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

Karen Simonyan, Andrea Vedaldi, Andrew Zisserman

Presenter: Ankur Garg

Motivation: Do CNNs understand aspects of Image?

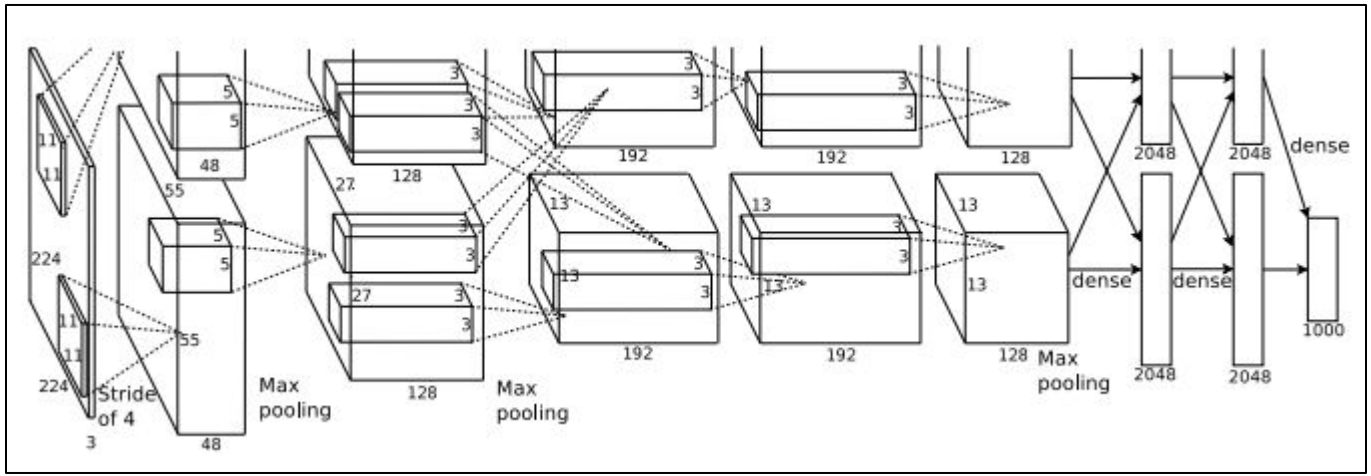


Image Source: Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems 2012 (pp. 1097-1105).

Contributions

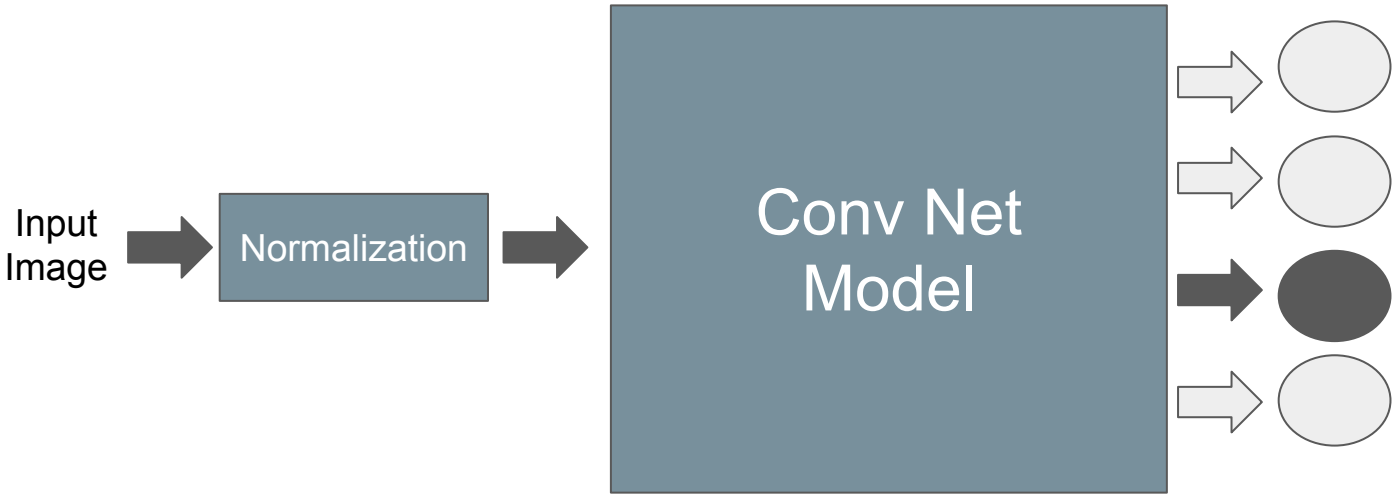
1. Obtain Understandable Visualisations of Conv Nets.
2. Image-specific class saliency map
    a. Can be used weakly supervised object localization

