

# 変化領域の推測による弱教師 あり領域分割の精度向上

電気通信大学 総合情報学科

下田 和, 柳井 啓司

# 弱教師あり領域分割

- 弱教師あり学習
  - ラベルのみで学習 (クラス分類と等価)

完全教師あり学習



弱教師あり学習



人

馬

車

# 提案手法

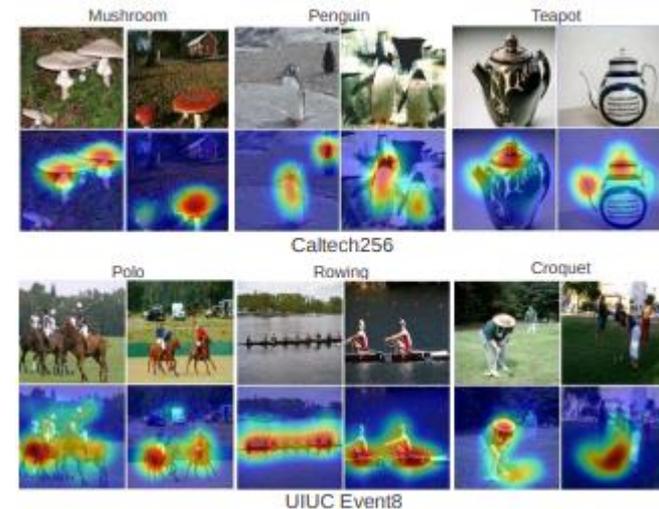
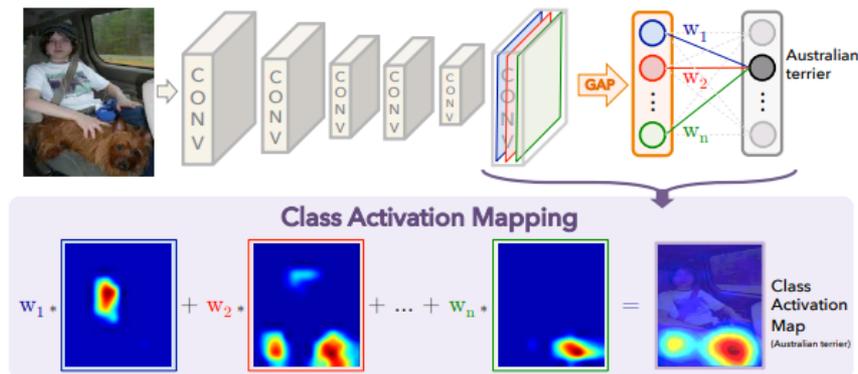
- 自己教師あり学習を活用した領域補正手法を提案
- 変化領域を推論するモデルを学習し、これを活用
- Pascal VOC 2012で弱教師あり領域分割の設定における最高精度を達成

# 既存の弱教師あり領域分割手法

- 分類結果の可視化
  - クラス分類モデルの認識結果を可視化
  - Simonyan et al, Oquab et al
- 弱教師あり領域分割の精度向上方法
  - 低次特徴量の活用
    - Conditional random field(CRF)
    - 色やエッジの情報など
  - 事前にPixel-levelの教師情報を生成し、領域分割モデルを学習
  - 学習中にPixel-levelの教師情報を生成し、領域分割モデルを学習

# Class Activation Map(CAM)

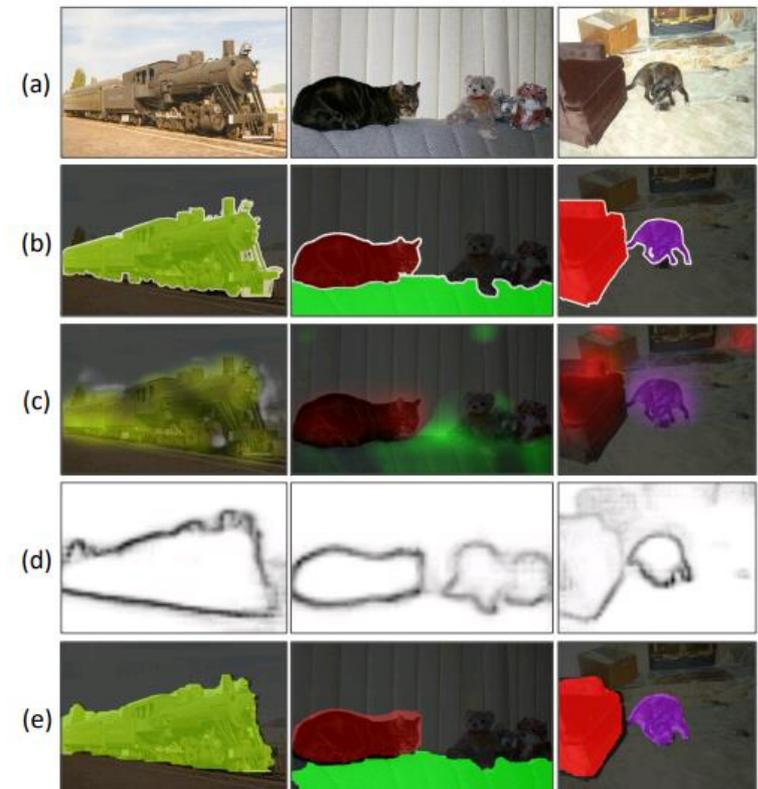
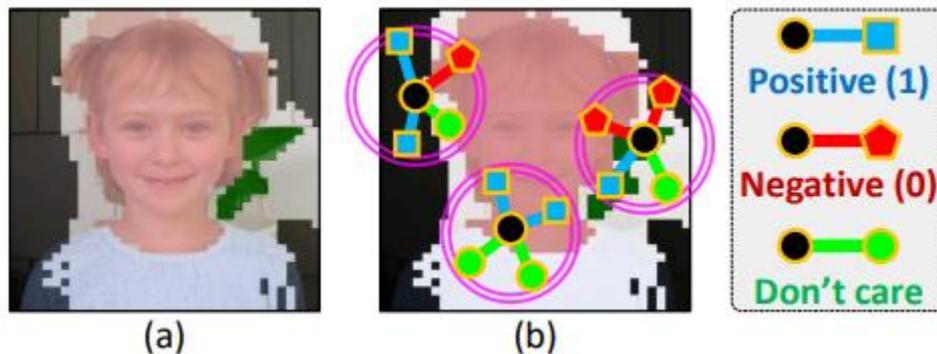
- Class activation map
  - クラス分類結果を可視化する手法
  - シンプルで実装が容易



[Zhou et al, CVPR 2016]

# Pixel-level Semantic Affinity(PSA)

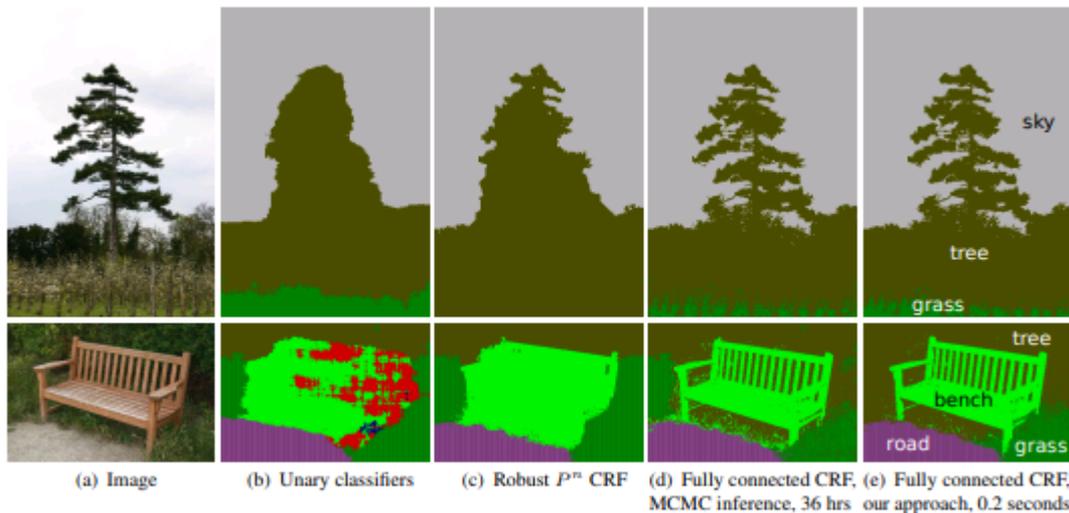
- Pixel-levelで物体の類似度を学習し、ラベルの伝播を行う手法
  - CRFの結果を活用
- CAMの精度を大幅に改善



