Modernizing Old Photos Using Multiple References via Photorealistic Style Transfer

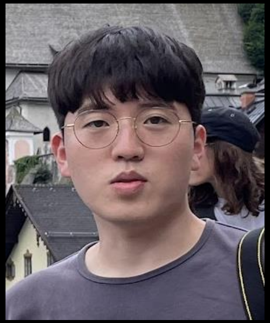
CVPR 2023 - WED-PM-011

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Agus Gunawan
KAIST

Soo Ye Kim
KAIST*
Adobe Research

Hyeonjun Sim
KAIST*



Jae-Ho Lee
ETRI

Munchurl Kim
KAIST

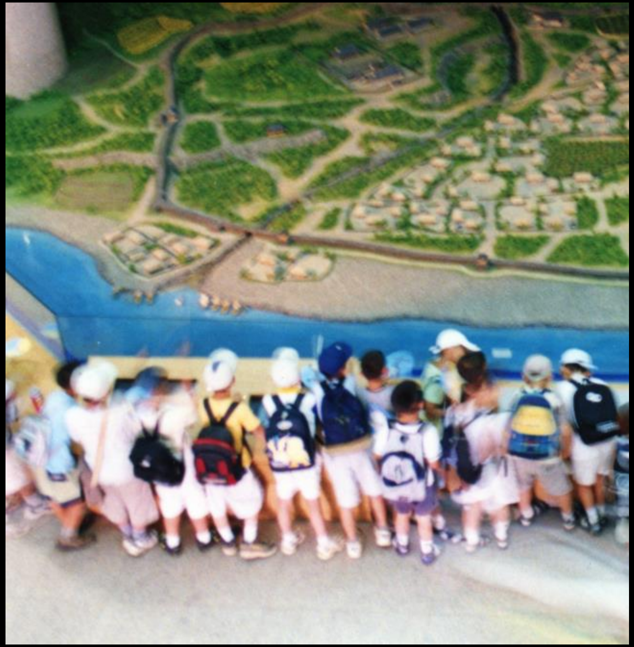
*some of the works are done while the author is in KAIST

Old Photos Characteristics

Structured degradation

Scratch, hole, …

Unstructured degradation



Blur, noise, …

Unique artifact

Sepia color, color fading, …

Previous Work's Solution

Old photo

OPR, Wan *et al*, CVPR 2020

Restoration

Mixed degradation restoration

Modernization

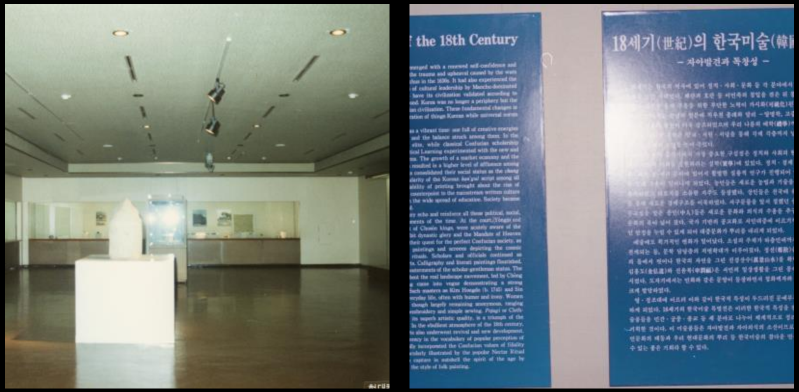
Implicit enhancement e.g., color

learned from prior of modern photos data

Output

Our Motivation

Old photos



OPR, Wan *et al*, CVPR 2020

implicit enhancement

Outputs

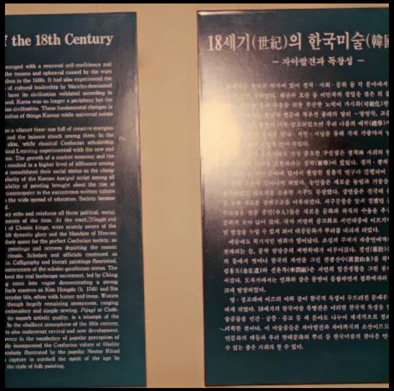
Overall look and style remain like old photos input or even worse

Our Motivation

Old photos


OPR, Wan *et al*, CVPR 2020

implicit enhancement

Outputs

Overall look and style remain like old photos input or even worse

Reference 1

Old photo

Reference 2

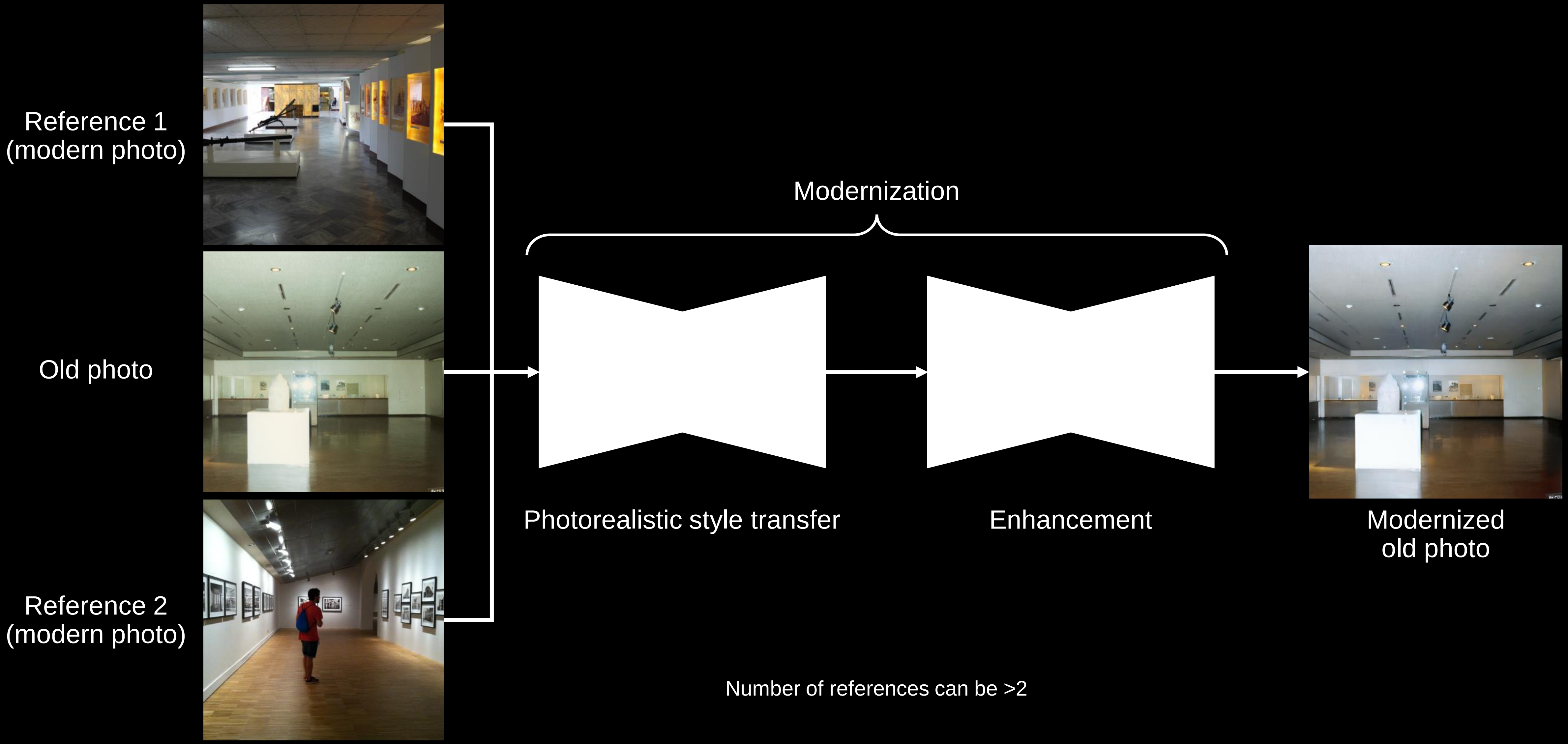
Ours

style transfer + enhancement

Output

Changing the style and enhancing the photo to look modern

Overview of Our Solution

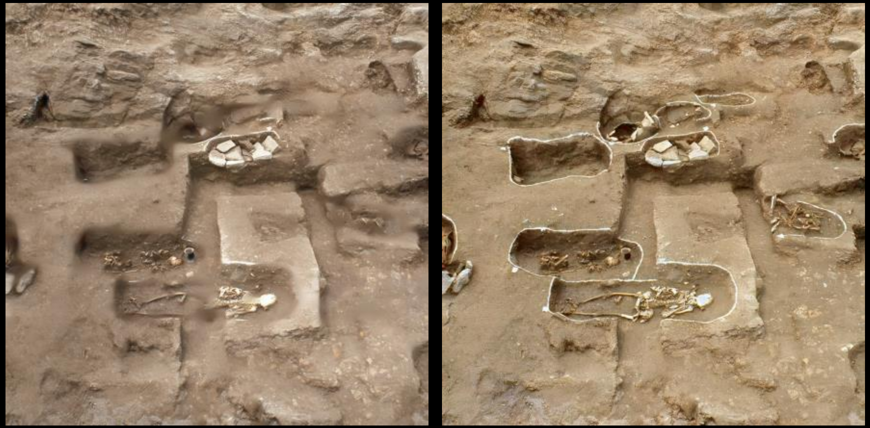


Number of references can be >2

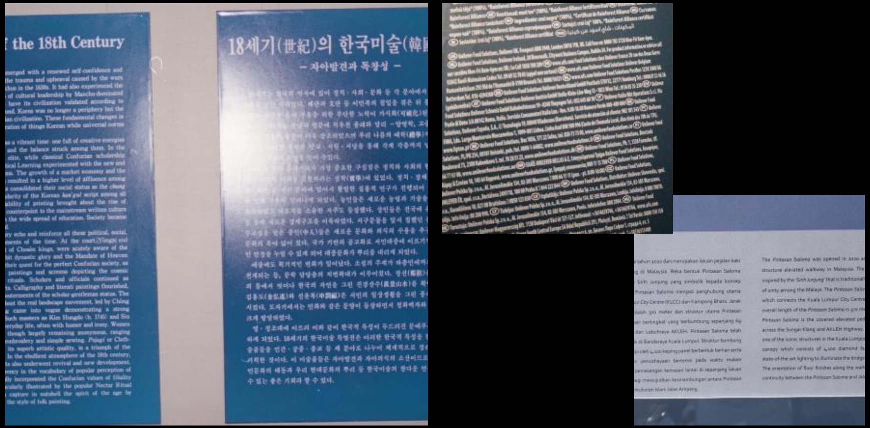
Overview of The Results

[42] Wan *et al*, CVPR 2020

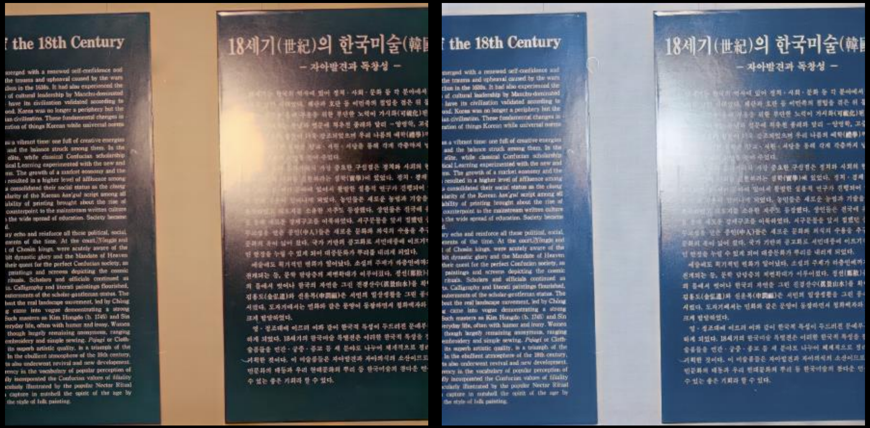
Old photo | References



OPR [42] | Ours



Old photo | References



OPR [42] | Ours

Better results even without using any old photos during training

Overview of The Results

[42] Wan *et al*, CVPR 2020


Old photo


References


OPR [42]


