

## 目的

### CNNを用いた弱教師あり領域分割

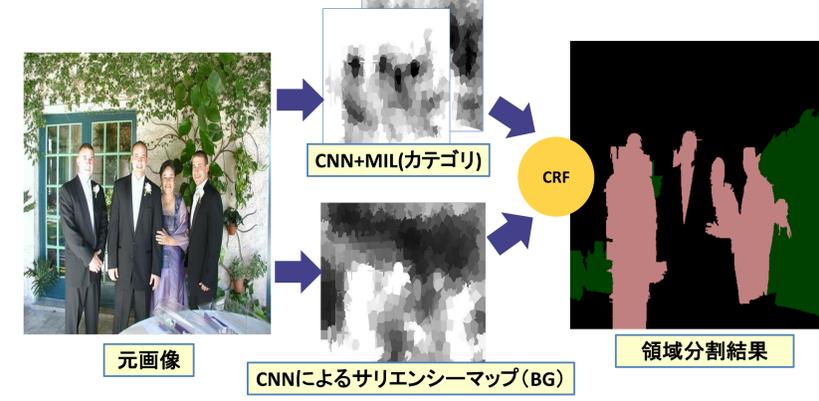
- 完全教師あり学習の領域分割
- ピクセル単位のアノテーションを用いて学習
- 弱教師あり学習の領域分割
- 車、バス、人のラベルのみで学習



## アプローチ

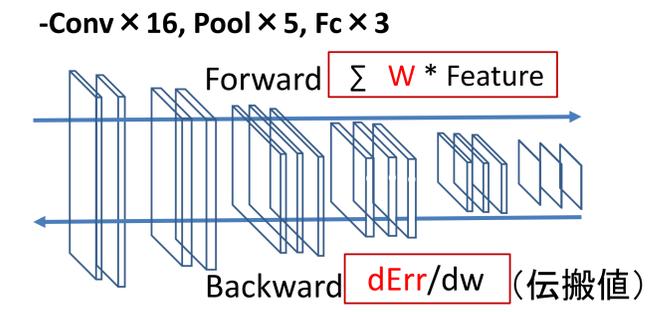
### 手順概要

- (1) CNNによるカテゴリの認識
- (2) BPによるサリエンスマップ [1]
- (背景領域の認識)
- (3) Zoom out feature [2]
- (カテゴリ領域の認識)
- (4) Super pixel + CRF



