2023 IEEE International Conference on Multimedia and Expo (ICME) | 978-1-6654-6891-6/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICME55011.2023.00282

Attention-Aware Anime Line Drawing Colorization

Yu Cao, Hao Tian and P. Y. Mok*
The Hong Kong Polytechnic University
Kowloon, Hong Kong
Email: tracy.mok@polyu.edu.hk

***Abstract*—Automatic colorization of anime line drawing has attracted much attention in recent years since it can substantially benefit the animation industry. User-hint based methods are the mainstream approach for line drawing colorization, while reference-based methods offer a more intuitive approach. Nevertheless, although reference-based methods can improve feature aggregation of the reference image and the line drawing, the colorization results are not compelling in terms of color consistency or semantic correspondence. In this paper, we introduce an attention-based model for anime line drawing colorization, in which a channel-wise and spatial-wise Convolutional Attention module is used to improve the ability of the encoder for feature extraction and key area perception, and a Stop-Gradient Attention module with cross-attention and self-attention is used to tackle the cross-domain long-range dependency problem. Extensive experiments show that our method outperforms other SOTA methods, with more accurate line structure and semantic color information.**

***Index Terms*—Attention Mechanism, Line Drawing Colorization, Conditional Generation**

I. Introduction

Image colorization has attracted extensive research attention [1]–[3] in the field of Computer Graphics and Multimedia. Line drawing colorization is an essential process in the animation industry. For example, manga and cartoon line drawing colorization is a practical application of this, while manual colorizing is time consuming, especially for line drawings with complex structure. In the past, different methods [4]–[6], mainly generative adversarial network (GAN)-based, have been proposed to speed up the process. Line drawing colorization is challenging, because line drawings, different from grayscale images, only contain structure content composing of a series of lines without any luminance or texture information.

The colorizing process can be generally regarded as a conditional image-to-image translation problem that directly converts the input line drawings to the output colored results. Early work [5] utilized a neural network to automatically colorize the cartoon images with random colors, which is the first deep learning based cartoon colorization method. Nevertheless, many interactions are usually needed to refine the colored results to satisfy what the user specified. To effectively control the color of the result, many user-hint based methods have been proposed successively, such as point colors [7]–[9], scribble colors [10], text-hint [11], and language-based [12]. These user-hint based methods are still not convenient or intuitive, especially for amateur users without aesthetic training.

Fig. 1. Sample images of line drawing, reference color image, and generated colored result of our model are from left to right. All three images have the resolution of 256×256.

Reference-based colorization methods [13]–[15], typically condtional GAN networks, provide a more convenient way, which can automatically complete the colorizing process without other manual intervention. Users only need to prepare a line drawing and a corresponding reference color image, see Fig. 1 as an illustration. In the typical approach, two encoders are used to extract the line drawing feature and reference color feature respectively, and then a feature aggregation block is designed to inject color from the corresponding position of reference image into line drawing. Lee *et al.* [14] proposed an attention-based Spatial Correspondence Feature Transfer (SCFT) module, and quantitatively proved its feature aggregation ability is superior to addition block and AdaIN [16] block. Li *et al.* [15] eliminated the gradient conflict among attention branches using Stop-Gradient Attention (SGA) module. Meanwhile, these methods bring new challenges in the color consistency and semantic correspondence between the colored image and reference image. It is attractive to design a line drawing colorization algorithm that meets the above two conditions, which will greatly reduce the tedious work of animators.

In this paper, we propose a novel conditional adversarial colorization network combining Convolutional Attention module and Stop-Gradient Attention module trained fully on anime line drawing dataset [17]. We first design a line drawing extraction network based on this dataset. Then, we train our colorization model through a proxy task of line drawing guided distorted image restoration. The encoder equipped with Con-

*Corresponding author.

The work described in this paper is supported in part by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Grant Number 152112/19E) and by the Innovation and Technology Commission of Hong Kong, under grant ITP/028/21TP.