Ours


Old photo

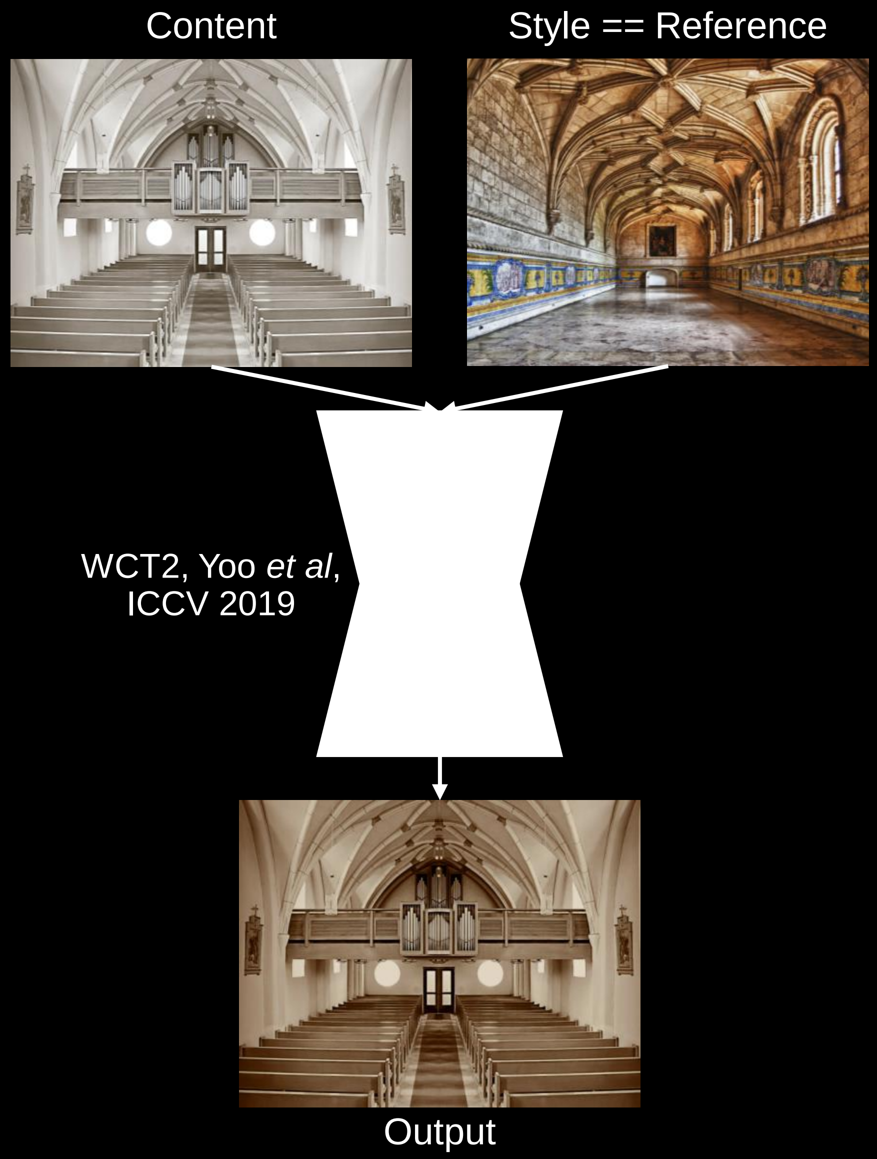

References

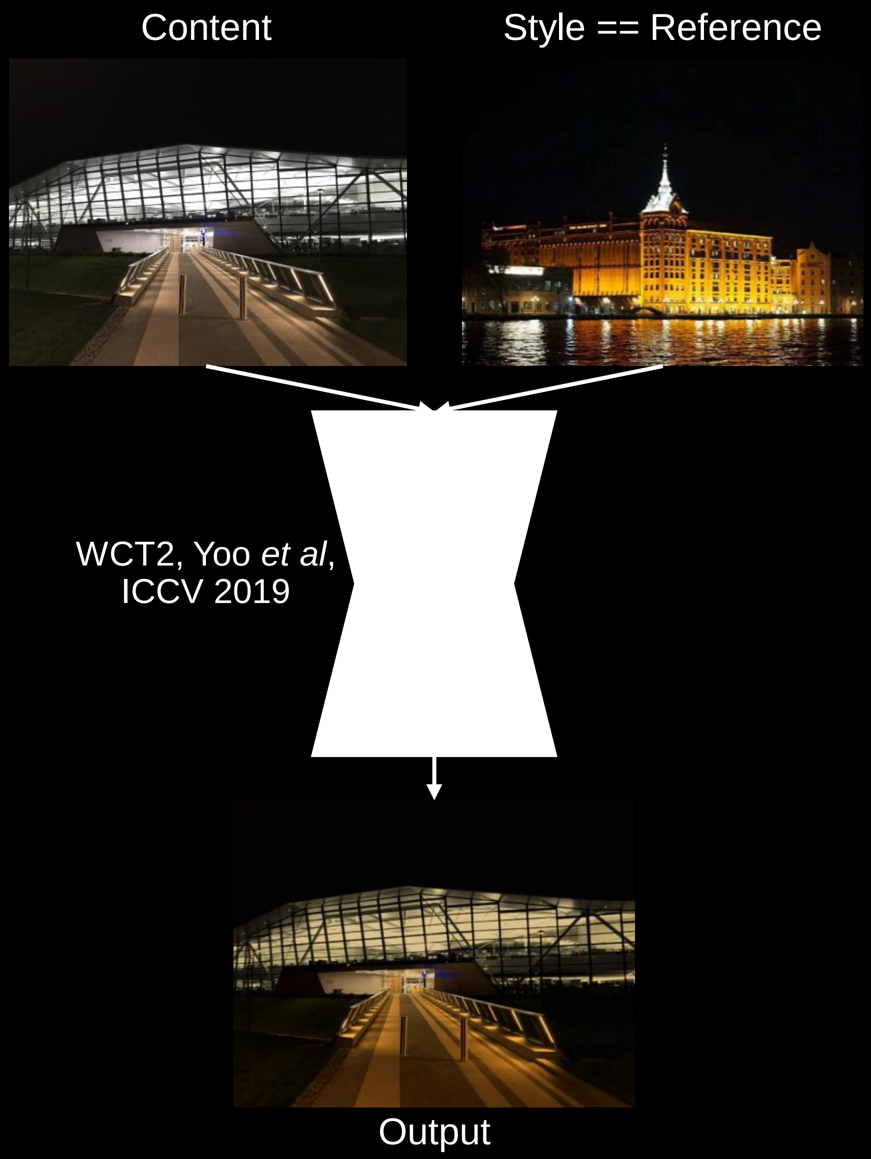

OPR [42]


Ours

Better results even without using any old photos during training

Inspiration from Photorealistic Style Transfer (PST)





Universal ability to perform style transfer for any photos without retraining to predefined styles

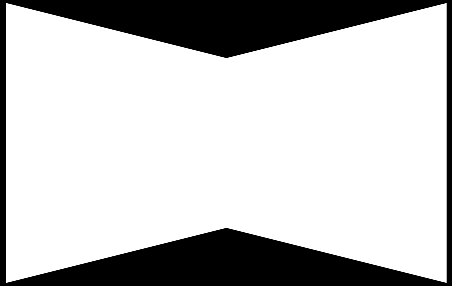
However, PST …

Problem 1 – Is only able to use a **single reference**

Old photo

No sky

Reference



PST
PhotoWCT2, Chiu *et al*,
WACV 2022

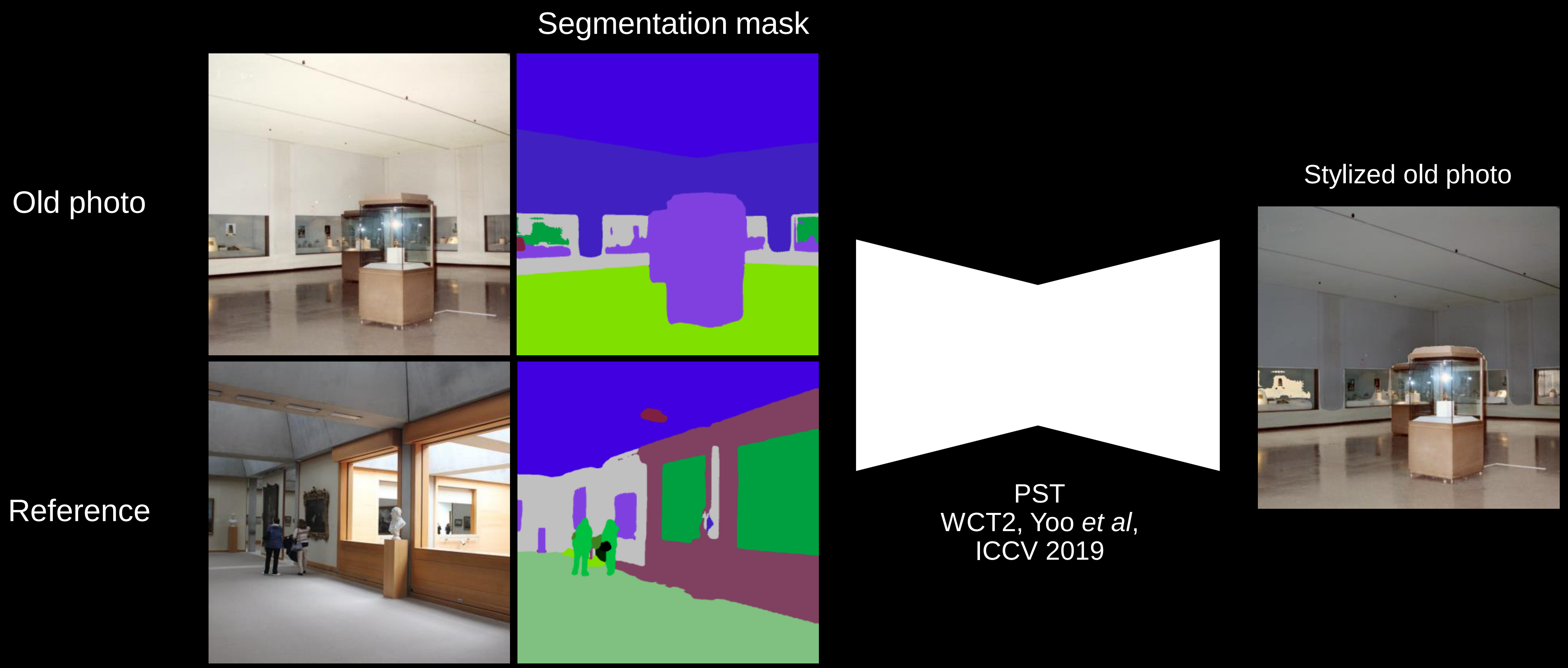
Stylized old photo

Unnatural result

- ✓ A single reference cannot match the whole semantics of a natural old photo e.g., no sky
- ✓ No corresponding region cause incorrect style transfer e.g., green sky

However, PST …

Problem 2 – Require **accurate segmentation mask** for **local style transfer**



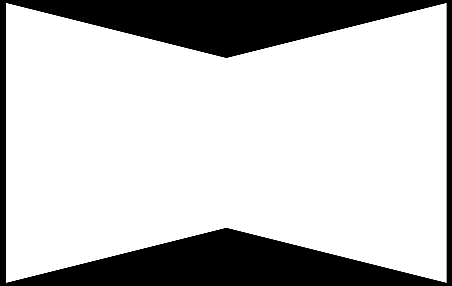
- ✓ Segmentation mask is obtained by using ViT-Adapter [Chen *et al*, ICLR 2023]
- ✓ Unnatural PST results due to unreliable segmentation mask for old photo

However, PST …

Problem 3 – Can produce **unnatural global style transfer** and **cannot perform enhancement** e.g., sharpening

Old photo

Reference



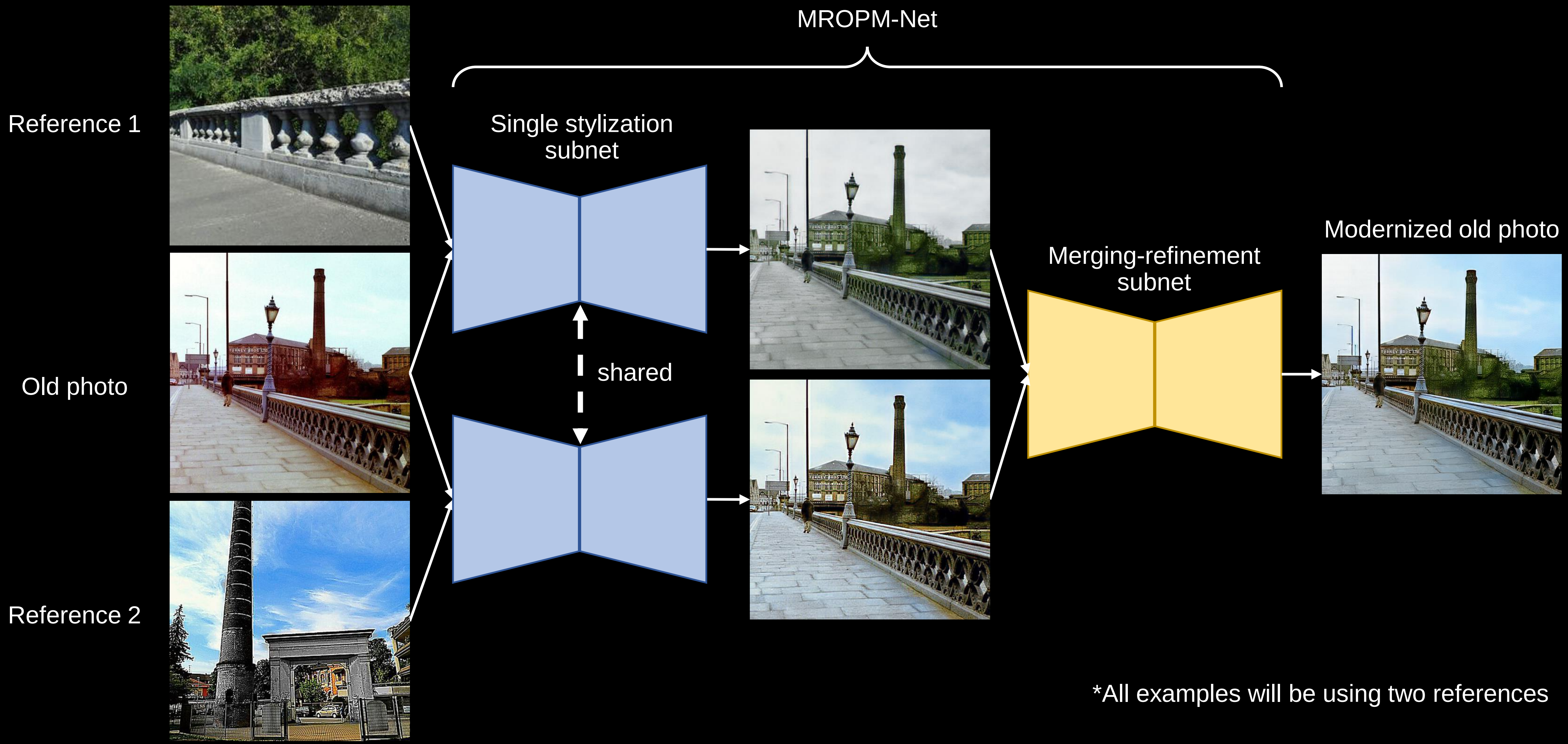
PST
WCT2, Yoo *et al*,
ICCV 2019

Stylized old photo

- ✓ PST produces an unnatural global style transfer result even though the reference is related
- ✓ PST cannot perform enhancement as can be seen by blurry and noisy stylization result

