Quantitative Analysis of Super-Resolution on Biomedical Images Pathology

M.Tech Thesis

Mayank Sharma

Department of Computer Science and Engineering
Indian Institute of Technology Jammu

June 25, 2021

भारतीय प्रौद्योगिकी
संस्थान जम्मू
INDIAN INSTITUTE OF
TECHNOLOGY JAMMU

Single Image super-resolution & Deep Learning Techniques

- Single image super-Resolution (SR) is the process of generating a higher resolution image from lower resolution (LR) images using a learning-based mechanism.
- High resolution images are desired in many applications such as satellite and aerial imaging, medical imaging, facial image analysis, text image analysis, sign and number plate reading, biometrics recognition, video surveillance, astronomy etc
- In recent years the field of biomedical image super-resolution has witnessed remarkable progress using deep learning (DL) techniques.

Motivation

- In real world, there is a considerable imbalance between samples of normal and diseased pathology images.
- Existing deep-learning-based single-image super-resolution works well over normal pathology images.
- However, in real-world clinical settings, diseased pathology images are of utmost importance for diagnosing physicians.
- Now, the question is do existing deep-learning based methods perform equally good over diseased pathology images? And can we deploy such algorithms in real-world clinical setting?

Problem Statement

- We attempt to quantitatively validate the performance of various existing deep-learning-based medical image super-resolution methods over diseased biomedical pathology images of different modalities (i,e MRI, X-Ray, histopathology and Retinopathy).

Interpolation Based Methods

- In medical imaging interpolation techniques are required for image generation, as well as are needed in image post-processing.
- Interpolation based methods such as Edge-directed interpolation (EDI) which utilize the local statistical and geometrical properties to interpolate the unknown pixel values [4], [10], [8] and Improved New Edge directed Interpolation (iNEDI) method [5] which uses a hybrid approach is used to switch between bilinear interpolation and covariance-based interpolation were quite popular for high-resolution images.
- Interpolation tend to over-smooth images with jagged and choppy effect. Although these methods are easy and fast to comprehend and operate, they are poor at processing complex and edge textures.

Deep Learning Based Methods

- Researchers see promising potential in deep-learning-based methods for medical image super-resolution with recent advancements in deep learning and image acquisition techniques propelled by the drive of big data.
- Existing deep-learning methods such as 4D-PET [6] Super-Resolution Using Deep Convolutional Networks (SRCNN) [2], Efficient Sub-Pixel Convolutional Neural Network (ESPCN) [9], Enhanced Deep Residual Networks for Single Image Super-Resolution (EDSR) [1], Wide Activation for Efficient and Accurate Image Super-Resolution (WDSR) [7] etc. have shown promising results over normal images.

Research Gap

- Many existing deep-learning methods show promising results in super-resolution over normal biomedical images.
- As per our knowledge, till the present date, the performance of super-resolution over pathological images has not been well analysed.

Datasets Used

- For this experiment, we have used four medical image datasets consisting of approximately 3000 images each. We have used 2000 images for training 800 for validation and we have tested our results on 200 images.
  - Chest X-Ray Pneumonia Dataset `https://data.mendeley.com/datasets/rscbjbr9sj/2`
  - Brain Tumor Dataset `https://figshare.com/articles/dataset/brain_tumor_dataset/1512427`
  - Breast Histopathology Image Dataset `http://gleason.case.edu/webdata/jpi-dl-tutorial/IDC_regular_ps50_idx5.zip`
  - Diabetic Retinopathy Image Dataset `https://www.kaggle.com/c/diabetic-retinopathy-detection/data`

Super-resolution Techniques

- Super-Resolution Using Deep Convolutional Networks (SRCNN) [10]
- Efficient Sub-Pixel Convolutional Neural Network (ESPCN) [1]
- Single Image Super-Resolution Using a Generative Adversarial Network (SRGAN) [3]
- Wide Activation for Efficient and Accurate Image Super-Resolution (WDSR) [2]
- Enhanced Deep Residual Networks for Single Image Super-Resolution (EDSR) [4]

Image Quality Assessment (IQA) Metrics

- Mean Absolute Error (MAE)
- Mean Absolute Percentage Error (MAPE)
- Mean Square Error (MSE)
- Root Mean Square Error (MSE)
- Peak Signal to Noise Ratio (PSNR)
- Structural Similarity Index (SSIM)
- Multi-Scale Structure Similarity Index (MS-SSIM)
- Spatial Corelation Coefficient (SCC)

Structure Similarity Index (SSIM)

- The Structural Similarity Index (SSIM) [11] is a perceptual metric that quantifies image quality degradation.
- SSIM measures the perceptual difference between two similar images.
- It is a complete reference metric that requires two images from the same image capture a reference image and a processed image.
- The structural similarity index (SSIM) metric extracts three critical characteristic features of an image:
  - Luminance or Brightness
  - Contrast
  - Structure

Luminance (Brightness)

- Luminance or Brightness can be defined as average of all the pixel values and it is denoted by μ.

$$\mu_y = \frac{1}{N} \sum_{i=1}^{N} y_i$$

- Here y_i is the i-th pixel value of the image y. And N is the total number of pixels.

Contrast

- Contrast is defined as standard deviation of all the pixel values. It is denoted by σ

$$\sigma_y = \left(\frac{1}{N-1} \sum_{i=1}^{N} (y_i - \mu_y)^2 \right)^{\frac{1}{2}}$$

- Here y_i is the i-th pixel value of the image y, μ is the mean of pixel values And N is the total number of pixels.

- The structural comparison is made by dividing the input image with the standard deviation so that the resulted image has unit standard deviation, which allows for a more robust comparison.

$$(\mathrm{y} - \mu_y) / \sigma_y$$

Luminance Comparison Function

- Luminance comparison is defined by the function $L(O, S)$ which is shown below. μ represents the mean of a given image. O is the original image and S is the super-resolved image being compared.

$$L(\mathrm{O}, \mathrm{S}) = \frac{2\mu_o\mu_s + M_1}{\mu_o^2 + \mu_s^2 + M_1}$$

- where M_1 is a constant to ensure stability in case if the denominator becomes 0. M_1 is given by,

$$M_1 = (P_1 R)^2$$

- R is the dynamic range for pixel values (we set it as 65535 since we are dealing with standard 16-bit images). P_1, P_2 are just normal constants.

Contrast Comparison Function

- Contrast comparison is defined by a function $C(O, S)$ which is shown below. σ denotes the standard deviation of a given image. O is the original image and S is the super-resolved image being compared.

$$C(\mathrm{O}, \mathrm{S}) = \frac{2\sigma_o\sigma_s + M_2}{\sigma_o^2 + \sigma_s^2 + M_2}$$

- where M_2 is given by,

$$M_2 = (P_2 R)^2$$

M_1

Structure Comparison Function

- Structure comparison is defined by the function $S(O, S)$, which is shown below. σ denotes the standard deviation of a given image. O is the original image and S is the super-resolved image being compared.

$$S(\mathrm{O}, \mathrm{S}) = \frac{\sigma_{os} + M_3}{\sigma_o \sigma_s + M_3}$$

- Here, σ (os) is defined as,

$$\sigma_{os} = \frac{1}{N-1} \sum_{i=1}^{N} (o_i - \mu_o)(s_i - \mu_s)$$

SSIM Formula

- And finally, the SSIM score for two images being compared is given by,

$$\text{SSIM}(o, s) = [L(o, s)]^{\alpha} \cdot [C(o, s)]^{\beta} \cdot [S(o, s)]^{\gamma}$$

- In the above SSIM score $\alpha > 0$, $\beta > 0$ and $\gamma > 0$ denote the relative importance of each metrics. In order to simplify the expression, if we assume, $\alpha = \beta = \gamma = 1$ and $M_3 = M_2/2$, we get the below simplified equation

$$\text{SSIM}(o, s) = \frac{(2\mu_o\mu_s + M_1)(2\sigma_{os} + M_2)}{(\mu_o^2 + \mu_s^2 + M_1)(\sigma_o^2 + \sigma_s^2 + M_2)}$$

