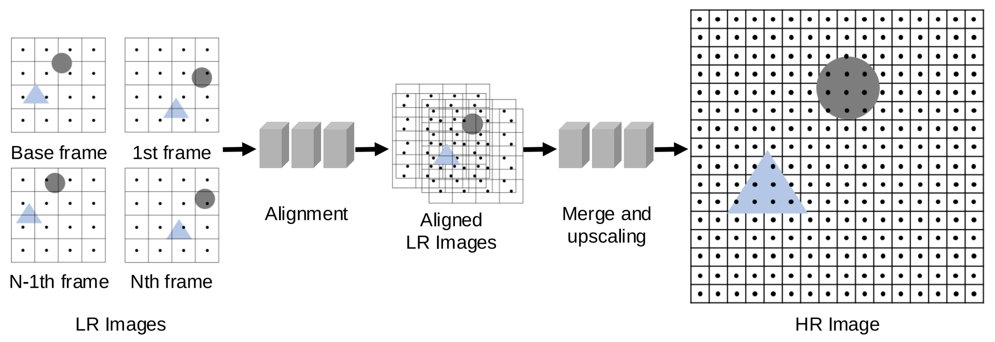
BurstM: Deep Burst Multi-scale SR using Fourier Space with Optical Flow

EungGu Kang[1] Byeonghun Lee[2] Sunghoon Im[1] Kyong Hwan Jin[2]
DGIST, Korea[1] Korea university, Korea[2]

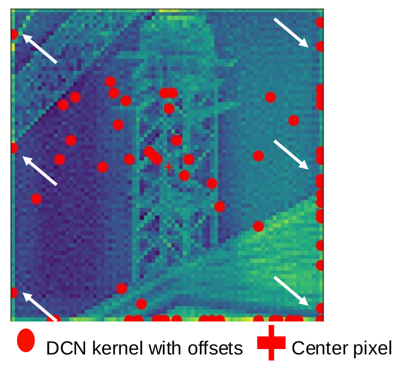
Introduction

Burst Super-Resolution

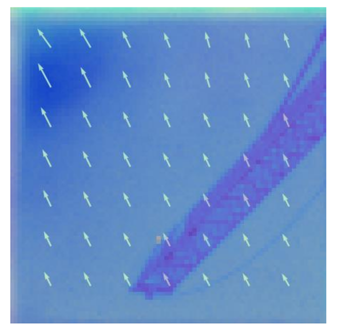
Burst super-resolution aims to restores high-resolution image from burst low-resolution images.



Motivation 1 – Limitation of Deformable Convolution Network

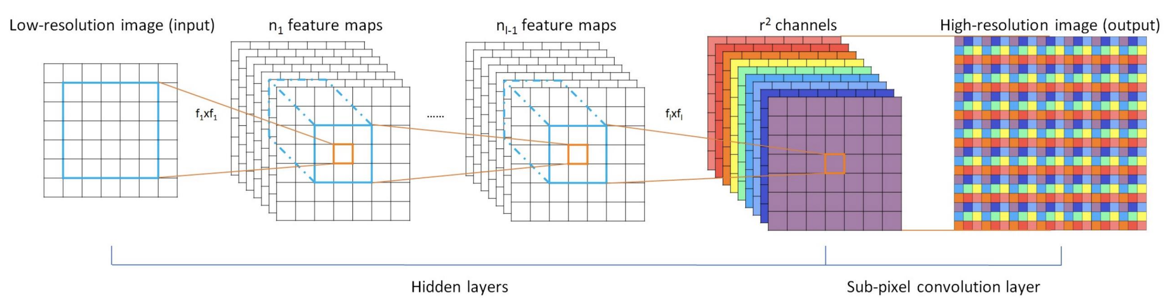


Deformable Convolution Network (DCN)
- Can't cover global alignment.
- Must use pre-defined number of kernels **regardless of pixel correlation.**

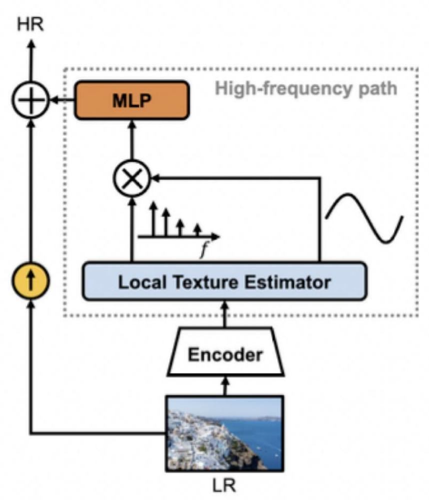


Optical Flow based warping
- Can cover global alignment.
- Based on the pixel correlation, **use only the necessary information**

