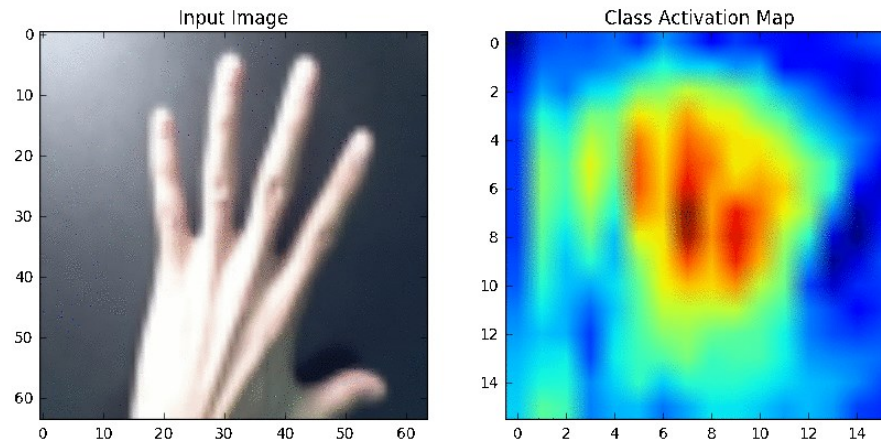


Visualization of CNN



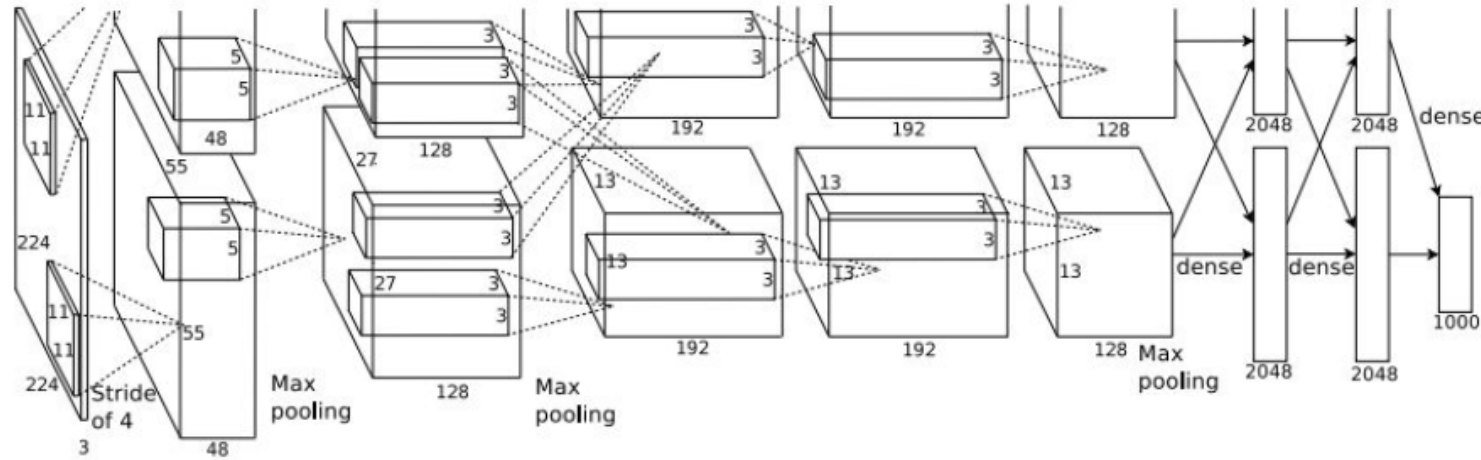
Fast Campus
Start Deep Learning with TensorFlow

What's going on inside CNN?

This image is CC0 public domain



Input Image:
3 x 224 x 224



What are the intermediate features looking for?

Class Scores:
1000 numbers

Visualize Patches that Maximally Activate Neurons

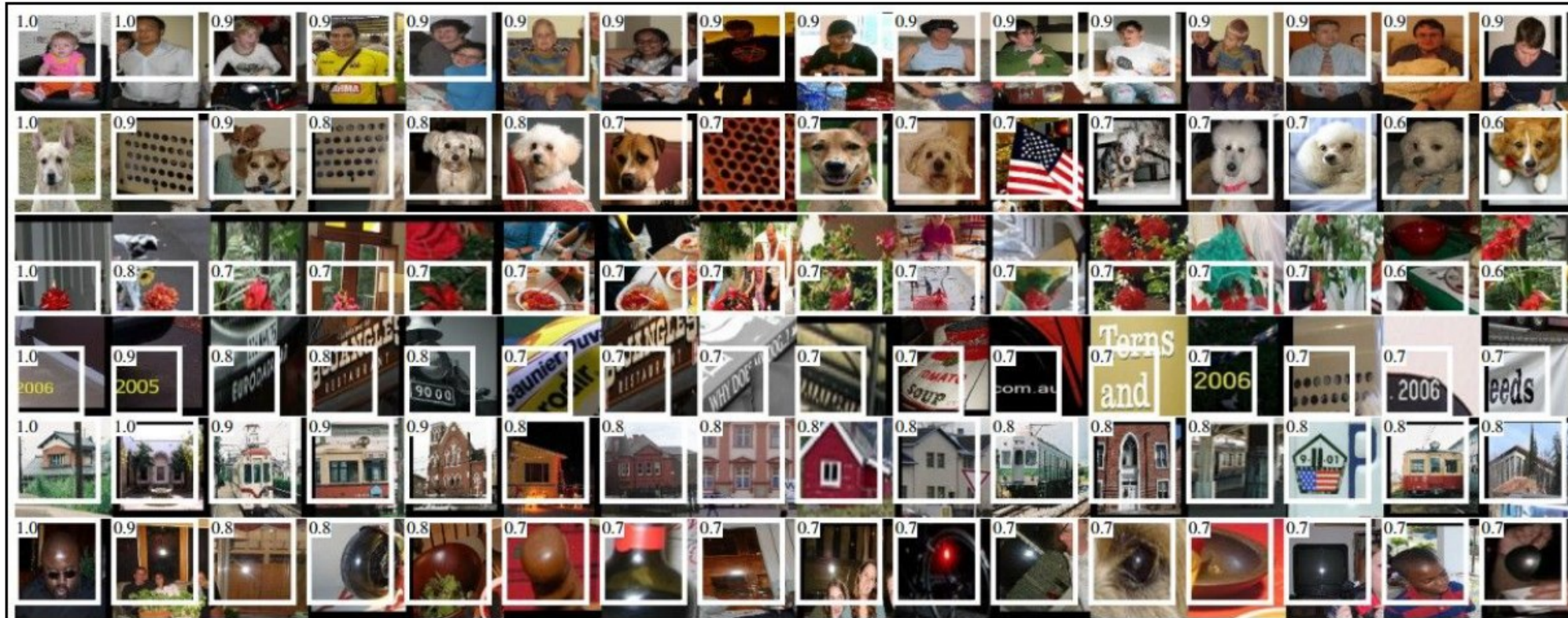
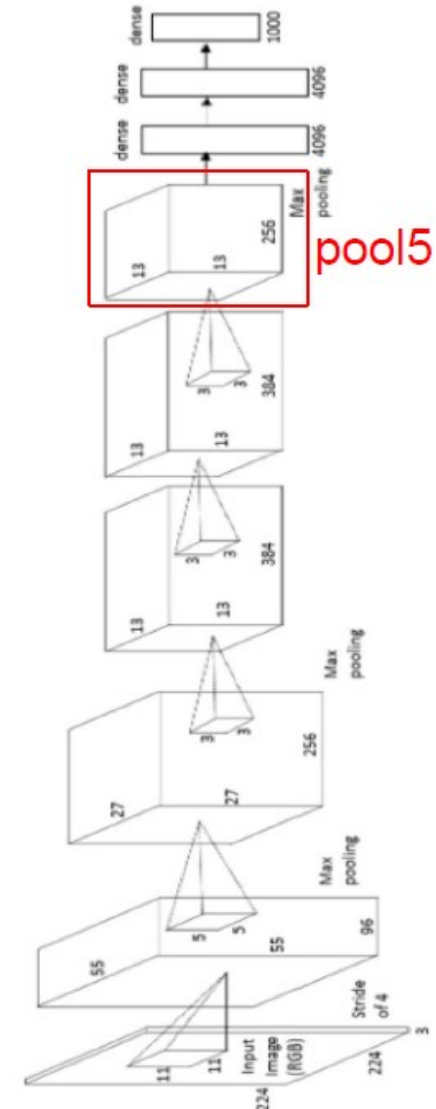