[Ahn et al. CVPR 2018]

# CRFを利用した弱教師あり領域分割の精度改善手法

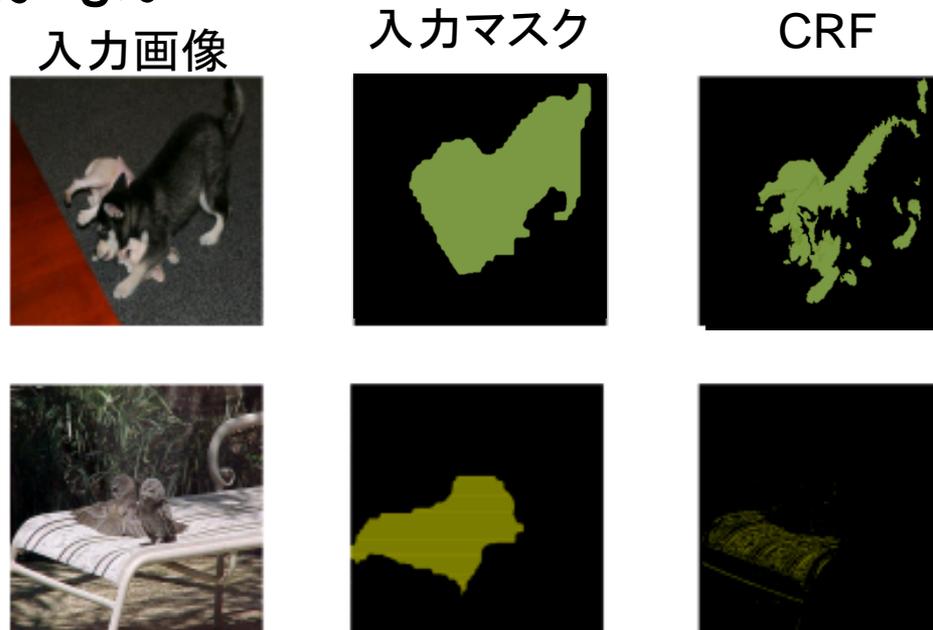
- Fully connected CRF
  - 色特徴量、ピクセルの位置情報を活用
  - 物体の輪郭に沿った領域が得られる



[Pathak et al. NIPS 2011]

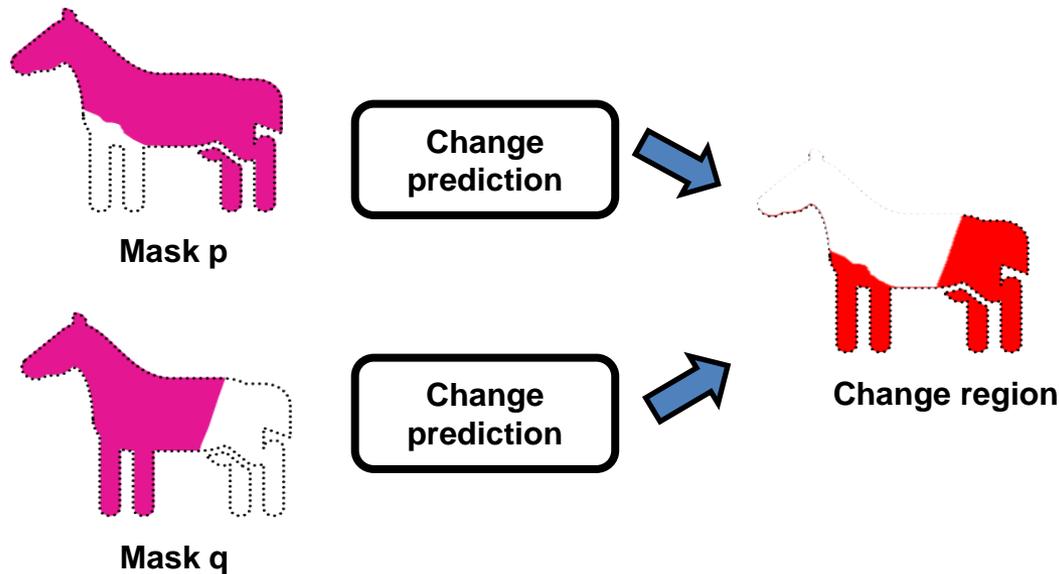
# 本研究の着眼点

- CRFの精度
  - 物体の輪郭に沿った領域が得られる
  - ただし、対象の物体に対応している領域が得られる保証がない



# 変化領域の推論

- 二つの候補領域の変化を推論するタスク
  - 自己教師あり学習 (Self-supervised learning)
    - 教師情報を必要としない擬似タスク (Pretext task)
    - Colorization や Frame ordering



# 変化領域の推論

- 入力
  - 任意の二つの領域分割結果のペア:  $m^p, m^q$
  - 画像  $x$  から得られる特徴量マップ
    - 高次の特徴量:  $e^h$ , 低次の特徴量  $e^l$

