

CNNの順・逆伝搬値とCRFを利用した弱教師領域分割

下田 和^{1,a)} 柳井 啓司¹

1. はじめに

領域分割は、画像における認識対象のカテゴリごとにピクセル単位でラベリングを行うタスクである。近年、Convolutional Neural Network(CNN)を応用する手法により領域分割タスクの精度は大きく向上した。しかし、高精度な領域分割手法の多くは、ピクセル単位の領域の教師情報を必要とする。このような高度な教師情報の付与には大きな労力を要し、認識対象の拡張など応用面における課題を抱えている。一方で、ピクセル単位の教師情報と比較して、画像レベルのラベル教師情報の付与は容易である。一般に、領域分割においてピクセル単位の教師情報による学習を完全教師あり学習、画像レベルのラベル教師情報のみを用いる学習を弱教師あり学習と呼ぶ。弱教師あり学習による高精度な領域分割を実現することができれば、学習データの収集が容易となり領域分割をより幅広く応用することが可能になる。本研究は、弱教師あり領域分割の精度向上に関する研究を行った。

CNNによる弱教師あり領域分割手法としては、Simonyanらの手法 [16] がある。Simonyanらは、カテゴリ c についての信号 S_c を Backpropagation により逆伝搬させたとき、画像レベル I_0 における逆伝搬値が物体の位置を反映していることを示した。

$$w_i = \left. \frac{\partial S_c}{\partial I} \right|_{I_0} \quad (1)$$

しかし、この逆伝搬値は異なるカテゴリによる応答の差が小さく(詳細はセクション 4 に後述)、実践的な領域分割手法として扱うには大きな課題を抱えている。そこで、本研究ではこの手法を改良し、逆伝搬値が位置推定において有効であることを示した。本研究においては、

- (1) 順伝搬値と逆伝搬値の統合による方法、
- (2) 逆伝搬値の差分による方法

の 2 つの逆伝搬値による位置推定手法を提案し、Pascal VOC 2012 データセットにおいて手法の性能を評価した。(1)の手法においては順伝搬値による位置推定結果を、逆伝搬値を活用することで改善可能であることを、[14]におけるプロポーザルによる平滑化手法と比較することにより示した。(2)の手法においては差分をとることで逆伝搬値のみで位置推定が可能であることを示した。特に、(2)の手法に関しては、Pascal VOC2012の弱教師領域分割タスクにおける State-of-the-art を達成した。

¹ 電気通信大学 大学院情報理工学研究所 総合情報学科 〒182-8585 東京都調布市 調布ケ丘 1 丁目 5-1

^{a)} shimoda-k@mm.inf.uec.ac.jp

2. 関連研究

2.1 完全教師あり領域分割

CNNを用いた領域分割手法で高精度を達成した手法として、Zoom Out Featres(ZOF)[10]とFully Convolutional Networks(FCN)[9]がある。Mostajabi et al.[10]はCNNにおける中間層の順伝搬値を Super pixel に割り当てることで、領域分割に応用可能な特徴量として扱った。また、中間層において圧縮されている位置情報を Upsampling により補間することで、各階層の順伝搬値を統合した。

Fully Convolutional Network(FCN)はSermanet et al.[15]によって提唱された手法である。CNNの高層における Inner product 層を Convolution 層の一種として扱うことで、入力を任意のサイズとする。入力画像のサイズを大きくすることで、CNNの最終的な出力はベクトルから位置情報を持つ行列となり、中間層における順伝搬値は密になる。また、Long et al.[9]はこの出力を直接領域分割の学習に用いることで高精度な領域分割が可能であることを示した。

2.2 弱教師あり領域分割

Pathak et al. [13]、Pinheiro et al.[14]は、FCN[9]を弱教師あり学習に応用した。FCNの出力は粗い位置情報を保持した行列となるが、Pathak et al.[13]はGlobal-Max-Poolingにより行列の出力をベクトルに変換することで、弱教師ありによる学習を行った。Pinheiro et al.[14]は、Overfeat [15]のネットワークを改良し、[13]と同様にFCNを弱教師で学習したが、後処理としてプロポーザルによる平滑化を行い領域分割結果の精度を向上させた。Papandreou et al. [11]、Pathak et al. CCNN [12]は、誤りを含む位置推定結果を教師情報として扱うことで、FCNの出力を直接学習可能であることを示した。

