논문 19-44-09-11 The Journal of Korean Institute of Communications and Information Sciences '19-09 Vol.44 No.09
https://doi.org/10.7840/kics.2019.44.9.1690

저정밀 레이블 지도를 이용한 의미론적 영상 분할

은 현 준•, 정 찬 호°, 김 창 익*

Sematic Image Segmentation Using Coarse Label Map

Hyunjun Eun•, Chanho Jung°, Changick Kim*

요 약

딥러닝 기반 의미적 영상 분할 방법은 학습을 위해 많은 비용이 요구되는 고정밀 레이블 지도가 필요하다. 이러한 문제점을 해결하고자 우리는 상대적으로 생성이 쉬운 저정밀 레이블 지도를 이용한 의미적 분할 방법을 제안한다. 제안하는 방법은 저정밀 레이블 지도를 활용하여 레이블 히트맵을 정의하고, 이를 CRF의 단항 퍼텐셜로 이용하여 의미적 분할을 수행한다. Cityscape 데이터 셋에 대한 성능평가에서, 제안하는 방법의 결과가 저정밀 레이블 지도 대비 더 향상된 성능을 내었고, 딥러닝 기반 방법 성능의 66%에 가까운 성능을 얻었다. 본 연구는 의미적 영상 분할 연구 및 고정밀 레이블 지도 생성에 도움이 될 것으로 판단된다.

Key Words : semantic segmentation, conditional random field, coarse label map

ABSTRACT

Deep learning based semantic segmentation methods require fine label maps, which costs expensive to generate, for training. To solve this problem, we propose a semantic segmentation method using coarse label maps easily obtained. The proposed method defines heat maps of labels from the coarse lable map utilized as unary potential in CRF for semantic segmentation. We used the Cityscape dataset for performance evaluation. Compared to a coarse label map, the proposed method achieved the improved performance. Also our method reaches 66% of the quality of a deep learning based method. We believe that this study guides for improving semantic segmentation and generating fine label maps.

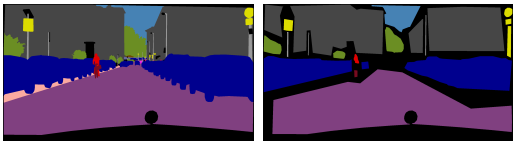
Ⅰ. 서 론

최근 높은 성능을 가지는 Fully Convolutional Network (FCN)이 의미적 영상 분할에서 많이 활용되고 있다.[1,2] 의미적 영상 분할은 다양한 물체의 모양을 고려하며 픽셀 단위 분류를 수행하기 때문에 어려운 작업이다. 그럼에도 불구하고 잘 구성된 데이터 셋의 공개는 FCN 기반 방법의 학습을 가능하게 한다. 그림 1은 여러 데이터 셋에서 영상과 해당 레이블 지도를 나타낸다. 그림 1(b)와 같이 학습에 필요한 고정밀 레이블 지도는 모든 픽셀의 레이블링이 필요하다. 고정밀 레이블링은 경계 영역에 대해 레이블링을 하지 않는 저정밀 레이블링과 비교해 약 1.5배의 시간이 소요된다. 이 때문에 많은 데이터가 필요한 딥러닝 기반 의미적 영상 분할은 데이터셋 구축에 많은 비용과 시간을 요구하며, 저정밀 레이블 지도를 이용하는 방법[3] 또한 연구가 되고 있다. 하지만 저정밀 레이블 지도를 이용하는 방법은 레이블 지도의 개선 없이 부분적으로 레이블이 없는 저정밀 지도를 그대로 사용하게 된다. 그 예로, [3]에서는 객체 검출 결과인 바운딩 박스 영역만을 레이블로 사용하고 있다. 이 때문에 고정밀 레이블 지도를 이용하는 방법에 비해 성능이 저하된다.

본 논문에서는 저정밀 레이블 지도의 레이블이 없는 영역에 대한 간단한 의미적 영상 분할 방법을 제안한다. 제안하는 방법은 경계 영역의 레이블이 없는 저정밀 레이블 지도를 대상으로 하여, 경계 영역의 의미적 영상 분할을 통해 고정밀 레이블 지도 생성을 목표

※ 본 연구는 NRF-2017R1C1B5015692 지원을 받아 수행되었습니다.
• First Author : (ORCID:0000-0001-7794-5377) School of Electrical Engineering, Korea Advanced Institute of Science and Technology, hj.eun@kaist.ac.kr, 학생회원
° Corresponding Author : (ORCID:0000-0003-3145-6732) Department of Electrical Engineering, Hanbat National University, peterjung@hanbat.ac.kr, 정회원
* (ORCID:0000-0001-9323-8488) School of Electrical Engineering, Korea Advanced Institute of Science and Technology, changick@kaist.ac.kr
논문번호 : 201904-051-A-LU, Received April 22, 2019; Revised June 3, 2019; Accepted June 9, 2019

(a) Image



(b) Fine label map (c) Coarse Label map

그림 1. 의미적 영상 분할을 위한 예제 영상 및 레이블 지도.
Fig. 1. Example images and label maps for semantic segmentation.

