· Semantic Segmentation :

← 주어진 이미지의 모든 픽셀에 대해 classification을 진행

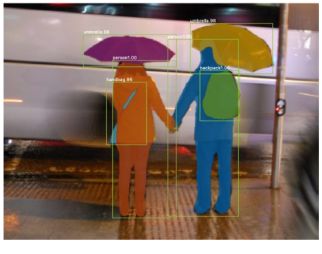
- 검정: background (foreground가 아닌 부분)
- 붉은색: 비행기
- 분홍색: 사람들

클래스가 같은 픽셀은 모두 같은 label을 갖고 있음

scene의 전체에 대한 이해가 필요할 때 유용

ex.) bio 분야에서 세포등을 관찰

· Instance Segmentation :



- 모든 픽셀에 대해 추론하지 않음
  (오직 foreground를 찾아내는 것에 집중)
- 객체마다 찾아낸 bounding box를 기준으로 하여 그 안의 한가지 object에 대해서 segmentation 작업을 진행한다.
- 결과적으로, object에 해당하는 픽셀에 대해서만 binary classification을 함.
- 같은 class에 해당하는 instance 하나하나를 구별함.

픽셀이 (class, id)의 pair을 추론

→ id는 같은 class의 서로 다른 instance를 구분지어줌

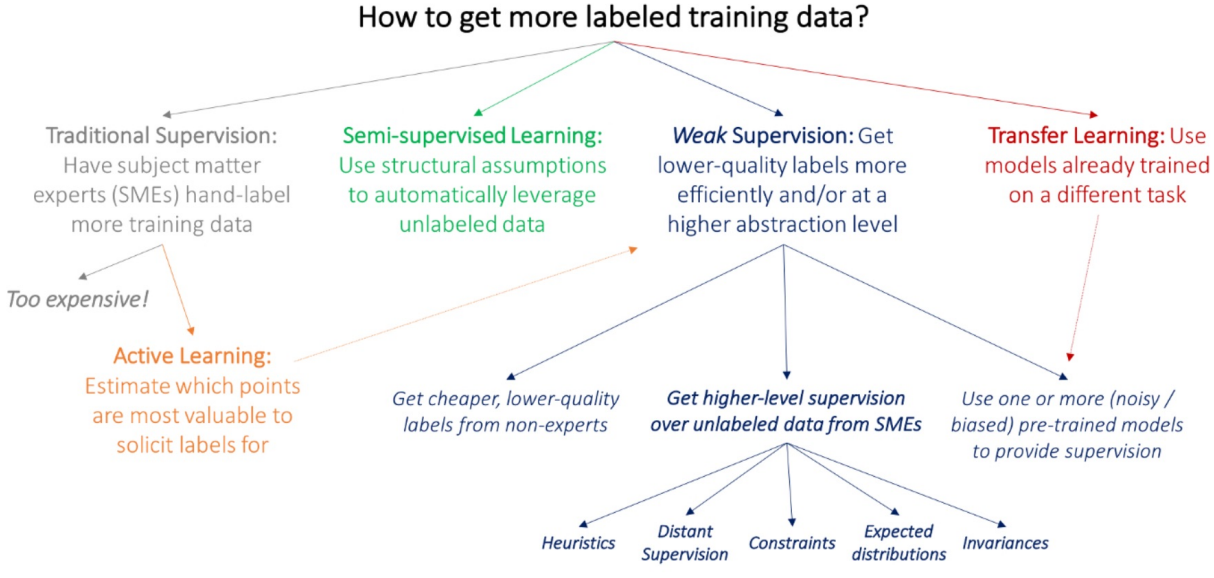
실시간으로 물체를 찾아내는 작업 같이 속도가 중요시되는 일에 유용

<Weakly Supervised Learning>

When to use:

- The available labeled data is insufficient to obtain a supervised model with good performance.
- The available labeled data is noisy or obtained from an imprecise source.
- There are insufficient domain experts, or getting them is very costly.
- The available time for manual annotation is very limited.

→ Weakly supervised learning is to decrease the cost of human experts annotation and increase the available labeled data for training.



Weak Supervision

① Incomplete supervision: only a subset of training data are labelled

② Inexact supervision: the training data are given with labels but not as exact as desired

③ Inaccurate supervision: in the training data, there are some labels with mistakes.

<Incomplete supervision>

- active learning: domain expert
- semi-supervised: generative models, label propagation, TSVM

CAM : bounding box가 없이도, CAM을 사용하면 bounding box가 위치한 곳을 가리킬 수 있어서 labeling 없이, bounding box를 정해주는 모국가 사용할 수 있다.

global max pooling: ordinary maxpooling layer with pool size equals to the size of the input.

ex) 0, 1, 2, 2, 5, 1, 2 → 5

input output

[Weakly Supervised Learning of Instance Segmentation with Inter-pixel Relations]

↑ Inexact supervision the training data are given labels but not as exact as desired.

