Learning Local Implicit Fourier Representation for Image Warping

DGIST 대구경북과학기술원 Daegu Gyeongbuk Institute of Science & Technology
\{IPL\}

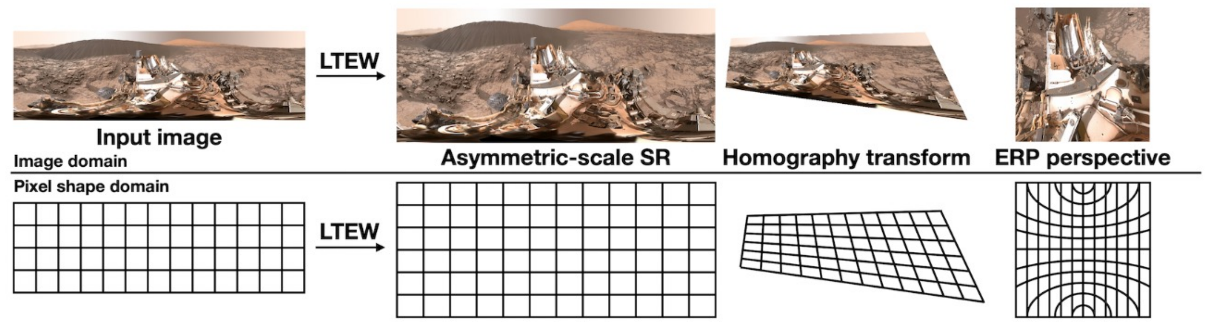
Jaewon Lee[1] Kwang Pyo Choi[2] Kyong Hwan Jin[1]
Image Processing Laboratory (IPL), DGIST, Korea[1]
Samsung Electronics, Korea[2]

Paper Code

Introduction

Image warping

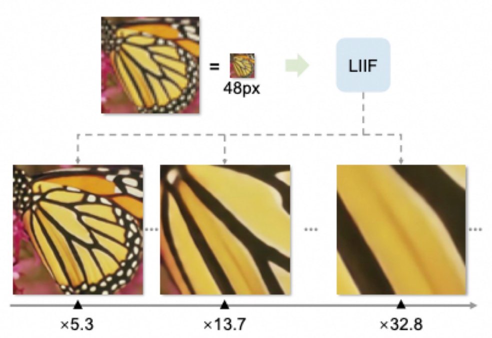
Image warping aims to reshape images defined on **rectangular grids** into **arbitrary shapes**



Motivation 1 – interpolation-based approach

Limited **generalization** into a large-scale representation (out of training range)

Motivation 2 – INR-based approach

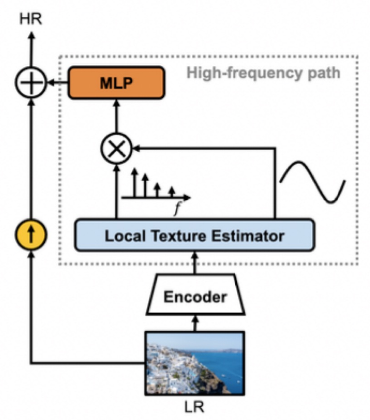


\+ Good generalization

\- Spectral bias

\+ Overcome spectral bias

\- Spatial-invariant function



Local texture estimator for image warping

Method	SRWarp (21')	LIIF (21')	LTE (22')	LTEW (ours)
Spatially-varying SR	✓	✗	✗	✓
Generalization	✗	✓	✓	✓
High-frequency detail	✓	✗	✓	✓

