Visual saliency: Methods of Identification and Applications

Group 7: 施長佑 管彥鳴 楊大昕 黃乾哲 葉詠富

Outline

- Introduction
  - What is Visual Saliency
  - Application
    - Object Detection
    - Robotic Vision
- Method
  - Static Salieny
  - Motion Saliency
  - Deconvolution
  - CAM
- Conclusion

Introduction

Introduction

- Visual saliency is the distinct subjective perceptual quality.
- It can make some items in the world stand out from their neighbors and immediately grab our attention.
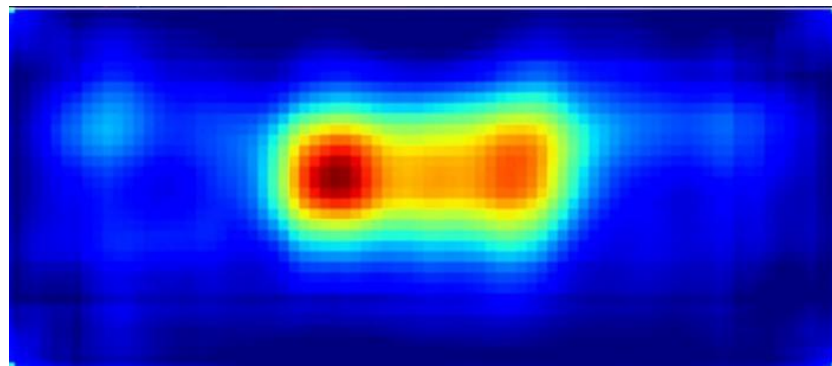
Why do we need saliency?

- It is important to rapidly detect potential prey, predators, or mates in a cluttered visual world.
- However, simultaneously identifying any and all interesting targets in one's visual field has prohibitive computational complexity.
- One solution is to restrict complex object recognition process to a small area or a few objects at any one time.
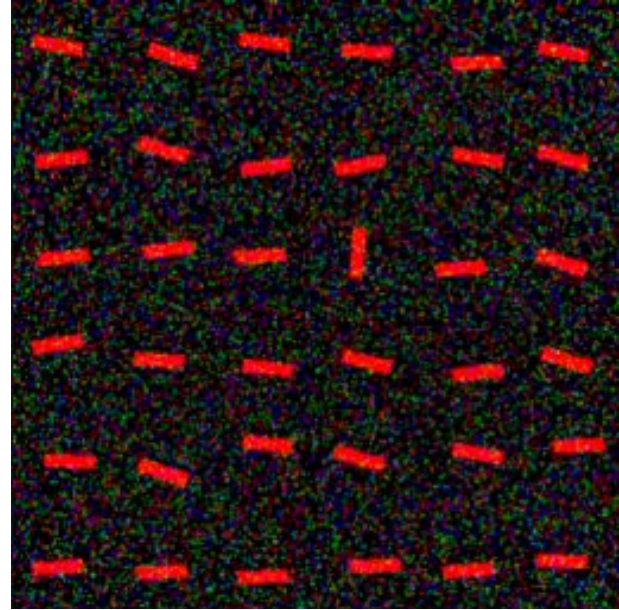
Example 1 - Color

- One item in the array of items strongly pops-out and effortlessly and immediately attracts attention.
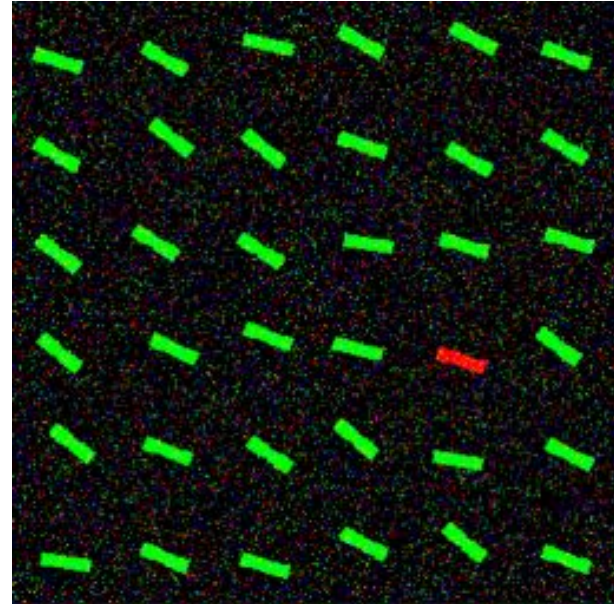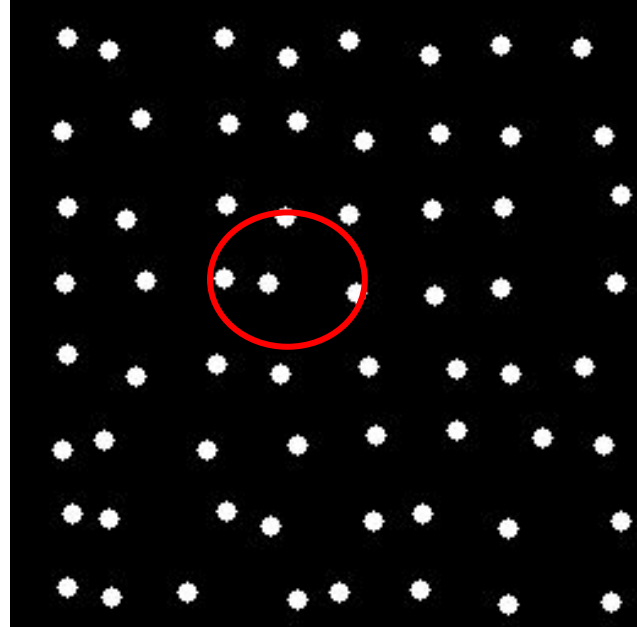
Example 2 - Direction

