# Object Detection Approaches in Images: A Weighted Scoring Model based Comparative Study

Hafsa Ouchra, Abdessamad Belangour
Laboratory of Information Technology and Modeling LTIM
Hassan II University, Faculty of Sciences Ben M'sik
Casablanca, Morocco

***Abstract*—Computer vision is a branch of artificial intelligence that trains computers to acquire high-level understanding of images and videos. Some of the most well-known areas in Computer Vision are object detection, object tracking and motion estimation among others. Our focus in this paper concerns object detection subarea of computer vision which aims at recognizing instances of predefined sets of objects classes using bounding boxes or object segmentation. Object detection relies on various algorithms belonging to various families that differs in term of speed and quality of results. Hence, we propose in this paper to provide a comparative study of these algorithms based on a set of criteria. In this comparative study we will start by presenting each of these algorithms, selecting a set of criteria for comparison and applying a comparative methodology to get results. The methodology we chose to this purpose is called WSM (Weighted Scoring Model) which fits exactly our needs. Indeed, WSM method allows us to assign a weight to each of our criterion to calculate a final score of each of our compared algorithms. The obtained results reveal the weaknesses and the strengths of each one of them and opened breaches for their future enhancement.**

***Keywords—Computer vision; object detection; images; WSM method; object detection algorithms***

## I. Introduction

Object detection consists of several subtasks such as face recognition, pedestrian detection, skeleton detection, etc., and has use cases such as surveillance systems, autonomous cars, etc.[1][2]. There are two types of approaches to object detection in images: one based on two-stage detectors and the other based on one-stage detectors. One-step object detection algorithms work by immediately detecting objects on a sample of possible locations such as Fast R-CNN [3], R-CNN [4], Faster R-CNN [5], etc. Two-step object detection algorithms will first propose a set of regions of interest and then rank the relevant regions such as SSD [6], YOLO [7], CenterNet [8], etc. The architectures of these algorithms differ from each other in terms of accuracy, speed, and required hardware resources.

These approaches rely on deep learning models that are capable of end-to-end object detection, as they use a multi-layer structure of algorithms called neural networks that allow to perform many tasks, such as clustering, classification, or regression. A neural network is composed of input, hidden and output layers, all of which are composed of "nodes" as shown in Fig. 1.

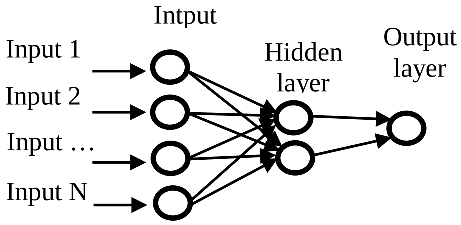

Fig. 1. A Simple Neural Network.

This paper, therefore, proposes to compare these algorithms based on the Weighted Scoring Model (WSM). Hence, we begin our comparative study by extracting the most relevant criteria for comparison and justify our choice for each criterion. Next, we present the WSM method before moving to assigning weights to each criterion and obtaining final scores for the compared object detection algorithms that are finally represented using a spider graph. The purpose of this spider graph is to show us the best detection model according to a set of scores for each criterion such as accuracy, speed, etc.

The document is organized as follows: Section II describes works related to our topic; Section III presents a background of object detection algorithms; Section IV presents our comparative study of object detection algorithms; Section V discusses the outcomes of this study and finally in Section 6 we draw a conclusion.

## II. Related Work

Object detection is a challenging task that arises in many image processing applications such as human-computer interaction, civil and military surveillance, virtual reality, and human motion analysis or image compression. This challenge rises in unconstrained environments where the tracking system will have to adapt to important variability of objects, to variations of luminosity, to occlusions (partial or total) as well as to motion detection problems.

The introduction of deep learning algorithms, especially CNNs, to computer vision problems has progressed rapidly which has led to very robust, efficient, and flexible vision systems. Since the results of the "ImageNet 2012 challenge" event [9], Deep Learning and especially convolution networks have become the best method to solve this kind of problem.

Object detection is a very active field of research that seeks to classify and locate regions/areas of an image. This field is at the crossroads of two others: image classification and object localization [11]. Research in object detection has naturally integrated image classification models, which has led to the creation of models such as SSD [6] and R-CNN [4], etc.

Many scientific works are aimed at discovering approaches to object detection in images and algorithms and techniques that exist in each approach. Many researchers have made scientific efforts in this area to compare and show advantages and disadvantages of each algorithm and techniques that detect objects in images. Several research works tried to compare algorithms and models for object detection in images [7][6][12][13][14][15].

