Visualization of CNN

Lina Hu

ShanghaiTech University

January 23, 2019

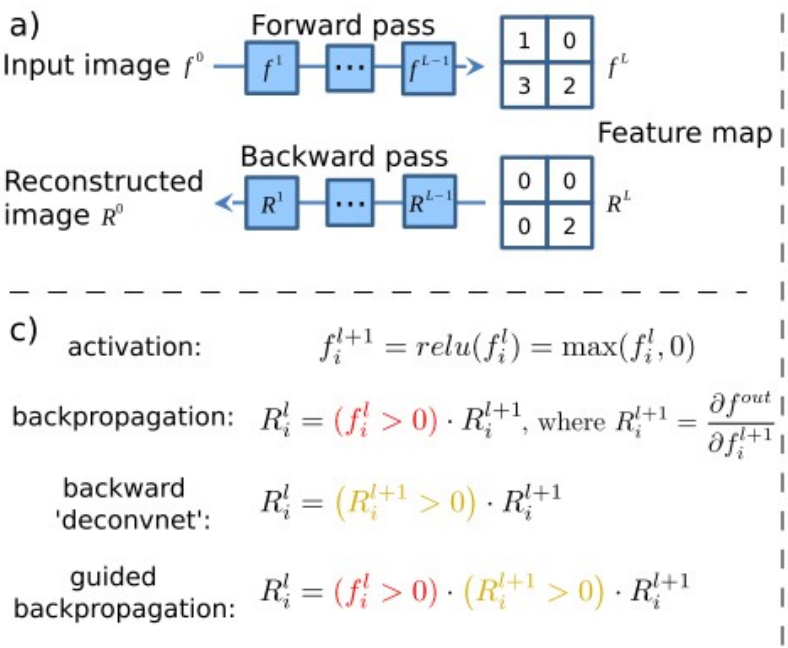
Overview

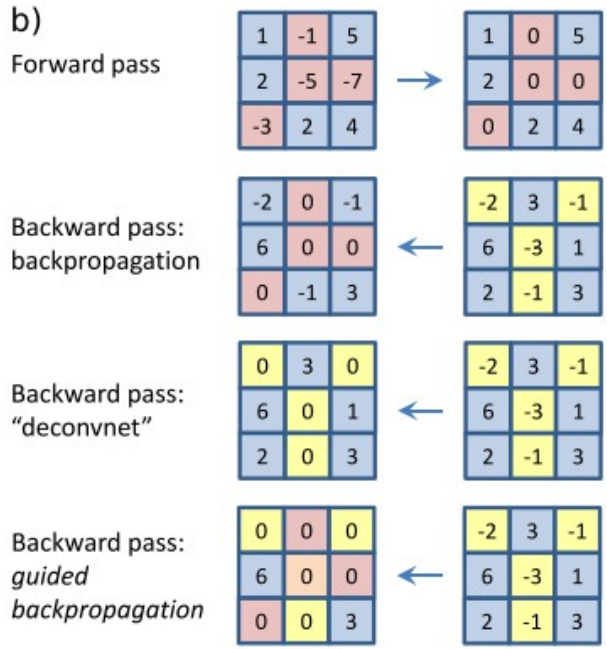
1. Different Backpropagation Methods[2]
2. Class Activation Mapping (CAM) [3]
3. Grad-CAM[1]

