Super-Resolution using GANs for Medical Imaging

Course Project (CS 736)

By: Ank Kumar Gupta (22B0623) & Veenus (22B0704)

Outline

- Introduction
- Background
- Problem Statement
- Motivation
- Mathematical Formulation
- Model Architecture
- Improved SE Block
- Loss & Optimization
- Training Strategy
- Results
- Conclusion
- References

Introduction

- We have implemented and fine-tuned a GAN based super resolution pipeline (Embedding an improved SE block and fusion loss) to upsample retinal images by 4x, 8x and 16x, achieving good PSNR/SSIM values.
- Motivation:
    - High-resolution detail (microvasculature, soft exudates) is critical for early disease detection in retinal and other medical scans
    - Simple interpolations (bicubic) introduce blurring and lose sharp edges
- Goal: Reviewing and Reproducing the results of “Medical Image SR using Improved GAN” (Bing et. al., IEEE Access 2019).
- Dataset: STARE (397 HR Retinal Images: 377 train/ 20 test)

Background

- Classical SR Approaches (Bicubic): Fast, but smooths out fine details.
- CNN-based SR:
  - SRCNN: 3 conv layers, trains end to end with MSE loss
  - VDSR: 20 layer residual CNN, adds global skip-connection
  - EDSR: Removes batch-norm, uses 64-channel residual blocks
- GAN-based SR:
  - SRGAN: Adds adversarial loss for perceptual realism
  - ESRGAN: Residual-in-residual dense blocks, improved texture fidelity
  - P-GAN: Progressive GAN specifically for medical images
- Squeeze-and-Excitation (SE) Blocks:
  - Recalibrates channel wise features via a learned sigmoid gate over global context
  - Limitation: Original (0, 1) scaling can attenuate features excessively when deeply stacked.

Problem Statement

- Let $I_{HR} \in R^{H \times W \times 3}$ be a ground truth high-resolution (HR) retinal image. We obtain the low-resolution (LR) input by bicubic downsampling:

$$I_{LR} = D_S(I_{HR}) \in R^{H/s \times W/s \times 3}, \quad s \in \{4, 8, 16\}.$$

  Where H & W are height and width of the HR image, s is the scaling factor and D_S represents downsampling operator.

  Our task is to learn a generator G such that

$$I_{SR} = G(I_{LR}) \in R^{H \times W \times 3}, \quad \text{such that} \quad I_{SR} \approx I_{HR}$$

- Baseline: Bicubic upsampling of I_{LR} back to $R^{H \times W \times 3}$.
- Evaluation Metrics:
  - PSNR (dB): Measures pixel-wise fidelity
  - SSIM(0-1): Quantifies perceptual/structural similarity
  - Visual Inspection to ensure qualitative results

Motivation for the Approach

- Approach: Embedding residual - scaled SE blocks into a simplified EDSR-GAN with fusion loss for sharper, detail-preserving medical super-resolution.
- Alone GAN+Pixel wise loss ensures high PSNR but reproduce over-smoothed output, while adversarial GAN loss sharpen textures but also introduce hallucinated artifacts.
- Original SE recalibrates each channel x_c by:

$$y_c = 1/HW \sum_{ij} x_c(i, j),\ s^{org} = \sigma(W_2\delta(W_1 y)) \in (0, 1),\quad x_c bar = s_c^{org} x_c.$$

    - Limitation: $s_c^{org} < 1$ always shrinks feature magnitudes leading to cumulative suppression when many SE blocks are stacked.
- Improved SE introduces a residual connection and allow scale up to 2, which preserves important features rather than attenuating them.
- Simplified EDSR generator strips out batch norm, add global residuals and embeds improved SE blocks to boost trainability and prominent vital anatomical features.
- Fusion loss balances pixel fidelity, perceptual similarity, and adversarial realism thus, preventing both over-smoothing and spurious artifacts.

