ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 3, March 2020

Action Detection Using Deep Learning and Performing Using Soft Hand

Mr. Abdulrashid Kamal Alaskar[1], Prof. Sachin. D. Ruikar[2], Dr.Kamal Alaskar[3]

PG Student, Department of Electronics Engineering Walchand College of Engineering Sangli, Maharahtra, India.[1]

Associate Professor, Department of Electronics Engineering Walchand College of Engineering Sangli, Maharashtra, India.[2]

Professor, Department of Computer Applications Bharati Vidyapeeth (Deemed to be University)Institute of Management Kadamwadi, Kolhapur, Maharashtra, India.[3]

ABSTRACT: There are lots of ways used to pick or grasp object through robotic hand but there are some hardly worked done with help of Deep Learning approaches. To solve this issue, a solution is proposed which involves human strategies of picking up an object using Neural Network classifier. Classifier uses help of object detection model to detect object in environment and classifier classifies into picking strategy as per objects shape and orientation. Strategy detected by classifier is can be used by soft hand as anticipatory action and reactive grasp. To increase accuracy number of primitive taken into consideration our bounded and some of limitation are taken into mind while proposing architecture.

KEYWORDS: Object Detection model, grasping strategies ,Neural Network.

I. INTRODUCTION

Soft hand as proven to be more efficient when used in supervision of human[1][2]. But such approach is still lag in performance hence Data Driven approach can be used to improve performance (see [3]).For detection of object on which grasping is done YOLO is embedded (You Look Once) technology YOLOv3 unlike used in [5] yolov2. yolov3 has shown good accuracy with respect to SSD. The code is online at https://pjreddie.com/yolo/. [4]. Yolov3 replaces soft max function with binary classification which helps in reducing complexity and output is same as multiclass classification only[6].

Machine Learning approach had proved positive results in detecting of object or say grasping details of object [7][8][9]. In [10]Neural network helps in predicting unseen object and action to be perform on that object with help of learning from pre labeled data. In [11] convolution neural network had played good role in detecting objects strategy output of whose network can be used to determine controlling robotic soft hand. Using such network objects are trained on 45 objects and tested with 10 objects which gives approx. accuracy of 84% [3].Especially focus on updating object detection technique used in [3] from YOLO9000 to YOLOv3 which are faster than as used in [3]Below Block Diagram shows overall system of project except change/updating in object detection technology to YOLOv3 [3]

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 3, March 2020

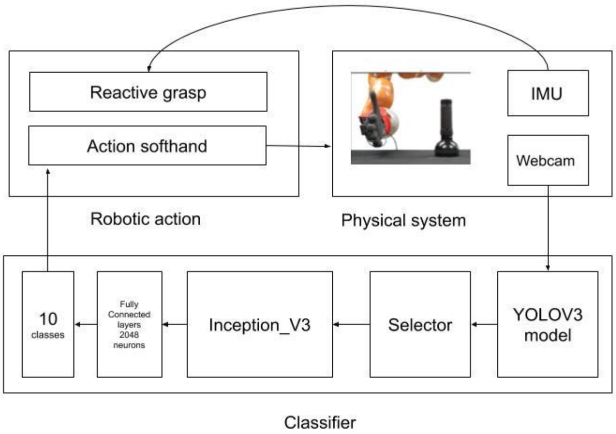


Fig I architechure for primitveselection

II. LITERATURE SURVEY

As a main reference to project [3] whose system is used for further increment in approach used in Data Driven approach to control Anthropomorphic soft hand Which uses Deep learning technology significance for increasing performance of control strategy of soft hand. It uses inceptionV3 module for grasping object and concluding strategy to be used by [11] reactive strategies have become significant in human robot exchanging of objects. It is tested over 10 objects by training on 45 objects which corresponds to accuracy of 84% on test objects. Where it is decided to update object detection method input given to model of inceptionv3. Paper achieves this goal by: i) Using Deep Neural Network model inceptionv3 model predicts or deicides which action human would take to pick certain type of objects , ii) Understanding through which action would be performed by human using robotic hand containing soft hand capability iii) Testing over on 10 objects which are not used in model training model would be able to predict action to perform on object by soft hand or robotic arm with accuracy of prediction 84%.

