SegNet

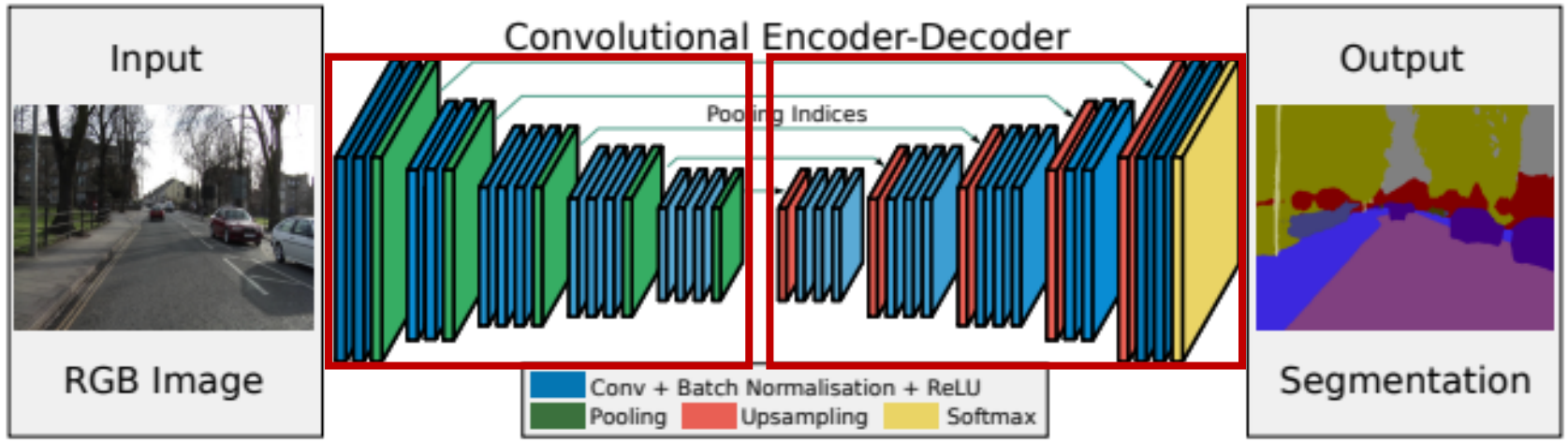
A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation

SegNet A Deep Convolutional Encoder Decoder 논문.pdf

SegNet

- 연산량 ↓
- Segmentation 성능 ↑

SegNet - Architecture



Encoder : VGG16에서 FC를 제거하여 Convolution layer들만 가져온 것을 사용

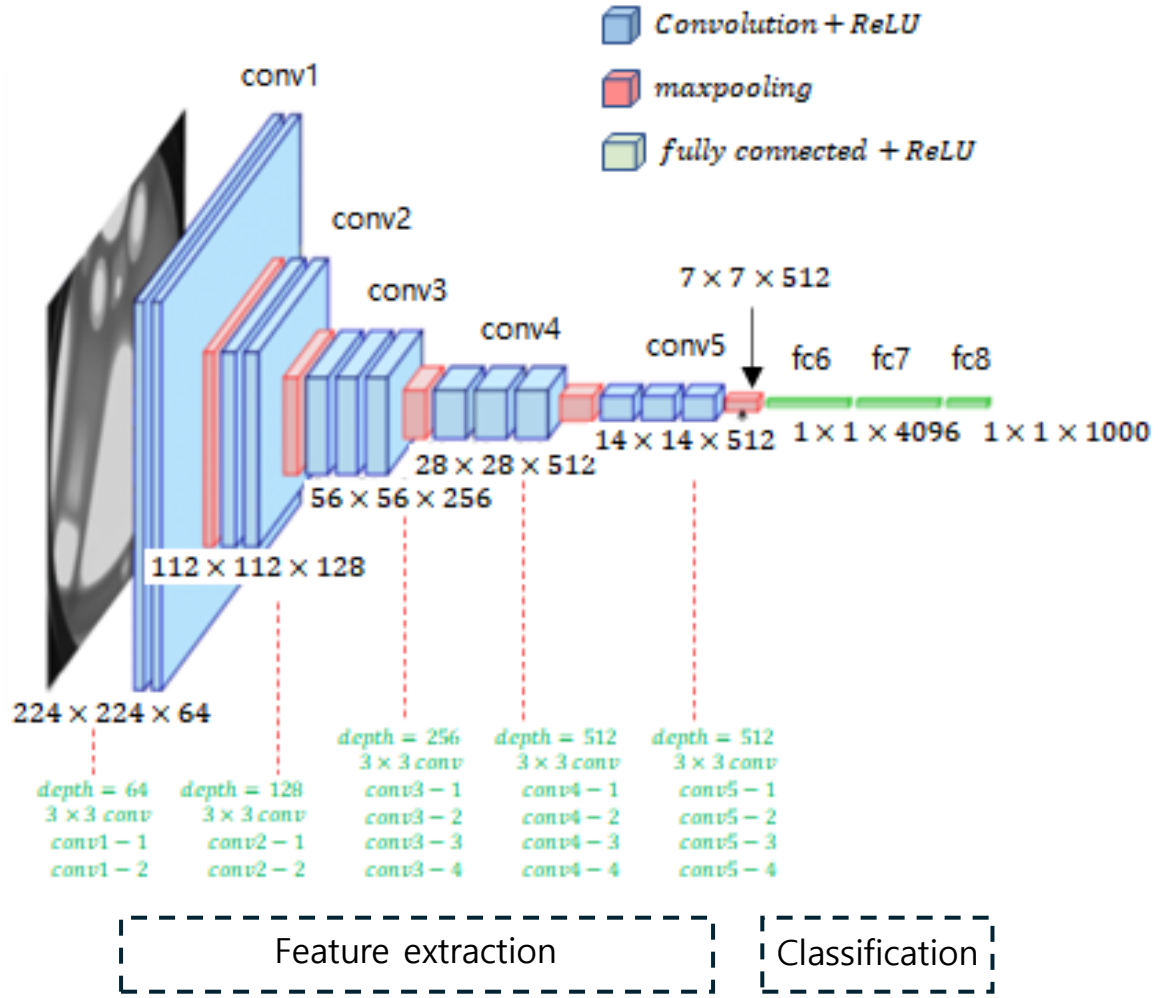
Decoder : Encoder를 좌우대칭 시킨 모양으로 설계

VGG16

- 신경망을 깊게
- 그러니 CONV에서 Filter 사이즈는 3x3으로 고정

- Filter 사이즈가 3x3 이니 생기는 장점 업데이트 해 줘야 하는 파라미터 수 줄어들었다.
- 여러 번 봐서 특징을 더 잘 기억한다.
- 그리고 cnn이 기본 구조여서 연산량 적다.

VGG16 - Architecture



Input Image

- 224x224x3 이미지 입력

Conv 1-1

- 64개의 3x3x3 필터로 입력 이미지를 convolution
- stride, zero padding = 1 -> 224x224x64 feature map
- ReLU 활성화 함수

Conv 1-2

- 64개의 3x3x64 필터로 feature map을 convolution
- Stride, zero padding = 1 -> 224x224x64 feature map
- ReLU 활성화 함수
- 2x2 max pooling을 stride=2 -> 112x112x64 feature map

Conv 2-1

- 128개의 3x3x64 필터로 feature map을 convolution
- Stride, zero padding = 1 -> 112x112x128 feature map
- ReLU 활성화 함수

⋮

Fc1

- 7x7x512의 feature map을 flatten 하여 25,088개의 뉴런 생성
- Fc1 층의 4,096개의 뉴런과 fully connected

