“Parts Recognition” CV Model Design Deep-Dive

Mar 5, 2019

Ron

Topics

First Part

- Scope, Design Goal
- Object Detection Algorithms
- Algorithm Comparison, Design Choice

Second Part

- Revisit Design Goal
- Semantic Segmentation Algorithms

Third Part

- Action Plan
- Open Domain Data
- Tools, Frameworks
- Custom Vision

Scope

- “Parts Recognition”
  - Where? -> recognize key parts in a given image -> we will provide a “base model”
  - What? -> classify (label) parts -> customer will fine-tune (i.e. retrain) the model

- “Parts” examples
  - HP: Toner, Fuser, Drum unit, Roller kit, etc. inside printer
  - AHFR: parts inside HVAC equipment
  - TEL: components inside coater, developer, etching system, etc.
  - Others
    - 1st Party data
    - Semiconductor
    - Hospital Instruments
    - HVAC
    - Oil & Gas (e.g. air flight)
    - MFG (manufacturing)

output from a pre-trained model

output from customer’s re-trained model

Challenges

- Difficult to acquire data from customers at this point
- Parts images are NOT commonly accessible.
- Need to identify accurate locations of objects (a.k.a. “Parts”)
- Need to build a common model for D365 customers’ various products.

Object Detection

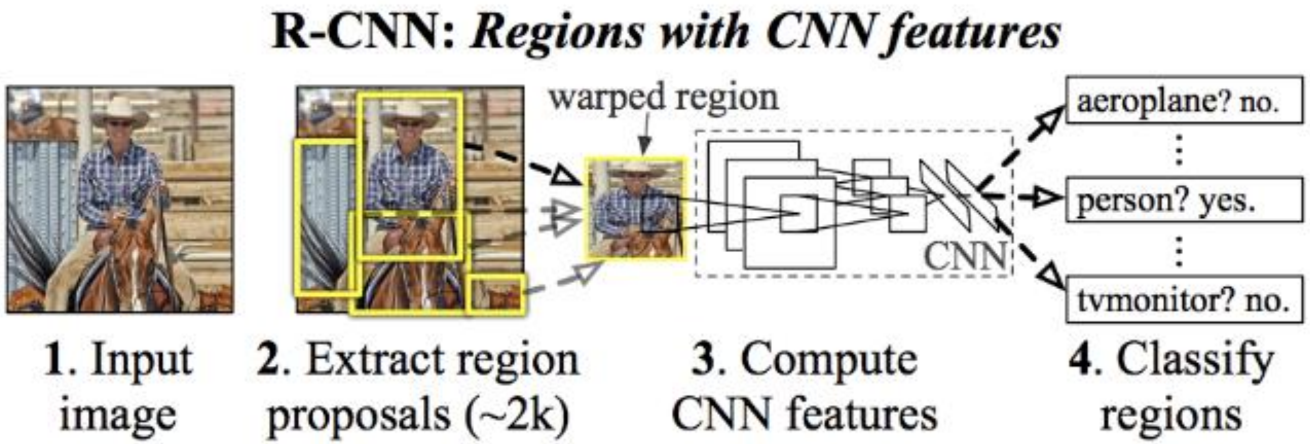
- Object Recognition
- Image Classification + Object Localization
- Localization is more important

Pre-trained model

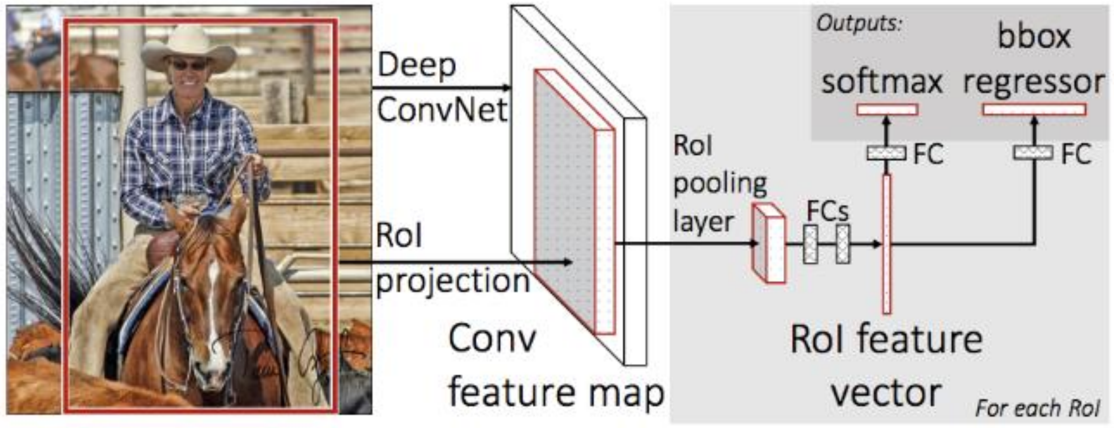
- No customer data contained
- Object Detection
- Transfer Learning
- Unsupervised Learning (in training)
- Continuous Learning through feedback loop (TBD)
- Performance bar
  - Accuracy
    - Bounding Box – **IOU** (Intersection over Union)
    - Predicted class (optional) – customer will fine-tune the model when it is retrained with labels
  - Latency
    - TBD