Sanchez et al. [16] have performed a review of state of the art related to the performance of pre-trained models for object detection in order to make a comparison of these algorithms in terms of reliability, accuracy, time processed, and problems detected. In this research [16], different pre-trained models using two datasets MS COCO[17] and PASCAL VOC[18] have been reviewed for object detection such as R-CNN, R-FCN, SSD, and YOLO, with different feature extractors such as VGG16, ResNet, Inception, MobileNet.

Srivastava et al. [19] were presented a comparative analysis of 3 major image processing algorithms: SSD, Faster R-CNN, and YOLO. In this analysis, they chose the COCO dataset to evaluate the performance and accuracy of the three algorithms and analyzed their strengths and weaknesses. They implemented these models and based on the results obtained, they determined the differences between the performance of each algorithm and the appropriate applications for each. The evaluation metrics are accuracy, precision, and F1 score.

In this work by Gupta et al.[20], the COCO dataset was used to extract sample images and then they used three architectures and three extractors to build different combinations of models in order to calculate the speed and accuracy (mAP) of each of these models.

All these comparative studies are based on the results of the implementation of each model used. For the comparison criteria, they focused on two criteria: accuracy (mAP) and speed. Most of these studies compared two-step detector-based approaches with one-step detector-based approaches.

The authors show that single-stage detectors are divided into two types [21][12]: Detectors with anchors and detectors without anchors, these two types each have advantages and disadvantages. The first major problem with anchored detectors is the involvement of several hyperparameters, which makes the algorithms very slow and computationally intensive [12]. This results in a low accuracy rate and makes the unanchored detectors outperform the anchored detectors. The best-known models in this approach are: YOLO [7], CenterNet [8], CornerNet [22], FCOS [12], etc. Then, they showed that the principle of two-stage detectors is that the first stage generates a set of regions of interest by the region proposal network (RPN) and the second stage for the classification of regions of interest. Among the algorithms that are based on this principle, we have R-CNN [21], Faster R-CNN [5], etc.

Our work is also based on the comparison of these approaches and algorithms based on the results of the implementations of previous works, and then according to a set of criteria, besides these two criteria: Average Accuracy (AP) and Speed (FPS), we have other criteria that we will discover in Section IV. What distinguishes our work from the others mentioned above is that this work uses the Weighted Scoring Model (WSM) which is one of the multi-criteria decision analysis methods. This method is adopted to make this comparison and get the work that is more favored for most of the criteria and discuss the result of this comparison.

## III. Object Detection in Images

Object detection and recognition are computer vision techniques that allow the detection of object instances in images or videos [23].

Models and techniques for object detection in images generally use extracted features and learning models to recognize instances of classes of objects. These models are based on convolutional neural networks known as CNNs, which can vary in architecture, which play a key role in the construction of algorithms, but the basic principle remains the same [23].

### A. Background

*1) Convolution neural network (CNN):* CNNs are one of the most popular classes of neural networks, especially for high-dimensional data (e.g., images and videos). CNNs work very similarly to standard neural networks. A key difference, however, is that each unit in a CNN layer is a two-dimensional (or high) filter convolved with the input to that layer. This is essential in cases where we wish to learn patterns from high-dimensional input media, such as images or videos. CNN filters incorporate spatial context by having a similar (but smaller) spatial shape as the input media, and use parameter sharing to significantly reduce the number of variables that can be learned [24].

CNNs can vary in architecture that play a key role in building algorithms, but the basic principle remains the same. Fig. 2 shows the Convolutional neural network architecture. It receives an input feature map, i.e., a three-dimensional matrix whose size of the first two dimensions corresponds to the length and width in pixels of the images.

