ATOM-N2020

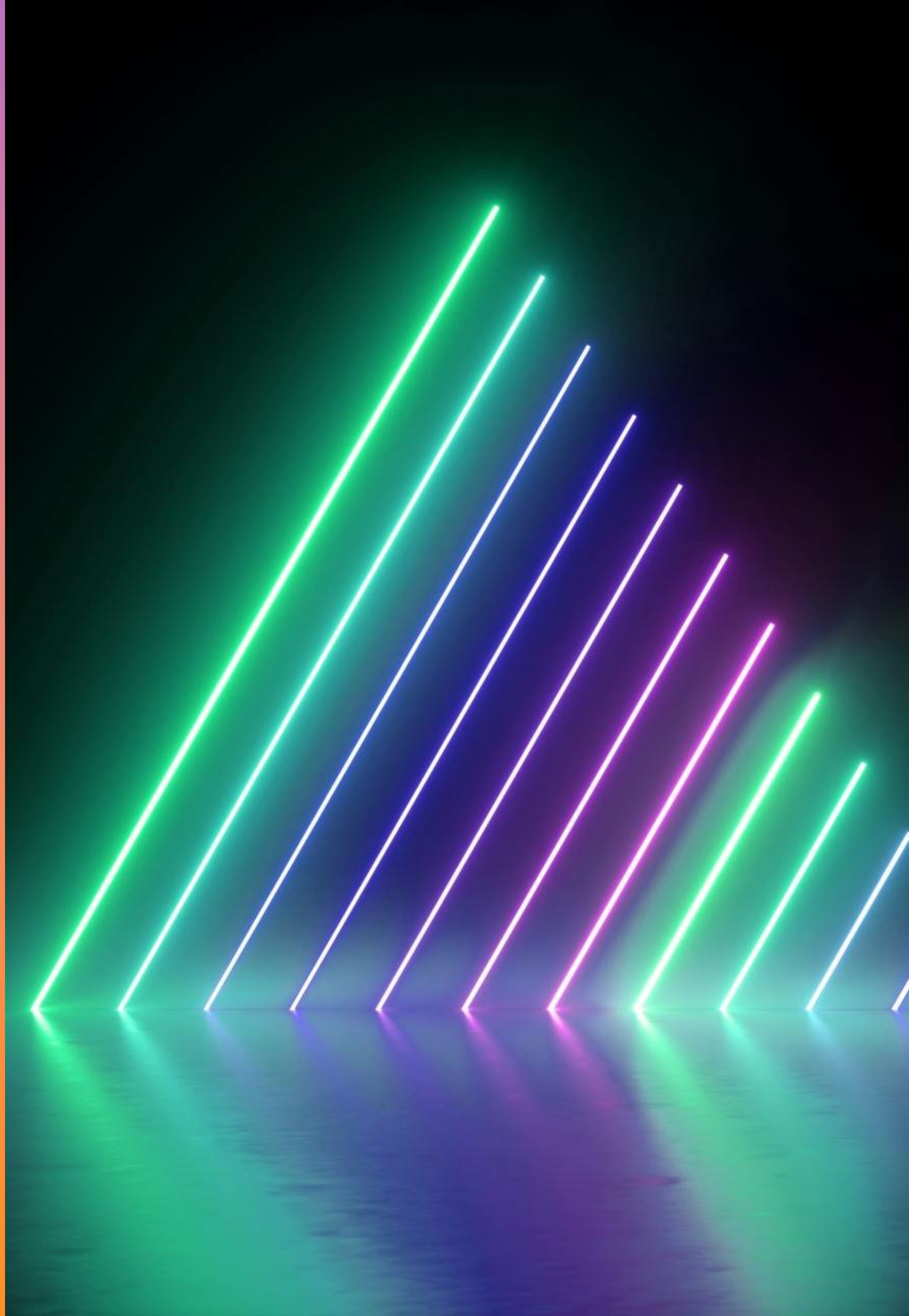
20 - 23 August 2020, Constanta, Romania

The 10th edition of the International Conference "Advanced Topics in Optoelectronics, Microelectronics and Nanotechnologies"

