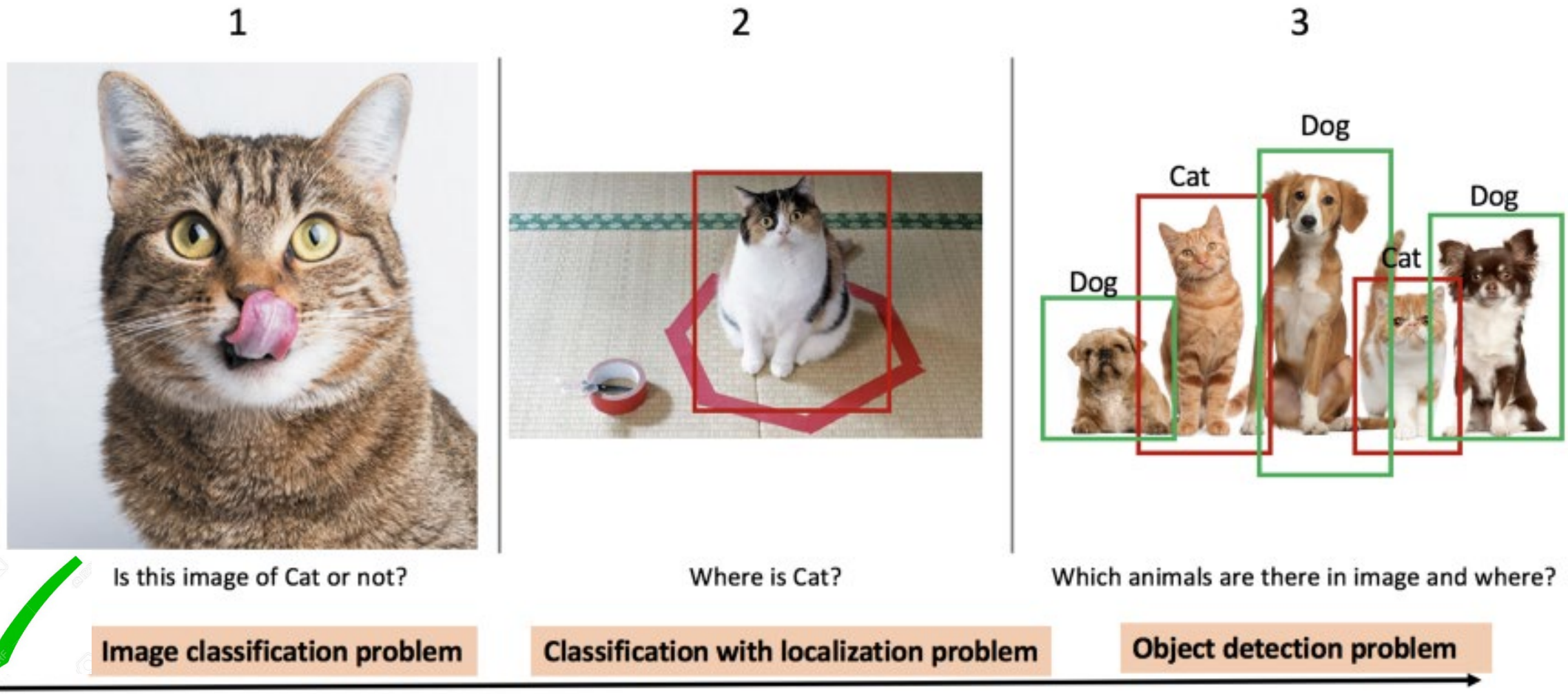
Deep Learning and Convolutional Neural Network (42028)

Object Detection- 1

Introduction



Images source: https://towardsdatascience.com/evolution-of-object-detection-and-localization-algorithms-e241021d8bad

Datasets and Performance Metrics

- The **PASCAL Visual Object Classification (PASCAL VOC)** is a popular dataset for object detection, classification and segmentation.
- 20 categories
- Link: http://host.robots.ox.ac.uk/pascal/VOC/

- **ImageNet** has released an object detection dataset in 2013
- Train set: 500,000 images, 200 categories.
- Not very popular due to large number of classes and dataset size!
- Large number classes complicates the task

Images source: https://towardsdatascience.com/evolution-of-object-detection-and-localization-algorithms-e241021d8bad

Datasets and Performance Metrics

Dataset Comparison

		PASCAL VOC 2012	ILSVRC 2014
Number of object classes		20	**200**
Training	Num images	5717	456567
	Num objects	13609	478807
Validation	Num images	5823	20121
	Num objects	13841	55502
Testing	Num images	10991	40152
	Num objects	---	---

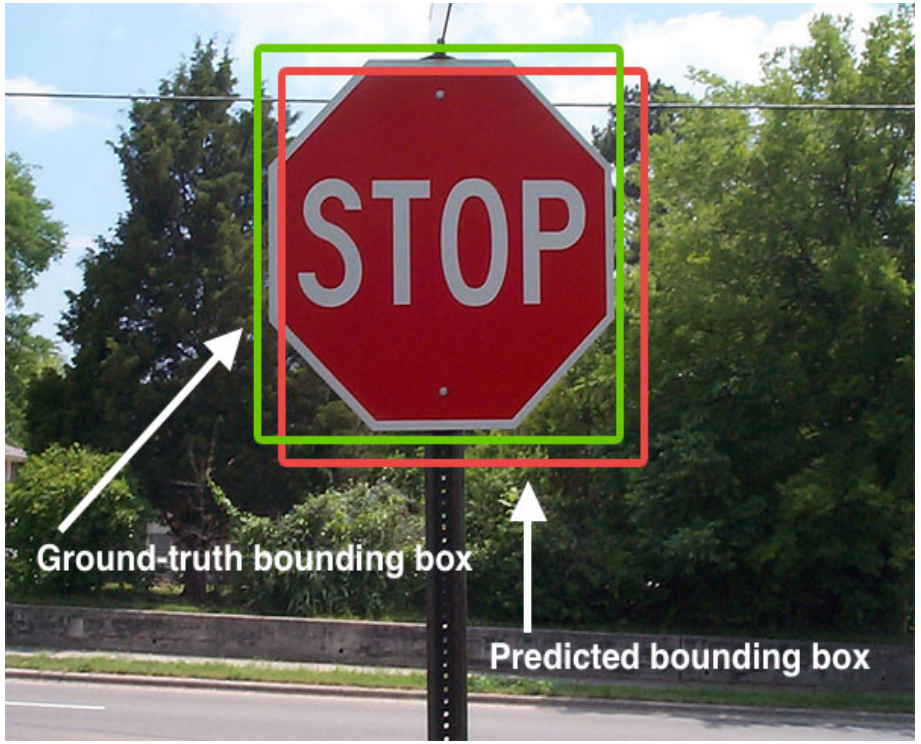
Images source: http://image-net.org/challenges/LSVRC/2014/

