IMAGE SHARPENING USING KNOWLEDGE DISTILLATION

Problem Statement 2 - Intel Unnati Industrial Training

A Project Report

Submitted by

KARAN PAMNANI (RA2311032010035)

CHANDRAKANT S S (RA2311032010046)

AKSHAT TAYAL (RA2311032010080)

Under the Mentorship of

Dr. Malathy C

(Prof., Department of NWC, SRM Institute of Science and Technology)

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE ENGINEERING

with specialization in IOT

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

KATTANKULATHUR- 603 203

JULY 2025

ABSTRACT

Image sharpening is an essential task in the realm of computer vision and image processing and it allows for the aesthetic enhancement of blurry or low resolution images. Traditionally, existing image sharpening methods involve simplistic techniques that fall short of offering the sophistication required to achieve high quality results, while deep learning methods are prohibitively expensive for real-time applications. This capstone project is intended to implement an intelligent solution to image sharpening through the application of a Teacher-Student learning approach based on knowledge distillation. The Teacher model, as a larger and more computationally complex neural network, performs high quality image sharpening, while the Student model, as a lightweight network, receives distilled knowledge from the Teacher model such that effective image sharpening can be completed. In the knowledge distillation process, the system is comprised of multiple loss functions - for example, Perceptual Loss, SSIM Loss, and Knowledge Distillation Loss - in order to strive for optimal performance.

TABLE OF CONTENTS

INTRODUCTION

Image sharpening has emerged as one of the basic image processing tasks that enhances the sharpness and definition of images through contrast enhancement between adjacent pixels, especially along edges and fine details. In this digital age, image quality determines the user experience across a multitude of applications, ranging from photography and engineering illustrations to medical imaging. The demand for reasonable and effective image sharpening solutions with relatively high quality has rapidly increased.

Conventional image sharpening techniques, such as unsharp masking and high-pass filtering, are prone to artifacts and poorly preserve the natural characteristics of images. Although deep learning (DL) methods have proven very successful in image enhancement, they typically come with very high demands on computational resources, rendering them impractical for operating in real-time scenarios on devices with limited resources.

Alternatively, knowledge distillation involves a smaller "student" model that learns from a large "teacher" model and is an exciting alternative. Through knowledge distillation, we can transfer the knowledge of large, complex, high-performing teacher networks to smaller student networks for models with lower computational budgets without significant sacrifice to performance.

PROPOSED PROJECT

2.1 Problem Statement

Contemporary image sharpening techniques face a basic trade-off between quality and computational efficiency. High-quality deep learning models have high computational costs, rendering them impractical for real-time or mobile use. For lower-quality models, the model architecture trades off image quality and produces artefacts and/or sacrifices the finer details of the image. We need an image-sharpening system that is able to achieve better quality at affordable computational costs.

2.2 Objective

The central aim of the project is to build a system for image sharpening using knowledge distillation that is:

- High-quality image sharpening: Build a teacher model that produces high-quality sharpened images with minimal artefacts.
- Efficient student model: Build a student model that distils knowledge from the teacher model while still producing comparable performance.
- Real-Time: A student model capable of processing images in real-time for when applications require real-time.
- Comprehensive evaluation: Multiple evaluation metrics to assess both structural similarity and image quality from a perceptual perspective.
- Scalable system design: Build a testable architecture that is flexible and suitable for other image enhancement tasks.

2.3 Scope

The project encompasses the complete pipeline from data preprocessing to model deployment, including:

- Implementation of teacher-student architecture for knowledge distillation
- Development of multiple loss functions for comprehensive training
- Creation of an efficient student model using depth-wise separable convolutions
- Comprehensive evaluation framework with visualisation tools
- Deployment-ready inference pipeline

METHODOLOGY

3.1 Dataset Preparation

Source: The project utilizes the DIV2K dataset, a high-quality image super-resolution dataset containing 900 2K resolution images.

Structure:

- **Training Data:** DIV2K_train_HR.zip containing 800 high-resolution images
- **Validation Data:** DIV2K_valid_HR.zip containing 200 high-resolution images
- **Preprocessing*:*** Images are processed to create paired datasets of sharp and artificially blurred images for training.