これらの弱教師あり領域分割の手法は、FCNにおける順伝搬値を用いており、Simonayn et al. により提唱された逆伝搬値を位置推定に用いていない。本研究は逆伝搬値を位置推定に活用し、ZOF、FCNなどの順伝搬値による位置推定手法と組み合わせることで、より詳細な位置推定が行えることを示した。また、逆伝搬値について差分をとることで、順伝搬値を用いずに高精度な位置推定を行った。

3. (1) 順伝搬値と逆伝搬値の統合による方法

順伝搬値の位置推定結果に、逆伝搬値を組み合わせることで、位置推定の改善を狙った。また、順伝搬値としては、ZOFとFCNの2種類の特徴を用い、異なる方法で逆伝搬値との統合を行った。

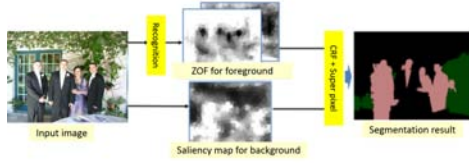


図 1 Processing flow of the ZOF base method.

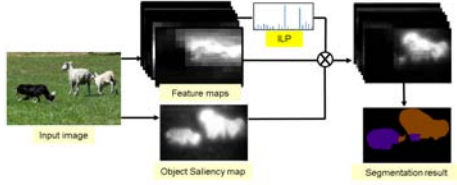


図 2 Processing flow of the FCN base method.

3.1 逆伝搬値による位置推定手法の改良

逆伝搬値を順伝搬値と組み合わせるうえで、Simonyan et al. [16] の手法について改良を行った。Simonyan et al. [16] により提唱された画像レベル I_0 における逆伝搬値は疎な位置推定結果となる傾向がある。そこで、本研究では Upsampling を用いて中間層の特徴量を補間しこれらを統合した。カテゴリ c における信号 S について、階層 L_i の逆伝搬値 v_i は以下の式で表すことができる。

$$v_i = \frac{\partial S_c}{\partial L} \Big|_{L_i} \quad (2)$$

v_i を Upsampling することで、画像と同じサイズ w_i とする。 $M_{i,x,y}^c$ を以下の式から計算した。

$$M_{i,x,y}^c = \max |w_{i,h_i(x,y,k)}| \quad (3)$$

$h_i(x,y,k)$ は w_i の要素のインデックスであり、 k はカーネルを表している。ただし、 $M_{i,x,y}^c$ はカテゴリごとの応答が小さいために $M_{i,x,y}^c \simeq M_{i,x,y}$ とした。この $M_{i,x,y}$ を統合することで、 $g_{x,y}$ を得る。 L は統合するためのレイヤー、 α はスカラー値である。

$$g_{x,y} = \frac{1}{L} \sum \tanh(\alpha \cdot M_{i,x,y}) \quad (4)$$

また、本研究では誤差を伝搬させる際に、単純な Backpropagation ではなく、Guided backpropagation [17] を採用した。

3.2 順伝搬値と逆伝搬値の統合

順伝搬値と逆伝搬値を統合する。ZOF と FCN それぞれの順伝搬値による特徴について異なる手順による統合を行った。図 1、図 2 にそれぞれの特徴量における統合の手順の概要を示した。

3.2.1 ZOF

Zoom Out Feature (ZOF) [10] は、近年の完全教師ありの領域分割において高い精度を達成した手法の一つである。本研究ではこの ZOF [10] を従来手法の Multiple instance learning (MIL) を用いて、弱教師あり学習に応用した。画像の各 Super pixel がポジティブ領域かネガティブ領域のどちらかに属していると考え、スタンダードな MIL の手法である miSVM [1] を適用し、各クラスにおける SVM のパラメータを得た。

また、単純な閾値処理でなく、Super pixel についての

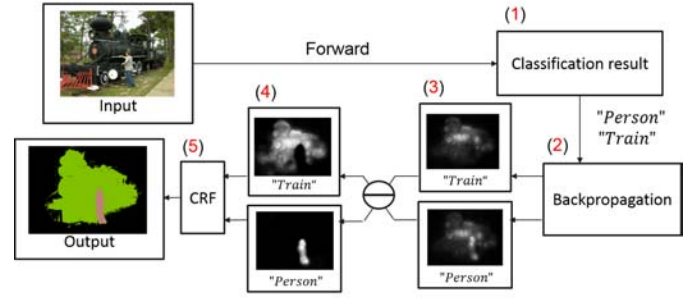


図 3 逆伝搬値の差分による弱教師あり領域分割の手順概要

CRF による統合を行った。画像 I の Super pixel におけるラベルを y_p であるとして、 y を全ての y_p のベクトルとするとエネルギー関数は以下の式で表すことができる。

$$E(y|I) = \sum_{p \in P} U(y_p|I) + \alpha \sum_{p,q \in N} V(y_p, y_q|I) \quad (5)$$

