Deep Learning based Super-Resolution

Zhifei Zhang
Keywords

• Deep learning
• Super-resolution (SR)
• Natural images
• Top conferences, e.g.,
  ▪ CVPR (held every year since 1985)
  ▪ ECCV (held on even years since 1990)
  ▪ ICCV (held on odd years since 1987)
Brief Statistics and Milestones

<table>
<thead>
<tr>
<th>Year</th>
<th>Conference</th>
<th>Papers</th>
<th>Citations</th>
</tr>
</thead>
<tbody>
<tr>
<td>2014</td>
<td>CVPR</td>
<td>SRCNN</td>
<td>743</td>
</tr>
<tr>
<td>2015</td>
<td>ECCV</td>
<td>SCN</td>
<td>171</td>
</tr>
<tr>
<td>2016</td>
<td>ICCV</td>
<td>ShCNN</td>
<td>32</td>
</tr>
<tr>
<td>2017</td>
<td>NIPS</td>
<td>SCKN</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>ICLR</td>
<td>Suff. Stat.</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Amortized MAP</td>
<td>71</td>
</tr>
</tbody>
</table>

Superscript: Number of citations from Google Scholar until Apr. 10, 2018 6:30 PM.
Trend (1/4) --- Trained Upscaling

Directly start from low-resolution image:
• Faster
• Less parameters
• Learn the upscaling process
Skip low-level features to high-level layers:

- Assist identity mapping
- Alleviate vanishing gradient problem
- Finer texture
The sub-pixel convolutional layer is faster than the deconvolution layer
Trend (4/4) --- Perceptual Loss

Original
SRCNN (L2 loss with 3-layer net)
SRResNet (L2 loss with deep ResNet)
VGG feature loss
SRGAN (Adversarial loss)

Non-perceptual loss

Perceptual loss
State-of-the-art Performance

CVPR
- SRCNN$^{743}$
- VDSR$^{413}$
- DRCN$^{168}$
- LapSRN$^{53}$
- DRRN$^{32}$

ECCV
- Perceptual Loss$^{465}$
- FSRCNN$^{137}$
- Sub-Pixel$^{219}$
- SRGAN$^{455}$
- EDSR$^{40}$

ICCV
- SCN$^{171}$
- ENet$^{32}$
- MemNet$^{9}$
- SRDenseNet$^{8}$

The state-of-art-work in visual quality, but with low PSNR

The state-of-art-work in PSNR, but visually blurry

They share similar structure: res blocks, skip connection, and post upscaling
State-of-the-art Performance

<table>
<thead>
<tr>
<th>Year</th>
<th>2014</th>
<th>2015</th>
<th>2016</th>
<th>2017</th>
</tr>
</thead>
<tbody>
<tr>
<td>CVPR</td>
<td>SRCNN</td>
<td>SCN</td>
<td>Perceptual Loss</td>
<td>SRGAN</td>
</tr>
<tr>
<td>ECCV</td>
<td>DRCN</td>
<td>FSRCNN</td>
<td>ENet, MemNet</td>
<td>DRRN</td>
</tr>
<tr>
<td>ICCV</td>
<td>SCN</td>
<td>SRDenseNet</td>
<td>EDSR</td>
<td></td>
</tr>
</tbody>
</table>

- **SRCNN**
- **SCN**
- **Perceptual Loss**
- **SRGAN**
- **DRCN**
- **FSRCNN**
- **ENet**
- **MemNet**
- **SRDenseNet**

**Performance Metrics**

- **PSNR**
- **Set14**
- **BSD100**
SRGAN --- First Introduce GAN to SR
EDSR --- Winner of NTIRE 2017

NTIRE 2017
New Trends in Image Restoration and Enhancement workshop and challenge on image super-resolution in conjunction with CVPR 2017
EDSR --- Network Structure

Input: CONV → ResBlock → CONV → ... → ResBlock → CONV → Upsampler → CONV → Output

- ResBlock: CONV + RelU + CONV
- Upsampler: X4
- Shuffle

Results:
- 0853 from DIV2K [26]
- Bicubic (30.80 dB / 0.9537)
- VDSR [11] (32.82 dB / 0.9623)
- SRResNet [14] (34.00 dB / 0.9679)
- EDSR+ (Ours) (34.78 dB / 0.9708)
Remove batch normalization:
• BN normalizes the features, it gets rid of range flexibility from networks by normalizing the features.
• Save approximately 40% of memory usage during training
Summary

Current Trends:
• Trained upscaling
• Skip connection
• Sup-pixel
• Perceptual loss

Existing Problems:
• Measurement metric, e.g., PSNR is not consistent to human evaluation
• Assumption on bicubic downscaling
• Lack of fine texture

It seems more and more difficult to make improvement to the traditional SR problem, especially in PSNR. It may be the time to explore new directions.
Interesting Papers on Single Image SR in CVPR 2018

Supervised learning → Unsupervised learning
A. Shocher et al., “Zero-Shot” Super-Resolution using Deep Internal Learning

Bicubic downscaling → Unknown downscaling
K. Zhang et al., Learning a Single Convolutional Super-Resolution Network for Multiple Degradations

One-way upscaling → Iterative up/downscaling
M. Haris et al., Deep Back-Projection Networks For Super-Resolution
Supervised learning → Unsupervised learning

A. Shocher et al., “Zero-Shot” Super-Resolution using Deep Internal Learning

Motivation: handling poor-quality low-resolution images, e.g., old photos, noisy images, biological data, and other images where the downscaling process is unknown or non-ideal.
Supervised learning $\rightarrow$ Unsupervised learning

A. Shocher et al., “Zero-Shot” Super-Resolution using Deep Internal Learning

Low-resolution image

EDSR

The proposed
Bicubic downscaling $\rightarrow$ Unknown downscaling

K. Zhang et al., Learning a Single Convolutional Super-Resolution Network for Multiple Degradations

Motivation:
Just like the paper title, breaking the assumption that a low-resolution image is bicubicly downsampled from a high-resolution image.
Motivation:
Iterative error feedback by back-projection, addressing the mutual dependencies of low- and high-resolution images.

The dense inter-layer connections alleviate the vanishing gradient problem, produce improved feature, and encourage feature reuse.
Thank you