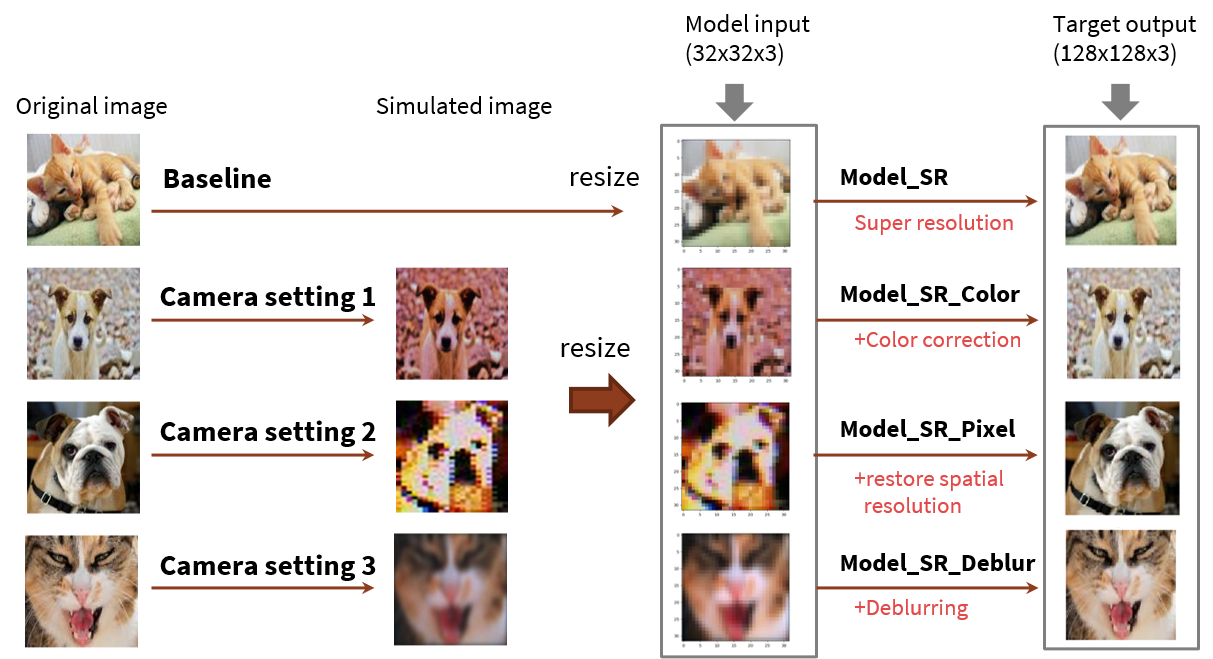


|  |
| --- |
| HAPPYMONK.AI  Assignment  TASK III |
|  |
| November 18  HAPPYMONK.AI TECH. PVT. LTD.  Authored by: REISHIKA GHOSH |

# SRGAN-for-Super-Resolution-and-Image-Enhancement

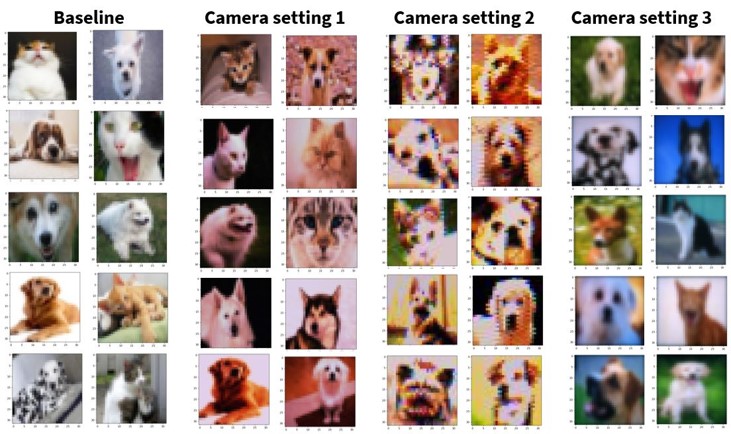
Tensorflow implementation of SRGAN-based [Photo-realistic single image super-resolution using a generative adversarial network](https://arxiv.org/abs/1609.04802) to generate high-resolution (HR) images from low-resolution (LR) images. In this work, we use SRGAN to upscale 32x32 images to 128x128 pixels. Meanwhile, we evaluate the impact of different camera parameters on the quality of final up-scaled (high resolution) images and infer from these stimuli to understand what the network can learn.

