“Optimization of annotation process for segmentation and detection model training”

Major Project Report

Submitted in Partial Fulfillment of the Requirements for the Degree of

BACHELOR OF TECHNOLOGY

IN

Instrumentation and Control Engineering

By
Deep Khut
(19BIC008)

Department of Electronics and Instrumentation Engineering
School of Technology
NIRMA UNIVERSITY
Ahmedabad 382 481

May 2023

CERTIFICATE

This is to certify that the Major Project Report entitled "**Optimization of annotation process for segmentation and detection model training** " submitted by Deep Khut (19BIC008) towards the partial fulfillment of the requirements for the award of degree in Bachelor of Technology in the field of IC Engineering of Nirma University is the record of work carried out by under our supervision and guidance. The work submitted has in our opinion reached a level required for being accepted for examination. The results embodied in this major project work to the best of our knowledge have not been submitted to any other University or Institution for award of any degree or diploma.

Date:

m pandya

Industry - Guide

Mr Mrudang Pandya
Research Lead
Abhishree Corporate Park,
Block no. 5-6, Abhishree,
Ambli – Bopal Rd.
Ahmedabad,380058

Institute - Guide

Prof. Harsh Kapadia
Department of Electronics & Instrumentation
Institute of Technology,
Nirma University,
Ahmedabad, 382 481

Head of Department

Department of EI Engineering
Institute of Technology
Nirma University
Ahmedabad

Director

Institute of Technology
Nirma University
Ahmedabad

Undertaking for Originality of the Work

I, Deep Khut , Roll No. 19BIC008 , give undertaking that the Major Project entitled "Optimization of annotation process for segmentation and detection model training " submitted by me, towards the partial fulfillment of the requirements for the degree of Bachelor of Technology in Instrumentation and Control of Nirma University, Ahmedabad, is the original work carried out by me and I give assurance that no attempt of plagiarism has been made. I understand that in the event of any similarity found subsequently with any other published work or any project report elsewhere; it will result in severe disciplinary action.

____

Signature of Student

Date: _______________

Place: _______________

Endorsed by:

m pandya

Mrudang Pandya

Harsh Kapadia

Certificate of Internship

This document is intended to certify that **Mr. Deep Khut** (full-time Bachelor's student of Instrumentation and Control Engineering at Nirma University, Ahmedabad, Gujarat, India) has completed internship for 16 weeks under my supervision and guidance at **Jekson Vision Pvt. Ltd.**

During this period, **Mr. Deep Khut** has successfully completed an Optimization of annotation process for segmentation and detection model training as per the requirement of the company.

In the work that was assigned to his, he was able to get over the hurdles in the coding which showcased his problem-solving abilities. He also thought of alternative and optimum solutions to handle the same, demonstrating good analytical abilities and logical bent of mind. He was sincere in his work and exhibited good team spirit.

Regards,

For Jekson Vision Private Limited,

Prakash

Prakash Sharma

Senior Manager -HR

Acknowledgement

I want to sincerely thank Mr. Mrudang Pandya and Jekson vision team for their contribution to discussion of “Optimization of annotation process for segmentation and detection model training”. The knowledge sharing has been considerably improved and made more thorough understanding of information and insight extraction.

I want to express my gratitude to Mr Pandya for kindly giving me the benefit of his time and expertise, as well as for giving me access to the materials and equipment I needed to perform this project. I am very Appreciative of his aid in helping me comprehend the complexities of Data Augmentation Techniques in an Computer Vision application because of his direction and support.

Additionally, I would like to thank Jekson Vision for their assistance and cooperation throughout this project. I feel privileged to have had the chance to work with them because of their commitment to excellence and dedication towards providing high-quality services.

In conclusion, I would like to express my sincere gratitude to Mr. Mrudang Pandya and Jekson Vision for their essential contributions to this subject. I also look forward to maintaining our future working relationship.

Deep Khut

Abstract

The annotation process is a critical step in training segmentation and detection models, but it can be time-consuming and tedious. To optimize this process, several strategies can be employed, including selecting the right annotation tool, using pre-annotation or semi-automatic annotation, developing a well-defined annotation protocol, providing training and feedback to annotators, using crowdsourcing, employing active learning, and monitoring the quality of the annotations. By implementing these strategies, the annotation process can be made more efficient and accurate, reducing the time and effort required while still achieving high accuracy.

The project starts off with an introduction to annotation process with different methodologies to perform it and having its specific advantages. By providing a labelled dataset for the model to learn and recognize patterns, annotations serve to increase the accuracy and efficacy of machine learning models. Machine learning models would be unable to discriminate between distinct items or regions of interest within an image or video without annotations, rendering them worthless for many real-world applications. The another phase of this project starts with data augmentation. Data augmentation is the process of creating additional training data from existing data using transformations such as rotation, scaling, and flipping. The purpose of data augmentation is to diversify the training data, making the model more resistant to fluctuations in real-world data.

Data augmentation and annotation can assist reduce the amount of manual labour necessary for training in addition to boosting the accuracy and effectiveness of machine learning models. The requirement for human annotation can be eliminated by creating more training data through data augmentation, making the training process more efficient. On the final stage, integrating these methodologies to create an final flow of the process for the data before going for training and testing purposes. It is feasible to improve the accuracy and efficacy of models while minimizing the amount of manual labour necessary for training by combining these strategies.

INDEX

List of Figures

Chapter 1: Introduction

1.1 Introduction to Company

A dynamic and forward-thinking business, Jekson Vision specialises in offering state-of-the-art vision inspection equipment and solutions. Jekson Vision has become recognised as a trusted authority in the area of machine vision technology thanks to a strong commitment to excellence and a customer-centric strategy.

Jekson Vision, a company founded with the goal of revolutionising quality control procedures, uses cutting-edge image technology, artificial intelligence, and machine learning algorithms to create cutting-edge inspection solutions. Pharmaceutical, food and beverage, automotive, and packaging sectors, among others, can use these systems to improve the quality of their products, boost productivity, and guarantee regulatory compliance.

In order to be on the cutting edge of technical breakthroughs, Jekson Vision consistently invests in research and development. We work hard to foresee market trends and a new problem, which enables us to provide innovative solutions that satisfy the changing needs of the market.

Jekson Vision is committed to providing exceptional customer care and support globally and has a strong network of partners and distributors. In addition to developing excellent products, we also offer thorough training, quick technical support, and pro-active maintenance services to guarantee top performance and client pleasure.

1.2 Overview

This project includes development and optimization of various methods for annotation and augmentation processes. Using image processing techniques we have developed semi-automatic annotation tool which can automatically detect the our region of interest and annotate in Json format and additionally improvised the augmentation method using Affine transformations. Also developed a method to convert this annotation in one format to another.

Overall, optimizing the annotation process for segmentation and detection model training include selecting the relevant tools and methodologies, developing a well-defined annotation protocol, providing annotators with training and feedback, and regularly checking the annotation quality. It is possible to improve the efficiency and accuracy of the annotation process by following these principles, resulting in higher performing models in real-world applications.

1.3 Motivation

Using Image processing technique: doesn't produce better output i.e less accurate for detection of different types of tablets in different filling trays.

So, by using Deep learning techniques to develop a full proof computer vision algorithm for tablet inspection which will not only increase the accuracy but also reduce the time.

To increase the accuracy of deep learning models we need to process the pre process data before it goes for training and testing purposes. So, the need of annotations and various tools for data augmentation is needed.

Also, to improve the models effectiveness, it should be able to recognize objects in various situations and environment such as low lighting. Through Data augmentation we can train our models through different visual variations so that they become more robust to these changes.

1.4 Literature Review

With the ongoing expansion of society and the economy, as well as the advancement of science and technology, the use of computer information technology in numerous fields has grown in scope. People have started to employ computer technology to create artificial intelligence systems capable of replacing humans. Among these, the use of computer vision technologies is essential. This type of technology can help to ensure product quality while also improving corporate production efficiency[1]. One of the most pressing issues in robotics is the safety of human-computer and human-machine interactions, which necessitates the "explainability" of algorithms, which frequently precludes the potential application of some deep learning solutions, regardless of their performance in pattern recognition applications. [2].

The study in [3] addresses the issue of image registration in printing defect inspection systems, as well as the selection of relevant feature regions. The suggested automatic feature region finding technique for printed image registration makes use of contour point distribution information as well as edge gradient direction, and it can also be used to detect printing defects online.

The [4] contribution describes a camera-based calibration method for optical see-through headsets used in augmented reality applications, as well as for consumer level systems. To estimate the projection parameters of the display model for a generic camera position, the proposed rapid automatic offline calibration approach is based on standard camera calibration and computer vision algorithms. They are then refined using planar homography, and the suggested method is validated using a custom MATLAB application.

Deep neural networks, particularly convolutional neural networks, have changed computer vision during the last decade. However, in order to provide satisfactory results, all deep learning models may want a vast amount of data. Unfortunately, significant amounts of data for real-world situations are not always available, and it is generally understood that a lack of data easily leads to overfitting. This problem can be solved in a variety of ways, one of which is data augmentation. We evaluate existing data augmentation techniques in computer vision applications, such as segmentation and classification, and propose novel strategies in this study. In specifically, [5] present a method for implementing data augmentation through the use of local information in photographs. The random local rotation approach, which involves randomly picking the position and size of circular sections in the image and rotating them with random angles, is parameter-free and simple to apply. It can be used instead of the standard rotation approach, which is prone to uneven image borders. It can also be used to supplement other data augmentation strategies. Extensive testing results and comparisons revealed that the new technique routinely beat its old counterparts in picture categorization, for example.

Deep convolutional neural networks have outperformed several computer vision tasks. However, in order to avoid overfitting, these networks rely extensively on huge data.

Overfitting is the phenomenon in which a network learns a function with extremely large variation in order to perfectly model the training data. Unfortunately, many application domains, such as medical image analysis, do not have access to huge data. This poll is about Data Augmentation, which is a data-space solution to the problem of limited data. Data Augmentation refers to a range of strategies for increasing the size and quality of training datasets in order to build stronger Deep Learning models.

[6] Geometric transformations, colour space augmentations, kernel filters, blending pictures, random erasing, feature space augmentation, adversarial training, generative adversarial networks, neural style transfer, and meta-learning are among the image augmentation technologies included in this survey. This survey extensively covers the use of augmentation methods based on GANs. This study will briefly explore other aspects of Data Augmentation, such as test-time augmentation, resolution impact, final dataset size, and curriculum learning, in addition to augmentation approaches. This survey will cover existing Data Augmentation methodologies, promising developments, and meta-level considerations for Data Augmentation implementation. Readers will grasp how Data Augmentation can improve model performance and expand constrained datasets to take use of big data possibilities.

1.5 Objective

The primary objectives of data augmentation in computer vision applications are:

1 Increase Dataset size: The goal of data augmentation is to increase the size of the training dataset by providing more examples with variances. By artificially boosting the dataset size, more diverse instances are provided for training models, potentially improving their performance and generalization.

