GLARE: A Dataset for Traffic Sign Detection in Sun Glare

Nicholas Gray, Megan Moraes, Jiang Bian, *Member, IEEE*, Alex Wang, Allen Tian, Kurt Wilson, Yan Huang, Haoyi Xiong, and Zhishan Guo, *Senior Member, IEEE*

***Abstract*— Real-time machine learning object detection algorithms are often found within autonomous vehicle technology and depend on quality datasets. It is essential that these algorithms work correctly in everyday conditions as well as under strong sun glare. Reports indicate glare is one of the two most prominent environment-related reasons for crashes. However, existing datasets, such as the Laboratory for Intelligent & Safe Automobiles Traffic Sign (LISA) Dataset and the German Traffic Sign Recognition Benchmark, do not reflect the existence of sun glare at all. This paper presents the GLARE (GLARE is available at: https://github.com/NicholasCG/GLARE_Dataset) traffic sign dataset: a collection of images with U.S-based traffic signs under heavy visual interference by sunlight. GLARE contains 2,157 images of traffic signs with sun glare, pulled from 33 videos of dashcam footage of roads in the United States. It provides an essential enrichment to the widely used LISA Traffic Sign dataset. Our experimental study shows that although several state-of-the-art baseline architectures have demonstrated good performance on traffic sign detection in conditions without sun glare in the past, they performed poorly when tested against GLARE (e.g., average mAP$_{0.5:0.95}$ of 19.4). We also notice that current architectures have better detection when trained on images of traffic signs in sun glare performance (e.g., average mAP$_{0.5:0.95}$ of 39.6), and perform best when trained on a mixture of conditions (e.g., average mAP$_{0.5:0.95}$ of 42.3).**

***Index Terms*— Traffic sign detection, public data set, sun glare.**

I. Introduction

DRIVING has seen its numerous phases of evolution, from being steam-propelled to becoming almost fully autonomous. Throughout these developments, the existence of one phenomenon has remained constant in the daily environment — intense sunlight which can obstruct the view of a vision sensor (either eyes or cameras) while maneuvering a vehicle. When the sun descends on the horizon, sun glare seeps below a car's visor and visually impairs vision sensors, causing difficulty in navigating everyday traffic. Temporary blindness (due to sun glare) causes difficulty in sensing other cars, and traffic signs, often leading to accidents. As a result of sun glare, a recent report [1] by the Department of Transportation has stated that as many as 9,000 glare-related accidents occur each year, making it one of the two most prominent environment-related reasons for crashes. The combination of harsh sun glare with common driving risk factors contributes to more crashes and congestion in day-to-day driving, leading to setbacks in implementing new automotive technologies.

There has been an upsurge of autonomous vehicles driving alongside everyday drivers, such as Tesla or Google's Waymo. These self-driving vehicles make their decisions through the use of object detection algorithms, which allows the autonomous system to locate objects (such as traffic signs on the road) using bounding boxes, classify them, and make a real-time decision (machine learning [2]) based on the algorithmic interpretation of the seen object. The functionality of these algorithms heavily depends on rich sets of data that are collected from real-world scenarios, annotated, and fed to "teach" algorithms what it may experience on the road. One set of data frequently used to teach algorithms within autonomous cars includes traffic signs—critical for navigating everyday traffic. While there are several datasets publicly available that focus on traffic signs in regular weather conditions, there is *few traffic sign dataset focusing on traffic signs with sun glare*. Our experiments indicate that when vehicles are continuously trained to recognize objects using data without sun glare, real-time algorithms within cars may fail to detect traffic signs and other objects when blinded by high-intensity visual noise, leading to catastrophe.

Datasets containing traffic signs with sun glare are often internal within autonomous driving companies and consequently are not publicly available for wider research purposes. While existing public datasets (such as the Laboratory for Intelligent & Safe Automobiles Traffic Sign (LISA) Dataset [3]) do not contain any sun glare at all, there is an emerging need to create a public dataset with a wide variety of traffic signs with sun glare interference to fill this disparity.

A. Contributions

As an addition to the LISA Traffic Sign dataset, we establish the GLARE dataset — a collection of images with traffic signs

Manuscript received 27 August 2022; revised 7 April 2023; accepted 14 June 2023. This work was supported in part by NSF under Grant CNS-1850851 and Grant CCF-2028481. The Associate Editor for this article was Z. Duric. *(Nicholas Gray, Megan Moraes, Jiang Bian, and Alex Wang contributed equally to this work.) (Corresponding author: Zhishan Guo.)*

Nicholas Gray, Megan Moraes, and Yan Huang are with the Department of Electrical and Computer Engineering (ECE), University of Central Florida, Orlando, FL 32816 USA.

Jiang Bian is with the Department of Electrical and Computer Engineering (ECE), University of Central Florida, Orlando, FL 32816 USA, and also with the Baidu Research Laboratory, Beijing 100084, China.

Alex Wang is with Trinity Preparatory School, Orlando, FL 32792 USA, and also with the Department of Computer Science, Georgia Tech University, Atlanta, GA 30332 USA.

Allen Tian is with Chapel Hill High School, Chapel Hill, NC 27516 USA, and also with the Department of Computer Science, The University of Chicago, Chicago, IL 60637 USA.

Kurt Wilson and Zhishan Guo are with the Department of Electrical and Computer Engineering (ECE), University of Central Florida, Orlando, FL 32816 USA, and also with the Department of Computer Science, North Carolina State University, Raleigh, NC 27695 USA (e-mail: zguo32@ncsu.edu).

