SOCAR: Socially-Obtained CAR Dataset for Image Recognition in the Wild

Taewon Seo*
SOCAR AI Research
Seoul, Republic of Korea
cillian@socar.kr

Kyung Ho Park*
SOCAR AI Research
Seoul, Republic of Korea
kp@socar.kr

Hyunhee Chung
SOCAR AI Research
Seoul, Republic of Korea
esther@socar.kr

Abstract

While cars have become a significant object in computer vision applications, there are fewer spotlights on publicly-available car-related datasets. Among previously-proposed car datasets, we discover several improvement avenues. As most previous car datasets consist of web-crawled or surveillance camera-taken images, they are insufficient to illustrate various attributes, such as points of view or parts. Moreover, prior datasets primarily deal with a car model recognition task; thus, the scope of applicative studies was limited. To improve these avenues, we propose a Socially-Obtained CAR (SOCAR) dataset, a real-world car image dataset consisting of car images with more prosperous attributes. The key contributions of our study are as follows. First, under coordination with a large-scale car-sharing platform, we retrieve user-taken car images on both external and internal attributes and establish a dataset consisting of 10K images on 14 classes. Second, we design each class to represent a particular car's state; therefore, the SOCAR dataset enables the practitioners to solve various image recognition tasks such as understanding defects, dirt, or car wash. Lastly, we suggest baseline experiment results on the proposed dataset and experimentally examine the trained model effectively capture discriminative regions similar to human vision. We highly expect practitioners to use our SOCAR dataset for academic research on understanding car attributes or computer vision applications.

1. Introduction

Background and Motivation As the car has become one of the essential transportation in society, it has become an important object in modern computer vision studies. Upon the rise of deep neural networks, numerous works proposed car-related applications that recognize car models [9, 3, 12], or identify defects on a car's surface to automate post-accident procedures [11, 16, 8]. To nourish car-related applications, numerous studies publicized large-scale car image datasets [9, 20, 10]; however, we claim these datasets are insufficient to be utilized in the real world, where unexpected novel attributes exist. We illustrate the limits of previously-proposed datasets below, which also become key motivations of our study.

First, most previously-proposed datasets consist of web-crawled or synthetically-taken car images. As the car images are usually taken outside, various factors (*i.e.*, illumination, angle, point of view, weather) impact the car images in the real world. While the car image recognition model should understand the car's attributes under the influence of these external factors, we claim that web-crawled or synthetically taken images are insufficient to meet these requirements. Second, several datasets have a limited scope of car attributes. Although a car has a variety of attributes (*i.e.*, external and internal parts), most prior datasets primarily focus on the external surface of the car; thus, computer vision practitioners can only let the model learn the external characteristics of the car. Last but not least, image recognition tasks based on the previous datasets are primarily limited to car model classification. Not only understanding a car's model, but we also presume that real world practitioners would have many interests in understanding a car's status. For example, given a car's external image, practitioners can automate post-accident insurance claims by classifying whether the car is damaged or not [11, 14, 16]. The practitioners at the car-sharing platform would effectively manage the car's hygiene status by classifying whether the given image shows dirty attributes (*i.e.*, dirt on the surface or trashes on the mats). However, prior datasets are insufficient to let practitioners establish these applications.

Key Contributions and Our Novelty To this end, we propose a Socially-Obtained CAR (SOCAR) dataset, which improves the aforementioned limits of previously-proposed datasets. The key contributions of our study are as follows. First, the practitioners can utilize our dataset to train the model to better understand the car's characteristics. We configure the SOCAR dataset with car images taken under various circumstances (i.e., background or point-of-view). Thus, we presume the model can acquire a more robust, practical inductive bias compared to the one trained under synthetically created, or web-crawled datasets. Second, the SOCAR dataset empowers practitioners to solve various car-related image recognition tasks in the real world. We establish our dataset with 14 labels, each describing a car's various statuses (i.e., normal, defect, or dirt) under both external and internal attributes. While prior datasets primarily focused on recognizing external attributes, based on our dataset, we expect the computer vision community can extend the target tasks, such as damage recognition, dirt recognition, or wash recognition. Third, we empower the practitioners to validate whether their models are robust to various weather conditions. While the SOCAR dataset's training set has car images taken on a sunny day, the test sets include car images under various weather conditions (Rainy, After Rainy day, Snowy, After Snowy day). As the influence of weather conditions (often unexpected in the real world) is the most significant but challenging hurdle in car-related applications, we expect our SOCAR dataset to become an effective evaluation suite. Fourth, we illustrate the procedures of establishing a dataset; thus, future practitioners who desire to make a similar dataset can refer to the proposed procedures. Lastly, we provide baseline experiment results on two image recognition tasks: damage recognition and dirt recognition. We aim to provide a solid guideline for future researchers by providing image recognition performances at various model architectures.

