ICANN19
28th International Conference on Artificial Neural Networks

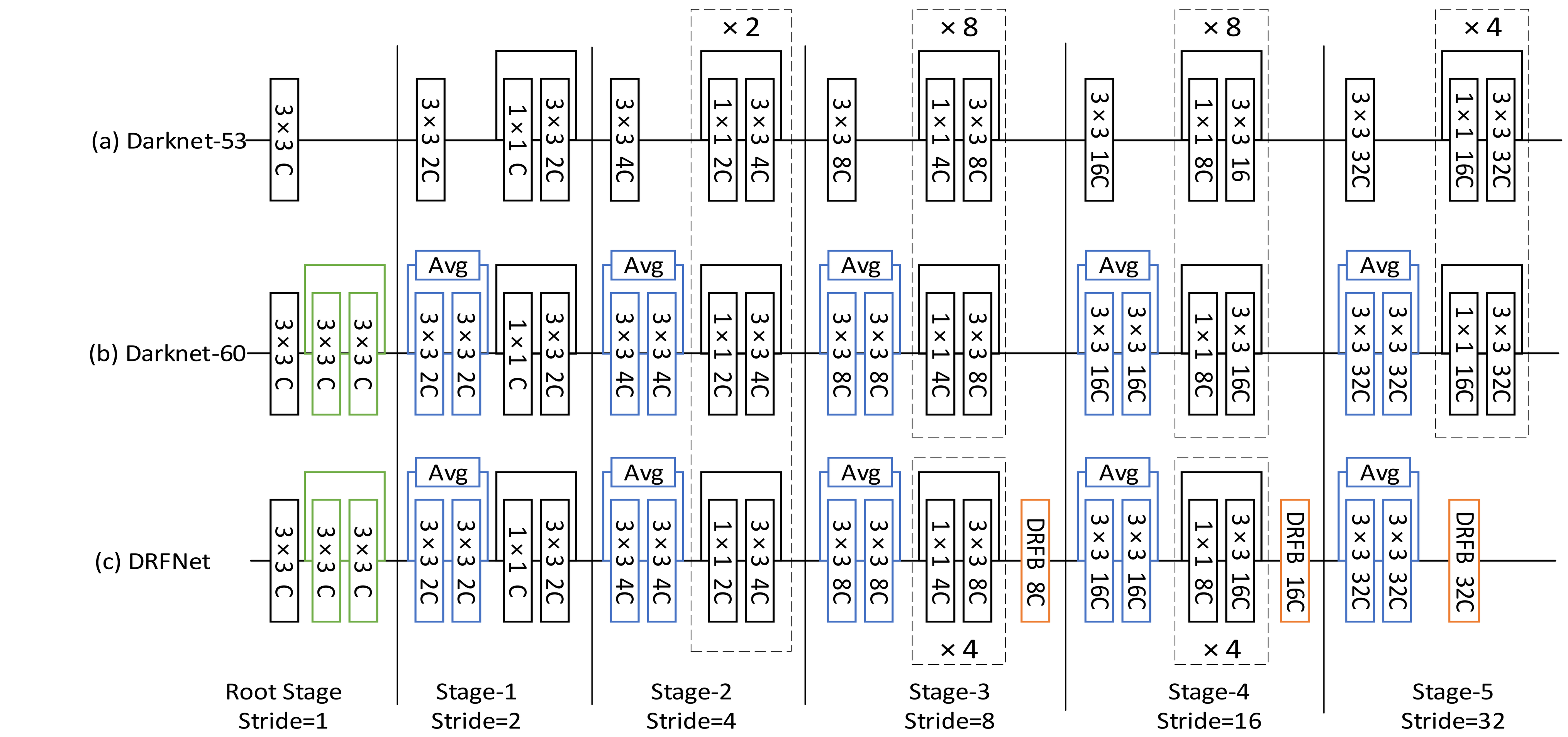
Dense Receptive Field Network: A Backbone Network for Object Detection

Fei Gao, Chengguang Yang, Yisu Ge, Shufang Lu and Qike Shao
College of Computer Science and Technology, Zhejiang University of Technology, HangZhou, China
Contact Author: Fei Gao, feig@zjut.edu.cn