Method

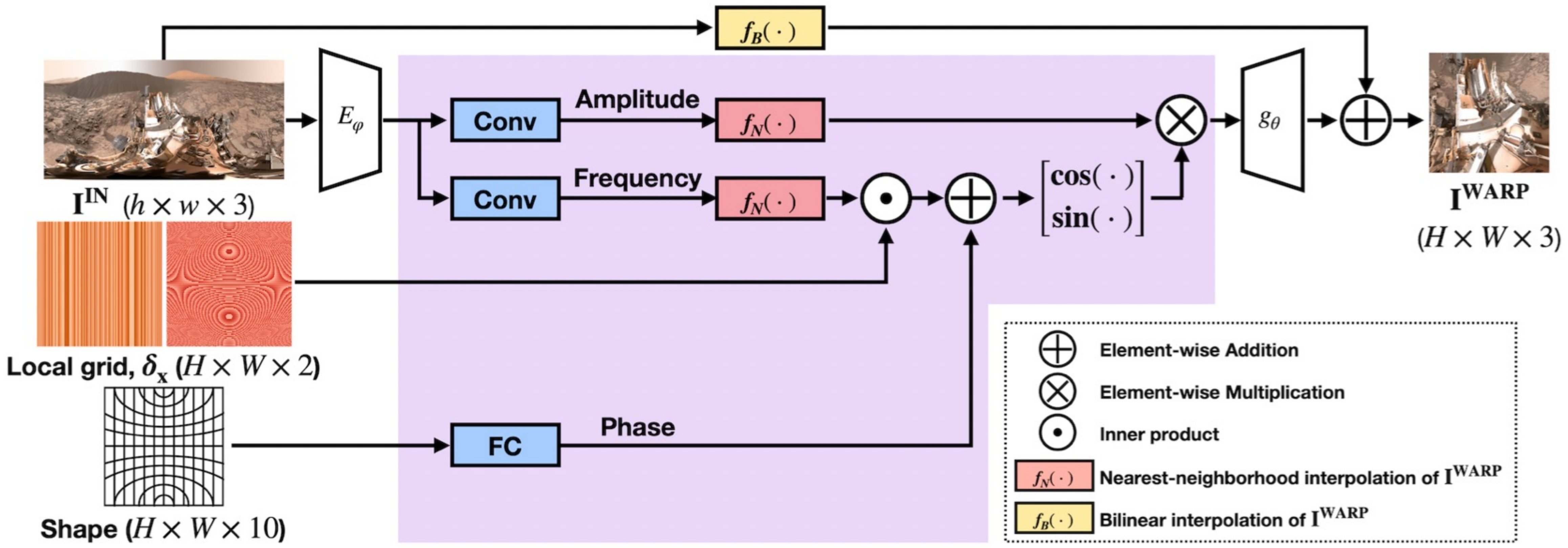


Image warping

Given an image $\mathbf{I}^{\mathrm{IN}}: \mathbf{X} \mapsto \mathbb{R}^3$, and a differentiable and invertible coordinate transformation $\boldsymbol{f}: \mathbf{X} \mapsto \mathbf{Y}$

we aim to represent a warped image $\mathbf{I}^{\mathrm{WARP}}: \mathbf{Y} \mapsto \mathbb{R}^3$

Local implicit representation function

$$\mathbf{I}^{\mathrm{WARP}}[\mathbf{y}; \boldsymbol{\theta}, \boldsymbol{\psi}] = \sum_{j \in \mathcal{T}} \boldsymbol{w}_j \boldsymbol{g}_\theta(\boldsymbol{h}_\psi(\mathbf{z}_j, \mathbf{x} - \boldsymbol{f}(\mathbf{x}_j)))$$

where $\mathbf{z} = E_\varphi(\mathbf{I}^{\mathrm{LR}})$

Local texture estimator

$$h_\psi(\mathbf{z}_j, \boldsymbol{\delta}, \mathbf{c}) = \underbrace{\mathbf{A}_j}_{\text{Amplitude}} \odot \begin{pmatrix} \cos\left(\pi\left(\underbrace{\mathbf{F}_j}_{\text{Frequency}}\boldsymbol{\delta} + \underbrace{h_p(\mathbf{c})}_{\text{Phase}}\right)\right) \\ \sin\left(\pi\left(\mathbf{F}_j\boldsymbol{\delta} + h_p(\mathbf{c})\right)\right) \end{pmatrix}$$

However, LTE fails to represent warped images since $\mathbf{F}_j$ is a frequency response of an input image, instead of of a warped image.

Learning Fourier information with coordinate transformations

1. Local grid relationship

$$\begin{aligned} \boldsymbol{\delta}_{\mathbf{y}} &= \mathbf{y} - f(\mathbf{x}_j) = f(\mathbf{x}) - f(\mathbf{x}_j) \\ &= \{f(\mathbf{x}_j) + \mathbf{J}_f(\mathbf{x}_j)(\mathbf{x} - \mathbf{x}_j) + \cancel{\mathcal{O}(\mathbf{x}^2)}\} - f(\mathbf{x}_j) \\ &\simeq \mathbf{J}_f(\mathbf{x}_j)(\mathbf{x} - \mathbf{x}_j) = \mathbf{J}_f(\mathbf{x}_j)\boldsymbol{\delta}_{\mathbf{x}} \end{aligned}$$

