Towards Weakly-Supervised Domain Adaptation for Lane Detection

Jingxing Zhou[1,4] Chongzhe Zhang[1,2] Jürgen Beyerer[3,4]
[1] Porsche Engineering Group GmbH [2] University of Stuttgart [3] Fraunhofer IOSB
[4] Vision and Fusion Laboratory (IES), Karlsruhe Institute of Technology

Abstract

Lane detection plays an indispensable role in automated driving functions and advanced driver assistance systems by providing fundamental spatial orientation, which is imperative for trajectory planning with traffic regulation compliance. The variability of lane structures across the world poses challenges for data-driven lane detection models. However, acquiring vast amounts of labeled data encompassing a wide variety of real-world scenarios for supervised learning is often cost-prohibitive. In this work, we propose a Weakly Supervised Domain Adaptation framework for Lane Detection (WSDAL), which requires easily-provided labels exclusively for the number of lanes *in the target domain to aid the adaptation process. WSDAL consists of a teacher-student network, an additional segmentation head as an auxiliary task during training, and a novel loss function that incorporates the number of lanes prediction. As a versatile framework, WSDAL can be applied to any anchor-based lane detector. Between three frequently-used lane detection datasets (TuSimple, CULane and CurveLanes) for domain adaptation, WSDAL framework demonstrates its effectiveness and efficiency over common unsupervised domain adaptation methods and fully supervised training. In addition, we discuss the quality requisites from the labels for the weakly-supervised domain adaptation, indicating that label errors at realistic scales still provide satisfactory results on the considered tasks.*

1. Introduction

Lane information is one of the most essential information required for assisted and automated driving functions, which gives vehicles a fundamental understanding of their position and orientation on the road. In addition, lane information is the cornerstone of path planning and control algorithms. For this reason, precise and reliable lane detection is indispensable in safe automated driving systems.

The research leading to these results is supported by the German Federal Ministry for Economic Affairs and Climate Action (BMWK) within the project AVEAS (www.aveas.org).

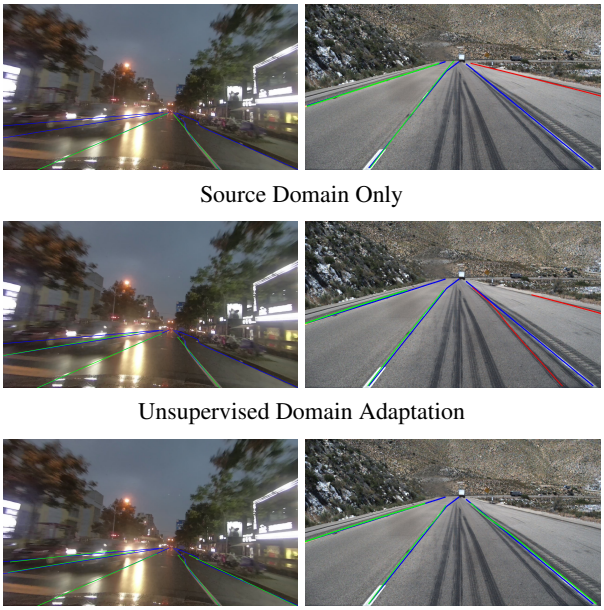

Figure 1. Qualitative results of the proposed WSDAL. On the left side of the figure, we show the advancement made by our proposed method on the task of *CULane* → *CurveLanes*, dealing with complex urban traffic scenes. On the right side, we show an example of the network trained on *CurveLanes* adapts to *TuSimple* that copes with object occlusion. Best viewed in color, true positive predictions shown in green, false positives in red, ground truth in blue.

However, lane detection poses unique challenges compared to other object detection tasks due to the variety and occlusion of the lanes [9, 30, 55, 62]. Therefore, detecting a complete lane requires the detector to bridge the discontinuous parts at the semantic level. In addition, lane markings span a wide area in images but have relatively few pixels, demanding models to possess both local and global perspectives to perceive lane features comprehensively [41, 61]. Furthermore, the task is especially challenging due to the variability in the environment and lane shape. For instance, the definition of lane markings shifts according to the driving context, as sometimes there are no center lane markings on narrow roads, but the vehicle is supposed still to drive on

the correct side of the road. Geographic location and terrain affect the distribution of the lanes as well. The traffic rules around the globe are significantly different: In some countries, people are allowed to drive on the hard shoulder in certain traffic situations, while it is prohibited in most countries. Those changes in the traffic regulations create even more challenges on top of the existing challenges in domain adaptation, as it is not trivial to derive the availability only based on appearance.

In this work, we introduce a novel domain adaptation framework for anchor-based lane detection models. With the help of simple meta labels from the target domain, the Number-of-Lanes (NoL) from the image, the network can be adapted to the target domain more efficiently. Compared to the full-supervised counterparts, such weak labels have cost advantages, as they are available via map services, for instance, OpenStreetMap[29], or from existing vehicle fleet that equips with sensor systems that have been quality assured. We perform our experiments on three datasets that are frequently used in lane detection research, including a wide range of driving situations from highway to urban driving, unveiling the advantages of the models that can be adapted to target domain using simple and low-cost labels from target domain. To the best knowledge of the authors, previous work in the lane detection domain concentrates mainly on the optimization of the network architecture and unsupervised domain adaptation methods [9, 11, 16, 62], which did not integrate publicly available world knowledge from metadata. Our main contributions are as follows:

- We propose a novel framework utilizing cost-effective number-of-lanes (NoL) labels from target domain images as weak supervision signals complementing the common unsupervised domain adaptation methods.
- We provide a thorough analysis of the performance based on three different datasets, including diverse driving scenarios, and discuss insights into the effectiveness of the proposed method compared with existing methods.
- We demonstrate the universal applicability of the proposed framework leveraging different anchor-based lane detection architectures and image encoders. Certainly related to real-world application, we also discuss the impact of potential label error in the target domain on the model's adaptation performance.

2. Related Work

2.1. Weakly-Supervised Learning

Despite the enormous success of data-driven computer vision models [2, 8, 24, 64], most of the ongoing work on the downstream tasks follows the fully supervised training scheme, which is related to high labeling costs of the data. Weakly supervised learning is used in several applications including object localization [6, 35], object detection [48, 58], semantic segmentation [4, 13, 20, 52]. Class Activation Map (CAM) techniques [33, 63] are widely used in weakly supervised learning with image-level labels, where the discriminative area generated by CAM can be further utilized for the supervision of computer vision tasks like semantic segmentation [13, 52]. For those cases where the expected predictions are dense, ground truth with sparse information is also frequently leveraged as weak label. For instance, image-level labels in object detection tasks or bounding boxes for segmentation tasks. Gao *et al*. [6] proposed a method that utilizes number of objects as supervision signal for object localization. Generating segmentation masks in the given bounding boxes [4, 15, 37] is used to generate unsupervised region proposals and iteratively updated to improve the segmentation capability of the network. Joon *et al*. [13] combined external saliency masks with image-level labels to train semantic segmentation models. Scribble [20] is also a low-cost annotation to propagate the category information to the unlabeled pixels for semantic segmentation. Those methods are complementary when various forms of weak supervision are used.

