INTERNATIONAL CONFERENCE
ON COMPUTATIONAL PHOTOGRAPHY 2022
August 1-3, Caltech, Pasadena

Time-of-Day Neural Style Transfer for Architectural Photographs

Yingshu Chen[1] Tuan-Anh Vu[1] Ka-Chun Shum[1] Binh-Son Hua[2] Sai-Kit Yeung[1]

[1]The Hong Kong University of Science and Technology [2]VinAI Research

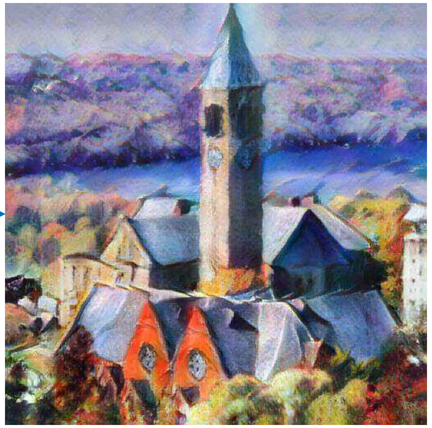
香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

VinAI

Motivation

Style Transfer

Artistic style transfer[1]



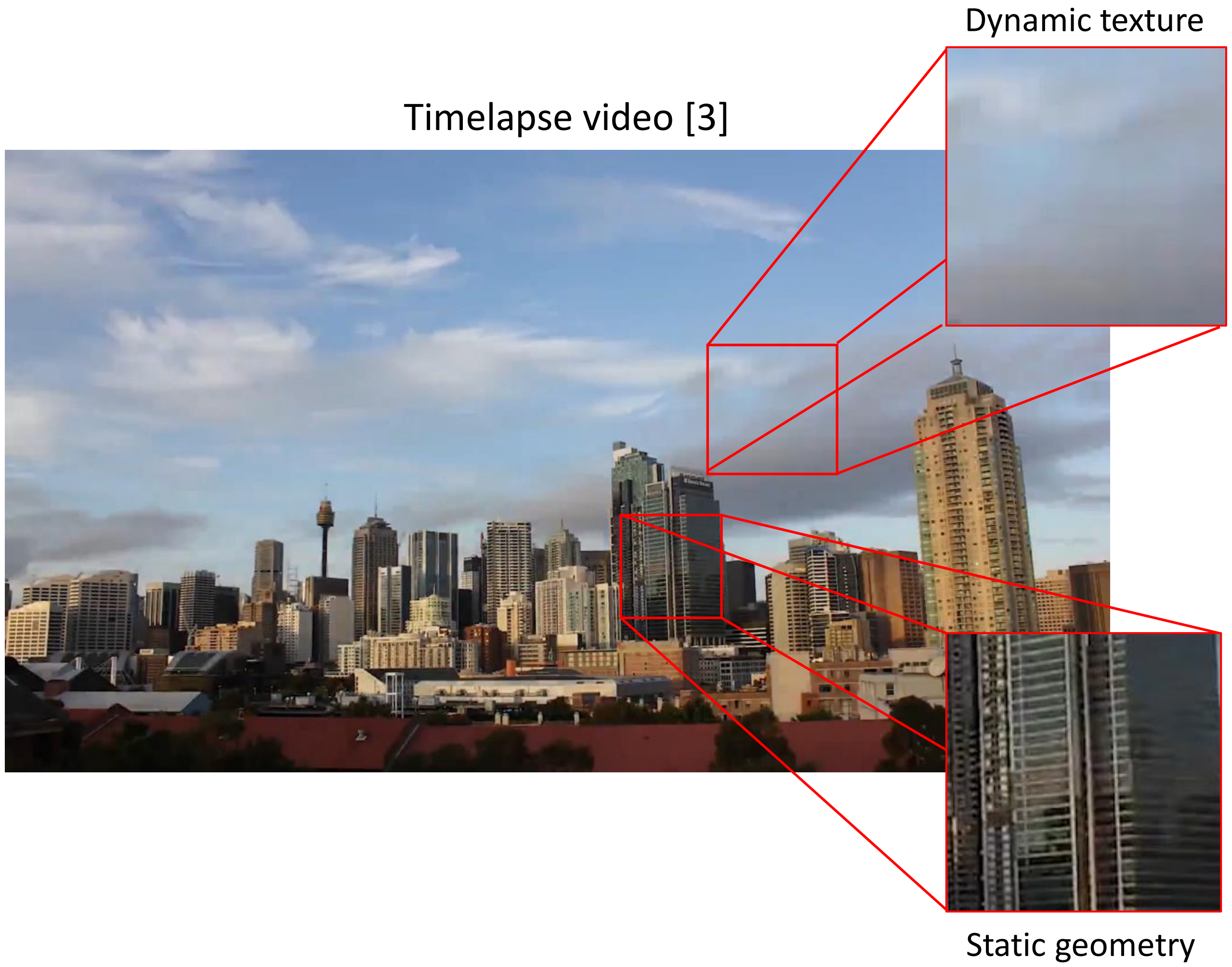
Domain translation[2]

Examples from
[1] Huang and Belongie, “Arbitrary style transfer in real-time with adaptive instance normalization”, ICCV 2017.
[2] Chang et al., “Domain-specific mappings for generative adversarial style transfer”, ECCV 2020.

Motivation

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

Timelapse video [3]



Video from
[1] Shih, et al. "Data-driven hallucination of different times of day from a single outdoor photo." Siggraph Asia 2013.

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

[2] Huang and Belongie, "Arbitrary style transfer in real-time with adaptive instance normalization", ICCV 2017.
[3] Chang et al., "Domain-specific mappings for generative adversarial style transfer", ECCV 2020.
Input images from pexels.com, 4147341 and pikwizard, 074a69d48e93c913aa718a929aea3b96.
Style images from unsplash.com, K4bvYKfXi3w and by *Ed Lofdahl* .