We hereby highlight our work's novelty in the following avenues. First, to the best of our knowledge, our work firstly enables the computer vision communities to solve such extended car-related image recognition tasks. Second, our dataset includes real world car images including various attributes. Furthermore, the SOCAR dataset is the first evaluation suite for evaluating the image recognition model's robustness to weather conditions in a car-related domain.

2. Related Works

2.1. Car-related Computer Vision Applications

Previously-proposed car-related applications primarily focus on car model recognition tasks. As an early approach, Krause *et al*., propose a benchmark car model dataset and define this task as a fine-grained classification [9]. Moreover, Fang *et al*., utilized CNNs to identify discriminative cues of each car model [3]. Lu *et al*., also proposed a hierarchical car model recognition method consisting of two stages: car logo classifier and sub-class classifiers [12]. Furthermore, recent car-related applications have started to focus on car damage recognition, which identifies whether a car image has a damaged area or not. Especially, car insurance companies and car sharing platforms actively contributed to these studies to automate post-accident procedures or prevent the over-claim of their customers. Prior works primarily utilize web-crawled car images to solve these tasks. Zhang *et al*., proposed a tire damage detection approach with CNN [23], and Patil *et al*., once utilized the transfer learning method to identify a car's damaged pattern [14]. Li *et al*., employed object detectors for predicting the damaged area at the given image [11]. Singh *et al*., proposed a car damage recognition method that classifies whether the car has damage or not, but also the severity of the damage [16]. Balci *et al*., once proposed a damage detection method under the surveillance camera-taken images, while most works are based on web-crawled images. Lastly, Khana *et al*., once suggested a car part recognition model with CNN [6].

2.2. Car-related Datasets

While numerous car-related application studies exist, there are few publicized benchmark datasets in the computer vision society. As a very classic contribution, Krause *et al*., proposed the Stanford-Cars dataset, which consists of 16,185 samples with 196 car models [9]. Based on this Stanford-Cars, numerous fine-grained classification studies have been proposed to identify car models. Along with the importance of car-related datasets, Yang *et al*., also proposed the CompCars dataset, which consists of 136,728 entire car images, 27,618 car parts images, and 50,000 front-view car images captured by a surveillance camera. [20]. While previously proposed Stanford-Cars only included web-crawled samples, CompCars has two improved contributions. First, it includes car images accumulated from the surveillance camera; thus, it empowers broader use cases for the practitioners. Second, car images in the CompCars embrace both external and internal attributes of the car, while Stanford-Cars only have external attributes. Along with this trend, Kuhn *et al*., suggested BRCars dataset consists of 300,325 samples [10] describing various car models' exterior and interior attributes. This dataset provides a more realistic problem setting regarding car-related applications as it bears real-world samples taken from various points of view. Our study shares similar motivation with the BRCars dataset (providing realistic problem settings with a real-world dataset). However, we claim that our SOCAR dataset has improved contribution because it enables practitioners to solve various image recognition tasks (*i.e*., car defect recognition, car dirt recognition), while previous datasets are focused on car model classification tasks.

3. Dataset Establishment Procedure

3.1. Source of Dataset

We hereby describe how we acquire the raw images from the real world car-sharing operations. The car-sharing platform enables users to borrow a car with a smartphone application. When users want to use the car, they search for the nearest parking station, reserve the car, open up the car

Figure 1. Sample images at the SOCAR dataset. From the upper left, in clockwise order, each image illustrates External Normal, External Defect, External Dirt, External Wash, Washing Machine, Dashboard, Cupholder, Cupholder Dirt, Glovebox, Water Fluid, Front Seat, Rear Seat, Trunk, and Mat Dirt. Note that the red box implies discriminative cues for recognizing the given label.

(a) SOCAR-Defect-Test (b) SOCAR-Dirt-Test

Figure 2. Sample images at the SOCAR-Damage-Test and SOCAR-Dirt-Test. The left image implies *External Defect* and *External Dirt* at SOCAR-Damage-Test and SOCAR-Dirt-Test, respectively. The right image illustrates *External Normal* sample. Note that the red box implies discriminative cues for recognizing the given label.

