Project Report

of

INTEL UNNATI

INDUSTRIAL TRAINING - 2025

Image Sharpening using Knowledge Distillation

SUBMITTED
BY

Sakshi Shivaprasad

(Reg.No: 230962006)

School of Computer Engineering,

Manipal Institute of Technology, Manipal

UNDER THE GUIDANCE OF

Mr. Abhishek Nandy
Industry Mentor
Intel Unnati
Bangalore

Dr. Roopashri Shetty
Faculty Mentor
School of Computer Engineering,
Manipal Institute of Technology,
Manipal

CHAPTER 1: INTRODUCTION

This project addresses the challenge of degraded video quality during real-time communication, particularly over low-bandwidth connections. With the increasing reliance on video conferencing platforms for professional and personal communication, the quality of visual content transmitted becomes a key determinant of user experience. However, many users face scenarios where limited bandwidth leads to blurry, low-quality video frames, reducing both clarity and usability.

To address this, the proposed project focuses on image sharpening using a deep learning-based knowledge distillation approach. Knowledge distillation enables a compact and efficient model (student) to learn from a larger, more accurate model (teacher), making it suitable for deployment in resource-constrained environments. By leveraging this paradigm, the project aims to develop a lightweight image sharpening model that enhances degraded visual frames in real-time video conferencing scenarios, without requiring high computational resources.

This project showcases the use of knowledge distillation to train a lightweight image sharpening model guided by a high-capacity teacher. The resulting student model is efficient and successfully enhances image sharpness, making it a practical solution for real-time deployment in video conferencing scenarios.

CHAPTER 2: PROBLEM STATEMENT

Video conferencing plays a vital role in communication across education, healthcare, and business sectors. However, poor network conditions often lead to blurred video frames, affecting user experience in detail-critical scenarios like reading text or recognizing facial expressions.

The core problem lies in restoring or enhancing image sharpness from these blurred frames in a way that is both perceptually effective and computationally efficient. Traditional sharpening techniques either fail to generalize or are too slow for real-time deployment. Therefore, there is a need for a fast and lightweight deep learning-based model that can improve the quality of blurry images using prior knowledge.

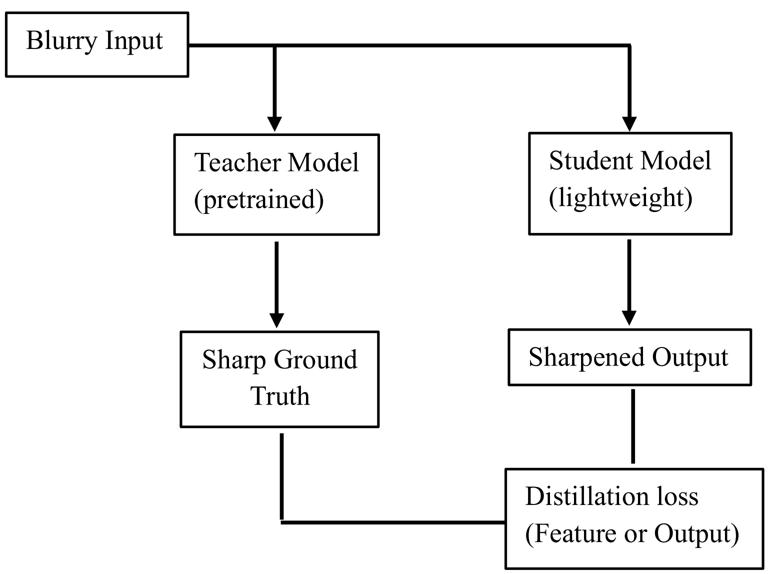
This project proposes a solution through a teacher-student learning framework, where a high-performance model transfers its learned knowledge to a compact student model. This enables real-time image enhancement on edge devices or low-end systems, providing a practical and scalable solution to image clarity degradation during video calls.

CHAPTER 3: ARCHITECTURE

This project uses a knowledge distillation approach for image sharpening. A pre-trained Restormer acts as the teacher model, providing high-quality sharpened outputs. The student model, a lightweight UNet, is trained to produce similar results by learning from the teacher's outputs and refining visual details from the input. This makes the model both efficient and suitable for improving image clarity in video conferencing.

