

Cellular Pyramidによる Semantic Segmentationのためのセルラ回路の自動構築

島崎 謙^{1,a)} 長尾 智晴^{1,b)}

概要：近年、セルと呼ばれる処理ユニットが相互結合したセルラ型の構造回路を用いた画像処理手法が提案され有効性が示されている。局所的な情報やその空間的な関係を考慮した処理構造は画像中の物体とその領域を認識する Semantic Segmentation においても有効と考えられる。しかし、Semantic Segmentation では画像の大局的な情報も考慮する必要があるため、既存のセルラ回路の処理構造では不十分である。そこで本稿では複数の解像度ごとにセルラ回路を構築し統合した結果を用いる多層型セルラ回路 Cellular Pyramid を提案し、Semantic Segmentation の問題に適用する。

1. はじめに

近年、画像認識の分野では従来の画像単位での物体認識ではなく画像中の複数の物体とその領域を認識する Semantic Segmentation に関する研究が盛んに行われている [1]。一方、画像処理の分野では Cellular Neural Network(CNN) に代表されるセルと呼ばれる処理ユニットが相互結合したセルラ型の構造回路を用いた画像処理手法が提案されており有効性が示されている [2]。局所的な情報やその空間的な関係を考慮した処理構造は Semantic Segmentation においても有効に働く考えられる。しかし、画像のより大局的な構造も考慮する必要があるため、既存のセルラ回路の処理では不十分である。そこで本稿では複数の解像度ごとに回路を構築し処理を行う多層セルラ回路 Cellular Pyramid を提案する。なお、回路の構築には進化計算法の一つである Cartesian Genetic Programming (CGP) [3] を用いる。

2. Cellular Pyramid

2.1 Cellular Pyramid の概要

Cellular Pyramid の処理構造を図 1 に示す。M×N[pixel] の入力画像を対象とした場合、M×N 個のユニットを格子状に配置する。このセルラ回路は入力画像の画素値と近傍のユニットの出力値を入力として、CGP による演算によって出力値を繰り返し計算を行うことで動作する。続いて入力画像を 1/2 倍に縮小した画像に対してユニットを配置し、出力値を計算する。入力画像を 1/4 倍に縮小した画像

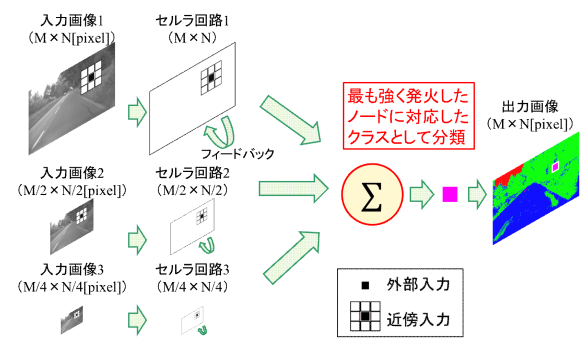


図 1 Cellular Pyramid の概要

に対しても同様の処理を行う。上記の複数のセルラ回路から得られた出力値を統合し、その結果を用いて画素の認識を行う。異なる解像度の出力値を考慮することで、画像の大局的な構造が考慮できることを期待している。ここで、各解像度で用いられるユニットの構造は CGP を用いて独立に最適化を行う。また、従来のセルラ回路ではグレースケールの画像が処理の対象とされてきた、視認性を向上させるために色は重要な要素である。そこで、本手法では文献 [4] の処理構造を参考に RGB の各色成分が相互に作用する処理構造を採用した。認識処理の詳細は次節で述べる。

2.2 セルラ回路の認識処理の流れ

本手法で用いる CGP は外部入力ノード、内部入力ノード、外部出力ノード、内部出力ノード、演算ノードの 5 種類のノードで構成される。ここで、外部出力ノードは認識対象とする物体のカテゴリ数 C と同じ数用意する。T を出力回数、 \mathbf{p} をセルラ回路上の座標、 \mathbf{r} を \mathbf{p} の近傍座標の集合として、各解像度での認識処理の手順を以下に示す。

(1) 入力画像の全ての画素値を $[0, 1]$ に正規化する。

¹ 横浜国立大学 大学院環境情報学府
^{a)} shimazaki-ken-yg@ynu.jp
^{b)} nagao@ynu.ac.jp

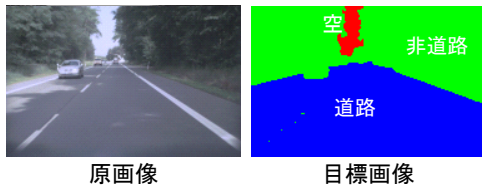


図 2 使用する画像の例

- (2) 全ての位置 \mathbf{p} の CGP を以下の通りに初期化する.
 - (a) 位置 \mathbf{p} の画素値を外部入力ノードに入力する.
 - (b) 位置 \mathbf{r} の画素値を内部入力ノードに入力する.
- (3) 以下の処理を T 回繰り返す.
 - (a) 全ての CGP を同期して 1 度だけ出力させる.
 - (b) 位置 \mathbf{p} の CGP の内部入力ノードに, 位置 \mathbf{r} の CGP の内部出力ノードの出力値を入力する.