1945-788X/23/$31.00 ©2023 IEEE
DOI 10.1109/ICME55011.2023.00282

volutional Attention module is introduced to help the encoder extract exact multi-scale features. Therefore, the colorization results of our model have both geometry details and accurate colors. Meanwhile, Stop-Gradient Attention module shows great feature aggregation performance and gives our model better semantic correspondence ability. With the attention-based network, we can generate colored results which are comparable to manual coloring results (see an example in Fig. 1). Our main contributions can be summarized as follows:

- We develop an anime line drawing colorization method based on attention-aware mechanism by integrating Convolutional Attention module and Stop-Gradient Attention module so as to improve the color consistency and semantic correspondence of coloring results.
- We design an U-Net [18] based network to extract line drawing from color images to simulate manually drawn ones. Since the number of available dataset for paired anime line drawing and color image is limited, this trained model can be used for data augmentation to facilitate colorization network training.
- Both qualitative and quantitative comparative analyses show that our method has outperformed other reference-based state-of-the-art methods in the task of anime line drawing colorization via training on a proxy task of line drawing guided distorted image restoration.

II. Related Work

A. Line Drawing Colorization

Since line drawing contains only structure information with sparse line sets, existing colorization methods for gray-scale images cannot be used directly. Traditional line drawing colorization approaches [4], [6] are commonly optimized-based, allowing users to use brushes to inject desired colors into specific regions. With the advancement of deep learning and for better control of color, many user-hint colorization methods [8], [11], [12] spring up. However, the complexity of such user-hint methods will become more labor-intensive as the number of line drawings increases, and many interactions are also needed. Therefore, many reference-based colorization methods [13]–[15], [19] have been proposed, which are very suitable for colorizing line drawing sets or videos of anime characters. Our proposed colorization model belongs to this reference-based method, and we can generate more faithful results with fantastic visual quality.

B. Semantic Correspondence

Semantic correspondence [20] is one of the fundamental problems in computer vision whose goal is to establish dense correspondences across images containing the targets of the same category or with similar semantic information. Reference-based line art colorization can also be viewed as cross-domain correspondence since the texture difference between line drawing and reference color image. Most semantic correspondence learning methods are designed for the grayscale image colorization. However, our proposed method relies on the reasonable design of the attention mechanism module, which can achieve plausible results for anime line drawing colorization.

C. Attention Mechanism

With the attention mechanism [21], [22] showing good performance in feature extraction and aggregation in image perception task, it makes neural model much closer to the human visual perception system. Recently some attention-based line drawing colorization methods [9], [14], [15] have been proposed for improving the model's colorization ability. For grayscale image colorization, Kumar *et al.* [23] presented the Colorization Transformer that entirely relies on self-attention for image colorization. Since few research work focuses on line drawing colorization using attention-based method, we mainly compare our method with Lee *et al.* [14] and Li *et al.* [15], and our method achieves the state-of-the-art result.

III. Method

Given a line drawing and a reference color image, our network can generate a colored result which contains clear geometry structure of the line drawing and accurate semantic color of the reference image. We formulate the problem as a novel conditional adversarial network in which the training process itself is regarded as line drawing guided distorted reference color image restoration, as shown in Fig. 2. Inspired by [14], we train our model in a self-augmented supervised manner, enabling the network to be trained on limited paired data. The trained model can perform reference-based colorization for other line drawing inputs during inference.

A. Model Architecture

As illustrated in Fig. 2, assuming I_{gt} is an original colored image, and an line drawing I_l is extracted from I_{gt} using line drawing extraction network LE. The distorted reference image I_r is extracted from I_{gt} using Thin Plate Spline (TPS) transformation [24]. It is important to note that there are usually large spatial structure discrepancy between the reference color image and the line drawing, thus a large volume of paired data of before and after colorization is needed to train a colorization network. Nevertheless, there are limited data of amine line drawing and it is very costly to prepare a dataset with such paired data. To address this problem, we first train a LE network to extract line drawings from colored amine images using an existing dataset [17]. With trained LE network, we augment data by generating corresponding anime line drawings from color images obtained from the internet.