Introduction

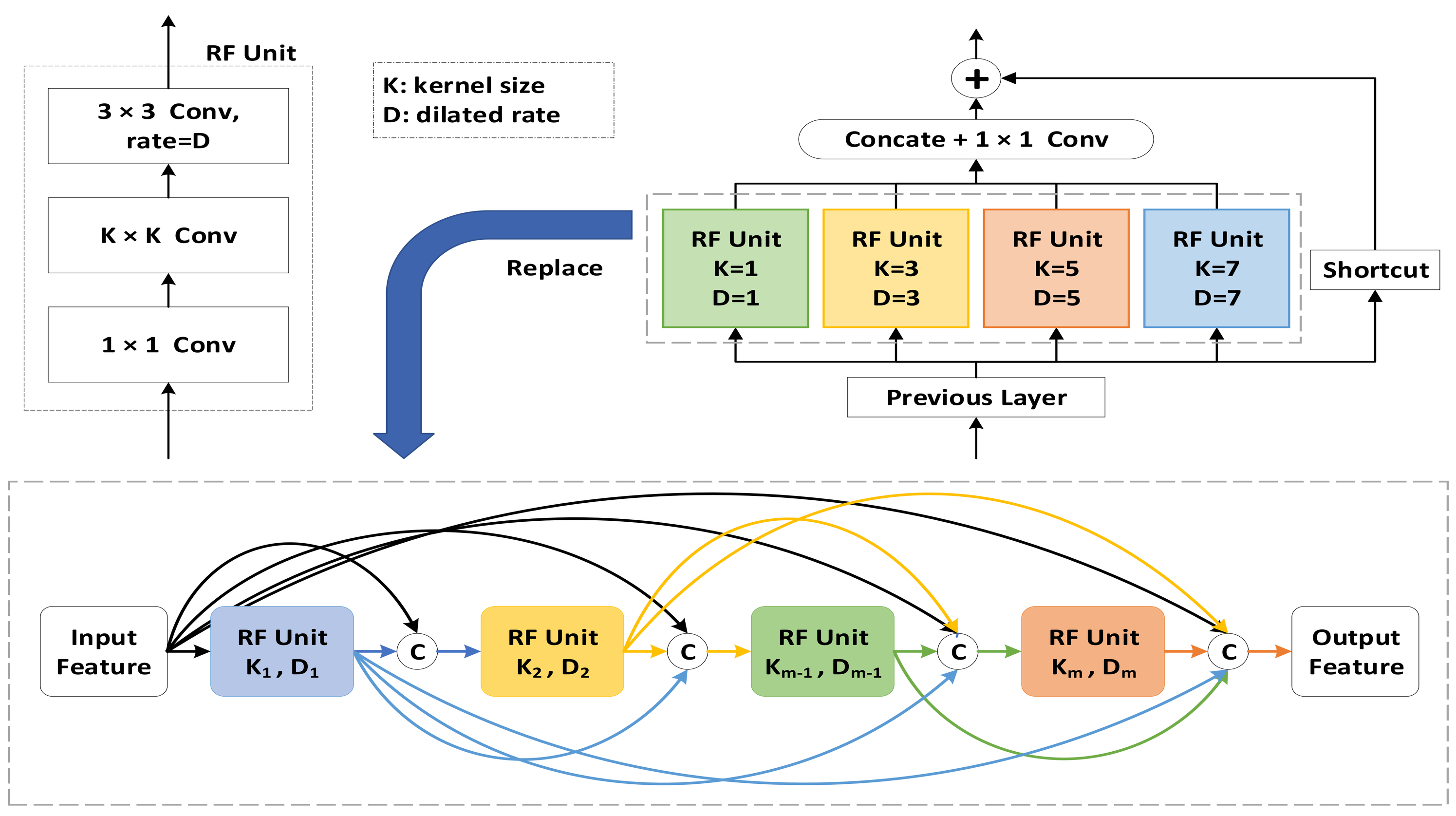
- Current backbone networks of object detection are mainly classification networks, which prevents the further improvement of the detector performance.
- Since the objects are distributed in different sizes and ratios, it is important for backbone networks to form dense receptive field for extracting discriminative features.
- Motivated by Receptive Filed Block (RFB) and Dense ASPP, we propose a more effective architecture named Dense Receptive Field Block (Dense RFB).
- With some modifications and proposed architecture, we present Dense Receptive Field Network (DRFNet) based on Darknet-53.
- DRFNet is a backbone network suitable for multi-scale detectors like ScratchDet and SSD.

