

HCS2019年10月研究会

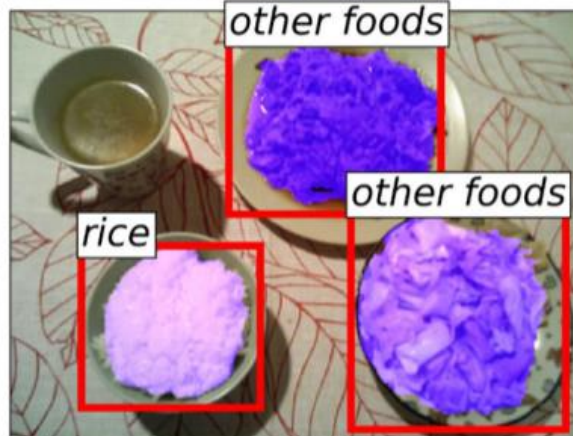
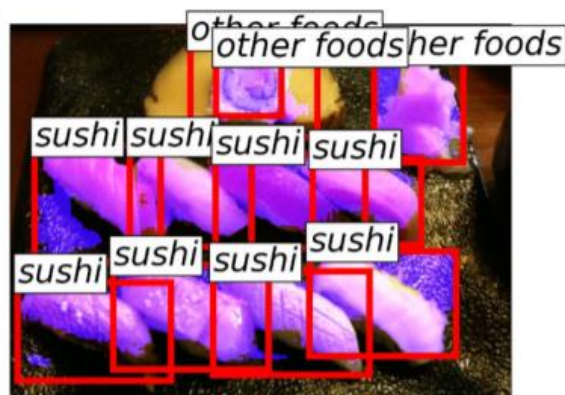
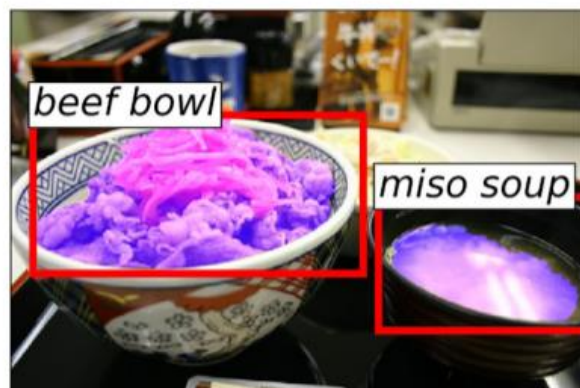
皿領域の推論を活用した食事の弱 教師あり領域分割