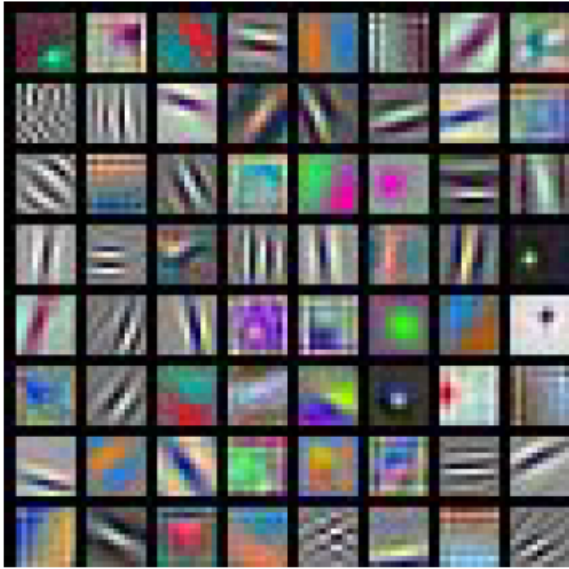
Figure 4: Top regions for six pool_5 units. Receptive fields and activation values are drawn in white. Some units are aligned to concepts, such as people (row 1) or text (4). Other units capture texture and material properties, such as dot arrays (2) and specular reflections (6).

Rich feature hierarchies for accurate object detection and semantic segmentation
 [Girshick, Donahue, Darrell, Malik]



Visualize Filters

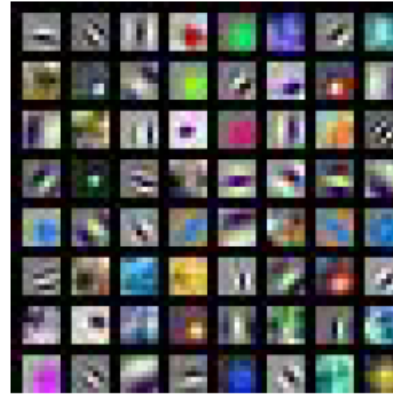
- Only interpretable on the first layer



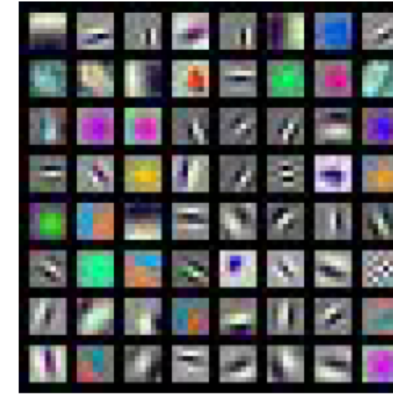
AlexNet:
 $64 \times 3 \times 11 \times 11$



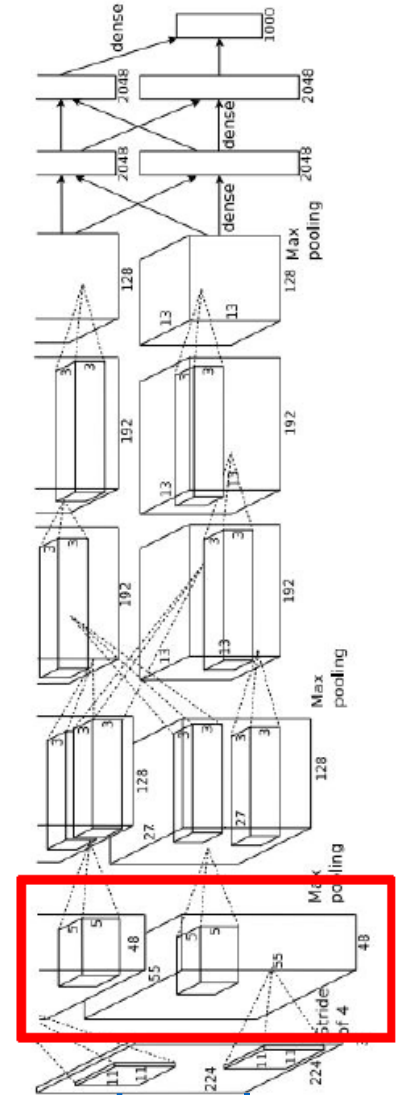
ResNet-18:
 $64 \times 3 \times 7 \times 7$