Algorithms for Object Detection

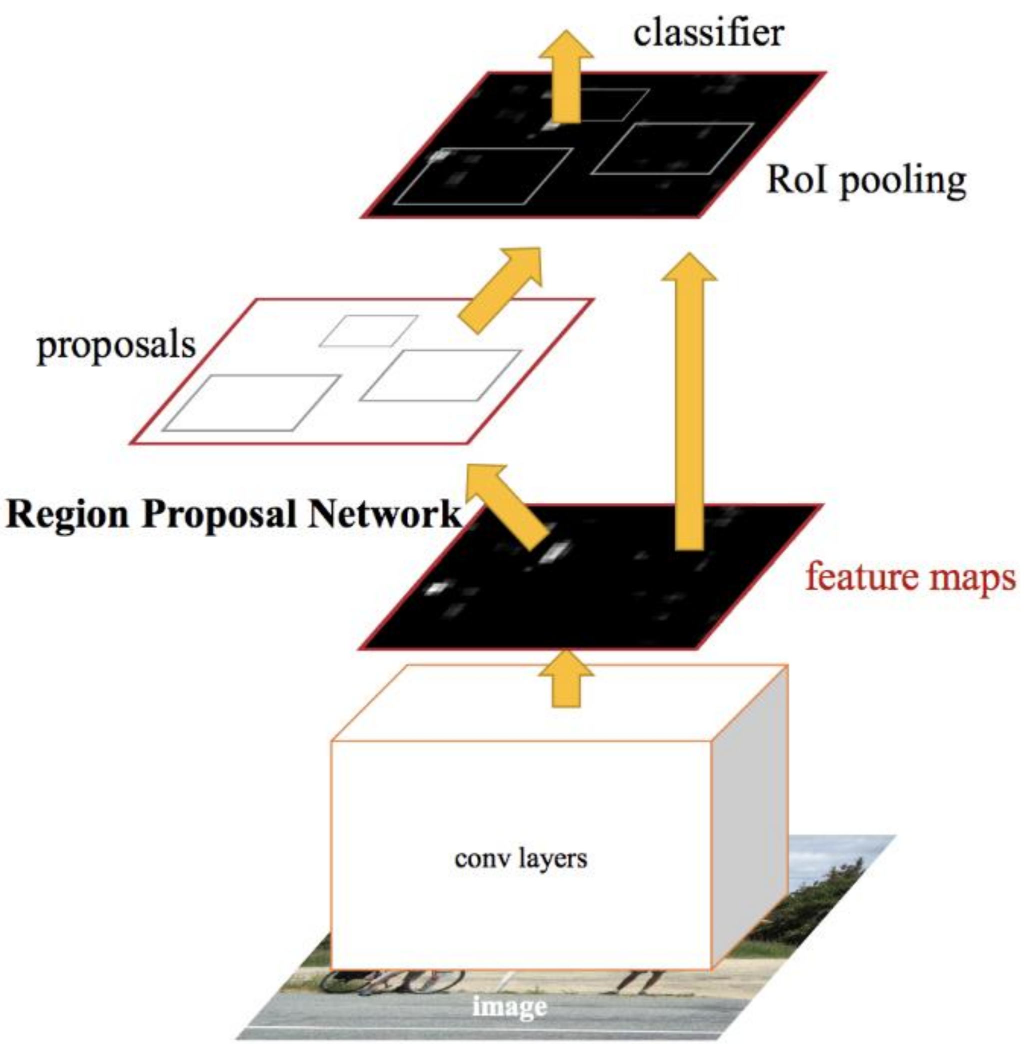
	Feature	Disadvantage	Note
CNN	Image -> multiple regions (tens of thousands) Region -> classes	High computational cost	
RCNN	Selective Search -> regions Classify 2k region proposals / image	High computational cost	
Fast RCNN	Still use Selective Search Image is fed to CNN only once.	High computational cost	
Faster RCNN	RPN (region proposal network) instead of Selective Search	RPN is much faster than Selective Search, but it still takes time.	
Yolo v2	Image -> x * x (grids) * m (bb), m = 5 Single network	Much faster, but struggles with small objects within the image	
Yolo v3	m = 3 x 3 scales, aspect ratios		
SSD	Good balance between speed and accuracy		*
RetinaNet	*Exploring…*		
CapsNet	*Dynamic routing of object-oriented neurons*	Research stage	*
M2Det	SSD based on multi-level FPN		*



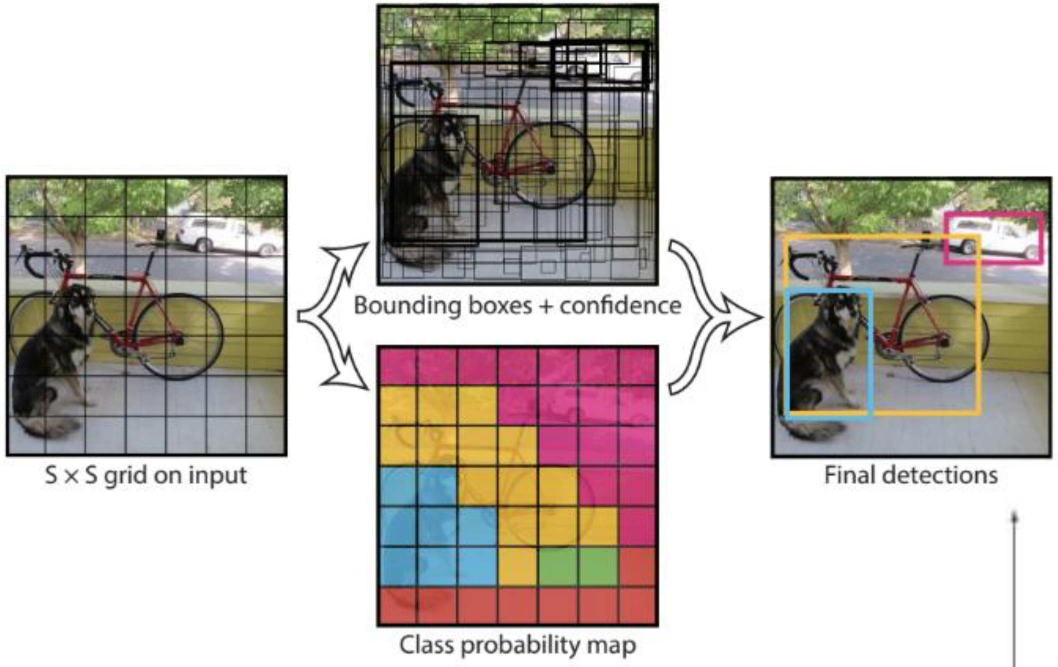
Fast R-CNN



Faster R-CNN



(From https://towardsdatascience.com/r-cnn-fast-r-cnn-faster-r-cnn-yolo-object-detection-algorithms-36d53571365e)