2.2. Data-Driven Lane Detection

Before deep learning methods were applied to lane detection, lane detection was heavily based on handcrafted features [1, 25], which contain pre-processing, feature extraction and curve fitting [42, 59]. In the era of data-driven lane detection, the most commonly used models can be categorized into *segmentation-based* detectors and *anchor-based* detectors. The vanilla approach treats lane detection as a segmentation task. With LaneNet [28], the lane detection task was accomplished by utilizing two lane segmentation branches for foreground segmentation and identifying the lane instances. Then, the segmentation maps were projected into the bird's eye view by transformation matrix. Further improvements like SCNN [30] and RESA [61] utilized spatial CNNs and added multi-stride connections, facilitating the perception of continuous structures. There are also experiments [55] that leveraged AutoML to explore and optimize the model structure for lane detection models. Liu *et al*. [22] utilized a style transfer model to enrich the training set and introduced a second branch that predicts the existence of lane based on pixel-level ground truth.

Line-CNN [17] introduced the line anchor in lane detection. Similar to box anchors in object detection, anchors extract feature representations to constitute an anchor feature, which is used for final prediction. LaneATT [41] extracts local feature maps to form anchor local features. With the help of an attention mechanism, a global feature vector is aggregated to deal with the sparse property of the lane detection task. SGNet [39] utilizes prior information like vanishing point and lane information to improve performance. As a row anchor-based method [31], it predicts the probable

cell for each predefined row on images. CondLaneNet [21] further improved row anchor-based lane detection by learning a heatmap of possible starting points of the lanes as prior information. Similar to many two-stage object detection models, CLRNet [62] adopts learnable anchor parameters for the regression tasks and the detection head gives prediction for each tensor. CLRerNet [9] further improves the performance by optimizing similarity cost that used during optimization to align the confidence score with IoU score.

2.3. Domain Adaptation for Lane Detection

Tackling domain shift for lane detection models is a fundamental yet less popular task. We started the evaluation of domain shift for some models introduced as a preliminary study in Appendix A showing the limited generalization ability of the models. As the lane detection task can be regarded as a *segmentation* task, the common unsupervised domain adaptation methods can be utilized including domain alignment with statistic divergence [14, 23, 40, 45], domain adversarial learning [3, 46, 47], normalization statistics alignment [18, 27], and self-training [10, 19, 34]. There is also work which involves utilizing weak labels for domain adaptation [5, 12, 26, 50, 51].

Ongoing research of lane detection research concentrates on unsupervised domain adaptation, in which the labels in the target domain are not available [38]. Domain adaptation techniques in lane *detection* contain viewpoint transformation alignment from different domains, adversarial training, and self-training methods. Yu *et al*. [56] proposed a method to project images into a uniform viewpoint by estimating the transformation matrix [53]. However, when the accuracy of the projection drops, the performance also degrades significantly. Besides that, the lane geometry needs to be similar as the model struggles when the label distribution shifts. Adversarial learning-based methods are similar to general unsupervised learning methods. Garnett *et al*. [7] used discriminator to distinguish between the model's predictions on the target and source domain. The gradient image of the input helps to limit the deep features extracted by the model. Hu *et al*. [11] further explored adversarial generative and adversarial discriminative methods for sim-to-real domain adaptation. MLDA [16] is a pseudo label-based lane domain adaptation method. It designs multi-level domain adaptation tasks including pixel-wise, instance-wise, and category-wise tasks to generate pseudo labels on target domain and improve model's global understanding by predicting the types of lanes in the images on the target domain. However, unsupervised adaptation methods are **not** able to deal with the regulation shifts and changes with annotation policy. Appearance-similar lane markings, which should be detected in some contexts, yet not in others, can only be addressed using costly, fully supervised learning methods prior to our work.

3. Methodology

3.1. Problem Formulation

The goal of lane detection is to achieve the lane detection task of taking the input image $x \in \mathbb{R}^{H\times W\times 3}$, where H and W are the height and width of the RGB image, and learn parameters θ for a mapping ϕ_θ to the lane prediction $y \in \mathbb{R}^{N\times L}$, where N denotes the number of lane *anchors*, i.e. the maximum number of possible lane predictions, and L the dimension of the lane description vector $\boldsymbol{l}$. Here, $\boldsymbol{l}$ contains $L-2$ elements $\boldsymbol{l}^{(1)} \ \ldots \ \boldsymbol{l}^{(L-2)}$ describing the lane position as coordinates, trained as regression and two elements $\boldsymbol{l}^{(L-1)}$, $\boldsymbol{l}^{(L)}$ for "foreground" and "background" respectively, indicating whether the particular lane likely exists or not, trained as classification. ϕ_θ^S is learned in source domain $\mathcal{D}_\mathcal{S} = \{(x_s, y_s)\}_{s\in\mathcal{S}}$. With unsupervised domain adaptation methods, we utilize only the images from the target dataset $\mathcal{D}_\mathcal{T}^{\text{UDA}} = \{(x_t)\}_{t\in\mathcal{T}}$, while for proposed *Weakly Supervised Domain Adaptation framework for Lane Detection* (WSDAL), the number of lanes $n_t \leq N$ in each target image is known during adaptation with $\mathcal{D}_\mathcal{T}^{\text{WSDAL}} = \{(x_t, n_t)\}_{t\in\mathcal{T}}$.

3.2. WSDAL Framework

Base Architecture: Similar to current work in unsupervised domain adaptation, we employ a self-training strategy based on the teacher–student model. Fig. 2 illustrates the proposed Weakly Supervised Domain Adaptation framework for Lane Detection (WSDAL), which contains a teacher model (red) and a student model (blue), whose architecture and initialization are shared. The teacher model receives original target images, while the student model is trained on data with strong augmentation on the target domain. In this case, the teacher model provides more reliable lane predictions, so it is a reasonable strategy to turn filtered predictions from the teacher model into pseudo labels to supervise the learning of the student model, as transforming low-quality lane predictions into pseudo labels for self-training has a detrimental effect on the adaptation process. In the lane detection task, we select only the lane predictions that have their scores $\boldsymbol{l}_n^{(L-1)} \overset{!}{>} \delta_c$ as the pseudo lane labels. In case the number of pseudo lane labels exceeds the maximum number of lanes in the dataset $N_\mathcal{D}$, we select lane predictions with the highest scores. Under the supervision of pseudo lane labels, the student model not only learns to extract more robust features to exclude the interference of data augmentation but also further adapts itself to the data pattern of the target domain. During the iterative training process, the teacher model is updated by the student model indirectly from the target domain using the Exponential Moving Average (EMA) strategy [43]:

