MicroAST: Towards Super-Fast Ultra-Resolution Arbitrary Style Transfer

Zhizhong Wang, Lei Zhao*, Zhiwen Zuo, Ailin Li, Haibo Chen, Wei Xing*, Dongming Lu

AAAI 2023 Oral

STRUCT Group Seminar
Presenter: Zhengbo Xu
2023.01.29

OUTLINE

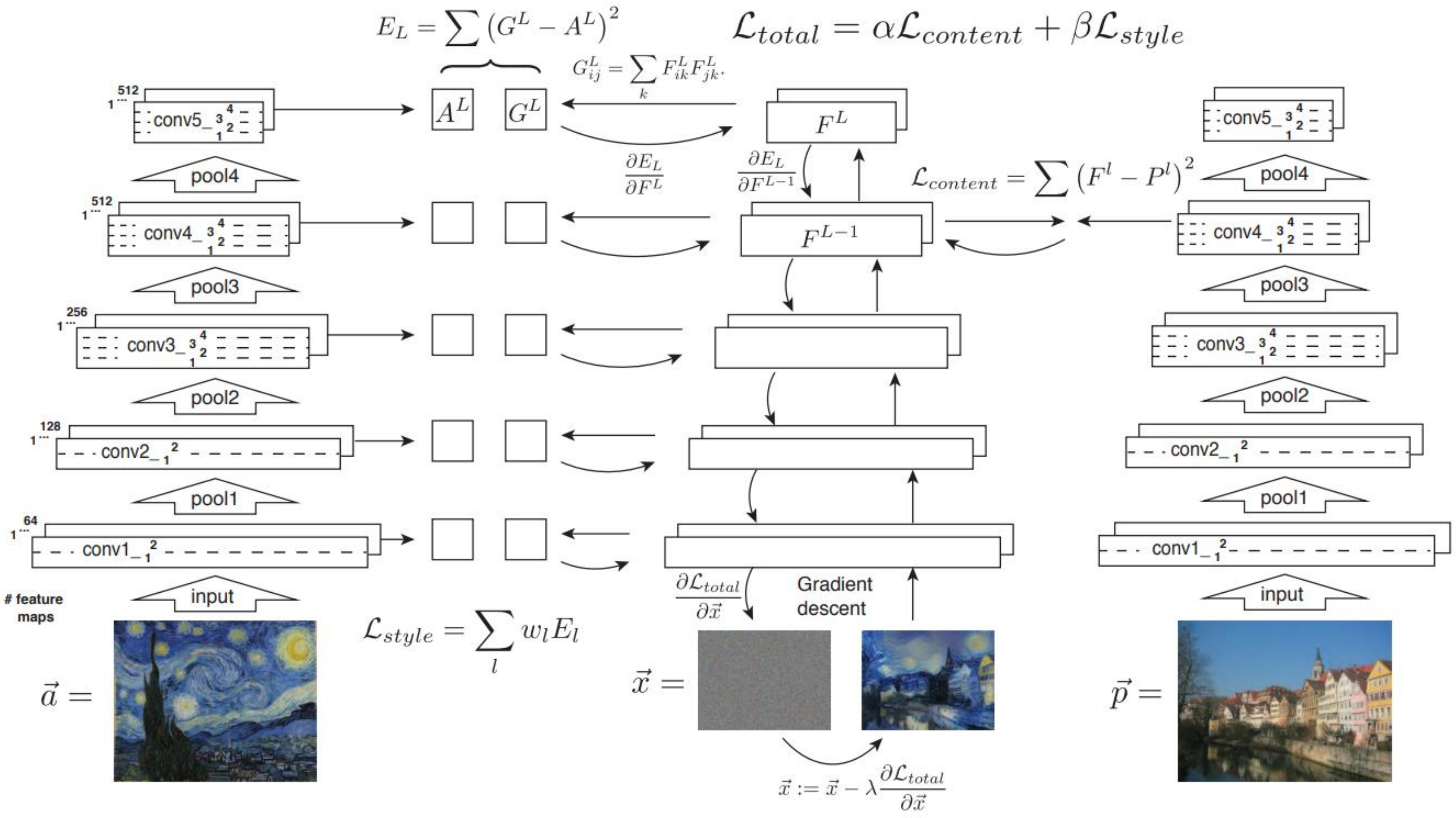
- Authorship
- Background
- Method
- Experiments
- Conclusion

