

再帰型畳み込みニューラルネットワークを用いた風景画像認識の調査

宮崎大学情報システム工学科

森 芳雄

指導教員 椋木雅之

2017年 2月16日

研究背景

大規模な畳み込みニューラルネットワークによる教師付き深層学習の手法は、一般画像認識などで非常に成功している

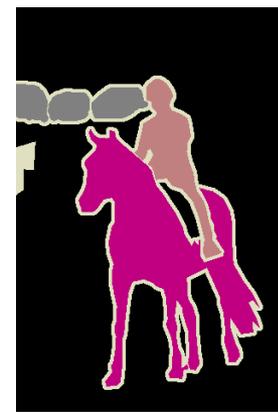


従来研究

セマンティック・セグメンテーション

画像に写っている対象のクラス名称とその対象の写っている位置をピクセルレベルで判別する

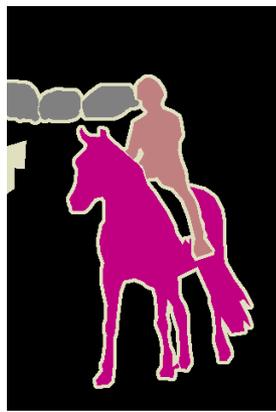
セマンティック・セグメンテーションのための手法にCRFasRNN[1]がある



[1]Conditional Random Fields as Recurrent Neural Networks, ICCV, 2015

問題点

- 物体を対象とした研究は多くされているが、風景画像を対象とした研究は少ない
- 風景画像に写っている同じクラスのもので、形が定まっていないものがある

