MG-MLP: Multi-gated MLP for Restoring Images from Spatially Variant Degradations

Jaihyun Koh[1], Jaihyun Lew[2], Jangho Lee[3] and Sungroh Yoon[2]

Samsung Display[1], Data Science and AI Laboratory, Seoul National University[2], Incheon National University[3]

1 Introduction and Background

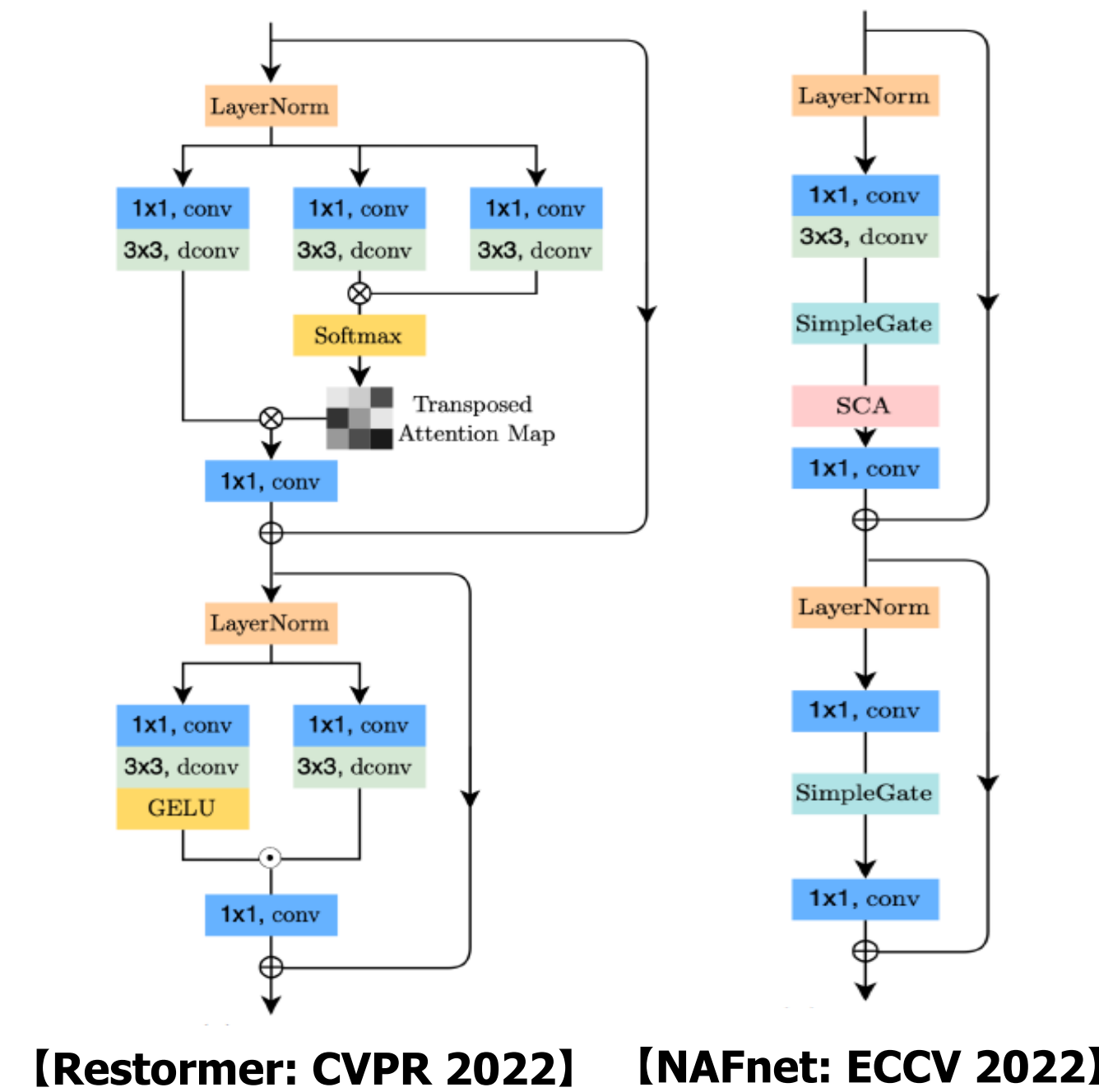- **Image Restoration Task**
