# Fully Convolutional Networks with Spectral Pooling Methods

CS 590 – Research Seminar

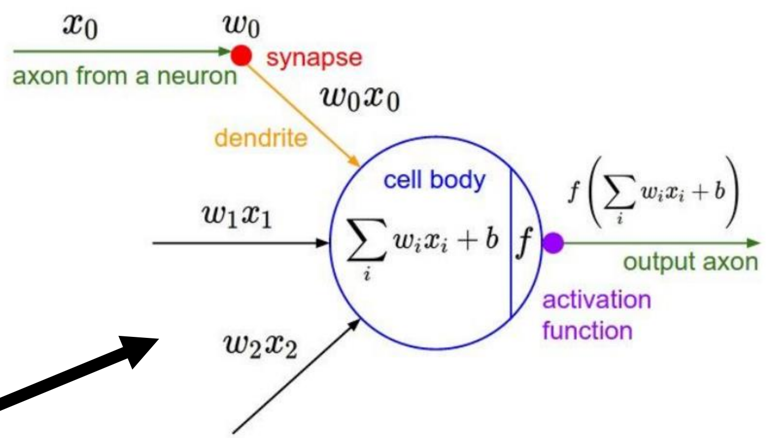
Presenter: Onur Aydın

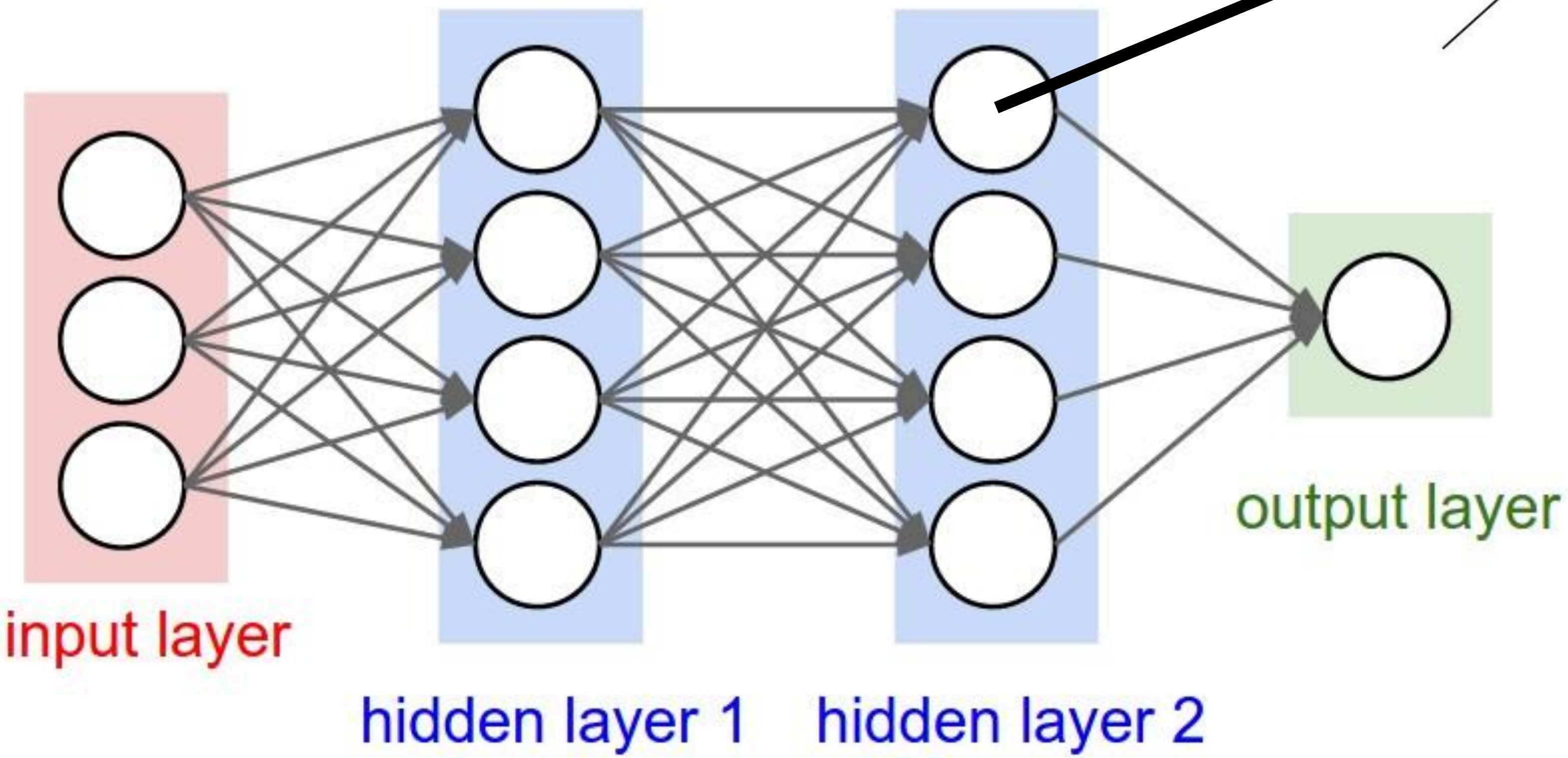
Advisor: Asst. Prof. Dr. R. Gökberk Cinbiş

# Outline


- 1. Introduction
  - 1.1. Neural Networks
  - 1.2. Deep Learning
- 2. Convolutional Neural Networks (CNN)
- 3. Fully Convolutional Networks (FCN)
- 4. Problem Description and Related Work
- 5. Spectral Pooling
- 6. Experiments and Results
- 7. Conclusion

# 1.1 Neural Networks



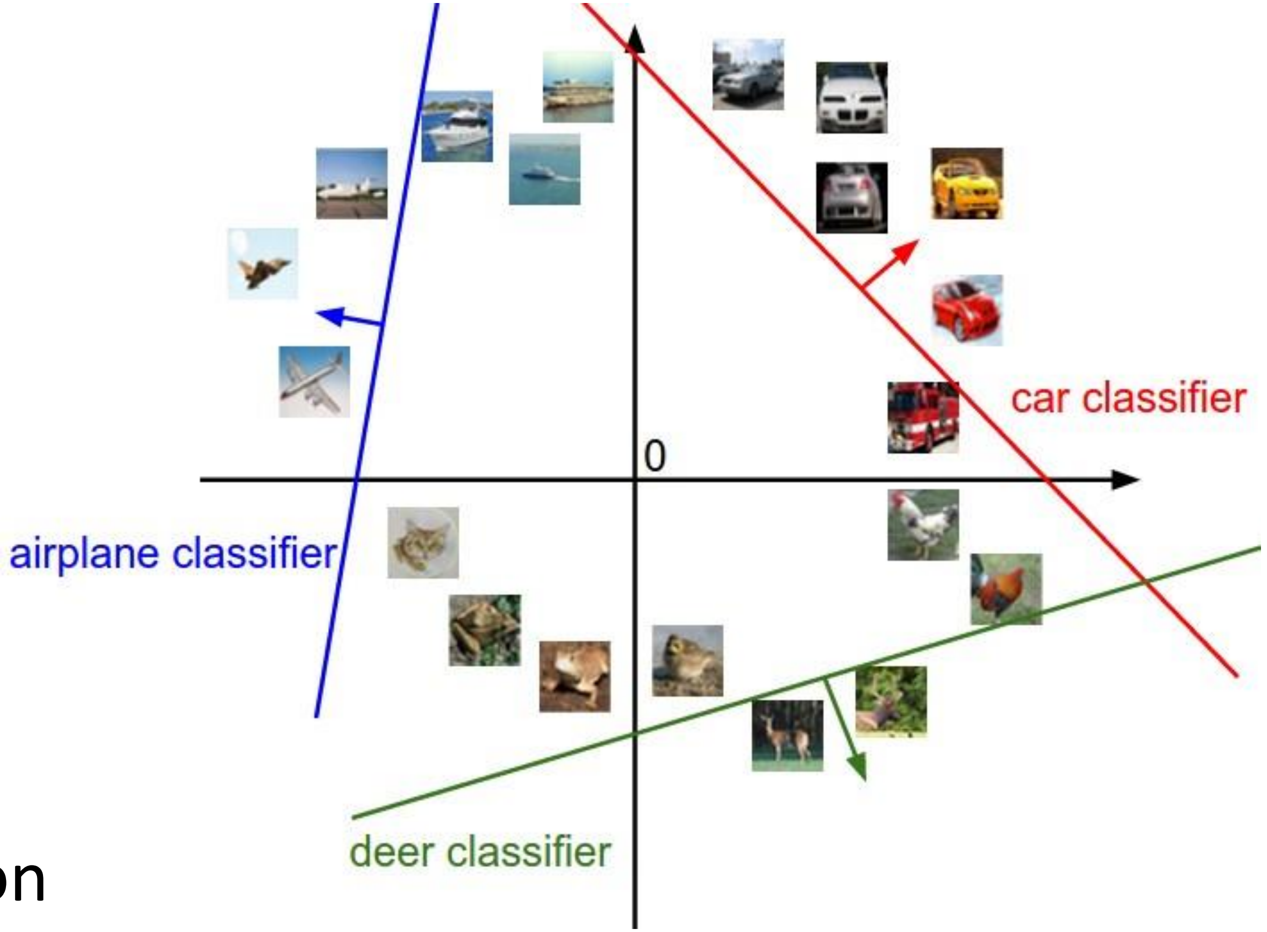




$$L(\theta, \hat{\theta})$$

$$\frac{\partial L}{\partial w}$$

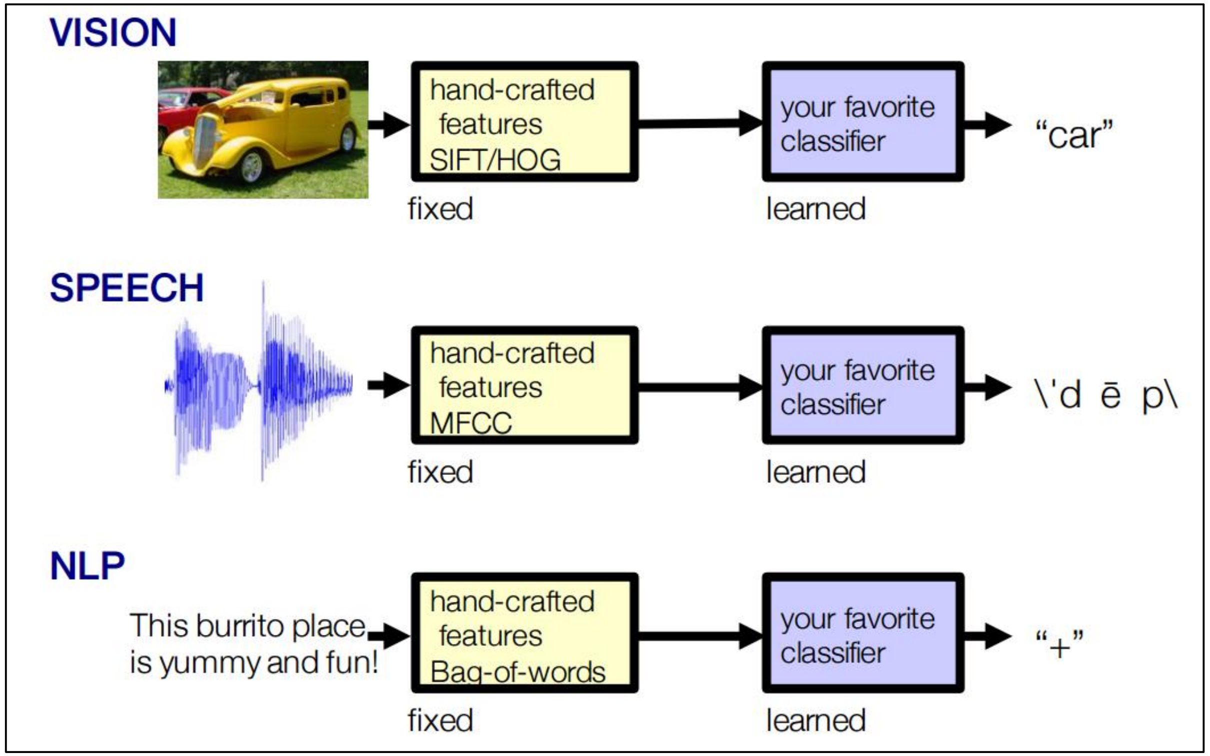
# 1.1 Neural Networks

- Neural Networks are good at:
  - Classification
  - Regression
  - Function Approximation