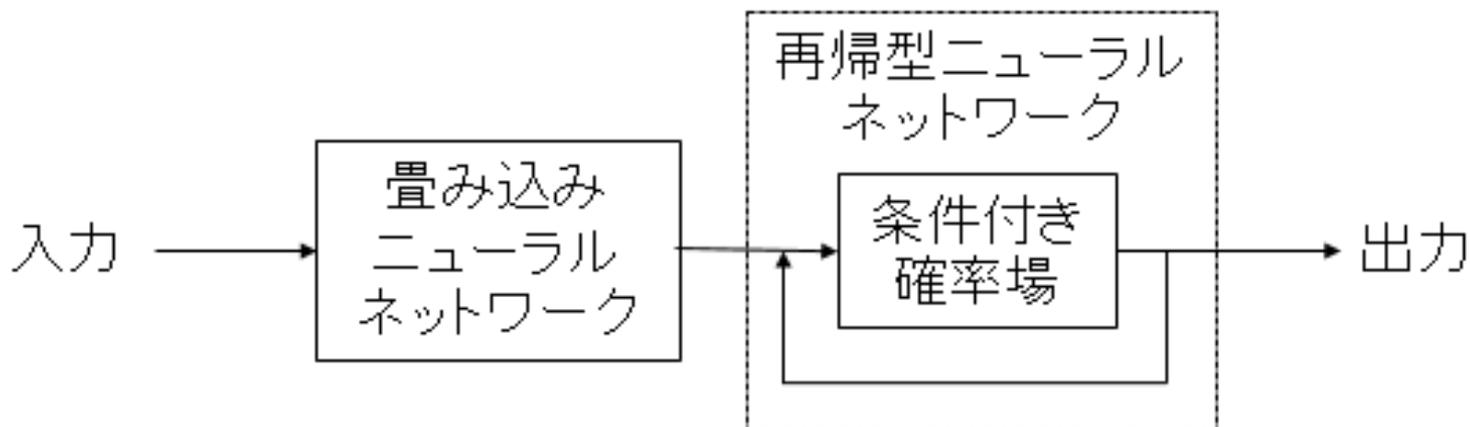


目的

CRFasRNNが風景画像に対しても有効であるか
調査する

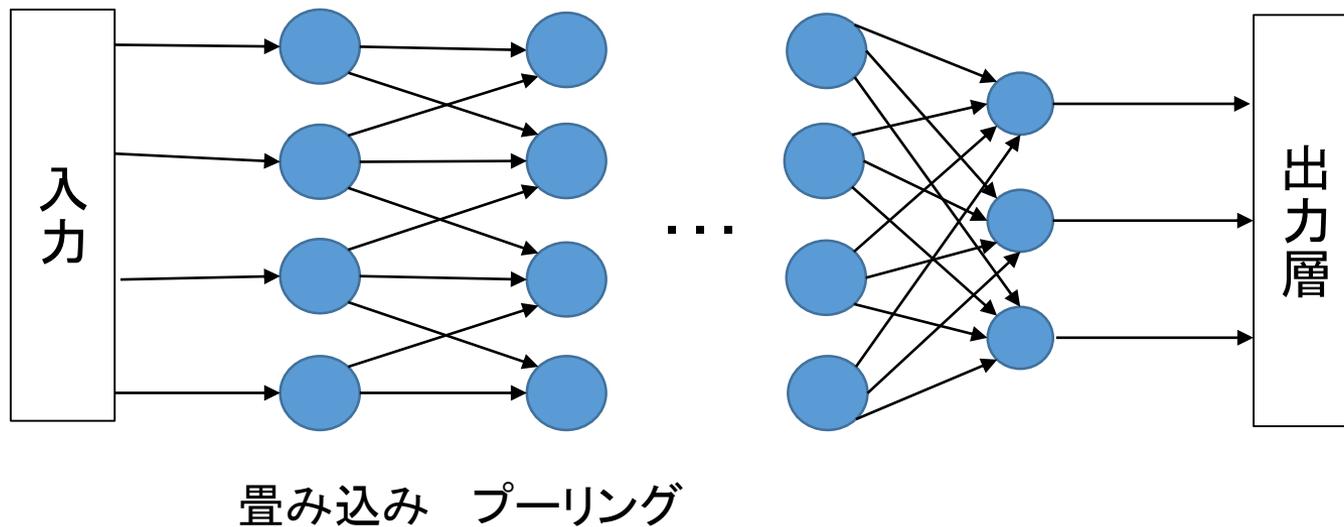
CRFasRNN

- 畳み込みニューラルネットワークに再帰型ニューラルネットワークを組み合わせたもの
- 畳み込みニューラルネットワークではラベル予測を行っている
- 条件付き確率場を再帰型ニューラルネットワークとして構築



畳み込みニューラルネットワーク

- 畳み込みニューラルネットワークは畳み込み層とプーリング層を含む順伝播型ネットワーク
- 画像認識で非常に有効である
- ピクセルレベルでラベル予測を行う

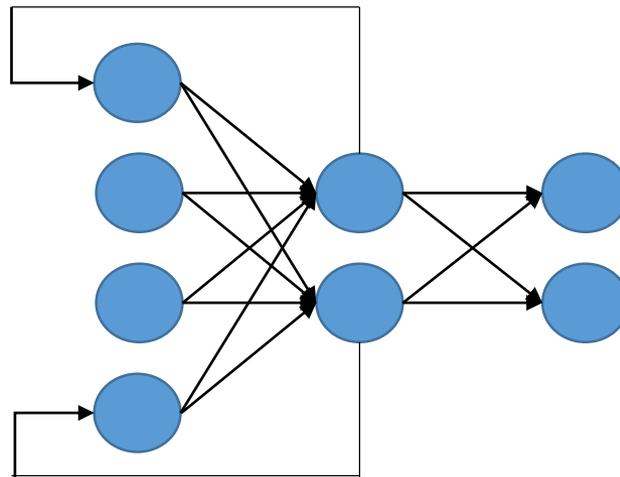


条件付き確率場

- 通常の物体認識のための畳み込みニューラルネットワークでは出力が粗くなることがある
- ピクセルレベルのラベル予測を精緻化し、粗い出力を改良する
- ピクセルレベルのラベル予測に畳み込みニューラルネットワークを利用する際の欠点を克服できる

