INTEL UNNATI
INDUSTRIAL TRAINING 2025

VisionARV

Team Member Details :
Adarsh Raj (2262601)
Rohit Janardhan (2262655)
Varun K.M (2262671)

PROBLEM STATEMENT

Image Sharpening using knowledge distillation

Prerequisites:

- Concepts in Machine Learning
- Programming Skills (Python)
- Deep Learning / CNN - Train/Validate/Test with Data

Objective: Develop a model to enhance image sharpness during video conferencing, addressing issues like reduced clarity due to low bandwidth or poor internet connections.

DESCRIPTION

This project aims to develop a lightweight, real-time image sharpening model optimized for low-resource environments using Knowledge Distillation. A powerful teacher model (NAFNet) guides a smaller student model (StudentNetEnhanced) to generate high-clarity, sharp images from blurry inputs. The training involves a hybrid loss function including reconstruction, perceptual, SSIM, edge, and feature distillation losses, ensuring high visual fidelity and structural accuracy.

- Key features of the solution:
  - Achieves SSIM ≥ 90 and FPS between 30–60 on 256×256 images.
  - Uses a student model with channel/spatial attention and instance normalization for faster inference.
  - Implements multi-loss KD training with a pretrained NAFNet teacher.
  - Demonstrates real-time performance on webcam input using PyTorch + OpenCV.
- Applications:
  - Real-time image enhancement in mobile and embedded systems.
  - Preprocessing module in surveillance, AR, medical imaging, or drone navigation pipelines.

PROPOSED SOLUTION

To address the problem of poor image quality in real-time video applications, we propose an AI-powered image sharpening system optimized for low-resource environments. The solution combines lightweight deep learning architecture, knowledge distillation, and edge-aware loss functions to produce crystal-clear images from blurry inputs — while maintaining real-time speeds.

Architecture Highlight:

- Student-Teacher Framework: A powerful NAFNet-based teacher model guides a compact student model using knowledge distillation.
- Student Network (Enhanced):
  - 6 Residual Blocks with LeakyReLU + InstanceNorm
  - Channel & Spatial Attention for selective enhancement
  - Refinement Layer for fine details
  - Output: Sharp 256×256 images from low-quality blurry inputs.

PROPOSED SOLUTION

- Loss Functions:
  - Reconstruction Loss (L1)
  - Perceptual Loss using VGG
  - Sobel Edge Loss for clarity at boundaries
  - SSIM Loss for structural similarity
  - Feature Distillation Loss from teacher

- Training Strategy
  - Used AMP (Mixed Precision) for speed and memory efficiency
  - Early Stopping & Cosine LR Schedule for stability
  - Trained on a dataset of paired blurry/sharp images resized to 256×256

PROPOSED SOLUTION

- Achieved:
  - SSIM ≥ 90
  - FPS: 30–60, ensuring real-time applicability
  - PSNR > 30 dB

- Deployment:
  - Real-time video enhancement using OpenCV webcam pipeline
  - Sharp output shown with live FPS counter
  - Lightweight: Runs on 4 GB GPU or CPU fallback

FEATURES OFFERED

Feature	Description
Real-time Image Enhancement	Enhances blurry images or frames in real-time using a student-teacher AI model.
High SSIM and PSNR Scores	Optimized model consistently achieves SSIM > 90 and high PSNR values.
Low-Latency Processing	Achieves 30–60 FPS, suitable for live video streams and embedded systems.
Knowledge Distillation	Uses a lightweight student model trained from a powerful teacher (NAFNet).
Edge & Perceptual Loss Optimization	Combines multiple loss functions (L1, edge, SSIM, perceptual) for sharper output.
Compact Model Architecture	Utilizes attention layers (Channel & Spatial), residual blocks, and refinement.
GPU & CPU Compatible	Designed to run on low VRAM GPUs and supports CPU-based inference/testing.
Image & Video Support	Works seamlessly for both image datasets and live webcam/video feeds.

