自己教師あり学習による変化領域の推論を活用した 弱教師あり領域分割

下田 和^{1,a)} 柳井 啓司^{1,b)}

概要

画像における物体のカテゴリ情報のみから領域分割を学 習する手法を弱教師あり領域分割と呼ぶ。弱教師あり学習 による領域分割が可能となれば,領域分割における学習 データを収集するための大幅なコストの削減が可能であ る。本研究においては,自己教師あり学習による変化領域 の推論を活用することにより,弱教師あり領域分割の精度 を向上させた。また,提案手法を領域分割モデルの学習に 組み込むことで,更なる精度向上を達成した。

1. はじめに

領域分割は画像認識において重要なタスクの一つである が,領域分割の学習に必要な教師情報はコストが膨大で あり,この問題は領域分割を応用するうえで大きな障害に なっている。領域分割における弱教師あり学習では,画像 のクラスラベルのみを用いて領域分割のためのモデルを学 習することで,安価なコストで領域分割を行うことができ る。代用が可能なほど弱教師あり領域分割の精度が向上す れば,大幅なコスト削減が期待できる。

領域分割結果の精度を改善する手法としては Conditional random field(CRF) [5] が広く知られている。CRF は色や 位置情報などを特徴量として活用し確率分布が領域の輪郭 に沿うように最適化する手法であるが、領域分割の精度向 上を保証する手法ではないため精度低下を招いてしまう場 合がある。Pixel-level affinity(PSA)[1] は、現在の弱教師あ り領域分割において高精度を達成している手法の一つであ るが、この手法においても CRF が取り入れられている。本 研究では PSA の領域分割結果に CRF を適用した際の精 度低下を防ぐために,自己教師あり学習による変化領域の 推論を活用した領域分割結果の統合手法を提案する。提案 手法においては PSA における CRF 適用前と CRF 適用後 の領域分割結果を入力として、これらの二つの領域分割結 果について変化領域を推論する。この変化領域の推論結果 から領域の信頼度を算出し、この信頼度を用いて領域分割 結果を統合する。変化領域の推論の学習と、変化領域の推 論結果を用いて領域分割結果を統合する方法の概要を図1 に示した。また、本研究においては、PSA の結果の改善の みでなく領域分割モデルの学習に提案手法を組み込む応用 方法を示す。この試みにおいては、学習中の教師情報にお けるノイズを推定し、正しいラベルに上書きをするという

難しい課題に挑戦したが, 提案手法は有効に働き, Pascal VOC 2012 データセットの弱教師あり領域分割手法におけ る最高精度を達成している。

2. 変化領域の推論を活用した領域分割結果の 統合手法

本論文においては、二つの候補領域の変化領域を推論す る新しいタスクを提案する。また、変化領域の推論結果を 用いて領域の信頼度を算出し、領域分割結果を統合する手 法を提案する。変化領域の推論においては、画像から得ら れる特徴量と二つの候補領域のうちの一方を入力として 与える。学習時にはどちらの候補領域を入力とした場合に おいても、同様に各ピクセルでラベルが異なっている領域 を出力するように学習するが、実際には入力の違いにより 結果が異なる。また,推論が容易になる変化領域の傾向は CRF 適用前と CRF 適用後で異なっている。図2にこれら の傾向の違いについて示した。学習サンプルに類似した変 化領域が多く存在すれば,変化領域の推論は容易になると 推測できる。このとき、CRF により変化した領域は基本的 に精度が向上していると仮定すれば, CRF 適用前の入力に おいて変化の推論が容易な領域は CRF により精度が向上 する可能性が高い。CRF 適用後においては, CRF により 大きく領域分割結果が変化すると、領域分割結果が入力の 特徴量とかけ離れた結果になり,変化領域の推論が容易に なる場合がある。このとき, CRF 適用前の入力が一定の精 度を保っており特徴量が信頼できると仮定すれば,変化領 域の推論が容易な領域は CRF により精度が低下している 可能性が高いと判断できる。CRF 適用後の推論において は、CRF 適用前の変化領域の推論と同様に、学習サンプル の頻度が多いと推論が容易になる傾向があるが、特徴量の 分布と領域分割マスクの違いの影響のほうが大きかった。 そこで、本提案手法においては、推論が容易な変化領域は入



