Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 7-Issue 1 January 2024
Paper: 24

Exploring Faster R-CNN Algorithms for Object Detection

1st ABHILASH JAIN
Department of Computer Science and Engineering Maulana Azad National Institute of Technology Bhopal, India
abhilashjain1919@gmail.com

2nd Dr. R. K. PATERIYA
Department of Computer Science and Engineering Maulana Azad National Institute of Technology Bhopal, India
pateriyark@gmail.com

***Abstract*—Object detection is an paramount task in computer vision that has applications in various fields, including robotics, autonomous vehicles, and augmented reality. In recent years, significant advancements have been made in computer vision techniques, leading to faster and more accurate object detection models. One such model is Faster R-CNN [11], which has become a popular choice among researchers due to its ability to detect objects with high accuracy while maintaining a fast detection speed. This review paper explores the latest Faster R-CNN-based models and their applications, providing a comprehensive wisdom of the existing shape of object detection in computer vision and its potential impact on various industries. Additionally, the paper discusses several other object detection ideals, including Sparse- RCNN, Cascade-RCNN, Loss-Guided Attention RCNN, Global Context Aggregation RCNN, iFS-RCNN, and HRRCNN, and their unique characteristics and advantages.**

***Index Terms*—Sparse R-CNN, Cascade R-CNN, LGA-RCNN, GCA RCNN, iFS-RCNN, DeFRCN, and HR RCNN**

I. Introduction

Computer vision is a rapidly evolving field with significant advancements in recent years. The latest developments in com- puter vision techniques have brought about faster and more accurate object detection models. With the growing need for efficient and reliable object detection, the latest Faster R-CNN- based models have become an important area of research. The origin of Faster R-CNN [9], [10] can be traced back to the need for faster and more accurate object detection. Since its inception, Faster R-CNN has become a popular choice among researchers due to its ability to detect objects with high accuracy while also maintaining a fast detection speed. Its development has undoubtedly helped move the field of computer vision forward. This review paper discusses the latest Faster R-CNN-based models and their applications. By examining these models, it hopes to provide a comprehensive wisdom of the existing shape of object detection in computer vision and the potential impact it can have on diverse industries.

Object detection is a fundamental requirement for many computer vision applications that have found applications in various fields, including robotics, autonomous vehicles, and augmented reality. Convolutional-neural-networks have shown remarkable progress in object detection in recent times. However, traditional object detection methods have several limitations that pose a challenge for object detection in complex scenarios. To address these challenges, researchers have proposed several object detection models, each with its unique characteristics and advantages. In this review paper, it discusses seven such models, namely Sparse R-CNN, Cascade R-CNN, Loss-Guided Attention RCNN (LGA-RCNN), Global Context Aggregation RCNN (GCA RCNN), ifs-RCNN, DE- FRCN, and HR RCNN. Sparse R-CNN [8] was developed to rethink the necessity of dense prior in object detection, which suffers from limitations such as repetitious results, heuristic assignment rules, and sensitivity Regarding the dimensions and shape of anchor boxes. The authors hope that their work could inspire exploring the next generation of object detec- tors. Cascade R-CNN [6] was developed to investigate The difficulty of creating precise object algorithms detectors that generate minimal false positives in close proximity. The model has several stages, each with a higher iou threshold, to improve detection accuracy. The LGA-RCNN [4] model integrates a loss-guided attention mechanism to emphasize discriminative regions of objects and improve detection performance. GCA RCNN [5] was developed to extract global features for the roi head, in which Inputs are cropped from the overall feature map in a partial manner. The model fuses global context and local features to boost and polish global context information.defrcn [1] is a model that has been proposed to overcome the shortcomings of Faster R-CNN, such as its failure to take into account few-shot scenarios and conflicts between its components, in order to enhance the performance of few-shot object identification. These tasks become challenging when there are few training examples available for new classes or When training on both base and new classes is not feasible due

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 7-Issue 1 January 2024
Paper: 24

to time constraints. To address these challenges, the ifs- RCNN [3] model was developed to learn from an infrequent training instances of new classes while not forgetting the previously learned knowledge of the base classes. On the other hand, the Hierarchical Relational framework for object detection HRRCNN [2] was developed to rectify the shortcomings of convolutional-neural-networks(convnets) in explicitly modeling as well as reasoning about contextual relationships in images. The HRRCNN model integrates three types of contextual relationships - pixel connections, scale dependencies, and object associations- in a unified model to improve object detection performance. This review paper

explores these models and their causes for development in greater detail.

