ISSN 2616-5775 Vol. 6, Issue 12: 33-44, DOI: 10.25236/AJCIS.2023.061204

An Improved Method for Digital Water Meter Reading Area Segmentation Based on U~2-Net

Rongrong Fei[1,a,*], Runze Cheng[1,b], Bin Yao[1,c], Feng Tian[1,d], Liuyang Gao[1,e], Zemu Men[1,f]

[1]School of Electronic Information and Artificial Intelligence, Shaanxi University of Science and Technology, Xi'an, China
[a]feirongrong@sust.edu.cn, [b]chengrunze666@163.com, [c]yaobin@sust.edu.cn, [d]723544546@qq.com, [e]supreme1060399828@163.com, [f]menzemu@yeah.net
[*]Corresponding author

Abstract: *With the continuous development of intelligent technology, the recognition of water meter readings has brought convenience to people's lives. By automatically reading water meter data, work efficiency can be improved. However, due to the varying shapes and sizes of the recognition areas, it is difficult to achieve good recognition results. Therefore, in response to the problem of varying sizes of water meter reading recognition areas, this paper proposes a method for digital water meter reading recognition area segmentation based on an improved U[2]-Net. Firstly, in terms of capturing data features, an improved Double-RSU module based on the RSU module is designed. This module increases depth and complexity compared to the original RSU module, thereby improving generalization and robustness. Secondly, in terms of model training, a combination of cross-entropy loss function, Jaccard coefficient and Dice loss function is used to comprehensively evaluate the entire model by considering binary classification, segmentation accuracy and segmentation area overlap. Experimental results show that the proposed method has good performance in MIoU, MAE and F1Score indicators, with the MIoU indicator increasing from 0.878 to 0.91, the MAE indicator increasing from 0.001 to 0.0009, and the F1Score indicator increasing from 0.9431 to 0.9455. These results indicate that the improved model has better performance compared to the original U[2]-Net model and can more accurately perform water meter reading area segmentation.*

Keywords: *Salient Object Detection, Deep Learning, Image Segmentation, U[2]-Net*

1. Introduction

Reading traditional mechanical water meters is an important and tedious task. Usually, water companies assign personnel to read the meter readings on site and record them in a dedicated ledger. This manual reading method is boring, inefficient, and easily affected by the physical and mental state of the staff, resulting in problems such as missed readings, estimated readings, incorrect readings, and even personal favor readings[1]. With the development of technology, thanks to the increasingly advanced computer and artificial intelligence technology and the improvement of computer hardware performance, the speed of computer image processing has made great breakthroughs. This has also enabled deep learning and computer vision to gradually enter various fields[2], making it possible to use intelligent devices for automatic meter reading of water meters, thereby avoiding the disadvantages of manual meter reading.

To eliminate the interference of digital characters in the non-reading area of the water meter dial on the recognition results of the water meter image reading, it is necessary to locate and segment the digit wheel area where the reading is located in the water meter image[3]. This task belongs to Salient Object Detection (SOD). Image salient object detection aims to simulate human visual mechanisms to identify and segment the parts of the scene that humans are interested in[4]. In the research of digital water meter reading area recognition, the reading area belongs to the part that humans are interested in.

In China, deep learning-based digital water meter reading area recognition technology has been widely applied. For example, a research team from Chongqing University of Posts and Telecommunications proposed a deep learning-based digital water meter reading area recognition method[5], which can quickly and accurately identify the reading area of digital water meters. In addition, a research team from Nanchang University proposed a deep learning-based digital water meter reading

ISSN 2616-5775 Vol. 6, Issue 12: 33-44, DOI: 10.25236/AJCIS.2023.061204

area recognition method[3], which can automatically adjust the recognition model to adapt to different environments and lighting conditions.

In foreign countries, deep learning-based digital water meter reading area recognition technology has also been widely applied. For example, in terms of segmentation algorithms, U-Net[6] is a very classic segmentation algorithm that uses an encoder-decoder structure and employs skip connections in the decoder part to better capture features. Later, scholars proposed a series of improved algorithms based on U-Net, such as improved U-Net, U-Net++[7], and so on. These algorithms have achieved good results in water meter reading area segmentation. In addition, some scholars have combined water meter reading area segmentation with other tasks, such as water meter reading recognition, water meter defect detection, etc. These methods can comprehensively understand the situation of water meters and improve the efficiency and accuracy of water meter management.

Currently, the digital water meter reading area recognition technology based on deep learning has been widely applied both domestically and internationally and is continuously being researched and developed. In the future, with the continuous advancement of deep learning technology and the widespread adoption of digital water management, the digital water meter reading area recognition technology will be applied and developed even more extensively.

Among the pixel segmentation methods currently used, U^2-Net[8] is a deep learning network structure designed by Qin et al. inspired by U-Net, which has better performance and effectiveness in the field of image segmentation and salient object detection tasks. This method proposes a new residual U-block (RSU) and uses a nested U-shaped structure, significantly improving the ability to extract multi-scale features from input feature maps of any resolution compared to the U-Net method[9].

