Supplementary Materials for Closing the Domain Gap in Manga Colorization via Aligned Paired Dataset

1. Additional Visualizations

This section provides additional visualizations that complement our main results. Figure 1 illustrates the effectiveness of our panel segmentation approach, which accurately identifies and separates individual manga panels, including non-rectangular and overlapping ones.

Figure 2 provides additional visual comparison for models trained on synthetic and real data.

Next, we present a comparison of baseline models trained on ImageNet and WikiArt on Figure 3 The visualizations clearly demonstrate that models trained on non-manga datasets, such as ImageNet and WikiArt, are not suitable for manga colorization. These models, including InstructPix2Pix (IP2P) and ControlNet (CN), perform much better when trained on manga-specific data, highlighting the importance of domain-specific training for achieving optimal colorization results.

Finally, Figure 4 presents a comparison of the manga-colorization-v2 model with several popular colorization applications that do not allow for fine-tuning, such as Petalica [1], Komiko [2], RecolorAI [3], and Style2Paints (S2P) [5]. The results indicate that these applications perform poorly in terms of color accuracy and structural preservation, emphasizing the advantages of using models specifically trained on manga datasets.

2. Synthetic data generation

In this section, we describe the synthetic data generation method employed in our experiments, based on the approach from [4]. This method was particularly effective when paired datasets of black-and-white (BW) and color images were not available, and the only way to train supervised models was to collect color images and generate corresponding BW images through algorithmic means. A comparison of synthetic BW images and real BW images is presented in Figure 6.

The exact algorithm for generating synthetic BW images is as follows:

1. Gather color images and BW images that are not matched and originate from different manga titles.
2. Apply the extended Difference of Gaussians (xDoG) [6] algorithm to the BW images to extract edges, resulting in images consisting of lines.
3. Train an image-to-image translation model to reconstruct the original BW images from the line images generated in step 2.
4. Apply the xDoG algorithm to the color images to generate line images from the color data.
5. Use the trained model from step 3 to generate BW images from the line images obtained in step 4.

This approach produces synthetic BW images that more closely resemble real manga BW images compared to simply converting color images to grayscale or using edge extraction alone.

3. MOS demographics

Our user study involved 25 participants who rated from 1 to 5. Fifteen were familiar with manga, and 10 were not. Participants were not professional artists, with diverse ages among those unfamiliar with manga, while those familiar were younger, reflecting manga's popularity among youth.

References

[1] https://https://petalica.com/. Accessed: 2024-09-09. 1, 3

[2] https://komiko.app/. Accessed: 2024-09-09. 1, 3

[3] https://recolorai.com/. Accessed: 2024-09-09. 1, 3

[4] Maksim Golyadkin and Ilya Makarov. Robust manga page colorization via coloring latent space. *IEEE Access*, 11:111581–111597, 2023. 1, 2

[5] lllyasviel. Style2paints. https://github.com/lllyasviel/style2paints, 2017. Accessed: 2024-09-09. 1, 3

[6] Holger Winnemöller, Jan Eric Kyprianidis, and Sven C. Olsen. Xdog: An extended difference-of-gaussians compendium including advanced image stylization. *Computers & Graphics*, 36(6):740–753, 2012. 2011 Joint Symposium on Computational Aesthetics (CAe), Non-Photorealistic Animation and Rendering (NPAR), and Sketch-Based Interfaces and Modeling (SBIM). 1

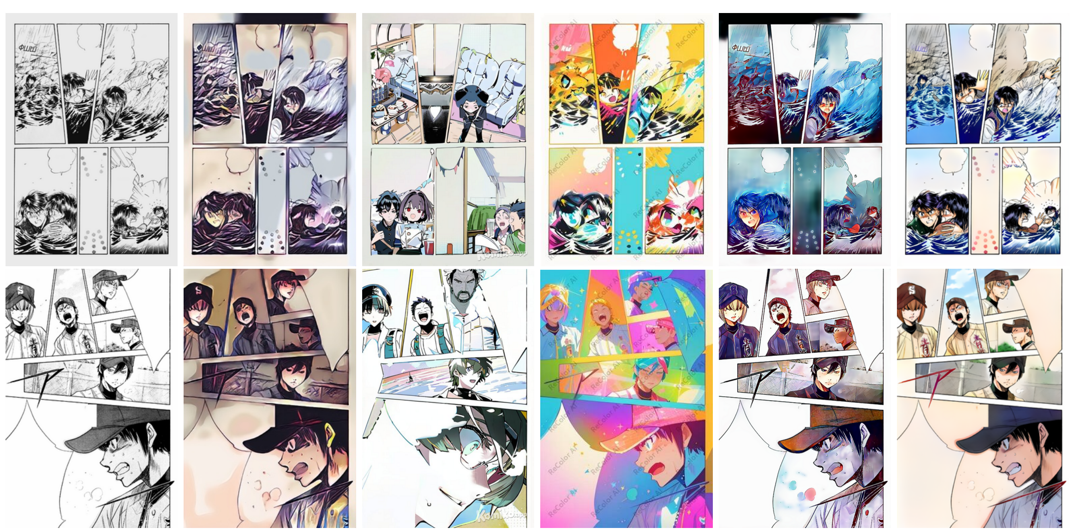
Figure 1. **Panel Segmentation results**. This figure illustrates the effectiveness of our approach in accurately identifying and separating individual panels. Such precise separation simplifies the colorization process by allowing each panel to be processed independently. The figure demonstrates that non-rectangular panels are segmented accurately, panels that intersect with others are correctly distinguished, and each panel is identified as an independent entity. This ensures that even complex and intersecting panels are handled correctly, maintaining the integrity of the artwork during colorization.