OUTLINE

- Authorship
- Background
- Method
- Experiments
- Conclusion

BACKGROUND: Neural Style Transfer

Overview



BACKGROUND: Neural Style Transfer

Improved aspects

- Efficiency
- Quality
- Generalization
- Diversity
- User Control

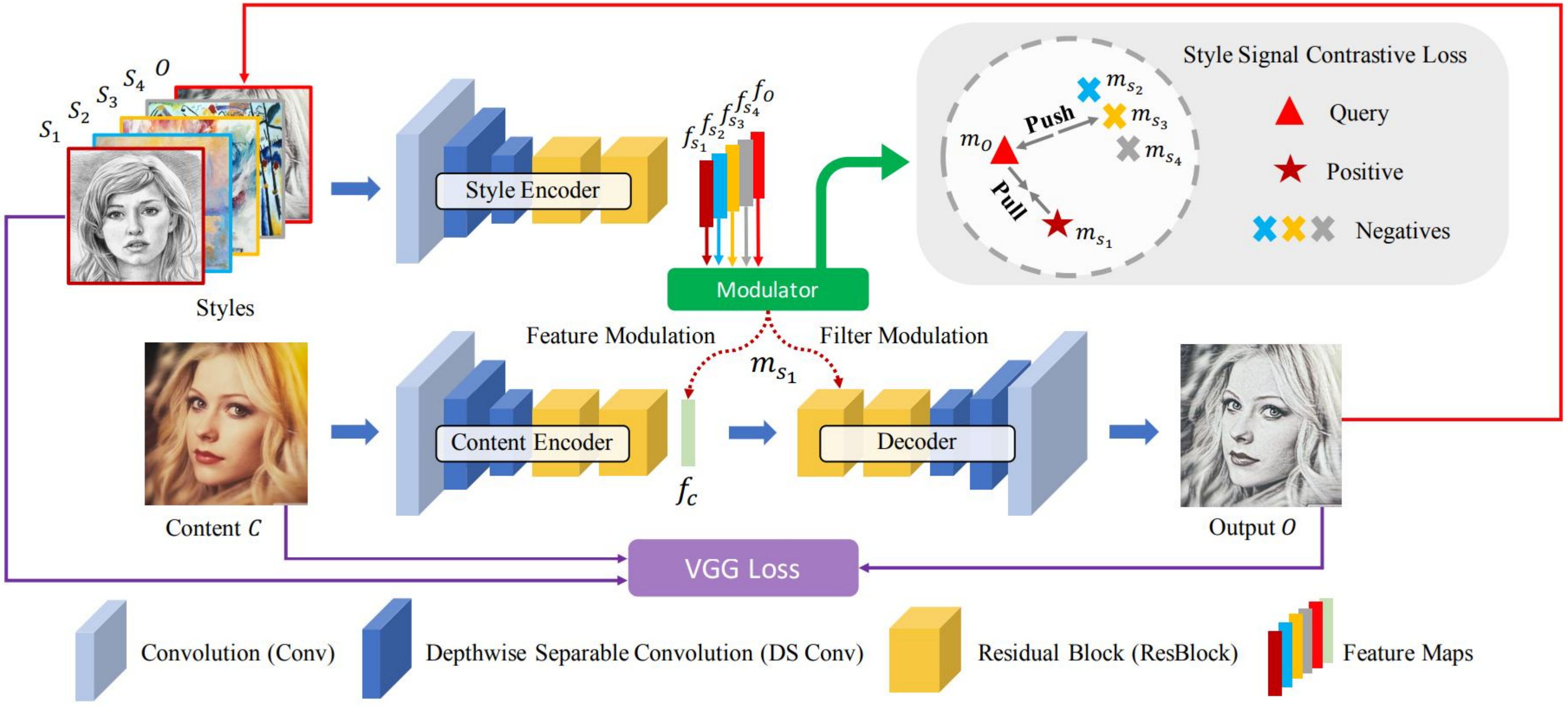
Generate 4K images with limited resources

OUTLINE

- Authorship
- Background
- Method
- Experiments
- Conclusion

METHOD

Overview



METHOD

Pipeline

- Extract features from content image C : $f_c := E_c(C)$.
- Extract features from style image S : $f_s := E_s(S)$.
- Convert f_s into style modulation signals: $m_s := \bar{\mathcal{M}}(f_s)$
- Stylize f_c using micro decoder D : $O := D(f_c, m_s)$

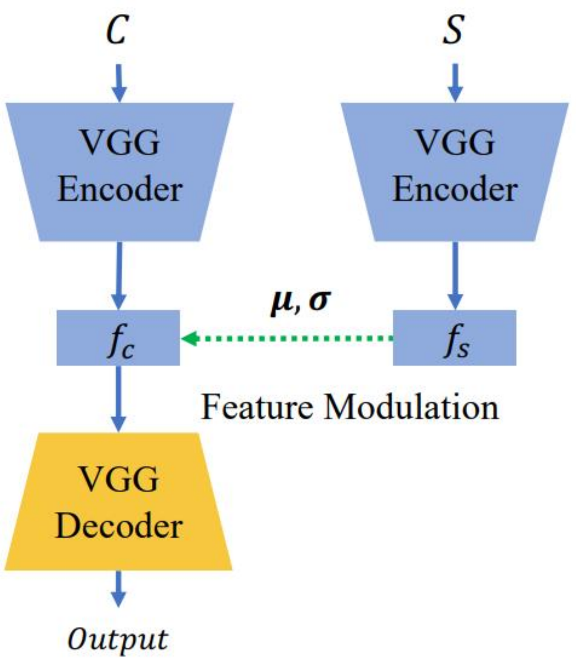
Training Loss

$$\mathcal{L}_{full} := \lambda_c \mathcal{L}_c + \lambda_s \mathcal{L}_s + \lambda_{ssc} \mathcal{L}_{ssc}.$$

Modulation Strategies in AST

AdaIN (Huang and Belongie 2017)

$$AdaIN(f_c, f_s) := \sigma(f_s)(\frac{f_c - \mu(f_c)}{\sigma(f_c)}) + \mu(f_s).$$



