H-024

識別モデルの不確かさを考慮した Dense CAE による Semantic Segmentation 法の提案

Proposal of Semantic Segmentation Method by Dense CAE Considering Uncertainty of Discriminative Model

磯部周哉 † Shuya Isobe
荒井秀一 † Shuichi Arai

1.Semantic Segmentation

画像認識の分野において，Semantic Segmentation という画像をピクセルごとにカテゴリ分類する技術に関する研究が盛んに行われている。この技術により、コンピュータがシーン中に存在する物体のカテゴリ、位置、形を認識できるため、シーンを理解する上で重要な技術である。近年では、CNN(Convolutional Neural Network) を学習に用いることで分類の性能が飛躍的に向上した。これらの手法は予め学習するカテゴリを定義するため、認識する対象を、それらのカテゴリが出現するようなシーンに限定する必要がある。このことから、既存の手法はいずれも、定義したカテゴリ以外の物体、すなわち未知の物体の存在を仮定していない。しかし、実際のシーンにおいては、限られた状況であっても、未知の物体は必ず出現するはずである。そのため、Semantic Segmentation の実用化に向けては、物体が既知か未知であるかを正しく認識する必要がある。そこで我々は、識別モデルが異なれば分類結果も異なることに着目し、様々なモデルから得られた分類結果の分布の分散が大きい領域は未知の物体に分類する手法を提案し、新たに既知と未知の分類を可能とすることで、分類の正確性の向上を目指す。

2. 提案手法

2.1.Dense CAE

学習に用いるネットワークとして、一般物体認識のタスクにおいて顕著な成績を残した Dense Net[1] と、畳み込みにより抽出され圧縮された特徴量を入力データのサイズに復元することが可能な CAE(Convolutional Auto-Encoder) を組み合わせた Dense CAE[2] を使用する。Dense Net は式 (1) のように前の層の特徴を連結していくことで、通常の深いネットワークでは失われてしまう、浅い層で抽出された特徴を保持することができ、汎化性能を向上させることができる。x_l は l 層目における出力を表し、H は畳み込みなどの非線形変換を表す。

$$x_l = H_l([x_0, x_1, ..., x_{l-1}]) \quad (1)$$

また、Semantic Segmentation のようなピクセルレベルの分類タスクにおいては、解像度は重要な要素であるため、CAE のように入力データのサイズに特徴量を復元することは有効である。今回我々は 77 層の Dense CAE を実装し、後述する実験に使用した。図 1 に実装したネットワーク構造を示す。

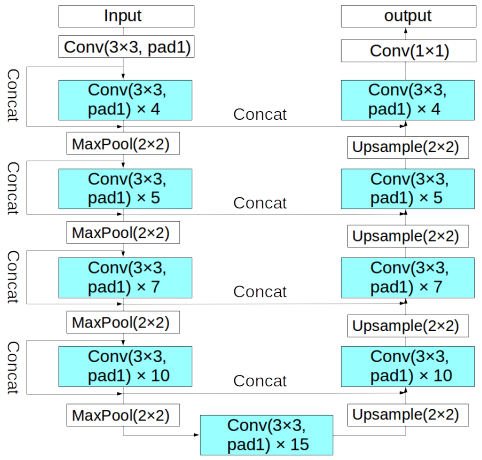

図 1: 77 層の Dense CAE のネットワーク構造

2.2. 識別モデルの不確かさを考慮した推論法

Dense CAE を用いて Semantic Segmentation を行うにあたり、我々は識別モデルが異なれば分類結果も異なることに着目した。例えば、ある画像に対して、モデル A が猫と分類しても、モデル B は犬と分類するケースが存在するということである。このように、識別モデルは不確かさを含んでいるため、我々は様々なモデルで推論した結果を用いて、より確からしい分類を行う手法を提案する。図 2 に提案手法の概要を示す。まず、様々なモデルを擬似的に構築する方法として、Dropout[3] を使用する。Dropout はランダムにニューロンを消去しながら学習を進めることで、過学習を抑える手法である。この手法は通常、学習時にのみに用いられるが、我々は推論時にも用いることで、ランダムにニューロンを消去して様々なモデルで推論を行う。そして、様々なモデルによる推論から得られた分類結果をピクセルごとに分布を取り、その平均を一時的な分類結果とする。そしてさらに、分布の分散を求めることで、未知の存在の可能性を明らかにする。モデルによって、あるピクセルにおける分類結果のばらつきが大きい場合、どのモデルにおいてもそのピクセルがどのカテゴリに属するか確信を持てていない状態であると考えられる。また、既知のカテゴリに関する学習が十分に行えていると仮定すると、分散が大きいピクセルは未知の物体である可能性が高いと考えられる。そこで我々は、分散に対する閾値を実験的に定め、分散が閾値より大きい場合は未知の物体に分類し、小さい場合は分布の平均で一番確率が高いカテゴリに分類することで、新た

