

食事画像に対する少数およびゼロショット領域分割 (21)

電気通信大学 本部 勇真 柳井 啓司

研究背景

現在のセグメンテーションモデルの学習には、大量の学習データを必要とし、タスクごとにデータを集めるのは非現実的。特に無数カテゴリが存在する食事データセットには、カテゴリごとのデータ不足の問題が存在する。

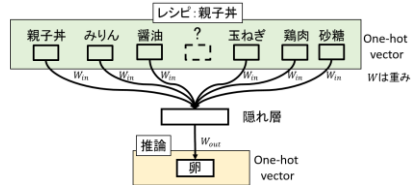
目的

- 数枚もしくは0枚の学習データで学習できる高精度なモデルの作成
- 食事学習データセットの量の不十分さを解消

提案手法

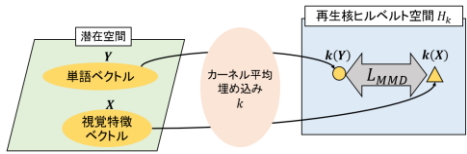
Few-shot SegmentationモデルのPFENetをベースとしたアーキテクチャ

- レシピの食材単語の共起関係に基づく単語埋め込みの追加
 - クックパッドデータセットに含まれる16万のレシピテキストを利用して料理単語のword2vecを学習
 - 埋め込み単語は、VAEによって潜在変数として変換



- Maximum Mean Discrepancyに基づく損失関数 L_{MMD} の追加
 - 2つの分布をGaussianカーネルで再生核ヒルベルト空間に写像した分布間の平均2乗誤差を使用することで分布の違いを定量

$$L_{MMD}^2(X_n, Y_m) = \|k(X) - k(Y)\|_{H_k}^2 = \frac{1}{n^2} \sum_{i,j=1}^n k(X_i, X_j) + \frac{1}{m^2} \sum_{i,j=1}^m k(Y_i, Y_j) - \frac{2}{mn} \sum_{i=1}^n \sum_{j=1}^m k(X_i, Y_j)$$



データセット

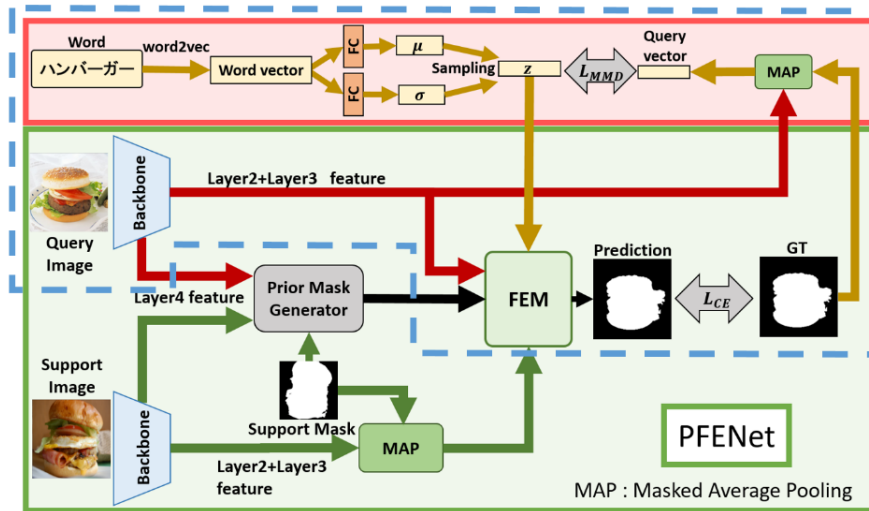
- UECFoodPix-25^l: 102カテゴリ合計1万枚の食事画像データセット
UEC-FoodPix CompleteのFew-shot版
- Pascal-5^l: PASCAL VOC2012データセットのFew-shot版



UECFoodPix-25^l



Pascal-5^l



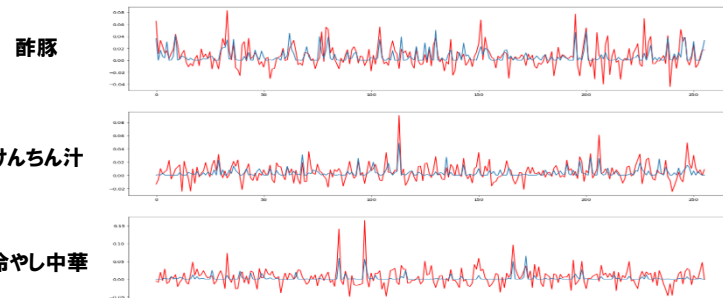
Few-shot モデル (wPFE), Zero-shotモデル (zPFE:青点線)

