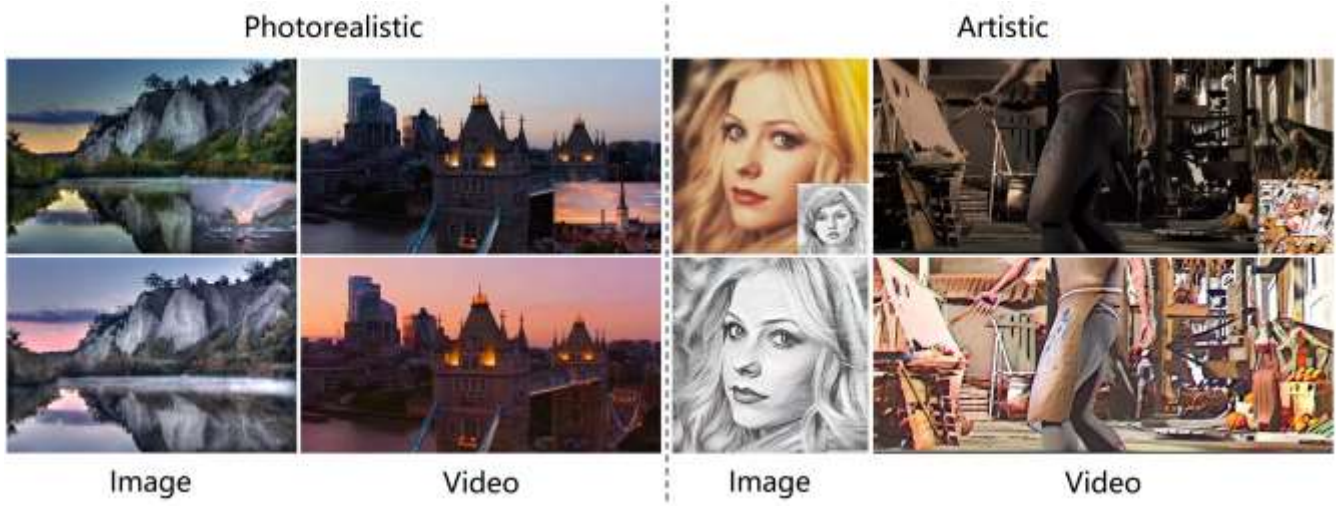
JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

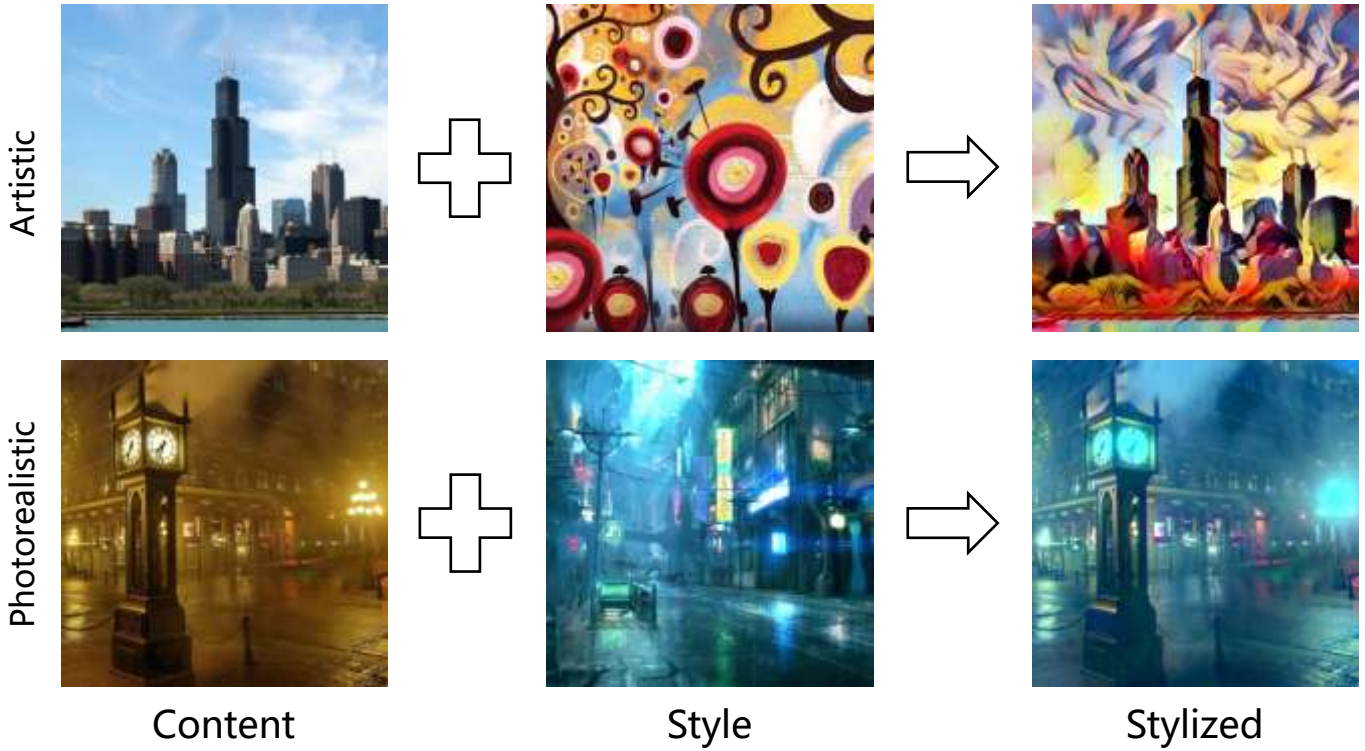
CAP-VSTNet: Content Affinity Preserved Versatile Style Transfer

