FoodGAN: SuperResolution on Low-Quality Food Images

Amay Aggarwal, Chinmay Patel, Federico Reyes Gomez {amayagg, chinmayp, frg100}@stanford.edu
Department of Computer Science, Stanford University

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.
- **Input:** A blurred low resolution (270 x 270) food image
- **Output:** A de-blurred higher resolution (1080 x 1080) version of the input image.

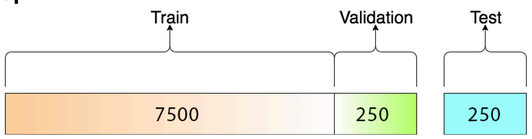
Approach:
- SRGAN: Vanilla baseline implementation
- VGGAN: Uses augmented VGG Loss
- SSIMGAN: Optimize for Structural Similarity

Dataset

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

Processing:
1. Blurring: Used a 3x3 Gaussian filter with σ = 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:



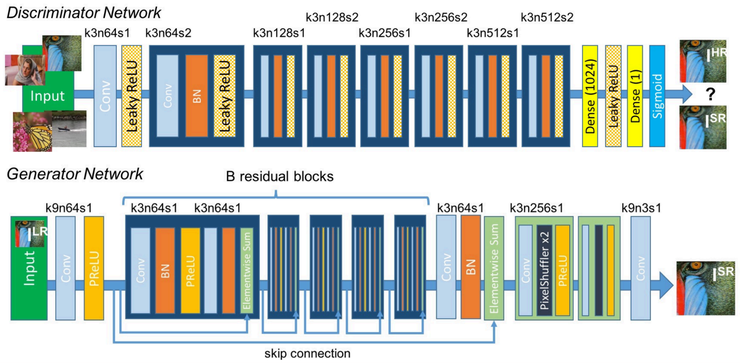
Methods

1. SRGAN:

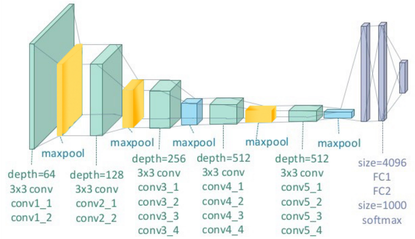
Reimplemented in Keras with TF 2.0

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:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

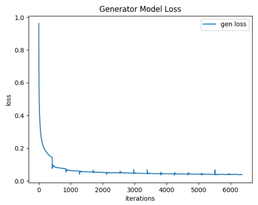
Results

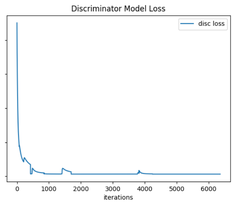
1: Downsampled Input

2: SRGAN Output

3: VGGAN Output

4: SSIMGAN Output





Model	SSIM	PSNR	LBD	RANK
SRGAN	0.804	67.7	138.5	1.94
VGGAN	0.781	66.0	150.9	2.3
SSIMGAN	0.848	68.6	30.2	1.76
HR	1	N/A	451.1	N/A

Quantitative Results

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 (σ = 0.8)

Metrics:
- SSIM: Structural Similarity between original and generated image
  - High level structure
- 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.

References

[1] Ledig, Christian, et al. "Photo-realistic single image super-resolution using a generative adversarial network." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
[2] https://thenextweb.com/opinion/2015/09/01/why-sharing-photos-of-food-is-about-more-than-whats-on-the-plate/
[3] Shu, Yujie. "Human Portrait Super Resolution Using GANs." CS 230.
[4] Yudai Nagano and Yohei Kikuta. 2018. SRGAN for super-resolving low-resolution food images. In Proceedings of the Joint Workshop on Multimedia for Cooking and Eating Activities and Multimedia Assisted Dietary Management (CEA/MADiMa '18). ACM, New York, NY, USA, 33-37. DOI: https://doi.org/10.1145/3230519.3230587
[5] Jaffe et al. "Super-Resolution to Improve Classification Accuracy of Low-Resolution Images"
[6] Nao Takano and Gita Alaghband. "SRGAN: Training Dataset Matters". 2019
[7] https://www.pyimagesearch.com/2015/09/07/blur-detection-with-opencv/
[8] Ke Li, Shichong Peng, and Jitendra Malik "Super-Resolution via Implicit Maximum Likelihood Estimation" 2018

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