Proposed Method – MROPM

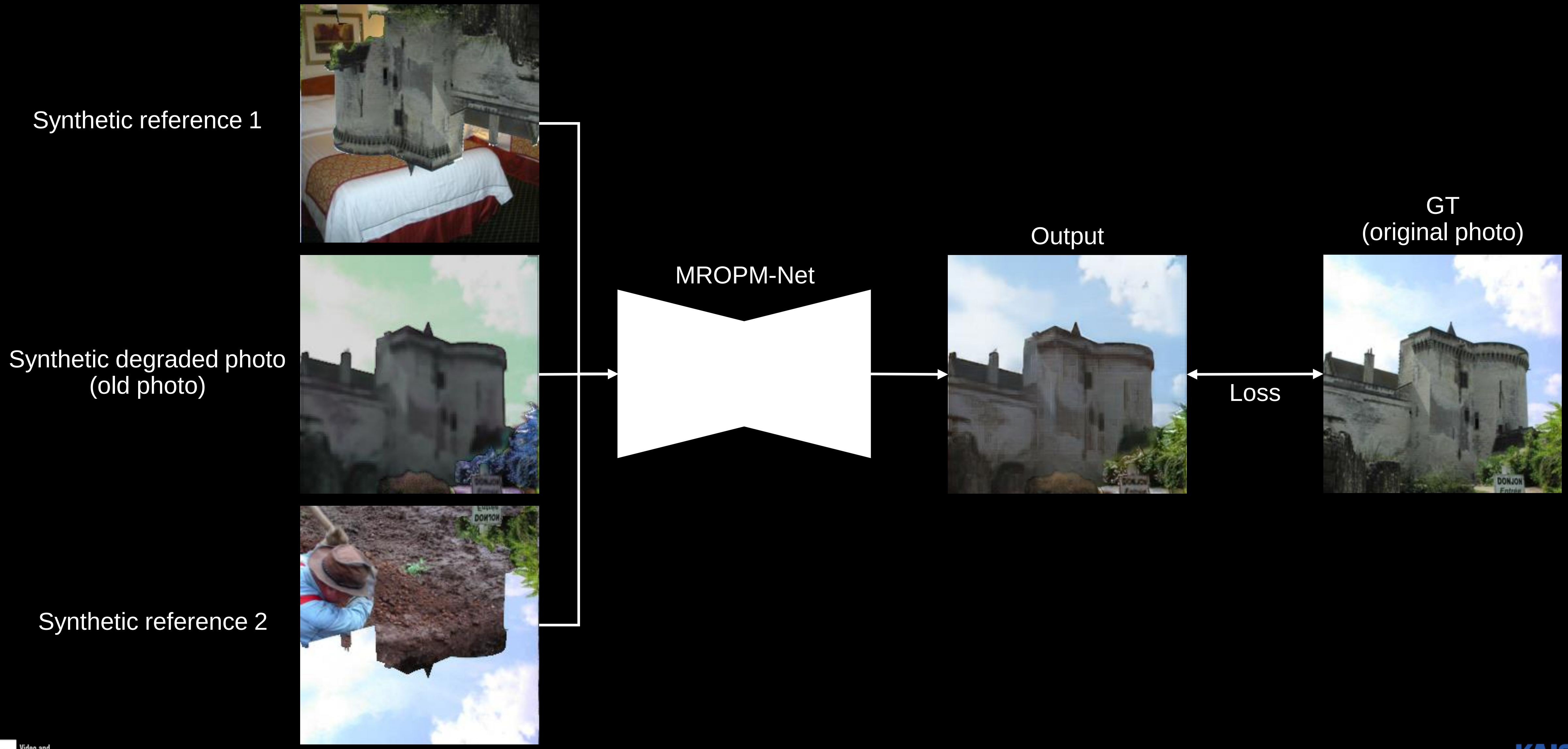
Part 1 – MROPM-Net



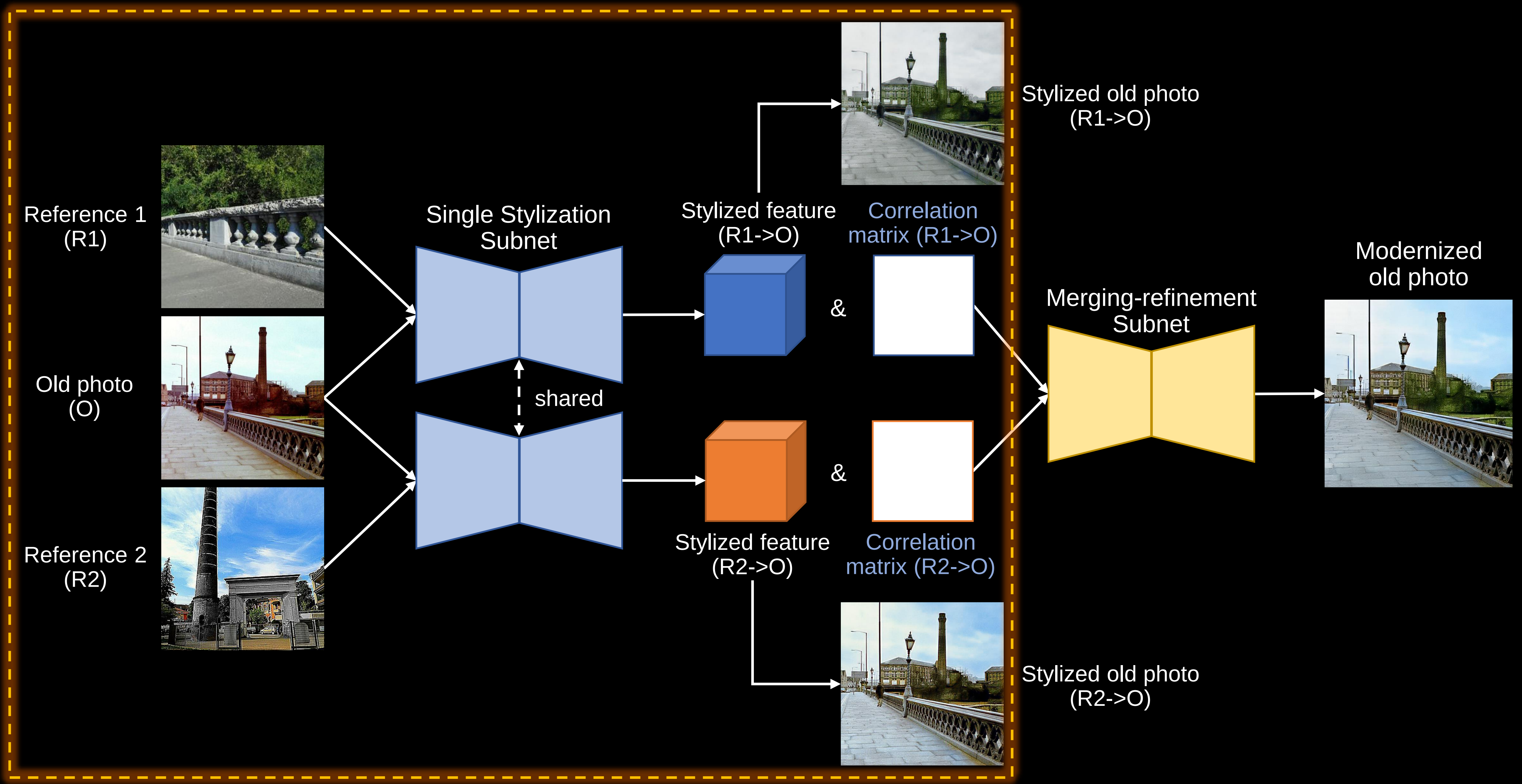
*All examples will be using two references