Mathematical Formulation

- Improved SE:
  - A residual Connection is introduced and scale up to 2 is allowed.
    $S^{imp} = [k_1\ \sigma(W_2\delta(W_1 y))+k_2\sigma(y)] \times 2$, $k_1 = 0.8$, $k_2 = 0.2$, $x_c bar = s_c^{org} x_c$, $s_c^{imp} \in (0, 2)$
    Where $[k_1\ \sigma(W_2\delta(W_1 y))+k_2\sigma(y)]$ is original SE.
  - Benefit: Preserves or even amplifies important features rather than only attenuating them, improving gradient flow in deep SR networks.
- Fusion Loss:
  - $L_{Fusion} = L_{VGG} + w_1 L_1 + w_{RG} L_{RG} + w_{MSE} L_{MSE}$
    Where, L_{VGG} = Perceptual Loss (VGG-based) for semantic content
    L_{RG} = Relativistic GAN loss for realistic textures
    L_1, L_{MSE} = Pixel losses for Pixel level fidelity (PSNR)
    w_i = Scaler Weights (Ensures not a single term dominates training)

Model Architecture

- Generator: Simplified EDSR+Improved SE
  - Conv → Improved SE → 16 ResBlocks (Conv-ReLU-Conv-SE)
  - Global residual skip connects input to deep features
  - $\log_2(s)$ × [PixelShuffle upsampling + Improved SE]
  - Final Conv outputs the super-resolved image
- Discriminator: Feature Fusion + Improved SE
  - Conv + LeakyReLU → 8 Conv-SE blocks with downsampling
  - Last 3 feature maps fused via spatial alignment
  - Global pooling → 1×1 Convs → Sigmoid output

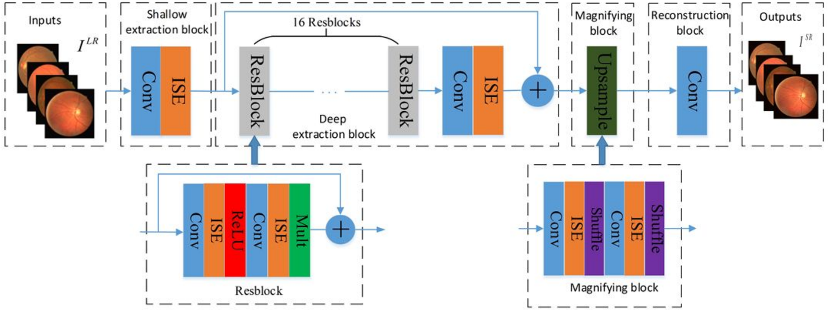


Fig 1: Generator for SE Model (Adopted from [1])

