HD Maps are Lane Detection Generalizers:
A Novel Generative Framework for Single-Source Domain Generalization

Daeun Lee*
Korea University
goodgpt@korea.ac.kr

Minhyeok Heo
NAVER LABS
heo.minhyeok@naverlabs.com

Jiwon Kim
NAVER LABS
g1.kim@naverlabs.com

Abstract

Lane detection is a vital task for vehicles to navigate and localize their position on the road. To ensure reliable driving, lane detection models must have robust generalization performance in various road environments. However, despite the advanced performance in the trained domain, their generalization performance still falls short of expectations due to the domain discrepancy. To bridge this gap, we propose a novel generative framework using HD Maps for Single-Source Domain Generalization (SSDG) in lane detection. We first generate numerous front-view images from lane markings of HD Maps. Next, we strategically select a core subset among the generated images using (i) lane structure and (ii) road surrounding criteria to maximize their diversity. In the end, utilizing this core set, we train lane detection models to boost their generalization performance. We validate that our generative framework from HD Maps outperforms the Domain Adaptation model MLDA with +3.01%p accuracy improvement, even though we do not access the target domain images.

1. Introduction

Lane detection is a fundamental task in autonomous driving. By detecting lane markings on the road, vehicles can localize their position and identify drivable spaces. However, current data-driven lane detection models still suffer from performance degradation when tested on unseen domains. Such discrepancies between the source and target datasets result in a severe performance decrease.

Recent studies [2, 6–8, 10] have attempted to address this performance degradation. These approaches include Domain Adaptation (DA), which involves making use of a limited number of data samples from the target domain, and Domain Generalization (DG), which leverages data from multiple source domains. However, in realistic driving scenarios, it is technically impossible to obtain even a small number of data samples from the target domain since they cannot be predetermined. Collecting and annotating data manually from numerous regions also have practical limitations. Therefore, it is more realistic to achieve generalization solely relying on a single-source domain without any prior information about the target domain.

From the Single-Source Domain Generalization (SSDG) perspective, it is imperative to obtain domain-invariant features, which necessitates a sufficiently *diverse* set of source domain data. To achieve this diversity, previous works [3, 11, 12, 14, 17, 23] actively adopt augmentation techniques such as Augmix [3] or adversarial methods [12, 17]. However, SSDG for lane detection has been unexplored, even though the significance for the real-world application.

Therefore, we propose a *model-agnostic* generative framework to generalize lane detection models adopting *HD Maps* as a *single-source domain*. To the best of our knowledge, this is the first paper addressing the domain shifts in lane detection with SSDG setup. Notably, we notify that HD Maps are effective lane-detection generalizers thanks to their flexibility in collecting numerous datasets depending on the camera parameters. Utilizing this advantage, we first get the 2D lane masks from pre-built HD Maps and generate lane-conditioned images with various surroundings. Next, we strategically select a core set among the generated images to maximize diversity by using *(i) lane structure* and *(ii) road surrounding* criteria. This can help to boost the robustness of our single-source domain from HD Maps generation. In the end, we train lane detection models regarding this core set as a single-source domain.

The main contributions of the paper are summarized as follows:

- We propose a model-agnostic generative framework to generalize a lane detection model utilizing a single-source domain from HD Maps, addressing the SSDG problem in lane detection for the first time.
- After decomposing the factor of domain shifts with *(i) lane structure* and *(ii) road surrounding*, we propose the core set selection criteria to maximize the diversity of single-source domain.

---

*Work done during the research internship at NAVER LABS.

- Diverse experimental validations have demonstrated that the proposed training framework outperforms previous DA methods without any target domain information.

2. Methodology

We first introduce the coreset selection criteria for lane and surroundings in Section 2.1. Next, we describe our total framework including the lane-conditioned generation from HD Maps and lane detection model training process in Section 2.2.

Problem Statement. For the Single-Source Domain Generalization (SSDG) in lane detection, we use source domain $\boldsymbol{D^s} = \{(I_i, M_i)\}_{i=1}^{N}$ which consists N pairs of front-view (i.e., *surrounding*) image I and lane mask M. Notably, our work demonstrates HD Maps provide a successful source for mitigating domain discrepancy in lane detection. To train the lane detection model as robustly as possible to the unseen target domains $\boldsymbol{D^t}$, we propose a diversity-aware framework by selecting a core set for the lane and surrounding.

2.1. Diversity-aware Coreset Selection Criteria

We first introduce a coreset selection method to boost the diversity of a single domain.

