H-031

受容野の範囲を制限可能な deformable convolution による semantic segmentation

Semantic segmentation with deformable convolution that can limit the range of receptive fields

折田 汐凪 † Sena Orita
荒井 秀一 † Shuichi Arai

1. はじめに

画像からシーンを理解することは自動運転において重要な要素である [1]. このシーン理解を実現する技術として semantic segmentation が期待されている. semantic segmentation は, 画像をピクセル単位でクラス分類することで物体を輪郭ごとに分割する技術である. これにより物体の詳細な形状まで認識できる.

2. 従来手法

Semantic segmentation の手法として, CNN を用いた手法が多く提案されている. しかし, CNN の畳み込みは固定形状のフィルタカーネルでサンプリングする手法であるため, 様々な形状, 大きさを持つ物体の特徴を表現しきれない. そこでフィルタカーネルの形状を入力に応じて可変にする “Deformable convolution”[2] が提案され, その後, deformable convolution を semantic segmentation の代表的な手法である High Resolution Network(HRNet)[3] に適用した DHRNet(Deform-conv HRNet)[4] が提案された.

2.1.Deformable convolution

Deformable convolution は, サンプリング位置からの変位である “offset” を推定することでフィルタカーネルのサンプリング位置を動的に変化させる. 畳み込みのサンプリング位置 p_n の集合を R, 中心位置を p_0, 重みを w, 入力特徴マップを x, 出力特徴マップを y としたとき, 畳み込みは式 (1) と表せる.

$$y(p_0) = \sum_{p_n \in R} w(p_n) x(p_0 + p_n) \tag{1}$$

Deformable convolution では offsetΔp が加わり, 式 (2) と表せる.

$$y(p_0) = \sum_{p_n \in R} w(p_n) x(p_0 + p_n + \Delta p) \tag{2}$$

ここで, offset は deformable convolution の入力特徴マップを図 1 の黒枠で示す CNN 部分の 3×3 畳み込みで推定する.

2.2.DHRNet

DHRNet は異なる解像度を処理するネットワークを組み合わせた 4 つのステージで構成されており, deformable convolution をネットワークの低解像度部分に導入している [4]. ここで, deformable convolution により変化したフィルタカーネルの位置を可視化したところ, 図 3 の赤い矢印で示す offset が明らかに大きいものが存在した.

†東京都市大学 総合理工学研究科

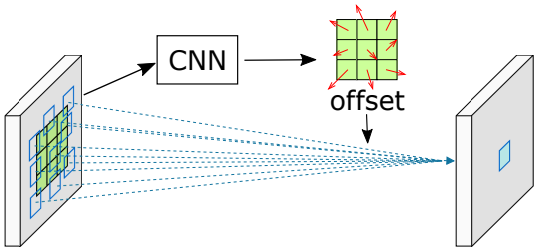


図 1: Deformable Convolution の流れ [2]

図 2: DHRNet の offset の可視化

3. 研究目的

CNN は受容野の範囲内の特徴を畳み込むことで特徴を抽出する. そして, 受容野の範囲は畳み込みを繰り返すことにより広がっていく. ここで, DHRNet における deformable convolution を導入した位置での受容野の範囲は 13x13 である. しかし, 図 3 で示した offset の範囲は, 緑枠で示す受容野の範囲の収まっていなかった. offset は CNN により推定されるため, 推定された offset

図 3: offset 受容野の範囲に収まっていない例

により移動するサンプリング位置が CNN の受容野外を示すべきでないと考えた. そこで推定する offset を CNN の受容野内に制限することで, より適切に offset を学習することを研究目的とする.

