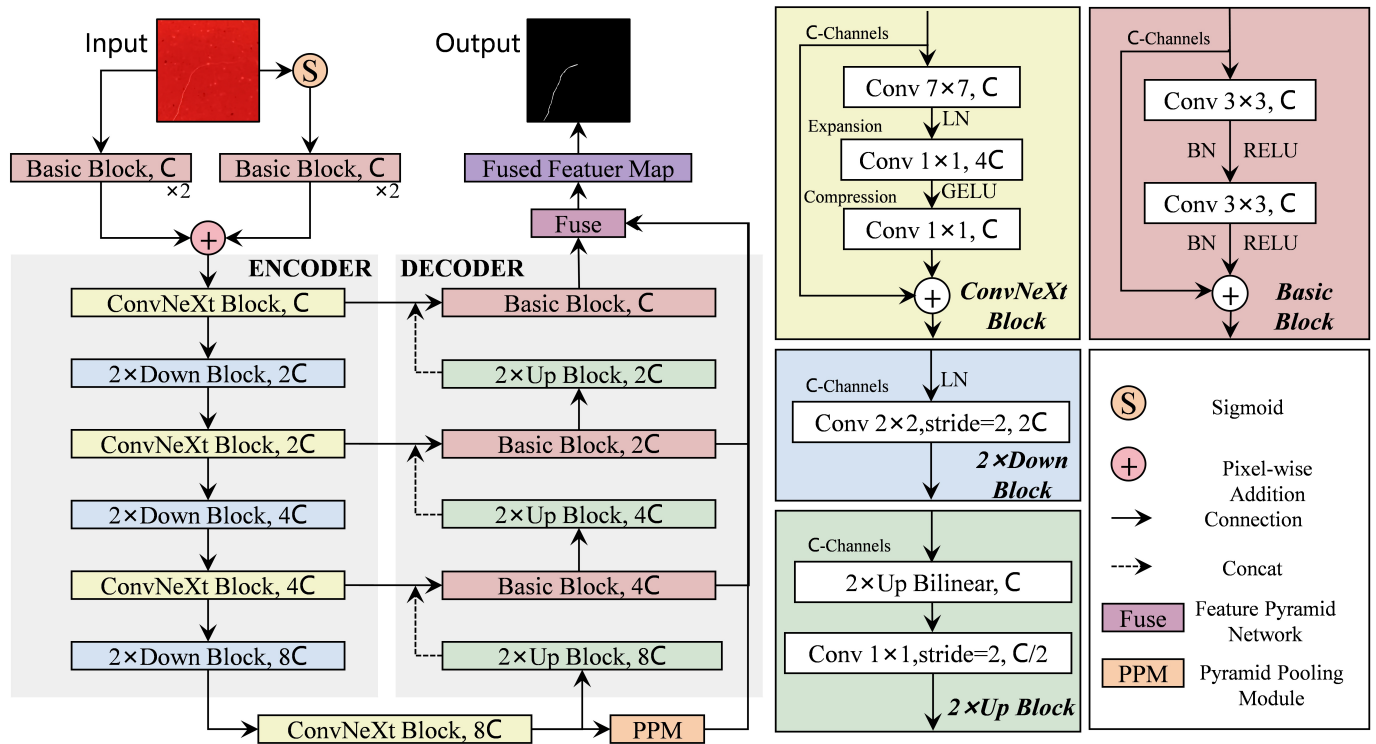


Figure 1. Architectural design of CBU-FI Net. The network is composed of three levels of encoders and decoders. The encoder includes three ConvNeXt blocks and downsampling layers, which progressively capture multi-scale features. The decoder consists of three basic blocks and upsampling layers that restore spatial resolution and integrate multi-scale features, enabling precise microcrack segmentation.

2.1.1. Overall Network Framework Design

Building upon the original U-Net architecture, CBU-FI Net achieves high-precision and high-speed semantic segmentation through its encoder–decoder framework. The improvement strategies for the encoder in this model are outlined as follows:

1. Drawing inspiration from dual-branch models, the traditional U-Net architecture is augmented with a feature infusion branch, enhancing the feature-space information of crack edges. To minimize the computational load of the fully convolutional network, crack images are processed through a dual-branch path before entering the encoder. Each branch employs a basic block to extract features from both the original image and the edge image. Consequently, the number of image channels increases from 3 to 48, and the image dimensions are reduced to one-fourth of their original size. The outputs from the two branches are then combined through pixel-wise summation before being input into the encoder.
2. The original U-Net architecture involves four downsampling operations, resulting in 1024 channels. To reduce model parameters and computational complexity, this model differs from the original U-Net by using ConvNeXt basic blocks as the backbone, replacing the traditional CBR structure with convolution operations. Additionally, the number of downsampling operations is reduced to three, and the number of channels is decreased to 384.
3. The original U-Net architecture employs max pooling for downsampling, which results in information loss—a significant drawback for microcrack segmentation tasks. To address this issue, the proposed model utilizes learnable dilated convolutions with a kernel size of 2 and a stride of 2 for downsampling. After three downsampling operations, the resolution of the feature maps is reduced to 1/32 of the original image size.

In the decoder stage, basic blocks from ResNet replace the two 3×3 convolutional layers of the original U-Net architecture, and bilinear interpolation is used for upsampling. In the final layer, a PPM (Pyramid Pooling Module) is applied to aggregate global context, expanding the receptive field, while an FPN (Feature Pyramid Network) combines multi-scale feature maps from basic blocks at various levels of the network. This approach allows the model to effectively capture both global and local context, which is crucial for accurately detecting and segmenting cracks in eggs. The multi-scale fusion of features, compared to using only the single highest-resolution feature map, minimizes information loss and enhances the precision of segmentation, particularly for small cracks and intricate boundary details.

2.1.2. Lightweight Trunk Feature Extraction Networks

ConvNeXt is a hybrid model that improves upon previous modules, leading to CNN performance in computer vision that surpasses that of the Swin Transformer. Specifically, it limits computational cost. In this model, the original CBR (Conv + BatchNorm + ReLU) structure is replaced with a more lightweight ConvNeXt Block, enhancing training and inference speed. As illustrated, the ConvNeXt Block begins with a 7×7 depthwise separable convolution, which captures more crack detail than a 3×3 convolution. Next, the expansion layer uses a 1×1 convolution to increase feature dimensions, followed by a compression layer, which also employs a 1×1 convolution to reduce the number of channels in the feature map, thereby decreasing the model's parameters. The output of the ConvNeXt block is a multi-channel feature map derived from the feature map output by the previous convolution module or its downsampled version. The feature maps output at various scales during the U-Net backbone stage are substituted with those produced by the ConvNeXt block at corresponding scales. Consequently, in the skip connection stage, these feature maps can still be concatenated with the upsampled feature maps, which are rich in abstract semantic information.

2.1.3. Feature Infusion Module

In microcrack detection, challenges arise due to the narrow width of the cracks and the low contrast between crack tips and the background texture. Edge features of cracks assist the model in distinguishing and locating crack defects. Therefore, incorporating edge information helps achieve clearer and more accurate crack shapes. Inspired by these methods, this paper introduces a feature infusion module based on the intrinsic characteristics of cracks. This module is integrated into the deep learning model to introduce edge information into crack detection, expanding the feature-space information of cracks. Consequently, it provides additional guidance to enhance model performance, particularly in cases of data imbalance.