Given the excellent performance of U^2-Net in the field of image segmentation, this paper proposes an improved algorithm based on U^2-Net for the task of digital water meter reading area segmentation, called Double-RSU U^2-Net. The RSU module of U^2-Net is optimized by introducing the Double-RSU module, which increases the depth and complexity, thereby improving the generalization and robustness of the algorithm. By integrating different loss functions, considering binary classification, segmentation accuracy, and overlap of segmented regions, the model is evaluated more comprehensively, thus improving the robustness and segmentation performance of the algorithm.

2. Related Work

2.1. Water Meter Image Dataset

The dataset mentioned in this article is a dataset for water meter image segmentation tasks, compiled by Kutsev Roman from TrainingData.pro in an article on Habr. The dataset contains 1244 water meter images and their corresponding 1244 segmentation masks, where the water meter images are RGB images and the masks are grayscale images.

In water meter image segmentation tasks, the goal is to segment the water meter reading part from the background in the water meter image. To achieve this goal, image segmentation algorithms are used to process the water meter image, resulting in a binary segmentation mask where the pixel values of the water meter reading part are 1 and the pixel values of the background are 0. The left image in Figure 1 is a water meter image, and the right one is a binary segmentation mask. In practical applications, this segmentation mask can be used for tasks such as automatic water meter reading and detection of water meter damage.

The water meter images in this dataset come from different sources, including different water meter brands, models, and shooting angles. The resolution and quality of these images also vary, providing a challenging dataset that can be used to evaluate the performance of different image segmentation algorithms.

ISSN 2616-5775 Vol. 6, Issue 12: 33-44, DOI: 10.25236/AJCIS.2023.061204

Figure 1: Digital water meter original image and mask.

2.2. Data preprocessing

During deep learning model training, it is usually necessary to preprocess the input data uniformly to enable training in the model. In this article, as the sizes of the images in the dataset are not consistent, they need to be uniformly set to 224×224 pixels before being input into the network model for training.

To achieve this goal, image processing libraries (such as OpenCV, Pillow, etc.) can be used to resize the water meter images to 224×224 pixels. At the same time, to ensure consistency between the training and testing data, the corresponding segmentation masks for each water meter image also need to be resized accordingly.

In addition, to convert the segmentation masks into binary images, thresholding can be used to set the pixel values greater than a certain threshold to 1 and the pixel values less than or equal to the threshold to 0. This can convert the segmentation masks into binary images, where the pixel values of the water meter reading part are 1 and the pixel values of the background part are 0.

By preprocessing and converting the dataset, it can be transformed into a format suitable for deep learning model training, while ensuring the consistency and usability of the data.

2.3. Data postprocessing

In water meter image segmentation tasks, it is usually necessary to overlay the segmentation mask with the original water meter image to determine the water meter reading area. Specifically, the areas with pixel values of 1 in the segmentation mask need to be overlaid onto the original water meter image, resulting in a new image that only contains the water meter reading area.

Figure 2: Selected reading area obtained by overlaying the mask with the original image.

After obtaining the new image shown in Figure 2, a Resize operation is performed to adjust its size

ISSN 2616-5775 Vol. 6, Issue 12: 33-44, DOI: 10.25236/AJCIS.2023.061204

to 224×224 pixels, which will serve as the input to the subsequent recognition network. When performing the Resize operation, it is necessary to ensure that the proportion and shape of the water meter reading area remain unchanged to ensure the performance and accuracy of the recognition network.

2.4. Basic Theory of U²-Net

In the field of salient object detection, Qin et al. [8] proposed U²-Net, a network that uses a two-level nested U-shaped network structure. This design uses the RSU structure[10] at the bottom level. The top level of U²-Net uses a structure similar to U-Net, with each layer filled with RSU blocks. The specific structures of RSU and U²-Net are shown in Figures 3 and 4, respectively. Using U²-Net to segment images has three advantages: it can capture more contextual information, the proposed RSU module can fuse features with different scale receptive fields; these RSU modules use a large number of pooling operations, which increases the depth of the entire architecture without significantly increasing the computational cost; the structure is only built on the RSU module and does not use any feature classification to train the backbone network, which is flexible, can adapt to different working environments, and has minimal performance loss[9].

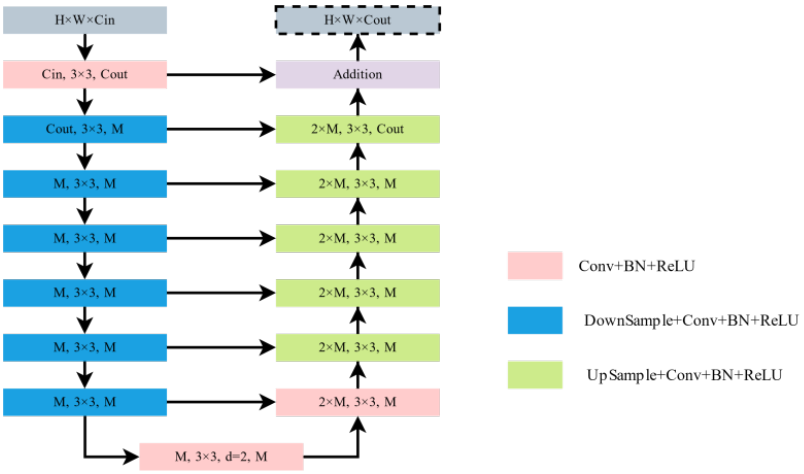

Figure 3: RSU Structure.