2 Improve model robustness: Variations in the training data, such as varied orientations, scales, lighting conditions, and deformations, are introduced through data augmentation approaches. This makes models more robust and adaptive to real-world settings with varied input conditions, resulting in improved performance on previously unreported data.

3 Generalization to new examples: By introducing models to multiple transformations and variations during training, data augmentation helps them generalize better to

previously unseen samples. This reduces overfitting, which occurs when models become overly specialized to training data and fail to perform effectively on fresh data.

4 Mitigating class imbalance: Data augmentation can be used to generate synthetic instances for minority classes in classification tasks with unequal class distributions, balancing the dataset and preventing bias towards majority classes. This enables models to learn more accurate representations for all classes while avoiding biassed predictions.

5 Handling limited training data: In situations when gathering significant amounts of labelled data is difficult or expensive, data augmentation allows for the expansion of the training dataset without the requirement for further manual annotation. This is especially useful when dealing with small or specialised datasets.

6 Enhancing model performance: Models can learn more diversified patterns and achieve better feature representation by introducing variations and enriching the training data. This can increase computer vision models' accuracy, precision, recall, and overall performance.

Overall, the goals of data augmentation are to improve the quality and diversity of training data, improve model performance and generalisation, reduce overfitting, and solve issues associated with limited or imbalanced datasets.

1.6 Scope of the Project

The project's scope, "Optimization of annotation process for segmentation and detection model training," entails enhancing the efficiency and efficacy of the annotation process in the field of computer vision, specifically for training segmentation and detection models. The project's goal is to investigate and develop optimization strategies to solve the issues and limits connected with the annotation process.

Here are some key aspects within the project scope:

1. Annotation work flow analysis: Identifying bottlenecks, inefficiencies, and areas for improvement in the present annotation procedure. Understanding the existing annotation process, tools, and methodologies is required.
2. Automation and tool selection: Investigating and analysing annotation tools and software that can help to streamline the annotation process, increase productivity, and improve annotation quality. This may entail conducting research and selecting

appropriate software, as well as designing unique tools tailored to the specific demands of segmentation and detection activities.

3. Quality Control and validation: Creating methods and methodologies for annotated data quality control and validation. This may entail developing criteria, executing tests for annotation accuracy, and establishing validation mechanisms to assure the annotations' dependability and consistency.
4. Performance evaluation: Evaluating the influence of the optimised annotation process on model performance. This entails using annotated datasets to train segmentation and detection models and evaluating the resulting model's accuracy, precision, recall, and other relevant performance metrics.
5. Documentation and best practices: Creating a reference for future projects by documenting the optimised annotation process, including tools, methodologies, and recommendations. Best practises, lessons gained, and ideas for effective annotation methods relevant to segmentation and detection model training may be included in this work.

The scope of the project will differ depending on the individual requirements, resources, and objectives. It is critical to specify specific goals and deliverables while taking into account realistic limits and the viability of implementing optimisations within the project timeline.

Chapter 2: Data Augmentation

2.1 Data Augmentation: Definition, uses and importance.

Data augmentation is a machine learning technique for generating new data from existing data. It entails transforming the original dataset in numerous ways, such as rotation, flipping, cropping, and changing the colour or brightness of images. The purpose of data augmentation is to improve the diversity of the dataset, allowing machine learning models to learn and generalize better from the given data.

One of the most common applications of data augmentation is to prevent overfitting. Overfitting happens when a machine learning model grows overly complicated and closely matches the training data, resulting in poor performance on new data. Machine learning models can be trained on a more diversified and broad set of data by producing fresh data through data augmentation techniques, lowering the risk of overfitting.

Another use of data augmentation is to improve the accuracy and effectiveness of machine learning models in real-world scenarios. It is also useful in situations where there is limited training data available. In such cases, data augmentation can help generate additional training data, making it possible to train a machine learning model on a more extensive and diverse set of data.

2.2 Data Augmentation Tools And its Features

***Open-source data Augmentation tools*:** Albumentations, imgaug, Augmentor, TensorFlow Data Augmentation, Keras ImageDataGenerator, etc.

Features of these above tools:

i. Augmentation techniques: Open-source data augmentation tools offer a wide range of techniques to modify input data, such as flipping, rotation, scaling, cropping, shearing, etc

ii. Batch processing: Open-source data augmentation tools often support batch processing of input data, allowing users to augment large datasets efficiently.

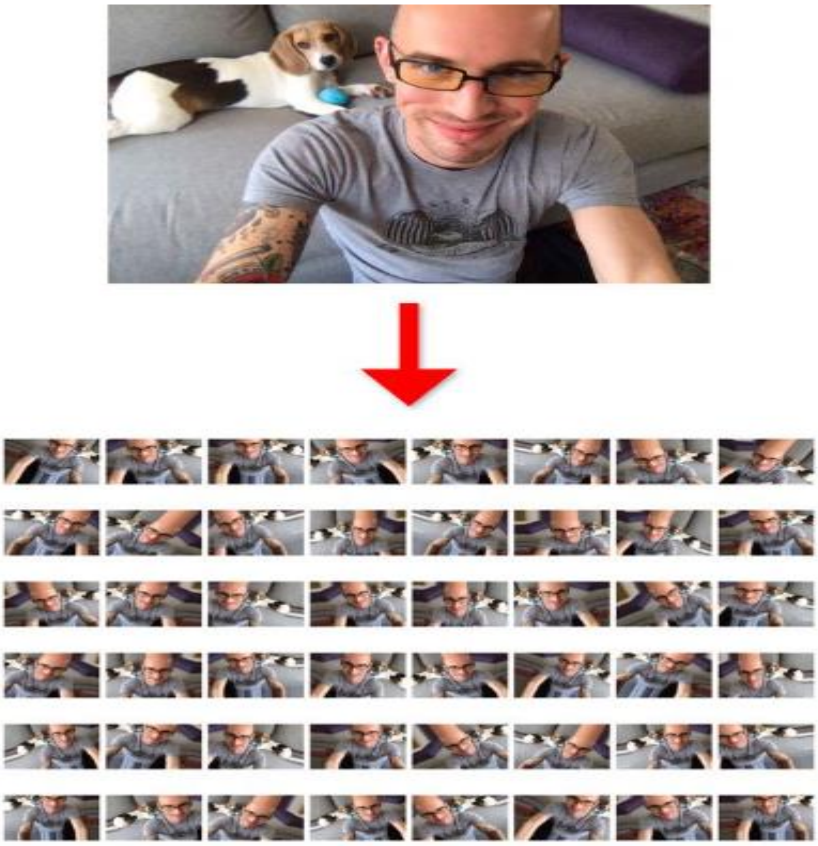
iii. Ease of use: Open-source data augmentation tools often have a user-friendly interface, making them accessible to users with little to no programming experience.

Paid Tools: Roboflow, Labelbox, SuperAnnotate, Hasty.ai, Databricks.

Features of these above tools:

i. Offers over 100 data augmentation techniques
ii. Offers team collaboration and project management features
iii. Supports multiple annotation types, including bounding boxes and segmentation masks
iv. Some of the tools provides a cloud-based machine learning platform with built-in data augmentation features

Overall, paid data augmentation tools offer advanced features, customization options, and greater convenience compared to open-source tools. So to get this advanced features we developed the code which will be useful in BIS(Blister Inspection System).



(i): Keras ImgDataGenerater

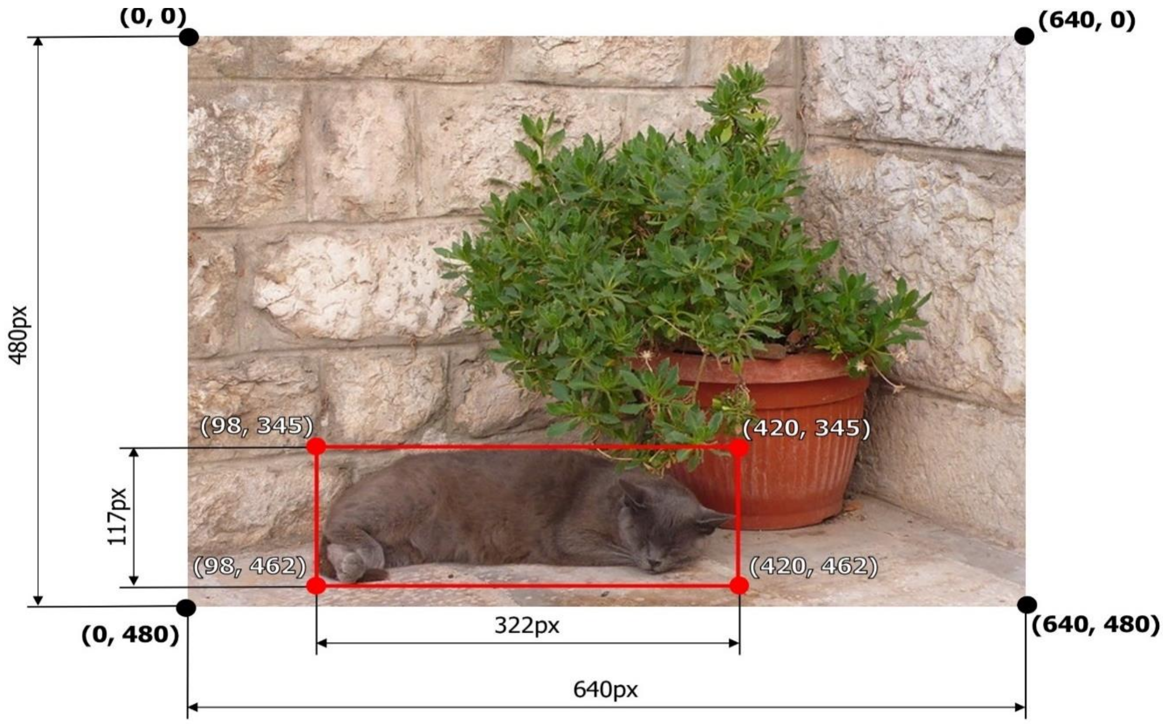
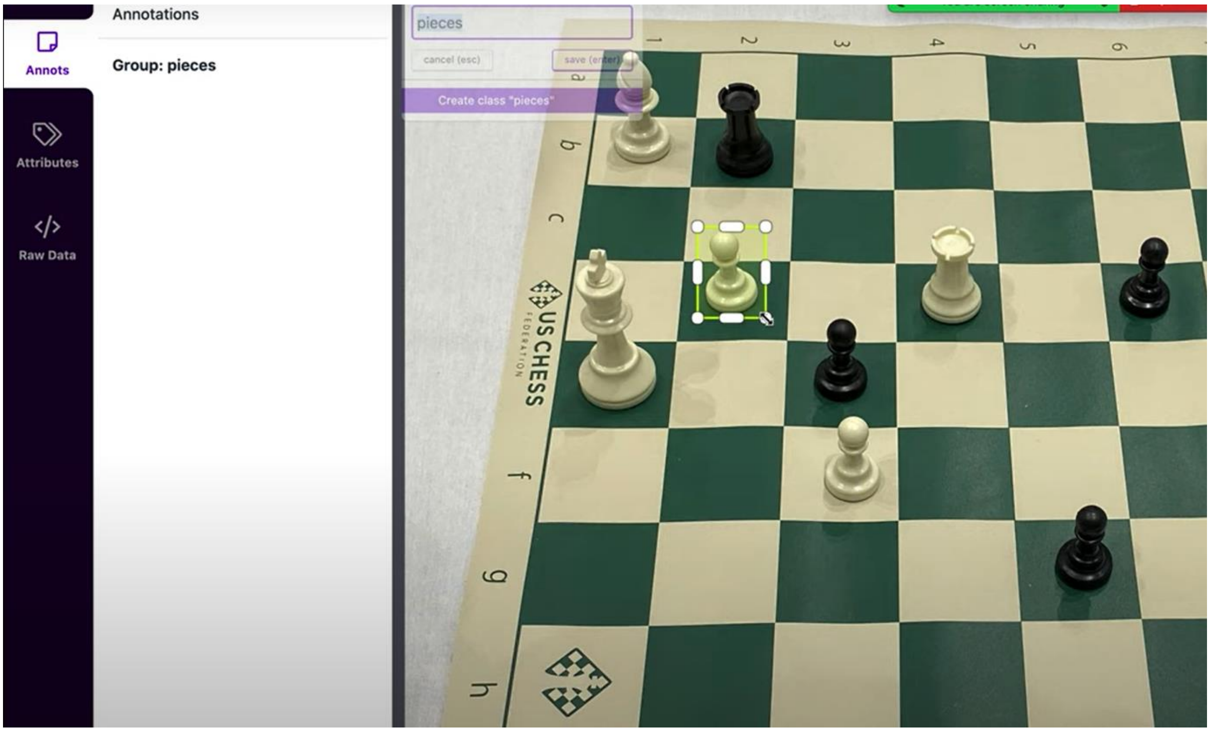


