

Generative Adversarial Networks for audio-visual forms

Dillig Marlene (MIM)

Felger Max (DIM)

Père Valentin (IMT Mines Albi)

Weiß Markus (MIM)

Betreuer: Prof. Dr. Ruxandra Lasowski, Prof. Dr. Norbert Schnell

Overview

- Research assignment
- Generative Adversarial Networks
- Audio-visual GAN
- Data collection
- Conclusion

Research assignment - AVGAN

Audio-visual Generative Adversarial Network



Research assignment - Data collection

make-a-face-for-science-test.web.app

MAKE A FACE



FOR SCIENCE

Generative Adversarial Networks

GAN

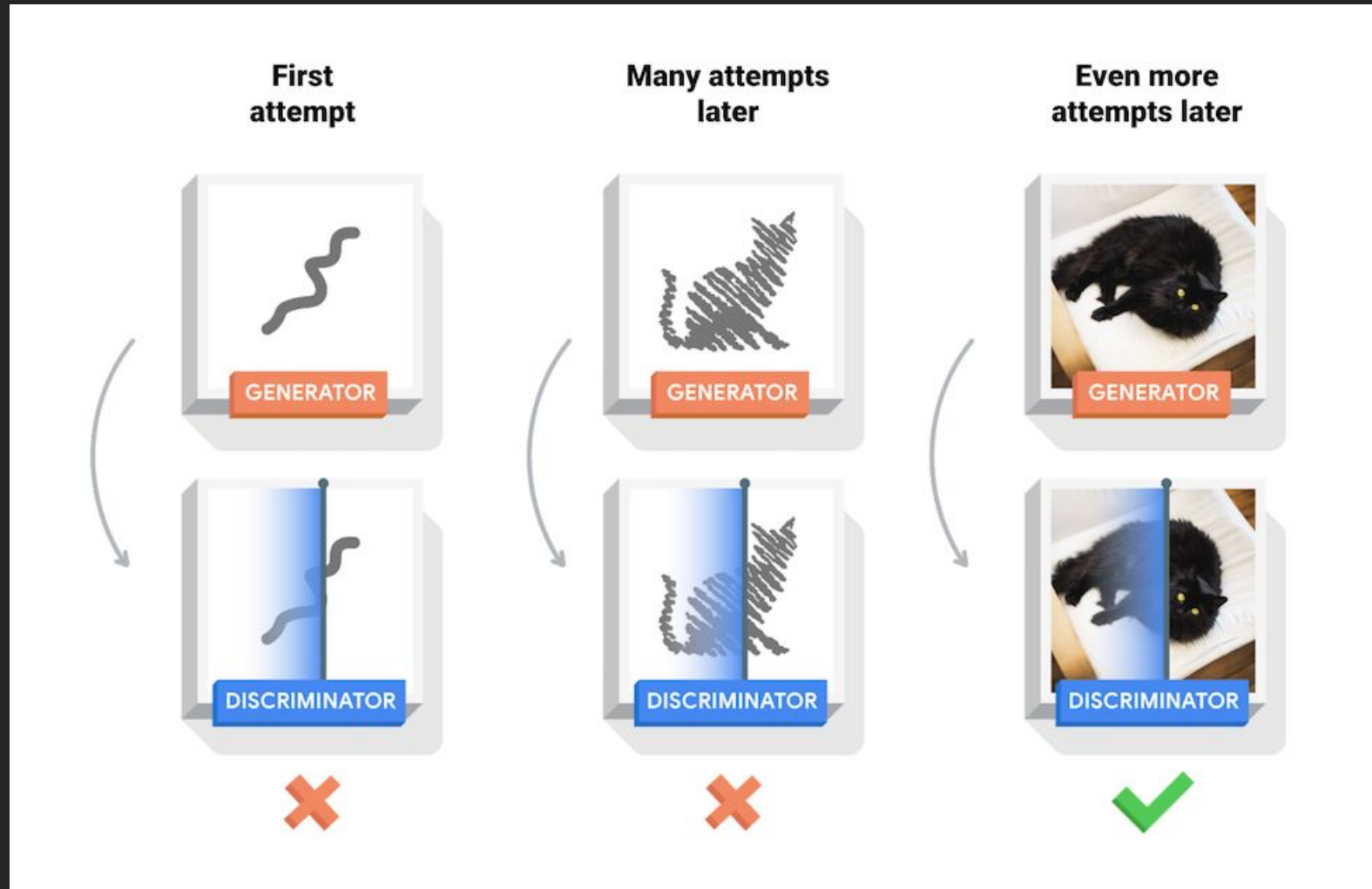
- Let a machine generate new text images and a sounds