Contributions

1. Obtain Understandable Visualisations of Conv Nets.
2. Image-specific class saliency map
    a. Can be used weakly supervised object localization

Class Model Visualization

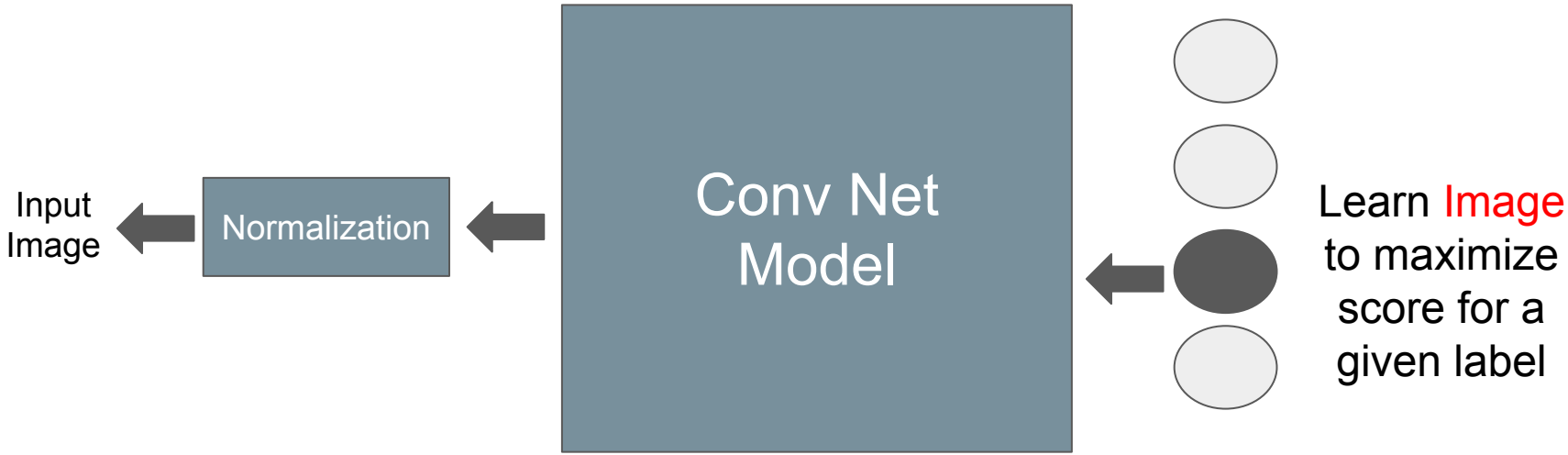
$$\underset{W}{arg\,max}\ S_C(W)$$



Learn weights to maximize score for true label

Class Model Visualization

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



Results: Class Model Visualization

- Key aspects of images in a class are captured

Bell Pepper

Dumbbell

Image Credit: Simonyan K, Vedaldi A, Zisserman A. Deep inside convolutional networks: Visualising image classification models and saliency maps. arXiv preprint arXiv:1312.6034. 2013 Dec 20.

