DOI: 10.53469/wjimt.2024.07(05).08

Matting Image Background Switching Application

Yueqi Shen, Dan Zhong

Jincheng College, Sichuan University, Chengdu 610000, Sichuan, China

Abstract: ***With the development of the times, people are increasingly pursuing an ideal lifestyle, and frequent photography often makes it difficult for people to get rid of the modified "perfect" image in the images. Crop, as a commonly used image processing technique, plays an important role in beautifying images. This article will analyze and compare this image cropping technique in depth, and briefly introduce its application methods. From a technical perspective, this cropping algorithm is independent of Trimap technology. By performing mask cropping on the image, it effectively extracts details and ultimately generates a more natural image with background changes.***

Keywords: Trimap; Image cutout algorithm; Mask algorithm.

1. MATTING CUTOUT OVERVIEW

Matting is a traditional way of image segmentation, and traditional (non learning) Matting methods typically require Trimap as input[1-3]. They can be roughly divided into sampling based techniques and propagation based techniques. The sampling based method uses sampling to establish color statistics of known foreground and background, and then solves the Matte of the "unknown" region[4-9]. The propagation based method refers to solving the Matting equation by propagating Matte from the foreground and background regions to the "unknown" region[10-12].

This article is based on the Matting image cutout algorithm, which performs a series of feature extractions by inputting background image B without foreground and original image I with foreground to obtain relatively good cutout results. Further merge with the newly added background to obtain the final image[13]. Trimap technology is mainly used for image segmentation and image editing tasks. It helps algorithms identify the boundaries between foreground and background in an image by labeling each pixel as foreground, background, or uncertain region. The image generated by this technique is a binary image, typically composed of three different values: foreground, background, and uncertain region, hence it is called a "trimap". There are various methods for generating trimap, including color threshold based segmentation, edge based segmentation, depth information based segmentation, and machine learning based segmentation[14]. Regardless of the method used, it is necessary to perform post-processing on the generated trimap to ensure that it accurately reflects the regional information in the image. The application of Trimap technology makes image segmentation and editing tasks more precise and efficient. Wu, Z. (2024). introduces the Meta-Path Guided Attention Aggregation Network (MPAAGN), which integrates meta-paths, attention mechanisms, and GraphSAGE for efficient node classification in heterogeneous networks, making it applicable to any large-scale graph data requiring advanced relationship modeling and node classification [15].

2. MATTING CUTOUT

2.1 Dataset

This dataset is mainly in the form of images, which are composed of software synthesis, handheld photography of daily life materials, and other methods. It mainly focuses on image selection and shooting for portraits, while also incorporating images of animals, plants, and other images for algorithmic analysis (for static images).

Input: Background image B 'with no foreground (using the. back. png extension), original image I (using the. img. png extension), and background image.

First output: Roughly segmented image S (using the _masksDL.png extension).

Second output: Final synthesized image (using _compose. png extension), foreground cutout (using _fg. png extension), masked image (using _matte. png extension), optimized segmented image (using _out. png extension)

The above are all the images that will appear in the dataset, based on the data from the original input folder and output file.

2.2 Specific Implementation Process

The process of this algorithm is based on a deep Matting network, and the biggest difference from other Matting image extraction algorithms is that it cancels the use of Trimap by taking an additional background image without a theme during photography. Although it increases the difficulty of shooting, it greatly reduces the training time of Trimap and achieves good results, optimizing the entire algorithm [3].

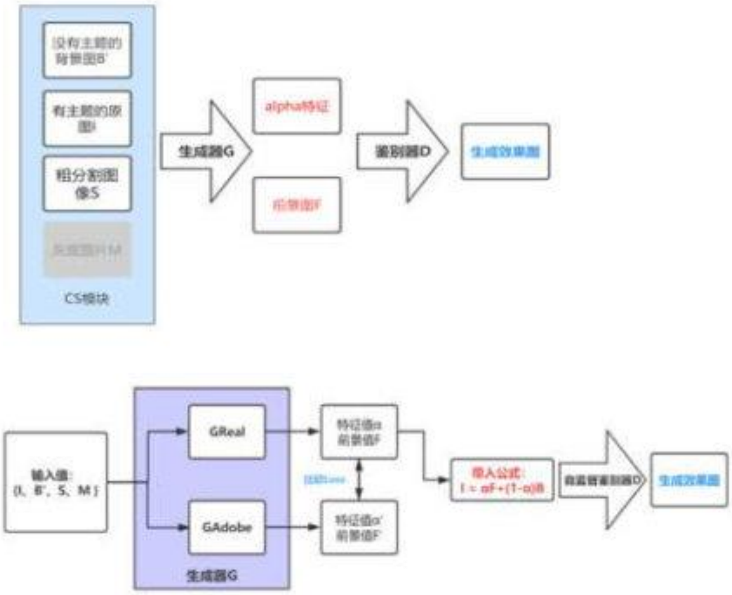

Figure 1: Algorithm specific flowchart

Firstly, by inputting a dataset of background image B 'without a theme and original image I with a theme, the images are subjected to erosion dilation and Gaussian blur operations to generate a coarse segmented image S of the character in the input folder. Here, M can be understood as having no video and is set to {I, I, I, I}, converted to grayscale values.

