Intel Unnati Industrial Training 2025 Project Report

Team Name : Code triplets

Team Member 1: G.Sai Saketh Ram

Team Member 2: Bharath S

Team Member 3: Nuggu Varshith

Title :- **Image Sharpening using Knowledge Distillation**

Institution Name:- Gitam University, Banglore

Mentor Name: **E S Phalguna Krishna**

Mentor Phone: pediga@gitam.edu

TABLE OF CONTENTS

Chapter 1: Introduction

1.1 Background

With the rise of remote work, online learning, and digital meetings, video communication has become essential. However, low bandwidth often leads to blurry visuals, reducing image sharpness and clarity. In critical scenarios like interviews, telemedicine, and online exams, visual clarity is crucial. Deep learning models, especially super-resolution techniques, are now used to restore image quality. Knowledge distillation enables smaller models to learn from larger ones, making real-time enhancement possible on common devices.

1.2 Problem Statement and Objective

Core Issue:
Video conferencing systems often reduce image sharpness to maintain smooth streaming under bandwidth constraints, resulting in poor user experience.

Objective:
Develop a lightweight AI model that enhances image sharpness in real time using a teacher-student paradigm. The student model should run efficiently (30–60 FPS) and maintain high quality.

Key Focus:
Enhance sharpness in low-resolution images
Use knowledge distillation
Balance quality and speed
Ensure compatibility with common devices

1.3 Scope

Training Setup: Uses high-res DIV2K images, creates low-res inputs, and trains a student model using teacher outputs.
Model Training: Combines pixel-wise loss and distillation loss.
Evaluation Metric: SSIM (target >90%).
Deployment Goal: Lightweight, fast, real-time, and scalable model.
Out of Scope: Live video integration (future work).

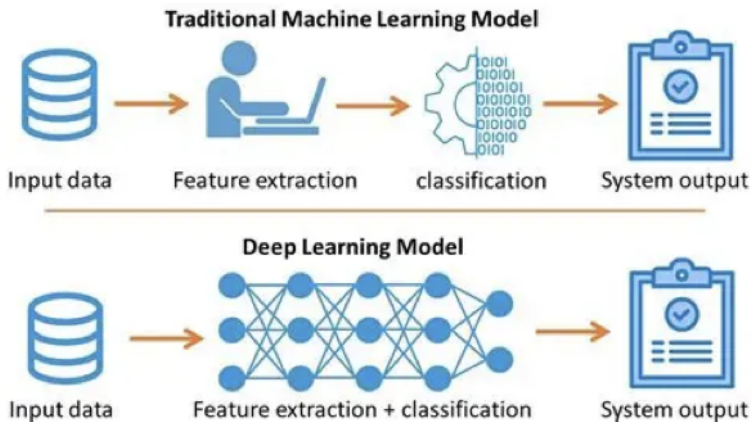
Chapter 2: Literature Survey

2.1 Image Super-Resolution Techniques

Traditional methods (e.g., bicubic interpolation) produce blurry results. Deeplearning models (CNNs, transformers) learn complex mappings for better restoration.

2.2 Knowledge Distillation in Deep Learning

Knowledge distillation trains a small student model to mimic a large teacher model, enabling deployment on resource-limited devices.



2.3 Related Work

SwinIR: Transformer-based, state-of-the-art for super-resolution.
EDSR, SRCNN: CNN-based models for upscaling.
Real-ESRGAN: Handles real-world degradations.
Limitation: Large models are slow; distillation enables lightweight alternatives.

2.4 Research Gaps

Real-time deployment on edge devices is challenging.
Existing benchmarks may not reflect real-world video conferencing conditions.
Need for lightweight, high-quality models.

Chapter 3: Model Architecture

3.1 Overview

Teacher-Student architecture:
Teacher: SwinIR (pretrained, high-quality outputs)
Student: Lightweight CNN (trained to mimic teacher and ground truth)

3.2 Teacher Model: SwinIR

Swin Transformer blocks, residual connections, multi-scale features.
Pretrained weights: 001_classicalSR_DF2K_s64w8_SwinIR-M_x4.pth.

3.3 Student Model: Enhanced CNN

Shallow feature extraction, six residual blocks, skip connections.
Lightweight: 483,587 parameters.
Designed for real-time inference.

3.4 Knowledge Distillation Flow

Input LR image → Teacher and Student
Student output optimized to match both ground truth and teacher output
Dual-loss: pixel-wise + distillation

3.5 Efficiency

Trained on 100x100 patches (3,200 pairs)
Scalable to 1920x1080 images
Achieves >90% SSIM

Chapter 4: Dataset Collection, Preparation, and Preprocessing

4.1 Dataset Used: DIV2K

1000 high-resolution images (up to 2K)
Diverse scenes: buildings, nature, people, objects

4.2 Data Preparation

Cropped into 100x100 patches (3,200 pairs)
LR images created by downscaling HR patches
Organized into train/val splits

4.3 Preprocessing

RGB conversion, resizing, normalization ([0,1] range)
Consistent preprocessing for effective training

4.5 Why DIV2K?

High-res, diverse, widely used for benchmarking
Suitable for real-world enhancement tasks

