Learning Deep Features for Discriminative Localization

CONTENTS

目录

Introduction

Introduction

- **FC替换成GAP**：减少参数，保存空间信息
- **Class Activation Mapping**：定位input image中的discriminative region
- CNN学到的generic localizable deep features可以应用于不同的任务中
- **结构固定：GAP与分类层直接相连**

引　　言

GAP结合CAP可以让CNN网络

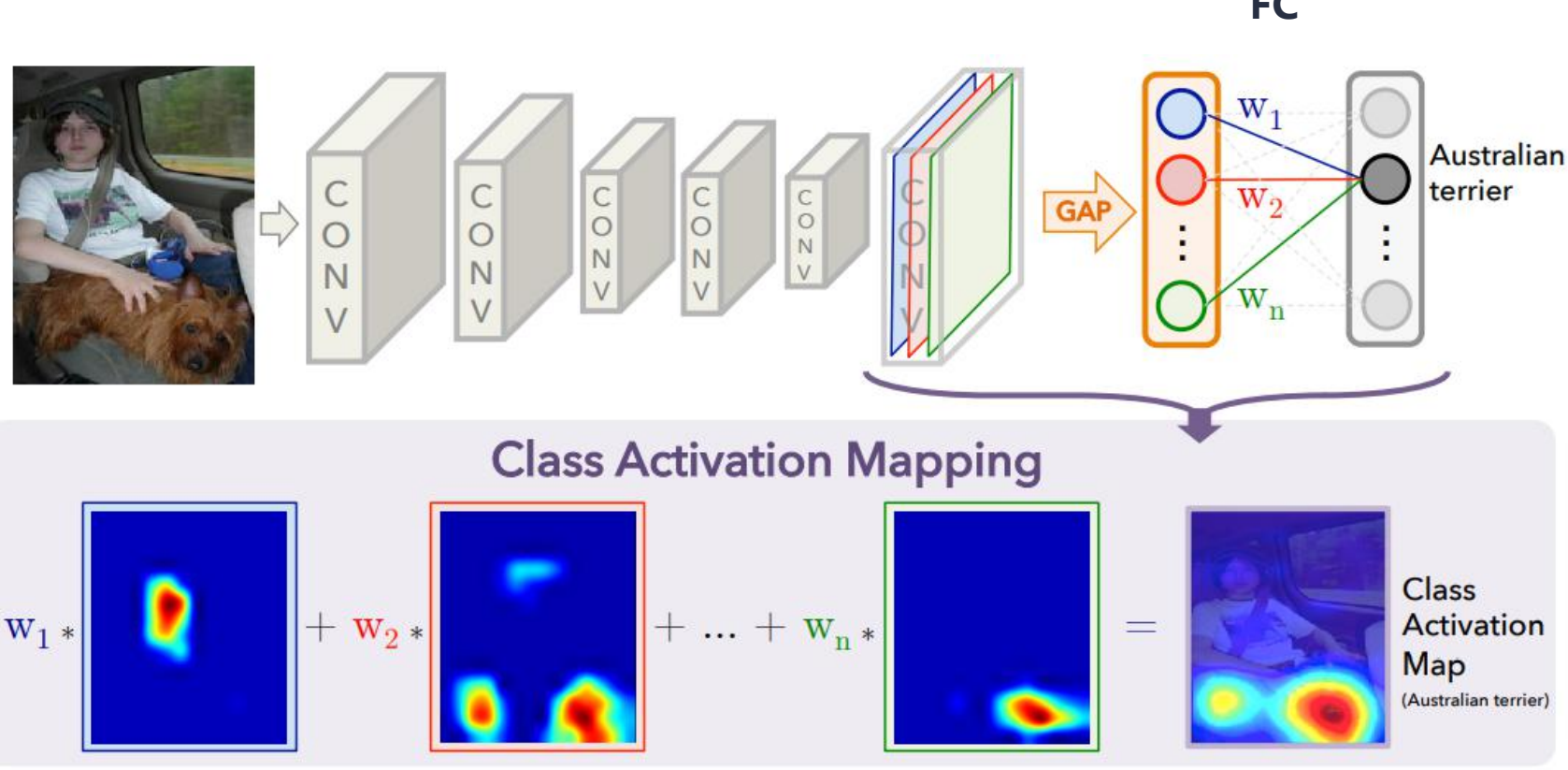
不仅进行图片分类，

而且可以定位图片中判别区域

（只通过一次前向传播）

C A M

C A M


FC
CONV
CONV
CONV
CONV
CONV
CONV
GAP
w_1
w_2
w_n
Australian terrier
Class Activation Mapping
$w_1 *$
+ $w_2 *$
+ ... + $w_n *$
=
Class Activation Map
(Australian terrier)

