Time-of-Day Neural Style Transfer for Architectural Photographs

Yingshu Chen\textsuperscript{1} Tuan-Anh Vu\textsuperscript{1} Ka-Chun Shum\textsuperscript{1} Binh-Son Hua\textsuperscript{2} Sai-Kit Yeung\textsuperscript{1}

\textsuperscript{1}The Hong Kong University of Science and Technology \quad \textsuperscript{2}VinAI Research
Motivation

Architectural photography style transfer is challenging due to its special composition of dynamic sky and static foreground.

Generic neural style transfer and image-to-image translation treat the image as a single entity without knowing the foreground and background, leading to mismatched chrominance and destroyed geometric features of the original architecture.

Examples from
Motivation

• **Architectural photography style transfer** is challenging due to its special composition of dynamic sky and static foreground.

Motivation

• **Architectural photography style transfer** is challenging due to its special composition of dynamic sky and static foreground.

• Generic neural style transfer and image-to-image translation treat the image as a single entity without knowing the foreground and background:
  • Destroy geometric features of the original architecture.
  • Lead to mismatched chrominance.

Input images from pexels.com, 4147341 and pikwizard, 074a69d48e93c913aa718a929aea3b96.
Style images from unsplash.com, K4bvYKfXi3w and by Ed Lofdahl.
Motivation

• **Architectural photography style transfer** is challenging due to its special composition of dynamic sky and static foreground.

• Generic neural style transfer and image-to-image translation treat the image as a single entity without knowing the foreground and background:
  • Destroy geometric features of the original architecture.
  • Lead to mismatched chrominance.


Input images from pexels.com, 4147341 and pikwizard, 074a69d48e93c913aa718a929aea3b96.
Style images from unsplash.com, K4bvYKfXi3w and by Ed Lofdahl.
Problem Overview

• Architectural Photo Style Transfer:
  • Given an architectural photo and a style reference, we transfer styles of background and foreground separately while keeping foreground geometry intact.

Input images from unsplash.com, Ncmd8uLe8HQ and unsplash.com, 5omwAMDxmkJU.
Methodology

• Overview
  • Architectural style transfer framework with three modules: segmentation, image translation and blending optimization.

Input background / foreground

Input (source) $x_1 \in X_1$

Style (target) $x_2 \in X_2$

Segmentation

Image Translation

Translated background / foreground

Alpha blending

Blending Optimization

Output

Image Gradient Loss
Spatial Luminance KL Loss
Methodology

• Step 1 - Segmentation
  • Explicitly represent foreground and background of source and style images.
Methodology

• Segmentation
  • Disentangle foreground and background for style transfer.
  • Foreground contains architecture, street, etc.
  • Background contains sky.
  • Use pretrained model (training stage) or manual labeling.
Methodology

• Step 2 – Image Translation
  • Train foreground and background translation models with different training hyperparameters according to their style transfer features.
Methodology

• Step 2 – Image Translation
  • Train foreground and background translation models with different training hyperparameters according to their style transfer features.
Methodology

• Image Translation
  • Bidirectional image-to-image translation for unpaired data.
  • Reconstruction, cycle-consistency, adversarial losses.

Fail to preserve primal geometry

Input image from pikwizard, 074a69d48e93c913aa718a9e3b96.
Methodology

- Image Translation
  - Bidirectional image-to-image translation for unpaired data.
  - Reconstruction, cycle-consistency, adversarial losses.
  - High-frequency geometry preservation.

\[
\begin{align*}
  &\text{Style} \quad E_s^1 \quad \text{Generator} \quad G_1 \quad \text{Discriminator} \quad D_1 \\
  &\text{Content} \quad E_c^1 \quad z_1 \quad \text{Generator} \quad G_2 \quad \text{Discriminator} \quad D_2 \\
  &\text{Style} \quad E_s^2 \\
  &\text{Content} \quad E_c^2 \\

\end{align*}
\]

Input image from pikwizard, 074a69d48e93c913aa718a929aea3b96.
Methodology

- Image Translation
- Bidirectional image-to-image translation for unpaired data.
- Reconstruction, cycle-consistency, adversarial losses.
- High-frequency geometry preservation.

