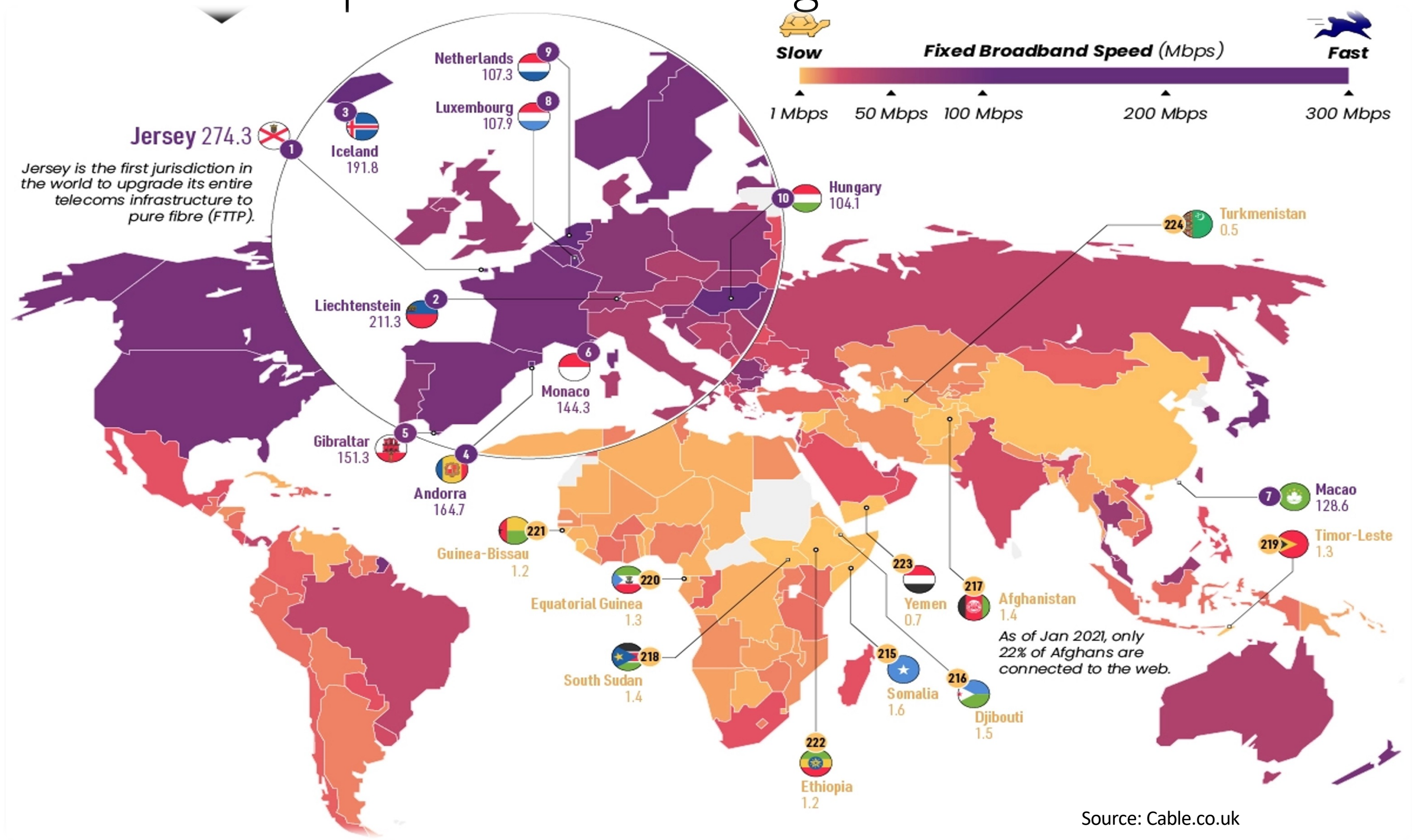
Gemino: Practical and Robust Neural Compression for Video Conferencing