- In this display, the vertical bar is visually salient.

fig.from http://www.scholarpedia.org/article/Visual_salience#External_Links
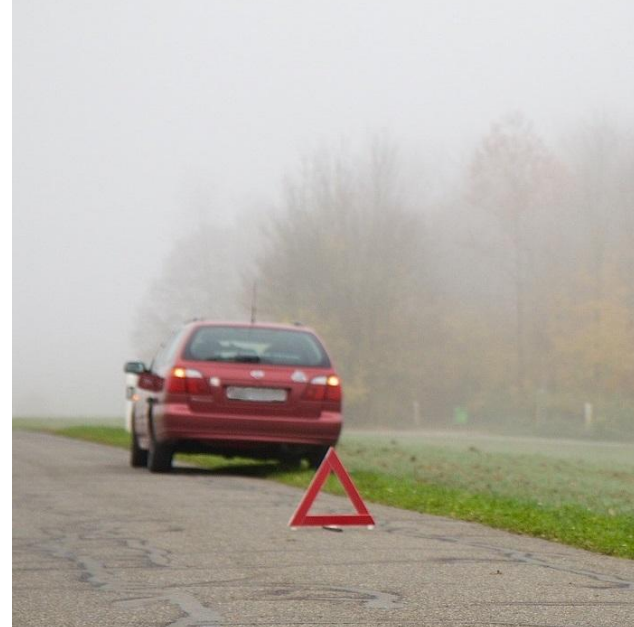
Example 3 - Speed

- Here is an example where a distinct pattern of motion is the only thing which distinguishes the salient dot from its neighboring distractor dots.

Example 4

- In natural environments, highly salient objects tend to automatically draw attention towards them.