Motivation 2 – Inflexible Super Resolution Scales



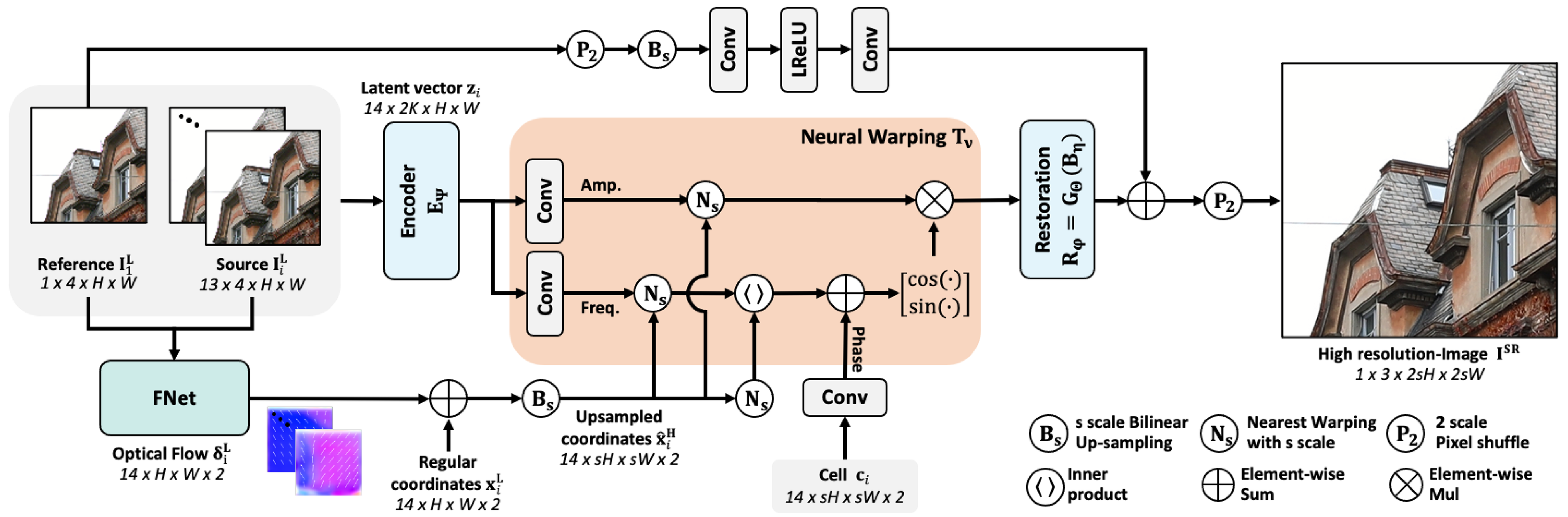
Recent researches utilize Pixel Shuffle method as up-sampler. Pixel Shuffle can reduce computational efficiency, but **Pixel Shuffle limits scale factors of up-sampling.**



To address fixed scale factors, using **Implicit Neural Representation (INR)**, which represents images in a continuous domain, is the best approach.

However, an MLP with ReLUs tends to be biased towards learning low-frequency components. We addressed this issue by **estimating Fourier information**, motivated by LTE.

Method



Burst Super-Resolution

Given a multiple low-resolution images $\mathbf{I}_i^{\mathbf{L}} \in \mathbb{R}^{N\times4\times H\times W}$, restore high-resolution image $\mathbf{I}^{\mathbf{SR}} \in \mathbb{R}^{1\times3\times sH\times sW}$.

$$\mathbf{I}^{\mathbf{SR}} = \mathbf{BurstM}(\mathbf{I}_i^{\mathbf{L}}), \qquad i = 1, \ldots, N$$

Optical flow estimation

The coordinates $\hat{\mathbf{x}}_i^{\mathbf{H}} \in \mathbb{R}^{sH\times sW\times2}$ are converted from estimated optical flow offsets with s scale bilinear interpolation.

$$\hat{\mathbf{x}}_i^{\mathbf{H}} = \mathbf{B_s}\left(FNet(\mathbf{I}_i^{\mathbf{L}}) + \mathbf{x}_i^{\mathbf{L}}\right), \qquad i = 1, \ldots, N$$

where $FNet(\cdot)$ is optical flow estimator, $\mathbf{B_s}$ is bilinear interpolation with s scale

