

1. はじめに

セマンティックセグメンテーションは、画像に含まれる物体をピクセル単位で識別する問題である。セマンティックセグメンテーションを高精度化する手法として、2次元拡張した RNN と DCNN を組み合わせた DAG-RNNs[1] が提案されている。しかしながら、RNN を組み合わせると GPU による並列演算ができないため、識別時間が増加するという課題がある。そこで、本研究では 2D-QRNN を提案し、DCNN に導入することでセマンティックセグメンテーションの高精度化及び高速化を実現する。

2. Quasi-Recurrent Neural Networks

Quasi-Recurrent Neural Networks(QRNN)[2] は、DCNN で用いられる畳み込みを LSTM に応用したネットワークである。QRNN は、LSTM における忘却ゲート、入力ゲート、出力ゲート、メモリセルへの入力候補値を畳み込みにより算出する。メモリセルの各ユニットは、メモリセルの他のユニットと独立して更新する。

3. 提案手法

QRNN をセマンティックセグメンテーションに適用するために、2次元拡張を行う必要がある。2次元拡張した QRNN を 2D-QRNN として定義する。2D-QRNN におけるメモリセルの更新は、図 1 で示すように $(+x, +y)$, $(-x, +y)$, $(+x, -y)$, $(-x, -y)$ の 4 通りの経路の順で行う。

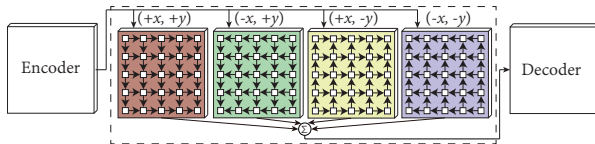


図 1：4 方向の 2D-QRNN の処理の流れ

2D-QRNN の 4 方向の出力は、最終的に 1×1 サイズの畳み込みを行い統合する。2D-QRNN における忘却ゲート \mathbf{f}_t 、入力ゲート \mathbf{i}_t 、出力ゲート \mathbf{o}_t とし、算出方法を式 (1)、式 (2)、式 (3) に示す。 d と d' は、 x または y 方向のいずれかである。 \mathbf{a} は前層からの入力、 t は座標を示す。忘却ゲートは、軸ごとに用意する。ここで、 \odot は内積計算、 \oplus はユニット同士の乗算と加算である。

$$\mathbf{f}_t^{d'} = \text{sigmoid}(\mathbf{W}_f^{d'} \cdot \mathbf{a}_{t-1} + \mathbf{W}_f \cdot \mathbf{a}_t) \quad (1)$$

$$\mathbf{i}_t = \text{sigmoid}(\Sigma_d(\mathbf{W}_i^d \cdot \mathbf{a}_{t-1}) + \mathbf{W}_i \cdot \mathbf{a}_t) \quad (2)$$

$$\mathbf{o}_t = \text{sigmoid}(\Sigma_d(\mathbf{W}_o^d \cdot \mathbf{a}_{t-1}) + \mathbf{W}_o \cdot \mathbf{a}_t) \quad (3)$$

メモリセルへの入力候補値 \mathbf{z}_t の算出方法を式 (4) に示す。

$$\mathbf{z}_t = \tanh(\Sigma_d(\mathbf{W}_z^d \cdot \mathbf{a}_{t-1}) + \mathbf{W}_z \cdot \mathbf{a}_t) \quad (4)$$

メモリセル \mathbf{c}_t 、2D-QRNN の出力 \mathbf{h}_t の算出方法を式 (5)、式 (6) に示す。

$$\mathbf{c}_t = \Sigma_d(\mathbf{f}_t^{d'} \odot \mathbf{c}_{t-1}^d) + \mathbf{i}_t \odot \mathbf{z}_t \quad (5)$$

$$\mathbf{h}_t = \mathbf{o}_t \odot \mathbf{c}_t \quad (6)$$

DCNN は局所的な領域からクラスを推論するのに対して、2D-QRNN は大局的な領域からクラス推論することが可能である。従って、DCNN と 2D-QRNN を組み合わせることで、局所領域と大局領域の 2 つの情報を考慮したセマンティックセグメンテーションを行うことが可能となる。本研究では、全結合層を除いた VGG-16 を Encoder と Decoder の構成としたネットワークをベースとする。2D-QRNN を図 2 のように、Encoder と Decoder の間の中間層として導入する。