Neural and computational mechanisms

● This significant difference could be in a number of simple visual feature dimensions which are believed to be represented in the early stages of cortical visual processing: edge orientation、motion direction or color.
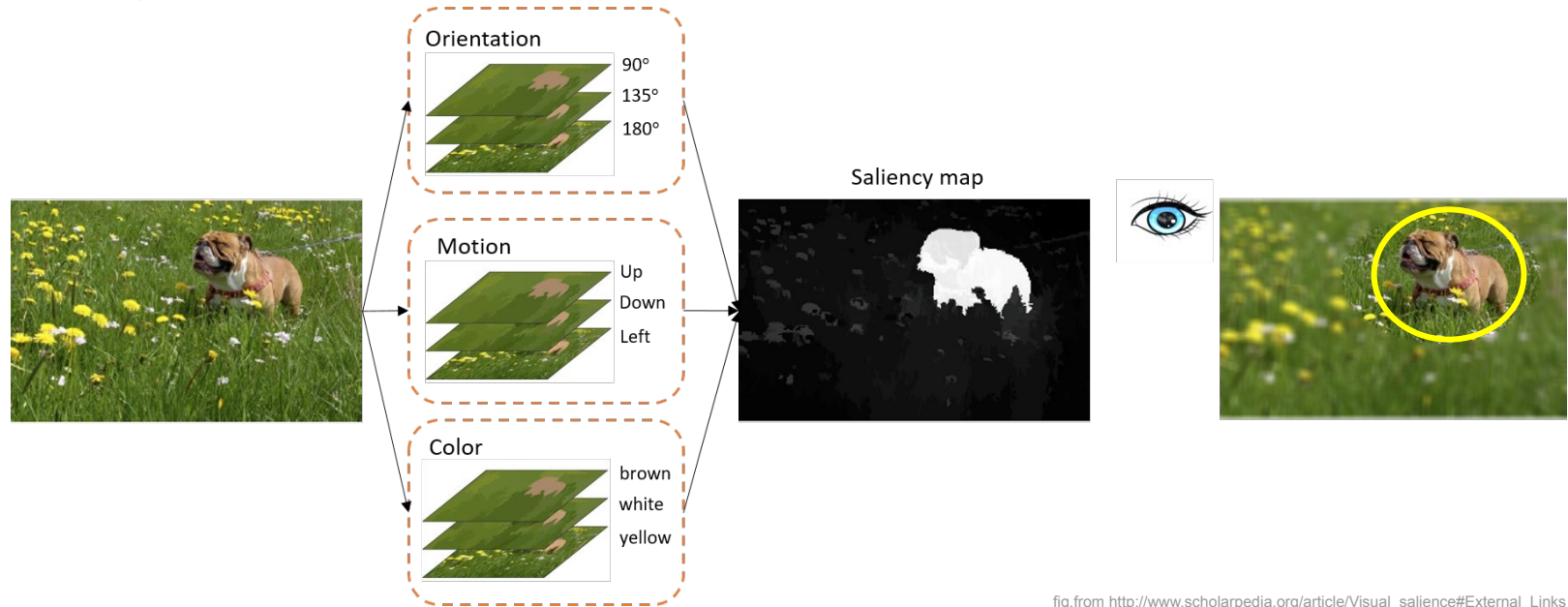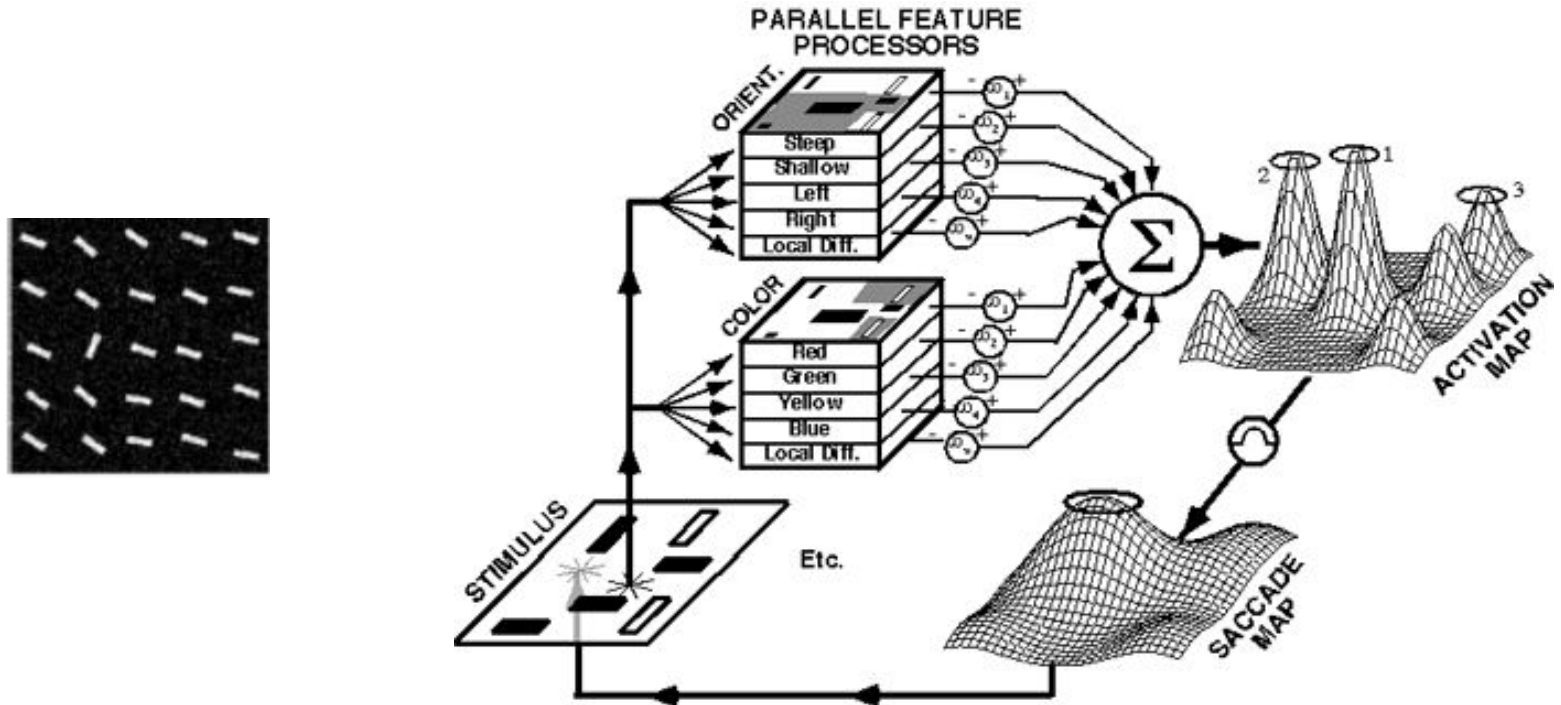
Feature maps

