Single Image Super-Resolution with the DIV2K Dataset: a case study

Steenhout, Iris; Asghari, Mahdi

Publication date:
2025

License:
GNU GPL

Document Version:
Final published version

Single Image Super-Resolution with the DIV2K Dataset: a case study

Iris Steenhout[1] and Mahdi Asghari[1]

[1] *Vrije Universiteit Brussel, Brussels, Belgium*

January 7, 2025

Abstract. Single-image super-resolution (SISR) reconstructs high-resolution (HR) images from low-resolution (LR) inputs, addressing hardware limitations and enabling detailed visualization for diverse applications. This study explores the Efficient Sub-Pixel Convolutional Neural Network (ESPCN) for SISR tasks using the DIV2K dataset, focusing on computational efficiency and reconstruction quality. Unlike traditional methods relying on interpolation or iterative processes, ESPCN employs sub-pixel convolution, providing an end-to-end, resource-efficient solution.

ESPCN demonstrates competitive performance, surpassing Bicubic, A+, and SRCNN in Peak Signal-to-Noise Ratio (PSNR) while maintaining a lightweight architecture suitable for real-time applications. However, its Structural Similarity Index (SSIM) suggests room for improvement in structural fidelity, and further evaluation of perceptual quality is needed due to the lack of comparative LPIPS scores.

This study highlights ESPCN's ability to balance efficiency and performance, making it a practical choice for applications like medical imaging, surveillance, and autonomous systems. It contributes to advancing SISR by optimizing accuracy and resource use, paving the way for future developments in deep learning-based super-resolution.

1 Introduction and Motivation

Super-resolution (SR) imaging, particularly single-image super-resolution (SISR), is an innovative computational technique designed to reconstruct high-resolution (HR) images from low-resolution (LR) inputs [1]. This approach addresses inherent limitations in imaging hardware, enabling detailed visualization and analysis across various disciplines. By reconstructing fine details and minimizing artifacts, SR imaging has become a transformative tool with applications in domains such as medical imaging, surveillance, video compression, astronomy, and cultural preservation [2, 3, 4].

In the medical field, SR imaging enhances critical diagnostic tasks by improving the visibility of small anatomical structures, facilitating the detection of tumors, vascular abnormalities, and other pathologies [5, 6]. In security and surveillance, it plays a crucial role in improving image clarity for facial recognition, license plate identification, and forensic investigations under suboptimal conditions [7, 8, 4]. In astronomy, SR has enabled breakthroughs such as refining deep-space images and reconstructing the first black hole image, advancing observational science. Furthermore, SR techniques are central to remote sensing, improving satellite imagery for applications such as disaster management and environmental monitoring [5, 6, 1].

The rapid advancements in machine learning, particularly deep learning, have propelled SR imaging into new realms of accuracy and efficiency [2, 3]. Techniques such as convolutional neural networks (CNNs), residual learning, and attention mechanisms have revolutionized SISR, allowing models to extract intricate details from LR images and generate HR outputs with unparalleled precision [9]. This project focuses on exploring SISR due to its potential to address critical challenges in image quality, resolution, and resource efficiency across diverse fields while contributing to the theoretical and practical advancements in this domain.

The choice of this project stems from its potential to address pressing challenges in various scientific and practical applications while advancing the theoretical boundaries of SISR. By bridging gaps in image quality and resolution, this research aims to provide impactful solutions that resonate across disciplines, reinforcing its role as a pivotal and scientifically relevant endeavor.

1.1 Literature review

Super-resolution (SR) imaging has evolved significantly over the past few decades, transitioning from traditional interpolation and reconstruction-based methods to sophisticated deep learning approaches [10]. These advancements have addressed critical challenges in resolution enhancement, artifact minimization, and computational efficiency, making SR

a cornerstone of modern imaging technologies.

Traditional SR methods, including interpolation-based approaches such as bilinear and bicubic interpolation, served as foundational techniques for upscaling low-resolution (LR) images. However, these methods often resulted in blurred outputs due to their inability to accurately reconstruct high-frequency details [5, 9]. Reconstruction-based methods, which utilize gradient profiles and non-local means, achieved better results by incorporating additional priors into the reconstruction process. Nonetheless, these approaches were computationally intensive and struggled with complex, noisy datasets, particularly in medical imaging and other high-precision applications.

