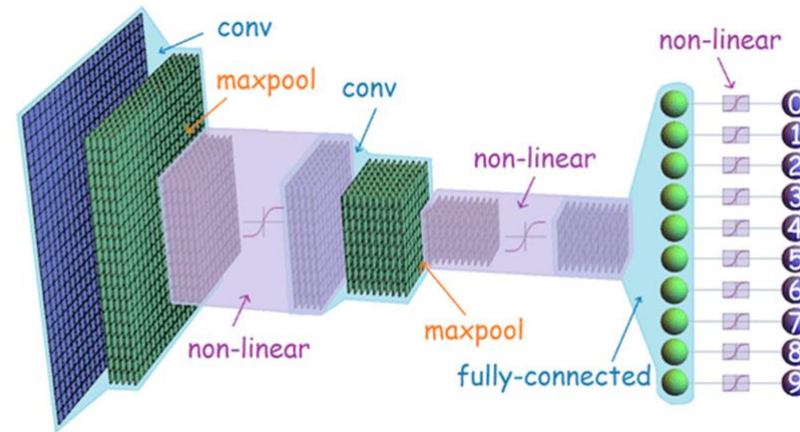


# Deep Learning Workshop

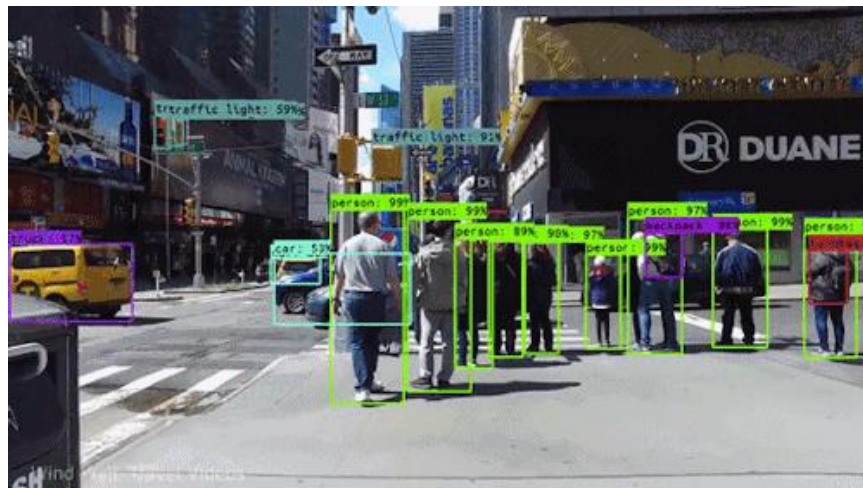
## Convolutional Neural Networks



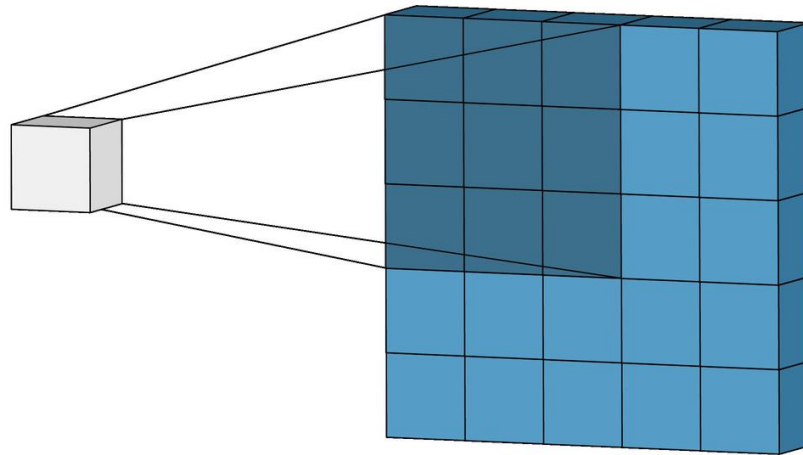
Instructor: Aaron Low

HELP University, Faculty of Computing and Digital Technology

# Computer Vision

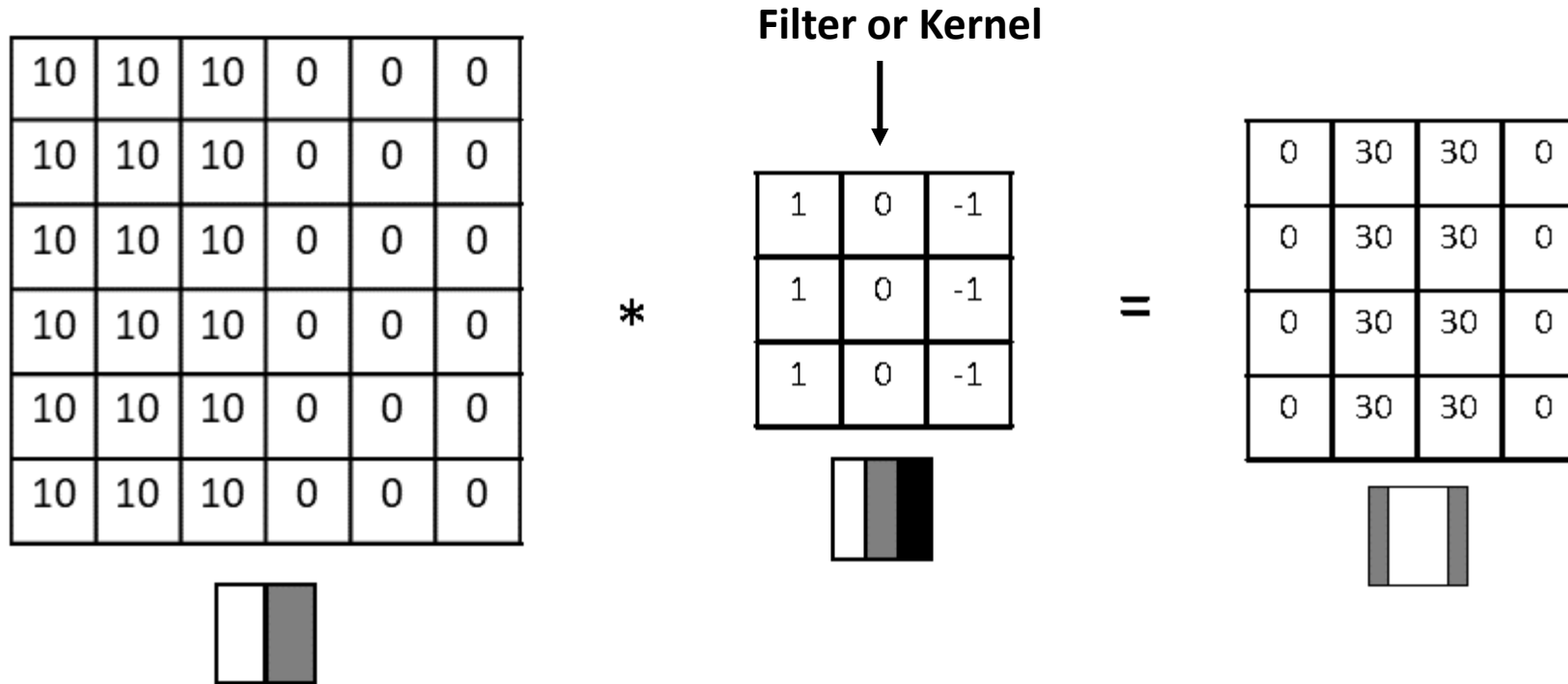


# Convolution Operation

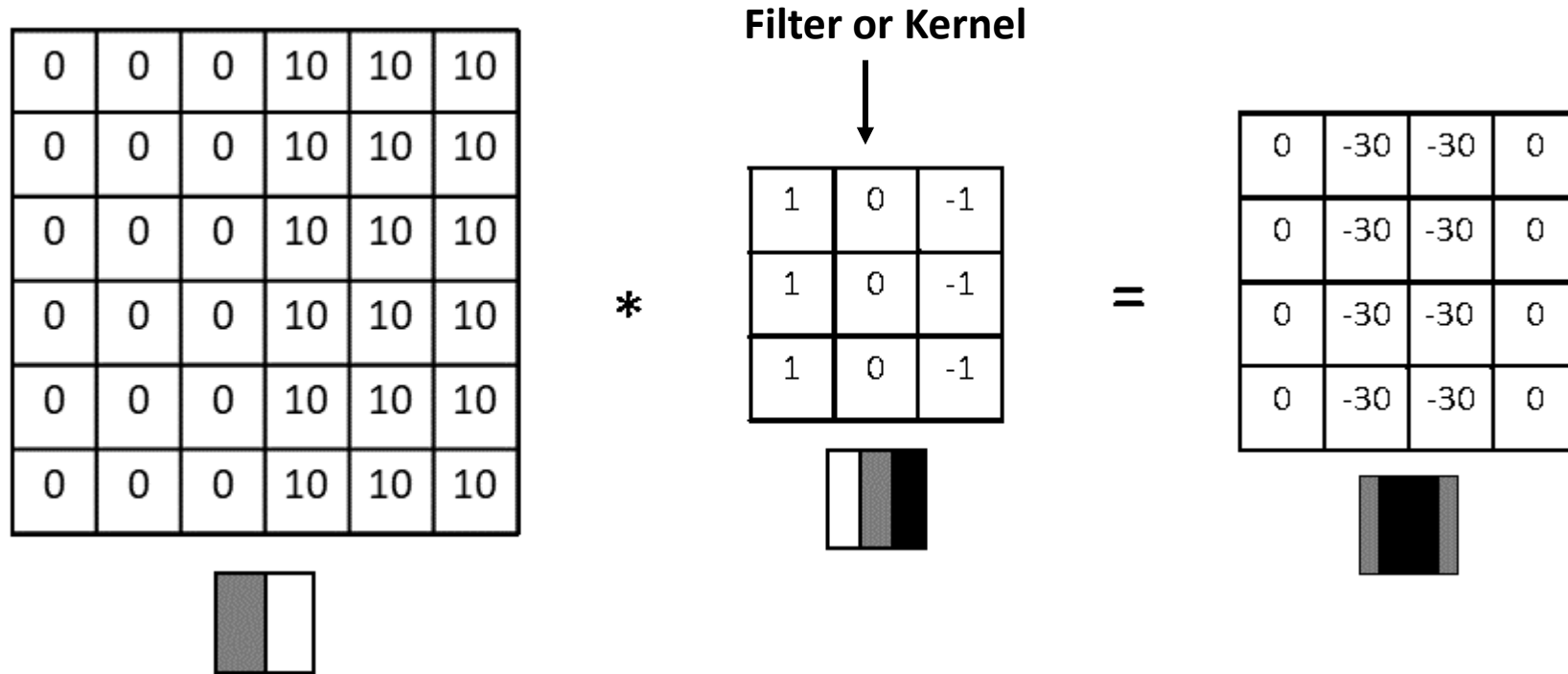


Convolutional Layer GIF from <https://blog.usejournal.com/convolutional-neural-networks-why-what-and-how-f8f6dbebb2f9?gi=3faa9b8cfe4c>