P は画像の全ての Super pixel を示しており、 N は隣接している Super pixel を示している。 $U(\cdot)$ は Unary 項、 $V(\cdot)$ は Pairwise 項である。係数 $\alpha = 15$ として重み付けをしている。 $Pr(y_p|I)$ は画像 I における各 Super pixel y_p のカテゴリの確率値としたときに、Unary 項は $U(y_p|I) = -\log Pr(y_p|I)$ として定めた。前景のカテゴリの確率値は ZOF と miSVM [1] から求め、背景カテゴリの確率値は、セクション 3.1 における逆伝搬値から求めた。

Pairwise 項は以下の式 [5] に従う。

$$V(y_p, y_q|I) = \left(\frac{Ld(p,q)}{1 + \|p - q\|} \right) [y_p \neq y_q] \quad (6)$$

$\|p - q\|$ は super pixel p と q における LUV 色空間における距離、 $Ld(p,q)$ は superpixel p と q で共通している境界の長さである。

3.2.2 FCN

完全教師ありで高精度を達成している Fully Convolutional Network (FCN) を弱教師あり領域分割に応用する。FCN においては、入力画像のサイズを大きくすると、出力が位置情報を持つ行列となる。この出力について、Global-Max-Pooling (GMP) を適用することで、行列をベクトルへと変換し、各カテゴリの確率値として扱うことで画像ラベルの教師情報から学習を行った。また、本研究においては FCN の粗い出力に対して、セクション 3.1 における逆伝搬値を適応することで、位置推定の精度向上を行った。ピクセル (x,y) における FCN の出力を $f_{x,y}$ 、逆伝搬値を $g_{x,y}$ とすると、位置推定結果 $h_{x,y}$ は以下の式に従う。

$$h_{x,y} = \begin{cases} k, & \text{if } \arg \max_{k \in C} f_{x,y}^k * g_{x,y} > \delta \\ k_{bg} & \text{otherwise} \end{cases} \quad (7)$$

C はクラス集合であり、 δ は閾値である。

4. (2) 逆伝搬値の差分による方法

本セクションにおいては、逆伝搬値の差分から Deep Class-specific Response Map (DCRM) を生成する。また、弱教師あり領域分割において有効であることを示す。図 3 に手法の概要を示した。

4.1 逆伝搬値の差分によるクラス応答の鮮明化

Simonyan et al.[16] の手法により得られる逆伝搬値は異なるカテゴリによる応答の差が小さい。これは、教師信号のクラスと CNN が学習している他のクラスで、認識に寄与している領域に共通部分があるためであると考えられる。そこで、本研究ではこの問題を解決するために、逆伝搬値の差をとることで、クラスごとに共通している領域を排除した。式 3 から、それぞれのクラス c における差分によるマップ \tilde{M}_i^c を得る。

$$\tilde{M}_{i,x,y}^c = \sum_{c' \in D} \max(M_{i,x,y}^c - M_{i,x,y}^{c'}, 0) [c \neq c'], \quad (8)$$

ただし、 D は選択したクラス集合である。また、本セクションにおいては、複数のサイズの入力画像により得られる逆伝搬値を統合し、これにより位置推定結果の改善が可能であることを示した。最終的なクラスマップ \hat{M}^c を以下の式により得る。

$$\hat{M}_{x,y}^c = \frac{1}{|T||L|} \sum_{j \in T} \sum_{i \in L} \tanh(\alpha \tilde{M}_{j,i,x,y}^c), \quad (9)$$

L は統合するためのレイヤー、 T はスケールについての集合、 α は固定値である。ただし、 $\tilde{M}_{j,i}$ は式 10 の計算の前に画像サイズにリサイズし、正規化を行っている。また、セクション 3.1 と同じく Guided Backpropagation による伝搬を行った。