The advent of machine learning marked a paradigm shift in SR imaging. Convolutional neural networks (CNNs) emerged as a powerful tool, capable of learning complex mappings between LR and high-resolution (HR) images [10, 5, 9]. Early CNN-based models, such as SRCNN, utilized shallow networks to extract features and upscale images. Subsequent innovations, such as the inclusion of residual learning and dense skip connections, addressed the vanishing gradient problem in deeper networks, enabling the extraction of intricate features and facilitating the generation of sharper HR images [1].

Incorporating generative adversarial networks (GANs) further enhanced the capabilities of SR imaging. GAN-based models, such as SRGAN and MedSRGAN, introduced perceptual loss functions to improve the visual quality of reconstructed images [5, 6]. These models not only preserved high-frequency details but also minimized noise and artifacts by leveraging adversarial training between generator and discriminator networks. The integration of multi-residual and attention mechanisms, such as in the Deep Residual Feature Distillation Channel Attention Network (DRFDCAN), further optimized feature extraction and reconstruction, significantly enhancing peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) metrics.

Medical imaging has been a primary beneficiary of these advancements. SR techniques are widely applied in modalities such as magnetic resonance imaging (MRI), computed tomography (CT), and ultrasound, where hardware constraints often limit resolution. Deep learning-based SR methods have enabled the enhancement of small anatomical structures critical for diagnosing tumors and vascular abnormalities [1, 9]. For instance, SR has been applied to retinal imaging to detect microvascular changes indicative of hypertensive retinopathy and to cardiac imaging to improve myocardial motion analysis. These improvements extend beyond diagnostics to support surgical planning, functional imaging, and therapeutic monitoring, cementing SR as a vital tool in healthcare.

Beyond the medical field, SR imaging plays a pivotal role in security and surveillance [1]. By enhancing the resolution of low-quality footage, SR improves the accuracy of facial recognition, license plate identification, and crowd monitoring. This capability is particularly valuable in forensic investigations and real-time threat detection, where image clarity is paramount. Additionally, SR improves biometric systems by enhancing iris and fingerprint recognition, bolstering security frameworks.

In remote sensing and satellite imaging, SR methods overcome hardware limitations to provide high-resolution terrestrial observations [1]. These techniques have been instrumental in applications ranging from urban planning and environmental monitoring to disaster management [5]. For example, SR algorithms refine imagery to detect structural damage after natural disasters or to monitor ecological changes. In astronomy, SR aids in refining deep-space images and exploring celestial phenomena, as demonstrated by its role in reconstructing the first black hole image.

Despite its advancements, the need for further development in SR imaging persists. The computational complexity of deep learning-based SR models often limits their deployment in real-time or resource-constrained environments. Hardware-based solutions, such as increasing pixel density or sensor size, are often cost-prohibitive and introduce additional challenges such as noise amplification and reduced signal-to-noise ratios [9]. Efforts to address these limitations have focused on lightweight network architectures, such as DRFDCAN, which optimize computational efficiency by leveraging residual learning and attention mechanisms. These architectures reduce the number of parameters and floating-point operations while maintaining high-quality reconstruction [5, 6]. Furthermore, the integration of vision transformers (ViTs) and hybrid CNN-ViT models has shown promise in capturing long-range dependencies, further improving the fidelity of SR outputs.

The continued evolution of SR imaging is driven by the demand for higher-resolution visuals across disciplines. Unlike traditional hardware-based solutions, which are costly and introduce challenges such as noise amplification, SR algorithms provide a cost-effective alternative that leverages existing systems to enhance resolution [5, 6, 1]. As deep learning techniques advance, the integration of novel architectures and loss functions promises to address current limitations, paving the way for more efficient and versatile SR applications.

In this research, the potential benefits of an Efficient Sub-Pixel Convolutional Neural Network (ESPCN) will be explored [11]. Our model is based on the work of Shi et al. [12]. By integrating advanced attention mechanisms and optimizing network efficiency, this study aims to achieve superior reconstruction quality while maintaining computational feasibility. It is expected that the findings will demonstrate improved PSNR and SSIM metrics that offer advantages for practical applications in fields such as medical imaging and surveillance. This investigation seeks to contribute to the ongoing evolution of SR technology by balancing performance and efficiency in addressing real-world imaging challenges.