$$\theta'_t = \alpha\theta'_{t-1} + (1-\alpha)\theta_t \tag{1}$$

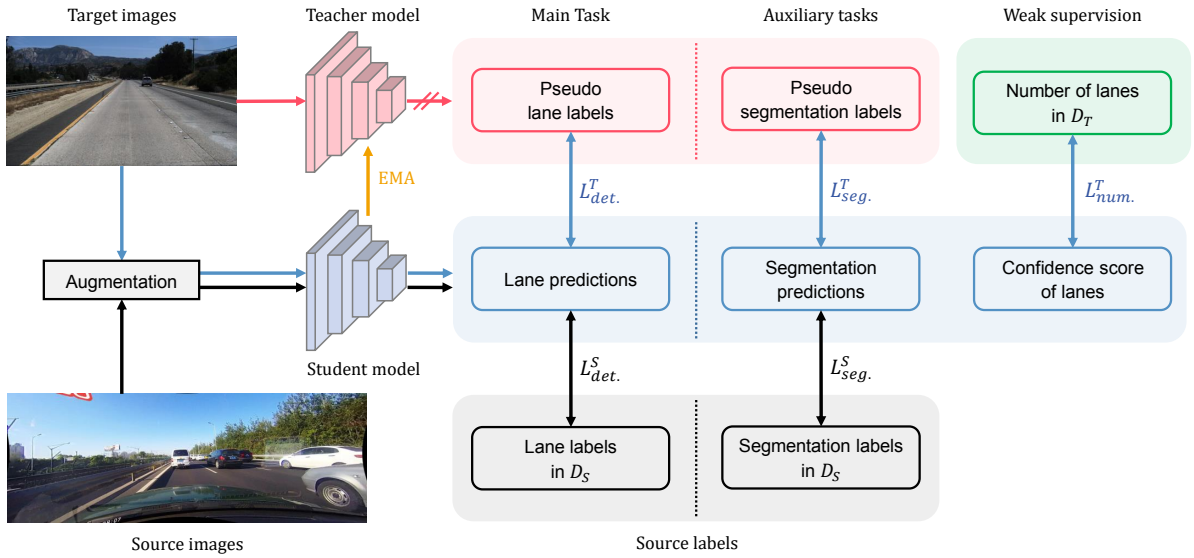

Figure 2. The framework of the proposed Weakly Supervised Domain Adaptation framework for Lane Detection (WSDAL). It consists of a teacher-student network, a segmentation head as an auxiliary task during training and weak supervision utilizing number-of-lanes labels.

where θ'_t represents the parameters of the teacher network at step t, θ_t represents the parameters of the student network at step t, and α is a smoothing coefficient hyperparameter. In this way, the teacher model is also able to acquire knowledge related to the target domain, enabling it to generate more accurate pseudo lane labels from the proposals [9] consistently. Using pseudo lane labels, we can calculate the loss in the lane detection task weighted by $\omega_{\text{det.}}$ with:

$$\mathcal{L}_{\text{det.}} = \omega_{\text{det.}} \mathcal{L}_D \left(\bar{y}_{\text{det.}}, \hat{y}_{\text{det.}} \right) \tag{2}$$

where $\mathcal{L}_D$ denotes the loss function from the anchor-based detector, $\bar{y}_{\text{det}}$ is the pseudo lane label generated by teacher model and $\hat{y}_{\text{det}}$ is the lane prediction by the student model.

Auxiliary Task: WSDAL framework can attach various segmentation tasks, which share the feature maps from the image encoder with the detection branch, as auxiliary tasks. In the case of having multi-scale feature maps, we utilize binary interpolation to derive the feature representation with the smallest downsampling rate. Typically, lane detectors perform instance segmentation tasks as auxiliary tasks where they categorize each lane instance [31, 62]. In our work, we compare it with binary segmentation as an additional auxiliary label strategy. The loss function $\mathcal{L}^S_{\text{seg.}}$ is calculated using cross-entropy to compare between predictions and ground truths on the source domain. The segmentation masks for training are generated by the existing lane marking annotations given the width of the lanes. Similar to the lane detection task, the segmentation pseudo label is also generated by the teacher model. The teacher model generates a probability map $\hat{y}_{\text{seg}}$ of c channels given the image without data augmentation, and filters $\hat{y}_{\text{seg}}$ to foreground points according to the probability threshold α_c:

$$\bar{y}_{\text{seg.}} = \begin{cases} \text{argmax}_c \, \hat{y}_{\text{seg.}}, & \hat{y}_{\text{seg.}} > \alpha_c \\ 0, & \text{else} \end{cases} \tag{3}$$

The lane segmentation predictions obtained from the teacher model are applied to the correct positional transformation operation into pseudo segmentation labels. This way it can be matched with the augmented image fed into the student model. The weighted loss function for the segmentation task is as follows:

$$\mathcal{L}_{\text{seg.}} = \omega_{\text{seg.}} \mathcal{L}_{\text{CE}} \left(\bar{y}_{\text{seg.}}, \hat{y}_{\text{seg.}} \right) \tag{4}$$

Weak Supervision Task: In target domains, unadapted detectors generate lane predictions with low confidence scores from out-of-distribution features, which hampers the process of filtering out correct lane predictions. In our work, we utilize the number of lanes to help the detector adapt to the target domain. Through our number of lane (NoL) loss we improve the confidence score of the lane predictions, which can match with the ground truth. It is necessary to ensure for any predicted lane description $\boldsymbol{l}_n$ in y that the foreground score $\boldsymbol{l}_n^{(L-1)}$ and the background score $\boldsymbol{l}_n^{(L)}$ satisfy $\max(\boldsymbol{l}_n^{(L-1)}, \boldsymbol{l}_n^{(L)}) \approx 1$ by softmax function, such that the confidence between positive and negative predictions remain as large as possible in the domain adaptation phase.

During the calculation of the number of lane loss, we filter the lane predictions with high confidence scores from the student model by $\boldsymbol{l}_n^{(L-1)} \overset{!}{>} \delta_c$ as positive classifications. The sum of their confidence scores is taken as the prediction of the number of lanes currently achieved by the model. In the initial stage, this value is usually smaller than the

number of lanes during domain adaptation. Therefore, the model under the supervision of weak labels increases the confidence score of lane predictions through the propagated gradients, especially the lane predictions whose confidence score is higher than δ_c in the initial stage, which helps the lane predictions with confidence scores above the threshold to diverge from the other lane predictions. After domain adaptation, the model does **not** acquire any meta information from environment during inference.

The weighted number of lanes loss is calculated by:

$$\mathcal{L}_{\text{num.}} = \omega_{\text{num.}} \mathcal{L}_{\text{smooth-l}_1} \left(n_t, \sum_{n \in N'} \boldsymbol{l}_n^{(L-1)} \right) \tag{5}$$

where n_t is the number of lanes and $N' = \{ n \mid \boldsymbol{l}_n^{(L-1)} > \delta_c \}$.

Optimization Process: As shown in Fig. 2, the images in the source and target domains have a total of three paths in the network:

1. In the first path shown in red, the images from the target domain are fed into the teacher model without data augmentation, where the teacher model produces detection and segmentation predictions. Subsequently, the predictions are filtered into pseudo labels according to the threshold δ_c and α_c. We do not calculate any gradient in this step.
2. In the second path shown in blue, the images after data augmentation from the target domain propagate through the student model generating predictions. Firstly, they are compared with the pseudo labels generated by the teacher model where $\mathcal{L}_{\text{det.}}^T$ and $\mathcal{L}_{\text{seg.}}^T$ can be derived. The sum of confidence scores above δ_c is taken as the prediction of the number of lanes to get the $\mathcal{L}_{\text{num.}}^T$.
3. For the third path shown in black, we calculate the losses $\mathcal{L}_{\text{det.}}^S$ and $\mathcal{L}_{\text{seg.}}^S$ on the detection and segmentation predictions of the student model given images with strong data augmentation.