図1 変化領域の推論の概略図。

¹ 電気通信大学 大学院情報理工学研究科 情報学専攻

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

 $^{^{\}rm b)}$ yanai@cs.uec.ac.jp

力の領域ラベルの信頼性が低いと考え,変化領域の推論結 果を入力マスクのラベルの信頼度マップとして扱いこれに 基づいて領域の統合を行う。

2.1 変化領域の推論による信頼度の算出とそれに基づいた領域分割結果の統合

本論文においては、二つの候補領域について割り当てら れたラベルが異なる領域のことを変化領域と定義し、この 変化領域を推論するタスクを提案する。まず、変化領域の 推論について定式化を行う。候補領域を仮に($m^{\text{before}}, m^{\text{after}}$) とし、これらの候補領域の変化を $M^{\text{before,after}}$ とする。ただ し、 $u \in \{1, 2, ..., n\}$ はピクセルの位置を表すものとする。 このとき、変化領域は以下の式で表すことができる。

$$M_u^{\text{before, after}} = \begin{cases} 1 & \text{if } (m_u^{\text{before}} = m_u^{\text{after}}) \\ 0 & \text{if } (m_u^{\text{before}} \neq m_u^{\text{after}}) \end{cases}$$
(1)

変化領域を推論する際は、二つの候補領域のペアから 片方の領域情報を与え、変化領域を推論する。また、変化 領域を推定する際には、 入力画像 x を Segmentation モ デルの Backbone network に与えた際に得られる高次の 特徴量 $e^h(x)$ と低次の特徴量 $e^l(x)$ を用いる。高次の特 徴量からは意味的な特徴量を低次の特徴量からは物体の 輪郭についての情報が得られることが期待できる。Segmentation モデルの Backbone network には, PSA[1] と同 様に Resnet38 を用いた。本論文においては、この推論を 行うネットワークモデルを Difference Detection network (DD-net)とした。推定された変化領域の分布を以下の式 で表す。 $d = \text{DDnet}(e^h(x; \theta_e), e^l(x; \theta_e), \hat{m}; \theta_d), d \in \mathbb{R}^{H \times W},$ た だし, *î* は領域分割結果 *m* について各ピクセルにおける ラベル情報を One-hot vector としてバイナリ化したもので ある。 θ_d は推論モデルのパラメータであり, 領域分割モデ ルの学習とは独立している。また、このパラメータは二つ の領域分割結果の組の変化の推論において共有されている。 これはパラメータを共有しなかった場合、領域の推論結果 にバイアスがかかるため推論結果を信頼度として平等に扱 う場合に問題となるためである。この変化の推論を二つの 候補領域 m^{before}, m^{after} についてそれぞれ適用し d^{before}, d^{after} を 得る。

この推論結果を領域分割結果の信頼度として考え,二つの候補領域の信頼度の差wを以下の式で定義する。

$$w_{u}^{\text{before, after}} = d_{u}^{\text{after}} - d_{u}^{\text{before}} + \hat{b}, \qquad (2)$$

ただし, *b*はそれぞれの領域分割マスクに対してのバイア スである。実験においてはグリッドサーチによる探索結果 から *b* = 0.4 とした。

信頼度の差 w に基づき, 領域の統合結果 m^{refine} を以下の 式から得る。

$$m_u^{\text{refine}} = \begin{cases} m_u^{\text{before}} & \text{if } (w_u^{\text{before, after}} > 0) \\ m_u^{\text{after}} & \text{if } (w_u^{\text{before, after}} < 0) \end{cases}$$
(3)

ただし、uはピクセルを表している。本研究では簡 単化のためにこの領域補正処理を Self-Supervised Difference Detection module(SSDD module) として、今後 $m^{\text{refine}} = \text{SSDD}(e(x; \theta_e), m^{\text{before}}, m^{\text{after}}; \theta_d)$ と記述する。また、 図 3 に本手法における領域の変化の推論のためのネット ワークを示した。



図 2 CRF 適用前と CRF 適用後における変化領域の推論の難易度 の変化の傾向についての説明図。



図 3 Difference Detection Network (DD-Net) の図。C はカテゴ リ数を示している。

変化領域の推論を活用した領域分割結果の 統合手法の弱教師あり領域分割への応用

本セクションでは,変化領域の推論を活用した領域分割 結果の統合手法を弱教師あり領域分割へ適用する方法につ いて述べる。PSA[1]の領域分割結果にCRFを適用した際 の精度低下を防ぐために,変化領域の推論を活用した領域 分割結果の統合を適用する。これを変化領域の推定を活用 した領域分割結果の静的補正として,セクション3.1 で述 べる。さらに,本手法を領域分割学習ネットワークに組み 込み変化領域の推論結果の学習と領域分割結果の学習を同 時に行う手法を変化領域の推定を活用した領域分割結果の 動的学習として,セクション3.2 において示す。

