Reference-based anime line art colorization by region correspondence using region features

Daisuke Nanya
Graduate School of Science and Technology
Meijo University
Nagoya, Japan
233426014 <at> ccmailg.meijo-u.ac.jp

Kouki Yonezawa
Faculty of Information Engineering
Meijo University
Nagoya, Japan
koukiy <at> meijo-u.ac.jp

Abstract—**The coloring process in anime (Japanese animation) production is a critical yet labor- and time-intensive phase in creating animated content. Reducing time and effort required for this step could significantly lower the time, cost, and effort required for anime production. Various body parts of anime characters are assigned corresponding colors in advance, and these colors often differ from character to character. Therefore, the characters in the line art need to be colored with these predetermined colors. However, as line art lacks information about the appropriate coloring, it becomes difficult to automatically color based solely on the line drawing. This study introduces a colorization system for anime line art based on region estimation using a reference colored image to address this issue. Through interviews with a Japanese animation studio, we identified that anime images are typically segmented into numerous distinct regions. The proposed system inputs a line art and reference colored image of the same cut as the line art and outputs the colored line art based on the features of corresponding regions in the two images. Experimental results using a dataset of real anime works confirm the effectiveness of our system.**

Keywords—anime, colorization, neural network

I. Introduction

The coloring phase in anime production is essential for creating animated content but is labor-intensive and time-consuming. Interviews with Studio Trigger [1], a Japanese animation studio, revealed four key insights into this process. First, the anime industry has been facing a decrease in the working population of the industry, an increase in the number of works being produced, and higher demands on the quality of work. In the future, animation studios will need to produce high-quality work with fewer staff members; therefore, efficiency is required. Second, the colors that exist in color anime images, including not only character parts such as eyes, hair, and skin, but also lines and props, have predetermined corresponding colors for all scenes, including dark and bright scenes. During the coloring process, the coloring staff color the line art with designated colors. Third, if there is even a slight change in a character's movement between the two frames, the coloring staff will color all changed regions. Generally, scenes with slight movement or line art with few regions tend to take less time in the coloring process, whereas scenes with excessive movement, such as battle scenes or line art with many regions, tend to take longer. Coloring staff say that coloring a single line of art can take up to 30 min. Finally, the anime industry has not yet introduced automated coloring systems; therefore, the coloring staff require manual coloring for line art. Therefore, an automated coloring system would be extremely valuable and would significantly benefit the coloring staff.

Additional findings from the interviews indicate that anime line art differs from live-action line images in that it features two types of lines: those present after coloring and those absent. Both types of lines often indicate the boundaries of a region, and the coloring staff use these lines as a guide to fill in the region and color the line art. Furthermore, anime images and line art are segmented into numerous regions, often exceeding 100, with each region typically assigned a single color, except in specific cases.

This paper introduces a novel colorization system for anime line art that uses a referenced colored image, to minimize the time and effort involved in the coloring process.

The remainder of this paper is organized as follows. Section 2 outlines the related work and Section 3 presents our approach. Section 4 outlines the experimental procedure and then the study is concluded in Section 5.

II. Related Works

A. GAN-based coloring methods

GAN-based coloring methods are based on generative adversarial networks (GANs) [2], a machine-learning technique that directly generates an output image using neural networks. These methods can color line art, even though the line art and reference images are not similar. However, they require large amounts of training data. Pix2Pix [3] is an image-to-image conversion method that converts line art into colored images. However, it remains uncertain whether the line art is colored according to the users' preferences, as the system currently lacks the ability to incorporate hint information. Tag2Pix [4] can automatically color the line art of illustration-style images using the tag information of the body part and its color as a hint. Users do not have to provide a reference image for colorization but cannot specify the pixel values of the color. A neural network-based coloring method [5] can automatically color the line art of illustration-style images using colors in a reference colored image as a hint. However, this method is unsuitable for anime because a single region may contain multiple colors.