(a) Input (b) mc-v2$_{sync}$ (c) ControlNet$_{sync}$ (d) mc-v2 (e) ControlNet

Figure 2. **Comparison of models trained on synthetic data and real data**. We trained ControlNet and manga-colorization-v2 (here denoted as mc-v2) on our dataset and on synthetic data generated using the color data generation method from [4]. The figure shows that models trained on real data are better at recognizing objects and their shapes, resulting in more proper colorings.

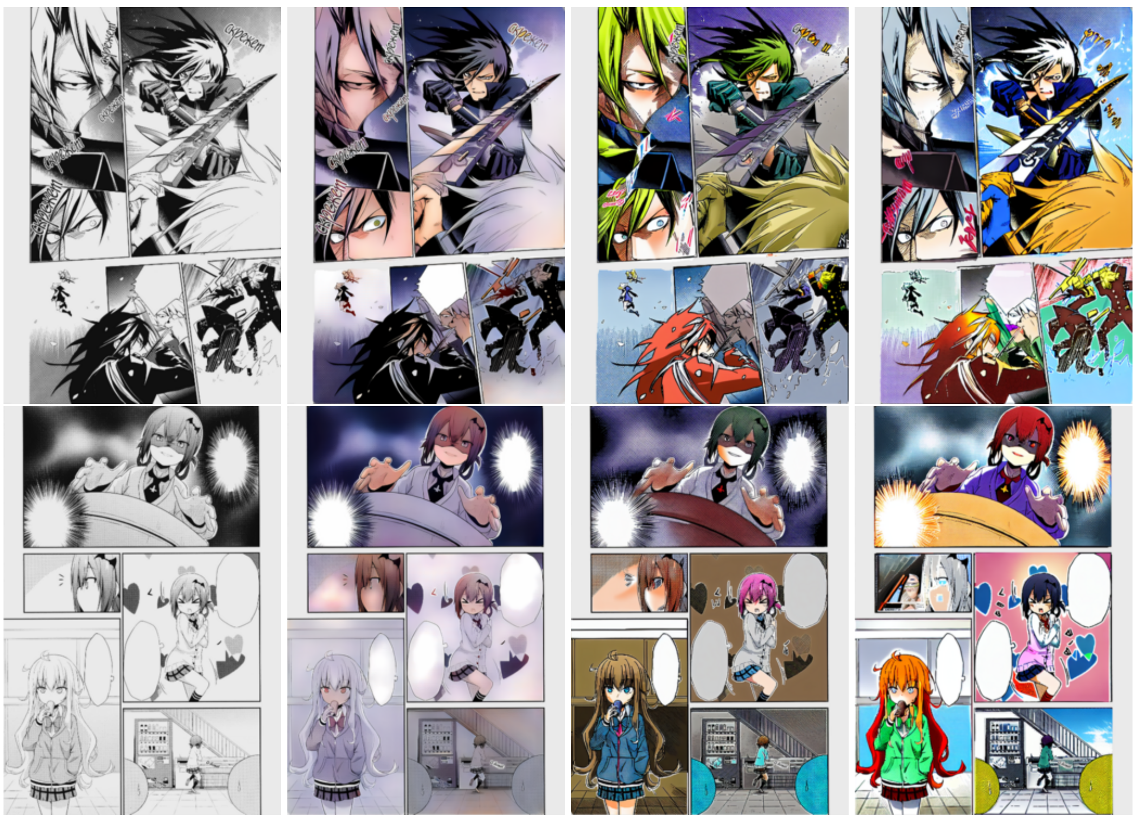
(a) Input (b) mv-v2 (IN) (c) IP2P (IN) (d) CN (IN) (e) mc-v2 (WA) (f) IP2P (WA) (g) CN (WA)

Figure 3. **Comparison of baseline models trained on ImageNet and WikiArt**. The figure demonstrates that models trained on non-manga datasets are not suitable for manga colorization.



(a) Input (b) Petalica (c) Komiko (d) RecolorAI (e) S2P (f) mc-v2

Figure 4. **Comparison of manga-colorization-v2 with existing colorization applications**. This figure compares the manga-colorization-v2 model trained on our dataset with several popular colorization applications that are not suitable for fine-tuning: Petalica [1], Komiko [2], RecolorAI [3], and Style2Paints (S2P) [5]. The results demonstrate that these applications perform poorly in terms of color accuracy and structural preservation, highlighting the advantages of using a model specifically trained for manga colorization.



(a) Input (b) mc-v2 (c) CN (d) IP2P

Figure 5. **Comparison of panel-level colorization results.** This figure compares the performance of manga-colorization-v2 (mc-v2), ControlNet (CN), and InstructPix2Pix (IP2P) models trained on panel-level data. The resulting pages are created by placing colorized panels back into the original image. Similar to the page-level results, diffusion-based approaches (ControlNet and InstructPix2Pix) demonstrate better color diversity, while the cGAN-based method (mc-v2) excels in structural preservation. However, since each panel is processed independently, allowing for a higher resolution within the same computational budget, the loss of structural information is reduced compared to page-level processing.

(a) Color (b) Real BW (c) Synthetic BW

Figure 6. **Comparison of real and synthetic black-and-white (BW) images**. This figure compares real BW images, matched to corresponding color images, with synthetic BW images generated from color ones. The synthetic images tend to overly rely on the content of the original color images, whereas the real BW images are more sparse and often filled with screentones. This demonstrates that synthetic data are not an exact replica of real BW images, which introduces a domain gap for models trained on synthetic data.