The model utilizes the computationally efficient Sobel operator as an edge detector to achieve real-time crack detection, meeting the requirements for real-time processing. Specifically, prior to the encoder, two 3×3 convolution kernels are employed to implement the Sobel edge detector, which detects crack edges and highlights crack contours to generate an edge map of the crack image. To align the features of the original image with those of the edge image, a dual-branch fusion strategy is adopted. This strategy uses identical basic feature extraction blocks to align and extract features from both the original and edge images. The features from these dual branches are then combined using element-wise addition, effectively integrating the extracted edge features into the segmentation process.

Seg-Grad-CAM is employed to visualize class activation maps [26] to better understand the impact of feature guidance on crack segmentation. Under identical conditions, feature heat maps of models with and without the feature infusion structure are compared on a microcrack dataset. With feature infusion, the model exhibits increased focus on fine cracks, particularly enhancing the prediction accuracy for crack tips. In contrast, without feature infusion, the model struggles to detect small crack tips. This illustrates that the dual-branch feature infusion strategy significantly expands the feature space, surpassing

the segmentation capabilities of conventional deep learning and mitigating the limitations posed by class imbalance (Figure 2).

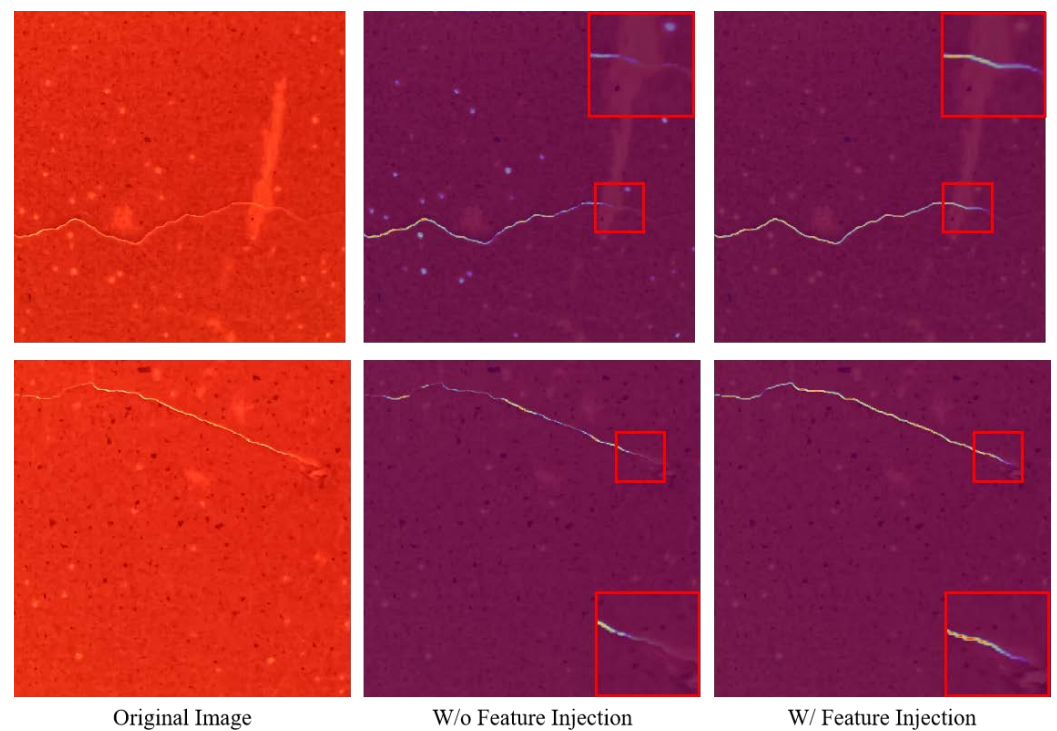


Figure 2. A visual comparison of the models with and without the feature infusion structure. The smaller red boxes in the figure highlights the end of the crack, and the zoomed-in view is displayed in the top-right/lower-right corner of the figure. By magnifying the segmentation results of microcrack tips, the figure emphasizes the significant advantages of the feature infusion structure in enhancing the accuracy of microcrack detection.

2.1.4. Composite Loss Function

In the training dataset, the imbalance between positive and negative samples is a significant issue. Table 1 shows that in eggshell crack images, the proportion of foreground (crack) pixels relative to background pixels is uneven, with the number of background pixels being 117 times that of crack pixels. This imbalance can cause deep learning models to overly focus on background pixels during training, leading to fluctuations in the loss function, affecting gradient stability, and hindering the improvement of model accuracy. Therefore, measures must be taken to address the imbalance between positive and negative samples to enhance model performance and accuracy.

Table 1. Average percentage of foreground and background pixels relative to total pixels in a cracked image.

Classes	Background	Crack
Percentage value (%)	99.441	0.559

The cross-entropy loss function is one of the most commonly used classification loss functions, measuring the discrepancy between the predicted values and the true labels. The definition of cross-entropy loss is shown in Equation (1).

$$L_{CE} = - \sum_i^i y_i \log \hat{y}_i, \quad (1)$$

where y_i represents the true label; \hat{y}_i represents the probability value predicted by the model, which ranges between 0 and 1; and the log function is the natural logarithm.

In this study, the number of microcrack pixels (target pixels) to be detected is significantly smaller than the number of background pixels, leading to a class imbalance problem. Dice loss, a region-related loss function, performs well in scenarios with class imbalance. Dice loss is defined by Equation (2), denoted as 1-Dice_coefficient, and represents the similarity between two samples.

$$L_{Dice} = 1 - Dice_{coefficient} = 1 - \frac{2 \times |X \cap Y|}{|X| + |Y|}, \quad (2)$$

It is computed by the intersection of the background pixel set (X) and the target pixel set (Y). A higher Dice coefficient indicates a more significant similarity between the two samples.

The network's performance in detecting microcracks is improved by proposing a composite loss function that integrates both cross-entropy loss and Dice loss. This approach leverages the strengths of both loss functions, addressing precision issues related to the small proportion of crack pixels and enhancing predictive accuracy. The expression for this composite loss function is provided in Equation (3).

$$Loss = L_{CE} + \alpha L_{Dice}, \quad (3)$$

where α denotes the adjustment factor.

Analysis of the segmentation results in Figure 3b shows that the original image in Figure 3a contains both large cracks and fine microcracks. When relying solely on cross-entropy loss, the model primarily focuses on segmenting prominent main cracks while neglecting smaller or less noticeable microcracks. Thus, cross-entropy loss mainly benefits the model in learning features related to background pixels but performs relatively poorly in capturing fine details. In contrast, as shown in Figure 3c, the model trained with Dice loss exhibits improved segmentation of details and local features, showing higher sensitivity to boundary details and better capturing the complex information related to microcracks. Finally, the composite loss function used in Figure 3d combines the strengths of both loss functions, resulting in superior overall segmentation performance.

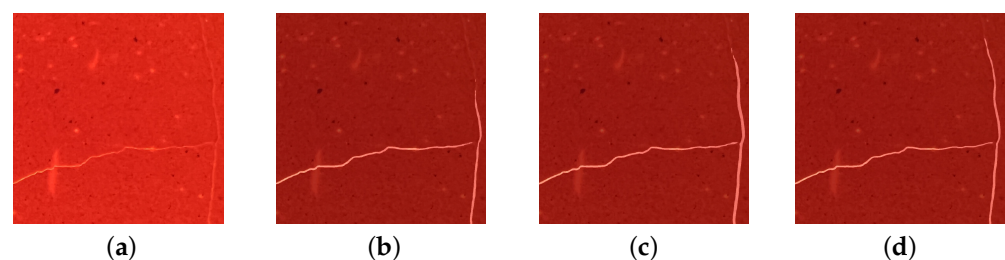


Figure 3. Crack segmentation results with different loss functions: (a) original image; (b) cross-entropy loss alone; (c) Dice loss alone; (d) composite loss function.

