Object Identification Model using Deep Reinforcement Machine Learning Concept for Image

Saurabh Tiwari¹, Dr S Veenadhari², Dr Sanjeev K Gupta³

¹Phd Research Scholar, Department of Computer Science and Engineering RNTU, Bhopal 462003, Madhya Pradesh India
²Associate Professor, RNTU Department of Computer Science and Engineering RNTU, Bhopal 462003, Madhya Pradesh India
³Professor, RNTU Department of Electronics and Communication Engineering RNTU, Bhopal 462003, Madhya Pradesh India

Email: saurabh_tietech@rediff.com

Abstract. This paper presents a model which gives the detailed process of object identification. We need to identify class and location of object in image for completing process of object identification. Proposed model works on the principle of reinforcement learning which takes action on the basis of rewards and experiences. Normally methods in literature uses sliding window which moves in same direction but proposed algorithm provides a variable mask which moves 360 degree for identifying object using action history vector proposed with RL also not only this work focuses on localization like other work but also used class information with Softmax classification able to classify multiple object in single image with efficient time which is novel. Proposed mask acts as agent and focuses on proposed candidate reason this saves time and works in efficient manner for identification. Agent depends on transformation action and by applying top down reasoning it gives location of object. Classification is done using Softmax classification as we are having features of image by CNN. Reinforcement learning concept used for training of agent and Pascal voc dataset used for testing. Analysis of only 10 to 25 regions is sufficient with proposed work to identify first instance of object. Experiment and performance evaluation shows the efficiency of proposed work.

1. Introduction

Now days deep machine learning is widely used for object identification it become more crucial if there is more than one object in single image we need to repeat the counter and process for identification. Definition of object says Objects [1] are standalone things which having well define boundary and centre and separates itself from background. Most common approach for object detection is to define problem as classification and with convolution or mask used on feature map for region classification. Two commonly used training algorithm known as supervised and unsupervised trainings are used for training model but proposed work used reinforcement training. Many techniques [3]-[4] try to improve computational time and space complexity with different algorithm which work well. Faster R CNN uses variable mask over feature map for region proposal with trained neural network. Proposed approach follows strategy that different path for different object. Also we not proposed candidate region with low level cues but high level concept of logical reasoning are used for candidate region selection. We proposed action strategy with Reinforcement learning which works by paying attention on current region and transform bounding box in a way that it can be more focused to target object. We define reward function which is equivalent to how much current bounding box covers the ground box. Deep-Q-network algorithm is used for learning policy for localization. We calculate precision and recall for Pascal Voc Dataset and obtain competitive results. Factor to compute processing
time for an image we define IAET image average evaluation which gives time required for complete process of identification.

2. Related Work
Several work proposed which define different algorithm and methodology for object detection. The research method given in [5] depends on neural network for identifying market products. For high accuracy and fast training inception v3 proposed in tensor flow library. Alex Krizhevsky proposed model named Alex net [6] in 2012 which uses CNN for image classification it consist five convolution and three fully connected layer. Within image RCNN [7] was first successful attempt to detect object efficiently but due to its slow speed and high computational cost different improvement proposed for detection process. FAST RCNN was next version of RCNN which uses ROI pooling for reducing amount of forward passes by passing it to sub regions. FASTER-RCNN [8] replaced selective search used in fast RCNN with region proposal network (RPN) and reduced time taken up to half a second. independent region proposal for detection of objects B. Schiele et al [9] given a detailed and structured analysis of eight region proposal on the basis of recall, ground truth annotation, precision keeping objective in mind that they want to give method among eight which is feasible to use for other projects. Every researcher in region proposal is working to reduce the number of evaluated region so detection process can be made fast and efficient. M. B. Blaschko et al proposed a theory using concept of branch and bound [11] which use few locations to predict region with high score. Aleksis [10] et al proposed deep reinforcement learning based object detection which replaces greedy ROI selection process with sequential attention mechanism using object detector with RPN. Reinforcement learning [12] work on the experiences objective is to define a policy that decide sequential action by enhancing future rewards. Recent trends [13]-[15] in reinforcement learning is to combine deep neural network with RL and obtaining policy from RL to solve the problem normally RL model defines policy or value function by evaluating deep features RL can successfully produced algorithm to play many Games like Atari[16] and others. Figure 1 shows the view of simple Reinforcement Learning concept to understand things clearly. Also by employing Deep RL many problems related to computer vision like object identification, localization, object tracking has solved efficiently.

![Reinforcement learning concepts](image)

Figure 1 Reinforcement learning concepts
Object identification problem solved by reinforcement leaning will follow normally two algorithm named policy gradient and Q-learning. Proposed work follows the concept of Q learning for policy framing David Silver [14] et al proposed policy gradient algorithm for RL with continuous action, comparative to stochastic policy deterministic gradient descents evaluated more efficiently and produce better computation results. We used Softmax classifier for class label because it gives probabilities for each class we compare Softmax with SVM classifier and we found that interpretation of probabilities is easier than marginal scores and we are calculating probabilities thus we elected Softmax classifier for proposed work.

