DRB-GAN: A Dynamic ResBlock Generative Adversarial Network for Artistic Style Transfer

Wenju Xu¹, Chengjiang Long², Ruisheng Wang³, Guanghui Wang⁴,

¹OPPO US Research Center, InnoPeak Technology Inc
²JD Finance America Corporation
³University of Calgary
⁴Ryerson University
Artistic style transfer

Picasso

Picasso’s self-portrait
Artistic style transfer

The Evolution of Picasso’s self portrait

Age 18  Age 25  Age 90

Style: Surrealism

Art collection
Arbitrary style transfer

- Cannot benefit from other style images sharing similar style.
- Cannot well obtain style consistency and maintain content structure similarity.

[1] Arbitrary style transfer (Huang et al., 2017)
[2] Neural style transfer (Gatys et al., 2016)
Collection style transfer

- Recognize and transfer the dominant style clues;
- Lack the flexibility of exploring style manifold.

[1] Adaptive Style Transfer (Sanakoyeu et al., 2018)
Insights

• Handle arbitrary style transfer and collection style transfer in a unified model.
• Ensure style consistency and content structural similarity.

“style codes” is modeled as the dynamic parameters within dynamic modules.
Insights

- Handle arbitrary style transfer and collection style transfer in a unified model.
- Ensure style consistency and content structural similarity.

“style codes” is modeled as the dynamic parameters within dynamic modules.
Insights

- Handle arbitrary style transfer and collection style transfer in a unified model.
- Ensure style consistency and content structural similarity.

- “style codes” is modeled as the dynamic parameters within dynamic modules.
DRB-GAN: A Dynamic ResBlock Generative Adversarial Network for Artistic Style Transfer

- Three Components: style encoding network, style transfer network and discriminative network.
Style Encoding Network

- Style encoder: learnable CNN & pretrained VGG.
Style Encoding Network

- Style recalibration: refine the style code with the class attention.
Style transfer network

- Dynamic ResBlock: dynamic convolutional layer and AdaIN.
Style code

• “style code” in dynamic ResBlocks:

$$\{\theta^c_\omega, \theta^c_{\gamma,\beta}\} = \{H_\omega(s^c), H_{\gamma,\beta}(s^c)\}$$
Collection style code

- “collection style code” as a weighted mean of the “style codes”:

\[
\{\bar{\theta}_\omega^c, \bar{\theta}_{\gamma,\beta}^c\} = \left\{ \frac{1}{K} \sum_{k=0}^{K} \pi_k \theta_{\omega_k}^c, \frac{1}{K} \sum_{k=0}^{K} \pi_k \theta_{\gamma_k,\beta_k}^c \mid c \sim N \right\}
\]
Style transfer network

- SW-LIN Decoder: spatial window layer-instance normalization layer.
- Preserve local feature and remove artifacts in generated images.
Style transfer network

\[
SW-LIN(\gamma, \beta, \rho) = \gamma(\rho \phi^c_{sw} + (1 - \rho) \phi^l_{sw}) + \beta
\]

\[
\phi_{sw} = \frac{h - \mathbb{E}_{x_i \in sw}[h(x_i)]}{\sqrt{\text{Var}_{x_i \in sw}[h(x_i)]}}
\]

- SW-LIN Decoder: spatial window layer-instance normalization layer.
- Preserve local feature and remove artifacts in generated images.
Discriminative network:

- Objective function
  \[ \mathcal{L} = \mathcal{L}_{adv} + \lambda_{per}\mathcal{L}_{per} + \lambda_{cls}\mathcal{L}_{cls} \]
Comparison with other approaches

- Dataset
  - Content image: Place365 dataset
  - Style image: Wikiart dataset
- Metrics: Deception rate, inference time and human study.
- Model is trained on 768x768 and inferred on arbitrary resolution.

<table>
<thead>
<tr>
<th>Method</th>
<th>Time (sec)</th>
<th>GPU memory (MiB)</th>
<th>Model</th>
<th>Deception rate</th>
<th>Human studies Content score</th>
<th>Style score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wikiart test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gatys et al.</td>
<td>200</td>
<td>3887</td>
<td>Pspm</td>
<td>0.251</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>AdaIN</td>
<td>0.16</td>
<td>8872</td>
<td>Aspm</td>
<td>0.061</td>
<td>67.1%</td>
<td>0.127</td>
</tr>
<tr>
<td>WCT</td>
<td>5.22</td>
<td>10720</td>
<td>Aspm</td>
<td>0.023</td>
<td>43.6%</td>
<td>0.019</td>
</tr>
<tr>
<td>PatchBased</td>
<td>8.70</td>
<td>4159</td>
<td>Aspm</td>
<td>0.063</td>
<td>39.2%</td>
<td>0.013</td>
</tr>
<tr>
<td>Johnson</td>
<td>0.06</td>
<td>671</td>
<td>Aspm</td>
<td>0.080</td>
<td>53.4%</td>
<td>0.043</td>
</tr>
<tr>
<td>CycleGAN</td>
<td>0.07</td>
<td>1391</td>
<td>Pdpm</td>
<td>0.130</td>
<td>38.5%</td>
<td>0.021</td>
</tr>
<tr>
<td>AST</td>
<td>0.07</td>
<td>1043</td>
<td>Pdpm</td>
<td>0.450</td>
<td>43.2%</td>
<td>0.012</td>
</tr>
<tr>
<td>DRB-GAN</td>
<td>0.08</td>
<td>1324</td>
<td>Mdpm</td>
<td><strong>0.573</strong></td>
<td><strong>72.2%</strong></td>
<td><strong>0.453</strong></td>
</tr>
</tbody>
</table>
Comparison with other approaches