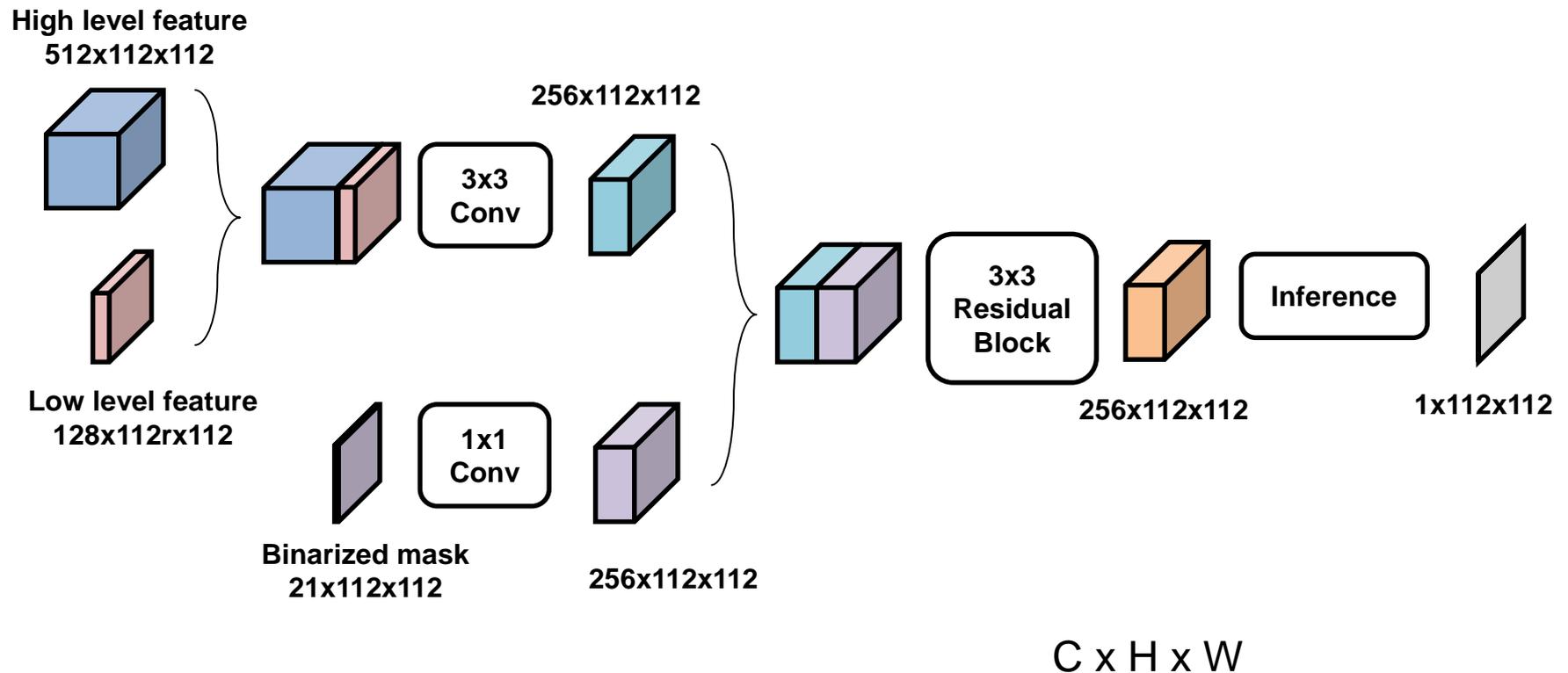
- 出力, 変化領域

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

- 変化領域の推論

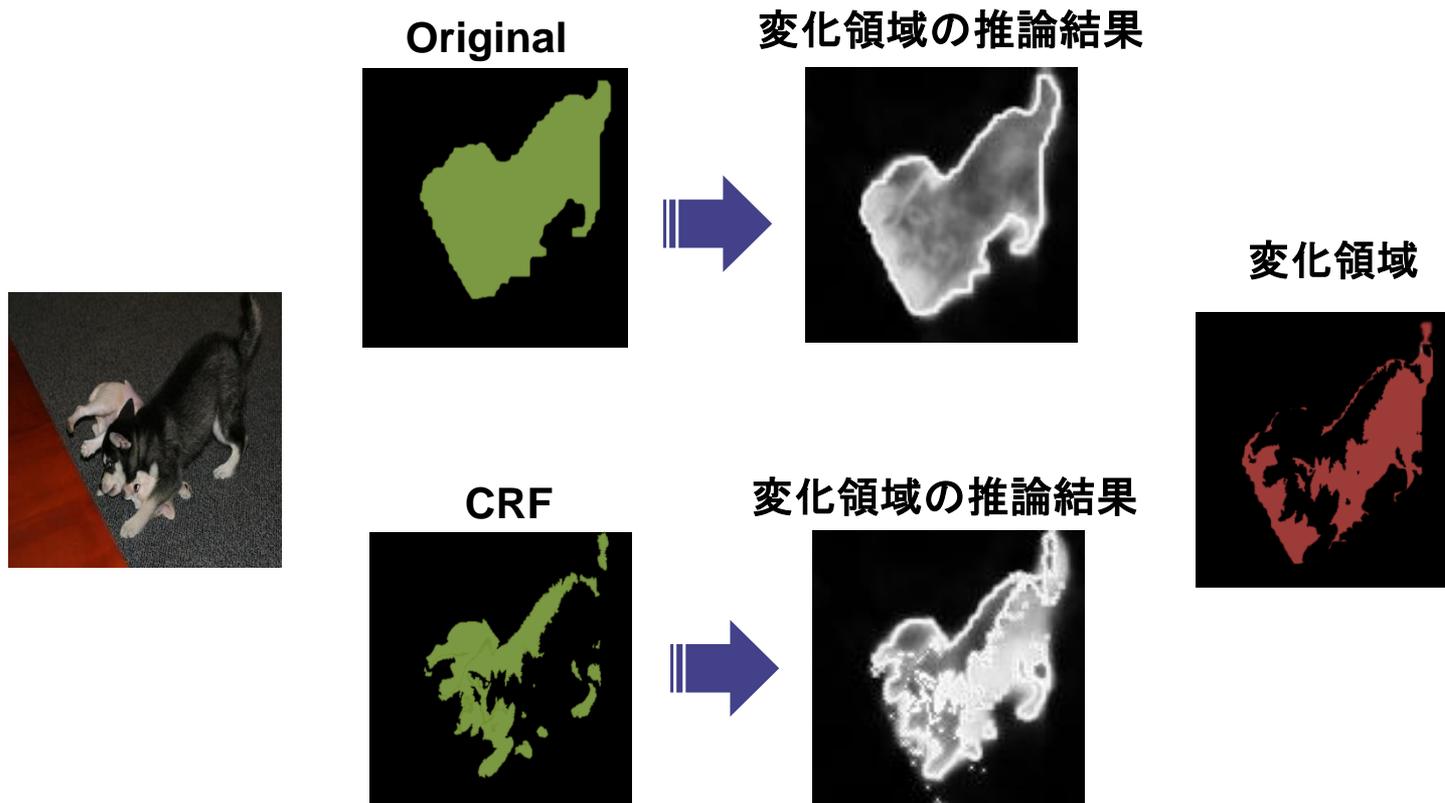
- $d = f(e^h(x; \theta_e), e^l(x; \theta_e), \hat{m}; \theta_d), d \in \mathbb{R}^{H \times W}$
- $\hat{m}$ : 各マスクにおけるクラスのチャンネル数分のバイナリ表現

# 変化領域の推論ネットワーク



# 変化領域の推論

- 推論対象は同じだが、異なる入力により、異なる推論結果が得られる(推論難易度の違いから)