Features & Requirements

- content and style encoder are identical
- encoders and decoder must be as complex as VGG
- encoders are fixed, decoder is trained

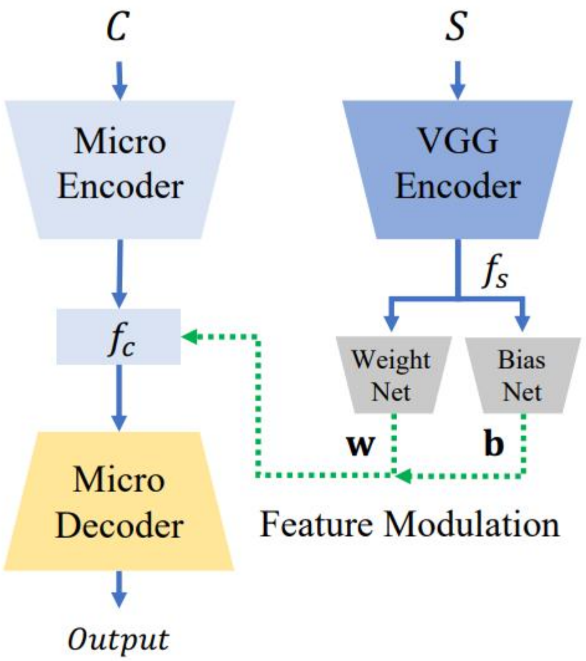
Modulation Strategies in AST

DIN (Jing et al. 2020)

$$DIN(f_c, f_s) := \mathbf{w}(\frac{f_c - \mu(f_c)}{\sigma(f_c)}) + \mathbf{b}.$$

Features & Requirements

- content encoder and decoder are lightweight
- style encoder is also high-cost VGG
- use subnets to predict weight and bias

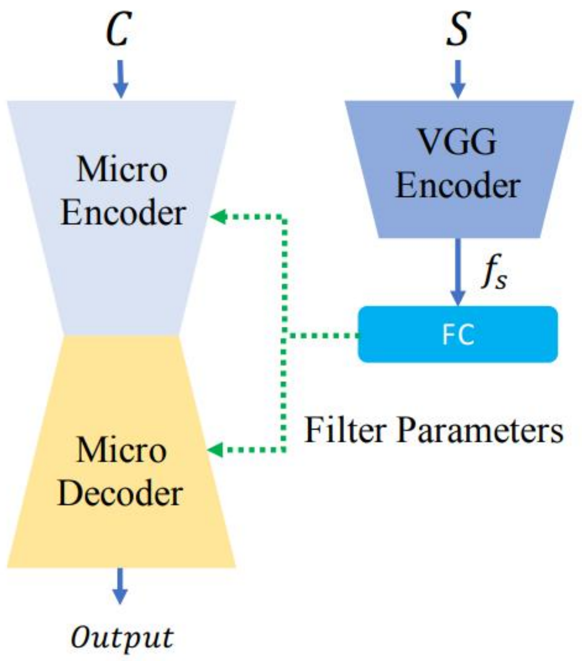


Modulation Strategies in AST

MetaNets (Shen, Yan, and Zeng 2018)

Features & Requirements

- content encoder and decoder are lightweight
- style encoder is also high-cost VGG
- style features help construct outputs
- FC layers lead to extra memory and slow inference time



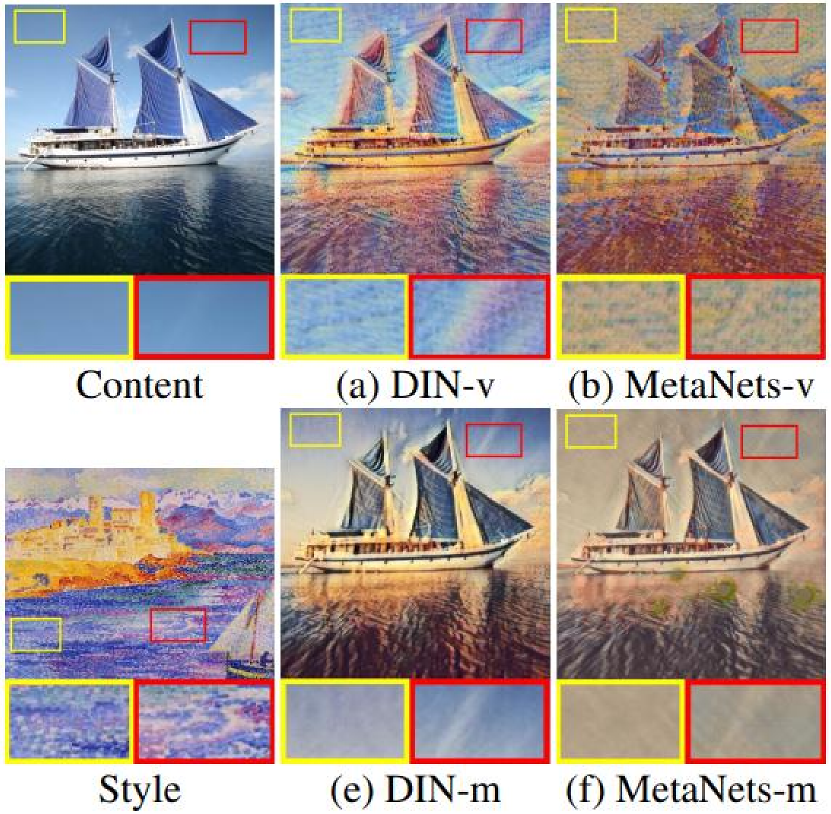
METHOD

Constrains

- The micro style encoder has limited ability to extract sufficiently complex style features
- The style signals are unitary and inflexible

