Animal Motion Detection Based on Mask-RCNN and COCOdataset

Keita Saito Anh Hoang

Dept.Computer Science Dept.Computer Science

Troy University Troy University

Troy,AL Troy,AL

ksaito@troy.edu ahoang@troy.edu

*Abstract*— In this paper, we introduce, simple and flexible framework for animal motion detection. Our approach efficiently detects objects in movie files. Also, we can generate a segmentation rectangle (mask). This method is called as Mask-RCNN. The advantage of this method is that mask-RCNN in COCO dataset is much faster than other motion detection method. Also, this method is easy to train and adds small overhead. Besides, Mask-RCNN is suitable for generalization. Mask-RCNN’s performance is better than any previous model. In this paper, we overview the fundamentals of mask-RCNN methods and review their successes in animal detection including (1)what is mask-RCNN, (2)the comparison among mask-RCNN and other methods, (3)our own method in animal detection. We conclude by discussing research issues and suggesting future directions for further improvement.

Keywords—Animal Motion Detection, Convolutional Neural Network, Deep Learning, COCOdataset

# Ⅰ.Introduction

Recently, the computer vision community has improved motion detection. Especially, these improvement was brought by the state of art deep learning concept such as the Fast R-CNN and Fully Convolutional Network framework for motion detection. These method’s advantages are robustness and flexibility. However, instance segmentation is still challenging parts for us because it needs the correct detection of all objects in an movie. The classifying individual objects and localizing each using a rectangle box are needed. In this point, we would like to show a simple and fast system using Mask-RCNN method.

Mask-RCNN method developed the method of Fast-RCNN by adding some branch for extending segmentation on each Region of Interest (RoI). Mask is a small matrix used for sharpening and blurring image. It is also called as a kernel, a filter, a convolution matrix. The small has a small FCN applied to each RoI. It predicts a segmentation mask in a pixel-to-pixel form. Mask RCNN is easy to implement and train given the Faster RCNN framework, which promotes a flexible architecture designs. Furthermore, the mask only needs a small overhead. Generally speaking, Mask RCNN is improved version of Faster RCNN, and the mask can produce better result than faster RCNN. Besides, Faster RCNN was not designed for pixel-to-pixel addition between network inputs and outputs. To improve this problem, He.et.al[1] proposed that a simple, quantization-free layer, named RoIAlign, that preserves the exact spatial locations. Actually, RoIAlign has a significant impact. It improved mask accuracy more than 10%. Also, it is necessary to decouple mask and class prediction.

As the dataset, we used the COCO dataset. COCO is a kind of a large-scale object detection dataset provided by Microsoft. In the environment on COCO dataset, this Mask-RCNN method is better than all the latest method’s result in terms of instance segmentation task

This review is organized as follows. Firstly, we explain the latest computational theories of neural networks and deep learning models and how they detect the motion from dataset. Secondly, we make comparison among the latest method and previous methods. Finally, we conclude by research trends and suggesting directions for further improvements.

# Ⅱ.Related work

1.R-CNN: The Region-based CNN is to attend to a mana

work with a single-layer architecture composed of an input layer and output layer. A perceptron is regarded as a linear model, so prohibit its application in tasks involving complicated data pattern. To overcome this limitation, a hidden layer was introduced between the input layer and the output layer. Keep in mind that in neural networks the units of the neighboring layers are fully connected, but there are no connections among units in the same layer.

# Ⅲ. Mask-RCNN

Mask-RCNN is a deep neural network aimed to solve instance segmentation problem in machine learning or computer vision. It can separate different object in an image or a video. If we give it an image, it gives us the object bounding boxes, classes and masks. There are two stages of mask-RCNN. The first stage is that it generates proposals about the regions where there might be an object based on the input image. The Second stage is that it predicts the class of the object, refines the bounding box and generates a mask in pixel level of the object based on the first stage proposal. Keep in mind that both stages are connected to the backbone structure. Backbone structure is kind of a FPN(Feature Pyramid Networks) style deep neural network. It consists of a bottom-up pathway, a top-bottom pathway and lateral connections. Bottom-up pathway can be any convolutional network which extracts features from raw images. Top-bottom pathway generates feature pyramid map which is similar in size to bottom-up pathway. Lateral connections are convolution and adding operations between two corresponding levels of the two path ways. FPN outperforms other single convolutional networks because it maintains strong semantically features at various-resolution scales.

