

Capstone Project

Extraction Information from Vietnamese ID Card Images

Members



Đỗ Công Duy (Leader)
HE150348



Vũ Đoan Quang
HE153583

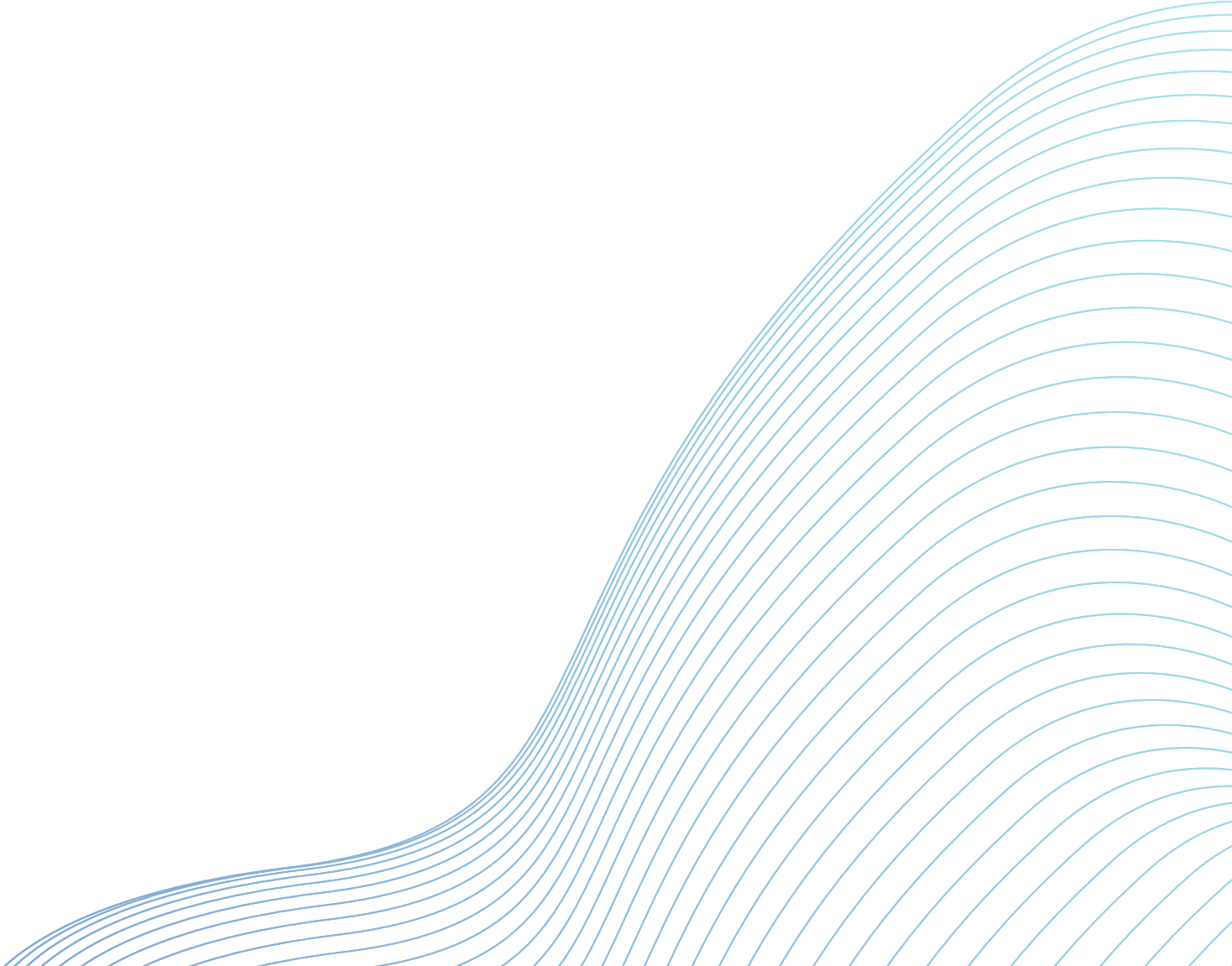


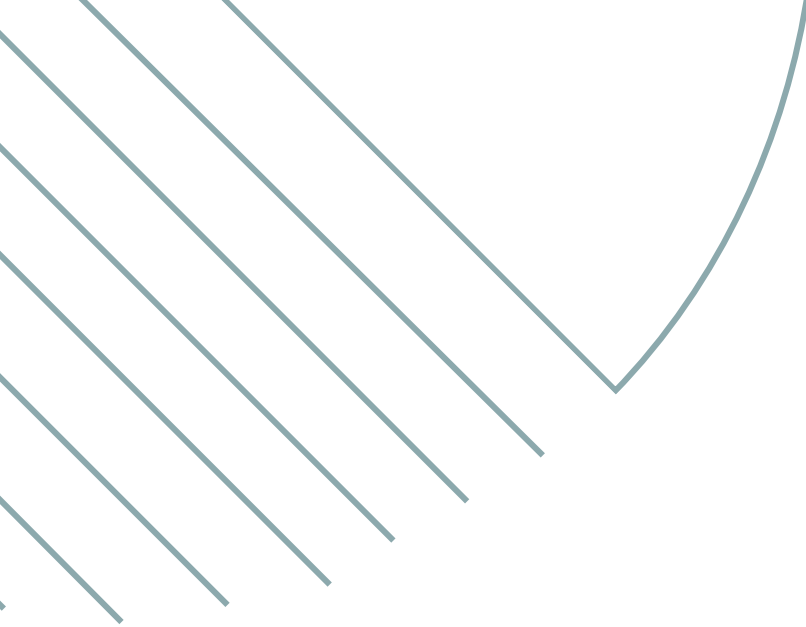
Vũ Hoàng Tài Toàn
HE150224

INTRODUCTION



Table of content

1. Introduction
 2. Approaches to the problem
 3. Proposal method
 4. Data preparation
 5. Results
 6. Demo Code
- 



Approaches to the problem



Approaches to the problem

1. Detect 4 corners of ID card



Approaches to the problem

