Neural Stereoscopic Image Style Transfer

Xinyu Gong, Haozhi Huang, Lin Ma, Fumin Shen, Wei Liu, Tong Zhang

ECCV 2018

Presented by Xutong Ren

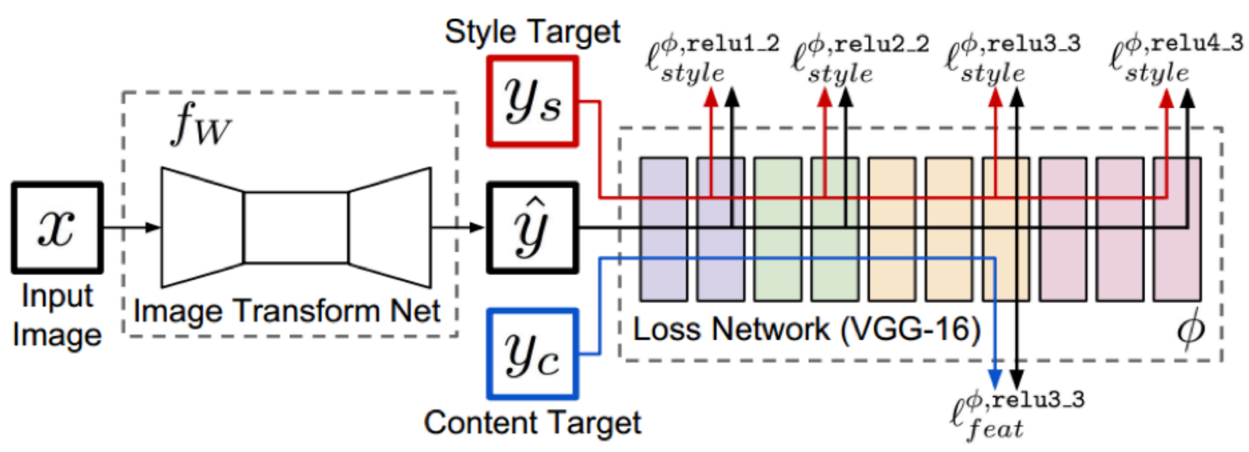
09/28/2018

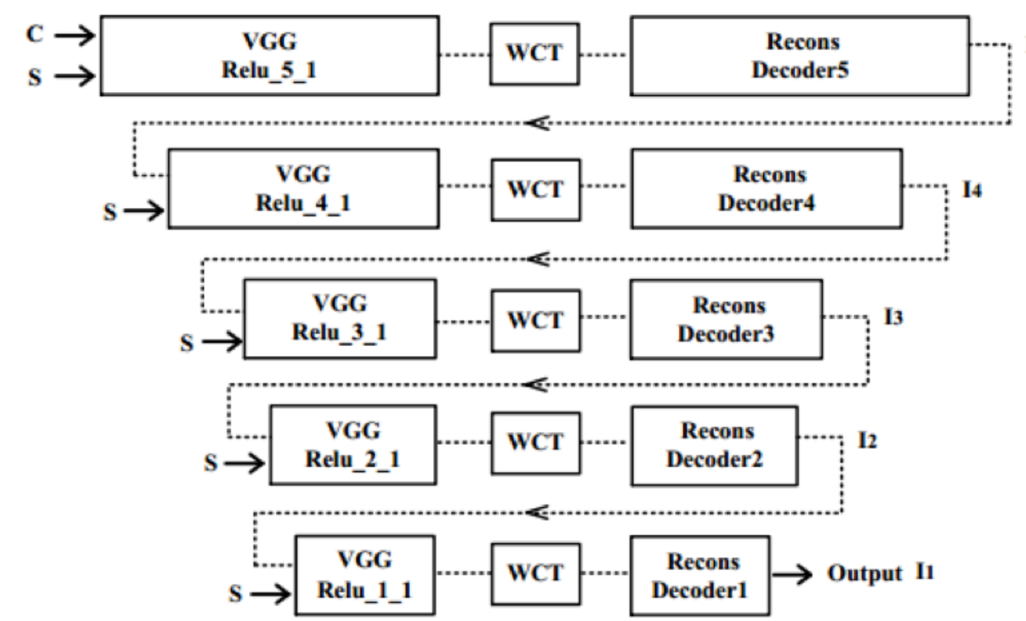
Outline

- Authors
- **Background**
- Proposed Method
- Experiments
- Conclusion

Stereoscopic Stylization

- Intuitively …





Johnson ECCV16

LI NIPS17

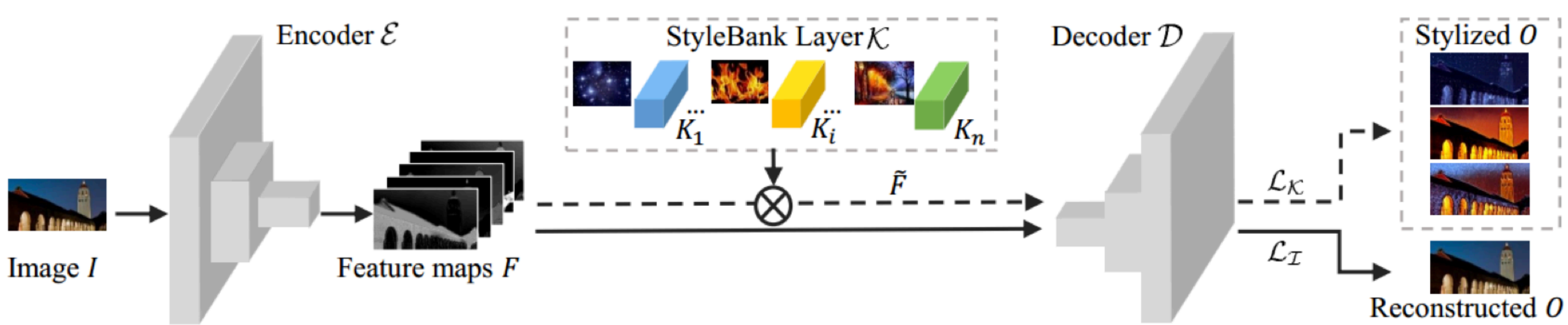


Figure 1. Our network architecture consists of three modules: image encoder $\mathcal{E}$, StyleBank layer $\mathcal{K}$ and image decoder $\mathcal{D}$

Chen CVPR17

Stereoscopic Stylization

- Intuitively …
  - Introduce view inconsistency in the stereo pair.

Stereoscopic Stylization

- Intuitively …
  - Introduce view inconsistency in the stereo pair.
- Wrap
  - Run single-image style transfer on the left view, and then warp it to the right view.

