**MINISTRY OF EDUCATION AND TRAINING**

**HCMC UNIVERSITY OF TECHNOLOGY AND EDUCATION**

**FACULTY FOR HIGH-QUALITY TRAINING**

*Logo

Description automatically generated*

**INTERNSHIP PROGRAM**

**AN OPTIMAL SOLUTION FOR EKYC SERVICE – VIETNAMESES ID CARD DETECTION IN UNICLOUD GROUP**

**Student:**

**PHAM MING LONG**

Student ID: 19119067

**MAJOR: COMPUTER ENGINEERING TECHNOLOGY**

**Advisor**: **Assoc. Prof. TRUONG NGOC SON**

Ho Chi Minh City, December 2022

**MINISTRY OF EDUCATION AND TRAINING**

**HCMC UNIVERSITY OF TECHNOLOGY AND EDUCATION**

**FACULTY FOR HIGH-QUALITY TRAINING**

*Logo

Description automatically generated*

**INTERNSHIP PROGRAM**

**AN OPTIMAL SOLUTION FOR EKYC SERVICE – VIETNAMESE ID CARD DETECTION IN UNICLOUD GROUP**

**Student:**

**PHAM MINH LONG**

Student ID: 19119067

**MAJOR: COMPUTER ENGINEERING TECHNOLOGY**

**Advisor**: **Assoc. Prof. TRUONG NGOC SON**

Ho Chi Minh City, December 2022

|  |  |
| --- | --- |
| Shape  Description automatically generated with medium confidence | THE SOCIALIST REPUBLIC OF VIETNAM  **Independence – Freedom– Happiness**  --------  *Ho Chi Minh City, December 30, 2022* |

**ADVISOR’S EVALUATION SHEET**

Student name: Pham Minh Long Student ID: 19119067

Major: Computer Engineering Technology

Project title: An optimal solution for eKYC service – Vietnamese ID Card Detection in Unicloud Group.

Advisor: Assoc. Prof. Truong Ngoc Son

**EVALUATION**

1. Content of the project:

1. Strengths:

1. Weaknesses:

1. Approval for oral defence? *(Approved or denied)*

1. Overall evaluation: (Excellent, Good, Fair, Poor)

1. Mark:……………….(*in words*: )

*Ho Chi Minh City, December 30, 2022*

**ADVISOR**

*(Sign with full name)*

**DISCLAIMER**

This is the final report for the Internship Program, **"An optimal solution for eKYC service – Vietnamese ID Card Detection in Unicloud Group"**. The report is accurate and is carried out entirely under the direction of the instructor, Assoc. Prof. TRUONG NGOC SON. The report does not duplicate any other sources either. Additionally, the paper includes a variety of cited and carefully labelled reference materials. Before the department, faculty, and school, we would like to fully accept responsibility for this promise.

**Student**

PHAM MINH LONG

**ACKNOWLEDGEMENT**

To complete this Internship Program, first of all, we would like to express our sincere thanks to all the Faculty for High-Quality training teachers, who have enthusiastically guided and equipped us with the necessary knowledge useful in the past semesters.

We would like to express our deep gratitude to Assoc. Prof. Truong Ngoc Son, directly guided and created all conditions to help us during the process of Internship Program.

Due to the short implementation time of the project, limited knowledge, limitations and errors in the topic have not been completely overcome. We look forward to receiving guidance and suggestions from you.

We sincerely thank you!

**Student**

PHAM MINH LONG

TABLE OF CONTENTS

[TABLE OF CONTENTS iii](#_Toc123256920)

[LIST OF FIGURES iv](#_Toc123256921)

[ABSTRACT v](#_Toc123256922)

[CHAPTER 1 general information of unicloud group 1](#_Toc123256923)

[1.1 HISTORY OF INFORMATION AND DEVELOPMENT 1](#_Toc123256924)

[1.2 VISIONS, MISSIONS, AND CORE VALUES 2](#_Toc123256925)

[1.3 MANUFACTURING PROCESS 2](#_Toc123256926)

[1.4 ORGANIZATIONAL STRUCTURE OF THE COMPANY 3](#_Toc123256927)

[1.5 ACHIEVEMENTS 4](#_Toc123256928)

[CHAPTER 2 GENERAL INFORMATION OF INTERNSHIP PROGRAM REPORT 6](#_Toc123256929)

[2.1 REASON FOR CHOOSING THE TOPIC 6](#_Toc123256930)

[2.2 RESEARCH OBJECTIVES 6](#_Toc123256931)

[2.3 RESEARCH METHODS 6](#_Toc123256932)

[2.4 THE SCOPE OF THIS RESEARCH 7](#_Toc123256933)

[CHAPTER 3 LITERATURE REVIEW 8](#_Toc123256934)

[3.1 IMAGE PROCESSING 8](#_Toc123256935)

[3.2 YOLOV7 9](#_Toc123256936)

[3.2.1 INTRODUCTION 9](#_Toc123256937)

[3.2.2 MODEL RE-PARAMETERIZATION 10](#_Toc123256938)

