Practical Gaze Tracking on Any Surface with Your Phone

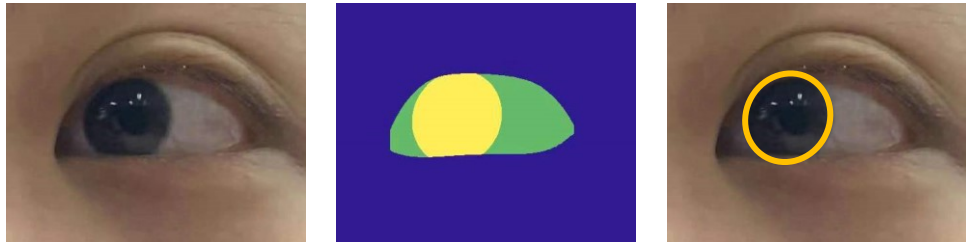
Jiani Cao[1], Jiesong Chen[1], Chengdong Lin[1], Yang Liu[2], Kun Wang[1], Zhenjiang Li[1]

City University of Hong Kong[1], University of Cambridge[2]

Outside the conference version...

Training Dataset Label

Raw image Label Iris boundary

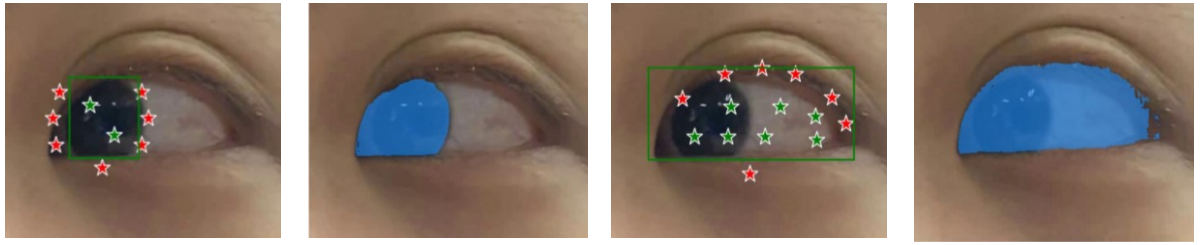


Annotate Train

- Label quality directly determines results.
- Manual labelling is time-consuming.

Foundation model?
Automatic?

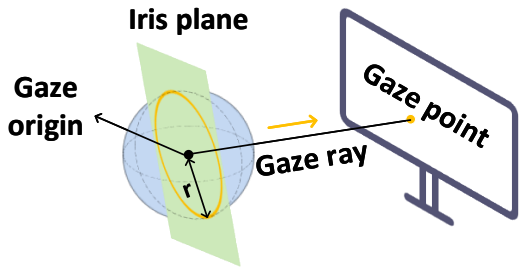
Using segment anything model (SAM)



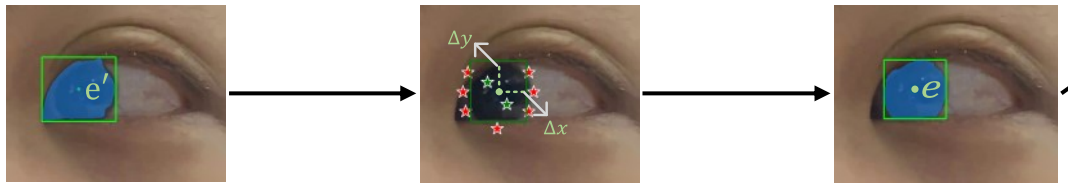
Prompt Prompt

- SAM can generally recognize the iris and sclera.
- However, the edge is quite unsmooth.
- The prompt needs to be provided manually.

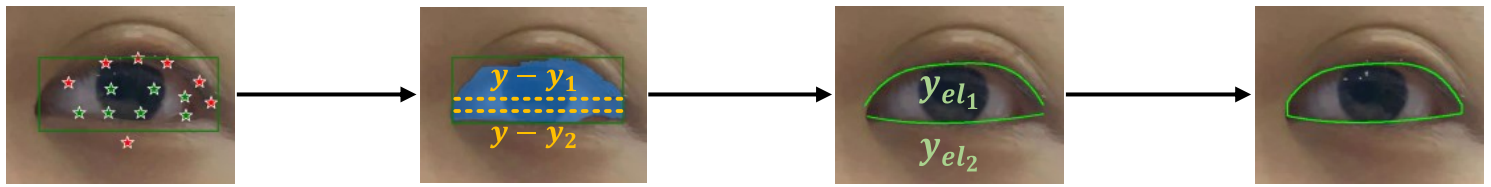
Training Dataset Label



Physical eye model for iris shape.



