Fully Automatic Colorization for Anime Character Considering Accurate Eye Colors

Kenta Akita
Kyushu University
2DS19064W@s.kyushu-u.ac.jp

Yuki Morimoto
Kyushu University
morimoto@design.kyushu-u.ac.jp

Reiji Tsuruno
Kyushu University
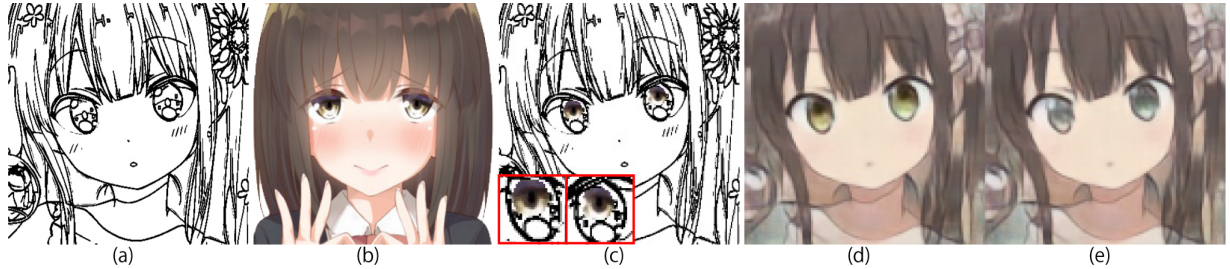tsuruno@design.kyushu-u.ac.jp

Figure 1: Highlights of our system: (a) input line drawing, (b) color reference image, (c) line drawing with auto-computed hints, (d) our colorization result, (e) result by [Furusawa et al. 2017]. All Images are from Nico-illust Dataset[Ikuta et al. 2016].

CCS CONCEPTS

• **Computing methodologies → Image processing**;

KEYWORDS

Colorization, Deep Learning, Line arts

ACM Reference Format:

1 INTRODUCTION

In this paper, we propose a method to colorize line drawings of anime characters' faces with colors from a reference image. Previous studies using reference images often fail to realize fully-automatic colorization, especially for small areas, e.g., eye colors in the resulting image may differ from the reference image. The proposed method accurately colorizes eyes in the input line drawing using automatically computed hints. The hints are round patches used to specify the positions and corresponding colors extracted from the eye areas of a reference image.

Recently, automatic colorization of line drawings using deep learning has been studied [Zhang et al. 2018][Ci et al. 2018]. Most of these studies require user postprocessing to correct the automatic colorization results. In such cases, color strokes and/or dots are given as hints on the result; therefore, multiple line drawings cannot be colorized simultaneously.

Comicolorization [Furusawa et al. 2017] is a semi-automatic colorization system for manga images. This method applies a color reference image to specify the color for each character image. However, with this method, it is difficult to reflect accurate colors for small regions, such as eyes and complicated hair areas. We propose a fully-automatic colorization method to accurately colorize eyes relative to a reference image.

2 METHOD

In the proposed method, based on the network of Comicolorization [Furusawa et al. 2017], hints are computed as color patches automatically using a network to estimate the positions of the eyes (Fig. 1). Here, the target line drawing to colorize is a facial raster image of an anime character. Figure 2 shows an overview of the proposed method.

2.1 Creating Dataset

The proposed method includes a colorization network and an eye position estimation network.

To train the colorization network, we created a color image dataset (CDS) and a paired raster line drawing dataset (LDS) from actual anime videos. First, 20000 facial square images greater than 256 x 256 pixels were detected [Nagadomi 2011] and cropped from the videos. These images were resized to 256 x 256. Then, the images were cropped to 224 x 224 pixels at random positions and flipped horizontally at 50% probability to increase estimation robustness.

The LDS was generated from CDS. First, CDS images were converted to grayscale. Next, morphological processing was applied to thin black lines by dilating white regions. Then, the subtracted image of the dilated image and the grayscale image resulted in white contours. Then, the image is inverted and binarized to obtain the line drawing and the LDS is created.

Paired datasets of line drawings (LDS^{eye}) and eye positions (PDS) were created for the eye position estimation network. The LDS^{eye} contains 2500 line drawings, including some of the LDS. The PDS
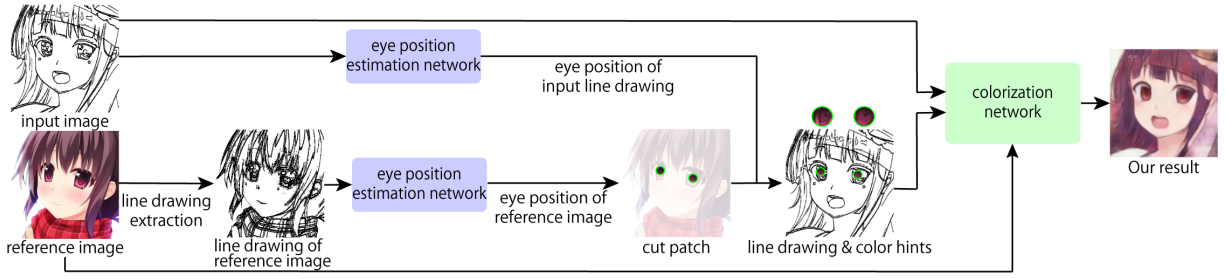
Figure 2: Overview of our method. All Images are from Nico-illust Dataset[Ikuta et al. 2016].

was created as the ground truth by manually specifying the positions of the right and left eyes in each LDS^{eye} image.