1.2 Methodology

1.2.1 Data

For this study, we utilize the DIVerse 2K (DIV2K) dataset, a widely used benchmark for image super-resolution tasks. This dataset provides paired high-resolution (HR) and low-resolution (LR) images, enabling effective supervised training and evaluation of super-resolution models. The low-resolution images are generated through bicubic interpolation downscaling, ensuring consistency across training and validation. Specifically, we use images with a downscaling factor of $\times 4$, where each LR image is derived by reducing the dimensions of the HR image by four times. The dataset is organized into four key folders required for this work:

- DIV2K/DIV2K train LR bicubic/X4/ - low-resolution training inputs,
- DIV2K/DIV2K train HR/ - high-resolution training ground truth,
- DIV2K/DIV2K valid LR bicubic/X4/ - low-resolution validation inputs,
- DIV2K/DIV2K valid HR/ - high-resolution validation ground truth.

1.2.2 Method and model: Background Enhanced Super-Resolution Convolutional Neural Network (ESPCN)

Given a single low-resolution image Y, the ESPCN model directly maps it to a high-resolution image $F(Y)$ through a lightweight, efficient network architecture [11]. Unlike methods that rely on bicubic upscaling as a pre-processing step, this method employs sub-pixel convolution to achieve end-to-end super-resolution. The model comprises three operations:

- **Patch Extraction and Feature Representation** The low-resolution input Y is processed through a convolutional layer, extracting overlapping patches and representing them as high-dimensional feature maps. Formally, the operation is represented as:

  $$F_1(Y) = \phi(W_1 * Y + b_1)$$

  where W_1 and b_1 are learnable weights and biases, and $\phi(\cdot)$ represents the ReLU activation function. Here, the model W_1 includes 64 filters of size 5×5, capturing fine-grained spatial features. To ensure stable training and normalized feature distributions, a batch normalization operation is applied after this layer, improving convergence speed and overall performance.
- **Non-linear Feature Mapping** The extracted features are transformed non-linearly through another convolutional layer:

  $$F_2(Y) = \phi(W_2 * F_1(Y) + b_2)$$

  In our model, W_2 includes 32 filters of size 3×3, enabling a reduction in feature dimensionality while retaining salient high-frequency information. A subsequent batch normalization layer further refines the activations, ensuring consistent feature scaling and enhancing generalization.
- **Sub-pixel Reconstruction** Instead of directly reconstructing the image in the spatial domain, the ESPCN uses a sub-pixel convolution approach:

  $$F(Y) = \text{PixelShuffle}(W_3 * F_2(Y) + b_3)$$

  Our model W_3 consists of $3 \times r^2$ filters (where r is the upscaling factor and equals to 4, in this case) with a kernel size of 3×3. The pixel shuffle operation rearranges the high-dimensional feature map into the desired high-resolution output.
- **Residual Path** Additionally, a bicubic interpolation operation serves as a residual connection, allowing the model to leverage traditional upscaling alongside learned features, thereby refining the final output. The ESPCN network then learns to add high-frequency details on top of this baseline. This helps the network focus on learning finer textures and details, rather than reconstructing the entire structure of the image from scratch.

This formulation offers computational efficiency by minimizing the network depth and leveraging sub-pixel convolution for image reconstruction, avoiding the need for pre-upscaling. The ESPCN model demonstrates competitive performance, particularly suited for real-time applications.

2 Convolutional Neural Networks for Super-Resolution: ESPCN Formulation

Implementation details High-resolution (HR) images were reconstructed from low-resolution (LR) counterparts using the Enhanced Super-Resolution Convolutional Neural Network (ESPCN) with an upscale factor of 4. Our architecture is based on the work of Shi et al. [12].The ESPCN architecture is specifically designed for super-resolution tasks and consists of three convolutional layers followed by a pixel-shuffling layer, which facilitates the upscaling process [11]. The first layer employs a 5×5 convolution with 64 filters and a padding of 2 to extract initial features from the LR inputs. The second layer utilizes a 3×3 convolution with 32 filters and a padding of 1 to enable deeper feature extraction. The third layer applies a 3×3 convolution with 48 filters, generating the output required for pixel shuffling. Finally, the pixel-shuffling operation rearranges low-resolution feature maps into the high-resolution output, effectively achieving the desired upscaling (Fig.1). Additionally, a bicubic interpolation-based residual connection is used within the architecture to assist the ESPCN in refining high-frequency details, without compromising its computational efficiency.

