CEG5304 C A 4

Zhang Bokun

A0263813E

1. Architectures

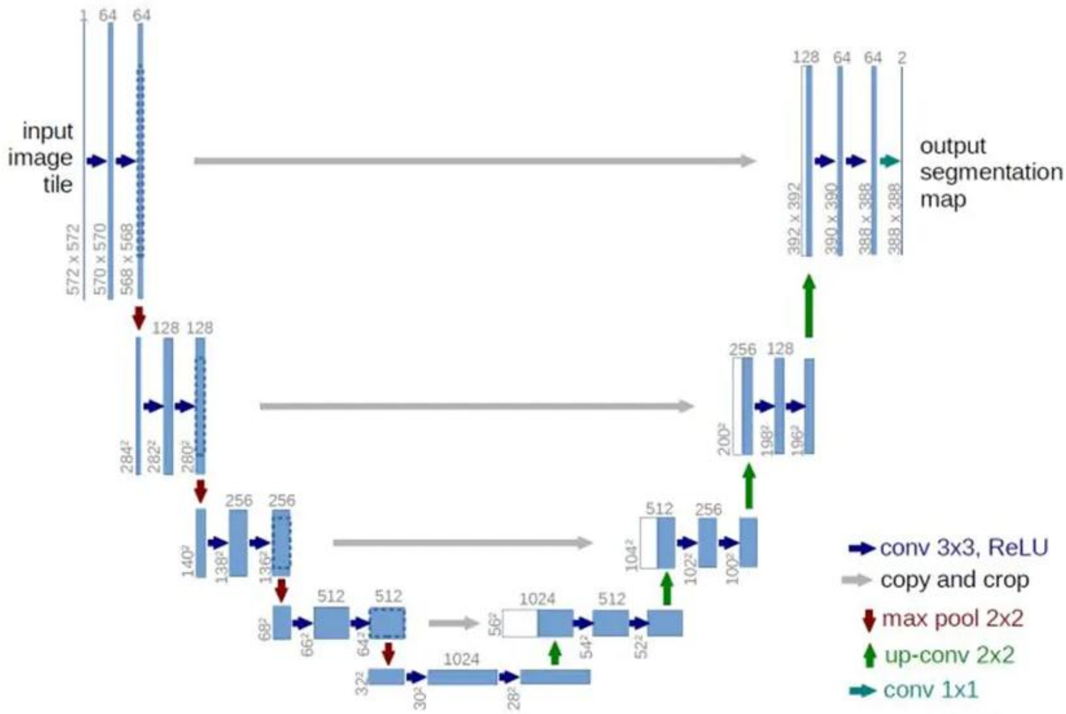


Figure 1 U-Net structure

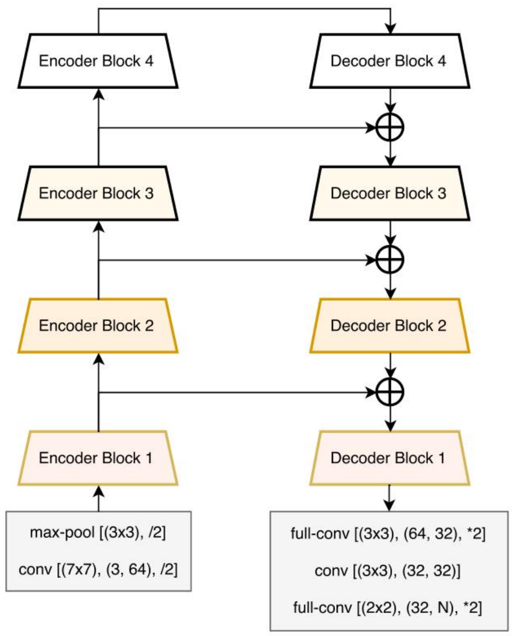
U-Net structure:

U-Net is a deep learning structure commonly used for image segmentation. It consists of an encoder and a decoder. The encoder is responsible for gradually Downsampling the input image and extracting high-level features, while the decoder is responsible for gradually upsampling and restoring the resolution, mapping high-level features back to the pixel level of the original image to generate segmentation results.