4. 提案

本稿では，offset の範囲を受容野の範囲に制限する hard tanh6 関数を deformable convolution に導入することを提案する．hard tanh6 関数は offset の値を-6 以上 6 以下に制限する関数であり，入力の特徴マップを x とすると式 (3) のように表せる．

$$hardtanh6(x) = \begin{cases} 6(x > 6) \\ x(-6 <= x <= 6) \\ -6(x < -6) \end{cases} \quad (3)$$

この hard tanh6 関数を図 4 の黒枠で示す CNN の後に導入する．

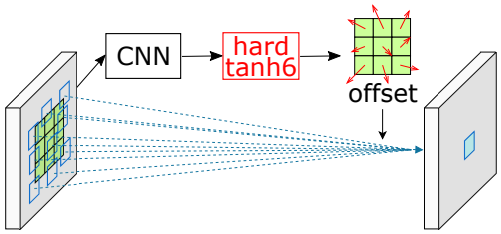


図 4: hard tanh6 を導入した Deformable Convolution

5. 実験及び実験結果

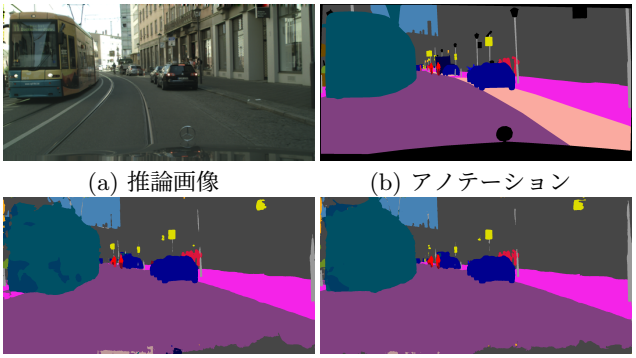
実験には，Cityscapes[1] データセットを使用した．学習用に 2975 枚，検証用に 500 枚，テスト用に 1525 枚用意されており，評価対象のクラスは 19 クラスである．本稿では学習用 2975 枚で学習し，検証用 500 枚で推論を行った．提案手法の有効性を示すために，semantic segmentation で主に用いられている評価指標である MIoU(Mean Intersection over Union) で評価した．MIoU による評価結果を表 1 に示す．表 1 より，提案手法は従来手法と比較して MIoU が向上している．

表 1: 従来手法と提案手法の比較

model	MIoU
DHRNet	73.52
提案手法 (DHRNet+hard tanh6)	73.85 (+0.33)

ここで，提案手法における offset の範囲を確認したところ，図 5 に示すようにすべての offset が受容野の範囲に収まっていることが確認できた．次に，提案手法と従来の DHRNet とのセグメンテーション結果を図 6 に示す．図 6 より，従来手法の DHRNet では左上の train の内部に他クラスが混在しているのに対し，提案手法では train の内部に他クラスが混在しなくなったことが確認できる．

図 5: offset が受容野の範囲に収まっている例



(a) 推論画像　(b) アノテーション

(c) 従来手法 (DHRNet)　(d) 提案手法

図 6: セグメンテーション結果

6. おわりに

本稿では，offset が CNN の受容野の範囲外を指すことは不適であると考え，deformable convolution に offset の範囲を制限する機構が存在しないことを問題点とした．そこで，入力の値を-6 以上，6 以下に制限する hard tanh6 関数を offset の推定部分に導入することを提案した．そして，MIoU の評価指標による従来手法との比較から，提案手法の有効性を確認した．

参考文献

[1] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3213–3223, 2016.

[2] Jifeng Dai, Haozhi Qi, Yuwen Xiong, Yi Li, Guodong Zhang, Han Hu, and Yichen Wei. Deformable convolutional networks. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

[3] Jingdong Wang, Ke Sun, Tianheng Cheng, Borui Jiang, Chaorui Deng, Yang Zhao, Dong Liu, Yadong Mu, Mingkui Tan, Xinggang Wang, Wenyu Liu, and Bin Xiao. Deep high-resolution representation learning for visual recognition. *TPAMI*, 2019.

[4] Daiki Ando and Shuichi Arai. Semantic segmentation using hrnet with deform-conv for feature extraction dependent on object shape. In *2021 3rd International Conference on Cybernetics and Intelligent System (ICORIS)*, pp. 1–5, 2021.