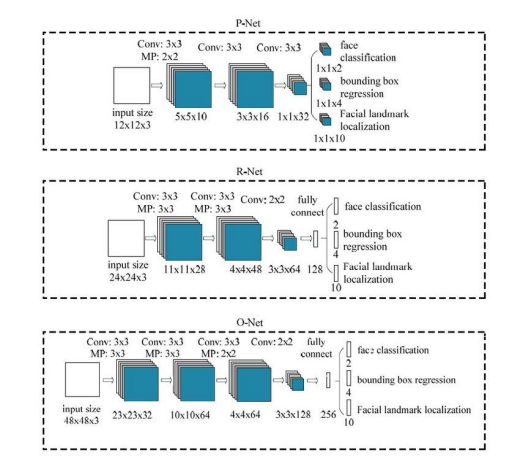
Efficient-ViT-Guard: Real-Time Face Detection and Blurring System for Video Privacy

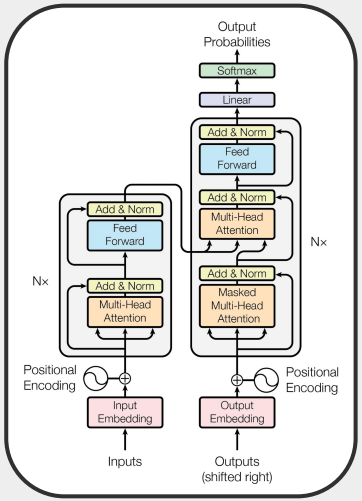
Dev Chheda, Divya Nori, Anirudh Valiveru

Pervasive video recording and sharing poses privacy concerns

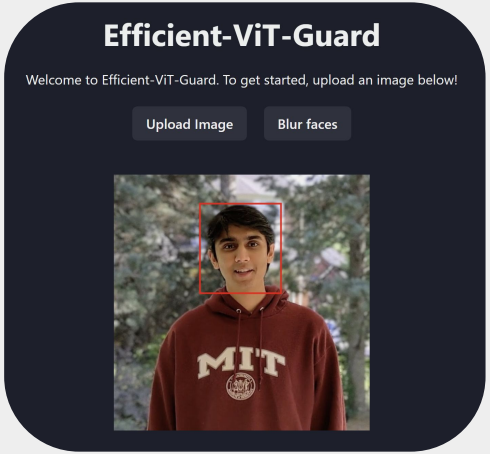
Potential solution: real-time face blurring