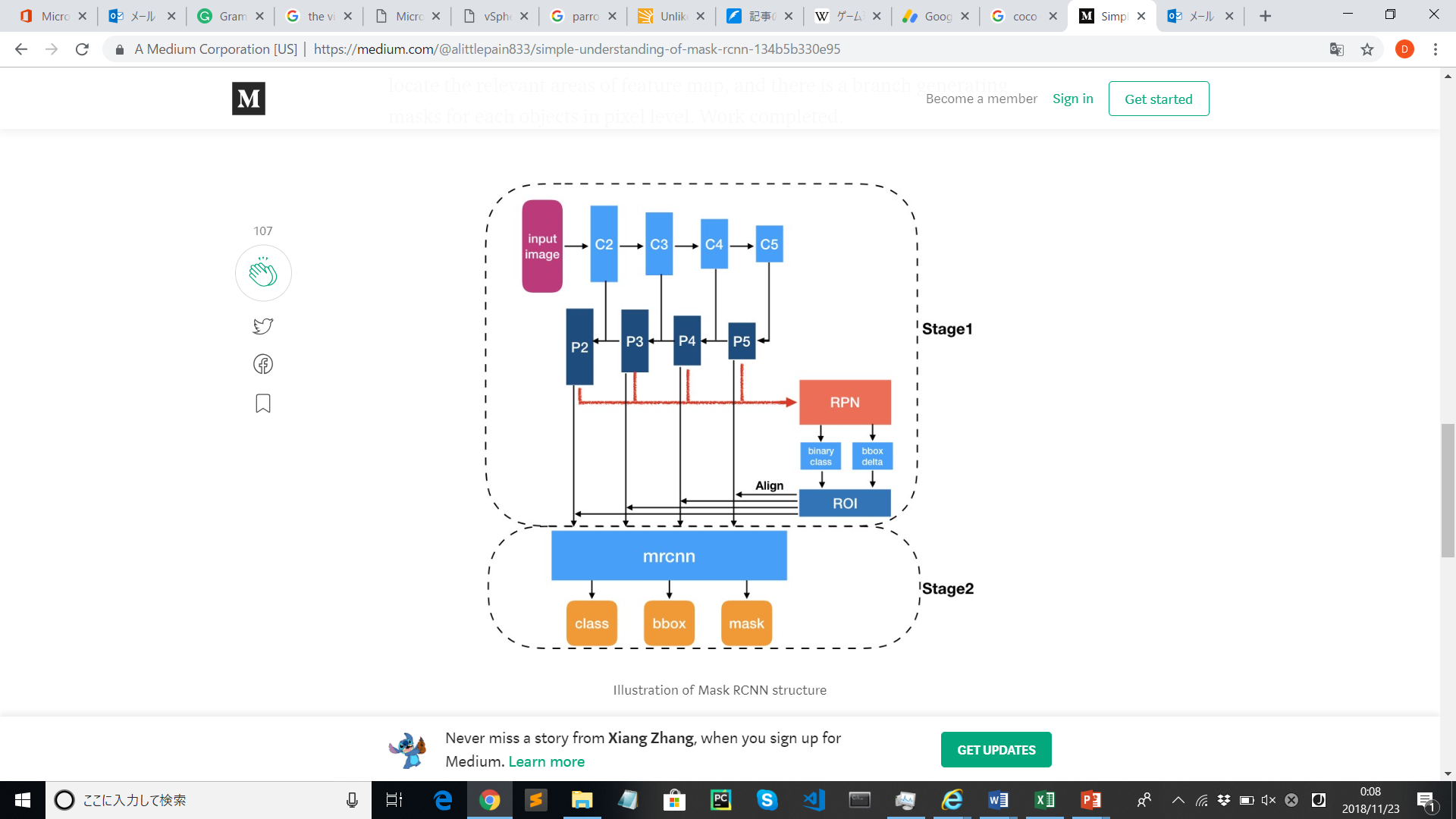


Table2.the illustration of Mask-RCNN architecture

Next, we introduce the Mask-RCNN basic architecture. In stage1, a light weight neural network called RPN scans all FPN top-bottom pathway and proposes regions, which may contain objects. In stage2, another neural network takes proposed regions by the first stage and assign them to several specific areas of feature map level, scans these areas, and generates objects classes (multi-categorial classified), bounding boxes and masks. And then use ROIallign to locate the relevant areas of feature map, and there is a branch generating masks for each objects in pixel level.

Actually, the concept of Mask-RCNN is simple. Not only the Faster-RCNN’s concept: two outputs for each candidate object, a class label and a bounding-box offset, but also outputs of the object mask is added. By doing so, Mask-RCNN is natural. However, the additional mask is different from the class and box outputs; therefore it needs extraction of much finer spatial layout of an object. Mask-RCNN also outputs a binary mask for each RoI. This is the big difference other systems which the classification depends on mask prediction. A mask have an input of spatial layout. Although, class labels and box offsets are into short output vectors by fully connected layers, mask is addressed by the pixel-to-pixedl correspondence supported by convolutions. This method’s fully convolutional representation needs fewer parameters and is more accurate than other previous methods.

Also, this pixel-to-pixel behavior needs RoI features, which are small feature maps to be aligned to faithfully preserve the explicit per-pixel spatial correspondence. This inspires us to develop the following RoIAlign layer that plays a important role in mask prediction. In terms of RoIAlign, RoIPool is a standard operation for extracting a small feature map from each RoI.