Data Selection Policy. In this work, we define *maximal diversity* as *minimum sum of similarities* between nodes in a graph. Let's assume a complete weighted graph $G = (V, E, W)$ where the vertex set $V = \{v_1, v_2, ..., v_N\}$ represents N data points, and each data point is connected to every other data point through an edge $E = \{(v_i, v_j)\}$. The weight of each edge $W = \{w(i, j)\}$ is defined as the similarity between the two data points connected by the edge. The near-optimal solution can be configured as Algorithm 1. This is freely applied to various types of data by defining each type of metric, hence it is easily expanded to both lane structure and surrounding image. Therefore, we define two similarity criteria for *(i) lane* and *(ii) surrounding images* as follows.

(i) Criteria for Lane Structure. We propose a lane similarity function f between two lane masks $\mathbf{M}_i, \mathbf{M}_j$ that consists of several lanes ℓ with its curvatures. Inspired by Eigenlanes [4], we project each lane ℓ onto the image plane and compare the similarity represented in pixel coordinates. More specifically, we express a lane as a set of 2D points uniformly sampled along the vertical direction in the image plane, i.e., $\ell = [x_1, x_2, ..., x_p]^T$, where x_i denotes the x-coordinate of the i-th sample, and P is the number of points. Then, we construct a lane matrix $\boldsymbol{A} = [\ell_1, \ell_2, ..., \ell_L]$ by utilizing all L lane curves from the entire training data. By applying the SVD to the constructed lane matrix $\boldsymbol{A} \in \mathbb{R}^{P \times L}$, we factorize it as follows:

$$\boldsymbol{A} = \boldsymbol{U} \Sigma \boldsymbol{V}^T, \tag{1}$$

Algorithm 1 Data Selection Policy
***
Require: Complete weighted graph G and $K = \#$ of vertices
 1: Find the vertex pair (i, j) which has minimum $w(i, j)$
 2: Add i and j to Q
 3: **while** $|Q| < K$ **do**
 4:     Find vertex k which minimizes $\sum_{l \in V \setminus Q} w(k, l)$
 5: **end while**
Ensure: Q

---

where $\boldsymbol{U} = [\mathbf{u}_1, ..., \mathbf{u}_P] \in \mathbb{R}^{P \times s}$ and $\boldsymbol{V} = [\mathbf{v}_1, ..., \mathbf{v}_L] \in \mathbb{R}^{s \times L}$ are orthogonal matrices, and $\Sigma \in \mathbb{R}^{s \times s}$ is a singular value matrix.

Using this, each lane ℓ is approximated as $\tilde{\ell}$ using R-rank approximation with singular vectors $\mathbf{u}_1, ..., \mathbf{u}_R$ as follows:

$$\tilde{\ell}_i = \mathbf{U}_R \mathbf{c}_i = [\mathbf{u}_1, ..., \mathbf{u}_R]\mathbf{c}_i. \tag{2}$$

where $\mathbf{U}_R \in \mathbb{R}^{P \times R}$. The coefficient vector $\mathbf{c}_i$ of each lane ℓ_i in the *eigenlane* space spanned by the singular vectors $[\mathbf{u}_1, ..., \mathbf{u}_R]$ is expressed as follows:

$$\mathbf{c}_i = \mathbf{U}_R^T \ell_i = [c_1, ..., c_R]. \tag{3}$$

Therefore, we simply define the similarity between two lanes ℓ_i and ℓ_j as Euclidean distance d by projecting them onto the *eigenlane* space as follows:

$$d(\ell_i, \ell_j) = |\mathbf{c}_i - \mathbf{c}_j|. \tag{4}$$

Moreover, our goal is to extend this lane-wise similarity to lane masks. Given two lane masks $\mathbf{M}_i = [\ell_0^i, ..., \ell_m^i]$ and $\mathbf{M}_j = [\ell_0^j, ..., \ell_n^j]$, each containing m and n lanes in their lane masks, respectively. Assuming that $m \geq n$ always holds, we define the similarity function f between lane masks $\mathbf{M}_i, \mathbf{M}_j$ as follows:

$$f(\mathbf{M}_i, \mathbf{M}_j) = \sum_k^n \min d(\ell_l^i, \ell_k^j) + |m - n|\, \kappa, \tag{5}$$

where κ denotes the penalty constant which we describe next. We believe that the number of lanes is also an important factor in determining the similarity. Therefore, when two-lane masks contain different numbers of lanes, we penalize their similarity by adding hyperparameter κ. With the above definition of similarity between lane masks, we select the K_ℓ number of lane masks by using the simple selection algorithm in Algorithm 1.

