Deep Learning

Lecture 13: Visualization & Understanding

심주용
숙명여자대학교
기계시스템학부

Milestone Meetings

Prepare powerpoint Slides for progress report, more or less containing:

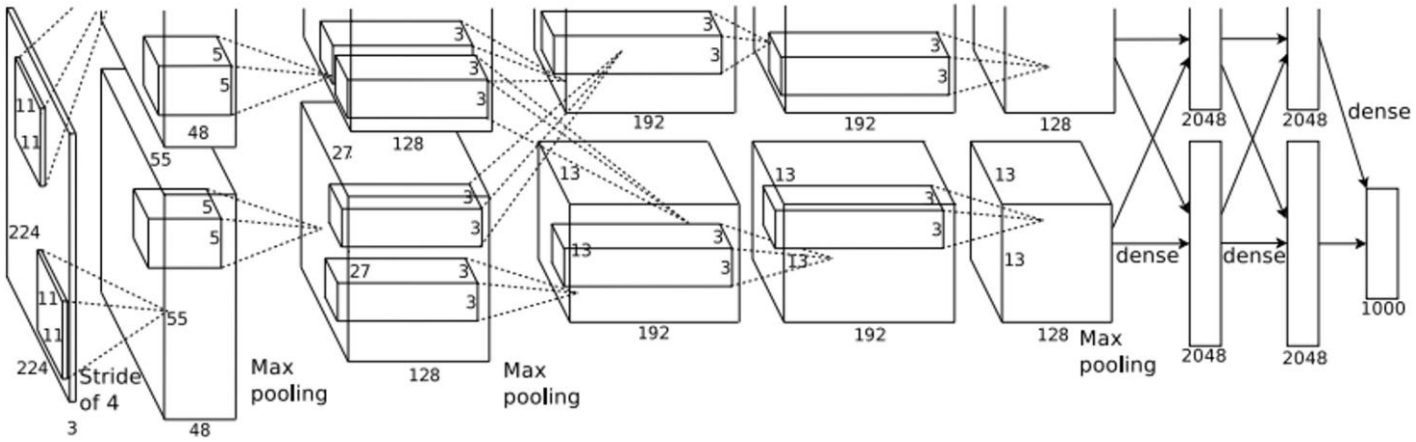
1. **Literature review (3+ sources)**
2. **Indication that code is up and running**
3. Data source explained correctly
4. What Github repo or other code you're basing your work off of
5. **Ran baseline model have results**
    a. Yes, points are taken off for no model running & no preliminary results
6. Data pipeline should be in place
7. Brief discussion of your preliminary results

5/22 (월) 1조, 2조, 3조, 4조
5/24 (수) 5조, 6조, 7조

What's going on inside Convolutional Networks?

This image is CC0 public domain

Input Image:
3 x 224 x 224

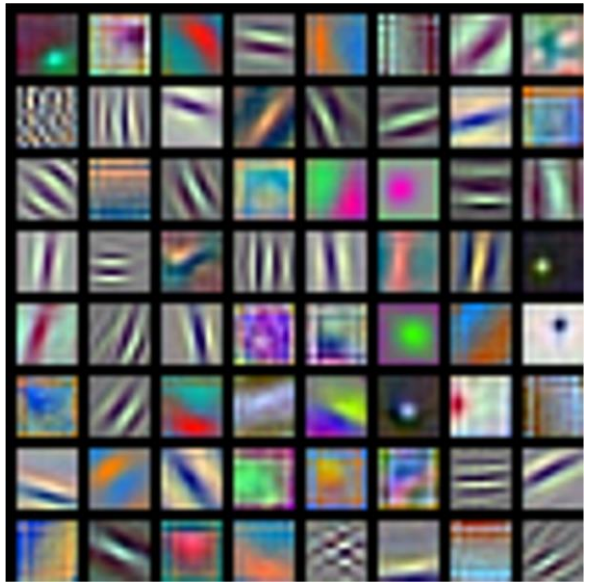


Class Scores:
1000 numbers

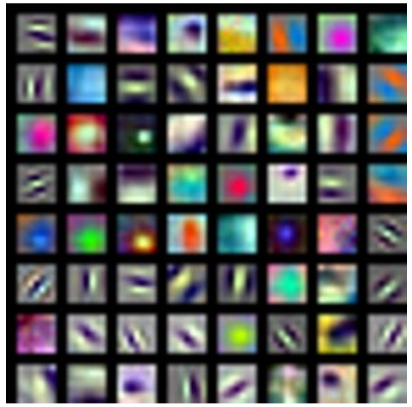
What are the intermediate features looking for?

Krizhevsky et al, "ImageNet Classification with Deep Convolutional Neural Networks", NeurIPS 2012.

First Layer: Visualize Filters

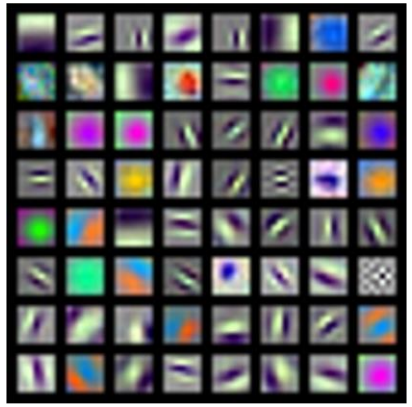


AlexNet:
64 x 3 x 11 x 11

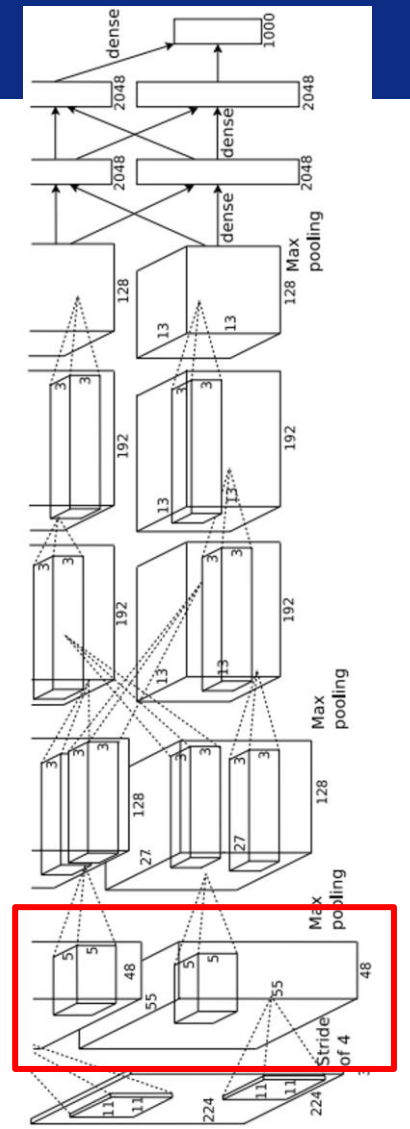


ResNet-18:
64 x 3 x 7 x 7

ResNet-101:
64 x 3 x 7 x 7



DenseNet-121:
64 x 3 x 7 x 7



Krizhevsky, “One weird trick for parallelizing convolutional neural networks”, arXiv 2014
He et al, “Deep Residual Learning for Image Recognition”, CVPR 2016
Huang et al, “Densely Connected Convolutional Networks”, CVPR 2017

Higher Layers: Visualize Filters

We can visualize filters at higher layers, but not that interesting

Source: ConvNetJS CIFAR-10 example
https://cs.stanford.edu/people/karpathy/convnetjs/demo/cifar10.html

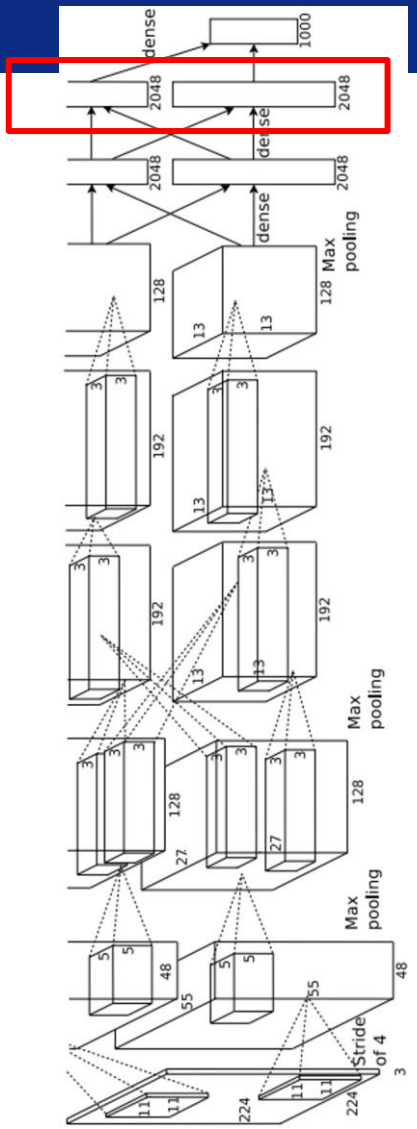
First layer weights: 16 x 3 x 7 x 7

Second layer weights: 20 x 16 x 7 x 7

Third layer weights: 20 x 20 x 7 x 7

Last Layer

FC7 layer



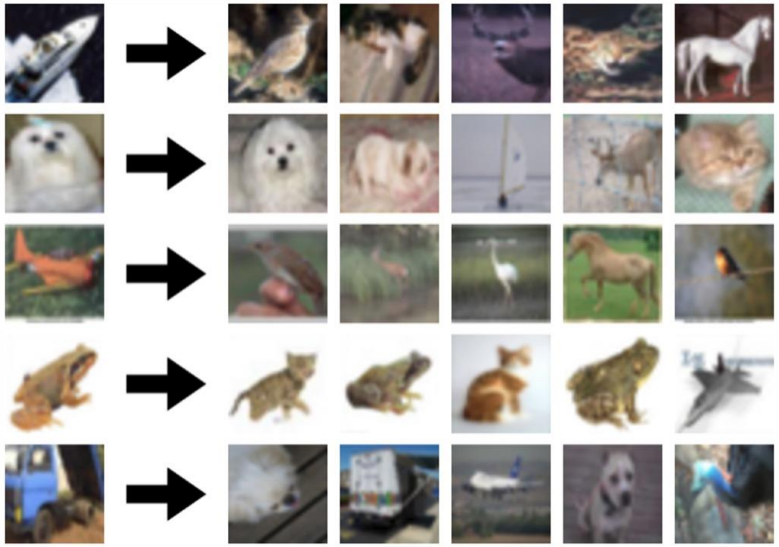
4096-dimensional feature vector for an image (layer immediately before the classifier)

Run the network on many images, collect the feature vectors

Krizhevsky et al, “ImageNet Classification with Deep Convolutional Neural Networks”, NeurIPS 2012.

