清華大學
Tsinghua University

i-VisionGroup

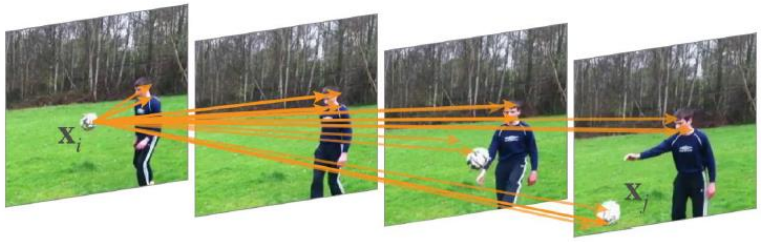
文献分享

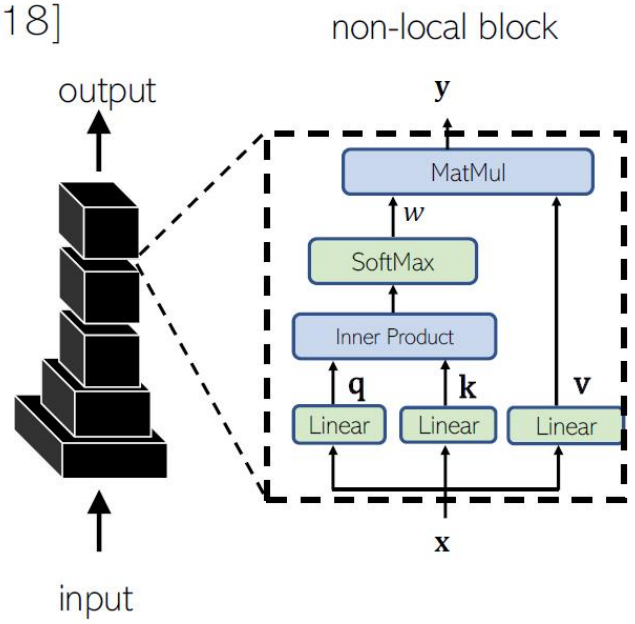
范博昊

Non-local Neural Networks

- 非局部网络

- Non-Local Networks [Wang et al, CVPR'2018]

