AI HACK ATHON

FACE ANALYSIS CHALLENGE

PIXTA VIETNAM · Cohost AI Smarter Sharing · VIBLO · HUB NETWORK · iHUB

SOLUTION FOR FACE ANALYSIS

A3N1 Team

Nguyễn Trọng Anh HCMUT

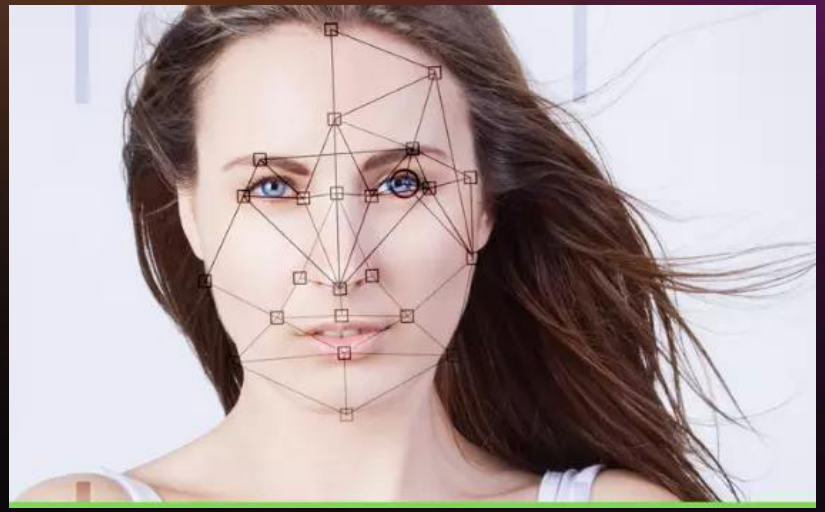
Nguyễn Đăng Khoa VGU

Nguyễn Văn Quân UIT

Ho Chi Minh, Jan 2024

I. Introduction

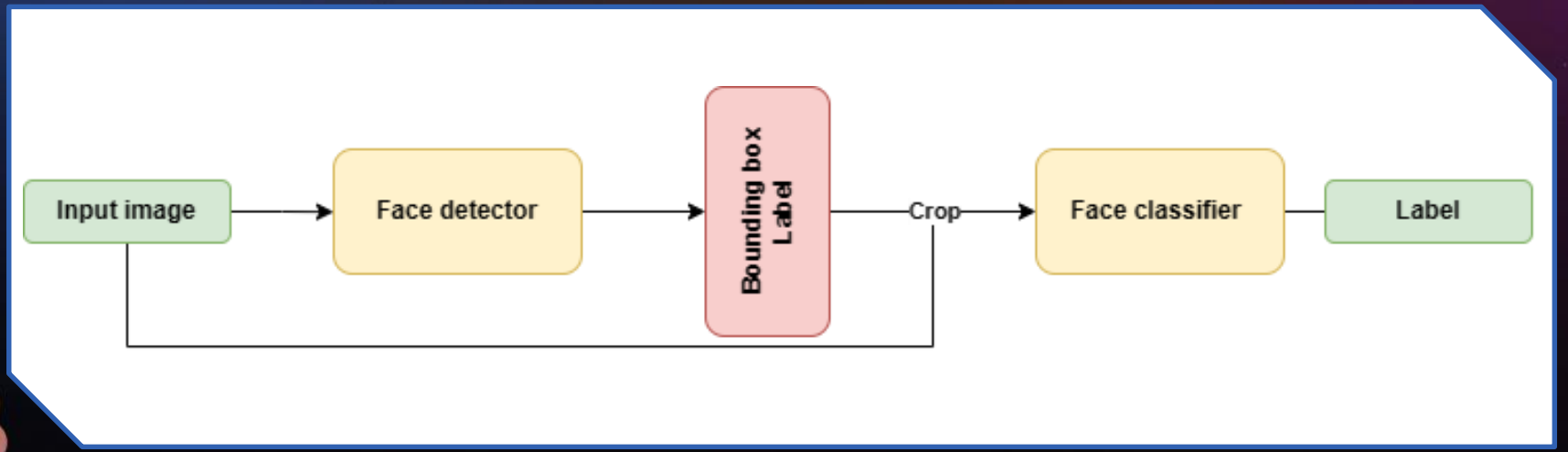
- Face analysis is the process of using technology to analyze human facial features to understand information such as age, gender, mood, and many other factors
- => applied in many fields. ranging, including industrial, medical, security, and entertainment