Our model consists of a generator and a discriminator as well as a specially designed "Attention-aware mechanism". For the generator, there are two encoders E_l and E_r, which are used to extract feature from I_l and I_r, respectively. Note that the input channel number of E_l is 1, while the input channel number of E_r is 3. By designing two feature extractors, the line feature and color feature of the image are disentangled [25]. The extracted line drawing feature maps F_l and color style feature maps F_r will go through a Convolutional Attention sub-module and a Feature Dimension

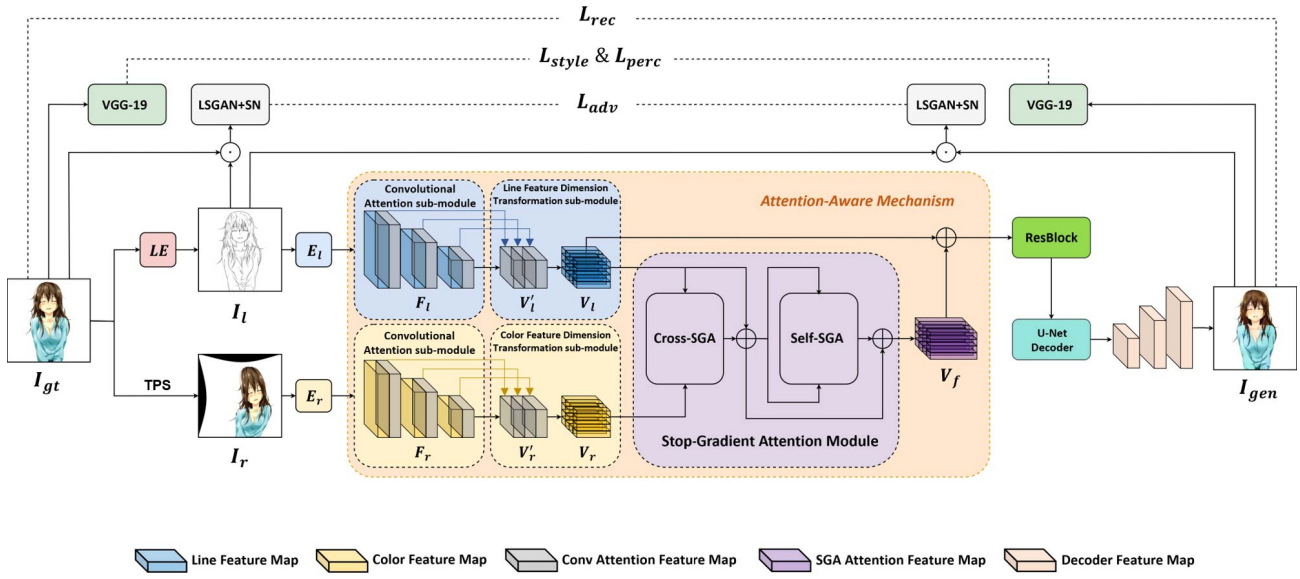

Fig. 2. Overview of the training pipeline for our network.

Transformation sub-module to get multi-scale attention feature map V_l and V_r. Next, we adopt the Stop-Gradient Attention (SGA) module to fuse V_l and V_r to get the fusion feature V_f. As shown in Fig. 2, the SGA module contains cross-attention and self-attention blocks to better acquire dense semantic correspondence and to tackle cross-domain long-range dependency problem. Then V_f and V_l will add together and pass through several residual blocks, and followed by a U-Net [18] decoder to generate restored result I_{gen}.

For the discriminator, we utilize conditional LSGAN [26] combined with Spectral Normalization(SN) [27] for stable training. Both generator and discriminator use Batch Normalization (BN) to accelerate the training speed, but there is no BN or SN in the final layer of the discriminator.

B. Attention-Aware Mechanism

Our model aims to generate the colored results, which contain semantic color and style information from the reference color image while present the same visual content as the target line drawing. To this end, an attention mechanism is designed to learn three essential features: 1) the content of the line drawing, 2) the color information of the reference image, and 3) the visual correspondences between the line drawing and the reference image.

1) Convolutional Attention (CA) Module: To improve the feature extraction ability of the two encoders E_l and E_r, which take into account both global and local features, we add spatial and channel attention mechanisms behind each convolution layer of the encoder. We use convolution attention (CA) module proposed in [28] as a sub-module of our attention operation. Such operation enables the network to adaptively extract important features from reference image and line drawing, thus the colored results will have accurate colors, natural transitions with clear line structure [28]. We design a Feature Dimension Transformation sub-module to integrate multi-scale features extracted from CA sub-module. More specifically, the features from the Convolutional Attention sub-module will be resized to the size of the final layer feature map, concatenate respectively on the channel dimension, and then transpose their spatial dimension with the channel dimension to generate V_l and V_r, which contains multi-scale feature.

