Journal of Physics: Conference Series **1448** (2020) 012003 doi:10.1088/1742-6596/1448/1/012003

# Use of deep learning algorithms for real-time detection of vessels in confined spaces using the Tensorflow framework

**S A Sánchez[1], J Campillo[1], and J C Martínez-Santos[1]**
[1]Universidad Tecnológica de Bolívar, Cartagena de Indias, Colombia

E-mail: sergio_sanchez@corposucre.edu.co

**Abstract.** Over 4515 small boat accidents were registered in the United State of America in 2012, resulting in 651 causalities and 22% of the accidents took place between two boats. It is, therefore, one of the most interesting applications for image analysis and recognition using deep learning, collision avoidance in passenger boats. Advances in parallel computing, graphic processing unit technology and deep learning have facilitated real-time image processing. The main objective of this study was to compare the performance metrics for different deep learning algorithms using pre-trained data sets. The algorithms used were: faster region-based convolutional neural networks, region-based fully convolutional network, and single shot multibox detector using the feature extractors: residual neural network, inception and convolutional neural networks for mobile vision applications to detect generic boats in confined waterways. These models were coded in Python programming language, using the framework Tensorflow and OpenCV library for image processing. The algorithms were pre-trained using the free images database posted on the web, Microsoft COCO. The use of these pre-trained models allowed making use of computers without graphic processing unit. As a result, it was found that the faster region-based convolutional neural networks and region-based fully convolutional network method compared to the single shot multibox detector method offer a small advantage precision if speed detection is not required, but the single shot multibox detector method is useful for case detectors in real time, however it did not perform as accurate when detecting small objects.

## 1. Introduction

Industry 4.0, better known as the fourth industrial revolution, promises great advances as well as technological challenges, the concept of artificial intelligence is the main character of this transformation, related to the analysis of large volumes of data "Big Data" and the use of smart processing algorithms. This creates great expectations, since this technology implies changing the paradigm of machines as simple instruction-based devices. The uses of these technologies offer great developments in the area of robotics, safety, health, transportation, among others [1].

Currently, artificial intelligence is related to other concepts such as machine learning and deep learning. Artificial Intelligence often refers to machines simulating the behavior and reasoning of the human brain. To achieve this, different techniques, including machine learning, are used. Machine learning refers to the computers' ability to learn, whether from data or techniques that use statistical methods that enable machines to learn by themselves. In its most complex use, learning uses deep learning. Deep learning, refers to a subset of techniques for classifying and relating large volumes of information that mimics neural networks [2].

Object detection is one of the main objectives when analyzing images, today, multiple technologies are based on deep learning. Typical analysis follows three different steps: Contour definition around shapes of interest, feature extraction and classification [3]. In this work, different algorithms used for feature extraction are compared, based on performance metrics between traditional methods and deep learning, applied to detection of boats in confined waterways. Detectors are trained according to the characteristics of the images and the different shapes are labeled for automatic recognition of different boat types.

For several years, researchers have been working on the detection and classification of different shapes using images analysis, applied to the fields of face recognition, traffic surveillance, collision prevention, object following, autonomous driving among many others. In the field of maritime navigation, its use has been less explored, but its potential is substantial. For instance, it can be used for remote sensing, maritime safety, protection against illegal fishing, pollution monitoring, etc. One off the most promising applications of deep learning in this field is its use for autonomous boats. Although, automatic identification of vessels offers great opportunities, the maritime environment presents many challenges, such as the movements of ships through the weather conditions [4].

A statistical analysis presented by David Rodgers about the accidents in small vessels registered by the US coast guard (USCG) reported 4515 accidents in 2012, with 651 causalities. In 73% of the fatal accidents, boat operators lacked the appropriate training by the USCG. Additionally, in 13% of the accidents, the cause was pilot's distraction and 22% of the accidents took place between two boats. It is, therefore, one of the most interesting applications for image analysis and recognition using deep learning for collision avoidance in passenger boats [5].