Haoyi Xiong is with the Baidu Research Laboratory, Beijing 100084, China.

Digital Object Identifier 10.1109/TITS.2023.3294411

TABLE I

COMPARISON OF EXISTING TRAFFIC SIGN DATASETS

Dataset	Images	Image Resolution	Classes	Tasks[*]	Features	Country	Year
STS [4]	3,777	1280×960	20	Both	w/ general occlusion	Sweden	2011
GTSRB [5]	51,839	15×15 ∼ 250×250	43	Recognition	w/ general occlusion	Germany	2011
GTSDB [6]	900	1360×1024	43	Localization	w/ general occlusion	Germany	2013
LISA [3]	6,610	640×680 ∼ 1024×522	49	Both	w/ general occlusion	USA	2012
TT-100K [7]	100,000[†]	2048×2048	221	Both	w/ general occlusion	China	2016
MTSD [8]	100,000	1000×1000 ∼ 2048×2048	313	Both	w/ general occlusion	Worldwide	2019
DFG [9][‡]	6,957	720 ×576∼1920×1080	200	Both	w/ general occlusion	Slovenia	2019
GLARE	2,157	720×480∼ 1920×1080	41	Both	**w/ heavy glares**	USA	2022

[*] We target localization and the recognition (classification) tasks here.

[†] Only 10,000 images contain traffic signs.

[‡] The DFG Traffic Sign Dataset uses polygon annotations, instead of bounding box annotations.

which have heavy visual interference as a result of strong sunlight. GLARE will be a publicly available set of images for training real-time object detection algorithms and more. This dataset and the proposed algorithms are intended to act as a baseline for upcoming researchers while developing, training, and examining their own models. The contributions of this work comes in three folds:

- We establish a fine-grained traffic sign dataset, GLARE, abundant with realistic glares on or near the traffic sign areas. To our knowledge, GLARE is the first traffic sign dataset with detailed annotations of sun glares, covering varied scenarios of glare conditions from daily driving. Compared to the commonly used dataset (e.g., LISA [3], GTSDB [6], and TT-100K [7]), GLARE provides pure observation of traffic sign with glares instead mixing with a sparse witness of general occlusions. We follow the standard format to annotate, calibrate, and reorganize the dataset for a wide range of research tasks (e.g., traffic sign localization, image classification, and temporal localization).
- We also have released the full procedures to step-by-step create the dataset and analyze its statistical features.
- We further showcase the research potentials of the GLARE dataset by testing it on a large group of benchmarks. Specifically, we observe that the performances of mainstream object detection architectures used in real-time traffic sign detection degrades drastically when trained on the LISA dataset, whereas training with the GLARE dataset shows a significant performance gain instead.

B. Organization

The rest of the paper is organized as follows: Section II summarizes the existing related work of traffic sign datasets and cutting-edge object detection architectures. Section III details the dataset including its collection, annotation, and statistics. Section IV reports the experiments to check the testing performance of the mainstream object detection architectures with both partially and entirely and without the GLARE dataset in the training phase. Section V concludes the paper and suggests ideas for future research.

II. RELATED WORK

A. Traffic Sign Datasets

With the advancement of autonomous driving, there has been an emphasis on collecting data with all types of road conditions, signs, and any factor to note while driving, leading to a plethora of datasets in the community specific to traffic sign detection.

Several datasets tend to focus on traffic signs found globally, each with variations. For example, the German Traffic Sign Recognition Benchmark [6] focuses on traffic signs from Germany and captured images in different environments under varied weather conditions. Others that follow a similar pattern include the Tsinghua-Tencent 100K dataset [7], the Swedish Traffic Sign dataset [4], and the Belgium Traffic Signs dataset [10]. It is advantageous to the computer vision community to have access to traffic signs from around the world, but there is a significant drawback common to public traffic sign datasets: a lack of sun glare within its images.

The use of convolutional neural networks (CNNs) is prevalent throughout traffic sign datasets, often for the tasks of localization, recognition, and joint localization and recognition, which is commonly referred to as object detection. To set the standard for these tasks, baselines are often attached to datasets in the form of varied CNNs. The Mapilliary dataset [8], for example, uses a Faster regional-based convolutional neural network (R-CNN) based detector to produce mean average precision (mAP) results over all of its classes. The DFG Traffic Sign dataset [9] is another example that uses such techniques to establish a baseline, utilizing a Faster R-CNN and a Mask R-CNN to provide mAP values ranging in the upper 90s. Although the dataset includes traffic-sign instances with synthetic distortions that may resemble sun glare, these images are incomparable to those with natural sun glare. Generalization performance may suffer from improper training when strictly utilizing datasets that lack natural sun glare. This is a phenomenon found often throughout datasets focused on traffic sign depiction: models are trained on data without obscuring conditions or with synthetic ones, such as sun glare, usually are unsatisfactory when tested in real driving scenarios.

Severe conditions, such as sun glare or heavy rain, impede the visibility of traffic signs while driving. Just as it hinders human drivers, it additionally interferes with algorithmic vision. The CURE-TSD-Real dataset [11], comprised of traffic sign images in simulated heavy road conditions, is an example of a dataset with severe conditions that resulted in a 29% drop in average precision. This motivates us to investigate the possibility of harsh sun glare causing a drop in algorithmic performance as well.

