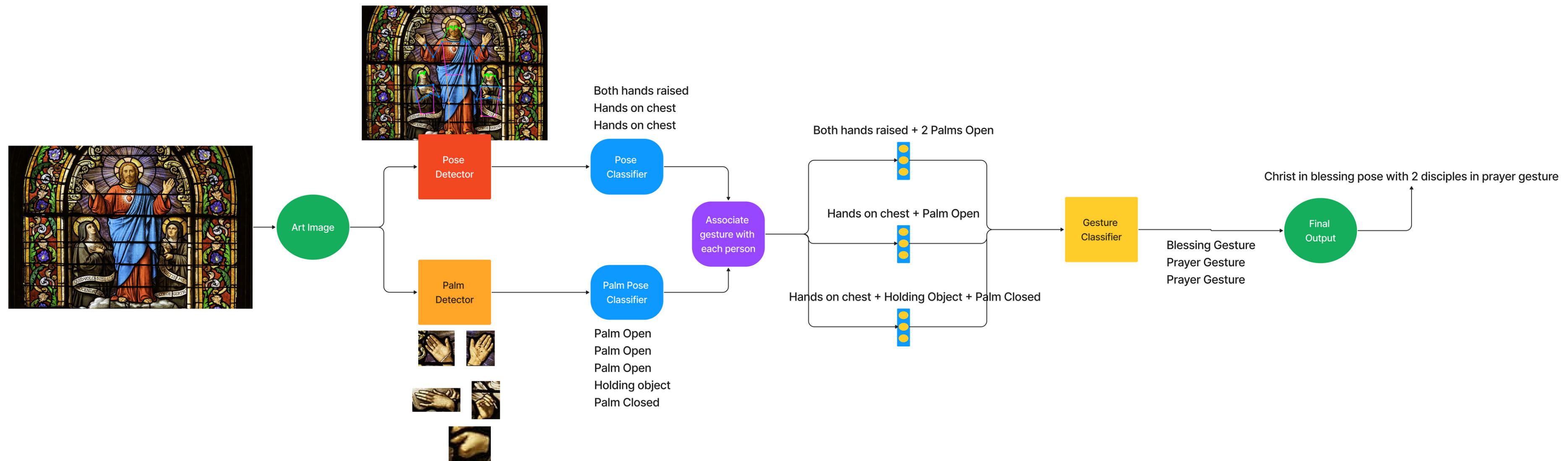


SPEECH GESTURE RECOGNITION

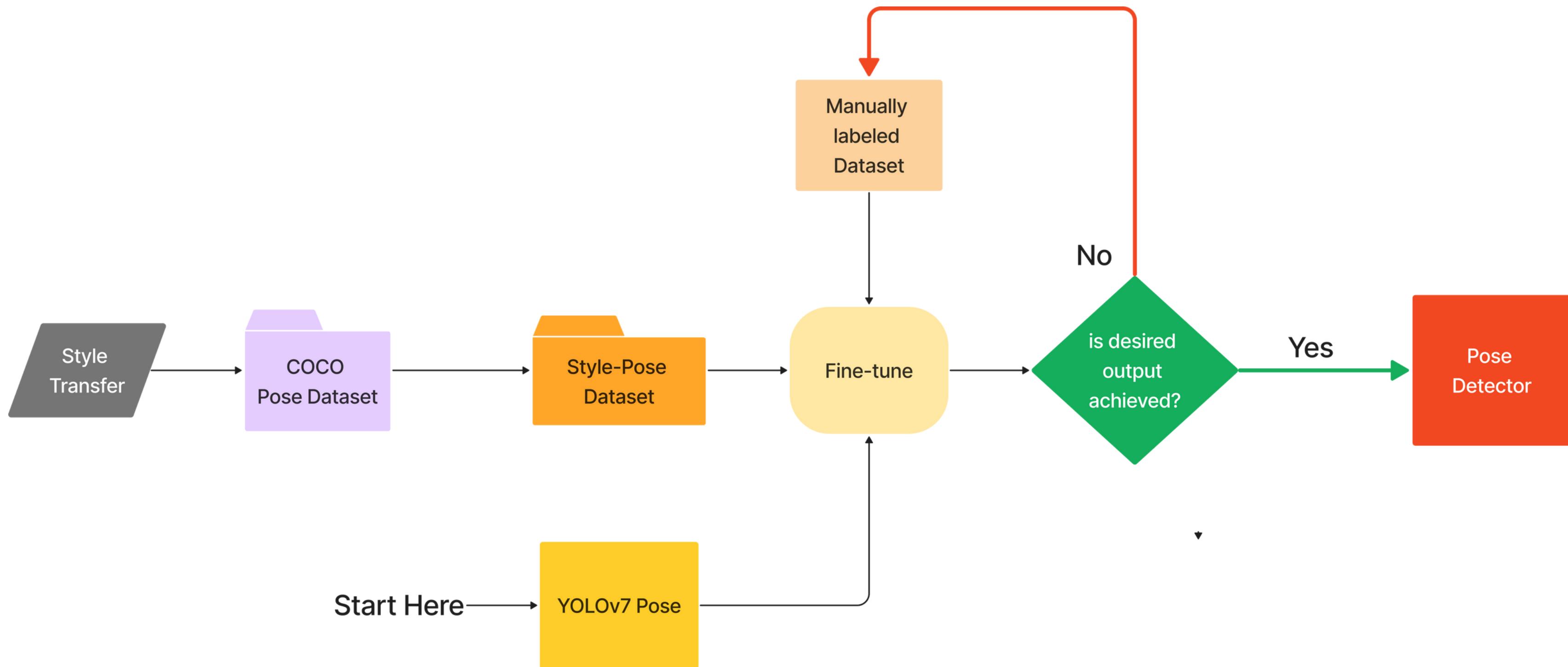
Progress PPT

RIGVED SHIRVALKAR

PROPOSED PIPELINE



POSE MODEL V1



Inferences:

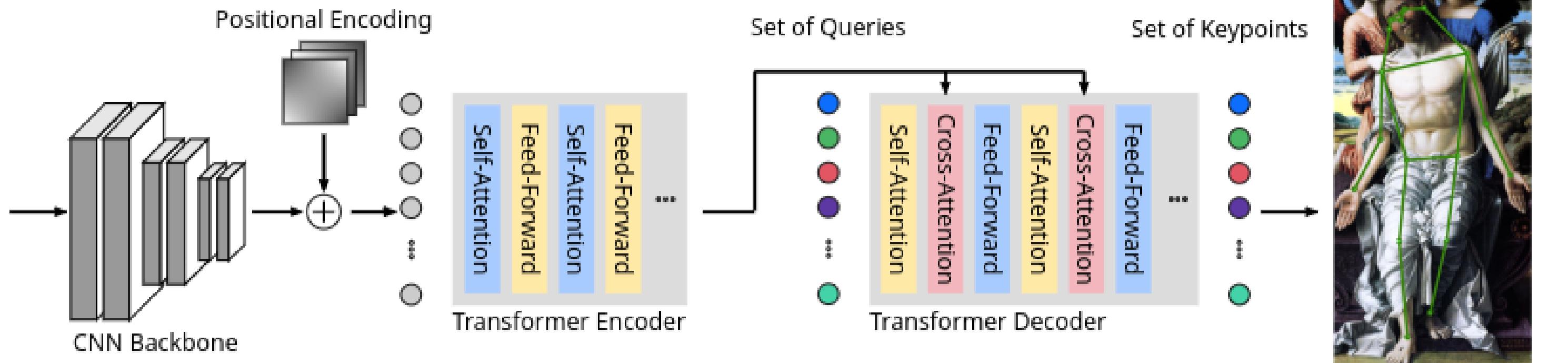
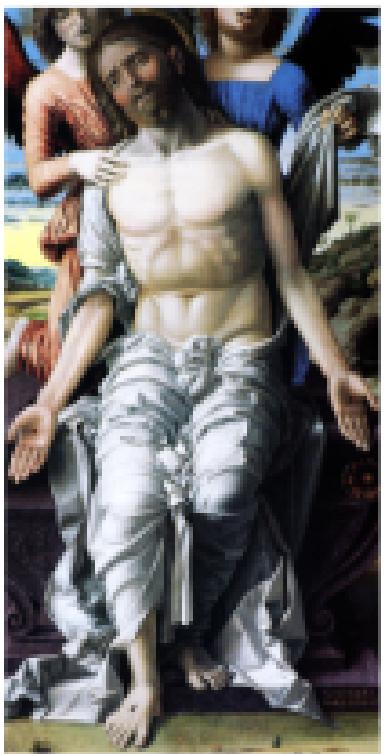
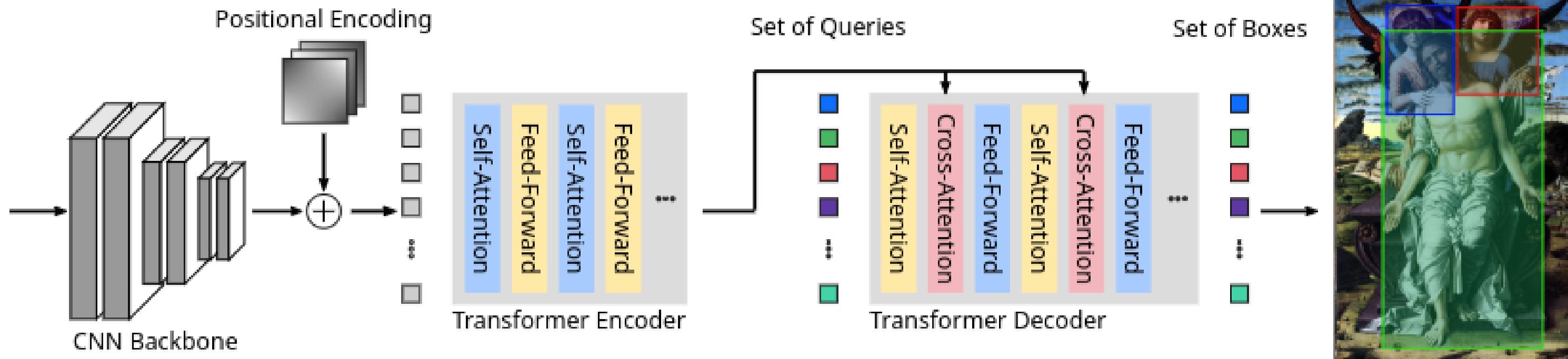
- Detection of small person improved
- Increase in number of detections (more confidence in complicated scenarios)
- Poor Metrics
- Some part of data is confusing the model due to high style transfer value.