B. Region-based coloring methods

Region-based coloring methods that focus on manga (Japanese comics) and anime images are segmented into numerous regions and color regions by estimating the regions in the line art corresponding to the regions in the reference image. These methods are suitable for coloring manga and anime because their line art is colored only with the colors contained in the reference image. A graph-correspondence-based coloring method [6] can automatically color manga using reference color images. Sato et al. developed a method for color regions contained in line art and reference images by graph correspondence using quadratic programming. However, the processing time can be extensive for anime images segmented into numerous regions. The component-based coloring method [7] can automatically color manga using a reference color image. Do et al. attempted to increase the accuracy of region estimation and reduce the processing time by adding region shape features to the graph correspondence-based coloring method [6]. An animation transformer [8] can automatically color animated images using reference-colored images. Casey et al. aimed to further improve the accuracy of region estimation using the Transformer model for region estimation. However, when a high-performance computer is used, training the model requires long processing time.

III. Our Approach

Fig. 1 presents an overview of the proposed method. The system inputs a reference colored image and grayscale line art as the target image and outputs the colored target image. This approach comprises three steps. First, the region information from the reference and target images was extracted. The regions were then estimated using the extracted information. Finally, a colored anime image obtained by coloring regions in the target image was completed and extraneous lines were removed. The proposed method does not directly generate colored images but produces images of many regions, each assigned a single color. The region estimation method in the second step is inspired by a component-based coloring method [7].

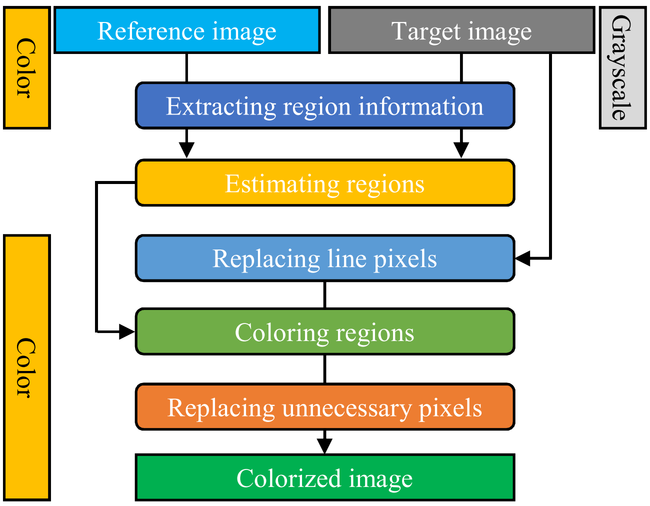


Fig. 1. Overview of our method. The system requires a color reference image and grayscale line art as the target image as inputs, and outputs a colorized image.

A. Extracting region information

Initially, the system gathers information on all the regions in both images, including the centers of gravity, areas, and shape features. The centers of gravity and areas were represented as numerical values ranging from 0 to 1, whereas shape features were encoded in 2048-dimensional vectors derived from a pre-trained ResNet50 [9].

B. Estimating regions

Second, the system uses three indices—center of gravity, area, and shape features—to match the corresponding regions in both images based on a cost function. The cost function assesses the similarity between the i-th region in the target image and the j-th region in the reference image and outputs smaller values that are more similar to the characteristics of the two regions. System corresponds j-th region, which minimizes the cost function for the i-th region. The cost function F_1 estimates the corresponding regions in large regions whose areas are over 0.1% of the entire target image. However, when the output value of the cost function F_1 exceeds the threshold, the corresponding region is not determined. We define the cost function F_1 as follows:

$$F_1(i,j) = D_{shape} \cdot D_{area} \cdot D_{length}, \quad (1)$$

$$D_{shape} = \exp(\alpha_1 \cdot \left\| f_{t_i} - f_{r_j} \right\|_1), \quad (2)$$

$$D_{area} = \exp(\beta_1 \cdot \left| a_{t_i} - a_{r_j} \right|), \quad (3)$$

$$D_{length} = \exp(\gamma_1 \cdot d_{t_i, r_j}), \quad (4)$$

$$d_{A,B} = |x_A - x_B| + |y_A - y_B|. \quad (5)$$

Here, D_{shape} represents the sum of the absolute differences between the shape feature vectors of the i-th and j-th regions. Lower values of D_{shape} indicate more similar shapes. Furthermore, D_{area} represents the absolute difference in the areas of the two regions, with lower values indicating more similar areas. D_{length} represents the distance between of the two regions, with lower values indicating more nearly. $d_{A,B}$ represents the distance between regions A and B, with lower values indicating more closeness. Using the Manhattan distance reduces the cost of calculation than the Euclidean distance. Parameters α, β, and γ measure the amount of importance given to each index in the cost function.