3.1 変化領域の推定を活用した領域分割結果の静的補正

Pixel-level affinity(PSA)[1] は Pixel level の類似度を学 習し、クラス分類の可視化結果を改善する手法である。本 手法においては PSA の CRF 適用前と適用後の結果につ いて変化領域を推論し、領域の統合を行う。まず、入力画 像 x について, PSA を適用した結果を $p^{\text{psa}}=PSA(x;\theta_{\mathcal{P}})$ と する。これについて CRF を適用した結果を perfl とする。 またこれらの確率分布に対して最大値のインデックスを取 ることにより領域分割マスクに変換したものを m^{psa}, m^{crf0} とする。これらの変化領域 M^{psa,crf0} についての推論結果を 以下の式から得る。 $d^{\text{psa}}=f(e^{h}(x;\theta_{e0}),e^{l}(x;\theta_{e0}),\hat{m}^{\text{psa}};\theta_{d0}),$ $d^{\text{crf0}} = f(e^h(x; \theta_{e0}), e^l(x; \theta_{e0}), \hat{m}^{\text{crf0}}; \theta_{d0}).$ また、 領 域補正結果 m^{ssdd0}を以下の式から求める。 $m^{\text{ssdd0}} = \text{SSDD}(e(x; \theta_{e0}), m^{\text{psa}}, m^{\text{crf0}}; \theta_{d0}).$ 図4に静的な領域 補正の概略図を示した。

Training 変化領域の推論の学習においては以下のロス を最適化する。



図 4 静的な領域候補の補正についての概略図。

$$\mathcal{L}_{\text{change}} = \frac{1}{|S|} \sum_{u \in S} (J(M^{\text{psa,crf0}}, d^{\text{psa}}, u) + J(M^{\text{psa,crf0}}, d^{\text{crf0}}, u)), \quad (4)$$

ただし, J(M, d, u) はピクセル uにおける $\sigma(d)$ と M につい ての二値交差エントロピー誤差であるとする。 $\sigma()$ はシグ モイド関数である。また、本研究においては Segmentation モデルの Backbone network からよい特徴量を得るために 以下の式で Segmentation モデルを領域分割マスク m^{psa} を 用いて学習し, $\theta_{e0} \in \theta_{s0}$ を最適化する。ただし, theta_{s0} と theta_{d0} の学習は独立している。

$$\mathcal{L}_{\text{seg}} = -\frac{1}{\sum_{k \in \mathcal{C}} |S_k^{\text{psa}}|} \sum_{k \in \mathcal{C}} \sum_{u \in S_k^{\text{psa}}} \log(h_u^k(x;\theta_{s0})), \qquad (5)$$

ただし, *h^ku* はピクセル *u* におけるクラス *k* についての確 率である。本手法においては最終的に以下のロスを最適化 する。

$$\mathcal{L}_{\text{static}} = \mathcal{L}_{\text{seg}} + \mathcal{L}_{\text{change}} \tag{6}$$

3.2 変化領域の推定を活用した領域分割ネットワークの 動的学習

近年の弱教師あり領域分割手法においては弱教師あり条 件下で生成した領域分割マスクで領域分割モデルを学習す ると精度が更に向上されることが知られており,この方法 は近年の弱教師あり領域分割において標準的に使われてい る。しかし,この際には生成した領域分割マスクの精度が 直接領域分割モデルの精度の限界に繋がる。もし,学習中 に教師情報をよりよいものに置き換えることができればさ らなる精度向上が期待できる。そこで,本手法においては セクション 3.1 において生成した領域分割マスクを用いて 領域分割モデルを学習させる際,学習中に更に提案手法を 適用し教師情報を上書きする。 特に,変化が推論可能な領 域のみ置き換えが可能であるという条件を設定することで 堅牢かつ精度の高い教師情報の上書きを実現した。

以下に提案手法の手順を示す。本手法においてはまず 入力画像 x から領域分割モデルの出力 $p^{\text{seg}}=g(x;\theta_{s1})$ とそ の出力についての CRF の適用結果 p^{crl} を得る。ただし, g は領域分割モデルの関数である。これらから得られるマ スクを ($m^{\text{seg}}, m^{\text{crl}}$) であるとする.そして,本手法において は補正された領域分割マスク m^{sedd1} を以下の式から得る。 $m^{\text{sedd1}}=\text{SSDD}(e(x;\theta_{e1}), m^{\text{seg}}, m^{\text{crl}};\theta_{d1})$.次に,得られた領域分 割マスク m^{sedd1} と学習前に弱教師あり条件化で事前に生成 したマスク m^{sedd1} からさらに補正されたマスク m^{sedd2} を以 下の式から得る。 $m^{\text{sedd2}}=\text{SSDD}(e(x;\theta_{e1}), m^{\text{sedd1}}, m^{\text{sedd2}};\theta_{d2})$.本 手法においては m^{sedd2} を学習中の各繰り返しにおいて生成 し、これを領域分割の学習に使う。図 5-(a) に動的な領域 補正の概略図を示した。