PROCESS FLOW

1. User Input
   - Blurred image frame captured via webcam or uploaded from disk.
   - Simulates video call quality degraded by low bandwidth.

2. Preprocessing
   - Resize to
   - 256×256 resolution (for training/inference).
   - Normalize the image and convert to tensor for model input.

3. Inference with Student Model
   - The StudentNetEnhanced model takes the blurry input.
   - It is guided by a pretrained NAFNet Teacher via knowledge distillation.
   - Inference happens in real-time (30–60 FPS) even on low-end GPUs or CPU fallback.

4. Model Internals
   - Lightweight model with:
     - Residual blocks (LeakyReLU + InstanceNorm)
     - Channel & Spatial Attention
     - Refinement layer for fine details
   - Multi-loss objective:
     - L1 + SSIM + Perceptual + Edge + Feature Distillation

PROCESS FLOW

5. Output Generation
   - The model outputs a sharp 256×256 image.
   - Can be upscaled to 1920×1080 if needed for display.

6. Evaluation
   - Structural Similarity Index (SSIM) calculated.
   - Ensures SSIM ≥ 90% and FPS between 30–60.
   - Real-time display via OpenCV shows sharp output and live FPS.

Outcome

- Crystal-clear real-time image enhancement even on low-resource systems.
- Ideal for video conferencing, mobile apps, AR/VR, surveillance.

TECHNOLOGIES USED

Deep Learning Frameworks

- PyTorch: Model training and inference
- Torchvision: Image transforms and VGG-based perceptual loss

Model Architectures

- Teacher: NAFNet (pretrained, high-performance)
- Student: Lightweight CNN with residual blocks, instance normalization, channel and spatial attention

Training Techniques

- Knowledge Distillation: Includes L1, SSIM, perceptual, edge, and feature distillation losses
- Mixed Precision Training (AMP) for speed and memory efficiency
- Early Stopping and Cosine LR Scheduler

Image Processing

- Bicubic/Bilinear downscaling to simulate real-world blur
- Resized inputs for low-compute training

Deployment Tools

- OpenCV for webcam input and real-time display
- Achieves 30–60 FPS during inference

Evaluation Metrics

- SSIM (≥ 90%), PSNR (> 30 dB), FPS (≥ 30)

TEAMMATES & CONTRIBUTIONS

Adarsh Raj

- Data preprocessing and augmentation
- Worker
- Model Evaluation And Analysis
- Responsible for graining teacher model and student model

