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.
  \[
  \mathcal{L}_{D_c} = \mathbb{E}_{y_c \sim p_{\text{data}}} \left[ \max(0, 1 - D_c(y_c)) \right] + \mathbb{E}_{(x,s) \sim p_{\text{data}}} \left[ \max(0, 1 + D_c(G(x,s))) \right], \]
  \[
  \mathcal{L}_{D_s} = \mathbb{E}_{y_s \sim p_{\text{data}}} \left[ \max(0, 1 - D_s(y_s)) \right] + \mathbb{E}_{(x,s) \sim p_{\text{data}}} \left[ \max(0, 1 + D_s(G(x,s,s))) \right],
  \]
• 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.
  \[
  \mathcal{L}_{\text{pix}} = \mathbb{E}_{(y_c,x,s) \sim p_{\text{data}}} || y_c - G(x,s) ||_1
  \]
• 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:
  \[
  \mathcal{L}_{\text{style}} = \mathbb{E}_{(x,s) \sim p_{\text{data}}} \left[ || c_s - C(G(x,s)) ||_2^2 \right],
  \]

In summary, our full objective function is:

\[
\mathcal{L} = \mathcal{L}_G + \lambda_{\text{pix}} \mathcal{L}_{\text{pix}} + \lambda_{\text{style}} \mathcal{L}_{\text{style}},
\]

where \( \lambda_{\text{pix}} \) and \( \lambda_{\text{style}} \) 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 Realistic, Galgame, Watercolor, Anime and Monochrome, respectively.