2019年10月26日

電気通信大学大学院情報学専攻

下田 和, 柳井啓司

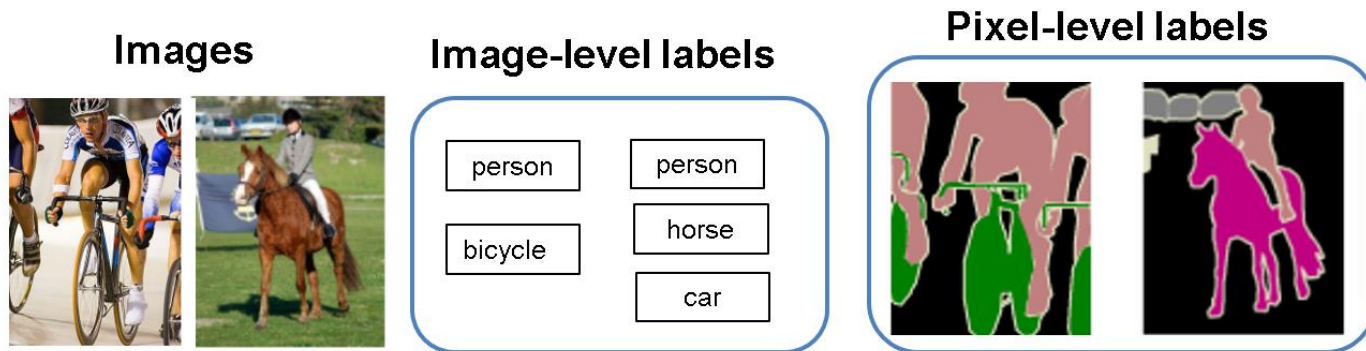
食事画像領域分割



食事量推定
カロリー推定

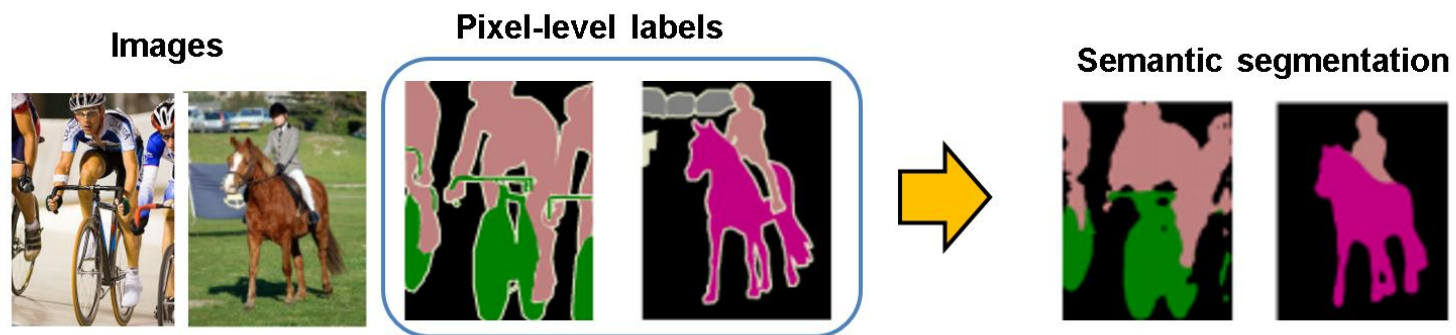
領域分割のアノテーション問題と 弱教師あり学習領域分割

- 画像ラベルのアノテーションと比較して領域分割のアノテーションコストは高コスト
- 画像ラベルから領域分割が学習できれば大幅な学習コスト削減につながる

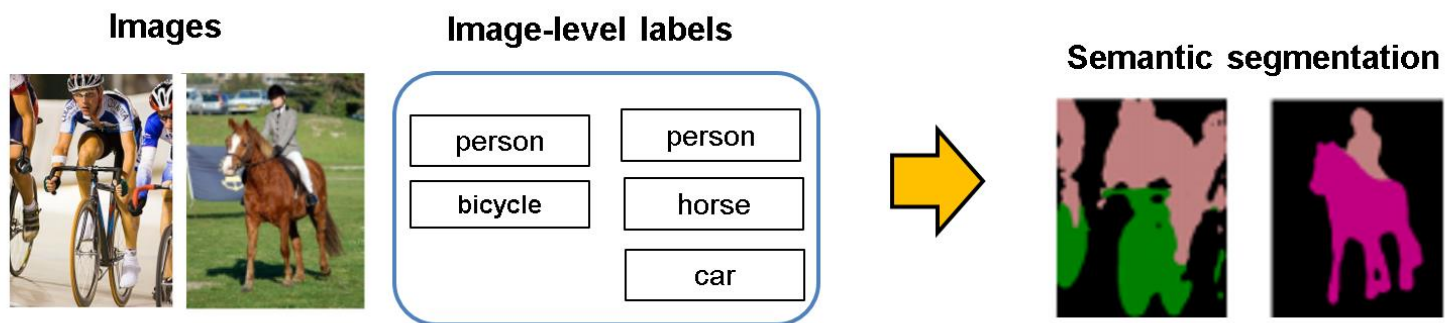


弱教師あり領域分割

Fully-supervised segmentation



Weakly-supervised segmentation



食事における弱教師あり領域分割

- 下田ら[Madima 2016, Madima 2017]
 - クラス分類の可視化を活用した弱教師あり領域分割



問題点

- 皿領域が食事領域に含まれてしまう



今回の研究の目的

- 皿領域を推論
 - アノテーションなし
- 皿領域を推論した結果を活用して、弱教師あり領域分割の精度向上

クラス分類モデルの可視化

- Class activation map(CAM)[Zhou et al. CVPR 2016]

