画像領域分割のための少量訓練データからの畳み込みニューラル ネットワークの解析的な初期化法

Analytical Initialization of Convolutional Neural Networks with Small Training Dataset for Image Segmentation

岳彦*2

Takehiko Sashida

指田

福田 竣^{*1} Shun Fukuda 中山 英樹 *1 Hideki Nakayama

*¹東京大学 大学院情報理工学系研究科 Graduate School of Information Science and Technology, The University of Tokyo *²コニカミノルタ株式会社 KONICA MINOLTA, INC.

3M2-3

Convolutional Neural Networks (CNN) have achieved high performances for many kinds of image recognition tasks. However, deep learning requires a large labeled dataset takes much time. One of the known solutions is using parameters pretrained with a large dataset, which makes it possible to use deep learning even if we only have a small dataset. Still, it has some problems. First, the effect is little when the domains of those datasets are largely different. Second, the architecture can be too large for training with an objective dataset because it is designed according to the pretrained model.

In this work, we propose a novel technique for image segmentation tasks, which calculates a pair of linear projections to minimize the distance between features of both RGB images and ground truths in the shared space. In experiments, our method achieves higher accuracies and faster convergences of training loss than random initialization.

1. はじめに

近年,画像認識の様々な問題で畳み込みニューラルネット ワーク(CNN)が大きな成果を挙げている.その理由の一つに 大規模な教師付きデータセットを用いて学習を行うことが挙げ られる.CNNは多層に渡って膨大なパラメータを持ち,学習 データに強く依存する.そのため,過学習を起こさないように 学習することは肝要である.多様な大規模データセットを用い ることで過学習を抑えられる.

小規模なデータセットを用いる場合には,過学習を防ぐ代 表的な方法に事前学習済みモデルを用いる方法がある.予め ImageNet など他の大規模データセットで学習したモデルを初 期値として,目的のデータセットで学習を行う方法である.こ の方法は,目的のデータセットの画像が事前学習に用いるデー タセットのものと近いほどその効果は高い.モデルは事前学習 に用いる大規模データセットに対して構築されたものを用いる ので,目的のデータセットにとっては大きすぎる可能性がある.

他のデータセットを用いず,小規模なデータセットのみを用 いて CNN の初期化を行う方法として,線形次元圧縮手法を活 用する方法がある.例えば[1]では,主成分分析やフィッシャー 判別分析を拡張させたフィッシャー重みマップを用いて畳み込 み層を初期化することで,小規模なデータセットに対するクラ ス分類問題で高い識別的な効果が得られている.

これまで画像領域分割問題に対しては線形次元圧縮手法に よる初期化法は十分に研究されていない.本研究では画像領域 分割問題に適した線形次元圧縮手法による初期化手法を新た に提案する.具体的には,RGB 画像と Ground truth 画像の 特徴ベクトルを射影した先でそれらの差の二乗ノルムを最小 にする射影を用いて畳み込み層の初期化を行う正準相関分析, 及びそれを入力が行列へと拡張させた正準重みマップを用いて 畳み込み層の初期化を行う.いくつかのデータセットを用いて 提案手法の有効性を示す.







図 1: 畳み込みと線形変換の関係.畳み込み処理は局所受容野 ごとにベクトル化した特徴マップと重みフィルタで内積をとる (線形変換)処理であると捉えられる.

2. 線形次元圧縮手法による初期化

小規模なクラス分類問題に対して,線形次元圧縮手法を用 いて畳み込み層を学習させる手法は高い成果をあげている.畳 み込み演算は,局所受容野ごとに特徴マップと重みフィルタと 内積をとっていると見なせることから,線形次元圧縮手法を用 いて初期化を行うことができる(図1).

その先行研究にフィッシャー重みマップ(FWM)を用いる ものがある[1].FWM[2]は,クラス間分散とクラス内分散の 比の距離を最大にする射影を求めるフィッシャー判別分析を, 入力が行列の場合へと拡張した方法である.教師クラスを用い る教師あり学習のため,クラス分類問題には適している手法で ある.しかし,画像領域分割問題では画像から対象クラスの領 域を正確に検出する必要があり,ピクセル間の位置関係を考慮 することは重要である.そのため,FWM は画像領域分割問題 には不向きである.

主成分分析 (PCA)を同様に拡張させた手法として固有重 みマップ (EWM)がある.FWM とは異なり一切教師情報を 用いない教師なしの手法で,様々な問題に対して用いることが できる.次節から EWM の詳細を説明する.

