Pedestrians Detection

~ MIRPR report ~

by Jurj Flaviu-Andrei
Maxim Tudor
Moisi Teofana-Ionela

Teacher:
Prof. Diosan Laura

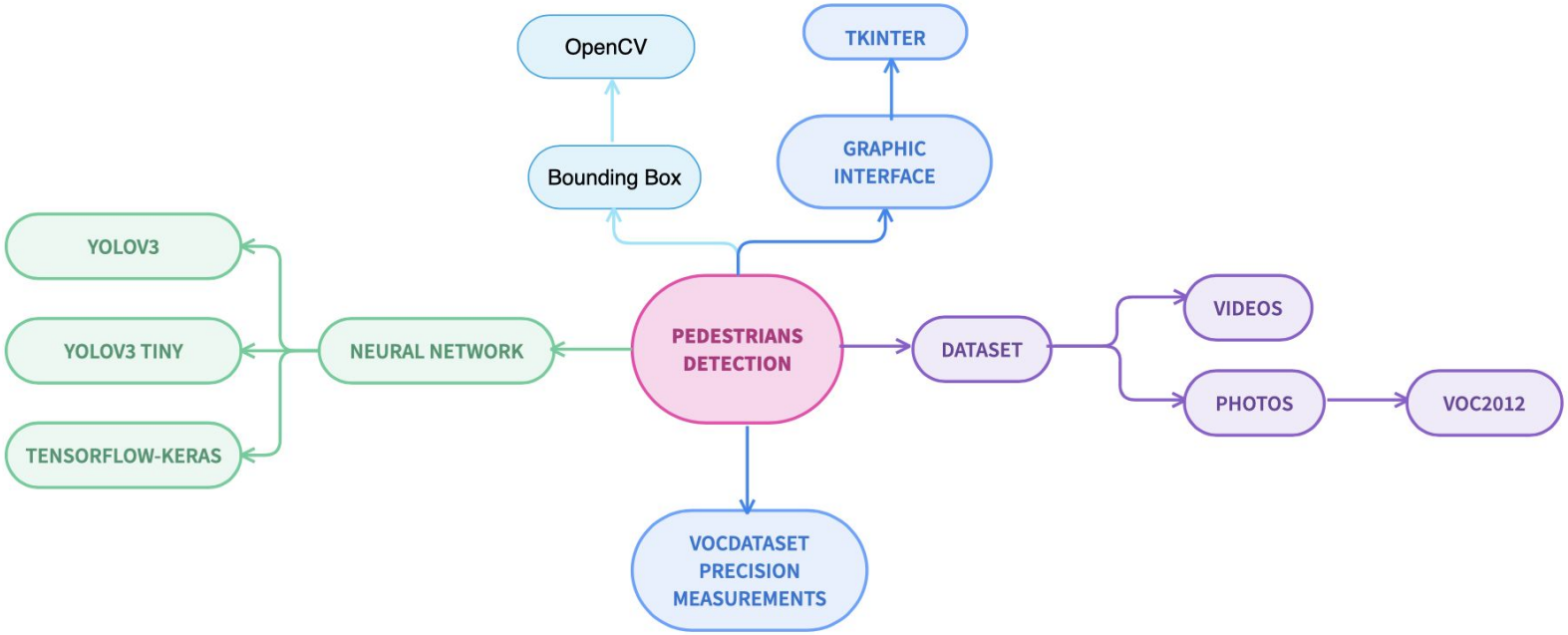
The Problem:

Statistics show that thousands of people are killed in traffic every year.

Looking at the reason of these accidents, it is almost always the lack of attention on the side of the driver.

There has been a great deal of interest in recent years in the development of pedestrian detection systems that could help reduce the number and impact of these accidents.

Initial Brainstorming



What is YOLO?

- YOLO stands for You Only Look Once
- It is amongst the most powerful detectors at the moment
- It’s A Fully Convolutional Neural Network

It's an object detector that uses features learned by a deep convolutional neural network to detect an object.