1. Detect 4 corners of ID card



Approaches to the problem

2. Image processing techniques

- To crop the ID card as a rectangle



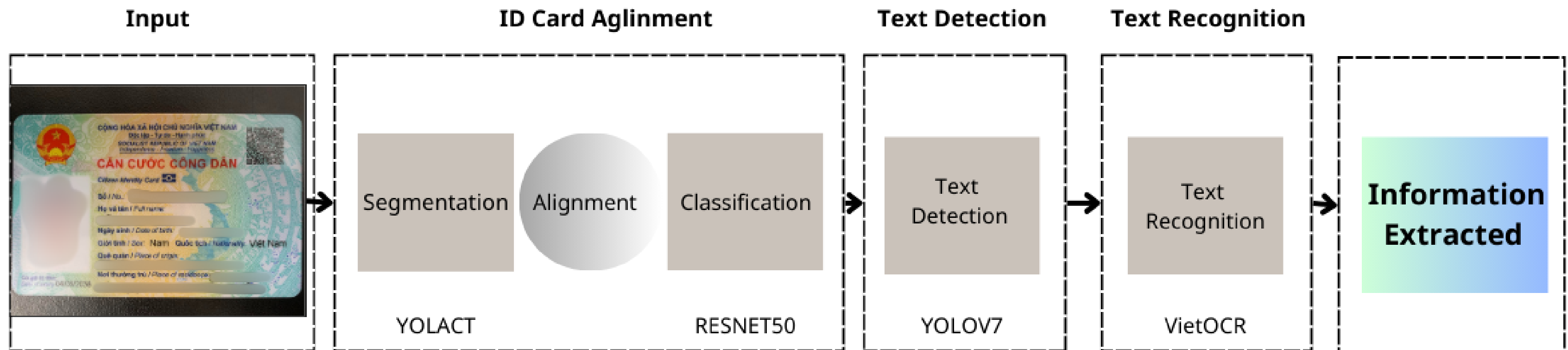
Approaches to the problem

2. Image processing techniques



Approaches to the problem

3. Our proposed method



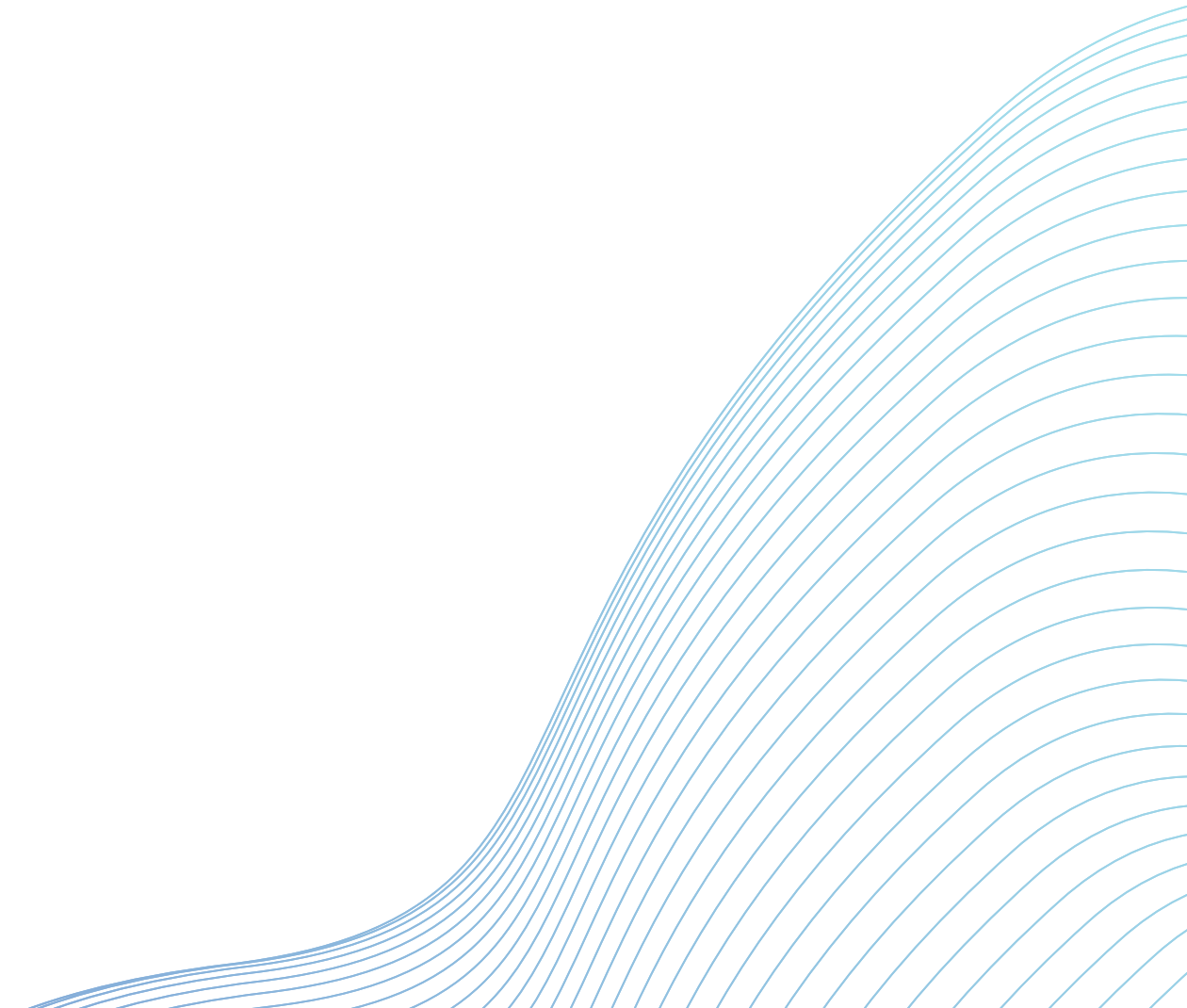


PROPOSAL METHOD

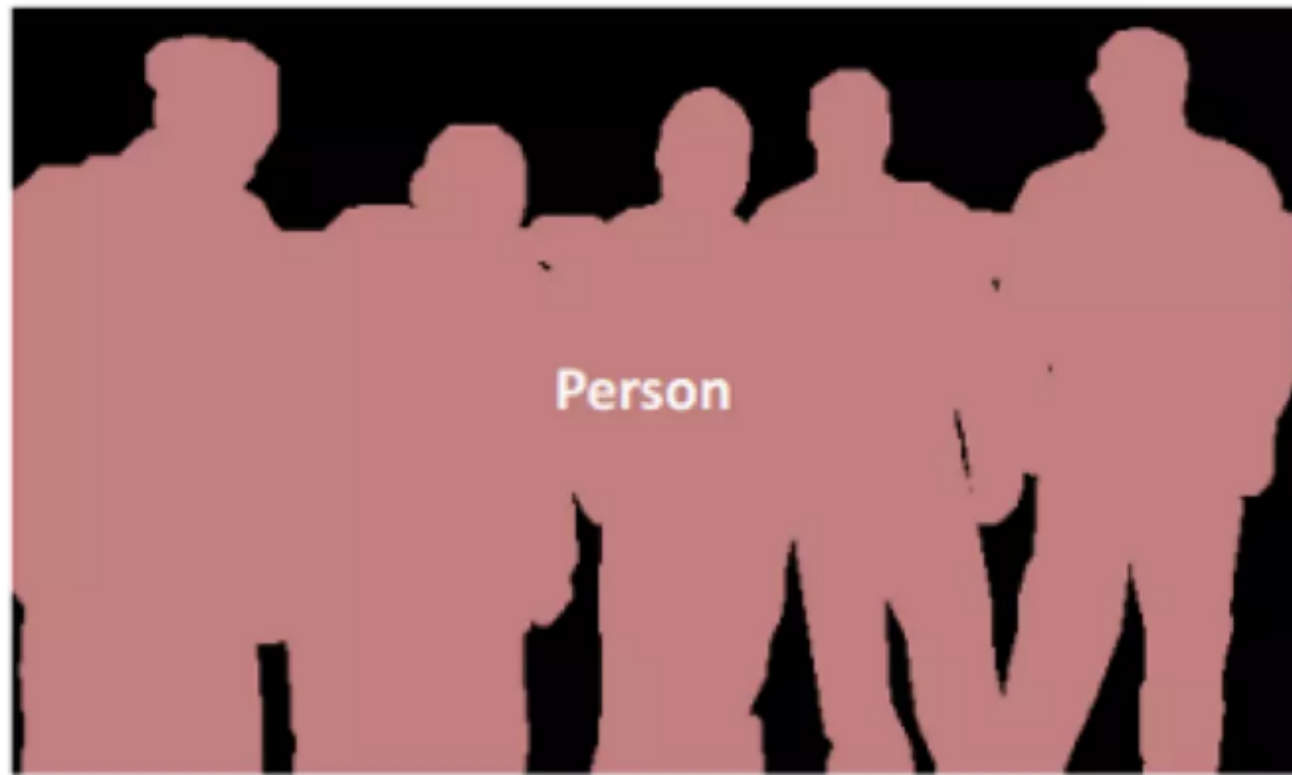
ID CARD ALIGNMENT

ID card alignment

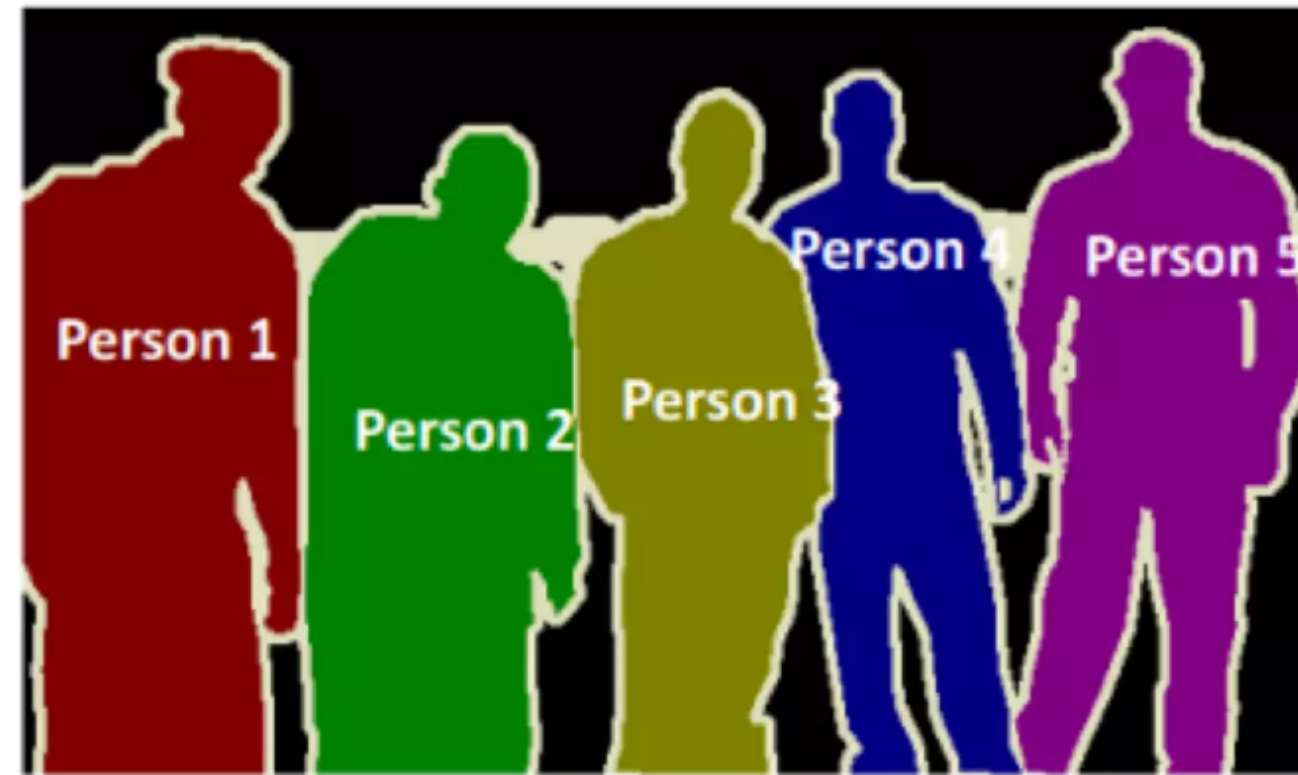
1. Segmentation model
2. Alignment algorithm
3. Classification model