In the Oxford IIIT Pet Dataset, U-Net can be used for semantic segmentation of pets in images, assigning each pixel to one of the three categories of "pets", "background", and "boundaries". Specifically, the U-Net encoder can extract features related to pets from images, such as edges, textures, and shapes, while the decoder can generate accurate segmentation results based on these features. In addition, U-Net can also solve possible blurring and discontinuity problems in segmentation results through techniques such as skip connections.

Code and pretrained models are available at:
https://github.com/yassouali/pytorch-segmentation

Link-Net structure:



LinkNet is a type of neural network architecture used for semantic segmentation tasks. Semantic segmentation is the process of assigning a label to each pixel in an image, with the goal of dividing the image into regions that correspond to different objects or parts of objects.

LinkNet works by using a series of convolutional and pooling layers to extract features from the input image. These features are then passed through an encoder-decoder network that predicts a segmentation mask for each pixel in the image.

One of the key features of LinkNet is the use of skip connections between the encoder and decoder layers. These skip connections allow the decoder to access features from earlier stages of the encoder, which can be useful for preserving spatial information and improving the accuracy of the segmentation.

LinkNet also uses a novel upsampling technique called upsampling blocks, which involve upsampling the feature maps using transposed convolutions and then applying a series of convolutional layers to refine the results.

2. How it works

U-net

The U-Net architecture consists of an encoder and a decoder network, with skip connections to improve the accuracy of the segmentation. The encoder network consists of a series of convolutional layers, followed by max-pooling layers to downsample the feature maps. The decoder network consists of a series of upsampling layers, followed by convolutional layers to generate the final segmentation mask. The skip connections allow information from the encoder to be directly passed to the decoder, helping to preserve spatial information and reduce the effect of information loss during the downsampling process.

The U-Net architecture is trained using a dataset of images with corresponding pixel-level labels. During training, the network learns to predict the correct class for each pixel in the image. Once the network is trained, it can be used to perform semantic segmentation on new images by passing them through the network and generating a segmentation mask.

Overall, the U-Net structure is an effective approach for semantic segmentation in the Oxford-IIIT Pet Dataset, allowing for accurate classification of each pixel in the image into one of three classes - "pet", "background", or "border".

Link-net

LinkNet is a type of neural network architecture used for semantic segmentation tasks. Semantic segmentation is the process of assigning a label to each pixel in an image, with the goal of dividing the image into regions that correspond to different objects or parts of objects.
LinkNet works by using a series of convolutional and pooling layers to extract features from the input image. These features are then passed through an encoder-decoder network that predicts a segmentation mask for each pixel in the image.

One of the key features of LinkNet is the use of skip connections between the encoder and decoder layers. These skip connections allow the decoder to access features from earlier stages of the encoder, which can be useful for preserving spatial information and improving the accuracy of the segmentation.

LinkNet also uses a novel upsampling technique called upsampling blocks, which involve upsampling the feature maps using transposed convolutions and then applying a series of convolutional layers to refine the results.

3. Training

Figure 2 Accuracy and Loss-U-net

We can see that we got a very good learning result, accuracy is over 80% and loss is under 3.5%.