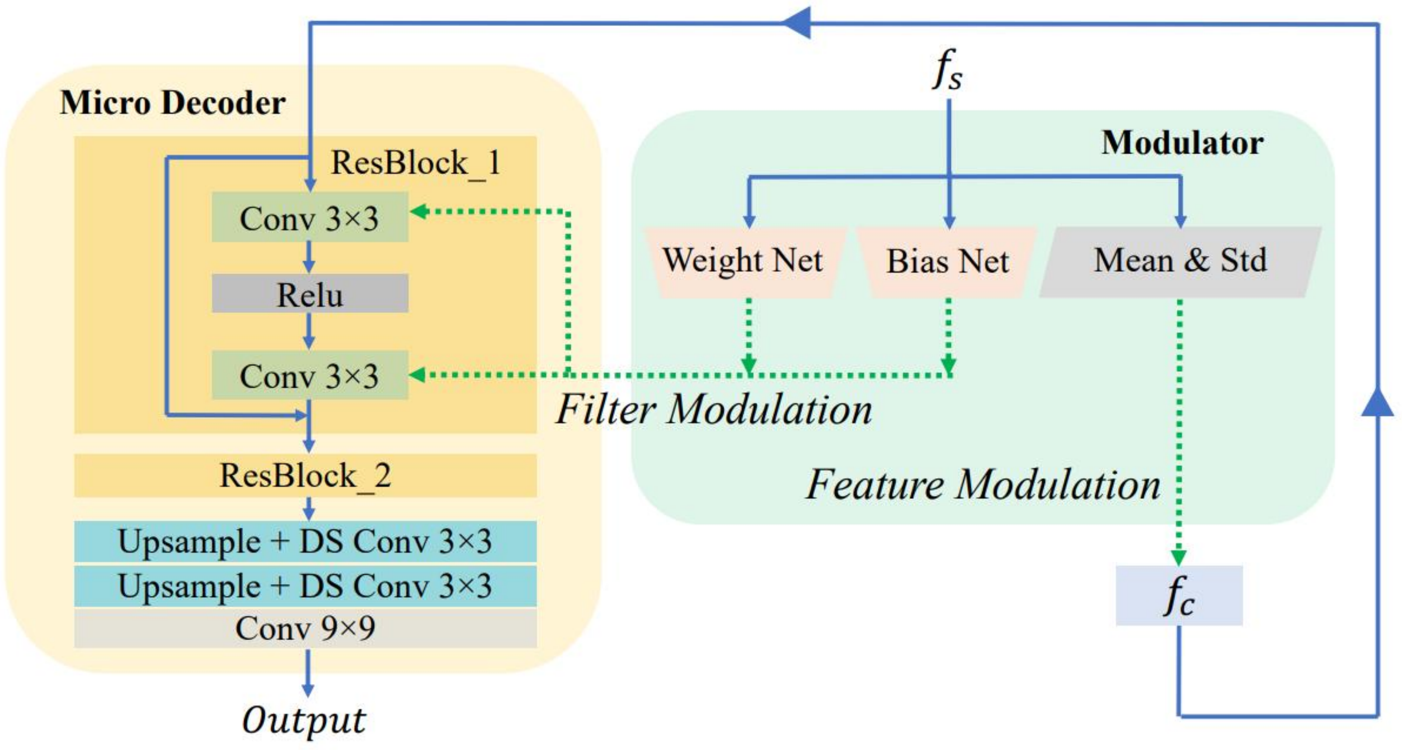
Solutions

- Introduce a dual-modulation strategy to inject more sophisticated and flexible style signals
- Propose a contrastive loss



METHOD

Dual-Modulation



$$m_s := (\boldsymbol{\mu}_s, \boldsymbol{\sigma}_s, \mathbf{w_s}, \mathbf{b_s}),$$

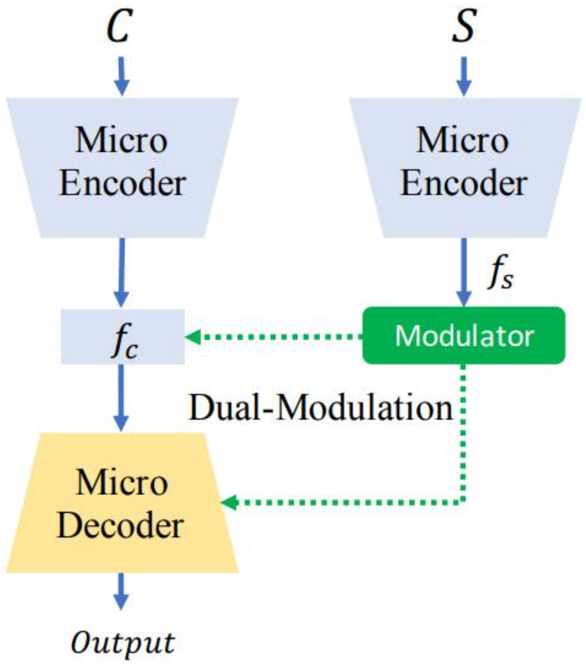
$$DualMod(D, f_c, m_s) := FeatMod(f_c, (\boldsymbol{\mu}_s, \boldsymbol{\sigma}_s)) + FilterMod(D, (\mathbf{w_s}, \mathbf{b_s})),$$

