Development of Web-Based Semi-Automatic Image Labeling Tools

M Rahadian Alamsyah P W
School of Electrical Engineering and Informatics
Bandung Institute of Technology
Bandung, Indonesia
rahadian.apw@gmail.com

Nugraha Priya Utama, S.T, M.A., Ph.D.
School of Electrical Engineering and Informatics
Bandung Institute of Technology
Bandung, Indonesia
utama@std.stei.itb.ac.id

***Abstract*— Computer vision has been further researched to assist human workers in various fields. The current computer vision approach uses machine learning, so the development of computer vision can take advantage of labeled image data. For the model built to process the image properly, a lot of labeled data is needed. Therefore, developing a computer vision model that utilizes machine learning has a long process of collecting and labeling image data.**

Many tools have been created to assist the process of labeling image data that have a graphical interface to make it easier for users to label images. In this final project, further development of the image data labeling tool is carried out to make the image data labeling process more straightforward. By utilizing a machine learning model in computer vision that has given good results in the detection of image data, the labeling tool is further modified by providing a label recommendation system when the labeling process is carried out. The recommendation system built makes the semi-automatic image data labeling process because the user does not need to provide a label from zero so that the efficiency of data labeling has the potential to increase. Kakas is built on a web-based basis, so the data integration process between labelers becomes more accessible and faster. The built-in tool can potentially increase the efficiency of labeling image data by 62.79% compared to the manual labeling process.

Keywords— machine learning; computer vision; image labeling; object detection

I. INTRODUCTION

Computer vision is a digital technology that a part of artificial intelligent. This technology is closely related to computational knowledge and understanding of the information contained in visual images [1]. The latest approach to computer vision for visual sensing uses machine learning which requires a set of image data that has been given information in the form of labels. By utilizing this data, computing devices can learn the information in the image data and generate new knowledge for other image inputs.

Algorithms and machine learning capabilities continue to evolve. The need for data as a source of information learned by computers is also increasing to produce machine learning models with better accuracy and capabilities. In computer vision, much data is also needed to create applications to process visual images. Thus, the output produced by the application can be better.

In the process of making machine learning models, data collection is the most time-consuming stage. Developing machine learning models for computer vision also takes a long time to collect, process, and label image data. This study intends to create a tool that can help increase efficiency in collecting and labeling image data so that the process of building machine learning models for computer vision becomes faster. The tool's design utilizes an existing machine learning model to provide label recommendations for the labeling process to be semi-automatic. The labeler only needs to correct the label recommended by the tools.

II. BASIC THEORY

A. Machine Learning

Machine learning is a scientific field that understands how to build computer programs that automatically improve performance based on the experience of the program [2]. At this time, machine learning has been used in various sectors of life, such as finance, business, education, health, and so on. Since computers were created, humans have been wondering what if computers could learn things in this world as humans learn. Imagine a computer can discover how to prevent a disease based on medical data humans have recorded. There have been many computational algorithms that have currently been developed, showing the effectiveness of computers in understanding data and generating conclusions. As science develops in understanding computers, machine learning is not impossible and can become the technology that impacts human life the most.

B. Data Train

The performance of machine learning can be improved by providing large amounts of training data. Training data is data that is entered into the computer robot at the beginning of the operation [3]. Training data is an essential source of experience for machine learning algorithms because the learning algorithm looks for relationships, develops understanding, and increases the accuracy of the automated models that will be created. Training data is an initial fact which is generally in the form of input data and the expected target results. The training data consists of rows that represent an input object and columns that represent features of that object.

C. Neural Network

Neural Network is a computer program that simulates the behavior of neural networks in the human brain to learn something and make decisions [4]. An artificial neural network is a machine learning algorithm that has robust results in providing predictive outcomes for various types of data. Each neuron in the artificial neural network consists of weight for the input value, which is then processed by the activating function and produces a value.

D. Instance Segmentation

Labeling the image has the meaning of providing the information contained in the image data. This data labeling is intended to use image data as training data in a machine learning model. The information embedded in the image can be in the form of classification, spatial information, square markers, etc. In general, for this type of labeling, instance segmentation is represented using polygons that bounding the object.

