SEMI-SUPERVISED LANE DETECTION WITH DEEP HOUGH TRANSFORM

Yancong Lin   *Silvia-Laura Pintea*   *Jan van Gemert*

Vision Lab, Delft University of Technology, The Netherlands

ABSTRACT

Current work on lane detection relies on large manually annotated datasets. We reduce the dependency on annotations by leveraging massive cheaply available unlabelled data. We propose a novel loss function exploiting geometric knowledge of lanes in Hough space, where a lane can be identified as a local maximum. By splitting lanes into separate channels, we can localize each lane via simple global max-pooling. The location of the maximum encodes the layout of a lane, while the intensity indicates the the probability of a lane being present. Maximizing the log-probability of the maximal bins helps neural networks find lanes without labels. On the CULane and TuSimple datasets, we show that the proposed Hough Transform loss improves performance significantly by learning from large amounts of unlabelled images.

Index Terms— Lane detection, Hough Transform, semi-supervised learning

1. INTRODUCTION

One key component of self-driving cars is the lane-keeping assist [1, 2], which actively keeps the vehicles in the marked lanes. The lane-keeping assist relies on accurate lane detection in the wild, which is a highly challenging task because of illumination and appearance variations, traffic flow, and new unseen driving scenarios [3].

State-of-the-art deep learning methods for lane detection perform remarkably well on benchmark datasets [3, 4, 5, 6]. However, they rely on deep networks powered by massive amounts of labelled data. Although the data itself can be obtained at relatively low cost, it's their annotations that are laborious and thus expensive [7]. Moreover, the existing curated datasets do not cover all the possible driving scenarios that could be encountered in real-world situations. Being able to leverage additional realistic unlabelled training data would allow for a more robust lane detection system.

To make effective use of additional unlabelled data, we propose a semi-supervised Hough Transform-based loss which exploits geometric prior knowledge of lanes in the Hough space [8, 9].

Lanes are lines, thus we propose a semi-supervised Hough Transform loss that parameterizes lines in Hough space, by mapping them to individual bins represented by an offset and an angle. Inspired by the work in [9], we rely on a trainable Hough Transform and Inverse Hough Transform (HT-IHT) module embedded into a neural network to learn Hough representations for lane detection. We subsequently extend its use for semi-supervised training, by noting that the presence of lanes leads to Hough bins with maximal votes. Maximizing the log-probability of these Hough bins requires no human supervision, enabling the network to detect lanes in unlabelled images.

This paper makes the following contributions: (1) we present an annotation-efficient approach for lane detection in a semi-supervised way; (2) to this end, we propose a novel loss function to exploit prior geometric knowledge of lanes in Hough space; (3) we experimentally show improved performance on the CULane [10] and TuSimple [11] datasets, given large amounts of unlabelled data.

2. RELATED WORK

Lane detection methods. Classic work on lane detection is based on knowledge-based manually designed geometric features. Examples include grouping image gradients [12, 13, 14, 15], or line detection techniques through Hough Transform [16, 17, 18, 19] relying on local edges extracted using image gradients. A main drawback of such knowledge-based methods is their inability to handle complex scenarios where traffic flow and illumination conditions change dramatically. Here, we address this by learning appearance variation of lanes in a deep network, while still relying on the Hough Transform as prior knowledge for line detection [8, 9].

Recently, deep neural networks have been employed for efficient lane detection, replacing well-engineered features. Typically, the learning-based methods treat the lane detection as a semantic segmentation task and learn semantic features from large datasets [1, 20, 21, 22, 23, 24]. In contrast to these works we improve the prediction accuracy by leveraging massive unlabelled data through semi-supervised learning.

Semi-supervised methods. Semi-supervised methods solve the learning task by relying on both labelled and unlabelled data [25], and are divided into: inductive approaches constructing a classifier over labelled and unlabelled data [26, 27, 28], and transductive approaches propagating where the task information is shared between data points [29, 30, 31].