Fc2

- Fc1층의 4,096개의 뉴런과 4,096개의 뉴런으로 fully connected

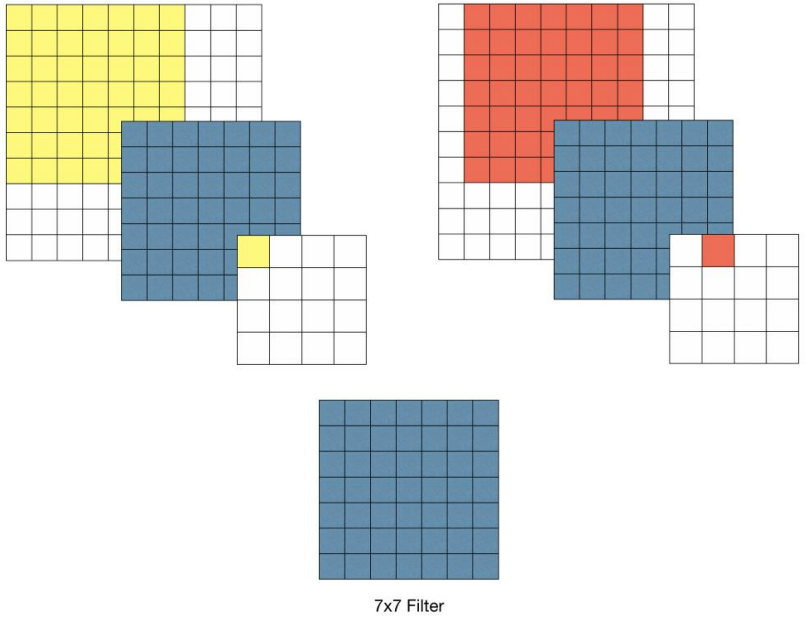
Fc3

- Fc2층의 4,096개의 뉴런과 fc3의 1,000개의 뉴런과 fully connected
- 출력값들은 softmax 함수로 활성화
- 마지막 뉴런의 수는 클래스의 개수를 의미

SegNet

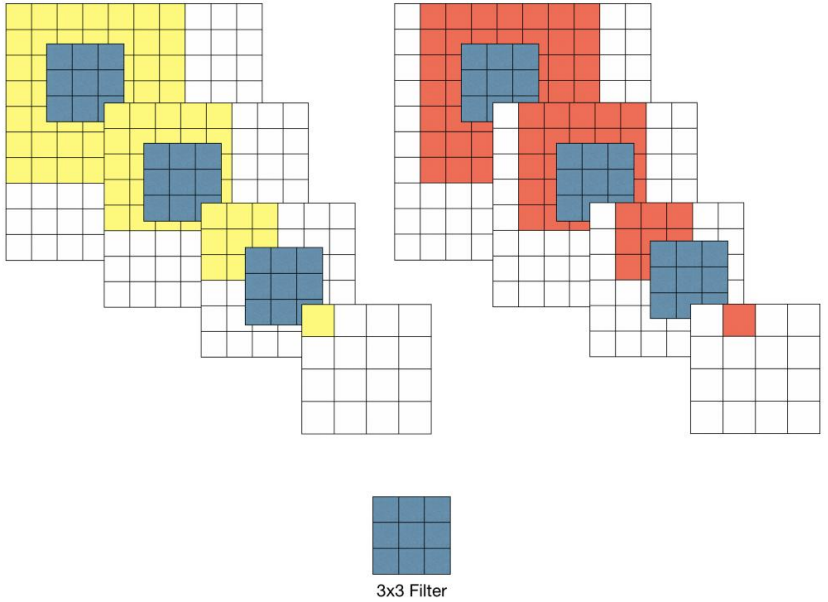
- 연산량 ↓
  - VGG16
    - CNN : parameter sharing
    - Fc 제거 : Encoder에서 classification 필요 x