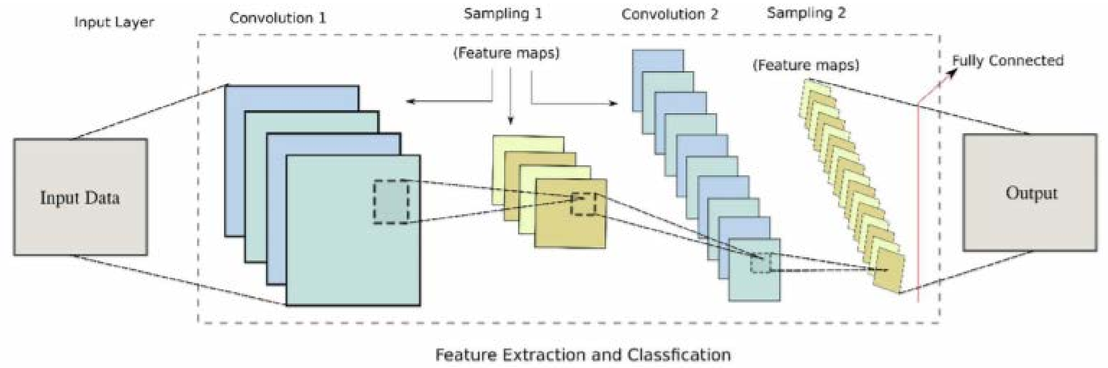

Fig. 2. CNN Architecture[10].

The size of the third dimension is three corresponding to the three channels of a color image: red, green, and blue. After this feature map, there are:

- Convolutional layer: The role of this layer is to extract fields from the input feature map and apply filters to them to compute new features, thus producing a convolved feature map.
- Max Pooling: The goal of this layer is to minimize the dimensions of the input while preserving as much information as possible because the processing of the convolutional layer is a very computationally expensive operation. And for this purpose, pooling is one of the techniques used to minimize the dimensions.
- Fully connected: After these previous layers, we have this layer whose role is to perform a classification based on the features extracted by the convolutions. And generally, this layer contains a SoftMax activation function, which provides a probability value between 0 and 1 for each of the classification labels that the model tries to predict.
- Output layer: this layer shows the result, i.e., the result of the classification algorithm used.

*2) Region proposal network (RPN):* Region Proposal Network (RPN) [5] is a fully convolutional network that simultaneously predicts object boundaries and objectivity scores at each position. It is trained end-to-end to generate high quality region proposals, used with the Fast R-CNN model that was proposed by[3] for object detection in images and reduces the number of candidate object locations by filtering out most background samples. This network is essential in the architecture of Faster R-CNN model which has been improved by researchers [5]. Fig. 4 shows the principle of Region Proposal Network.

*3) Feature pyramid network (FPN):* Feature Pyramid Network (FPN) [25] is a feature extractor and generates multiple layers of feature maps with better quality information than the regular feature pyramid for object detection. FPN consists of a bottom-up and top-down path. For the bottom-up path, it is the usual convolutional network for feature extraction. As we go up, the spatial resolution decreases, and with more high-level structures detected, the semantic value of each layer increases. For the downward path, it is to build higher resolution layers from a semantically rich layer. Some simultaneous works like RetinaNet [13] also use this type of network. Fig. 3 shows the Feature Pyramid Network architecture.

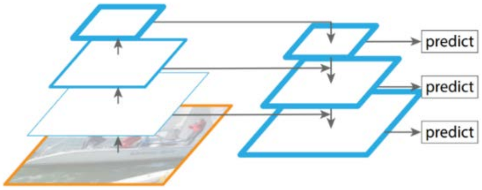

Fig. 3. Feature Pyramid Network (FPN).

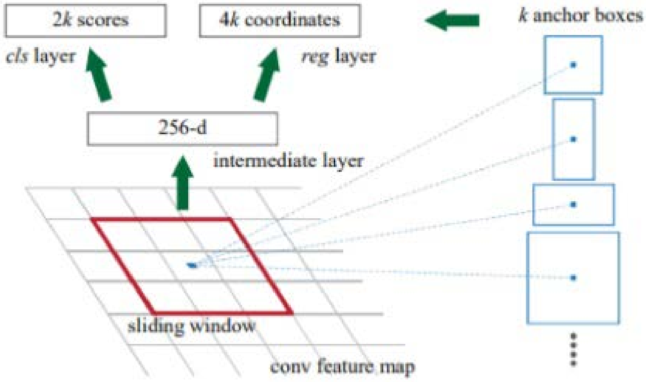

Fig. 4. Region Proposal Network (RPN).

### B. Principle of Staged Detectors and their Advantages and Disadvantages

Object detection approaches in images are divided into two categories: approach based on two-stage detectors and approach based on one-stage detectors. Each approaches contains many models, in Section IV, we will detail and compare these models.

*1) Two-stage detectors:* The principle of two-stage detectors is that the first stage generates a set of regions of interest using Region Proposal Network (RPN), which reduces the number of candidate object locations by filtering most of the background samples. The second step is the classification of regions of interest among candidate object locations extracted in the first step [23].