SR Model Flow

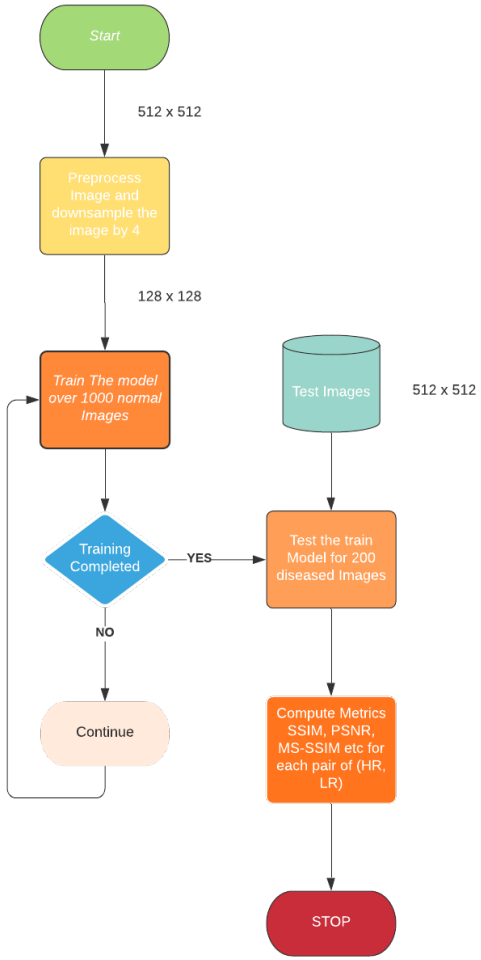


Figure 1: SR Model Flow

Experimental Results for Chest X-Ray Pneumonia Dataset

- For the chest x-ray pneumonia dataset, we have generated results for five super-resolution models.
- Trained ESPCN and SRCNN models on 1000 standard chest x-ray normal images and evaluated the performance of the trained models on 200 diseased chest x-ray images.

Experimental Results for Chest X-Ray Pneumonia Dataset

Figure 2: Normal Chest X-Ray SR Result Example

SSIM Score - 0.93883

Figure 3: Diseased Chest X-Ray SR Result Example

SSIM Score - 0.89905

Experimental Results for Chest X-Ray Pneumonia Dataset

Table 1: ESPCN Super-resolution Results

	Presence of Pneumonia in Images		Absence of Pneumonia in Images	
	mean	std	mean	std
MAE	0.0003	0.0006	0.0002	0.0003
MPE	0.035	0.067	0.029	0.032
MSE	13.3497	11.1276	20.2264	27.824
RMSE	3.4799	1.1134	4.1822	1.6538
PSNR	37.6666	2.4561	36.0589	2.1595
SSIM	0.9295	0.0357	0.9401	0.0537
MS-SSIM	0.9819	0.0085	0.9854	0.0102
SCC	0.1506	0.0533	0.1747	0.0476

Table 2: SRCNN Super-resolution Results

	Presence of Pneumonia in Images		Absence of Pneumonia in Images	
	mean	std	mean	std
MAE	0.000422	0.00077	0.000318	0.00029
MPE	0.04223	0.07776	0.03187	0.0296
MSE	118.543	33.0124	110.0492	40.5618
RMSE	10.7861	1.4843	10.3481	1.7219
PSNR	27.5552	1.1922	27.9444	1.3721
SSIM	0.8551	0.045097	0.8956	0.05589
MS-SSIM	0.9493	0.01431	0.95032	0.0129
SCC	0.00955	0.0045	0.00972	0.00402

Experimental Results for Brain Tumor Dataset

- For the brain tumour dataset, we have generated results for two super-resolution models.
- Trained ESPCN and SRCNN models on 1000 standard brain tumour images and evaluated the performance of the trained models on 200 brain tumour images separately for brain and tumour images each.
- Did pre-processing for all the 200 brain tumour images before we could perform the image quality assessment. We separated the tumour region from the brain region using polygon curve approximation contour algorithm.

Experimental Results for Brain Tumor Dataset

Figure 4: Tumour Region SR Result Example

Figure 5: Brain Region SR Result Example

Table 3: ESPCN Super-resolution Results

	Tumor Region		Brain Region	
	mean	std	mean	std
MAE	0.05051	0.01603	0.00025	0.000006
MPE	5.05168	1.603	0.025	0.0063
MSE	0.00258	0.0015	21.5182	8.1236
RMSE	0.0025	0.0015	4.5576	0.8636
PSNR	26.6957	2.7794	83.3143	1.6844
SSIM	0.9292	0.0364	0.9999	2.9888
MS-SSIM	0.9823	0.0019	1	0
SCC	0.9408	0.071	0.3058	0.0506

Table 4: SRCNN Super-resolution Results

	Tumor Region		Brain Region	
	mean	std	mean	std
MAE	0.05568	0.01843	0.001319	0.000303
MPE	5.56846	1.8435	0.1319	0.03034
MSE	0.00269	0.00161	446.861	161.617
RMSE	0.04947	0.01561	20.7943	3.8023
PSNR	26.5714	2.8761	70.1199	1.628
SSIM	0.9178	0.04281	0.9925	0.00031
MS-SSIM	0.9922	0.00014	0.99	0.0001
SCC	0.8985	0.0916	0.49891	0.00523

Experimental Results for Breast Histopathology Image Dataset

- For the breast histopathology image dataset, we have generated results for five super-resolution models.
- Trained ESPCN and SRCNN models on 1000 normal breast histopathology images and evaluated the performance of the trained models on 200 diseased breast histopathology images.
- We have evaluated results using pretrained single-image super-resolution models such SRGAN, WDSR, and EDSR. These models have been trained on DIV2K dataset [60] which contains 1000 2K resolution RGB images divided into: 800 images for training, 100 images for validation, 100 images for testing.
- Tested the performance of these pre-trained models on 200 breast histopathology images.

Experimental Results for Breast Histopathology Image Dataset

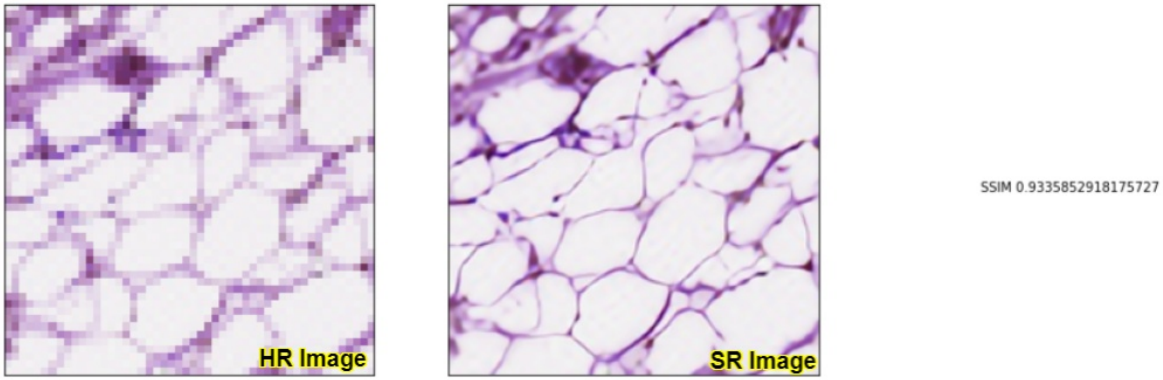


Figure 6: Normal Breast Tissue SR Result Example

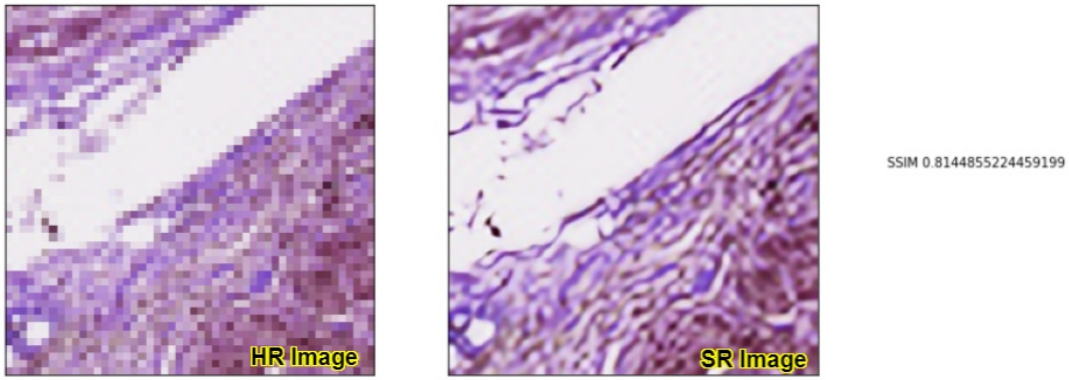


Figure 7: Diseased Breast Tissue SR Result Example

Table 5: EDSR Super-resolution Results

	Presence of Breast Cancer in Images		Absence of Breast Cancer in Images	
	mean	std	mean	std
MAE	0.0019	0.0018	0.0019	0.0018
MPE	0.1931	0.1894	0.1931	0.1894
MSE	5874.278	2857.132	279.9498	80.8171
RMSE	74.2147	19.1427	16.5259	2.6157
PSNR	11.0273	2.3435	23.8952	1.56
SSIM	0.3477	0.021	0.866	0.0789
MS-SSIM	0.2357	0.108	0.979	0.0089
SCC	0.0026	0.0197	0.9152	0.0298
VIFP	0.0026	0.0018	0.3837	0.0478

