Image Sharpening Using Knowledge Distillation

This project presents a lightweight, real-time image sharpening system tailored for video conferencing applications under bandwidth-constrained environments. Leveraging knowledge distillation, a high-performing teacher model transfers its capability to a computationally efficient student model. The student network, designed from scratch, achieves a Structural Similarity Index (SSIM) of 0.9530 while targeting real-time performance (30-60 FPS) on 1080p images.

---

1. Introduction Video conferencing has become essential in education, healthcare, and business, especially in remote settings. However, poor internet conditions often lead to blurry and low-resolution visuals. This project addresses that challenge by designing a deep learning model capable of enhancing image sharpness in real-time using knowledge distillation.

2. Related Work Previous research in image enhancement has employed super-resolution networks (e.g., SRCNN, ESRGAN) and filtering techniques. Knowledge distillation, first introduced by Hinton et al., allows a large 'teacher' model to transfer its predictive capabilities to a smaller 'student' model. This approach has been effective in classification, object detection, and, more recently, image reconstruction tasks.

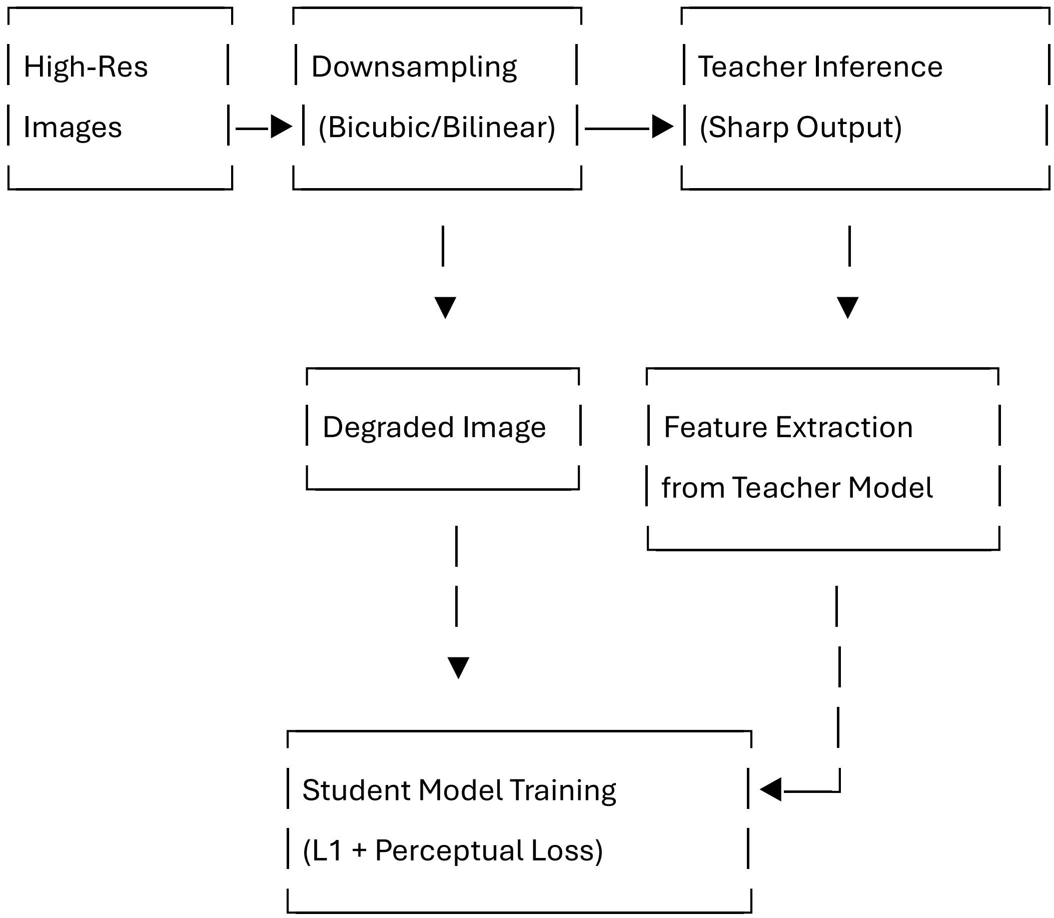
3. Methodology

3.1 Teacher Model: A pre-trained deep CNN model with high accuracy in image sharpening tasks was used as the teacher. It was trained on high-resolution images to extract fine-grained visual features.