The GLARE dataset intends to be an extension of the LISA dataset, which is one of the most commonly used American traffic sign datasets with an emphasis on large variations within urban landscapes. The dataset is comprised of videos and stand-alone images of traffic signs, amounting to about 6,610 images and 7,855 annotations. Source data for the LISA dataset comes with color, in grayscale, and does not include images with excessive sun glare. The LISA dataset answered a need for a public dataset with US-based traffic signs and notably contributed more as it includes full traceability of its dataset by providing full annotations of all images, and includes all associated tracks. We provide full comparisons of the existing related traffic sign datasets in Table I with GLARE, which shows that GLARE is the latest traffic dataset (with two years gap from DFG and MTSD and nine years gap from original LISA) and is formed by high-resolution images with heavy/harsh glares on traffic signs.

B. Traffic Sign Classification

One of the most popular applications with the aforementioned datasets is traffic sign classification, where tremendous efforts are accomplished from statistical learning to deep learning paradigms. For example, Soendoro and Supriana [12] first adopt the SVMs with sparse representation to recognize the class of traffic signs in images. With the rise the deep learning, convolutional neural networks begin to dominate the performances of recognition/classification in the traffic sign domain. Specifically, on the GTSRB dataset, a large amount of CNN variants [13] shows a powerful ability for generalization, where the classification accuracy on the testing set is even better than the performance of human experts (e.g., CNNs with spatial transformers [14] can achieve roughly 99.7% in terms of top-1 accuracy). Note that, the reported high classification accuracy is based on the cropped traffic sign images, where we can obtain these images via a specifically designed object detection task.

C. End-to-End Traffic Sign Detection

At first, the localization and the classification are two independent tasks, where the classification is built upon the properly localized bounding boxes (i.e., the traffic sign is located and extracted intact). With the rise of the CNNs, the original powerful performance of image classification/recognition [15] rapidly transfers to object detection domain. Furthermore, it has been a consensus that the family of CNNs is capable of detecting a bounding box for a specific object while classifying its category simultaneously. The well-known RCNN/Fast-RCNN [16], [17] first generates potential bounding boxes on the frames and then classifies the object only in these bounding boxes. However, the final performance of object detection depends on the performances of multiple stages during the complex pipeline (i.e., preprocessing, classification, and post-processing to re-score the proposed bounding boxes), where the whole process is slow as well. To address the efficiency and complexity issue, [18] proposes the first edition of You Only Look Once (YOLO) series, v1 to v5, and treats the object detection as a single regression task to directly establish the connection among the image pixels, the bounding box coordinates and the labels with probabilities. To further improve the performance, YOLOX [19] is proposed as incorporating the anchor-free manner and several cutting-edge detection techniques (e.g., decoupled head, dynamic label assignment strategy).

Another branch of object detection strategies leverages the popular transformer [20] encoder-decoder architectures by removing the complicated hand-designed components such as non-maximum suppression or anchor generation while optimizing a global loss that enables unique classifications via bipartite matching. Similar ideas are brought into traditional RCNN architecture that a special Swin Transformer [21] shows a great performance gain when replacing the ResNet50 backbone. Almost all the aforementioned object detection algorithms rely on supervised learning with labeled traffic signs and it is rarely considered that the possible strong localized noises (e.g., sun glare) in the testing phase may degrade the detection performance.

III. GLARE Dataset

This section presents the GLARE traffic sign dataset, a sun glare focused dataset to assist researchers and developers in building real-time autonomous traffic sign detection and classification system in sun glare conditions. The dataset includes 2157 images and annotations, each containing a single traffic sign annotation. This dataset can be used for object localization, recognition, and object detection tasks.

A. Video Collection and Processing

The initial collection process started with three dashboard cameras recording approximately 38 hours of footage around the Orlando area. Two cameras were forward facing with one filmed at 1920 × 1080 (1080p) and the other rearward facing one filmed at 720 × 480 (480p). In total, the cameras filmed 463 initial videos of 40 hours and 25 minutes of footage.

The first step in video processing was to remove all videos that did not meet the criteria of having both sun glare and traffic signs at the time. This resulted in 163 videos that contained some amount of sun glare. The second step was to extract the sections of video with sun glare, referred to as *clips*. The clips each contained a continuous presence of sun glare, with about a half second of extra time at the start of each clip to allow for ease of finding the beginning of the sun glare. Since only footage that concurrently contains sun glare and traffic signs matters, any clips that did not contain traffic signs during the follow-up screening section were discarded. Following a similar procedure as the LISA dataset [3], these short videos are referred to as *tracks*. 189 tracks were used in the creation of the GLARE dataset, totaling 18 minutes and 11 seconds of footage. The tracks were organized by their original source video, with 33 original source videos being used in total to produce the GLARE dataset.

