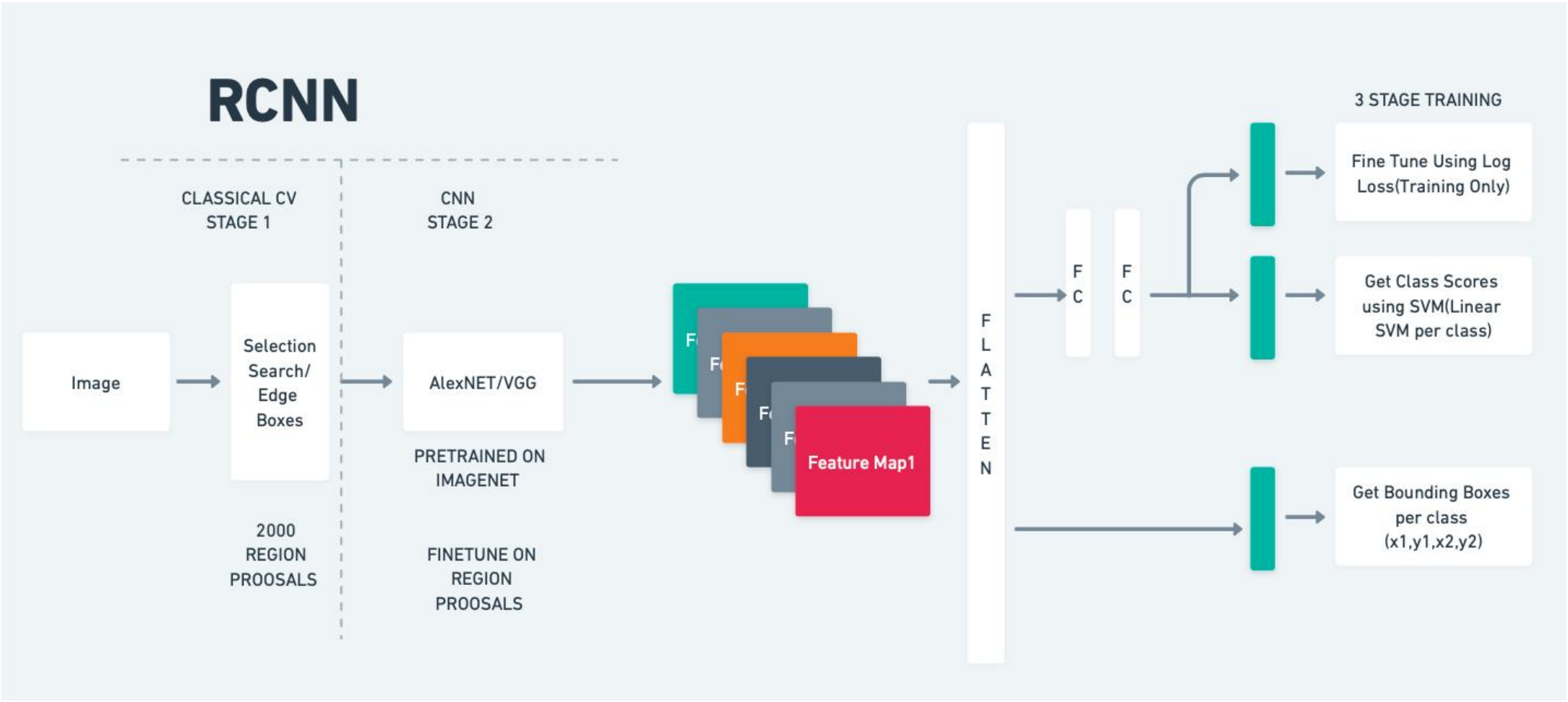
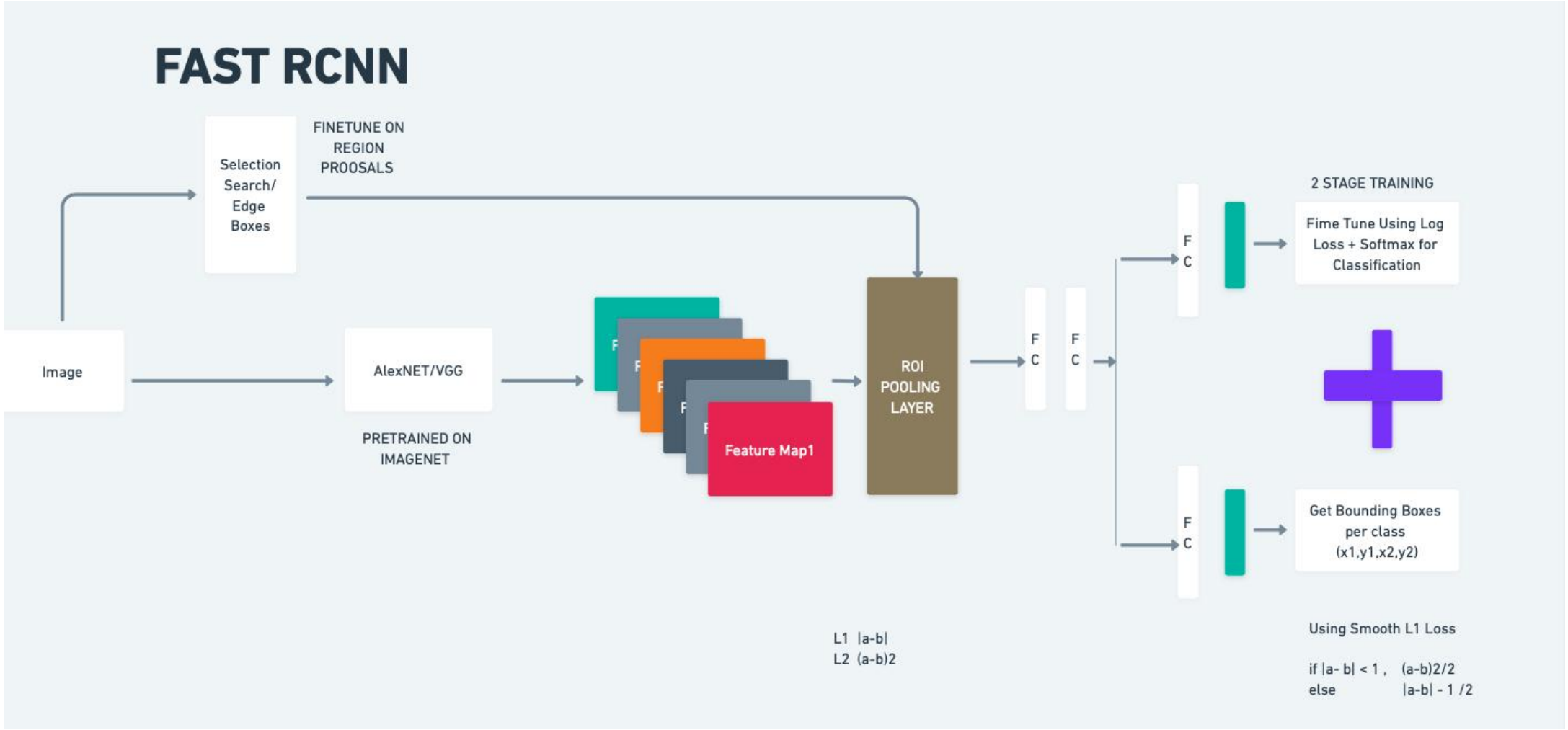
