Comparative Analysis of Image Super-Resolution Models

Group:52 - Gritta Joshy (gusjoshgr@student.gu.se)
Qi Chen (gusqichr@student.gu.se)

***Abstract*—Image super-resolution (SR) plays an essential role in enhancing low-resolution images, with applications in fields such as medical imaging, satellite imagery, and entertainment. In this report, we analyze three deep learning models for SR: Enhanced Deep Super-Resolution (EDSR), Very Deep Super-Resolution (VDSR), and Deep Recursive Convolutional Network (DRCN). Each model uses a distinct architecture to achieve high-quality image upscaling. We trained the models on the DIV2K dataset and evaluated their performance on the Set5 benchmark using PSNR, SSIM, and Perceptual Index (PI) metrics. Results show that DRCN achieved the highest PSNR, while VDSR demonstrated superior SSIM. EDSR provided a balanced performance with competitive results in both metrics. Our findings highlight key trade-offs between complexity, computational cost, and SR performance, offering insights into the best use cases for each model.**

I. Introduction

Image super-resolution (SR) addresses the challenge of reconstructing high-resolution (HR) images from low-resolution (LR) inputs. It is critical for applications requiring detailed imagery, such as medical imaging for accurate diagnosis, satellite imagery for land mapping, and digital media for enhancing image quality. Challenges include maintaining image details, minimizing image distortions, and improving computational efficiency.

In this project we are experimenting with three deep learning models: EDSR, VDSR, and DRCN to upscale low resolution images to high resolution images. DRCN employs a recursive architecture that allows it to learn from multiple iterations of the same convolutional layers, effectively enhancing low-resolution images while maintaining a compact model size. VDSR utilizes a deep network structure to learn residuals between low and high resolution images, achieving accurate upscaling with a focus on preserving structural details. EDSR optimizes residual learning by removing unnecessary modules, resulting in superior performance for high-resolution image reconstruction, making all three models highly relevant for improving image quality in various applications.

The report aims to compare the performance of these models based on several quantitative and qualitative metrics: PSNR,SSIM,PI. Through this comparison, we aim to identify the model that best balances image quality and computational efficiency for super-resolution tasks.

II. Related Work

Image super-resolution (SR) is a technique to reconstruct high-resolution images from low-resolution inputs. Traditional methods, such as bilinear and bicubic interpolation [1], are computationally efficient but often fail to restore high-frequency details, resulting in smoother images and lack sharpness. These limitations make traditional approaches less effective for applications requiring detailed resolution such as medical imaging. Recent advancements in deep learning have significantly enhanced SR capabilities, enabling models to learn complex mappings from low-resolution to high-resolution images. Unlike interpolation methods, deep learning approaches can capture high-frequency information, generating images with finer textures and sharper edges. The introduction of residual learning and recursive structures has further improved model performance and efficiency, leading to advanced SR techniques.

Among deep learning-based SR models, the Enhanced Deep Super-Resolution (EDSR) model, proposed by Lim et al. (2017) [2] has gained attention for its effectiveness in image reconstruction. EDSR employs a deep residual network architecture with 32 convolutional layers avoiding batch normalization, which the authors argue reduces computational complexity without compromising performance. The model leverages both local and global residual learning techniques to significantly improve image reconstruction, particularly at high-scale factors, where retaining fine details is crucial. However, EDSR's depth and reliance on multiple residual blocks make it computationally expensive and memory-intensive compared to lighter SR models.

Another prominent model is the Very Deep Super-Resolution (VDSR) model, proposed by Kim et al. (2016), [3]. This is an evolving deep SR model that applies a residual learning framework. This paper implements 20 convolutional layers with a single global skip connection, enabling the model to focus on learning the residuals—the differences between low- and high-resolution images. This approach speeds up convergence during training, making VDSR relatively efficient despite its depth. The model performs well at moderate scaling factors but may struggle with high-scale SR tasks compared to models with more smooth residual or recursive mechanisms. However, VDSR's lightweight architecture makes it a strong candidate for real-time applications requiring fast processing with balanced image quality.

The Deep Recursive Convolutional Network (DRCN), also proposed by Kim et al. (2016), [4] introduces a novel

recursive architecture that leverages weight sharing across layers. This structure applies the same set of filters multiple times, allowing DRCN to achieve a high effective depth without increasing model parameters. DRCN loss function is designed to use both intermediate and final outputs, allowing for a more stable and effective training process. By balancing the contributions of the recursive outputs and the final prediction, the model can achieve high-quality super-resolution results while maintaining computational efficiency. However, the recursive applications may result in longer training times, making DRCN less suited for tasks with tight computational constraints.