Linfeng Wen Chengying Gao Changqing Zou

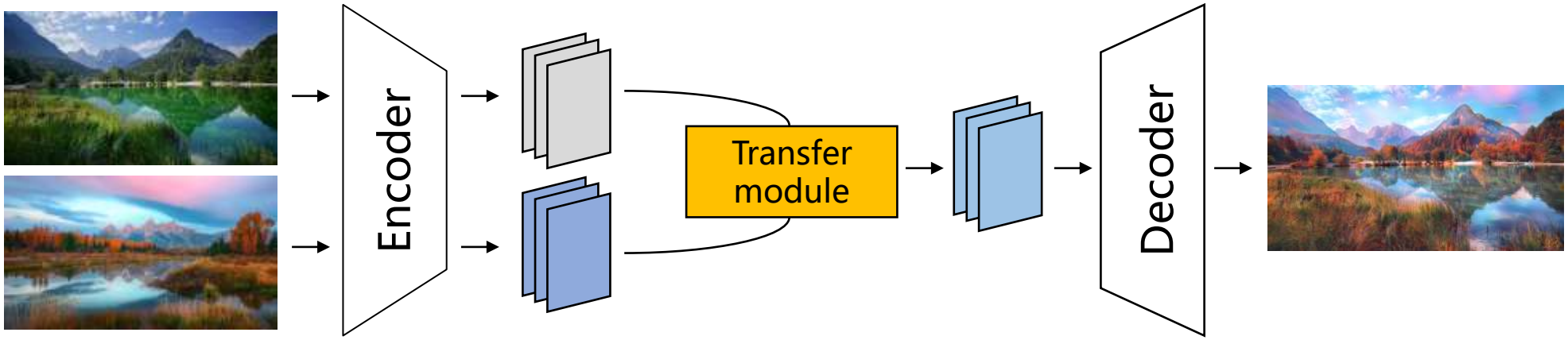
Sun Yat-Sen University, Guangzhou, China
CAD&CG, Zhejiang University and Zhejiang Lab, Hangzhou, China



The task of style transfer



Popular approach



Problem: content affinity loss

Inconsistent stylization

Unclear details

Noticeable seams

Ours

CAP-VSTNet

① Reversible residual network based on channel refinement module
- avoid content affinity loss

