Deep Transfer Learning for Intelligent Vehicle Perception: a Survey

Xinyu Liu[1], Jinlong Li[1], Jin Ma[1], Huiming Sun[1], Zhigang Xu[2], Tianyun Zhang[1], Hongkai Yu[1]*

[1] *Department of Electrical Engineering and Computer Science, Cleveland State University, Cleveland, OH 44115, USA*

[2] *School of Information Engineering, Chang'an University, Xi'an 710064, China*

Abstract

Deep learning-based intelligent vehicle perception has been developing prominently in recent years to provide a reliable source for motion planning and decision making in autonomous driving. A large number of powerful deep learning-based methods can achieve excellent performance in solving various perception problems of autonomous driving. However, these deep learning methods still have several limitations, for example, the assumption that lab-training (source domain) and real-testing (target domain) data follow the same feature distribution may not be practical in the real world. There is often a dramatic domain gap between them in many real-world cases. As a solution to this challenge, deep transfer learning can handle situations excellently by transferring the knowledge from one domain to another. Deep transfer learning aims to improve task performance in a new domain by leveraging the knowledge of similar tasks learned in another domain before. Nevertheless, there are currently no survey papers on the topic of deep transfer learning for intelligent vehicle perception. To the best of our knowledge, this paper represents the first comprehensive survey on the topic of the deep transfer learning for intelligent vehicle perception. This paper discusses the domain gaps related to the differences of sensor, data, and model for the intelligent vehicle perception. The recent applications, challenges, future researches in intelligent vehicle perception are also explored.

*Corresponding author: Hongkai Yu, Email: h.yu19@csuohio.edu.

Keywords: deep transfer learning, domain gap, intelligent vehicle perception, autonomous driving

1. Introduction

In recent years, perception has been viewed as a critical component in intelligent vehicles for precise localization, safe motion planning, and robust control Li et al. (2020a), Yurtsever et al. (2020), Huang and Chen (2020). The perception system provides intelligent vehicles with immediate environmental information about surrounding pedestrians, vehicles, traffic signs, and other items and helps to avoid possible collisions. Therefore, the perception tasks play an indispensable role in intelligent vehicles and autonomous driving Arnold et al. (2019). Recently, the deep learning methods have gained significant traction in the intelligent vehicle perception and have achieved great successes Grigorescu et al. (2020), Wen and Jo (2022), Chen et al. (2022).

However, as shown in Fig. 1, there are lots of complex cases where the deep learning methods might fail in the real world. For example, a deep learning based vehicle detection model pre-trained on sunny weather data might be then tested in the foggy weather or night condition, leading to a large performance drop. This degradation is influenced by the domain gap (shift) between diverse driving environments Hnewa and Radha (2020), Mirza et al. (2022), Mohammed et al. (2020), *e.g.*, different weather and illumination conditions. Moreover, different types and settings of the sensors Rist et al. (2019) installed on vehicles and various deep learning model structures Xu et al. (2023a) Khalil and Mouftah (2022) during the Vehicle-to-Vehicle (V2V) cooperative perception might result in the domain gap as well.

The above mentioned performance drop for intelligent vehicle perception because of the domain gap can be relieved via the Transfer Learning (TL) Zhuang et al. (2020) methods. The TL techniques include two goals: 1) fully using the prior knowledge obtained from the source domain to guide the inference in the related target domain, 2) largely reducing the feature distribution discrepancy caused by the domain gap. Due to these two goals, the performance for deep learning based intelligent vehicle perception systems in related but different domains can be enhanced. The deep learning model's

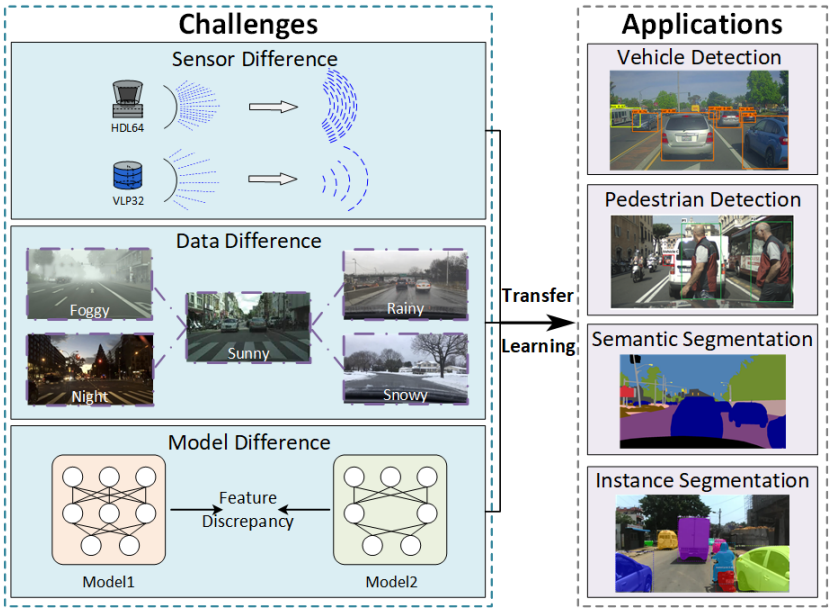

Figure 1: Illustration of Challenges and Applications of Intelligent Vehicle Perception. Transfer Learning (TL) methods can be applied to reduce the domain gaps by sensor difference, data difference, and model difference.

generalization capability can be improved for the intelligent vehicle perception under different challenging scenarios as shown in Fig. 1.

In this paper, we focus on the transfer learning methods for the intelligent vehicle perception in the deep learning era. This paper first reviews the related tasks and benchmark datasets for intelligent vehicle perception, and then classifies the domain gaps to three differences of sensor, data, and model during the vehicle driving. Next, we carefully review about 150 related published papers of deep transfer learning since the deep learning research is started, then we classify the deep learning based transfer learning methods into four types: 1) Supervised TL, 2) Unsupervised TL, 3) Weakly-and-semi Supervised TL, and 4) Domain Generalization. For the first three types, the transfer learning is implemented from one source domain to one target domain, where our classification depends on whether the target domain has labeled data or not. For the last type, the transfer learning is conducted from one source domain to multiple target domains for the generalization in many seen or unseen driving scenarios. In addition,

several subdivisions of each type of transfer learning methods are also reviewed and analyzed in this survey.

The contributions of this paper can be outlined as follows.

- To the best of our knowledge, this paper is the first in-depth survey on the topic of the deep transfer learning for intelligent vehicle perception.
- This paper summarizes the domain gap for intelligent vehicle perception into three types (differences of sensor, data, model) and gives detailed explanations to the related tasks and benchmark datasets.
- After reviewing about 150 related published papers, we classify the deep transfer learning methods for intelligent vehicle perception into four types and explain each of them in details.

The subsequent sections of this paper are structured as follows. Section 2 provides a general overview about the related tasks and benchmark datasets for intelligent vehicle perception. Section 3 presents the domain distribution discrepancy and the three kinds of domain gap. Section 4 details the different methodologies of the deep transfer learning techniques. Sections 5, 6 explain the challenges and future research, followed by a conclusion in Section 7.

2. Intelligent Vehicle Perception

For intelligent vehicles or autonomous driving, perception plays a crucial role in receiving data from sensors and extracting meaningful information from the surrounding environment, so as to make meaningful decisions for the precise motion planning by identifying obstacles, traffic signs/markers, and available driving areas. Two types of mainstream sensors (Camera, LiDAR) are widely used in self-driving or intelligent driving vehicles Cao et al. (2019) Fadadu et al. (2022) Liu et al. (2021b) Liu et al. (2022b) Gholamhosseinian and Seitz (2021). These sensors installed on vehicles are utilized for the intelligent vehicle perception tasks.

The intelligent vehicle perception tasks include discovering the surrounding vehicles and pedestrians, recognizing traffic signs and markers, finding the driving areas (*e.g.*, road regions), and so on. In the real world, sometimes the objects may be similar to each other or the background, and the challenging scenarios (*e.g.*, diverse weather, dark illumination) might affect the performance of sensors, making the perception tasks even more difficult Hnewa and Radha (2020), Li et al. (2023b). This paper groups these intelligent vehicle perception tasks into two classes (Object Detection, Semantic/Instance Segmentation) and further discusses these challenges for intelligent vehicle perception in the real world.

2.1. Object Detection

To achieve autonomous driving safely and successfully, it is necessary to have a reliable object detection system. Considering the complex road conditions, it is essential to detect (localize and recognize) other vehicles, pedestrians, and obstacles to prevent potential accidents. However, detecting objects in urban areas is challenging due to the diverse types of objects and unknown road situations Arnold et al. (2019), Feng et al. (2020).

2D Object Detection: By only using the relatively cheap camera sensor(s), deep learning models can be easily applied to efficiently detect (localize and recognize) the surrounding objects from the 2D image data Yeong et al. (2021). The output will be the identified 2D bounding boxes (2D coordinates) with the recognized object classes for the surrounding objects on each camera image, with a real-time or near real-time inference speed. However, 2D object detection alone can only provide the object's position on a 2D plane, which does not provide enough information Wang et al. (2019b), *e.g.*, object depth, object 3D size.

3D Object Detection: Considering the limitations of 2D object detection, the object 3D information might equip the intelligent vehicle with the capability to more robustly and accurately perceive and recognize surrounding objects. The output will be the identified 3D bounding boxes (3D coordinates) with the recognized object classes for the surrounding objects, with a reasonable inference time. Because the images of camera sensors and the point clouds of LiDAR sensors could provide the depth cues,

the 3D object detection task could be achieved via three sensor settings: 1) Camera only Wang et al. (2023a), 2) LiDAR only Xu et al. (2023b) Xu et al. (2022b), 3) Camera + LiDAR Zhao et al. (2020).

2.2. Semantic/Instance Segmentation

Different with the object detection task, the segmentation task not only discovers the object regions but also give the pixel-level labels (masks) for everything (object and background) in the driving scenarios. For the intelligent vehicle perception, the segmentation task can be classified into two types: Semantic Segmentation, Instance Segmentation.

Semantic Segmentation: Semantic segmentation involves the assignment of a semantic label to every pixel within an image, such as "road", "vehicle" or "pedestrian", "traffic sign", and so on. This technique enables the intelligent vehicle to perceive the surrounding environment and understand the scene more comprehensively Feng et al. (2020), Mo et al. (2022). The identification of specific regions within an image can aid the self-driving vehicles in making informed decisions, *e.g.*, determining where the driving road region is.

Instance Segmentation: Instance segmentation outputs the boundaries (pixel-level masks) of each object and assigns a unique label to each discovered object Zhou et al. (2020a), which seems like a integration of object detection and semantic segmentation. It is particularly useful for identifying the shape, location, and number of surrounding objects in autonomous driving Rashed et al. (2021), Ko et al. (2021).