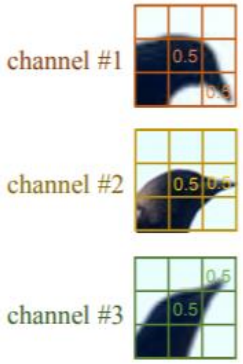


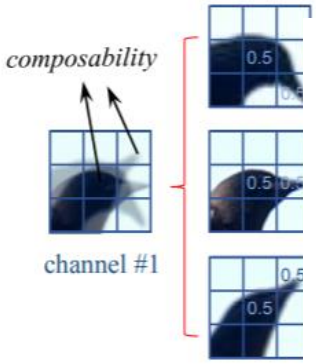


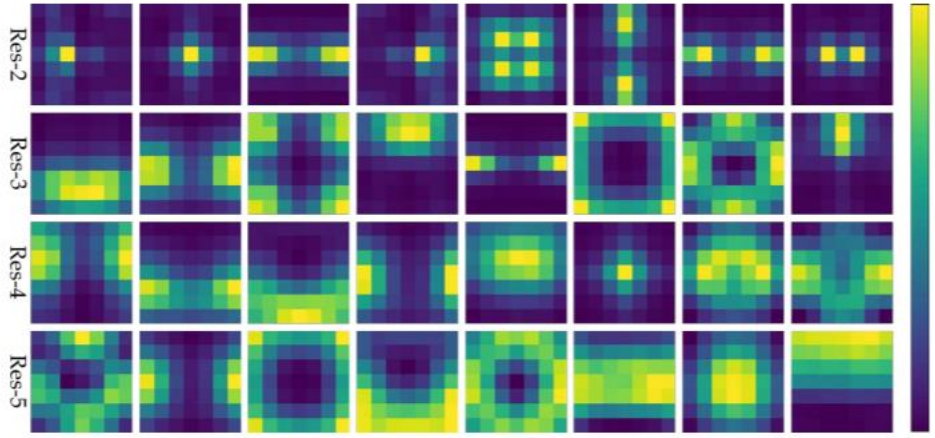
Non-local与全连接的区别在于其权重和特征值有关

归一化：将softmax替换成了N，也就是求解的总个体数

Local Relation Networks for Image Recognition







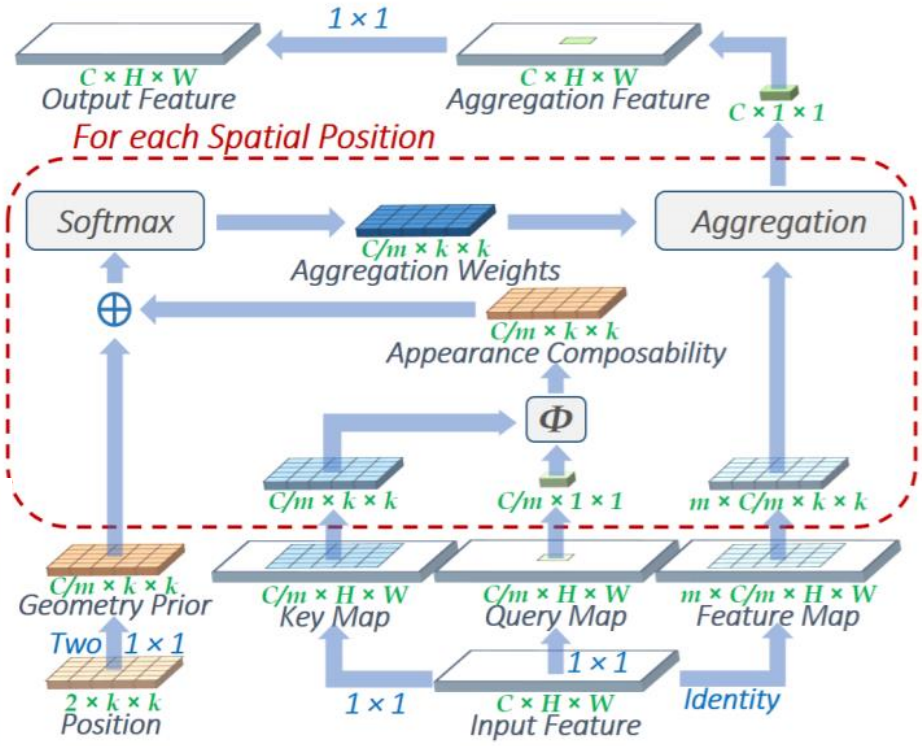
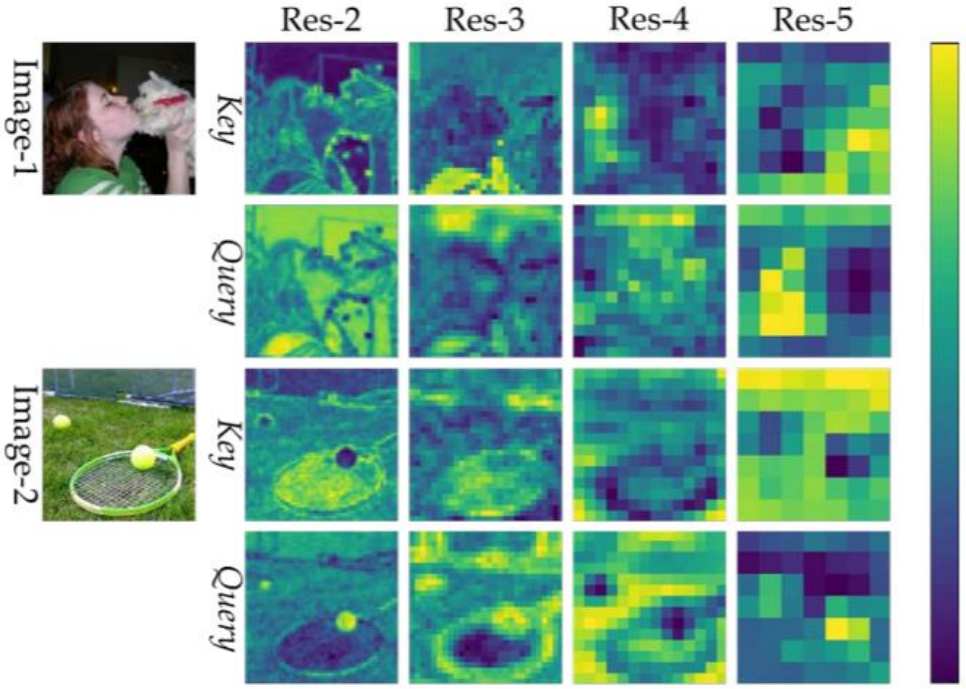


