DEEP CNN WITH COLOR LINES MODEL FOR UNMARKED ROAD SEGMENTATION

Shashank Yadav[1] *Suvam Patra*[1] *Chetan Arora*[2] *Subhashis Banerjee*[1]

[1]Indian Institute of Technology Delhi
New Delhi 110016

[2]Indraprastha Institute of Information Technology Delhi
New Delhi 110020

ABSTRACT

Road detection from a monocular camera is an important perception module in any advanced driver assistance or autonomous driving system. Traditional techniques [1, 2, 3, 4, 5, 6] work reasonably well for this problem, when the roads are well maintained and the boundaries are clearly marked. However, in many developing countries or even for the rural areas in the developed countries, the assumption does not hold which leads to failure of such techniques. In this paper we propose a novel technique based on the combination of deep convolutional neural networks (CNNs), along with color lines model [7] based prior in a conditional random field (CRF) framework. While the CNN learns the road texture, the color lines model allows to adapt to varying illumination conditions. We show that our technique outperforms the state of the art segmentation techniques on the unmarked road segmentation problem. Though, not a focus of this paper, we show that even on the standard benchmark datasets like KITTI [8] and CamVid [9], where the road boundaries are well marked, the proposed technique performs competitively to the contemporary techniques.

Index Terms— Road segmentation, road detection, graph cuts, CNN, CRF

1. INTRODUCTION

With the rapid progress in machine learning techniques, researchers are now looking towards autonomous navigation of a vehicle in all kinds of road and environmental conditions. While, many of the problems in autonomous driving look easy in sanitised city or highway environments of developed countries, the same problems become extremely hard in cluttered and chaotic scenarios particularly in developing countries. Road segmentation is one such problem, where many techniques work successfully when the roads are well maintained and boundaries are clearly marked. In many parts of the world, such as rural and undeveloped areas, the assumption is not valid, which leads to failure of these techniques. The focus of this paper is on road segmentation for such difficult cases where the roads are not maintained (potholes or different textured patches) and the road boundaries are not marked.

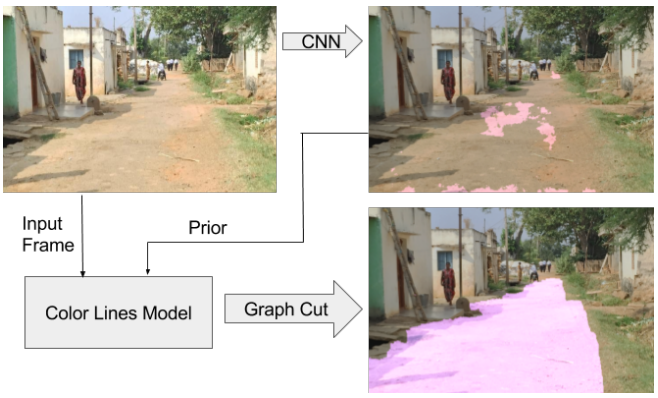

Fig. 1: The focus of this paper is on segmentation of unmaintained roads without markings. We propose color lines model [7], in conjunction with a CNN model in a CRF based framework. The color lines model helps CNN adapt better to varying illumination and road conditions. The proposed model outperforms state of the art on the benchmark as well as the dataset captured by us on the targeted road conditions.

Road segmentation is a challenging problem given the observed variability with different times of the day, changing lighting conditions, weather, and variable road conditions etc. Presence of potholes, and even the filled potholes, which lead to variability in the road texture further complicate the problem. Road markings often act as a secondary cue, but may not be available in all situations. In the last decade, computer vision researchers have proposed broadly two styles of techniques to detect the roads from images as well as videos. Feature based techniques like [1, 2] use lane markings and road boundaries, whereas model based techniques [3] use image or structure based priors and geometrical assumptions for detecting roads. Learning methods like [4, 5, 6] try to learn features based on image information like image histograms.

Despite tremendous progress in deep learning based techniques for semantic image segmentation, detecting road boundary remains a challenging problem due to the wide variability of roads and marking conditions in different countries. It has been observed that the network trained on a particular kind of dataset, often do not perform well in different road conditions [9, 10].