The advantage is that the models and algorithms that exist in this approach such as: : R-CNN [21], R-FCN [15], Fast R-CNN [3], Mask R-CNN [26], etc., have a high accuracy but the disadvantage is that at the level of processing time of each image is very slow due to repetitive detection and classification [23]. In Table II, we will list advantages and disadvantages of one and two-stage detectors.

*2) One-stage detectors:* The main goal of one-step detectors is to unify detection and classification in one step which will result in a single pass through the neural network. Thus, it predicts all candidate object locations at once which increases the speed of object detection [23] but on the other hand, they suffer low accuracy. The most known models in this approach are: YOLO [7], RetinaNet [13], SSD [6], CenterNet [8], etc.

One-stage detectors are divided into two types: Anchor-based detectors [27] and others anchor not based detectors. For Anchor-based detectors, the detection is done by a frame around the detected objects [23]. This frame is called an anchor or bounding box as seen in this figure. Concerning, anchor-not-based detectors, the detection is made by a point in the center of the detected object. The first major problem with anchor detectors is the involvement of several hyperparameters, which makes the algorithms very slow and requires much computation power. This results in a low rate of accuracy and makes anchorless detectors outperform anchor detectors [23].

## IV. Comparison of Object Detection Approaches in Images

Comparison of object detection models in images is based on the WSM method which allows to differentiate between compared objects according to a set of criteria [28].

### A. Weighted Scoring Model

Weighted scoring model [29] is a project management technique that combines quantitative and qualitative measures to facilitate operational decision-making and allows multiple criteria to be considered. Specific scoring criteria can be selected based on well-defined objectives and product metrics.

This technique assigns a weight to each criterion based on its relative importance, with the most important criterion being assigned the highest weight. To realize the application of the WSM method, we have four steps to follow [28]:

- Step 1: Select a list of features and other initiatives being considered.
- Step 2: Select criteria, including costs and benefits, on which you will evaluate each of these initiatives.
- Step 3: Determine the respective weighting of each criterion that you will use to evaluate your competing initiatives.
- Step 4: Assign individual scores to each feature, for all your cost and benefit criteria, and then calculate these overall scores to determine the ranking of the list of features.

*1) Comparison criteria:* Each approach contains many different models as presented in the sections above, and each model has a set of criteria that help the user to make the decision to choose the best performing and most adaptable model for his project. The choice of criteria for comparing each model is extracted from previous work that has enriched the object detection domain. We have identified five criteria:

- Average Accuracy (AP): This criterion indicates the accuracy value of each object detection model. It depends on the quality of the input images, the number of training samples, the model parameters, and the required accuracy threshold. The values presented in Table IV of this criterion are scores from 1 to 5 according to the following intervals to which the accuracy value (AP) of each model belongs.
- Detection time (FPS): This criterion indicates the number of frames processed per second for each object detection model. It expresses the processing speed of the model. In Table IV, the values presented of this criterion are scores from 1 to 5 according to the following intervals to which the SPF value of each model belongs.
- Real-time: This criterion shows how well the model works for real-time object detection. This criterion is evaluated in three values: 1 which means poor, 4 which means fair, 5 which means good.
- Number of stages: this criterion shows the number of stages of each model and designs the category of detectors. We have two categories: One-stage detectors and two-stage detectors.
- Simple network structure: This criterion shows if the model is easy and simple to use for object detection. This criterion is evaluated as a Boolean value that shows the availability of this criterion for each model.

Table I shows the intervals to which the accuracy value (AP) and detection time (FPS) of each model belong and their scores.

Table IV shows the scores from 0 to 5 of each criterion corresponding to each model according to the evaluation of each criterion that we explained in section 5.1.1.

TABLE I. AP and FPS Scores for each Interval

| **Intervals** | **Scores** |
|---|---|
| [0,10] | 1 |
| [10,20] | 2 |
| [20,30] | 3 |
| [30,40] | 4 |
| [40,50] | 5 |

TABLE II. ADVANTAGES AND DISADVANTAGES OF OBJECTS DETECTION MODELS

| **Model / method** | **Category** | **Advantages** | **Disadvantages** |
|---|---|---|---|
| Fast R-CNN | Two-stage detectors | High detection accuracy, Low misdetection rate | Non-real-time detection, speed of object detection per image is low |
| Faster R-CNN | | | |
| R-FCN | | | |
| Mask R-CNN | | | |
| CenterNet | One-stage detectors | Object detection can be in real-time, Simple network structure, speed of object detection per image is high. | Low detection accuracy, Poor results for small and dense objects, Easy to mislocate |
| CornerNet | | | |
| RetinaNet | | | |
| YOLO v2 | | | |
| YOLO v3 | | | |
| SSD | | | |
| FCOS | | | |

