Identifying ID Submissions using Computer Vision

Improving the customer experience in applying for a new digital bank account

Contents

01

Introduction

Indonesia has a Huge Untapped Banking Market, but ID Verification is Laborious & Time-Consuming

Scenario: A digital bank in Singapore wants to expand its banking business into the Indonesian market

Opportunity:

- **Large under-served market:** 4th most populous nation, 3rd largest unbanked population[1]
- **Young, tech-savvy & tech-hungry population:** median age is 30 years; has the 2nd highest interest level in digital financial services (Southeast Asia)[2]

Challenge:

- **No national digital ID:** for account opening, laborious & time-consuming ID verification must be done with digital submission of physical documents (i.e. photos)

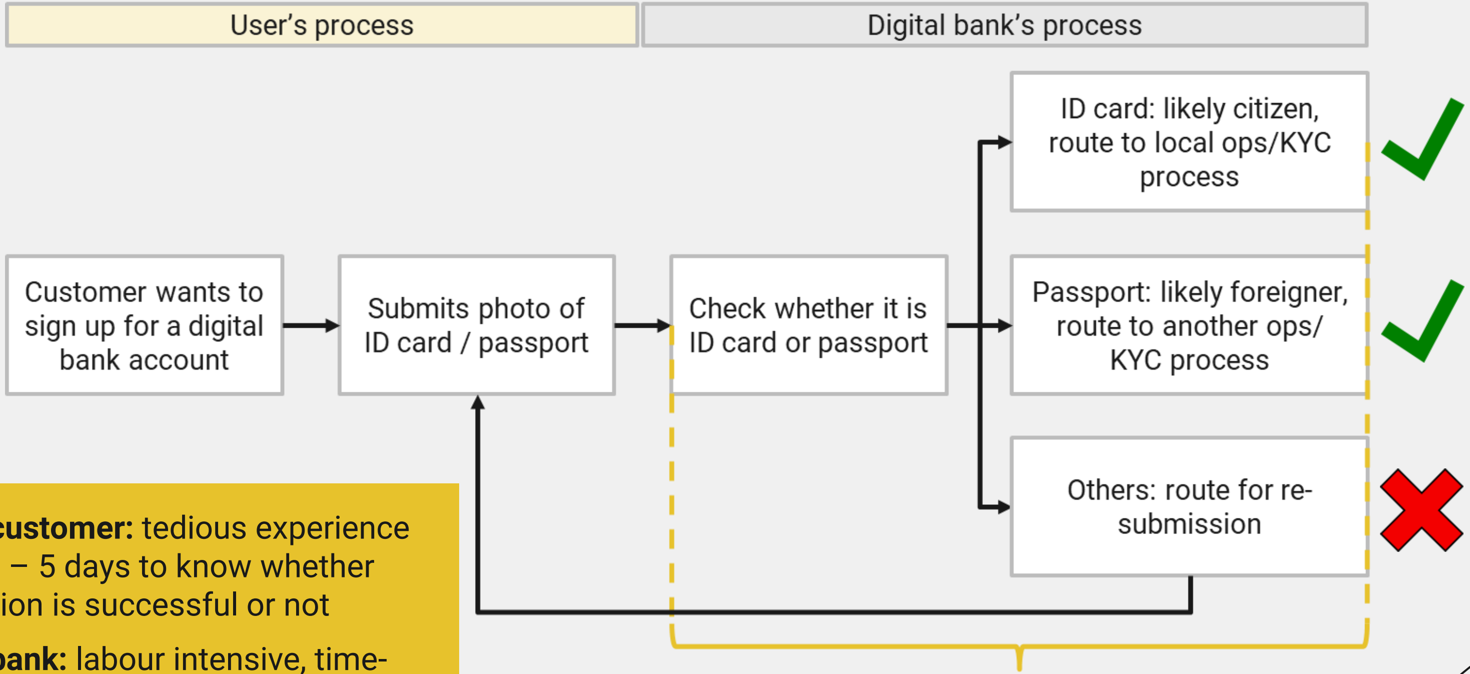
50% (~140m) of population yet to have a bank account[1]

146m people are between 15 – 49 years of age[3]

Problem Statement: manual verification of ID submissions results in poor customer experience

[1] World Economic Forum, Feb 2022, [2] EY, Dec 2022, [3] United Nations Statistics Division Yearbook, 2020

Currently, ID Verification takes: 3 – 5 Working Days

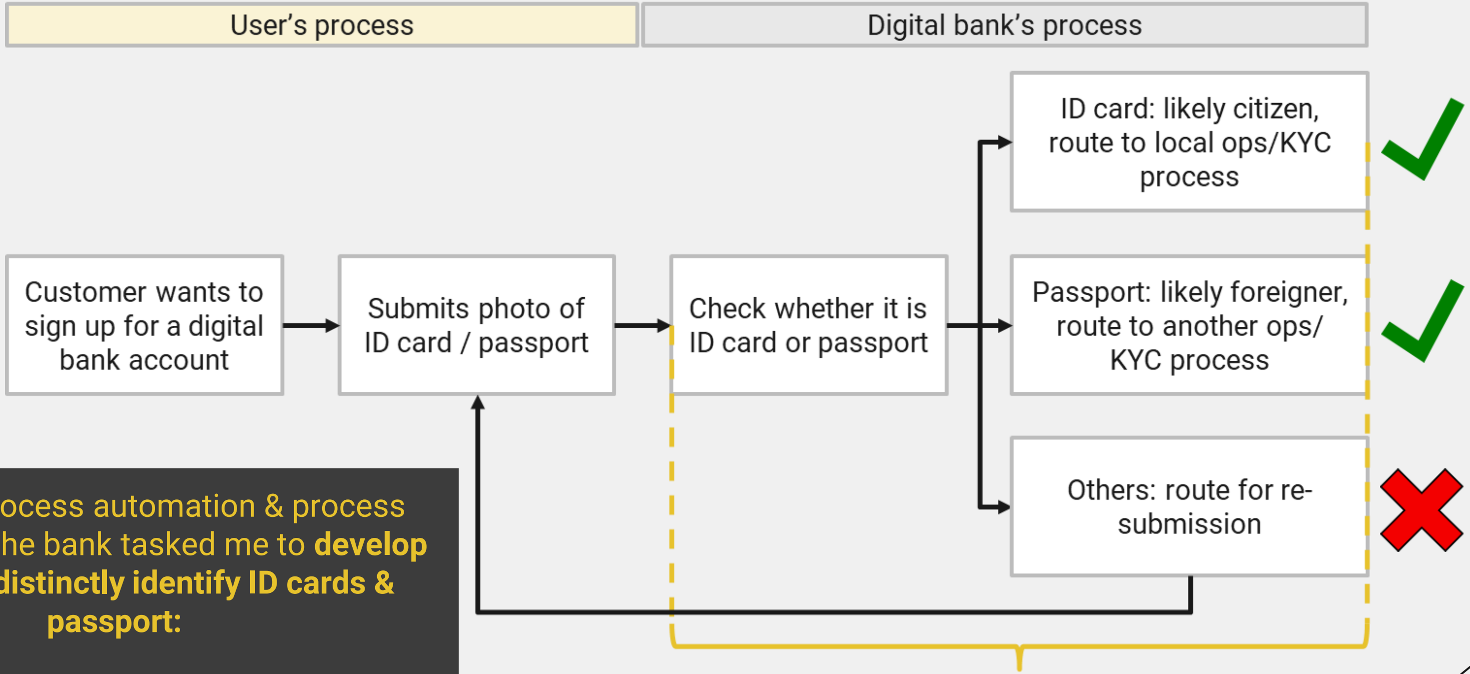


- **For the customer:** tedious experience to wait 3 – 5 days to know whether submission is successful or not
- **For the bank:** labour intensive, time-consuming

Processing time: 3 – 5 working days[4]

[4] as reference: an incorrect photo submission to Singapore's ICA may double the process time (The Straits Times, May 22): 2 weeks for NRIC, 4 weeks for passport (ICA, n.d.)