# 変化領域の推論

- 目的
  - 変化領域の推論結果を用いて、二つの領域候補からより多くの正解ラベルを抽出する

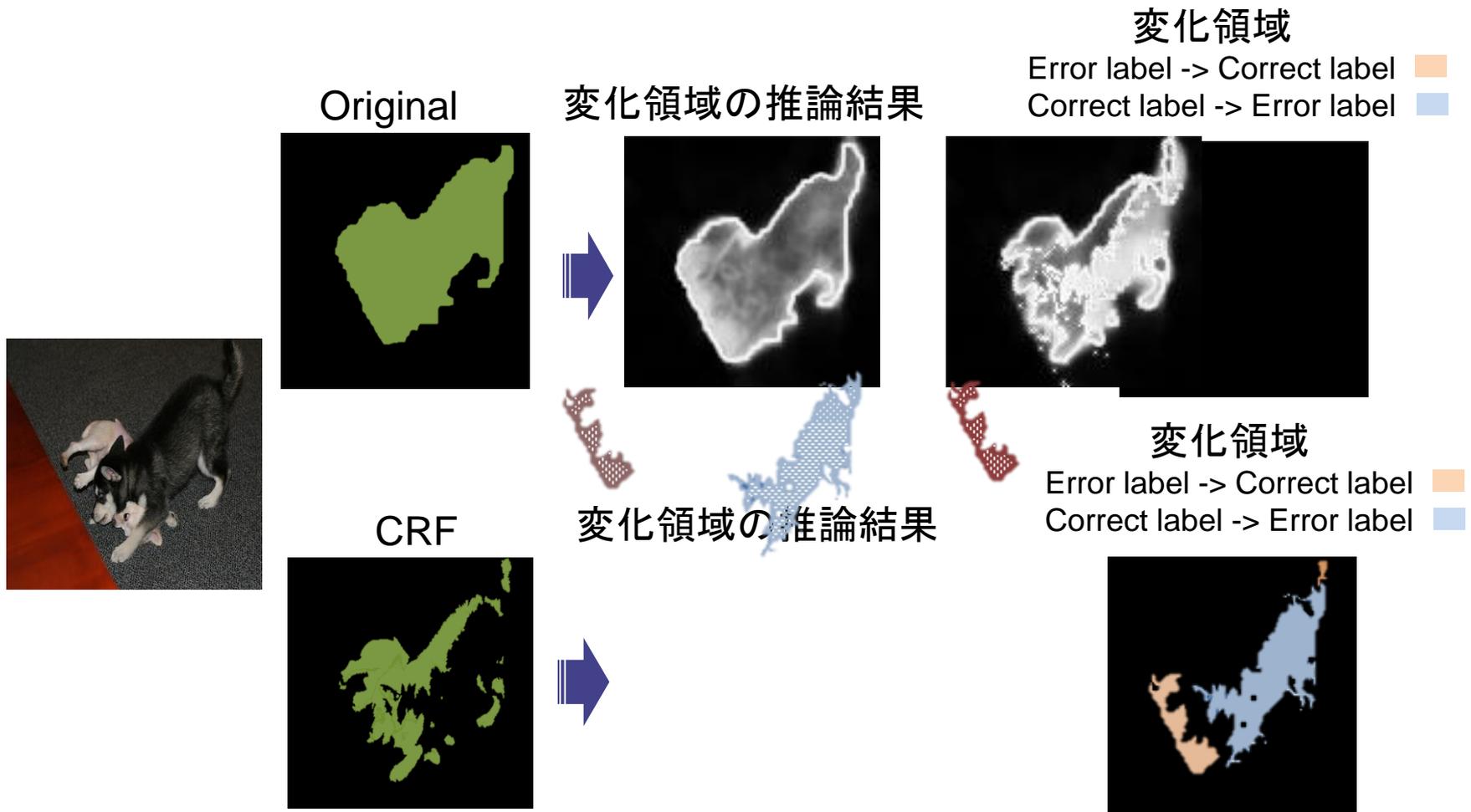
- 仮定

誤りラベルから正解ラベルを求めるほうが

正解ラベルから誤りラベルを求めるより難しい

- 正解ラベルの領域は規則的
- 誤りラベルの領域は不規則的

# 変化領域の推論



# 変化領域の推論による領域補正

- 変化領域における変化の推論結果
  - 大きい値: 誤りラベルの可能性が高い
  - 小さい値: 正解ラベルの可能性が高い
- 変化領域の推論から得られる確信度

$$w_u^{p,q} = (d_u^q + b^{m_u^q}) - (d_u^p + b^{m_u^p}) + \hat{b}$$

- $\hat{b}$ : 二つのmaskに対してのバイアス
- $b^c$ : class c に対してのバイアス
- 領域補正結果  $m^r$ 

$$m_u^r = \begin{cases} m_u^p & \text{if } (w_u^{p,q} > 0) \\ m_u^q & \text{if } (w_u^{p,q} < 0) \end{cases}$$

# 変化領域の推論を活用した領域 補正の応用

- 変化領域の推論を活用した静的な領域補正
  - Pixel-level Semantic Affinity(PSA)で得られた結果に対しCRFを適用
  - PSA+CRFの結果を提案手法で改善
- 変化領域の推論を活用した動的な領域補正
  - 提案手法を領域分割モデルの学習へ組み込み
  - 学習中に提案手法により補正した領域分割結果を使って学習

# 変化領域の推論を活用した動的な領域補正

- Pixel-level Semantic Affinity(PSA)
- PSAの結果にCRFを適用
- これらの領域分割結果をペアとして扱い、変化領域を学習
- 変化領域の推論結果を用いて領域候補を補正

# 変化領域の推論モデルの学習

- 学習ロス: cross entropy

- $m^p$ : PSAマスク  $\mathcal{L}_{change}(d^p, d^{\tilde{p}}, M^{p, \tilde{p}}) = \frac{1}{|S|} \sum_{u \in S} J(M^{p, \tilde{p}}, d^p, u)$
- $m^{\tilde{p}}$ : CRFマスク  $+ J(M^{p, \tilde{p}}, d^{\tilde{p}}, u)$

$$J(M^{p, \tilde{p}}, d^p, u) = M_u^{p, \tilde{p}} \log d_u^p + (1 - M_u^{p, \tilde{p}}) \log(1 - d_u^p), \quad (5)$$

$$J(M^{p, \tilde{p}}, d^{\tilde{p}}, u) = M_u^{p, \tilde{p}} \log d_u^{\tilde{p}} + (1 - M_u^{p, \tilde{p}}) \log(1 - d_u^{\tilde{p}})$$

- 明らかに悪いサンプルを除外

- マスクにおけるクラスラベルの変化量  $\delta_c^{p, \tilde{p}} = N_c^{m^{\tilde{p}}} / N_c^{m^p}$
- 閾値で処理  $\text{if } \delta_c > 0.5 \text{ then } \gamma_c = 1 \text{ else } \gamma_c = 0$
- 処理が反映されたロス  $\text{if } |\sum \gamma_k| = 0 \text{ then } \bar{\mathcal{L}}_{change} = 0 \text{ else } \bar{\mathcal{L}}_{change} = \mathcal{L}_{change}$
- バイアスの決定  $\text{if } \gamma_c^{p, \tilde{p}} = 1 \text{ then } b^c = 1 \text{ else } b^c = 0$

# 変化領域の推論モデルの学習

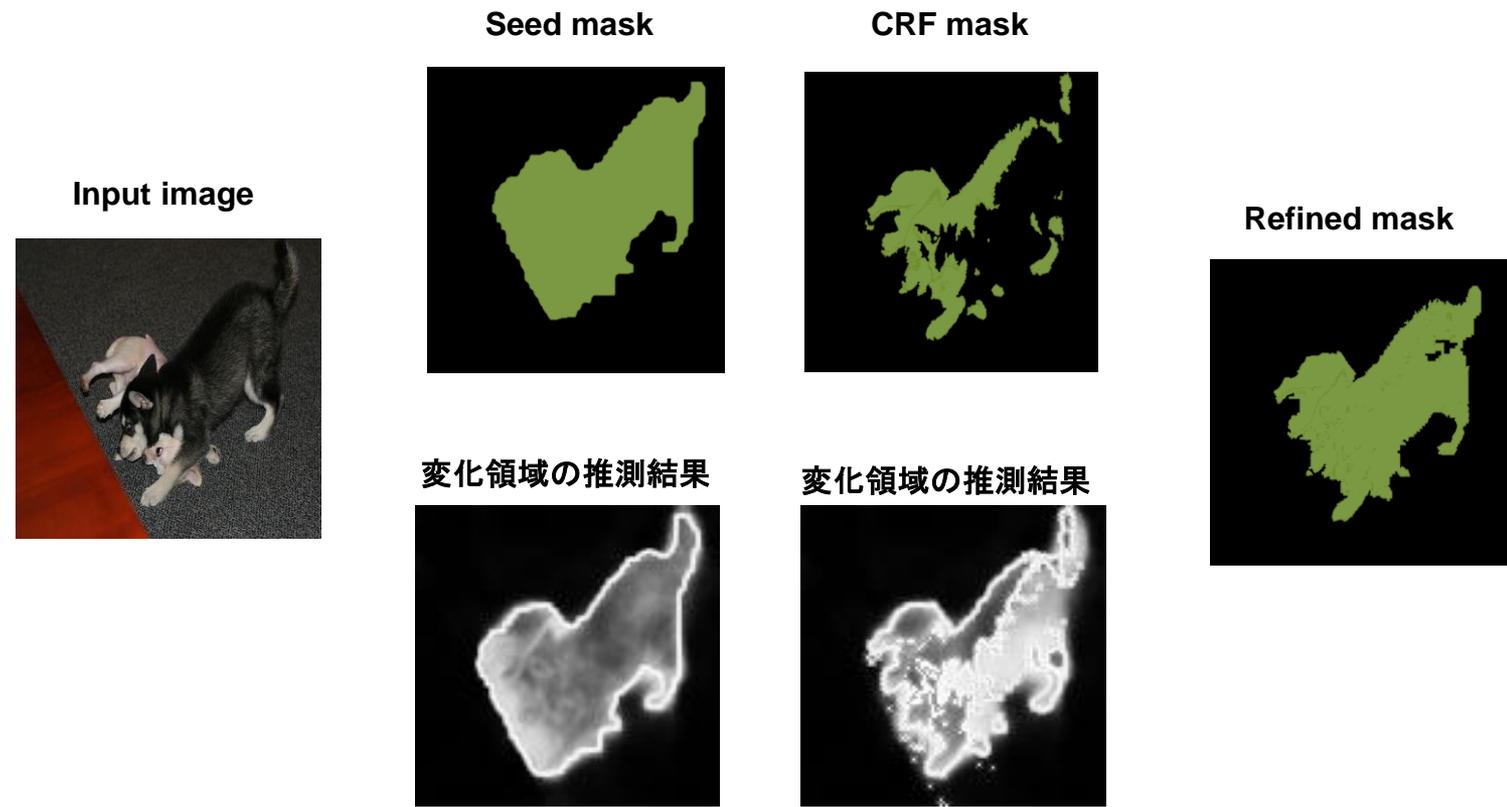
- よい特徴量を得るために領域分割モデルを同時に学習

$$\mathcal{L}_{seg} = -\frac{1}{\sum_{k \in K} |S_c^k|} \sum_{k \in K} \sum_{u \in S_c^k} \log(h_u^k(x; \theta_s))$$