The three inputs obtained are used for feature extraction through the encoder of generator G, and a discriminator D is used to guide training and generate real results.

The CS module is used in the input process, and the weight of each input is calculated based on the degree of fit between the input image and the background. For example, if the theme and background fit well, the weight mainly exists in the processing of graphic edges; If the theme is affected by background lighting, the weight mainly lies in adjusting the theme lighting, etc. Choosing the CS module can effectively compare the weights of different features to obtain the optimal selection effect. In the CS module, four feature values form three channel features obtained from 1x1 convolution, BatchNorm, and ReLU during the mapping process. These features are combined with the channel features of the original image and trained under supervision in the Abobe dataset to obtain the feature value α and foreground value F. If the loss value is within a controllable range by comparing the true value with the processed value α and F, it will be passed to the remaining blocks and decoders [2].

If there is no CS module to assist in training the Adobe dataset, it will often result in errors in the obtained network. Although it greatly reduces the gap between real images and synthetic images created using Adobe datasets to some extent, it still poses certain difficulties compared to directly processing actual images. Therefore, based on the fusion of foreground F and background B', a self supervised discriminator D is formed to determine the authenticity of the graphics [1].

The self supervisor D is generated from supervised training on the Adobe dataset, and is a self supervised part like Real. It is a GAN network that performs authenticity self supervised discrimination on theme images generated in different backgrounds, extracting the most realistic foreground F.

Due to the fact that the most basic image extraction algorithms extract foreground F, background B, and feature value α and substitute them into the formula:

$I = \alpha F + (1 - \alpha) B$

F represents the color of the foreground at (x, y), and B represents the color of the background at (x, y). α is the transparency of the foreground, with a value range of 0 to 1

For this equation, we are not familiar with α, F, and B, so it has always been considered a pathological equation that cannot directly obtain parameter values through calculation. Instead, algorithms are used to continuously calculate and compare images to obtain the true values of these parameters.

Through the above steps, we obtained the values of α, F, and B, and extracted the desired foreground image.

2.3 Results Analysis and Comparison

2.3.1 Comparison of Results from the Original Dataset

From the image in the first row of Figure 2, there is no significant contour difference, so it is effective for foreground segmentation. But when the image is enlarged, it can be seen through comparison that the coarse segmented 0012MaskSDl.png image is far inferior to the depth loss optimized network processed 0012Mout.png in terms of details. The latter focuses on the processing of subtle parts, accurately segmenting the fine details of hair strands naturally. By observing the depth and color of the hair strands, it can be seen that this effect is mainly the result of the light extinction network during cutting, which makes the entire cut image more layered when combined with other background images.

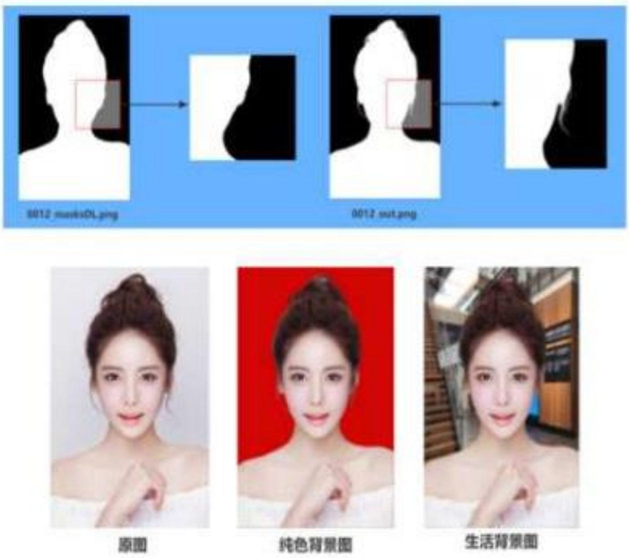

Figure 2: Image comparison before and after algorithm optimization

After changing to a new background, it can be seen that whether it is a solid color background or a living background, the extracted image is integrated with the background. At the same time, it also reasonably cropped the background images with different sizes to meet the needs. On this basis, we have developed a certain curiosity about the content of the images themselves and conducted comparative research on them.

