Mathematical Foundations of TinyESPCN-Enhanced

Bidyendu Das, Rishi Kumar Saawarn, Subhranshu Sarkar

November 12, 2025

Abstract

This paper presents the formal mathematical foundation of the **TinyESPCN-Enhanced** model, a compact and efficient convolutional neural network designed for single-image super-resolution. The work rigorously derives the theoretical underpinnings of the architecture, loss formulation, optimization process, and gradient dynamics that justify its design choices. The analysis demonstrates how sub-pixel convolution, residual connections, and channel attention collectively improve performance and convergence properties under a constrained parameter budget of approximately 2.8×10^5 parameters.

1 Introduction

Super-resolution tasks seek a function $f_\theta : \mathbb{R}^{H\times W\times 3} \to \mathbb{R}^{(sH)\times(sW)\times 3}$ that reconstructs a high-resolution (HR) image from a low-resolution (LR) input. Classical interpolation methods fail to approximate high-frequency components, whereas deep networks approximate f_θ as a non-linear operator learned through empirical risk minimization.

TinyESPCN-Enhanced optimizes this process through a compact architecture featuring residual convolutional blocks and channel attention. Despite its small parameter footprint (≈ 0.28M), it achieves high perceptual quality by optimizing a multi-term loss function that balances perceptual fidelity, edge sharpness, and color accuracy.

2 Mathematical Formulation of the Model

Given an input image $I_{LR} \in \mathbb{R}^{H\times W\times 3}$, the network aims to reconstruct $I_{SR} = f_\theta(I_{LR})$ such that $I_{SR} \approx I_{HR}$. The TinyESPCN-Enhanced model defines f_θ as a composition of functions:

$$I_{SR} = \Phi_{\text{global}}(I_{LR}) = \text{Clamp}(\text{PixelShuffle}(W_2 * g(W_1 * I_{LR})) + \text{Up}_{bicubic}(I_{LR}), 0, 1), \quad (1)$$

where $g(\cdot)$ is the residual trunk composed of $N = 10$ convolutional-ReLU units and $*$ denotes convolution.

Each convolutional layer can be represented as an affine transformation in a finite-dimensional function space:

$$\mathcal{C}_k(x) = \sigma(W_k * x + b_k), \quad W_k \in \mathbb{R}^{C_{out}\times C_{in}\times K\times K}, \quad (2)$$

where $\sigma(\cdot)$ denotes the ReLU activation. For the input $x_1 = \mathcal{C}_1(I_{LR})$ and trunk output $x_2 = g(x_1)$, the residual connection yields:

$$x_f = x_1 + x_2. \tag{3}$$

The final sub-pixel upscaling step employs the operator Π_s defined as:

$$\Pi_s : \mathbb{R}^{3s^2 \times H \times W} \to \mathbb{R}^{3 \times sH \times sW}, \tag{4}$$

which rearranges feature maps into spatial positions, ensuring efficient low-dimensional processing.

3 Channel Attention Mechanism

Channel attention introduces an adaptive weighting mechanism for feature channels. Formally, given feature tensor $x \in \mathbb{R}^{B \times C \times H \times W}$, the attention weights are computed as:

$$\alpha = \sigma(W_2 \cdot \mathrm{ReLU}(W_1 \cdot \mathrm{GAP}(x))), \tag{5}$$

where GAP denotes global average pooling. The recalibrated feature tensor is then:

$$\tilde{x} = x \odot \alpha, \tag{6}$$

with $\odot$ denoting element-wise multiplication.

This formulation ensures that the learned weights α optimize the mutual information between channels, prioritizing those contributing to high-frequency content.

4 Loss Function Derivation

The total objective function combines perceptual, edge, and color-space consistency terms:

$$\mathcal{L}_{total} = \mathcal{L}_{VGG} + \lambda_e \mathcal{L}_{edge} + \lambda_c \mathcal{L}_{Lab}, \tag{7}$$

with weights $\lambda_e = 0.2$, $\lambda_c = 0.1$. Each term is defined as follows.