実験:埋め込み単語の有効性の検証

- word2vecに用いる学習データにwikiの4千万文を用いる場合と、クックパッドのレシピ量(1万と16万)の違いによる実験
- 日本語wikiの辞書に存在しない単語は、食材のベクトルの平均使用

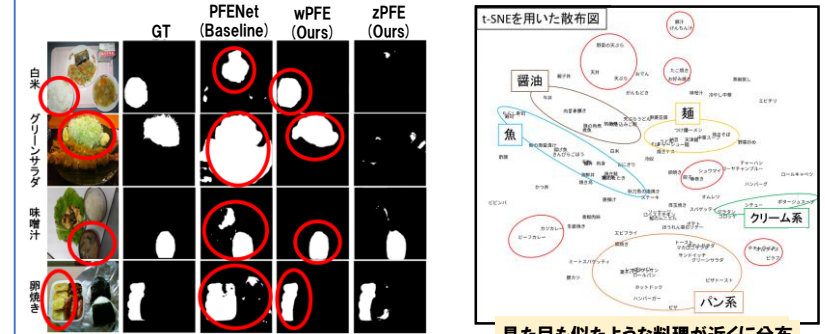
	学習データの違いによる比較			単語の視覚特徴への再構手法の比較		
	CookPad (10K)	CookPad (160K)	Wiki (40M)	+GMMN +MMD	+VAE +MMD	None
Mean	0.796	0.816	0.802	0.834	0.843	0.832

- wikiの文章データよりもレシピの食材データを使用するほうが料理の視覚情報を反映
- VAEを使用したほうがより複雑な表現を獲得できる



赤色:再構成した特徴量、青色:クエリのMAPベクトル

実験:各データセットによる定量/定性分析

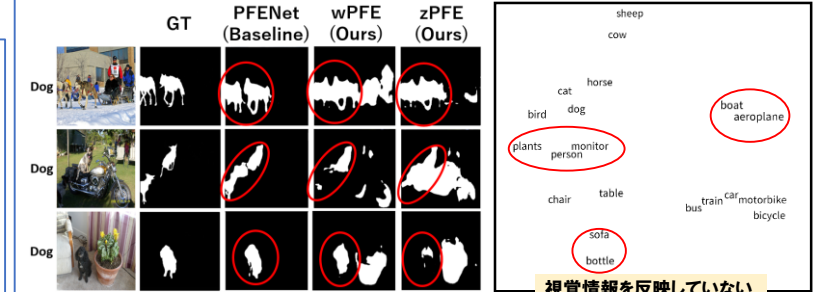


見た目も似たような料理が近くに分布



	UECFoodPix-25 ^l						
	Five-shot		One-shot		Zero-shot		
	wPFE (Ours)	PFE	wPFE (Ours)	PFE	zPFE (Ours)	PFE	Kato
Mean	0.845	0.838	0.843	0.832	0.816	0.786	0.736