ResNet-101:
 $64 \times 3 \times 7 \times 7$



DenseNet-121:
 $64 \times 3 \times 7 \times 7$



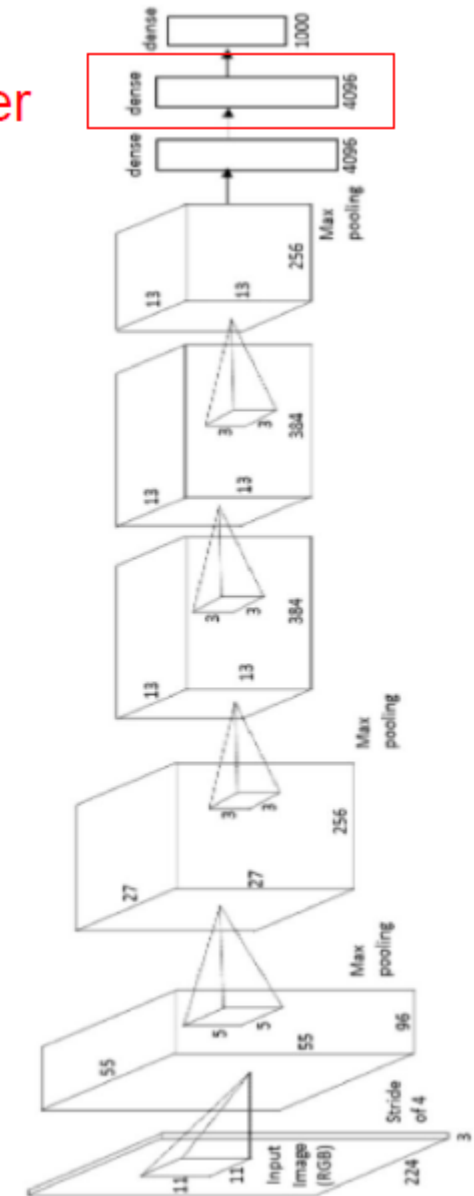
- <http://cs.stanford.edu/people/karpathy/convnetjs/demo/cifar10.html>

Visualizing the Representation

fc7 layer

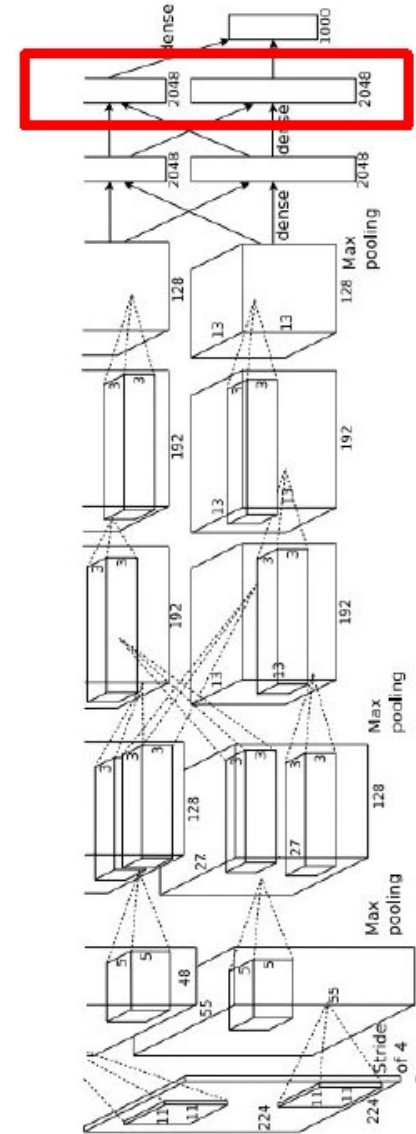
4096-dimensional “code” for an image
(layer immediately before the classifier)

can collect the code for many images



Last Layer : Nearest Neighbors

4096-dim vector



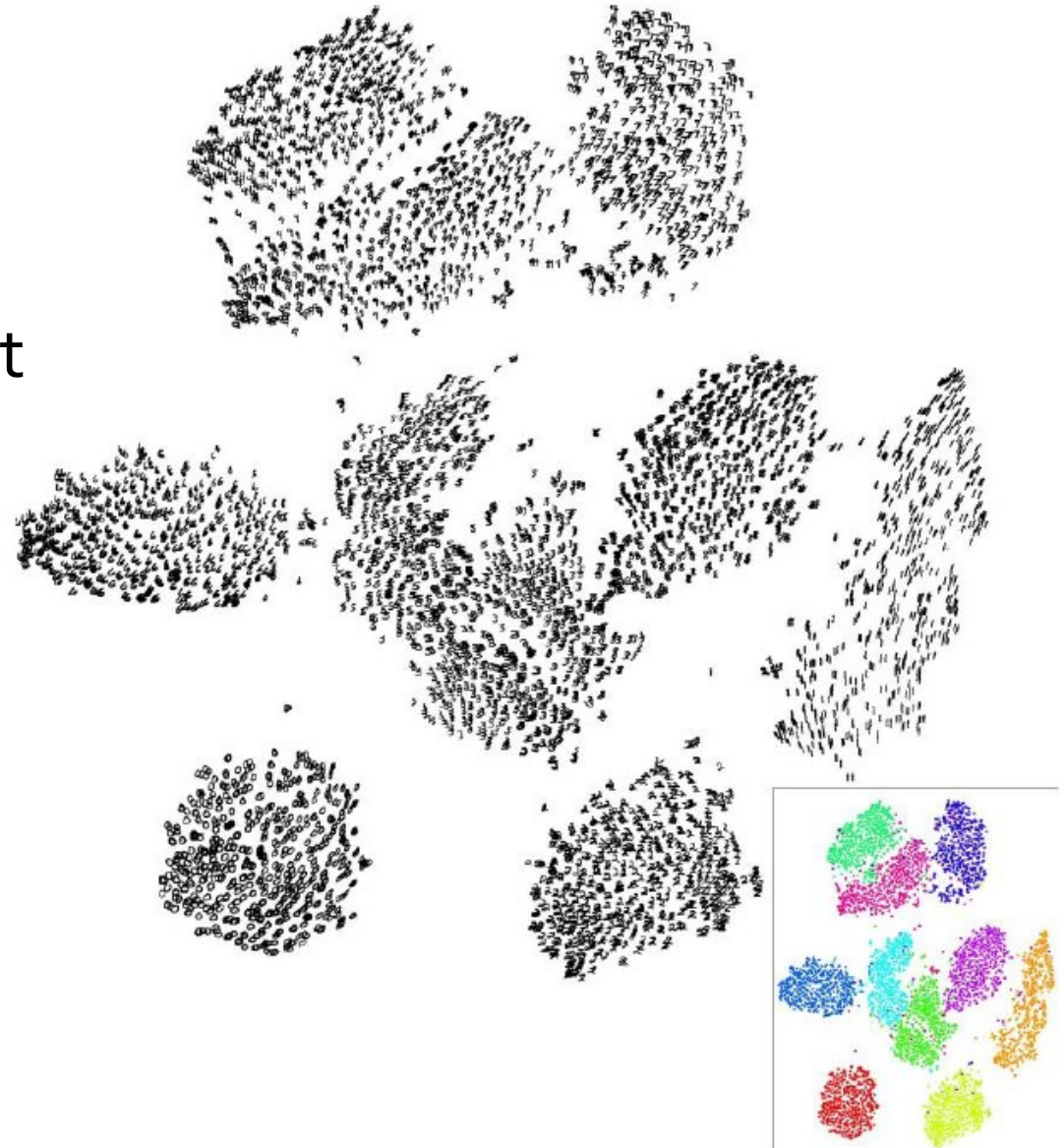
Test image L2 Nearest neighbors in feature space