- 最終的なロス

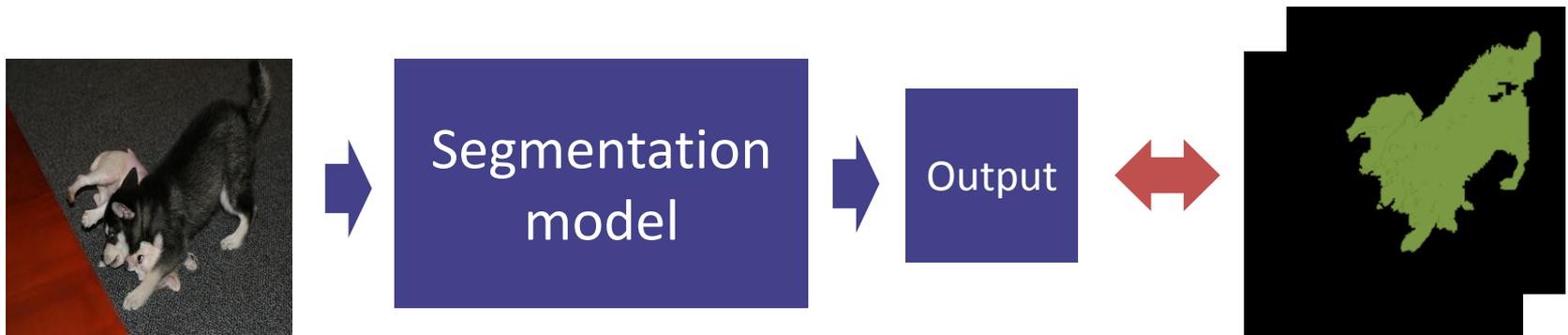
$$\mathcal{L}_{static} = \mathcal{L}_{seg} + \bar{\mathcal{L}}_{change}$$

# 変化領域の推論を活用した静的な領域補正



# 変化領域の推論を活用した動的な領域補正

- 領域分割モデルの学習への提案手法の組み込み
  - 固定された領域Seedで学習すると領域Seedの精度が直接精度の限界に繋がる
  - 領域Seedを学習中に柔軟に置き換えることができれば精度向上が期待できる

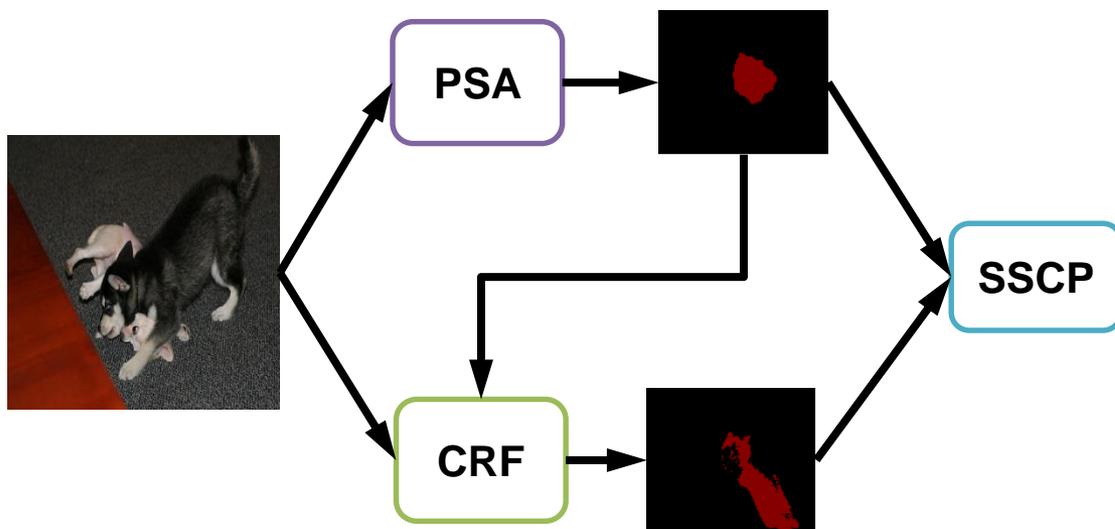


# 変化領域の推論を活用した動的な領域補正

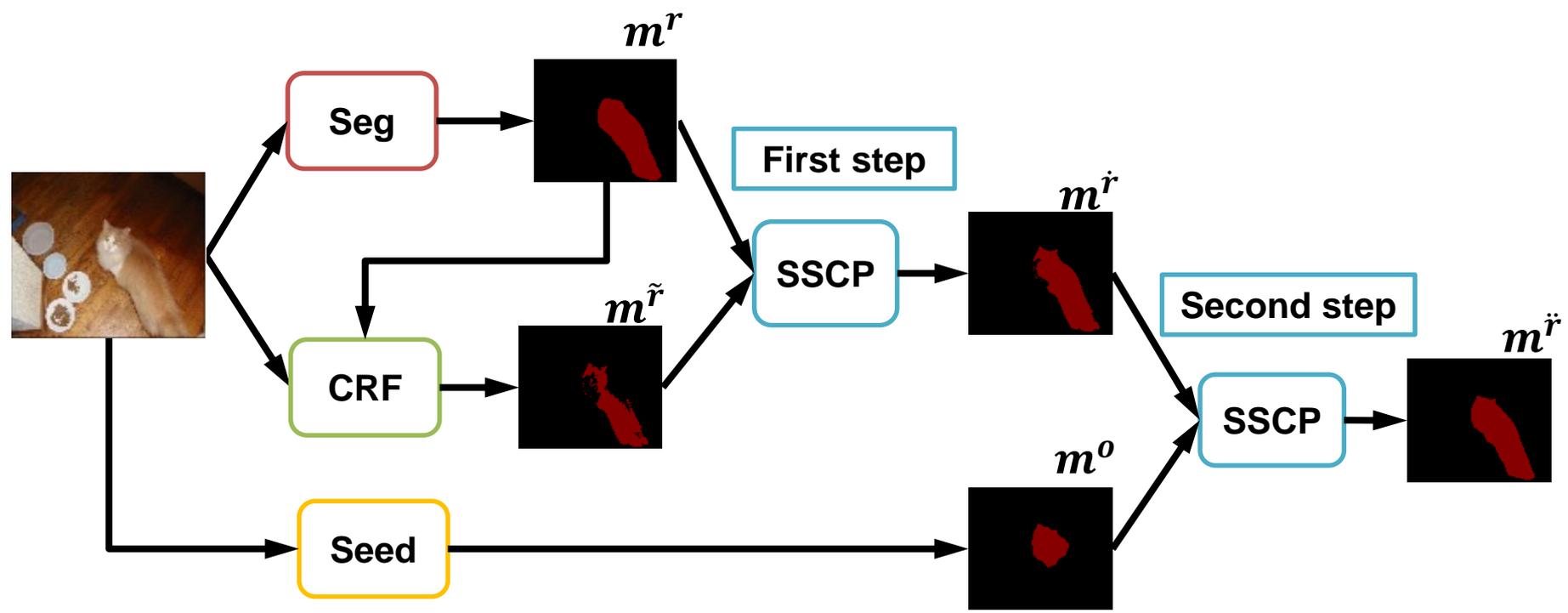
- 既存研究
  - 予めラベルの比率を決定
    - 前景40%背景60%程度 CCNN (ICCV 2015)
  - CRFの結果と類似するように距離学習を行う
    - SEC (ECCV 2016)
  - 上書きできる領域を予め決め打ち
    - DSRG (CVPR 2018)
- 提案手法
  - Seedから変化を推論できる範囲内で教師情報を上書き

