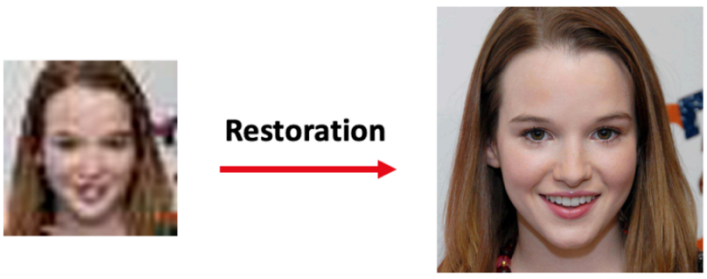
AI6126 ADVANCED COMPUTER VISION

PROJECT 2: BLIND FACE SUPER RESOLUTION



NANYANG TECHNOLOGICAL UNIVERSITY
SINGAPORE

APRIL 26, 2024

CHITHRA RAMESH ASSWIN
G2302832A

PROJECT 2 : BLIND FACE SUPER-RESOLUTION

AI6126 : ADVANCED COMPUTER VISION

1. PREPROCESSING OF DATA

The FFHQ dataset provided for the project includes 5,000 high-quality (HQ) face images for training and 400 HQ-LQ image pairs for validation, with LQ images pre-processed to 128 x 128 pixels through a **second-order degradation pipeline** that involves Gaussian blur, downsampling, noise, and compression. In the absence of training LQ images, these are generated by applying the same degradation pipeline to HQ images, crucial for training the super-resolution model to enhance image quality from LQ to HQ. Two test sets are issued closer to the project deadline: 400 synthetic LQ images and 6 real-world LQ images, with the latter upscaled using bicubic interpolation to create fake GT images for local PSNR testing, although actual PSNR evaluation occurs post-submission on CodaLab using true HQ images as reference.

2. MODELS USED & TECHNIQUES:

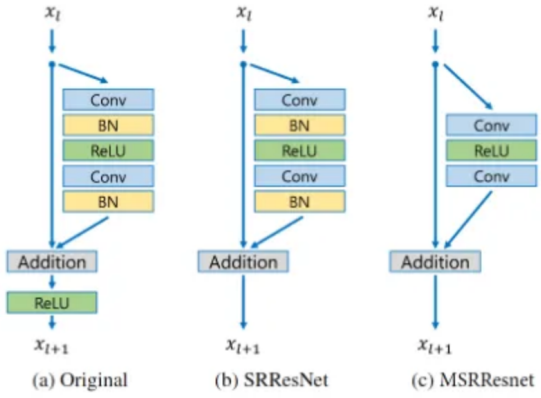


Fig.1 comparison of SRResNet vs MSRResNet

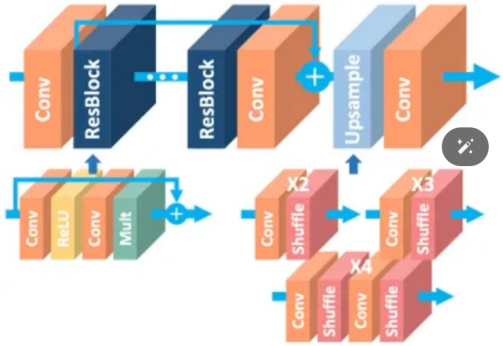


Fig.2 EDSR Architecture

2.1. Model Architecture 1: Modified SRResNet.

The chosen architecture is a refined variant of SRResNet, strategically modified by removing Batch Normalization (BN) layers as observed from Figure 1 to enhance the super-resolution process. This modification aims to maintain the integrity of high-frequency details, which are essential for reconstructing high-resolution images with clarity and sharpness. The network is constructed to process RGB images using 3 input and 3 output channels. Within the network, 64 feature maps are used for each convolutional layer to capture diverse and intricate features. The model deepens its capacity to capture details with an increased count of 24 residual blocks, compared to the standard 16, allowing for a substantial learning of the residuals between low and high-resolution images without information loss. To achieve the desired upscaling, the network implements a factor of 4, perfectly suited for the task of enhancing images from a resolution of 128x128 to 512x512 pixels.

2.2. Model Architecture 2: Enhanced Deep Super-Resolution Network EDSR

Building upon the foundation of deep learning-based super-resolution, the Enhanced Deep Super-Resolution Network (EDSR) for this project excludes Batch Normalization (BN) from its architecture. This decision is instrumental in avoiding BN-induced artifacts, ensuring the produced high-resolution images are free from unnecessary blurring and preserve the fine details. The architecture utilizes 3 input and output channels compatible with standard RGB images and incorporates 64 feature maps across the convolutional layers to robustly extract and process image features. The EDSR structure is fortified with 24 powerful residual blocks, augmenting the typical 16-block setup, to enhance the network's ability to learn intricate image details critical for super-resolution. This augmented design facilitates an upscaling by a factor of 4, transforming 128x128 pixel images into crisp, clear 512x512 pixel outputs, aligning with the project's objectives for high-resolution image generation.

a. Data Augmentation:

In the preprocessing pipeline of the FFHQsubDataset, various image augmentations are applied to the ground truth (GT) images to enhance the robustness of the super-resolution model. These include a 50% chance of horizontal flip, rotation by 90, 180, or 270 degrees, 50% probability of brightness adjustment (0.7 to 1.3 factor), and 50% chance of contrast adjustment (same factor range). These augmentation strategies are crucial for preparing the model to handle a wide range of image qualities and degradation types encountered in real-world applications.

