

Image Localization with Weakly Supervised Learning

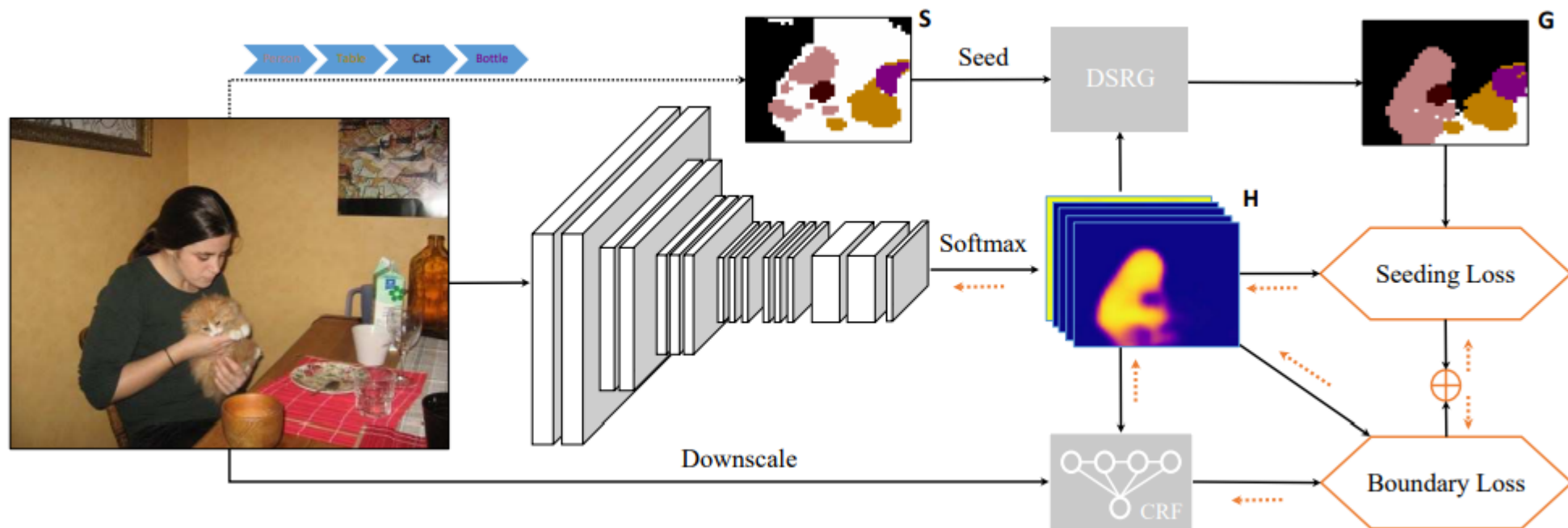
Weakly-supervised learning

- 적고 불완전한 데이터로 간단한 classifier를 만들고, 그 classifier의 출력 결과를 학습 set으로 보고 train을 하는 방법.
- self-training, bootstrapping이라고도 한다.
- label된 데이터가 적거나, 아니면 label이 불완전한 데이터가 있어도 효과적으로 학습이 가능하다는 장점이 있다.
- 그러나 label이 잘못된 비중이 너무 크면 학습이 잘 안 된다는 단점도 있다.
- Unsupervised Learning은 label이 붙여진 데이터를 사용하지 않는다는 차이점이 있다.

논문 소개

- CAM 모델과 label만 매겨진 이미지 데이터를 가지고 object의 region을 추출하는 모델 제안
- Huang, Z., Wang, X., Wang, J., Liu, W., & Wang, J. (2018). Weakly-supervised semantic segmentation network with deep seeded region growing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 7014-7023).
- Paper:
http://openaccess.thecvf.com/content_cvpr_2018/papers/Huang_Wweakly-Supervised_Semantic_Segmentation_CVPR_2018_paper.pdf