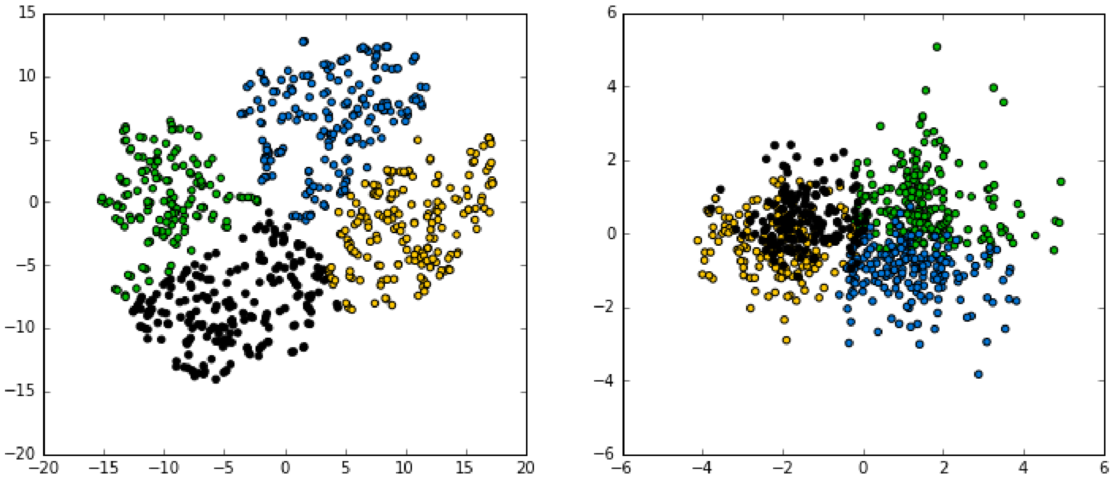



Better Classification ⟵ Better Feature Extraction

# 1.1 Neural Networks



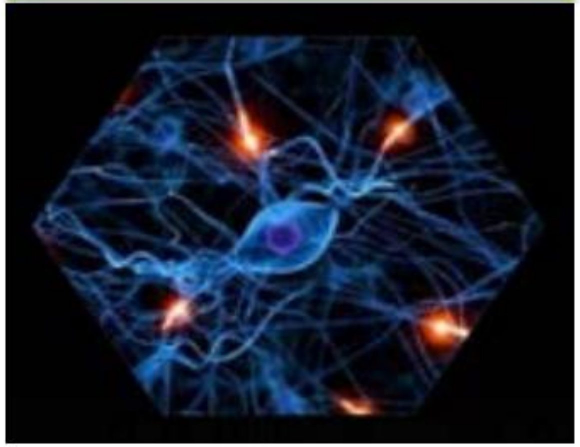

*Traditional Machine Learning*

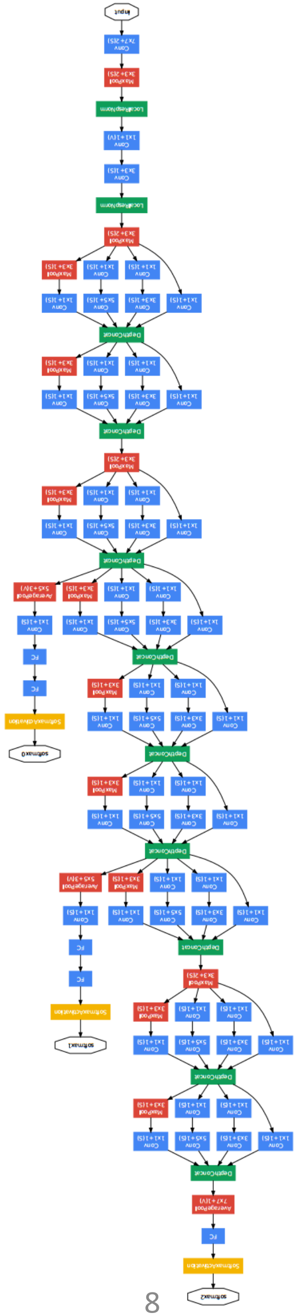


Better Classification ← Better Feature Extraction

# 1.2. Deep Learning

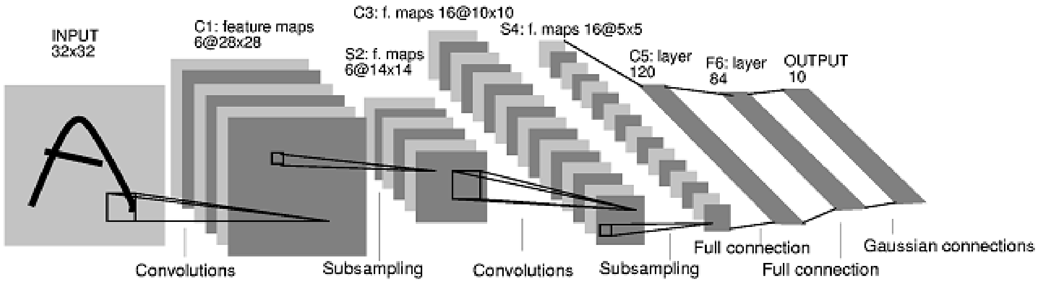
- *'… To Learn representations of data with multiple levels of abstraction.'*

| Big Data Availability | New DL Techniques | GPU acceleration |
|---|---|---|
| 350 millions images uploaded per day<br>2.5 Petabytes of customer data hourly<br>100 hours of video uploaded every minute |  |  |

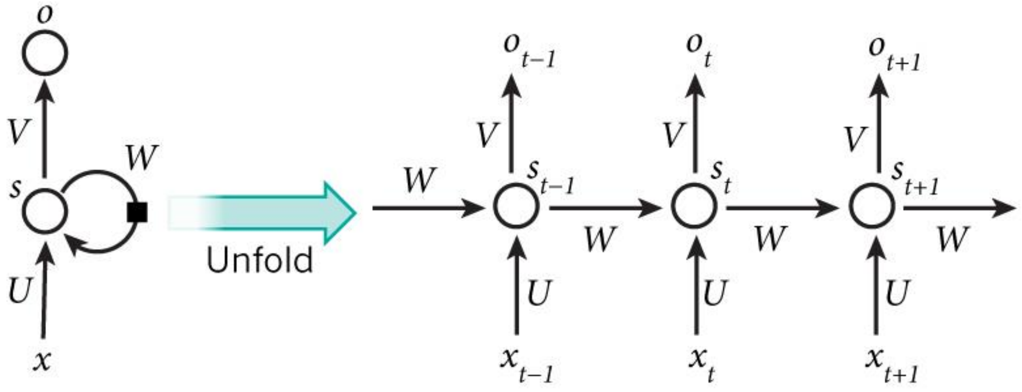


# 1.2. Deep Learning

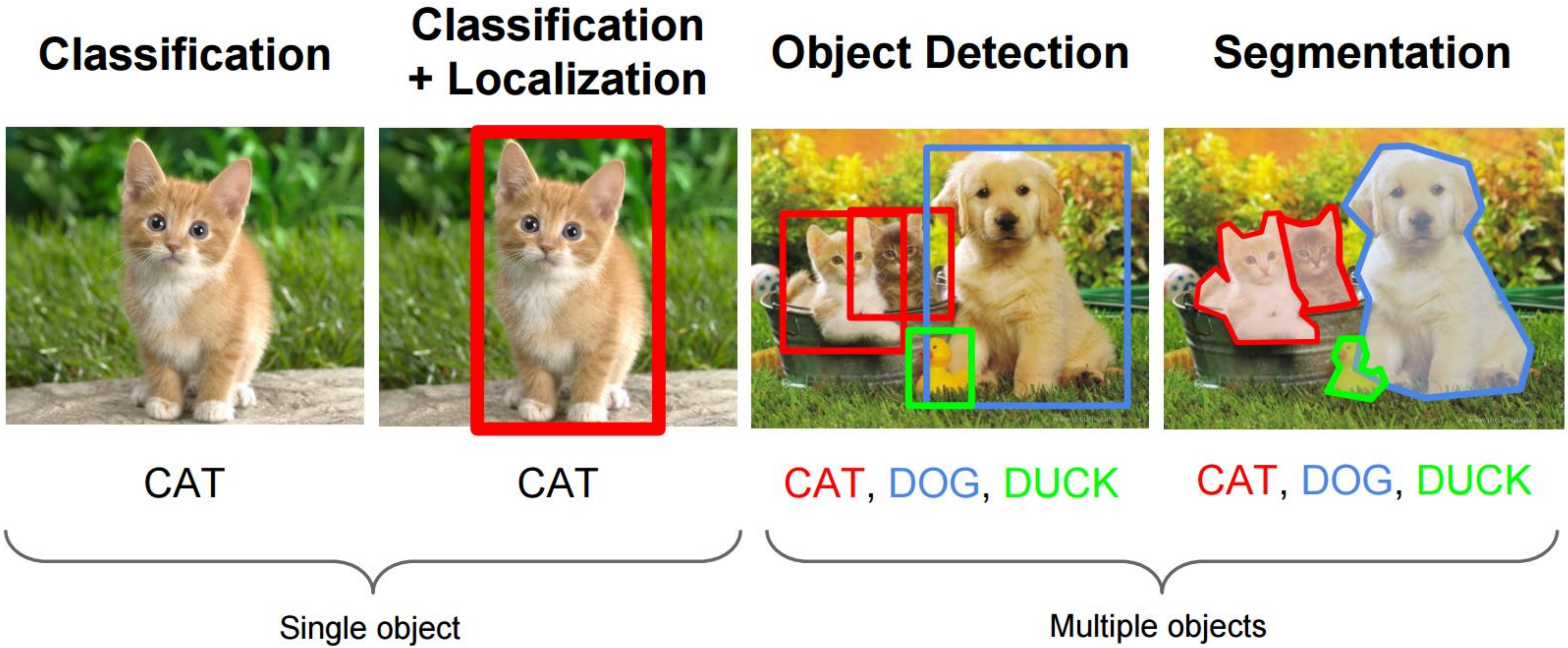
- Convolutional Neural Networks (CNN)
  - Image, video, speech and audio