Network Modification



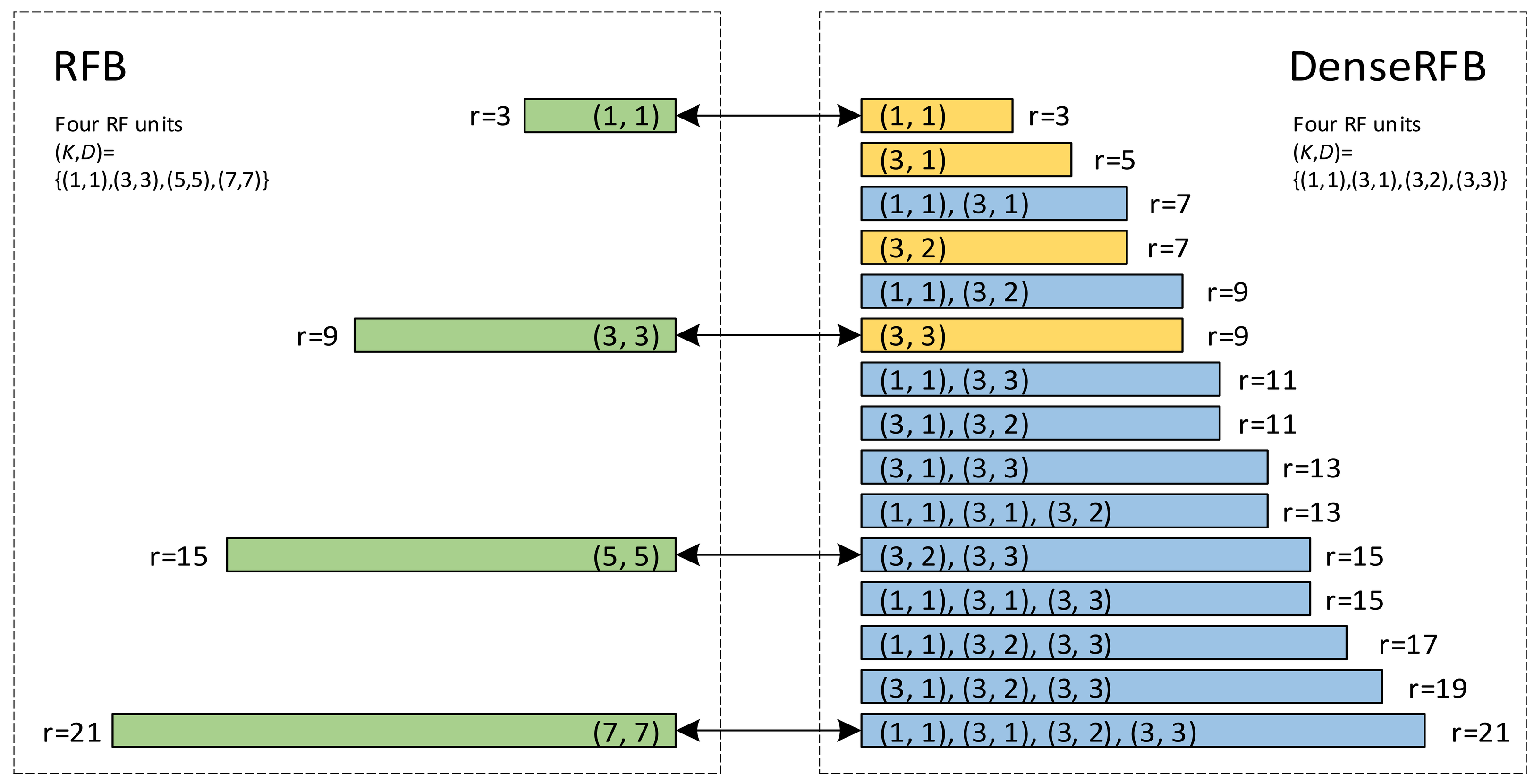
- *Based on effective network -- Darknet-53*
- *Additional Residual Block in Root Stage (input size)*
- *Alternative Down-sampling Block*
- *Less traditional conv block*
- *Dense Receptive Field Blocks for detection stage (stride = 8 / 16 / 32)*

