EECS 504: Foundations of Computer Vision

Team: Pixel Polyglots

Saket Pradhan (saketp@umich.edu)
Kanishka Gabel (kgabel@umich.edu)
Srushti Hippargi (shipparg@umich.edu)
Shrey Shah (shreyzz@umich.edu)

presents

Linguistic Avatars


Babbel


tandem


italki


MEM
RiSE


HelloTalk


duolingo


Busuu

But they all lack a key component of the ideal language experience...

SPEECH

AND

PRONUNCIATION

4 IN A ROW

4

Listen for the missing word

Meine Katze steht nie vor zehn Uhr morgens auf.

1

2

Nice! Meaning:

My cat never gets up before ten AM.

REPORT

CONTINUE

Listen for the missing word

This is text

Meine Katze steht nie vor zehn Uhr morgens auf.

1

2

Nice! Meaning:

My cat never gets up before ten AM.

REPORT

CONTINUE

Listen for the missing word

But how to actually speak this??

Meine Katze steht nie vor zehn Uhr morgens auf.

1

2

Nice! Meaning:

My cat never gets up before ten AM.

REPORT

CONTINUE

Okay, so what options do we have to improve our speaking skills?

synthesia.io
Relaunch to update

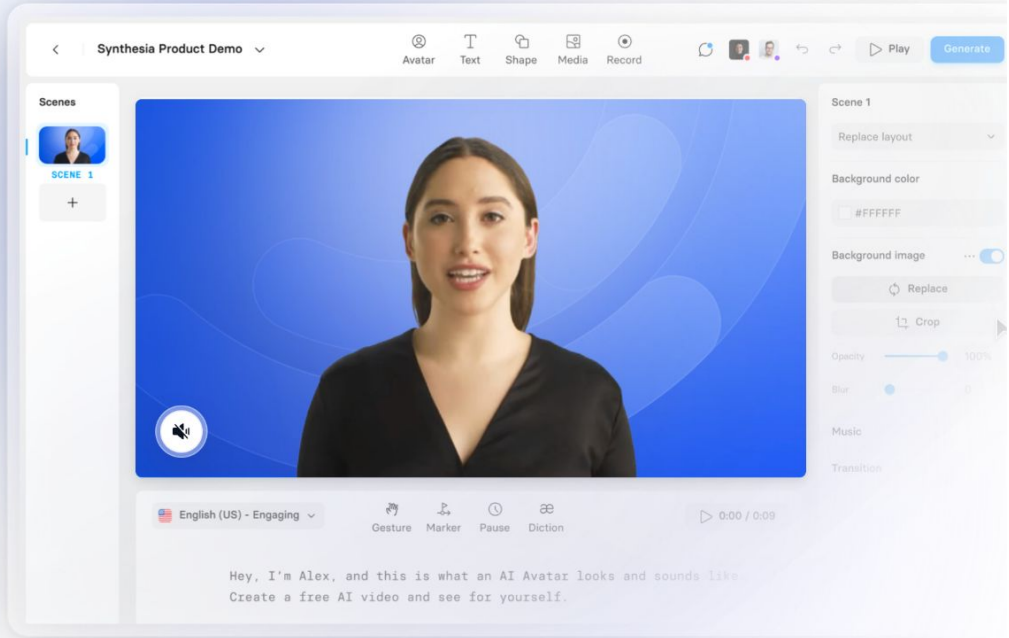
#1 AI VIDEO GENERATION PLATFORM

Turn your text into videos in minutes

- Get natural sounding AI voices in 120+ languages
- Make your videos more engaging with 140+ AI avatars
- Edit as easily as a slide deck, no experience required

Create a free AI video

No credit card required.



Trusted by over 50,000 companies of all sizes

So, what’s wrong with it?

