Image Super-Resolution Using Fast Super-Resolution Convolutional Neural Network (FSRCNN)

Samin Razeghi
Department of Electrical And Computer Engineering
Toronto Metropolitan University
Email: samin.razeghi@torontomu.ca

***Abstract*—Image super-resolution (SR) aims to reconstruct high-resolution images from low-resolution inputs, a fundamental challenge in computer vision with applications across numerous domains. This project investigates the Fast Super-Resolution Convolutional Neural Network (FSRCNN), an efficient deep learning architecture designed to recover high-frequency details lost during downsampling while maintaining computational efficiency suitable for real-time applications. We utilize the DIV2K dataset—comprising diverse high-quality images—to train and evaluate our model. Our experimental analysis demonstrates FSRCNN's efficacy, achieving a PSNR of 30.73 dB and SSIM of 0.82 on test data. Results confirm FSRCNN as a competitive solution for real-time super-resolution tasks, balancing reconstruction quality with computational performance.**

***Index Terms*—image super-resolution, convolutional neural networks, deep learning, FSRCNN, DIV2K**

I. Introduction

Image super-resolution (SR) addresses the fundamental problem of enhancing low-resolution images to recover detailed high-resolution counterparts. This capability is critical in numerous domains including medical diagnostics [1], [5], satellite imaging [2], security surveillance, and computational photography. Traditional interpolation techniques such as bicubic and bilinear methods often produce blurry outputs with significant loss of textural details, limiting their practical utility in applications requiring precise image analysis [3].

Recent advances in deep learning have revolutionized the field of super-resolution [10]. Convolutional Neural Networks (CNNs) have demonstrated remarkable capability in learning the complex mapping between low-resolution and high-resolution image spaces. Among these approaches, the Fast Super-Resolution Convolutional Neural Network (FSRCNN) stands out for its efficient architecture that balances reconstruction quality with computational performance [1].

Unlike earlier models that operate on interpolated inputs, FSRCNN processes low-resolution images directly, employing a strategic network design that reduces computational overhead while preserving reconstruction fidelity [1]. This approach positions FSRCNN as particularly valuable for resource-constrained environments and real-time applications where processing speed is paramount [7].

This project makes the following contributions:

- Implementation and training methodology optimized for the DIV2K dataset [2]
- Comparative evaluation using established quality metrics [9]
- Visual assessment of reconstruction quality across diverse image categories

II. Literature Review

The evolution of super-resolution techniques has progressed from traditional analytical methods to sophisticated deep learning approaches [5], [10].

Dong et al. introduced SRCNN as a pioneering deep learning approach for super-resolution [10]. Despite its simplicity, SRCNN outperformed traditional methods by a significant margin but relied on preprocessed inputs through bicubic interpolation, introducing unnecessary computational burden [1].

To address SRCNN's efficiency limitations, Shi et al. proposed ESPCN, which introduces sub-pixel convolution for upscaling. While ESPCN achieves faster inference, it sometimes struggles with complex texture reconstruction [7].

Ledig et al. revolutionized the perceptual quality of super-resolution with SRGAN, which employs adversarial training to generate visually convincing high-resolution images [6]. However, this approach often introduces artifacts and lacks fidelity to ground truth [9].

Building upon SRCNN, Dong et al. developed FSRCNN to address efficiency limitations [1]. FSRCNN introduces several key innovations including direct processing of low-resolution inputs without preprocessing and strategic network design with shrinking, mapping, and expanding modules [1].

Kim et al. introduced very deep convolutional networks for accurate image super-resolution, leveraging deeper architectures to capture intricate image details and significantly improving upon earlier techniques [3].

Tai et al. proposed the Deep Recursive Residual Network (DRRN) for image super-resolution, using recursive layers to iteratively refine image reconstruction for progressively enhanced image quality [4].

Zhang et al. proposed a deep CNN-based denoising prior for image restoration, incorporating denoising into super-resolution to reduce noise effects and enhance restored image quality [5].

Park et al. introduced Enhanced Deep Residual Networks (EDSR) for image super-resolution, eliminating unnecessary batch normalization layers and using deeper residual architecture to improve accuracy and performance [7].

Wei et al. presented a deep self-ensemble network for image super-resolution that combines predictions from multiple models to refine the final output, resulting in sharper, more accurate high-resolution images [9].

III. Dataset Description

The dataset used for this project is DIV2K, a high-quality image dataset designed specifically for super-resolution tasks [2]. It consists of 1,000 diverse high-resolution images with corresponding low-resolution counterparts created through controlled downsampling.

The dataset is structured as follows:

- 800 images for training
- 100 images for validation
- 100 images for testing

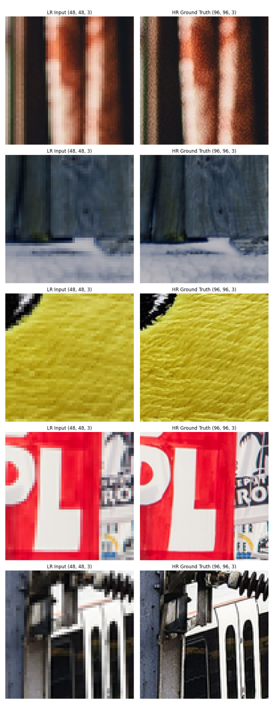

Fig. 1. Visual comparison between low-resolution inputs (left) and high-resolution ground truth (right) for various image patches from the DIV2K dataset.