We note that modern machine learning based techniques do not exploit physics based color model which can help pre-

dicting wide variations in appearance of road under different illumination conditions. In a different setting, Omer and Werman [7] has showed that the effect of an illuminant on the color of an object can be approximately modelled as additive. This implies that the color of the object approximately follows a linear curve under different illumination conditions. We use their model, hereinafter referred to as color lines model, to model the visual appearance of a road. We use the model as prior in conjunction with likelihood from a standard CNN model (SegNet) within a CRF framework. The novelty of the proposed model is that it allows the deep network to adapt to roads with varying texture and appearance. We validate our model on standard benchmarks as well as on datasets of unmarked roads captured by us.

2. RELATED WORK

Alvarez et al. [11] assume that roads appear in the lower part of the images to model road appearance in the whole image. Low-level cues such as color [12, 13, 14, 15], or a combination of color and texture [16] have also been used for road segmentation. However, the assumptions fail often. For example, color based models can not model large intra-class variability usually observed among pixels of the road. Also, the bottom part of the image often include back part of a vehicle in the front. Shadows on the road pose additional challenge in identifying the road regions. Alvarez et al. [12] propose a training-based illumination-invariant color space to handle shadows. Sturgess et al. [17] use semantic labeling of pixels, where a pixel classifier is learnt from motion based features and smoothness is encouraged by modeling the image with a Conditional Random Field (CRF). In [18], a model of the road is learnt from training data which is then refined from unsupervised images. Other notable techniques use temporal consistency and use detection results in previous images as training data for the current frame [13, 15]. Zuo et al. [19] has proposed a combination of model and feature based ideas to adapt to various road conditions. Deep neural networks with fully convolutional model has been proposed by Long et al. [20] for the problem. Researchers have observed that the deep networks trained on a particular dataset fail to adapt to a different domain restricting the wider applicability of these models [9, 10]. Alvarez et al. suggest non-parametric models to circumvent such problems [21]. Fritsch et al. [22] have recently proposed a dataset to compare various road detection techniques. Other larger and more general datasets such as CamVid [9], KITTI[8] and Cityscapes [23] have road as one of the labels.

3. PROPOSED METHODOLOGY

In the absence of any explicit road markings, color and texture of the road represent two dominant cues for the road segmentation. Deep convolutional neural networks have been shown to be adept at learning various texture patterns for object detection and image segmentation problems. However, the models are typically trained to ignore the color to achieve invariance against illumination. The color of the road is an important cue for the segmentation but is useful only if we can learn the appearance model under various illumination and weather conditions. We use color lines model [7] to predict road appearance under varying illumination.

3.1. Learning Road Texture

In this paper we use pre-trained SegNet [25] model which segments the image into 12 different classes including a label for the road. We club all the other labels, except for the road, into a background class. The pre-trained model, as is, does not perform well on the unmarked road segmentation because of its inability to exploit road color cues. However, we make up for this weakness by inserting color lines based appearance model. It may be noted that we use SegNet because of its ability to learn road texture efficiently. Other similar techniques (based on deep learning or without) could have been equivalently used without changing the proposed formulation.

3.2. Learning Road Color Model

We learn frame specific color lines model for road appearance in each video frame. We use SegNet output to bootstrap and initialize the road pixels for learning the appearance model. The entire RGB space is quantized into bins using discrete concentric spheres centered at origin. The radius increases in multiples of some constant integer. The volume between two consecutive spheres is defined as a bin. Bin k contains the set of RGB values of image points lying in the volume between k^{th} and the $k + 1^{th}$ concentric spheres.

We use the segmentation from the SegNet to create two color line models, one for the road and the other for the background. Here we have made a simplifying assumption of single color line for the background as well. Each of the lines is modelled as the set of representative RGB points falling on this line. We compute a representative point, each for road and background, corresponding to every bin. The representative point for road (background) is computed as the mean RGB color of the road (background) pixels in that bin. We also compute the variance of the RGB colors which we use later to compute the probability of a test color originating from the color line.