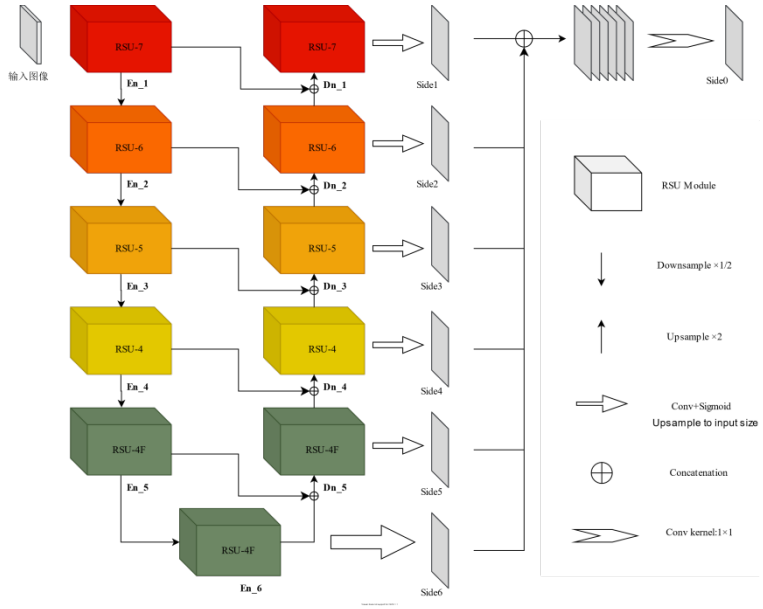

Figure 4: Top-level structure of U²-Net.

To meet different deployment requirements, the original paper proposed two versions of U²-Net: the standard version and the lightweight version. Table 1 shows the parameter distribution of each encoding and decoding block for the two versions. In each module, I represents the number of input channels, M represents the number of intermediate channels, and O represents the number of output channels.

ISSN 2616-5775 Vol. 6, Issue 12: 33-44, DOI: 10.25236/AJCIS.2023.061204

Table 1: Parameter distribution of encoding and decoding blocks for standard U²-Net and lightweight U²-Net†.

Architecture with different blocks	Stages										
	En_1	**En_2**	**En_3**	**En_4**	**En_5**	**En_6**	**De_5**	**De_4**	**De_3**	**De_2**	**De_1**
U²-Net	RSU-7	RSU-6	RSU-5	RSU-4	RSU-4F	RSU-4F	RSU-4F	RSU-4	RSU-5	RSU-6	RSU-7
	I:3	I:64	I:128	I:256	I:512	I:512	I:1024	I:1024	I:512	I:256	I:128
	M:32	M:32	M:64	M:128	M:256	M:256	M:256	M:128	M:64	M:32	M:16
	O:64	O:128	O:256	O:512	O:512	O:512	O:512	O:256	O:128	O:64	O:64
U²-Net†	RSU-7	RSU-6	RSU-5	RSU-4	RSU-4F	RSU-4F	RSU-4F	RSU-4	RSU-5	RSU-6	RSU-7
	I:3	I:64	I:64	I:64	I:64	I:64	I:128	I:128	I:128	I:128	I:128
	M:16	M:16	M:16	M:16	M:16	M:16	M:16	M:16	M:16	M:16	M:16
	O:64	O:64	O:64	O:64	O:64	O:64	O:64	O:64	O:64	O:64	O:64

2.5. Basic Theory of DoubleU-Net

To address the problem of class imbalance in medical imaging datasets, Debesh Jha et al. [11] proposed the DoubleU-Net based on the U-Net network and improved with VGG19[12] for medical image detection in 2020, which has strong robustness and generalization ability[13]. The network structure of DoubleU-Net is shown in Figure 5.