Previous studies have compared EDSR, VDSR, and DRCN regrading of their architectural strengths, weaknesses, and applicability to different SR tasks. While EDSR often achieves the highest performance in terms of image quality, it requires more computational resources, making it less feasible for real-time applications. VDSR, although less computationally demanding, provides competitive SR performance and is widely applied in scenarios requiring moderate-scale SR. DRCN, with its recursive weight-sharing structure, offers a unique balance between model complexity and efficiency, making it suitable for tasks where flexible scaling is essential. Our study builds on this body of research by directly comparing these models under identical training and evaluation conditions. By focusing on their performance on the DIV2K and Set5 datasets, we aim to provide insights into the trade-offs between model complexity, computational cost, and SR performance, offering guidance on their practical use cases.

III. Methodology

A. Dataset

The DIV2K dataset is a high-quality image dataset widely used for training and evaluating image super-resolution models. It consists of 900 images in 2K resolution, with 800 images designated for training, 100 for validation. The dataset provides both high resolution versions of each image, enabling models to learn and be evaluated on a range of image enhancement tasks.

For model evaluation, we use the Set5 as test dataset, which is a popular benchmark for evaluating image super-resolution models, containing just five high-quality images. It is commonly used to test model performance and benchmark results in super-resolution research.

B. EDSR

1) Model description: EDSR is built on a deep residual network architecture with 32 convolutional layers [2], allowing for complex feature extraction. The architecture begins with a MeanShift layer, which normalizes the input images by adjusting the RGB values based on predefined mean and standard deviation parameters. This preprocessing step is crucial for improving the model's performance. The model consists of a head module that includes a convolutional layer

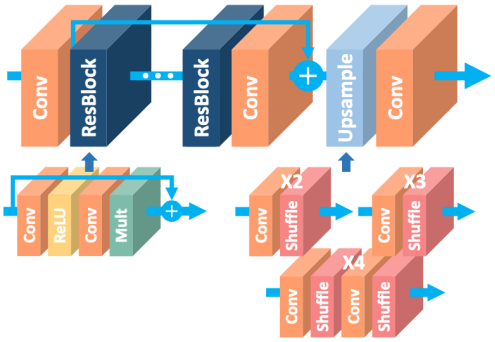


Fig. 1. EDSR Architecture [2]

to extract initial features from the input image. Following this, the body module is composed of multiple Residual Blocks each containing two convolutional layers with ReLU activations. These residual connections allow the model to learn the residuals between low-resolution and high-resolution images, facilitating faster convergence during training. The architecture also incorporates a tail module, which includes an Upsampler that utilizes pixel shuffling to increase the spatial resolution of the feature maps. Finally, the output is processed through another convolutional layer to produce the final high-resolution image. The use of residual scaling within the residual blocks helps stabilize training when using a large number of filters, enhancing the overall strength of the model.

2) Training Setup: For training, we resized DIV2K images to 512x512 pixels as high-resolution targets and 128x128 pixels as low-resolution inputs, ensuring a consistent format and size, with a batch size of 8. The network was trained from scratch using L1 loss, which measures the absolute difference between the predicted super-resolved images and the ground truth high-resolution images. Our ×4 model was trained over 20 epochs using the Adaptive Moment Estimation (Adam) optimizer with a learning rate of 0.0001.

C. VDSR

1) Model description: The VDSR model is characterized by its deep architecture, consisting of 20 convolutional layers that facilitate the learning of complex mappings from low-resolution to high-resolution images. [3] The architecture begins with a convolutional layer that transforms the input RGB image into a feature map with 64 channels, effectively capturing essential image features. This is followed by 18 additional convolutional layers, all maintaining 64 channels, which further clarify the feature representations. A key aspect of VDSR is its use of a single global skip connection, which allows the model to learn the residuals—the differences between the low-resolution input and the high-resolution target. [5] This skip connection not only speeds up the convergence during training but also enhances the model's ability to reconstruct fine details in the output image. The ReLU activation function is applied after each convolutional layer, except for the final output layer, ensuring non-linearity in the model's learning process.