Goal: Reduce to <30 Seconds for Each Submission



To enable process automation & process acceleration, the bank tasked me to **develop a model to distinctly identify ID cards & passport:**

Multi-class object detection in images

Processing time: <30 seconds

Approach: To Train YOLOv5 Model on ID Documents

01

Label & Analyse Dataset

As proof-of-concept, to use **artificially-generated ID cards and passports** from East Europe, North-West Asia

02

Model Training

Use **YOLOv5** for its state-of-the-art inference speed, good accuracy performance & ease of use

03

Model Deployment via Streamlit

Use **Streamlit to illustrate model prediction capability** – users can choose from a selected set of photos, or upload their own

02

Exploratory Data Analysis (EDA)

Dataset: 1K photos from East Europe & North-West Asia[5]

	Train 600 photos	**Validation** 200 photos	**Test** 200 photos
ID cards	300 from Slovakia, Spain, Finland (100 each)	100 from Estonia	100 from Albania
Passports	300 from Latvia, Russia, Greece (100 each)	100 from Serbia	100 from Azerbaijan

- **Mock ID documents:** from East Europe and North-West Asia, each with unique text fields and artificially generated faces
- **Separated nationalities:** Train, validation & test datasets each have unique nationalities to prevent data leakages
- **Resized:** all photos were resized to 640 x 640 pixels for YOLO

[5] More details in Annex

Dataset: Illustration* (1)

Normal

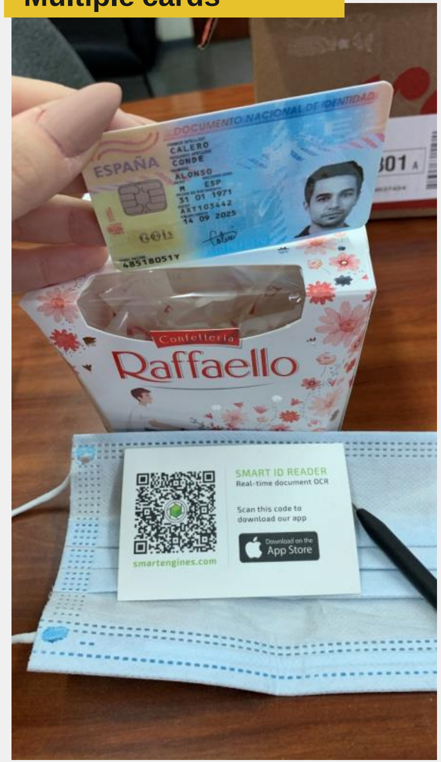
Different background

Different brightness

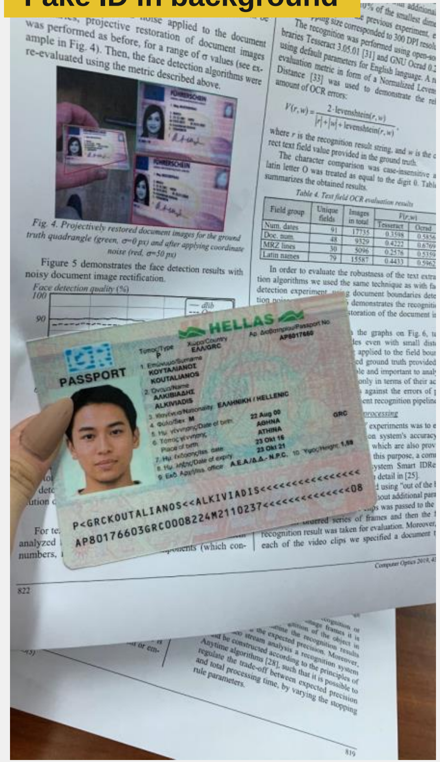
*Illustrated before photo resize to 640 x 640 pixels

Dataset: Illustration (2)

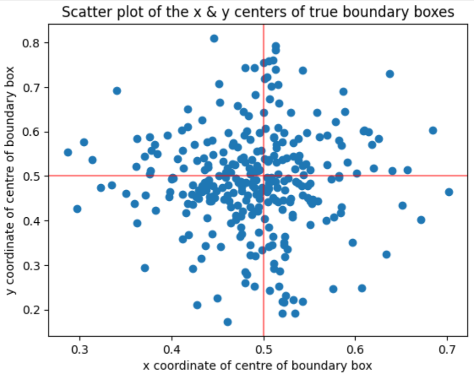
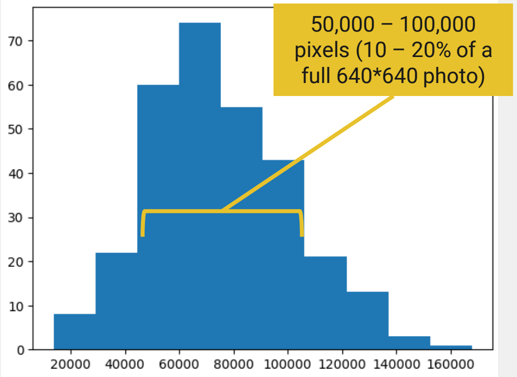
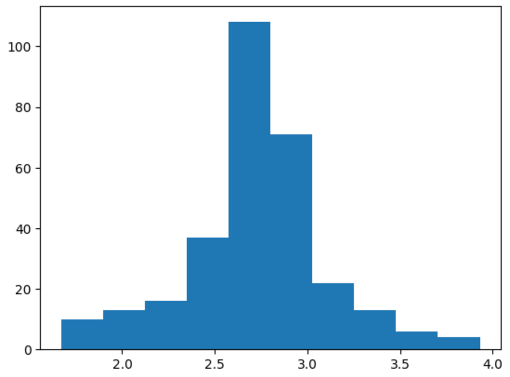
Multiple cards