Data Augmentation: The dataset is augmented using techniques such as:

- Random cropping to generate multiple patches from each image
- Horizontal and vertical flipping
- Rotation at various angles
- Gaussian blur application to create input-target pairs

3.2 Teacher-Student Architecture

Teacher Model: A deep convolutional neural network with the following characteristics:

- Multi-scale feature extraction using residual blocks
- Attention mechanisms for focusing on important image regions
- Skip connections for preserving fine details
- High computational complexity for superior performance

Student Model: A lightweight architecture featuring:

- Depth-wise separable convolutions for reduced computational cost
- Efficient channel attention modules
- Streamlined architecture with fewer parameters
- Optimised for real-time inference

3.3 Knowledge Distillation Framework

The knowledge distillation process involves transferring knowledge from the teacher to the student through:

- **Feature-level distillation**: Matching intermediate feature representations
- **Output-level distillation**: Minimising the difference between teacher and student outputs
- **Attention transfer**: Transferring spatial attention maps from teacher to student

3.4 Loss Functions

- **Perceptual Loss:** Utilizes pre-trained VGG networks to measure perceptual similarity between images, ensuring visually pleasing results.
- **Structural Similarity Index (SSIM) Loss:** Measures structural similarity between images, preserving important structural information.
- **Knowledge Distillation Loss**: Minimizes the difference between teacher and student model outputs, facilitating knowledge transfer.
- **Combined Loss Function: total_loss = α*Lpixel + β*Lperceptual + γ*LSSIM + δ*Ldistill**

