avatars4all

Eyal Gruss
<3
Tel-Aviv, Israel
eyalgruss@gmail.com

Abstract

We present an environment [11] for running First Order Motion Model [13] using a live webcam feed in the browser over Google Colaboratory. This allows novice users to experience almost real-time live head puppeteering, or so called "deep fake avatars", with no need of dedicated hardware, software installation or technical know how. A rich GUI allows extensive control of model and media options, as well as some unique innovations including fast auto-calibration and a Muppets generator [9]. This, and other accompanying notebooks, serve in practice as educational, creative and activist tools.

1 Main

With the advance of the coronavirus pandemic in the beginning of 2020, The majority of human social activity has been forced online to the virtual realm. Only a few months earlier, First Order Motion Model (FOMM) [13] was released, introducing the ability of one-shot video-driven image animation. This was soon followed by [10], a real-time environment for FOMM allowing using "Avatars for Zoom, Skype and other video-conferencing apps". Is the time ripe to claim the once promised cybernetic utopia? Could we at last shed our physical shells and be whoever we want to be in Zoom-space?

> People come to the Oasis for all the things they can do, but they stay because of all the things they can be: tall, beautiful, scary, a different sex, a different species, live action, cartoon, it's all your call. [Ready Player One film, 2018]

The repository [11] contains a few Colab notebooks that attempt to make the technology accessible for all. Requiring only a browser and a Google account, these notebooks can be operated with one click ("run all"). However, they are also flexible tools, allowing users to use and manipulate their own selected media. The live webcam environment is based on WebSocket similar to [8]. To the author's best knowledge, it is the fastest purely online solution for live FOMM avatars, as well as one of the very few real-time webcam Colab implementations. The GUI in figure 1 shows a multitude of controls for zooming, calibration, switching between avatars, generating new avatars, and various model and display parameters.

A novel fast auto-calibration mode that works in real-time, finds the best alignment between driver and avatar based on model keypoints (rather than facial landmarks). Following Avatarify [10], which inspired this project, the user can generate new avatars based on StyleGAN "This Person Does Not Exist" [6] website. Taking the idea further, one can also generate avatars specifically of men, women, boys, girls [2], Waifus [7], Fursonas [3] and Muppets [9], the latter developed especially for this project by Doron Adler, in collaboration with the author. One can also drag and drop local or web images on the GUI to upload new avatars, as inspired by [1]. Other innovations include an exaggeration factor slider to lever stronger keypoint motions, an option to take your own snapshot and puppeteer it, reminiscent to Nvidia Maxine [4], which may help understanding the mechanism, an optional post-process step for the pipeline for offline videos, using Wav2Lip [12] following FOMM,
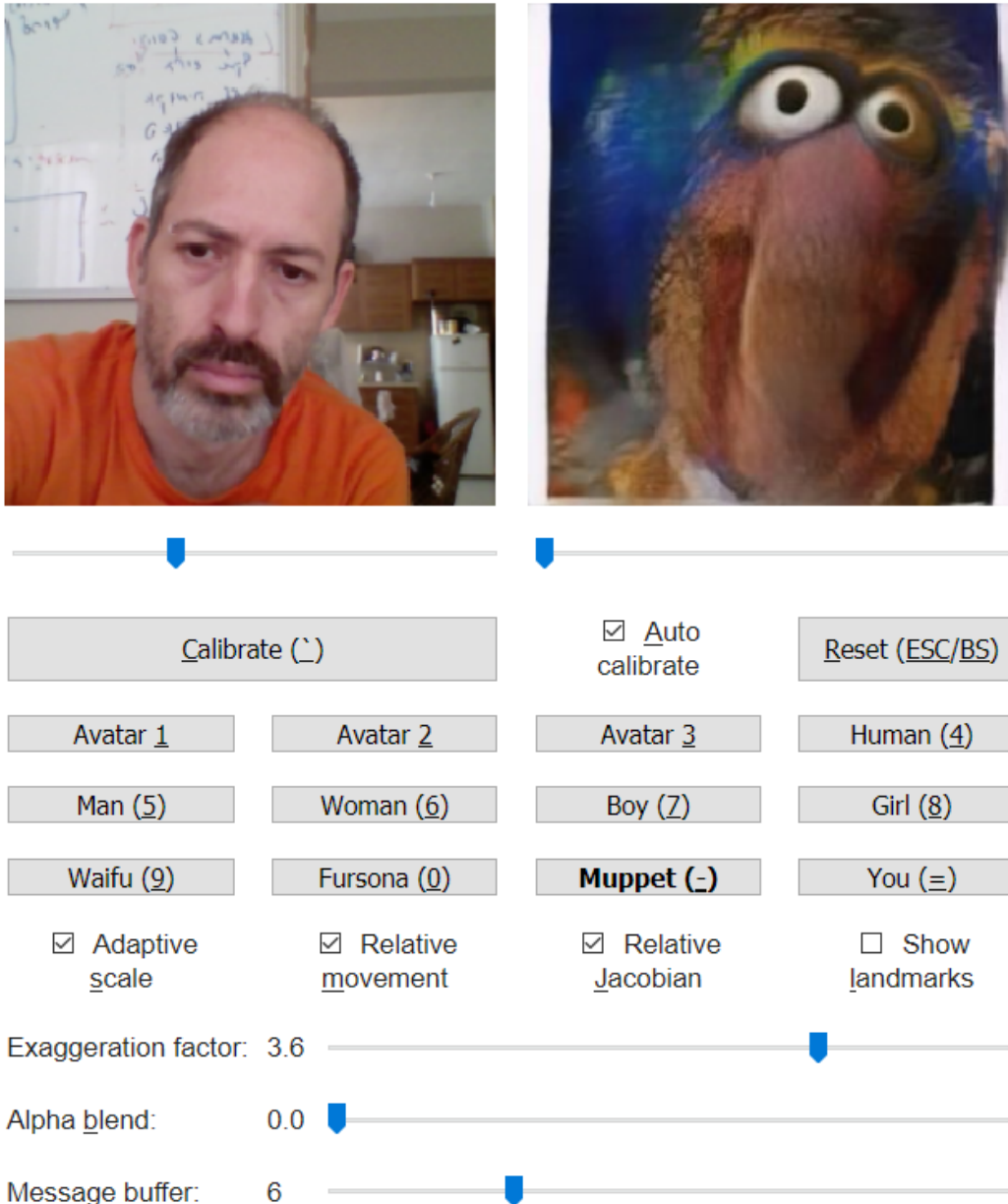
payload=11756 fps=12.3