INTELLIGENT LIGHTING SYSTEM USING SINGLE SHOT MULTIBOX DETECTOR

OROIAN-BOCA Maria Loredana, BOCA Liana Luminița & CEBUC Radu

"1 Decembrie 1918" University of Alba Iulia, Department of Computer Science, Mathematics and Engeneering, Romania



Content

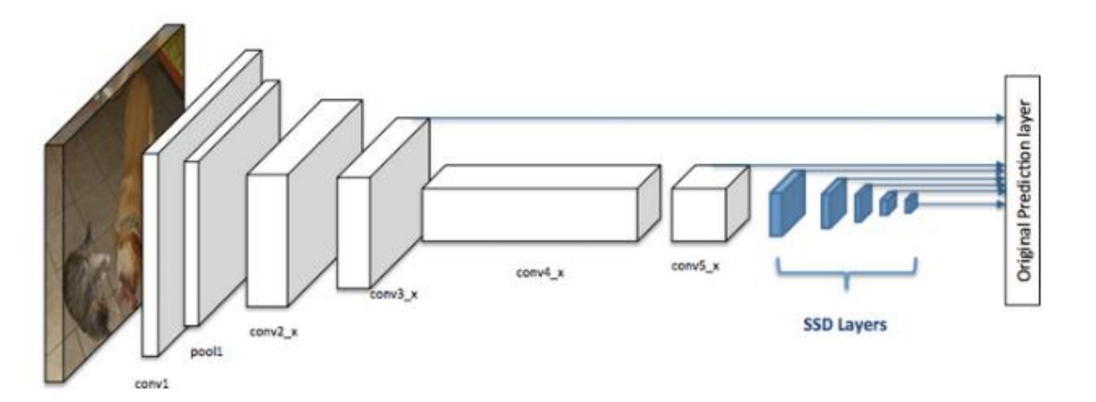
- INTRODUCTION
- METHODOLOGY
- CASE STUDY
- CONCLUSIONS
- FUTURE WORK

INTRODUCTION

Safety is an important aspect of automated self-driving vehicles. Safety is not only for drivers and passengers, but also for pedestrians, other vehicles, and two wheelers.

In this paper we developed one intelligent lighting system for motorcycles, using a convolutional neural network Single Shot Multibox Detector to detect cars at night on the road for motorcycle.

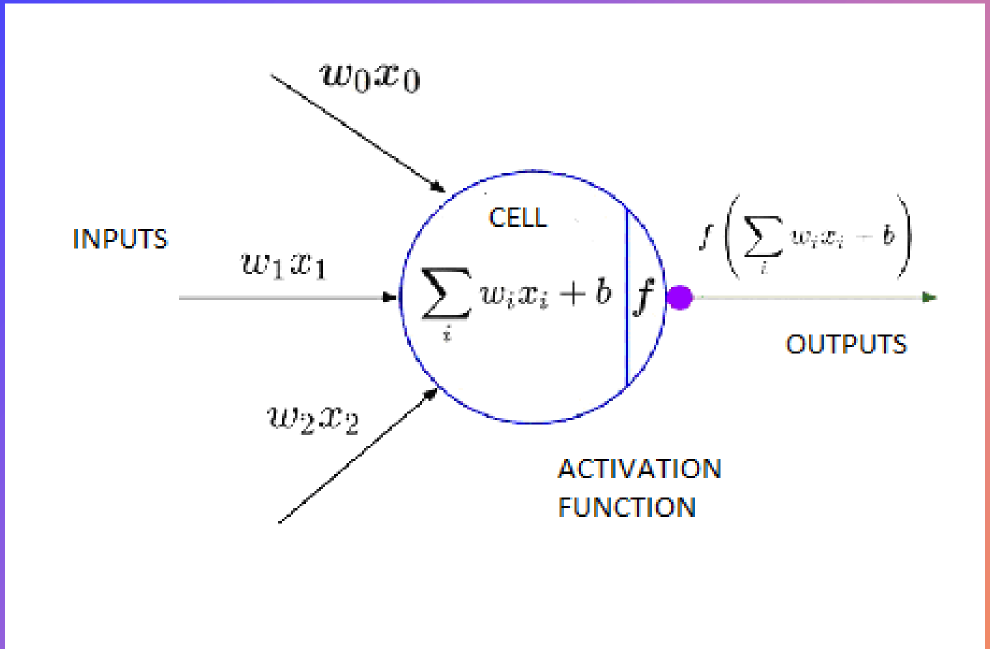
The main purpose of this paper is to reduce the number of lives lost. The system improves the visibility of a motorcyclist, especially on the highway, where speed is high and early observation of the motorcyclist is vital.