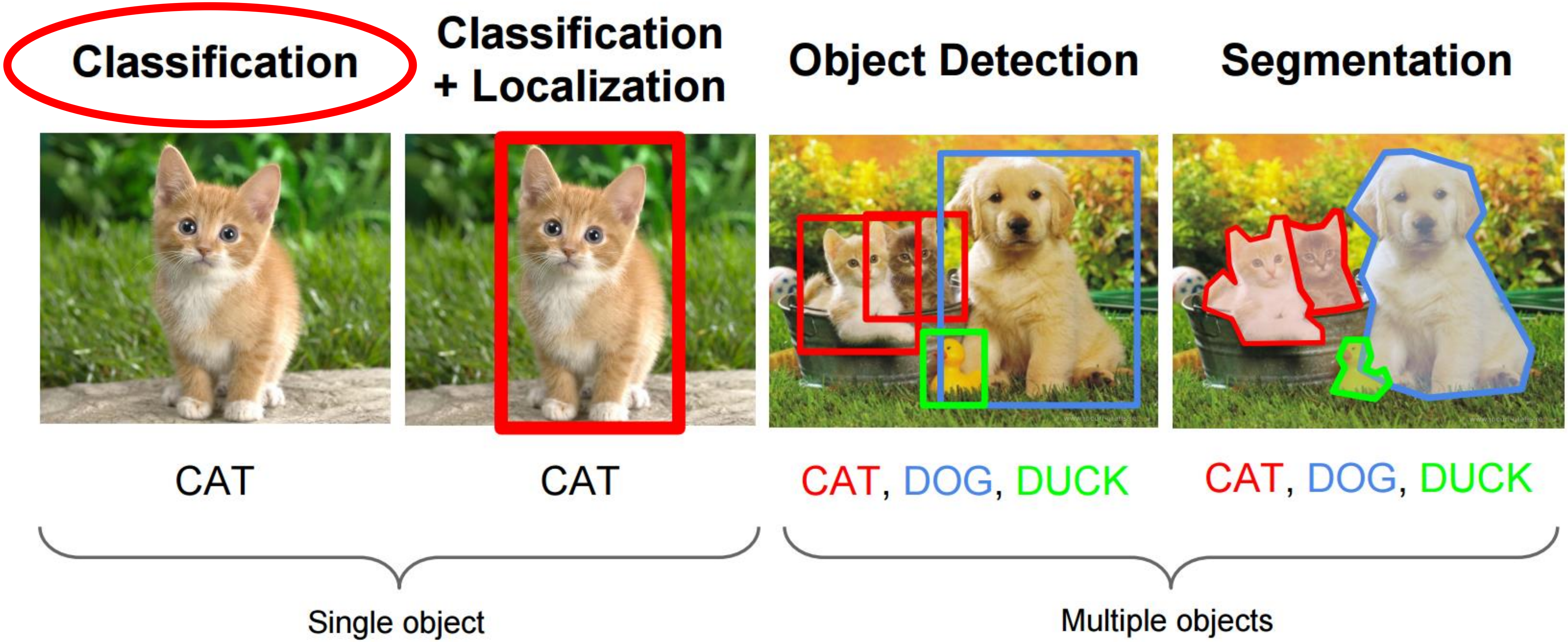

- Recurrent Neural Networks (RNN)
  - Text, speech and time series



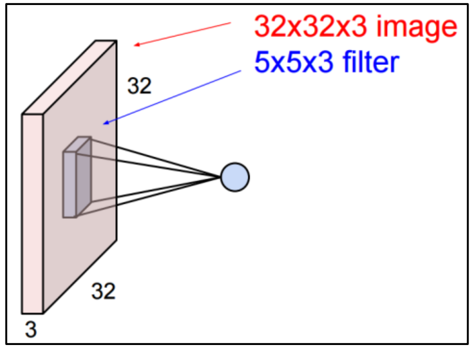
# 1.2. Deep Learning – Computer Vision



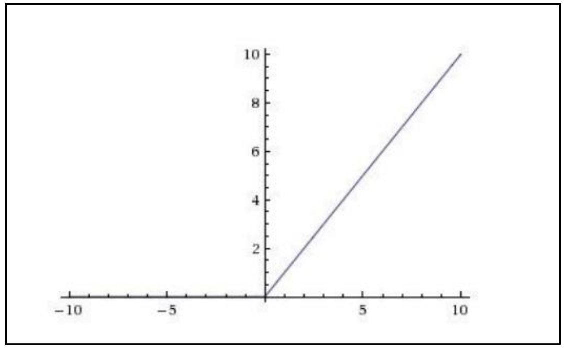
# 2. Convolutional Neural Networks



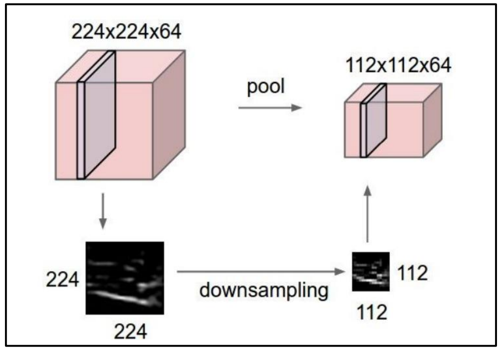
# 2. Convolutional Neural Networks



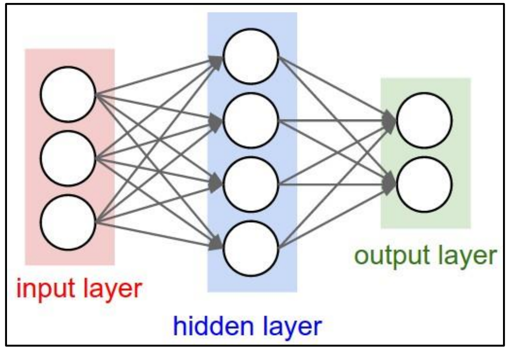

CONV



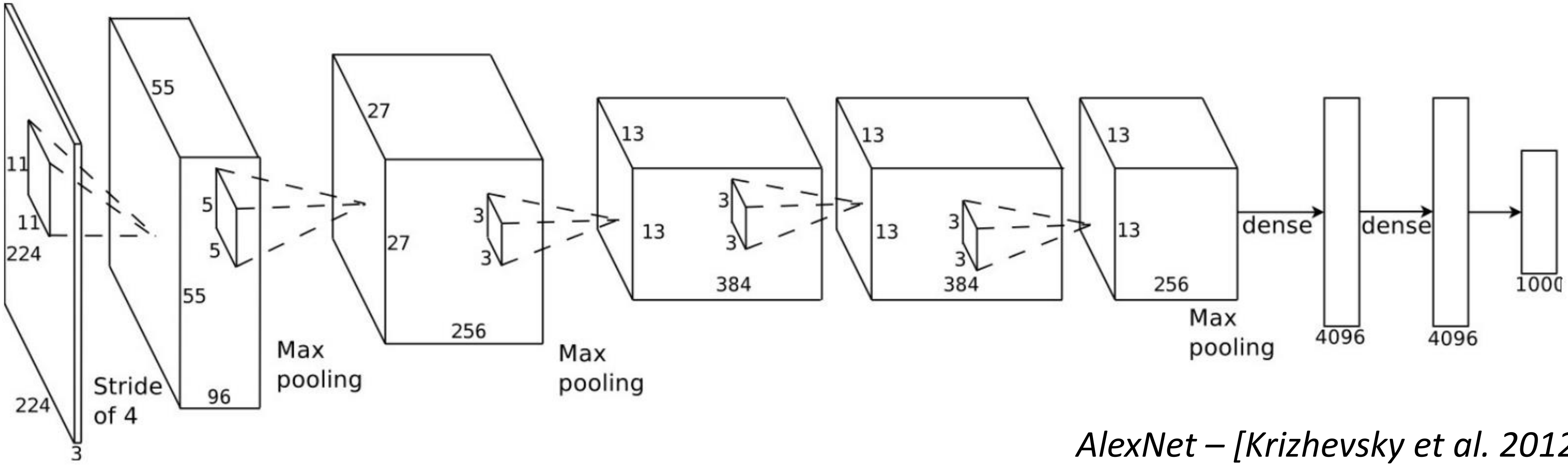

RELU




POOL

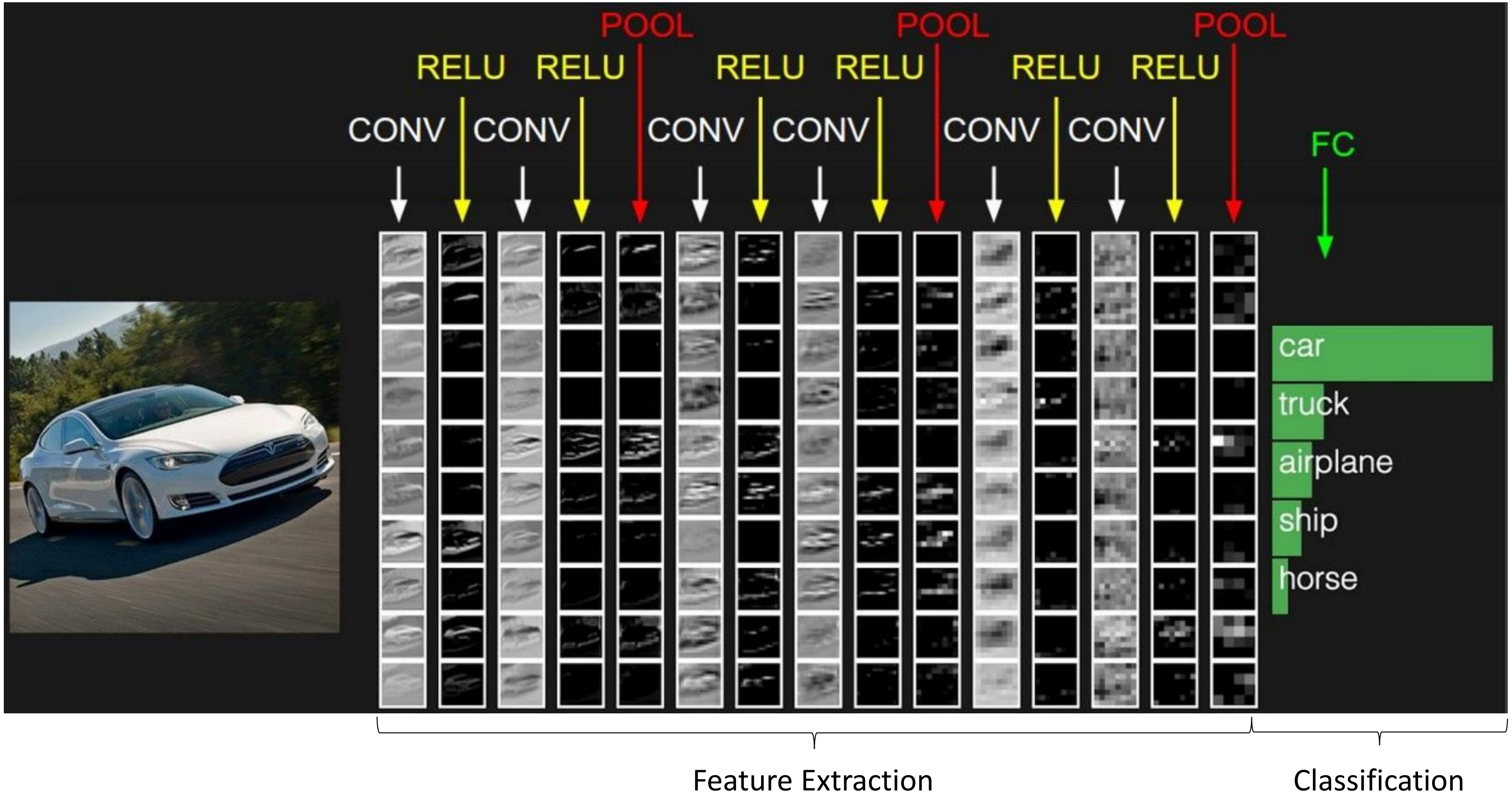



