

3D BAT: A Semi-Automatic, Web-based 3D Annotation Toolbox for Full-Surround, Multi-Modal Data Streams

Walter Zimmer, Akshay Rangesh and Mohan Trivedi

Abstract—In this paper, we focus on obtaining 2D and 3D labels, as well as track IDs for objects on the road with the help of a novel 3D Bounding Box Annotation Toolbox (3D BAT). Our open source, web-based 3D BAT incorporates several smart features to improve usability and efficiency. For instance, this annotation toolbox supports semi-automatic labeling of tracks using interpolation, which is vital for downstream tasks like tracking, motion planning and motion prediction. Moreover, annotations for all camera images are automatically obtained by projecting annotations from 3D space into the image domain. In addition to the raw image and point cloud feeds, a Masterview consisting of the top view (bird’s-eye-view), side view and front views is made available to observe objects of interest from different perspectives. Comparisons of our method with other publicly available annotation tools reveal that 3D annotations can be obtained faster and more efficiently by using our toolbox.

I. INTRODUCTION

There is a growing need for broadly available 3D annotation platforms that have a labeling user interface that is intuitive enough for non-domain experts to use. Many existing tools can annotate 2D images, but there is a growing need to also annotate 3D LiDAR data.

Navigating/labeling in 3D space requires a carefully designed user interface. Annotating large amounts of 3D data also requires a big workforce that may need training. Another challenge is that the resolution and the clarity may be limited, making it hard to differentiate between objects. In this paper, we propose an open source tool that alleviates the need to buy end-to-end “Training Data as a Service” solutions. Annotators can label their own data very accurately and have full control of the toolbox and their data. Our toolbox also provides functionality to evaluate annotated data (built-in quality control) in order to obtain labels with high quality. Finally, human annotation tasks are very labor-intensive and sometimes require additional human resources. As a remedy, one can even consider using our toolbox to outsource the annotation tasks to Amazon Mechanical Turks (AMTs) - a popular crowd-sourcing platform. This is possible because of the web-based, platform-independent nature of our toolbox, capable of running on most modern browsers, irrespective of the operating system involved.

Recently, applications related to autonomous and partially automated cars have attracted significant attention. For research in these applications, a full-surround multi-modal dataset with 2D and 3D annotations, as well as assigned track IDs would be very useful [1]–[3]. Motivated by these

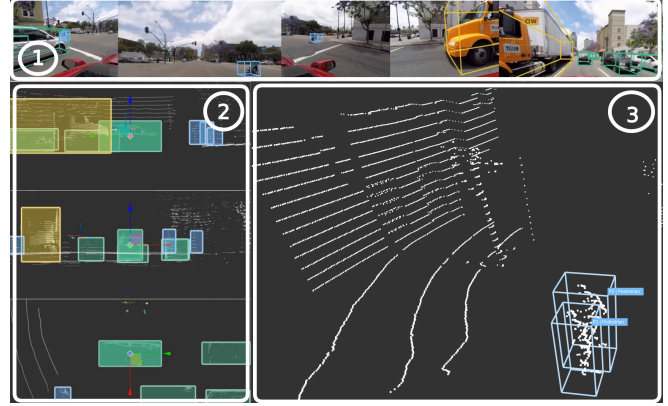


Fig. 1: Overview of the annotation toolbox. (1) A horizontal scrollable and vertical resizable panoramic camera image provides a full-surround view. (2) The Masterview that consists of a side view (top), a front view (middle) and a top view (bottom) supports the user during the annotation process. (3) 3D view in that the user can navigate and place annotations.

needs, our work focuses on the challenging task of 2D image and 3D point cloud annotation of street scenes. Inspired by the easy usage of 3D modeling tools (e.g. Blender [4]), we make the user annotate scenes directly in 3D and transfer those annotations back into the image domain. Laser data is first annotated with rough bounding primitives, and then a geometric model is used to transfer these labels into the image space. The advantage of this approach is that the label in 3D is automatically projected into multiple camera views, thus lowering the annotation time considerably. We also include additional features like smart interpolation for users to semi-automatically label sequential data streams. All these features put together make our tool easy to use, accurate and efficient.

