CrowdNotate: An Approach to Crowdsourcing the Annotation of Image Data in an Iterative Pipeline

Joel Ilao*, Vincent Ilagan, Rayvhen Mico Rosima and Ralph Matthew Sanson
College of Computer Studies, De La Salle University
**Corresponding Author: joel.ilao@dlsu.edu.ph*

Abstract: Obtaining an appropriate dataset is crucial for computer vision tasks, but it can be challenging due to labor or financial constraints, especially when specialized labels are needed. This paper explores using crowdsourcing to gather datasets for machine learning and data science applications. The researchers developed a web-based system, CrowdNotate, which offers a user-friendly environment for easily repeating activities to generate crucial data for creating suitable datasets. Usability and user acceptance testing using the MEEGA+ Framework indicate that users find the web application adequate for image annotation tasks, though its interface aesthetics can be further improved. Overall, the system benefits both end-users and researchers in terms of ease of use and reliability.
Key Words: computer vision; image annotation; crowdsourcing; image datasets

1. INTRODUCTION

Obtaining large-scale, high-quality annotated image datasets is essential for many applications, including data integration and machine learning (ML). For instance, deep learning (DL) has significantly advanced machine learning and now performs at the cutting edge in a wide range of tasks, including image identification, natural language processing, and Generative Adversarial Networks (GANs) (Kramberger & Potočnik, 2020); however, in order to obtain greater performance, the majority of deep learning approaches necessitate large training sets, which typically entail high annotation costs or labor.

The quality of labeled training data is crucial to the success of supervised machine learning algorithms in the field of computer vision. The main method for obtaining labeled training data is through manual human annotation; this method of obtaining the labels can be time-consuming and expensive, and can often prove difficult for those needing huge amounts of data. Additionally, the labels must be approximated based on the arbitrary judgment of a small group of experts, who frequently have different opinions on the labels.

There have been many studies that tackled the annotation of images of all kinds. An enormous number of studies have focused on annotating image data that spans three decades, which began with the manual extraction of features followed by some classifiers to annotate images, too increasingly complex deep learning methodologies for both the annotation and retrieval process (Bhagat & Choudhary, 2018). Additionally, within these decades, various tools have also emerged, such as LabelMe (Russel et al., 2007), ByLabel (Qin et al., 2018), VIA Annotation Software (Dutta & Zisserman, 2019), and the object- and event-level video annotation tool for anomaly detection (Go, 2022), to name a few. Examples of efforts in the field include studies by (Jiu & Sahbi, 2017; Little et al., 2010; Moradi et al., 2016; Yakovlev & Lisovychenko, 2020).

