

Citation	D. Temel, M-H. Chen, and G. AlRegib, "Traffic Sign Detection under Challenging Conditions: A Deeper Look Into Performance Variations and Spectral Characteristics," IEEE Transactions on Intelligent Transportation Systems, 2019.
Dataset	https://github.com/olivesgatech/CURE-TSD
Bib	@INPROCEEDINGS{Temel2019_ITS, author={D. Temel and M.-H. Chen and G. AlRegib}, booktitle={IEEE Transactions on Intelligent Transportation Systems}, title={Traffic Sign Detection under Challenging Conditions: A Deeper Look Into Performance Variations and Spectral Characteristics}, year={2019}, }
DOI	10.1109/TITS.2019.2931429
ISSN	1524-9050
Contact	alregib@gatech.edu https://ghassanalregib.info/ dcantemel@gmail.com http://cantemel.com/

Traffic Sign Detection under Challenging Conditions: A Deeper Look Into Performance Variations and Spectral Characteristics

Dogancan Temel, Min-Hung Chen, and Ghassan AlRegib



Fig. 1: Challenging scene examples from the CURE-TSD-Real dataset: dirty lens, haze, snow, blur, and rain, respectively.

Abstract—Traffic signs are critical for maintaining the safety and efficiency of our roads. Therefore, we need to carefully assess the capabilities and limitations of automated traffic sign detection systems. Existing traffic sign datasets are limited in terms of type and severity of challenging conditions. Metadata corresponding to these conditions are unavailable and it is not possible to investigate the effect of a single factor because of simultaneous changes in numerous conditions. To overcome the shortcomings in existing datasets, we introduced the CURE-TSD-Real dataset, which is based on simulated challenging conditions that correspond to adversaries that can occur in real-world environments and systems. We test the performance of two benchmark algorithms and show that severe conditions can result in an average performance degradation of 29% in precision and 68% in recall. We investigate the effect of challenging conditions through spectral analysis and show that challenging conditions can lead to distinct magnitude spectrum characteristics. Moreover, we show that mean magnitude spectrum of changes in video sequences under challenging conditions can be an indicator of detection performance. CURE-TSD-Real dataset is available online at <https://github.com/olivesgatech/CURE-TSD>.

Keywords—Autonomous vehicles, traffic sign detection and recognition, challenging conditions dataset, magnitude spectrum, machine learning

I. INTRODUCTION

Transportation systems are transformed by disruptive technologies that are based on autonomous systems. In order for autonomous systems to seamlessly operate in real-world conditions, they need to be robust under challenging conditions. In this study, we focus on automated traffic sign detection systems and investigate the effect of challenging conditions in algorithmic performance. Currently, the performance of traffic sign detection algorithms are tested with existing traffic sign datasets in the literature [1–12], which have been very useful to develop and evaluate state-of-the-art traffic sign recognition and detection algorithms. However, these datasets are usually very limited in terms of type and severity of challenging

conditions. Moreover, there is no metadata corresponding to the types and the levels of these conditions. Traffic sign size can be considered as the only metadata corresponding to the challenge level, which can degrade the detection performance significantly [13]. Besides the limitations of challenging conditions, it is not feasible to assess the effect of most of the challenging conditions in existing datasets because of limited control over the acquisition process. A number of conditions change simultaneously, which makes it impossible to investigate the effect of a specific condition. To overcome the shortcomings of existing datasets and enable assessing the effect of challenging conditions one at a time, we introduced traffic sign recognition [14] and detection datasets [15, 16]. We hosted the first IEEE Video and Image Processing (VIP) Cup [15] within the IEEE Signal Processing Society and obtained an algorithmic benchmark for the CURE-TSD dataset, which is based on video sequences corresponding to real-world as well as simulator environments. In this study, we focus on the real-world sequences denoted as CURE-TSD-Real dataset. Specifically, we investigate the effect of challenging conditions over the average performance of top two benchmark algorithms that are based on deep neural networks.

Recently, adversarial examples have been commonly used in the literature to test the vulnerability of recognition and detection systems [17, 18]. Even though adversarial examples have been successful in deceiving these generic object recognition and detection systems, they have not been effective in traffic sign recognition and detection systems [19, 20]. Lu *et al.* [21] showed that adversarial examples deceive traffic sign detection systems only in limited scenarios and Das *et al.* [20] showed that a simple compression stage can minimize the effect of adversarial attacks in traffic sign recognition. Adversarial examples are useful to assess the limits of existing systems with special inputs that are optimized for deception. However, such adversarial examples do not necessarily correspond to real-world challenging conditions. Moreover, previous studies mainly focused on feeding adversarial data directly into the classification model. However, in real world, challenging conditions can directly affect the data acquired by sensors rather than classifiers. In this study, we differentiate

Authors are with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, 30332-0250 USA. Corresponding author email: cantemel@gatech.edu.

from previous studies by focusing on challenging conditions corresponding to adversaries that can naturally occur in real-world environments and systems as shown in Fig. 1. On contrary to adversarial studies in the literature that require model information to design input images, we designed challenging conditions independent of the detection algorithms, which enables a black-box performance assessment. Previously, we performed black-box assessment of object detection APIs with realistic challenging conditions in [22, 23], and investigated the robustness and out-of-distribution classification performance of traffic sign classifiers in [24, 25].

In addition to investigating the effect of challenging conditions in traffic sign detection performance, we also analyze the effect of challenging conditions in terms of spectral characteristics. In [26], Van der Schaaf and Van Hateren analyzed the power spectrum of natural images and showed that there is a common characteristic followed by natural images. In [27], Torralba and Olivia extracted more information based on spectrum related to the openness of images, the semantic category of scenes, the recognition of objects, and the depth of scenes. Based on these observations and findings, a direct comparison between spectrum of challenge-free sequences and challenging sequences can be affected by the context of the scene. However, if we first obtain the difference between challenge-free and challenging sequences and then obtain the power spectrum, we can limit the effect of context and concentrate on the change with respect to reference conditions, which is the methodology pursued in this study.

The rest of this paper is organized as follows: In Section II, we analyze existing traffic sign datasets. In Section III, we provide a general description of the CURE-TSD-Real dataset, describe challenging conditions, and briefly introduce benchmark algorithms. We discuss traffic sign detection performance under challenging conditions and analyze spectral characteristics of these conditions in Section IV. Specifically, we explain the performance metrics in Section IV-A, describe the training and test datasets in Section IV-B, report performance variation under challenging conditions in Section IV-C, perform a spectral analysis of challenging conditions in Section IV-D, and analyze the relationship between detection performance and spectral characteristics in Section IV-E. Finally, we conclude our work in Section V.

II. EXISTING DATASETS

We summarize the main characteristics of existing traffic sign datasets in Table I, which includes number of video sequences, number of annotated frames/images, number of annotated signs, annotation information, resolution, number of sign types, annotated sign size, acquisition location, and challenging conditions. When a category does not apply to a specific dataset, there is a ‘not applicable’ abbreviation (N/A). We report the characteristics of publicly available datasets based on the reference papers as well as available dataset files. The majority of listed datasets are based on images whereas LISA provides short tracks up to 30 frames per track and BelgiumTS provides 4 videos whose number of frames varies between 2,001 and 6,001. Total number of annotated

images varies from 273 to 100,000 in which number of traffic signs is in between 273 and 51,840. Annotations are mainly based on sign types and bounding box coordinates but 3D location, pixel map, visibility status, occlusion condition, and road status are also provided in certain datasets. The resolution of the traffic sign detection datasets is in between 360x270 and 2,048x2,048 and the sign size in all listed datasets vary from 2x7 to 573x557. The number of traffic sign types vary from 3 to 62. Challenging conditions are not annotated or explicitly described in majority of the datasets. Therefore, we visually inspected these datasets to list apparent challenging conditions, which include illumination, occlusion, shadow, blur, reflection, codec error, dirty lens, overcast, haze, and deformation of the traffic signs. The majority of the listed datasets were captured in Europe but recent studies also include China and USA. All of the datasets include images captured with color cameras but LISA dataset also includes grayscale images directly obtained from car cameras.



