Camera-based Context-aware Traffic Light Detection for Self-Driving Vehicles

Daiki Shiotsuka[1], Yuto Nakamura[1,2], Kohei Iwamasa[1], Yu Yamaguchi[1], and Shunsuke Aoki[1,3]
[1] Turing Inc., Tokyo, Japan,
[2] The University of Tokyo, Tokyo, Japan,
[3] National Institute of Informatics, Tokyo, Japan,
daiki.shiotsuka@turing-motors.com

Abstract

In the rapidly evolving field of autonomous driving, accurately detecting traffic lights is a paramount challenge. This paper introduces a novel dataset and an accompanying model designed specifically for detecting traffic signals crucial to autonomous vehicles. Our approach focuses on lights directly relevant to the ego vehicle, incorporating 44,000 meticulously annotated images that represent diverse signal states. The development of this dataset involved an iterative process, guided by the initial model's performance, resulting in significant enhancements in efficiency and accuracy. The model exhibits a robust capability to discern pertinent traffic signals in complex environments, marking a significant advancement in improving navigation and safety for autonomous vehicles.

1 Introduction

In autonomous driving, traffic light detection is one of the key technologies. Recently, with the advancement of deep learning, many traffic light detection methods have been proposed[1,2].

Deep learning requires large datasets, and there are several datasets for traffic lights. The Bosch Small Traffic Lights Dataset (BSTLD) [3] includes 8,334 images from video data recorded on University Avenue in Palo Alto, California, labeling signals as "off", "green", "red", and "yellow". The VIVA dataset [4] contains over 40,000 images with bounding boxes for traffic signals, classified into seven categories such as GO, WARNING, and STOP. Using images from the COCO dataset [5], the COCO Traffic dataset [6] provides a traffic signal detection dataset. While these datasets cover different signal aspects, none offer information about the signal relevant to the ego vehicle. In the context of autonomous driving, if the ego car fails to correctly identify the traffic light it should follow, this could potentially lead to significant accidents. Addressing this gap, several studies have developed datasets that emphasize the relationship between traffic lights and self-driving vehicles. The DriveU Traffic Light Dataset (DTLD) [7] boasts over 230,000 annotations, capturing various situations using an array of tags encompassing state (red, yellow, green, red-yellow, off), pictograms (circle, arrow left, pedestrian, etc.), among others. It also includes a relevance tag to denote the significance of traffic lights to the ego vehicle. Cityscapes TL++ [8] is an extension of the Cityscapes dataset [9], specifically developed for traffic signal detection and also considers the relevance of signals to the ego vehicle. The LAVA Salient Lights Dataset [10] not only annotates the state, color, directionality, and obstruction of signals but also includes a salience attribute, indicating a traffic signal's importance. It labels 30,566 signals, identifying 9,051 as important and 21,515 as not. Our dataset is uniquely focused on annotating only those traffic signals directly relevant to the self-driving vehicle's lane or path, thereby capturing the most critical signals for the vehicle's navigation. Also, our dataset comprises over 1000 hours of driving footage primarily from Tokyo, where the majority of traffic signals are horizontal.

Symposium on Machine Learning for Autonomous Driving (ML4AD 2023).

With the development of object detection techniques, neural network-based frameworks are widely used. Two-step frameworks have been proposed for conducting detection and classification separately [11-14]. Otherwise, One-step frameworks are efficient at performing detection and classification tasks simultaneously, including those based on YOLO[15-19], Faster R-CNN[20,21], and SSD[22,23]. In this paper, we employ a method based on YOLOX[24]. YOLOX is a simple detector that adopts an anchor-free manner and achieves high performance by improving the learning method.

Our contributions are as follows:

- We annotated traffic signals on 44,000 images, each labeled with bounding boxes and seven attributes: red, green, yellow, unknown, left-arrow, straight-arrow, and right-arrow. Notably, these attributes are not mutually exclusive and can be combined to accurately represent the diverse states of traffic signals.
- Our dataset exclusively focuses on annotating traffic signals that are imperative for the ego vehicle's immediate attention.
- We adopted an iterative approach in dataset development. Starting with a provisional dataset to train an initial model, we then identified and targeted the model's weaknesses, refining the dataset accordingly. This method markedly enhanced our model's efficiency and performance.
- The model's detection accuracy was impressive in quantitative assessments, demonstrating high precision. In qualitative evaluations, we confirmed that the model effectively detected traffic signals most relevant to the ego vehicle.