MTCNN as face detection model

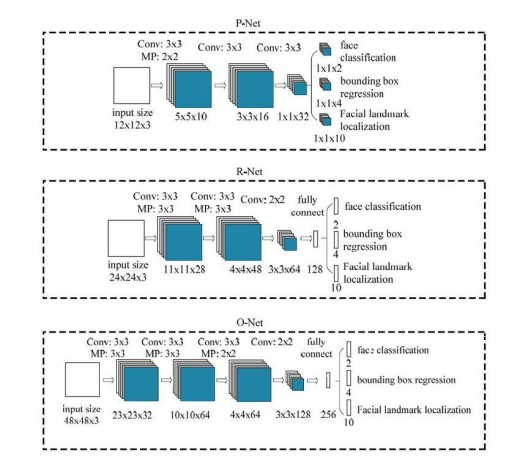


EfficientViT SAM as face segmentation model

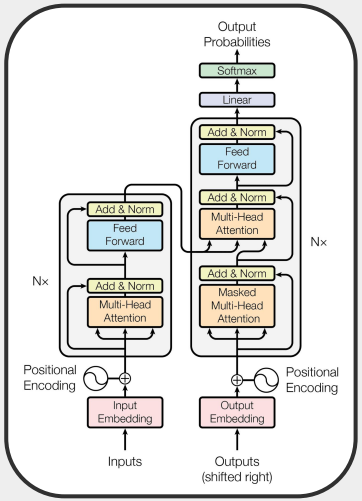


Full-stack web app on Node.js + Flask

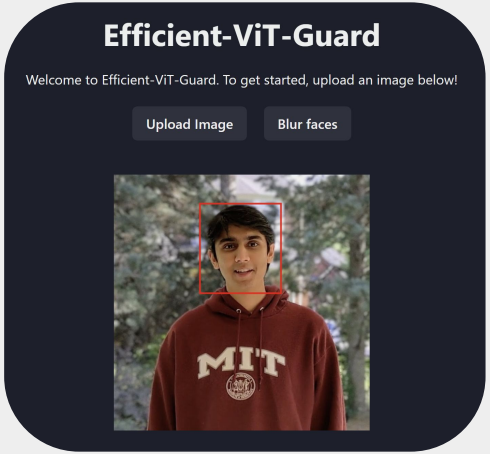
Step 1: Benchmark face detection model



MTCNN as face detection model



EfficientViT SAM as face segmentation model



Full-stack web app on Node.js + Flask

What do realistic video frames look like?

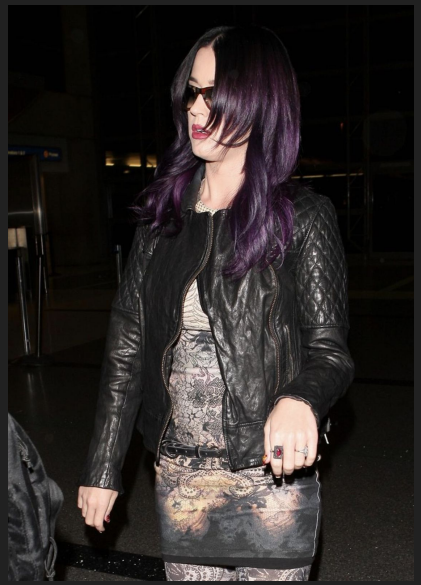
What do realistic video frames look like?