# Convolution Operation: Edge Detection

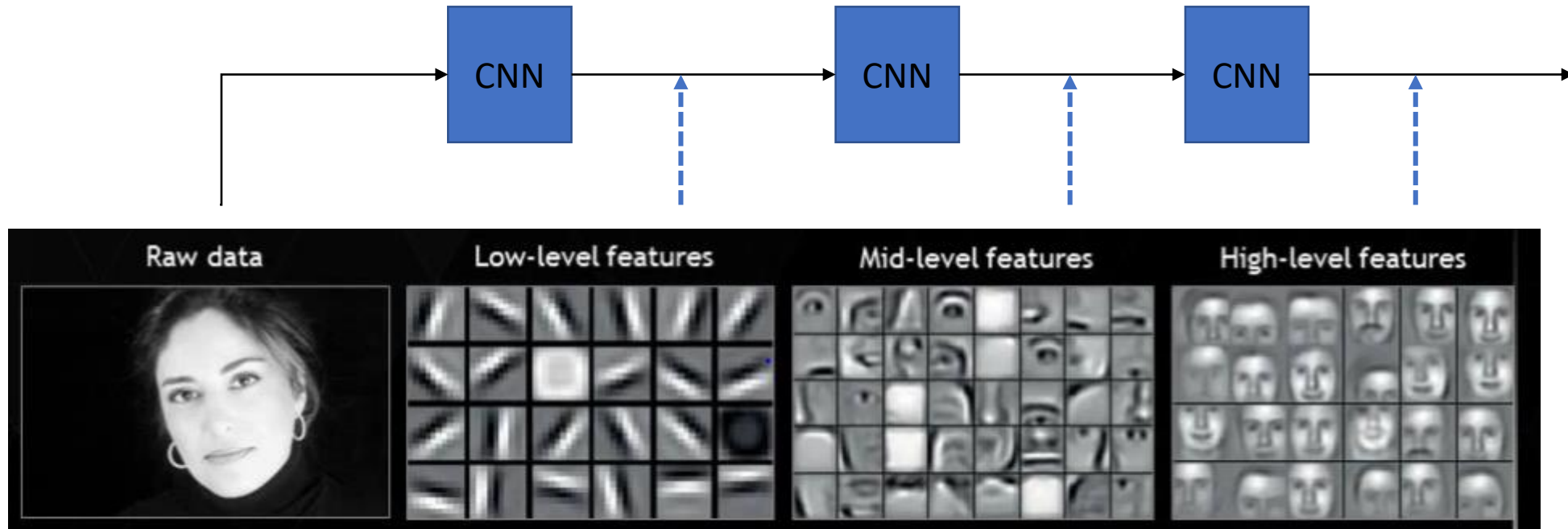


# Convolution Operation: Edge Detection



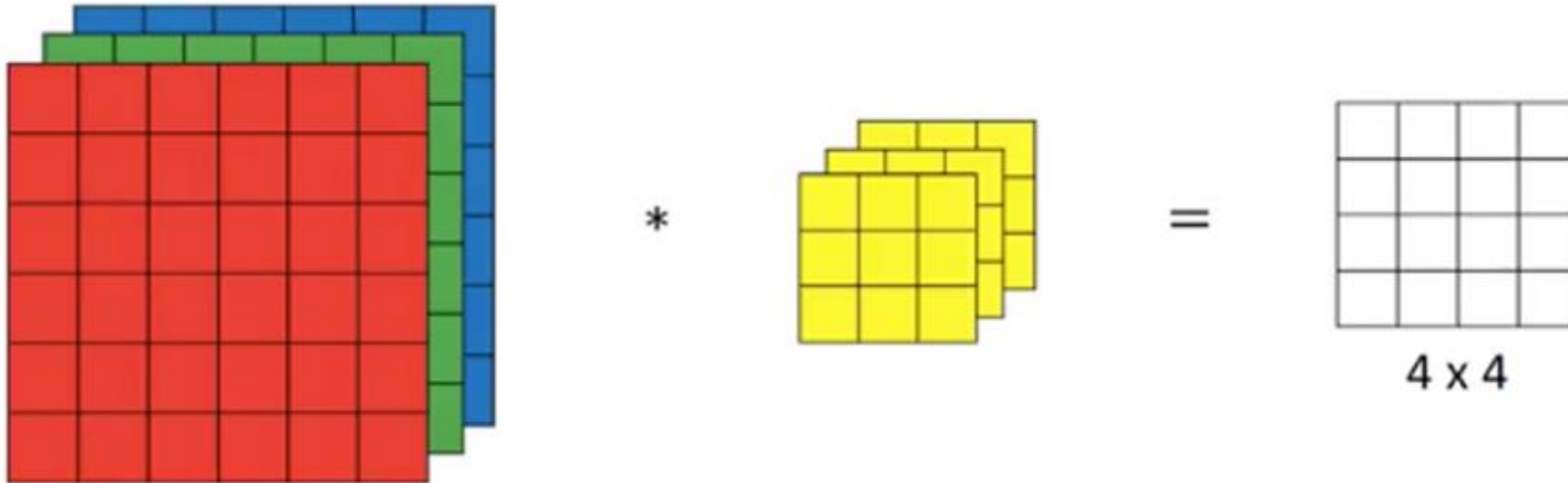
# Feature Learning

- Deep Learning is **Representation/Feature Learning**

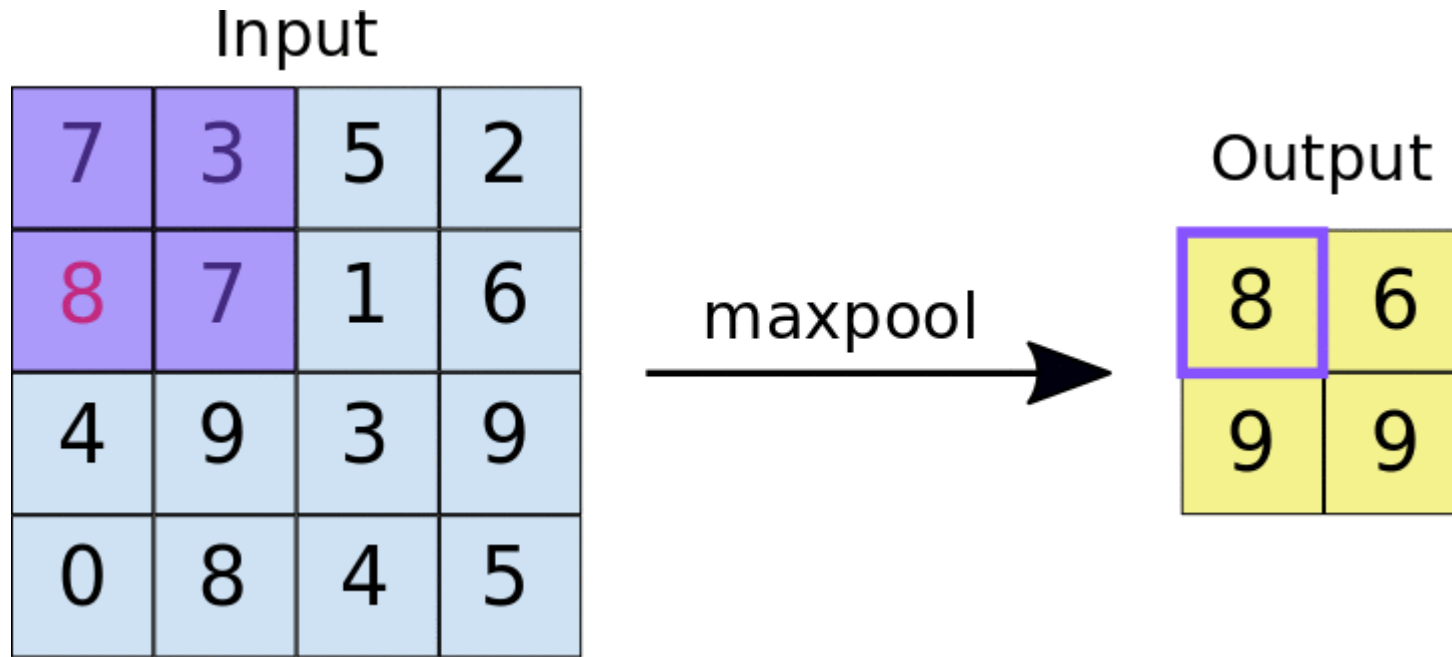


# Convolutional Neural Network: Convolutional Layer

- No need to hand choose filters just **learn the correct weights**



# Convolutional Neural Networks: Max Pool



Max Pool layer from <https://developers.google.com/machine-learning/practica/image-classification/convolutional-neural-networks>



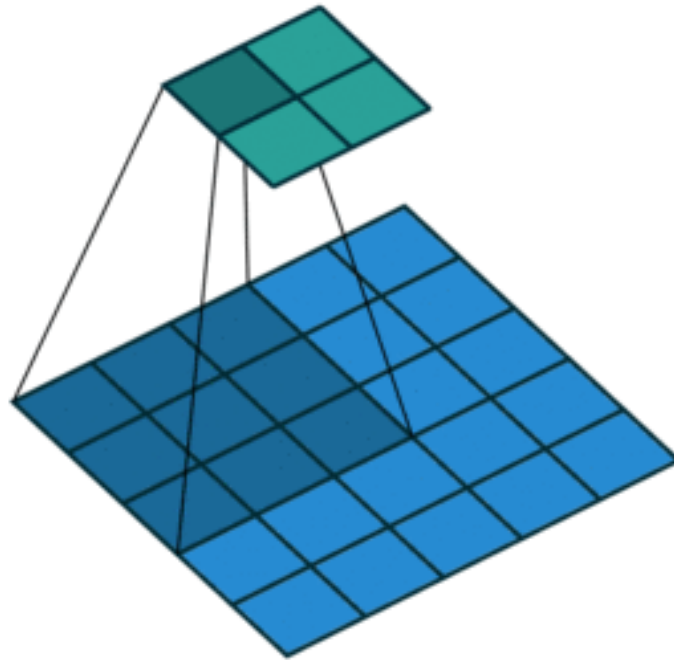
# Convolutional Neural Networks: Padding

0 <sub>2</sub>	0 <sub>0</sub>	0 <sub>1</sub>	0	0	0	0
0 <sub>1</sub>	2 <sub>0</sub>	2 <sub>0</sub>	3	3	3	0
0 <sub>0</sub>	0 <sub>1</sub>	1 <sub>1</sub>	3	0	3	0
0	2	3	0	1	3	0
0	3	3	2	1	2	0
0	3	3	0	2	3	0
0	0	0	0	0	0	0

1	6	5
7	10	9
7	10	8

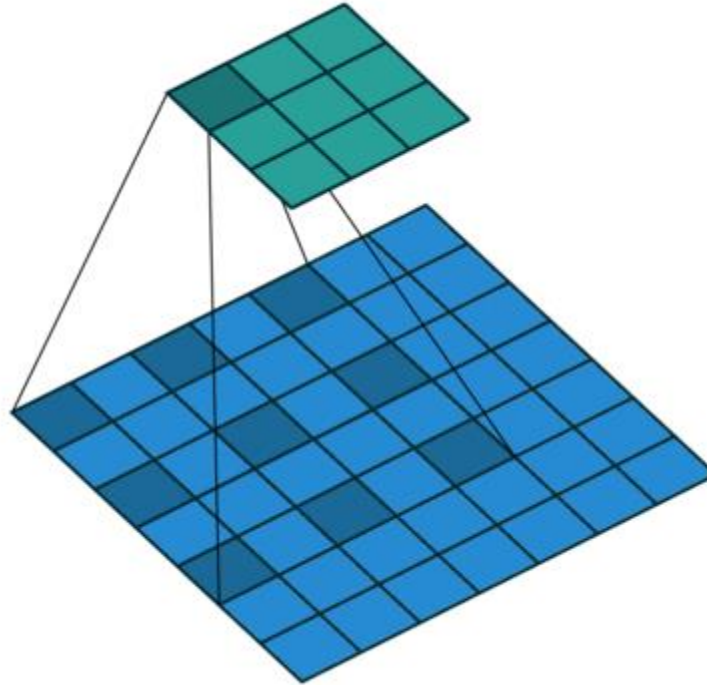
Padding gif from [http://deeplearning.net/software/theano/tutorial/conv\\_arithmetic.html](http://deeplearning.net/software/theano/tutorial/conv_arithmetic.html)

# Convolutional Neural Networks: Strided Convolution



Strided convolution gif from [http://deeplearning.net/software/theano/tutorial/conv\\_arithmetic.html](http://deeplearning.net/software/theano/tutorial/conv_arithmetic.html)

# Convolutional Neural Networks: Dilated Convolution



Dilated convolution gif from [http://deeplearning.net/software/theano/tutorial/conv\\_arithmetic.html](http://deeplearning.net/software/theano/tutorial/conv_arithmetic.html)

## Convolution Output

$$H_{out} = \frac{H_{in} + 2 \times \text{padding}[0] - \text{kernel\_size}[0]}{\text{stride}[0]} + 1$$

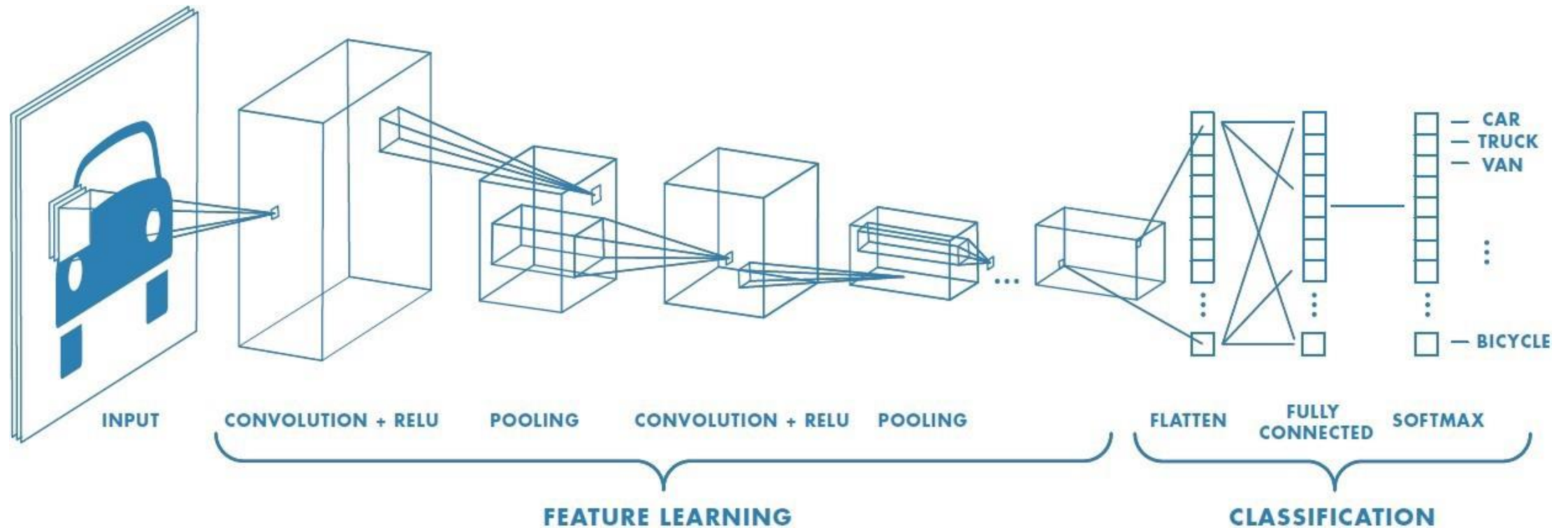
$$W_{out} = \frac{W_{in} + 2 \times \text{padding}[1] - \text{kernel\_size}[1]}{\text{stride}[1]} + 1$$