The pixel-shuffling layer plays a crucial role in the ESPCN model as it enables sub-pixel convolution, a computationally efficient technique for upscaling. Instead of directly learning

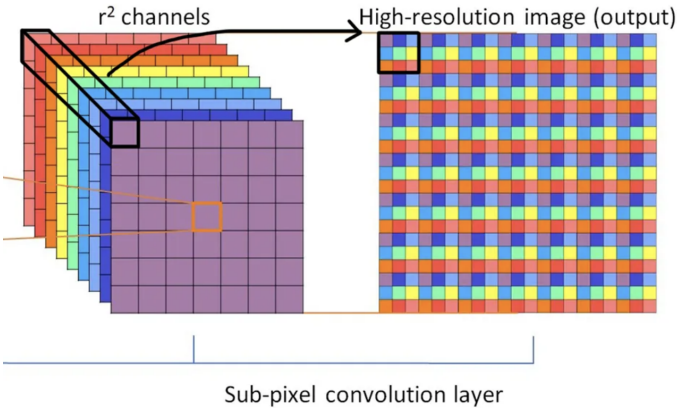

Figure 1: *Operation of pixel shuffle [11].*

HR image features, the network predicts multiple LR feature maps that are spatially rearranged to form the HR image. This approach mitigates the computational overhead associated with traditional upsampling methods, such as deconvolution, while preserving image details. The use of sub-pixel convolution ensures that the information extracted by the earlier convolutional layers is mapped effectively to the desired resolution, significantly improving reconstruction quality.

```
class ESPCN(nn.Module):
    def __init__(self, upscale_factor=4):
        super(ESPCN, self).__init__()

        self.conv1 = nn.Conv2d(3, 64, kernel_size=5, padding=2)
        self.bn1 = nn.BatchNorm2d(64)
        self.conv2 = nn.Conv2d(64, 32, kernel_size=3, padding=1)
        self.bn2 = nn.BatchNorm2d(32)
        self.conv3 = nn.Conv2d(32, upscale_factor**2 * 3,
        kernel_size=3, padding=1)
        self.pixel_shuffle = nn.PixelShuffle(upscale_factor)

    def forward(self, x):
        residual = F.interpolate(x, scale_factor=4, mode='bicubic',
        align_corners=False)
        x = F.relu(self.bn1(self.conv1(x)))
        x = F.relu(self.bn2(self.conv2(x)))
        x = self.conv3(x)
        x = self.pixel_shuffle(x)
        return x + residual
```

Preprocessing of HR and LR image pairs was performed using the DIV2K dataset. Patches of size 40x40 (LR) and 160×160 (HR) were extracted, where the HR patches corresponded to the scaled versions of their LR counterparts by a factor of 4. To enhance the robustness of the model, optional data augmentations were applied during the preprocessing stage with a probability of 50 percent. These augmentations included random horizontal and vertical flips, rotations by angles such as 90°, 180°, or 270°, and combinations of these transformations. The following code snippet illustrates the augmentation process:

```
if augmentations:
    if random() > 0.5:
        LR_patch = LR_patch.transpose(Image.Transpose.FLIP_LEFT_RIGHT)
        HR_patch = HR_patch.transpose(Image.Transpose.FLIP_LEFT_RIGHT)
    if random() > 0.5:
        LR_patch = LR_patch.transpose(Image.Transpose.FLIP_TOP_BOTTOM)
        HR_patch = HR_patch.transpose(Image.Transpose.FLIP_TOP_BOTTOM)
    if random() > 0.5:
        angle = choice([90, 180, 270])
        LR_patch = LR_patch.rotate(angle)
        HR_patch = HR_patch.rotate(angle)
```

For each image pair, 20 augmented patches were generated. LR patches were directly cropped from the LR images, while HR patches were obtained from the corresponding coordinates of the HR images. This augmentation strategy introduced variability into the training dataset, thereby improving the generalizability of the model.

To train the ESPCN model, we estimate the parameters of the network $\Theta = \{W_1, W_2, W_3, b_1, b_2, b_3\}$ by minimizing the loss between the reconstructed high-resolution images $F(Y;\Theta)$ and the corresponding ground truth high-resolution images X. The training process follows a supervised learning paradigm and is optimized for the given training dataset containing high-resolution (HR) image samples I_n^{HR}, $n = 1, \ldots, N$. We generate the corresponding low-resolution (LR) images I_n^{LR}, $n = 1, \ldots, N$. The network is trained using the pixel-wise mean squared error (MSE) of the reconstructed images, defined as the objective function:

$$\mathcal{L}(\hat{Y}, Y) = \frac{1}{N} \sum_{i=1}^{N} \left(\hat{Y}_i - Y_i \right)^2$$

It averages the sum of the squared differences between the predicted pixel value and the ground truth pixel value for the i-th pixel. This ensures that larger errors are penalized more heavily.