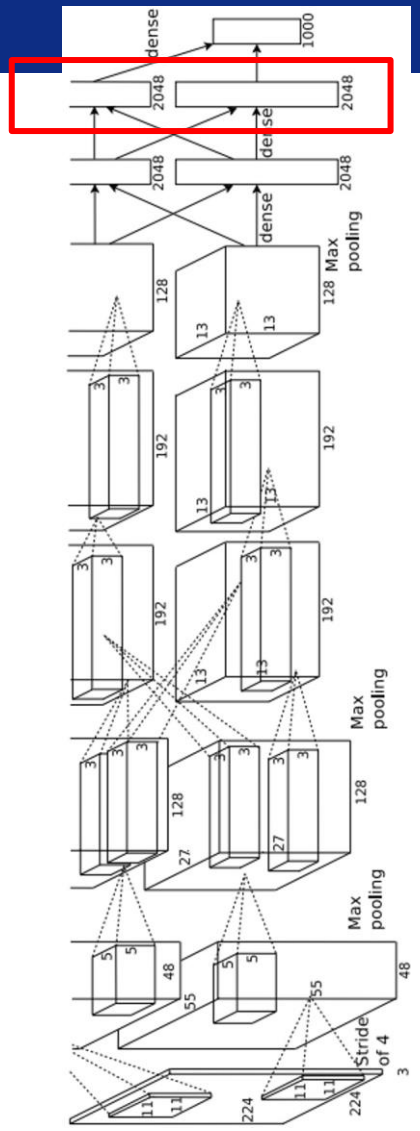
Last Layer: Nearest Neighbors

Recall: Nearest



Test image

L2 Nearest neighbors in <u>feature</u> space



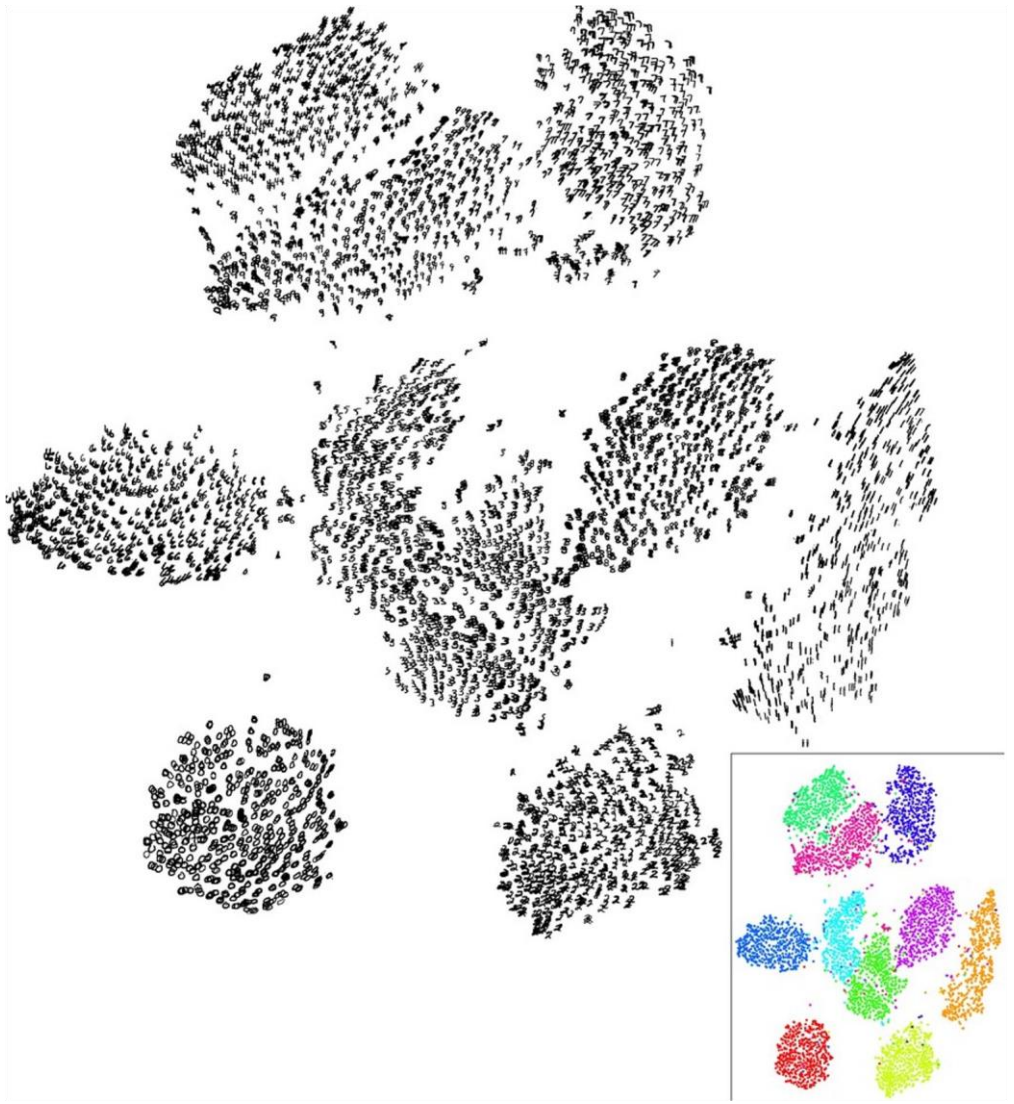
Krizhevsky et al, “ImageNet Classification with Deep Convolutional Neural Networks”, NeurIPS 2012.

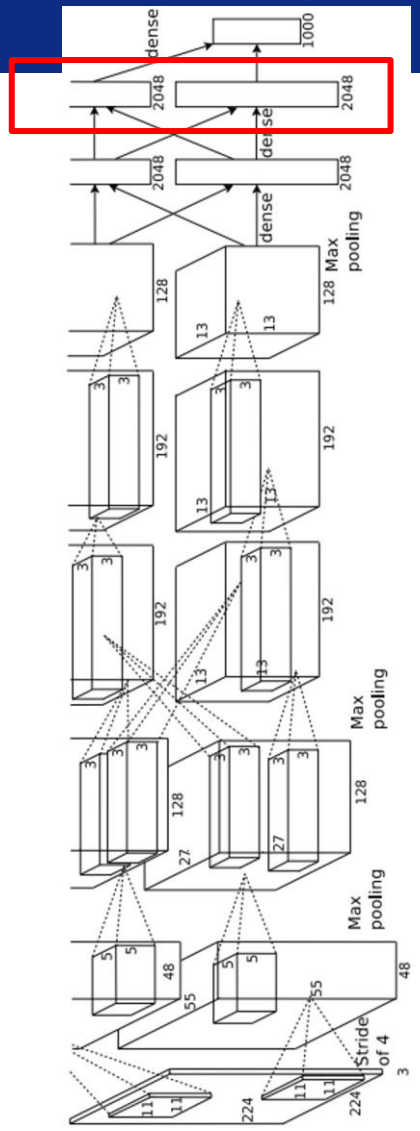
Last Layer: Dimensionality Reduction

Visualize the “space” of FC7 feature vectors by reducing dimensionality of vectors from 4096 to 2 dimensions

Simple algorithm: Principal Component Analysis (PCA)

