**Image Segmentation(2) - 기술과 딥러닝 기반 모델(FCN, U-Net, SegNet, DeepLab, ResNet)**

Q. Image Segmentation?

A. 이미지 분할은 디지털 이미지를 여러 개의 의미 있는 부분 또는 객체로 나누는 과정입니다. 컴퓨터 비전과 이미지 처리에서 중요한 기술로, 각 부분이 유사한 속성을 가지거나 동일한 객체를 나타내도록 합니다.

Q. 주요 목적은?

A. 객체를 인식하고 검출하여 객체의 위치와 경계를 정확히 파악합니다. 그리고 다양한 분야에서 이미지 내의 특정 영역을 분석합니다. 추가로 특정 영역을 선택하여 색상 변경, 제거, 합성 등의 작업을 수행합니다.

Q. 주요 방법들은?

A. Thresholding, Clustering, Edge Detection, Region Growing, Graph Cuts & Energy Minimization, Watershed Based Method들이 있습니다.

Q. Thresholding?

A. 임계값 처리 방법으로 픽셀 값이 특정 임계값(threshold)보다 크거나 작은지에 따라 이미지를 이진화 하는 간단한 방법입니다.

Q. Clustering?

A. 픽셀을 유사한 속성을 가진 그룹으로 분할합니다. K-means 클러스터링, Mean Shift 클러스터링 방법이 있습니다.

Q. Edge Detection?

A. 에지 검출 방법은 이미지에서 경계선을 감지하여 객체를 분할합니다. 검출하는 방법에는 Canny Edge Detector, Sobel Filter, Laplacian of Gaussian이 있습니다.

Q. Region Growing?

A. 초기 시드(seed) 픽셀에서 시작하여 유사한 속성을 가진 인접 픽셀을 확장합니다. 시드 포인트를 선택하고 확장 기준을 세운 후 확장합니다.

Q. Graph Cut & Energy Minimization?

A. 그래프 컷 방법은 이미지를 그래프로 표현하고 최소 컷을 찾아 분할하는 방법입니다. 각 픽셀을 그래프의 노드로, 픽셀 간의 유사성을 엣지로 표현합니다. 그리고 그래프를 두 개의 서브 그래프로 나누는 최소 컷을 찾습니다. 최대 유량 – 최소 컷 정리를 사용하여 최적의 분할을 수행합니다.

Q. Watershed Based Method

A. 이미지에서 그래디언트의 크기로 구성된 그래디언트 magnitude 영상을 구하고, 그 영상으로부터 Watershed(산의 봉우리, peak)를 구해 그 watershed를 영역을 구분해주는 역할로 이용하고 이미지를 분할하는 방식입니다.

추가적으로 Deep learning based method도 있습니다. 머신러닝 알고리즘을 사용해 이미지에 있는 물체의 특징을 학습한 뒤, 알고리즘을 사용하여 세그먼트를 객체 또는 배경으로 분류하는 방법입니다. Deep learning based method에 대해서 자세히 살펴보도록 하겠습니다.

1. FCN(Fully Convolutional networks)

* Convolution Layer를 통해 Feature 추출
* 1x1 Convolution Layer를 통해, 낮은 해상도의 Class Presence Heat Map 추출
* Transposed Convolution을 통해 낮은 해상도의 heatmap을 Upsampling한 뒤, input과 같은 크기의 Map 생성
* Map의 각 pixel class에 따라 색칠한 뒤, Segmentation 결과 반환

1. U-Net

* Contracting Path
  1. Encoder의 역할을 수행하는 부분으로, 전형적인 Convolution Network로 구성
  2. 입력을 Feature Map으로 변형해 이미지의 Context를 파악
  3. 점진적으로 Spatial dimension을 줄여가며 고차원의 semantic 정보를 convolution filter가 추출해낼 수 있게 됨
  4. Contracting Path의 앞단에 이미 잘 학습된 모델을 Backbone으로 사용해 학습 효율과 성능을 높일 수 있으며, 주로 ResNet 등의 모델을 사용함.
* Expanding Path
  1. Decoder의 역할을 수행하는 부분으로, 전형적인 Upsampling + Convolution Network로 구성
  2. Convolution 연산을 거치기 전, Contracting Path에서 줄어든 사이즈를 다시 복원(Upsampling)하는 형태
  3. Expanding Path에서는 Contracting을 통해 얻은 Feature Map을 Upsampling하고, 각 Expanding 단계에 대응되는 Contracting 단계에서의 Feature Map과 결합해서(Skip-Connection Concatenate) 더 정확한 Localization을 수행
  4. encoder에서 spatial dimension 축소로 인해 손실된 spatial 정보를 점진적으로 복원하여 정교한 boundary segmentation을 완성
  5. Multi-Scale Object Segmentation을 위해 DownSampling과 UpSampling을 순서대로 반복하는 구조

1. SegNet

* Encoder에서는 VGG16의 13개 Convolution Layer를 동일하게 사용한다. Convolution, Pooling 구조들을 통해 Input Image의 feature map을 추출한다.
* Decoder에서는 Encoder에서 뽑은 feature map을 Upsampling과 Convolution하여 원래 크기로 되돌린다. Upsampling 후 마지막 Layer에서는 각 Class의 픽셀 단위 분류를 위한 Softmax Layer가 존재한다.

1. DeepLab

* 기존 convolution과 달리 필터 내부에 빈 공간을 둔 채로 작동한다.
* 얼마나 빈 공간을 둘지 결정하는 파라미터인 rate가 1일 때는 기존 convolution과 동일하고 rate가 커질수록 빈 공간이 넓어진다.
* 기존 convolution과 동일한 양의 파라미터와 계산량을 유지하면서 한 픽셀이 볼 수 있는 영역을 크게 할 수 있게 된다.

1. ResNet

* VCG-19의 구조를 뼈대로 하여 convolution layer를 추가해 깊게 만든 후, shortcut들을 추가한다.