Vibhaalakshmi Sivaraman[1], Pantea Karimi[1], Vedantha Venkatapathy[1], Mehrdad Khani[1], Sadjad Fouladi[2], Mohammad Alizadeh[1], Frédo Durand[1], Vivienne Sze[1]

[1]MIT CSAIL [2]Microsoft Research

Motivation

- State of the art codecs' search-based methods (i.e., VP8) have limited bitrate range
- Many countries fall below bandwidth recommendations for video conferencing
- Poor user experience even in high-bandwidth areas



Source: Cable.co.uk

Gemino Design

Warping approaches

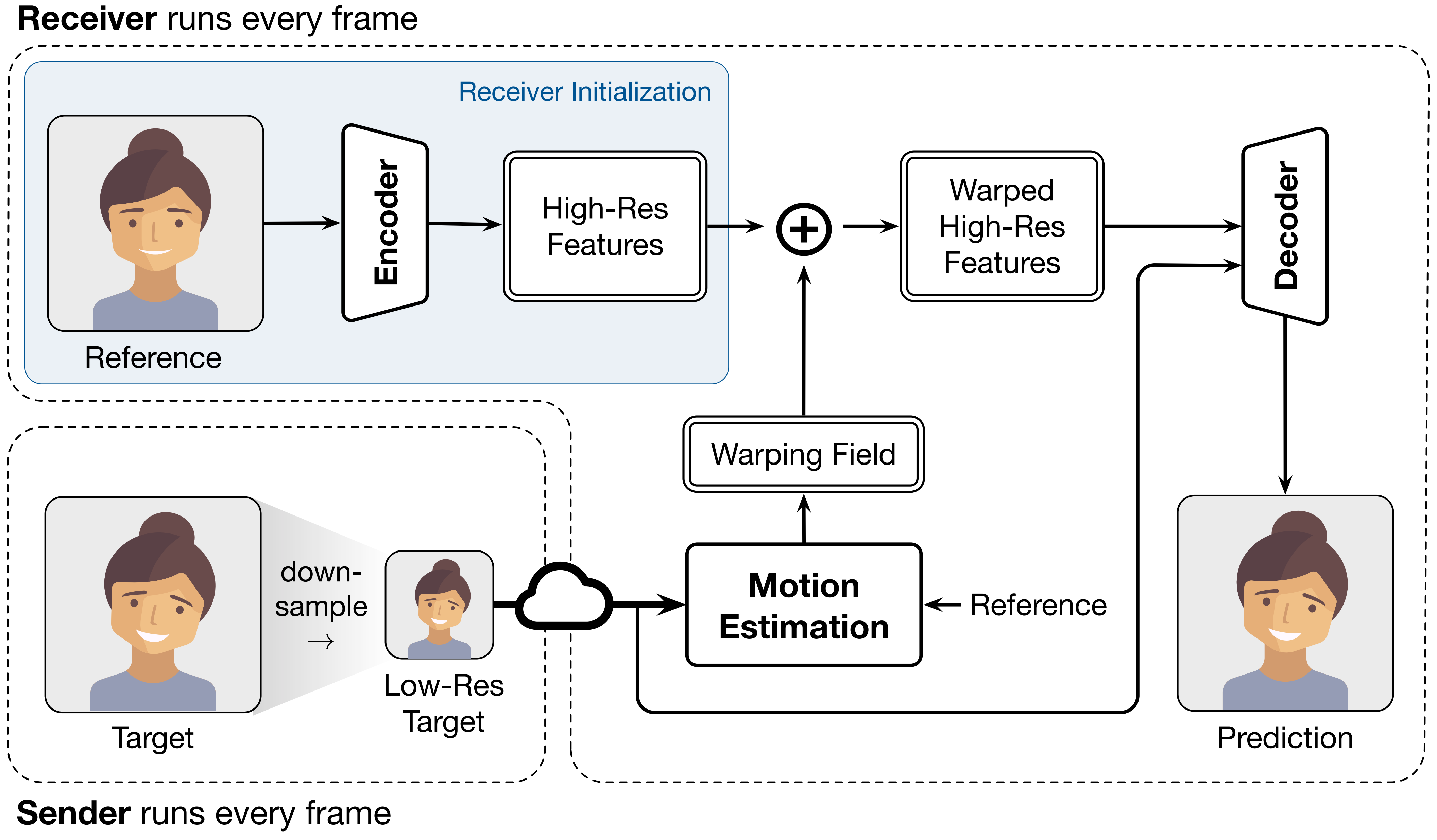
- Use sparse facial landmarks to estimate pose
- Catastrophic failure with large motion
- Good high-frequency fidelity for small motion