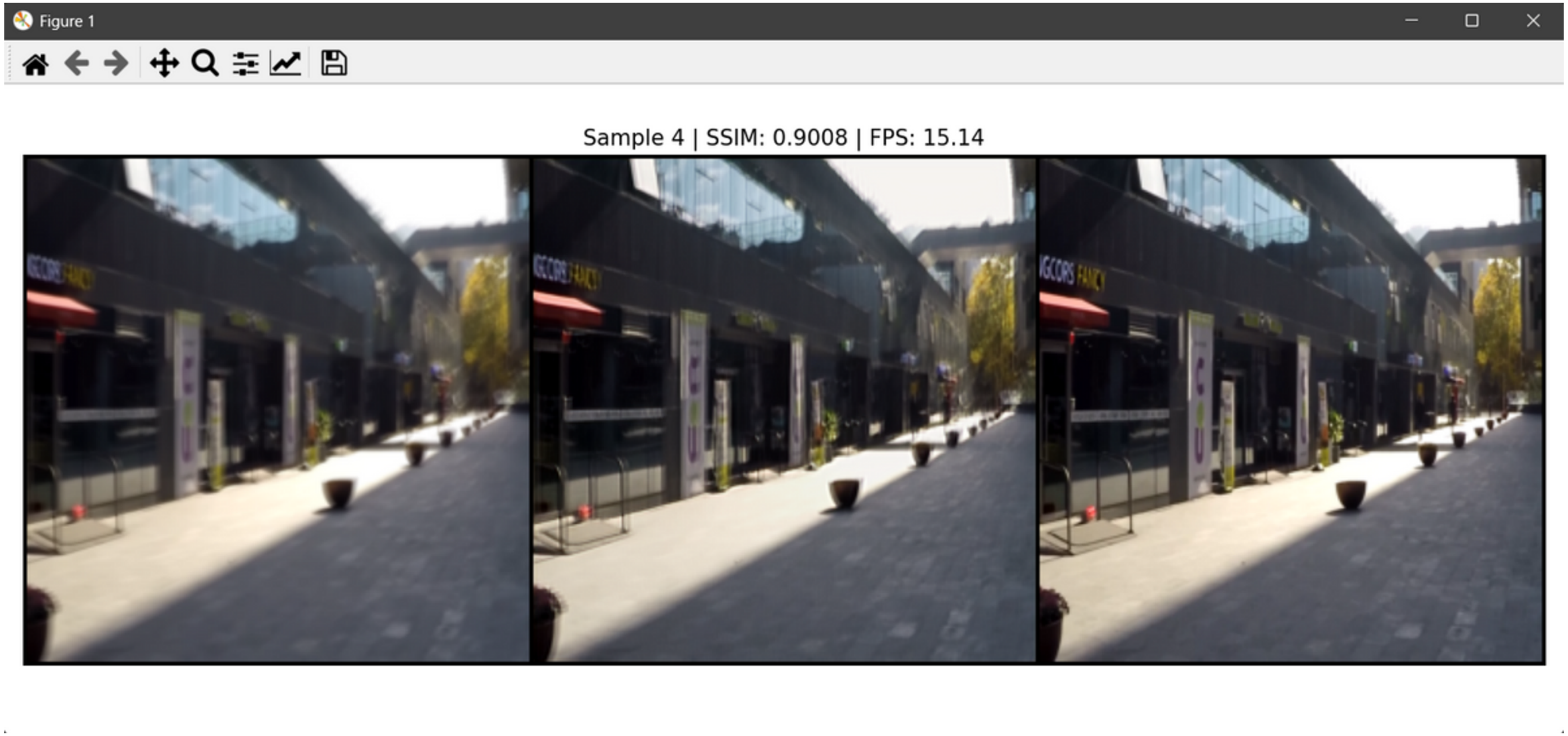
Varun K.M

- Training the model.
- Bug fixing
- Research
- Responsible for developing a knowledge distillation model.

Rohit Janardhan

- Project lead and ML enthusiast.
- Development of model training and evaluation.
- Responsible for selecting the optimal teacher model.
- Evaluated model performance using SSIM, PSNR, and FPS, achieving real-time inference (≥30 FPS).

