Single Image Super Resolution of Textures via CNNs

Andrew Palmer

What is Super Resolution (SR) ?

Simple: Obtain one or more high-resolution images from one or more low-resolution ones

Low resolution image

High resolution image
8x Upscaling

Many, many applications

Biometric recognition, including resolution enhancement for faces fingerprints, and iris images

Medical diagnosis

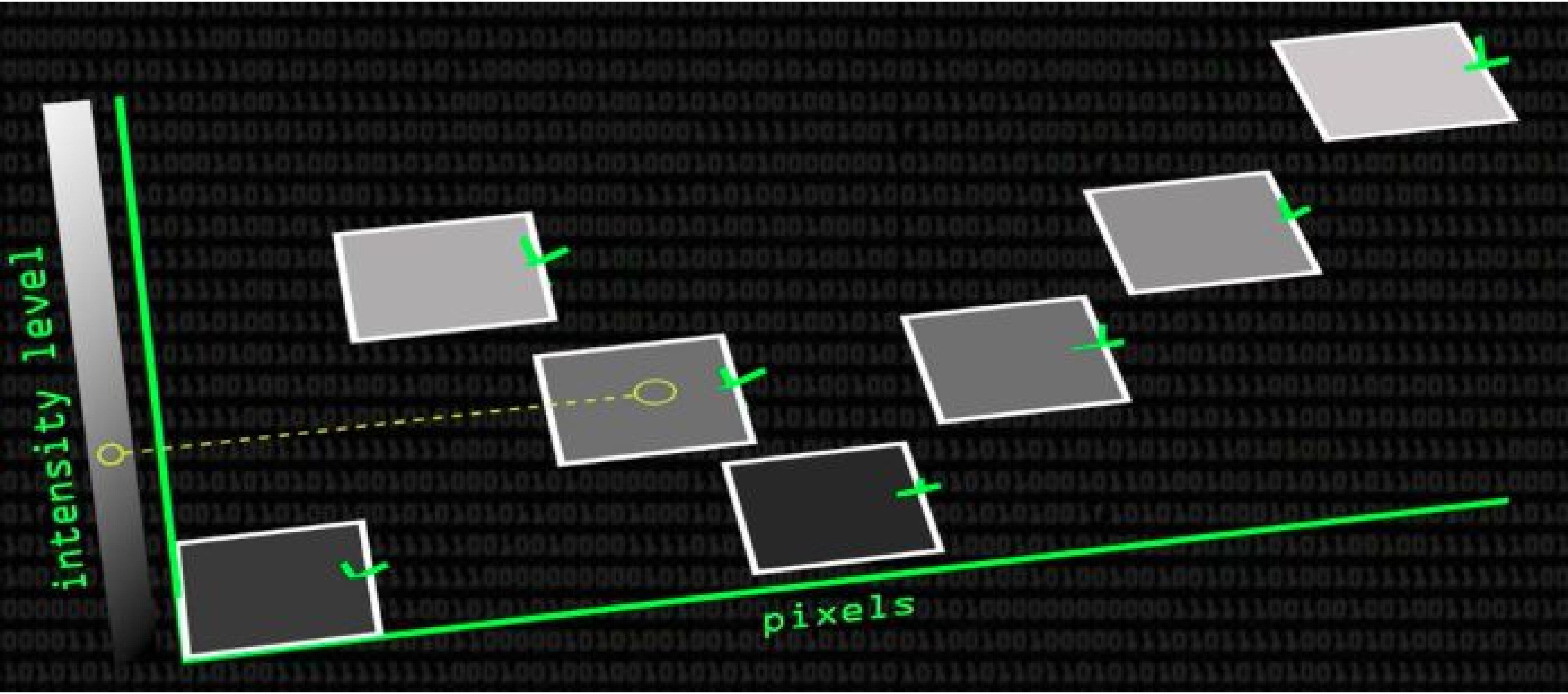
Image compression

Text enhancement; preprocessing step for optical character recognition

SR has diverse challenges

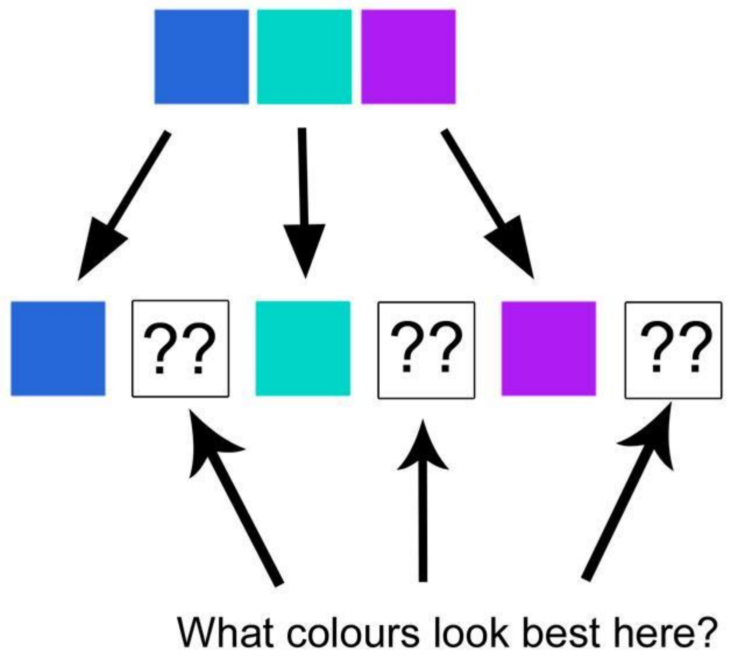
Underspecified problem; many solutions

No solid theory for determining what is 'good' enhancement. If it looks good, it looks good