Table 1. Descriptions on each label and the number of samples at each class in SOCAR-Main

Class Name	Description	Number of samples
External Normal	Car's exterior surface without any defects or dirt	2000
External Defect	Car's exterior surface with defects such as scratch, dents, spacing, or breakage	2000
External Dirt	Car's exterior surface with dirt	1759
External Wash	Car's exterior surface with bubbles during the car wash	2000
Washing Machine	User-taken car images taken inside of the car at the washing machine	2000
Dashboard	Car's interior dashboard	2000
Cupholder	Car's interior cupholder without any dirt	2000
Cupholder Dirt	Car's interior cuphollder with dirt or trash	539
Glovebox	Car's interior glovebox at the passenger seat	2000
Washer Fluid	Washer fluid in the bonnet	2000
Front Seat	Car's interior front seat	1671
Rear Seat	Car's interior rear seat	1957
Trunk	Car's interior trunk	1748
Mat Dirt	Car's interior mat with dirt or trash	1628

Table 2. The number of samples at SOCAR-Defect-Test and SOCAR-Dirt-Test

	SOCAR-Defect-Test					**SOCAR-Dirt-Test**	
	Snowy Day	**After Snowy Day**	**Rainy Day**	**After Rainy Day**		**After Rainy Day**	**After Snowy Day**
External Normal	1009	920	1216	823	**External Normal**	1667	1273
External Defect	136	146	70	90	**External Dirt**	386	642

with a smartphone application, and start their trips. After they finish their use, the users park the car at the parking station where they borrowed it and lock the car, and end the reservation. During these procedures of car use, *anonymous company name* forces the users to take pictures of the car and upload them through the application at particular events, such as before and after using the car, during an accident, car wash, repair, or charging washer fluids. The users take pictures of the car's exterior and interior attributes under various circumstances, and these images are accumulated in the central database of the car-sharing platform. As shown in Figure 1, these samples become a source of the SOCAR dataset, and we manually annotate these samples.

3.2. Annotation

As the source of car images are user-generated ones, irrelevant or less-qualified samples frequently exist. We presume providing every sample to the annotators would bear a particular waste of resources; thus, we employ a simple but effective method to reduce annotation costs. First, given a set of unlabeled data, we establish a seed dataset consisting of 500 samples at each label. Second, we train a seed classifier based on ResNet-50 with conventional cross-entropy loss, which solves 14-class classification. Supposing this trained model understands the discriminative cues of given labels, we then implement an out-of-distribution (OOD) detector following the method proposed in [18]. Note that we utilize the OOD detector in [18] due to its state-of-the-art performance in OOD detection studies. Given the unlabeled sample, we provide it to the OOD detector and eliminate it from the annotation procedure when it is identified as an OOD. We expect the samples predicted as OOD would presumably increase the annotation cost without adding labeled samples to the dataset. For the samples that are not identified as OOD, we assign pseudo-labels for them following the prediction results yielded by the trained seed classifier. Lastly, we let the labelers annotate these pseudo-labeled samples by checking whether the pseudo-label is correct or not. Note that we do not highlight the quantitative impact of these annotation procedures in this work as our study's aim is proposing the labeled dataset, not proposing novel data annotation procedure.

3.3. Preventing Privacy Concern

As the car images are generated from the real world, it is crucial to ensure that the dataset does not include any privacy issues, especially human's faces or bodies. Thus, we employ a pre-trained human face detector [24] and human body detector [22] and check every annotated sample. We then eliminate the sample that has either detected face or body to prevent privacy concerns. After we deleted face or body-detected samples, we manually checked every sample to ensure our dataset's sanity regarding privacy concerns.

4. The SOCAR dataset

4.1. Key Characteristics

The SOCAR dataset has 14 classes describing the car's various statuses, and we summarize brief descriptions of each class in Table 1. We hereby highlight that our SOCAR dataset reflects various realistic attributes of the car by adding diverse circumstances in the real world. The detailed descriptions are provided below.

Background Variety The images are taken under various backgrounds. As various studies proposed earlier, image recognition models are easily biased to the background, not the primary object or characteristics in a given image [13]. To evade this risk, we intentionally include samples with various backgrounds, as shown in Figure 3 (a). The samples at Figure 3 (a) belong to the *External Normal* label. The samples in each class in the SOCAR dataset are taken at indoor parking lots or outside roads; thus, we presume samples in the SOCAR dataset sufficiently reflect various backgrounds.