YOLO makes use of only convolutional layers, making it a fully convolutional network (FCN).

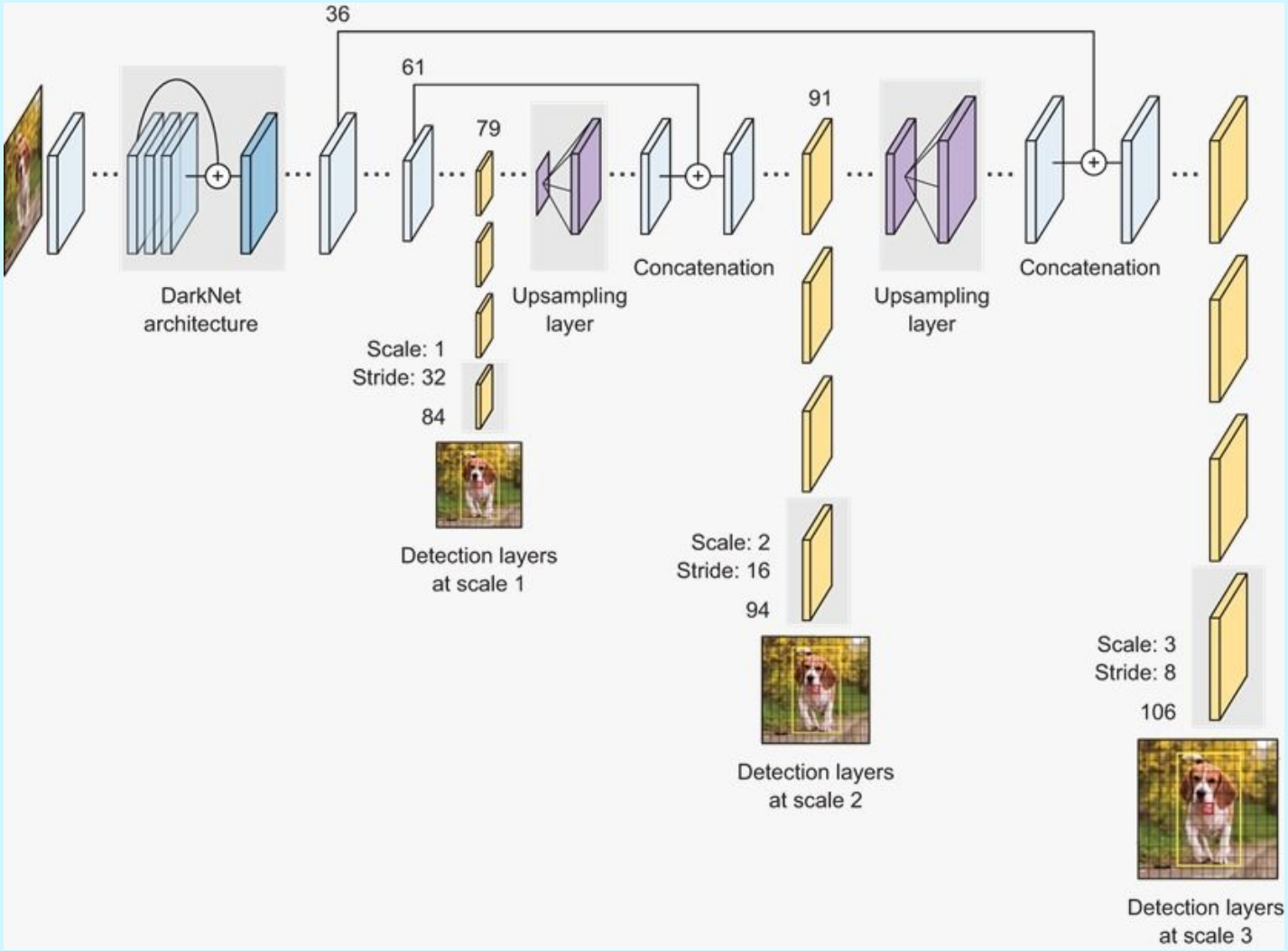
It has 75 convolutional layers, with skip connections and upsampling layers.