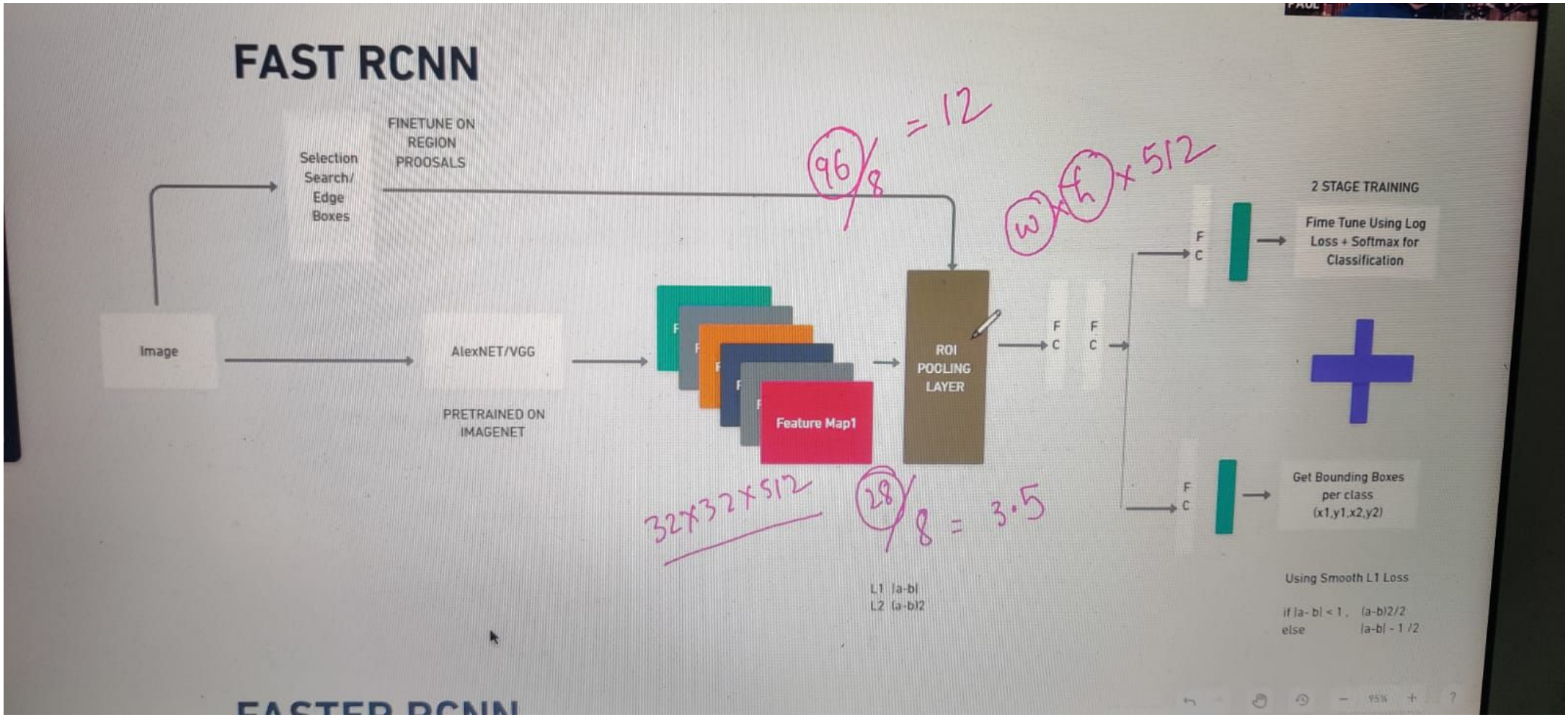