Brushing teeth



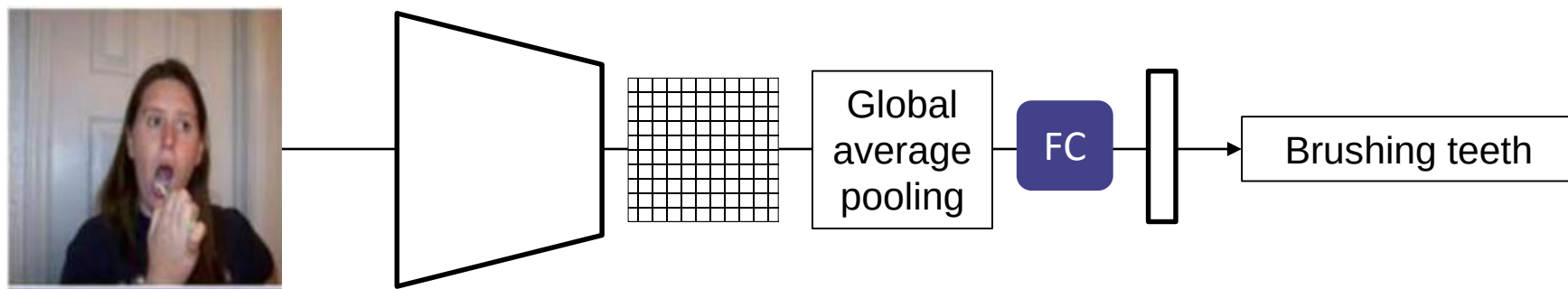
Cutting trees



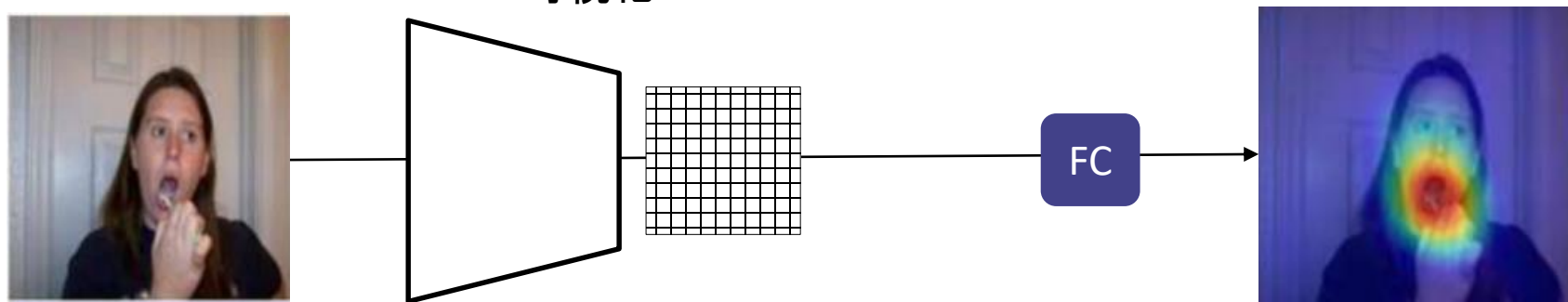
クラス分類モデルの可視化

- Class activation map(CAM)[Zhou et al. CVPR 2016]

学習



可視化



食事画像における クラス分類モデルの可視化

- クラス分類に重要な領域が反応
- 食事のクラス分類器
 - 食事のクラス分類に重要な領域が反応
- 食事/非食事画像の分類器の可視化
 - 食事画像であると判断するうえで重要な領域が反応

食事識別器の可視化と 皿領域の関係

- 食事のクラス分類器の可視化
 - 食事の皿領域は反応しない
 - ほとんどの食事画像に皿領域は含まれているので食事のクラスの認識には寄与しないから
- 食事/非食事画像の分類器の可視化
 - 食事の皿領域が反応する
 - ほとんどの食事画像に皿領域は含まれているので、食事かどうかの認識に寄与するから

皿領域の推論アイデア

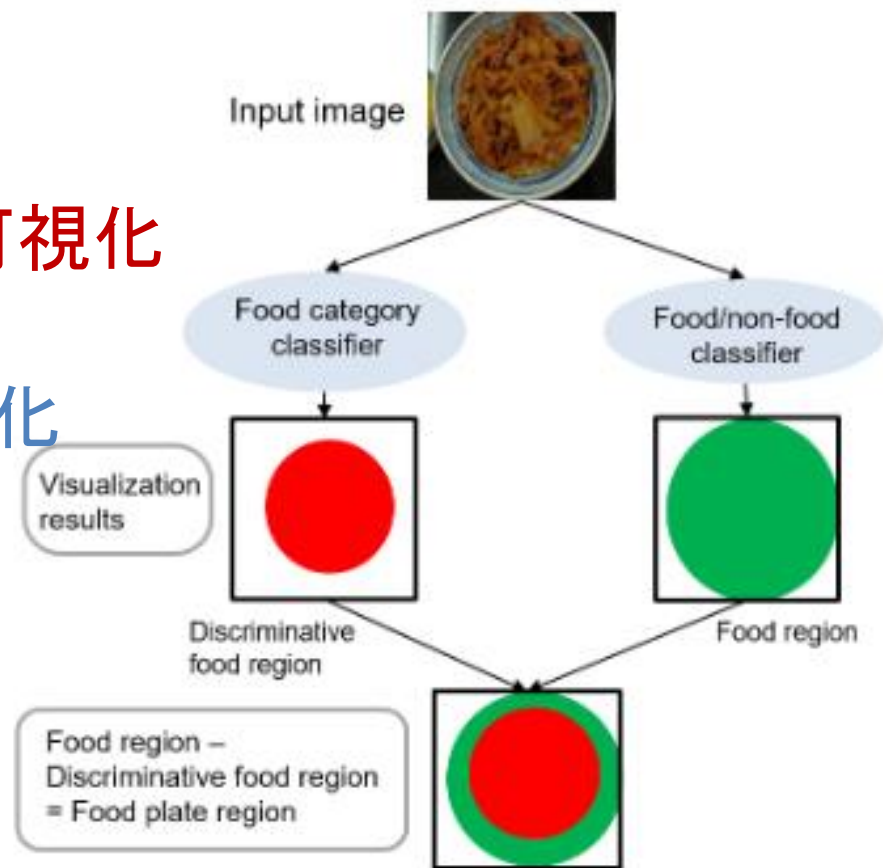
食事/非食事画像の分類器の可視化

—

食事のクラス分類器の可視化

=

食事の皿領域



皿領域マスクの生成

Class activation map (CAM)