Datasets and Performance Metrics

- **Intersection over Union (IoU):**

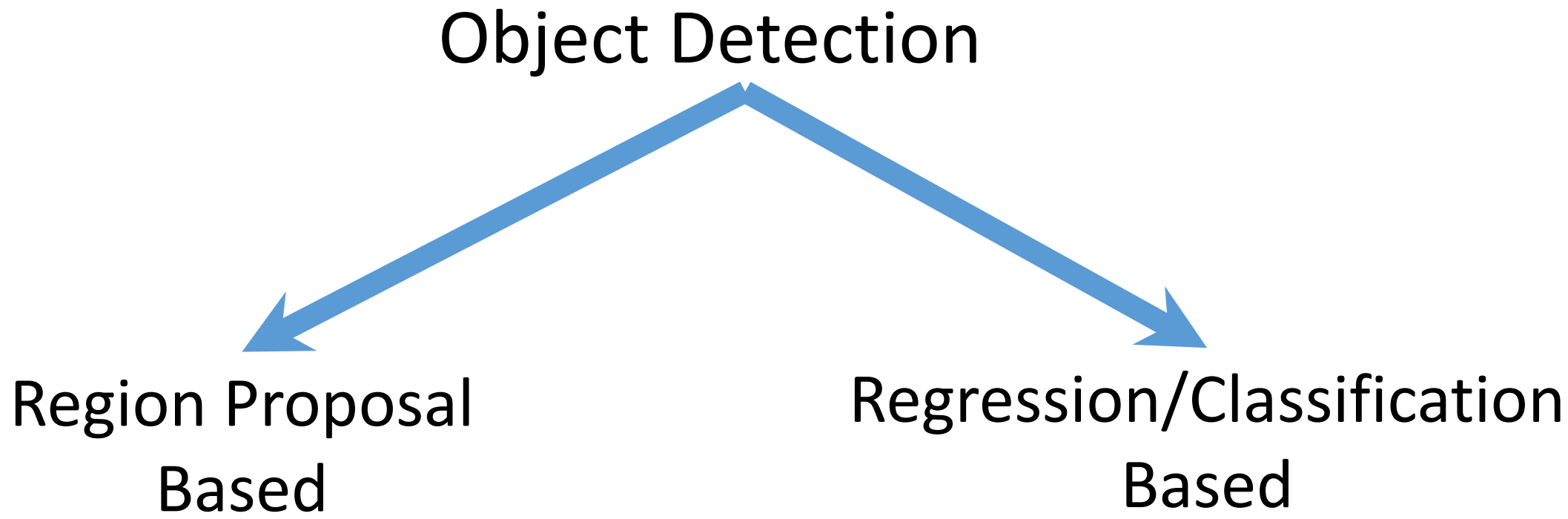
Intersection over Union is a metric used for the evaluation of an object detector, i.e. how good is the predicted bounding box for an object detected closely matches

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



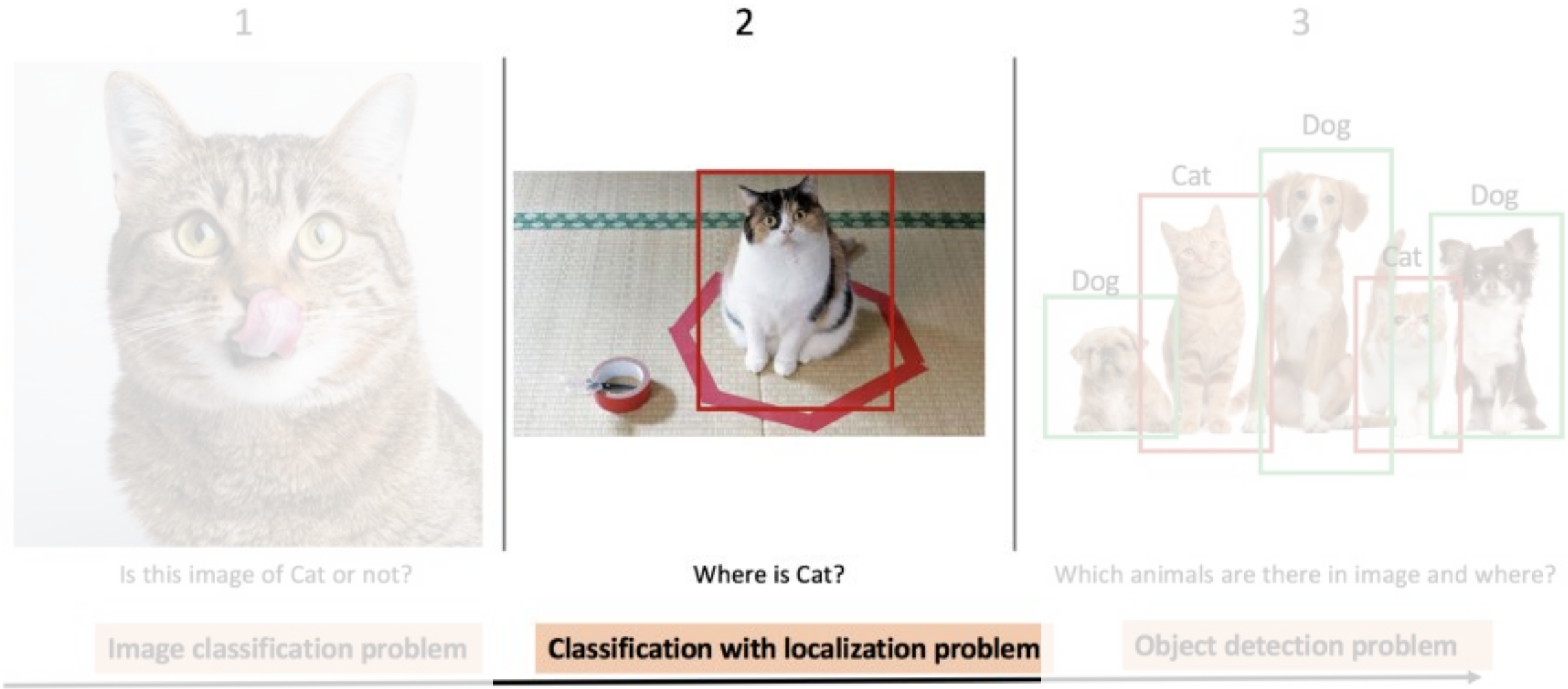
Reference: https://www.pyimagesearch.com/2016/11/07/intersection-over-union-iou-for-object-detection/

