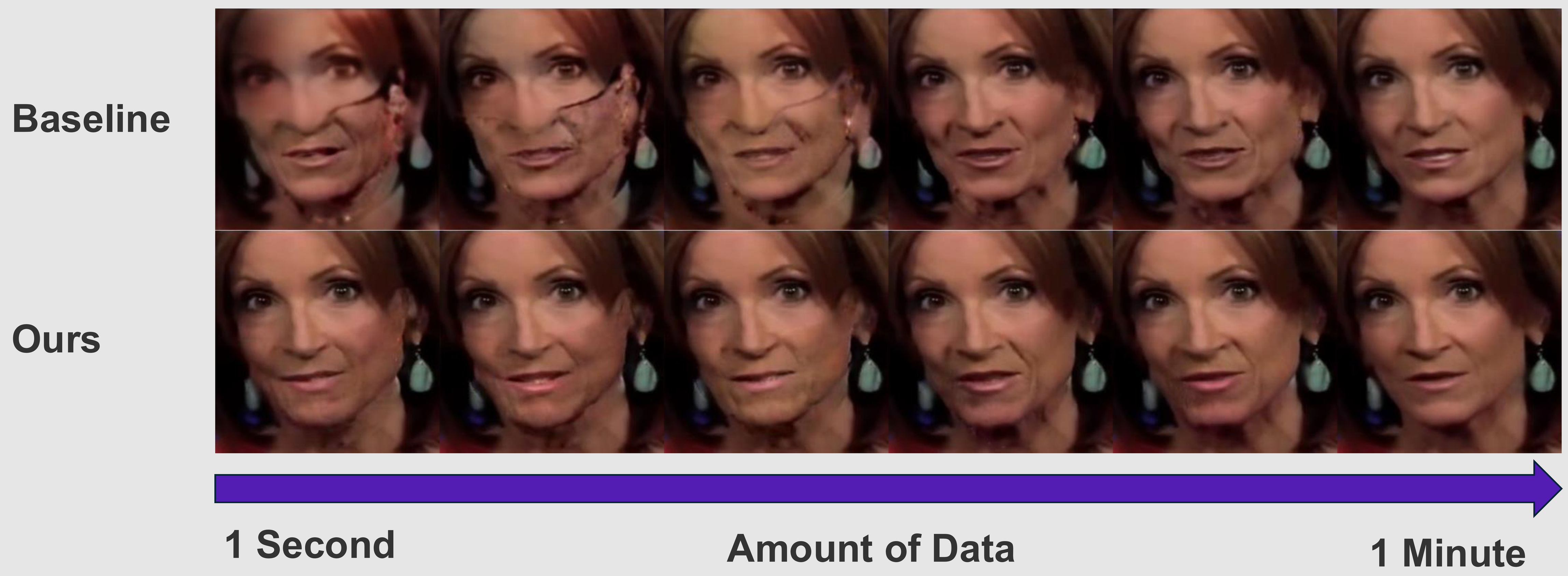
DEAD: Data-Efficient Audiovisual Dubbing using Neural Rendering Priors

Jack Saunders & Vinay P. Namboodiri
University of Bath



TLDR: We present a method for High-quality visual dubbing with only a few seconds of actor data.

Problem

- Visual Dubbing is the process of changing the lips of a video to match a given audio
- Extras appear in movies for as little as just a few seconds. How can we dub them?

Existing Solutions

Existing solutions are either person-generic/zero-shot and low quality, or require several minutes of person-specific data

Person-Generic Models

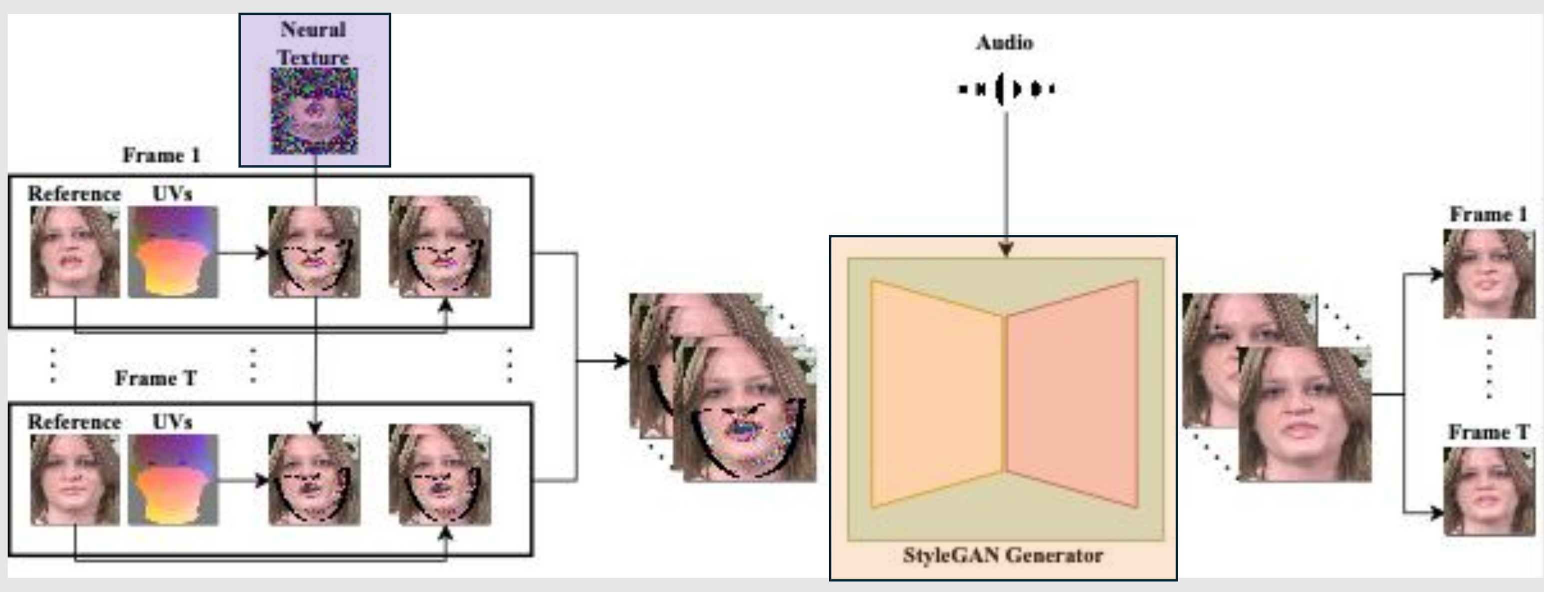
- Trained on a large multi-person dataset
- Works zero-shot with a reference image
- Usually robust but not high quality