TABLE III. VALUES OF THE CRITERIA CORRESPONDING TO EACH MODEL ACCORDING TO THE RESULT OF THEIR IMPLEMENTATION

| **Criteria** | **Fast R-CNN [3] (VGG-16)** | **Faster R-CNN** [14] **(ResNet-101)** | **R-FCN** [15] **(ResNet-101)** | **Mask R-CNN** [26] [30] **(ResNet-101-FPN)** | **CenterNet [8] (DLA-34)** | **CornerNet [14] (Hourglass-104)** | **RetinaNet** [13] **(ResNet-101-FPN)** | **YOLO v2** [31] | **YOLO v3** [32]**(DarkNet-53)** | **SSD [6]** | **FCOS [11] (ResNet-101-FPN)** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| AP | 35,9 | 34,9 | 31,5 | 35,7 | 41,6 | 42,1 | 40,8 | 44 | 33 | 46,5 | 41,5 |
| FPS | 0,5 | 5 | 12 | 5 | 28 | 4,1 | 5,4 | 40 | 20 | 22 | 8,1 |

TABLE IV. COMPARATIVE STUDY OF CRITERIA FOR EACH MODEL

| **Criteria** | **Values of criteria corresponding to each model** | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Fast R-CNN [3] (VGG-16)** | **Faster R-CNN** [14] **(ResNet-101)** | **R-FCN** [15] **(ResNet-101)** | **Mask R-CNN** [26] [30] **(ResNet-101-FPN)** | **CenterNet [8] (DLA-34)** | **CornerNet [14] (Hourglass-104)** | **RetinaNet** [13] **(ResNet-101-FPN)** | **YOLO v2** [31] | **YOLO v3** [32] **(DarkNet-53)** | **SSD [6]** | **FCOS [11] (ResNet-101-FPN)** |
| AP | 4 | 4 | 4 | 4 | 5 | 5 | 4 | 5 | 4 | 5 | 5 |
| FPS | 1 | 1 | 2 | 1 | 3 | 1 | 1 | 5 | 3 | 3 | 1 |
| Real-time | 1 | 1 | 4 | 1 | 5 | 1 | 1 | 5 | 5 | 5 | 1 |
| Number of stages | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Simple network structure | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

*2) Comparative study:* The values of the average accuracy (AP) and detection time (FPS) criteria as shown in Table III are extracted from the studies of many researchers that we have already cited in the previous sections. All these models are trained and tested on the same dataset called MS COCO.

MS COCO dataset has been described by Lin et al [17] as a set of data for large-scale object detection, segmentation, and subtitling. It contains a large, richly annotated dataset of images representing complex everyday scenes of common objects in their natural context: photos of 91 types of objects that would be easily recognizable by a 4-year-old child.

### *B. Application of Weighted Scoring Model*

The Table V shows the WSM results for each object detection model in the image. The allocation of weighting percentages is done according to the importance of the criterion. Because of their mandatory nature, priority is given to these two criteria: Average precision (AP) and detection time (FPS), a weight of 0,3 is assigned to each of these criteria. The second category of importance is given to this criterion: Real-time, a weight of 0,2 is assigned to this criterion. The next two criteria are not of great importance: number of stages and simple network structure, each of these two criteria have a weight of 0,1. The weight of the total scores is equal to 1.