Fake ID in background



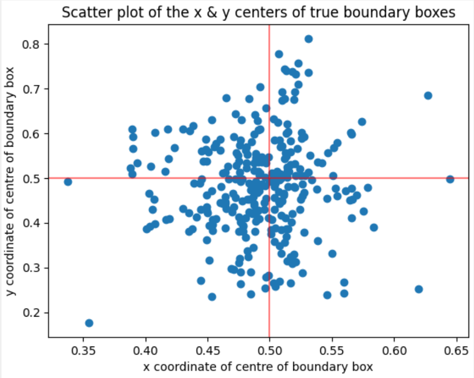
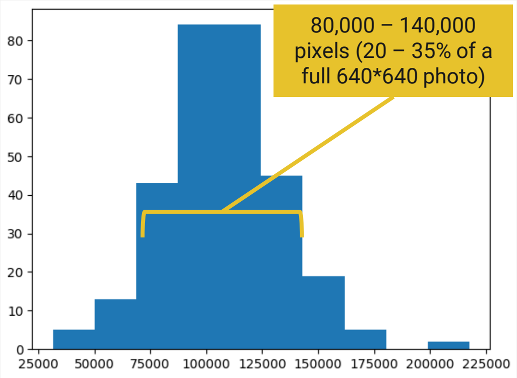
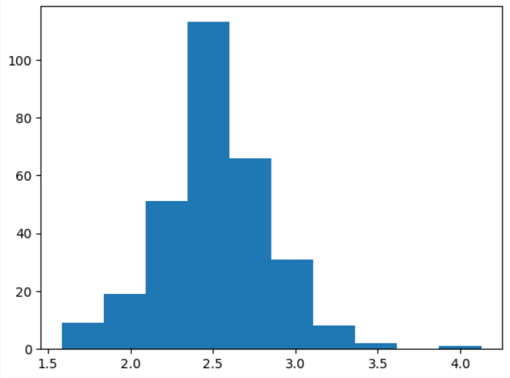
EDA: Training ID Cards are Usually in the Center, 10-20% of Full Photo Size, have Standard ID Aspect Ratio

	Scatterplot of Boundary Box centres	Histogram of Boundary Box Area	Histogram of Boundary Box Aspect Ratio
Graphs			
Analysis	**ID cards usually in centre of photo & along the red lines** → trained model may be **weak in identifying cards in photo corners**	**No samples near 409,600 pixels** (640 * 640) → trained model may be **weak in identifying close-up photos of ID cards**	**Most hover between 2.5 – 3.0** (standard is 2.6 for an ID card*) → trained model may be **weak in identifying rotated ID cards**

*ID card's standard aspect ratio of 2.6 is after resizing photos to 640 * 640 pixels

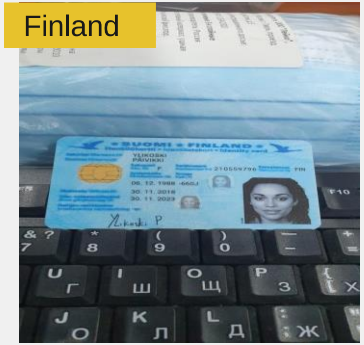
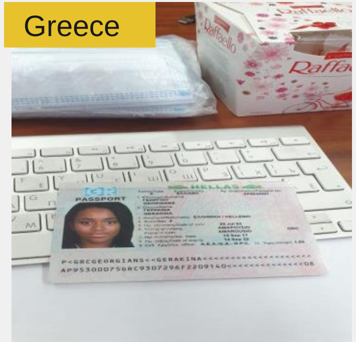
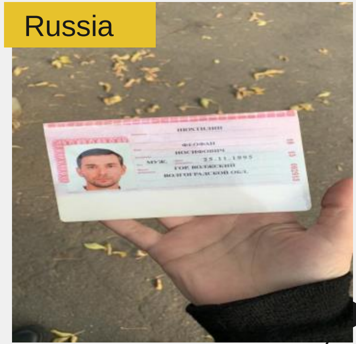
EDA references: Neptune.AI, Kaggle samples

EDA : Training Passports are Similar to ID Cards

<table>
<tr><th></th><th>Scatterplot of
Boundary Box centres</th><th>Histogram of
Boundary Box Area</th><th>Histogram of
Boundary Box Aspect Ratio</th></tr>
<tr><td>Graphs</td><td>
</td><td>
</td><td>
</td></tr>
<tr><td>Analysis</td><td>Passports usually in centre of photo & along the vertical red line → trained model may be weak in identifying passports in left, right & photo corners</td><td>No samples near 409,600 pixels (640 * 640) → trained model may be weak in identifying close-up photos of passports</td><td>Most hover around 2.5 (standard is 2.6 for an passport*) → trained model may be weak in identifying rotated passports</td></tr>
</table>

*Passport's standard aspect ratio of 2.6 is after resizing photos to 640 * 640 pixels

EDA : Training ID Cards, Passports’ Colour Themes

ID Cards: **Blue, Yellow**	Spain 	Finland 	Slovakia 
Passports: **White, Pink**	Latvia 	Greece 	Russia 

EDA Takeaway: to explore **data augmentation for training dataset** to increase sample variations