Recall: Nearest neighbors in pixel space

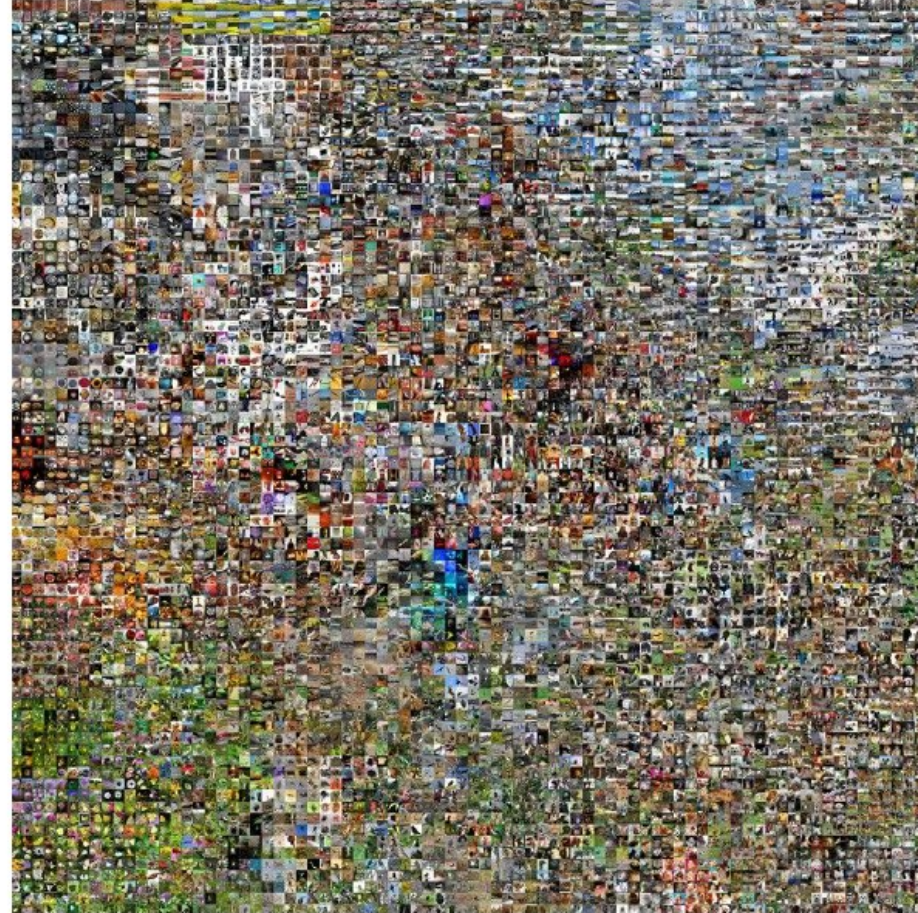


Last Layer: Dimensionality Reduction

- Visualize the “space” of FC7 feature vectors by reducing dimensionality of vectors from 4096 to 2 dimensions
- Simple algorithm: Principle Component Analysis(PCA)
- More complex: t-SNE

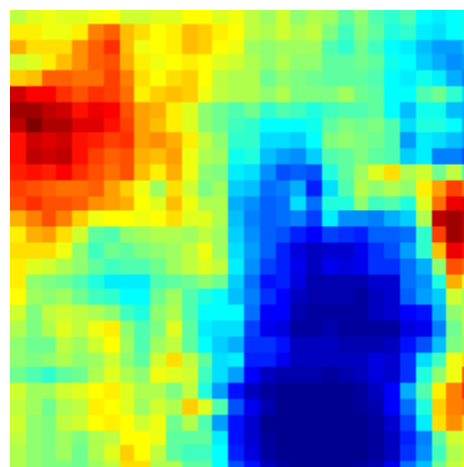
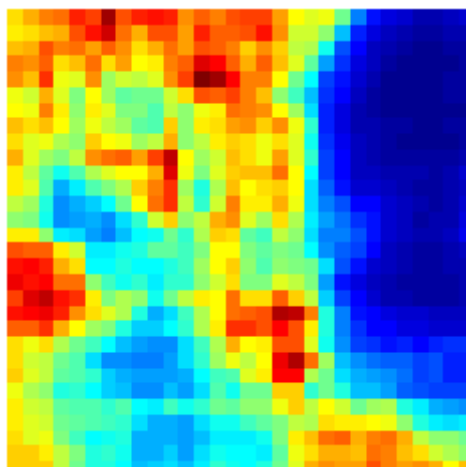
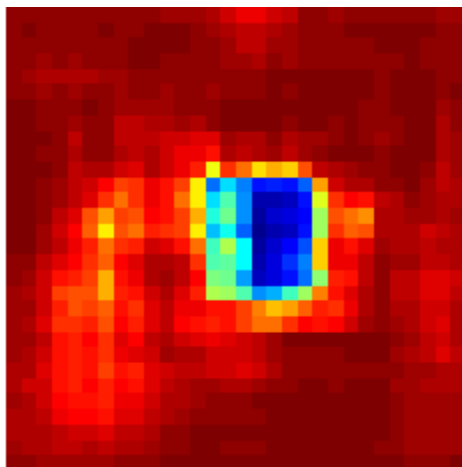
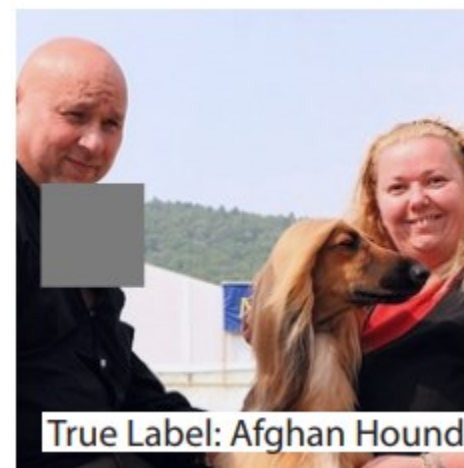