The Adam optimizer, with a learning rate of 0.001, was employed for parameter updates over a total of 50 epochs. Data loading and batching were facilitated by PyTorch's DataLoader to ensure efficient processing and shuffling of input data.

```
device = torch.device("cuda" if torch.cuda.is_available() else "cpu")
upscale_factor = 4
num_epochs = 50
batch_size = 16
learning_rate = 0.001
checkpoint_dir = "checkpoints"

criterion = nn.MSELoss()
optimizer = optim.Adam(model.parameters(), lr=learning_rate)

train_model_with_validation(model, train_loader, val_loader,
    criterion, optimizer, num_epochs, device, checkpoint_dir)
```

Evaluation of the model was performed on the validation set. The model's performance was assessed quantitatively using Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM) metrics and Learned Perceptual Image Patch Similarity (LPIPS). Super-resolved images generated from the LR validation inputs were compared against their corresponding HR ground truth images.

3 Analysis: Experimental Results

The training and validation loss curves for the ESPCN model demonstrate effective learning and generalization during the super-resolution task (Fig.2). Both losses exhibit a decreasing

trend, with the training loss declining steadily as the model learns to map low-resolution to high-resolution images using the Mean Squared Error (MSE) loss function. The validation loss consistently remains lower than the training loss, indicating good generalization to unseen data. Minor fluctuations in the validation loss are observed, likely due to variations in the validation dataset. After approximately 28 epochs, the rate of loss reduction diminishes, suggesting the model has nearly converged.

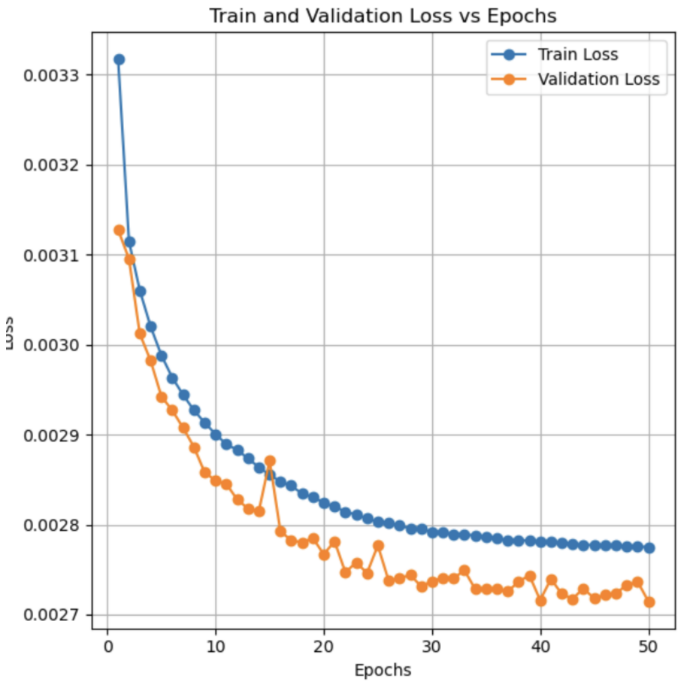

Figure 2: *Training and validation loss over 50 epochs (ESPCN) [13].*

The performance of the ESPCN model is compared against the Bicubic, A+, and SRCNN methods using three evaluation metrics: PSNR, SSIM, and LPIPS (table 1). The Bicubic, A+, and SRCNN results are benchmarked from the paper by Dong et al. [5]. The results of the ESPCN approach, showcasing its super-resolution capabilities, can be visually assessed in Figure 3, which highlights qualitative examples of reconstructed high-resolution images.

The PSNR values show that ESPCN achieves the highest performance, with a score of 30.57, surpassing SRCNN (30.49), A+ (30.28), and Bicubic interpolation (28.42). However, despite the superior PSNR, ESPCN underperforms in SSIM with a value of 0.7855 compared to SRCNN (0.8628), A+ (0.8603) and bicubic (0.8104). The LPIPS score, available only for ESPCN, is 0.2163. LPIPS measures perceptual similarity and lower values indicate higher perceptual quality.