More complex: **t-SNE**

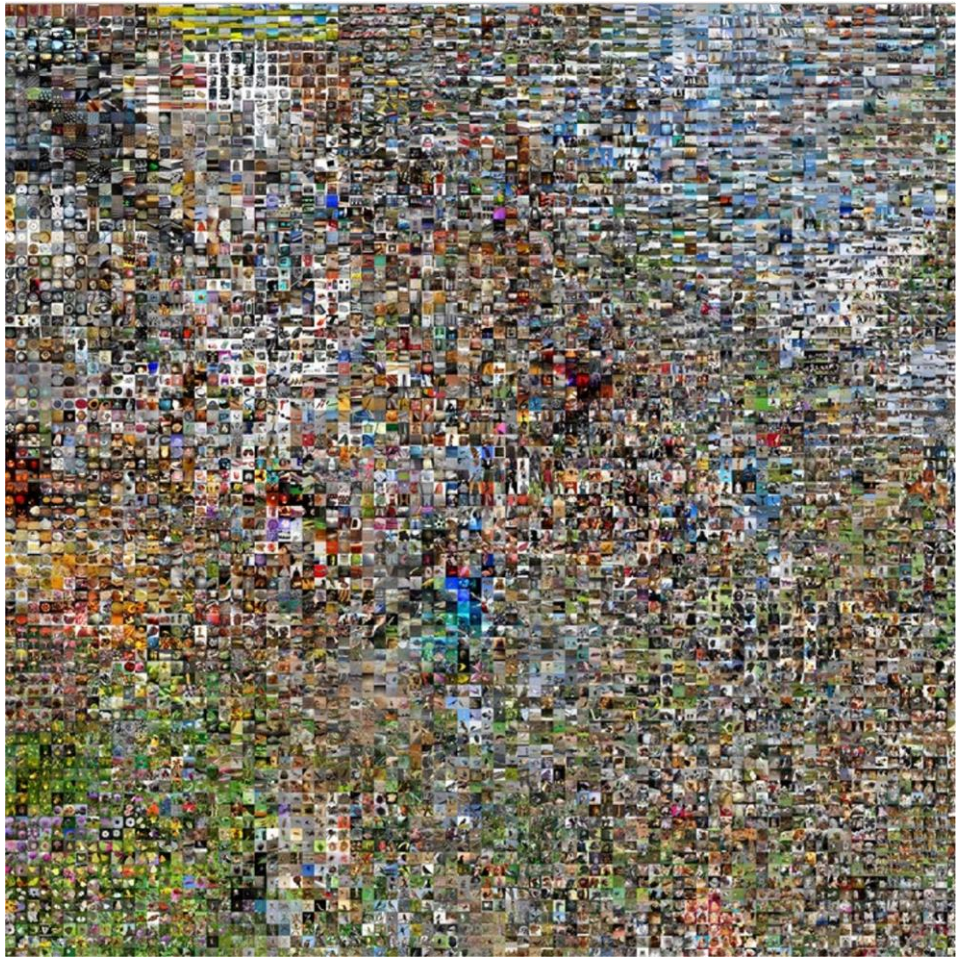




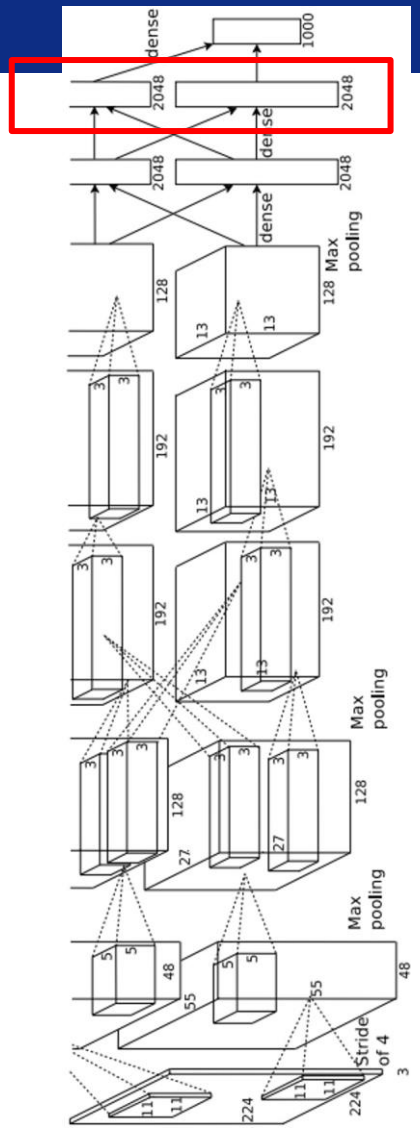
Van der Maaten and Hinton, “Visualizing Data using t-SNE”, JMLR 2008
Figure copyright Laurens van der Maaten and Geoff Hinton, 2008. Reproduced with permission.

Last Layer: Dimensionality Reduction

Van der Maaten and Hinton, “Visualizing Data using t-SNE”, JMLR 2008
Krizhevsky et al, “ImageNet Classification with Deep Convolutional Neural Networks”, NIPS 2012. Figure reproduced with permission.

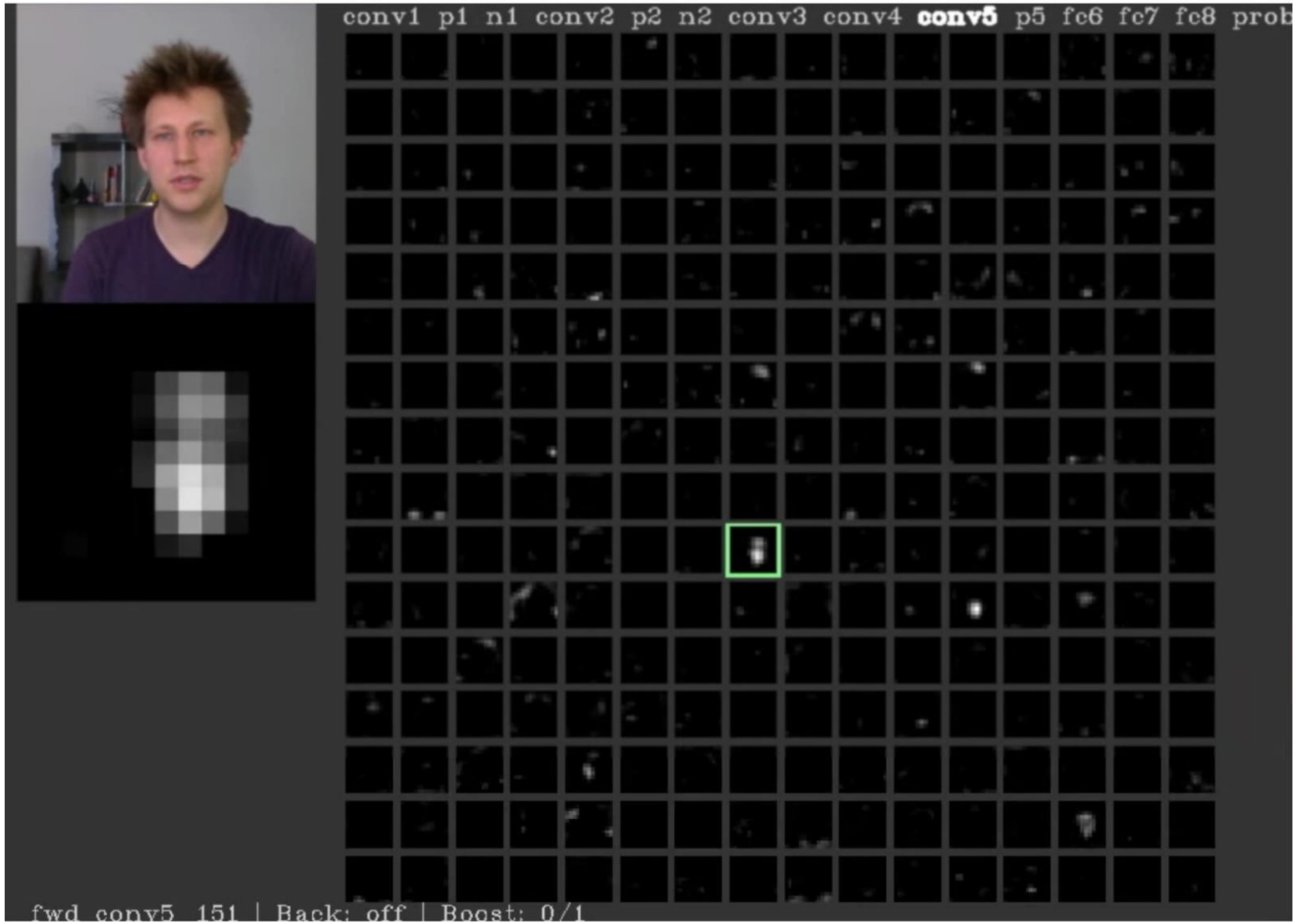


See high-resolution versions at
http://cs.stanford.edu/people/karpathy/cnnembed/



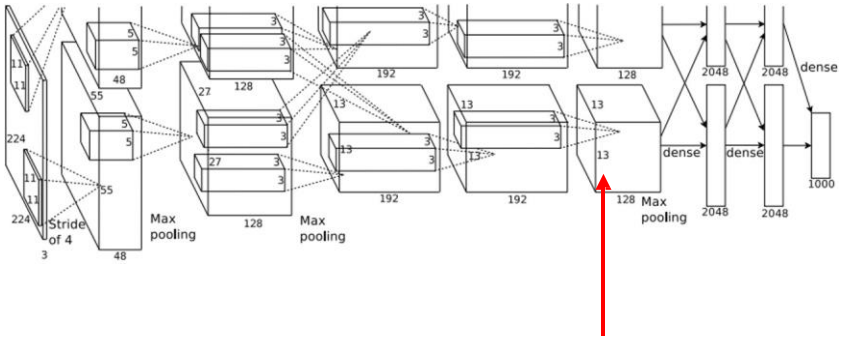
Visualizing Activations

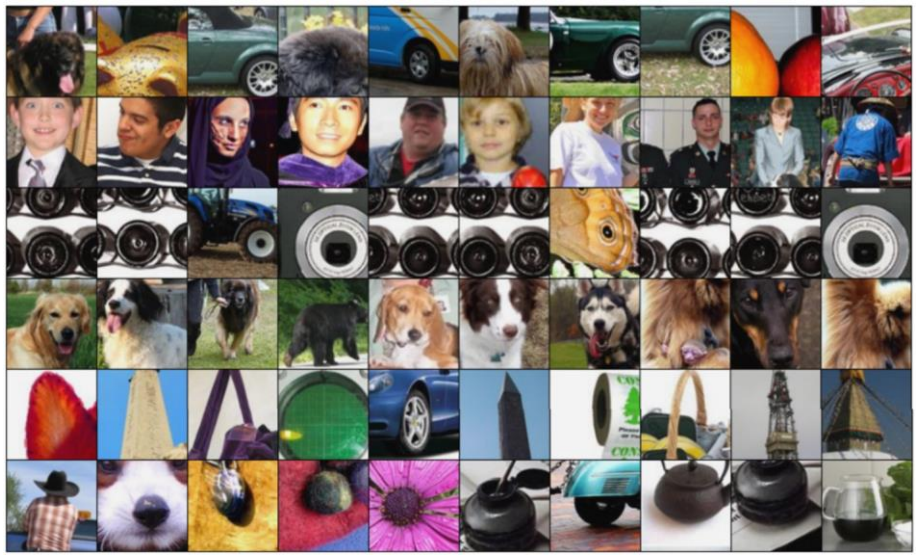
conv5 feature map is 128x13x13; visualize as 128 13x13 grayscale images



Yosinski et al, "Understanding Neural Networks Through Deep Visualization", ICML DL Workshop 2014. Figure copyright Jason Yosinski, 2014. Reproduced with permission.

Maximally Activating Patches





Springenberg et al, “Striving for Simplicity: The All Convolutional Net”, ICLR Workshop 2015

Pick a layer and a channel; e.g. conv5 is 128 x 13 x 13, pick channel 17/128

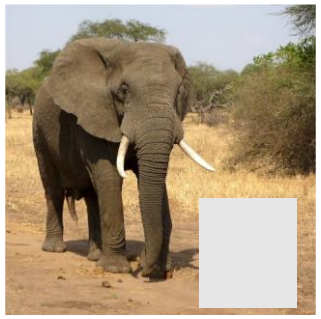
Run many images through the network, record values of chosen channel

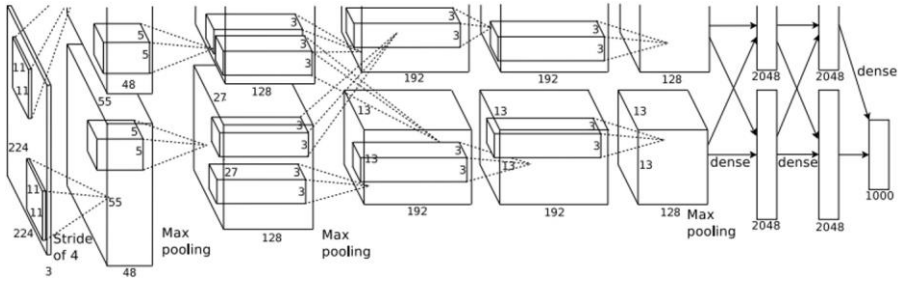
Visualize image patches that correspond to maximal activations

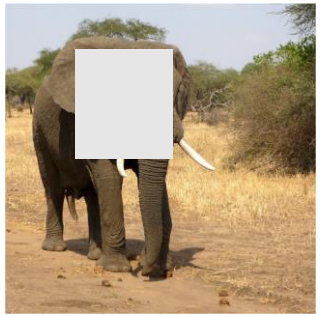
Which pixels matter?