Overview

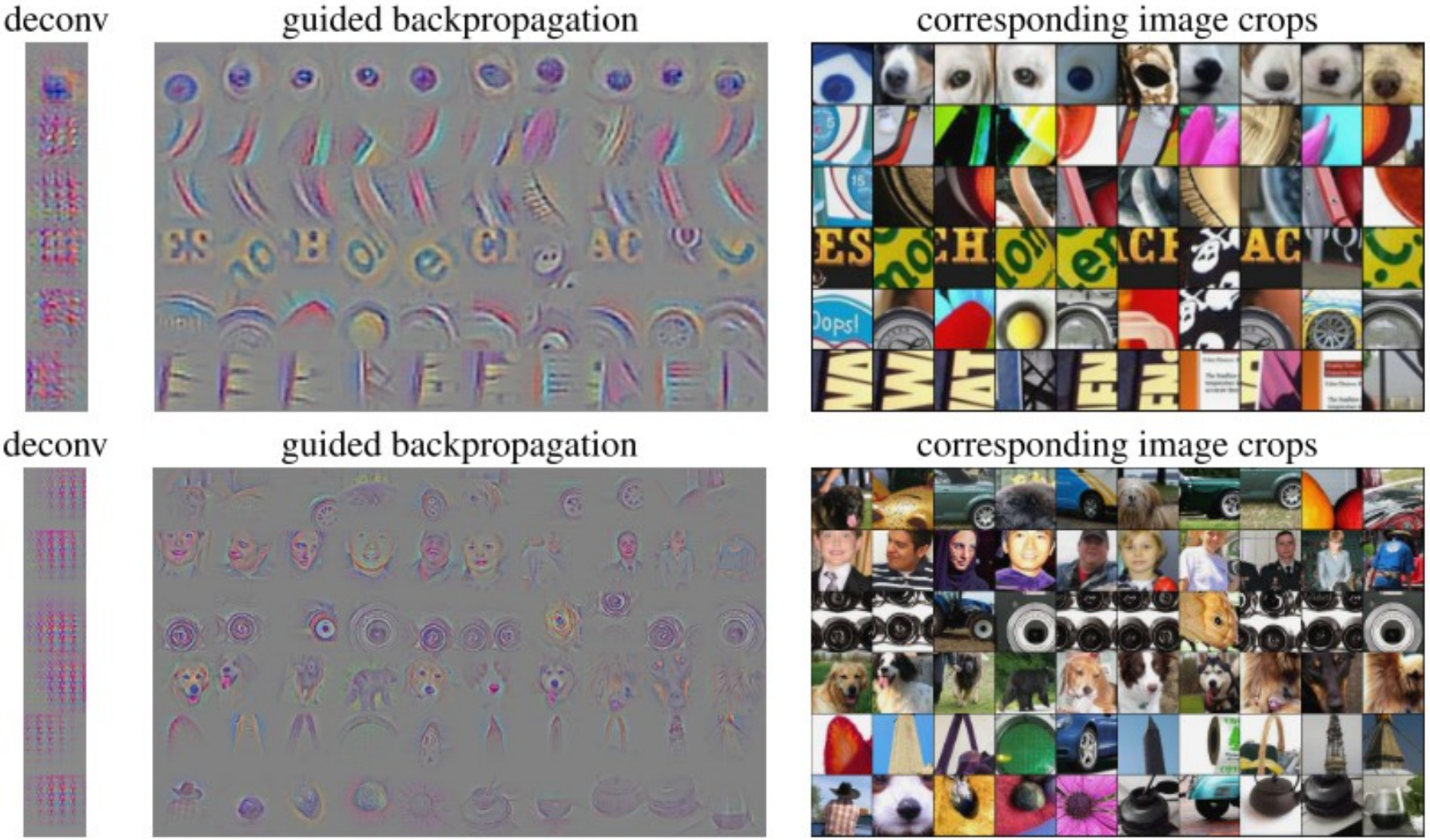
1. Common backpropagation
2. Deconvolution
3. Guided-backpropagation