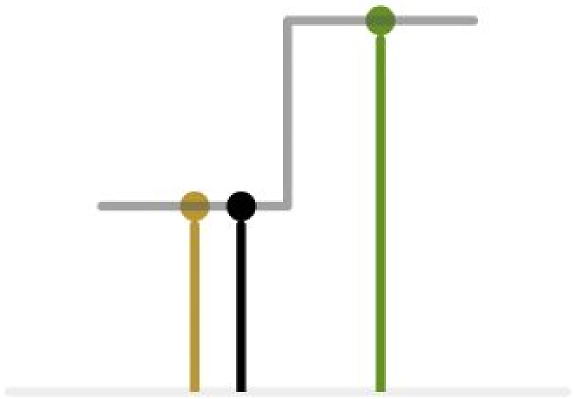


Computerphile: Bicubic Interpolation - https://goo.gl/wAvNtM

Visualizing the problem for 2x upscaling

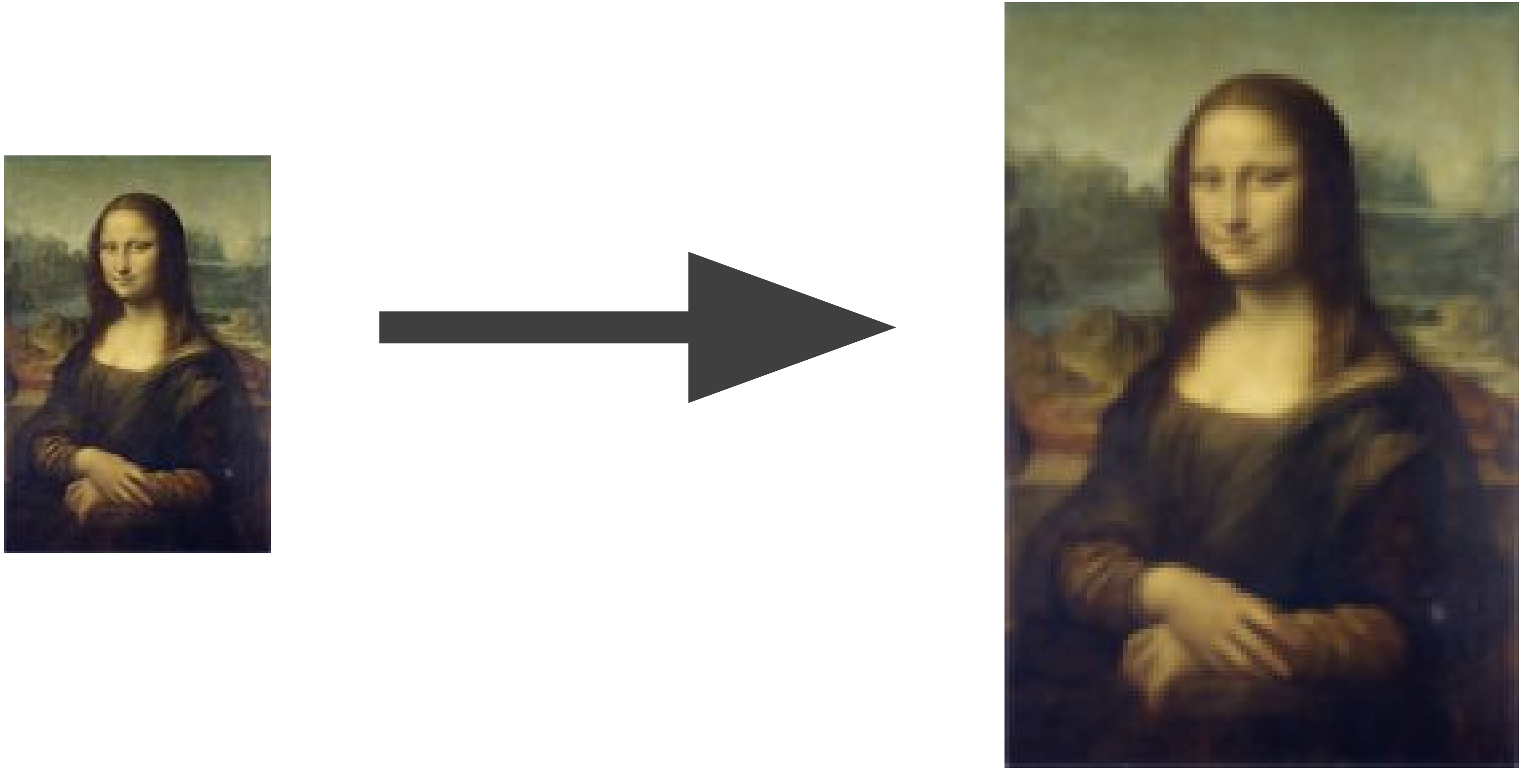


Nearest Neighbour Upscaling

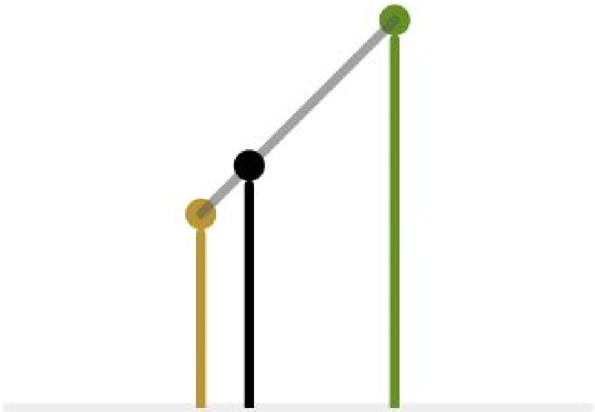


1D nearest-
neighbour

Nearest Neighbour Upscaling

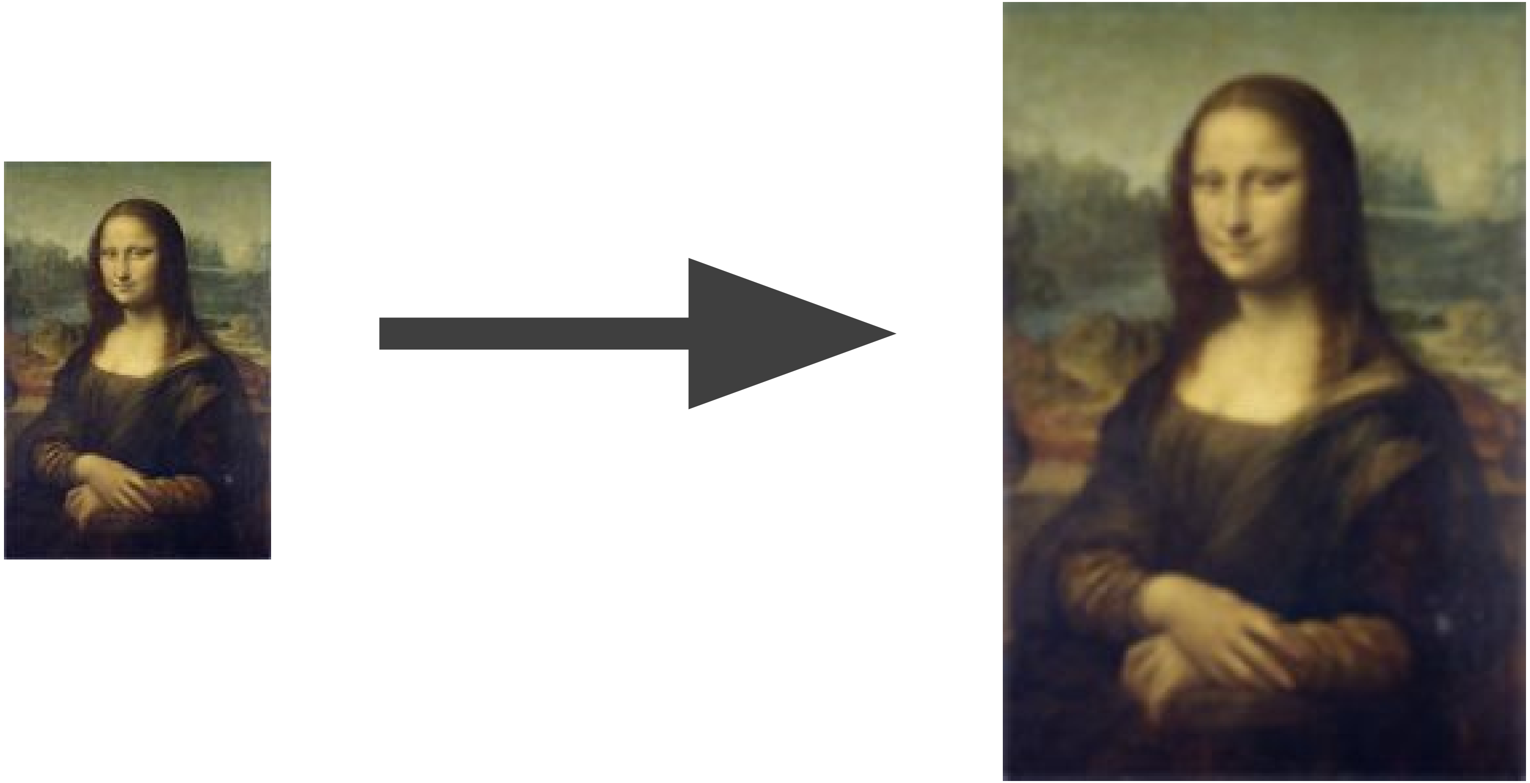


Linear Upscaling

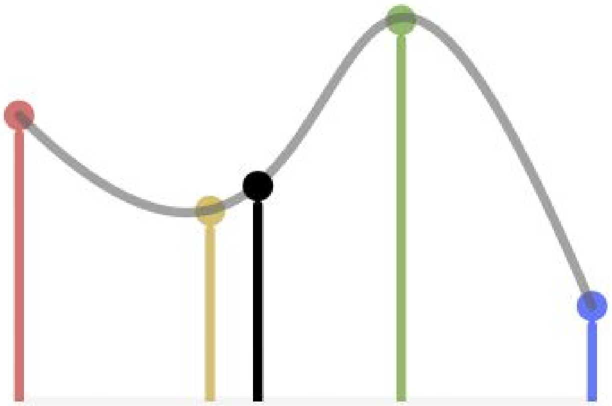


Linear

Linear Upscaling

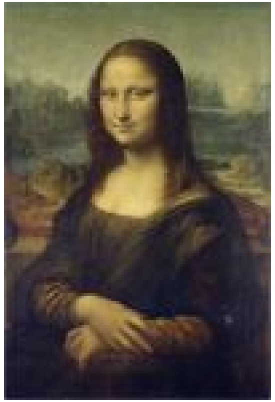


Cubic Upscaling



Cubic

Cubic Upscaling



Classical interpolation; most seen in practice

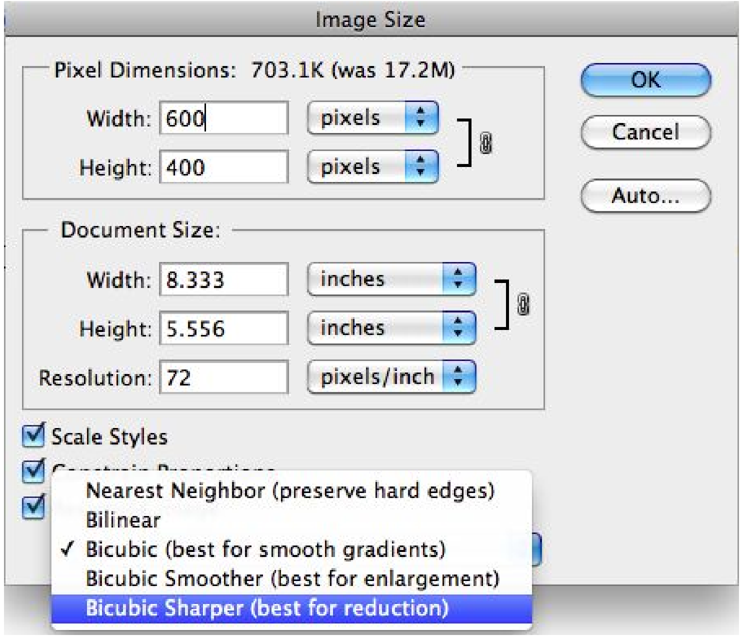


Image resizing options in Photoshop

Other traditional methods

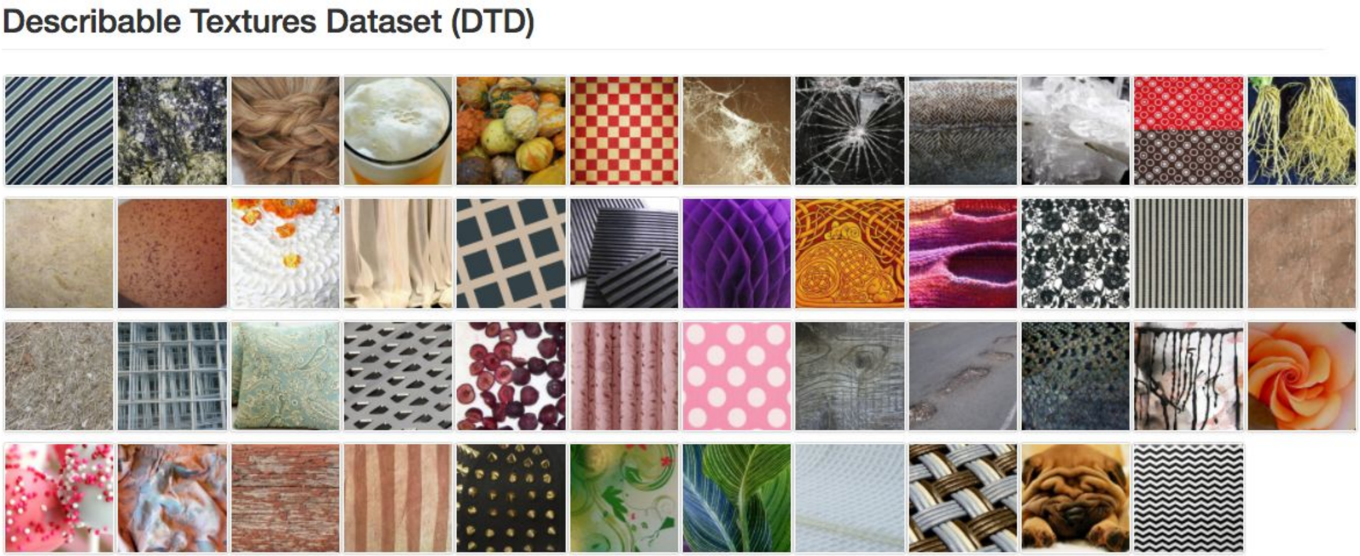