4.2 全結合 CRF

CRF はラベルの拡張手法であり、粗い領域分割結果の平滑化手法として用いられる。本研究では、全結合 CRF[8] により、領域推定結果を改善した。エネルギー関数は以下の式で表すことができる。

$$E(c) = \sum_i \theta_i(c_i) + \sum_{ij} \theta_{ij}(c_i, c_j) \quad (10)$$

Unary 項は逆伝搬値 M_i^c から求め、 $\theta_i(c_i) = -\log(\tanh(\alpha \cdot M_i^c))$ とした。 c はピクセルに割り当てられたラベルである。

本研究では *target* クラス + 背景クラスのラベルの領域拡張を行う。背景クラス以外の逆伝搬値の情報を統合することで、背景クラスの確率値 M^{bg} とした。 $M^{bg} = 1 - \max_{c \in \text{target}} M_{x,y}^c$ ただし、*target* はグローバルプーリングにおける分類結果の閾値から決定した。Pairwise 項は [8] に従った。

5. 実験

5.1 Dataset

領域分割のベンチマークのデータセットとして PASCAL VOC 2012 [4] を用いた。PASCAL は 20 の異なるクラスからなるデータセットであり、特に領域分割においては Background クラスを含む 21 のクラスを認識する必要がある。PASCAL には 1464 枚の train 画像、1449 枚の validation 画像、1456 枚の test 画像が領域分割として割り当てられている。また、[6] により提供された 10582 枚の train_aug 画像と 12031 枚の trainval_aug 画像を用いた。結果は一般に公開されている Pascal のサーバーにおける評価結果に基づいており、Mean IoU における精度となっている。

表 1 Results on PASCAL VOC 2012 val set.

Methods	additional image																			mIoU		
	big	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	motor	person	plant	sheep	sofa		train	tv
Weakly Supervised:																						
MIL-FCN [13]	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	25.7	
EM-Adapt [11]	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	38.2	
CCNN [12]	-	65.9	23.8	17.6	22.8	19.4	36.2	47.3	46.9	47.0	16.3	36.1	22.2	43.2	33.7	44.9	39.8	29.9	33.4	22.2	38.8	34.5
MIL-sppd [14]	-	77.2	37.3	18.4	25.4	28.2	31.9	41.6	48.1	50.7	12.7	45.7	14.6	50.9	44.1	39.2	37.9	28.3	44.0	19.6	37.6	35.0
MIL-bb [14]	-	78.6	46.9	18.6	27.9	30.7	38.4	44.0	49.6	49.8	11.6	44.7	14.6	50.4	44.7	40.8	38.5	26.0	45.0	20.5	36.9	34.8
MIL-seg [14]	-	79.6	50.2	21.6	40.6	34.9	40.5	45.9	51.5	60.6	12.6	51.2	11.6	56.8	52.9	44.8	42.7	31.2	55.4	21.5	38.8	36.9
【方法 (1)】																						
ZOF with GBP (ours)	-	70.6	44.4	24.7	37.5	16.4	33.3	60.6	35.5	58.8	5.5	45.5	15.9	53.4	41.1	54.8	39.6	24.2	52.1	18.4	38.6	27.5
FCN with MCG (ours)	-	71.0	21.9	18.5	22.0	12.8	34.6	37.5	43.3	47.1	17.7	38.5	29.4	40.9	43.3	40.7	38.7	29.0	35.6	22.8	36.6	25.0
FCN with GBP (ours)	-	76.8	40.0	28.1	38.6	24.6	39.7	37.3	50.2	51.4	23.9	47.2	25.8	53.6	49.1	53.9	45.1	36.0	48.1	30.0	35.8	33.9
【方法 (2)】																						
DCRM w/o CRF (ours)	-	72.5	35.0	23.7	33.6	20.9	32.8	57.6	49.2	51.4	14.7	42.9	32.8	51.9	43.1	53.2	47.6	35.2	47.7	28.6	40.3	36.7
DCRM w/ CRF (ours)	-	76.7	45.1	24.6	40.8	23.0	34.8	61.0	51.9	52.4	15.5	45.9	32.7	54.9	48.6	57.4	51.8	38.2	55.4	32.2	42.6	39.6

表 2 Results on PASCAL VOC 2012 test set.

