

ゼロショット学習を用いた一般物体セグメンテーション

谷田 啓一, Tristan Hascoet, 滝口 哲也, 有木 康雄

神戸大学

1. はじめに

近年の物体認識の分野において, 自然にある全ての物体に対する物体認識は, 多くのオブジェクトクラスが含まれるため学習が困難である. また大規模な物体認識を行う場合, それぞれのクラスに対する十分な学習データを集めることは難しく, また集められるデータの数にもばらつきが生じることがある. こうした問題に対して, ゼロショット学習法の考え方を用いて学習データのないクラスの画像をセグメンテーションする方法を提案する.

2. ゼロショットセグメンテーション

ゼロショット学習法 (ZSL) では, 学習データのあるクラス (既知クラス) の特徴を用いて, 学習データのないクラス (未知クラス) を識別する. 物体認識の手法と ZSL[1] の関係は図 1 のように定義される. ZSL を用いない手法では, 識別クラスを 1 とした one hot vector が与えられ, ZSL を用いる手法では識別クラスに関する semantic vector が与えられる. またセグメンテーション (Segmentation) は入力画像の画素ごとに分類化 (Classification) を行うことから画素ごとの分類化と定義できる. 以上のことから, ゼロショットセグメンテーション (ZS_Segmentation) を画素ごとの ZSL を用いた分類化と定義する.

3. 提案手法

3.1 ゼロショットセグメンテーションの問題点

ZSL では未知クラスを識別するので, クラスごとの学習が偏りやすく, 入力画像のクラスに関わらず識別されるクラス (ハブ) が発生する.

3.2 学習モデル

今回の手法では, ベースモデルとして SegNet[2] を用いる. 図 1 にあるように, 入力画像の各画素のクラスラベルに, クラス名についての 100 次元ワードベクトルを埋め込んだものを用いる. 以降このモデルを ZS-SegNet と呼ぶ.

3.3 学習方法

ラベルマップの大きさを (H, W) (H : 縦方向の画素数, W : 横方向の画素数) とする. 3.1 で挙げた問題点を解決するために, ワードベクトルの正規化を行う. ワードベクトルを v とすると, 正規化ワードベクトル V は以下のようになる.

$$V = \frac{v}{\|v\|_2} \quad (1)$$

また画素 (i, j) ($i = 1, \dots, H$) ($j = 1, \dots, W$) に埋め込まれた正規化ワードベクトルを $\mathbf{v}_{i,j}$, ZS-SegNet からの出力を $\hat{\mathbf{v}}_{i,j}$ とすると, ZS-SegNet は以下の誤差関数を最小化するように学習する.

$$L = \sum_i^H \sum_j^W \|\mathbf{v}_{i,j} - \hat{\mathbf{v}}_{i,j}\|_2 \quad (2)$$

この誤差関数により, ZS-SegNet は各画素のクラスラベルに対応するワードベクトルを出力できるようになる.

3.4 クラスラベルの推定方法

ZS-SegNet の出力から未知クラスのラベルを推定するために, ZS-SegNet の出力に未知クラスのラベルに対応するワードベクトルを用いた k 近傍法を適用する.

4. 実験

4.1 実験条件

評価指標として画像全体での平均識別率 [%] を用いる. データセットとして COCO-Stuff[3] の学習データと評価データを用いる. 用いるクラスは COCO-Stuff の全クラスとし, 既知クラスと未知クラスの分割を COCO-Stuff の Label Hierarchy に従って図 2 のように行う.