Training 生成された領域分割マスク *m*^{ssdd2} を用いて領域 分割モデルを以下の式から学習する。



図 5 (a) 動的な領域補正の概略図。(b) Second SSDD module の 学習図。(c) 領域分割モデルの学習図。)

$$\mathcal{L}_{\text{seg-main}} = -\frac{1}{\sum_{k \in \mathcal{C}} |S_k^{\text{ssdd2}}|} \sum_{k \in \mathcal{C}} \sum_{u \in S_k^{\text{ssdd2}}} \log(h_u^k(x; \theta_{s1})).$$
(7)

(m^{seg} と m^{crf1})の変化の推論モデルは以下の式で学習する。

$$\mathcal{L}_{\text{dd-crf}} = \frac{1}{|S|} \sum_{u \in S} (J(M^{\text{seg,crf1}}, d^{\text{seg}}, u) + J(M^{\text{seg,crf1}}, d^{\text{crf1}}, u)),$$
(8)

次に (m^{stdll} , m^{stdl0})の変化の推論モデルの学習について 述べる。(m^{stdll} , m^{stdl0})の変化の推論は m^{stdl0} が学習の間常 に一定であるので、変化の推論が容易である。この場合変 化の推論の難易度の差を作るのが困難になり、提案手法の 適用が難しくなる。そこで、学習を安定化させるために本 手法では新しい領域分割ブランチを作成する。このブラン チから得られる確率分布を p^{sub} , 領域分割マスクを m^{sub} と する。そして、変化領域の推論の学習においては、以下のペ アの領域候補 (m^{stdl} , m^{sub})と (m^{sub} , m^{stdl1})の変化の推論時におけ る変化領域の推論を難しくし学習が発散するのを防ぐのに 貢献した。この変化の推論は以下のロスを最適化し学習さ せる。

$$\mathcal{L}_{\rm dd-seed} = \frac{1}{|S|} \sum_{u \in S} (J(M^{\rm ssdd0, sub}, d^{\rm ssdd0}, u) + J(M^{\rm ssdd1, sub}, d^{\rm ssdd1}, u)), \ (9)$$

また, パラメータ $\theta_{s1'}$ は以下の式で学習する。

$$\mathcal{L}_{\text{seg-sub}} = \alpha \mathcal{L}^{\text{ssdd0}} + (1 - \alpha) \mathcal{L}^{\text{ssdd2}}, \quad (10)$$

$$\begin{split} \mathcal{L}^{\text{ssdd0}} &= -\frac{1}{\sum_{k \in \mathcal{C}} |S^{\text{ssdd0}}_k|} \sum_{k \in \mathcal{C}} \sum_{u \in S^{\text{ssdd0}}_k} \log(h^k_u(x; \theta_{s1'})) \\ \mathcal{L}^{\text{ssdd2}} &= -\frac{1}{\sum_{k \in \mathcal{C}} |S^{\text{ssdd2}}_k|} \sum_{k \in \mathcal{C}} \sum_{u \in S^{\text{ssdd2}}_k} \log(h^k_u(x; \theta_{s1'})). \end{split}$$

 α は学習に用いるマスクについてのハイパーパラメータである。実験においては $\alpha = 0.5$ とした。最終的なロスは以下の式に従う。

$$\mathcal{L}_{dynamic} = \mathcal{L}_{dd-crf} + \mathcal{L}_{dd-seed} + \mathcal{L}_{seg-main} + \mathcal{L}_{seg-sub}$$
(11)

図 5-(b)(c) に動的な領域補正の学習方法を示した。

表 1 変化領域の推論モデルを使った静的な領域補正の効果の検証 († 再現実装における結果)

	Raw seed	Trained model
	(train set)	(validation set)
PSA	52.5	58.4†
PSA+CRF	48.0	59.0†
提案手法(静的な領域補正)	53.4	61.4