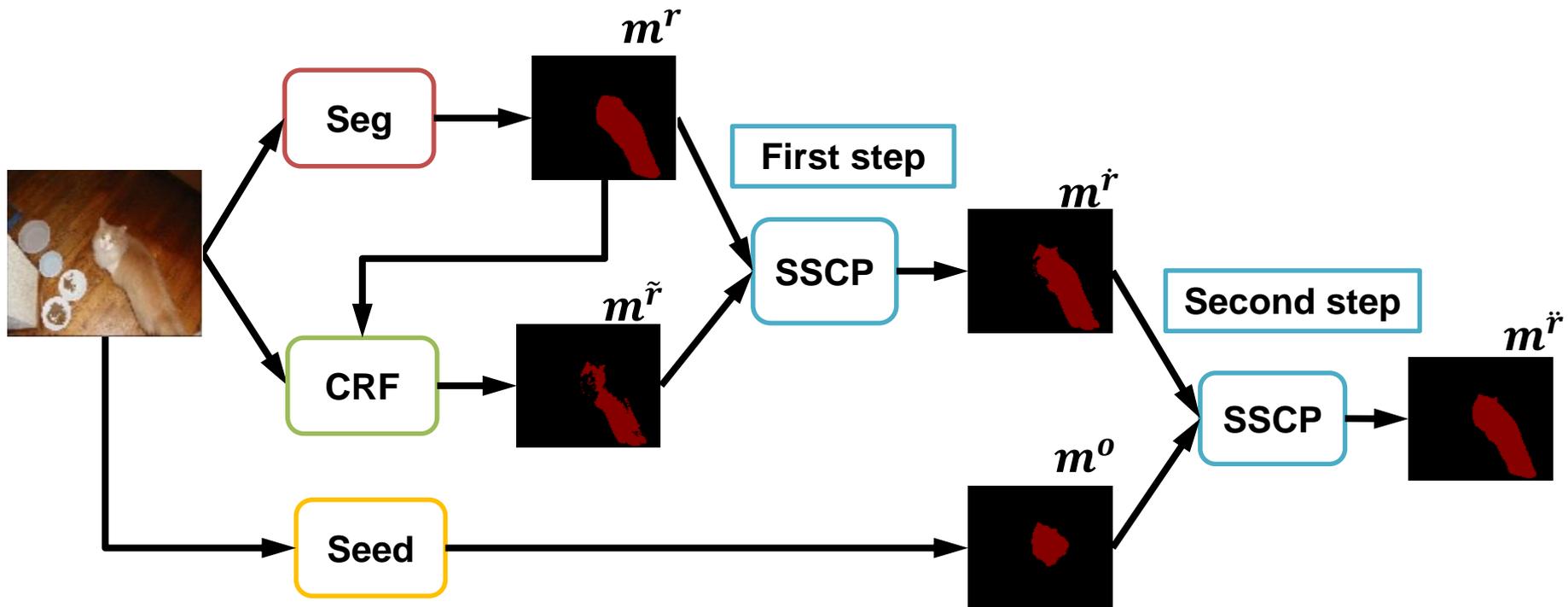
# 変化領域の推論を活用した動的な領域補正の手順

- ステップ1
  - 領域分割モデルから領域分割結果を得る
  - 領域分割結果にCRFを適用する
  - 領域分割結果とCRFについて変化領域を推論し、領域の補正を行う
- ステップ2
  - ステップ1で得られた領域の補正結果と領域Seedの変化領域を推論し、領域の補正を行う
- 二度めの補正結果を使って領域分割モデルを学習