METHOD

FeatMod

$$\boldsymbol{\mu}_s := \mu(f_s), \quad \boldsymbol{\sigma}_s := \sigma(f_s),$$

$$FeatMod(f_c, (\boldsymbol{\mu}_s, \boldsymbol{\sigma}_s)) := \boldsymbol{\sigma}_s\left(\frac{f_c - \mu(f_c)}{\sigma(f_c)}\right) + \boldsymbol{\mu}_s$$



Difference

- Generate from the micro encoder
- Just use mean and deviation

METHOD

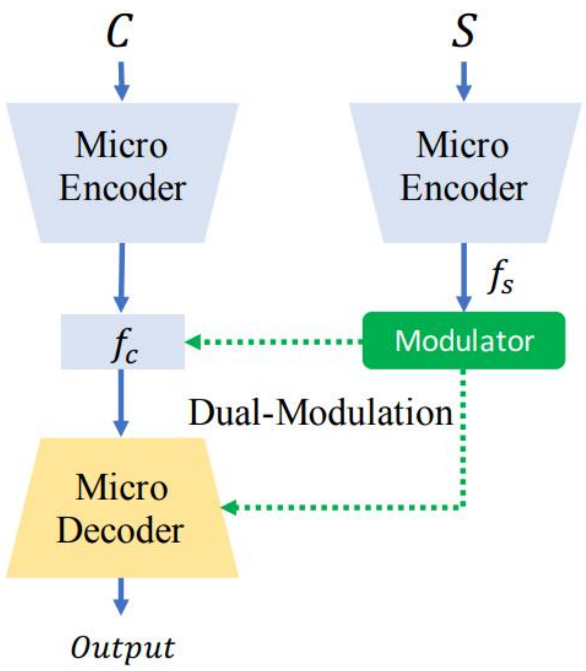
FilterMod

$$\mathbf{w_s} := \xi_w(f_s), \quad \mathbf{b_s} := \xi_b(f_s),$$

$$\begin{aligned}
&FilterMod(D, (\mathbf{w_s}, \mathbf{b_s})) \\
&:= ResBlock(f_c, (\mathbf{w_s}, \mathbf{b_s})) \\
&:= Conv(Relu(Conv(f_c, (\mathbf{w_s}, \mathbf{b_s}))), (\mathbf{w_s}, \mathbf{b_s})) + f_c.
\end{aligned}$$

- Use simple subnets to predict weight and bias

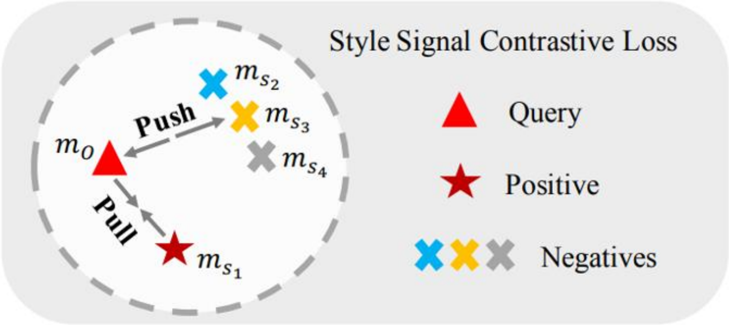
$$\begin{aligned}
Conv(f_c, (\mathbf{w_s}, \mathbf{b_s})) &:= (\mathbf{w_s} * \mathcal{F} + \mathbf{b_s}) \circledast f_c \\
&:= (\mathbf{w_s} * \mathcal{F}) \circledast f_c + \mathbf{b_s} \circledast f_c \\
&:= \mathbf{w_s} * (\mathcal{F} \circledast f_c) + \mathbf{b_s} * f_c,
\end{aligned}$$



METHOD

Contrastive Learning

$$\mathcal{L}_{ssc} := \sum_{i=1}^{N} \frac{\| m_{o_i} - m_{s_i} \|_2}{\sum_{j \neq i}^{N} \| m_{o_i} - m_{s_j} \|_2}.$$