기존 모델



→ (7x7 filter) x convolution 연산 1회 = 49개의 Parameter

VGG 16 모델

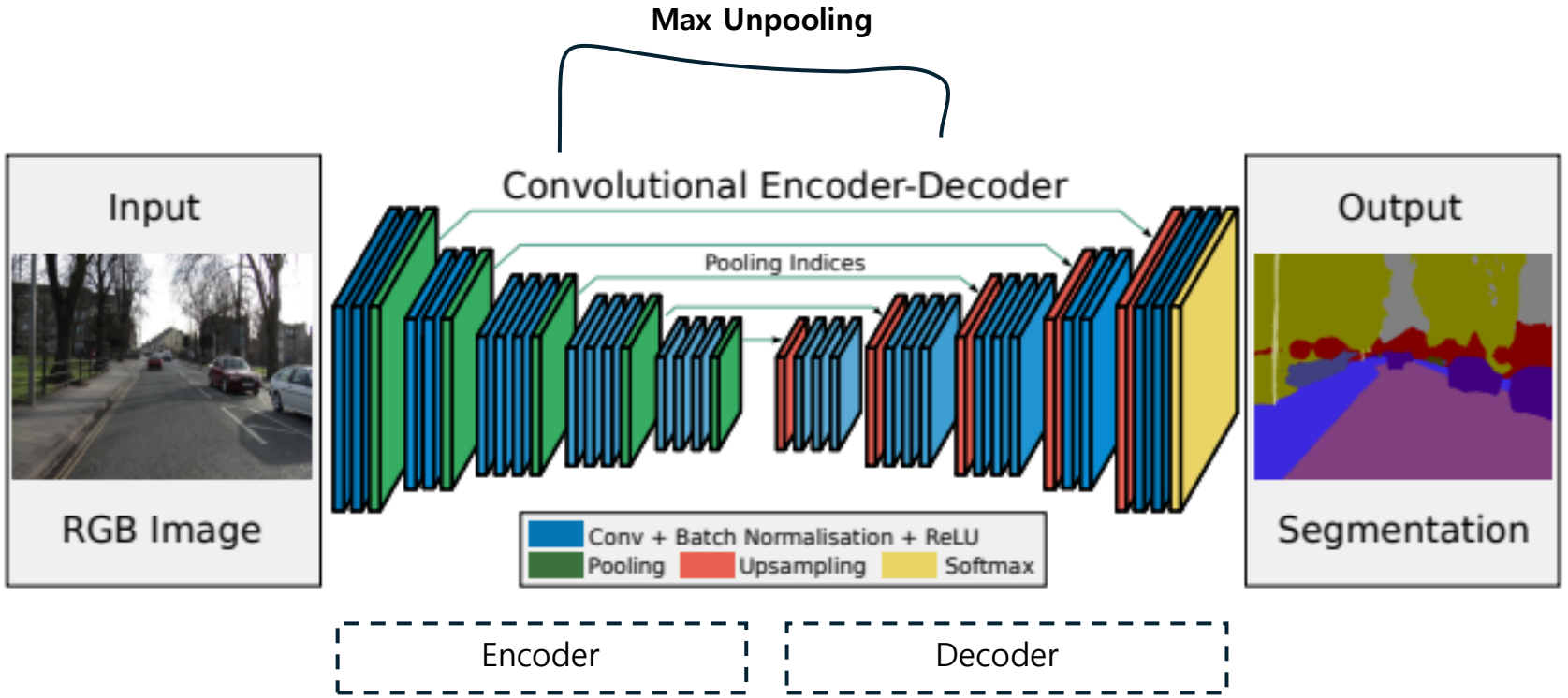


→ (3x3 filter) x convolution 연산 3회 = 27개의 Parameter

SegNet

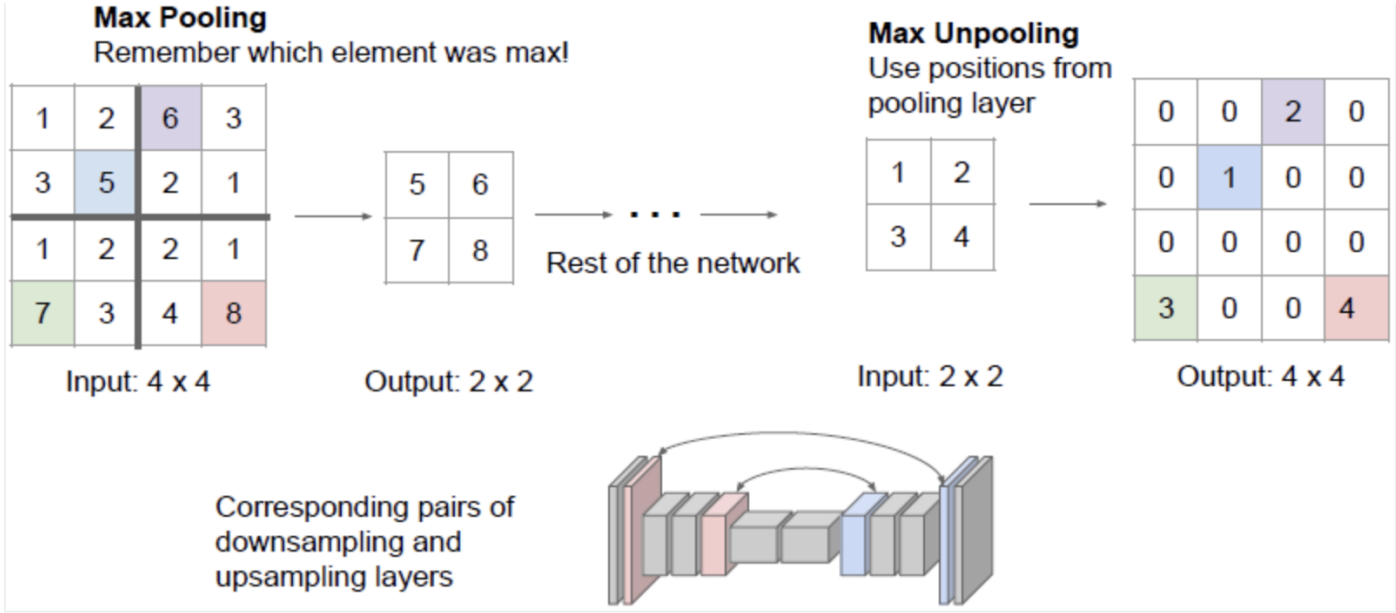
- 연산량 ↓
  - VGG16
    - CNN : parameter sharing
    - Fc 제거 : Encoder에서 classification 필요 x
    - Filter : 작은 필터의 이점