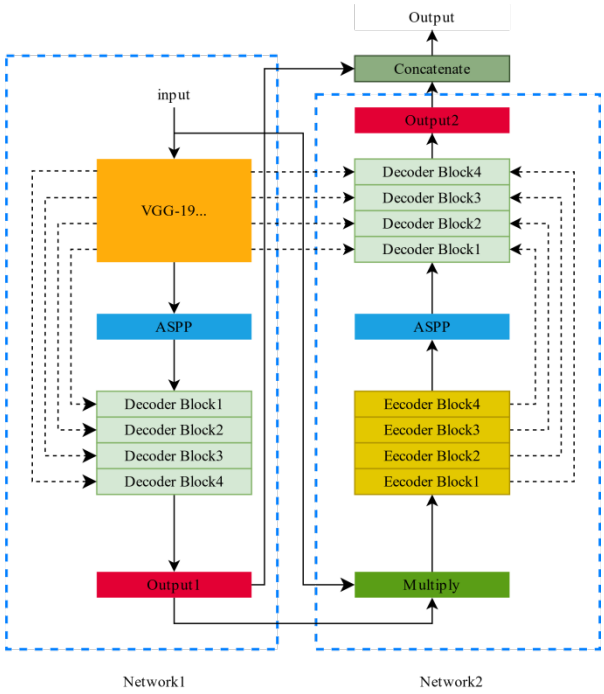

Figure 5: DoubleU-Net Structure.

The first encoder-decoder structure of DoubleU-Net uses VGG-19 as the feature extraction network. The extracted feature maps are passed through ASPP[14] (Atrous Spatial Pyramid Pooling) and then connected to the intermediate feature maps of VGG-19 before being input into the decoder block. The original input feature map is multiplied with the output feature map of the first encoder-decoder structure and serves as the input feature map for the second encoder-decoder structure. The second structure is similar to the first, but its decoder input is also associated with the output of the first encoder. The output feature maps of both structures are connected and passed through an output layer to obtain the final result.

The advantage of DoubleU-Net is that it can handle both low-resolution and high-resolution images simultaneously, thereby improving the accuracy and effectiveness of segmentation. Additionally, DoubleU-Net can further optimize segmentation results by adjusting the parameters and structure of the two U-Net networks.

ISSN 2616-5775 Vol. 6, Issue 12: 33-44, DOI: 10.25236/AJCIS.2023.061204

3. Double-RSU U^2-Net

The previous section introduced the dataset used in this paper, the preprocessing and postprocessing of data, and the model structure related to this paper. There are two improvements in this paper: one is the improved RSU module, and the other is the fusion loss function. This chapter will provide a detailed introduction to these two improvements.

3.1. Double-RSU Module

For salient object detection, it is easy to achieve accurate detection results when the texture and color features of the salient object are significantly different from the background. However, in real-world scenarios, there are often many cases where the foreground is similar to the background, which will cause the model to be unable to accurately identify the precise location of the salient object, resulting in a large number of false positive and false negative predictions, hindering the performance improvement of salient object detection models[7]. To solve this problem, this paper draws on the double encoder-decoder structure of DoubleU-Net and designs the Double-RSU module based on the RSU module. The specific structure is shown in Figure 6.

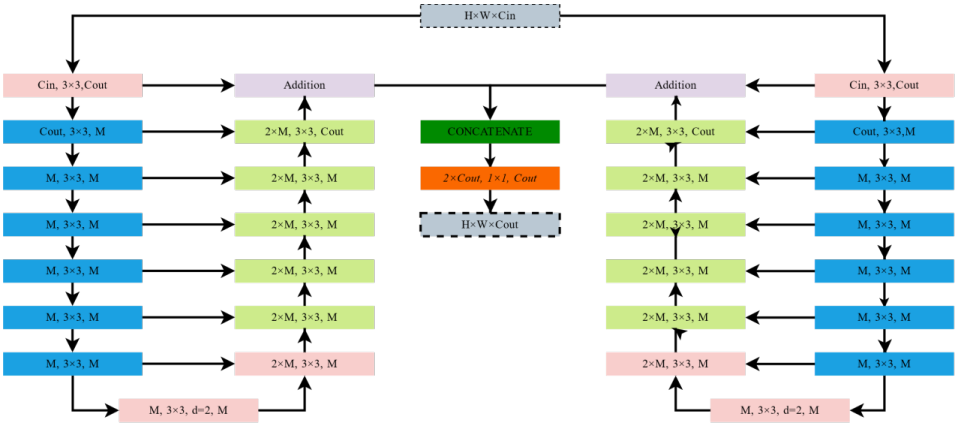

Figure 6: Double-RSU Module.

The Double-RSU module has two RSU structures, each of which learns the semantic features of the input feature maps. Their outputs are concatenated and passed through a 1×1 Conv layer to output the output feature map of the module. Due to its deeper network structure, the Double-RSU module can capture a larger receptive field, which can improve the model's understanding of contextual information in images. With two RSU modules, the Double-RSU module can extract more and richer features, which can help the model better distinguish foreground and background. With its deeper network structure and better feature extraction capabilities, the Double-RSU module can better generalize to new image data, which can help the model better adapt to different scenarios and tasks, thereby improving generalization and robustness.

