Experimental results and analysis

In this experiment, the experimental results will be evaluated from two aspects: experimental data and subjective perception. Among them, the experimental data is used to evaluate the effect of the SegNet model on the training CamVid dataset through the above multiple evaluation indicators, such as different mIoU and accuracy curves. Subjective perception is to intuitively observe the advantages and disadvantages of SegNet network segmentation through the predicted images.

After changing the number of epochs, we found that when the epochs reached 50, the effect of the training model was basically stable and reached the best, and then the value of the epochs continued to increase, and the value of the evaluation index mIoU basically did not change much, stabilizing at about 60%. This shows that choosing the right number of epochs is critical to how well the model is trained. If there are too few epochs, the model may not have learned the patterns in the training data sufficiently, resulting in underfitting. On the other hand, if there are too many epochs, the model may overfit the training data, i.e., the model is too sensitive to noise and detail in the training data, resulting in poor performance on unseen data. In order to avoid overfitting, we continuously increase the value of epochs after epochs is 50, and find that the mIoU value does not decrease and remains stable, and the Loss value also remains stable, so we can stop training and retain the model weight at this time. The mIoU-epoches curve is shown in Figure 3.2, and the Loss-epoches curve is shown in Figure 3.3.

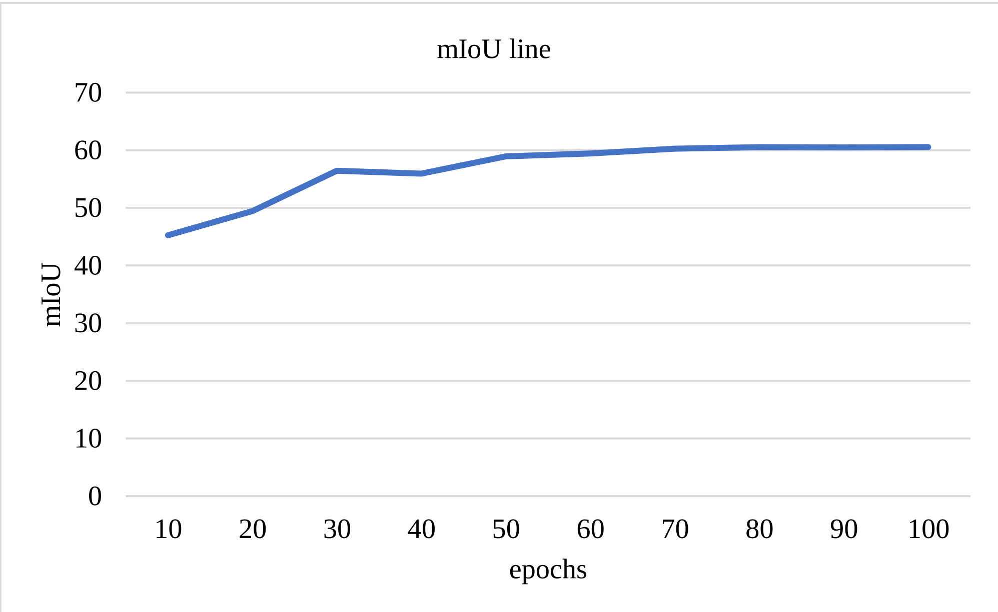

figure 3.2 mIoU-epoches line

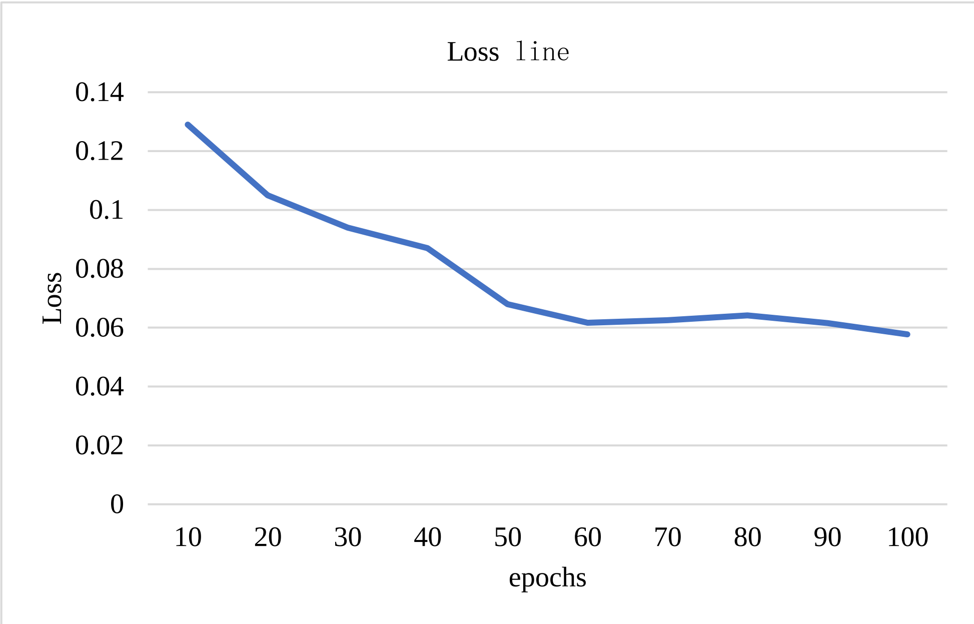

figure 3.3 Loss-epoches line

After understanding the optimal epochs for training, we take the epochs value as 50 and export the values of each indicator generated by the training, and get the following results:

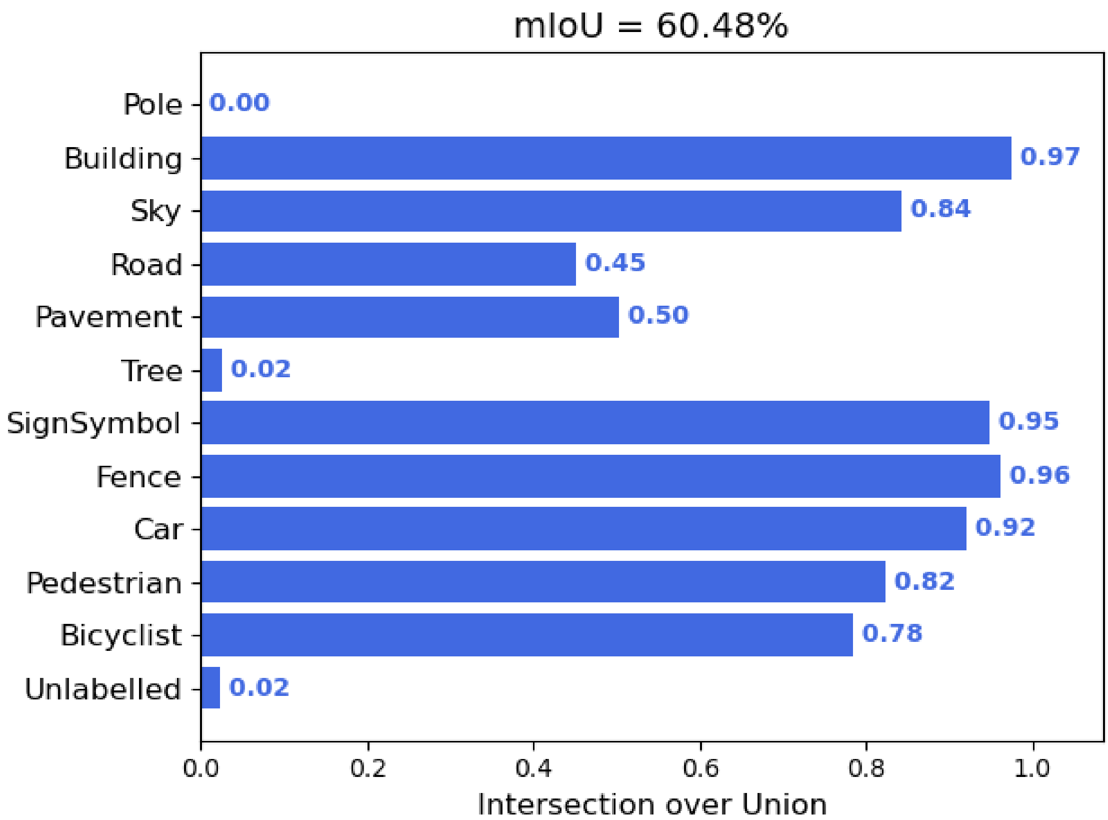

Figure 3.4 Io of each tag when epochs=50

As can be seen from Figure 3.4, the intersection and union ratios of most of the 12 categories in the CamVid dataset are very high compared to IoU. Among them, the IoU value of the four categories of Building, SingSymbol, Fence, and Car can even reach more than 90%; The intersection ratio of Sky, Pedestrain, and Bicyclist can also be more than 75%; The rest of the Road and Pavement are basically 50 percent combined; The overall average intersection of the model is more than 60% higher than that of mIoU. It is generally agreed that in a computer detection task, if the result is acceptable, then the detection is said to be correct. This shows that the classification effect of the model is still considerable.loU≥0.5