The performance comparisons of deep learning algorithms allow developing a selection criterion of the most suitable method in the area of autonomous vessels in order to help avoid collisions in confined waterways. As the number of boats in the ports increases, the waterways will considerably increase its density. This situation will increase the collision risk; therefore, it is important to develop systems that make it possible to detect these objects and be able to add navigation systems, the ability to make automatic decisions for collision avoidance. This requires combining artificial imaging methods and deep learning algorithms that allow to incorporate the maneuverability action of the boat, human experience and navigation rules.

## 2. Methodology

In recent years, the architecture of the deep networks has had a very significant progress, for the moment, Keras and TensorFlow framework hold a dominant position, with different-pre-trained models already included in several libraries. These libraries include: VGG16, VGG1, ResNet50, Inception V3, Xception and MobileNet. The VGG networks follow a typical pattern of classical convolutional networks. MobileNet is a simplified Xception architecture, optimized for mobile applications. The following architectures; ResNet, Inception and Xception have become a reference point for subsequent studies of artificial vision and deep learning for its versatility [6].

There are many factors that explain the recent interest in deep learning, among those factors, the following ones can be highlighted: the availability of the data sets, GPU capacity, activation functions, new architectures, regularization techniques, redistribution of data, optimization of networks, optimization of systems, system optimization and software tools with large communities. Some of the most active communities include: Tensorflow, Theano, Keras, CNTK, PyTorch, Chainer and Mxnet. All of these, have allowed solving problems using deep learning techniques in a simpler way. Nowadays, the Python language has gained a great importance in automatic learning, in comparison with other languages due to its support for the deep learning framework. Within these frameworks, TensorFlow by Google draws special attention, due to its open source stack that uses data flow graphics. PyTorch uses the Python language. Theano is a Python library that supports mathematical expressions involving tensors. CNTK is a set of tools developed by Microsoft, Open source for deep learning, Keras is a high-level bookstore created by Francis Chollet, member of the Google Brain team that allows you to choose the models that are built in Theano, Tensorflow or CNTK. Finally, MXNet is a bookstore that also focuses on deep learning with support in several languages [7].

It is complex to define a fair set of characteristics for the different object detectors; each real-life scenario may require a different approach. In order to make a decision based on the required accuracy and speed, it is necessary to know other factors that affect the performance. For instance; the type of extractors (VGG16, ResNet, Inception, MobileNet), output steps of the extractor, input of different image resolutions, matching strategy and threshold.

There are different metrics that can improve object detection algorithms based on more accurate positioning, faster speed and more accurate classification; metrics that stand out are: Intersection over union (IoU), mean average precision (mAP) and rendered frames per second (FPS). IoU is an indicator that determines how close the predicted picture is from the real picture [8]. The average metric average accuracy is the product accuracy and recovery detection bounding boxes. The higher the score the map, the more precise the network is, at the cost of execution speed [9]. Processed frames per second (FPS) is used to judge how fast the system is [10].

### *2.1. Data set*
The ResNet, Mobilnet and Inception architectures and Faster R-CNN, R-FCN and SSD models used for this research were pre-trained on the data set posted on the free web, called Microsoft COCO (common objects in context). Microsoft COCO is a data set of 300.000 images with 90 common objects [11]. The API provides different models of object detection, which compensates for the speed and accuracy of the location of delimiters boxes. For the specific case of this research, objects related to class boat that improve the performance of the algorithms, were used.

### *2.2. Comparison between deep learning algorithms for object detection*
A review of the state of the art was conducted to validate the performance of the Fast-RCNN, Faster-RCNN, R-FCN and SSD algorithms using different free databases on the web, highlighting MS COCO, IMAGEnet and PAS-CAL VOC with the purpose of reviewing parameters of speed and accuracy of these methods using different image resolutions in different contexts. In Table 1, the advantages and disadvantages of Fast-RCNN, Faster-RCNN, R-FCN and SSD methods are presented. In order to obtain the year of the method's creation, the authors use a significant contribution of these detectors of objects in images.