2.3. Benchmark dataset

Based on different sensor types, the intelligent vehicles could have the image data from the camera sensor and the point cloud data from the LiDAR sensor.

Camera data: The 3-channel color images in Green, Red, Blue primary colors of light (*i.e.*, RGB images) are commonly acquired by monocular or multiple cameras, which are simple and reliable sensors that closely resemble human eyes Feng et al. (2020). One of the main benefits of RGB cameras is their high resolution and relatively low cost. However, their performance can deteriorate significantly under the challenging weather and illumination conditions Feng et al. (2021).

LiDAR data: Unlike cameras, laser sensors offer direct and precise 3D information, making it easier to extract object candidates and aiding in the classification task by providing 3D shape information. LiDAR, also known as light detection and ranging, is a sensor technology that is capable of detecting targets in all lighting conditions and creating a distance map of the targets with high spatial coverage Li and Ibanez-Guzman (2020), Li et al. (2020b). LiDAR could work in some challenging weather and dark illumination scenarios, but it is quite expensive with high cost. Its high cost is a major obstacle to wider adoption Li et al. (2020b) Pham et al. (2020).

Benchmark for 2D Object Detection: KITTI Geiger et al. (2013), Cityscapes Cordts et al. (2016) , SIM10k Johnson-Roberson et al. (2016), Foggy Cityscapes Sakaridis et al. (2018), Syn2Real-D Peng et al. (2018), BDD100k Yu et al. (2018), GTA5 Richter et al. (2016), nuScenes Caesar et al. (2020), Waymo Open Sun et al. (2020), A*3D Pham et al. (2020), ApolloScape Huang et al. (2018), Ford Agarwal et al. (2020), A2D2 Geyer et al. (2020), ONCE Mao et al. (2021), and Automine Li et al. (2022c).

Benchmark for 3D Object Detection: KITTI Geiger et al. (2013), Cityscapes Cordts et al. (2016), Foggy Cityscapes Sakaridis et al. (2018), GTA5-LiDAR Wu et al. (2019), nuScenes Caesar et al. (2020), Waymo Open Sun et al. (2020), A*3D Pham et al. (2020), ApolloScape Huang et al. (2018), Ford Agarwal et al. (2020), A2D2 Geyer et al. (2020), ONCE Mao et al. (2021), Automine Li et al. (2022c), OPV2V Xu et al. (2022b) and V2V4Real Xu et al. (2023b).

Benchmark for Semantic Segmentation: KITTI Geiger et al. (2013), Cityscapes Cordts et al. (2016), Waymo Open Sun et al. (2020), ApolloScape Huang et al. (2018), BDD100k Yu et al. (2018), and A2D2 Geyer et al. (2020).

Benchmark for Instance Segmentation: Cityscapes Cordts et al. (2016), nuScenes Caesar et al. (2020), BDD100k Yu et al. (2018), and KITTI-360 Liao et al. (2022).

The Table 1 summarizes the current widely-used benchmark dataset details for the intelligent vehicle perception tasks, including the image resolution, image numbers, LiDAR frame numbers, task types, real or synthetic information of each benchmark dataset.

Benchmark	Image Resolution	Image #	LiDAR Frame #	Tasks	Real/Syn
KITTI Geiger et al. (2013)	1,392×512	15K	1.3M	D, S	R
Cityscapes Cordts et al. (2016)	2,048×1,024	25K	-	D, S	Syn
SIM10k Johnson-Roberson et al. (2016)	1,914×1,052	10K	-	D	Syn
Foggy Cityscapes Sakaridis et al. (2018)	2,048×1,024	3,475	-	D, S	Syn
Syn2Real-D Peng et al. (2018)	-	248K	-	D	Syn, R
BDD100K Yu et al. (2018)	1,280×720	8K	-	D, S	R
GTA Richter et al. (2016)	1,914×1,052	24,966	-	S	Syn
GTA-LiDAR Wu et al. (2019)	64×512	100K	-	S	Syn
H3D Patil et al. (2019)	1,920×1,200	27,721	-	D	R
nuScenes Caesar et al. (2020)	1,600×900	40K	-	D	R
Waymo Open Sun et al. (2020)	1,920×1,280	600K	-	D, S	R
ApolloCar3D Song et al. (2019b)	3,384×2,710	5,277	-	D, S	R
A*3D Pham et al. (2020)	2,048×1,536	39K	39,179	D	R
ApolloScape Huang et al. (2018)	3,384×2,710	143,906	-	S	R
SYNTHIA Ros et al. (2016)	960×720	13.4K	-	S	Syn
Lyft Level 5 Houston et al. (2021)	-	55K	-	S	R
Ford Agarwal et al. (2020)	-	200K	-	D	R
A2D2 Geyer et al. (2020)	1,928×1,208	12K		D, S	R
ONCE Mao et al. (2021)	1,920×1,020	1M	-	D	R
AutoMine Li et al. (2022c)	2,048×1,536	18K	-	D	R
OPV2V Xu et al. (2022b)	800×600	44K	11K	D	Syn
V2V4Real Xu et al. (2023b)	2,064×1,544	40K	20K	D	R

Table 1: Benchmark datasets for intelligent vehicle perception. D: object detection in 2D or 3D, S: semantic or instance segmentation, Syn: synthetic data, R: real data.

3. Domain Distribution Discrepancy

Despite the remarkable achievements of the intelligent vehicle perception algorithms on benchmark datasets, there are still significant challenges in the real world due to the large variations in the sensor types and settings, data in diverse style, environment, weather and illumination, trained epoch, and architecture Li et al. (2022b), Feng et al. (2021), Schutera et al. (2020), Song et al. (2023). Based on these observations, we divide the domain distribution discrepancy for intelligent vehicle perception into three types: sensor difference, data difference, and model difference, as shown in Table 2.

3.1. Sensor difference

First of all, the domain gap shows up when the sensors are different in types and settings. Let us explain the sensor difference for camera and LiDAR separately. The camera sensor is cheap but not robust to different types and settings, for example, angle difference from horizontal to oblique Rist et al. (2019), placement dissimilarity from front view to rear view Alonso et al. (2020), image resolution diversity Carranza-García

et al. (2020), and so on. The LiDAR sensors might also have different types and settings, for example, different laser beam numbers Yi et al. (2021), various LiDAR equipment from different companies Xu et al. (2023a), LiDAR placement dissimilarity Hu et al. (2022a), and so on. These real-world challenges due to the sensor difference may generate the heterogeneous feature distribution between different domains Triess et al. (2021), Zhou et al. (2022b), Chakeri et al. (2021).

3.2. Data difference

In addition, the domain gap exists when the data itself is different in style and format (*e.g.*, transfer learning from synthetic to real), or the data collected by the sensors are different. 1) Researchers are recently interested in learning the prior knowledge from synthetic data to help the learning on real data. The synthetic data is normally generated by computer game engines, like SYNTHIA Ros et al. (2016), GTA5 Richter et al. (2016). Although utilizing synthetic data has been becoming a popular alternative solution, models trained with synthetic data still suffer from the incapability of generalization in the real world Wu et al. (2019), Yue et al. (2019). 2) The data collected by the sensors in different urban or highway environments Shenaj et al. (2023), diverse weather (foggy, rainy, snowy, sunny, *etc*.) Miglani and Kumar (2019), Xu et al. (2021), Mirza et al. (2022), Bogdoll et al. (2022), Li et al. (2023b), dissimilar illumination conditions (daytime, nighttime, tunnel, *etc*.) Wu et al. (2021) might result in the heterogeneous data distribution between different domains.

3.3. Model difference

Finally, the model difference is also one possible reason of the domain gap. When the perception architecture is the same, diverse models may still exist because of different trained epochs You et al. (2022). When the perception architecture is the same, diverse models may exist due to different CNN frameworks Xu et al. (2021). When the perception architecture is diverse, the model is obviously different, for example, from CNN architecture to transformer architecture Sun et al. (2022). The detection features extracted from diverse deep learning models are quite different, as shown in Xu et al. (2023a), leading to the heterogeneous feature distribution between different domains.

Types	Differences	Examples
Sensor Difference	Setup Placement Angle	64-beam LiDAR → 32-beam LiDAR Yi et al. (2021) Front → Rear Alonso et al. (2020) Horizontal → Oblique Rist et al. (2019)
Data Difference	Synthetic/Real Environment Weather Illumination	GTA5 → Cityscapes Murez et al. (2018) KITTI → Cityscapes He and Zhang (2019) Cityscapes → Foggy Cityscapes Li et al. (2023b) Cityscapes → Dark Zurich Wu et al. (2021)
Model Difference	Epoch Old/Upgraded Architecture	Epoch 50 → Epoch 80 You et al. (2022) PointPillars → PV-RCNN Xu et al. (2021) CNN → Transformer Sun et al. (2022)

Table 2: Domain distribution discrepancy with three types of differences for the intelligent vehicle perception: sensor, data and model. "→" means the model training with the left data and testing on the right data.

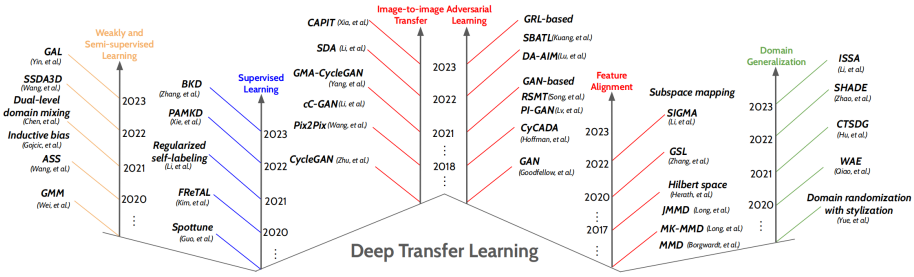

Figure 2: The Chronological Overview of Transfer Learning Research in the Deep Learning Era.

4. Deep Transfer Learning Methodology

With the rapid advancement of autonomous driving techniques, there is now an abundance of driving scene images available. Deep learning methods are booming in the application of autonomous driving with high performance of perception. This paper is focused on the transfer learning methods for the intelligent vehicle perception in the deep learning era.

Transfer Learning (TL) is a machine learning method to largely apply the knowledge acquired from one task or domain to another related task or domain Zhuang et al. (2020). This paper classifies the deep transfer learning into several main types: Supervised TL, Unsupervised TL, Weakly-and-semi Supervised TL, Domain Generalization. The chronological overview of the transfer learning research development in the deep learning era is shown in Fig. 2.

4.1. Supervised TL

In the transfer learning research, the source domain normally has the manually annotated ground truth. If the target domain also has the manually annotated ground truth, the machine learning technique that transfers knowledge from a labeled source domain to the labeled target domain is named as Supervised TL Drews et al. (2017), Yu et al. (2018), Zhou et al. (2019). Gathering such manually annotated data requires substantial human involvement, which is labor-intensive and time-consuming Carvalho et al. (2015).