- 従来手法を上回った
- 食事画像同士の類似度が高く、対象クラスの知識を補完しやすい



視覚情報を反映していない

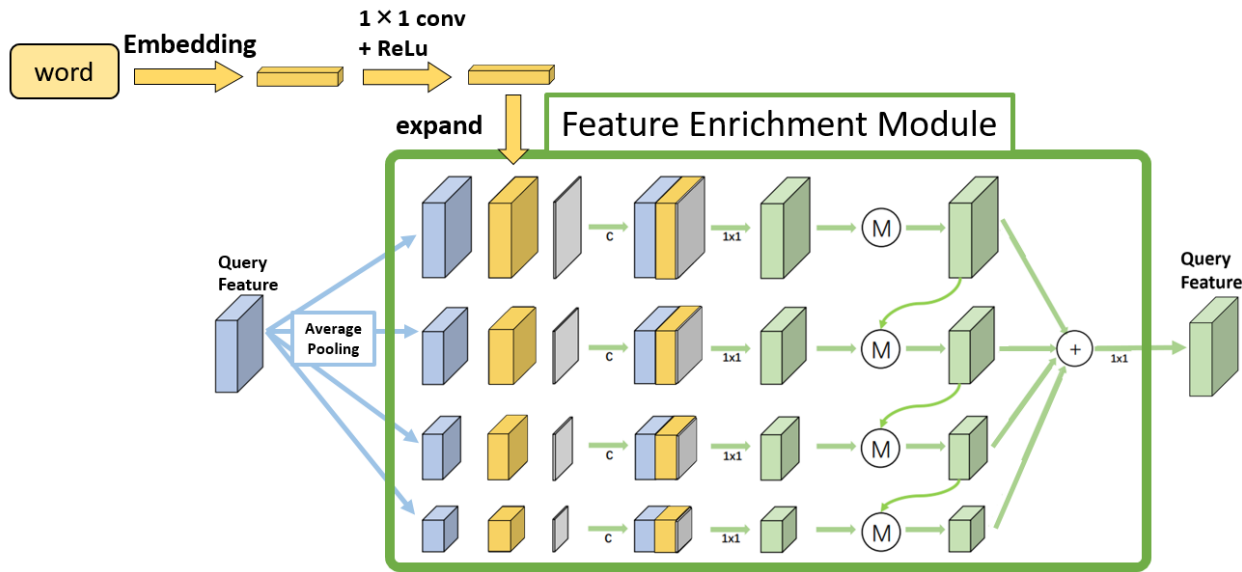
	Pascal-5 ^l						
	Five-shot		One-shot		Zero-shot		
	wPFE (Ours)	PFE	wPFE (Ours)	PFE	zPFE (Ours)	PFE	Kato
Mean	0.591	0.621	0.586	0.607	0.517	0.551	0.454

- 従来手法を下回った
- wikiの文章で学習したword2vecの分散表現には視覚情報が反映されていない

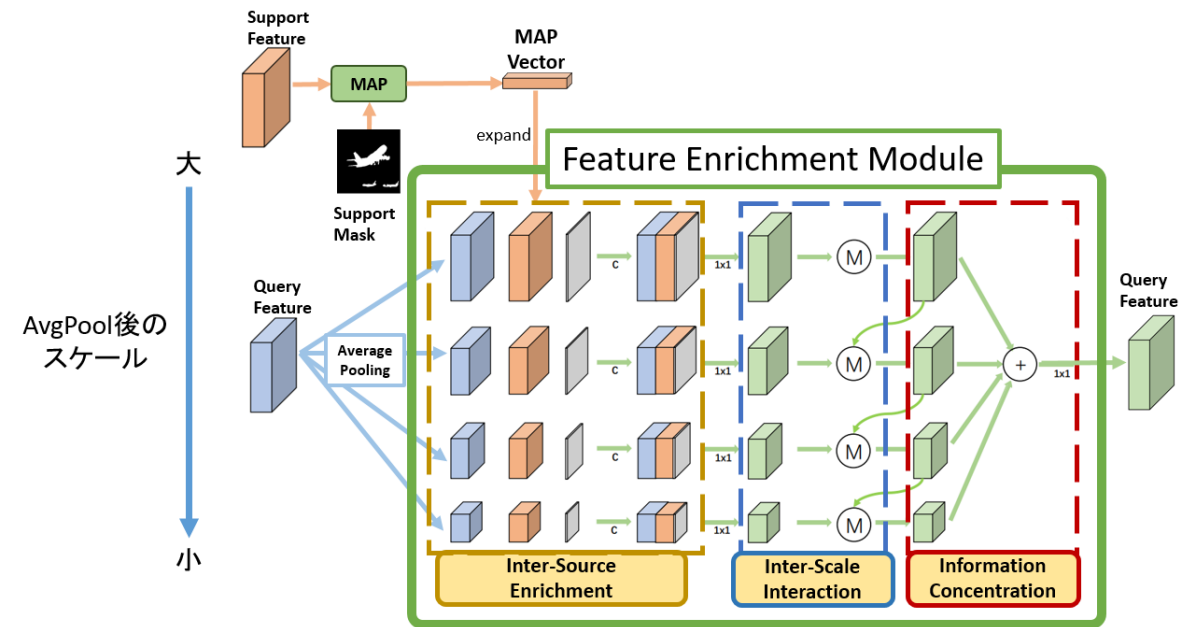
まとめと今後の課題

- PFENetをベースに単語埋め込みと新たな損失を加えた新しいモデルを提案
- UECFoodPix-25^l: 精度改善 → 食事画像同士の類似度が高く、食材単語の共起関係が効果的
- Pascal-5^l: 精度低下 → wikiの文章によるword2vecの学習は、視覚情報を反映していない
- BERTによるより良い埋め込み表現、データセットの拡張