[3.2.3 MODEL SCALING 10](#_Toc123256939)

[3.2.4 EXTENDED EFFICIENT LAYER AGGREGATION NETWORKS 11](#_Toc123256940)

[3.2.5 MODEL SCALING FOR CONCATENATION-BASED MODELS 12](#_Toc123256941)

[3.2.6 COARSE FOR AUXILIARY AND FIN FOR LEAD LOSS 13](#_Toc123256942)

[3.3 PROGRAMMING LANGUAGE AND FRAMEWORKS 14](#_Toc123256943)

[CHAPTER 4 RESEARCH, DESIGN AND CONSTRUCT AN OPTIMAL SOLUTION FOR ID CARD DETECTION 18](#_Toc123256944)

[4.1 DATA ANALYSIS AND PREPROCESSING 18](#_Toc123256945)

[4.2 CORRECTING IMAGE ORIENTATION 20](#_Toc123256946)

[4.3 MODULES CONCATENATION 22](#_Toc123256947)

[CHAPTER 5 RESULTS AND EVALUATIONS 24](#_Toc123256948)

[5.1 YOLOV7 INFERENCE AND COMPARATION. 24](#_Toc123256949)

[5.2 RESPONSES FROM FAST API 25](#_Toc123256950)

[5.3 RESULTS IN CORRECTING IMAGE ORIENTATION 26](#_Toc123256951)

[5.4 RESULTS IN PERSPECTIVE TRANSFORM 27](#_Toc123256952)

[CHAPTER 6 CONCLUSION AND RECOMMENDATIONS 29](#_Toc123256953)

[6.1 CONCLUSION 29](#_Toc123256954)

[6.2 RECOMMENDATIONS 29](#_Toc123256955)

[REFERENCES 30](#_Toc123256956)

[APPENDICES 31](#_Toc123256957)

LIST OF FIGURES

[*Figure 1.1. Unicloud Group Logo [1]* 1](#_Toc123256877)

[*Figure 1.2. Unicloud's STM automated banking system. [2]* 3](#_Toc123256878)

[*Figure 1.3. Digital Banking Solution [3]* 4](#_Toc123256879)

[*Figure 2.1. Research progress* 7](#_Toc123256880)

[*Figure 3.1. Getting a "top-down" perspective of an image by using an OpenCV perspective transform.[3]* 9](#_Toc123256881)

[*Figure 3.2. The comparisons between YOLOv7 and other methods.[4]* 10](#_Toc123256882)

[*Figure 3.3. Extended efficient layer aggregation networks.[4]* 11](#_Toc123256883)

[*Figure 3.4. Model scaling for concatenation-based models.[4]* 12](#_Toc123256884)

[*Figure 3.5. Planned re-parameterized model.[4]* 13](#_Toc123256885)

[*Figure 3.6. Coarse for auxiliary and fine for lead head label assigner.[4]* 13](#_Toc123256886)

[*Figure 3.7. Python programming logo.[6]* 14](#_Toc123256887)

[*Figure 3.8. Pytorch framework logo.[5]* 15](#_Toc123256888)

[*Figure 3.9. Jira Software logo.[7]* 16](#_Toc123256889)

[*Figure 3.10. Label studio logo.[8]* 17](#_Toc123256890)

[*Figure 4.1. The label format is exported by Label Studio* 19](#_Toc123256891)

[*Figure 4.2. Original orienetation in ID Card.* 21](#_Toc123256892)

[*Figure 4.3. Correcting Image Orientation in ID Card.* 22](#_Toc123256893)

[*Figure 4.4. Module combination-based YOLOv7, Correcting Image Orientation, and Perspective Transform.* 23](#_Toc123256894)

[*Figure 5.1. Comparation in YOLO* 24](#_Toc123256895)

[*Figure 5.2. Polygon detection in YOLOv7x* 25](#_Toc123256896)

[*Figure 5.3. Response from FastAPI* 26](#_Toc123256897)

[*Figure 5.4. Result in correcting image orientation* 27](#_Toc123256898)

[*Figure 5.5. Image Alignment* 28](#_Toc123256899)

ABSTRACT

In this day and age, we have many model detectors such as Faster-RCNN, SDD, YOLO, and so on. More specifically, we will apply the latest version of YOLO, namely YOLOv7. In order to take ROI in ID Card, we additionally use Perspective Transform based on 4 orientations of the image, namely top-left, top-right, bottom-left, and bottom-right. However, when we cut the ROI in the image completely, the orientation of the image is not correct. Moreover, many applications have used classification models to categorise the corners such as CNN, ResNet50, AlexNet, and so on. But this method will be low inference. Therefore, we decide to apply mathematics so as to calculate the corner replies on the orientated vector of top-left and top-right that we will describe in this repository.

*Keywords: YOLOv7, Correcting Image Orientation, Perspective Transform, FastAPI*.