1. It’s not open-source

1. It’s not open-source
2. It’s damn too expensive

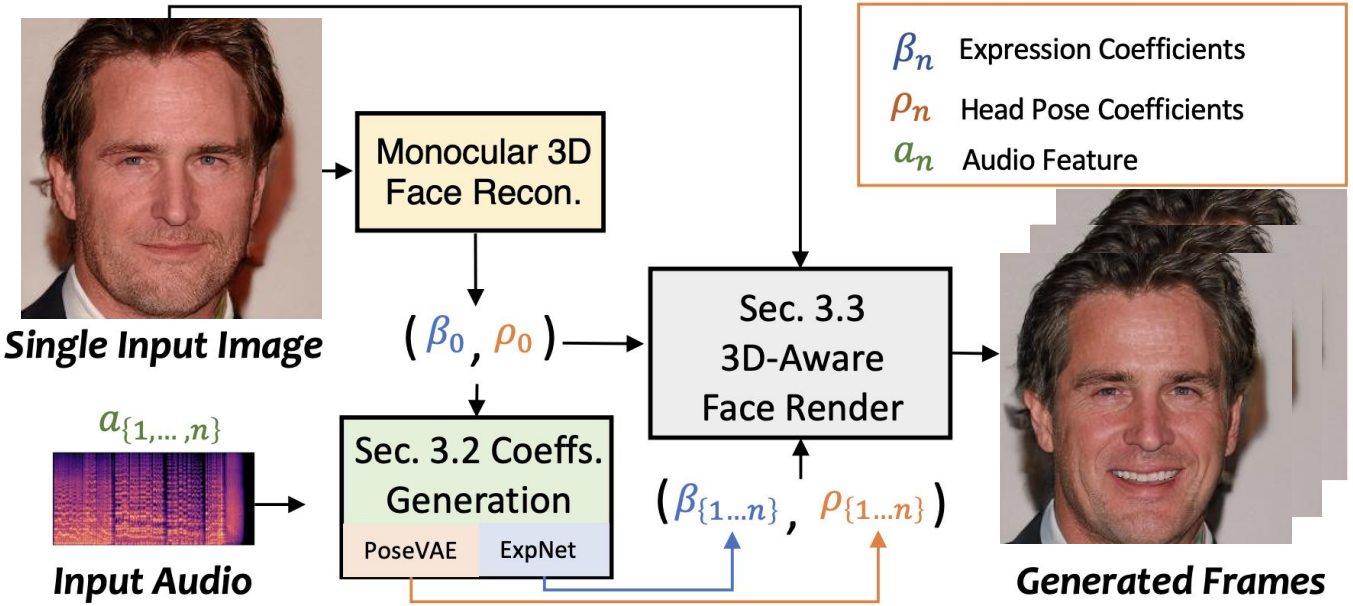
What can we do about it?

Create an alternative service that can produce high quality deepfakes with minimal GPU usage

