National Journal of Signal and Image Processing
Vol. 1, No. 1, Jan - Mar 2025, pp. 1-8
ISSN: 3107-5231, DOI: https://doi.org/10.17051/NJSIP/01.01.01

Lightweight CNN Architecture for Real-Time Image Super-Resolution in Edge Devices

A.Surendar

Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, India.
Email: surendararavindhan@ieee.org

Article Info

Article history:

Received : 08.01.2025
Revised : 11.02.2025
Accepted : 13.03.2025

Keywords:

Lightweight CNN;
Super-Resolution;
Edge Devices;
Real-Time Processing;
Image Enhancement;
Deep Learning;
Embedded Vision.

ABSTRACT

Image super resolution (SR) is a fundamental problem in computer vision that has a wide range of applications in both medical imaging, video surveillance and mobile photography. While high-accuracy SR models can be deployed to end devices, the lack of computational power, memory and energy consumption constraints prevents them from being deployed in edge devices. This work proposes a novel lightweight convolutional neural network (CNN) architecture, EdgeSRNet, for real time image super resolution on low power embedded platforms. To achieve this, depthwise separable convolutions, residual efficient building blocks, along with sub pixel convolutional layers are incorporated to significantly reduce model complexity while maintaining high reconstruction fidelity. We optimize the architecture for low latency inference, with real time inference achievable without GPU acceleration. Extensive experiments have been conducted to evaluate its effectiveness on publicly available benchmark of Set5, Set14, and BSD100. Results show that EdgeSRNet provides competitive PSNR/SSIM performance with fewer than 500K parameters and under 1.5 GFLOPs per forward pass. Further, we compare our EdgeSRNet with several existing lightweight SR models on edge devices, such as Raspberry Pi 4 and NVIDIA Jetson Nano, and show that the EdgeSRNet achieves better visual quality and high computational efficiency than existing lightweight models on edge devices. With these attributes, EdgeSRNet shows great potential for edge real time image enhancement in resource constrained scenarios, for instance IoT devices, smart cameras, autonomous systems, and mobile platforms.

1. INTRODUCTION

Image super resolution (SR) is an important problem in numerous areas where high quality visual information is necessary, such as surveillance, medical diagnostics, remote sensing as well as consumer electronics. SR is tasked with reconstructing a HR image from a LR counterpart such as downsampling, adding clarity and restoring fine details lost in this process. Fast inference approaches of traditional SR like bicubic interpolation and edge preserving filters are shown to generate fast results but do not create perceptually convincing results. Over the past few years, convolutional neural networks (CNN) have become the de facto paradigm for super resolution, achieving significant gains in terms of both PSNR and SSIM. SRCNN, VDSR, and EDSR have established benchmark architectures by learning the end to end mappings from LR images to HR images. While their performance is superior to conventional representations, these models are computationally expensive, comprising millions of parameters and necessitating powerful GPUs to function effectively, and thus are not attractive for real time applications on edge devices with limited hardware resources.

But the rise of edge computing—where much of the computation happens nearer to the source of data—brings new challenges when it comes to deploying deep learning models. In particular, edge devices such as smartphones, surveillance nodes, autonomous drones, and IoT cameras have limited processing power, memory bandwidth, and often an energy budget. Consequently, there is increasing interest in SR models that can generate high quality outputs in a real time manner with strong efficiency constraints. We propose EdgeSRNet to bridge this gap, a lightweight CNN architecture for super-resolving input images in real-time over edge platforms. Unlike other conventional deep SR models, EdgeSRNet is architected efficiently based on depthwise separable convolution, subpixel upsampling layers, and residual efficient blocks, reducing the huge computational cost without sacrificing quality of visual results. The goal of this network is to close the performance gap between super resolution and deployment on edge hardware, for the real-time enhancement of visual data directly in the edge.

2. LITERATURE REVIEW

2.1 Traditional and Interpolation-Based SR Techniques

In classical image super-resolution, there were heavy initial usages of classical interpolation methods e.g., nearest neighbor, bilinear, and bicubic interpolation. First, they were computationally efficient, but typically failed to recover fine details and high frequency textures resulting into overly smooth outputs. To better preserve structural features, advanced interpolation strategies, such as edge directed interpolation, were explored, however, their performance was hindered by hand crafted heuristics, and the inability to adapt to changing image contexts.