Variety at POVs and Car Parts The images depicting the exterior part have various POV(Point of View)s and parts to prevent the model from being overly biased toward these attributes. For example, referring to Figure 3 (b), we show the samples in the *External Dirt* label. While these samples are annotated as the same class, they are taken at various POVs. Furthermore, upon the Figure 3 (c), they also belong to the *External Dirt* as well as depicting various car parts. By diversifying POVs and car parts in the SOCAR dataset, we presume the inductive biases learned in the model would become similar to the human vision, not simply biased to the particular attributes.

Defect Variety As various defect types exist in the real world, we aim to depict most of these characteristics in the dataset. As shown in Figure 3, we empirically discover four defect types: scratch, dent, spacing, and harsh breakage. As we aim to evade letting the model be biased to the particular defect type, we retrieve damaged car images at each defect type and manually add them to the *External Defect* class. Note that samples at each damage type are not balanced.

Dirt Variety We further diversified the characteristics of dirt in the SOCAR dataset. Especially, for *Cupholder Dirt* and *Mat Dirt*, we aim to include various dirt characteristics. Referring to Figure 3 (d), we show samples belonging to the *Mat Dirt*. These images include dust, cookie crumbs, or any trash. We aim to empower the learned model to understand various patterns of dirty attributes of the car, which frequently exist in the real world.

4.2. Description on Subsets

The SOCAR dataset consists of three subsets: SOCAR-Main, SOCAR-Defect-Test, and SOCAR-Dirt-Test. First, the SOCAR-Main are the subsets where practitioners can

(a) Background Variety (b) POV Variety (c) Part Variety (d) Defect Variety (e) Dirt Variety

Figure 3. Various characteristics of the SOCAR dataset

train their image recognition models and test their performance. In this study, we split the SOCAR-Main into the training and test sets following the ratio of 9:1 with a random seed. Furthermore, SOCAR-Defect-Test and SOCAR-Dirt-Test are the external validation sets that the practitioner can use to check the robustness of trained models on images taken under various weather conditions. For SOCAR-Defect-Test, there exist four weather conditions: Snowy, After Snowy, Rainy, and After Rainy. *Snowy* and *Rainy* include car images under snowy and rainy weather, where rain drops and snow confuses the model in recognizing the damaged area, respectively. We additionally add car images taken after these snowy and rainy days because specks of dirt remain after the rain and snow yield dirty attributes on the car's surface, which confuses the model in recognizing the damaged surface. For SOCAR-Dirt-Test, there exist two weather conditions Snowy and Rainy. We exclude *After Rainy* and *After Snowy* option for SOCAR-Dirt-Test because there's no tangible differences between *Rainy* and *After Snowy*. Please refer to Table 1 and Table 2 for the detailed numbers.

5. Experiments

5.1. Image Recognition at SOCAR-Main

Objective and Setup First and foremost, we provide image recognition performances of various deep neural networks. We aim to examine whether the SOCAR dataset's label space is well-designed enough to be classified by conventional deep neural networks. We also expect these results would be a solid baseline of models trained under the SOCAR dataset; thus, future works might refer to these results to check whether they correctly implement their models. We employ widely-utilized convolutional neural network (CNN) architectures and vision transformers (ViT) for deep neural network architectures. The utilized networks are ResNet [4], DenseNet [5], ResNext [19], Wide ResNet [21], EfficientNet [17], and ViT [1]. In addition, we also utilize Progressive Multi-Granularity (PMG) [2] network in the experiment, as we presume discriminative cues between given labels could be interpreted as fine-grained ones. Note that PMG is state-of-the-art in fine-grained classification studies. We set the learning objective as minimizing cross-entropy loss under the Adam [7] optimizer. We highlight that no data augmentation strategies have been used for the fair evaluation. Please refer to the supplementary materials for more detailed implementation details. Furthermore, given the trained classifier, we investigate whether it correctly captures the discriminative cues of the given class. We employ Grad-CAM [15] to visualize the region where the trained model primarily focuses on. By exploring mostly activated regions at the given image, we aim to validate whether the inductive bias is correctly conveyed to the image recognition model. Upon these setups, experiment results and Grad-CAM results are shown in Table **??** and Figure 4, respectively. Note that the Grad-CAM results are based on ResNet-50 classifier.