03

Modelling & Deployment

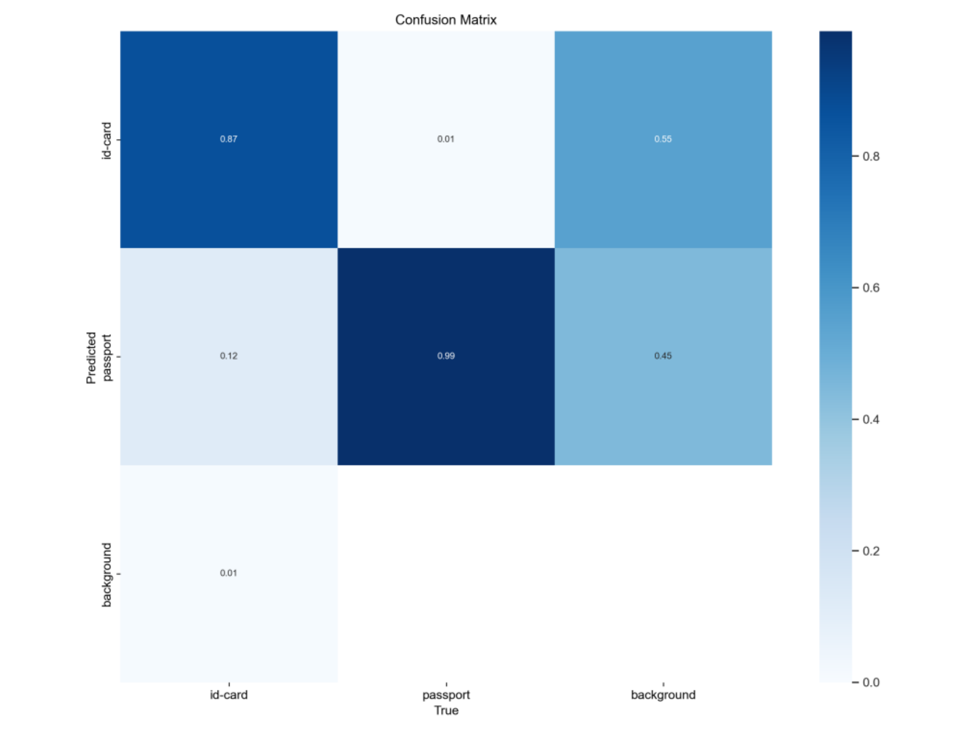
Modelling Overview

	Baseline Model	Improved Model 1	Improved Model 2	Improved Model 3
Dataset	1000 photos (600 train, 200 validation, 200 test)			1600 photos (1200 train, 200 validation, 200 test)
Augmentation on training data	None	▪ Colours: hue, saturation, brightness[6] ▪ Image translation, scale, flip left-right, mosaic[7]	▪ Colours: hue, saturation, brightness ▪ Image translation, scale, flip left-right, mosaic[7] ▪ Rotation, shear[7]	▪ Colours: hue, saturation, brightness ▪ Image translation, scale, flip left-right, mosaic[7]
Model training, validation	**YOLOv5 Large**[8,9]: ▪ Batch size = 8 (limited by hardware[10]) ▪ Epochs = 100 ▪ Starting weights = pretrained weights on COCO 2017			
Model testing	▪ Best weight from training-validation ▪ Confidence threshold = 0.7 ▪ With test-time augmentation			

[6] Detailed explanation of colour augmentation in annex
[7] Detailed explanation of dimensional augmentation in annex
[8] YOLO performance over other models in annex, YOLOv5l performance over other variants in annex
[9] Reference for training parameters for small dataset by Ultralytics, May 2022
[10] Hardware used: Intel i5-12400F, RAM 16GB, Nvidia RTX 3060

Baseline Model

Validation Scores



Test Scores

	ID Cards	Passport
Correctly identified	**99**	**62**
Wrongly identified as the other class	**1**	**25**
Wrongly identified as background	0	**13**

Analysis for Train-Val:

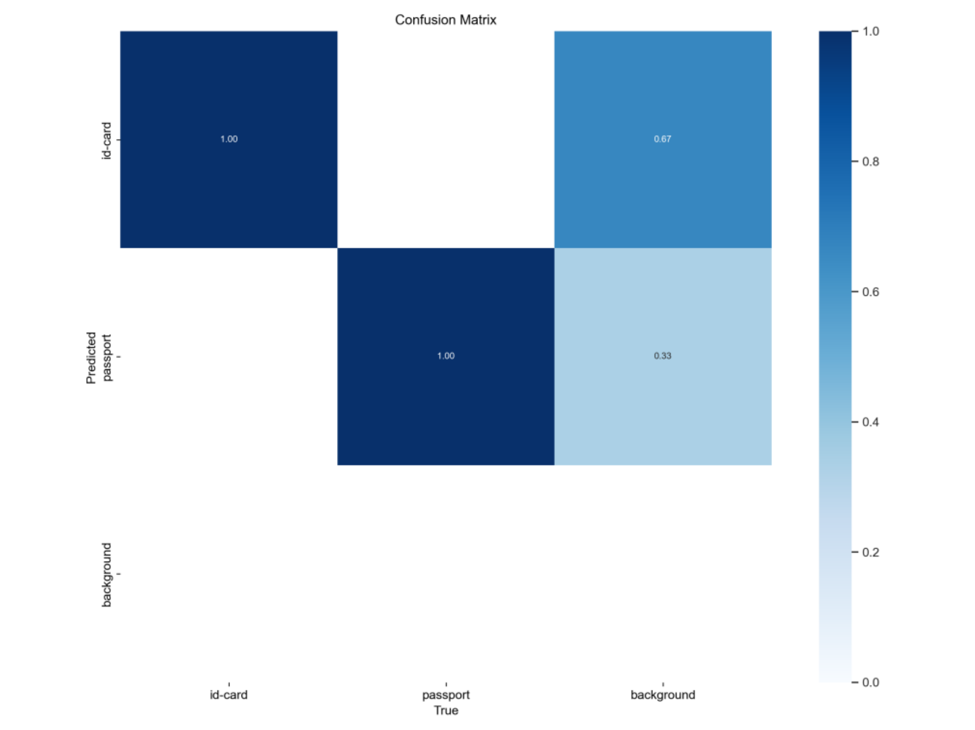
- **Did well for passport, did decently for ID card**
- Identified some background as ID card/Passport

Analysis for Test:

- **Performed poorly for passport prediction**

Improved Model 1 (+ Augmentation)

Validation Scores



Test Scores

	ID Cards	Passport
Correctly identified	**95**	**99**
Wrongly identified as the other class	**5**	**1**
Wrongly identified as background	0	0

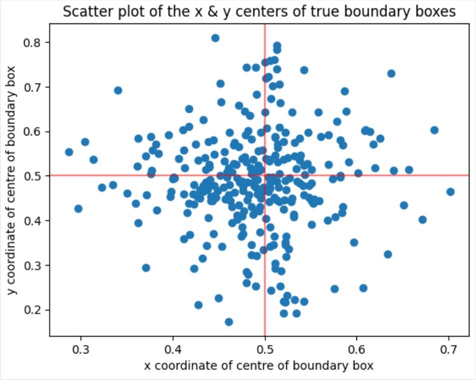
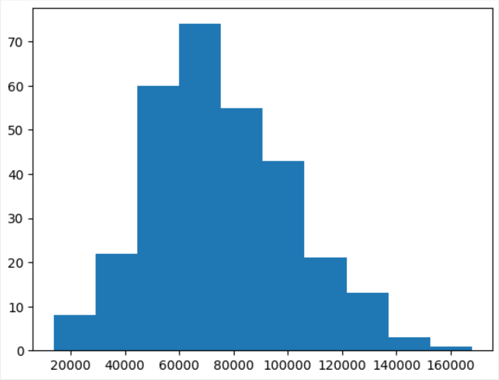
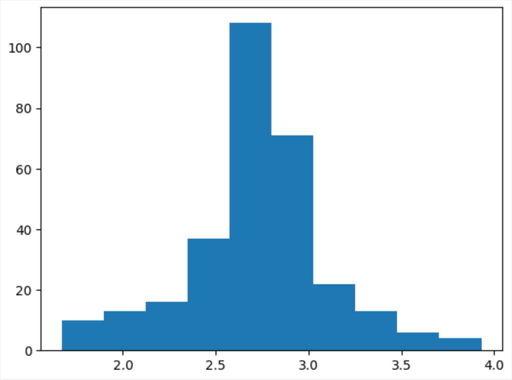
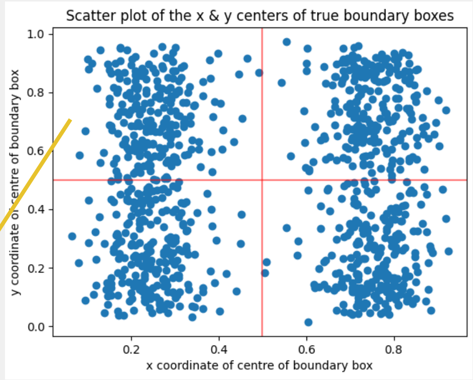
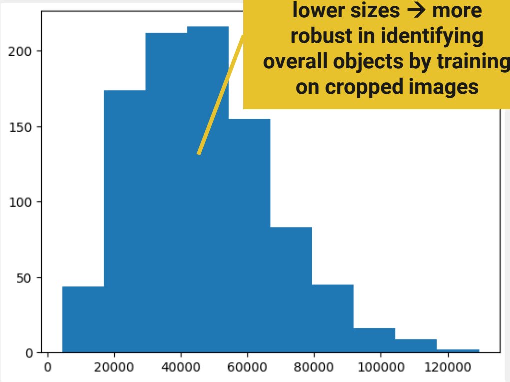
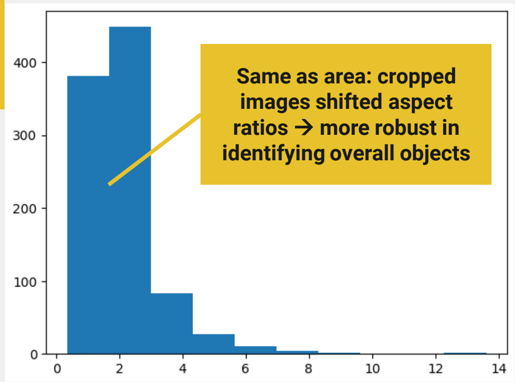
Analysis for Train-Val:

- **This model performed much better –** likely due to **training data augmentation** (see following EDA)**.**
- **Still predicting background as object**

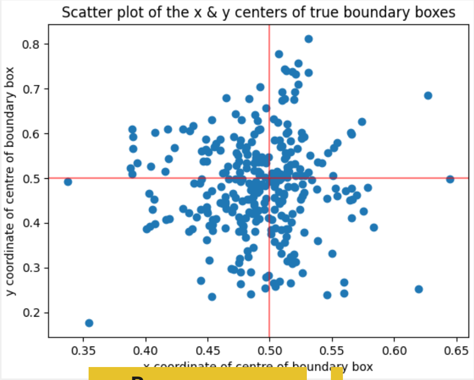
Analysis for Test:

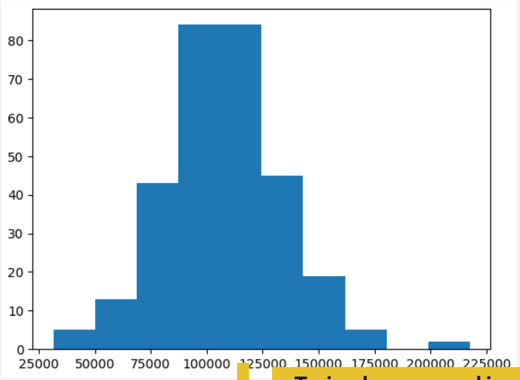
- **Better overall performance;** performed **better for passport prediction**

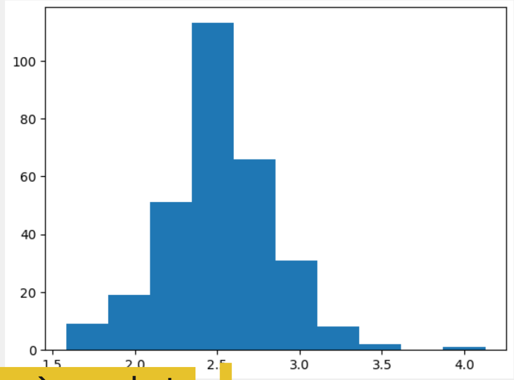
Augmented ID Cards has Better Distributed Samples

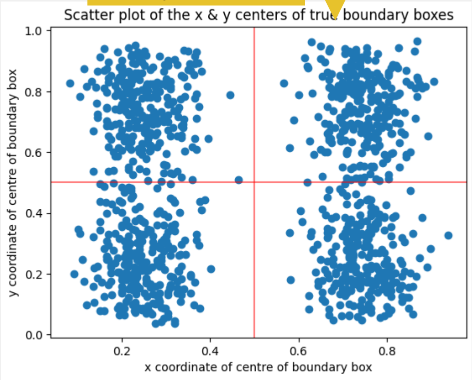
	Scatterplot of Boundary Box centres	Histogram of Boundary Box Area	Histogram of Boundary Box Aspect Ratio
ID Cards (**non-augmented**)			
ID Cards (**augmented**) **Became more evenly distributed due to translation, mosaic, scale**		**Shifted area towards lower sizes → more robust in identifying overall objects by training on cropped images** 	 **Same as area: cropped images shifted aspect ratios → more robust in identifying overall objects**

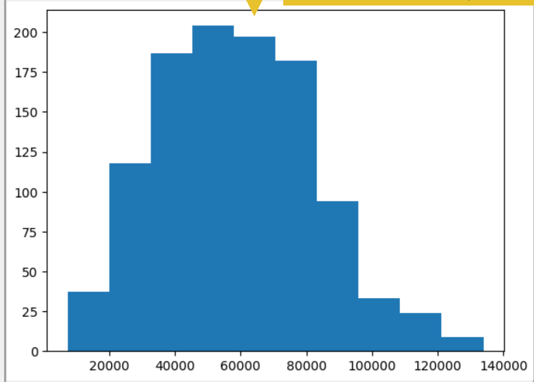
Same Improvements for Passports

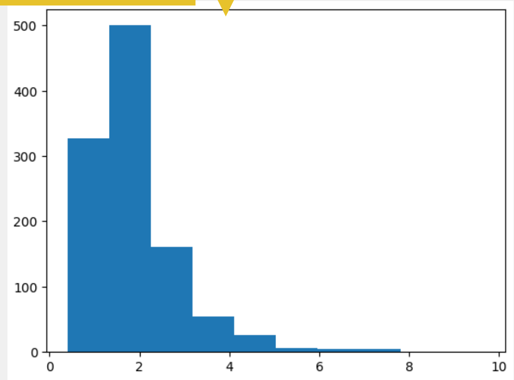
<table>
<tr><th></th><th>Scatterplot of Boundary Box centres</th><th>Histogram of Boundary Box Area</th><th>Histogram of Boundary Box Aspect Ratio</th></tr>
<tr><td>Passport (**non-augmented**)</td><td>