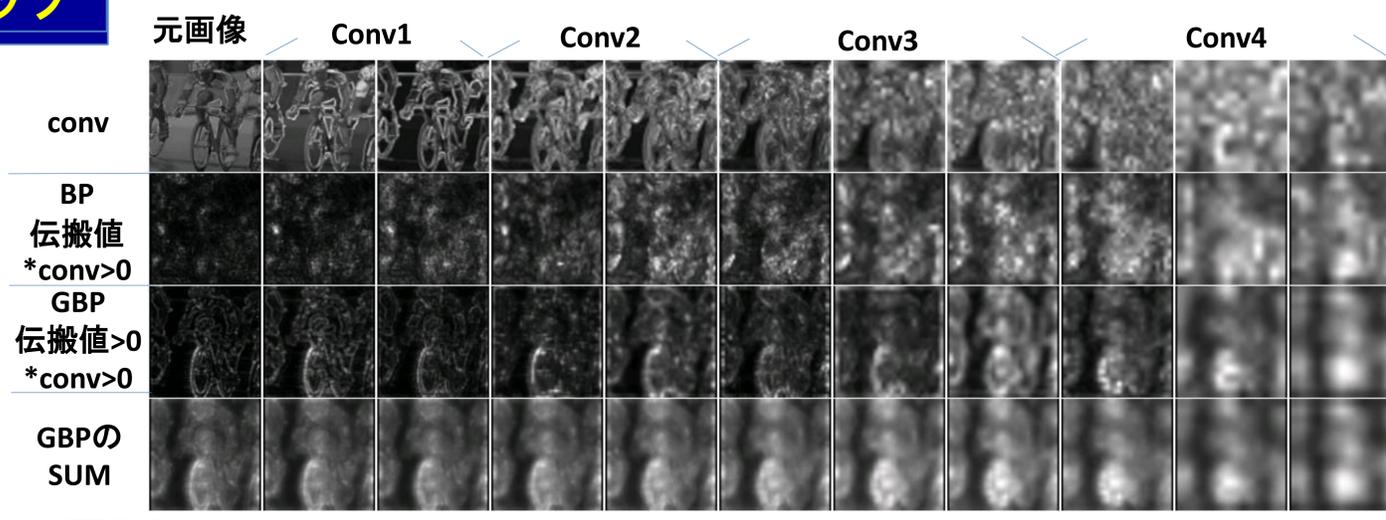
## VGG16のBPによるサリエンスマップ

### VGG16 [5]



### マルチラベルによる学習

- 通常 : 自転車個人↓ + 自転車↓人↑
  - マルチラベル: 自転車個人↑
  - Sigmoid cross entropy loss
- $$E = -\frac{1}{n} \sum_{n=1}^N [p_n \log \hat{p}_n + (1 - p_n) \log(1 - \hat{p}_n)]$$



-従来手法では、元画像レベルのみのサリエンスマップを利用  
 -本研究では、中間層もサリエンスマップとして活用  
 -バイリニア変換でリサイズをして縦横比を統一し、カーネルのMaxをとる

## Zoom out feature + MIL

Zoom out feature [2]  
 完全教師あり学習  
 CNNを局所特徴量に (Super pixel単位)

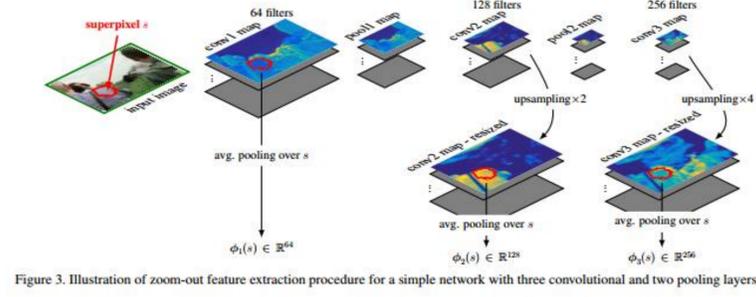
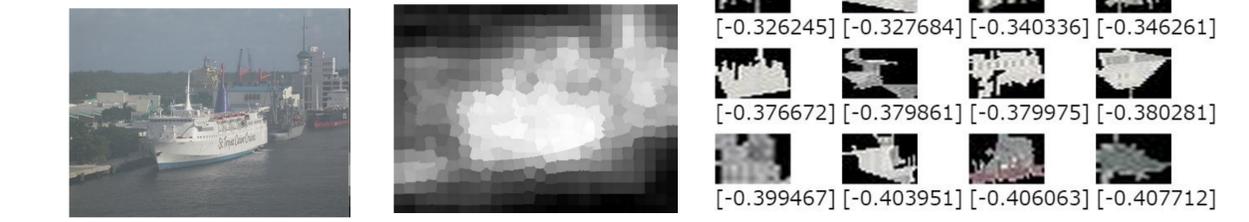


Figure 3. Illustration of zoom-out feature extraction procedure for a simple network with three convolutional and two pooling layers.

### 完全教師あり学習→弱教師あり学習に適用

- Multiple Instance Learning (MIL) [4]
- 既存手法 (SiftやHogでも使われていた)
- SVMを使ったカテゴリ領域の学習
- ポジティブ画像 ネガティブ領域とポジティブ領域
- ネガティブ画像 ネガティブ領域
- {各カテゴリで学習、リランキングを繰り返し、ネガティブを減らしていく}
- Super pixel領域をリランキングした例