Gaussian Smoothing, Wiener, Median filters (good at denoising)

Sharpening by amplifying existing image details (need to ensure that noise isn't amplified)

Texture Super Resolution

Results could use some improvement in texture quality

The Describable Textures Dataset

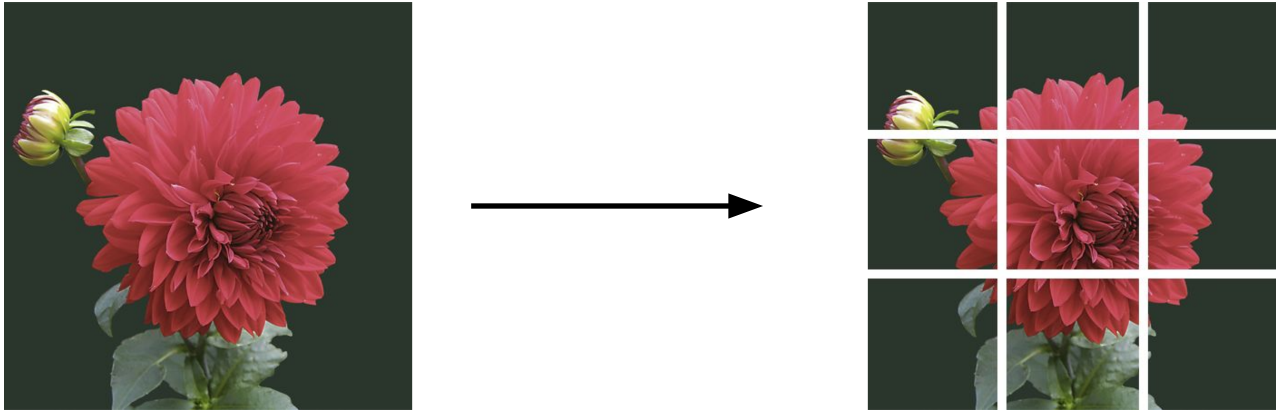


The Describable Textures Dataset (DTD): textural images in the wild, annotated with a series of human-centric attributes, inspired by the perceptual properties of textures

5640 images, split into 47 classes, e.g. blotchy, polka dot, grainy

Describable Textures Dataset - https://goo.gl/YLtw2S

Setup: Data synthesis by cropping

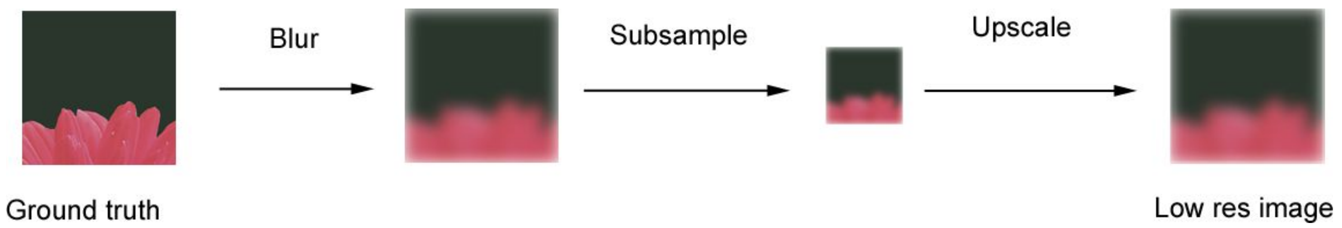


Low res (LR) image synthesis: crop ground truth images into several sub-images

Setup: Generate a LR image from each patch

For each HR cropped image:

1. Apply a Gaussian convolution
2. Sub-sample by the upscaling factor (produces a smaller image)
3. Bicubic upscaling



Input data going into the CNN

Total number of generated patches: 2,436,258

A subset (500) of the original images was used:

Training set: 127,744 patches

Validation: 12,544 patches

Common Evaluation Metrics

Peak Signal to Noise (PSNR); >=30 dB for restoration = very good

Structural Similarity (SSIM)

Subjective perception

Time efficiency

$$
\begin{aligned}
PSNR &= 10 \cdot \log_{10}\left(\frac{MAX_I^2}{MSE}\right) \\
&= 20 \cdot \log_{10}\left(\frac{MAX_I}{\sqrt{MSE}}\right) \\
&= 20 \cdot \log_{10}(MAX_I) - 10 \cdot \log_{10}(MSE)
\end{aligned}
$$