In the process of domain adaptation, we iteratively update both student and teacher models. The student network rapidly transfers the knowledge of the model from the source domain to the target domain, driven by a combination of self-supervised and weakly supervised processes. The parameters of the student model are updated to the teacher model via EMA, which enables the teacher model to generate accurate predictions consistently during training. In the WSDAL framework, the overall loss function is stated as follows:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{det.}}^T + \mathcal{L}_{\text{seg.}}^T + \mathcal{L}_{\text{num.}}^T + \mathcal{L}_{\text{det.}}^S + \mathcal{L}_{\text{seg.}}^S \tag{6}$$

4. Experiments

4.1. Datasets and Evaluation Metrics

TuSimple: TuSimple [44] is a lane dataset focusing on highway scenarios collected in the United States containing around 6000 images. The dataset is relatively homogeneous as the data were collected under good and moderate weather conditions and only covers highways. For the evaluation, lanes are evaluated using sampled points. When the distance between the ground truth point and the lane prediction point is less than 20 pixels, the prediction is considered as correct. For a lane prediction, it is regarded as a true positive if the percentage of correct point predictions is higher than 85%. We report besides the False Positive Rate (FPR), False Negative Rate (FNR) and accuracy, also the F1-score:

$$\text{F1-score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{7}$$

where $\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$, $\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$.

CULane: CULane [30] is a large scale dataset that contains diverse driving situations in Beijing with 133,235 frames. In addition to normal scenario, there are eight challenging subsets including traffic jams, nights, intersections. For evaluation, we strictly follow the evaluation scheme from [30]: A true positive prediction is considered as the prediction with a certain width that has an Intersection-of-Union over a predefined threshold of the ground truth. We report the F1-scores of the subsets and overall dataset. For the subset *cross*, we report the number of false positives.

CurveLanes: CurveLanes [55] is a large scale lane detection dataset with 150,000 images for complex driving scenarios and up to 14 lanes per image. Like the CULane dataset, this dataset was also collected in China with high urban bias. Besides that, this dataset also concentrates on curved lane markings different from the natural distribution of curvature from other datasets. The evaluation scheme is similar to CULane and we report F1-scores, Precision and Recall values based on predefined IoU thresholds.

4.2. Implementation

To simplify the experiment, we select LaneATT [41] and CLRNet [62], two anchor-based detection architectures that are frequently used as baseline model for the performance comparison [9, 49]. When CULane is used as the source domain, following the default settings of LaneATT, all input images are resized to 640 $\times$ 360 and using AdamW optimizer with a cosine annealing learning rate strategy. CLRNet follows its default settings by scaling the input image to 800 $\times$ 320 and using the AdamW optimizer using the cosine annealing learning rate strategy. We trained the models from pre-trained ImageNet-1k weights.

In the case of domain adaptation, we first train the model on the source datasets and the weights are used as starting point for the domain adaptation. We repeat our experiments three times using different seeds. We primarily choose ResNet18 as the encoder. For the data augmentation, we use horizontal flip, channel shuffle, color jitter, motion blur, median blur, random rotation and random scaling for the

student model. We utilize the binary segmentation task as an auxiliary task for the detection network. For TSUDA, we empirically utilize a lower pseudo label threshold $\delta_c = 0.2$ to guarantee that enough labels can be generated, while for WSDAL we can employ a higher threshold $\delta_c = 0.5$. For segmentation tasks, we take the threshold $\alpha_c = 0.8$. Further implementation details are in Appendix B. For ablation, we report the results by deactivating the NoL loss to perform only the unsupervised domain adaptation without using any information from the target domain except the images. We compare this variant, noted as Teacher–Student Unsupervised Domain Adaptation (TSUDA) together with Domain Generalization (DG) without utilizing any adaptation methods with our proposed method.

4.3. Target Dataset TuSimple

We start our investigation with *TuSimple* as the target dataset, which focuses on highway driving scenarios. The comparison of the domain adaptation performance is shown in Tab. 1. As introduced in Sec. 4.2, we evaluate two different base detectors. The proposed weakly-supervised method significantly boosts performance from the *CULane* dataset, as lower FPRs and FNRs can be observed while F1-score and accuracy increase compared with unsupervised method and without adaptation. If the network was previously trained on a more challenging dataset like *CurveLanes*, the model's ability to detect lanes can be observed by the high F1-score even if the model does not have any domain adaptation methods applied. Similar to the *CULane* dataset as source dataset, the proposed WSDAL achieves better detection performance on the target domain with *CurveLanes*. Compared with the upper bound of the model's capacity, namely the model trained on target dataset *TuSimple* only, the model shows an average F1-score difference of only 4% without needing labels on the target domain that are indispensable for fully supervised learning. It shows that the model can still re-use the learned feature representations from the complex source domain since it is trained on data that include many object occlusions. With the help of the generated pseudo labels and the number of lanes constraint for the model's output, the model is able to comprehend the priority of the lane candidates to the explicit lane markings and to suppress the structures that may be considered false positives, for instance, road boundaries and road surface shifts. We provide further in-depth analysis and qualitative results in Appendices E and G.

4.4. Target Dataset CULane

When we consider adapting from the *TuSimple* dataset, we observe as shown in Tab. 2 that the models suffer from the shifted lighting condition from daytime only to also include night driving scenarios. Besides that, the model also needs to cope with typical urban driving situations where the lanes

Table 1. Comparison of the LaneATT and CLRNet models which are trained on *CULane* and *CurveLanes* datasets and adapt to *TuSimple*. We report average score in % according to three runs.

$\mathcal{D}_s$	Model	Backbone	Method	F1	FPR	FNR	Accuracy
CULane	LaneATT	ResNet-18	DG	54.9	49.6	39.4	81.29
			TSUDA	$69.9_{\pm 0.7}$	$32.6_{\pm 0.7}$	$25.4_{\pm 0.6}$	$84.7_{\pm 0.2}$
			WSDAL	$81.7_{\pm 0.3}$	$18.8_{\pm 1.0}$	$17.5_{\pm 0.7}$	$88.8_{\pm 0.4}$
	CLRNet		DG	70.3	32.6	26.35	85.89
			TSUDA	$84.8_{\pm 0.3}$	$12.6_{\pm 0.1}$	$18.9_{\pm 0.9}$	$85.5_{\pm 0.7}$
			WSDAL	$86.9_{\pm 0.3}$	$11.5_{\pm 0.2}$	$15.2_{\pm 0.5}$	$89.0_{\pm 0.4}$
CurveLanes	CLRNet	ResNet-18	DG	80.7	8.1	35.8	72.9
			TSUDA	$84.5_{\pm 0.2}$	$15.7_{\pm 0.5}$	$15.3_{\pm 0.6}$	$86.5_{\pm 0.4}$
			WSDAL	$91.2_{\pm 0.5}$	$8.1_{\pm 0.5}$	$9.7_{\pm 0.5}$	$91.2_{\pm 0.3}$
		ConvNeXt-atto	DG	82.4	10.7	27.1	80.0
			TSUDA	$78.0_{\pm 2.1}$	$23.1_{\pm 1.9}$	$15.5_{\pm 1.9}$	$84.2_{\pm 0.6}$
			WSDAL	$88.9_{\pm 0.7}$	$10.7_{\pm 0.8}$	$11.6_{\pm 0.7}$	$90.4_{\pm 0.4}$
$\mathcal{D}_T$	CLRNet	ResNet-18		95.3	5.5	3.8	95.1
$\mathcal{D}_T$	LaneATT	ResNet-18		95.1	5.9	3.7	94.9