- Different observational aspects of a picture will have different degrees of saliency.
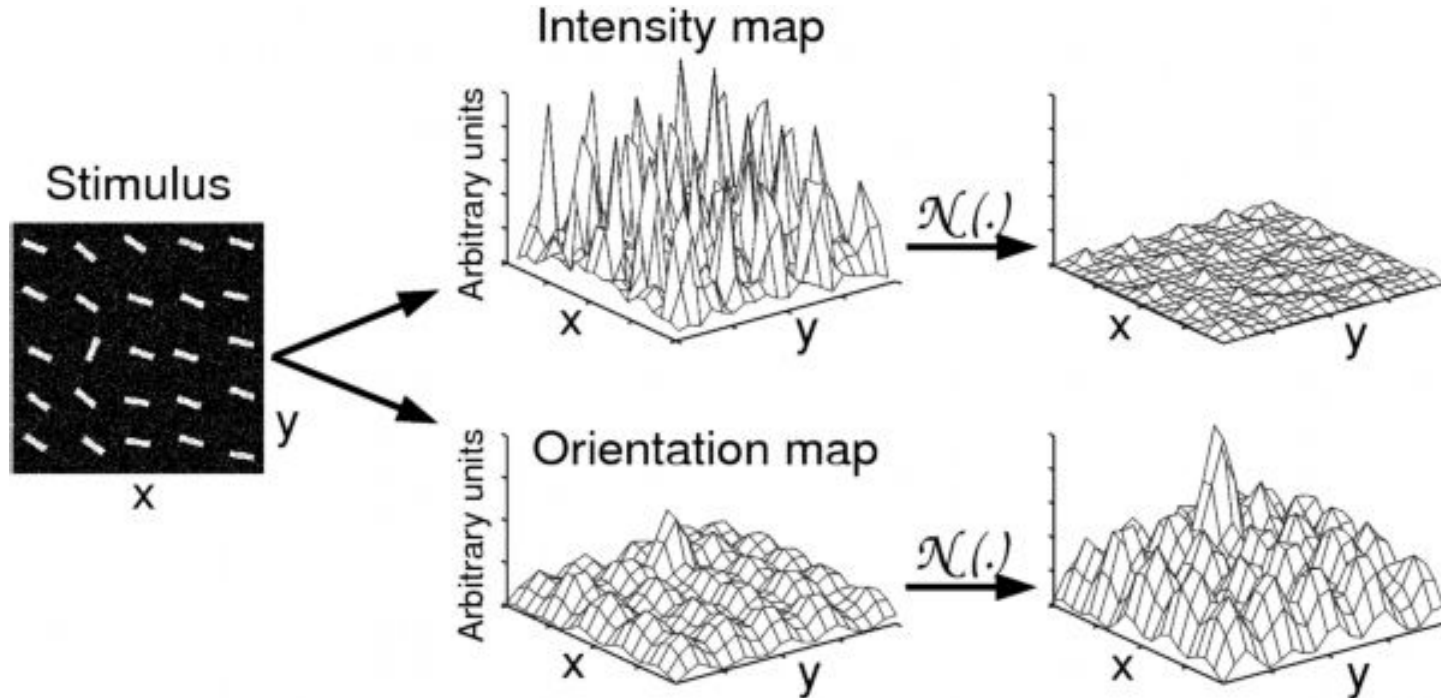
Top-down modulation by task demands

- The more Top, the more important.

Applications

- Automatic target detection

- Salient objects in the environment as navigation landmarks.

Applications

- Finding tumors in mammograms.