Table 6: SRGAN Super-resolution Results

	Presence of Breast Cancer in Images		Absence of Breast Cancer in Images	
	mean	std	mean	std
MAE	0.0002	0.0003	0.0002	0.0003
MPE	0.0278	0.0378	0.0278	0.0378
MSE	5184.923	2339.6083	628.8999	216.2162
RMSE	70.1548	16.2241	24.6629	4.5429
PSNR	11.4462	2.0394	20.4565	1.7664
SSIM	0.3532	0.02113	0.7626	0.0616
MS-SSIM	0.2322	0.1067	0.9276	0.0423
SCC	0.0028	0.0202	0.7447	0.0664

Table 7: WDSR Super-resolution Results

	Presence of Breast Cancer in Images		Absence of Breast Cancer in Images	
	mean	std	mean	std
MAE	0.002	0.0018	0.002	0.0018
MPE	0.2028	0.1869	0.2028	0.1869
MSE	5890.321	2837.1937	294.253	89.9002
RMSE	74.3657	18.9752	16.9202	2.8209
PSNR	11.0029	2.3171	23.7045	1.6461
SSIM	0.3437	0.0209	0.861	0.0329
MS-SSIM	0.2328	0.1085	0.9778	0.0104
SCC	0.0023	0.0195	0.9122	0.0332

Table 8: ESPCN Super-resolution Results

	Presence of Breast Cancer in Images		Absence of Breast Cancer in Images	
	mean	std	mean	std
MAE	0.0007	0.0011	0.0007	0.0005
MPE	0.0713	0.1122	0.0712	0.0576
MSE	731.1954	244.2163	699.5786	273.9321
RMSE	26.6593	4.525	25.9618	5.0555
PSNR	19.7422	1.5073	20.0098	1.7073
SSIM	0.6229	0.0827	0.7456	0.0789
MS-SSIM	0.8981	0.0308	0.9378	0.0256
SCC	0.3402	0.0814	0.4022	0.09

Experimental Results for Breast Histopathology Image Dataset

Table 9: SRCNN Super-resolution Results

	Presence of Breast Cancer in Images		Absence of Breast Cancer in Images	
	mean	std	mean	std
MAE	0.0268	0.0463	0.0797	0.0466
MPE	2.6797	4.6308	7.9713	4.6687
MSE	285.0829	179.8302	297.061	182.173
RMSE	51.016	14.9145	51.991	16.55
PSNR	14.3036	2.3161	14.2834	2.9353
SSIM	0.5746	0.0454	0.6	0.0648
MS-SSIM	0.5568	0.0753	0.5769	0.0506
SCC	0.5104	0.0096	0.5164	0.0115

Experimental Results for Diabetic Retinopathy Image Dataset

- For diabetic retinopathy image dataset, we have generated results for five super-resolution models.
- Trained ESPCN and SRCNN models on 1000 normal diabetic retinopathy images and evaluated the performance of the trained models on 200 anomalous diabetic retinopathy images.
- We have evaluated results using pretrained single-image super-resolution models such SRGAN, WDSR, and EDSR.
- Tested the performance of these pre-trained models on 200 diabetic retinopathy images.

Experimental Results for Diabetic Retinopathy Image Dataset

Figure 8: Normal Diabetic Retinopathy SR Result Example

Figure 9: Diseased Diabetic Retinopathy SR Result Example

Table 10: EDSR Super-resolution Results

	Presence of Diabetic Retinopathy		Absence of Diabetic Retinopathy	
	mean	std	mean	std
MAE	0.00075	0.0031	0.00084	0.0024
MPE	0.0752	0.3108	0.0841	0.2447
MSE	7.588	6.1985	9.0764	6.8541
RMSE	2.5381	1.0704	2.8055	1.0977
PSNR	40.8551	3.8829	39.8481	3.464
SSIM	0.9661	0.0075	0.9844	0.0071
MS-SSIM	0.9966	0.00091	0.9975	0.00083
SCC	0.6793	0.1753	0.6935	0.1588

Table 11: SRGAN Super-resolution Results

	Presence of Diabetic Retinopathy		Absence of Diabetic Retinopathy	
	mean	std	mean	std
MAE	0.00256	0.01059	0.00282	0.00839
MPE	0.2563	1.0598	0.2825	0.8396
MSE	54.0392	18.755	58.3966	18.1808
RMSE	7.24	1.2732	7.55193	1.16825
PSNR	31.08	1.626	30.6746	1.3593
SSIM	0.66426	0.0875	0.6982	0.0785
MS-SSIM	0.9634	0.0137	0.9626	0.0128
SCC	0.3596	0.101	0.3729	0.0978

Table 12: WDSR Super-resolution Results

	Presence of Diabetic Retinopathy		Absence of Diabetic Retinopathy	
	mean	std	mean	std
MAE	0.00063	0.0022	0.00075	0.0018
MPE	0.0637	0.2206	0.07563	0.1806
MSE	8.5602	7.105	10.24704	7.7949
RMSE	2.6883	1.1546	2.9738	1.1845
PSNR	40.3826	3.9375	39.37209	3.54467
SSIM	0.95369	0.00806	0.9719	0.0075
MS-SSIM	0.9965	0.000933	0.9974	0.00081
SCC	0.67576	0.1748	0.69028	0.1582

Table 13: ESPCN Super-resolution Results

	Presence of Diabetic Retinopathy		Absence of Diabetic Retinopathy	
	mean	std	mean	std
MAE	0.001687	0.00726	0.001904	0.00616
MPE	0.16877	0.7263	0.19041	0.61601
MSE	6.6742	13.0547	6.49026	4.40397
RMSE	2.39915	0.9583	2.47855	0.5891
PSNR	40.8229	1.9122	40.4236	1.6472
SSIM	0.9499	0.01497	0.9712	0.00813
MS-SSIM	0.99059	0.00291	0.99048	0.00208
SCC	0.37914	0.1047	0.40238	0.10158

Experimental Results for Diabetic Retinopathy Image Dataset

Table 14: SRCNN Super-resolution Results

	Presence of Diabetic Retinopathy		Absence of Diabetic Retinopathy	
	mean	std	mean	std
MAE	0.32041	0.036479	0.31437	0.03965
MPE	32.04103	3.6479	31.4373	3.9655
MSE	123.4825	25.375	102.892	24.287
RMSE	11.1123	1.95301	10.4218	1.9444
PSNR	18.1671	1.1179	18.2562	1.1094
SSIM	0.640806	0.08808	0.6658	0.07843
MS-SSIM	0.86204	0.03857	0.86442	0.03477
SCC	0.01027	0.00627	0.01094	0.005462

Discussion

- On analyzing the metrics for normal images, we could see that the obtained super-resolved images are far more detailed and can retain most of the original image's structure, contrast and intensity pattern across the entire image.
- On analyzing the metrics for diseased images, we could see a loss of texture and structure while pixel intensity pattern is preserved across the entire image.

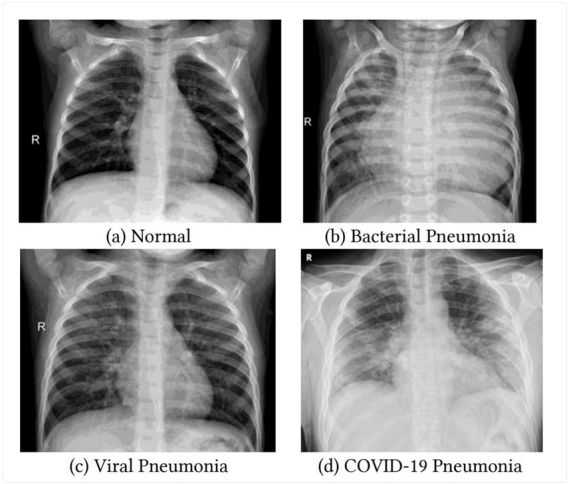


Figure 10: Chest X-Ray of a Patient

- Preserving texture and structure is very important for radiologists while making the decision. For example, when looking into the chest x-ray image of the patient to check whether he as of (c) viral pneumonia, (b) bacterial pneumonia or (d) covid-19 pneumonia, the radiologists need intensity and structure information to make the decision; loss of any of this information can lead to the wrong diagnosis.