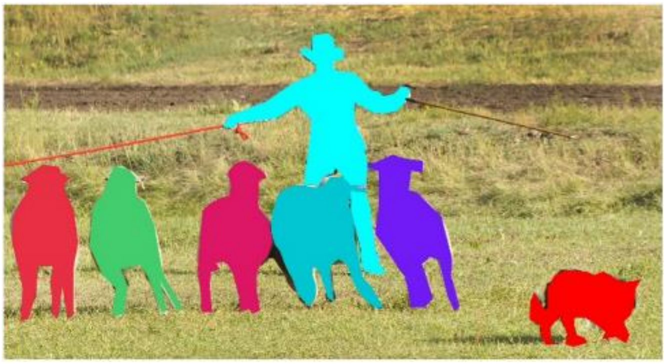


Fig. 2. Example of instance segmentation on image data.

E. Mask R-CNN

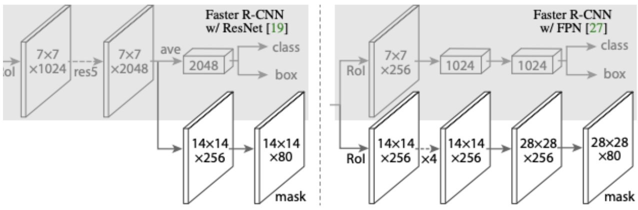


Fig. 3. Mask R-CNN architecture extend two existing Faster RCNN heads.

Mask R-CNN is a convolutional neural network model developed from Faster R-CNN [5] by adding segmentation aspects to objects that have been detected [6]. In this model, a pixel-to-pixel alignment process is added that is not contained in the Faster R-CNN to produce a binary closure for each RoI. Binary mask represents the part of the pixels in the image which is the area of the detected object. This model generates a label with type instance segmentation.

F. Ramer-Douglas-Peucker

In image processing applications, it is desirable to show the boundaries of regions in the image using polygons consisting of several representative pixels which are commonly referred to as dominant points. Through the polygon approach, points are searched to represent the curve of the shape using fewer points so that the representation of the shape is not sensitive to noise caused by image digitization and processing is easier because it uses fewer points. Ramer-Douglas-Peucker (RDP) is one of the algorithms proposed by Ramer, Douglas, and Peucker to detect the dominant point on a complex polygon to produce a simpler polygon [7]. This algorithm calculates the deviation of each point of the polygon and eliminates the points whose deviation exceeds the tolerance deviation.

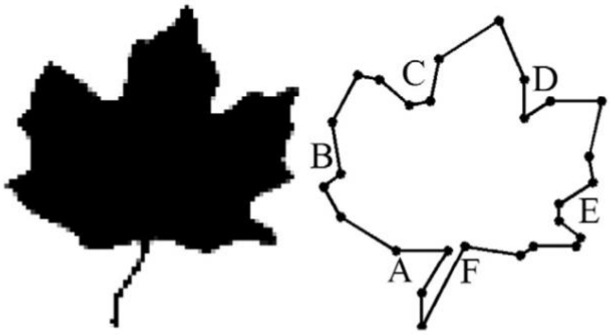


Fig. 1. Example of object shape in pixel and polygon.

III. Design Solution

To solve the problem of collecting labeled image data to build a machine learning model, we present a labeling tool that can perform automatic detection of objects to be labeled. Of course, the object given by the system from annotation tools is not necessarily correct. Therefore, we called this annotation tools semi-automatic not fully automatic. The automatic detection for object label in the image utilizes matured instance segmentation model Mask R-CNN.

The output from Mask R-CNN model is a binary mask that represent instance of object in the image. Binary mask is data type that hard to use in annotation editor instead of using polygonal structure. So, we need to post process the output of instance segmentation model to simplify data and make labeling object in the image easier. Ramer-Douglas-Peucker (RDP) is used to get polygonal structure from binary mask output. The process to get label recommendation figured in Fig. 4.