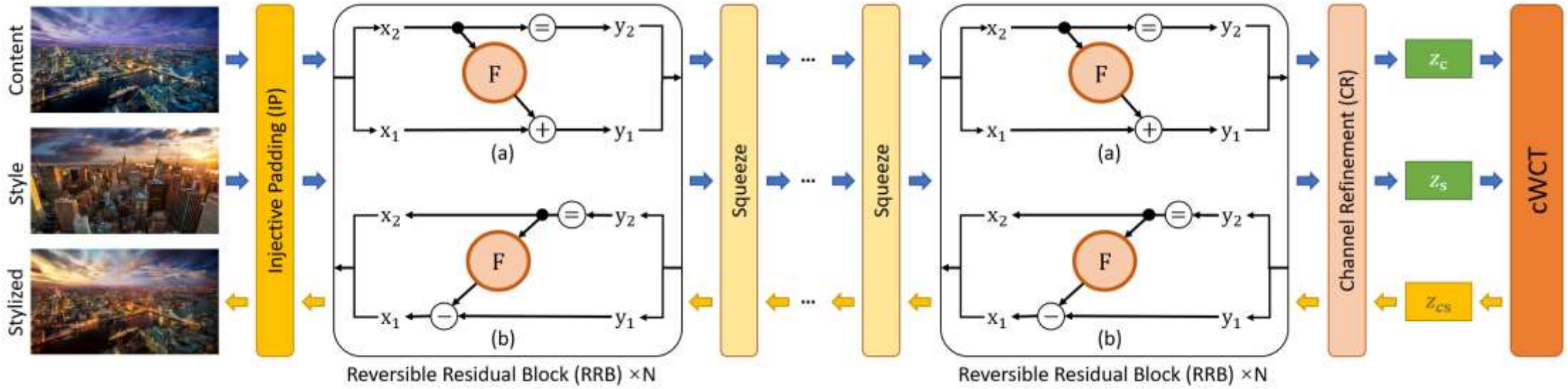
② Unbiased linear transform based on Cholesky decomposition
- preserve feature affinity

③ Training loss based on Matting Laplacian
- preserve pixel affinity

CAP-VSTNet

- State-of-the-art performance
- Real-time image/video style transfer
- Style interpolation
- Ultra-resolution (4K)

Illustration of CAP-VSTNet



Three steps

- Forward inference (➡)
- Transfer (cWCT)
- Backward inference (⬅)

Details of CAP-VSTNet

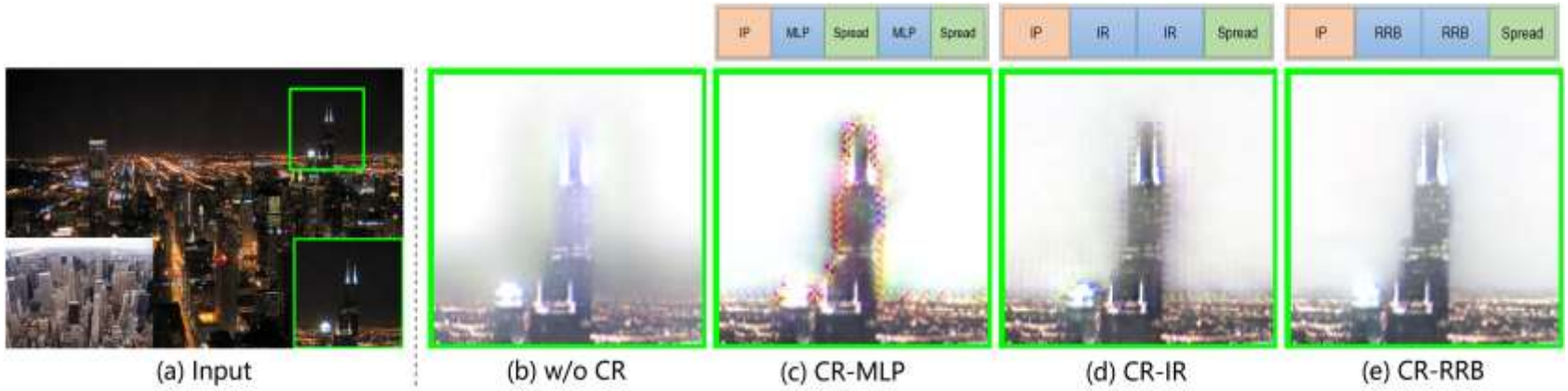
- Reversible residual network

  - Bijective transformation: preserve all information to avoid content affinity loss