The writing is organized into three leading units. The first unit provides a synopsis of different Faster R-CNN algorithms for object detection, including, DEFRCN, HR-RCNN, ifs- RCNN, Global Context Aware RCNN, LGA-RCNN, Cascade R-CNN, and Sparse R-CNN. The second unit compares the performances of these algorithms. Finally, the third section presents the study's conclusion, summarizing the findings and suggesting future research directions for improving object detection performance.

II. Methods

A. Revisiting faster rcnn

Faster R-CNN is a two-stage grouping design that em- ploys three mechanism blocks for end-to-end training and a joint convolutional backbone [15] for retrieving general- ized features, an effective Region-Proposal-Network (RPN) for producing class-agnostics proposals, and a task-specific RCNN head for carrying out class-relevant classification and localization. In further detail, the input picture is delivered to the RPN and RCNN modules in parallel after being processed into the backbone, which creates a high-level feature map. Second, RPN begins a small number of high-quality region recommendations by concurrently identifying and regressing a set of scale-varying anchors from the feature map. Next, RCNN conducts a box classifier and a regressor for enhancing the item classification probabilities and bounding box refine- ment based on the collective feature map and suggestions and RoI pooling, which combines individual region-of-interest into such a set dimension of the feature map.

B. DEFRCN: A Few-Shot Object Detection Method based on Decoupled-Faster-RCNN

Current detection frameworks, such as Faster R-CNN, often need help in scenarios with limited data and when dealing with multi-task learning [14] and shared backbone challenges. It proposes a new, efficient architecture called Decoupled Faster R-CNN (DEFRCN) to address these issues. In particular, It enhances Faster R-CNN by incorporating a Gradient-Decoupled-Layer(GDL) for multistage separation and a Prototypical-Calibration-Block(PCB) for multitask sep- aration. The model is illustrated in the accompanying Fig.

1. The Gradient-Decoupled-Layer is an innovative deep layer that alters the feature propagation and gradient computation procedures to separate it is preceding and succeeding layers. The Prototypical-Calibration-Block is a classification model based on an offline prototypes model that utilizes proposals from the detector as information and improves the calibrating classification scores using pairwise comparisons

C. HRRCNN: Novel Hierarchical-Relational-Reasoning Ap- proach

Extraction of feature pyramids is accomplished through a backbone network and generate region proposals for an image. Instead of processing features computed for each region of the

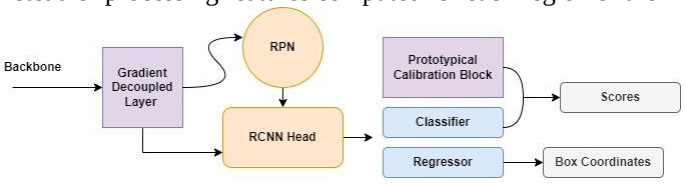

Fig. 1. The architecture of DeFRCN.

image separately in a box head, in-between the feature pyramid and box head, the HR-RCNN introduces a hierarchical relational reasoning [16] (HR) component. This component retains three relational reasoning elements: a pixel, scaled, and region-

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 7-Issue 1 January 2024
Paper: 24

of-interests graph. To make the model more efficient, It uses a novel graph-attention-module (GAM) that can gather intake from different types of heterogeneous graphs by computing attention weight based on the quantifying the semantic [17] and spatial proximities of nodes . The subsequent Fig. 2 depicts the given model. This empowers the model to yield a more sophisticated rendering of the image.