Current frameworks



Images source: https://towardsdatascience.com/evolution-of-object-detection-and-localization-algorithms-e241021d8bad

Task: Classification with Localization



Images source: https://towardsdatascience.com/evolution-of-object-detection-and-localization-algorithms-e241021d8bad

Task: Classification with Localization

Classification Task:

Input : Image

Output: Label

Performance Evaluation: Accuracy

Output : Dog

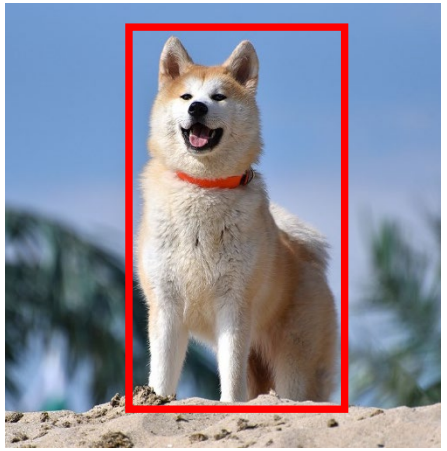
Localization Task:

Input : Image

Output: Bounding Box in the image

(x, y, Ht, Wd) or (x, y, x', y')

Performance Evaluation: IoU



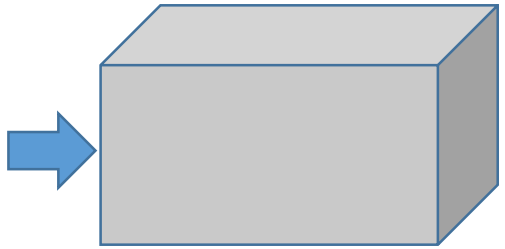
Output : (x, y, Ht, Wd)

Source and Reference: http://cs231n.stanford.edu/slides/2016/winter1516_lecture8.pdf

Image source: https://en.wikipedia.org/wiki/Akita_(dog)

Localization as a regression problem

Input Image



CNN

Output : 4 numbers
(x', y', Ht', Wd')

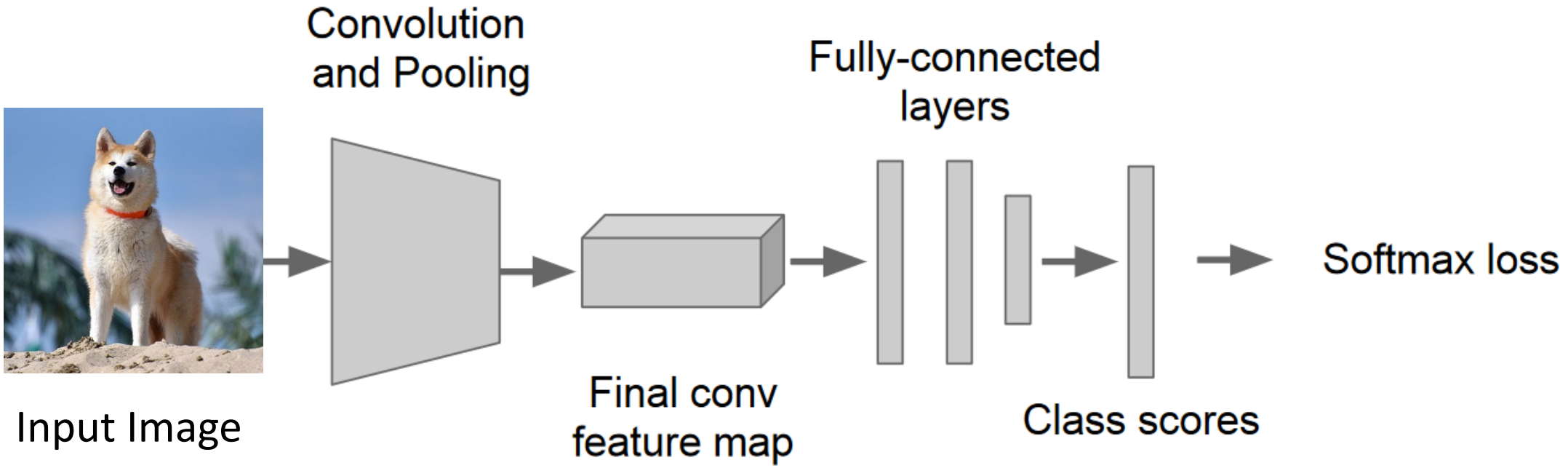
Ground Truth: 4 numbers
(x, y, Ht, Wd)

Calculate Loss
L2 Loss

Source and Reference: http://cs231n.stanford.edu/slides/2016/winter1516_lecture8.pdf

Image source: https://en.wikipedia.org/wiki/Akita_(dog)

Localization as a regression problem



Source and Reference: http://cs231n.stanford.edu/slides/2016/winter1516_lecture8.pdf

Image source: https://en.wikipedia.org/wiki/Akita_(dog)