FC




*AlexNet – [Krizhevsky et al. 2012]*

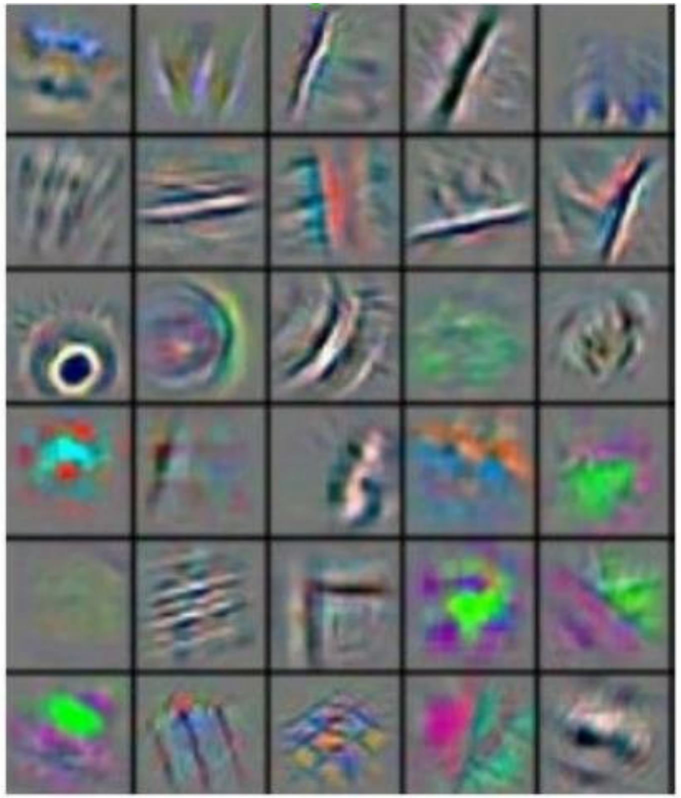
# 2. Convolutional Neural Networks



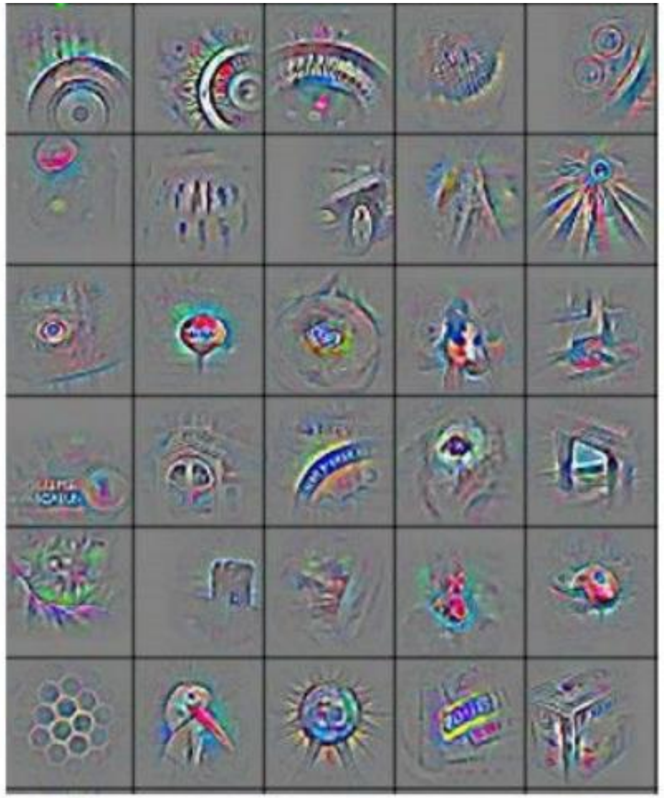
# 2. Convolutional Neural Networks

- Filters Learned by CNN

Low level Structures

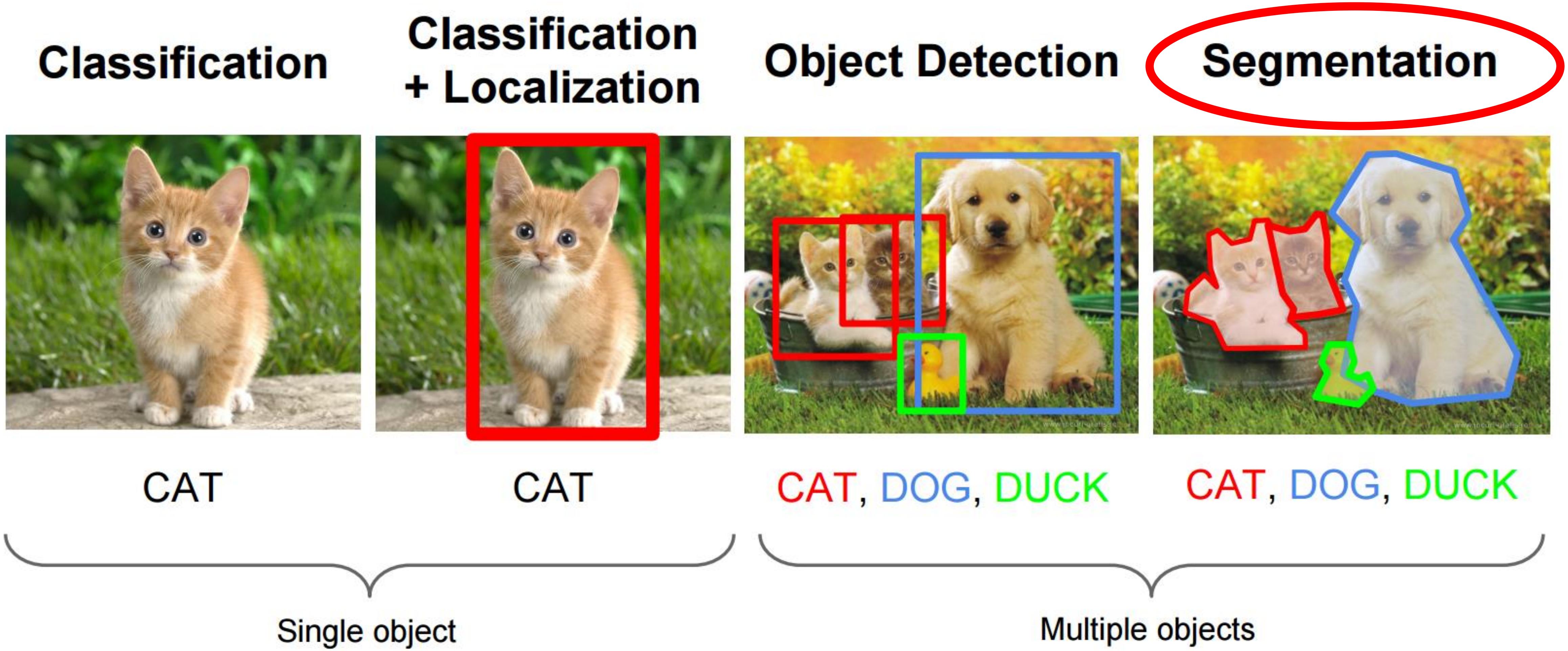


More Complex Structures

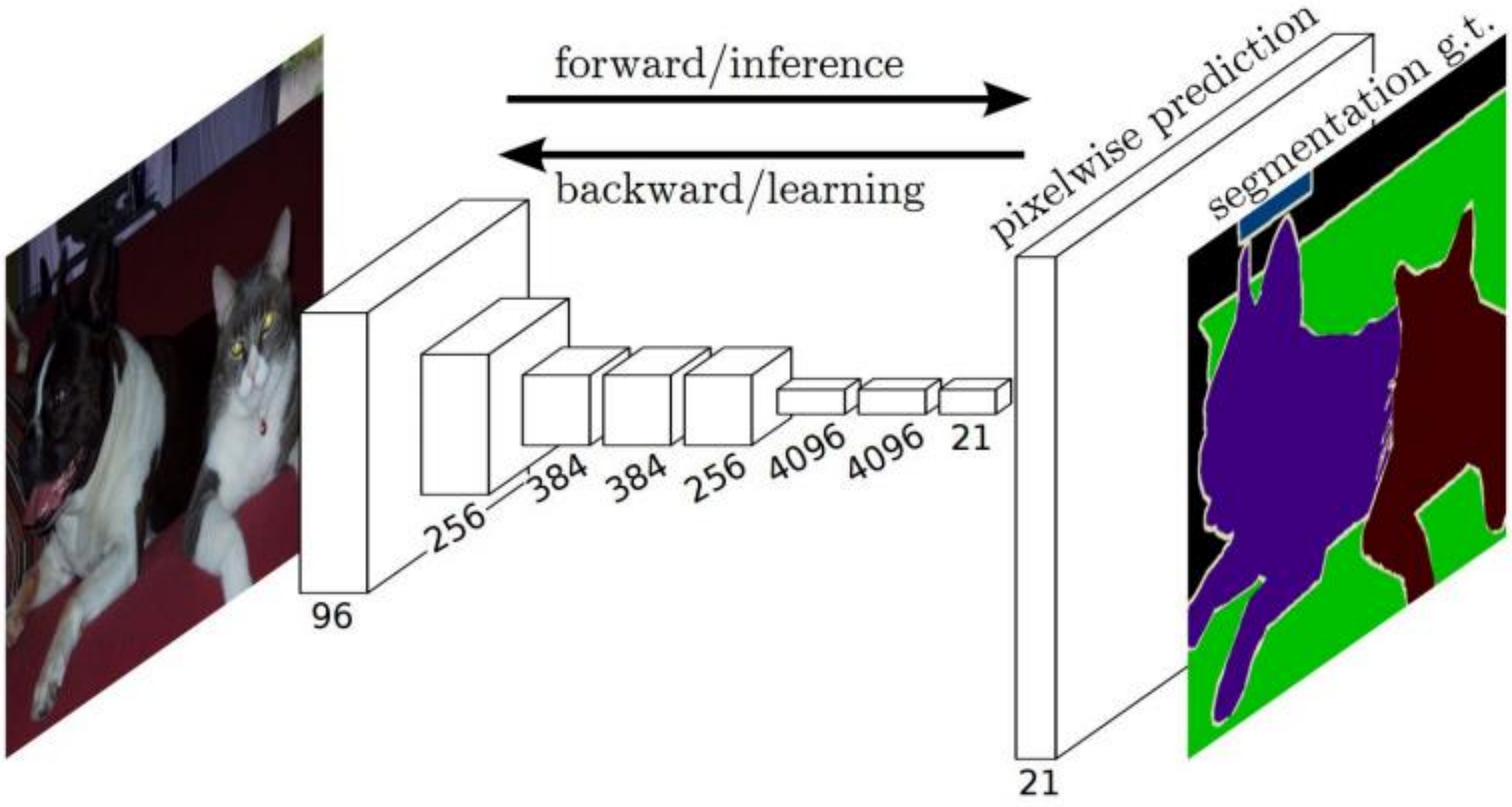


Highly Complex, Abstract Concepts

# 3. Fully Convolutional Networks



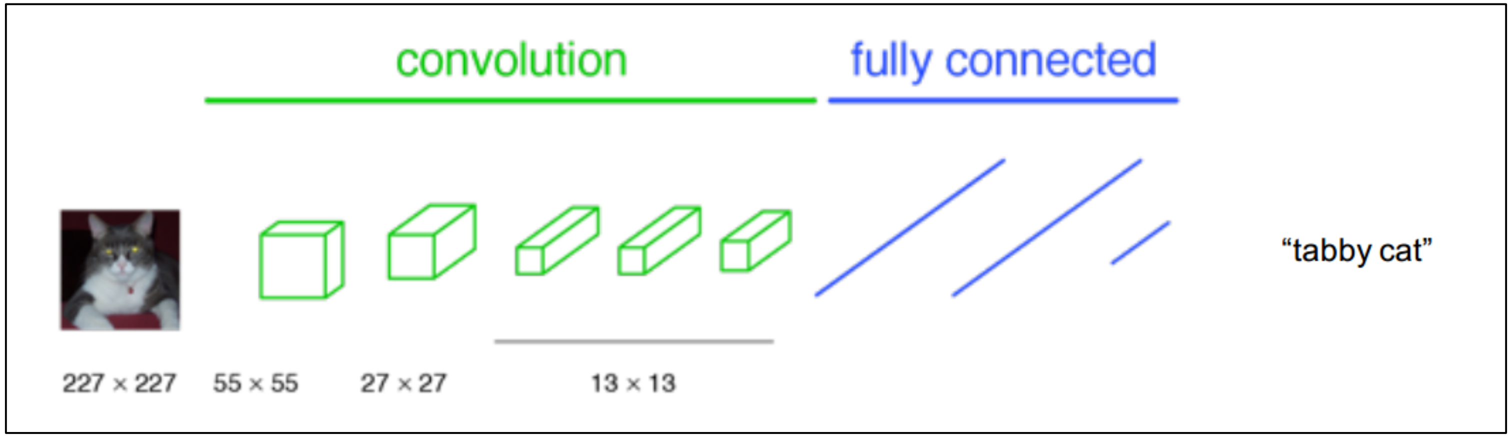
