Mouth Animation

Action Plan

Bachelor Applied Computer Science

Yori Verbist

Academic year 2020-2021

Campus Geel, Kleinhoefstraat 4, BE-2440 Geel

LID VAN
ASSOCIATIE
KU LEUVEN

THOMAS
MORE

Contents

1 Introduction

This is my action plan that I made to get a better view what is expected of me during my internship. The goal of my internship is to let the mouth of a person in a picture or video move on the basis of an audio sample. This is done by artificially generating mouth movements on top of the image or frame.

I will start by introducing the company where I'm doing my internship. After this I'll give some more information about the goal of this internship, the different stages of this project and a last how I'll be doing my reporting.

2 Company

Brainjar is a small company of ±10 employees that is mainly focused in Artificial Intelligence located in Leuven, Belgium. They mostly do AI consulting and make end-to-end Machine Learning applications.
Brainjar is part of the Raccoons group which is part of the Cronos Group.
Some examples of their previous works are: for MOW Vlaanderen they implemented deep learning for automated crack detection in bridges using drones. They also have wide range of different internships in AI. These internships range from CV to NLP.

brainjar

3 Motive and Background

The motive of this internship is to show clients what is possible with AI. These showcases are mostly projects that most people working in AI don't have time to do. So it's perfect to give these projects to interns so they get more experience with AI and the company gets more examples of the possibilities of AI.
I chose this particular internship because I'm mostly interested in the Computer Vision part of deep learning. It's also a topic that requires more research than most topics, which is more challenging and appealing to me. This way I can figure out if a research heavy topic is my cup of tea, or not, because I'm considering a follow-up Masters degree after I completed my Bachelors education.

4 Goal

The goal of this project is to make a generative model that can generate a mouth on a face when it's given a audio file and a picture/video. I'll be building on top of an existing project(Wav2Lip[2]) where Ill be adding some extra functionalities. Since this project is build by a research group that has multiple years of experience in this field it would be dumb to start from scratch. What these extra functionalities are, you can find in the section below, Milestones.

5 Milestones

The following milestones are ordered in chronological order in which they'll be implemented.

5.1 Dutch Data

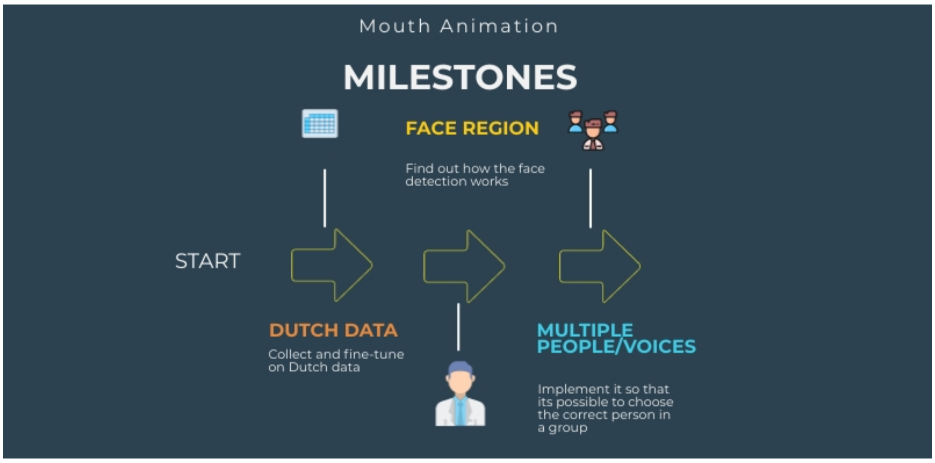
The first milestone is to fine-tune the model on Dutch data. Since it's trained on English data it will not always work as well on Dutch audio files. This is a tricky part since the English data is not publicly available and the data is also heavily modified. I'll be looking deeper in how they modified their data and if it's possible to do this in the amount of time I have.

5.2 Face Region

One of the weak points of the original project is that the chin is not always included in the generated region of the face. This results in a moving mouth but a chin that's just standing still. My goal is to adjust this part of the project so the chin is always included in the generated region. Wav2Lip uses the Face-Alignment[1] project which I'll adjust to make this possible.

5.3 Multiple people/voices

Now it's only possible to generate the mouth when there is one person in the frame. The goal is to make it possible so it generates the mouth of the correct person when there are multiple people in the frame. If there is time left it's also possible to add a functionality to use an audio file where multiple people are talking and that the audio splits on the correct persons. So you can identify which voice corresponds with which person in the frames.



Mouth Animation

MILESTONES

FACE REGION

Find out how the face detection works

START

DUTCH DATA

Collect and fine-tune on Dutch data

MULTIPLE PEOPLE/VOICES

Implement it so that its possible to choose the correct person in a group

6 Reporting

Every week there will be a stand-up meeting with my mentors. Also at the end of every week there will be a meeting in which each intern shows what he has done that week. I'll be using Confluence to document everything since it's very easy to use.
There will also be two meeting which the supervising tutor of the school to follow up on how I'm doing. The first one will be to present this action plan, while the second one will be to show how much I have accomplished until then. The first meeting will be held on week 4, the second on week 8.

At the end of my internship my mentors will be evaluating my performance on different criteria.

References

[1] Adrian Bulat and Georgios Tzimiropoulos. How far are we from solving the 2d & 3d face alignment problem? (and a dataset of 230,000 3d facial landmarks). In *International Conference on Computer Vision*, 2017.

[2] K R Prajwal, Rudrabha Mukhopadhyay, Vinay P. Namboodiri, and C.V. Jawahar. A lip sync expert is all you need for speech to lip generation in the wild. In *Proceedings of the 28th ACM International Conference on Multimedia*, MM '20, page 484–492, New York, NY, USA, 2020. Association for Computing Machinery.