More about Yolo

Before detection systems repurposed classifiers or localizers to perform detection. They apply the model to an image at multiple locations and scales. High scoring regions of the image are considered detections.

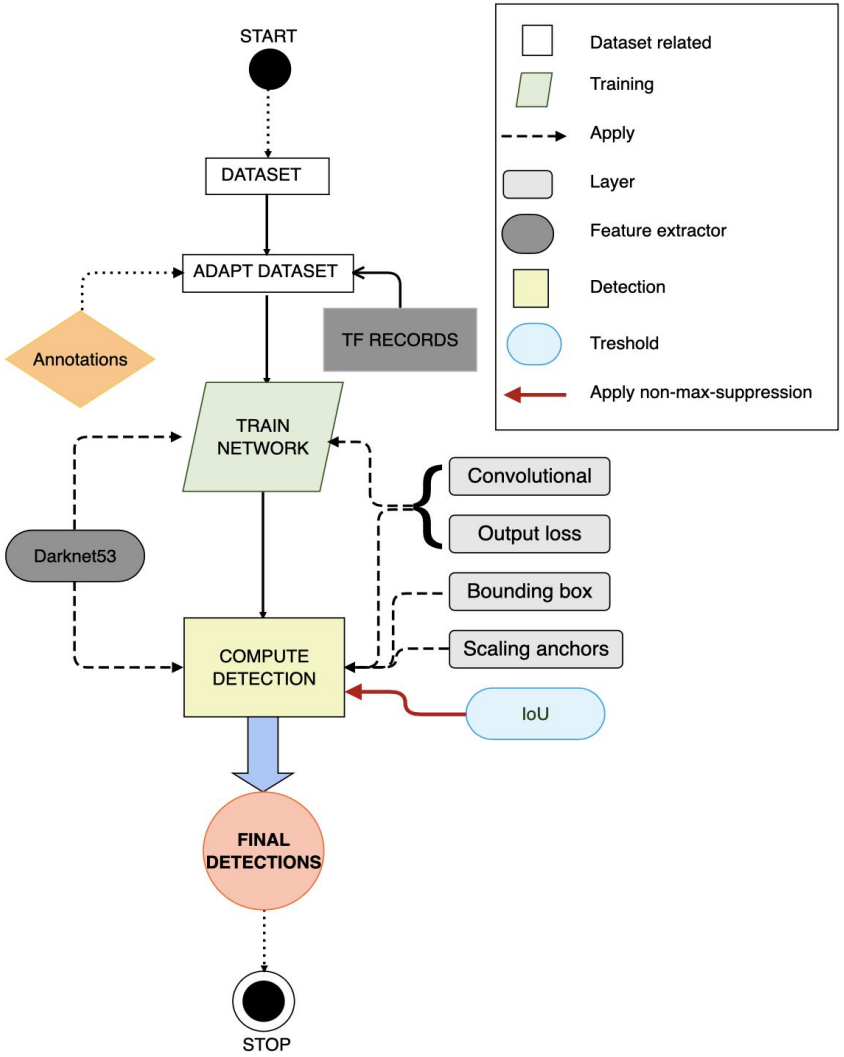
Yolo applies a single neural network to the full image. The network divides the image into regions and predicts bounding boxes and probabilities for each region. These bounding boxes are weighted by the predicted probabilities.

Each new version of Yolo has brought improvements towards accuracy and performance.