Experimental methodology

Dataset: Hugging Face Wider Face Dataset

High illumination no occlusion

Low illumination no occlusion



High illumination with occlusion

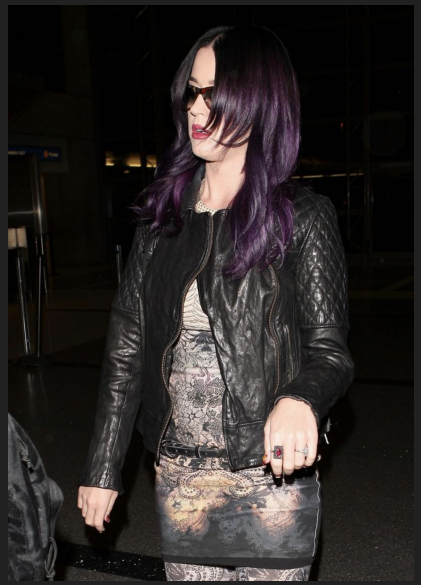
Low illumination with occlusion

Experimental methodology

We compare time efficiency and accuracy of face detection across these settings

High illumination no occlusion

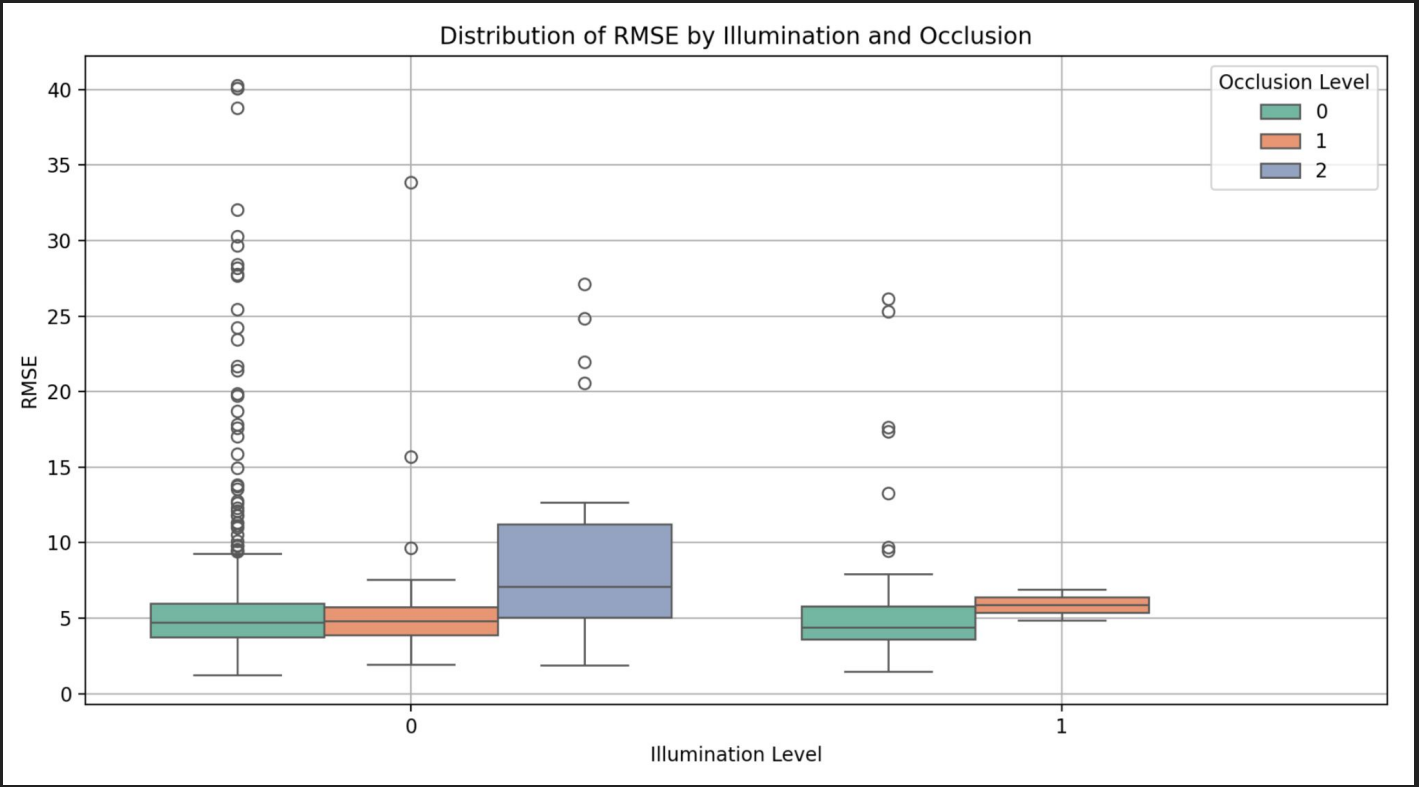