We divide the Supervise TL methods into Fine-tuning and Knowledge distillation via teacher-student network in this paper.

4.1.1. Fine-tuning

Fine-tuning is a common technique in the transfer learning Guo et al. (2019b), Li and Zhang (2021), Hu et al. (2022a), which has been widely used in intelligent vehicle perception Wang et al. (2019b), Luo et al. (2021), Liang et al. (2022), Xu et al. (2019), Doan et al. (2019). Fine-tuning takes an existing neural network model pre-trained on a source domain dataset and further trains it on a new target domain dataset. By the fine-tuning, the knowledge learned from the source domain can be leveraged to improve the performance on the target domain. It is worth mentioning that fine-tuning a pre-trained neural network model could obtain better performance than directly training from scratch. Typically, the pre-trained neural network model is trained on a large-scale dataset, enabling to acquire the knowledge from a wide range. The learning rate of fine-tuning on the target domain is relatively small as a fine adjustment for the neural network model pre-trained on source domain.

The fine-tuning based transfer learning is a simple but effective way to transfer the knowledge gained from the pre-training on source domain to enhance the performance on the target domain with less data and computational resources than training from scratch. However, as a supervised method, fine-tuning requires the manually annotated ground truth on the target domain, which might be not available in some real-world applications.

The fine-tuning methods could be roughly classified into two types: 1) *Whole Fine-tuning*: it trains all the layers of the whole neural network model. 2) *Partial Fine-tuning*: it allows us to train only the some interested layers of the pre-trained neural network while keeping the some layers frozen.

Whole Fine-tuning: All the layers of the entire neural network model are fine-tuned to obtain the spatial–temporal interactions Ye et al. (2021) among autonomous vehicles and the 3D perception in autonomous driving Sautier et al. (2022).

Partial Fine-tuning: Guo et al. (2018) only fine-tunes the encoder-decoder based semantic segmentation model, by fixing a pre-trained sub-network to ensure the multi-class boundary constraint.

4.1.2. Knowledge distillation via teacher-student network

Knowledge distillation Hinton et al. (2015), Gou et al. (2021), Wang and Yoon (2021), Chen et al. (2021a), Xie and Du (2022), Beyer et al. (2022) is an advanced technique in deep learning, which is also referred to as teacher-student learning, where a student neural network is trained on target domain to imitate the knowledge of a teacher neural network trained on source domain. Knowledge distillation has been widely utilized in intelligent vehicle perception Kothandaraman et al. (2021), Gao et al. (2022), Hou et al. (2022), Yang et al. (2022), Sautier et al. (2022).

Knowledge distillation could be beneficial to model generalization, model compression, model transferability. It improves the model generalization so that the student network can generalize better on unseen examples, especially in scenarios with limited training data. It allows to compress a large teacher model into a smaller student model. It enables the knowledge transferability from the teacher model (source domain) to the student model (target domain) even with different deep learning architectures. The teacher network is typically trained on a large-scale dataset for the next knowledge transferability to the student network, however the large-scale dataset might be not available in the source domain of some intelligent vehicle perception tasks.

Inspired by Lan and Tian (2022), the knowledge distillation methods could be roughly classified into two types: 1) *Response Knowledge Distillation*: It focuses on the final output layer of the teacher model so as to teach a student model to mimic its

predictions. The core concept is to use a loss function called the distillation loss, which measures the difference between the output activations of the student and teacher models. By minimizing this loss during training, the student model gradually improves its ability to generate predictions that closely resemble those of the teacher model. 2) *Intermediate Knowledge Distillation*: It focuses on aligning the intermediate representations of the teacher and student models. The intermediate layers learn to recognize and distinguish specific features in the data, and this knowledge distilled in teacher network can be leveraged to train the student model effectively.

Response Knowledge Distillation: Gao et al. (2022) proposes the cross-domain correlation distillation loss to transfer knowledge from daytime to nighttime domains, thereby improving nighttime semantic segmentation performance.

Intermediate Knowledge Distillation: Hou et al. (2022) proposes an approach of transferring distilled knowledge from a larger source teacher model to a smaller target student network to conduct LiDAR semantic segmentation. Specifically, the intermediate Point-to-Voxel Knowledge Distillation approach is utilized to transfer latent knowledge from both point level and voxel level to complement sparse supervision signals.

4.2. Unsupervised TL

In the intelligent vehicle perception, data labeling is a time-consuming and labor-intensive process in real-world scenarios. Generally, supervised algorithms struggle when there is a scarcity of labeled data in the source domains Niu et al. (2020), Pan and Yang (2010). To overcome these challenges, Unsupervised Transfer Learning (TL) has emerged as a promising approach for addressing such specific cases in the intelligent vehicle perception tasks. Unsupervised TL refers to a scenario where there is unlabeled target data besides labeled data available in source domain. Unsupervised TL approaches offer promising solutions to overcome the limitations of limited labeled data availability, enabling more efficient and effective perception in intelligent vehicles.

In this survey, the Unsupervised TL methods are divided into four types: image-to-image transfer, adversarial learning, feature alignment, self-learning. They are explained in details as below.

4.2.1. Image-to-image transfer

Image-to-image transfer, also known as image-to-image translation, is a computer vision task that involves converting an input image to a different domain. It aims to establish a learned correspondence between two visual domains, where the input image originates from the source domain, while the desired output image that resembles the target domain. The goal is to generate a corresponding image with similar style of the target domain and simultaneously preserve the relevant characteristics and semantic contents of the input image. It has found extensive application in the field of autonomous driving as well as intelligent transportation systems, including semantic segmentation Murez et al. (2018), Pizzati et al. (2020), lane recognition Hou et al. (2019), Liu et al. (2021a), data augmentation Zhang et al. (2022), Yang et al. (2020) Mușat et al. (2021) and object detection Schutera et al. (2020), Li et al. (2021, 2022b), Shan et al. (2019).

Image-to-image transfer neural networks are commonly implemented using two different approaches: 1) *Paired Image-to-Image Transfer* and 2) *Unpaired Image-to-Image Transfer*. The first approach utilizes generative adversarial networks trained on paired images Wang et al. (2018). This type of network learns a mapping that transforms an input image from its original domain to desired output domain Isola et al. (2017). The second approach addresses scenarios where unpaired images are used to establish a more general framework Zhu et al. (2017), Park et al. (2020), inspiring the unsupervised image-to-image translation methods Liu et al. (2017), Baek et al. (2021).

Paired Image-to-Image Transfer: Isola et al. (2017) investigated the utilization of conditional Generative Adversarial Networks (GAN) namely pix2pix for paired image-to-image translation tasks Hao et al. (2019). The GAN with condition learns a generative model of data but with the added condition of an input image to produce a corresponding output image. This approach strives to produce plausible images in target domain. The adversarial loss is utilized to train a Generator Network which is updated using l_1 loss, which quantifies the disparity between the generated image as well as predicted output. By incorporating additional loss, the Generator Network can produce plausible translations of the source images. Conversely, the Discriminator Network is

designed to perform generated image classification. With the paired training data, these methods could translate the image of similar styles in different domains. However, in practical applications of intelligent vehicle perception, the requirement for paired training data poses a limitation.

Unpaired Image-to-Image Transfer: Cycle-consistency GAN (CycleGAN) Zhu et al. (2017) is a type of GAN model that enables image translation between unpaired datasets Mușat et al. (2021), Uricar et al. (2021), Shan et al. (2019), Liu et al. (2022a). The training process of a CycleGAN involves optimizing two generators and two discriminators simultaneously. One generator is responsible for learning the mapping function G from domain $Xto(\rightarrow)Y$, while the other generator F learns the mapping from domain $Yto(\rightarrow)X$.

Both G and F are trained simultaneously, incorporating a cycle consistency loss that enforces the cycle consistency to ensure that $F(G(x)) \approx x$ and $G(F(y)) \approx y$. This loss combined with adversarial losses on domains X and Y yields objective for unpaired image-to-image translation. Unpaired Image-to-Image Transfer release the requirement of paired training data, which is more general in the real-world applications of intelligent vehicle perception. By incorporating adversarial losses on domains X and Y, the objective for unpaired image-to-image translation is obtained. Unpaired Image-to-Image Transfer release the need for paired training data, making them more general in real-life applications of intelligent vehicle perception.

However, these image-to-image transfer approaches rely on task-specific and predefined similarity functions between inputs and outputs and do not consider the reliability and robustness of the translation frameworks, which might be disrupted by the perturbations added to input and targeted images. This issue is particularly crucial for autonomous driving.

4.2.2. Adversarial learning

Adversarial learning refers to a machine learning technique that involves training two neural networks in a competitive manner, which is initially introduced in the context of Generative Adversarial Networks (GAN) by Goodfellow et al. (2020) and also mentioned in Gradient Reversal Layer (GRL) framework Ganin and Lempitsky (2015),

and provides a promising approach for generating target-similar samples at the pixel-level or target-similar representations at the feature-level by training robust deep neural networks. It has become popular for addressing transfer learning challenges by minimizing the domain discrepancy using adversarial objectives, such as fooling a domain discriminator/classifier. During training, the feature extractor and the domain discriminator are engaged in an adversarial game. The feature extractor tries to produce representations that confuse the domain discriminator, making it difficult for the discriminator to differentiate between the domains. Meanwhile, the objective of the domain discriminator is to correctly classify the samples into their respective domains. This adversarial process encourages the learning of domain-invariant features by the feature extractor, thereby minimizing the differences between domains. By minimizing the domain disparities through adversarial learning, the model learns representations that capture the underlying domain-invariant information shared across domains. This approach helps to address TL challenges by effectively reducing the disparities between two different domains, improving the model's generalization capabilities across different domains.

The adversarial learning based transfer learning methods for intelligent vehicle perception consists of two types: 1) *GRL based Methods* and 2) *GAN based Methods*. The first method relies on minimizing the domain distribution discrepancy through a gradient reversal in the back propagation of feature extraction to confuse the domain discriminator. In contrast, the last one focuses on training the Generator Network and Discriminator Network alternately using a Min-Max adversarial loss function, with the goal of acquiring domain-invariant features.

GRL based Methods: Domain adaptation in different vehicle perception domains can be achieved through the addition of a Gradient Reversal Layer (GRL) to the deep learning architecture Xu et al. (2023a), Li et al. (2023b). The mechanism of domain adversarial embedding involves using a discriminator with a GRL to differentiate between samples from two domains. The discriminator is a binary classifier, while the GRL can reverse the training gradient in the back propagation of feature extraction. Both the discriminator and the GRL work together to align the feature distributions across different domains. It is worth mentioning that the GRL only comes into ef-

fect during the backpropagation phase and does not affect the forward propagation process Ganin and Lempitsky (2015). Let us give a detailed example for better understanding. Li et al. (2023b) introduces a new framework for domain adaptive object detection in autonomous driving during challenging foggy weather. The approach addresses the domain gap between clear and foggy weather in vehicle driving by incorporating image-level and object-level adaptation techniques, which aim to minimize differences in object appearance and image style. Additionally, a novel Adversarial Gradient Reversal Layer (AdvGRL) has been proposed to enable adversarial mining for difficult examples along with domain adaptation.