2 Dataset

For the creation of a traffic signal detection model, we gathered extensive data and manually annotated a large number of images.

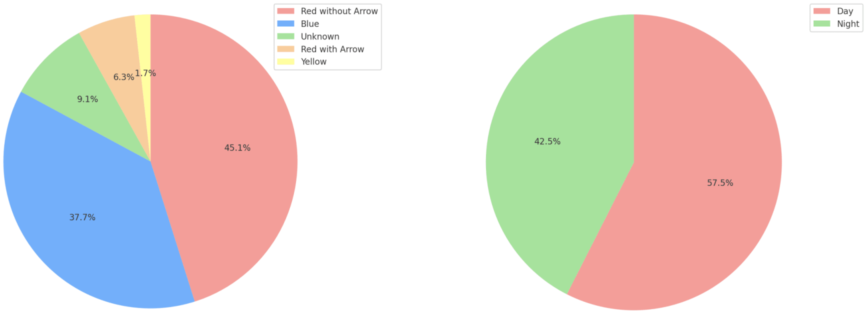
2.1 Data Collection

We collected data that consists of approximately 1,000 hours of video footage recorded by a forward-facing camera mounted on the front of a car. The data covers daytime and nighttime from June 2022 to March 2023, primarily in Tokyo. It includes a variety of locations and situations such as urban areas, highways, and mountainous regions. The videos were recorded at 20Hz, with an image resolution of 1928 x 1208 pixels.

2.2 Data Annotation

We conducted annotation in two stages. Initially, we randomly selected 20,000 images from the collected data. Figure 1 shows the distribution of images in the first annotation dataset. As seen in Figure 1-a, 57.5% of the images are taken during the day and 42.5% at night. Figure 1-b represents the distribution of traffic signal attributes. Here, *Red without Arrow* refers to red signals without arrow indicators, and *Red with Arrow* refers to signals including arrow indicators. During annotation, we labeled each object in the images with its position [x, y, w, h] and attributes (red, green, yellow, unknown, left-arrow, straight-arrow, right-arrow). Here, x and y denote the top-left coordinates of the object's bounding box, while w and h represent its width and height, respectively, following the same format as the COCO dataset. Moreover, the attributes are not mutually exclusive and can be labeled simultaneously as characteristics of the signal. In this process, we did not annotate every traffic signal in the images. Instead, our annotators manually selected and annotated only the signal deemed most crucial for the vehicle to follow at that moment. This approach ensures that our traffic signal model can accurately identify the appropriate signal to follow, even in complex situations.

We trained the initial model using the initial dataset composed of the 20,000 images annotated as described above. We chose the YOLOX-based model. The details of the training process are explained in the section Model Training. Training results indicated a comparatively lower detection rate for yellow signals relative to other attributes. Hence, we used the initial model to extract images containing yellow signals from the data that had not yet been annotated. YOLOX allows for adjusting the detection ease via a threshold value, which we significantly lowered to gather more images with yellow signals. Using this approach, we compiled a dataset of 24,000 images, ensuring an increased data of yellow signals. Figure 2 shows the distribution of the second dataset. The imbalance of attributes within the dataset has shown improvement compared to the first dataset. Ultimately, we trained the YOLOX-x model using a combined dataset of 44,000 images from both datasets. The dataset includes labels for 19,509 traffic signals.