Person-Specific Models

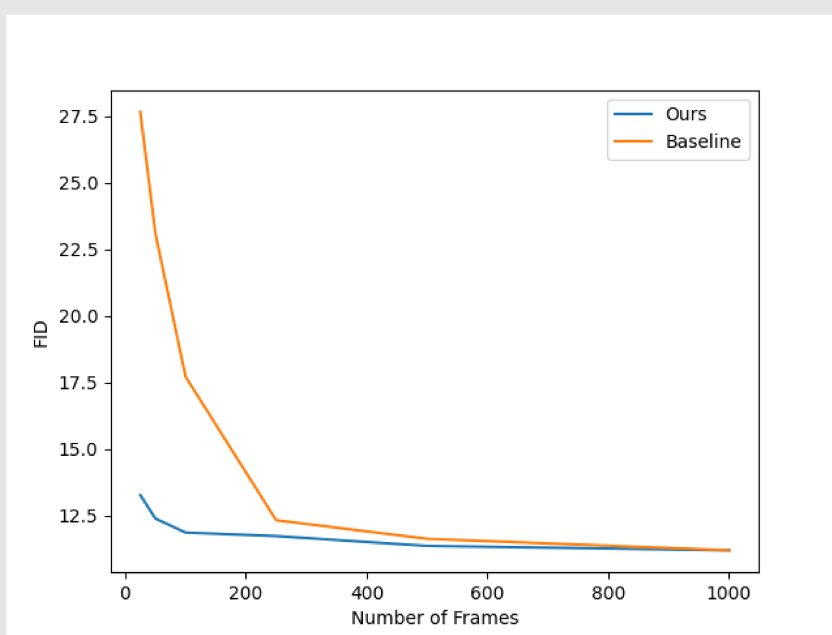
- Trained on 2-5 minutes of data for a single person
- Works only on the person it was trained on
- Usually high quality but not robust

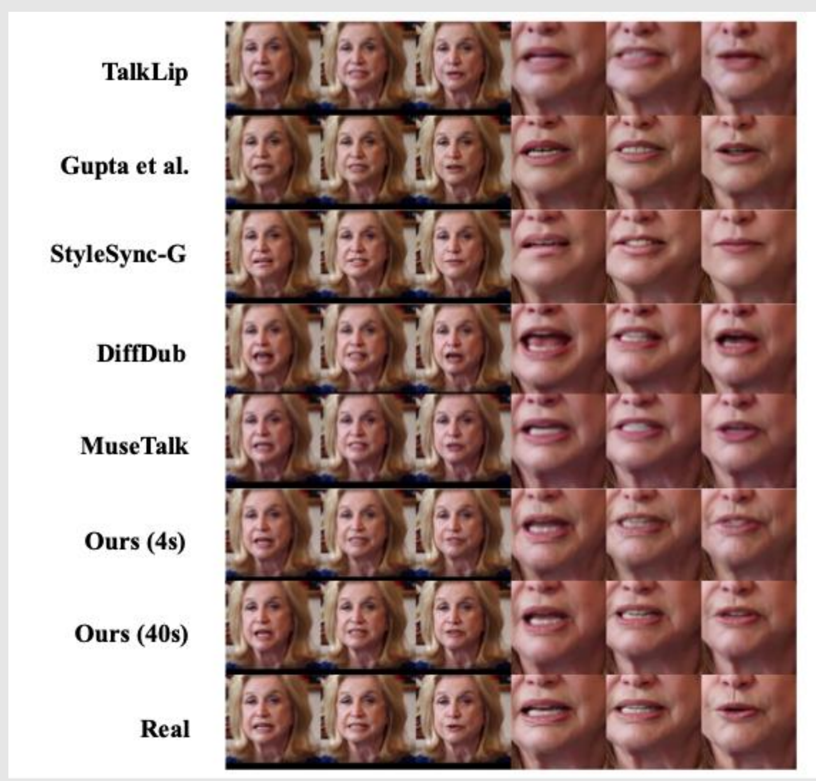
Our Solution

Best of both worlds. We build on **deferred neural rendering**, splitting it into person-generic (image-to-image) and person-specific (neural textures) components.



Results





BMVC 2025

UNIVERSITY OF BATH