## Convolution Output

$$H_{out} = \frac{H_{in} + 2 \times \text{padding}[0] - \text{kernel\_size}[0]}{\text{stride}[0]} + 1$$

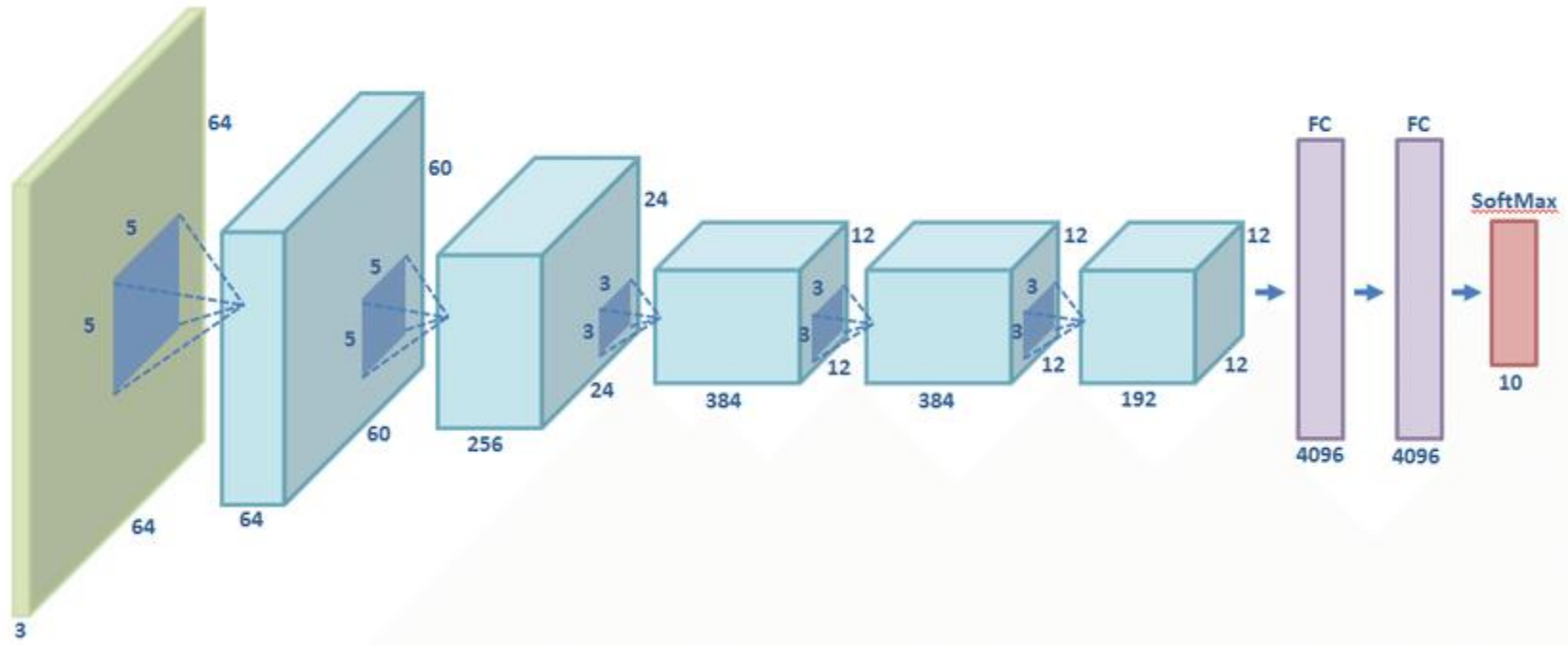
$$W_{out} = \frac{W_{in} + 2 \times \text{padding}[1] - \text{kernel\_size}[1]}{\text{stride}[1]} + 1$$

# Convolutional Neural Networks: Putting it all together



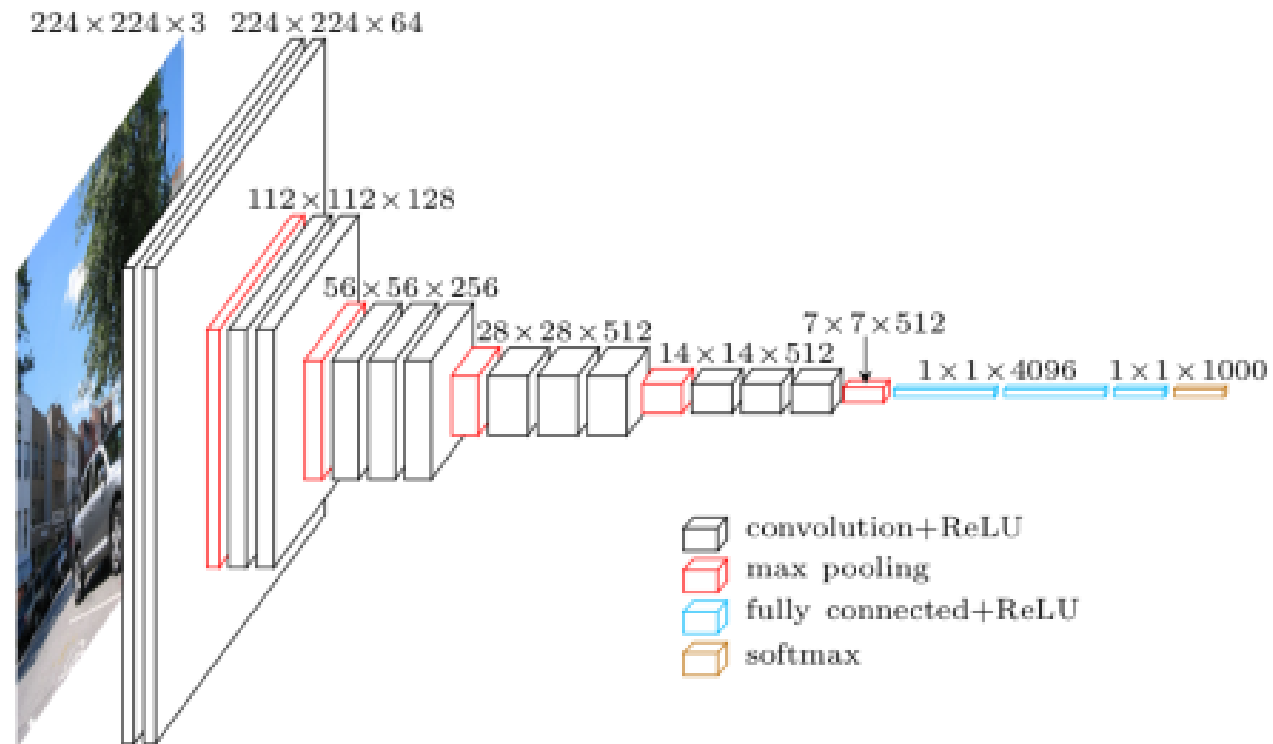
Convolutional Network from <https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53>

# Popular Conv Net: AlexNet



Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems. 2012.

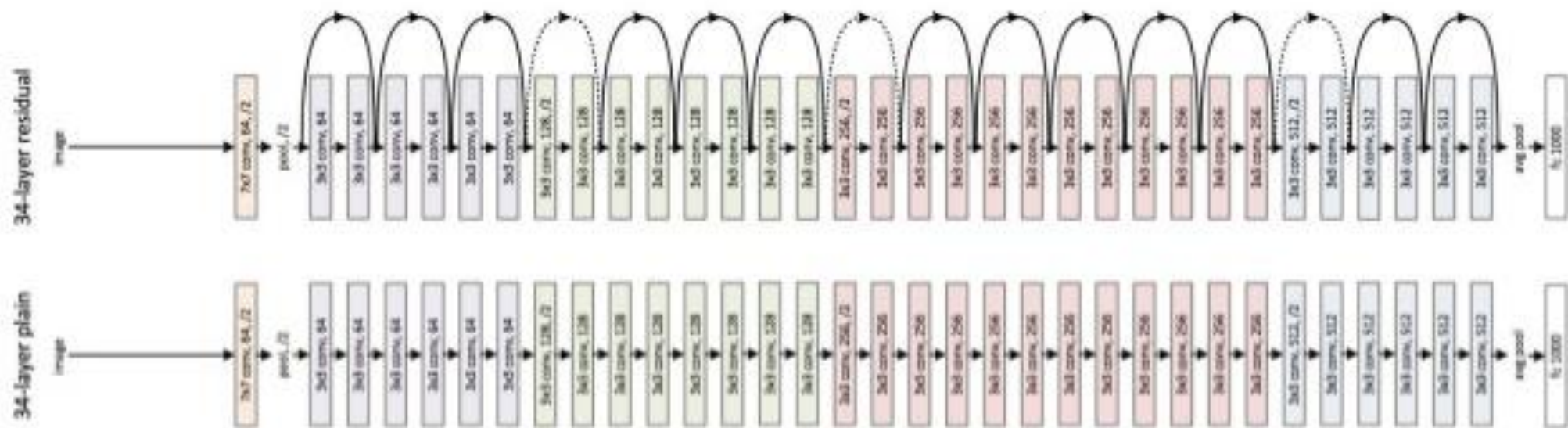
# Popular Conv Net: VGG



Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).