Fig. 2: Traffic signs types in the CURE-TSD-Real dataset.

III. CURE-TSD-REAL DATASET

A. General Information

Among the datasets analyzed in Section II, BelgiumTS [2] and LISA [9] are the only datasets that provide partial video sequences. When this study was conducted, tracks in the LISA dataset were available but not the video sequences. Therefore, we utilized the BelgiumTS [2] dataset to obtain video sequences. We selected a subset of the traffic signs in the BelgiumTS dataset as shown in Fig. 2 and labeled consecutive frames only for these sign types. Sign types were selected according to the compatibility with the synthesized part of the CURE-TSD dataset, which is not considered in this paper. In total, there were four main sequences in BelgiumTS, which included 3,001, 6,201, 2,001, and 4,001 frames. We grouped 300 consecutive frames as individual videos and obtained a total of 49 videos. In the BelgiumTS dataset, annotations were provided for specific frames and annotating the missing frames

TABLE I: Main characteristics of publicly available datasets and CURE-TSD-Real dataset*

	Number of videos	Number of annotated images	Number of annotated signs	Annotation information	Resolution	Number of sign types	Annotated sign size	Acquisition location	Challenging conditions (*=annotated)
RUG [1]	N/A	N/A	N/A	N/A	360X270	3	N/A	Netherlands	illumination overcast blur
BelgiumTS [2]	4	9,006	13,444	sign type bounding box 3D location	1,628x1,236	62	9x10 to 206x277	Belgium	illumination occlusion overcast
BelgiumTSC [3]	N/A	7,125	7,125	sign type bounding box	22x21 to 674x527	62	11x10 to 562x438	Belgium	illumination occlusion deformation
Stereopolis [4]	N/A	273	273	sign type bounding box	1,920x1,080	10	25x25 to 204x159	France	illumination occlusion overcast
STS [5]	N/A	3,488	3,488	sign type bounding box visibility status road status	1,280x960	7	3x5 to 263x248	Sweden	illumination occlusion* shadow, blur* overcast, rain
GTSRB [6, 7]	N/A	51,840	51,840	sign type	15x15 to 250x250	43	15x15 to 250x250	Germany	illumination occlusion, blur deformation perspective
LISA [9]	17 tracks	6,610	7,855	sign type bounding box occlusion status road status	640x480 to 1,024x522	47	6x6 to 167x168	USA	illumination occlusion* shadow, blur reflection codec error dirty lens decolorization* overcast
GTSDB [8]	N/A	900	1,206	sign type bounding box	1,360X800	43	16-128 longer edge	Germany	illumination occlusion blur, shadow haze, rain overcast
TT-100K [10]	N/A	100,000	30,000	sign type bounding box pixel map	2,048x2,048	45	2x7 to 397x394	China	illumination occlusion overcast haze, shadow
CTSD [11]	N/A	1,100	1,574	sign type bounding box	1,024x768 & 1,270x800	48	26x26 to 573x557	China	illumination occlusion shadow, rain, overcast dirty lens reflection haze, blur
CCTSDB [12]	N/A	10,000	13,361	sign category bounding box	1,024x768 & 1,270x800 & 1,000x350	3 classes	10x11 to 380x378	China	illumination occlusion shadow, rain overcast dirty lens reflection haze, blur
CURE-TSD-Real	2,989	896,700	648,186	sign type bounding box challenge type challenge level	1,628x1,236	14	10x11 to 206x277	Belgium	rain*, snow* shadow*, haze* illumination* decolorization* blur*, noise* codec error* dirty lens* occlusion overcast

* Online sources of the datasets are hyperlinked to the dataset names in the first column of this table.

could have resulted in an inconsistency among the labels. Therefore, we annotated all the frames including the ones with existing labels and extended the number of annotated frames from 9,006 to 14,700 using the Video Annotation Tool from Irvine, California (VATIC) [28]. Specifically, we utilized the JavaScript version on the browser and labeled a traffic sign if half of the sign was visible. We considered the original video sequences as baseline and added challenging conditions at different levels to test the performance limits of traffic sign detection algorithms.

B. Challenging Conditions

We processed original video sequences with 12 challenge types to obtain challenging video sequences as illustrated in Fig. 3. Postproduction of challenging conditions scaled up the dataset size from 14,700 images to 896,700 images. We adjusted the level of challenging conditions through visual inspection rather than numerical progression. Challenge-free sequences were considered as level 0 and we added five different levels for each challenge type. Levels were adjusted according to the following rules: *level 1 does not affect the visibility of traffic signs perceptually, level 2 affects the visibility of small and distant traffic signs, level 3 makes the visibility of small and distant traffic signs difficult, level 4 makes the visibility of small and distant traffic signs challenging, and level 5 makes the visibility of small and distant traffic signs nearly impossible.* Simulated conditions and implementation details are listed as follows:

- *Decolorization* tests the effect of color acquisition error, which was implemented using Black & White color correction filter version 1.0. The filter settings were: Reds= 40, Yellows= 60, Greens= 40, Cyans= 60, Blues= 20, and Magentas= 80. We utilized multiple adjustment layers to compound the effect of the color correction filter and created multiple distinct levels.
- *Lens/Gaussian blur* tests the effect of dynamic scene acquisition. *Lens blur* was implemented with the Camera Lens Blur filter version 1.0 whose radius was set to 2, 4, 6, 8, and 10 along with Hexagan Iris Shape. For the *Gaussian blur* challenge, Gaussian Blur filter version 3.0 was used whose Bluriness levels were set to 5, 10, 15, 20, and 25. On contrary to lens blur, Gaussian blur is distributed in all directions, which leads to less structured blurred objects.
- *Codec error* tests the effect of encoder/decoder error, which was implemented using Time Displacement filter version 1.6. Max Displacement Time was set to 0.1, 0.2, 0.3, 0.4, and 0.5.
- *Darkening (underexposure)* tests the effect of underexposure, which was implemented using Exposure filter version 1.0. The master channel Exposure parameter was set to $-1, -3, -5, -7$, and -9 .
- *Dirty lens* tests the effect of occlusion because of dirt over camera lens, which was implemented by overlaying a set of dusty lens images.
- *Exposure (overexposure)* tests the effect of overexposure in acquisition, which was implemented with

the Exposure filter version 1.0. The master channel Exposure parameter was set to 1, 3, 5, 7, and 9.

- *Noise* tests the effect of acquisition noise, which was implemented using the Noise filter version 2.6. The Amount of Noise parameter was set to 20, 40, 60, 70, and 71.
- *Rain* tests the effect of occlusion due to rain, which was implemented using the Gradient Ramp generator version 3.2 with colors # 0F1E2D and # 5A7492 to create a blueish hue over the video, and CC Rainfall generator from Cycore Effects HD 1.8.2 version 1.1. The Opacity level was set to 25% with 5 adjustment layers and the Drops option was set to 10000, 20000, 50000, and 100000.
- *Shadow (occlusion)* tests the effect of non-uniform lighting due to shadow. Based on the description of Merriam Webster [29], shadow refers to *partial darkness or obscurity within a part of space from which rays from a source of light are cut off by an interposed opaque body*. In this study, darkness/obscurity refers to the vertical patterns and space refers to the traffic sign. In Fig. 3, we can observe that shadow partially occludes the traffic sign and the levels corresponds to the darkness of the occluded region. The *shadow* condition was implemented using Venetian Blinds filter version 2.3. Transition Completeness was 47%, Direction was $0x + 0.0^\circ$, and Width was 142. Finally, Opacity was set to 15%, 30%, 45%, 60%, and 75%.
- *Snow* tests the effect of occlusion due to snow, which was implemented using the Glow filter version 2.6 with color # FFFFFF to create a white hue over the video, and CC Snowfall generator from Cycore Effects HD 1.8.2 version 1.1. Glow Threshold was 55%, Glow Intensity was 1.4, Glow Operation was Screen, and Glow Dimension was Horizontal. Drops option in the CC Snowfall generator was set to 10000, 50000, 100000, and 140000 using 9 adjustment layers.
- *Haze* tests the effect of occlusion due to haze, which was implemented using the Ellipse Shape Layer filter version 1.0 with radial gradient fill using color # D6D6D6 in the center with 100% opacity and color # 000000 at the edges with 0% opacity. The shape and focal point location of the ellipse was manually controlled to closely follow the furthest point in the video, which created a sense of depth to the scene and emulated the behaviour of haze. In addition to Ellipse filter, Smart Blur version 1.0, Exposure version 1.0, and Brightness & Contrast version 1.0 were utilized. In the Smart Blur filter, Radius was set to 3, and Threshold was set to 25. In the Exposure filter, Radius was -1 and Gamma Correction was 1. In the Brightness & Contrast filter, Brightness was set to -34 and Contrast was set to -13 . Solid Layer was used with a color code of # CECECE and opacity was set to 10%, 20%, 30%, 40%, and 50%.

