

조건부 무작위장을 사용한 세그넷의 시맨틱 영상 분할 성능 향상

김 윤 형*, 은 현 준*, 정 찬 호°, 김 창 익*

Improving the Semantic Image Segmentation Performance of SegNet Using Conditional Random Fields

Yoonhyung Kim*, Hyunjun Eun*,
Chanho Jung°, Changick Kim*

요 약

본 논문에서는 CRF를 이용한 SegNet의 시맨틱 영상 분할 성능 향상 방법을 제안한다. 심층 합성곱 신경망 구조로 이루어진 SegNet은 특징 맵에 대한 반복적인 스케일링을 수반하기 때문에 영상 분할 결과의 선명도가 낮다. CRF 알고리즘은 원본 영상의 구조 및 질감 패턴에 따라 SegNet의 분할 결과를 보완함으로써 선명도 및 정확도를 개선한다. CamVid 데이터 셋에 대해 성능 비교 평가를 수행하였으며, 제안하는 방법이 더 높은 성능을 가짐을 보였다. 본 연구는 시맨틱 영상 분할 방법 개선에 실질적인 도움이 될 것으로 판단된다.

Key Words : Semantic segmentation, Conditional random field, Deep convolutional neural network

ABSTRACT

In this paper, we propose a method to enhance the semantic segmentation performance of SegNet by using Conditional Random Fields (CRFs). As a deep convolutional neural network architecture, SegNet

involves multiple scaling operations of feature maps, degrading the output sharpness. The SegNet results can be improved by using CRFs which interpolate output labels based on structures and textures of original images. We use the CamVid dataset to conduct performance evaluations. The proposed method achieves the higher performance than the original SegNet approach. We believe that this study guides for improving semantic segmentation.

1. 서 론

시맨틱 영상 분할 (Semantic segmentation)은 입력 영상의 모든 픽셀점에 대해 레이블을 할당하는 기술로써 자동차 자율주행 기술, 관심 객체 검출^[1], 무인 감시 시스템 등 여러 컴퓨터 비전 분야에서 활용된다. 최근에는 심층 합성곱 인공 신경망 (Convolutional neural network)과 같은 학습 기반 분류기의 발전에 힘입어 시맨틱 영상 분할 문제를 딥러닝 기반 접근법으로 해결하려는 시도가 있었다. 다양한 딥러닝 기반 시맨틱 영상 분할 방법들 중 하나인 SegNet^[2]은 입력 영상에 대해 여러 층에 걸쳐 높은 수준의 특징 맵으로 인코딩하고, 이를 다시 낮은 수준의 특징 맵으로 디코딩한 후 Softmax 함수를 적용함으로써 모든 픽셀점에 대한 레이블을 결정한다. 인코딩 시에는 다운 샘플링을, 디코딩 시에는 업 샘플링을 수반하는데, 이 과정에서 여러 번에 걸친 스케일링으로 인해 그림 1(c)와 같이 원본 영상의 선명도가 손실되는 문제가 발생한다.

본 논문에서는 CRF (Conditional Random Field)^[3]를 이용한 SegNet의 시맨틱 영상 분할 성능 향상 방법을 제안한다. CRF는 입력 영상의 구조적, 질감적 패턴 정보를 근거로 시맨틱 영상 분할결과 값을 보완한다. 이를 통해 SegNet 결과 영상의 선명도와 시맨틱 영상 분할의 정확도를 그림 1(d)와 같이 향상시킬 수 있다. 기존 SegNet 방법과 SegNet 결과에 CRF를 적용하는 방법의 성능 비교 평가를 위하여 CamVid 데이터 셋^[4]을 이용하였다. 그 결과 제안하는 CRF 적용 방법이 픽셀 정확도, IoU (Intersection over Union), BF (Boundary F1) 수치에 대해 전반적으로

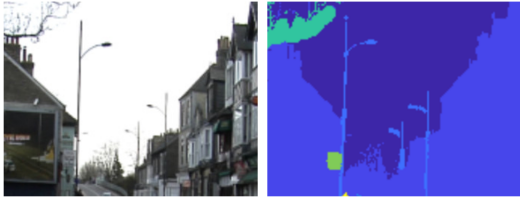
* 이 논문은 2016년도 한밭대학교 신입교수연구비의 지원을 받았다.