After estimating large regions with the cost function F_1, the cost function F_2 estimates all regions that are not determined by the cost function F_1. We define the cost function F_2 as follows:

$$F_2(i,j) = D_{shape} \cdot D_{area} \cdot D_{point}, \quad (6)$$

$$D_{point} = \exp(\gamma_2 \cdot \sum_{m=0}^{M} \left| d_{t_i, t_m} - d_{r_j, r_m} \right|). \quad (7)$$

Here, D_{point} denotes the total difference between the Manhattan distance from the i-th region to the region associated with the cost function F_1 and the Manhattan distance from the j-th region to the region associated with F_1. The parameters of D_{shape} and D_{area} in the cost function F_2 are α_2 and β_2, respectively. We set the parameters in the cost functions as follows: α_1 and β_1 are 1, α_2 is 0.5, β_2 is 3, γ_1 is 0.0005, and γ_2 is 0.0001. We set the threshold with the cost function F_1 to 5.

C. Coloring regions

In the final step, the system produces a completely colored image while performing three processes: replacing the line pixels of the target image, coloring the regions, and replacing unnecessary pixels.

In the first process, the system replaces the pixel values of the lines that remain, even in colored anime images, with a predetermined value. The color of the character's body parts and the color of the line were predetermined. In the second process, the system fills each region in the target image with a color corresponding to the estimated region in the reference image. After coloring, the system eliminates unnecessary lines that do not appear in the animated images. An unnecessary pixel value is replaced with the most frequently located pixel value among the eight surrounding pixels. If an unnecessary pixel is located at the edge of the image, its value is replaced by the most frequently located pixel value among the three or five surrounding pixels.

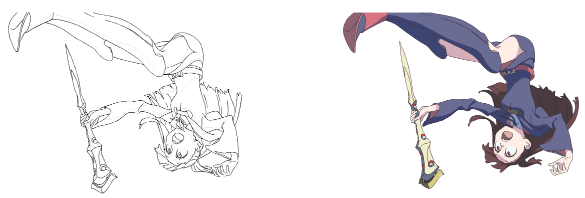
IV. Experiment

We evaluated our method using a dataset [10] comprising real animated works [11] and generated 92 images from 46 cuts. These images have movement throughout, and the line art is the previous and next frames of the reference images in the same cut. We compared the accuracy rate of our proposed method with that of existing methods, the component-based coloring method [7], and the animation transformer [8]. The software proposed in reference [12] was used as the animation transformer. The proposed and component-based coloring methods colorized all 92 images; however, the animation transformer was unable to colorize five images. Table 1 shows the average region-based accuracy rates of our method and existing methods for 87 images, excluding five images that could not be generated by the animation transformer. Our method outperformed these algorithms with $P < 0.03$ using the paired t test.

TABLE I. AVERAGE ACCURACY RATE OF OUR METHOD AND EXISTING METHODS

Method	Average accuracy rate
Our method	77.13%
Component-based coloring method	69.05%
The animation transformer	76.32%