</td><td>


</td><td>


</td></tr>
<tr><td>Passport (**augmented**)</td><td>


</td><td>


</td><td>


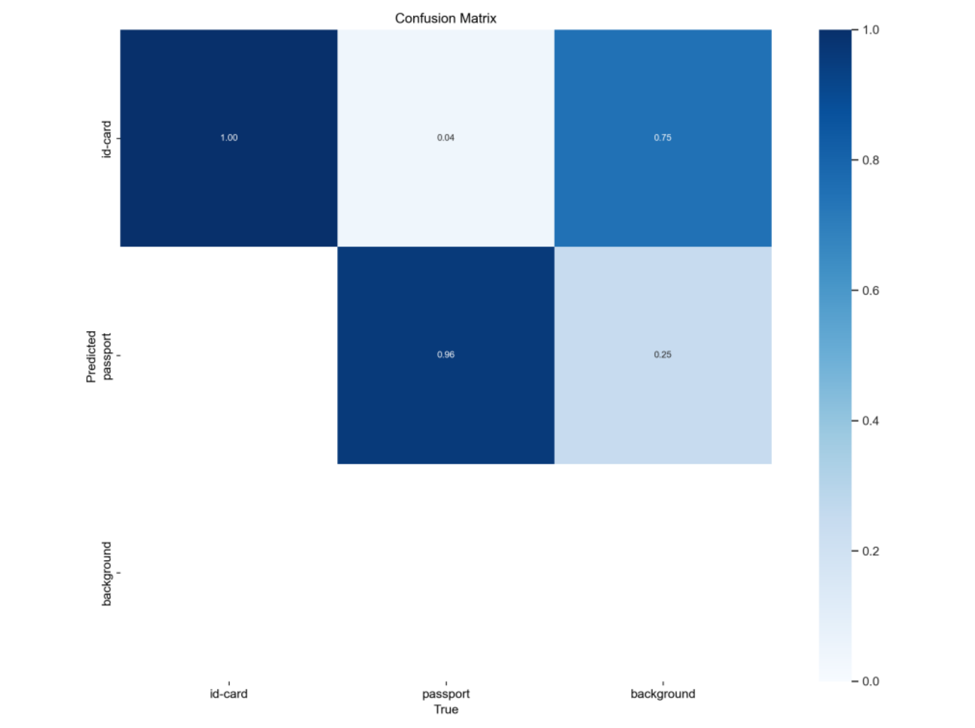
</td></tr>
</table>

Became more evenly distributed

Trained on cropped images → more robust in object detection

Improved Model 2 (+ Augmentation + Rotate + Shear)

Validation Scores



Test Scores

	ID Cards	Passport
Correctly identified	**95**	**98**
Wrongly identified as the other class	**3**	**2**
Wrongly identified as background	**2**	0

Analysis for Train-Val:

- This model performed similar to improved model 1
- Still predicting background as object

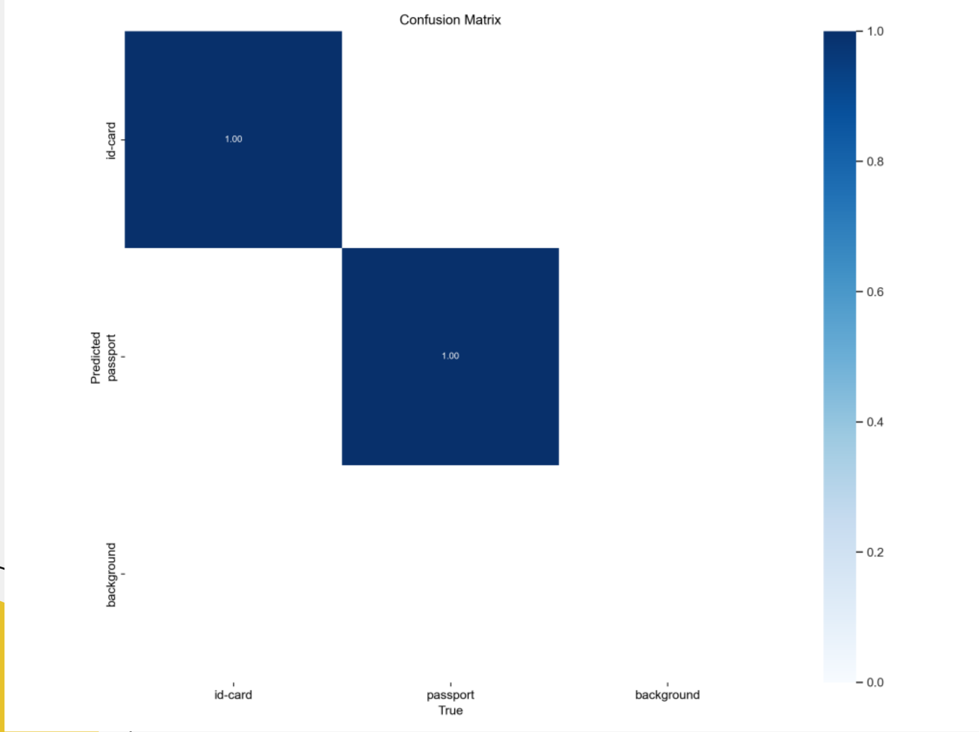
Analysis for Test:

- **Poorer performance than improved model 1** – likely due to excessive data augmentation on small dataset[11]

[11] Excessive data augmentation lead to poorer model performance, because it introduces data noise (Ultralytics, Tencent)

Improved Model 3 (+ Augmentation + 2x Training Data)

Validation Scores



Test Scores

	ID Cards	Passport
Correctly identified	**96**	**100**
Wrongly identified as the other class	**1**	0
Wrongly identified as background	**3**	0

Analysis for Train-Val:

- This model performed well

Analysis for Test:

- **Best performance thus far,** although only a small improvement from improved model 1
- **Increasing data size through augmentation can lead to some improvements**[12]

[12] Small datasets can benefit from data augmentation techniques (The Analytics Club, Datacamp)

Model Test Performance Comparison

	Baseline		Improved 1		Improved 2		Improved 3 *best performance*	
	ID Cards	Passport	ID Cards	Passport	ID Cards	Passport	ID Cards	Passport
Correctly identified	99	62	95	99	95	98	96	100
Wrongly identified as the other class	1	25	5	1	3	2	1	0
Wrongly identified as background	0	13	0	0	2	0	3	0
Total correct predictions	161		194		193		196	
Total wrong predictions	39		6		7		4	

Streamlit Demonstration

Please scan
the QR code to access
my Streamlit app!

04

Conclusion

Summary

Problem statement	A digital bank in Singapore wants to expand its banking business into Indonesia, but **faced the challenge of Indonesia not having digital national ID cards.** The bank wants to **avoid manual ID verification for account opening**, because it leads to a poor customer experience
Results	As tasked by the bank, an **object detection model was successfully developed** with a test **accuracy of 96% for ID cards and 100% for passports.** The streamlit deployment also **achieved a classification speed of ~3 seconds** for each photo submission[13].

[13] tested on both PC (using CPU) and mobile phone (Samsung Galaxy S22+)

Limitations & Areas for Improvement

Limitations

- **Dataset used is too small** even for transfer learning, likely leading to **poor generalisation**
- **Possibly limited performance on Southeast Asian ID documents** (due to nature of dataset)
- Current model is **unable to discern quality of submissions** (e.g. blur, too dark, too bright etc.)

Areas for improvement

- **Conduct proof-of-concept using real data** (>= 1,500 images per class)
- **Use transfer learning to train on local dataset** (i.e. start from developed weights)
- **Develop a multi-head**[14] **model to conduct quality classification** (each head to classify 1 type of quality)

[14] Cornell, Sep 2021, Marvik, Dec 2022

Thank You!