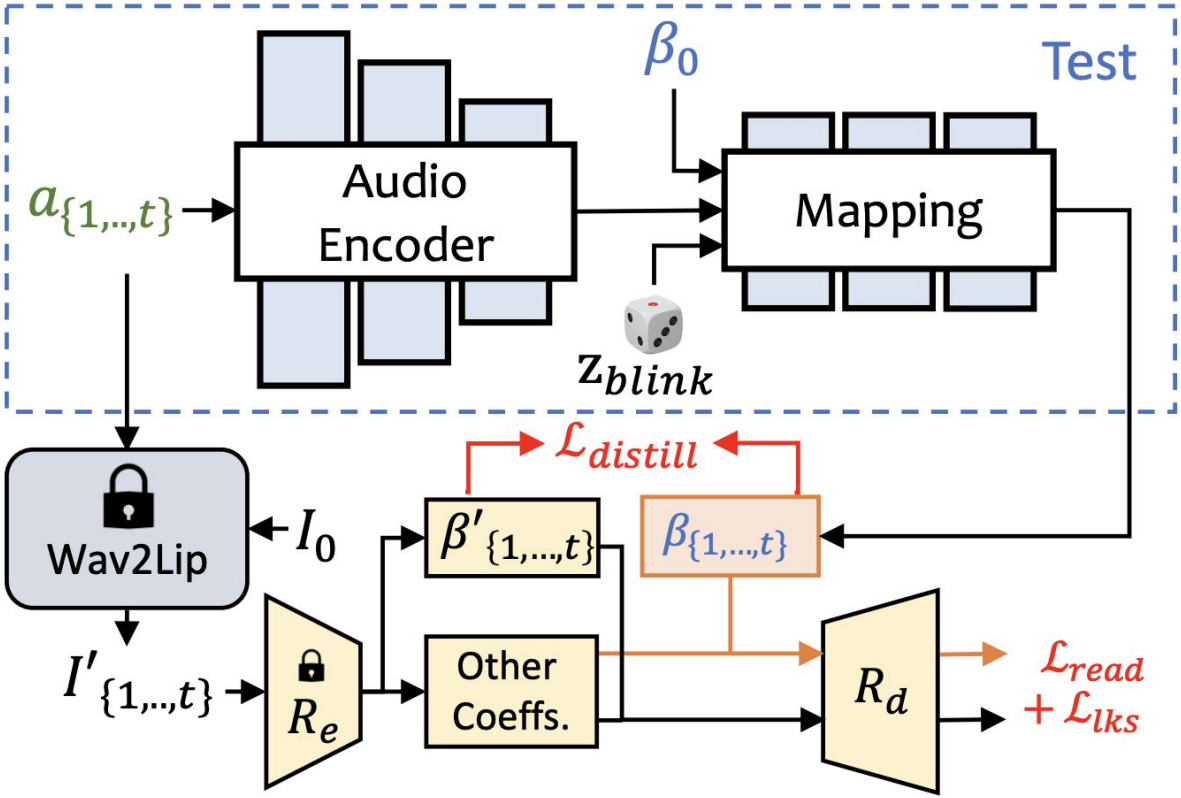
Original Image

After Processing

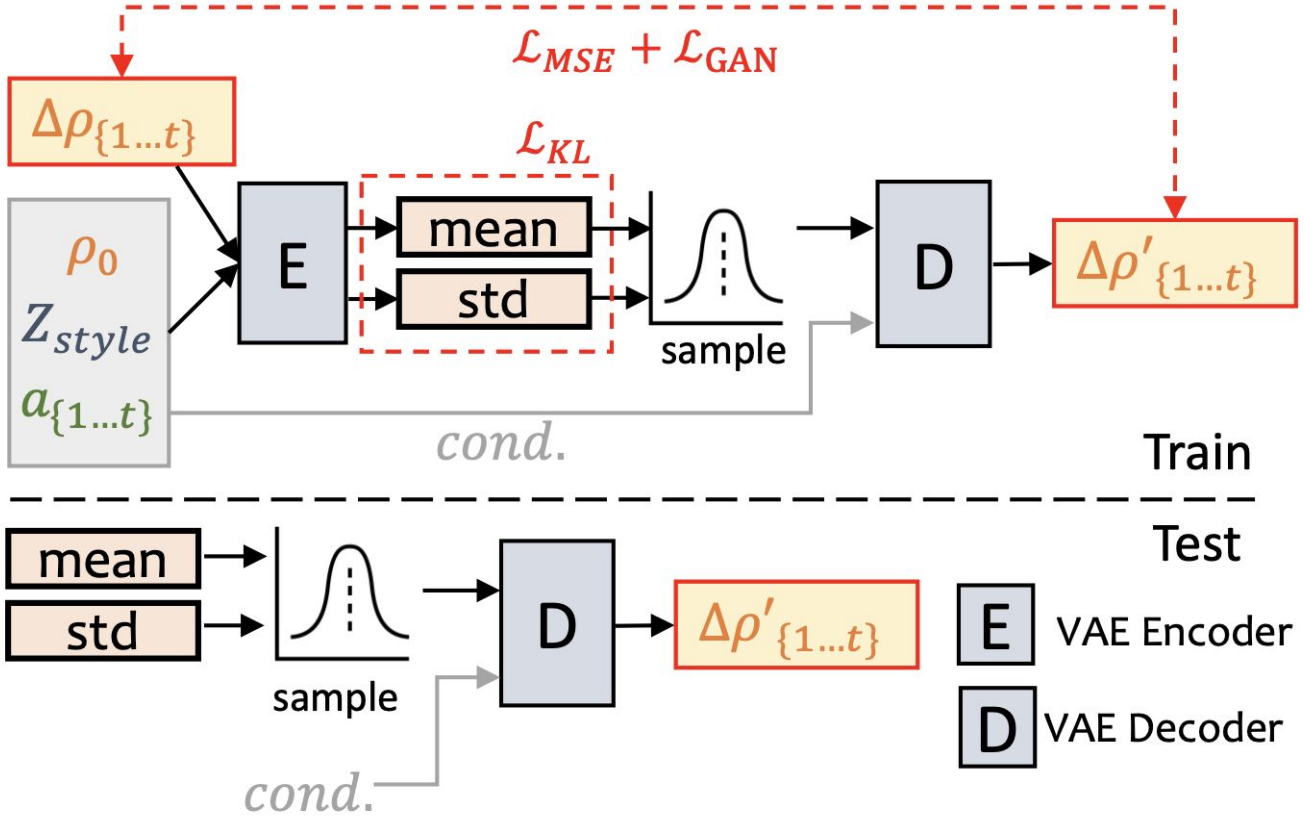
SadTalker



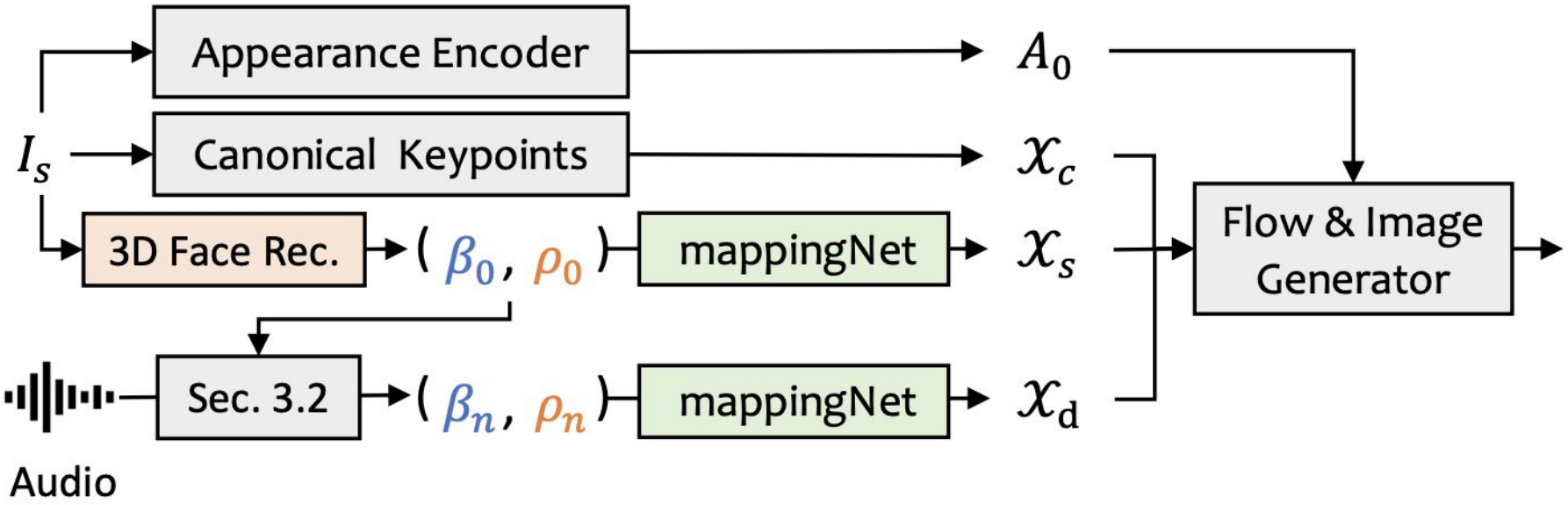
Overall workflow



ExpNet Structure



PoseVAE Pipeline


Appearance Encoder
A_0
I_s
Canonical Keypoints
x_c
3D Face Rec.
(β_0, ρ_0)
mappingNet
x_s
Flow & Image Generator
Sec. 3.2
(β_n, ρ_n)
mappingNet
x_d
Audio

FaceRender

What are the problems with this approach?

- **The input images should have a neutral expression.**

- **When the subject has some other expression, the facial landmarks are unable to register properly by the model.**