Contributions

1. Obtain Understandable Visualisations of Conv Nets.
2. Image-specific class saliency map
    a. Can be used weakly supervised object localization

Image-Specific Class Saliency Visualisation

Linear Score Model

$$S_C = w_C^T I + b_C$$

Taylor approximation of Classification Function

$$S_C \approx w^T I + b$$

Find ***w*** through **single backpropagation** step

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

Image Credit: Simonyan K, Vedaldi A, Zisserman A. Deep inside convolutional networks: Visualising image classification models and saliency maps. arXiv preprint arXiv:1312.6034. 2013 Dec 20.

Results: Image-Specific Class Saliency Visualisation

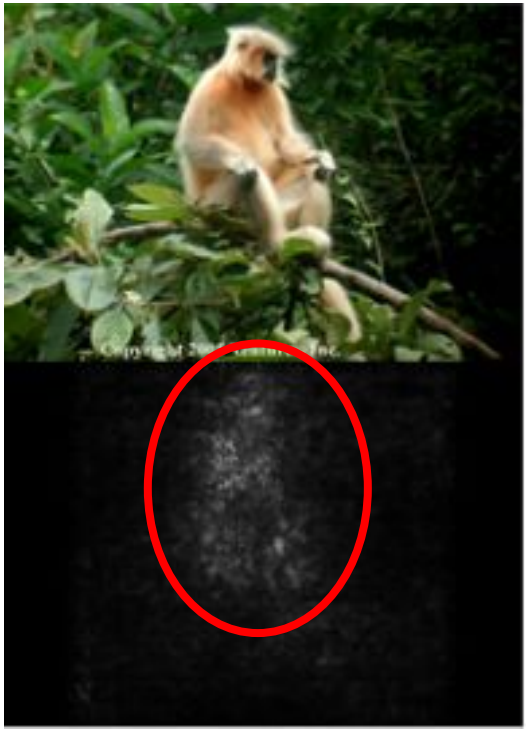


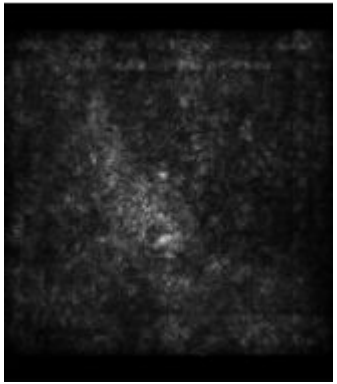
Image Credit: Simonyan K, Vedaldi A, Zisserman A. Deep inside convolutional networks: Visualising image classification models and saliency maps. arXiv preprint arXiv:1312.6034. 2013 Dec 20.

Contributions

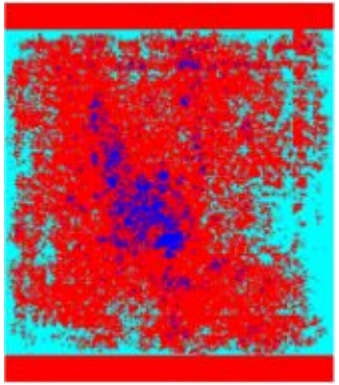
1. Obtain Understandable Visualisations of Conv Nets.
2. Image-specific class saliency map
    a. Can be used weakly supervised object localization

Weakly Supervised Object Localisation

Input Image



Saliency Map



Graph Cut Color Segmentation Output

Foreground Segmentation Mask

Image Credit: Simonyan K, Vedaldi A, Zisserman A. Deep inside convolutional networks: Visualising image classification models and saliency maps. arXiv preprint arXiv:1312.6034. 2013 Dec 20.