Stavrakis ICIP05

Stereoscopic Stylization

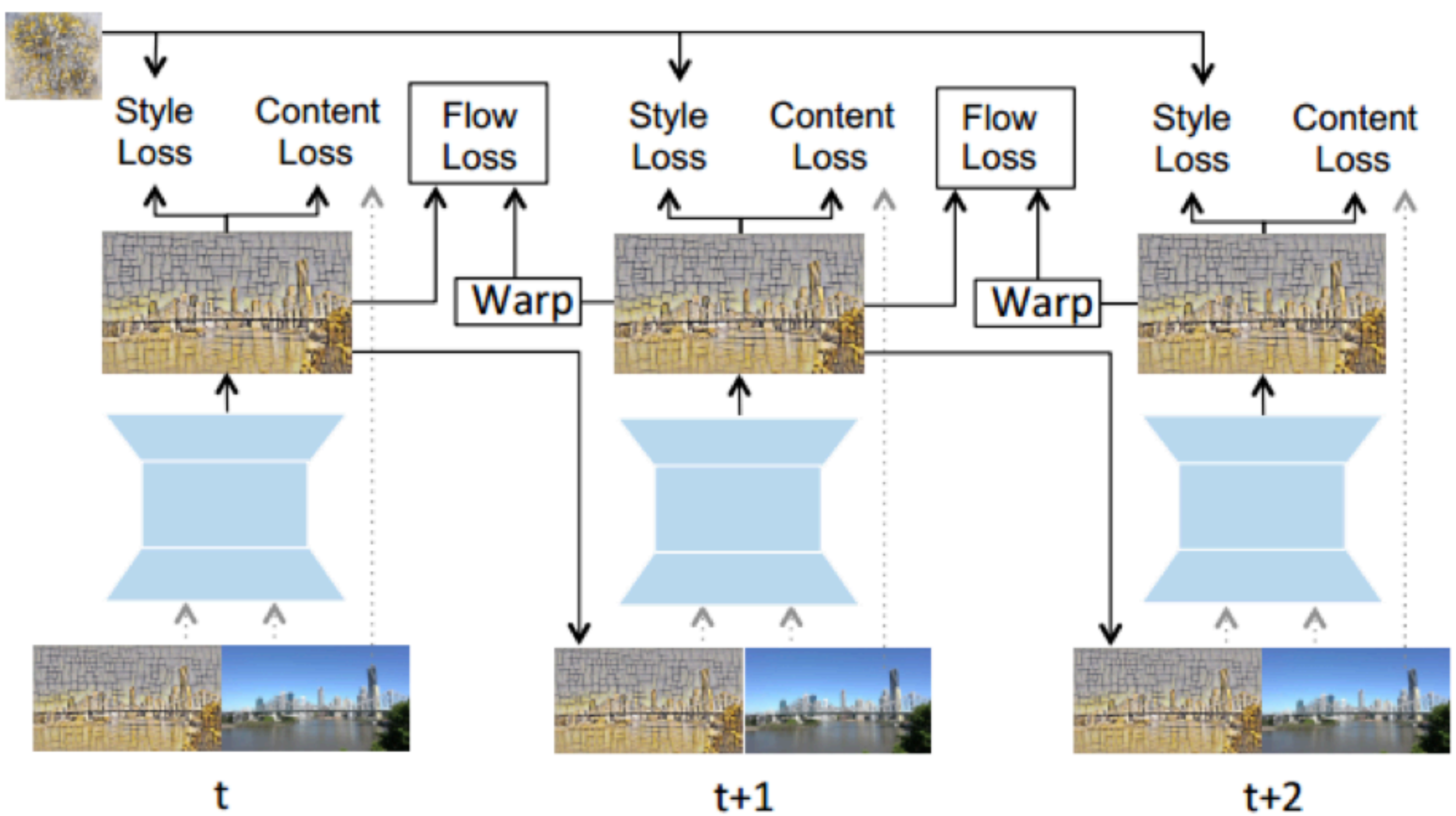
- Intuitively …
  - Introduce view inconsistency in the stereo pair.
- Wrap
  - Run single-image style transfer on the left view, and then warp it to the right view.
  - Introduce black regions due to the occluded regions in a stereo pair.

Stereoscopic Stylization

- Proposal
  - Takes a stereo pair and processes each view in an individual path.
  - Use a feature aggregation block to share feature level information between the two paths.

Video Stylization

Theorem 1. *Let γ be a sphere centered at the origin with radius* $\mathrm{tr}(\Phi_s \Phi_s^T)^{\frac{1}{2}}$. *Then, Φ_p minimizes the objective* $J(\Phi_p) = \|\Phi_p \Phi_p^T - \Phi_s \Phi_s^T\|_F^2$ *iff* $\Phi_p \in \gamma$.

