**Fully Convolutional Networks for Semantic Segmentation**

Jonathan Long, Evan Shelhamer, Trevor Darrell (UC Berkeley, 2015)

**1. 연구 배경 (왜 이 논문을 썼는가?)**

* 이미지는 여러 객체들로 구성되어 있어. 이 객체들을 **각 픽셀 단위로 분류**하는 걸 “semantic segmentation(의미 기반 분할)”이라고 해.
* 그전까지의 CNN(예: VGG, AlexNet)은 한 장의 이미지에 대해 하나의 클래스(예: 고양이, 자동차)만 예측했어. 즉, **전체 이미지에 대한 예측만 가능**했지.
* 하지만 우리는 각 픽셀마다 “이건 사람, 이건 배경, 이건 고양이”처럼 정확히 분리해내야 해.
* 그래서 저자들은 CNN을 픽셀 단위로 예측할 수 있게 바꾼 \*\*FCN(Fully Convolutional Network)\*\*을 제안했어!

**2. 주요 내용 및 방법 (어떻게 했는가?)**

**✅ 핵심 아이디어:**

**기존 분류 CNN을 ‘픽셀 단위 예측용 구조’로 바꾼다.**

**1) Fully Convolutional Network (FCN)**

* 기존 CNN은 마지막에 Fully Connected Layer (고정된 크기 입력만 가능)가 있어.
* FCN은 이 FC Layer를 **1×1 Convolution Layer**로 바꿔서, 어떤 크기의 입력 이미지든 처리할 수 있어!
* 이렇게 하면 **한 번에 전체 이미지에 대한 예측 가능** + 학습도 효율적이야.

**2) Deconvolution (Upsampling)**

* CNN 구조는 보통 feature map의 해상도를 줄이면서 진행되잖아? (예: Max Pooling)
* 그래서 나온 결과는 **원래 이미지보다 작고 흐릿해.**
* 이를 해결하려고 \*\*Upsampling(Deconvolution)\*\*을 도입해서 원래 이미지 크기로 복원했어!

**3) Skip Architecture (FCN-32s → FCN-16s → FCN-8s)**

* 낮은 레이어(pool3, pool4)는 \*\*디테일 정보(모양, 위치)\*\*를 잘 담고 있어.
* 높은 레이어(conv7)는 \*\*의미 정보(이게 사람인지, 자동차인지)\*\*를 잘 담고 있어.
* 그래서 두 정보를 **합쳐서** 더 정확하게 예측하는 구조를 만들었어!
  + FCN-32s: 가장 coarse한 결과
  + FCN-16s: pool4 정보 추가 → 더 세밀
  + FCN-8s: pool3까지 추가 → 가장 정밀한 segmentation

**3. 실험 결과**

* **데이터셋**: PASCAL VOC 2011
* **성능 측정 지표**: mean IU (Intersection over Union, 높을수록 좋음)
* 성능 비교:
  + FCN-32s: 59.4%
  + FCN-16s: 62.4%
  + FCN-8s: 62.7%  
    → 점점 더 세밀한 구조일수록 성능 향상!

**4. 논문의 의의 (이 논문이 왜 중요한가?)**

* 이 논문은 "CNN을 segmentation에도 쓸 수 있다!"는 가능성을 처음 보여줌.
* 복잡한 후처리 없이, **end-to-end로 학습 가능**하게 함.
* 이후 등장한 U-Net, DeepLab, PSPNet 등 수많은 모델의 기반이 됨.
* 말 그대로 **semantic segmentation의 시작점이 된 논문**임.