Measures:

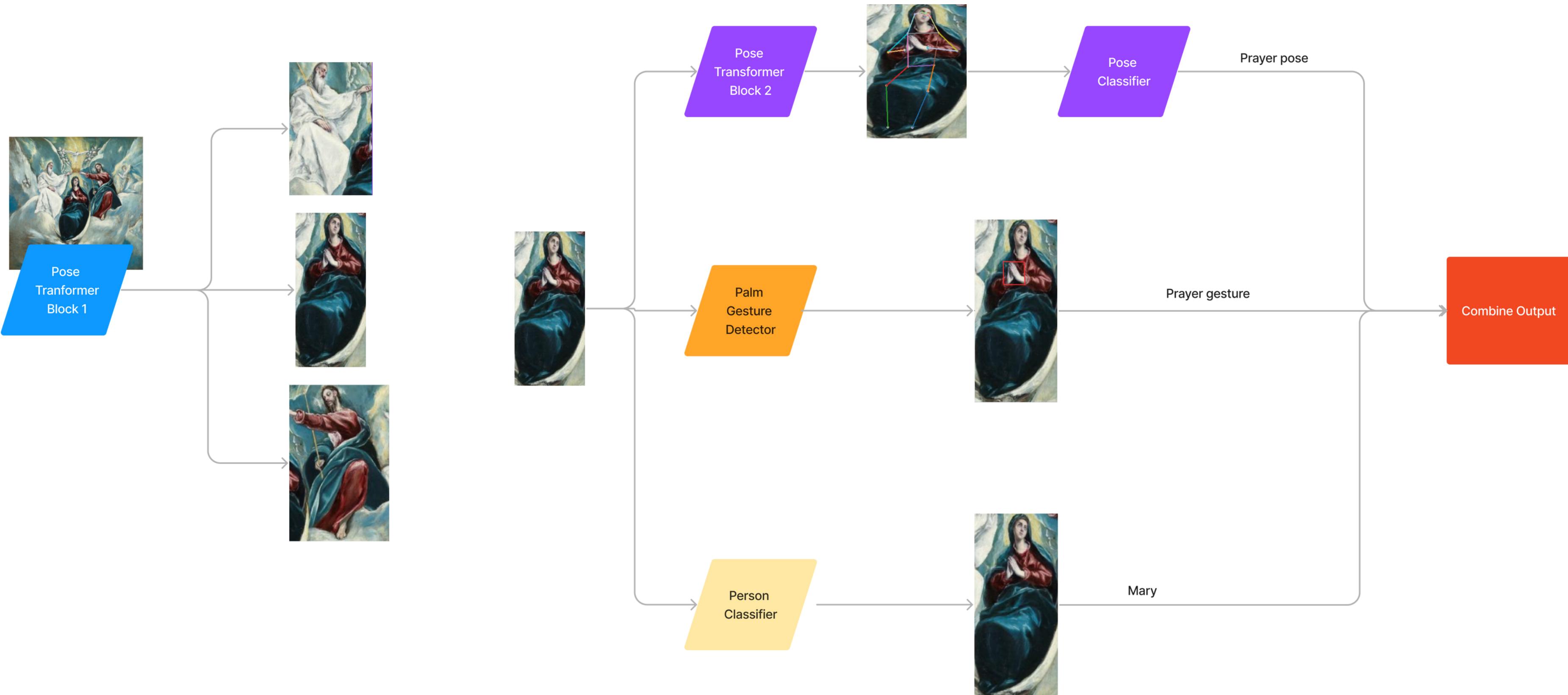
- Reduce the amount of style transfer value and regenerate the data.
- Train the model again