C A M

$$F_k = \sum_{x,y} f_k(x,y)$$

$$S_c = \sum_k w_k^c \sum_{x,y} f_k(x,y) = \sum_{x,y} \sum_k w_k^c f_k(x,y).$$

class confidence score

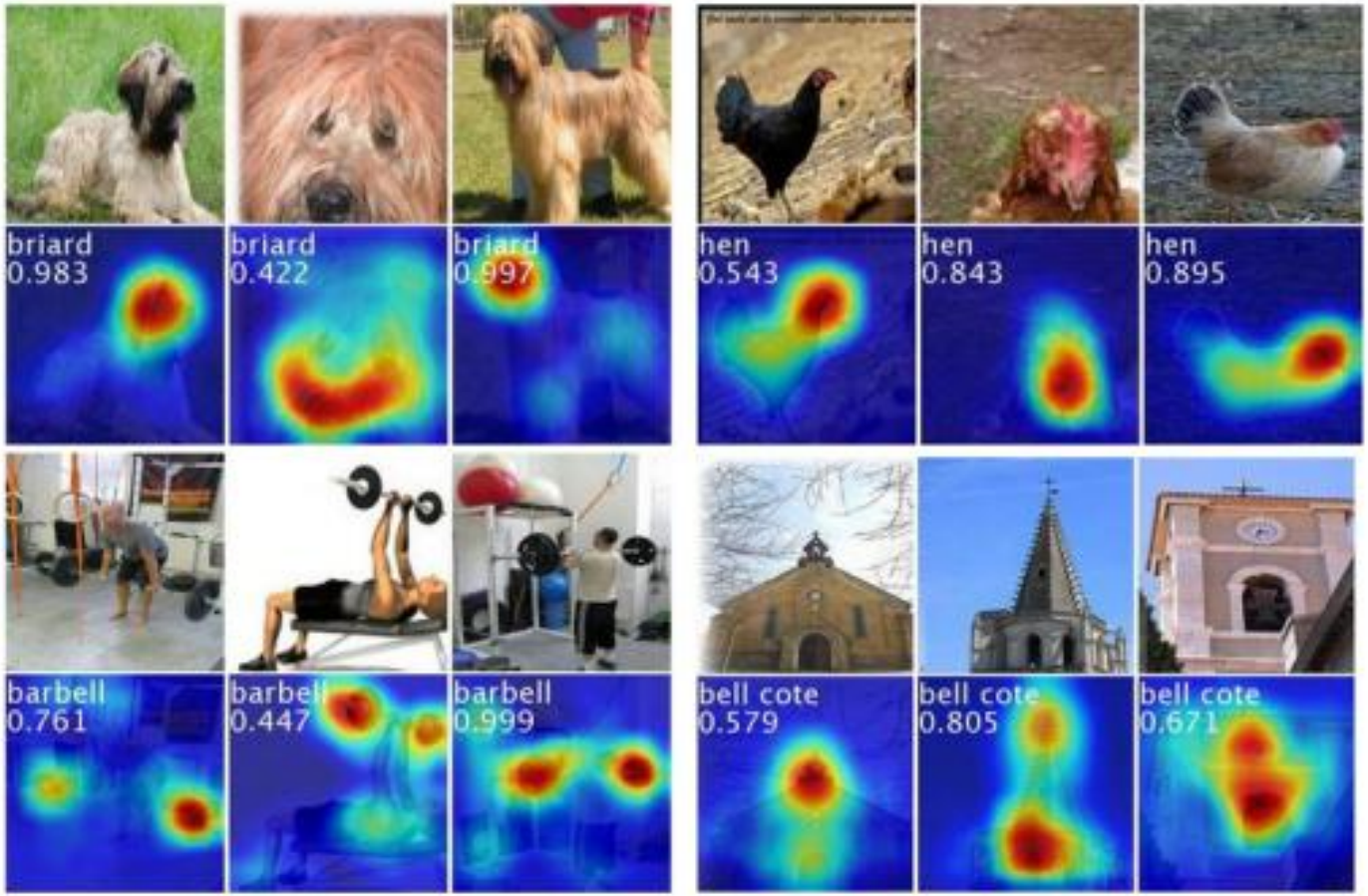
$$M_c(x,y) = \sum_k w_k^c f_k(x,y).$$

CAM

$$S_c = \sum_{x,y} M_c(x,y)$$

Mc(x,y):让图片分为c类时激活（x,y)位置上的值的重要性

C A M


briard
0.983
briard
0.422
briard
0.997
hen
0.543
hen
0.843
hen
0.895
barbell
0.761
barbell
0.447
barbell
0.999
bell cote
0.579
bell cote
0.805
bell cote
0.671

3 Application

Weakly-supervised Object Localization

◆**setup**

1、AlexNet , VGGnet , and GoogLeNet（使用GAP替代FC）

2、GAP有更高的空间分辨率——mapping resolution

3、对不同模型移除（增加）不同的卷积层

Weakly-supervised Object Localization

◆ **classify :**

original AlexNet , VGGnet , and GoogLeNet ,andNetwork in Network

◆**localize :**

original GoogLeNet3，NIN and using backpropagation

◆**GAP vs GMP :**

GoogLeNet-GAP，GoogLeNet-GMP

Weakly-supervised Object Localization

Table 1. Classification error on the ILSVRC validation set.