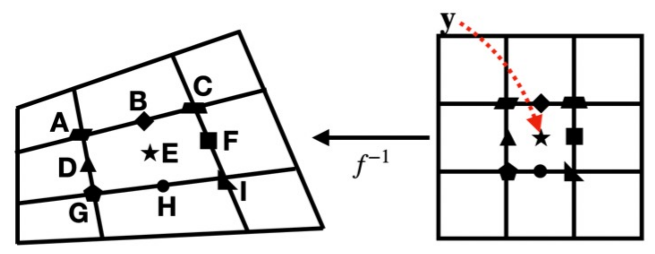
2. Frequency response of a warped image

$$\mathbf{F}'_j \simeq \mathbf{J}_f^{-T}(\mathbf{x}_j)\mathbf{F}_j.$$

3. Redefine the LTE (LTE Warp)

$$\begin{aligned} h_\psi(\mathbf{z}_j, \boldsymbol{\delta}_{\mathbf{y}}, f) &= \mathbf{A}_j \odot \begin{bmatrix} \cos(\pi < \mathbf{F}'_j, \boldsymbol{\delta}_{\mathbf{y}} >) \\ \sin(\pi < \mathbf{F}'_j, \boldsymbol{\delta}_{\mathbf{y}} >) \end{bmatrix} \\ &\simeq \mathbf{A}_j \odot \begin{bmatrix} \cos(\pi < \mathbf{J}_f^{-T}(\mathbf{x}_j)\mathbf{F}_j, \mathbf{J}_f(\mathbf{x}_j)\boldsymbol{\delta}_{\mathbf{x}} >) \\ \sin(\pi < \mathbf{J}_f^{-T}(\mathbf{x}_j)\mathbf{F}_j, \mathbf{J}_f(\mathbf{x}_j)\boldsymbol{\delta}_{\mathbf{x}} >) \end{bmatrix} \\ &= \mathbf{A}_j \odot \begin{bmatrix} \cos(\pi < \mathbf{F}_j, \boldsymbol{\delta}_{\mathbf{x}} >) \\ \sin(\pi < \mathbf{F}_j, \boldsymbol{\delta}_{\mathbf{x}} >) \end{bmatrix} \\ &= h_\psi(\mathbf{z}_j, \boldsymbol{\delta}_{\mathbf{x}}). \end{aligned}$$

Shape-dependent phase estimation



(a) Inverse mapping

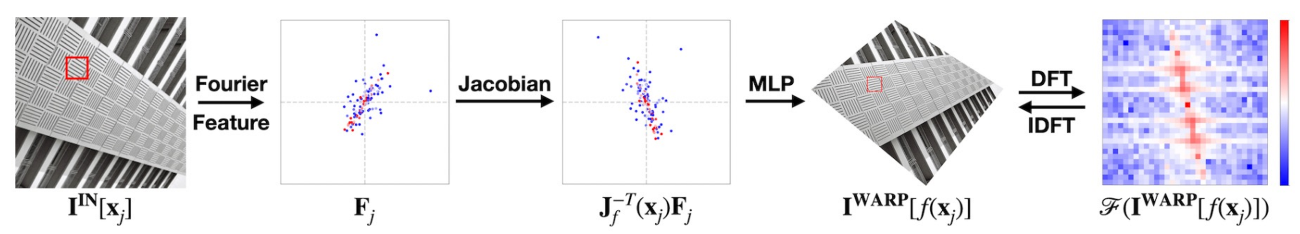
$$\tilde{J}_{f^{-1}}(\mathbf{y}) = \begin{pmatrix} F - D \\ B - H \end{pmatrix}$$

$$\tilde{H}_{f^{-1}}(\mathbf{y}) = \begin{pmatrix} 4(F - 2E + D) & C - A - I + G \\ C - A - I + G & 4(B - 2E + H) \end{pmatrix}$$

(b) Numerical derivatives calculation

$$\boldsymbol{s}(\mathbf{y}) = [\tilde{\mathbf{J}}_{f^{-1}}(\mathbf{y}), \tilde{\mathbf{H}}_{f^{-1}}(\mathbf{y})]$$

Fourier space visualization



Fourier space is transformed by the Jacobian matrix

Results

Quantitative comparison

Asymmetric-scale SR within in-scale (top) and out-of-scale (bottom)

Method	Set5 ×3/×8	Set5 ×3/×7	Set5 ×3.2/×6.1	Set14 ×8/×4	Set14 ×7/×4	Set14 ×7/×3.5	B100 ×8/×2.8	B100 ×3/×6	B100 ×7/×2.9	Urban100 ×3.2/×6	Urban100 ×3.2/×7.6	Urban100 ×7.1/×3.1
Bicubic	25.69	26.35	26.84	24.27	24.62	24.79	24.67	25.58	24.98	22.55	21.92	22.15
RCAN [51]	29.00	30.01	30.46	26.48	26.94	27.11	26.06	27.19	26.47	25.52	24.50	24.84
MetaSR-RCAN [19]	28.75	29.74	30.38	26.32	26.85	27.03	26.07	27.15	26.45	25.50	24.47	24.84
Arb-RCAN [46]	28.37	29.35	30.08	26.06	26.63	26.84	25.91	27.14	26.40	25.36	24.12	24.61
LTEW-RCAN (ours)	29.26	30.16	30.64	26.60	27.06	27.25	26.25	27.28	26.62	25.85	24.79	25.18