The main contributions of this work are as follows: First, we propose a novel annotation system with tools for efficient, accurate 3D localization and tracking of objects using full-surround multi-modal data streams (see Fig. 1). Second, a systematic comparison with 33 annotation tools is carried out to highlight improvements and benefits of our approach (see Table I). Finally, we conduct both qualitative and quantitative evaluations of the efficiency and accuracy of our system using four human annotators with diverse background and skills, resulting in more than 13,800 object annotations within one hour of usage. The annotation toolbox is available for public use at: <https://github.com/walzimmer/3d-bat>.

The authors are with the Laboratory for Intelligent and Safe Automobiles, University of California, San Diego, CA 92092, USA.
{wzimmer, arangesh, mtrivedi}@ucsd.edu

TABLE I: Comparison between different annotation tools.
 ○ Integrates feature ○ Feature only available for 2D images ○ Feature not available

Feature		F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	Score
open source	3D BAT (OUR)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	14/14
	TUBS [5]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	7/14
	VAST [6]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	6.0/14
	CVAT [7]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	5.5/14
	LabelMe [8]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	5.5/14
	CLEAN [9]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	5.5/14
	BRAT [10]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	3.5/14
	WebAnno [11]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	3.5/14
	Image Tagger [12]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	3.5/14
	3D BAT (OUR)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	14/14
commercial	scale.ai [13]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	10/14
	playment.io [14]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	9/14
	surfing.ai [15]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	9/14
	dataturks.com [16]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	5.5/14
	supervise.ly [17]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	4.5/14
	labelbox.com [18]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	4.5/14
	neurula.com [19]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	4.5/14
	prodi.gy [20]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	4.5/14
	3D BAT (OUR)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	14/14
	scale.ai [13]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	10/14

- | | | | |
|----|---|-----|---|
| F1 | Full-surround annotations | F8 | 3D transform controls |
| F2 | Semi-automatic labeling | F9 | 2D and 3D annotations |
| F3 | 3D to 2D label transfer (projections) | F10 | Web-based (online accessible & platform ind.) |
| F4 | Automatic tracking | F11 | Redo/undo functionality |
| F5 | Masterview (side, front, top and 3D view) | F12 | Keyboard only annotation mode |
| F6 | Navigation in 3D | F13 | Auto save function |
| F7 | Auto ground detection | F14 | Review annotations |

II. RELATED STUDIES

In this section, we provide a comprehensive comparison between a variety of popular annotation tools. Although 33 annotation tools were evaluated in total, only 16 of the highest scoring ones are shown in Table I for the sake of brevity. More than half of the evaluated tools are open source. The comparison is split into two tables to separate the open source tools from the commercial tools.

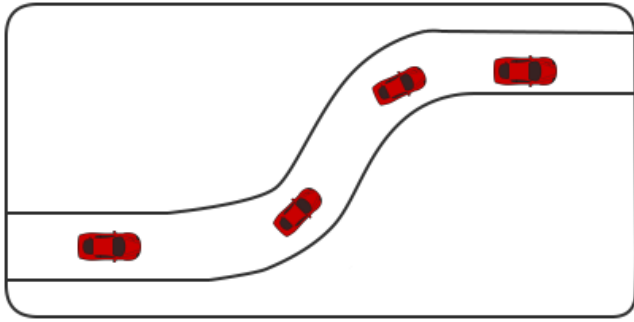
Looking at the open source tools, only the TUBS [5] annotation tool comes close to our toolbox. It provides full-surround annotations, semi-automatic labeling and annotations are projected to all cameras. Besides an automatic tracking functionality, the TUBS labeling tool has also possibility to detect and remove the ground to better locate road users on the ground. Our tool was also compared to commercial tools and achieves the highest ranking, given the features that are listed in Table I. Among the commercial options, three tools come close to our annotation tool: `scale.ai` [13], `playment.io` [14] and `surftech.ai` [15]. All of them provide automatic or semi-automatic 3D annotation techniques, a 3D to 2D label transfer as well as the possibility to navigate in 3D space. Most of the annotation tools listed in the table are web-based, offer a redo/undo functionality and automatically save changes that are done by the user. However, none of the tools support keyboard only annotations, which is a very efficient way to create labels. Full-surround annotations are only offered by `scale.ai`, `surfing.ai` and the TUBS annotation tool, whereas only `playment.io` and the TUBS annotation tool support the annotation of tracks. Compared to annotating