[](https://user-images.githubusercontent.com/65942005/100526323-93c81080-317c-11eb-991a-2299057ba5b0.png)Data processing and model training pipeline: The original image is processed with different camera parameters using [ISET Camera Designer](https://github.com/iset/isetcam/wiki). These images are resized to 32x32x3 and serve as the (LR) input to the generator. The target HR images are the original ones that are not processed. A total of four models were trained:  
Model\_SR: SRGAN model that does super-resolution only  
Model\_SR\_Color: SRGAN model that does super-resolution and color correction  
Model\_SR\_Pixel: SRGAN model that does super-resolution and restores spatial resolution due to reduction of system MTF  
Model\_SR\_Deblur: SRGAN model that does super-resolution and deblur

DataSet

Training

1800 cat and dog images were downloaded from Flickr and Pixabay. These images were processed using three different camera settings using ISET Camera Designer developed by David Cardinal, 2020, ''F20-PSYCH-221'', Stanford University, under the framework of [ISETCAM](https://github.com/iset/isetcam/wiki)

[](https://user-images.githubusercontent.com/65942005/100526342-b8bc8380-317c-11eb-83d6-06f0e971b66e.jpg)

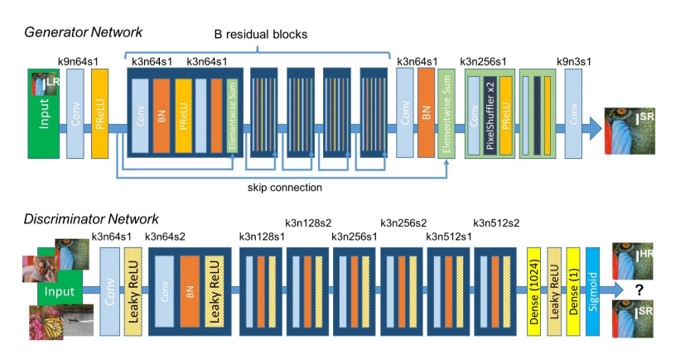
Camera setting 1- F/4; aperture diameter: 1mm, pixel size: 1umx1um; under exposing; default image processor. To produce images that require color correction as we use the default image processor while under exposure, these images have a warm tone  
Camera setting 2- F/4; aperture diameter: 1mm, pixel size: 25umx25um. In general, a large pixel size is desirable because it results in a higher dynamic range and signal-to-noise ratio. However, the reduction in spatial resolution and system MTF introduce severe pixelated effect  
Camera setting 3- F/22; aperture diameter: 0.176mm; pixel size: 1umx1um. Images look much blurry compared with the original images because they become diffraction-limited at a small aperture value

Testing

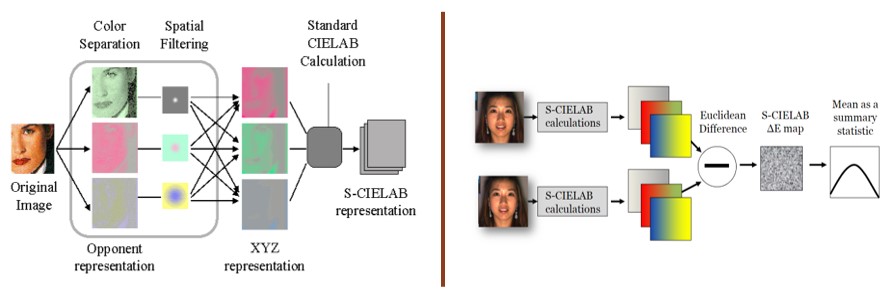
About 100 images, not necessarily cats and dogs, are used for evaluation. As an input to four trained models (Model\_SR, Model\_SR\_Color, Model\_SR\_Pixel, and Model\_SR\_Deblur), these images went through the same camera settings and resized to 32x32x3 similar to the training dataset.