Dense Receptive Field Block



- **The RF unit is the basic component to generate effective receptive field.**
- **The rearrangement of RF units from parallel to dense connection helps Dense Receptive Field Block to form denser ERF for better performance.**

Receptive Field Pyramid



- **The comparison between RFB and DenseRFB.**
- **With smaller kernel size and dilated rate, DenseRFB forms a dense receptive field pyramid, it helps to capture more discriminative features.**

Results

Trained From Scratch on Pascal VOC

One-stage Method	Backbone	Input size	mAP(%)	FPS
SSD300 [1]	VGG-16	300×300	77.2	120
DSSD321 [16]	ResNet-101	321×321	78.6	10
STDN321 [17]	DenseNet-169	321×321	79.3	40
DES300 [25]	VGG-16	300×300	79.7	30
RefineDet320 [26]	VGG-16	320×320	80.0	40
RFBNet300 [10]	VGG-16	300×300	80.5	83
ScratchDet300 [7]	**DRFNet**	300×300	**81.0**	61

Trained From Scratch on MS COCO

One-stage Method	Backbone	AP	AP_{50}	AP_{75}	AP_s	AP_m	AP_l	FPS
DSSD321 [16]	ResNet-101	28.0	46.1	29.2	7.4	28.1	47.6	9.5
STDN300 [17]	DenseNet-169	28.0	45.6	29.4	7.9	29.7	45.1	38
RefineDet320 [26]	VGG-16	29.4	49.2	31.3	10.0	32.0	44.4	36
RFBNet300 [10]	VGG-16	30.3	49.3	31.8	11.8	31.9	45.9	66
RetinaNet400 [3]	ResNet-50-FPN	30.5	47.8	32.7	11.2	33.8	46.1	16
RetinaNet400 [3]	ResNet-101-FPN	31.9	49.5	34.1	11.6	35.8	48.5	12
SratchDet300 [7]	Root-ResNet-34	32.7	52.0	34.9	13.0	35.6	49.0	25
SratchDet300 [7]	**DRFNet**	32.2	50.4	34.4	**13.2**	**37.2**	48.2	54

Trained on ImageNet then trained on MS COCO

Model	Top-1	Top-5	FPS
ResNet-101 [18]	77.4	93.7	112
ResNet-152 [18]	78.4	94.1	70
DRFNet	78.5	94.1	143

One-stage Method	Backbone	AP	AP_{50}	AP_{75}	AP_s	AP_m	AP_l	FPS
STDN513 [17]	DenseNet-169	31.8	51.0	33.6	14.4	36.1	43.4	25
RefineDet512 [26]	VGG-16	33.0	54.5	35.5	16.3	36.3	44	22
YOLOv3 [5]	Darknet-53	33.0	**57.9**	34.4	18.3	35.4	41.9	20
DSSD513 [16]	ResNet-101	33.2	53.3	35.2	13.0	35.4	**51.1**	6
RFBNet512 [10]	VGG-16	33.8	54.2	35.9	16.2	37.1	47.4	**33**
RetinaNet500 [3]	ResNet-101-FPN	34.4	53.1	36.8	14.7	38.5	49.1	11
SSD512 [1]	**DRFNet**	**34.7**	53.2	**37.2**	**18.7**	**43.2**	50.2	29

Conclusion

- We present an effective and efficient backbone network called DRFNet for object detection.
- DRFNet is based on Darknet-53 with some modifications and a new architecture.
- The new architecture called Dense Receptive Field Block (Dense RFB) can form much denser RF as well as ERF for better performance.
- Our DRFNet can achieve competitive results on the simple detection Framework – SSD.

Reference

[1] Liu, S., Huang, D., et al. Receptive field block net for accurate and fast object detection. ECCV,2018.
[2] M. Yang, K. Yu, C. Zhang, Z. Li, and K. Yang. Denseaspp for semantic segmentation in street scenes. CVPR,2018.
[3] Zhu, R., Zhang, S., Wang, X., Wen, L., Shi, H., Bo, L., Mei, T. Scratchdet: Exploring to train single-shot object detectors from scratch. CVPR,2019.
[4] Redmon, J., Farhadi, A.: Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767 (2018)

Other contact Information:
Fei Gao, gfei_jack@163.com, WeChat: Kevin-FG
Chengguang Yang, chg_yang@qq.com, WeChat: Light_ChgY

ManuscriptID: 225