Results: Weakly Supervised Object Localisation

- ILSVRC-2013 Localisation Challenge
- Top 5 Error: 46.4%
  - Best submission: 29.9% (Fully Supervised Model)
- Authors' previous full supervised submission achieved 50%

Contributions

1. Obtain Understandable Visualisations of Conv Nets.
2. Image-specific class saliency map
   a. Can be used weakly supervised object localization
3. Generalization to existing deconvolutional network reconstruction procedure

Relation to Deconvolutional Networks

- Flipped Kernel in Backpropagation matches exactly to reconstruction layer
- Argmax during backpropagation of a max pool layer corresponds to max-pool in reconstruction

Discussion of Pros

Decoding the Black Box

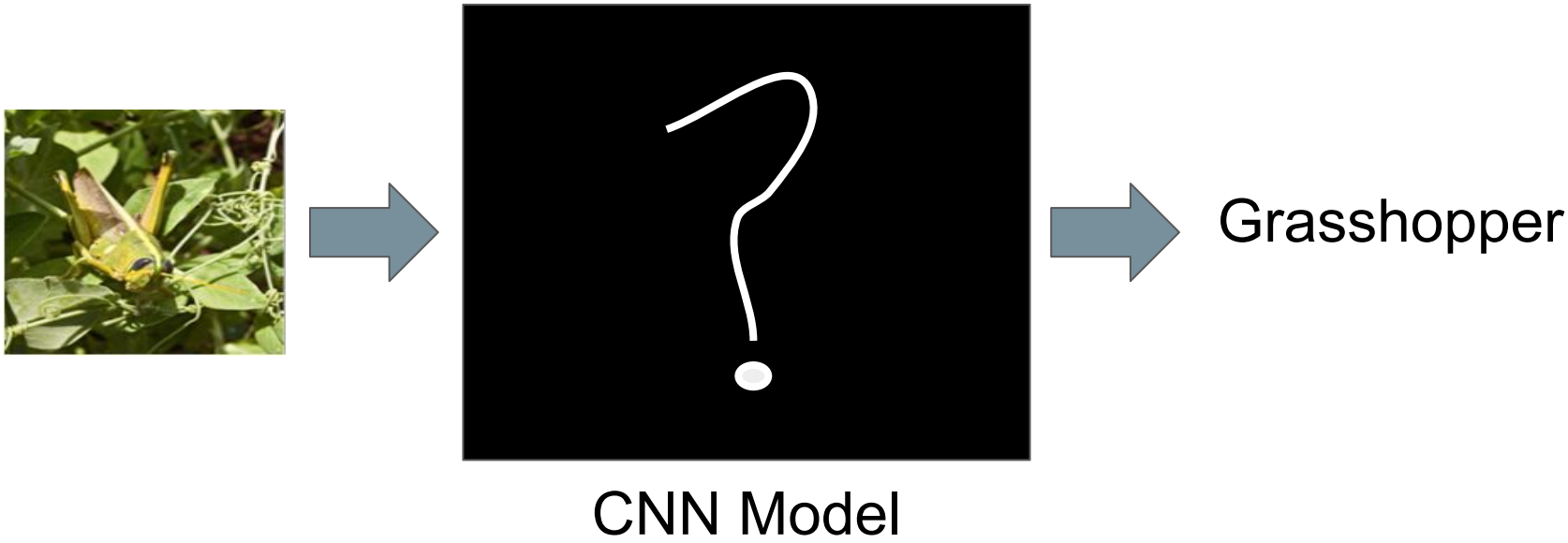


Image Credit: Simonyan K, Vedaldi A, Zisserman A. Deep inside convolutional networks: Visualising image classification models and saliency maps. arXiv preprint arXiv:1312.6034. 2013 Dec 20.