# 3. Fully Convolutional Networks



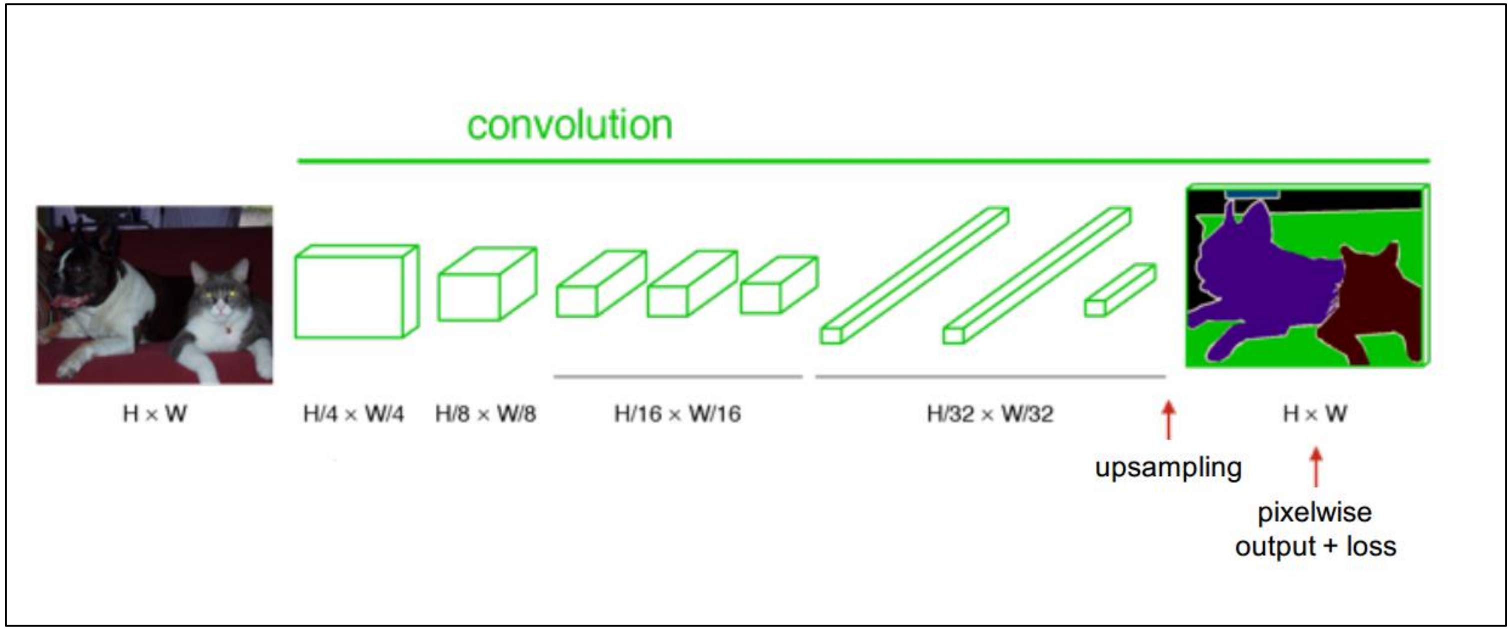

*Fully Convolutional Networks for Semantic Segmentation – [Long et al. 2015]*

# 3. Fully Convolutional Networks

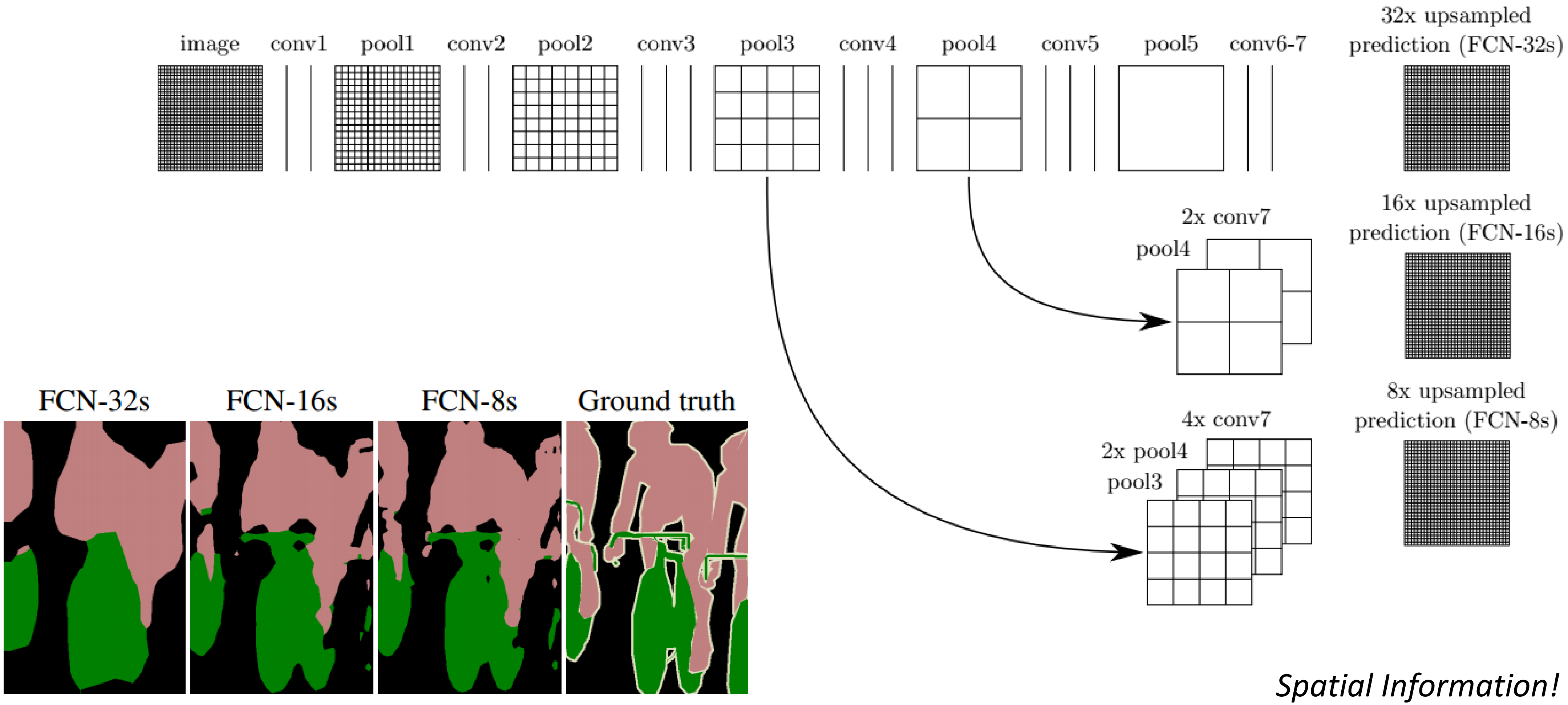
CNN




FCN



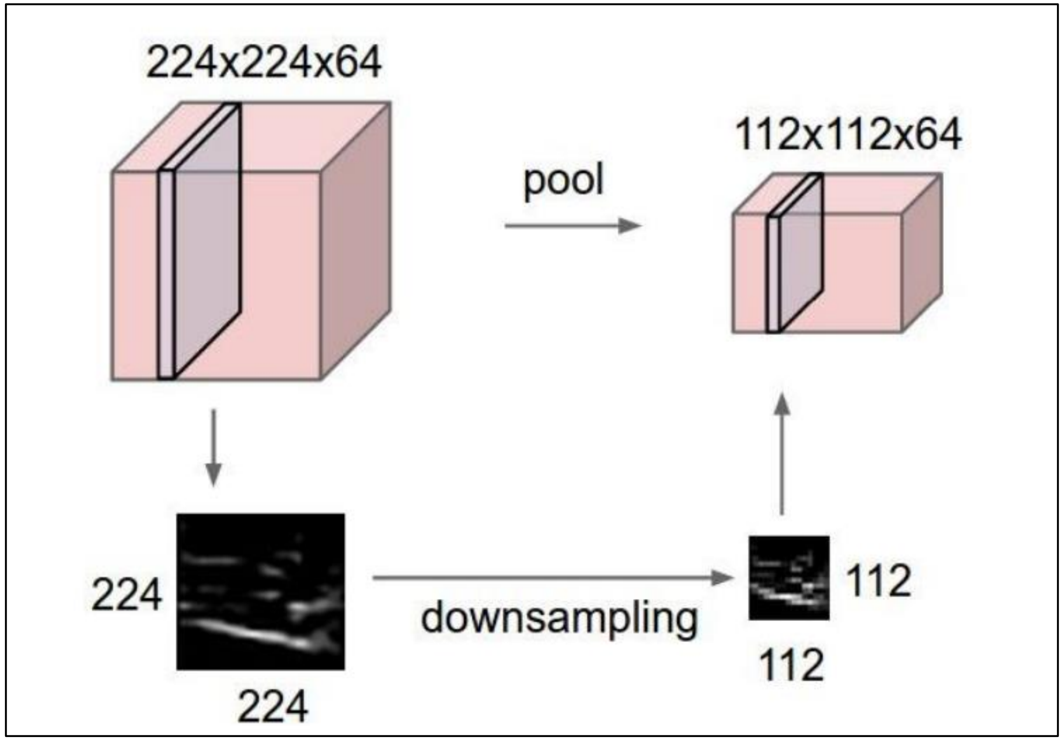
# 3. Fully Convolutional Networks

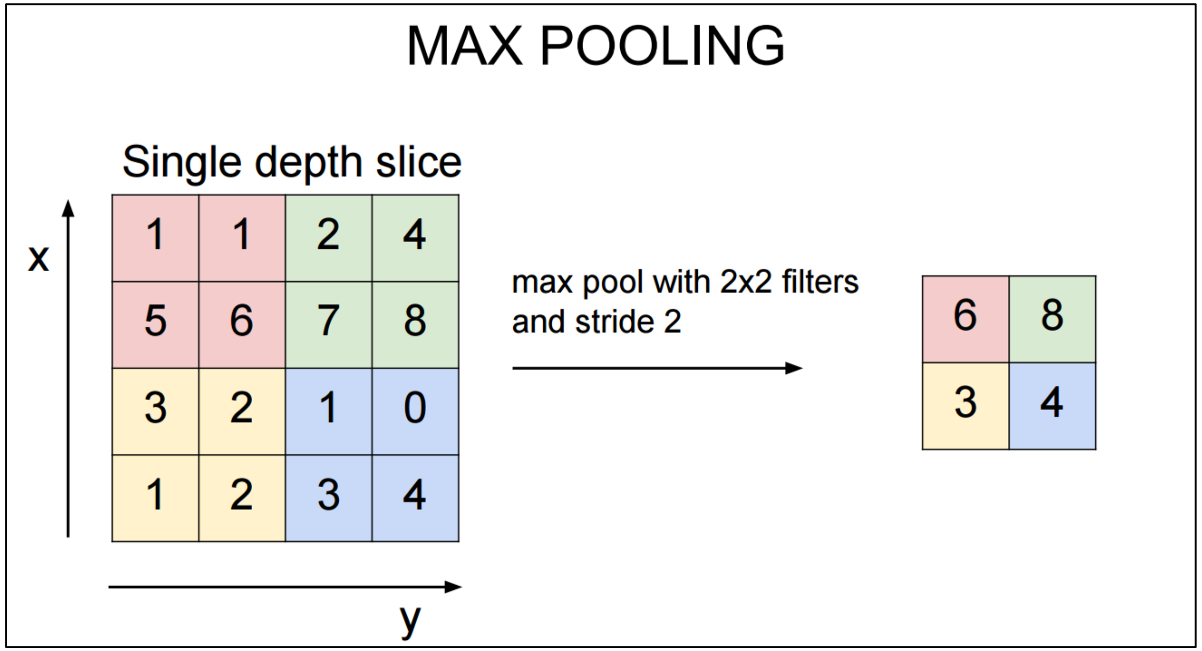



*Spatial Information!*

# 4. Problem Description

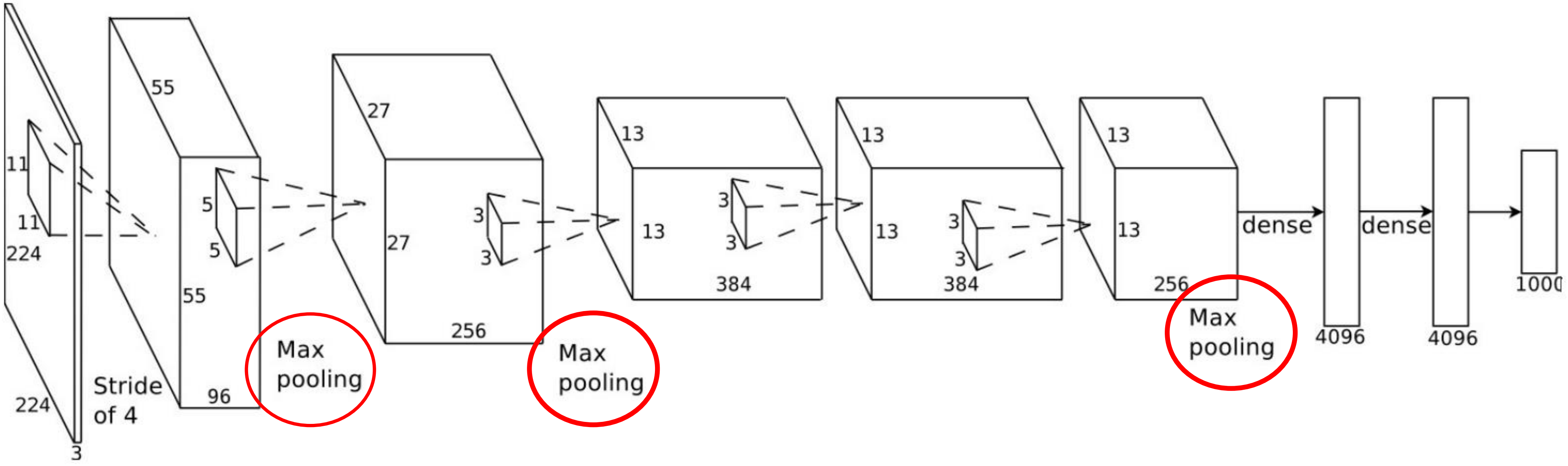
- Pooling Layer:



# 4. Problem Description

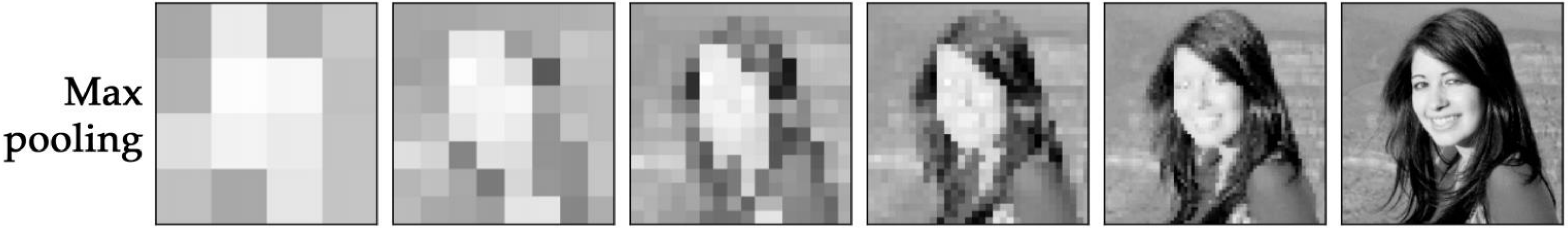
- Max Pooling




*AlexNet – [Krizhevsky et al. 2012]*

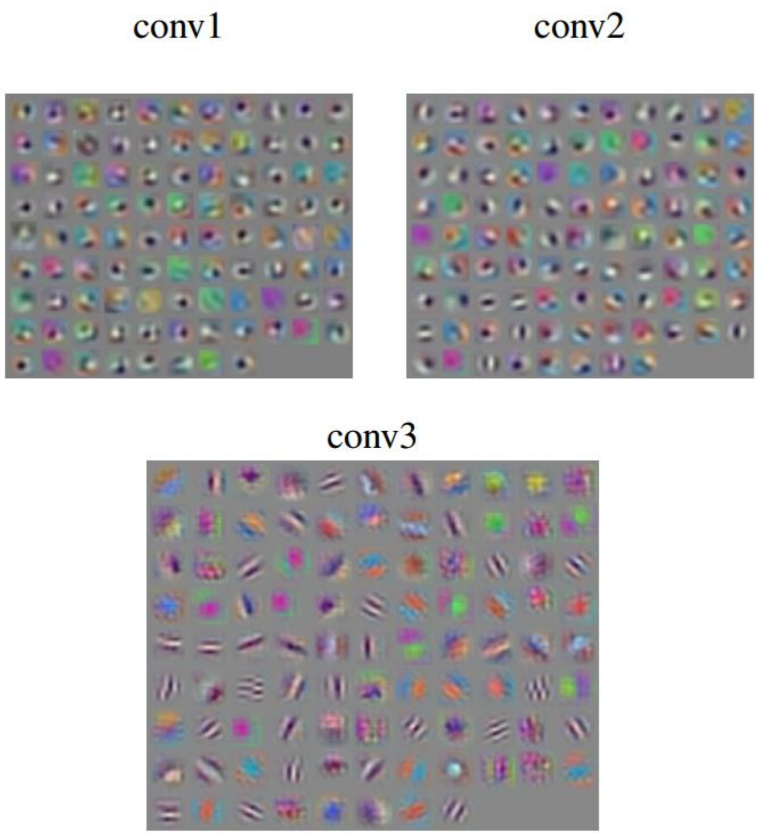
# 4. Problem Description

- Max Pooling – Problems and Limitations
  - Considerable amount of spatial information is lost
  - Restriction in output dimensionality



# STRIVING FOR SIMPLICITY: THE ALL CONVOLUTIONAL NET

**Jost Tobias Springenberg***, **Alexey Dosovitskiy***, **Thomas Brox, Martin Riedmiller**
Department of Computer Science
University of Freiburg
Freiburg, 79110, Germany
{springj, dosovits, brox, riedmiller}@cs.uni-freiburg.de




## ABSTRACT

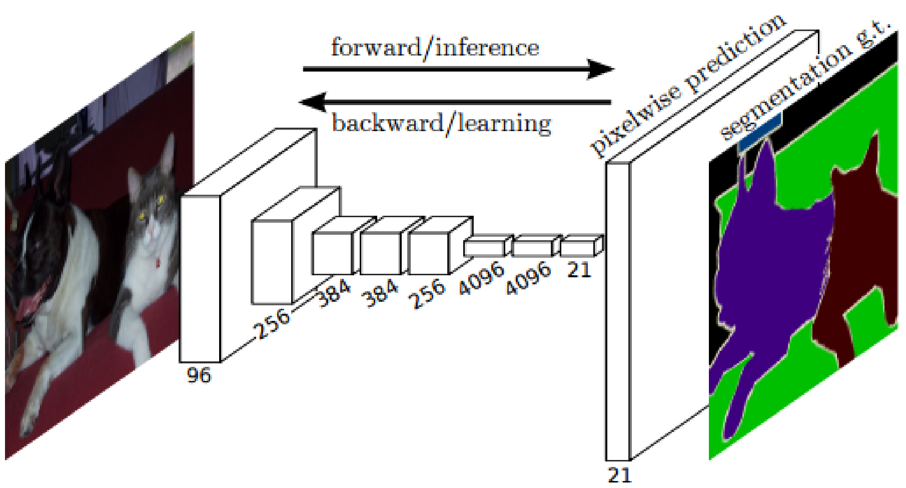
Most modern convolutional neural networks (CNNs) used for object recognition are built using the same principles: Alternating convolution and max-pooling layers followed by a small number of fully connected layers. We re-evaluate the state of the art for object recognition from small images with convolutional networks, questioning the necessity of different components in the pipeline. We find that max-pooling can simply be replaced by a convolutional layer with increased stride without loss in accuracy on several image recognition benchmarks. Following this finding – and building on other recent work for finding simple network structures – we propose a new architecture that consists solely of convolutional layers and yields competitive or state of the art performance on several object recognition datasets (CIFAR-10, CIFAR-100, ImageNet). To analyze the network we introduce a new variant of the "deconvolution approach" for visualizing features learned by CNNs, which can be applied to a broader range of network structures than existing approaches.

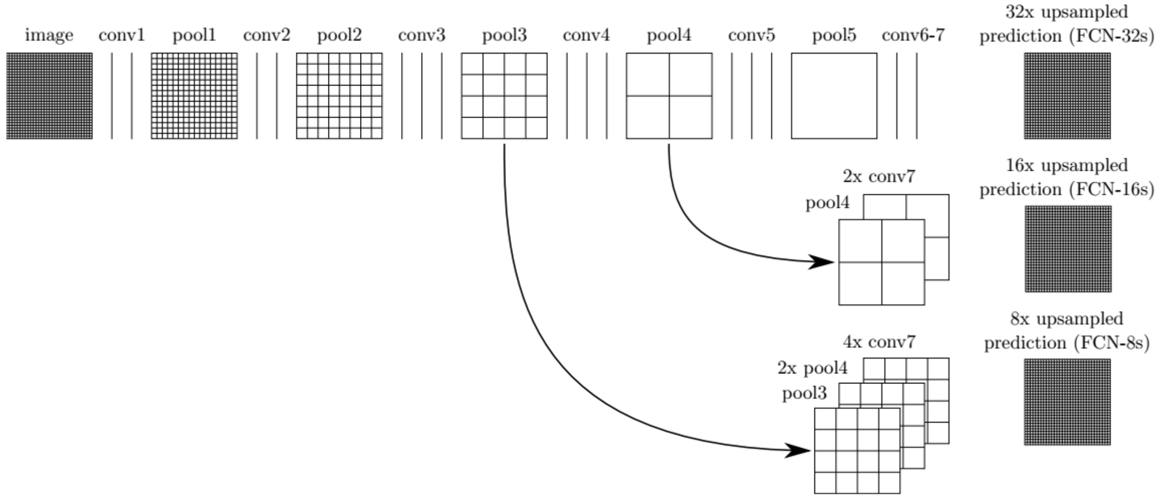
# Fully Convolutional Networks for Semantic Segmentation

Jonathan Long* Evan Shelhamer* Trevor Darrell
UC Berkeley
{jonlong,shelhamer,trevor}@cs.berkeley.edu

## Abstract

*Convolutional networks are powerful visual models that yield hierarchies of features. We show that convolutional networks by themselves, trained end-to-end, pixels-to-pixels, exceed the state-of-the-art in semantic segmentation. Our key insight is to build "fully convolutional" networks that take input of arbitrary size and produce correspondingly-sized output with efficient inference and learning. We define and detail the space of fully convolutional networks, explain their application to spatially dense prediction tasks, and draw connections to prior models. We adapt contemporary classification networks (AlexNet [22], the VGG net [34], and GoogLeNet [35]) into fully convolutional networks and transfer their learned representations by fine-tuning [5] to the segmentation task. We then define a skip architecture that combines semantic information from a deep, coarse layer with appearance information from a shallow, fine layer to produce accurate and detailed segmentations. Our fully convolutional network achieves state-of-the-art segmentation of PASCAL VOC (20% relative improvement to 62.2% mean IU on 2012), NYUDv2, and SIFT Flow, while inference takes less than one fifth of a second for a typical image.*



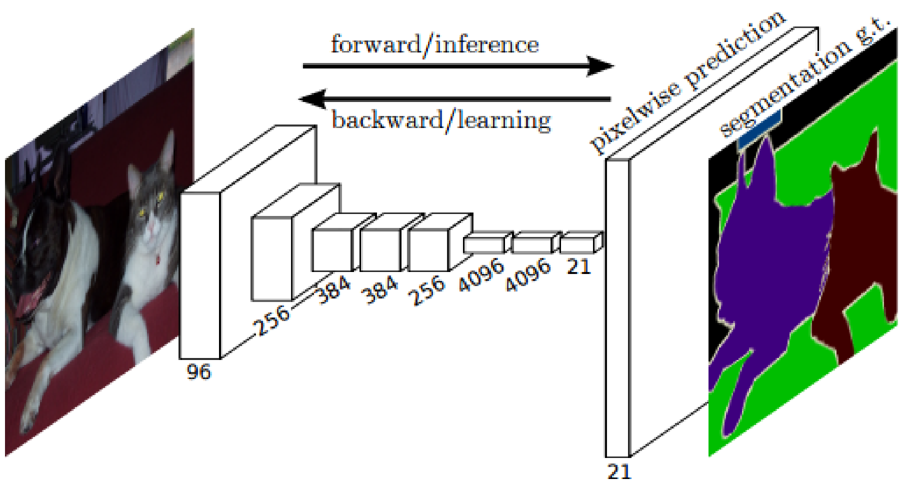



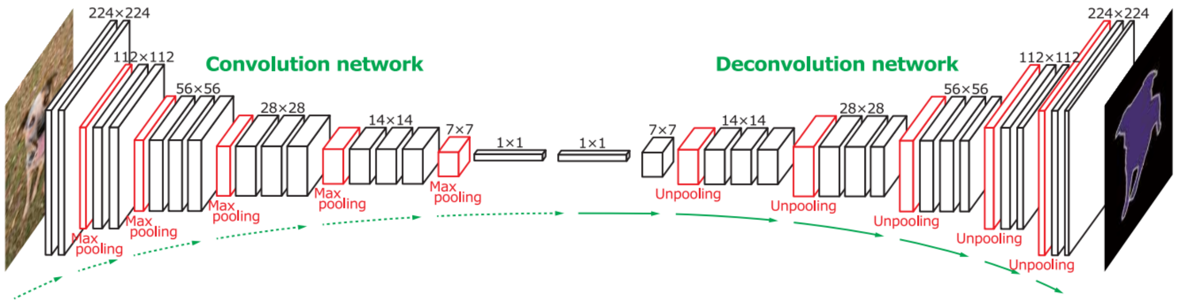
# Learning Deconvolution Network for Semantic Segmentation

Hyeonwoo Noh Seunghoon Hong Bohyung Han
Department of Computer Science and Engineering, POSTECH, Korea
{hyeonwoonoh_,maga33,bhhan}@postech.ac.kr

## Abstract

*We propose a novel semantic segmentation algorithm by learning a deep deconvolution network. We learn the network on top of the convolutional layers adopted from VGG 16-layer net. The deconvolution network is composed of deconvolution and unpooling layers, which identify pixel-wise class labels and predict segmentation masks. We apply the trained network to each proposal in an input image, and construct the final semantic segmentation map by combining the results from all proposals in a simple manner. The proposed algorithm mitigates the limitations of the existing methods based on fully convolutional networks by integrating deep deconvolution network and proposal-wise prediction; our segmentation method typically identifies detailed structures and handles objects in multiple scales naturally. Our network demonstrates outstanding performance in PASCAL VOC 2012 dataset, and we achieve the best accuracy (72.5%) among the methods trained without using Microsoft COCO dataset through ensemble with the fully convolutional network.*



# Seed, Expand and Constrain: Three Principles for Weakly-Supervised Image Segmentation

Alexander Kolesnikov — akolesnikov@ist.ac.at
Christoph H. Lampert — chl@ist.ac.at