Challenge types were selected and synthesized based on

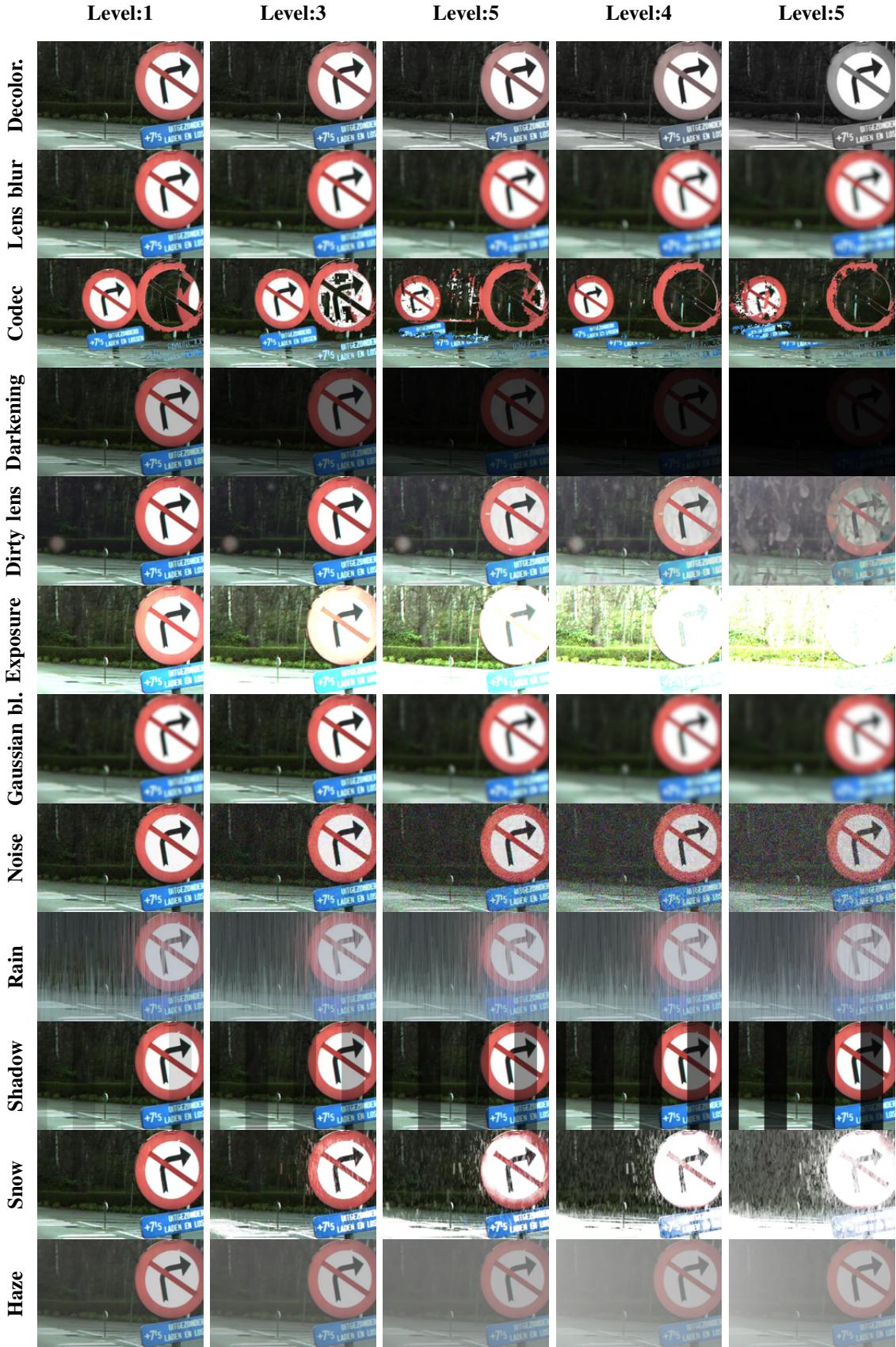


Fig. 3: All challenge types and levels corresponding to a sample frame in the CURE-TSD-Real video sequences.

the discussion with the Multimedia Signal Processing Technical Committee and IEEE Signal Processing Society during the VIP Cup 2017 competition process. In the original competition proposal, we proposed challenging conditions including *blur*, *exposure*, *rain*, and *snow*. Based on the recommendations of the committee members and follow-up discussions, we added the remaining challenging conditions including *decolorization*, *codec error*, *dirty lens*, *noise*, *shadow*, and *haze*. All of the challenging conditions other than *snow* were observed in the prior real-world traffic datasets [1–12] as summarized in Table I, which indicates the relevance and significance of these conditions. To simulate challenging conditions, we utilized the state-of-the-art visual effects and motion graphics software Adobe(c) After Effects, which has been commonly used for realistic image and video processing and synthesis in the literature [30–32]. We provided the details of challenge generation process including the operators and parameters so that challenging condition synthesis can be reproducible, and researchers can build on top of the initial configurations. Challenging conditions do not have to be used all together and researchers can select the challenging conditions that are relevant and sufficient for their target application. This dataset can be considered as an initial step to assess the robustness of traffic sign detection algorithms under controlled challenging conditions, which can be further enhanced by the research community.

C. Benchmark Algorithms

In this study, we analyze the average performance of two top performing algorithms in the VIP Cup traffic sign detection challenge [15, 16]. Both of the algorithms are based on state-of-the-art convolutional neural networks (CNNs) including U-Net [33], ResNet [34], VGG[35], and GoogLeNet [36]. In both algorithms, localization and recognition of the traffic signs are performed by separate CNNs. Details of the finalist algorithms are summarized as follows: First algorithm includes a VGG-based challenge type detection stage followed by a histogram equalization and a ResNet-based denoising depending on the challenge type. Traffic signs are localized by a U-Net architecture and recognized by a custom CNN architecture. Second algorithm extracts features with a pretrained GoogLeNet architecture whose features at the end of Inception 5B layer are fed into the regression layer to obtain sign locations, which are further classified by a custom CNN architecture.

TABLE II: Detection performance metrics. β (0.5 or 2.0) is used to adjust the relative importance of *precision* and *recall*.

Term	Description/Formulation
Positive (P)	Total number of traffic signs
True positive (TP)	Total number of correct traffic sign detections
False positive (FP)	Total number of false traffic sign detections
False negative(FN)	Number of undetected traffic signs
Precision (prec)	$TP/(TP + FP)$
Recall (rec)	$TP/(TP + FN)$
F_β score	$(1 + \beta^2)(prec \cdot rec)/(\beta^2 \cdot prec + rec)$

IV. TRAFFIC SIGN DETECTION UNDER CHALLENGING CONDITIONS

A. Performance Metrics

We calculate precision, recall, $F_{0.5}$, and F_2 metrics to measure the traffic sign detection performance as described in Table II. A detection is considered correct if intersection over union (IoU), also known as Jaccard index, is at least 0.5. IoU is obtained by calculating the overlapping area between the ground truth bounding box and the estimated bounding box, and diving the overlapping area by the area of the union between these boxes.