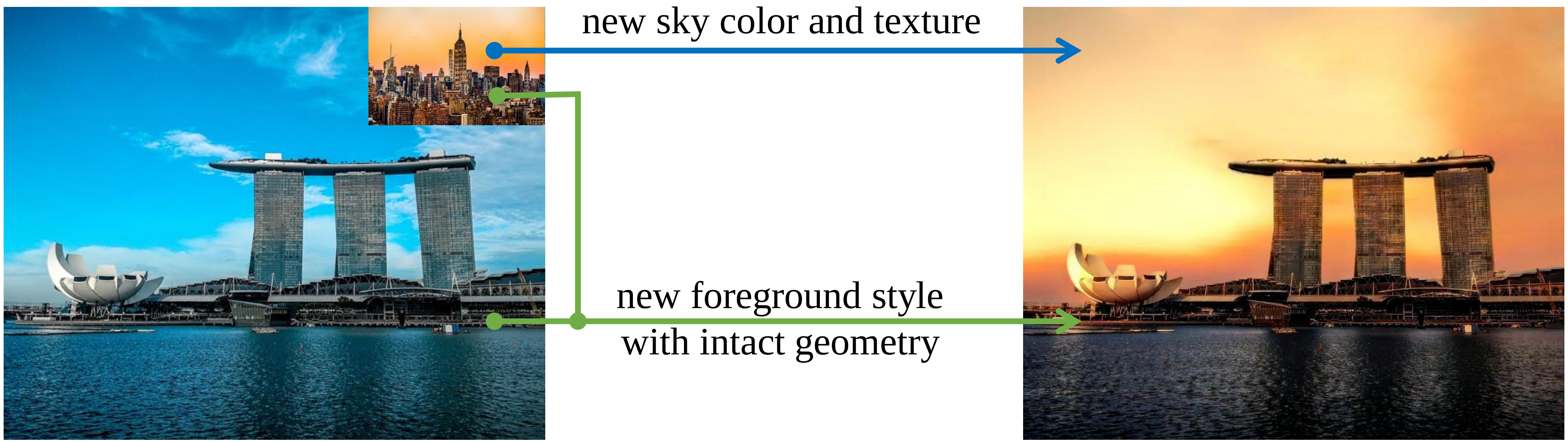
Motivation

- ***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

[1] Huang and Belongie, “Arbitrary style transfer in real-time with adaptive instance normalization”, ICCV 2017.
[2] Chang et al., “Domain-specific mappings for generative adversarial style transfer”, ECCV 2020.
Input images from pexels.com, 4147341 and pikwizard, 074a69d48e93c913aa718a929aea3b96.
Style images from unsplash.com, K4bvYKfXi3w and by *Ed Lofdahl* .

Problem Overview

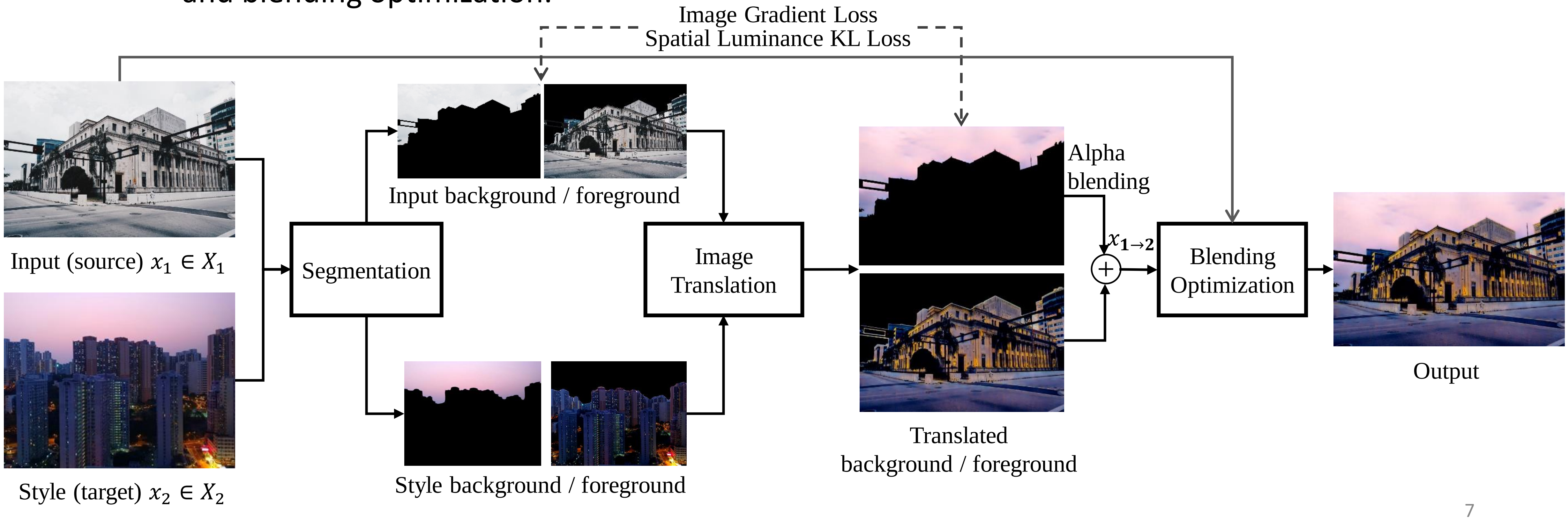
- Architectural Photo Style Transfer:
  - Given an architectural photo and a style reference, we transfer styles of background and foreground separately while keeping foreground geometry intact.



Input images from unsplash.com, Ncmd8uLe8H0 and unsplash.com, 5omwAMDxmkU.