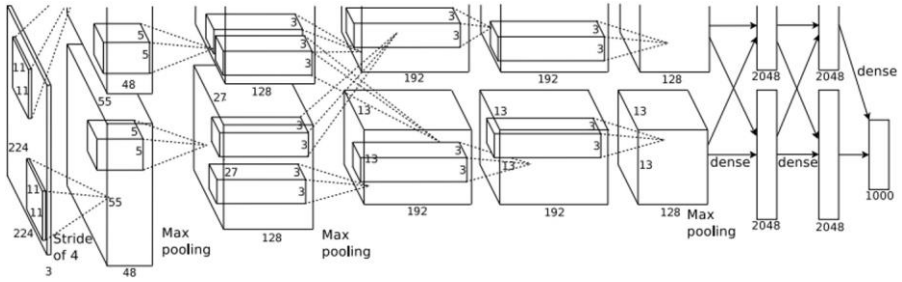
Saliency via Occlusion

Mask part of the image before feeding to CNN, check how much predicted probabilities change

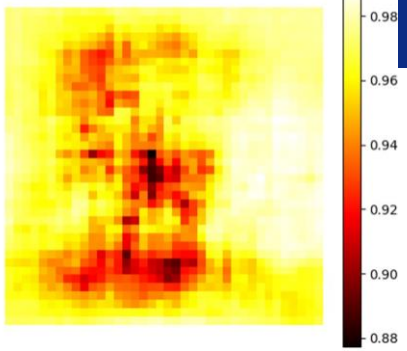




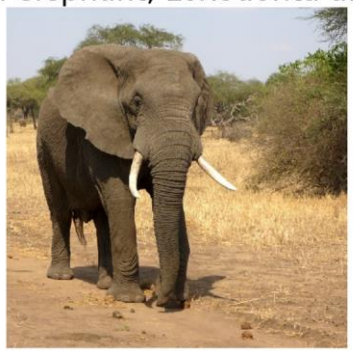


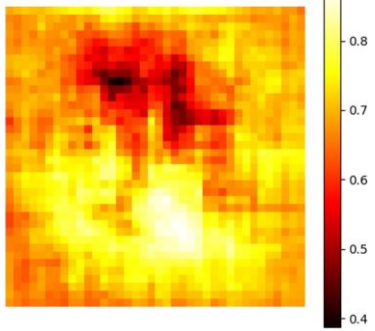


schooner

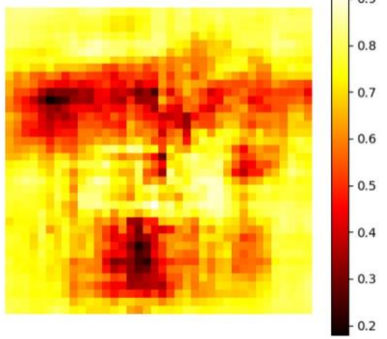


African elephant, Loxodonta africana





go-kart

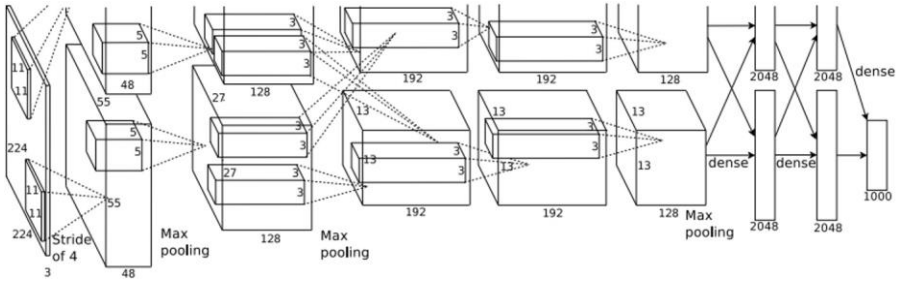


Boat image is CC0 public domain
Elephant image is CC0 public domain
Go-Karts image is CC0 public domain

Zeiler and Fergus, "Visualizing and Understanding Convolutional Networks", ECCV 2014

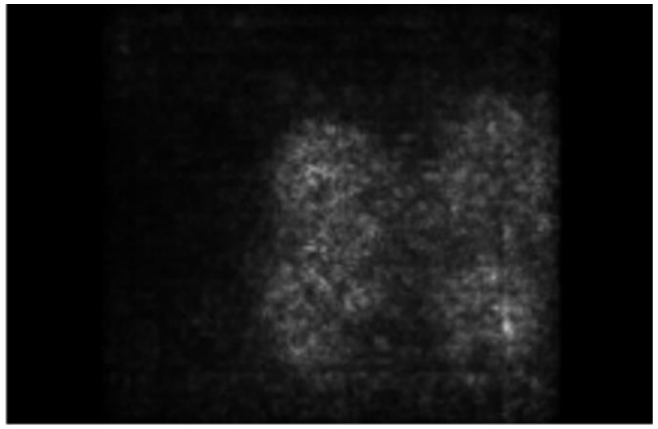
Which pixels matter? Saliency via Backprop

Forward pass: Compute probabilities



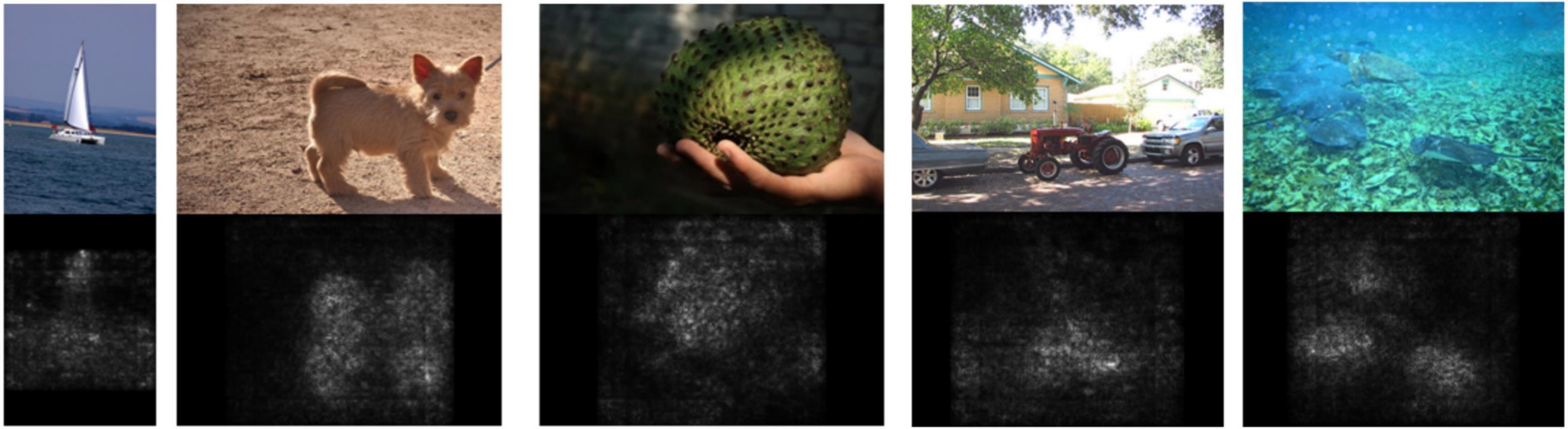
Dog

Compute gradient of (unnormalized) class score with respect to image pixels, take absolute value and max over RGB channels



Simonyan, Vedaldi, and Zisserman, “Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps”, ICLR Workshop 2014

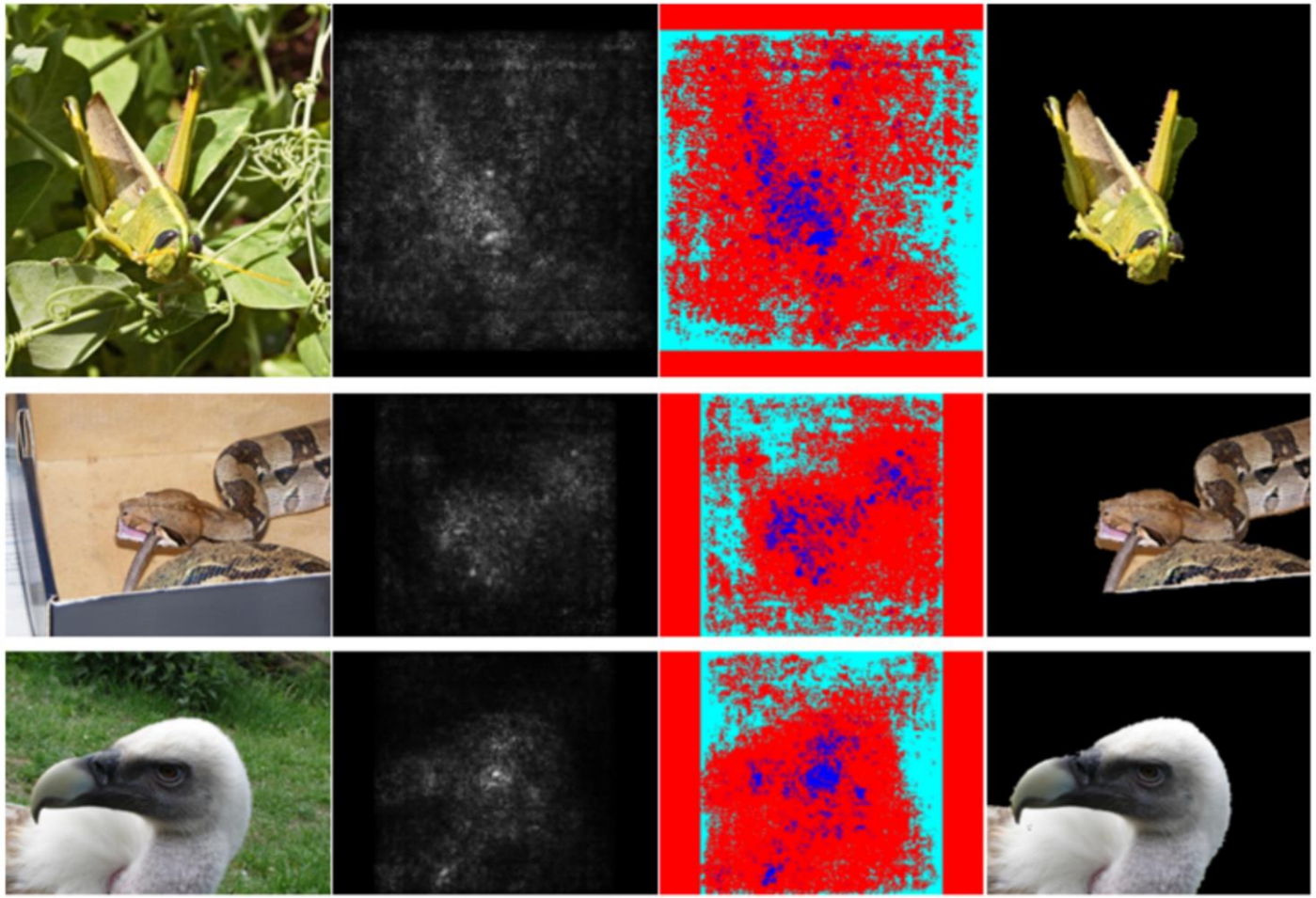
Which pixels matter? Saliency via Backprop