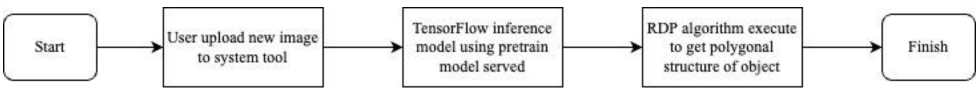


Fig. 4. Flow chart of recommendation system in annotation tools to help process of labeling image.

IV. Implementation and Evaluation

A. Implementation

In building a web-based image labeling tool we utilize several library and framework to simplify process of making the annotation tools. To handle data processing on the backend, the python-language Django framework is used so that it is easier to use TensorFlow as an inference engine in machine learning models. For database using SQLite which is easy to use and configure on backend framework. So, the backend architecture will be looks like Fig. 5.

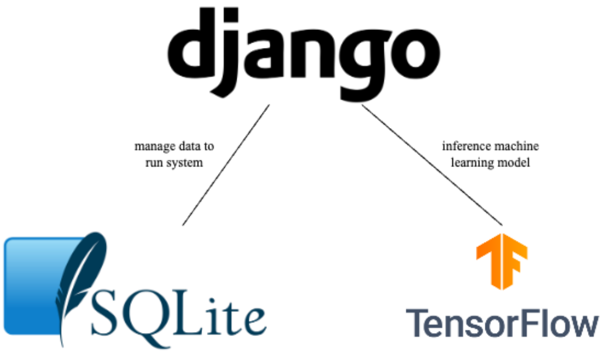


Fig. 5. Backend structure that has responsibilities in business layer of application.

For the frontend of application, we use Next.JS that a framework of React.JS which has responsibilities to serve user interface through the web. User must be using web browser to access image labeling tools across http. The frontend access data which show to user from backend with REST API. Request data using API call based on user behavior to interface and then backend will response data to frontend. Next.JS will render page based on data received from Django. So, large architecture of image labeling tools will be like Fig. 6.

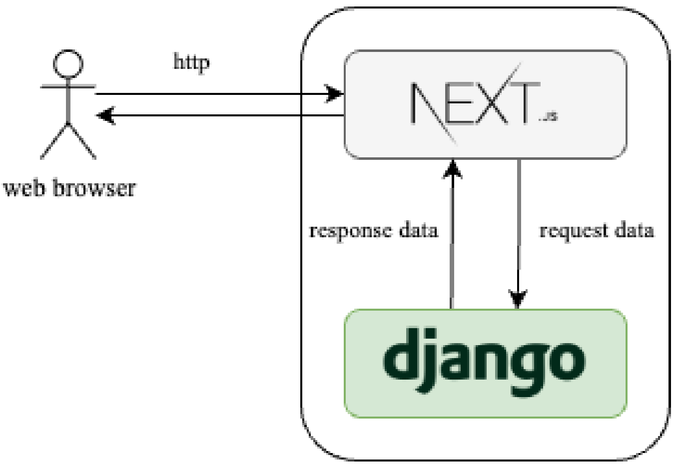
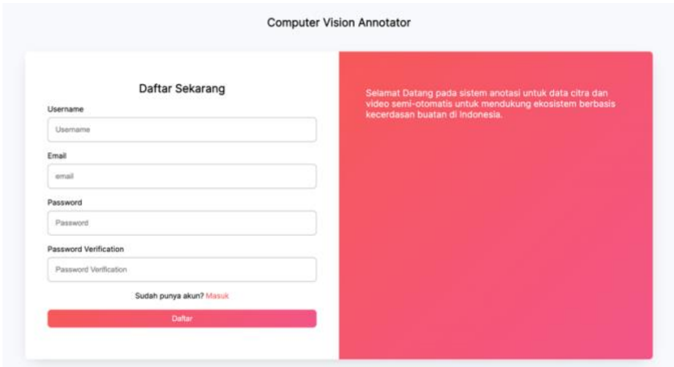


Fig. 6. Web architecture of image labeling tools.

After the implementation is done, features such as labeling tools are produced, but with additional label recommendations and user access control. The tools have several pages with they own utility. Each page will be explained after this paragraph.