GAN based Methods: GAN Goodfellow et al. (2020), Song et al. (2020) is a popular deep learning framework that can be used to teach a model to capture the distribution patterns present within the training data, enabling the generation of new data from that same distribution. A GAN consists of two separate models, namely the generator G and the discriminator D. The generator G's job is to create "fake" images that resemble the training images so as to confuse the discriminator D. The applications of GAN in autonomous driving have been recently explored owing to its remarkable progress in generating realistic images. Specifically, GAN has been leveraged to generate image or subspace feature undistinguished by domain classifier based discriminator, for example, GAN could generate aligned/similar features between clear weather and foggy weather Li et al. (2023b, 2022a), between synthetic game data and real-world data Biasetton et al. (2019), Zhang et al. (2021b), between daytime data and nighttime data Wang et al. (2022a), Li et al. (2022a). Let us give a detailed example for better understanding. Hoffman et al. (2018) proposes a domain adaptation model which combines generative image space alignment, latent feature space alignment, and the vehicle perception task. By considering the vehicle perception task (semantic segmentation of urban driving scenes), the image-level features, latent features, and the task-related semantic features are aligned across different domains by an adversarial learning via a GAN-based framework.

4.2.3. Feature alignment

To minimize the domain distribution discrepancy, the objective of feature alignment in transfer learning is to discover an aligned feature representation from multiple domains. Typically, the feature distribution difference between different domains can be defined as loss functions during the deep neural network training, so minimizing the loss functions of the feature distribution difference across multiple domains will reduce the domain gap.

Feature alignment-based transfer learning can be classified into two main categories: 1) *Subspace Feature Alignment*, and 2) *Attention-guided Feature Alignment*. The first one focuses on aligning the feature distribution in the lower-dimensional subspace representation by using different metrics of distribution distances. The second one uses the attention mechanism to extract the attention maps first and then enforce the attention maps from multiple domains to be the same.

Subspace Feature Alignment: By projecting the features from different domains to a lower-dimensional subspace, several metrics to describe the distance of feature distribution across source and target domains can be defined as the loss functions in the deep learning framework. Minimizing these metric distances (loss functions) will align the features of different domains in the subspace. The widely used metric to describe the feature distribution distances are Principal Component Analysis (PCA) projected subspace feature distance Song et al. (2019a), Maximum Mean Discrepancy (MMD) Borgwardt et al. (2006), Kullback–Leibler Divergence Zhang et al. (2018), Gram Matrix Guo et al. (2019a), Multi-Kernel MMD Gretton et al. (2012), Long et al. (2015), Joint MMD Long et al. (2017), Wasserstein distance Arjovsky et al. (2017), *etc*. For example, let us take a close look at the definition of the MMD metric, which is formulated as

$$MMD(\mathcal{X}_s, \mathcal{X}_t) = \| \frac{1}{n_s} \sum_{i=1}^{n_s} k(\mathbf{x}_i^s) - \frac{1}{n_t} \sum_{j=1}^{n_t} k(\mathbf{x}_j^t) \|_H, \tag{1}$$

where $\mathcal{X}_s$ and $\mathcal{X}_t$ denote the sets of samples obtained from the source and target domains, $\mathbf{x}_i^s$ and $\mathbf{x}_j^t$ are individual samples from the respective domains, and n_s and n_t denote the sample sizes of the source and target domains respectively, k denotes the

kernel functions, and H indicates the Reproducing Kernel Hilbert Space (RKHS).

Attention-guided Feature Alignment: Taking inspiration from the attention mechanism Zhou et al. (2016), Vaswani et al. (2017), the most informative components of specific importance can be focused for the intelligent vehicle perception. The deep learning frameworks can first extract the attention maps, then the distance of attention maps between two domains can be defined as loss function to be minimized during the neural network training Zhou et al. (2020b), Zagoruyko and Komodakis (2016). By employing this approach, it becomes possible to align the feature distribution across both the source and target domains via the attention map consistency constraint. For example, in Cho et al. (2023), the relation-aware knowledge captured by multiple detection heads can be transferred using a specially designed attention head loss for the improved LiDAR-based 3D object detection in the context of autonomous driving.

4.2.4. Self-learning

Autonomous vehicles continuously collect unlabeled data during their operation, creating an opportunity for self-learning Liu et al. (2021c), Zhang et al. (2021a), Kumar et al. (2021), Luo et al. (2021), Ziegler and Asano (2022), which offers a promising approach to reduce the reliance on labeled data and enhance model flexibility. Given the absence of labeled data in target domain using Unsupervised TL, the self-learning methods use the additional cues to evaluate the neural network prediction in an unsupervised setting, so some prediction results with high confidence are used as the pseudo-labels in the further training or testing.

The following shows some representative examples of self-learning methods for the Unsupervised TL based intelligent vehicle perception. The entropy based uncertainty can be used to define the hardness of a specific training sample so as to implement an easy-to-hard curriculum learning for semantic segmentation Pan et al. (2020). Wang et al. (2021a) utilizes self-supervised learning to enhance the semantic segmentation performance by using depth estimation as guidance to overcome the domain gap between the source and target domains. They explicitly capture the correlation between task features and use target depth estimation to enhance target semantic predictions. The adaptation difficulty, as inferred from depth information, is subsequently utilized

to enhance the quality of pseudo-labels for target semantic segmentation. Shin et al. (2022) proposes a multi-modal extension of test-time adaptation in the context of 3D semantic segmentation. To improve the unstable performance of models at test time, they design both intra-modal and inter-modal modules together to acquire more dependable self-learning signals of pseudo-labels. Zhang et al. (2021a) utilizes the multiple classifiers with attention heads to evaluate the uncertainty associated with the pseudo-labels. The panoramic pseudo-labels with high confidences are then used to improve the panoramic semantic segmentation prediction in an iterative fashion.

By leveraging self-learning in autonomous driving, the need for extensive manual annotation of data is reduced, enabling more cost-effective and efficient training of models. The iterative process of incorporating high-confidence classified samples and generating pseudo-labels facilitates the development of a promising classifier using only unlabeled target domain data. Meanwhile, the robustness and convergence of the self-learning methods is still an open question for the reliable intelligent vehicle perception.

4.3. Weakly-and-semi Supervised TL

Although impressive results have been achieved by unsupervised TL methods, the domain gap cannot be completely eliminated due to the lack of supervision on the target domain. There is still a relative performance gap compared with supervised TL methods. Another way in addressing the domain gap is by using the weakly-and-semi supervised learning method that utilizes both weakly labeled and some labeled/unlabeled data in target domain.

By involving some supervisions in the target domain, the weakly-and-semi supervised learning methods could achieve a better performance than the unsupervised TL methods while still worse than the supervised TL methods. While various methods have been proposed for weakly-and-semi supervised transfer learning, how to leverage the unlabeled target data with the help of available labeled data under different situations is still a challenging open question.

Based on the available supervision, the weakly-and-semi supervised transfer learning methods could be roughly classified into two types: 1) *Weakly-Supervised TL*:

There are only weakly supervised labels in the target domain. 2) *Semi-Supervised TL*: There are only semi-supervised labels in the target domain, including some labeled data and the remaining unlabeled data on target domain.

Weakly-Supervised TL: Theories of weakly supervised learning have been applied in autonomous driving Barnes et al. (2017), Gojcic et al. (2021), such as object detection, semantic segmentation, and instance segmentation. The transfer learning techniques can be applied simultaneously with the weakly supervised learning. For example, when an instance-level task only has image-level annotations in target domain but with instance-level annotations in source domain, the pseudo annotations can be predicted Inoue et al. (2018) for the object detection task. Given a source domain (synthetic data) with pixel/object- level labels, a target domain (real-world scenes) might only have object-level labels, where the pixel-level and object-level domain classifiers can be used in transfer learning to learn domain-invariant features for the semantic segmentation task in driving scenes Wang et al. (2019a).

Semi-Supervised TL: There are three types of training data (labeled source data, labeled target data, and unlabeled target data) in the semi-supervised TL setting Wang et al. (2020), Chen et al. (2021b), Wang et al. (2022c). The key point for improving semi-supervised TL is to effectively use available unlabeled data from target domain and limited labeled data from different domains. For example, Wang et al. (2020) aligns feature distribution across two domains by introducing an extra semantic-level adaptation module, which leverages a few labeled images from the target domain to supervise the segmentation and feature adaptation tasks. Other works focus on generating pseudo labels for unlabeled target data by using labeled source data and labeled target data. For example, Wang et al. (2022c) solves this problem by two-stage learning that includes inter-domain adaptation stage and intra-domain generalization stage. While Chen et al. (2021b) uses the domain-mixed teacher models and knowledge distillation to train a good student model, then the good student model will generate pseudo labels for the next round of teacher model training.

4.4. Domain generalization

Domain Generalization (DG) for intelligent vehicle perception offers a solution to the challenge of enhancing the resilience of deep neural networks against arbitrary unseen driving scenes Zhou et al. (2022a). Unlike Domain Adaptation (DA), DG methods typically focus on learning a shared representation across multiple source domains. This approach aims to enhance the model ability to generalize across various domains, enabling it to perform well in an unknown target domain of driving. Nevertheless, the collection of multi-domain datasets is a laborious and costly endeavor, and the efficacy of DG methods is significantly influenced by the quantity of source datasets Wang et al. (2022b).

The concept of domain generalization (DG) has emerged as a solution to address the lack of target data in domain gap Blanchard et al. (2011) Wang et al. (2022b). The primary distinction between DA and DG lies in the fact that DG does not require access to the target domain during the training phase. DG aims to develop a model by using data from one or multiple related but distinct source domains to generate any out-of-distribution target domain data Shen et al. (2021). The existing methods for DG can be divided into two main groups according to the number of source domains: *Multi-source DG* and *Single-source DG*.

Multi-source DG: Its primary motivation is to utilize data from multiple sources to learn representations that are invariant to different marginal distributions Wilson and Cook (2020) Luo et al. (2022) Zhao et al. (2022). Due to the absence of target data, it is challenging for a model trained on a single source to achieve generalization effectively. By leveraging multiple domains, a model can discover stable patterns across the source domains, leading to better generalization results on unseen domains. The underlying concept behind this category is to minimize the difference between the representations of various source domains, thus learn domain-invariant representations Yue et al. (2019), Hu et al. (2022b), Xu et al. (2022a), Li et al. (2022b), Choi et al. (2021), Lin et al. (2021), Acuna et al. (2021).