Backpropagation Methods





Backpropagation Methods



CAM

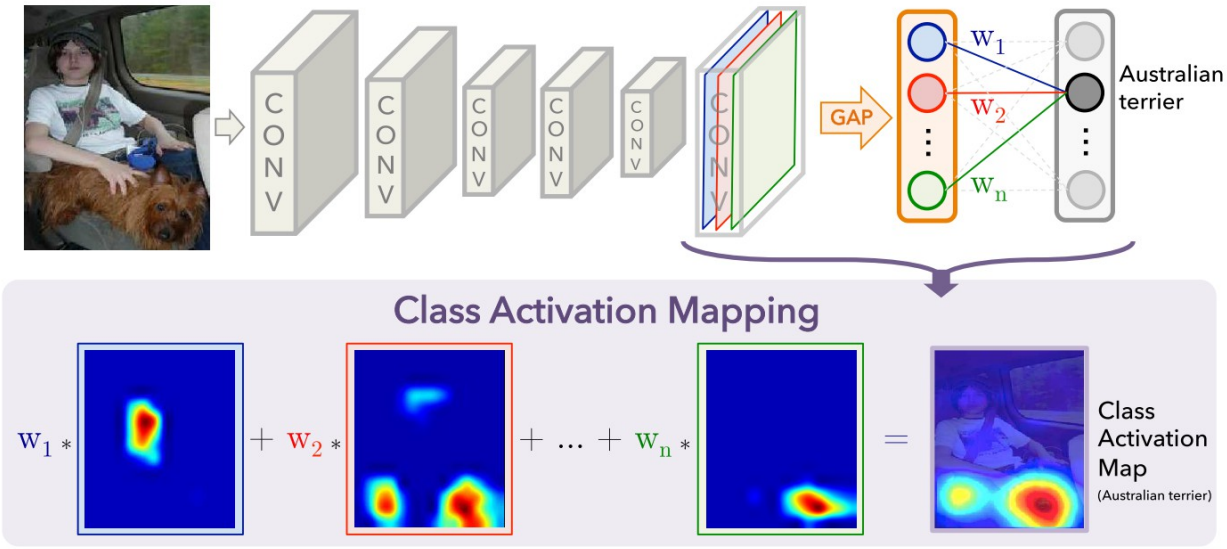
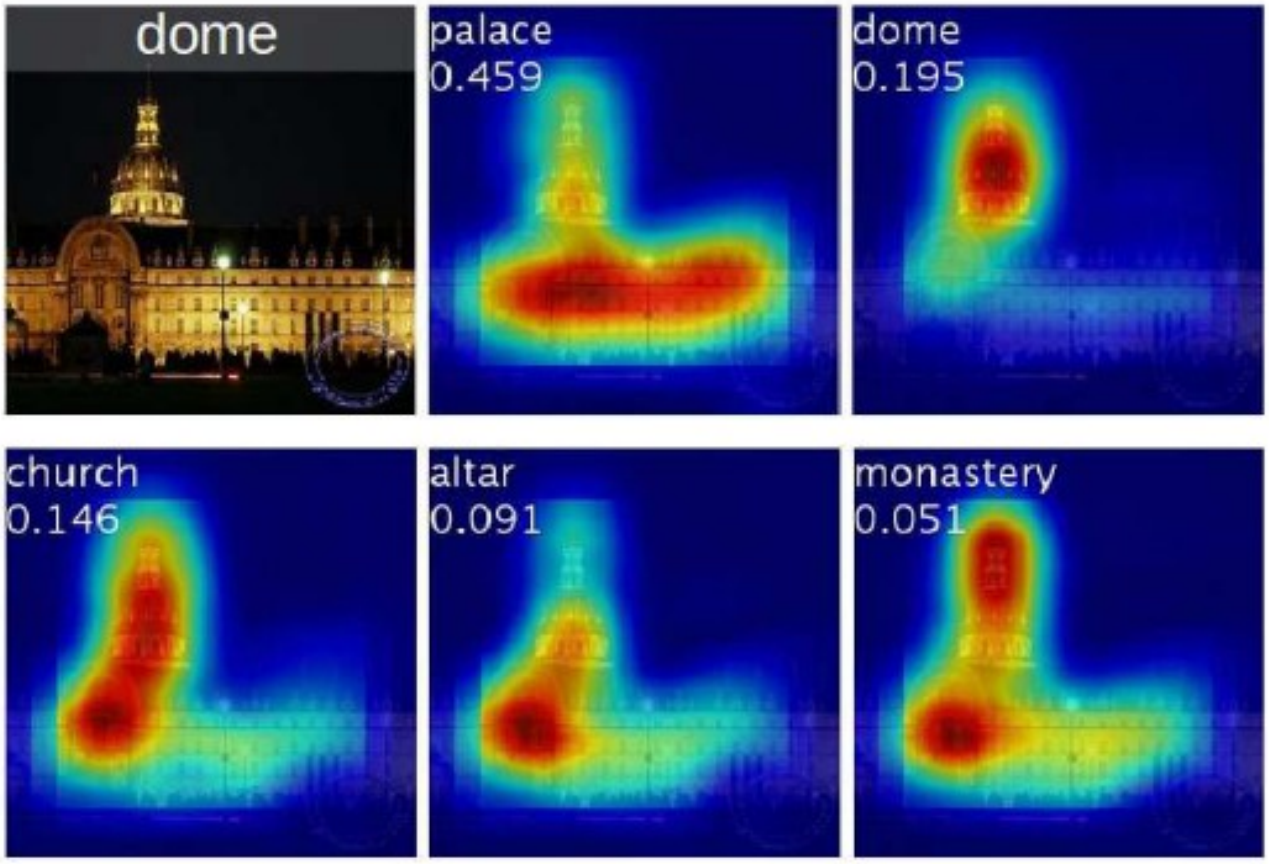