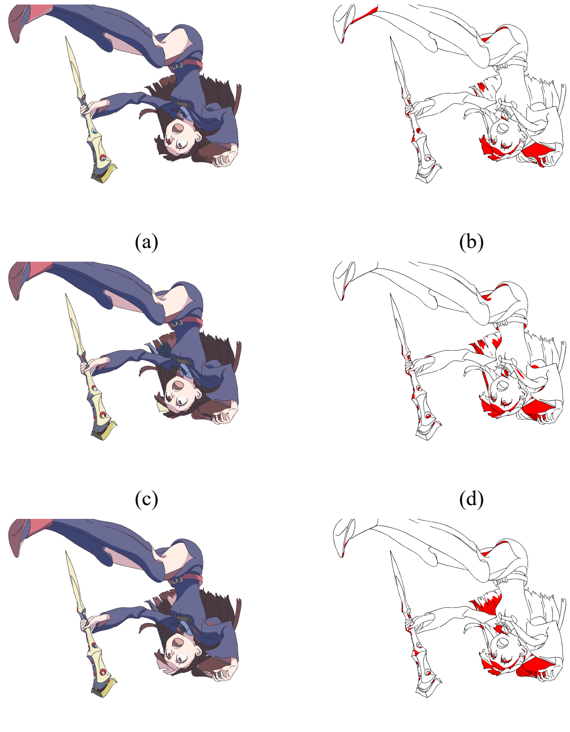
Fig. 2 shows an example of an input image. Unnecessary lines in the colored image are shown as gray lines in the line art. The line art has 343 regions, and the reference image has 294 regions. These images are among the most complex cuts in anime images because they contain over 200 regions.



(a) (b)

Fig. 2. (a) input line art and (b) reference image. The image in (b) is the next frame of the line art.

Fig. 3 shows an example of the colorization results and the difference images between the colorization results and their ground truth. The images in Figs. 3(a), (c), and (e) are outputted by inputting Fig. 2 into each system. The accuracy rates are 67.64% (232 343 regions) for our method, 54.52% (187 out of 343 regions) for the component-based coloring method, and 62.68% (215 343 regions) for the animation transformer. In these difference images, larger regions can be accurately estimated, and estimation errors are common in smaller regions.



(a) (b) (c) (d) (e) (f)

Fig. 3. Examples of colorization results and difference images between the colorization results and their ground truths. (a) Colorization results of the proposed method. (b) Difference between (a) and Fig. 2(c). (c) Colorization results of component-based coloring method. (d) Difference image between (c) and Fig. 2(c). (e) Colorization results of animation transformer. (f) Difference between (e) and Fig. 2(c).

Fig. 4 shows an example of colorization that results in the appearance of a monster. The line art has 65 regions, and the reference image has 71 regions. The accuracy rates were 92.31% (60 of 65 regions) for our method, 90.77% (59 of 65 regions) for the component-based coloring method, and 89.23% (58 of 65 regions) for the animation transformer. In scenes that contain non-human characters, such as monsters, the area of the regions is larger, and the number of regions tends to be smaller than in scenes that contain humans; therefore, the centers of gravity of the regions are not dense. Therefore, the system is less likely to cause estimation errors, and the accuracy rate of the region estimation will be higher. The results demonstrated that large regions are accurately colored, whereas smaller regions are occasionally assigned incorrect colors. Figs. 5 and 6 present examples of consecutive colorizations and a comparison of the colorization results, respectively.

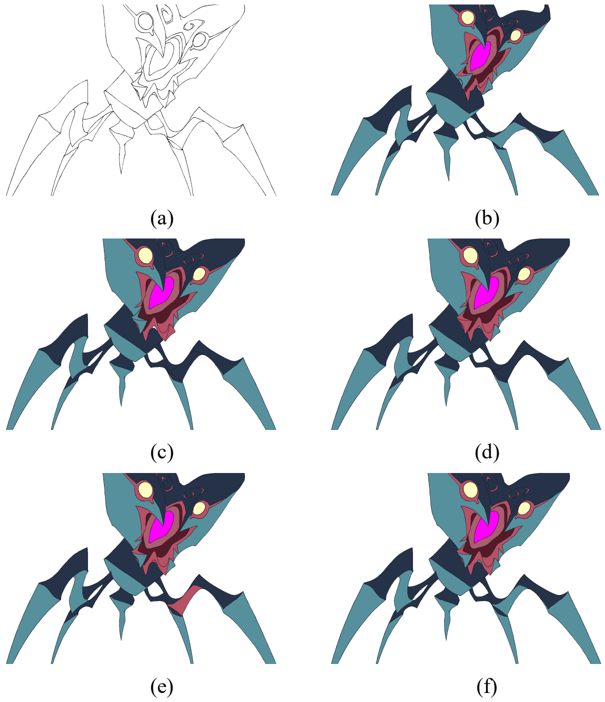


Fig. 4. An example of colorization results in the appearance of a monster and the ground truth. (a) Line art. (b) Reference images. This image is the next frame of line art. (c) Colorization results of the proposed method. (d) Colorization result of component-based coloring method. (e) Colorization results of animation transformer. (f) Ground truth.