Discussion

- Similarity texture information pay key role while analysing histopathology (tissue) or tumor regions in brain. Even a slightest loss of information can lead to the wrong diagnosis which can be fatal for the patient.

Conclusion

- Super-resolved image cannot completely map the original structure of diseased region.
- We can not deploy such algorithms in real-world clinical settings.
- Till the present date, the final decision is made by the radiologists.
- Scope for research in developing better algorithms for dealing with heterogeneous medical pathology images.
- Scope for research in coming up with better deep-learning-based models that perform on par on abnormal images.

[1] H. K. S. N. K. M. L. Bee Lim, Sanghyun Son.
Enhanced deep residual networks for single image super-resolution.
2013.

[2] K. H. X. T. Chao Dong, Chen Change Loy.
Image super-resolution using deep convolutional networks.
2017.

[3] F. H. J. C. A. C. A. A. A. A. A. T. J. T. Z. W. W. S. Christian Ledig, Lucas Theis.

[4] A. Giachetti and N. Asuni.
Fast artifacts- free image interpolation.
2008.

[5] icola Asuni and A. Giachett.
Accuracy improvements and artifact removalin edge based image interpolation.
2008.

[6] K. M. L. J. Kim, J. Kwon Lee.
Deeply-recursive convolutional network for image super-resolution.
2016.

[7] J. Y. N. X. Z. W. X. W. T. H. Jiahui Yu, Yuchen Fan.
Wide activation for efficient and accurate image super-resolution.
2010.

[8] X. Li and M. T. Orchard.
New edge-directed interpolation".
2001.

[9] F. H. J. T. A. P. A. R. B. D. R. Z. W. Wenzhe Shi, Jose Caballero.
Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network.
2015.

[10] N. A. W.S. Lai, J.B. Huang and M. Yang.
Deep laplacian pyramid networks for fast and accurate superresolution.
2017.

[11] I. A. C. B. F. I. H. R. S. S. M. I. Zhou Wang, Member and I. Eero P. Simoncelli, Senior Member.
Image quality assessment: From error visibility to structural similarity.
2004.

Thanks!

Supporting Material: Chest X-Ray Pneumonia Dataset

- The chest x-ray dataset is organized into three directories (Train, Test and Val) and contains subdirectories for each image class (Pneumonia/Normal).
- There are 5,863 X-Ray images of JPEG format and two categories (Pneumonia/Normal)

Supporting Material: Chest X-Ray Pneumonia Dataset

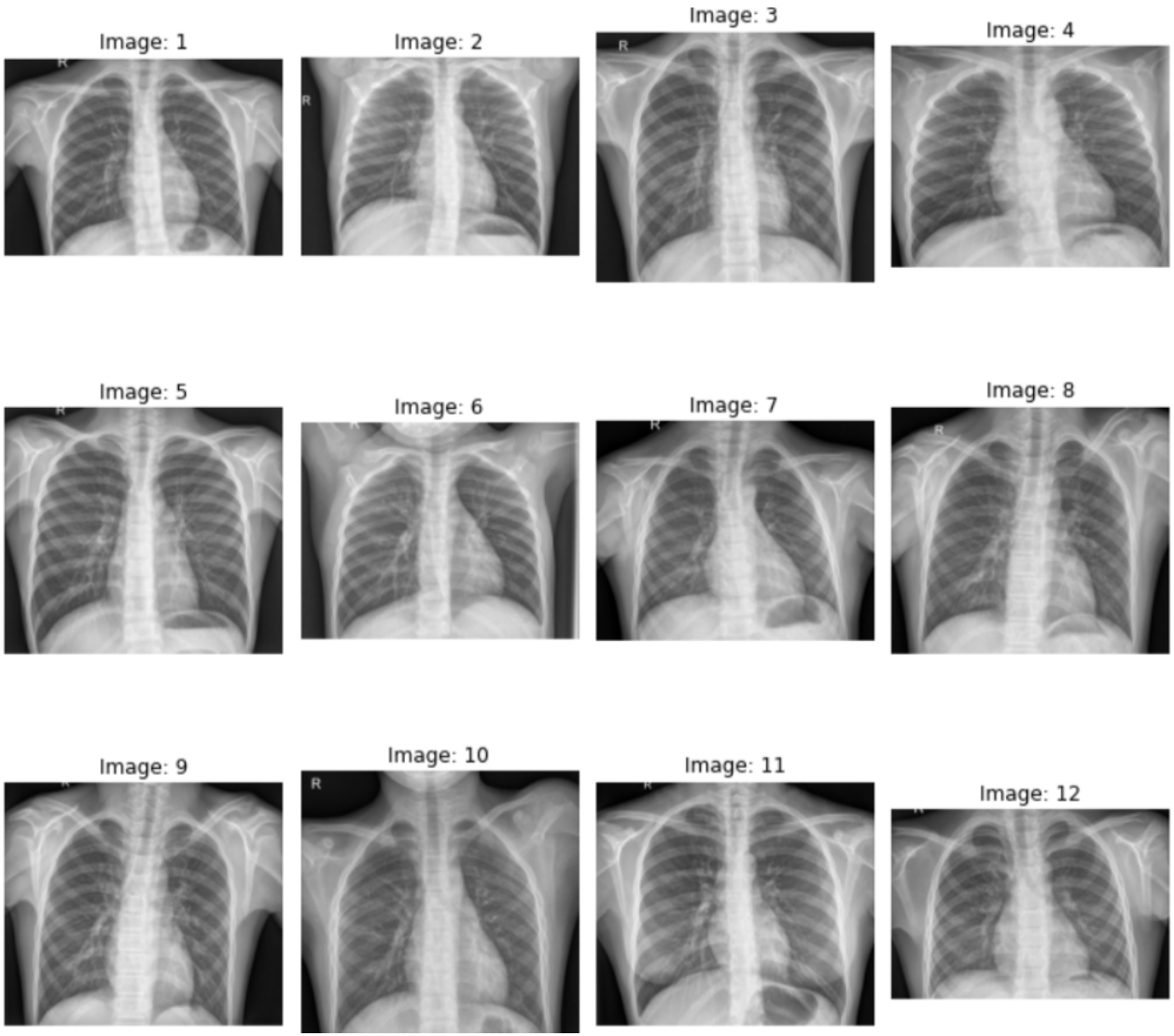


Figure 11: Example Images from Chest X-Ray Pneumonia Dataset

Supporting Material: Brain Tumor Dataset

- This brain tumour dataset containing 3064 T1-weighted contrast-enhanced images from 233 patients with three kinds of brain tumour meningioma (708 slices), glioma (1426 slices), and pituitary tumour (930 pieces).
- The whole dataset is split into four subsets, each containing 766 slices.
- This dataset is organized in the Matlab data format (.mat file). Each file stores a struct containing the following fields for an image:
  - cjdata.label: 1 for meningioma, 2 for glioma, 3 for pituitary tumor
  - cjdata.PID: patient ID
  - cjdata.image: image data
  - cjdata.tumorBorder: a vector storing the coordinates of discrete points on the tumour border. For example, [x1, y1, x2, y2,...] in which x1 and y1 are planar coordinates on the tumour border. It was generated by manually delineating the tumour border. So we can use it to generate the binary image