However, there are several tag categories that deserve our attention, such as Pole and Tree, which have very small intersection and union ratios, and the Pole category is even 0, which is mainly due to the following three reasons:

First, the dataset selected in this experiment is not large, and the label categories in the dataset are not sufficient when learning the model, which will lead to the inability to learn the categories of some labels, so the intersection union ratio will be small or even zero.

Second, in some cases, the size of the target object in the image itself is small. Since the prediction box needs to cover the entire target object, even if the positioning of the prediction box is relatively accurate, the intersection area of the prediction box and the real box will be small because the target object itself is small, resulting in a low IoU value.

Third, the model will prefer a more efficient and convenient classification method in the process of training, if there is only a small part of a class in recognition, the model will be more inclined to classify this small category into a large block category that is more convenient to identify, in layman's terms, that is, the model will occasionally be "lazy", which will also lead to a low IoU value of some categories.

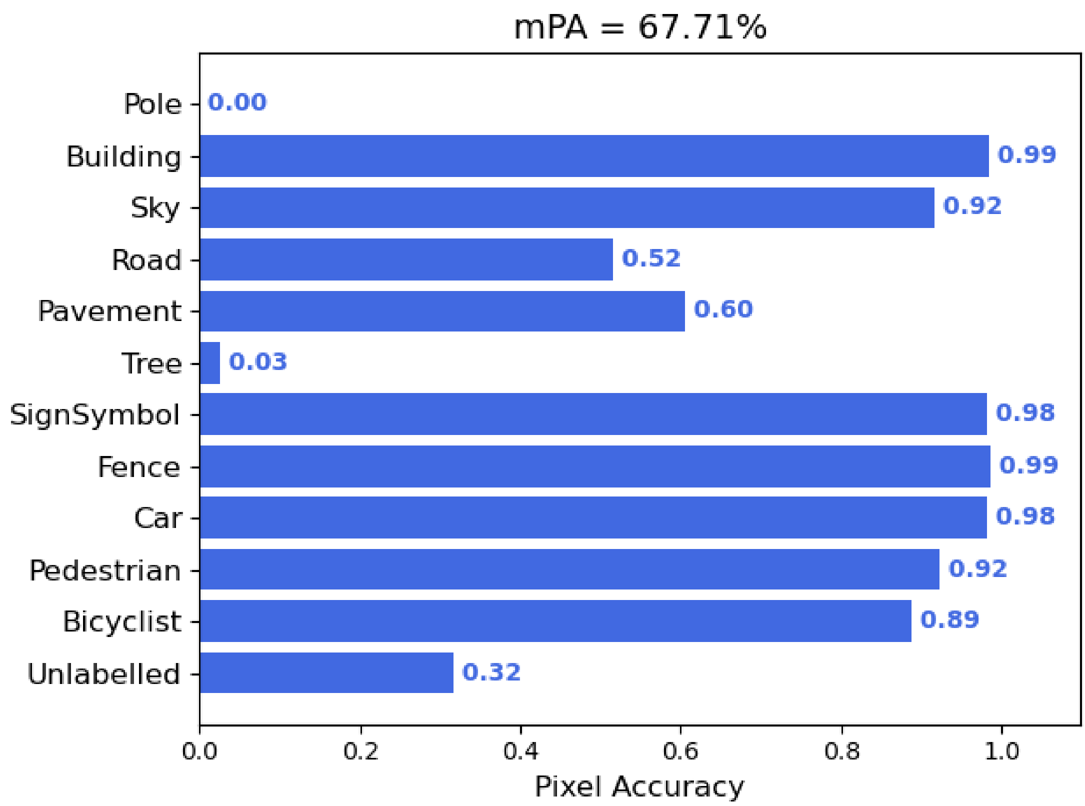

Figure 35 The PA value of each label when epochs=50

The PA value reflects the ratio of the number of pixels that the model predicts correctly to the total number of pixels. The higher the PA value, the higher the number of pixels that the model predicts correctly, and the better the model performance. For example, in Figure 3.5, the PA values of several categories such as "Building", "Sky", "SignSymbol", "Fence", "Car", and "Pedestrain" are all very high. However, PA values are susceptible to category imbalances in the image. For example, if a category occupies the vast majority of pixels in an image, the PA value may be high even if the model's prediction performance on that category is poor.