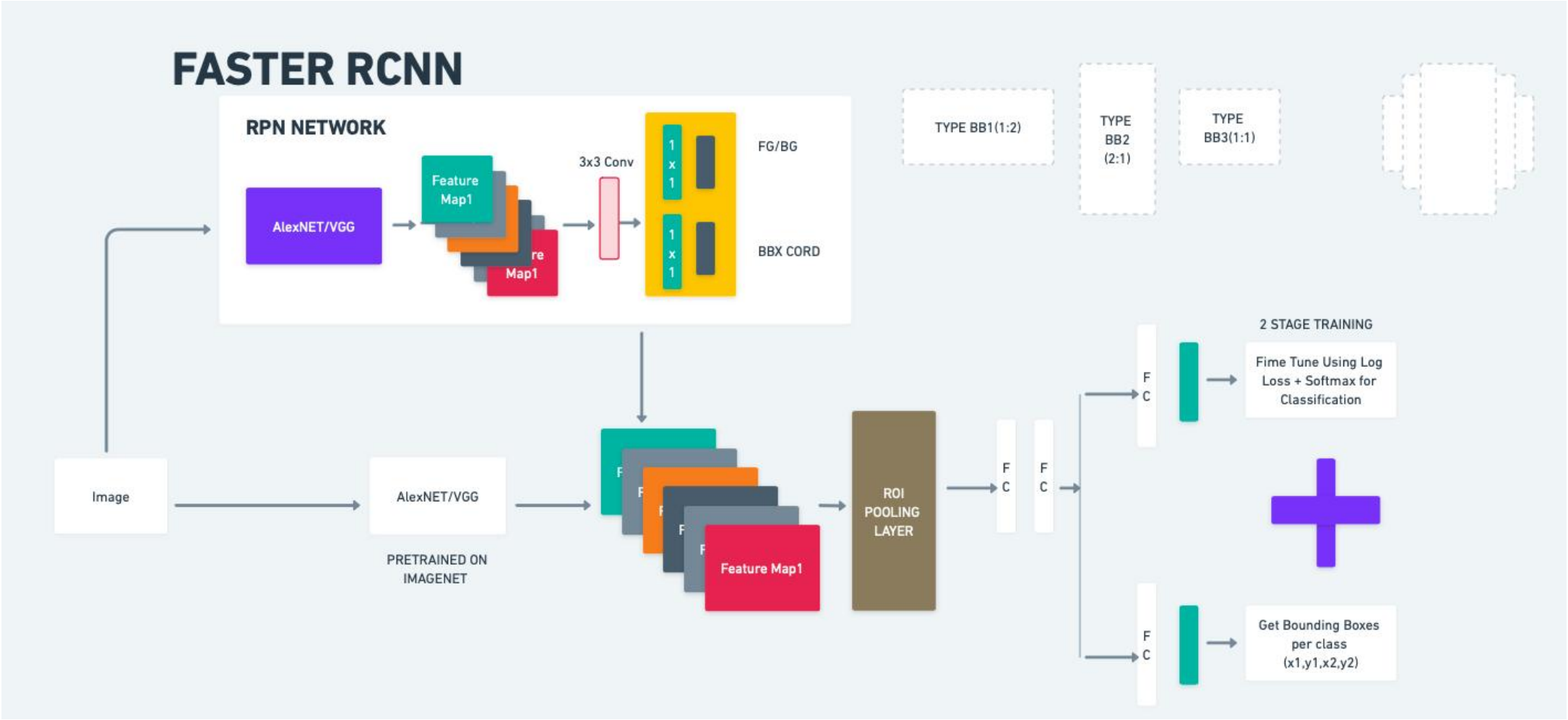
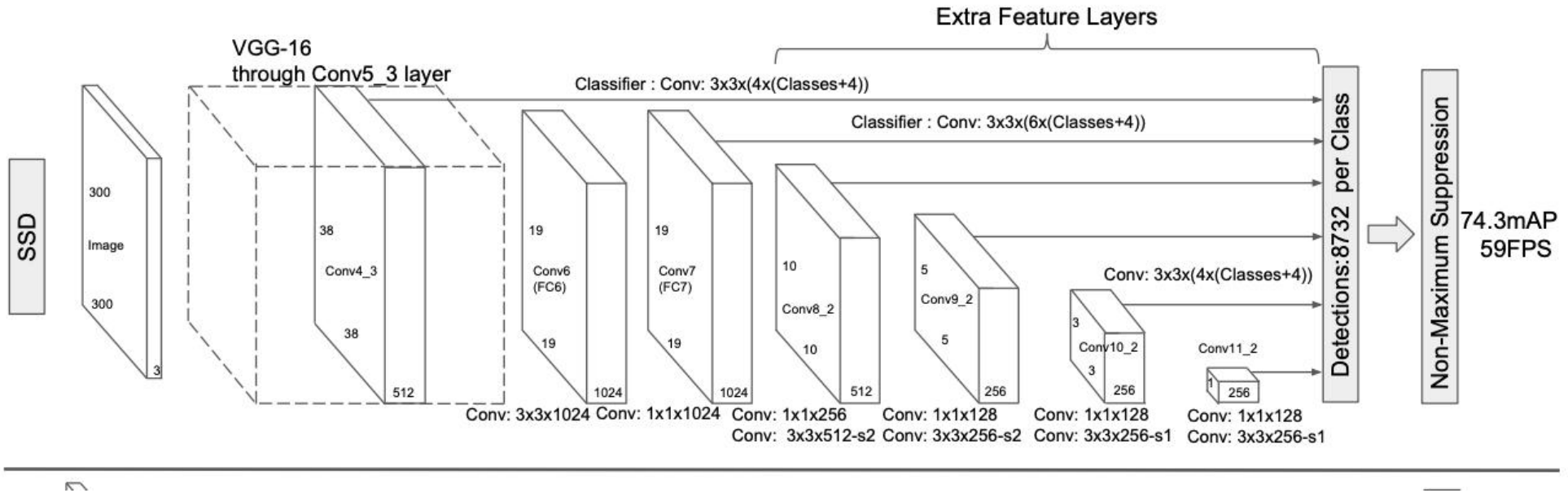