Simonyan, Vedaldi, and Zisserman, “Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps”, ICLR Workshop 2014

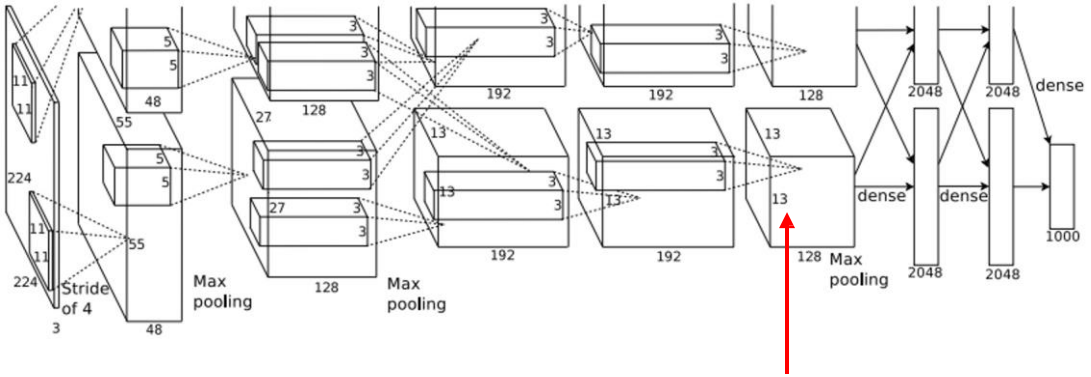
Saliency Maps: Segmentation without Supervision

Use GrabCut on saliency map



Simonyan, Vedaldi, and Zisserman, "Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps", ICLR Workshop 2014.
Rother et al, "Grabcut: Interactive foreground extraction using iterated graph cuts", ACM TOG 2004.

Intermediate Features via backprop



Pick a single intermediate neuron, e.g. one value in 128 x 13 x 13 conv5 feature map

Compute gradient of neuron value with respect to image pixels

Zeiler and Fergus, “Visualizing and Understanding Convolutional Networks”, ECCV 2014
Springenberg et al, “Striving for Simplicity: The All Convolutional Net”, ICLR Workshop 2015

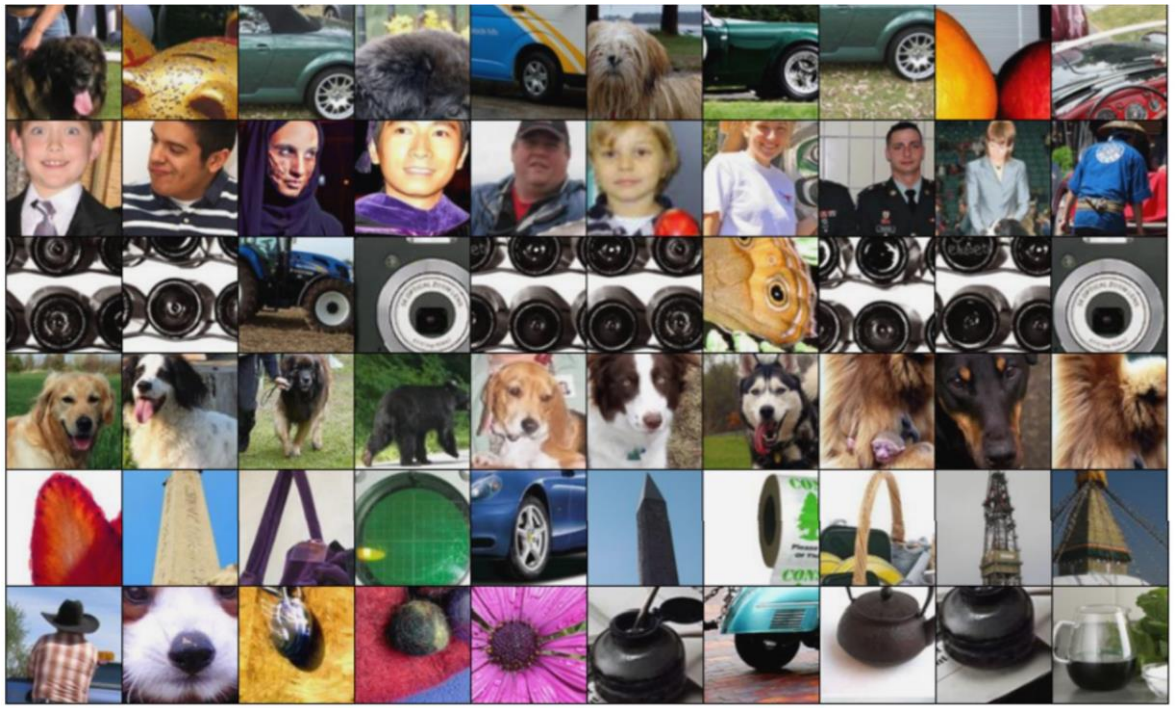
Intermediate Features via backprop

Maximally activating patches
(Each row is a different neuron)

Guided Backprop

Zeiler and Fergus, "Visualizing and Understanding Convolutional Networks", ECCV 2014
Springenberg et al, "Striving for Simplicity: The All Convolutional Net", ICLR Workshop 2015

Intermediate Features via backprop

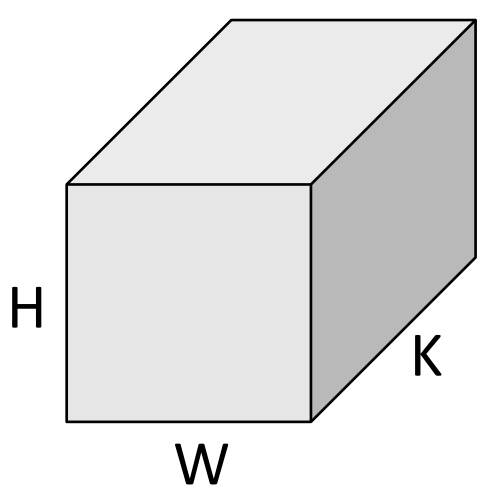


Maximally activating patches
(Each row is a different neuron)

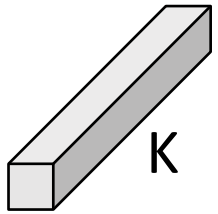
Guided Backprop

Zeiler and Fergus, "Visualizing and Understanding Convolutional Networks", ECCV 2014
Springenberg et al, "Striving for Simplicity: The All Convolutional Net", ICLR Workshop 2015

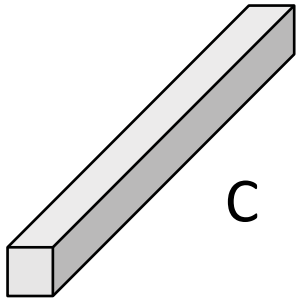
Class Activation Mapping (CAM)



Global Average Pooling



Fully Connected Layer, weights $w \in \mathbb{R}^{K \times C}$



Last layer CNN features:
$f \in \mathbb{R}^{H \times W \times K}$

Pooled features:
$F \in \mathbb{R}^{K}$

Class Scores:
$S \in \mathbb{R}^{C}$

$$F_k = \frac{1}{HW} \sum_{h,w} f_{h,w,k} \qquad S_c = \sum_k w_{k,c} F_k = \frac{1}{HW} \sum_k w_{k,c} \sum_{h,w} f_{h,w,k}$$