개략적 방법

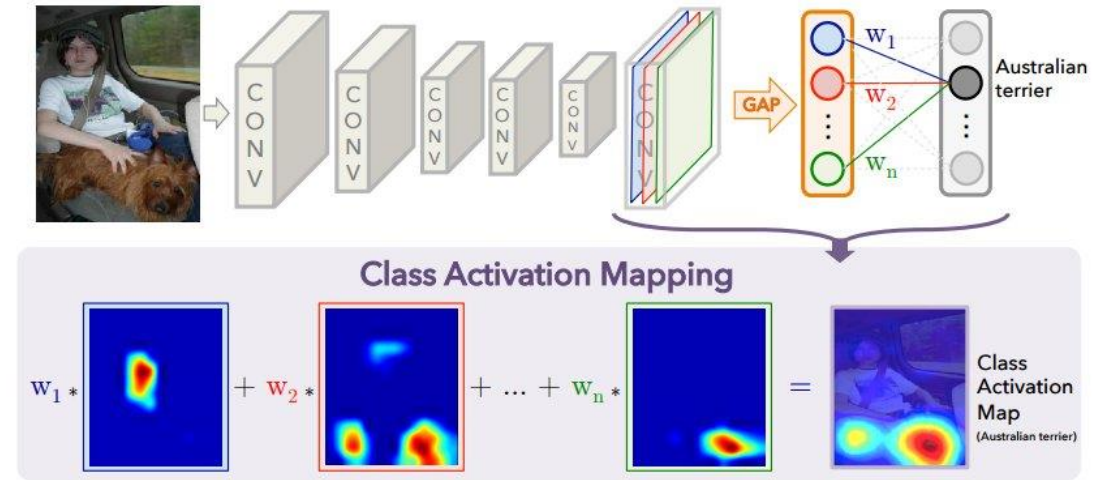


방법

- 불완전한 segmentation 데이터 만들기
 - Pre-trained Classifier를 이용해 CAM 추출 후, 일정 임계점 이상의 값을 각각 map으로 만듦 (S)
- DeepLab-ASPP 기반의 segmentation model 구현
 - segmentation model의 predict 값 추출 (H)
- S값, H값을 이용해 Region Growing 수행 (DSRG), 그리고 fully-connected CRF를 통과한 값을, 다시 segmentation model의 학습 데이터로 사용

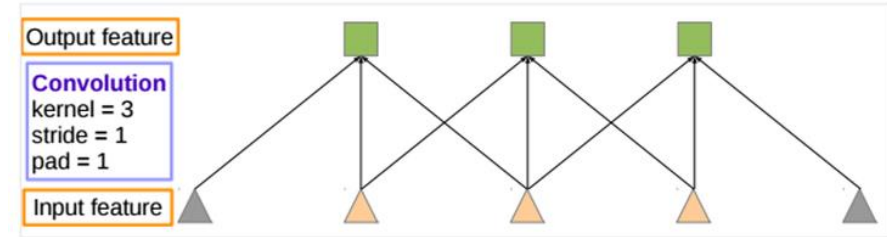
불완전한 segmentation 데이터 만들기

- CAM(Class Activation Map)을 이용.
- VGG16 등 image classification 모델의 feature layer를 Global Average Pooling하여 Activation 추출.
- 특정 픽셀에 있는 class는 activation 값이 임계점 이상이면서 그 픽셀에서 최대한 class로 선택한다. 임계점을 넘는 activation이 없으면 unlabeled로 둔다.

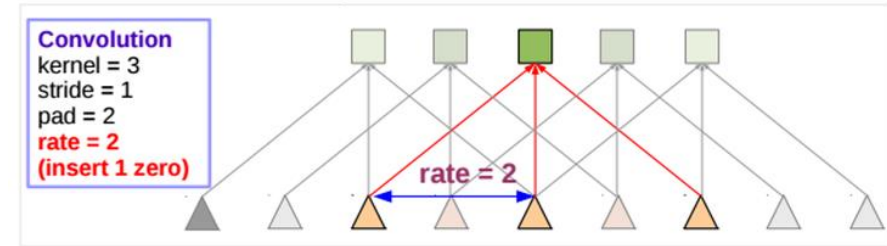


DeepLab-ASPP 기반의 segmentation model

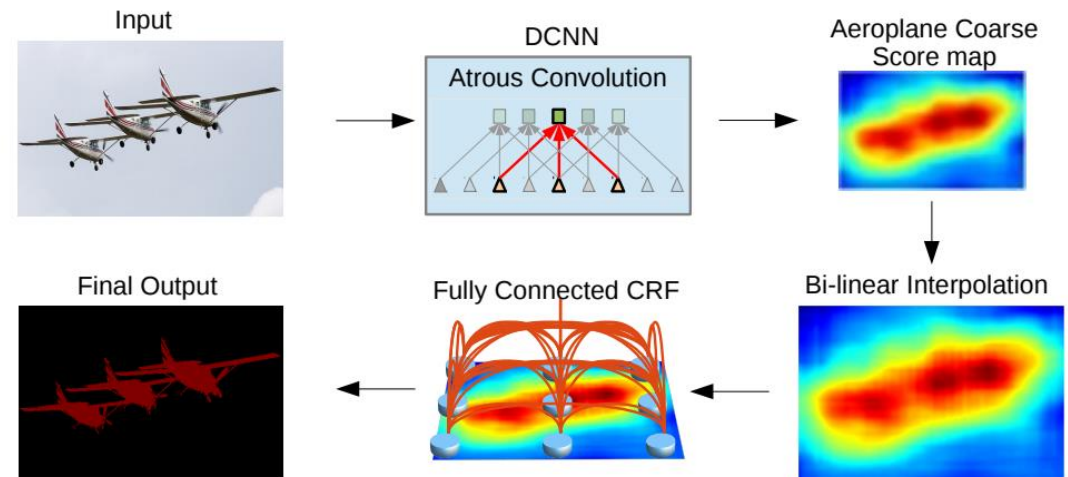
- <https://arxiv.org/pdf/1606.00915.pdf>
- segmentation model은 이 모델의 구현을 그대로 사용하였다.
- Atrous Convolution을 여러 겹 쌓은 모델의 출력값을 score로 보고 fully-connected CRF를 통과시켜 segmentation을 하는 모델
- Atrous Convolution: 중간에 빈 공간(hole)을 두어, 넓은 간격을 두고 데이터를 읽는 방법