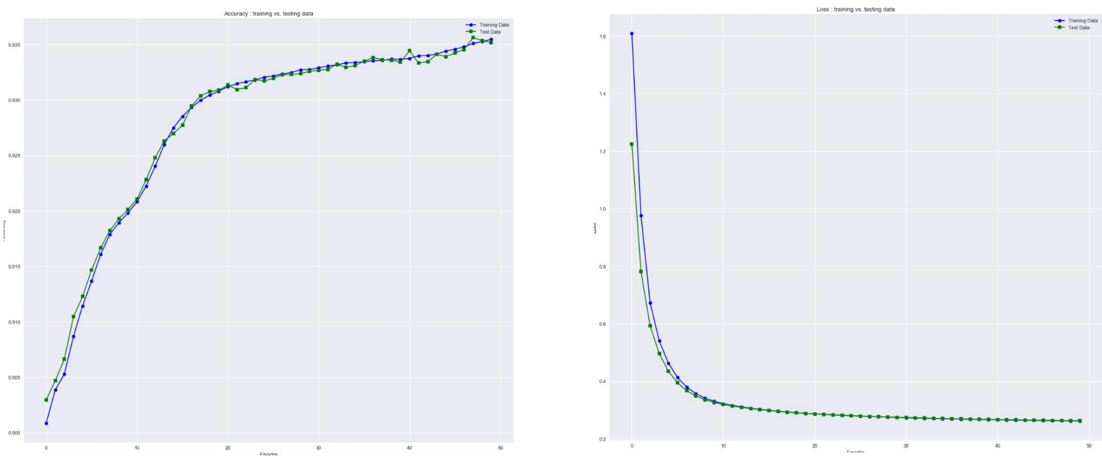


Figure 3 Accuracy and Loss-Link-net

We can see that we got a learning result, accuracy is over 90% and loss is under 15%.

4. Testing

In this part, I use two different network models, U-net and link-net to test. And all models are training myself.

Cartoon

Figure 4 Part2-5-1

Figure 5 Part2-5-1-seg-U-net

Figure 6 Part2-5-1-seg-link-net

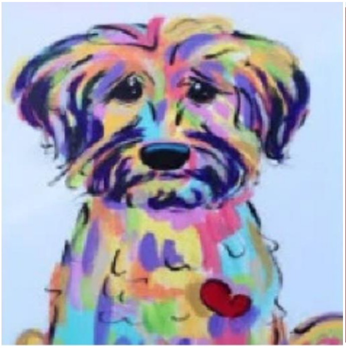


Figure 7 Part2-5-2

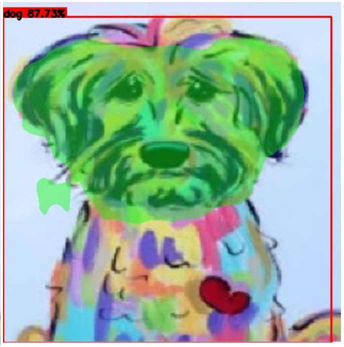


Figure 8 Part2-5-2-seg-U-net

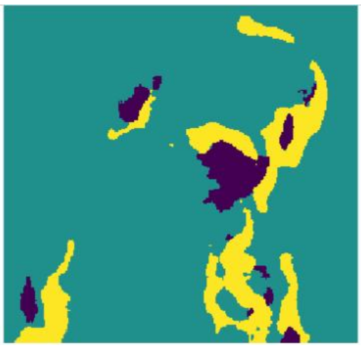


Figure 9 Part2-5-2-seg-link-net

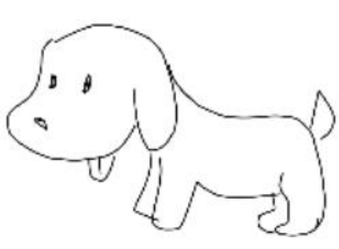


Figure 10 Part2-5-3

Figure 11 Part2-5-3-seg-U-net

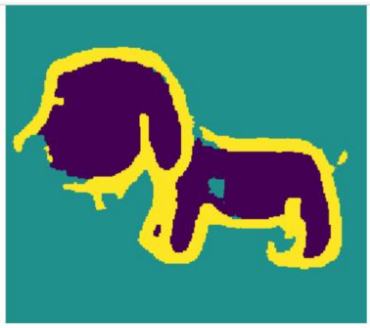


Figure 12 Part2-5-3-seg-link-net

In cartoon part, We can see that U-net is difficult to recognize the body parts of cartoon pets, but it can effectively recognize the face. Once the face of a cartoon pet is small, it is difficult to segment the pet from the background.