Segmentation model



Semantic Segmentation



Instance Segmentation

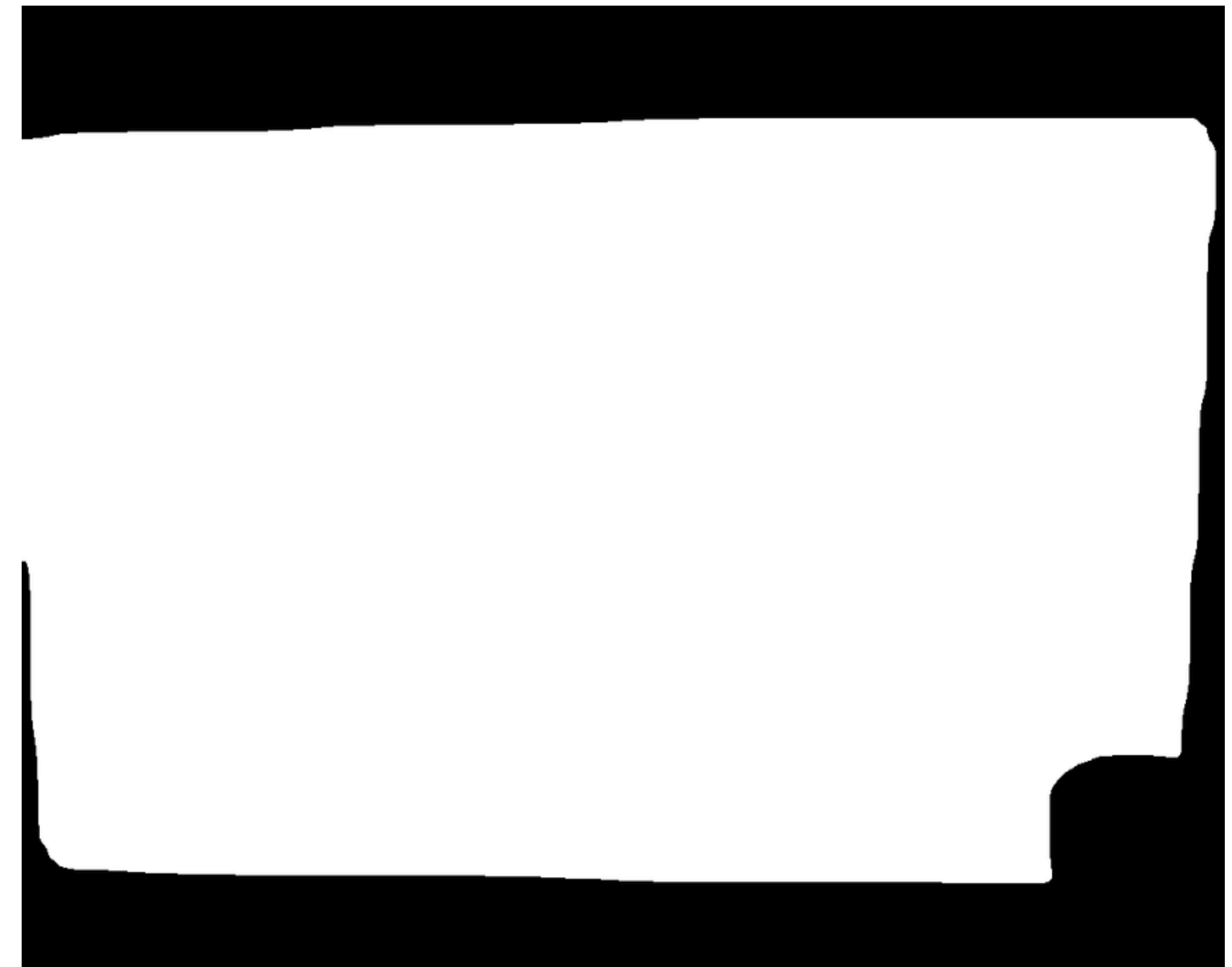
=> Instance segmentation model

Segmentation model

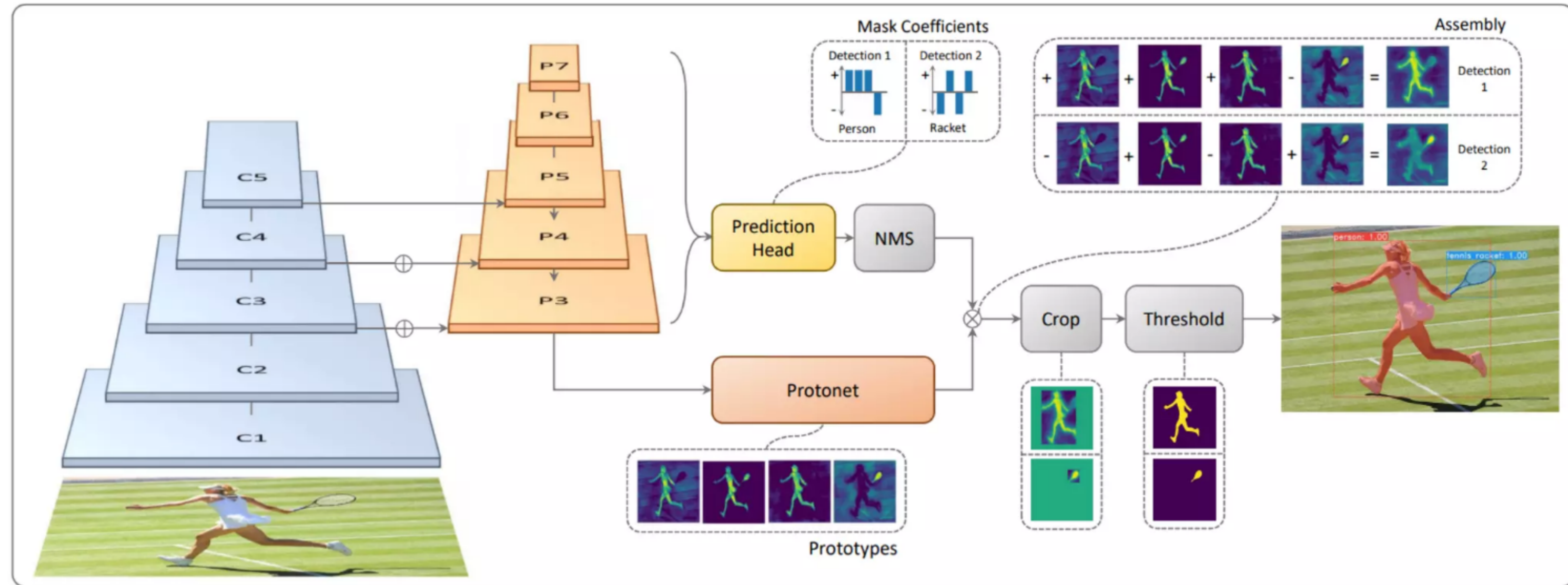
Input



Expected output

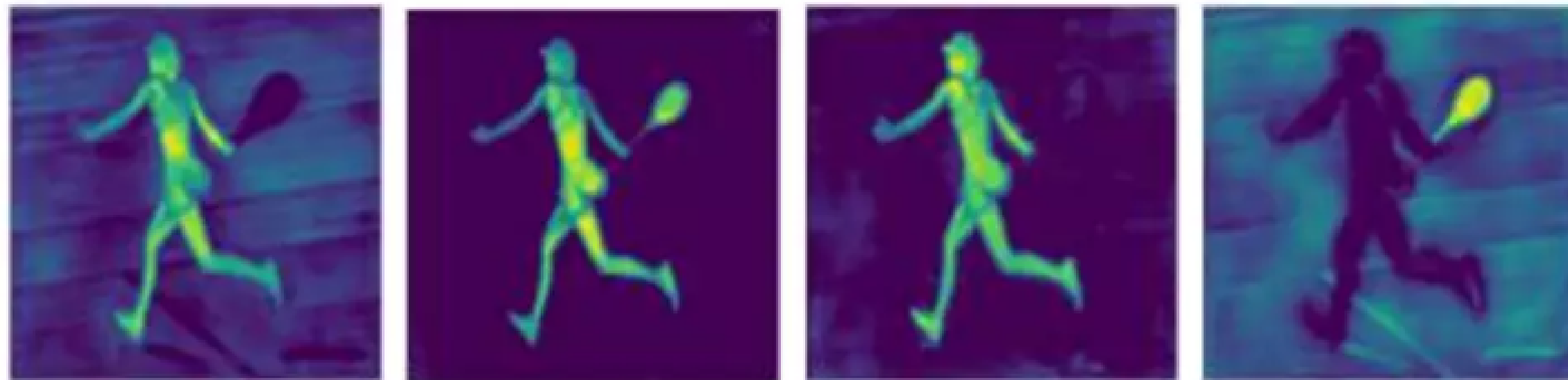


YOLOACT model architecture



Propotype net

- Prototype net is 1 Full-Conv
- Return K prototype mask
- One point to note is that larger K does not mean better output quality

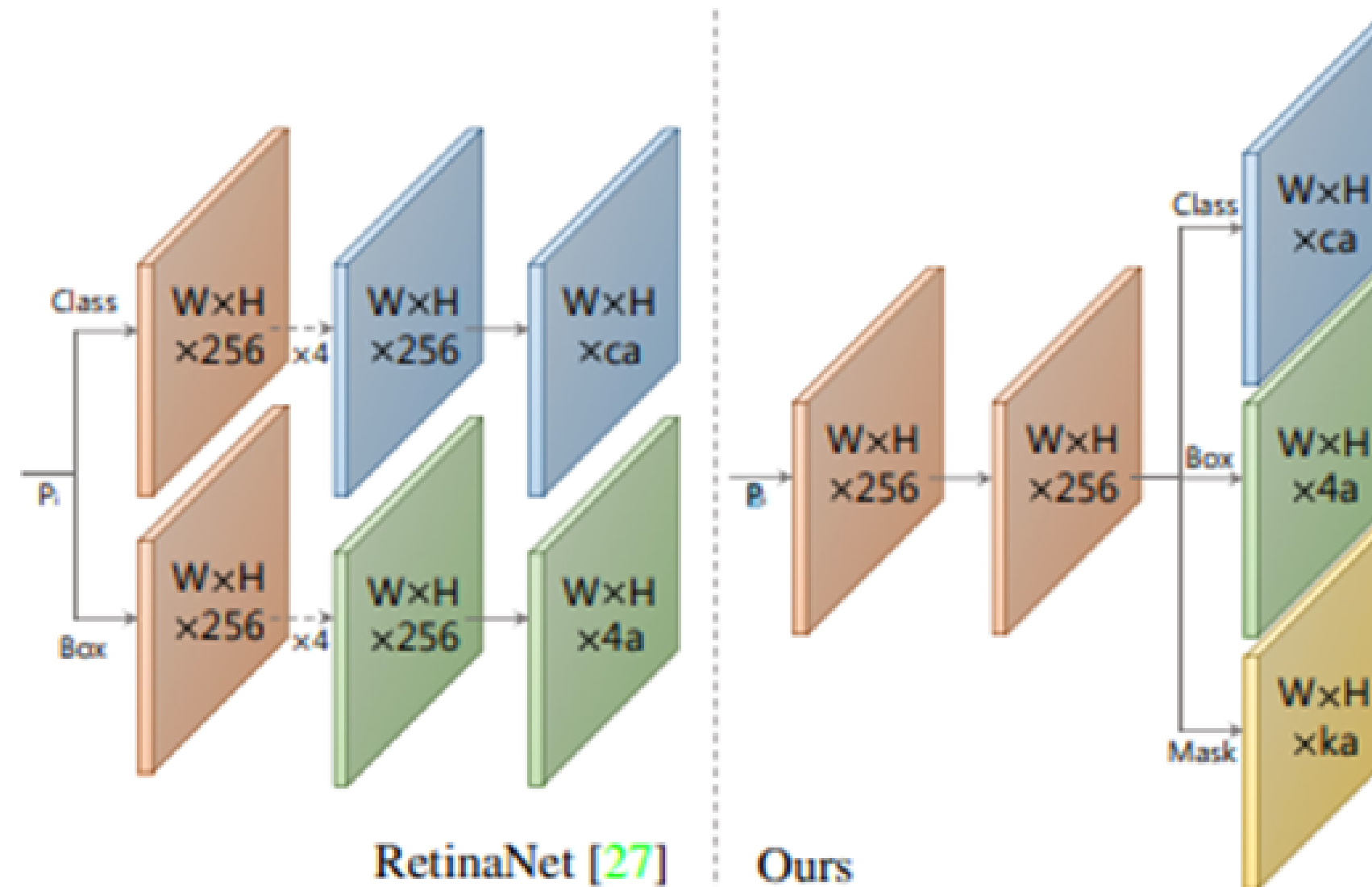


Prediction head

Object detector:

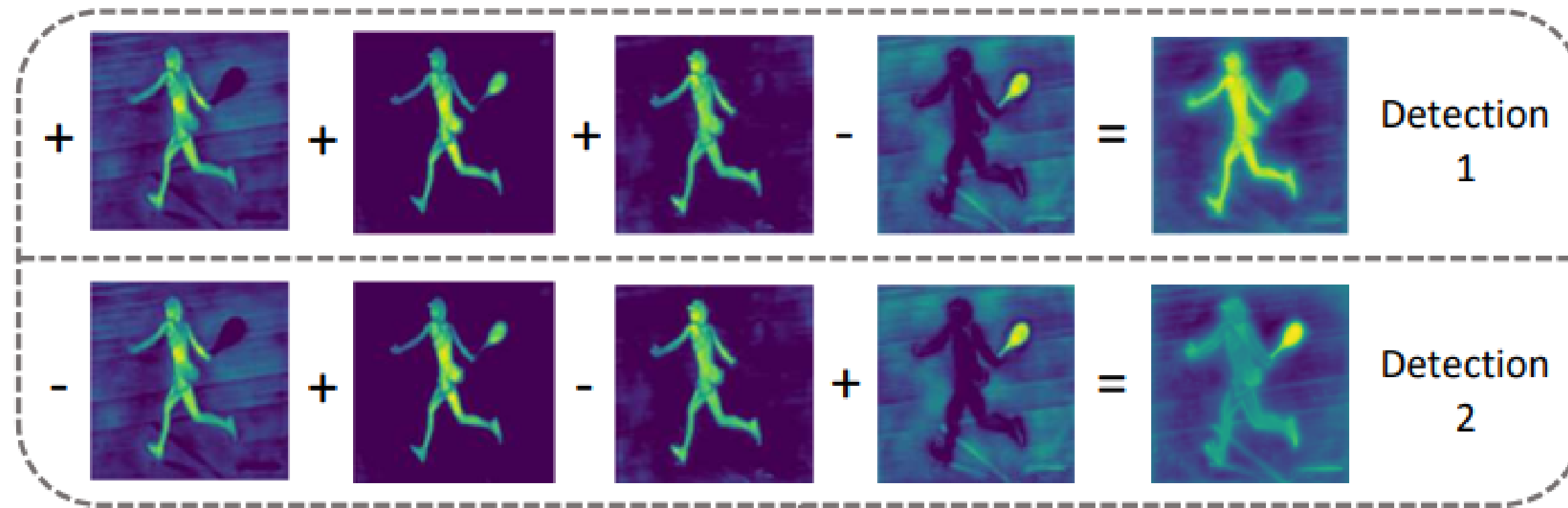
- Predict C class confidences
- Predict bounding 4 offset

Yolact adds a third branch to predict K mask coefficients



Mask Assembly

- Example with $K = 4$: get Prototype Mask as image below and $[1, 1, 1, -1]$ and $[-1, 1, -1, 1]$.