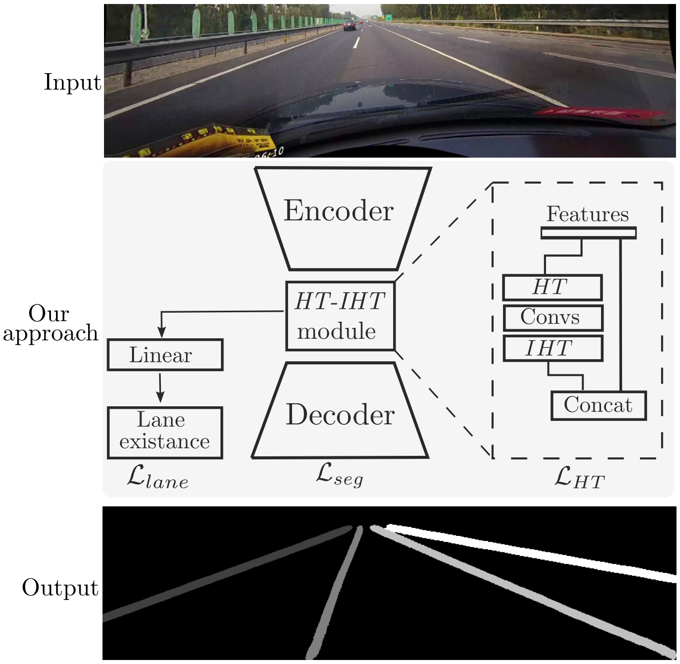

Fig. 1: **Overview of our model.** We have an encoder, a decoder, and a fully connected layer inspired by the ERFNet [4, 5] with a trainable Hough-Transform (HT) and Inverse Hough-Transform (IHT) module [9], on top of which we build our HT-based semi-supervised loss maximizing the probability of the maximal bins in Hough space, where $\mathcal{L}_{lane}$, $\mathcal{L}_{seg}$, and $\mathcal{L}_{HT}$ are the optimized loss functions.

A self-driving car has no access to the test statistics, therefore we consider the inductive case.

3. SEMI-SUPERVISED LANE DETECTION

Given an input image, our model outputs a lane probability and a semantic segmentation mask of lane pixels. We use as a starting point the popular ERFNet [5][1]. The ERFNet contains a convolutional encoder for deep feature extraction, a convolutional decoder for lane predictions, and a fully connected layer for predicting the probability of a lane. We insert a trainable Hough Transform and Inverse Hough Transform (*HT-IHT*) block [9] between the encoder and decoder, and utilize the Hough representations of lanes for semi-supervised learning. Fig. 1 depicts the overall structure of our model.

3.1. Hough Transform line priors

We encode an input image to a semantic feature representations F which is mapped to the Hough space, through a trainable Hough Transform module [9]. The Hough transform HT maps a feature map F of size $[H \times W]$ to an $[N_\rho \times N_\theta]$ Hough histogram, where N_ρ and N_θ are the number of discrete offsets and angles. Pixels along lines in F are mapped into discrete pairs of offsets ρ and angles θ. Specifically, given a line direction indexed by i with its corresponding pixels (x_i, y_i), they all vote in the Hough space for the closest bin (ρ, θ):

$$HT(\rho, \theta) = \sum_i F(x_i, y_i), \tag{1}$$

where the mapping is given by $\rho = x_i \cos\theta + y_i \sin\theta$.

We perform a set of 1D convolutions in Hough space over the offset direction and apply an Inverse Hough Transform IHT module mapping the $[N_\rho \times N_\theta]$ Hough histogram back to an $[H \times W]$ feature map [9]. The IHT maps bins (ρ, θ) to pixels (x_i, y_i) by averaging all the HT bins where a certain pixel has voted:

$$IHT(x_i, y_i) = \frac{1}{N_\theta} \sum_\theta HT(x_i \cos\theta + y_i \sin\theta, \theta). \tag{2}$$

We concatenate the features F with the IHT features, followed by a convolutional layer merging these two branches. We set $H = 26$, $W = 122$, $N_\rho = 125$ and $N_\theta = 60$.