POSE MODEL V2

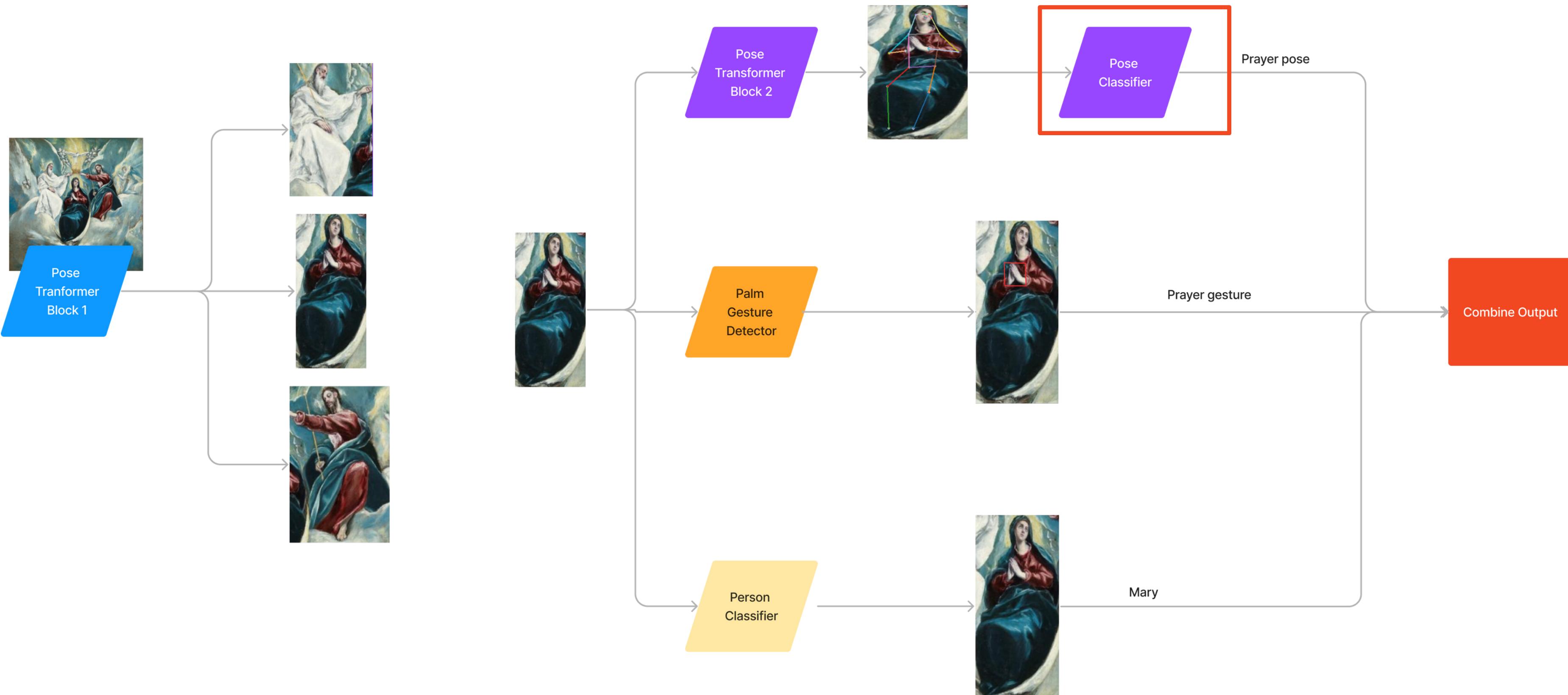
- **MODEL USED:**
 - **POSE TRANSFORMER**
 - **[HTTPS://ARXIV.ORG/ABS/2207.02976](https://arxiv.org/abs/2207.02976)**
- **ADVANTAGES OF MODEL:**
 - **RECENTLY RELEASED AND OPEN SOURCE (GITHUB HAS NO LICENSE)**
 - **USES SEMI SUPERVISED LEARNING SO TRAINED ON LARGE DATA**
 - **TRAINED ON STYLE TRANSFER DATA AS WELL AS LOT OF OTHER DATA**
 - **CURRENT SOTA BY A SIGNIFICANT MARGIN**
 - **SEPARATES EACH PERSON FROM THE IMAGE AND DETECTS KEYPOINTS FOR EACH**