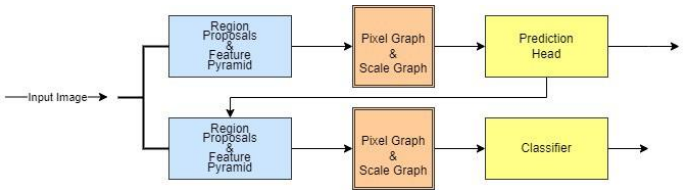

Fig. 2. The HR-RCNN structure.

D. iFS RCNN: Segmentation of Few-Shot Instances with In- cremental Learning

In order to ensure optimum efficiency on all old and fresh classes, this model presents a technique for incremental few-shot instance segmentation in which new object classes are presented when training instances of old classes are not there. The procedure constructs the Mask-RCNN framework by providing an untried class classification established on the probit operation, which addresses the lack of training examples for new classes through Bayesian learning [18]. The paper proposes a novel bounding box predictor guided by uncertainty estimation. It estimates the uncertainty in localization on new classes and uses it to refine bounding box predictions and height the loss. It feeds a new bounding-box of the calculated uncertainty together with the ROI-aligned pooled feature map. The segmentation head then receives the refined bounding box. A new loss is created to penalize errors on training examples with specific bounding-box predictions more severely by being designed to be less for extremely uncertain predictions described in Fig.3.

E. Global Context Aware RCNN for Object Detection

The proposed system is contextually conscious and facili- tates the integration of global context information with specific information from Regions of Interest (ROIs) in a two-stage object detection network. In traditional two-stage methods,

TABLE I FASTER R-CNN ALGORITHMS

S.no	Topic	Year
1	DeFRCN: (A Few-Shot Object Detection Method based on Decoupled-Faster-RCNN)	2021
2	HRRCNN: (Novel Hierarchical-Relational-Reasoning Approach)	2021
3	iFS RCNN (Segmentation of Few-Shot Instances with Incremental Learning)	2022
4	Global Context Aware RCNN for Object Detection	2020
5	LGA-RCNN: (Loss-Guided Attention for Object Detection)	2021
6	Cascade R-CNN	2017
7	Sparse R-CNN	2021

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 7-Issue 1 January 2024
Paper: 24

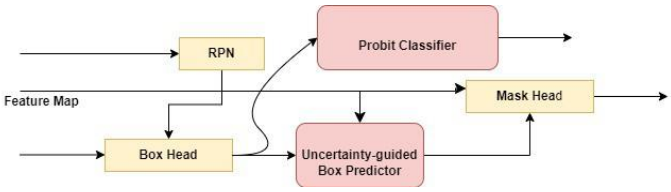

Fig. 3. The iFS-RCNN structure.

the initial stage (rpn head) separates the foreground from the background and forecasts the anchor box's regression coefficient. The second stage (roi head) forecasts the specific category of RoI and computes to enhance the precision of the bounding box. An offset value is employed. It fine-tunes the features using global statistics for the rpn head to enhance the network's global feature recognition capabilities. Nevertheless, implementing a more elaborate design for extracting global features for the cropped features of the full feature map is employed in the R0I head. Our method employs the global context information extracted from diverse levels of the fea- ture pyramid in the model by merging them through thick connections in the R0I head, generating higher-dimensional global descriptors using The module designed to incorporate contextual information, as illustrated in Figure 4. Furthermore, like FPN [19], It uses shared, fully connected several layers to extract features at different processing stages. Moreover, ultimately merge prediction information from these stages to make the final decision.

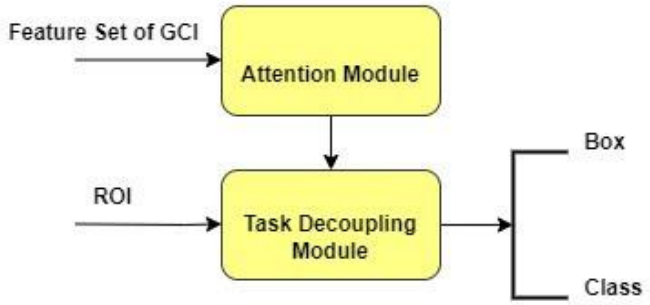

Fig. 4. Visual depiction of GCA RCNN.

