EnhanceNet

Single Image Super-Resolution
Through Automated Texture Synthesis

state of the art by PSNR
our result

Mehdi S. M. Sajjadi, Bernhard Schölkopf, Michael Hirsch

Max Planck Institute for Intelligent Systems

Single Image Super-Resolution

Low-resolution input  →  ?  →  Original high-res image
Method: Classical approach

- Minimize mean-squared error
  \[ \text{MSE} = \frac{1}{nm} \sum_{i}^{n} \sum_{j}^{m} (\tilde{I}_{ij} - I_{ij})^2 \]

- Evaluation by Peak signal-to-noise ratio (PSNR)
  \[ \text{PSNR} = -10 \log_{10} (\text{MSE}) \]

- The field is dominated by convolutional neural nets
  - SRCNN (Dong et al., ECCV 2014)
  - DRCN, VDSR (Kim et al., CVPR 2016)
  - NTIRE challenge on image super-resolution (CVPR 2017)
State of the art by PSNR

Low-resolution input ➔ CNN ➔ High-resolution output
Is PSNR the right metric?

Low-resolution input  Generated image  Original image

EnhanceNet’s output

Low-resolution input  Our result  Original image
PSNR vs. visual similarity

- - - -
- - - -
- - - -
- - - -

High-resolution image

- - - -
- - - -
- - - -
- - - -

Low-resolution image

- - - -
- - - -
- - - -
- - - -

Optimal PSNR

- - - -
- - - -
- - - -
- - - -

Realistic image

- - - -
- - - -
- - - -
- - - -

Low PSNR
Method: residual CNN with loss

- **Euclidean distance / mean squared error (MSE)**
  \[ \|\bar{I} - I\|_2^2 = \frac{1}{NMC} \sum_{ij} (\bar{I}_{ij} - \bar{I}_{ij})^2 \]

- **Perceptual loss** (Dosovitskiy and Brox 2016, Johnson et al. 2016)
  \[ \|\phi(\bar{I}) - \phi(I)\|_2^2 \text{ MSE in VGG feature space} \]

- **Texture loss / style transfer** (Gatys et al. 2015)
  \[ \|G(\phi(\bar{I})) - G(\phi(I))\|_2^2 \text{ MSE of correlation in VGG feature space} \]

- **Adversarial loss / GAN** (Goodfellow et al. 2014)
  \[ D(\bar{I}), D(I) \in [0, 1] \text{ discriminator rates realism of image patches} \]
Best result by PSNR vs. our best
Evaluation

- PSNR/SSIM/IFC: **ENet-E SOTA**; ENet-PAT low scores
- Survey: **ENet-PAT** is preferred over ENet-E in **91.0%** images
- Object recognition image quality benchmark
  - Feed ImageNet through super-resolution models
  - Run pre-trained object recognition network on results
  - **ENet-PAT leads to lowest error**

<table>
<thead>
<tr>
<th>Evaluation</th>
<th>Bicubic</th>
<th>DRCN [26]</th>
<th>PSyCo [40]</th>
<th>ENet-E</th>
<th>ENet-EA</th>
<th>ENet-PA</th>
<th>ENet-PAT</th>
<th>Baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top-1 error</td>
<td>0.506</td>
<td>0.477</td>
<td>0.454</td>
<td>0.449</td>
<td>0.407</td>
<td>0.429</td>
<td><strong>0.399</strong></td>
<td>0.260</td>
</tr>
<tr>
<td>Top-5 error</td>
<td>0.266</td>
<td>0.242</td>
<td>0.224</td>
<td>0.214</td>
<td>0.185</td>
<td>0.199</td>
<td><strong>0.171</strong></td>
<td>0.072</td>
</tr>
<tr>
<td>Confidence</td>
<td>0.754</td>
<td>0.727</td>
<td>0.728</td>
<td>0.754</td>
<td>0.760</td>
<td>0.783</td>
<td><strong>0.797</strong></td>
<td>0.882</td>
</tr>
</tbody>
</table>
Comparison w/ other methods


PSyCo [40]  VDSR [25]  DRCN [26]  ENet-E  ENet-PAT  \( I_{HR} \)
A closer look at ENet-PAT

Bicubic

ENet-PAT

$I_{HR}$
Conclusion

- Introduce a novel combination of loss functions for single image super-resolution
- State of the art in quantitative + qualitative benchmarks
- Propose object detection image quality benchmark
  - Let’s see if it works in other domains as well
- Outlook: lots of room for improvements
  - Perceptual evaluation still an unsolved problem