- Prior Guided Feature Enrichment Network(PFENet)[2][Tian, TPAMI2020]

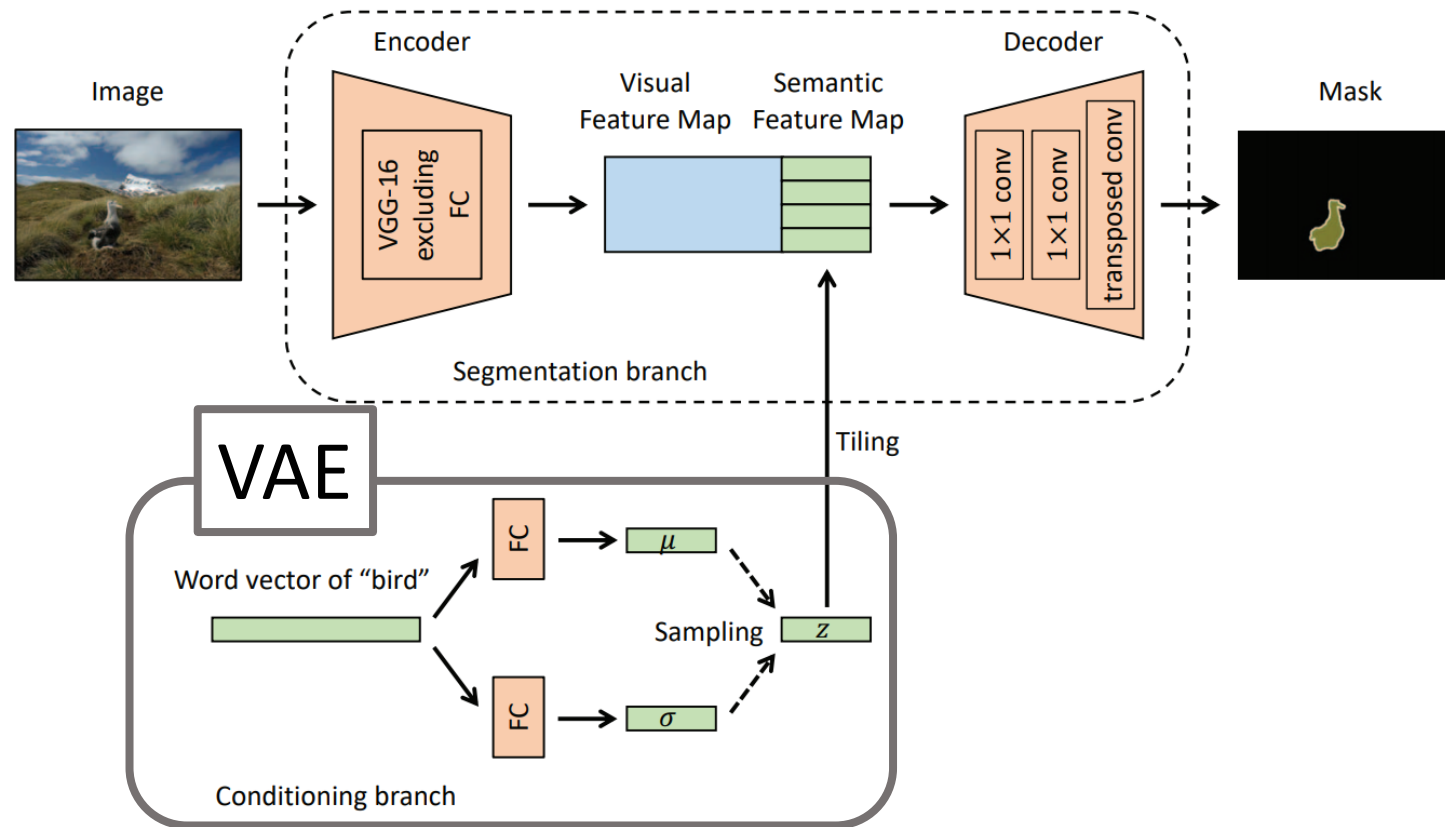


PFENetのZero-shotモデル



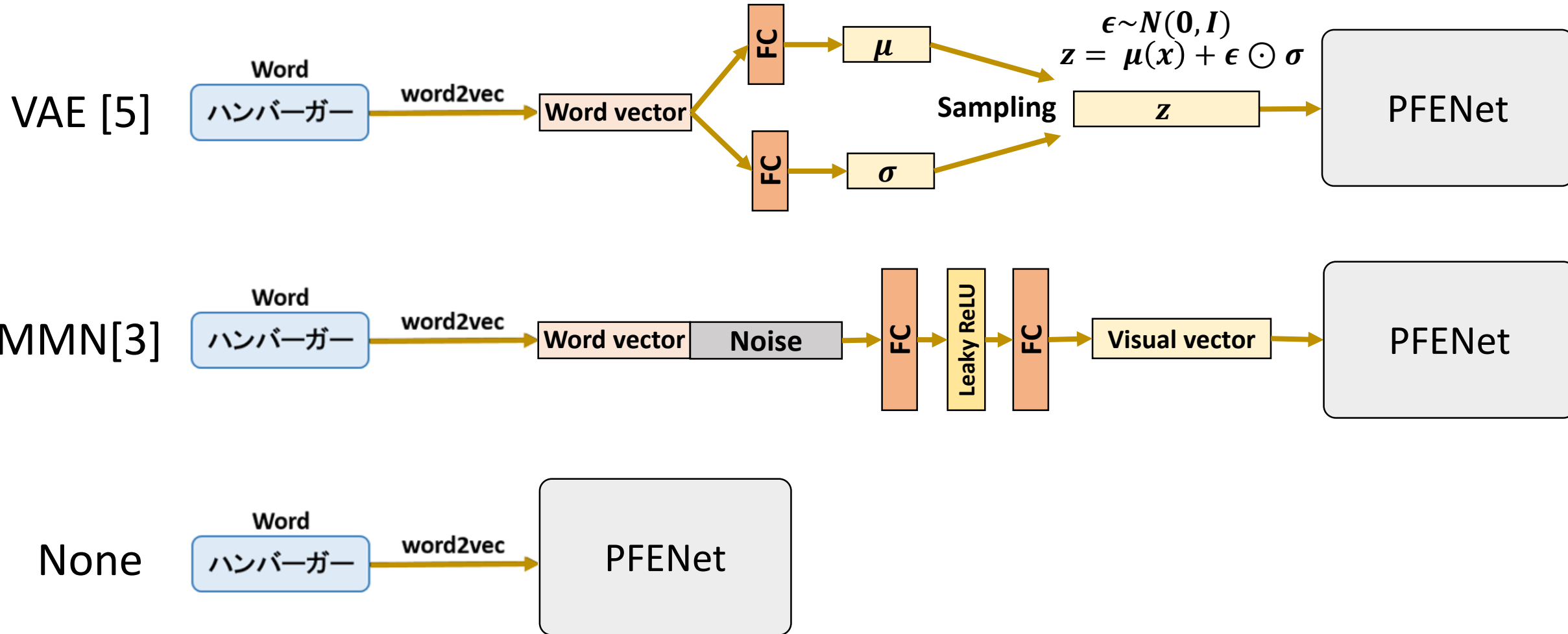
Feature Enrichment Module(FEM):
異なるスケールの特徴を順に畳み込む

- Zero-shot semantic segmentation via variational mapping [5]^[Kato, CVPR Workshops 2019]
 - 単語ベクトルにVAEを用いて潜在空間の特徴 z に変換し視覚特徴として使用
 - 本研究でも, 同様にVAEを使用し視覚特徴へと変換する.



