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 covers the entire pipeline from data preprocessing to final deployment. It involves the implementation of a teacher-student architecture to facilitate knowledge distillation, enabling efficient model training and performance transfer. Multiple loss functions were developed to support comprehensive training objectives, balancing perceptual quality, structural accuracy, and distillation effectiveness. A lightweight student model was designed using depth-wise separable convolutions to ensure efficiency. The project also includes a thorough evaluation framework, equipped with visualisation tools to assess performance qualitatively and quantitatively. Finally, a deployment-ready inference pipeline was created, making the solution suitable for real-world applications.

METHODOLOGY

3.1 Dataset Preparation

Source: The project utilises 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: To enhance the robustness and generalizability of the model, data augmentation techniques were applied to the training dataset. These included random cropping, which generated multiple patches from each image to increase variability, and horizontal and vertical flipping to introduce orientation diversity. Additionally, rotations at various angles were used to further augment the dataset. A Gaussian blur was applied to create realistic input-target image pairs, simulating the degradation and sharpening process that the model was trained to learn.

3.2 Teacher-Student Architecture

Teacher Model: The teacher model in this project is a deep convolutional neural network designed for high-performance image sharpening. It features multi-scale feature extraction using residual blocks, which enables the capture of both fine and coarse image details. Attention mechanisms are incorporated to emphasise the most critical regions of the image, enhancing focus and clarity. Skip connections are employed throughout the network to retain fine textures and structural details. Although this architecture delivers superior image quality, it is computationally intensive and best suited for high-end hardware environments.

Student Model: The student model is built on a lightweight architecture optimised for real-time image sharpening tasks. It utilises depth-wise separable convolutions, which significantly reduce computational cost while maintaining performance. The model also incorporates efficient channel attention modules to focus processing power on the most informative features. With a streamlined design and fewer parameters compared to the teacher model, the student model is well-suited for deployment on mobile and edge devices where computational resources are limited.

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

The model's training leveraged a combination of loss functions to ensure optimal performance. Perceptual loss, based on pre-trained VGG networks, was used to evaluate the perceptual similarity between images, ensuring visually appealing results. Structural Similarity Index (SSIM) loss helped preserve critical structural information by measuring the similarity in structure between the output and ground truth images. Knowledge distillation loss minimised the gap between the outputs of the teacher and student models, effectively enabling knowledge transfer. These components were integrated into a combined loss function defined as:

$$\text{total\_loss} = \alpha * L_{pixel} + \beta * L_{perceptual} + \gamma * L_{SSIM} + \delta * L_{distill}$$