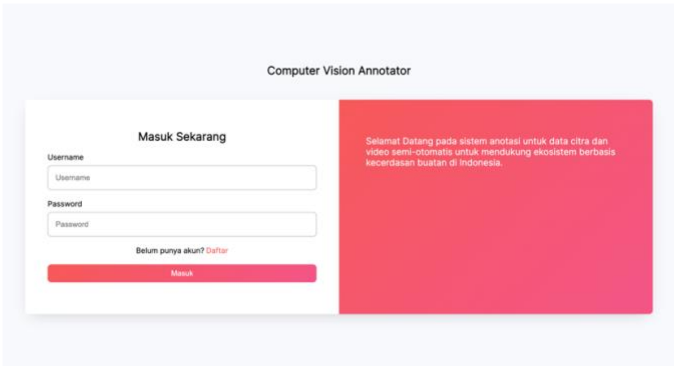
1) Register Page

Page for user register new account to access feature on image labeling tools. Just fill the fields and click button "Daftar".



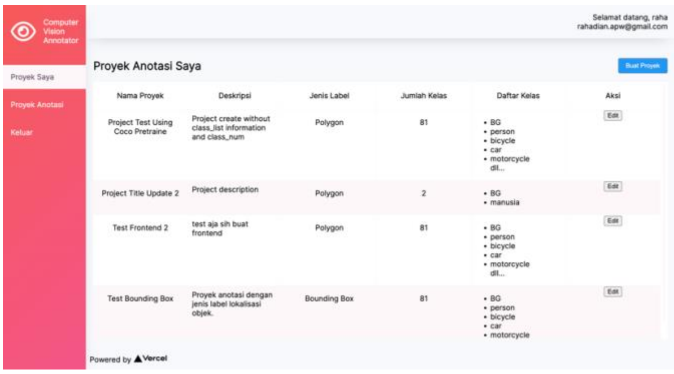
2) Login Page

Page for user authentication to the application using username and password.



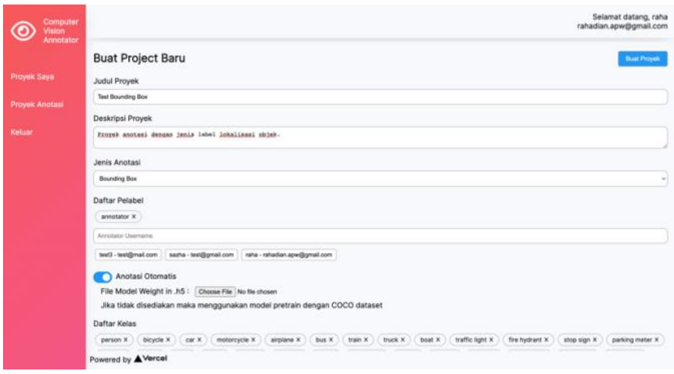
3) My Project

Page for showing all image labeling project that logged in user have.



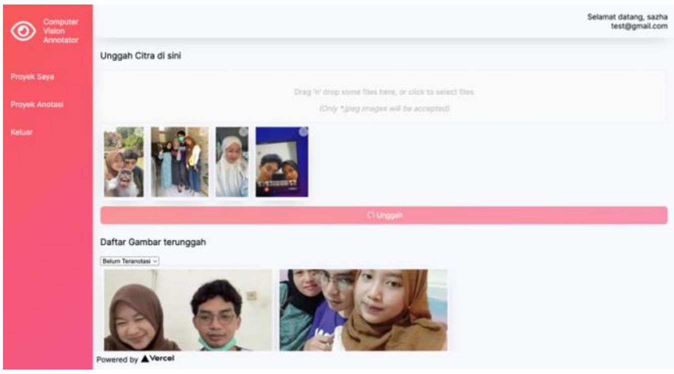
4) Create and Edit Project

Form page for create or edit image labeling project that logged in user have.