3.2. Hough Transform loss for unlabelled data

A lane is composed of a set of line segments with a certain width, that share the same orientation. For unlabelled images we rely on the observation that lanes correspond to local maxima in the Hough space. Since the ERFNet [4, 5] predicts a single lane in each output channel, the mapping to Hough space recovers the lanes as global maxima in their respective channels. Having a large global maximum indicates that pixels along that line direction are well aligned, thus falling in the same bin. Based on this observation, we provide supervision to unlabelled inputs by maximizing the log-probability of the maximum bin $(\hat{\rho}, \hat{\theta})$ in Hough domain. To give the HT bins a probabilistic interpretation, we rescale the HT maps between $[0, 1]$ for each angle direction independently by applying an L_1 normalization over the offset dimension:

$$\mathcal{L}_{HT} = -\log\left(\frac{HT(\hat{\rho}, \hat{\theta})}{\sum_{k=0}^{N_\rho} HT(\rho_k, \hat{\theta})}\right), \tag{3}$$

where $(\hat{\rho}, \hat{\theta})$ is the positions of the global maximum in Hough space, calculated from the predicted segmentation masks.

3.3. Training with both labelled and unlabelled data

We train our model with both labelled and unlabelled data. As in [3, 4] the network predicts for each channel a mask used in a cross entropy loss $\mathcal{L}_{\text{seg}}$ over labelled data for predicting the semantic segmentation. Additionally, the network predicts lane probabilities p which are used in a binary cross entropy loss over labelled data $\mathcal{L}_{\text{lane}}$ for optimizing for the existence of a lane. We also optimize the proposed $\mathcal{L}_{HT}$, only

[1] We rely on the implementation in [4]: `https://github.com/cardwing/Codes-for-Lane-Detection`

when the predicted probability p of a lane is larger than a threshold τ; otherwise, we skip the corresponding lane. We set $\tau = 0.9$. The total loss is the combination of the three losses:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{seg}} + \alpha \mathcal{L}_{\text{lane}} + \beta \mathcal{L}_{HT}(p, \tau), \tag{4}$$

where α and β are used to balance different loss terms.

4. EXPERIMENTAL ANALYSIS

Datasets. We evaluate our models on the TuSimple dataset [11] and CULane dataset [10]. All video clips in TuSimple dataset are taken on highways. There are 3,626 frames for training and 2,782 frames for testing. The CULane dataset contains images from 9 different driving scenarios, such as lanes in shadow and at night with poor lighting conditions. There are 88,880 images for training, 9,675 for validation, and 34,680 images for testing. We follow the official evaluation protocol to measure accuracy on the TuSimple, and use F_1 measure on the CULane dataset.

Baselines. We compare with the baseline ERFNet [5], and with the ERFNet-HT using the *HT-IHT* block [9]. Both models are trained from scratch with labelled data only. For semi-supervised learning, we consider the ERFNet-*pseudo* pseudo-labeling baseline, and our proposed ERFNet-HT-$\mathcal{L}_{HT}$. The ERFNet-*pseudo* baseline first learns to predict lanes on annotated data only, and subsequently uses the predicted pseudo-labels to annotate unlabelled data, and then retrains the model on all data. ERFNet-*pseudo* treats the prediction with a confidence score larger than 0.9 as ”ground truth” and optimize the $\mathcal{L}_{\text{seg}}$ with pseudo-labels. ERFNet-HT-$\mathcal{L}_{HT}$ uses our proposed $\mathcal{L}_{HT}$ loss. ERFNet-HT-*pseudo*+$\mathcal{L}_{HT}$ combines both pseudo-labelling and our proposed $\mathcal{L}_{HT}$ loss. Additionally, we also compare with s4GAN [32], a state-of-the-art semi-supervised learning model for semantic segmentation.

