Abstract

Introduction

- The goal of **SR**(Super Resolution) is to enhance a **LR**(Low Resolution) image to a **HR**(High Resolution) image.
- The domain of SR research can be divided into 3 main areas: **SISR**(Single Image SR), **MISR**(Multi Image SR) and **VSR**(Video SR).
- Consider an LR video which consists of a sequence of frames: $LR_{t-n}, \dots, LR_t, \dots, LR_{t+n}$, where we super-resolve a target frame LR_t.
  - SISR: Only use LR_t, so obviously it fails to utilize temporal information.
  - MISR: Align all the neighboring frames and fuses them to utilize the mising details observed only in one or some of the frames.
  - VSR: Almost similar to **MISR**, but the alignment progress is done with concern for temporal smoothness, which is not concerned in **MISR**.
- **IQA**(Image Quality Assessment): **PSNR**(Peak Signal to Noise Ratio) and **SSIM**(Structural Similarity).
  - PSNR:

$$MSE = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} (X(i,j) - Y(i,j))^2$$

$$PSNR = 10 \lg(\frac{(2^n - 1)^2}{MSE})$$

    - Larger is better, the unit is decible(db).
    - **n** is number of bits per pixel.
  - SSIM:

$$\mu_X = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} X(i,j),\ \sigma_X^2 = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} (X(i,j) - \mu_X)^2$$

$$\sigma_{XY} = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} ((X(i,j) - \mu_X)(Y(i,j) - \mu_Y))$$

$$l(X,Y) = \frac{2\mu_X \mu_Y + C_1}{\mu_X^2 + \mu_Y^2 + C_1},\ c(X,Y) = \frac{2\sigma_X \sigma_Y + C_2}{\sigma_X^2 + \sigma_Y^2 + C_2},\ s(X,Y) = \frac{\sigma_{XY} + C_3}{\sigma_X \sigma_Y + C_3}$$

$$SSIM(X,Y) = l(X,Y) \cdot c(X,Y) \cdot s(X,Y)$$

    - Larger is better.
    - **L**ight, **C**ontrast and **S**tructure.
    - To avoid the case where the denominator is 0, normally we use $C_1 = (K_1 \cdot L)^2, C_2 = (K_2 \cdot L)^2, C_3 = C_2/2$, and $K1 = 0.01, K2 = 0.03, L = 255$.
- In most cases, **CNNs** are well qualified for the job, but once the upscaling factors are large enough, **GAN** approach is nearly indispensable to render the "naturality".
- Deep VSR can be divided into 5 types based on the approach to preserving temporal information.
  - Temporal Concatenation: `DUF`
  - Temporal Aggregation: `Robust Video Super-Resolution with Learned Temporal Dynamics`

- RNNs: `Detail-revealing deep video super-resolution`, `Bidirectional recurrent convolutional networks for multi-frame super-resolution`
    - many-to-one
    - many-to-many
- Optical Flow: `Framerecurrent video super-resolution`
    - warp
        - `Video super-resolution with convolutional neural networks`
        - `Total variation regularization of local-global optical flow`
    - `VESPCN`: replace the optical flow model with a trainable motion compensation network.
- Pre-training then Fine-tuning
    - `VESPCN`
    - `Detail-revealing deep video super-resolution`
    - `Robust Video Super-Rsolution with Learned Temporal Dynamics`

- 2D-conv & 3D-conv: https://www.zhihu.com/question/266352189
- Deep SR can be divided into 4 types of upsampling.
    - Predefined upsampling: `SRCNN`
        - LR-(interpolation)-MR-(CNNs)-HR
    - Single upsampling: `FSRCNN`, `ESPCN`, `EDSR`
        - Lengthy training time.
- Progressive upsampling: `LapSRN`
    - Iterative up and downsampling: `DBPN`, `RBPN`

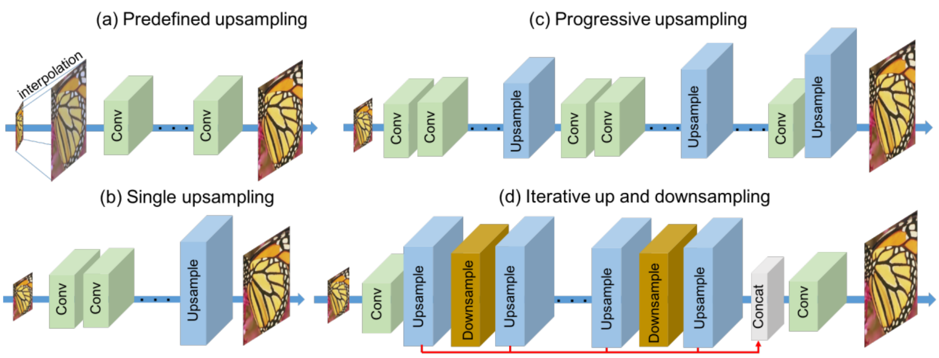

Figure 2. Comparisons of Deep Network SR. (a) Predefined upsampling (e.g., SRCNN [6], VDSR [22], DRRN [43]) commonly uses the conventional interpolation, such as Bicubic, to upscale LR input images before entering the network. (b) Single upsampling (e.g., FSRCNN [7], ESPCN [38]) propagates the LR features, then construct the SR image at the last step. (c) Progressive upsampling uses a Laplacian pyramid network which gradually predicts SR images [25]. (d) Iterative up and downsampling approach is proposed by our DBPN which exploit the mutually connected up- (blue box) and down-sampling (gold box) stages to obtain numerous HR features in different depths.