A. DIFFERENT MODELS

For Deep Learning InceptionV3 module will be used which is pretrained model for object grasping and third version of inception module (see [12]) . Module is trained over ImageNet Dataset .InceprionV3 model with 144 crops gained top-5 error rate is 4.2%, which pullback PReLU-Net and Inception-v2 which were used in 2015. With 42 layers deep, the parameter complexity increases by only 2.5% google Net [12].It consists of 313 layers of neuron where some of layers will be retrained to get our performance output. PyTorch versionInception-v3:[14] https://github.com/pytorch/vision/blob/master/torchvision/models/inception.py

B. OBJECT DETECTION MODELS

Yolov3 are proved to be faster than YOLOv2 used in [3] as said in[4]It's very fine on the old detection metric of .5 IOU.Yolov3 uses multiclass classification method for object detection. It uses boxes which are predefined, and model are trained on it with lots of images label with boxes. While Detecting object it uses 10,000 approx. boxes to predict one of which having more significance level and use that box as detected object box and label that box depending on pretrained images.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

	backbone	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
Two-stage methods							
Faster R-CNN+++ [5]	ResNet-101-C4	34.9	55.7	37.4	15.6	38.7	50.9
Faster R-CNN w FPN [8]	ResNet-101-FPN	36.2	59.1	39.0	18.2	39.0	48.2
Faster R-CNN by G-RMI [6]	Inception-ResNet-v2 [21]	34.7	55.5	36.7	13.5	38.1	52.0
Faster R-CNN w TDM [20]	Inception-ResNet-v2-TDM	36.8	57.7	39.2	16.2	39.8	**52.1**
One-stage methods							
YOLOv2 [15]	DarkNet-19 [15]	21.6	44.0	19.2	5.0	22.4	35.5
SSD513 [11, 3]	ResNet-101-SSD	31.2	50.4	33.3	10.2	34.5	49.8
DSSD513 [3]	ResNet-101-DSSD	33.2	53.3	35.2	13.0	35.4	51.1
RetinaNet [9]	ResNet-101-FPN	39.1	59.1	42.3	21.8	42.7	50.2
RetinaNet [9]	ResNeXt-101-FPN	**40.8**	**61.1**	**44.1**	**24.1**	**44.2**	51.2
YOLOv3 608 × 608	Darknet-53	33.0	57.9	34.4	18.3	35.4	41.9

Table I Object Detection score

III. METHODOLOGY

working of block in detail for figure shown above (Figure) .

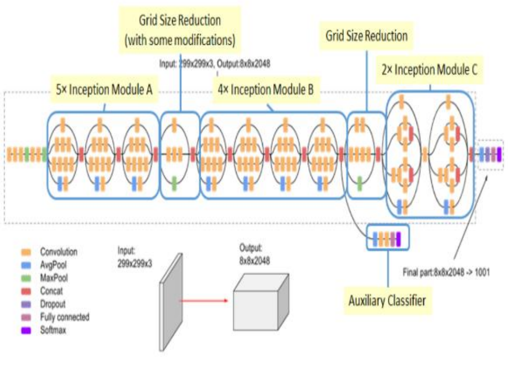
- Training Phase
- Test Phase

A. TRAINING PHASE

To reduce training time, pretrained models is used to get more accuracy in less time
YOLOv3 pre trained model weights and model to detect object which is been trained on COCO datasets detects 80 classes if require to train can train using link to code (https://github.com/qqwweee/keras- yolo3).Training InceptionV3module which is best for classification which can perform better than human vision also.