3)



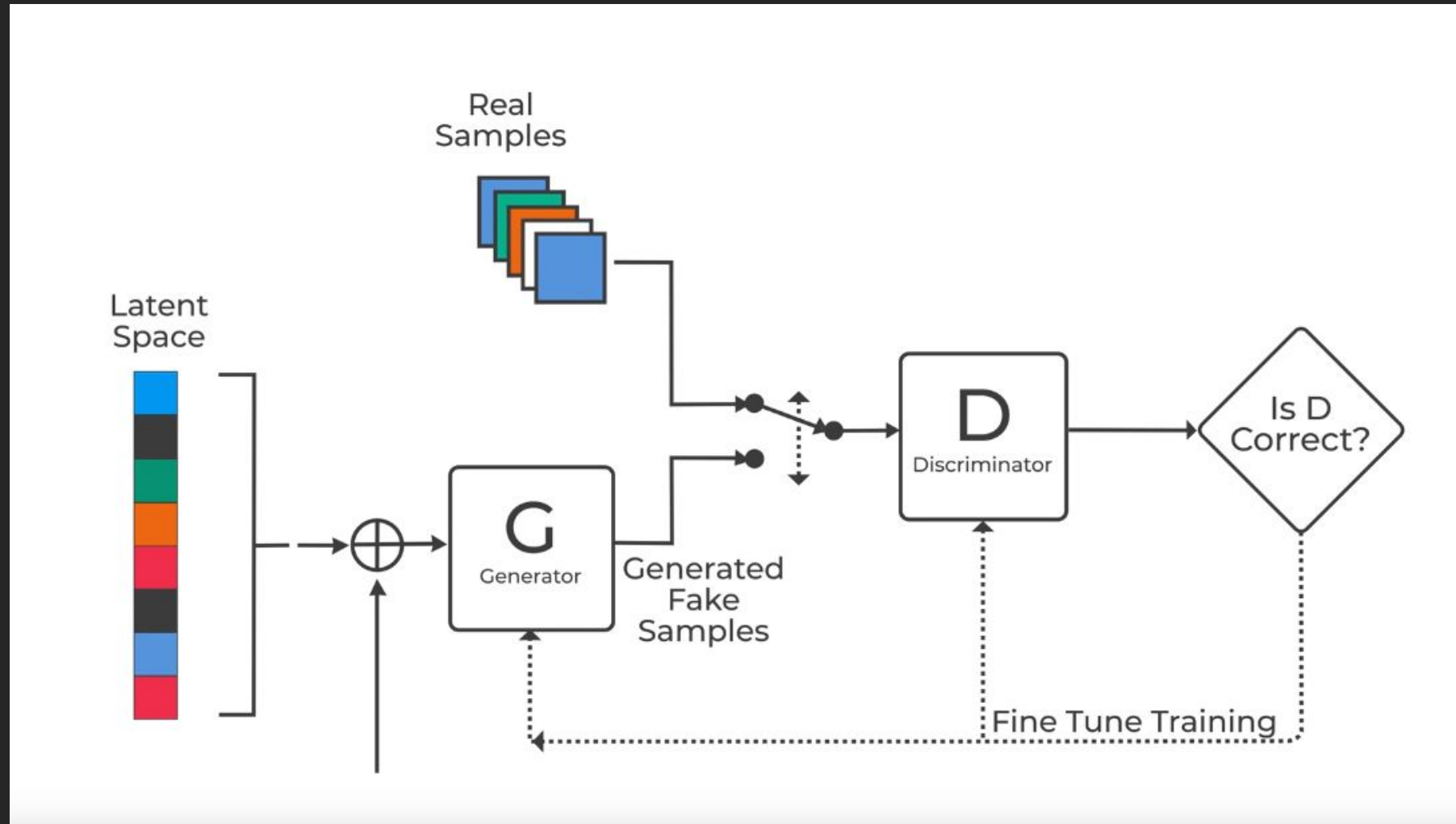
GAN - Generator vs. Discriminator

4)