We selected a side view of the human face from the dataset for our study as an example and conducted the following analysis.

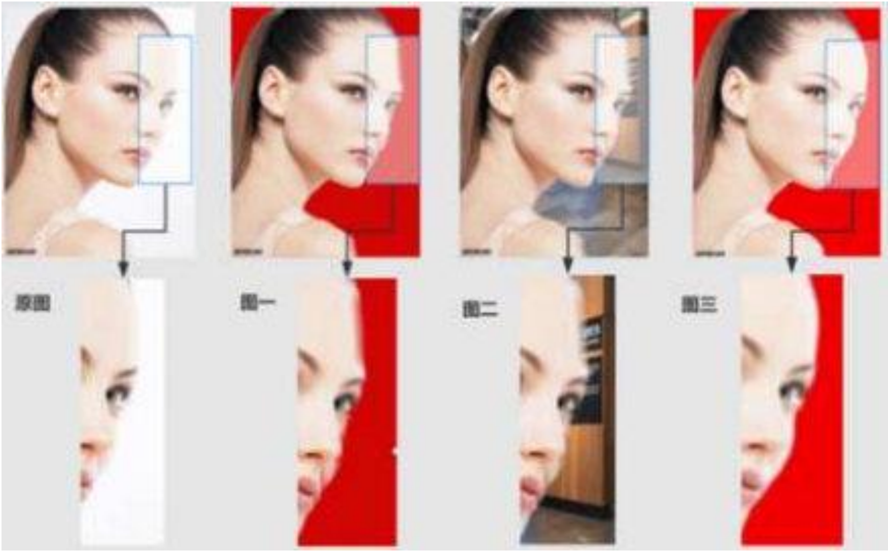

Figure 3: Comparison of Background Switching Details in Side View Original Image

The first image in Figure 3 is the original image, Figure 1 is a solid color background image, Figure 2 is a life background image, and Figure 3 is an optimized version of the solid color background image.

Firstly, following our standard operating procedure, we replaced some background images to combine different backgrounds, resulting in the images shown in Figure 1 and Figure 2. Through observation, there is an overfitting phenomenon around the facial contours in Figures 1 and 2, which is caused by blindly fitting the background image and eliminating edge boundaries, resulting in the distortion mentioned above.

Figure 3 shows appropriate handling, clear edge boundaries, and beautiful contour lines. This is because, based on the lessons learned from Figure 1 and Figure 2, we began to search for the root of the problem. We found that starting from the original image, due to the high gloss on the side of the character in the original image, the edges of the portrait are fundamentally similar in color to the background whiteboard, making it difficult to segment. Therefore, in the process of optimizing foreground segmentation (Figure 1), it may not be as good as rough foreground segmentation. Taking into account the overall situation, we have not yet adopted a certain approach or method for foreground segmentation. Therefore, we have started to modify the post image by adopting a similar approach to alpha post-processing, which involves setting a certain threshold to make the image a binary image. Then, we combine the binary image with the foreground to obtain the effect shown in Figure 3.

And next, we will try to observe some single or multiple person cutout methods:

We conducted a cutout analysis of single and double individuals standing and sitting down. By comparison, it can be found that using the original threshold is a good choice when extracting images of the human body, and can achieve good background fitting through image extraction.

In the above processing, it was found that the algorithm cannot perform image cutout on some flat or real portrait images. Currently, it cannot be determined whether it is due to the inability of the original image to segment foreground and background or the inability to recognize images other than the human body. We are also constantly exploring and working together with everyone.