CURRENT PIPELINE



CURRENT PIPELINE

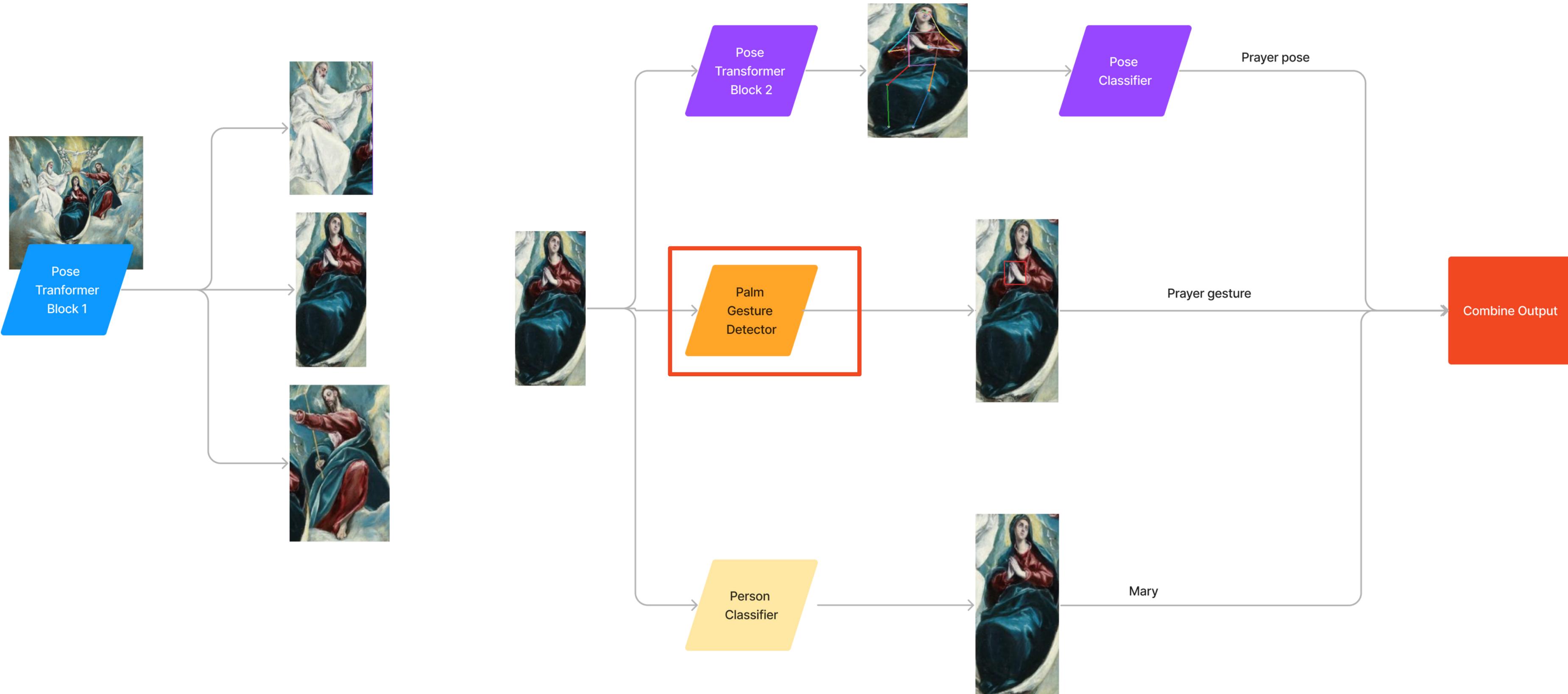


POSE CLASSIFIER

- SIMPLE NEURAL NETWORK
- TRAINED ON KEY-POINT LABELS OBTAINED FROM POSE MODEL (ONLY HANDS)
- CURRENT TRAINED BLOCKS:
 - BOTH HANDS RAISED
 - PRAYER POSE
 - NULL
- (WILL ADD HAND ON CHEST)

