

Fully Convolutional Networks for Semantic Segmentation

2015

Jonathan Long/Evan Shelhamer/Trevor Darrell

<https://arxiv.org/pdf/1411.4038.pdf>

どんなもの？

- ・ FCNは、畳み込み層のみのネットワーク構造を用いてセマンティックセグメンテーションを行う手法
- ・ AlexNetやVGGの全結合層を畳み込み層に置き換えることで画像をピクセル単位で推定

議論はある？

- ・ なし

どうやって有効だと検証した？

- ・ PASCAL VOC 2011および2012のテストセットでFCN、SDS、R-CNNを比較して20%ほどの時間短縮を実現した

先行研究と比べて何がすごい？

- ・ 全結合層を無くすことで、従来のCNNのように入力画像のサイズを固定する制約がなくなった。
- ・ 全結合層を畳み込み層に置き換えると、クラス分類の結果がヒートマップとして出力されるようになる

技術の手法や肝は？

- ・ CNNの最後の全結合層を畳み込み層に置き換えることで出力を分類クラスではなく二次元マップに変換
- ・ 畳み込み層で得られた特徴マップをそのままセマンティックセグメンテーションの結果として使用できる

次に読むべき論文は？

U-Net: Convolutional Networks for Biomedical Image Segmentation

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

U-Net: Convolutional Networks for Biomedical Image Segmentation

MICCAI 2015

Olaf Ronneberger / Philipp Fischer / Thomas Brox

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

どんなもの？

- ・ FCN(fully convolution network)の1つであり、画像のセグメンテーション（物体がどこにあるか）を推定するためのネットワーク
- ・ 畳み込みのみで表現されたネットワークで全結合層がない
- ・ モデルの構造が左右対象でアルファベットの「U」に似ていることから、「U-net」と呼ばれる

どうやって有効だと検証した？

- ・ ISBIチャレンジのセグメンテーションタスクにおいて、FCNの考え方をベースに拡張することでSOTAを更新した

技術の手法や肝は？

- ・ 画像全体のクラス分類からFCNとdeconvolutionを使い、物体の位置情報を出力できるようになった
- ・ ダウンサンプリングからアップサンプリングへのスキップアーキテクチャーによってローカル特徴を保持している

議論はある？

- ・ なし

先行研究と比べて何がすごい？

- ・ U-Netでは少ないサンプル画像をベースに学習が可能になった

次に読むべき論文は？

SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation
<https://arxiv.org/pdf/1411.4038.pdf>

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

IEEE 2015

Vijay Badrinarayanan / Alex Kendall / Roberto Cipolla / Senior Member

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

どんなもの？

- ・ SegNet はセマンティックセグメンテーションのための Encoder と Decoder のネットワーク構成を有するモデル

議論はある？

- ・ セグメンテーションの精度と高速化、省メモリ化はトレードオフの関係

どうやって有効だと検証した？

- ・ Caffeで実装したモデルで検証（FCN、LargeFOV、DeconvNetと比較）
- ・ 従来の手法（FCN、DeconvNet）よりもセグメンテーションの精度が高く学習時間やメモリの消費量も小さい。（ただし、LargeFOVよりは学習時間やメモリの消費量が多い）

先行研究と比べて何がすごい？

- ・ 初期の Encoder と Decoder のネットワーク構成を用いたモデル
- ・ Decoder（アップサンプリング）が単純なため、Encoder 部分を変えて応用が可能

技術の手法や肝は？

- ・ SegNet は Encoder（ダウンサンプリング）と Decoder（アップサンプリング）の左右対称なネットワークとそれに続くピクセル単位の分類レイヤーで構成されている
- ・ SegNet の特徴はアップサンプリングの際に、対応する Encoder のMAXプーリングステップで計算したプーリングインデックスを使用して、非線形アップサンプリングを実行していること → アップサンプリング時の学習が不要

次に読むべき論文は？

Mask R-CNN

<https://arxiv.org/pdf/1411.4038.pdf>

Pyramid Scene Parsing Network

CVPR 2017

Hengshuang Zhao / Jianping Shi / Xiaojuan Qi / Xiaogang Wang / Jiaya Jia

<https://arxiv.org/pdf/1612.01105.pdf>

どんなもの？

- ・ PSPNetは、CVPR2017で発表されたセマンティックセグメンテーションの手法
- ・ エンコーダーとデコーダーの間にPyramid Pooling Moduleを追加したモデル

議論はある？

- ・ なし

どうやって有効だと検証した？

- ・ 「ImageNet scene parsing challenge 2016」 「PASCAL VOC 2012 semantic segmentation」 「urban scene understanding dataset Cityscapes」 のデータセットで評価

先行研究と比べて何がすごい？

- ・ 異なる大きさのsub-regionごとの特徴を大域的なコンテキストを考慮するために活用することで、 ImageNet 2017 Scene Parsing Challengeで優勝したモデル
- ・ クラス間のサンプル数の大小だけでなく、 難しいクラスと簡単なクラスがある場合にそれらを考慮した損失関数を使うことで性能をあげた

技術の手法や肝は？

- ・ PSPNetはEncoderに学習済のResNetのモデルを利用しており、EncoderとDecoderの間に複数の解像度でプーリングを行うPyramid Pooling Moduleを追加している

次に読むべき論文は？

Mask R-CNN

<https://arxiv.org/pdf/1411.4038.pdf>

Mask R-CNN

2015

Jonathan Long/Evan Shelhamer/Trevor Darrell

<https://arxiv.org/pdf/1411.4038.pdf>

どんなもの？

- ・ Mask R-CNNは、一般物体検出(Generic Object Detection)と Instance Segmentationを同時に行うマルチタスクの手法
- ・ 2017年に開催されたコンピュータビジョン分野のトップカンファレンス「ICCV2017」でBest Paper Awardを受賞した

議論はある？

- ・ 議論なし

どうやって有効だと検証した？

- ・ この論文では、Mask R-CNNの基本設計の有用性を示すために、画像から特徴抽出を行うCNN(論文ではbackbone networkと呼んでいる)に複数種類のネットワーク(ResNet, ResNeXt, FPN)を試して結果を比較したり、違うデータセットを使っての比較、別のタスクへの応用例として人体の姿勢推定(OpenPoseのような骨格検出)の実装を示している

先行研究と比べて何がすごい？

- ・ Mask R-CNNでは以下の3種類の出力が可能
 - 物体クラス(カテゴリー)の分類
 - Bounding Boxの回帰
 - 物体領域マスクの推定(pixelの分類)

技術の手法や肝は？

- ・ Mask R-CNNのネットワークではSegmentationのマスクを推定するための分岐が追加され、特徴マップの各RoI(Region of Interest : 関心領域)を固定サイズのベクトルに収めるRoI Pooling処理の代わりに、それを改良したRoI Alignという新しい手法が導入されている

次に読むべき論文は？

YOLOv3: An Incremental Improvement
<https://arxiv.org/pdf/1804.02767.pdf>