Single-source DG: It assumes that the training data is homogeneous, which is sampled from a single domain Qiao et al. (2020), Wang et al. (2021b). Single-source DG methods revolve around data augmentation, they aim to create samples that are out of

the domain and utilize them to train the network in conjunction with the source samples, enhancing the generalization capability Li et al. (2023c), Lehner et al. (2022), Hu et al. (2022b), Khosravian et al. (2021), Chuah et al. (2022), Sanchez et al. (2022), Zhang et al. (2020), Wu and Deng (2022). Although single-source DG methods are not robust as multi-source domain method due to the limited information from source domain, they do not rely on domain labels for learning, which makes them applicable to both single-source and multi-source scenarios.

5. Challenges

This section outlines the main challenges of the deep transfer learning for the current intelligent vehicle perception as below.

- **Sensor Robustness**: The current camera and LiDAR sensors are not robust enough in the extreme driving scenarios, like diverse weather, dark illumination, various environments. In addition, for the V2V cooperative perception, the V2V communication sensors might have the issues of lossy communication Li et al. (2023a) due to the fast speed, obstacles, *etc* Schlager et al. (2022a).

- **Methodology Limitation**: The current unsupervised transfer learning methods are worse than the supervised transfer learning methods with a relative performance insufficiency. In addition, how to fully utilize the knowledge of the source domain and the human prior cognition and experience is still a question to be answered. How to effectively use the weakly and partially labeled data is still a open question.

- **Realism of Synthetic Data**: By eliminating the need for manual annotation, the synthetic data generated by computer game engines is quite helpful to improve the training data size, but it still has significant differences with the real-world data in styles, lighting conditions, viewpoints, and vehicle behaviors, *etc*.

- **Scarcity of Annotated Benchmarks in Complex Scenarios**: There are infinite complex scenarios in the real-world driving, but the current benchmark

datasets in the complex driving scenarios are still limited. For example, the Foggy Cityscapes dataset Sakaridis et al. (2018) only has 2,975 training images during the foggy weather, whose small size poses a clear hurdle for the accurate perception of the intelligent vehicle in the foggy weather.

- **International Standards for Hardware Sensors**: The hardware sensors might be provided from multiple companies of different countries, but there are no unified international standards for the hardware sensors for intelligent vehicle perception. For example, the different hardware sensor types and settings will enlarge the domain gap in different environments.

- **International Standards for Software Packages**: The software package might be provided from multiple companies of different countries as well, but there are no unified international standards for the software packes for intelligent vehicle perception. For example, sharing the features of models trained in different epochs, *e.g.*, from different companies, will result in the performance drop in V2V cooperative perception Xu et al. (2023a).

6. Future Research

This section describes the future research directions of the deep transfer learning for the current intelligent vehicle perception as below.

- **Improving Sensor Robustness**: More future research can be focused on improving the sensor robustness, for example, the camera and LiDAR sensors in diverse weather, dark illumination, various environments, and the communication sensors in the V2V system Tahir et al. (2021).

- **Developing More Advanced Methodologies**: Researchers could make efforts to develop more advanced deep transfer learning methods in the future, for example, largely reducing the performance disparity between unsupervised and supervised approaches, incorporating the Vehicle-to-Everything (V2X) techniques to communicate with connected vehicles and smart infrastructures, involving the

Large Language Models, like ChatGPT Gao et al. (2023), to better simulate the human cognition and knowledge so as to guide the transfer learning methods, accurately self-learning the unlabeled data, effectively and efficiently using the weakly and partially supervised data.

- **Enhancing Realism of Synthetic Data**: The realism of the synthetic data can be improved by more advanced computer game engines in the future. The customized synthetic data can be better simulated via a digital twin simulation system Wang et al. (2023b).

- **Encouraging Benchmarks in Complex Scenarios**: We expect that more high-quality benchmark datasets in complex driving scenarios could be collected and publicized in the future. We also encourage more advanced physical models to simulate the benchmark data (Camera, LiDAR) in complex driving scenarios in the future, such as simulating the fog, rain, snow, lighting changes, *etc*.

- **Promoting International Standards for Hardware Sensors**: We hope that the multiple companies of different countries can collaborate together to promote the international standards for hardware sensors in the future, including the types, settings, parameters of the hardware sensors in different driving environments Schlager et al. (2022b), Masmoudi et al. (2021).

- **Promoting International Standards for Software Packages**: The multiple companies of different countries are expected to collaborate together to promote the international standards for software packages in the future, including the deep learning model architectures and frameworks, hyper parameters, privacy and safety preservation, *etc*.

7. Conclusion

In this survey paper, we presented a comprehensive review of deep transfer learning for intelligent vehicle perception. We reviewed the perception tasks and the related benchmark datasets and then divided the domain distribution discrepancy of the intelligent vehicle perception in the real world into sensor, data, and model differences. Then,

we provided clearly classified and summarized definition and description of numerous representative deep transfer learning approaches and related works in intelligent vehicle perception. Through our intensive analysis and review, we have identified several potential challenges and directions for future research. Overall, this survey paper aims to make contributions to introduce and explain the deep transfer learning techniques for intelligent vehicle perception, offering invaluable insights and directions for the future research.

8. Acknowledgement

This work was supported by NSF 2215388 and CSU FRD grants.

References

D. Acuna, J. Philion, and S. Fidler. Towards optimal strategies for training self-driving perception models in simulation. *Advances in Neural Information Processing Systems*, 34:1686–1699, 2021.

S. Agarwal, A. Vora, G. Pandey, W. Williams, H. Kourous, and J. McBride. Ford multi-av seasonal dataset. *The International Journal of Robotics Research*, 39(12): 1367–1376, 2020.

I. Alonso, L. Riazuelo, L. Montesano, and A. C. Murillo. Domain adaptation in lidar semantic segmentation by aligning class distributions. *arXiv preprint arXiv:2010.12239*, 2020.

M. Arjovsky, S. Chintala, and L. Bottou. Wasserstein generative adversarial networks. In *International conference on machine learning*, pages 214–223. PMLR, 2017.

E. Arnold, O. Y. Al-Jarrah, M. Dianati, S. Fallah, D. Oxtoby, and A. Mouzakitis. A survey on 3d object detection methods for autonomous driving applications. *IEEE Transactions on Intelligent Transportation Systems*, 20(10):3782–3795, 2019.

K. Baek, Y. Choi, Y. Uh, J. Yoo, and H. Shim. Rethinking the truly unsupervised image-to-image translation. In *IEEE/CVF International Conference on Computer Vision*, pages 14154–14163, 2021.

D. Barnes, W. Maddern, and I. Posner. Find your own way: Weakly-supervised segmentation of path proposals for urban autonomy. In *IEEE International Conference on Robotics and Automation*, pages 203–210. IEEE, 2017.

L. Beyer, X. Zhai, A. Royer, L. Markeeva, R. Anil, and A. Kolesnikov. Knowledge distillation: A good teacher is patient and consistent. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10925–10934, 2022.

M. Biasetton, U. Michieli, G. Agresti, and P. Zanuttigh. Unsupervised domain adaptation for semantic segmentation of urban scenes. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019.

G. Blanchard, G. Lee, and C. Scott. Generalizing from several related classification tasks to a new unlabeled sample. *Advances in neural information processing systems*, 24, 2011.

D. Bogdoll, M. Nitsche, and J. M. Zöllner. Anomaly detection in autonomous driving: A survey. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4488–4499, 2022.

K. M. Borgwardt, A. Gretton, M. J. Rasch, H.-P. Kriegel, B. Schölkopf, and A. J. Smola. Integrating structured biological data by kernel maximum mean discrepancy. *Bioinformatics*, 22(14):e49–e57, 2006.

H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom. nuscenes: A multimodal dataset for autonomous driving. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11621–11631, 2020.

Y. Cao, C. Xiao, B. Cyr, Y. Zhou, W. Park, S. Rampazzi, Q. A. Chen, K. Fu, and Z. M. Mao. Adversarial sensor attack on lidar-based perception in autonomous driving. In

ACM SIGSAC Conference on Computer and Communications Security, pages 2267–2281, 2019.

M. Carranza-García, J. Torres-Mateo, P. Lara-Benítez, and J. García-Gutiérrez. On the performance of one-stage and two-stage object detectors in autonomous vehicles using camera data. *Remote Sensing*, 13(1):89, 2020.

A. Carvalho, S. Lefévre, G. Schildbach, J. Kong, and F. Borrelli. Automated driving: The role of forecasts and uncertainty—a control perspective. *European Journal of Control*, 24:14–32, 2015.

A. Chakeri, X. Wang, Q. Goss, M. I. Akbas, and L. G. Jaimes. A platform-based incentive mechanism for autonomous vehicle crowdsensing. *IEEE Open Journal of Intelligent Transportation Systems*, 2:13–23, 2021.

L. Chen, Y. Li, C. Huang, B. Li, Y. Xing, D. Tian, L. Li, Z. Hu, X. Na, Z. Li, et al. Milestones in autonomous driving and intelligent vehicles: Survey of surveys. *IEEE Transactions on Intelligent Vehicles*, 8(2):1046–1056, 2022.

P. Chen, S. Liu, H. Zhao, and J. Jia. Distilling knowledge via knowledge review. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5008–5017, 2021a.

S. Chen, X. Jia, J. He, Y. Shi, and J. Liu. Semi-supervised domain adaptation based on dual-level domain mixing for semantic segmentation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11018–11027, 2021b.

H. Cho, J. Choi, G. Baek, and W. Hwang. itkd: Interchange transfer-based knowledge distillation for 3d object detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13540–13549, 2023.

S. Choi, S. Jung, H. Yun, J. T. Kim, S. Kim, and J. Choo. Robustnet: Improving domain generalization in urban-scene segmentation via instance selective whitening. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11580–11590, 2021.

W. Chuah, R. Tennakoon, R. Hoseinnezhad, A. Bab-Hadiashar, and D. Suter. Itsa: An information-theoretic approach to automatic shortcut avoidance and domain generalization in stereo matching networks. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13022–13032, 2022.

M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele. The cityscapes dataset for semantic urban scene understanding. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3213–3223, 2016.

A.-D. Doan, Y. Latif, T.-J. Chin, Y. Liu, T.-T. Do, and I. Reid. Scalable place recognition under appearance change for autonomous driving. In *IEEE/CVF International Conference on Computer Vision*, pages 9319–9328, 2019.

P. Drews, G. Williams, B. Goldfain, E. A. Theodorou, and J. M. Rehg. Aggressive deep driving: Combining convolutional neural networks and model predictive control. In *Conference on Robot Learning*, pages 133–142. PMLR, 2017.

S. Fadadu, S. Pandey, D. Hegde, Y. Shi, F.-C. Chou, N. Djuric, and C. Vallespi-Gonzalez. Multi-view fusion of sensor data for improved perception and prediction in autonomous driving. In *IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 2349–2357, 2022.