RCNN
CLASSICAL CV
STAGE 1
CNN
STAGE 2
Image
Selection
Search/
Edge
Boxes
2000
REGION
PROOSALS
AlexNET/VGG
PRETRAINED ON
IMAGENET
FINETUNE ON
REGION
PROOSALS
Feature Map1
FLATTEN
FC
FC
3 STAGE TRAINING
Fine Tune Using Log
Loss(Training Only)
Get Class Scores
using SVM(Linear
SVM per class)
Get Bounding Boxes
per class
(x1,y1,x2,y2)

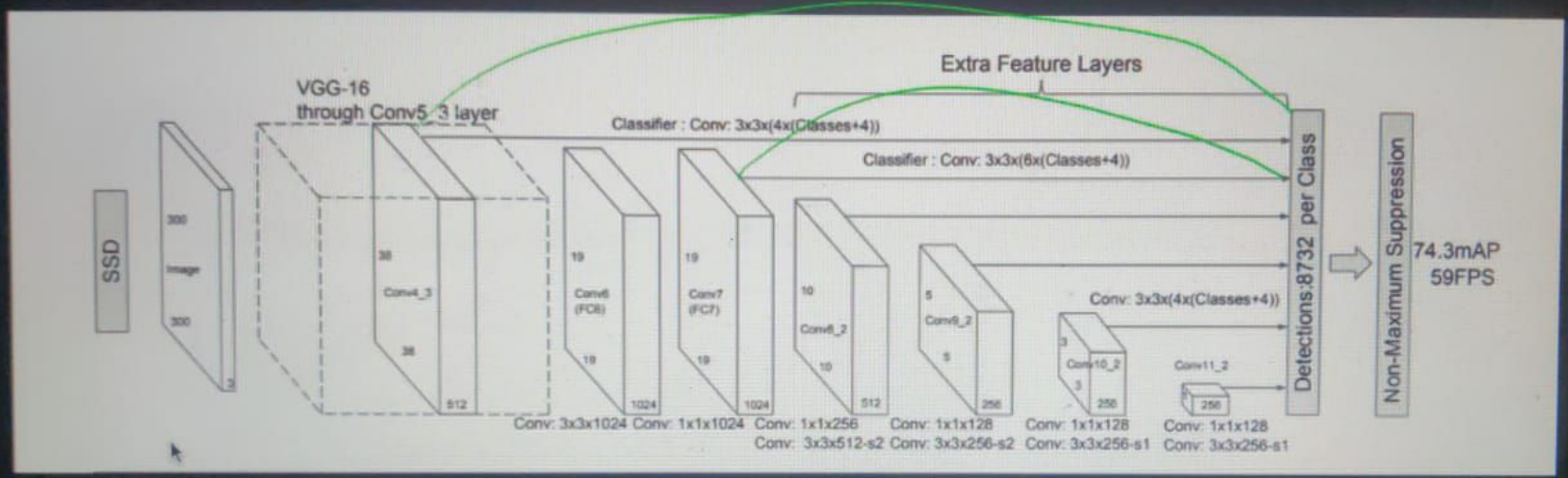

FAST RCNN
Selection Search/ Edge Boxes
FINETUNE ON REGION PROOSALS
Image
AlexNET/VGG
PRETRAINED ON IMAGENET
Feature Map1
ROI POOLING LAYER
F C
F C
F C
F C
2 STAGE TRAINING
Fime Tune Using Log Loss + Softmax for Classification
Get Bounding Boxes per class (x1,y1,x2,y2)
L1 |a-b|
L2 (a-b)2
Using Smooth L1 Loss
if |a- b| < 1 , (a-b)2/2
else |a-b| - 1 /2


FAST RCNN
Selection Search/ Edge Boxes
FINETUNE ON REGION PROOSALS
Image
AlexNET/VGG
PRETRAINED ON IMAGENET
Feature Map1
ROI POOLING LAYER
F C
F C
F C
F C
2 STAGE TRAINING
Fime Tune Using Log Loss + Softmax for Classification
Get Bounding Boxes per class (x1,y1,x2,y2)
Using Smooth L1 Loss
if |a- b| < 1 , (a-b)2/2
else |a-b| - 1 /2
L1 |a-b|
L2 (a-b)2
96/8 = 12
w x h x 512
32x32x512
28/8 = 3.5


FASTER RCNN
RPN NETWORK
AlexNET/VGG
Feature Map1
Map1
3x3 Conv
1 x 1
1 x 1
FG/BG
BBX CORD
TYPE BB1(1:2)
TYPE BB2 (2:1)
TYPE BB3(1:1)
Image
AlexNET/VGG
PRETRAINED ON IMAGENET
Feature Map1
ROI POOLING LAYER
F C
F C
F C
F C
2 STAGE TRAINING
Fime Tune Using Log Loss + Softmax for Classification
Get Bounding Boxes per class (x1,y1,x2,y2)

Liu *et al.*


Extra Feature Layers
VGG-16
through Conv5_3 layer
Classifier : Conv: 3x3x(4x(Classes+4))
Classifier : Conv: 3x3x(6x(Classes+4))
Conv: 3x3x(4x(Classes+4))
SSD
300
Image
300
3
38
Conv4_3
38
512
19
Conv6
(FC6)
19
1024
19
Conv7
(FC7)
19
1024
10
Conv8_2
10
512
5
Conv9_2
5
256
3
Conv10_2
3
256
Conv11_2
1
256
Detections:8732 per Class
Non-Maximum Suppression
74.3mAP
59FPS
Conv: 3x3x1024
Conv: 1x1x1024
Conv: 1x1x256
Conv: 3x3x512-s2
Conv: 1x1x128
Conv: 3x3x256-s2
Conv: 1x1x128
Conv: 3x3x256-s1
Conv: 1x1x128
Conv: 3x3x256-s1