Related Work

timeline

- SRCNN[2014], VDSR[2016]
- DUF[2018], TDAN[2018], EDVR[2019]
- ESPCN(calc in lr space)[2016], VESPCN(video espcn)[2017]
- DBPN[2018], RBPN[2019]
- SRGAN[2017]
- iSeeBetter(RBPN+SRGAN)[2020]

NTIRE

- 2017: `EDSR`
- 2018: `RBPN`

results

- SRCNN

Video Super-Resolution	Ultra Video Group HD - 4x upscaling	SRCNN	Average PSNR	37.52	# 2		Compare
Image Super-Resolution	Urban100 - 4x upscaling	SRCNN	PSNR	24.52	# 35		Compare
			SSIM	0.7221	# 31		Compare
Video Super-Resolution	Vid4 - 4x upscaling	SRCNN	PSNR	24.68	# 14		Compare
			SSIM	0.7158	# 13		Compare
			MOVIE	6.90	# 4		Compare
Video Super-Resolution	Xiph HD - 4x upscaling	SRCNN	Average PSNR	31.47	# 2		Compare

- EDVR

Results from the Paper

Edit

Ranked #1 on Deblurring on REDS

Get a GitHub badge

TASK	DATASET	MODEL	METRIC NAME	METRIC VALUE	GLOBAL RANK	RESULT	BENCHMARK
Deblurring	REDS	EDVR_Deblur	Average PSNR	34.80	# 1		Compare
Video Super-Resolution	Vid4 - 4x upscaling	EDVR	PSNR	27.35	# 4		Compare
			SSIM	0.8264	# 3		Compare

- ESPCN

Video Super-Resolution	Ultra Video Group HD - 4x upscaling	ESPCN	Average PSNR	37.91	# 1		Compare
Video Super-Resolution	Ultra Video Group HD - 4x upscaling	bicubic	Average PSNR	36.20	# 3		Compare
Video Super-Resolution	Vid4 - 4x upscaling	ESPCN	PSNR	25.06	# 13		Compare
			SSIM	0.7394	# 11		Compare
			MOVIE	6.54	# 3		Compare
Video Super-Resolution	Xiph HD - 4x upscaling	bicubic	Average PSNR	30.30	# 3		Compare
Video Super-Resolution	Xiph HD - 4x upscaling	ESPCN	Average PSNR	31.67	# 1		Compare

- VESPCN

Ranked #10 on Video Super-Resolution on Vid4 - 4x upscaling

Get a GitHub badge

TASK	DATASET	MODEL	METRIC NAME	METRIC VALUE	GLOBAL RANK	RESULT	BENCHMARK
Video Super-Resolution	Vid4 - 4x upscaling	VESPCN	PSNR	25.35	# 12		Compare
			SSIM	0.7557	# 10		Compare
			MOVIE	5.82	# 2		Compare
Video Super-Resolution	Vid4 - 4x upscaling	bicubic	PSNR	23.82	# 16		Compare
			SSIM	0.6548	# 15		Compare
			MOVIE	9.31	# 5		Compare

- DBPN

TASK	DATASET	MODEL	METRIC NAME	METRIC VALUE	GLOBAL RANK	RESULT	BENCHMARK
Image Super-Resolution	BSD100 - 4x upscaling	D-DBPN	PSNR	27.72	# 11		Compare
			SSIM	0.740	# 18		Compare
Image Super-Resolution	Manga109 - 4x upscaling	D-DBPN	SSIM	0.914	# 14		Compare
Image Super-Resolution	Set14 - 4x upscaling	D-DBPN	PSNR	28.82	# 11		Compare
			SSIM	0.786	# 19		Compare
Image Super-Resolution	Set5 - 4x upscaling	D-DBPN	PSNR	32.47	# 12		Compare
			SSIM	0.898	# 16		Compare
Image Super-Resolution	Urban100 - 4x upscaling	D-DBPN	SSIM	0.795	# 15		Compare
Video Super-Resolution	Vid4 - 4x upscaling	DBPN	PSNR	25.37	# 11		Compare
			SSIM	0.737	# 12		Compare

- RBPN

Results from the Paper

Edit

Ranked #6 on Video Super-Resolution on Vid4 - 4x upscaling

Get a GitHub badge

TASK	DATASET	MODEL	METRIC NAME	METRIC VALUE	GLOBAL RANK	RESULT	BENCHMARK
Video Super-Resolution	Vid4 - 4x upscaling	RBPN/6-PF	PSNR	27.12	# 6		Compare
			SSIM	0.818	# 6		Compare

- iSeeBetter

Results from the Paper

Edit

Ranked #1 on Video Super-Resolution on Vimeo90k

Get a GitHub badge

TASK	DATASET	MODEL	METRIC NAME	METRIC VALUE	GLOBAL RANK	BENCHMARK
Video Super-Resolution	Vid4 - 4x upscaling	iSeeBetter	PSNR	27.43	# 3	Compare
			SSIM	0.835	# 1	Compare
Video Super-Resolution	Vimeo90k	iSeeBetter	PSNR	40.17	# 1	Compare