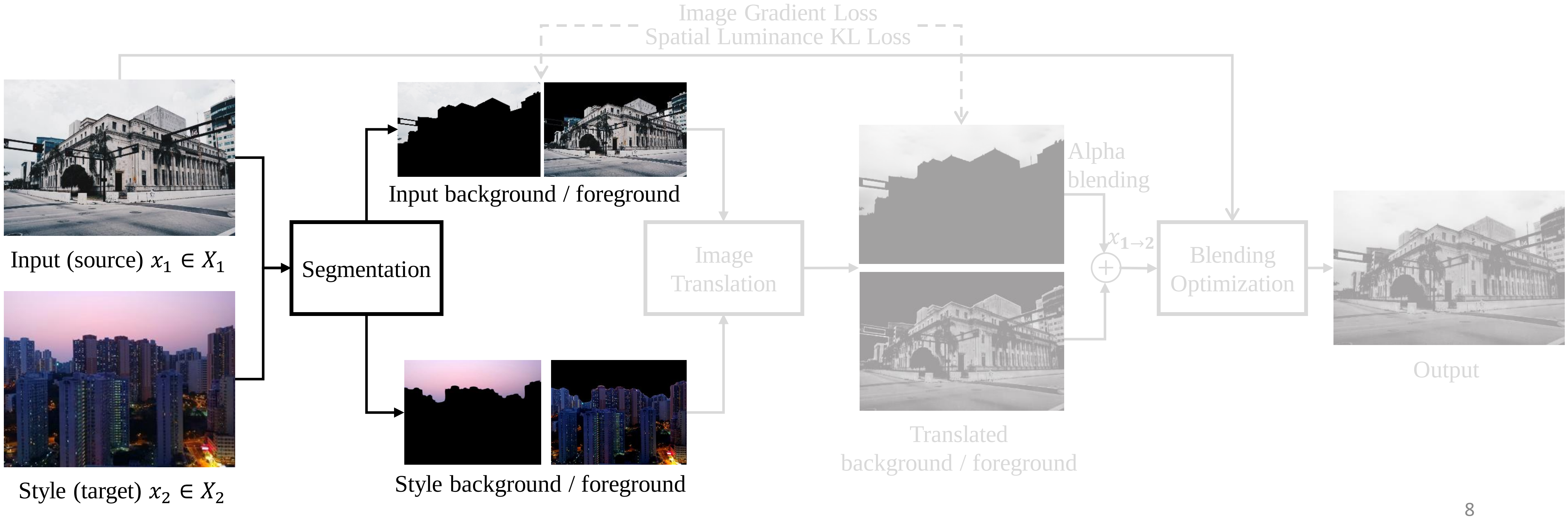
Methodology

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



Methodology

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



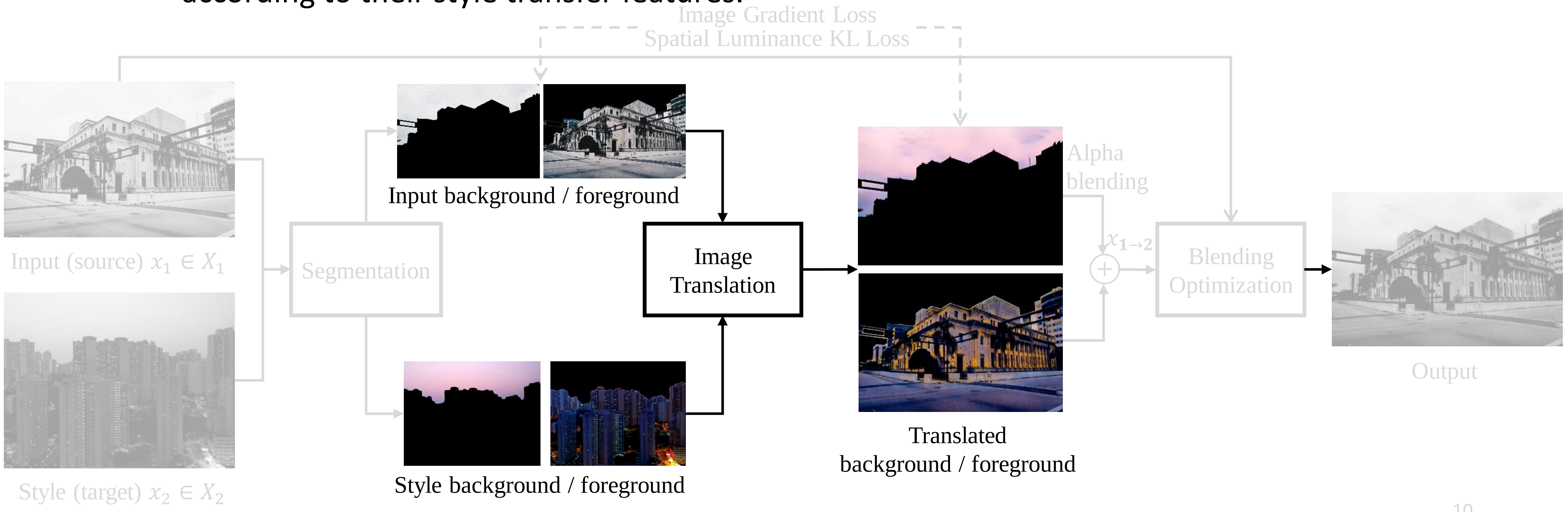
Methodology

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

Input image from pikwizard, 9cda9a1396248a477dce3fd81baad5e9.
Style image from pexels.com, city-buildings-1904605.