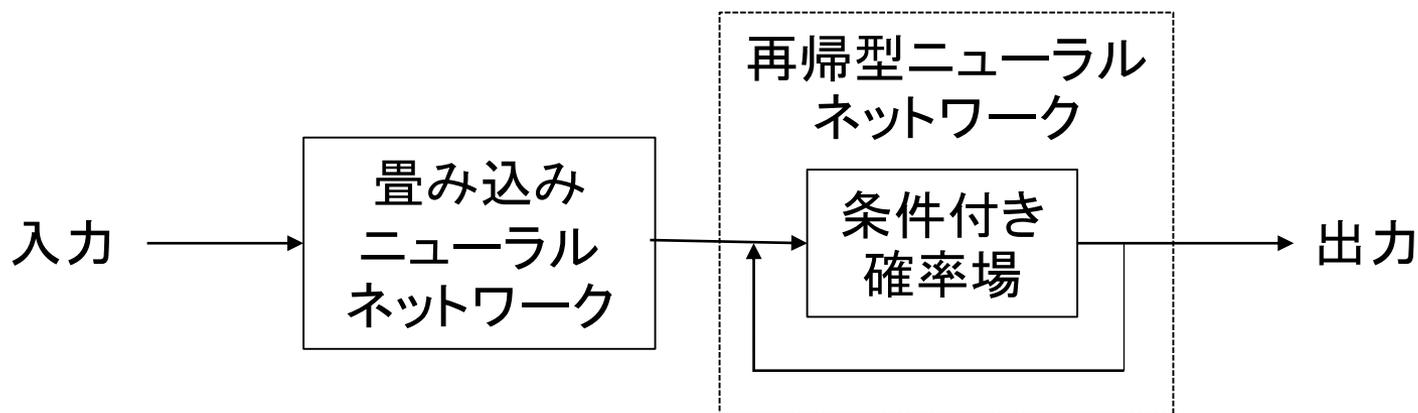
再帰型ニューラルネットワーク

- 情報を一時的に記憶し、振る舞いを動的に変化させることができる
- 画像上で隣り合うピクセルの並びをとらえ、分類問題をうまく処理できる
- 条件付き確率場を再帰型ニューラルネットワークとして構築することで細かい境界を生成できる



CRFasRNN

- 畳み込みニューラルネットワークは画像認識で非常に有効である
- 条件付き確率場はピクセルレベルのラベル予測を精緻化できる
- CRFasRNNは畳み込みニューラルネットワークと条件付き確率場の両方の望ましい特性を有する



風景を対象とした実験

- 風景画像にCRFasRNNを適用して、どれだけ識別できるかを調査
- 公開データセットを使用して学習・評価

実験で用いるデータセット

- SUN2012データセットを用いる
- 16,873枚の画像があり、4,919種類のクラスが含まれている

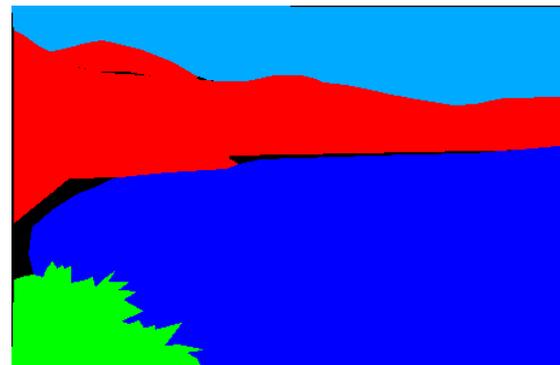


訓練用画像の作成

- クラスごとに色分けした画像を作成
- 分類するクラスは空、木、水、山、岩、道路の6クラス
- 正解画像と元画像のペアを訓練用画像とする
- 訓練用画像は931組作成 841枚を使用