로 한다. 상대적으로 경계 영역은 다른 영역들에 비해 레이블링이 어렵고 많은 시간이 소요된다. 그림 1.(c)는 저정밀 레이블 지도를 나타내며, 검정색 영역이 비레이블링 픽셀이다. 일반적으로 사용되는 세밀한 레이블 지도 (그림 1.(b))와 다르게 저정밀 레이블 지도는 상대적으로 적은 시간 및 비용으로 생성할 수 있다. 우리는 이와 같은 저정밀 레이블 지도를 Conditional Random Field (CRF)[4]의 단항 포텐셜 (Unary potential)로 RGB 영상을 컬레 퍼텐셜 (Pair-wise potential)로 활용하여 의미적 영상 분할을 수행한다. 제안하는 방법의 성능 평가를 위하여 Cityscape[5] 데이터 셋을 이용하였으며, 그 결과 정확도, Intersection over Union (IoU) 수치에 대해서 저정밀 레이블 지도 대비 향상된 성능 및 고정밀 레이블 지도를 사용하는 딥러닝 기반 방법 중 하나인 SegNet[6] 성능의 66%에 가까운 성능을 얻었다.

Ⅱ. 제안하는 방법

본 논문에서는 RGB 영상 $I^{RGB} \in R^{WH\times 3}$와 픽셀의 레이블 분류 확률 히트맵 (Heat map) $I^{Prob} \in R^{WH\times C}$를 CRF의 입력으로 사용하여 의미적 영상 분할을 수행한다. 제안하는 방법의 전체 흐름도는 그림 2에 나타난다. H, C는 각각 영상의 높이, 영상의 너비, 레이블 수를 나타낸다. I^{Prob}는 각 픽셀이 각 레이블로 분류 될 확률로 정의되며 이는 저정밀 레이블 지도 $L^C \in R^{WH}$를 이용하여 정의 할 수 있다.

$$I^{Prob}(p,:) = \begin{cases} L^{CE}(p,:) & p \; is \; labeld \; on \; L^C \\ \boldsymbol{u} & \text{otherwise} \end{cases}, \tag{1}$$

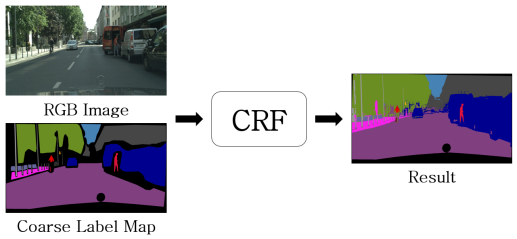


그림 2. 제안하는 방법의 흐름도.
Fig. 2. Flowchart of the proposed method.

$L^{CE} \in R^{WH\times C}$는 L^C의 원-핫 인코딩 (One-hot encoding) 매트릭스이다. 식 (1)은 L^C에서 레이블링 픽셀의 경우는 원-핫 인코딩 매트릭스에서 해당 픽셀의 벡터를 그대로 사용하고, 비레이블링 픽셀은 확률 벡터 $\boldsymbol{u} \in R^{1\times C}$로 정의하는 것을 나타낸다. $\boldsymbol{u}$는 해당 비레이블링 픽셀이 각 레이블로 분류 될 확률 벡터로 해당 비레이블링 픽셀 p와 주변 레이블링 픽셀 p_n의 유사도로 정의한다. 즉, p가 레이블 c로 분류될 확률은 아래 식으로 정의 한다.

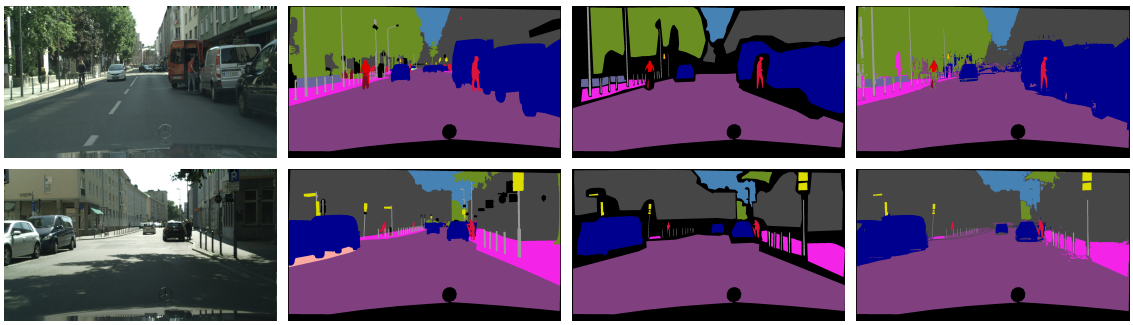
$$u_c = \max(\exp(-D^{RGB}(p,p_1^c) - D^{XY}(p,p_1^c), ..., \exp(-D^{RGB}(p,p_N^c) - D^{XY}(p,p_N^c))), \tag{2}$$

