Dejavu: Enhancing Videoconferencing with Prior Knowledge

Pan Hu, Rakesh Misra, Sachin Katti
Stanford University and Uhana Inc.

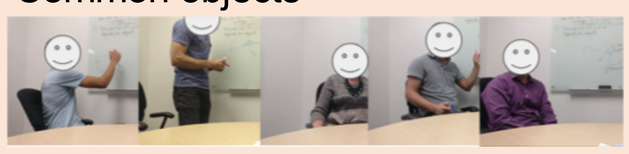
Motivation

Heavily compressed, low quality videoconferencing video

- Limited uploading bandwidth on mobile network
- Conservative quality to ensure interactivity

Visual similarities across videoconferencing sessions

- Same room / person
- Common objects



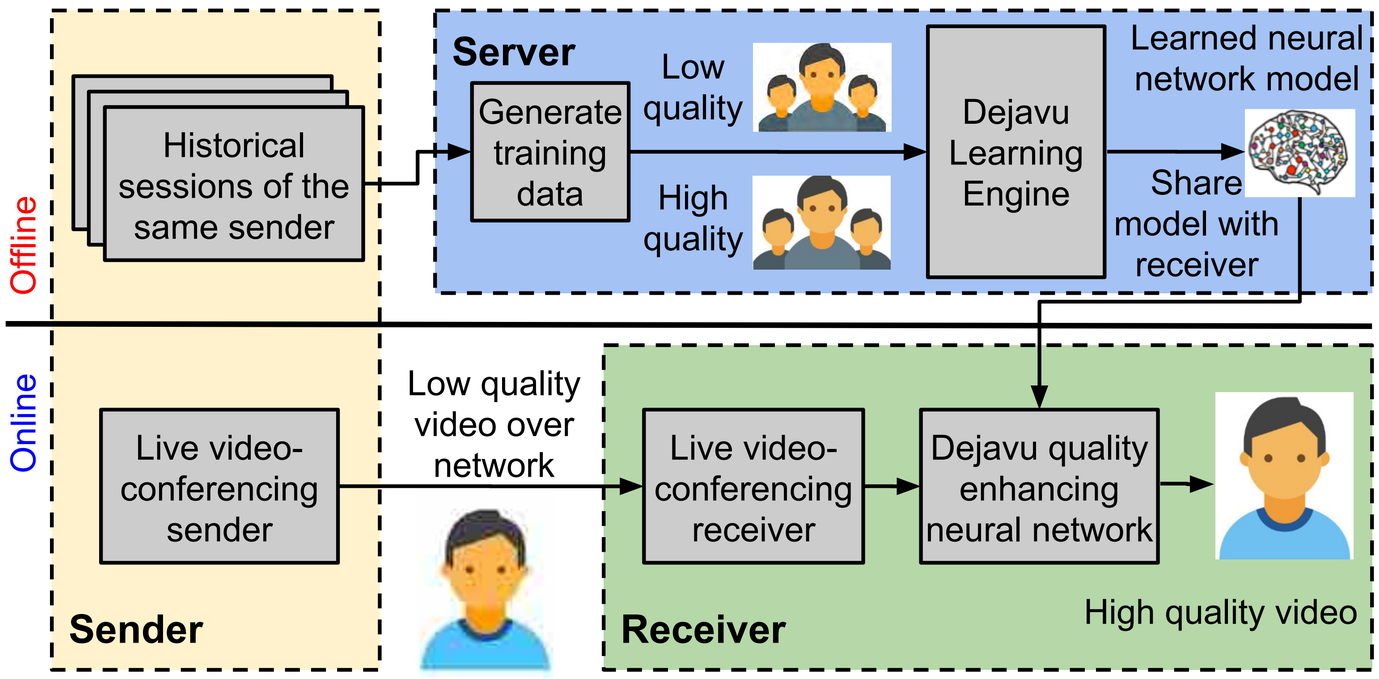
Encode similarities (prior knowledge) into neural network

Challenge

- Neural network design
  - Fast (ideally running 30fps @1080p, even on cellphones)
  - High PSNR gain, even on very blurry videos
- Analyze performance gain in different scenarios