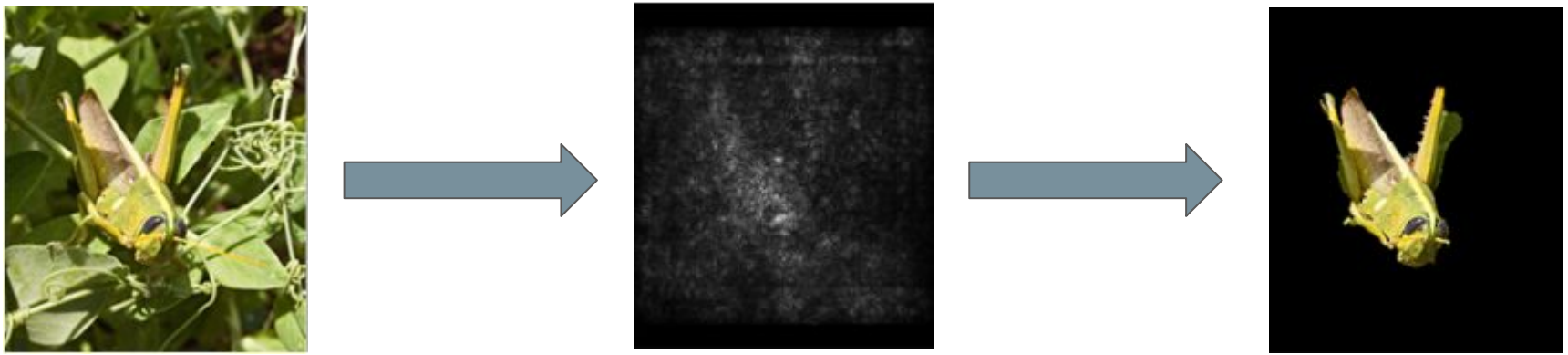
Intuitive Visual Results

Dumbbell

Image Credit: Simonyan K, Vedaldi A, Zisserman A. Deep inside convolutional networks: Visualising image classification models and saliency maps. arXiv preprint arXiv:1312.6034. 2013 Dec 20.

Importance of Image Saliency Maps

- Extending to Object Localisation shows usefulness of Saliency Maps
- Opened up research avenues for incorporating saliency maps.



Input Image

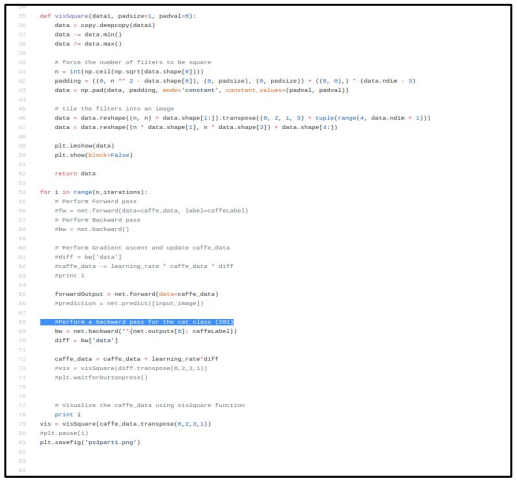
Saliency Map

Foreground Segmentation Mask

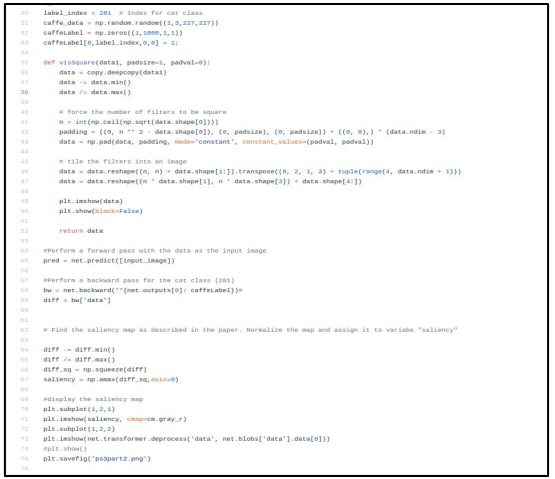
Image Credit: Simonyan K, Vedaldi A, Zisserman A. Deep inside convolutional networks: Visualising image classification models and saliency maps. arXiv preprint arXiv:1312.6034. 2013 Dec 20.

Easy to Understand & Implement

- Saliency Map extraction requires only a single backpropagation step
- Techniques easily translatable to code
  - https://raghakot.github.io/keras-vis/visualizations/saliency/



Class Model Visualization



Saliency Map Extraction

Image Credit: https://github.com/smajida/Deep_Inside_Convolutional_Networks

References

- Visualizing and Understanding Deep Neural Networks by Matt Zeiler