Low illumination no occlusion



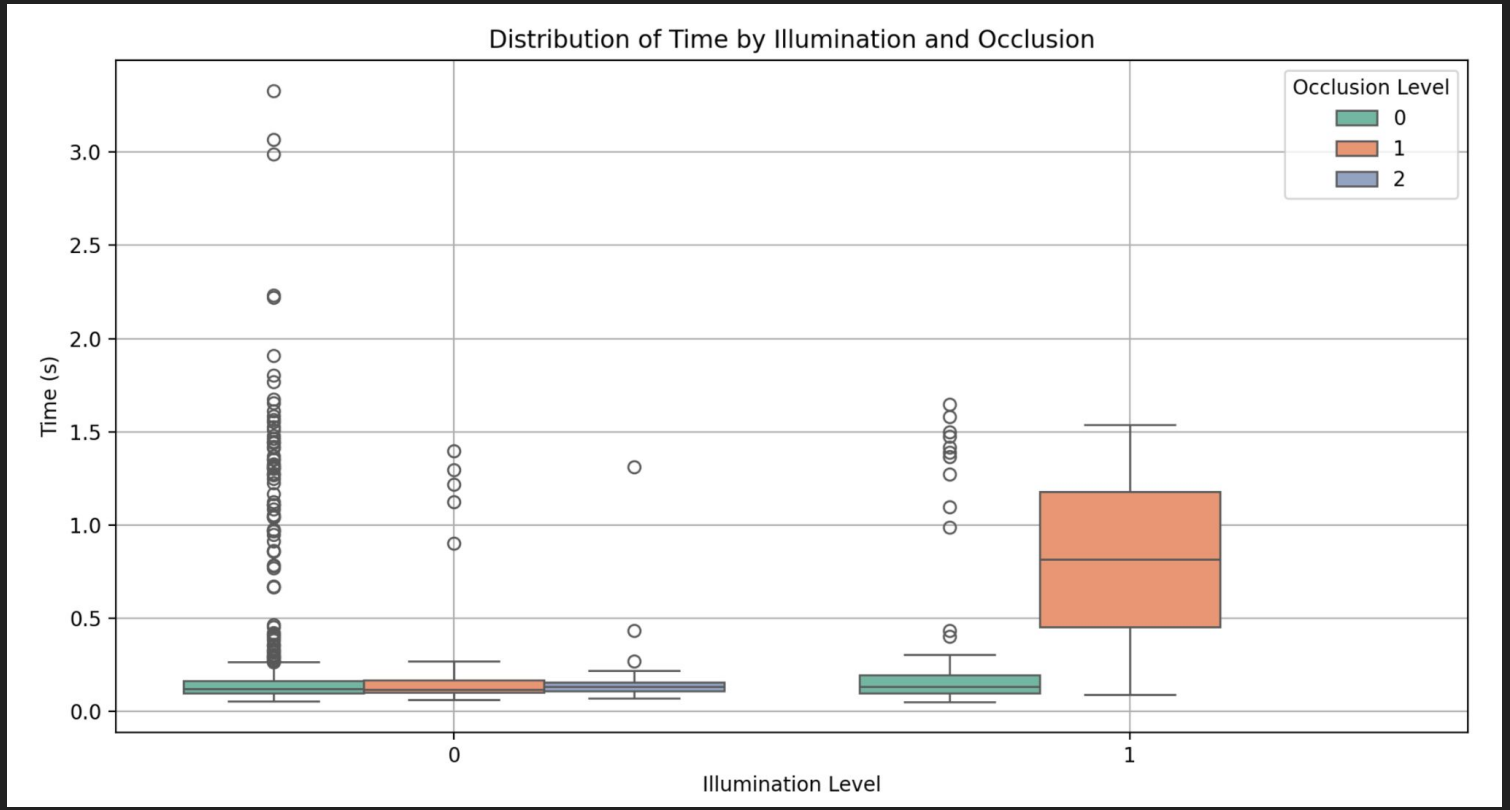
High illumination with occlusion

Low illumination with occlusion

RMSE of predicted face location is significantly higher under high occlusion



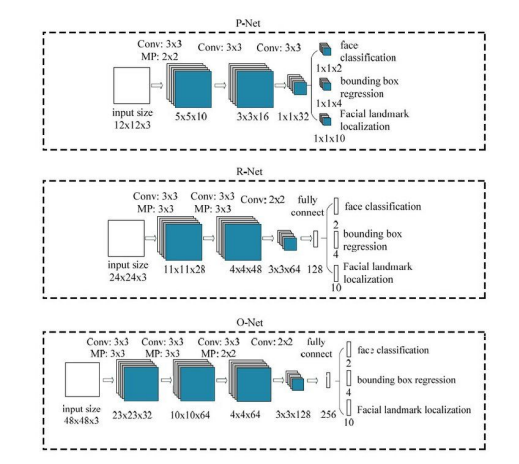
Low illumination and occlusions decrease time efficiency



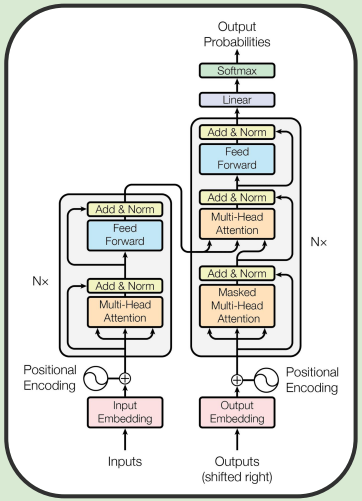
Successful example in low illumination setting

