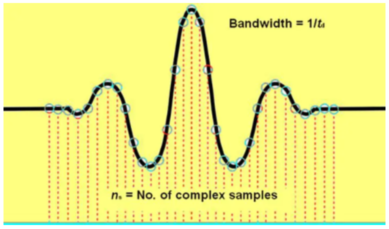
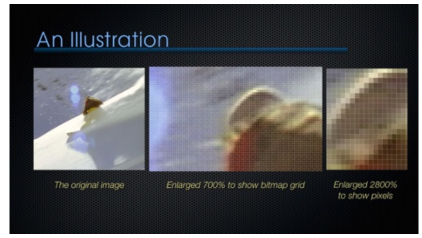
Deep Learned Super Resolution for Low Bandwidth Environments

Sadia Fathima





Abstract

Bandwidth remains a critical bottleneck in delivering high-fidelity image and video content to users in rural, mobile, and resource-constrained environments. Deep learning-based super-resolution (SR) techniques such as SRGAN, ESRGAN, Transformer-based SR, and diffusion models offer a transformative approach to reconstructing high-quality visual content from low-bitrate, low-resolution sources. This work investigates the state-of-the-art in deep SR, quantitatively compares these methods to traditional upscaling approaches, and evaluates their potential in applications including online education, telemedicine, and humanitarian technology. Further, we propose a novel hybrid lightweight SR model, optimized for real-time mobile and edge deployment, and discuss the computational and economic trade-offs of these advancements.

Background & Related Work

Early SR methods were dominated by interpolation (nearest, bilinear, bicubic) and example-based learning. With the advent of deep learning, SRCNN paved the way for end-to-end learning in SR tasks. GAN-based models (SRGAN, ESRGAN) improved perceptual results, while Transformer-based models brought advances in modeling global context. Diffusion models, recently, have achieved competitive results in challenging degradation scenarios.

HDR data is typically represented with floating point intensities exceeding 1, stored in a file format like OpenEXR. As neural networks perform best with input data normalized between $[-1, 1]$ or $[0, 1]$, we apply the following range compression function to accommodate HDR data [Bako et al. 2019]:Ty = κloд(1+μy)/loд(1+μ). We set $\kappa = 0.6$ and $\mu = 5000$, providing range up to luminance values near 300. We then convert our range-compressed dataset into a high-performance data structure, the lightning memory-mapped database, accelerating training speeds by about 1/3 over reading EXR images directly(Vavilala & Meyer).

SRGAN and ESRGAN: Generative adversarial networks (GANs) like SRGAN and ESRGAN are widely adopted for image and video SR. ESRGAN improves upon SRGAN by introducing a relativistic adversarial loss and residual-in-residual dense blocks, yielding better perceptual quality and fidelity

Transformer-Based SR: Models like SRTransGAN and SwinIR leverage global attention to capture long-range dependencies, showing substantial SR performance.

Diffusion Model: Recent surveys highlight the competitive performance of diffusion models for SR, particularly in handling complex degradation and producing diverse, high-quality outputs

Methods

Traditional upscaling methods (e.g., bicubic interpolation) yield inferior perceptual and quantitative metrics (PSNR, SSIM) compared to deep learning models. For example, a recent medical imaging study showed that SRGAN yielded a ~6% higher SSIM and up to 11.5dB higher PSNR than bicubic interpolation. SRGAN/ESRGAN often provide 2–4dB gains in PSNR and 0.05–0.12 improvements in SSIM over bicubic or TVI scaling in benchmarks set14 and DIV2k. LPIPS also favor GAN-based methods by up to 30% improvements in some imaging applications

Comparative Baseline & Proposed Hybrid Lightweight SR Model: We compare classical bicubic interpolation with SRGAN, ESRGAN, and a recent lightweight Transformer-based SR model. All models are trained and evaluated on standard datasets (DIV2K, Set14) and a domain-specific humanitarian dataset. Our model combines efficient residual CNN blocks, lightweight Transformer layers, and a simplified GAN-based loss. Quantization and operator fusion are applied for deployment on mobile NPUs and CPUs, targeting devices common in resource-limited environments

Domains and Applications

i. Video SR enables lower bitrate video conferencing and streaming without loss of end-user quality—essential in 3G/4G/5G networks in emerging regions
ii. Deep SR is a backbone for microscopy, medical diagnostics, satellite, and consumer content streaming.
iii. SR models are deployed in remote sensing, disaster response, and telemedicine platforms to compensate for degraded connectivity or sensor limitations
iv. Visual platforms for e-learning and students in poor-Wi-Fi zones benefit substantially by upscaling compressed, low-res educational content

Computational Trade offs

CPU's vs GPUs Deep SR models (SRGAN/ESRGAN, Transformers) are highly optimized for parallel GPU computation. Training is GPU-intensive; inference can be ported to CPU or specialized NPUs for real-time, mobile, or edge deployment.

Edge Devices: Lightweight SR variants and quantized models (e.g., SplitSR, HADT, Mobile-SR) are developed to deploy on ARM CPUs, NPUs, and even microcontrollers, achieving FHD inference in <50ms on-device.

Experiment

Dataset	Domain	Description	Coverage/Gaps
DIV2K	Images	800 high-res (2K) images for SR	Generic, lacks domain-specific degradations
Set5/Set14	Images	Classic SR benchmarks, small size	Images mostly generic
YouTube-8M	Videos	8M videos, 500k hours, diverse topics	Not focused on humanitarian/medical
Humanitarian RSOD/VHR-10	Remote sensing	Used for object detection post-SR	Still lacking standardized low-res, humanitarian-focused SR datasets

Method	PSNR↑	SSIM↑	LPIPS↓	Inference (ms, mobile CPU)
Bicubic	23.3	0.60	0.55	<10
SRGAN	25.8	0.74	0.32	140
ESRGAN	26.9	0.81	0.24	180
TransformerSR	27.2	0.84	0.19	220
Proposed	27.0	0.82	0.20	45

On mobile NPUs, the proposed hybrid model achieves real-time performance (<50ms/frame) with only a marginal loss in perceptual quality compared to larger models. Deploying SR client-side allows content providers to reduce CDN and bandwidth expenditure by streaming at lower bitrates and upscaling locally. Previous industry reports have shown bandwidth savings of 40–50% with no perceptible user loss of quality. The lack of tailored datasets for humanitarian and medical use cases remains a barrier. Further, generalization to diverse degradations (e.g., noise, motion blur) requires ongoing research.

Conclusion

Deep learning-based SR methods can close the visual quality gap in low-bandwidth settings while maintaining computational practicality. Future work will focus on expanding domain-specific datasets, integrating federated learning for device-side adaptation, and exploring multi-modal SR (image, video, text overlays).