individual images, video sequences offer the advantage of temporal coherence between adjacent frames, thereby making it possible to use propagation techniques to transfer labels from the current frame to subsequent frames. Note that two open source tools (TUBS and VAST) and only three commercial tools (`scale.ai`, `playment.io` and `surftech.ai`) support the annotation of 3D objects. All other listed tools provide only the functionality to label in image space, hence only half a point (yellow checkmark) was added to the final score.

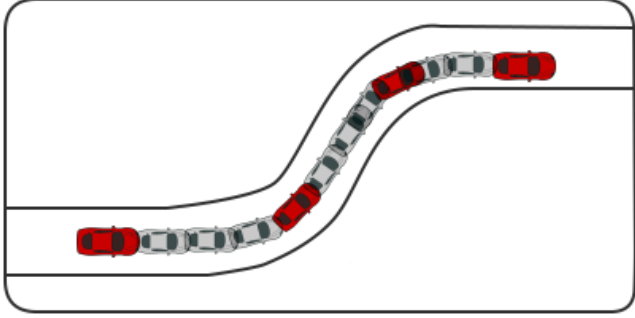
To evaluate our tool in realistic use cases, we consider different publicly available real world datasets. Compared to the popular KITTI dataset which only has camera images that were obtained by two stereo cameras pointing forward, the LISA-T [21] test vehicle has six full-surround cameras, thus more data has to be annotated per timestamp. We also consider the newly released NuScenes dataset [22] comprising of a similar full-surround sensor configuration. Since this dataset consists of vast amounts of labelled sequences, we use it to carry out quantitative evaluations of our tool.

III. SYSTEM ARCHITECTURE

In this section, the proposed methodology as well as the implemented features will be described in detail. The goal was to develop a limited and simple interface, since this leads to superior annotation experience. The 3D annotation toolbox is based on WebGL (see Fig. 1) to allow collaborative annotating. The toolbox was designed to annotate one object at a time because it is more efficient and strongly preferred by the workers.



(a) Creating control points for interpolation.



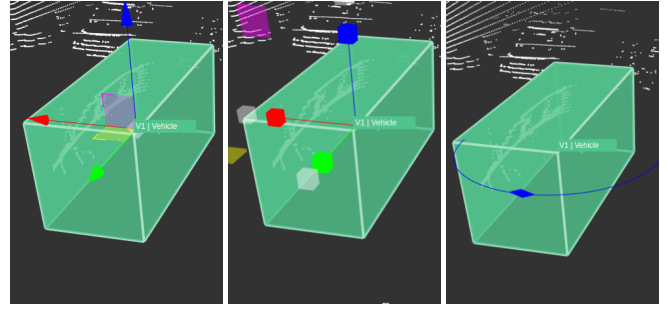
(b) Interpolation of objects. The track of the car is displayed transparent (gray). The interpolation takes also rotation and scaling into account.

Fig. 2: Illustration of object interpolation within a curve.

First, all available camera images are displayed at once (full-surround view, see Fig. 1) so that objects that are covered by multiple cameras can always be seen by the user. Without this feature, the user would have to switch between all camera images to find an object, leading to inefficiency.

