

輪講資料 7.22(水)

IHPC M1 Zhu hanyu シュカンウ

輪講論文: **Mask R-CNN**

Published in: IEEE Transactions on Pattern Analysis and Machine Intelligence (Volume:42, Issue:2, Feb.1 2020)

Kaiming He

Georgia Gkioxari

Piotr Dollar

Ross Girshick

序言 **Abstract**

この論文には、Mask R-CNN という object instance segmentation(対象事例分割)の手段が提出された。

Mask R-CNNは既存のR-CNNに基づく、新しい分岐予測を使用し、本来のbounding boxのと並行する。

特性：

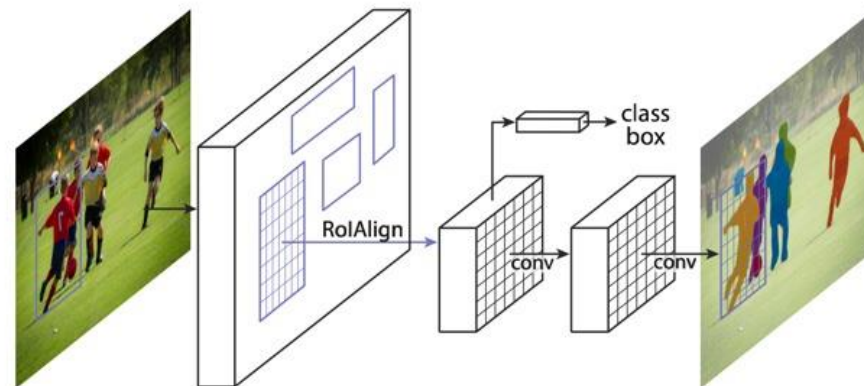
- ・非常に使いやすく、良い柔軟性と汎用性を持っている。
- ・モデルの訓練は簡単で、R-CNNをベースにして、少しだけコストを加えた。
- ・処理のフレームは5fpsである。
- ・事例分割、枠の検出、人の要点の検出を含んだ。
- ・互換性がいい、COCOでも最適な表現がある。
(COCO 2016 challenge winner)

紹介 Introduction

Mask R-CNN: R-CNN

+ 新しい分岐 >>

- 分割されたマスクの分岐予測
- **FCN** 全畳み込みニューラルネットワーク
- 各**RoI**域に応用される (**Region of Interest**)
- 画素一画素の方式です



Mask R-CNNは**Faster R-CNN**を易く実現し、訓練できて、柔軟な構造設計と拡張にとっても良い。

Mask R-CNN は Faster R-CNNの直観的な拡張である。しかし、Faster R-CNNは、ネットワークの入力と出力の間に、画素よりの校正がないので、RoIPoolの効果は不十分である。