Figure 1: Open Source Tools: (ii) Albumentations



(i) Labelbox

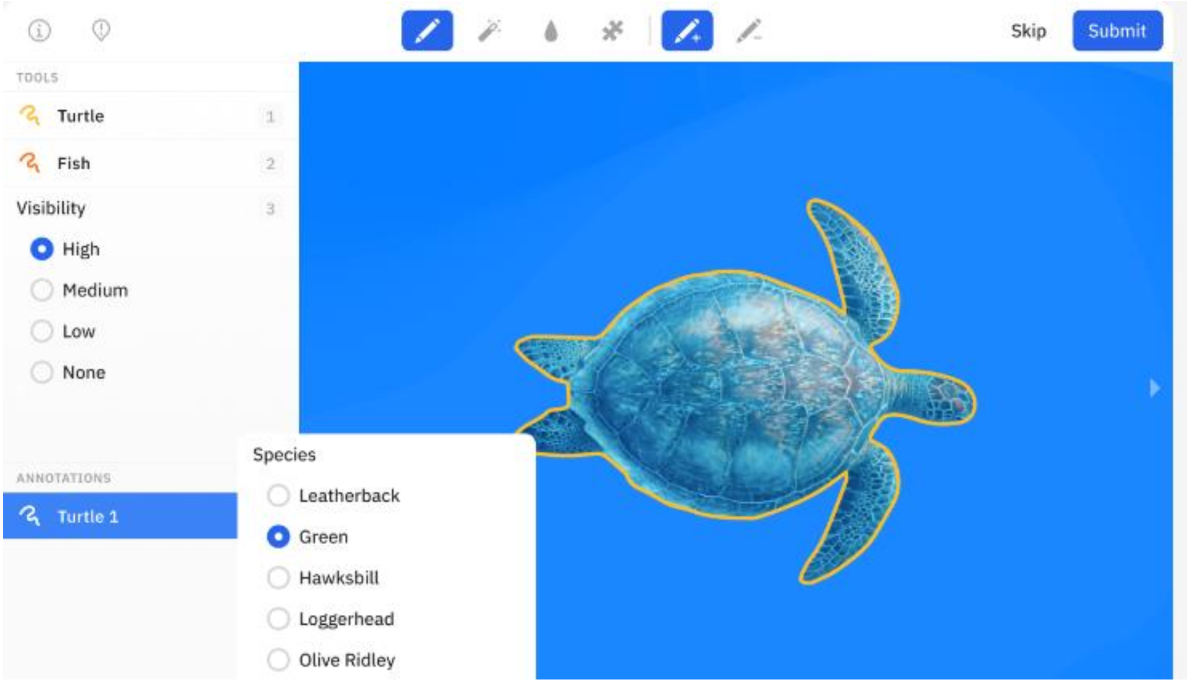


Figure 2: Paid tools: (ii) Roboflow

Chapter 3: Developing the features.

3.1 Overview

This project is the result of our summer internship training which is based on work in the field of image processing and machine learning. The major goal of this project is to reduce the dependency on the annotation tool which was used during the implementation.

During this internship, Developed various features for annotation tools which is implemented in plenty of medicinal Tablet and Capsule images which is generated by company's Tablet Inspection System

3.2 Feature1: Contour based annotation: semi-automatic annotation.

Contour-based annotation is a type of image annotation technique used in computer vision and machine learning to label the boundaries of an object or region of interest in an image.

In this contour-based annotation method we create a series of curves or lines around the item or region of interest in the image using image processing techniques, forming a closed contour that outlines the object. The contour is a set of (x,y) coordinates that describe the location of the curve or line segments.

Then extracting this set of coordinates and dump this data into a Labelme Json format-based annotation files.

The Labelme JSON format consists of a dictionary that contains the following keys:

(i) "version": indicates the version of the annotation format, which is currently set to "4.5.7"

(ii) "flags": contains optional flags for the annotation

(iii)"shapes": contains a list of shape annotations, where each shape is defined as a dictionary with the following keys:

(iv)"label": the label or class name of the shape

(v) "points": a list of (x, y) coordinates that define the vertices of the shape

(vi)"group_id": an optional ID that can be used to group shapes together

(vii) "shape_type": the type of shape, which can be "polygon", "rectangle", "circle", "line", or "point"

(viii) "flags": contains optional flags for the shape annotation

When compared to other annotation approaches such as bounding boxes or polygons, contour-based annotation delivers more accurate and detailed information on the shape and location of items in an image. Furthermore, contour-based annotation is more efficient than pixel-level annotation, which entails assigning a class name to each pixel in a picture.

Other advantages of this methods are

- ➔ Reduced annotation time: Contour-based annotation just requires annotators to sketch the boundary of an object, which is faster and less time-consuming than other approaches such as pixel-level annotation.
- ➔ Improved generalization: Contour-based annotation offers more information about the shape of objects in a picture, which can aid machine learning models in generalizing to previously unseen examples.
- ➔ Consistency: Contour-based annotation ensures that objects are labelled consistently across numerous photos, which is critical for training correct machine learning models.

Flow Of this Algorithm

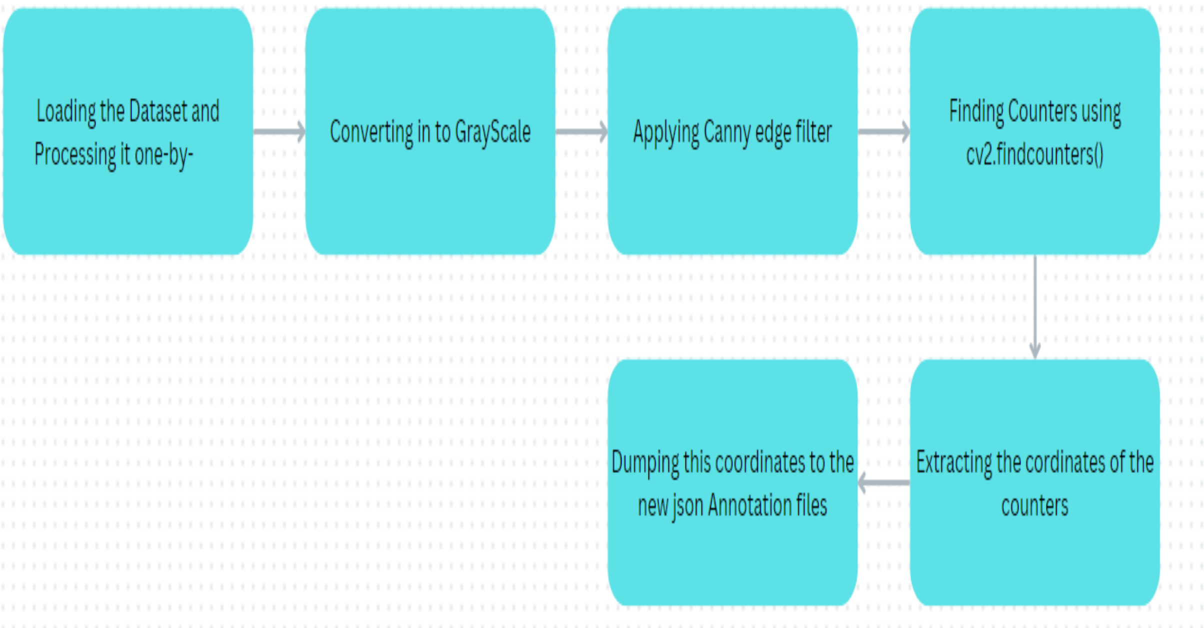


Figure 3: Flow chart for Contour based annotation.

1. Loading the whole data of images for contour-based annotations.
2. To find contours we need to first convert images into gray scale.
3. So that we can apply canny edge filter and find contours using cv2.findcontours()
4. Creating an json file using python scripts and dumping the coordinates of contours into it.