Supporting Material: Brain Tumor Dataset

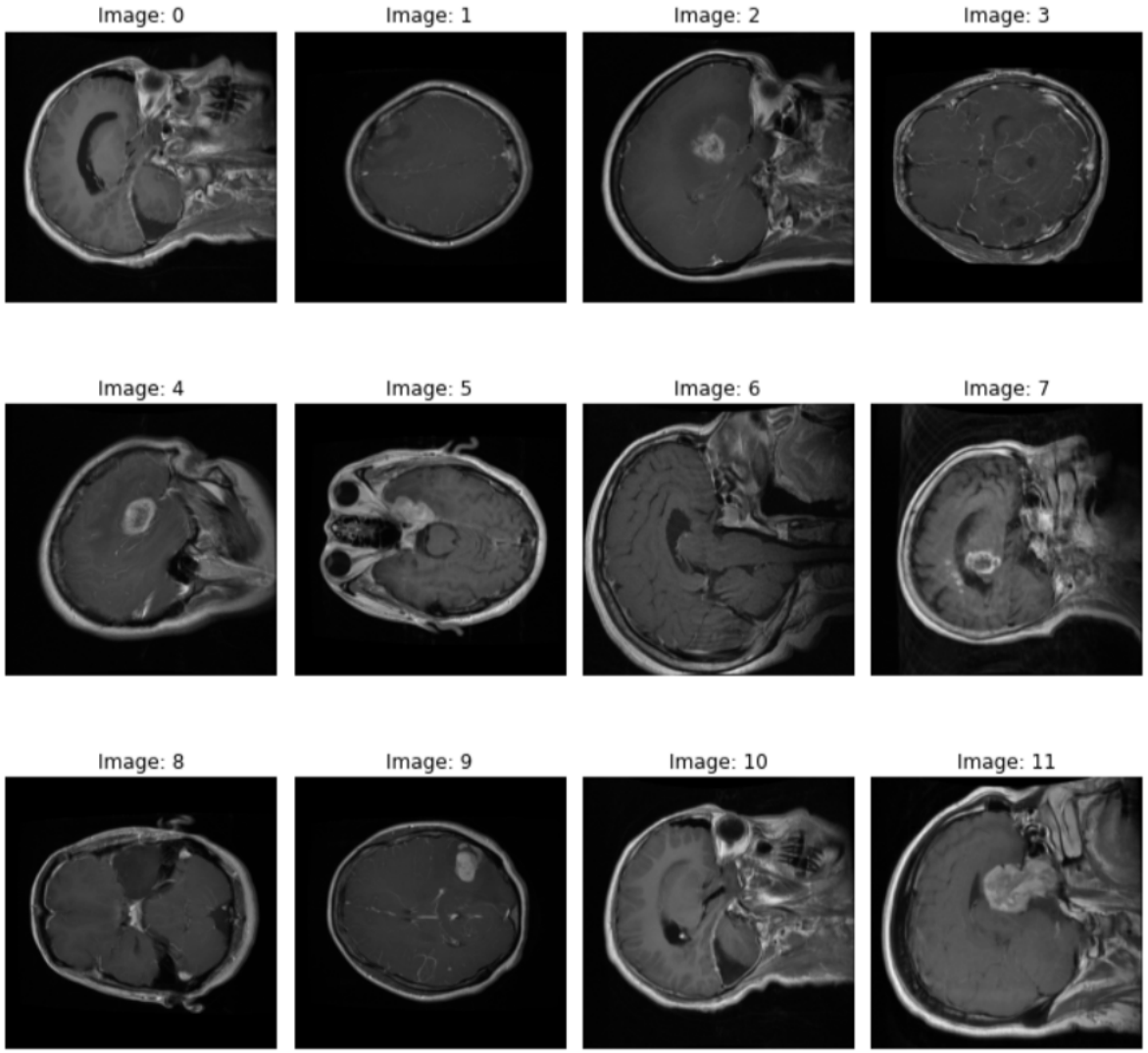


Figure 12: Example Images from Brain Tumor Dataset

Supporting Material: Breast Histopathology Image Dataset

- The dataset consists of 162 whole mount slide images of Breast Cancer (BCa) samples scanned at 40x. From these samples, 277,524 patches of size 50 x 50 were extracted consisting of 198,738 negative and 78,786 positive patches.

Supporting Material: Breast Histopathology Image Dataset

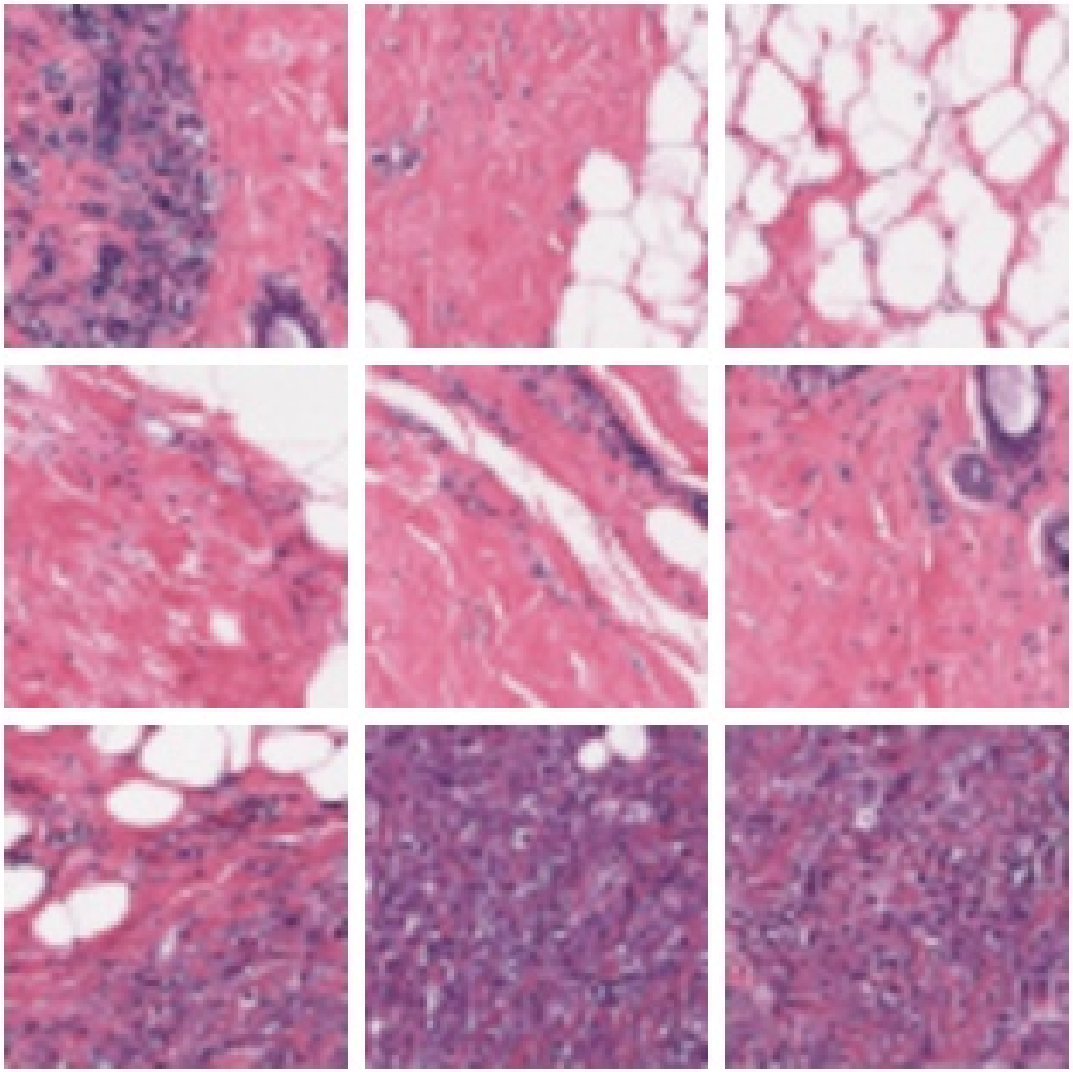


Figure 13: Example Images from Breast Histopathology Image Dataset

Supporting Material: Diabetic Retinopathy Image Dataset

- This dataset was taken from the kaggle competition, and EyePACS provided the retinal images.
- The dataset consists of a large set of high-resolution retina images taken under a variety of imaging conditions. Left and right field are given for every patient.Images are labelled with patient id and are either left or right eye.
- The clinician has evaluated the presence of diabetic retinopathy in each image on a scale of 0 to 4.
  - 0 - No DR
  - 1 - Mild
  - 2 - Moderate
  - 3 - Severe
  - 4 - Proliferative DR