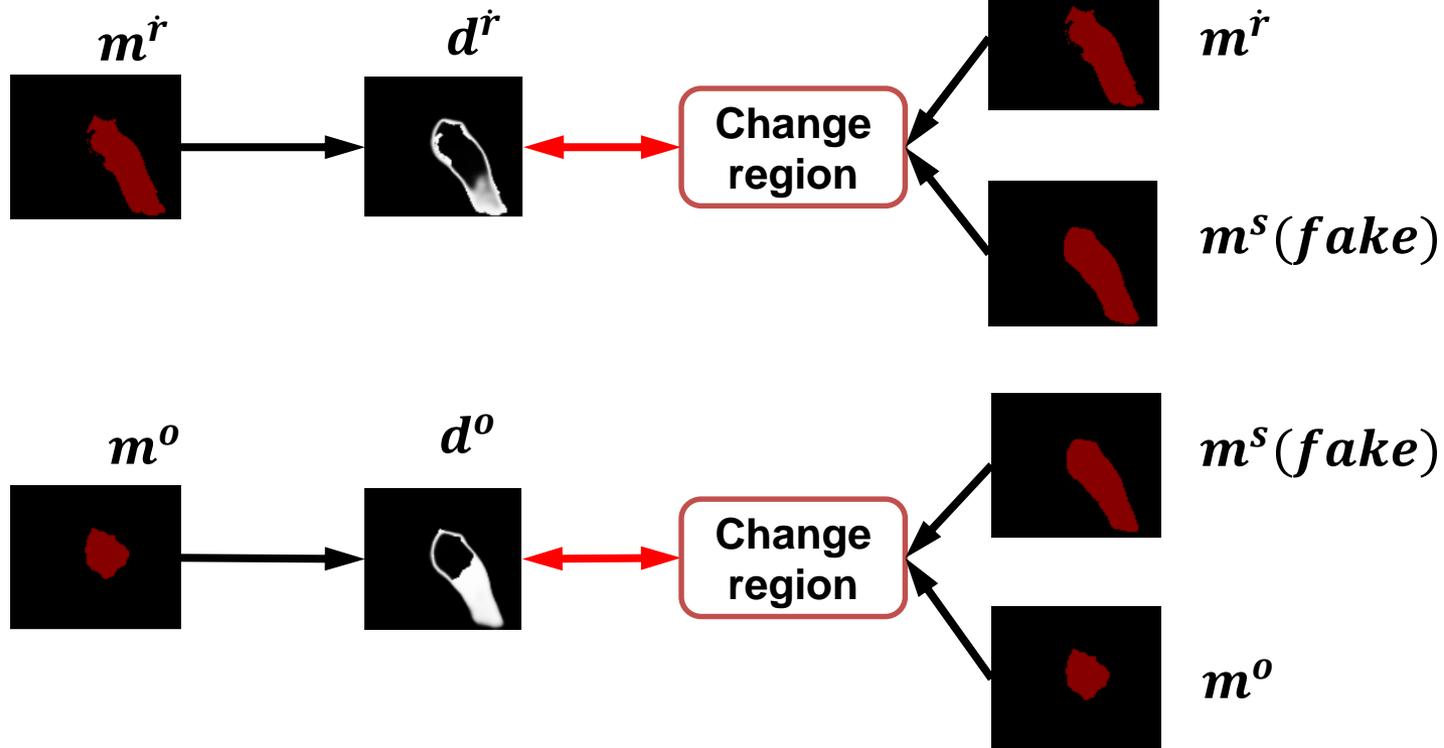


# 変化領域の推論を活用した動的な領域補正の概要図

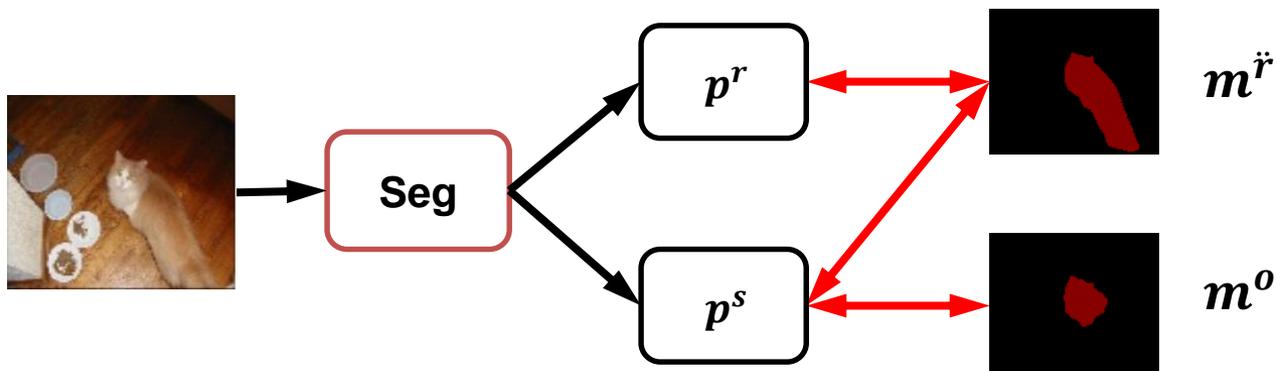




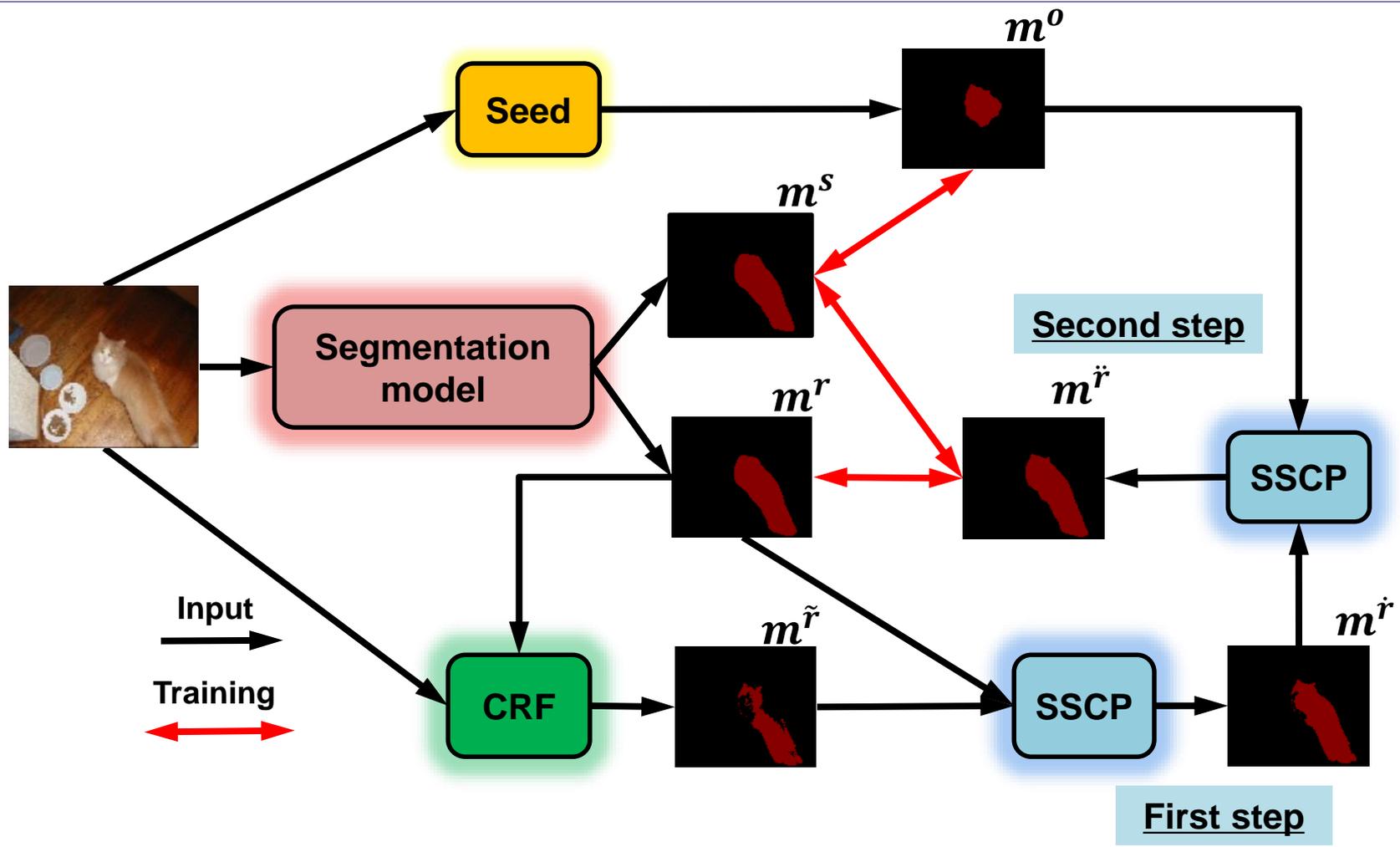
### Training of the second SCPP module



### Training of segmentation network



# 変化領域の推論を活用した動的な領域補正の概要図



# 変化領域の推論モデルの学習

- **First step**

- $m^r$ : Segmentation modelの出力マスク
- $m^{\tilde{r}}$ : CRFの出力結果

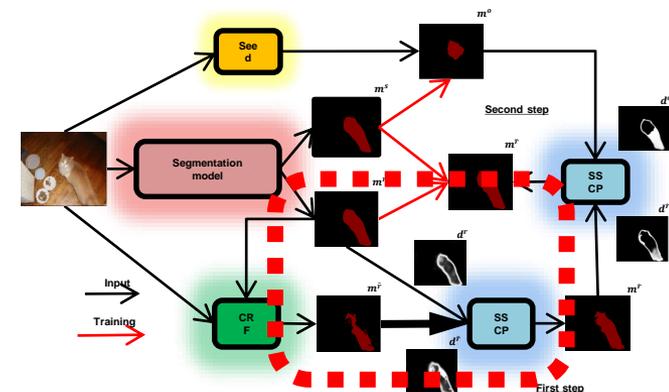
$$\mathcal{L}_{cp-crf}(d^r, d^{\tilde{r}}, M^{r, \tilde{r}}) = \frac{1}{|S|} \sum_{u \in S} J(M^{r, \tilde{r}}, d^r, u) + J(M^{r, \tilde{r}}, d^{\tilde{r}}, u)$$

$$J(M^{r, \tilde{r}}, d^r, u) = M_u^{r, \tilde{r}} \log d_u^r + (1 - M_u^{r, \tilde{r}}) \log(1 - d_u^r)$$

$$J(M^{r, \tilde{r}}, d^{\tilde{r}}, u) = M_u^{r, \tilde{r}} \log d_u^{\tilde{r}} + (1 - M_u^{r, \tilde{r}}) \log(1 - d_u^{\tilde{r}})$$

- 明らかに悪いサンプルを除外

$$\text{if } |\sum \gamma_k^{r, \tilde{r}}| > 0 \text{ then } \bar{\mathcal{L}}_{cp-crf} = 0 \text{ else } \bar{\mathcal{L}}_{cp-crf} = \mathcal{L}_{cp-crf}$$



# 変化領域の推論モデルの学習

## • Second step

### – 入力

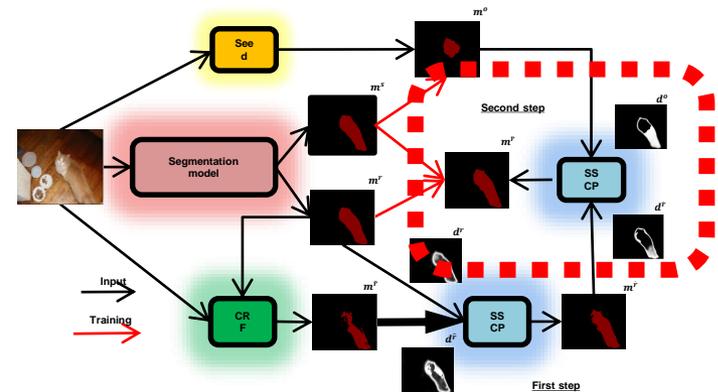
- 一度目の補正結果の出力
- Seed 領域 (固定)