TABLE V. RESULTS OF WSM

| Criteria | Weight | Requirement score | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Fast R-CNN [3] (VGG-16) | Faster R-CNN [14] (ResNet-101) | R-FCN [15] (ResNet-101) | Mask R-CNN [26] [30](ResNet-101-FPN) | CenterNet [8] (DLA-34) | CornerNet [14] (Hourglass-104) | RetinaNet [13] (ResNet-101-FPN) | YOLO v2 [31] | YOLO v3 [32] (DarkNet-53) | SSD [6] | FCOS [11] (ResNet-101-FPN) |
| AP | 0,3 | 1,2 | 1,2 | 1,2 | 1,2 | 1,5 | 1,5 | 1,2 | 1,5 | 1,2 | 1,5 | 1,5 |
| FPS | 0,3 | 0,3 | 0,3 | 0,6 | 0,3 | 0,9 | 0,3 | 0,3 | 1,5 | 0,9 | 0,9 | 0,3 |
| Real-time | 0,2 | 0,2 | 0,2 | 0,8 | 0,2 | 1 | 0,2 | 0,2 | 1 | 1 | 1 | 0,2 |
| Number of stages | 0,1 | 0,2 | 0,2 | 0,2 | 0,2 | 0,1 | 0,1 | 0,1 | 0,1 | 0,1 | 0,1 | 0,1 |
| Simple network structure | 0,1 | 0 | 0 | 0 | 0 | 0,1 | 0,1 | 0,1 | 0,1 | 0,1 | 0,1 | 0,1 |
| Weighted scores | 1 | 1,9 | 1,9 | 2,8 | 2,8 | 3,6 | 2,2 | 1,9 | 4,2 | 3,3 | 3,6 | 2,2 |

## V. DISCUSSION

According to the previous results, the YOLO v2 model is the best performing model for object detection in images. It is a model that was proposed by J. Redmon and A. Farhadi in 2016 in [31].

In terms of speed, this model is one of the best object detection and recognition models, capable of recognizing objects and processing frames up to 40 FPS (in our case) and sometimes up to 150 FPS depending on the architecture used. However, in terms of AP accuracy, YOLO v2 was not the top model but has good accuracy (AP) of 44% when trained on the MS COCO dataset. However, Fast R-CNN, Faster R-CNN which was the state of the art at that time has an AP of 35.9% and 34.9% successively. YOLO v2 belongs to single-stage detectors. Generally, the architecture of this type of detector is simple and easy to use compared to the other models based on two-stage detectors such as Fast R-CNN, Faster R-CNN, R-FCN, and Mask R-CNN.