**Table 1.** Advantages and disadvantages of some methods for detecting objects in images.

| Method | Advantage | Disadvantages |
|---|---|---|
| Fast-RCNN [12] | The calculation of the characteristics of CNN is performed in one iteration, making the object detection is 25 times | Using a generator external candidate region created a bottleneck in the detection process |
| Faster-RCNN [13] | The RPN method allows object detection to be almost real-time, approximately 0.12 seconds per image | Despite the efficiency of the algorithm, it is not enough fast to be used in applications that require time real, as in the case of autonomous vehicles and boats |
| R-FCN [14] | The test time of R-FCN is much faster than that of R-CNN | R-FCN has a competitive mAP but lower than that of Faster R-CNN |
| SSD [15] | The use of a single network, makes the location of objects be faster than the Fast-RCNN and Faster-RCNN methods | The accuracy of object detection is less in comparison with the Fast-RCNN and Faster-RCNN methods |

## 3. Experimental results
The tests were performed using the TensorFlow framework, the Python programming language version 3.7 and the OpenCV library, all these tools are completely open software. 20 types of images and 5 videos related to ships in confined spaces are used, with modifications of 300x300 and videos of 1280x720 per frame. The methods used as generic ship detectors were: faster RCNN, R-FCN FCN and

Journal of Physics: Conference Series **1448** (2020) 012003 doi:10.1088/1742-6596/1448/1/012003

SSD-R, ResNet, Mobilnet and Inception architectures. Metrics related to the accuracy and speed of the detectors, are summarized in Table 2. The aforementioned methods were pre-trained with the free database MS COCO. These models are not only focused on recognition and classification of the images, but also the location of objects within it, drawing a bounding box around them. Figure 1 shows the accuracy of object detection algorithms.

**Table 2.** The performance metrics Faster-RCNN, R-FCN FCN-R and methods SSD with different metrics.

| Method | mAP | Speed | Input resolution |
|---|---|---|---|
| SSD Mobilnet V1 | 21 | fast | 300x300 |
| SSD Inception V2 | 24 | fast | 300x300 |
| R-FCN Resnet101 | 30 | medium | 300x300 |
| Faster R-CNN Resnet101 | 32 | medium | 300x300 |
| Faster R-CNN Inception-R | 37 | slow | 300x300 |


**Figure 1.** Performance of (a) SSD method using the inception, (b) R-FCN method using the Resnet101, (c) Faster R-CNN method using the Resnet101 and (d) Faster R-CNN method using the Inception and Resnet101.

With the results obtained, it is evident that SSD-based models with Mobilnet V1 and Inception V2 feature extractors become faster than R-CNN and R-FCN models with feature extractors based on Resnet101 and Inception V2. However, the latter methods offer greater accuracy than SSD.

For the particular case of this pilot test, which goal was to adjust a model that would allow prediction of vessels in confined spaces, it was essential to ensure accuracy and speed because these spaces are very small, the braking time of water transport means. It is totally different compared to those of land vehicles, so it is essential to make decisions with considerable time, to avoid collisions in vessels. The best fitting model was the R-FCN with the Resnet101 feature extractor in relation to the amount of hits related to vessels detected by image and the detection speed, which can be evidenced in Figure 1(b).

It is important to highlight the results, that depending on the type of application, a faster, slower or more precise model must be selected. In addition, combinations can be made between feature extractors and other parameters mentioned above, which will allow better performance depending on the application.

## 4. Conclusions

Deep learning algorithms have significantly improved their performance over the years, however, there are still several challenges. Parallel computing with powerful GPUs has reduced the training time of artificial neural networks, but pre-trained models have allowed the use of models without the need to retrain them. These are stored in the devices reducing the computing time, without having to use a GPU. This allows a starting point for researchers seeking to enter this field without having to build and train an object detector from scratch, which would require a long time. Keras and TensorFlow provide a variety of pre-trained models that can be used for this purpose.