元画像



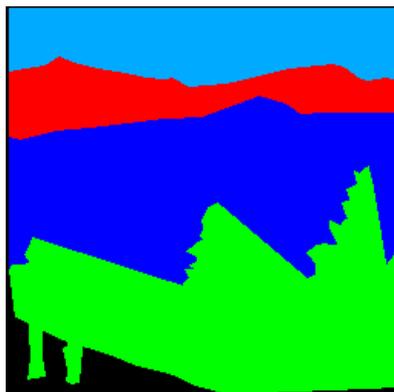
正解画像

実験結果

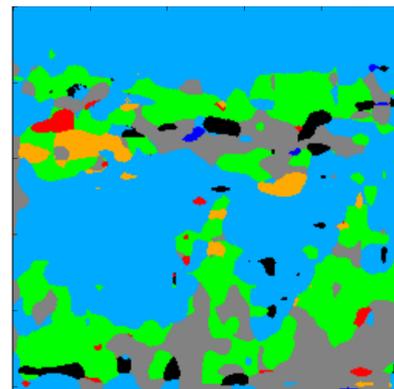
15,000回学習を行った



入力画像



正解画像

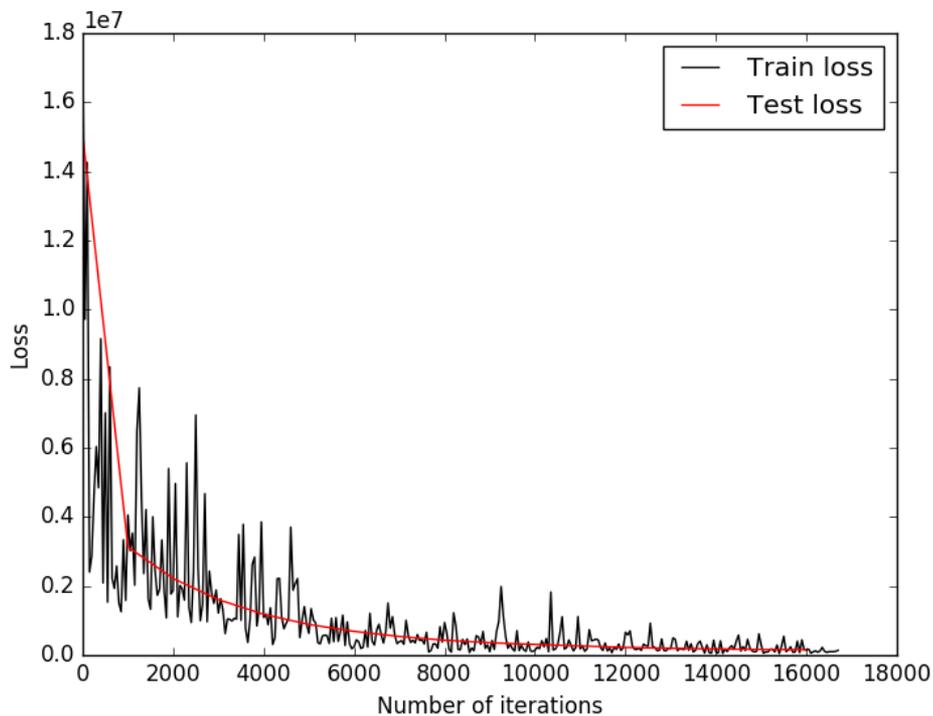


学習 15,000回

空: 	木: 	水: 
山: 	岩: 	道路: 