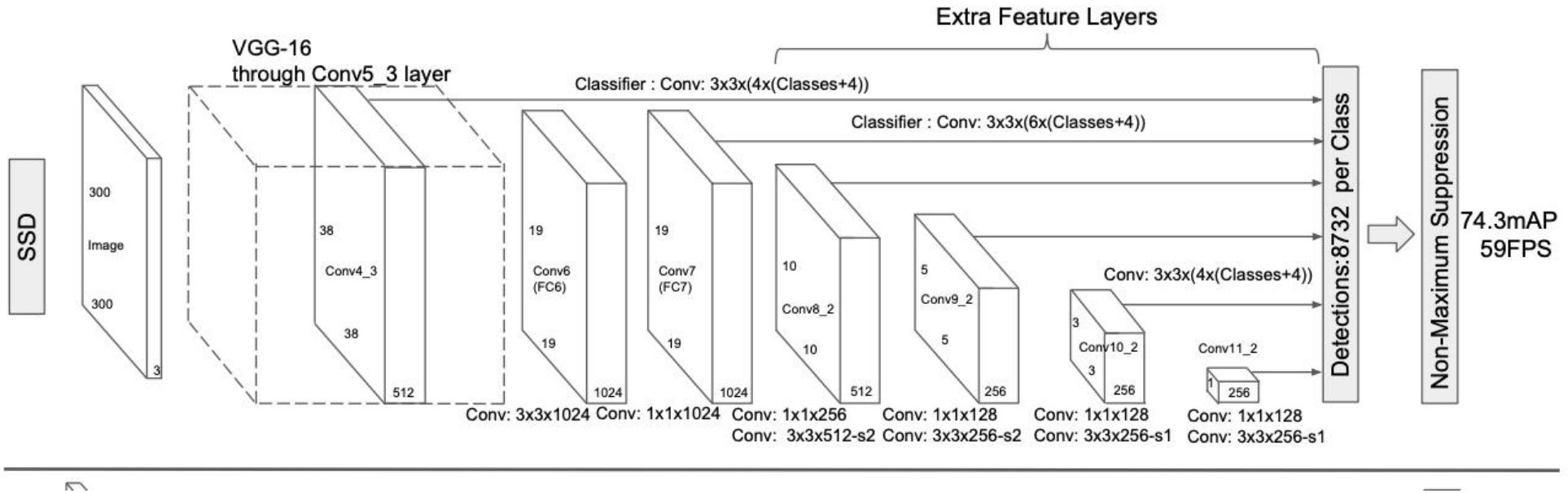


Edges/Gradients → Patterns → Textures → Parts of objects → Object

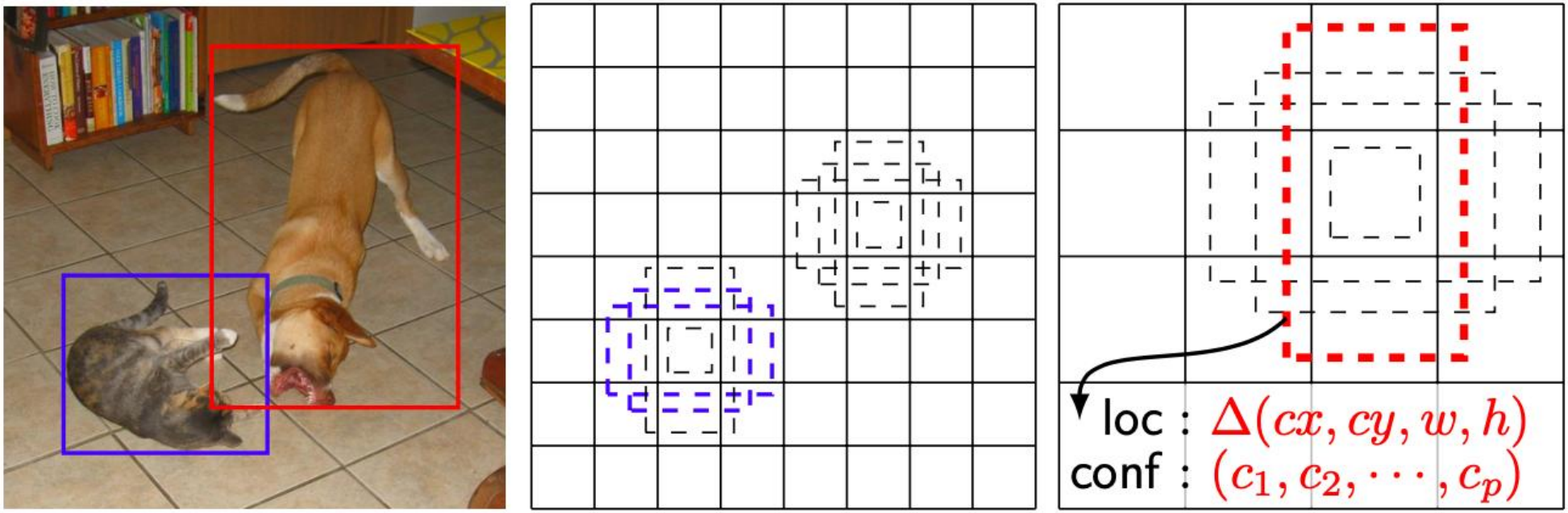
C_1 C_2 C_3 C_4 C_5

$\{38 \times 38 \times 512,\ 19 \times 19 \times 1024\}$ → Classification (Conf Score)