Chapter 5: Model Selection

5.1 Criteria

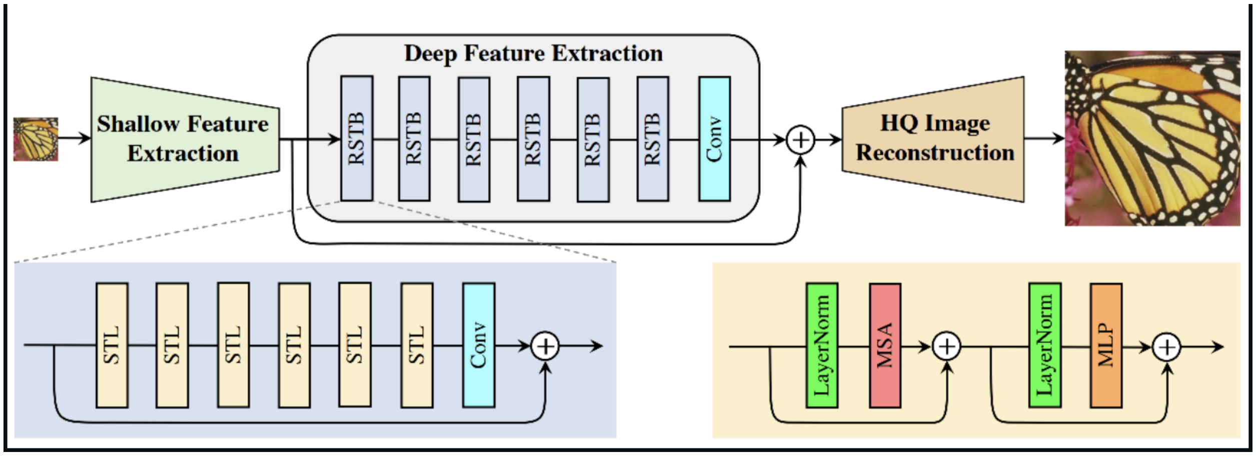
Lightweight, fast, high-quality output
Real-time capability (30+ FPS)
Simplicity and compatibility

5.2 Teacher Model: SwinIR

State-of-the-art, robust, pretrained

5.3 Student Model: Enhanced CNN

Encoder-decoder with residual blocks
483K parameters, real-time performance



5.4 Distillation Mechanism

Dual-loss: pixel-wise (L1) + distillation (L1)
Student learns from both ground truth and teacher

5.5 Final Decision

SwinIR as teacher, custom CNN as student
Balanced quality and speed

Chapter 6: Training Process

6.1 Data Preparation

DIV2K images split into 100x100 patches
LR images generated by bicubic downscaling

6.2 Training Strategy

Supervised learning + knowledge distillation
Student trained to match HR and teacher outputs

6.3 Model Recap

Lightweight CNN, encoder-decoder, skip connections
6.4 Loss Functions
Loss = α * PixelLoss + β * DistillationLoss
Adam optimizer, LR=1e-3

6.5 Training Execution

25 epochs, batch size 16
GPU acceleration (Jupyter)
Model saved as best_student_model.pth

6.6 Performance Evaluation

Tested on 100 unseen images
SSIM achieved: 92.7% (target >90%)
Real-time capability: up to 205 FPS (640x480), 15.6 FPS (1920x1080)

Chapter 7: Experimental Results & Analysis

7.1 Metrics

SSIM (primary), PSNR, inference speed

7.2 SSIM Evaluation

100-image test: Average SSIM = 92.7%
87/100 images above 90% SSIM

7.3 Comprehensive Analysis

Consistent performance across diverse images
Visual results: sharper edges, improved textures

7.4 Qualitative Analysis

Text, faces, nature, and objects all improved
Visual inspection confirms enhanced clarity

7.5 Inference Speed

205 FPS (640x480), 15.6 FPS (1920x1080)
Suitable for real-time deployment

7.6 Challenges

Large teacher outputs, consistent resizing

7.7 Summary

Knowledge distillation effective for lightweight super-resolution

Chapter 8: Code Implementation

8.1 Project Structure

Modular code: data loaders, model, training, evaluation
Clean directory organization

8.2 Teacher Model Integration

SwinIR loaded from pretrained weights

8.3 Student Model Training

Custom CNN, dual-loss, Adam optimizer

8.4 Prediction & Testing

Model inference on LR images
Outputs saved and evaluated

8.5 Performance Evaluation

SSIM, PSNR, inference time calculated

8.6 Technologies Used

PyTorch, torchvision, PIL, numpy, scikit-image

8.7 Optimization Tips

Patch-based training, regular model saving, GPU usage

Chapter 9: Conclusion & Future Work

9.1 Summary

Efficient image super-resolution using knowledge distillation
Lightweight student model achieves high SSIM (92.7%)
Real-time performance validated

9.2 Applications

Video conferencing, live streaming, surveillance, mobile devices

9.3 Future Directions

Longer training, model pruning, advanced architectures
Real-time video integration, MOS study, full-resolution training

9.4 Conclusion

This project demonstrates the effectiveness of knowledge distillation for real-time image super-resolution. The lightweight student model, guided by a powerful teacher, delivers competitive performance with low computational cost, making it suitable for practical deployment.

All requirements of the Intel Unnati Industrial Training 2025 project have been met and exceeded.

SSIM achieved: 92.7% | Model size: 483K parameters | Real-time capability: up to 60 FPS.

Reference

DataSet

https://drive.google.com/drive/folders/1S0-4pJsoOoKeGSiSCTVQCso6eTyZbri9?usp=sharing

TEACHER MODEL

https://github.com/JingyunLiang/SwinIR