Table 3. Experiment results on SOCAR dataset at various models

Model	Accuracy	Precision	Recall	F1 Score
ResNet-50	0.9757	0.9443	0.9761	0.9538
ResNet-100	0.9778	0.9330	0.9796	0.9441
DenseNet-169	0.9768	0.9445	0.9763	0.9541
DenseNet-201	0.9755	0.9306	0.9770	0.9414
ResNext-50	0.9778	0.9462	0.9791	0.9561
Wide ResNet-50	0.9768	0.9270	0.9778	0.9358
EfficientNet	0.9571	0.8310	0.8286	0.8287
PMG	0.9730	0.9233	0.9740	0.9319
ViT-16	0.9483	0.8964	0.9529	0.9082
ViT-32	0.9223	0.8618	0.9284	0.8674

Table 4. Experiment results on car defect recognition and car dirt recognition at SOCAR-Main.

SOCAR-Test	Car Defect Recognition		Car Dirt Recognition	
	Accuracy	F1 Score	Accuracy	F1 Score
Binary Classifier (ResNet50)	0.9190	0.9187	0.9618	0.9568
Binary Classifier (PMG)	0.9290	0.9289	0.9684	0.9645
14-Class Classifier (ResNet50)	0.9290	0.9288	0.9578	0.9521
14-Class Classifier (PMG)	**0.9220**	0.9217	0.9618	0.9566

Analysis We observe that conventional deep neural networks sufficiently accomplish precise image classification performances on the SOCAR dataset; thus, we conclude that the label space is well-designed. Furthermore, we discover the most activated region in each class corresponds to the correct discriminative cues. Upon Figure 4, for ex-