Fig. 5. Example of consecutive colorization results. Line art 1 ($t=1$) was colorized using the reference image ($t=0$), and line art 2 ($t=2$) was colorized using the colorized line art ($t=2$).

V. Conclusion

This study introduces a novel method for color anime line art using reference images. Users do not need to prepare large amounts of data obtained from copyright owners because our method does not use machine-learning techniques. When the line art and reference images have similar compositions, such as when both images are consecutive frames, our method can accurately color numerous regions. The incorrect painting results are similar to those obtained for smaller areas. Aggregating these areas improves the performance of the proposed method. By introducing the proposed system into the coloring process, the coloring staff can entrust the coloring of large regions to the system and only color regions with wrong colors.

Acknowledgment

The dataset used in this study was provided by the National Institute of Informatics and the Studio Trigger.

References

[1] Trigger, Inc., "TRIGGER official website." [Online]. Available: https://www.st-trigger.co.jp/

[2] I. Goodfellow *et al.*, "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, pp. 139–144, Oct. 2020, doi: 10.1145/3422622.

[3] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI: IEEE, Jul. 2017, pp. 5967–5976. doi: 10.1109/CVPR.2017.632.

[4] H. Kim, H. Y. Jhoo, E. Park, and S. Yoo, "Tag2Pix: Line Art Colorization Using Text Tag With SECat and Changing Loss," presented at the Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 9056–9065. [Online]. Available: https://openaccess.thecvf.com/content_ICCV_2019/html/Kim_Tag2Pix_Line_Art_Colorization_Using_Text_Tag_With_SECat_and_ICCV_2019_paper.html

[5] Y. Cao, H. Tian, and P. Y. Mok, "Attention-Aware Anime Line Drawing Colorization," in *2023 IEEE International Conference on Multimedia and Expo (ICME)*, Brisbane, Australia: IEEE, Jul. 2023, pp. 1637–1642. doi: 10.1109/ICME55011.2023.00282.

[6] K. Sato, Y. Matsui, T. Yamasaki, and K. Aizawa, "Reference-Based Manga Colorization by Graph Correspondence Using Quadratic Programming," in *SIGGRAPH Asia 2014 Technical Briefs*, in SA '14. Association for Computing Machinery, Nov. 2014, pp. 1–4. doi: 10.1145/2669024.2669037.

[7] T. Do *et al.*, "Anime Sketch Colorization by Component-Based Matching using Deep Appearance Features and Graph Representation," in *2020 25th International Conference on Pattern Recognition (ICPR)*, Milan, Italy: IEEE, Jan. 2021, pp. 3154–3161. doi: 10.1109/ICPR48806.2021.9412507.