36
61
79
91
DarkNet
architecture
Upsampling
layer
Concatenation
Upsampling
layer
Concatenation
Scale: 1
Stride: 32
84
Detection layers
at scale 1
Scale: 2
Stride: 16
94
Detection layers
at scale 2
Scale: 3
Stride: 8
106
Detection layers
at scale 3

Algorithm Flow

➔ Ontology map

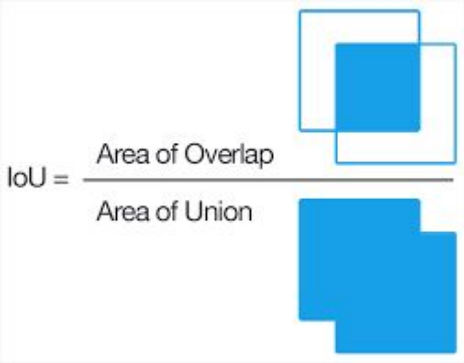


Darknet-53

	Type	Filters	Size	Output
	Convolutional	32	3 × 3	256 × 256
	Convolutional	64	3 × 3 / 2	128 × 128
1×	Convolutional	32	1 × 1	
	Convolutional	64	3 × 3	
	Residual			128 × 128
	Convolutional	128	3 × 3 / 2	64 × 64
2×	Convolutional	64	1 × 1	
	Convolutional	128	3 × 3	
	Residual			64 × 64
	Convolutional	256	3 × 3 / 2	32 × 32
8×	Convolutional	128	1 × 1	
	Convolutional	256	3 × 3	
	Residual			32 × 32
	Convolutional	512	3 × 3 / 2	16 × 16
8×	Convolutional	256	1 × 1	
	Convolutional	512	3 × 3	
	Residual			16 × 16
	Convolutional	1024	3 × 3 / 2	8 × 8
4×	Convolutional	512	1 × 1	
	Convolutional	1024	3 × 3	
	Residual			8 × 8
	Avgpool		Global	
	Connected		1000	
	Softmax			

Table 1. **Darknet-53.**

IoU



Our Dataset: Pascal VOC2012

The PASCAL VOC project provides standardised image data sets for object class recognition. VOC2012 is the largest available one containing annotations for 20 classes.

We have used the entire dataset to train our network, but only for the annotation of one class: person.

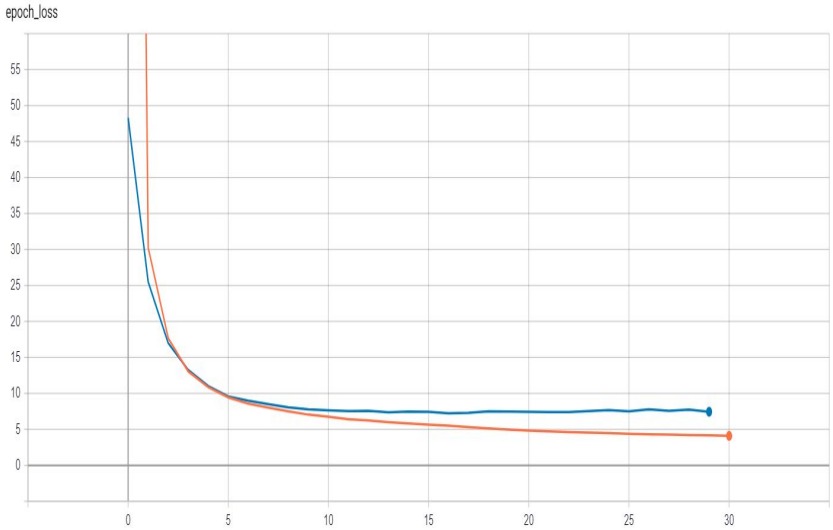
The dataset was split by VOC creators in a train-validation manner. The percentages were close to a 50%-50% division.

The total number of images was a little over 10.000.