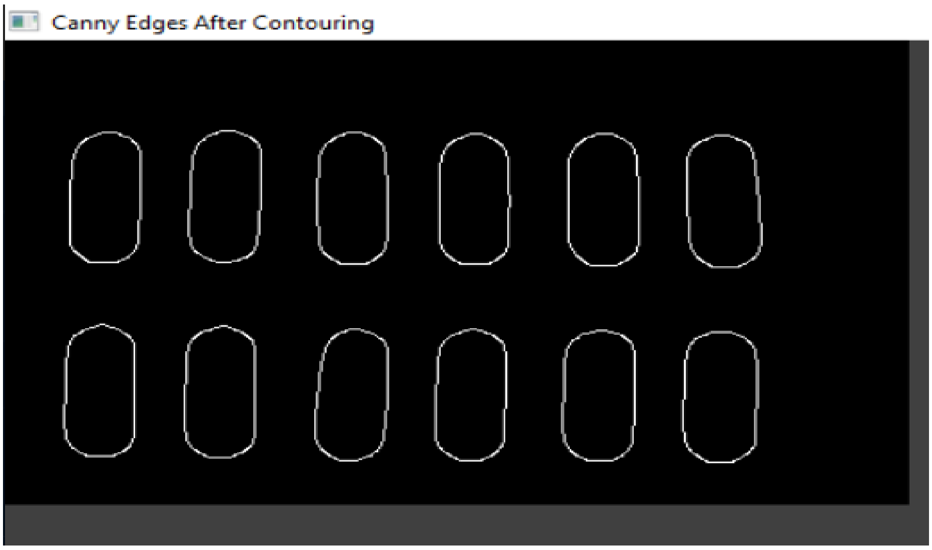


Figure 4(i): Result of Canny edge filter

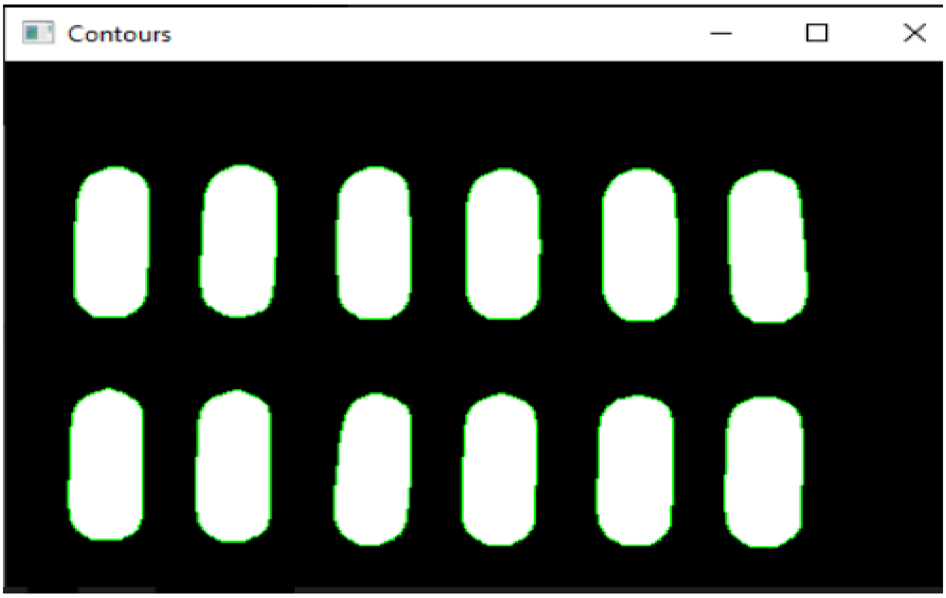


Figure 4(ii): Finding contours

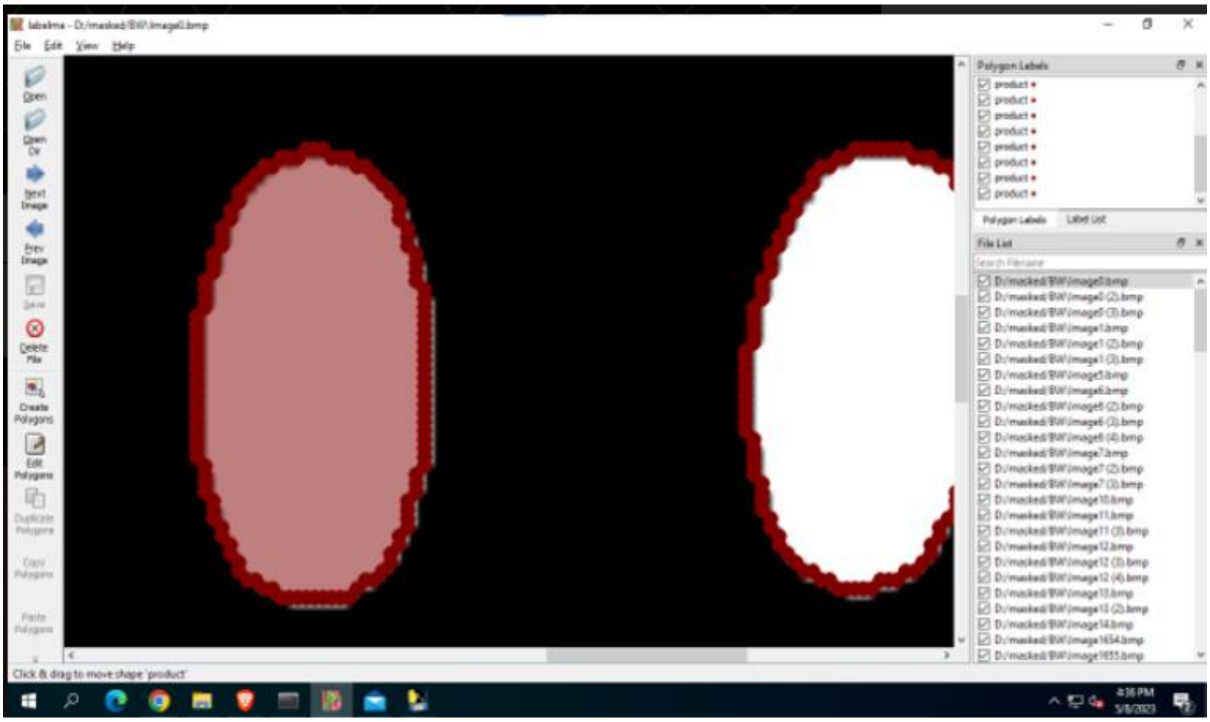


Figure4(iii): Contour based Annotations

3.3 Feature2: COCO to Labelme

Figure 5: Sample of Annotation in Labelme

The COCO (Common Objects in Context) format is a popular format for annotating object recognition and segmentation information. The COCO format defines a JSON file format that includes information about the images, annotations, and categories in the dataset. key components of the COCO format:

I. Image information of every single image in dataset.

II. Annotations: including the ID of the image and the coordinates of annotations

III. Categories: each category has assigned a unique ID and name.

Developed a feature for converting multiple Labelme json file to single COCO format file and Vice versa.

Need/Advantages of this feature.

1. ***Better organization and management of annotations:*** COCO format allows for a single annotation file to contain all annotations for a dataset, making it easier to manage and organize annotations.
2. ***Compatibility with popular deep learning frameworks:*** Many deep learning frameworks such as TensorFlow, PyTorch, and MXNet support COCO format natively, making it easier to use the annotated data in these frameworks.


```
{
    "images": [
        {
            "height": 270,
            "width": 1276,
            "id": 1,
            "file_name": "final_1_BISM_23012022_193304_R42.bmp"
        },
        {
            "height": 270,
            "width": 1276,
            "id": 2,
            "file_name": "final_1_good (2).bmp"
        },
        {
            "height": 570,
            "width": 1408,
            "id": 3,
            "file_name": "final_1_good (3).bmp"
        },
        {
            "height": 626,
            "width": 1430,
            "id": 4,
            "file_name": "final_1_good (4).bmp"
        }
    ],
    "categories": [
        {
            "supercategory": "product",
            "id": 1,
            "name": "product"
        }
    ],
    "annotations": [
        {
            "segmentation": [
                [
                    65.21739130434783,
                    78.71853546910755,
                    114.8741418764302,
                    49.88558352402746,
                    123.11212814645309,
                    46.224256292906176,
                    130.43478260869566,
                    45.995423340961096,
                    136.38443935926773,
                    50.34324942791762,
```

3. ***Improved training performance:*** Since COCO format supports annotations for segmentation, bounding boxes, and key points, it can provide more detailed information to the deep learning model, leading to improved performance.
4. ***Efficient data loading:*** With COCO format, it is possible to load annotations directly from a JSON file, which can save time and reduce the chances of errors during the loading process.