For any pixel with RGB value x, we first compute the bin it lies in. We then compute the score of it lying on the road (or background) as the probability of it originating from a Gaussian distribution with mean as the representative point of road (background) in that bin and variance computed as described above. In case there is no representative point in that bin we use the color lines model to extrapolate the line to that bin. This helps us in attaining color constancy under varying illumination

Method	Benchmark	MaxF	AP	PRE	REC	FPR	FNR
Segnet	UM_ROAD	82.17 %	76.46 %	84.03 %	80.40 %	6.97 %	19.60 %
	UMM_ROAD	88.59 %	83.54 %	88.35 %	88.84 %	12.88 %	11.16 %
	UU_ROAD	77.23 %	69.23 %	82.29 %	72.76 %	5.10 %	27.24 %
	URBAN_ROAD	84.04 %	78.76 %	85.50 %	82.63 %	7.72 %	17.37 %
Ours	UM_ROAD	83.50 %	72.28 %	76.37 %	92.09 %	12.98 %	7.91 %
	UMM_ROAD	90.30 %	83.33 %	86.43 %	94.53 %	16.32 %	5.47 %
	UU_ROAD	79.89 %	67.48 %	77.01 %	82.99 %	8.07 %	17.01 %
	URBAN_ROAD	85.73 %	76.89 %	81.02 %	91.01 %	11.74 %	8.99 %

Table 1: Segmentation Results on the KITTI [8] dataset. MaxF: Maximum F1-measure, AP: Average precision as used in PASCAL VOC [24] challenges, PRE: Precision, REC: Recall, FPR: False Positive Rate, FNR: False Negative Rate (the four latter measures are evaluated at the working point MaxF). This is as mentioned in the KITTI Benchamrk Suite [8]

Method	PRE	REC	F val
Segnet	93.07 %	94.86 %	93.95 %
Ours	93.31 %	94.99 %	94.14 %

Table 2: Segmentation Results on Camvid [9] Testing dataset

3.3. CRF Formulation

We use the scores for the road and the background to create a conditional random field based graphical model [26, 27]. For the CRF formulation, the problem is modeled as a 2 label, corresponding to road and background, labeling problem. The CRF formulation creates a graph corresponding to the problem with nodes corresponding to each pixel, along with two special nodes s and t referred to as source and sink respectively [26, 27]. There are edges between pixel nodes (called inter-pixel edges) and between source/sink and pixel nodes (called terminal edges). The terminal edge capacities correspond to the data term of the energy minimization problem as in [26, 27] and the inter-pixel edge capacities correspond to the smoothing cost.

We set the capacity of the terminal edge from each node to the source as the score for that pixel originating from road computed (as described in Section 3.2). Similarly, we set the capacity of the terminal edge from each node to the sink as the score for that pixel originating from the background. We also multiply these capacities by the confidence of the pixel being on road that we get in the form of the probability of a label by SegNet. We compute the capacity of inter-pixel edge between pixel x and y as: $V_{xy} = P_r(x|i) * P_r(y|j) + P_b(x|i) * P_b(y|j)$. Here we assume x and y belongs to bin i and j respectively. $P_r(x|i)$ and $P_b(x|i)$ represent the score of pixel x for originating from road and background respectively. We infer the best labeling having the maximum posterior probability by finding the minimum cut in this specially created graph.

4. RESULTS

We evaluate our method on images from different datasets. For urban roads we use images from the KITTI [8] and CamVid [9] datasets. For rural roads, we use the dataset captured by us on the Indian road conditions. We present quantitative as well as qualitative results. We use a SegNet network which was trained on an ensemble of 3433 images [25]. The choice of SegNet was made owing to its state of art performance and availability of the pretrained model.

Table 1 shows a comparison of our method with SegNet on the KITTI [8] dataset obtained via submission on their online interface. As seen from the results, we get improvement of several points over the SegNet model. A few qualitative results showing significant visual improvement are shown in Fig 2.

We also evaluate our method on CamVid [9] dataset, a publicly available dataset for semantic segmentation. Since we are only interested in the detection of roads, we use precision and recall for the road as a metric. The results are shown in Table 2. As the performance of SegNet is already quite good and saturated on this dataset, we only achieve small improvement on this dataset.

For rural road conditions, we present some qualitative results on the images captured by us. Fig. 3 shows some results. Fig. 4 shows results on some unmarked road images downloaded from the internet. We observe significant improvement in such images.

5. CONCLUSION

In this paper we have proposed a novel algorithm for road detection under varying illumination using a monocular camera. The proposed alorithm uses SegNet for modelling road texture and color lines model for learning appearance of unmarked and un-maintained roads commonly found in developing countries as well as rural areas in the developed world. Experiments show superiority of the proposed approach over state of the art.

Acknowledgement: This work was supported by a research grant from Continental Automotive Components (India) Pvt. Ltd. Chetan Arora is supported by Visvesaraya Young Faculty Fellowship and Infosys Center for AI.