are often occluded by other traffic participants. That is the possible reason that the models trained on *TuSimple*, despite the improvement brought by weakly-supervised domain adaptation, still have high disparity compared with the oracle models. Especially with LaneATT, the use of WSDAL can only bring a slight overall performance improvement in the *TuSimple* to *CULane* task. However, WSDAL can be very helpful to improve CLRNet's performance in the target domain.

From the models that are previously trained on *CurveLanes*, we observe – as in Sec. 4.3 – relatively high scores in the target domain without any adaptation. Besides the normal F1-score indications that are reported in the eight splits, we also observe a noticeable decrease of false positives in the category *Cross Roads*, where there should be no lanes detected according to the labeling policy of *CULane*, compared with the base model without adaptation. Furthermore, it can be noticed that the unsupervised domain adaptation methods do not perform well in this setup. Our explanation is that due to the dilemma in selecting the threshold δ_c for pseudo lane label generation, it is tough to balance the quality and quantity of the lane labels [9]. It is not reliable to differentiate the predictions (often more than the maximum number of lanes in CULane) purely based on the scores, as the selected pseudo labels may not represent the label policy of the CULane. With WSDAL, we incorporate the number of lane loss giving the network hints that encourage the networks to focus on the most apparent visual structures available on the target domain, which can be observed by the improvement in the split *night*.

4.5. Target Dataset CurveLanes

We investigate the shift *CULane* $\rightarrow$ *CurveLanes* based on the CLRNet architecture, as we notice the limitations of the other architecture and training on datasets that do not have high variance. Tab. 3 shows the model's performance based on various IoU thresholds besides the standard 0.5 during evaluation. Due to the significant changes in data

Table 2. Comparison of the LaneATT and CLRNet models which are trained on the *TuSimple* and *CurveLanes* dataset and adapt to *CULane*. We report the average score in % according to three runs.

$\mathcal{D}_s$	Model	Backbone	Method	Normal	Crowd	Dazzle	Shadow	Noline	Arrow	Curve	Cross	Night	**Total**
TuSimple	CLRNet	ResNet-18	DG	42.0	20.4	11.6	5.6	8.2	28.1	22.3	1950	2.9	23.0
			TSUDA	59.9$_{\pm 0.7}$	40.2$_{\pm 0.5}$	32.3$_{\pm 1.4}$	24.9$_{\pm 0.8}$	26.3$_{\pm 0.8}$	54.1$_{\pm 0.5}$	44.6$_{\pm 0.7}$	4510$_{\pm 388}$	30.2$_{\pm 1.5}$	43.0$_{\pm 0.9}$
			WSDAL	65.2$_{\pm 0.4}$	43.4$_{\pm 0.7}$	35.6$_{\pm 1.4}$	28.2$_{\pm 1.5}$	29.3$_{\pm 0.5}$	57.5$_{\pm 0.7}$	48.7$_{\pm 0.2}$	1975$_{\pm 178}$	33.4$_{\pm 0.8}$	47.2$_{\pm 0.6}$
	LaneATT		DG	38.9	18.1	10.9	3.9	6.5	26.1	21.4	569	2.6	21.5
			TSUDA	33.6$_{\pm 0.8}$	19.8$_{\pm 0.4}$	19.9$_{\pm 0.7}$	10.8$_{\pm 0.2}$	13.5$_{\pm 0.3}$	25.6$_{\pm 0.3}$	21.3$_{\pm 0.8}$	7378$_{\pm 250}$	9.0$_{\pm 0.3}$	21.4$_{\pm 0.5}$
			WSDAL	49.0$_{\pm 0.9}$	28.0$_{\pm 0.7}$	25.2$_{\pm 1.4}$	15.6$_{\pm 0.9}$	22.0$_{\pm 0.2}$	37.3$_{\pm 0.8}$	36.6$_{\pm 0.3}$	2560$_{\pm 90}$	21.6$_{\pm 0.7}$	32.8$_{\pm 0.6}$
CurveLanes	CLRNet	ResNet-18	DG	82.2	66.0	60.2	70.1	48.0	76.8	72.0	6730	62.9	66.7
			TSUDA	68.5$_{\pm 0.5}$	50.1$_{\pm 0.6}$	42.4$_{\pm 0.4}$	48.5$_{\pm 0.7}$	30.8$_{\pm 0.4}$	63.6$_{\pm 0.7}$	56.5$_{\pm 0.3}$	889$_{\pm 21}$	51.7$_{\pm 0.5}$	55.0$_{\pm 0.5}$
			WSDAL	86.2$_{\pm 0.0}$	68.5$_{\pm 0.2}$	60.1$_{\pm 0.2}$	72.7$_{\pm 0.1}$	47.3$_{\pm 0.1}$	80.5$_{\pm 0.2}$	73.0$_{\pm 0.4}$	2370$_{\pm 15}$	66.6$_{\pm 0.1}$	70.9$_{\pm 0.1}$
		ConvNext	DG	78.5	63.6	54.6	66.3	42.4	72.7	69.0	3208	59.3	64.7
			TSUDA	49.9$_{\pm 0.4}$	27.5$_{\pm 0.2}$	26.1$_{\pm 0.4}$	40.4$_{\pm 0.8}$	20.8$_{\pm 0.2}$	39.6$_{\pm 0.5}$	47.4$_{\pm 0.2}$	5803$_{\pm 144}$	34.5$_{\pm 0.2}$	35.5$_{\pm 0.2}$
			WSDAL	86.7$_{\pm 0.2}$	68.2$_{\pm 0.2}$	62.8$_{\pm 0.3}$	71.1$_{\pm 0.2}$	46.7$_{\pm 0.2}$	81.5$_{\pm 0.3}$	70.8$_{\pm 0.2}$	2311$_{\pm 12}$	67.2$_{\pm 0.3}$	71.0$_{\pm 0.2}$
$\mathcal{D}_T$	CLRNet	ResNet-18		93.0	77.5	70.1	77.4	52.7	89.3	68.2	1239	77.5	78.5
$\mathcal{D}_T$	LaneATT	ResNet-18		91.0	73.0	65.6	76.7	48.8	86.7	65.1	1173	70.1	75.5

Table 3. Comparison of task *CULane* ($\mathcal{D}_s$) $\rightarrow$ *CurveLanes* ($\mathcal{D}_T$) of CLRNet with ResNet-18 based on various IoU thresholds. We report the average score in % according to three runs.