# GENERAL INFORMATION OF UNICLOUD GROUP

## HISTORY OF INFORMATION AND DEVELOPMENT

Unicloud Group was established in 2016, and currently has more than 300 members at the Hanoi headquarters and Ho Chi Minh City branch. UNICLOUD is a strategic partner of Sunshine - a multi-industry economic group; KSFinance (KSF); Smart Construction Group (SCG); ODE; Sunshine Homes (SSH).



*Figure 1.1. Unicloud Group Logo [1]*

Unicloud owns a comprehensive technology ecosystem that covers all key areas of the economy such as "Digital Banking Platform, Dgital Transformation, IoT Smart City, VR" Unicloud is waiting for potential employees to join hands to solve difficult problems, bringing real value to improve the quality of life not only of Vietnamese people but of the whole world. Along with the success of UNICLOUD are important strategic partners such as Sunshine Group - a multi-industry economic group; Real Estate and Fintech Corporation KSFinance (KSF); Smart Construction Group Joint Stock Company (SCG); ODE International Media and Entertainment Corporation; Branded real estate developer Sunshine Homes (SSH).

## VISIONS, MISSIONS, AND CORE VALUES

Unicloud Group researches develop and provide important technology products and solutions for most of the "backbone" sectors of the economy. Aim to strengthen capacity and promote the development of professions, improve the quality of life.

Constantly innovating, creating to solve the problems of users and the world's technology industry with 5,000 employees in at least 10 countries around the globe.

Unicloud has the mission of pioneering the digital transformation of Vietnam and the world, helping to raise the value of the Vietnamese people's brains to serve the civilized community to help improve the quality of life not only of Vietnamese people but all over the world. gender.

Owning core systems and core technologies to exploit Big Data through sharing users with other platforms. Focusing resources on ecosystems that create millions of users, and billions of transactions around Real Estate, Banking, Finance, Insurance and daily essential services.

## MANUFACTURING PROCESS

Using 100% self-built software, it took Unicloud's engineers more than 6 months to write and test the software for the machine's operation, which has been verified in practice for similarity compared to other models. most modern on the market. Even more, STM is superior when it can "seize, measure" flexibly depending on the needs of each user, thanks to its self-built software and customizable hardware.

Accordingly, by applying the eKYC electronic identification feature, customers can make transactions entirely by themselves using a chip-based CCCD, applying biometric identification, supporting video calls, opening cards with QRCode, issuing cards, and submitting payments. - Withdraw money on STM machines, with an extremely strict monitoring system for all operations, security and no fees. With the fresh signature feature, besides performing banking transactions, STM also easily solves problems related to public services.



*Figure 1.2. Unicloud's STM automated banking system. [2]*

A representative of Unicloud said that up to now, together with a system of strategic partner banks, the Group has completed the installation of about 300 machines, and continued to deploy and conquer the goal of installing 1,000 STMs for areas. in rural, remote and isolated areas, or localities that are lacking technological solutions in the field of Finance - Banking in order to accelerate the roadmap of a "cashless society".

## ORGANIZATIONAL STRUCTURE OF THE COMPANY

Unicloud has been applying the method of organizing human resources according to functional structure to ensure creativity, understanding and meeting customer needs better. At the same time can monitor and evaluate the operation in the most effective way. Currently, Unicloud has 4 main working groups in the company in Ho Chi Minh City, which are internal affairs, finance - accounting, technology, and business. The technology block includes teams such as AI, Full Stack Developer, DevOps, FrontEnd, BackEnd, and so on.

## ACHIEVEMENTS

Owning core systems and core technologies to exploit Big Data through sharing users with other platforms. Focusing resources on ecosystems that create millions of users, and billions of transactions around Real Estate, Banking, Finance, Insurance and daily essential services.

Applying digital technology in financial and banking activities, thereby providing a variety of financial products on many digital channels. Personalize customer experience, shorten the time and minimize procedures to increase customer satisfaction.



*Figure 1.3. Digital Banking Solution [3]*

Bank branch solution 4.0, serving daily transactions: account opening, money transfer, bill payment, public services, and financial services. Support a variety of authentication methods: EMV card, finger vein, face, QR Code, NFC, chip CCCD reader.

The all-inclusive eKYC server solution for banks, AI solutions & eKYC servers, allows financial institutions to authenticate customers in the most natural way on online transaction channels. Machine Learning self-improves continuously during operation because AI is constantly “in contact” with customers.

# GENERAL INFORMATION OF INTERNSHIP PROGRAM REPORT

## REASON FOR CHOOSING THE TOPIC

Currently, the trading software at Unicloud, specifically the eKYC service, is almost completed. However, the identification of identity cards is not optimized automatically. Because customers often tend to give the ID card the wrong angle or direction with the identity recognition model. Therefore, the model will predict and return the wrong results to the user. So, the leader of the AI team planned to improve the model by figuring out how to make the model close to the ID card and automatically rotate it back in the right direction.

## RESEARCH OBJECTIVES

As mentioned above, the eKYC model of the identity card identification software has not determined the wrong angle of the input image. Therefore, the model is improved to help confirm the identity conveniently without forcing the user to be in the right direction as the machine requires, and to increase the convergence speed to return the fastest result to the user by applying the algorithm and optimising performance.