D. Feng, C. Haase-Schütz, L. Rosenbaum, H. Hertlein, C. Glaeser, F. Timm, W. Wiesbeck, and K. Dietmayer. Deep multi-modal object detection and semantic segmentation for autonomous driving: Datasets, methods, and challenges. *IEEE Transactions on Intelligent Transportation Systems*, 22(3):1341–1360, 2020.

D. Feng, A. Harakeh, S. L. Waslander, and K. Dietmayer. A review and comparative study on probabilistic object detection in autonomous driving. *IEEE Transactions on Intelligent Transportation Systems*, 23(8):9961–9980, 2021.

Y. Ganin and V. Lempitsky. Unsupervised domain adaptation by backpropagation. In *International Conference on Machine Learning*, pages 1180–1189. PMLR, 2015.

H. Gao, J. Guo, G. Wang, and Q. Zhang. Cross-domain correlation distillation for unsupervised domain adaptation in nighttime semantic segmentation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9913–9923, 2022.

Y. Gao, W. Tong, E. Q. Wu, W. Chen, G. Zhu, and F.-Y. Wang. Chat with chatgpt on interactive engines for intelligent driving. *IEEE Transactions on Intelligent Vehicles*, 2023.

A. Geiger, P. Lenz, C. Stiller, and R. Urtasun. Vision meets robotics: The kitti dataset. *The International Journal of Robotics Research*, 32(11):1231–1237, 2013.

J. Geyer, Y. Kassahun, M. Mahmudi, X. Ricou, R. Durgesh, A. S. Chung, L. Hauswald, V. H. Pham, M. Mühlegg, S. Dorn, et al. A2d2: Audi autonomous driving dataset. *arXiv preprint arXiv:2004.06320*, 2020.

A. Gholamhosseinian and J. Seitz. Vehicle classification in intelligent transport systems: An overview, methods and software perspective. *IEEE Open Journal of Intelligent Transportation Systems*, 2:173–194, 2021.

Z. Gojcic, O. Litany, A. Wieser, L. J. Guibas, and T. Birdal. Weakly supervised learning of rigid 3d scene flow. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5692–5703, 2021.

I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020.

J. Gou, B. Yu, S. J. Maybank, and D. Tao. Knowledge distillation: A survey. *International Journal of Computer Vision*, 129:1789–1819, 2021.

A. Gretton, D. Sejdinovic, H. Strathmann, S. Balakrishnan, M. Pontil, K. Fukumizu, and B. K. Sriperumbudur. Optimal kernel choice for large-scale two-sample tests. *Advances in neural information processing systems*, 25, 2012.

S. Grigorescu, B. Trasnea, T. Cocias, and G. Macesanu. A survey of deep learning techniques for autonomous driving. *Journal of Field Robotics*, 37(3):362–386, 2020.

D. Guo, L. Zhu, Y. Lu, H. Yu, and S. Wang. Small object sensitive segmentation of urban street scene with spatial adjacency between object classes. *IEEE Transactions on Image Processing*, 28(6):2643–2653, 2018.

D. Guo, Y. Pei, K. Zheng, H. Yu, Y. Lu, and S. Wang. Degraded image semantic segmentation with dense-gram networks. *IEEE Transactions on Image Processing*, 29:782–795, 2019a.

Y. Guo, H. Shi, A. Kumar, K. Grauman, T. Rosing, and R. Feris. Spottune: transfer learning through adaptive fine-tuning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4805–4814, 2019b.

Z. Hao, S. You, Y. Li, K. Li, and F. Lu. Learning from synthetic photorealistic raindrop for single image raindrop removal. In *IEEE/CVF International Conference on Computer Vision Workshops*, pages 0–0, 2019.

Z. He and L. Zhang. Multi-adversarial faster-rcnn for unrestricted object detection. In *IEEE/CVF International Conference on Computer Vision*, pages 6668–6677, 2019.

G. Hinton, O. Vinyals, and J. Dean. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015.

M. Hnewa and H. Radha. Object detection under rainy conditions for autonomous vehicles: A review of state-of-the-art and emerging techniques. *IEEE Signal Processing Magazine*, 38(1):53–67, 2020.

J. Hoffman, E. Tzeng, T. Park, J.-Y. Zhu, P. Isola, K. Saenko, A. Efros, and T. Darrell. Cycada: Cycle-consistent adversarial domain adaptation. In *International conference on machine learning*, pages 1989–1998. Pmlr, 2018.

Y. Hou, Z. Ma, C. Liu, and C. C. Loy. Learning lightweight lane detection cnns by self attention distillation. In *IEEE/CVF International Conference on Computer Vision*, pages 1013–1021, 2019.

Y. Hou, X. Zhu, Y. Ma, C. C. Loy, and Y. Li. Point-to-voxel knowledge distillation for lidar semantic segmentation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8479–8488, 2022.

J. Houston, G. Zuidhof, L. Bergamini, Y. Ye, L. Chen, A. Jain, S. Omari, V. Iglovikov, and P. Ondruska. One thousand and one hours: Self-driving motion prediction dataset. In *Conference on Robot Learning*, pages 409–418. PMLR, 2021.

H. Hu, Z. Liu, S. Chitlangia, A. Agnihotri, and D. Zhao. Investigating the impact of multi-lidar placement on object detection for autonomous driving. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2550–2559, 2022a.

Y. Hu, X. Jia, M. Tomizuka, and W. Zhan. Causal-based time series domain generalization for vehicle intention prediction. In *International Conference on Robotics and Automation*, pages 7806–7813. IEEE, 2022b.

X. Huang, X. Cheng, Q. Geng, B. Cao, D. Zhou, P. Wang, Y. Lin, and R. Yang. The apolloscape dataset for autonomous driving. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 954–960, 2018.

Y. Huang and Y. Chen. Autonomous driving with deep learning: A survey of state-of-art technologies. *arXiv preprint arXiv:2006.06091*, 2020.

N. Inoue, R. Furuta, T. Yamasaki, and K. Aizawa. Cross-domain weakly-supervised object detection through progressive domain adaptation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 5001–5009, 2018.

P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1125–1134, 2017.

M. Johnson-Roberson, C. Barto, R. Mehta, S. N. Sridhar, K. Rosaen, and R. Vasudevan. Driving in the matrix: Can virtual worlds replace human-generated annotations for real world tasks? *arXiv preprint arXiv:1610.01983*, 2016.

Y. H. Khalil and H. T. Mouftah. Licanet: Further enhancement of joint perception and motion prediction based on multi-modal fusion. *IEEE Open Journal of Intelligent Transportation Systems*, 3:222–235, 2022.

A. Khosravian, A. Amirkhani, H. Kashiani, and M. Masih-Tehrani. Generalizing state-of-the-art object detectors for autonomous vehicles in unseen environments. *Expert Systems with Applications*, 183:115417, 2021.

Y. Ko, Y. Lee, S. Azam, F. Munir, M. Jeon, and W. Pedrycz. Key points estimation and point instance segmentation approach for lane detection. *IEEE Transactions on Intelligent Transportation Systems*, 23(7):8949–8958, 2021.

D. Kothandaraman, A. Nambiar, and A. Mittal. Domain adaptive knowledge distillation for driving scene semantic segmentation. In *IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 134–143, 2021.

V. R. Kumar, M. Klingner, S. Yogamani, S. Milz, T. Fingscheidt, and P. Mader. Syndistnet: Self-supervised monocular fisheye camera distance estimation synergized with semantic segmentation for autonomous driving. In *IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 61–71, 2021.

Q. Lan and Q. Tian. Instance, scale, and teacher adaptive knowledge distillation for visual detection in autonomous driving. *IEEE Transactions on Intelligent Vehicles*, 2022.

A. Lehner, S. Gasperini, A. Marcos-Ramiro, M. Schmidt, M.-A. N. Mahani, N. Navab, B. Busam, and F. Tombari. 3d-vfield: Adversarial augmentation of point clouds for domain generalization in 3d object detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17295–17304, 2022.

D. Li and H. Zhang. Improved regularization and robustness for fine-tuning in neural networks. *Advances in Neural Information Processing Systems*, 34:27249–27262, 2021.

D. Li, L. Deng, and Z. Cai. Intelligent vehicle network system and smart city management based on genetic algorithms and image perception. *Mechanical Systems and Signal Processing*, 141:106623, 2020a.

G. Li, Z. Ji, and X. Qu. Stepwise domain adaptation (sda) for object detection in autonomous vehicles using an adaptive centernet. *IEEE Transactions on Intelligent Transportation Systems*, 23(10):17729–17743, 2022a.

G. Li, Z. Ji, X. Qu, R. Zhou, and D. Cao. Cross-domain object detection for autonomous driving: A stepwise domain adaptative yolo approach. *IEEE Transactions on Intelligent Vehicles*, 7(3):603–615, 2022b.

J. Li, Z. Xu, L. Fu, X. Zhou, and H. Yu. Domain adaptation from daytime to nighttime: A situation-sensitive vehicle detection and traffic flow parameter estimation framework. *Transportation Research Part C: Emerging Technologies*, 124:102946, 2021.

J. Li, R. Xu, X. Liu, J. Ma, Z. Chi, J. Ma, and H. Yu. Learning for vehicle-to-vehicle cooperative perception under lossy communication. *IEEE Transactions on Intelligent Vehicles*, 2023a.

J. Li, R. Xu, J. Ma, Q. Zou, J. Ma, and H. Yu. Domain adaptive object detection for autonomous driving under foggy weather. In *IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 612–622, 2023b.

Y. Li and J. Ibanez-Guzman. Lidar for autonomous driving: The principles, challenges, and trends for automotive lidar and perception systems. *IEEE Signal Processing Magazine*, 37(4):50–61, 2020.

Y. Li, L. Ma, Z. Zhong, F. Liu, M. A. Chapman, D. Cao, and J. Li. Deep learning for lidar point clouds in autonomous driving: A review. *IEEE Transactions on Neural Networks and Learning Systems*, 32(8):3412–3432, 2020b.

Y. Li, Z. Li, S. Teng, Y. Zhang, Y. Zhou, Y. Zhu, D. Cao, B. Tian, Y. Ai, Z. Xuanyuan, et al. Automine: An unmanned mine dataset. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21308–21317, 2022c.

Y. Li, D. Zhang, M. Keuper, and A. Khoreva. Intra-source style augmentation for improved domain generalization. In *IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 509–519, 2023c.

X. Liang, Y. Liu, T. Chen, M. Liu, and Q. Yang. Federated transfer reinforcement learning for autonomous driving. In *Federated and Transfer Learning*, pages 357–371. Springer, 2022.