2) Stop-Gradient Attention (SGA) Module: Since there exists a gradient conflict issue in self-attention based feature aggregation module, we utilize the Stop-Gradient Attention module [15] to integrate the line drawing feature and reference color image feature. The key operation is truncating the back propagation of the gradient of attention matrix. Inside the SGA module, the attention matrix is normalized through row dimension and column dimension respectively, and we use two SGA modules equipped with short connections for cross-attention and self-attention separately. By integrating SGA module with convolutional attention module, our model shows good performance in semantic correspondence and generates colored results in high quality.

C. Loss Function

To train our network for high quality colorization results, namely with better color consistency and semantic correspondence, we define the loss function as follows.

L1 Loss. Since the main training task of our model is distorted image restoration, the pixel-level reconstruction loss is needed. We use L1 loss rather than L2 loss as L1 encourages less blurring [29]. We directly use it to penalize the model for the pixel-level loss between the restored image I_{gen} and

original image I_{gt}:

$$L_{rec} = E[\| G(I_l, I_r) - I_{gt} \|_1] \quad (1)$$
$$= E[\| I_{gen} - I_{gt} \|_1], \quad (2)$$

where G denotes the generator.

Perceptual Loss. We use the VGG-19 model pretrained on ImageNet as feature extractor to compute the high-level semantic and low-level perceptual loss [30] through multiple feature layers. The loss function is defined as follows:

$$L_{perc} = E[\sum_{l} \| \phi_l(I_{gen}) - \phi_l(I_{gt}) \|_1], \quad (3)$$

where ϕ_l represents the l^{th} feature layer after ReLU activation function of VGG-19.

Style Loss. To generate image with more clear appearance and to address checkboard artifacts, a style loss [30] is introduced as follows:

$$L_{style} = E[\sum_{l} \| \mathcal{G}(\phi_l(I_{gen})) - \mathcal{G}(\phi_l(I_{gt})) \|_1], \quad (4)$$

where $\mathcal{G}$ means the gram matrix.

Adversarial Loss. Since our model is a kind of generative model, in order to improve the generation ability and make the training process more stable, we leverage the conditional LSGAN [26] objective function equipped with Spectral Normalization (SN) to calculate an adversarial loss:

$$L_{adv} = E[\| D(I_{gt}, I_l) \|_2^2] + E[\| 1 - D(G(I_l, I_r), I_l) \|_2^2], \quad (5)$$

where D denotes the discriminator.

Total loss function. We set the optimization goal by combining all the above losses as follows:

$$\min_G \max_D L = \lambda_{adv} L_{adv} + \lambda_{rec} L_{rec} + \lambda_{perc} L_{perc} + \lambda_{style} L_{style}, \quad (6)$$

where λ_{adv}, λ_{rec}, λ_{perc} and λ_{style} represent the corresponding weights of the adversarial loss, L1 loss, perceptual loss and style loss. We set $\lambda_{adv} = 1$, $\lambda_{rec} = 30$, $\lambda_{perc} = 0.01$, and $\lambda_{style} = 50$.

IV. Experiments

A. Data Preparation

We train our model with an anime dataset consisting of 17769 images shared on Kaggle website [17]. We first train a line drawing extraction network based on this dataset. We use TPS transformation [24] to convert the original color anime image to a geometrically distorted image online along with the training of our colorization network (Fig. 2). The full data are used to train our model without separating validation data. We can use our trained line drawing extraction network to prepare extra data for testing, similar data augmentation idea is used in colorization task [8], [9].

B. Implementation Details

We implement our method using PyTorch framework and the model is trained on a NVIDIA 3090 GPU with a batch size of 16. We set the total iteration number as 250000, and both the generator and the discriminator were alternately updated in every iteration. The input size of all images is set at 256×256 and the pixel value is normalized to the range of [-1, 1]. Weight parameters for five ReLU activated feature layers are all set to 1 for computing the perceptual loss and style loss. We use Adam optimizer with β_1=0.5, β_2=0.999. The learning rates of generator and discriminator are set to 0.0001 and 0.0002, respectively, and we do not use extra learning rate update strategy during training.