Super-resolution approaches

- Preserve low-frequency content
- Poor high-frequency fidelity

Gemino

- Uses high-frequency-conditional super-resolution to combine both approaches
- Extreme super-resolution (8x) to achieve audio like bitrates


Receiver runs every frame

Receiver Initialization

Reference

Encoder

High-Res Features

Warped High-Res Features

Decoder

Warping Field

Motion Estimation

Reference

Prediction

Target

down-sample →

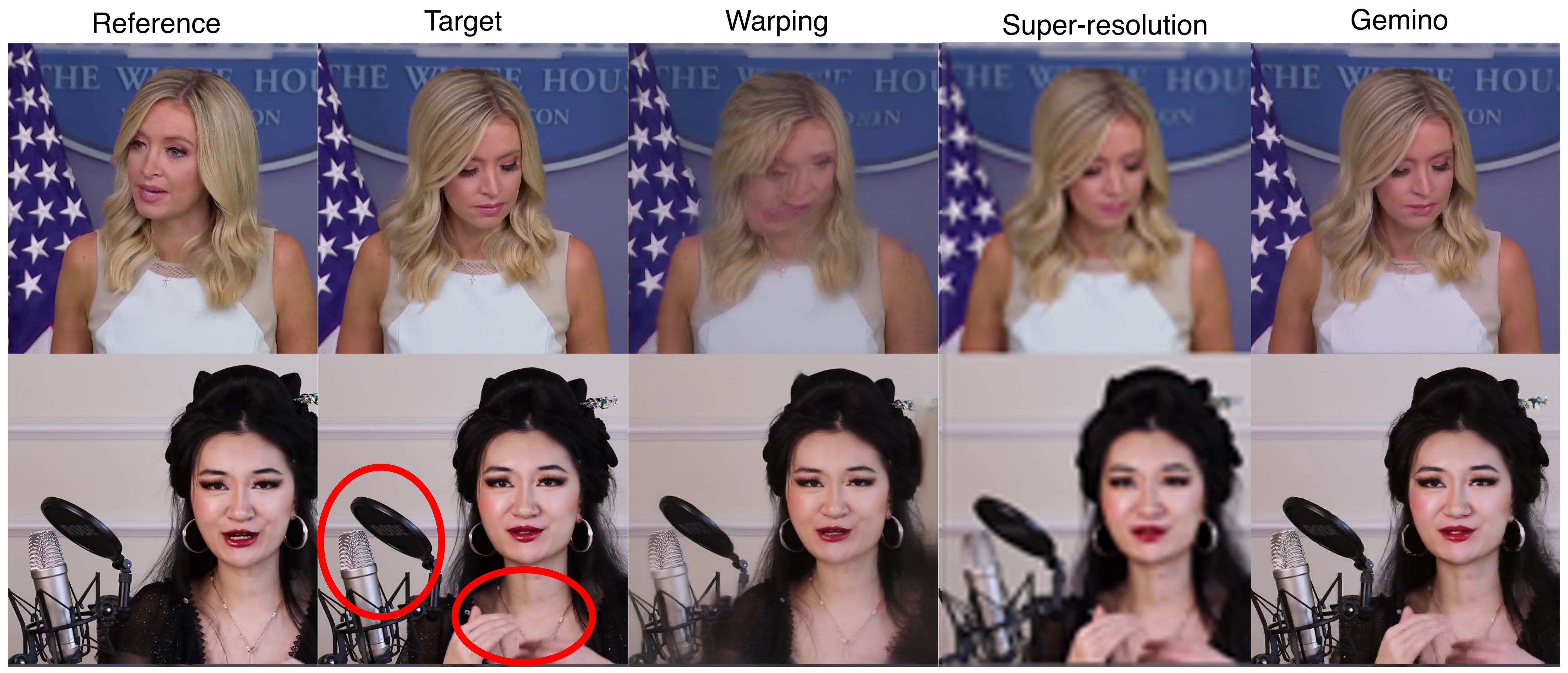
Low-Res Target

Sender runs every frame

Optimizations

- Codec-in-the-loop training to overcome artifacts produced by video codecs at low bitrates and low resolutions
- Per-person fine-tuning for improved fidelity
- Multi-scale architecture to reduce operations per pixel at higher resolutions
- Depth-wise separable convolutions to reduce MACs
- Channel pruning further allows for real-time inference on a TitanX GPU at 1024x1024 resolution.