Last Layer: Dimensionality Reduction



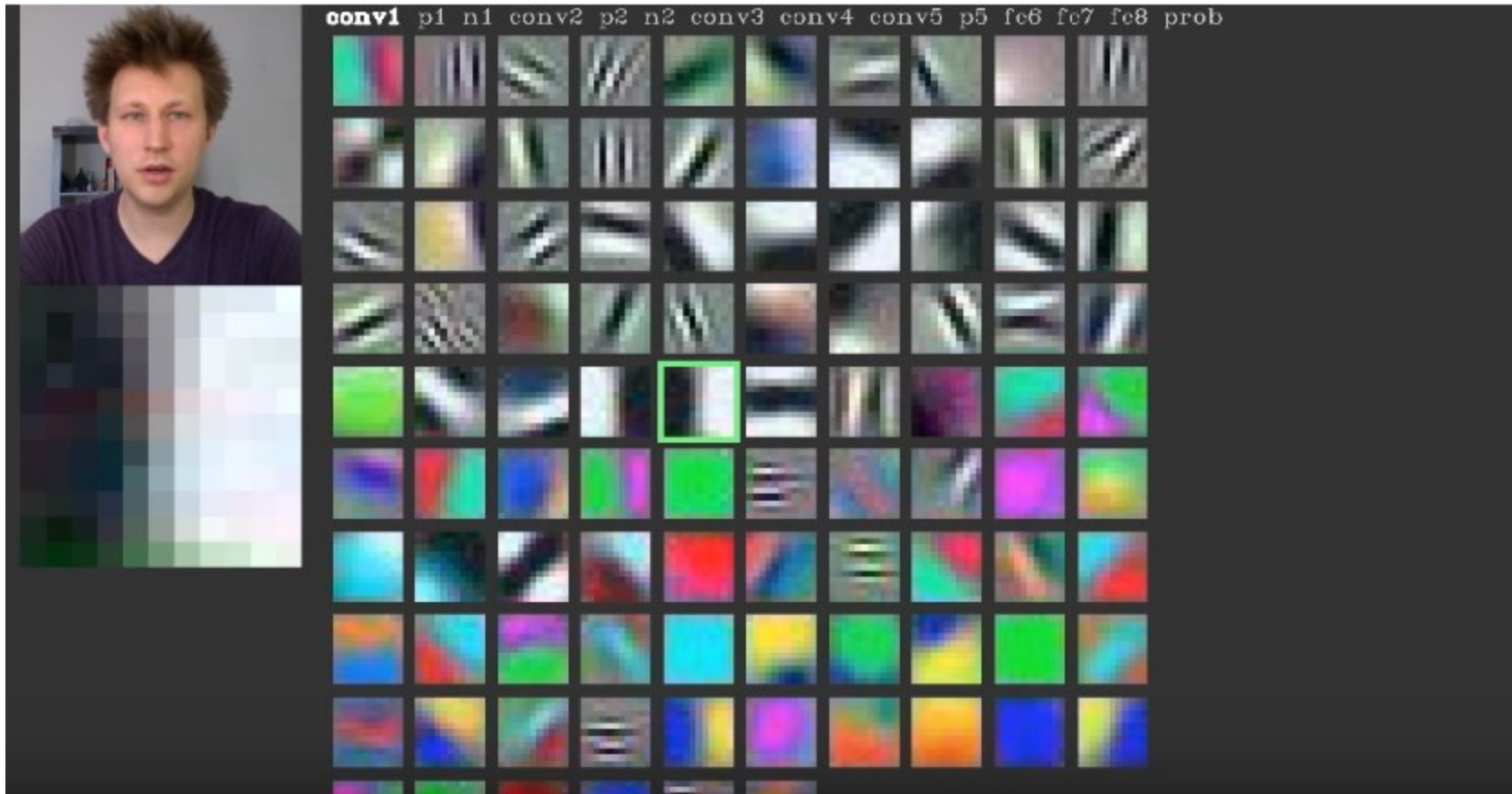
- <http://cs.stanford.edu/people/karpathy/cnnembed/>