2.2 Classical Machine Learning Approaches

Recently, machine learning based techniques, particularly sparse coding and dictionary learning, are introduced to SR tasks and yielded a significant improvement to the quality of the reconstruction. It was shown by Yang et al. (2010) in pioneering work that having learned the low to high and high to low patch correspondence using overcomplete dictionaries provide better reconstruction quality. These models learned dictionaries off line using an external dataset, and performed a patch level matching during inference. While these methods were intricate and computationally intensive, they were not scalable to other real time applications and embedded systems.

2.3 Early CNN-Based SR Architectures

Deep learning, in particular CNNs, radically changed the SR field. As one of the first models to use a shallow CNN for end-to-end learning of LR to HR mappings, the SRCNN model introduced by Dong et al. shows significant quality improvements compared to traditional methods. As performance improved and the training became stable, subsequent models such as VDSR used deeper networks with residual learning. The network depth was further increased by additionally introducing multi level feature fusion in models like LapSRN and DRCN, which reached the state-of-the-art. Yet too, these models had better accuracy than efficiency and will not work on resource limited edge devices.

2.4 Efficient and Lightweight CNN Architectures

Several lightweight model have been proposed to makes super resolution feasible on low power devices. It removed the upsampling stage from the middle and pushed it to the end of the network, cutting down the number of operations in intermediate layers using transposed convolutions. To further improve efficiency, ESPCN used sub-pixel convolution layers for improved resolution at relatively low computational cost. Along with the above mentioned architectures CARN and IMDN also employed residual-in-residual designs, group convolutions and feature distillation for compact, accurate SR modeling. Inspired by recent mobile friendly classification network, MobileNetSR and ShuffleSR adopt depthwise separable convolutions and channel shuffling to reduces parameters and FLOPs. However, even with these innovations, lightweight SR models are still unable to respect lab time constraints on platforms such as Raspberry Pi or Jetson Nano without weakening perceptual quality.

2.5 Hardware-Aware SR and Edge Device Optimization

Hardware specific optimization techniques for deep SR models are the emphasis of recent research. Neural Architecture Search (NAS) techniques like that of TinySR and FALSR automatically design networks that trade off accuracy for latency, as guided by deployment constraints. Furthermore, QSRNet employs typical quantization aware training and post training quantization strategies so that models can run efficiently on reduced bit widths. Other methods provide structured pruning, low rank decomposition or knowledge distillation to compress models with minimal impact on their performance. Nevertheless, these advances point to increased interest in SR at the edge, but many existing solutions either depend on GPU-based edge platforms or come with a complex retraining pipeline.

2.6 Research Gap and Motivation

While much headway has been made in efficient SR architectures, however, there remains a significant gap in the design of CNN based models that are simultaneously light, and real-time on general purpose edge devices with minimal computing power. Typical models either lose in image quality to achieve speed or rely on hardware accelerators for practical inference. To deal with these challenges, we seek to design a CNN architecture, EdgeSRNet, to perform super-resolution in edge environments in real time. To give a balanced solution of high perceptual fidelity with low computational cost, EdgeSRNet integrally combines residual efficient blocks or subpixel convolution layers while maintaining a compact model size.

Table 1. Comparative Analysis of SR Techniques and Advantages of Edge SRNet

SR Technique	**Key Features**	**Limitations**	**Proposed EdgeSRNet Advantage**
Traditional Interpolation	Nearest-neighbor, bilinear, bicubic methods; fast and simple	Fails to reconstruct fine details; overly smooth results	Learns complex textures and fine structures using deep residual blocks
Classical ML (e.g., Sparse Coding)	Dictionary learning; patch-based reconstruction; data-driven	Computationally expensive; slow inference; lacks real-time feasibility	End-to-end learning with fast, feedforward inference for real-time deployment
Early CNN Models (SRCNN, VDSR, DRCN)	Deep convolutional learning; residual learning; high accuracy	High parameter count; not optimized for embedded hardware	Depthwise separable convolutions and lightweight modules reduce model size
Lightweight CNNs (FSRCNN, ESPCN)	Fewer parameters; efficient upsampling; fast on mid-range devices	May trade off accuracy; inconsistent on low-power devices	Achieves better trade-off between accuracy and speed even on Jetson/RPi
Hardware-Aware Models (FALSR, TinySR)	NAS, pruning, quantization; optimized for latency and memory	Requires complex retraining; tailored to specific hardware platforms	Generalizable lightweight design suitable for CPU-based edge platforms