Figure 2. Class Activation Mapping: the predicted class score is mapped back to the previous convolutional layer to generate the class activation maps (CAMs). The CAM highlights the class-specific discriminative regions.

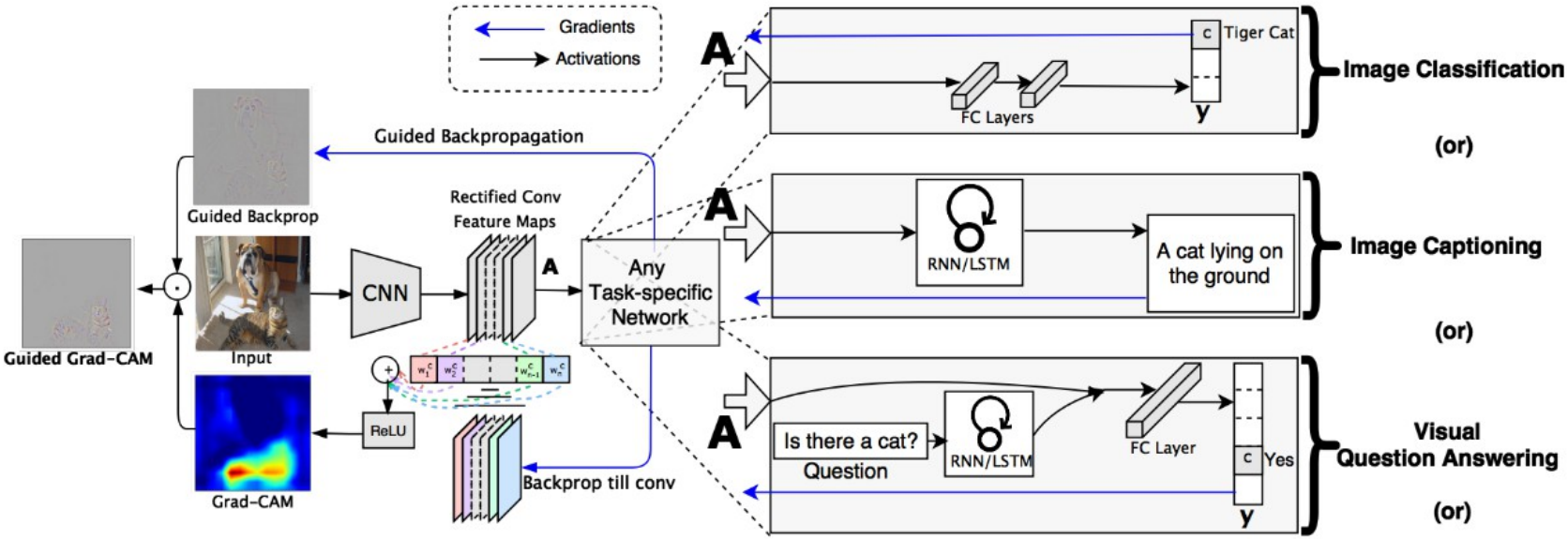
$$S_c = \sum_k w_k^c \sum_{x,y} f_k(x,y) = \sum_{x,y} \sum_k w_k^c f_k(x,y).$$