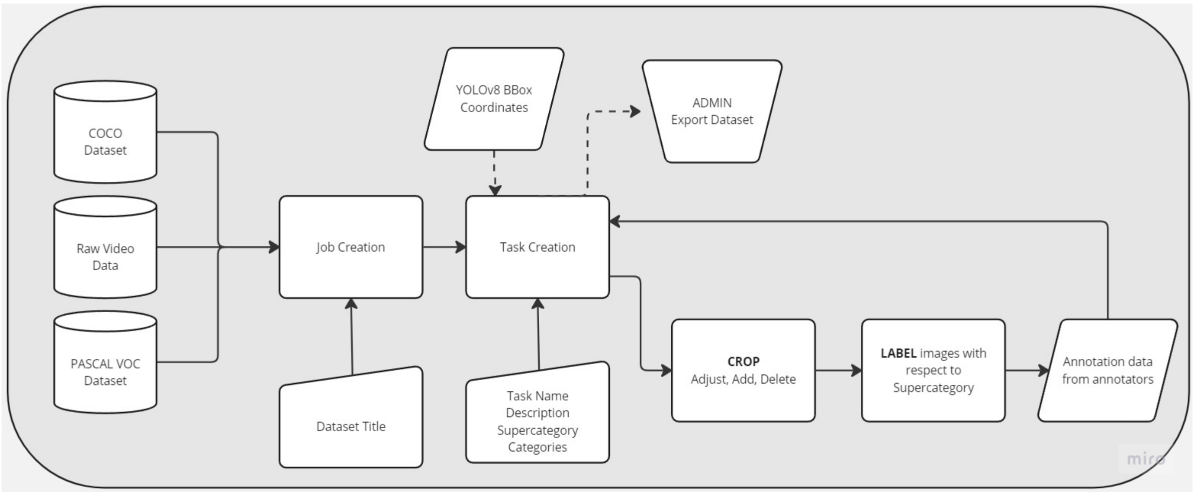

Figure 1. Process Flowchart

Computer Vision projects with specific application scenarios, such as vehicle counting, tend to require only a subset of features and functionalities of general-purpose annotation tools. Furthermore, the complexity of these tools could overwhelm recruited human annotators who do not have sufficient ICT competency skills to navigate the software interface, even for simple tasks. This would require extended human annotators' orientation, and further Quality Assurance checking by the project team, thereby resulting in longer project durations and potentially incurring corresponding additional expenses. Finally, the aforementioned annotation tools need minimum device specifications for installations, such as PC workstations and laptops. Addressing these issues requires the development of dedicated software, since existing annotation tools are either closed or difficult to modify. Indeed, custom-made annotation tools for specific purposes have been developed, such as a video splitter and image annotator by (Cempron & Ilao, 2020) for a vehicle classification task, and Go (2022)'s video annotation tool used to train traffic objects and incident detectors. These tools, however, have limited accessibility, thereby hampering annotation productivity due to limited annotators' reach, thus resulting in fewer labeled datasets. Providing insufficient amounts of datasets to any deep learning algorithm will inhibit it from learning more features, preventing its ability to predict various real-world test data as it fails to generalize when using a test set.

This paper proposes CrowdNotate, a web-based application for crowdsourcing image annotations from regular, i.e. non-expert, users. A single image could potentially have different labels and attributes depending on the type of dataset a specific researcher needs. This highlights the need to guide these expectations to control the potential label description of an image. For example, two labels such as "ford focus" and "parked outside" are valid descriptions for the same vehicle category but present a challenge of assigning multiple attributes for labels. The application provides instructions that will reduce the risk of ambiguous label categorization. The main flow of CrowdNotate would follow the CROP activity followed by the LABEL activity. In order to source the unlabeled datasets, a YOLOv8 object detection and classification model was used.

2. METHODOLOGY

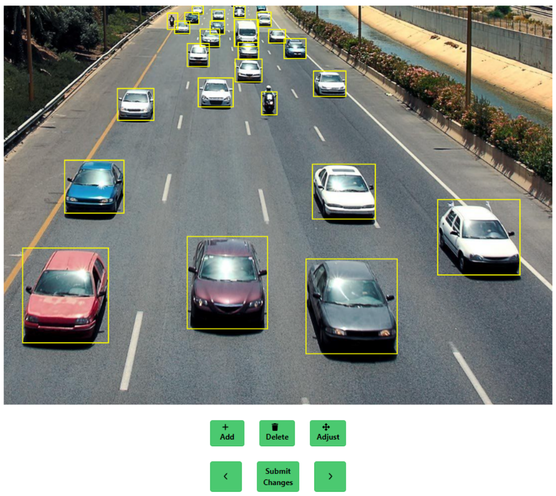

Figure 2. The image 'Crop' Interface accessible via the Admin account. Initial object instances can be detected using a built-in YOLOv8 detector, which can be further refined prior to sending to Annotators.

CrowdNotate is a web application that is accessible to both PC and mobile platforms; upon visiting the web app, users will be required to log in or create an account. Please refer to Figure 1 for the Process Flowchart. Two types of accounts exist within CrowdNotate: Admin and Annotator. Annotator accounts can provide labels to existing images. Admin accounts have access to more features, such as the CROP activity, meant to govern how the data is presented to annotators and extract meaningful data from the interactions of the annotator accounts.

2.1 Semi-Automatic Labeling

In order to aid researchers in creating annotations at a faster rate, CrowdNotate employs the help of YOLOv8 by ultralytics (Jocher, Chaurasia & Qiu, 2023). YOLOv8 is an object detector that was trained on the Microsoft COCO Dataset (Lin et al., 2014), which contains 80 object classes. The object detector provides the initial labels, which the human annotators can confirm eventually through the interactive user interface.