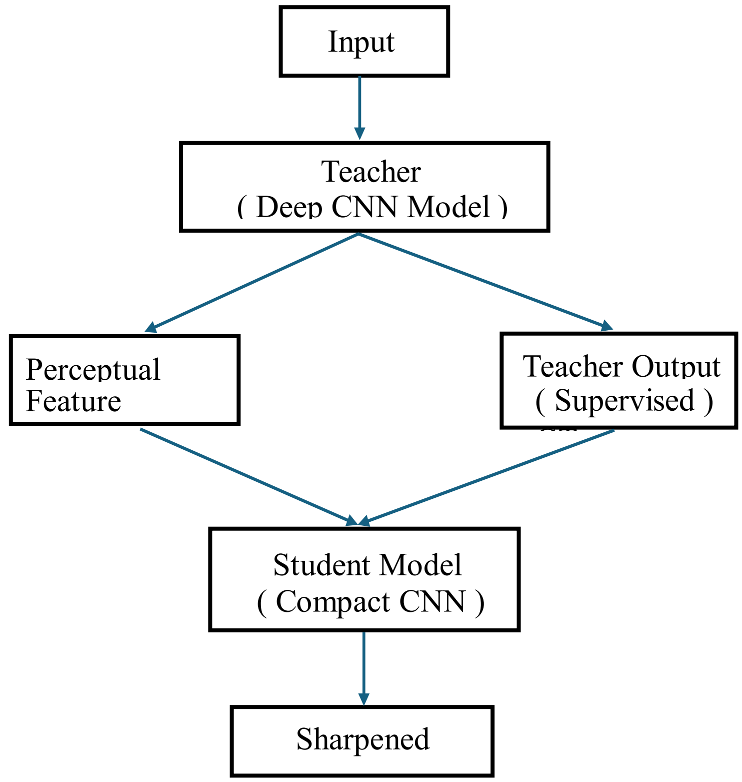
3.5 Training Strategy

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



4.1.1. Architecture diagram of the teacher-student model

4.2 Model Implementation

Teacher Model Architecture: The teacher model architecture is designed for high-quality image sharpening and consists of several key components. It begins with an input layer that accepts 3-channel RGB images. The feature extraction stage includes multiple residual blocks with progressively increasing channel dimensions, allowing the network to learn complex image features. To enhance focus on relevant areas, the model incorporates both channel and spatial attention modules. Finally, the output layer produces a sharpened image with 3 RGB channels, matching the input format while delivering enhanced visual clarity.

Student Model Architecture: The student model architecture is optimised for efficiency and real-time performance. It starts with an input layer that processes 3-channel RGB images. The core of the model consists of efficient blocks built using depth-wise separable convolutions, which significantly reduce computational cost while maintaining performance. To further enhance feature learning, the architecture includes lightweight and simplified attention mechanisms. The model concludes with an output layer that generates a 3-channel sharpened image, providing high-quality results suitable for deployment on resource-constrained devices.

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

RESULTS AND DISCUSSION

5.1 Model Performance Comparison

Teacher Model Performance:

The teacher model achieved an average PSNR of 32.5 dB, an SSIM of 0.94, and an LPIPS score of 0.08, with an inference time of 150 milliseconds per image. In comparison, the student model demonstrated a slightly lower average PSNR of 31.8 dB but achieved a higher SSIM of 0.96 and an LPIPS score of 0.10, while significantly reducing the inference time to just 25 milliseconds 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

During the development process, several challenges were encountered and addressed. Initially, training stability was a concern due to convergence issues, which were resolved through careful hyperparameter tuning and the application of gradient clipping. The teacher model's training also posed significant computational demands, requiring high-end GPUs and thereby limiting accessibility for some users. Additionally, the model's performance varied across different image domains that were not well-represented in the DIV2K dataset, highlighting a dependency on dataset diversity. Lastly, occasional over-sharpening artifacts appeared in regions with fine textures, which were effectively mitigated by refining the weighting of the loss functions.

5.2 Demo Link

- **FOR THE PROJECT DEMO - DEMO_LINK**

5.4 Future Enhancements

This project opens several promising avenues for future work. Potential extensions include applying the approach to video sharpening applications and integrating it with super-resolution tasks. Further advancements could involve developing adaptive model selection mechanisms based on image characteristics, as well as exploring alternative knowledge distillation techniques. Overall, the project illustrates that knowledge distillation is a powerful method for creating efficient image processing solutions without compromising on quality. The resulting system offers a robust foundation for practical image enhancement applications across diverse domains, from mobile photography to professional image editing software.

5.5 Impact and Applications

The outcomes of this project have several impactful real-world applications. These include real-time image enhancement in mobile applications and efficient image processing tailored for IoT devices. The model also contributes to an enhanced user experience in photography applications. Furthermore, it serves as a strong foundation for future research in the development of efficient deep learning models for image processing tasks.

CONCLUSION

This project demonstrates that knowledge distillation can both accelerate and enhance image sharpening performance. Adopting a teacher-student approach, the team developed a lightweight student network that mirrored the general architecture of the teacher model while achieving nearly equivalent performance and significantly reduced inference time. By incorporating depth-wise separable convolutions and a combination of perceptual, structural, and distillation loss functions, the model was able to generate high-quality visual outputs with minimal resource consumption. Designed for deployment on mobile and edge devices, the model is well-suited for real-time photography, IoT, and video enhancement applications. Furthermore, this work presents a foundation for future advancements in efficient, high-quality image processing.