SRGAN Model

In this work, we use Tensorflow implementation of ["Photo-realistic single image super-resolution using a generative adversarial network."](https://arxiv.org/abs/1609.04802). Small modifications are done as follows:  
PixelShuffler x2: This is for feature map upscaling. We use 2x ‘deconv2d’ Keras built-in function for implementation.  
PRelu(Parameterized Relu): PRelu introduces a learnable parameter that makes it possible to adaptively learn the negative part coefficient. We use Relu as an activation function for simplicity.

[](https://user-images.githubusercontent.com/65942005/100526325-975b9780-317c-11eb-8528-1785a6659b10.jpg)  
Generator and Discriminator in SRGAN.

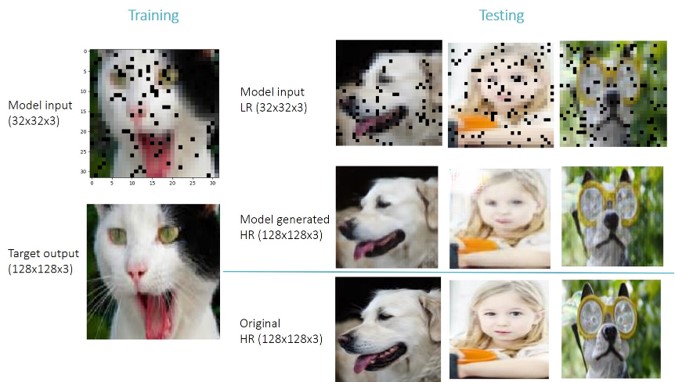
S-CIELAB representation

We use it as an evaluation matrix of image quality. [S-CIELAB](http://scarlet.stanford.edu/~brian/scielab/introduction.html) is an extension of the CIE L*a*b\* DeltaE color difference formula and provides human spatial sensitivity difference between a reference and a corresponding test image. The key components of calculating S-CIELAB representation include color transformation and the spatial filtering steps that simulate the human visual system before the standard CIELAB Delta E calculations.  
[](https://user-images.githubusercontent.com/65942005/100526525-7bf18c00-317e-11eb-9b3f-887526b5aa36.jpg)

Result

Effect of missing pixels

SRGAN model can deal with missing/noise pixels (about 10% in our experiment) and generate HR images that not only have smooth edges but also restore the details.

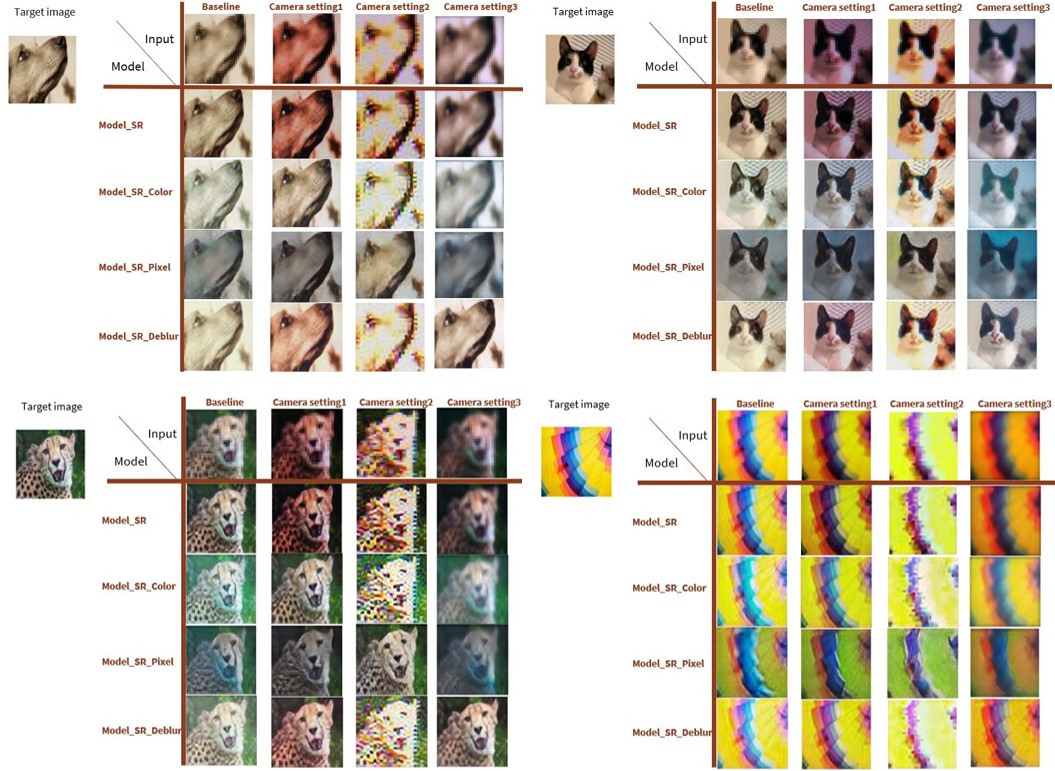
[](https://user-images.githubusercontent.com/65942005/100526718-71d08d00-3180-11eb-8f10-63aa751a9317.jpg)