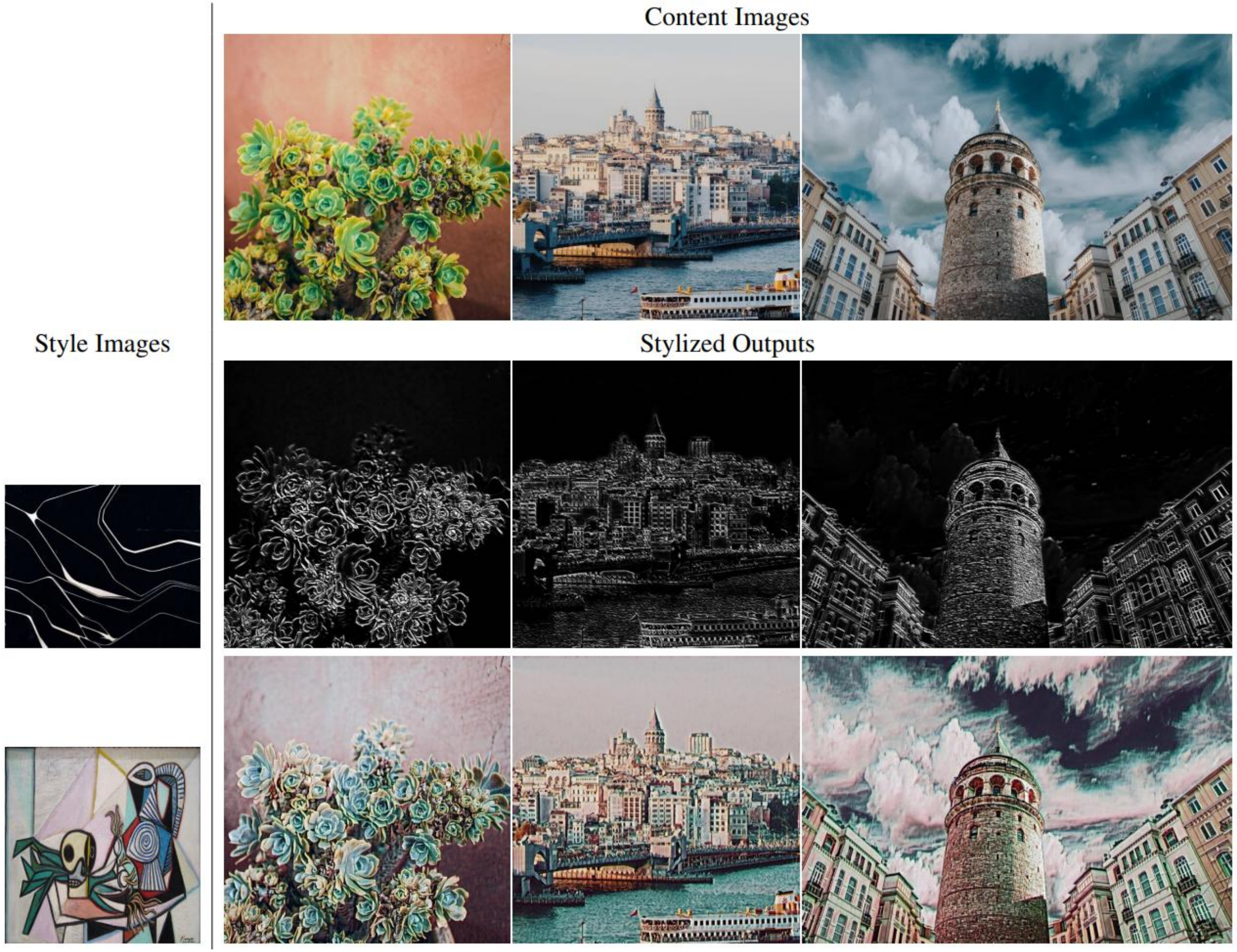
Difference

- Compare output images with style images
- A different form instead of vanilla loss (BCE)