Supporting Material: Diabetic Retinopathy Image Dataset

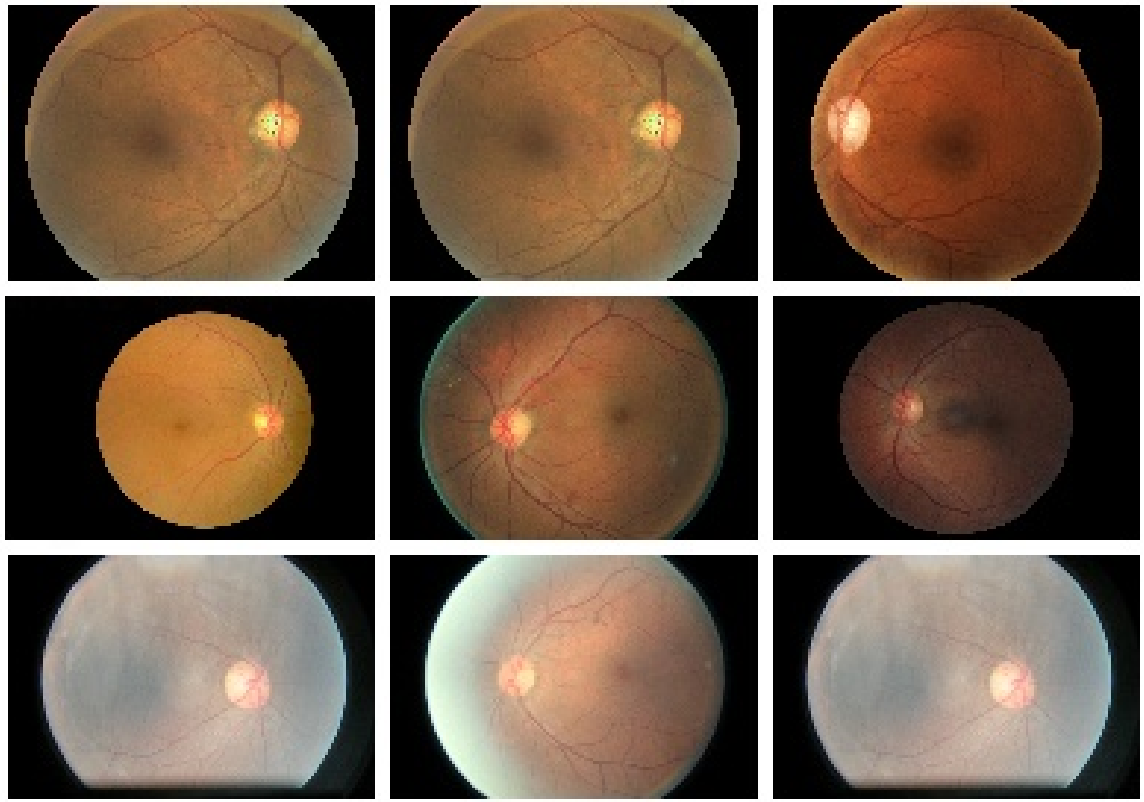


Figure 14: Example Images from Diabetic Retinopathy Image Dataset

Supporting Material: Mean Square Error (MSE)

- The MSE is defined as the mean or average of the squared differences between the target value and expected predicted values in a dataset.

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2$$

- Where Y_i is the i-th target value in the dataset and $\hat{Y}_i$ is the i-th predicted value. The difference between target value and predicted value is squared, which removes the sign, and results in a positive error value. Squaring also results in significant magnifying errors. The bigger the difference between the predicted and target values, the larger the resulting squared positive error. This leads to punishing models more for larger errors when MSE is used as a loss function.

Supporting Material: Root Mean Square Error (RMSE)

- The root mean squared error can be considered as an extension of the mean squared error. Here we need to make a note that the square root of the error is calculated, which means the units of the RMSE are the same as the original units of the target value that is being predicted. Therefore we can use MSE loss to train the model and RMSE to evaluate and report its performance.

$$\mathrm{RMSD} = \sqrt{\frac{\sum_{i=1}^{N} \left(Y_i - \hat{Y}_i\right)^2}{N}}$$

- Where Y_i is the i-th sample target value , $\hat{Y}_i$ is the i-th sample predicted value.

Supporting Material: Mean Absolute Error (MAE)

- The MAE is defined as the mean or average of the absolute differences between the target value and expected predicted values in a dataset. The MAE can be calculated as follows:

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n} \left| Y_i - \hat{Y}_i \right|}{n}$$

- Where Y_i is the i-th sample expected value, $\hat{Y}_i$ is the i-th sample predicted value.

Supporting Material:Mean Absolute Percentage Error (MAPE)

- Mean absolute percentage error (MAPE) is frequently used as a loss function for regression problems and in model evaluation because of its very intuitive interpretation in terms of relative error. It usually expresses the accuracy as a ratio defined by the formula:

$$M = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right|$$

- where Y_i is the target value and $\hat{Y}_i$ is the predicted value. The MAPE can also be reported as a percentage. The difference between $\hat{Y}_i$ and Y_t is divided by the target value Y_i. Now, the absolute value is summed for every sample and divided by the total number of samples n.

Supporting Material:Peak Signal to Noise Ratio (PSNR)

- The peak signal-to-noise ratio (PSNR) can be defined as the ratio between the maximum possible power of a photo and the power of corrupting noise which affects the quality of its representation. To estimate the PSNR of a picture, it is necessary to compare that picture to a perfect clean picture with the maximum possible power. PSNR is defined as follows:

$$PSNR = 10 \log_{10} \left(\frac{(I-1)^2}{MSE} \right) = 20 \log_{10} \left(\frac{I-1}{RMSE} \right)$$

- Here, I is the number of maximum possible intensity levels (minimum intensity level supposed to be 0) in an image.

Supporting Material:Peak Signal to Noise Ratio (PSNR)

- Here, MSE is the mean squared error which is defined as:

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} (Original(i,j) - Degraded(i,j))^2$$

- Where Original(i,j) represents the matrix data of the original image. Degraded(i,j) means the matrix data of the degraded photo. m represents the numbers of rows of pixels, and i represents the index of that row of the picture. n represents the number of columns of pixels and j represents the index of that column of the photo. Here, RMSE is the root mean squared error.

Supporting Material:Multi-Scale Structure Similarity Index (MS-SSIM)

- MS-SSIM is extended, and a better version form of SSIM, called Multiscale SSIM (MS-SSIM) [33], is executed over multiple scales through a process of various stages of sub-sampling reminiscent of multiscale processing in the early visual system. It has been shown that MS-SSIM performs equally well or better than SSIM on different subjective image and video databases.

$$\mathrm{SSIM}(o, s) = [L_m(o, s)]^{\alpha M} \cdot \prod_{j=1}^{M} [C_j(o, s)]^{\beta_j} [S_j(o, s)]^{\gamma_j}$$

- Here, the exponents αM, β_j and γ_j are used to adjust the relative importance of different components

Supporting Material:Spatial Correlation Coefficient (SCC)

- Spatial correlation refers to the degree to which one object is similar to other nearby objects. The term Spatial means space, and correlation means association. In naive terms, it measures how close objects are similar to other comparable objects. SCC is used to measure the overall spatial correlation in the data set. SCC can be classified as positive, negative and no spatial correlation.

Supporting Material:Pixel Based Visual Information Fidelity (vif-p)

- Visual Information Fidelity (VIF) is a complete reference image quality assessment index that relies on natural scene statistics and the concept of photo information extracted by the human optical system (HVS). The VIF index employs wild scene statistical (NSS) models in occurrence with a distortion (channel) model to evaluate the information shared between the test and the reference images.

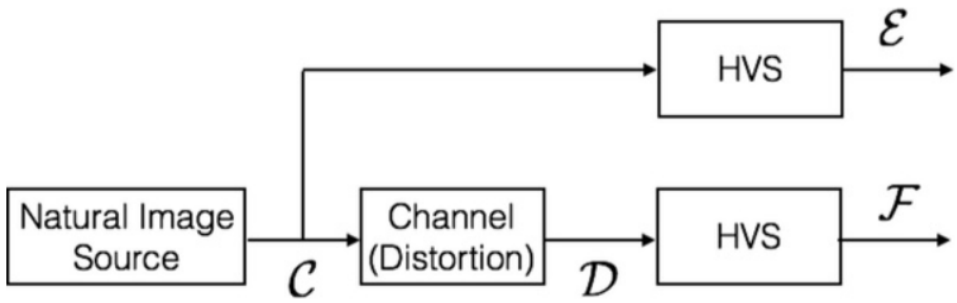


Figure 15: shows two information measures are combined to form a visual information fidelity measure that relates visual quality to relative image information

Supporting Material: Enhanced Deep Residual Networks for Single Image Super-Resolution (EDSR)

- EDSR is a super-resolution model proposed after SRResNet. SRResNet successfully solved processing time and memory consumption problems, but ResNet used in SRResNet is a model architecture for image classification, which is not optimal for super-resolution. Therefore, EDSR builds a more optimal model for super-resolution by removing unnecessary modules from ResNet. EDSR introduces a constant scaling layer of 0.1 at the output of the last convolution layer for the residual block to make the training more stable.