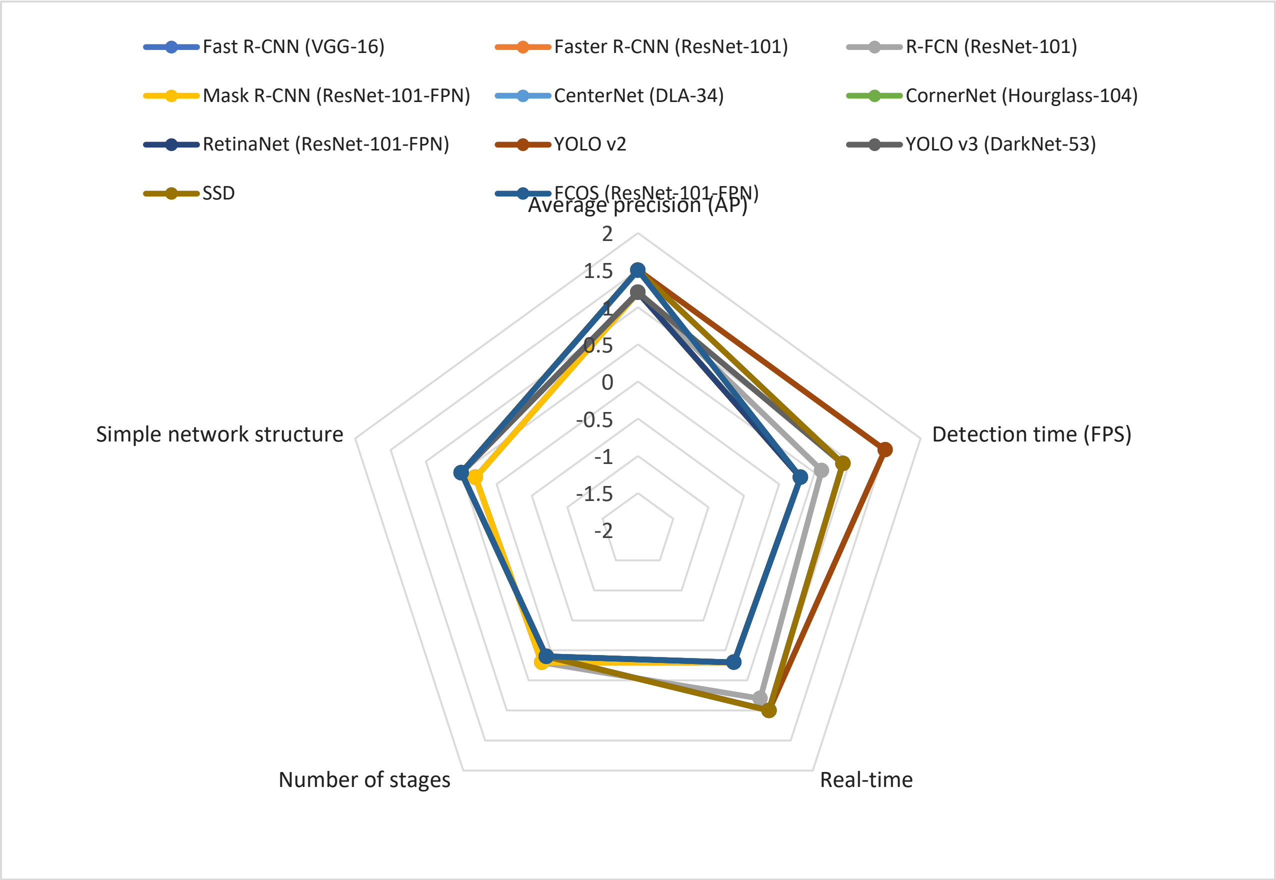

Fig. 5. Spider Chart Multi-Criteria Decision.

The other models CenterNet [8], CornerNet [22], RetinaNet [13], YOLO v3 [32], SSD [6], and FCOS [12] are quite efficient but not as efficient as YOLO v2 in terms of speed (FPS) and accuracy (AP) because these models have sometimes low accuracy and slow speed compared to YOLO v2 so they do not work well in real-time. This result is reflected in the multi-criteria radar graph presented in Fig. 5.

Many works do not consider the use of multi-criteria decision analysis (MCDA) methods which is a valuable tool that can be applied to many complex decisions.

It can solve complex problems that include qualitative and/or quantitative aspects in the decision-making process [33]. In the literature of this field, there is a great lack of this kind of comparison like this article which uses one of the multi-criteria analysis methods such as AHP, WSM, MAUT, and WPM, etc.

## VI. Conclusion and Future Work

In this paper, a comparative study based on a weighted scoring model is presented. This study is a comparison of models for object detection in images. This work starts by identifying a set of relevant works that adopt the different models of object detection in images. Then, we described the main architectures of the models that are cited in these works. We have also seen the advantages and disadvantages of the models studied in this article. Thus, we defined the WSM method and then we identified a set of criteria of each model to realize this comparison.

According to the result of our comparison, the best performing model is YOLO v2 because it has high accuracy (AP) and the frame rate per second (FPS) is fast, this means that this model works well in real-time against other models sometimes are slow and they have low accuracy. Based on the Weighted Scoring Model method, the scores of each of the studied models are obtained.

These scores helped us to establish a general classification between these models, but they also showed their strengths and weaknesses concerning each studied criterion.

In future work, we will study the algorithms and models that are effective for the classification of satellite images, and we will try to make an implementation of the most efficient model for the detection and classification of images, especially satellite images. This work provides a contribution to computer scientists and data scientists to help them choose between the different existing models and algorithms, according to their needs and the criteria that matter most to them. The aim of this study is to help the user to make the decision to choose the most efficient model for his project.

## References

[1] G. Informatique, M. D. Nuzillard, U. R. Champagne-ardenne, M. J. Boonaert, and M. De Douai, “SCIENCES ET TECHNOLOGIES Spécialité Application de techniques d ’ apprentissage pour la détection et la reconnaissance d ’ individus Université Lille 1,” vol. 072, 2012.

[2] M. Saadia, “Détection et suivi d ’ objets Remerciements,” 2019.

[3] R. Girshick, “Fast R-CNN,” Proc. IEEE Int. Conf. Comput. Vis., vol. 2015 Inter, pp. 1440–1448, 2015.

[4] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Region-Based Convolutional Networks for Accurate Object Detection and Segmentation,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 38, no. 1, pp. 142–158, 2016.

[5] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 6, pp. 1137–1149, 2017.

[6] W. Liu et al., “SSD: Single shot multibox detector,” Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 9905 LNCS, pp. 21–37, 2016.

[7] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 2016-Decem, pp. 779–788, 2016.

[8] X. Zhou, D. Wang, and P. Krähenbühl, “Objects as Points,” 2019.

[9] O. Russakovsky et al., “ImageNet Large Scale Visual Recognition Challenge.”

[10] B. Jan et al., “Deep learning in big data Analytics: A comparative study,” Comput. Electr. Eng., vol. 75, no. September 2018, pp. 275–287, 2019.