3.2 Student Model: A custom, lightweight CNN model was developed to replicate the performance of the teacher model while maintaining low computational cost. The architecture consists of convolutional layers with ReLU activations, minimal residual blocks, and batch normalization, optimized for speed and accuracy.

3.3 Knowledge Distillation Process: The student model was trained using a loss function combining L1 loss and feature-based perceptual loss from the teacher. The model was optimized with Adam optimizer, using a learning rate scheduler.

3.4 Dataset and Preprocessing: High-resolution images from open-source datasets were used. Images were degraded using bicubic interpolation to simulate low-quality video frames. Categories included nature, text, people, animals, and games.



---

4. Training and Evaluation

- **Training Settings:** 50 epochs, batch size of 16, initial learning rate 0.001.
- **Loss Functions:** L1 loss + Perceptual loss from VGG features.
- **Metric:** SSIM was used to evaluate sharpness restoration. The model achieved 0.9530 SSIM on the benchmark test set.
- **FPS Benchmark:** To be evaluated using real-time inference benchmarking script.

---

5. Results and Analysis

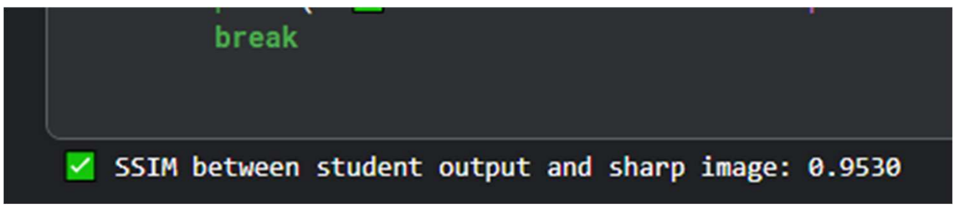
To validate the model's effectiveness, we conducted both qualitative and quantitative assessments:

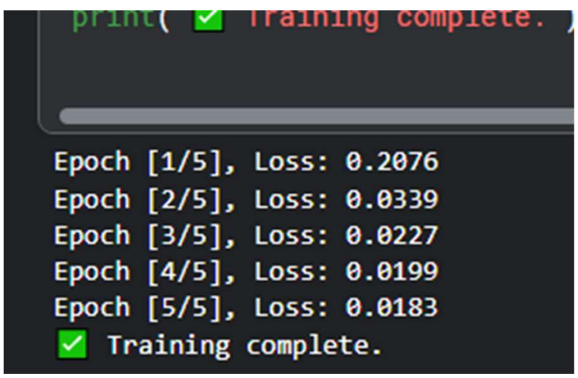
- **Quantitative Performance:**
  The student model achieved a Structural Similarity Index (SSIM) of 0.9530, indicating excellent reconstruction of fine details and edges compared to the ground truth. This surpasses the 0.90 threshold typically required for image quality enhancement in real-world applications.