Supporting Material: Enhanced Deep Residual Networks for Single Image Super-Resolution (EDSR)

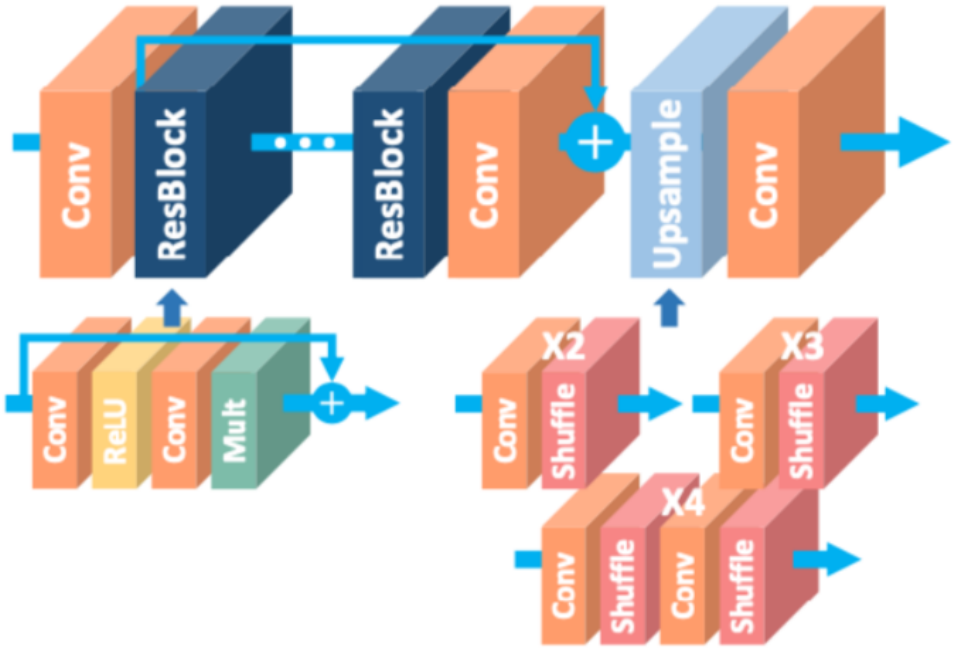


Figure 16: EDSR Architecture

- In EDSR, the authors have proposed different architecture of ResBlock, which more efficient to train the model.

Supporting Material: Enhanced Deep Residual Networks for Single Image Super-Resolution (EDSR)

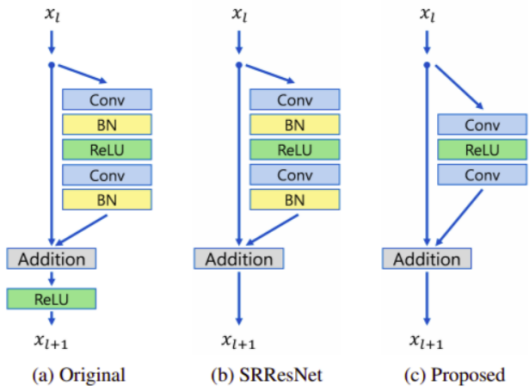


Figure 17: Residual Block Comparison with ResNet and SRResNet

- In this proposed network, the authors have removed the batch normalization layers; since batch normalization layers normalize the features, they eliminate range flexibility from networks by standardising the features. Therefore it is better to remove them. Moreover, GPU memory usage is sufficiently reduced since the batch normalization layers consume the same amount of memory as the preceding convolutional layers.

Supporting Material: Wide Activation for Efficient and Accurate Image Super-Resolution (WDSR)

- WDSR model uses the concept of expanding features before ReLU activation, which allows more information to pass through; at the same time, it capable of keeping the highly non-linearity of deep neural networks. Thus allowing low-level SR features from shallow layers to easier propagate to the final layer for better dense pixel value predictions.The general architecture is represented in fig 14

Figure 18: WDSR Network Architecture Improvement over EDSR Network

Supporting Material: Wide Activation for Efficient and Accurate Image Super-Resolution (WDSR)

- The resulting SR residual network has a narrow identity mapping pathway with wider (2x to 4x) channels before activating each residual block.
- To further widen activation (6x to 9x) without computational overhead, we introduce linear low-rank convolution into SR networks and achieve even better accuracy-efficiency trade-offs.
- The Residual block architecture is attached below in fig 15

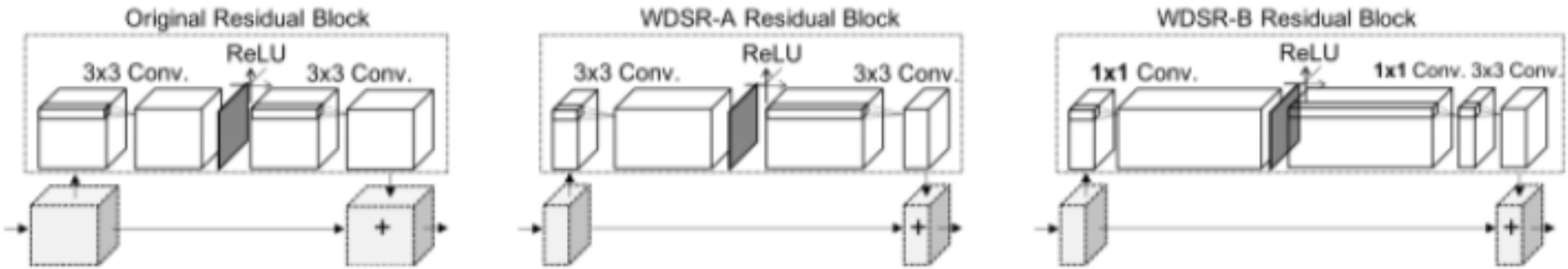


Figure 19: Residual Block Architecture

Supporting Material: Wide Activation for Efficient and Accurate Image Super-Resolution (WDSR)

- Starting from left there is vanilla residual block.In middle we have WDSR-A: residual block with broad activation and at right we have WDSR-B: residual block with broader activation and linear low-rank convolution. We demonstrate different residual building blocks for image super-resolution networks.
- Compared with vanilla residual blocks used in EDSR, authors have introduce WDSR-A, which has a narrow identity mapping pathway with wider (2x to 4x) channels before activation in each residual block. Further present WDSR-B with linear low-rank convolution stack and even widen activation (6x to 9x) without the computational overhead.
- In WDSR-A and WDSR-B, all ReLU activation layers are only applied between two comprehensive features (features with larger channel numbers).

Supporting Material: Single Image Super-Resolution Using a Generative Adversarial Network (SRGAN)

- Super-resolution GAN applies a deep network combined with an adversary network to produce higher resolution images. During the training, A high-resolution image (HR) is downsampled to a low-resolution image (LR). A GAN generator upsamples LR images to super-resolution images (SR). We use a discriminator to distinguish the HR images and backpropagate the GAN loss to train the discriminator and the generator.

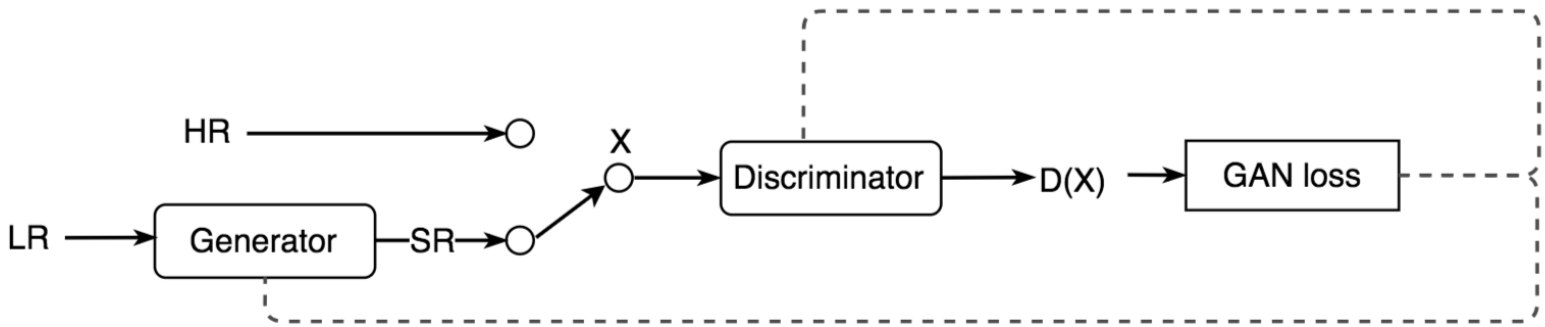


Figure 20: SRGAN Back-Propagation Model

Supporting Material: Single Image Super-Resolution Using a Generative Adversarial Network (SRGAN)

- Below we have presented the network design for the generator and the discriminator. It is mainly composed of convolution layers, batch normalization and parameterized ReLU (PRelU). The generator also implements skip connections similar to ResNet. The convolution layer with k3n64s1 stands for 3x3 kernel filters outputting 64 channels with stride 1.