Next, an optional semi-automatic labeling method is used to annotate all frames between two specific frames (start and end frame) that are determined by the annotator. After the implementation of this semi-automatic interpolation technique (see Fig. 2a and 2b), the annotation time was seen to drastically decrease. This interpolation is based on a linear model and is to be used primarily for labelling small clips of the data sequence.

Additionally, we use 3D and projective geometry to create a 3D to 2D label transfer option for the user. A 3D to 2D label transfer is very useful to obtain automatically annotated camera images from already labelled 3D point clouds. The annotator places a 3D label in the point cloud which is then projected into all six camera images. This method reduces annotation time and helps the user to place the annotation very accurately into the point cloud, since the projection is updated in real-time. The projection of 3D annotations into the camera images (3D to 2D label transfer) does not require the user to label each camera image which increases the speed by many orders. An automatic track assignment of objects simplifies the annotation further. Those tracks can later be used for motion planning and motion prediction. Adding the *Masterview* (see Fig. 1) which consists of the side, front and bird's-eye-view, the user doesn't have to



(a) Translation controls. (b) Scale Controls. (c) Rotation Controls.

Fig. 3: The user can switch between three transformation modes for which short cuts are provided. Transformation controls can be hidden as well as changed in size. They allow to translate/scale an object along two axis simultaneously.

change into the 3D view anymore to adjust the dimensions of an object.

Once the user selects an object in the scene, transform controls appear on next to it (see Fig. 3a, 3b and 3c). These controls allow the user to translate, scale and rotate the selected object. The user can interactively correct wrong positions and orientations and change between those three modes using keyboard shortcuts. This transformation can also be done using the keyboard. Shortcuts are provided to switch between the translation, scaling and rotation mode.

Furthermore, graphical elements such as buttons and checkboxes were kept at a minimum and keyboard shortcuts are provided for every operation. Having a limited number of choices is shown to save time and remove confusion, as it diminishes the cognitive load of the workers. Fig. 4 shows two example frames that were labeled from the LISA-T dataset.

IV. EXPERIMENTAL EVALUATION

The most common strategy to evaluate annotated data is to compare the obtained annotations with ground truth, this is, annotations from experts. After all features were implemented an experiment was performed in that several user had the task to annotate a sequence of 10 seconds which consists of 300 key frames and another very crowded sequence of 35 key frames. Annotators were first instructed to watch the tutorial videos as well as the raw video to get familiar with the scene and to see where interpolation makes sense. The annotation results were compared against the ground truth that was created by an experienced user (User1) who was familiar with the annotation toolbox. Table II shows how much time, mouse clicks and keystrokes the annotation user needed to annotate a sequence. Furthermore the total number of annotated objects is listed as well as the number of objects for each class.

Annotating 1800 camera images and 300 laser scans requires on average about one hour. This results in annotating 30 camera images and 5 laser scans per second. To improve the speed of the annotation process one can evaluate



Fig. 4: Example annotations of two sequences from the LISA-T dataset.

TABLE II: Annotation statistics of different users on the LISA-T dataset.

	Ground truth	User1	User2	User3
Time (hh:mm):	02:00	01:02	01:00	00:52
Clicks:	3324	1388	984	780
Keystrokes:	1032	112	255	1663
Annotated obj.:	6851	1722	3078	2212
CAR:	5051	1112	2329	1481
PEDESTRIAN:	1500	610	749	431
MOTORCYCLE:	0	0	0	0
BICYCLE:	0	0	0	0
TRUCK:	300	0	0	300
3D-IoU:	-	0.073	0.206	0.135
3D-IoU>0.6:	-	0.058	0.091	0.049
Precision:	-	0.012	0.013	0.010
Recall:	-	0.012	0.012	0.010
F_1 score:	-	0.012	0.091	0.010

heatmaps of both, mouse and keyboard. Fig. 5a shows the average heatmap of user interactions with the mouse and Fig. 5b visualizes the heatmap of the average keystrokes the annotation users performed during their task.