具体的には、RoIPoolは事例処理の核心となるが、特徴抽出に対して空間の定量化が粗末である。例えば、校正の精度。このため、

RoIAlign が提出された。

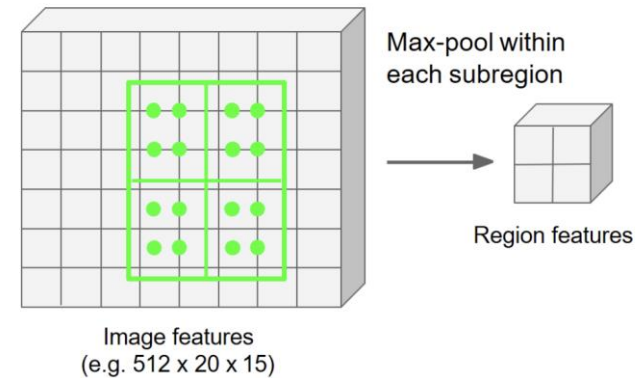
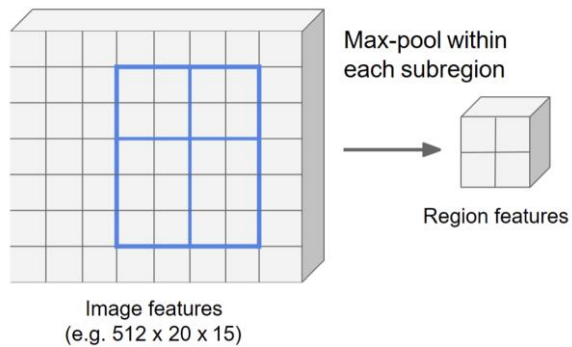
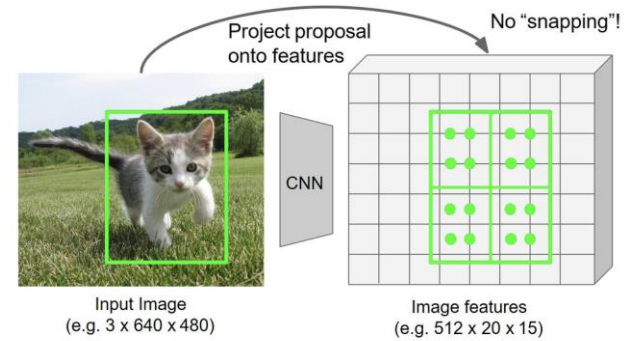
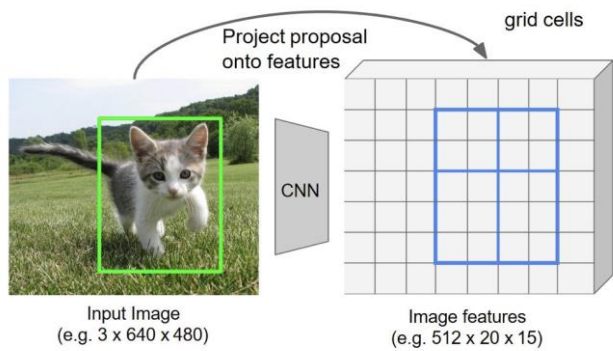
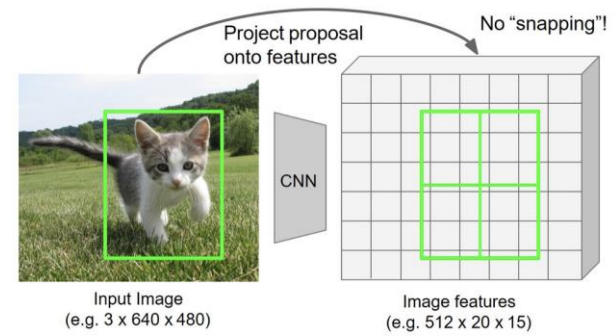
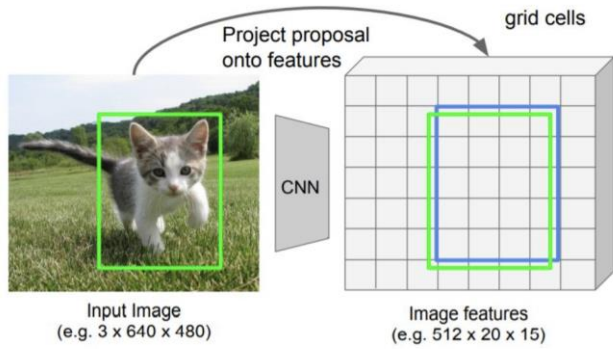
Faster R-CNNの RoIPool