食事のクラス分類器の可視化

$$v_F = CAM(x; \theta_F) \in \mathbb{R}^{2 \times H \times W}$$

食事/非食事画像の分類器の可視化

$$v_L = CAM(x; \theta_L) \in \mathbb{R}^{C \times H \times W}$$

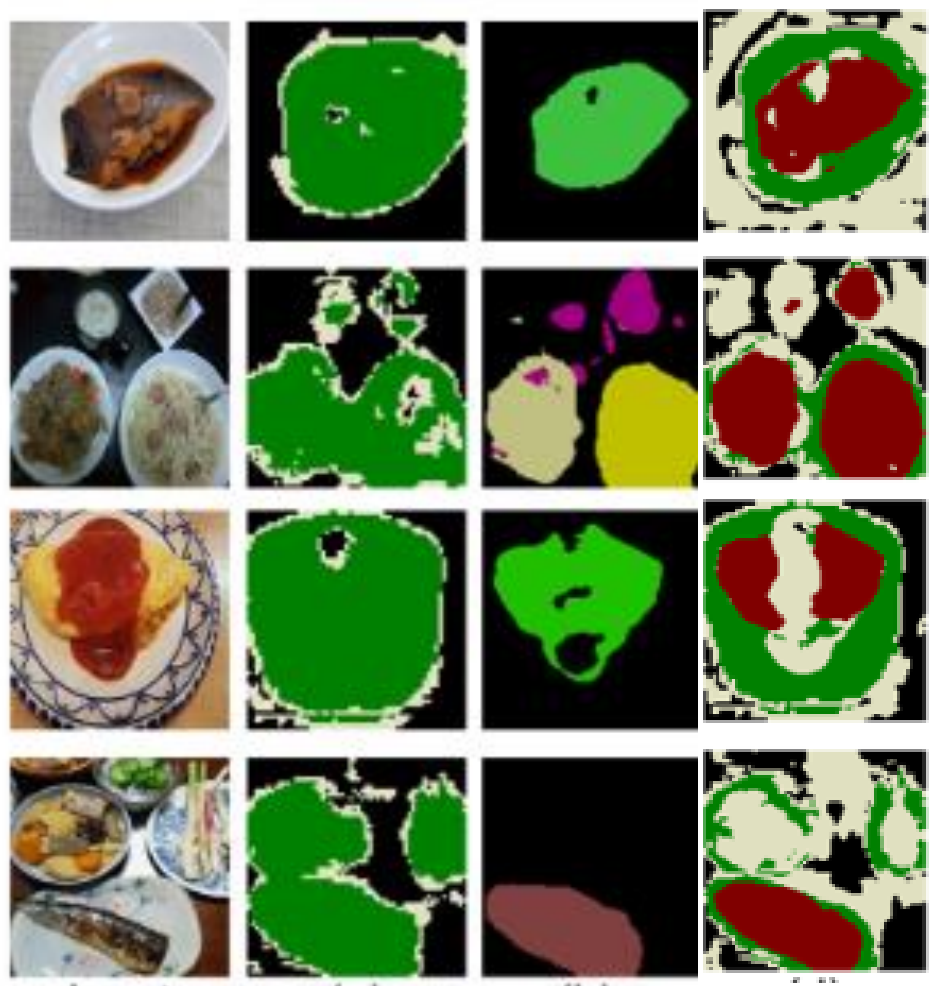
皿領域のピクセル集合: S_P

食事のクラス分類器の可視化における食事領域のピクセル集合: S_F^{fg}

食事/非食事画像の分類器の可視化における食事領域のピクセル集合: S_y^{fg}

$$S_P = S_F^{fg} - S_y^{fg}, y \in L$$

皿領域マスクの生成例



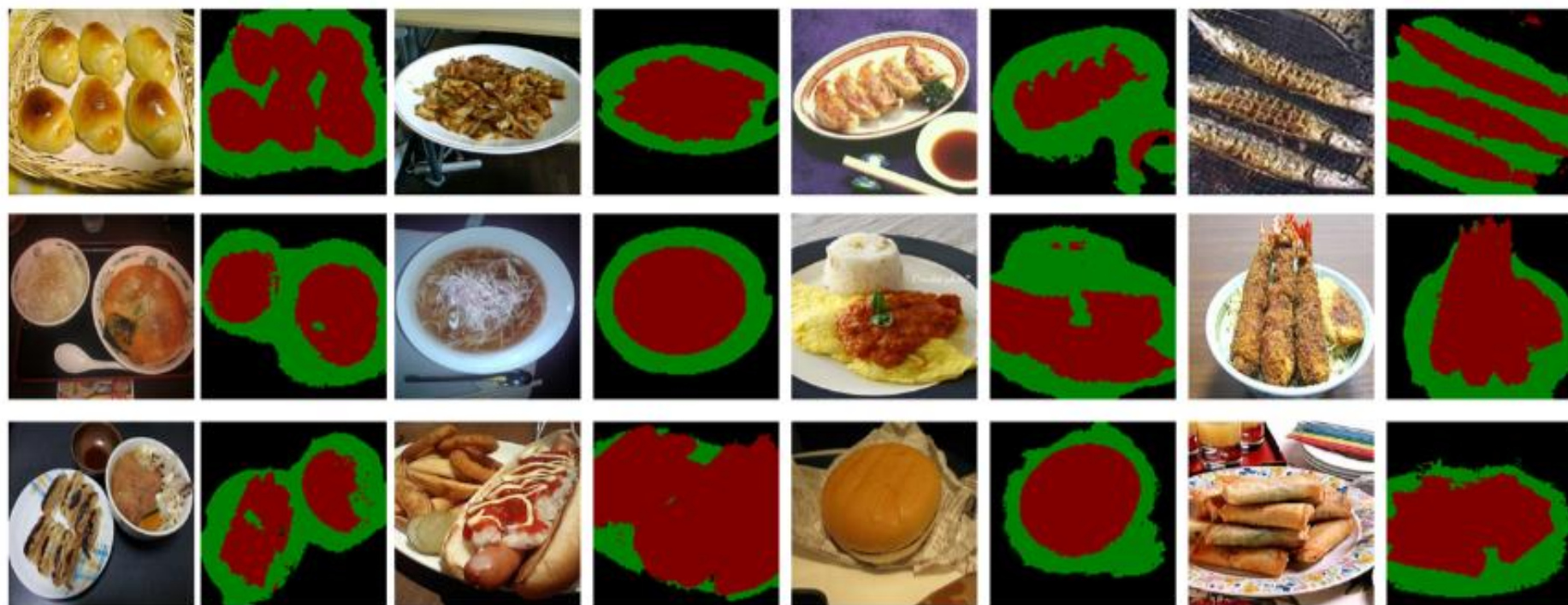