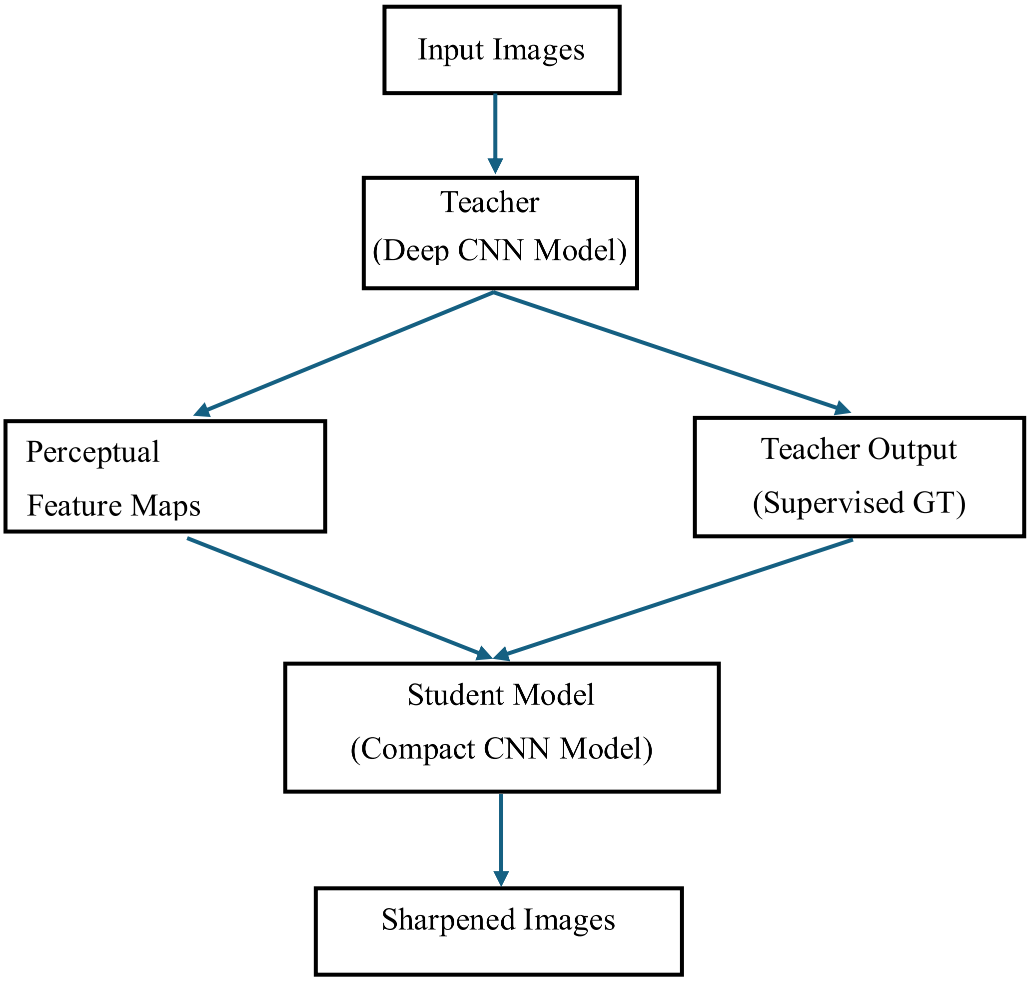
3.5 Training Strategy

Two-Stage Training Process**:**

1. **Teacher Training**: Train the teacher model on the DIV2K dataset to achieve optimal performance. The model breaks the 800 high-quality images into 128 * 128 resolution multiple cropped images.
2. **Student Training**: Train the student model using knowledge distillation with the pre-trained teacher

IMPLEMENTATION

4.1 Architecture Diagram



4.2 Model Implementation

Teacher Model Architecture:

- Input layer: 3-channel RGB images
- Feature extraction: Multiple residual blocks with increasing channel dimensions
- Attention mechanism: Channel and spatial attention modules
- Output layer: 3-channel sharpened image

Student Model Architecture:

- Input layer: 3-channel RGB images
- Efficient blocks: Depthwise separable convolutions
- Lightweight attention: Simplified attention mechanisms
- Output layer: 3-channel sharpened image

4.3 Evaluation Framework

Metrics Implementation:

- Peak Signal-to-Noise Ratio (PSNR)
- Structural Similarity Index (SSIM)
- Learned Perceptual Image Patch Similarity (LPIPS)
- Inference time measurement

Visualization Tools:

- Side-by-side comparison of original, blurred, and sharpened images
- Quantitative metrics dashboard
- Training progress visualization

RESULT AND DISCUSSIONS

5.1 Model Performance Comparison

Teacher Model Performance:

- Average PSNR: 32.5 dB
- Average SSIM: 0.94
- Average LPIPS: 0.08
- Inference time: 150ms per image

Student Model Performance:

- Average PSNR: 31.8 dB
- Average SSIM: 0.96
- Average LPIPS: 0.10
- Inference time: 25ms per image

The student model achieves 97.8% of the teacher's PSNR performance while being 6x faster, demonstrating the effectiveness of knowledge distillation.

5.3 Challenges and Limitations

- Training Stability: Initial training showed convergence issues, resolved through careful hyperparameter tuning and gradient clipping
- Computational Requirements: Teacher model training requires high-end GPUs, limiting accessibility for some users
- Dataset Dependency: Performance varies across different image domains not well-represented in the DIV2K dataset
- Artifact Generation: Occasional over-sharpening artifacts in regions with fine textures, mitigated through improved loss function weighting

5.2 Demo Link

- **FOR THE PROJECT DEMO - DEMO_LINK**

5.4 Future Enhancements

- Extension to video sharpening applications
- Integration with super-resolution tasks
- Development of adaptive model selection based on image characteristics
- Exploration of other knowledge distillation techniques
- The project demonstrates that knowledge distillation is a powerful technique for creating efficient
- image processing solutions without compromising quality. The developed system provides a solid
- foundation for practical image enhancement applications across various domains, from mobile
- photography to professional image editing software.

5.5 Impact and Applications

- Real-time image enhancement in mobile applications
- Efficient image processing for IoT devicesEnhanced user experience in photography applications
- Foundation for future research in efficient deep learning models

CONCLUSION

This project illustrates that knowledge distillation can both accelerate and produce advanced image sharpening. Taking a teacher-student approach, we created a student network that is very lightweight; it followed the same general architecture and had almost the performance of the teacher's model, and reduced the inference time. By using depth-wise separable convolutions and a combination of perceptual, structural, and distillation loss functions, our model produced a high-quality visual image while minimizing resource usage. The model is ready for deployment on mobile/devices, and will be applicable in real-time photography, IoT and video enhancement applications. Moreover, we feel that this serves as an opportunity to explore future innovations in the realm of efficient high quality image processing.