(ii) Criteria for Surrounding Image. By training the model on RGB images that have the *same* lane masks but *different* image surroundings, we aim to make the model less sensitive to variations in those surroundings. However, unlike the case of lane layouts, it is difficult to obtain a quantitative similarity metric between surroundings of RGB images. Therefore, we adopt a comparison metric
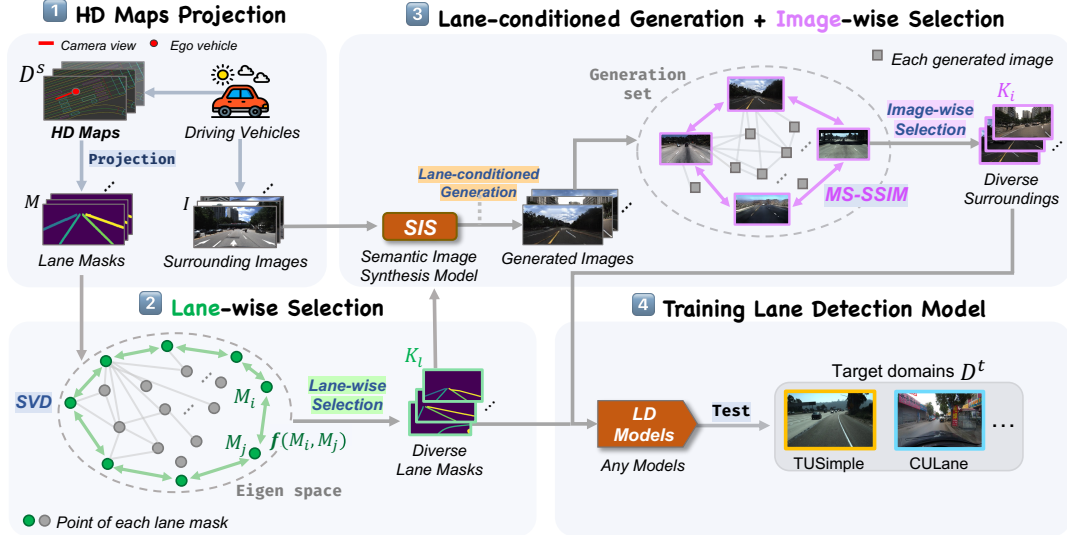
Figure 1. **The overall framework.** Our framework adopts HD Maps as a single-source domain to build robust lane detection models. We first extract numerous lane masks from HD Maps with surrounding images. Then, we train SIS (Semantic Image Synthesis) generative models to synthesize lane-conditioned images with diverse surroundings. We select a core set among synthesized images with two criteria to maximize its diversity: *(i) lane structure* and *(ii) surroundings*. Then, we finally train the lane detection model and obtain enhanced generalization performance.

commonly used to check the quality of generated images by image synthesis models.

The Structural Similarity Index Measure (SSIM) [22] and its Multi-Scale extension (MS-SSIM) [21] are common metrics to evaluate the perceived quality or similarity of images. Since MS-SSIM has also been used to compare the quality of generated images against the reference images [15], we choose the MS-SSIM as the similarity metric between two surrounding images. Therefore, we select K_i surrounding images using MS-SSIM using the Algorithm 1.

2.2. Overall Framework

Combining with our coreset selection criteria, we now introduce a diversity-aware framework to raise the robustness of the lane detection model.

Extracting Lane Masks with Lane Selection. We note that the huge potential of HD Maps is the flexibility of extracting diverse lane masks according to camera parameters. Using them to pre-built HD Maps, we extract the pair of lane masks M with their surrounding image I by projecting the lane objects onto the camera image plane at the current location. It is worth noting that this process does not require any additional annotation to obtain the pixel-wise labels. Also, with this projection, we acquire a substantial number of lane masks from HD Maps easily.

After getting these N pairs of lane masks and surrounding images, we adopt the aforementioned lane-wise selection criteria to boost lane diversity. Finally, we select K_ℓ number of lane masks among N where $K_\ell < N$.

Lane-conditioned Generation with Image Selection.

Next, utilizing these K_ℓ selected lane masks, we adopt Semantic Image Synthesis (SIS) methods [16, 20] to generate lane-conditioned images. Unlike conventional SIS approaches, we use only lane masks as a condition to generate images to make diverse surroundings. We generate 100 RGB images for each lane label mask.

After that, we now adopt the aforementioned image selection criteria to select diverse surroundings among generated images. We finally select K_i number of generations.

Training Lane Detection Model with Coreset. Combining selections from both lane masks and surroundings, we build the coreset which maximized diversity. Please note that this coreset is originally from HD Maps. Then, we finally train the lane detection model using this dataset. It is noteworthy that our framework is model-agnostic, which means it applies all lane detection models like data augmentation.