B. Training and Test Sets

There are 49 reference video sequences with 300 frames per video as described in Section III-A. Video sequences were split into training and test sequences by following 7:3 splitting ratio, which led to 34 training video sequences and 15 test video sequences. For each reference video sequence, we generated 60 challenging sequences based on 12 challenge types and 5 challenge levels. For each challenge type, there are 170 (34x5) video sequences (51,000 frames) in the training set and 75 (15x5) video sequences (22,500 frames) in the test set. We set the number of video sequences same for each challenge type and level to eliminate any bias towards a specific challenge type or level. Overall, there are 2,074 video sequences (622,200 frames) in the training set and 915 video sequences (274,500 frames) in the test set.

C. Performance Variation under Challenging Conditions

As reference performance, we calculate the average detection performance of benchmarked algorithms for challenge-free sequences. Then, we calculate the detection performance under varying challenging conditions and levels. Average detection performance under different challenge levels are reported in Fig. 4. Y-axis corresponds to detection performance and x-axis corresponds to challenge levels. In addition to reporting average detection performance for each challenge level, we also report the percentage performance degradation under severe challenging conditions (level 5). Based on the results in Fig. 4, detection performance significantly decreases with respect to reference challenge-free conditions. Specifically, detection performance degrades by 20% in precision, 65% in recall, 38% in $F_{0.5}$ score and 61% in F_2 score.

To understand the effect of challenging condition types over traffic sign detection, we report the performance of top-I team, top-II team, and their average (top-I-II) in Table III. The detection performance of top-I team degrades by 55% in precision, 80% in recall, 70% in $F_{0.5}$ score, and 78% in F_2 score. For top-II team, the performance degradation is 18% in precision, 22% in recall, 21% in $F_{0.5}$ score, and 22% in F_2 score. We can observe that overall performance degradation for each team vary between 18% and 80% in which the variation is higher for top-I team. When team results are averaged (Top-I-II), overall performance degrades by 29% in precision, 68% in recall, 48% in $F_{0.5}$ score, and 65% in F_2 score. Precision under *noise* challenge is the only category in which performance

TABLE III: Detection performance of top-I, top-II, and average top-I-II algorithms under challenging conditions for each challenge type and performance metric. First row reports the detection performance over reference video sequences without simulated challenging conditions. For each challenge type, we report the detection performance in the corresponding first row and percentage performance degradation (\downarrow) with respect to the challenge-free conditions in the corresponding second row. At the last two rows, we report the average detection performances and performance degradations (\downarrow) with bold font over all challenge types.

Algorithms	Top-I				Top-II				Average Top-I-II			
	Precision	Recall	$F_{0.5}$	F_2	Precision	Recall	$F_{0.5}$	F_2	Precision	Recall	$F_{0.5}$	F_2
-	0.35	0.29	0.34	0.30	0.65	0.07	0.25	0.09	0.50	0.18	0.30	0.20
Decolorization	0.00	0.00	0.00	0.00	0.67	0.07	0.24	0.08	0.34	0.03	0.12	0.04
	$\downarrow 100$	$\downarrow 100$	$\downarrow 100$	$\downarrow 100$	4	$\downarrow 10$	$\downarrow 6$	$\downarrow 10$	$\downarrow 32$	$\downarrow 82$	$\downarrow 60$	$\downarrow 79$
Lens blur	0.19	0.12	0.17	0.13	0.54	0.08	0.25	0.10	0.44	0.10	0.22	0.12
	$\downarrow 45$	$\downarrow 59$	$\downarrow 49$	$\downarrow 57$	$\downarrow 17$	11	1	10	$\downarrow 12$	$\downarrow 45$	$\downarrow 24$	$\downarrow 41$
Codec error	0.04	0.02	0.03	0.02	0.17	0.01	0.04	0.01	0.11	0.01	0.03	0.01
	$\downarrow 89$	$\downarrow 95$	$\downarrow 91$	$\downarrow 94$	$\downarrow 73$	$\downarrow 89$	$\downarrow 86$	$\downarrow 88$	$\downarrow 79$	$\downarrow 93$	$\downarrow 89$	$\downarrow 93$
Darkening	0.19	0.14	0.18	0.15	0.62	0.07	0.25	0.09	0.41	0.11	0.22	0.12
	$\downarrow 47$	$\downarrow 52$	$\downarrow 48$	$\downarrow 51$	$\downarrow 4$	$\downarrow 0$	$\downarrow 0$	$\downarrow 0$	$\downarrow 19$	$\downarrow 41$	$\downarrow 27$	$\downarrow 39$
Dirty lens	0.10	0.04	0.07	0.04	0.60	0.07	0.24	0.08	0.36	0.05	0.16	0.06
	$\downarrow 72$	$\downarrow 87$	$\downarrow 78$	$\downarrow 86$	$\downarrow 7$	$\downarrow 8$	$\downarrow 7$	$\downarrow 8$	$\downarrow 28$	$\downarrow 71$	$\downarrow 48$	$\downarrow 68$
Exposure	0.04	0.01	0.02	0.01	0.25	0.02	0.09	0.03	0.14	0.02	0.05	0.02
	$\downarrow 90$	$\downarrow 98$	$\downarrow 95$	$\downarrow 98$	$\downarrow 61$	$\downarrow 66$	$\downarrow 65$	$\downarrow 66$	$\downarrow 71$	$\downarrow 92$	$\downarrow 82$	$\downarrow 91$
Gaussian blur	0.24	0.03	0.11	0.04	0.54	0.08	0.25	0.09	0.41	0.06	0.18	0.07
	$\downarrow 33$	$\downarrow 89$	$\downarrow 68$	$\downarrow 87$	$\downarrow 17$	6	$\downarrow 2$	5	$\downarrow 18$	$\downarrow 70$	$\downarrow 40$	$\downarrow 66$
Noise	0.33	0.10	0.24	0.11	0.65	0.04	0.18	0.06	0.50	0.07	0.21	0.08
	$\downarrow 6$	$\downarrow 68$	$\downarrow 30$	$\downarrow 63$	1	$\downarrow 39$	$\downarrow 27$	$\downarrow 38$	$\downarrow 1$	$\downarrow 62$	$\downarrow 29$	$\downarrow 58$
Rain	0.00	0.00	0.00	0.00	0.45	0.05	0.17	0.06	0.22	0.02	0.09	0.03
	$\downarrow 100$	$\downarrow 100$	$\downarrow 100$	$\downarrow 100$	$\downarrow 31$	$\downarrow 35$	$\downarrow 31$	$\downarrow 35$	$\downarrow 55$	$\downarrow 87$	$\downarrow 71$	$\downarrow 85$
Shadow	0.27	0.23	0.26	0.24	0.64	0.06	0.22	0.07	0.46	0.15	0.25	0.16
	$\downarrow 24$	$\downarrow 21$	$\downarrow 22$	$\downarrow 21$	$\downarrow 1$	$\downarrow 17$	$\downarrow 13$	$\downarrow 17$	$\downarrow 8$	$\downarrow 21$	$\downarrow 16$	$\downarrow 20$
Snow	0.28	0.04	0.13	0.05	0.60	0.06	0.22	0.08	0.44	0.05	0.17	0.06
	$\downarrow 20$	$\downarrow 87$	$\downarrow 63$	$\downarrow 85$	$\downarrow 6$	$\downarrow 14$	$\downarrow 12$	$\downarrow 14$	$\downarrow 11$	$\downarrow 72$	$\downarrow 41$	$\downarrow 68$
Haze	0.22	0.00	0.00	0.00	0.64	0.07	0.25	0.09	0.44	0.04	0.13	0.05
	$\downarrow 37$	$\downarrow 100$	$\downarrow 99$	$\downarrow 100$	$\downarrow 1$	$\downarrow 2$	$\downarrow 0$	$\downarrow 1$	$\downarrow 11$	$\downarrow 79$	$\downarrow 56$	$\downarrow 77$
Average	0.16	0.06	0.10	0.07	0.53	0.06	0.20	0.07	0.36	0.06	0.15	0.07
	$\downarrow 55$	$\downarrow 80$	$\downarrow 70$	$\downarrow 78$	$\downarrow 18$	$\downarrow 22$	$\downarrow 21$	$\downarrow 22$	$\downarrow 29$	$\downarrow 68$	$\downarrow 48$	$\downarrow 65$

remains almost same as reported in top-I-II results. In other challenge categories, precision degradation varies between 8% and 79%. Performance degradation in recall is more significant than precision in all challenge categories, which varies between 21% and 93%. Performance variation in F score is in between 16% and 93%.