Katoらのモデル

- 実験 :埋め込み単語の有効性の検証
- 単語の再構成手法の比較



食事画像に対する少数およびゼロショット領域分割 (21)

本部勇真 柳井 啓司

UECFoodPix-25ⁱの結果

Fold-0						Fold-1						Fold-2						Fold-3					
	PFE	PFE_5	wPFE	wPFE_5	zPFE		PFE	PFE_5	wPFE	wPFE_5	zPFE		PFE	PFE_5	wPFE	wPFE_5	zPFE		PFE	PFE_5	wPFE	wPFE_5	zPFE
親子丼	0.95	0.95	0.90	0.89	0.94	ピラフ	0.94	0.94	0.93	0.93	0.92	鰻丼	0.83	0.82	0.86	0.86	0.89	白米	0.51	0.51	0.75	0.76	0.41
チキンライス	0.94	0.93	0.94	0.94	0.92	寿司	0.77	0.76	0.77	0.77	0.76	ビーフカレー	0.75	0.77	0.79	0.79	0.78	かつ丼	0.89	0.90	0.91	0.90	0.91
トースト	0.87	0.88	0.88	0.88	0.88	ピビンバ	0.91	0.94	0.94	0.94	0.94	天丼	0.90	0.90	0.91	0.91	0.91	チャーハン	0.91	0.91	0.93	0.94	0.90
菓子パン	0.78	0.78	0.82	0.81	0.79	レーズンパン	0.92	0.93	0.92	0.93	0.93	ロールパン	0.89	0.90	0.90	0.90	0.89	クロワッサン	0.91	0.91	0.90	0.91	0.91
うどん	0.89	0.89	0.91	0.91	0.90	サンドイッチ	0.82	0.82	0.82	0.83	0.81	ピザ	0.96	0.96	0.95	0.95	0.95	ハンバーガー	0.79	0.81	0.81	0.82	0.79
チャーシュー麺	0.92	0.92	0.93	0.93	0.92	ラーメン	0.91	0.91	0.91	0.91	0.89	そば	0.80	0.81	0.84	0.84	0.75	天ぷらうどん	0.90	0.91	0.90	0.90	0.91
お好み焼き	0.91	0.91	0.90	0.90	0.90	スパゲッティ	0.92	0.93	0.94	0.94	0.93	焼きそば	0.91	0.91	0.91	0.91	0.91	天津麺	0.92	0.92	0.85	0.85	0.83
コロッケ	0.88	0.88	0.91	0.90	0.89	野菜炒め	0.85	0.86	0.85	0.85	0.84	グラタン	0.88	0.88	0.87	0.87	0.84	たこ焼き	0.84	0.85	0.87	0.87	0.81
味噌汁	0.67	0.72	0.79	0.81	0.55	野菜の天ぷら	0.84	0.86	0.84	0.85	0.84	ほうれん草のソテー	0.85	0.87	0.88	0.89	0.73	焼きナス	0.85	0.85	0.82	0.82	0.81
オムレツ	0.80	0.79	0.81	0.81	0.80	おでん	0.83	0.81	0.80	0.80	0.79	ソーセージ	0.59	0.59	0.61	0.61	0.57	ポタージュスープ	0.85	0.90	0.90	0.92	0.83
焼き魚	0.81	0.85	0.87	0.87	0.84	シチュー	0.92	0.91	0.91	0.90	0.91	餃子	0.78	0.77	0.81	0.81	0.76	がんもどき	0.69	0.71	0.72	0.74	0.70
刺身	0.67	0.66	0.72	0.73	0.67	鮭のムニエル	0.73	0.74	0.76	0.75	0.75	焼き鮭	0.73	0.75	0.70	0.71	0.62	魚のフライ	0.74	0.76	0.70	0.72	0.71
鯉のたたき	0.77	0.80	0.78	0.80	0.78	酢豚	0.93	0.94	0.93	0.93	0.95	すき焼き	0.94	0.93	0.93	0.92	0.90	秋刀魚の塩焼き	0.76	0.77	0.78	0.79	0.76
豚カツ	0.91	0.91	0.91	0.91	0.89	唐揚げ	0.80	0.82	0.84	0.83	0.83	天ぷら	0.89	0.89	0.89	0.89	0.88	茶碗蒸し	0.78	0.78	0.78	0.78	0.67
ハンバーグ	0.74	0.75	0.66	0.64	0.70	肉じゃが	0.94	0.94	0.93	0.93	0.93	煮魚	0.75	0.77	0.78	0.77	0.77	鯿の南蛮漬け	0.90	0.90	0.89	0.88	0.89
麻婆豆腐	0.96	0.96	0.94	0.96	0.94	生姜焼き	0.79	0.83	0.82	0.85	0.85	干物	0.88	0.88	0.87	0.87	0.86	ステーキ	0.81	0.81	0.78	0.78	0.84
目玉焼き	0.51	0.51	0.57	0.58	0.41	卵焼き	0.76	0.75	0.87	0.87	0.84	ロールキャベツ	0.73	0.72	0.72	0.73	0.72	焼き鳥	0.80	0.81	0.82	0.82	0.79
冷やし中華	0.90	0.92	0.92	0.93	0.93	春巻き	0.86	0.87	0.86	0.86	0.85	冷奴	0.55	0.54	0.52	0.51	0.48	納豆	0.78	0.78	0.79	0.80	0.73
海鮮丼	0.82	0.84	0.85	0.87	0.70	筑前煮	0.92	0.93	0.93	0.94	0.94	豚の角煮	0.85	0.86	0.85	0.85	0.86	青椒肉絲	0.93	0.94	0.94	0.94	0.91
ローストチキン	0.84	0.84	0.84	0.83	0.83	エビチリ	0.88	0.89	0.90	0.90	0.91	鯛焼き	0.93	0.93	0.93	0.93	0.93	ちらし寿司	0.92	0.91	0.92	0.91	0.92
ミートスパゲッティ	0.89	0.91	0.93	0.93	0.91	カツカレー	0.89	0.90	0.88	0.88	0.85	オムライス	0.78	0.81	0.84	0.84	0.83	シュウマイ	0.84	0.84	0.84	0.85	0.82
マカロニサラダ	0.85	0.86	0.86	0.87	0.85	グリーンサラダ	0.50	0.50	0.64	0.65	0.53	ポテトサラダ	0.80	0.82	0.81	0.81	0.80	エビフライ	0.82	0.82	0.81	0.81	0.75
牛丼	0.93	0.93	0.93	0.93	0.92	中華スープ	0.88	0.88	0.81	0.81	0.81	豚汁	0.92	0.94	0.94	0.94	0.93	けんちん汁	0.91	0.93	0.87	0.84	0.94
つけ麺	0.66	0.65	0.67	0.71	0.39	ピザトースト	0.95	0.94	0.94	0.94	0.93	おにぎり	0.76	0.75	0.78	0.76	0.69	きんぴらごぼう	0.88	0.89	0.87	0.88	0.86
ゴーヤチャンプルー	0.93	0.94	0.93	0.93	0.94	炊き込みご飯	0.91	0.90	0.88	0.89	0.89	ポテト	0.55	0.57	0.57	0.56	0.44	ホットドック	0.88	0.89	0.89	0.89	0.90
mIoU	0.832	0.839	0.847	0.851	0.808	mIoU	0.855	0.860	0.865	0.867	0.857	mIoU	0.807	0.814	0.818	0.817	0.788	mIoU	0.832	0.840	0.842	0.844	0.811