C. Qualitative Evaluation

Fig. 3 illustrates the visual comparisons between our method and two other state-of-the-art reference-based colorization methods [14], [15] trained on the same dataset [17] and two non-learning based user-hint colorization tools [7], [10]. For a fair comparison, all three learning based models are pretrained to the extent of convergence with batch size of 16. As shown in Fig. 3, Lee *et al.* [14] generates results with color bleeding, especially in the region of hair and clothes. Li *et al.* [15] generates some results with distorted color compared with the reference images. The two user-hint colorization methods using color scribbles [10] and color points [7] as inputs, respectively. They are not very convenient to use because manually specifying the colors from the reference image is imprecise and prone to color aliasing and semantic interference problems during the coloring process. In contrast, reference-based colorization method has obvious advantage in terms of easy operation. Comparatively, our method offers good semantic correspondence in the parts of hair, eyes and clothes. In addition, we also find from the experiments that our model with Convolutional Attention mechanism converges faster than [15] does during the training stage.

D. Quantitative Evaluation and Ablation Study

We set two kinds of coloring cases to evaluate the performance of our network, self-reference colorization and random-reference colorization. For self-reference colorization, the line drawing and reference image are paired, ideally the colorized output should be exactly the same as the reference image. In order to evaluate the colorization quality, we measure the structure similarity and perceptual similarity between the colored result and reference image using Peak Signal-to-Noise Ratio (PSNR), Multi-Scale Structural Similarity Index Measure (MS-SSIM) and Learned Perceptual Image Patch Similarity (LPIPS) [31]. For the random-reference colorization, it is more like the common practical usage when using exemplar-based colorization method. To evaluate the generative ability of GAN-based network, we perform a quantitative study by calculating the Fréchet Inception Distance (FID) score between the colorized result and reference image. A smaller FID indicates that the distribution of the colored image is closer to the reference color image.