Augmentation	Probability	Description
Horizontal Flip	0.5 (if `use_hflip` is True)	Flip the image horizontally with a 50% chance.
Rotation	Dependent on `use_rot` option	Rotate the image by 90, 180, or 270 degrees if enabled.
Brightness Adjustment	0.5	Randomly adjust brightness with a factor between 0.7 and 1.3.
Contrast Adjustment	0.5	Randomly adjust contrast with a factor between 0.7 and 1.3.

b. Optimizers and Scheduler :

The Adam optimizer was configured with a learning rate of $2 \times 10^{-4}$and beta coefficients set to 0.9 and 0.99. A Cosine Annealing scheduler was used to adjust the learning rate, featuring a strategy with two periods of 150,000 iterations each.

LOSS FUNCTION:

For the training of the super-resolution model, two metrics are integral to the optimization process: the L1 loss and the Peak Signal-to-Noise Ratio (PSNR). The L1 loss function is employed as the primary loss during training, guiding the model to minimize the average absolute differences between the predicted high-resolution images and the ground truth:

$$L_1(\hat{y}, y) = \frac{1}{N}\sum_{i=1}^{N}|\hat{y}_i - y_i|$$

Where $(\hat{y}_i)$ represents the pixel values of the predicted image,(y_i)denotes the pixel values of the HQ ground truth image, and (N) is the total number of pixels in each image.

Simultaneously, the model's performance is also evaluated using the PSNR, which is a common metric in image processing that measures the ratio between the maximum possible power of a signal (in this case, pixel values) and the power of corrupting noise that affects the fidelity of its representation. The PSNR is calculated based on the mean squared error (MSE) between the predicted and ground truth images, as given by the following equation:

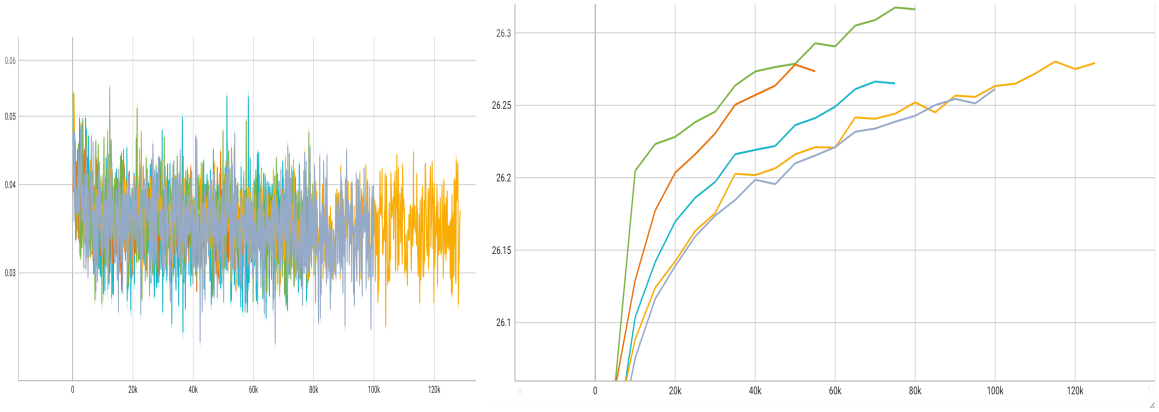
$$\text{PSNR} = 20 \cdot \log_{10}\left(\frac{255}{\sqrt{\text{MSE}}}\right)$$

The MSE is the average of the squares of the differences between the predicted and actual pixel values. Model checkpoints are saved based on the highest PSNR value achieved, indicating the most accurate image restoration at that point in training. By balancing the direct pixel-wise accuracy (via L1 loss) with the PSNR for quality assessment, the model is fine-tuned to produce high-quality outputs that closely match the original high-resolution images.