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

CAM

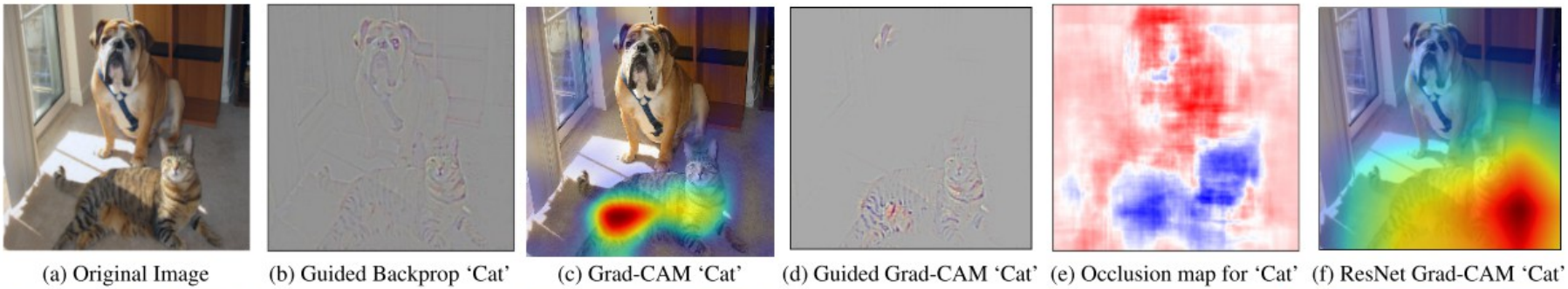

dome
palace
0.459
dome
0.195
church
0.146
altar
0.091
monastery
0.051

Grad-CAM



$$\alpha_k^c = \overbrace{\frac{1}{Z}\sum_i\sum_j}^{\text{global average pooling}} \underbrace{\frac{\partial y^c}{\partial A_{ij}^k}}_{\text{gradients via backprop}}$$

Grad-CAM



(a) Original Image (b) Guided Backprop 'Cat' (c) Grad-CAM 'Cat' (d) Guided Grad-CAM 'Cat' (e) Occlusion map for 'Cat' (f) ResNet Grad-CAM 'Cat'

(g) Original Image (h) Guided Backprop 'Dog' (i) Grad-CAM 'Dog' (j) Guided Grad-CAM 'Dog' (k) Occlusion map for 'Dog' (l) ResNet Grad-CAM 'Dog'

Reference I

Ramprasaath R Selvaraju et al. "Grad-cam: Visual explanations from deep networks via gradient-based localization". In: *2017 IEEE International Conference on Computer Vision (ICCV)*. IEEE. 2017, pp. 618–626.

Jost Tobias Springenberg et al. "Striving for simplicity: The all convolutional net". In: *arXiv preprint arXiv:1412.6806* (2014).

Bolei Zhou et al. "Learning deep features for discriminative localization". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016, pp. 2921–2929.

Thanks

Thanks for Attention!