GAN - Architecture

5)

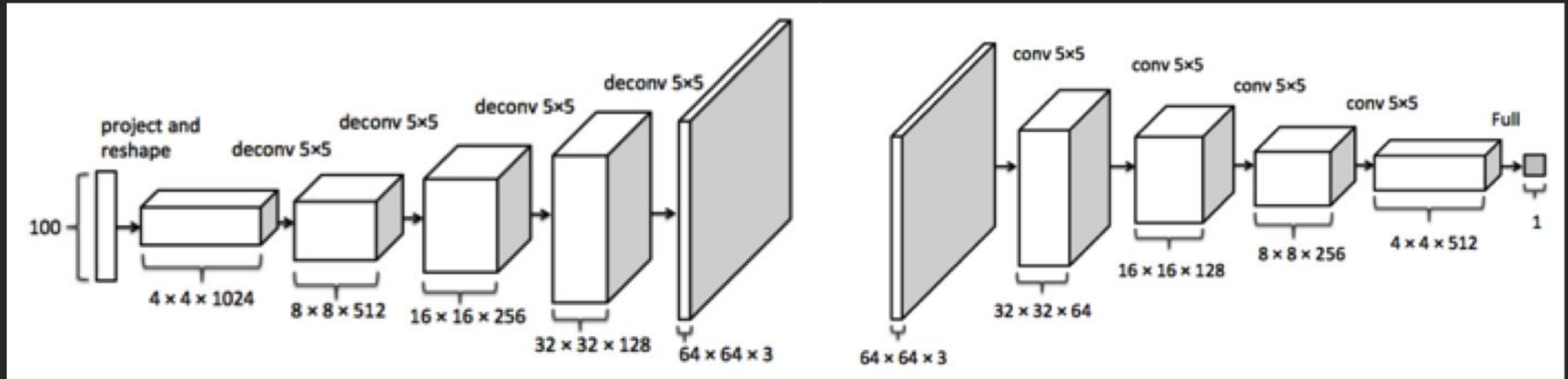


Architecture of a GAN

Generator

Discriminator

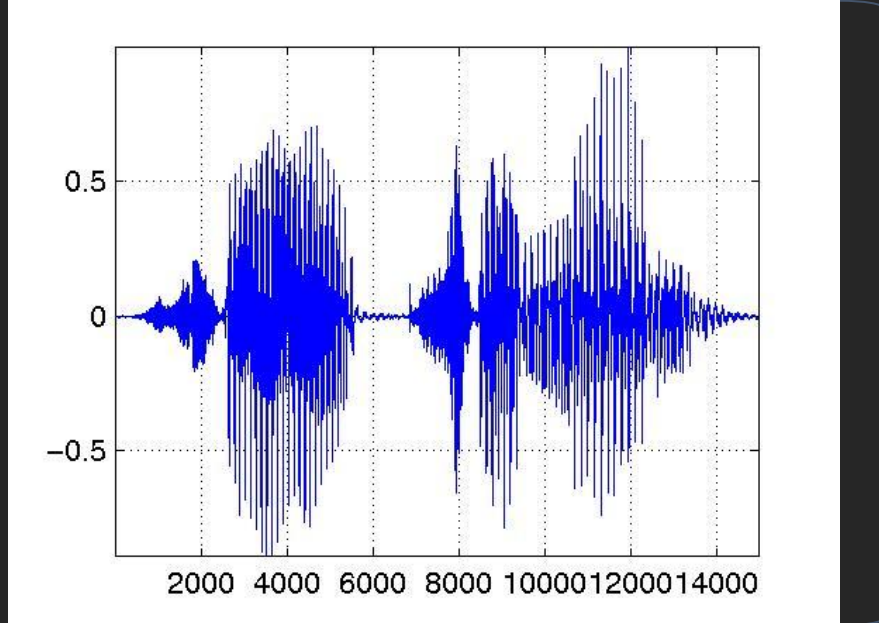
6)