Eval. Mat	Bicubic	A+	SRCNN	ESPCN
PSNR	28.42	30.28	30.49	**30.57**
SSIM	0.8104	0.8603	**0.8628**	0.7855
LPIPS	-	-	-	**0.2163**

Table 1: *Comparison of PSNR, SSIM and LPIPS across methods.*

4 Discussion

The observed behavior of the training and validation loss functions highlights the effectiveness of ESPCN for super-resolution tasks, as the model minimizes reconstruction error consistently during training. This reinforces ESPCN's capability to produce high-quality outputs. However, further refinement could be achieved by incorporating perceptual or adversarial loss components, which focus on optimizing features aligned with human perception rather than solely pixel-level accuracy, thereby improving the perceptual quality of the reconstructed images.

Our results demonstrate that ESPCN achieves superior PSNR compared to Bicubic, A+, and SRCNN methods as benchmarked in Dong et al. [5], indicating that it provides the best reconstruction fidelity at the pixel level. This suggests that ESPCN excels at minimizing pixel-wise differences, making it highly effective for tasks requiring precise numerical reconstruction. However, when evaluated using SSIM, which measures perceived image quality by assessing structural similarity, ESPCN underperforms compared to SRCNN and A+. This suggests potential limitations in preserving structural fidelity, despite its strong numerical accuracy.

The LPIPS score, reported only for ESPCN, indicates competitive perceptual performance with a relatively low value, aligning with the model's high PSNR. Although LPIPS scores for other methods are unavailable in the baseline paper [11], the results suggest that ESPCN produces visually coherent outputs. This trade-off between numerical accuracy and structural similarity highlights the potential for further improvements by integrating perceptual and adversarial losses.

Perceptual loss functions aim to improve the quality of generated images by comparing feature representations in a pre-trained deep neural network (e.g., VGG). Unlike traditional pixel-level losses such as Mean Squared Error (MSE) or Mean Absolute Error (MAE), perceptual loss evaluates discrepancies in feature space, enabling the model to capture high-level semantic differences. The perceptual loss can be represented by:

$$\mathcal{L}_{\text{perceptual}} = \sum_{l=1}^{L} \frac{1}{N_l} \|\phi_l(y) - \phi_l(\hat{y})\|^2,$$

where ϕ_l represents the feature maps of layer l in the pre-trained network, y is the ground truth image, $\hat{y}$ is the predicted image, and N_l is the number of elements in the feature map. By comparing feature representations at various levels, this approach leads to sharper, more realistic images, especially in tasks where fine details or textures are crucial.

Adversarial loss, on the other hand, leverages the framework of Generative Adversarial Networks (GANs) to enhance visual realism. By training the generator to create images indistinguishable from real ones and simultaneously training a discriminator to differentiate between real and generated