Do you have any questions?

Annex: Dataset

- The digitally generated photos were printed and made into physical documents
- These documents were then taken under these conditions:
    - Low lighting conditions (20 documents of each type)
    - Keyboard as a background (10 documents of each type)
    - Natural lighting, captured outdoors (10 documents of each type)
    - Table as a background (10 documents of each type)
    - Cloth with various textures as a background (10 documents of each type)
    - Text document as a background (10 documents of each type)
    - High projective distortions of the document (20 documents of each type)
    - Highlight from the sun or lamp hides a portion of the document (10 documents of each type)
- Each country abide to these pre-set parameters:
    - 80% are adults (18 – 60 years old)
    - 10% are seniors (>60)
    - 10% are children, adolescents (<= 17)
    - 50:50 male-female ratio
- Source: MIDV-2020: A Comprehensive Benchmark Dataset for Identity Document Analysis

Annex: Brief History of YOLO

- **YOLO (You Only Look Once) is a popular object detection and image segmentation model** developed by Joseph Redmon and Ali Farhadi at the University of Washington. The first version of YOLO was released in 2015 and **quickly gained popularity due to its high speed and accuracy**.

- YOLOv2 was released in 2016 and improved upon the original model by incorporating batch normalization, anchor boxes, and dimension clusters. YOLOv3 was released in 2018 and further improved the model's performance by using a more efficient backbone network, adding a feature pyramid, and making use of focal loss.

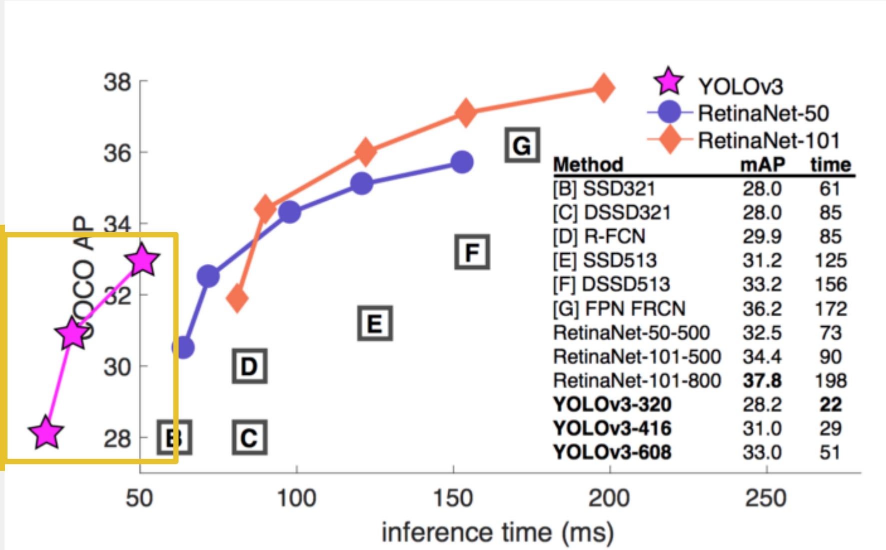
- In 2020, YOLOv4 was released which introduced a number of innovations such as the use of Mosaic data augmentation, a new anchor-free detection head, and a new loss function.

- In 2021, Ultralytics released YOLOv5, which further improved the model's performance and added new features such as support for panoptic segmentation and object tracking. In addition, Ultralytics maintains a public open-source repository to provide guides and tutorials on the installation and use of YOLOv5 – **making YOLOv5 easy to build and use**.

Source: Ultralytics

Annex: Why YOLO and Not Other Models?

YOLO excels in detection speed, enabling object detection in real-time videos



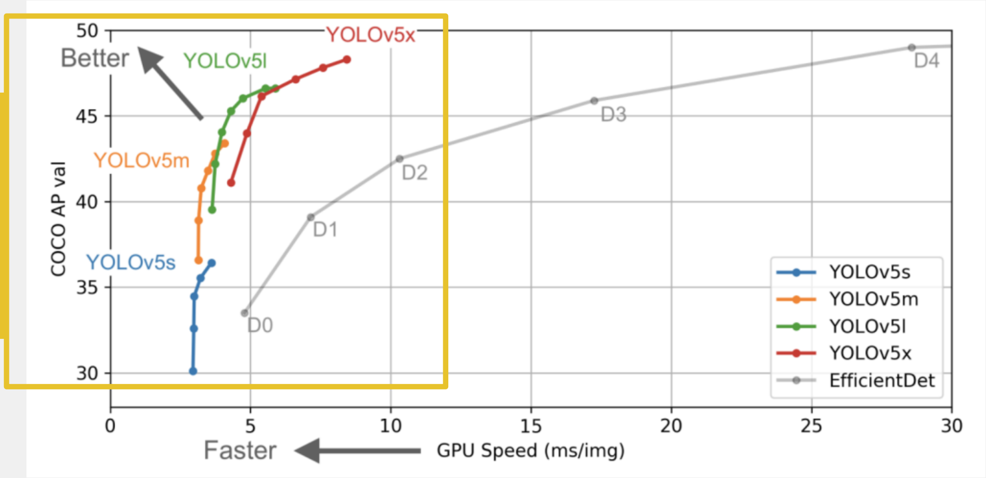
Source: YOLOv3 paper

YOLO is a **single-shot detector*** that uses a fully convolutional neural network (CNN) to process an image. It processes an entire image in a single pass, making them computationally efficient.

*Single-shot object detection uses a single pass of the input image to make predictions about the presence and location of objects in the image. Whereas a two-shot object detection uses two passes of the input image to make predictions about the presence and location of objects: the first pass is used to generate a set of proposals or potential object locations, and the second pass is used to refine these proposals and make final predictions

Annex: Why YOLOv5l (Large)?

YOLOv5l offers a good balance between performance and speed of inference



Source: Ultraytics

Annex: YOLOv5 Architecture

The YOLO network consists of three main pieces.

