Sketch Colorization Using Diffusion Models and Photo-Sketch Correspondence

Axel Delaval, Adama Koïta
École Polytechnique — Telecom-Paris
Email: axel.delaval@gmail.com, adama.koita@telecom-paris.fr

GitHub Repository

***Abstract*—Sketch colorization has gained significant attention in recent years, particularly in the context of anime and manga artwork. Recent advancements in diffusion models and deep learning have led to substantial improvements in automated colorization techniques. This project explores various approaches to anime sketch colorization, leveraging diffusion models and deep learning architectures.**

Our final model draws inspiration primarily from AnimeDiffusion [1], MangaNinja [2], and photo-sketch correspondence models [3]. We also explored ideas from feature matching and robust visual representation models [4]–[10].

Due to computational constraints, we trained our model on a subset of the Danbooru dataset [11]. However, we also curated an extensive dataset combining AnimeFace [12] and AnimeDiffusion [1], enriched with augmented deformation flows to generate diverse training samples.

***Index Terms*—Sketch Colorization, Diffusion Models, Anime Art, Deep Learning, Feature Matching, Image Processing**

Contents

I. Introduction

The process of colorizing sketches, particularly in the domain of anime and manga, presents a challenging problem at the intersection of computer vision and deep learning. Traditional methods relied heavily on manual intervention and artistic expertise. However, with the rise of generative models and diffusion-based approaches, automated sketch colorization has become an increasingly viable solution.

Recent breakthroughs in diffusion models [1], [2] have significantly improved the accuracy and expressiveness of colorized outputs. Additionally, learning dense correspondences between photos and sketches [3] enhances structural alignment and color consistency in the final renders. While these methods form the core of our approach, we also investigated techniques for robust visual feature extraction [6], [7] and feature matching [4], [5].

Our dataset comprises anime face images labeled with character names. For training, we utilized a subset of the Danbooru dataset [11] due to memory and computation constraints. However, we prepared a larger dataset by combining AnimeFace [12] and AnimeDiffusion [1], applying procedural augmentation techniques to generate additional variations per character. In AnimeDiffusion's dataset [1], pre-sketched images were available, whereas for the other datasets, we generated sketches using procedural techniques.

This article will be further expanded to include detailed methodology, evaluation, and results. For now, it serves as a reference for our approach and cited works.

References

[1] Y. Cao, X. Meng, P. Mok, X. Liu, T.-Y. Lee, and P. Li, "Animediffusion: Anime face line drawing colorization via diffusion models," *arXiv preprint arXiv:2303.11137*, 2023.

[2] Z. Liu, K. L. Cheng, X. Chen, J. Xiao, H. Ouyang, K. Zhu, Y. Liu, Y. Shen, Q. Chen, and P. Luo, "Manganinja: Line art colorization with precise reference following," *arXiv preprint arXiv:2501.08332*, 2025.

[3] X. Lu, X. Wang, and J. E. Fan, "Learning dense correspondences between photos and sketches," 2023. [Online]. Available: https://arxiv.org/abs/2307.12967

[4] P.-E. Sarlin, D. DeTone, T. Malisiewicz, and A. Rabinovich, "Superglue: Learning feature matching with graph neural networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 4938–4947.

[5] J. Sun, Z. Shen, Y. Wang, H. Bao, and X. Zhou, "Loftr: Detector-free local feature matching with transformers," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 8922–8931.

[6] M. Oquab, T. Darcet, T. Moutakanni, H. Vo, M. Szafraniec, V. Khalidov, P. Fernandez, D. Haziza, F. Massa, A. El-Nouby *et al.*, "Dinov2: Learning robust visual features without supervision," *arXiv preprint arXiv:2304.07193*, 2023.

[7] L. Tang, M. Jia, Q. Wang, C. P. Phoo, and B. Hariharan, "Emergent correspondence from image diffusion," *Advances in Neural Information Processing Systems*, vol. 36, pp. 1363–1389, 2023.

[8] S. Koley, A. K. Bhunia, A. Sain, P. N. Chowdhury, T. Xiang, and Y.-Z. Song, "Text-to-image diffusion models are great sketch-photo matchmakers," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 16 826–16 837.

[9] K. Gupta, V. Jampani, C. Esteves, A. Shrivastava, A. Makadia, N. Snavely, and A. Kar, "Asic: Aligning sparse in-the-wild image collections," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 4134–4145.

[10] J. Zhang, C. Herrmann, J. Hur, L. Polania Cabrera, V. Jampani, D. Sun, and M.-H. Yang, "A tale of two features: Stable diffusion complements dino for zero-shot semantic correspondence," *Advances in Neural Information Processing Systems*, vol. 36, pp. 45 533–45 547, 2023.

[11] M. O'Neill, "Tagged anime illustrations," 2019, dataset containing tagged anime illustrations from Danbooru. [Online]. Available: https://www.kaggle.com/datasets/mylesoneill/tagged-anime-illustrations/data

[12] TheDevastator, "Anime face dataset by character name," 2023, dataset suitable for image classification, containing 130 anime characters with 75 images each, scrapped from Danbooru. [Online]. Available: https://www.kaggle.com/datasets/thedevastator/anime-face-dataset-by-character-name?resource=download