Fig. 2: Each column depicts results on a different image taken from the KITTI dataset [8]. The first row for each column denotes the output after running Segnet [25] and second row shows the results by our method.

Fig. 3: Each column depicts results on a different image taken by us on Indian roads. The first row for each column denotes the output after running Segnet [25] and second row shows the results by our method.

Fig. 4: Each column depicts results on a different image taken from the internet. The first row for each column denotes the output after running Segnet [25] and second row shows the results by our method.

6. REFERENCES

[1] L.W. Tsai, J.W. Hsieh, C.H. Chuang, and K.C. Fan, ": Lane detection using directional random walks.," in *IEEE Intelligent Vehicles Symposium*, 2008.

[2] Q. Li, N. Zheng, and H. Cheng, "Spring robot: A prototype autonomous vehicle and its algorithms for lane detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 5, pp. 300–308, 2004.

[3] Y. Wang, E. Teoh, and D. Shen, "Lane detection and tracking using b-snake," *Image Vision Computing*, vol. 22, pp. 269–280, 2004.

[4] J. Wang, Z. Ji, and Y. Su, "Unstructured road detection using hybrid features," in *International Conference on Machine Learning and Cybernetics*, 2009.

[5] S. Yun, Z. Guo-ying, and Y. Yong, "A road detection algorithm by boosting using feature combination," in *Intelligent Vehicles Symposium*, 2007.

[6] M. Foedisch and A. Takeuchi, "Adaptive road detection through continuous environment learning," in *Applied Imagery Pattern Recognition Workshop*, 2004.

[7] I. Omer and M. Werman, "Color lines: image specific color representation," in *CVPR*, 2004.

[8] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *CVPR*, 2012.

[9] G. J. Brostow, J. Shotton, J. Fauqueur, and R. Cipolla, "Segmentation and recognition using structure from motion point clouds," in *ECCV*, 2008.

[10] C. Guo, S. Mita, and D. McAllester, "Robust road detection and tracking in challenging scenarios based on markov random fields with unsupervised learning," *IEEE Trans. on ITS*, vol. 13, pp. 1338–1354, 2012.

[11] J. M. Alvarez, M. Salzmann, and N. Barnes, "Learning appearance models for road detection," in *IV*, 2013.

[12] J. M. Alvarez and A. Lopez, "Road detection based on illuminant invariance," *IEEE Trans. on ITS*, vol. 12, pp. 184–193, 2011.

[13] M. Sotelo, F. Rodriguez, and L. Magdalena, "Virtuous: vision-based road transp. for unmanned operation on urbanlike scenarios.," *IEEE Trans. on ITS*, vol. 5, 2004.

[14] B. Kim, J. Son, and K Sohn, "Illumination invariant road detection based on learning method," in *ITSC*, 2011.

[15] C. Tan, T. Hong, T Chang, and M. Shneier, "Color model based real-time learning for road following," in *ITSC*, 2006.

[16] P. Lombardi, M. Zanin, and S. Messelodi, "Switching models for vision-based onboard road detection," in *ITSC*, 2005.

[17] P. Sturgess, K. Alahari, L. Ladicky, and P. H. S. Torr, "Combining appearance and structure from motion features for road scene understanding," in *BMVC*, 2009.

[18] J. M. Alvarez, T. Gevers, Y. LeCun, and A. M. Lopez, "Road scene segmentation from a single image," in *ECCV*, 2012.

[19] W. H. Zuo and T. Z. Yao, "Road model prediction based unstructured road detection," *Journal of Zhejiang university science C(comput & Electron)*, vol. 11, pp. 822–834, 2014.

[20] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *CVPR*, 2015.

[21] J. M. Alvarez, M. Salzmann, and N Barnes, "Data-driven road detection," in *WACV*, 2014.

[22] J. Fritsch, T. Kuehnl, and A Geiger, "A new performance measure and evaluation benchmark for road detection algorithms," in *ITSC*, 2013.

[23] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," in *CVPR*, 2016.

[24] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *International Journal of Computer Vision*, vol. 88, pp. 303–338, 2010.

[25] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *CoRR*, vol. abs/1511.00561, 2015.

[26] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *IEEE Transactions on PAMI*, vol. 26, pp. 1124–1137, 2004.

[27] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Transactions on PAMI*, vol. 23, pp. 1222–1239, 2001.