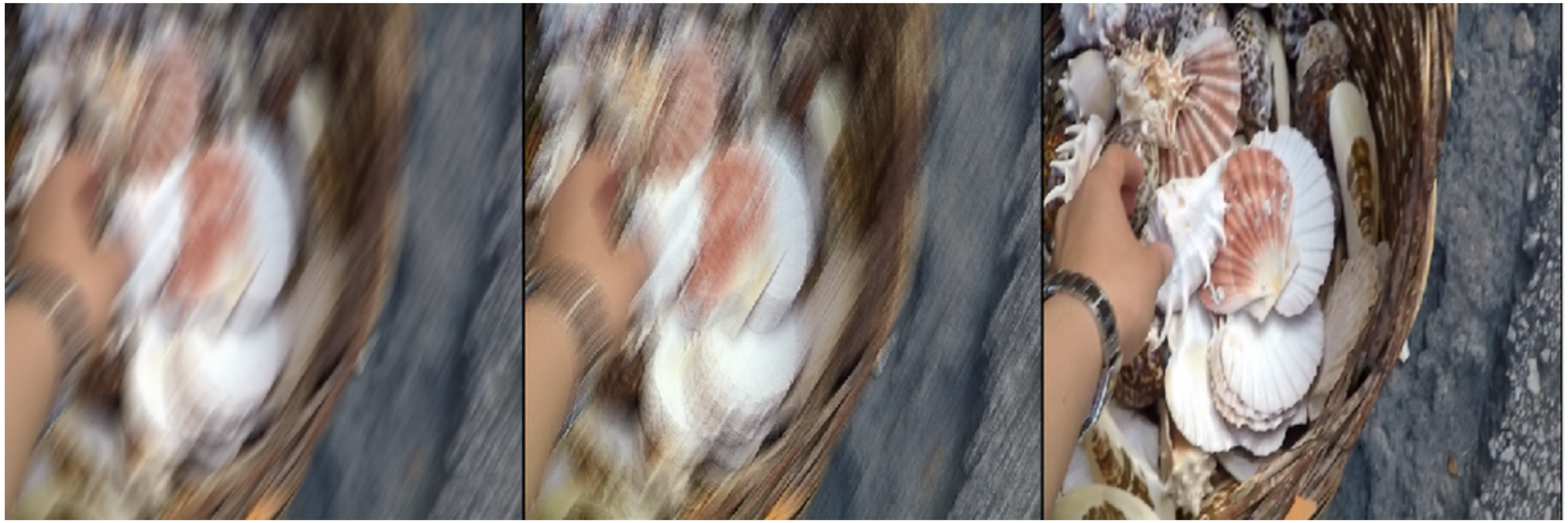
IMAGES



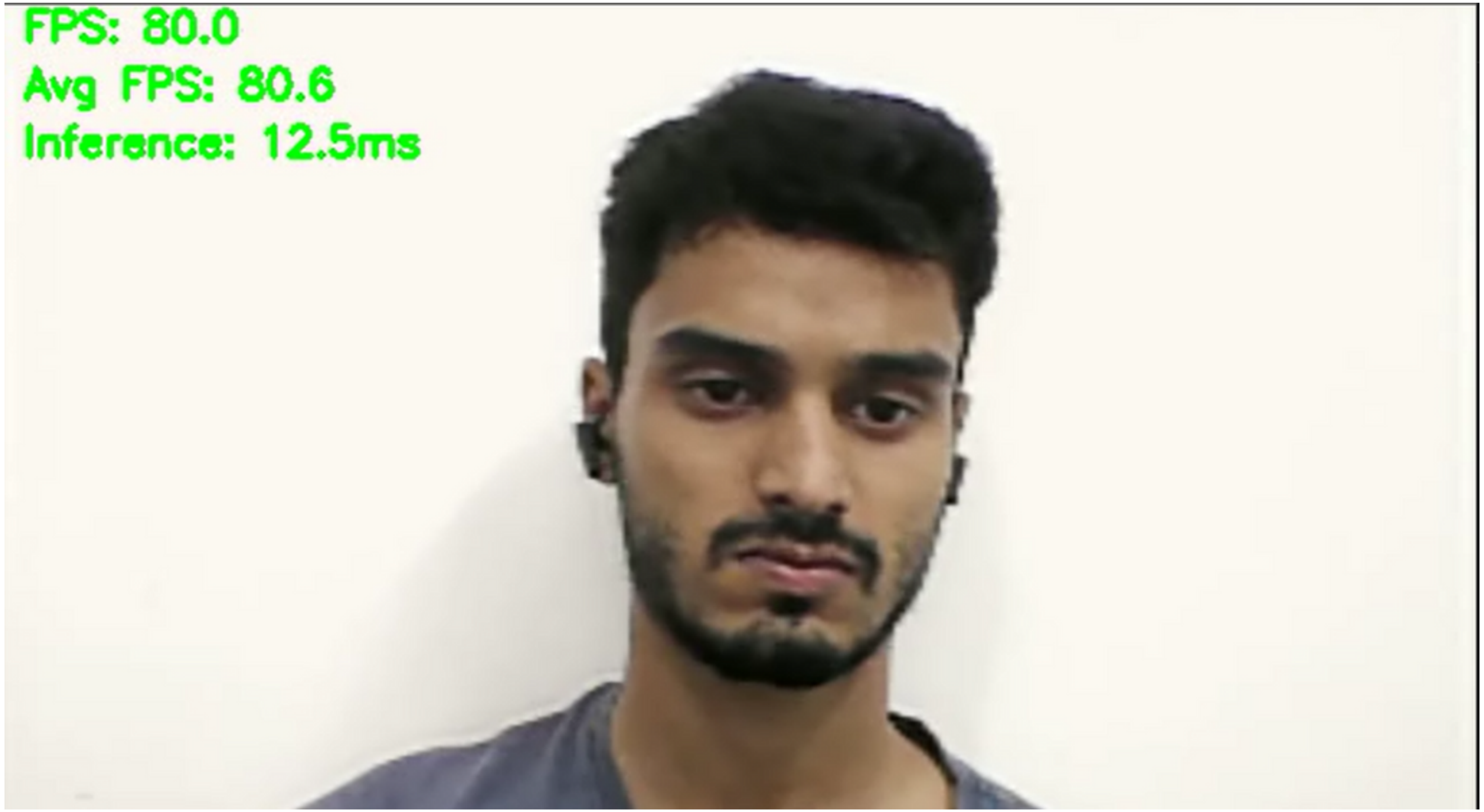
IMAGES


Figure 1
x=465. y=142
[82, 70, 67
Sample 6 | SSIM: 0.9282 | FPS: 3.22
Input | Prediction | Ground Truth
FRANGCORS FANCY

IMAGES



IMAGES



COMPONENTS USED FOR STUDENT OUTPUT

1. Student Model (StudentNetEnhanced)
   - A lightweight CNN optimized for real-time inference
   - Includes:
     - 6 Residual Blocks (LeakyReLU + InstanceNorm)
     - Channel Attention & Spatial Attention
     - Final Refinement Layer
   - Trained using Knowledge Distillation (KD) from a high-capacity teacher model (NAFNet)
2. Knowledge Distillation Losses
   - L1 Reconstruction Loss – Ensures output is close to ground truth sharp image
   - Perceptual Loss (VGG-based) – Ensures perceptual similarity with the teacher's output
   - SSIM Loss – Improves structural similarity and clarity
   - Edge (Sobel) Loss – Preserves and enhances edges
   - Feature Distillation Loss – Aligns intermediate features between teacher and student

COMPONENTS USED FOR STUDENT OUTPUT

3. Real-Time Inference Pipeline

- **Input: Live webcam/video frame (blurry)**
- **Model Output: Sharp image tensor (256×256 or 512×512)**
- **Post-processing:**
  - **Unsharp Masking using a 3×3 Gaussian blur kernel:**
    - **Formula: output = input + strength * (input - blurred)**
    - **Implemented using torch.nn.functional.conv2d**
    - **Controlled via a strength parameter (e.g., 0.15)**

4. OpenCV Pipeline

- **Captures input frames from webcam**
- **Converts BGR to RGB before model inference**
- **Overlays FPS on output frame**
- **Displays and records output video with sharp visuals**

CONCLUSION

- Developed a lightweight, real-time image sharpening system using knowledge distillation.
- Successfully trained a student model guided by a high-performance teacher (NAFNet).
- Achieved SSIM ≥ 90% and FPS between 30–60, ensuring both quality and speed.
- Used a multi-loss training approach for enhanced detail and structural similarity.
- Demonstrated effective performance on webcam input, suitable for real-world deployment.
- Applicable in areas like video conferencing, surveillance, AR/VR, and mobile devices.
- Proved that high-quality image enhancement is possible on low-resource hardware.

THANK YOU

Team : VisionARV