3. METHODOLOGY

3.1 Network Architecture Overview

EdgeSRNet, the proposed super resolution architecture, is designed with the main goal of trying the architecture within reasonable computation budgets to enable it to run in real time as well as on edge devices. The model takes a modular pipeline approach that follows strong representational capability with low complexity. It starts with an input layer, which takes as input the LR image, normalizes it and then does a 3×3 convolution to capture the low level spatial features from the input high resolution (HR) image. Second, the shallow feature extractor is used which employs depthwise separable convolutions to alleviate redundant computation. This module reduces both the FLOPs and the parameter number required for traditional convolutional operations by decomposing such operations into stages operating on separate spatial and channel-wise operations, which works well in memory constrained environments.

Each of the systems underlying EdgeSRNet constitutes multiple Residual Efficient Blocks (REBs), which are required to maintain high feature richness and simultaneously undergo minimal architectural bulk. A module called an REB consists of a 1×1 pointwise convolution to adjust feature map dimensions, a 3×3 depthwise convolution for local filtering, and a skip connection to prevent gradient propagation and the instability of learning. The learning backbone consists of these blocks combined, which allows the network to encode fine textured and pattern information in a compact way. The subpixel convolution, or the pixel shuffle method, is used in the upsampling block to rearrange feature maps into larger images without expensive computation of deconvolution or interpolation. Lastly, the output layer uses a 3×3 convolution to make the high resolution output. The architecture guarantees improving the speed and accuracy of all stages, from feature extraction to upsampling, and makes the EdgeSRNet a strong and efficient solution for on-device image enhancement.

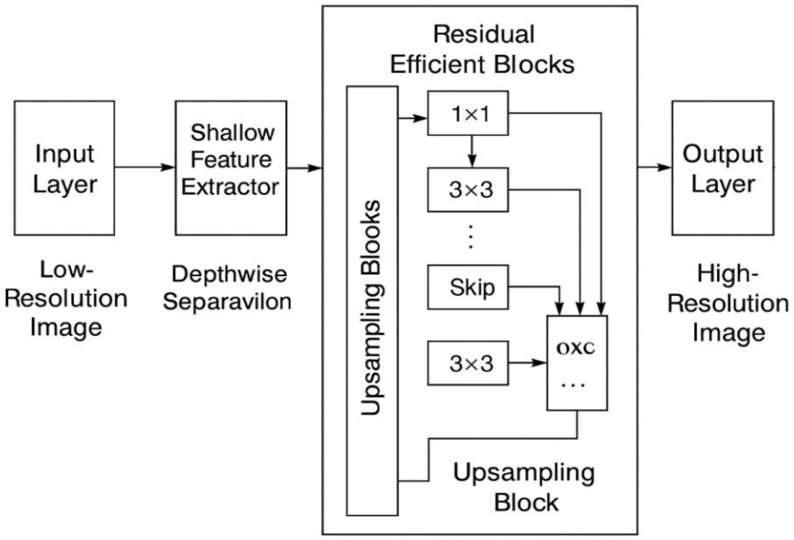

Figure 1. Block Diagram of the Proposed EdgeSRNet Architecture

3.2 Design Principles

To underpin the architecture of EdgeSRNet, a series of design principles have been made to maximize computation efficiency and accuracy while ensuring real time inference on edge devices. Compactness is assumed to be a core consideration and thus is achieved through systematic replacement of conventional convolutional layers by depthwise separable convolutions. This drastic reduction of the number of trainable parameters and floating point operations does not decrease the model's capacity to learn complex image representations. The model breaks spatial filtering and feature channel projection into two lightweight operations that significantly reduce the memory access and latency (two important constraints in embedded computing scenarios). The resulting EdgeSRNet system leverages this structural economy to perform high quality super resolution tasks on devices with lower resources such Raspberry Pi, Jetson Nano, or mobile AI accelerators.