Successful example in occluded setting

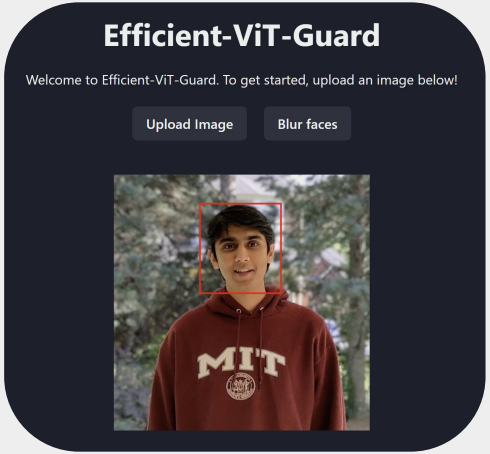
Step 2: Benchmark face segmentation model



MTCNN as face detection model

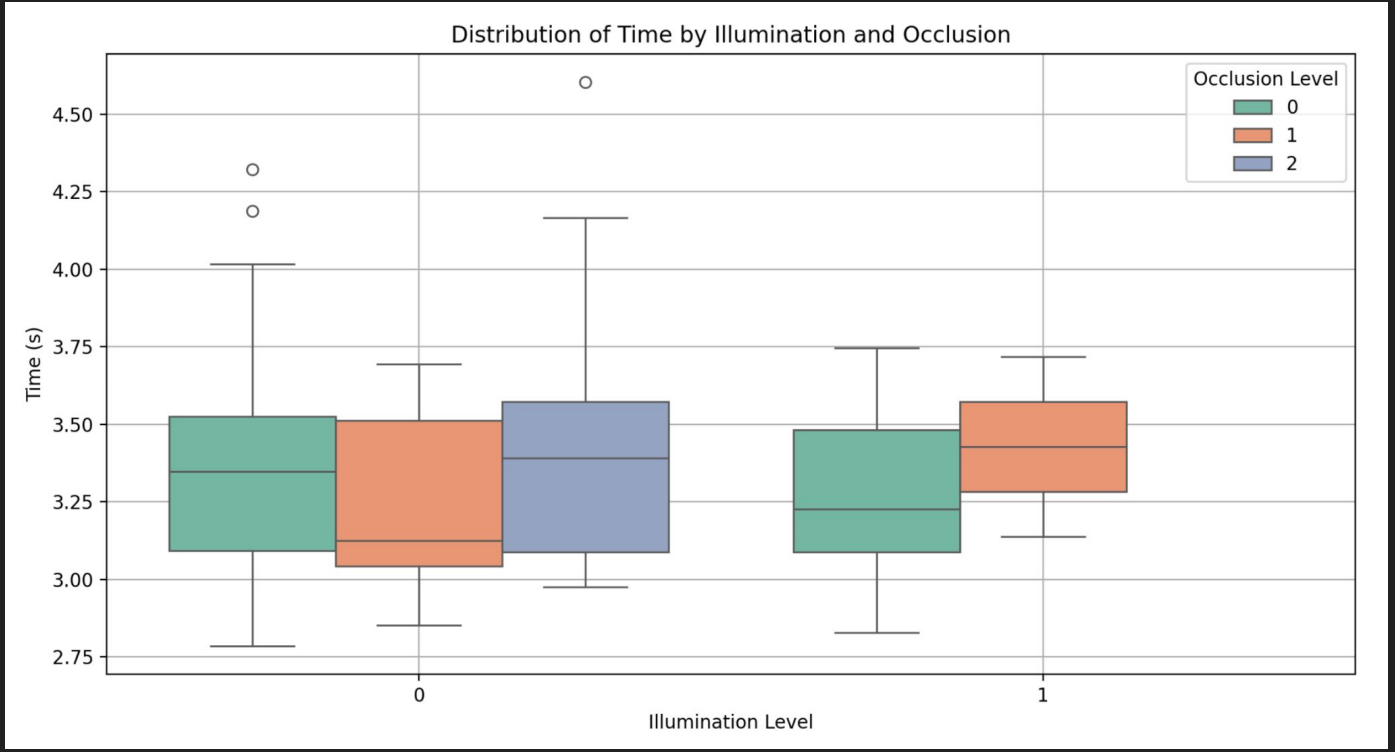


EfficientViT SAM as face segmentation model



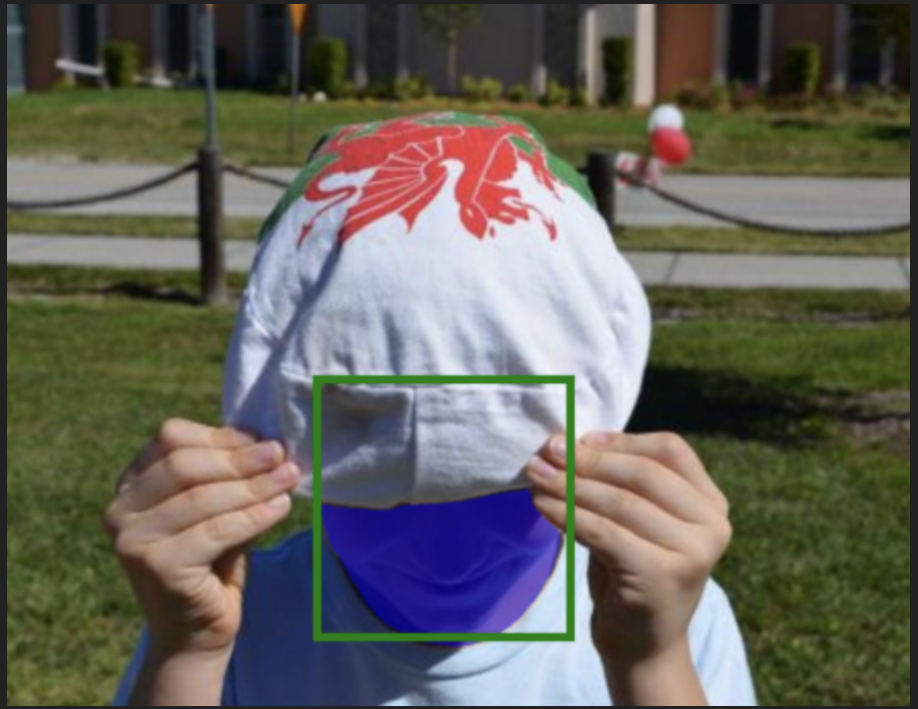
Full-stack web app on Node.js + Flask