## RESEARCH METHODS

To complete an optimal model, data sets need to be collected and labelled first. Then, we will explore the optimal deep learning models in terms of accuracy, recall, precision, and f1-score to select training data for machine learning. In order for the data to be accurate and consistent with the selected model, the data needs to be reformatted to the correct model of the training model and visualize the data for testing.

Diagram

Description automatically generated

*Figure 2.1. Research progress*

After successful preprocessing, the fine and filtered data will be trained in deep learning models to give good accuracy. After successful training, the model's results may not predict well, so we need to post-process the data. Then we use the weights to test with another dataset to evaluate the darkness. superior. Finally, the model will be fine-tuned or cleaned the data to optimize accuracy and speed inference.

## THE SCOPE OF THIS RESEARCH

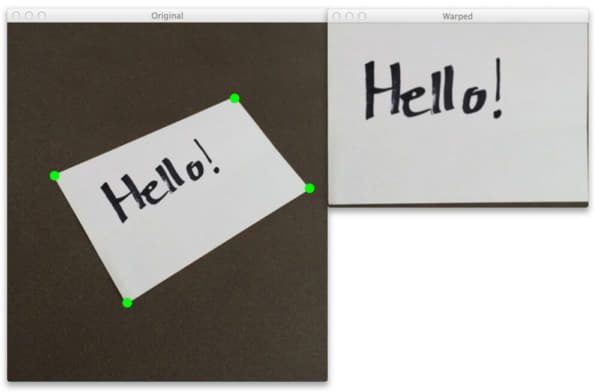
There are many scientific papers on algorithms that are very popular these days. This is a valuable resource for research. Besides, there is a lot of open-source code that is very convenient to edit and apply. However, the data in the company is not public and there is very little data to learn. In return, the AI community is very popular and wide, and easy to exchange and research on strong social sites.

# LITERATURE REVIEW

## IMAGE PROCESSING

Image processing is a technique for applying various operations to an image in order to improve it or to extract some relevant information from it. It is a form of signal processing where a picture serves as the input and the output could be another image or features related to the original image. Imaging technology is one of those that is expanding quickly today. The fields of engineering and computer science also include it as a core research subject.

With the help of Perspective Transformation, the perspective of a picture or video can be altered to better understand the necessary information. In order to change the viewpoint, we must specify the places on the image from which we wish to acquire information. The points inside which we wish to display our image must also be provided. The perspective transform is then extracted from the two sets of points provided and wrapped around the original picture.



*Figure 3.1. Getting a "top-down" perspective of an image by using an OpenCV perspective transform.[3]*

## YOLOV7

### INTRODUCTION

These days, real-time object detection plays an important role in computer vision such as multi-object tracking, autonomous driving, and so on. In this section, the proposed modules and optimization methods for trainable bag-of-freebies will be discussed. These methods can enhance detection accuracy without boosting the inference cost. Basically, model re-parameterization and dynamic label assignment will be applied in network training and objection detection to scale original parameters and assign dynamic targets for the outputs of different branches. To solve these state-of-the-art methods, extend and compound scaling methods will be proposed. The advantages of this method are to reduce about 40% of parameters and 50% computation of state-of-the-art real-time object detection and boost inference speed and detection accuracy.

Between 5 and 160 frames per second, YOLOv7 outperforms all other object detectors in terms of both speed and accuracy. and among all real-time object detectors with 30 FPS or higher on GPU V100, has the highest accuracy (56.8% AP). YOLOv7-E6 object detector (56 FPS V100, 55.9% AP) outperforms both transformer-based detector SWINL Cascade-Mask R-CNN (9.2 FPS A100, 53.9% AP) and convolutional-based detector ConvNeXt-XL Cascade-Mask R-CNN (8.6 FPS A100, 55.2% AP) by 501% in speed and 0.7%

Chart, line chart

Description automatically generated

*Figure 3.2. The comparisons between YOLOv7 and other methods.[4]*

### MODEL RE-PARAMETERIZATION

Techniques for model re-parametrization combine various computational modules into a single one at the inference step. Model re-parameterization can be categorized into two types: model-level ensemble and module-level ensemble. Both of these types of ensembles can be thought of as ensemble techniques. For model-level reparameterization in order to produce the final inference model, there are two typical methods. The first is to train numerous, identical models using various training datasets, and then average the weights of the multiple trained models. The second method involves calculating a weighted average of model weights across various iteration counts.

### MODEL SCALING

A model can be scaled up or down to fit a variety of computing devices by using this technique. The model scaling method typically makes use of a variety of scaling factors, including resolution (size of the input image), depth (number of layers), width (number of channels), and stage (number of feature pyramids), in order to achieve a good trade-off for the number of network parameters, computation, inference speed, and accuracy. One of the often employed model scaling techniques is network architecture search (NAS). Without establishing overly complex criteria, NAS may automatically look upscaling factors from the search space.