IoU Threshold	Method	F1	Precision	Recall
0.3	DG	67.6	92.0	53.4
	TSUDA	68.8$_{\pm 0.1}$	92.5$_{\pm 0.1}$	54.8$_{\pm 0.1}$
	WSDAL	74.2$_{\pm 0.1}$	83.2$_{\pm 0.2}$	66.9$_{\pm 0.2}$
	$\mathcal{D}_T$	83.3	89.8	77.7
0.5	DG	61.4	82.5	48.5
	TSUDA	62.4$_{\pm 0.0}$	83.9$_{\pm 0.1}$	49.7$_{\pm 0.1}$
	WSDAL	64.1$_{\pm 0.1}$	71.9$_{\pm 0.1}$	57.8$_{\pm 0.2}$
	$\mathcal{D}_T$	75.2	81.1	70.1

distribution from *CurveLanes* dataset, as there are up to 14 annotated lanes and small radius of curvatures have high presence in the images. In spite of the difficulties with the proposed weakly-supervised method, improvements can be observed as the F1-score increases by 2%. Furthermore, similar to [30], we consider loose evaluation that tolerates higher positional deviations with IoU threshold 0.3 as the weak labels do not provide any direct supervision regarding positions of the lanes to ego vehicle, which are highly variational according to the annotator. The increasing disparity between the scores based on different IoU thresholds indicates that the weak labels provide helpful information for the adaptation to a more challenging dataset. Qualitative results can be found in Fig. 1. Besides that, the improved F1 score is due to the fact that the recall of the model increases significantly after applying WSDAL. The decrease in precision can be explained by the fact that the model increases the number of predictions on the target domain with possible positional drifts due to the lack of regression constraint during adaptation, shown in Fig. 10 of the Appendix.

Table 4. Applying WSDAL with 3 different backbones with CLRNet architecture from *CurveLanes* to *TuSimple*. Scores in brackets are the corresponding score trained directly on target domain. We report the average score in % according to three runs.

Backbone	#params	F1	FPR	FNR	Accuracy
ResNet18	11.7M	91.2 (95.3)	8.1 (5.5)	9.7 (3.8)	91.2 (95.1)
DLA34	15.8M	91.0 (96.3)	8.4 (4.1)	9.7 (3.2)	90.4 (95.6)
ConvNext	3.9M	88.9 (95.8)	11.0 (5.6)	11.6 (2.8)	90.6 (95.6)
ERFnet	2.4M	87.2 (95.2)	12.4 (6.0)	13.2 (3.4)	90.5 (95.5)

5. Ablation Study

5.1. Backbones

In order to verify that WSDAL improves the performance of models with different backbones on the target domain, we design experiments on CLRNet, as it achieves the best adaptation results from the previous experiments. For the ablation study, we choose in addition to ResNet three different backbones: DLA34 [57], ConvNextv2-atto [54] and the ERFNet [32] encoder, as they are representative backbones that are frequently used in research. Besides the ConvNext models that we reported in the previous tables, we show the models' performance on the task *CurveLanes* to *TuSimple* in Tab. 4. All of the proposed backbones improve their performance on the target domain using WSDAL compared with their unsupervised counterparts and close the gap to the upper bound. It is worth noticing that the trade-off between performance and learning capacity is not significant as the smallest model has only 15% of the parameters of the DLA34 model but they can achieve similar accuracy and 4% F1-score difference. But smaller models struggle to generate high quality pseudo labels, leading to higher performance disparity to the upper bound. Extensive results of the experiments on *TuSimple*, *CULane* and *CurveLanes* with various image encoders can be found in Appendix E.

Table 5. Ablation study for auxiliary segmentation task *CULane* → *TuSimple*. All values in %.

Model	Segmentation	F1	FPR	FNR	Accuracy
CLRNet		80.2	16.5	22.8	85.1
CLRNet	Instance	83.2	14.5	19.0	86.5
CLRNet	Binary	86.9	11.5	15.2	89.0
LaneATT		76.0	25.5	22.5	87.3
LaneATT	Instance	79.8	20.4	19.9	87.7
LaneATT	Binary	81.7	18.8	17.5	88.8

5.2. Auxiliary Task

In lane detectors, the lane segmentation task can enhance the model's performance in the source domain, but its impact on model generalization is less understood. We conducted an ablation study for the lane segmentation task within the WSDAL framework, with results presented in Tab. 5. It is evident that both auxiliary segmentation tasks significantly improve the performance of CLRNet and LaneATT in the target domain. This noteworthy improvement in the model's adaptation ability to the target domain is more encouraging compared to the limited performance boost observed in the source domain, which is extensively discussed in Appendix C. Anchor-based lane detection focuses on leveraging the intrinsic correlations of local information to predict overall lane parameters, while the lane segmentation task emphasizes the model's ability to identify semantic information of pixels discretely. The lane segmentation capability acquired through training on the source domain assists the model in better discovering lane foreground points through pixel-wise semantic features in the target domain and improving lane prediction accuracy. In both lane segmentation tasks, binary segmentation possesses greater universality across different domains and can be less influenced by the lane classification strategy in the source domain. This property makes the binary segmentation task better at enhancing the performance of lane detectors in the target domain. In addition, we also ablate the necessity of utilizing teacher–student network and auxiliary segmentation task for supporting WSDAL in Appendix D.

5.3. Unreliable Labels

As the NoL labels may come from map providers or other means of inferencing the NoL labels might be inaccurate, we ablate in this section the impact of utilizing unreliable NoL labels during weakly-supervised domain adaptation for lane detection models. We choose CLRNet as the base architecture for the evaluation. Similar to Sec. 5.1, we utilize various backbones to verify the results. For the experiments, we use an identical training setup to the ones that we deployed in the previous sections. During domain adaptation, instead of providing the correct NoL label from the image n, we replace the NoL label with a random number between zero and the known maximum NoL label from the dataset. We gradually increase the proportion of incorrect labels up to 100% and repeat the experiment for each error rate configuration three times using different seeds.

Fig. 3 depicts the F1-score of the model affected by the incorrect number of weak supervision labels. Overall, the transition point is located at around 30% error rate, indicating the proposed method outperforms the conventional UDA methods with suboptimal label quality from the target domain. In addition, we show in Appendix F that the transition point is also related to the properties of source and target datasets and the model architecture.

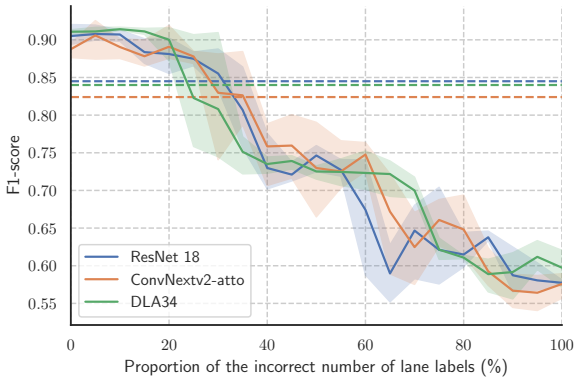

Figure 3. Results on *CurveLanes* → *Tusimple* with incorrect NoL labels on various image encoders with CLRNet. The dashed lines show the results from TSUDA.