Evaluating the mouse heatmap one can see that annotators preferred to use the button to navigate to the next frame instead of using a keyboard shortcut which would be faster. A solution would be to provide more instructions to the user a also provide more time to get familiar with the toolbox. The majority of the annotators translated and rescaled the 3D bounding box after they positioned it into the pointcloud. Hence the keyboard key 'T' (to switch into translation mode) and 'S' (to switch into scaling mode) was pressed often. The average intersection over union in 3D (3D-IoU) between the manually created ground truth and the user annotations is 0.138 and 0.066 (considering a boundary of 0.6).

Fig. 6 shows evaluation metrics (3D-IoU, precision, recall and F_1 -score) for the annotated LISA-T dataset. We provide the mathematical descriptions of each of these metrics below:

$$precision = \frac{TP}{TP + FP}, \quad (1)$$

$$recall = \frac{TP}{TP + FN}, \quad (2)$$

$$F_1 - score = \frac{1}{\frac{1}{recall} + (\frac{1}{precision})} \cdot \quad (3)$$



(a) Mouse interactions with the annotation tool. Red areas visualize a high number of clicks.



(b) A keyboard heatmap visualizes the keystrokes of all annotators.

Fig. 5: Heatmaps of keyboard and mouse were created using the toolbox WhatPulse [23].

Precision and Recall are standard metrics that evaluate the overall performance of statistical models, and the F_1 -score calculates the harmonic mean of the precision and recall (harmonic mean because the precision and recall are ratios) [24].

After the experiment, all participants were asked to fill out a feedback form to evaluate the performance and user experience with the annotation toolbox. Table III shows the average results. From these results, we gather that the annotators described the toolbox as efficient, helpful, intelligent and easy to learn. The average score is 4.12 out of 5.00 which is 82.40%.

Evaluation was also performed on the NuScenes dataset since groundtruth annotations were already provided. In the second experiment again user statistics were created (see Table IV). The average 3D-IoU between the manually created ground truth and the user annotations is 0.0748 and 0.0393 (considering the boundary of 0.6). Fig. 7 shows

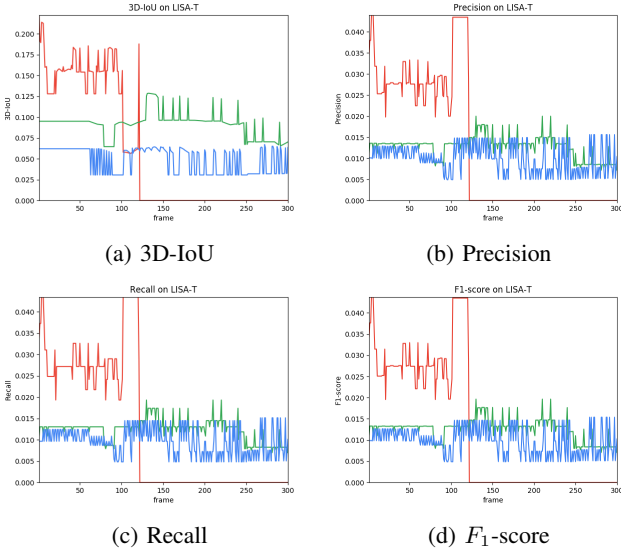


Fig. 6: Evaluation metrics on the LISA-T dataset. *User1* annotated the objects very accurately in the beginning.