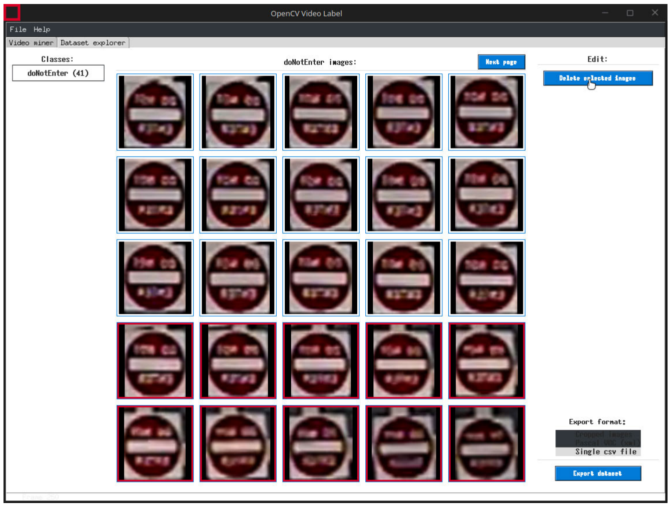
B. Annotation Process

The image annotation process was separated into two main steps: bounding box localization of traffic signs, and bounding box approval with cleanup. The first step was completed by two individuals at the same time with a single open-source tool [22] to allow for efficient labeling. The second step was completed after the initial processing of all the images as a quality assurance step by two individuals who worked on labeling and processing the initial bounding boxes in the LISA dataset format [3]. The automatic bounding box tracking algorithms available with the tool were Re3 [23] and CMT (Consensus-based Matching and Tracking of Keypoints for Object Tracking) [24], and we used Re3 for all our annotations due to stable tracking at all sizes. When annotations were saved, each image that was exported had a single associated annotation.

1) Traffic Sign Localization: In the first step, tracks were processed together based on the original source video. Each track would then be played to completion, with the bounding box labeling occurring on the first frame with sun glare that was a multiple of 5. The process would continue with the current bounding box being saved on every subsequent multiple of 5 until the track finished playing or sun glare was no longer on the screen. When labeling traffic signs using the bounding box localization tool, you could automatically choose the classification of the traffic sign to allow for increased efficiency. After the initial labeling, the annotation tool would continue automatically annotating until the current user deemed the automatic annotation to have drifted too far. The annotator would then delete and reapply the bounding box, and continue annotating until all the tracks were processed. For each traffic sign in a track, no more than 30 annotations of that traffic sign would be saved to decrease the overexposure of that sign.

2) Bounding Box Approval and Cleanup: After a track was labeled, the annotations were reviewed and either approved or rejected. Any bounding boxes with background noise that can be removed with manual selection are removed and relabeled. After all tracks and annotations were processed, the video was exported in a single CSV file (similar to the LISA dataset) for further processing, as demonstrated by Figure 1.

After all the tracks were processed and exported by the original source video, the annotations were further processed to remove previously uncaught errors and extract statistical information from the entire dataset. Any bounding box that did not localize a sign or contained significant background noise was rejected, and any improperly labeled annotations were renamed. After removing the improper annotations, the remaining annotations were then categorized on if the traffic signs were covered in any way, and if they were on the current road or a side road [3]. These "Occluded" and "On another road" annotations [3] were then pooled.



Classes: doNotEnter (41)

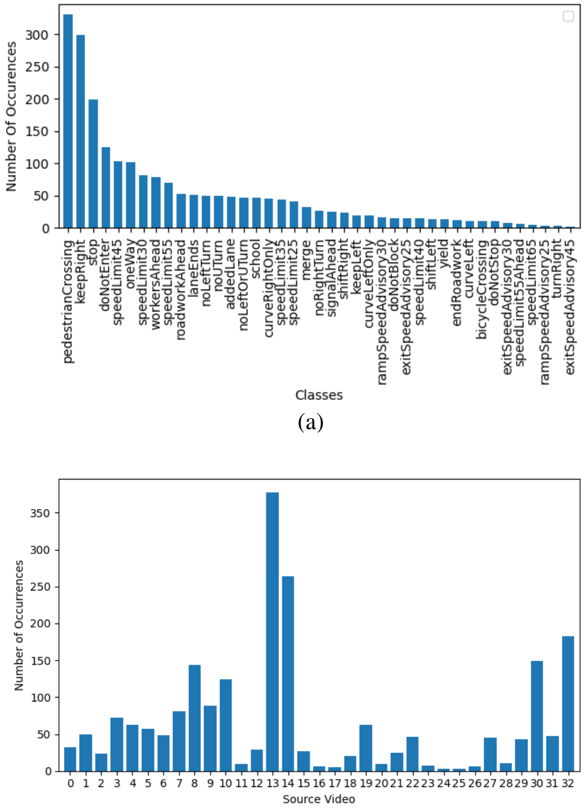
Export format: Single csv file

Fig. 1. Bounding box processing and exporting.

C. Dataset Statistics

The GLARE dataset contains 2,157 bounding box annotations and associated images distributed across 41 classes. Figure 2a shows the distributions of the annotations per class. The annotations were created from multiple videos to ensure a variety in the location in glare conditions. For each track in each source video, a maximum of 30 frames for each traffic sign class were allowed to minimize over-exposure of traffic signs in specific positions and sun glare conditions. Figure 2b shows the distribution of annotations across the 33 source videos processed. The size of the bounding box annotations varies between 6 × 14 and 137 × 178 pixels, and the size of the images is either 810 × 540 or 937 × 540 pixels. The dataset works with existing scripts released alongside the LISA dataset for annotation, extraction, and splitting [3].