Even though almost all the simulated challenging conditions degrade detection performance, there are few exceptions for top-II algorithm in *decolorization* (+4% precision), *noise* (+1% precision) *lens blur* (+11% recall, +10% F_2 , and +1% $F_{0.5}$), and *Gaussian blur* (+6% recall and +5% F_2). The effect of *decolorization* significantly depends on the benchmark algorithm and the increase in precision under *noise* for top-II algorithm does not exceed 1%. In the *decolorization* category, detection performance of top-I algorithm degrades 100% in term of all metrics whereas performance variation of top-II algorithm does not exceed 10%. When traffic signs under severe challenging conditions are compared in Fig. 3, we can observe that structural information including high frequency components mostly remain intact in *decolorization* category whereas chroma information is distorted. Therefore, we can conclude that top-I algorithm significantly relies on color

information whereas top-II does not rely as much. Challenging conditions based on blur filter out high frequency components that can be used to recognize and localize traffic signs. But, at the same time, filtering out high frequency components can eliminate certain false detections, which can explain the minor performance enhancement in the aforementioned exceptional categories.

We report the average performance degradation over all the performance metrics and algorithms to understand the overall effect of challenging conditions types in Fig. 5. *Codec error* and *exposure* result in the highest performance degradation in the challenge categories. As observed in sixth row from top in Fig. 3, *exposure* condition can significantly saturate descriptive regions of traffic signs that are critical for recognition, which results in a high performance degradation. In case of *codec error*, we observe visual artifacts that corrupt the structural characteristics of the sign. In addition to the structural artifacts, *codec error* challenge can relocate a significant portion of the traffic sign to a new location as shown in the third row from top in Fig. 3, which would not satisfy the required overlap between the ground truth location and the detected location even though the sign can be recognized accu-

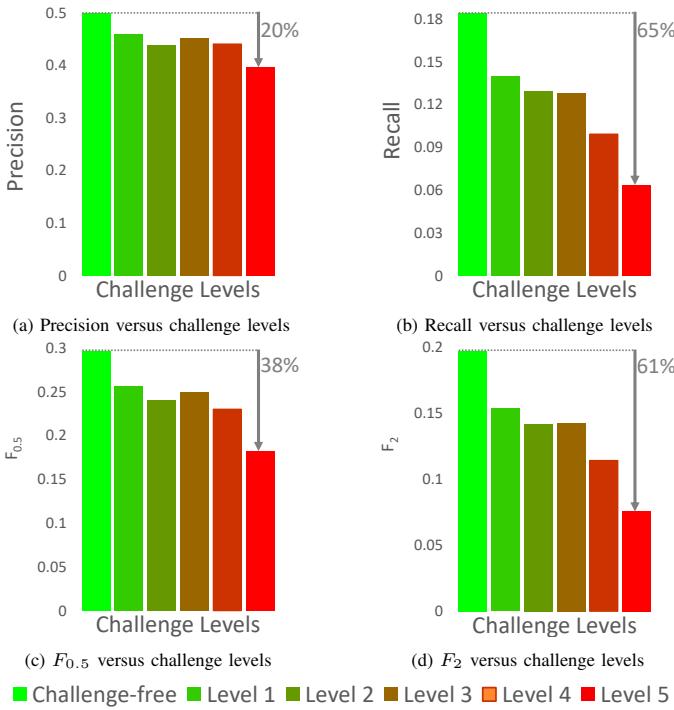


Fig. 4: Average traffic sign detection performance of top two algorithms with respect to challenge levels over all categories. Performance variations between challenge-free and severe conditions are reported for each metric in percentage.

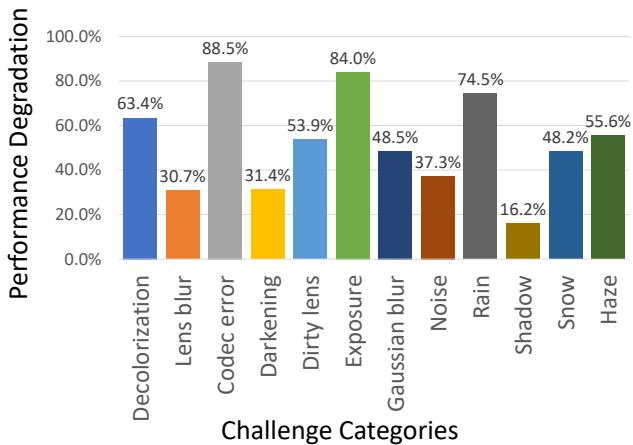


Fig. 5: Average detection performance degradation under challenging conditions for each challenge type.

rately. Benchmark algorithms can be vulnerable to challenging weather conditions including *rain*, *haze*, and *snow*, which lead to a performance degradation between 48.2% and 74.5%. In real-world scenarios, outer surface of the camera lens or the window surface in front of the camera can get dirty because of the weather and road conditions, which can affect the visibility of traffic signs due to occlusion. Based on the *dirty lens* experiments, occlusions can reduce the overall traffic sign detection performance by half. *Blur*, *darkening*, and *noise* challenges degrade the detection performance between 30.7 and 48.5% whereas *shadow* results in the least performance degradation by 16.2%. Because of the difficulty of realistic shadow generation, we used a simple periodic pattern to simulate local *shadow* effect, which can be considered as a partial *darkening* effect. We can observe that the performance degradation in the *darkening* category is almost double the degradation in the *shadow* category, which is proportional to the ratio of the darkened regions when degraded images are compared in both categories as observed in fourth row and tenth row from top in Fig. 3. When traffic sign images with high level *darkening* conditions are observed in Fig. 3, images may appear almost entirely dark and can be considered as the most challenging condition perceptually. However, the perceptual level of darkness depends on the display settings and if images are observed under different display configurations, it can be possible to observe descriptive parts of the traffic sign even under high level *darkening* conditions.

D. Spectral Analysis of Challenging Conditions

In this section, we investigate the effect of challenging conditions over the spectral characteristics of video sequences. In the CURE-TSD-Real dataset, there are 49 challenge-free sequences. Corresponding to each challenge-free reference video, there are 60 video sequences with distinct challenge conditions (12 types \times 5 levels). For each challenge-free reference video, we obtain the pixel-wise and frame-wise difference between the reference video and challenging video, which results in the residual video. In total, we obtain 2,940 (49 videos \times 60 challenge configurations) residual videos, which correspond to 245 residual videos per challenge type. For each residual video, we calculate the power spectrum per frame to obtain a power spectrum sequence. We calculate the power spectrum of a residual frame as

$$\text{Log} \{ |\mathcal{F} \{ |R| \}| \}, \quad (1)$$

where R is the residual frame, $| |$ is the absolute value, \mathcal{F} is the 2-D discrete Fourier transform, and Log is the logarithm. We calculate the pixel-wise average of 245 power spectrum sequences for each challenge type and quantize them to obtain the average magnitude spectrum maps in Fig. 6.