F. LGA-RCNN: Loss-Guided Attention for Object Detection

LGA-RCNN is a proposed method that uses a Loss-Guided Attention (LGA) module to identify critical regions of an object, which are then fused with global information for improved classification and localization. The LGA module leverages a k Gaussian architecture to generate masks that highlight the most informative regions in the R0I feature maps. It is supervised by an additional classification loss to ensure optimal locations. LGA modules are employed to predict rep- resentative regions and highlight them to enhance classification accuracy. By combining highlighted features with the first RoI feature, an improvement in both class-ification and regression is achieved, as shown in Fig. 5. The fused RoI feature maps combine local and global information to improve detection results, and the Gaussian masks [20] also focus on the marginal regions of the object to enhance location accuracy.

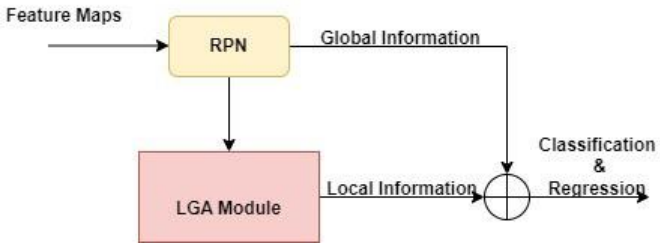

Fig. 5. Illustration of LGA R-CNN.

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 7-Issue 1 January 2024
Paper: 24

G. Cascade-RCNN: SOTA Object Detection

A novel detector architecture called Cascade R-CNN has been presented to deal with the problem of precise object detection. It is an R-CNN extension with many stages, with detector stages further down the cascade, more discriminating against nearby false positives. Resampling gradually raises the quality of hypotheses, ensuring an equal-sized positive training set for each detector and reducing overfitting. The subsequent R-CNN stage in the cascade utilizes the output of the previous stage for training. Three acute effects of this cascade learning on detector training are decreased likelihood of overfitting at high IoU thresholds [21], the optimality of deeper stage detectors for higher IoU thresholds, and an increase in the understanding efficiency of Regression of bounding box parameters are enhanced in the later stages of the model. Refer to Fig. 6 below for a depiction of the model.

H. Sparse RCNN

Sparse R-CNN is a method for detecting entities in illus- trations that strictly utilizes sparsity. In several works, the

TABLE II COMPARISON OF THE METHODS

Methods	AP	AP_{50}	AP_{75}	AP_s	AP_m	AP_l
DeFRCN	41.9	62.3	45.1	22.3	46.6	57.8
DeFRCN + GDL	36.5	57.6	39.2	19.8	41.7	50.3
HR-RCNN (R101-DCN-FPN)	47.7	68.2	51.7	30.8	50.4	59.4
HR-RCNN (R101-FPN)	44.9	65.1	48.4	26.7	47.6	56.5
GCA-RCNN Double-Head (GCA)	42.1	63.0	45.9	24.4	45.2	53.2
GCA-RCNN	40	61.6	43.5	22.8	43.2	50.3
Cascade-RCNN with R1O1	42.8	62.1	46.3	23.7	45.5	55.2
Sparse-RCNN with RX1O1	46.9	66.3	51.2	28.6	49.2	58.7
Sparse-RCNN with RX1O1 and DCN	48.9	68.3	53.4	29.9	50.9	62.4
Sparse-RCNN with RX1O1 and DCN X	51.5	71.1	57.1	34.2	53.4	64.1

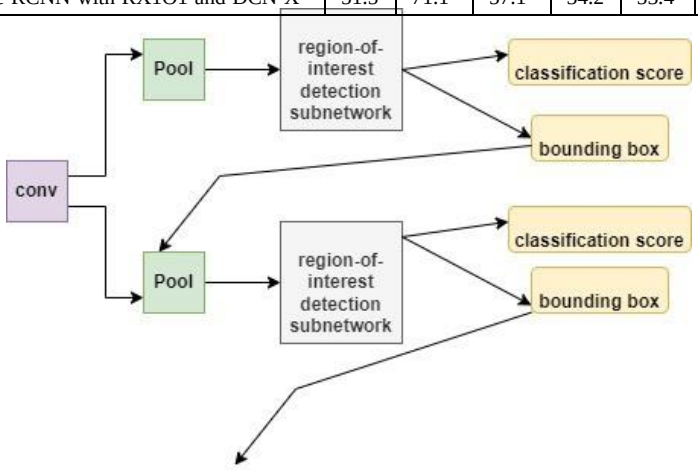

Fig. 6. A graphical representation of Cascade R-CNN.

