B-spline Texture Coefficients Estimator for Screen Content Image Super-Resolution

Byeonghyun Pak Jaewon Lee Kyong Hwan Jin
Daegu Gyeongbuk Institute of Science and Technology (DGIST), Korea

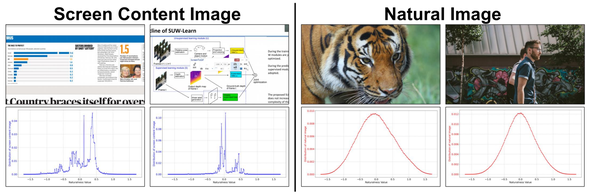
CVPR JUNE 19-24 2022 NEW ORLEANS • LOUISIANA

1. Overview

Problem) Screen Content Image Super-Resolution

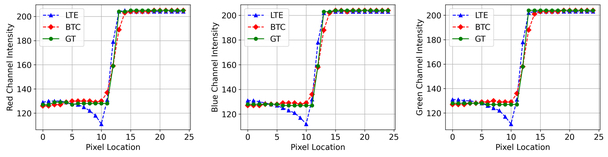
aims to **reconstruct high-resolution screen content image** w/o information distortion on arbitrary magnification. **(SCI SR)**

Motivation 1) Screen content image ↔ Natural image



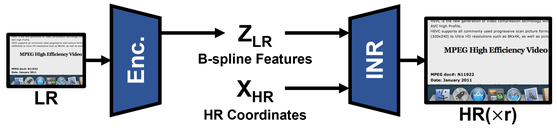
Previous works focused on natural image SR.
We need to **consider the distributional property of SCI.**

Motivation 2) Limitation of Fourier basis for SCI SR



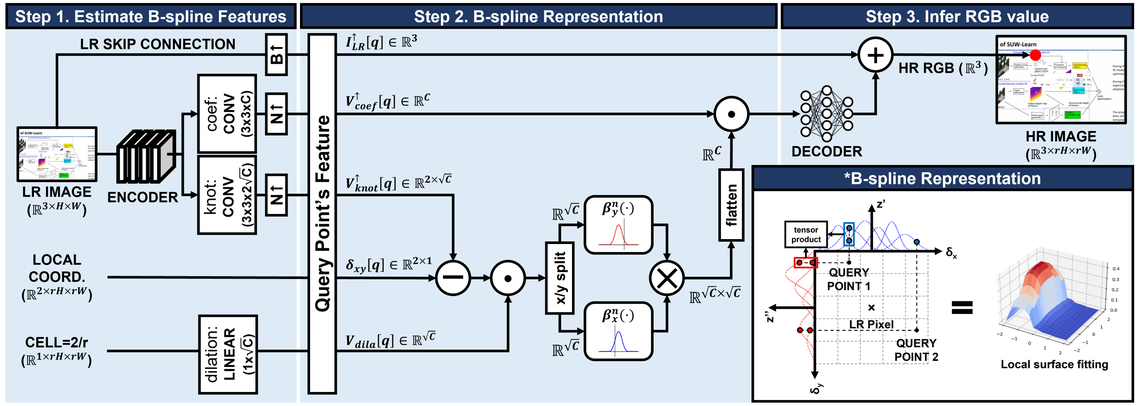
The SR method with Fourier feature (LTE) shows **under/overshooting aliasing at SCI's discontinuities.**

B-spline Texture Coefficients estimator (BTC)



utilizes **implicit neural representation (INR) with B-splines.**
» **performance↑ & aliasing↓ & model complexity↓**

2. Method



▲ ⊕/⊖/⊙ : element-wise addition/subtraction/multiplication, ⊗ : tensor product, B↑/N↑ : Bilinear/Nearest-Neighborhood up

Implicit Neural Function for SCI SR

$$I_{HR}[q;\theta,\psi,\varphi]=I^{\uparrow}_{LR}[q]+\sum_{t\in\mathcal{N}}w_t f_\theta\left(g_\psi(z_t,\delta_t,s)\right);\ z=E_\varphi(I_{LR})$$