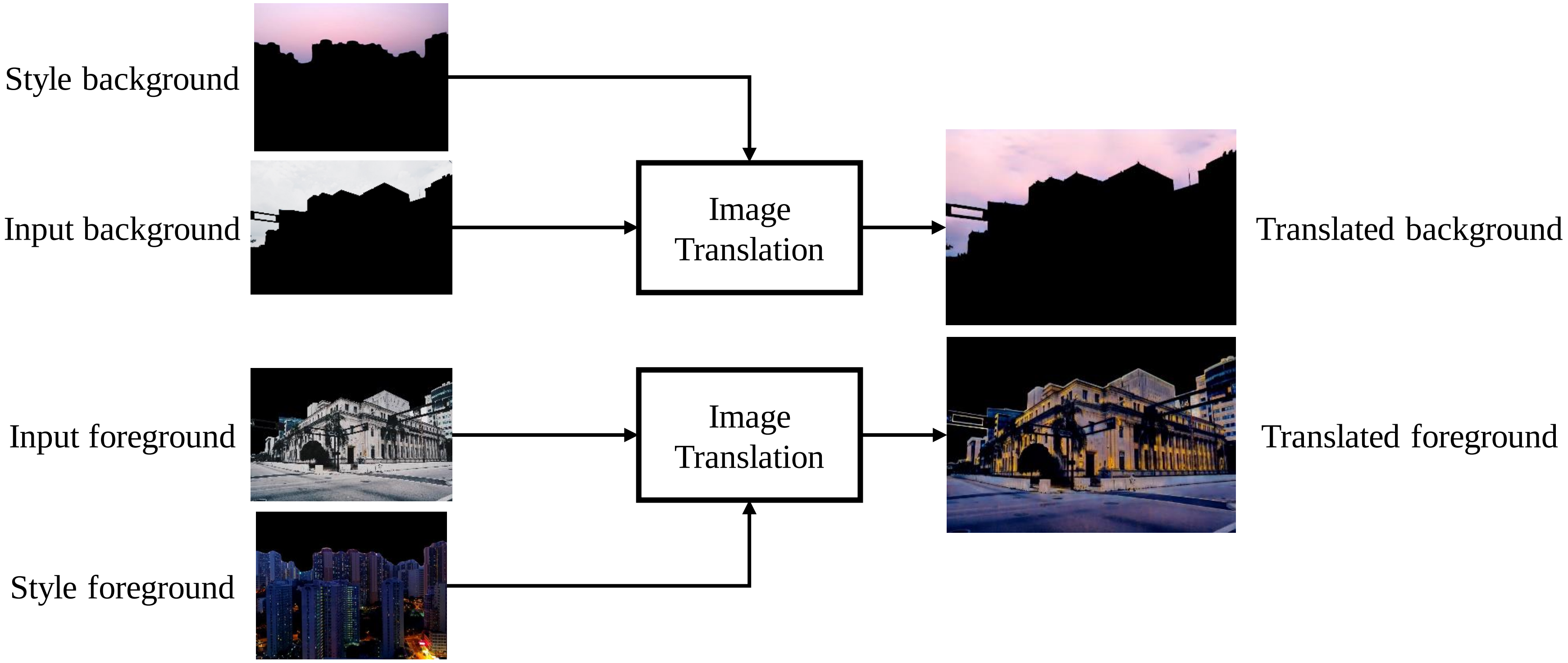
Methodology

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



Methodology

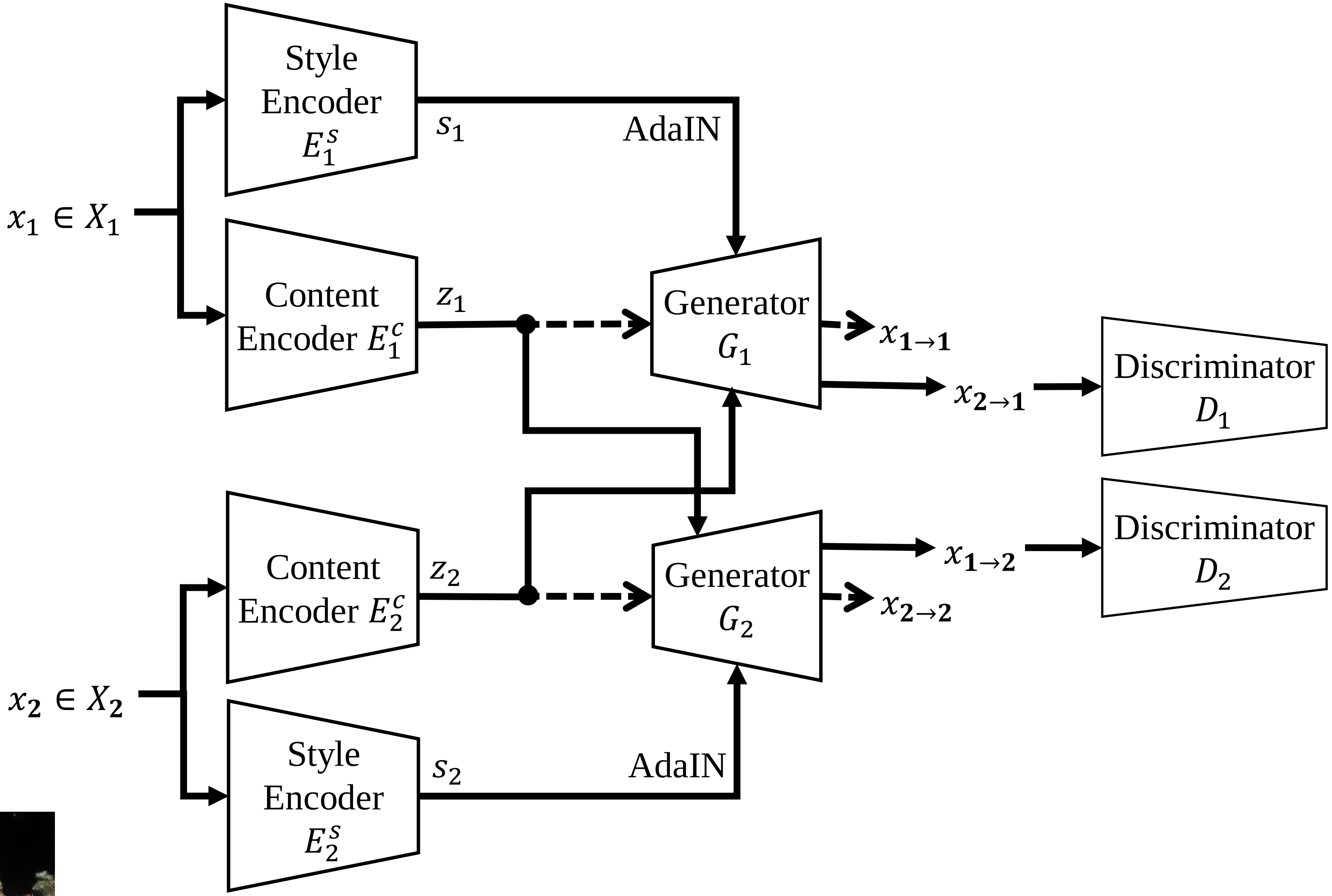
- Step 2 – Image Translation
  - Train foreground and background translation models with different training hyperparameters according to their style transfer features.



Methodology

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

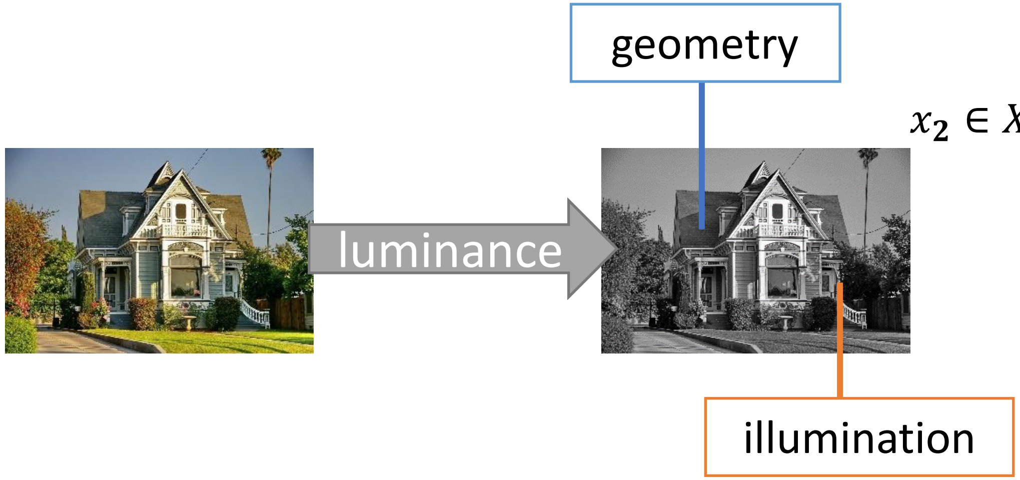
Fail to preserve primal geometry

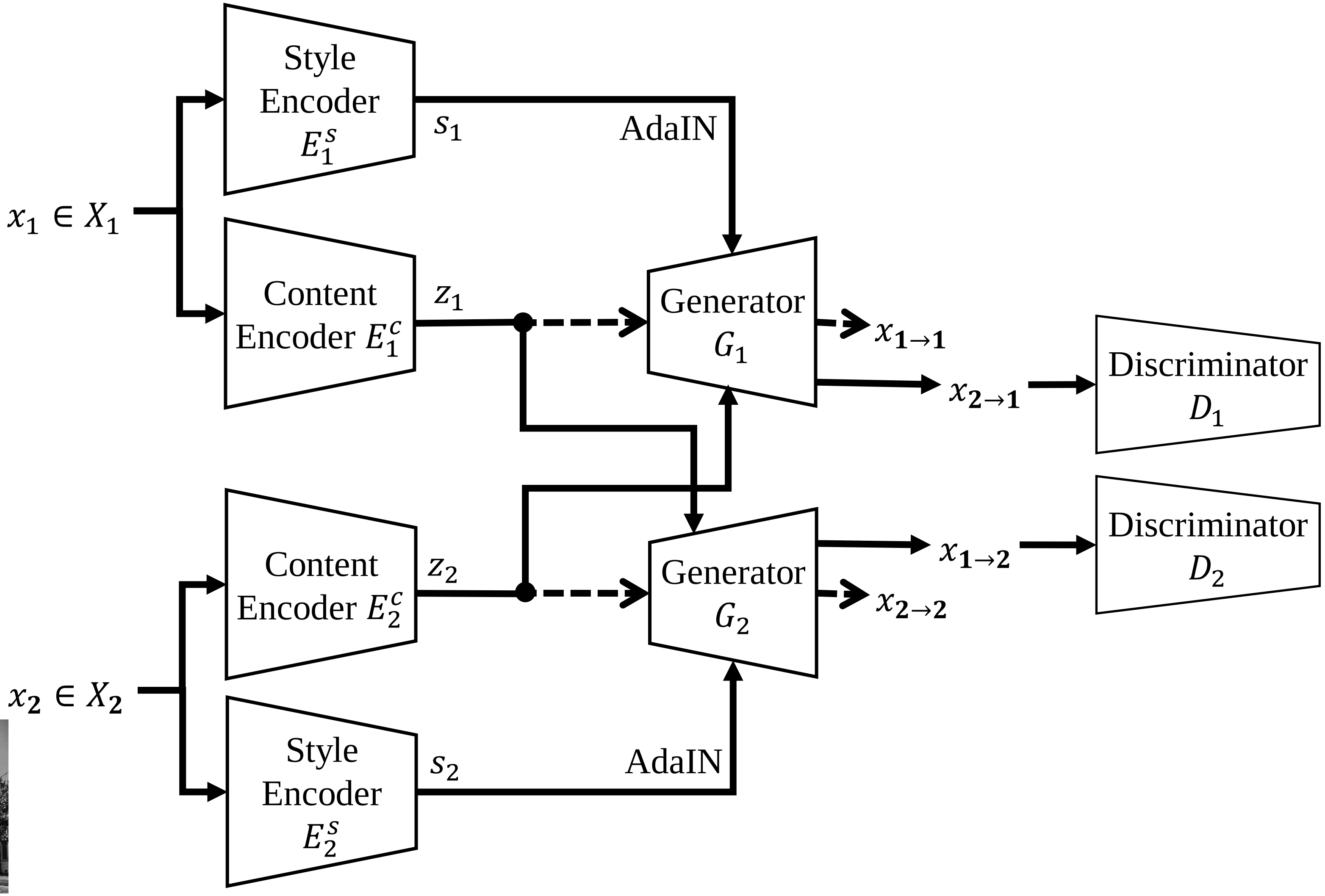


[2] Chang et al. "Domain-specific mappings for generative adversarial style transfer", ECCV 2020.
Input image from pikwizard, 074a69d48e93c913aa718a929aea3b96.