SegNet - Architecture



SegNet – Max unpolling

Max pooling의 경우 Unpooling 할 때 Max pooled된 값의 위치를 알 수 없다.



Max Pooling

- 특정 영역 내의 최대값 선택
- 해당 영역을 대표
- 나머지 정보 버림

→ 해상도 감소하게 되고 객체의 경계 흐려짐.

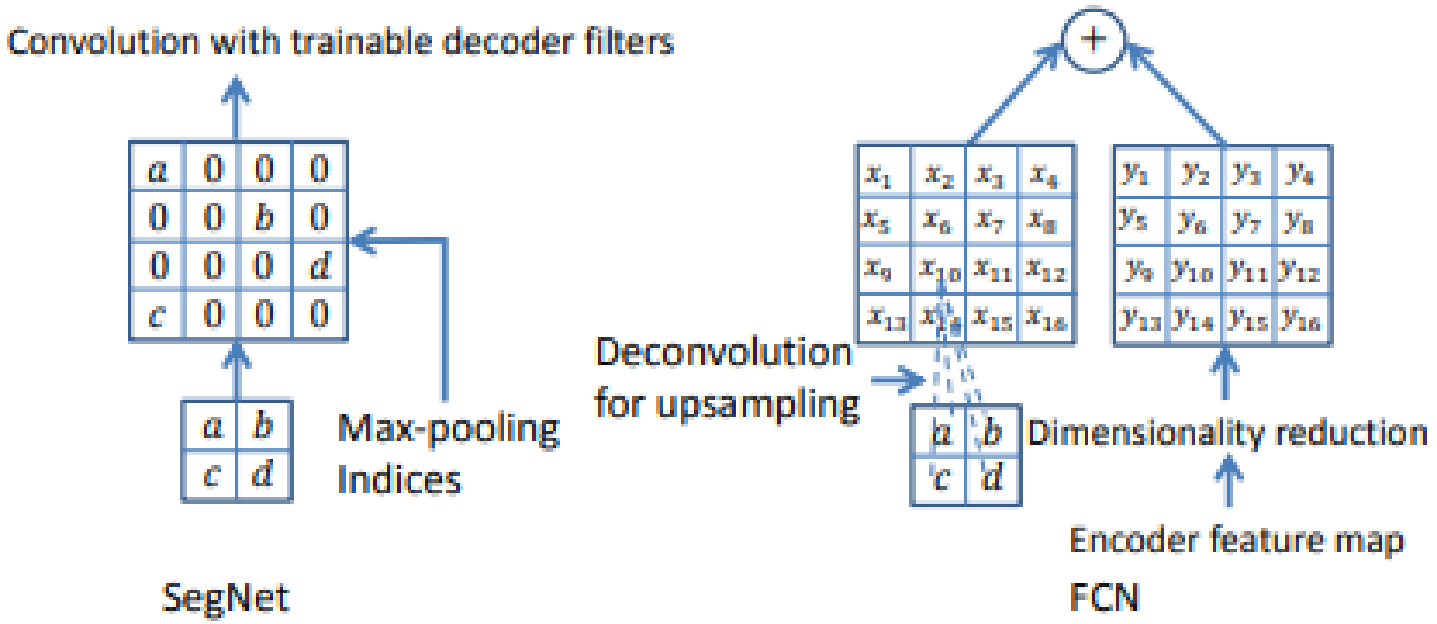
Max UnPooling

- Max Pooling에서 사용된 위치 정보 기억
- UpSampling 과정에서 해당 위치에만 값을 배치
- 나머지는 0으로 채움

→ Max pooling에서 선택된 최대값들의 정확한 위치를 유지하고 원본 이미지의 경계 부근에서 선택되었던 특성들이 그 위치에 정확하게 복원되어 경계가 선명하게 보존