Localization as a regression problem

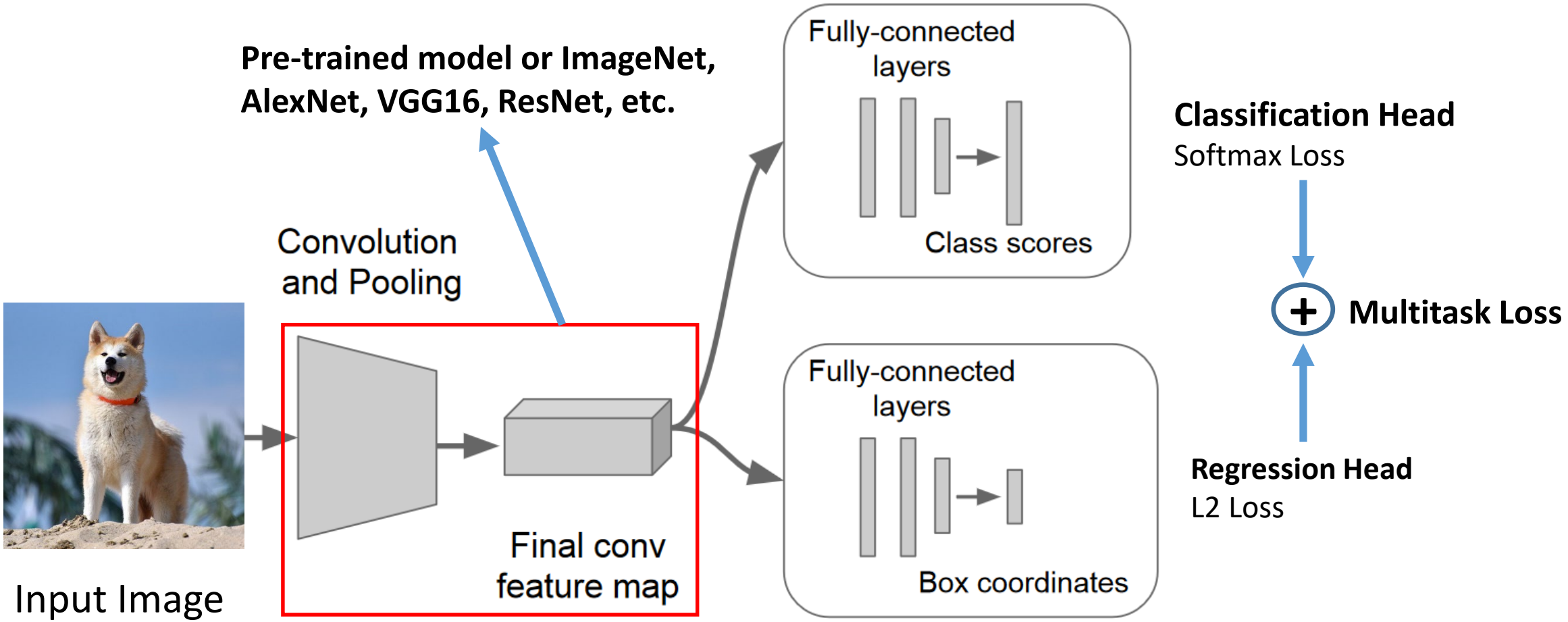
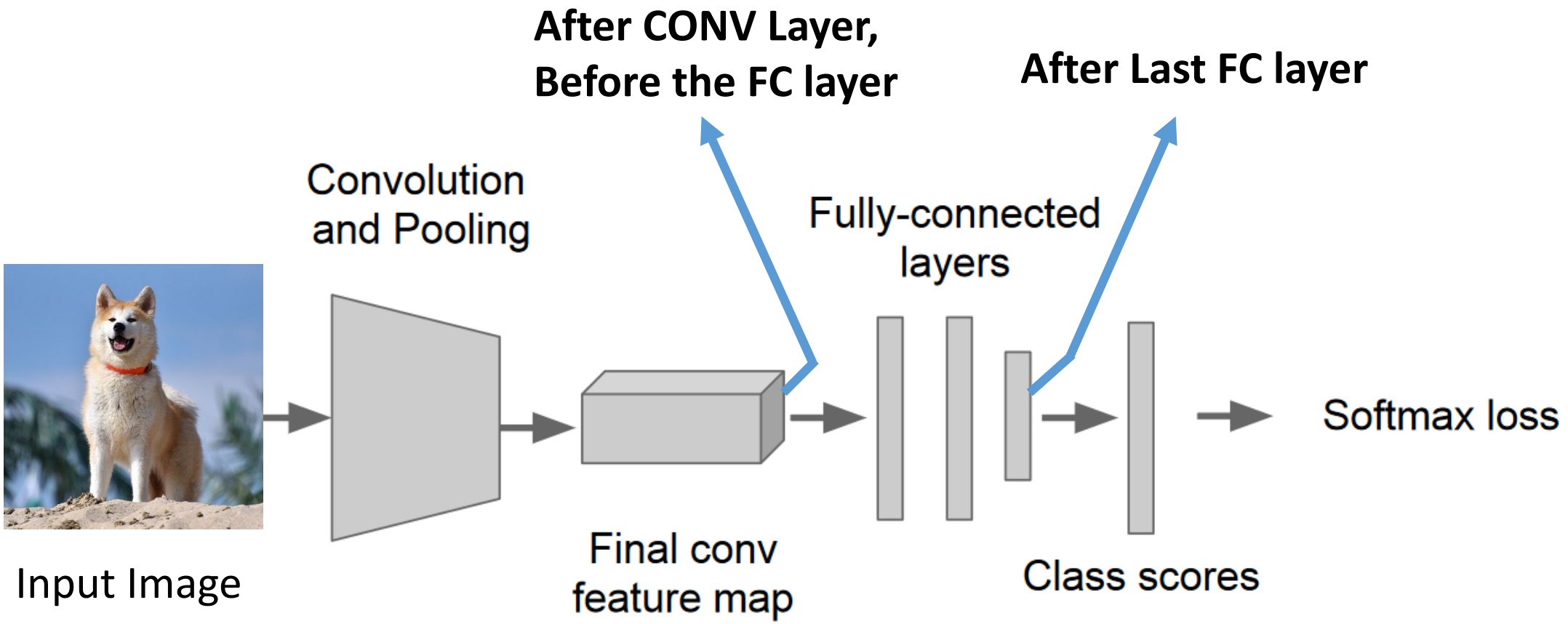


Image Source and Reference: http://cs231n.stanford.edu/slides/2016/winter1516_lecture8.pdf