Methodology

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





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:

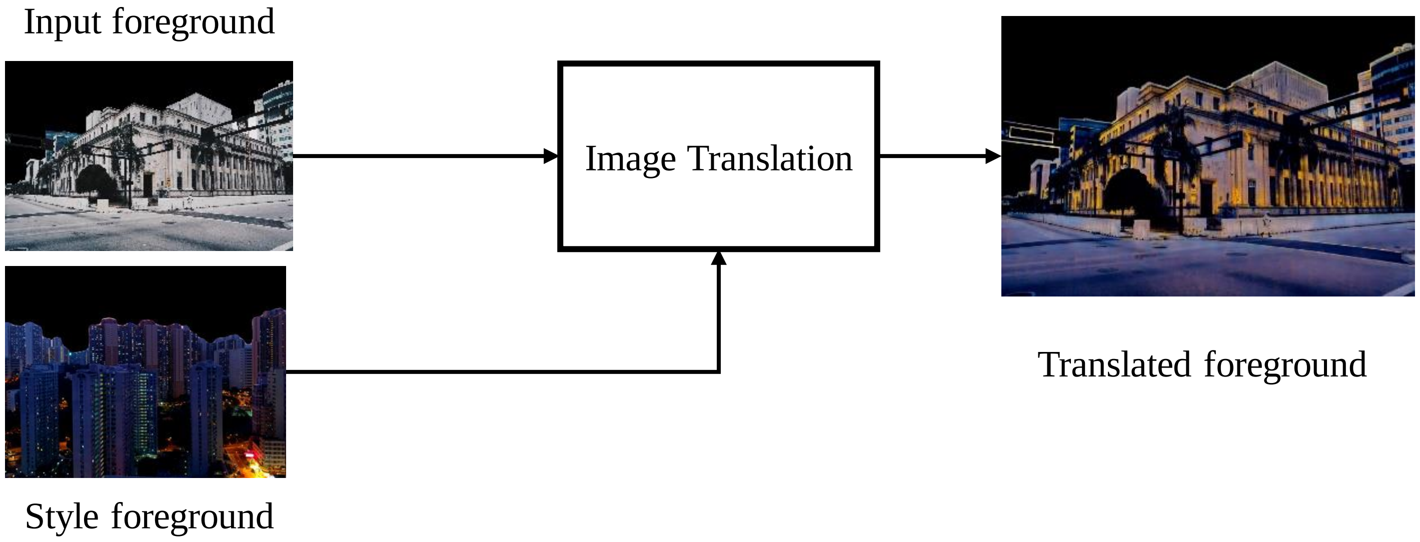
- Image Gradient loss:

$$\mathcal{L}_{gd} = \mathbb{E}_{x_1,x_2}[\|\nabla(Y(x_{1\to 2}) - \nabla(Y(x_1)\|_1]$$

- Spatial luminance KL loss:

$$\mathcal{L}_{kl} = \mathbb{E}_{x_1,x_2}[KL(Y(x_{1\to 2})\|Y(x_1))]$$

$*Y(\cdot)$ 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:

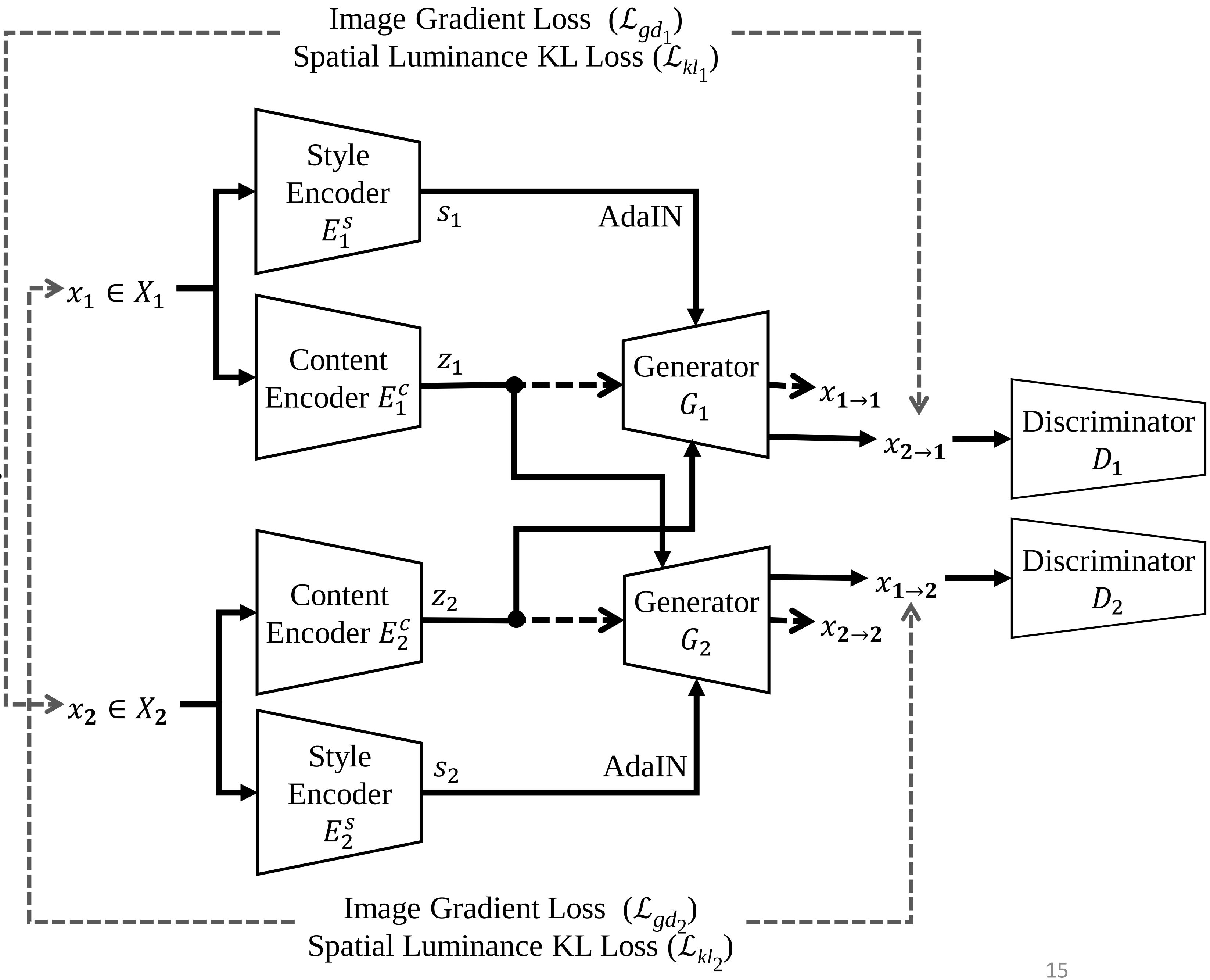
- Image Gradient loss:

$$\mathcal{L}_{gd} = \mathbb{E}_{x_1,x_2}[\|\nabla(Y(x_{1\to 2}) - \nabla(Y(x_1)\|_1]$$

