Image specific class saliency map

Soumyadeep Banik, CS-II
RKMVERI, Belur
Term Project, 2021

Problem Statement

Given an Image I, and the corresponding class model learnt by image-classification ConvNet.

Our goal is to find the set of pixels which are taking more human visual attention.

Saliency Map

- What is saliency map?
- Some real world applications
- Method in this paper
- Quadtree based approach to improve complexity

Saliency Map

- What is saliency map?
- Some real world applications
- Method in this paper

  “Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps”

- Quadtree based approach to improve complexity

Saliency Map

- What is saliency map?

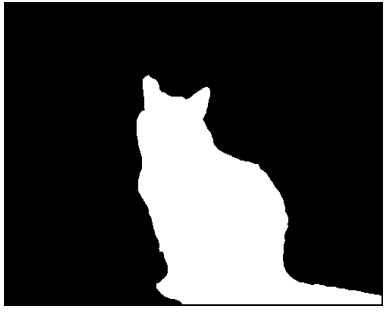
- The most important/salient set of pixels in an image.

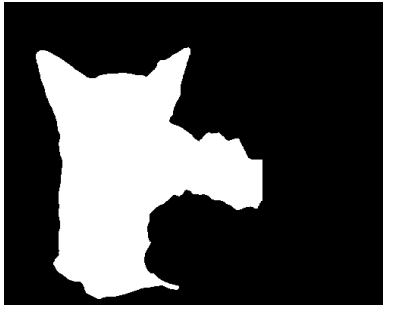
- Saliency map can be viewed as a Image segmentation problem, where each pixel has a unique influnce in the image.

Saliency Map







PASCAL-S dataset

Real world applications

- Medical Imaging
- Image cropping
- Image Captioning
- Robot active vision
- Region-of-interest extraction

Methodology

- We have score S_c of a class c for an image I. The goal is to find a Image where the score is maximized.

$$\arg\max_I S_c(I) - \lambda \|I\|_2^2$$

- Probability of each class score => $P_c = \frac{\exp S_c}{\sum_c \exp S_c}$

Maximum prob. can be achieved by minimizig the scores of other classes. So, by optimizing the class score S_c we can focus more on the image specific classes .

Methodology

- So the class score can be written as $S_c(I) = w_c^T I + b_c$

where w is the derivative of S_c with respect to the image I at the point (image) I_0