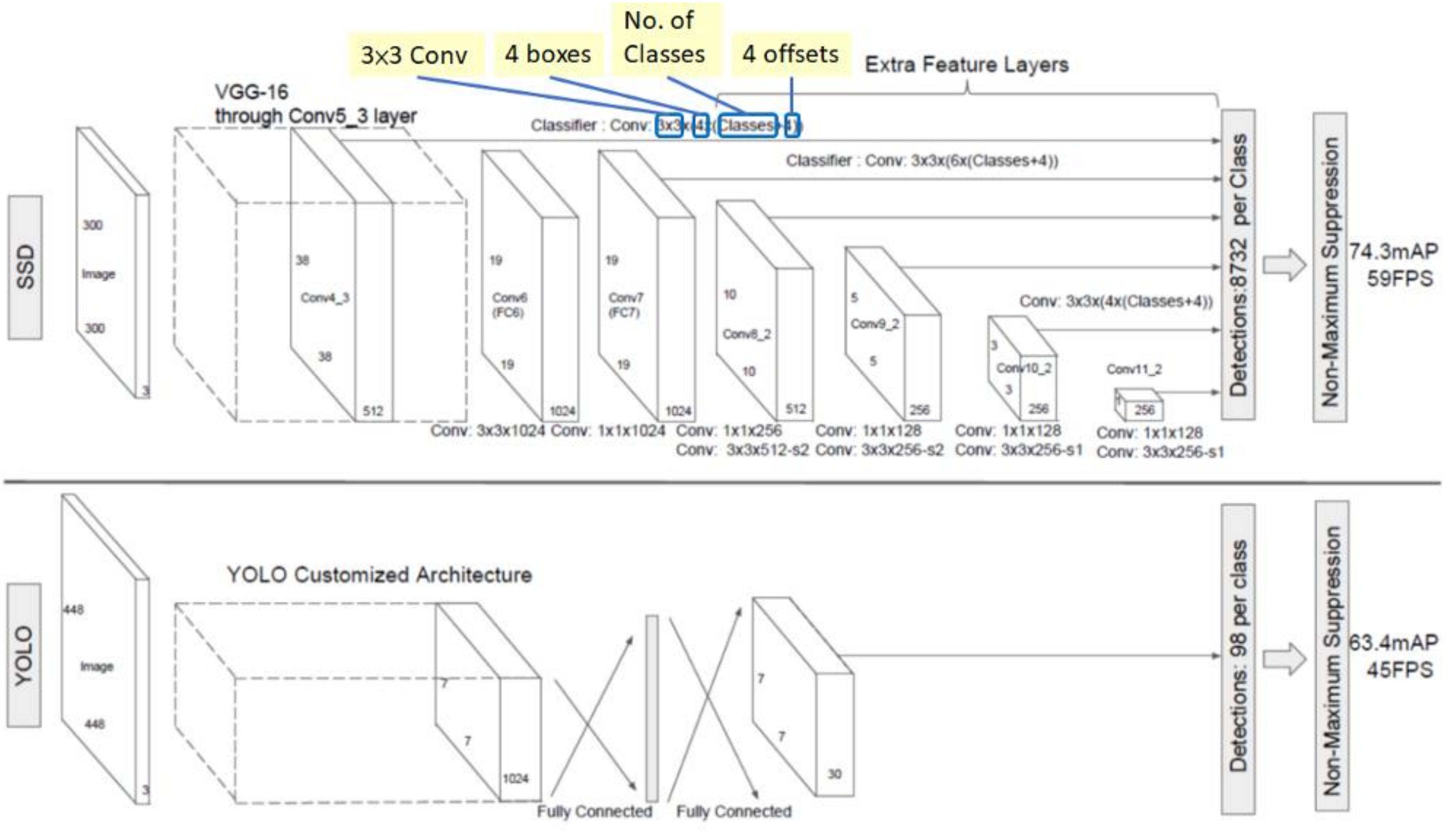
YOLO

(From https://towardsdatascience.com/r-cnn-fast-r-cnn-faster-r-cnn-yolo-object-detection-algorithms-36d53571365e, https://cv-tricks.com/object-detection/faster-r-cnn-yolo-ssd/)

SSD – Single-Short MultiBox Detector

- Multi-scale sliding window detector
- Feature sharing between classification and localization
- Priorbox – decides how local the detector is
  - Different types of priorbox with different scale or aspect ratio
- Data augmentation strategies
  - “zoom in” and “zoom out”
- Post-processing
  - E.g. filter out from 24,564 predictions on SSD512; filter out from 8,732 predictions on SSD300

SSD Network Architecture vs. Yolo's

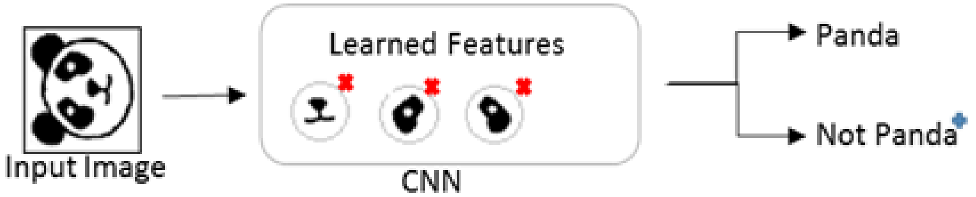


SSD (Top) vs YOLO (Bottom)

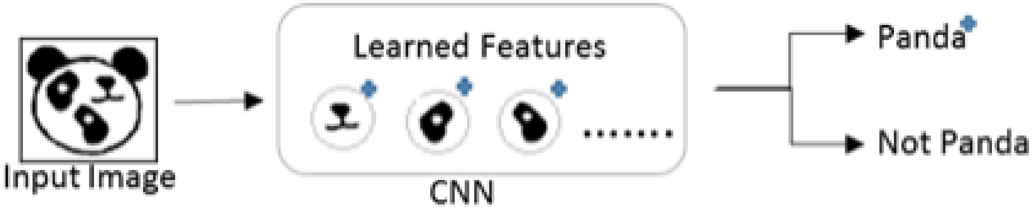
From https://towardsdatascience.com/review-ssd-single-shot-detector-object-detection-851a94607d11

Limitations in CNN

- Neurons don't consider the properties of a feature, like orientation, size, velocity, color, and etc.