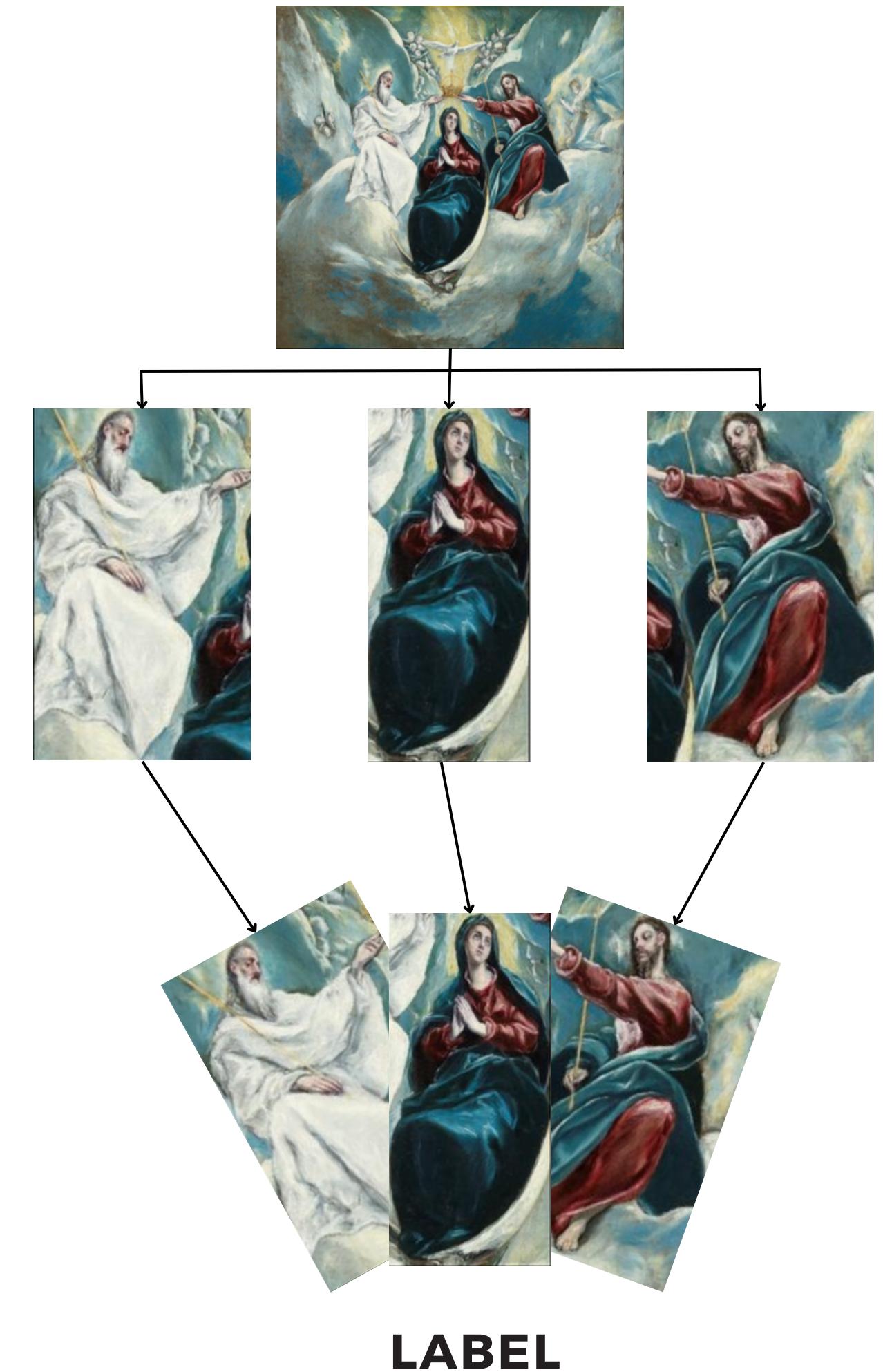


CURRENT PIPELINE



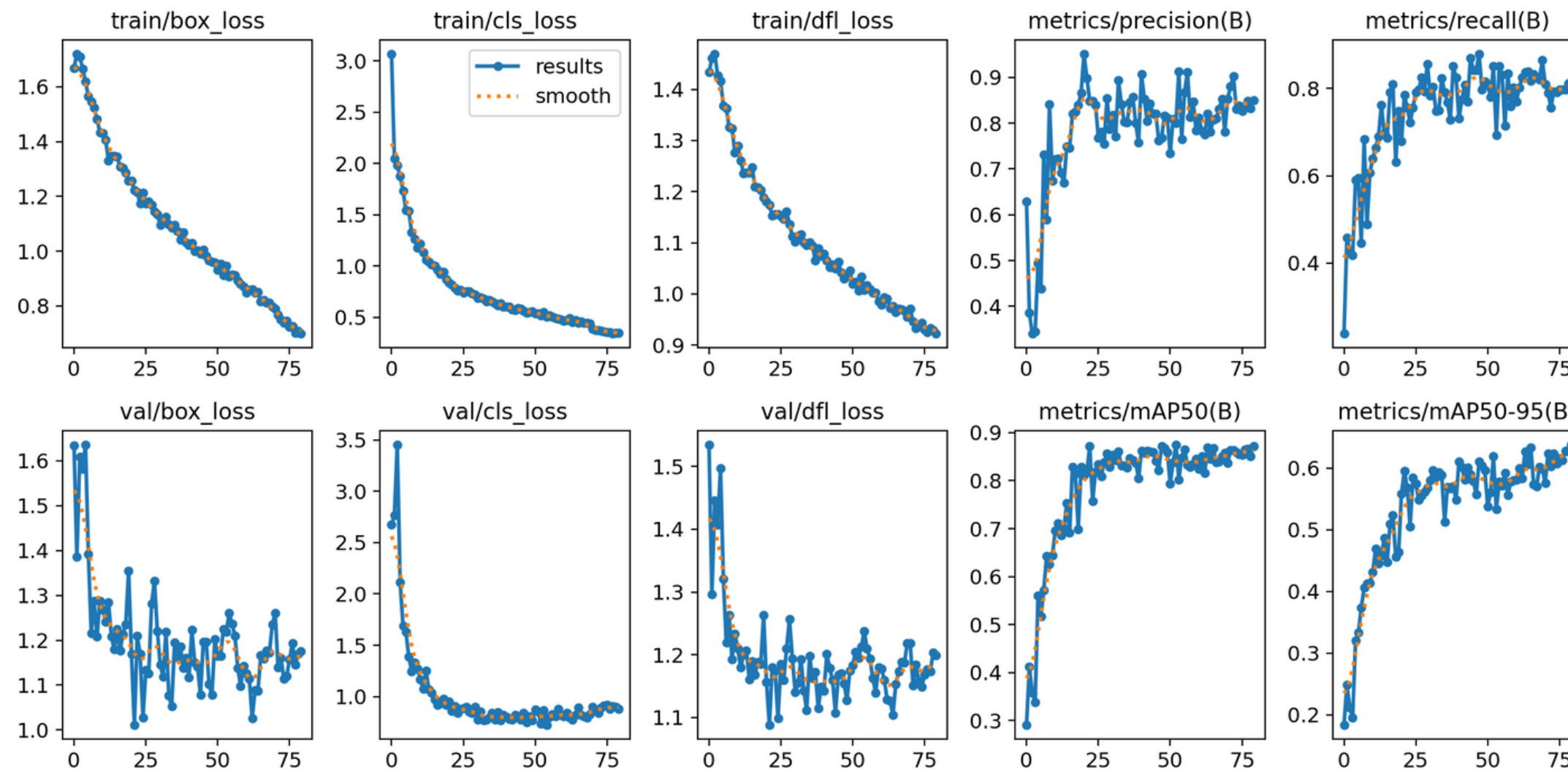
THE DATA

- GENERATED BY EXTRACTING EACH PERSON USING THE TRANSFORMER MODEL
- 2.2K FROM PARTH'S DATA
- 15K FROM RISHAB'S DATA (1.5K FULL IMGS)
- BROUGHT IT DOWN TO 2.2K BY TAKING ONLY THE POSITIVE FULL IMAGES
- FINALLY OBTAINED CLOSE TO 4500 IMAGES (FROM RISHAB AND PARTH'S DATA) AND ADDED THE FOLLOWING CLASSES
 - 415 NULL IMAGES
 - 165 BLESSING IMAGES
 - 260 PRAYER GESTURES
 - 125 IMAGES WITH FRUIT
 - 35 IMAGES WITH GRAPES