Neural warping

Warping multiple LR images by estimated coordinates $\hat{\mathbf{x}}_i^{\mathbf{H}}$ on **Fourier space.**

$$\mathbf{F}_i = \mathbf{N_s}\left(\mathbf{g_f}(\mathbf{z}_i), \hat{\mathbf{x}}_i^{\mathbf{H}}\right) \text{ : Frequency estimator}$$

$$\mathbf{A}_i = \mathbf{N_s}\left(\mathbf{g_a}(\mathbf{z}_i), \hat{\mathbf{x}}_i^{\mathbf{H}}\right) \text{ : Amplitude estimator}$$

$$\mathbf{ph}_i = \mathbf{g_p}(\mathbf{c}_i) \text{ : Phase estimator}$$

$$\mathbf{m}_i = \mathbf{A}_i \odot \begin{bmatrix} \cos(\pi(\langle \mathbf{F}_i, \boldsymbol{\delta}_i \rangle + \mathbf{ph}_i)) \\ \sin(\pi(\langle \mathbf{F}_i, \boldsymbol{\delta}_i \rangle + \mathbf{ph}_i)) \end{bmatrix}, \quad i = 1, \ldots, N,$$

where $\mathbf{z}_i = \mathbf{E}_\psi(\mathbf{I}_i^{\mathbf{L}})$, $\boldsymbol{\delta}_i = \hat{\mathbf{x}}_i^{\mathbf{H}} - \mathbf{x}_i^{\mathbf{H}}$

Reconstruction & Skip connection

Reconstruct from multiple LR images to HR image with skip connection.

$$\mathbf{h} = \mathbf{R}_\varphi(\{\mathbf{m}_i\}_{i=1}^N) = \mathbf{G}_\Theta\left(\mathbf{B}_\eta(\{\mathbf{m}_i\}_{i=1}^N)\right)$$

$$\mathbf{I}^{\mathbf{SR}} = \mathbf{P}_2\left(\mathbf{h} + \mathbf{Q}_\xi(\mathbf{I}_1^{\mathbf{L}})\right)$$

where $\mathbf{Q}_\xi(\cdot)$ is skip connection network, $\mathbf{P}_2$ is x2 pixel shuffle