$$= \frac{1}{HW} \sum_{h,w} \sum_k w_{k,c} f_{h,w,k}$$

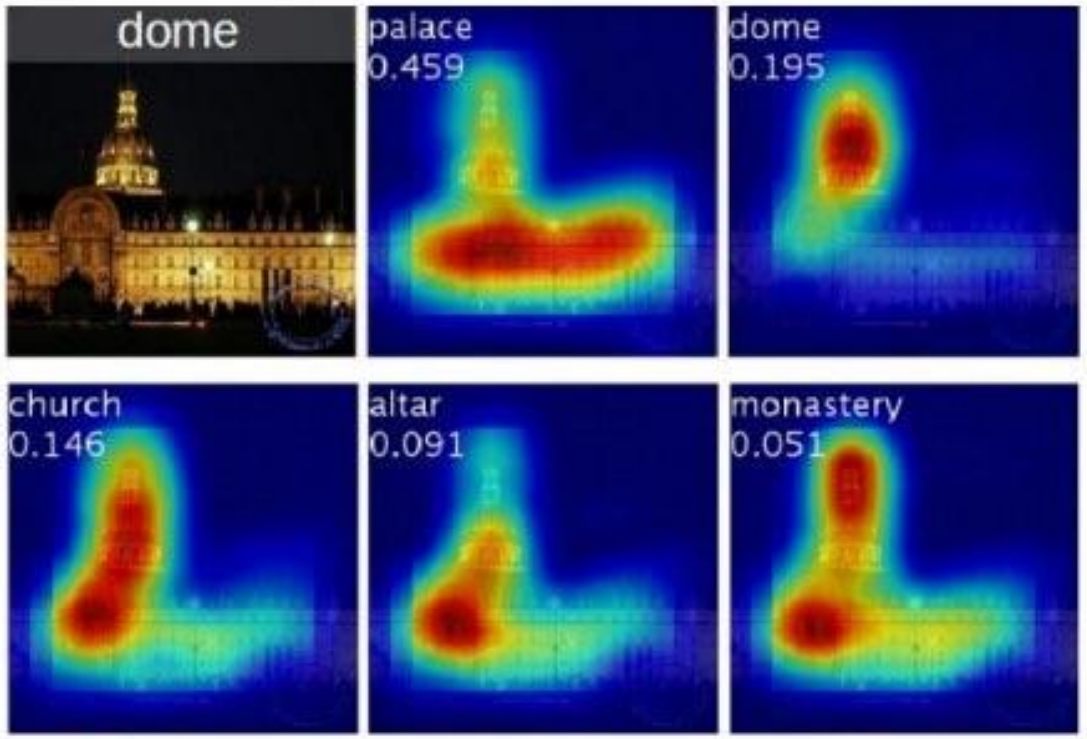
Class Activation Maps:
$M \in \mathbb{R}^{C,H,W}$

$$M_{c,h,w} = \sum_k w_{k,c} f_{h,w,k}$$

Zhou et al, "Learning Deep Features for Discriminative Localization", CVPR 2016

Class Activation Mapping (CAM)

Problem 1: Can only apply to last conv layer
Problem 2: Can only work with GAP



Class activation maps of top 5 predictions

Class activation maps for one object class

Zhou et al, “Learning Deep Features for Discriminative Localization”, CVPR 2016

Gradient-Weighted Class Activation Mapping (Grad-CAM)

1. Pick any layer, with activations $A \in \mathbb{R}^{H \times W \times K}$
2. Compute gradient of class score S_c with respect to A:

$$\frac{\partial S_c}{\partial A} \in \mathbb{R}^{H \times W \times K}$$

3. Global Average Pool the gradients to get weights $\alpha \in \mathbb{R}^{K}$:

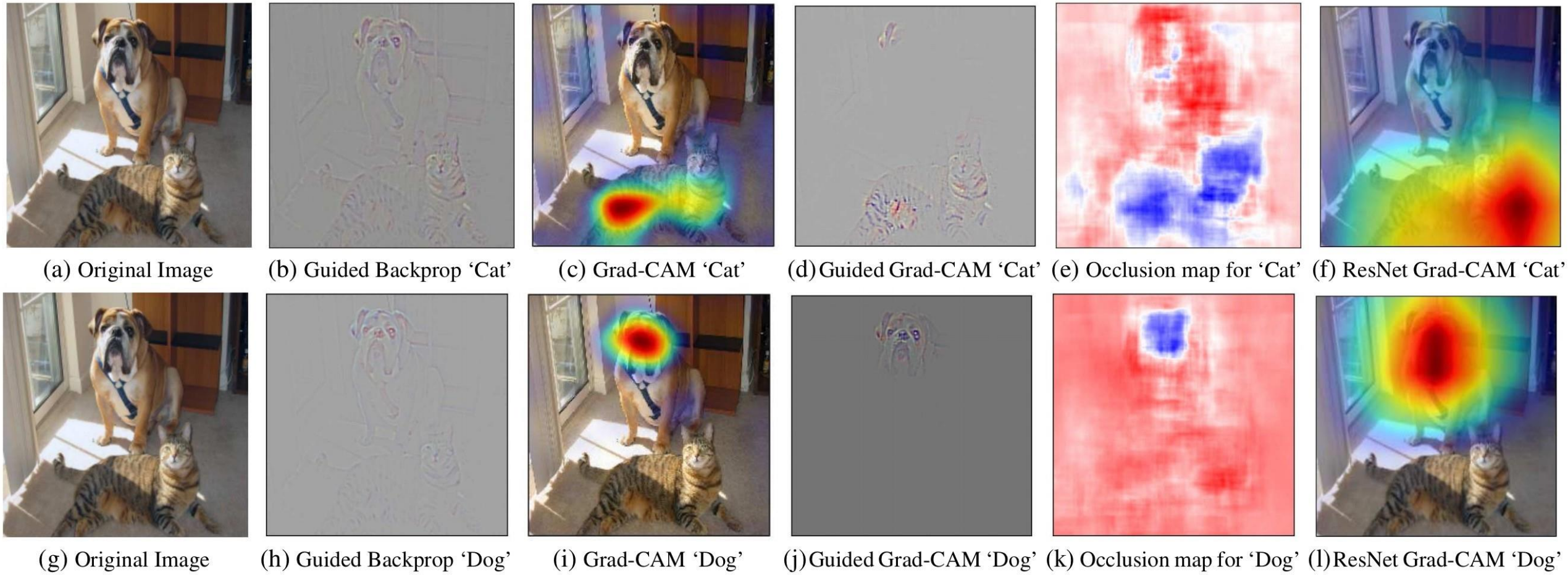
$$\alpha_k = \frac{1}{HW} \sum_{h,w} \frac{\partial S_c}{\partial A_{h,w,k}}$$

4. Compute activation map $M^c \in \mathbb{R}^{H,W}$:

$$M^c_{h,w} = ReLU\left(\sum_k \alpha_k A_{h,w,k}\right)$$

Selvaraju et al, “Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization”, CVPR 2017

Gradient-Weighted Class Activation Mapping (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'

Selvaraju et al, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization", CVPR 2017

Gradient-Weighted Class Activation Mapping (Grad-CAM)

Can also be applied beyond classification models, e.g. image captioning

Selvaraju et al, “Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization”, CVPR 2017

Weekly Supervised Semantic Segmentation

Label Annotation Cost

Cost Expensive for Semantic Segmentation Label Annotation

78 x cost of pixel level than Image level

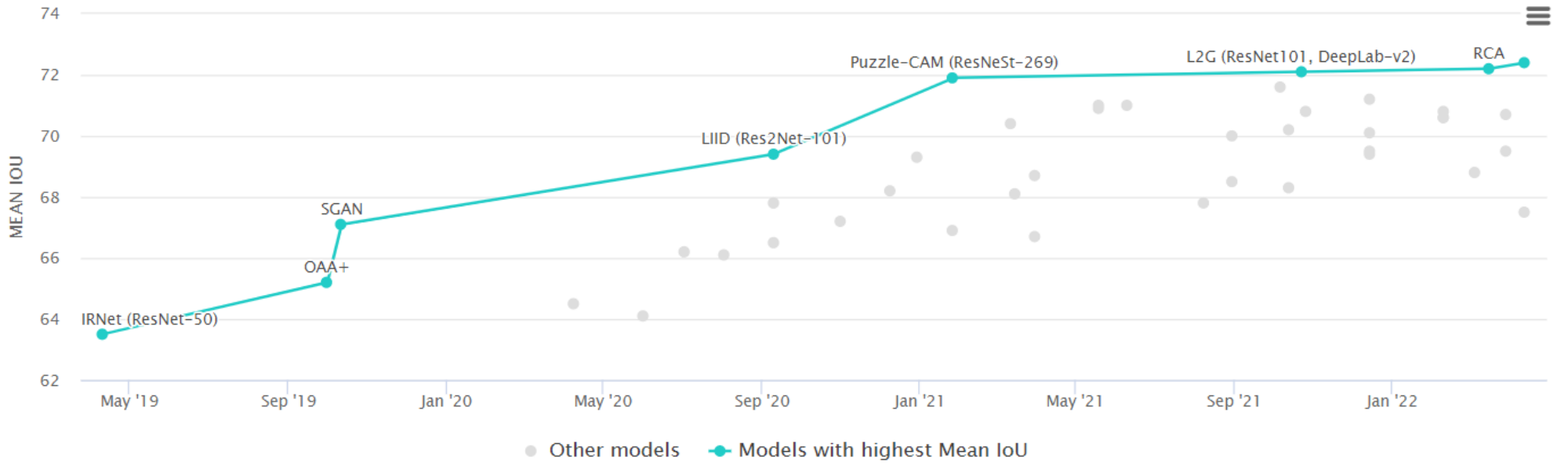
출처 : Weakly Supervised Learning for Computer Vision, CVPR18 Tutorial