皿領域モデルの学習

The loss of the plate segmentation

$$\mathcal{L}_{plate} = - \frac{1}{\sum_{k=(0,1,2)} |S_k|} \sum_{k=(0,1,2)} \sum_{u \in S_k} \log(h_u^k(x; \theta_P))$$

$$S_0 = S_F^{bg}, S_1 = S_y^{fg} \text{ and } S_2 = S_P$$

皿領域の推論結果

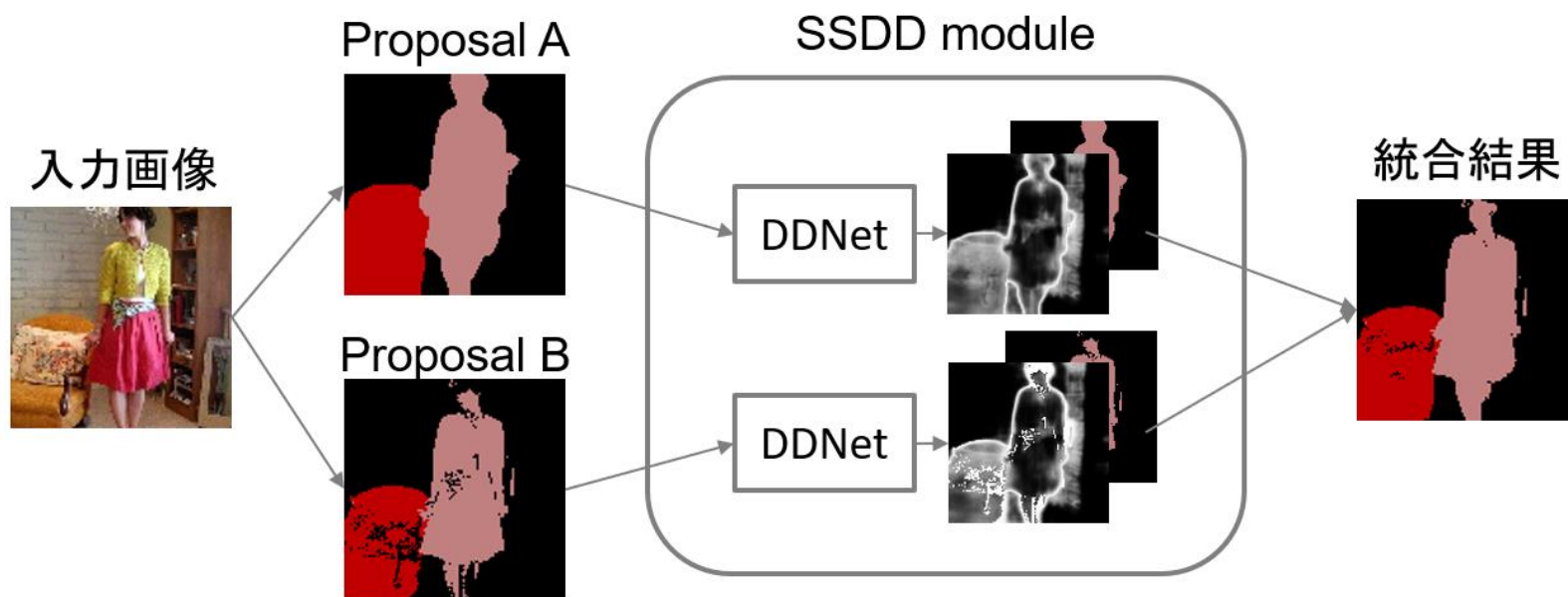


弱教師あり食事領域分割

- 目的
- 皿領域の推論を行う
- 皿領域の推論結果を活用して弱教師あり食事領域分割の精度を向上

弱教師あり領域分割

- 下田ら [ICCV 2019]