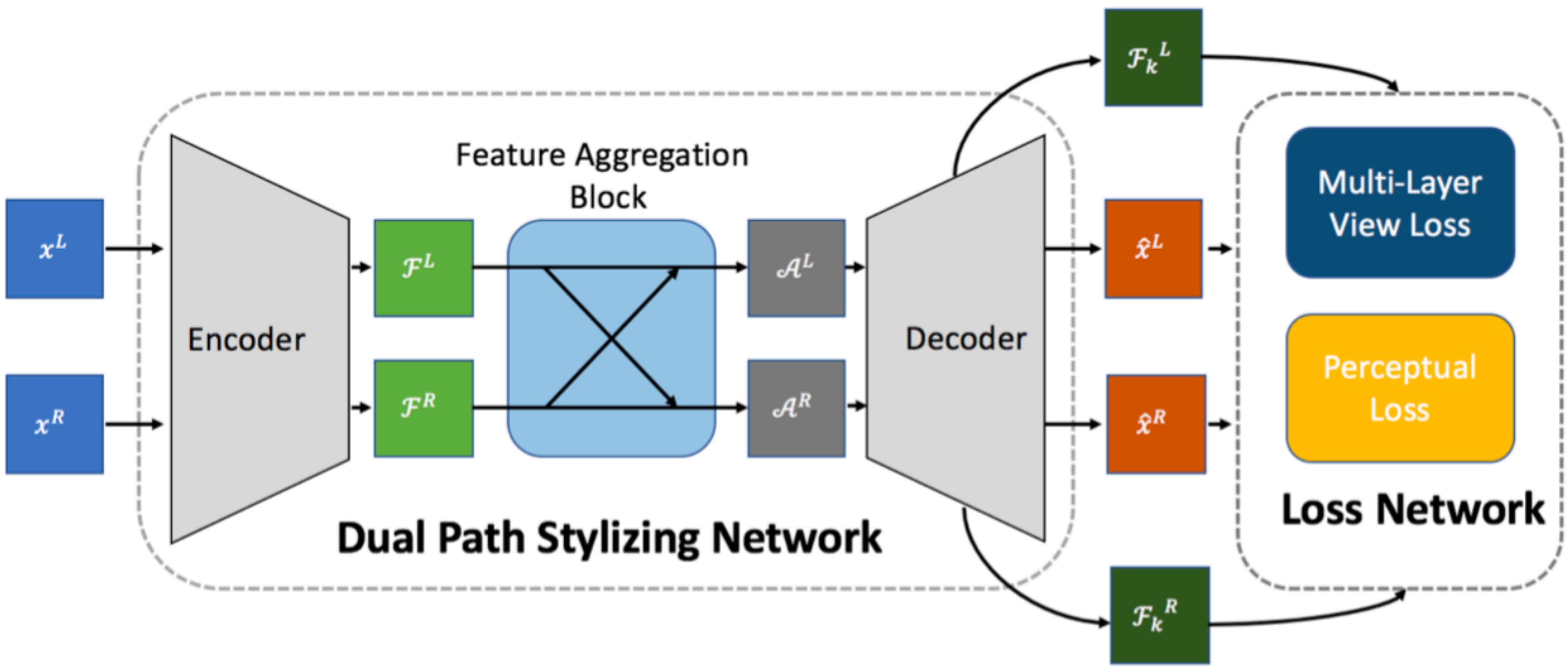


Gupta ICCV17

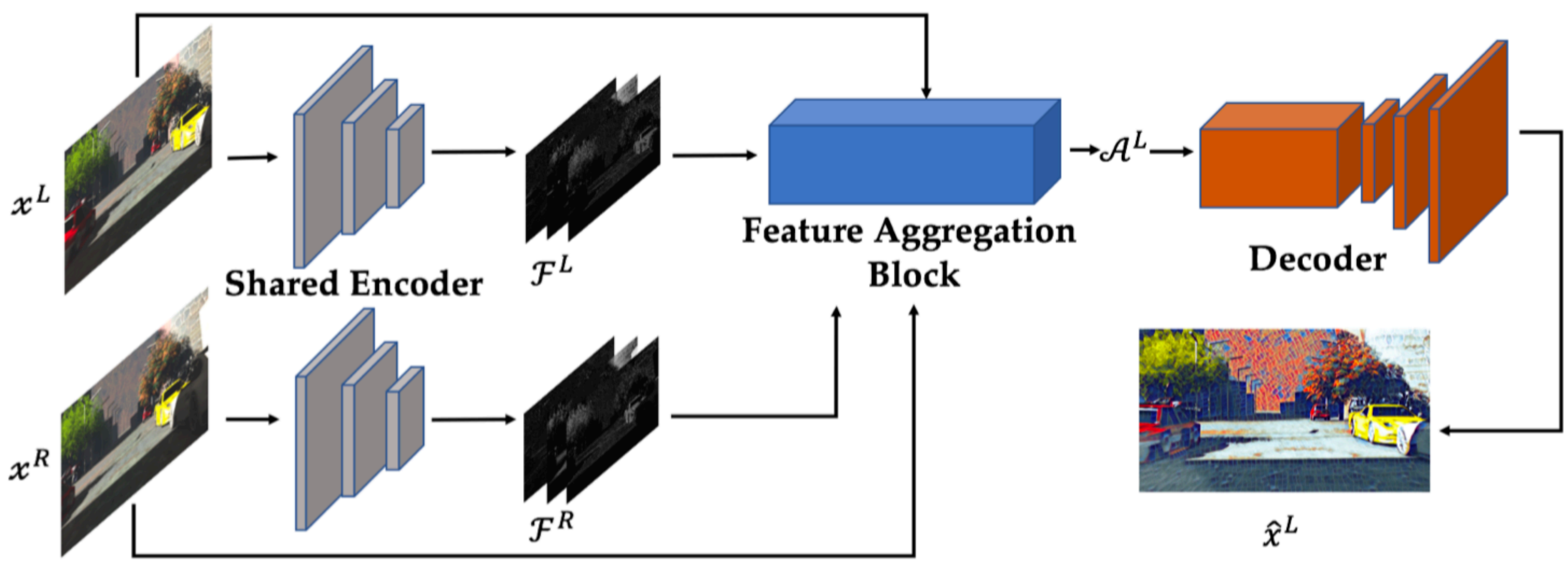
Outline

- Authors
- Background
- **Proposed Method**
- Experiments
- Conclusion

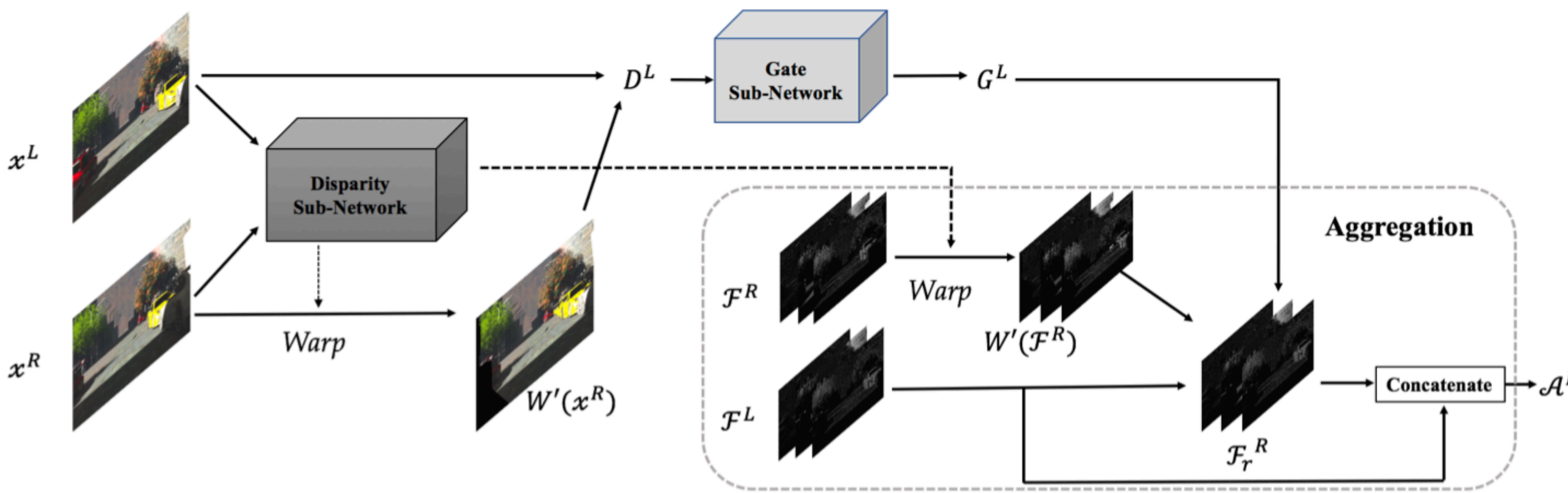