3. Object Identification Model and scheme
This section will give detail about model and scheme used for region proposal, features extraction, action and reward of RL and proposed model with algorithm.

3.1. Object identification model
We proposed an improved and boosting model for identifying object figure 2 will give theoretical view of proposed model. we named Model ODDRL-NET that is object detection deep reinforcement learning model. It having three convolution layer based on CNN responsible for features generation also we are using kernel Relu and strides as shown in figure to generate features map. Input of model is a pre-processed image taken from Pascal Voc dataset. Model has two fully connected layers named FC4 and FC5. FC4 is responsible to processed input from features map and provides output with 512 nodes after dropout it concatenated with action history vector which finally provides the state representation. After this we use Q-learning for correct localization on the basis of probabilities calculated Softmax classifier will predict class information of object.

3.2. Problem statement with action dynamics

We defined problem statement as Markov Decision Process (MDP) which helps work with decision making. Reinforcement learning concept works on environments, state, action, agent and reward we have image as environment and on different state with action history agent will try to transform a tight mask called bounding box around object with the help of different action history. The agent makes transition from one state to another it marks current region which is visible and past region. Agent mark rewards as positive and negative as well for getting visible it mark 1 and other for 0 which helps the algorithm to calculate score function for further accuracy of algorithm.

Proposed work uses eleven action for movement which includes scaling, translation and special action stop which finishes the first pass of algorithm if image having single object. Translation consist of four moves (left right up down) with fast moves with double capacity for each four, scale changes includes (up Scale, down Scale) which maintain and set aspect ratio. Bounding box in implementation simply represented by \( B = (x_1, y_1, x_2, y_2) \) it changes shapes with a factor \( \beta \) given by following equation

\[
B_w = \beta^* (x_2 - x_1) \quad \beta_h = \beta^* (y_2 - y_1)
\]

Where the value of \( \beta \) lies in range of 0 and 1 given by \( \beta \in (0, 1) \) new position of bounding box is defined by adding or eliminating \( \beta_w \) or \( \beta_h \) from x or y co-ordinates. Finally stop action is the only action which stops the process and indicates successful identification of single object and terminates process and place the mask on initial position for multiple objects present in the environment it will restart process . With the stop action mask will overlap ground truth and intersection over union [17] (IoU) computed, which is used in many visual attention model to calculate the effective identification.
3.3. State transition and Reward calculation

State is a factor which moves the process further. We define some parameters like observed region vector (Or) and action history vector (Ha). Transition of two states depends on (Or, Ha) we used state generalization for effective transition. Feature vector or is the vector extracted with current region using first three convolution layers of ODDRL-Net model acting as a pretrained model for vector extraction as proposed by Girshick et al. Any region attended by the agent deformed to match the required input size (284x284) by using technique provided by [18] J. Donahue et al.

Function which leads to search towards objective refer as reward its count one gain when agent identifies an instance of object after having one appropriate action. Measuring gain or improvement depends on intersection over union (IoU) calculated on predicted and target bounding box. We get reward when the state changes mathematical computation of reward will done with moving one state to other. We can define it say current state s is changes to s’ while applying action and every state have current box b than reward will be defined as

\[ R_a(s, s') = \text{sign} \left( \text{IoU}(b', g) - \text{IoU}(b, g) \right) \]

If reward is improving with current to next state than we get reward positive and otherwise negative. The reward scheme is binary because without quantization difference in IoU will be very small and hard to predict which action is bad or good. When we found target object in visible region it means time to have stop action and identification phase. The stop action is a special action because it lead the box in a terminal state which doesn’t change in this case we get IoU zero so at termination we can define reward function as

\[ R(s, s') = \begin{cases} 1 & \text{if } \text{IoU}(b, g) \geq 0.6 \\ -1 & \text{otherwise} \end{cases} \]

Where IoU (b, g) is the overlap ratio between terminal box b and ground truth g. Proposed agent follows Greedy strategies which provide a minimum number of steps which helpful to make proposed scheme efficient and systematic.

4. Training and Testing

Training and testing is done using Pascal Voc Dataset for the very first step we pre-processed images using a python script to get suitable dimension and space. We already know that an image is collection of neuron and if we reduce size it means less neuron to handle. We created XML file with label image software than a CSV file which contains map with that we can classify images easily. We implemented our model in tensor flow [19] using anaconda python version 3 this make easy this model to use with different architectures devices.