Methods	additional image																			mIoU		
	big	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	motor	person	plant	sheep	sofa		train	tv
Fully Supervised:																						
O2P [2]	85.4	69.7	22.3	45.2	44.4	49.6	66.7	57.8	56.2	13.5	46.1	32.3	41.2	59.1	55.3	51.0	36.2	50.4	27.8	46.9	44.6	
SDS [7]	86.3	63.3	25.7	63.0	39.8	59.2	70.9	61.4	54.9	16.8	45.0	48.2	50.5	51.0	57.7	63.3	31.8	58.7	31.2	55.7	48.5	
FCN-8s [9]	-	76.8	34.2	68.9	49.4	60.3	75.3	74.7	77.6	21.4	62.5	46.8	71.8	63.9	76.5	73.9	45.2	72.4	37.4	70.9	55.1	
DeepLab [3]	93.1	84.4	54.5	81.5	63.6	65.9	85.1	79.1	83.4	30.7	74.1	59.8	79.0	76.1	83.2	80.8	59.7	82.2	50.4	73.1	63.7	
Size tag Supervised:																						
CCNN w/size [12]	-	42.3	24.5	56.0	30.6	39.0	58.8	52.7	54.8	14.6	48.4	34.2	52.7	46.9	61.1	44.8	37.4	48.8	30.6	47.7	41.7	
Weakly Supervised:																						
MIL-FCN [13]	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	24.9	
EM-Adapt [11]	-	76.3	37.1	21.9	41.6	26.1	38.5	50.8	44.9	48.9	16.7	40.8	29.4	47.1	45.8	54.8	28.2	30.0	44.0	29.2	34.3	46.0
CCNN [12]	-	21.3	17.7	22.8	17.9	38.3	51.3	43.9	51.4	15.6	38.4	17.4	46.5	38.6	53.3	40.6	34.3	36.8	20.1	32.9	38.0	
MIL-ILP-seg [14]	78.7	48.0	21.2	31.1	28.4	35.1	51.4	55.5	52.8	7.8	56.2	19.9	53.8	50.3	40.0	38.6	27.8	51.8	24.7	33.3	46.3	
【方法 (1)】																						
ZOF with GBP (ours)	71.1	48.4	24.4	48.5	15.2	38.2	65.6	32.8	57.9	5.1	43.8	18.2	46.2	48.7	50.4	35.7	22.5	41.7	19.1	29.2	27.1	
FCN with MCG (ours)	71.9	21.8	18.4	25.4	14.9	35.2	40.0	39.7	41.5	13.4	36.4	29.9	36.5	45.4	41.3	38.7	26.9	34.5	19.7	29.8	33.3	
FCN with GBP (ours)	78.0	35.8	28.5	45.7	25.9	43.1	40.1	46.9	49.1	16.3	42.4	29.6	50.8	51.3	57.2	44.4	28.9	44.8	27.5	31.6	36.2	
【方法 (2)】																						
DCRM w/o CRF (ours)	73.9	34.1	24.4	39.6	18.0	37.6	57.8	49.0	51.5	13.3	42.3	33.5	47.8	44.2	63.7	44.3	34.5	48.3	31.2	35.7	37.1	
DCRM w/ CRF (ours)	78.1	43.8	26.3	49.8	19.5	40.3	61.6	53.9	52.7	13.7	47.3	34.8	50.3	48.9	69.0	49.7	38.4	45.7	134.0	38.0	40.0	

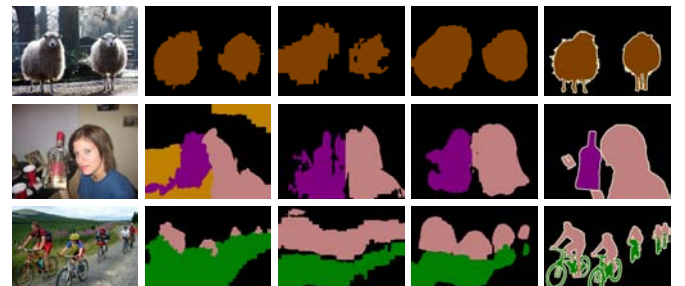


図 4 それぞれ、左から入力画像、ZOF with GBP, FCN with MCG, FCN with GBP の結果、正解領域となっている。

5.2 順伝搬値と逆伝搬値の統合による領域分割の結果

表 1、2 に Pascal データセットにおける結果を示した。Feature maps と GBP を組み合わせた手法である ZOF with GBP、FCN with GBP は、他の弱教師あり領域分割と比較して高い精度を達成し、追加画像を用いている [14] に近い精度となっている。本手法と [14] の間には学習画像枚数の違いや学習画像枚数などいくつかの相違点がある。そこで、本研究は、VGG16 ネットワークの FCN を Global-Max-Pooling を用いて学習したモデルについて、MIL-seg[14] と同様の手法で実験を行った (FCN with MCG)。FCN with GBP と FCN with MCG における相違点は平滑化に用いる Priors の部分のみであるが、FCN with GBP は 7% 以上精度が向上しており、逆伝搬値がプロポーザルによる平滑化と比較して有効であったことがわかる。図 5 に GBP と MCG による Priors の比較を示す。