4. 評価実験

2D-QRNN を DCNN に導入する効果を評価実験により検証する。

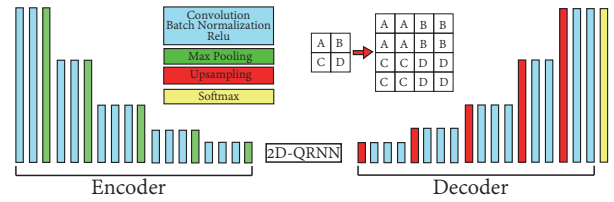


図 2：2D-QRNN + VGG-16

4.1. 実験条件

本実験では、VGG-16、DAG-RNNs+VGG-16、2D-QRNN+VGG-16 の 3 つのネットワークによるセマンティックセグメンテーションの性能比較を行う。データセットは、Clothing-Co-Parsing を使用する。識別クラスは、髪、肌、メガネ、上半身、下半身、カバン、くつ、帽子、背景の 9 クラスである。学習用に 456 枚、評価用に 20 枚を用いる。学習時は、5 種類のサイズ (320×160 , 512×256 , 640×320 , 960×480 , $1,280 \times 640$) で、マルチスケール学習を行う。評価時の画像サイズは、学習時と等しい 5 種類とする。使用機器は、GPU が Tesla P100、CPU が E5-2643 v4 @3.40GHz である。

4.2. 実験結果

識別精度の比較を図 3 に示す。提案手法は、評価サイズが 320×160 , 512×256 , 640×320 の時、DAG-RNNs+VGG-16 の Mean IoU よりも良い精度を達成している。また、全評価サイズにおいて、提案手法は VGG-16 の Mean IoU よりも良い精度となっている。

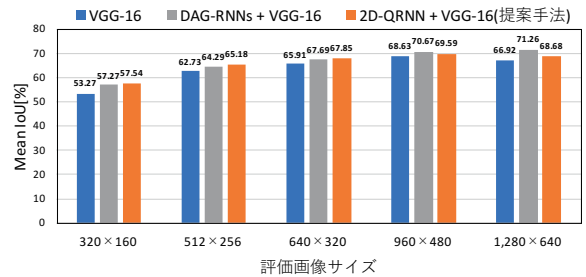


図 3：各評価サイズにおける Mean IoU の比較

識別時間の比較を図 4 に示す。VGG-16 の識別時間は、極めて短く評価画像サイズに依存していないことがわかる。提案手法は、全評価サイズにおいて、DAG-RNNs+VGG-16 と比較すると約 1/2 以下の識別時間となった。

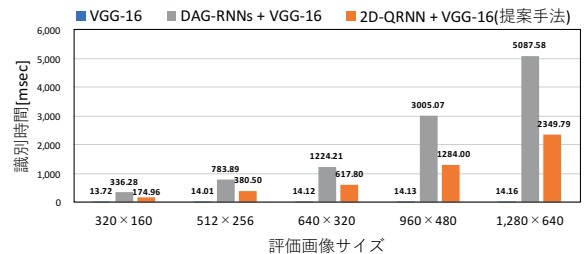


図 4：各評価サイズにおける識別時間の比較

5. おわりに

本研究では、2D-QRNN を DCNN に導入することで、従来の DAG-RNNs を用いた手法と同等の精度を保ちつつ、識別時間を半分に抑えることができた。今後は、時系列データを利用したセマンティックセグメンテーションの高精度化を行う。

参考文献

- [1] B. Shuaia, et al, "DAG-Recurrent Neural Networks For Scene Labeling", CVPR, 2015.
- [2] J. Bradbury, et al, "Quasi-Recurrent Neural Networks", ICLP, 2016.