Weekly Supervised Semantic Segmentation

Weakly-Supervised Semantic Segmentation on PASCAL VOC 2012 val

Leaderboard Dataset

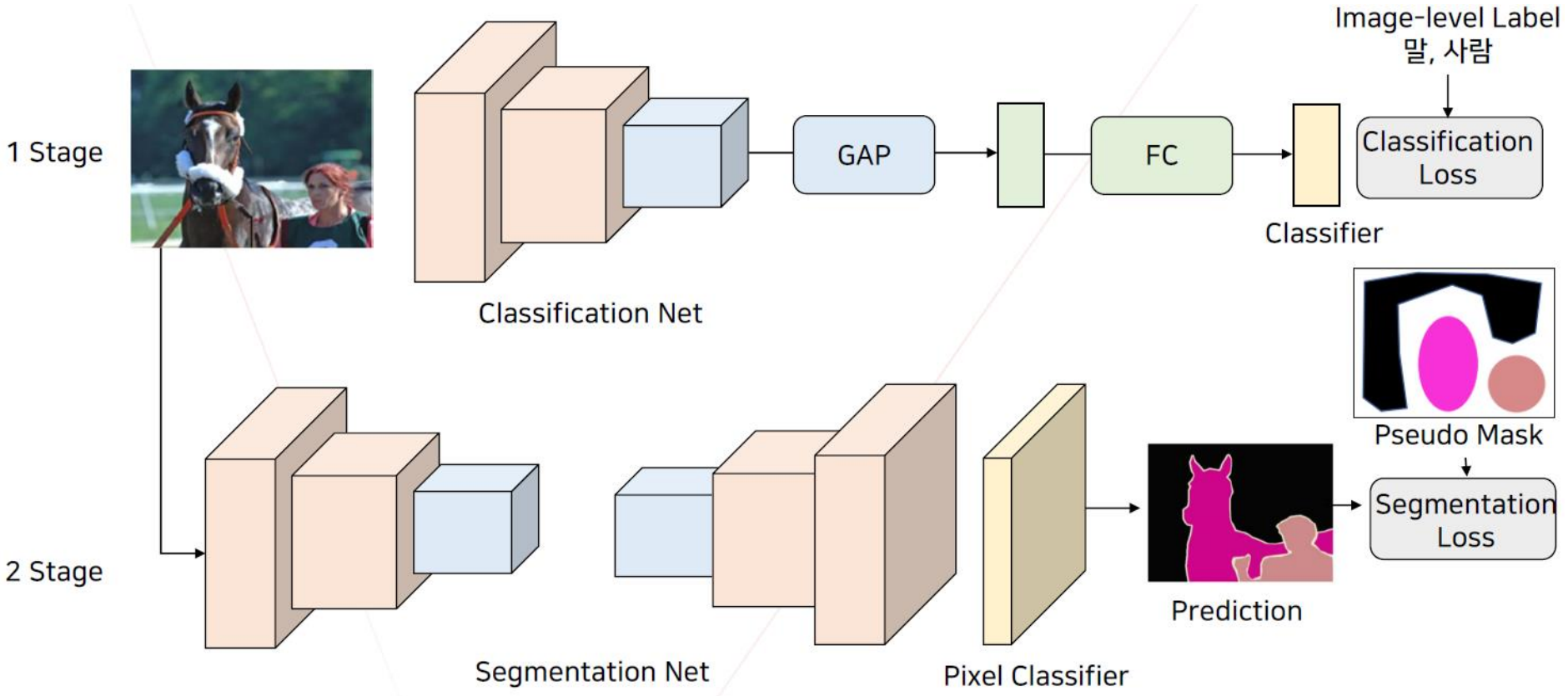
View Mean IoU by Date for All models



Weekly Supervised Semantic Segmentation

Step 1. CAM Extraction from Classifier

Step 2. Trained Segmentation model by using Pseudo Pixel Label on CAM extracted from Step 1.



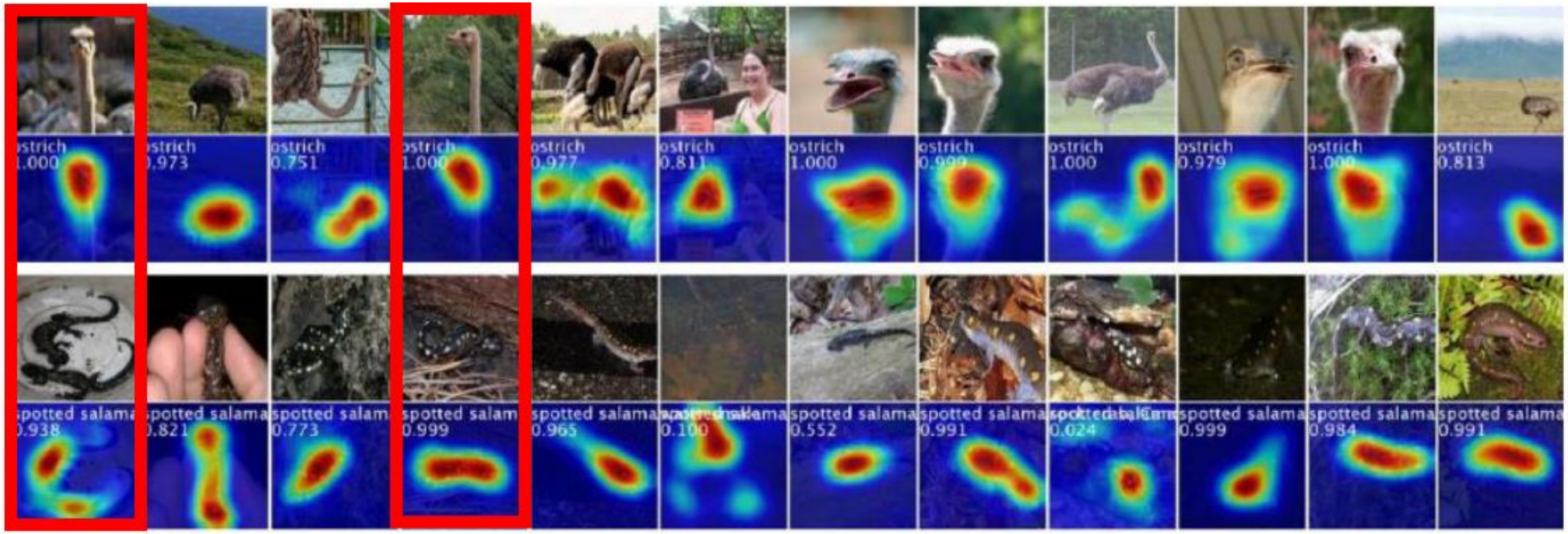
출처 : 김현우 Weekly Supervised Semantic Segmentation

Weekly Supervised Semantic Segmentation

Limitation

1. Poor & Unsharp Result of CAM
2. CAM results are concentrated only in the discriminative area
3. CAM not calculated for background