2.1 固有重みマップ

重みマップによる初期化では,特徴マップから得られた受容野ベクトル(図1参照) $x_{(x,y)}$ を列挙して構成した行列 $X = (x_{(1,1)}, \cdots, x_{(w-k+1,h-k+1)})$ を用いて行う.この行列を特徴行列と呼ぶ.また,h, wはそれぞれ特徴マップの縦,横の長さ,kはフィルタサイズを指す.

特徴行列の集合を ${X_i}_{i=1}^N$ に対して, $z_i = (X_i - \overline{X})^T a$ と 線形射影を行う場合を考える.固有重みマップ (EWM) では, 射影先で分散が最大とする規準のもとで射影ベクトル a の学 習を行う. EWM の規準 J_{EWM} は次式で定義される.

$$J_{\text{EWM}}(\boldsymbol{a}) = \frac{1}{N} \sum_{i=1}^{N} (\boldsymbol{z}_i - \overline{\boldsymbol{z}})^T (\boldsymbol{z}_i - \overline{\boldsymbol{z}})$$
(1)

$$= \boldsymbol{a}^{T} \left\{ \frac{1}{N} \sum_{i=1}^{N} (X_{i} - \overline{X}) (X_{i} - \overline{X})^{T} \right\} \boldsymbol{a} \qquad (2)$$

$$= \boldsymbol{a}^T \boldsymbol{\Sigma}_X \boldsymbol{a}. \tag{3}$$

 J_{EWM} を $a^T a = 1$ の制約の下で最大化させる時,射影ベクト ルaは次の一般化固有方程式の固有ベクトルとして得られる.

$$\Sigma_X \boldsymbol{a} = \lambda_{\text{EWM}} \boldsymbol{a}.$$
 (4)

この方程式から,大きい固有値に対応する固有ベクトルを順に 複数選び射影を構築する.

3. 提案手法

3.1 正準重みマップ

フィッシャー重みマップはピクセル間の関係を教師として与 えることが難しい問題がある.そこで本研究では,画像領域分 割に適した線形次元圧縮手法として正準重みマップ(CWM) を提案する.CWM はEWM やFWM と同様に正準相関分析 (CCA)を入力が行列の場合へと拡張した手法で,2つの異な るドメインに存在する特徴行列のペアに対して,これらの射影 先での差の二乗ノルムを小さくする線形射影を求める手法であ る.片方のドメインのデータとして Ground truth を与えるこ とで,ピクセル間の関係を教師情報として考慮した,画像領域 分割に適した教師あり学習による初期化を行うことができる. 3.1.1 定式化

CWM では,二つの特徴マップ F_i, G_i から 2.1 節で述べた 方法で特徴行列 X_i , Y_i を構成し,これらを共通空間へと射影 した際に差の二乗ノルムを最小にする射影ベクトル a,bを求 める.それぞれの射影を $s_i = (X_i - \overline{X})^T a$, $t_i = (Y_i - \overline{Y})^T b$ とする時,CWM の規準 J_{CWM} は次の通りに定義される.

$$J_{\text{CWM}}(\boldsymbol{a}, \boldsymbol{b}) = \frac{1}{N} \sum_{i=1}^{N} \|\boldsymbol{s}_{i} - \boldsymbol{t}_{i}\|^{2}$$
(5)
$$= \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{s}_{i}^{T} \boldsymbol{s}_{i} + \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{t}_{i}^{T} \boldsymbol{t}_{i} - \frac{2}{N} \sum_{i=1}^{N} \boldsymbol{s}_{i}^{T} \boldsymbol{t}_{i}$$
(6)

$$= \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{a}^{T} \left\{ (X_{i} - \overline{X})(X_{i} - \overline{X})^{T} \right\} \boldsymbol{a}$$

$$+ \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{a}^{T} \left\{ (X_{i} - \overline{X})(X_{i} - \overline{X})^{T} \right\} \boldsymbol{a}$$
(7)

$$+ \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{b}^{T} \left\{ (Y_{i} - Y)(Y_{i} - Y)^{T} \right\} \boldsymbol{b}$$
(7)
$$- \frac{2}{N} \sum_{i=1}^{N} \boldsymbol{a}^{T} \left\{ (X_{i} - \overline{X})(Y_{i} - \overline{Y})^{T} \right\} \boldsymbol{b}$$

$$= \frac{1}{N} \sum_{i=1}^{L} a^{T} \left\{ (X_{i} - X)(Y_{i} - Y) \right\} b$$
$$= a^{T} \Sigma_{XX} a + b^{T} \Sigma_{YY} b - 2a^{T} \Sigma_{XY} b.$$
(8)



図 2: CWM の入力に必要な Ground truth 特徴量の構成方法 1(GT 構成 1). Ground truth を各ピクセルごとにそのクラ スの教師信号に関して one-hot ベクトルにし, クラス数のチャ ネルを持つ特徴量を与える.