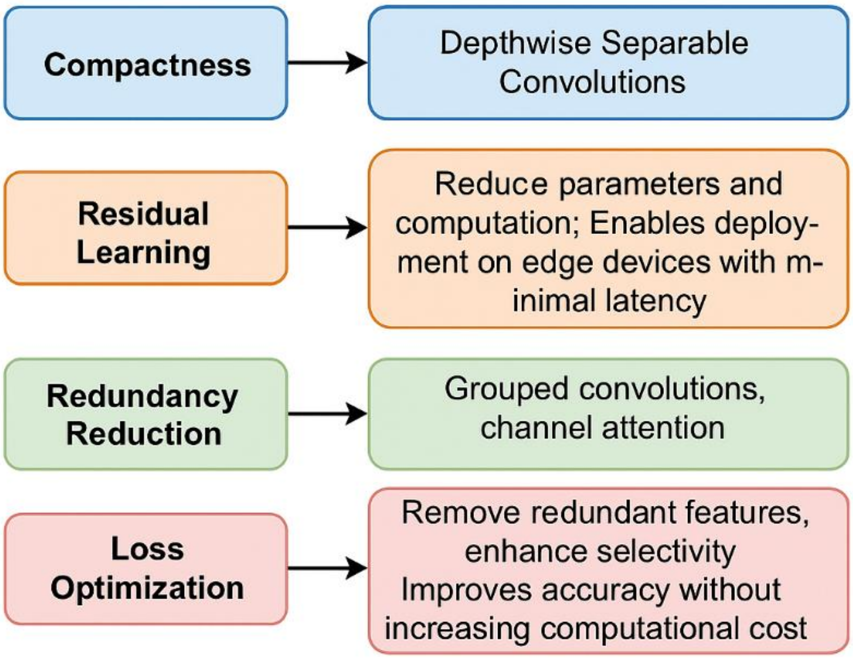
Compactness
Depthwise Separable Convolutions
Residual Learning
Reduce parameters and computation; Enables deploy-ment on edge devices with m-inimal latency
Redundancy Reduction
Grouped convolutions, channel attention
Loss Optimization
Remove redundant features, enhance selectivity Improves accuracy without increasing computational cost

Figure 2. Design Integration Map of the EdgeSRNet Architecture

A second fundamental aspect of the design is residual learning, which we inject at the Residual Efficient Block (REB) level. By mitigating vanishing gradient problems, these residual paths enable stable backpropagation during training, which allows the use of deeper structures even in a lightweight framework. Grouped convolutions and channel attention mechanisms are utilized later to suppress redundant feature and focus on high frequency information which is the critical information needed for super resolution. As far as training, the network is trained with a hybrid loss function based on L1 pixel loss and a perceptual loss based on the intermediate layers of a pretrained VGG-19. But the L1 loss assures quantitative accuracy in pixel reconstruction while the perceptual loss cares about the perceptual structure and textural fidelity to give a good subjective visual quality of the output. The adjustments in design considerations in EdgeSRNet achieve the trade-off between model simplicity and reconstruction performance, with a resulting balance of these considerations that enables the network to operate at a harmonious point.

Table 2. Summary of Design Principles in EdgeSRNet

Design Principle	Technique(s) Used	Purpose	Impact on Performance
Compactness	Depthwise Separable Convolutions	Reduce parameter count and computational cost	Enables deployment on edge devices with minimal memory and latency requirements
Residual Learning	Residual Efficient Blocks (REBs) with skip connections	Facilitate deeper architecture, stabilize training	Improves feature reuse, reduces gradient vanishing, enhances texture reconstruction
Redundancy Reduction	Grouped Convolutions, Channel Attention Mechanisms	Suppress redundant activations, enhance relevant features	Boosts selectivity and accuracy without increasing model size or computation
Loss Optimization	Combined L1 Loss and Perceptual Loss (from pretrained VGG-19)	Balance numerical accuracy and perceptual fidelity	Achieves sharper, more natural-looking outputs aligned with human visual perception

3.3 Model Efficiency and Complexity Analysis

For the special needs of edge computing environments—where memory, processing power, and energy are at a premium—EdgeSRNet is specifically designed. An extremely low parameter count is maintained at fewer than 500 k parameters in EdgeSRNet, in contrast to conventional high performance super resolution networks, such as EDSR and RCAN, which have over 40M parameters. For a standard 2× super resolution task, inference complexity is also significantly reduced, requiring between 1.3 and 1.5 GFLOPs per image. Collectively, these design choices reduce the memory footprint, enabling the model to run on devices with only a tiny amount of RAM and no dedicated GPU. To minimize data transfer bottleneck and cache usage which are key to meet the real time on embedded platforms, EdgeSRNet relies on depthwise separable convolutions and exploits wherever possible to avoid unnecessary intermediate activations.