High-frequency geometry losses:
- Image Gradient loss:
  \[ \mathcal{L}_{gd} = \mathbb{E}_{x_1,x_2} [\| \nabla(Y(x_{1\rightarrow2}) - \nabla(Y(x_1)) \|_1] \]
- Spatial luminance KL loss:
  \[ \mathcal{L}_{kl} = \mathbb{E}_{x_1,x_2} [KL(Y(x_{1\rightarrow2})\|Y(x_1))] \]

*\(Y(\cdot)\) is luminance channel.
Methodology

- **Image Translation**
  - Bidirectional image-to-image translation for unpaired data.
- Reconstruction, cycle-consistency, adversarial losses.
- High-frequency geometry preservation.

**High-frequency geometry losses:**

- **Image Gradient loss:**
  \[ \mathcal{L}_{gd} = \mathbb{E}_{x_1, x_2} [||\nabla(Y(x_{1\rightarrow2}) - \nabla(Y(x_1))||_1] \]
- **Spatial luminance KL loss:**
  \[ \mathcal{L}_{kl} = \mathbb{E}_{x_1, x_2} [KL(Y(x_{1\rightarrow2}) || Y(x_1))] \]

*\( Y(\cdot) \) is luminance channel.*
Methodology

• Step 3 – Blending Optimization
  • With input high-resolution source geometry information, we optimize blended results with perfect gradient information.

Input (source) $x_1 \in X_1$

Input background / foreground

Segmentation

Style (target) $x_2 \in X_2$

Style background / foreground

Translated background / foreground

Blending Optimization

Image Translation

Output

High-res perfect geometry

Image Gradient Loss

Spatial Luminance KL Loss

Alpha blending

$X_1 \rightarrow X_2$
Methodology

• Blending Optimization
  • Restore high-fidelity gradient information of input content.
  • Optional: new background sky texture gradient.

Source image by David Spender.
Dataset

- Unpaired dataset from the Internet and time-lapse video frames.
- 21,000 architectural photos for training.
- 1,000 photos for evaluation.
- 4 labels for time-of-day styles: *day, golden, blue, night*, with diverse styles of architectures and sky.

Photos by Unsplash users *lisanto_12, bartmynameisbart, lanceanderson, christopher__burns*. 
Experiments

- Ablation study
  - Segmentation

<table>
<thead>
<tr>
<th></th>
<th>e-SSIM↑</th>
<th>Acc↑</th>
<th>IS↑</th>
<th>IoU↑</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ours-whole</td>
<td>0.6838</td>
<td>0.8282</td>
<td>2.5240</td>
<td>0.7410</td>
</tr>
<tr>
<td>Ours</td>
<td>0.6359</td>
<td>0.9486</td>
<td>2.7290</td>
<td>0.7257</td>
</tr>
</tbody>
</table>

Ours-whole: our translation model trained with whole images.

Input source and style reference

Ours without segmentation

Ours with segmentation

Input image from unsplash.com, Zr-ZnTNekEY. Style image from unsplash.com, sw_ePW1sYnU.

*e-SSIM: SSIM on image edges.
Experiments

• Ablation study
  • Geometry Losses

\[ \mathcal{L}_{kl} \]: spatial luminance KL loss.
\[ \mathcal{L}_{gd} \]: image gradient loss.
\[ \mathcal{L}_{total} \]: all losses.