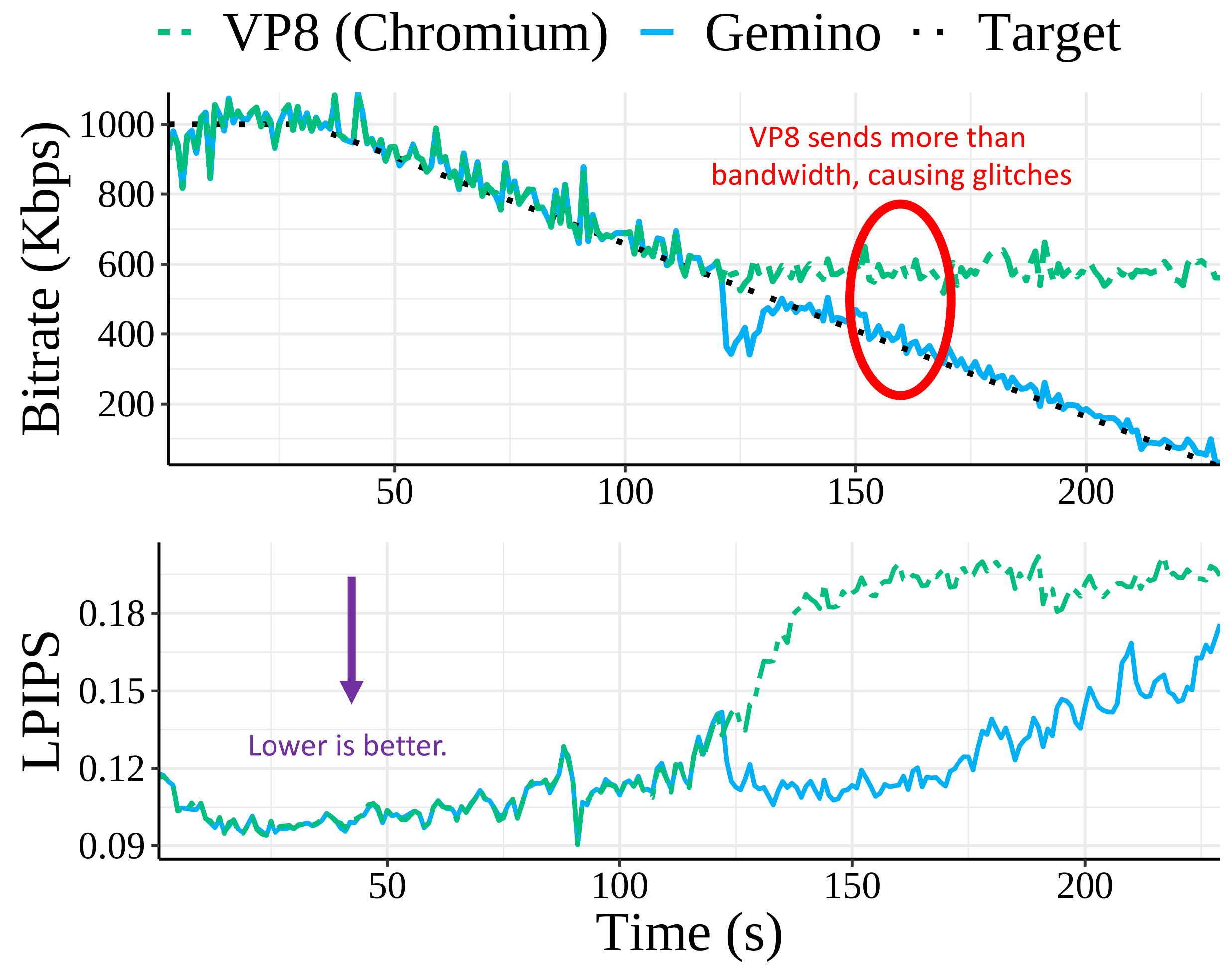
Qualitative Comparisons


Reference Target Warping Super-resolution Gemino

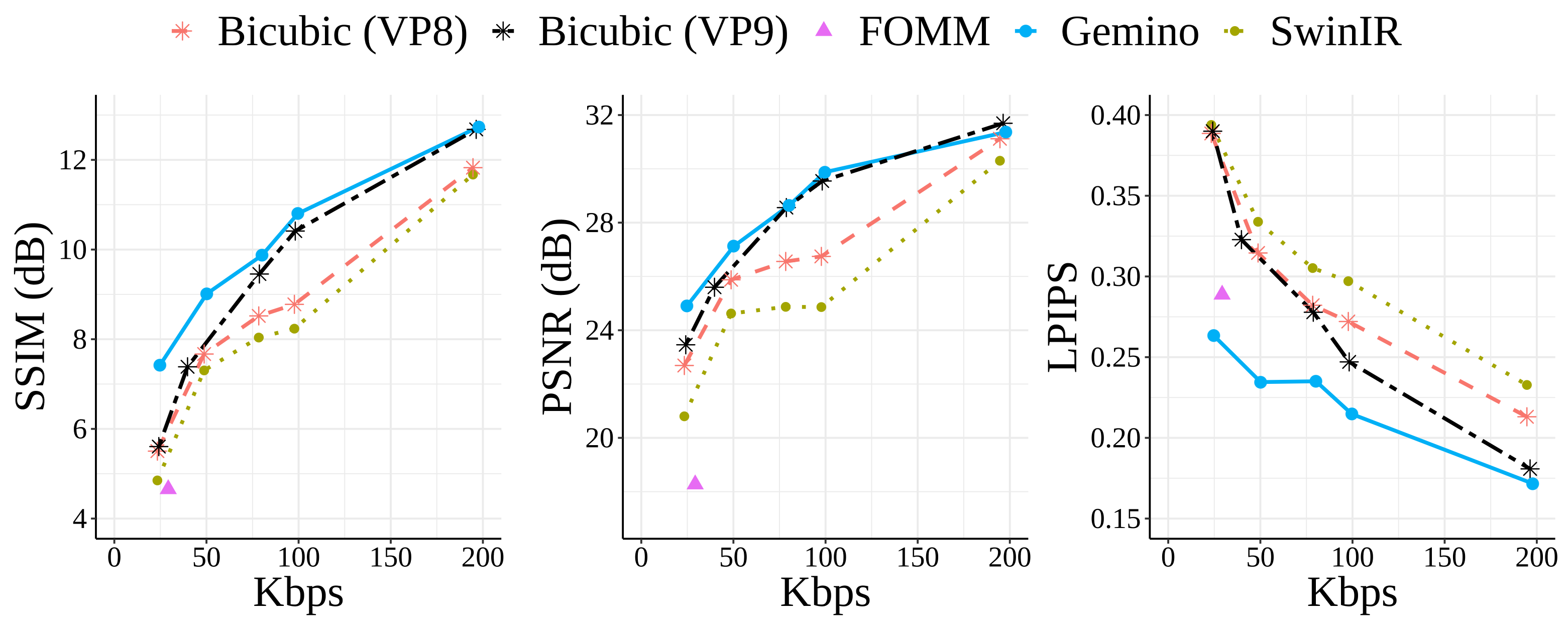
Quantitative Comparisons

