Yolo Algorithm for object detection

ESLAM MAGDY REZK EBRAHIM HASSANIN  
CT002-3-2-AIM-L-4 APU2F2111CS(IS) / APD2F2111CS(IS)  
Asia Pacific TechnologyTechnology Park, Malaysia  
TP062816

*Abstract*—This topic deals with how the Yolo algorithm works and how to determine things through it. It also focuses on the applications within it and what is its backbone. Many questions we must ask when we talk about the Yolo algorithm, one of the most important algorithms ever used in artificial intelligence.

# Summary

The

Object detection is gaining traction across industries, from personal security to corporate productivity. Object detection is often used to identify faces. This section addresses current research on object detecting algorithms. Object detection is used widely. An overview of deep learning for broad object detection covers recent achievements in deep learning for object detection. Object recognition finds objects in natural pictures that fit into predefined categories. Around 300 research entries included topics such object proposal generation, feature representation, training approaches, context modelling, popular datasets, and assessment measures. CNN used to build mobile robots for navigation, surveillance, and EOD. The use of vision systems in robots allows them to identify their surroundings and what they see. According to the results, Faster R-CNN can accurately identify objects detected by SSD in real-time applications. Object detection has been popular recently.

substantial research since it relates to image and video interpretation. R-CNN can manage occlusion, low pixel density, and turbulence. Deep neural networks have excelled in picture classification. Regression is used to get bounding boxes. Multi-scale inference is used to identify objects at cheap cost. Les performances sont évaluées à l'aide du PASCAL- He et al. proposed a deep learning framework to simplify deep neural network training. Deep network depth is crucial in research. When trained with higher network depth, deep networks may be able to converge. Similarly, accuracy quickly reaches saturation. The degradation problem has a built-in solution that doesn't increase training error. The authors couldn't identify an honest or better solution using ImageNet and COCO. Visual identification skills have improved in the previous two years. Huge CNNs' properties are unknown compared to HOG and SIFT. To examine fine-tuning, we ran many CNN feature learning experiments on two datasets: detection and classification. Pre-training boosts performance says the author. Inside-Outside Net (ION) promises effective multi-scale and contextual visual recognition. This object detector employs both internal and external data. Spatial recurrent neural networks for external environment. A skip pooling algorithm extracts data at various scales and pitches. Results rose from 73.9 to 76.4 percent map using PASCAL VOC 2012. From 19% to 33% on the COCO dataset.

Several researchers are now interested in AUVs. Fish recognition using CNN and three optimization methods. Simplify the network and train faster. Data augmentation is used to supply extra data samples and to keep up with the training procedures. Overfitting was addressed using the Drop Out approach. The network was reconditioned using the loss function. Finally, the author demonstrated the model's underwater item expansion. All smart automobiles are designed to perceive their environment. All these vehicles must take precautions to avoid road dangers. No deep learning model can predict its accuracy. The author provided many approaches for assessing object detector uncertainty. The automotive pedestrian dataset is often used to identify base earn. Tensor flow employs Yolov3 to help beginners. Projections should be less explicit to decrease uncertainty. Quick deep learning object identification is Yolo (You Only Look Once). The neural network outperforms DPM, SSD, and R-CNN in one assessment. Preprocessing at 45 FPS. YOLO9000 and YOLOv2. On the Pa 2007 dataset, our models beat real-time detection approaches with 78.6% accuracy at 67 FPS. YOLOv2 is now YOLOv3. YOLOv3 works well. This method constructs a weights model from all photographs and assigns a class name to each item. To use this strategy, split each image into layers. The pre-trained weight model always gets the best photo label. YOLOv3 runs at 28.2 mappings in 22 MS, three times faster than SSD. It can't distinguish little things in high-resolution photographs, for example. ACF-PR-YOLO was created by Liu et al. ACF takes elements from images and mixes them into YOLO net area suggestions. On the public TDBC, our approach outperforms YOLO by 13.69% and Single shots multi box detector by 25.27%. Many academics are interested in utilizing aerial photographs to identify vehicles. The YOLO approach combines three public aerial image datasets. Testing with rotating, small, and thick objects shows good performance. From above, the car's features are hard to see. Faster R-CNN and YOLOv3 are the fastest methods. Two models were evaluated using data from Stanford and PSU. Vidyavani et al. propose real-time detection using YOLOv3 and deep learning. On the COCO dataset, this YOLOv3 technique performed better. Sang et al used YOLOv2 to detect objects in photographs. K-means ++ clustering is used to group the training dataset bounding boxes into six groups. Fusion of many features promotes feature extraction. Datasets for testing and training include Compcars and BIT - vehicle validation. A YOLO security camera monitors human movement. Model evaluated using Liris human activity dataset [1].

Regression based on method for YOLO:

YOLO's algorithm analyzes the images you submit to it, videos, or live broadcasts, and then divides them into S grids multiplied by S. Also, each alum inside the Yolo algorithm has its own work and task, if the network structure within the algorithm consists of two layers, the layer and those two layers are connected to each other. Surrounding it are wrapping layers of 24 layers, If, as shown in the picture, the output process is (C +B x 5) x S x S, as the variable B represents the goals of the categories that the network wants to reach, or in the correct sense of the expected goals and are specific to each network, while the variable C is specific to the number of classes within the network as well, in At the end, boxes are placed around the expected classes, indicating the probability of the classes for the tensor data, Although the Yolo algorithm relies on speed in setting goals, it faces a problem in identifying small goals, because the network is not divided in detail because there are many goals in the same network, Because of these problems, the fifth version of the Yolo algorithm was released, which solves these problems and focuses and develops on the color of the image or the colors within the broadcast as well as the frames, and trains them on the categories that are present within them, and thus the mosaic and data are improved, especially the measurements within the data and spaces and their modification And by using the Faster R-CNN mechanism, through which targets or small objects were captured and the algorithm was trained to deal with all sizes of objects, As for the remote sensing mechanism, the R-FCN mechanism was used to analyze images and increase accuracy by identifying targets [2].