Networks	top-1 val. error	top-5 val. error
VGGnet-GAP	33.4	12.2
GoogLeNet-GAP	35.0	13.2
AlexNet*-GAP	44.9	20.9
AlexNet-GAP	51.1	26.3
GoogLeNet	31.9	11.3
VGGnet	31.2	11.4
AlexNet	42.6	19.5
NIN	41.9	19.6
GoogLeNet-GMP	35.6	13.9

◆**classification results：**

表现下降了大约1-2%，观察到AlexNet对移除FC最敏感；总的来说，移除FC对分类性能影响不大

Weakly-supervised Object Localization

Table 2. Localization error on the ILSVRC validation set. *Back-prop* refers to using [22] for localization instead of CAM.

Method	top-1 val.error	top-5 val. error
GoogLeNet-GAP	**56.40**	**43.00**
VGGnet-GAP	57.20	45.14
GoogLeNet	60.09	49.34
AlexNet*-GAP	63.75	49.53
AlexNet-GAP	67.19	52.16
NIN	65.47	54.19
Backprop on GoogLeNet	61.31	50.55
Backprop on VGGnet	61.12	51.46
Backprop on AlexNet	65.17	52.64
GoogLeNet-GMP	57.78	45.26

◆**localization results**：在CAM中生成bounding box——阈值技术；**与Backprop比较**

Weakly-supervised Object Localization

Table 3. Localization error on the ILSVRC test set for various weakly- and fully- supervised methods.