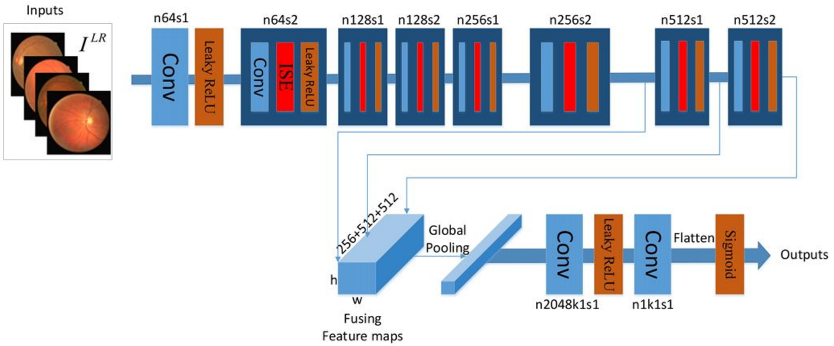


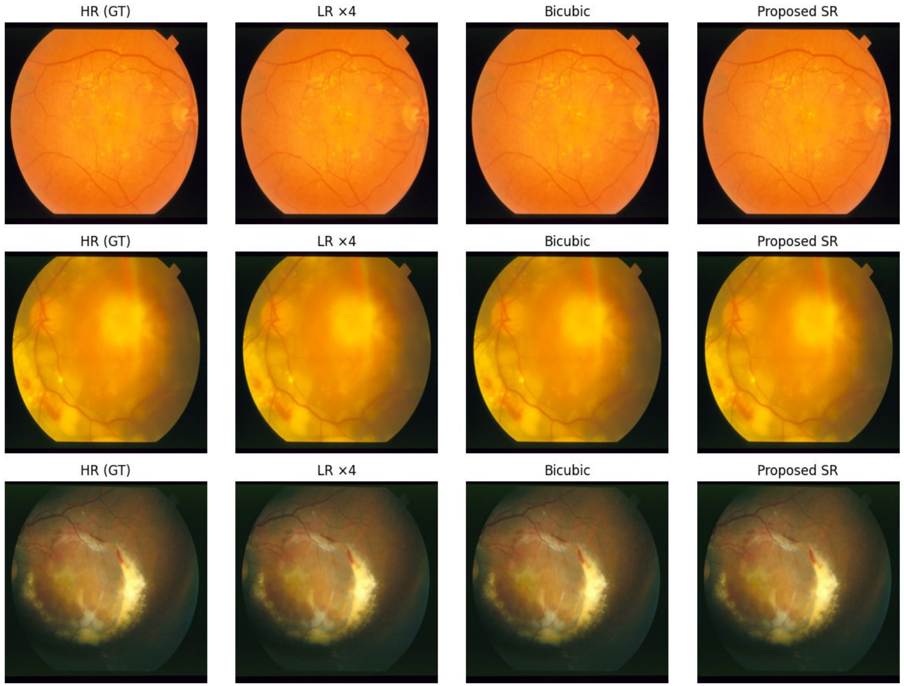
Fig 2: Discriminator for SE Model (Adopted from [1])

Loss & Optimization

- Loss weights (In Fusion Loss):
  w_1 = 1, w_{RG} = 0.005, w_{MSE} = 0.5
- Optimizers: Adam (lr = 1e-4, betas=(0.9,0.999), eps=1e-8)
- Training schedule:
  - Pretrain baselines only for inference
  - Train Proposed model: 100 epochs, batch 16 on 1024×1024 patches
  - Augment: random flip & rotations
- Hardware constraint: fits in 15 GB GPU (Colab T4 GPU)

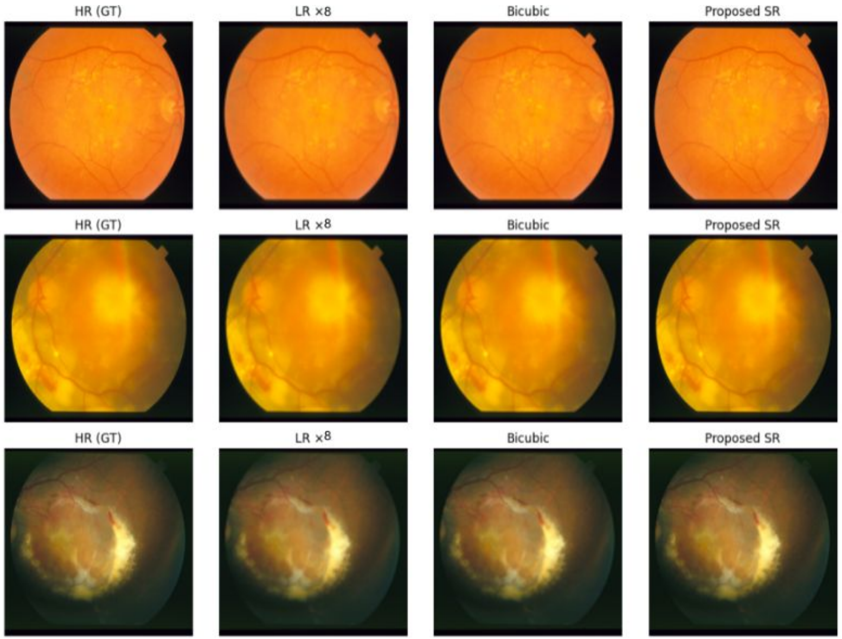
Results

- 4x Upscaling:
  - Bicubic:
    - PSNR: 40.03 dB
    - SSIM: 0.9512
  - Proposed:
    - PSNR: 42.65 dB
    - SSIM: 0.9567