Implementation details. We follow the implementation and hyper-parameters in [4]. We use SGD [33] to train ERFNet and ERFNet-HT for 24 epochs. ERFNet-*pseudo*, ERFNet-HT-$\mathcal{L}_{HT}$ and ERFNet-HT-*pseudo*+$\mathcal{L}_{HT}$ are trained with extra unlabelled data for another 12 epochs. The initial learning rate is 1×10^{-2}, and is decreased by a factor of $(1 - t/T)^{0.9}$, where t is the current training epoch and T is the total number of epochs, as in [4]. The batch size is set to be 16. For our $\mathcal{L}_{\text{total}}$, we set the weights $\alpha = 0.1$ and $\beta = 0.01$ to ensure that all loss terms have similar magnitudes. Following [4], we multiply the $\mathcal{L}_{\text{seg}}$ for the background class by 0.4 to counter the large number of background pixels. For s4GAN [32], we directly use the official implementation [2].

Results analysis. To evaluate the effectiveness of our $\mathcal{L}_{HT}$ in utilizing unlabelled data, we randomly split the CULane training data into $\{100/0, 50/50, 10/90, 5/95, 1/99\}$ sets, where the first digit indicates the proportion of labelled data, while the second one is the proportion of unlabelled data. The TuSimple dataset is split into $\{100/0, 50/50, 10/90\}$ sets, as it contains only 3,626 images. We use the same splits for all models. We report accuracy on TuSimple and F_1-measure on the CULane dataset.

Table 1: **Performance on TuSimple and CULane datasets with various amounts of labelled and unlabelled data.** The first column indicates the proportion of labelled data for training. The remaining data is treated as unlabelled for semi-supervised learning. ERFNet-HT-$\mathcal{L}_{HT}$ and ERFNet-HT-*pseudo*+$\mathcal{L}_{HT}$ show performance improvements on both datasets. When the number of labelled samples decreases, the advantage of ERFNet-HT-$\mathcal{L}_{HT}$ is more pronounced.

Labels	s4GAN [32]	ERFNet models				
		Baseline [5]	HT [9]	*pseudo*	HT-$\mathcal{L}_{HT}$	HT-*pseudo* + $\mathcal{L}_{HT}$
		Accuracy (%) on the TuSimple dataset				
100%	-	**93.71**	**93.71**	-	-	-
50%	88.82	92.97	93.47	93.37	93.63	**93.70**
10%	86.25	82.97	77.71	92.12	92.98	**93.05**
		F_1 scores on the CULane dataset				
100%	-	69.86	**70.52**	-	-	-
50%	-	69.39	68.59	69.68	**70.75**	70.41
10%	-	60.99	61.46	65.56	64.04	**66.10**
5%	-	56.61	57.78	61.99	62.32	**63.67**
1%	-	32.99	32.48	51.38	**55.10**	52.80

Table 1 compares all models on various training sets. ERFNet-HT-*pseudo*+$\mathcal{L}_{HT}$ achieves the best performance on both 50% and 10% subsets of TuSimple dataset. The improvement over the supervised baseline is more than 15% on the 10% subset. All semi-supervised ERFNet models improve accuracy, indicating the potential of exploiting massive unlabelled data. Pseudo-labeling allows learning from high confidence predictions explicitly, while $\mathcal{L}_{HT}$ optimizes line feature representations in Hough space in an implicit way. However, s4GAN [32] shows inferior performance to other models, due to the fact that s4GAN is not specifically optimized for lane detection, where image content differs substantially from its origin usage. In general, semi-supervised models perform similar on the TuSimple dataset as it only includes the highway scenario. On the CULane dataset, ERFNet-HT-*pseudo*+$\mathcal{L}_{HT}$ consistently outperforms ERFNet-*pseudo*, validating the usefulness of the Hough priors ($\mathcal{L}_{HT}$) in exploiting lane representations in the semi-supervised setting. The s4GAN is lacking since we are unable to produce reliable prediction on this dataset.

We observe that ERFNet-HT-$\mathcal{L}_{HT}$ improves over all other

[2] https://github.com/sud0301/semisup-semseg

[3] For cross-road, we show only the number of false-positives, as in [10].