  - Low-level semantic, the relationship between pixels rather than abstract.
  - Sufficient receptive field is essential to search for similar patterns.
  - Architecture should be "fully convolutional": Not fixed resolution.
- **Restormer (CVPR 2022)**
  - Channel-wise multi-head self-attention with quadratic complexities.
  - Leveraging 3rd order interaction of transformer architectures.
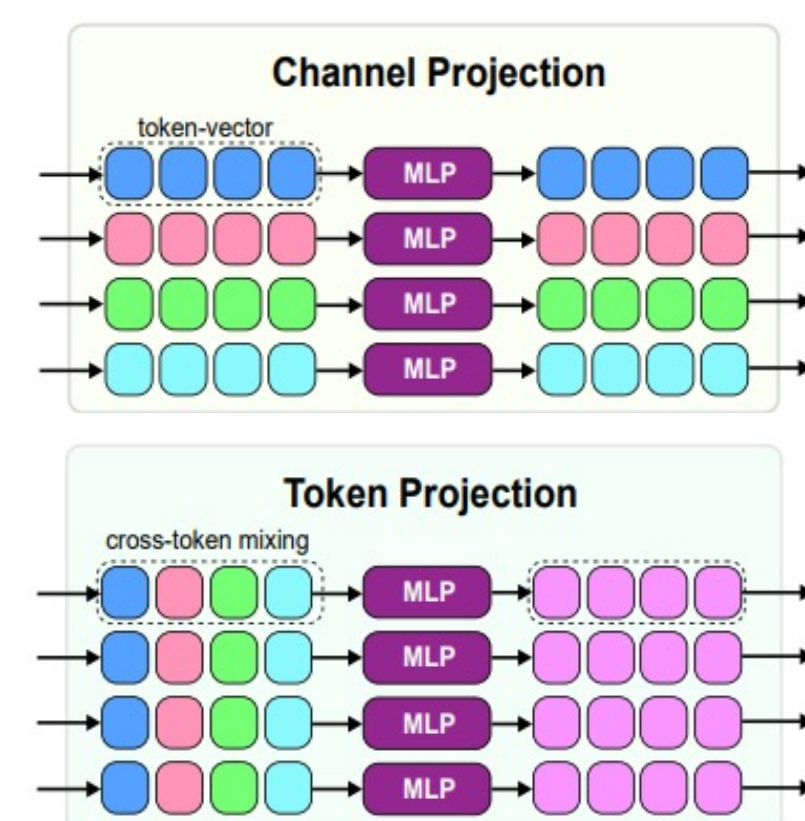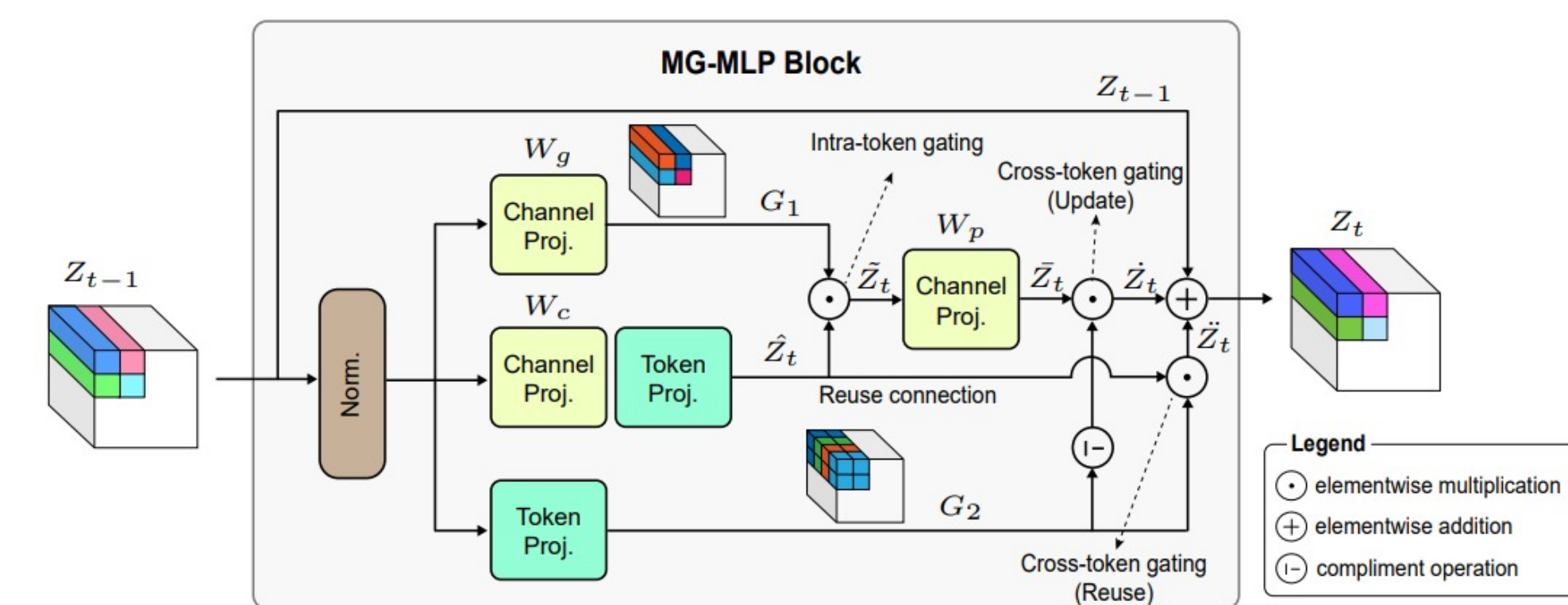 - Invariant architecture to the resolution of an input image.
- **NAFnet (ECCV 2022)**
  - MLP architecture using 1x1 conv. and 3x3 depth-wise conv.
  - Gating operation using element-wise multiplication.
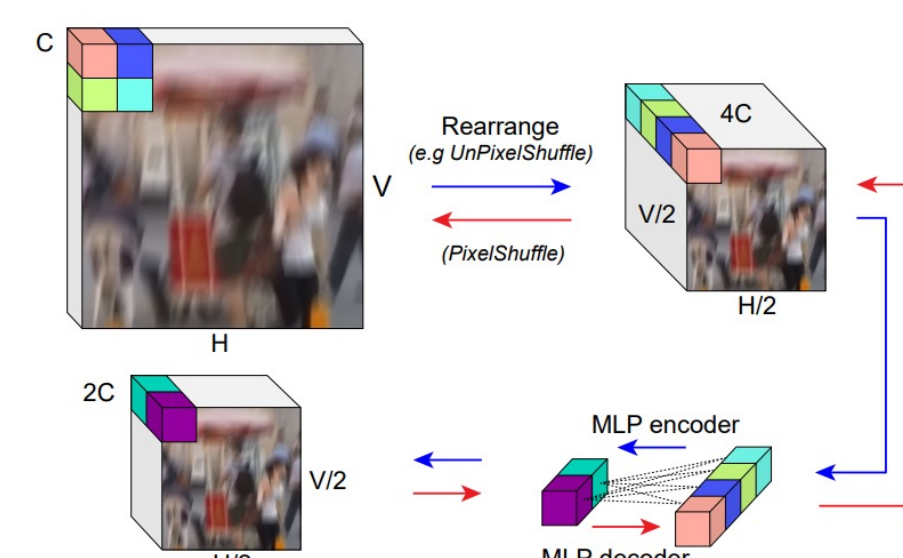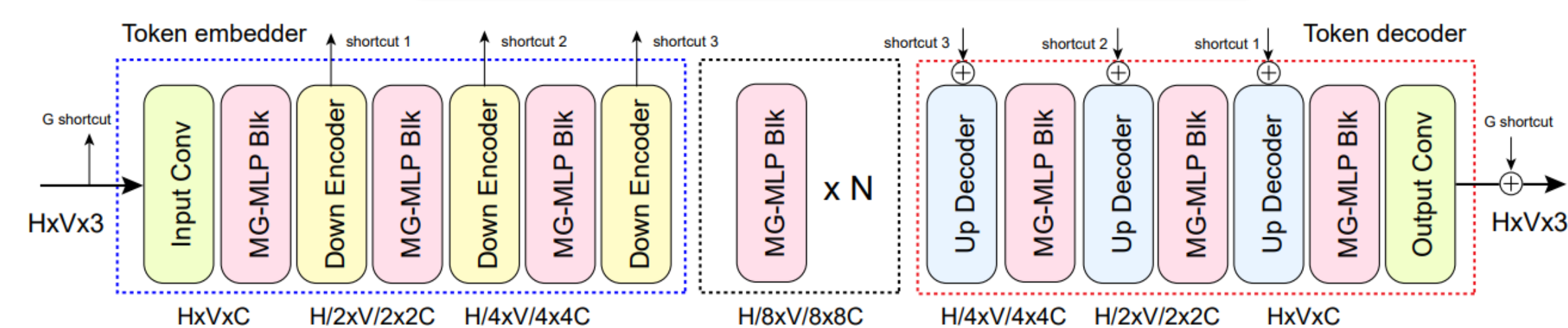 - Channel mixing (1x1) and token mixing (3x3) are critical components.