Y. Liao, J. Xie, and A. Geiger. Kitti-360: A novel dataset and benchmarks for urban scene understanding in 2d and 3d. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

C. Lin, Z. Yuan, S. Zhao, P. Sun, C. Wang, and J. Cai. Domain-invariant disentangled network for generalizable object detection. In *IEEE/CVF International Conference on Computer Vision*, pages 8771–8780, 2021.

L. Liu, X. Chen, S. Zhu, and P. Tan. Condlanenet: a top-to-down lane detection framework based on conditional convolution. In *IEEE/CVF International Conference on Computer Vision*, pages 3773–3782, 2021a.

M.-Y. Liu, T. Breuel, and J. Kautz. Unsupervised image-to-image translation networks. *Advances in neural information processing systems*, 30, 2017.

P. Liu, C. Zhang, H. Qi, G. Wang, and H. Zheng. Multi-attention densenet: A scattering medium imaging optimization framework for visual data pre-processing of autonomous driving systems. *IEEE Transactions on Intelligent Transportation Systems*, 23(12):25396–25407, 2022a.

W. Liu, X. Xia, L. Xiong, Y. Lu, L. Gao, and Z. Yu. Automated vehicle sideslip angle estimation considering signal measurement characteristic. *IEEE Sensors Journal*, 21 (19):21675–21687, 2021b.

W. Liu, K. Quijano, and M. M. Crawford. Yolov5-tassel: detecting tassels in rgb uav imagery with improved yolov5 based on transfer learning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15:8085–8094, 2022b.

Y. Liu, W. Zhang, and J. Wang. Source-free domain adaptation for semantic segmentation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1215–1224, 2021c.

M. Long, Y. Cao, J. Wang, and M. Jordan. Learning transferable features with deep adaptation networks. In *International conference on machine learning*, pages 97–105. PMLR, 2015.

M. Long, H. Zhu, J. Wang, and M. I. Jordan. Deep transfer learning with joint adaptation networks. In *International conference on machine learning*, pages 2208–2217. PMLR, 2017.

C. Luo, X. Yang, and A. Yuille. Self-supervised pillar motion learning for autonomous driving. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3183–3192, 2021.

X. Luo, J. Zhang, K. Yang, A. Roitberg, K. Peng, and R. Stiefelhagen. Towards robust semantic segmentation of accident scenes via multi-source mixed sampling and meta-learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4429–4439, 2022.

J. Mao, M. Niu, C. Jiang, H. Liang, J. Chen, X. Liang, Y. Li, C. Ye, W. Zhang, Z. Li, et al. One million scenes for autonomous driving: Once dataset. *arXiv preprint arXiv:2106.11037*, 2021.

M. Masmoudi, H. Friji, H. Ghazzai, and Y. Massoud. A reinforcement learning framework for video frame-based autonomous car-following. *IEEE Open Journal of Intelligent Transportation Systems*, 2:111–127, 2021.

A. Miglani and N. Kumar. Deep learning models for traffic flow prediction in autonomous vehicles: A review, solutions, and challenges. *Vehicular Communications*, 20:100184, 2019.

M. J. Mirza, M. Masana, H. Possegger, and H. Bischof. An efficient domain-incremental learning approach to drive in all weather conditions. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3001–3011, 2022.

Y. Mo, Y. Wu, X. Yang, F. Liu, and Y. Liao. Review the state-of-the-art technologies of semantic segmentation based on deep learning. *Neurocomputing*, 493:626–646, 2022.

A. S. Mohammed, A. Amamou, F. K. Ayevide, S. Kelouwani, K. Agbossou, and N. Zioui. The perception system of intelligent ground vehicles in all weather conditions: A systematic literature review. *Sensors*, 20(22):6532, 2020.

Z. Murez, S. Kolouri, D. Kriegman, R. Ramamoorthi, and K. Kim. Image to image translation for domain adaptation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 4500–4509, 2018.

V. Muşat, I. Fursa, P. Newman, F. Cuzzolin, and A. Bradley. Multi-weather city: Adverse weather stacking for autonomous driving. In *IEEE/CVF International Conference on Computer Vision*, pages 2906–2915, 2021.

S. Niu, Y. Liu, J. Wang, and H. Song. A decade survey of transfer learning (2010–2020). *IEEE Transactions on Artificial Intelligence*, 1(2):151–166, 2020.

F. Pan, I. Shin, F. Rameau, S. Lee, and I. S. Kweon. Unsupervised intra-domain adaptation for semantic segmentation through self-supervision. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3764–3773, 2020.

S. J. Pan and Q. Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2010.

T. Park, A. A. Efros, R. Zhang, and J.-Y. Zhu. Contrastive learning for unpaired image-to-image translation. In *Computer Vision–16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IX 16*, pages 319–345. Springer, 2020.

A. Patil, S. Malla, H. Gang, and Y.-T. Chen. The h3d dataset for full-surround 3d multi-object detection and tracking in crowded urban scenes. In *International Conference on Robotics and Automation*, pages 9552–9557. IEEE, 2019.

X. Peng, B. Usman, K. Saito, N. Kaushik, J. Hoffman, and K. Saenko. Syn2real: A new benchmark forsynthetic-to-real visual domain adaptation. *arXiv preprint arXiv:1806.09755*, 2018.

Q.-H. Pham, P. Sevestre, R. S. Pahwa, H. Zhan, C. H. Pang, Y. Chen, A. Mustafa, V. Chandrasekhar, and J. Lin. A 3d dataset: Towards autonomous driving in chal-

lenging environments. In *IEEE International Conference on Robotics and Automation*, pages 2267–2273. IEEE, 2020.

F. Pizzati, R. d. Charette, M. Zaccaria, and P. Cerri. Domain bridge for unpaired image-to-image translation and unsupervised domain adaptation. In *IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 2990–2998, 2020.

F. Qiao, L. Zhao, and X. Peng. Learning to learn single domain generalization. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12556–12565, 2020.

H. Rashed, E. Mohamed, G. Sistu, V. R. Kumar, C. Eising, A. El-Sallab, and S. Yogamani. Generalized object detection on fisheye cameras for autonomous driving: Dataset, representations and baseline. In *IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 2272–2280, 2021.

S. R. Richter, V. Vineet, S. Roth, and V. Koltun. Playing for data: Ground truth from computer games. In *Computer Vision–14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14*, pages 102–118. Springer, 2016.

C. B. Rist, M. Enzweiler, and D. M. Gavrila. Cross-sensor deep domain adaptation for lidar detection and segmentation. In *IEEE Intelligent Vehicles Symposium*, pages 1535–1542. IEEE, 2019.

G. Ros, L. Sellart, J. Materzynska, D. Vazquez, and A. M. Lopez. The synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3234–3243, 2016.

C. Sakaridis, D. Dai, and L. Van Gool. Semantic foggy scene understanding with synthetic data. *International Journal of Computer Vision*, 126:973–992, 2018.

J. Sanchez, J.-E. Deschaud, and F. Goulette. Domain generalization of 3d semantic segmentation in autonomous driving. *arXiv preprint arXiv:2212.04245*, 2022.

C. Sautier, G. Puy, S. Gidaris, A. Boulch, A. Bursuc, and R. Marlet. Image-to-lidar self-supervised distillation for autonomous driving data. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9891–9901, 2022.

B. Schlager, T. Goelles, M. Behmer, S. Muckenhuber, J. Payer, and D. Watzenig. Automotive lidar and vibration: Resonance, inertial measurement unit, and effects on the point cloud. *IEEE Open Journal of Intelligent Transportation Systems*, 3:426–434, 2022a.

B. Schlager, T. Goelles, S. Muckenhuber, and D. Watzenig. Contaminations on lidar sensor covers: Performance degradation including fault detection and modeling as potential applications. *IEEE Open Journal of Intelligent Transportation Systems*, 3: 738–747, 2022b.

M. Schutera, M. Hussein, J. Abhau, R. Mikut, and M. Reischl. Night-to-day: Online image-to-image translation for object detection within autonomous driving by night. *IEEE Transactions on Intelligent Vehicles*, 6(3):480–489, 2020.

Y. Shan, W. F. Lu, and C. M. Chew. Pixel and feature level based domain adaptation for object detection in autonomous driving. *Neurocomputing*, 367:31–38, 2019.

Z. Shen, J. Liu, Y. He, X. Zhang, R. Xu, H. Yu, and P. Cui. Towards out-of-distribution generalization: A survey. *arXiv preprint arXiv:2108.13624*, 2021.

D. Shenaj, E. Fanì, M. Toldo, D. Caldarola, A. Tavera, U. Michieli, M. Ciccone, P. Zanuttigh, and B. Caputo. Learning across domains and devices: Style-driven source-free domain adaptation in clustered federated learning. In *IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 444–454, 2023.

I. Shin, Y.-H. Tsai, B. Zhuang, S. Schulter, B. Liu, S. Garg, I. S. Kweon, and K.-J. Yoon. Mm-tta: multi-modal test-time adaptation for 3d semantic segmentation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16928–16937, 2022.

S. Song, H. Yu, Z. Miao, Q. Zhang, Y. Lin, and S. Wang. Domain adaptation for convolutional neural networks-based remote sensing scene classification. *IEEE Geoscience and Remote Sensing Letters*, 16(8):1324–1328, 2019a.

S. Song, H. Yu, Z. Miao, J. Fang, K. Zheng, C. Ma, and S. Wang. Multi-spectral salient object detection by adversarial domain adaptation. In *AAAI Conference on Artificial Intelligence*, volume 34, pages 12023–12030, 2020.

X. Song, P. Wang, D. Zhou, R. Zhu, C. Guan, Y. Dai, H. Su, H. Li, and R. Yang. Apollocar3d: A large 3d car instance understanding benchmark for autonomous driving. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5452–5462, 2019b.

Z. Song, Z. He, X. Li, Q. Ma, R. Ming, Z. Mao, H. Pei, L. Peng, J. Hu, D. Yao, et al. Synthetic datasets for autonomous driving: A survey. *arXiv preprint arXiv:2304.12205*, 2023.

P. Sun, H. Kretzschmar, X. Dotiwalla, A. Chouard, V. Patnaik, P. Tsui, J. Guo, Y. Zhou, Y. Chai, B. Caine, et al. Scalability in perception for autonomous driving: Waymo open dataset. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2446–2454, 2020.

T. Sun, C. Lu, T. Zhang, and H. Ling. Safe self-refinement for transformer-based domain adaptation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7191–7200, 2022.

M. N. Tahir, K. Mäenpää, T. Sukuvaara, and P. Leviäkangas. Deployment and analysis of cooperative intelligent transport system pilot service alerts in real environment. *IEEE Open Journal of Intelligent Transportation Systems*, 2:140–148, 2021.

L. T. Triess, M. Dreissig, C. B. Rist, and J. M. Zöllner. A survey on deep domain adaptation for lidar perception. In *IEEE Intelligent Vehicles Symposium Workshops*, pages 350–357. IEEE, 2021.