3. Experiments

3.1. Datasets

LabsLane (HD Maps). We use Naver Labs HD Maps[1] to create lane label masks, which are publicly available for research purposes and provide rich information about the road environment in cities of South Korea. During the projection of the lane object onto the camera view, we only consider a maximum of four lanes near the ego vehicle, which is consistent with other benchmark datasets that also have a maxi-

---

[1]https://naverlabs.com/en/datasets

Table 1. **Quantitative results on CULane target domain**. †Note that we attempted to reproduce [5], but were not even able to reproduce their performance. In this table, we present their results as they are reported.

Experiment setting	Trained data			Total	Normal	Crowded	Night	No line	Shadow	Arrow	Dazzle light	Curve	Crossroad
	TUSimple	CULane	LabsLane										
Source only - ERFNet	✔			24.2	41	19.6	9.1	12.7	12.7	28.8	11.4	32.6	1240
Source only - GANet	✔			30.5	50	25.5	19.6	18.5	16.8	36	25.4	34.2	7405
Advent [18]	✔	✔		30.4	49.3	24.7	20.5	18.4	16.4	34.4	26.1	34.9	6527
PyCDA [9]	✔	✔		25.1	41.8	19.9	13.6	15.1	13.7	27.8	18.2	29.6	4422
Maximum Squares [1]	✔	✔		31	50.5	27.2	20.8	19	20.4	40.1	27.4	38.8	10324
MLDA [5]†	✔	✔		38.4	**61.4**	**36.3**	27.4	21.3	**23.4**	49.1	30.3	43.4	11386
Source only - GANet			✔	34.9	52.4	25.9	27.1	18.8	20.3	44.5	25.9	42.7	**939**
Ours - ERFNet			✔	38.2	56.8	32.5	24.4	**23.8**	22.3	51.2	**32.2**	40.9	4572
Ours - GANet			✔	**39.6**	56.7	34	**30.1**	21.2	22	**53.2**	27.1	**46.9**	1275

Table 2. **Quantitative results on TUSimple target domain.** Our framework using LabsLane shows similar performance with Target only.

Experiment setting	Trained data			Acc	FP	FN
	CULane	TUSimple	LabsLane			
Source only - ERFNet	✔			60.9	31.6	55.2
Source only - GANet	✔			76.9	25.8	35
Target only - GANet		✔		95.9	1.97	2.62
Advent [18]	✔	✔		77.1	39.7	43.9
PyCDA [9]	✔	✔		80.9	51.9	45.1
Maximum Squares [1]	✔	✔		76	38.2	42.8
MLDA [5]	✔	✔		89.7	29.5	18.4
Ours			✔	**92.4**	**6.7**	**9.4**

(a) Lane label mask   (b) Selected 5 synthesized RGB images

Figure 2. **Examples of generated dataset.** We utilized OA-SIS [16] as our generative model to visualize results.

mum of four or five lanes. As a result, the generated dataset, dubbed LabsLane, consists of 7,184 RGB images and corresponding 6 DoF poses.

CULane. It consists of complex *urban road environments* and 133,235 frames, of which 88,880, 9,675, and 34,680 are used for training, validation, and testing, respectively.

TUSimple. It mostly consists of *highway* driving scenes and around 7,000 video clips, each containing 20 frames, with labeled lanes on the final frame of each clip.

3.2. Implementation details

We used GANet-small [19] and ERFNet [13] as lane detection models which are respectively keypoint-based and segmentation-based methods. For the ERFNet setting, we faithfully followed the MLDA [5] setup. We use the F1 score and accuracy respectively to measure the performance metric for CULane and TUSimple like the official evaluation code.

3.3. Main results

Results on CULane. In Table 1, we summarize the performances of each lane detection model on CULane dataset, which includes various scenario cases. We compared our methods with domain adaptation methods such as Advent [18], PyCDA [9], Maximum Squares [1], and MLDA [5]. Although our approach is an SSDG setup, we

observe that our total performance outperforms MLDA. In particular, the performance improvement is greater when it is not a typical surrounding scene (e.g. Night, Dazzle light). This demonstrates that our HD Map-based diversity maximization framework robustly handles large domain shifts.

Results on TUSimple. We also validate our proposed framework on the TUSimple dataset in Table 2. We found that our framework also outperforms the DA-based methods even though we do not use target domain information. It is especially notable that both MLDA [5] and the proposed method showed significant performance improvements compared to the results on the CULane dataset.