- Cannot correctly clarify deformed image



- Translation Invariance

- Max pooling loses location info.

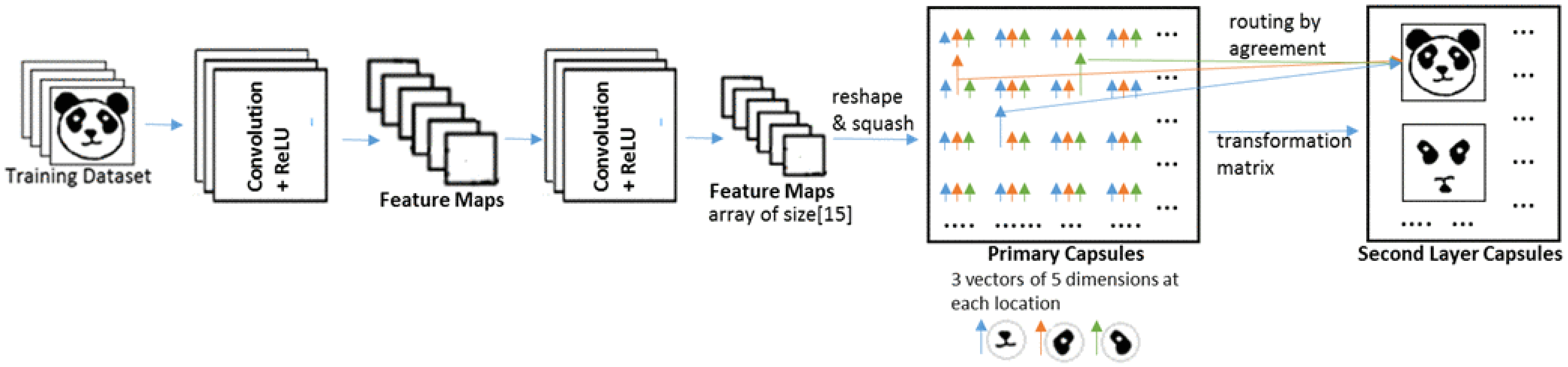
(From https://becominghuman.ai/understanding-capsnet-part-1-e274943a018d)

CapsNet – Capsule Network

- "Capsule", invented by Geoffrey Hinton
  - Group of neurons
  - Object-oriented neuron

- Being equivariant to the spatial setup of each entity inside an image

- Deliver rotational and other invariances

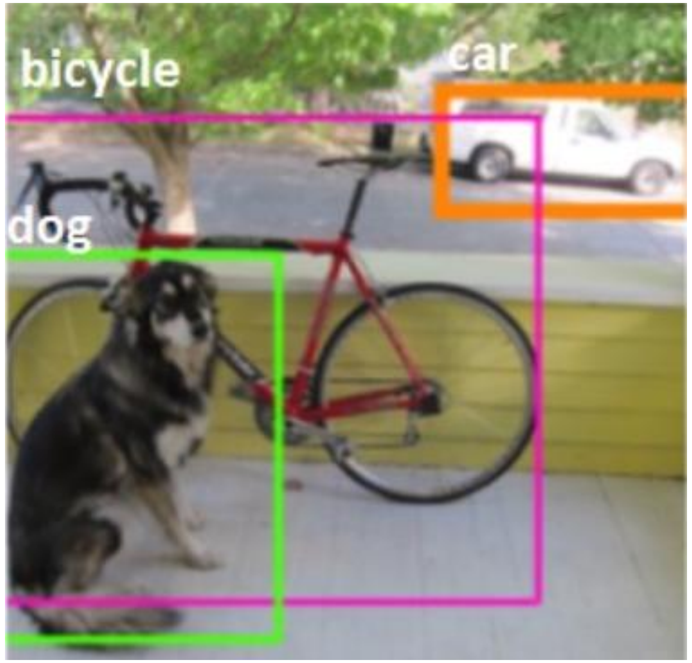


(From https://becominghuman.ai/understanding-capsnet-part-1-e274943a018d)

