MACQUARIE
University
SYDNEY·AUSTRALIA

Progress Report

Personalized Facial Expression Imitation

Peizhen Li

Dec 19, 2024

Outline

- **Work completed in the past four weeks**
  1. Dataset collection
     - results from intern students
     - dataset statistics
     - control values rollout: high-level vs low-level
  2. Coding and experiments in virtual environment
  3. Mapping network training and toy experiments on physical Ameca
- **Expected outcomes in the following four weeks**
  1. Current issues and possible solutions
     - ineffectiveness of the mapping network
     - insufficient action space exploration/coverage
     - real-time requirement
  2. Draft for IROS25 submission

Dataset Collection

- Results from intern students
- Dataset statistics

📈 **Dataset Statistics**

Statistics	Train	Validation
# of images	9341	2336
# of control sequences	9341	2336
# of (image, control sequence)	9341	2336

```
Jaw Pitch: 0.524652
Jaw Yaw: 0.6763
Lip Bottom Curl: 0.670025
Lip Bottom Depress Left: 0.541817
Lip Bottom Depress Middle: 0.6158985
Lip Bottom Depress Right: 0.533704
Lip Corner Raise Left: 0.446363
Lip Corner Raise Right: 0.525653
Lip Corner Stretch Left: 0.431248
Lip Corner Stretch Right: 0.327669
Lip Top Curl: 0.3929235
```

Dataset Collection

- High-level vs low-level controls
  - Tune high-level controls in the animator tool
  - Convert to low-level controls to better explore the action space (fine-grained control)
- Matrix formulation

$$\begin{pmatrix} H_1, & H_2, & \cdots, & H_{12} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{12} \end{pmatrix} = x_1 H_1 + x_2 H_2 + \cdots + x_{12} H_{12} = H_{\Delta}$$

$$H_i \in \mathbb{R}^{13 \times 1}, \quad x_1 + x_2 + \cdots + x_{12} = 1, \quad i \in \{1, 2, \cdots, 12\}$$

Coding and Experiments in Virtual Environment

- Scripts for dataset curation and model training

```
$ python misc/construct_dataset.py
$ python main.py train.batch_size=16
```

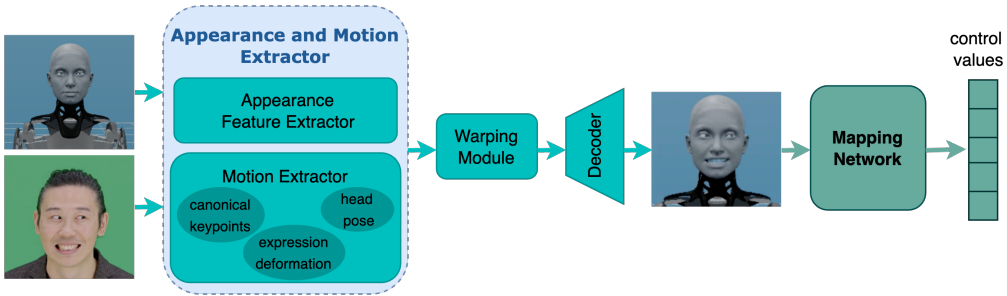
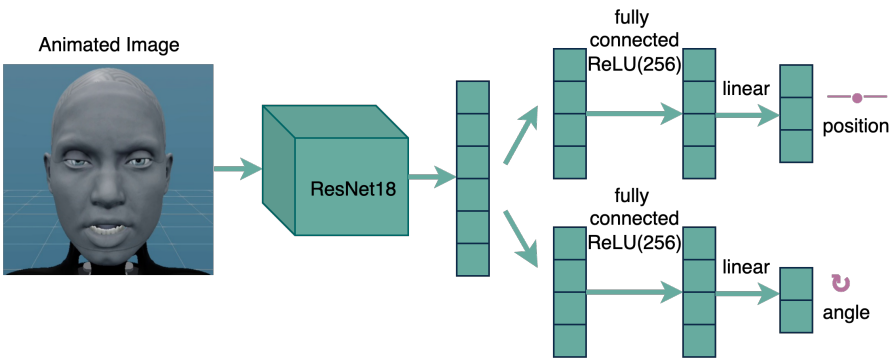


Figure: An overview of the proposed facial expression imitation framework.

Mapping Network Training and Toy Experiments

- From image space to action space
- Output a list of 30 low-level control values



Mapping Network Training and Toy Experiments

- **Goal:** Generate robotic facial expressions that mimic those of a human performer.

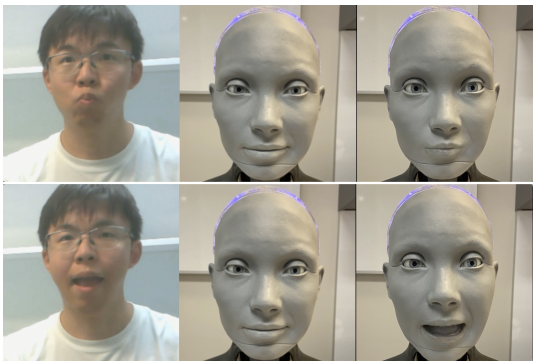


Figure: Left: Human performer. Middle: Ameca in a neutral state. Right: Ameca displaying a facial expression to mimic the human performer.

Expected Outcomes in the following four weeks

- Current Issues:
  - ineffectiveness of the mapping network
  - insufficient action space exploration/coverage
  - real-time requirement
- Possible Solutions
  - sanity check on the dataset (temporal alignment/granularity/dataset scale)
  - inspect control value distribution
  - adjust network architecture
  - predict deviations from neutral state
- Draft for RSS25/IROS submission