食事画像に対する少数およびゼロショット領域分割 (21)

本部勇真 柳井 啓司

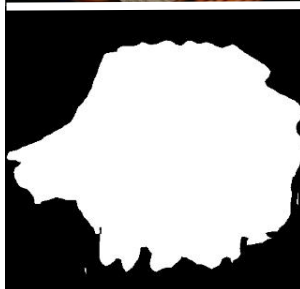
UECFoodPix-25ⁱの結果

support



ポテト(mIoU0.57)

成功例

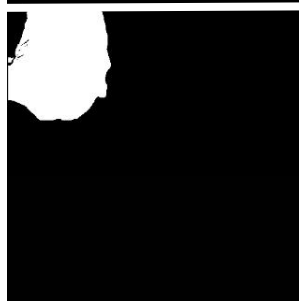


Pred



GT

失敗例

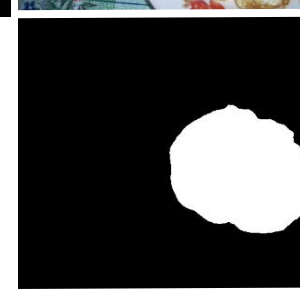


support

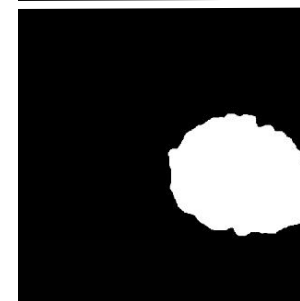


目玉焼き(mIoU0.58)