3.2. Loss Function

U^2-Net uses BCE (binary cross-entropy) as the loss function, and the loss formula for U^2-Net is shown in equation 1:

$$L = \sum_{m=1}^{M} \omega_{side}^{(m)} l_{side}^{(m)} + \omega_{fuse} l_{fuse} \quad (1)$$

The loss function can be divided into two parts. One part is the loss between side1, side2, side3, side4, side5, side6, and the ground truth (GT) label image. Before calculating the loss, each side needs to be passed through a sigmoid activation function to obtain the corresponding probability map, which is $\sum_{m=1}^{M} \omega_{side}^{(m)} l_{side}^{(m)}$. The other part is the loss between the final fused probability map and the GT, which is $\omega_{fuse} l_{fuse}$. Here, l is the BCE loss, ω is the balance coefficient between the losses, and in the original paper, ω is equal to l for all losses, and M is equal to 6, which is side1 to side6.

However, for pixel-level segmentation tasks, the ratio of positive and negative samples is usually imbalanced. Using only the BCE loss function can easily lead to class imbalance problems, resulting in poor segmentation results for minority classes. Therefore, this paper proposes a fusion loss function that

ISSN 2616-5775 Vol. 6, Issue 12: 33-44, DOI: 10.25236/AJCIS.2023.061204

combines BCE, Jaccard, and Dice losses, denoted as BJD-LOSS. The Jaccard index (also known as the Jaccard similarity coefficient or IoU) is a statistical method used to compare the similarity between limited sample sets. The Dice coefficient is used to measure the accuracy of the prediction results. This fusion loss function comprehensively considers binary classification, segmentation accuracy, and segmentation region overlap, which more comprehensively evaluates the entire model and improves the robustness and effectiveness of the model. The proportions of each part can be adjusted according to the specific needs of the task to achieve better training results. The improved loss function is shown in equation 2:

$$L_{BJD} = \sum_{m=0}^{M} \omega_1 l_{jaccard}^{(m)} + \omega_2 l_{bce}^{(m)} + \omega_3 l_{dice}^{(m)} \quad (2)$$

Here, M is equal to 6, which includes side0 and side1 to side6. ω represents the proportion of each loss function, and in this paper, they are all set to 1.

4. Experimental Results and Analysis

4.1. Evaluation Metrics

This paper uses three evaluation metrics for experiments, including Mean Intersection over Union (MIoU), Mean Absolute Error (MAE), and F1Score.

First, several concepts need to be clarified: TP (True Positive) is a sample that is judged as positive but is actually negative; TN (True Negative) is a sample that is judged as negative and is actually negative; FP (False Positive) is a sample that is judged as positive but is actually negative; FN (False Negative) is a sample that is judged as negative but is actually positive.

Accuracy is the proportion of correctly classified samples among all samples, i.e., the proportion of correctly classified samples among all actual positive and negative samples. The formula is shown in equation 3:

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

Precision is the proportion of truly positive samples among all samples judged as positive, i.e., the proportion of truly positive samples among all samples judged as positive. The formula is shown in equation 4:

$$precision = \frac{TP}{TP+FP} \quad (4)$$

Recall is the proportion of truly positive samples among all actual positive samples, i.e., the proportion of truly positive samples among all actual positive samples. The formula is shown in equation 5:

$$recall = \frac{TP}{TP+FN} \quad (5)$$

MIoU is a commonly used metric for semantic segmentation, which represents the average intersection over union for each class in the dataset. The formula for calculation is shown in equation 6:

$$MIoU = \frac{1}{k+1} \sum_{i=0}^{k} \frac{TP}{FN+FP+TP} \quad (6)$$

MAE is one of the commonly used evaluation metrics in regression problems, which represents the average absolute error between the predicted values and the true values. It is used to measure the magnitude of the difference between the model's predicted results and the true values. The formula for calculation is shown in equation 7:

$$MAE = \frac{1}{k} \sum_{i=1}^{k} |h(x_i) - y_i| \quad (7)$$

F1Score is a metric used to measure classification problems. In some machine learning competitions for multi-classification problems, F1Score is often used as the final evaluation method. It is the harmonic mean of precision and recall, with a value between 0 and 1. The formula for calculation is shown in equation 8:

$$F_1 = 2 \times \frac{precision \times recall}{precision + recall} \quad (8)$$

ISSN 2616-5775 Vol. 6, Issue 12: 33-44, DOI: 10.25236/AJCIS.2023.061204

4.2. Data Augmentation

Data augmentation is a data preprocessing technique that generates more training data by applying a series of transformations and augmentations to the original data, thereby improving the performance and robustness of machine learning and deep learning models.

The data augmentation methods used in this article include scaling, distortion of length and width, and color space transformation. These data augmentation methods help increase the diversity of the dataset, thereby improving the model's robustness and generalization ability.

After data augmentation, each image in the original dataset was expanded into 8 images, resulting in a new dataset containing 11,196 images. Specifically, for each original image, scaling, distortion of length and width, and color space transformation were performed to obtain 8 new images, as shown in Figure 7. During data augmentation, the corresponding segmentation mask for each image also needs to undergo position transformation to ensure that the mask corresponds to the augmented image, which ensures consistency between training and testing data.

Figure 7: Original image (middle) and augmented images.