2.2. Dataset

2.2.1. Microcrack Dataset for Eggs

In previous work, we created a large-scale pixel-level annotated dataset comprising 3436 samples of microcrack defects in eggs, with each image having a size of 1024×1024 pixels. The crack types in these image blocks include curved, branched, and reticulated uneven cracks. Microcracks predominantly appear at the terminals and branch ends of the cracks, and they are mainly distributed along the edges of the feature maps. Example images from the dataset are shown in Figure 4. In this study, we quantitatively classified the dataset based on the width and average length of the cracks, dividing it into two levels—L1 patch and L2 patch—with detailed information provided in Table 2.

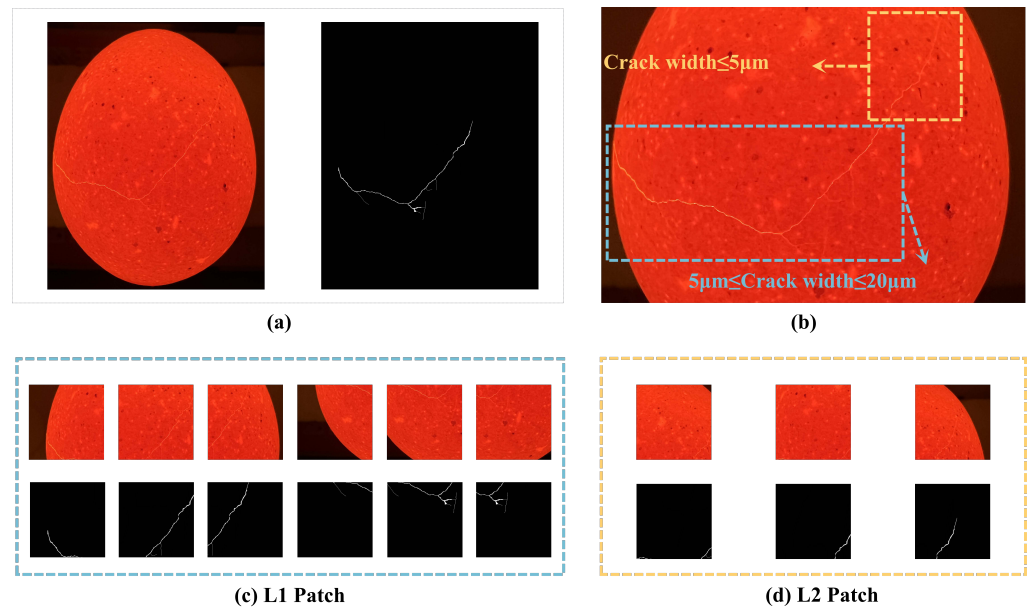


Figure 4. Example images from the dataset: (a) original egg image and corresponding mask map; (b) crack grading map; (c) L1 patch (crack width less than 20 μm and more than 5 μm); (d) L2 patch (crack width less than 5 μm).

Table 2. Crack size grading quantization table.

Level	Crack Width	Average Crack Length	Count
L1 Patch	5 μm ~20 μm	10 mm	2486
L2 Patch	≤ 5 μm	30 μm	950

2.2.2. Public Segmentation Datasets

We selected the publicly available CrackSeg9k [27] dataset for performance assessment to evaluate the reliability and generalizability of the model. This dataset comprises a total of 9255 images (400×400 resolution), combining seven different smaller, open-source datasets. It includes various structural cracks in masonry walls, ceramics, concrete bridge decks, walls, and pavement, featuring microcracks with widths ranging from 0.06 mm to 25 mm. Shreyas Kulkarni et al. [27] optimized the dataset by standardizing the masks across multiple classes, creating CrackSeg9k, the largest, most diverse, and most consistent crack segmentation dataset constructed to date.

2.3. Evaluation Metrics

The performance of the trained model in detecting microcracks in poultry eggs is assessed from the following two perspectives: model performance and real-time efficiency. Model performance is evaluated using metrics such as the confusion matrix, pixel accuracy, crack Intersection over Union (Crack-IoU), and Mean Intersection over Union (MIoU) to gauge classification accuracy and segmentation precision. As shown in Equations (4)–(6), these indices take values in the interval of [0, 1], with higher values indicating better segmentation results. Real-time efficiency is evaluated based on inference time, parameter count, and FLOPs (floating-point operations per second). Higher values in these three metrics indicate greater processing power and lower computational complexity for the model. A comprehensive assessment of these metrics offers a thorough evaluation of the model's effectiveness and efficiency in practical applications.

$$\text{Pixel Accuracy} = \frac{TP + TN}{TP + TN + FP + FN'} \quad (4)$$

$$Crack - IoU = \frac{TP}{TP + FP + FN'} \quad (5)$$

$$MIoU = \frac{1}{N} \sum_{k=1}^n IoU_k, \quad (6)$$

where k is the number of categories, IoU_k represents the IoU value for type k , and n represents the total number of categories in the dataset.

TP denotes the number of pixels correctly classified as cracks, FP denotes the number of pixels incorrectly classified as cracks, TN denotes the number of pixels correctly classified as background, and FN denotes the number of pixels incorrectly classified as background.

2.4. Experimental Environment and Parameter Settings

Information about the server hardware and software used in this study is shown in Table 3.

Table 3. Server hardware and software information.

Attribute Name	Attribute Content
CPU	12th Gen Intel(R) Core(TM) i7-12700K
GPU	GeForce RTX 3090
Graphics Memory	48G
Operating System	Ubuntu 22.04.1
Deep Learning Framework	MMSegmentation
MMSegmentation Version	1.2.2
Python Version	3.8
Pytorch Version	1.10.1

The egg microcrack dataset contains a total of 3436 sheets, and the ratio of the division of the training set, the validation set, and the test set is 6:1:3. We employed a learning rate strategy described by the following Equation (7):

$$l_r = b_{l_r} \times \left(1 - \frac{e}{\max e}\right)^p \quad (7)$$

where b_{l_r} represents the base learning rate, e represents the current iteration number, $\max e$ denotes the maximum number of iterations for the model, and p is a hyperparameter used to control the learning rate. The hyperparameter values used for network training are presented in Table 4.

Table 4. Hyperparameter information.

Hyperparameter	Values
l_r	Type = Poly *, $p = 0.9$
Optimizer	Type = SGD **, $b_{l_r} = 0.01$, Momentum *** = 0.9
Batch size	8
Max Iterations	80,000

* An exponential transformation strategy. ** Stochastic gradient descent. *** The momentum algorithm is employed as the optimization algorithm.

3. Results

3.1. Data Analysis

This paper compares six advanced deep learning-based segmentation architectures to evaluate the performance of the CBU-FI Net model in microcrack detection tasks. Extensive experiments were conducted on a microcrack defect detection dataset for eggs and the publicly available CrackSeg9k segmentation dataset. The selected models include three classic semantic segmentation models (U-Net [28], PSPNet (Res50) [22], and DeepLabV3

(Res50) [29]), as well as three state-of-the-art dual-branch models (BiseNetV2 [30], DDRNet [31], and HrSegNet-B48 [32]). These comparative experiments aim to comprehensively assess and verify the superior performance of CBU-FI Net in detecting microcracks.