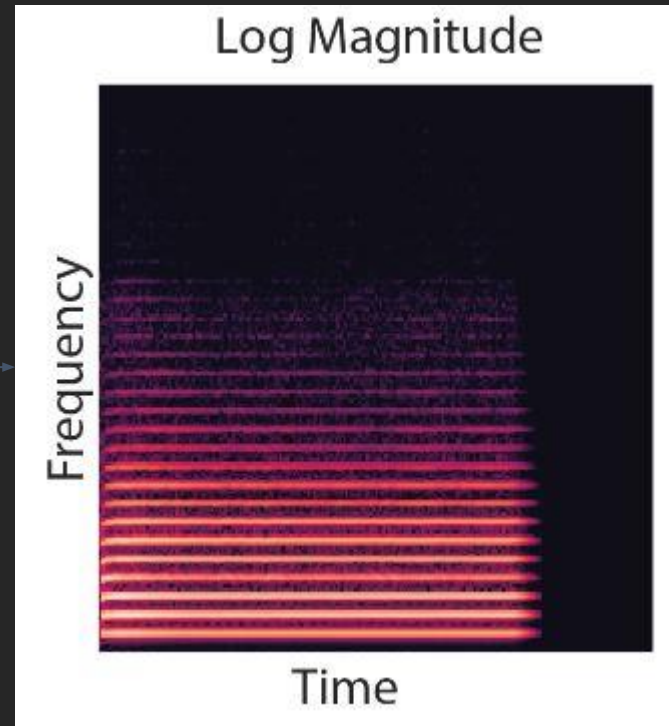
Audio-visual GAN

Wave transform

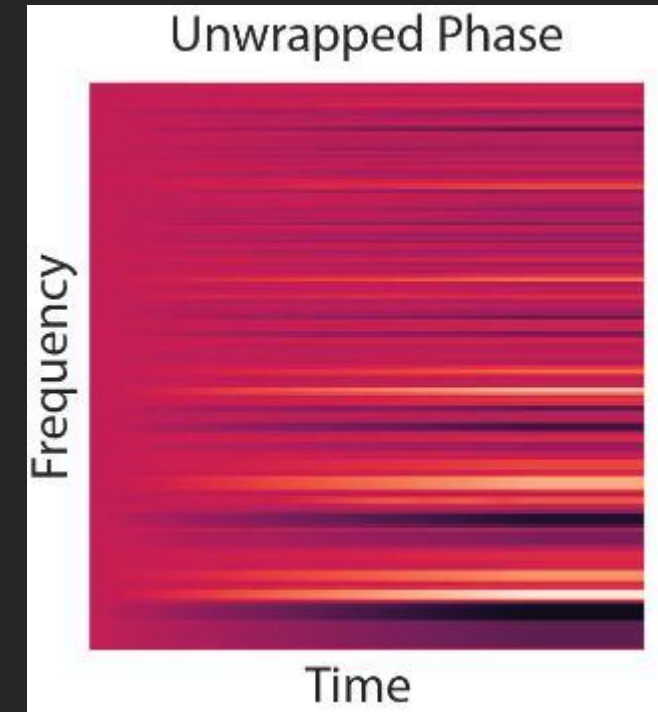
7)



8)