SegNet – Max unpolling

Max-pooling 위치를 저장해 두었다가 이후 Max Unpooling 진행



FCN에서는 Transposed convolution으로 upsampling 진행

SegNet

- 연산량 ↓
  - VGG16
    - CNN : parameter sharing
    - Filter : 작은 필터의 이점
    - Fc 제거 : Encoder에서 classification 필요 x
  - Max Unpooling : parameter 개수 감소

- Segmentation 성능 ↑
  - Max Unpooling : boundary delineation 향상

비교 Architecture

- FCN
- DeepLab
- DeconvNet

사용 데이터

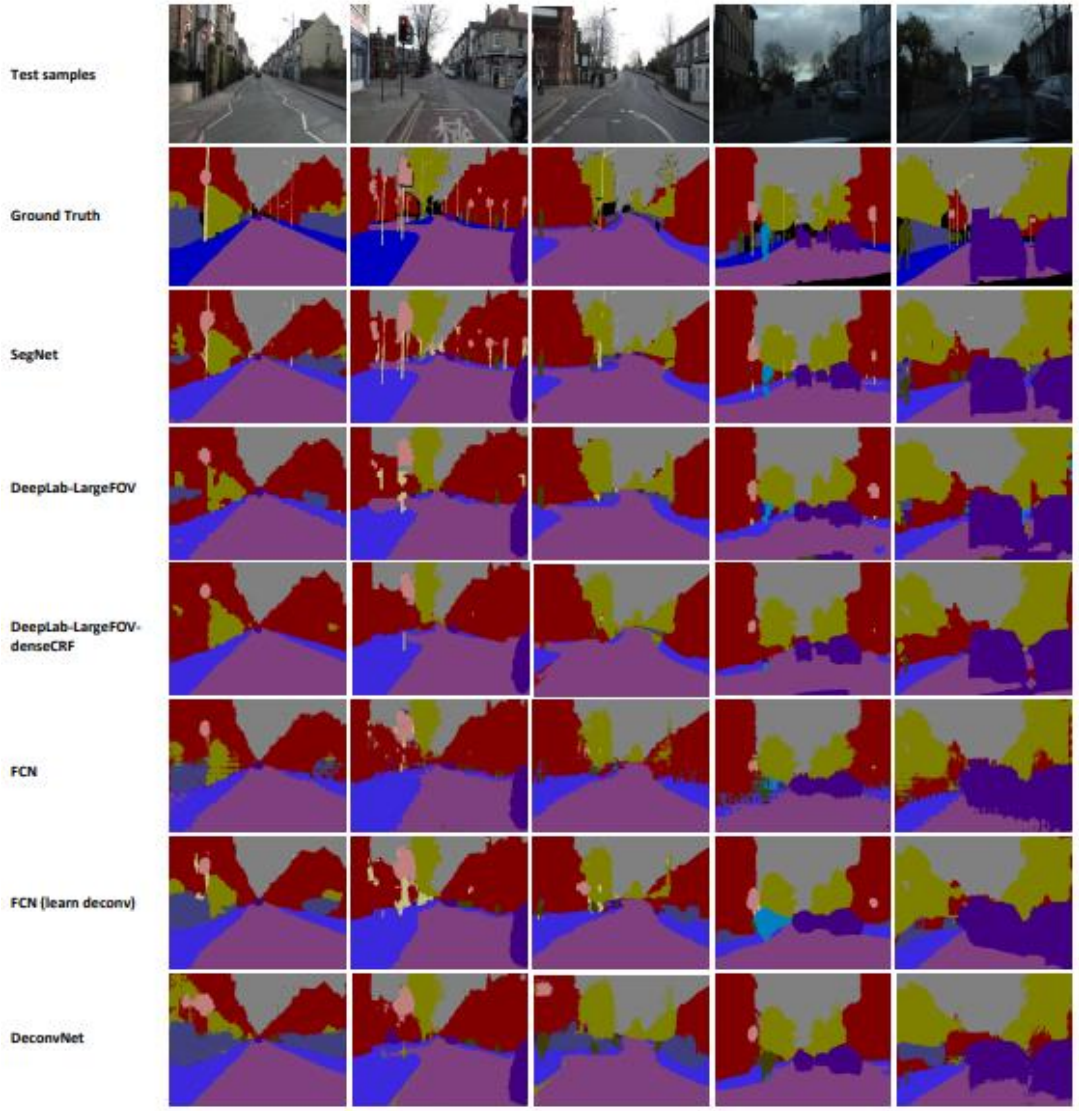
- Cam Vid dataset - road scene segmentation
- SUN RGB-D dataset - indoor scene segmentation