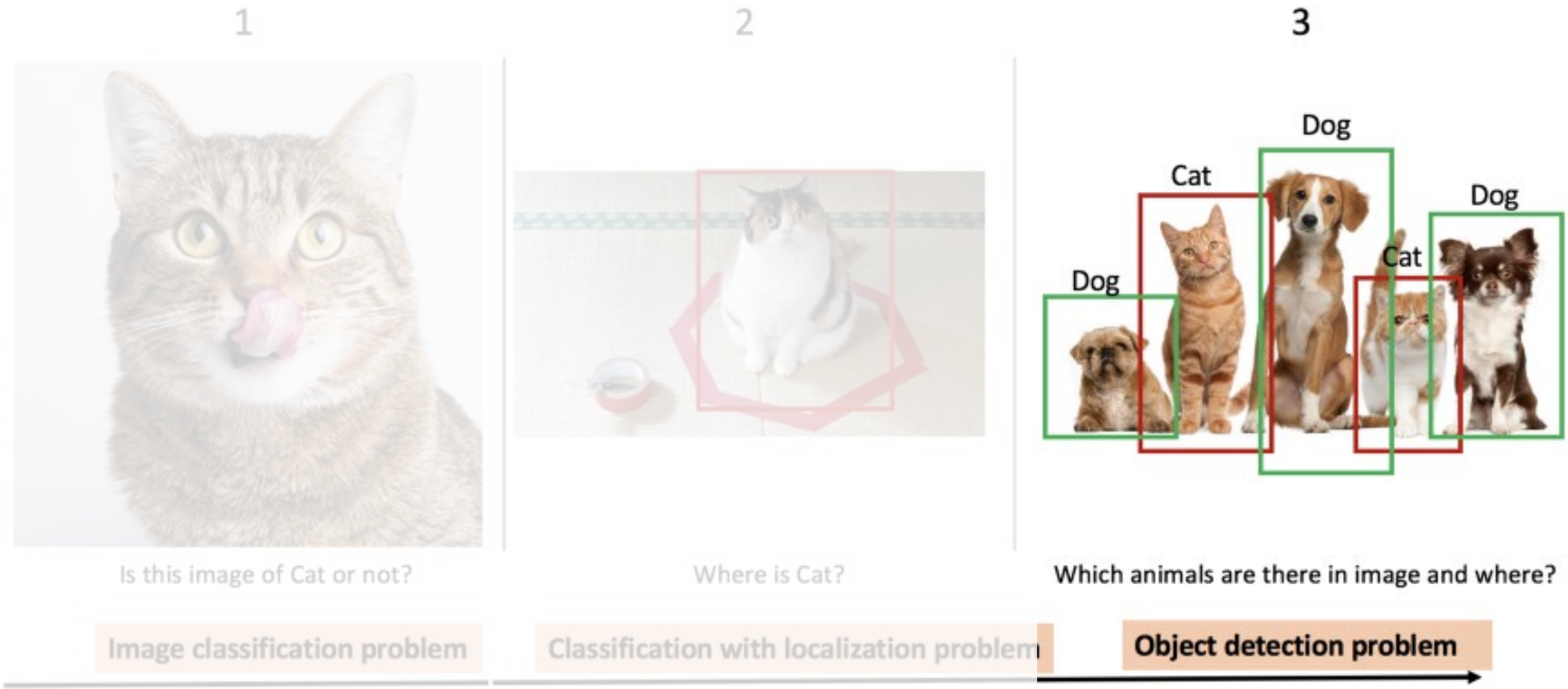
Potential locations for Regression head in CNN



Source and Reference: http://cs231n.stanford.edu/slides/2016/winter1516_lecture8.pdf

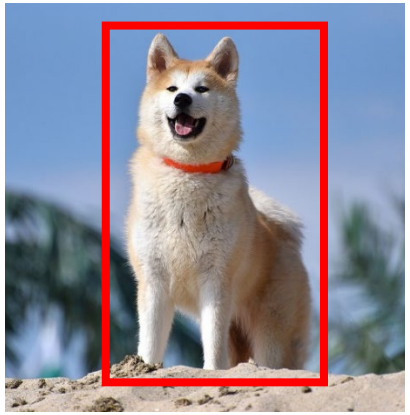
Image source: https://en.wikipedia.org/wiki/Akita_(dog)

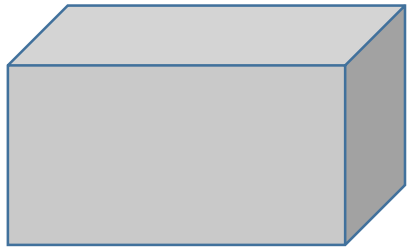
Task: Object Detection Problem



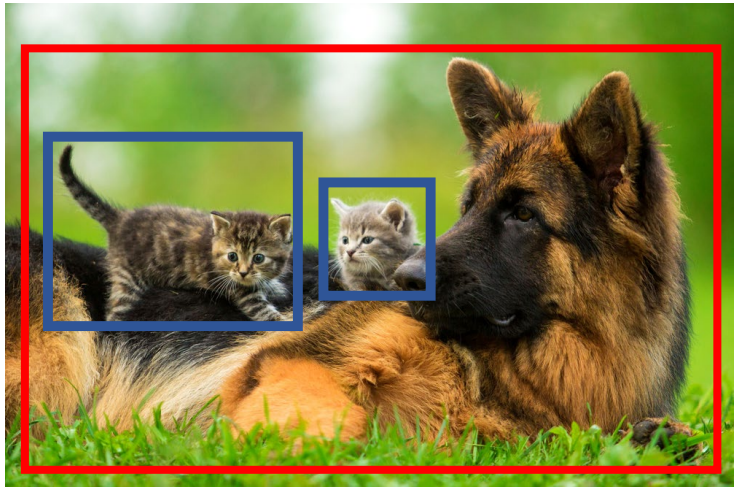
Images source: https://towardsdatascience.com/evolution-of-object-detection-and-localization-algorithms-e241021d8bad

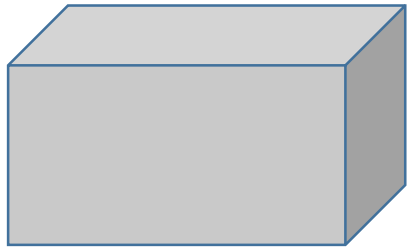
Detection as a regression problem





Output : Dog, (x, y, Ht, Wd)





Output :
Dog, (x, y, Ht, Wd)
Cat, (x, y, Ht, Wd)
Cat, (x, y, Ht, Wd)

Source and Reference: http://cs231n.stanford.edu/slides/2016/winter1516_lecture8.pdf

Image source: https://en.wikipedia.org/wiki/Akita_(dog)
http://www.kamloopspropertyforsale.com/blog/does-owning-a-cat-depreciate-your-home.html

Detection as a classification problem

1. **Apply Sliding Window technique**
2. **Apply CNN to different Windows and get a prediction**

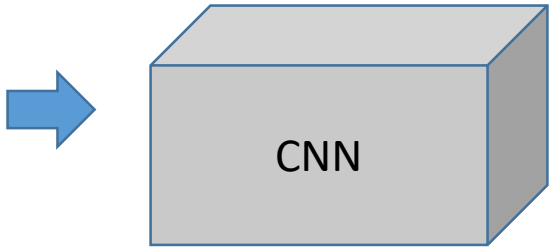
Output :
Dog? No
Cat? No
Background? Yes