• First Author : School of Electrical Engineering, Korea Advanced Institute of Science and Technology, yhkim1127@kaist.ac.kr, 학생회원

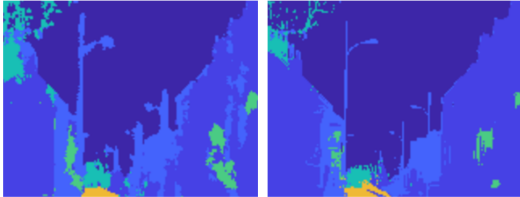
° Corresponding Author : Department of Electrical Engineering, Hanbat National University, peterjung@hanbat.ac.kr, 정회원

* School of Electrical Engineering, Korea Advanced Institute of Science and Technology, hj.eun@kaist.ac.kr, 학생회원, changick@kaist.ac.kr

논문번호 : KICS2018-02-033, Received February 2, 2018; Revised March 1, 2018; Accepted March 2, 2018



(a) Input image (b) Ground truth label



(c) Result with SegNet (d) Result with SegNet&CRF

그림 1. CRF 적용 여부에 따른 SegNet의 시맨틱 영상 분할 결과 비교

Fig. 1. Comparison of semantic segmentation results obtained by SegNet only and SegNet followed by CRFs.

더 높은 성능을 나타내었다.

II. 제안하는 방법

SegNet은 너비 W , 높이 H , 채널 3의 RGB 영상 I_O 를 입력받아 너비 W , 높이 H , 채널 L 의 결과 영상 I_R 을 출력한다. 여기서 L 은 사전에 정의된 레이블의 개수를 나타낸다. I_R 의 각 채널 영상은 해당 레이블에 대한 유사도를 나타내는 히트맵 (Heat map)이다. 기존의 SegNet은 I_R 의 공간적 위치에서 가장 큰 유사도를 가지는 채널의 레이블을 할당하는 방식으로 레이

블 맵 I_S 를 생성한다. 본 논문에서 제안하는 CRF 적용 방식은 I_O 와 I_R 을 입력받아 높은 선명도의 레이블 맵 I_S^{CRF} 를 생성한다.

구하고자 하는 I_S^{CRF} 의 벡터 형태를 \mathbf{x} 라 할 때, CRF 최적화 목적 함수는 다음과 같이 정의된다.

$$\mathbf{x}^* = \operatorname{argmax}_{\mathbf{x}} E(\mathbf{x}),$$

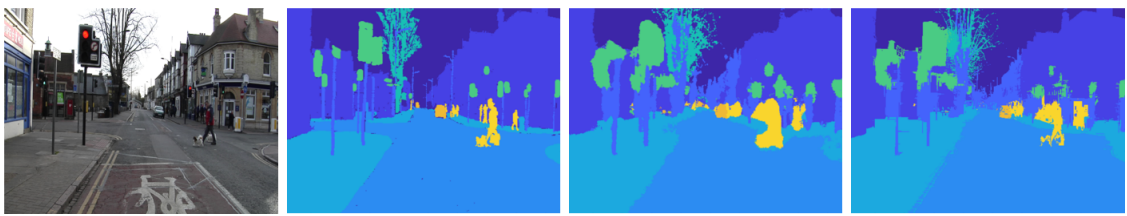
$$E(\mathbf{x}) = \sum_i \theta_i(x_i) + \sum_{i,j} \theta_{i,j}(x_i, x_j). \quad (1)$$

식 (1)에서 θ_i 는 i 번째 픽셀에서의 단항 퍼텐셜 (Unary potential)이다. 이는 SegNet의 결과 영상 I_R 에 의해 각 픽셀 위치마다 독립적으로 결정되며, CRF에 의해 결정될 레이블의 사전 단서로써 활용된다. $\theta_{i,j}$ 는 (i, j) 번째 픽셀 쌍에 대한 켈레 퍼텐셜 (Pairwise potential)이며 다음과 같이 정의된다.

$$\theta_{i,j}(x_i, x_j) = \exp\left(-\frac{|p_i - p_j|^2}{2\theta_\alpha^2} - \frac{|I_{O_i} - I_{O_j}|^2}{2\theta_\beta^2}\right)$$

$$+ \exp\left(-\frac{|p_i - p_j|^2}{2\theta_\gamma^2}\right). \quad (2)$$

식 (2)에서 p_i 와 I_{O_i} 는 i 번째 픽셀의 위치와 컬러값을 각각 나타낸다. $\theta_\alpha, \theta_\beta, \theta_\gamma$ 는 균형 상수이며, 제안하는 방법에서는 $\theta_\alpha = 5, \theta_\beta = 3, \theta_\gamma = 0.5$ 의 값을 적용하였다. 식 (2)의 켈레 퍼텐셜에 의해, 각 픽셀에서의 레이블은 거리가 가깝고 원본 영상의 컬러값이 비



(a) Input image (b) Ground truth label (c) Result with SegNet (d) Result with SegNet&CRF

그림 2. CamVid 데이터 셋의 시험 영상에 대한 시맨틱 영상 분할 결과 비교

Fig. 2. Comparison of semantic segmentation results for several test images of the CamVid dataset

표 1. 시맨틱 영상 분할 결과의 정량적 비교
Table 1. Quantitative comparison on semantic segmentation results.

