INTEL UNNATI
INDUSTRIAL TRAINING 2025

A PRESENTATION BY ASHISH M PRABHU (TEAM INTELLECT)

PROBLEM STATEMENT

Image Sharpening using knowledge distillation

The main aim of this project is to create a deep learning model that improves the sharpness of video images during online meetings. In many cases, low bandwidth or unstable internet connections can cause videos to appear blurry or unclear. This project addresses that issue by applying knowledge distillation, a technique that helps train a lightweight yet effective model. The goal is to deliver better visual quality in real time, ensuring clearer video calls even when network conditions are not ideal.

DESCRIPTION

This project aims to develop an efficient real-time image sharpening system for low-power devices using Knowledge Distillation. A high-performance UNetTeacher teacher model trains a compact UNetStudent network to convert blurry inputs into clear outputs. Training employs a multi-loss approach combining reconstruction accuracy, perceptual quality, SSIM metrics, edge preservation, and teacher feature distillation.

Key capabilities:

- Maintains SSIM scores ≥90 at 30–60 frames per second for 256×256 images
- Student model uses channel/spatial attention and instance normalization for speed
- Implements multi-loss Knowledge Distillation with a pretrained UNetTeacher
- Shows live sharpening on webcam feeds using PyTorch and OpenCV

Practical uses:

- Real-time enhancement for mobile phones and embedded devices
- Preprocessing for security cameras, medical imaging, augmented reality apps, and drone navigation systems

METHOD OVERVIEW

I present an AI-based image enhancement system designed for real-time use in low-resource settings. It sharpens blurry frames using a lightweight deep learning model, combining knowledge distillation and edge-aware loss for efficient performance.

Key Architecture Features

- **Teacher–Student learning with UNetTeacher-based teacher**
- **6 Residual blocks with LeakyReLU and InstanceNorm**
- **Channel and Spatial Attention for focused enhancement**
- **Refinement layer to enhance fine details**

Output:

Generates sharp 256×256 images from blurry inputs with real-time efficiency.

METHOD OVERVIEW

Loss Functions & Training Strategy :

- Used multiple loss functions for effective learning:

    1. L1 Loss for pixel-wise reconstruction
    2. Perceptual Loss (VGG) to maintain visual quality
    3. Sobel Edge Loss to enhance edge sharpness
    4. SSIM Loss for structural similarity
    5. Feature Distillation Loss to guide student with teacher features

- Training used AMP (Mixed Precision) for faster, memory-efficient training

- Applied Early Stopping and Cosine Learning Rate Schedule for training stability

- Dataset: Paired blurry and sharp images, resized to 256×256

METHOD OVERVIEW

Performance & Deployment:

- SSIM ≥ 90 and PSNR > 30 dB achieved
- Runs at 30–60 FPS — suitable for real-time use
- Deployed using OpenCV webcam pipeline
- Shows live sharpened output with FPS counter
- Lightweight: Works on 4 GB GPU or CPU fallback

FEATURES OF THE MODEL

Feature	Description
Real-time Image Enhancement	Instantly sharpens blurry images and video frames using a student-teacher AI approach.
High SSIM and PSNR Scores	Delivers consistently high SSIM (>90) and PSNR values for quality enhancement.
Low-Latency Processing	Processes 30–60 FPS, making it ideal for live streams and embedded systems.
Knowledge Distillation	Trains a lightweight student model under guidance from a powerful UNet-based teacher.
Edge & Perceptual Loss Optimization	Uses combined L1, edge, SSIM, and perceptual losses for crisp and accurate results.

FEATURES OF THE MODEL

Feature	Description
Compact Model Architecture	Includes residual blocks and attention layers (channel & spatial) for better feature focus.
GPU & CPU Compatible	Optimized for low-VRAM GPUs and supports inference on CPU devices as well.
Image & Video Support	Works seamlessly with image datasets, live webcam input, or video files.

SYSTEM WORKFLOW

- **User Input:**
  Blurry image frame captured from webcam or uploaded; simulates low-bandwidth video quality.
- **Preprocessing:**
  Image resized to 256×256, normalized, and converted to tensor format for model input.
- **Inference (Student Model):**
  Enhanced UNetStudent processes input in real-time (30–60 FPS), guided by a pretrained UNetTeacher teacher via knowledge distillation.
- **Model Internals:**
  Contains residual blocks (LeakyReLU + InstanceNorm)
  Uses channel & spatial attention for targeted sharpening
  Refinement layer enhances fine details
  Trained using combined losses: L1 + SSIM + Perceptual + Edge + Feature Distillation

SYSTEM WORKFLOW

- **Output Generation:**
  Sharp 256×256 output, upscalable to 1920×1080 for display.

- **Evaluation:**
  Achieves SSIM ≥ 90%, maintains 30–60 FPS, and displays output in real time via OpenCV.

- **Outcome:**
  Real-time high-quality image enhancement on low-resource systems — suitable for video calls, AR/VR, mobile apps, and surveillance.

TECHNOLOGIES USED

Deep Learning Frameworks

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

Model Architectures

- Teacher: UNetTeacher (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)

CONCLUSION

- Built a lightweight image enhancement system capable of real-time sharpening using knowledge distillation.
- Trained an efficient student model under the supervision of a high-performing teacher model (UNetTeacher).
- Reached SSIM scores of 90% or higher and maintained 30–60 FPS, balancing quality with performance.
- Implemented a multi-loss strategy during training to preserve fine details and structural integrity.
- Validated performance using live webcam input, confirming its readiness for real-world applications.
- Suited for use in video calls, security systems, augmented/virtual reality, and mobile platforms.
- Demonstrated that high-quality image sharpening can be achieved on devices with limited resources.

Thank You
For Your Time