Source and Reference: http://cs231n.stanford.edu/slides/2016/winter1516_lecture8.pdf

Image source: https://en.wikipedia.org/wiki/Akita_(dog)
http://www.kamloopspropertyforsale.com/blog/does-owning-a-cat-depreciate-your-home.html

Detection as a classification problem

1. **Apply Sliding Window technique**
2. **Apply CNN to different Windows and get a prediction**



Output :
Dog? No
Cat? Yes
Background? No

Source and Reference: http://cs231n.stanford.edu/slides/2016/winter1516_lecture8.pdf

Image source: https://en.wikipedia.org/wiki/Akita_(dog)
http://www.kamloopspropertyforsale.com/blog/does-owning-a-cat-depreciate-your-home.html

Detection as a classification problem

1. **Apply Sliding Window technique**
2. **Apply CNN to different Windows and get a prediction**

Output :
Dog? No
Cat? Yes
Background? No

Source and Reference: http://cs231n.stanford.edu/slides/2016/winter1516_lecture8.pdf

Image source: https://en.wikipedia.org/wiki/Akita_(dog)
http://www.kamloopspropertyforsale.com/blog/does-owning-a-cat-depreciate-your-home.html

Detection as a classification problem

Issue with Sliding Window technique

1. Apply CNN on large number of windows
2. Multiple scale and locations of windows
3. Inaccurate bounding boxes
4. Computationally expensive

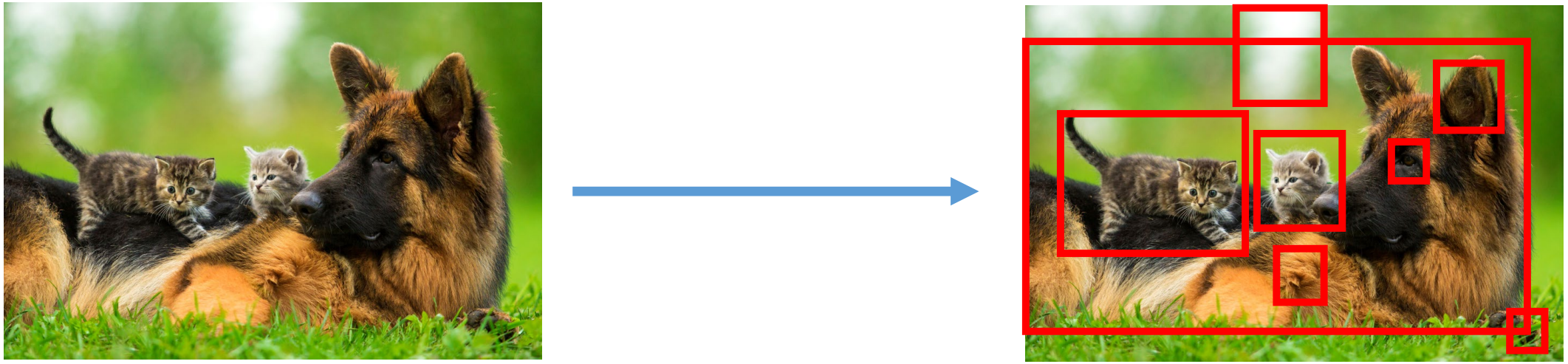
Source and Reference: http://cs231n.stanford.edu/slides/2016/winter1516_lecture8.pdf

Image source: https://en.wikipedia.org/wiki/Akita_(dog)
http://www.kamloopspropertyforsale.com/blog/does-owning-a-cat-depreciate-your-home.html

Detection as a classification problem

Region Proposal Technique:

- Find blobs in the image that are most likely to contain objects
- E.g: Selective search → ~1000-2000 region proposals using CPU!

