Confidence Estimation of Few-shot Patch-based Learning for Anime-style Colorization

Yuexiang Ji
The University of Tokyo
Tokyo, Japan
ji-yuexiang080@g.ecc.u-tokyo.ac.jp

Akinobu Maejima
OLM Digital, Inc., IMAGICA GROUP Inc.
Tokyo, Japan
akinobu.maejima@olm.co.jp

Yotam Sechayk
The University of Tokyo
Tokyo, Japan
sechayk-yotam@g.ecc.u-tokyo.ac.jp

Yuki Koyama
The University of Tokyo
Tokyo, Japan
koyama@pe.t.u-tokyo.ac.jp

Takeo Igarashi
The University of Tokyo
Tokyo, Japan
takeo@acm.org

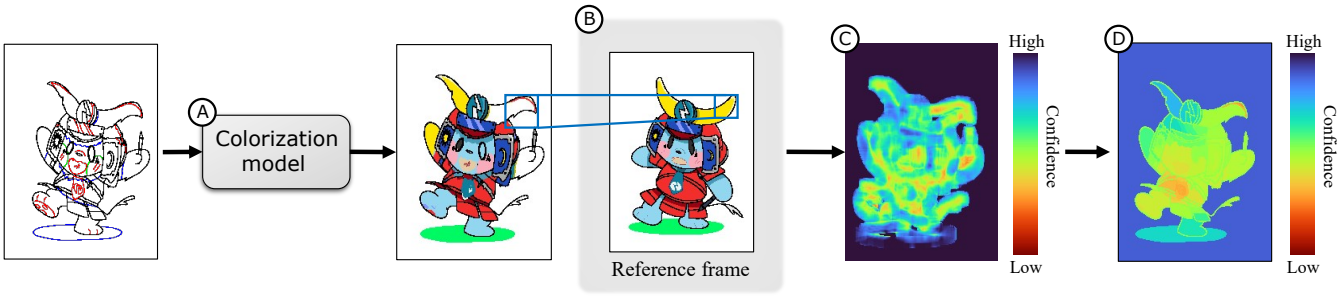

Figure 1: Overview of the proposed pipeline. First we get the output (A) of the colorization model, then we (B) perform patch matching against a reference frame, then we (C) detect confidence per pixel, and finally (D) calculate region-based confidence scores. Images originate from *Deadline*, ©OLM Asia SDN BHD.

Abstract

In hand-drawn anime production, automatic colorization is used to boost productivity, where line drawings are automatically colored based on reference frames. However, the results sometimes include wrong color estimations, requiring artists to carefully inspect each region and correct colors—a time-consuming and labor-intensive task. To support this process, we propose a confidence estimation method that indicates the confidence level of colorization for each region of the image. Our method compares local patches in the colorized result and the reference frame.

Keywords

Automatic colorization, Line drawing, Confidence estimation

ACM Reference Format:
Yuexiang Ji, Akinobu Maejima, Yotam Sechayk, Yuki Koyama, and Takeo Igarashi. 2025. Confidence Estimation of Few-shot Patch-based Learning for Anime-style Colorization. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Posters (SIGGRAPH Posters '25), August 10–14, 2025, Vancouver, BC, Canada.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3721250.3742964

SIGGRAPH Posters '25, Vancouver, BC, Canada

ACM ISBN 979-8-4007-1549-5/25/08
https://doi.org/10.1145/3721250.3742964

1 Introduction

Automatic anime-style line drawing colorization has gained attention as a time-saving tool in digital illustration workflows [Cao et al. 2023b; Liu et al. 2025; Maejima et al. 2024]. While recent models improve efficiency and accuracy using reference images [Liu et al. 2025; Maejima et al. 2024], they still produce incorrect color assignments—for example, changing a character's eye color across frames. Animators must then manually scan the colorized result side by side with the references, such as other frames or character sheets, and correct errors by hand. This tedious process can outweigh the time saved by automation, sometimes exceeding the time required for manual coloring [Cao et al. 2023a].

To assist animators and address this challenge, we investigate tailored approaches to estimate the confidence of colorized result, and visualization methods to convey colorization confidence. Using our pipeline, artists can easily identify regions in the colorized result that may require manual corrections, thereby saving time and effort on scanning the colorized result for possible errors.

2 Method

We developed a pipeline to estimate the color-confidence of colorization models. Our pipeline consists of two stages: pixel-wise

confidence computation, and region-wise aggregation. For each pixel in the colorized result, we extract a patch centered at the pixel and search for the most similar patch of the same size within the reference frame [Maejima et al. 2024]. We define a reference frame as an adjacent frame previously marked as color-accurate. The similarity score between the two patches directly determines the pixel-wise confidence at the pixel in the colorized result. Similarity between patches centered at $p = (p_1, p_2)$ in the colorized result and $q = (q_1, q_2)$ in the reference image is defined as:

$$\mathrm{R}(p,q) = \frac{\sum_{x,y}(P'_{\mathrm{out}}(p_1+x, p_2+y) \cdot P'_{\mathrm{ref}}(q_1+x, q_2+y))}{\sqrt{\sum_{x,y}(P'_{\mathrm{out}}(p_1+x, p_2+y))^2 \cdot \sum_{x,y}(P'_{\mathrm{ref}}(q_1+x, q_2+y))^2}}$$

Here, P'_{out} and P'_{ref} denote zero-mean pixel values (normalized from the 0–255 range) in the colorized result and reference image, respectively. The summation ranges over integer values $x \in [-\frac{w}{2}, \frac{w}{2}]$ and $y \in [-\frac{h}{2}, \frac{h}{2}]$, where w and h are the width and height of the patch.