5.3 逆伝搬値の差分による領域分割の結果

本研究は、中間層の逆伝搬値、入力画像のサイズについ

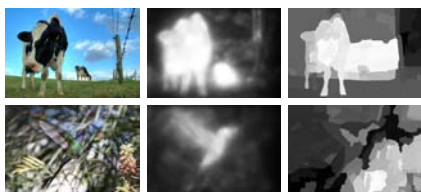


図 5 それぞれ、(左) 入力画像、(中央) 逆伝搬値、(右)MCG priors.

表 3 中間層における比較

layer	mean IoU
block1 (conv1.1, conv1.2)	5.5
block2 (conv2.1, conv2.2)	21.5
block3 (conv3.1, conv3.2, conv3.3)	32.5
block4 (conv4.1, conv4.2, conv4.3)	40.3
block5 (conv5.1, conv5.2, conv5.3)	26.3
average block 1,2,3,4,5	41.3
average block 2,3,4,5	42.2
average block 3,4,5	42.8
average block 4,5	42.5
average block 3,4	37.97

表 4 入力画像サイズにおける比較

input image size	mean IoU
(1) 300 × 300	34.5
(2) 400 × 400	41.0
(3) 500 × 500	42.4
(4) 600 × 600	41.8
(5) 700 × 700	40.0
(6) 800 × 800	34.5
average (1),(2),(3)	41.1
average (2),(3)	42.9
average (2),(3),(4)	43.5
average (3),(4)	42.9
average (3),(4),(5)	42.5
average (3),(4),(5),(6)	42.8

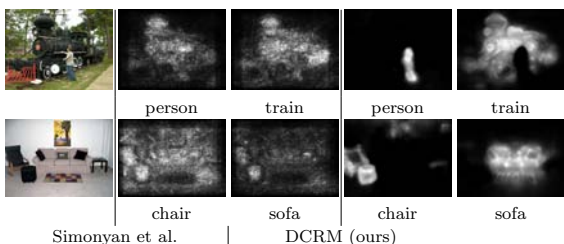


図 6 本手法と Simonyan et al. [16] の比較

表 5 逆伝搬値の差分による効果

class N	0	1	2	3	4	5	10	15
mean IoU	38.2	42.2	43.5	44.1	44.2	44.0	43.7	43.3

表 6 Simonyan et al. [16] との比較。

method \ α	2	2.5	3	4	5	6	7	8	9	10	15
Simonyan et al.	-	-	10.0	20.6	28.3	32.7	33.4	33.8	33.8	33.3	28.7
DCRM (ours)	40.0	44.0	44.1	40.6	36.4	-	-	-	-	-	-

て、複数の組み合わせを試し、結果を比較した。表 4、3 に結果を示す。Conv3、Conv4、Conv5 の逆伝搬値が位置推定において有効であり、入力画像のサイズを変化させ、それを統合することで精度を向上させることが可能であることを示した。

また、本研究では、逆伝搬値の差分による位置推定を行った。このとき、式 8 におけるクラス集合 D は認識結果における各クラスの確率値に基づいて選択した。差分に用いる D クラスの数における領域分割結果の比較を表 5 に示した。クラス数 N が 0 の場合が差分を適用しなかった結果であり、逆伝搬値の差分をとることでより精度の高い位置推定を行うことができた。

本研究は、Simonyan et al.[16] の手法を改良を行ったが、図 6 に逆伝搬値の視覚化を比較した結果、表 6 に逆伝搬値を用いた領域分割結果の精度の比較結果を示した。本研究は Simonyan et al. の手法と比較して、クラスごとに詳細な物体の位置推定を行うことができ、領域分割精度が向上していることがわかる。