(a) Traffic light attribute distribution

(b) Traffic light Day/Night distribution

Figure 1: Distribution of data in the first dataset

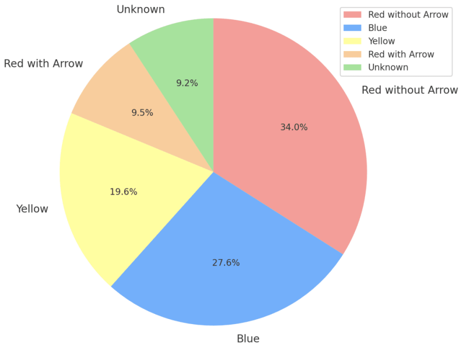


Figure 2: Traffic light attribute distribution of data in the second dataset

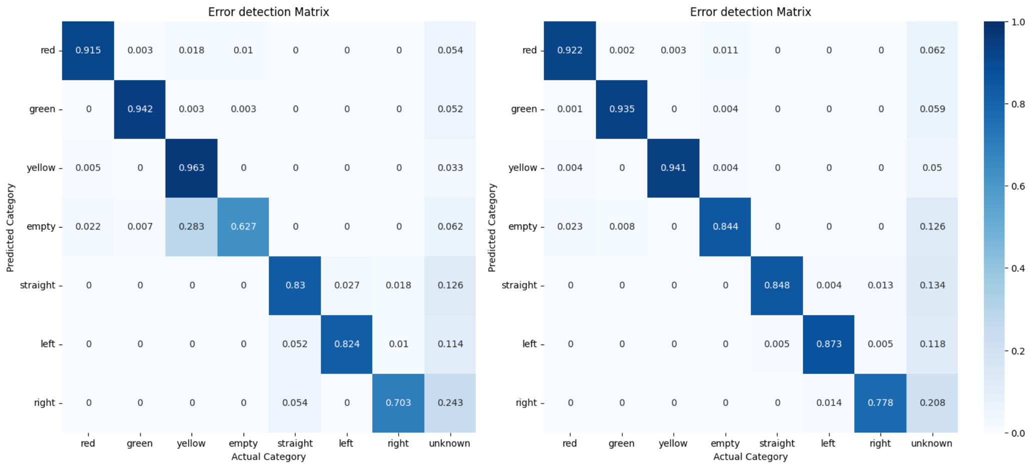
3 Model Training

In our model, we utilized the YOLOX-X architecture as a foundation, enhancing it with a newly added head to classify the attributes of traffic signals. To achieve this, we employed transfer learning for traffic light detection tasks from a pre-trained YOLOX-X model. For our loss function, we chose Binary Cross Entropy Loss. This adaptation was crucial in enabling our model to discern complex states, such as identifying a red traffic signal that simultaneously indicates both straight and right arrows.

Our training approach was a two-stage strategy, outlined as follows:

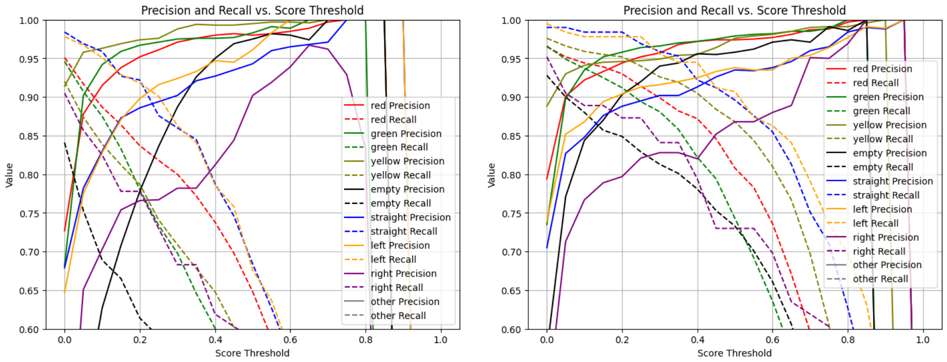
First, we trained the core model using the initial dataset. The dataset was split with 80% for training and 20% for validation. We used the YOLOX training framework, resizing input images to 640×840 pixels and adjusting bounding box coordinates to this resolution. Our training parameters followed YOLOX's recommendations, including a 30-epoch training period. Notably, we set the horizontal flip probability to 0, a modification tailored to our dataset which predominantly features horizontal traffic signals, as is common in Japan.

Second, we trained another model using the second dataset, keeping the training parameters consistent with the initial setup.



(a) First training model (b) Second training model

Figure 3: Confusion matrix of these models



(a) First training model (b) Second training model

Figure 4: The relationship between the Precision/Recall and the score threshold

4 Experiment

4.1 Quantitative Evaluation

Figure 3 presents the confusion matrix of precision for the two models, using validation data and setting the score threshold at 0.1 and the Intersection over Union (IoU) threshold at 0.4. From this figure, the second model exhibits enhanced accuracy in detecting traffic signals when compared to the first model.