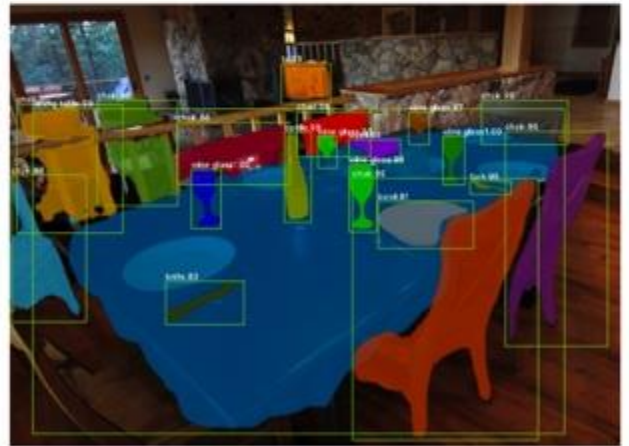
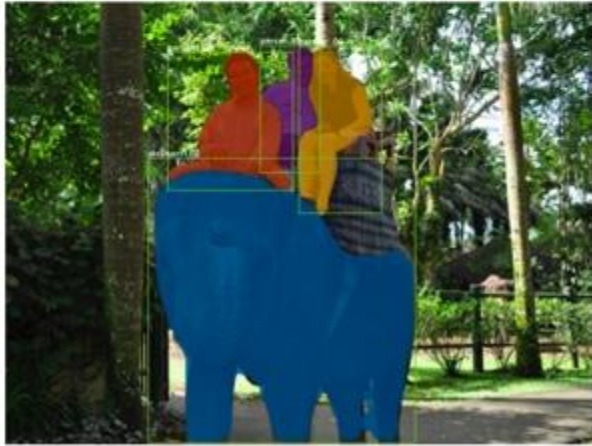
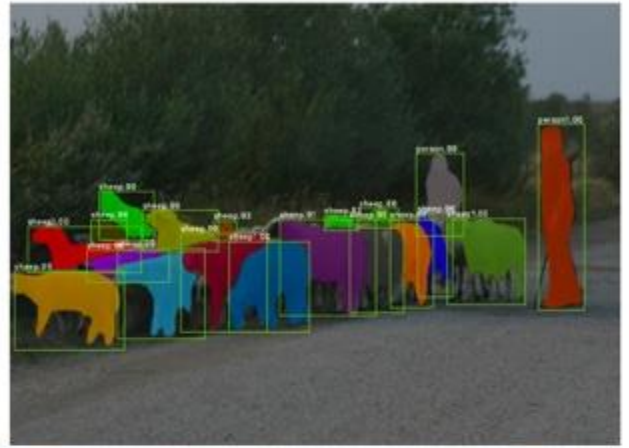
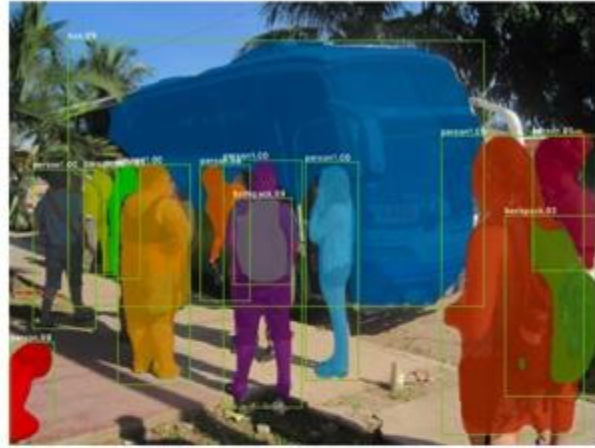
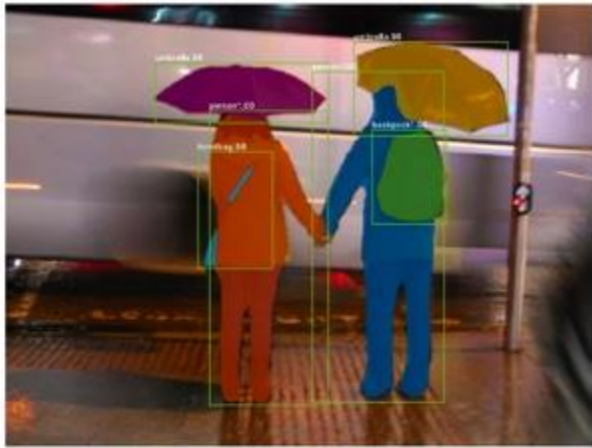
- ・ グリッドユニットに整列する
- ・ グリッドユニットを
7 x 7の独立のユニットに分ける
- ・ max pooling

Mask R-CNNのRoIAlign

- ・ 各区域を巡り、そのまま定量化をしない
- ・ 各区域に4つサンプリング点を選択する
- ・ 各点の画素を計算する
- ・ max pooling

RoIAlignは10% – 50%の精度が向上できて、classの比較も要らない。





COCOで実行した結果である。効果良い

Mask R-CNN

ネットワークの構造

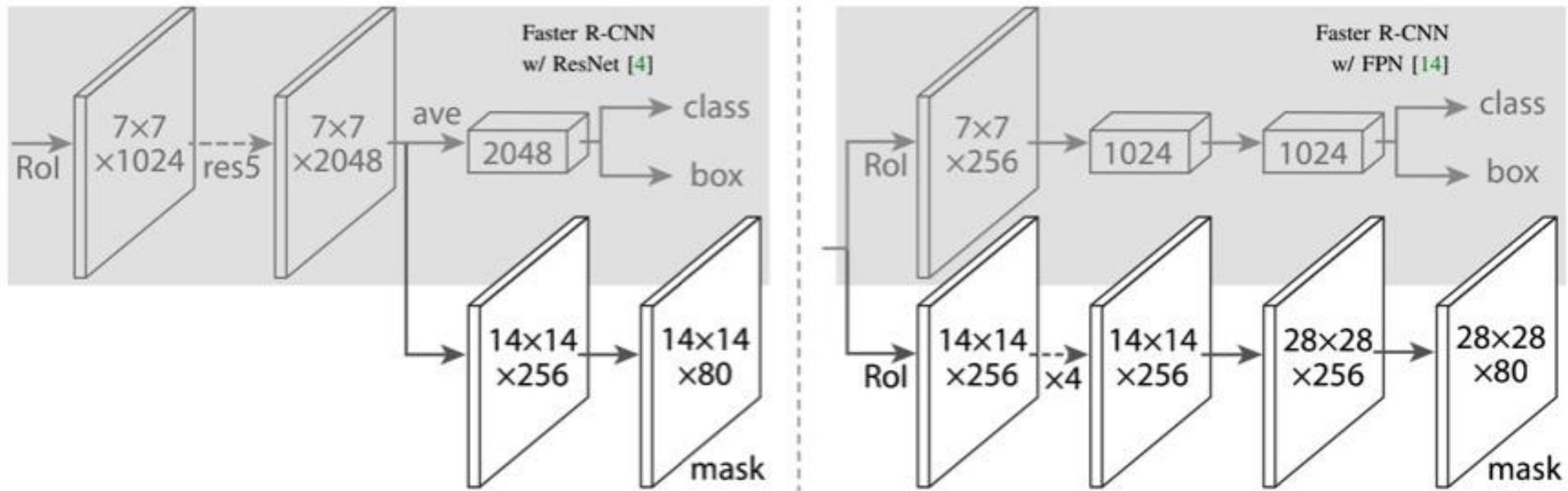
- ・ backbone 基幹 – 畳み込みニューラルの仕組みで、全部のイメージの特徴が抽出できる。
- ・ head – ネットワークで、識別分類、bounding-box回帰、マスク予測ができる。