In the average magnitude spectrum maps, central region corresponds to low frequency components and corners represent high frequency components. Color coding is based on the magnitude of the frequency components. Color of the spectrum elements varies from dark blue to yellow as their magnitude increases. We can observe that challenging conditions lead to characteristic spectral shapes that can be used to analyze

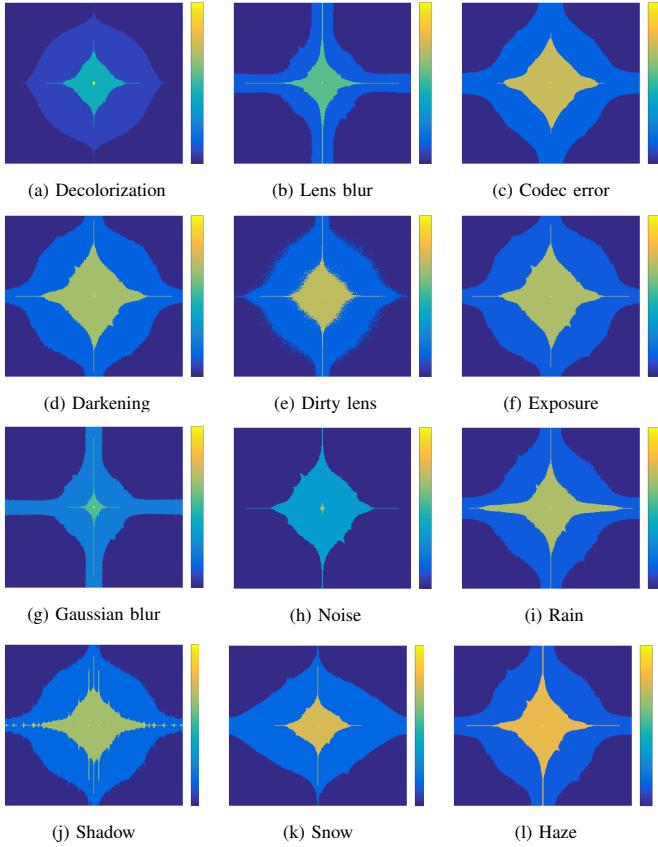


Fig. 6: Average magnitude spectrum maps of video sequences corresponding to different challenge types.

the effect of these conditions. Even though *exposure* and *darkening* correspond to perceptually very distinct images as observed in Fig. 3, their spectral representations correspond to an almost identical pattern in Fig. 6(d) and Fig. 6(f). Spectral representation of *exposure* challenge indicates that high-frequency components remain similar to the challenge-free sequences whereas low-frequency components get significantly distorted. *Exposure* and *haze* result in non-uniform deformations that affect the visibility of certain regions in an image, which lead to similar spectral representations. In the *blur* challenge, *lens blur* and *Gaussian blur* lead to a similar pattern in which boundaries of horizontal and vertical regions correspond to cutoff frequencies as observed in Fig. 6(b) and Fig. 6(g). Challenging conditions result in dominant vertical patterns in the *rain* and the *shadow* challenges as observed in Fig. 3, which correspond to a more predominant horizontal pattern in spectral representations as shown in Fig. 6(i) and Fig. 6(j). Moreover, we observe discrete peaks in the spectral representations of the *shadow* challenge in Fig. 6(j) because of the periodic shadow patterns. In the *rain* challenge, falling particles are the main occluding factor whereas in the *snow* challenge, piled up snow significantly occludes certain regions, which limits the highest spectral components to a more central region as observed in Fig. 6(k). *Noise* and *decolorization*

challenge lead to a peak at DC along with minor low frequency degradations as shown in Fig. 6(a) and Fig. 6(h). In the *codec error* challenge, we observe local shifts of certain regions in the images as shown in Fig. 3, which leads to an almost symmetric spectral representation in Fig. 6(c) without the sharp horizontal and vertical lines. On contrary to all other challenges, *dirty lens* challenge results in deformation over the images that varies in terms of shape and size, which leads to a granular structure in the spectral representation as shown in Fig. 6(e).

Previously, we investigated the effect of challenge types in spectral representations of residual videos. To understand the effect of challenge levels in spectral representations, we also need to analyze the changes in spectrum with respect to the severity of challenging conditions. In Fig. 7, we show the average magnitude spectrum maps of video sequences for different challenge levels and types. Minor condition corresponds to level one, medium condition corresponds to level three, and major condition corresponds to level five challenges. Each spectrum corresponds to the average of 49 video sequences with a specific challenge type and level. The most significant spectral change with respect to challenge levels occur in case of *decolorization* and *noise* whereas the least change occur in case of *codec error*, *darkening*, and *rain*.

In the aforementioned experiments and analysis, we focused on the effect of individual challenging conditions and levels because video sequences in the CURE-TSD-Real dataset include one challenging condition at a time. Therefore, it is not possible to directly assess the effect of concurrent challenging conditions. In order to test the capability of spectral representations under concurrent challenging conditions with an example, we combined *rain* and *exposure* conditions and obtained their magnitude spectrums as shown in Fig. 8. We obtained each magnitude spectrum map by averaging frame-level magnitude spectrums of 49 video sequences (14,700 frames). Spectral maps corresponding to concurrent *rain* and *exposure* conditions are shown in Fig. 8(c) and Fig. 8(f). In addition to spectral maps of concurrent *rain* and *exposure* conditions, we included the spectral maps of isolated *rain* and *exposure* conditions in Fig. 8(a), Fig. 8(b), Fig. 8(d), and Fig. 8(e) to visually compare them next to each other.

In case of concurrent conditions, we can observe that major condition dominates the spectral representation. Spectral map of concurrent major *rain* and minor *exposure* (Fig. 8(f)) is similar to the spectral maps of *rain* (Fig. 8(d-e)) in terms of asymmetry between horizontal and vertical components. Moreover, spectral map of concurrent major *exposure* and minor *rain* (Fig. 8(c)) is similar to the spectral map of major *exposure* (Fig. 8(b)). Thus, we can mention that dominant conditions mostly determine the shape of the spectral maps. However, we can still observe differences in the spectral maps when we compare concurrent major and minor condition with solely major condition. For example, spectral map of major *rain* (Fig. 8(e)) and spectral map of concurrent major *rain* and minor *exposure* (Fig. 8(f)) are still separable from each other in terms of shape and color. Meanwhile, spectral map of major *exposure* and minor *rain* (Fig. 8(c)) and spectral map of major *exposure* (Fig. 8(b)) are separable from each