PALM POSE MODEL

- YOLO V8 OBJECT DETECTION MODEL



*CAN IMPROVISE WITH A LARGER MODEL FURTHER

CURRENT INTERFACE

```
(yolo) rigvedrs@Rigveds-Lenovo:~/AI/Red Hen/Full_Pipeline_V2$ python3 B1_detect_person.py
Loading model...
loading image from ./img.jpg...
Detecting persons...
Saving image dictionaries in ./data/image_data.pickle...
Successfully Completed!
(yolo) rigvedrs@Rigveds-Lenovo:~/AI/Red Hen/Full_Pipeline_V2$ python3 B2_detect_keypoints.py
Loading pickle file...
Loading Model
Detecting Keypoints
Saving image dictionaries in ./data/image_data.pickle...
Successfully Completed!
(yolo) rigvedrs@Rigveds-Lenovo:~/AI/Red Hen/Full_Pipeline_V2$ python3 B3_classify_poses.py
Loading pickle file...
Pickle file successfully loaded
Loading palm detection model...
Detection Gestures...

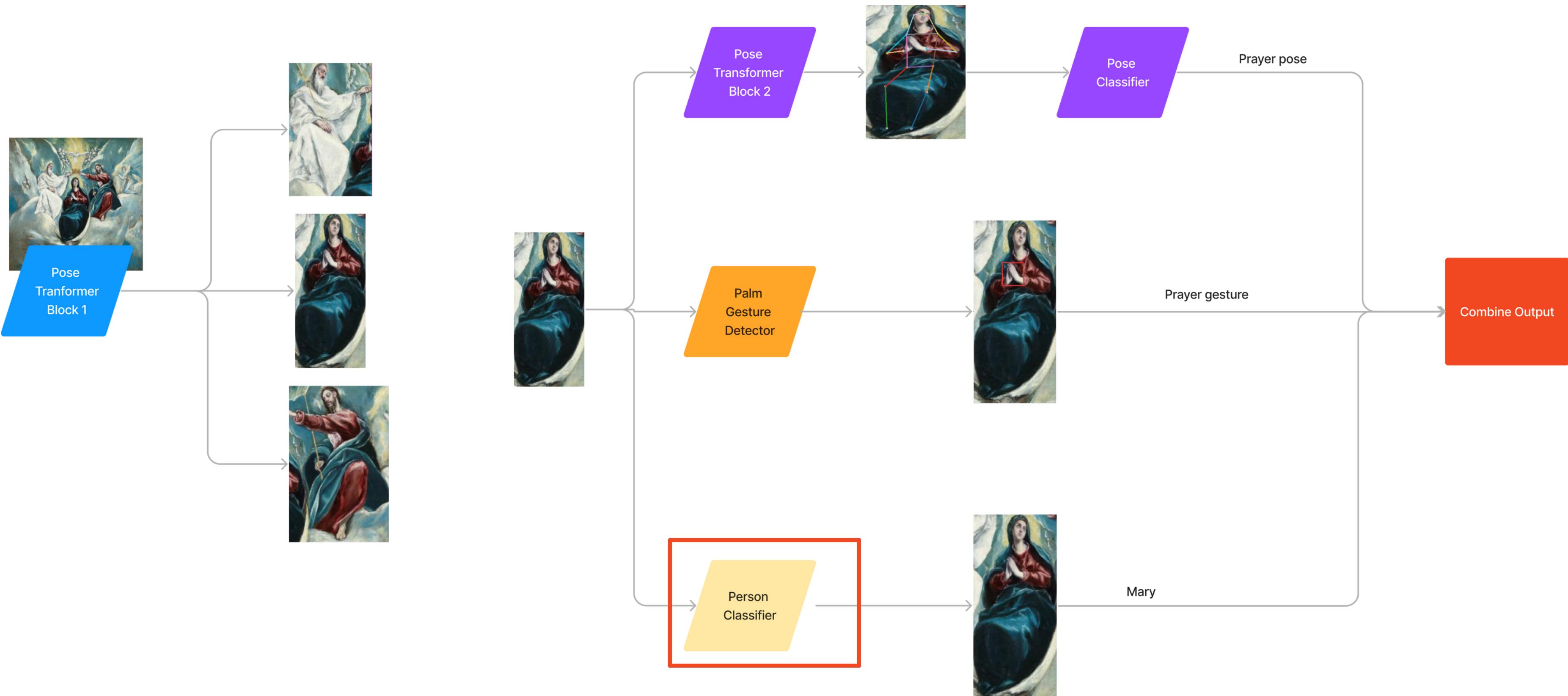
WARNING !NMS time limit 0.550s exceeded
0: 512x352 1 blessing, 163.7ms
Speed: 825.0ms preprocess, 163.7ms inference, 1253.5ms postprocess per image at shape (1, 3, 512, 352)

0: 512x416 1 blessing, 75.4ms
Speed: 6.4ms preprocess, 75.4ms inference, 1.1ms postprocess per image at shape (1, 3, 512, 416)

0: 512x224 (no detections), 51.2ms
Speed: 0.9ms preprocess, 51.2ms inference, 0.8ms postprocess per image at shape (1, 3, 512, 224)
Detected Blessing Gesture
Detected Blessing Gesture
Saving image dictionaries in ./data/image_data.pickle...
Pickle file successfully saved
Successfully Completed
```