ResNetのもとで、

- ・ 深層の数が50や101
- ・ 特徴抽出は最後の4番目のCNNを用いてC4と呼ばれる。

FPNのもとで、

- ・ feature pyramid network – 特徴ピラミッド構造のネットワーク
- ・ top-downの方式で、上位層に得られたfeature mapをサンプリングし、下位に転送する。
これから、下位層の特徴も上位の情報が共有できることである。



既存のR-CNNを2つ拡張した。左/右パネルにはResNet C4とFPNのbackboneが表示されている。1つのマスク分岐が加えられた。数字は空間の画素とチャンネルを表す。矢印は、conv、deconv、または上下位から推定される全畳み込みニューラル層を表す(convは空間的次元を保持し、deconvは空間的次元を増加させる)。すべてのconvsは 3×3 であり、でも出力convは 1×1 , deconvsは 2×2 である。隠れ層に活性化関数ReLUを用いる。左の「res5」はResNetの5番目の段階を表しており、簡単のため1番目のconvに 7×7 RoIを使用するように修正した。右: 「 $\times 4$ 」は4つの連続してスタックを表す。

処理の詳細 Implementation Details

訓練の時は、

- IoUが0.5以上であると、RoIはPLUSになる。この場合、マスクの loss L_{mask} は定義される。
- マスクの目標はRoIと測定したマスクの共同部分である。
- イメージとしての訓練、画素は800で
- 一つGPUは2つイメージが対応して、各イメージはN個RoIがあり（PLUS,MINUSの比率が1：3）
- 8GPUを使って、160 k epoch、learning rateは0.2、重みdecayは0.0001、モメンタムは0.9

テストの時は、

- ・ マスク分岐は最優の100個検出ボックスにしか応用しない。
- ・ RoIごとのマスク分岐はK個のマスクを予測できるが、K番目のマスクのみを用いる。

ここで、Kが分岐予測を分類するクラスである。

- ・ $m \times m$ のマスク出力サイズをRoIサイズに調整し、0.5の閾値で二値化する

	backbone	AP	AP ₅₀	AP ₇₅	AP _S	AP _M	AP _L
MNC [23]	ResNet-101-C4	24.6	44.3	24.8	4.7	25.9	43.6
FCIS [24] +OHEM	ResNet-101-C5-dilated	29.2	49.5	-	7.1	31.3	50.0
FCIS+++ [24] +OHEM	ResNet-101-C5-dilated	33.6	54.5	-	-	-	-
Mask R-CNN	ResNet-101-C4	33.1	54.9	34.8	12.1	35.6	51.1
Mask R-CNN	ResNet-101-FPN	35.7	58.0	37.8	15.5	38.1	52.4
Mask R-CNN	ResNeXt-101-FPN	37.1	60.0	39.4	16.9	39.9	53.5

- ・ インスタンス・セグメンテーション結果の一例
- ・ テストはマルチスケール、水平反転などを含んだ。

・ APはIoU閾値を超えた平均精度である。数値範囲がある。例えば、AP50:閾値-0.5

結論 Conclusion

この論文には、ROI Poolの不足を説明して、効果的なセグメンテーションのフレームワークを提案した。Mask R-CNNを既存のR-CNNと比較し、前者よりの改良を詳しく紹介した。

資料を調べ、この論文に通じて、Mask R-CNNの構造や利点を理解してみる。今実装環境もしてみるところである。

以上 どうぞよろしくお願ひ致します。