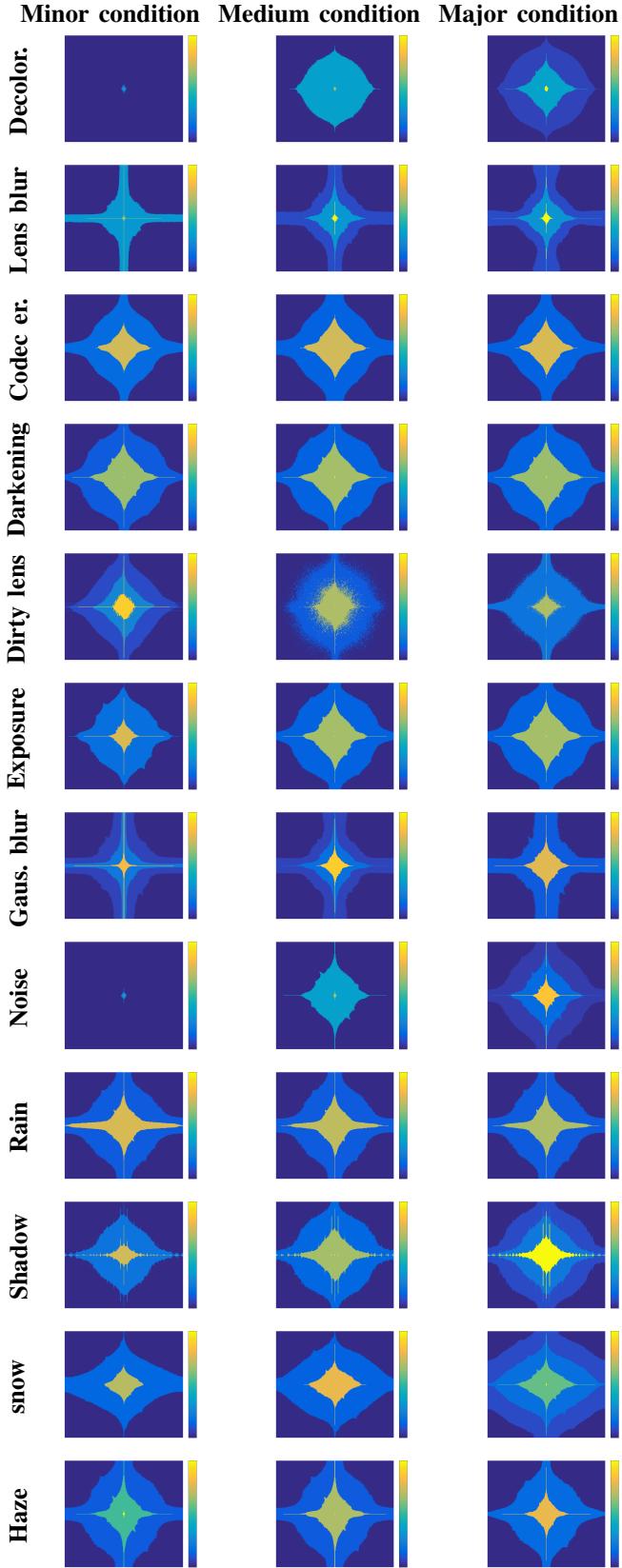


Fig. 7: Average magnitude spectrum maps of video sequences for different challenge levels and types.

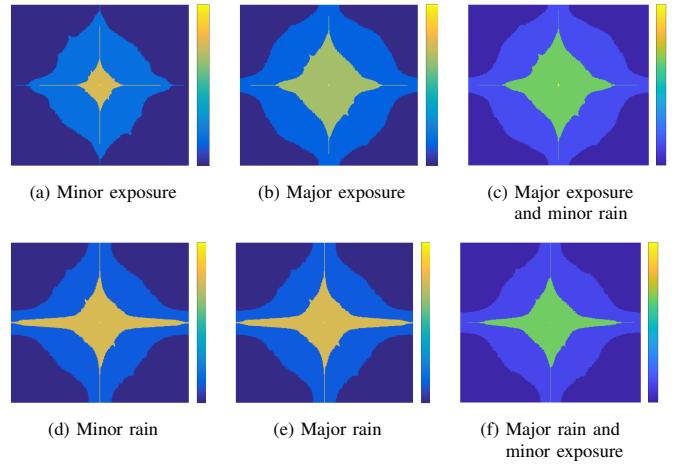


Fig. 8: Average magnitude spectrum maps of video sequences corresponding to rain, exposure, and a combination of rain and exposure at different challenge levels.

other in terms of color, which reflects differences in terms of spectral magnitude. Based on this example, we can express that spectral maps can reflect the impact of two concurrent conditions, but identification of the concurrent conditions may not be as straightforward as the identification of individual conditions.

E. Detection Performance versus Spectral Characteristics

Even though challenge levels affect the spectral representations, high level spectral shapes remain similar in majority of the challenging conditions. The intensity of the magnitude spectrums can be used to quantify the changes in spectral representations, which can be an indicator of the detection performance degradations. In Fig. 9, we show the relationship between detection performance and mean magnitude spectrum. Specifically, we computed the detection performance under varying challenge levels and calculated the mean magnitude spectrum corresponding to the varying challenge levels. We can observe that an increase in mean magnitude spectrum generally corresponds to a decrease in detection performance. To measure the correlation between traffic sign detection performance and mean magnitude spectrum, we calculated the Spearman rank order correlation coefficient, which is reported for each detection performance metric in Table IV. Specifically, we measured the correlation between mean magnitude spectrum and detection performance for each challenge category and obtained the average of these correlation coefficients. Based on the experiments, correlation between detection performance and mean magnitude spectrum is 0.643 for precision, 0.848 for recall, 0.657 for $F_{0.5}$ score, and 0.810 for F_2 score.

Spectral representations can be used to analyze the changes in images and videos and these changes can be quantified by measuring the changes in spectral representations. A direct mean pooling operation is a straightforward approach to quantify spectrums of residual sequences. However, detection algorithms do not necessarily react identically to changes

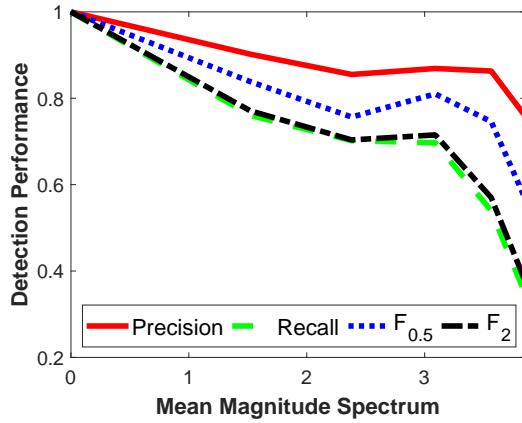


Fig. 9: Detection performance versus mean magnitude spectrum of residual video sequences.

at different frequencies. Therefore, instead of a direct mean pooling operation, a weighted pooling can be performed by considering the relative importance of frequency bands for traffic sign detection. For example, in JPEG compression [37], the objective is to compress the image as much as possible without visual artifacts. To achieve this objective, quantization tables were designed based on psychovisual experiments to compress signal components according to their perceptual significance. Similarly, a significance map can be designed for traffic sign detection application to quantify the changes in spectral components according to their algorithmic significance. Spectral analysis approach investigated in this study requires a reference video. Therefore, to estimate the traffic sign detection performance, we need to obtain the images of the same scene at different conditions. Such a system is feasible for a fixed camera setup in which we can capture the same region at different times. To deploy such systems to mobile platforms, we need to focus on no-reference spectral representations in which there is no need for a reference video.

TABLE IV: Detection performance degradation estimation with mean magnitude spectrum under challenging conditions.

Estimated Metric	Precision	Recall	$F_{0.5}$	F_2
Estimation Performance (Spearman Correlation)	0.643	0.848	0.657	0.810

V. CONCLUSION

We analyzed the average performance of benchmark algorithms in the CURE-TSD-Real dataset and showed that detection performance can significantly degrade under challenging conditions. *Codec error* and *exposure* resulted in the most significant performance degradation with more than 80% whereas *shadow* resulted in the least degradation with around 16%. Challenging weather conditions *snow*, *haze*, and *rain* resulted in at least 48% performance degradation. Around 63% performance degradation in *decolorization* highlighted the importance of color information for certain algorithms in

sign detection. Detection performance degradation based on *darkening*, *noise* and *blur* is in between 30% and 48% whereas *dirty lens* exceeds 50%. Our frequency domain analysis showed that simulated challenging conditions can correspond to distinct spectral patterns and magnitude of these spectral patterns can be used to estimate the detection performance under challenging conditions. Degradation estimation performance based on spectral representations was in between 0.64 and 0.84 in terms of Spearman correlation. As future work, adaptive pooling and no-reference spectral analysis are promising research directions that can be further investigated to estimate detection performance of algorithms by solely considering the environmental conditions.