Dual Path Stylizing Network



Dual Path Stylizing Network



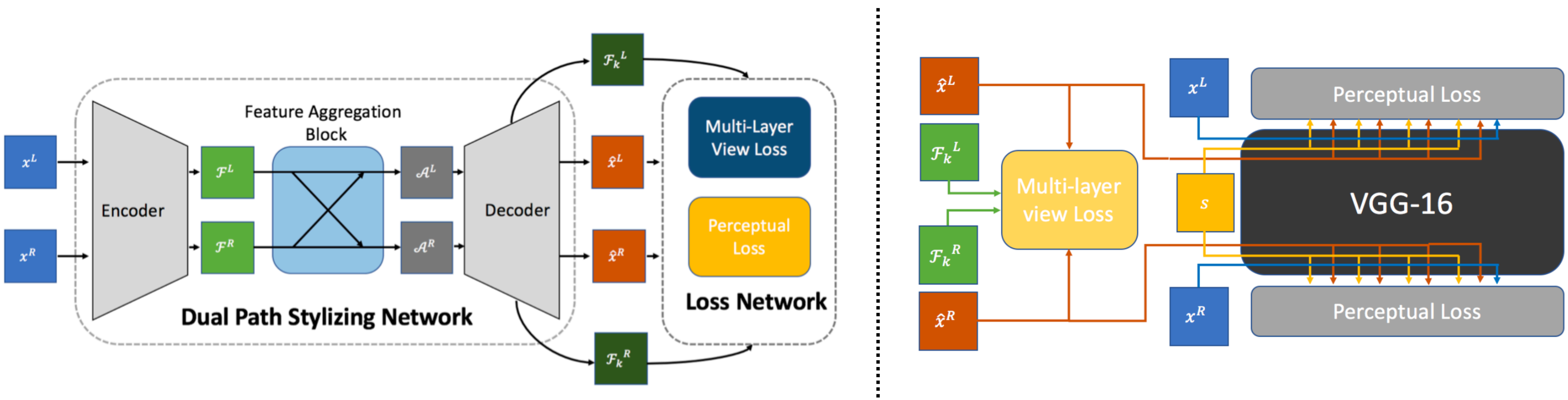
Feature Aggregation Block



$$\mathcal{F}_r^R = W'(\mathcal{F}^R) \odot G^L + \mathcal{F}^L \odot (1 - G^L),$$

Loss Network

$$\mathcal{L}_{\text{total}} = \sum_{d \in \{L,R\}} \mathcal{L}_{\text{perceptual}}(s, x^d, \widehat{x}^d) + \lambda \mathcal{L}_{\text{view}}(\widehat{x}^L, \widehat{x}^R, \mathcal{F}_k^L, \mathcal{F}_k^R),$$



