

## 進捗報告

## 1 今週やったこと

Semantic Segmentation 関連の論文を読んでまとめた。各手法のまとめの前に Semantic Segmentation で用いられる評価指標を説明する。

## 1.1 pixel accuracy

1 枚の画像に対する正解画素の比率。  $N$  はピクセルの数。  $n_{ij}$  はクラス  $i$  を  $j$  と推定した個数。

$$\frac{1}{N} \sum_{i=0}^k n_{ii} \quad (1)$$

## 1.2 mean accuracy

各クラスの recall をクラス数で平均したもの。  $k$  はクラス数。  $t_i$  はクラス  $i$  の個数。

$$\frac{1}{k} \sum_{i=0}^k \frac{n_{ii}}{t_i} \quad (2)$$

## 1.3 mean IU(IoU)

IoU のクラス平均。

$$\frac{1}{k} \sum_{i=0}^k \frac{n_{ii}}{(t_i + \sum_{j=0}^k n_{ji} - n_{ii})} \quad (3)$$

また、  
 GT(Ground Truth): 正しいクラス  $i$  のピクセル  
 PR(Prediction):  $i$  と予測したピクセル  
 TP(True Positive): 正しく  $i$  と推定されたピクセル  
 とすると、

$$\text{IoU} = \frac{\text{TP}}{\text{GT} + \text{PR} - \text{TP}}$$

とも表すことができる。

## 1.4 frequency weighted IU

クラス間の個数に差がある場合を考慮した IoU。

$$\frac{1}{N} \sum_{i=0}^k \frac{t_i n_{ii}}{(t_i + \sum_{j=0}^k n_{ji} - n_{ii})} \quad (4)$$

## 2 Fully Convolutional Networks for Semantic Segmentation(FCN) [1]

すべてが Convolution 層によって構成されている (VGG, AlexNet, GoogLeNet のモデルの全結合層を Convolution 層にしている)。ローカルの情報とグローバルな情報を保つため (pixel の情報が重要であるため) に skip 構造を採用している。上層から skip してきた複数の出力を最終層で結合する。skip においては結合先でチャンネルに対する和をとっている。用いているデータセットは PASCAL VOC [2], NYUDv2 [3], SIFT Flow [4] の 3 種類。

## 3 来週以降のタスク

引き続き論文を読み進める。ある程度読めたら 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] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. *CoRR*, abs/1411.4038, 2014.
- [2] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The pascal visual object classes (voc) challenge. *International Journal of Computer Vision*, 88(2):303–338, June 2010.
- [3] Pushmeet Kohli Nathan Silberman, Derek Hoiem and Rob Fergus. Indoor segmentation and support inference from rgbd images. In *ECCV*, 2012.
- [4] Ce Liu, Jenny Yuen, and Antonio Torralba. Sift flow: Dense correspondence across scenes and its applications. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(5):978–994, May 2011.