3.1.1. Comparative Evaluation of Model Performance on Ungraded Datasets

In the comparative experiments for egg microcrack detection, CBU-FI Net and six other models were trained and tested on the egg microcrack defect detection dataset with identical proportions, with each image having a resolution of 1024×1024 . Additionally, this model was compared with our previous research result obtained using the MobileUNet-CBAM model [33]. The experiment evaluated the performance of each model in detecting egg microcracks with a crack width of less than $20 \mu\text{m}$, with the results presented in Table 5.

Table 5. Results of different models on the ungraded dataset.

Network Structure	Pixel Accuracy (%)	Crack-IoU (%)	MIoU (%)	Inference Time (s)	Parameters (M)	FLOPs (GFLOPs)
U-Net	71.38	61.56	80.28	0.079	29.06	811.64
PSPNet (Res50)	75.93	64.22	81.84	0.069	48.98	713.76
DeepLabV3 (Res50)	76.30	64.30	81.88	0.091	68.01	1078.57
BiseNetV2	76.84	64.43	81.92	0.007	3.341	49.142
DDRNet	77.31	64.95	82.18	0.009	20.299	71.759
HrSegNet-B48	75.15	64.45	81.94	0.008	5.465	32.672
MobileUNet-CBAM	77.04	65.01	82.24	0.080	14.09	96.38
CBU-FI Net	78.70	65.51	82.47	0.021	7.696	102.88

According to the results on the ungraded dataset, the proposed CBU-FI Net achieved the highest pixel accuracy (78.70%), Crack-IoU (65.51%), and MIoU (82.47%) with a small parameter count (7.696M). Classic semantic segmentation models like U-Net and PSPNet (Res50) underperformed, primarily due to their reliance on standard convolutions, which limit their ability to capture the fine-grained contextual details needed for small crack detection. DeepLab-V3 (Res50) performed better due to the use of atrous convolutions that expanded the receptive field, but this came at the cost of a longer inference time and increased computational complexity, making it inefficient for real-time tasks.

Recent advancements have introduced dual-branch architectures, such as BiseNetV2, DDRNet, and HrSegNet-B48, which aim to balance semantic information and fine detail extraction while reducing computational load. While these networks significantly reduced model size and complexity, their light weights designs led to less accurate segmentation of microcracks. BiseNetV2 and DDRNet achieved fast inference by simplifying feature extraction, but this compromised their ability to detect microcracks under $20 \mu\text{m}$. HrSegNet-B48, despite improving detail processing with a more complex encoder and modules, still struggled to achieve the precision needed for microcrack detection when compared to CBU-FI Net.

CBU-FI Net's lightweight design surpassed MobileUNet-CBAM, achieving a scale similar to that of dual-branch networks while offering competitive inference speed and delivering superior segmentation accuracy. Its ability to aggregate multi-scale features allowed it to excel in detecting microcracks under $20 \mu\text{m}$, striking a balance between model complexity and segmentation performance.

3.1.2. Comparative Analysis of Models on the Graded Dataset (L2 Patch)

The model's performance was evaluated using the L2 patch dataset, where all crack widths are below $5 \mu\text{m}$. The results are presented in Figure 5. The decrease in accuracy compared to Experiment 3.1.1 can be attributed to the dataset's finer microcracks, which cause a loss of detail and texture features in low-resolution or downsampled images. This makes it more challenging for models to capture critical microcrack features.

U-Net and PSPNet struggled due to their inability to effectively detect such small details. While DeepLab-V3 (Res50) benefited from atrous convolutions to capture more global information, it still lacked the precision needed for fine cracks under 5 μm . Lightweight models like BiSeNetV2, DDRNet, and HrSegNet-B48 prioritized efficiency but sacrificed fine-grained feature extraction, leading to less accurate segmentation in this context. Despite these challenges, the improved CBU-FI Net showed superior performance by effectively retaining fine textures through multi-scale feature aggregation and enhanced upsampling. This allowed it to consistently detect microcracks with greater accuracy, demonstrating its adaptability and generalization capabilities, even for cracks smaller than 5 μm .

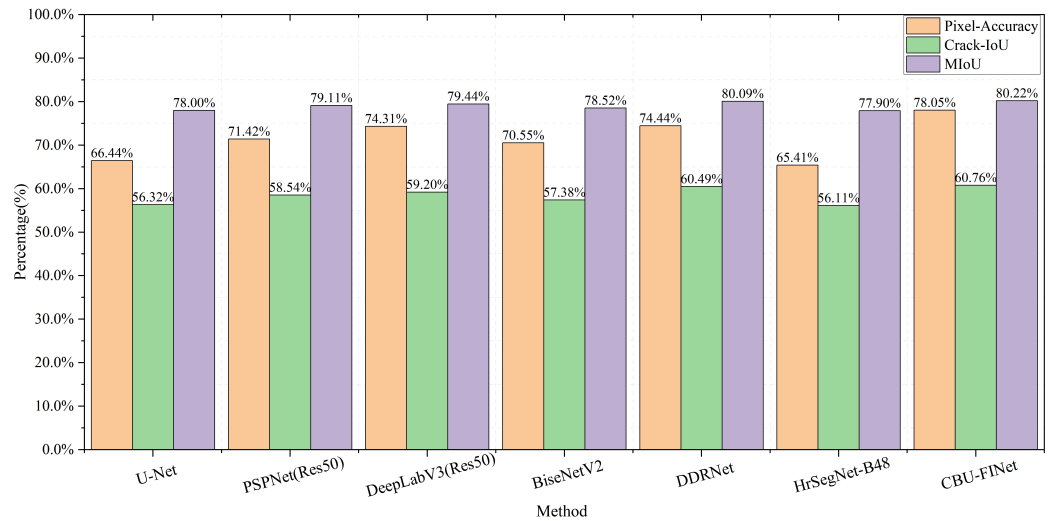


Figure 5. Results of model performance on the L2 patch dataset.

3.1.3. Comparative Analysis of Model Performance on a Public Dataset (CrackSeg9k)

We conducted experiments on the publicly available CrackSeg9k dataset to assess the robustness and generalization capabilities of various semantic segmentation models for crack detection. Our study included CBU-FI Net and six other models. All models were trained and tested on the same proportion of training and testing sets, with each image having a resolution of 400×400 pixels. To ensure a valid comparison, all models were trained from scratch without using pre-trained weights from other datasets. The results of these experiments are shown in Table 6.

Table 6. Results of different models on the CrackSeg9k dataset.

Network Structure	MIoU (%)	Inference Time (s)	FLOPs (GFLOPs)
U-Net	80.69	0.013	124.01
PSPNet(Res50)	80.80	0.012	109.22
DeepLabV3(Res50)	80.22	0.015	165.57
BiSeNetV2	78.85	0.002	7.513
DDRNet	75.39	0.003	11.122
HrSegNet-B48	79.26	0.002	5.036
CBU-FI Net	81.38	0.004	15.692

CBU-FI Net achieved an MIoU of 82.47% and an inference time of 0.021 s per image, setting a benchmark in crack detection for structures like masonry walls, ceramics, and pavement. It performed well on cracks ranging from 0.06 mm to 25 mm, demonstrating strong generalization across diverse surfaces. In contrast, U-Net and PSPNet struggled with finer cracks due to limitations in feature extraction. DeepLab-V3 (Res50) handled larger cracks better but was slower and more computationally demanding. Lightweight models like BiSeNetV2, DDRNet, and HrSegNet-B48, while faster, sacrificed accuracy on narrower cracks. Their simpler architectures favored speed over precision, limiting their

effectiveness in detecting microcracks. CBU-FI Net, by balancing model complexity and detail retention, demonstrated superior accuracy and adaptability across a wide range of crack sizes and backgrounds.

