**Motivation**

A low-resolution (LR) image corresponds to infinite high-resolution (HR) solutions. Yet, existing super-resolution (SR) methods (e.g., ESRGAN) output only one.

**Outputs of SR networks**

**Goal:**

Methods (e.g., ESRGAN) output only one resolution.

But could it be a Zebra?

But could it be a ‘0’ or ‘7’?

ESRGAN

Other perfectly consistent reconstructions

Outputting one arbitrary HR solution problematic in many scenarios, e.g.,

- **Motivation**
  - Can only reduce reconstruction error (increase PSNR):
    - EDSR
    - EDSR + CEM
  - Accommodating realistic, image-specific downsampling kernels:

<table>
<thead>
<tr>
<th>SR factor</th>
<th>2x</th>
<th>3x</th>
<th>4x</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSNR</td>
<td>35.97</td>
<td>32.27</td>
<td>30.30</td>
</tr>
</tbody>
</table>

**Consistency Enforcing Module (CEM)**

Outputs of SR networks not guaranteed to match LR input when downsampled

The SR problem

Conventional SR networks

Inconsistent high-res

High res

Low res

The CEM

Promote periodic patterns

Promote local brightness

Propagate patches from source to target regions

Embed hand scribble using various tools

Manipulate local smoothness

Our CEM

A Fixed (non-learnable) architectural module

Consistent high-res

Our outputs, all consistent with input (precisely match it when downsampled)

**Exploration tools**

- Manually controlling spatial derivatives
- Promote local brightness
- Manually controlling periodic patterns
- Spread scribbles

**Use case #1 - Manual texture editing**

Exploring different possible digits:

“Acromiolumbral distance” (between bones), indicative of tendon tear when < 7 mm

**Other example use cases**

Accommodating existing user knowledge regarding image subject:

- Forensics - Examining possible HR solutions to a captured LR image:
- Medical – Checking how likely is a certain pathology:

- Exploring different possible digits:
  - Low-res input: Outputs seem plausible only for 0,1,8, suggesting other digits less likely
  - Low-res suspected
  - Low-res input: Outputs seem plausible only for 0,1,8, suggesting other digits less likely

Pre-edited SR

Exploring different possible digits:

- Low-res suspected
- Low-res input: Outputs seem plausible only for 0,1,8, suggesting other digits less likely

“Acromiolumbral distance” (between bones), indicative of tendon tear when < 7 mm

Pre-edited SR

Exploring different possible digits:

- Low-res input: Outputs seem plausible only for 0,1,8, suggesting other digits less likely

**Model training**

Adversarial loss

Control-enabling loss

LR inputs

SR outputs

Random control signals

SR Network

CEM

In contrast to existing methods, using CEM allows us to use no full reference loss terms (e.g., $\ell_1$, VGG), that promote overly smoothed, non-diverse outputs.

**Exploration tools**

- Manually controlling spatial derivatives
- Promote local brightness
- Manually controlling periodic patterns
- Spread scribbles

**Consistency Enforcing Module (CEM)**

Outputs of SR networks not guaranteed to match LR input when downsampled

The SR problem

Conventional SR networks

Inconsistent high-res

High res

Low res

The CEM

Promote periodic patterns

Promote local brightness

Propagate patches from source to target regions

Embed hand scribble using various tools

Manipulate local smoothness

Our CEM

A Fixed (non-learnable) architectural module

Consistent high-res

Our outputs, all consistent with input (precisely match it when downsampled)