 J_{CWM} が $\frac{1}{N} \sum_{i=1}^{N} \|\boldsymbol{s}_i\|^2 = 1, \frac{1}{N} \sum_{i=1}^{N} \|\boldsymbol{t}_i\|^2 = 1$ の制約の下で最小となる時,射影ベクトル $\boldsymbol{a}, \boldsymbol{b}$ は次の一般化固有方程式の固有ベクトルとして得られる.

$$\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}\boldsymbol{a} = \lambda_{\text{CWM}}\Sigma_{XX}\boldsymbol{a},\tag{9}$$

$$\Sigma_{YX} \Sigma_{XX}^{-1} \Sigma_{XY} \boldsymbol{b} = \lambda_{\text{CWM}} \Sigma_{YY} \boldsymbol{b}.$$
 (10)

これらの方程式から,大きい固有値に対応する固有ベクトルを 複数選び射影を構築する.しかし,射影は高々*X_i*,*Y_i*の小さい 方の要素数までしか得られないので,片方の特徴行列の要素数 が小すぎる場合には十分な数の射影を得られない.

また, $\Sigma_{XX} \Sigma_{YY}$ がフルランクでない場合には,正則化項 として αI を加える.これにより安定して学習を行える.本研 究では $\alpha = 0.001$ とした.

3.1.2 Ground truth 特徴量の構成方法

CWM は二つの特徴量を近づける操作を行うため,効果的な 教師あり学習を行うためには適した Ground truth (GT)特 徴量を構成することが重要である.本研究ではエンコーダー・ デコーダー型のモデル [3] に対して GT の構成方法を二つ提案 する.

一つ目の方法(GT構成1)は,GTを各ピクセルごとにそのクラスの教師信号に関して one-hot ベクトルにし,クラス数のチャネルを持つ特徴量を与える方法である(図2).対象の層によってはRGB 画像側の特徴量は Max pooling によりサイズが元のものより小さくなっているが,その場合にはGT側も同じ回数 Max pooling を適用させて同等のサイズにしてから用いる.どの層の初期化に対しても直接GTが与えられる.CWMで得られる射影の数は最大でも2つの特徴行列の小さい方の要素数となるため,この構成方法ではGTの特徴行列の要素数が小さく,十分な数の重みフィルタを得られない.そのため,他の方法で補う必要がある.

二つ目の方法(GT構成2)は,GTに対してもそれを入力 とするモデルを構築し,特徴抽出を行ったものを用いる方法で ある(図3).まず,モデルのエンコーダーと同じ構造を持つ エンコーダーからなるモデルをGTに対して構成する.ただ し,最初の畳み込み層の入力チャネル数はクラス数にする.エ ンコーダーの畳み込み層に対して初期化を行う際には,GT側 の畳み込みも同位置にあるものを用いる.デコーダーの畳み込 み層に対して初期化を行う際には,GT側はそれと対称に位置 するエンコーダーの畳み込み層を用いる.



図 3: CWM の入力に必要な Ground truth 特徴量の構成方法 2(GT 構成 2). Ground truth 側にも同等のエンコーダーで 構成され,対称的に位置する畳み込みペアに対して正準重み マップを用いて初期化を行う.

GT 構成 1 と比較すると GT の特徴行列の要素数は大きい ので,多くの重みフィルタを得ることができる.しかし,GT 側のモデルに対しても初期化を行うので,一つの層につき二つ の固有方程式を解く必要がある.そのため,GT 構成1より2 倍以上の計算コストがかかる.

3.2 ReLUによる情報欠落に対する対処

CNNでは活性化関数として ReLU がよく用いられる.ReLU は各畳み込み層の直後に置かれ,負値の情報を欠落させる処理 を行う.そのため,線形次元圧縮手法を用いて畳み込み層の初 期化を行ったとしても,ReLUにより半分の情報を失うため, 初期化の効果は半減される.

線形次元圧縮手法で得られた射影ベクトル a に加えて -a を用いることで両符号の情報を残せる.また,この操作を行う ことで2倍の数の射影ベクトル数が得られるため,前節で述 べた重みフィルタ数が不足する問題についても緩和される.