Components:

- Input: Blurry images
- Teacher: Restormer model (pretrained)
- Student: Lightweight Unet variant
- Losses: L1 Loss, Perceptual Loss, SSIM Loss, Charbonnier Loss
- Output: Sharpened Image



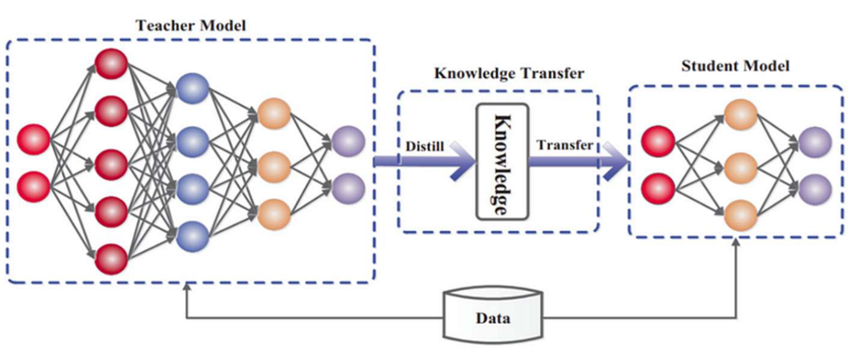


Fig 1: Teacher-Student Model Architecture

CHAPTER 3: METHODOLOGY

1. Dataset and Preprocessing

- **Dataset Used**: GoPro Deblurring Dataset
- **Image Size**: 1,280×720 resolution (HD)
- **Total Images**: 3,214 (2,103 training, 1,111 test)
- **Input Format**: Blurry-Sharp image pairs
- **Training Samples**: 100x100 cropped patches from full-resolution images
- **Testing Samples**: Test dataset was used as provided for evaluating model performance
- **Preprocessing**:
  - Training images were cropped into 100x100 patches.
  - Pixel values were normalized to [0,1] using PyTorch tensor conversion.

2. Teacher and Student Model

- **Teacher Model**: Restormer
  - Role: Generates high-quality sharpened outputs used as the target reference for the student model during training
- **Student Model**: Improved lightweight Unet (base=32)
  - Features: Compact architecture enabling fast inference and low resource usage

3. Loss Functions

A composite loss function was implemented to guide the model's learning process:

- **L1 Loss**: Measures pixel-wise difference between the student output and the sharp image.
- **Perceptual Loss**: Extracted using a pre-trained VGG network to compare high-level features between student output and ground truth.
- **SSIM Loss**: Structural similarity index used to ensure perceptual closeness to the target sharp image.
- **Charbonnier Loss**: A robust variation of L1 loss, improves stability and sharpness by penalizing small errors effectively.

4. Code Implementation

- **Data Loader**: A custom PyTorch dataset class was built to load triplet data—blurred inputs, ground truth sharp images, and teacher outputs.
- **Training Loop**: The training loop iterates over batches of triplets, computing a composite loss and updating model weights using the Adam optimizer. SSIM scores are logged per epoch for tracking.
- **Model Checkpointing**: Periodic saving of model weights ensures training can be resumed and the best-performing model (based on SSIM) is retained.
- **Learning Rate Scheduling**: A ReduceLROnPlateau scheduler adjusts learning rate when SSIM plateaus, aiding convergence.

CHAPTER 5: RESULTS AND EVALUATION

1. Performance Analysis

- **Final SSIM Achieved**: The student model attained a Structural Similarity Index of **0.8673**, indicating effective sharpening performance closely aligned with the ground truth sharp images.

2. Visual Results

- Improved definition of fine details
- Smoother textures and reduced visual distortions compared to blurred inputs

Fig. 2: Comparison of Blur Input, Student Model Output, and Ground Truth Sharp Image

CHAPTER 6: CONCLUSION

This project demonstrates a compact and effective image sharpening system trained using knowledge distillation. By using the Restormer as the teacher and a streamlined UNet-based model as the student, the approach balances image quality and model efficiency. The student model achieved an SSIM of **0.8673**, showing notable improvements in visual clarity while remaining lightweight.

The training pipeline incorporates efficient patch-based learning, robust validation, and structured supervision from the teacher model. These choices enable fast convergence and good generalization. The results confirm the potential of lightweight CNNs trained with knowledge distillation to improve image quality under limited computational constraints.

Future Enhancements

While the current solution delivers promising results, a couple of focused improvements could further enhance its effectiveness:

- **Temporal Consistency**: Adapting the model for video input can help maintain consistency across frames and prevent flickering effects.
- **Cross-Domain Adaptability**: Future work could explore training the model on varied domains such as low-light images, compressed video frames, or real-time webcam streams to enhance generalization and usability.

Overall, the teacher-student framework effectively delivers image sharpening by balancing visual quality and runtime efficiency, making it suitable for practical use in bandwidth-constrained environments.

CHAPTER 7: REFERENCES

- Research paper: https://arxiv.org/abs/2111.09881
- Teacher-Student Model: https://miro.medium.com/v2/resize:fit:1200/1*vlyCVw7QGelaF2MOcfRqlQ.png
- Dataset: GoPro Image Deblurring Dataset
- Pretrained Restormer: swz30/Restormer: [CVPR 2022--Oral] Restormer: Efficient Transformer for High-Resolution Image Restoration. SOTA for motion deblurring, image deraining, denoising (Gaussian/real data), and defocus deblurring.