M. Uricar, G. Sistu, H. Rashed, A. Vobecky, V. R. Kumar, P. Krizek, F. Burger, and S. Yogamani. Let's get dirty: Gan based data augmentation for camera lens soiling

detection in autonomous driving. In *IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 766–775, 2021.

A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. Attention is all you need. *Advances In Neural Information Processing Systems*, 30, 2017.

H. Wang, Y. Chen, Y. Cai, L. Chen, Y. Li, M. A. Sotelo, and Z. Li. Sfnet-n: An improved sfnet algorithm for semantic segmentation of low-light autonomous driving road scenes. *IEEE Transactions on Intelligent Transportation Systems*, 23(11): 21405–21417, 2022a.

J. Wang, C. Lan, C. Liu, Y. Ouyang, T. Qin, W. Lu, Y. Chen, W. Zeng, and P. Yu. Generalizing to unseen domains: A survey on domain generalization. *IEEE Transactions on Knowledge and Data Engineering*, 2022b.

L. Wang and K.-J. Yoon. Knowledge distillation and student-teacher learning for visual intelligence: A review and new outlooks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

Q. Wang, J. Gao, and X. Li. Weakly supervised adversarial domain adaptation for semantic segmentation in urban scenes. *IEEE Transactions on Image Processing*, 28(9):4376–4386, 2019a.

Q. Wang, D. Dai, L. Hoyer, L. Van Gool, and O. Fink. Domain adaptive semantic segmentation with self-supervised depth estimation. In *IEEE/CVF International Conference on Computer Vision*, pages 8515–8525, 2021a.

T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 8798–8807, 2018.

Y. Wang, W.-L. Chao, D. Garg, B. Hariharan, M. Campbell, and K. Q. Weinberger. Pseudo-lidar from visual depth estimation: Bridging the gap in 3d object detection for autonomous driving. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8445–8453, 2019b.

Y. Wang, J. Yin, W. Li, P. Frossard, R. Yang, and J. Shen. Ssda3d: Semi-supervised domain adaptation for 3d object detection from point cloud. *arXiv preprint arXiv:2212.02845*, 2022c.

Y. Wang, Q. Mao, H. Zhu, J. Deng, Y. Zhang, J. Ji, H. Li, and Y. Zhang. Multi-modal 3d object detection in autonomous driving: a survey. *International Journal of Computer Vision*, pages 1–31, 2023a.

Z. Wang, Y. Wei, R. Feris, J. Xiong, W.-M. Hwu, T. S. Huang, and H. Shi. Alleviating semantic-level shift: A semi-supervised domain adaptation method for semantic segmentation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 936–937, 2020.

Z. Wang, Y. Luo, R. Qiu, Z. Huang, and M. Baktashmotlagh. Learning to diversify for single domain generalization. In *IEEE/CVF International Conference on Computer Vision*, pages 834–843, 2021b.

Z. Wang, C. Lv, and F.-Y. Wang. A new era of intelligent vehicles and intelligent transportation systems: Digital twins and parallel intelligence. *IEEE Transactions on Intelligent Vehicles*, 2023b.

L.-H. Wen and K.-H. Jo. Deep learning-based perception systems for autonomous driving: A comprehensive survey. *Neurocomputing*, 2022.

G. Wilson and D. J. Cook. A survey of unsupervised deep domain adaptation. *ACM Transactions on Intelligent Systems and Technology*, 11(5):1–46, 2020.

A. Wu and C. Deng. Single-domain generalized object detection in urban scene via cyclic-disentangled self-distillation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 847–856, 2022.

B. Wu, X. Zhou, S. Zhao, X. Yue, and K. Keutzer. Squeezesegv2: Improved model structure and unsupervised domain adaptation for road-object segmentation from a lidar point cloud. In *International Conference on Robotics and Automation*, pages 4376–4382. IEEE, 2019.

X. Wu, Z. Wu, H. Guo, L. Ju, and S. Wang. Dannet: A one-stage domain adaptation network for unsupervised nighttime semantic segmentation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15769–15778, 2021.

P. Xie and X. Du. Performance-aware mutual knowledge distillation for improving neural architecture search. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11922–11932, 2022.

J. Xu, Y. Nie, P. Wang, and A. M. López. Training a binary weight object detector by knowledge transfer for autonomous driving. In *International Conference on Robotics and Automation*, pages 2379–2384. IEEE, 2019.

Q. Xu, Y. Zhou, W. Wang, C. R. Qi, and D. Anguelov. Spg: Unsupervised domain adaptation for 3d object detection via semantic point generation. In *IEEE/CVF International Conference on Computer Vision*, pages 15446–15456, 2021.

Q. Xu, L. Yao, Z. Jiang, G. Jiang, W. Chu, W. Han, W. Zhang, C. Wang, and Y. Tai. Dirl: Domain-invariant representation learning for generalizable semantic segmentation. In *AAAI Conference on Artificial Intelligence*, volume 36, pages 2884–2892, 2022a.

R. Xu, H. Xiang, X. Xia, X. Han, J. Li, and J. Ma. Opv2v: An open benchmark dataset and fusion pipeline for perception with vehicle-to-vehicle communication. In *International Conference on Robotics and Automation*, pages 2583–2589. IEEE, 2022b.

R. Xu, J. Li, X. Dong, H. Yu, and J. Ma. Bridging the domain gap for multi-agent perception. *IEEE International Conference on Robotics and Automation*, 2023a.

R. Xu, X. Xia, J. Li, H. Li, S. Zhang, Z. Tu, Z. Meng, H. Xiang, X. Dong, R. Song, et al. V2v4real: A real-world large-scale dataset for vehicle-to-vehicle cooperative perception. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13712–13722, 2023b.

C. Yang, H. Zhou, Z. An, X. Jiang, Y. Xu, and Q. Zhang. Cross-image relational knowledge distillation for semantic segmentation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12319–12328, 2022.

Z. Yang, Y. Chai, D. Anguelov, Y. Zhou, P. Sun, D. Erhan, S. Rafferty, and H. Kretzschmar. Surfelgan: Synthesizing realistic sensor data for autonomous driving. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11118–11127, 2020.

L. Ye, Z. Wang, X. Chen, J. Wang, K. Wu, and K. Lu. Gsan: Graph self-attention network for learning spatial–temporal interaction representation in autonomous driving. *IEEE Internet of Things Journal*, 9(12):9190–9204, 2021.

D. J. Yeong, G. Velasco-Hernandez, J. Barry, and J. Walsh. Sensor and sensor fusion technology in autonomous vehicles: A review. *Sensors*, 21(6):2140, 2021.

L. Yi, B. Gong, and T. Funkhouser. Complete & label: A domain adaptation approach to semantic segmentation of lidar point clouds. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15363–15373, 2021.

Y. You, C. P. Phoo, K. Luo, T. Zhang, W.-L. Chao, B. Hariharan, M. Campbell, and K. Q. Weinberger. Unsupervised adaptation from repeated traversals for autonomous driving. *Advances in Neural Information Processing Systems*, 35:27716–27729, 2022.

F. Yu, W. Xian, Y. Chen, F. Liu, M. Liao, V. Madhavan, and T. Darrell. Bdd100k: A diverse driving video database with scalable annotation tooling. *arXiv preprint arXiv:1805.04687*, 2(5):6, 2018.

X. Yue, Y. Zhang, S. Zhao, A. Sangiovanni-Vincentelli, K. Keutzer, and B. Gong. Domain randomization and pyramid consistency: Simulation-to-real generalization without accessing target domain data. In *IEEE/CVF International Conference on Computer Vision*, pages 2100–2110, 2019.

E. Yurtsever, J. Lambert, A. Carballo, and K. Takeda. A survey of autonomous driving: Common practices and emerging technologies. *IEEE access*, 8:58443–58469, 2020.

S. Zagoruyko and N. Komodakis. Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer. *arXiv preprint arXiv:1612.03928*, 2016.

J. Zhang, C. Ma, K. Yang, A. Roitberg, K. Peng, and R. Stiefelhagen. Transfer beyond the field of view: Dense panoramic semantic segmentation via unsupervised domain adaptation. *IEEE Transactions on Intelligent Transportation Systems*, 23(7):9478–9491, 2021a.

W. Zhang, Z. Wang, and C. C. Loy. Exploring data augmentation for multi-modality 3d object detection. *arXiv preprint arXiv:2012.12741*, 2020.

X. Zhang, H. Zhang, J. Lu, L. Shao, and J. Yang. Target-targeted domain adaptation for unsupervised semantic segmentation. In *IEEE International Conference on Robotics and Automation*, pages 13560–13566. IEEE, 2021b.

X. Zhang, N. Tseng, A. Syed, R. Bhasin, and N. Jaipuria. Simbar: Single image-based scene relighting for effective data augmentation for automated driving vision tasks. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3718–3728, 2022.

Y. Zhang, T. Xiang, T. M. Hospedales, and H. Lu. Deep mutual learning. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 4320–4328, 2018.

X. Zhao, P. Sun, Z. Xu, H. Min, and H. Yu. Fusion of 3d lidar and camera data for object detection in autonomous vehicle applications. *IEEE Sensors Journal*, 20(9): 4901–4913, 2020.

Y. Zhao, Z. Zhong, N. Zhao, N. Sebe, and G. H. Lee. Style-hallucinated dual consistency learning for domain generalized semantic segmentation. In *Computer Vision–17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXVIII*, pages 535–552. Springer, 2022.

B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba. Learning deep features for discriminative localization. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2921–2929, 2016.

D. Zhou, Z. Ma, and J. Sun. Autonomous vehicles' turning motion planning for conflict areas at mixed-flow intersections. *IEEE Transactions on Intelligent Vehicles*, 5(2): 204–216, 2019.

D. Zhou, J. Fang, X. Song, L. Liu, J. Yin, Y. Dai, H. Li, and R. Yang. Joint 3d instance segmentation and object detection for autonomous driving. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1839–1849, 2020a.

K. Zhou, Z. Liu, Y. Qiao, T. Xiang, and C. C. Loy. Domain generalization: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022a.

Y. Zhou, L. Liu, H. Zhao, M. López-Benítez, L. Yu, and Y. Yue. Towards deep radar perception for autonomous driving: Datasets, methods, and challenges. *Sensors*, 22 (11):4208, 2022b.

Z. Zhou, Z. Wang, H. Lu, S. Wang, and M. Sun. Multi-type self-attention guided degraded saliency detection. In *AAAI Conference on Artificial Intelligence*, volume 34, pages 13082–13089, 2020b.

J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *IEEE International Conference on Computer Vision*, pages 2223–2232, 2017.

F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He. A comprehensive survey on transfer learning. *IEEE*, 109(1):43–76, 2020.

A. Ziegler and Y. M. Asano. Self-supervised learning of object parts for semantic segmentation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14502–14511, 2022.