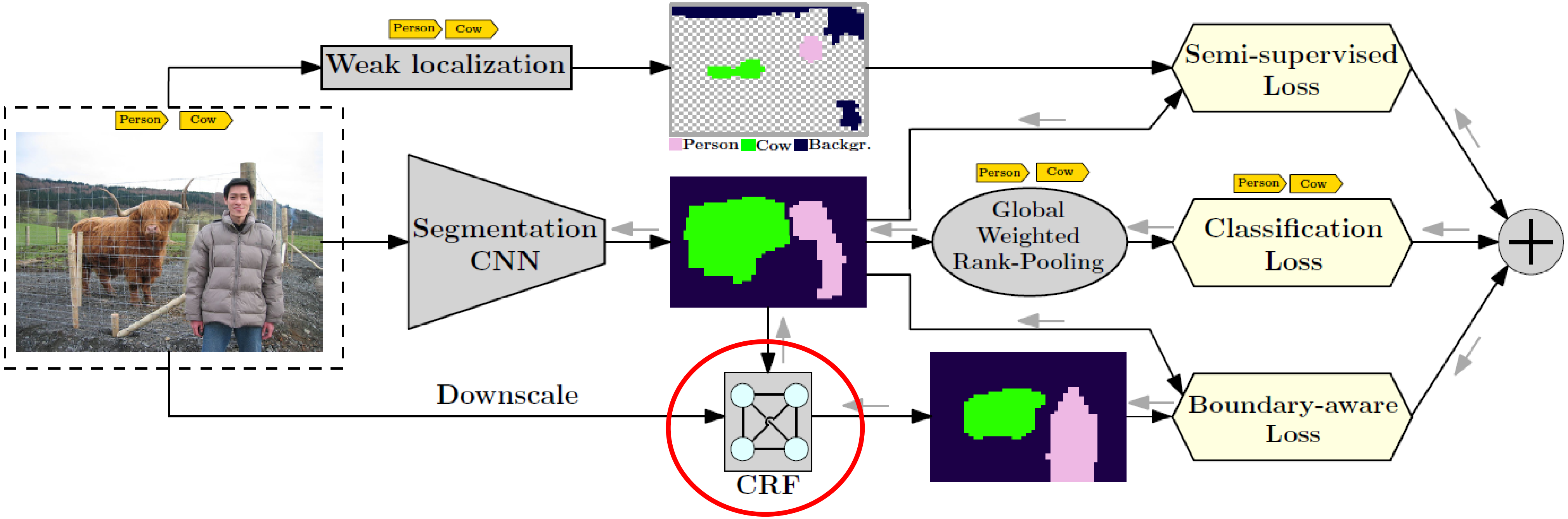
IST Austria

**Abstract.** We introduce a new loss function for the weakly-supervised training of semantic image segmentation models based on three guiding principles: to *seed* with weak location cues, to *expand* objects based on the information about which classes can occur, and to *constrain* the segmentations to coincide with image boundaries. We show experimentally that training a deep convolutional neural network using the proposed loss function leads to substantially better segmentations than previous state-of-the-art methods on the challenging PASCAL VOC 2012 dataset. We furthermore give insight into the working mechanism of our method by a detailed experimental study that illustrates how the segmentation quality is affected by each term of the proposed loss function as well as their combinations.

# Seed, Expand and Constrain: Three Principles for Weakly-Supervised Image Segmentation


Alexander Kolesnikov akolesnikov@ist.ac.at
Christoph H. Lampert chl@ist.ac.at

IST Austria


*Conditional Random Field*

## Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun

## Multi-Scale Orderless Pooling of Deep Convolutional Activation Features

Yunchao Gong[1], Liwei Wang[2], Ruiqi Guo[2], and Svetlana Lazebnik[2]

[1]University of North Carolina at Chapel Hill
yunchao@cs.unc.edu
[2]University of Illinois at Urbana-Champaign
{lwang97,guo29,slazebni}@illinois.edu

## Stochastic Pooling for Regularization of Deep Convolutional Neural Networks

**Matthew D. Zeiler**
Department of Computer Science
Courant Institute, New York University
zeiler@cs.nyu.edu

**Rob Fergus**
Department of Computer Science
Courant Institute, New York University
fergus@cs.nyu.edu

## Generalizing Pooling Functions in Convolutional Neural Networks: Mixed, Gated, and Tree

**Chen-Yu Lee**
UCSD ECE
chl260@ucsd.edu

**Patrick W. Gallagher**
UCSD Cognitive Science
patrick.w.gallagher@gmail.com

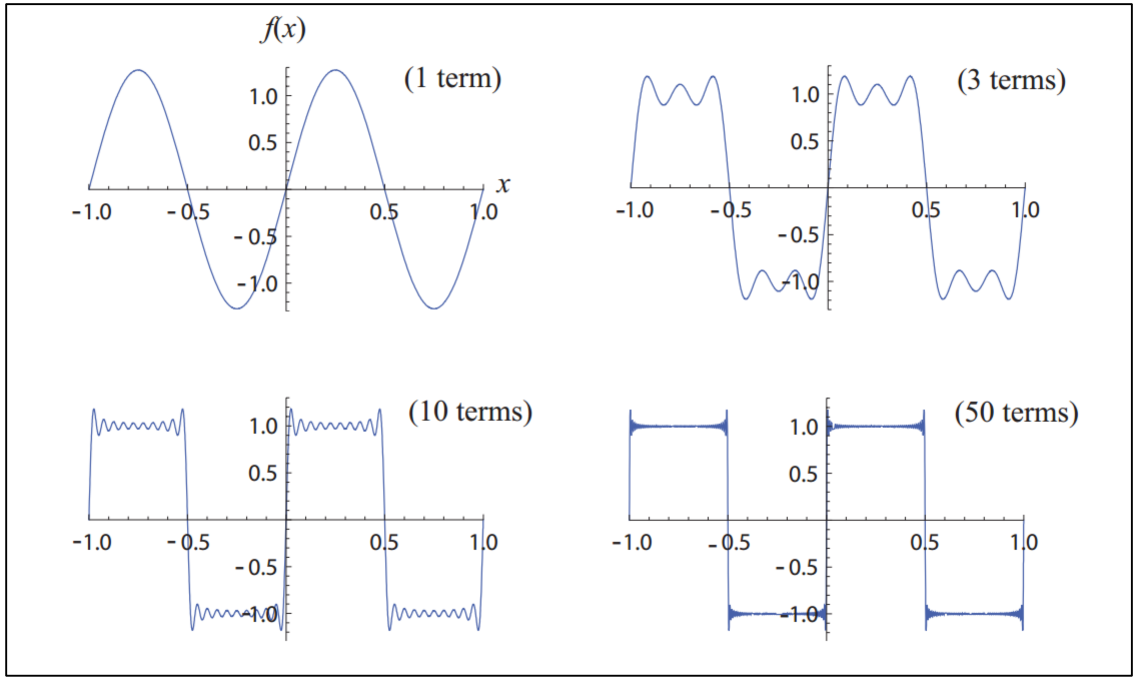
**Zhuowen Tu**
UCSD Cognitive Science
ztu@ucsd.edu

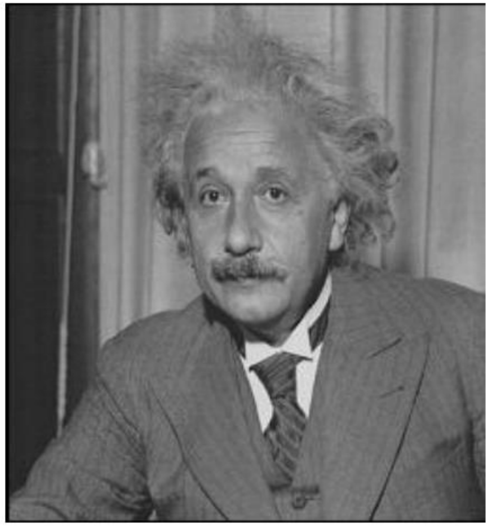
# 5. Spectral Pooling

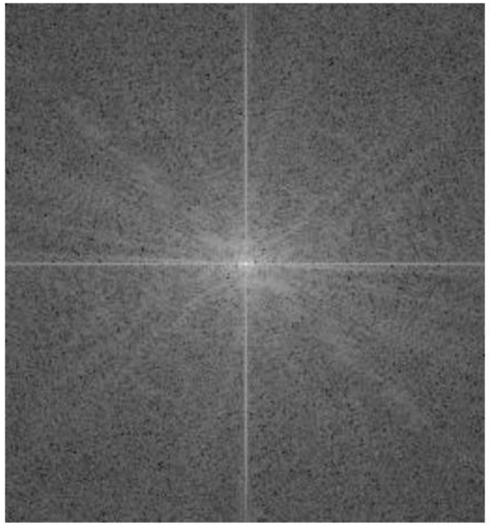
- Our Solution is '*Fourier Transform*'

$$\boxed{F(w) = \frac{1}{2\pi}\int_{-\infty}^{\infty} f(x)e^{-jwx}dx}$$

$$\boxed{f(x) = \int_{-\infty}^{\infty} F(w)e^{jwx}d\tau}$$

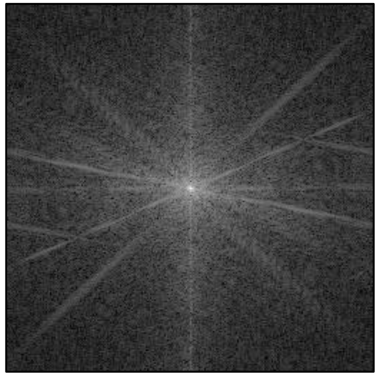


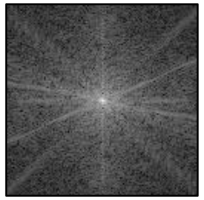


# 5. Spectral Pooling

N-Point
2D DFT



Crop High
Frequency Terms



N/2-Point
2D IDFT

$$F(u,v) = \frac{1}{NM}\sum_{x=0}^{N-1}\sum_{y=0}^{M-1} f(x,y)e^{-2\pi j\left(\frac{xu}{N}+\frac{yv}{M}\right)}$$

$$f(x,y) = \sum_{u=0}^{N-1}\sum_{v=0}^{M-1} F(u,v)e^{2\pi j\left(\frac{xu}{N}+\frac{yv}{M}\right)}$$

# Spectral Representations for Convolutional Neural Networks

**Oren Rippel**
Department of Mathematics
Massachusetts Institute of Technology
rippel@math.mit.edu

**Jasper Snoek**
School of Engineering and Applied Sciences
Harvard University
jsnoek@seas.harvard.edu

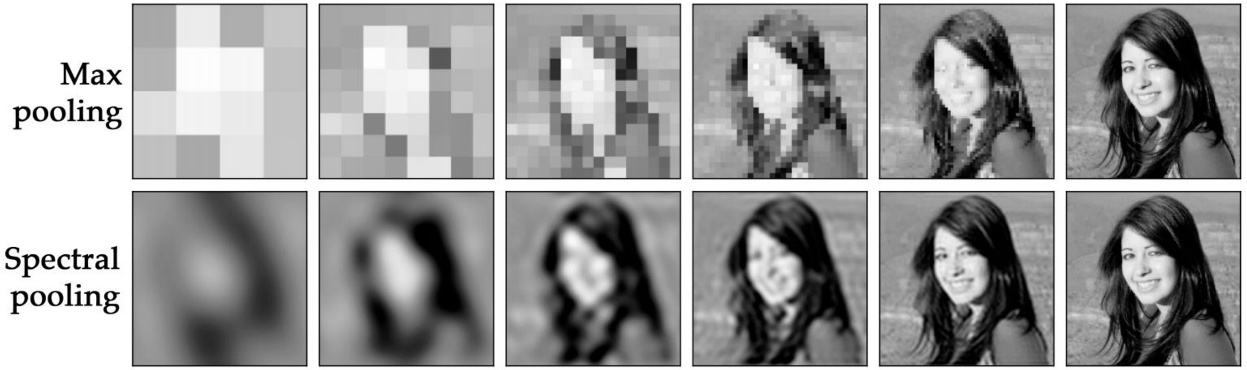
**Ryan P. Adams**
School of Engineering and Applied Sciences
Harvard University
rpa@seas.harvard.edu

## Abstract

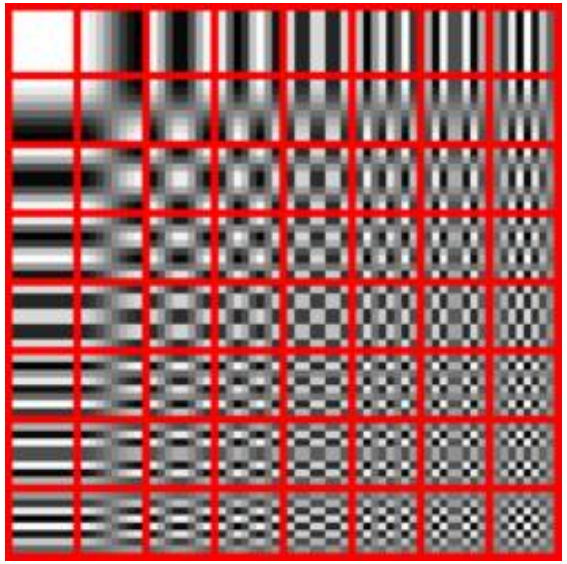
Discrete Fourier transforms provide a significant speedup in the computation of convolutions in deep learning. In this work, we demonstrate that, beyond its advantages for efficient computation, the spectral domain also provides a powerful representation in which to model and train convolutional neural networks (CNNs).