- Spatial luminance KL loss:

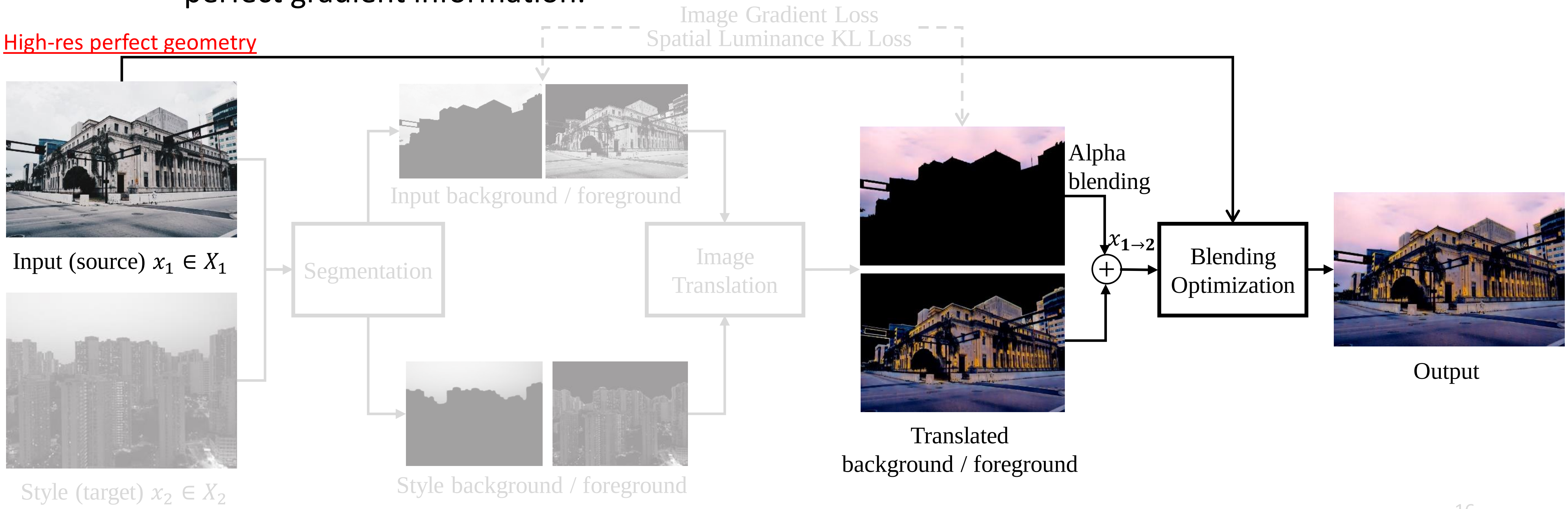
$$\mathcal{L}_{kl} = \mathbb{E}_{x_1,x_2}[KL(Y(x_{1\to 2})\|Y(x_2))]$$

**Y*(·) *is luminance channel.*



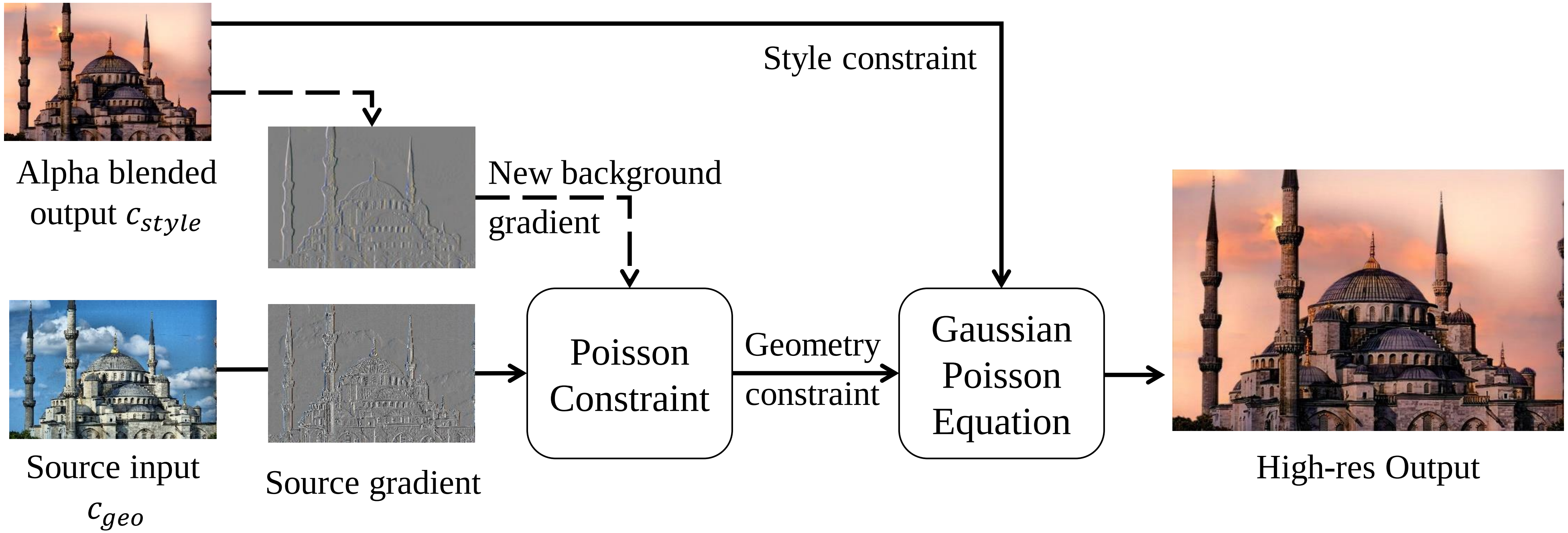
Methodology

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



Methodology

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



[4] Wu et al., "GP-GAN: Towards realistic high-resolution image blending," ACMMM 2019.
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.

day golden blue night

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

Experiments

- Ablation study
  - Segmentation

	e-SSIM↑	Acc↑	IS↑	IoU↑
Ours-whole	0.6838	0.8282	2.5240	0.7410
Ours	0.6359	0.9486	2.7290	0.7257

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.