- The competition poses a challenge to build ideas and apply technology to identify all facial features such as: Gender, Age, Ethnicity, Skin color, Wearing a mask/ Not wearing a mask, Feeling emoji, Face recognition, ...

AI HACK ATHON

II. Proposed method

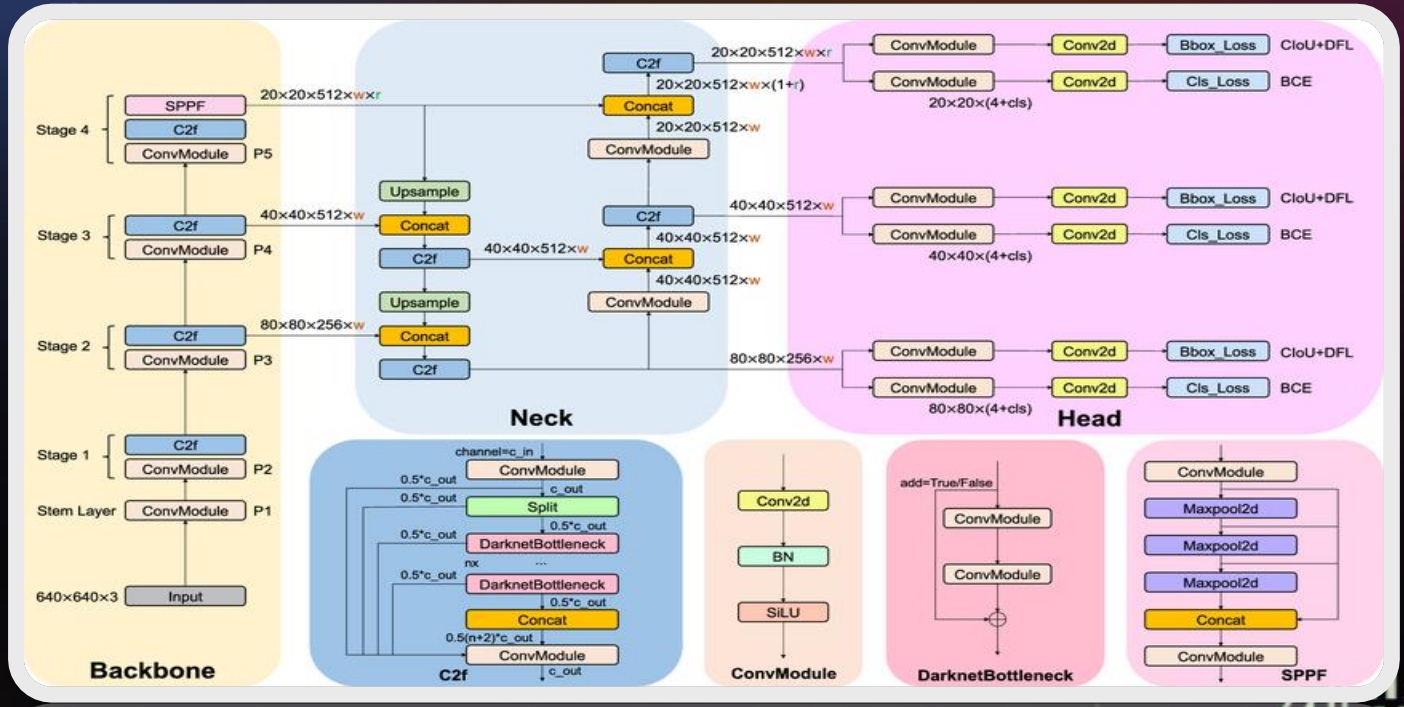


Overview of our pipeline

II. Proposed method

a. Face detector

Face detector: YOLOv8

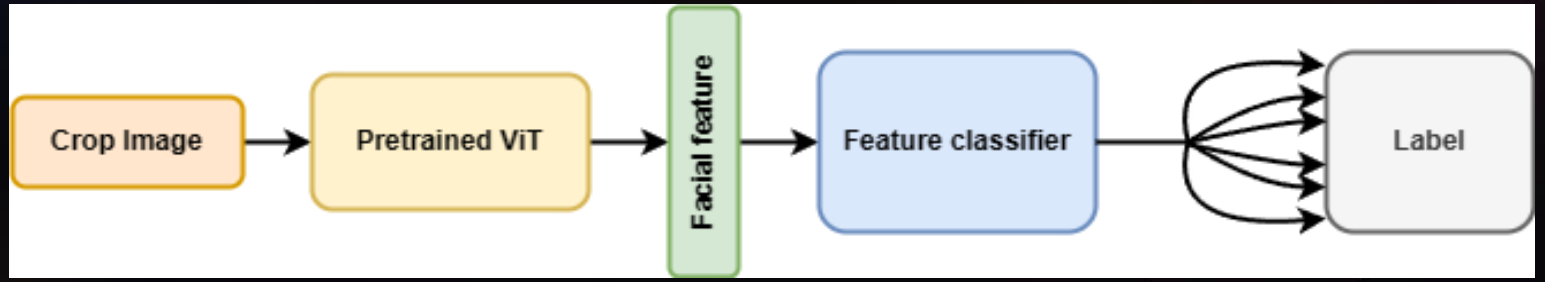


II. Proposed method

b. Face Attributes Classifier

Multi-model for classification task:
- Enormous memory
- Difficult to implement for real application

=> Using single model with multi-task classifier



Overview Face Attributes classifier

II. Proposed method

b. Face Attributes Classifier

Type of classifier:

- Custom SVM: $K(x, x') = (-\gamma \|x - x'\|_2^2 + c)^d$
- Transformer: 4 layers (encoder) – 8 heads
- Fully Connected Layer (MLP): 3 layers

Training strategy :

- Freeze VIT backbone and train Face Attributes Classifier
- Do not use augment data or extra data
- Use Balanced Binary Cross Entropy Loss for ViT (CVPR2023) for 6 attributes

AI HACK ATHON

II. Proposed method

b. Face Attributes Classifier

$$\mathcal{L}_{\text{Bal-BCE}} = -\sum_{\mathbf{y}_i \in \mathcal{C}} w_i [\mathbb{1}(\mathbf{y}_i) \cdot \log \frac{1}{1 + e^{-[\mathbf{z}_{\mathbf{y}_i} + \log \pi_{\mathbf{y}_i} - \log(1-\pi_{\mathbf{y}_i})]}}$$

$$+ (1 - \mathbb{1}(\mathbf{y}_i)) \cdot \log(1 - \frac{1}{1 + e^{-[\mathbf{z}_{\mathbf{y}_i} + \log \pi_{\mathbf{y}_i} - \log(1-\pi_{\mathbf{y}_i})]}})]$$

Bal-BCE loss for ViT (adapted from [1])

[1] Zhengzhuo Xu et al. Learning Imbalanced Data with Vision Transformers.
In CVPR, pages 15793 – 15800, IEEE, 2023.