### EXTENDED EFFICIENT LAYER AGGREGATION NETWORKS

A variant of VoVNet is the CSPVoVNet concept shown in Figure 2. The architecture of CSPVoVNet takes into account the aforementioned fundamental designing considerations in addition to analyzing the gradient route to allow the weights of various layers to learn more varied features. The above-mentioned gradient analysis method produces inferences more quickly and accurately. How to create an efficient network? is the design technique taken into account by ELAN. They got to the idea that a deeper network can effectively train and converge if the shortest and longest gradient path is controlled. The main architecture of the Extended-ELAN (E-ELAN).

Timeline

Description automatically generated

*Figure 3.3. Extended efficient layer aggregation networks.[4]*

The suggested E-ELAN employs expand, shuffle, and merge cardinality to continuously improve the network's capacity for learning while preserving the original gradient path. E-ELAN solely modifies the architecture of the computational block, leaving the architecture of the transition layer untouched. In order to increase the channel and cardinality of computing blocks, we plan to employ group convolution. To all the computational blocks in a computational layer, the same group parameter and channel multiplier will be applied. In accordance with the specified group parameter g, the feature map produced by each computing block is subsequently divided into g groups and concatenated.

The aforementioned techniques are primarily employed in PlainNet and ResNet topologies. The in-degree and out-degree of each layer will not change while scaling up or down in these topologies, allowing us to independently examine the effects of each scaling factor on the number of parameters and computation. However, when these techniques are used with a concatenation-based architecture, we discover that whether scaling up or scaling down is done on depth, the in-degree of a translation layer that follows a concatenation-based computational block will change.

Chart, box and whisker chart

Description automatically generated with medium confidence

*Figure 3.4. Model scaling for concatenation-based models.[4]*

Therefore, for a concatenation-based model, we must suggest the matching compound model scaling approach. Calculating the change in the output channel of a computational block is necessary for scaling the depth factor of that block. The outcome is depicted after performing width factor scaling on the transition layers with the same degree of change.

### MODEL SCALING FOR CONCATENATION-BASED MODELS

RepConv really integrates identity connection, 1x1, and 3x3 convolutions in a single convolutional layer. Following an examination of the interaction and corresponding performance of RepConv and various architectures, we discover that the identity connection in RepConv eliminates the residual in ResNet and the concatenation in DenseNet, resulting in a greater variety of gradients for various feature maps. The architecture of the anticipated re-parameterized convolution is designed using RepConv without identity connection (RepConvN) for the aforementioned reasons.

Diagram

Description automatically generated

*Figure 3.5. Planned re-parameterized model.[4]*

### COARSE FOR AUXILIARY AND FIN FOR LEAD LOSS

A method that is frequently employed while training deep networks is deep supervision . The shallow network weights using assistant loss as the guidance, and the basic idea is to add extra auxiliary head in the middle levels of the network. Deep supervision can nevertheless greatly enhance the performance of the model on a variety of tasks, even for architectures like ResNet and DenseNet that often converge well. The object detector architecture is depicted in Figure 5(a) and (b), respectively, with and without deep supervision.

Diagram

Description automatically generated

*Figure 3.6. Coarse for auxiliary and fine for lead head label assigner.[4]*

The ground truth and the lead head's prediction results serve as the major inputs to the lead head guided label assigner, which then uses optimization to produce soft labels. This collection of soft labels will serve as the lead head and auxiliary head's target training model.

In order to generate soft labels, the coarse-to-fine lead head guided label assigner additionally employed the ground truth and the lead head's expected outcome. However, as part of the process, we produce two distinct sets of soft labels, referred to as coarse labels and fine labels. Whereas the fine label is the same as the soft label produced by the lead head guided label assigner, the coarse label is produced by loosening the restrictions placed on the positive sample assignment process, which allows for the treatment of more grids as positive targets.

## PROGRAMMING LANGUAGE AND FRAMEWORKS

High-level, all-purpose programming languages like Python are very popular. The Python Software Foundation continued to develop it after it was first created by Guido van Rossum in 1991. Its syntax lets programmers to convey their notions in less lines of code because code readability was a key consideration in its creation. Working swiftly and integrating systems more effectively are made possible by the programming language Python.



*Figure 3.7. Python programming logo.[6]*

Facebook created the Deep Learning framework Pytorch. Facebook is recognized as a technology company that devotes significant resources to the development of artificial intelligence (in addition to Amazon, Google, or Apple). Using free sources to create Pytorch A sizable sharing community has been developed by Facebook. Take a moment to search since the problems you're experiencing might already have been solved and shared with the community given the extensive resources on our site.

Logo, company name

Description automatically generated

*Figure 3.8. Pytorch framework logo.[5]*

* 1. PROJECT MANAGEMENT TOOLS

Jira is an application used to manage errors, and track problems that occur in the company's project activities. This software is developed by Atlassian company, Australia.