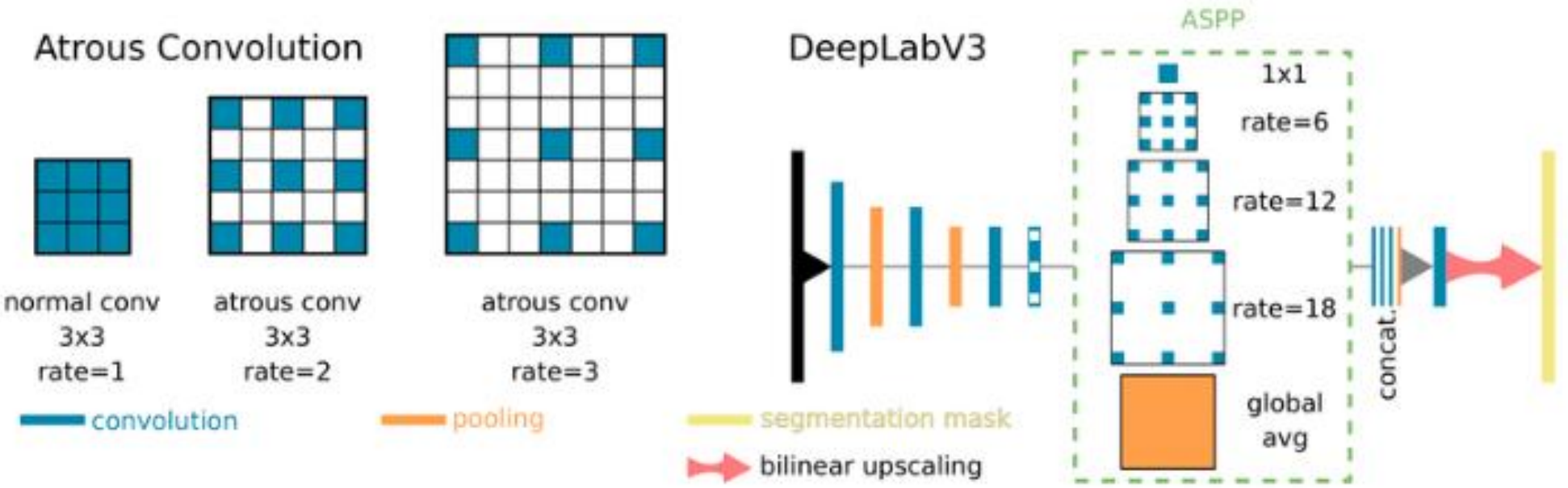
Liu *et al.*


Extra Feature Layers
VGG-16
through Conv5_3 layer
Classifier : Conv: 3x3x(4x(Classes+4))
Classifier : Conv: 3x3x(6x(Classes+4))
Conv: 3x3x(4x(Classes+4))
SSD
300
Image
300
3
38
Conv4_3
38
512
19
Conv6
(FC6)
19
1024
19
Conv7
(FC7)
19
1024
10
Conv8_2
10
512
5
Conv9_2
5
256
3
Conv10_2
3
256
Conv11_2
1
256
Detections:8732 per Class
Non-Maximum Suppression
74.3mAP
59FPS
Conv: 3x3x1024
Conv: 1x1x1024
Conv: 1x1x256
Conv: 3x3x512-s2
Conv: 1x1x128
Conv: 3x3x256-s2
Conv: 1x1x128
Conv: 3x3x256-s1
Conv: 1x1x128
Conv: 3x3x256-s1

SSD: Single Shot MultiBox Detector



(a) Image with GT boxes (b) 8×8 feature map (c) 4×4 feature map



Conceptual overview of atrous convolution (left) and the Atrous Spatial Pyramid Pooling module (ASPP) within the DeepLabV3 [59] architecture (right). In atrous convolution, "holes" are inserted into the kernel. They provide a larger receptive field and maintain resolution at the same time. The ASPP module combines atrous convolutions of different rates and global image context via global average pooling in order to exploit feature maps on different scales efficiently [56,58,59].

Semantic Segmentation vs. Instance Segmentation vs. Panoptic Segmentation

(a) Image

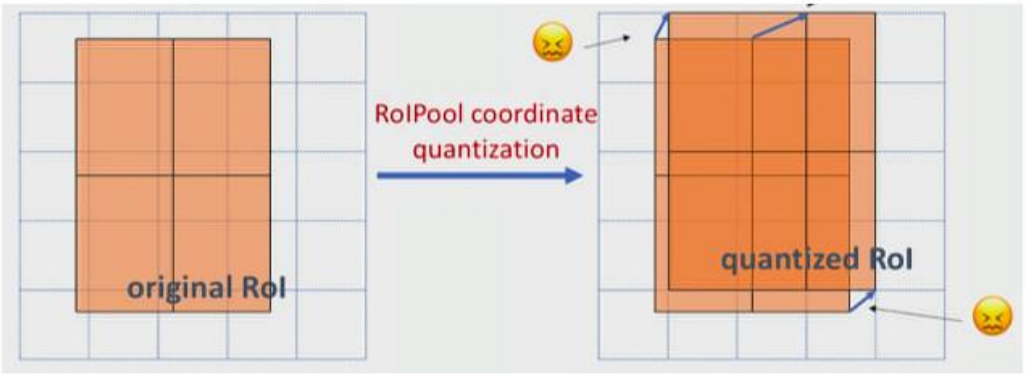
(b) Semantic Segmentation

(c) Instance Segmentation

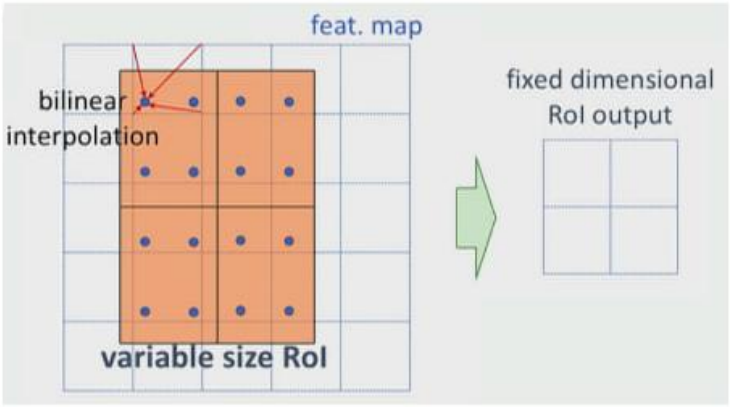
(d) Panoptic Segmentation

RoiAlign

As discussed above, RoIPool layer extracts small feature maps from each RoI. The problem with RoIPool is quantization. If the RoI doesn't perfectly align with the grid in feature map as shown, the quantization breaks pixel-to-pixel alignment. It isn't much of a problem in object detection, but in case of predicting masks, which require finer spatial localization, it matters.