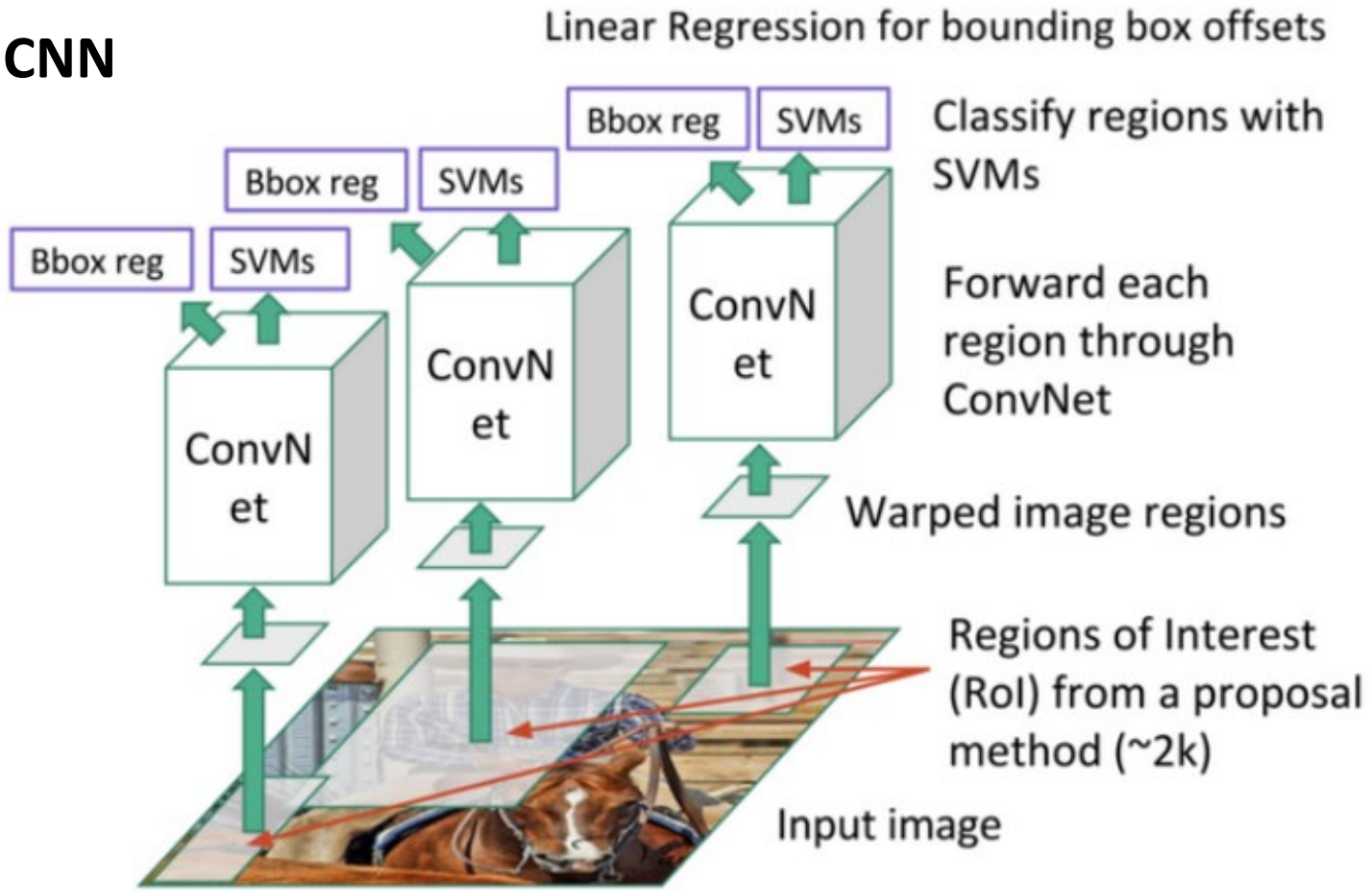


Source and Reference: http://cs231n.stanford.edu/slides/2016/winter1516_lecture8.pdf

Case Study: R-CNN

R-CNN: Region based CNN

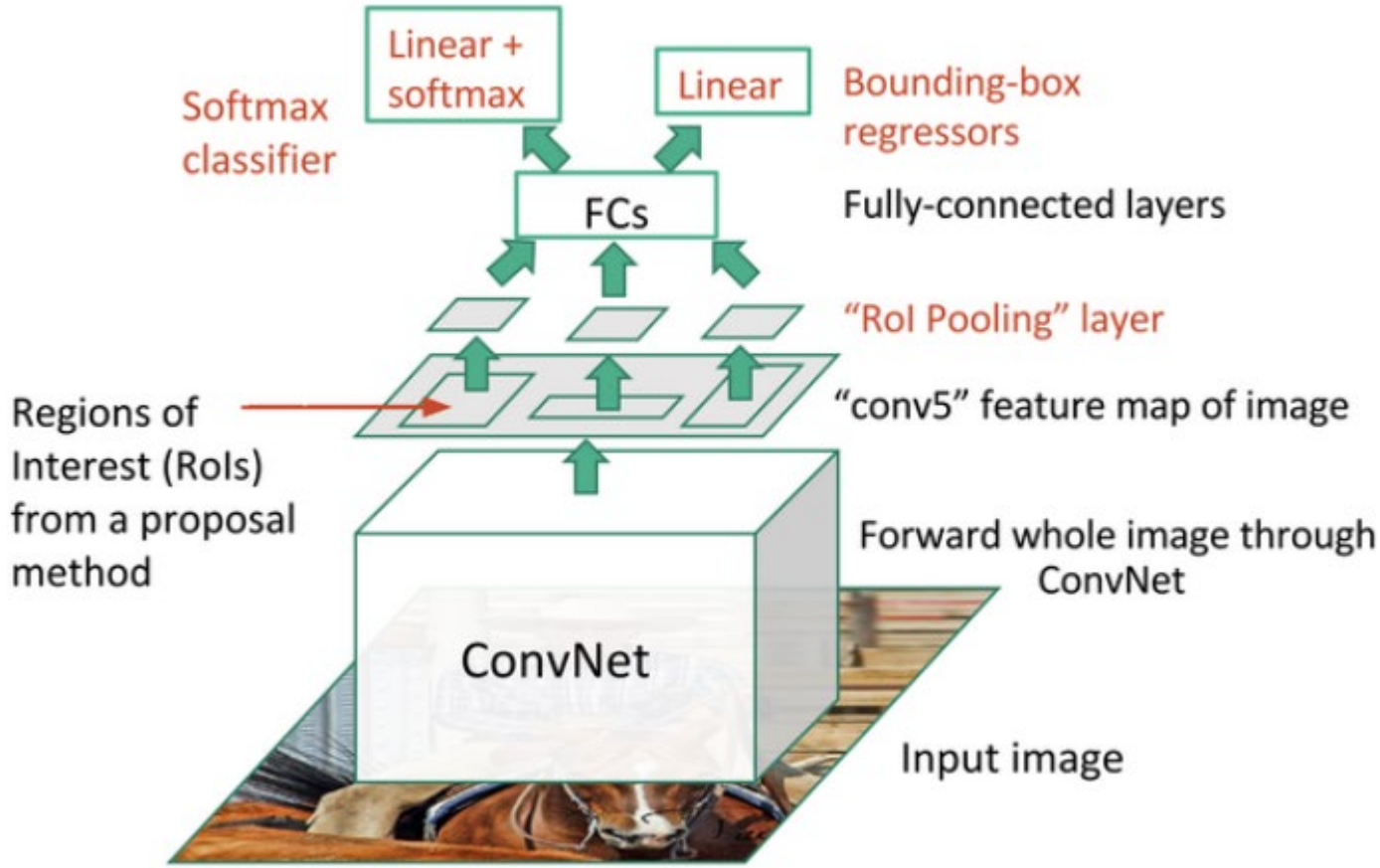
1. Resized to match the input to CNN requiremen.
2. mAP: 62.4% for 2007 PASCAL VOC



Source and Reference: http://cs231n.stanford.edu/slides/2016/winter1516_lecture8.pdf
https://medium.com/zylapp/review-of-deep-learning-algorithms-for-object-detection-c1f3d437b852