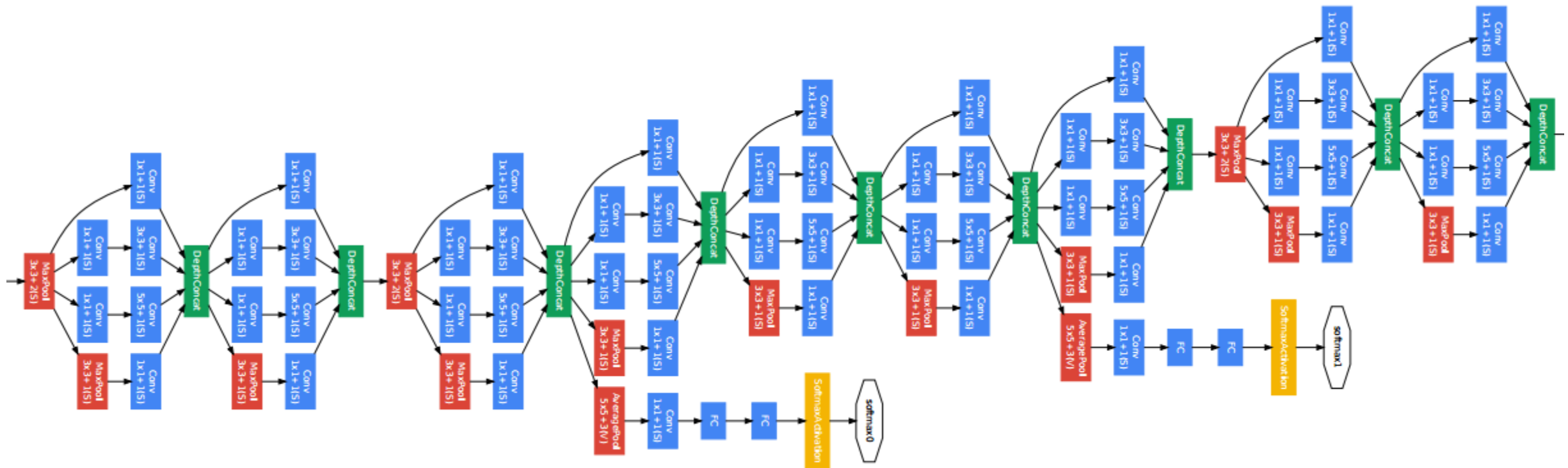


## Popular Conv Net: ResNet



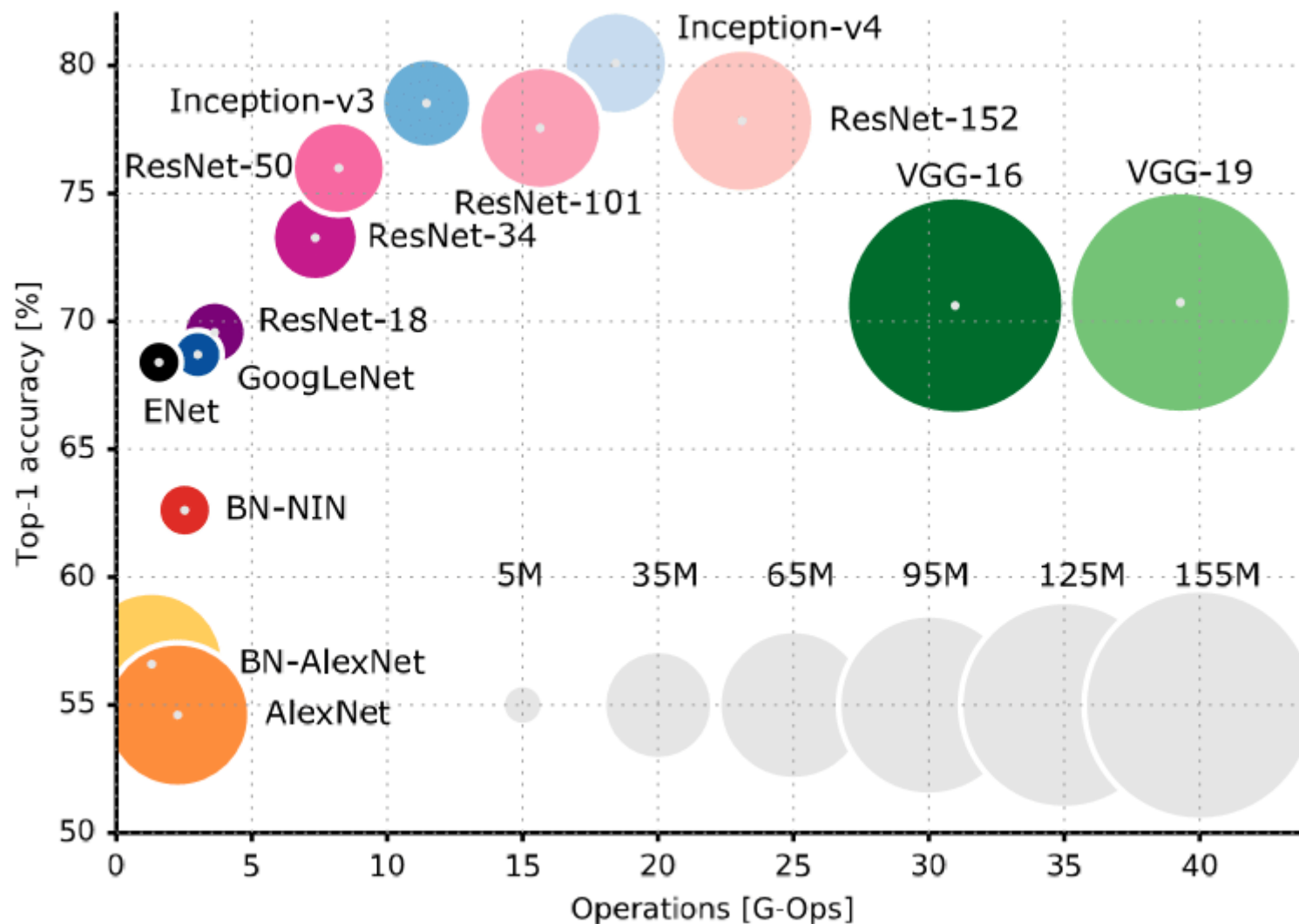
He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

# Popular Conv Net: Inception Net



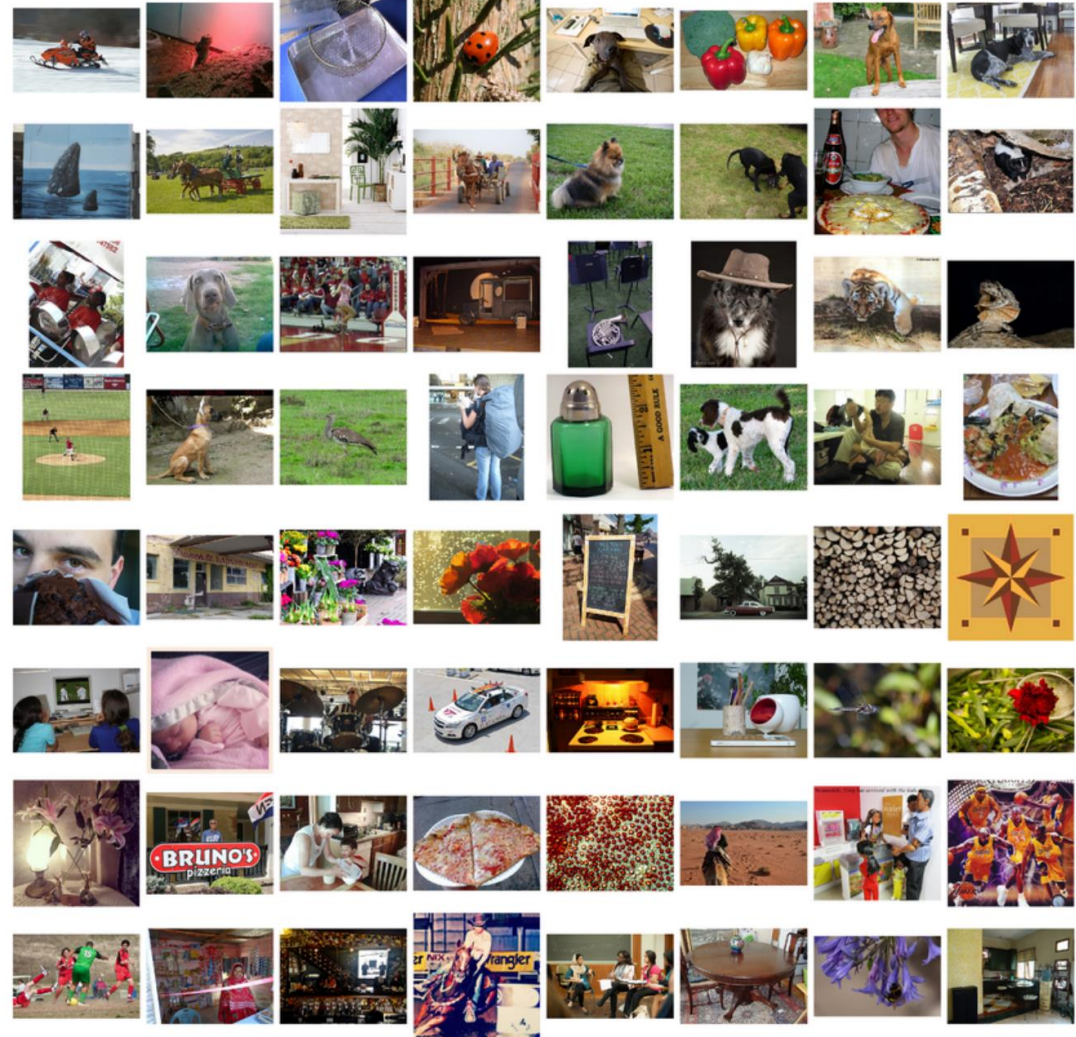
Szegedy, Christian, et al. "Going deeper with convolutions." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.