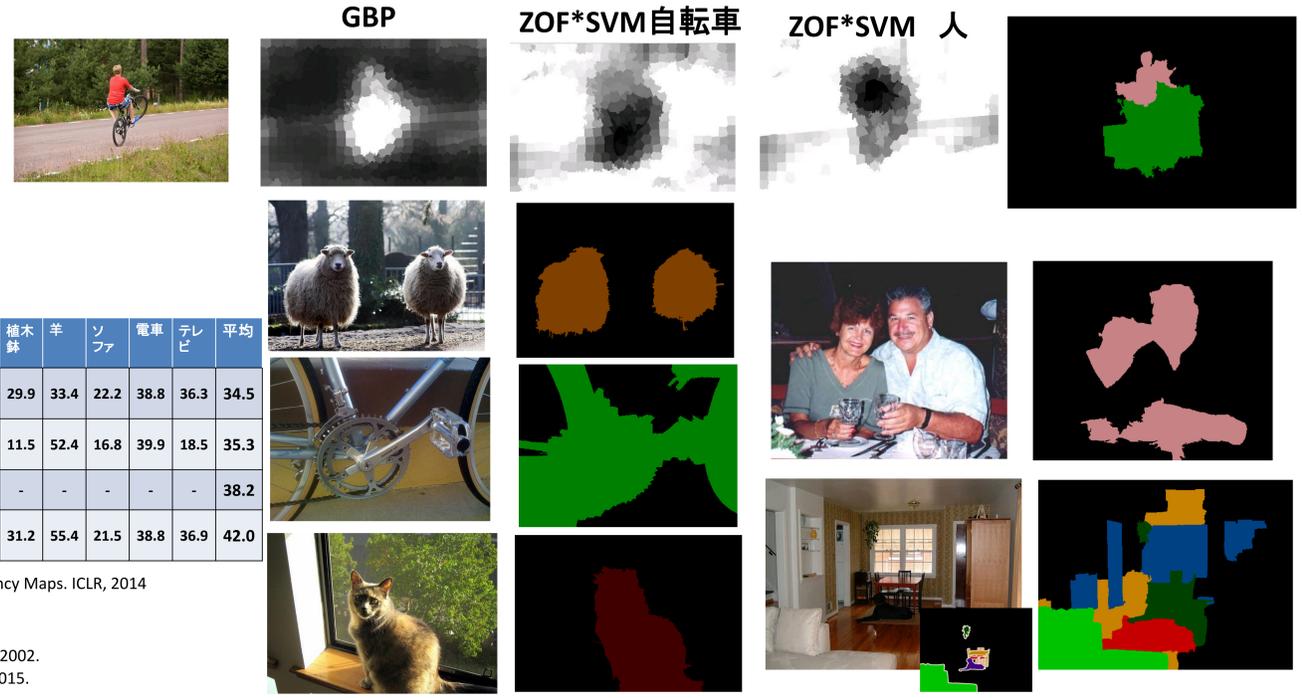
## 評価

Pascal VOC 2012 (20クラス+BG)  
 MIL-ILP [3] は70000枚の追加画像あり  
 mean IU:  $(1/n_d) \sum_i n_{ii} / (t_i + \sum_j n_{ji} - n_{ii})$   $t_i = \sum_j n_{ij}$

考察・室内の物体の領域分割が困難  
 ・特定のインスタンスに反応することも (顔の領域など)

	背景	飛行機	自転車	鳥	船	ボトル	バス	車	猫	椅子	牛	机	犬	馬	バイク	人	植木鉢	羊	ソファ	電車	テレビ	平均	
CCNN	65.9	23.8	17.6	22.8	19.4	36.2	47.3	46.9	47.0	16.3	36.1	22.2	43.2	33.7	44.9	39.8	29.9	33.4	22.2	38.8	36.3	34.5	
提案手法	63.4	40.8	16.1	44.6	15.2	26.5	46.9	29.9	64.3	8.0	45.5	14.3	54.8	41.5	49.9	40.6	11.5	52.4	16.8	39.9	18.5	35.3	
EM-Adapt	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	38.2
MIL-ILP[3]*	79.6	50.2	21.6	40.6	34.9	40.5	45.9	51.5	60.6	12.6	51.2	11.6	56.8	52.9	44.8	42.7	31.2	55.4	21.5	38.8	36.9	42.0	

## 領域分割結果



[1] K. Simonyan et al. Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps. ICLR, 2014  
 [2] Mohammadreza et al. Feedforward semantic segmentation with zoom-out features, CVPR, (2015)  
 [3] Pedro et al. From Image-level to Pixel-level Labeling with Convolutional Networks, CVPR, (2015)  
 [4] S. Andrews, I. Tschantzaris, T. Hofmann. Support Vector Machines for Multiple-Instance Learning. NIPS 2002.  
 [5] K. Simonyan, A. Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition, ICLR 2015.