정량적 평가 척도

- Global accuracy (G): 데이터셋 전체의 픽셀 수에서 올바르게 분류된 픽셀의 수
- Class average accuracy (C): 각 클래스마다 accuracy를 계산한 뒤, 평균낸 것
- mIoU
- Boundary F1 Score (BF): 2 * precision * recall / (recall + precision)

Result

Cam Vid dataset

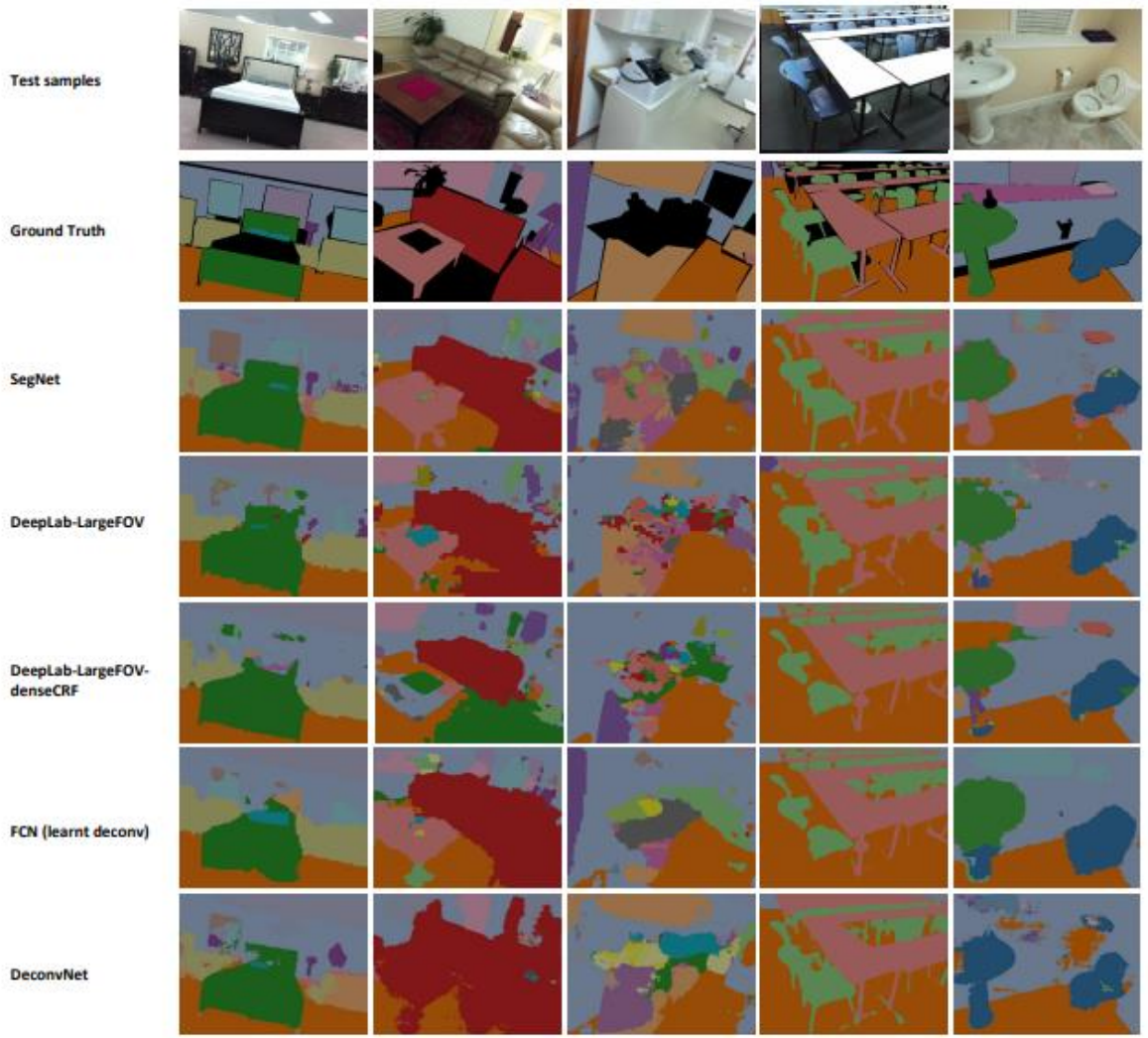


Method	Building	Tree	Sky	Car	Sign-Symbol	Road	Pedestrian	Fence	Column-Pole	Side-walk	Bicyclist	Class avg.	Global avg.	mIoU	BF
SfM+Appearance [28]	46.2	61.9	89.7	68.6	42.9	89.5	53.6	46.6	0.7	60.5	22.5	53.0	69.1	n/a*	
Boosting [29]	61.9	67.3	91.1	71.1	58.5	92.9	49.5	37.6	25.8	77.8	24.7	59.8	76.4	n/a*	
Dense Depth Maps [32]	85.3	57.3	95.4	69.2	46.5	**98.5**	23.8	44.3	22.0	38.1	28.7	55.4	82.1	n/a*	
Structured Random Forests [31]	n/a											51.4	72.5	n/a*	
Neural Decision Forests [64]	n/a											56.1	82.1	n/a*	
Local Label Descriptors [65]	80.7	61.5	88.8	16.4	n/a	98.0	1.09	0.05	4.13	12.4	0.07	36.3	73.6	n/a*	
Super Parsing [33]	87.0	67.1	96.9	62.7	30.1	95.9	14.7	17.9	1.7	70.0	19.4	51.2	83.3	n/a*	
SegNet (3.5K dataset training - 140K)	**89.6**	**83.4**	96.1	**87.7**	52.7	96.4	**62.2**	**53.45**	**32.1**	**93.3**	**36.5**	**71.20**	**90.40**	60.10	46.84
CRF based approaches															
Boosting + pairwise CRF [29]	70.7	70.8	94.7	74.4	55.9	94.1	45.7	37.2	13.0	79.3	23.1	59.9	79.8	n/a*	
Boosting+Higher order [29]	84.5	72.6	**97.5**	72.7	34.1	95.3	34.2	45.7	8.1	77.6	28.5	59.2	83.8	n/a*	