: Generates psuedo instance segmentation labels of training images

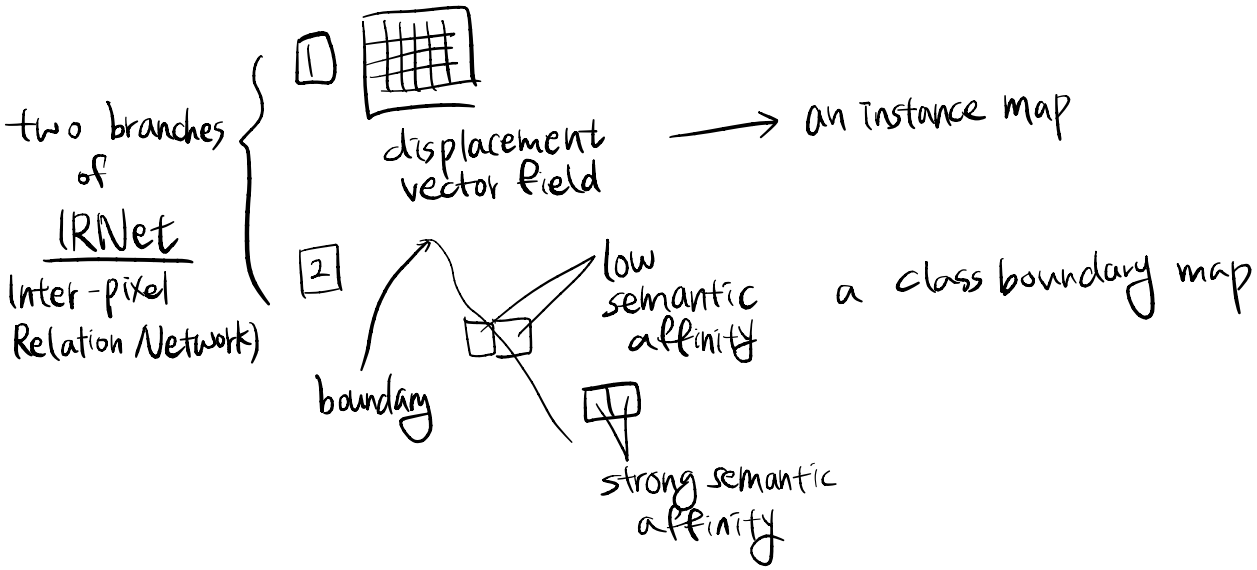
Complement CAM
- ✗ distinguish different instances (instance-agnostic)
- ✗ find entire instance areas with accurate boundaries.

→ a class-agnostic instance map

→ pairwise semantic affinity

⇩

Generate a psuedo instance segmentation label by selecting the instance label with the highest attention score in the instance-wise CAMs at each pixel.



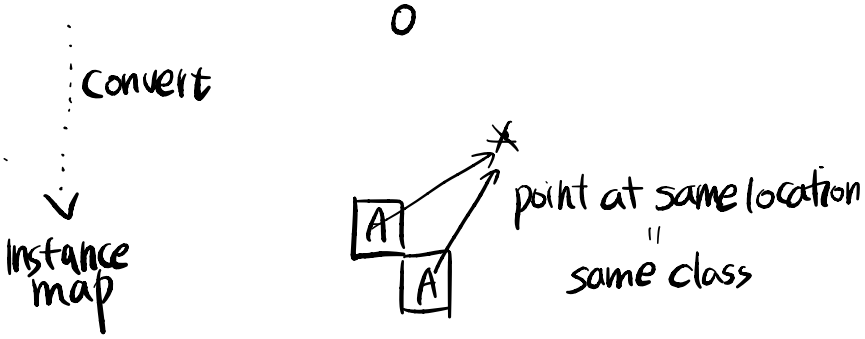
⇩

Collect pixels with high attention scores
& train IRNet with the displacements and class equivalence between the collected pixels.

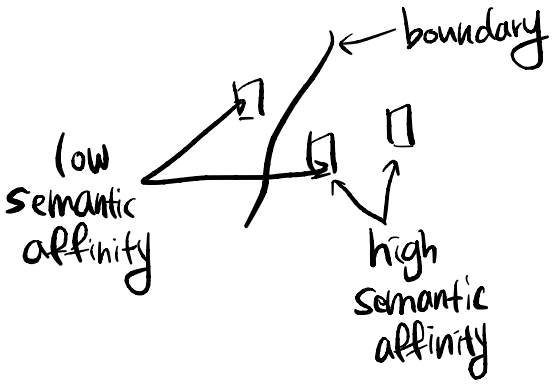
∴ No supervision in addition to image-level class labels is required.

→ weak supervised learning

Displacement vector field: the centroid of the instance the pixel belongs to.



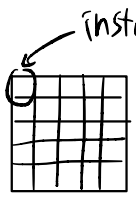
Class Boundaries : compute semantic affinities between pairs



<Roles of Class Attention Map (CAM)>

① define seed areas of instances ------> entire instance areas.
reliable later

② extract inter-pixel relations ------> source of supervision for IRNet



Displacement D → class-agnostic instance map (I)

: a rough instance segmentation mask without class labels nor accurate boundaries.

Class Boundary B → pairwise affinities

↓

semantic affinities

Confidence score for class equivalence between a pair of pixels.

incorporation of instance-agnostic CAMs ⟶ instance-wise CAMs