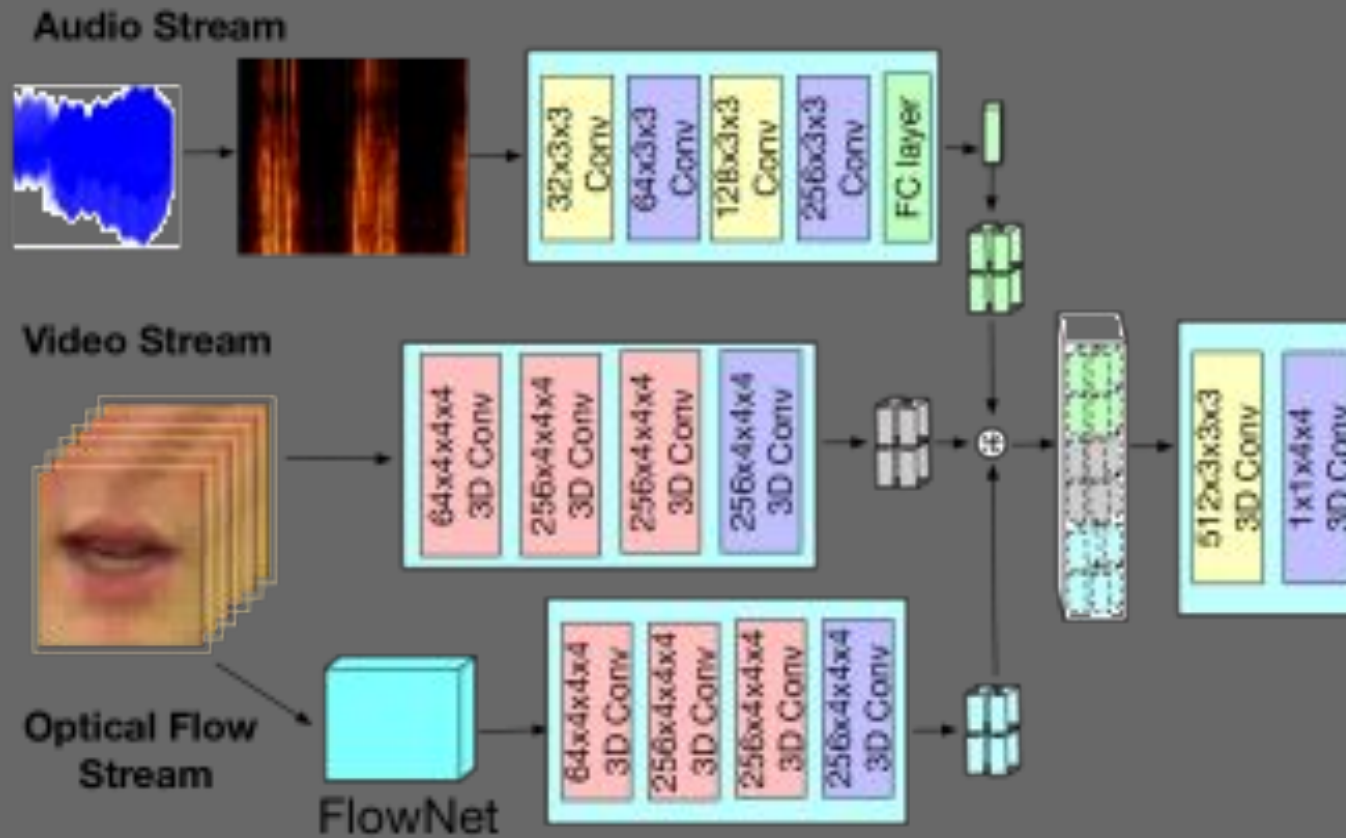


8)