Boosting+Detectors+CRF [30]	81.5	76.6	96.2	78.7	40.2	93.9	43.0	47.6	14.3	81.5	33.9	62.5	83.8	n/a*	

SegNet, DeconvNet이 가장 좋은 성능을 보입니다.

Result

SUN RGB-D dataset



Network/Iterations	40K				80K				>80K				Max iter
	G	C	mIoU	BF	G	C	mIoU	BF	G	C	mIoU	BF	
SegNet	88.81	59.93	50.02	35.78	89.68	69.82	57.18	42.08	90.40	71.20	60.10	46.84	140K
DeepLab-LargeFOV [3]	85.95	60.41	50.18	26.25	87.76	62.57	53.34	32.04	88.20	62.53	53.88	32.77	140K
DeepLab-LargeFOV-denseCRF [3]	not computed								89.71	60.67	54.74	40.79	140K
FCN	81.97	54.38	46.59	22.86	82.71	56.22	47.95	24.76	83.27	59.56	49.83	27.99	200K
FCN (learnt deconv) [2]	83.21	56.05	48.68	27.40	83.71	59.64	50.80	31.01	83.14	64.21	51.96	33.18	160K
DeconvNet [4]	85.26	46.40	39.69	27.36	85.19	54.08	43.74	29.33	89.58	70.24	59.77	52.23	260K

Wall	Floor	Cabinet	Bed	Chair	Sofa	Table	Door	Window	Bookshelf	Picture	Counter	Blinds
83.42	93.43	63.37	73.18	75.92	59.57	64.18	52.50	57.51	42.05	56.17	37.66	40.29
Desk	Shelves	Curtain	Dresser	Pillow	Mirror	Floor mat	Clothes	Ceiling	Books	Fridge	TV	Paper
11.92	11.45	66.56	52.73	43.80	26.30	0.00	34.31	74.11	53.77	29.85	33.76	22.73
Towel	Shower curtain	Box	Whiteboard	Person	Night stand	Toilet	Sink	Lamp	Bathtub	Bag		
19.83	0.03	23.14	60.25	27.27	29.88	76.00	58.10	35.27	48.86	16.76		

SegNet은 다른 모델들과 비교해서 G, C, mIou, BF 모두에서 우수한 성능을 보입니다.