2.3.2 Comparison with Photoshop cutout

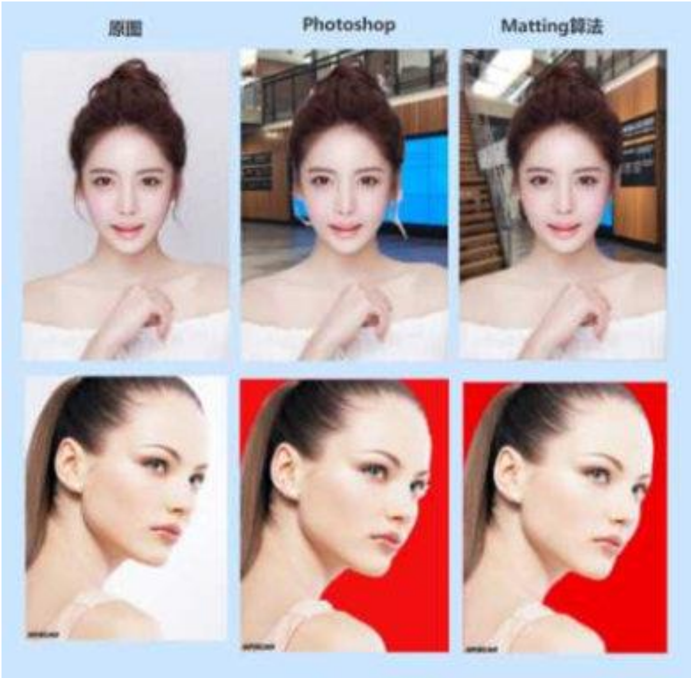

Figure 4: Comparison of cutout effects between Photoshop and Matting

Observing the left image, comparing the three images in the first group, we found that the images processed in Photoshop have a rough background on the hair strands, and the original background fragments around the hair are clearly visible, which greatly affects the overall aesthetics. After Matting algorithm processing, the hair in the image is clearly visible, which is significantly better than Photoshop, and the foreground and background are well integrated. We found that the advantage of Photoshop is that it preserves the pixels of the original image, making the portrait appear clearer; The Matting algorithm has affected the original image quality due to the use of matte processing. This reflects their respective advantages and disadvantages.

The main problem with the second set of images is that the segmentation of the portrait is not natural enough. As mentioned earlier, this set of images, due to the difficulty in finding boundaries when cropping, was ultimately optimized by adjusting the parameters. Observing the image processed by Photoshop, the problem also exists at the facial edges, where the jagged segmentation lines can be clearly seen, making the image appear more rigid.

So overall, the main problem with Photoshop is improper image and background processing, which makes the segmented images very stiff. However, the Matting algorithm performs better based on this point, not only processing the edges of the image, but also processing both the image and background simultaneously. Maintaining the original clarity of images in Photoshop to make them more vivid is also something that Matting algorithm needs to learn.

2.3.3 Comparison with Paddlehub image cutout algorithm

Paddlehub is a popular way for audiences to simultaneously capture multiple images. Compared to Photoshop, it greatly reduces manpower and time. Its intelligent module can recognize the theme well and intelligently extract images. At the same time, the simple code clearly and concisely indicates the meaning.

Next, we will compare Paddlehub's image cutout method with Matting's image cutout method

Figure 5: Comparison of Paddlehub and Matting cutout images

The image processed by the left paddlehub in Figure 5 clearly shows that the processing is relatively rough. It cannot effectively process fine hair strands, so its algorithm directly removes the unprocessed images as backgrounds during the processing; By processing the character edges of the image through Paddlehub, it can be seen that the edge processing of the entire foreground is rougher and there are residual signs of the original background. Compared with the former, Matting algorithm has better processing results, and its foreground and background are more integrated. However, looking at the entire image from another perspective, the Matting algorithm's processed images exhibit overfitting during the generation process, which can cause partial distortion of the characters. Therefore, it is necessary to further adjust parameters such as thresholds during the image generation process to reduce some distortion.

While using the same Paddlehub algorithm, we found that the compatibility between the extracted image and the new background still needs to be improved.

3. PRACTICAL APPLICATION IN DAILY LIFE

In short, the most practical application of cutout technology is to switch the background of people's photos, which can also be well achieved for making ID photos at home.

For the Matting algorithm, I think this practical application is simply a process of replacing the background with a solid color background. People can take photos of themselves through their mobile phones and use this algorithm for background replacement. I have tried some ID photo software on the market, and most of them generally reflect the situation that the hair details are not processed properly, resulting in the original background being cut in when cropping, or the cropped image only shows the hair closely attached to the scalp, which looks very unnatural. A large amount of post production makes the acquisition of ID photos more complicated.

In the process of trying the Matting algorithm for ID photos, I replaced the image background with a red screen and cropped it to generate a natural and simple electronic ID photo. As shown in Figure 6.