Occlusion Experiments

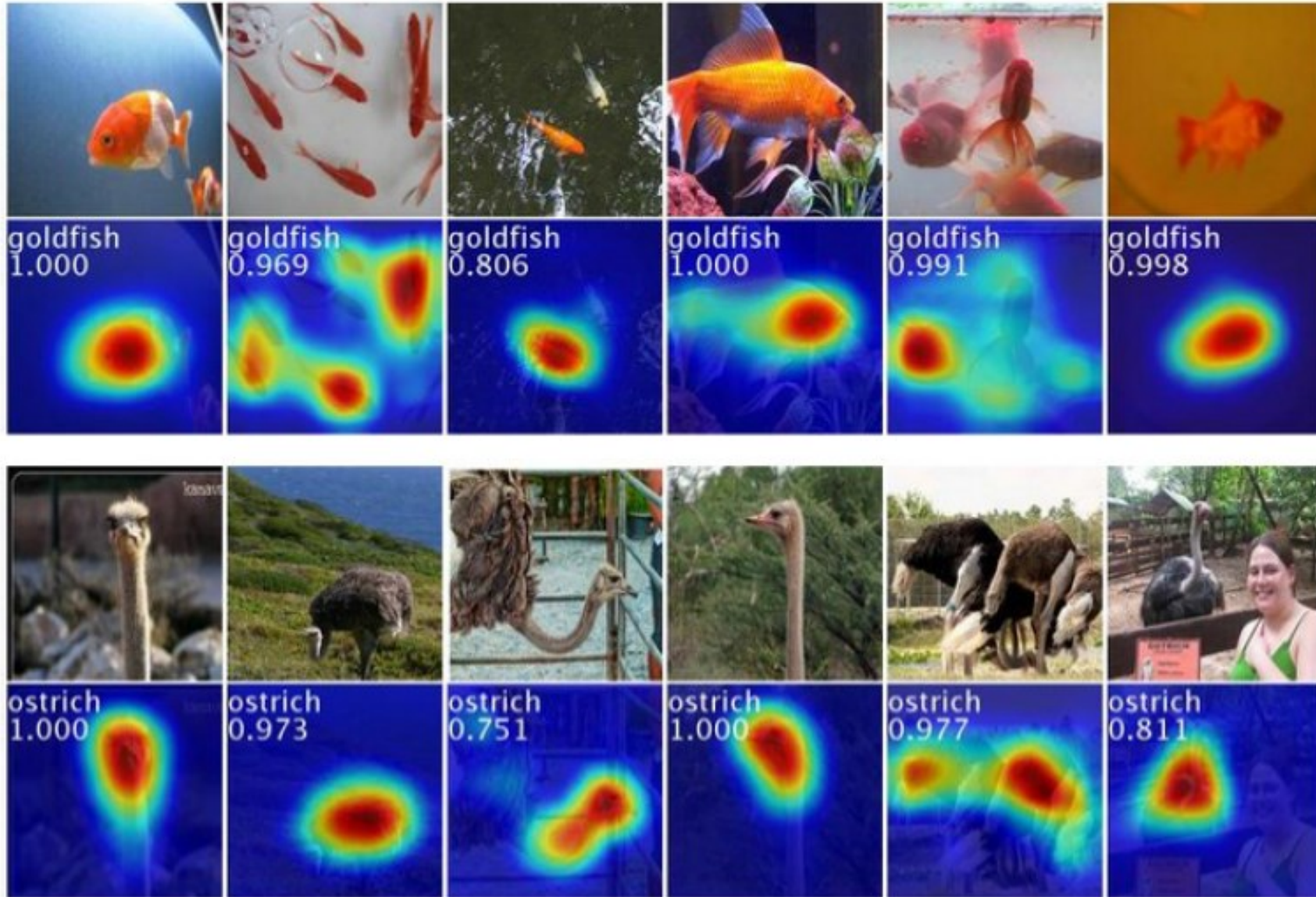


Visualizing Activations

- <https://www.youtube.com/watch?v=AgkflQ4lGaM>



Weakly Supervised Learning



Class activation map (CAM)

- **Identify important image regions** by projecting back the weights of output layer to convolutional feature maps.
- CAMs can be generated for each class in single image.
- Regions for each categories are different in given image.
 - palace, dome, church ...

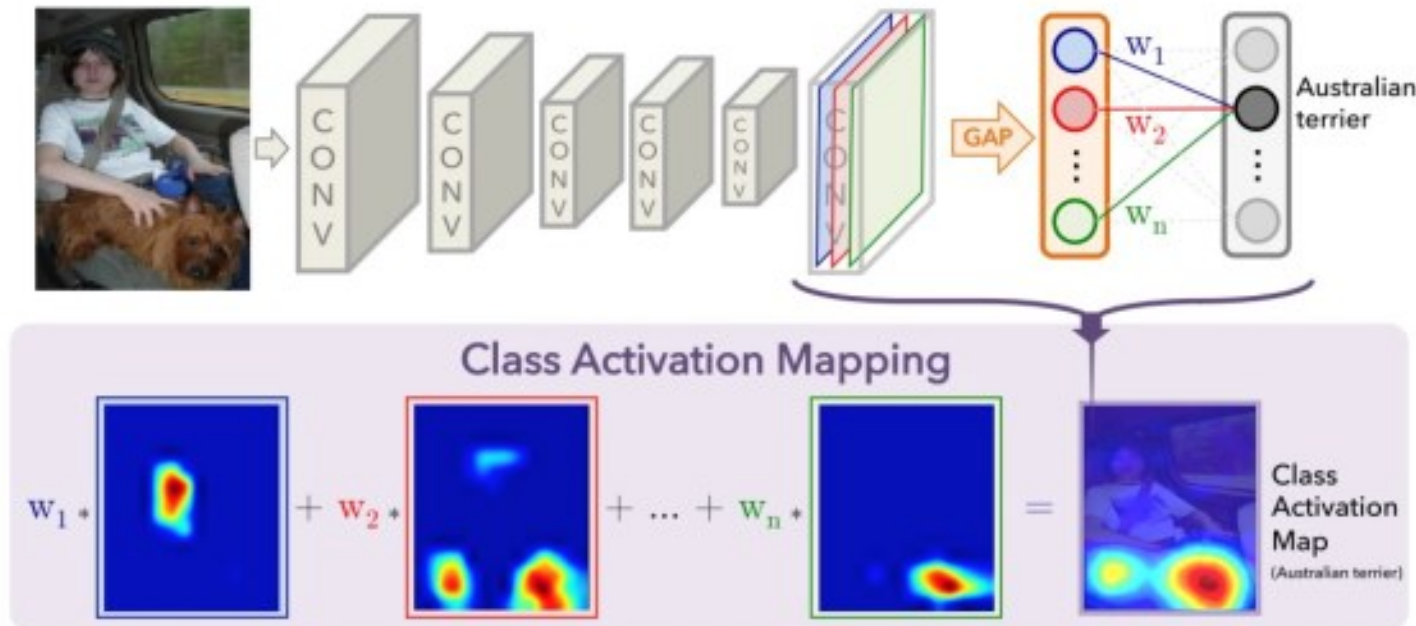


Figure 2. Class Activation Mapping: the predicted class score is mapped back to the previous convolutional layer to generate the class activation maps (CAMs). The CAM highlights the class-specific discriminative regions.