METHODOLOGY- Neural networks

- The captured image will go through the filtering process with the help of the convolution operation, as in figure [1].
- A filter is a 3x3 matrix that maps to different regions of the image. This mapping is followed by the convolution operation which involves multiplying and summing the pixel values of the image with those of the filter, thus obtaining a single number.
- The rest of the values are dropped, and then the filter will move to a right position, applying the same reasoning after which the resulting matrix is a reproduction of the initial image. Thus, the image information is preserved, and the size is reduced.
- On each small image obtained, a filter called max pooling will be applied, having the value of the matrix of 2 x 2, and permanently, of the four pixels, the value of the largest pixel will be kept, because it is considered that it has the most information. importance. (Max comes from the maximum value).
- The vector resulting from the repetition of the resulting operations is transmitted as the input of an artificial neural network.

The neuron



A neuron receives a number vector as input. Each number in the vector is assigned a weight (w - weight). Weight is the importance of an input, and the weight vector is automatically learned by the neuron. The weighted sum, to which the value of the Bias is added (automatically learned parameter) is nothing but the equation of the line (e.g. in plane 2d: y = mx + b). [3]

Thus, the neuron at the end of the learning process will be a linear classifier: it can separate the data with a straight line.

The number of inputs (x1, x2, x3...) represents the dimensionality of the problem. For a problem space with more than two dimensions, the data are separated linearly by the classifier using a hyper-plan. For an inseparable, linear data set, several levels of neurons are densely linked thus being able to classify based on nonlinear functions

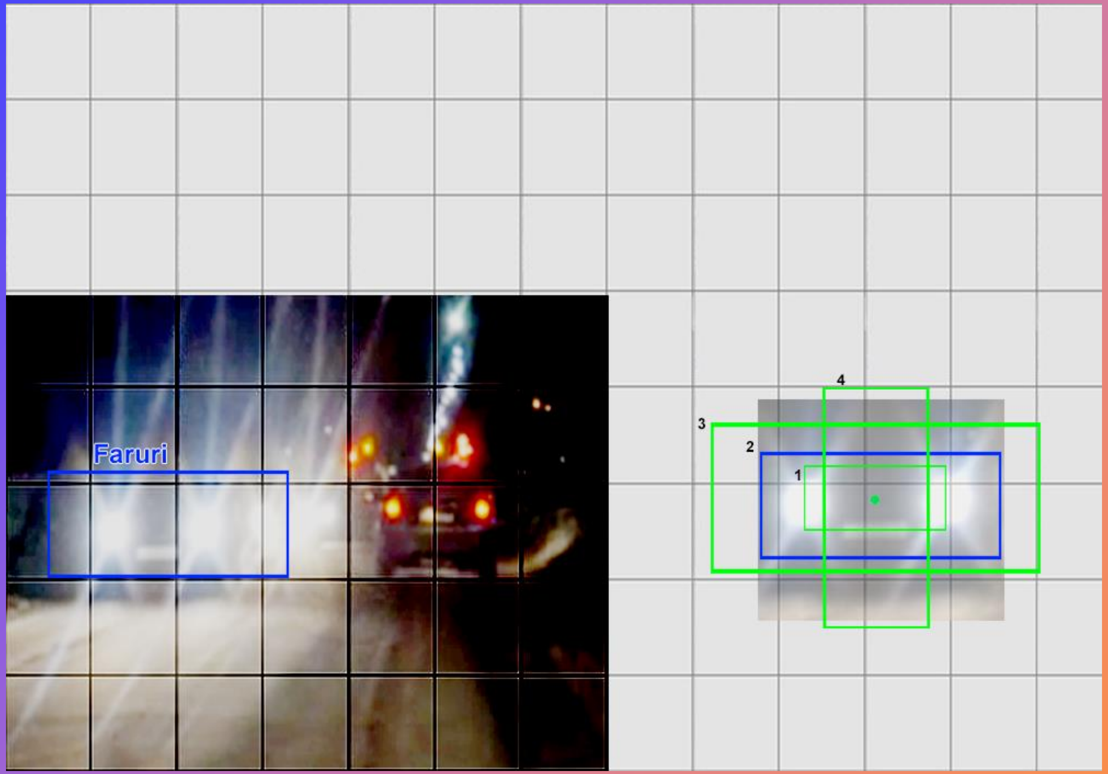
The sigmoid function maps any number (between - infinity and + infinity) between 0 and 1. This function makes a probability and is the most basic activation function