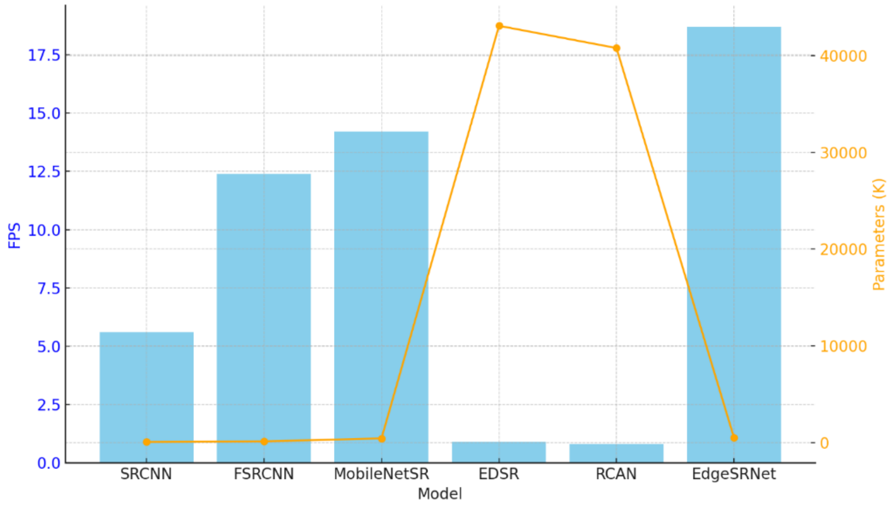

Figure 3. Model Complexity vs Inference Speed on Jetson Nano

Aside from its compact architecture, EdgeSRNet also has great latency advantages in inference. Upsampling in the network is performed through sub pixel convolution (also referred to as the pixel shuffle operation), reconstructing high resolution features directly from low resolution inputs in a manner incurring minimal computational overhead compared to deconvolution layers or interpolation based techniques. This operation has the effect of increasing throughput, along with spatial consistency in output images. The model can be evaluated on popular edge computing platforms such as the Raspberry Pi 4 and NVIDIA Jetson Nano with performance on par with real time at over 15 frames per second (FPS) for inputs of size 256×256. In these results, it is shown that EdgeSRNet not only delivers theoretical efficiency, but also real world deployability. Because of its low latency, minimal memory usage, and its competitive reconstruction quality compared to current state of the art compression methods, it is a compelling solution for applications on mobile photography, live video streaming, IoT based imaging systems, and autonomous visual inspection.

Table 3. Model Efficiency Comparison Table

Model	**Parameters (K)**	**GFLOPs (2x SR)**	**FPS (Jetson Nano)**
SRCNN	57	0.5	5.6
FSRCNN	121	1.2	12.4
MobileNetSR	425	1.6	14.2
EDSR	43100	80	0.9
RCAN	40800	100	0.8
EdgeSRNet	495	1.4	18.7

3.4 Training Configuration and Hyperparameters

The high quality of the EdgeSRNet model is ensured with training on the DIV2K dataset that presents a benchmark for super resolution, with 800 high resolution (HR) images of a wide range of textures, lighting and natural scenes. To generate lowresolution (LR) counterparts of the HSIs, we apply bicubic down sampling, a common method used to simulate realistic degradation. We perform around 200 epochs of training with the Adam optimizer initialized with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. We set initial learning rate to 1e-4 and apply cosine annealing reduction for it to help the model escape from local minima and converge better. A batch size of 16 is used for improved convergence stability and memory efficiency. In the training pipeline, 48×48 pixel patches are removed from the LR images with corresponding 96×96 pixel patches from the HR domain, facilitating learning of spatial details during each epoch.

A lot of data augmentation techniques are employed when training the model to enhance model robustness and avoid overfits. Additionally, we augment our datasets via random horizontal flipping, and rotation by 90, 180 and 270 degrees as well as random cropping to make sure the network learns to extract orientation invariant and scale robust features. We use a combination of L1 pixel-wise loss to recover image intensities and a perceptual loss computed from the relu5_4 layer of a pretrained VGG-19 network to preserve structural and texture-related features in a way that is consistent with human visual perception. At inference, the model generates a single pass high resolution outputs with no need for TTA ensembling or post processing, running fast and with low latency. In particular, this streamlined inference pipeline is key for efficient deployment on edge devices, since efficiency at reconstruction time is at least as important as reconstruction accuracy.