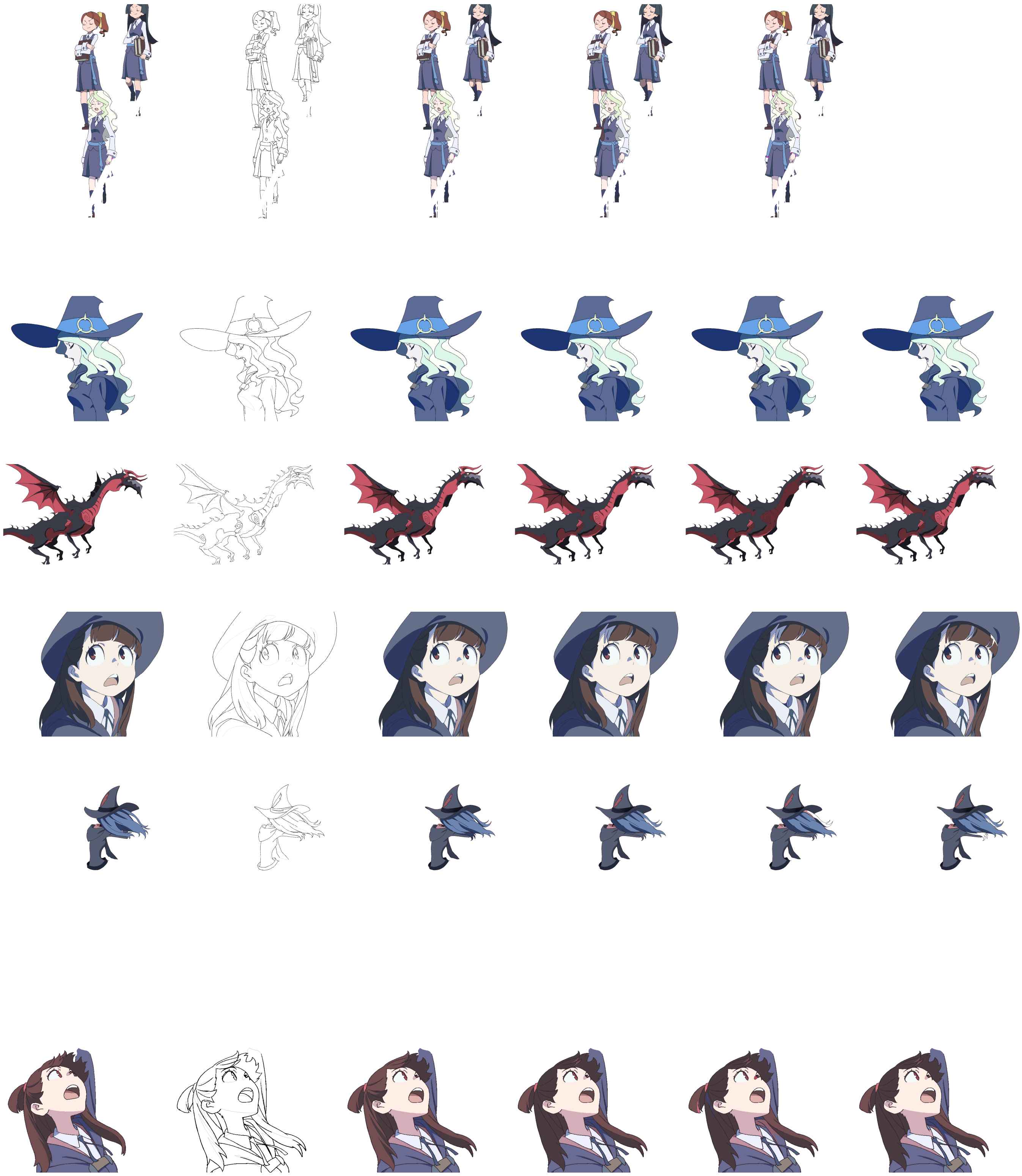
[8] E. Casey, V. Perez, and Z. Li, "The Animation Transformer: Visual Correspondence via Segment Matching," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, QC, Canada: IEEE, Oct. 2021, pp. 11303–11312. doi: 10.1109/ICCV48922.2021.01113.

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA: IEEE, Jun. 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.

[10] National Institute of Informatics, "Trigger dataset," IDR Informatics Research Data Repository. [Online]. Available: https://www.nii.ac.jp/dsc/idr/trigger/

[11] TRIGGER, Y. Yoshinari, and GOOD SMILE COMPANY, "'Little Witch Academia: The Enchanted Parade' official website." [Online]. Available: http://littlewitchacademia.jp/

[12] Cadmium, *Cadmium App*. (2021). [Online]. Available: https://cadmium.app/



(a) (b) (c) (d) (e) (f)

Fig. 6. Comparison of the colorization results. These lines of art are the previous or next frame of the reference image. Blank spaces are images that cannot be colored. The animation transformer seems unable to color line art that contains a large number of pixels or many regions with over 350 pixels. (a) Reference image. (b) Line art. (c) Ground truth. (d) Colorization result of our method. (e) Colorization results of component-based coloring method. (f) Colorization results of animation transformer.