Alignment algorithm

Input



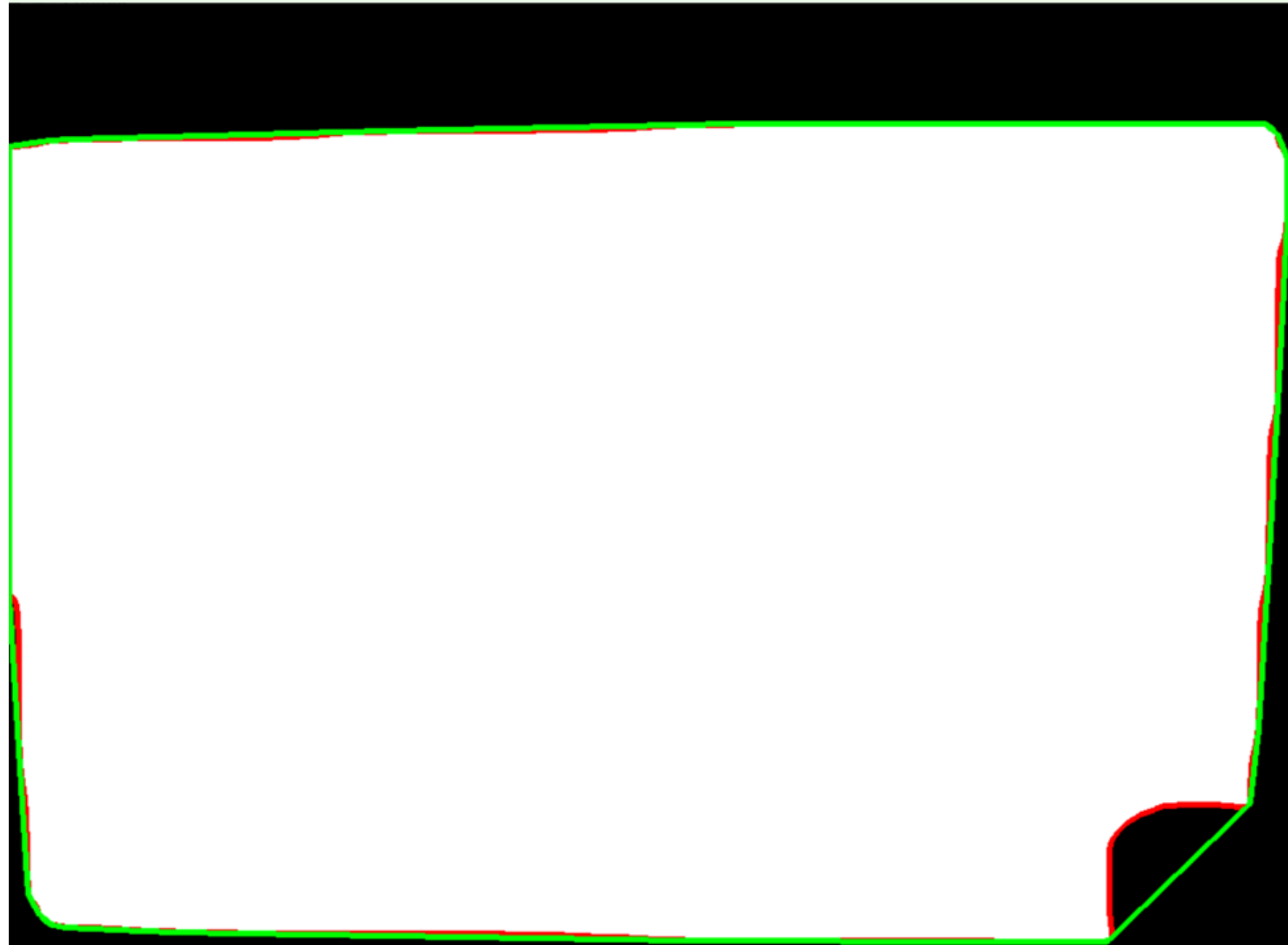
Expected output



Image size 1026x640

Alignment algorithm

- Contour and Convex hull

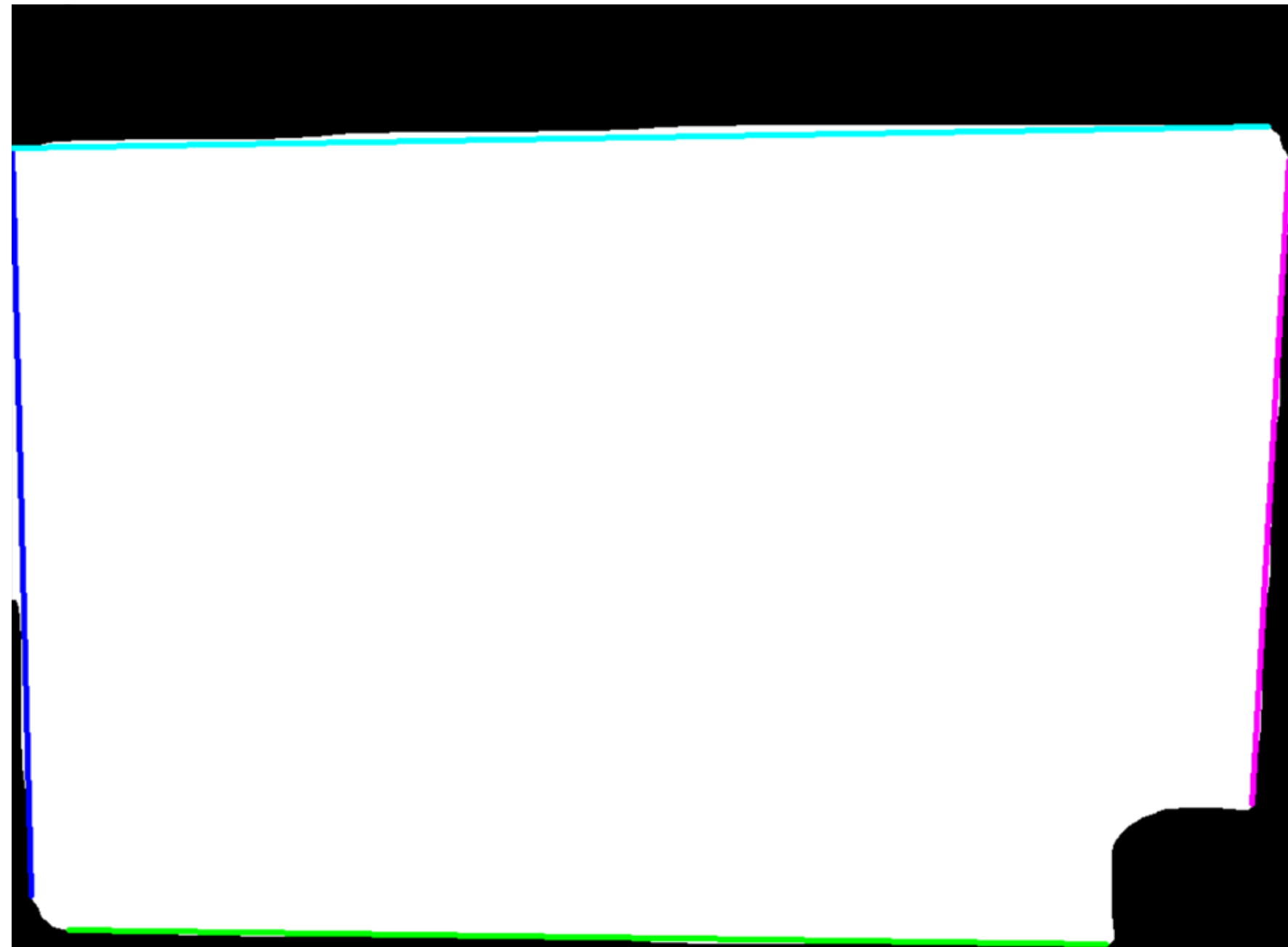


Alignment algorithm

- Remove points on line with this $\cos \theta = \frac{\mathbf{v}_1 \cdot \mathbf{v}_2}{\|\mathbf{v}_1\| \|\mathbf{v}_2\|}$