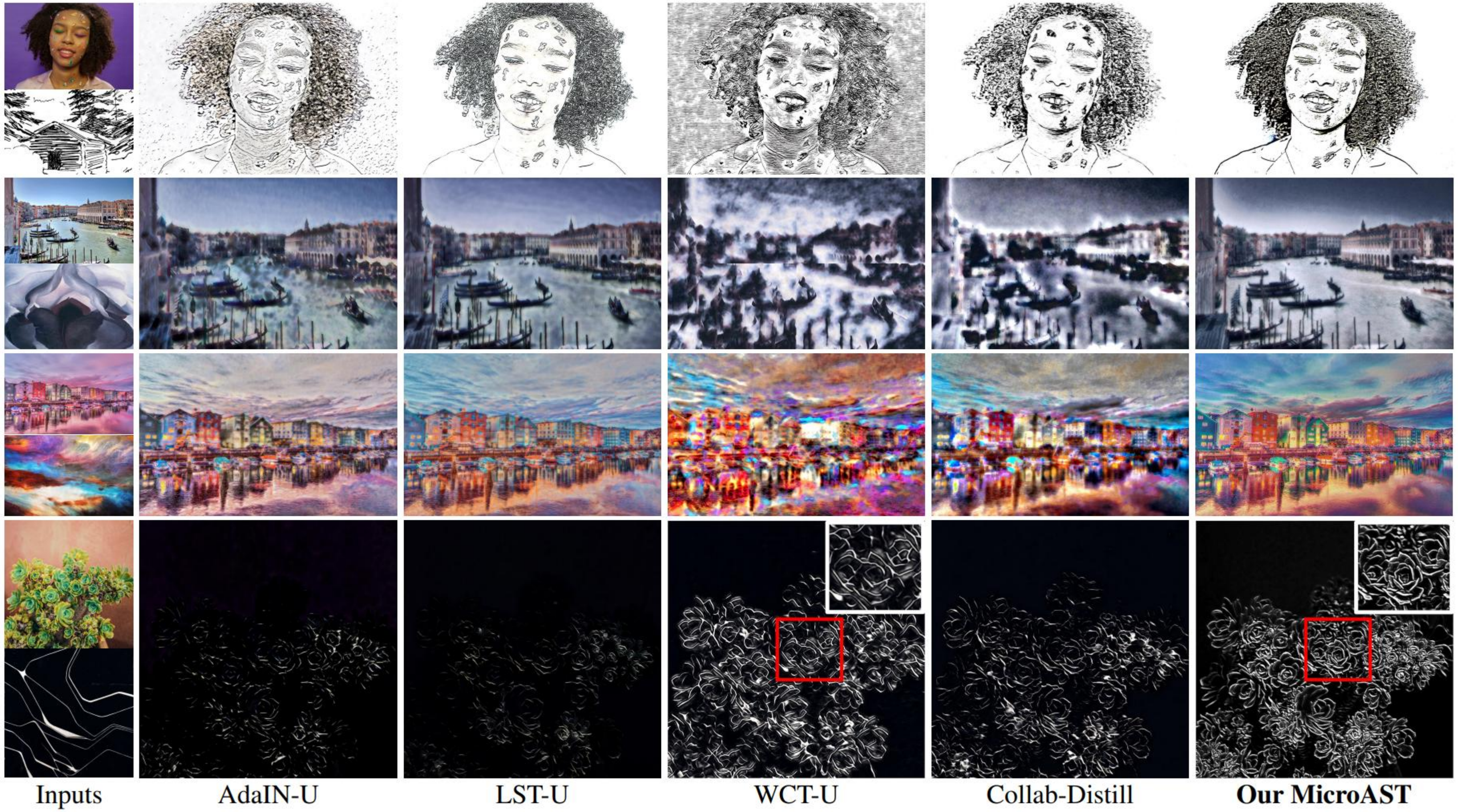
OUTLINE

- Authorship
- Background
- Method
- Experiments
- Conclusion

EXPERIMENTS

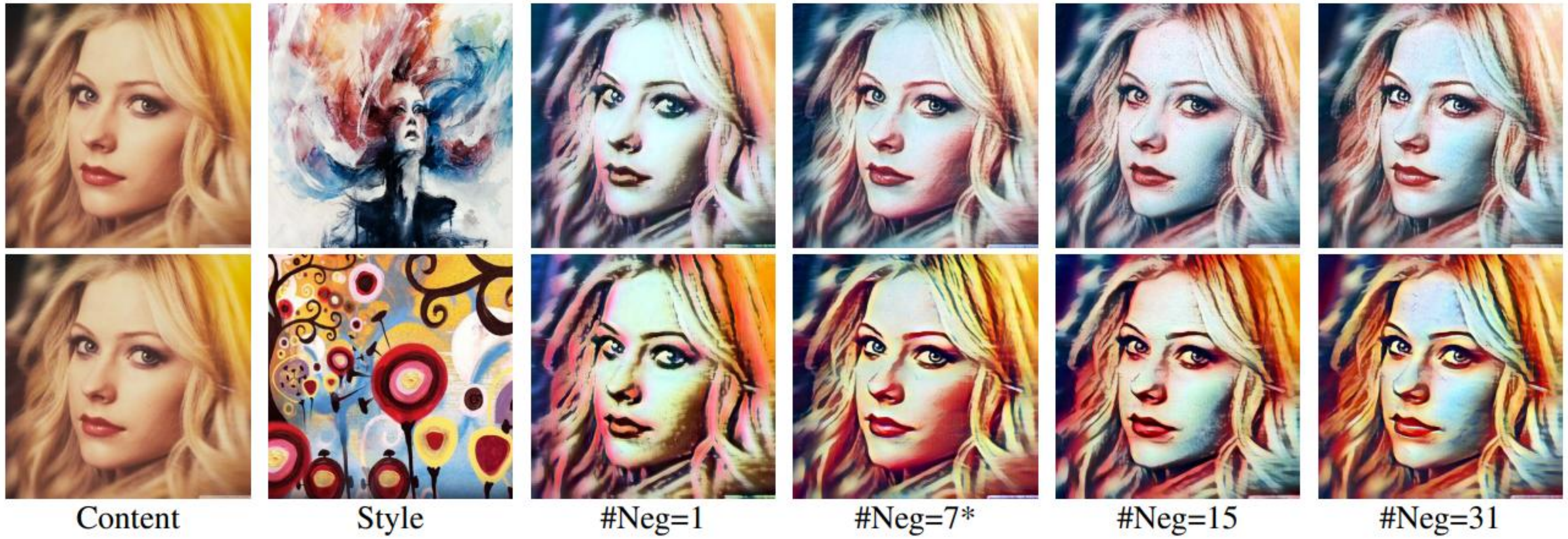


EXPERIMENTS



EXPERIMENTS

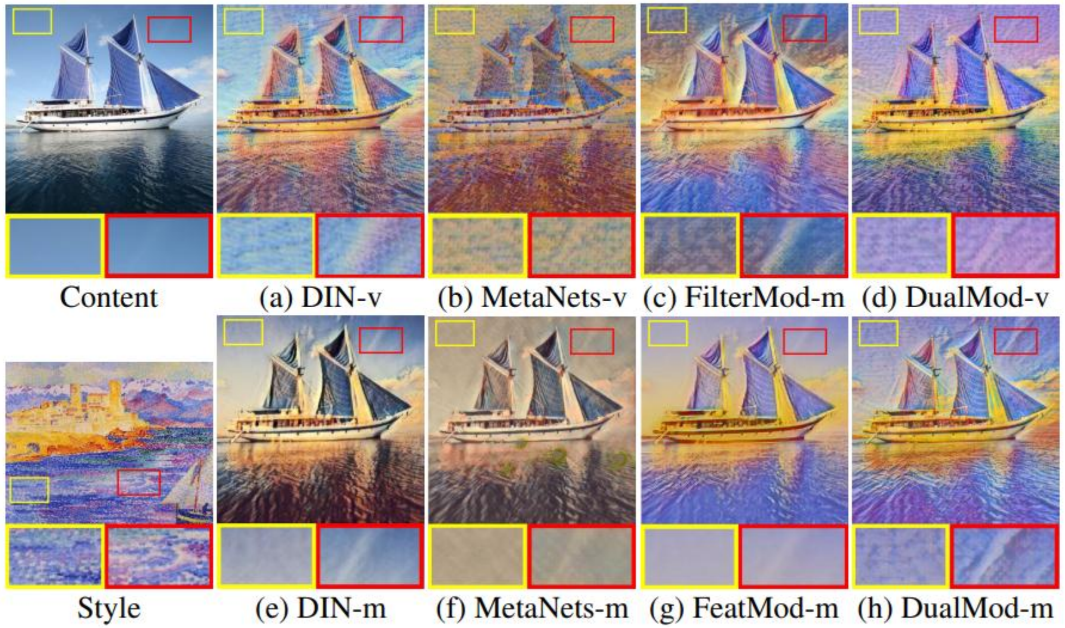
EXPERIMENTS



#Negative	1	7*	15	31
SSIM ↑	0.440	**0.531**	0.504	0.492
Style Loss ↓	2.547	2.342	2.333	**2.332**

EXPERIMENTS

(a) Training Curves	(b) Full Model	(c) w/o $\mathcal{L}_{ssc}$	(d) Single-level
SSIM ↑ / Style Loss ↓ :	0.531 / 2.342	0.482 / 2.897	0.554 / 2.686



OUTLINE