Case Study: FAST- R-CNN

1. Reduce computation
2. ROIs from feature maps using selective search
3. mAP: 70% for 2007 PASCAL VOC

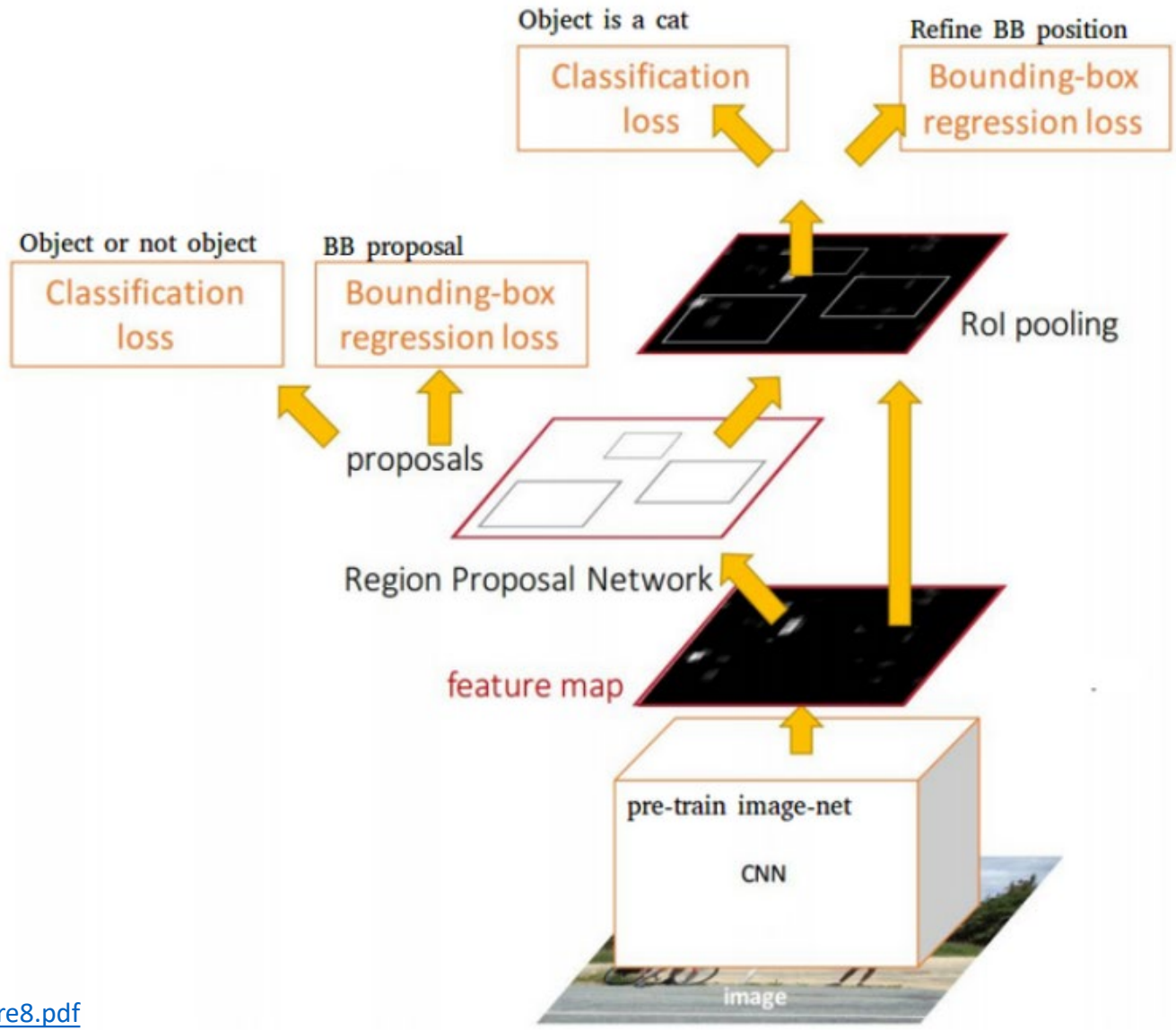


Source and Reference: http://cs231n.stanford.edu/slides/2016/winter1516_lecture8.pdf
https://medium.com/zylapp/review-of-deep-learning-algorithms-for-object-detection-c1f3d437b852

Case Study: FASTER- R-CNN

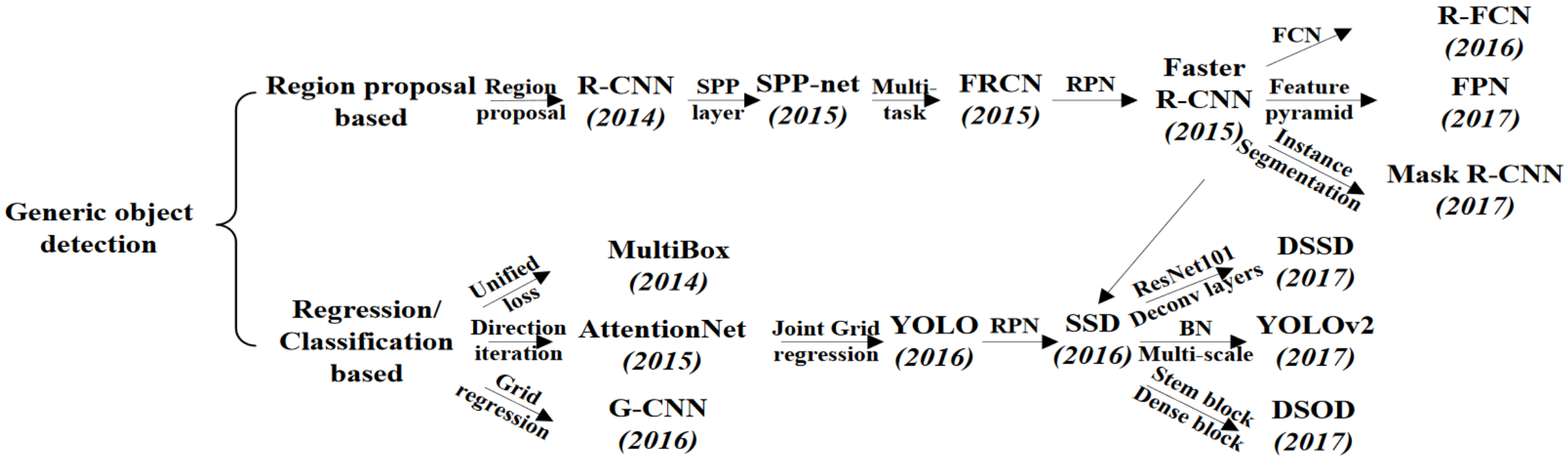
1. Introduced RPN (Region Proposal Network)

2. mAP: 78.8% for 2007 PASCAL VOC



Source and Reference: http://cs231n.stanford.edu/slides/2016/winter1516_lecture8.pdf
https://medium.com/zylapp/review-of-deep-learning-algorithms-for-object-detection-c1f3d437b852

Object Detection Techniques History



Images source: https://arxiv.org/pdf/1807.05511.pdf