● User1 ● User2 ● User3

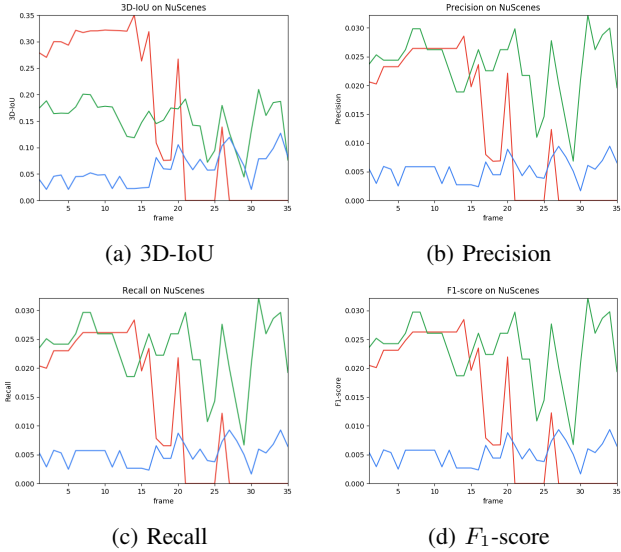


Fig. 7: Evaluation metrics on the NuScenes dataset. *User2* achieved the best accuracy in total.

● User1 ● User2 ● User3

evaluation metrics (3D-IoU, precision, recall and F_1 -score) for each annotated frame on the first NuScenes sequence.

Next, we focus on the qualitative evaluation of our annotation tool. Fig. 8 shows a comparison between manually annotated objects (top) and annotations provided by NuScenes (bottom). One can see that NuScenes provides a lot of annotations. Even objects that can't be seen in the camera image were annotated. The reason for that is that objects containing at least one point of the point clouds were annotated [22].

TABLE III: Evaluation of the user experience. All annotators rated the *Masterview* with the highest score because it simplified the annotation task a lot.

Criteria	Score
Does this toolbox increase your productivity?	4.33/5
How would you rate the performance of the annotation tool?	3.67/5
Is the annotation toolbox useful for you?	4.50/5
Is the toolbox efficient?	4.00/5
Is the interaction with the toolbox clear and understandable?	3.67/5
Are screen message consistent?	3.67/5
Are input prompts clear?	4.00/5
Are error messages helpful?	3.67/5
Was it easy to learn to operate the annotation toolbox?	4.00/5
How easy was the exploration of new features?	3.67/5
Was it easy to remember all commands?	4.33/5
Were help messages on the screen helpful?	3.67/5
Was the supplemental reference material clear?	3.67/5
Was it easy to read and understand the instructions?	4.33/5
The Masterview (side view, front view and bird's-eye-view) simplifies the task.	5.00/5
The organization of information is very clear.	4.00/5
The system speed is fast enough to run the toolbox.	3.67/5
The projection of 3D labels into the image domain is useful.	4.33/5
The interpolation mode helps me to annotate more labels in the same amount of time.	4.67/5
The auto-save functionality was helpful to save time.	4.67/5
The undo/redo functionality follows an important principle of user interface design and increases the satisfaction of users.	4.00/5
	4.12/5

TABLE IV: Annotation statistics of different users on the NuScenes dataset. Annotators who spent more time on the annotation task achieved better accuracy.

	Ground truth	User1	User2	User3
Time (hh:mm):	04:00	00:49	02:30	01:30
Clicks:	4933	2815	4418	2440
Keystrokes:	5812	1521	1488	134
Annotated obj.:	1303	388	395	585
CAR:	1048	374	395	585
PEDESTRIAN:	97	14	0	0
MOTORCYCLE:	0	0	0	0
BICYCLE:	0	0	0	0
TRUCK:	158	0	0	0
3D-IoU:	-	0.068	0.084	0.072
3D-IoU>0.6:	-	0.047	0.054	0.017
Precision:	-	0.004	0.015	0.002
Recall:	-	0.004	0.014	0.002
F_1 score:	-	0.004	0.015	0.002

V. CONCLUDING REMARKS & FUTURE DIRECTIONS

In this research we have contributed to three specific areas. First, we proposed a novel annotation system with tools for efficient, accurate 3D localization of objects and their dynamic movement using full-surround multi-modal (camera and LiDAR) data streams. A robust, scalable and efficient way to annotate 2D and 3D data was provided. The main goal was to create an annotation tool that enables the user to obtain full-surround annotations as well as annotations of tracks. Second, a systematic comparison with 33 annotation tools was performed. Finally, an evaluation of the efficiency and accuracy of the system with four human annotators with diverse backgrounds and skills was carried out, resulting in more than 13,800 object annotations within one hour.