3.4 Feature 3: Flipping and rotating the images and its Annotations

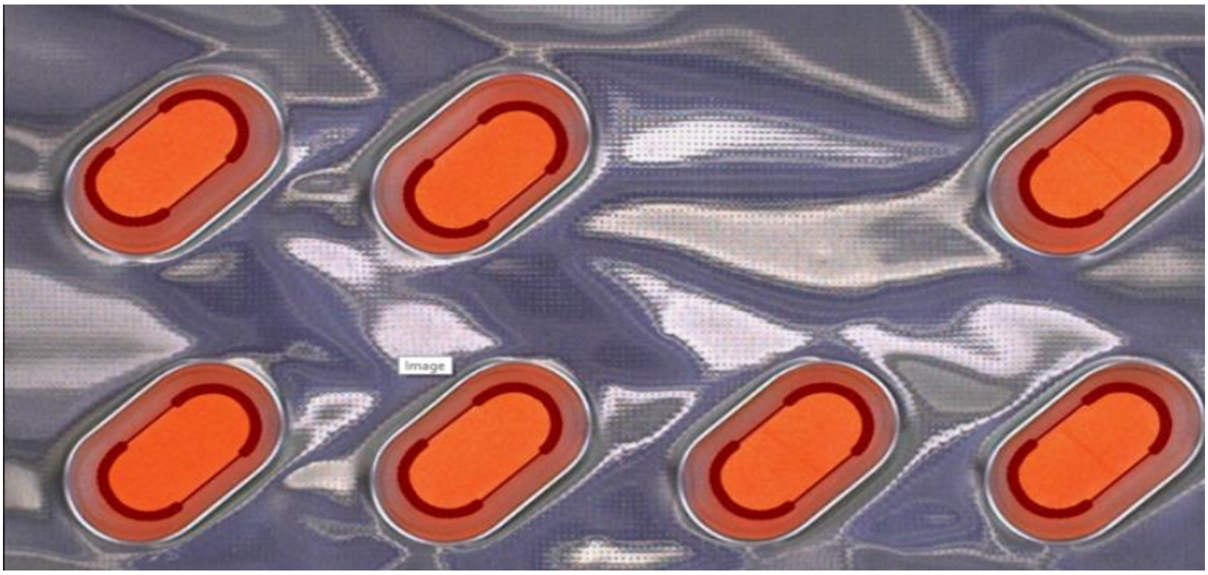


Figure 7(i): Original Annotated Image

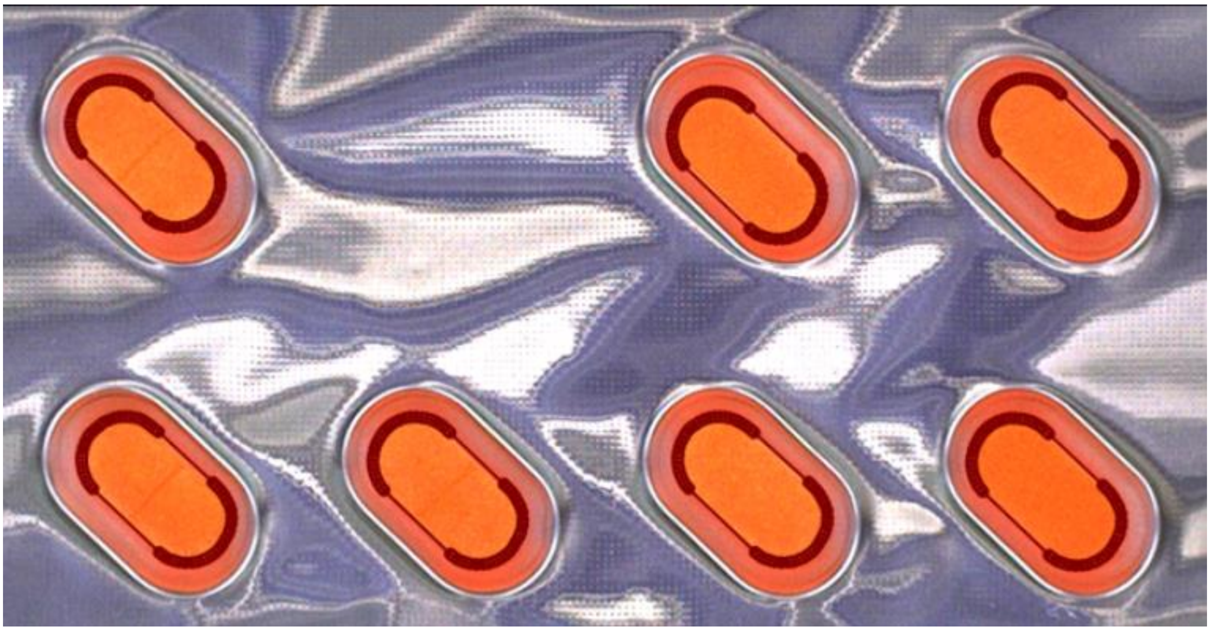


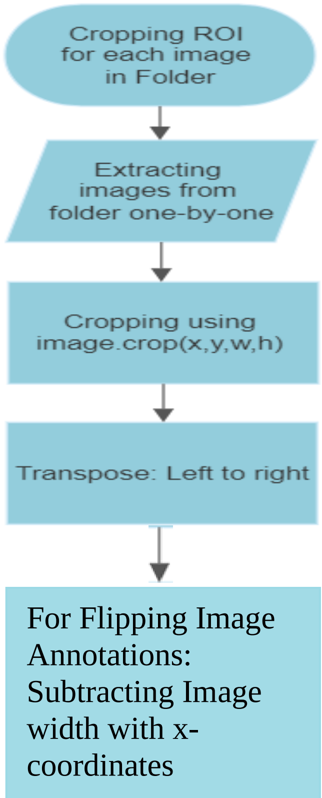
Figure 7(ii): Flipped Image along with Annotations

Flipping is a transformation that creates a mirror image of the original image along the horizontal or vertical axis. The pixels in the image are reorganized so that the original image's left side becomes the flipped image's right side, and vice versa.

Because it increases the number of images in the dataset without requiring additional data gathering activities, the flipping technique is excellent for providing more varied and diversified training data. Furthermore, flipping can aid in the reduction of overfitting by introducing differences in the training data that the model must learn to generalise to.

Flow the this Method:

Figure 8: Flow of Flipping image annotations



1. Cropping the Image to get Proper ROI (region of interest)
2. Extracting images One-by-one and apply transpose operation for Flipping Images vertically.
3. To also rotate annotation file we first extract coordinates from the json file and update its x-coordinate by subtracting image width with itself

Rotating:

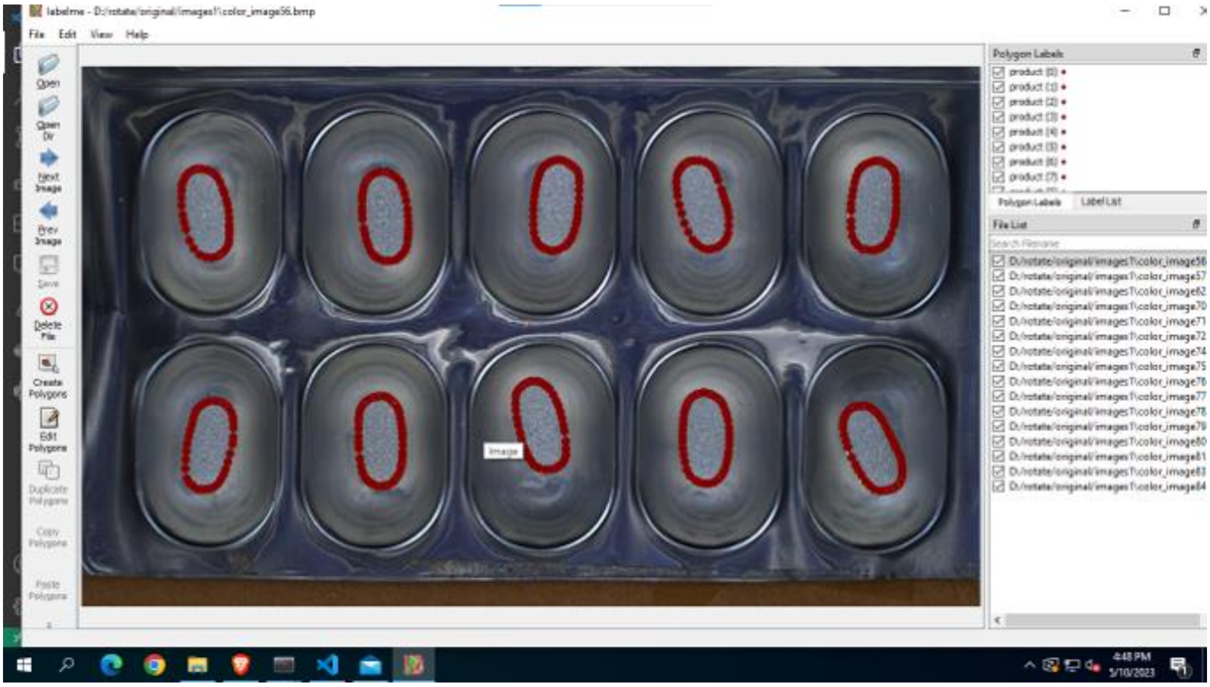


Figure 7(iii): Original Annotated image

Rotating Images along with its annotations there's a method named Affine transformation. through this method Images can be rotated in any angle. For rotating Annotations same method of Affine transformation is used. Extracting annotation Points from the json file and applying affine transformation to it.

Affine transformation is a geometric transformation technique used in computer vision and image processing to apply to an image or object a combination of translation, rotation, scaling, and shearing operations. Straight lines, parallelism, and distance ratios between points are all preserved. picture registration, picture alignment, and data augmentation are all frequent applications for affine transformations.

The key operations involved in an affine transformation are:

(i) Rotation: Rotating the image or object by a certain angle around a specified center point.

(ii) Scaling: Resizing the image or object by multiplying the coordinates by scaling factors in the x and y directions.

(iii) Shearing: Distorting the image or object by adding a shearing effect in the x and/or y direction, causing the image or object to stretch in one direction while compressing in the perpendicular direction.

Image registration, panorama stitching, object detection, and data augmentation are all common uses for affine transformations in computer vision. Affine transformations can be performed to training images in data augmentation to yield augmented samples with variations in position, rotation, scale, and shear. This helps to diversify the training dataset and improves the model's capacity to generalise to new contexts and points of view.

Figure 7(iv): Rotated Image along with is annotations

3.5 Feature 4: converting annotations from Labelme Json to LabelImg

If you are working with a deep learning framework that supports YOLO format annotations, such as Darknet or YOLOv5, converting Labelme JSON annotations to YOLO format can be handy. You can train your object identification models using these frameworks by converting your annotations to YOLO format.

Furthermore, YOLO format annotations are represented differently than Labelme JSON annotations, which are based on polygonal shapes. In some circumstances, YOLO format annotations use bounding boxes to identify object locations, which can make training easier and faster.

Convolutional neural networks (CNNs) are used by YOLO to estimate the likelihood of object existence and the bounding box coordinates of each item by splitting the input picture into a grid of pixels. The output in its ultimate form is a list of objects together with their respective bounding box coordinates, confidence scores, and class labels.

YOLO is a single-shot detection approach that completes object recognition and classification in one forward pass over the network, as opposed to conventional object detection algorithms, which call for numerous passes over an image and a time-consuming post-processing phase. Because of this, YOLO is substantially quicker than other object identification algorithms, making it ideal for real-time applications like robotics, video monitoring, and self-driving automobiles

There are several advantages to using YOLO format annotations over other annotation formats like Labelme JSON:

I. Simplicity: YOLO format annotations indicate item locations using bounding boxes, which is a simpler and more straightforward approach than polygonal shapes or other more sophisticated annotations.
II. Speed: Because YOLO style annotations are simpler and more compact than other annotation formats, deep learning frameworks and computer vision

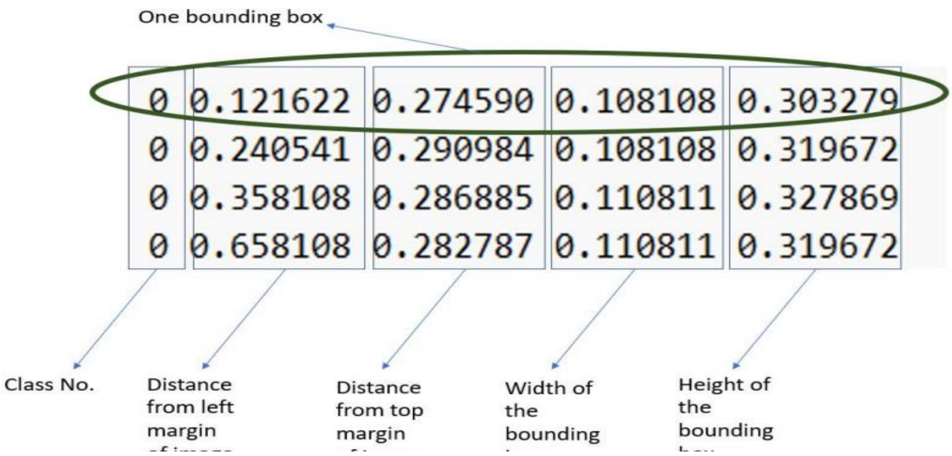


Figure 9: YOLO File Format

algorithms can analyze them more quickly. This may result in speedier training and improved performance on huge datasets.

III. Compatibility: Many major deep learning frameworks and computer vision libraries, including Darknet, YOLOv5, TensorFlow, and PyTorch, support YOLO format annotations. This enables working with YOLO format annotations across platforms and programming languages simple.