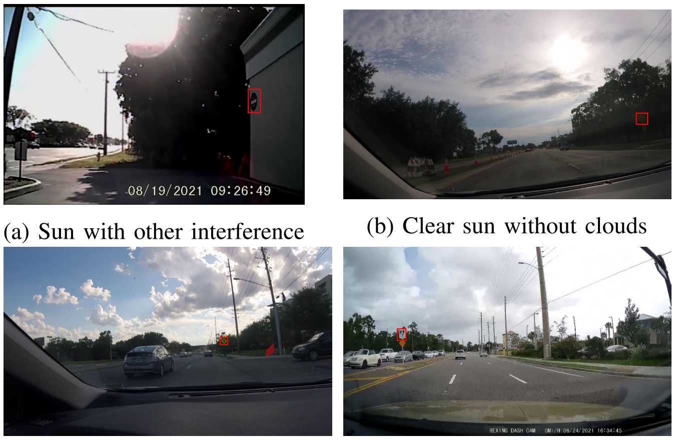
The types of visual interference labeled as sun glare can be broadly categorized into four categories. The categories described are subjective but are recorded to allow for a greater understanding of how we evaluated sun glare during the initial video processing for the dataset. Examples of each category can be seen in Figure 3. The first category is where there is a clear sun without any significant additional bright cloud noise or brightness interference from the camera. The sun appears as a bright ball, excluding any obstruction by either clouds or other objects. In an upcoming section, we will describe a naïve detector for this type of sun glare to improve traffic sign detection results. The second category is where there is a visible sun, but there are additional clouds that add to the overall brightness of the image. The third category is where there is minimal to no visible sun due to cloud interference. Although the sun is not visible in the frame, there is still visual interference that causes traffic signs to be less visible than in clear conditions, decreasing detection. The fourth category is sun glare due to other interference. The sun being visible is not a requirement, as there is visual interference due to the camera settings. Either way, the visual interference appears similar to the interference caused by the other types of sun glare. The images themselves have not been labeled based on the type



(a)

(b)

Fig. 2. Statistics/Distributions of GLARE dataset.



(a) Sun with other interference

(b) Clear sun without clouds

(c) Clouds with non-visible sun

(d) Clear sun with clouds

Fig. 3. Examples of images form the GLARE dataset with bounding boxes highlighted.

of sun glare due to the subjective nature of the categories and that some images can fit into multiple categories.

IV. Benchmarks

In order to test how sun glare conditions affect the ability to detect traffic signs, we performed multiple tests comparing how different state-of-the-art object detection architectures perform detecting traffic signs in sun glare conditions. Each architecture was tested by being trained on only the LISA dataset, on only the GLARE dataset, and on the LISA and GLARE datasets combined. The trained models were tested on similar testing sets, with the LISA-trained models tested on a subset of the testing set for the GLARE and combined models. We also performed a supplementary test using the YOLOv8 architecture for comparing how using initially random weights and using pre-trained weights affects the ability to detect traffic signs in sun glare conditions.

A. Splitting Methodology

To fairly test all the trained models and minimize similar frames in the training sets and testing set, the initial testing set was created using the frames from 5 videos in the GLARE dataset with a focus on maximizing traffic sign coverage for sign types present in multiple videos while minimizing the percentage of the GLARE dataset used. However, it was found that about half of the traffic sign types are only present in a single video. Therefore, to give coverage of these traffic sign types in the testing set while minimizing the likelihood of similar frames across the sets, the last frames for traffic sign types with less than 5 frames, in the order they appear in the original video, were selected for the testing set. This resulted in about 26.52% of the GLARE dataset being used for the testing set.

To train each architecture, 3 different training sets were created for the GLARE-only models, the LISA-only models, and the combined dataset models. For the LISA dataset models, the training set was created by selecting a subset of the LISA dataset with only the frames with traffic sign types that are present in both the LISA dataset and the GLARE dataset. Similarly, a custom testing set was also created for these models by subsetting the larger testing set with only the frames containing traffic sign types present in both GLARE and LISA. This was to not test the models on traffic sign types not present during training. For the GLARE dataset models, 3 augmented copies for each frame in GLARE not selected for testing were created with alterations using noise, color jitter, and blur, and added to the training set along with the original image. The GLARE training set was then split into 3 folds to account for the added bias of similar frames in the training set with 3 fold cross-validation performed. The resulting models were tested using the entire testing set with the scoring results averaged. For the combined models, the training sets for the LISA and GLARE models were combined, shuffled, and split into 3 folds similar to the GLARE training set for 3 fold cross-validation. Similarly, the resulting models were tested using the entire testing set with the scoring results averaged.

B. Scoring Methodology

The scoring metric used for comparing the performance of the models is the Mean Average Precision (mAP) of the classes as defined in COCO [25] and implemented in Ultralytics' YOLOv5 [26] and YOLOv8 [27] releases and OpenMMLab's mmdetection toolbox [28]. For each ground truth bounding box in an image, multiple predicted bounding boxes are produced, and the ratio between the area of the intersection between each prediction and the ground truth and the area of the union between each prediction and the ground truth is calculated. If a predicted bounding box's Intersection over Union (IoU) is equal to over a specified threshold, then the prediction is kept with the IoU used as a confidence score. The Average Precision (AP) is then calculated as the area under the precision-recall curve for the kept predictions, and the Mean Average Precision is the average of the Average Precision over all the different label types. We calculate $\mathrm{mAP}_{0.5}$ as the Mean Average Precision where the specified threshold for the IoU is 0.5. $\mathrm{mAP}_{0.5:0.95}$ is calculated as the average of the Mean Average Precision with the specified thresholds being over a range from 0.5 to 0.95 inclusive, with a step of 0.05 [25].