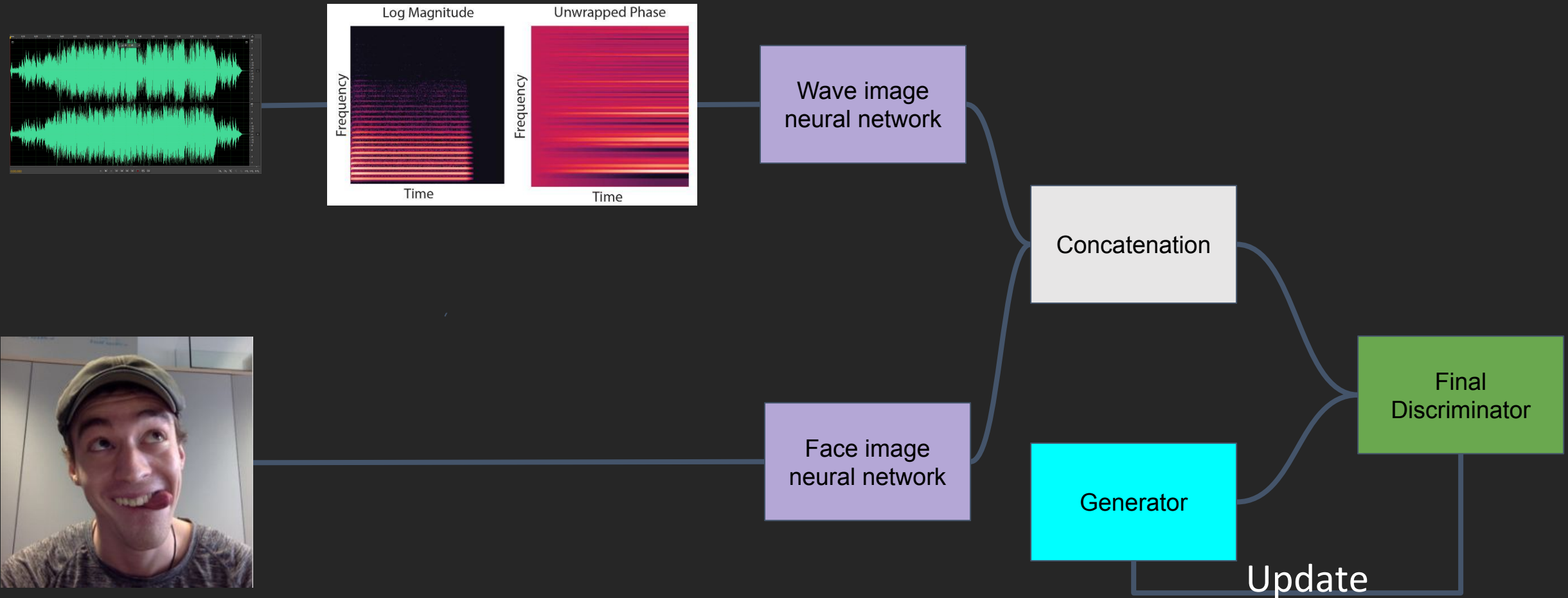


Lip Movement GAN (2018)

9)



Architecture



Data Collection

Website

- Users can record faces (image and audio)
- They can also rate other contributions
- The website should be:
 - fully responsive
 - platform independent
 - easy to use

Website

- Frameworks:
 - jQuery
 - Bootstrap
- APIs & SDKs:
 - Cloud Storage
 - Cloud Firestore
 - Web Audio API

Website

- Hosting: Firebase (Spark Plan)
 - Created as a startup
 - Grew into a next-generation app-development platform on Google Cloud Platform
- Storage: Cloud Storage for Firebase
 - Simple and cost-effective object storage service by Google

Website

- Database: Cloud Firestore
 - Supports flexible, hierarchical data structure (Collections > Documents > Data)
 - Uses data synchronization to update data on any connected device

File format

- Image:
 - JPG
 - 320x320 px
 - High quality
 - Avg. file size: 110 KB (depending on device)