空、水の領域が空として識別され、他はほとんど識別されなかった

風景画像の学習の損失

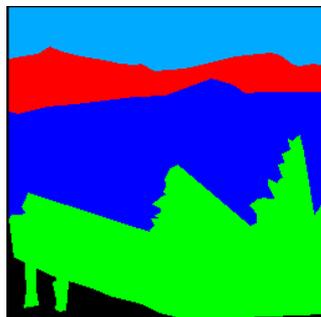


損失は小さい値に収束している

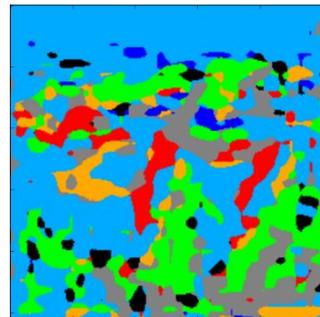
学習の進捗状況の確認



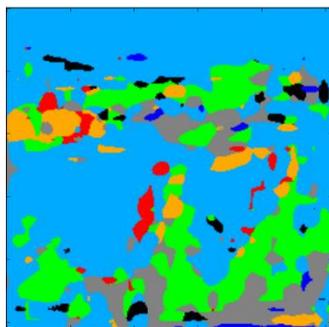
入力画像



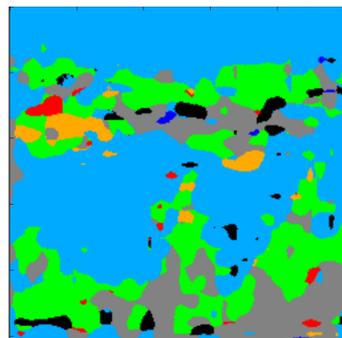
正解画像



学習5,000回



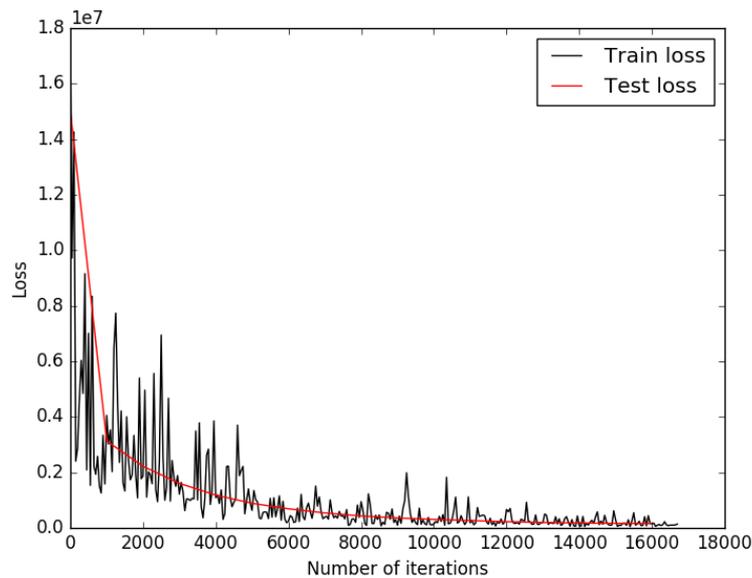
学習10,000回



学習15,000回

学習5,000回では木も識別されていたが、10,000回ではその部分は減り、識別結果が単調に向上してはいなかった

風景画像の学習の損失



収束の仕方は一様でなく、損失が大きくなる箇所が存在

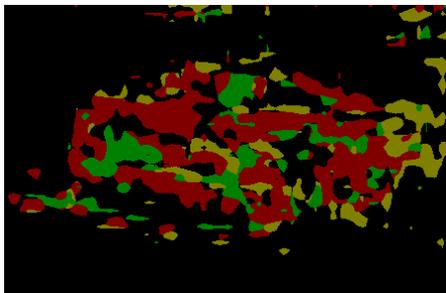
物体を対象とした確認実験

- 識別結果が良くなかったのは対象が風景であったことが原因かを調査
- データセットはCRFasRNNの論文でも利用していたVOC2012を使用
- 訓練用画像は車、バス、バイクの画像計449組
- 学習回数5,000回、10,000回、15,000回

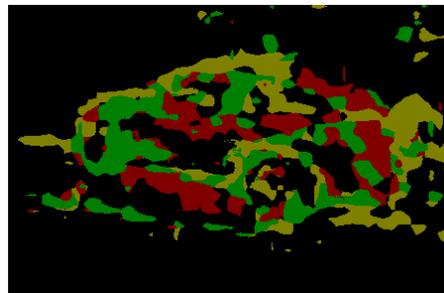
実験結果



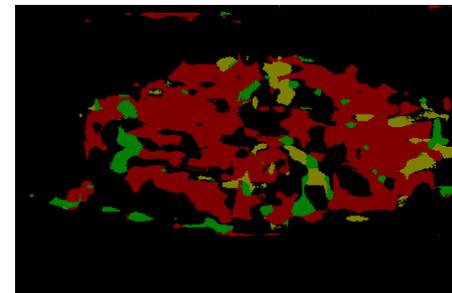
入力画像



学習5,000回



学習10,000回

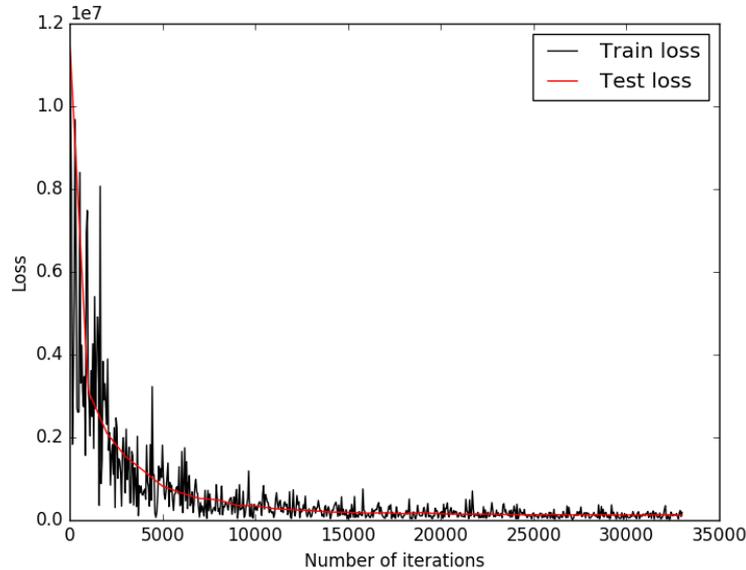


学習15,000回



学習回数によって多く識別されているクラスが異なっていた

物体識別の学習の損失



- 風景画像同様、収束の仕方は一様ではなかった
- 対象が風景だからうまくいかなかったわけではなかった
- 学習の損失が一様に収束しないのはCRFasRNNの特性と考えられる

まとめ

- CRFasRNNの特性として、学習の損失が一様に収束せず、識別結果が単調に向上しないことが分かった
- これは風景画像特有の現象ではないことを確認できた

今後の課題

学習回数を多くし、風景画像の識別結果がどのように変化していくか調査する