重み付け誤差関数を導入した DCNN による顔領域セグメンテーションの高精度化

EP11086 中村隆也　　指導教授：藤吉弘亘，山下隆義

1. はじめに

画像のセマンティックセグメンテーションは，画素毎に属するクラスを求める問題であり，多くの手法が提案されている．中でも，Deep Convolutional Neural Network(以下 DCNN) を用いた手法は，高い認識性能を持つことが報告されている [1]．しかし，面積の小さいクラスに対して認識精度が低下する問題がある．そこで，本研究では DCNN に重み付け誤差関数を導入することで，面積の小さいクラスに対して高精度なセグメンテーションを実現する．

2. 提案手法

DCNN を用いたセグメンテーションにおいて，面積の小さいクラスは，認識精度が低下する傾向がある．そこで，本研究では DCNN の学習に重み付け誤差関数を導入する．重み付け誤差関数とは，クラス毎に学習誤差の重みを設定する手法である．クラス c の学習誤差 E_c は式 (1) のように表される．

$$E_c = \frac{1}{2}\sum_{x\in X}\sum_{y\in Y}(T_{cxy} - O_{cxy})^2 \tag{1}$$

ここで，T は教師ラベル，X と Y はラベルの縦と横の幅，O は出力である．この各クラスの学習誤差に重み v を設定し，確率的勾配降下法により結合重み $\boldsymbol{W}$ を式 (2) のように更新する．

$$\begin{aligned}\boldsymbol{W}_{t+1} &= \boldsymbol{W}_t - \eta\frac{\partial\sum_{c\in C}E_c}{\partial\boldsymbol{W}_t}\\ &= \boldsymbol{W}_t - \eta\left(\frac{\partial E_0\cdot v_0}{\partial\boldsymbol{W}_t} + \frac{\partial E_1\cdot v_1}{\partial\boldsymbol{W}_t} + \cdots + \frac{\partial E_C\cdot v_C}{\partial\boldsymbol{W}_t}\right)\end{aligned} \tag{2}$$

ここで，η は学習係数，t は更新回数，C はクラス数である．学習係数 η は定数であるため，式 (2) は式 (3) のように変形できる．

$$\begin{aligned}\boldsymbol{W}_{t+1} = \boldsymbol{W}_t - \Bigg(&\eta\cdot v_0\cdot\frac{\partial E_0}{\partial\boldsymbol{W}_t} + \eta\cdot v_1\cdot\frac{\partial E_1}{\partial\boldsymbol{W}_t} + \\ &\cdots + \eta\cdot v_C\cdot\frac{\partial E_C}{\partial\boldsymbol{W}_t}\Bigg)\end{aligned} \tag{3}$$

式 (3) より，学習誤差に対してクラス毎の重みを設定することは，各クラスに対して学習係数を変化させることと等価である．すなわち，重み付け誤差関数を導入し，確率的勾配降下法によるパラメータの変化量を増減させることで，任意のクラスの学習速度を調節することができる．

顔領域セグメンテーションにおける重み付け誤差関数を導入した際の学習誤差の推移を図 1 に示す．ここでは，面積の小さいクラスとして目のクラスを対象とした．図 1 より，設定する重みを小さくすることで学習の更新速度が遅くなることがわかる．

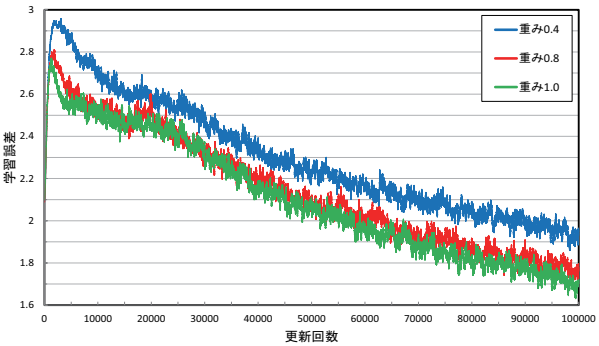

図 1：更新回数毎の学習誤差

3. 実験

DCNN によるセグメンテーションにおける重み付け誤差関数の有効性を評価実験により示す．

3.1. 実験概要

本実験で用いるデータセットは，Labeled Face in the Wild(LFW) である．LFW は，顔の各パーツについてラベルが与えられており，目，鼻，口，サングラス，帽子，髪と髭，顔の肌，背景の 8 クラスに対してラベルが与えられている．予備実験より，目の領域のセグメンテーションの性能が低いため，本実験では，目のクラスの重みを変更して学習する．

3.2. 実験結果

図 2 に更新回数毎の正解率を示す．図 1 の学習速度とは逆に，重みを小さくするほど正解率の向上が早まることがわかる．一方，顔パーツ全ての平均である Face Accuracy は重みの変更による影響を受けないことがわかる．また，重みを 0.8 に設定したとき，従来の DCNN より目の正解率が向上することが確認できる．

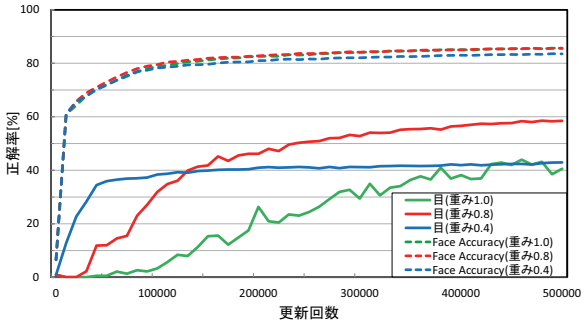

図 2：更新回数毎の認識精度

目の重みを変化させたときの目と顔全体の認識精度を表 1 に示す．表 1 より，重みの変更による顔全体の認識精度への影響は小さく，目の重みを 0.8 にしたとき，目の認識精度は従来の DCNN より約 18%上昇した．これは，学習の更新速度を遅らせることで，学習後半における誤差の逆伝播を，認識の困難なクラスに集中して行うためと考えられる．

表 1：目の重みを変更したときの認識精度 [%]

目の重み	1.2	1.0	0.8	0.6
目	10.46	40.55	58.49	52.68
顔全体	84.36	84.38	84.08	83.67

提案手法によるセグメンテーション例を図 3 に示す．図 3 より，目の重みの変更により，目の検出領域が増減することが確認できる．

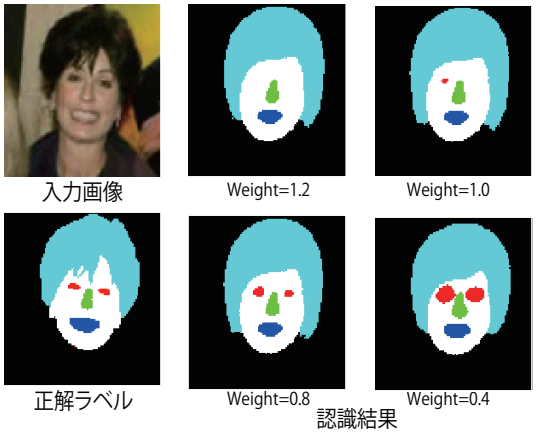

図 3：顔領域セグメンテーション例

4. おわりに

本稿では，DCNN による顔領域セグメンテーションに重み付け誤差関数を導入することで，学習速度を調節し，面積の小さいクラスである目の高精度な認識を実現した．今後は，他の画像認識問題に対して重み付け誤差関数を導入し，高精度な認識を実現する．

参考文献

[1] C. Farabet, *et al*., “Learning Hierarchical Features for Scene Labeling”, IEEE of PAMI, 2013