emphasis on dense object candidates is pronounced in object detection. Each grid of the image feature map with dimensions H x W contains a predetermined set of k anchor boxes. To perform classification and location, Rather than generating dense object candidates, it employs a sparse set of learned object proposals with a fixed length of N to feed into the object recognition head for object recognition purposes. Fig. 7 below illustrates the model. In Sparse R-CNN, all efforts associated with the design and assignment of labels to object candidates are completely avoided by restricting the number of handcrafted object candidates and utilizing a fixed set of N (100) learnable propositions instead. Final predictions are immediately

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 7-Issue 1 January 2024
Paper: 24

output without unnecessary post-procedure suppressions.

Fig. 7. Sparse RCNN in a nutshell.

III. Comparison of the methods

Table 2 compares several models established on Faster R- CNN for object recognition tested on the COCO [13] dataset. The measurement metrics used are average precision [25] (AP), AP_{50}, AP_{75}, AP_s, AP_m, AP_l, representing the precision at different intersections of union (IOU) thresholds and for vari- ous object sizes. The DeFRCN model with GDL improves the consistency of Faster R-CNN and achieves better productivity than baselines. The HR-RCNN with ResNet-101 backbone and a 2x training scheme achieves an overall AP of 44.9 while using deformable ResNet-101 as the backbone increases the AP to 47.7. The GCA-RCNN with FPN baselines achieves an overall AP of 40.0, while the Double-Head RCNN [22] strategy achieves an overall AP of 42.1. A total AP of 46.9 is achieved by sparse R-CNN with ResNeXt-101(RX101) [24] in its pure form, 48.9 with DCN [22], and 51.5 with ad- ditional test-time augmentations. Using COCO-style Average Precision, Cascade R-CNN with ResNet-101(R101) [25] as its base achieves a total AP of 42.8. These models show that to achieve better detection accuracy and inference speed and satisfy the real-time requirements of video processing, deep learning-based object detection models require an improved architecture.

IV. Conclusion

A comprehensive look delves into the world of object detection and provides an in-depth examination of various models constructed using Faster R-CNN. The comparison of these models on the COCO dataset reveals that Sparse R- CNN reigns supreme with an impressive overall AP of 51.5. The article sheds light on the architecture of each model and the improvements made to overcome the challenges of speed and accuracy. As object detection is a complex problem, deep learning models must continuously strive to improve their architecture to meet the demands of real-time video processing. With advancements in technology and innovative approaches, we can expect even greater progress in the field of object detection.

References

[1] Qiao, Limeng, Yuxuan Zhao, Zhiyuan Li, Xi Qiu, Jianan Wu, and Chi Zhang. ”Defrcn: Decoupled faster r-cnn for few-shot object detection.” In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 8681-8690. 2021.

[2] Chen, Hao, and Abhinav Shrivastava. ”HR-RCNN: Hierarchical Rela- tional Reasoning for Object Detection.” arXiv preprint arXiv:2110.13892 (2021).

[3] Nguyen, Khoi, and Sinisa Todorovic. ”ifs-rcnn: An incremental few- shot instance segmenter.” In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7010-7019. 2022.

[4] Yi, Xin, Jiahao Wu, Bo Ma, Yangtong Ou, and Longyao Liu. ”LGA- RCNN: Loss-Guided Attention for Object Detection.” arXiv preprint arXiv:2104.13763 (2021).

[5] Zhang, Wenchao, Chong Fu, Haoyu Xie, Mai Zhu, Ming Tie, and Junxin Chen. ”Global context aware RCNN for object detection.” Neural Computing and Applications 33 (2021): 11627-11639.

[6] Cai, Zhaowei, and Nuno Vasconcelos. ”Cascade r-cnn: Delving into high quality object detection.” In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 6154-6162. 2018.