3.2. Results of Ablation Experiments

3.2.1. Experimental Results on Composite Loss Parameters

We observed the impact of varying the weight of the adjustment coefficient on model performance under the same experimental conditions to investigate the setting of the adjustment coefficient in the proposed composite loss function. The experimental results are illustrated in Figure 6, where the horizontal axis represents the adjustment coefficient (α) for the Dice loss term in the composite loss function. The left vertical axis represents the Crack-IoU value, while the right vertical axis represents the pixel accuracy value.

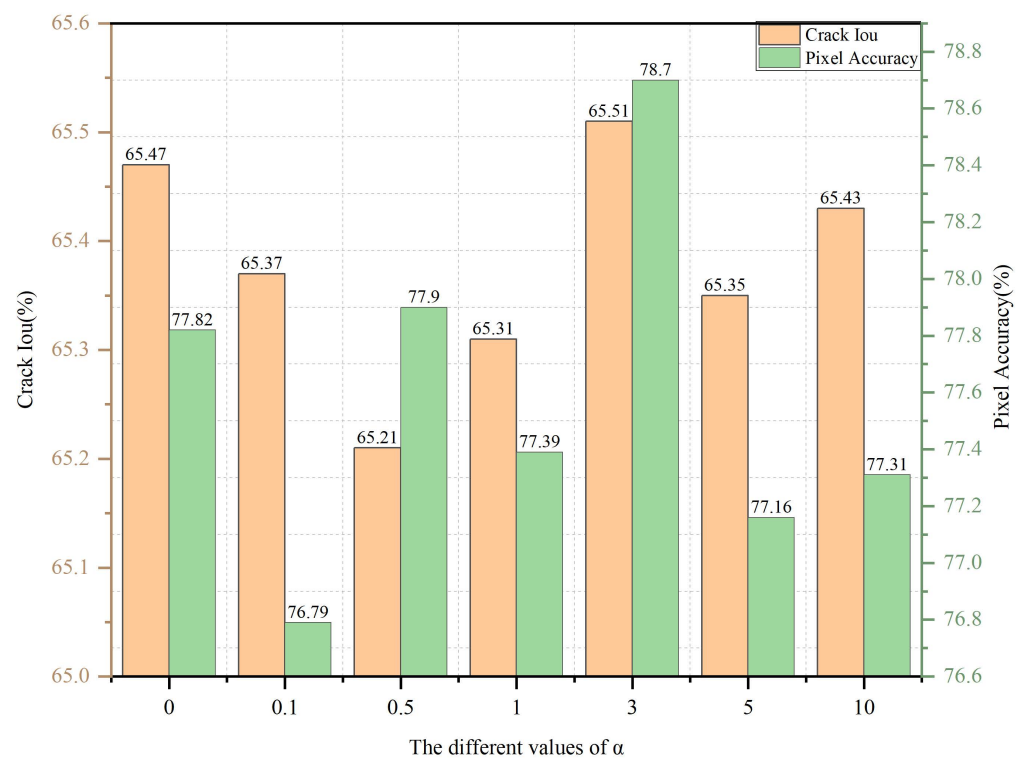


Figure 6. Results of different settings for the adjustment coefficient (α) in the composite loss function.

Based on our analysis, we can conclude the following. When the weight of the cross-entropy loss is relatively high, the model's accuracy decreases. This is because cross-entropy loss focuses more on the overall segmentation results and is less effective at capturing detailed features in microcrack detection. As the weight of the Dice loss increases, the model's performance improves. Under the given experimental conditions, setting the parameter weight to 3 (with a ratio of cross-entropy loss to Dice loss weight of 1:3) yields the optimal max Crack-IoU, indicating that the model achieves its best performance at this ratio. However, as the Dice loss weight increases further, the model's performance gradually declines slightly. This suggests that the model's performance is relatively insensitive to changes in the Dice loss, indicating a certain robustness in the selection of loss function weights.

3.2.2. Verification of Model Ablation Experiments

We conducted ablation experiments to further verify the effectiveness of the CBU-FI Net model in crack segmentation tasks. By incrementally adding various components to the model, we assessed their impact on performance. Table 7 summarizes the mean

intersection over union (MIoU), parameter count, and floating-point operations (FLOPs) for each ablation experiment configuration.

Table 7. Ablation results of each module of the CBU-FI Net model.

U-Net	ConvNeXt Block	Feature Injection	Fuse	Cross-Entropy Loss	Dice Loss	MIoU (%)	Parameters (M)	FLOPs (GFLOPs)
✓				✓		80.28	29.06	811.64
✓	✓			✓		81.68	5.055	27.34
✓	✓	✓		✓		82.10	5.430	52.54
✓	✓	✓	✓	✓		82.45	7.696	102.88
✓	✓	✓	✓	✓	✓	82.47	7.696	102.88

Analysis of the ablation experiments clearly demonstrates the impact of each component on the performance enhancement of the CBU-FI Net model. Adding the ConvNeXt block resulted in an increase in MIoU to 81.68%, while significantly reducing the parameter count to 5.055 M and FLOPs to 27.34 GFLOPs. This indicates that the ConvNeXt block effectively improves segmentation accuracy while reducing model complexity. The subsequent inclusion of the feature infusion module increased the MIoU to 82.10%, with a slight rise in parameter count to 5.430 M and FLOPs to 52.54 GFLOPs, highlighting the contribution of feature infusion to segmentation accuracy while maintaining a low computational complexity. The introduction of the Fuse module further raised the MIoU to 82.45%, with the parameter count and FLOPs increasing to 7.696 M and 102.88 GFLOPs, respectively, demonstrating the critical role of the Fuse module in feature fusion, significantly enhancing model performance. Finally, refining the composite loss function led to a slight MIoU increase to 82.47%, with no change in parameter count or FLOPs, indicating that the designed loss function further optimized model performance during the final fine-tuning stage, achieving the best MIoU. Overall, the results of the ablation experiments further validate the effectiveness of the introduced improvements.

3.3. Detection Visualization and Analysis

This section presents a qualitative analysis of the performance of CBU-FI Net and six other models across various datasets. Through visualization, we comprehensively evaluate each model's segmentation performance under different conditions. The qualitative analysis provides insights into how each model handles diverse types of cracks, varying background complexities, and changes in image quality, thereby enhancing our understanding of their robustness and generalization capabilities.

3.3.1. Visualization Analysis of the Poultry Egg Dataset

To evaluate the performance of each model in the microcrack detection task, we used samples characterized by lower clarity, the presence of texture interference, and finer cracks. The experimental results are illustrated in Figure 7, with red boxes indicating the areas of missed detection.

The original U-Net network exhibits relatively poor performance in crack segmentation, with some crack regions appearing fragmented and unable to be completely segmented. Additionally, the finer parts at the ends of cracks are not accurately segmented. This may be due to the structural limitations of the U-Net network in extracting microcrack features, leading to less detailed segmentation and a significant number of missed detections. Although other advanced models have shown improvements in microcrack detection, they still exhibit breakage, especially when crack features are not prominent, making it difficult to fully segment the crack regions. In contrast, the CBU-FI Net model proposed in this paper performs exceptionally well in crack segmentation tasks across various scales and shapes. Its overall detection results are closest to the ground truth, with significantly reduced breakage in crack segmentation, accurately segmenting the fine parts at the ends of cracks, thereby making the segmented crack details and boundaries clearer.