HR images from trained SRGAN models

SRGAN model can be trained to perform super-resolution and image enhancement (including color correction and de-blurring) simultaneously

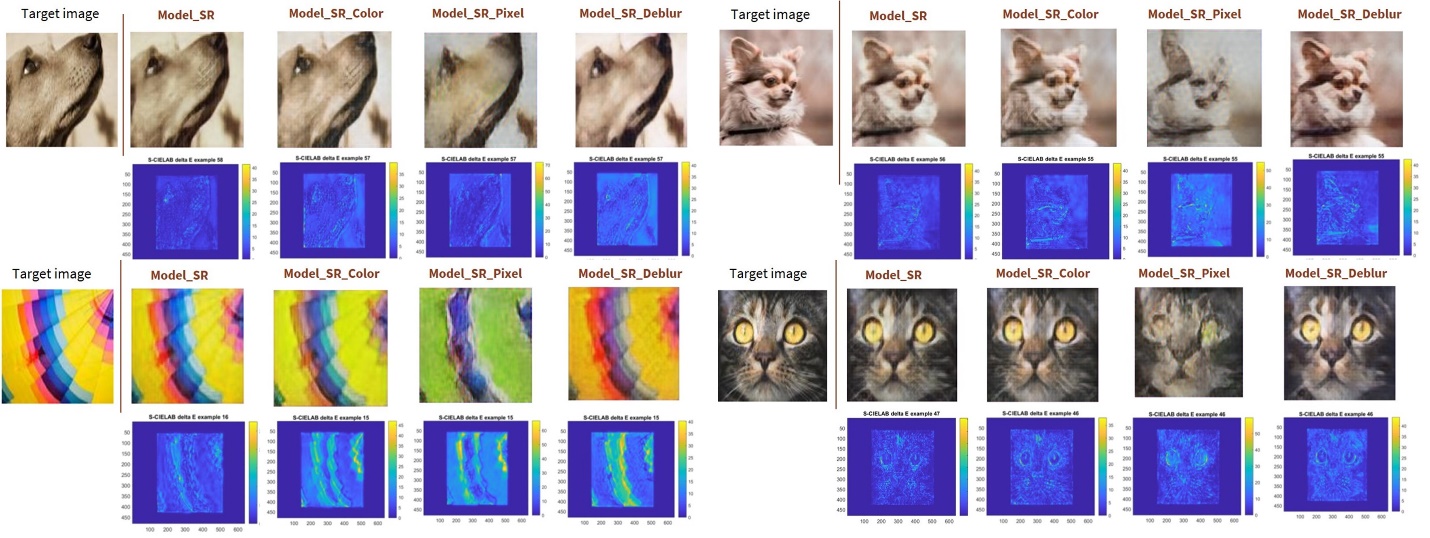
The following matrices summarize the results from all four trained SRGAN models. The 'input' rows show the input LR images (original and processed images using three different camera settings as described in the method section). Each cell in the matrix represents the result of HR images generated by a given model from a given input LR. For example, in each 4x4 matrix, the 3rd row, 2nd column image is the result generated by Model\_SR\_Pixel using LR images processed by camera setting1. The images in the diagonal have higher qualities (closer to the target images) compared with the off-diagonal images because these are how the models were originally trained.