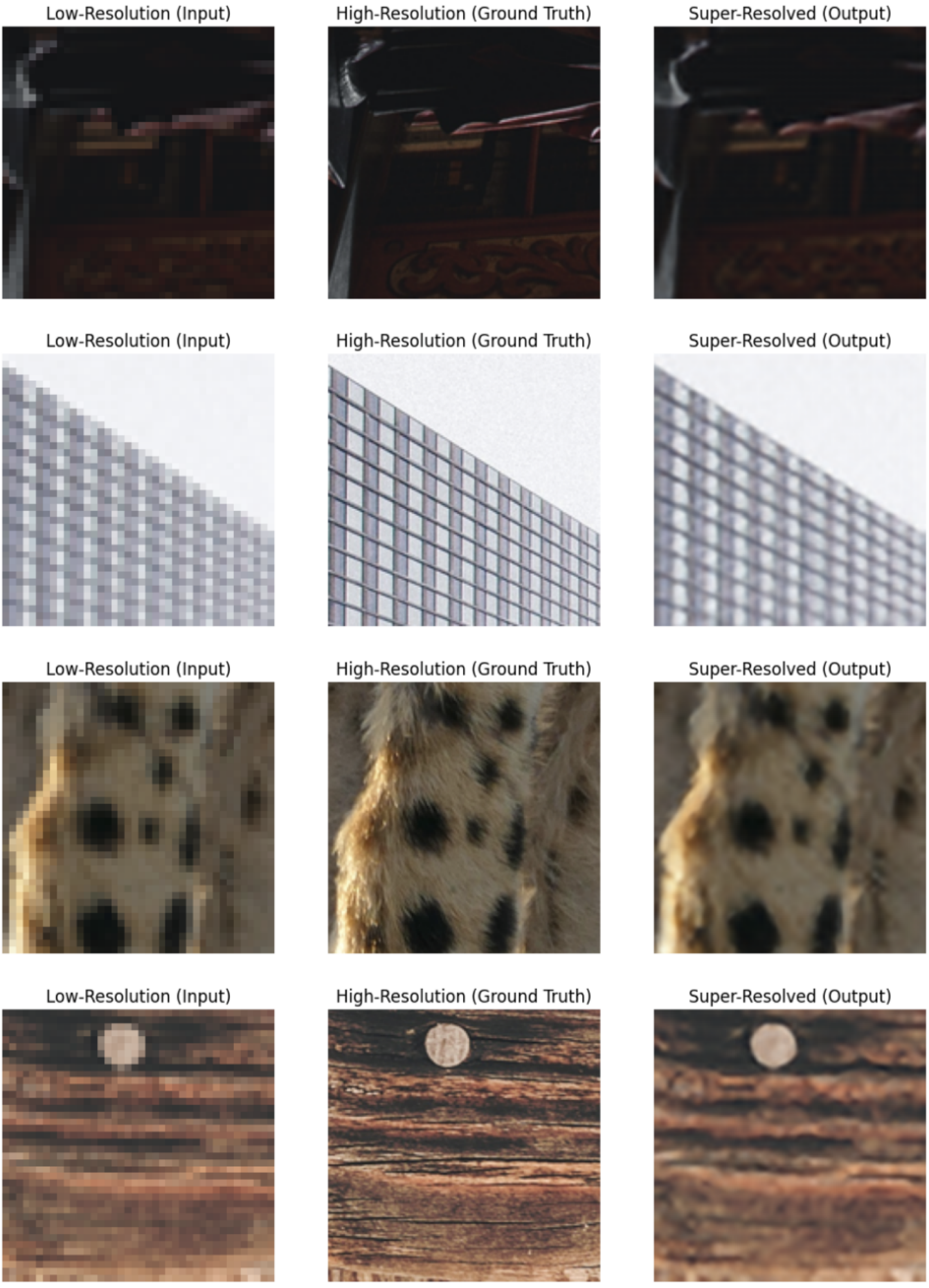

Figure 3: *Random super-resolution examples from DIV2K with an upscaling factor of 4.*

images, adversarial loss encourages the model to produce sharper outputs with improved texture and detail. The adversarial loss is defined as:

$$\mathcal{L}_{\text{adv}} = -\mathbb{E}_{\hat{y}\sim G(x)}[\log D(\hat{y})],$$

where $G(x)$ is the generator output given the input x, $\hat{y} = G(x)$ is the generated image, and $D(\hat{y})$ is the discriminator's probability that $\hat{y}$ is a real image. The generator minimizes this loss while the discriminator maximizes it, creating a min-max optimization problem that leads to more photorealistic outputs.

5 Conclusion and Limitations

From an engineering perspective, ESPCN's design is particularly advantageous for real-time applications, such as video streaming, medical imaging, or autonomous systems. Its ability to eliminate the need for separate optimization steps, coupled with its computational efficiency and scalability, makes it a robust, reliable, and practical solution for image super-resolution tasks in resource-constrained environments. The lightweight architecture ensures low latency during inference, while the end-to-end optimization framework simplifies implementation and reduces computational overhead. These qualities position ESPCN as a promising tool for a wide range of engineering applications, particularly in scenarios where processing speed and hardware limitations are critical.

Despite its advantages, our ESPCN model has notable limitations that warrant further exploration. While the model achieves superior pixel-level accuracy, as indicated by its high PSNR, it underperforms in structural similarity, with lower SSIM scores compared to other methods such as SRCNN and A+. This indicates that ESPCN may struggle to preserve fine structural details and textures in the reconstructed images, which can limit its applicability in tasks where structural fidelity is critical, such as high-resolution medical imaging or geospatial analysis.

Additionally, while the relatively low LPIPS score suggests competitive perceptual performance, the lack of comparative LPIPS data for other methods in the reference paper in the baseline paper [11] prevents a thorough evaluation of ESPCN's perceptual quality. This underscores the need for future studies to benchmark ESPCN against other state-of-the-art models using perceptual quality metrics.

To address these limitations, future work should focus on integrating perceptual loss functions, which prioritize high-level semantic features aligned with human perception, and adversarial loss functions, which encourage the generation of sharper and more realistic outputs. These enhancements could improve both the structural and perceptual consistency of ESPCN, enabling it to produce images that not only achieve high numerical accuracy but also align more closely with human visual expectations.