$$w = \left.\frac{\partial S_c}{\partial I}\right|_{I_0}$$

- We take the maximum magnitude of w across all colour channels

$$M_{ij} = \max_c |w_{h(i,j,c)}|$$

Methodology

- So the class score can be written as $S_c(I) \approx w^T I + b$

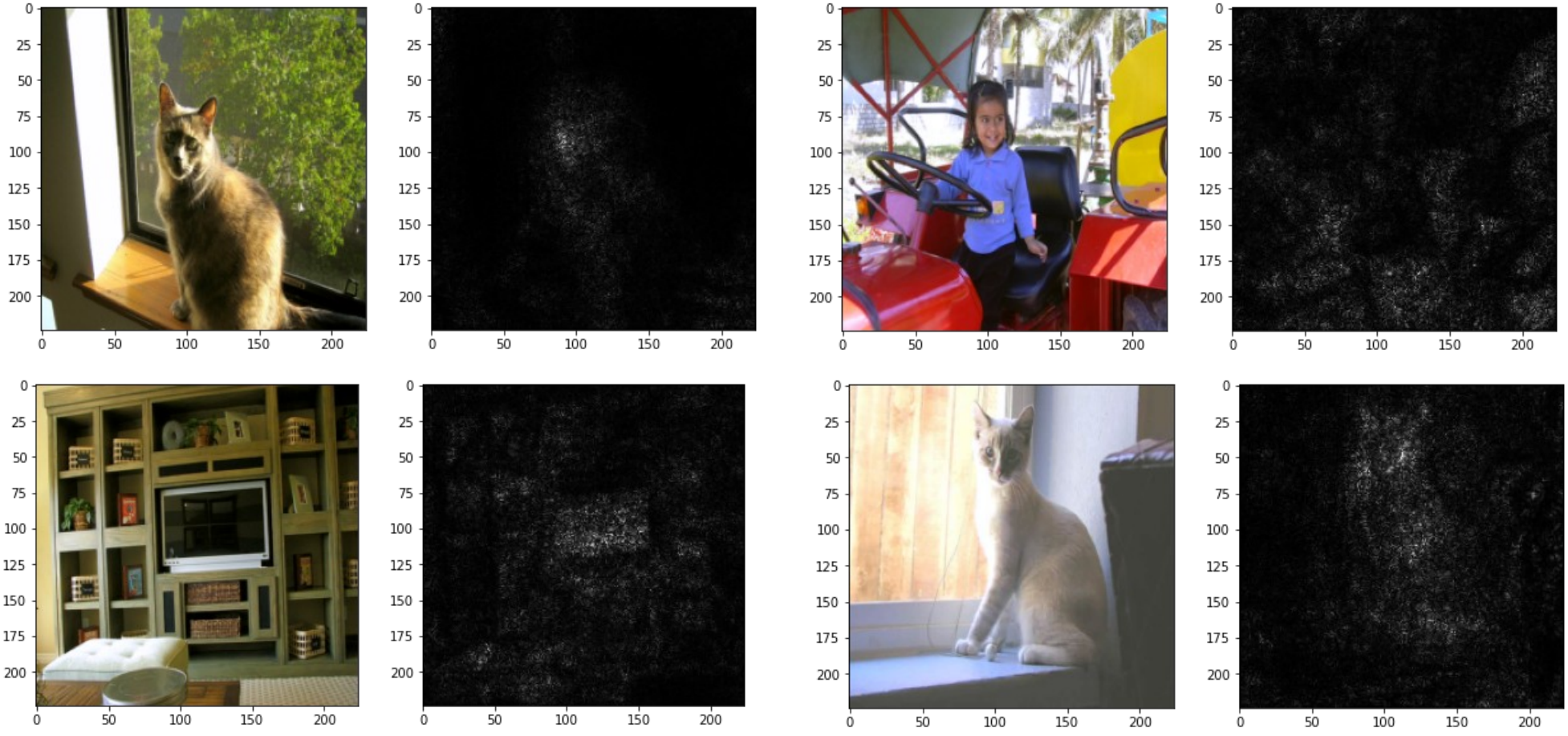
where w is the derivative of S_c with respect to the image I at the point (image) I_0

$$w = \left. \frac{\partial S_c}{\partial I} \right|_{I_0}$$

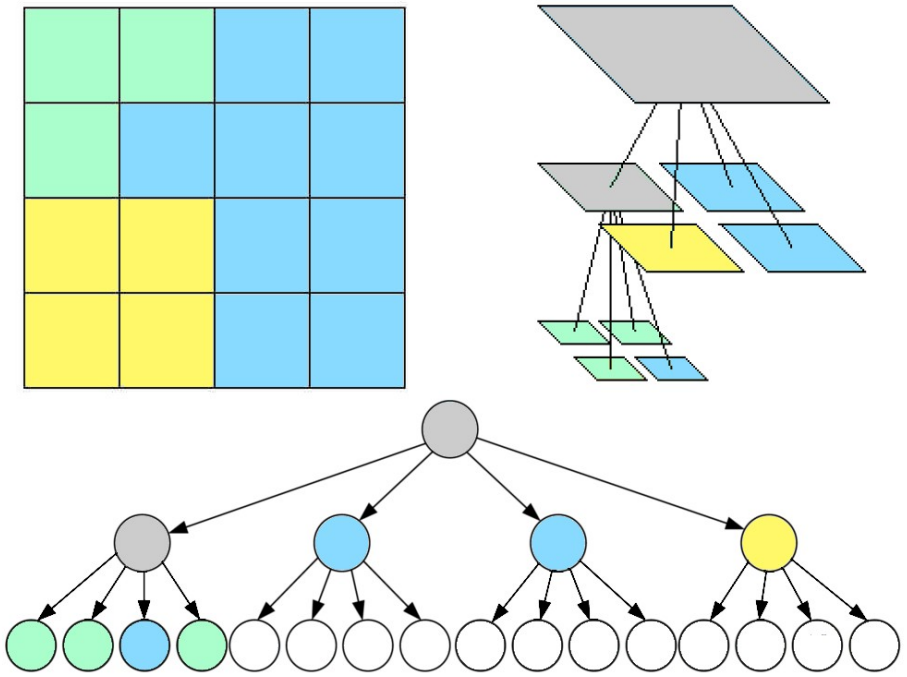
- We take the maximum magnitude of w across all colour channels