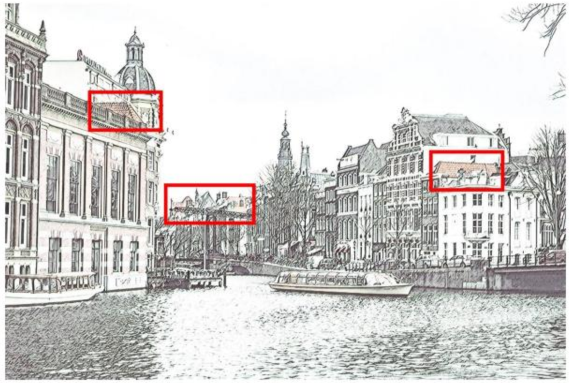
- Authorship
- Background
- Method
- Experiments
- Conclusion

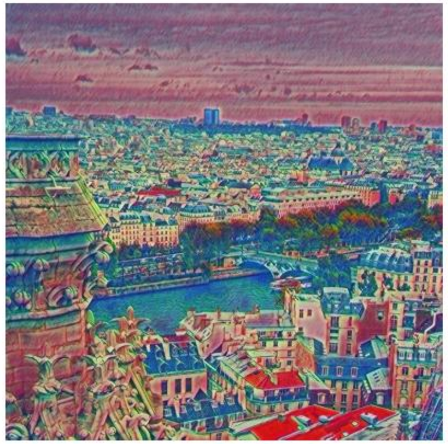
CONCLUSION

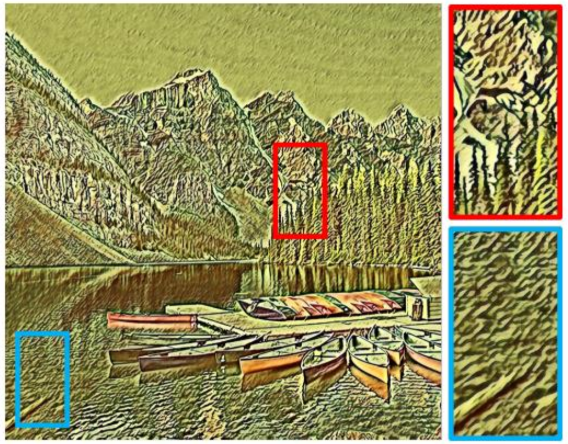
- Propose lightweight MicroAST to achieve super-fast ultra-resolution arbitrary style transfer
- Introduce the dual-modulation strategy
- Introduce a new style signal contrastive loss

LIMITATION

- Under-stylized results
- Fail to transfer complicated texture details
- Fail to deal with small stroke size







Thanks for listening!