Metric	Method	Label											Mean
		Sky	Buil ding	Pole	Walk way	Road	Tree	Traffic light	Fence	Vehi cle	Pede strian	Bike	
Accu.	SegNet ¹	0.925	0.771	0.800	0.935	0.942	0.857	0.853	0.887	0.944	0.877	0.707	0.863
	+CRF ²	0.959	0.790	0.784	0.939	0.947	0.864	0.858	0.901	0.945	0.881	0.687	0.869
IoU	SegNet	0.886	0.730	0.193	0.898	0.785	0.720	0.378	0.482	0.791	0.373	0.446	0.607
	+CRF	0.911	0.751	0.200	0.901	0.790	0.732	0.410	0.494	0.796	0.382	0.452	0.620
BF	SegNet	0.877	0.597	0.542	0.699	0.765	0.623	0.437	0.384	0.630	0.492	0.446	0.616
	+CRF	0.887	0.609	0.577	0.740	0.801	0.646	0.506	0.423	0.676	0.549	0.493	0.654

SegNet¹: Conventional SegNet
+CRF²: SegNet with CRF

슷한 픽셀과 동일한 레이블을 갖도록 유도된다. 식 (1)에 정의된 대로 컬러 퍼텐셜은 단항 퍼텐셜과 함께 최적화되며, 이를 통해 I_R 의 골격을 유지함과 동시에 높은 선명도의 I_S^{CRF} 를 얻을 수 있다. 식 (1)의 최적화에는 평균장 근사화 (Mean field approximation) 기법을 적용하였다.

III. 실험 결과

제안하는 방법의 성능 비교를 위하여 CamVid 데이터 셋을 사용하였다. CamVid 데이터 셋은 주간 및 야간에 촬영된 자동차 주행 영상으로 구성되어 있으며, 시맨틱 영상 분할 성능 평가에 활용된다.

학습 영상 367장과 시험 영상 233장에 대해 총 11종의 레이블로 구분된 Ground truth 영상이 제공된다. 모든 영상의 크기는 480×360 이다.

비교 평가를 위한 실험은 시험 영상 233장을 이용하여 수행하였다. 그림 2는 기존 SegNet과 제안하는 CRF 적용 방법에 대한 정성적 비교를 보여준다. Ground truth 영상과 비교했을 때, 제안하는 방법이 더 높은 선명도를 보임을 확인할 수 있다. 표 1은 11개 레이블에 대한 정량적 성능 결과 비교를 나타낸다. IoU, BF 수치는 제안하는 방법이 모든 레이블에 대해서 우수함을 확인할 수 있다. 픽셀 정확도의 경우 9종의 레이블에 대해서는 성능 개선이 이루어졌으나, 2종의 레이블에 대해서는 성능 저하가 발생하였다. 성능 저하가 발생한 레이블은 기둥, 자전거와 같은 작은 물체들이었으며, 이는 작은 영역의 레이블에 대해 낮은 신뢰도를 주는 CRF의 모폴로지적 특성으로부터 기인하였다고 볼 수 있다. 각 측정법별 평균 성능은 제안하는 방법이 우수함을 확인할 수 있다.

IV. 결론

본 논문에서는 SegNet의 선명도 및 정확도를 개선하기 위해 CRF를 적용하는 방법을 제안하였다. CamVid 데이터 셋을 이용하여 기존 SegNet 방법의 결과와 제안하는 CRF 적용 방법의 결과에 대해 비교 평가를 수행하였으며, 제안하는 방법이 평균적으로 더 나은 성능을 가짐을 확인하였다. 본 연구결과는 시맨틱 영상 분할 기법 특히 자율주행 차량에서의 주행상황 인식 성능 개선에 실질적인 도움이 될 것으로 판단된다.

References

- [1] S. Hwang, et al., "Aircraft detection using deep convolutional neural network-based semantic segmentation," *J. Inst. Control, Robotics, and Syst.*, vol. 23, no. 8, pp. 625-634, Aug. 2017.
- [2] V. Badrinarayanan, et al., "SegNet: A deep convolutional encoder-decoder architecture for scene segmentation," *IEEE Trans. Pat. Anal. Mach. Intell.*, accepted, 2017.
- [3] P. Krahenbuhl and V. Koltun, "Efficient inference in fully connected CRFs with Gaussian edge potentials," *Advances in Neural Info. Process. Syst.*, vol. 24, pp. 109-117, Dec. 2011.
- [4] G. Brostow, et al., "Segmentation and recognition using structure from motion point clouds," in *Proc. 10th Eur. Conf. Computer Vision*, pp. 44-57, 2014.