$D^{RGB}(p,p_n^c)$와 $D^{XY}(p,p_n^c)$는 p와 p_n^c의 RGB 컬러와 XY 거리의 유클리디안 거리 (Euclidean distance)로 정의되며, p^{c_n}는 L^C에서 레이블 c를 가지는 픽셀을 의미한다. 최종 $\boldsymbol{u}$는 L^C에서 레이블이 존재하여 정의된 u_c에 대해서만 소프트맥스 (softmax)를 취하여 확률 값을 사용하고, 레이블이 존재하지 않는 경우 u_c를 0으로 정의하게 된다. 즉, 식 (2)는 p가 p^{c_n}와 가깝고 RGB 컬러가 유사할수록 c 레이블을 가질 확률이 높게 만든다.

다음으로 I^{RGB}와 위에서 정의한 I^{Prob}를 CRF의 컬레 퍼텐셜과 단항 퍼텐셜에 활용하여 의미적 영상 분할을 수행한다. 구하고자 하는 의미적 영상 분할 레이블 지도 L^F의 벡터 형태를 $\boldsymbol{p}$라 할 때, CRF 최적화 목적 함수는 다음과 같이 정의된다.

$$\boldsymbol{p}^* = \text{argmax}_{\mathbf{p}} \mathrm{E}(\mathbf{p}),$$
$$E(\boldsymbol{p}) = \sum_p \theta_{unary}(p) + \sum_{p,q} \theta_{pair}(p,q), \tag{3}$$

$\theta_{unary}(p)$는 p번째 픽셀에서의 단항 퍼텐셜이다.



(a) Image (b) Fine label map (c) Coarse label map (d) Ours

그림 3. Cityscape 데이터셋에 대한 의미론적 영상 분할 결과의 정성적 비교.
Fig. 3. Qualitative comparison on sematic segmentation on the Cityscape dataset.

표 1. Cityscape 데이터셋에 대한 의미론적 영상 분할 결과의 정량적 비교.
Table 1. Quantitative comparison on semantic segmentation on the Cityscape dataset.

		Class									
		road	sidewalk	building	wall	fence	pole	traffic light	train	person	terrain
		sky	motorcycle	vegetation	car	truck	bus	traffic sign	rider	bicycle	avg.
Acc.	Coarse	.870	.338	.541	.169	.295	.307	.193	.195	.316	.104
		.580	.188	.547	.524	.243	.350	.373	.260	.253	.350
	Ours	**.966**	**.434**	**.770**	**.208**	**.351**	**.322**	**.205**	**.244**	**.363**	**.141**
		.856	**.213**	**.822**	**.655**	**.293**	**.415**	**.410**	**.289**	**.280**	**.433**
IoU	SegNet	.956	.701	.828	.299	.319	.381	.431	.347	.673	.623
		.917	.405	.873	.879	.217	.290	.446	.507	.566	.561
	Coarse	**.864**	.330	.531	.166	**.265**	**.297**	.189	**.195**	**.312**	.101
		.575	.181	.541	.522	.240	.346	**.361**	**.255**	.246	.343
	Ours	.861	**.379**	**.669**	**.170**	.255	.288	.189	.191	.309	**.118**
		.758	**.185**	**.672**	**.552**	**.259**	**.358**	.360	.254	.246	**.372**

이는 I^{Prob}에 의해 각 픽셀 위치마다 독립적으로 결정되며, CRF에 의해 결정될 레이블의 사전 단서로써 활용된다. $\theta_{pair}(p,q)$는 픽셀 p와 픽셀 q 쌍에 대한 켤레 퍼텐셜이며 다음과 같이 정의된다.

$$\theta_{pair}(p,q) = \exp\left(-D^{RGB}(p,q) - D^{XY}(p,q)\right) + \exp\left(-D^{RGB}(p,q)\right). \quad (4)$$