### – 問題点

- Seed領域が固定のため変化領域の推論が容易
- 難易度の差を作るのが困難

### – 解決方法

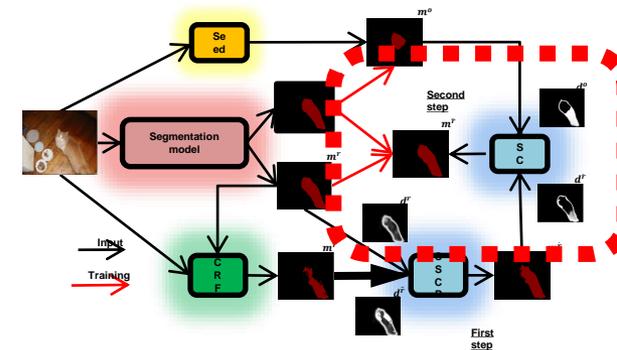
- 学習時には固定されたSeed領域ではなく、Seed領域で学習した出力結果を使用



# 変化領域の推論モデルの学習

- **Second step**

- 学習時には推論と異なるSegmentation maskを使用
- $m^0$ : Seed 領域
- $m^s$ : Seed領域の学習結果
- $m^r$ : 1度目の領域補正結果



$$\mathcal{L}_{cp-seed}(d^o, d^r, M^{o,\hat{r}}) = \frac{1}{|S|} \sum_{u \in S} J(M^{o,s}, d^o, u) + J(M^{\hat{r},o}, d^r, u)$$

$$J(M^{r,\tilde{r}}, d^r, u) = M_u^{r,\tilde{r}} \log d_u^r + (1 - M_u^{r,\tilde{r}}) \log(1 - d_u^r)$$

$$J(M^{r,\tilde{r}}, d^{\tilde{r}}, u) = M_u^{r,\tilde{r}} \log d_u^{\tilde{r}} + (1 - M_u^{r,\tilde{r}}) \log(1 - d_u^{\tilde{r}})$$

# 領域分割モデルの学習

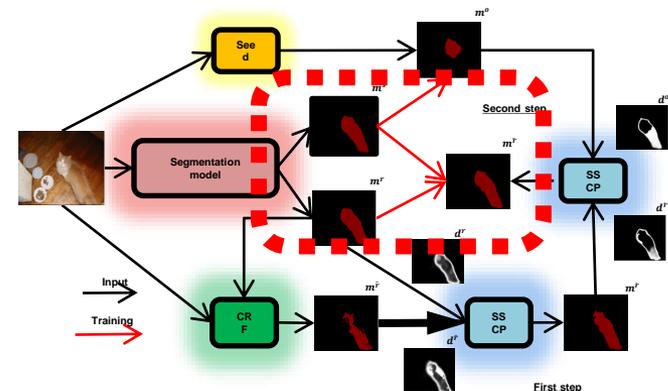
- 二つのSegmentation ブランチを学習

$$\mathcal{L}_{refine} = -\frac{1}{\sum_{k \in K} |S_c^{\bar{r}}|} \sum_{k \in K} \sum_{u \in S_c^{\bar{r}}} \log(h_u^k(x; \theta_r))$$

$$\mathcal{L}_{seed} = 0.5\mathcal{L}^o + 0.5\mathcal{L}^{\bar{r}}$$

$$\mathcal{L}^s = -\frac{1}{\sum_{k \in K} |S_c^o|} \sum_{k \in K} \sum_{u \in S_c^o} \log(h_u^k(x; \theta_o))$$

$$\mathcal{L}^{\bar{r}} = -\frac{1}{\sum_{k \in K} |S_c^{\bar{r}}|} \sum_{k \in K} \sum_{u \in S_c^{\bar{r}}} \log(h_u^k(x; \theta_o))$$



# 学習

- 最終的なロス
  - 二つのブランチの領域分割のロス
  - 二つの変化領域のロス

$$\mathcal{L}_{dynamic} = \mathcal{L}_{refine} + \mathcal{L}_{seed} + \bar{\mathcal{L}}_{cp-crj} + \mathcal{L}_{cp-seed}$$

# 実験

- Pascal VOC dataset 2012を使って実験
  - 学習画像枚数 10582
  - クラス数 21(前景クラス20、背景クラス1)
- 評価指標
  - Mean IoU
  - ピクセルの重なり率

# 静的な領域分割結果の補正

手法	Raw seed (Train set)	Trained segmentation model (Val set)
PSA	52.5	58.4
PSA+CRF	48.0	59.0
静的な領域分割結果の補正	53.4	61.4

# 動的な領域分割結果の補正

手法	Val set
PSA (再現値)	58.4
PSA (文献値)	61.7
動的な領域分割結果の補正	64.7

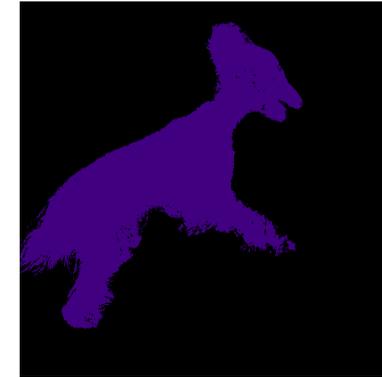
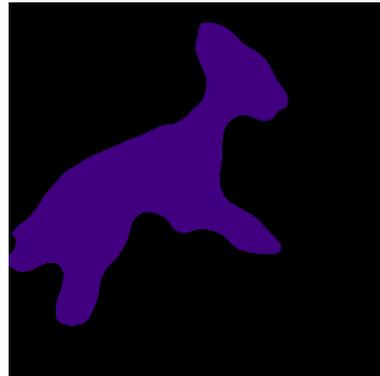
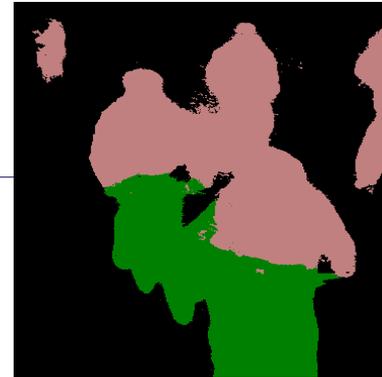
# 同じ条件における他の弱教師あり 領域分割との精度比較

Pascal VOC test set

手法	Val set	Test set
FCN-MIL(ICLR2015)	25.7	24.9
CCNN(ICCV2015)	35.3	35.6
EM-Adapt(ICCV2015)	38.2	39.6
DCSM(ECCV2016)	44.1	45.1
BFBP(ECCV2016)	46.6	48.0
SEC (ECCV2016)	50.7	51.7
TPL(ICCV2016)	53.1	53.8
CBTS(CVPR2017)	52.8	53.7
PSA(CVPR2018)	61.7	63.7
提案手法	64.7	65.3

# 低コストな追加の教師情報を用いた 弱教師あり領域分割との精度比較

手法	追加の教師情報	Val set	Test set
MIL-seg(CVPR2015)	Saliency mask + Imagenet images	42.0	40.6
STC (PAMI2017)	Saliency mask + Web images	49.8	51.2
Hong et al. CVPR2017	Web videos	58.1	58.7
Shen et al. (CVPR2018)	Web images	63.0	63.9
SeeNet(NIPS2018)	Saliency mask	63.1	62.8
AISI(ECCV2018)	Instance saliency mask	63.6	64.5
提案手法	-	64.6	65.3



# まとめ・今後の課題

- まとめ
  - 自己教師あり学習による変化領域の推論モデルを用いた領域補正手法を提案した
  - 提案手法の応用として静的な領域補正手法と動的な領域補正手法の有効性を示した
- 今後の課題
  - 提案手法の更なる改良
    - スケールの変化に対して堅牢な領域分割
  - 半教師あり学習など他の学習手法に応用