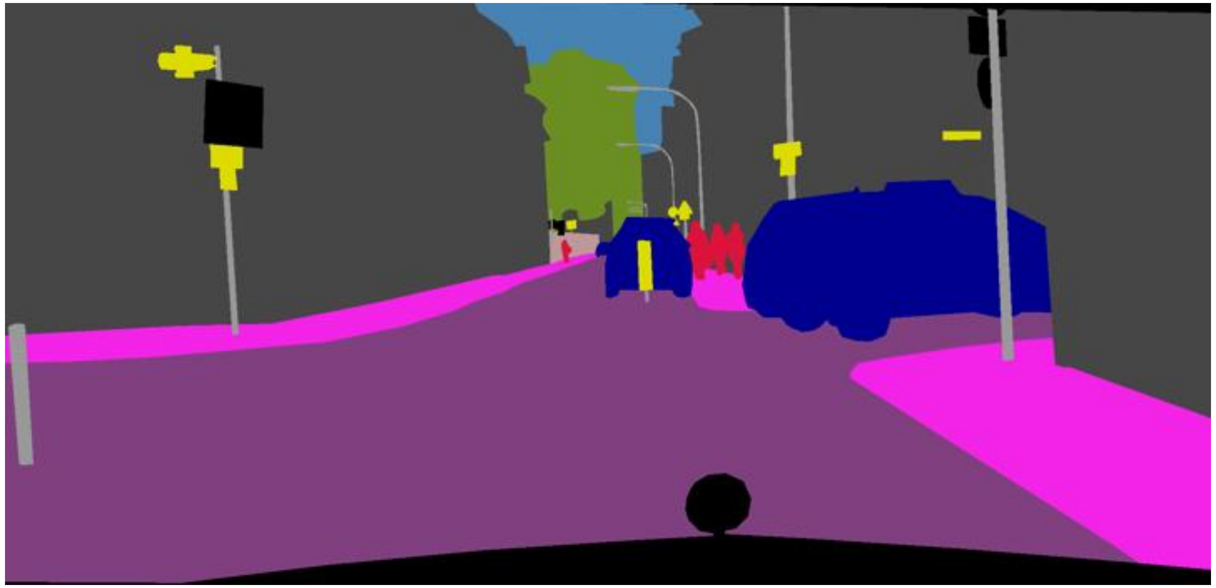
Revisit Design Goal

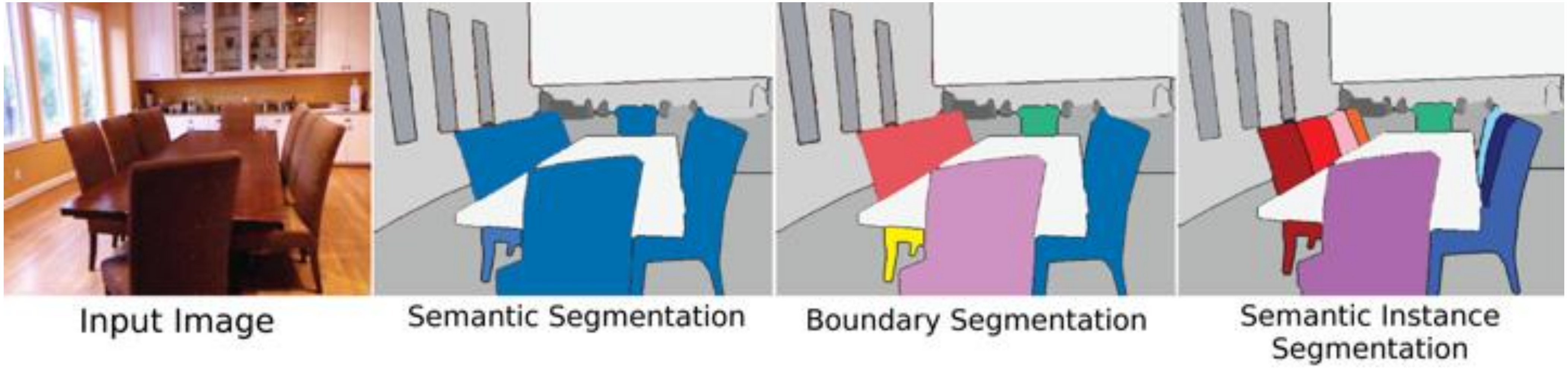
- Object localization is much more important.
- Customer will fine-tune model
  - Label detected objects
  - Adjust positions
- May focus on object localization without classification, if it improves performance signiticantly

Object Detection



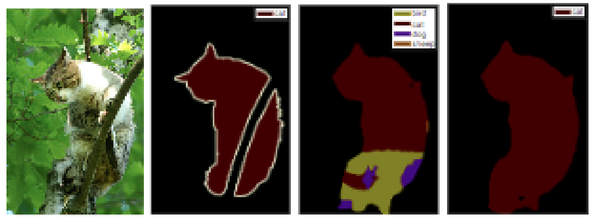
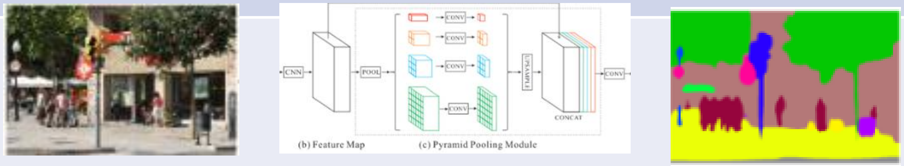
Semantic Segmentation