成功例

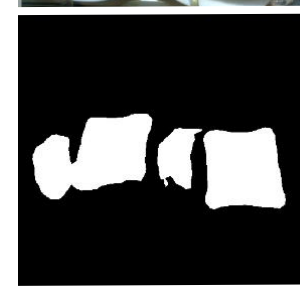
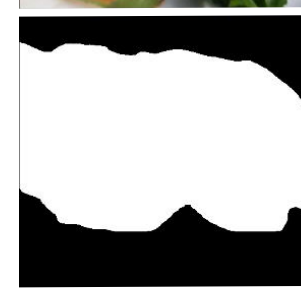


Pred



GT

失敗例



食事画像に対する少数およびゼロショット領域分割 (21)

本部勇真 柳井 啓司

UECFoodPix-25ⁱの結果: サポートの違いによる比較
グリーンサラダ(mIoU0.65)

サポートの情報が弱い場合

サポートの情報が強い場合

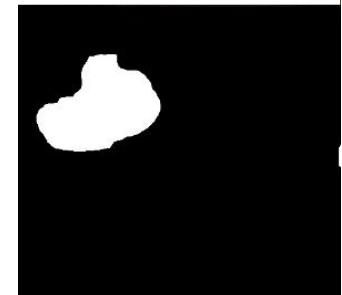
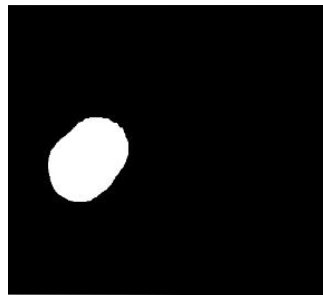
Support

Query

Support

Query

Query



Pred

Pred

GT

GT

食事画像に対する少数およびゼロショット領域分割 (21)

本部勇真 柳井 啓司

Pascal-5ⁱの結果

Fold-0						
	One-shot		Zero-shot		five-shot	
	wPFE	PFE	zPFE	PFE	wPFE	PFE
aeroplane	0.780	0.816	0.680	0.719	0.787	0.82
bicycle	0.336	0.369	0.234	0.307	0.325	0.368
bird	0.807	0.818	0.742	0.738	0.823	0.826
boat	0.583	0.593	0.516	0.430	0.611	0.618
bottle	0.490	0.494	0.449	0.417	0.523	0.509
Mean	0.599	0.618	0.524	0.522	0.614	0.628

Fold-1						
	One-shot		Zero-shot		five-shot	
	wPFE	PFE	zPFE	PFE	wPFE	PFE
bus	0.875	0.887	0.845	0.847	0.879	0.888
car	0.629	0.602	0.552	0.625	0.616	0.611
cat	0.843	0.877	0.747	0.829	0.846	0.880
chair	0.196	0.203	0.176	0.239	0.219	0.233
cow	0.863	0.926	0.866	0.910	0.856	0.932
Mean	0.681	0.699	0.637	0.690	0.683	0.709

Fold-2						
	One-shot		Zero-shot		five-shot	
	wPFE	PFE	zPFE	PFE	wPFE	PFE
table	0.155	0.151	0.127	0.132	0.174	0.150
dog	0.783	0.865	0.633	0.812	0.797	0.875
horse	0.853	0.870	0.803	0.851	0.862	0.878
motorbike	0.753	0.822	0.722	0.778	0.744	0.821
person	0.074	0.106	0.043	0.050	0.065	0.097
Mean	0.523	0.563	0.465	0.524	0.528	0.564

Fold-3						
	One-shot		Zero-shot		five-shot	
	wPFE	PFE	zPFE	PFE	wPFE	PFE
plants	0.292	0.229	0.236	0.269	0.265	0.245
sheep	0.877	0.919	0.787	0.863	0.882	0.922
sofa	0.462	0.565	0.227	0.397	0.459	0.602
train	0.751	0.797	0.736	0.704	0.752	0.806
monitor	0.323	0.322	0.226	0.099	0.336	0.345
Mean	0.541	0.566	0.442	0.467	0.539	0.584