[11] “A Gentle Introduction to Object Recognition with Deep Learning.” [Online]. Available: https://machinelearningmastery.com/object-recognition-with-deep-learning/. [Accessed: 19-Mar-2021].

[12] Z. Tian, C. Shen, H. Chen, and T. He, “FCOS: Fully convolutional one-stage object detection,” Proc. IEEE Int. Conf. Comput. Vis., vol. 2019-Octob, pp. 9626–9635, 2019.

[13] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, “Focal Loss for Dense Object Detection,” Proc. IEEE Int. Conf. Comput. Vis., vol. 2019-Octob, pp. 502–511, 2019.

[14] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 2016-Decem, pp. 770–778, 2016.

[15] J. Dai, Y. Li, K. He, and J. Sun, “R-FCN: Object Detection via Region-based Fully Convolutional Networks,” Adv. Neural Inf. Process. Syst., pp. 379–387, 2016.

[16] S. A. Sanchez, H. J. Romero, and A. D. Morales, “A review: Comparison of performance metrics of pretrained models for object detection using the TensorFlow framework,” IOP Conf. Ser. Mater. Sci. Eng., vol. 844, no. 1, 2020.

[17] T. Y. Lin et al., “Microsoft COCO: Common objects in context,” in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2014, vol. 8693 LNCS, no. PART 5, pp. 740–755.

[18] M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, “The Pascal Visual Object Classes Challenge: A Retrospective,” Int. J. Comput. Vis., vol. 111, no. 1, pp. 98–136, 2015.

[19] S. Srivastava, A. V. Divekar, C. Anilkumar, I. Naik, V. Kulkarni, and V. Pattabiraman, “Comparative analysis of deep learning image detection algorithms,” J. Big Data, vol. 8, no. 1, pp. 1–22, 2021.

[20] A. Gupta, R. Puri, M. Verma, S. Gunjyal, and A. Kumar, “Performance Comparison of Object Detection Algorithms with different Feature Extractors,” 2019 6th Int. Conf. Signal Process. Integr. Networks, SPIN 2019, pp. 472–477, 2019.

[21] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., pp. 580–587, 2014.

[22] H. Law and J. Deng, “CornerNet: Detecting Objects as Paired Keypoints,” Int. J. Comput. Vis., vol. 128, no. 3, pp. 642–656, 2020.

[23] H. Ouchra and A. Belangour, “Object detection approaches in images: a survey,” Thirteen. Int. Conf. Digit. Image Process. (ICDIP 2021), vol. 11878, p. 118780H, Jun. 2021.

[24] S. Khan, H. Rahmani, S. A. A. Shah, and M. Bennamoun, “A Guide to Convolutional Neural Networks for Computer Vision,” Synth. Lect. Comput. Vis., vol. 8, no. 1, pp. 1–207, 2018.

[25] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, “Feature Pyramid Networks for Object Detection,” 2017.

[26] P. Doll, R. Girshick, and F. Ai, “Mask_R-CNN_ICCV_2017_paper,” arXiv Prepr. arXiv1703.06870, pp. 2961–2969, 2017.

[27] K. Duan, L. Xie, H. Qi, S. Bai, Q. Huang, and Q. Tian, “Corner Proposal Network for Anchor-Free, Two-Stage Object Detection,” Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 12348 LNCS, pp. 399–416, 2020.

[28] B. El Khalyly, A. Belangour, M. Banane, and A. Erraissi, “A comparative study of microservices-based IoT platforms,” Int. J. Adv. Comput. Sci. Appl., vol. 11, no. 8, pp. 389–398, 2020.

[29] A. Griffith and J. D. Headley, “Using a weighted score model as an aid to selecting procurement methods for small building works,” Constr. Manag. Econ., vol. 15, no. 4, pp. 341–348, 1997.

[30] X. Wang, R. Zhang, T. Kong, L. Li, and C. Shen, “SOLOv2: Dynamic and Fast Instance Segmentation,” 2020.

[31] J. Redmon and A. Farhadi, “YOLO9000: Better, faster, stronger,” Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017, vol. 2017-Janua, pp. 6517–6525, 2017.

[32] J. Redmon and A. Farhadi, “YOLOv3: An Incremental Improvement,” 2018.

[33] I. Hamzane and B. Abdessamad, “A built-in criteria analysis for best IT governance framework,” Int. J. Adv. Comput. Sci. Appl., vol. 10, no. 10, pp. 185–190, 2019.