- Automatic cropping/centering of images for display on small portable screens
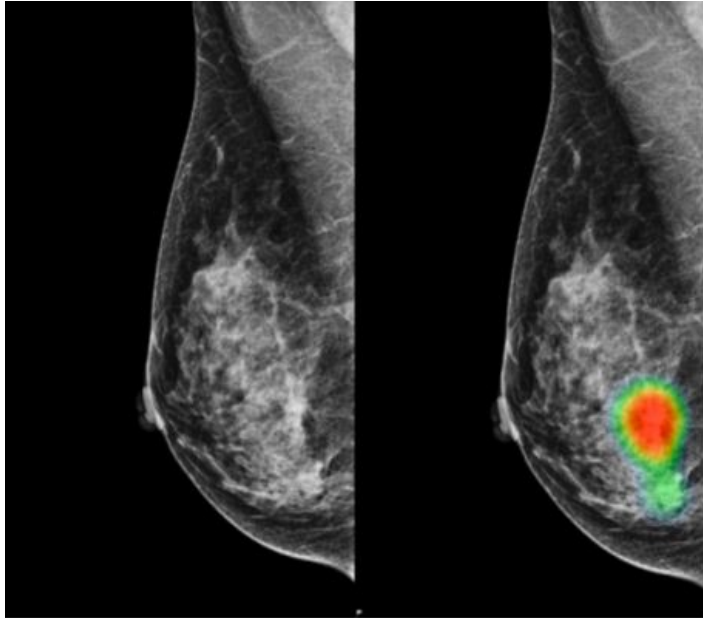
Saliency and CNN

- Why does the neural network predict a particular label? What did it see in the image that led to a particular prediction? We can answer these questions through saliency maps.

Convolution

Pooling

Flatten
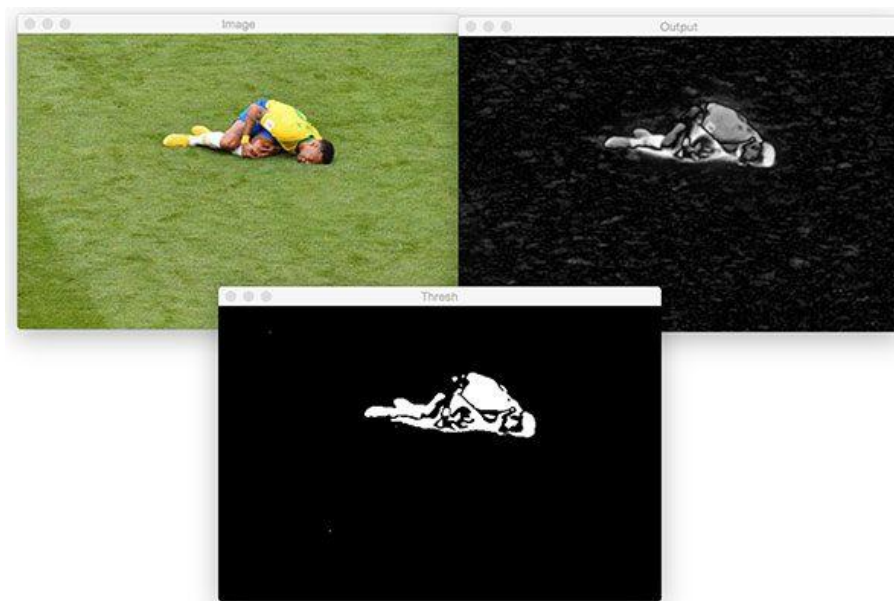
How do?

