

FOMM

そもそも、FOMMとは？

自分の顔に合わせて目、口、顔の動きや向きまでも同期できるようなモデルのことである。

正式名称は**First Order Motion Model**。

FOMM

まずは、フリーで拾ってきた動画を使って、モナリザを動かしてみた。（GoogleColabを使用）（動画）



FOMM

次に、リアルタイムで自分の顔を検知してモナリザを動かしてみた。

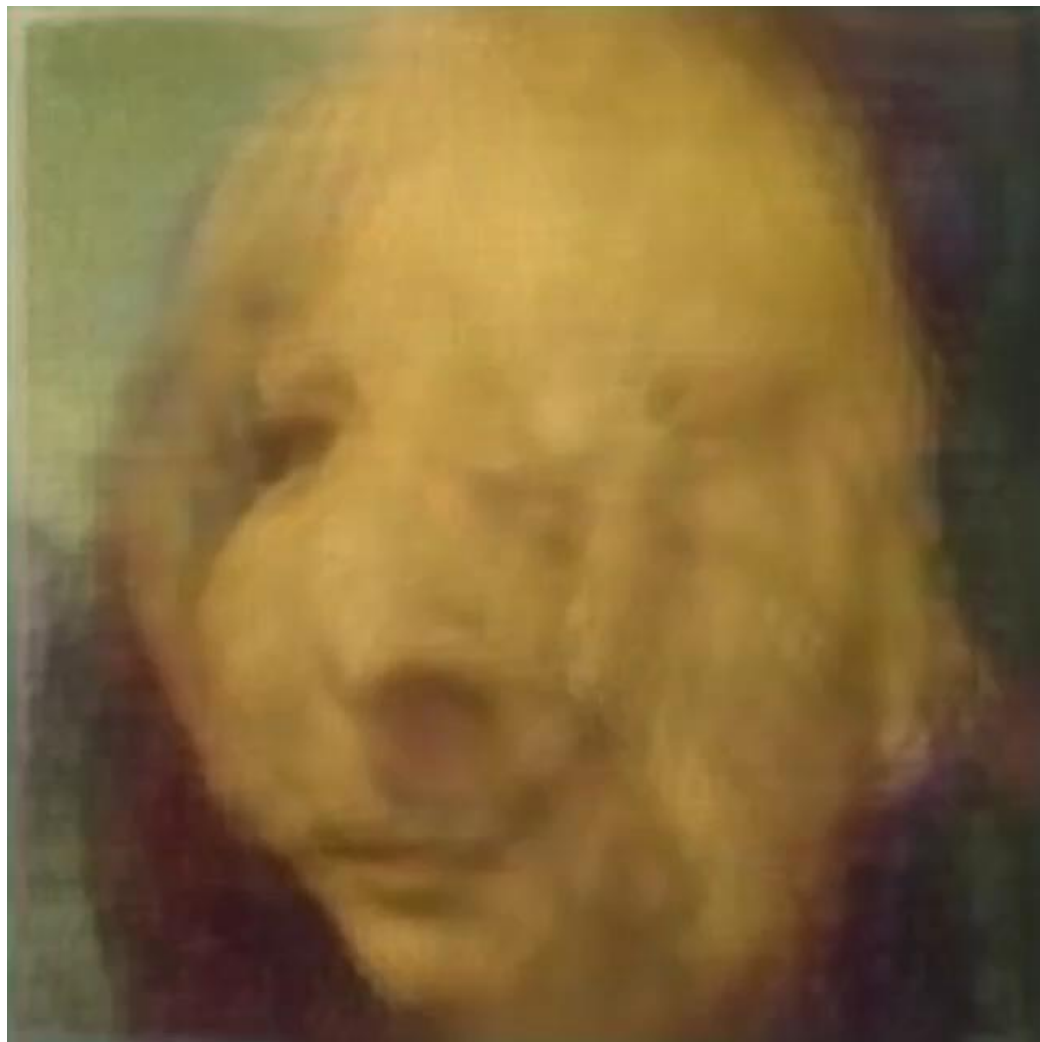
FOMMはかなり重い上に、リアルタイムで動かしたいのでGPUを使わないとだめだと考えた。（CPUで動かしていない）

FOMMのディープラーニングのところはGPU(WSL)で動かして、OpenCVでWebカメラをキャプチャするところはCPU(Windows)で動かした。その間のやり取りはWebSocketを使った。

環境構築のやり方が調べてもわからなかったためこのようなめんどくさいやり方をした。

FOMM

結果は以下の通り。（動画）
怖い！！！！
生成AIは難しい！！！！
4月に文書生成にチャレンジ
したが、その時も失敗。



FOMM

GoogleColabで実際に扱うコードがまともに動くことはわかっていたのに...

このような結果になった要因として考えられるのは

- ・ Webカメラの性能が悪い
- ・ FOMMは新しいモデルではない(少数のフレームに対応)

本質的な要因はFOMMについてしっかり理解していないためわからない。今後、勉強したいと思う。

(自分の顔をWebカメラで映して、その動画を使ってみるべきだった。)

しかし、目や口の動きは一応表現できている？ので思ったようなものは作れなかったが、全くダメではなかった。