The next immediate step is to test the annotation toolbox

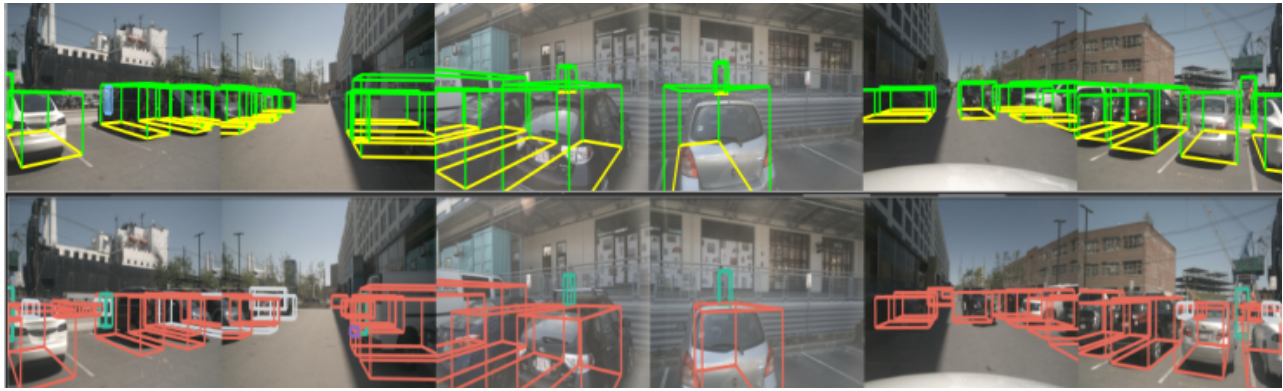


Fig. 8: Comparison between manually created annotations (top) and annotations provided by NuScenes (bottom).

with multiple users, to perform stress tests and make the toolbox ready to be used by multiple users in parallel. A server-side application will be implemented that is connected to a database to obtain and store labels from the crowd.

The second goal is to use a small amount of the dataset that was manually annotated to train object detection algorithms to provide an inference feature that predicts the pose of objects. Using this feature, annotators only need to perform small adjustments and sometimes delete false positives. Annotators also mentioned that interpolation for multiple objects would be very helpful which would further decrease annotation time.

Next, a point cloud aggregation mode will be added to facilitate the labeling of distant objects. Also, a variant of the Iterative Closest Point (ICP) algorithm will be used to ensure that labels move in the correct direction when the user switches to the next frame. Moreover, further annotation modes (attribute annotation, key point annotation, semantic annotation as well as behavior annotation) will be added to the toolbox. Finally, a highly automatic annotation mode will enable the user to obtain annotations very efficiently so that only small corrections by the human are needed.