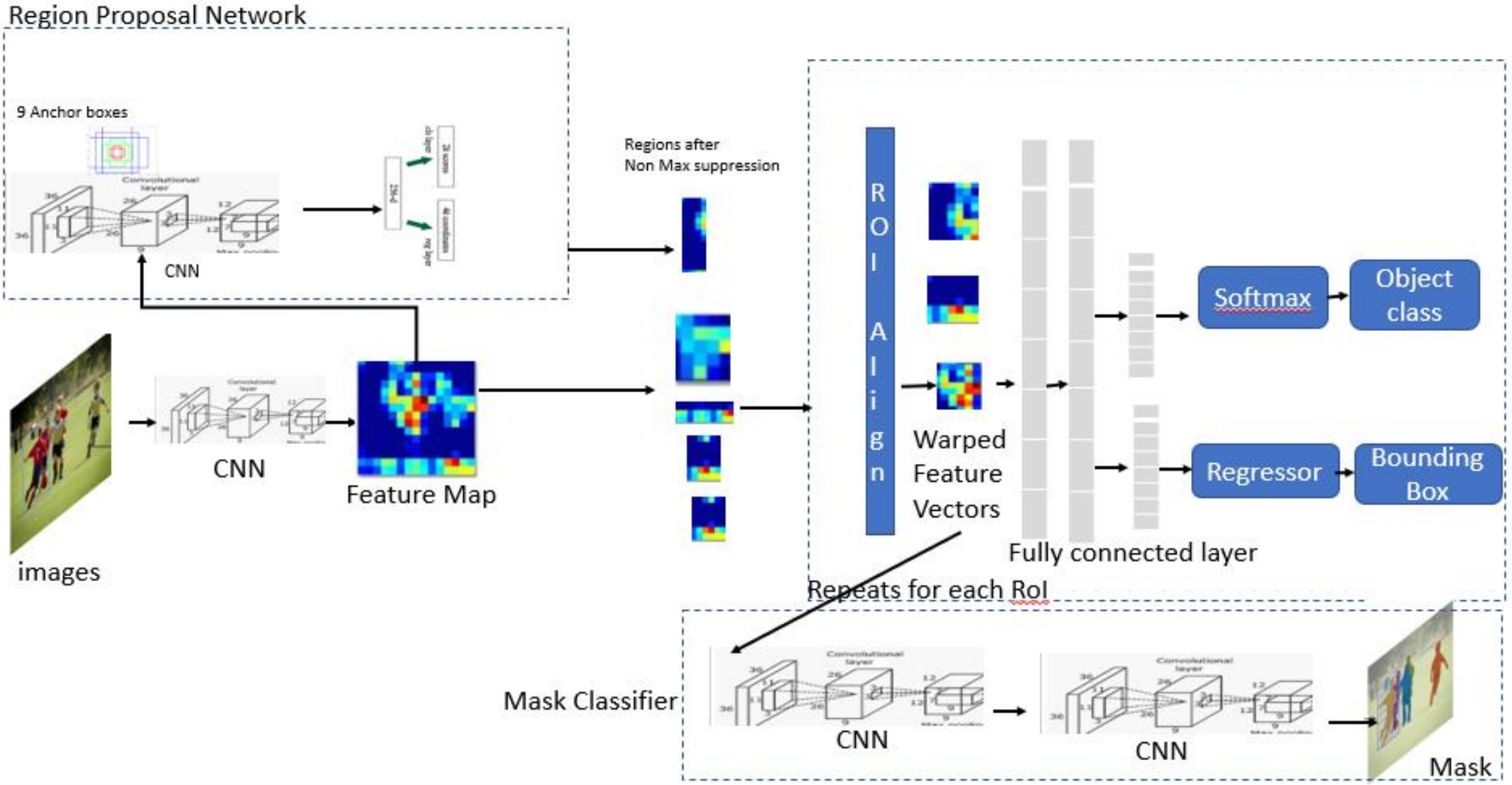


RoIAlign is an improvement over the RoIPool operation. What RoIAlign does is to smoothly transform features from the RoIs (which has different aspect sizes) into fixed size feature vectors without using *quantization*. It uses bilinear interpolation to do. A grid of sampling points are used within each bin of RoI, which are used to interpolate the features at its nearest neighbors as shown.



For example, in the above figure, you can't apply the max-pooling directly due to the misalignment of RoI with the feature map grids, thus in case of RoIAlign, four points are sampled in each bin using bilinear interpolation from its nearest neighbors. Finally, the max value from these points is chosen to get the required 2x2 feature map.

Mask RCNN


Region Proposal Network
9 Anchor boxes
CNN
Regions after
Non Max suppression
images
CNN
Feature Map
ROI Align
Warped
Feature
Vectors
Softmax
Object
class
Regressor
Bounding
Box
Fully connected layer
Repeats for each RoI
Mask Classifier
CNN
CNN
Mask