Results

- CAM on top 5 predictions on an image
- CAM for one object class in images

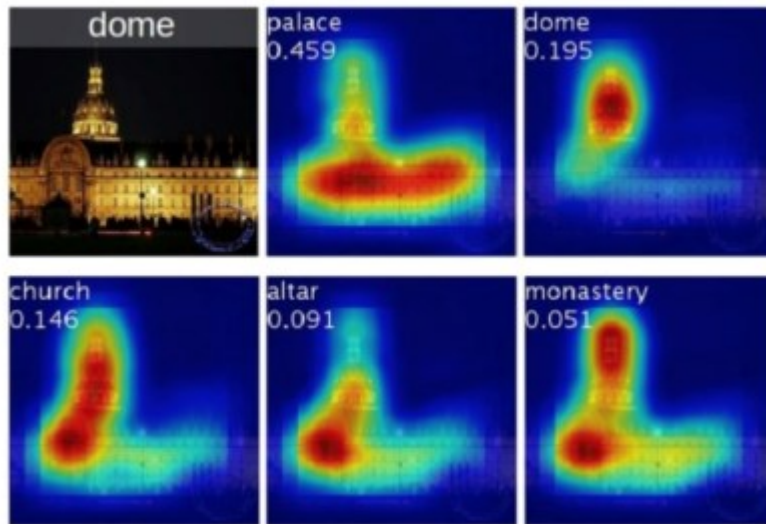


Figure 4. Examples of the CAMs generated from the top 5 predicted categories for the given image with ground-truth as dome. The predicted class and its score are shown above each class activation map. We observe that the highlighted regions vary across predicted classes e.g., *dome* activates the upper round part while *palace* activates the lower flat part of the compound.

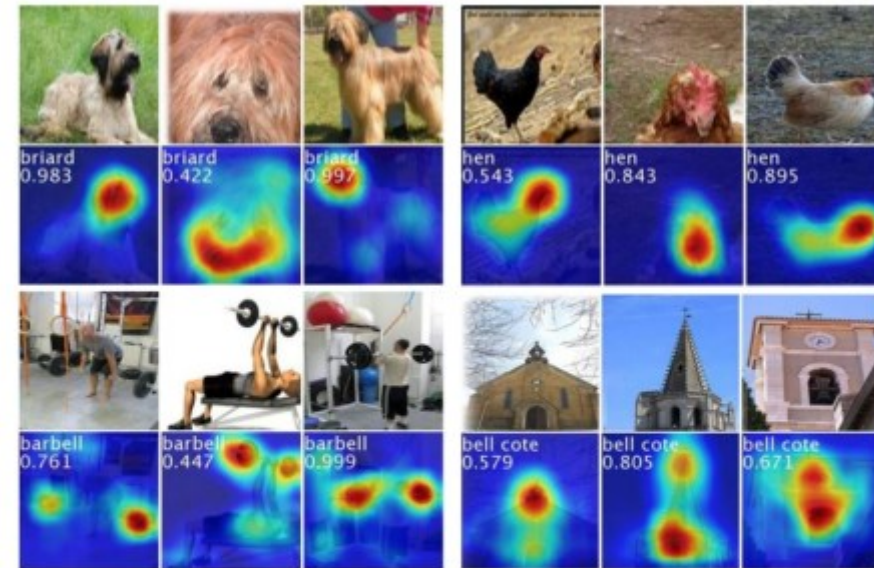


Figure 3. The CAMs of four classes from ILSVRC [20]. The maps highlight the discriminative image regions used for image classification e.g., the head of the animal for *briard* and *hen*, the plates in *barbell*, and the bell in *bell cote*.

GAP & GMP

- GAP (upper) vs GMP (lower)
- GAP outperforms GMP
- GAP highlights more **complete** object regions and less background noise.
- Loss for average pooling benefits when the network identifies **all discriminative** regions of an object

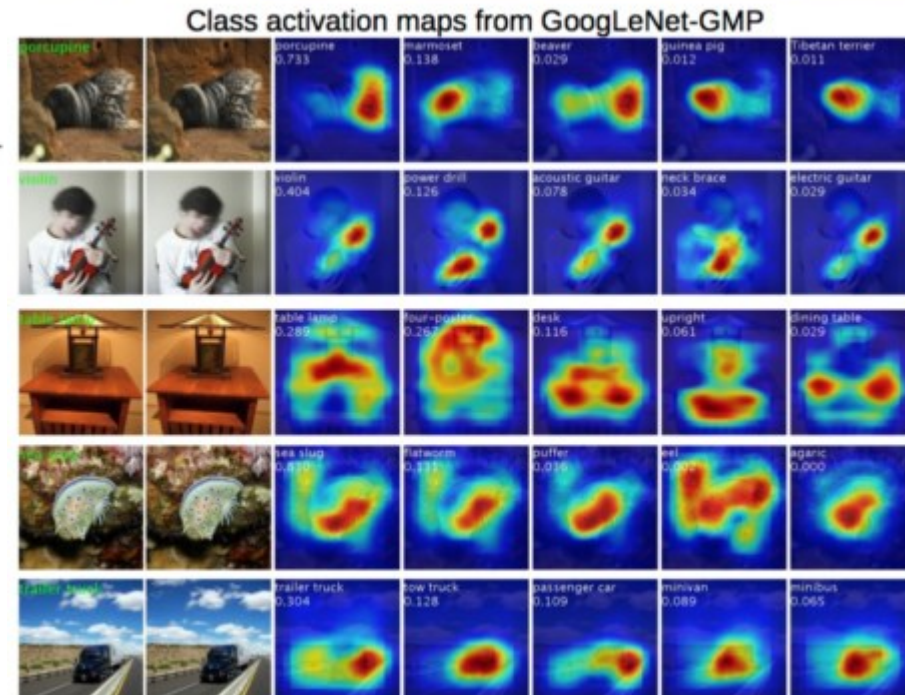
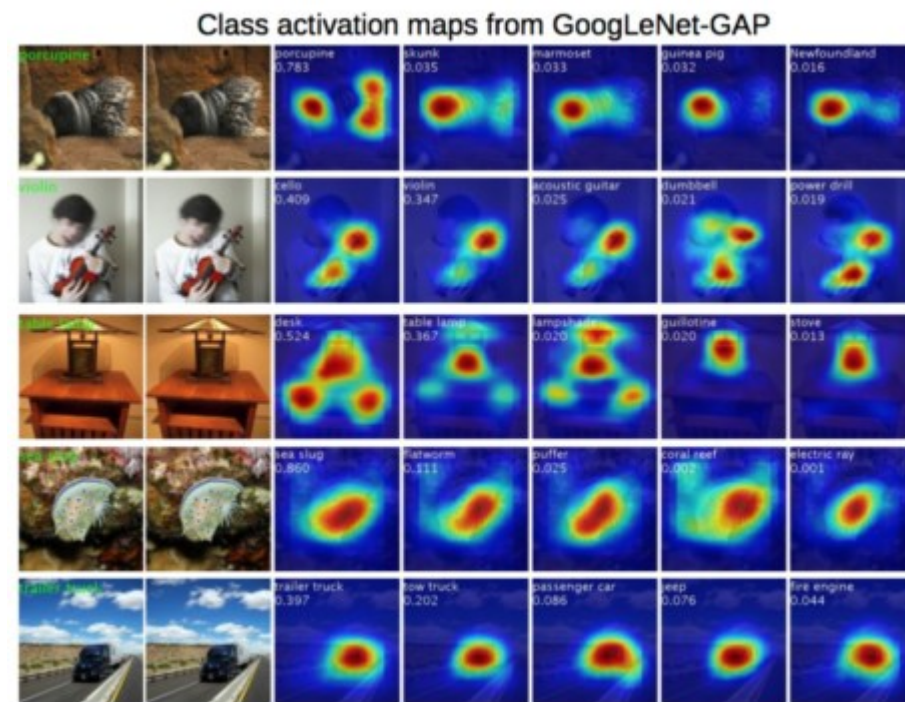