Working of this algorithm

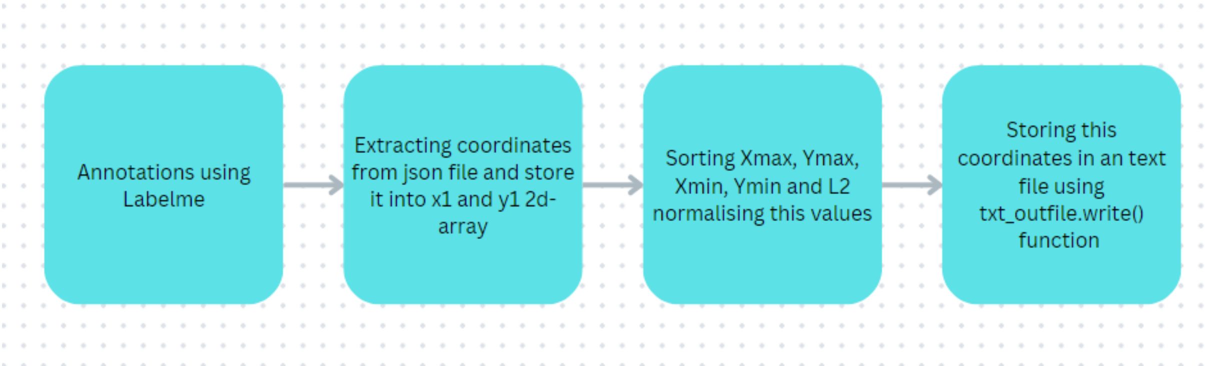


Figure 10: Flow of Labelme Json to LabelImg Txt

1. Using our first feature we can annotate the images and get the Json files.
2. This json files contains coordinates of contours or labeled products polygonal share.
3. In Yolo format we need to generate bounding boxes around our product.
4. So to generate bounding boxes we extract the coordinates from json files, and filter the x-min, x-max, y-min, y-max from each labeled product to dump this data into a TXT format.
5. Before Dumping YOLO format has normalized Values, so we need to normalize x-min, x-max, y-min, y-max these values using L-2 normalization.
6. And then storing this values into TXT file

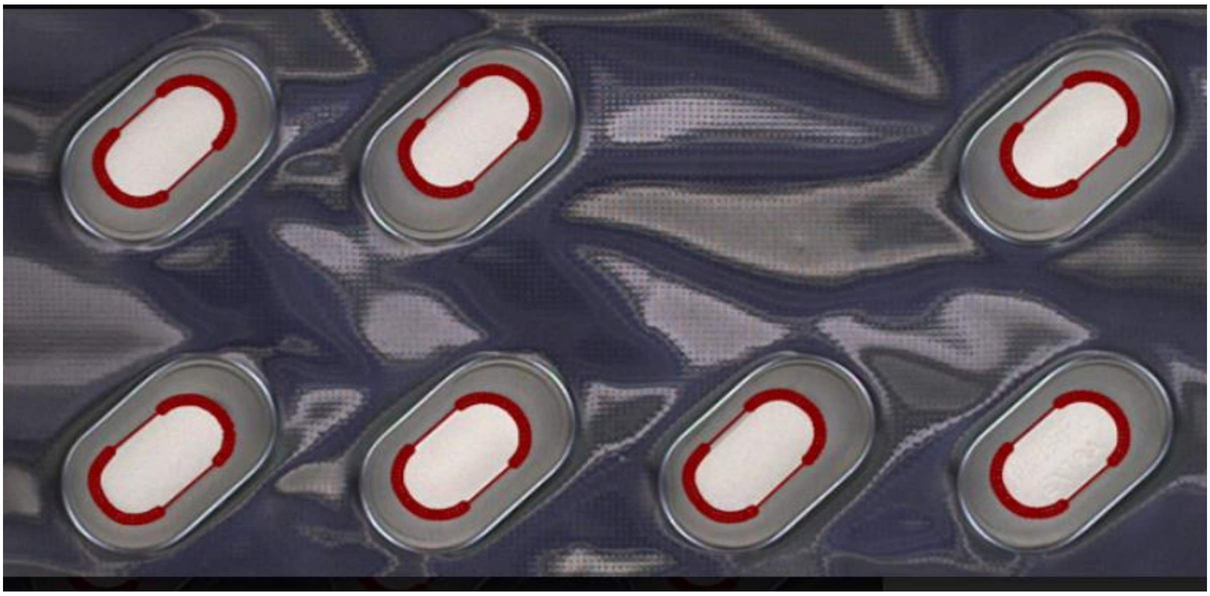


Figure 11(i): 1st Image with annotations in Json

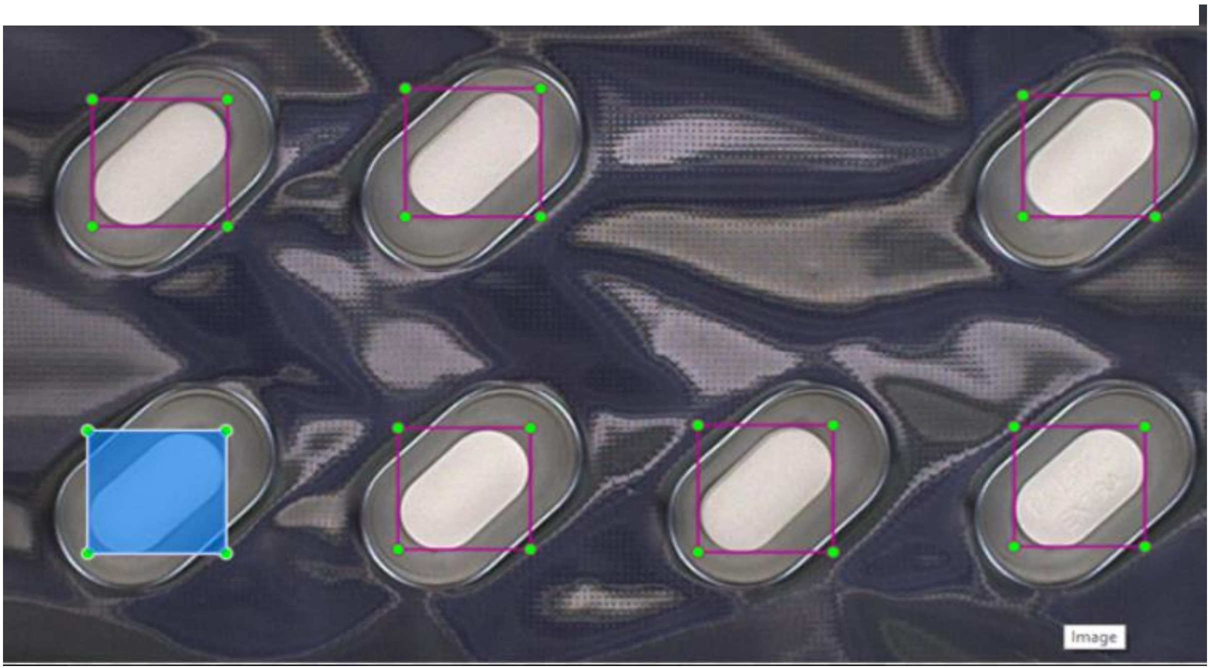


Figure 11(ii): 1st Image Annotations Converted to YOLO format From Json

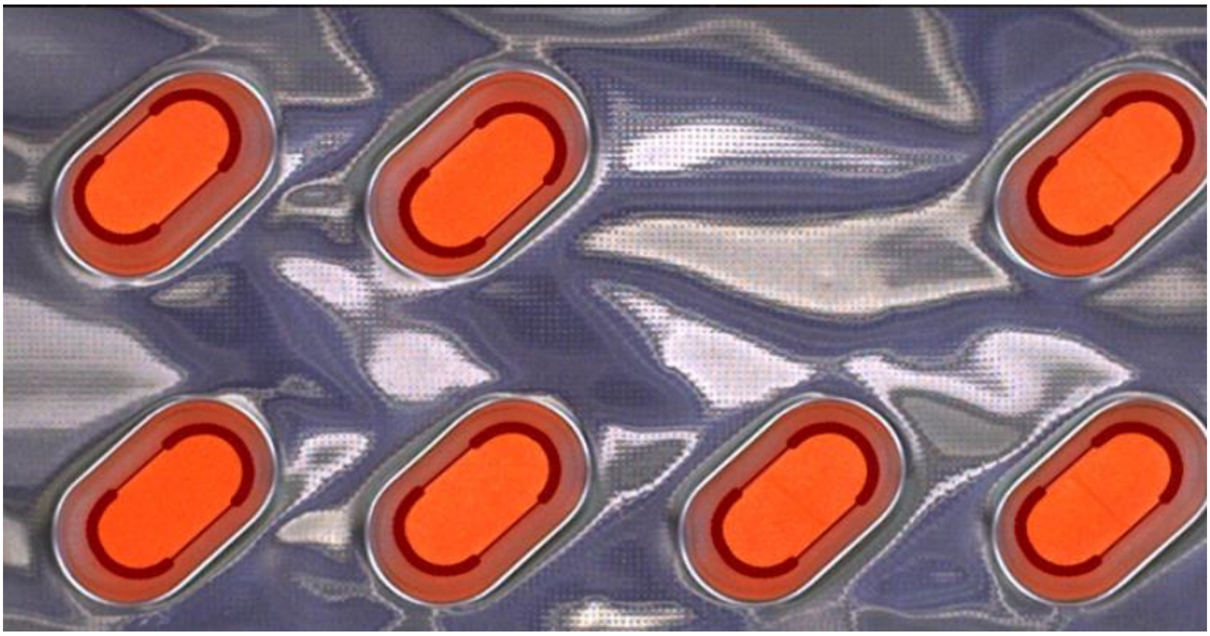


Figure 11(iii): 2nd Image with annotations in Json

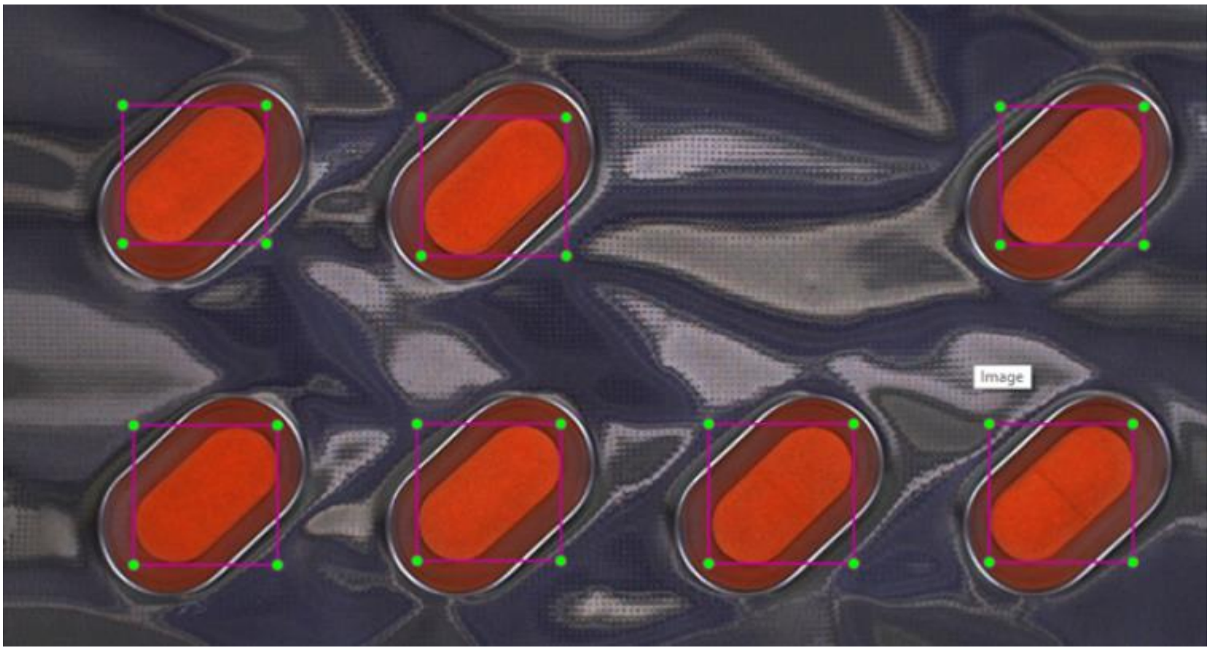


Figure 11(iv): Labelme Json to LabelImg YOLO

3.6 Feature5: Cropping for Multiple product Image using YOLO format Annotations.

This feature simply crops the images from their annotations which we have converted from above feature.