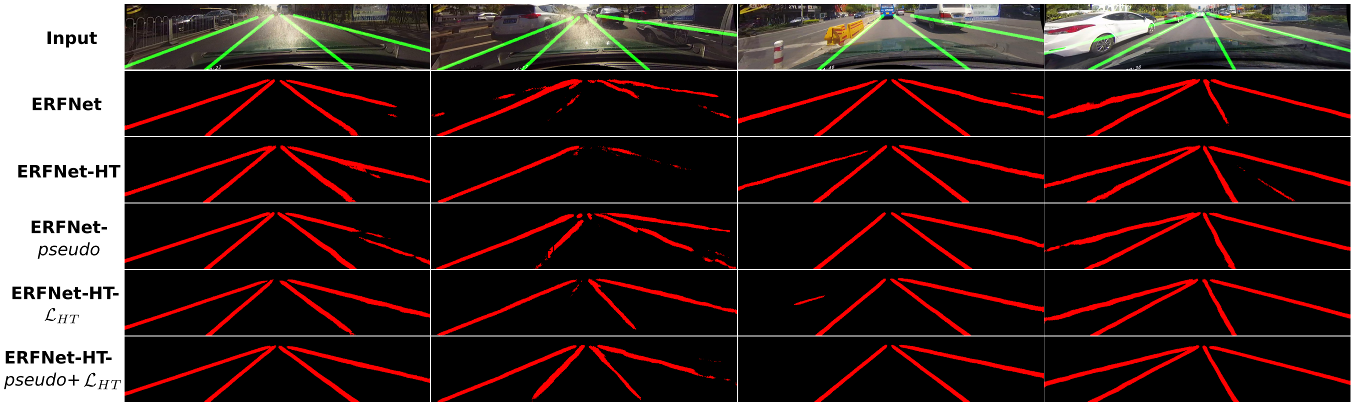

Fig. 2: **Visualizations of predicted lanes on the CULane dataset.** Only 10% annotated data is used for training. ERFNet-HT-*pseudo*+$\mathcal{L}_{HT}$ performs better on challenging samples and better localizes lane boundaries. The inference speed of the ERFNet-HT is around 13 frames per second on a NVIDIA GTX1080Ti GPU.

Table 2: **F_1 scores for different scenarios, with 1% labelled data.** ERFNet-HT-$\mathcal{L}_{HT}$ outperforms other models in most scenarios, indicating that the $\mathcal{L}_{HT}$ loss exploits useful geometric knowledge of lanes when adding unlabelled samples.

ERFNet models	Baseline [5]	HT [9]	*pseudo*	HT-$\mathcal{L}_{HT}$	HT-*pseudo* +$\mathcal{L}_{HT}$
Normal	49.24	51.25	69.72	**75.06**	71.83
Crowded	31.74	31.49	49.53	**52.52**	50.97
Night	22.36	21.67	45.27	**50.77**	45.42
No line	18.78	17.54	28.05	**32.02**	30.12
Shadow	24.71	17.69	36.63	**38.50**	35.97
Arrow	39.39	38.22	57.28	**63.33**	59.18
Dazzle	26.25	23.76	**40.29**	40.28	39.42
Curve	33.62	34.53	46.56	**50.52**	46.42
Cross [3]	6949	8711	**3355**	5292	3676
Avg F_1	33.00	32.48	51.38	**55.10**	52.80

models on the 1% subset by a large margin. On the 1% subset, there is not sufficient labelled data (less than $1K$ training images), and therefore the "ground truth" produced by pseudo-labelling in ERFNet-*pseudo* is noisy and imperfect. In this case, learning from pseudo-labelled data explicitly can be harmful, while the $\mathcal{L}_{HT}$ avoids this problem by exploiting useful prior geometric knowledge about lines, in Hough space. In comparison, on the 50% subset, the differences among all models are marginal, when ample training data is available. The experiment demonstrates the potential of our $\mathcal{L}_{HT}$ loss for data-efficient learning in Hough space in a semi-supervised setting.

We compare the performance of all ERFNet models in various driving scenarios in Table 2. ERFNet-HT-$\mathcal{L}_{HT}$ shows considerable improvement over other models in most scenarios in Table 2, and the advantage accentuates (up to 5%), where the amount of labelled data is decreased to 1% only. The superiority of ERFNet-HT-$\mathcal{L}_{HT}$ demonstrates the capability of $\mathcal{L}_{HT}$ to exploit geometric lanes information from unlabelled data. We also notice that the "No line", "Shadow" and "Dazzle" scenarios are more challenging for all methods, compared with the other scenarios.