- **Backbone**: A convolutional neural network that aggregates and forms image features at different granularities.
- **Neck**: A series of layers to mix and combine image features to pass them forward to prediction.
- **Head**: Consumes features from the neck and takes box and class prediction steps.

https://towardsai.net/p/computer-vision/yolo-v5%E2%80%8A-%E2%80%8Aexplained-and-demystified

https://iq.opengenus.org/yolov5/

https://blog.roboflow.com/yolov5-improvements-and-evaluation/

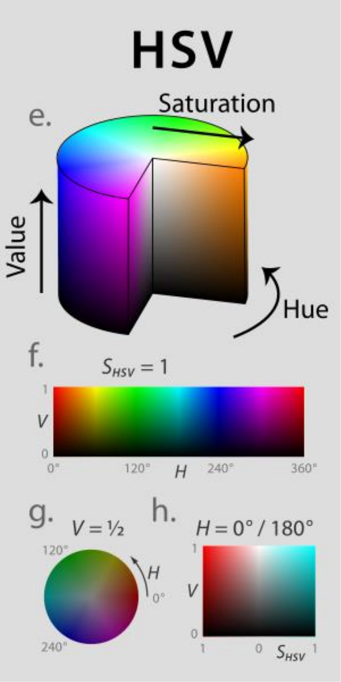
Annex: Single-Shot Object Detection

- asd

Annex: YOLOv5 Loss Functions

- Classes loss(BCE loss)
- Objectness loss(BCE loss)
- Location loss(CIoU loss)

Annex: Colour Augmentation



- **Saturation: intensity of colour in an image**. Highly saturated images have vivid, rich colours and lowly saturated images are pale and washed-out
- **Hue**: **attribute of a visible light** due to which it is differentiated from or similar to the primary colors: red, green and blue
- **Value** (a.k.a brightness): the **perception of how intense the light coming from a screen is**

Source: Techopedia for saturation, hue, value; Wikipedia

Annex: Dimension Augmentation (1)

- **Translation**: vertical / horizontal shift of object

From the left, we have the original image, the image translated to the right, and the image translated upwards.

- **Scale**: altering the image size to be larger (i.e. zoom in) or smaller (i.e. zoom out) than the original

From the left, we have the original image, the image scaled outward by 10%, and the image scaled outward by 20%

Source: Nanonets

Annex: Dimension Augmentation (2)

- **Flip left-right**: flipping the image horizontally

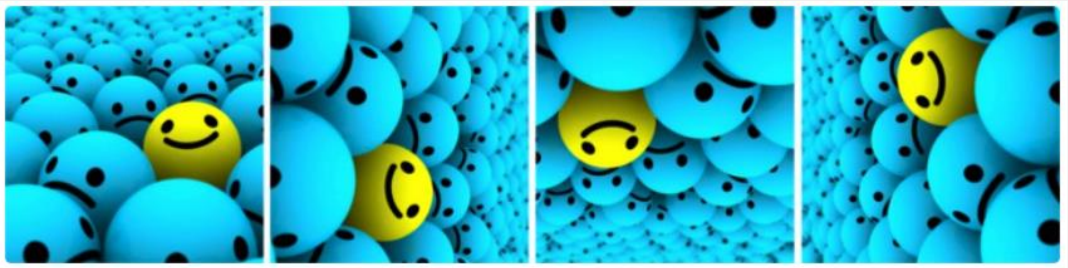
From the left, we have the original image, followed by the image flipped horizontally, and then the image flipped vertically.

- **Mosaic**: combining a corner of an image with 3 separate corners of 3 other images

Source: Nanonets, Roboflow

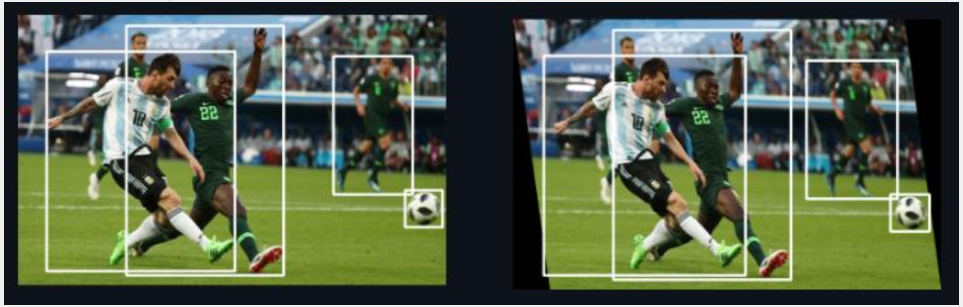
Annex: Dimension Augmentation (3)

- **Rotate**: turn the image around the centre of the image as its rotating axis



The images are rotated by 90 degrees clockwise with respect to the previous one, as we move from left to right.

- **Shear**: displace the image horizontally to different degree, transforming the image into a parallelogram



Source: Nanonets, Paperspaceblog

Annex: Auto Learning Bounding Box Anchors

- asd

Annex: Non-Maximum Suppression

- asd

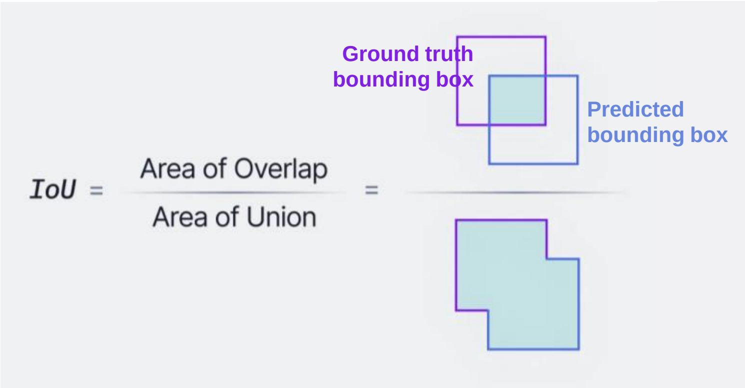
Annex: Intersection over Union (IoU)

Overview

IoU is a popular metric to **measure localization accuracy** and calculate localization errors in object detection models.

Detailed explanation

- The formula for IoU is show on the right
- E.g. IoU score of 0.54 = 54% overlap between the two boxes
- A threshold can be set for the IoU score, to decide how much overlap constitutes a ‘positive’ classification



Source: v7labs

Annex: Mean Average Precision (mAP)

Overview	mAP is a popular metric that measures the
Detailed explanation	• **Average Precision (AP) is calculated as the area under a precision vs. recall curve for a set of predictions**. • Recall and precision offer a trade-off that is graphically represented into a curve by varying the classification threshold. The area under this precision vs. recall curve gives us the Average Precision per class for the model. **The average of this value, taken over all classes, is called mean Average Precision (mAP)**.

Source: v7labs

- https://blog.roboflow.com/mean-average-precision/
- https://blog.paperspace.com/mean-average-precision/#:~:text=To%20evaluate%20object%20detection%20models,model%20is%20in%20its%20detections.

Annex: Modelling Train-Val Results

- Baseline Model:

Class	Images	Instances	P	R	mAP50	mAP50-95:
all	200	200	0.945	0.954	0.983	0.963
id-card	200	100	0.979	0.927	0.983	0.958
passport	200	100	0.91	0.98	0.983	0.968

- Improved Model 1:

Class	Images	Instances	P	R	mAP50	mAP50-95:
all	200	200	0.998	0.995	0.995	0.988
id-card	200	100	0.998	0.99	0.995	0.986
passport	200	100	0.999	1	0.995	0.99

- Improved Model 2:

Class	Images	Instances	P	R	mAP50	mAP50-95:
all	200	200	0.985	0.982	0.992	0.859
id-card	200	100	0.97	1	0.989	0.854
passport	200	100	1	0.964	0.995	0.865

- Improved Model 3:

Class	Images	Instances	P	R	mAP50	mAP50-95:
all	200	200	0.999	1	0.995	0.992
id-card	200	100	0.998	1	0.995	0.994
passport	200	100	1	0.999	0.995	0.991