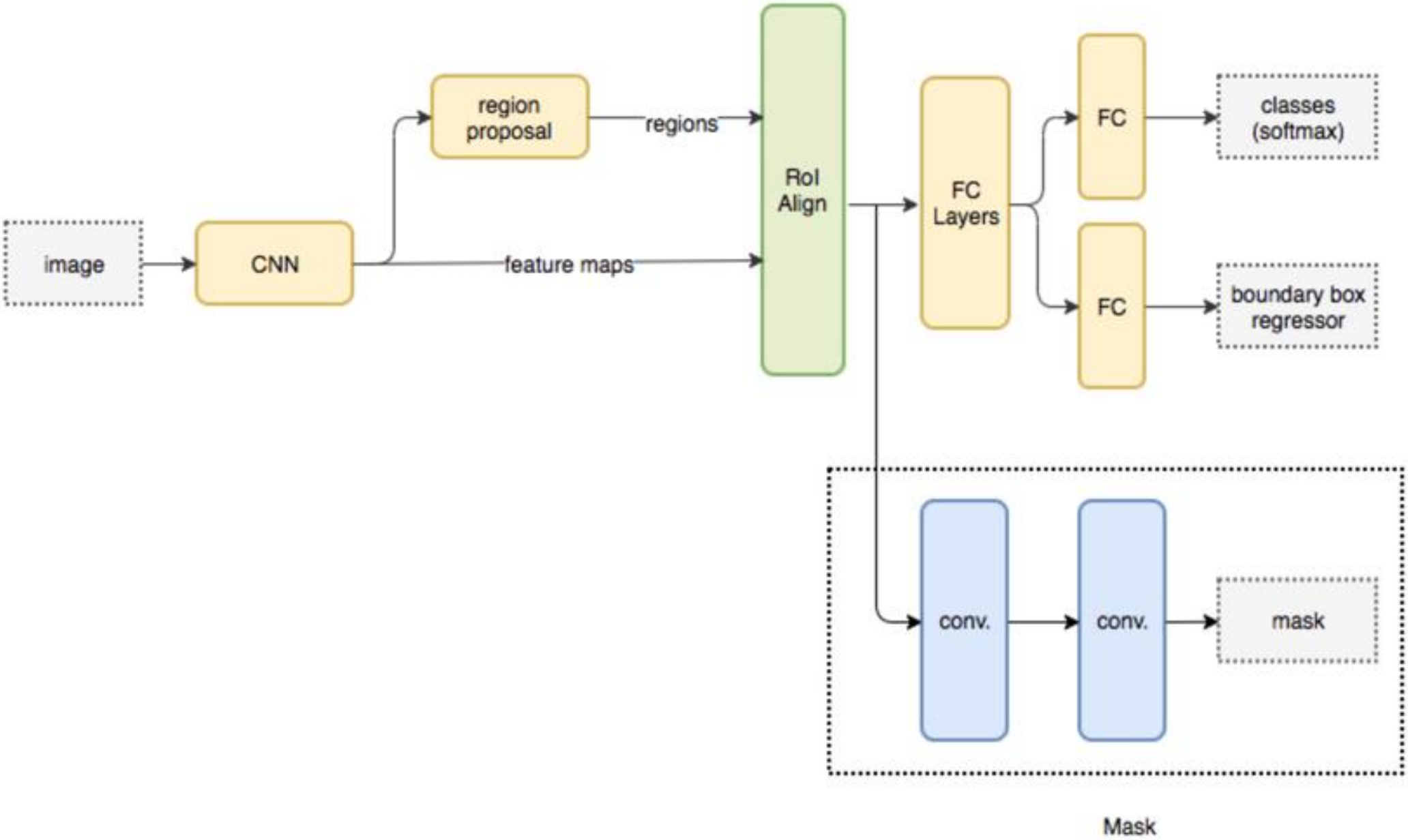
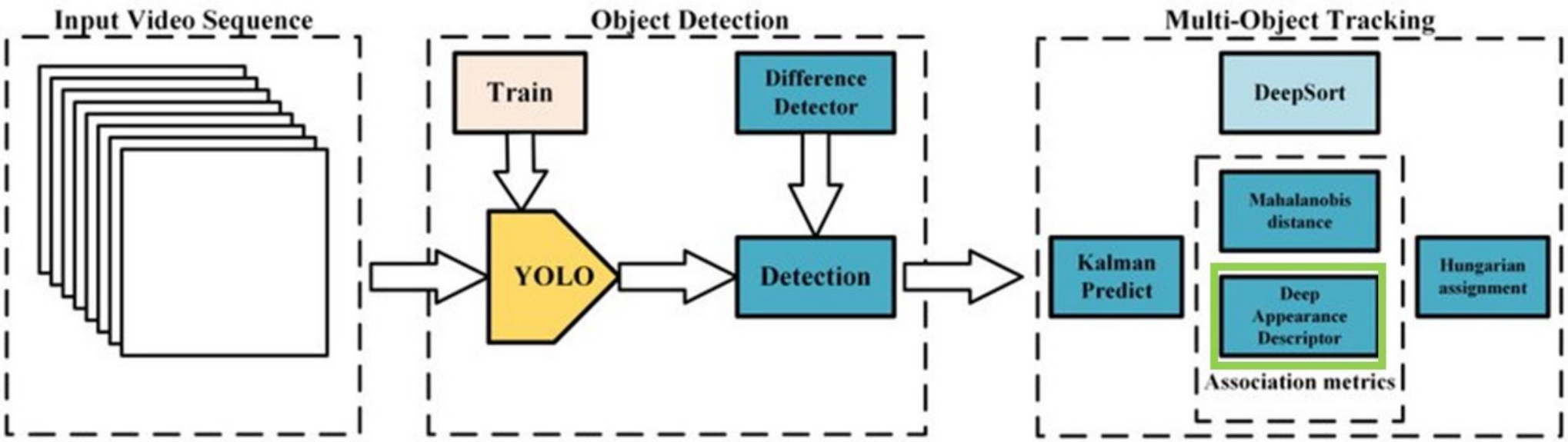


image
CNN
region proposal
regions
feature maps
RoI Align
FC Layers
FC
classes (softmax)
FC
boundary box regressor
conv.
conv.
mask
Mask

DeepSORT

Where is the Deep Learning in all of this?



Computer Vision

Face Recognition

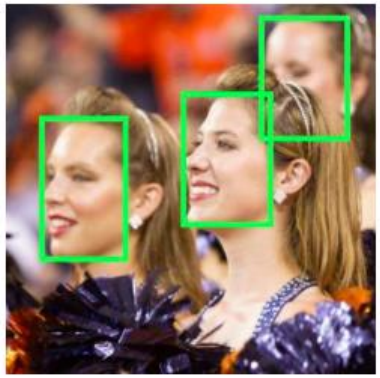
Edit

440 papers with code • 23 benchmarks • 79 datasets

Facial recognition is the task of making a positive identification of a face in a photo or video image against a pre-existing database of faces. It begins with detection - distinguishing human faces from other objects in the image - and then works on identification of those detected faces.

The state of the art tables for this task are contained mainly in the consistent parts of the task : the face verification and face identification tasks.

(Image credit: Face Verification)



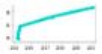
Benchmarks

Add a Result

These leaderboards are used to track progress in Face Recognition

Trend	Dataset	Best Model	Paper	Code	Compare
	LFW	Prodpoly			See all
	CFP-FP	Prodpoly			See all
	MLFW	MS1MV2, R100, SFace			See all
	CelebA+masks	Fine-tuned ArcFace			See all

Content

	CASIA-WebFace+masks	🏆 Fine-tuned ArcFace			See all
	AgeDB-30	🏆 QMagFace			See all
	Adience	🏆 PIC - MagFace			See all
	Color FERET	🏆 PIC - QMagFace			See all
	MORPH	🏆 PIC - ArcFace			See all
	IJB-B	🏆 ArcFace+Ours cleaned MS1MV2, 3.9M(+70K)			See all

Show all 23 benchmarks

Libraries

Use these libraries to find Face Recognition models and implementations

serengil/deepface	10 papers	5,233 ★
deepinsight/insightface	9 papers	13,210 ★
Tencent/TFace	5 papers	958 ★
tomas-gajarsky/facetorch	3 papers	88 ★

See all 5 libraries.

Libraries

Use these libraries to find Face Recognition models and implementations

serengil/deepface	10 papers	5,233 ★
deepinsight/insightface	9 papers	13,210 ★
Tencent/TFace	5 papers	958 ★
tomas-gajarsky/facetorch	3 papers	88 ★

See all 5 libraries.

Datasets

LFW | CASIA-WebFace | MS-Celeb-1M | Extended Yale B | IJB-B | Adience

VGG Face | RFW | Partial-REID | Color FERET

See all 79 face recognition datasets

Subtasks

Age-Invariant Face Recognition | Face Quality Assessement | Unsupervised face recognition