This tool can run on any database, hardware platform as well as the operating system. Integrate multiple systems easily, including Excel, RSS and Email… A quick search through Jira Query Language. Support to build a suitable and fast workflow with the requirements of the project. Besides, with this software, you can keep a close eye on the progress of each project.



*Figure 3.9. Jira Software logo.[7]*

Support to manage Bugs, Tasks, work improvement or project errors. Create and save highly configurable filters, and share them with other users via periodic email. In addition, Jira also provides a variety of statistics, and summary reports with many charts. These charts are suitable for a wide range of users and project types, making it easier to view information.

An open-source data labelling tool for investigating and classifying various sorts of data is called Label Studio. With several different data formats, you can execute various labelling operations. Additionally, you can use continuous active learning with Label Studio to connect it with machine learning models and provide predictions for labels (pre-labels).

A picture containing diagram

Description automatically generated

*Figure 3.10. Label studio logo.[8]*

# RESEARCH, DESIGN AND CONSTRUCT AN OPTIMAL SOLUTION FOR ID CARD DETECTION

## DATA ANALYSIS AND PREPROCESSING

The data will be stripped of the IDs/national identity cards labelled 'rotate' and take the remaining 7 labelled IDs/ID cards to train the model to learn the correct oriented shapes to proceed with the method. the flipping method can be easier with fixed angles. The data is divided at the rate of 50 20 30 corresponding to the set of train, Val, and test. After dividing the number of train, Val, and test sets, they are 1,502, 603, and 898, respectively. The data sets are dropped because there are no labelled coordinates.

After we label all the data, we will export the JSON file as follows:

Graphical user interface, text, application

Description automatically generated

*Figure 4.1. The label format is exported by Label Studio*

In order to get the ID/ID model on the right track, we need to filter out the images labeled “rotate”, keeping the images with the remaining labels. The goal is to train using the YOLO model, so it is necessary to convert to a custom label YOLO as follows: [class, , , width, height]. In addition, to get the polygon for identity / identity card, we need to convert the "points" coordinates in Figure 6, from 1 point to 4 points to get 4 bounding boxes at the corners of the image. The goal is that after training, YOLO returns the coordinates of , , , , based on these coordinates we will find the center point in each bounding, then connect these center points we will have get 1 polygon. The corresponding points in Figure 6 will be {(, ), (, ), , ), (, )}. The formula to convert 4 bounding at 4 corners would be:

(4.1)

(4.2)

(4.3)

(4.4)

Since YOLO coordinates require normalization to the interval (0,1), the formula to convert to bounding box YOLO is:

(4.5)

(4.6)

(4.7)

(4.8)

## CORRECTING IMAGE ORIENTATION

Based on the predicted bounding box, we will flip the image with 3 cases 90, 180, 270 degrees by calculating the angle between vector Ox and vector containing coordinates top left and top right for vector AB with A as top left, B is the top right of the image as shown below.

Text, whiteboard

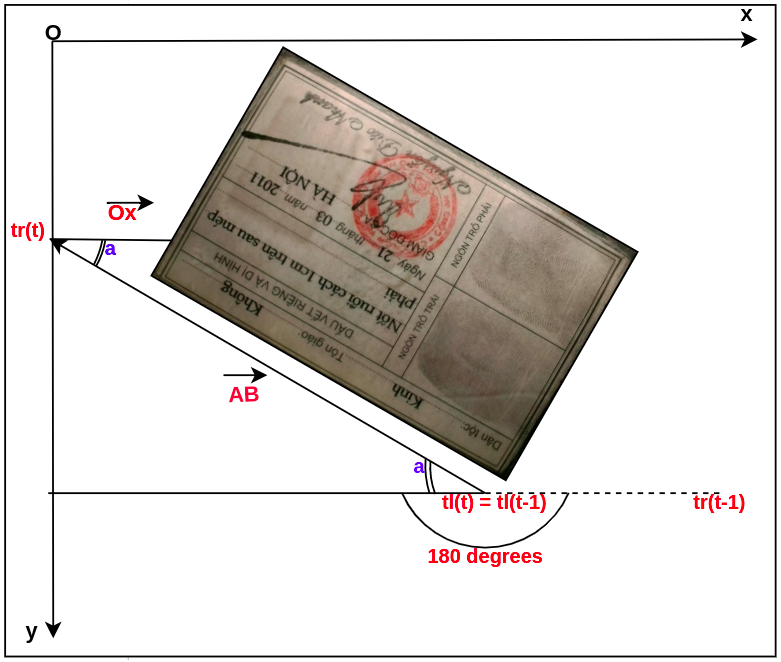
Description automatically generated

*Figure 4.2. Original orienetation in ID Card.*

Let's assume that vector AB( - , - ) is the combination between top\_left(tl) and top\_right(tr) coordination. Therefore, we will have the equation to rotate.