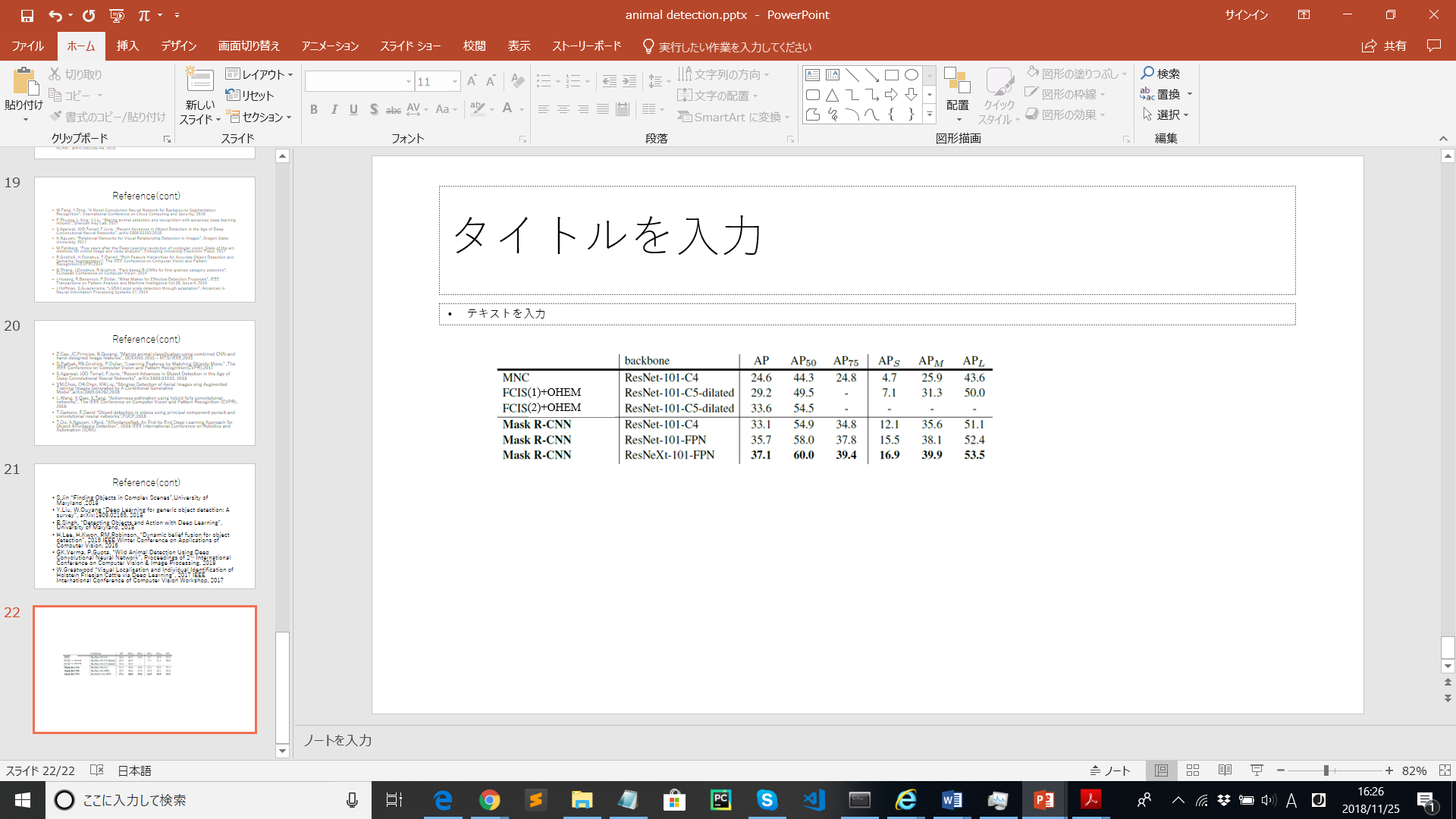


Table1. The comparison among Mask-RCNN and other methods. MNC means the winner method on 2015 COCO image segmentation challenges, FCIS means the winner method on 2016 COCO image segmentation challenge. FCIS(2) means the improved version of FCIS(1). It includes multi-scale train/test and horizontal flip test, and OHEM.

# Ⅳ. Innovative parts

In order to apply previous research into our animal motion detection method, we tried the following modification:

# Ⅴ. Experiments

　According to the He’s paper[1], they performed a full comparison of Mask-RCNN to the latest method on the COCO dataset. They trained the union of 80k train images and a 35k sets of val image. As we can see above table1, the performance of Mask-RCNN outperformed than other latest methods.

In our method, we tried to train ●● datasets,

# Ⅵ.Results

We found that we

# Ⅵ.Conclusion And Future work

C

# Ⅶ. disclosure statement

The author is not belong to any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

# Ⅷ. Acknowledgment

We would like to appreciate Prof. Zhao for her continuous guidance, support ,and help throughout this class.