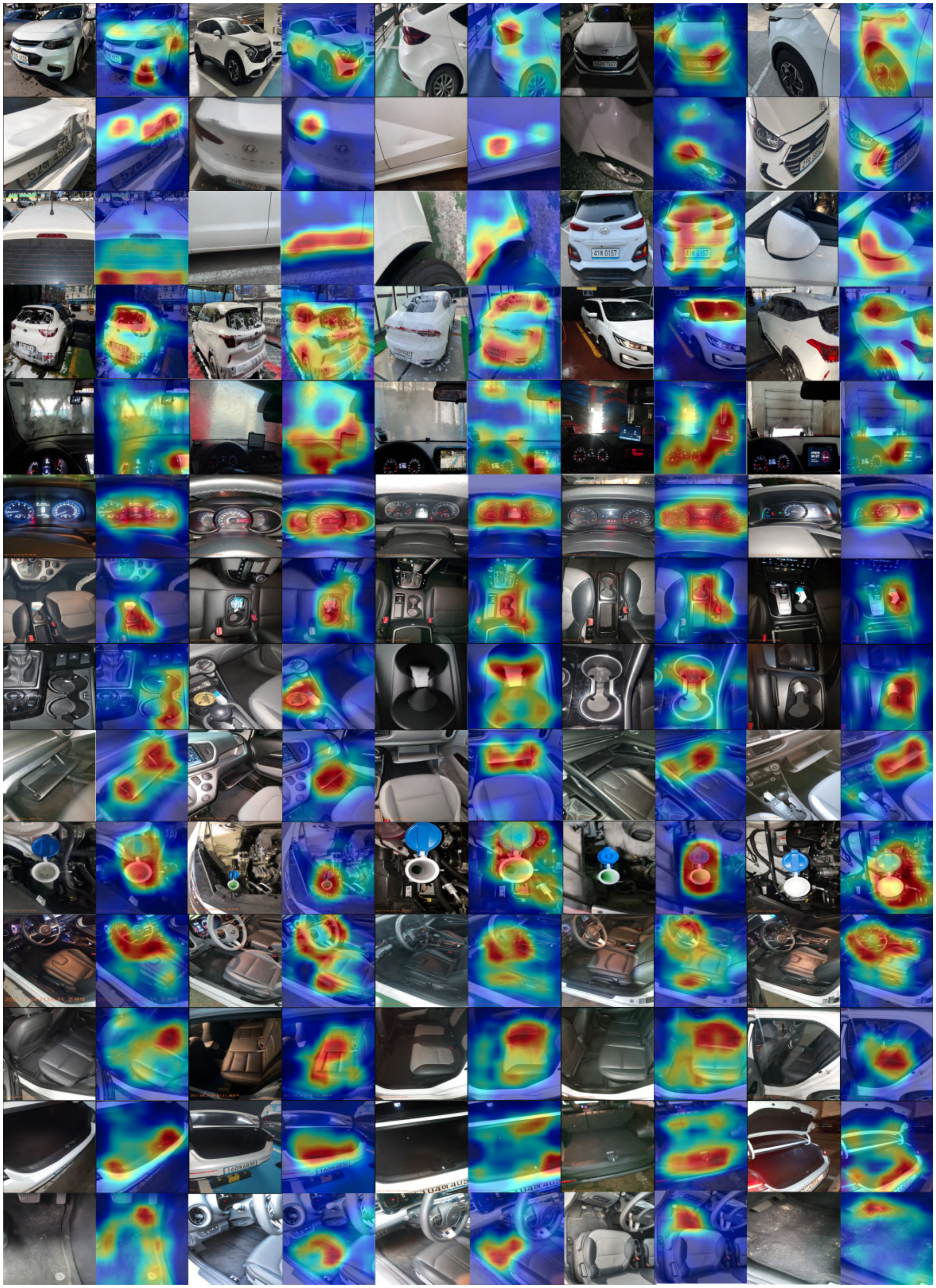


Figure 4. Grad-CAM results yielded by the trained classifier based on ResNet-50. From the first row to the last one, samples at each row are *External Defect*, *External Dirt*, *External Wash*, *Glovebox*, and *Mat Dirt*. We discover that the trained model correctly captures discriminative cues of the given classes. Please refer to the supplementary materials for the Grad-CAM results on the other classes.

ample, we observe the trained model correctly identifies the damaged area in *External Defect*. For *External-Dirt* and *External-Wash*, the model correctly captures relevant area at the car's surface; dirty area at *External Dirt* and bubbles at *External Wash*. Not only the external attributes, but the model also precisely identifies relevant discriminative regions at *Glovebox* and *Mat Dirt*; therefore, we conclude that the inductive bias is correctly acquired in the model.

Interestingly, we discover that ViT models achieve inferior performance compared to CNN-based models. We presume an underlying reason is the number of training samples in the dataset. As prior studies once noted [1], transformer-based models require an enormous amount of training samples to learn discriminative cues on given labels. We interpret that the given training samples are insufficient to convey adequate inductive bias to the ViTs; therefore, we highly recommend practitioners utilize CNN-based models to achieve supreme performance on the SOCAR dataset.

5.2. Image Recognition under Various Weathers

Objective and Setup We then examine whether utilizing irrelevant labels is advantageous in the target image recognition task. Among various image recognition tasks, we employ two tasks: car defect recognition and car dirt recognition. The car defect recognition is a binary classification between *External Normal* and *External Defect*, while the car dirt recognition is also a binary classification between *External Normal* and *External Dirt*. One possible solution to these problems is acquiring training samples from the binary labels and training the model (denoted as *Binary Classifier*). However, at this point, we question whether utilizing irrelevant labels is advantageous in learning discriminative cues on given labels. For example, for car defect recognition, samples at the *External Normal* and *External Defect* are relevant labels, and the other 12 labels are irrelevant. We hypothesize that utilizing irrelevant labels would enhance the inductive bias of the trained model, as the model can learn discriminative cues of given labels compared to the other labels. We denote the model trained under the 14-class as *14-Class Classifier*.

Given the image recognition task, we trained both *Binary Classifier* and *14-Class Classifier* with the samples at SOCAR-Training. For the test sets, we utilized both SOCAR-Test and SOCAR-Defect-Test for car defect recognition, while SOCAR-Test and SOCAR-Dirt-Test for car dirt recognition. We followed the same implementation details shown in Section 5.1, and evaluated the trained model with four evaluation metrics: Accuracy, Precision, Recall, and F1-score. Furthermore, to examine whether using irrelevant labels convey adequate inductive biases, we compare Grad-CAM results at *Binary Classifier* and *14-Class Classifier*. We acquired several samples at SOCAR-Defect-Test and SOCAR-Dirt-Test and applied Grad-CAM with trained models for car defect recognition and car dirt recognition, respectively. The experiment results are shown in Table 4, Table 5, and Table 6. We also visualize compared Grad-CAM results at *Binary Classifier* and *14-Class Classifier* in Figure 5.