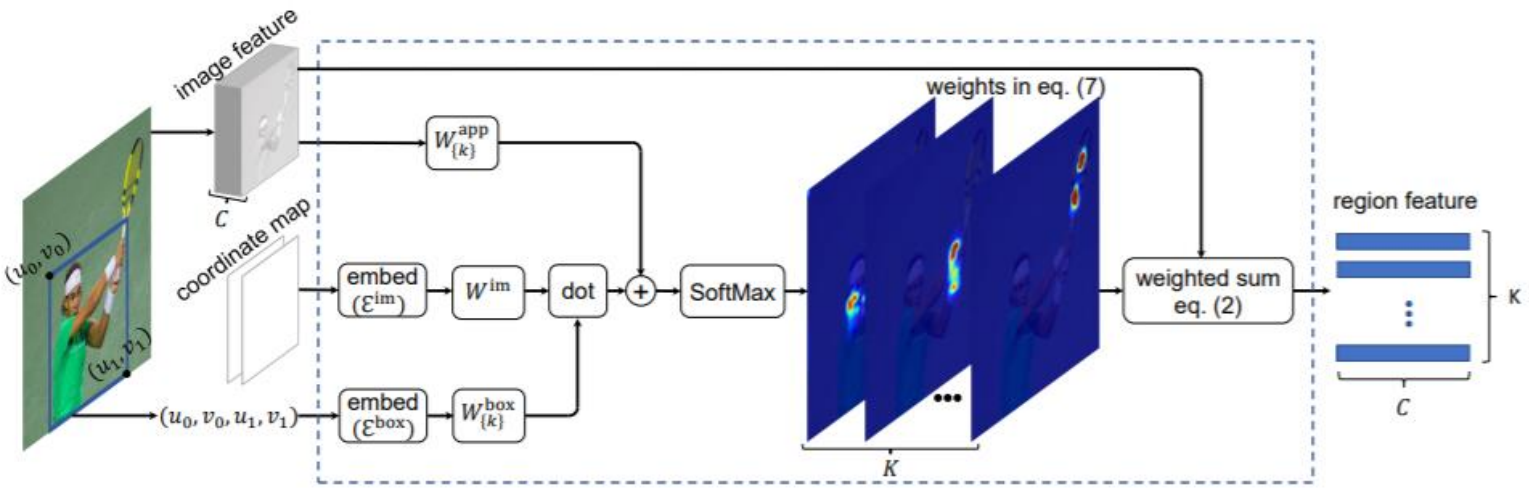
Figure 2. The local relation layer.