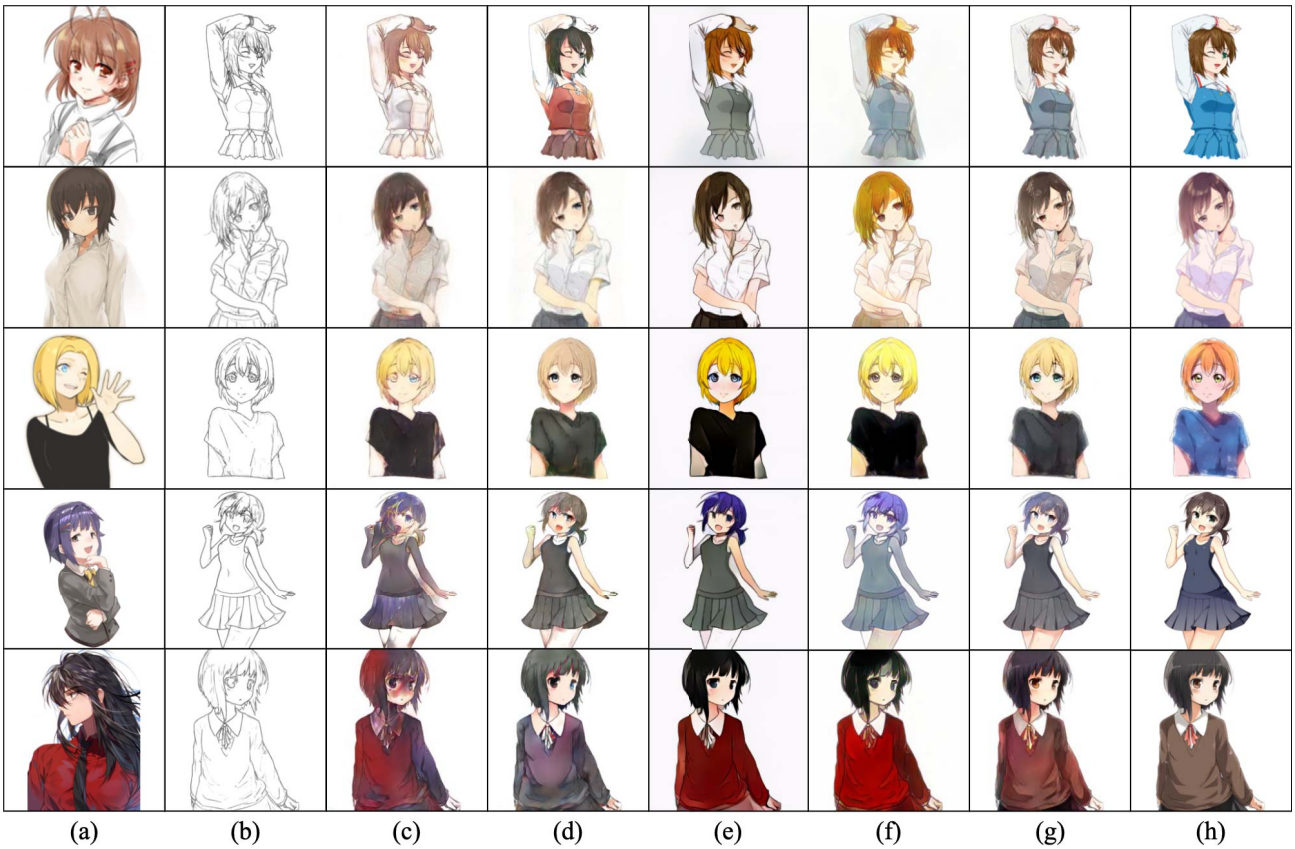

Fig. 3. Qualitative result comparison: (a) reference color images, (b) line drawings, the results of (c) Lee *et al.* [14], (d) Li *et al.* [15], (e) Petalica [10], (f) Adeleine [7], (g) ours, and (h) original color images

As is shown in Table I, our method achieves the best FID, 17.35% improvement in comparison to [15] and 50.55% improvement in comparison to [14]. For three metrics of self-reference colorization, our method is still the best one. It indicates that through training on the proxy task of image restoration, our model not only acquires good image restoration performance but also can be used to line drawing colorization.

Table I can also serve as ablation study, in which Lee *et al.* [14] can be regarded as baseline, while Li *et al.* [15] utilizes SGA module only, our method integrates both CA and SGA modules together. It shows that CA module itself (i.e. [14]+CA) or SGA module itself (i.e. [15]) are not as effective as the current proposal of attention-aware mechanism that integrates both.

TABLE I
QUANTITATIVE COMPARISON WITH STATE-OF-THE-ART METHODS

Method	FID↓	PSNR↑	MS-SSIM↑	LPIPS↓
Lee *et al.* [14]	47.20	26.36	0.95	0.06
Lee *et al.* [14] + CA	40.08	24.70	0.92	0.09
Li *et al.* [15]	28.24	21.01	0.94	0.07
Ours	**23.34**	**27.21**	**0.96**	**0.05**

V. CONCLUSION

In this paper, we propose a new attention-based colorization algorithm for anime line drawings. To do so, we first train a line drawing extractor using existing anime dataset available from the internet. The trained line drawing extractor is integrated in the model training of our proposed attention-based colorization network for data augmentation. Our method achieves faithful colorization results using a novel attention mechanism by integrating (1) a Convolutional Attention module containing channel-wise and spatial-wise attention block, and (2) a Stop Gradient-Attention module including cross-attention and self-attention block. Through extensive experiments, our method has demonstrated better performance both qualitatively and quantitatively, outperforming other state-of-the-art methods, with more accurate line structure and semantic color information.

ACKNOWLEDGMENT

The authors wish to thank Leonard Chen, Mandy Wong and Rachel Liu for their contribution in dataset preparation.

REFERENCES

[1] Menghan Xia, Wenbo Hu, Tien-Tsin Wong, and Jue Wang, "Disentangled image colorization via global anchors," *ACM Transactions on Graphics (TOG)*, vol. 41, no. 6, pp. 204:1–204:13, 2022.

[2] Zhitong Huang, Nanxuan Zhao, and Jing Liao, "Unicolor: A unified framework for multi-modal colorization with transformer," *arXiv preprint arXiv:2209.11223*, 2022.

[3] Peng Lu, Jinbei Yu, Xujun Peng, Zhaoran Zhao, and Xiaojie Wang, "Gray2colornet: Transfer more colors from reference image," in *Proceedings of the 28th ACM International Conference on Multimedia*, 2020, pp. 3210–3218.

[4] Yingge Qu, Tien-Tsin Wong, and Pheng-Ann Heng, "Manga colorization," *ACM Transactions on Graphics (TOG)*, vol. 25, no. 3, pp. 1214–1220, 2006.

[5] Domonkos Varga, Csaba Attila Szabo, and Tamas Sziranyi, "Automatic cartoon colorization based on convolutional neural network," in *Proceedings of the 15th International Workshop on Content-Based Multimedia Indexing*, 2017, pp. 1–6.