The initial weights are chosen at random, and the network is not trained. The network will perform a forward pass to obtain a result that it will compare to the image tag (manually assigned in the image tagging process in the dataset). This comparison is made using an error function. If the error value is high, the neural network initiates the back-propagation operation by updating the weight values based on the gradient descent algorithm. The processing of each image in the data set is called the epoch, and a neural network needs an epoch number greater than or equal to 1 to be considered entrained.

Single shot detection -SSD

- SSD is an image detector that uses a VGG16 network as a feature extractor. Faster R-CNN uses a network of region proposals to create Boundary boxes and uses those boxes to classify objects.
- Although it has the highest level of accuracy, the whole process takes place at 7 frames per second.
- Far below what real-time processing needs. SSD speeds up the process by eliminating the need for a network of proposals in the region. To recover the decrease in accuracy, the SSD applies several improvements, including functions on multiple scales and boxes.
- These enhancements allow the SSD to match the accuracy of Faster R-CNN using lower resolution images, which further speeds up processing speed.
- Accuracy is measured as average and measured in mAP (average accuracy).

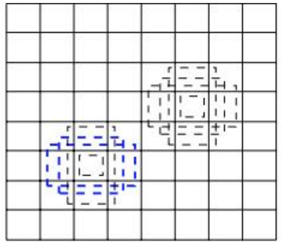
CASE STUDY



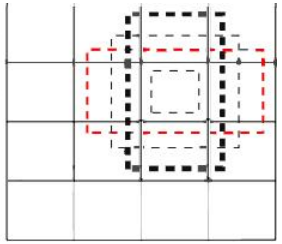
- SSD predictions are classified as positive or negative results. The SSD uses only positive results in calculating the location cost (mismatch box). If the boundary box corresponds (not the limiting box provided) it has an IoU higher than 0.5 with the basic truth, the result being positive.
- The IoU calculation is used to measure the overlap between two proposals. Otherwise, it is negative. (IoU, the intersection over the union, is the ratio of the intersected area over the united area for two regions).
- Figure 5. Boundary box selection
- Default boxes 1 and 2 (but not 3) have an IoU greater than 0.5 with the ground truth box above (blue box). So only box 1 and 2 are positive matches. Once is was identify the positive matches, will be use the limit boxes provided to calculate the cost.
- This matching strategy encourages each prediction to predict the shapes closer to the appropriate default box. Therefore, these predictions are more diverse and more stable in formation

Feature maps on multiple scales and default border boxes

Image with objects detection



8x8 map layer [1]



4x4 map layer [1]

Filtering application

- How the SSD combines feature maps on multiple scales and default delimitation boxes to detect objects on a different scale. The picture below matches a default box (in red) in the 4 × 4 map layer, but not all the default boxes in the higher resolution 8 × 8 feature map.[1]
- Maps with higher resolution characteristics are responsible for detecting small objects. The first layer for detecting the object has a spatial size of 38 × 38, a fairly large reduction in the input image. Therefore, SSDs usually have lower accuracy for detecting small objects compared to other detection methods.
- If it is a problem, it can be reduced by using higher resolution images.
- Data Increasing is important for improving accuracy. Growth data with color reversal, cropping and distortion. To manage variations in different object sizes and shapes, each training image is randomly sampled using one of the following options:
    - Using the original
    - Sampling of the original with IoU of 0.1, 0.3, 0.5, 0.7 or 0.9.
    - Samples a part of the image at random.