2.2 Colorization Network

The colorization network is based on one of the Comicalorization networks; however, we only apply this network's histogram model. The input and output of the network are the LDS and CDS for training the colorization network. While training the colorization model, the loss function and optimizer of our network are same as those in the Comicolorization network. Here the model is trained over 273000 iterations with a batch size of 30.

The colorization network takes a line drawing and color reference image as input data. In addition, round hint patches are given at the eyes in the input line drawing to generate intermediate input data such as the Fig. 1c. The eye positions in the input line drawing and reference image are estimated as described in Section 2.3. The hint patches are cut out from eyes in the reference image at the computed eye positions. Then, these patches are pasted at the eye positions in the line drawing (Fig. 1c). The colorization network refers to this image to reflect eyes colors accurately.

2.3 Eye Position Estimation Network

The input and output used to train the eye position estimation network are the LDS^{eye} and PDS, respectively. With the exception of the final layer, the network is based on ResNet-18 [He et al. 2016]. In addition, we change the filter size of the convolutional layers from 3 × 3 to 4 × 4. The outputs of the network are the x and y coordinates of the right and left eyes; thus, in our model, the final output layer is a fully-connected layer with four channels. The network is trained without using the activation function of the final output layer and outputs the eye position values.

The loss function of our network is mean squared error, and the optimizer is the same as the colorization network. Here, the model is trained over 55000 iterations with a batch size of 80.

The eye positions are estimated for both the line drawing and color reference images. The input to this network is a facial line drawing of a character. Before eye position estimation of a reference color image, the color image is converted to a line drawing using the same process used to create the LDS.

3 RESULTS AND CONCLUSION

We compare results obtained using the proposed method and an existing method [Furusawa et al. 2017] without user postprocessing

(Figs. 1 and 3). As can be seen, the eye colors in the proposed method's results are more accurately reflect the eye colors in the reference image. The proposed method enables us to accurately colorize eyes even if there is given a detailed specification, such as different colors of eyes (Fig. 3, bottom). In addition, the proposed method can colorize many line drawings simultaneously without requiring user postprocessing.

However, there are some limitations relative to the hint patches. For example, non-existent content, such as highlights in the input line drawing, can occur in the colorized results because the hint patch includes the content in the reference image. In contrast, if the patch does not include the entire eye gradation, the proposed method cannot reflect an accurate gradation. In such cases, the user can edit the image using hints, such as color dots; however, in future, we expect to improve these limitations.
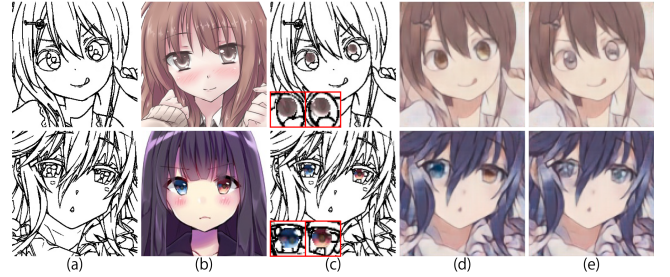

Figure 3: Result comparison: input line drawing (a), reference image (b), input image with hint patches (c), proposed method's result (d), existing method's result [Furusawa et al. 2017] without post processing (e). All Images are from Nico-illust Dataset[Ikuta et al. 2016].

REFERENCES

Yuanzheng Ci, Xinzhu Ma, Zhihui Wang, Haojie Li, and Zhongxuan Luo. 2018. User-Guided Deep Anime Line Art Colorization with Conditional Adversarial Networks *(MM '18)*. 1536–1544.

Chie Furusawa, Kazuyuki Hiroshiba, Keisuke Ogaki, and Yuri Odagiri. 2017. Comicolorization: Semi-automatic Manga Colorization. *SIGGRAPH Asia 2017 Technical Briefs* (2017).

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep Residual Learning for Image Recognition. *CVPR* (2016).

Hikaru Ikuta, Keisuke Ogaki, and Yuri Odagiri. 2016. Blending Texture Features from Multiple Reference Images for Style Transfer. In *SIGGRAPH Asia Technical Briefs*.

Nagadomi. 2011. lbpcascade_animeface. https://github.com/nagadomi/lbpcascade_animeface

Lvmin Zhang, Y I Ji, Chengze Li, Tien-Tsin Wong, Yi Ji, and Chunping Liu. 2018. Two-stage Sketch Colorization. *ACM Trans. Graph* 37, 6 (2018).