

進捗報告

1 今週やったこと

先週に引き続き Semantic Segmentation 関連の論文を読んでまとめた。

2 DeepLab [1, 2]

Google の出しているオープンソースの Semantic Segmentation のためのモデル。図 1 に DeepLab v3+ のモデル概要を示す。

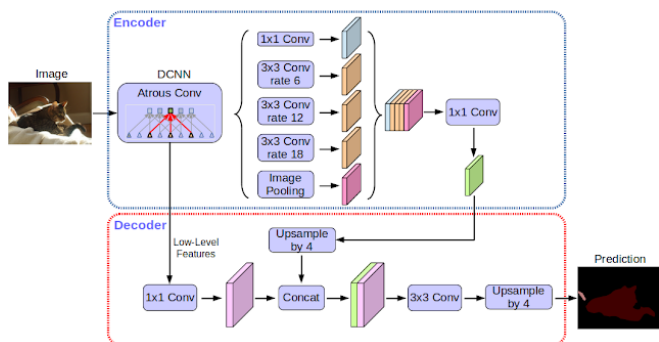


図 1: DeepLab モデル概要図

現在は DeepLab v3+ が最新バージョン。DeepLab には 3 つの特徴が紹介されている。1 つ目が atrous convolution, 2 つ目が atrous spatial pyramid pooling(ASPP), 3 つ目が CRF による条件付き確率場の利用である。基本的には DCNN と CRF を組み合わせたモデルになっており、DCNN では VGG16 や ResNet を fine tuning したものをを用いている。論文では ResNet を用いた時のほうが精度が良かったとある。

3 atrous convolution

拡張畳み込みとも呼ばれる、深い畳み込みニューラルネットワークによって計算された特徴の解像度を明示的に制御し、マルチスケール情報をキャプチャするためにフィルタの視野を調整することを可能にする畳み込み方式。レート r が存在し、 $r = 1$ のときは標準

的な畳み込み演算に一致する。特に 2 次元の場合、以下のように定式化される。また、図 2 に概要図を示す。

$$y[i] = \sum_k x[i + r * k]w[k] \quad (1)$$

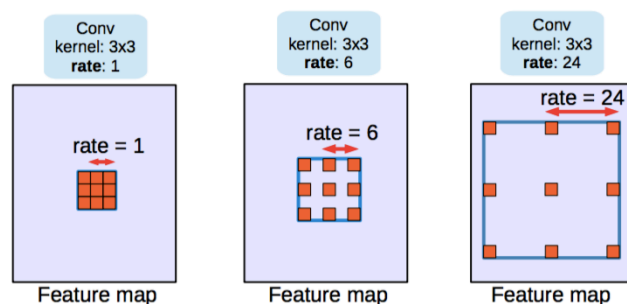


図 2: atrous 畳み込み概要図

4 atrous spatial pyramid pooling(ASPP)

scale invariance を考慮するために atrous 畳み込みのレート r を様々に設定する方法。図 3 に ASPP の概要図を示す。様々なレートで畳みこまれたレイヤは重ね合わせて 1×1 の畳み込み演算によって 1 枚のレイヤとなる。

5 Conditional Random Field(CRF)

CRF は画像の隣り合った画素のペアを考えて、その 2 つが同じ object に入るか否かを計算していく手法である。DeepLab では bilinear upsampling によって feature map を引きのばし、その直後に fully-connected Conditional Random Field を用いる。下に数式を示す。 θ_{ij} における $\mu(x_i, x_j)$ はフィルターであり、 $x_i \neq x_j$ のとき 1 であり、同じときに 0 をとる。カッコ内は重みづけされた 2 つのカーネルの和であり、1 つめのカーネルは値の差とピクセルの位置情報を扱い (a kind of bilateral filter), エッジの抽出などを行う。2 つめの

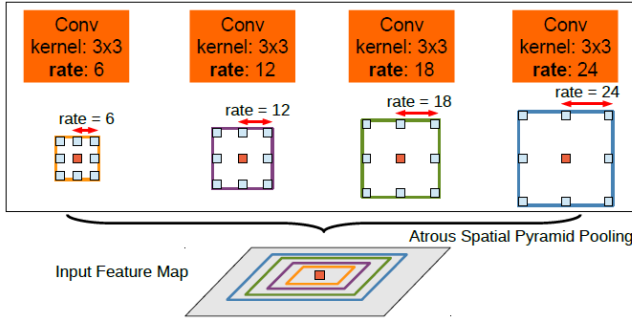


図 3: ASPP 概要図

カーネルはピクセルの位置情報のみを扱う (Gaussian filter). σ と w は cross validation によって得られる.

$$E[x] = \sum_i \theta_i(x_i) + \sum_{ij} \theta_{ij}(x_i, x_j)$$

$$\theta_i(x_i) = -\log P(x_i)$$

$$\theta_{ij}(x_i, x_j) = \mu(x_i, x_j) \left[w_1 \exp\left(-\frac{\|p_i - p_j\|^2}{2\sigma_\alpha^2} - \frac{\|I_i - I_j\|^2}{2\sigma_\beta^2}\right) + w_2 \exp\left(-\frac{\|p_i - p_j\|^2}{2\sigma_\gamma^2}\right) \right]$$

ただし, DeepLab v3, v3+ ではもう使われていない (CRF は前処理が必要で end-to-end な学習ができないため).

6 来週以降のタスク

github にいろいろなフレームワークでの実装が紹介されているのでひとまず動かせるかどうか試してみる.

参考文献

- [1] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *CoRR*, abs/1606.00915, 2016.

- [2] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. *CoRR*, abs/1802.02611, 2018.