$$M_{ij} = \max_c |w_{h(i,j,c)}|$$

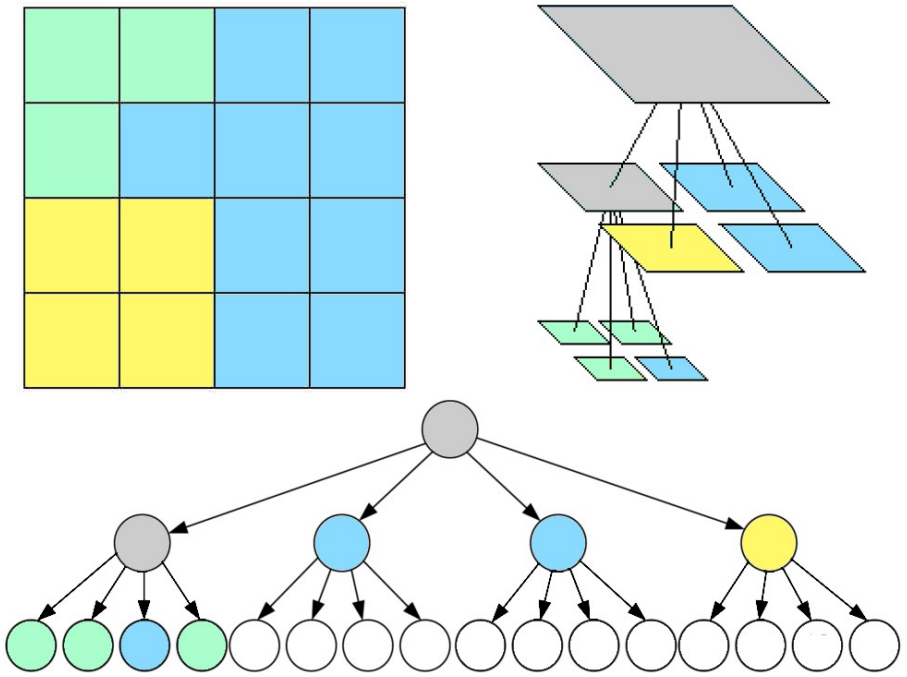
Results

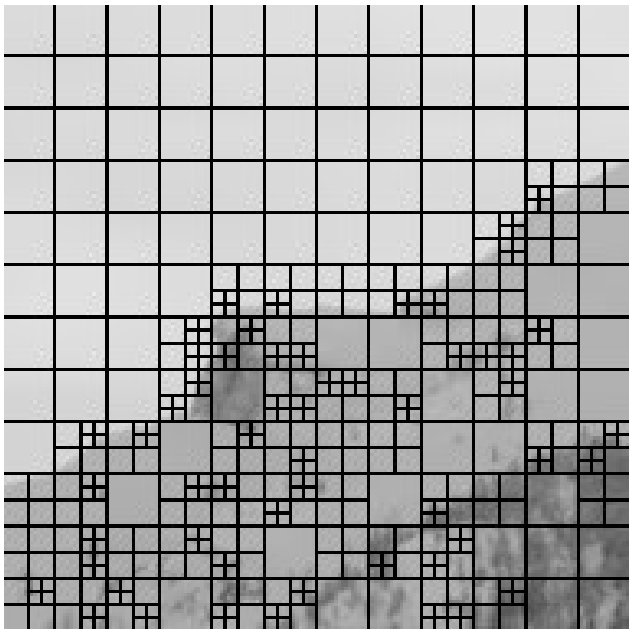


Quadtree based approach

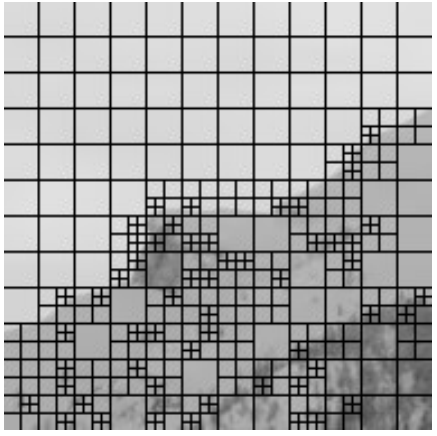


Quadtree based approach

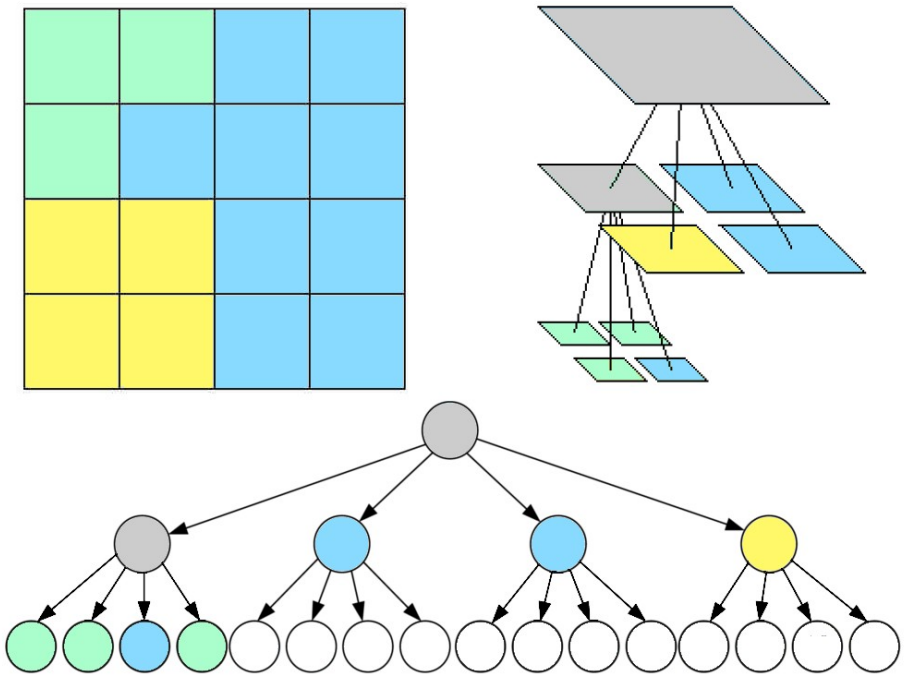




Quadtree based approach



Quadtree based approach



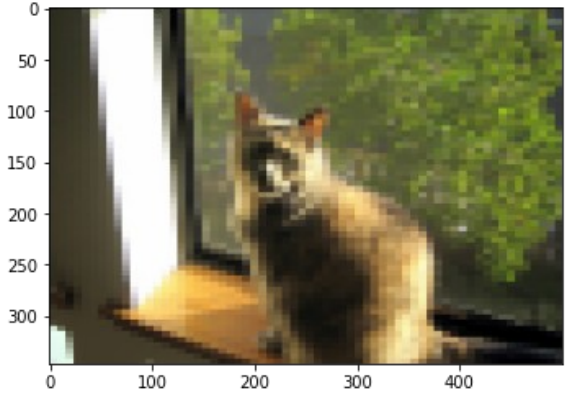
In case of CNN:

Complexity: $M^2 N^2 C$
for NxN image
convolving with MxM kernel

For Quadtree:

Complexity: $M^2 N' C$

Quadtree based approach



In case of CNN:

Complexity: $M^2 N^2 C$
for NxN image
convolving with MxM kernel

For Quadtree:

Complexity: $M^2 N' C$

References:

I. Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps: Karen Simonyan Andrea Vedaldi Andrew Zisserman Visual Geometry Group, University of Oxford

II. Quadtree Generating Networks: Efficient Hierarchical Scene Parsing with Sparse Convolutions. Kashyap Chitta Jose M. Alvarez Martial Hebert

III. REPRESENTING IMAGES USING THE QUADTREE DATA STRUCTURE, ISRAEL SHUVAL, Bachelor of Science. Oklahoma City University, Oklahoma, 1986