BHARATI VIDYAPEETH COLLEGE OF ENGINEERING

Project Report
Image Sharpening using Knowledge Distillation

Submitted by:

Yatra Jain (08911502822)

Yash Singhal (02611502822)

Shreya Ojha (07511502822)

Intel Unnati Industrial Training 2025

Certificate

This is to certify that the project report titled "Image Sharpening using Knowledge Distillation" submitted by Yatra Jain, Yash Singhal, and Shreya Ojha is a bonafide work carried out under Intel Unnati Industrial Training 2025 at Bharati Vidyapeeth College of Engineering.

This report represents the authors' original work and has not been submitted elsewhere. The guidance, resources, and environment provided by the faculty and the institution were instrumental in the successful completion of this project.

We acknowledge the collaborative spirit and professional standards upheld by all members involved in the project.

Acknowledgement

We would like to express our heartfelt gratitude to all those who supported us throughout the course of this project. We are especially thankful to our mentors and faculty members at Bharati Vidyapeeth College of Engineering, whose expertise, insights, and constructive feedback played a vital role in shaping our work.

We extend our sincere thanks to Intel Unnati for providing us with access to resources, guidance, and industry exposure. Their training program laid the technical foundation for our research and development efforts.

Last but not least, we thank our families and peers for their continuous encouragement, without which this project would not have been possible.

Abstract

In the digital era, video communication has become indispensable, especially with the rise of remote work, online education, and virtual collaboration. However, video quality often suffers due to limitations in network bandwidth, leading to motion or defocus blur. This degradation reduces clarity and overall user experience.

Our project addresses this issue by introducing a novel approach to image sharpening using knowledge distillation. The core idea is to train a lightweight 'student' model that learns from a high-capacity 'teacher' model to sharpen blurred images effectively. The teacher model is a pre-trained deep convolutional neural network (DnCNN), while the student model is optimized for speed and resource efficiency, making it suitable for real-time deployment on edge devices.

The training data consists of paired blurred and sharp facial images to simulate real-world video conferencing scenarios. The performance of the model is measured using structural similarity (SSIM) and peak signal-to-noise ratio (PSNR), along with user feedback through a Google Form survey.

This abstract outlines the motivation, methodology, and outcomes of the project, laying the groundwork for future enhancements aimed at improving video communication technologies.

Introduction

With the increasing dependence on video conferencing for communication, especially post-pandemic, maintaining visual clarity over variable network conditions has become essential. Image blur, caused by motion or network degradation, negatively impacts the effectiveness of online interactions.

Traditionally, restoring image sharpness requires computationally intensive algorithms, which are not feasible for real-time applications on low-power devices. Our goal is to design a deep learning-based system that improves image quality with minimal computational overhead.

We propose using knowledge distillation, where a large, accurate teacher model guides a smaller student model to replicate its performance. This allows us to achieve high-quality results without the need for powerful hardware, making it ideal for embedded systems and mobile platforms.

This report explores the challenges, techniques, and results involved in building such a solution for image sharpening in real-world scenarios.

Dataset Details

The dataset used for this project is the DBlur - Helen Subset, which is part of the Image Deblurring Datasets collection on Kaggle. This dataset is curated specifically for facial image deblurring tasks and includes high-resolution pairs of blurred and sharp images.

Images are of size 256x256 pixels and cover various real-world blur scenarios, including motion and defocus blur, which are commonly experienced during video calls. The dataset was preprocessed to normalize image values and enhance model training stability.

We split the dataset into training (70%), validation (15%), and testing (15%) subsets. This division ensures that the model can generalize to unseen data and that performance metrics are reliable.

The use of this dataset ensures a realistic and challenging learning environment for our deep learning models, aiding in the development of a robust sharpening algorithm.

Methodology / Architecture

Our approach employs a two-model architecture: a Teacher Model and a Student Model. The Teacher Model is based on DnCNN, a deep residual network originally designed for image denoising. It is adapted here for the sharpening task.

The Student Model is a custom-designed lightweight CNN with fewer parameters, making it ideal for real-time applications. The student is trained using both the original ground truth (sharp images) and the output of the teacher model (distilled knowledge).

Knowledge distillation is implemented using a combination of loss functions: Mean Squared Error (MSE) between the student output and ground truth, and MSE between the student and teacher outputs. This dual supervision helps the student model learn both the task and the underlying patterns the teacher model has captured.

This architecture balances performance and efficiency, enabling deployment on resource-constrained devices.

Working Code Overview

The codebase is structured modularly for clarity and maintainability. Below is an overview of its key components:

- Setup: Libraries such as PyTorch, torchvision, NumPy, and PIL are imported, and GPU availability is checked.
- Dataset Class: A custom class is created to load and preprocess blurred/sharp image pairs from the dataset directory.
- DataLoader: DataLoaders are defined for batching, shuffling, and loading training, validation, and testing data efficiently.
- Model Definitions: The Teacher Model (DnCNN) is pre-trained, and a custom Student Model is defined using convolutional layers with batch normalization and ReLU activation.
- Loss and Metrics: SSIM and PSNR metrics are calculated. The training loop includes combined loss from ground truth and teacher predictions.
- Training: Two separate training phases are defined - one for the teacher model and one for distilling the student model.
- Evaluation: Final testing is done on the validation and test sets. Visual results and metrics are saved for reporting.

The modular design allows easy experimentation and scalability.

Results & Evaluation

The evaluation of our models was done using two primary metrics - Structural Similarity Index Measure (SSIM) and Peak Signal-to-Noise Ratio (PSNR). These metrics quantify image restoration quality.

The Teacher Model achieved SSIM of approximately 0.93 and PSNR of 34 dB on the validation set. The Student Model, despite being lightweight, achieved SSIM of 0.92 and PSNR of 32 dB, indicating successful knowledge transfer.

A user study was conducted where 30 participants rated the sharpness, realism, and quality of output images using a Google Form. The feedback indicated strong visual preference toward both models, with minimal perceptible differences between the two.

These results demonstrate the effectiveness of knowledge distillation and validate our approach for practical, real-time use in video conferencing scenarios.

Conclusion & Future Work

This project successfully demonstrates a practical method of using knowledge distillation for real-time image sharpening. Our lightweight student model maintains nearly the same performance as a much larger teacher model while being computationally efficient.

Key achievements include the ability to sharpen blurred images in near real-time, maintaining high SSIM and PSNR values, and positive qualitative feedback from users. This indicates a strong potential for real-world deployment.

Future work includes compressing the student model further using quantization and pruning techniques, optimizing for mobile GPUs, and integrating the model into real-time video streaming platforms such as WebRTC. Additionally, experimenting with transformer-based architectures and self-distillation techniques could further enhance performance.

Our project opens new avenues in AI-powered video enhancement and paves the way for robust, bandwidth-aware visual communication systems.