C. Configuration and Implementation

To demonstrate the necessity of viability of the GLARE dataset, we experimented with seven state-of-the-art methods comparing the results when training on only GLARE, only LISA, and GLARE and LISA together. We chose to test on a variety of different architecture families, with a focus on including state-of-the-art one-stage architectures commonly used in real-time object detection tasks and architectures based on the transformer architecture, which has demonstrated greatly improved results on several deep learning tasks, such as object detection. For this, we chose YOLOv5 [26], YOLOX [19], YOLOv8 [27], and TOOD [29] as our one-stage detection architectures, and chose Deformable DETR [30] and the Faster-RCNN [31] with a Swin Transformer backbone [21] as our transformer architectures. Finally, we also selected the Faster-RCNN with a ResNet50 backbone [32] as a baseline to compare the more recent and state-of-the-art architectures against.

The three YOLO architectures were chosen due to YOLOv8 and YOLOX having architecture changes to improve results compared to YOLOv5. For example, YOLOv8 and YOLOX are anchor-free architectures while YOLOv5 is anchor based, which tends to increase the number of bounding box predictions. YOLOv8 has also done other changes compared to YOLOv5, such as changing the first convolution layer in the stem from 6×6 to 3×3 kernel, reducing the CSP Bottleneck from 3 convolutions to 2, and changing the kernel size of the first convolution layer in the bottleneck to 3×3. YOLOX also differs from YOLOv5 by utilizing a decoupled head and a novel dynamic label assignment strategy named simOTA. All 3 YOLO architectures were trained using the small model size, as the GLARE dataset is designed to be used in real-time object detection. TOOD was chosen to demonstrate the performance of a one-stage detection architecture that performs object detection differently from the YOLO series, as TOOD uses a "new head structure and alignment-based learning approach" [29] to align the classification and localization tasks for object detection.

The YOLOv5 and YOLOv8 models were trained and tested on Ultralytics' releases [27], [33], and the rest of the architectures were trained and tested using OpenMMLab's MMDetection toolbox [28]. All architectures used the given training and testing pipelines for images to not bias the detection results to any training set. As MMDetection provides different configurations for the available architectures, TOOD was trained with multi-scale training, Deformable DETR used the 2-stage version with iterative box refinement, and the Faster-RCNN with a Swin Transformer backbone used multi-scale cropping.

TABLE II
Training Configuration for Each Architecture

	Batch Size	LR	Epochs
Faster-RCNN$_{ResNet50}$	16	0.02	24
YOLOv5	32	0.01	150
Deformable DETR	8	5e-5	50
Faster-RCNN$_{SwinT\text{-}Base}$	16	1e-4	36
YOLOX	32	5e-3	150
TOOD	8	5e-3	24
YOLOv8	32	0.01	150

The Faster-RCNN, Deformable DETR, and TOOD architectures used the ResNet50 backbone with pre-trained weights trained on the COCO dataset [25], and the Faster-RCNN with a Swin Transformer backbone used pre-trained weights trained from the ImageNet-1k dataset [34] for the backbone. All the YOLO architectures were trained with completely random weights in the backbone. To account for the possible decrease in performance compared to the other architectures with pre-trained weights, we performed a supplementary test comparing the performance of YOLOv8 when trained using random weights and using weights pre-trained on the COCO dataset, which the results are described in the following section.

The YOLOv5 models were trained on an RTX 3070, the YOLOv8 and TOOD models trained on LISA and GLARE were trained on an RTX 3090, and the rest of the models were trained on two NVIDIA Tesla V100 GPUs. For all the architectures, we used the default hyperparameters provided, with alterations to the training batch size, initial learning rate, and the number of epochs to fit our datasets. We only altered the learning schedule for the Faster-RCNN with a ResNet50 backbone, where we did not have any warm-up epochs. For the architectures trained using MMDetection, the auto-scale-learning-rate flag was used to automatically adjust the learning rate to the batch size. The architecture training configurations are shown in Table II.

D. Benchmark Results

The results of testing the benchmark architectures trained on the GLARE, LISA, and combined training sets are shown in Tables III and IV.[1] The results of testing the YOLOv8 architecture on random weights and pre-trained weights from training on the COCO dataset are shown in Table V.

For the models trained on the GLARE dataset, the average $\mathrm{mAP}_{0.5}$ is 59.2, and the average $\mathrm{mAP}_{0.5:0.95}$ is 39.6. For the models trained on the LISA dataset, the average $\mathrm{mAP}_{0.5}$ is 34.7, and the average $\mathrm{mAP}_{0.5:0.95}$ is 19.4. For the models trained on the combined GLARE and LISA datasets,

[1] The architectures are listed by the initial release of the associated publication or code itself if no publication is available.