Setup: Hyperparameters

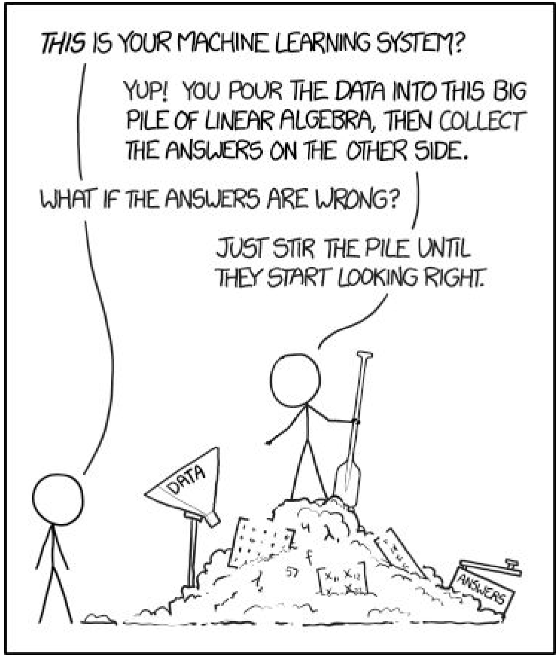
Activations: ReLU, ELU, tanh (popular choice for SR)

Loss function: MSE (old, but most common for SR)

Optimizer: ADAM

Iterations: 20,000

Learning rate: 10^{-3}



https://xkcd.com/1838/

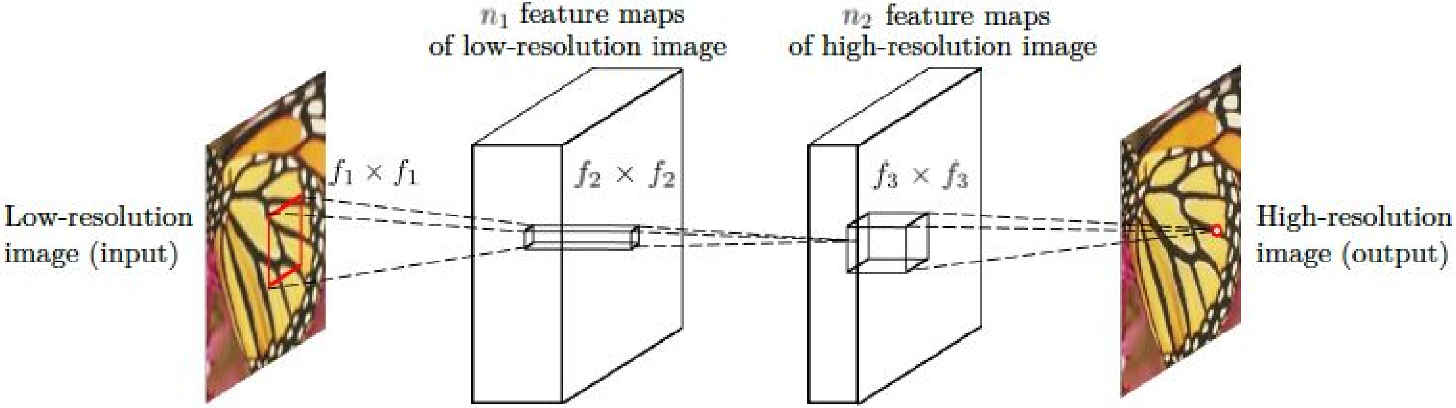
Setup: Misc.

Over the RGB colour space

Zero padded edges; each layer outputs same dimensions

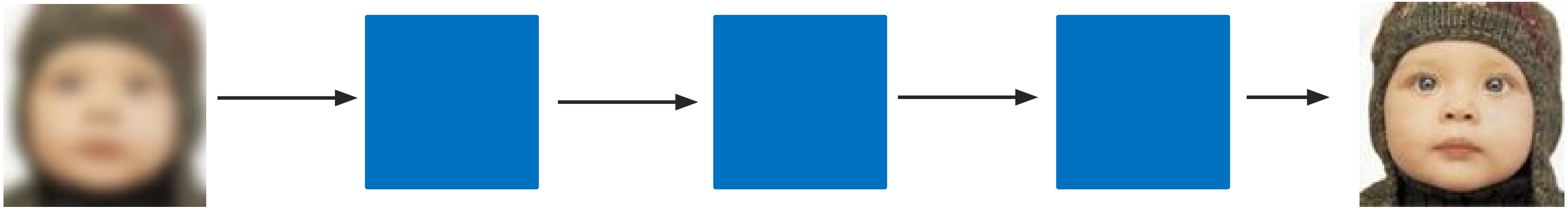
No max pooling layers (might be bad for denoising and SR tasks https://arxiv.org/abs/1511.04491)

SRCNN



The effectiveness of deeper structures for super resolution is found not as apparent as that shown in image classification