Pascal2012 の Validation set における他の弱教師の領域分割手法との比較を表 1、2 に示した。他の既存の弱教師あり領域分割手法の精度を Validation、Test セットともに上回り高い精度を達成していることがわかる。また、70 倍の量の追加画像を用いて学習した、[14] を上回る精度となり、特に Test set においては、物体についての大きさの教師情報を追加で付与した手法 [12] と同程度の精度を達成した。逆伝搬値の差分による物体の位置推定が弱教師あり領域分割において有効であることが分かる。

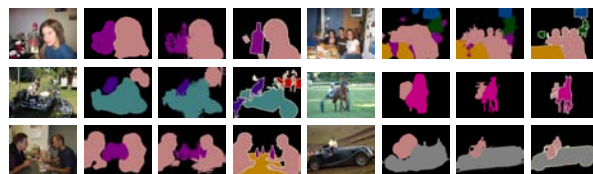


図 7 それぞれ、入力画像 (左)、領域分割結果 (中央)、正解領域 (右) である。

6. 結論

本研究は Simonyan et al. [16] の手法を改良し、逆伝搬値を用いることで、順伝搬値の位置推定結果を改善可能であることを示した。また、逆伝搬値について差分をとることで、順伝搬値を用いずに高精度な位置推定が可能であることを示した。特に Pascal Voc 2012 データセットにおける弱教師あり領域分割タスクにおいて State-of-the-art を達成した。謝辞：本研究は JSPS 科研費 15H05915(新学術領域研究「多元質感知」) の助成を受けたものです。

参考文献

- [1] Andrews, S., Tsochantaridis, I. and Hofmann, T.: Support Vector Machines for Multiple-Instance Learning, *Advances in Neural Information Processing Systems* (2002).
- [2] Carreira, J., Caseiro, R., Batista, J. and Sminchisescu, C.: Semantic Segmentation with Second-Order Pooling, *ECCV* (2012).
- [3] Chen, L., Papandreou, G., Kokkinos, I., Murphy, K. and L., Y. A.: Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs, *ICLR* (2015).
- [4] Everingham, M., Eslami, S. M. A., Van Gool, L., Williams, C. K. I., Winn, J. and Zisserman, A.: The Pascal Visual Object Classes Challenge: A Retrospective, *International Journal of Computer Vision*, Vol. 111, No. 1, pp. 98–136 (2015).
- [5] Fulkerson, B., Vedaldi, A. and Soatto, S.: Class segmentation and object localization with superpixel neighborhoods, *ICCV* (2009).
- [6] Hariharan, B., Arbelaez, P., Bourdev, L., Maji, S. and J., M.: Semantic contours from inverse detectors, *ICCV* (2011).
- [7] Hariharan, B., Arbeláez, P., Girshick, R. and Malik, J.: Simultaneous Detection and Segmentation, *ECCV* (2014).
- [8] Krahenbuhl, P. and Koltun, V.: Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials, *Advances in Neural Information Processing Systems* (2011).
- [9] Long, J., Shelhamer, E. and Darrell, T.: Fully Convolutional Networks for Semantic Segmentation, *CVPR* (2015).
- [10] Mostajabi, M., Yadollahpour, P. and Shakhnarovich, G.: Feedforward semantic segmentation with zoom-out features, *CVPR* (2015).
- [11] Papandreou, G., Chen, L.-C., Murphy, K. and Yuille, A. L.: Weakly-and semi-supervised learning of a dcnn for semantic image segmentation, *ICCV* (2015).
- [12] Pathak, D., Krahenbuhl, P. and Darrell, T.: Constrained convolutional neural networks for weakly supervised segmentation, *ICCV* (2015).
- [13] Pathak, D., Shelhamer, E., Long, J. and Darrell, T.: Fully convolutional multi-class multiple instance learning, *ICLR* (2015).
- [14] Pedro, P. and Ronan, C.: From Image-level to Pixel-level Labeling with Convolutional Networks, *CVPR* (2015).
- [15] Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R. and LeCun, Y.: Overfeat: Integrated Recognition, Localization and Detection using Convolutional Networks, *ICLR* (2014).
- [16] Simonyan, K., Vedaldi, A. and Zisserman, A.: Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps, *ICLR* (2014).
- [17] Springenberg, J. T., Dosovitskiy, A., Brox, T. and Riedmiller, M.: Striving for Simplicity: The All Convolutional Net, *ICLR* (2015).