TABLE III

$\text{mAP}_{0.5}$ Scoring Results After Training With Random Initial Weights

	GLARE	LISA	Combined
Faster-RCNN$_{\textit{ResNet50}}$	55.1	35.3	68.3
YOLOv5	54.8	26.9	62.6
Deformable DETR	62.0	42.1	71.6
Faster-RCNN$_{\textit{SwinT-Base}}$	**68.7**	39.3	**73.9**
YOLOX	51.5	17.0	60.7
TOOD	63.5	**57.4**	73.6
YOLOv8	58.7	24.8	64.6

TABLE IV

$\text{mAP}_{0.5:0.95}$ Scoring Results After Training With Random Initial Weights

	GLARE	LISA	Combined
Faster-RCNN$_{\textit{ResNet50}}$	35.8	19.7	44.1
YOLOv5	37.5	14.9	42.4
Deformable DETR	40.8	24.1	46.8
Faster-RCNN$_{\textit{SwinT-Base}}$	**45.2**	20.6	48.7
YOLOX	34.6	10.0	41.1
TOOD	42.6	**31.2**	**49.8**
YOLOv8	41.0	15.0	43.9

the average $\text{mAP}_{0.5}$ is 67.9, and the average $\text{mAP}_{0.5:0.95}$ is 42.3. The difference between the GLARE-trained models and the LISA-trained models is 24.5 for $\text{mAP}_{0.5}$ and 20.2 for $\text{mAP}_{0.5:0.95}$, and the difference between combined-dataset-trained models and GLARE-trained models is 8.7 for $\text{mAP}_{0.5}$ and 2.7 for $\text{mAP}_{0.5:0.95}$.

The best-performing architectures overall were the Faster-RCNN with a Swin Transformer backbone and TOOD, with the Swin architecture performing best when trained on only GLARE for both $\text{mAP}_{0.5}$ and $\text{mAP}_{0.5:0.95}$, TOOD performed best when trained on only LISA for both $\text{mAP}_{0.5}$ and $\text{mAP}_{0.5:0.95}$, and among the models trained on the combined dataset, the Swin architecture performed best on $\text{mAP}_{0.5}$ and TOOD performed best on $\text{mAP}_{0.5:0.95}$. Among transformer-based models, the Swin architecture performed best when trained on only GLARE or the combined dataset, and Deformable DETR performed best when trained on only the LISA dataset. For single-stage architectures, TOOD outperformed every architecture in the YOLO family for GLARE-trained, LISA-trained, and combined-dataset-trained models. Compared to the Faster-RCNN with a ResNet50 backbone, YOLOv5 and YOLOX performed worse when trained on only the GLARE dataset, and all architectures in the YOLO family performed worse when trained on only the LISA dataset and when trained on the combined dataset.

These results indicate that sun glare has a noticeable effect on the ability of object detection architecture to detect traffic signs in sun glare conditions, especially for the YOLO family of architectures, which are frequently used in real-time object detection tasks. When the architectures were trained on LISA alone, there was a significant decrease in performance compared to when they are trained on GLARE alone or GLARE and LISA together. Our results also validate the GLARE dataset as a useful extension of the LISA dataset, as training using the LISA and GLARE datasets together resulted in superior performance compared to both the architectures trained only on GLARE and the architectures trained only on LISA. The unusually high performance of the TOOD architecture, especially when trained on only the LISA dataset, warrants future investigation on the ability of architectures with similar structures to detect traffic signs in sun glare conditions when trained on datasets without a significant presence of sun glare.

TABLE V

Scoring Results After Training YOLOv8 With Random Weights and Pre-Trained Weights

	$\text{mAP}_{0.5}$		$\text{mAP}_{0.5:0.95}$	
	Random	COCO	Random	COCO
GLARE	58.7	**61.7**	41.0	**44.4**
LISA	**24.7**	22.3	**15.0**	12.6
Combined	64.6	**66.0**	43.9	**45.8**

For the test using YOLOv8 comparing using random weights and COCO pre-trained weights, using pre-trained weights led to an increase in performance by 3 in $\text{mAP}_{0.5}$ and 3.4 in $\text{mAP}_{0.5:0.95}$ for the models trained on the GLARE dataset and by 1.4 in $\text{mAP}_{0.5}$ and 1.9 in $\text{mAP}_{0.5}$ for the models trained on the combined dataset. For the models trained on the LISA dataset, training using pre-trained weights led to a decrease in performance by 2.4 in $\text{mAP}_{0.5}$ and $\text{mAP}_{0.5:0.95}$. A reasonable explanation for these results is that the pre-trained weights allow for YOLOv8, and by extension possibly other architectures in the YOLO family and similar architectures, to better fit the training set. This would lead to the models trained on LISA having more difficulty detecting traffic signs in sun glare conditions, while models trained with at least part of the data containing traffic signs in sun glare lead to increased performance, further indicating the usefulness of the GLARE dataset in traffic sign detection.

V. Conclusion and Future Works

This paper introduces GLARE, a traffic sign dataset with a focus on sun glare and how it affects the recognition of traffic signs in such conditions. The dataset includes 2,157 images with corresponding bounding box annotations of traffic signs across 41 classes from the Orlando area. The GLARE dataset has a specific focus on images with sun glare present, which affects both human drivers and cameras for autonomous driving systems. Our baseline benchmarks have shown that sun glare has a noticeable effect on the ability of current architectures to detect traffic signs.

The GLARE dataset is the beginning of future research on traffic sign detection in naturally noisy conditions and the removal of sun glare as well. We believe this dataset can be used as a testing set for entire sun glare removal using the

U-Net architecture [35], as seen in previous work removing sun flares from images [36]. We also believe this dataset can be extended to include traffic signs in other noisy or abnormal conditions, such as rain, fog, and night-time driving. Such an extension could be used to create and train architectures that can detect traffic signs with greater precision in a wider variety of conditions, whether through image restoration or detection and recognition alone. Finally, we believe this dataset can be used to assist in the development of object detection architectures that can detect objects well despite localized noise, as possibly evidenced by TOOD.

Acknowledgment

The authors would like to thank Sudharsan Vaidhun's contributions to the raw footage.

