

# 画像セグメンテーション の技術について

# セグメンテーション大まかな分類

---

- ・ Semantic Segmentation(セマンティック セグメンテーション)

塗り絵のように画像上全てのピクセルに対して、クラスカテゴリーをつける。

手法：U-Net（MICCAI 2015）、SegNet（arXiv 2015）

- ・ Instance Segmentation(インスタンス セグメンテーション)

Object detection のような物体の認識をピクセルレベルで行う。Semantic Segmentationと異なり、重なりのある同一物体などを正しく別々に検出する一方、物体候補領域、RoI（region of interest）に対して segmentation を行う。

手法：Mask R-CNN（ICCV2017）、FCIS（ICCV2017）

- ・ Panoptic Segmentation(パノプティック セグメンテーション)

Semantic Segmentation と Instance Segmentation を足し合わせたような手法。

# SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation

(10 Oct 2016) Vijay Badrinarayanan, Alex Kendall, Roberto Cipolla, Senior Member, IEEE,

<https://arxiv.org/pdf/1511.00561.pdf>

## どんなもの？

SegNetは、深層畳み込みネットワークを活用した手法で、その当時競合する手法よりも大幅に少ないパラメータ数で学習可能。最急降下法を利用しend-to-end で学習ができる。また、構造はエンコーダとそれに対応するデコーダがあり最後に分類用のレイヤーがある。

## どうやって有効だと検証した？

道路のシーンのセグメンテーション実験とSUN RGB-Dという室内のデータセットによる実験をしている。他の比較を行う手法としては、FCNやDeconvNetなど。評価指標はGT: (Ground Truth)PR: (Prediction)TP: (True Positive)からIoU =  $TP / (GT + PR - TP)$ で求めている。

## 技術の手法や肝は？

ネットワークの構成に畳み込み層のみ利用し、エンコーダとデコーダという構成にしている。最終的には、end-to-end

## 議論はある？

SegNetは自動運転技術やARアプリケーションなどのリアルタイムへの利用を想定していた為、そのため今までの手法よりも計算の高速化することに注目し作られている。

## 先行研究と比べて何がすごい？

当時のセグメンテーション手法は全結合での学習を行っていたため、パラメータ数など膨大になっていたが、Segnetでは、全結合は利用せず、畳み込みネットワークのみを利用しているため、パラメータの数が大幅に減っている点。

## 次に読むべき論文は？

同じセマンティックセグメンテーション手法であるU-net

<https://arxiv.org/pdf/1505.04597.pdf>

# U-Net: Convolutional Networks for Biomedical Image Segmentation

(18 Aug 2015) Olaf Ronneberger, Philipp Fischer, and Thomas Brox

<https://arxiv.org/abs/1505.04597>

## どんなもの？

U-netは少ないトレーニングデータで学習ができISBI2015のチャレンジにおいても他に大きな差をつけて圧勝した。

## どうやって有効だと検証した？

3つの異なるセグメンテーションのタスクで検証。タスクは、the EM segmentation challenge、the ISBI cell tracking challenge 2015のデータで検証。  
評価指標はIOUを使い

## 技術の手法や肝は？

U-netは、FCNから着想を得て構築されている。FCNとの違いは、endoder(downsampling)とDecoder(upsampling)が対称的に行うようにしている。  
upsamplingの際は対称となるdownsamplingのパスから切り出した画像を足し合わせている。



## 先行研究と比べて何がすごい？

当時の先行の手法だと、膨大なデータを用意し学習に使用しなければ行けなかったが、U-netは少ないデータでend-to-endの学習ができる

## 次に読むべき論文は？

PSPnet : <https://arxiv.org/abs/1612.01105>

**Encoder**と**Decoder**の間に**Pyramid Pooling Module**を追加している

# Mask R-CNN

(24 jan 2018) Kaiming He Georgia Gkioxari Piotr Dollar Ross Girshick

<https://arxiv.org/pdf/1703.06870.pdf>

## どんなもの？

Mask R-CNNはFaster R-CNNを元に物体検出に加えSegmentation maskを生成できる手法。Faster R-CNNの2ステージによる学習の2ステージ目にsegmentation maskを行う分類器を追加することで実現している。また、さらなる拡張として人のポーズも読み取ることも可能。

## どうやって有効だと検証した？

COCOデータセットをを利用し、the COCO 2015 and 2016 segmentation challengesの勝者であるMNCとFCISと比較した結果、gAP (averaged over IoU thresholds), AP50, AP75, APS, APM, APLの全ての評価において2手法を上回る結果となっている。\*APは平均適合率

## 技術の手法や肝は？

U-netは、faster R-CNNのモデルを参考にHead部分にを追加するとともに、Roi\_PoolingではなくRoi\_Alignを利用している。これは物体検出ではなく、Segmentationではpixel-to-pixelの対応が重要になる為。

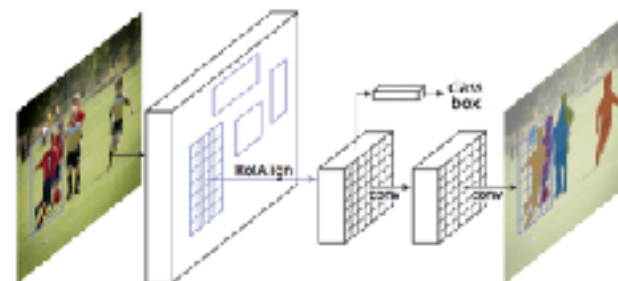


Figure 1. The Mask R-CNN framework for instance segmentation.