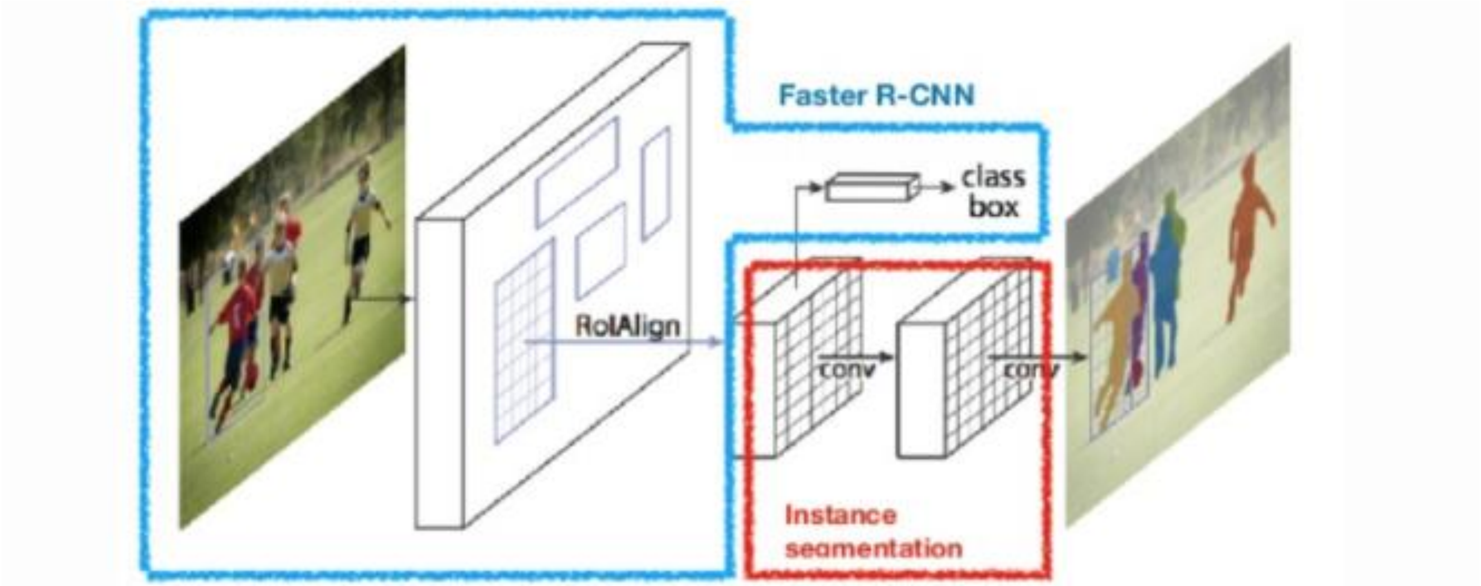




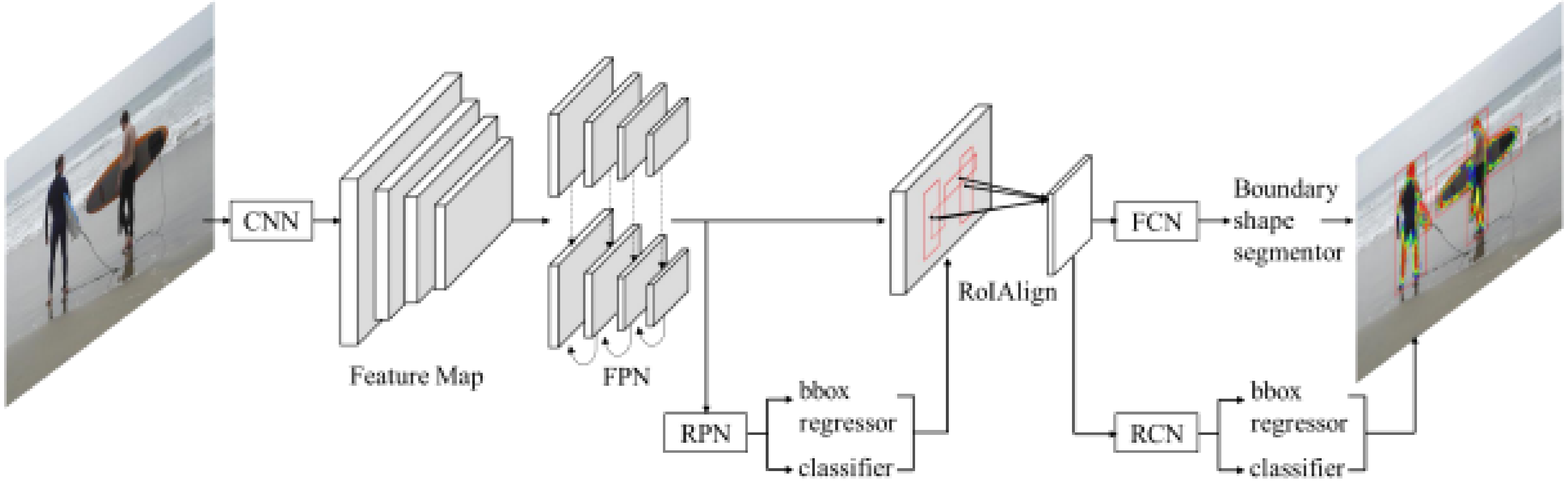
(From https://stackoverflow.com/questions/33947823/what-is-semantic-segmentation-compared-to-segmentation-and-scene-labeling)

Algorithms for Image Segmentation

	Feature	Disadvantage	
FCN (Fully Convolutional Network)	 (a) Image (b) Truth (c) FCN (d) ParseNet		
ParseNet			
FPN (Feature Pyramid Network)	*exploring…*		
PSPNet (Pyramid Scene Parsing Network)			
Mask RCNN	Instance segmentation (at pixel level) Derived from Faster RCNN Decouple classification and pixel-level mask prediction		
DPM (Deformable Parts Model)	A root filter, multiple part filters, and a spatial model		
DeepLab, DeepLabv3, DeepLabv3+	*exploring…*		
PANet (Path Aggregation Network)	Based on Mask RCNN and FPN		



Mask R-CNN is Faster R-CNN model with image segmentation. (Image source: He et al., 2017)



(From https://arxiv.org/pdf/1810.10327.pdf)

Actions

Approach 1

- Collect data, create labels, and build a pretrained model using new algorithm (e.g. SSD, or CapsNet, or M2Det)
- Assist with an unsupervised learning method

Approach 2

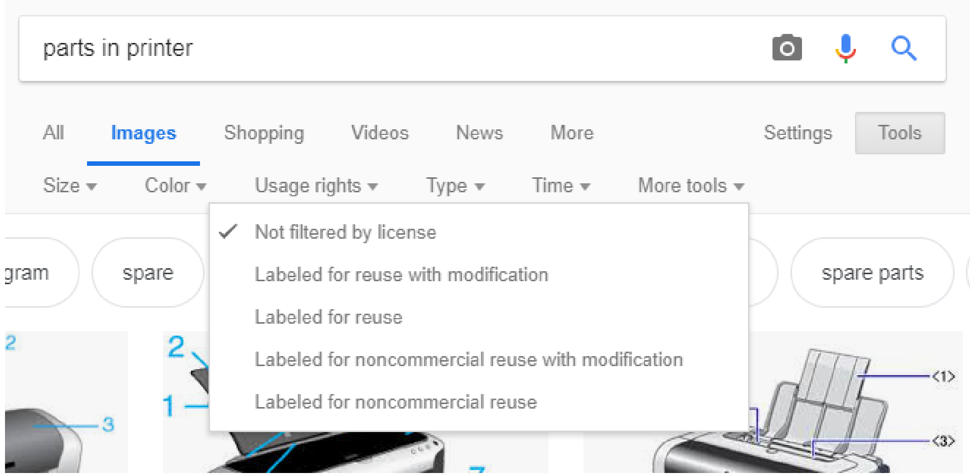
- Explore semantic segmentation modeling method