<table>
<thead>
<tr>
<th></th>
<th>w/o ( \mathcal{L}<em>{kl} + \mathcal{L}</em>{gd} )</th>
<th>w/o ( \mathcal{L}_{kl} )</th>
<th>w/o ( \mathcal{L}_{gd} )</th>
<th>( \mathcal{L}_{total} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>e-SSIM(\uparrow)</td>
<td>0.4800</td>
<td>0.5539</td>
<td>0.5159</td>
<td>0.6359</td>
</tr>
<tr>
<td>Acc(\uparrow)</td>
<td>0.8934</td>
<td>0.9201</td>
<td>0.9265</td>
<td>0.9486</td>
</tr>
<tr>
<td>IS(\uparrow)</td>
<td>2.6858</td>
<td>2.7183</td>
<td>2.7214</td>
<td>2.7290</td>
</tr>
<tr>
<td>IoU(\uparrow)</td>
<td>0.6056</td>
<td>0.6536</td>
<td>0.6612</td>
<td>0.7257</td>
</tr>
</tbody>
</table>

Only taking foreground as an example. Original image from unsplash.3onN7CKCrH8.

*e-SSIM: SSIM on image edges.
Experiments

• Ablation study
  • Blending Optimization

Ours (or Ours-opt): our translation models trained with segmented images.

<table>
<thead>
<tr>
<th></th>
<th>e-SSIM↑</th>
<th>Acc↑</th>
<th>IS↑</th>
<th>IoU↑</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ours</td>
<td>0.6359</td>
<td>0.9486</td>
<td>2.7290</td>
<td>0.7257</td>
</tr>
<tr>
<td>Ours-opt</td>
<td>0.8094</td>
<td>0.9007</td>
<td>2.6127</td>
<td>0.7715</td>
</tr>
</tbody>
</table>

Input image from pikwizard.81dde04c1a0a2ac3f3682d680f2374bf. Style image from pexels.almudena-cathedral-madrid-423932.

*e-SSIM: SSIM on image edges
## Experiments

### image-to-image translation

<table>
<thead>
<tr>
<th></th>
<th>DRIT++</th>
<th>MUNIT</th>
<th>FUNIT</th>
<th>DSMAP</th>
<th>StarGANv2</th>
<th>AdaIN</th>
<th>SANet</th>
<th>AdaAttN</th>
<th>LST</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>e-SSIM↑</td>
<td>0.5214</td>
<td>0.5653</td>
<td>0.4959</td>
<td>0.4790</td>
<td>0.4778</td>
<td>0.4962</td>
<td>0.4854</td>
<td>0.5194</td>
<td>0.4903</td>
<td><strong>0.6359</strong></td>
</tr>
<tr>
<td>Acc↑</td>
<td>0.8903</td>
<td>0.8678</td>
<td>0.7714</td>
<td>0.9106</td>
<td>0.8788</td>
<td>0.7352</td>
<td>0.6193</td>
<td>0.6443</td>
<td>0.7071</td>
<td><strong>0.9486</strong></td>
</tr>
<tr>
<td>IS↑</td>
<td>2.6160</td>
<td>2.5916</td>
<td>2.5903</td>
<td>2.6580</td>
<td>2.6088</td>
<td>2.4082</td>
<td>2.1062</td>
<td>2.0928</td>
<td>1.7299</td>
<td><strong>2.7290</strong></td>
</tr>
<tr>
<td>IoU↑</td>
<td>0.6915</td>
<td><strong>0.7382</strong></td>
<td>0.5473</td>
<td>0.4975</td>
<td>0.4100</td>
<td>0.6642</td>
<td>0.7183</td>
<td>0.6532</td>
<td>0.6264</td>
<td>0.7257</td>
</tr>
</tbody>
</table>

### generic neural style transfer

#### Photorealism

<table>
<thead>
<tr>
<th></th>
<th>Real</th>
<th>Fake</th>
<th>Not Sure</th>
</tr>
</thead>
<tbody>
<tr>
<td>GT</td>
<td>89.47</td>
<td>64.91</td>
<td>52.63</td>
</tr>
<tr>
<td>Ours</td>
<td>73.68</td>
<td>82.46</td>
<td>63.22</td>
</tr>
<tr>
<td>DRIT++</td>
<td>12.28</td>
<td>5.75</td>
<td>3.45</td>
</tr>
<tr>
<td>MUNIT</td>
<td>14.04</td>
<td>6.72</td>
<td>3.45</td>
</tr>
<tr>
<td>FUNIT</td>
<td>5.26</td>
<td>5.75</td>
<td>3.45</td>
</tr>
<tr>
<td>DSMAP</td>
<td>10.53</td>
<td>47.36</td>
<td>3.45</td>
</tr>
<tr>
<td>StarGANv2</td>
<td>11.15</td>
<td>11.15</td>
<td>11.15</td>
</tr>
</tbody>
</table>