In summary, when comparing the results in the figures, it is evident that the improved algorithm proposed in this paper demonstrates outstanding segmentation performance in the poultry egg microcrack detection task. Compared to the original U-Net and other advanced models, the CBU-FI Net model effectively segments cracks. The segmentation visualization results indicate that this model has a significant advantage in handling microcracks of various scales and shapes, better capturing the detailed features of cracks and fully segmenting microcrack regions. This effectively reduces segmentation breakage and enhances the clarity of crack boundaries, making the shape of the cracks more realistic and accurate. This improvement has significant application value for microcrack detection tasks in the poultry egg industry.

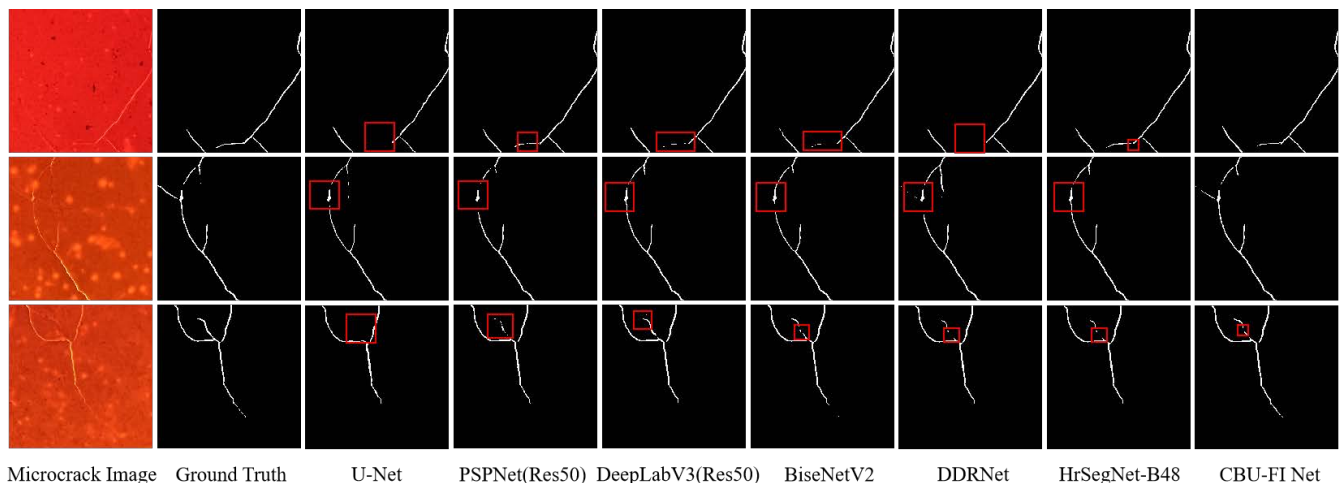


Figure 7. Visualization results of different models. The red boxes in the figure indicate missed detections.

3.3.2. Visualization Analysis of the CrackSeg9k Dataset

We selected five samples from the CrackSeg9k test set for visualization analysis to highlight the effectiveness of the proposed method. These samples encompass a range of crack scenarios, including asphalt pavement, concrete pavement, stone brick walls, and tile floors. By analyzing the crack segmentation results across these diverse scenarios, we can comprehensively assess the robustness and generalization capabilities of the proposed CBU-FI Net model in real-world applications.

The results indicate that in multi-scenario crack segmentation tasks, larger models perform slightly better than lightweight models, suggesting that computational limitations can affect model segmentation capabilities. When detecting pavement cracks, even though the materials differ, most models can segment the main shape of the cracks, since the crack features are distinctly different from the background texture. However, issues such as missed detections at the crack ends and crack breakage still occur. In contrast, when detecting cracks in tiles and stone bricks, the high similarity between the crack features and the background texture makes them harder to distinguish. Only PSPNet (Res50), DeepLab (Res50), and the proposed CBU-FI Net successfully detected the main cracks. Notably, PSPNet (Res50) and DeepLab (Res50) still exhibit crack breakage and missed ends. The proposed CBU-FI Net model significantly improves microcrack detection performance against low-contrast backgrounds through the design of feature infusion and feature fusion strategies. Compared to other models, CBU-FI Net excels in crack segmentation tasks for various scales and shapes, with detection results closest to the ground truth. This model notably reduces crack breakage and enhances the clarity of crack details and boundaries, demonstrating exceptional robustness and generalization capabilities in handling complex backgrounds and diverse crack features. In summary, the proposed CBU-FI Net model demonstrates superior segmentation performance across various crack scenarios, validating its effectiveness and practical value in crack segmentation tasks. Its design strategies not only improve detection accuracy but also significantly enhance the segmentation of crack

details and boundaries, providing a reliable solution for crack detection in real-world applications (Figure 8).

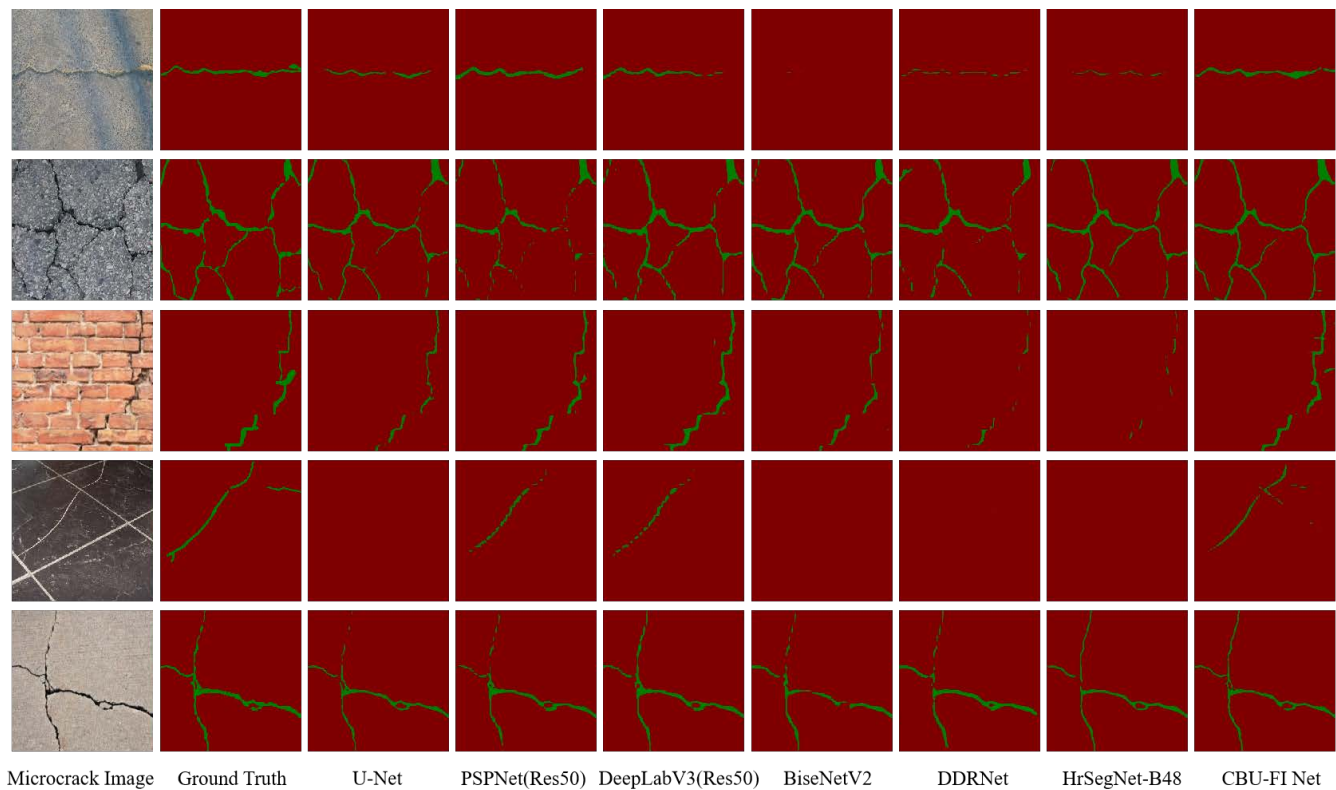


Figure 8. Visualization analysis of various models on the CrackSeg9k dataset.