- E_φ : Encoder
- f_θ : Decoder
- g_ψ : BTC
- s : Cell value (=2/r)
- δ_t : Local coordinate
- w_t : Local ensemble weight

B-spline Representation of BTC

$$g_\psi(z_t,\delta_t,s)=c_t\odot vec\left[\beta^n\left(\frac{\delta^y_t-k^y_t}{d}\right)\otimes\beta^n\left(\frac{\delta^x_t-k^x_t}{d}\right)^T\right]$$

- c_t : Coefficient feature
- d : Dilation feature
- k_t : Knot feature
- β^n : B-spline basis function

3. Results

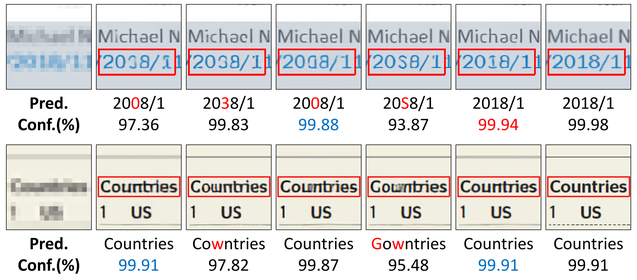
Quantitative comparison

Test set	Method	# Params.	PSNR (dB) In-training-scale ×2	×3	×4	Out-of-scale ×5	×7	×9
SCI1K (n = 200)	Bicubic	-	28.81	25.15	23.18	22.02	20.72	19.96
	RDN (18')	21.97M	38.45	33.59	29.81	-	-	-
	MetaSR (19')	22.42M	38.57	33.67	30.12	27.52	23.91	22.02
	LIIF (21')	22.32M	38.65	33.97	30.55	27.77	23.99	22.18
	ITSRN (21')	22.62M	38.74	34.32	30.82	28.15	24.36	22.36
	LTE (22')	22.53M	39.14	34.50	30.93	28.22	24.28	22.39
	BTC (ours)	22.40M	39.17	34.58	31.10	28.33	24.47	22.48

▲ **Quantitative comparison on SR results (PSNR (dB)).** All methods use RDN (18') as an encoder.

Method	(a) Computation cost Mem.	Time (mean±std)	(b) Acc./Conf. (%) ×4	×5
MetaSR (19')	15.1 GB	655.91±2.09 ms	90.89/98.19	87.07/94.89
LIIF (21')	12.3 GB	1024.02±3.33 ms	91.00/98.26	87.17/94.69
ITSRN (21')	21.0 GB	1185.45±4.95 ms	90.89/98.13	87.07/96.80
LTE (22')	8.7 GB	1099.34±2.54 ms	90.89/98.00	87.07/96.18
BTC (ours)	7.8 GB	958.77±3.28 ms	93.63/98.55	89.70/97.88

Qualitative comparison

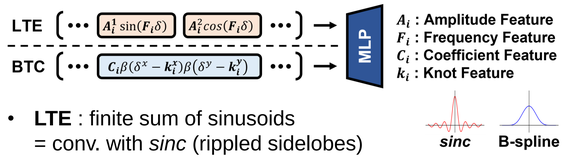


▲ **Visual comparison for x5 and x7, respectively.** Scene Text Recognition (STR) results for the red box are reported.

➤ **(a) Computation cost comparison.** We use 480x480 sized input on x2 SR. For computation time, we iterated 300 times.
(b) Scene text recognition (STR) comparison. Per each scale, we randomly crop text regions from SCID dataset.

4. Discussion

B-spline (BTC) vs. Fourier (LTE) for SCI SR



- **LTE** : finite sum of sinusoids
  = conv. with *sinc* (rippled sidelobes)
  = (-) under/overshooting↑ at discontinuity
- **BTC** : conv. with B-spline (positive and integrated to 1)
  = map the value btw. min/max of input signal
  = (+) under/overshooting↓ at discontinuity

Under/Overshooting at discontinuity