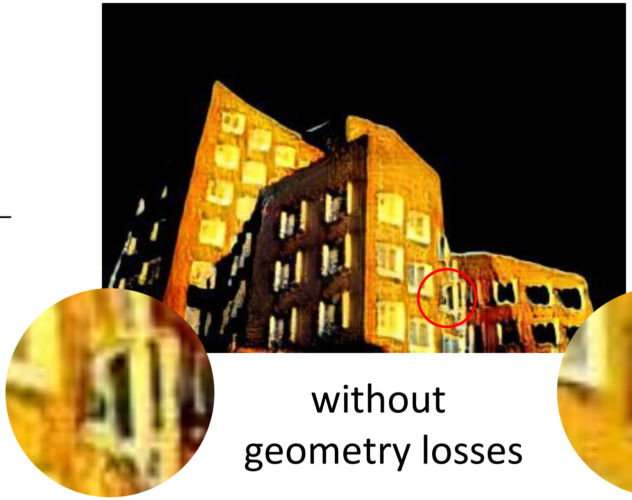
Experiments

- Ablation study
  - Geometry Losses

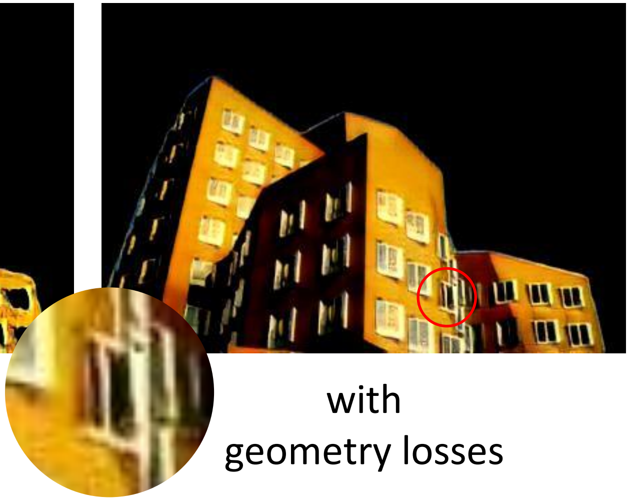
	w/o $\mathcal{L}_{kl} + \mathcal{L}_{gd}$	w/o $\mathcal{L}_{kl}$	w/o $\mathcal{L}_{gd}$	$\mathcal{L}_{total}$
e-SSIM↑	0.4800	0.5539	0.5159	**0.6359**
Acc↑	0.8934	0.9201	0.9265	**0.9486**
IS↑	2.6858	2.7183	2.7241	**2.7290**
IoU↑	0.6056	0.6536	0.6612	**0.7257**

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

Input source and style reference



without geometry losses



with geometry losses

Only taking foreground as an example. Original image from unsplash,3onN7CKCrH8.
*e-SSIM: SSIM on image edges.

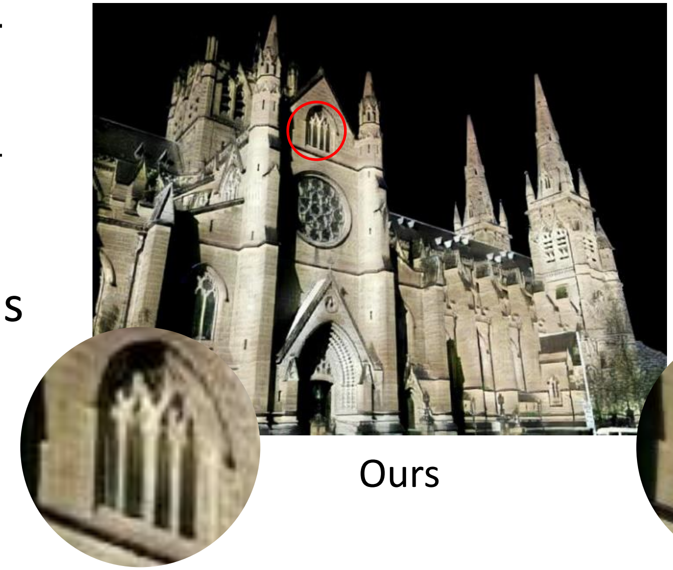
Experiments

- Ablation study
    - Blending Optimization

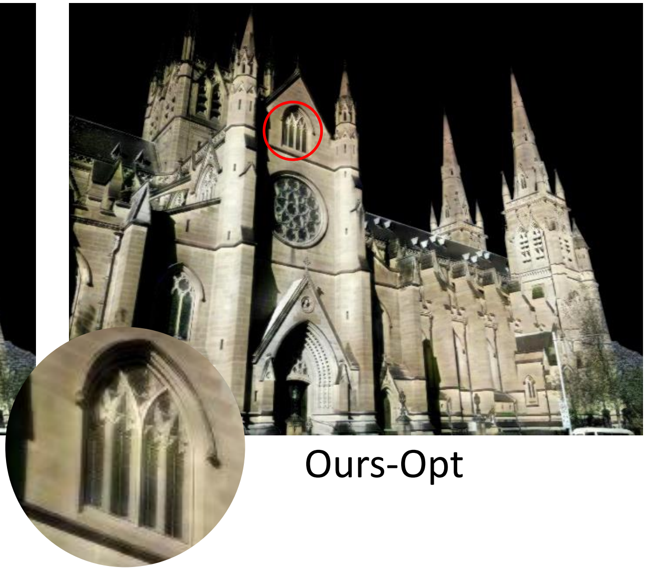
	e-SSIM↑	Acc↑	IS↑	IoU↑
Ours	0.6359	0.9486	2.7290	0.7257
Ours-opt	0.8094	0.9007	2.6127	0.7715