4. Discussion

This study addresses the low efficiency and high cost of traditional poultry egg microcrack detection by proposing a real-time U-Net architecture based on feature infusion (CBU-FI Net). Using U-Net as the baseline model, we improved it to address the issues of large model parameters and slow image processing speed. The proposed model combines U-Net and ConvNeXt, enhancing the extraction of local detail information and global semantic information by injecting image edge features into the encoder and aggregating multi-scale features in the decoder. Additionally, a hybrid loss function based on cross-entropy loss and Dice loss was designed to further improve the segmentation performance of microcrack images, achieving real-time, high-precision, pixel-level microcrack detection. Compared to the original U-Net model, the improved model reduces the number of parameters and accelerates image processing speed while optimizing the model's focus on microcracks. This enhances the proportion of crack edge features within the overall semantic information, strengthening the model's ability to extract crack features.

The proposed real-time poultry egg microcrack detection model based on feature infusion achieves a real-time detection speed of 21 ms for input images with a resolution of 1024×1024 . When detecting microcracks smaller than $20 \mu\text{m}$, the model achieved an MIoU of 82.47%. On the benchmark CrackSeg9k dataset, it achieved an inference speed of 4 ms per image with a resolution of 400×400 and an MIoU of 81.38%, with a model parameter count of 7.696 M. Compared to state-of-the-art segmentation models, it achieves leading accuracy with fewer parameters. Using industrial microscopic measurement, the smallest detectable crack size is $3 \mu\text{m}$. According to the visualization results presented in Section 3.3, although microcrack segmentation is relatively complete, there are still small parts of microcrack ends that are not fully detected. This may be due to the microcrack width being less than $3 \mu\text{m}$ or insufficient image resolution, resulting in the cracks occupying fewer

pixels in the image and making it difficult for the algorithm to accurately segment the complete outline of the cracks.

The high-efficiency microcrack detection system we propose can help to improve production efficiency on poultry egg production lines, reduce economic losses caused by defective eggs, and enhance agricultural production efficiency. It can positively impact the upgrading and development of the entire poultry egg industry.

5. Conclusions

This study proposes a real-time ConvNext-based U-Net with feature infusion for egg microcrack detection (CBU-FI Net). This model integrates the advantages of ConvNext's large convolution kernels and depthwise separable convolutions, addressing the issues of large parameter size and slow inference speed in the U-Net model. To tackle challenges in microcrack detection, such as narrow crack width and low contrast between crack ends and the background texture, the model infuses image edge features in the encoder and aggregates multi-scale features in the decoder, enhancing the extraction of both local detail information and global semantic information. Additionally, a composite loss function is constructed to address the imbalance between positive and negative samples. Using a single-point backlight source collection device, over 3400 graded poultry egg microcrack image patches were created for model training and validation.

Experimental results demonstrate that CBU-FI Net's parameter size is only one-third of that of the original U-Net, and the inference speed is 21 ms per image (1 million pixels). This model exhibits strong robustness and generalization ability, adapting to different types of cracks and complex background environments. On the public benchmark CrackSeg9k dataset, CBU-FI Net achieved an inference speed of 4 ms per 400×400 image and an MIoU of 81.38%. Additionally, on the poultry egg microcrack defect dataset, the model's Crack-IoU for microcracks (less than $20 \mu\text{m}$) was 65.51%; for smaller cracks (less than $5 \mu\text{m}$), the detection results show a Crack-IoU and MIoU of 60.76% and 80.22%, respectively, achieving real-time, high-precision real-microcrack detection. These results are superior to those achieved by the traditional U-Net model and other advanced semantic segmentation models, laying a foundation for online microcrack detection in poultry eggs based on semantic segmentation.

The proposed method for detecting microcracks in poultry eggs can significantly enhance production efficiency on egg processing lines, reducing economic losses associated with undetected defects. By improving the accuracy and reliability of crack detection, this method contributes to higher standards of food safety and quality control. Additionally, the method can be extended to other industries requiring precise defect detection, such as materials science, automotive manufacturing, and aerospace. As the technology continues to develop, it may find broader applications, driving innovation and efficiency across various sectors and supporting sustainable growth and advancement in these industries. In the future, we will explore the integration of discharge methods with imaging techniques to develop a multi-sensor fusion approach for the detection of microcracks in poultry eggs. By combining discharge signals with image data, we aim to enhance detection accuracy and robustness, thereby further improving the effectiveness of microcrack detection.

Author Contributions: Conceptualization, C.S. and C.Z. (Chun Zhang); methodology, C.S., C.Z. (Chun Zhang) and Y.L.; software, Y.L. and X.J.; validation, C.S., C.Z. (Chun Zhang), C.Z. (Changsheng Zhu), Y.L. and Y.M.; formal analysis, C.S., C.Z. (Chun Zhang), Y.L., X.J., Y.M., S.Y. and W.S.; investigation, C.S., Y.L., and C.Z. (Changsheng Zhu); resources, C.S. and Y.M.; data curation, Y.L., X.J. and W.S.; writing—original draft preparation, C.S., C.Z. (Chun Zhang) and Y.L.; writing—review and editing, C.S., Y.L., C.Z. (Changsheng Zhu), X.J. and W.S.; visualization, C.S., C.Z. (Chun Zhang), Y.L., C.Z. (Changsheng Zhu), Y.L., W.S. and Y.M.; supervision, C.S., C.Z. (Chun Zhang) and C.Z. (Changsheng Zhu). All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Shandong Province Science and Technology SME Innovation Capability Enhancement Project (No. 2023TSGC0576 and No. 2023TSGC0605) and the Tai'an Science and Technology Innovation Development Plan (No. 2021GX050 and No. 2020GX055).

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The datasets presented in this article are not readily available because the data are part of an ongoing study. Requests to access the datasets should be directed to skd996523@sdust.edu.cn. The data derived from public-domain resources in this study are available in the CrackSeg9k dataset at https://doi.org/10.1007/978-3-031-25082-8_12 [27].

Acknowledgments: Special thanks to the Intelligent Perception and Cognition Laboratory at Shandong University of Science and Technology for providing the necessary facilities and resources to support this research.