Moreover, in link net, the body parts of cartoon pets are better recognized, but their faces cannot be well segmented from the background.

The vital reason is that we do not have any cartoon dataset for training.

Real

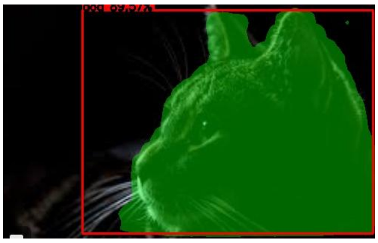




Figure 13 Part2-7-1 Figure 14 Part2-7-1-seg-U-net Figure 15 Part2-7-1-seg-link-net



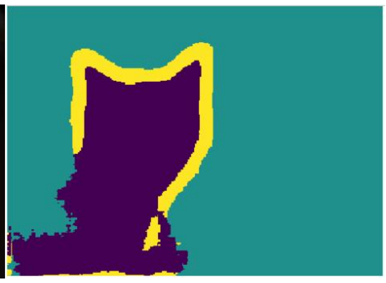
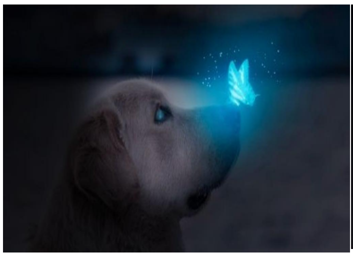


Figure 16 Part2-7-2 Figure 17 Part2-7-2-seg-U-net Figure 18 Part2-7-2-seg-link-net



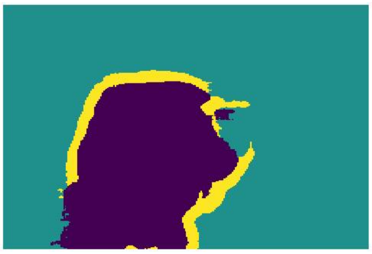


Figure 19 Part2-7-3 Figure 20 Part2-7-3-seg-U-net Figure 21 Part2-7-3-seg-link-net

Obviously, U-net is good at recognize the real pets pictures and link-net is good at cartoons. I think the reason is that U-net is focus on pet's face and link-net is concentrate on pet's body.

Thus in this real part, U-net almost perfectly segment the pets from the background. But link-net cannot do that.

5. Solutions

Segmentation of images taken at night can be challenging due to factors such as low light, noise, and lack of contrast. However, there are several approaches that can be used to address this problem:

Data augmentation: One way to improve the performance of segmentation models on nighttime images is to use data augmentation techniques, such as brightness adjustments or adding artificial noise. These techniques can help the model learn to recognize features that are specific to nighttime images and improve its ability to generalize to new, unseen nighttime data.

Pre-processing: Pre-processing techniques such as denoising, contrast enhancement, and image filtering can help to improve the quality of nighttime images before they are fed into the segmentation model. These techniques can help to reduce the impact of noise and increase the visibility of important features in the image.

Transfer learning: Transfer learning is a technique in which a pre-trained model is used as a starting point for a new task. In the case of nighttime segmentation, a pre-trained model that has been trained on daytime images can be fine-tuned on nighttime images to improve its performance.

Use of infrared imaging: Infrared imaging can provide better visibility of nighttime scenes, as it is not affected by ambient light and can detect temperature differences in the scene. Using infrared imaging can improve the quality of input images for segmentation models, leading to more accurate results.

Multi-modal fusion: In some cases, combining information from multiple sources can improve the accuracy of nighttime segmentation. For example, combining visible light images with thermal imaging or using other types of sensors can provide additional information that can help the model to distinguish between objects in low-light conditions.

Overall, there are several approaches that can be used to address the challenges of nighttime segmentation, including data augmentation, pre-processing, transfer learning, use of infrared imaging, and multi-modal fusion. The best approach will depend on the specific application and the available data and resources.