The difference between the detectors is shrinking. Single shot detectors use more complex designs in order to increase their precision and regions-based detectors accelerate the operations to be faster. Detectors based on regions such as CNN and Faster R-R-FCN show a small advantage if speed precision is needed in real-time, single-shot detectors are used for real-time processing. However, applications must verify if it meets the requirements accurately.

The SSD method had difficulties detecting small objects, the accuracy in the detection of objects is lower compared to the methods R-FCN, Fast-RCNN and Faster R-CNN. For large objects, SSD can overcome a faster R-CNN and R-FCN with precision using lighter and faster extractors. Although many have been successful in object detection methods, there are still many challenges that must be overcome. Deep learning will have a more prospective future in a wide range of applications. As future work, it is recommended to pre-train detectors for different types of boats and elements that use deep learning methods to generate a greater contribution to collision avoidance in vessels.

**References**

[1] Wang L & Sng D 2018 Deep learning algorithms with applications to video analytics for a smart city: A survey *SciRate* **11** 8

[2] Val Román J L 2012 *Industria 4.0. La transformación digital de la industria* (Valencia: Conferencia de Directores y Decanos de Ingeniería Informática, Informes CODDII)

[3] Pérez E 2017 *Diseño de una metodología para el procesamiento de imágenes mamográficas basada en técnicas de aprendizaje profundo* (Madrid: Universidad Politécnica de Madrid)

[4] Medina C, Cuellar S & Mojica P 2017 *Inteligencia artificial y control del espacio aéreo* (Colombia: Superintendencia de Industria y Comercio)

[5] Department of Homeland Security and Coast Guard 2014 A*ccidents in small vessels Recreational Boating Statistics* (United State of America: Department of Homeland Security and Coast Guard)

[6] Sundar K S, Bonta L R, Baruah P K & Sankara S S 2018 Evaluating training time of Inception-v3 and Resnet-50,101 models using TensorFlow across CPU and GPU *Second International Conference on Electronics, Communication and Aerospace Technology* (Coimbatore: IEEE)

[7] J Huang, *et al.* 2017 Speed/Accuracy trade-offs for modern convolutional object detectors *Conference on Computer Vision and Pattern Recognition* (Honolulu: IEEE)

[8] Rezatofighi H, *et al.* 2019 Generalized intersection over union: a metric and a loss for bounding box regression *Conference on Computer Vision and Pattern Recognition* (California: IEEE)

[9] Li Y, Huang C, Ding L, Li Z, Pan Y & Gao X 2019 Deep learning in bioinformatics: Introduction, application, and perspective in the big data era *Methods* **166** 4

[10] Blanco-Filgueira B, *et al.* 2019 Deep learning-based multiple object visual tracking on embedded system for iot and mobile edge computing applications *IEEE Internet of Things Journal* **6(3)** 5423

[11] Lin T Y, *et al.* 2014 Microsoft coco: Common objects in context *European Conference on Computer Vision* (Zurich: Springer)

[12] Redmon J & Farhadi A 2016 YOLO9000: Better, faster, stronger *Conference on Computer Vision and Pattern Recognition* (Amsterdam: IEEE)

[13] Ren S, *et al.*2017 Faster R-CNN: To- wards real-time object detection with region proposal networks *IEEE Transactions on Pattern Analysis and Machine Intelligence* **39(6)** 1137

[14] Dai J, *et al.* 2016 R-FCN: Object detection region-based fully via convolutional networks *30th Conference on Neural Information Processing Systems* (Barcelona: Neural Information Processing Systems)

[15] W Liu, D Angelov, D Erhan, C Szegedy 2015 SSD: single shot multibox detector CoRR *Conference on Computer Vision and Pattern Recognition* (Amsterdam: IEEE)