AI HACK ATHON

III. Results

a. Results on scoreboard

We use VIT base for all following model with different classifier

Type of classifier	Training time (RTX 4090 24g)	Public test	Private test
Custom SVM	2 m/epoch	26.08	25.30
Transformer	3 m/epoch	26.06	**25.37**
Fully Connected Layer	2 m/epoch	**26.13 (2)**	25.06

III. Results

b. Performance and inference time

Model	mAP50/avg dev acc	Capacity	params	Time (s) on CPU i5 1135g7	Time (s) on GPU Tesla T4	Time (s) on GPU Tesla T4(quantized)
Yolov8 n	~0.9401	6mb	~3m	~0.45	~0.06	-
Yolov8 m	~0.9484	50mb	~26m	~0.61	~0.10	0.015 ± 0.005 (640 scale)
Yolov8 x	~0.9525	130mb	~68m	~1.45	~0.23	
VIT small + SVM	0.9088	85mb	~22m	~0.10	~0.02	-
VIT base + SVM	0.9123	140mb	~87m	~0.21	~0.04	-
VIT large +SVM	0.9096	1.2gb	~305m	~0.81	~0.19	-

IV. Self-evaluation

Aspect	Our method
Accuracy	• Top 2 on public test • Achieve 25.37 final score on private test
Business	• Both Yolov8 (face detector) and VIT(face attributes classifier) **have commercial use license** • Can improve performance with augment or extra data • Smaller variants can optimize capacity and inference time
Creativity	• Design custom SVM kernel for face attributes classifier with competitive result and save resources
Possibility	• Face detector has been quantized for faster inference time • Pipeline can run on CPU in less than 1s/frame

V. Limitations

- Challenge in imbalanced data
- Having difficulty in challenging attributes such as age, emotion, especially skintone.
- Face Attributes Classifier is not designed to automatically augment data.
- Face Attributes Classifier has not been quantized.
- Faces with mask are still impossible to predict emotion (bias to neutral)

=> Addressed in the future

AI HACK ATHON

VI. Product & Demo

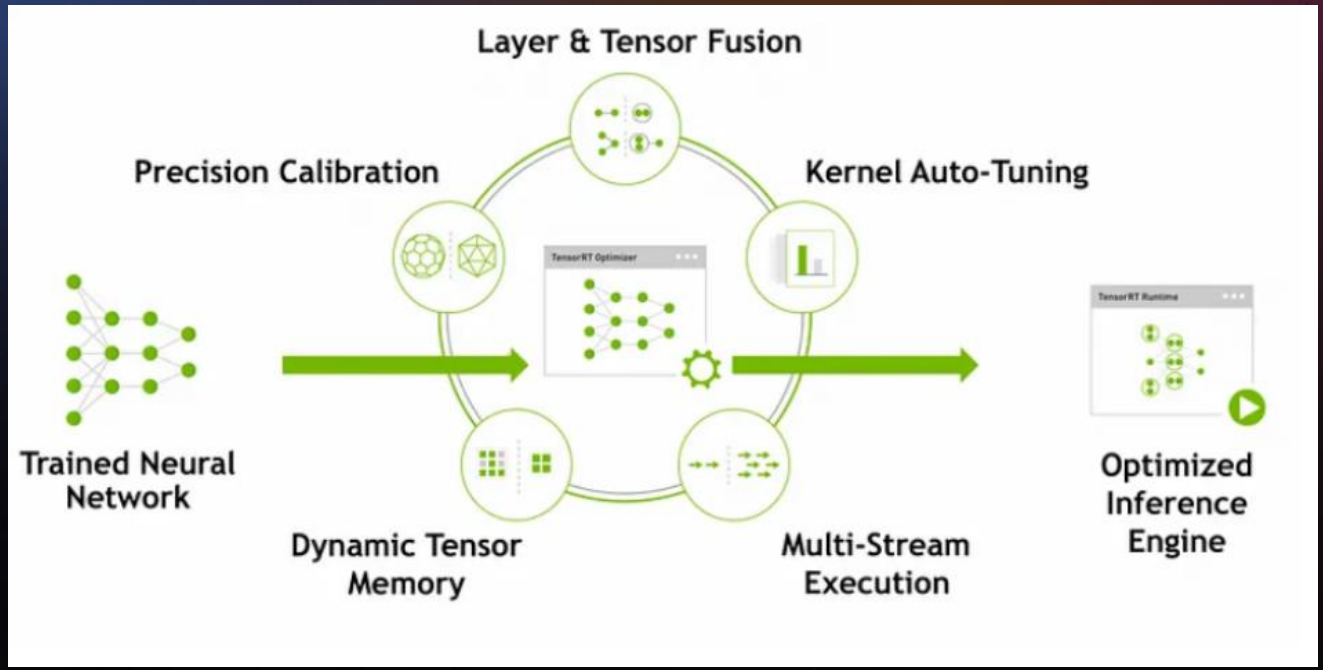
Quantization step: reduce **about 74.07%** inference time of YOLO (jetson Xavier)