##### References

1. K.He, G.Gkioxari, P.Dollar, R.Girshick “Mask R-CNN”, Facehook AI Research (FAIR),2018
2. X.Wang, A.Shrivastava, A.Gupta “A-Fast-RCNN:Hard Positive Generation via Adversary for Object Detection” ,2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),2017
3. J.Hosang, R.Benenson, P.Dollar “What makes for effective detection proposals?” ,arXiv:1502.05082, 2015
4. G.Papandreou, T.zhu, N.Kanazawa “Towards accurate multi-person pose estimation in the wild”, IEEE Conference on Computer Vision and Pattern Recognition (CVPR),2015
5. RA.Guler “Densepose:Dense human pose estimation in the wild”, arXiv:1802.00434,2018
6. M.Danish “Beef Cattle Instance Segmentation Using Mask-RCNN”, Dublin Institute of Technology, 2018
7. Z.Zhang, Z.He, G.Cao, W.Cao “Animal Detection From Highly Cluttered Natural Scenes Using Spatiotemporal Object Region Proposals and Verification”, IEEE Transactions on Multimedia Vol.18, Issue:10, 2016
8. B.Cheng, Y.Wei, H.Shi “Revisiting RCNN: On Awakening the Classification Power of Faster RCNN”, arXiv:1803.06799, 2018
9. W.Fang, Y.Ding, “A Novel Convolution Neural Network for Background Segmentation Recognition”, International Conference on Cloud Computing and Security, 2018
10. P.Zhuang, L.Xing, Y.Liu, “Maring animal detection and recognition with advanced deep learning models”, Shenzen Key Lab, 2017
11. S.Agarwal, JOD.Terrail, F.Jurie, “Recent Advances in Object Detection in the Age of Deep Convolutional Neural Networks”, arXiv:1809.03193,2018
12. K.Nguyen, “Relational Networks for Visual Relationship Detection in Images”, Oregon State University, 2017
13. M.Felsberg, “Five years after the Deep Learning revolution of computer vision: State of the art methods for online image and video analysis”, Linkoping University Electronic Press, 2017
14. R.Girshick, H.Donahue, T.Darrell, “Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation”, The IEEE Conference on Computer Vision and Pattern Recognition(CVPR),2014
15. N.Zhang, J.Donahue, R.Girshick, “Part-based R-CNNs for fine-grained category detection”, European Conference on Computer Vision ,2014
16. J.Hosang, R.Benenson, P.Dollar, “What Makes for Effective Detection Proposals”, IEEE Transactions on Pattern Analysis and Machine Intelligence Vol.38, Issue:4, 2016
17. J.Hoffman, S.Guadarrama, “LSDA:Large scale detection through adaptation”, Advances in Neural Information Processing Systems 27, 2014
18. Z.Cao, JC.Principe, B.Ouyang, “Marine animal classification using combined CNN and hand-designed image features”, OCEANS 2015 – MTS/IEEE,2015
19. D.Pathak, RB.Girshick, P.Dollar, “Learning Features by Watching Objects Move.” ,The IEEE Conference on Computer Vision and Pattern Recognition(CVPR),2017
20. S.Agarwal, JOD Terrail, F.Jurie, “Recent Advances in Object Detection in the Age of Deep Convolutional Neural Networks”, arXiv:1809.03193, 2018
21. YM.Chou, CH.Chen, KH.Liu, “Stingray Detection of Aerial Images sing Augmented Training Images Generated by A Conditional Generative Model”,arXiv:1805.04262,2018
22. L.Wang, Y.Qiao, X.Tang, “Actionness estimation using hybrid fully convolutional networks”, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016
23. T.Gamero, E.David “Object detection in videos using principal component pursuit and convolutional neural networks”,PUCP,2018
24. T.Do, A.Nguyen, I.Reid, “AffordanceNet: An End-to-End Deep Learning Approach for Object Affordance Detection”, 2018 IEEE International Conference on Robotics and Automation (ICRA)
25. S.Jin “Finding Objects in Complex Scenes”,University of Maryland ,2018
26. Y.Liu, W.Ouyang “Deep Learning for generic object detection: A survey”, arXiv:1809.02165, 2018
27. B.Singh, “Detecting Objects and Action with Deep Learning”, University of Maryland, 2018
28. H.Lee, H.Kwon, RM.Robinson, “Dynamic belief fusion for object detection”, 2016 IEEE Winter Conference on Applications of Computer Vision, 2016
29. GK.Verma, P.Gupta, “Wild Animal Detection Using Deep Convolutional Neural Network”, Proceedings of 2nd International Conference on Computer Vision & Image Processing, 2018
30. W.Greatwood “Visual Localisation and Individual Identification of Holstein Friesian Cattle via Deep Learning”, 2017 IEEE International Conference of Computer Vision Workshop, 2017