

MMSegmentation 논문 정리

2018년에 발표된 아래의 논문들의 abstract를 정리

- [PSANet \(ECCV'2018\)](#)
- [DeepLabV3+ \(CVPR'2018\)](#)
- [UPerNet \(ECCV'2018\)](#)
- [NonLocal Net \(CVPR'2018\)](#)
- [EncNet \(CVPR'2018\)](#)

1. PSANet

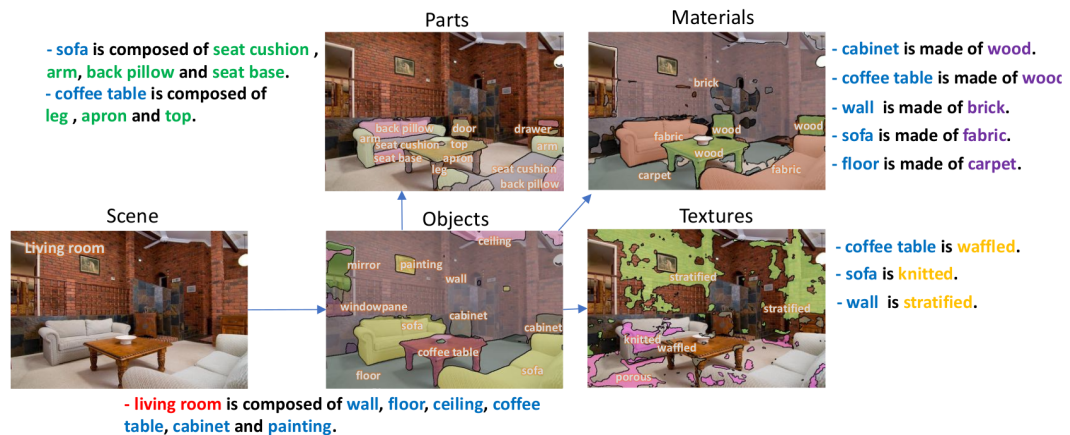
- 인용수 : 344
- Abstract
 - CNN은 local neighborhood region으로만 information flow가 제한되어 complex scenes의 전체적인 이해가 어렵다.
 - 본 논문에서 제안하는 *point-wise spacial attention network (PSANet)*는 neighborhood constraint를 완화한다.
 - feature map의 각 position들은 self-adaptively learned attention mask를 통해 다른 모든 position들과 연결되며, 양방향의 information propagation이 가능하다.
 - 다른 position의 정보들은 현재 position의 prediction에 도움이 되고, 반대로 현재 position의 정보는 다른 position들의 prediction에 도움을 주기 위해 distributed될 수도 있다.
 - 다음의 benchmarks에서 top performance를 기록하였다.
 - ADE20K, PASCAL VOC 2012, Cityscapes에서 top performance를 기록했다.
 - [ADE20K](#) : **Indoor** scenes
 - [PASCAL VOC 2012](#) : **Object**-centric images
 - [Cityscapes](#) : **Urban street** scenes

2. DeepLabV3+

- 인용수 : 3448
- Abstract
 - 기존의 네트워크는 다음과 같은 구조를 가진다.
 - 앞단의 network는 filter나 pooling등을 통해 contextual information을 encoding
 - 뒷단의 network는 spatial information을 점진적으로 복원하여 object의 boundary를 선명하게 capture
 - 본 논문에서 제안하는 *DeepLabV3+*는 위의 2가지 network의 장점을 모두 통합한다.
 - DeepLabV3에 Simple decoder module을 추가하여 segmentation 결과(특히 object의 boundary)를 refine하였다.
 - Xception의 depthwise separable convolution을 Atrous Spatial Pyramid Pooling 및 decoder module에 적용하여 빠르고 강력한 encoder-decoder network를 만들었다.
 - PASCAL VOC 2012 및 Cityscapes dataset에서 아무 전처리 없이 좋은 성능을 보였다.
 - [PASCAL VOC 2012](#) : **Object**-centric images
 - [Cityscapes](#) : **Urban street** scenes

3. UPerNet

- 인용수 : 196
- Abstract
 - 사람은 visual word를 multiple level로 인식한다.
 - scene categorization / object detection / texture or surface identification
 - 본 논문에서는 주어진 이미지에서 가능한 많은 visual concepts들을 인식하도록 하는 *Unified Perceptual Parsing*이라는 시스템을 연구하였고(아래 이미지 참고), heterogeneous image annotations로부터 모델을 학습시키기 위해 training strategy 및 multi-task 프레임워크 UPerNet을 개발하였다.



- 아래의 여러 데이터셋들을 합친 Broden Dataset (Broadly and Densely Labeled Dataset)에서 학습 및 테스트하였고, 이는 scenes, objects, object parts, materials and textures in a variety of contexts를 포함한다. 이때, objects, object parts, materials는 segmented down to pixel level / texture와 scene은 image level로 annotation 되어 있다.
 - [ADE20K](#) : Indoor **scenes**
 - [Pascal-Context](#) : additional annotations for PASCAL VOC 2010 (beyond the original PASCAL semantic segmentation task by providing **annotations for the whole scene**)
 - [Pascal-Part](#) : additional annotations for PASCAL VOC 2010. (beyond the original PASCAL object detection task by providing segmentation masks for **each body part of the object**)
 - [OpenSurfaces](#) : large database of annotated **surfaces**
 - [Describable Textures Dataset](#) : collection of **textural** images in the wild

4. NonLocalNet

- 인용수 : 2777
- Abstract
 - Convolutional, recurrent 연산은 한번에 하나의 local neighborhood를 처리하는 building block이다.
 - 본 논문에서는 long-range dependencies를 위한 generic family of building block으로서 *non-local operation*을 제안한다.
 - non-local operation은 한 위치에서의 response를 모든 position에서의 feature들의 weighted sum으로 계산하는데, 이 building block은 많은 computer vision 모델들에 붙여서 사용될 수 있다.
 - Video classification에서는 non-local model이 Kinetics, Charades 데이터셋에서 우수한 성능을 보였다.

- Static image recognition에서는 object detection / segmentation / pose estimation 모두에서 기존의 모델에 non-local block을 추가하여 성능이 향상됨을 확인하였다.
 - [MS COCO \(segmentation\)](#) : addresses both **stuff** and **thing** classes (Object-centric images로 보임)

5. EncNet

- 인용수 : 554
- Abstract
 - 본 논문에서는, scene의 semantic context를 capture하고 class-dependent feature map을 selectively highlight하는 *Context Encoding Module*을 제안하고, 이를 통해 global contextual information이 semantic segmentation에 미치는 영향을 연구하였다.
 - Context Encoding Module은 아래 dataset들에서 semantic segmentation 성능을 크게 향상시켰다. (약간의 extra computational cost만 추가됨)
 - [Pascal-Context](#) : additional annotations for PASCAL VOC 2010 (beyond the original PASCAL semantic segmentation task by providing **annotations for the whole scene**)
 - [PASCAL VOC 2012](#) : **Object**-centric images
 - [ADE20K](#) : **Indoor** scenes