MODEL CARD

MediaPipe Hand Gesture Classification

SOLUTION DETAILS

Hand Gesture Classification model uses hand landmarks produced by MediaPipe Hands Model to classify a hand pose as one of the 8 hand gesture classes, namely,

- Closed Fist
- Open Palm
- Pointing Up
- Thumb Down
- Thumb Up
- Victory
- I Love You
- None of the above gestures

SOLUTION SPECIFICATIONS

Solution Architecture

- Two step neural network pipeline with an embedding model followed by a classification model. This pipeline runs on hand landmarks and related information for a single hand, but does not directly process any images (i.e. RGB pixel data).

Inputs

This pipeline consumes MediaPipe Hands model's outputs:

- 21 3-dimensional screen landmarks represented as a 1 x 63 tensor and normalized by image size.
- A float scalar represents the handedness probability of the predicted hand.
- 21 3-dimensional metric scale world landmarks represented as a 1 x 63 tensor and normalized by image size.
- Refer to this model card for more details.

No image data was directly input into the model.

Output(s)

An 8 element vector that predicts the probability of each of the following classes:

- 0th-element: probability that hand pose is not a known hand gesture to the model
- 1st-8th: probability of hand pose is one of the 7 known gestures.

Based on Model Cards for Model Reporting, In Proceedings of FAT* Conference (FAT*2019). ACM, New York, NY, USA, 9 pages

EMBEDDING MODEL SPECIFICATIONS

Model Type

- Fully Connected Neural Network with residual blocks

Model Architecture

- Regression model

Inputs

- 21 3-dimensional screen landmarks represented as a 1 x 63 tensor and normalized by image size.
- A float scalar represents the handedness probability of the predicted hand.
- 21 3-dimensional metric scale world landmarks represented as a 1 x 63 tensor and normalized by image size.

Output(s)

- A float tensor 128x1 embedding tensor of predicted embedding representing the hand landmarks, which is further used in the classification model head, described in the next section.

CLASSIFICATION MODEL SPECIFICATIONS

Model Type

- Fully Connected Neural Network

Model Architecture

- Classification model

Inputs

- A float tensor 128x1 embedding tensor of predicted embedding representing the hand landmarks.

Output(s)

- An 8 element vector that predicts the probability for each of the 8 above mentioned gesture classes.

Intended Uses

APPLICATION

Predict if and what the hand gesture of a given hand's landmark information.

DOMAIN & USERS

Mobile AR (augmented reality) applications
Gesture recognition
Hand control

OUT-OF-SCOPE APPLICATIONS

Not appropriate for:

- Hand gestures involving multiple hands (e.g. two handed heart shape)
- Hand gestures involving motion (e.g. waving goodbye)
- Translate sign language
- Any form of surveillance or identity recognition is explicitly out of scope and not enabled by this technology.