Figure 4 shows the relationship between the Precision/Recall and the score threshold for the two models. Based on this observation, it's evident that the second model, with its notably improved recall, experiences fewer undetected cases even at higher score thresholds. This reduced dependency on the threshold value compared to the first model suggests an enhancement in the detection capabilities of the second model, indicating a more robust and reliable performance across various threshold settings.

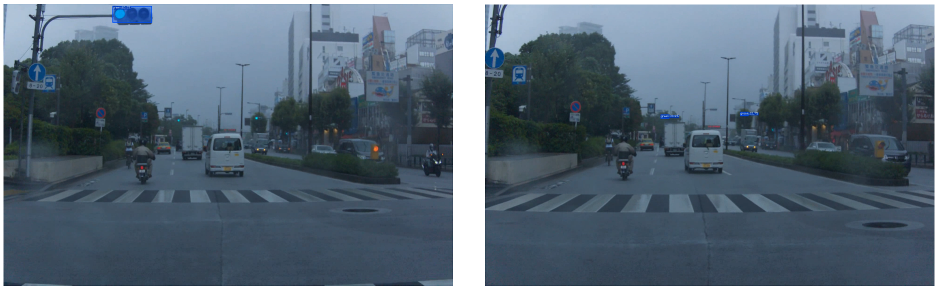
4.2 Qualitative Evaluation

In the proposed method, during the dataset creation phase, we annotated only the traffic signals that were relevant to follow. Figure 5 demonstrates an example of detection by our second model at times t and t+1. At time t, the model detects only the nearest traffic signal, ignoring the distant one. At t+1, when the nearer traffic signal is no longer present in the image, the model successfully detects the previously ignored distant signal. This indicates that our proposed model can discern and focus on the traffic signals that are most relevant in the image.

Also, Figure 6 shows an example of an urban area intersection. Our model successfully detects only the relevant green traffic signal ahead, which the driver should follow, and does not detect the yellow signal on the right side of the image. This example further demonstrates the effectiveness of our approach in detecting the traffic signal that requires attention.

Moreover, Figure 7 in the article illustrates the detection results for nighttime data. In subfigure 7-a, the proposed model successfully ignores various elements, including pedestrian signals and opposing vehicle headlights, amidst multiple light sources, and even correctly identifies arrow signals. In subfigure 7-b, the model accurately recognizes a red signal despite the scattering of light caused by rain. These observations demonstrate the model's capability to accurately infer information in nighttime conditions. The effectiveness of the model in complex lighting scenarios, such as those presented by night conditions with additional challenges like rain, underscores its robustness and adaptability in diverse environments. This reinforces the model's applicability for real-world scenarios where variable lighting conditions are a common occurrence.

Finally, we present examples of detection failures. In Subfigure 8-a, the system detects a traffic light at a merging point and mistakenly recognizes it as a green signal. However, in this scenario, it should have detected the red signal in the vehicle's own lane. Subfigure 8-b shows the system incorrectly detecting an irrelevant traffic signal at an intersection. These situations present significant judgment challenges, underscoring the need for integrating methods such as lane topology estimation.



(a) Result at time t

(b) Result at time t+1

Figure 5: Detection results at times t and t+1, demonstrating the model's capability to adjust focus from the nearer to the distant traffic signal as the scene changes.