- MODEL ARCHITECTURE

Instead of creating whole model from scratch pretrained model InceptionV3 is used and use its weights and retrain some of its layers. Figure shows detail description of v3 module. To train model Dataset is required. Pop last layer of inceptionv3 module and add two layers each of 2048 neuron and SoftMax layer of 6 classes indication probability of each object which strategy to be used more as shown in figure



Inception-v3 Architecture (Batch Norm and ReLU are used after Conv)

Figure II Inceptionv3 model

- DATASET

For creation of dataset as suggested in paper [3] objects are trained on 45 objects that are different mostly different then used in [3] . Objects are kept on table at center position different people are asked to pick up object and keep it onside and record this video and each video is labeled as per primitive taken by human in that video and identified six actions namely unlike of [4] to increase accuracy of model as top left and right pinches are can be replaced

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 3, March 2020

Top: In top grasp approach hand is approached from top with palm parallel to object on top of it this approach is used to pick object which are big in size like box used in below example
Bottom: In this primitive hand are approach from right side and picked object with 4 fingers and thumb in opposite of 4 fingers providing negative force too hold object e.g. bowl
Pinch: This type of approach is like top approach only but have some difference which are commonly used for small objects like match stick
Slide: In this approach object which are thin in size like cd are slidden to corner of table and flipped and picked up
Flip: This is primitive use for small object with thin in size like coin where slide primitive can't be used here and just flip object perpendicular to table.
Lateral: In this primitive hand approaches towards objects from right side with palm perpendicular to table and parallel to object like picking up bottle
This are 45 objects that are used here
Note: Objects which are used are also orientation dependent.

Figure III Training Data images

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 3, March 2020

- TRANSFER LEARNING

Transfer Learning approach is where used weights of pre trained model to predict our aim can add some of more layers at the end of model that are fully connected or expand model and train concatenate part. Can also retrain some of the layers of models.
Adam as our optimizer.

- STEPS IN TRAINING

Initially it is set are some of the parameters of our build model taking from [3] paper outcomes
BATCH_SIZE =10
EPOCHS = 10
Regularizer = 0.01
Pdrop = not used in our case

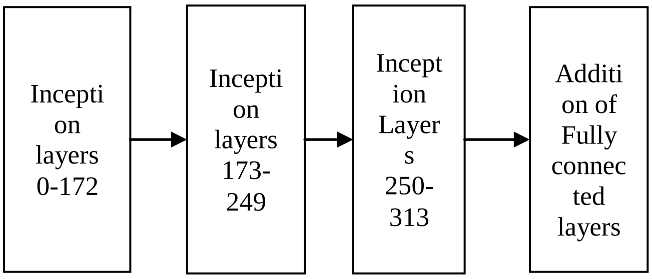


Fig IV inception block diagram

Learning rate = Used Adam doesn't require learning rate.
Learning rate for fine tuning = Used Adam doesn't require learning rate.
Optimizer = Adam.
Activation function for last FFC Layers = relu.
Prediction layer activation function = softmax.
Next freeze all layers in model except the last three layers in model as shown in figure and train last three layers with learning rate initialized in step1.
After training of last layers freeze all layers in model except unfreezing middle layers from 173-249 This layers are used for fine tuning to get inner attributes.
Using less amount of time using Kaggle platform to run our network on GPU which is freely provided to us by kaggle of 12gb ram of NVIDIA graphics.

A. TEST PHASE

Once our model is train can now integrate it in our system to test new objects.

B. FLOWCHART

For implementation two flow charts are suggested one for training and other for real time implementation. Dataset is created using Videos generated by human strategy for picking up objects Instead of generating videos we have used objects image directly images. Model is created using pre trained model of inceptionV3 popping last layer and adding 2 Fully connected layers and SoftMax layer for multi class classification.Dataset is resized to input require size of model i.e 416*416 and then split into 80% as training data and 20% as validation data.
Freeze all layers except added layers and compile model with batch size of 10 and learning rate of 0.001 Adam as optimizer and train model with 30 epochs on gpu.Freeze all layers except layers from 72-249 layer and compile model with the same parameters as mentioned above except learning rate of 0.00001 and train model.
Model is trained and ready for prediction.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 3, March 2020

IV. IMPLEMENTATION FLOWCHART

Input is taken from Webcam and use OpenCV library to process on frames got from webcam and capture video using webcam is setup the size of frames as per YOLOv3 model requirement and send frames to Object Detection model.YOLOv3 method is selected for detecting of object frames of webcam are given to YOLOv3 whereby it randomly generates boxes in frames and each frame is passed through Darknet 53.Model which has convolution layers which gives 3d tensors giving parameter for detected object box. After completion of object detection object is selected which is bounded by boxes present near to center of frames Image is selected and resized to required size of trained model i.e. 416*416Model which train is trained previously is used here to predict one of 6 primitives defined earlier. Class is selected which gives more probability below graph shows confusion matrix for trained model.

