

1. Architecture details for experiments on MS-COCO 2017

Table 1 details the architecture the experiments used on MS-COCO 2017.

2. Fine-grain metrics for experiments on MS-COCO 2017

Table 2 presents fine-grain object detection metrics on MS-COCO 2017. Table 3 presents fine-grain instance segmentation metrics on MS-COCO 2017.

References

- [1] Xiaolong Wang, Ross Girshick, Abhinav Gupta, and Kaiming He. Non-local neural networks. In *CVPR*, 2018.

Table 1. **Architecture details for experiments on MS-COCO 2017 object detection and instance segmentation.** This table assumes the backbone architecture is ResNet-50. For ResNet-101 and ResNeXt-101, the only difference will be the number of ResBlocks in each ResBlock group (res1-4) and/or the type of the blocks (ResNeXtBlock (32x4d) instead of ResBlock)

Block	Type	Input	Output size
input	Input	N/A	896×1280
conv1	Conv 3×3	input	448×640
maxpool	Maxpool 2×2	conv1	224×320
res1	ResBlock $\times 3$	maxpool	224×320
res2	ResBlock $\times 4$	res1	112×160
res3	ResBlock $\times 6$	res2	56×80
res4	ResBlock $\times 3$	res3	28×40
fpn5	conv 3×3	res4	14×20
fpn4	conv 3×3	res4 + fpn5 (upsampled)	28×40
fpn3	conv 3×3	res3 + fpn4 (upsampled)	56×80
fpn2	conv 3×3	res2 + fpn3 (upsampled)	112×160
fpn1	conv 3×3	res1 + fpn2 (upsampled)	224×320
rpn	RPN	fpn1-4	N/A
roi	RoI Align	fpn1-4	N/A

Table 2. **Fine-grain metrics for experiments on MS-COCO 2017 object detection.** $+n$ NL means adding n non-local [1] blocks to the backbone and FPN. $+n$ EA means adding n EA modules to the backbone and FPN. OOM indicates out-of-memory errors

Backbone	AP	AP-50	AP-75	AP-small	AP-medium	AP-large
ResNet-50	39.4	60.6	42.8	24.7	43.0	50.9
+1 NL	40.3	61.9	43.6	24.3	43.8	52.2
+1 EA	40.2	61.9	43.6	24.9	44.0	51.5
+5 NL	40.7	62.1	44.2	25.3	44.5	52.0
+5 EA	40.6	62.8	44.2	25.0	44.6	52.3
+7 NL	OOM	OOM	OOM	OOM	OOM	OOM
+7 EA	41.2	62.7	44.8	25.8	44.9	52.5
ResNeXt-101	43.5	65.4	47.5	27.0	47.9	55.3
+7 EA	44.9	66.8	48.7	27.1	49.1	57.6

Table 3. **Fine-grain metrics for experiments on MS-COCO 2017 instance segmentation.** $+n$ NL means adding n non-local [1] blocks to the backbone and FPN. $+n$ EA means adding n EA modules to the backbone and FPN. OOM indicates out-of-memory errors

Backbone	AP	AP-50	AP-75	AP-small	AP-medium	AP-large
ResNet-50	35.1	57.0	37.2	19.9	38.1	47.7
+1 NL	35.9	58.2	38.1	20.9	39.2	47.9
+1 EA	36.0	58.0	38.3	20.3	39.0	47.9
+5 NL	36.3	59.4	38.5	19.7	39.8	50.4
+5 EA	36.2	59.2	38.2	19.9	39.9	49.4
+7 NL	OOM	OOM	OOM	OOM	OOM	OOM
+7 EA	36.7	59.1	39.2	21.6	40.2	48.6
ResNeXt-101	38.5	61.8	41.0	21.7	42.5	51.7
+7 EA	39.3	63.0	42.0	22.6	43.0	52.4