Figure 1: GUI for live webcam avatar in Colab. The author (left) is puppeteering a generated Muppet.

to fix the lip sync, and combining Wav2Lip with speaker diarization for automatic animated skit creation from audio ("Wav2Skit").

These tools were the basis for several workshops and tutorials at international festivals and conferences in 2020, including Suoja/Shelter, South Africa NAF, ADAF, Reclaim Futures, Fubar, ISEA, Technarte, EVA London, Piksel, Stuttgarter Filmwinter, Dorot-Con, MozFest, HAI-GEN and ICCC [11]. They are now being introduced in elementary and middle schools in Israel with the Pisga-Cyber excellence program [5]. A pleasantly surprising first real-world usage of the described system.

2

Broader impact and ethical implications

This is a dangerous time. The ability to synthesize and manipulate media is improving by the day: in the quality of outcome, in the mediums, modalities and conditions dealt with, in the required computational and data resources, and in the availability and accessibility of the technology. We are in the midst of a transition period, where these facilities are still accessible mostly to the tech savvy or to those with the means to hire them. It may not be long before we have ubiquitous and seamless smartphone apps that can create perfect deep fakes. However, it is the author's opinion that precisely in this interim, it is imperative to liberate and democratize the technology.

The advancement of technology cannot be stopped. AI and synthetic media, like electricity, fire and other technologies, can be used for good and for bad. It can be used both to infringe one's privacy and to protect one's privacy. It can be used to bully and to harass, or to promote self expression and self acceptance. Fake news is not a new problem. Blood libels have existed throughout the last millennia. Photomontage technology has been used to fake photographs as early as 1857. Videos are harder to fake, but Hollywood, Disney and government agencies have been doing so for the last century. Contemporary examples show that it is enough to change the label on an image, or slightly edit an audiovisual recording, to achieve a strong effect. The solution to combat this is in education. Making the technology accessible to educators, artists, journalists as well as the general public, will serve to raise awareness, healthy skepticism and critical thinking, toward media and the spectrum of contemporary possibilities in media creation and manipulation.

References

[1] URL `https://terryky.github.io/tfjs_webgl_app/face_landmark`.

[2] URL `https://fakeface.rest`.

[3] URL `https://thisfursonadoesnotexist.com`.

[4] URL `https://developer.nvidia.com/MAXINE`.

[5] URL `https://pisgacyber.co.il`.

[6] URL `https://thispersondoesnotexist.com`.

[7] URL `https://www.thiswaifudoesnotexist.net`.

[8] a2kiti. Webcam google colab, 2020. URL `https://github.com/a2kiti/webCamGoogleColab`.

[9] D. Adler and E. Gruss. This muppet does not exist, 2020. URL `https://thismuppetdoesnotexist.com`.

[10] A. Aliev and K. Iskakov. Avatarify, 2020. URL `https://github.com/alievk/avatarify`.

[11] E. Gruss. avatars4all, 2020. URL `https://github.com/eyaler/avatars4all`.

[12] K. R. Prajwal, R. Mukhopadhyay, V. Namboodiri, and C. V. Jawahar. A lip sync expert is all you need for speech to lip generation in the wild, 2020. URL `http://bhaasha.iiit.ac.in/lipsync`.

[13] A. Siarohin, S. Lathuilière, S. Tulyakov, E. Ricci, and N. Sebe. First order motion model for image animation. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 7137–7147. Curran Associates, Inc., 2019. URL `https://aliaksandrsiarohin.github.io/first-order-model-website`.