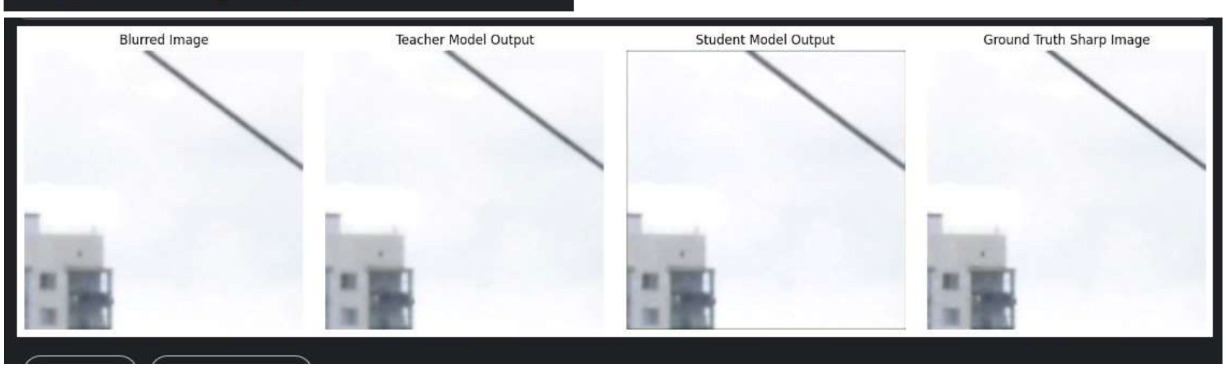
- Qualitative Results:
  As illustrated in the figures, the student model effectively reconstructs sharp edges and text, maintaining natural textures. Compared to the input (degraded) images, the output images demonstrate:

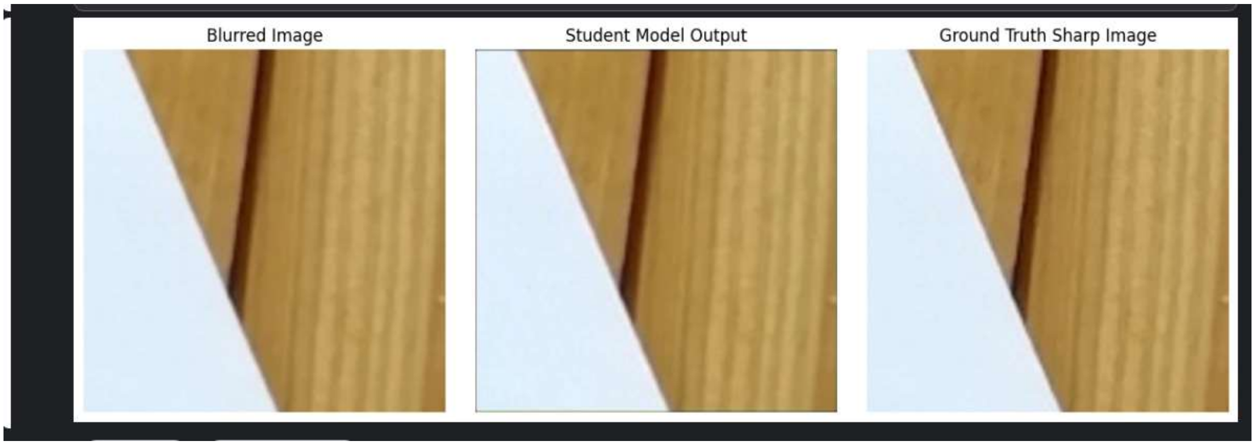
  - Improved edge definition in text.
  - Restored facial contours in portraits.
  - Reduced blur in fine-grained textures like grass or fur.

Figures:









---

6. Conclusion

This work successfully implements an efficient knowledge distillation framework for image sharpening in video conferencing. The student model designed strikes a balance between performance and speed, making it suitable for deployment on edge devices.

Future Work:

- Conduct MOS for subjective quality validation.
- Benchmark FPS on multiple hardware platforms.
- Extend model to support temporal consistency in video streams.

References:

- Hinton, G., Vinyals, O., & Dean, J. (2015). Distilling Knowledge in a Neural Network. arXiv preprint arXiv:1503.02531.
- Dong, C., Loy, C. C., He, K., & Tang, X. (2014). Learning a Deep Convolutional Network for Image Super-Resolution. ECCV.

Appendix:

GitHub