CNN Architecture

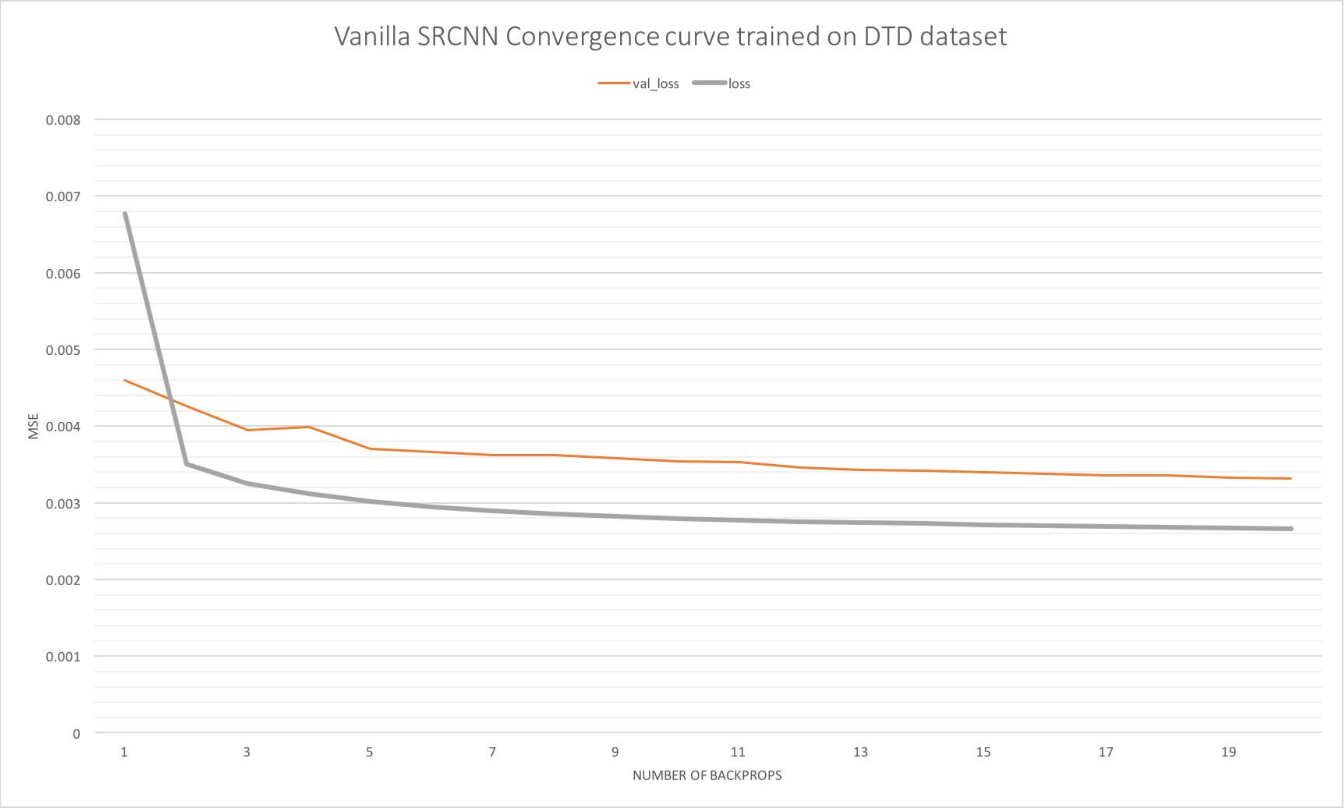


Conv
Activation: Relu
Filter size: 9x9
Filters: 64
Padding: same
Dim (out): 32 x 32

Conv
Activation: Relu
Filter size: 3x3
Filters: 32
Padding: same
Dim (out): 32 x 32

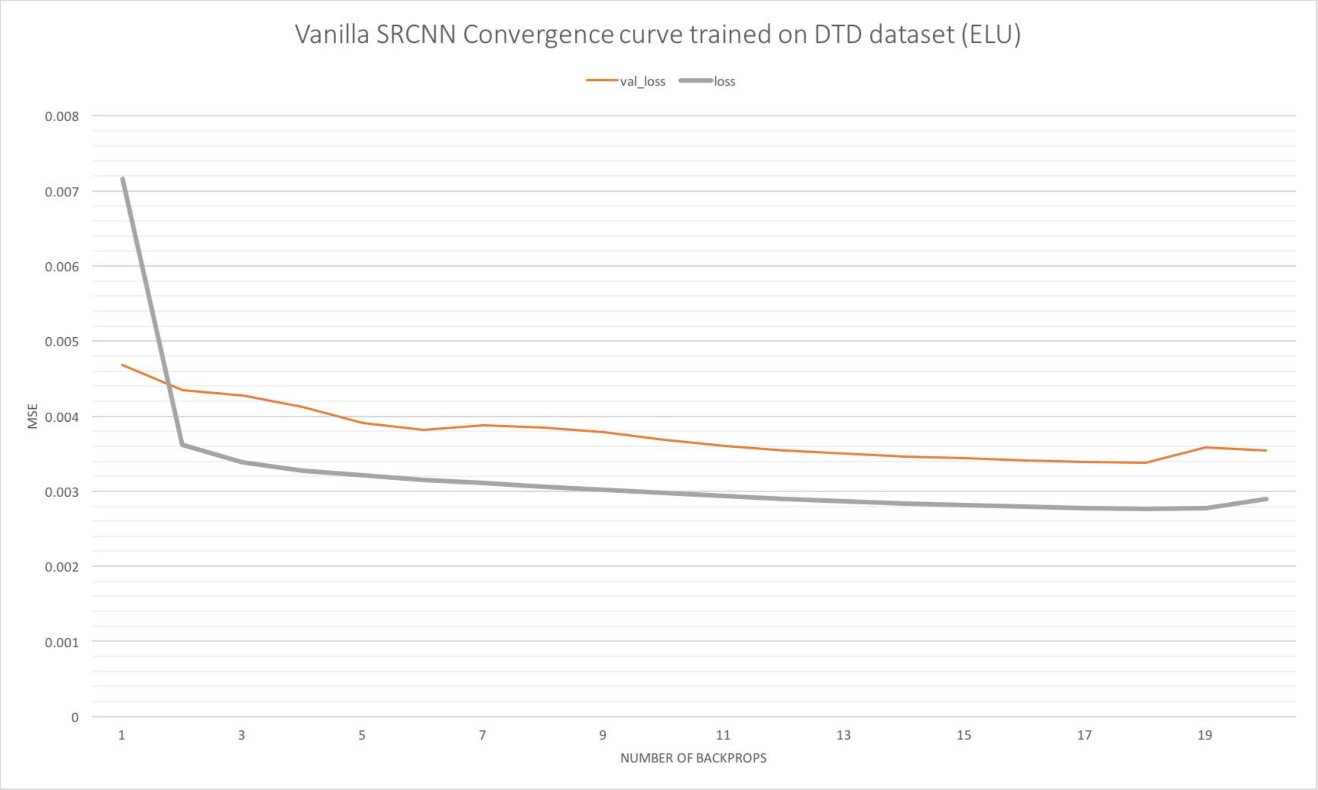
Conv
Activation: Relu
Filter size: 5x5
Filters: 32
Padding: same
Dim (out): 32 x 32

Vanilla SRCNN (RELU)