- **For example, a person smiling in the photo would produce a poor facial expression in the video.**

Smiling Expression

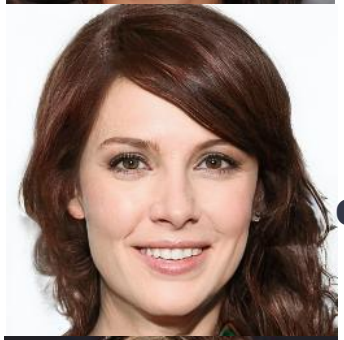
Original Image

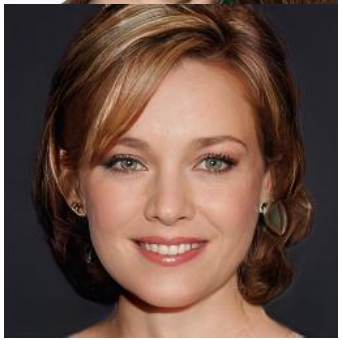
After Processing

How do we solve this?

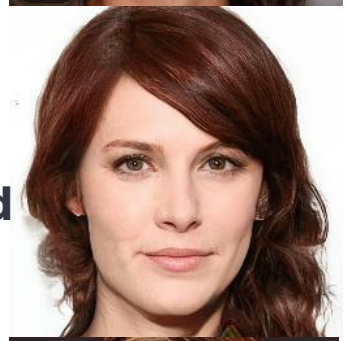

StyleGAN

StyleGAN
CelebA 64x64





revise latent codes, for expression and lips



Better video on SadTalker with natural Expression

- **Extracting expressive image’s latent code with the appropriate StyleGAN encoder.**

- **Manipulating latent vector to reduce expression, then generating the neutral image.**

***Future scope**

THANK YOU!!

PLZ VOTE :)