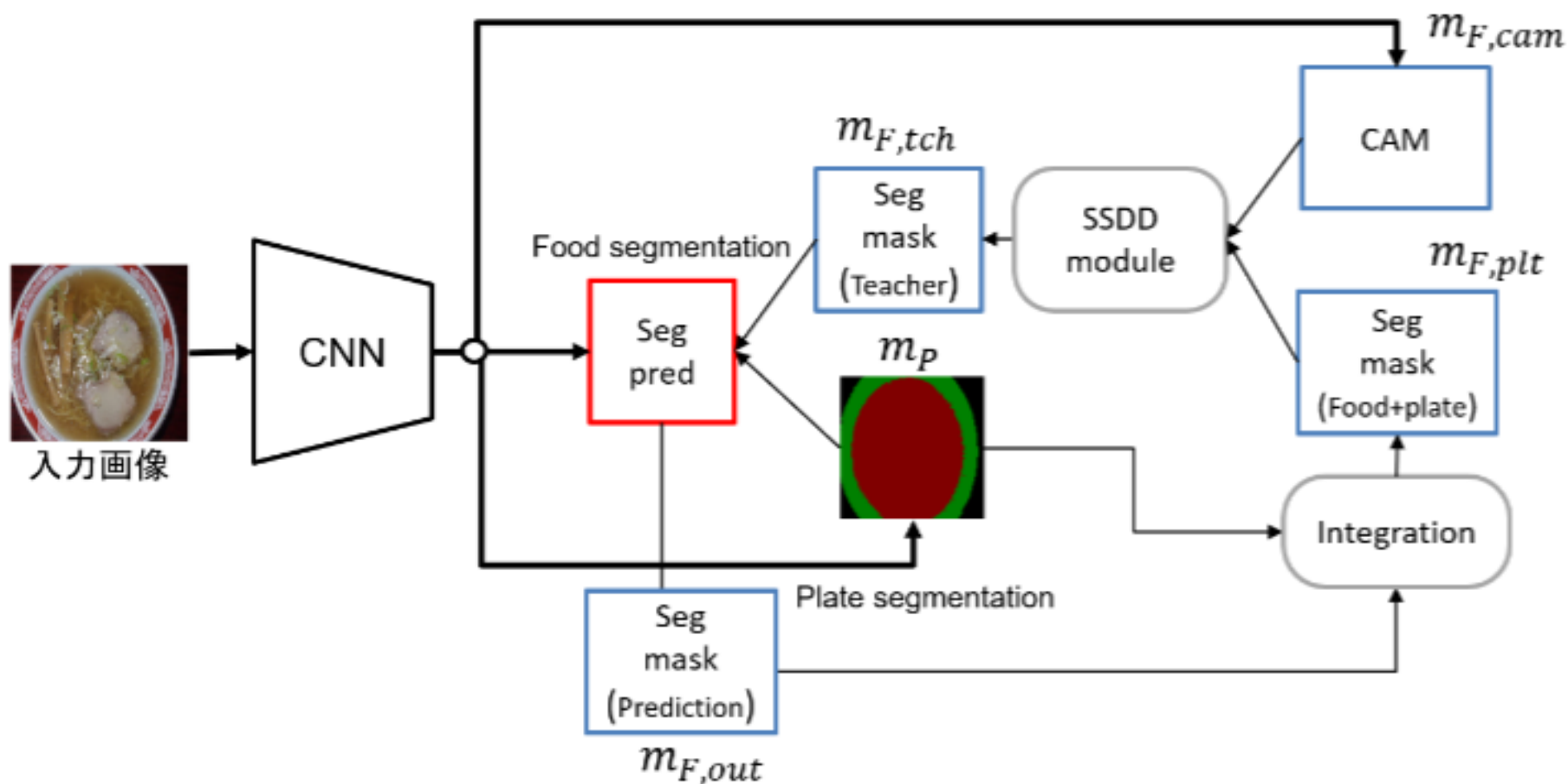


二つの種類の領域分割マスクを統合
今回

Proposal A → CAM

Proposal B → 皿領域 + 食事領域分割モデルの出力

提案手法の概要図



領域分割モデルの学習

- SSDDモジュール(領域の統合結果) S_{tch} を使って領域分割モデルを学習

$$\mathcal{L}_{main} = - \frac{1}{\sum_{k \in \hat{y}} |S_{F,tch}^k|} \sum_{k \in \hat{y}} \sum_{u \in S_{F,tch}^k} \log(h_u^k(x; \theta_{seg}))$$