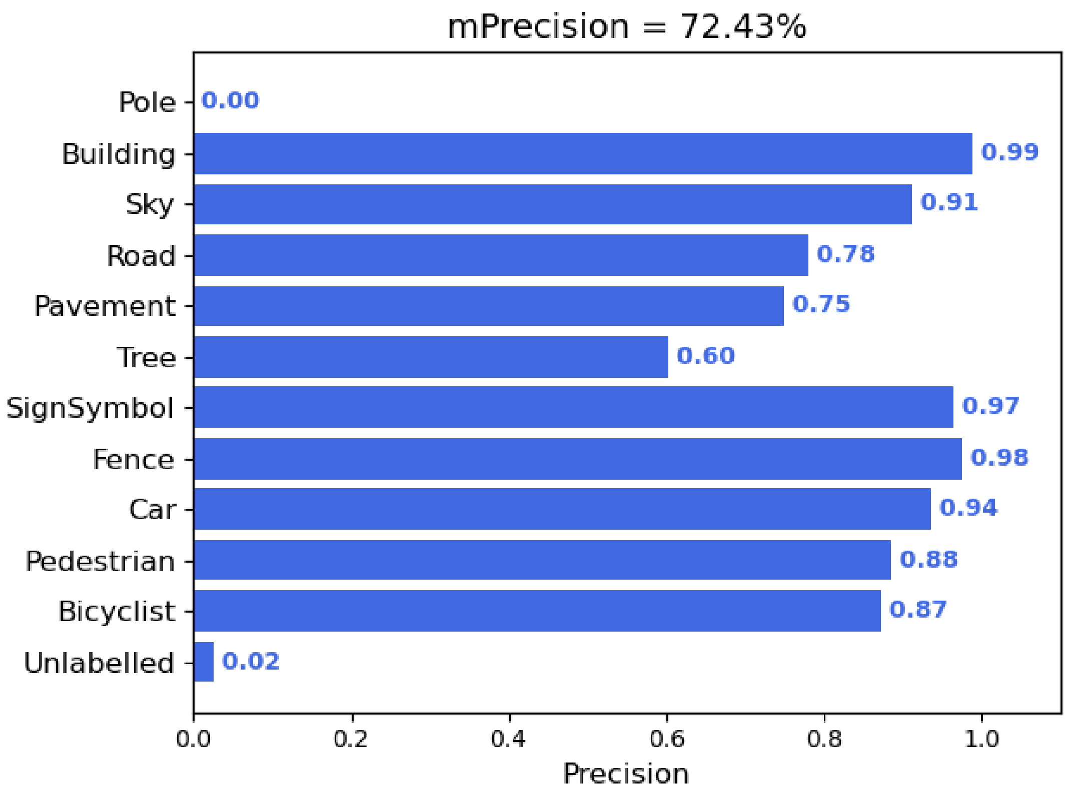

Figure 36 epochs=50 The Precision value of each tag

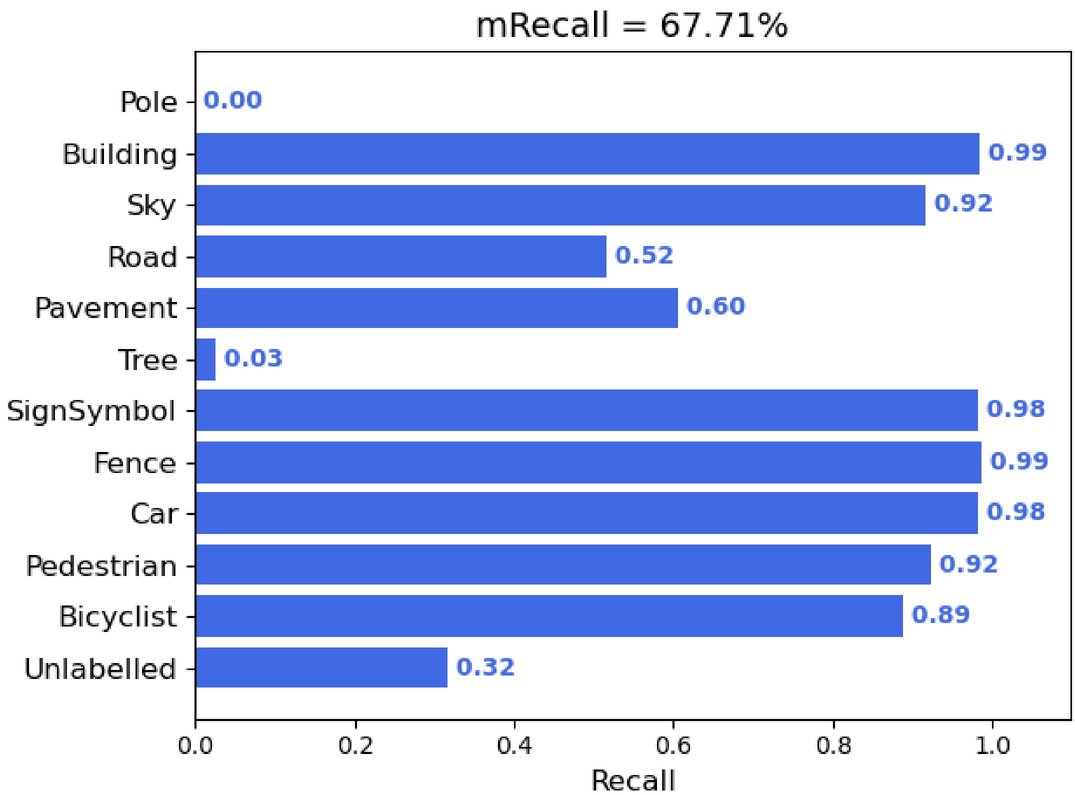

Figure 37 The Recall value of each tag when epochs=50

The higher the Precision value, that is, the higher the precision, the better the prediction result of the model. At the same time, the higher the precision of the model, the fewer false positives. However, in some cases, relying solely on precision may not be comprehensive, because when the negative sample in the dataset is much larger than the positive example, the model may predict all the samples as negative, resulting in very high precision, but a low recall. As shown in Figure 3.6, the accuracy of the "Tree" classification is 60%, but as can be seen from Figure 3.7, the recall rate of the "Tree" label category is very low, only 3%.

The higher the recall value, the higher the recall rate, the stronger the model's ability to identify positive examples, that is, the more the model can find all the real positive examples. However, in some cases, relying solely on recall may also be incomplete, because when the model is too lenient in predicting a sample as a positive example, the recall is high, but the precision can be low, resulting in a large number of false positives in the prediction results. As shown in Figure 3.7, the recall rate for "Unlabelled" is 32%, but its accuracy rate is only 2%. There are also several types of samples with very low recall, such as "Pole" and "Tree", which may be affected by the large difference in the number of positive and negative samples in the dataset. Sample classification with a high recall rate can also be the result of false positives.

It can be seen that precision and recall are two mutually restrictive indicators, and in general, improving precision may reduce recall, and vice versa. Therefore, in practice, these two indicators need to be weighed according to specific needs.