[7] Sun, Peize, Rufeng Zhang, Yi Jiang, Tao Kong, Chenfeng Xu, Wei Zhan, Masayoshi Tomizuka et al. ”Sparse r-cnn: End-to-end object detection with learnable proposals.” In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 14454-14463. 2021.

[8] Sun, Peize, Rufeng Zhang, Yi Jiang, Tao Kong, Chenfeng Xu, Wei Zhan, Masayoshi Tomizuka et al. ”Sparse r-cnn: End-to-end object detection with learnable proposals.” In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 14454-14463. 2021.

[9] Girshick, Ross, Jeff Donahue, Trevor Darrell, and Jitendra Malik. ”Rich feature hierarchies for accurate object detection and semantic segmentation.” In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 580-587. 2014.

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 7-Issue 1 January 2024
Paper: 24

[10] Girshick, Ross. ”Fast r-cnn.” In Proceedings of the IEEE international conference on computer vision, pp. 1440-1448. 2015.

[11] Ren, Shaoqing, Kaiming He, Ross Girshick, and Jian Sun. ”Faster r- cnn: Towards real-time object detection with region proposal networks.” Advances in neural information processing systems 28 (2015).

[12] Lin, Tsung-Yi, Michael Maire, Serge Belongie, James Hays, Pietro Per- ona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. ”Microsoft coco: Common objects in context.” In Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13, pp. 740-755. Springer International Publishing, 2014.

[13] Zhang, Yu, and Qiang Yang. ”A survey on multi-task learning.” IEEE Transactions on Knowledge and Data Engineering 34, no. 12 (2021): 5586-5609.

[14] Ajit, Arohan, Koustav Acharya, and Abhishek Samanta. ”A review of convolutional neural networks.” In 2020 international conference on emerging trends in information technology and engineering (ic-ETITE), pp. 1-5. IEEE, 2020.

[15] Santoro, Adam, David Raposo, David G. Barrett, Mateusz Malinowski, Razvan Pascanu, Peter Battaglia, and Timothy Lillicrap. ”A simple neural network module for relational reasoning.” Advances in neural information processing systems 30 (2017).

[16] Rips, Lance J., Edward J. Shoben, and Edward E. Smith. ”Semantic distance and the verification of semantic relations.” Journal of verbal learning and verbal behavior 12, no. 1 (1973): 1-20.

[17] Neal, Radford M. Bayesian learning for neural networks. Vol. 118. Springer Science Business Media, 2012.

[18] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, S. Belongie, Feature pyramid networks for object detection, in: Proceedings of the ´

[19] IEEE conference on computer vision and pattern recognition, 2017, pp. 2117–2125. Chen, Jer-Sen, Andres Huertas, and G. Medioni. ”Fast convolution with Laplacian-of-Gaussian masks.” IEEE Transactions on Pattern Analysis and Machine Intelligence 4 (1987): 584-590.

[20] Rezatofighi, Hamid, Nathan Tsoi, JunYoung Gwak, Amir Sadeghian, Ian Reid, and Silvio Savarese. ”Generalized intersection over union: A metric and a loss for bounding box regression.” In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 658-666. 2019.

[21] Jifeng Dai, Haozhi Qi, Yuwen Xiong, Yi Li, Guodong Zhang, Han Hu, and Yichen Wei. Deformable convolutional networks. In ICCV, 2017.

[22] H. Hu, J. Gu, Z. Zhang, J. Dai, Y. Wei, Relation networks for object detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 3588–3597.

[23] Xie, Saining, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. ”Aggregated residual transformations for deep neural networks.” In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1492-1500. 2017.

[24] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. ”Deep residual learning for image recognition.” In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778. 2016.

[25] Henderson, Paul, and Vittorio Ferrari. ”End-to-end training of object class detectors for mean average precision.” In Computer Vision–ACCV 2016: 13th Asian Conference on Computer Vision, Taipei, Taiwan, November 20-24, 2016, Revised Selected Papers, Part V 13, pp. 198-213. Springer International Publishing, 2017.

.