Adaptation to network variability

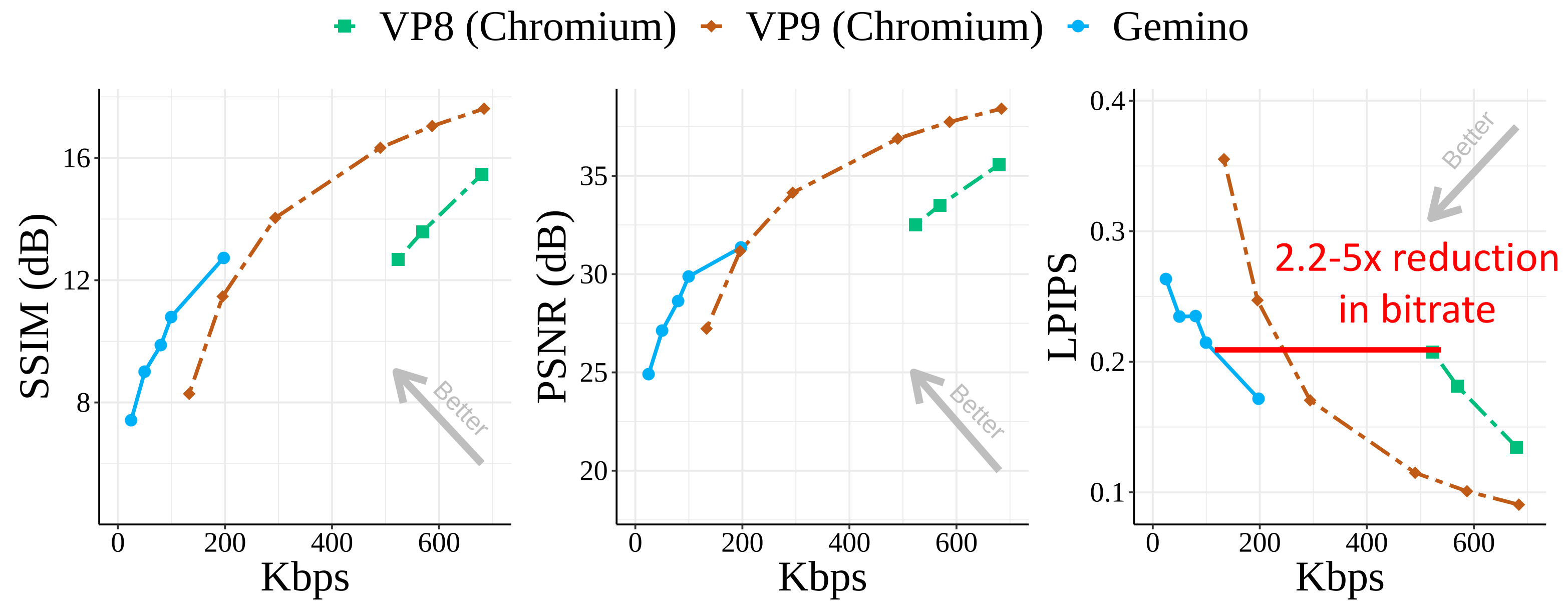
- VP8 saturates at few 100 Kbps
- Gemino responds to target bitrate and smoothly trades off compression for visual quality.



Low-bitrate Regime



Gemino vs. VP8/9



Gemino provides same video quality at 2.2-5x lower bandwidth!

Contact: vibhaa@mit.edu