REFERENCES

- [1] A. Rangesh and M. M. Trivedi, "No blind spots: Full-surround multi-object tracking for autonomous vehicles using cameras & lidars," *arXiv preprint arXiv:1802.08755*, 2018.
- [2] N. Deo, A. Rangesh, and M. M. Trivedi, "How would surround vehicles move? a unified framework for maneuver classification and motion prediction," *IEEE Transactions on Intelligent Vehicles*, vol. 3, no. 2, pp. 129–140, 2018.
- [3] A. Rangesh and M. M. Trivedi, "Ground plane polling for 6dof pose estimation of objects on the road," *arXiv preprint arXiv:1811.06666*, 2018.
- [4] Blender. Blender.org - Home of the Blender project - Free and Open 3D Creation Software. 00000. [Online]. Available: <https://www.blender.org/>
- [5] C. Plachetka, J. Ricken, and M. Maurer, "The TUBS Road User Dataset: A New LiDAR Dataset and its Application to CNN-based Road User Classification for Automated Vehicles," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pp. 2623–2630, 00000.
- [6] D. R. Berger, H. S. Seung, and J. W. Lichtman, "VAST (Volume Annotation and Segmentation Tool): Efficient Manual and Semi-Automatic Labeling of Large 3D Image Stacks," vol. 12, 00001. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6198149/>
- [7] C. Vondrick, D. Patterson, and D. Ramanan, "Efficiently Scaling up Crowdsourced Video Annotation: A Set of Best Practices for High Quality, Economical Video Labeling," vol. 101, no. 1, pp. 184–204, 00000. [Online]. Available: <http://link.springer.com/10.1007/s11263-012-0564-1>
- [8] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman, "LabelMe: A Database and Web-Based Tool for Image Annotation," vol. 77, no. 1-3, pp. 157–173, 02439. [Online]. Available: <http://link.springer.com/10.1007/s11263-007-0090-8>
- [9] T.-T. Kuo, J. Huh, J. Kim, R. El-Kareh, S. Singh, S. F. Feupe, V. Kuri, G. Lin, M. E. Day, L. Ohno-Machado, and C.-N. Hsu, "The Impact of Automatic Pre-annotation in Clinical Note Data Element Extraction - the CLEAN Tool," 00000. [Online]. Available: <http://arxiv.org/abs/1808.03806>
- [10] P. Stenetorp, S. Pyysalo, G. Topic, T. Ohta, S. Ananiadou, and J. Tsujii, "Brat: A Web-based Tool for NLP-Assisted Text Annotation," p. 6.
- [11] R. E. de Castilho, S. M. Yimam, S. Hartmann, I. Gurevych, A. Frank, and C. Biemann, "A Web-based Tool for the Integrated Annotation of Semantic and Syntactic Structures," p. 9, 00020.
- [12] N. Fiedler, M. Bestmann, and N. Hendrich, "ImageTagger: An Open Source Online Platform for Collaborative Image Labeling," p. 8.
- [13] ScaleAi. Sensor Fusion: 3D Point Cloud Annotation - Scale. 00000. [Online]. Available: <https://scale.ai/sensor-fusion-annotation>
- [14] playment. 3D Point Cloud Annotation — Data Labeling Platform. 00000. [Online]. Available: <https://playment.io>
- [15] J. Xie, M. Kiefel, M.-T. Sun, and A. Geiger, "Semantic Instance Annotation of Street Scenes by 3D to 2D Label Transfer," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, pp. 3688–3697, 00048. [Online]. Available: <http://ieeexplore.ieee.org/document/7780770/>
- [16] Dataturks. Best online platform for your ML data annotation needs. 00000. [Online]. Available: <https://dataturks.com/>
- [17] supervisely. Supervisely - Web platform for computer vision. Annotation, training and deploy. 00000. [Online]. Available: <https://supervise.ly/>
- [18] labelbox. Labelbox: The best way to create and manage training data. 00000. [Online]. Available: <https://labelbox.com/automotive-industry>
- [19] Neurala. Brain Builder Beta — Neurala. 00000. [Online]. Available: <https://www.neurala.com/solutions/brain>
- [20] Prodigy. Prodigy An annotation tool for AI, Machine Learning & NLP. 00000. [Online]. Available: <https://prodigy.ai/index>
- [21] A. Rangesh, N. Deo, K. Yuen, K. Pirozhenko, P. Gunaratne, and M. M. Trivedi, "Exploring the Situational Awareness of Humans Inside Autonomous Vehicles," p. 8.
- [22] H. Caesar, "The devkit of the nuScenes dataset. Contribute to nutonomy/nuScenes-devkit development by creating an account on GitHub," nuTonomy, 00000. [Online]. Available: <https://github.com/nutonomy/nuScenes-devkit>
- [23] M. Smit. WhatPulse. 00000. [Online]. Available: <https://whatpulse.org/>
- [24] C. Goutte and E. Gaussier, "A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation," in *Advances in Information Retrieval*, ser. Lecture Notes in Computer Science, D. E. Losada and J. M. Fernandez-Luna, Eds. Springer Berlin Heidelberg, pp. 345–359, 00469.