SRFlow-DA: Super-Resolution Using Normalizing Flow with Deep Convolutional Block

Younghyun Jo  Sejong Yang  Seon Joo Kim

Yonsei University

NTIRE 2021
Learning the Super-Resolution Space Challenge
Learning SR Space: Concept

- Many high-resolution images can be downsampled to the same low-resolution image.
- Ill-posed.

- Most of current SR methods only map one LR to one HR.
- Deterministic mapping.

https://github.com/andreas128/NTIRE21_Learning_SR_Space
Learning SR Space: Concept

- Many high-resolution images can be downsampled to the same low-resolution image.
- Ill-posed.

- One can frame the SR problem as learning a stochastic mapping.
- Challenge: How to make it?
VAE vs GAN vs Flow

Implicit density.

GAN: minimax the classification error loss.

Explicit approximation.

VAE: maximize ELBO.

Explicit density.

Flow-based generative models: minimize the negative log-likelihood

VAE vs GAN vs Flow


• “VarSR”, VAE like, stochastic output.
• For faces, numbers.


• “DeepSEE”, GAN, stochastic output.
• For faces.


• “SRFlow”, normalizing flow, stochastic output.
• For general images.
Our Baseline: SRFlow

1x1 Conv \( y_{i,j} = W x_{i,j} \)

Actnorm \( y_{i,j} = s \odot x_{i,j} + b \)

Reverse \( x_{i,j} = W^{-1} y_{i,j} \)

Reverse \( x_{i,j} = (y_{i,j} - b)/s \)
Our Baseline: SRFlow

- For x8 SR.
- 16 flow steps for each level ($K=16$).
- Total 4 scale level ($L=4$).

Affine coupling (affine injector)

Reverse
Our Approach: SRFlow-DA (Deep convolutional block in the Affine couplings)

Original affine coupling

Our approach (SRFlow-DA)

- Stack 6 3x3 conv layers.
- Remove actnorm.

- 5x5 Receptive field (RF).
- $K=16 \rightarrow$ Total 129x129 RF.
- $L=4$.
- Params: 34.1M.
- Training time: 120h (RTX 2080 Ti).
- Inference time: 1.97s (1920x1080 output).

- 13x13 RF.
- $K=6 \rightarrow$ Total 145x145 RF.
- $L=3$.
- Params: 13.3M.
- Training time: 47h (RTX 2080 Ti).
- Inference time: 1.01s (1920x1080 output).
Variants of SRFlow-DA

- **SRFlow-DA-R**
  - 13x13 RF.
  - \(K=6\).
  - \(L=3\).
  - Add a skip connection.

- **SRFlow-DA-S**
  - 13x13 RF.
  - \(K=18\).
  - \(L=1\).
  - For a single-scale architecture.

- **SRFlow-DA-D**
  - 77x77 RF.
  - \(K=1\).
  - \(L=3\).
  - To see if the performance can increase using only a deeper block without many flow steps.
## Table 1: SR Results on DIV2K Validation Set

### (a) ×4 SR Results on DIV2K Validation Set.

<table>
<thead>
<tr>
<th>Method</th>
<th>Params</th>
<th>Time</th>
<th>PSNR</th>
<th>LPIPS</th>
<th>Diversity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>LR</td>
<td>HR</td>
<td>mean</td>
</tr>
<tr>
<td>Bicubic</td>
<td>-</td>
<td>-</td>
<td>38.70</td>
<td>26.70</td>
<td>0.409</td>
</tr>
<tr>
<td>RRDB</td>
<td>16.7M</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.253</td>
</tr>
<tr>
<td>ESRGAN</td>
<td>16.7M</td>
<td>-</td>
<td>39.03</td>
<td>26.22</td>
<td>0.124</td>
</tr>
<tr>
<td>SRFLOW</td>
<td>22.8M</td>
<td>120h</td>
<td>50.64</td>
<td>27.09</td>
<td>0.120</td>
</tr>
<tr>
<td>SRFLOW-DA</td>
<td>8.7M</td>
<td>33h</td>
<td>50.88</td>
<td>27.57</td>
<td>0.121</td>
</tr>
<tr>
<td>SRFLOW-DA-R</td>
<td>8.7M</td>
<td>33h</td>
<td>50.92</td>
<td>27.23</td>
<td>0.120</td>
</tr>
<tr>
<td>SRFLOW-DA-S</td>
<td>8.8M</td>
<td>25h</td>
<td>50.48</td>
<td>27.43</td>
<td>0.130</td>
</tr>
<tr>
<td>SRFLOW-DA-D</td>
<td>6.4M</td>
<td>30h</td>
<td>49.00</td>
<td>26.78</td>
<td>0.132</td>
</tr>
</tbody>
</table>

### (b) ×8 SR Results on DIV2K Validation Set.

<table>
<thead>
<tr>
<th>Method</th>
<th>Params</th>
<th>Time</th>
<th>PSNR</th>
<th>LPIPS</th>
<th>Diversity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>LR</td>
<td>HR</td>
<td>mean</td>
</tr>
<tr>
<td>Bicubic</td>
<td>-</td>
<td>-</td>
<td>37.16</td>
<td>23.74</td>
<td>0.584</td>
</tr>
<tr>
<td>RRDB</td>
<td>16.7M</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.419</td>
</tr>
<tr>
<td>ESRGAN</td>
<td>16.7M</td>
<td>-</td>
<td>31.35</td>
<td>22.18</td>
<td>0.277</td>
</tr>
<tr>
<td>SRFLOW</td>
<td>34.1M</td>
<td>120h</td>
<td>50.09</td>
<td>23.03</td>
<td>0.272</td>
</tr>
<tr>
<td>SRFLOW-DA</td>
<td>13.3M</td>
<td>47h</td>
<td>50.91</td>
<td>23.75</td>
<td>0.261</td>
</tr>
<tr>
<td>SRFLOW-DA-R</td>
<td>13.3M</td>
<td>47h</td>
<td>50.73</td>
<td>23.42</td>
<td>0.256</td>
</tr>
<tr>
<td>SRFLOW-DA-S</td>
<td>15.0M</td>
<td>30h</td>
<td>52.21</td>
<td>23.71</td>
<td>0.268</td>
</tr>
<tr>
<td>SRFLOW-DA-D</td>
<td>9.5M</td>
<td>37h</td>
<td>45.18</td>
<td>22.54</td>
<td>0.306</td>
</tr>
</tbody>
</table>

**Challenge submission version**

**Our best**

**Diversity: span of global and local values**

**Worse diversity, but better global, local values**

10
Thank you

https://github.com/yhjo09/SRFlow-DA