Low illumination and occlusions don’t significantly affect time efficiency

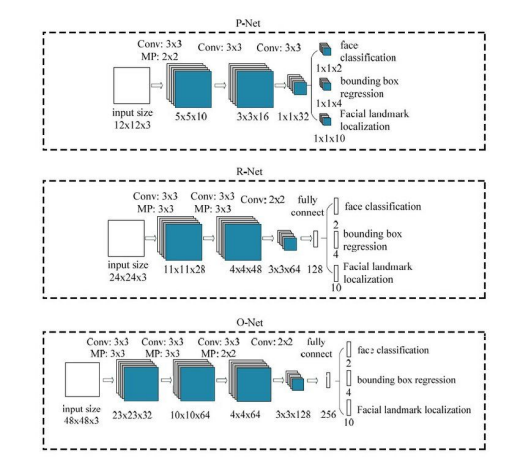


Successful example in low illumination setting

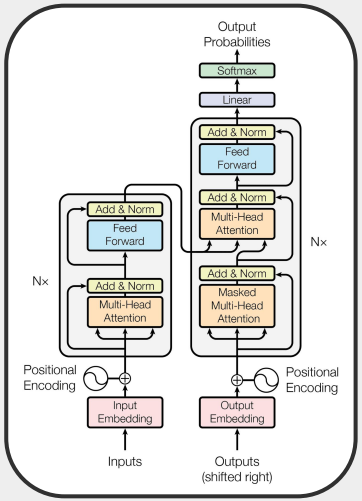
Successful example in occluded setting



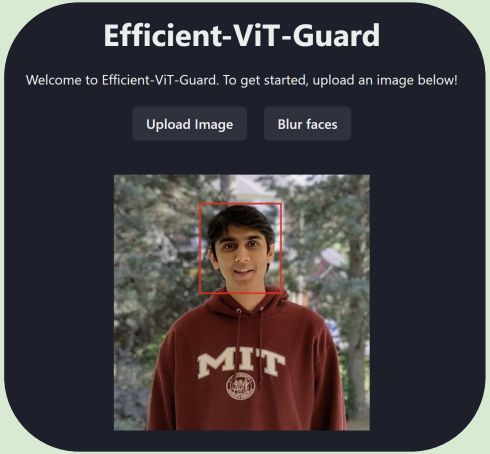
Step 3: Integrate into full-stack system