The objective of the agent is to find a policy to transform bounding box in a manner that it find appropriate action which enhance the reward when it interact with the environments. We Framed policy for training with reinforcement learning using Q-learning algorithm and solved problem of Q-learning with algorithm given by D silver et al [20]. We used shallower faster RCNN concept in our features and proposals which is trained in a way that is acts as pretrained network for Q network it having advantages that learning of Q function is relatively faster and RCNN will acts as feed forward for Q network. Also we are having full features learning as we using customize Pascal-Voc Dataset learned features improved performance of Q network. We trained agent by initializing Q-network parameter randomly and agent interacted with the environment with Greedy strategy. Environment having information about ground truth box and for calculating Reward we use IoU with the current bounding box so here we get action with positive and negative record. We used stochastic gradient descent and back propagation algorithm also for handling dropout we use [21] to regularized dropout.

Now when we completed phase of training it’s time to test model for identification .behaviour of agent is more important whether it’s learned to find region out of proposed region by algorithm. Initially we don’t know how much object are present in single image so we allow the agent to run with 200 steps it means we are going to evaluate 200 region per image . The search will be restart in two conditions if agent has used stop action or 30 steps passed without stop it means agent is trapped in obscure region we decided 30 steps because we getting results in fewer than thirty steps also it make the search time efficient.
5. Performance Evaluation and Discussion
We setup performance evaluation based on detection quality, precision and recall. Every object identification algorithm works to improve these parameters; evaluations shows that proposed work have improved technique compared to other methods in literature. We evaluated proposed scheme with customize Pascal VOC dataset and also added many other images which used for training and testing purposes for proposed model. We implemented this work with Anaconda python 3.5 combined with tensor flow. All operation executed in 8th generation Intel core i7 8557U processor, 8 GB DDR-4 RAM and hard drive with 500 GB capacity with windows10. Implementation details of this work shown in figure 4 and figure 5 which gives identification of two sample test image for giving the quality of project. As shown in figure 4 and 5 the value of detection is more than 80 percent which clears the proposed box and ground truth box having more than 70 percent intersection over union which shows mask is properly placed by agent and based on score class calculated by Softmax algorithm is correct.

5.1. Estimation of precision
Precision tells how accurate is our prediction proposed approach will considers all attended region which predict object we have a detector which scores all region on single search step. Normally in object detection precision depends on intersection over union (IoU) given by ratio between area of intersection and area of union of predicted bounding box and ground truth box. Precision means positive predicted value which given for class in classification process. Precision is calculated by

\[ \text{Precision} = \frac{\text{True positive}}{\text{True positive} + \text{False positive}} \]

| Methods          | Aro  | Bike | Bird | Boat | Bottle | Bus  | Car  | Cat  | Chair | Cow  | Table | Dog  | Horse | Bike | Person | Plant | Sheep | Sofa | Train | Tv   | MAP   |
|------------------|------|------|------|------|--------|------|------|------|-------|------|-------|------|-------|------|--------|-------|-------|------|------|-----|------|
| R-CNN[10]        | 64.2 | 69.7 | 50.0 | 41.9 | 32.0   | 62.6 | 71.0 | 60.7 | 32.7  | 58.5 | 46.5  | 56.1 | 60.6  | 66.8 | 54.2   | 31.5 | 52.8  | 49.9 | 57.9 | 64.7 | 54.2 |
| Region [38]      | 44.6 | 55.6 | 24.7 | 23.5 | 6.3    | 49.4 | 51.0 | 57.5 | 14.3  | 35.9 | 45.9  | 41.3 | 61.9  | 54.7 | 44.1   | 16.0 | 28.6  | 41.7 | 63.2 | 44.2 | 40.2 |
| RPN-300[39]      | 77.0 | 78.1 | 69.3 | 59.4 | 38.3   | 81.6 | 78.6 | 86.7 | 42.8  | 78.8 | 68.9  | 84.7 | 82.0  | 76.6 | 69.9   | 31.8 | 70.1  | 74.8 | 80.4 | 70.4 | 70.4 |
| Proposed method  | 87.4 | 83.6 | 76.8 | 62.9 | 59.6   | 81.9 | 82.0 | 91.3 | 54.9  | 82.6 | 59.0  | 89.0 | 85.5  | 84.7 | 84.1   | 52.2 | 78.9  | 65.5 | 85.4 | 70.2 | 75.9 |

Table 1 Calculation of AP (Performances with bold font present best results)

Precision tells how many selected items are relevant and for many class we take MAP that is mean average precision. Above table 1 will give comparison of precision and average of 20 class images taken from Pascal Voc Dataset. Also proposed method compared with different methods and bold font showing best results.
5.2. Recall
In simple words recall defined as how many relevant items are selected Recall formula for a given class in classification is given by

\[
\text{Recall} = \frac{\text{True positive}}{\text{True positive} + \text{False negative}}
\]

Region attended by agent can be consider as object proposal candidate for evaluating these we follow approach given by B. Schiele et al. [22]. We process 20 category and running 200 steps per category it means 4000 region will be evaluated and get 78 percent recall. Evaluation of recall is done by Q values predicted by agent to score attended region, when agent initiate stop we add constant to score those region so they can clearly get position in top for clearly identify region. Figure 6 will show the plot between recall vs. number of proposals also we compared proposed method with other discussed earlier. With ten proposals we having more than 40 percent recall while other methods reached only 20-30 percentage.