Proposed Method – MROPM

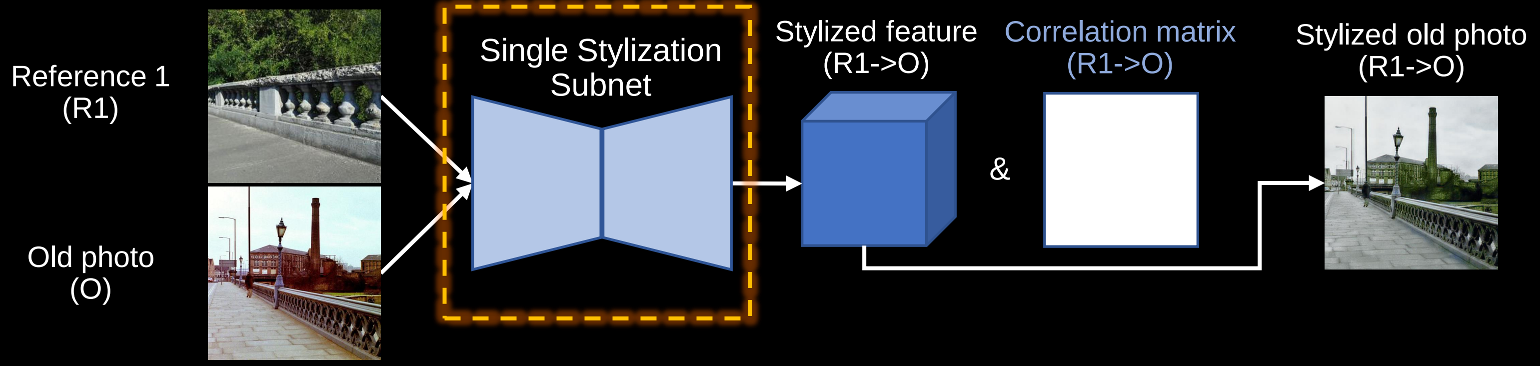
Part 2 – Novel training strategy



Proposed Method – MROPM-Net

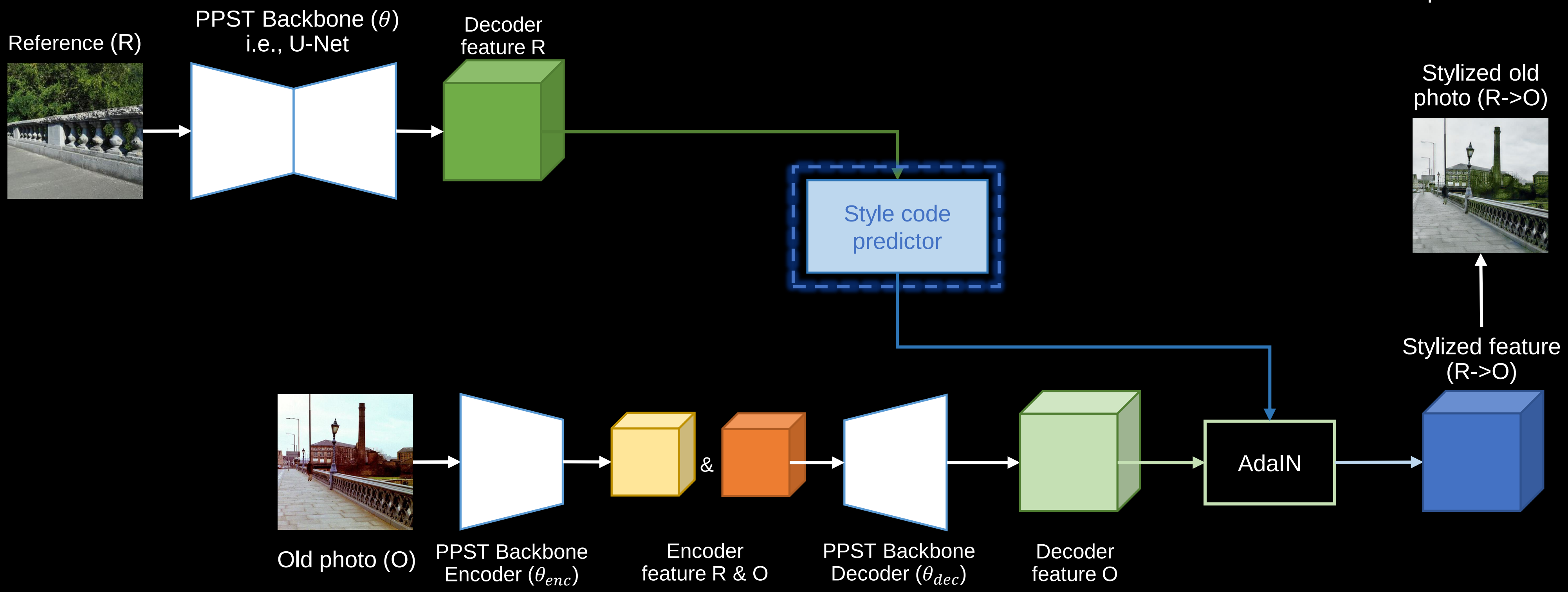


Proposed Method – MROPM-Net



Single stylization subnet details

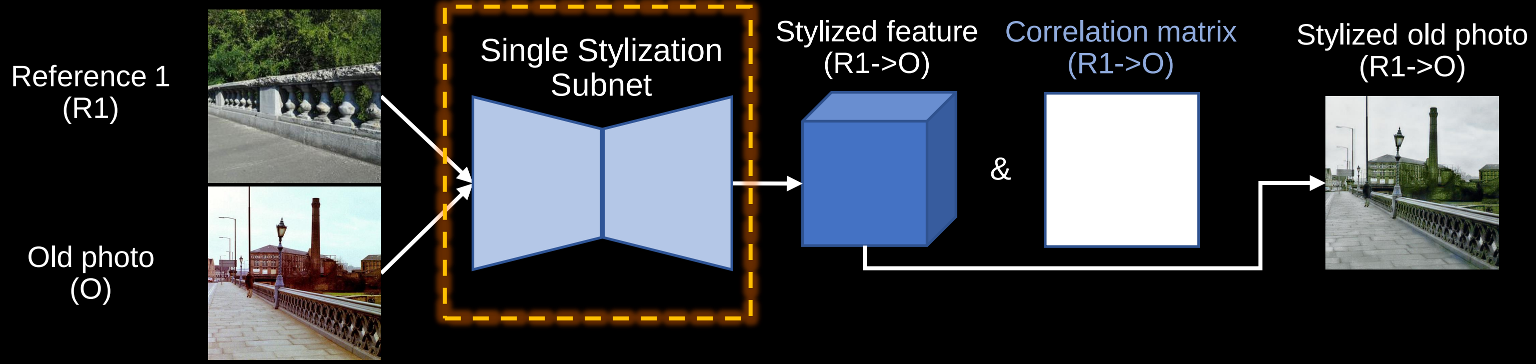
style code: feature's mean and std
same color: same weight parameters
PPST: pretrained PST



VICLab Video and Image Computing Lab

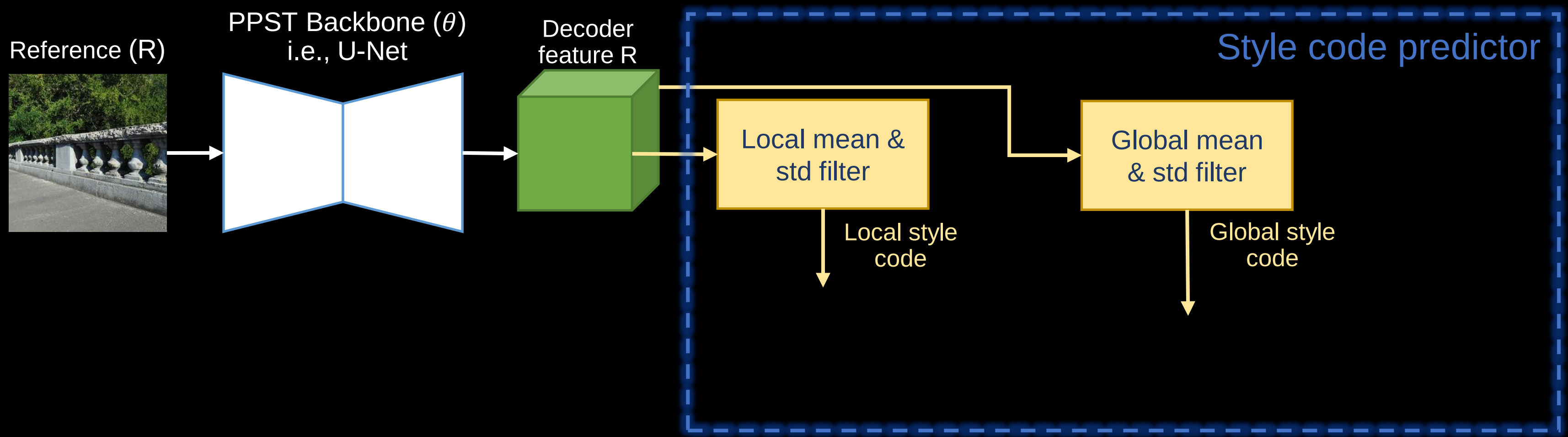
KAIST

Proposed Method – MROPM-Net

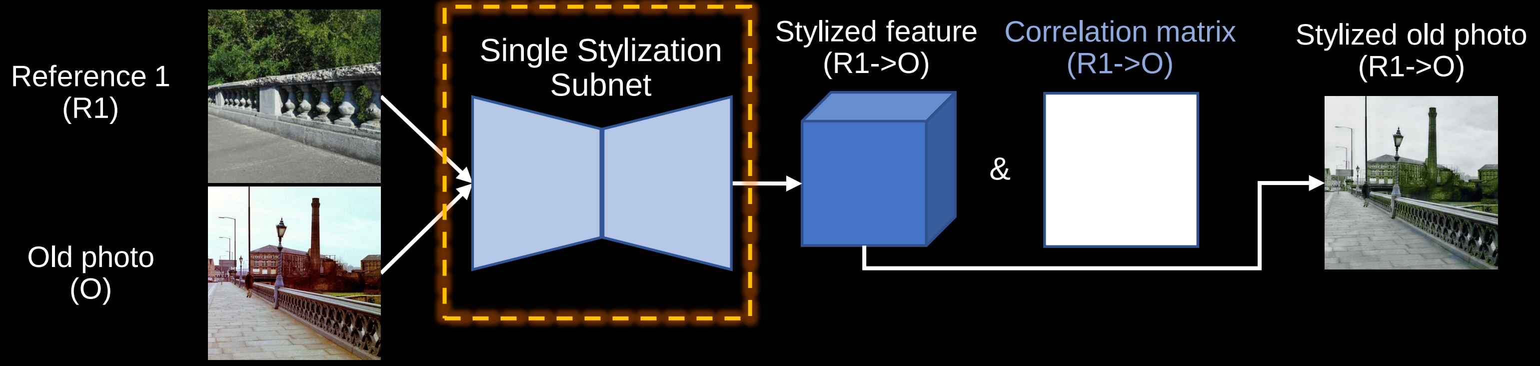


Single stylization subnet details

style code: feature's mean and std
same color: same weight parameters
PPST: pretrained PST

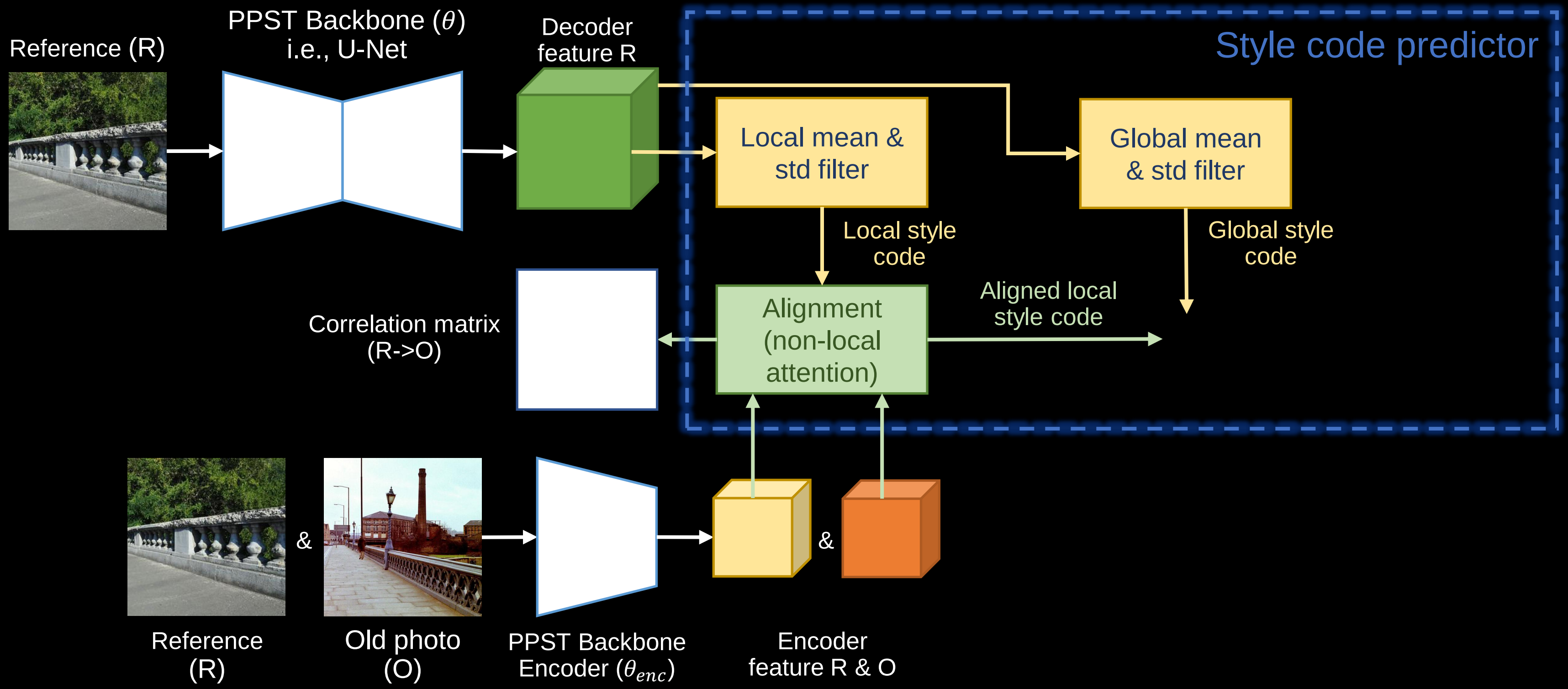


Proposed Method – MROPM-Net

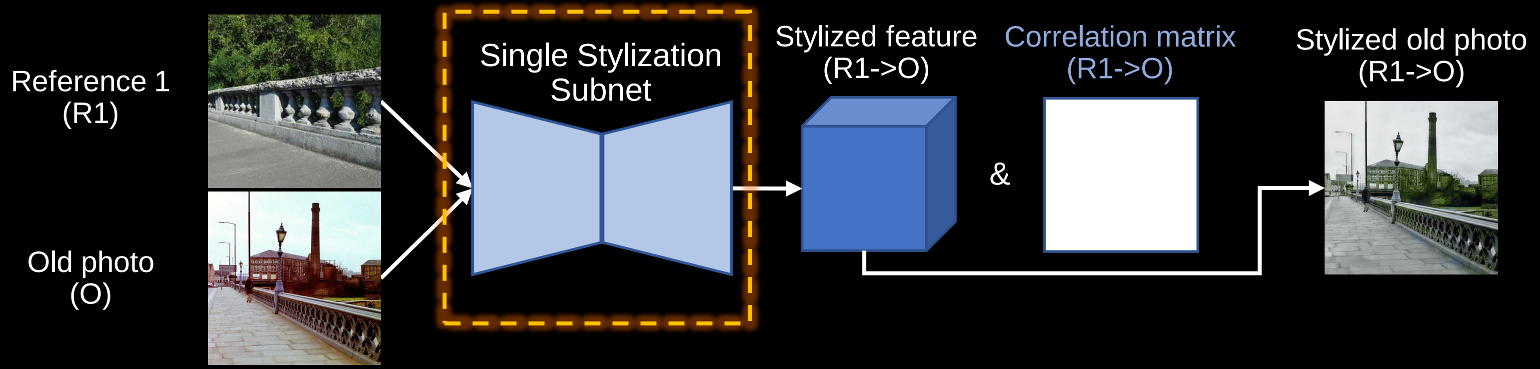


Single stylization subnet details

style code: feature's mean and std
same color: same weight parameters
PPST: pretrained PST