## 先行研究と比べて何がすごい？

Segmentation maskを生成しているにも関わらず、ベースのFaster R-CNNと比べても、同程度の処理速度。

## 次に読むべき論文は？

Mask Scoring R-CNN : <https://arxiv.org/pdf/1903.00241.pdf>

# Panoptic Segmentation

(10 Apr 2019) Alexander Kirillov Kaiming He Ross Girshick Carsten Rother Piotr Dollar

<https://arxiv.org/pdf/1801.00868.pdf>

## どんなもの？

Panoptic Segmentationは、Semantic Segmentationと Instance Segmentationを統合したものである。画像内の全てのピクセルに対してラベリングを行う。Stuffとthingsと故障するセグメンテーションを行なっている。StuffではSemantic Segmentationから得られる数えられないまとまりのないものを示している。Thingsについては物体検知もしくはInstance Segmentationから得られる

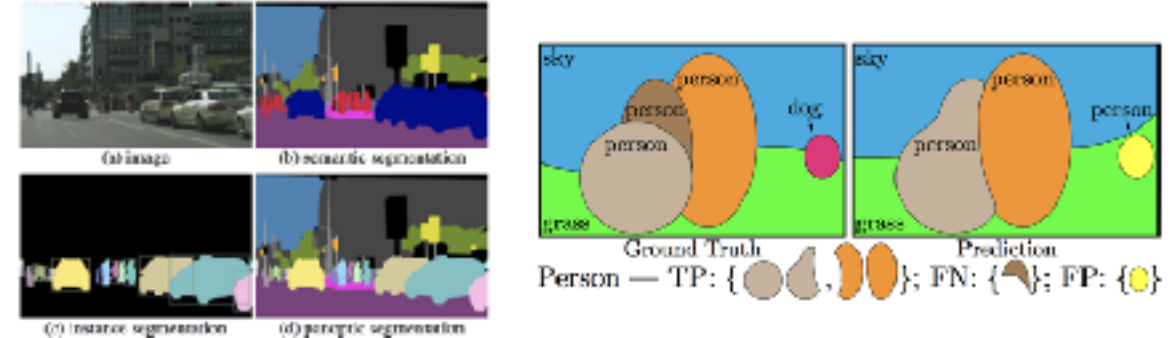
## どうやって有効だと検証した？

既存の Instance Segmentationと Semantic Segmentationのネットワークを使用し、Cityscapes, ADE20k, Vistas データセットでの評価。PanopticQuality (PQ) という新しい評価指標を使い検証。

$$PQ = \underbrace{\frac{\sum_{(p,g) \in TP} \text{IoU}(p,g)}{|TP|}}_{\text{segmentation quality (SQ)}} \times \underbrace{\frac{|TP|}{|TP| + \frac{1}{2}|FP| + \frac{1}{2}|FN|}}_{\text{recognition quality (RQ)}}.$$

## 技術の手法や肝は？

Segmentationの2つの手法を組み合わせと、このモデルのための新しい評価指標の活用



## 先行研究と比べて何がすごい？

場蔵内の全てのピクセルに対する、ラベリングしセグメンテーションを行なっている点

## 次に読むべき論文は？

Panoptic Feature Pyramid Networks : <https://arxiv.org/pdf/1901.02446.pdf>

# UPNet: A Unified Panoptic Segmentation Network

(3 Apr 2019) Yuwen Xiong Renjie Liao Hengshuang Zhao Rui Hu1 Min Bai Ersin Yumer Raquel Urtasun

Uber ATG University of Toronto The Chinese University of Hong Kong

<https://arxiv.org/pdf/1901.03784.pdf>

## どんなもの？

UPNetでは、Panoptic Segmentationでの課題の解決を提案している。  
単一のバックボーンでend-to-end学習する。

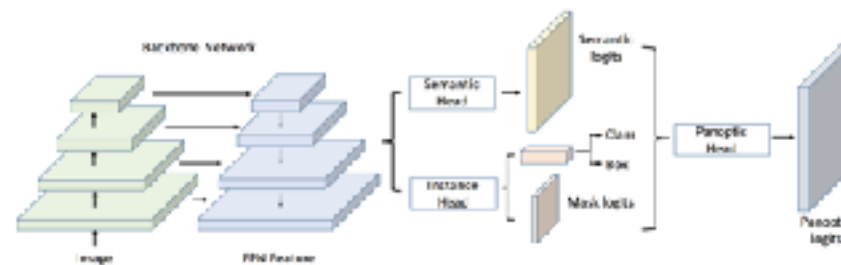


Figure 1: Overall architecture of our UPNet.

## どうやって有効だと検証した？

COCOとCityscapesで既存の手法より高速で高精度だった。

## 先行研究と比べて何がすごい？

PanopticSegmentationでは、各ピクセルに対し固有の答えを出す必要があったのに対し

## 技術の手法や肝は？

InstanceとSemanticのヘッドで抽出した特徴からPanopticHeadを使って結果をマージする手法を提案。

ResNet50 FPNをバックボーンとして使用している。

## 次に読むべき論文は？

Panoptic Feature Pyramid Networks : <https://arxiv.org/pdf/1901.02446.pdf>