Table 2. Localization error on the ILSVRC validation set. *Backprop* refers to using [22] for localization instead of CAM.

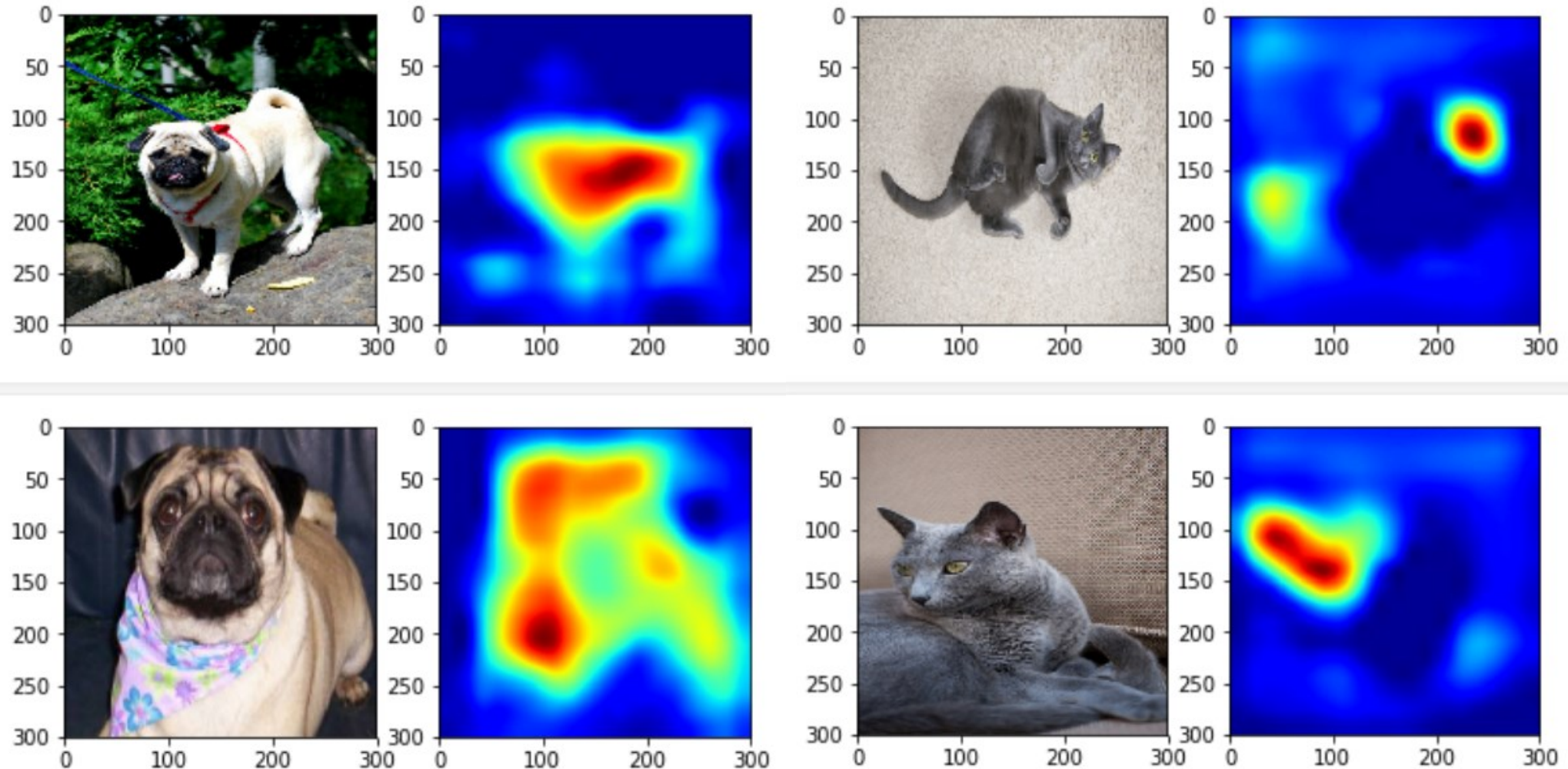
Method	top-1 val. error	top-5 val. error
GoogLeNet-GAP	56.40	43.00
VGGnet-GAP	57.20	45.14
GoogLeNet	60.09	49.34
AlexNet*-GAP	63.75	49.53
AlexNet-GAP	67.19	52.16
NIN	65.47	54.19
Backprop on GoogLeNet	61.31	50.55
Backprop on VGGnet	61.12	51.46
Backprop on AlexNet	65.17	52.64
GoogLeNet-GMP	57.78	45.26

Table 1. Classification error on the ILSVRC validation

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

Weakness of CAM (Weakly Supervised Localicztion)

- Focusing on discriminative features



Weakness of CAM (Weakly Supervised Localicztion)

- Focusing on discriminative features