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

References

1. Brake, J.; Walsh, T.; Benton, C., Jr.; Petite, J.; Meijerhof, R.; Penalva, G. Egg handling and storage. *Poult. Sci.* **1997**, *76*, 144–151. [[CrossRef](#)] [[PubMed](#)]
2. de Abreu Fernandes, E.; Litz, F.H. The eggshell and its commercial and production importance. In *Egg Innovations and Strategies for Improvements*; Elsevier: Amsterdam, The Netherlands, 2017; pp. 261–270.
3. Mazzuco, H.; Bertechini, A.G. Critical points on egg production: Causes, importance and incidence of eggshell breakage and defects. *Ciência e Agrotecnologia* **2014**, *38*, 7–14. [[CrossRef](#)]
4. Sun, L.; Zhang, P.; Feng, S.; Qiang, M.; Cai, J. Eggshell crack detection based on the transient impact analysis and cross-correlation method. *Curr. Res. Food Sci.* **2021**, *4*, 716–723. [[CrossRef](#)] [[PubMed](#)]
5. Sun, L.; Feng, S.; Chen, C.; Liu, X.; Cai, J. Identification of eggshell crack for hen egg and duck egg using correlation analysis based on acoustic resonance method. *J. Food Process. Eng.* **2020**, *43*, e13430. [[CrossRef](#)]
6. Bain, M.; MacLeod, N.; Thomson, R.; Hancock, J.W. Microcracks in eggs. *Poult. Sci.* **2006**, *85*, 2001–2008. [[CrossRef](#)]
7. Shi, C.; Wang, Y.; Zhang, C.; Yuan, J.; Cheng, Y.; Jia, B.; Zhu, C. Nondestructive Detection of Microcracks in Poultry Eggs Based on the Electrical Characteristics Model. *Agriculture* **2022**, *12*, 1137. [[CrossRef](#)]
8. Shi, C.; Cheng, Y.; Zhang, C.; Yuan, J.; Wang, Y.; Jiang, X.; Zhu, C. Wavelet Scattering Convolution Network-Based Detection Algorithm on Nondestructive Microcrack Electrical Signals of Eggs. *Agriculture* **2023**, *13*, 730. [[CrossRef](#)]
9. Purahong, B.; Chaowalittawin, V.; Krungseanmuang, W.; Sathaporn, P.; Anuwongpinit, T.; Lasakul, A. Crack Detection of Eggshell using Image Processing and Computer Vision. *J. Phys. Conf. Ser.* **2022**, *2261*, 012021. [[CrossRef](#)]
10. Guanjun, B.; Mimi, J.; Yi, X.; Shibo, C.; Qinghua, Y. Cracked egg recognition based on machine vision. *Comput. Electron. Agric.* **2019**, *158*, 159–166. [[CrossRef](#)]
11. Yao, K.; Sun, J.; Chen, C.; Xu, M.; Zhou, X.; Cao, Y.; Tian, Y. Non-destructive detection of egg qualities based on hyperspectral imaging. *J. Food Eng.* **2022**, *325*, 111024. [[CrossRef](#)]
12. Botta, B.; Gattam, S.S.R.; Datta, A.K. Eggshell crack detection using deep convolutional neural networks. *J. Food Eng.* **2022**, *315*, 110798. [[CrossRef](#)]
13. Tang, W.; Hu, J.; Wang, Q. High-throughput online visual detection method of cracked preserved eggs based on deep learning. *Appl. Sci.* **2022**, *12*, 952. [[CrossRef](#)]
14. Nasiri, A.; Omid, M.; Taheri-Garavand, A. An automatic sorting system for unwashed eggs using deep learning. *J. Food Eng.* **2020**, *283*, 110036. [[CrossRef](#)]
15. Xu, X.; Gao, Y.; Fu, C.; Qiu, J.; Zhang, W. Research on the Corn Stover Image Segmentation Method via an Unmanned Aerial Vehicle (UAV) and Improved U-Net Network. *Agriculture* **2024**, *14*, 217. [[CrossRef](#)]
16. Liu, C.; Su, J.; Wang, L.; Lu, S.; Li, L. LA-DeepLab V3+: A Novel Counting network for pigs. *Agriculture* **2022**, *12*, 284. [[CrossRef](#)]
17. Liu, Z.; Mao, H.; Wu, C.Y.; Feichtenhofer, C.; Darrell, T.; Xie, S. A convnet for the 2020s. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 11976–11986.
18. Han, Z.; Jian, M.; Wang, G.G. ConvUNeXt: An efficient convolution neural network for medical image segmentation. *Knowl.-Based Syst.* **2022**, *253*, 109512. [[CrossRef](#)]
19. He, Z.; Chen, W.; Zhang, J.; Wang, Y.H. Infrastructure crack segmentation: Boundary guidance method and benchmark dataset. *arXiv* **2023**, arXiv:2306.09196.
20. Qin, X.; Zhang, Z.; Huang, C.; Gao, C.; Dehghan, M.; Jagersand, M. Basnet: Boundary-aware salient object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 7479–7489.
21. Alshawi, R.; Hoque, M.T.; Ferdous, M.M.; Abdelguerfi, M.; Niles, K.; Prathak, K.; Tom, J.; Klein, J.; Mousa, M.; Lopez, J.J. Dual Attention U-Net with Feature Infusion: Pushing the Boundaries of Multiclass Defect Segmentation. *arXiv* **2023**, arXiv:2312.14053.
22. 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.

23. Lin, G.; Milan, A.; Shen, C.; Reid, I. Refinenet: Multi-path refinement networks for high-resolution semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1925–1934.
24. Milletari, F.; Navab, N.; Ahmadi, S.A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In Proceedings of the 2016 Fourth International Conference on 3D Vision (3DV), Stanford, CA, USA, 25–28 October 2016; pp. 565–571.
25. Yeung, M.; Sala, E.; Schönlieb, C.B.; Rundo, L. Unified focal loss: Generalising dice and cross entropy-based losses to handle class imbalanced medical image segmentation. *Comput. Med. Imaging Graph.* **2022**, *95*, 102026. [[CrossRef](#)]
26. Vinogradova, K.; Dibrov, A.; Myers, G. Towards Interpretable Semantic Segmentation via Gradient-Weighted Class Activation Mapping (Student Abstract). *Proc. AAAI Conf. Artif. Intell.* **2020**, *34*, 13943–13944. [[CrossRef](#)]
27. Kulkarni, S.; Singh, S.; Balakrishnan, D.; Sharma, S.; Devunuri, S.; Korlapati, S.C.R. CrackSeg9k: A collection and benchmark for crack segmentation datasets and frameworks. In Proceedings of the European Conference on Computer Vision, Tel Aviv, Israel, 23–27 October 2022; Springer: Berlin/Heidelberg, Germany, 2022; pp. 179–195.
28. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015; proceedings, part III 18; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.
29. 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.
30. Yu, C.; Gao, C.; Wang, J.; Yu, G.; Shen, C.; Sang, N. Bisenet v2: Bilateral network with guided aggregation for real-time semantic segmentation. *Int. J. Comput. Vis.* **2021**, *129*, 3051–3068. [[CrossRef](#)]
31. Hong, Y.; Pan, H.; Sun, W.; Jia, Y. Deep dual-resolution networks for real-time and accurate semantic segmentation of road scenes. *arXiv* **2021**, arXiv:2101.06085.
32. Li, Y.; Ma, R.; Liu, H.; Cheng, G. HrSegNet: Real-time High-Resolution Neural Network with Semantic Guidance for Crack Segmentation. *arXiv* **2023**, arXiv:2307.00270.
33. Shi, C.; Li, Y.; Cheng, Y.; Wang, Y.; Zhu, C.; Wang, K.; Zhang, C. Detection of Microcrack in Eggs Based on Improved U-Net. In Proceedings of the 2023 IEEE 9th International Conference on Cloud Computing and Intelligent Systems (CCIS), Dali, China, 12–13 August 2023; pp. 409–413.

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.