We employ spectral representations to introduce a number of innovations to CNN design. First, we propose *spectral pooling*, which performs dimensionality reduction by truncating the representation in the frequency domain. This approach preserves considerably more information per parameter than other pooling strategies and enables flexibility in the choice of pooling output dimensionality. This representation also enables a new form of stochastic regularization by randomized modification of resolution. We show that these methods achieve competitive results on classification and approximation tasks, without using any dropout or max-pooling.

Finally, we demonstrate the effectiveness of complex-coefficient spectral parameterization of convolutional filters. While this leaves the underlying model unchanged, it results in a representation that greatly facilitates optimization. We observe on a variety of popular CNN configurations that this leads to significantly faster convergence during training.



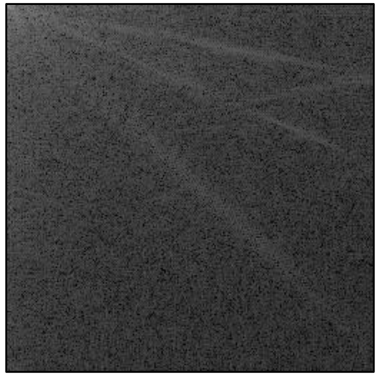
# 5. Spectral Pooling



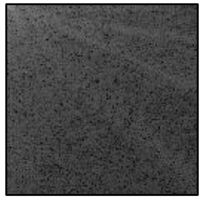
- **Further Improvement:** *Discrete Cosine Transform* (DCT)
    - Very similar to Discrete Fourier Transform
    - A real-valued transform unlike complex-valued DFT
    - Has energy compaction property
    - Very similar to Karhunen-Loève Transform (KLT) and Principal Component Analysis (PCA)
    - Basis functions are constant, real-valued cosine functions

# 5. Spectral Pooling

N-Point
2D DCT



Crop High
Frequency Terms



N/2-Point
2D IDCT

$$F(u,v) = a(u)a(v)\sum_{x=0}^{N-1}\sum_{y=0}^{M-1} f(x,y)\cos\left(\frac{\pi(2x+1)u}{2N}\right)\cos\left(\frac{\pi(2y+1)v}{2M}\right)$$

$$f(x,y) = \sum_{u=0}^{N-1}\sum_{v=0}^{M-1} a(u)a(v)F(x,y)\cos\left(\frac{\pi(2x+1)u}{2N}\right)\cos\left(\frac{\pi(2y+1)v}{2M}\right)$$

$$a(x) = \begin{cases}\sqrt{{}^{1}/_{N}}, & x = 0\\ \sqrt{{}^{2}/_{N}}, & x \neq 0\end{cases}$$

# 5. Spectral Pooling

## Forward Propagation

$I \rightarrow NxM\ Input\ Image$

$I_{DCT} \rightarrow NxM\ DCT\ of\ Input\ Image$

$I_{OUT} \rightarrow N/2\,x\,M/2\ \ Output\ Image$

$D_1 \rightarrow NxN\ DCT\ Matrix$ $\quad D_{1c} \rightarrow N/2\,xN$

$D_2 \rightarrow MxM\ DCT\ Matrix$ $\quad D_{2c} \rightarrow M/2\,xM$

$D_{1s} \rightarrow N/2\,x\,N/2\ DCT\ Matrix$

$D_{2s} \rightarrow M/2\,x\,M/2\ DCT\ Matrix$

$$I_{DCT} = D_1\ x\ I\ xD_2^T$$

$$I'_{DCT} = D_{1c}\ x\ I\ xD_{2c}^T$$

$$I_{OUT} = D_{1s}^T\ x\ I'_{DCT}\ xD_{2s}$$

$$I_{OUT} = D_{1s}^T\ xD_{1c}\ x\ I\ xD_{2c}^T\ xD_{2s}$$

$$\boxed{\begin{array}{l} I_{OUT} = A\ x\ I\ x\ B \\ A = D_{1s}^T\ xD_{1c} \qquad B = D_{2c}^T\ xD_{2s} \end{array}}$$

# 5. Spectral Pooling

## Backward Propagation

$I \rightarrow NxM\ Input\ Image$

$I_{DCT} \rightarrow NxM\ DCT\ of\ Input\ Image$

$I_{OUT} \rightarrow N/2\, x\, M/2\ \ Output\ Image$

$D_1 \rightarrow NxN\ DCT\ Matrix$ $\quad D_{1c} \rightarrow N/2\, xN$

$D_2 \rightarrow MxM\ DCT\ Matrix$ $\quad D_{2c} \rightarrow M/2\, xM$

$D_{1s} \rightarrow N/2\, x\, N/2\ DCT\ Matrix$

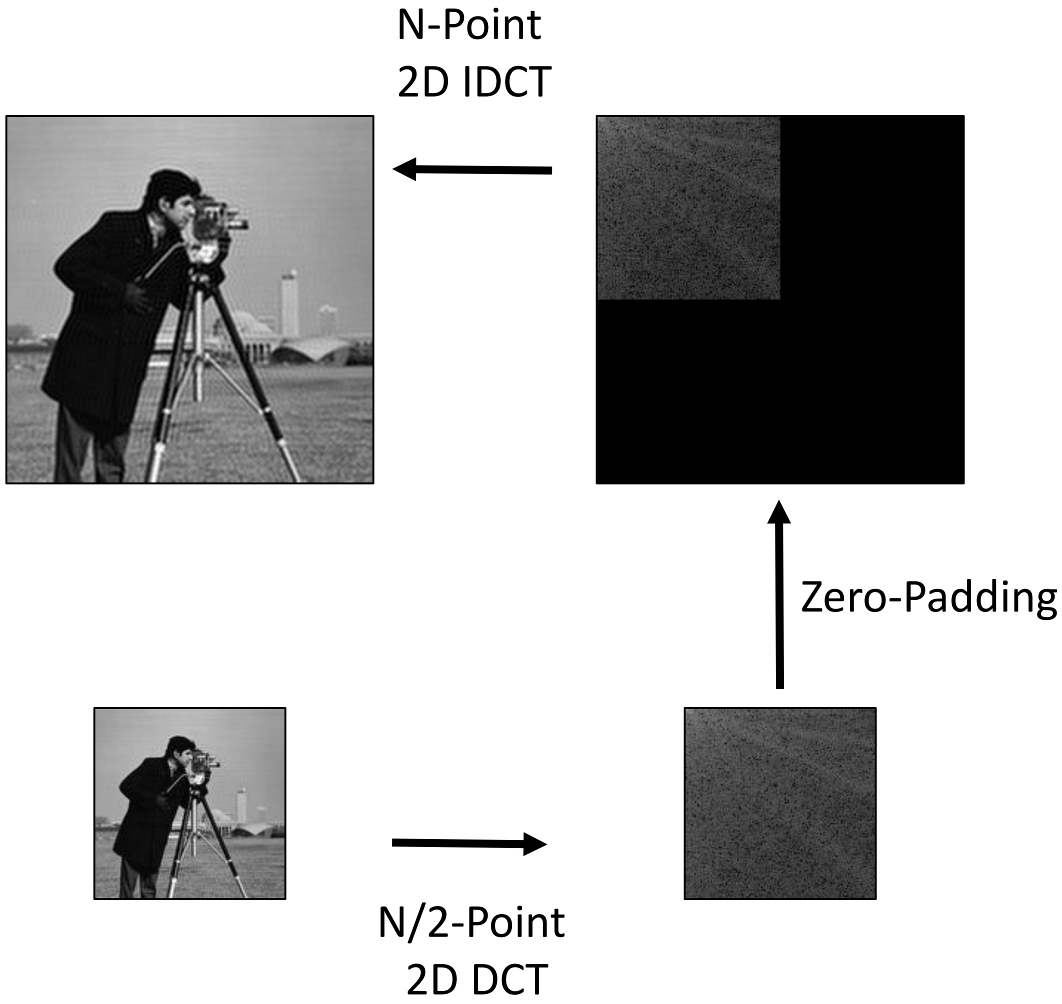
$D_{2s} \rightarrow M/2\, x\, M/2\ DCT\ Matrix$

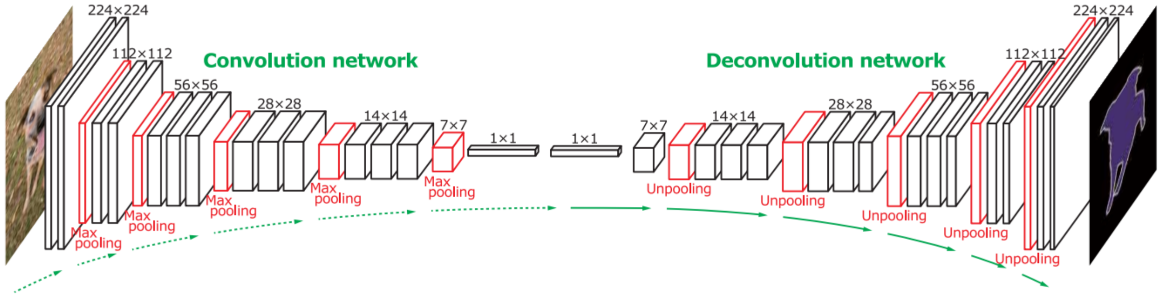
$$\boxed{\begin{array}{l} I_{OUT} = A\ x\ I\ x\ B \\ A = D_{1s}^{T}\ xD_{1c} \qquad B = D_{2c}^{T}\ xD_{2s} \end{array}}$$

$$\frac{\partial y}{\partial x} = \sum_{i=1}^{N}\sum_{j=1}^{M}\frac{\partial y}{\partial p_{ij}}\frac{\partial p_{ij}}{\partial x}$$

$$\frac{\partial y}{\partial x} = \sum_{i=1}^{N}\sum_{j=1}^{M}\frac{\partial y}{\partial p_{ij}}(a_i b_j^T) = A^T\ x\ I_{OUT}\ x\ B^T$$

# 5. Spectral Pooling / Unpooling



# 6. Experiments and Results

| | pixel acc. | mean acc. | mean IU | f.w. IU |
|---|---|---|---|---|
| **FCN-32s** | 89.1 | 73.3 | 59.4 | 81.4 |
| **FCN-16s** | 90.0 | 75.7 | 62.4 | 83.0 |
| **FCN-8s** | 90.3 | 75.9 | 62.7 | 83.2 |
| **FCN-32s DCT** | - | - | - | - |
| **FCN-16s DCT** | - | - | - | - |
| **FCN-8s DCT** | - | - | - | - |

*PASCAL VOC 2011 Dataset*

# 7. Conclusion


- Spectral Pooling using Discrete Wavelet Transform (DWT)
- Construct Deeper Networks Completely in Frequency Domain
- Use Spectral Unpooling in Deconvolutional Networks
- Use Spectral Pooling in Different Architectures

# References


- A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In NIPS, 2012.
- B. B. Le Cun, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, “Handwritten digit recognition with a backpropagation network,” in NIPS. Citeseer, 1990.
- C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. CoRR, abs/1409.4842, 2014.
- M. D. Zeiler and R. Fergus, “Visualizing and understanding convolutional networks,” in ECCV, 2014.
- LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. Nature 521, 436–444 (2015)
- J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In CVPR, 2015.
- H. Noh, S. Hong, and B. Han. Learning deconvolution network for semantic segmentation. In Proc. Int. Conf. Comp. Vis., 2015.
- J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller. Striving for simplicity: The all convolutional net. In ICLR 2015 Workshop Track, 2015.
- Kolesnikov, A., Lampert, C.: Seed, expand and constrain: Three principles for weakly-supervised image segmentation, 2016.
- S. Ren, K. He, R. Girshick, and J. Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. arXiv preprint arXiv:1506.01497, 2015.
- M. Lin, Q. Chen, and S. Yan, “Network in network,” arXiv preprint arXiv:1312.4400, 2013.
- C. Gulcehre, K. Cho, R. Pascanu, and Y. Bengio, “Learned-norm pooling for deep feedforward and recurrent neural networks,” in ML and KDD, 2014.
- L. Wan, M. Zeiler, S. Zhang, Y. L. Cun, and R. Fergus, “Regularization of neural networks using dropconnect,” in ICML, 2013.
- D. Yu, H. Wang, P. Chen, and Z. Wei, “Mixed pooling for convolutional neural networks,” in Rough Sets and Knowledge Technology, 2014.
- M. D. Zeiler and R. Fergus, “Stochastic pooling for regularization of deep convolutional neural networks,” CoRR, 2013.
- O. Rippel, J. Snoek, and R. P. Adams, “Spectral representations for convolutional neural networks,” arXiv preprint arXiv:1506.03767, 2015.
- K. He, X. Zhang, S. Ren, and J. Sun, “Spatial pyramid pooling in deep convolutional networks for visual recognition,” in ECCV, 2014.
- J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, and G. Wang, “Recent advances in convolutional neural networks,” arXiv preprint arXiv:1512.07108, 2015.

*Thank you!*