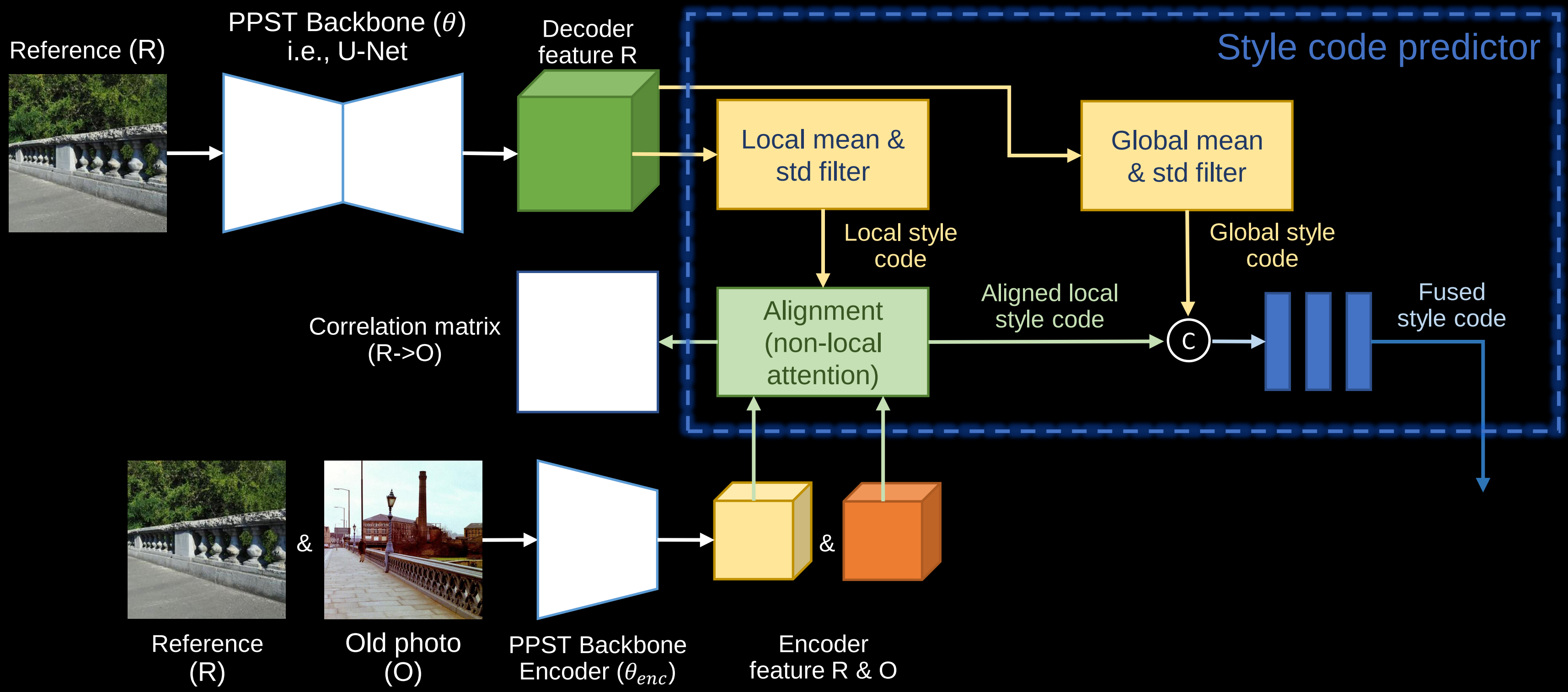


Proposed Method – MROPM-Net

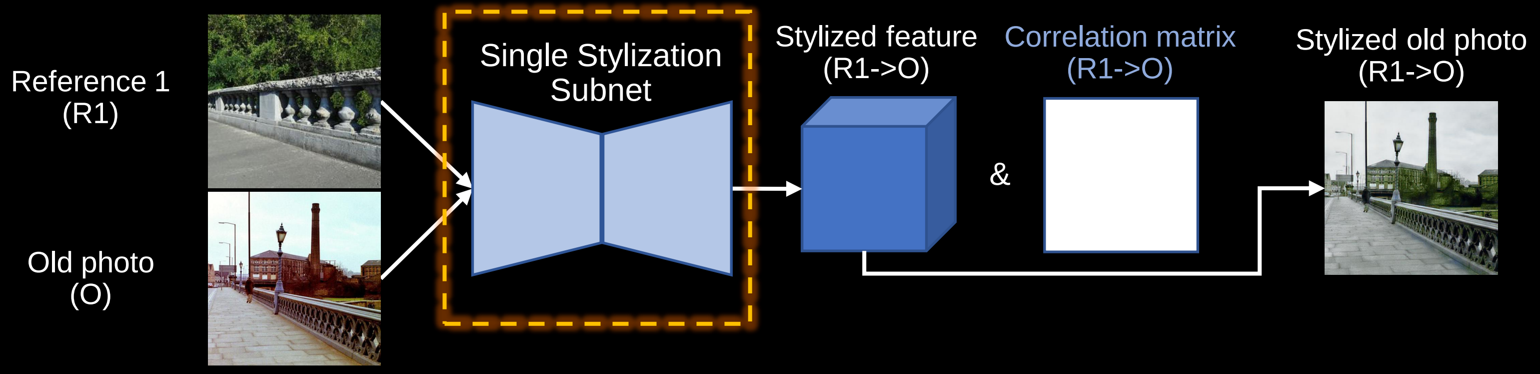


Single stylization subnet details

style code: feature's mean and std
same color: same weight parameters
PPST: pretrained PST

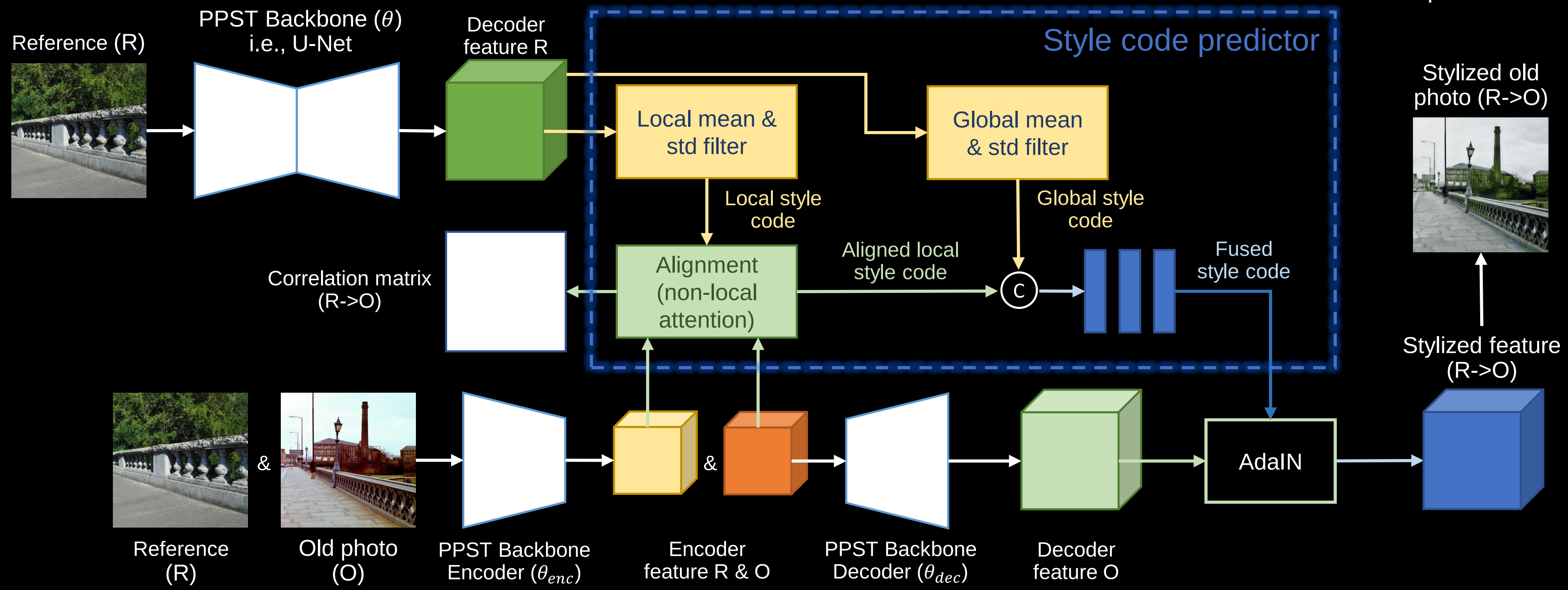


Proposed Method – MROPM-Net

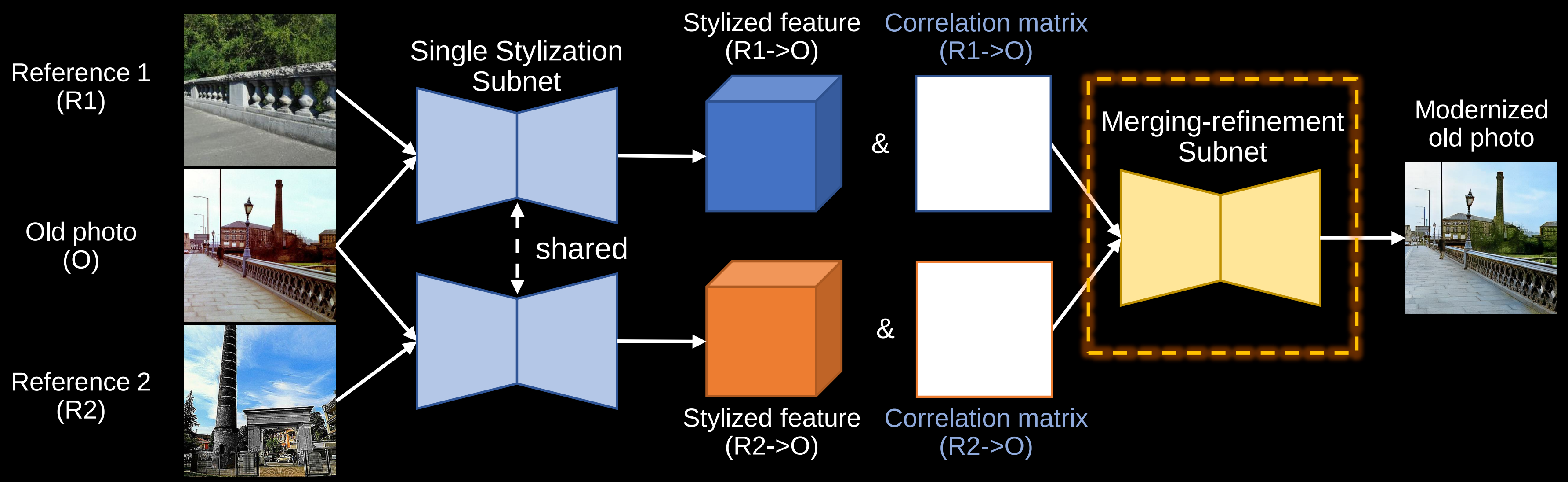


Single stylization subnet details

style code: feature's mean and std
same color: same weight parameters
PPST: pretrained PST