Alignment algorithm



Alignment algorithm

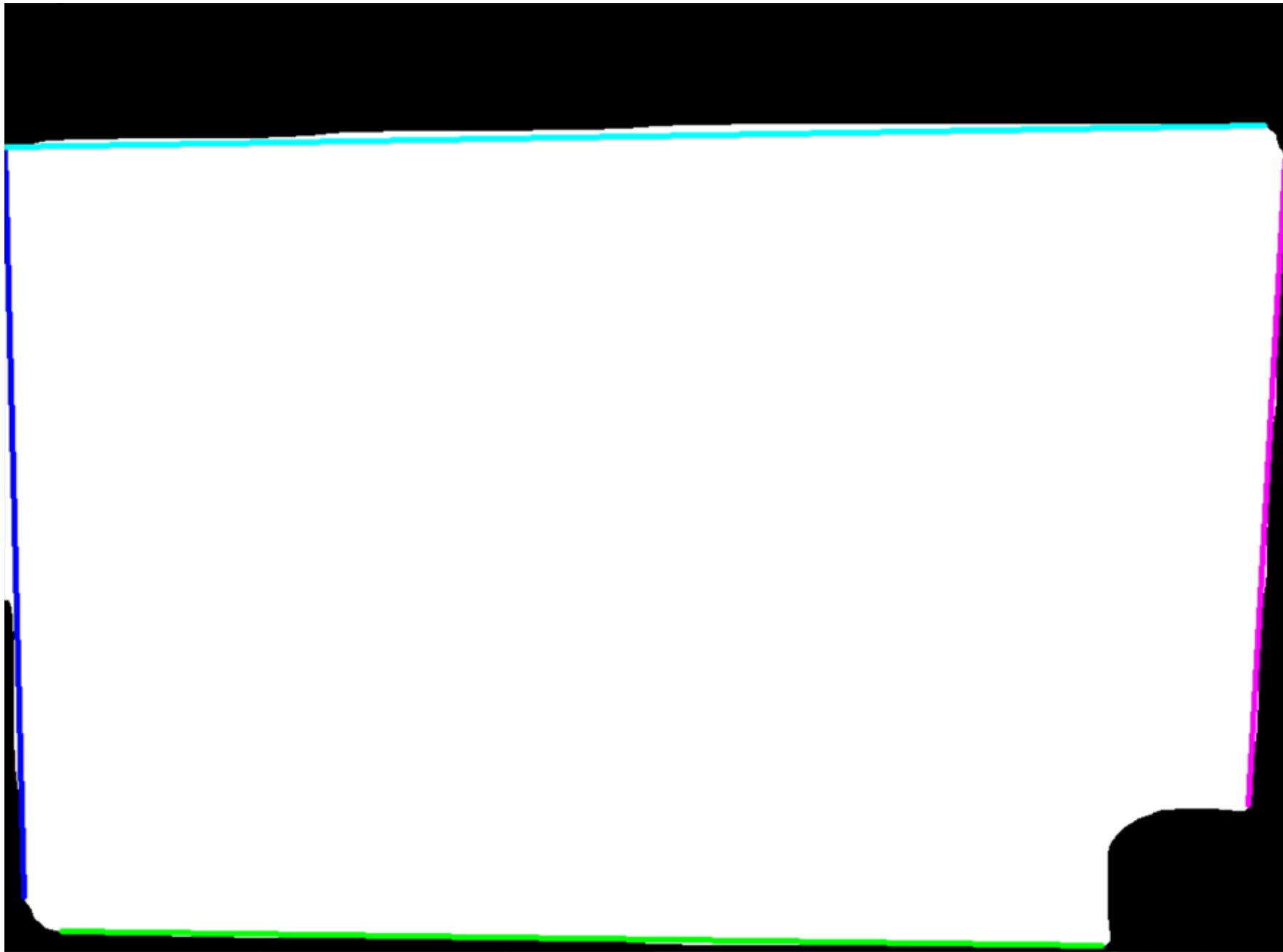
- getPerspectiveTransform

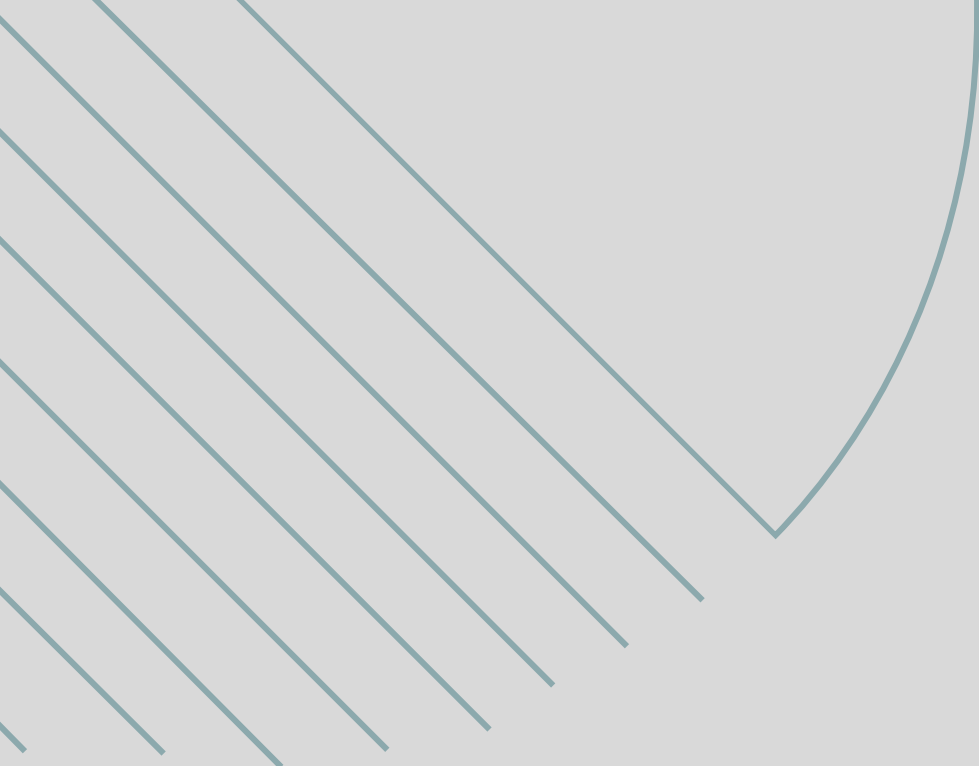


Image size 1026x640

Classification model

- Transfer learning ResNet50 for classify in 2 cases





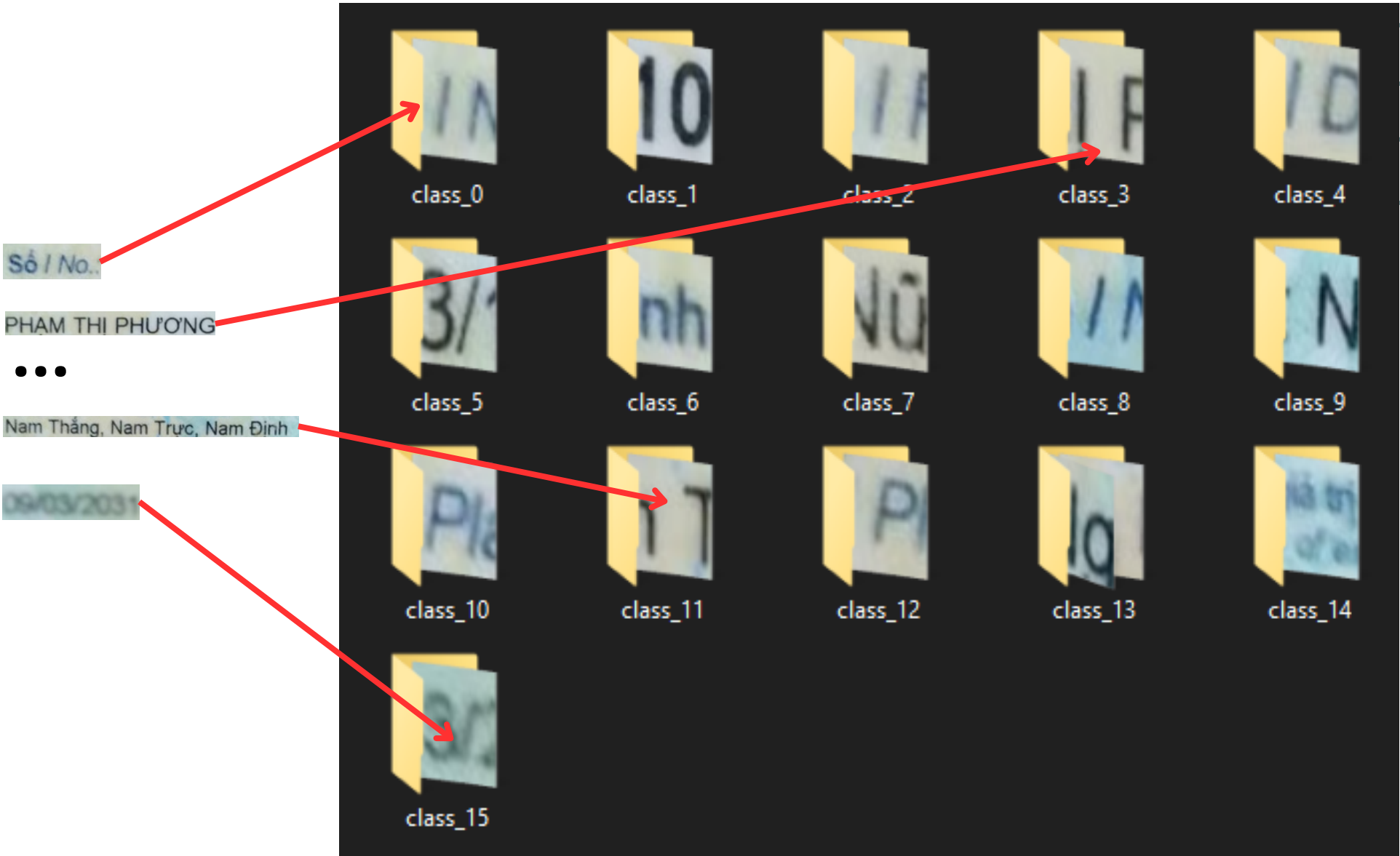
TEXT DETECTION

Input



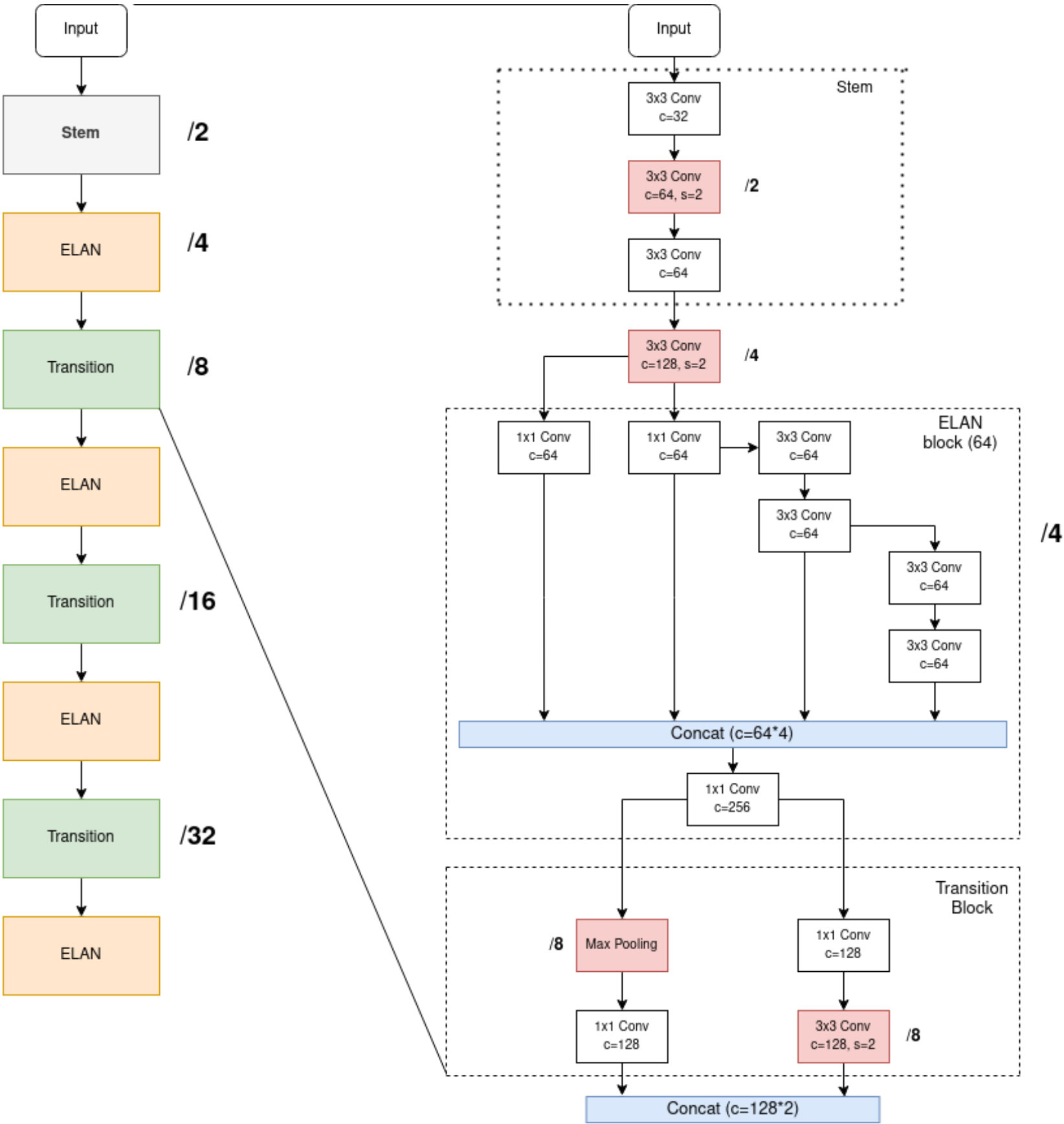
Image size 1026x640

Expected output



YOLOv7's Model Backbone

Backbone



YOLOv7's Model

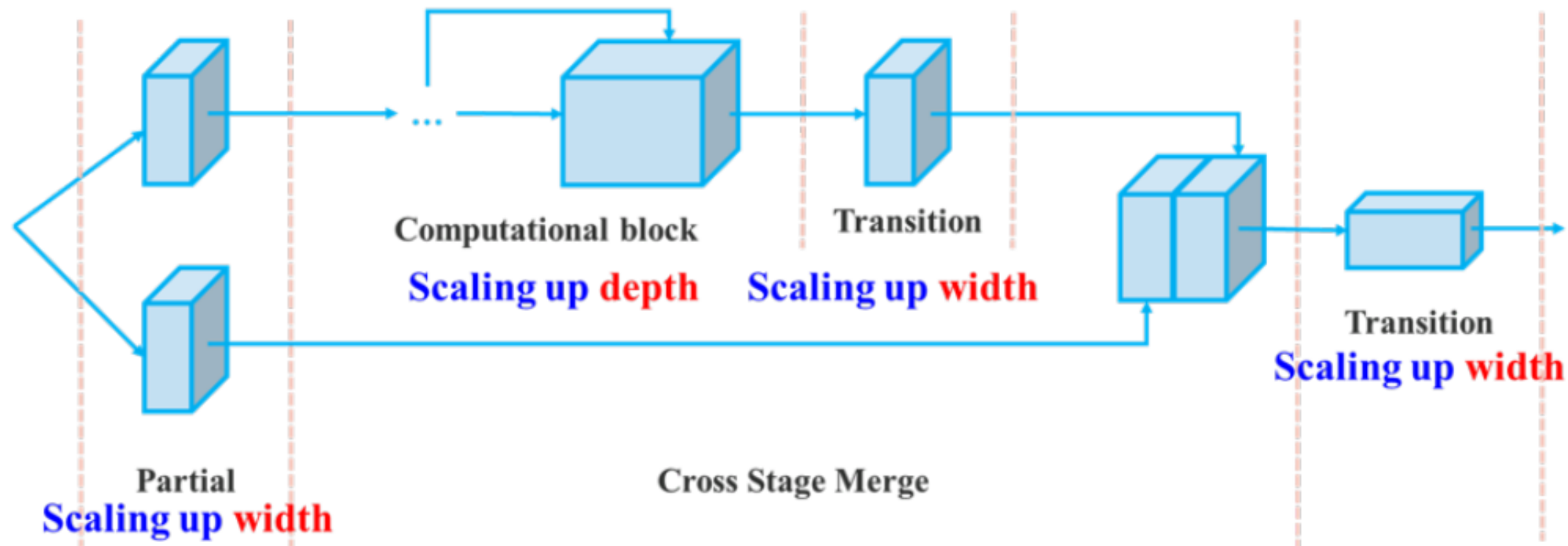
Compound Model Scaling

Scaling up depth: Increase the number of convolutional layers in each stage

Ex: Computational block 2x2 $\longrightarrow$ scaling up depth $\longrightarrow$ Computational block 2x2

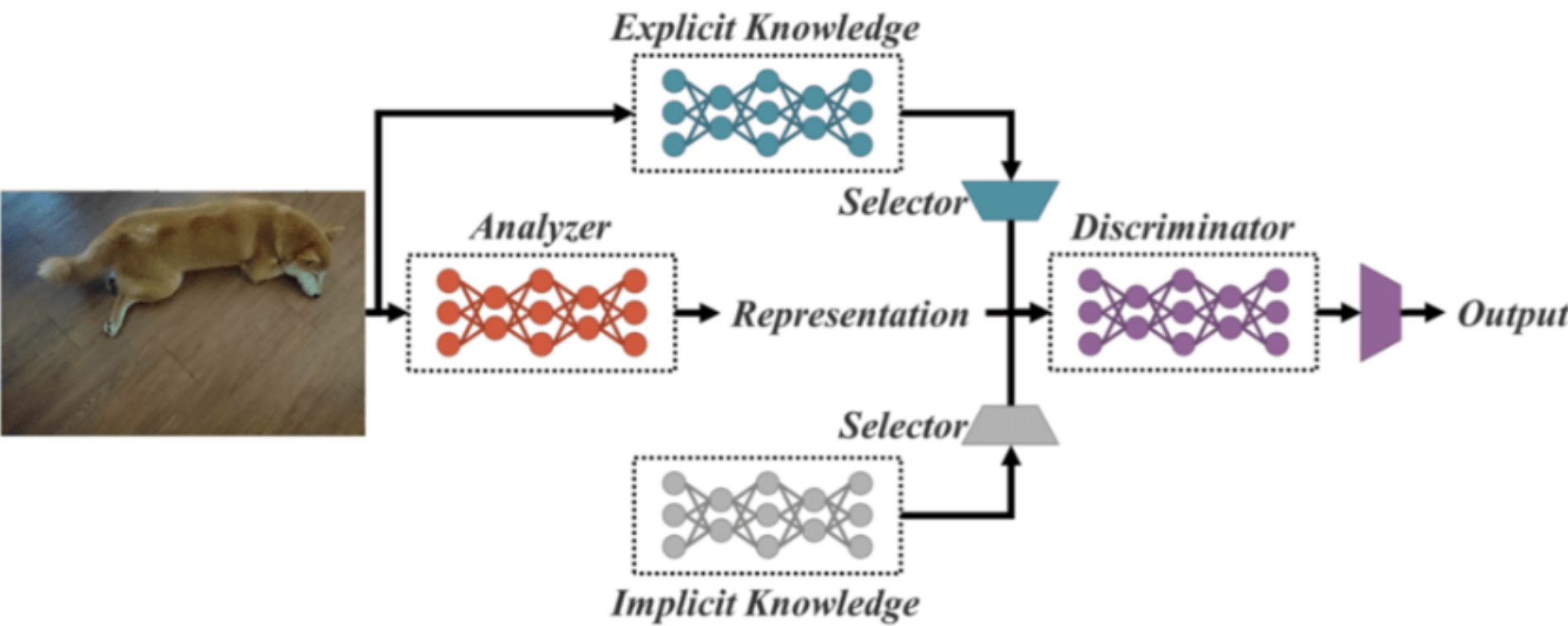