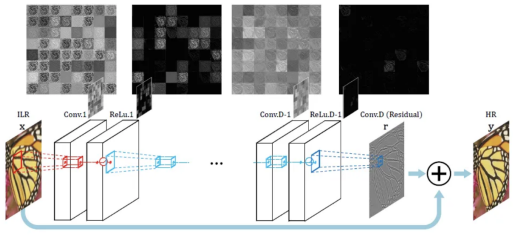


Fig. 2. VDSR Architecture [3]

2) Training Setup: For training, we resized DIV2K images to 128x128 pixels as high-resolution targets and downscaled them to 64x64 pixels, then upscaled back to 128x128 using bicubic interpolation to create low-resolution inputs, with a batch size of 8. The VDSR model was trained from scratch using L2 loss, which calculates the average squared difference between the super-resolved and ground truth high-resolution images. The ×2 model was trained over 50 epochs using the Stochastic Gradient Descent optimizer with a learning rate of 0.001 and momentum of 0.9.

D. DRCN

1) Model description: The DRCN model consist of an architecture that uses a recursive structure to enhance its learning capacity while maintaining computational efficiency. The model begins with an embedding network composed of two convolutional layers, with ReLU activation, which processes the input low-resolution image to extract essential features. [4] Following this, the recursive inference network applies the same convolutional layer—multiple times , allowing the model to generate multiple intermediate outputs through recursive applications of convolutional filters. This weight-sharing mechanism not only increases the effective depth of the network but also helps in capturing intricate details without significantly increasing the number of parameters. After obtaining the recursive outputs, a reconstruction network combines these outputs using additional convolutional layers, including a final layer that maps the features back to the original channel size, while applying a depth-to-space operation to upscale the resolution by a factor of 4. This combination of layers, filters, and recursive learning enables DRCN to effectively perform image super-resolution, making it a powerful tool in the field.

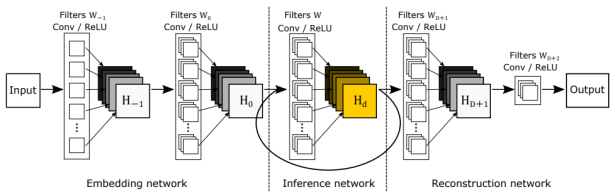


Fig. 3. DRCN Architecture [4]

2) Training Setup: For training, DIV2K images were resized to 512x512 pixels as high-resolution targets and 128x128 pixels as low-resolution inputs, ensuring a consistent format and size, with a batch size of 8. The network was trained from scratch using a custom loss function that combines multiple components, including mean squared error, color loss, recursive loss, and L2 regularization. The ×4 model was trained from scratch for 70 epochs using the Adam optimizer with a learning rate of 0.0006.

E. Evaluation

For evaluation, we tested all three models on the Set5 dataset, using PSNR, SSIM, and PI metrics to assess image quality. Additionally, we recorded inference time and memory usage to evaluate the model's efficiency.

IV. Results/Discussion

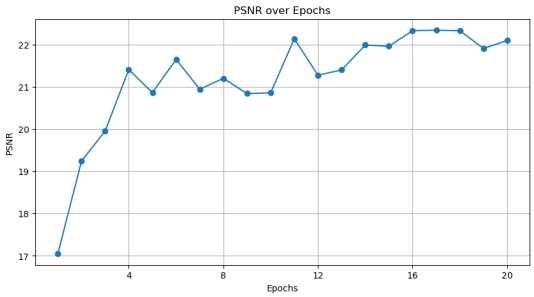


Fig. 4. PSNR over epochs in EDSR

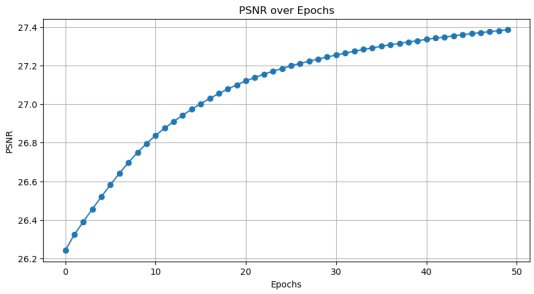


Fig. 5. PSNR over epochs in VDSR

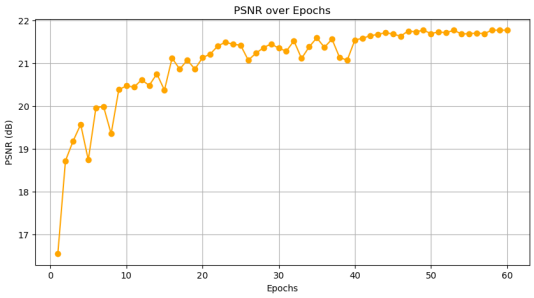


Fig. 6. PSNR over epochs in DRCN

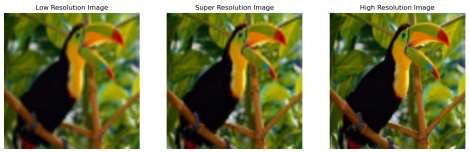
The PSNR for three models generally increases over the epochs, suggesting that the model is learning and improving its

performance. There are some fluctuations, which are common as the model adjusts its parameters. Towards the later epochs, the PSNR values stabilize, indicating that the models are converging and further training might not lead to significant improvements. Overall, these three graphs suggest that the model is effectively learning to enhance image quality over time.