<table>
<thead>
<tr>
<th>Content</th>
<th>Style</th>
<th>CSD</th>
<th>AST</th>
<th>Gatys</th>
<th>CycleGAN</th>
<th>AdaIN</th>
<th>MetaNet</th>
<th>CST</th>
<th>Our</th>
</tr>
</thead>
</table>

- Our method: no artifacts in the regions and preserve the structural similarity.
Arbitrary style transfer

- Style consistency & Content structural similarity.
Collection style transfer

Table 2. Quantitative comparison of different methods. SD stands for style distance metric; DS represents deception score.

<table>
<thead>
<tr>
<th>Setting</th>
<th>Arbitrary Style (SD↓)</th>
<th>Collection style (DS↑)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>K=2</td>
</tr>
<tr>
<td>AdaIN</td>
<td>263.4</td>
<td>0.066</td>
</tr>
<tr>
<td>MetaNet</td>
<td>271.8</td>
<td>0.032</td>
</tr>
<tr>
<td>DRB-GAN</td>
<td><strong>241.2</strong></td>
<td><strong>0.576</strong></td>
</tr>
</tbody>
</table>

- The number of style images used to calculate the mean style code.

DRB-GAN: A Dynamic ResBlock Generative Adversarial Network for Artistic Style Transfer
Collection style transfer

Table 2. Quantitative comparison of different methods. SD stands for style distance metric; DS represents deception score.

<table>
<thead>
<tr>
<th>Setting</th>
<th>Arbitrary Style (SD↓)</th>
<th>Collection style (DS↑)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>K=2</td>
<td>5</td>
</tr>
<tr>
<td>AdaIN</td>
<td>263.4</td>
<td>0.066</td>
</tr>
<tr>
<td>MetaNet</td>
<td>271.8</td>
<td>0.032</td>
</tr>
<tr>
<td>DRB-GAN</td>
<td>241.2</td>
<td><strong>0.576</strong></td>
</tr>
</tbody>
</table>

- The number of style images used to calculate the mean style code.

DRB-GAN: A Dynamic ResBlock Generative Adversarial Network for Artistic Style Transfer
Collection style transfer

Table 2. Quantitative comparison of different methods. SD stands for style distance metric; DS represents deception score.

<table>
<thead>
<tr>
<th>Setting</th>
<th>Arbitrary Style (SD↓)</th>
<th>Collection style (DS↑)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>K=2</td>
</tr>
<tr>
<td>AdaIN</td>
<td>263.4</td>
<td>0.066</td>
</tr>
<tr>
<td>MetaNet</td>
<td>271.8</td>
<td>0.032</td>
</tr>
<tr>
<td>DRB-GAN</td>
<td>241.2</td>
<td>0.576</td>
</tr>
</tbody>
</table>

- The number of style images used to calculate the mean style code.
Collection style transfer

Collection style transfer

Content

AST

Ours

2021 ICCV VIRTUAL
Ablation study

- SW-LIN Decoder: preserve local feature and remove artifacts.
- \(\text{w/o } \mathcal{L}_{adv}\) : improve the style consistency.
- \(\text{w/o } \text{vgg}\) : capture the dominant style clues without subtle details.
- \(\text{w/o } \mathcal{L}_{cls}\) : causes slight degradation on stroke size variations.
Discriminative network

- Collection discriminator: improve style consistency.

[1] CST (Jan Svoboda, 2020)
Evaluation with unseen styles

- (c) (f) (i): arbitrary style transfer.
- (d) (g) (j): collection style transfer.
Evaluation with different resolutions

- Style consistency.
- Structural similarity.
HD Stylization

Content

Style

1024x2560

3072x7680

768x1920

2048x5120
Four-Way Style Interpolation

- Our model creates a smooth manifold structure.
Video Style Transfer @1920x1080

- All stylizations come from one trained model.
Conclusions

• A unified Model that handle arbitrary style transfer and collection style transfer.
• “style codes” is modeled as the dynamic parameters within Dynamic ResBlocks.
• Style consistency & Content structural similarity.
QR Code for our project:
https://github.com/xuwenju123/DRB-GAN

Thank you!