We visualize line predictions from different models in Fig. 2. Our ERFNet-HT-*pseudo*+$\mathcal{L}_{HT}$ better localizes lanes, especially when a lane extends away from the image boundary, as in the first two examples. As shown in the second example, due to occlusion, ERFNet and ERFNet-HT miss the two middle lanes, while ERFNet-HT-$\mathcal{L}_{HT}$ only predicts one. In the third example there is an annotation inconsistency, where the opposite lane at the image border is not annotated. Overall, ERFNet-HT-*pseudo*+$\mathcal{L}_{HT}$ produces sharper and more precise predictions, in both simple and challenging scenarios.

5. LIMITATIONS AND CONCLUSIONS

We propose semi-supervised lane detection by exploiting global line priors in Hough space through the use of an additional loss. We can incorporate unlabelled data during training thus overcoming the need for expensive and error-prone annotations. Currently our method assumes a single lane in each channel, and therefore we can optimize for the global maximum in Hough space. This assumption may not always hold and an extension to multiple local maxima is future research. However, our proposed Hough loss adds valuable prior geometric knowledge about lanes when annotations are too scarce even for pseudo-labelling based methods. We experimentally demonstrate the added value of our proposed loss on TuSimple and CULane datasets for limited annotated data.

6. REFERENCES

[1] Bei He, Rui Ai, Yang Yan, and Xianpeng Lang, "Accurate and robust lane detection based on dual-view convolutional neutral network," in *Intelligent Vehicles Symposium*, 2016, pp. 1041–1046.

[2] Jochen Pohl, Wolfgang Birk, and Lena Westervall, "A driver-distraction-based lane-keeping assistance system," *Journal of Systems and Control Engineering*, vol. 221, no. 4, pp. 541–552, 2007.

[3] Yuenan Hou, Zheng Ma, Chunxiao Liu, and Chen Change Loy, "Learning lightweight lane detection cnns by self attention distillation," in *ICCV*, 2019, pp. 1013–1021.

[4] Yuenan Hou, Zheng Ma, Chunxiao Liu, Tak-Wai Hui, and Chen Change Loy, "Inter-region affinity distillation for road marking segmentation," in *CVPR*, 2020, pp. 12486–12495.

[5] Eduardo Romera, José M Alvarez, Luis M Bergasa, and Roberto Arroyo, "Erfnet: Efficient residual factorized convnet for real-time semantic segmentation," *Transactions on Intelligent Transportation Systems*, vol. 19, no. 1, pp. 263–272, 2017.

[6] Zequn Qin, Huanyu Wang, and Xi Li, "Ultra fast structure-aware deep lane detection," in *The European Conference on Computer Vision (ECCV)*, 2020.

[7] Miriam Huijser and Jan C van Gemert, "Active decision boundary annotation with deep generative models," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 5286–5295.

[8] Richard O Duda and Peter E Hart, "Use of the hough transformation to detect lines and curves in pictures," *Communications of the ACM*, vol. 15, no. 1, pp. 11–15, 1972.

[9] Yancong Lin, Silvia L Pintea, and Jan C van Gemert, "Deep hough-transform line priors," *EECV*, 2020.

[10] Xingang Pan, Jianping Shi, Ping Luo, Xiaogang Wang, and Xiaoou Tang, "Spatial as deep: Spatial cnn for traffic scene understanding," *CoRR*, 2017.

[11] "Tusimple lane detection challenge," `https://github.com/TuSimple/tusimple-benchmark`.

[12] ZuWhan Kim, "Robust lane detection and tracking in challenging scenarios," *Transactions on Intelligent Transportation Systems*, vol. 9, no. 1, pp. 16–26, 2008.

[13] Hendrik Deusch, Jürgen Wiest, Stephan Reuter, Magdalena Szczot, Marcus Konrad, and Klaus Dietmayer, "A random finite set approach to multiple lane detection," in *International Conference on Intelligent Transportation Systems*, 2012, pp. 270–275.