# Comparison of Networks on ImageNet Classification



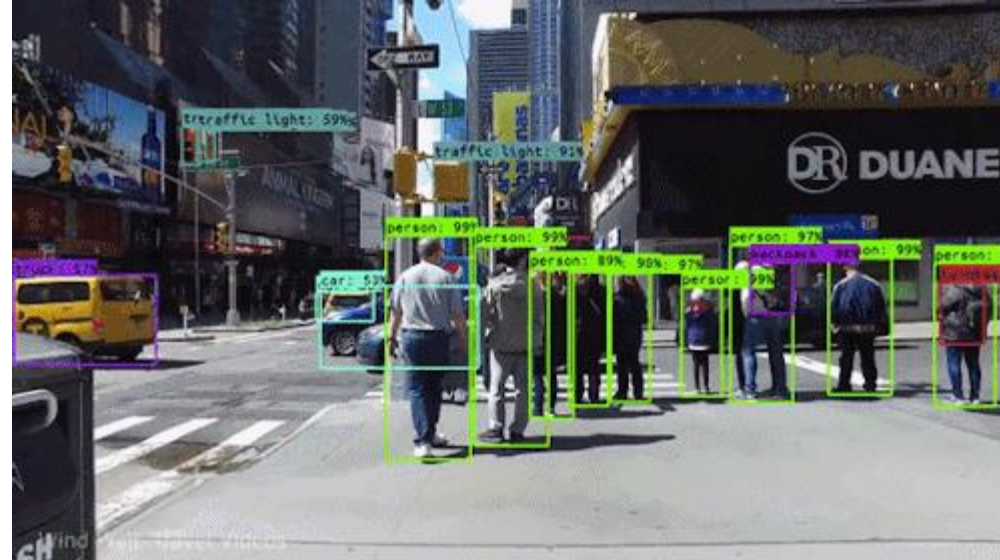
# Computer Vision Tasks: Image Classification

- ImageNet is a popular dataset
- 1.2 million training images
- 100 thousand testing images
- 1000 classes



# Computer Vision Tasks: Object Detection

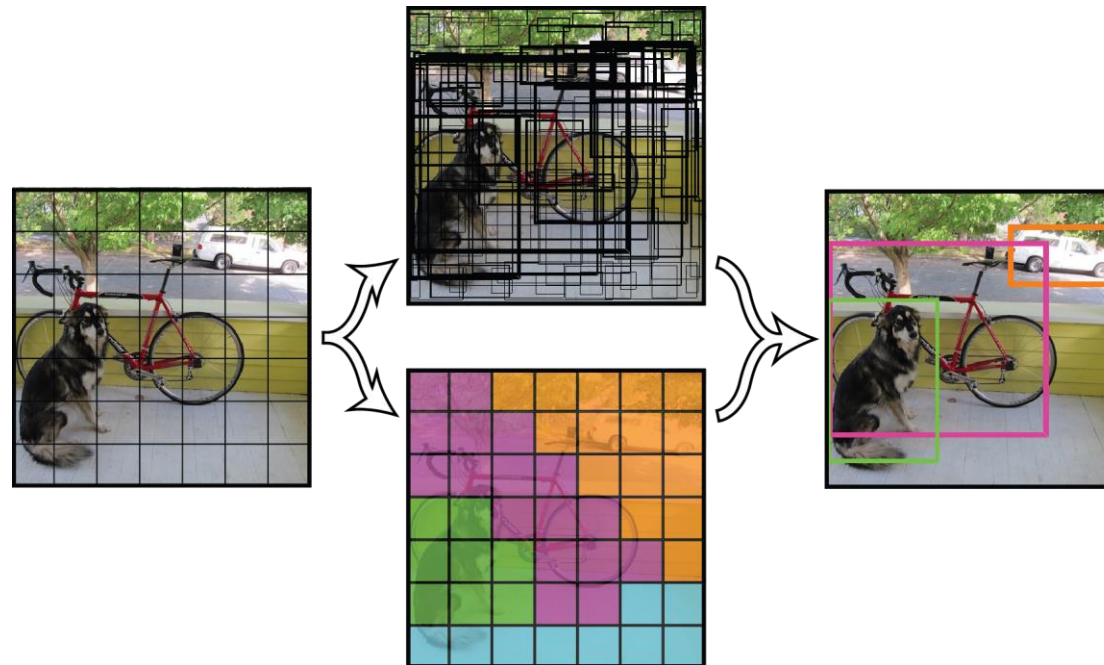
- Detect one or multiple object bounding boxes in an image
- Typically multi-task with image classification





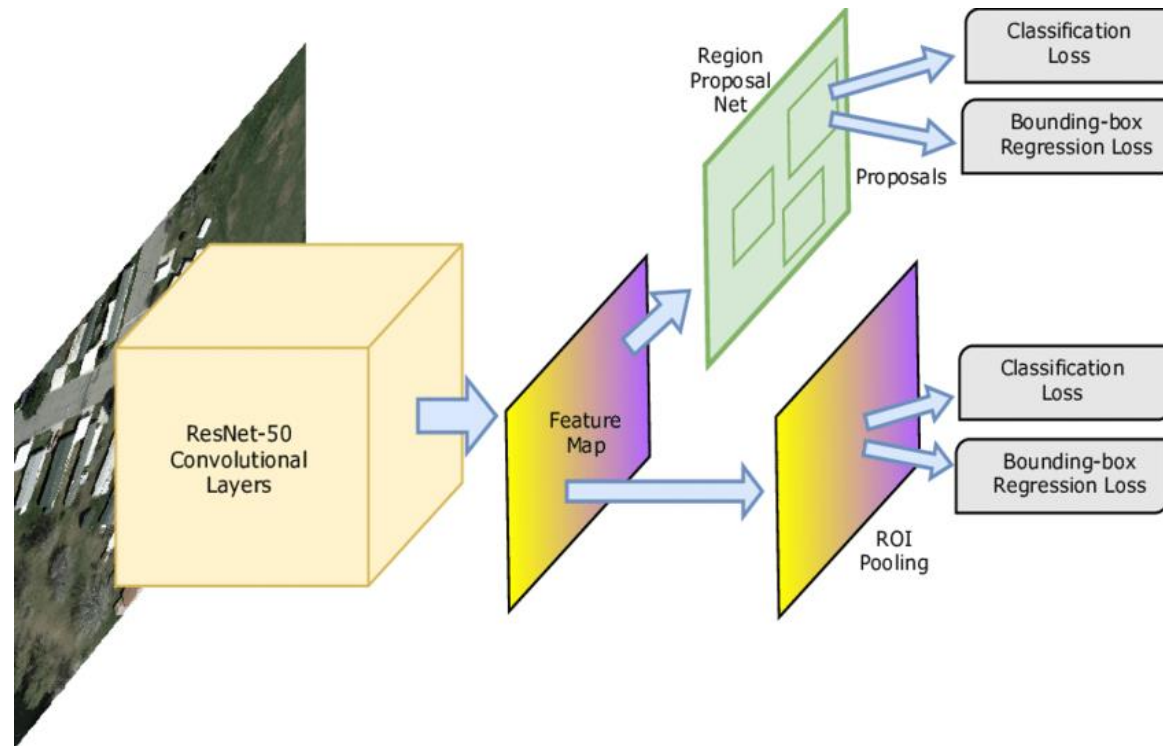
# Object Detection Methods: YOLO (You Only Look Once)