#### Structure & Style

<table>
<thead>
<tr>
<th></th>
<th>Ours vs DRIT++</th>
<th>Ours vs MUNIT</th>
<th>Ours vs FUNIT</th>
<th>Ours vs DSMAP</th>
<th>Ours vs StarGANv2</th>
<th>Ours vs AdaIN</th>
<th>Ours vs SANet</th>
<th>Ours vs AdaAttN</th>
<th>Ours vs LST</th>
</tr>
</thead>
<tbody>
<tr>
<td>GT</td>
<td>70.83</td>
<td>65.28</td>
<td>72.22</td>
<td>90.28</td>
<td>84.48</td>
<td>89.66</td>
<td>83.91</td>
<td>87.36</td>
<td>91.38</td>
</tr>
</tbody>
</table>

---

22
Experiments

Comparison to image-to-image translation methods

Input image from pikwizard, 074a69d48e93c913aa718a929ae3b96. Style images from pexels.com,buildings-under-cloudy-sky-during-sunset-462331, by Ed Lofdahl and pexels,almudena-cathedral-madrid-423932.
Experiments

Comparison to image-to-image translation methods

Experiments

Comparison to image-to-image translation methods

Experiments

Comparison to image-to-image translation methods

Experiments

Comparison to image-to-image translation methods

Style references

Ours-Opt

Experiments

Comparison to neural style transfer methods

Input

Style references

Input image from unsplash.com, Ncmd8uLe8H0.
Style images from unsplash.com, 5omwAMDxmU, unsplash.com, K4bvYkFxi3w, pexels.com, city-skyline-across-body-of-water-during-night-time-3586966/.
Experiments

Comparison to neural style transfer methods

Style images from unsplash.com, SomwAMDxmkU, unsplash.com, K4bvYKfXi3w, pexels.com, city-skyline-across-body-of-water-during-night-time-3586966/.
Experiments

Comparison to neural style transfer methods

Style images from unsplash.com, SomwAMDxmkU, unsplash.com, K4bvYKfX3w, pexels.com, city-skyline-across-body-of-water-during-night-time-3586966/.
Experiments

Comparison to neural style transfer methods

Style references

Style images from unsplash.com, SomwAMDxmkU, unsplash.com, K4bvYKfXi3w, pexels.com, city-skyline-across-body-of-water-during-night-time-3586966/.

Ours
Experiments

Comparison to neural style transfer methods

Style references

Ours-Opt

Style images from unsplash.com, SomwAMDxmkU, unsplash.com, K4bYKfXi3w, pexels.com, city-skyline-across-body-of-water-during-night-time-3586966/.
More Results

Input image from pexels.com, 4147341. Style image from unsplash.com, 5omwAMDxmkU.
More Results

Input image from pexels.com, 4147341. Style image from unsplash.com, K4bvYkFXi3w.
More Results

Contribution

1) A new problem setting for style transfer: **photorealistic style transfer for architectural photographs** of different times of day.

2) An image-to-image translation neural network with disentanglement representation that separately **considers style transfer for image foreground and background respectively**, accompanied with simple but effective **geometry losses** designed for image content preservation.

3) A **new dataset of architectural photographs** and an extensive benchmark for architectural style transfer.
Time-of-Day Neural Style Transfer for Architectural Photographs

Yingshu Chen¹  Tuan-Anh Vu¹  Ka-Chun Shum¹
Binh-Son Hua²  Sai-Kit Yeung¹
¹The Hong Kong University of Science and Technology
²VinAI Research