4. 実験

4.1 データセット

網膜検出用の DRIVE[4] と細胞画像領域分割用の Warwick-QU[5] の二つのデータセットを用いて実験を行った.DRIVE は訓練画像が 20 枚,テスト画像が 20 枚で構成されており,Warwick-QU は訓練画像が 85 枚,テスト画像が 80 枚で構成 されている.DRIVE のクラスは背景か網膜かの 2 クラスであ る.Warwick-QU は元は背景か細胞かの 2 クラスであったが, 細胞クラスの領域に対して 4 近傍に膨張させ,細胞領域の増え た部分を輪郭クラスとする.それをさらに輪郭クラスに対して 3 回 8 近傍に膨張させたものを実験に用いる.評価指標として, DRIVE に対しては PR 曲線に対する AUC,Warwick-QU に 対しては Dice 係数を用いる.

学習を行う際には,予め訓練データから検証用データとして Warwick-QU, DRIVE に対してそれぞれランダムに 4,15 枚選び抜く.残りのデータを訓練データとして扱う.

4.2 モデル

画像領域分割の既存モデルに SegNet がある [3]. SegNet は VGGNet の最初の 5 層とそれを反転したものを連結して構成 されている.本実験ではこれを参考にして, VGGNet の最初 の 4 層のもの及びそれを反転させて連結して構成したモデル, SegNet4 を用いる. SegNet4 の構成図を図 4 に示す.

加えて, SegNet4 の中間のチャネル数を $2 \times$ クラス数 $\times 3^2$



図 4: SegNet4 のモデル図.

へと縮小させた SegNetMini4 も用いる.DRIVE, Wariwick-QU に対してはクラス数がそれぞれ2,3のため, SegNetMini4 の中間のチャネル数は36,54となる.この数はGT構成1を 用いた際に得られる重みフィルタの最大数である.GT構成1 を用いて CWM で初期化を行った際の純粋な効果を図るため に SegNetMini4 を用いて検証を行う.

4.3 初期化方法

SegNetMini4 の畳み込み層に対しては, ランダム, 主成分分 析(PCA), 固有重みマップ(EWM), 正準相関分析(CCA), 正準重みマップ(CWM)を用いて, 比較評価する. ランダム 初期化法として, Xavier[6]を用いる. Xavier は CNN のラン ダム初期化法として広く使われている. CWM で初期化を行 う際には, GT 構成1,2 で GT 特徴量を構成する場合をそれ ぞれ, CWM, CWM(GT2)とする.

PCA, CCAは、それぞれ EWM、CWM のベースの手法で、 PCA、CCA ではベクトルを入力とするのに対して、EWM、 CWM では行列を入力とする.この違いによる差分を検証す るために、これらの初期化法を比較評価する.なお、CCA で 初期化を行う際には GT 構成1の方法で GT 特徴量を構成し たものを用いる.

SegNet4 の畳み込みに対しては, ランダム, CWM&EWM, CWM(GT2), 事前学習済みモデルを用いて初期化する. CWM&EWM は GT 構成 1 を用いて CWM で初期化した 際に得られる射影ベクトルの不足する分を EWM で補う方法 を指す.CWM(GT2)においても, 1 段目の畳み込みに対して 初期化を行う際には射影ベクトルが不足するが,その分はラ ンダム初期化で補う.事前学習済みモデルによる初期化では, PASCAL VOC2012 で学習されたパラメータを用いる.

線形次元圧縮手法で初期化を行う際には,各画像からサイズ が 192×192 のパッチを中心が等間隔に離れさせて,DRIVE に対しては $2 \times 2 = 4$ 枚,Warwick-QUに対しては $4 \times 3 = 12$ 枚切り出したもの,及びそれらを 90 度ずつ回転させたもの,左 右反転させたものを用いた.また,バッチサイズは 18 とした.

4.4 評価結果

訓練データ数を変化させて,初期化手法の有効性の検証実 験を行った.バッチサイズを,DRIVE に対しては SegNet4, SegNetMini4 のいずれを用いる場合においても4,Warwick-QU に対しては SegNet4 を用いる場合では3,SegNetmini4 を用いる場合では5とした.

SegNetMini4 を用いた際の DRIVE 及び Warwick-QU の予 測結果を表1,2 に示す. これらの結果から訓練データ数が小 さい時に CWM による初期化が有効であることが読み取れる. しかし,CCA と CWM,GT 構成1とGT 構成2に有意差は なかった.

次に SegNet4 を用いた際の DRIVE 及び Warwick-QU の 予測結果を表3,4 に示す.先の結果と同様にいずれのデータ セットにおいても訓練データ数が少ない場合において,CWM による初期化が効果的であると言える.しかし,いずれのケー

表 1: DRIVE (データセット) & SegNetMini4 (モデル)を用 いた時の AUC.