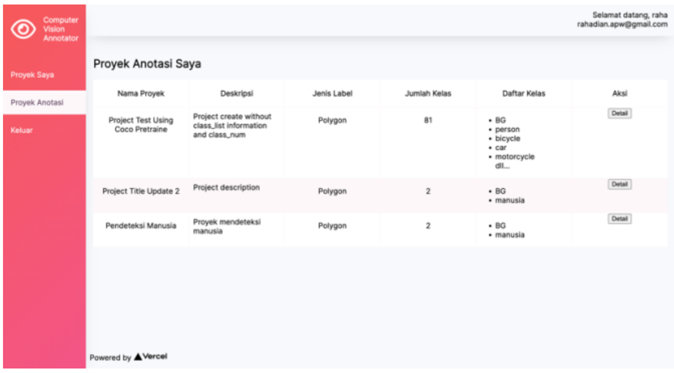
5) Project Images

Page for list all images and upload image for image labeling project that user have.



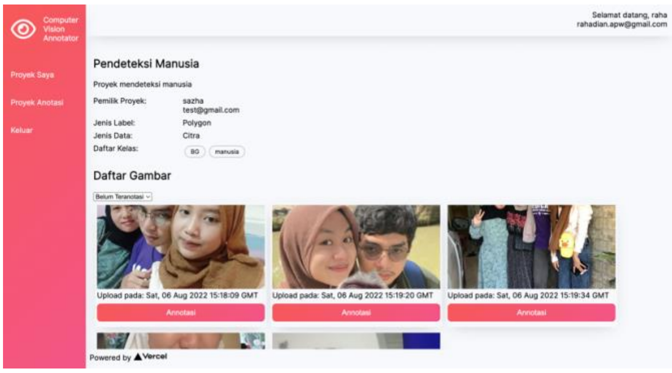
6) My Annotating Project

Page for showing project that logged in user is an annotator in image labeling project.



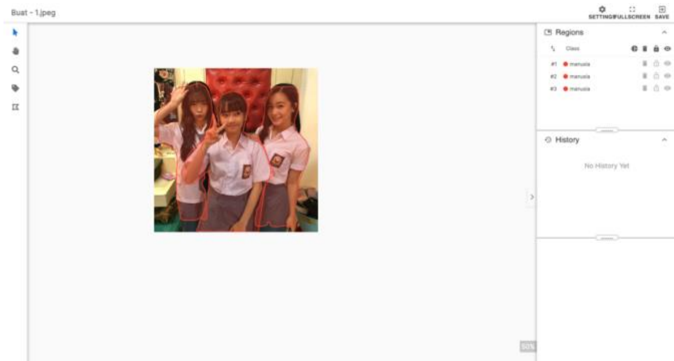
7) Detail Project

Page for annotator user get information about image labeling project and get list of images in the project.



8) Image Label Editor

Page for user create or update label in an image.



B. Evaluation

The development of a semi-automatic image data labeling tool is expected to increase the efficiency of the image data labeling process to build a machine learning model for computer vision. To prove an increase in efficiency, a measurement is needed to test it. In labeling an image data containing several objects, it means that the annotator must carry out the process of marking as many objects as there are objects in the image. With a recommendation system that utilizes Mask R-CNN, the annotator does not need to mark objects as much as the number of objects in the image data. Therefore, the efficiency of the recommendation system on the image data labeling tool can be represented by the number of objects detected divided by the number of objects that should be labeled. It is the same as the recall function, as in (1), with true positive (TP) is object mark detected by system and false negative (FN) is object must be detected by system but not detected.

$$recall = \frac{TP}{TP+FN} \times 100\% \qquad (1)$$

After determining the measurement method to test the efficiency of the semi-automatic image data labeling tool, then the experiment was carried out. The experiment was carried out using a tool that had been built by providing input of five images and then evaluated from the results given.

Fig. 9. Experiment 1 with six objects in the image. The system successfully detects five objects so that the value of TP is 5 and the value of FN is 1. Therefore, the recall value is 83.3%.