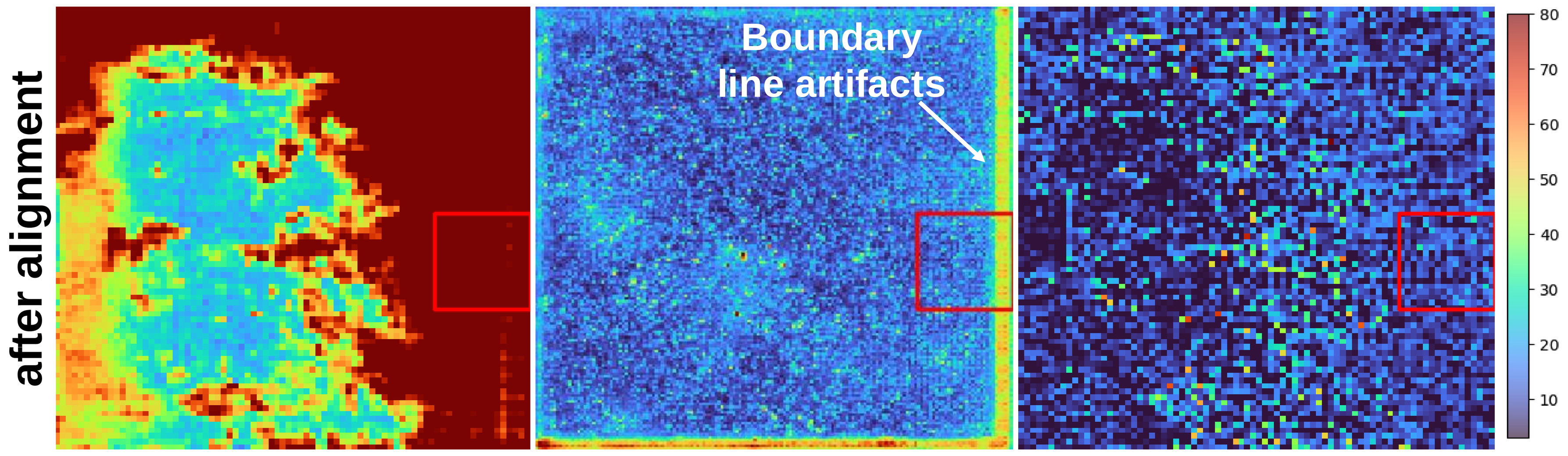
Results

Quantitative comparison

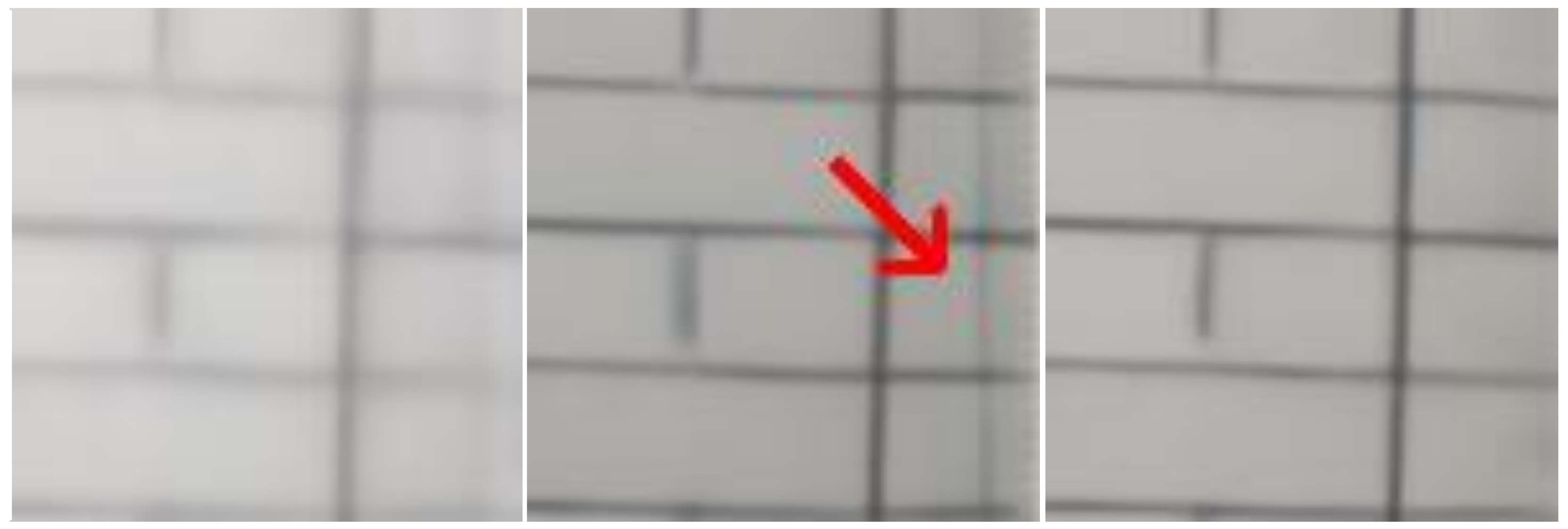
	# Params. M	SyntheticBurst x2		SyntheticBurst x3		SyntheticBurst x4		BurstSR x4	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Bicubic	-	38.30	0.948	33.94	0.886	33.02	0.862	42.55	0.962
DBSR [3]	13.01	40.51	0.965	40.11	0.959	40.76	0.959	48.05	0.984
MFIR [4]	12.13	41.25	0.971	41.81	0.972	41.56	0.964	48.33	0.985
BIPNet† [11]	6.7	37.58	0.928	40.83	0.955	41.93	0.960	48.49	0.985
GMTNet† [31]	-	-	-	-	-	42.36	0.961	48.95	0.986
Burstormer† [12]	2.5	37.06	0.925	40.26	0.953	42.83	0.973	48.82	0.986
BSRT-Small† [30]	4.92	40.64	0.966	42.30	0.975	42.72	0.971	48.48	0.985
BSRT-Large† [30]	20.71	40.33	0.965	42.87	0.979	43.62	0.975	48.57	0.986
BurstM (Ours)	14.0	46.01	0.985	44.79	0.982	42.87	0.973	49.12	0.987

Qualitative comparison (BurstSR - Real world data x4)

Residual map after alignment



Estimated image



Burstormer (DCN & transformer-based method) | BSRT-L | **BurstM (Ours)**

Inference time comparison

Scale : x4	BIPNet†	Burstormer†	BSRT-S†	BSRT-L†	BurstM
Params. (M)	6.7	**2.5**	4.92	20.71	14.9
Runtime (ms)	20.8	20.8	148.0	266.1	**11.6**
PSNR (dB)	48.49	48.82	48.48	48.57	**49.12**

† indicates Transformer-based method.

Conclusion

The proposed BurstM overcomes both quality and computational complexity for real-world datasets by utilizing INR and Optical flow.

EUROPEAN CONFERENCE ON COMPUTER VISION
MILANO 2024