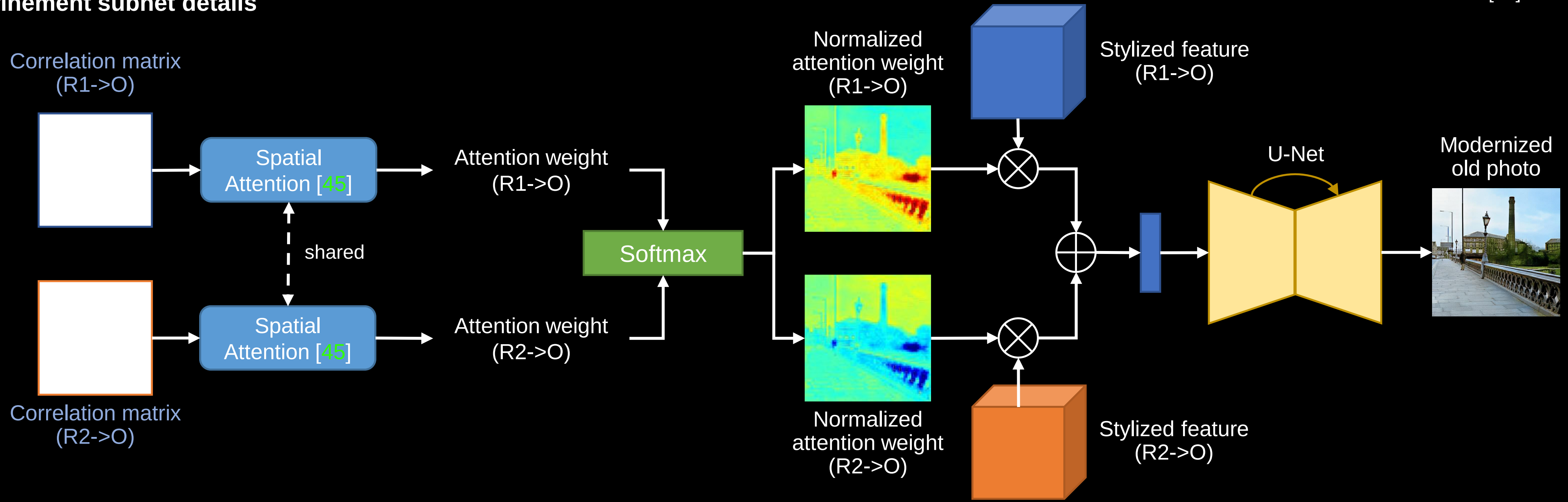


Proposed Method – MROPM-Net

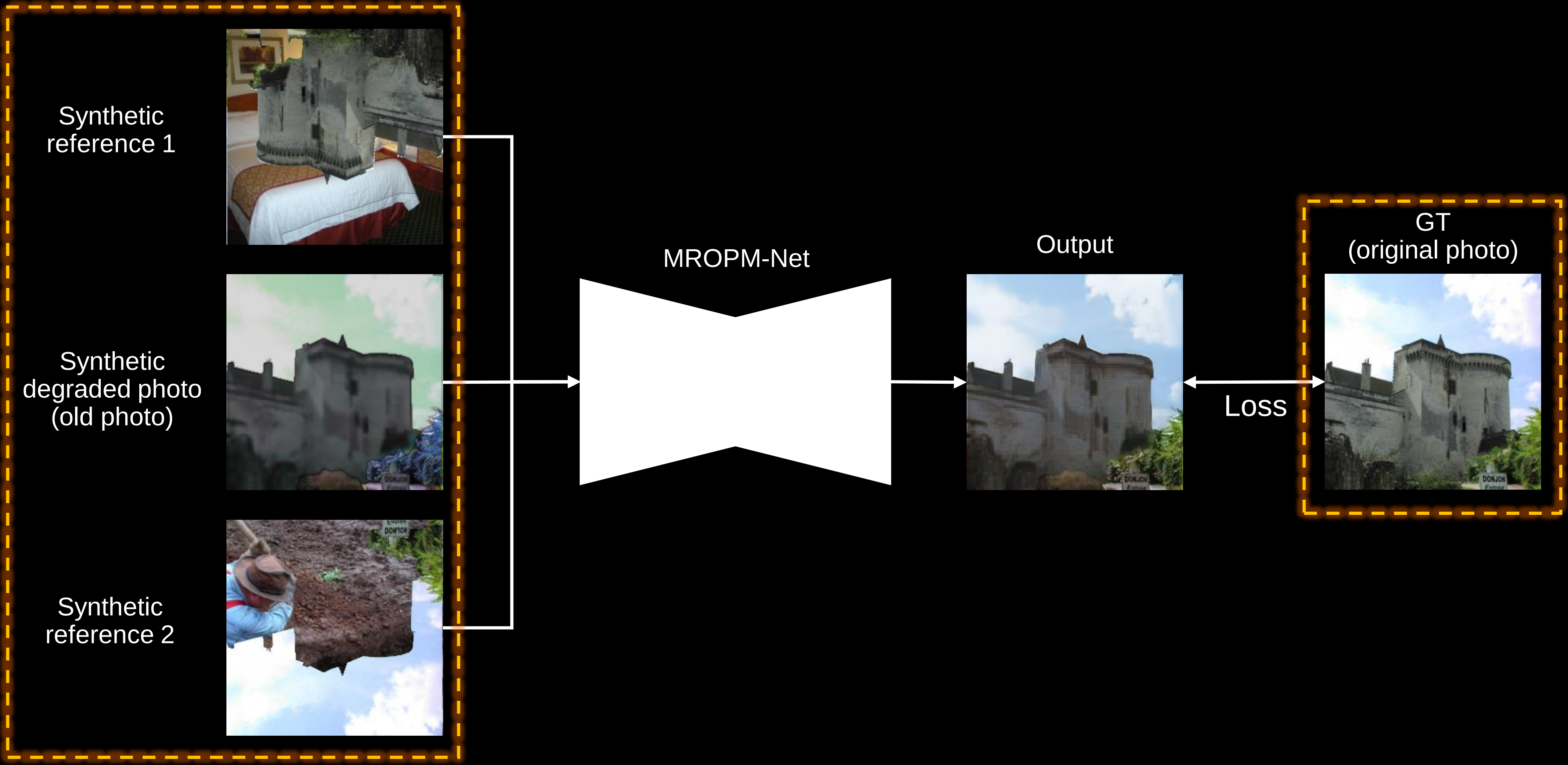


Merging-refinement subnet details

[45] Woo *et al*, ECCV 2018

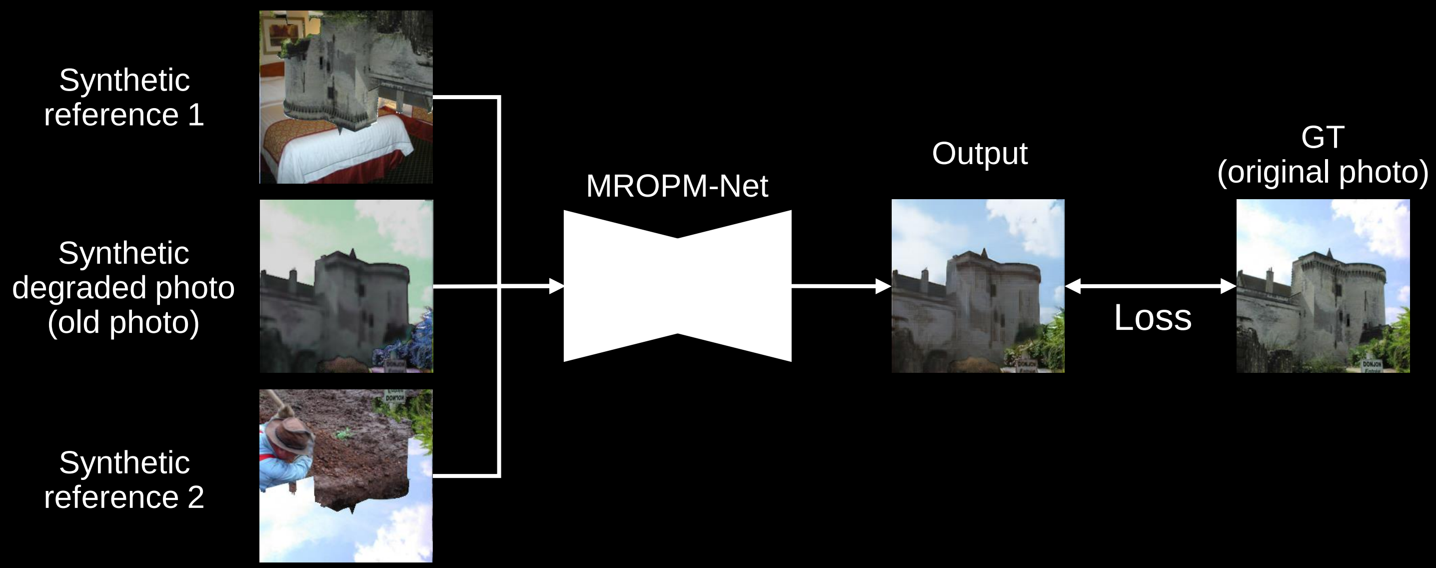


Proposed Method – Training Strategy



Proposed Method – Training Strategy

Data generation



Dataset with dense segmentation

[5] Caesar et al, CVPR 2018
[54] Zhou et al, CVPR 2017

Dataset* i.e., COCO-Stuff [5]

$D \sim$

Image

Corresponding segmentation mask

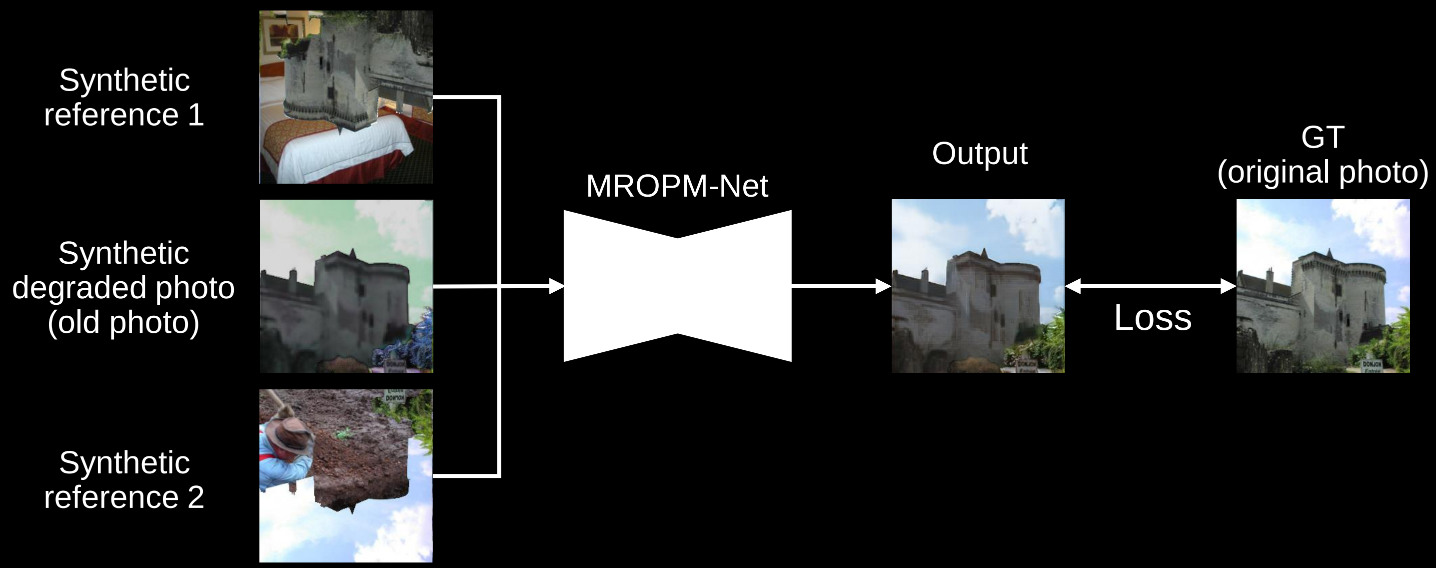
*Other datasets that have dense semantic segmentation masks can also be used e.g., ADE20K [54]

VICLab Video and Image Computing Lab

KAIST

Proposed Method – Training Strategy

Data generation



Style variant- / invariant- transformation

Style Variant Transformation (SVT)

Random color jittering, synthetic unstructured degradation i.e., blur, noise, resizing, and compression artifact*

Can change the statistics, i.e., mean and std, of certain regions

Style Invariant Transformation (SIT)

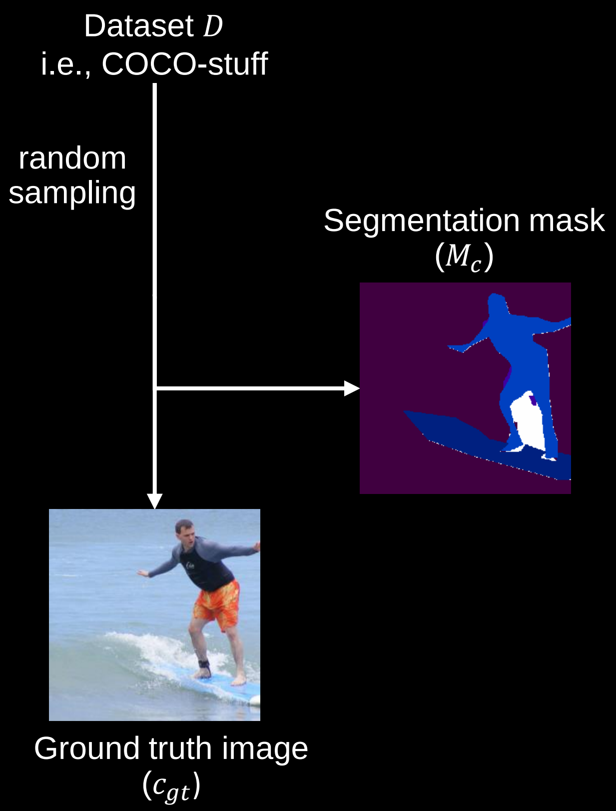
Random rigid transformation, i.e., rotation, flipping, translation (only for the region that can be translated)

Cannot change the statistics, i.e., mean and std, of certain regions

*Other types of degradations, e.g., scratches, can be included for better generalization to these types of degradations