REFERENCES

- [1] C. Grigorescu and N. Petkov, "Distance sets for shape filters and shape recognition," *IEEE Trans. Im. Proc.*, vol. 12, no. 10, pp. 1274–1286, Oct 2003.
- [2] R. Timofte, K. Zimmermann, and L. V. Gool, "Multi-view traffic sign detection, recognition, and 3D localisation," in *Works. Applic. Comp. Vis.*, Dec 2009, pp. 1–8.
- [3] R. Timofte, K. Zimmermann, and L. Van Gool, "Multi-view traffic sign detection, recognition, and 3D localisation," *Mach. Vis. Applic.*, vol. 25, no. 3, pp. 633–647, 2014.
- [4] R. Belaroussi, P. Foucher, J. P. Tarel, B. Soheilian, P. Charbonnier, and N. Paparoditis, "Road sign detection in images: A case study," in *Int. Conf. on Patt. Recog.*, Aug 2010, pp. 484–488.
- [5] F. Larsson and M. 'Felsberg, "Using fourier descriptors and spatial models for traffic sign recognition," in *Scand. Conf. Im. Analy.*, Berlin, Heidelberg, 2011, SCIA'11, pp. 238–249, Springer-Verlag.
- [6] J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel, "The German Traffic Sign Recognition Benchmark: A multi-class classification competition," in *Int. Joi. Conf. Neur. Netw.*, July 2011, pp. 1453–1460.
- [7] J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel, "Man versus computer: Benchmarking machine learning algorithms for traffic sign recognition," *Neur. Netw.*, vol. 32, pp. 323 – 332, 2012.
- [8] 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 *Int. Joi. Conf. Neur. Netw.*, Aug 2013, pp. 1–8.
- [9] 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.
- [10] Z. Zhu, D. Liang, S. Zhang, X. Huang, B. Li, and S. Hu, "Traffic-sign detection and classification in the wild," in *IEEE Conf. Comp. Vis. Patt. Recog.*, June 2016, pp. 2110–2118.
- [11] Y. Yang, H. Luo, H. Xu, and F. Wu, "Towards real-time traffic sign detection and classification," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 7, pp. 2022–2031, July 2016.

- [12] J. Zhang, M. Huang, X. Jin, and X. Li, "A real-time chinese traffic sign detection algorithm based on modified yolov2," *Algorithms*, vol. 10, no. 4, 2017.
- [13] K. Yi, Z. Jian, S. Chen, Y. Yang, and N. Zheng, "Knowledge-based recurrent attentive neural network for small object detection," in *arXiv:1803.05263*, 2018.
- [14] D. Temel, G. Kwon, M. Prabhushankar, and G. AlRegib, "CURE-TSR: Challenging unreal and real environments for traffic sign recognition," in *Neur. Inform. Proces. Syst. MLITS Works.*, 2017.
- [15] D. Temel and G. AlRegib, "Traffic signs in the wild: Highlights from the ieee video and image processing cup 2017 student competition [sp competitions]," *IEEE Sig. Proc. Mag.*, vol. 35, no. 2, pp. 154–161, 2018.
- [16] D. Temel, T. Alshawi, M.-H. Chen, and G. AlRegib, "Challenging environments for traffic sign detection: Reliability assessment under inclement conditions," in *arXiv:1902.06857*, 2019.
- [17] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus, "Intriguing properties of neural networks," in *Int. Conf. on Learn. Rep.*, 2014.
- [18] I. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," in *Int. Conf. on Learn. Rep.*, 2015.
- [19] J. Lu, H. Sibai, E. Fabry, and D. Forsyth, "No need to worry about adversarial examples in object detection in autonomous vehicles," in *IEEE Conf. Comp. Vis. Patt. Recog. Spot. Oral Works.*, 2017.
- [20] N. Das, M. Shanbhogue, S.-T. Chen, F. Hohman, L. Chen, M. E. Kounavis, and D. H. Chau, "Keeping the bad guys out: Protecting and vaccinating deep learning with jpeg compression," in *arXiv:1705.02900*, 2017.
- [21] H. Luo, Y. Yang, B. Tong, F. Wu, and B. Fan, "Traffic sign recognition using a multi-task convolutional neural network," *IEEE Trans. Intell. Transp. Syst.*, vol. PP, no. 99, pp. 1–12, 2017.
- [22] D. Temel, J. Lee, and G. AlRegib, "CURE-OR: Challenging unreal and real environments for object recognition," in *IEEE Int. Conf. Mach. Learn. Appl.*, Dec 2018, pp. 137–144.
- [23] D. Temel, J. Lee, and G. AlRegib, "Object recognition under multifarious conditions: A reliability analysis and a feature similarity-based performance estimation," in *IEEE Int. Conf. Im. Proces.*, Sept 2019.
- [24] M. Prabhushankar, G. Kwon, D. Temel, and G. AlRegib, "Semantically interpretable and controllable filter sets," in *IEEE Int. Conf. Im. Proces.*, Oct 2018, pp. 1053–1057.
- [25] G. Kwon, M. Prabhushankar, D. Temel, and G. AlRegib, "Distorted representation space characterization through backpropagated gradients," in *IEEE Int. Conf. Im. Proces.*, Sep. 2019, pp. 2651–2655.
- [26] A. Van der Schaaf and J.H. van Hateren, "Modelling the power spectra of natural images: Statistics and information," *Vis. Res.*, vol. 36, no. 17, pp. 2759 – 2770, 1996.
- [27] A. Torralba and A. Oliva, "Statistics of natural image categories," *Netw.: Comp.Neur. Syst.*, vol. 14, no. 3, pp. 391–412, 2003, PMID: 12938764.
- [28] C. Vondrick, D. Patterson, and D. Ramanan, "Efficiently scaling up crowdsourced video annotation," *Int. Jour. Comp. Vis.*, pp. 1–21, 10.1007/s11263-012-0564-1.
- [29] Merriam-Webster, "Definition of shadow," *Merriam-Webster Online Dictionary*, <https://www.merriam-webster.com/dictionary/shadow>.
- [30] S. Liu, L. Yuan, P. Tan, and J. Sun, "Bundled camera paths for video stabilization," *ACM Trans. Graph.*, vol. 32, no. 4, pp. 78:1–78:10, July 2013.
- [31] S. Liu, L. Yuan, P. Tan, and J. Sun, "Steadyflow: Spatially smooth optical flow for video stabilization," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014.
- [32] B. Zhuang, L. Cheong, and G. H. Lee, "Rolling-shutter-aware differential sfm and image rectification," in *International Conference on Computer Vision*, Oct 2017, pp. 948–956.
- [33] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *arXiv:1505.04597*, 2015.
- [34] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *arXiv:1512.03385*, 2015.
- [35] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *arXiv:1409.1556*, 2014.
- [36] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *arXiv:1409.4842*, 2014.
- [37] G. K. Wallace, "The jpeg still picture compression standard," *IEEE Transactions on Consumer Electronics*, vol. 38, no. 1, pp. xviii–xxxiv, Feb 1992.



Dogancan Temel received his M.S. and Ph.D. degrees in electrical and computer engineering (ECE) from the Georgia Institute of Technology (Georgia Tech) in 2013 and 2016, respectively, where he is currently a postdoctoral fellow. During his PhD, he received Texas Instruments Leadership Universities Fellowship for four consecutive years. He was the recipient of the Best Doctoral Dissertation Award from the Sigma Xi honor society, the Graduate Research Assistant Excellence Award from the School of ECE, and the Outstanding Research Award from the Center for Signal and Information Processing at Georgia Tech. His research interests are focused on human and machine vision.



Min-Hung Chen received an M.S. degree in 2012 from National Taiwan University. He worked as a research assistant in Academia Sinica, Taiwan in 2013. He is currently working towards PhD in the school of Electrical and Computer Engineering in Georgia Institute of Technology, Atlanta. During his PhD, he received Ministry of Education Technologies Incubation Scholarship from Taiwan for three years. His research interests include computer vision, deep learning, image and video processing. Recently, Min-Hung Chen is investigating the temporal dynamics through deep learning techniques.



Ghassan AlRegib is currently a Professor with the School of Electrical and Computer Engineering, Georgia Institute of Technology. He was a recipient of the ECE Outstanding Graduate Teaching Award in 2001 and both the CSIP Research and the CSIP Service Awards in 2003, the ECE Outstanding Junior Faculty Member Award, in 2008, and the 2017 Denning Faculty Award for Global Engagement. His research group, the Omni Lab for Intelligent Visual Engineering and Science (OLIVES) works on research projects related to machine learning,

image and video processing, image and video understanding, seismic interpretation, healthcare intelligence, machine learning for ophthalmology, and video analytics. He participated in several service activities within the IEEE including the organization of the First IEEE VIP Cup (2017), Area Editor for the IEEE Signal Processing Magazine, and the Technical Program Chair of GlobalSIP14 and ICIP20. He has provided services and consultation to several firms, companies, and international educational and Research and Development organizations. He has been a witness expert in a number of patents infringement cases.