MTCNN as face detection model

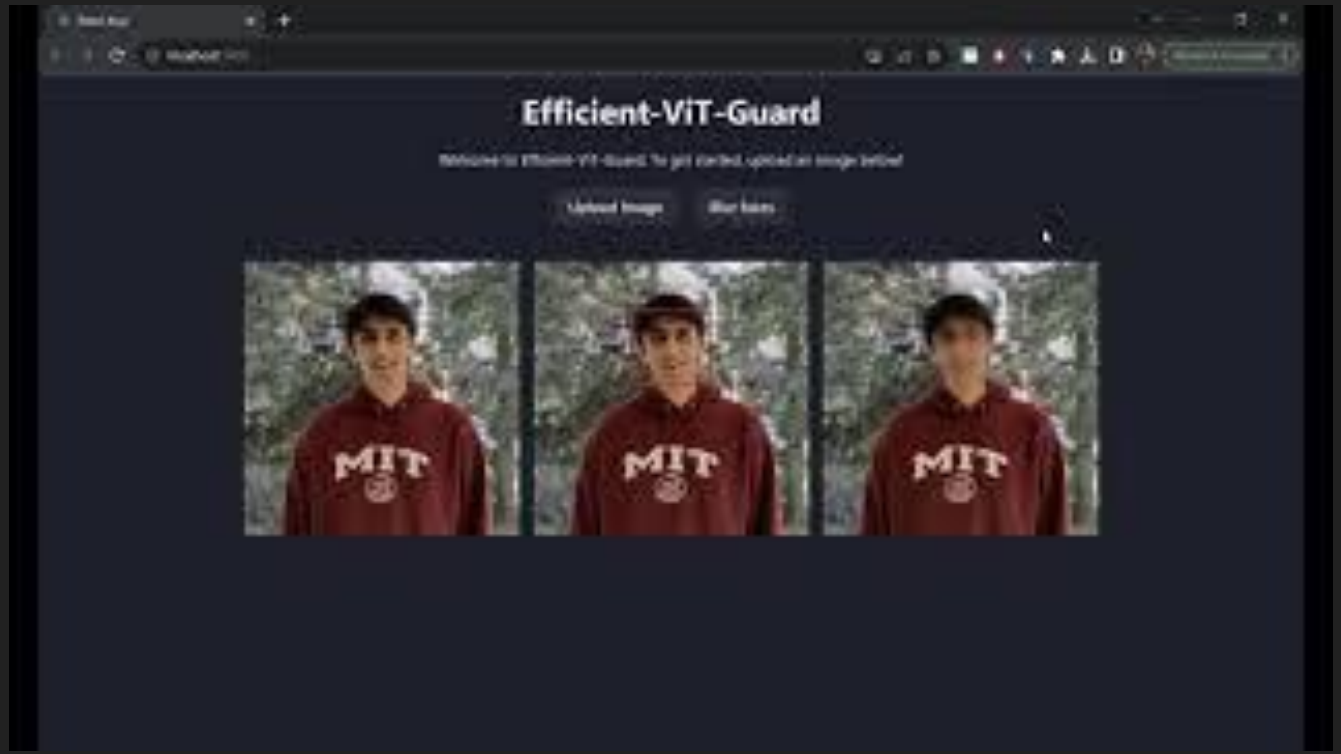


EfficientViT SAM as face segmentation model



Full-stack web app on Node.js + Flask

Efficient-ViT-Guard demo



https://www.youtube.com/watch?v=VTCGm4N-_k8

Takeaways and next steps

- ***MTCNN and Efficient-ViT SAM can be used to build a fast and reliable face detection/blurring privacy preserving system***
- High occlusion affects accuracy of MTCNN, but illumination does not have significant effect
- Occlusion and low illumination together affect the time efficiency of MTCNN and Efficient-ViT SAM
- In general, time efficiency of both models is consistent across conditions

Potential extensions:

- Use optical flow to track faces over several video frames
- Only a single bounding box from user is needed to track and blur a face