Model\_SR: this model is trained to perform super-resolution only. As expected, the result looks simply upscaled without changing other characteristics of the input images such as color and focus.  
Model\_SR\_Color: the outline and details look similar to Model\_SR. Also, because this model is trained to do color correction, the color tone is different between the input and output images (becomes 'brighter' in general).  
Model\_SR\_Pixel: unlike Model\_SR and Model\_SR\_Color, the result from this model looks relatively unnatural. However, when the input image is from camera setting 3 (reduction of system MTF due to large sensor pixel), the resulting HR image improved a lot - it learns how to restore spatial resolution to some extent.  
Model\_SR\_Deblur: This model successfully learned how to de-blur. It is also interesting that all of its output images seem to remain in good focus regardless of whether the input image is within/or out of focus.

[](https://user-images.githubusercontent.com/65942005/100526346-beb26480-317c-11eb-9a84-55fcf1beb2a6.jpg)

S-CIELAB delta E maps

The S-CIELAB delta E maps show the difference between the target images and the model-generated images. Consider these differences as 'residue' (mainly at the edges) that the model can improve, it might be interesting in future work to replace/add S-CIELAB representation to the generator's loss function. The reason is, that one of the major changes that a more advanced version of the SRGAN model (called Enhanced SRGAN, SRGAN) has made is to use feature maps before activation for calculating content loss. As we extract feature maps from a relatively deep layer in the VGG19 layer, some of the features after activation become inactive and contain less information. It is possible that S-CIELAB can provide additional information, especially from a human spatial sensitivity point of view, to the generator during training and create a new class of super-resolution images that focus more on how accurate the reproduction of color is to the original when viewed by a human observer.

[](https://user-images.githubusercontent.com/65942005/100526352-cffb7100-317c-11eb-83a0-1e664f367add.jpg)

Proposed future work for model improvement

We proposed in future work to use S-CIELAB delta E maps as generator loss functions to incorporate human spatial color sensitivity during training. This could enable a new class of super-resolution images that focus more on the reproduction of color patterns when viewed by a human observer.

# SwinIR: Image Restoration Using Swin Transformer

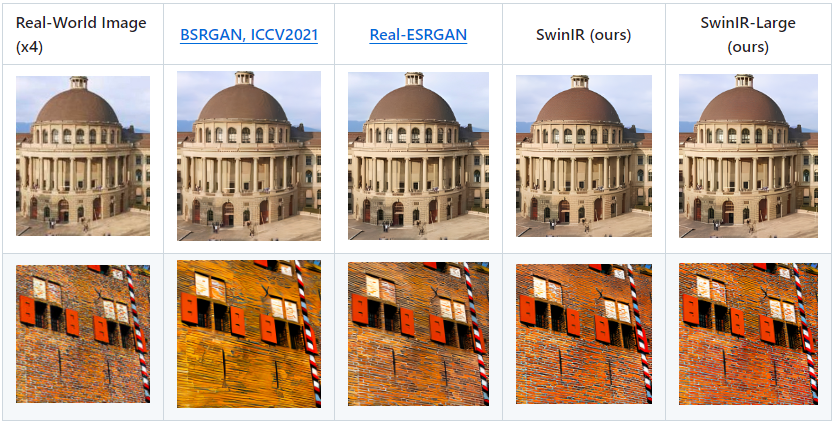


Image restoration is a long-standing low-level vision problem that aims to restore high-quality images from low-quality images (e.g., downscaled, noisy and compressed images). While state-of-the-art image restoration methods are based on convolutional neural networks, few attempts have been made with Transformers which show impressive performance on high-level vision tasks. In this paper, we propose a strong baseline model SwinIR for image restoration based on the Swin Transformer. SwinIR consists of three parts: shallow feature extraction, deep feature extraction and high-quality image reconstruction. In particular, the deep feature extraction module is composed of several residual Swin Transformer blocks (RSTB), each of which has several Swin Transformer layers together with a residual connection. We conduct experiments on three representative tasks: image super-resolution (including classical, lightweight and real-world image super-resolution), image denoising (including grayscale and color image denoising) and JPEG compression artifact reduction. Experimental results demonstrate that SwinIR outperforms state-of-the-art methods on different tasks by up to 0.14~0.45dB, while the total number of parameters can be reduced by up to 67%.