Dog

Cat

Convolution
Layer

Pooling
Layer
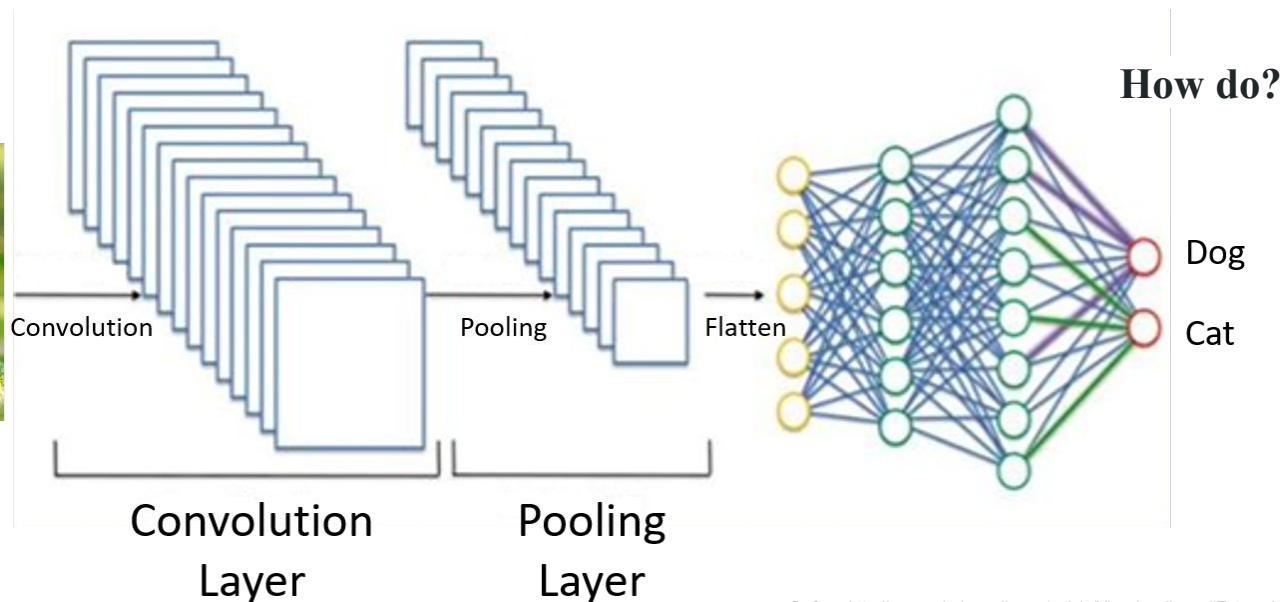
Method

Static Saliency

- This class of saliency detection algorithms relies on image features and statistics to localize the most interesting regions of an image.

Saliency Detection By Gaussian Pyramid

- We have an image and the basic features like colour, orientation, the intensity is extracted from the image.
- These processed images are used to create Gaussian pyramids to create features Map.
- Saliency map is created by taking the mean of all the feature maps.
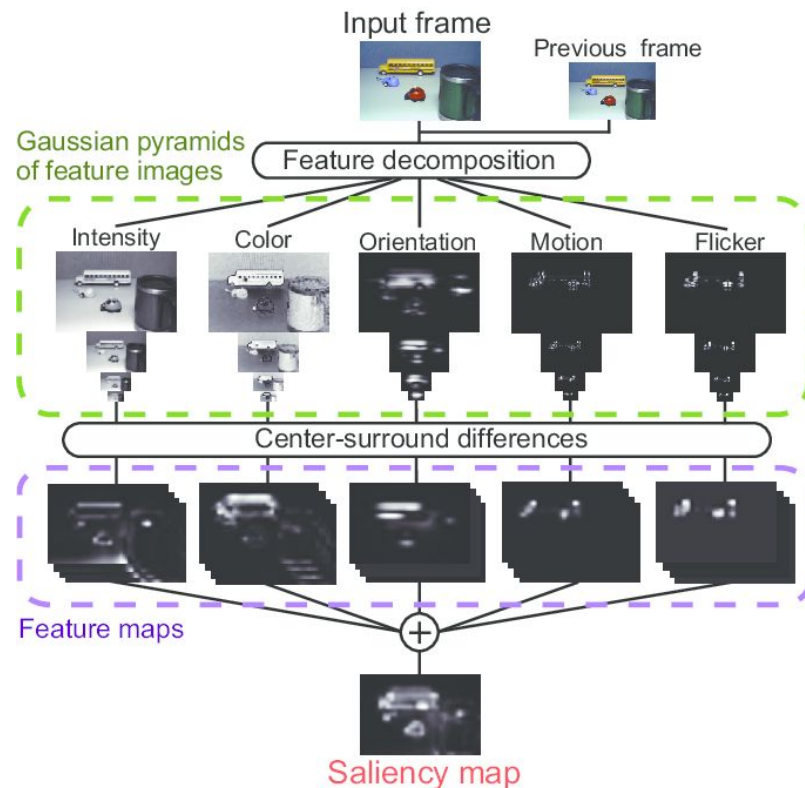
Motion Saliency

- Algorithms in this class typically rely on video or frame-by-frame inputs.
- The motion saliency algorithms process the frames, keeping track of objects that "move".
- Objects that move are considered salient.