- Single-stage detector



# Object Detection Methods: Faster R-CNN

- Multi-stage detector



# Computer Vision Tasks: Segmentation



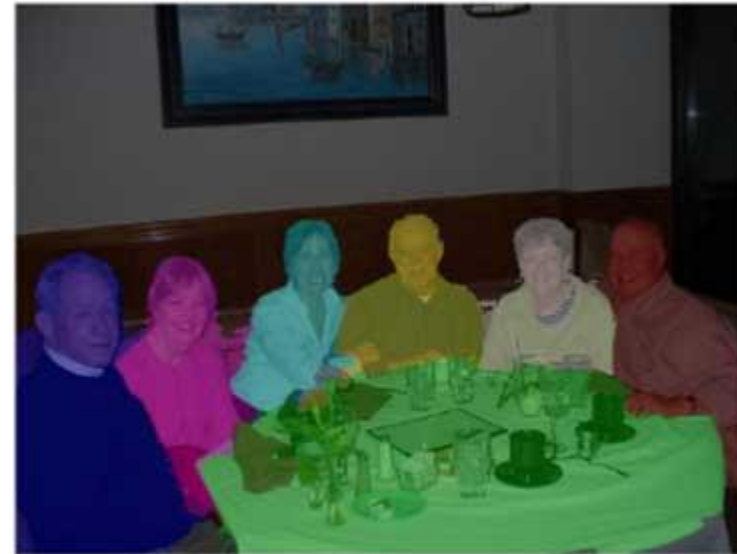


# Computer Vision Tasks: Segmentation

- **Semantic Segmentation:** Detects each pixel and overall object category
- **Instance Segmentation:** Detects each pixel and identifies individual instances of the object