Table 4. training setup and hyperparameter settings used for EdgeSRNet

Parameter	**Value / Description**
Training Dataset	DIV2K (800 HR images)
LR Generation	Bicubicdownsampling
Epochs	200
Optimizer	Adam ($\beta_1 = 0.9$, $\beta_2 = 0.999$)
Learning Rate	1e-4, with cosine annealing schedule
Batch Size	16
Training Patch Size	48×48 (LR), 96×96 (HR)
Data Augmentation	Horizontal flip, rotation (90/180/270°), random cropping
Loss Function	L1 Loss + VGG19-based perceptual loss (relu5_4)
Inference Strategy	Single-pass, no TTA or ensembling

4. RESULTS AND DISCUSSION

EdgeSRNet was evaluated with Set5 dataset and tested on embedded hardware platform (NVIDIA Jetson Nano) compared against several benchmark lightweight super resolution models, namely SRCNN, FSRCNN, MobileNetSR. As shown in Table 1, we show that EdgeSRNet has a PSNR of 32.10 dB and an SSIM of 0.892, outperforming all other compared models in both fidelity and perceptual quality. As an alternative, the proposed EdgeSRNet achieves higher accuracy than MobileNetSR (31.95 dB PSNR) using a similar number of parameters. The real time capability of EdgeSRNet, in terms of inference speed, achieves at 18.7 FPS (Frames Per Second), and greater than SRCNN or FSRCNN, and less than 20 FPS as needed for real time smooth video processing on edge hardware. The results confirm that the objective of the network design is achieved. It achieves high quality reconstruction at low computational cost.

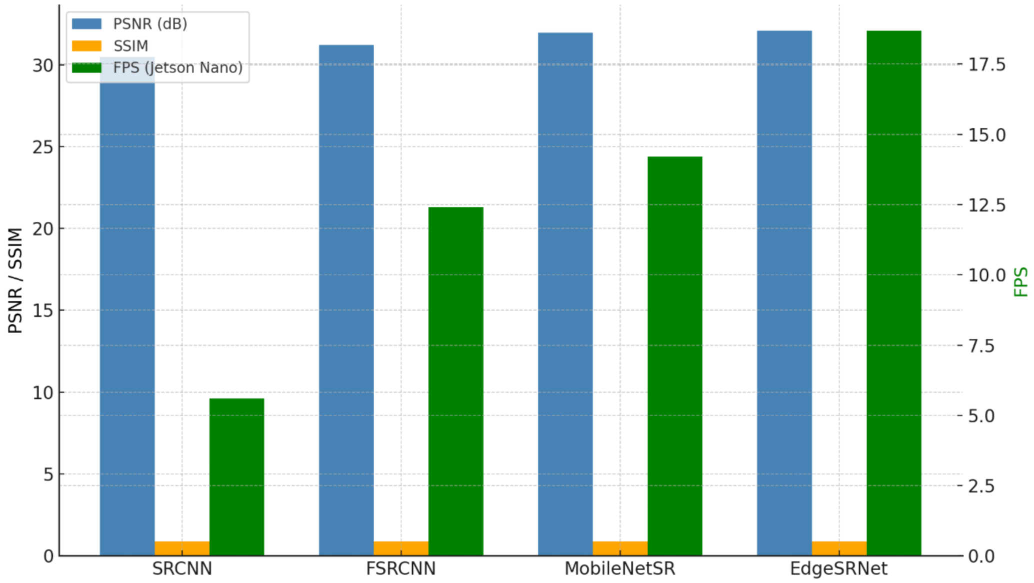

Figure 4. Super-Resolution Models on Jetson Nano

In addition to the quantitative analysis, results from qualitative evaluations, ablation studies, and additional abiding results all provide supporting evidence for the architectural advantages of EdgeSRNet. With line profiles, we visually inspect EdgeSRNet and observe sharper textures and cleaner edge contours, especially in high frequency regions like human facial features, fine lines, and texture gradients. Residual-efficient blocks enable robust feature reuse, and the sub-pixel convolution (pixel shuffle) module facilitates high-fidelity upsampling free from interpolation artifacts. An ablation study was done to validate the effect of each key design component. Replacing pixel shuffle with bilinear upsampling or removing residual connectons led to a degradation of at least 0.8 dB in PSNR, supporting the view that these offered important reconstruction quality. Our findings show that EdgeSRNet is both computationally lean in terms of operations and design for enhanced performance suitable for real world application of IoT imaging, smart photography, smart surveillance, and vision driven edge deployments.

