SS4-18

CNNの逆伝搬を利用した食事画像の領域分割

下田 和, 柳井 啓司
電気通信大学 大学院情報理工学研究科 総合情報学専攻

目的

食事画像の領域分割 FOOD

CNNの逆伝搬を用いた高精度な物体検出＋領域分割

－入力は画像とCNNのモデルのみ

－ピクセル単位のアノテーションが不要

ゆくゆくはカロリー計算なども

アプローチ

RCNN [1]

大量の領域候補をプロポーザルし、それぞれの領域をCNNで認識

（2014 PASCALの物体検出タスクにおいてトップの精度）

本手法

大量の領域候補をプロポーザルし、

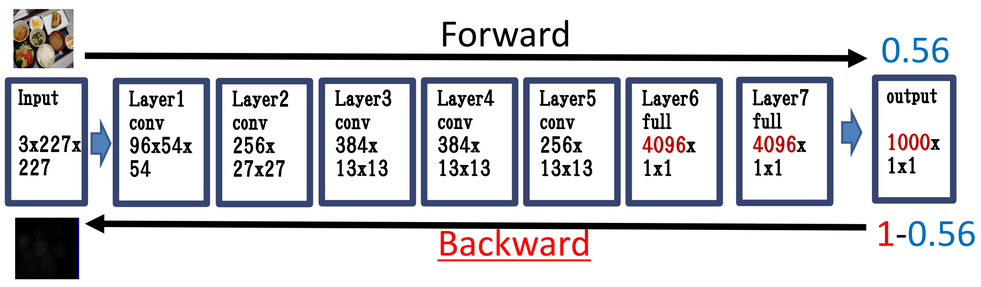
それぞれの領域でCNNによる逆伝搬を利用した領域分割

手法概要

逆伝搬（バックプロパゲーションBP）とは

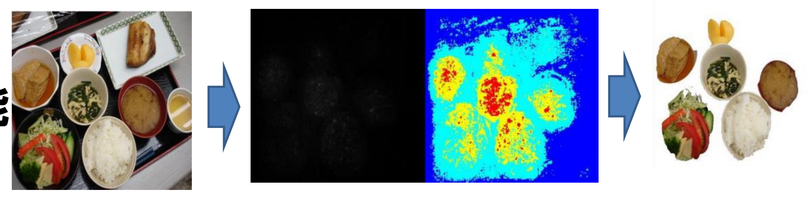
CNNの階層的なパラメータを学習する際の手法

誤差を画像レベルにまで伝搬させることで可視化が可能



逆伝搬を用いた領域分割 [2]