Anticipatory action is divided into two phase

- Approach Phase
- Grasp Phase

Approach phase can be studied from [3] and further robotic action is not performed

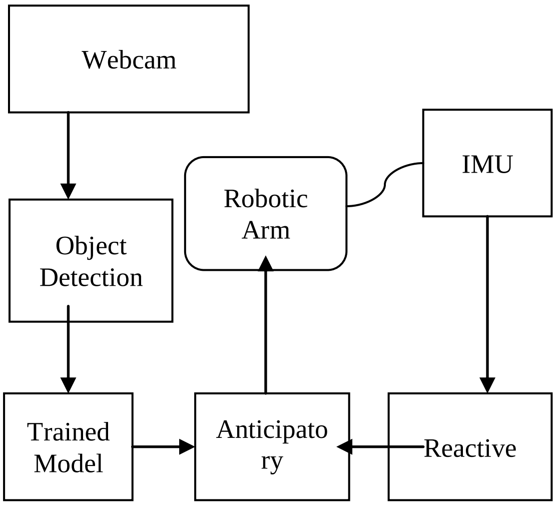


Figure VI implementation Flowchart

C. RESULTS

- Model is trained on 45 objects which gives high accuracy to 97% and validation accuracy to 93% as shown above
- Model is Tested our trained model on 10 objects which gives results of overall average of 84% which works very good with lateral top and top pinch but slight less accuracy as shown in confusion matrix with slide and flip and works worst with bottom primitive giving accuracy of 46%
- YOLOv3 which gave 51ms of response which is more than YOLOv2 and show good performance on gpu

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 3, March 2020

```
Train for 584 steps, validate for 156 steps
Epoch 1/5
584/584 [==============================] - 144s 247ms/step - loss: 2.8217
- accuracy: 0.5205 - val_loss: 2.6481 - val_accuracy: 0.3359
Epoch 2/5
584/584 [==============================] - 124s 212ms/step - loss: 0.6282
- accuracy: 0.8021 - val_loss: 2.2706 - val_accuracy: 0.6205
Epoch 3/5
584/584 [==============================] - 124s 213ms/step - loss: 0.3554
- accuracy: 0.9332 - val_loss: 0.1948 - val_accuracy: 1.0000
Epoch 4/5
584/584 [==============================] - 124s 212ms/step - loss: 0.1832
- accuracy: 0.9800 - val_loss: 0.2815 - val_accuracy: 0.9109
Epoch 5/5
584/584 [==============================] - 124s 213ms/step - loss: 0.1783
- accuracy: 0.9783 - val_loss: 0.1976 - val_accuracy: 0.9333
```