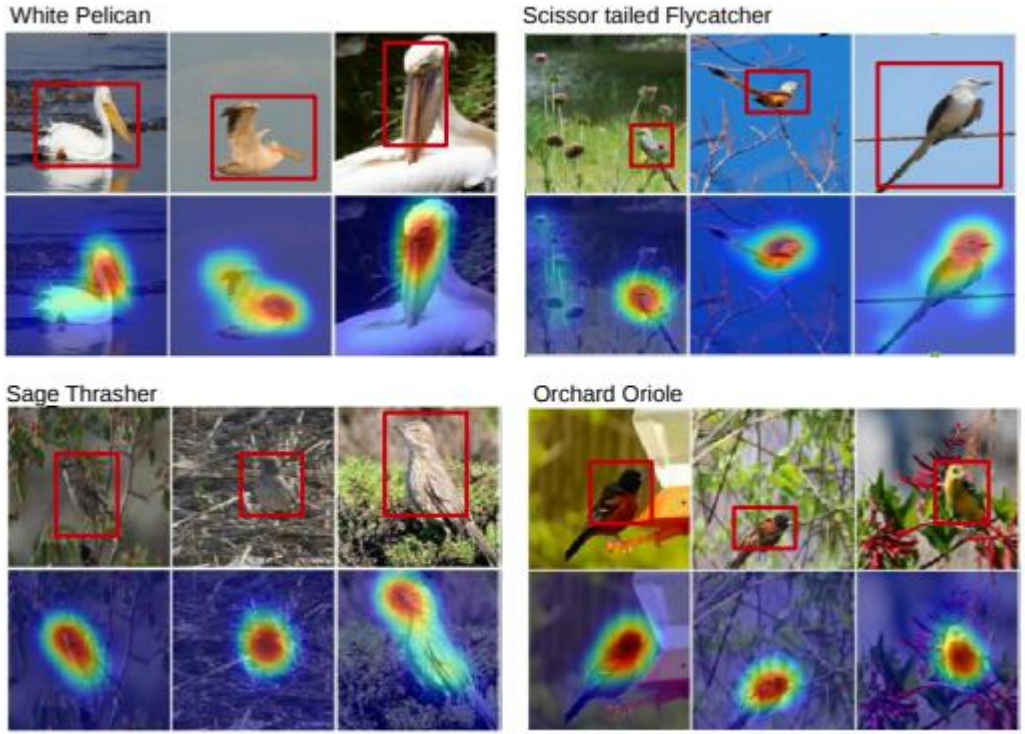
Method	supervision	top-5 test error
GoogLeNet-GAP (heuristics)	weakly	**37.1**
GoogLeNet-GAP	weakly	42.9
Backprop [22]	weakly	46.4
GoogLeNet [24]	full	26.7
OverFeat [21]	full	29.9
AlexNet [24]	full	34.2