2.2 CROP Activity

The CROP activity (refer to Figure 2) begins with data from the bounding box coordinates from an incomplete dataset and/or coordinates from YOLOv8. The coordinates are superimposed on the presented frame, and the admin has 4 possible actions for each existing image: Assert, Adjust, Add, or Delete, which are used to remedy the mistakes that may have been done by the YOLOv8 detector. After all adjustments have been performed, these are saved to the database. Only the Admin account can access this functionality, since only this account can select images from an image collection, before they are fielded to Annotator accounts.

2.2 LABEL Activity

Refer to the corresponding flowchart in Figure 3 and the interface available for Annotator accounts, shown in Figure 4. The admin provides the context for the expected types of labels in the form of Supercategories and Categories, as well as bounding box data from the CROP Activity. From that point, annotators are able to choose a Label with respect to the category and is stored to the database as annotations from annotator with the chosen category and confidence in the chosen category. For every submission, there is a backend calculation of the Fleiss Kappa and Confidence Levels of the annotation; after every calculation, the annotations' category is updated in the database.

3. RESULTS AND ANALYSIS

To measure users' perception on the functionality and aesthetic of the developed tool, a questionnaire which contains several modified questions from the

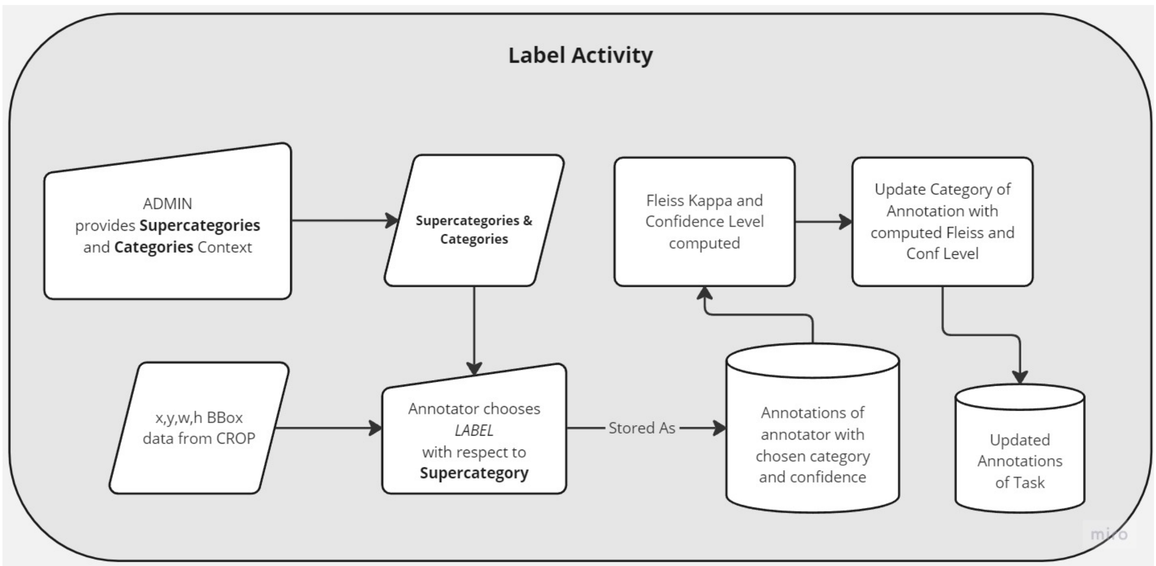

Figure 3. LABEL Activity Flowchart

MEEGA+ measurement instrument[1]. The experiments were conducted with 30 users in closed testing, with 21 respondents in the survey form. Further experiments were tested with 15 users from the Philippine High School System intern students (PSHSS) and 16 users from the closed beta testing that is used to survey CVAT, a comparable annotation tool to CrowdNotate.

Based on the initial user acceptance testing results, the majority of users have expressed a positive outlook on the web application. While there may be room for improvement in terms of aesthetics, the feedback indicates that the web application excels in other quality factors, particularly usability and player experience. The users agreed that the web application can be used and shared among peers due to its straightforward approach, and enhanced productivity achieved by keeping the application as simple as possible.