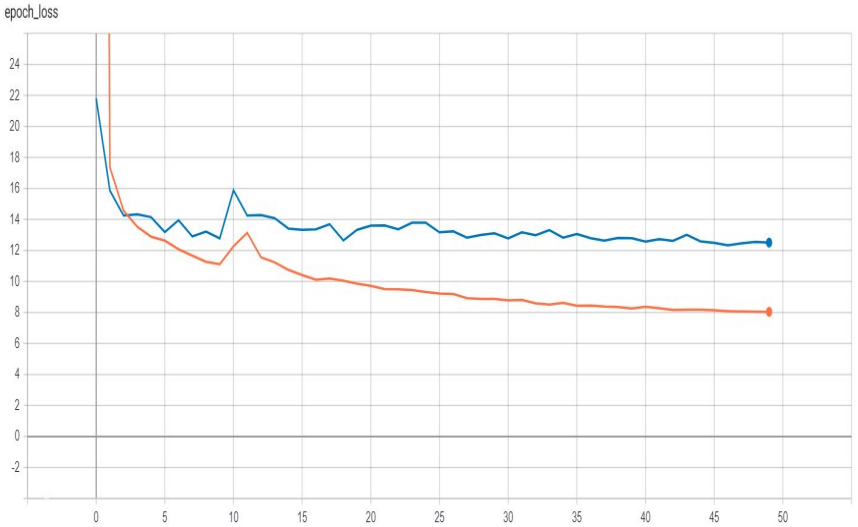
Official site: http://host.robots.ox.ac.uk/pascal/VOC/voc2012/

