**Introduction**

**Motivation:**
- Over 300 million photos under food-related hashtags on Instagram. A large portion of these images are low-quality images.
- Potential application of efficiently transmitting low-resolution images and then upscaling using our model.

**Goal:**
- To train a generative model to enhance low-quality food images by performing super-resolution and de-blurring.

**Source:**
- Scraped 8000 images from Instagram tagged with #food, #foodporn, #yum, #yummy.

**Approach:**
- SRGAN: Vanilla baseline implementation
- VGGAN: Uses augmented VGG Loss
- SSIMGAN: Optimize for Structural Similarity

**Dataset**

**Source:**
- Scraped 8000 images from Instagram tagged with #food, #foodporn, #yum, #yummy.

**Processing:**
1. Blurring: Used a 3x3 Gaussian filter with $\sigma = 0.5$ and then applied a depth-wise 2D convolution.
2. Downsampling: Applied a 4x4 average pooling layer.
3. Cropping: Random 96x96 image crops

**Split:**
- Training: 7500
- Validation: 250
- Test: 250

**Methods**

1. **SRGAN:**
   - Reimplemented in Keras with TF 2.0
   - No modifications, trained from scratch on our high quality food image dataset
   - Preprocessed with a non-trainable Keras model to implement the gaussian blur and downsampling

2. **VGGAN:**
   - VGG Loss is the MSE between image features of generated and original HR images
   - Augmented Loss: Use pool1 and pool2 features in addition to pool4 (highlighted)

3. **SSIMGAN:**
   - Added structural similarity as a maximizing objective to the generator loss function based on the following equation:
   $$SSIM(x, y) = \frac{(2\mu_x \mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

**Evaluation**

**Conditions:**
- Each model was run for at least 15 epochs for testing
- Most of them converged
- High res images were downsampled (4x) and blurred ($\sigma = 0.8$)

**Metrics:**
- SSIM: Structural Similarity between original and generated image
- PSNR: Peak Signal to Noise Ratio
- Approximation to human perception of reconstruction quality
- LBD: “Variance of Laplacian” Blur Detector
- Measures sharpness of an image
- RANK
  - Qualitative average ranking of each model within one image

**Conclusion**

**Conclusions:**
- SSIMGAN gives the best qualitative results while VGGAN has the best sharpness results
- Small modifications, such as changing the generator loss function, can create noticeably different outputs
- Targeted improvements in the loss function can help better performance of SRGAN-like models for a specific task

**Next Steps:**
- Include color difference as a minimizing objective in the loss function in order to minimize the color fading
- Use pre-trained models other than VGG for feature representations.
- Try out different architectures for the generator.

**Results**

**Quantitative Results**

<table>
<thead>
<tr>
<th></th>
<th>SSIM</th>
<th>PSNR</th>
<th>LBD</th>
<th>RANK</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRGAN</td>
<td>0.804</td>
<td>67.7</td>
<td>138.5</td>
<td>1.94</td>
</tr>
<tr>
<td>VGGAN</td>
<td>0.781</td>
<td>66.0</td>
<td>150.9</td>
<td>2.3</td>
</tr>
<tr>
<td>SSIMGAN</td>
<td>0.848</td>
<td>68.6</td>
<td>30.2</td>
<td>1.76</td>
</tr>
</tbody>
</table>

**References**


[7] https://doi.org/10.1145/3230519.3230587


Acknowledgements: Thanks to our Project Mentor Jo Chuang for his helpful guidance and advice. Thank you to AWS for the computational resources.