Generated Images from SIS. In Figure 2, we show the generated images from simple SIS models OASIS [16]. Our framework succeeds in generating diverse surroundings from numerous lane markings from HD Maps. We believe there is more room for improvement when we adopt diffusion-based models.

4. Conclusion

In this paper, we propose a novel training framework that utilizes HD Maps for robust lane detection. We generated a large number of images from HD Maps with varying surroundings and lane structures to mitigate domain discrepancy on lane detection.

References

[1] Minghao Chen, Hongyang Xue, and Deng Cai. Domain adaptation for semantic segmentation with maximum squares loss. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2090–2099, 2019. 4

[2] Qi Dou, Daniel Coelho de Castro, Konstantinos Kamnitsas, and Ben Glocker. Domain generalization via model-agnostic learning of semantic features. *Advances in neural information processing systems*, 32, 2019. 1

[3] Dan Hendrycks, Norman Mu, Ekin D Cubuk, Barret Zoph, Justin Gilmer, and Balaji Lakshminarayanan. Augmix: A simple data processing method to improve robustness and uncertainty. *arXiv preprint arXiv:1912.02781*, 2019. 1

[4] Dongkwon Jin, Wonhui Park, Seong-Gyun Jeong, Heeyeon Kwon, and Chang-Su Kim. Eigenlanes: Data-driven lane descriptors for structurally diverse lanes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17163–17171, 2022. 2

[5] Chenguang Li, Boheng Zhang, Jia Shi, and Guangliang Cheng. Multi-level domain adaptation for lane detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4380–4389, 2022. 4

[6] Da Li, Yongxin Yang, Yi-Zhe Song, and Timothy M Hospedales. Deeper, broader and artier domain generalization. In *Proceedings of the IEEE international conference on computer vision*, pages 5542–5550, 2017. 1

[7] Haoliang Li, Sinno Jialin Pan, Shiqi Wang, and Alex C Kot. Domain generalization with adversarial feature learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5400–5409, 2018.

[8] Ya Li, Xinmei Tian, Mingming Gong, Yajing Liu, Tongliang Liu, Kun Zhang, and Dacheng Tao. Deep domain generalization via conditional invariant adversarial networks. In *Proceedings of the European conference on computer vision (ECCV)*, pages 624–639, 2018. 1

[9] Qing Lian, Fengmao Lv, Lixin Duan, and Boqing Gong. Constructing self-motivated pyramid curriculums for cross-domain semantic segmentation: A non-adversarial approach. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6758–6767, 2019. 4

[10] Krikamol Muandet, David Balduzzi, and Bernhard Schölkopf. Domain generalization via invariant feature representation. In *International conference on machine learning*, pages 10–18. PMLR, 2013. 1

[11] Xi Peng, Fengchun Qiao, and Long Zhao. Out-of-domain generalization from a single source: An uncertainty quantification approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022. 1

[12] Fengchun Qiao, Long Zhao, and Xi Peng. Learning to learn single domain generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12556–12565, 2020. 1

[13] 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. 4

[14] Seonguk Seo, Yumin Suh, Dongwan Kim, Geeho Kim, Jongwoo Han, and Bohyung Han. Learning to optimize domain specific normalization for domain generalization. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXII 16*, pages 68–83. Springer, 2020. 1

[15] Jake Snell, Karl Ridgeway, Renjie Liao, Brett D Roads, Michael C Mozer, and Richard S Zemel. Learning to generate images with perceptual similarity metrics. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 4277–4281. IEEE, 2017. 3

[16] Vadim Sushko, Edgar Schönfeld, Dan Zhang, Juergen Gall, Bernt Schiele, and Anna Khoreva. You only need adversarial supervision for semantic image synthesis. *arXiv preprint arXiv:2012.04781*, 2020. 3, 4

[17] Riccardo Volpi, Hongseok Namkoong, Ozan Sener, John C Duchi, Vittorio Murino, and Silvio Savarese. Generalizing to unseen domains via adversarial data augmentation. *Advances in neural information processing systems*, 31, 2018. 1

[18] 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. 4

[19] 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. 4

[20] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8798–8807, 2018. 3

[21] Zhou Wang, Eero P Simoncelli, and Alan C Bovik. Multiscale structural similarity for image quality assessment. In *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, pages 1398–1402. Ieee, 2003. 3

[22] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 3

[23] Qinwei Xu, Ruipeng Zhang, Yi-Yan Wu, Ya Zhang, Ning Liu, and Yanfeng Wang. Simde: A simple domain expansion approach for single-source domain generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4797–4807, 2023. 1