식 (4)의 첫 번째 텀은 Appearance Kernel로 거리가 가깝고 유사한 RGB 컬러값을 가지는 픽셀들이 동일한 레이블을 갖도록 유도한다. 두 번째 텀은 Smoothness Kernel로 크기가 작은 영역들은 제거하도록 유도한다. 식 (3)의 최적화 결과는 각 픽셀들이 I^{Prob}에서 높은 확률을 가지며 레이블을 가지며, 주변 RGB 컬러가 유사한 픽셀들과 동일한 레이블을 가지도록 한다. 식 (3)은 평균장 근사화 (Mean field approximation)를 이용하여 최적화 할 수 있다. CRF의 결과인 L^F는 L^C의 레이블링 픽셀의 레이블이 다를 수 있다. 이를 해결하기 위하여 최종 의미적 분할 레이블 지도 L을 다음과 정의한다.

$$L(p) = \begin{cases} L^C(p) & p \text{ is labeld on } L^C \\ L^F(p) & otherwise \end{cases}. \quad (5)$$

즉, 저정밀 레이블 지도에서 비레이블링 픽셀의 레이블은 CRF의 결과인 L^F에서 다른 픽셀의 레이블은 L^C를 따르도록 한다.

Ⅲ. 실험 결과

제안하는 방법의 성능 평가를 위하여 Cityscape 데이터 셋의 500장의 검증 영상 (2048×1024 크기)를 사용하였다. 해당 데이터 셋은 거리 영상을 다루고 있으며 35개의 클래스를 포함한다. 또한 영상 모두에 대해 고정밀 레이블 지도와 저정밀 레이블 지도를 제공한다. 그림 3은 제안하는 방법의 결과의 정성적 비교를 보여준다. 그림 3.(d)에서 보듯이 제안하는 방법으

로 생성된 결과가 세밀한 레이블 지도 (그림 3.(b)와 유사하며 저정밀 레이블 지도 (그림 3.(c))의 비레이블링 픽셀이 잘 레이블링 된 것을 확인할 수 있다. 표 1에는 저정밀 레이블 지도와 제안하는 방법의 정량적 성능 비교가 나타나 있다. 제안하는 방법의 저정밀 레이블 지도 개선 정도를 판단하기 위하여 정확도, Intersection over Union (IoU)를 성능 평가 지표로 사용하였다. IoU는 다음과 같이 정의된다.

$$IoU = (A_P \cap A_L)/(A_P \cup A_L), \quad (3)$$

A_P, A_L은 예측한 영역과 실제 레이블 영역을 나타내며, 값이 클수록 예측한 영역과 레이블 영역이 유사하다는 것을 의미한다.

Cityscape 데이터 셋에서는 표 1에 기재된 19개의 클래스를 정량적 성능 평가 대상으로 하며, 표 1의 성능은 500장 영상에 대한 평균을 나타낸다. 정확도에서는 모든 클래스에서 제안하는 방법이 저정밀 레이블 지도 대비 우수함을 확인할 수 있다. IoU 수치에 대해서는 크기가 작은 클래스의 경우 제안하는 방법과 저정밀 레이블 지도가 유사한 성능을 보이지만, 이외 클래스에 대해서는 제안하는 방법이 나은 성능을 나타낸다. 또한 저정밀 지도를 사용하며, 비학습 기반의 제안하는 방법은 고정밀 레이블 지도를 학습에 사용한 SegNet의 IoU 수치의 66%에 가까운 성능을 얻었다.

Ⅳ. 결 론

본 논문에서는 적은 비용이 드는 저정밀 레이블 지도 기반 의미적 분할 방법을 제안하였다. 제안하는 방법은 저정밀 레이블 지도를 CRF의 단항 퍼텐셜로 활용함으로써 의미적 분할 결과를 얻어낸다. Cityscape 데이터 셋을 이용하여 성능 비교를 수행하였으며 저정밀 레이블 지도 대비 향상된 성능을 가짐을 보였으며, 고정밀 지도를 학습에 이용하는 딥러닝 기반 방법 성능의 66%에 가까운 성능을 얻었다. 본 연구는 저정밀 지도 기반 의미적 영상 분할 연구에 실질적인 도움이 될 것으로 판단된다.

References

[1] E. Shelhamer, et al., “Fully convolutional networks for semantic segmentation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640-651, May 2017.

[2] S. Lim and D. Kim, “Semantic segmentation using convolutional neural network with conditional random field,” *J. KIECS*, vol. 12, no. 3, pp. 451-456, Jun. 2017.

[3] A. Khoreva, et al., “Simple does it: Weakly supervised instance and segmentation,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 876-885, Jul. 2017.

[4] P. Krahenbuhl and V. Koltun, “Efficient inference in fully connected CRFs with Gaussian edge potentials,” *Advances in Neural Inf. Process. Syst.*, vol. 24, pp. 109-117, Dec. 2011.

[5] M. Cordts, et al., “The cityscapes dataset for semantic urban scene understanding,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 3213-3223, Jun. 2016.

[6] V. Badrinarayanan, et al., “SegNet: A deep convolutional encoder-decoder architecture for image segmentation,” arXiv preprint arXiv: 1511.00561, 2016.