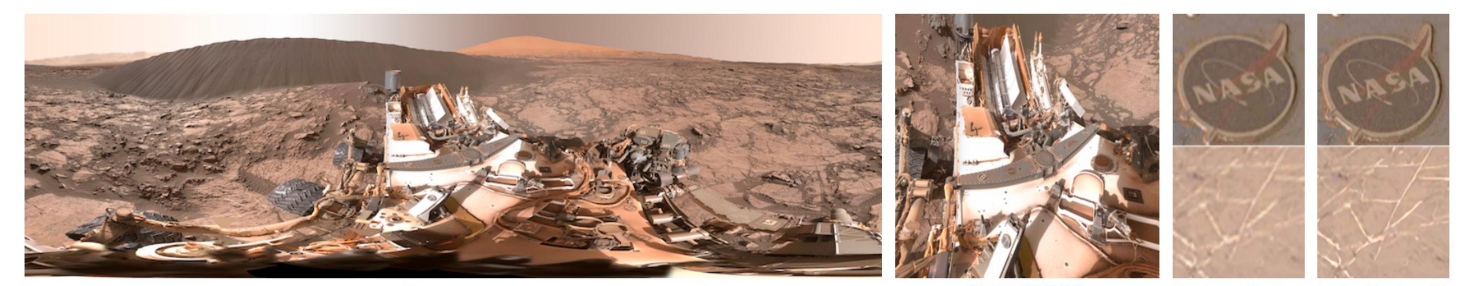
Method	Set5 ×1.5/×4	Set5 ×1.5/×3.5	Set5 ×1.6/×3.05	Set14 ×4/×2	Set14 ×3.5/×2	Set14 ×3.5/×1.75	B100 ×4/×1.4	B100 ×1.5/×3	B100 ×3.5/×1.45	Urban100 ×1.6/×3	Urban100 ×1.6/×3.8	Urban100 ×3.55/×1.55
Bicubic	30.01	30.83	31.40	27.25	27.88	27.27	27.45	28.86	27.94	25.93	24.92	25.19
RCAN [51]	34.14	35.05	35.67	30.35	31.02	31.21	29.35	31.30	29.98	30.72	28.81	29.34
MetaSR-RCAN [19]	34.20	35.17	35.81	30.40	31.05	31.33	29.43	31.26	30.09	30.73	29.03	29.67
Arb-RCAN [46]	34.37	35.40	36.05	30.55	31.27	31.54	29.54	31.40	30.22	31.13	29.36	30.04
LTEW-RCAN (ours)	34.45	35.46	36.12	30.57	31.21	31.55	29.62	31.40	30.24	31.25	29.57	30.21

Homography transformation

Method	DIV2KW *isc*	DIV2KW *osc*	Set5W *isc*	Set5W *osc*	Set14W *isc*	Set14W *osc*	B100W *isc*	B100W *osc*	Urban100W *isc*	Urban100W *osc*
Bicubic	27.85	25.03	35.00	28.75	28.79	24.57	28.67	25.02	24.84	21.89
RRDB [48]	30.76	26.84	37.40	30.34	31.56	25.95	30.29	26.32	28.83	23.94
SRWarp-RRDB [42]	31.04	26.75	37.93	29.90	32.11	25.35	30.48	26.10	29.45	24.04
LTEW-RRDB (ours)	31.10	26.92	38.20	31.07	32.15	26.02	30.56	26.41	29.50	24.25

Qualitative comparison (ERP2Perspective)

ERP Image   Bicubic   RRDB [48]   SRWarp [42]   **LTEW**



ERP Image (8190px × 2529px)   LTEW   Bicubic **LTEW**

Model complexity (x2 SR with 256x256 input)

Method	Training task	#Params.	Runtime	Memory	*in-scale* ×2	*in-scale* ×3	*in-scale* ×4	*out-of-scale* ×6	*out-of-scale* ×8
Arb-RCAN [46]	Asymmetric -scale SR	16.6M	**160ms**	**1.39GB**	**32.39**	**29.32**	27.76	25.74	24.55
LTEW-RCAN (ours)	Asymmetric -scale SR	**15.8M**	283ms	1.77GB	32.36	29.30	**27.78**	**26.01**	**24.95**
SRWarp-RRDB [42]	Homography transform	18.3M	328ms	2.34GB	32.31	29.27	**27.77**	25.33	24.45
LTEW-RRDB (ours)	Homography transform	**17.1M**	**285ms**	**1.79GB**	**32.35**	**29.29**	27.76	**25.98**	**24.95**

Conclusion

LTEW-based neural function : Fourier basis + MLP
➔ Arbitrary transformation with high-frequency details