Deconvolutional Network

- Deconvolution uses an unsupervised method to find a set of kernels and feature maps, and let them reconstruct the image.
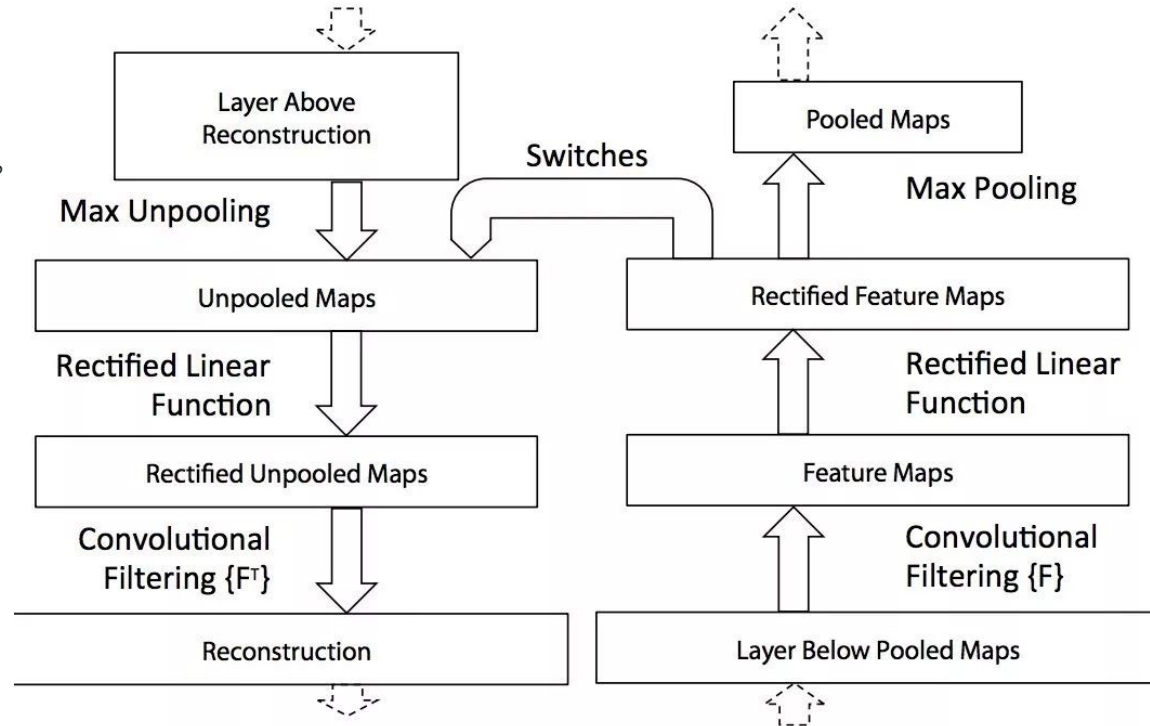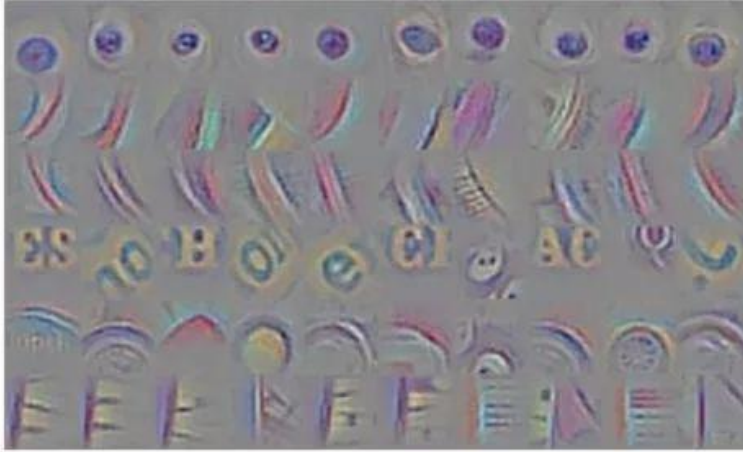
Deconvolutional Network - Guided Backpropagation

- Guided Backpropagation is one of the deconvolution methods.

