# Anime Character Segmentation and Alpha Chanel Reconstruction

Zepeng Liu† Hiroshi Watanabe†

Waseda University†

**Abstract**: The alpha channel is essential for graphic design and game production. In general, the alpha channel often retains a lot of transparency information, making it easy to remove the background. With alpha channel, designers can move the character position arbitrarily without considering the changing background. However, since the alpha channel is invisible, it is often discarded in real life (e.g. photography, printing or screenshots). For illustration, one anime character in an illustration that has lost the alpha channel is hard to reuse. This study is dedicated to reconstructing the lost alpha channel of illustrations in order to extract anime characters and make them can be reused.

## 1 Introduction

Alpha channel is very important auxiliary data of images. Especially in the field of graphic design and game production, the alpha channel retains a lot of transparency information, making it easy to remove the background and be reused in various scenes. However, because alpha channel can only be saved in the computer storage and is invisible in the real world, it is often be lost in various situations. For example, save a picture by photography, printing, screenshot or converting a picture's format to JPEG. One anime character in an illustration that has lost the alpha channel is hard to reuse. So, we have proposed a way to reconstruct the lost alpha channel.

Although the reconstruction problem sounds very difficult, in a way, it is only a sub-topic of semantic segmentation. Just think that the alpha channel represents a region with a high opacity of 1, while the other regions are 0, which becomes a two-class semantic segmentation problem. But this topic is not exactly the same as semantic segmentation, it requires very high precision and not just the general location of the target.

To achieve this goal, we created a custom dataset. The semantic segmentation model is based on U-Net, which is a representative FCN (Fully Convolutional Networks).

## 2 Dataset

This customized dataset (Fig. 1) contains 3588 images, of which 3408 are training data and the remaining 180 are test data. All the images are standings (日本語:立ち絵)

Figure 1: Samples of the dataset

contained alpha channel, they are extracted from various games or downloaded from sharing websites. For each picture, the RGB channel is used as the input and the alpha channel is the target.

Data Augmentation is necessary because the original dataset is too little to successfully train a network. Moreover, a reasonable data augmentation can effectively strengthen the robustness and invariance of the network.

When training, we adopted rotation, horizontal flip, zoom and shift to extension our dataset. Special, we replace the background color of the image with a certain probability. Empirically, the white background is 40%, the light background is 40%, the dark background is 10%, and the black background is 10%.

## 3 Network

We use U-Net [1] as the Network. U-Net is a typical FCN model, without full connected layers, it can accept arbitrarily input size. [2] U-Net have successfully used in many advanced researches [3] [4].

## 4 Training

The optimizer is Adam with β1=0.5, β2=0.999. The learning Rate is 0.0002, input size is 256 x 256 and batch size is 16. The adversarial loss is not used, it is just a least MSE (mean squares error) function. In order to ensure convergence, we trained this model for 1000 epochs in 5 days by one RTX 2080 TI.

## 5 Result

We find that although traditional semantic segmentation problems tend to use PA or IoU as benchmarks, their numerical differences are too small on this topic. PSNR is a more stringent metric to evaluate the quality of the resulting images.

Note that although our model is trained at 256 x 256, it does a good job in test set with the resolution of 512 x 512, which proves that it has some scaling invariance.

In Fig.2, we show 2 results, their inputs are both 512 x 512. The 1st one almost perfectly restored the image with the PSNR of 49.6896, PA of 0.9998 and IoU of 0.9989. The 2nd one mis-mapping skirt which the same color as background. Still, it has a PSNR of 16.2556, PA of 0.9751 and IoU of 0.9295.

Table 1: Test Result

| | PSNR | PA | IoU |
|---|---|---|---|
| 256×256 | 36.9945 | 0.9973 | 0.9855 |
| 512×512 | 33.8018 | 0.9947 | 0.9785 |

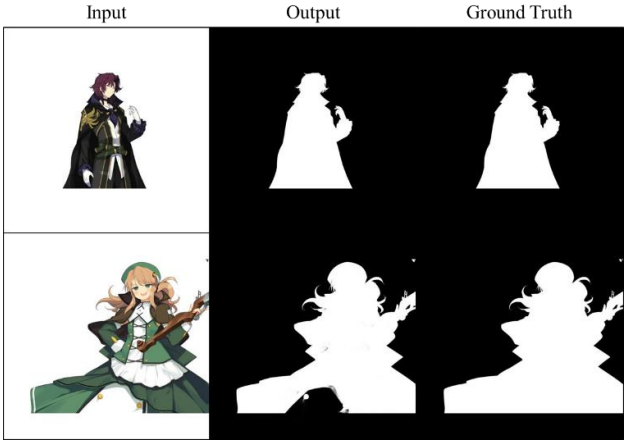


Figure 2: Samples of the Results

## 6 Conclusion

Our research explores the possibility of using semantic segmentation to solve alpha channel reconstruction problem. It is very promising for removing the background of illustrations in web for reusing in graphic design and game production.

But now, some issues still need to be resolved. For example, the scaling invariance is limited. Then, because of memory limitations, it is difficult for the network to process large pictures.

## References

[1] Olaf Ronneberger, Philipp Fischer, Thomas Brox: “U-Net: Convolutional Networks for Biomedical Image Segmentation”, arXiv:1505.04597 [cs.CV]

[2] Jonathan Long, Evan Shelhamer, Trevor Darrell: “Fully convolutional networks for semantic segmentation”, arXiv:1411.4038 [cs.CV]

[3] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, Alexei A. Efros: “Image-to-Image Translation with Conditional Adversarial Networks”, arXiv:1611.07004 [cs.CV]

[4] Jun-Yan Zhu, Taesung Park, Phillip Isola, Alexei A. Efros: “Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks”, arXiv: 1703.10593 [cs.CV]

*Communications and Computer Eng., Waseda Univ.*
*ADVANCED MULTIMEDIA SYSTEMS LAB.*
*Shillman Hall 401, 3-14-9 Okubo, Shinjuku, Tokyo, 169-0072 Japan*
*Phone: +81-3-5286-2509, Fax: +81-3-5286-3488*
*E-mail: takuho@ruri.waseda.jp*