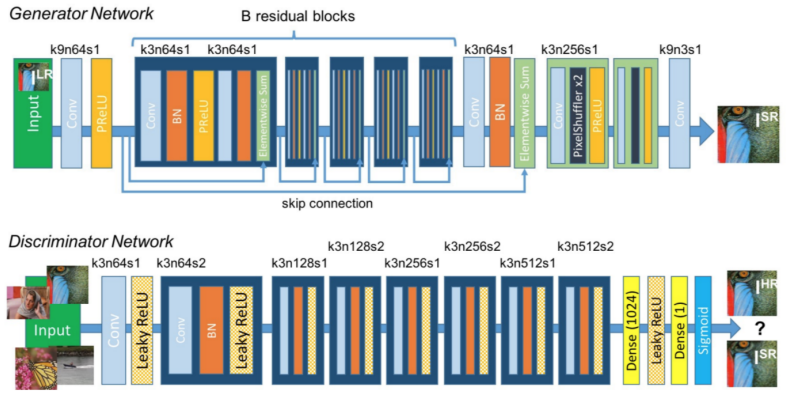
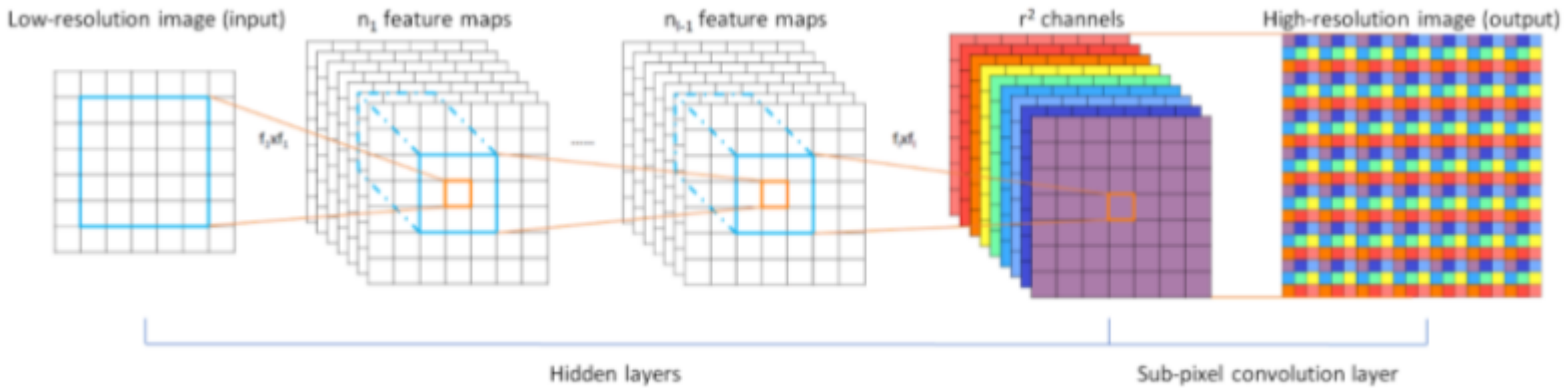


Figure 21: Generator and Discriminator Architecture

Supporting Material: Efficient Sub-Pixel Convolutional Neural Network (ESPCN)

- ESPCN is a CNN based deep learning architecture where the feature maps are extracted in the LR space. The authors have introduced an efficient sub-pixel convolution layer that learns an array of upscaling filters to upscale the final LR feature maps into the HR output. In this process, we effectively replace the handcrafted bicubic filter in the SR pipeline with more complex upscaling filters specifically trained for each feature map whilst also reducing the computational complexity of the overall SR operation.



Supporting Material: Efficient Sub-Pixel Convolutional Neural Network (ESPCN)

- In ESPCN, the low-resolution (LR) image is upscaled at the very last stage as the high-resolution (HR) image. Therefore, the number of computations are reduced for the network because small-size feature maps are used. Consequently, real-time performance can be achieved. ESPCN uses the concept of pixel-shuffle; image reshaping using phase shift is also called pixel shuffle, which rearranges the elements of H x W x C . r^2 tensor to form rH x rW x C tensor as shown below.

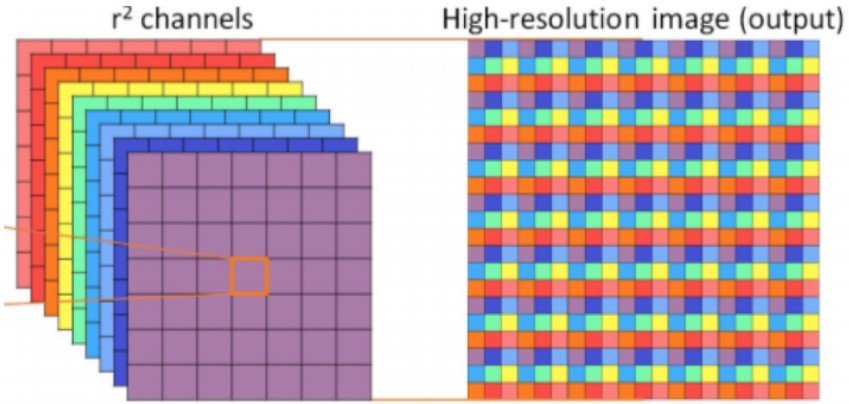


Figure 23: ESPCN Architecture

Supporting Material:Super-Resolution Using Deep Convolutional Networks (SRCNN)

- The SRCNN is a simple network consisting of two hidden convolutional layers, as shown in Fig. 20. The input image is an upscaled version of the LR image, so the architecture corresponds to the pre-upsampling SR.

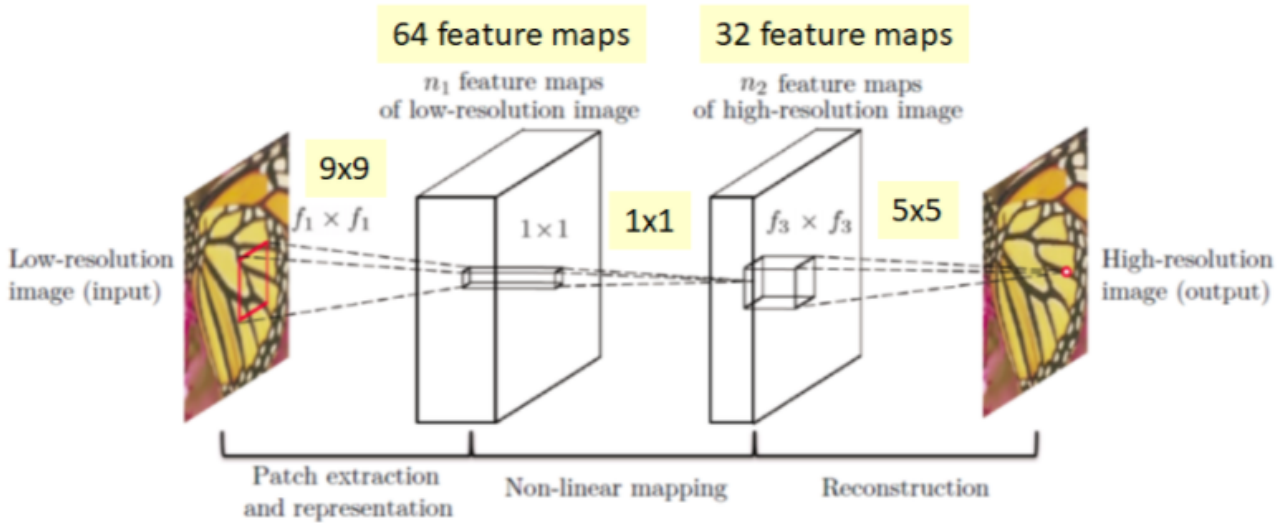


Figure 24: SRCNN Architecture

Supporting Material: Efficient Sub-Pixel Convolutional Neural Network (ESPCN)

- There are three parts, patch extraction and representation, non-linear mapping, and reconstruction.
  - Patch Extraction and Representation : First of all, we upscale low-resolution input to the desired size using bicubic interpolation before we input it to the SRCNN network.
  - Non-Linear Mapping : After patch extraction, now, a non-linear mapping is performed.
  - Reconstruction : After we do mapping, we need to reconstruct the image. Hence, we do convolution again.