Image Data (1)

- From open domain images
  - Open Image Dataset – Google from Flickr
    https://storage.googleapis.com/openimages/web/download.html
  - COCO(Common Objects in Context) – CVDF(Common Visual Data Foundation), MS, Facebook, etc.
    http://cocodataset.org
  - ImageNet – Stanford, Princeton
    http://image-net.org/download
  - PASCAL VOC (Pattern Analysis, Statistical Modeling and Computational Learning, Visual Object Classes)
    http://host.robots.ox.ac.uk/pascal/VOC/
  - Tiny Images Dataset - NYU, MIT
    http://horatio.cs.nyu.edu/mit/tiny/data/index.html
  - The CIFAR-10 dataset - Professor Hinton, and co.
    https://www.cs.toronto.edu/~kriz/cifar.html

Image Data (2)

- From image search engines – need to pay attention to privacy
  - Bing
  - Google on Chrome
  - Yahoo, Baidu, ...



- From video data
  - A large-scale database of object videos from YouTube
    https://data.vision.ee.ethz.ch/cvl/youtube-objects/

Training Tools

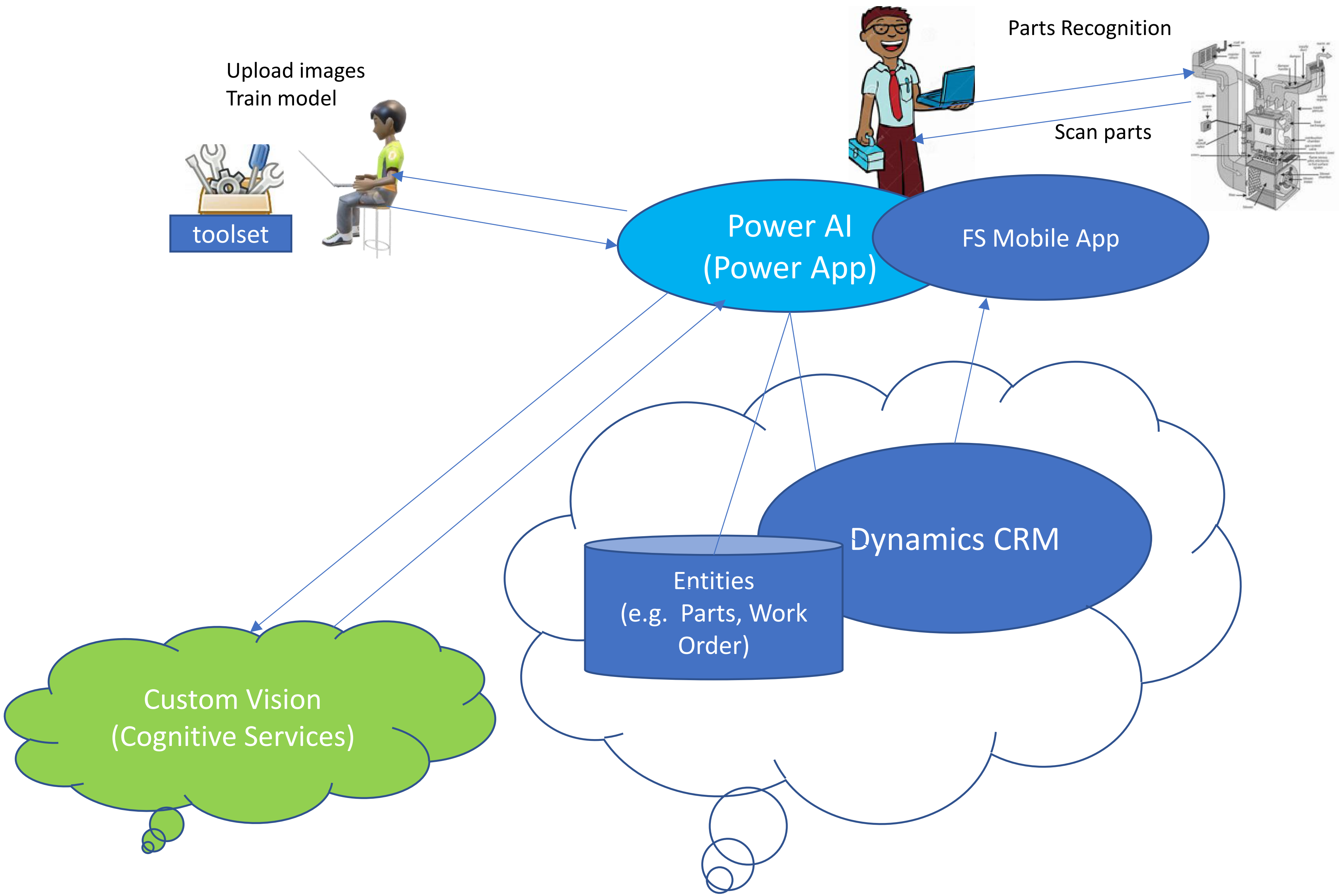
- Labeling
  - Lableimg
  - VoTT
  - Custom Vision UI

