



MODEL CARD

Hand Detection/Tracking



MODEL DETAILS

Two lightweight models (palm detector, hand landmark model) to detect palm and predict hand landmarks within an image on a smartphone. Palm detector returns bounding boxes for each palm and hand landmark model predicts [keypoints](#) for each hand from cropped image.



MODEL SPECIFICATIONS

Model Type:

Convolutional Neural Network

Model Architecture:

- Palm detector: [SSD](#)-like with a custom encoder
- Hand landmark model: regression model

Input(s):

- Image (possibly a video frame) resized to 256×256 pixels.

Output(s):

- Palm bounding box coordinates with 7 Palm 2D keypoints ([illustration](#))
- 21 3D hand keypoints
- Detection confidence score



AUTHORS

Fan Zhang (zhafang@google.com)
Valentin Bazarevsky (bazarevsky@google.com)



CITATION

Citation here



DOCUMENTATION

- [Google AI blog post 19 Aug 2019](#)
- Example usage included as part of the open source MediaPipe example documentation hosted at <http://github.com/google/mediapipe>



LICENSED UNDER

[Apache License, Version 2.0](#)

Intended Uses



APPLICATION

Detecting prominently displayed hands within images or videos captured by a smartphone camera. This front-facing camera model focuses on relatively large hands.



USERS

Mobile AR (augmented reality) applications.



NOTES

Not appropriate for:

- Counting the number of hands in a crowd
- Detecting hands with gloves or occlusions
- Detecting hands too far away from the camera (e.g. further than 2 meters)

Limitations



TRAINING

The models have been trained on limited dataset and are meant for experimental usage.



PERFORMANCE

The models have not been tested in “in-the-wild” smartphone camera conditions, including low-end devices, low light, motion blur etc., that can affect performance.

Ethical Considerations



DATA

This model was trained and evaluated on images of people using a mobile AR application captured with smartphone cameras in various “in-the-wild” conditions.



HUMAN LIFE

The model is not intended for human life-critical decisions. The primary intended application is for research and entertainment purposes.



AUTOMATION BIAS

Educate end users on model limitations to discourage overreliance (‘automation bias’).

Key Principles*

*Based on [Google's AI Principles](#) that are most relevant to this model.

Potential Research & Social Benefits

- Useful for finding relevant images/video frames in content that is laborious to sort manually
- Could potentially be extended for human accessibility applications like gesture detection

Bias

- As with many human sensing tools, performance may vary across skin tones, gender, age, and potentially other sensitive demographic characteristics.
- The current experimental version of the model uses a preliminary dataset that is set to be expanded in the future to address diversity more thoroughly. It can be used for building experimental setups but is not ready for being published in wide-audience consumer-facing apps until a newer version of the model is released. Google acknowledges that there is a potential for bias due to the limited nature of the dataset but takes no responsibility for it in this initial version of the model.
- Users should remain alert to fairness concerns in use, and in particular, the potential impact of false negatives (failure to detect a hand) on individuals or groups.

Safety

- Consider any potential safety implications of false-positives (detecting a hand where none is present) and/or false negatives (failure to detect a face) for the use case.
- This model is not appropriate for safety-critical applications.

Privacy

- Hand detection is not hand recognition; the model does not identify the owner of the hand or store a uniquely identifying hand representation.
- Users should remain alert to privacy risks; hand detection combined with metadata associated with the image/video could expose personally identifying information.

Definitions

AUGMENTED REALITY (AR)

Augmented reality, a technology that superimposes a computer-generated image on a user's view of the real world, thus providing a composite view.

KEYPOINTS

Hand "keypoints" or "landmarks" are (x, y, z) coordinate locations of hand features.