弱监督：标签为整张图片的类别，不包括BB的定位标签

◆localization results：

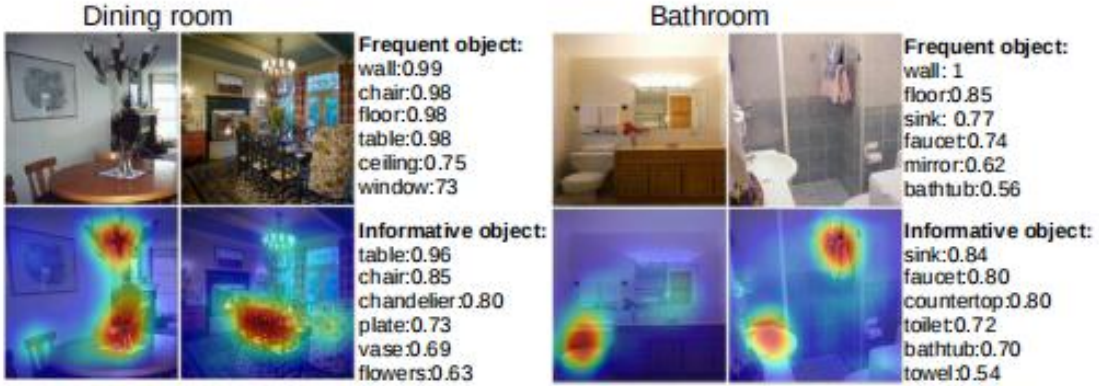
weakly vs full

Deep Features for Generic Localization



- **Fine-grained Recognition**：从判别性区域（CAM BBox）中提取特征进行分类会提升模型表现

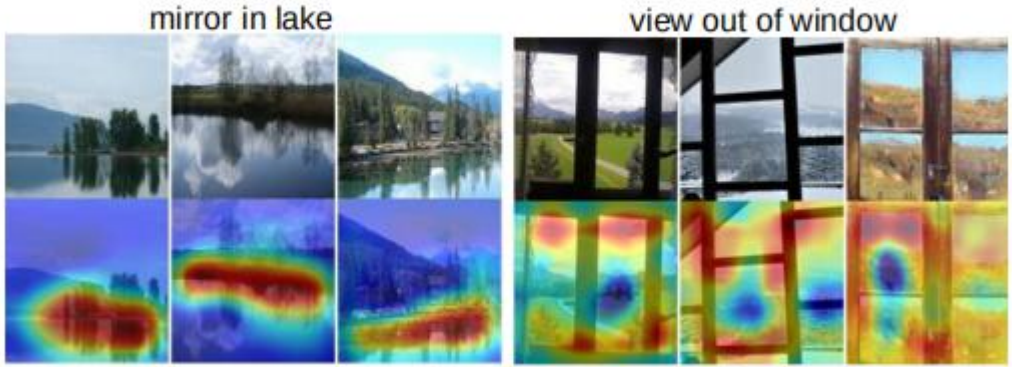
Deep Features for Generic Localization



高激活区域与特定场景的标志性物体一致

- **Discovering informative objects in the scenes**：对于每个场景种类训练了一个 SVM 并且使用线性SVM的权重计算 CAMs；

Deep Features for Generic Localization



- **Concept localization in weakly labeled images :**

  正样本：图片+概念短语；

  负样本：图片

  使用hard-negative mining学习到concept detectors，并用CAM技术定位图片中的概念；

Deep Features for Generic Localization

◆**Weakly supervised text detector：**

正样本：图片+文本；

负样本：图片（只包括风景）

没有BB标注训练的情况下可以准确标注出文本

Deep Features for Generic Localization

◆Interpreting visual question answering：

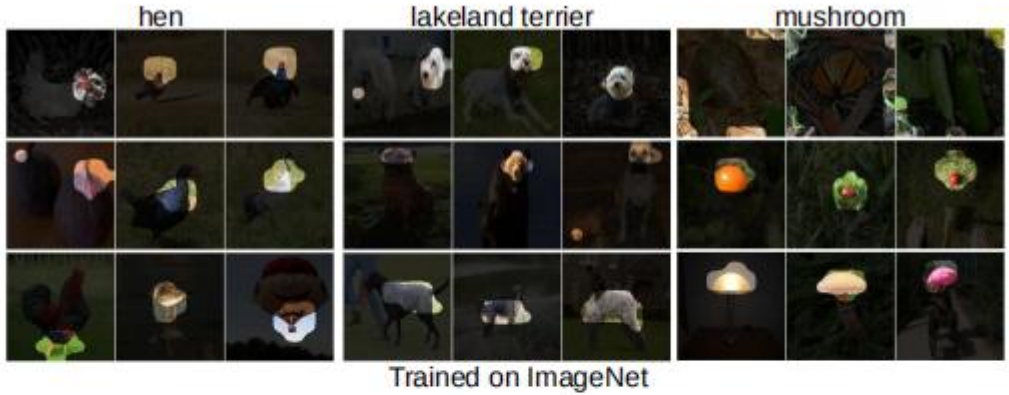
准确标注出与预测回答相关的图像区域

Visualizing Class-Specific Units

- 即可视化哪一些channel的**feature map**（final conv）对给定类别最具有判别性
- 使用GAP与softmax之间生成的权重对feature map进行**排序**
- 观察出物体的哪些部分最具有**判别性**以及是哪些feature map探测到了这些部分
- 具有判别性的feature map的**组合**引导CNN进行分类

Visualizing Class-Specific Units

object



scene

TOP-3 units

conclusion

总　　结

◆提出了一种基于普通分类CNN的GAP改良，并提出CAM技术使得定位任务可以**融合**进普通的分类任务中。

◆对于只有分类标签而**无定位标签**的数据集，网络的训练不仅能保持原基础网络的分类性能，还能得到分类物体在原图的定位。（弱监督）

◆其中CAM可以**可视化**预测类在任何给定图片上的**得分**，并标出CNN检测到的物体的判别性区域。

THANKS!

LIVE AND LEARN