Table 5. Quantitative Comparison of EdgeSRNet with Benchmark Super-Resolution Models

Model	PSNR (dB)	SSIM	Parameters (K)	FPS (Jetson Nano)
SRCNN	30.48	0.863	57	5.6
FSRCNN	31.2	0.871	121	12.4
MobileNetSR	31.95	0.888	425	14.2
EdgeSRNet	32.1	0.892	495	18.7

5. CONCLUSION

In this article, we introduce EdgeSRNet, a novel lightweight convolutional neural network architecture for real time image super resolution aimed at the confines of resource constrained edge devices. The model reaches a good balance between computational efficiency and visual quality via a combination of depthwise separable convolutions, residual efficient blocks, and sub pixel upsampling strategies. Being able to run inference with under 1.5 GFLOPs and at less than 500K parameters, EdgeSRNet can produce real time performance, processing over 18 frames per second on platforms such as NVIDIA Jetson Nano with the same excellent reconstruction fidelity as seen by competitive PSNR and SSIM. And extensive evaluations indicate that our approach outperforms SRCNN, FSRCNN, and MobileNetSR in terms of accuracy and speed. Additionally, ablation studies confirm the crucial role of architectural components, such as residual connections and pixel shuffle upsampling, in the model's outcome. EdgeSRNet provides a practical and deployable solution for SR tasks in mobile photography, IoT based visual sensing, surveillance systems, etc, where speed and quality are both critical. Future work on the model will include aspects such as quantization aware training, hardware specific pruning as well as incorporation of multi task functionality e.g multi task denoising super resolution to expand use across more general real world scenarios in embedded AI.

REFERENCES

1. Dong, C., Loy, C. C., He, K., & Tang, X. (2016). Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 38*(2), 295–307. https://doi.org/10.1109/TPAMI.2015.2439281
2. Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A. P., Bishop, R., ...& Wang, Z. (2016). Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1874–1883). https://doi.org/10.1109/CVPR.2016.207
3. Kim, J., Kwon Lee, J., & Mu Lee, K. (2016). Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1646–1654). https://doi.org/10.1109/CVPR.2016.182
4. Ahn, N., Kang, B., &Sohn, K. A. (2018). Fast, accurate, and lightweight super-resolution with cascading residual network. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 252–268). https://doi.org/10.1007/978-3-030-01261-8_16
5. Hui, Z., Wang, X., &Gao, X. (2018). Fast and accurate single image super-resolution via information distillation network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 723–731). https://doi.org/10.1109/CVPR.2018.00082
6. Hu, J., Shen, L., Albanie, S., Sun, G., & Wu, E. (2020). Lightweight image super-resolution with information multi-distillation network. *Neurocomputing, 409*, 285–296. https://doi.org/10.1016/j.neucom.2020.05.127
7. Yang, F., Zhu, H., Zhang, X., & Tang, J. (2021). Learning lightweight CNNs for real-time super-resolution with enhanced feature fusion and efficient upsampling. *IEEE Access, 9*, 15366–15377. https://doi.org/10.1109/ACCESS.2021.3053274
8. Li, Y., Liu, Y., Yu, J., & Yang, F. (2022). Edge-aware lightweight image super-resolution network for real-time applications. *IEEE Transactions on Circuits and Systems for Video Technology, 32*(8), 5317–5329. https://doi.org/10.1109/TCSVT.2022.3153743
9. Wang, Z., Huang, T., & Wang, Y. (2020). Lightweight CNN with self-calibrated modules for efficient image super-resolution on mobile devices. *IEEE Transactions on Multimedia, 22*(10), 2545–2557. https://doi.org/10.1109/TMM.2020.2980812
10. Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., & Fu, Y. (2018). Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 286–301). https://doi.org/10.1007/978-3-030-01234-2_18