Loss Network

$$\mathcal{L}_{\text{total}} = \sum_{d \in \{L,R\}} \mathcal{L}_{\text{perceptual}}(s, x^d, \widehat{x}^d) + \lambda \mathcal{L}_{\text{view}}(\widehat{x}^L, \widehat{x}^R, \mathcal{F}_k^L, \mathcal{F}_k^R),$$

$$\mathcal{L}_{\text{content}}(x^d, \widehat{x}^d) = \sum_l \frac{1}{H^l W^l C^l} \left\| \mathcal{F}^l(x^d) - \mathcal{F}^l(\widehat{x}^d) \right\|_2^2,$$

$$G_{ij}^l(x^d) = \frac{1}{H^l W^l} \sum_h^{H^l} \sum_w^{W^l} \mathcal{F}^l(x^d)_{h,w,i} \mathcal{F}^l(x^d)_{h,w,j},$$

$$\mathcal{L}_{\text{style}}(s, \widehat{x}^d) = \sum_l {\frac{1}{C^l}}^2 \left\| G^l(s) - G^l(\widehat{x}^d) \right\|_2^2.$$

Loss Network

$$\mathcal{L}_{\text{total}} = \sum_{d \in \{L,R\}} \mathcal{L}_{\text{perceptual}}(s, x^d, \widehat{x}^d) + \lambda \mathcal{L}_{\text{view}}(\widehat{x}^L, \widehat{x}^R, \mathcal{F}_k^L, \mathcal{F}_k^R),$$

$$\begin{aligned}\mathcal{L}_{\text{view}}^{\text{img}} = & \frac{1}{\sum_{i,j} M_{i,j}^L} \left\| M^L \odot (\widehat{x}^L - W(\widehat{x}^R)) \right\|_2^2 \\ & + \frac{1}{\sum_{i,j} M_{i,j}^R} \left\| M^R \odot (\widehat{x}^R - W(\widehat{x}^L)) \right\|_2^2,\end{aligned}$$

$$\begin{aligned}\mathcal{L}_{\text{view}}^{\text{feat}} = & \frac{1}{\sum_{i,j} m_{i,j}^L} \left\| m^L \odot [\mathcal{F}_k^L - W(\mathcal{F}_k^R)] \right\|_2^2 \\ & + \frac{1}{\sum_{i,j} m_{i,j}^R} \left\| m^R \odot [\mathcal{F}_k^R - W(\mathcal{F}_k^L)] \right\|_2^2,\end{aligned}$$

Outline

- Authors
- Background
- Proposed Method
- **Experiments**
- Conclusion

Model Configuration

Table 1. Model configuration.

Layer	Kernel	Stride	C_{in}	C_{out}	Acitivation
Encoder					
Conv	3×3	1	3	16	ReLU
Conv	3×3	2	16	32	ReLU
Conv	3×3	2	32	48	ReLU
Decoder					
Conv	3×3	1	96	96	ReLU
Conv	3×3	1	96	48	ReLU
Res × 5			48	48	ReLU
Deconv	3×3	0.5	48	32	ReLU
Deconv	3×3	0.5	32	16	ReLU
Conv	3×3	1	16	3	tanh

Layer	Kernel	Stride	C_{in}	C_{out}	Acitivation
Disparity Sub-network					
Conv	3×3	1	6	32	ReLU
Conv	3×3	2	32	64	ReLU
Conv	3×3	2	64	48	ReLU
Res × 5			48	48	ReLU
Deconv	3×3	0.5	48	24	ReLU
Deconv	3×3	0.5	24	8	ReLU
Conv	3×3	1	8	3	ReLU
Conv	3×3	1	3	1	-
Gate Sub-network					
Conv	3×3	1	3	6	ReLU
Conv	1×1	1	6	12	ReLU
Conv	1×1	1	12	6	ReLU
Conv	1×1	1	6	3	ReLU
Conv	1×1	1	3	1	tanh