Scaling Up With Cross Stage Merge: Add cross stage merge layers between stages

Ex: scaling up depth $\longrightarrow$ Add cross stage merge layers between computational block
 $\longrightarrow$ cross stage merge combines feature maps from two computational block

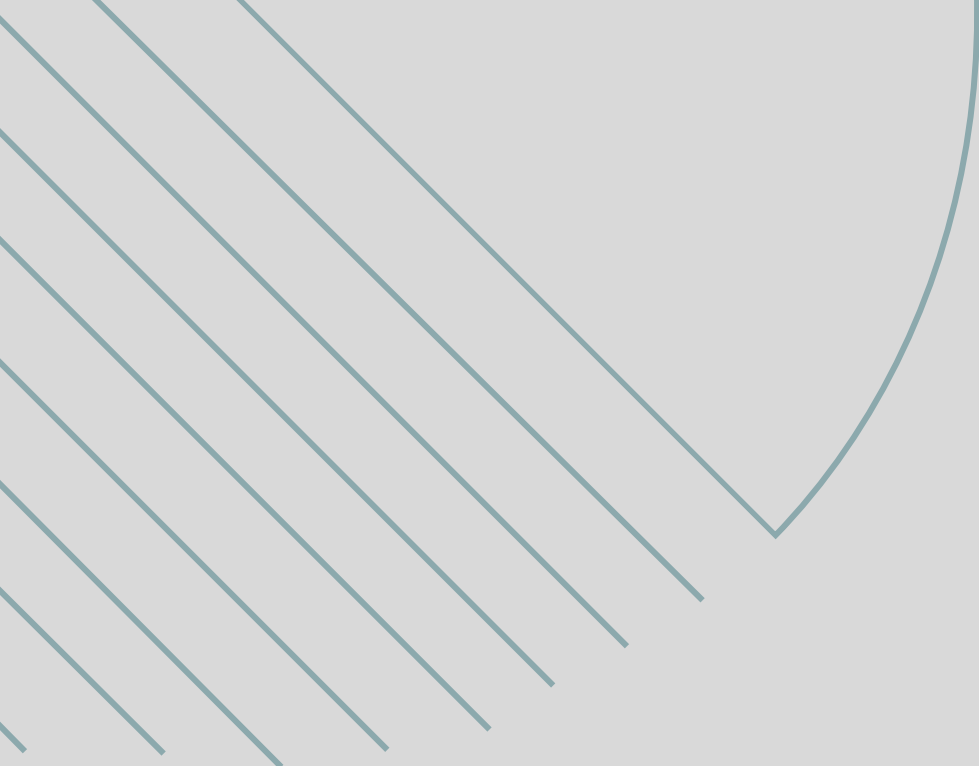


YOLOv7's Model

Implicit Knowledge

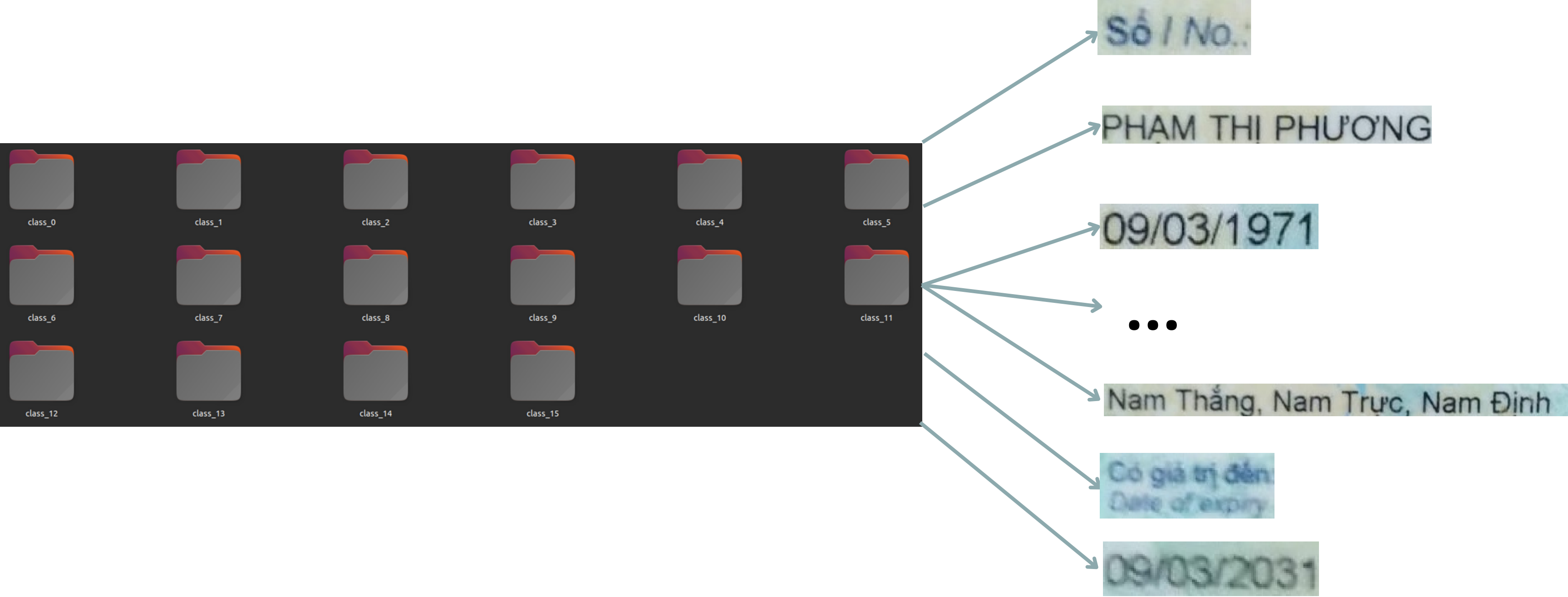


Feature	Implicit Knowledge	Explicit Knowledge
Source of knowledge	Training process	Labeled data
Representation	Vector, neural network, combination	Features
Incorporation methods	Addition, multiplication	Network architecture
Benefits	Learns complex relationships, improves performance	Enables object classification and localization