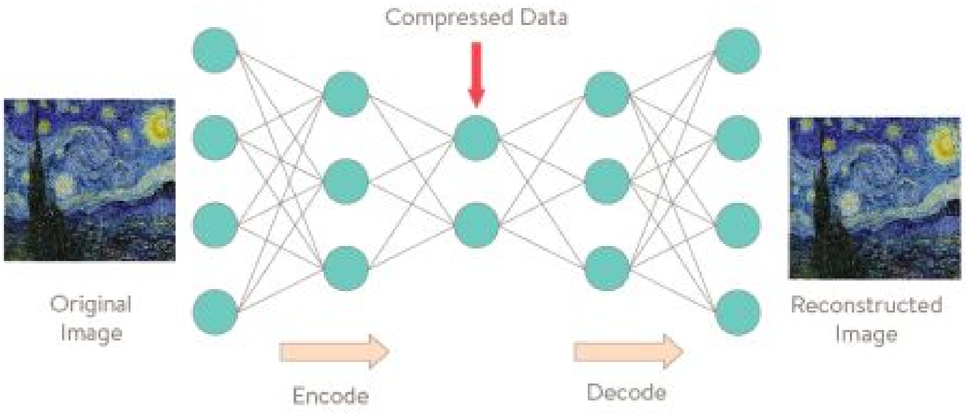
Filter sizes: 9-1-5
Max PSNR: 25.77

Vanilla SRCNN (ELU)



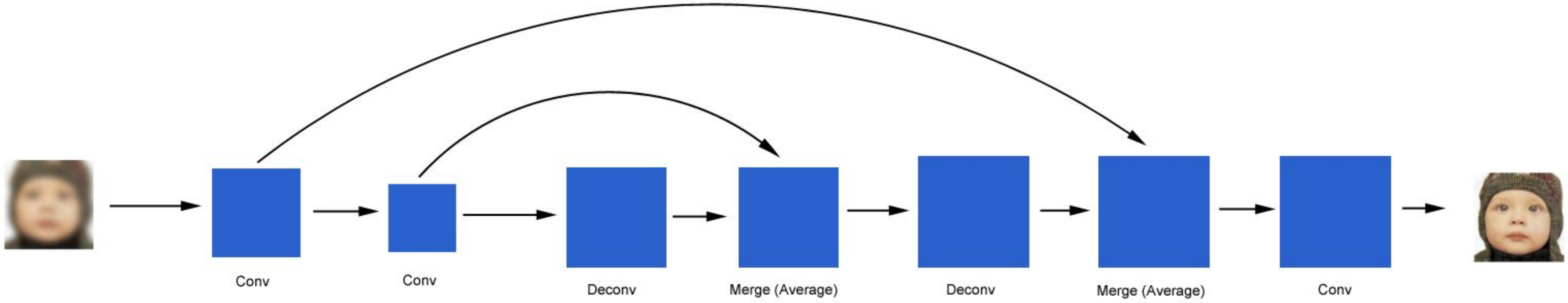
Filter sizes: 9-1-5
Max PSNR: 24.52

Autoencoder



Source: https://goo.gl/R7itkL

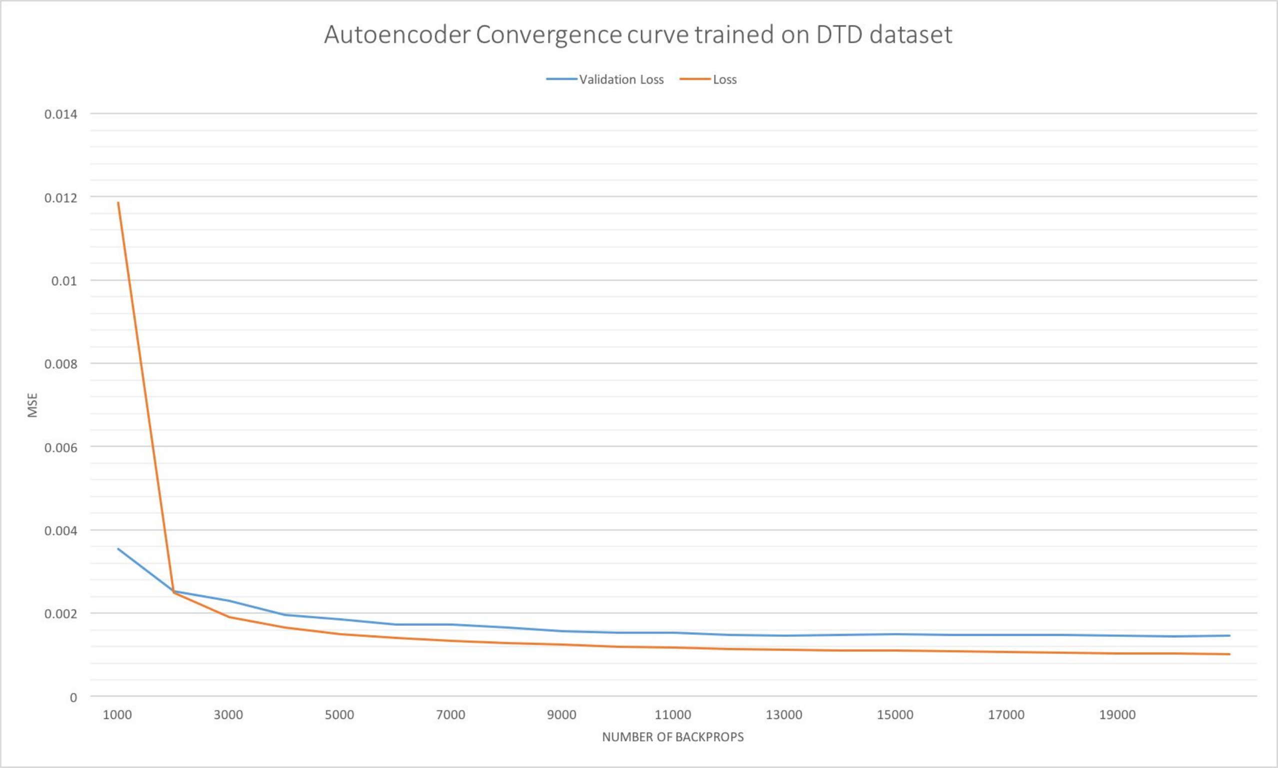
Autoencoder with skip connections



All filter sizes: 3x3
Number of filters: 64 (except the last Conv layer)

- Symmetric Skip connections
  Helps on recovering clean images
  Converges much faster and attains a higher-quality local optimum.