6 Acknowledgements

Mahdi Asghari (MA) initiated and meticulously executed the coding process, drawing inspiration from the foundational work of Shi et al. [12]. He established the initial model architecture.

Subsequently, Iris Steenhout (IS) advanced the project by designing and implementing additional visualizations, which complemented and enriched the modeling process.

Together, MA and IS rigorously tested the model, collaboratively exploring various hyperparameters and refining its performance. As part of this joint optimization effort, IS implemented a scheduler to terminate training when no improvement was observed after 10 epochs. This process effectively reduces resource usage, particularly for larger datasets, ensuring a robust and efficient final model. MA further enhanced the model by incorporating batch normalization and a residual path, and he also created a GitHub repository for the project and all of its necessary files for easy execution via CLI with a simple README file to explain how everything works. Grigorii Korotych (GK) contributed by integrating the LPIPS testing metric from the lpips library. MA aligned this metric with the self-coded metrics outlined in the task description, ensuring a cohesive and accurate evaluation process.

IS initiated the writing of this paper, crafting the initial framework for its structure and presentation, and conducted the comprehensive literature review to provide a robust theoretical foundation. MA subsequently reviewed the manuscript, ensuring its coherence, clarity, and overall quality.

We also acknowledge the role of ChatGPT in assisting with proofreading and language refinement, enhancing the readability of the text. The final version underwent manual review to ensure accuracy, consistency, and alignment with our intended message.

References

1. Al-Olofi WA and Rushdi MA. Medical Image Super-Resolution. *Artificial Intelligence and Image Processing in Medical Imaging*. Ed. by available) EN(. Elsevier, 2024 :321–9. DOI: 10.1016/B978-0-323-95462-4.00013-3
2. Kim J, Lee JK, and Lee KM. Accurate Image Super-Resolution Using Very Deep Convolutional Networks. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016 :1646–54
3. Zhang Y, Li K, Li K, Zhong B, and Fu Y. Image Super-Resolution Using Very Deep Residual Channel Attention Networks. arXiv preprint 2018 Jul; arXiv:1807.02758:1–16. Available from: https://arxiv.org/pdf/1807.02758

4. Kundu R. Deep Learning for Image Super-Resolution [incl. Architectures]. V7 Labs Computer Vision Blog. 2022. Available from: https://www.v7labs.com/blog/image-super-resolution-guide
5. Ahmad W, Ali H, Shah Z, and Azmat S. A new generative adversarial network for medical images super resolution. Nature Scientific Reports 2022; 12. DOI: 10.1038/s41598-022-13658-4
6. Umirzakova S, Mardieva S, Muksimova S, Ahmad S, and Whangbo T. Enhancing the Super-Resolution of Medical Images: Introducing the Deep Residual Feature Distillation Channel Attention Network for Optimized Performance and Efficiency. Bioengineering 2023; 10:1332. DOI: 10.3390/bioengineering10111332
7. Dong C, Loy CC, He K, and Tang X. Image Super-Resolution Using Deep Convolutional Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence 2015; 38:295–307
8. Kim J, Lee JK, and Lee KM. Deeply-Recursive Convolutional Network for Image Super-Resolution. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016 :1637–45
9. Qiu D, Cheng Y, and Wang X. Medical image super-resolution reconstruction algorithms based on deep learning: A survey. Computer Methods and Programs in Biomedicine 2023; 238:107590. DOI: 10.1016/j.cmpb.2023.107590
10. He J, Ma H, Guo M, Wang J, Wang Z, and Fan G. Research into super-resolution in medical imaging from 2000 to 2023: bibliometric analysis and visualization. Quantitative Imaging in Medicine and Surgery 2024; 14:5109–30
11. Cen zhuo. An Overview of ESPCN: An Efficient Sub-pixel Convolutional Neural Network. Medium. 2020. Available from: https://medium.com/@zhuocen93/an-overview-of-espcn-an-efficient-sub-pixel-convolutional-neural-network-b76d0a6c875e
12. Shi W et al. Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Network. Computer Vision Foundation. 2016. Available from: https://www.cv-foundation.org/openaccess/content_cvpr_2016/papers/Shi_Real-Time_Single_Image_CVPR_2016_paper.pdf
13. J. Bradley E. Schelbert GL et al. A cox-based risk prediction model for early detection of cardiovascular disease: identification of key risk factors for the development of a 10-year CVD risk prediction. Advances in Preventive Medicine 2022 :1–19