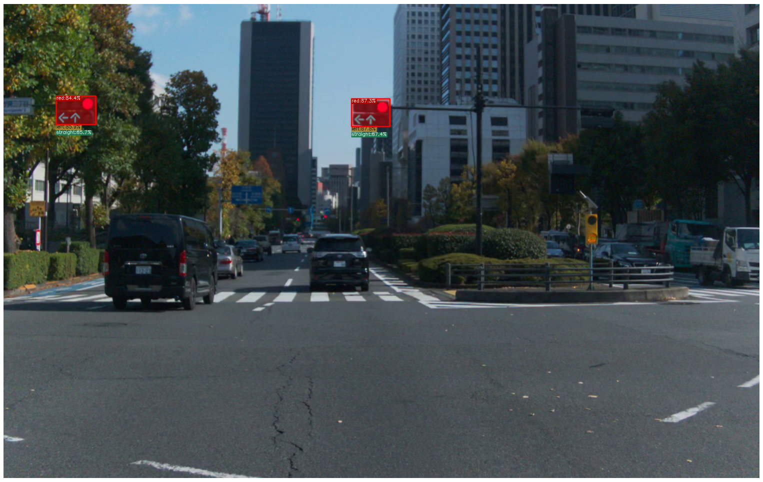
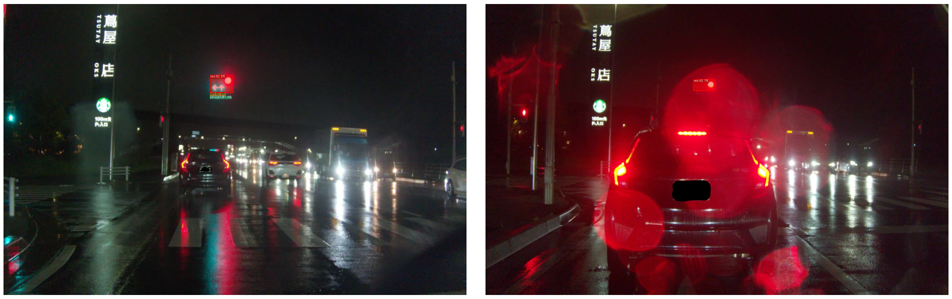


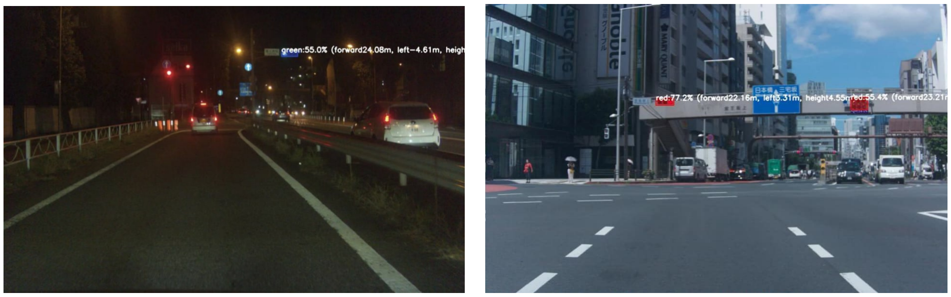
Figure 6: Example of detection at an urban intersection, where the model identifies only the relevant green traffic signal ahead, effectively ignoring the yellow signal on the right side of the image.



(a) Detection of traffic signals and headlights among various night light sources.

(b) Recognition of a red signal in rainy, light-scattering conditions.

Figure 7: Nighttime Detection Results of the Proposed Model.


green:55.0% (forward24.08m, left-4.61m, heigh
red:77.2% (forward22.16m, left3.31m, height4.55m
red:55.4% (forward23.21r
日本橋
三宅坂

(a) Fail result at a merging point.

(b) Fail result at an Intersection.

Figure 8: Examples of Traffic Signal Detection Failures.

5 Conclusion

In this paper, we introduced a dataset and an associated model specifically designed to detect the traffic signals that are most relevant to the ego vehicle. Our quantitative evaluation shows the efficacy of our multi-stage dataset creation process, which included annotating additional data based on the performance of the initial model. The qualitative evaluation confirms our model's ability to accurately detect relevant traffic signals. In this study, the traffic signals for the vehicle to follow were identified through human annotation. In future work, we can anticipate the development of a more robust traffic signal detection model that explicitly considers factors such as lane topology to detect signals on the ego vehicle's lane, enhancing its relevance.

References

[1] Cabrera, M. D., Cerri, P., Pirlo, G., Ferrer, M. A., Impedovo, D. (2015). A survey on traffic light detection. In *New Trends in Image Analysis and Processing - ICIAP - Workshops*. Springer.

[2] Pavlitska, S., Lambing, N., Bangaru, A. K., Zöllner, J. M. (2023). Traffic Light Recognition using Convolutional Neural Networks: A Survey. arXiv:2309.02158.

[3] Behrendt, K., Novak, L., Botros, R. (2017). A deep learning approach to traffic lights: Detection, tracking, and classification. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE.

[4] Jensen, M. B., Philipsen, M. P., Møgelmose, A., Moeslund, T. B., Trivedi, M. M. (2016). Vision for looking at traffic lights: Issues, survey, and perspectives. *IEEE transactions on intelligent transportation systems*.