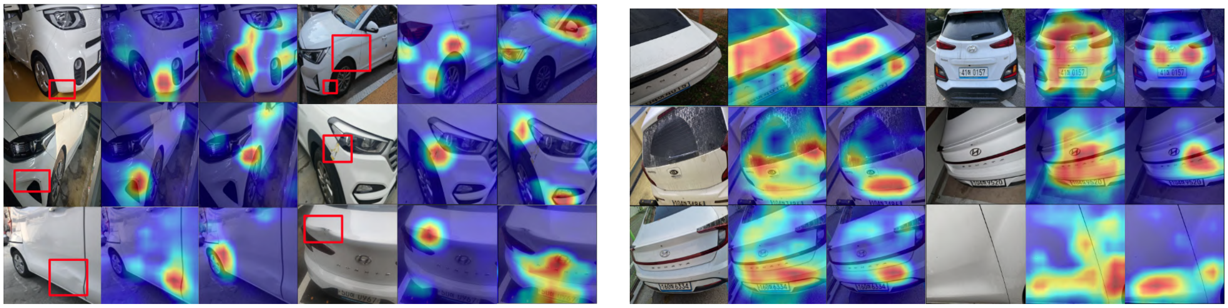
Analysis Upon the experiment results, we discover that using irrelevant samples is advantageous in recognizing target labels. Moreover, *14-Class Classifier*, which is the model trained under every label of the SOCAR dataset, achieved supreme performance in every setting. We analyze that an underlying reason for these results exists at the learned inductive bias. Upon Figure 5, we observe the most activated region at the *Binary Classifier* insufficiently captures the discriminative cues. For samples in the *External Defect*, *Binary Classiifer* frequently captures the correct discriminative region while *14-Class Classifier* correctly identifies corresponding area. For samples in the *External Dirt*, we observe that *Binary Classifier* fails to fully recognize discriminative regions while the *14-Class Classifier* sufficiently covers the relevant area. We interpret that the binary classifiers acquired an inductive bias particularly focused on discriminating given two labels, but the use of irrelevant labels at *14-Class Classifier* contributes to the acquisition of representation power similar to human vision. Consequentially, we recommend practitioners utilize every label in the SOCAR dataset to acquire better image recognition performances and precise inductive biases similar to human vision.

6. Conclusions

We propose the SOCAR dataset, a novel car-related dataset consisting of ten thousand real world car images. Compared to previously-proposed car-related datasets, our dataset has the following novelties. First, the SOCAR dataset includes car images retrieved from real world car-sharing operations; thus, it provides richer attributes to the practitioners. Second, classes in the SOCAR dataset describe various car statuses, such as defect, dirt, or wash. We highly expect the practitioners can establish various academic contributions or computer vision applications related to the car. Lastly, for reproducibility, we perform a series of experiments based on the proposed dataset and report corresponding results. We also examine that the trained model precisely captures discriminative cues of the given sample; thus, the inductive bias is correctly conveyed to the model.

7. Acknowledgement

This work is supported by the Korea Agency for Infrastructure Technology Advancement grant funded by the Ministry of Land, Infrastructure, and Transport (Grant RS-2022-0014579).



(a) Car Defect Recognition (b) Car Dirt Recognition

Figure 5. Grad-CAM results yielded by the trained *Binary Classifier* and *14-Class Classifier* based on ResNet-50. The 1st and 4rd column imply the original image, the 2nd and 5th column describes the Grad-CAM result at *14-Class Classifier*, and the third and sixth columns illustrate Grad-CAM results at *Binary Classifier*. We observe the *14-Class Classifier* effectively captures discriminative cues of the given classes while *Binary Classifier* insufficiently recognizes them. Thus, we recommend that the practitioners use every label in the SOCAR dataset to convey inductive bias in the model similar to human vision.

Table 5. Experiment results on car defect recognition at SOCAR-Defect-Test.

Car Defect Recognition	Accuracy				F1 Score			
	Snowy Day	After Snowy Day	Rainy Day	After Rainy Day	Snowy Day	After Snowy Day	Rainy Day	After Rainy Day
Binary Classifier (ResNet50)	0.7799	0.8204	0.8096	0.8705	0.5768	0.6521	0.6737	0.7704
Binary Classifier (PMG)	0.7683	0.8039	0.8009	0.8593	0.5677	0.6514	0.6770	0.7547
14-Class Classifier (ResNet50)	**0.8927**	**0.8981**	**0.8594**	**0.8902**	**0.6626**	**0.7230**	**0.6972**	**0.7798**
14-Class Classifier (PMG)	0.8849	0.8817	0.8428	**0.8902**	0.6381	0.6968	0.6642	0.7762

Table 6. Experiment results on car dirt recognition at SOCAR-Dirt-Test.