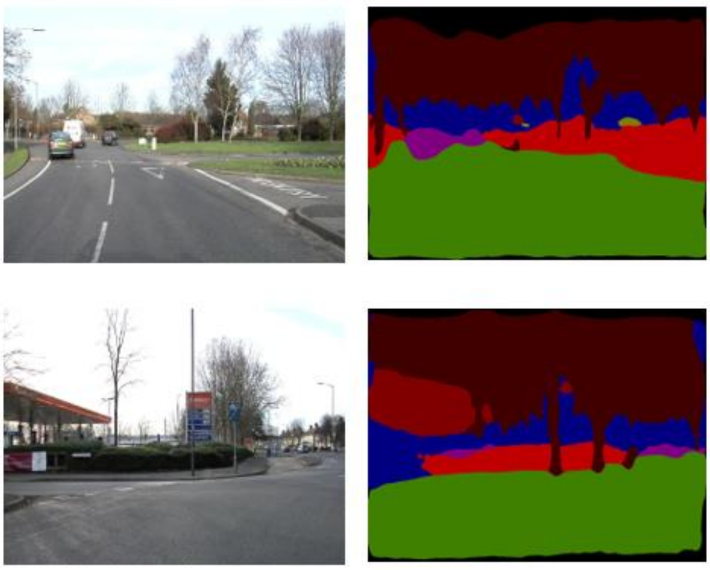

Figure 38 Random road scenario prediction results

From the results of Figure 3.8, it can be observed that the effect of the SegNet model is still ideal for the prediction of road scenes, and the model can clearly distinguish the obvious categories such as roads, grass, and cars, and the segmentation edges are relatively smooth. This is a significant improvement over the segmentation effect that SegNet first proposed.

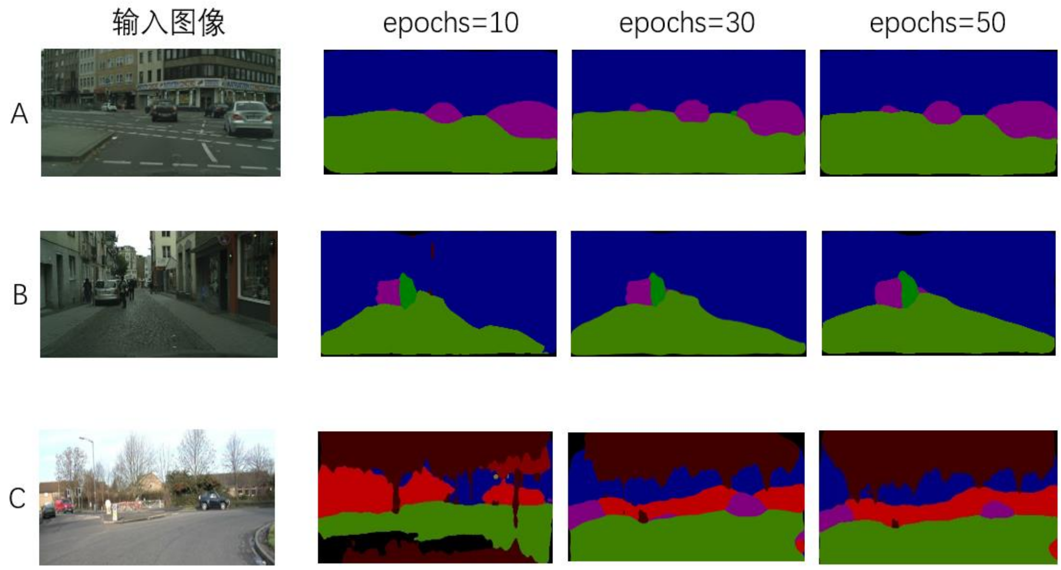

Figure 39 Predictions under different epochs

In this experiment, the weights under different epochs values are imported when predicting the images, and it can be found that when the epochs values are small, the prediction images of the model are still relatively abstract. There are many burrs on the edge of the segment, which is not clear and meticulous enough; Occasionally, there will be misidentification, such as recognizing a large tree as a background. In Figure 3.9, we can see that in the input image A, when the epochs value is 10, the car in front of the building in the far left is almost not recognized, the edges are blurry, and the segmented image is small. When we increase the epochs to 20, we find that the model can identify the car more clearly, and when the epochs are increased to 50, the segmentation edges are smoother and more detailed. It is obvious to the naked eye that the effect of segmentation gradually improves as the epochs increase. The same is true in Figure B, where the edge of the split between the pavement and the building becomes smoother as the epochs value increases. In Figure C, it can be seen that the epochs are very small, the road surface is widely misrecognized, the trees are also largely unrecognized, and the car is almost not recognized at all, and the segmentation effect is particularly poor. However, when the epochs increase to 20 or 50, these poor results get better little by little, and finally produce a more ideal segmented image.

Finally, the training efficiency of the model was analyzed. From the perspective of time, the time of training is basically controlled within one hour, and it can be seen that compared with PSPnet and FCN, the time efficiency of the model has been greatly improved, which proves the effectiveness of the SegNet model.

This chapter first describes the process of building a SegNet network, including the frameworks and functions used. The training experiment is carried out using the built network, and the experimental results are analyzed, and the conclusion shows that the SegNet network has certain advantages.

Results from: Tang Xinya

Findings from: Xinya Tang