Achieve > 30 FPS for Face detection only

```
root@tronganh-jetsonxavier:/home/tronganh/AIChallenge# yolo detect val model=/home/tronganh/AIChallenge/log/checkpoint_yolov8m/
best.pt data=/home/tronganh/AIChallenge/datasets/mydataset.yaml batch=1 imgsz=640
Ultralytics YOLOv8.0.200 🚀 Python-3.8.10 torch-2.0.0a0+ec3941ad.nv23.02 CUDA:0 (Xavier, 6857MiB)
Model summary (fused): 218 layers, 25840339 parameters, 0 gradients, 78.7 GFLOPs
val: Scanning /home/tronganh/AIChallenge/datasets/data_yolo/dev/labels.cache... 703 images, 0 backgrounds, 0 corrupt: 100%|
                 Class     Images  Instances      Box(P          R      mAP50  mAP50-95): 100%|| 703/703 [03:47<00:00
                   all        703        703          1          1      0.995      0.987
Speed: 1.4ms preprocess, 121.5ms inference, 0.1ms loss, 4.4ms postprocess per image
Results saved to runs/detect/val5
 Learn more at https://docs.ultralytics.com/modes/val
root@tronganh-jetsonxavier:/home/tronganh/AIChallenge# yolo detect val model=/home/tronganh/AIChallenge/log/checkpoint_yolov8m/
bestt.pt data=/home/tronganh/AIChallenge/datasets/mydataset.yaml batch=1 imgsz=640
best.onnx         best.pt           best_fp16.engine
root@tronganh-jetsonxavier:/home/tronganh/AIChallenge# yolo detect val model=/home/tronganh/AIChallenge/log/checkpoint_yolov8m/
best_fp16.engine data=/home/tronganh/AIChallenge/datasets/mydataset.yaml batch=1 imgsz=640
Ultralytics YOLOv8.0.200 🚀 Python-3.8.10 torch-2.0.0a0+ec3941ad.nv23.02 CUDA:0 (Xavier, 6857MiB)
Loading /home/tronganh/AIChallenge/log/checkpoint_yolov8m/best_fp16.engine for TensorRT inference...
[01/23/2024-10:50:19] [TRT] [I] Loaded engine size: 51 MiB
[01/23/2024-10:50:22] [TRT] [I] [MemUsageChange] Init cuDNN: CPU +343, GPU +325, now: CPU 741, GPU 3360 (MiB)
[01/23/2024-10:50:22] [TRT] [I] [MemUsageChange] TensorRT-managed allocation in engine deserialization: CPU +0, GPU +49, now: C
PU 0, GPU 49 (MiB)
[01/23/2024-10:50:23] [TRT] [I] [MemUsageChange] Init cuDNN: CPU +0, GPU +0, now: CPU 690, GPU 3312 (MiB)
[01/23/2024-10:50:23] [TRT] [I] [MemUsageChange] TensorRT-managed allocation in IExecutionContext creation: CPU +0, GPU +32, no
w: CPU 0, GPU 81 (MiB)
val: Scanning /home/tronganh/AIChallenge/datasets/data_yolo/dev/labels.cache... 703 images, 0 backgrounds, 0 corrupt: 100%|
                 Class     Images  Instances      Box(P          R      mAP50  mAP50-95): 100%|| 703/703 [01:22<00:00
                   all        703        703          1          1      0.995      0.988
Speed: 1.2ms preprocess, 31.4ms inference, 0.1ms loss, 4.2ms postprocess per image
Results saved to runs/detect/val6
 Learn more at https://docs.ultralytics.com/modes/val
```