6. Conclusion

In this work, we propose a novel Weakly Supervised Domain Adaptation framework for Lane Detection (WSDAL). The WSDAL framework consists of a teacher-student network and involves three tasks: lane detection, lane segmentation, and number of lanes prediction. The teacher network takes unaugmented images as input, while the student network takes augmented images for self-training. For lane detection and lane segmentation tasks, the teacher network can generate relatively accurate predictions, which are then converted into pseudo-labels to supervise the learning of the student network in the target domain. The number of lanes in target domain images serves as weak labels that affect the loss of the student's performance on the teacher-generated pseudo-labels. With an extensive range of experiments, we show that these weak labels enhance the confidence of lane predictions, enabling the detector to generate more high-confidence correct predictions in the target domain. Domain adaptation through weak supervision can reduce the false predictions of the model in the case of misaligned label distributions in the source and target domains. Furthermore, we show that the proposed method is realistically deployable due to its tolerance to imperfect labels.

References

[1] John Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(6):679–698, 1986. 2

[2] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*, 2017. 2

[3] Minghao Chen, Hongyang Xue, and Deng Cai. Domain adaptation for semantic segmentation with maximum squares loss. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 2090–2099, 2019. 3

[4] Jifeng Dai, Kaiming He, and Jian Sun. Boxsup: Exploiting bounding boxes to supervise convolutional networks for semantic segmentation. In *Proceedings of the IEEE international conference on computer vision*, pages 1635–1643, 2015. 2

[5] Anurag Das, Yongqin Xian, Dengxin Dai, and Bernt Schiele. Weakly-supervised domain adaptive semantic segmentation with prototypical contrastive learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 15434–15443, 2023. 3

[6] Mingfei Gao, Ang Li, Ruichi Yu, Vlad I Morariu, and Larry S Davis. C-wsl: Count-guided weakly supervised localization. In *Proceedings of the European conference on computer vision (ECCV)*, pages 152–168, 2018. 2

[7] Noa Garnett, Roy Uziel, Netalee Efrat, and Dan Levi. Synthetic-to-real domain adaptation for lane detection. In *Proceedings of the Asian Conference on Computer Vision*, 2020. 3

[8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 2

[9] Hiroto Honda and Yusuke Uchida. Clrernet: improving confidence of lane detection with laneiou. In *Proc. of the WACV*, 2024. 1, 2, 3, 4, 5, 6

[10] Lukas Hoyer, Dengxin Dai, Haoran Wang, and Luc Van Gool. Mic: Masked image consistency for context-enhanced domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11721–11732, 2023. 3

[11] Chuqing Hu, Sinclair Hudson, Martin Ethier, Mohammad Al-Sharman, Derek Rayside, and William Melek. Sim-to-real domain adaptation for lane detection and classification in autonomous driving. In *2022 IEEE Intelligent Vehicles Symposium (IV)*, pages 457–463, 2022. 2, 3

[12] Naoto Inoue, Ryosuke Furuta, Toshihiko Yamasaki, and Kiyoharu Aizawa. Cross-domain weakly-supervised object detection through progressive domain adaptation. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5001–5009, 2018. 3

[13] Seong Joon Oh, Rodrigo Benenson, Anna Khoreva, Zeynep Akata, Mario Fritz, and Bernt Schiele. Exploiting saliency for object segmentation from image level labels. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4410–4419, 2017. 2

[14] Guoliang Kang, Lu Jiang, Yi Yang, and Alexander G Hauptmann. Contrastive adaptation network for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4893–4902, 2019. 3

[15] Anna Khoreva, Rodrigo Benenson, Jan Hosang, Matthias Hein, and Bernt Schiele. Simple does it: Weakly supervised instance and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 876–885, 2017. 2

[16] Chenguang Li, Boheng Zhang, Jia Shi, and Guangliang Cheng. Multi-level domain adaptation for lane detection. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 4379–4388, 2022. 2, 3

[17] Xiang Li, Jun Li, Xiaolin Hu, and Jian Yang. Line-cnn: End-to-end traffic line detection with line proposal unit. *IEEE Transactions on Intelligent Transportation Systems*, 21:248–258, 2020. 2

[18] Yanghao Li, Naiyan Wang, Jianping Shi, Xiaodi Hou, and Jiaying Liu. Adaptive batch normalization for practical domain adaptation. *Pattern Recognition*, 80:109–117, 2018. 3

[19] Qing Lian, Lixin Duan, Fengmao Lv, and Boqing Gong. Constructing self-motivated pyramid curriculums for cross-domain semantic segmentation: A non-adversarial approach. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 6757–6766, 2019. 3

[20] Di Lin, Jifeng Dai, Jiaya Jia, Kaiming He, and Jian Sun. Scribblesup: Scribble-supervised convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3159–3167, 2016. 2

[21] Lizhe Liu, Xiaohao Chen, Siyu Zhu, and Ping Tan. Condlanenet: A top-to-down lane detection framework based on conditional convolution. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 3753–3762, 2021. 3, 2

[22] Tong Liu, Zhaowei Chen, Yi Yang, Zehao Wu, and Haowei Li. Lane detection in low-light conditions using an efficient data enhancement: Light conditions style transfer. In *2020 IEEE intelligent vehicles symposium (IV)*, pages 1394–1399. IEEE, 2020. 2

[23] Xiaofeng Liu, Chaehwa Yoo, Fangxu Xing, Hyejin Oh, Georges El Fakhri, Je-Won Kang, Jonghye Woo, et al. Deep unsupervised domain adaptation: A review of recent advances and perspectives. *APSIPA Transactions on Signal and Information Processing*, 11(1), 2022. 3

[24] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A convnet for the 2020s. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11966–11976, 2022. 2

[25] David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60:91–110, 2004. 2

[26] Fengmao Lv, Guosheng Lin, Peng Liu, Guowu Yang, Sinno Jialin Pan, and Lixin Duan. Weakly-supervised cross-domain road scene segmentation via multi-level curriculum

adaptation. *IEEE Transactions on Circuits and Systems for Video Technology*, 31:3493–3503, 2021. 3

[27] Fabio Maria Carlucci, Lorenzo Porzi, Barbara Caputo, Elisa Ricci, and Samuel Rota Bulo. Autodial: Automatic domain alignment layers. In *Proceedings of the IEEE international conference on computer vision*, pages 5067–5075, 2017. 3

[28] Davy Neven, Bert De Brabandere, Stamatios Georgoulis, Marc Proesmans, and Luc Van Gool. Towards end-to-end lane detection: An instance segmentation approach, 2018. 2

[29] OpenStreetMap contributors. Planet dump retrieved from https://planet.osm.org . https://www.openstreetmap.org, 2017. 2

[30] Xingang Pan, Jianping Shi, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Spatial as deep: Spatial cnn for traffic scene understanding. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32, 2018. 1, 2, 5, 7

[31] Zequn Qin, Huanyu Wang, and Xi Li. Ultra fast structure-aware deep lane detection. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIV 16*, pages 276–291. Springer, 2020. 2, 4