4.1 Perceptual Loss

Using a feature extractor ϕ_j corresponding to VGG19 layers, the perceptual loss is defined by:

$$\mathcal{L}_{VGG} = \sum_{j \in J} \|\phi_j(I_{SR}) - \phi_j(I_{HR})\|_1. \tag{8}$$

This loss minimizes the L_1 distance between feature representations in a pre-trained feature space, approximating perceptual similarity through non-linear feature correlation.

4.2 Edge Loss

The edge loss compares gradient and Laplacian maps of SR and HR images. Let S_x, S_y, and L be Sobel and Laplacian operators, respectively. Then,

$$\mathcal{L}_{edge} = \|S_x * I_{SR} - S_x * I_{HR}\|_1 + \|S_y * I_{SR} - S_y * I_{HR}\|_1 + \|L * I_{SR} - L * I_{HR}\|_1. \tag{9}$$

This formulation ensures local gradient alignment, preserving fine structures.

4.3 Color Loss

For color preservation, both SR and HR images are transformed to the Lab color space via a differentiable mapping T_{Lab}. The color loss is then:

$$\mathcal{L}_{Lab} = \|T_{Lab}(I_{SR}) - T_{Lab}(I_{HR})\|_1. \tag{10}$$

This term penalizes chromatic deviations that are perceptually salient in human vision.

5 Optimization and Gradient Flow

The model parameters θ are optimized by minimizing $\mathcal{L}_{total}$ via gradient descent:

$$\theta^{(t+1)} = \theta^{(t)} - \eta \nabla_\theta \mathcal{L}_{total}(f_\theta(I_{LR}), I_{HR}), \tag{11}$$

where η is the learning rate. The presence of residual and global skip connections ensures that the Jacobian $J = \frac{\partial f_\theta}{\partial I_{LR}}$ maintains stable eigenvalues, avoiding gradient explosion or vanishing effects.

The overall optimization objective can be viewed as minimizing a functional in the Sobolev space $W^{1,1}(\Omega)$, where the energy functional $E(f_\theta)$ satisfies:

$$E(f_\theta) = \int_\Omega \|\nabla(f_\theta(I_{LR}) - I_{HR})\|_1 \, dx + \sum_j \|\phi_j(f_\theta(I_{LR})) - \phi_j(I_{HR})\|_1. \tag{12}$$

6 Parameter Efficiency Analysis

The TinyESPCN-Enhanced model has approximately 2.8×10^5 parameters, primarily distributed as:

- Initial 7×7 convolution: $3 \times 64 \times 7 \times 7 = 9{,}408$ parameters.
- Ten residual blocks: each $64 \times 64 \times 3 \times 3 = 36{,}864$, totaling 368,640 parameters.
- Channel attention layers (two FC layers with reduction ratio 16): $64 \times 4 + 4 \times 64 = 512$ parameters.
- Final convolution: $64 \times (3s^2) \times 3 \times 3$ for $s = 2$, giving $64 \times 12 \times 9 = 6{,}912$ parameters.

After accounting for shared weights and batch normalization, the total trainable parameters are approximately 280,000 (0.28M). This compactness provides excellent inference efficiency while maintaining strong representational capacity.

7 Discussion

The proposed mathematical structure demonstrates that TinyESPCN-Enhanced achieves a balance between model expressiveness and efficiency. Sub-pixel rearrangement confines most operations to low-resolution space, significantly reducing computational complexity from $O((sH)(sW))$ to $O(HW)$. The combined residual and attention structures yield improved gradient propagation and selective channel reinforcement, ensuring convergence under a limited parameter budget.

8 Conclusion

This paper has formalized the mathematical rationale behind TinyESPCN-Enhanced. By defining its components through explicit mappings, gradient flows, and loss-space interactions, we demonstrate that the architecture maintains theoretical stability while achieving high perceptual reconstruction quality. The model's efficiency stems from its constrained parameterization and the principled combination of residual learning, channel attention, and perceptual-edge-aware optimization.