[5] Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C. L. (2014). Microsoft coco: Common objects in context. In *European Conference on Computer Vision (ECCV)*. Springer.

[6] Kirchhoff, D., Hoang, P. (2023). COCO Dataset Extensions for Driving Tasks. Available at: https://www.neuralception.com/cocodatasetextension/. Accessed: 2023-11-15.

[7] Fregin, A., Muller, J., Krebel, U., Dietmayer, K. (2018). The driveu traffic light dataset: Introduction and comparison with existing datasets. In *International Conference on Robotics and Automation (ICRA)*. IEEE.

[8] Janosovits, J. (2022). Cityscapes tl++: Semantic traffic light annotations for the cityscapes dataset. In *International Conference on Robotics and Automation (ICRA)*. IEEE.

[9] Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., Schiele, B. (2016). The cityscapes dataset for semantic urban scene understanding. In *Conference on Computer Vision and Pattern Recognition (CVPR).*

[10] Greer, R., et al. (2023). Robust Traffic Light Detection Using Salience-Sensitive Loss: Computational Framework and Evaluations. arXiv preprint arXiv:2305.04516.

[11] Vitas, D., Tomic, M., Burul, M. (2020). Traffic light detection in autonomous driving systems. *IEEE Consumer Electronics Magazine*.

[12] Bali, S., Kumar, T., Tyagi, S. (2022). Lightweight deep learning model for traffic light detection. In *International Conference on Technological Advancements in Computational Sciences (ICTACS)*. IEEE.

[13] Jayasinghe, O., Hemachandra, S., Anhettigama, D., Kariyawasam, S., Wickremasinghe, T., Ekanayake, C., Rodrigo, R., Jayasekara, P. (2022). Towards real-time traffic sign and traffic light detection on embedded systems. In *Intelligent Vehicles Symposium (IV)*. IEEE.

[14] Lin, S.-Y., Lin, H.-Y. (2022). A two-stage framework for diverse traffic light recognition based on individual signal detection. In *Pattern Recognition and Artificial Intelligence: Mediterranean Conference (MedPRAI)*. Springer.

[15] Yan, S., Liu, X., Qian, W., Chen, Q. (2021). An end-to-end traffic light detection algorithm based on deep learning. In *International conference on security, pattern analysis, and cybernetics (SPAC)*. IEEE.

[16] Xiang, N., Cao, Z., Wang, Y., Jia, Q. (2021). A real-time vehicle traffic light detection algorithm based on modified yolov3. In *International Conference on Electronics Technology (ICET)*. IEEE.

[17] DeRong, M., ZhongMei, T. (2023). Remote traffic light detection and recognition based on deep learning. In *World Conference on Computing and Communication Technologies (WCCCT)*. IEEE.

[18] Liu, P., Li, T. (2023). Traffic light detection based on depth improved yolov5. In *International Conference on Neural Networks, Information and Communication Engineering (NNICE)*. IEEE.

[19] Wang, Q., Zhang, Q., Liang, X., Wang, Y., Zhou, C., Mikulovich, V. I. (2022). Traffic lights detection and recognition method based on the improved yolov4 algorithm. *Sensors*.

[20] Pon, A., Adrienko, O., Harakeh, A., Waslander, S. L. (2018). A hierarchical deep architecture and mini-batch selection method for joint traffic sign and light detection. In *Conference on Computer and Robot Vision (CRV)*. IEEE.

[21] Bach, M., Stumper, D., Dietmayer, K. (2018). Deep convolutional traffic light recognition for automated driving. In *International Conference on Intelligent Transportation Systems (ITSC)*. IEEE.

[22] Müller, J., Dietmayer, K. (2018). Detecting traffic lights by single shot detection. In *International Conference on Intelligent Transportation Systems (ITSC)*. IEEE.

[23] Naimi, H., Akilan, T., Khalid, M. A. (2021). Fast traffic sign and light detection using deep learning for automotive applications. In *IEEE Western New York Image and Signal Processing Workshop (WNYISPW)*. IEEE.

[24] Ge, Z., et al. (2021). YOLOX: Exceeding YOLO series in 2021. In *arXiv preprint arXiv:2107.08430*.