TEXT RECOGNITION

Input

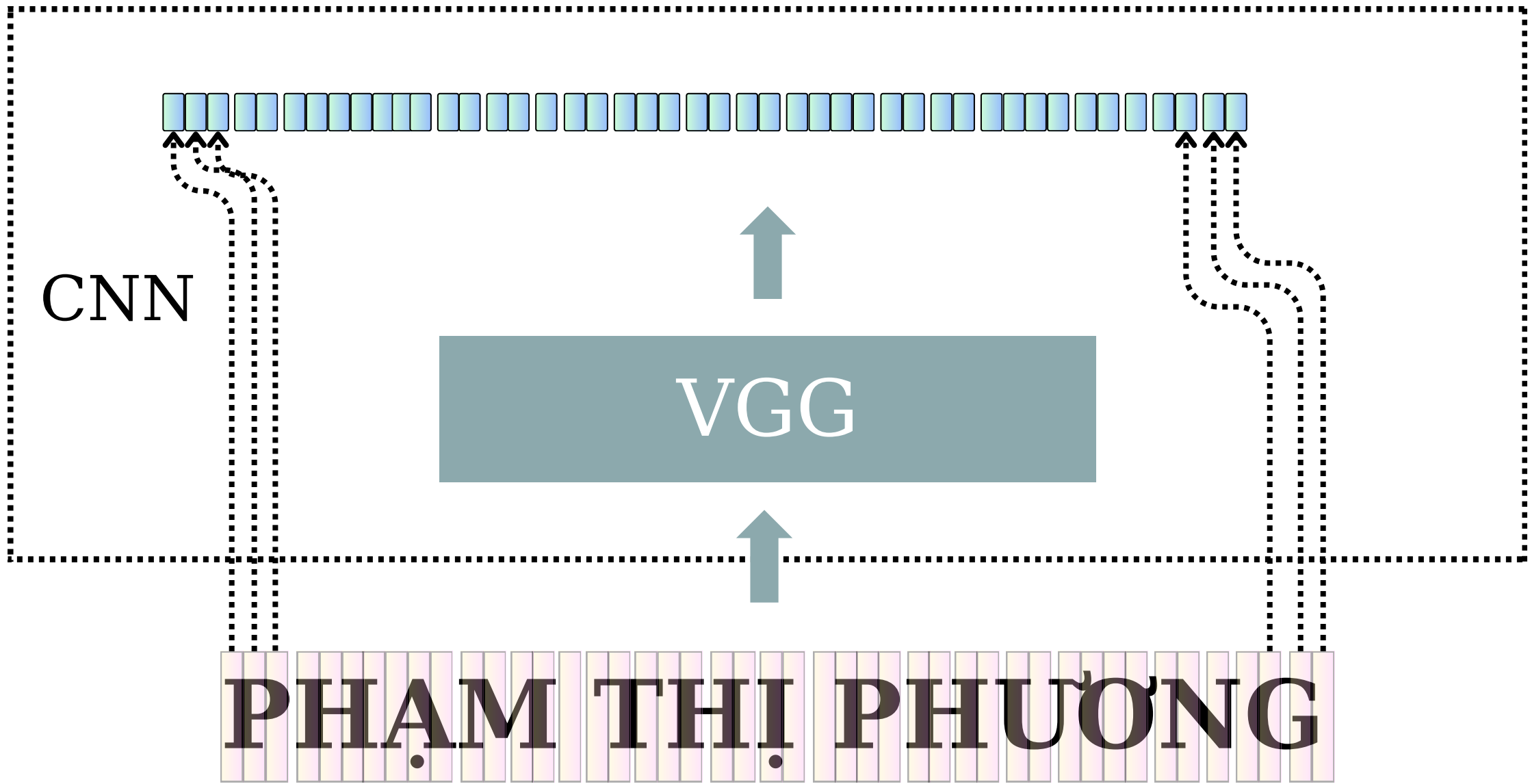


Expected Output

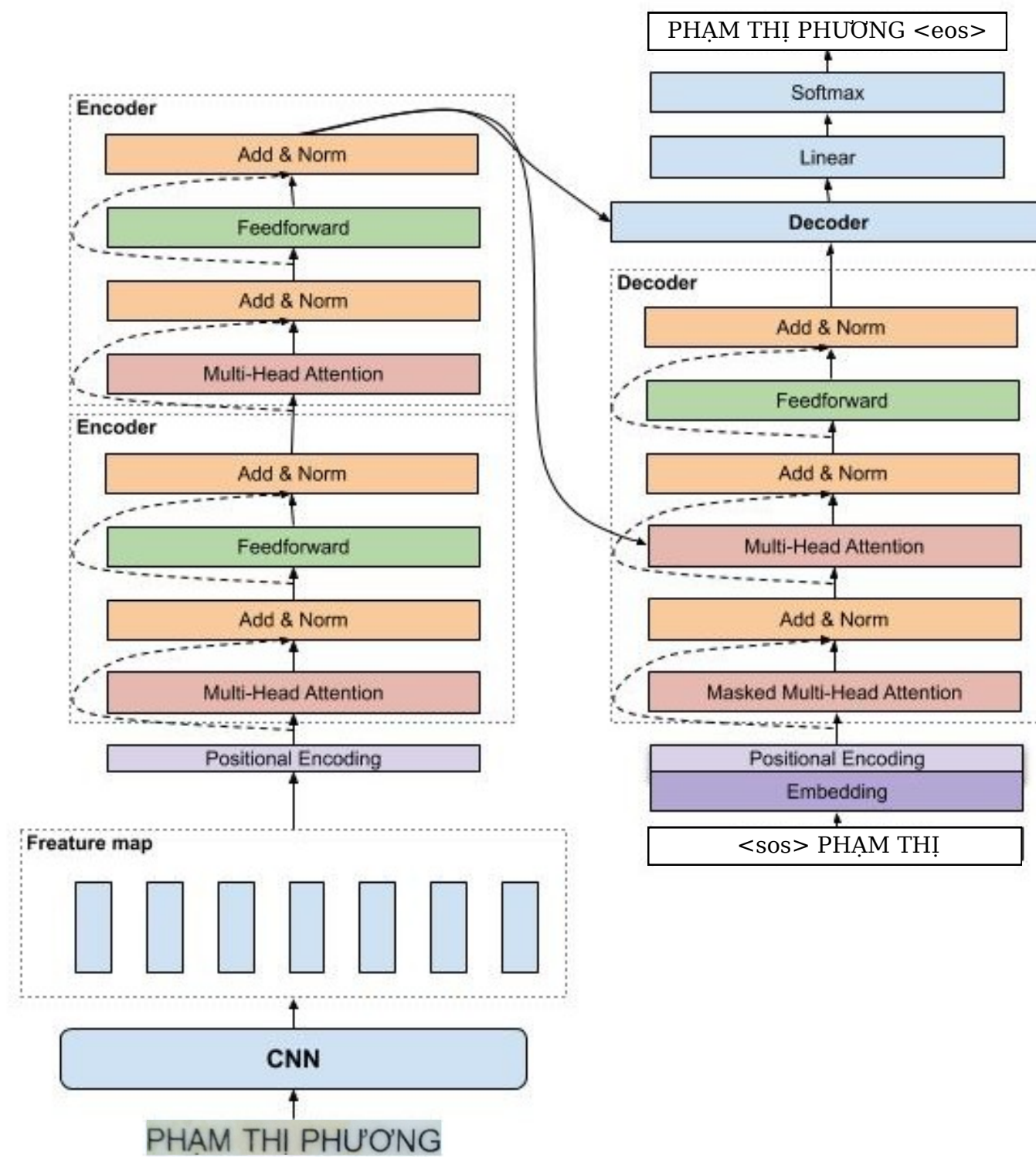


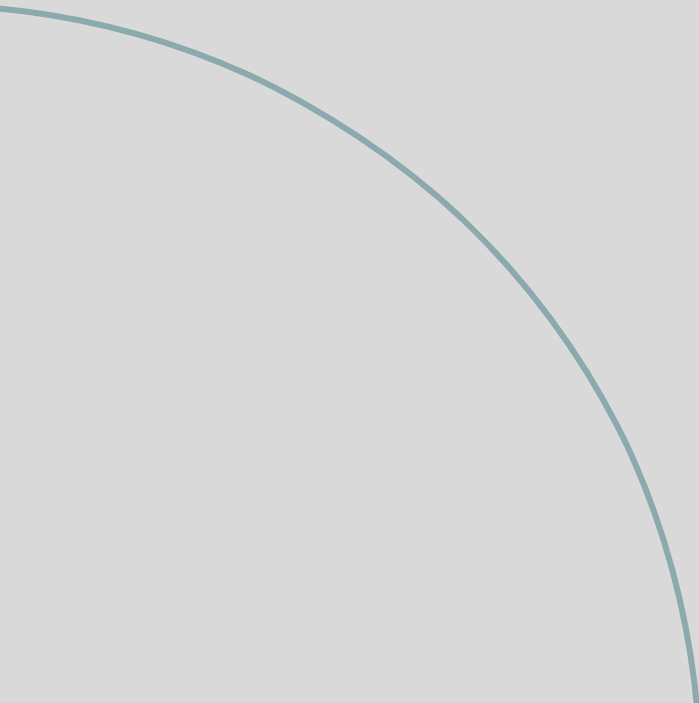
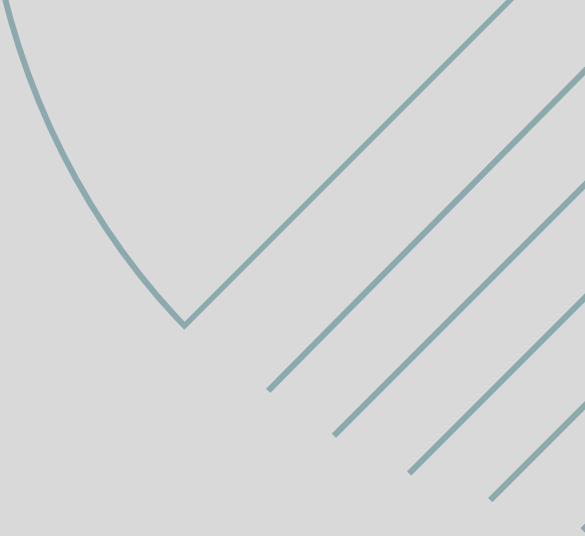
```
Số / No.: 036171002843
Họ và tên / Full name: PHẠM THỊ PHƯƠNG
Ngày sinh / Date of birth: 09/03/1971
Giới tính / Sex: Nữ
Quốc tịch / Nationality: Việt Nam
Quê quán / Place of origin: Nam Thắng, Nam Trực, Nam Định
Nơi thường trú / Place of residence: 12 Ngõ 7, Cầu Đơ 4, Hà Cầu, Hà Đông, Hà Nội
Có giá trị đến / Date of expiry: 09/03/2031
```

VietOCR



VietOCR





DATA

ID Cards Dataset: 644 images, Photos are collected from relatives and social media

Synthetic Images: Half of dataset were rotated ID cards for classification model.

Synthetic Text Images: 500 alignment images for text area detection.

Synthetic Cropped Images: 8000 images, 16 classes for OCR transfer learning.