*JN

CURRENT PIPELINE



PERSON CLASSIFIER

- **YOLO CLASSIFICATION MODEL**
- **CLASSIFY IMAGE DATA TO DETECT CLASSES**
 - **MARY**
 - **BABY JESUS**
 - **(ALSO ADDING SOME ANGEL IMAGES)**
- **CURRENTLY LABELING THE DATA**
- **TRAINING WILL BE COMPLETED BY THIS WEEKEND**

FINAL OUTPUT

- COMBINE THE PREDICTIONS AND GIVE FINAL OUTPUT
- CURRENT PRED CLASSES:
 - FOR HANDS:
 - BOTH ARMS RAISED (ORANS GESTURE: PRAISE AND SUBMISSION TO GOD)
 - PRAYER POSE (POSTURE OF PRAYER AND DEVOTION TO GOD)
 - FOR PALM:
 - PRAYER POSE
 - TWO FINGERS RAISED (BLESSING GESTURE)
 - FOR PERSON CLASSIFIER:
 - MARY
 - BABY JESUS
 - (IF BOTH IN ONE THEN IT'S MARY HOLDING BABY JESUS)

POSE MODEL

- 2 BLOCKS
 - PERSON EXTRACTION
 - POSE DETECTION

POSE CLASSIFIER

- NEURAL NETWORK
- TRAINED ON HAND AND SHOULDER KEY-POINT VALUES

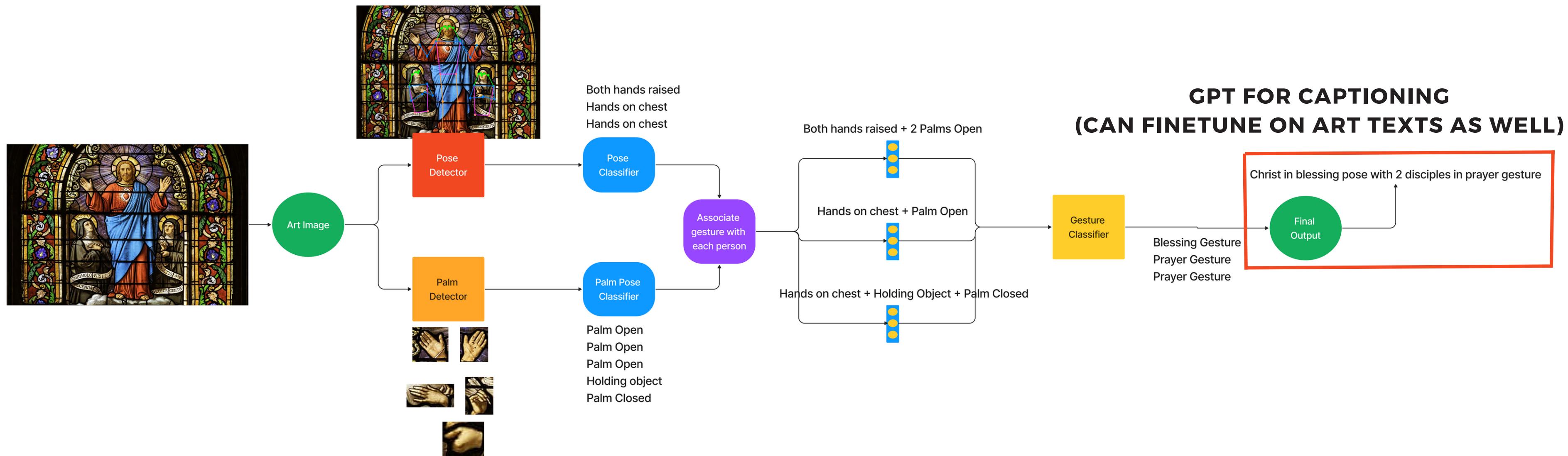
PALM POSE MODEL

- YOLO OBJECT DETECTION MODEL
- TRAINED ON MANUALLY LABELLED DATA

PERSON CLASSIFIER

- YOLO CLASSIFICATION MODEL
- TRAINED ON MANUALLY LABELLED DATA

FURTHER WORK



FUTURE WORK

- **FUTURE WORK:**
 - **IMAGE CAPTIONING**
 - **FINETUNE AN IMAGE CAPTIONING MODEL USING THE CAPTIONS OBTAINED FROM THE PIPELINE**

⚡ Hosted inference API ⓘ

🕒 Image-to-Text



Computation time on Intel Xeon 3rd Gen Scalable cpu: 2.863 s

painting of a woman holding a child in her lap

JSON Output Maximize

- **PROBLEMS:**

- **MULTIPLE IMAGES DETECTED FOR SAME PERSON**
- **DECIDING MORE POSE BLOCKS**
- **SOME EXCEPTIONS FOR BOTH HANDS RAISED**
- **WILL HAVE TO LABEL A LOT OF DATA MOVING FORWARD AS WELL (SEMI SUPERVISED LEARNING APPROACH)**