The quality score indicates the quality of the application based on the MEEGA+ scale (note: higher values are better).

Table 1. Quality Level

Survey	Quality Level Score
Closed Beta Testing	58
PSHSS	72
CVAT	71

4. CONCLUSIONS

In addition to creating an engaging and iterative tool, the researchers have accomplished the development of a system that is inherently easy to use for both the end-users and the researchers who rely on its services. Recognizing the importance of user experience, they aim to design intuitive interfaces, streamline workflows, and provide comprehensive documentation and support to ensure a smooth and user-friendly

[1] http://www.gqs.ufsc.br/files/2020/02/MEEGA-Scale-instructions.pdf

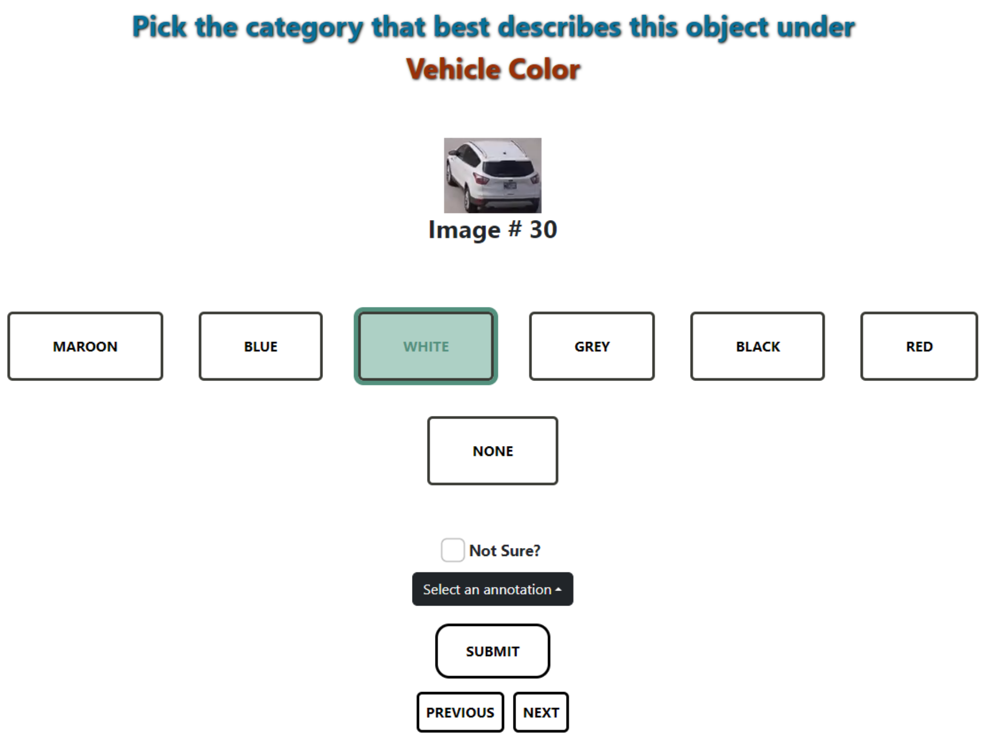

Figure 4. The Image LABEL Activity accessible via the Annotator account. Pre-defined object categories are set by the Administrator account. Annotators can check the 'Note Sure' checkbox if they are unsure of their choice.

experience. This focus on usability will enable users and researchers to interact with the system effortlessly, maximizing productivity and minimizing learning curves.

Furthermore, the findings of this study will benefit the sectors of research in the field of computer science specializing in computer vision and areas of human-computer interaction. As data annotation is treated as an increasingly specific part of machine learning as a whole, the quality of data collected should be regarded as a top priority. Multiple data were collected in different fields, such as environmental (Russo et al., 2021), audio interaction (Hantke et al., 2015), and vehicles (J. Sang, 2018). The use of these dedicated tools can be used to further improve their feasibility when it comes to different aspects of computer vision, as it also introduces an element of interaction that can benefit researchers looking to yield a dataset from their raw data. Furthermore, future researchers can access other models that are not YOLO related, like Fast Region-based Convolutional Neural Networks (Fast R-CNN) or Single Shot Detectors (SSD). Lastly, other datasets that are not COCO dataset-related should be tested in order to test formatting between the future researcher's preferred object detector format and COCO format when inputting them in a training dataset.