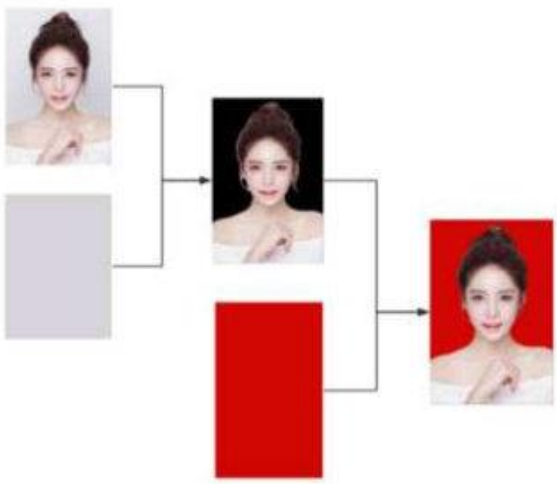

Figure 6: Flow Chart for Generating Application Effects

4. CONCLUSION

In general, the entire Matting algorithm is divided into two parts: one is the deep network used to predict Matting, which is obtained through loss training; And another Matting network is self supervised to determine image quality. In such a network, continuous loss training is performed on images to achieve a more perfect presentation and eliminate the need for green screens in image cutout techniques. The central point of the entire Matting algorithm is that it did not use the common ternary graph Trimap, but instead replaced it with an untitled background graph. Although it increased the basic workload, this approach greatly reduced the running time and space consumption, giving the entire algorithm the most obvious advantage. Meanwhile, although this article did not mention the production of videos, the algorithm is still applicable to videos and the presentation effect is not inferior to the production of images.

On this basis, based on the Matting algorithm, the problem of image distortion caused by the fusion of foreground and background is still a point that needs to be overcome to improve the entire algorithm.

REFERENCES

[1] Jiang, L., Yu, C., Wu, Z., & Wang, Y. (2024). Advanced AI framework for enhanced detection and assessment of abdominal trauma: Integrating 3D segmentation with 2D CNN and RNN models. arXiv preprint arXiv:2407.16165.
[2] Yan, H., Wang, Z., Xu, Z., Wang, Z., Wu, Z., & Lyu, R. (2024). Research on image super-resolution reconstruction mechanism based on convolutional neural network. arXiv preprint arXiv:2407.13211.
[3] Ji, H., Xu, X., Su, G., Wang, J., & Wang, Y. (2024). Utilizing Machine Learning for Precise Audience Targeting in Data Science and Targeted Advertising. Academic Journal of Science and Technology, 9(2), 215-220.
[4] Ma, Y., Shen, Z., & Shen, J. (2024). Cloud Computing and Hyperscale Data Centers: A Comparative Study of Usage Patterns. Journal of Theory and Practice of Engineering Science, 4(06), 11-19.
[5] Ren, Z. (2024). VGCN: An Enhanced Graph Convolutional Network Model for Text Classification. Journal of Industrial Engineering and Applied Science, 2(4), 110-115.
[6] Ren, Z. (2024). Enhanced YOLOv8 Infrared Image Object Detection Method with SPD Module. Journal of Theory and Practice in Engineering and Technology, 1(2), 1–7. Retrieved from https://woodyinternational.com/index.php/jtpet/article/view/42
[7] Yuan, B., & Song, T. (2023, November). Structural Resilience and Connectivity of the IPv6 Internet: An AS-level Topology Examination. In Proceedings of the 4th International Conference on Artificial Intelligence and Computer Engineering (pp. 853-856).
[8] Yuan, B., Song, T., & Yao, J. (2024, January). Identification of important nodes in the information propagation network based on the artificial intelligence method. In 2024 4th International Conference on Consumer Electronics and Computer Engineering (ICCECE) (pp. 11-14). IEEE.
[9] Wang, Z. (2024, August). CausalBench: A Comprehensive Benchmark for Evaluating Causal Reasoning Capabilities of Large Language Models. In Proceedings of the 10th SIGHAN Workshop on Chinese Language Processing (SIGHAN-10) (pp. 143-151).
[10] Lyu, H., Wang, Z., & Babakhani, A. (2020). A UHF/UWB hybrid RFID tag with a 51-m energy-harvesting sensitivity for remote vital-sign monitoring. IEEE transactions on microwave theory and techniques, 68(11), 4886-4895.
[11] Wu, Z., Wang, X., Huang, S., Yang, H., & Ma, D. (2024). Research on Prediction Recommendation System Based on Improved Markov Model. Advances in Computer, Signals and Systems, 8(5), 87-97.
[12] Wu, Z. (2024). MPGAAN: Effective and Efficient Heterogeneous Information Network Classification. Journal of Computer Science and Technology Studies, 6(4), 08-16.
[13] Xu, S., Wu, P., & Chen, Y. (2024). Interfacial thermal conductance in 2D WS2/MoSe2 and MoS2/MoSe2 lateral heterostructures. Computational Materials Science, 245, 113282.
[14] Wu, P., Iquebal, A. S., & Ankit, K. (2023). Emulating microstructural evolution during spinodal decomposition using a tensor decomposed convolutional and recurrent neural network. Computational Materials Science, 224, 112187.
[15] Wu, Z. (2024). MPGAAN: Effective and Efficient Heterogeneous Information Network Classification. Journal of Computer Science and Technology Studies, 6(4), 08-16.