(4.9)

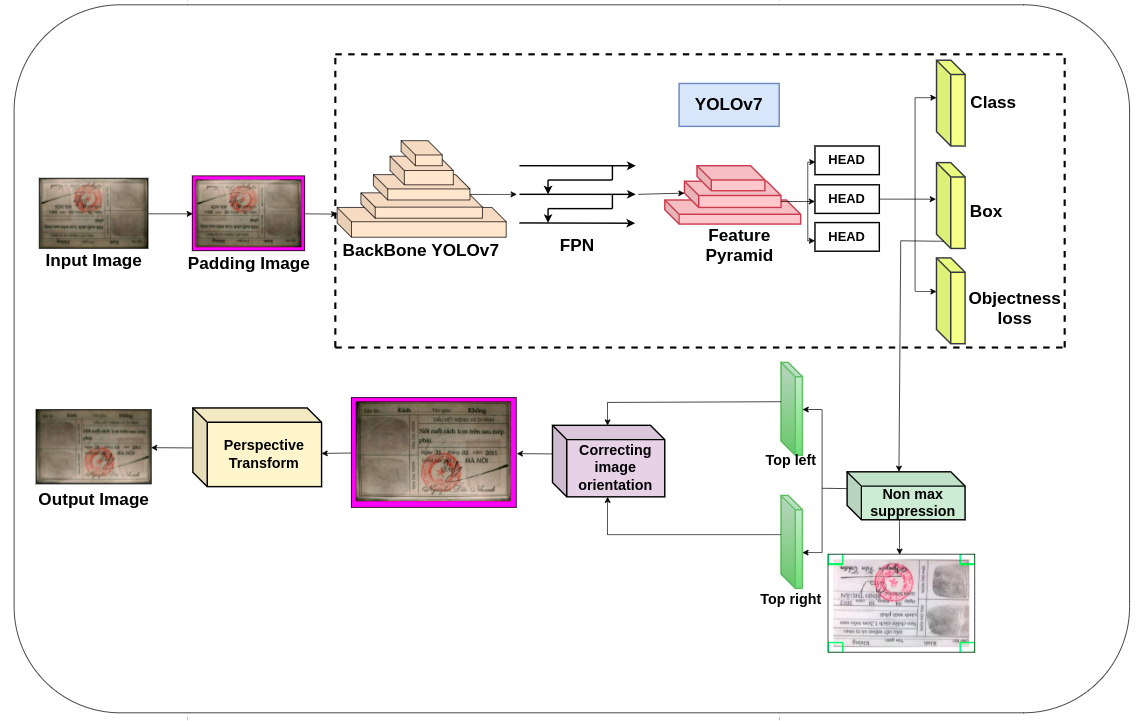
On the other hand, if the image has the angle which is different with zero and greater than 180 degrees, the image will be considered with the condition below to rotate suitably. Otherwise, the angle will be rotated. Finally we will flip in an anti-clockwise angle.



*Figure 4.3. Correcting Image Orientation in ID Card.*

## MODULES CONCATENATION

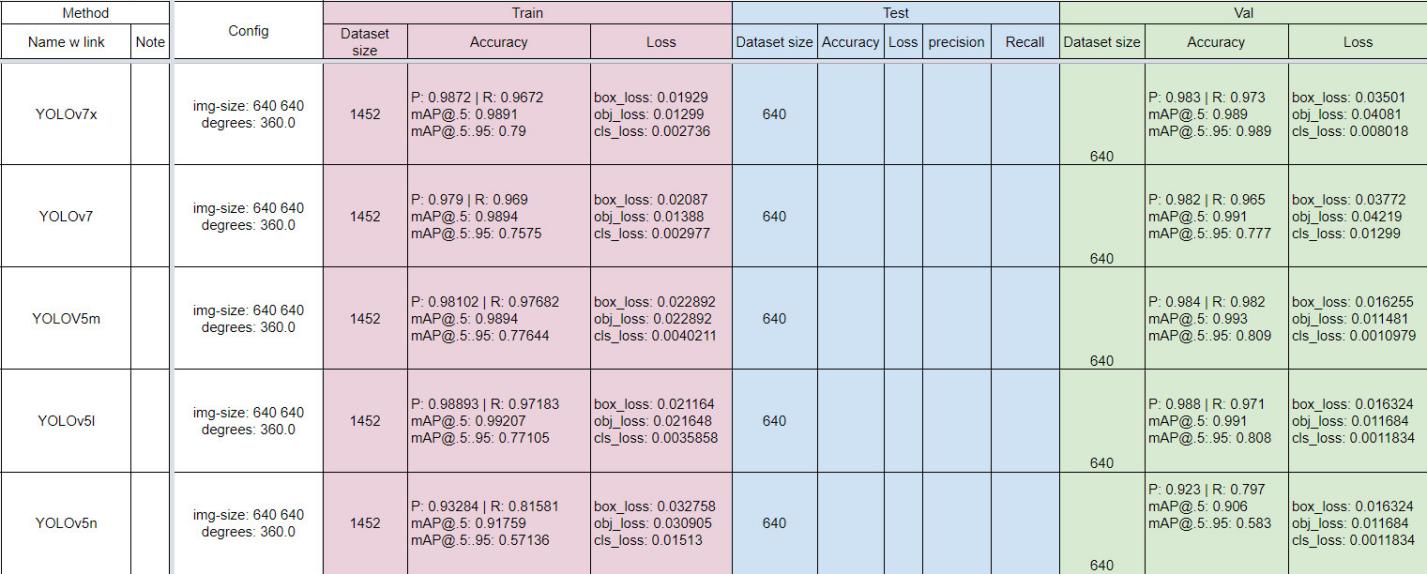
After completing the submodules, these models will be combined. Specifically, the input image will be padded with a purple area so that when using perspective transform, cutting information will not be lost. Then, this image will be fed into the YOLOv7 model for training. Since the bounding boxes of YOLOv7 are expected to overlap, they will be rectified by using non max suppression. The output bounding box of YOLOv7 will use the top left and top right points to calculate the angle to rotate in the right direction. Next, the flipped image in the right direction will be cropped based on the 4 center points of the bounding box to get the ideal area and also the final output.