[14] Hunjae Yoo, Ukil Yang, and Kwanghoon Sohn, "Gradient-enhancing conversion for illumination-robust lane detection," *Transactions on Intelligent Transportation Systems*, vol. 14, no. 3, pp. 1083–1094, 2013.

[15] Tao Wu and Ananth Ranganathan, "A practical system for road marking detection and recognition," in *Intelligent Vehicles Symposium*, 2012, pp. 25–30.

[16] Byambaa Dorj and Deok Jin Lee, "A precise lane detection algorithm based on top view image transformation and least-square approaches," *Journal of Sensors*, vol. 2016, 2016.

[17] Kamarul Ghazali, Rui Xiao, and Jie Ma, "Road lane detection using h-maxima and improved hough transform," in *International Conference on Computational Intelligence, Modelling and Simulation*, 2012, pp. 205–208.

[18] Yue Wang, Dinggang Shen, and Eam Khwang Teoh, "Lane detection using spline model," *Pattern Recognition Letters*, vol. 21, no. 8, pp. 677–689, 2000.

[19] Bin Yu and Anil K Jain, "Lane boundary detection using a multiresolution hough transform," in *ICIP*, 1997, vol. 2, pp. 748–751.

[20] Seokju Lee, Junsik Kim, Jae Shin Yoon, Seunghak Shin, Oleksandr Bailo, Namil Kim, Tae-Hee Lee, Hyun Seok Hong, Seung-Hoon Han, and In So Kweon, "Vpgnet: Vanishing point guided network for lane and road marking detection and recognition," in *ICCV*, 2017, pp. 1947–1955.

[21] Jun Li, Xue Mei, Danil Prokhorov, and Dacheng Tao, "Deep neural network for structural prediction and lane detection in traffic scene," *Transactions on neural networks and learning systems*, vol. 28, no. 3, pp. 690–703, 2016.

[22] Davy Neven, Bert De Brabandere, Stamatios Georgoulis, Marc Proesmans, and Luc Van Gool, "Towards end-to-end lane detection: an instance segmentation approach," in *Intelligent vehicles symposium*, 2018, pp. 286–291.

[23] Jigang Tang, Songbin Li, and Peng Liu, "A review of lane detection methods based on deep learning," *Pattern Recognition*, p. 107623, 2020.

[24] Jie Zhang, Yi Xu, Bingbing Ni, and Zhenyu Duan, "Geometric constrained joint lane segmentation and lane boundary detection," in *ECCV*, 2018, pp. 486–502.

[25] Jesper E Van Engelen and Holger H Hoos, "A survey on semi-supervised learning," *Machine Learning*, vol. 109, no. 2, pp. 373–440, 2020.

[26] Avrim Blum, John Lafferty, Mugizi Robert Rwebangira, and Rajashekar Reddy, "Semi-supervised learning using randomized mincuts," in *ICML*, 2004, p. 13.

[27] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton, "A simple framework for contrastive learning of visual representations," *ICML*, 2020.

[28] Durk P Kingma, Shakir Mohamed, Danilo Jimenez Rezende, and Max Welling, "Semi-supervised learning with deep generative models," *NeurIPS*, vol. 27, pp. 3581–3589, 2014.

[29] Avrim Blum and Shuchi Chawla, "Learning from labeled and unlabeled data using graph mincuts," *ICML*, 2001.

[30] Tony Jebara, Jun Wang, and Shih-Fu Chang, "Graph construction and b-matching for semi-supervised learning," in *ICML*, 2009, pp. 441–448.

[31] Xiaojin Zhu, Zoubin Ghahramani, and John D Lafferty, "Semi-supervised learning using gaussian fields and harmonic functions," in *ICML*, 2003, pp. 912–919.

[32] Sudhanshu Mittal, Maxim Tatarchenko, and Thomas Brox, "Semi-supervised semantic segmentation with high-and low-level consistency," *IEEE transactions on pattern analysis and machine intelligence*, 2019.

[33] Léon Bottou, "Large-scale machine learning with stochastic gradient descent," in *Proceedings of COMPSTAT'2010*, pp. 177–186. 2010.