University College Cork

Master's Thesis

Pedestrian Action Recognition

Classification and tracking

Dhanya Sringeri Jayachandra

supervised by
Dr. Gregory Provan

August 2, 2018

Contents

1 Abstract **3**

2 Introduction **3**

3 Literature Review **3**
3.1 Detection . . . 3
3.2 Convolutional neural networks . . . 3
3.3 Region Proposal network . . . 4

4 Datasets **4**

5 State of Art **4**
5.1 Parameter optimization . . . 4
5.2 Region Proposal network . . . 5
5.3 Regularization: . . . 5
5.4 L2 Regularization . . . 5

6 Accuracy **5**

7 Future work: **6**

List of Tables

List of Figures

1 Abstract

In this paper, we will discuss and try to classify pedestrians from cyclists as accurately and quickly as possible. Accuracy and Speed are both very important aspects that could highly influence the outcome in road-safety systems. Various types of Convolutional Neural Networks approches will be applied and we can compare the advantages of one model over the other. Improving accuray of object detection and tracking are our main goal.

2 Introduction

Road-Safety is a primary concern in all Autonomous and Advanced Driver Assistance systems. The smart cars are facing many challenges in Vulnerable road users recognition [VRUs]. We have seen examples of Computer Vision failing to recognize cyclists, and failure in recognizing pedestrians in dim light. Its important to classify between pedestrians and cyclists as traffic rules are different for these two. Though most autonomous cars are able to classify pedestrians, it would largely help us deal with the cyclists differently. Vulnerable road users can be detected using 2 methods. (1) Sensor based approach (2)Vision based approach In sensor based approach, the autonomous vehicles are mounted with sensors such as LiDAR systems. The depth information of the sensor based systems result in higher accuracy and these are considered supe- rior and known to have high accuaracy in todays research field. However, these methods will not be portable as they are dependent on the hardware of the car. Sensors are also expensive and some car manufacturers are reluctant to compramise the look of the car for smart features. In this paper, we will discuss and try to identify cyclists as accurately and quickly as possible using vision based approach. Accuracy and Speed are both very important aspects that could highly influence the outcome in road-safety systems. In this papers, we will compare the performance of different Convolutional neural network architectures, their performance on the vision-based networks. We can compare the perfor- mance of one model over the other. Improving accuray of object detection with each new network tried is the goal.

3 Literature Review

3.1 Detection

Pedestrian detection has taken off really well since the introduction of deep learning neural networks. Architectures like Convolutional neural networks have done exceptionally well. Early methods like VJ detector by Viola-Jones [tobecitedcitexiaofei li new 2016], Histogram of Oriented Gradients by Dalal and Triggs in 2005 [12], Based on HOG detector, Deformable Part Model (DPM) was designed to weaken the deformation effect of non-rigid objects by Felzenswalb et al. in 2008 [13]. Another variant method ChnFtrs was applied to deploy multiple registered images channels for classification by Dollar et al. in 2009 [14]. In 2013, ConvNet model was introduced to yield compet- itive results on major pedestrian detection benchmarks by Sermanet et al. [15]. In paper,

3.2 Convolutional neural networks

Convolutional neural networks were introduced in 2014, ever since then they have bet all the previous methods in terms of object detection. In 2014, R-CNN [Regional Convolutional neural network] won the imagenet

challenge by a significant margin compared to previous methods. It used a region proposal method on the image, then each proposed region was fed to seperate CNN and finally the object was detected. Fast-RCNN introduced using the regional proposal method on the feature map that was extractred by the CNN, hence saving a lot of computation time. Faster-RCNN as we have used in this model, uses its own region proposal method based on the values computed by during the feature map generator instead of using an external regional proposal algorithm. This makes the system 100 times faster than RCNN. Faster RCNN consists of 2 parts: (1)Feature Extractor
(2)Region Proposal network

Feature Extractor

: The feature extractor localises the parts of the image based on features. Some of the popular feature extractors are VGGnet, Resnet and Inception. So far, Resnet is known to have the highest accuracy. It is a very deep model with 101 hidden layers. Due to gradient explosion problem, it was impossible to have such deep layers up until the invention of Resnet. Resnet won the imagenet object detection challenge in 2015, beating other competitors in the field by significant margin. We will be using Resnet as our feature extractor in our model.

3.3 Region Proposal network

: Another CNN model that is considered in this setting is, Faster Region based convolutional neural networks:

4 Datasets

The dataset that is considered in this experiment is available for public as tsinga-dailmer dataset/

5 State of Art

: I have divided the dataset of cyclists in 3 different views. Narrow, Intermediate and wide based on the aspect ratio of the bounding box of cyclists. Its important to train the network on the different views of cyclists seperately as the cyclists look completely different from different views.

5.1 Parameter optimization

Stride size: Reduced the stride size from 16 to 8, even though the computation time is longer, more information needs to processed.

5.2 Region Proposal network

Region proposal networks generate proposals from the image fed to the network which might contain the object. So, RPN is responsible for the network to decide if an object is in foreground or background. In the first stage anchor generator, I am reducing the stride, so that more regions are proposed and none of the small cyclists are missed out. I have 4 anchors of scales 0.25, 0.5, 0.75 and 1.0. Each anchor will have 3 boxes with aspect ratio of 1:1(Intermidiate), 1:2(Narrow view), 2:1(Wide view). The maximum number of box proposals from the first stage was changed to 300. Used drop-out to avoid overfitting of the neurons and improve accuracy. Drop out increases the number of possible iterations required to converge. Hence, we increase the number of iterations.

5.3 Regularization:

Regularization modifies the objective function that we minimize by adding additional terms that penalize large weights. In other words, we change the objective function so that it becomes Error+f(), where f() grows larger as the components of grow larger and is the regularization strength (a hyper-parameter for the learning algorithm).

5.4 L2 Regularization

It can be implemented by augmenting the error function with the squared magnitude of all weights in the neural network. In other words, for every weight w in the neural network, we add 1/2 w**2 to the error function. The L2 regularization has the intuitive interpretation of heavily penalizing peaky weight vectors and preferring diffuse weight vectors. This has the appealing property of encouraging the network to use all of its inputs a little rather than using only some of its inputs a lot. Of particular note is that during the gradient descent update, using the L2 regularization ultimately means that every weight is decayed linearly to zero. Because of this phenomenon, L2 regularization is also commonly referred to as weight decay.

6 Accuracy

Model	Modifications	Narrow	Intermediate Wide
FasterRCNN +Resnet 787	No finetuning	53.2	22
FasterRCNN +Resnet 5415	Increased region proposals	55	33
FasterRCNN + Resnet	dropout	59	32
R-FCN + Resnet	No finetuning	50	33

7 Future work:

We could calculate the distance between car and cyclists to determine if any action must be taken. We could also conduct the same experiments mentioned above in dim light and night time environments to understand how well the models would perform under challenging circumstances.