逆伝搬＋グラブカット



逆伝搬　グラブカット

（ただし、複数のピークに対応できない　プロポーザル）

CNNの学習済みモデルのみ

ピクセル単位のアノテーションが不要

Google

他の応用事例

Deep dream

逆伝搬を複数回繰返

Food dream

手法詳細

1 Selective Serchを用いて候補領域をプロポーザル

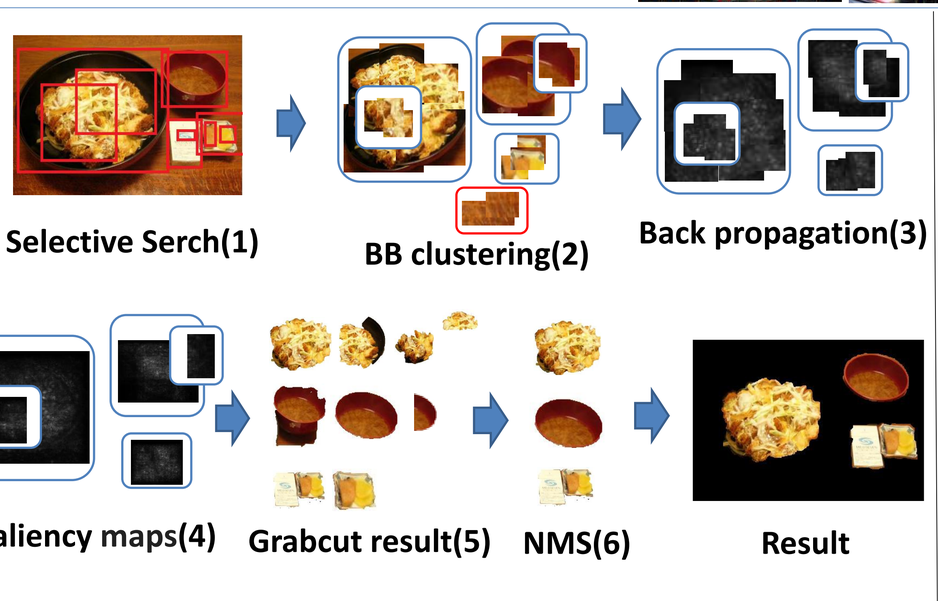
2 BBを重なり率からクラスタリング（NMS）

3 各バウンディングボックスでバックプロパゲーション

4 それぞれのグループで平均をとる

5 サリエンシーマップを用いてグラブカット

6 Non Maximum Suppresion（NMS）で結果の統合



結果

PASCAL 2012 一般画像の領域分割（弱教師有学習）ピクセル単位のずれで評価

VOC 2012	bg	aero	bike	bird	boat	btl	bus	car	cat	chair	cow	dtable	dog	horse	mbike	person	plant	sheep	sofa	train	tv	mAP
Zhang et al[3]	75	47	36	65	15	35	82	43	62	27	47	36	41	73	50	36	46	32	13	42	33	**44.6**
Our method	77.2	42.0	19.2	27.5	21.2	33.9	44.9	47.2	41.3	12.4	35.7	18.8	42.3	29.7	43.6	40.6	27.4	50.5	19.7	46.9	41.7	**36.4**

PASCAL 2007 一般画像の物体検出

評価基準（mAP）

重なり率50％以上なら正解、それ以下なら不正解

	aero	bike	bird	boat	btl	bus	car	cat	chair	cow	dtable	dog	horse	mbike	person	plant	sheep	sofa	train	tv	mAP
RCNN	64.2	69.7	50.0	41.9	32.0	62.6	71.0	60.7	32.7	58.5	46.5	56.1	60.6	66.8	54.2	31.5	52.8	48.9	57.9	64.7	**54.2**
Our method	81.5	70.2	65.2	39.7	37.8	63.9	83.2	67.8	27.0	65.3	39.5	63.6	63.2	73.2	61.2	37.3	63.5	39.8	70.0	60.8	**58.7**

UECFOOD101 複数食事画像の物体検出

UECFOOD100 mAP	100 class (all)	53 class (#item >= 10)	11 class (#item >= 50)
RCNN	26.0	21.8	25.7
Ours	**49.9**	55.3	**55.4**

複数食事画像のデータセットに偏りがあったので、

条件をわけて精度を算出。

（例 ごはんは300枚以上あるが、うな重は10枚以下など）

FOOD

それぞれReluの際の伝搬手法を変更　Relu back (forward max (x,0))

guided back propagation　y = z * (x > 0)*(z > 0)

deconvolution　y = z * (z > 0)

back propagation　y = z * (x > 0)

UECFOOD100 mAP	100 class (all)	53 class (#item >= 10)	11 class (#item >= 50)
guided back propagation	**50.7**	52.5	51.4
deconvolution	48.0	54.1	**55.4**
back propagation	49.9	**55.3**	**55.4**

他の伝搬法でも評価

[1]R. Girshick et al. R ich feature hierarchies for accurate object detection and semantic segmentation.CVPR, 2014
[2]K. Simonyan et al. Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps. ICLR, 2014
[3]W. Zhang et al. Weakly Supervised Semantic Segmentation for Social Images. CVPR, 2015