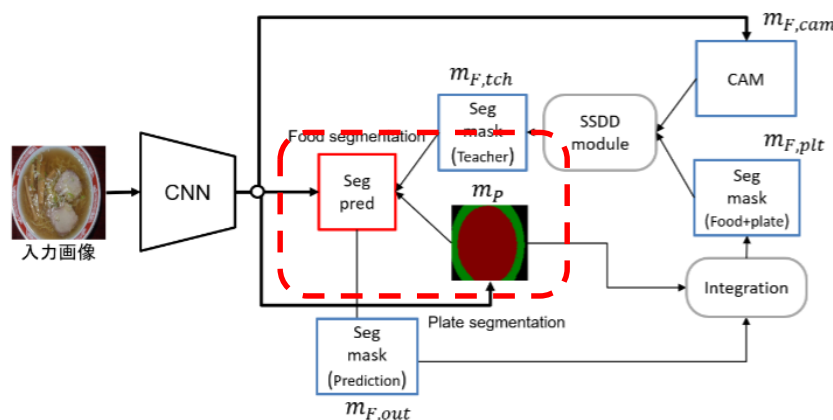
皿領域の推論結果を活用した 弱教師あり食事領域分割

- ① 皿領域の推論結果による食事領域の制限
- ② SSDD moduleの出力結果のCAMへのフィードバック
- ③ 皿領域の推論結果による背景領域の出力についての罰則

①皿領域の推論結果による背景領域の出力についての罰則

- 前景領域が皿領域の推論結果と一致するように学習

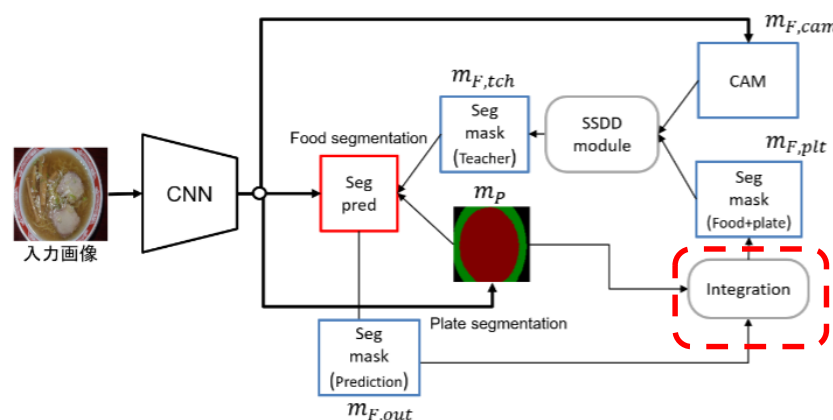
$$\mathcal{L}_{penalty} = -\frac{1}{|S_{P,out}^{food}|} \sum_{u \in S_{P,out}^{food}} \log(-h_u^{bg}(x; \theta_{seg}))$$



② 皿領域の推論結果による食事領域の制限

- 背景領域が皿領域と同一になるように領域分割マスクを修正

$$m_{F,plt} = \begin{cases} m_{F,out} & \text{if } (m_{P,out} = \text{food class}) \\ BG \text{ class} & \text{if } (m_{P,out} = BG \text{ or plate class}) \end{cases}$$

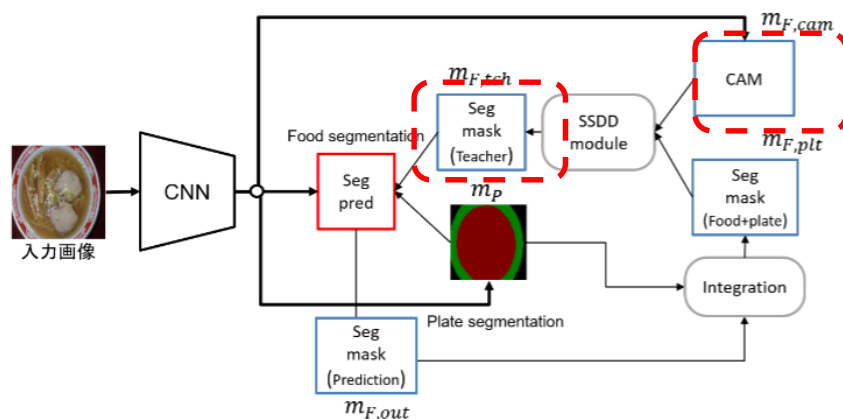


③SSDD moduleの出力結果のCAMへのフィードバック

- CAMの結果とSSDDの出力が同じになるようにクラス分類を学習

$$\mathcal{L}_{feedback} = -\frac{1}{|\hat{y}|} \sum_{k \in \hat{y}} \log(p_d^k(x; \theta_{cl}))$$

$$e_d^k(x; \theta_e) = -\frac{1}{|S_{F,df}^k|} \sum_{u \in S_{F,df}^k} e_h(x; \theta_e)$$



最終的なロス関数

$$\mathcal{L}_{final} = \mathcal{L}_{main} + 0.1\mathcal{L}_{feedback} + 0.1\mathcal{L}_{penalty}$$

実験

- データセット
 - UEC FOOD101
 - 100 classes
 - 10000 images
 - 評価のために領域分割ラベルを1000枚分アノテーション