RESULTS

- YOLACT for instance segmentation
- mAP@[0.5:0.05:0.95]

Saving state, iter: 11000

```
[171] 11000 || B: 0.028 | C: 0.005 | M: 0.087 | S: 0.009 | T: 0.129 || ETA: 11 days, 20:20:14 || timer: 1.971
```

[illegible]

Results

- YOLOv7 for text detection

Class	P	R	mAP@.5	mAP@.5:.95
All	98.5%	96.80%	96.10%	74%
No_Title	100%	99.20%	99.60%	67.30%
No	99.60%	100%	99.60%	81.70%
Name_Title	94.30%	92.60%	90.50%	70.70%
Name	96%	90.90%	88.60%	75.30%
Date_Title	97.70%	96.20%	94.70%	74.90%
Date	99.90%	100%	99.60%	77%
Sex_Title	99.90%	100%	99.50%	75.30%
Sex	99.90%	100%	99.60%	76.80%
Nation_Title	99.50%	98.10%	98.50%	75.50%
Nation	100%	98.10%	98.80%	75.80%
Origin_Title	95.70%	96.20%	93.80%	73.20%
Origin	95.80%	89.30%	88.30%	68.50%
Residence_Title	97.70%	92.70%	91.60%	62.90%
Residence	98.60%	96.20%	96.50%	76.10%
Expiry_Title	99.70%	100%	99.50%	78.10%
Expiry	100%	100%	99.60%	75.40%

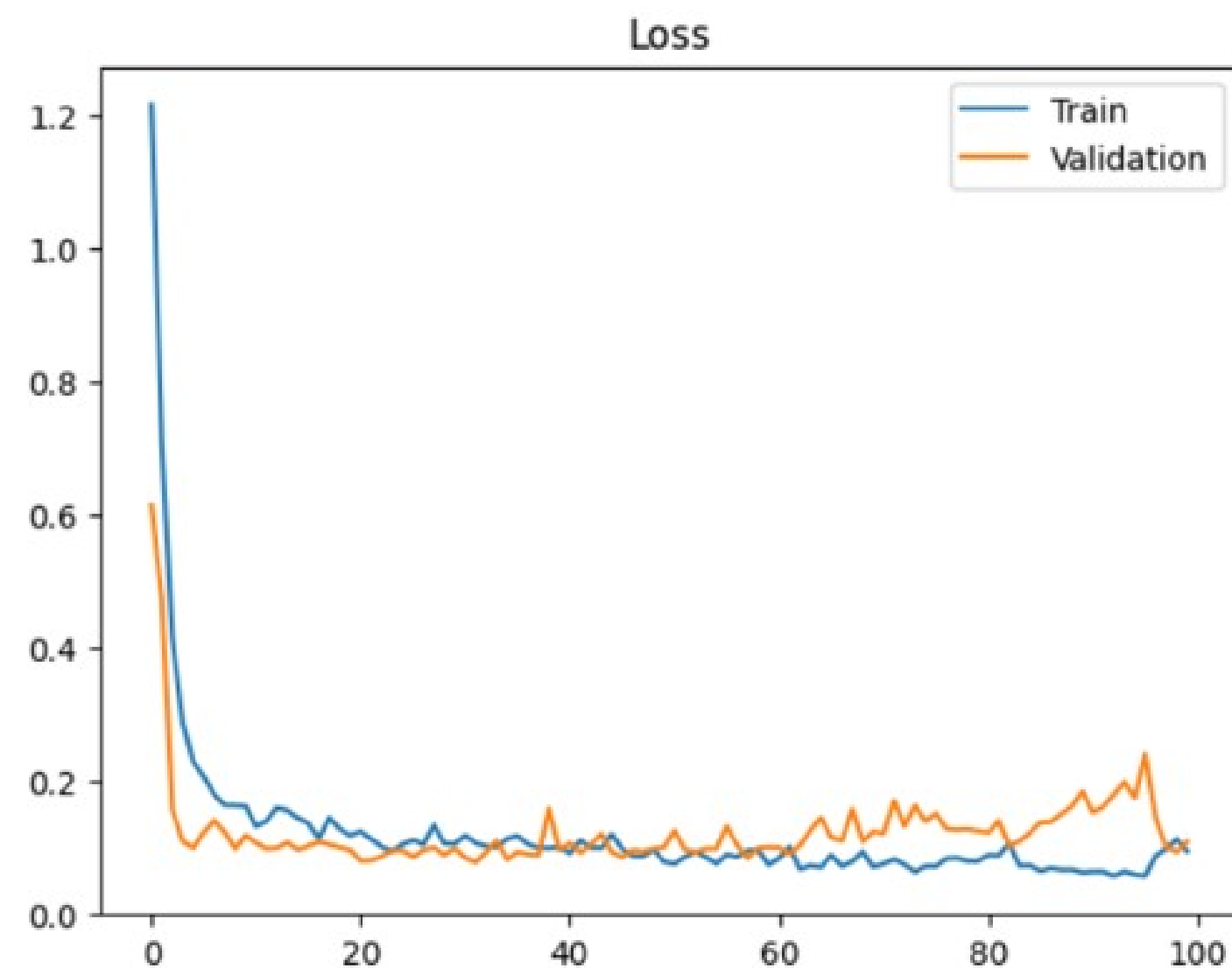
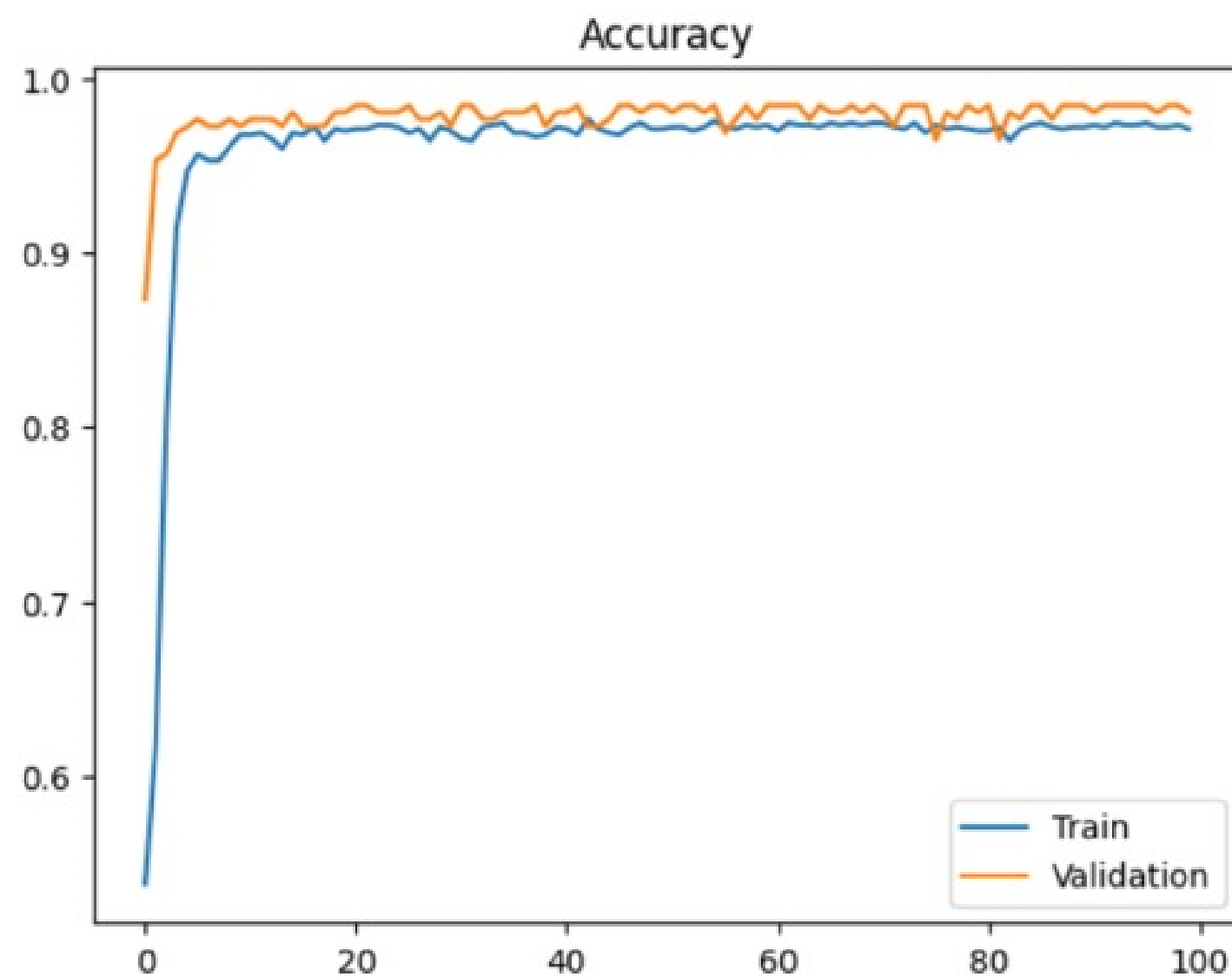
Results

- Transfer learning VietOCR for text recognition

Class	Pre-Train Model	Transfer Learning Model
All	64.23%	98.19%
ID_Title	73.82%	100%
ID	59.45%	97.65%
Name_Title	75.55%	100%
Name	49.83%	95.71%
DOB_Title	76%	100%
Date of birth	57.47%	96.12%
Sex_Title	74.66%	100%
Sex	57.44%	99.29%
Nation_Title	76.39%	100%
Nation	71.85%	100%
Origin_Title	70.06%	100%
Origin	50.24%	89.56%
Residence_Title	71.68%	100%
Residence	49.33%	94.07%
Expiry_Title	72.04%	100%
Expiry	53.17%	98.73%

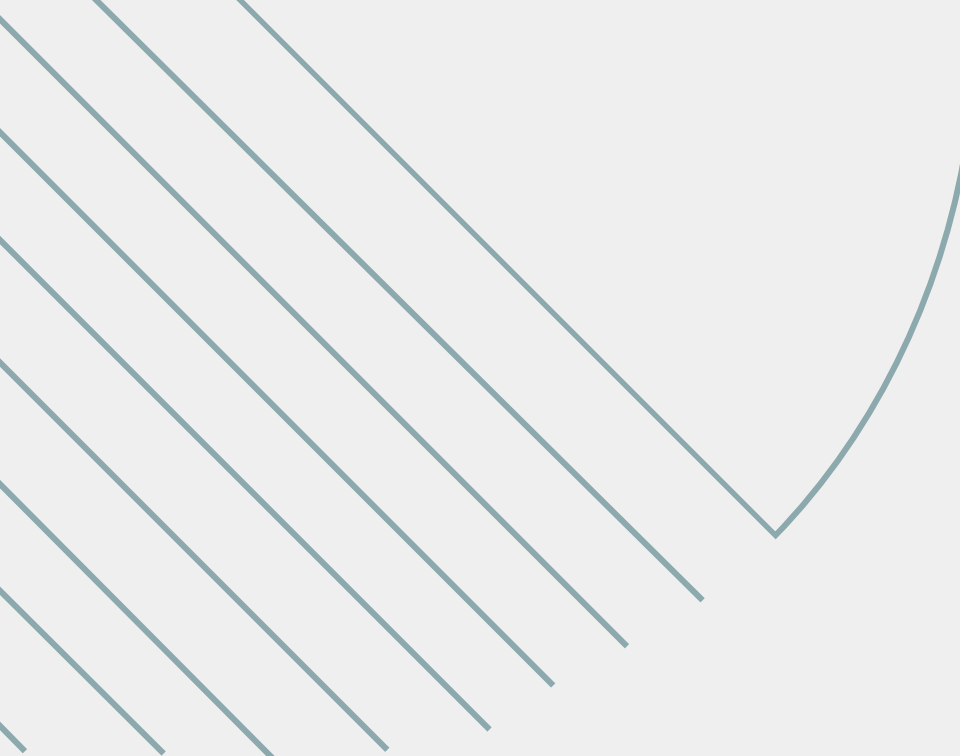
Results

- Transfer learning ResNet50 for classification model



Product Demo





THANK YOU