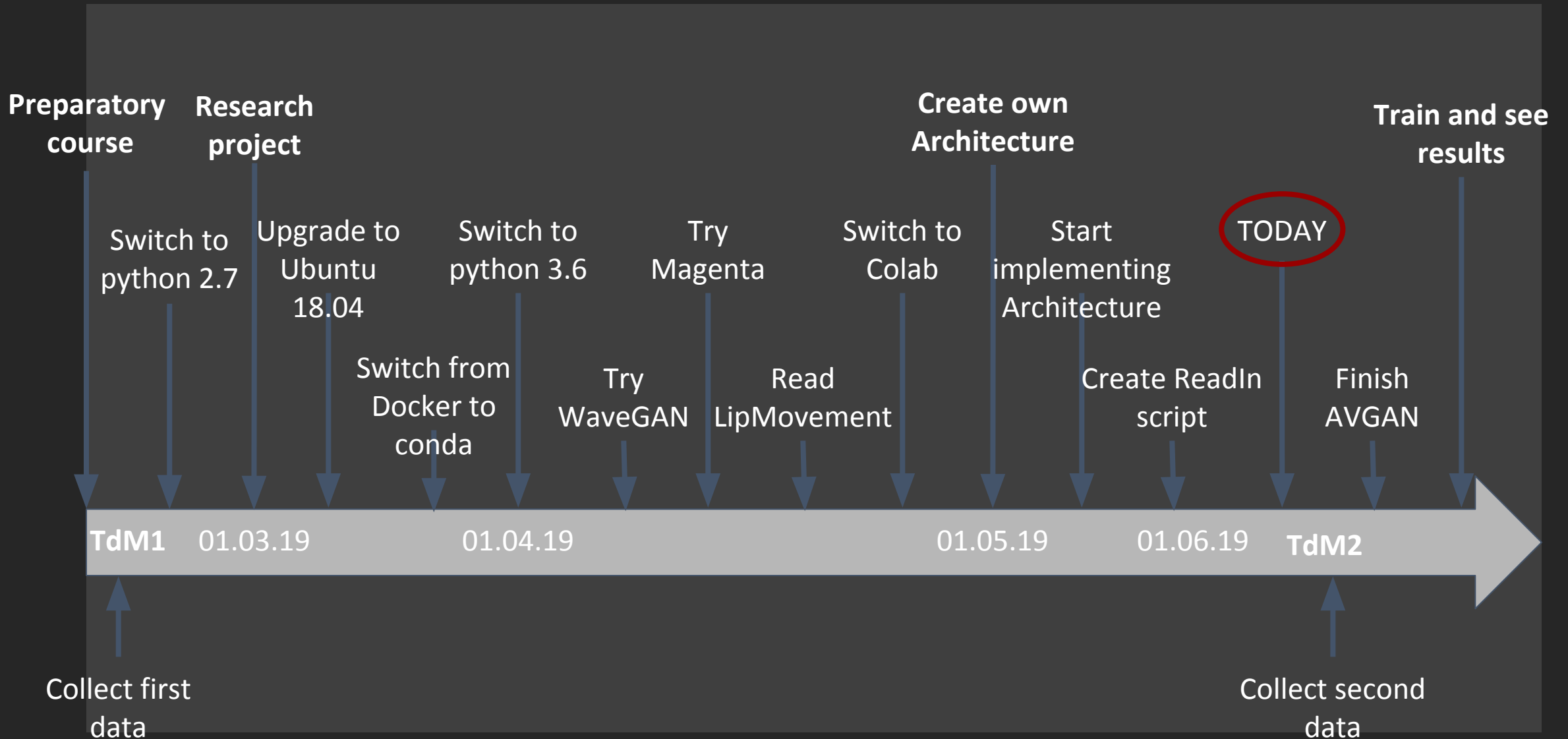
File format

- Audio:
 - WAV
 - 1 sec
 - Avg. file size: 90 KB (depending on device)

make-a-face-for-science-test.web.app

Conclusion

Progression and future work



Thank you!

You may ask questions now.

References

Introduction

- 1) <https://de.clipdealer.com/photo/media/A:59226972>
- 2) <https://arxiv.org/abs/1812.04948>

DCGAN

- 3) <https://github.com/tensorflow/docs/blob/master/site/en/r2/tutorials/generative/dcgan.ipynb>
- 4) <https://github.com/tensorflow/docs/blob/master/site/en/r2/tutorials/generative/dcgan.ipynb>
- 5) <https://hub.packtpub.com/what-are-generative-adversarial-networks-gans-and-how-do-they-work/>
- 6) http://www.timzhangyuxuan.com/project_dcgan/

Wave transform

- 7) <https://arxiv.org/pdf/1902.08710.pdf>
- 8) <http://soundfile.sapp.org/doc/WaveFormat/>

Lip Movement

- 9) https://www.groundai.com/media/arxiv_projects/191092/x5.png.750x0_q75_crop.png

References

- 1) <https://de.clipdealer.com/photo/media/A:59226972>
- 2) <https://arxiv.org/abs/1812.04948>
- 3) <https://hub.packtpub.com/what-are-generative-adversarial-networks-gans-and-how-do-they-work/>
- 4) https://www.groundai.com/media/arxiv_projects/191092/x5.png.750x0_q75_crop.png