To develop this we extracted the Bounding box coordinates from TXT file and then applying img.crop() function from PIL library to the Images so that we can get images of every single capsule so that we can use it for training and testing purposes in future.

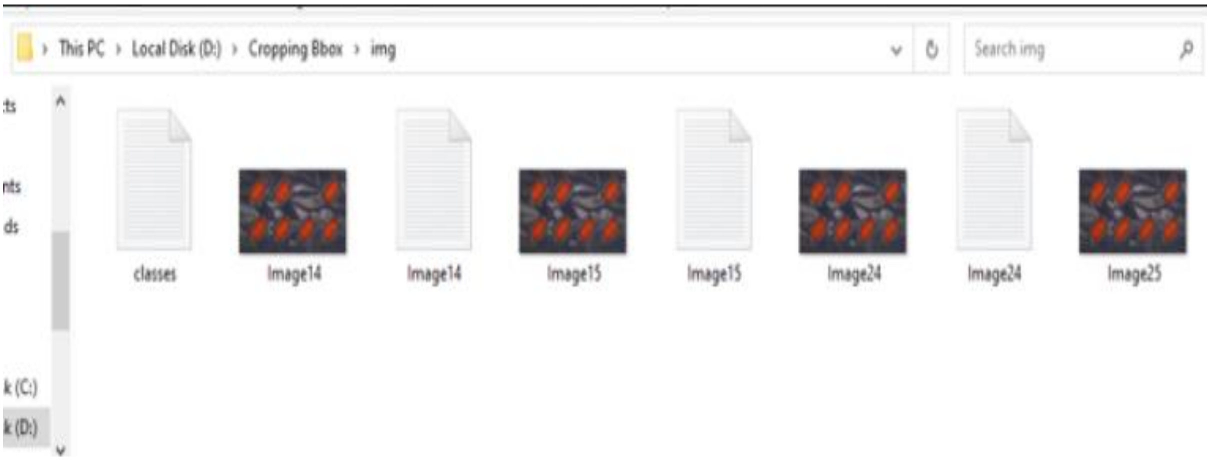


Figure 12(i): Image with its YOLO annotation file

Figure 12(ii): Cropping for Multiple product Image.

Chapter 4: Implementation of tools

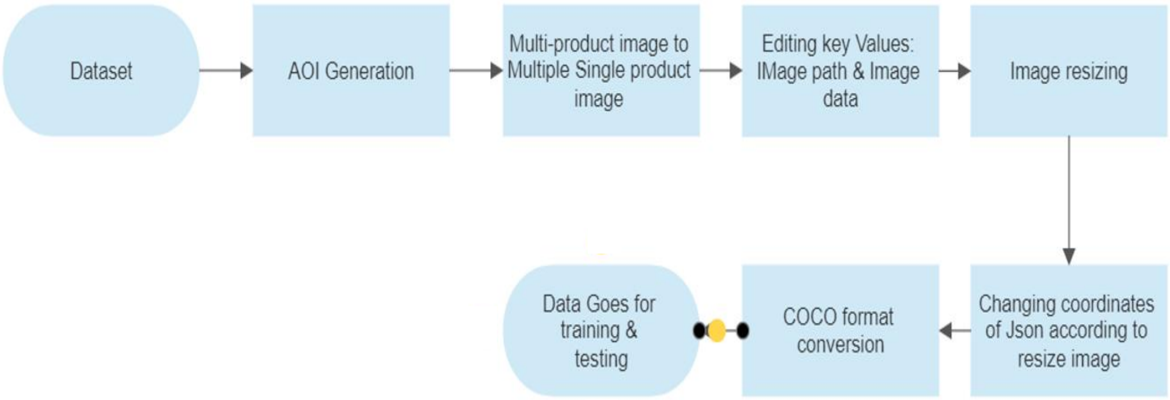


Figure 11: Final Pipeline for Data Augmentations

Before the Data goes of training and testing, we need to do various pre-processing tasks which also include Data Augmentation. The Above features which we have developed can be used in various stages of data pre-processing. And Currently these features are provided by the paid software/platforms.

4.1 AOI Generation

AOI stands for Area of Interest which can also be referred to as ROI (Region of Interest). These are referred to as the sample within the given data set. The Concept of AOI is used in various application areas. For example: In the field of medical (medical imaging) or pharmaceutical companies. The images are taken and using the AOI techniques some specific boundaries/segments are extracted. And that extracted samples can be further used/given for training in a Machine learning model. The benefit of using AOI is that it helps to reduce unnecessary interference in the images and also helps to track the images fast as the pixel size is reduced.

AOI can be in any shape such as rectangular, elliptical, or freehand shape. And can be taken from any point of the image. But during the time of analysis, we used the center AOI method for cropping purposes because it matches our requirements. The product given were aligned to the center so it was easy to cut them through the center.

The implementation was done using **OpenCV/python**. So below is the process to obtain it.

i. Import necessary libraries.
ii. Then take images as input.
iii. Use function "Select ROI".
iv. Select the rectangular points.
v. Cropping can be accomplished with the help of these points.
vi. After that Display the images
vii. Then save the AOI images.

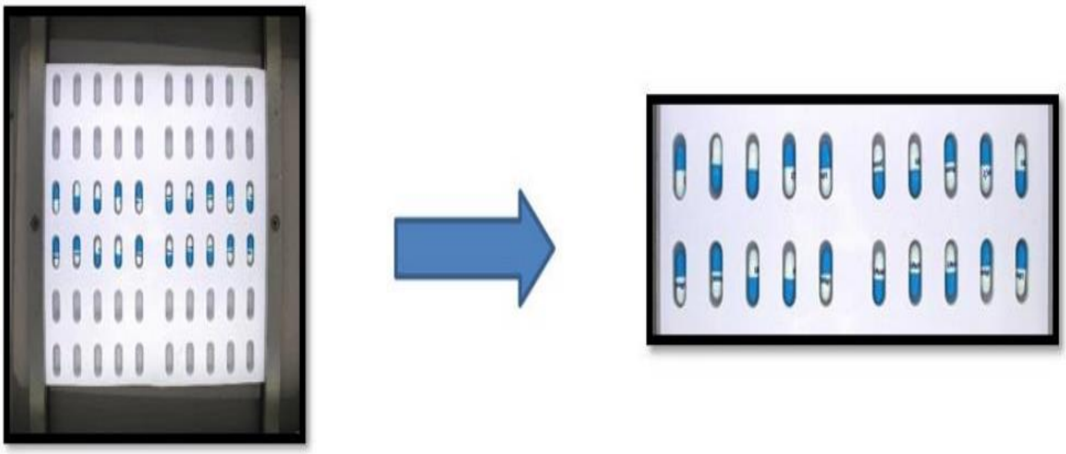


Figure 12: Result of AOI generation

4.2 Image resizing And flipping and change in coordinates accordingly

In this process, images were resized according to the aspect ratio like 50%,75%, or 200% of the original images that is compression or expansion of the images.

But its corresponding JSON has the coordinate point according to the old image's pixels which will lead to an error in the annotation. The solution to this problem was to annotate the images again. But instead, it was found that if we take the aspect ratio that is by how much ratio the height and width of the image are changed will be multiplied by the coordinates of the old JSON which will lead to a similar annotation.

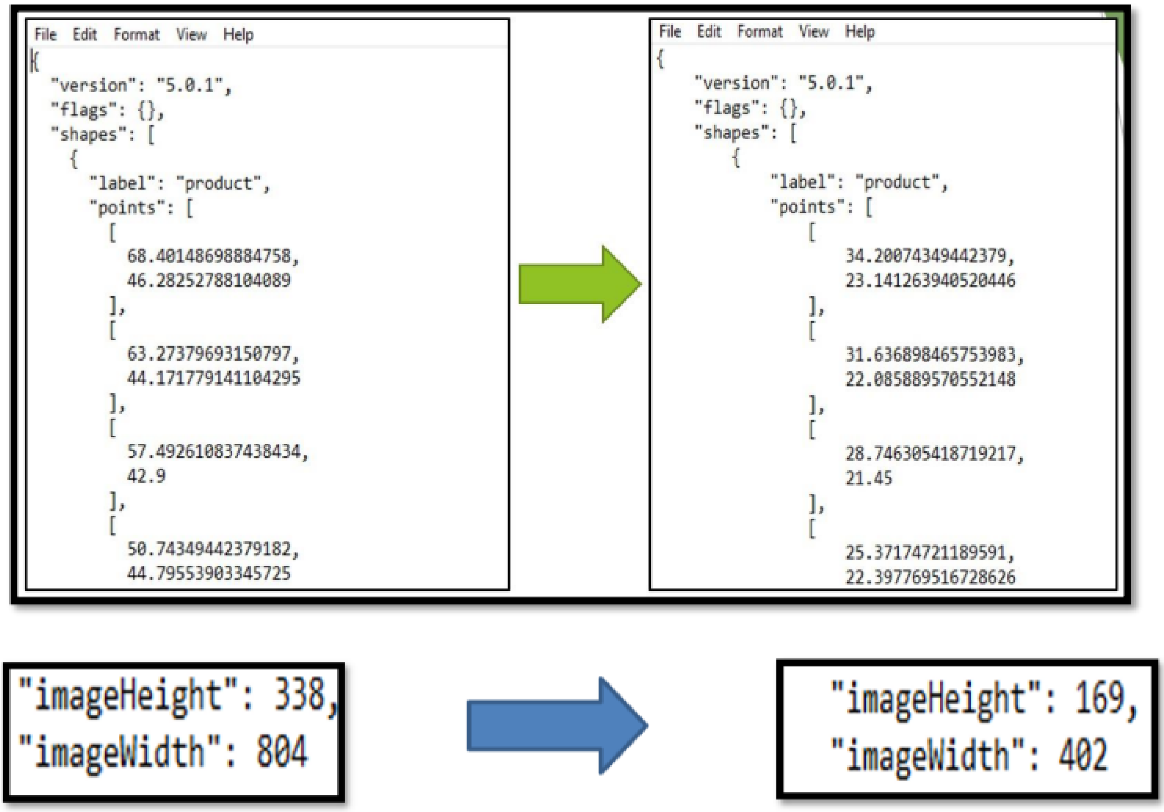


Figure 13: Change in Coordinated after Resizing

4.3 Labelme JSON to COCO JSON

COCO (Common Objects in Context) is a large dataset used for computer vision applications such as object detection, segmentation, and captioning. It was created by and is one of the most often used benchmarks for object identification and image interpretation systems. In 2014, Microsoft and Cornell University partnered.