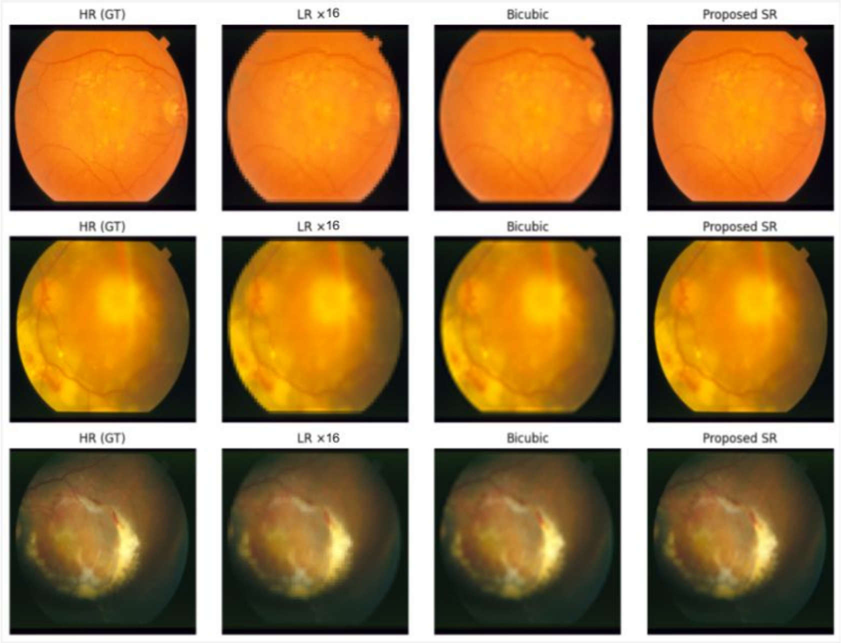
Results

- 8x Upscaling:
  - Bicubic:
    - PSNR: 35.41 dB
    - SSIM: 0.9195
  - Proposed:
    - PSNR: 39.88 dB
    - SSIM: 0.9330



Results

- 16x Upscaling:
  - Bicubic:
    - PSNR: 31.60 dB
    - SSIM: 0.8915
  - Proposed:
    - PSNR: 36.62 dB
    - SSIM: 0.9181



Conclusion

- Key Comparisons:
  - 4×: PSNR Increased by 2.6 dB (42.65 vs. 40.03), SSIM Increased by 0.9567
  - 8×: PSNR Increased by 4.5 dB (39.88 vs. 35.41), SSIM Increased by 0.9330
  - 16×: PSNR Increased by 5.0 dB (36.62 vs. 31.60), SSIM Increased by 0.9181
- Sharper edges are produced using proposed method even at high scaling factors
- Therefore, it can be concluded that Improved SE blocks preserve essential channels in deep networks, and Fusion loss balances fidelity, realism, and perceptual quality.
- Further work can be in direction of integrating more data, and even it can be extended to 3D medical SR for CT/MRI applications.

Reference

[1] Xinyang Bing, Wenwu Zhang, Liying Zheng and Yanbo Zhang, “Medical Image Super-Resolution Using Improved Generative Adversarial Networks,” IEEE Access, 10.1109/ACCESS.2019.2944862, Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8854062

[2] Structured Analysis of the Retina. Accessed: 2000. [Online]. Available: http://cecas.clemson.edu/~ahoover/stare/

Thank You!