(a) Sparse feature extraction



(b) Dense feature extraction



Loss Function

- seed loss의 의미: foreground에 있는 class의 probability, background에 있는 class의 probability가 최대가 되도록 함
 - $H_{u,c}$: segmentation map H 에 대하여, position u 에서 class c 일 확률
- boundary loss: 테두리를 더 명확하게 해주기 위해 사용

$$\ell_{seed} = -\frac{1}{\sum_{c \in \mathcal{C}} |S_c|} \sum_{c \in \mathcal{C}} \sum_{u \in S_c} \log H_{u,c} - \frac{1}{\sum_{c \in \bar{\mathcal{C}}} |S_c|} \sum_{c \in \bar{\mathcal{C}}} \sum_{u \in S_c} \log H_{u,c}, \quad (1)$$

Deep Seed Region Growing

- seed(CAM 값)을 보니 label된 픽셀이 sparse하게 퍼져 있었다.
따라서 segment를 묶어 주기 위해 이 방법을 사용한다.
 - $H_{u,c}$ 값이 임계점 이상이고, class c 의 확률값이 픽셀 u 에서 최대가 맞으면 그 class로 바꾼다.
 - 아니면 그냥 unlabeled된 채로 둔다.

fully-connected CRF

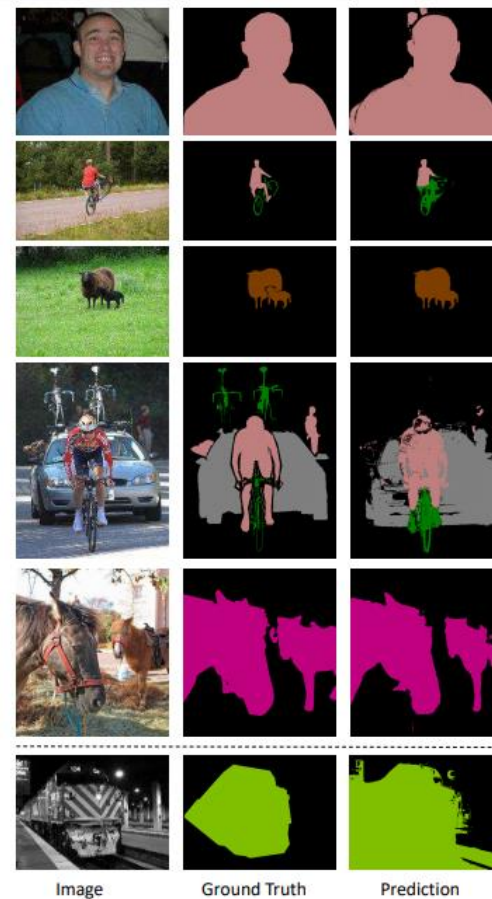
- 컴퓨터 비전 등에서 이미지나 region의 노이즈를 줄이기 위해 사용하는 방법
- 복잡한 수식을 이용해 노이즈를 줄이는 방법으로, 자세한 내용은 생략.

Re-train

- 위에서 얻은 segmentation data를 다시 학습 데이터로 보고, 다시 segmentation model을 학습한다.
- 처음에 seed로 부정확한 CAM 값을 사용한 뒤, 그 다음부터 모델의 출력값을 다시 학습 데이터로 사용하는 것이 핵심

Result

- 성능이 아주 좋은 것은 아니지만, 부정확한 값인 CAM 값을 가지고 이미지 segmentation을 수행할 수 있다는 가능성을 보여주었다.



Reference

- http://openaccess.thecvf.com/content_cvpr_2018/papers/Huang_Weakly-Supervised_Semantic_Segmentation_CVPR_2018_paper.pdf
- <https://laonple.blog.me/221000648527>