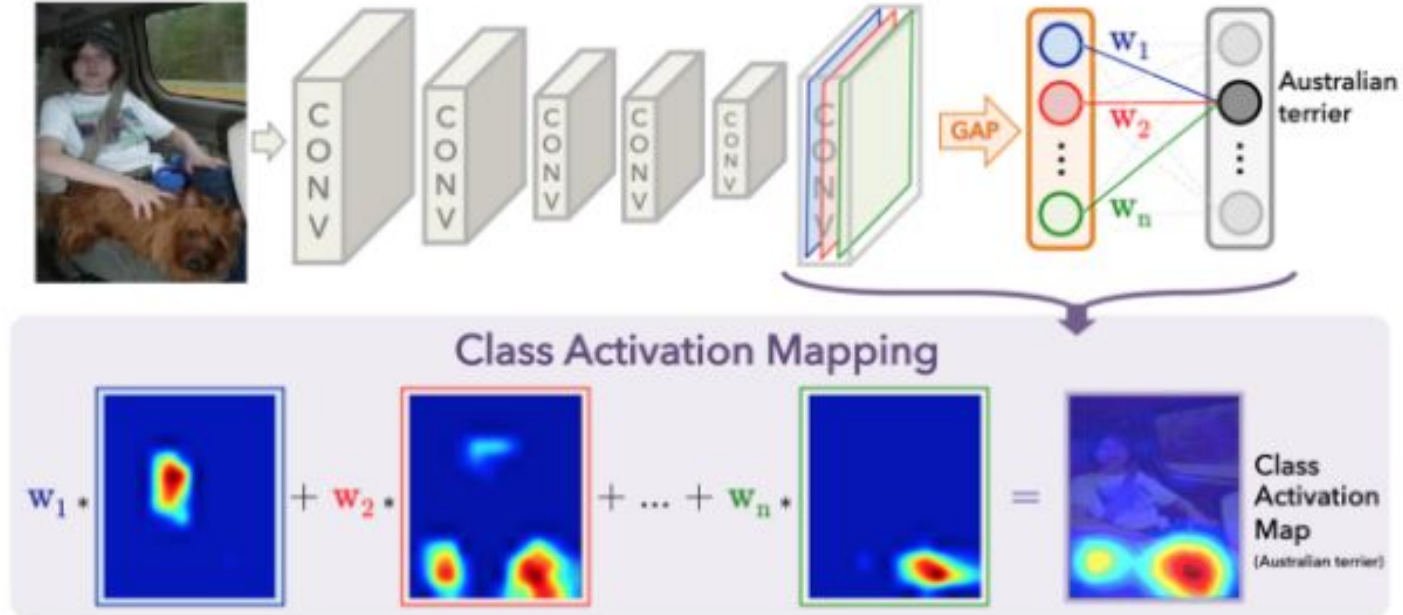

deconv   guided backpropagation   corresponding image crops

Deconvolutional Network - Class Activation Mapping

- The pixel array after GAP will be multiplied by the weight w. The larger the value of the weight w, the greater the influence of the image represented by the pixel.

Class Activation Mapping

$$w_1 * \quad + \quad w_2 * \quad + \ldots + \quad w_n * \quad = \quad$$

Class Activation Map (Australian terrier)

CAM Applications

- CAM can show the parts that CNN cares about when classifying cat and dog.

| Original Image | Guided Backprop 'Cat' | CAM 'Cat' |
| Original Image | Guided Backprop 'Dog' | CAM 'Dog' |
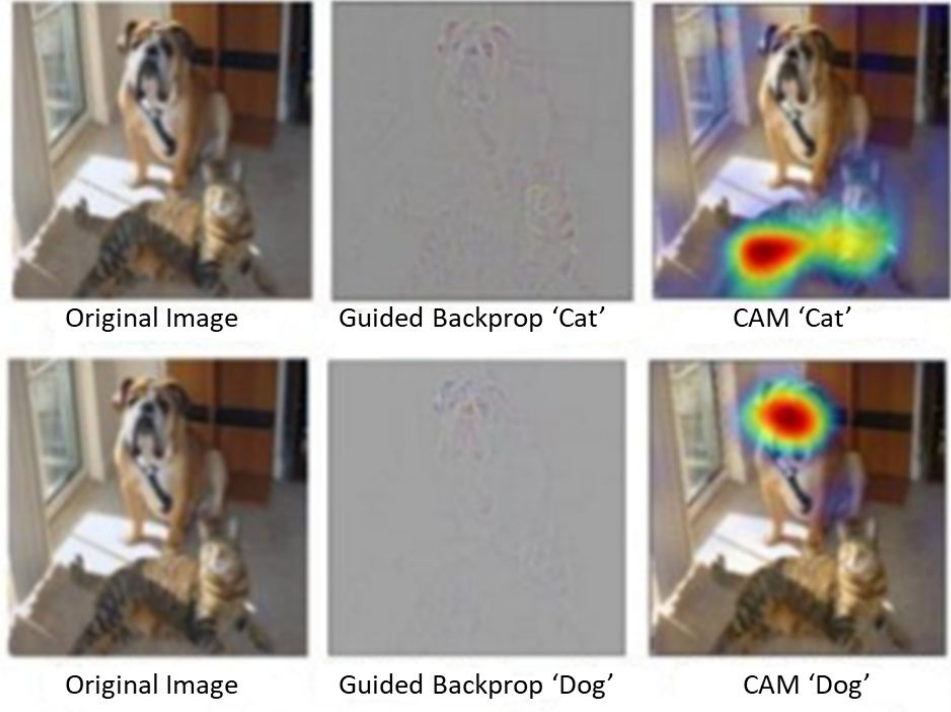
Conclusion

- It is important to rapidly detect potential prey, predators, or mates in a cluttered visual world.
- However, simultaneously identifying any and all interesting targets in one's visual field has prohibitive computational complexity.
- One solution is to restrict complex object recognition process to a small area or a few objects at any one time.

Thank You for Listening