4.3. Experimental environment and training strategy

The experimental environment consists of Python 3.9, PyTorch 2.0.0+cu11.7, Windows 11 operating system, and an NVIDIA GeForce RTX 4060 graphics card.

The training strategy involves splitting the dataset before and after data augmentation into training, validation, and test sets in a ratio of 6:2:2. The batch size is set to 4, and the learning rate is set to 0.0001. The monitoring metric is set to mean intersection over union (MIoU), and only the model with the highest MIoU is saved during training. The patience is set to 30, which means that if the loss function does not decrease after 30 training iterations, the current model is considered optimal and the training is automatically stopped.

4.4. Comparative analysis of experimental results

Comparative analysis of the digital water meter reading area segmentation method based on U2-Net† in this article with other commonly used image pixel segmentation methods, including U-Net, PSPNet[15], UNet++, LinkNet[16], Deeplabv3[14], Deeplabv3+[17], MANet[18], and PAN[19]. To ensure fairness, all methods were evaluated using the same evaluation code.

4.4.1. Quantitative evaluation

Table 2 presents a comparison of the proposed method and the other 8 methods on different indicators on the old dataset (before data augmentation), as well as the number of parameters of each model. The

ISSN 2616-5775 Vol. 6, Issue 12: 33-44, DOI: 10.25236/AJCIS.2023.061204

results in black font indicate the best performance.

Table 2: Performance comparison of the proposed method and the other 8 methods on the old dataset.

Architecture	MIoU	MAE	F1Score	Params
U-Net	0.89	**0.0026**	**0.9391**	34527041
PSPNet	0.7994	0.0044	0.8816	**1501217**
UNet++	0.8749	0.0029	0.9358	26078609
LinkNet	0.8618	0.0034	0.9205	21771937
Deeplabv3	0.831	0.0037	0.9034	26007105
Deeplabv3+	0.8635	0.0031	0.922	22437457
MANet	0.8737	0.0029	0.928	31783633
PAN	0.8374	0.0036	0.9138	21475816
The proposed method	**0.8992**	**0.0026**	0.9321	3065613

As shown in the table, the proposed method outperforms other methods significantly in terms of the MIoU metric, almost reaching 0.9. The performance of the proposed method and U-Net is the best in terms of MAE, reaching 0.0026. Although the proposed method's performance in F1Score is not the best, it still ranks among the top three. The number of parameters in the proposed method is among the top three, with only 3,065,613 parameters. Although the number of parameters is about twice that of PSPNet, it is still a lightweight model, and PSPNet's other indicators are far inferior to the proposed method. The number of parameters in U-Net, which outperforms the proposed method in terms of MAE and UNet++, which outperforms the proposed method in terms of F1Score, is in the tens of millions, far greater than the number of parameters in the proposed method.

Table 3: Performance comparison of the proposed method and U^2-Net† on the old and new datasets, respectively.

Architecture	Old Dataset			New Dataset		
	MIoU	MAE	F1Score	MIoU	MAE	F1Score
U^2-Net†	0.8776	0.003	0.9254	0.9011	0.001	0.9431
The proposed method	**0.8992**	**0.0026**	**0.9321**	**0.9104**	**0.0009**	**0.9455**

Table 3 presents a comparison of the proposed method and the U^2-Net† model on different indicators on the old and new datasets (after data augmentation), with the results in black font indicating the best performance. After data augmentation, U^2-Net† showed an improvement in MIoU by 0.0235, a decrease in MAE by 0.002, and an improvement in F1Score by 0.0177. The proposed method showed an improvement in MIoU by 0.012, a decrease in MAE by 0.0017, and an improvement in F1Score by 0.0134. It can be obviously seen that data augmentation significantly improved the performance of both the proposed method and U^2-Net† in all indicators. The proposed method outperformed U^2-Net† by 0.0216 in terms of MIoU, decreased MAE by 0.0004, and improved F1Score by 0.0067 on the old dataset. Similarly, the proposed method also showed varying degrees of improvement over U^2-Net† on the new dataset. Therefore, the proposed method is superior to U^2-Net†.

4.4.2. Qualitative evaluation

In order to further verify the excellent performance of the proposed method, Figure 8 shows a visual comparison of the proposed method and other methods. The Image in Figure 8(a) and Figure 8(b) is the input original image of the water meter, and the Mask is the corresponding label. The other images are the segmentation results of various methods applied to the Image.

In Figure 8(a), other methods failed to effectively segment the reading area of the water meter and were greatly affected by the characters and numbers near the reading area, resulting in irregularly shaped segmentation results. The proposed method was not affected by noise and successfully segmented the reading area of the water meter.

In Figure 8(b), most methods only partially segmented the reading area, and were also affected by the QR code number in the water meter, which was mistakenly segmented as part of the reading area. For example, U-Net segmented the reading area into disconnected parts and also segmented the QR code number. The segmentation results of U^2-Net† and PAN were closer, as they segmented most of the reading area, but with some errors in other areas. The proposed method was not affected by noise or the QR code number and successfully segmented the reading area with good boundary details and continuity.