Training

Main Loss, Tiny Architecture, No Smoothing

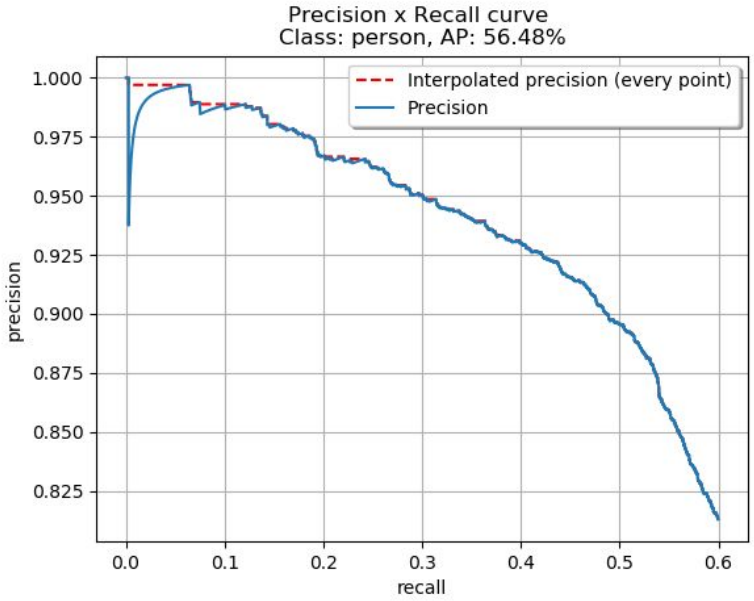


Main loss, Normal Architecture, No Smoothing

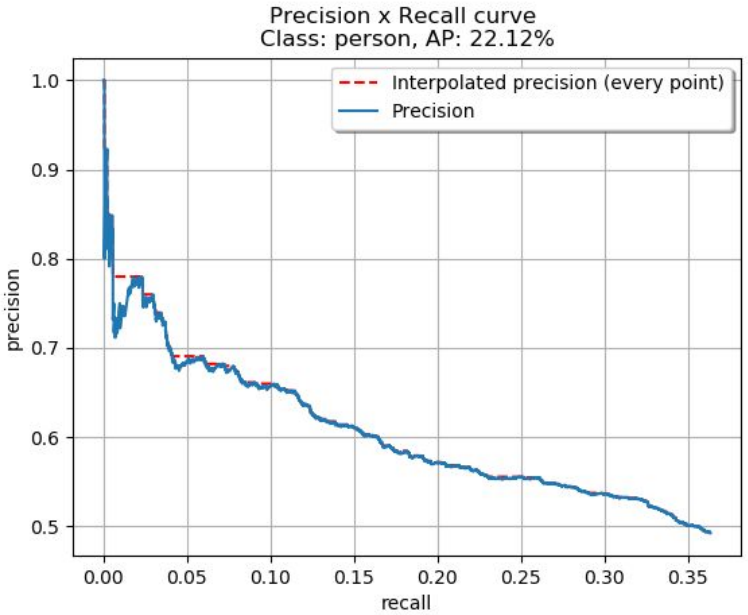


Results Tiny Architecture

Yolo_IOU=0.3, Yolo_Score=0.3, Metric_IOU=0.5



Yolo_IOU=0.3, Yolo_Score=0.3, Metric_IOU=0.75

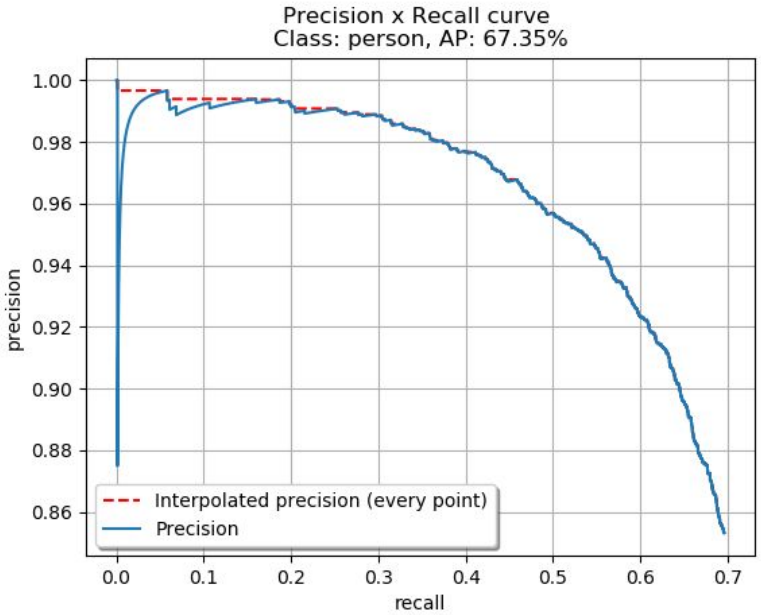


Future Improvements

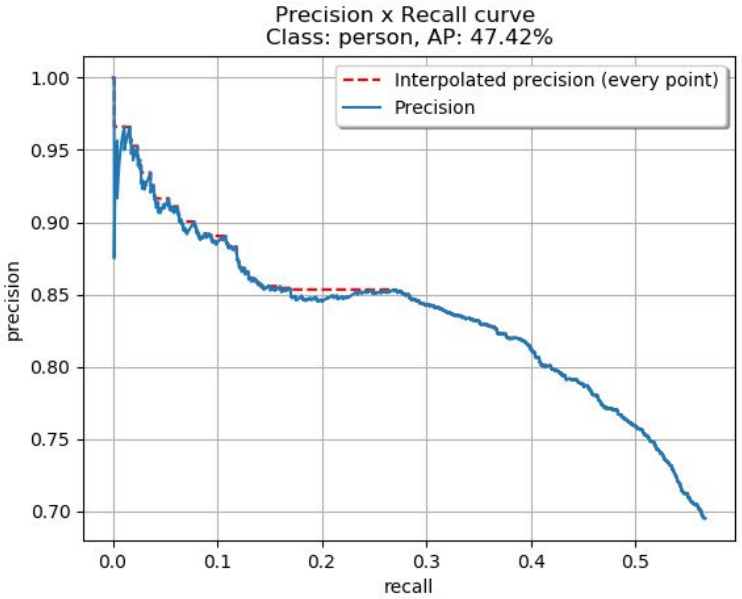
- More sophisticated algorithms for image resizing
- Data augmentation
- Larger and more various dataset
- Increasing input size to retain more detail
- Friendlier user interface

Results Normal Architecture

Yolo_IOU=0.5, Yolo_Score=0.5, Metric_IOU=0.5



Yolo_IOU=0.5, Yolo_Score=0.5, Metric_IOU=0.75



Q&A