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

Input source and style reference



Ours



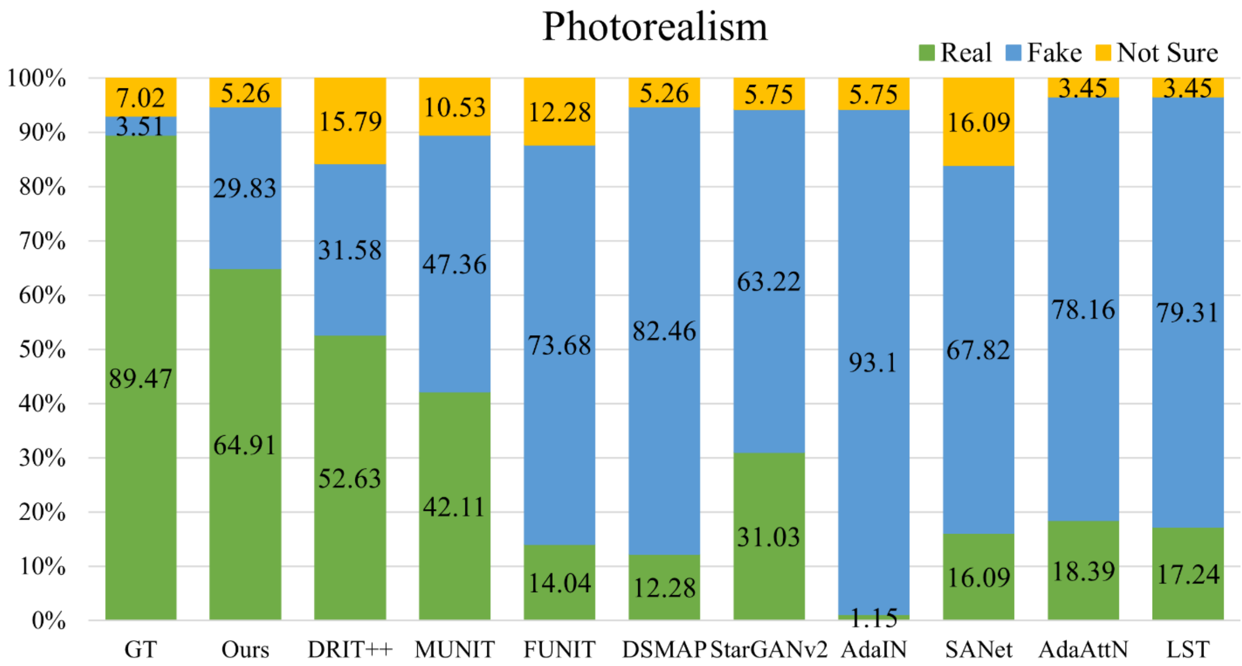
Ours-Opt

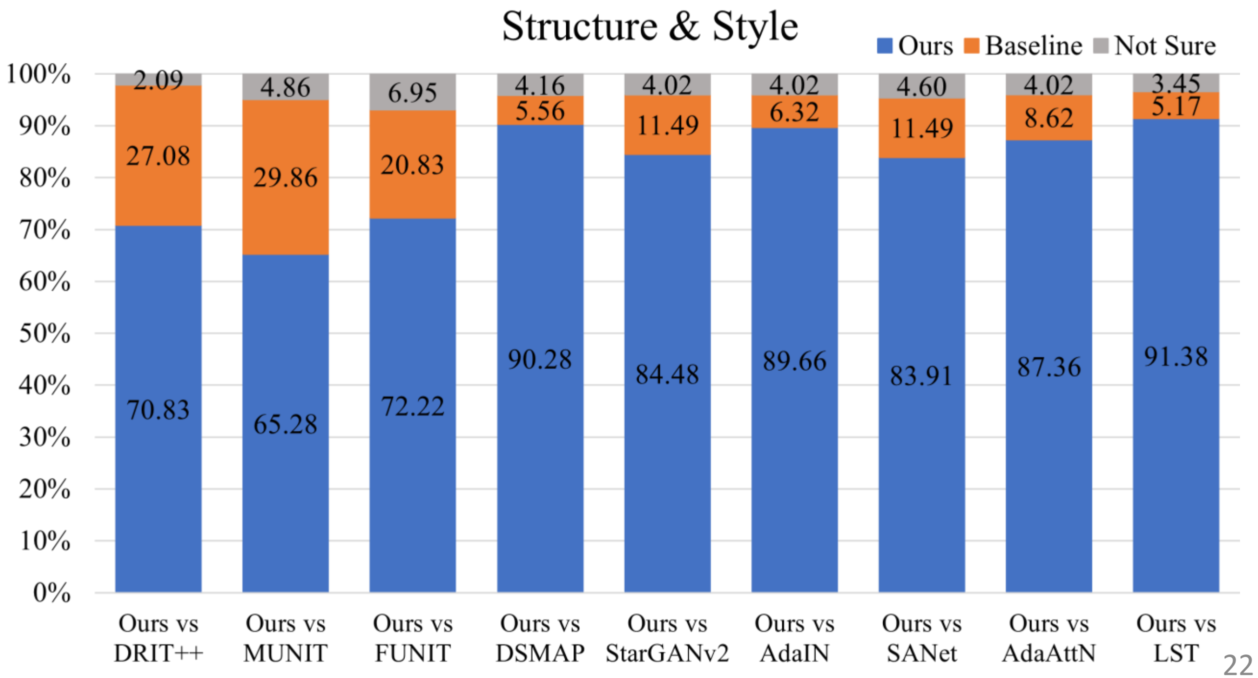
Input image from pikwizard,81dde04c1a0a2ac3f3682d680f2374bf. Style image from pexels,almudena-cathedral-madrid-423932.

*e-SSIM: SSIM on image edges

Experiments

	image-to-image translation					generic neural style transfer				
	DRIT++	MUNIT	FUNIT	DSMAP	StarGANv2	AdaIN	SANet	AdaAttN	LST	Ours
e-SSIM↑	0.5214	<u>0.5653</u>	0.4959	0.4790	0.4778	0.4962	0.4854	0.5194	0.4903	**0.6359**
Acc↑	0.8903	0.8678	0.77.14	<u>0.9106</u>	0.8788	0.7352	0.6193	0.6443	0.7071	**0.9486**
IS↑	2.6160	2.5916	2.5903	<u>2.6580</u>	2.6088	2.4082	2.1062	2.0928	1.7299	**2.7290**
IoU↑	0.6915	**0.7382**	0.5473	0.4975	0.4100	0.6642	0.7183	0.6532	0.6264	<u>0.7257</u>





Experiments

Comparison to image-to-image translation methods

Style references

Input

Input image from pikwizard, 074a69d48e93c913aa718a929aea3b96.
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

Style references

MUNIT

Huang et al., “Multimodal unsupervised image-to-image translation,” ECCV 2018.
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

Style references

DSMAP

Chang et al., "Domain-specific mappings for generative adversarial style transfer," ECCV 2020.
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

Style references

Ours

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

Style references

Ours-Opt

Style images from pexels.com,buildings-under-cloudy-sky-during-sunset-462331, by *Ed Lofdahl* and pexels,almudena-cathedral-madrid-423932.

Experiments

Comparison to neural style transfer methods

Style references

Input

Input image from unsplash.com, Ncmd8uLe8H0.
Style images from unsplash.com, 5omwAMDxmkU, unsplash.com, K4bvYKfXi3w, pexels.com, city-skyline-across-body-of-water-during-night-time-3586966/.

Experiments

Comparison to neural style transfer methods

Style references

AdaIN

Huang and Belongie, “Arbitrary style transfer in real-time with adaptive instance normalization,” ICCV 2017.
Style images from unsplash.com, 5omwAMDxmkU, unsplash.com, K4bvYKfXi3w, pexels.com, city-skyline-across-body-of-water-during-night-time-3586966/.

Experiments

Comparison to neural style transfer methods

Style references

AdaAttN

Liu et al., “AdaAttN: Revisit attention mechanism in arbitrary neural style transfer,” ICCV 2021.
Style images from unsplash.com, 5omwAMDxmkU, unsplash.com, K4bvYKfXi3w, pexels.com, city-skyline-across-body-of-water-during-night-time-3586966/.

Experiments

Comparison to neural style transfer methods

Style references

Ours

Style images from unsplash.com, 5omwAMDxmkU, unsplash.com, K4bvYKfXi3w, pexels.com, city-skyline-across-body-of-water-during-night-time-3586966/.

Experiments

Comparison to neural style transfer methods

Style references

Ours-Opt

Style images from unsplash.com, 5omwAMDxmkU, unsplash.com, K4bvYKfXi3w, pexels.com, city-skyline-across-body-of-water-during-night-time-3586966/.

More Results

Input

Style reference

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

More Results

Input

Style reference

Input image from pexels.com, 4147341. Style image from unsplash.com, K4bvYKfXi3w.

More Results

Input

Style reference

Input image from pexels.com, 4147341. Style image from pexels.com, city-skyline-across-body-of-water-during-night-time-3586966/.

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.

INTERNATIONAL CONFERENCE
ON COMPUTATIONAL PHOTOGRAPHY 2022
August 1-3, Caltech, Pasadena

Time-of-Day Neural Style Transfer for Architectural Photographs

Yingshu Chen[1] Tuan-Anh Vu[1] Ka-Chun Shum[1]
Binh-Son Hua[2] Sai-Kit Yeung[1]
[1]The Hong Kong University of Science and Technology
[2]VinAI Research

Project page