References

[1] *Traffic Safety Facts: Crash Stats—Sun Glare and Slick Roads Are the Two Most Environment-Related Causing Circumstances*, National Highway Traffic Safety Administration U.S. Department of Transportation, Washington, DC, USA, 2015.

[2] J. Bian et al., "Machine learning in real-time Internet of Things (IoT) systems: A survey," *IEEE Internet Things J.*, vol. 9, no. 11, pp. 8364–8386, Jun. 2022.

[3] A. Mogelmose, M. M. Trivedi, and T. B. Moeslund, "Vision-based traffic sign detection and analysis for intelligent driver assistance systems: Perspectives and survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 4, pp. 1484–1497, Dec. 2012.

[4] F. Larsson, M. Felsberg, and P. E. Forssen, "Correlating Fourier descriptors of local patches for road sign recognition," *IET Comput. Vis.*, vol. 5, no. 4, pp. 244–254, Jul. 2011.

[5] J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel, "Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition," *Neural Netw.*, vol. 32, pp. 323–332, Aug. 2012.

[6] S. Houben, J. Stallkamp, J. Salmen, M. Schlipsing, and C. Igel, "Detection of traffic signs in real-world images: The German traffic sign detection benchmark," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Aug. 2013, pp. 1–8.

[7] Z. Zhu, D. Liang, S. Zhang, X. Huang, B. Li, and S. Hu, "Traffic-sign detection and classification in the wild," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2110–2118.

[8] C. Ertler, J. Mislej, T. Ollmann, L. Porzi, G. Neuhold, and Y. Kuang, "The mapillary traffic sign dataset for detection and classification on a global scale," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2020, pp. 68–84.

[9] D. Tabernik and D. Skocaj, "Deep learning for large-scale traffic-sign detection and recognition," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 4, pp. 1427–1440, Apr. 2020.

[10] R. Timofte, K. Zimmermann, and L. Van Gool, "Multi-view traffic sign detection, recognition, and 3D localisation," *Mach. Vis. Appl.*, vol. 25, no. 3, pp. 633–647, Apr. 2014.

[11] D. Temel, M. Chen, and G. AlRegib, "Traffic sign detection under challenging conditions: A deeper look into performance variations and spectral characteristics," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 9, pp. 3663–3673, Sep. 2020.

[12] D. Soendoro and I. Supriana, "Traffic sign recognition with color-based method, shape-arc estimation and SVM," in *Proc. Int. Conf. Electr. Eng. Informat.*, Jul. 2011, pp. 1–6.

[13] X. Mao, S. Hijazi, R. Casas, P. Kaul, R. Kumar, and C. Rowen, "Hierarchical CNN for traffic sign recognition," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2016, pp. 130–135.

[14] Á. Arcos-García, J. A. Álvarez-García, and L. M. Soria-Morillo, "Deep neural network for traffic sign recognition systems: An analysis of spatial transformers and stochastic optimisation methods," *Neural Netw.*, vol. 99, pp. 158–165, Mar. 2018.

[15] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 25, 2012, pp. 1–12.

[16] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587.

[17] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.

[18] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.

[19] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "YOLOX: Exceeding YOLO series in 2021," 2021, *arXiv:2107.08430*.

[20] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2020, pp. 213–229.

[21] Z. Liu et al., "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 9992–10002.

[22] N. D. Bruijn. (2019). *OpenCV-Video-Label: An OpenCV Video Player Which Allows the User to Easily Generate Labeled Images From videos*. [Online]. Available: https://github.com/natdebru/opencv-video-label

[23] D. Gordon, A. Farhadi, and D. Fox, "Re^3: Real-time recurrent regression networks for visual tracking of generic objects," *IEEE Robot. Autom. Lett.*, vol. 3, no. 2, pp. 788–795, Apr. 2018.

[24] G. Nebehay and R. Pflugfelder, "Clustering of static-adaptive correspondences for deformable object tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 2784–2791.

[25] T.-Y. Lin et al., "Microsoft COCO: Common objects in context," in *Proc. ECCV*, Sep. 2014, pp. 1–15.

[26] G. Jocher et al. (Feb. 2022). *TensorRT, TensorFlow Edge TPU and OpenVINO Export and Inference*. [Online]. Available: http://ultralytics/yolov5

[27] G. Jocher, A. Chaurasia, and J. Qiu, "YOLO by ultralytics," Tech. Rep., 2023. [Online]. Available: https://github.com/ultralytics/ultralytics

[28] K. Chen et al., "MMDetection: Open MMLab detection toolbox and benchmark," 2019, *arXiv:1906.07155*.

[29] C. Feng, Y. Zhong, Y. Gao, M. R. Scott, and W. Huang, "TOOD: Task-aligned one-stage object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 3490–3499.

[30] X. Zhu, W. Su, L. Lu, B. Li, X. Wang, and J. Dai, "Deformable DETR: Deformable transformers for end-to-end object detection," in *Proc. Int. Conf. Learn. Represent.*, 2021, pp. 1–15.

[31] 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, Jun. 2017.

[32] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 770–778.

[33] G. Jocher, "YOLOv5 by ultralytics," Tech. Rep., 2020. [Online]. Available: https://github.com/ultralytics/yolov5

[34] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.

[35] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2015, pp. 234–241.

[36] Y. Wu et al., "How to train neural networks for flare removal," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 2219–2227.