Figure VII Training Output

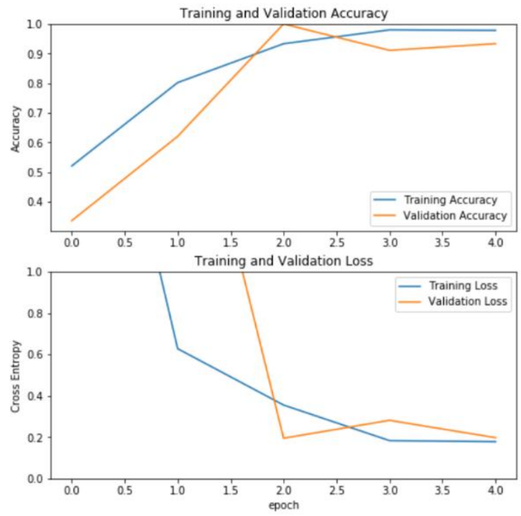


Figure VIII Training Graph

```
Confusion Matrix
[[ 9  0  0  0  0  0]
 [ 0  4  0  3  4  0]
 [ 0  0  9  0  0  0]
 [ 0  0  0 12  0  0]
 [ 0  0  3  0  6  0]
 [ 0  0  0  0  3  6]]
Classification Report
              precision    recall  f1-score   support

     lateral       1.00      1.00      1.00         9
         top       1.00      0.36      0.53        11
        flip       0.75      1.00      0.86         9
       slide       0.80      1.00      0.89        12
      bottom       0.46      0.67      0.55         9
   top_pinch       1.00      0.67      0.80         9

    accuracy                           0.78        59
   macro avg       0.84      0.78      0.77        59
weighted avg       0.84      0.78      0.77        59
```

Figure IX Predicted Data

V. CONCLUSION

Our work includes use of YOLOv3 instead of YOLOv2 which detects objects more accurately. Proposed and validated a Primitive detectionusing deep learning and performing using soft hand Goal can be gained by:
i) Creating new model with using transfer learning methodology and using new Adam optimizer ii) Creating new primitive to increase accuracy of model to predict actions and reducing primitive used in previous work, iii) Using Neural Network methodology in robotic system to make it more intelligent iv) Testing on wide range of object that are not used in training data.Further in next approach or future work stereo camera will be used to get more accuracy

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 3, March 2020

REFERENCES

1. C. Erdogan, A. Schroder, and O. Brock, "Coordination of intrinsic and extrinsic degrees of freedom in soft robotic grasping," in 2018 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2018, pp. 1–6.
2. M. Haas, W. Friedl, G. Stillfried, and H. Hoppner, "Human-robotic variable-stiffness grasps of small-fruit containers are successful even under severely impaired sensory feedback," Frontiers in neurorobotics, vol. 12, p. 70, 2018.
3. J. Zhou *et al*., "A Soft-Robotic Approach to Anthropomorphic Robotic Hand Dexterity," in *IEEE Access*, vol. 7, pp. 101483-101495, 2019.
4. YOLOv3: An Incremental Improvement Joseph Redmon, Ali Farhadi University of Washington
5. J. Redmon and A. Farhadi, "Yolo9000: better, faster, stronger," arXiv preprint, 2017.
6. https://medium.com/@jonathan_hui/real-time-object-detection-with-yolo-yolov2-28b1b93e2088
7. X. Yan, J. Hsu, M. Khansari, Y. Bai, A. Pathak, A. Gupta, J. Davidson, and H. Lee, "Learning 6-dof grasping interaction via deep geometryaware 3d representations," in 2018 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2018, pp. 1–9.
8. P. Schmidt, N. Vahrenkamp, M. W¨achter, and T. Asfour, "Grasping of unknown objects using deep convolutional neural networks based on depth images," in 2018 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2018, pp. 6831–6838.
9. A. Zeng, S. Song, K.-T. Yu, E. Donlon, F. R. Hogan, M. Bauza, D. Ma, O. Taylor, M. Liu, E. Romo et al., "Robotic pick-and-place of novel objects in clutter with multi-affordance grasping and cross-domain image matching," in 2018 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2018, pp. 1–8.
10. T. Nishimura, K. Mizushima, Y. Suzuki, T. Tsuji, and T. Watanabe, "Thin plate manipulation by an under-actuated robotic soft gripper utilizing the environment," in Intelligent Robots and Systems (IROS), 2017 IEEE/RSJ International Conference on. IEEE, 2017, pp. 1236–1243.
11. M. Bianchi, G. Averta, E. Battaglia, C. Rosales, A. Tondo, M. Poggiani, G. Santaera, S. Ciotti, M. G. Catalano, and A. Bicchi, "Tactilebased grasp primitives for soft hands: Applications to human-to-robothandover tasks and beyond," in Robotics and Automation (ICRA), 2018 IEEE International Conference on. IEEE, 2019.
12. [2015 CVPR] [GoogLeNet / Inception-v1] Going Deeper with Convolutions
13. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 2818–2826.
14. https://medium.com/@sh.tsang/review-inception-v3-1st-runner-up-image-classification-in-ilsvrc-2015-17915421f77c