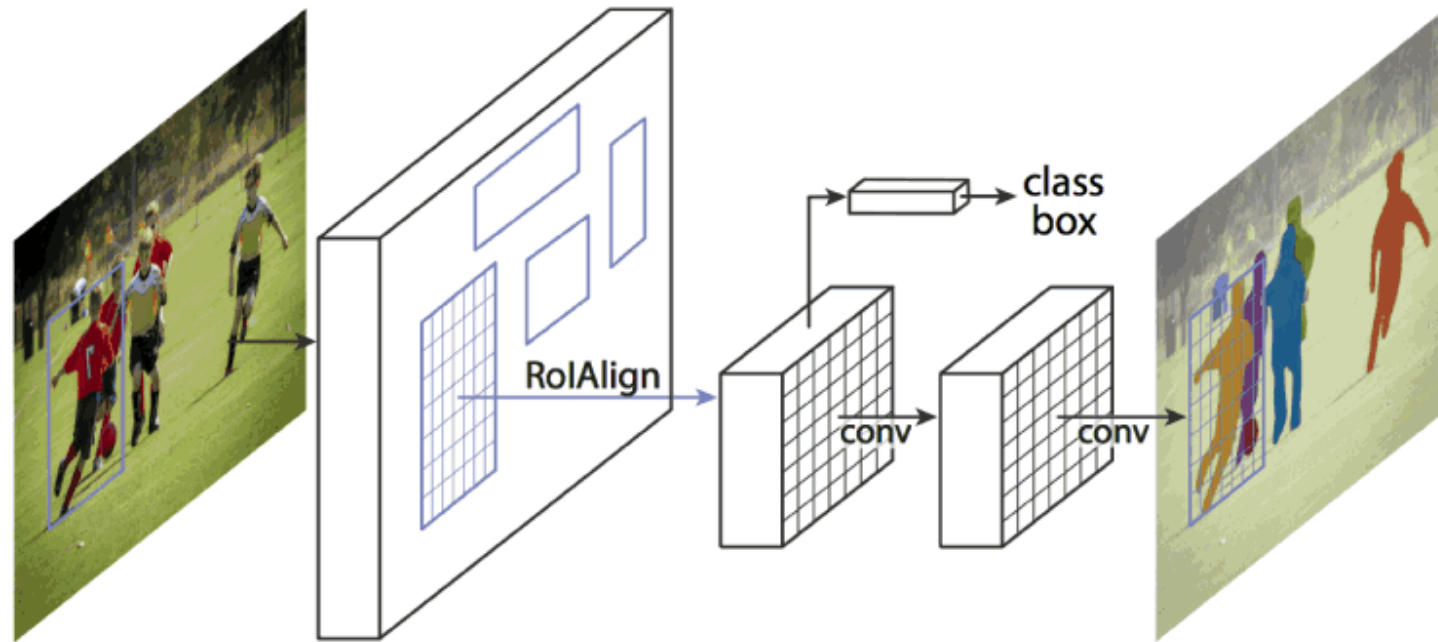


Semantic Segmentation



Instance Segmentation

# Instance Segmentation Methods: Mask R-CNN

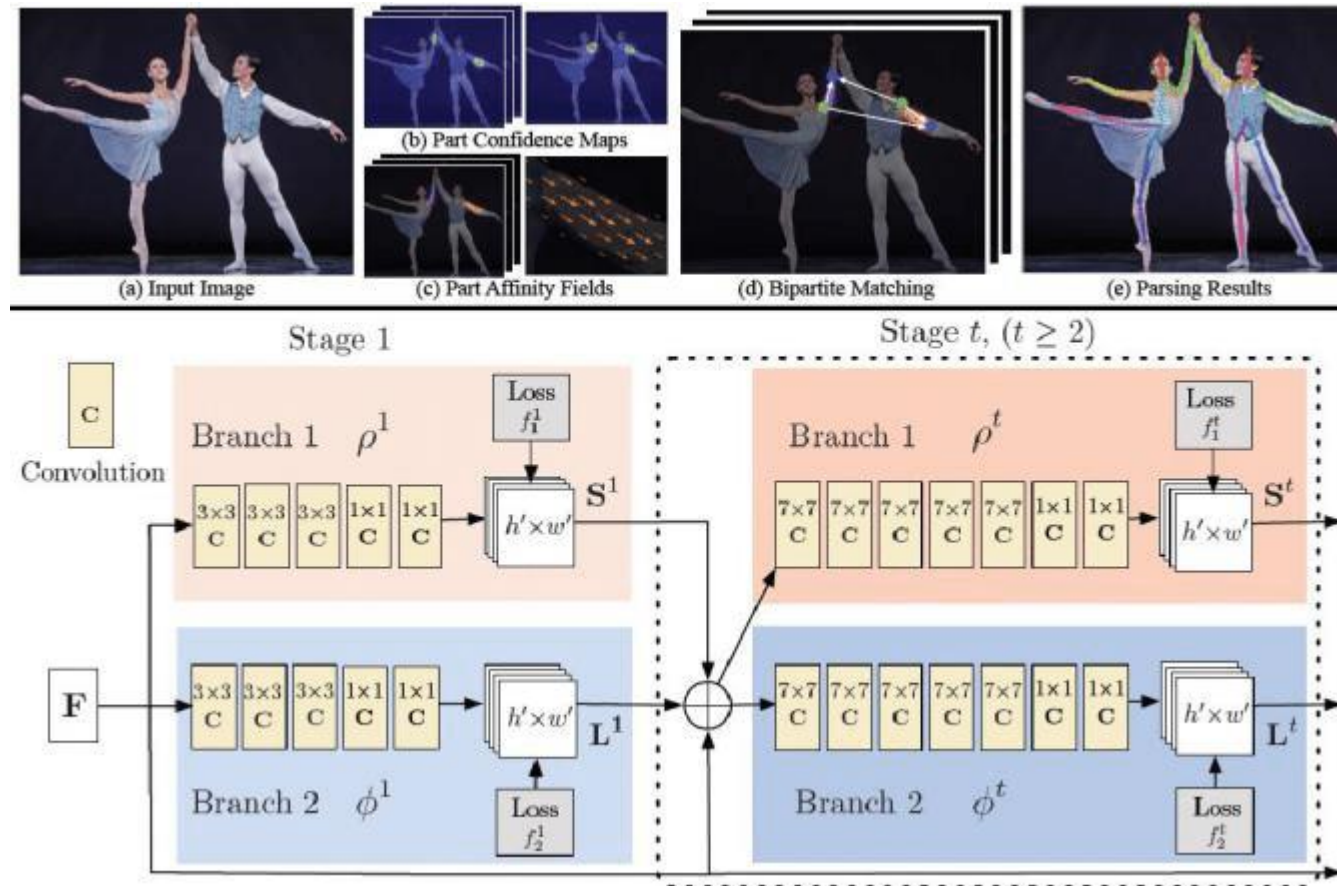


The Mask R-CNN framework for instance segmentation

# Computer Vision Tasks: Pose Estimation



# Pose Estimation Methods: OpenPose





# Computer Vision Tasks: Style Transfer

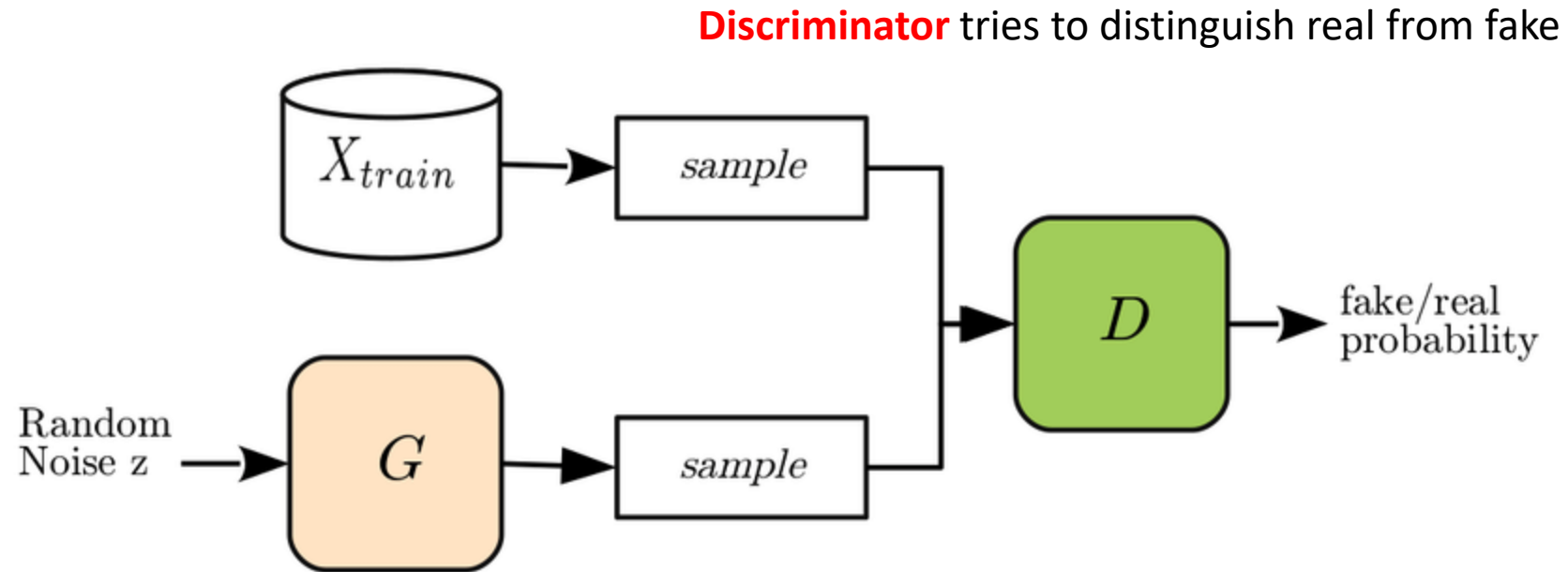


1: Gatys, Leon A., Alexander S. Ecker, and Matthias Bethge. "A neural algorithm of artistic style." arXiv preprint arXiv:1508.06576 (2015).

# Computer Vision Tasks: Image Generation



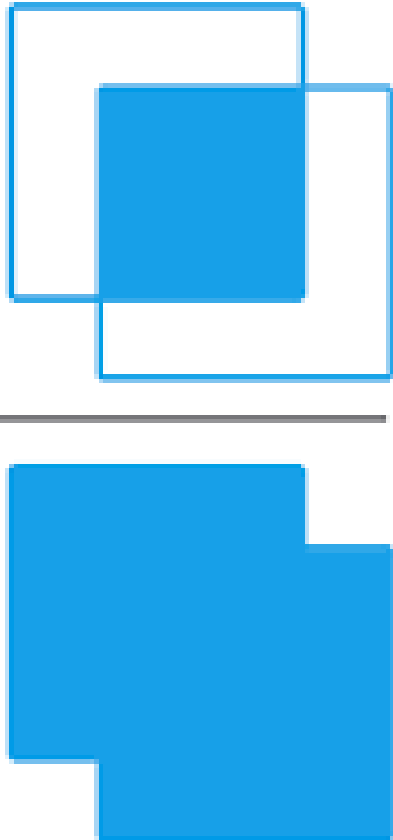
# Popular Networks: Generative Adversarial Networks



**Generator** tries to fool the **Discriminator**

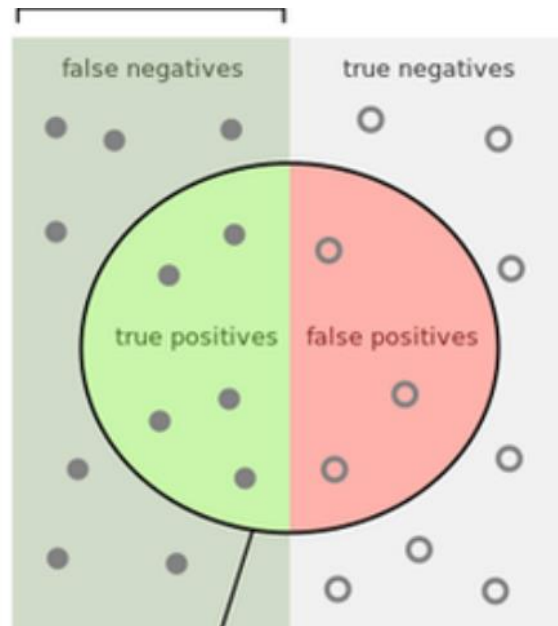