表 2 最新の弱教師あり領域分割手法との比較。

Method	Val	Test
FCN-MIL [8]ICLR2015	25.7	24.9
CCNN [7]ICCV2015	35.3	35.6
EM-Adapt [6]ICCV2015	38.2	39.6
DCSM [11]ECCV2016	44.1	45.1
BFBP [10]ECCV2016	46.6	48.0
SEC [4]eccv2016	50.7	51.7
CBTS [9]CVPR2017	52.8	53.7
TPL [3]1CCV2017	53.1	53.8
MEFF [2]CVPR2018	-	55.6
PSA [1]CVPR2018	61.7	63.7
SSDD(提案手法, 動的な領域補正)	64.9	65.5



図 6 (a) 入力画像, (b)PSA の領域分割結果, (c) (b) についての変 化領域の推論結果, (d) (b) についての CRF の適用結果, (e) (d) についての変化領域の推論結果, (f) 手印手法による領域 補正結果, (g) 正解マスク. 最下段は提案手法を適用しても精 度が改善しない場合の例である。

4. 実験

領域分割のベンチマークのデータセットとして PASCAL VOC 2012 (21 class) を用いた。Pascal は 10582 枚の train 画像, 1449 枚の validation 画像, 1456 枚の test 画像から成 る。評価の指標としては Mean IoU を用いた。

表1の左列にPSAを学習画像に適用した結果と静的な 領域補正(セクション3.1)の適用結果を示した。単純に本 手法を適用した場合CRFの結果は元のPSAの領域分割結 果と比較して大きく精度が落ちてしまったものの,本手法 を適用することで精度が向上している。右列には左列の領 域分割マスクで領域分割モデルを学習しこれをValidation set に適用した際の精度を示している。CRFの結果は左列 の場合とは異なりCRF適用前の結果を上回っているが, 提案手法はこれをさらに上回っている。この結果から本手 法がCRFの結果の改善に貢献していることがわかる。図 6に提案手法による領域分割の補正結果を示した。

もうひとつの実験として動的領域補正の精度(セクション 3.2)を近年の弱教師あり領域分割手法と比較した。表2の結果から提案手法が弱教師あり領域分割において最高精度を達成していることがわかる。図4に本手法により得られた領域分割結果の例を示す。



図 7 Pascal VOC 2012 における領域分割結果.

5. おわりに

本手法においては, 弱教師あり領域分割において重要な 役割を担っている CRF の問題点に着目し変化領域の推論 を活用することで新しい領域補正手法を提案した。また, 提案手法を Segmentation モデルの学習に組み込むことが 可能であることを示した。さらに, 弱教師あり条件下にお ける PASCAL VOC 2012 の最高精度を達成し提案手法の 有効性を示した。

謝辞 本研究は JSPS 科研費 17J10261, 15H05915, 17H01745, 19H04929, 17H06100の助成を受けたものです。

参考文献

- Ahn, J. and Kwak, S.: Learning Pixel-level Semantic Affinity with Image-level Supervision for Weakly Supervised Semantic Segmentation, CVPR (2018).
- [2] Ge, W., Yang, S. and Yu, Y.: Multi-evidence filtering and fusion for multi-label classification, object detection and semantic segmentation based on weakly supervised learning, *CVPR* (2018).
- [3] Kim, D., Cho, D., Yoo, D. and Kweon, I.: Two-Phase Learning for Weakly Supervised Object Localization, *ICCV* (2017).
- [4] Kolesnikov, A. and H.Lampert, C.: Seed, Expand and Constrain: Three Principles for Weakly-Supervised Image Segmentation, ECCV (2016).
- [5] Krahenbuhl, P. and Koltun, V.: Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials, Advances in Neural Information Processing Systems (2011).
- [6] 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).
- [7] Pathak, D., Krahenbuhl, P. and Darrell, T.: Constrained convolutional neural networks for weakly supervised segmentation, *ICCV* (2015).
- [8] Pathak, D., Shelhamer, E., Long, J. and Darrell, T.: Fully convolutional multi-class multiple instance learning, *ICLR* (2015).
- [9] Roy, A. and Todorovic, S.: Combining bottom-up, topdown, and smoothness cues for weakly supervised image segmentation, CVPR (2017).
- [10] Saleh, F., Akbarian, M., Salzmann, M., Petersson, L., Gould, S. and M.Alvares, J.: Built-in Foreground/Background Prior for Weakly-Supervised Semantic Segmentation, ECCV (2016).
- [11] Shimoda, W. and Yanai, K.: Distinct Class Saliency Maps for Weakly Supervised Semantic Segmentation, ECCV (2016).