![Figure 6 Plot Recall @threshold IoU 0.50](image)

For better illustration we plot distribution among correctly identified object and number of steps required to identify object correctly. Proposed method identify more than 85 percent object in less than 40 steps and average of 18.

5.3. Image Evaluation time
Image evaluation time refers to complete time required to process one image for identifying object. Single Image Average Evaluation Time (SIAET) [23] and storage used by Carlos Augusto et al. to show the time taken to execute an image. Following table will compare the different method with proposed method for image evaluation time and storage requirement.

| Method                        | Storage Required | IET          |
|-------------------------------|------------------|--------------|
| R-CNN                         | Data Dependent   | 47302ms      |
| FAST R-CNN                    | 477 MB           | 2185ms       |
| FASTER R-CNN                  | 523 MB           | 385ms        |
| FLODNET + SLIDING WINDOW      | 105 MB           | 6304ms       |
| FLODNET+CONVOLUTION FEATURE MAP | 95 MB         | 113ms        |
| ODDRL-NET                     | 96 MB            | 110ms        |

6. Conclusion and Future work
Thus this work present a novel object identification model based on faster RCNN concept and Reinforcement Learning. We used the action history vector to calculate reward and used Q learning policy for placing the box at correct location. Reinforcement learning is used in literature for object localization but proposed work not only localize object but defines class information as well. Precision
and recall gives the quality of identification very high with this scheme. All results implemented in python which is the language of machine learning now days. Overall method gives better results and time efficiency for identification.

References

[1] Deselaers T et al what is an object Proceedings IEEE Conference Computer Vision and Pattern Recognition (CVPR), Jun 2010, pp 73-80.
[2] Lapte I Improvements of object detection using boosted histograms in Proceedings BMVC, vol. 3, 2006.
[3] Dollár P et al Edge boxes, Locating object proposals from edges in Proceedings European Conference Computer Visi 391-405.
[4] Hoffman S C and Thiagarajan D Continuity report Revisiting grocery recognition using tensor flow publication vision.cornell.edu.
[5] Krizhevsky A and Hinton G E Image Net classification with deep convolutional neural networks in Proceedings Adv. Neural Infor. Proce. System, 2012, 1097-1105.
[6] Darrell T and Malik J et al Rich feature hierarchies for accurate object detection and semantic segmentation in Proc. IEEE Confer. Comp Vi Patt Reco Jun. 2014, 580-587.
[7] Girshick R et al Faster R-CNN Towards real-time object detection with region proposal networks in Proc. Advance Neural Info. Proce. System 2015 pp. 91-99.
[8] Schiele B and Hosang J et al How good are detection proposals really arxiv 1406-6962, 2014.
[9] Aleksis P and Christian Deep Learning of Region Proposal Networks for Object Detection IEEE conferen. CVF 2018.
[10] Hofmann T et al Beyond sliding windows objects localization by efficient sub window search CVPR IEEE, 2008.
[11] Sutton R S Introduction to reinforcement learning volume – 135 books on Reinforcement Learning.
[12] Silver D and Wierstra D et al Playing Atari with deep reinforcement learning arxiv pre print 1312-5602.
[13] Lever G et al Deterministic policy gradient algorithms Intern. Confere. On Machine Learning ICML., 2014.
[14] Guez A and VanHasselt H Deep reinforcement learning with double q-learning CORR, 1509.06461,2015.
[15] Silver D et al Playing Atari with deep reinforcement learning arxiv preprint arxiv-1312-5602, 2013.
[16] Itti L et al. Computational modelling of visual attention Nature reviews Neu. Scie. 194-203, April 2001.
[17] Donahue J, Girshick R et al Rich feature hierarchies for accurate object detection and semantic segmentation CVPR 2014.
[18] Abadi M, Paul B et al. Tensor Flow Large-scale machine learning on heterogeneous distributed system arxiv.org/abs/1603.04467.
[19] Silver D, Veness J et al. Human level control through deep reinforcement learning Nature 518-529-533, Feb 2015.
[20] Hinton G, Salakhutdinov R et al Dropout A simple way to prevent neural networks from over fitting The Journal of Machine Learning (ML) Research, 15-1929-1958, Jun 2014.
[21] Hosang J et al How good are detection proposals really arxiv 1406-6962, 2014.
[22] Augusto Carlos et al. Fast and Lightweight Object Detection Network Detection and Recognition on Resource Constrained Devices IEEE 2169-3536 2018.