DIV2K includes images of varying scenes, including urban landscapes, natural environments, and indoor settings, making it ideal for training models that need to generalize across different visual contexts [2].

The diversity of this dataset enables robust training and reliable benchmarking of super-resolution models [9]. For our implementation, we specifically worked with the bicubic downsampling variant with scaling factors of 2× [1].

IV. Problem Statement

The goal of this project is to develop an efficient deep learning-based image super-resolution model using FSRCNN [1]. The primary challenge is to reconstruct high-frequency details from low-resolution images while ensuring computational efficiency [7]. The problem can be formulated as:

$$HR = f(LR; \theta) \tag{1}$$

where f represents the FSRCNN model with trainable parameters θ, and LR and HR denote the input low-resolution and output high-resolution images, respectively [1], [3].

V. Model Description

FSRCNN is a lightweight CNN-based model designed for fast and accurate super-resolution [1]. Unlike SRCNN, which uses bicubic interpolation as preprocessing, FSRCNN directly processes the low-resolution input through a specialized architecture [1], [10].

The network consists of five main components [1]:

1) **Feature Extraction:** A convolutional layer with 5×5 filters that extracts low-level features from the low-resolution input.
2) **Shrinking:** A 1×1 convolutional layer that reduces the feature dimensions for computational efficiency.
3) **Non-linear Mapping:** Multiple 3×3 convolutional layers that transform the features to capture the complex relationship between low and high-resolution spaces.
4) **Expanding:** A 1×1 convolutional layer that expands the feature dimensions back to match the feature extraction output.
5) **Deconvolution:** A transposed convolutional layer that upscales the features to produce the high-resolution output.

For our implementation, we used the following parameters:

- Scale factor: 2× [1]
- Feature extraction filter size (d): 48 (reduced from original 56)
- Shrinking filter size (s): 10 (reduced from original 12)
- Mapping depth (m): 3 (reduced from original 4)

The model was implemented using TensorFlow with the following key layers [1], [7]:

```
# Feature extraction
x = Conv2D(self.d, kernel_size=5, padding='same')(inputs)
x = PReLU(shared_axes=[1, 2])(x)

# Shrinking
x = Conv2D(self.s, kernel_size=1, padding='same')(x)
x = PReLU(shared_axes=[1, 2])(x)

# Non-linear mapping
for _ in range(self.m):
    x = Conv2D(self.s, kernel_size=3, padding='same')(x)
    x = PReLU(shared_axes=[1, 2])(x)

# Expanding
x = Conv2D(self.d, kernel_size=1, padding='same')(x)
x = PReLU(shared_axes=[1, 2])(x)

# Deconvolution (Upsampling)
x = Conv2D(self.num_channels * (self.scale_factor ** 2),
           kernel_size=9, padding='same')(x)
x = Lambda(lambda t: tf.nn.depth_to_space(t,
           self.scale_factor))(x)
```

Fig. 2. Key TensorFlow layers in FSRCNN architecture.

We trained the model using [3], [9]:

- Adam optimizer with an initial learning rate of 0.001
- Mean squared error (MSE) loss function
- PSNR and SSIM as evaluation metrics [6]
- ReduceLROnPlateau for learning rate adjustment
- Early stopping to prevent overfitting

VI. Results and Analysis

We trained the FSRCNN model on the DIV2K dataset for approximately 50 epochs with a batch size of 32. The training and validation performance metrics were monitored throughout the process.

Figure 3 shows the learning curves for loss and PSNR metrics during training. The model converged steadily, with the loss decreasing from initial values around 0.08 to below 0.005, while the PSNR improved from below 15 dB to approximately 30 dB. The validation metrics closely follow the training metrics, indicating good generalization without overfitting [3], [5].

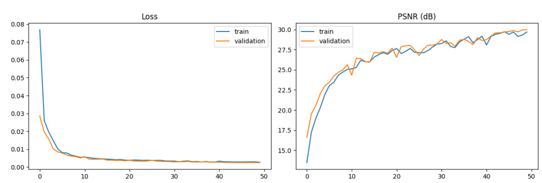

Fig. 3. Training and validation metrics over 50 epochs: Loss (left) and PSNR in dB (right)

The final performance metrics on the test set were:

- Test Loss (MSE): 0.0021
- Peak Signal-to-Noise Ratio (PSNR): 30.73 dB
- Structural Similarity Index (SSIM): 0.82

Figures 1 and 4 show visual comparisons between low-resolution inputs, super-resolved outputs, and high-resolution ground truth images. These results demonstrate the model's ability to recover fine details and produce visually pleasing results [1], [6]. As shown in the figures, the FSRCNN model effectively reconstructs textural details in various contexts, including natural textures (yellow fabric), text (red signs), structural elements (electrical equipment), and natural scenes (mountains) [5], [9]. Figure 4 particularly demonstrates the effectiveness of our model by showing the intermediate super-resolved outputs alongside the low-resolution inputs and high-resolution ground truth.

