**Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation**

Liang-Chieh Chen et al. (Google, 2018)

DeepLabv3+

**1. 연구 배경**

* DeepLab 시리즈는 semantic segmentation 성능이 우수하지만, 기존 DeepLabv3는 **경계선이 흐릿한 문제**가 있었어.
* 반면에 **encoder-decoder 구조**는 경계는 잘 복원하지만 전체적인 의미 정보가 부족했지.
* 그래서 이 논문에서는 **DeepLabv3의 좋은 점 + decoder의 장점**을 결합해서 DeepLabv3+를 만들었어!

**2. 주요 내용 및 방법**

**✅ 핵심 구조:**

👉 **DeepLabv3 (encoder) + 얇고 효율적인 decoder**

**1) Atrous Spatial Pyramid Pooling (ASPP)**

* 다양한 크기의 receptive field를 가지는 **Atrous Convolution**을 병렬로 사용해서 **멀티스케일 컨텍스트 정보**를 추출하는 구조.
* DeepLabv3에서는 ASPP로 encoder 기능을 강화했어.

**2) Decoder 구조 추가**

* 기존 DeepLabv3는 마지막 feature map을 단순히 bilinear upsampling 했어 (너무 단순).
* DeepLabv3+에서는 \*\*low-level feature(poor4 같은 디테일 정보)\*\*와 high-level feature(ASPP 결과)를 **합쳐서 디코딩**함.
* 이렇게 하면 **경계선이 더 날카롭고 정확해짐**!

**3) Depthwise Separable Convolution**

* 연산량을 줄이기 위해 standard conv를 두 단계로 분리:
  1. **Depthwise convolution** (채널별 필터)
  2. **Pointwise (1x1) convolution** (채널 통합)
* 이 구조를 ASPP와 decoder에 적용해서 **속도 ↑, 성능 유지 or 개선**!

**3. 실험 결과**

* **PASCAL VOC 2012**: 89.0% (SOTA 성능!)
* **Cityscapes**: 82.1%
* **Output stride 8 or 16** 선택 가능 (16은 빠르고, 8은 더 정확)
* **Xception backbone + atrous separable conv** 조합이 가장 강력함

**4. 논문의 의의**

* semantic segmentation에서 **정확도와 경계 품질을 동시에 잡음!**
* **ASPP + decoder + 속도까지 잡은 구조**
* 현재까지도 segmentation 모델 구현할 때 **기본 템플릿처럼 사용됨**