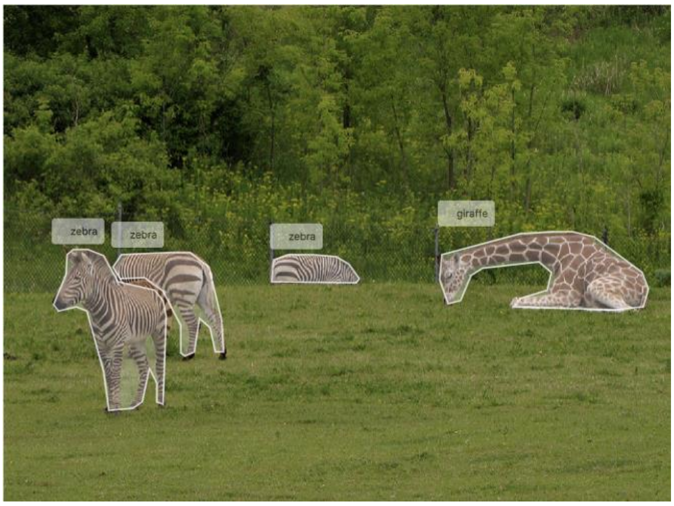


Fig. 11. Experiment 2 with four objects in the image. The system successfully detects for objects so that the value of TP is 4 and the value of FN is 0. Therefore, the recall value is 100%.

Fig. 7. Experiment 3 with ten objects in the image. The system successfully detects seven objects so that the value of TP is 7 and the value of FN is 3. Therefore, the recall value is 70%.

Fig. 8. Experiment 4 with ten objects in the image. The system successfully detects six objects so that the value of TP is 6 and the value of FN is 4. Therefore, the recall value is 60%.

Fig. 10. Experiment 5 with three objects in the image. The system successfully detects two objects so that the value of TP is 2 and the value of FN is 1. Therefore, the recall value is 66.67%.

V. CONCLUSION

After development of image labeling tools and doing experiment for evaluation, we get the recall value for five image samples. The recall value from experiments have average 62.79%. Therefore, image annotation tools that we develop potentially improve efficiency of labeling image dataset about 62.79% instead doing manually. So, the conclusion is recommendation system that attached in image annotation tool can help building image dataset faster and efficient.

ACKNOWLEDGMENT

The completion of this paper could not have been possible without the help from the Almighty God. The author thanks Dr. Nugraha Priya Utama, S.T, M.A., Ph.D. as supervisor of this whole research and project, as all of this would not be possible without his guidance. The author also thanks his family, friends, and many more people for always supporting him. The author also thanks Institut Teknologi Bandung for giving him the opportunity to learn and grow much more.

REFERENCES

[1] N. Ahuja, "Computer vision." McGraw-Hill Education, Aug. 01, 2021. doi: 10.1036/1097-8542.154050 OP - AccessScience.

[2] [2] T. M. Mitchell, *Machine learning*, vol. 45. 1997. [Online]. Available: https://books.google.ca/books?id=EoYBngEACAAJ&dq=mitchell+machine+learning+1997&hl=en&sa=X&ved=0ahUKEwiomdqfj8TkAhWGslkKHRCbAtoQ6AEIKjAA

[3] [3] McGraw-Hill Education, *McGraw-Hill Dictionary of Scientific and Technical Terms*, 6th ed. New York, 2003.

[4] [4] D. A. Gustafson, "Neural network." McGraw-Hill Education, Aug. 03, 2018. doi: 10.1036/1097-8542.449750 OP - AccessScience.

[5] [5] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Trans Pattern Anal Mach Intell*, vol. 39, no. 6, pp. 1137–1149, 2017, doi: 10.1109/TPAMI.2016.2577031.

[6] [6] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," *IEEE Trans Pattern Anal Mach Intell*, vol. 42, no. 2, pp. 386–397, 2020, doi: 10.1109/TPAMI.2018.2844175.

[7] [7] D. K. Prasad, M. K. H. Leung, C. Quek, and S. Y. Cho, "A novel framework for making dominant point detection methods non-parametric," *Image Vis Comput*, vol. 30, no. 11, pp. 843–859, 2012, doi: 10.1016/j.imavis.2012.06.010.