精度評価

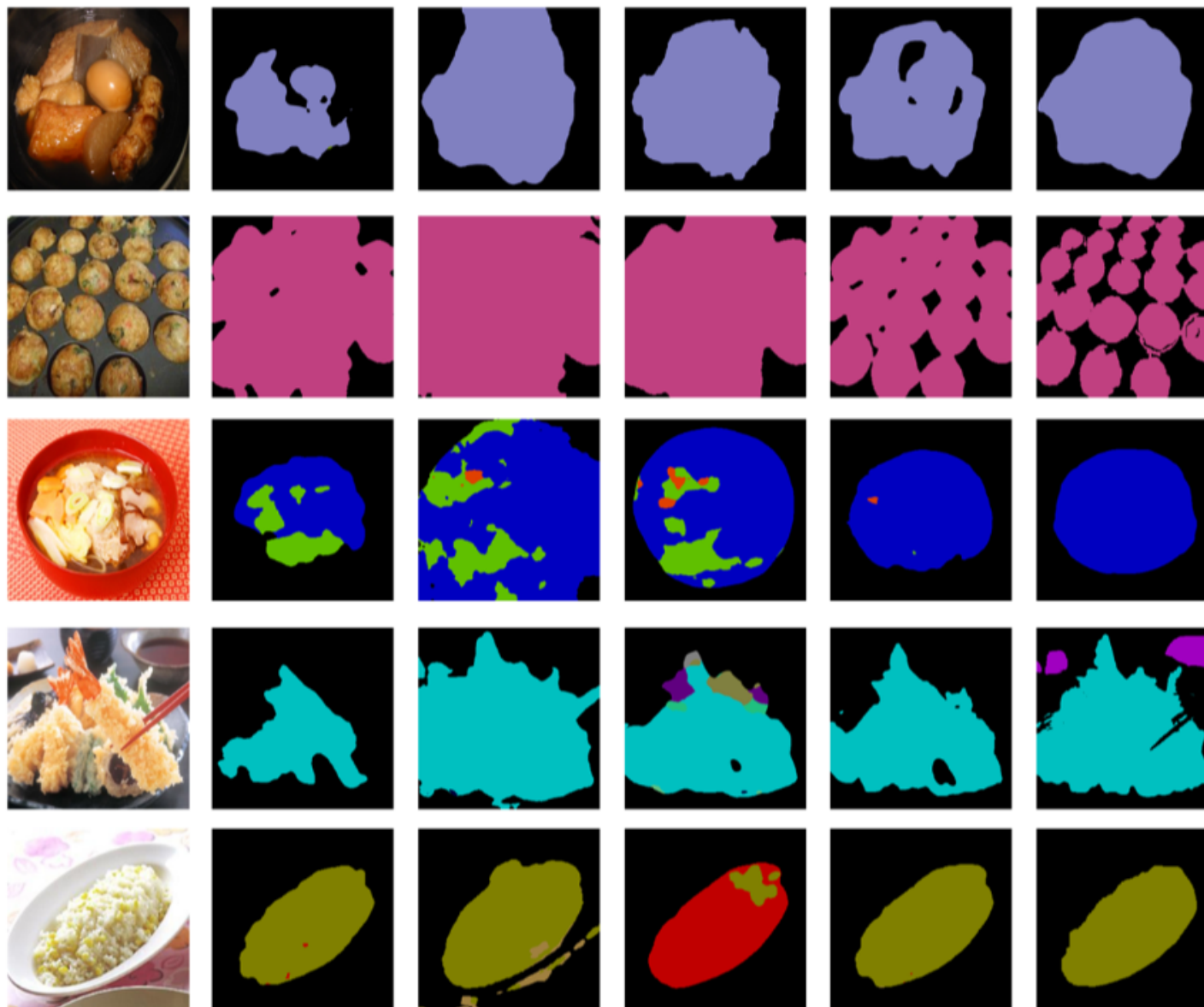
- [1] Simple does it, CVPR 2017
 - Bounding boxのアノテーションを学習に使用

	mIoU	Pixel Acc
Base method	50.2	77.5
BB annotation + GrabCut [1]	51.1	81.9
proposed	52.3	80.4

各手法を取り入れた際の 精度評価

- A: 皿領域の推論結果による食事領域の制限
- B: 皿領域の推論結果による背景領域の出力についての罰則
- C: SSDD moduleの出力結果のCAMへのフィードバック

	(A)	(B)	(C)	mIoU	Pacc
(I)	-	-	-	50.2	77.5
(II)	-	✓	✓	49.8	78.9
(III)	✓	-	✓	46.0	67.3
(IV)	✓	✓	-	51.2	78.2
(V)	✓	✓	✓	52.3	80.4



Input

(II)

(III)

(IV)

(V)

Human supervision

まとめ

- 皿についてのアノテーションを使わずに食事分類器の可視化のみから皿領域の推論を行った
- 皿領域の推論結果から弱教師あり食事領域分割の精度向上を達成した