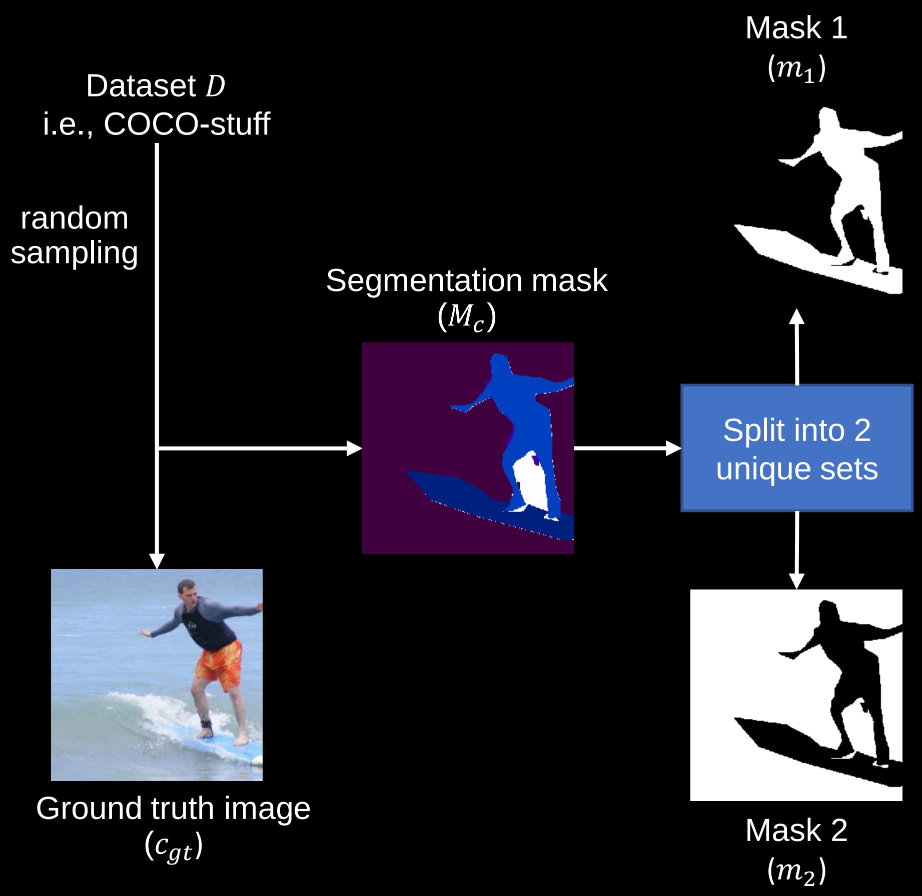
Proposed Method – Training Strategy

Data generation examples for two references



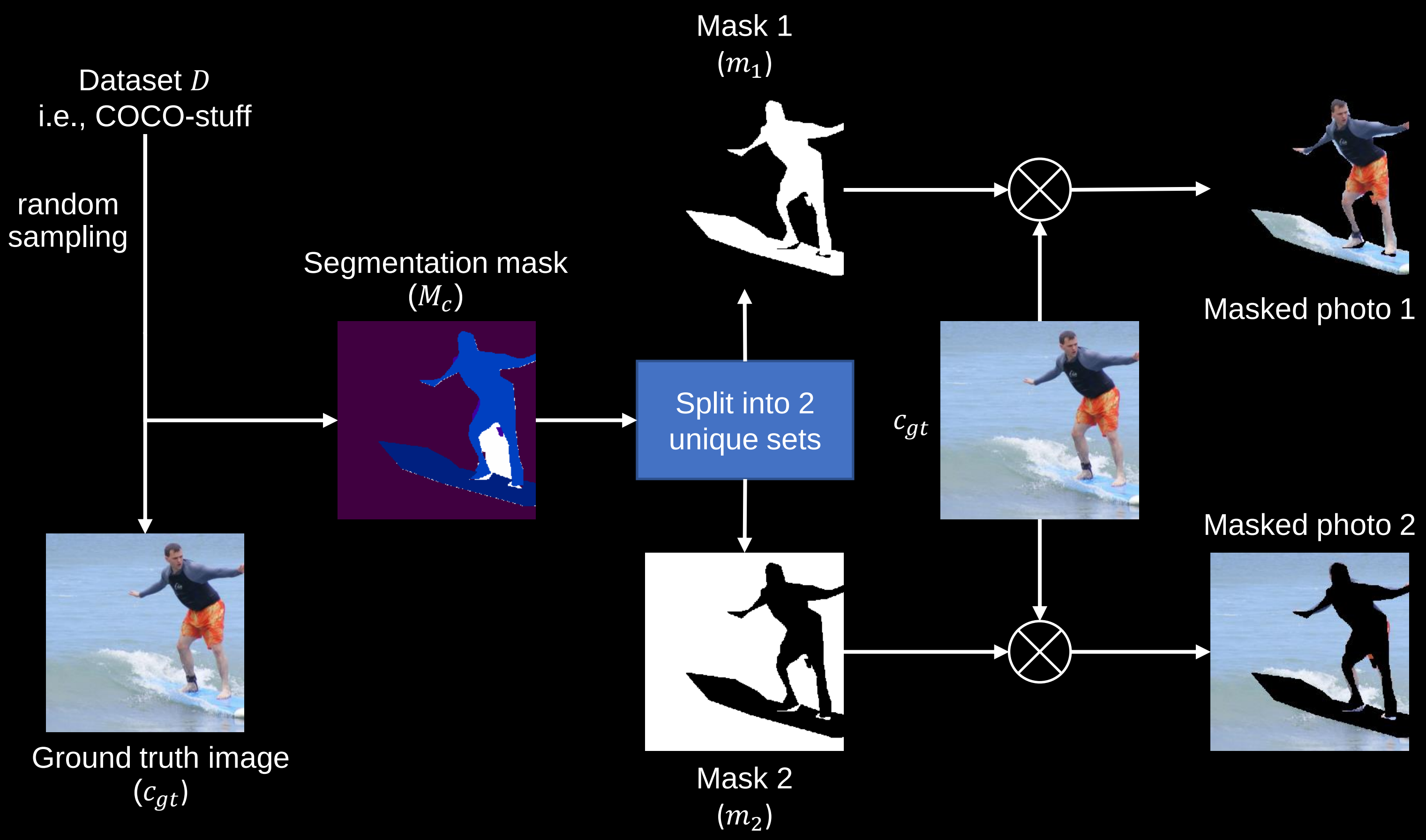
Proposed Method – Training Strategy

Data generation examples for two references



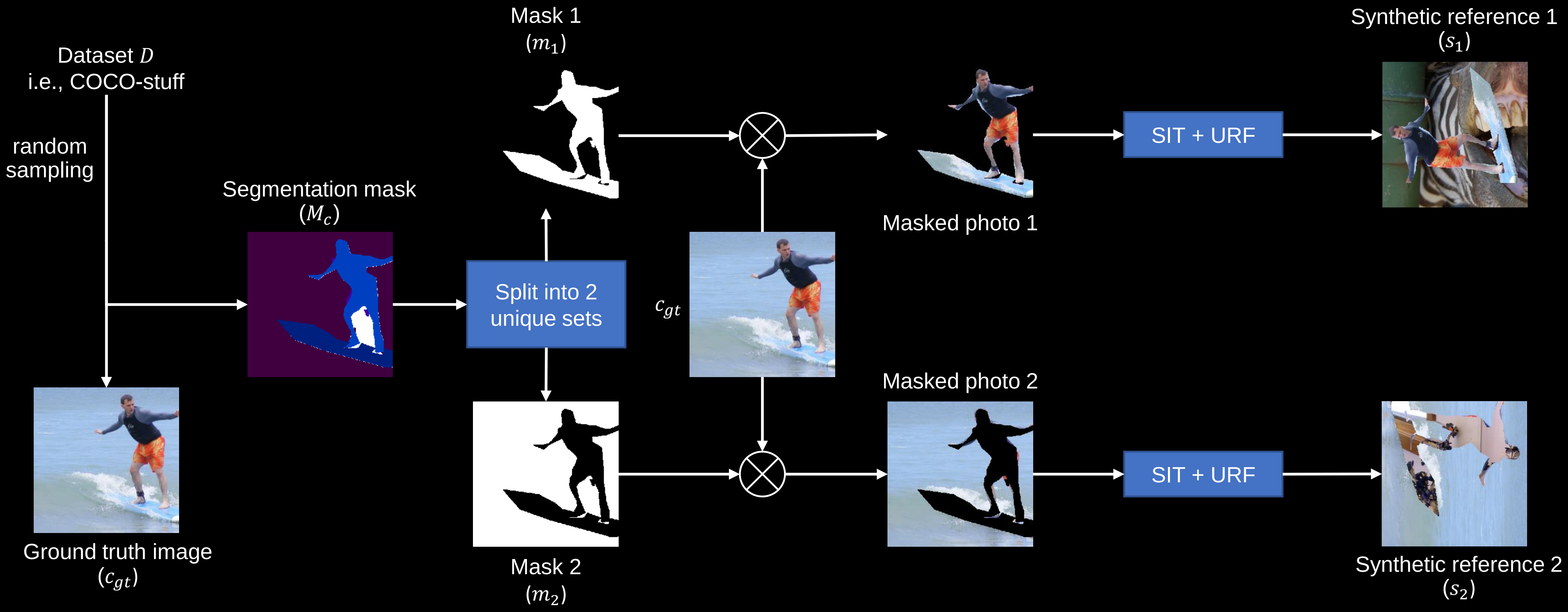
Proposed Method – Training Strategy

Data generation examples for two references



Proposed Method – Training Strategy

Data generation examples for two references



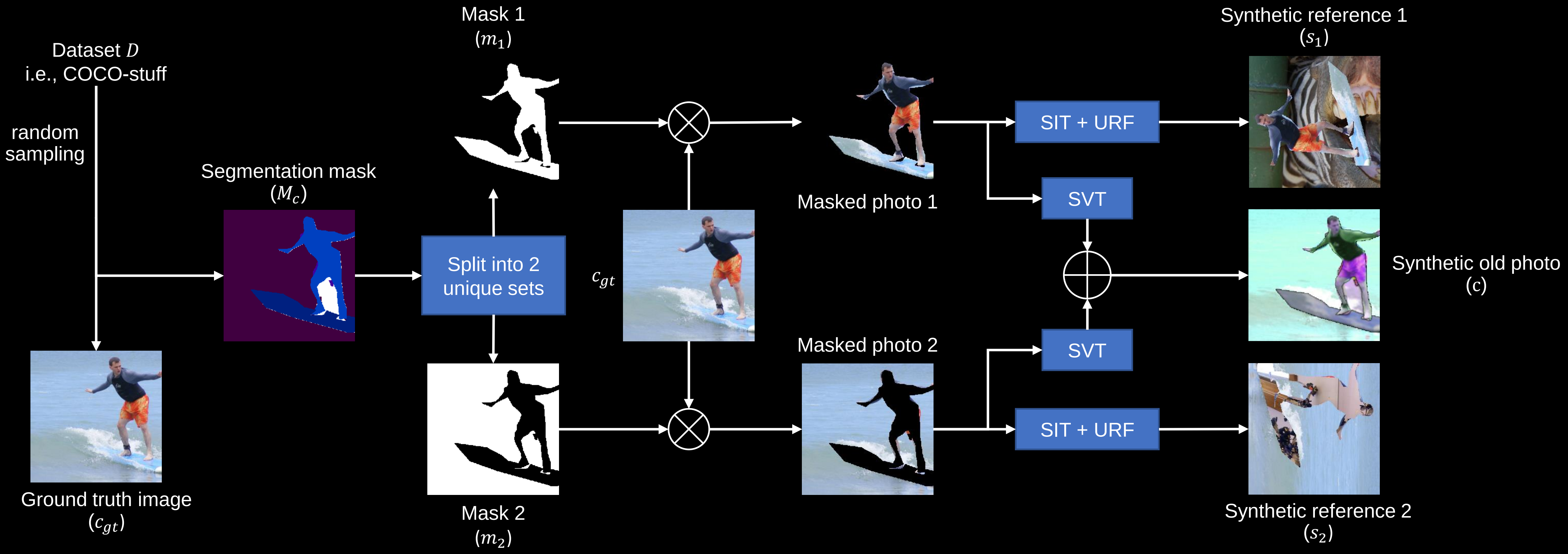
SVT: Style variant transformation

SIT: Style invariant transformation

URF: Unmasked region filling using a different random image from the same dataset

Proposed Method – Training Strategy

Data generation examples for two references



SVT: Style variant transformation

SIT: Style invariant transformation

URF: Unmasked region filling using a different random image from the same dataset

*Additional details related to the network and training strategy can be found in the paper and supplementary material

Proposed Cultural Heritage Dataset (CHD)

- ✓ 644 old color photos from the 20th century
- ✓ Three national museums in South Korea
- ✓ Outdoor and indoor natural scenes of cultural heritage e.g., special exhibitions and excavation ruins
- ✓ Containing little structured degradations, e.g., scratches, but varying unstructured degradations and color degradations

Proposed Dataset – Comparison

HWFD [11]

RealOld [21]

Ours

	HWFD [11]	RealOld [21]	Ours
Number of images	224	200	644
Era	19-20th century	-	20th century
Content type	Face	Portrait	Indoor & outdoor natural scenes
Color space	Greyscale	Greyscale	Color
Resolution	133 x 133 until 1024 x 1024	-	1024 x 1024
Expert ground-truth	✕	✓	✕

[11] Time-travel rephotography, Luo *et al*, TOG 2021
[21] Pik-fix, Xu *et al*, WACV 2023

- ✓ Our dataset has the **greatest number of images**
- ✓ Our dataset contains **indoor & outdoor natural scenes** which are more complex than face and portrait photos
- ✓ 130 old photos (~20% of the data) are augmented with references crawled from the internet for the evaluation

Experiments – Quantitative Evaluation

Synthetic data evaluation – ADE20K (different from training data)

Method	**PSNR ↑**	**SSIM ↑**	**LPIPS ↓**
ExColTran [50] + OPR-R	19.5796	0.7885	0.2563
ReHistoGAN [1] + OPR-R	<u>20.0458</u>	**0.7987**	<u>0.2109</u>
MAST [15] + OPR-R	19.0148	0.7853	0.2270
PCAPST [6] + OPR-R	19.1731	0.7908	0.2197
Ours	**21.2212**	<u>0.7919</u>	**0.2027**

Best PSNR and LPIPS, and comparable SSIM score

The method can **effectively utilize the reference** to jointly **stylize** and **enhance** the synthetically degraded images while preserving the structure

Experiments – Quantitative Evaluation

Real old photo evaluation

Method	**NIQE ↓**	**BRISQUE ↓**
OPR [42]	4.8705	21.4588
ExColTran [50] + OPR	4.9415	18.8971
ReHistoGAN [1] + OPR	4.8051	26.2557
MAST [15] + OPR	4.8111	18.9555
PCAPST [6] + OPR	4.7094	18.9860
Ours – Single Reference	<u>3.4737</u>	<u>15.5152</u>
Ours – Multiple References	**3.4487**	**15.4180**

- ✓ Best NIQE and BRISQUE compared to other baselines in single-reference setting
- ✓ Further improves the performance by using multiple references

Experiments – Quantitative Evaluation

User study

Method	**Top 1 (%)**	**Top 2 (%)**	**Top 3 (%)**	**Top 4 (%)**	**Top 5 (%)**
OPR [42]	<u>17.44</u>	<u>39.83</u>	57.05	70.90	87.22
ExColTran [50] + OPR	1.62	5.13	10.77	24.87	47.27
ReHistoGAN [1] + OPR	7.91	32.27	<u>61.84</u>	<u>83.80</u>	<u>96.92</u>
MAST [15] + OPR	5.68	21.62	41.92	66.33	86.20
PCAPST [6] + OPR	10.98	28.50	44.87	61.97	84.66
Ours	**56.37**	**72.69**	**83.55**	**92.14**	**97.74**