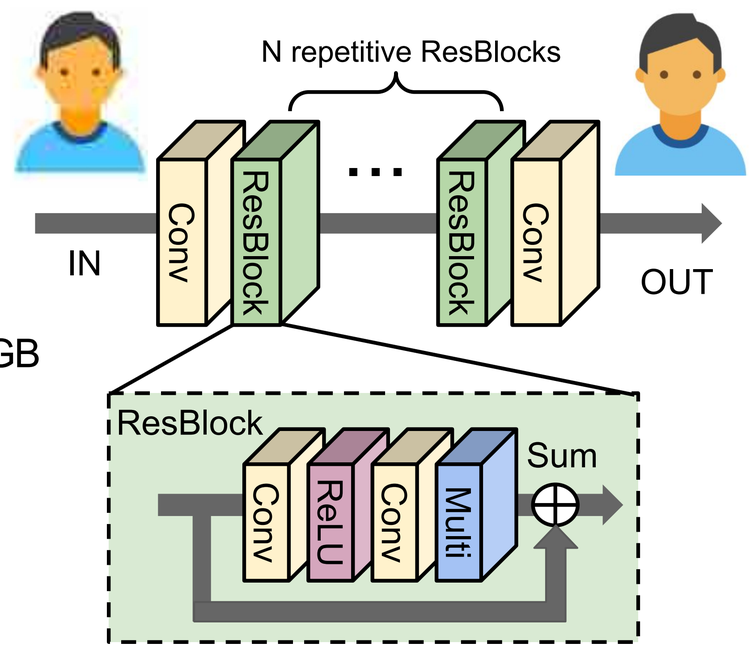
System Design

System overview

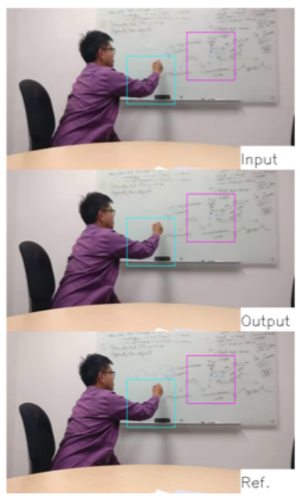


Neural network architecture

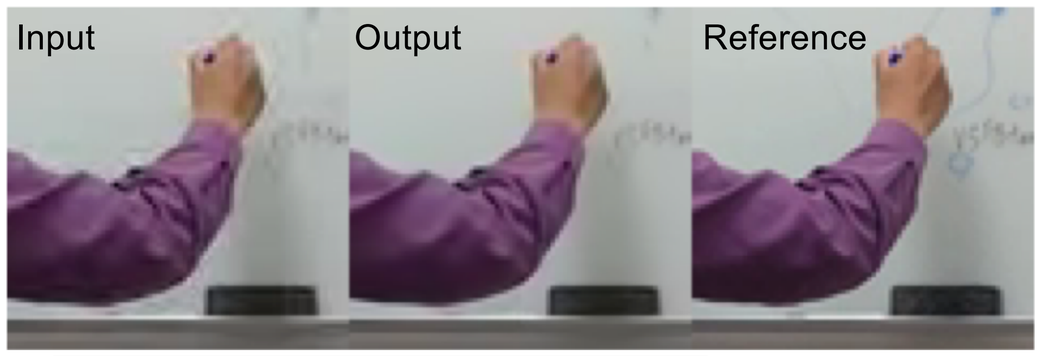
- Reuses EDSR
- Process frame by frame
- Convert RGB into YUV, processing on Y only