[6] Daniel Sỳkora, John Dingliana, and Steven Collins, "Lazybrush: Flexible painting tool for hand-drawn cartoons," in *Computer Graphics Forum*. Wiley Online Library, 2009, vol. 28, pp. 599–608.

[7] Adeleine, "Adeleine colorization," https://github.com/SerialLain3170/Colorization/tree/master/Adeleine, 2021.

[8] Yuanzheng Ci, Xinzhu Ma, Zhihui Wang, Haojie Li, and Zhongxuan Luo, "User-guided deep anime line art colorization with conditional adversarial networks," in *Proceedings of the 26th ACM international conference on Multimedia*, 2018, pp. 1536–1544.

[9] Mingcheng Yuan and Edgar Simo-Serra, "Line art colorization with concatenated spatial attention," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 3946–3950.

[10] Petalica, "Petalica paint," https://petalica.com/index$_e n.html$, 2019.

[11] Hyunsu Kim, Ho Young Jhoo, Eunhyeok Park, and Sungjoo Yoo, "Tag2pix: Line art colorization using text tag with secat and changing loss," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 9056–9065.

[12] Changqing Zou, Haoran Mo, Chengying Gao, Ruofei Du, and Hongbo Fu, "Language-based colorization of scene sketches," *ACM Transactions on Graphics (TOG)*, vol. 38, no. 6, pp. 1–16, 2019.

[13] Shu-Yu Chen, Jia-Qi Zhang, Lin Gao, Yue He, Shihong Xia, Min Shi, and Fang-Lue Zhang, "Active colorization for cartoon line drawings," *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 2, pp. 1198–1208, 2020.

[14] Junsoo Lee, Eungyeup Kim, Yunsung Lee, Dongjun Kim, Jaehyuk Chang, and Jaegul Choo, "Reference-based sketch image colorization using augmented-self reference and dense semantic correspondence," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 5801–5810.

[15] Zekun Li, Zhengyang Geng, Zhao Kang, Wenyu Chen, and Yibo Yang, "Eliminating gradient conflict in reference-based line-art colorization," *arXiv preprint arXiv:2207.06095*, 2022.

[16] Xun Huang and Serge Belongie, "Arbitrary style transfer in real-time with adaptive instance normalization," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 1501–1510.

[17] Taebum Kim, "Anime sketch colorization paired dataset from danbooru," https://www.kaggle.com/ktaebum/anime-sketch-colorization-pair, 2018.

[18] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[19] Chie Furusawa, Kazuyuki Hiroshiba, Keisuke Ogaki, and Yuri Odagiri, "Comicolorization: semi-automatic manga colorization," in *SIGGRAPH Asia 2017 Technical Briefs*, pp. 1–4. 2017.

[20] Taihong Xiao, Sifei Liu, Shalini De Mello, Zhiding Yu, Jan Kautz, and Ming-Hsuan Yang, "Learning contrastive representation for semantic correspondence," *International Journal of Computer Vision*, vol. 130, no. 5, pp. 1293–1309, 2022.

[21] Jie Hu, Li Shen, and Gang Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.

[22] Miao Hu, Yali Li, Lu Fang, and Shengjin Wang, "A2-fpn: Attention aggregation based feature pyramid network for instance segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 15343–15352.

[23] Manoj Kumar, Dirk Weissenborn, and Nal Kalchbrenner, "Colorization transformer," *arXiv preprint arXiv:2102.04432*, 2021.

[24] Fred L. Bookstein, "Principal warps: Thin-plate splines and the decomposition of deformations," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 11, no. 6, pp. 567–585, 1989.

[25] Xun Huang, Ming-Yu Liu, Serge Belongie, and Jan Kautz, "Multimodal unsupervised image-to-image translation," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 172–189.

[26] Xudong Mao, Qing Li, Haoran Xie, Raymond YK Lau, Zhen Wang, and Stephen Paul Smolley, "Least squares generative adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2794–2802.

[27] Takeru Miyato, Toshiki Kataoka, Masanori Koyama, and Yuichi Yoshida, "Spectral normalization for generative adversarial networks," *arXiv preprint arXiv:1802.05957*, 2018.

[28] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3–19.

[29] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.

[30] Justin Johnson, Alexandre Alahi, and Li Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14*. Springer, 2016, pp. 694–711.

[31] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 586–595.