Car Dirt Recognition	Accuracy		F1 Score	
	After Rainy Day	After Snowy Day	After Rainy Day	After Snowy Day
Binary Classifier (ResNet50)	0.9123	0.8444	0.8474	0.8173
Binary Classifier (PMG)	0.9065	0.8548	0.8404	0.8293
14-Class Classifier (ResNet50)	**0.9157**	0.8470	**0.8519**	0.8152
14-Class Classifier (PMG)	0.9133	**0.8595**	0.8457	**0.8311**

References

[1] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

[2] Ruoyi Du, Dongliang Chang, Ayan Kumar Bhunia, Jiyang Xie, Zhanyu Ma, Yi-Zhe Song, and Jun Guo. Fine-grained visual classification via progressive multi-granularity training of jigsaw patches. In *European Conference on Computer Vision*, pages 153–168. Springer, 2020.

[3] Jie Fang, Yu Zhou, Yao Yu, and Sidan Du. Fine-grained vehicle model recognition using a coarse-to-fine convolutional neural network architecture. *IEEE Transactions on Intelligent Transportation Systems*, 18(7):1782–1792, 2016.

[4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[5] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.

[6] Salik Ram Khanal, Eurico Vasco Amorim, and Vitor Filipe. Classification of car parts using deep neural network. In *Portuguese Conference on Automatic Control*, pages 582–591. Springer, 2020.

[7] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[8] Alex Kost. Applying neural networks for tire pressure monitoring systems. 2018.

[9] Jonathan Krause, Michael Stark, Jia Deng, and Li Fei-Fei. 3d object representations for fine-grained categorization. In *Proceedings of the IEEE international conference on computer vision workshops*, pages 554–561, 2013.

[10] Daniel M Kuhn and Viviane P Moreira. Brcars: a dataset for fine-grained classification of car images. In *2021 34th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 231–238. IEEE, 2021.

[11] Pei Li, Bingyu Shen, and Weishan Dong. An anti-fraud system for car insurance claim based on visual evidence. *arXiv preprint arXiv:1804.11207*, 2018.

[12] Lei Lu and Hua Huang. A hierarchical scheme for vehicle make and model recognition from frontal images of vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 20(5):1774–1786, 2018.

[13] Thao Nguyen, Maithra Raghu, and Simon Kornblith. On the origins of the block structure phenomenon in neural network representations. *arXiv preprint arXiv:2202.07184*, 2022.

[14] Kalpesh Patil, Mandar Kulkarni, Anand Sriraman, and Shirish Karande. Deep learning based car damage classification. In *2017 16th IEEE international conference on machine learning and applications (ICMLA)*, pages 50–54. IEEE, 2017.

[15] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, pages 618–626, 2017.

[16] Ranjodh Singh, Meghna P Ayyar, Tata Venkata Sri Pavan, Sandeep Gosain, and Rajiv Ratn Shah. Automating car insurance claims using deep learning techniques. In *2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM)*, pages 199–207. IEEE, 2019.

[17] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pages 6105–6114. PMLR, 2019.

[18] Sagar Vaze, Kai Han, Andrea Vedaldi, and Andrew Zisserman. Open-set recognition: A good closed-set classifier is all you need. *arXiv preprint arXiv:2110.06207*, 2021.

[19] Saining Xie, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. Aggregated residual transformations for deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1492–1500, 2017.

[20] Linjie Yang, Ping Luo, Chen Change Loy, and Xiaoou Tang. A large-scale car dataset for fine-grained categorization and verification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3973–3981, 2015.

[21] Sergey Zagoruyko and Nikos Komodakis. Wide residual networks. *arXiv preprint arXiv:1605.07146*, 2016.

[22] Hao Zhang, Feng Li, Shilong Liu, Lei Zhang, Hang Su, Jun Zhu, Lionel M Ni, and Heung-Yeung Shum. Dino: Detr with improved denoising anchor boxes for end-to-end object detection. *arXiv preprint arXiv:2203.03605*, 2022.

[23] Yan Zhang, Xuehong Cui, Yun Liu, and Bin Yu. Tire defects classification using convolution architecture for fast feature embedding. *International Journal of Computational Intelligence Systems*, 11(1):1056–1066, 2018.

[24] Yanjia Zhu, Hongxiang Cai, Shuhan Zhang, Chenhao Wang, and Yichao Xiong. Tinaface: Strong but simple baseline for face detection. *arXiv preprint arXiv:2011.13183*, 2020.