ISSN 2616-5775 Vol. 6, Issue 12: 33-44, DOI: 10.25236/AJCIS.2023.061204

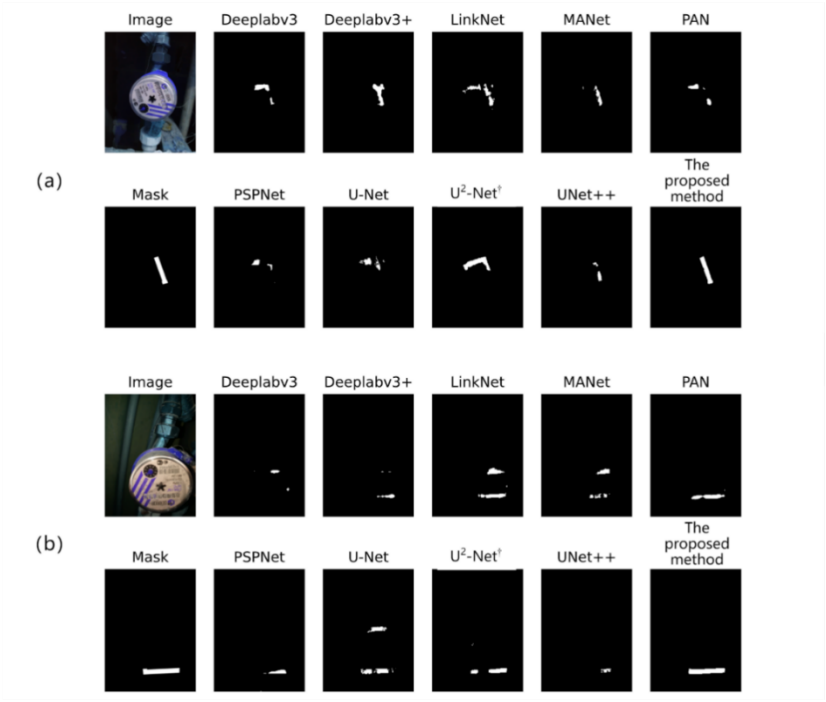

Figure 8: Visual comparison of the proposed method and other methods.

4.4.3. Result analysis

Both the quantitative and qualitative evaluations show that the proposed method has significant advantages over other commonly used pixel segmentation methods in segmenting the reading area of digital water meters. The quantitative evaluation intuitively shows the evaluation indicators of various algorithms' segmentation performance on the reading area in the old dataset, such as MIoU, MAE, F1Score, and the number of parameters. Through comprehensive comparison, the proposed method has a significant advantage. This is because the proposed method uses the Double-RSU module, where two groups of RSU structures are trained separately and then fused to obtain the output features, improving the robustness of the model. The proposed method also uses a fusion loss function for training, which more comprehensively evaluates the segmentation effect. The qualitative evaluation shows that the proposed method outperforms other methods in all aspects, with the most significant segmentation effect.

4.5. Ablation Study

To validate the effectiveness of the proposed Double-RSU module and BJD-LOSS, ablation experiments were conducted on the old dataset. The experimental results are shown in Table 4. First, Double-RSU and BJD-LOSS were separately introduced into U^2-Net†, and then both were introduced together into U^2-Net†. The experimental results are shown in the table, with boldface indicating the optimal results. From the table, it can be seen that after replacing the RSU module with the Double-RSU module, all indicators of the algorithm were improved. Similarly, there was a significant improvement in the algorithm when the loss function was replaced with BJD-LOSS. When both were introduced, the algorithm indicators were greatly improved, especially MIoU.

Table 4: Ablation experiments of different modules on the old dataset.

Double-RSU	BJD-LOSS	MIoU	MAE	F1Score
		0.8776	0.003	0.9254
✓		0.8874	0.0028	0.9298
	✓	0.8852	0.0029	0.9291
✓	✓	**0.8992**	**0.0026**	**0.9321**

Ablation experiments were conducted on the new dataset, and the experimental results are shown in Table 5. It can be seen from the table that after introducing the Double-RSU module, there was a significant improvement in MIoU and F1Score. However, after introducing BJD-LOSS, there was a slight decrease in both MIoU and F1Score. After introducing both, there was a significant improvement in MIoU and an improvement in MAE, while F1Score slightly decreased but was still better than the

ISSN 2616-5775 Vol. 6, Issue 12: 33-44, DOI: 10.25236/AJCIS.2023.061204

original model.

Table 5: Ablation experiments of different modules on the new dataset.

Double-RSU	BJD-LOSS	MIoU	MAE	F1Score
		0.9011	0.001	0.9431
✓		0.9083	0.001	**0.9479**
	✓	0.8985	0.001	0.9384
✓	✓	**0.9104**	**0.0009**	0.9455

4.6. Experiments on Optimization of Training Strategy

In order to find the optimal training strategy, this paper combined different Batch Sizes with different Learn Rates based on U^2-Net† and conducted training on the new dataset. The experimental results are shown in Table 6. According to the experimental results, the optimal Batch Size was determined to be 4 and the optimal Learn Rate was 0.0001. These two parameters were then used as the training strategy for the experiments in this paper.

