Stylized-Colorization for Line Arts

Tzu-Ting Fang
National Tsing Hua University
Hsinchu, Taiwan

Duc Minh Vo
SOKENDAI
Tokyo, Japan

Akihiro Sugimoto
National Institute of Informatics
Tokyo, Japan

Shang-Hong Lai
National Tsing Hua University
Hsinchu, Taiwan

Introduction

- The styles give different impression to the same anime character illustration.
- We address a novel problem of stylized-colorization which colorizes a given line art using a given coloring style in text.
- It can be stated as multi-domain image translation.
- Our task is challenging because it has 2 aspects: colorization and style transfer.
- We proposed a novel style feature loss to enhance the dissimilarity between different coloring styles.

Objective Function

- **Adversarial loss**
  - **Hinge loss** helps the adversarial learning to be strong and stable.

$$\mathcal{L}_{D_c} = \mathbb{E}_{y_c \sim p_{\text{data}}}\left[\max(0, 1 - D_c(y_c))\right] + \mathbb{E}_{(x,s)\sim p_{\text{data}}}\left[\max(0, 1 + D_c(G(x,s))\right], \quad (1)$$

$$\mathcal{L}_{D_s} = \mathbb{E}_{(y_s,s) \sim p_{\text{data}}}\left[\max(0, 1 - D_s(y_s, s))\right] + \mathbb{E}_{(x,s)\sim p_{\text{data}}}\left[\max(0, 1 + D_s(G(x,s), s)\right], \quad (2)$$

$$\mathcal{L}_G = -\mathbb{E}_{(x,s)\sim p_{\text{data}}}\left[D_c(G(x,s)\right] - \mathbb{E}_{(x,s)\sim p_{\text{data}}}\left[D_s(G(x,s), s))\right]. \quad (3)$$

- **Per-pixel loss**
  - **L1 distance** between the generated image and ground truth colored illustration.
  - Enforcing the generator learning the color and keeping the image structure.

$$\mathcal{L}_{\text{pix}} = \mathbb{E}_{(y_c,x,s)\sim p_{\text{data}}} \|y_c - G(x,s)\|_1 \quad (4)$$

- **Style feature loss**
  - Firstly, we pre-trained a style classifier **C** by employing the **center loss**.
  - Then, we obtained the feature representation c_s for coloring style s to define our style feature loss:

$$\mathcal{L}_{\text{style}} = \mathbb{E}_{(x,s)\sim p_{\text{data}}}\left[\|\mathbf{c}_s - C(G(x,s))\|_2^2\right], \quad (5)$$

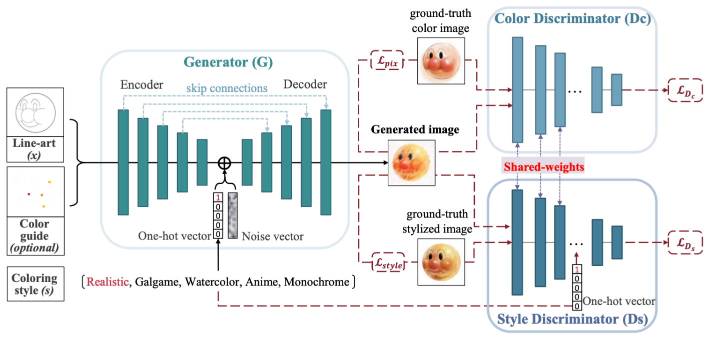
x: input line-art
s: the coloring style
y_c: ground-truth colored image
y_s: ground-truth stylized image
s: feature representation of style s
$C(G(\cdot))$: the style feature of generated image

- In summary, our full objective function is:

$$\mathcal{L} = \mathcal{L}_G + \lambda_{\text{pix}}\mathcal{L}_{\text{pix}} + \lambda_{\text{style}}\mathcal{L}_{\text{style}}, \quad (6)$$

, where λ_{pix} and λ_{style} are the hyper-parameters.

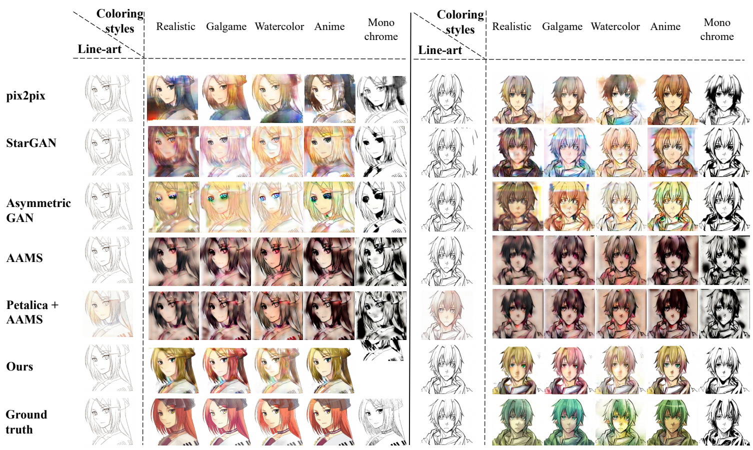
Proposed Method



- We propose a GAN-based end-to-end model.
- The model has one generator and two discriminators.
- **Generator:**
  - Based on the U-Net architecture.
  - Receive a pair of a line art and a coloring style in text as its inputs.
- **Discriminators:**
  - Two discriminators share weights at early layers.
  - Judge the generated image in two aspects: one for color and the other for style.

Results

- Visual comparison against other methods:



Quantitative Comparison

- We adopted three metrics for evaluation.
  - PSNR, SSIM and FID
- R, G, W, A, M stand for **R**ealistic, **G**algame, **W**atercolor, **A**nime, **M**onochrome, respectively.

Metric	PSNR ↑					SSIM ↑					FID ↓			
Coloring style	R	G	W	A	M	R	G	W	A	M	R	W	A	M
pix2pix [10]	8.378	11.028	11.488	10.535	7.357	0.389	0.516	0.501	0.506	0.407	146.62	123.91	129.03	93.87
StarGAN [12]	4.978	5.062	6.288	5.112	3.649	0.145	0.158	0.222	0.184	0.165	165.83	161.03	176.79	106.89
AsymmetricGAN [13]	5.180	5.095	6.047	4.927	3.486	0.165	0.151	0.226	0.201	0.157	173.20	163.75	197.24	114.30
AAMS [23]	8.995	9.379	10.867	8.873	6.866	0.363	0.436	0.414	0.362	0.297	167.27	151.42	187.63	189.27
Petalica+AAMS	9.266	9.504	10.764	9.454	6.688	0.377	0.459	0.442	0.412	0.302	151.86	126.66	160.61	141.69
Ours	**10.753**	**11.511**	**12.957**	**11.165**	**7.799**	**0.455**	**0.573**	**0.563**	**0.548**	**0.445**	**118.88**	**117.38**	**125.49**	**88.34**