	訓練データサイズ			
初期化法	4	10	16	
ランダム	0.790	0.891	0.897	
PCA	0.770	0.889	0.893	
EWM	0.757	0.884	0.895	
CCA	0.873	0.898	0.897	
CWM	0.872	0.894	0.903	
CWM(GT2)	0.872	0.895	0.897	

表 2: Warwick-QU (データセット) & SegNetMini4 (モデル) を用いた時の Dice 係数.

	訓練データサイズ			
初期化法	15	45	70	
ランダム	0.582	0.710	0.706	
PCA	0.567	0.716	0.718	
EWM	0.556	0.684	0.711	
CCA	0.646	0.671	0.7188	
CWM	0.583	0.703	0.753	
CWM(GT2)	0.662	0.721	0.727	

表 3: DRIVE (データセット) & SegNet4 (モデル)を用いた 時の AUC.

	訓練データサイズ			
初期化法	4	10	16	
ランダム	0.777	0.881	0.890	
CWM&EWM	0.858	0.892	0.897	
CWM(GT2)	0.872	0.893	0.902	
事前学習済み	0.896	0.897	0.905	

表 4: Warwick-QU(データセット)&SegNet4(モデル)を 用いた時の Dice 係数.

	訓練データサイズ			
初期化法	15	45	70	
ランダム	0.558	0.754	0.747	
CWM&EWM	0.659	0.734	0.758	
CWM(GT2)	0.616	0.751	0.736	
事前学習済み	0.737	0.777	0.771	

スにおいても事前学習済みモデルで初期化した場合より劣る結 果となった.

最後に初期化による収束の速さについて検証をする.図5に DRIVE に対して訓練データサイズを4とした時の損失の推移 を載せる.CCA,CWM で初期化を行った際には,ランダム で初期化した時より速く収束することが読み取れる.

4.5 定性的な評価

SegNetMini4 を用いて, DRIVE, Warwick-QU に対して, 訓練データ数がそれぞれ4,15として学習を行った際の予測の 結果例を図6,7に示す.DRIVEの例では,AUCの結果が低 いほどエッジの過検出が起きており,光による明暗の差のある 箇所で過検出が起きている.Warwick-QUの例では,Dice係 数が低いほど,大きな細胞に対して大きくその領域を予測する ことを苦手としている傾向が読み取れる.



図 5: DRIVE (訓練データ数:4)を用いて SegNetMini4 を学 習させた際の検証誤差の推移.



図 6: DRIVE の予測結果例.訓練データ数が4で,モデルは SegNetMini4を用いた.

RGB	Random	EWM	CWM	CWM2	Ground truth
RA			the c		

図 7: Wariwick-QUの予測結果例.訓練データ数が15で,モ デルは SegNetMini4 を用いた.

5. まとめ・今後の展望

本研究では,画像領域分割問題に対する CNN の初期化方法 として,正準相関分析をベースとした正準重みマップ(CWM) を提案した.2つの小規模な医用画像データセットを用いて検 証実験を行ったところ,データ数が少ない場合に対して収束の 高速化効果,評価指標の向上が確認された.

本研究では,どの層に対して提案手法による初期化が効果 的であったかについては検証を行えていない.今後の展望とし てまずこの部分を解明し,さらに事前学習済みモデルなどの他 の初期化法と組み合わせて,より効果的な方法について検証を 行っていきたい.

参考文献

- H. Nakayama, "Efficient discriminative convolution using Fisher weight map", In *Proc. BMVC*, 2013.
- [2] Y. Shinohara, N. Otsu, "Facial expression recognition using Fisher weight maps", In *IEEE FG*, 2004.
- [3] V. Badrinarayanan, A. Kendall, R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation", In *Proc. CVPR*, 2015.
- [4] J. J. Staal, M. D. Abramoff, M. Niemeijer, M. A. Viergever, and B. van Ginneken. "Ridge based vessel segmentation in color images of the retina", *IEEE Transactions on Medical Imaging*, 2004.
- [5] K. Sirinukunwattana, J. P. W. Pluim, H. Chen, X. Qi, P.-A. Heng, Y. B. Guo, L. Y. Wang, B. J. Matuszewski, E. Bruni, U. Sanchez, et al. "Gland segmentation in colon histology images: The glas challenge contest", arXiv preprint arXiv:1603.00275, 2016.
- [6] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks", In Proc. AISTATS, 2010.