Automatic SAM for iris location.



Automatic SAM for sclera.

Iris label

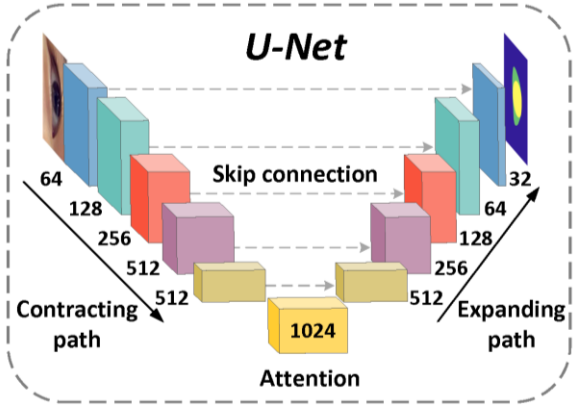
Sclera label

Manual label
(2.37cm)

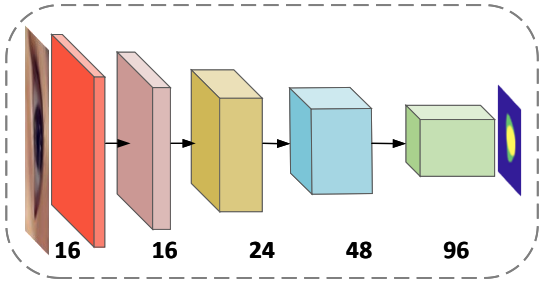
Close

Generated label
(2.57 cm)

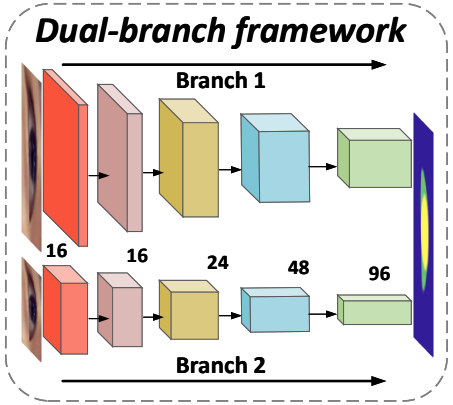
Latency on Mobile Devices



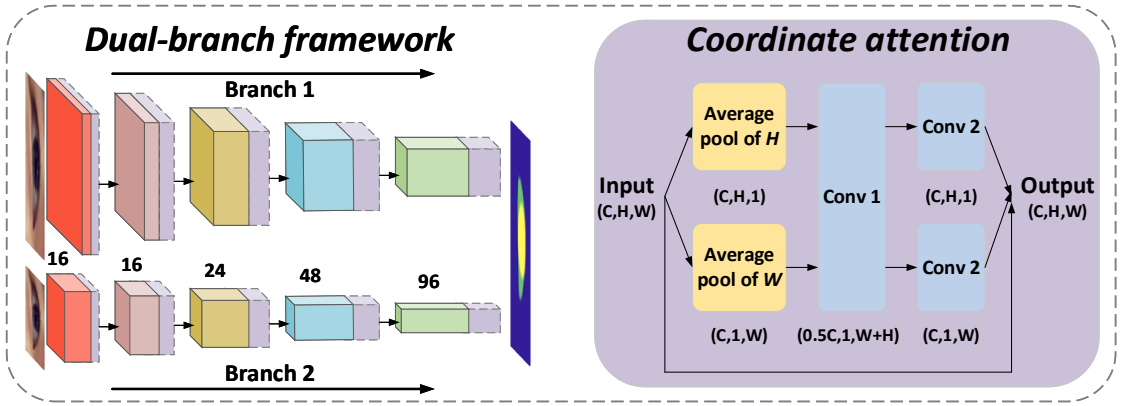
Good performance (2.37cm)
Bad latency (2.854s)



Bad performance (3.29cm)
Good latency (0.021s)



Maintain performance (2.97cm)
Good latency (0.024s)



Further maintain performance (2.87cm)
Good latency (0.026s)