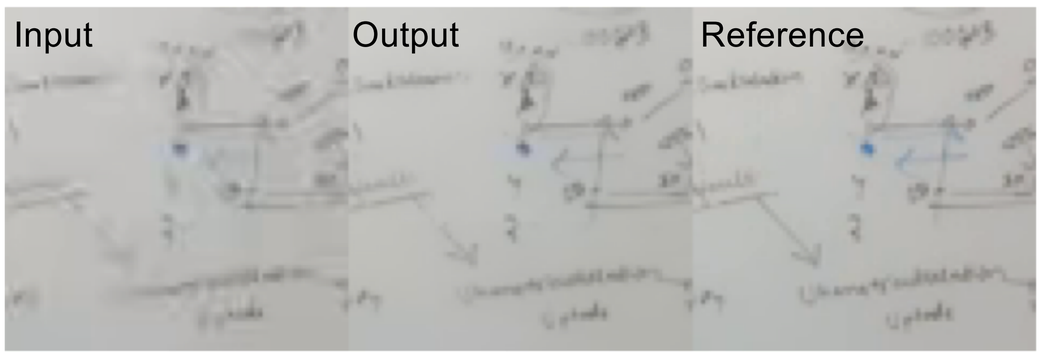


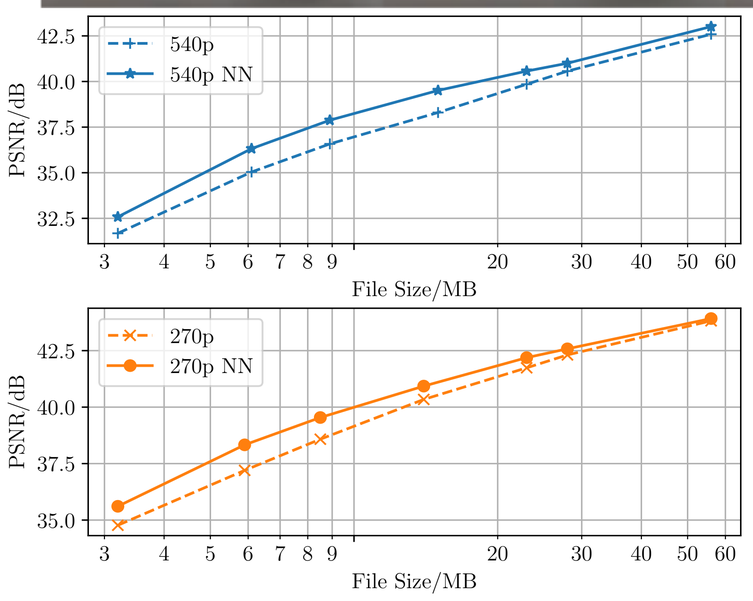
Experiment Result



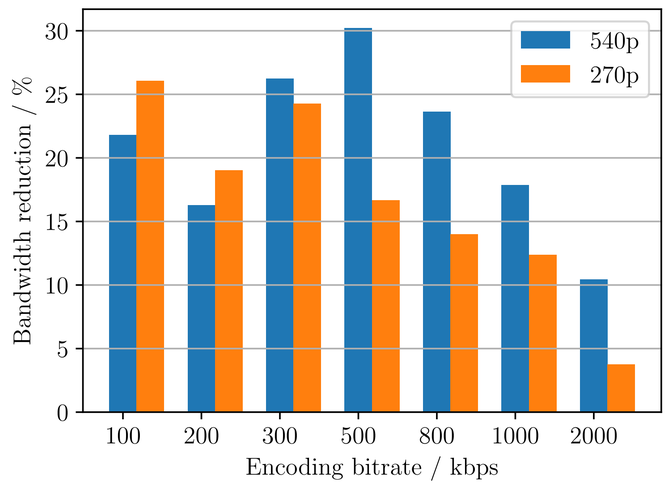
- Up to 1.3dB PSNR Gain on 540p, 1.1dB on 270p
- Save up to 30% bandwidth



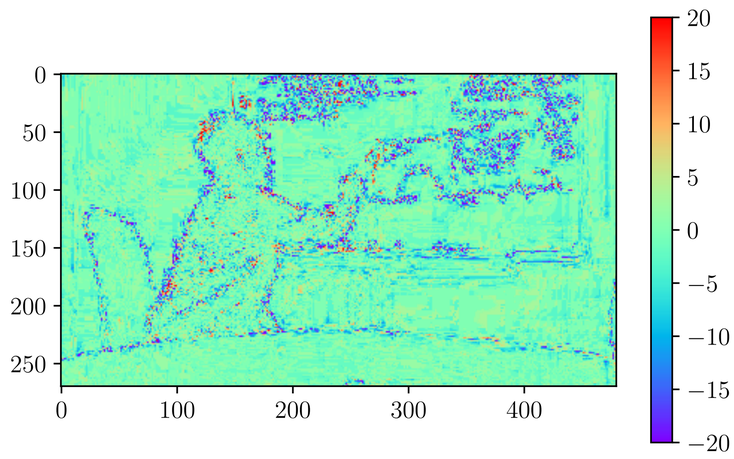




- PSNR gain on 270p/540p



- Calculated bandwidth saving



- Error map between Dejavu input and output

Discussion

Problems need to be solved for a practical Dejavu:

1. **Evaluate real-world performance**
   - Collect large-scale, real-user dataset.
   - Evaluate Quality of Experience that includes processing delay in real system
2. **More efficient inference**
   - Exploit inter-frame similarity based on motion estimation (from codec) or reuse part of NN (clockwork RNN)
   - Knowledge distillation or model compression / quantization to speed up / fit in small RAM

Similar mechanism could be applied to a wide range of video streaming apps (like Twitch / Youtube) to **improve quality** or **reduce CDN cost**!