Chart

Description automatically generated

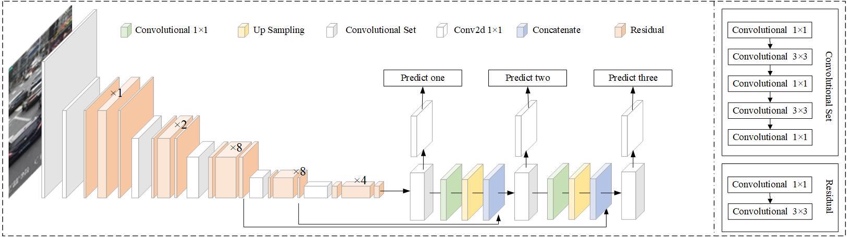
1. Yolo network structure

Backbone network for YOLO Algorithm:

Computer vision is rapidly being used in devices like smartphones, self-driving cars, and drones. Autopilot technology, for example, necessitates cars' ability to recognize their surroundings, analyze the view, and respond appropriately; AI editing and other smartphone functionalities need precise object identification in photographs; Object detection is the most extensively used fundamental functional component for scene processing in embedded systems, and as a result, it has attracted a lot of attention.

Object detection is a key task in computer vision. It is required for VCR and Image Understanding (IU). Certain manual feature-based item recognition techniques have been phased out in preference of neural networks due to their poor accuracy and restricted environmental adaptability. Because of its rapid, end-to-end learning, and, scalable architecture, CNN have gotten a lot of attention. R-CNN, Fast, Faster, and Mask R-CNN are only a few of the RPN-based object identification algorithms that update their object detection accuracy. To ensure great accuracy, these techniques rely on extensive GPU compute capabilities. Many practical applications that need actual improvement on a cognitively restricted platform, on the other hand, may not profit from such accuracy increases. In addition, with newer technologies such as autonomous driving, models' ability to detect things in real time is crucial. A speedy and precise object detector is required for autonomous vehicles, Augmented Reality (AR), and other devices. SSD, DSSD, YOLO series methods, and Retina Net are among of the most recent suggestions. Because of its balance of detection precision and reliability, YOLOv3 is the most widely utilized object detector in practical applications. YOLOv3 still needs a significant amount of computational overhead and battery usage to maintain reasonable detection performance. The model can only run on the embedded device if it is shrunk. The YOLO and SSD networks are available in smaller variants, but their detection accuracy is reduced. Reducing model size and FLOPs without reducing detection accuracy becomes a major challenge when employing object detectors on embedded systems. To decrease parameters and model size, the most popular method is to redesign a more effective system. As a result, SqueezeNe, and other similar algorithms may maintain detection accuracy while lowering size of model and floating point of float calculations (FLOPs).

To increase detection accuracy, we created a lightweight network model based on YOLOv3. A backbone network for feature extraction is developed utilizing just 16 percent of darknet-53's specifications. Meanwhile, we created the Mini-YOLOv3 Multi-Scale Feature Pyramid network, which is based on a basic U-shaped architecture and improves multi-scale object detection performance. In comparison to YOLOv3, it has fewer learnable parameters and FLOPs. Mini-YOLOv3 achieves a 77.6% decrease in FLOPs, a 76.76 percent reduction in parameters size, and equal accuracy rate to YOLOv3 on the MS-COCO test data. At 67 frames per second, Mini-YOLOv3 achieves mAP-52.1 [3].



1. YoloV3 Backbone Structure
2. Limitations

Things can be analyzed and known, but Yolo algorithm faces some problems, including that it will not be able to identify small things such as rows of ants, for example, because the network is not divided and fully focused on the thing in front of it, but with the development of Yolo versions and the addition of Faster-CNN to it Trained on small things, she can discover small things quickly and with high accuracy, such as YOLO version 5, and also add some categories to it so that you can know all the classes or things that are presented to them.

So, the question here, what is the purpose from studying yolo algorithms?

It is possible to learn how to track and know things, and also can develop on small things, knowing all the details inside the place, and through the Yolo algorithm, it is also possible, for example, to know how many people have visited the place.

1. Conclusion

In the summary, it was understood how the Yolo algorithm works, what is its backbone, and what is the difference between it and the rest of the algorithms of the same type as well, how it can be developed, and what problems it encounters. Yolo algorithm is one of the most important algorithms used to determine things, whether they are small or large and contain on high frames, great accuracy, and speed as well.

##### References

1. Venkata, K, Saia, P, Ramya, C, Venkatramaphanikumar, S (2020, November). *Object detection using stacked YOLOv3* retrieved from ResearchGate: DOI:[10.18280/isi.250517](http://dx.doi.org/10.18280/isi.250517)
2. W. Wu, H. Liu, L. Li, X. Wang, Z. Wang, J.Li, Y. Chang, (2021, October 29) *Application of local fully Convolutional Neural Networks combined with YOLO v5 algorithm in small target detection of remote sensing image* Retrieved from plosone: <https://doi.org/10.1371/journal.pone.0259283>
3. Q. -C. Mao, H. -M. Sun, Y. -B. Liu and R. -S. Jia, "Mini-YOLOv3: Real-Time Object Detector for Embedded Applications," in IEEE Access, vol. 7, pp. 133529-133538, 2019, doi: 10.1109/ACCESS.2019.2941547.