[32] Eduardo Romera, José M Alvarez, Luis M Bergasa, and Roberto Arroyo. Erfnet: Efficient residual factorized convnet for real-time semantic segmentation. *IEEE Transactions on Intelligent Transportation Systems*, 19(1):263–272, 2017. 7, 1

[33] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, pages 618–626, 2017. 2

[34] Inkyu Shin, Sanghyun Woo, Fei Pan, and In So Kweon. Two-phase pseudo label densification for self-training based domain adaptation. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIII 16*, pages 532–548. Springer, 2020. 3

[35] Krishna Kumar Singh, Hao Yu, Aron Sarmasi, Gautam Pradeep, and Yong Jae Lee. Hide-and-seek: A data augmentation technique for weakly-supervised localization and beyond. *arXiv preprint arXiv:1811.02545*, 2018. 2

[36] Kihyuk Sohn, David Berthelot, Nicholas Carlini, Zizhao Zhang, Han Zhang, Colin A Raffel, Ekin Dogus Cubuk, Alexey Kurakin, and Chun-Liang Li. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. *Advances in neural information processing systems*, 33:596–608, 2020. 1

[37] Chunfeng Song, Yan Huang, Wanli Ouyang, and Liang Wang. Box-driven class-wise region masking and filling rate guided loss for weakly supervised semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3136–3145, 2019. 2

[38] Bonifaz Stuhr, Johann Haselberger, and Julian Gebele. Carlane: A lane detection benchmark for unsupervised domain adaptation from simulation to multiple real-world domains. In *Advances in Neural Information Processing Systems*, pages 4046–4058. Curran Associates, Inc., 2022. 3

[39] Jinming Su, Chao Chen, Ke Zhang, Junfeng Luo, Xiaoming Wei, and Xiaolin Wei. Structure guided lane detection, 2021. 2

[40] Baochen Sun, Jiashi Feng, and Kate Saenko. Return of frustratingly easy domain adaptation. In *Proceedings of the AAAI conference on artificial intelligence*, 2016. 3

[41] Lucas Tabelini, Rodrigo Berriel, Thiago M. Paixao, Claudine Badue, Alberto F. De Souza, and Thiago Oliveira-Santos. Keep your eyes on the lane: Real-time attention-guided lane detection. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 294–302, 2021. 1, 2, 5

[42] Jigang Tang, Songbin Li, and Peng Liu. A review of lane detection methods based on deep learning. *Pattern Recognition*, 111:107623, 2021. 2

[43] Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *Advances in neural information processing systems*, 30, 2017. 3

[44] TuSimple. Tusimple dataset benchmark. https://github.com/TuSimple/tusimple-benchmark, 2017. 5

[45] Eric Tzeng, Judy Hoffman, Ning Zhang, Kate Saenko, and Trevor Darrell. Deep domain confusion: Maximizing for domain invariance. *arXiv preprint arXiv:1412.3474*, 2014. 3

[46] Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell. Adversarial discriminative domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7167–7176, 2017. 3

[47] Tuan-Hung Vu, Himalaya Jain, Maxime Bucher, Matthieu Cord, and Patrick Pérez. Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2517–2526, 2019. 3

[48] Fang Wan, Pengxu Wei, Jianbin Jiao, Zhenjun Han, and Qixiang Ye. Min-entropy latent model for weakly supervised object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1297–1306, 2018. 2

[49] Jinsheng Wang, Yinchao Ma, Shaofei Huang, Tianrui Hui, Fei Wang, Chen Qian, and Tianzhu Zhang. A keypoint-based global association network for lane detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1392–1401, 2022. 5, 2

[50] Qi Wang, Junyu Gao, and Xuelong Li. Weakly supervised adversarial domain adaptation for semantic segmentation in urban scenes. *IEEE Transactions on Image Processing*, 28: 4376–4386, 2019. 3

[51] Yuting Wang, Ricardo Guerrero, and Vladimir Pavlovic. D2f2wod: Learning object proposals for weakly-supervised object detection via progressive domain adaptation. In *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 22–31, 2023. 3

[52] Yunchao Wei, Jiashi Feng, Xiaodan Liang, Ming-Ming Cheng, Yao Zhao, and Shuicheng Yan. Object region mining with adversarial erasing: A simple classification to semantic segmentation approach. In *2017 IEEE Conference on Com-*

puter Vision and Pattern Recognition (CVPR), pages 6488–6496, 2017. 2

[53] Tuopu Wen, Diange Yang, Kun Jiang, Chunlei Yu, Jiaxin Lin, Benny Wijaya, and Xinyu Jiao. Bridging the gap of lane detection performance between different datasets: Unified viewpoint transformation. *IEEE Transactions on Intelligent Transportation Systems*, 22:6198–6207, 2021. 3

[54] Sanghyun Woo, Shoubhik Debnath, Ronghang Hu, Xinlei Chen, Zhuang Liu, In So Kweon, and Saining Xie. Convnext v2: Co-designing and scaling convnets with masked autoencoders. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16133–16142, 2023. 7, 1

[55] Hang Xu, Shaoju Wang, Xinyue Cai, Wei Zhang, Xiaodan Liang, and Zhenguo Li. Curvelane-nas: Unifying lane-sensitive architecture search and adaptive point blending, 2020. 1, 2, 5

[56] Chunlei Yu, Tuopu Wen, Long Chen, and Kun Jiang. Common bird-view transformation for robust lane detection. In *2019 IEEE 9th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER)*, pages 665–669, 2019. 3

[57] Fisher Yu, Dequan Wang, Evan Shelhamer, and Trevor Darrell. Deep layer aggregation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2403–2412, 2018. 7, 1

[58] Dingwen Zhang, Wenyuan Zeng, Jieru Yao, and Junwei Han. Weakly supervised object detection using proposal- and semantic-level relationships. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(6):3349–3363, 2022. 2

[59] Youcheng Zhang, Zongqing Lu, Xuechen Zhang, Jing-Hao Xue, and Qingmin Liao. Deep learning in lane marking detection: A survey. *IEEE Transactions on Intelligent Transportation Systems*, 23:5976–5992, 2022. 2

[60] Tu Zheng. Lanedet. https://github.com/Turoad/lanedet, 2021. 1

[61] Tu Zheng, Hao Fang, Yi Zhang, Wenjian Tang, Zheng Yang, Haifeng Liu, and Deng Cai. Resa: Recurrent feature-shift aggregator for lane detection. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35:3547–3554, 2021. 1, 2

[62] Tu Zheng, Yifei Huang, Yang Liu, Wenjian Tang, Zheng Yang, Deng Cai, and Xiaofei He. Clrnet: Cross layer refinement network for lane detection. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 888–897, 2022. 1, 2, 3, 4, 5

[63] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba. Learning deep features for discriminative localization. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2921–2929, 2016. 2

[64] Xizhou Zhu, Weijie Su, Lewei Lu, Bin Li, Xiaogang Wang, and Jifeng Dai. Deformable detr: Deformable transformers for end-to-end object detection. In *Proceedings of ICLR*, 2021. 2