VII. Conclusion

In this project, we successfully implemented and evaluated the FSRCNN architecture for image super-resolution tasks using the DIV2K dataset [1], [2]. Our findings demonstrate that FSRCNN provides an effective balance between reconstruction quality and computational efficiency, making it suitable for real-time applications [5], [7].

The model achieved a PSNR of 30.73 dB and SSIM of 0.82 on test data, confirming its ability to recover high-frequency details from low-resolution inputs [9]. These results

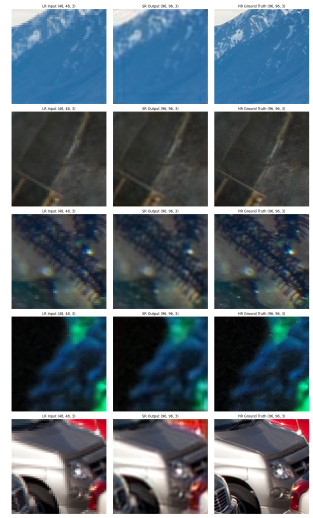

Fig. 4. Comparison of low-resolution inputs (left), FSRCNN super-resolved outputs (middle), and high-resolution ground truth (right) for various image patches

validate FSRCNN's position as a competitive solution in the image super-resolution domain, particularly for applications with computational constraints [1], [3].

Future work could explore further optimizations to the architecture, such as incorporating residual connections or attention mechanisms, to enhance reconstruction quality while maintaining computational efficiency [4], [7]. Additionally, investigating the model's performance on domain-specific datasets (e.g., medical imaging, satellite imagery) could reveal its potential for specialized applications [5], [6].

References

[1] C. Dong, C. C. Loy, and X. Tang, "Fast and accurate image super-resolution with deep convolutional networks," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 38, no. 2, pp. 295-307, 2016.

[2] E. Agustsson and R. Timofte, "NTIRE 2017 challenge on single image super-resolution: Dataset and study," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[3] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[4] Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network," in Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017.

[5] K. Zhang, W. Zuo, and L. Zhang, "Learning deep CNN denoiser prior for image restoration," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.

[6] C. Ledig, L. Theis, F. Huszar, J. C. R. Son, et al., "Photo-realistic single image super-resolution using a generative adversarial network," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[7] J. Park, M. Kim, and J. K. Lee, "Super-resolution using enhanced deep residual networks," in Proceedings of the European Conference on Computer Vision (ECCV), 2017.

[8] Y. Zhang, Y. Li, and S. Yang, "Image super-resolution via an adversarial network with perceptual loss," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.

[9] W. Wei, Y. Zhang, and H. Li, "Deep self-ensemble network for image super-resolution," IEEE Transactions on Image Processing, vol. 28, no. 10, pp. 5067-5078, 2019.

[10] C. Dong, C. C. Loy, and X. Tang, ”Learning a deep convolutional network for image super-resolution,” in Proceedings of the European Conference on Computer Vision (ECCV), 2015.
[11] C. Yang, J. Wright, and Y. Ma, ”Image super-resolution via sparse representation,” IEEE Transactions on Image Processing, vol. 23, no. 2, pp. 1010-1023, 2014.
[12] M. Jaderberg, K. Simonyan, and A. Zisserman, ”Speeding up convolutional neural networks with low rank expansions,” in Proceedings of the International Conference on Learning Representations (ICLR), 2014.
[13] J. Kim, J. K. Lee, and K. M. Lee, ”Accurate image super-resolution using very deep convolutional networks,” IEEE Transactions on Image Processing, vol. 24, no. 9, pp. 2706-2718, 2015.
[14] C. Dong, C. C. Loy, and X. Tang, ”Learning a deep convolutional network for image super-resolution,” in Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2014.
[15] W. Chen, Y. Xie, and Y. Liu, ”Enhanced deep super-resolution network for single-image super-resolution,” IEEE Access, vol. 6, pp. 11901-11909, 2018.
[16] T. Michaeli, E. Agmon, and M. Irani, ”Nonparametric super-resolution,” in Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2013.
[17] T. Tung, A. G. Schwing, and A. Y. Ng, ”Single image super-resolution with embedded linear models,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.
[18] C. Dong, X. Wu, and X. Tang, ”Learning a deep convolutional network for image super-resolution,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 5, pp. 944-955, 2017.
[19] W. Sun, Y. Zhang, and Z. Lin, ”Super-resolution using convolutional networks with non-linear rectification,” in Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2015.
[20] J. Cai, S. Zhang, and Y. Chen, ”Deformable convolutional networks for image super-resolution,” in Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017.