On suggestions provided by the users, the UI and UX of the web application can be further improved by having a clear and predictable design. Visual cues can be appropriately distinguished through color, and users can quickly identify critical actions or information without relying solely on text. Additionally, an improvement on the landing page can attract more users to the website by giving better visual clarity and a concise explanation of purpose and goals.

6. REFERENCES

PK Bhagat and Prakash Choudhary. 2018. Image annotation: Then and now. Image and Vision Computing 80 (2018), 1–23.

Jonathan Cempron and Joel Ilao. 2020. Development of annotation tools for building labeled datasets for a vision-based traffic monitoring system. In De La Salle University Research Congress, Manila, Philippines.

Abhishek Dutta and Andrew Zisserman. 2019. The VIA annotation software for images, audio and video. In Proceedings of the 27th ACM international conference on multimedia. 2276–2279.

Jeffrey C Go. 2022. The object and event-level video annotation tool for anomaly detection. (2022).

Simone Hantke, Florian Eyben, Tobias Appel, and Björn Schuller. 2015. iHEARu-PLAY: Introducing a game for crowdsourced data collection for affective computing. In 2015 International Conference on Affective Computing and Intelligent Interaction (ACII). IEEE, 891–897.

Mingyuan Jiu and Hichem Sahbi. 2017. Nonlinear deep kernel learning for image annotation. IEEE Transactions on Image Processing 26, 4 (2017), 1820–1832.

Glenn Jocher, Ayush Chaurasia, and Jing Qiu. 2023. YOLO by Ultralytics. https://github.com/ultralytics/ultralytics

Tin Kramberger and Božidar Potočnik. 2020. LSUN-stanford car dataset: Enhancing large-scale car image datasets using deep learning for usage in GAN training. Applied Sciences 10, 14 (2020), 4913.

Tsung-Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, and Piotr Dollár. 2014. Microsoft COCO: Common Objects in Context. CoRR abs/1405.0312 (2014).

Greg Little, Lydia B Chilton, Max Goldman, and Robert C Miller. 2010. Turkit: human computation algorithms on mechanical turk. In Proceedings of the 23nd annual ACM symposium on User interface software and technology. 57–66.

Mehdi Moradi, Yufan Guo, Yaniv Gur, Mohammadreza Negahdar, and Tanveer Syeda-Mahmood. 2016. A cross-modality neural network transform for semi-automatic medical image annotation. In International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 300–307.

Xuebin Qin, Shida He, Zichen Zhang, Masood Dehghan, and Martin Jagersand. 2018. Bylabel: A boundary based semi-automatic image annotation tool. In 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 1804–1813.

B. C. Russel, A. Torralba, K. P. Murphy, and W. T. Freeman. 2007. LabelMe: A Database and Web-Based Tool for Image Annotation. International Journal of Computer Vision 77, 1–3 (2007), 156–173.

Stefania Russo, Michael D. Besmer, Frank Blumensaat, Damien Bouffard, Andy Disch, Frederik Hammes, Angelika Hess, Moritz Lürig, Blake Matthews, Camille Minaudo, Eberhard Morgenroth, Viet Tran-Khac, and Kris Villez. 2021. The value of human data annotation for machine learning based anomaly detection in environmental systems. Water Research 206 (2021), 117695. https://doi.org/10.1016/j.watres.2021.117695

Jun Sang, Zhongyuan Wu, Pei Guo, Haibo Hu, Hong Xiang, Qian Zhang, and Bin Cai. 2018. An Improved YOLOv2 for Vehicle Detection. Sensors 18, 12 (2018). https://doi.org/10.3390/s18124272

A Yakovlev and O Lisovychenko. 2020. An approach for image annotation automatization for artificial intelligence models learning. 1, 36 (2020), 32–40.