Then, we compute **region-wise confidence**. We segment the colorized result into **regions** (an area enclosed by line segments) using a flood-fill algorithm to yield coherent regions. The confidence of each region is the average confidence over each of its pixels.

3 Result

Our method successfully convey confidence of colorized result. For example, in Figure 2, we effectively highlight incorrect color assignments with low confidence score, such as the bottom of the character's feet which correctly receiving low confidence score 0.33. However, we still experience misleading confidence scores, such as with the right eye of the character, where the confidence is still relatively high (such as 0.62), although the color is incorrect based on the ground truth.

We further explore discretized visualization of confidence, where we clamp continuous values to fixed thresholds. We use the following thresholds which our investigation found useful: 0.0–0.50 → 0.50, 0.62–0.75 → 0.75, and 0.82–1.0 → 0.90. Reducing color variety improves clarity, making it easier for animators to identify problematic areas. Using both discretizations and region based confidence maps results in easy-to-interpret confidence estimation, as we demonstrate in Figure 2.

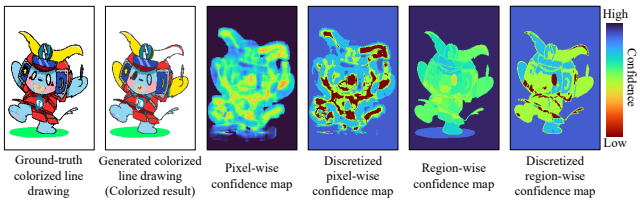

Figure 2: Discretized pixel-wise and region-wise confidence map, compared with original continuous confidence maps. Images originate from *Deadline*, ©OLM Asia SDN BHD.

Finally, we investigate how patch size influences accurate patch matching, a critical factor in our pipeline. Our results indicate that patch size directly affects the quality of the resulting confidence map. An 8×8 patch results in a miscolored region—the right ear, which should be yellow—being erroneously assigned high confidence. In contrast, a 16×16 patch produces a more accurate confidence map. Smaller patches primarily capture local edge information, which can lead to false correspondences. For instance, the model may incorrectly match the ear region to a distant hand based on similar boundary features (Figure 3). Larger patches, which capture more global context, are more likely to yield semantically meaningful correspondences, such as ear-to-ear, thereby enhancing the reliability of confidence estimates.

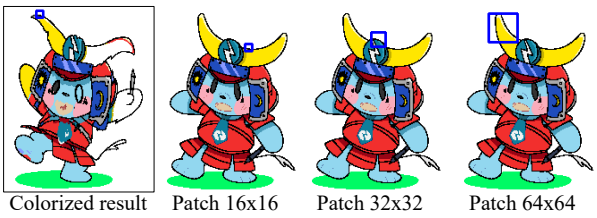

Figure 3: Patch matching results with different patch sizes. The blue marker indicates the target patch (colorized result), and the best-match patch candidate results. Images originate from *Deadline*, ©OLM Asia SDN BHD.

4 Limitations and Future Work

Sometimes, inaccurate matches occur due to distant, unrelated regions. We mitigate this by using larger patches and can further improve reliability by incorporating spatial information or dividing the image into smaller grids. Our pipeline is computationally expensive (7–8 minutes for a 400×500 image with 16×16 patches), but runtime can be reduced by skipping regions unlikely to yield useful matches.

5 Conclusion

We introduced a patch-based confidence estimation method for anime-style colorization that helps animators identify areas likely needing correction. By comparing the colorized result with a reference image, our method enables evaluation without ground-truth data, and also offers intuitive visualizations of low-confidence regions.

Acknowledgments

This work is funded by New Energy and Industrial Technology Development Organization (NEDO) JPNP20017.

References

Yu Cao, Xiangqiao Meng, PY Mok, Xueting Liu, Tong-Yee Lee, and Ping Li. 2023a. Animediffusion: Anime face line drawing colorization via diffusion models. *arXiv preprint arXiv:2303.11137* (2023).

Yu Cao, Hao Tian, and PY Mok. 2023b. Attention-aware anime line drawing colorization. In *2023 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 1637–1642.

Zhiheng Liu, Ka Leong Cheng, Xi Chen, Jie Xiao, Hao Ouyang, Kai Zhu, Yu Liu, Yujun Shen, Qifeng Chen, and Ping Luo. 2025. MangaNinja: Line Art Colorization with Precise Reference Following. *arXiv preprint arXiv:2501.08332* (2025).

Akinobu Maejima, Seitaro Shinagawa, Hiroyuki Kubo, Takuya Funatomi, Tatsuo Yotsukura, Satoshi Nakamura, and Yasuhiro Mukaigawa. 2024. Continual few-shot patch-based learning for anime-style colorization. *Computational Visual Media* 10, 4 (2024), 705–723.