

IMAGE SUPER-RESOLUTION AND RECONSTRUCTION BY GENERATIVE NEURAL NETWORKS

Roman Rapco | Supervisor: Ing. Miron Kuzma, PhD.

Motivation and Problem

Image super-resolution reconstructs a high-resolution image from a low-resolution input. It is an ill-posed inverse problem because image degradation removes information and fine details.

Classical interpolation methods are fast, but they only estimate new pixel values from local ones. At larger scale factors, this often produces blur and loss of fine textures.

Generative adversarial networks can synthesize more convincing details, introducing a trade-off between pixel fidelity and perceptual quality.

Proposed Method

The main practical contribution is **ConvNeXt-SRGAN**. Residual blocks in the original SRGAN generator were replaced by adapted ConvNeXt blocks, while the discriminator, loss formulation and overall adversarial training principle were intentionally kept unchanged.

- **7 × 7 depthwise convolution** for a larger receptive field
- **LayerNorm** instead of BatchNorm in the generator
- **1 × 1 pointwise convolutions**, GELU, layer scale and residual skip connection

Evaluation and Results

All models were evaluated on 100 **DIV2K validation images** at a **4× scale factor**. The reference SRGAN achieved the highest **PSNR** and **SSIM**, while the main **ConvNeXt-SRGAN** variant provides strong perceptual quality and substantially lower generator-core complexity.

Method	PSNR	SSIM	LPIPS ↓
Bicubic	25.52	0.7584	0.4249
SRGAN	26.48	0.7705	0.1687
ConvNeXt-SRGAN	24.29	0.6686	0.1741

PERCEPTUAL QUALITY - LPIPS 0.1741

SRGAN: 0.1687 — the two GAN-based methods are perceptually comparable

GENERATOR CORE - 2.878 GMACs

SRGAN: 10.872 GMACs • approximately 0.26×

BLOCK PARAMETERS - 19,968

SRGAN residual block: 74,112 • approximately 0.27×

Interpretation: lower PSNR and SSIM do not necessarily imply lower perceived quality in perceptually oriented GAN methods. The proposed model shifts the balance toward perceptual fidelity and computational efficiency.



Conclusion

Main contribution of the thesis is **replacing SRGAN residual blocks with adapted ConvNeXt blocks** demonstrating a useful perception-distortion trade-off.

ConvNeXt-SRGAN model uses about 0.26x generator-core operations and 0.27x block parameters of SRGAN. Perceptual quality is comparable to original SRGAN, while some scenes show sharper, more realistic fine textures.