3. Experiments, Training Curves & Results Discussion:

Architecture	Augmentation	Parameters	Loss	Iterations Trained	Best Val PSNR
MSRResNet-B16	No	1,517,571	L1	125,000	26.51530
MSRResNet-B16	Yes	1,517,571	L1	100,000	26.44951
EDSR B-16	No	1,517,571	L1	80,000	26.420882
MSRResNet-B24	Yes	2,108,419	L1	55,000	26.508414
MSRResNet-B20	**Yes**	**1,812,995**	**L1**	**75,000**	**26.561710**

Test PSNR for the best Model- MSRResNet-B20 : 26.63871



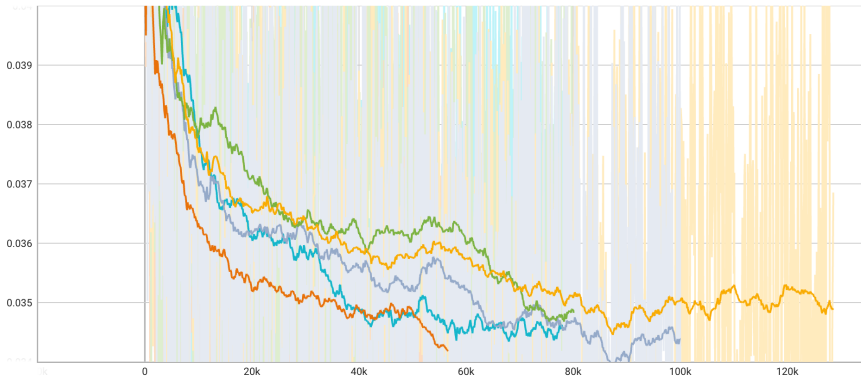


Fig.4 a) PSNR training curves b) l_pix c) smoothed l_pix training curve

In the conducted experiments, several variations of the Super-Resolution model based on a modified SRResNet architecture and EDSR framework were trained and evaluated. Lets see anlyse each of them below

3.1. Experiment 1 :Analysis of Residual Blocks

The experiments conducted show a distinct relationship between the number of residual blocks in the network and the model's performance. Models designated as B16, B20, and B24 correspond to the number of residual blocks within the architecture. The validation PSNR reveals a nuanced trend where MSRResNet-B20, with 20 residual blocks, outperforms both the 16-block and 24-block variants, achieving the highest PSNR. This indicates that an intermediate level of network depth may offer the most advantageous balance between model complexity and performance, suggesting that beyond a certain point, additional residual blocks do not yield proportional gains in image quality and could potentially lead to overfitting.

3.2. Experiment 2 & 3: Impact of Architecture and Augmentation

Upon examining the models' architectural foundation, it is evident that augmentation plays a pivotal role in enhancing the super-resolution performance. Models that utilized augmentation strategies, such as random horizontal flips and brightness/contrast adjustments, consistently achieved higher PSNR values. This improvement underscores the efficacy of augmentations in creating a robust model capable of handling a variety of real-world imaging conditions. Furthermore, the consistent use of the L1 loss across all models firmly establishes its utility in guiding the super-resolution process, striking an optimal balance between precision and practicality in reconstruction.

The comparison between the modified SRResNet architectures and the EDSR framework, while employing a similar number of residual blocks, shows that the specific configurations and inherent characteristics of the network design can have substantial effects on the super-resolution task. As evidenced by the training curves and resulting PSNR values, the intricate relationship between architecture design, augmentation, and the number of residual blocks collectively influences the model's performance. This holistic understanding emphasizes the need for careful consideration of each aspect during model development to optimize super-resolution outcomes.

4. CODALAB SUBMISSION:

#	SCORE	FILENAME	SUBMISSION DATE	SIZE (BYTES)	STATUS	✔	
1	26.4537268343	Archive.zip	04/24/2024 13:34:34	29259	Finished		+
2	26.5132676076	Archive_30k.zip	04/25/2024 12:08:07	42322	Finished		+
3	26.5084141772	Archive_30k_bs16.zip	04/25/2024 12:44:18	43513	Finished		+
4	26.5084141772	Archive_30k_bs16.zip	04/25/2024 12:49:44	43513	Finished		+
5	26.3893818678	Archive_125k.zip	04/25/2024 15:37:51	47301	Finished		+
6	26.6387103565	Archive_2035k.zip	04/25/2024 20:26:28	50914	Finished	✔	+

5. GPU SPECIFICAITIONS :

Table Hardware Specs: SCSE GPU Cluster to train the Models

Cluster	GPU	CPU	Memory	MaxWall
q_amsai	7	1	12G	8Hrs
q_dmsai	10	1	30G	8Hrs

6. Output Visualization :

Fig Low-Quality(LQ) vs. Enhanced Images by MSRResnet with 20 Blocks & Augmentations

Sample Images taken from the low quality test samples provided and compared witht Test images got from the best model which is MSR Resnet , 20 block with augmentations.

7. References：

[1] Wang, Xintao, et al. "Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data." Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. 2021.

[2] Wang, Xintao, et al. "GFP-GAN: Towards Real-World Blind Face Restoration with Generative Facial Prior." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.

[3] https://github.com/xinntao/Real-ESRGAN/tree/master

[4] https://bozliu.medium.com/single-image-super-resolution-challenge-6f4835e5a156

[5] https://github.com/XPixelGroup/BasicSR