Table 6: Experimental results of optimization of training strategy.

Batch Size	Learn Rate	MIoU	MAE	F1Score
4	0.0001	**0.9011**	**0.001**	0.9431
4	0.00001	0.9006	**0.001**	**0.945**
8	0.0001	0.8927	0.0011	0.9364
8	0.00001	0.8952	**0.001**	0.9353
16	0.0001	0.8839	**0.001**	0.9276
16	0.00001	0.817	0.0031	0.832
32	0.01	0.893	0.0011	0.9352
32	0.001	0.8986	**0.001**	0.9355
32	0.0001	0.8745	0.0013	0.9183

5. Conclusion

The aim of this study is to address the salient object detection task in water meter dial area detection, and propose an improved model named Double-RSU U^2-Net based on U^2-Net†. Comparative experimental results show that the Double-RSU U^2-Net model has significantly improved detection accuracy and robustness compared to the U^2-Net† model.

The improvement of the Double-RSU U^2-Net model is mainly reflected in the model structure. This paper uses two RSU modules and performs Concatenate operation at the end. This design can allow the model to better learn different features of the image and merge them together, thereby improving the accuracy of detection. In addition, this paper also expands the dataset, which is also an important factor in improving model performance.

Experimental results show that the Double-RSU U^2-Net model performs better than the U^2-Net† model on both the old and new datasets after dataset expansion. This indicates that the model has high detection accuracy and robustness, and can be applied to practical water meter dial area detection tasks. Additionally, the dataset expansion method proposed in this paper can provide reference for other related tasks.

References

[1] Shen J F. Research and application of water meter reading recognition method based on deep learning [D]. Jishou University, 2022.
[2] Xiong L. Water meter reading recognition method based on deep learning in real scene[D]. Nanchang University, 2022.
[3] Cui Y. Reading recognition method of digital water meter image based on deep learning[D]. Nanchang University, 2022.
[4] Li Y Z, Zhao J S. A review of image saliency object detection based on deep learning[J]. Software Engineering, 2023, 26(01): 1-4.
[5] Zhang L L. Research and implementation of water meter reading recognition algorithm[D].

ISSN 2616-5775 Vol. 6, Issue 12: 33-44, DOI: 10.25236/AJCIS.2023.061204

Chongqing University of Posts and Telecommunications, 2022.
[6] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation [J]. International Conference on Medical Image Computing and Computer-assisted Intervention, 2015, 9351: 234-241.
[7] Zhou Z W, Siddiquee M M R, Tajbakhsh N, et al. UNet++: A Nested U-Net Architecture for Medical Image Segmentation[C]. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, 2018, 11045: 3-11.
[8] Qin X, Zhang Z, Huang C, et al. U2-Net: Going deeper with nested U-structure for salient object detection [J]. Pattern Recognition, 2020, 106.
[9] Wang Y X, Ge H W. Metal surface defect detection algorithm based on U2-Net[J]. Journal of Nanjing University (Natural Sciences), 2023, 59(03): 413-424.
[10] Ye M, Li X C, Liu K, et al. An impedance imaging method based on the U2-Net model[J]. Chinese Journal of Scientific Instrument, 2021, 42(02): 235-243.
[11] Jha D, Riegler M A, Johansen D, et al. DoubleU-Net: A Deep Convolutional Neural Network for Medical Image Segmentation[J]. IEEE 33rd International Symposium on Computer-Based Medical Systems, 2020: 558-564.
[12] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition [J]. Computer Science, 2014.
[13] Hou X X. Research on highland lake extraction of medium-resolution remote sensing images based on DoubleU-Net[D]. Guangzhou University, 2022.
[14] Chen L C, Papandreou G, Schroff F, et al. Rethinking Atrous Convolution for Semantic Image Segmentation [J]. Computer Science, 2017.
[15] Zhao H S, Shi J P, Qi X J, et al. Pyramid Scene Parsing Network[J]. Computer Science, 2017: 6230-6239.
[16] Abhishek C, Eugenio C. LinkNet: Exploiting Encoder Representations for Efficient Semantic Segmentation [J]. IEEE Visual Communications and Image Processing, 2017.
[17] Chen L C, Zhu Y k, Papandreou G, et al. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation[J]. European Conference on Computer Vision, 2018, 11211: 833-851.
[18] Liang J Y, Sun G L, Zhang K, et al. Mutual Affine Network for Spatially Variant Kernel Estimation in Blind Image Super-Resolution[J]. IEEE/CVF International Conference on Computer Vision, 2021: 4076-4085.
[19] Li H C, Xiong P F, An J, et al. Pyramid Attention Network for Semantic Segmentation[J]. Computer Science, 2018.