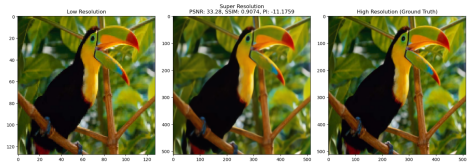
A. Image quality

- EDSR

- VDSR



- DRCN



Model	PSNR	SSIM	PI
EDSR	30.14	0.8666	-5.5048
VDSR	29.82	0.9574	-5.3873
DRCN	31.86	0.8876	-10.3662

TABLE I
PSNR, SSIM, AND PI VALUES FOR THREE MODELS.

EDSR provides a balanced performance with decent PSNR and PI scores but lags slightly in SSIM, suggesting it might not be as good in preserving fine details.
VDSR offers a high SSIM and the best PI, which implies it is highly effective at maintaining structural similarity and perceptual quality, making it ideal for applications prioritizing visual appeal and structural integrity.
DRCN excels in terms of PSNR, indicating strong reconstruction quality, however it does not prioritize perceptual quality a lot as indicated by the lowest PI score.

B. Computational efficiency

EDSR is the largest model with over 43 million parameters which typically indicates higher precision and potentially better quality outputs. However, this resulted in high memory usage (512 MB) and slower inference time, making it less suitable for real-time or resource-constrained applications.
VDSR offers a low parameter count, relatively fast inference time, and moderate memory usage (6MB) that allows for efficient processing without sacrificing much on quality.
DRCN is the most efficient in terms of both speed and memory, making it ideal for applications requiring real-time processing and low memory usage. However, it may slightly underperform in detail-intensive tasks due to its relatively smaller parameter count.

Model	Parameters	Inference Time (ms)	Memory Usage (MB)
EDSR	43,089,923	4126.71	512.48
VDSR	668,227	136.63	6.02
DRCN	1,888,148	101.54	0.21

TABLE II
PARAMETERS, INFERENCE TIME, AND MEMORY USAGE FOR THREE MODELS.

V. CONCLUSION

In this project, we evaluated three super-resolution models—EDSR, VDSR, and DRCN—across reconstruction quality, perceptual similarity, and computational efficiency. DRCN demonstrated the highest PSNR, confirming strong reconstruction capabilities, though it had a lower PI, suggesting it may not emphasize perceptual quality as much. Its efficiency in inference time and memory usage makes it suitable for real-time applications with limited resources. VDSR achieved the highest SSIM and PI, balancing structural and perceptual quality with low computational demands, making it ideal for applications prioritizing visual correctness alongside efficiency. EDSR, with the highest parameter count, produced balanced reconstruction quality but had slower inference and higher memory consumption, making it best for high-resource environments focused on detail preservation. These findings highlight the trade-offs between quality and efficiency in super-resolution, relevant to ongoing research on resource-conscious model design. Future work could explore hybrid approaches that combine strengths of multiple architectures, improve training techniques for faster convergence, or adapt models to diverse datasets for broader applications.

REFERENCES

[1] D. Khaledyan, A. Amirany, K. Jafari, M. H. Moaiyeri, A. Z. Khuzani, and N. Mashhadi, "Low-cost implementation of bilinear and bicubic image interpolation for real-time image super-resolution," *arXiv preprint arXiv:2009.09622*, 2020. [Online]. Available: https://doi.org/10.48550/arXiv.2009.09622

[2] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," *arXiv preprint arXiv:1707.02921*, 2017. [Online]. Available: https://doi.org/10.48550/arXiv.1707.02921

[3] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," *arXiv preprint arXiv:1511.04587*, 2016. [Online]. Available: https://doi.org/10.48550/arXiv.1511.04587

[4] ——, "Deeply-recursive convolutional network for image super-resolution," *arXiv preprint arXiv:1511.04491*, 2016. [Online]. Available: https://doi.org/10.48550/arXiv.1511.04491

[5] V. Romanuke, "An improvement of the vdsr network for single image super-resolution by truncation and adjustment of the learning rate parameters," *Applied Computer Systems*, vol. 24, no. 1, pp. 61–68, 2019. [Online]. Available: https://doi.org/10.2478/acss-2019-0008