# Evaluation Metrics: Intersection over Union

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$




# Evaluation Metrics: Precision and Recall



How many selected  
items are relevant?

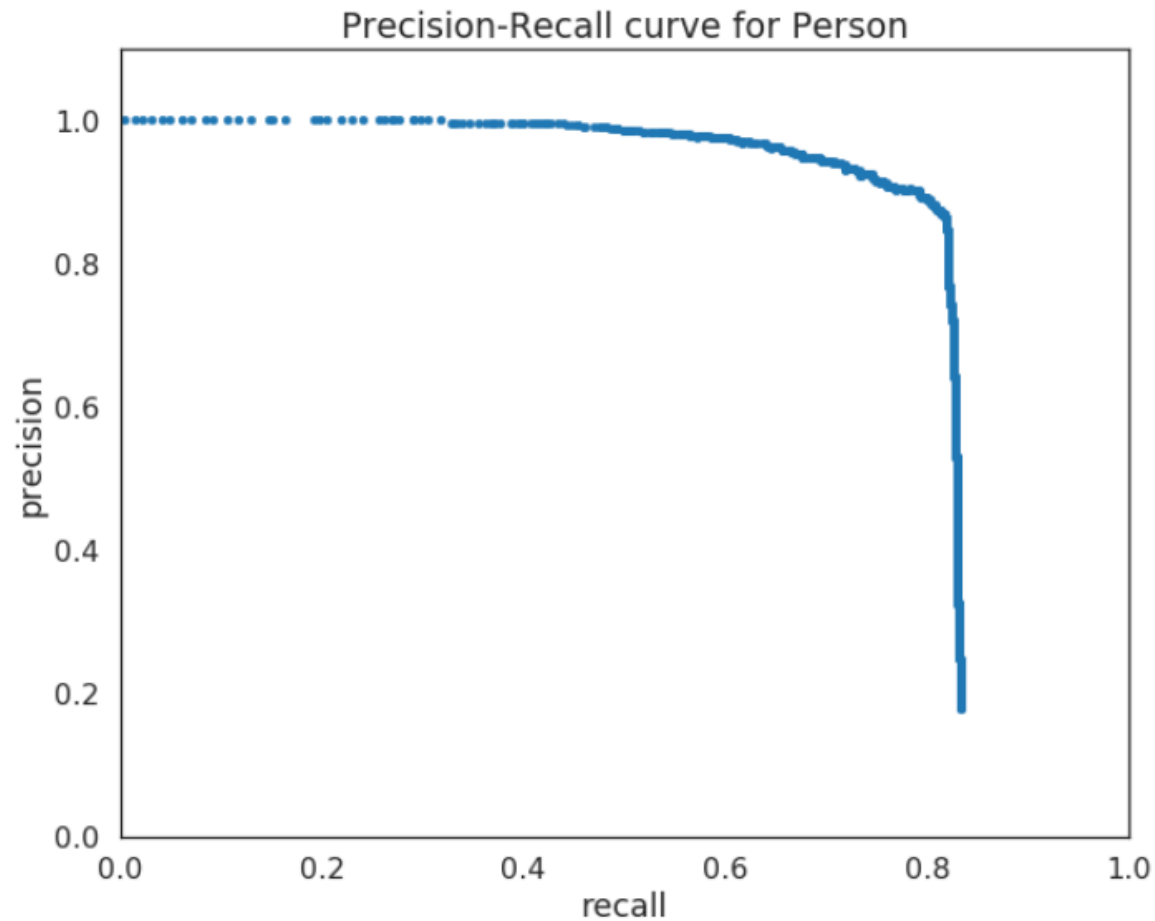
$$\text{Precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}}$$

How many relevant  
items are selected?

$$\text{Recall} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$

# Evaluation Metrics: Precision and Recall

- Plot for varying positive prediction thresholds



# Evaluation Metrics: F1 Score

- Single metric considering both precision and recall
- Good for single metric evaluation
- Gives equal importance to precision and recall (this may cause issues when you regard precision more over recall or vice versa)

$$F1 = 2 \left( \frac{precision \times recall}{precision + recall} \right)$$

Questions?