Paper Reading Report-02

Han Zhang
u7235649

Abstract

*This is my reading report for the paper titled: "***Real-Time High-Resolution Background Matting***", authored by Shanchuan Lin et al, and published in CVPR 2021.*

All ENGN8501 submissions will be subject to ANU's TurnitIn plagiarism check, against both the original paper, internet resources, as well as all other students' submissions. So please make sure that you must write your own reports, and declare the following originality statement:

I, Han Zhang, hereby confirm that I am the sole author of this report and that I have compiled it in my own words.

1. Problem Statement

The research problem the paper attempts to address is to allow users to replace the background of the video in real time without extra input or special backgrounds such as a green screen, so as to improve entertainment or enhance privacy.

Background replacement can be achieved through methods such as segmentation or matting. The segmentation-based method runs very fast, but it lacks in resolution and detail. Although the traditional trimap-based matting method has better performance, it cannot run in real time at high resolution and often requires manual input by the user. The matting method without any external input has many limitations. The method of matting with known natural backgrounds has a better quality but cannot achieve real-time and high-resolution.

The method proposed in this paper can perform background replacement in real-time and high resolution without extra input from users, and in the same time give a better quality.

2. Summarise the paper's main contributions

The method proposed in this paper can provide high-quality background replacement in real-time at high resolution and does not require any extra input by the user. This method is not traditional segmentation or trimap-based matting, but uses two neural networks and an additional captured background image.

However, real-time processing and good results cannot be achieved on all devices, it requires good hardware support. This hardware limitation can be balanced by reducing the matting quality.

3. Method and Experiment

The training of the method proposed in this paper is based on multiple data sets, and the pictures on the network are crawled to generate the data set.

This method is based on two neural networks. The first is the basic network to provide low-resolution matting to the entire image. This is a fully-convolutional encoder-decoder network with three modules: Backbone, ASPP and Decoder, to estimate alpha matte and foreground residual. The second refinement network performs high-resolution detail processing on the first k parts with higher prediction errors in the results of the basic network, so as to provide higher quality and increase running speed.

During training, the authors used a variety of data augmentation techniques to avoid over-fitting and adapting to the real situation, and cropped in each batch to support any resolution and aspect ratio. In the verification, synthetic data sets and real pictures were used.

The authors found through experiments that the use of the AIM dataset will worsen the results due to low resolution and small number of samples. Also, in the basic network, using ResNet-50 as the encoder backbone could be more efficient.

4. Critical Analysis

4.1. Are the paper's contributions significant?

I think the contribution of this paper is significant. The traditional segmentation or trimap-based matting methods and other methods have insufficient matting quality and speed, or require additional input from the users, thus cannot work at high resolution in real time. In contrast, the method proposed in this paper changes the thinking and uses neural networks, so as to achieve real-time, high-resolution, and high-quality background matting.

In addition, it is hard to obtain a large number of high-quality and high-resolution matting data sets. The authors

collected, organized and published many data sets, which will be of great help to people who will study this topic in the future.

4.2. Are the authors' main claims valid?

The authors validated their main claims. They introduced the structure of their neural network and the data set used in detail, described the process of their experiment, and from the given result pictures, compared with other methods, the effect of the method they proposed is better than the real-time matting methods and some non-real-time methods methods. They also did a user research, and the result showed that most people thought the method they proposed was more effective.

4.3. Limitation and weaknesses

Their method needs to capture additional background photos, and needs to ensure that the background only moves slightly or does not move.

Also, to achieve real-time, high-resolution, high-quality background matting requires great hardware support, which is limited by the current hardware. However this can be balanced by reducing the extinction quality.

In addition, there are some other common limitations, like when the subject is projected heavily on the background or the color is similar to the background, or the background has a strong texture, this method may fail.

4.4. Extension and future work

I think the extra work is mainly aimed at the limitations and weaknesses mentioned above.

For the problem that additional background pictures are required and the background cannot be moved significantly, due to the relative motion of the subject and the background, the part of the occluded background can be displayed at different times. And usually the background movement is continuous, and the backgrounds of adjacent frames will overlap, so the backgrounds should be able to be estimated and spliced by comparing similar parts. I think they can try to compare several adjacent frames of pictures, with the help of image similarity matching methods and recognition algorithms (such as human body recognition) to identify the approximate subject and background positions, thereby stitching and matte the background.

For other problems where the algorithm may fail under certain backgrounds, more and better training samples and additional training may be required for better results.

4.5. Conclusion and personal reflection

This paper is exciting. The authors proposed a new direction to solve the background matting problem, that is, to use a two-layer neural network to achieve real-time, high-resolution and high-quality processing, which is a great progress. The first layer of the neural network performs rough processing on the entire image, and the second layer performs detailed processing on the parts with large prediction errors in the results of the previous layer, thereby improving the quality and speed, which is an inspiration to me. I think this idea can also be applied to other multi-layer neural networks.

To solve this problem, I may combine image similarity algorithm and human body recognition algorithm on the basis of neural network to recognize and stitch the background, so as to improve the adaptability to the moving background or even avoid using additional background images.

References

[1] Shanchuan Lin, Andrey Ryabtsev, Soumyadip Sengupta, Brian Curless, Steve Seitz and Ira Kemelmacher-Shlizerman. *Real-Time High-Resolution Background Matting*. CVPR 2021.

[2] Peter Rogelj and Stanislav Kovacic. *Local Similarity Measures for Multimodal Image Matching*. IWISPA 2000. Proceedings of the First International Workshop on Image and Signal Processing and Analysis. in conjunction with 22nd International Conference on Information Technology Interfaces.