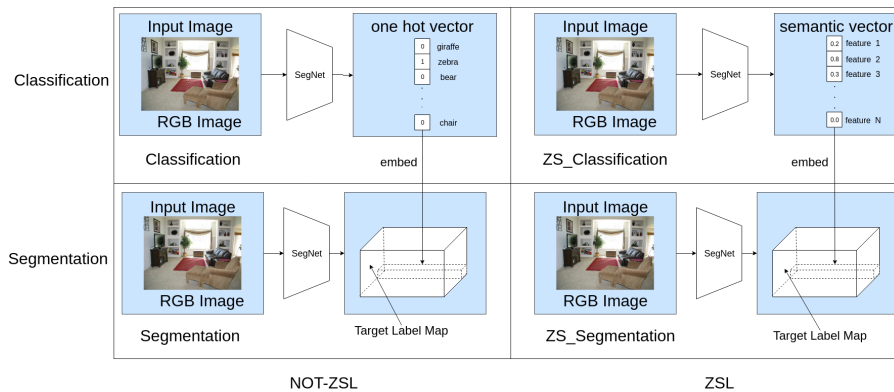


図1 物体認識の手法と ZSL の関係

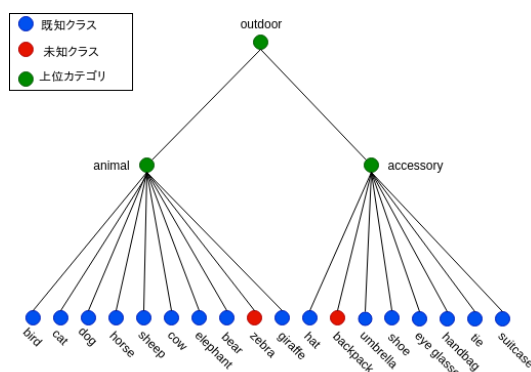


図2 既知クラスと未知クラスの分割例

未知クラスを上位カテゴリに対して一つ定義し、それ以外を既知クラスと定義した。クラス名からワードベクトルに変換するにあたり GloVe[4] を用いる。モデルの学習は 3.3 で示した方法で行い、ラベルの推定は 3.4 で示した方法で行う。

4.2 評価方法

ZS-SegNet の性能を評価するために、既存の SegNet[2] と ZS-SegNet に対して既知クラス、ZSL、generalized-ZSL(G-ZSL) の 3 つの手法に対して実験を行う。ここで、ZSL ではテスト時に未知クラスのワードベクトルのみを用いて未知クラスを識別するものとし、G-ZSL ではテスト時に既知クラスと未知クラスの両方のワードベクトルを用いて未知クラスを識別するものとする。

4.3 実験結果

結果は表 1 の通りとなった。SegNet では未知クラスを識別できないが、ZS-SegNet では未知クラスを識別できる。これは ZS-SegNet が入力画像から既知クラスの特

表1 各手法での識別率 [%]

手法	既知クラス	ZSL	G-ZSL
SegNet[2] の識別率	54.8	0.0	0.0
ZS-SegNet の識別率	52.4	17.1	0.01

徴を学習できたからである。また ZS-SegNet では、ZSL の手法において学習できない未知クラスが存在することも確認でき、G-ZSL の手法においてほとんど学習が来ないことも確認できる。

5. おわりに

本稿では学習データのないクラスの画像をセグメンテーションする方法を提案した。また実験を行い、未知クラスの画像に対して 17% の精度で識別できることを示した。今後はゼロショットセグメンテーションにおけるより良いワードベクトルの与え方や、グラフを用いた特徴の学習を行う予定である。

参考文献

- [1] M. Palatucci, D. Pomerleau, G. Hinton, and T. Mitchell, “Zero-Shot Learning with Semantic Output Codes,” in Neural Information Processing Systems (NIPS), 2009.
- [2] V. Badrinarayanan, A. Kendall, and R. Cipolla, “SegNet: A deep convolutional encoder-decoder architecture for image segmentation,” IEEE Transactions on Pattern Analysis and Machine Intelligence, .
- [3] H. Caesar, J. Uijlings, and V. Ferrari, “COCO-Stuff: Thing and stuff classes in context,” in Computer vision and pattern recognition (CVPR), 2018 IEEE conference on, IEEE, 2018.
- [4] J. Pennington, R. Socher, and C. D. Manning, “GloVe: Global vectors for word representation,” in EMNLP, 2014.