*Figure 4.4. Module combination-based YOLOv7, Correcting Image Orientation, and Perspective Transform.*

# RESULTS AND EVALUATIONS

## YOLOV7 INFERENCE AND COMPARATION.

****

*Figure 5.1. Comparation in YOLO*

It can be seen that the versions of YOLOv7 have high inference speed and high accuracy because of the re-parameterization technique. However, YOLOv7 has a higher speed than YOLOv7x by more than 2ms, but in terms of accuracy YOLOv7x has higher accuracy. The speed difference is negligible, so YOLOv7x is the ideal choice.



*Figure 5.2. Polygon detection in YOLOv7x*

The above ID Card detection result is drawn from 4 bounding boxes with 4 corners and finding the center of each bounding box we will get the same result. It can be seen that YOLOv7x detects very well.

## RESPONSES FROM FAST API

The results of the Fast API return are the image name, the coordinates of the 4-corner bounding box in clockwise order, confidence score, class name, object name. The API is used to communicate with other teams' software, namely full stack and frontend teams.

**A picture containing table

Description automatically generated**

*Figure 5.3. Response from FastAPI*

## RESULTS IN CORRECTING IMAGE ORIENTATION

The results below show that the mathematical model, instead of using the deep learning model, gives high accuracy results and high inference speed, accelerating FPS to avoid software lag.



*Figure 5.4. Result in correcting image orientation*

## RESULTS IN PERSPECTIVE TRANSFORM

After that, we will get the bounding box coordinate detection result to get the ROI area applying the perpective transform image processing algorithm.



*Figure 5.5. Image Alignment*

# CONCLUSION AND RECOMMENDATIONS

## CONCLUSION

This is one of the cool methods to get the polygon of cccd/cmnd instead of taking a bounding box of the shape. Modeling for high accuracy and fast inference speed thanks to model reparameterization and model scaling techniques. However, in the process of data collection and filtering, there are still many difficulties and inaccuracies, so some panels will be flipped and rotated, leading to inaccurate learning models and detecting wrong angles. We will filter the data carefully and will try other versions of YOLOv7 such as ELAN, E-LAN, etc.

## RECOMMENDATIONS

The model has overcome the phenomenon of image being flipped due to the wrong shooting angle and cropping the ideal area for standardization. next device for character recognition model in cccd/cmnd. In addition, also know how to use Git: Gitlab, Github, Jira: project manager, proficient in pytorch, tensorflow, numpy, pandas, seaborn, opencv, matplotlib, learn more from colleagues, listen, give gentle suggestions.

REFERENCES

1. Công ty cổ phần tập đoàn công nghệ Unicloud, Dẫn đầu kỉ nguyên của lối sống thông minh, <http://career.unicloud.com.vn/>, 2022.
2. Báo điện tử dân Việt, Máy giao máy giao dịch ngân hàng tự động STM - sản phẩm chủ lực của Tập đoàn công nghệ Unicloud, <https://danviet.vn/may-giao-may-giao-dich-ngan-hang-tu-dong-stm-san-pham-chu-luc-cua-tap-doan-cong-nghe-unicloud-20220406163021043.htm>, 2020.
3. Pyimagesearch, 4 Point OpenCV getPerspective Transform Example, <https://pyimagesearch.com/2014/08/25/4-point-opencv-getperspective-transform-example/>, 2022.
4. Chien-Yao Wang, YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors, <https://arxiv.org/abs/2207.02696>, 2022.
5. The Linux Foundation, From Research to Production, <https://pytorch.org/>, 2022.
6. Vietnix JSC, Python là gì? Ứng dụng của ngôn ngữ lập trình Python, <https://vietnix.vn/python-la-gi/>, 2022.
7. Besthard, JIRA Software Commercial 10 Users [JSCP-ATL-10], <https://besthard.ru/jira-software-commercial-10-users/>, 2012 – 2022.
8. Heartex, Get started with Label Studio, <https://labelstud.io/guide/>, 2022.

APPENDICES

**Code github**: https://github.com/Syun1208/eKYC-ID-Card-Detection.git