Descriptions, object instances, and object categories are associated to over 330,000 pictures in the COCO dataset. The collection has 80 different types of objects, including humans, animals, cars, and furniture. In addition to object recognition and segmentation, the collection includes annotations for key point detection, which is the process of localizing specific features of an item, such as a person's joints.

The COCO dataset has been used to train and evaluate a wide range of cutting-edge object identification and segmentation approaches, including Faster R-CNN, Mask R-CNN, and YOLO. The dataset has also been used in a number of challenges and contests, including the COCO Captioning Challenge and the COCO Detection Challenge, which have encouraged more research and improvement in the field of computer vision.

The availability of COCO, which is widely considered as a benchmark dataset for object detection and image interpretation, has significantly aided computer vision research.

A number of machine learning techniques and frameworks have been developed to interface with the COCO dataset for a variety of computer vision applications such as object detection, segmentation, and captioning. Here are a few examples:
Facebook AI Research (FAIR) developed Detectron2, a well-known open-source software framework for object detection and segmentation. Detectron2 provides pre-trained models on the COCO dataset that can be customised for specific uses.

TensorFlow Object Identification API: Created by Google, this framework provides a collection of models that have already been trained to perform tasks such as object identification, segmentation, and classification. The COCO dataset training is facilitated by the API, which also includes analysis and assessment tools.
Mask R-CNN is a deep learning architecture developed by FAIR for example segmentation, which involves locating and classifying each occurrence of an object in an image. Mask R-CNN was trained on the COCO dataset and now performs at the cutting edge of object segmentation.

Difference between Labelme keys and COCO keys are as follow:

Labelme JSON keys:

- Version
- Flags
- Shapes
- Labels
- Points
- Group_id
- Imageheight
- Imagewidth
- Imagedata

COCO JSON keys:

- Images
- Categories
- Annotation
- Segmentation
- Height
- Width
- Id
- File_name
- Category and their ids

```
        }
    ],
    "categories": [
        {
            "supercategory": "product",
            "id": 1,
            "name": "product"
        }
    ],
    "annotations": [
        {
            "segmentation": [
                [
                    41.6,
                    26.0,
                    44.0,
                    25.1,
                    46.4,
                    25.0,
                    49.0,
                    25.7,
                    51.7,
                    27.1,
                    54.3,
                    28.7,
                    55.9,
```

```
{
    "images": [
        {
            "height": 122,
            "width": 370,
            "id": 1,
            "file_name": "Screenshot 2023-03-03 215639.jpg"
```

Figure 14(i): COCO format Json file

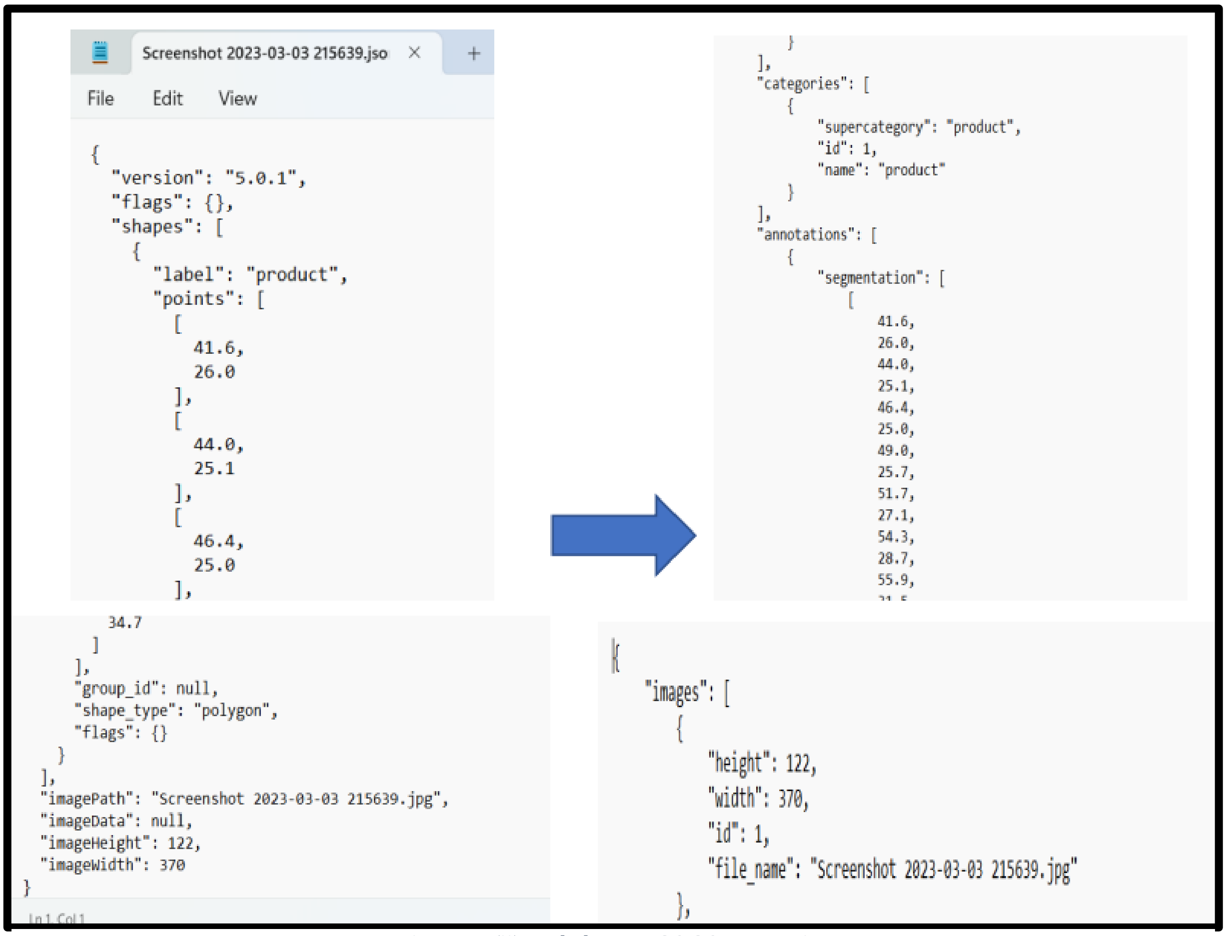


Figure 16(ii): LabelME to COCO conversion

Chapter 5

5.1 Conclusion

Finally, the annotation process must be optimized for successful segmentation and detection model training. Because human annotation can be time-consuming and error-prone, there are various annotation tools available to automate and streamline the process. The annotation tool to be used is determined by the type of annotations required and the specific project needs.

In addition to choosing the correct annotation tool, other factors such as dataset size, annotation quality, and consistency should be considered. The employment of many annotators and quality control mechanisms can help to reduce annotation errors and assure high-quality annotations.

Finally, choosing the right segmentation, detection, and hyperparameters might have an impact on how successfully the model is trained. Experimentation and repeated hyperparameter adjustment can improve the model's performance. By speeding the annotation process and model training, the accuracy and efficacy of segmentation and detection tasks in computer vision can be increased overall, resulting in higher performance in real-world applications.

As a result, we use the COCO and YOLO formats for model training and testing, which are among the most widely used formats worldwide.

5.2 Future Scope

In the optimisation of the annotation process for segmentation and detection model training, there is a lot of space for future growth and progress. Here are some ideas for future research and innovation:

1. Annotation tools can be coupled with well-known deep learning frameworks such as TensorFlow and PyTorch, resulting in more fluid and effective workflows for training and annotating data.
2. Improved automation: The development of more complex automated annotation systems capable of correctly recognising and naming objects in still and moving

images can significantly reduce the requirement for manual annotation while increasing the efficiency and precision of the annotation process.

3. Improves Quality Control: Quality control measures such as automated annotation verification and inter-annotator agreement may be developed further to ensure that annotations are correct and consistent.
4. Integration with new modalities: Annotation tools can be improved to allow for the annotation of new modalities, such as 3D images and point clouds, bringing up new opportunities for applications in areas such as robotics and augmented reality.
5. Multimodal annotation: The ability to annotate multiple modalities simultaneously, such as photographs and sounds, may enable new applications such as multimodal object recognition and scene understanding.

In general, research on the optimisation of the annotation process for segmentation and detection model training is ongoing, and future advances could lead to enhanced precision, efficacy, and new applications in computer vision and other domains.

References

1. Y. Li and Y. Zhang, "Application Research of Computer Vision Technology in Automation," 2020 International Conference on Computer Information and Big Data Applications (CIBDA), Guiyang, China, 2020, pp. 374-377, doi: 10.1109/CIBDA50819.2020.00090.
2. Okarma, Krzysztof. (2020). Applications of Computer Vision in Automation and Robotics. Applied Sciences. 10. 6783. 10.3390/app10196783.
3. Chen, Y.; He, P.; Gao, M.; Zhang, E. Automatic Feature Region Searching Algorithm for Image Registration in Printing Defect Inspection Systems. Appl. Sci. 2019, 9, 4838
4. Cutolo, F.; Fontana, U.; Cattari, N.; Ferrari, V. Off-Line Camera-Based Calibration for Optical See-Through Head-Mounted Displays. Appl. Sci. 2020, 10, 193
5. Alomar, K.; Aysel, H.I.; Cai, X. Data Augmentation in Classification and Segmentation: A Survey and New Strategies. J. Imaging 2023, 9, 46. https://doi.org/ 10.3390/jimaging9020046
6. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Adv Neural Inf Process Syst. 2012;25:1106–14.
7. https://docs.opencv.org/4.x/d5/daf/tutorial_py_histogram_equalization.html
8. https://stackoverflow.com/questions/25008458/how-to-apply-clahe-on-rgb-color-images
9. https://pyimagesearch.com/2021/01/20/opencv-flip-image-cv2-flip/
10. https://github.com/rooneysh/Labelme2YOLO
11. https://roboflow.com/convert/labelme-json-to-yolo-darknet-txt
12. EFFICIENT APPROACH OF DETECTION AND VISUALIZATION OF THE DAMAGED TABLETS RIHAB HAZIM QASIM#1, AND MUZHIR SHABAN. AL-ANI*
13. Medical Tablet Damage Detection Based on Circle Hough Transform

DOI:10.4206/aus.2019.n26.2.58

14 Single-Image Reflection Removal Using Deep Learning: A Systematic Review

Ali Amanlou; Amir Abolfazl_Suratgar; Jafar_Tavoosi; Ardashir_Mohammadzadeh; Amir Mosavi

19BIC008_Report

ORIGINALITY REPORT

16%	11%	9%	9%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.mdpi.com Internet Source	5%
2	Submitted to Institute of Technology, Nirma University Student Paper	4%
3	journalofbigdata.springeropen.com Internet Source	1%
4	Submitted to University College London Student Paper	1%
5	www.irjmets.com Internet Source	1%
6	Submitted to Liverpool John Moores University Student Paper	1%
7	Submitted to CSU, San Jose State University Student Paper	1%
8	www.researchgate.net Internet Source	1%
9	archive.org	