Our method outperforms other baselines with a 56.37% chance selected as the best method

Experiments – Ablation Study

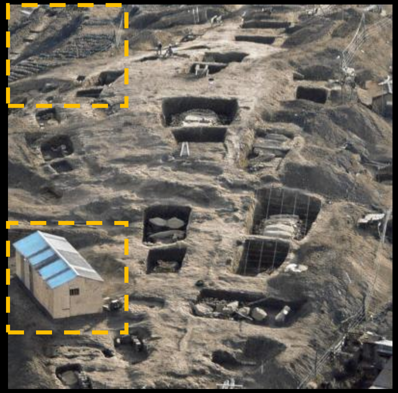
Ablation study for single stylization subnet


Old photo


Reference


w/o alignment


w/o fusion


Full model

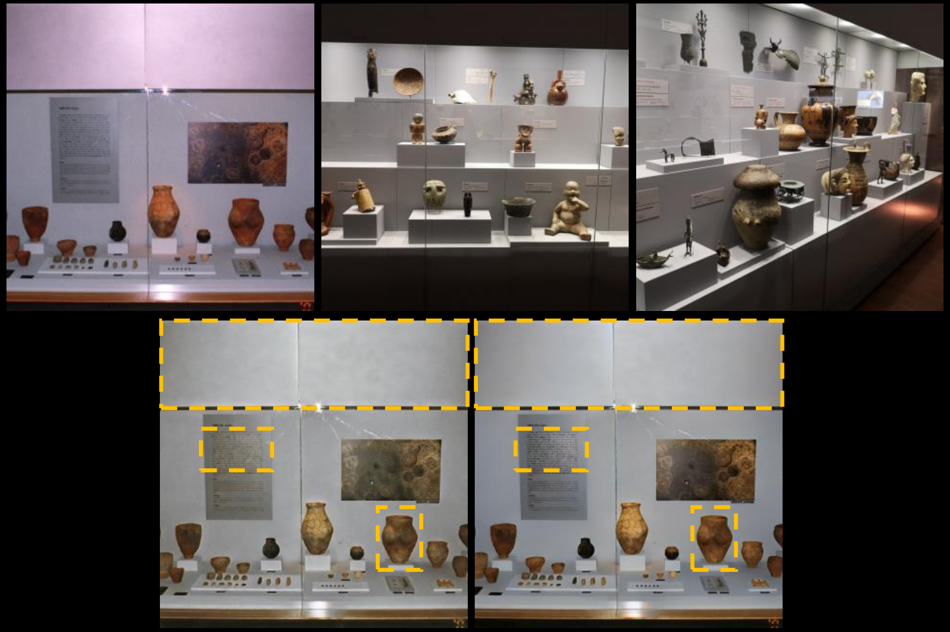
- ✓ w/o alignment: fails to accurately transfer the local styles of objects
- ✓ w/o fusion: coarse local and global stylization

Experiments – Ablation Study

Ablation study for merging-refinement subnet

Real old photo evaluation

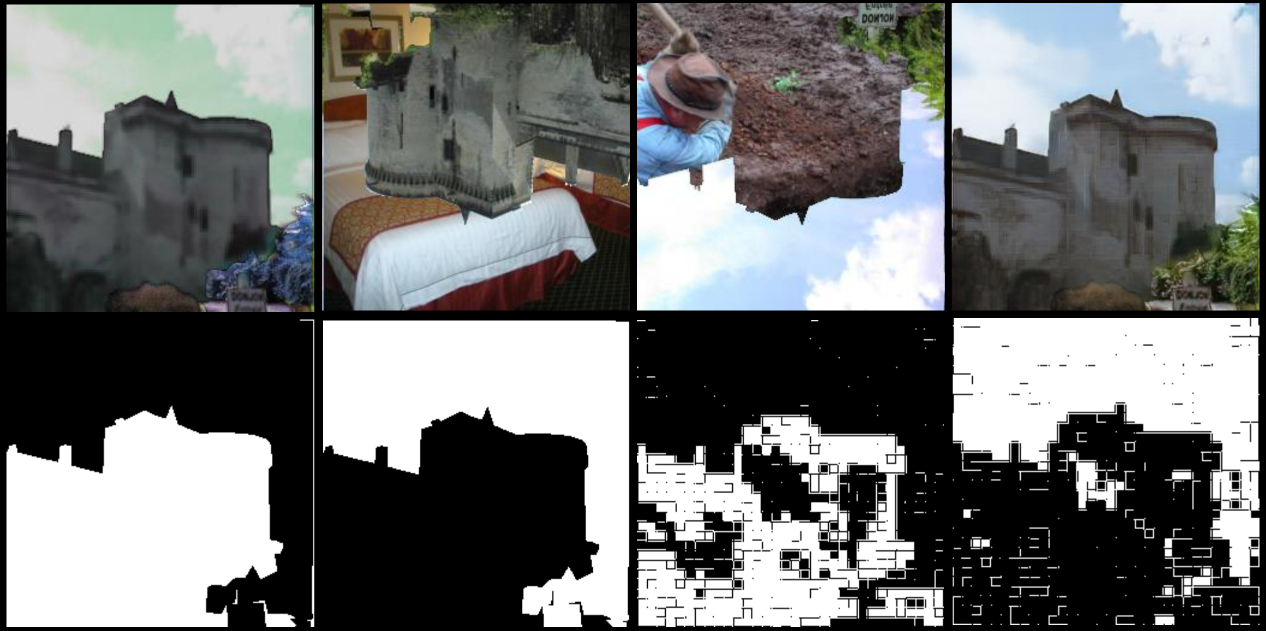
Old photo | Reference 1 | Reference 2



w/o merging-refinement subnet | Full model

Effectivity of merging-refinement subnet in synthetic data

Input | Reference 1 | Reference 2 | Output



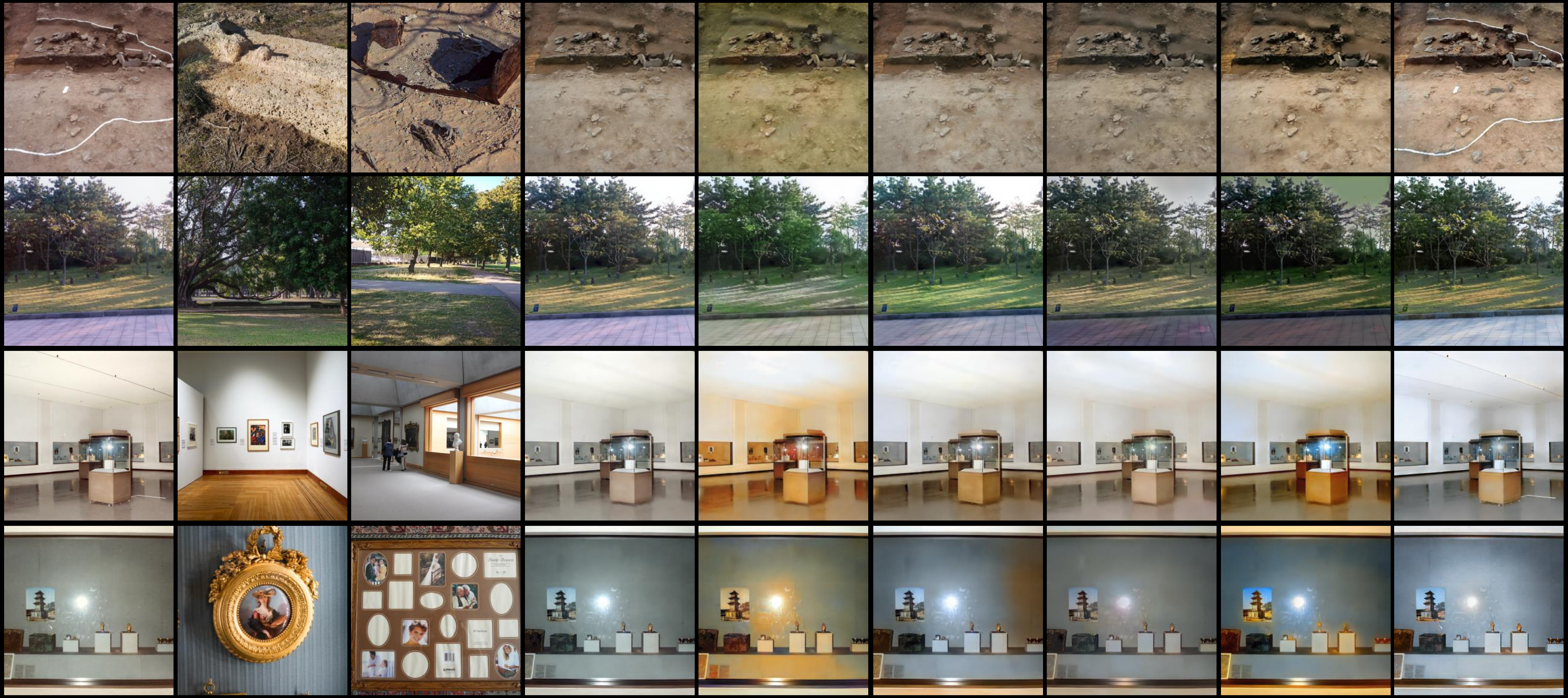
GT mask 1 | GT mask 2 | Predicted mask 1 | Predicted mask 2

Predicted masks are generated by thresholding spatial attention output

The subnet can select relevant regions from multiple references to transfer their styles to the corresponding region in the old photo input

Experiments – Qualitative Evaluation

Old photo	Reference 1	Reference 2	OPR [42]	ExColTran [50] + OPR	ReHistoGAN [1] + OPR	MAST [15] + OPR	PCAPST [6] + OPR	Ours



Better qualitative results compared to the baselines (baselines use reference 1)

Additional Results – Old Photos in The Wild

Old photo

OPR

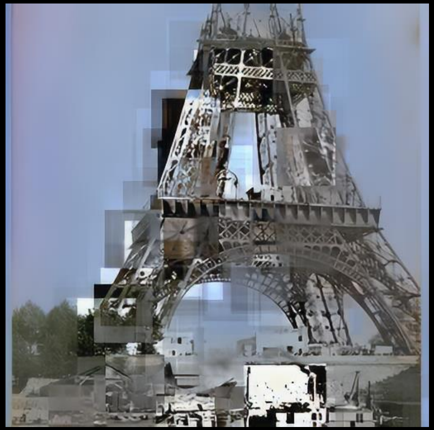
ExColTran + OPR

ReHistoGAN + OPR

Reference 1

Reference 2

MAST + OPR



PCAPST + OPR

Ours

Our method can better match the corresponding semantic regions between the old photo and multiple references even when the **viewpoint and scale are different**

Additional Results – Semantic Photorealistic Style Transfer

Content | Style | ExColTran | ReHistoGAN | MAST | PCAPST | Ours – Only PST | Ours – Full

Our method can be used for semantic photorealistic style transfer (PST)

Additional Results – User Study

Old photo

OPR (11.11%)

ExColTran + OPR (0.00%)

ReHistoGAN + OPR (5.55%)

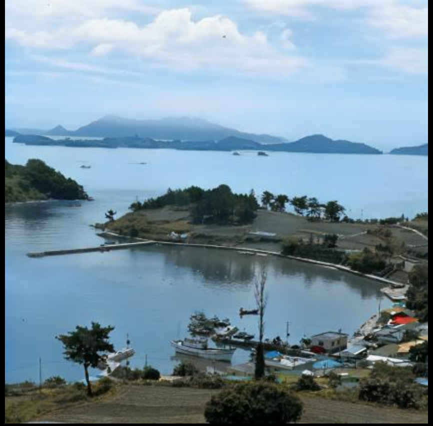
Reference 1

Reference 2

MAST + OPR (0.00%)

PCAPST + OPR (0.00%)

Ours (83.34%)



Baselines use reference 1 as their reference

Additional Results – User Study

Old photo

OPR (16.67%)

ExColTran +
OPR (0.00%)

ReHistoGAN +
OPR (5.55%)

Reference 1

Reference 2

MAST +
OPR (0.00%)

PCAPST +
OPR (5.55%)

Ours (72.23%)

Baselines use reference 1 as their reference

Additional Results – User Study

Old photo

OPR (11.11%)

ExColTran +
OPR (0.00%)

ReHistoGAN +
OPR (5.55%)

Reference 1

Reference 2

MAST +
OPR (0.00%)

PCAPST +
OPR (16.67%)

Ours (66.67%)

Baselines use reference 1 as their reference

Additional Results – User Study

Old photo

OPR (44.44%)

ExColTran + OPR (0.00%)

ReHistoGAN + OPR (16.67%)

Reference 1

Reference 2

MAST + OPR (11.11%)

PCAPST + OPR (11.11%)

Ours (16.67%)

Baselines use reference 1 as their reference

*Additional results and ablation studies can be found in the paper and supplementary material

SCAN ME

Thank you! For more details, please visit:

kaist-viclab.github.io/old-photo-modernization

Modernizing Old Photos Using Multiple References via Photorealistic Style Transfer

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

June 21, 2023 – PM – Number 11

WED-PM-011

VICLab Video and Image Computing Lab

KAIST