Details of CAP-VSTNet

- Reversible residual network

- Bijective transformation: preserve all information to avoid content affinity loss
- Channel Refinement (CR): remove redundant information in reversible network



Alternative design choices of CR module

Details of CAP-VSTNet

- cWCT

  - Unbiased linear transform: preserve feature affinity

Details of CAP-VSTNet

- cWCT

- Unbiased linear transform: preserve feature affinity
- Cholesky decomposition: derivable, stable and efficient

Details of CAP-VSTNet

- cWCT

- Unbiased linear transform: preserve feature affinity
- Cholesky decomposition: derivable, stable and efficient

Method		AdaIN	WCT	LinearWCT	cWCT
Reversible		✓	✓		✓
Stability		✓		✓	✓
Learning-free			✓		✓
Time	C=32	0.066	1.186	0.288	0.097
	C=256	0.424	3.205	2.419	0.808

Details of CAP-VSTNet

- Training loss

- $L_{total} = L_s + \lambda_m L_m + \lambda_{cyc} L_{cyc}$,
- Matting Laplacian loss L_m: address pixel affinity problem led by the linear transform

Details of CAP-VSTNet

- Training loss

- $L_{total} = L_s + \lambda_m L_m + \lambda_{cyc} L_{cyc}$,
- Matting Laplacian loss L_m: address pixel affinity problem led by the linear transform
- Cycle reconstruction loss L_{cyc}: alleviate numerical error problem of reversible network

Details of CAP-VSTNet

- Video style transfer

- Content of stylized video: stable
- Style of stylized video : need more constraint

Details of CAP-VSTNet

- Video style transfer

- Content of stylized video: stable
- Style of stylized video : need more constraint ⇨ ① adjust the style loss L_s

  ② add regularization in L_{total}

Experiments

- Photorealistic image style transfer

Method	PhotoWCT [25]	WCT^2 [40]	PhotoNet [2]	DSTN [14]	PCA-KD [8]	Ours
SSIM↑	0.582	<u>0.644</u>	0.608	0.566	0.634	**0.650**
Gram loss↓	1.539	<u>0.796</u>	1.970	0.996	1.162	**0.750**
Time↓	16.88	0.32	0.19	0.92	**0.05**	<u>0.12</u>
Parameters	8.35M	10.12M	40.24M	103.45M	334K	4.09M

- The execution time is evaluated on 1024 × 512 resolution.

Experiments

- Photorealistic image style transfer

Experiments

- Photorealistic/Artistic video style transfer

- Photorealistic video

Method	Gram loss↓	Temporal loss↓	
		i=1	i=10
WCT^2 [40]	0.665	<u>0.040</u>	<u>0.108</u>
CCPL [39]	<u>0.527</u>	0.069	0.132
Ours	**0.435**	**0.039**	**0.107**

- Artistic video

Method	Gram loss↓	Temporal loss↓	
		i=1	i=10
LinearWCT [23]	0.473	0.117	0.237
ReReVST [37]	0.815	<u>0.108</u>	<u>0.235</u>
IEContraAST [6]	1.062	0.141	0.262
CCPL [39]	**0.371**	0.128	0.251
Ours	<u>0.436</u>	**0.104**	**0.228**

Experiments

- Video style transfer
- Style interpolation
- Ultra-resolution

Experiments

[Video]

CAP-VSTNet: Content Affinity Preserved Versatile Style Transfer

Linfeng Wen Chengying Gao Changqing Zou

Sun Yat-Sen University, Guangzhou, China
CAD&CG, Zhejiang University and Zhejiang Lab, Hangzhou, China

For more information: https://github.com/linfengWen98/CAP-VSTNet

Thank you for your watching!