The sampled patch will have an aspect ratio between 1/2 and 2. Then it is resized to a fixed size and we will remove half of the training data. In addition, we can apply photo distortion.

Partial results after first testing

- SSD has poorer performance than Faster R-CNN for small scale objects. In SSDs, small objects can only be detected in higher resolution layers. These layers contain low-level features, such as edges or parts of colour, that are less informative for classification.
- Accuracy increases with the number of default limit boxes at the cost of speed.
- Multi-scale feature maps improve the detection of objects at different scales.
- Designing better default boundary boxes helps accuracy.
- SSD has a lower location error compared to R-CNN, but more classification error dealing with similar categories. Higher classification errors may use the same bounding box to make multiple class predictions.
- SSD512 has better accuracy (2.5%) than SSD300 but runs at 22 FPS instead of 59.
- The SSD can be trained end-to-end for better accuracy. Makes more predictions and has better coverage in terms of location, scaling and aspect ratios. With the above improvements, the SSD can reduce the resolution of the input image to 300 × 300 with a comparative precision performance

Non maximum suppression

- The SSD uses non-maximum deletion was applied to eliminate duplicate predictions to the same object. SSD sorts predictions by reliable scores. Starting with the most reliable prediction, the SSD evaluates the previously viewed limitation boxes if they have an IoU greater than 0.45 with the current prediction for the same class. If found, the current prediction will be ignored. At most, it was kept the first 200 predictions on the image.
- A picture is received. The picture comes in VGG16, and from there it goes on, applying convolution filters. Each layer reaches one detection per class. This way you can get a lot more features. If only the pixels in the picture are taken (width x height) a much smaller number will be obtained than the number of features extracted by each layer (layer). Much more detection / class is obtained.
- The entire filtering process depends on the single threshold value. Selecting the threshold value is essential for the performance of the model. Setting this threshold is difficult.

Functionality of non maximum suppression

- Assume that the overlap threshold N is 0.5. If there is a proposal with 0.51 IoU and it has a good trust score, the box will be removed, even if the trust is higher than many other houses with less IoU.
- Because of this, if there are two objects next to each other, one of them would be removed. A proposal with a 0.9 intersection rate is still maintained, although its confidence is very low. Of course, this is the known problem with any threshold-based technique. Below is an example of such a case.
- Only the 0.9 proposal is retained, and others will be removed. This reduces the accuracy of the model.
- The simple and effective way to deal with this case is to use Soft-NMS. We apply this idea to the example above. Instead of the case with complete elimination of the proposals with a score of 0.8, we keep the proposals, but reduce the score as shown in the second figure.
- The 0.4 scores of both proposals are calculated based on the intersection values.

Probability calculation

Example with 0.4 score

CONCLUSIONS AND FUTURE WORK

- The convolutional model for predicting detections is different for each YOLO feature layer operating on a single-scale feature map. The output values of the bounding box are measured against an implicit value. We use the VGG-16 network as a basis, but other networks should produce good results.
- To solve the problem that the system can distinguish street lighting from the headlight or stop of the vehicle in front, we used deep learning techniques. This approach is the most flexible option in terms of object recognition and the cost of production is not high.
- So, with the help of the webcam mounted in front of the motorcycle, near the headlight, the neural network monitors and captures pictures in real time. These are then processed by obtaining the class (Far / Stop) and the position of the objects found in the image. Based on the result, the long phase of the motorcycle is automatically controlled according to the other traffic participants. This network will be implemented on a Raspberry PI3 computer.
- **Future work -** With the help of a GPS module, a GPRS module and the accelerometer already mounted, the motorcycle can benefit from an intelligent global anti-theft system. This system can detect the change in the position of the motorcycle when the key is not in contact and can send an SMS to alert the owner and can send real-time GPS coordinates if the motorcycle is moving. PI 4 can monitor the battery charge level in real time and alert the user. All these implementations are possible by adding small parts of the decoding to the already existing software.