PROJECT DETAILS

IMAGE SHARPENING USING KNOWLEDGE DISTILLATION

TIMING Jun 5, 2025 to Jul 12, 2025

STATUS COMPLETED

OWNERS debangshikundu03@gmail.com

TABLE OF CONTENTS

PROJECT DETAILS
RESOURCES

➔ PROJECT OVERVIEW

This project tackles the challenge of restoring image clarity in low-quality images—such as those seen during video conferencing or low-bandwidth streaming—by leveraging deep learning. A high-capacity **Teacher Model** (ResNet-based) is trained to learn fine-grained features, while a lightweight **Student Model** is distilled from it to enable faster inference with comparable quality.

The goal is to enhance the **perceived sharpness and visual fidelity** of blurry or low-resolution input images while maintaining real-time performance capabilities.

➔ OBJECTIVES

- Improve the sharpness of low-resolution images.
- Leverage **knowledge distillation** to compress a high-performance model into a smaller one.
- Evaluate image quality using SSIM and PSNR.
- Enable a lightweight model to operate in constrained environments (e.g., edge devices).

➔ Problem Statement:

With the rise of video conferencing, live streams, and mobile photography, image quality often suffers due to compression and limited bandwidth, resulting in blurry visuals. High-end super-resolution models are computationally expensive and unsuitable for deployment on low-power devices.

This project proposes a **two-stage KD-based solution** where:

- A heavy, accurate Teacher model is trained to perform image sharpening.
- A compact Student model learns to mimic the teacher’s output using distilled knowledge.

➔ Key Features:

➔ ResNet-based **Teacher model** for high-fidelity sharpening.
➔ Lightweight **Student model** with reduced parameters.
➔ Data preprocessing pipeline using **DIV2K** dataset (HR-LR patch generation).
➔ Evaluation using perceptual quality metrics: **SSIM** and **PSNR**.
➔ Inference-ready code with visualization and batch prediction.

➔ Literature Review:

➔ Briefly mention previous works in:
  - ◆ Super-resolution (e.g., SRCNN, ESRGAN)
  - ◆ Knowledge distillation in computer vision

➔ Compare your approach's simplicity and deployment-friendliness.

➔ Methodology:

1. Data Preparation

- Source: DIV2K dataset
- 500 training and 50 validation images
- Random 128×128 HR crops → Bicubic downsampled to LR → Resized to 128×128
- Output: paired LR-HR patches

2. Model Architecture:

Teacher Model (ResNet18-based)

- Pretrained ResNet18 as encoder
- Custom decoder with upsampling layers
- Learns to enhance LR images to HR quality

Student Model (3-layer CNN)

- Minimal architecture: 3 conv layers with ReLU
- Learns from teacher's output, not ground-truth directly
- Ideal for deployment on resource-constrained devices

3. Training Procedure

➔ **Teacher:** Trained on (LR → HR) with MSE loss

- **Student:** Trained on (LR → Teacher Output) using MSE loss (Knowledge Distillation)
- Optimizer: Adam
- Epochs: 10 for each model

➔ Evaluation Metrics

Metric	Description
SSIM	Structural Similarity Index (perceptual quality)
PSNR	Peak Signal-to-Noise Ratio (pixel accuracy)

→ Visual comparisons of:

- **Input LR image**
- **Teacher output**
- **Student output**
- **Ground Truth HR**

Results Summary:

Dataset	Avg SSIM	Avg PSNR (dB)
Validation	~0.9154	~22.86
Test	~0.8962	~23.14

➔ Requirements

Python Libraries

- `torch`, `torchvision`
- `numpy`, `pillow`
- `matplotlib`
- `scikit-image`
- `tqdm`

➔ Potential Applications:

➔ Video call sharpening filters
➔ Smartphone camera enhancement
➔ Image upscaling in medical or surveillance footage
➔ Compression artifact removal in low-bandwidth systems

➔ Limitations and Future Work:

- ◆ Currently trained only on 128×128 patches
- ◆ Student performance drops on unseen datasets

➔ Future improvements:

- ◆ Use of feature distillation (intermediate layers)
- ◆ Use of newer teacher models like Swin Transformers
- ◆ Integration into video streams

➔ LINKS AND FILES

References:

- DIV2K Dataset
- ResNet Paper
- TorchVision Models
- SSIM and PSNR - scikit-image
- Knowledge Distillation - Hinton et al.