Autoencoder



Max PSNR: 28.4

Other results

Tanh: Max PSNR: 24.68

Filter sizes: 9-3-5 max PSNR: 26.03

Todo: Denoising autoencoder (at the reconstruction phase)

Standard dataset using same configuration max PSNR: 30.49

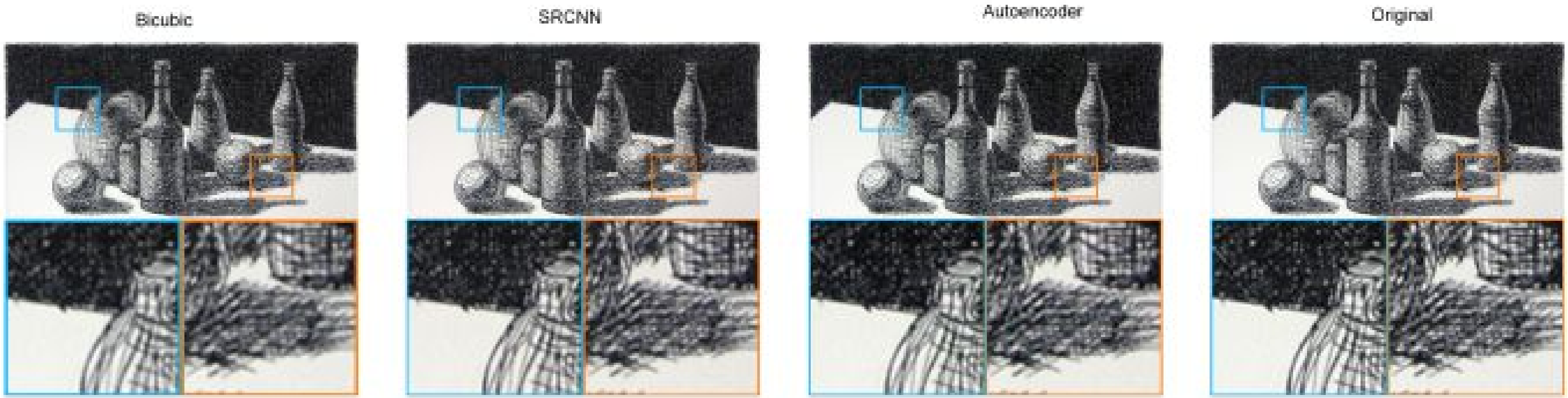
Todo: GAN

From Twitter

PSNR values **worse** than bicubic interpolation

Perceptual quality is the best by far

Results for a sample cross-hatched texture



Fibrous texture; 2x upscaling

Bicubic

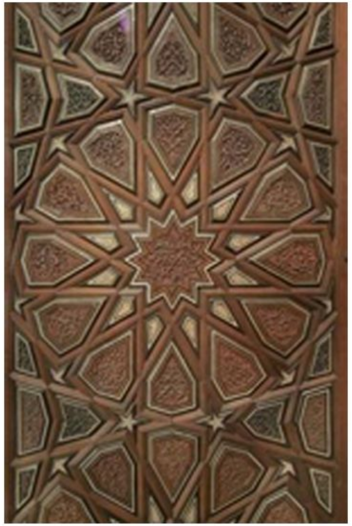
SRCNN

Autoencoder

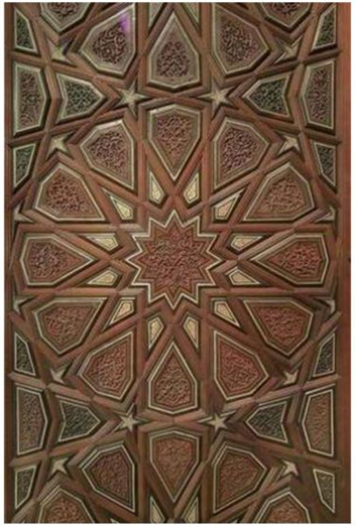
High Resolution

Interlaced texture; 2x upscaling

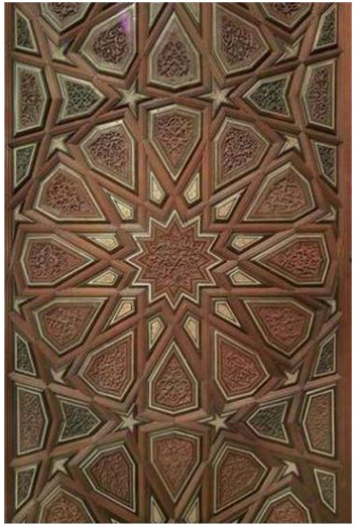
Bicubic



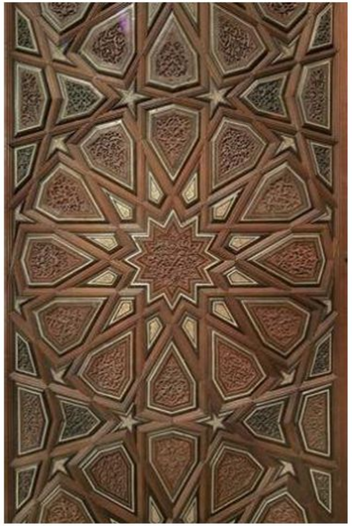
SRCNN



Autoencoder



High Resolution



Striped texture; 2x upscaling

Bicubic

SRCNN

Autoencoder

High Resolution

Porous texture; 2x upscaling

Bicubic

SRCNN

Autoencoder

High Resolution

Todo

More in-depth metrics on the performance against standard datasets

Evaluate performance on larger scaling factors, i.e. 4x, 8x etc.

Fix the GAN

(Maybe) Try a different down-sampling technique (some argue against bicubic preprocessing)

(Probably) Load a pre-trained model e.g. ResNet

Conclusion and Future work

Standard, shallow CNNs work alright for single image texture super resolution

Shallow autoencoders work better than CNNs

Quantify the performance for each texture class

Evaluate how effective these models are for classification

References

Boosting Optical Character Recognition: A Super-Resolution Approach - https://arxiv.org/abs/1506.02211

Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network - https://arxiv.org/abs/1609.04802

Image Super-Resolution Using Deep Convolutional Networks - https://arxiv.org/abs/1501.00092

Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion - http://www.jmlr.org/papers/volume11/vincent10a/vincent10a.pdf

Super-Resolution via Deep Learning - https://arxiv.org/pdf/1706.09077.pdf

Image Restoration Using Convolutional Auto-encoders with Symmetric Skip Connections - https://arxiv.org/pdf/1606.08921.pdf

Questions, suggestions, ideas?