[Restormer: CVPR 2022]   [NAFnet: ECCV 2022]

2 Proposed MG-MLP Block

- **Intra–token gating**: controls the flow of information through the interaction of data in each token.
- **Cross–token gating**: updates the resulting token from intra–token gating and simultaneously brings back the data discarded by intra–token gating by referring to the adjacent tokens.

Macro Architecture

4 Benefit / Experiments and Results

- **More Flexible Information Flow control**
  - Update gate, reuse gate inspired by LSTM, GRU.
  - Input and receptive field adaptive gating
- **3rd order interaction like ViT**
  - ViT: inner product of query, key and value
  - MG-MLP: multiplication of three projection of input
- **Multi-modal Gaussian Mixture**
  - By two gating unit

$$p(z_t|z_{t-1}) = p(z_t, g_1|z_{t-1}) + p(z_t, g_2|z_{t-1})$$
$$= p(g_1|z_{t-1})p(z_t|z_{t-1}, g_1)$$
$$+ p(g_2|z_{t-1})p(z_t|z_{t-1}, g_2).$$

Complexity and Performance

Models	Computational Budget					GoPro [30]		REDS [31]		RealBlur-J [32]		DVD [35]	
	C	N	MACs	Params	Mems	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Restormer [53]	64	16	75.9G	54.3M	3.6G	31.96	0.9527	28.51	0.8570	28.87	0.9094	31.28	0.9367
NAFNet [6]	64	28	59.0G	54.5M	2.2G	32.34	0.9558	28.65	0.8588	28.99	0.9089	31.48	0.9393
MG-MLP (ours)	64	56	54.4G	46.7M	2.1G	**32.87**	**0.9604**	**28.80**	**0.8613**	**29.14**	**0.9110**	**31.59**	**0.9409**

Generalization

RealBlur-J trained	GoPro [30]		HIDE [33]		RealBlur-R [32]		REDS [31]		DVD [35]		Throughput (img/s)
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	
Restormer [53]	24.29	0.8468	23.58	0.8197	35.02	0.9298	25.01	0.7700	25.79	0.8687	0.575
NAFNet [6]	24.17	0.8578	23.31	0.8241	35.31	0.9309	24.51	**0.7757**	25.20	0.8783	0.690
MG-MLP (ours)	**24.54**	**0.8596**	**23.95**	**0.8298**	**35.46**	**0.9337**	**25.16**	0.7745	**26.20**	**0.8854**	**0.787**

GoPro Dataset

Input   Restormer   NAFnet   MG-MLP (ours)   Ground truth

Real Blur-J and DVD Dataset

Input   Restormer   NAFnet   MG-MLP (ours)   Ground truth
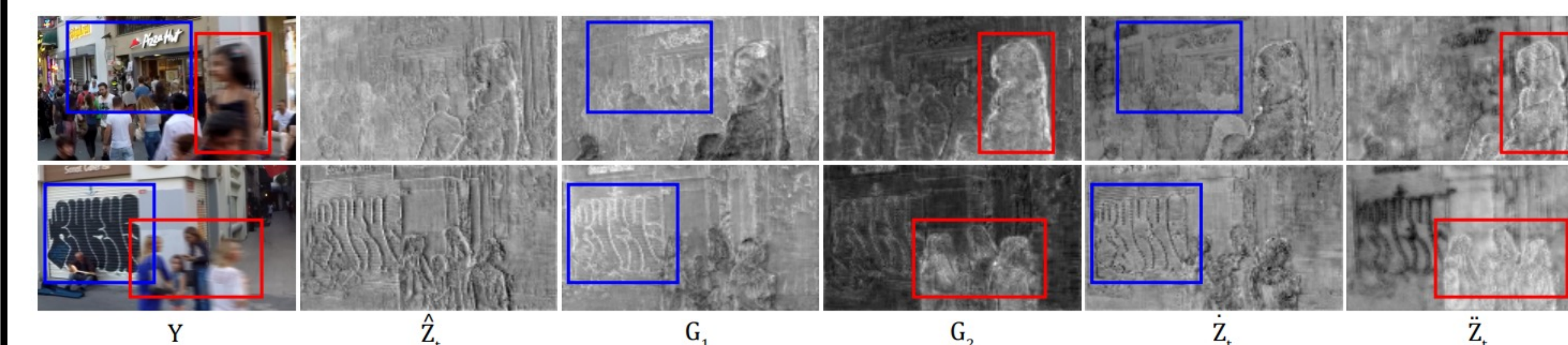
5 Feature Visualization and Ablation

- The two residuals refined by the two gating mechanisms contain complementary features. The two gating paths perform distinct feature refinements by dividing a region into more blurry and less blurry regions.

Feature Visualization

Y $\hat{Z}_t$ G_1 G_2 $\tilde{Z}_t$ $\dot{Z}_t$

Ablation

	Ablation (C = 32)			GoPro testset [30]	
G_1	G_2	Reuse	Update	PSNR	SSIM
Channel	N/A			31.54	0.9465
Channel	Token	✓		31.09	0.9433
Channel	Token		✓	31.30	0.9452
Token	Channel	✓	✓	31.36	0.9459
Channel	Token	✓	✓	31.76	0.9500