Problem of CAM Guides: Bigger than the object, blob shapes

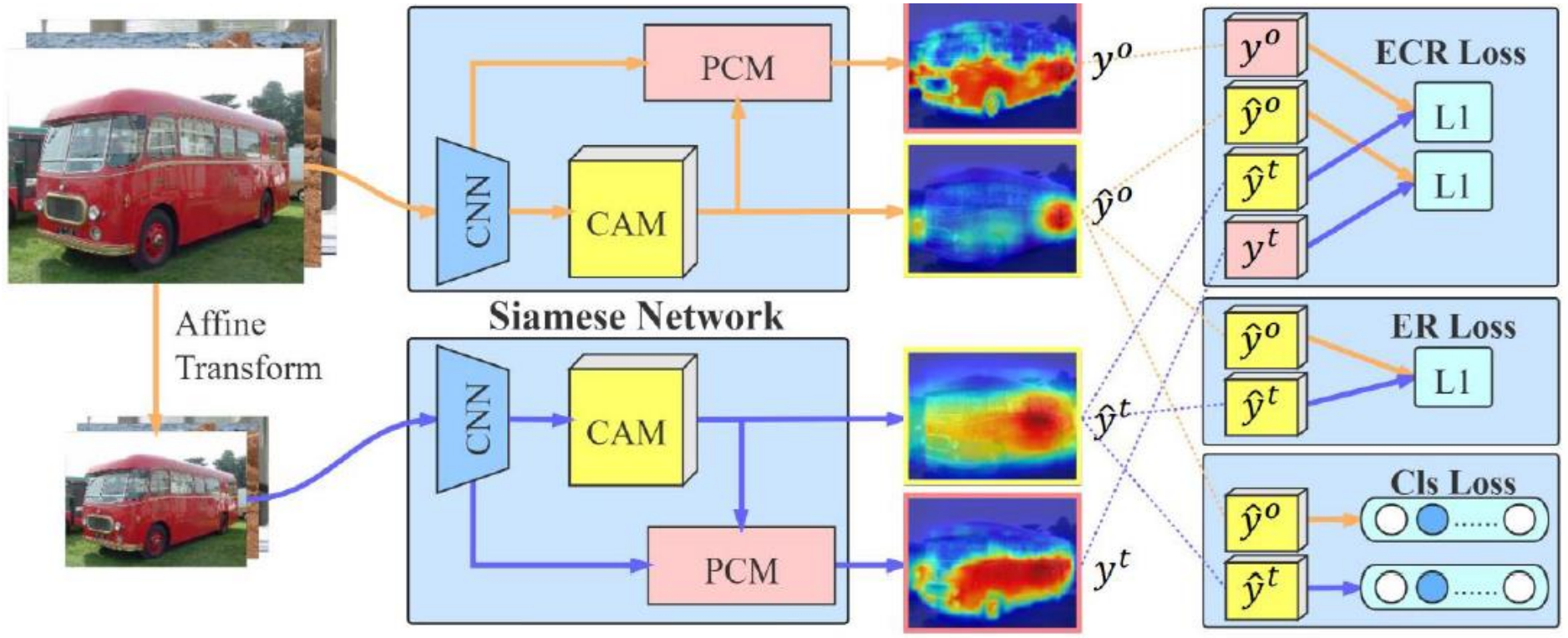


출처 : Weakly Supervised Learning for Computer Vision, CVPR18 Tutorial

Weekly Supervised Semantic Segmentation

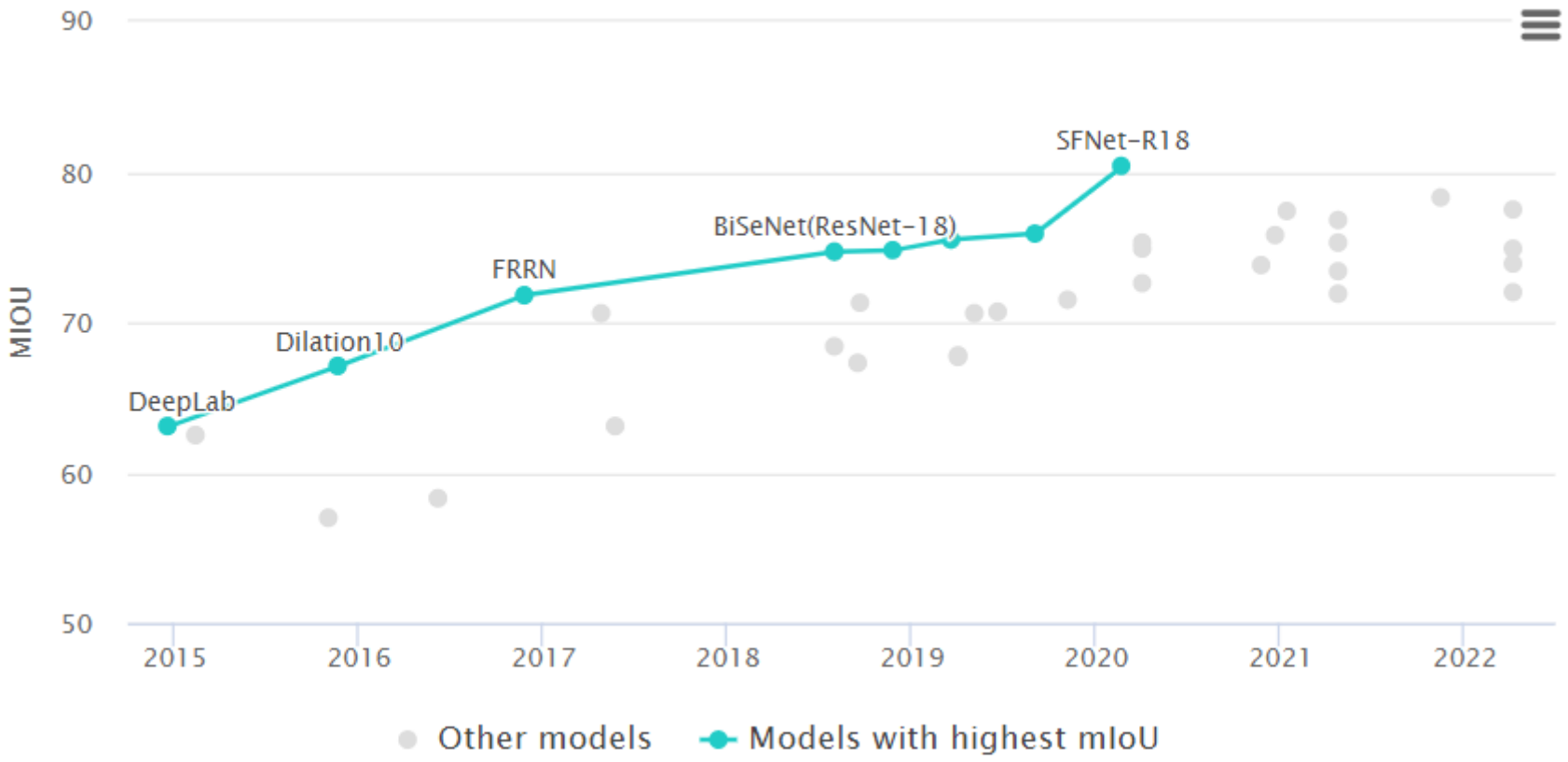
Solutions

1. Enhancing Boundaries
2. Use Saliency as well
3. Self-supervised learning



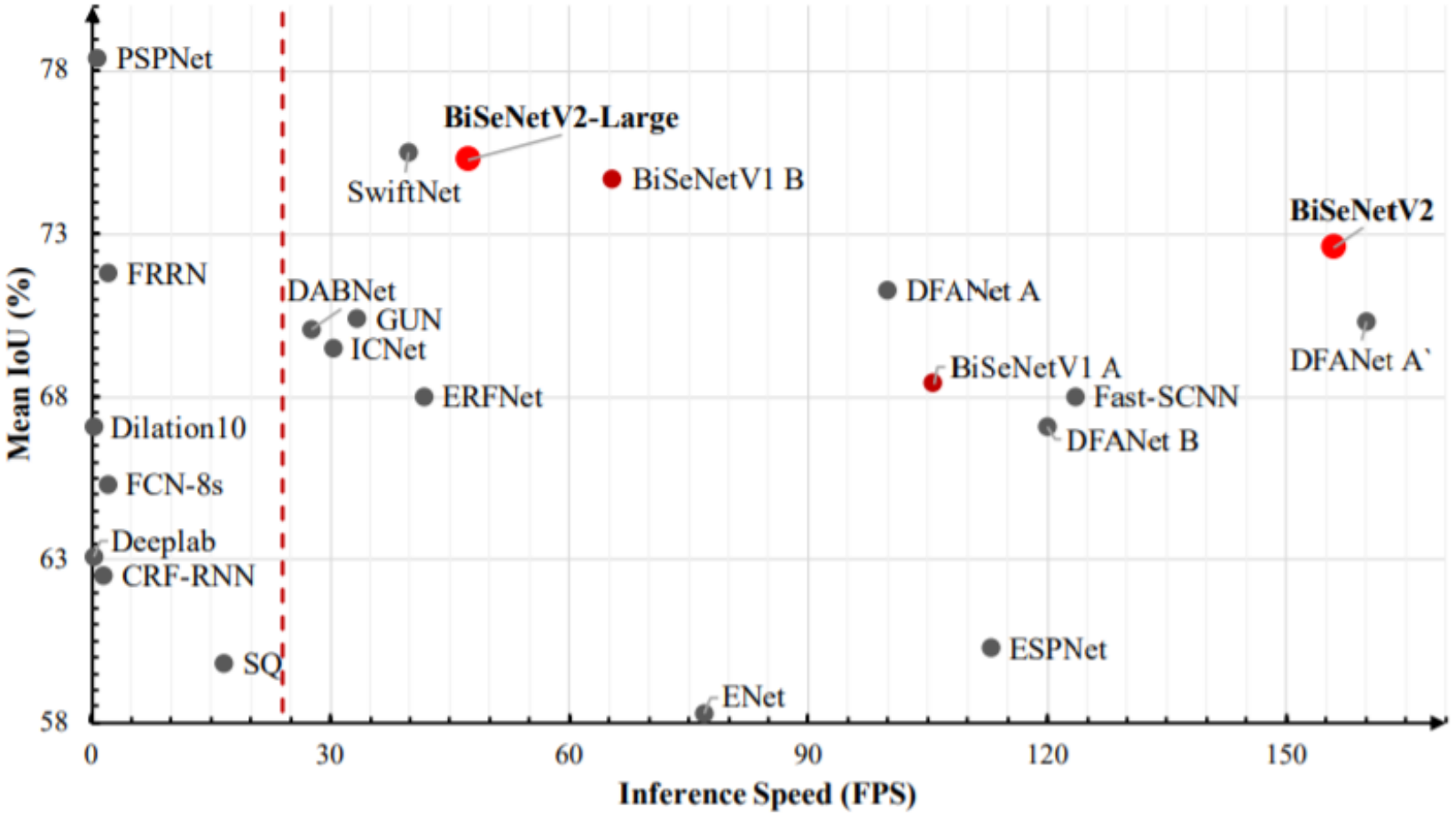
출처 : "Self-supervised Equivariant Attention Mechanism for Weakly Supervised Semantic Segmentation", CVPR 2020

Real Time Semantic Segmentation



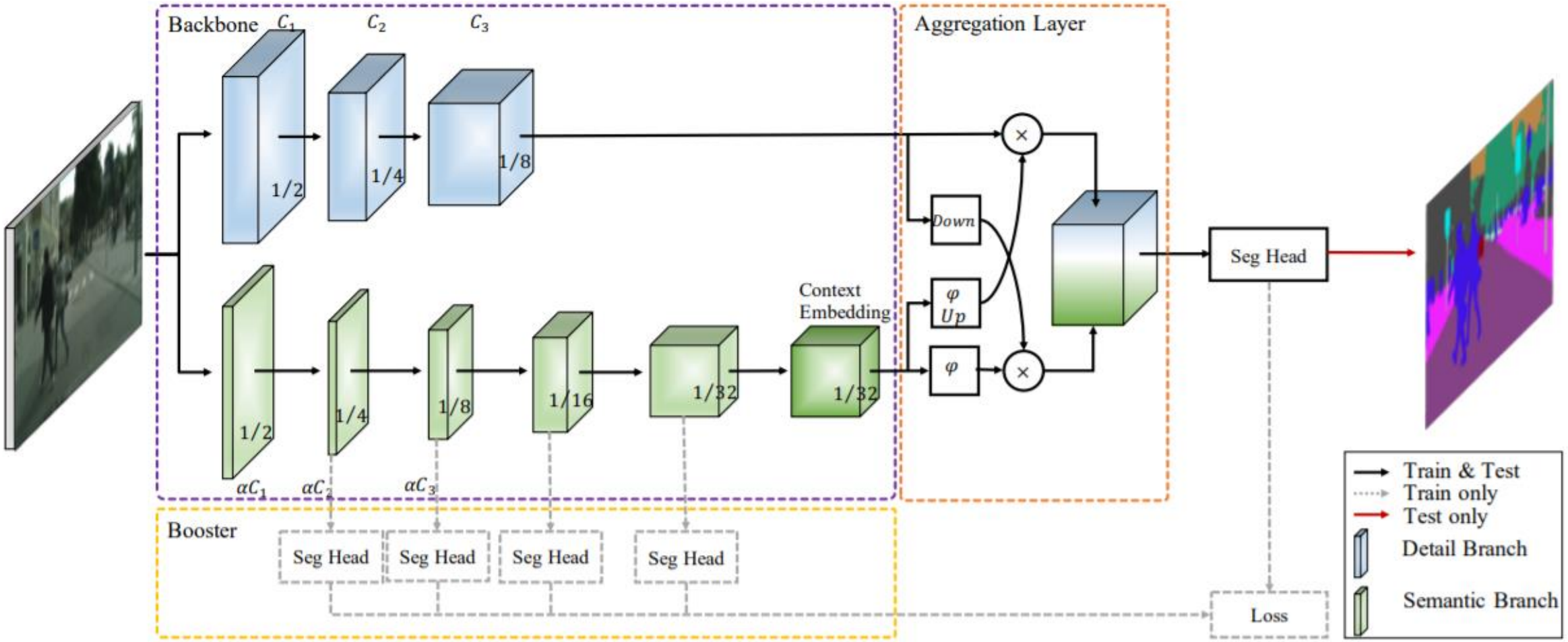
출처 : https://paperswithcode.com/sota/real-time-semantic-segmentation-on-cityscapes

Real Time Semantic Segmentation



출처 : Bilateral Network with Guided Aggregation for Real-time Semantic Segmentation, 2020

Real Time Semantic Segmentation



출처 : Bilateral Network with Guided Aggregation for Real-time Semantic Segmentation, 2020

Today's Summary

Lecture 14: Visualization & Understanding

- Visualize Filters
- Visualize Activations
- Saliency Maps

+ Weekly Supervised Semantic Segmentation

+ Real Time Semantic Segmentation

Team Meeting

Lecture 14: Visualization & Understanding

Team Meeting

3:45 PM~ 1조 이가은, 김윤진, 송유리, 이선명

3:55 PM~ 2조 박채은, 박현정, 오시은, 김진희

4:05 PM~ 3조 김민지, 이재인, 황예진, 정하늘

4:15 PM~ 4조 이민영, 이은재, 이지나, 이태림

4:25 PM~ 5조 이경진, 구태현, 주성경, 이한희, 최은서