**[Semantic Segmentation Model]**

- 출처:<https://paperswithcode.com/sota/semantic-segmentation-on-pascal-voc-2012>

⇒ 평가된 데이터셋의 종류에 따라 높은 성능을 보인 모델을 정리해 놓은 사이트

**1.** ***PASCAL VOC 2012 test*** (성능 높은순으로 나열)

|  |  |  |  |
| --- | --- | --- | --- |
| 모델 명 | 논문 abstract | 코드(github) | 논문 연도 |
| DeepLab v3+ | - 20년1월을 기준으로 현재까지 가장 성능이 뛰어난 모델  - Atrous convolution 구조 | [링크1](https://github.com/tensorflow/models/tree/master/research/deeplab)  [링크2](https://github.com/jfzhang95/pytorch-deeplab-xception) (pytorch) | 2018년 |
| [ExFuse](https://paperswithcode.com/paper/exfuse-enhancing-feature-fusion-for-semantic)  [(ResNeXt-131)](https://paperswithcode.com/paper/exfuse-enhancing-feature-fusion-for-semantic) | - 기존 모델과는 달리, low level feature와 high level feature를 단순히 혼합하는 것이 아니라 semantic 정보는 low level feature에 도입하고 high resolution detail은 high level feature에 도입함. 이렇게 함으로써 low level과 high level feature사이의 gap을 극복함. | 공개 안됨 | 2018년 |
| DPC | - meta-learning technique사용  - scene parsing, person-part segmentation, semantic image segmentation처럼 dense한 이미지 예측에 초점을 둠  - 이전 최신 알고리즘들 보다 parameter개수가 절반이고 computatational cost도 절반으로 줄어서 계산적인 측면에서 더욱 효율적임 | [링크1](https://github.com/tensorflow/models/tree/master/research/deeplab) | 2018년 |
| EncNet | - Context Encoding Module 개념을 도입하여 semantic segmentation에 global contextual information의 영향력을 부여함.  - 상대적으로 얕은 network (14 layer)를 갖고 있어 속도 향상 | [링크1](https://github.com/zhanghang1989/PyTorch-Encoding) (pytorch) | 2018년 |
| PSPNet | - 이전 SPP(spatial pyramid pooling)방식과는 달리, global feature를 얻기 위해  pyramid scene parsing network(PSPNet)방식을 사용. 또한 pixel-level feature를 global pyramid pooling 까지 확장함  è Local clue와 global clue가 결합하여 결과를 더욱 reliable하게 만듬 | [링크1](https://github.com/tensorflow/models/tree/master/research/deeplab)  [링크2](https://github.com/CSAILVision/semantic-segmentation-pytorch) (pytorch) | 2016년 |

***2.*** ***Cityscapes Dataset***

|  |  |  |  |
| --- | --- | --- | --- |
| 모델 명 | 논문 abstract | 코드(github) | 논문 연도 |
| OCR | - 객체의 class에 상응하는 픽셀을 characterize  à 1) ground-truth segmentation으로 객체의 region을 탐색  2) 객체의 region 내에 있는 픽셀을 수집  3) 각 픽셀과 객체 region사이의 관계를 계산하고, 픽셀과 객체 region 사이의 관계에 따라 weighted aggregation 실행 | [링크1](https://github.com/HRNet/HRNet-Semantic-Segmentation) (pytorch)  [링크2](https://github.com/PkuRainBow/OCNet.pytorch) (pytorch) | 2020년 |
| DCNAS | - 대규모의 target dataset에서 최적의 네트워크 구조를 직접 찾아냄  by connecting cells with each other using learnable weights, we introduce a densely connected search space to cover an abundance of mainstream network designs | 공개 안됨 | 2020년 |
| HANet | - urban-scene image에서의 semantic segmentation 성능을 향상시키기 위해 고안됨  urban-scene images have their own distinct characteristics, but most semantic segmentation networks do not reflect such unique attributes in the architecture. The proposed network architecture incorporates the capability exploiting the attributes to handle the urban scene dataset effectively | [링크1](https://github.com/shachoi/HANet) (pytorch) | 2020년 |
| **Gated-SCNN** | - 최근 방식들은 color, shape, texture의 정보를 deepCNN내에서 한번에 처리.  하지만 이런 방식은 서로 다른 유형의 정보를 가질때는 적절하지 않음  è two-stream CNN architecture : shape정보를 가진 stream과 기존의 stream을 병행으로 처리함. 두개의 stream을 연결하는 intermediate layer를 갖는다 | [링크1](https://github.com/nv-tlabs/GSCNN) | 2019년 |

***3.*** ***ADE20K Dataset***

|  |  |  |  |
| --- | --- | --- | --- |
| 모델 명 | 논문 abstract | 코드(github) | 논문 연도 |
| CPN | - Intra-class와 inter-class context를 명확히 구별하기 위해 feature aggregation를 고안.  Given an input image and corresponding ground truth, Affinity Loss constructs an ideal affinity map to supervise the learning of Context Prior. The learned Context Prior extracts the pixels belonging to the same category, while the reversed prior focuses on the pixels of different classes | [링크1](https://github.com/ycszen/ContextPrior) | 2020년 |
| LaU-regression-loss | we present a Location-aware Upsampling (LaU) that adaptively refines the interpolating coordinates with trainable offsets. Then, location-aware losses are established by encouraging pixels to move towards well-classified location | [링크1](https://github.com/HolmesShuan/Location-aware-Upsampling-for-Semantic-Segmentation) (pytorch) | 2019년 |
| PSPNet | 위와 동일 |  |  |
| **EncNet** | 위와 동일 |  |  |

이 외에도 21가지의 서로 다른 dataset에 따른 각 모델의 성능이 순위별로 공개 되어 있음.

(KITTI, Cityscapes val, Semantic3D 등등…)

Video, 3D image data를 제외하고

모든 dataset에서 공통적으로 높은 성능을 보인 모델을 차출하면,

1. OCR – 7개의 dataset에서 높은 순위를 차지

2. DCNAS – 4개의 dataset에서 높은 순위를 차지

3. DeepLabv3+, HRNetV2, CPN, LaU-regression-loss, PSPNet, EncNet - 3개의 dataset에서 높은 순위를 차지

*※※ 주의: backbone을 무엇으로 사용했는 지에 따라 성능이 다름 (ex ORC모델 이라고 하더라도 backbone이 HRNetV2, ResNet-101 등 다양할 수 있음)*

모든 데이터셋을 종합해 보았을 때, OCR의 성능이 공통적으로 높게 평가되었으며 그 다음은 DCNAS이다.

두 모델의 논문과 코드를 중점적으로 자세히 살펴볼 예정.

[Segmentation in OCR]

Sementation에 관한 초기 연구에서는 주로 객체의 spatial scale(공간적 규모)만 고려하였음 (ASPP, PPM 등)

최근 여러 논문들 (DANet, CFNet, OCNet 등)은 객체의 위치와 그것의 contextual position(문맥적 위치)과의 관계를 고려하고, 유사한 표현들에 대해서는 더 높은 가중치를 부여하여 contextual position을 집계함.

Motivation – 하나의 픽셀에 부여된 class는 픽셀이 속해있는 객체의 category라는 점

목표는 픽셀에 상응하는 class의 object region에 대한 표현을 반영함으로써 하나의 픽셀에 대한 표현을 확장(augementation)시키는 것.