Fish
Mosaic
Candy
Dream

Ablation Study

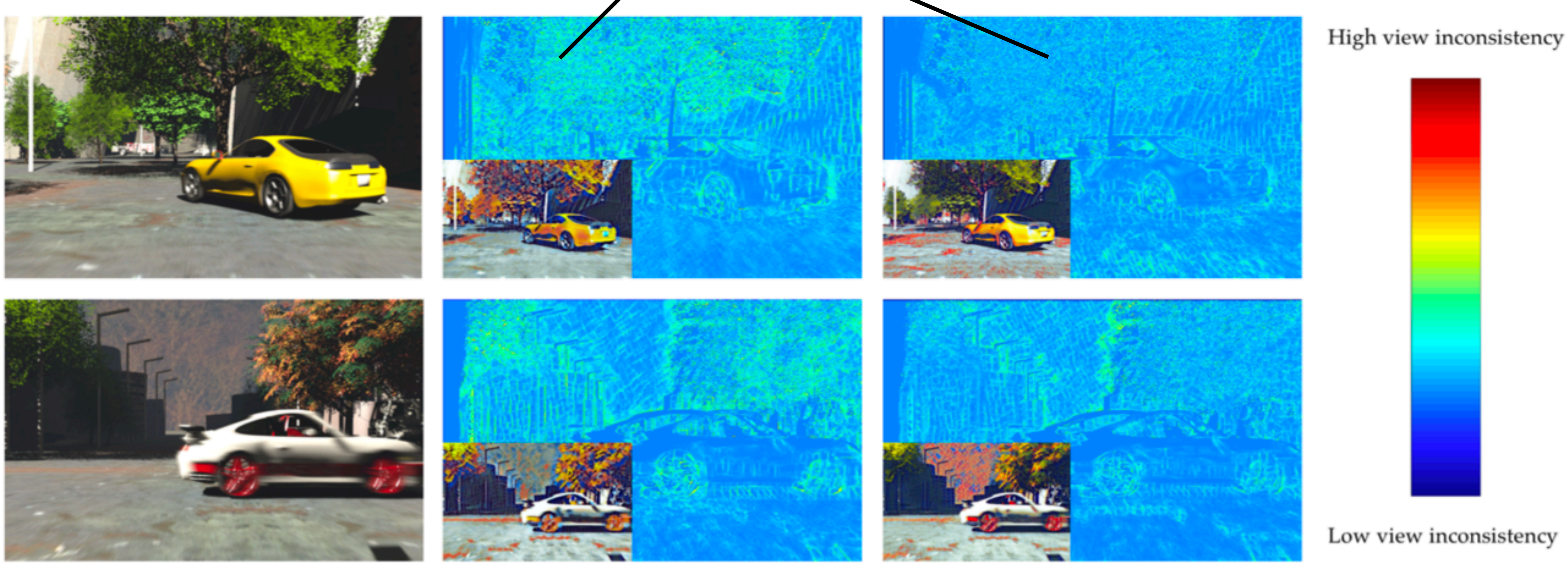
Model	*SingleImage*	*SingleImage-IV*	*Stereo-FA-IV*	*Stereo-FA-dp-IV*	*Stereo-FA-MV*
MSL	426	424	410	**407**	417
MVL	2033	1121	1028	1022	**1014**
MCL	**424153**	485089	481056	478413	445336

Model	*SingleImage-IV*	*CON-IV*	*W-G-CON-IV*
MSL	424	**328**	410
MVL	1121	1068	**1028**
MCL	485089	489555	**481056**

Comparison

Table 3. User preferences.

Style	Prefer ours	Prefer Johnson *et al.*'s	Equal
Candy	**143**	29	38
Fish	**166**	14	30
Mosaic	**152**	24	34




Style Image
Left view
Zoom-in area
Right view

Conclusion

- Natural Stereoscopic Stylization
- Detailed Network
- Different Experiments

- Lack of Visual Comparisons
- Complicated Framework