HACK ATHON

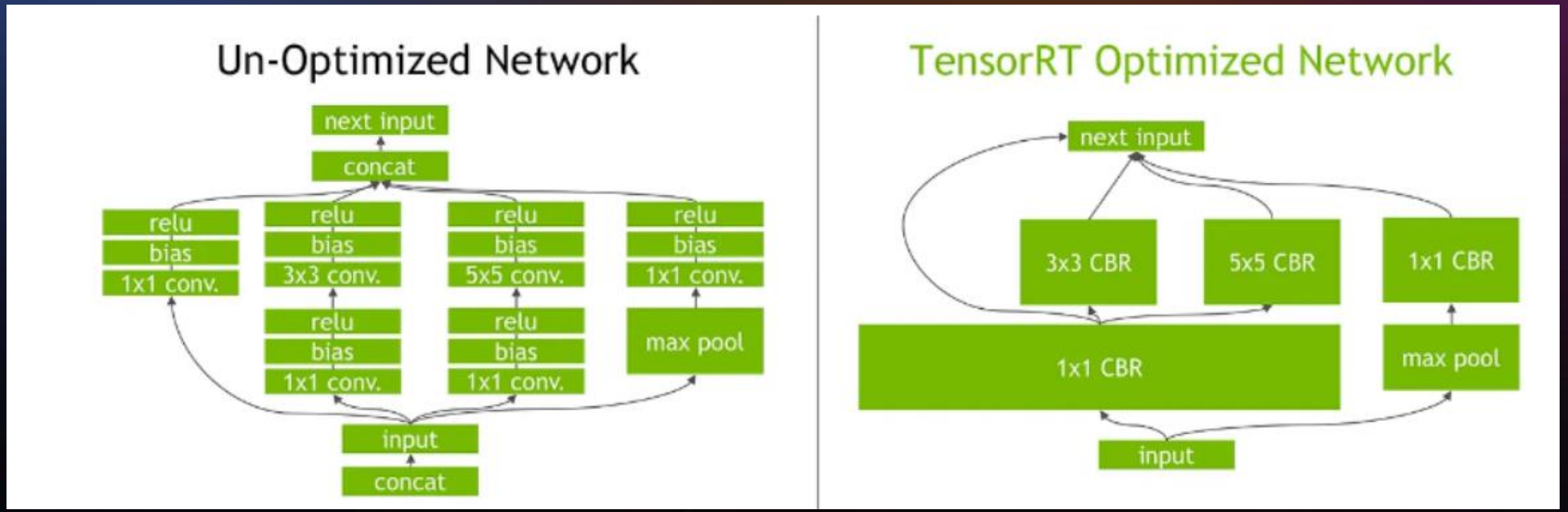
VI. Product & Demo



Quantization and optimization using TensorRT

AI HACK ATHON

VI. Product & Demo



Layer fusion method optimization

VI. Product & Demo

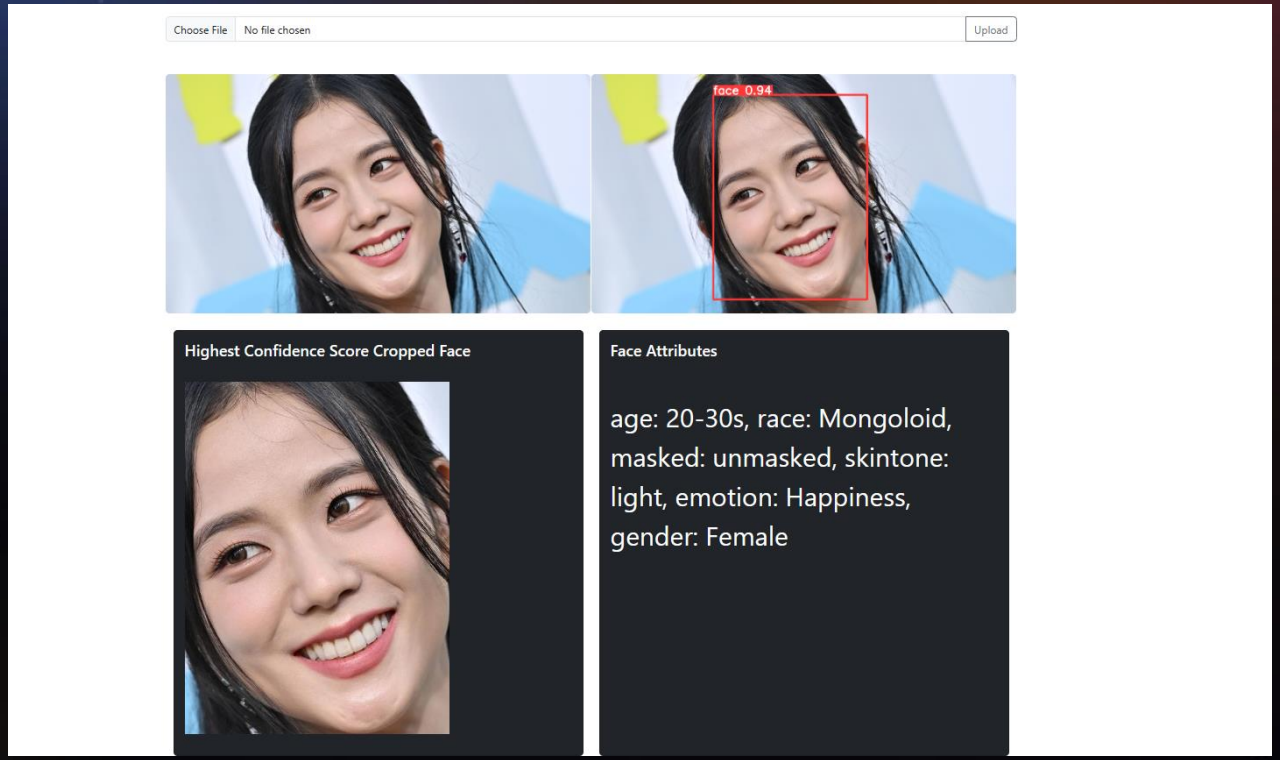
Quantization step: reduce **about 74.07%** inference time of YOLO (jetson Xavier)
Face classifier model has not been quantized yet
Achieve 17 FPS (including pre-processing and post-processing steps)
GPU cost: 2.2 GB
RAM cost: 4.8GB

```
Thời gian yolov8 thực hiện: 0.019832611083984375 s
Thời gian face analysis model thực hiện: 0.038395404815673828 s
Tổng thời gian thực hiện: 0.059618473052978516 s
```

AI HACK ATHON

VI. Product & Demo

Make a simple website that can detect faces in image and analyze face attributes (age, race, masked, skintone, emotion, gender)



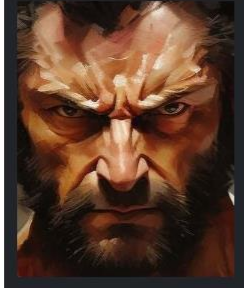
AI HACK ATHON

VI. Product & Demo

⚠ Note:
To get the best result, please upload a vertical photo, without tilt or angle, and with good lighting conditions

Choose File | No file chosen | Upload

Highest Confidence Score Cropped Face



Face Attributes

age: 20-30s, race: Caucasian, masked: unmasked, skintone: mid-dark, emotion: Anger, gender: Male

HACK ATHON

Thank you for listening !

AI HACK ATHON