2.3 各層の結果の統合

前節の処理を各解像度で行い, 外部出力ノードが出力した値を用いて位置 \mathbf{p} の画素の認識を行う. 外部出力ノードの出力値の統合方法を式 (1) に示す.

$$\mathbf{p}_{\text{cat}} = \arg \max_{c \in C} \left(\sum_{l=0}^{L-1} w_l O_c(\mathbf{p}_l, c) \right) \quad (1)$$

$$\mathbf{p}_l = \left(\lceil \frac{x}{2^l} \rceil, \lceil \frac{y}{2^l} \rceil \right) \quad (2)$$

ここで, \mathbf{p}_{cat} は認識結果のカテゴリ, L は用いたセルラ回路の数 (本稿では $L = 3$ で固定), $O_c(\mathbf{p}_l, c)$ は l 番目の縮小画像の位置 \mathbf{p}_l におけるカテゴリ c に対応した外部出力ノードの出力値, w_l は重み係数をそれぞれ表す. 各カテゴリの外部出力値の総和の内, 最も強く発火したノードに対応したカテゴリとして位置 \mathbf{p} の画素値を認識する.

3. Semantic Segmentation 実験

本稿では, 車載カメラで撮影した画像の大局的な構造を把握するために「道路」, 「非道路」, 「空」へ分類する 3 クラスの Semantic Segmentation 問題を用いて提案手法の性能を検証する. 本実験で用いる画像の例を図 2 に示す. ここで, 目標画像中の色は認識対象となるカテゴリを表している. ユニットの構築のための学習画像として 5 枚, 性能を評価するための未知画像として 5 枚の画像を用意した. また, 原画像を用いて構築した回路のみを用いた手法と比較することで提案手法の有効性を検証する.

4. 結果と考察

学習, および未知における認識結果の例を図 3, 図 4 にそれぞれ示す. 目標画像と類似しているほど高精度な認識結果だといえる. 図 3 上部の画像ではどちらの手法においても比較的良好な認識結果が得られている. 図 3 下部の画像では縮小画像なしの結果において「道路」の日蔭の領域の認識結果が不十分なのに対して提案手法では良好な認識結果が得られている. これはフィードバック処理を用いても

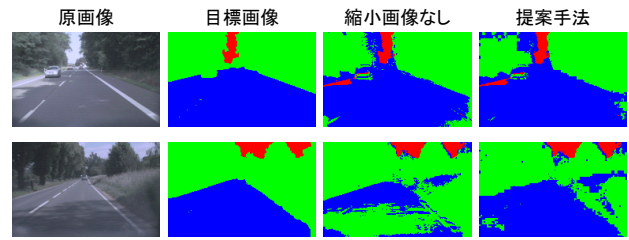


図 3 学習画像の認識結果の例

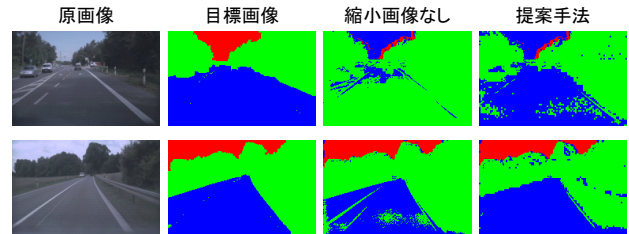


図 4 未知画像の認識結果の例

日蔭とそうではない道路領域を考慮した認識処理ができなかったためと考えられる. 一方, 縮小画像を処理するセルラ回路では原画像を処理するセルラ回路に比べてより広い範囲の画素を考慮した認識処理が可能である. その結果を統合することで良好な認識結果が得られたと考えられる.

同様の結果が図 4 の上部の画像でも確認できる. しかし, 「空」の領域の認識結果が不十分である. これは全ての解像度に対して汎化性能の高い認識処理が構築できなかったためと考えられる. そのため, 回路の処理構造について精査する必要がある. また, 出力値を統合したことで誤った認識が行われている領域が存在することから認識結果の統合方法についても検討する必要がある.

5. おわりに

本稿では複数の解像度ごとにセルラ回路を構築し, それらの結果を統合し認識処理を行う多層セルラ回路 Cellular Pyramid を提案した. 今後, 精度向上のための回路構築方法の検討, 各回路の出力の統合方法について検討する予定である.

参考文献

- [1] J. Shotton, M. Johnson, and R. Cipolla. Semantic tex-ton forests for image categorization and segmentation. In Computer Vision and Pattern Recognition, 2008.
- [2] 永田智洋, 長尾智晴. "セルラ進化型ニューラルネットワークによるシングルフレーム超解像処理." 電子情報通信学会論文誌 D 95.10 (2012): 1859-1868.
- [3] J. F. Miller and P. Thomson, "Cartesian Genetic Programming", Processing of the Third European Conference on Genetic Programming, LNCS, Vol.1802, pp121-132, 2000.
- [4] 井上貴志, 西尾芳文. "カラー画像を扱う 3 層 Cellular Neural Networks の研究." 電子情報通信学会技術研究報告. NLP, 非線形問題 108.174 (2008): 33-36.