†東京都市大学大学院 工学研究科

に既知と未知の分類を可能とした。

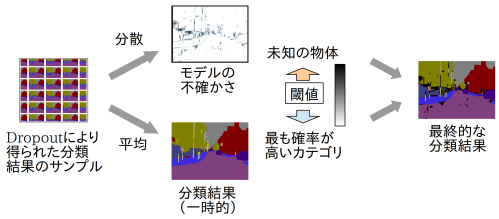

図 2: 提案手法の概要

3. 実験

3.1. 学習条件

提案手法による推論を実現するために、2 章で示した Dense CAE による学習を行う。データセットは屋外シーンの画像群である CamVid[4] を使用した。学習用に 367 枚、テスト用に 233 枚用意されており、対象とするカテゴリは Sky、Building、Pole、Road、Pavement、Tree、Sign symbol、Fence、Car、Pedestrian、Bicyclist の 11 カテゴリである。表 1 にハイパーパラメータ等の学習条件を示す。

表 1: 使用したハイパーパラメータ

Epoch	Optimizer	Learning rate	Weight decay
450	RMSProp	0.001	0.0001

3.2. 推論結果

表 1 に示した学習条件で学習させた Dense CAE を用いて、提案手法により推論した結果を図 3 に示す。

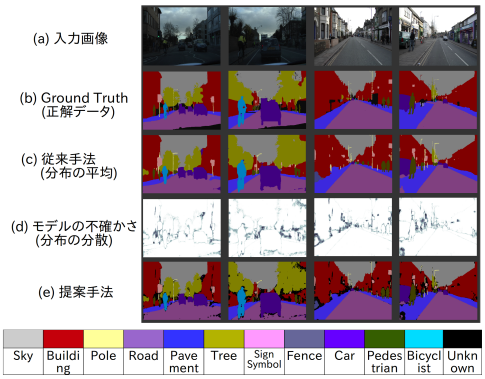

図 3: 提案手法による推論結果

図中の分布の分散から得られるモデルの不確かさ (d) は、黒いほうが分散が大きいことを表している。提案手法の推論結果 (e) を見ると、分散が大きい領域においては未知の物体に分類できていることがわかる。

3.3. 推論結果の評価

提案手法の有効性を示すために、図 3 に示したような推論結果を、このタスクで広く用いられている式 (2)(3)(4) の 3 つの評価尺度：Global acc(Global accuracy)、Class avg(Class average)、Mean IoU(Mean intersection over union) で評価を行う。3 つとも推論結果が Ground truth に近いほど高い値となる。i はカテゴリのインデックス、C はカテゴリ数を表し、GT は Ground truth、TP は True positive、FP は False positive のピクセル数を表す。

$$Global\ acc = \frac{TP}{GT} \tag{2}$$

$$Class\ avg = \frac{\sum_i^C \frac{TP_i}{GT_i}}{C} \tag{3}$$

$$Mean\ IoU = \frac{\sum_i^C \frac{TP_i}{GT_i + FP_i}}{C} \tag{4}$$

3 つの尺度による評価結果を表 2 に示す。ここで比較対象とする従来手法は、分布の平均から得られる推論結果である。3 つの評価尺度による比較結果から、提案手法により分類の正確性が向上していることがわかる。

表 2: 提案手法と従来手法の正確度の比較

	Global acc	Class avg	Mean IoU
従来手法	90.6	82.1	65.5
提案手法	91.7	83.4	67.9

4. 結論

本稿では、既存の Semantic Segmentation の手法がいずれも未知の物体の存在を仮定していない点に着目した。そして、識別モデルの不確かさを用いて、新たに推論する対象の物体が既知か未知かの分類を可能とする手法を提案し、分類の正確性の向上を図った。その結果、3 つの評価尺度による従来手法との比較から、提案手法の有効性を示した。

参考文献

[1] Huang, G., et al. (2017, July). Densely connected convolutional networks. CVPR, (Vol. 1, No. 2, p. 3).

[2] Jegou, S., et al. (2017, July). The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. CVPRW, (pp. 1175-1183).

[3] Srivastava, N., et al. (2014). Dropout: A simple way to prevent neural networks from overfitting. JMLR, 15(1), 1929-1958.

[4] Brostow, G. J., et al. (2009). Semantic object classes in video: A high-definition ground truth database. Pattern Recognition Letters, 30(2), 88-97.