- Format
  - Utility to convert JSON to Xml (PASCAL VOC format)
  - Others

- Frameworks
  - Caffe
  - PyTorch
  - Tensorflow / Keras
  - Darknet / Darkflow

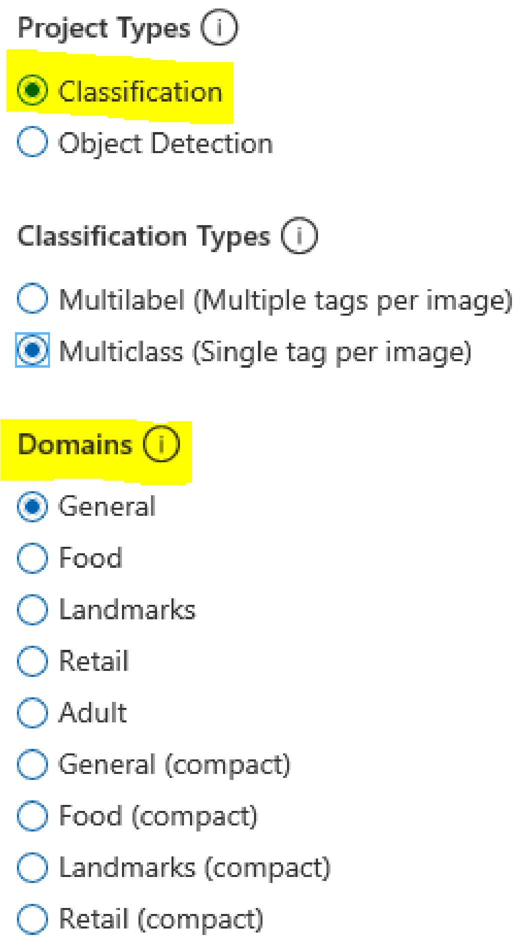
Performance Factors

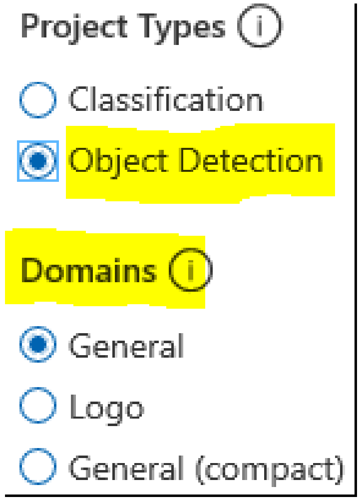
- Input image resolutions
- Image preprocessing
- Data Augmentation
- Feature extractors
- IOU threshold
- Localization loss function
- Deep learning platform to be used
- Training parameters
  - e.g. batch size, learning rate, image resize, etc.



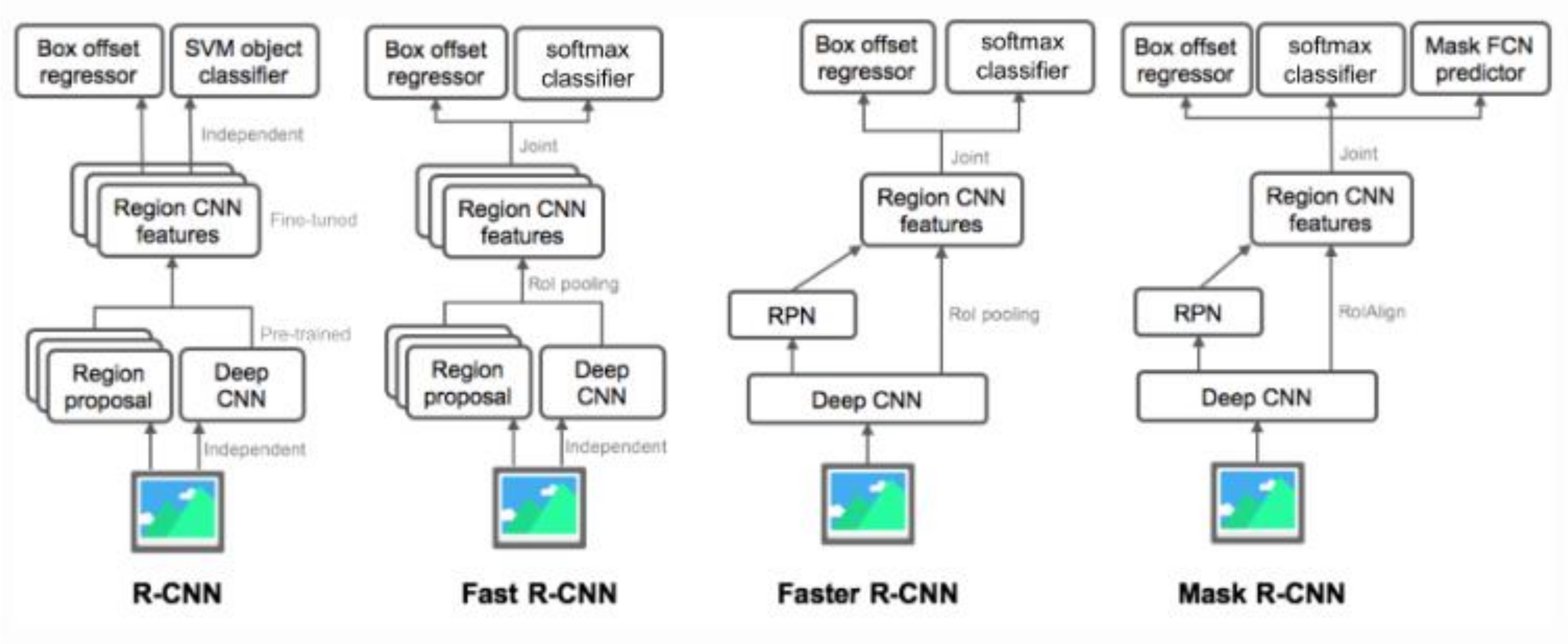
Upload images
Train model
toolset
Parts Recognition
Scan parts
Power AI
(Power App)
FS Mobile App
Dynamics CRM
Entities
(e.g. Parts, Work Order)
Custom Vision
(Cognitive Services)

Opportunities in Custom Vision





Appendix



(From https://lilianweng.github.io/lil-log/2017/12/31/object-recognition-for-dummies-part-3.html)

Transfer Learning

Unsupervised Learning