Local Relation Networks for Image Recognition



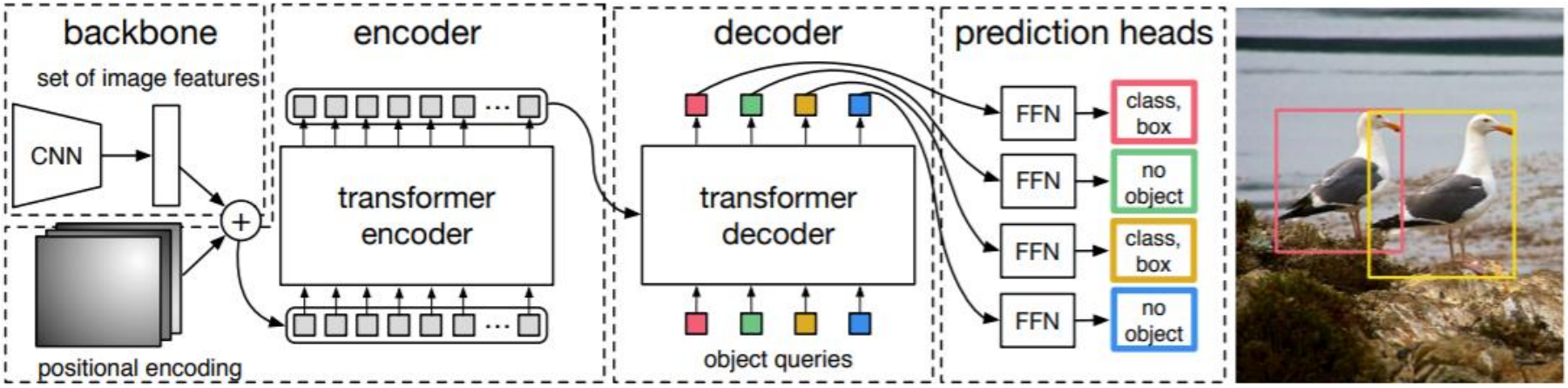
stage	output	ResNet-50	**LR-Net-50 (7×7, m=8)**
res1	112×112	7×7 conv, 64, stride 2	**1×1, 64** **7×7 LR, 64, stride 2**
res2	56×56	3×3 max pool, stride 2	3×3 max pool, stride 2
		[1×1, 64 3×3 conv, 64 1×1, 256] ×3	[1×1, 100 **7×7 LR, 100** 1×1, 256] ×3
res3	28×28	[1×1, 128 3×3 conv, 128 1×1, 512] ×4	[1×1, 200 **7×7 LR, 200** 1×1, 512] ×4
res4	14×14	[1×1, 256 3×3 conv, 256 1×1, 1024] ×6	[1×1, 400 **7×7 LR, 400** 1×1, 1024] ×6
res5	7×7	[1×1, 512 3×3 conv, 512 1×1, 2048] ×3	[1×1, 800 **7×7 LR, 800** 1×1, 2048] ×3
	1×1	global average pool 1000-d fc, softmax	global average pool 1000-d fc, softmax
# params		**25.5**×10^6	**23.3**×10^6
FLOPs		**4.3**×10^9	**4.3**×10^9

Learning Region Features for Object Detection



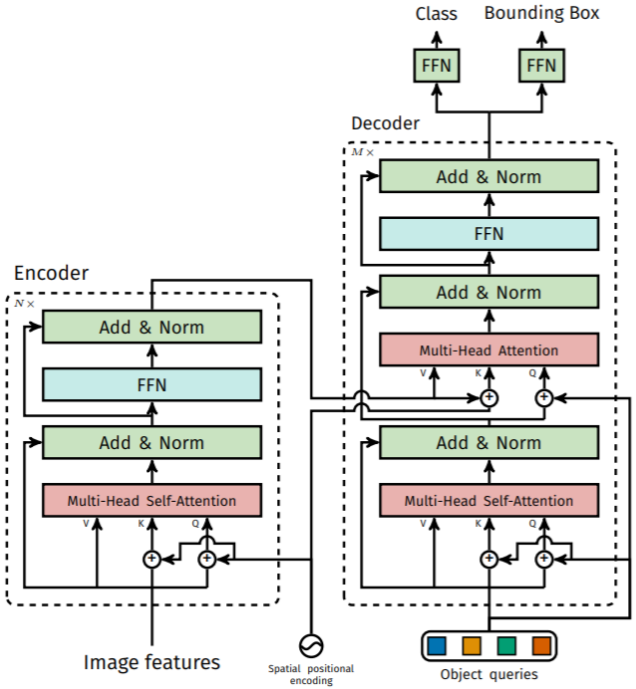
$$w_k(b, p, \mathbf{x}) \propto \exp(G_k(b, p) + A_k(\mathbf{x}, p)).$$

End-to-End Object Detection with Transformers

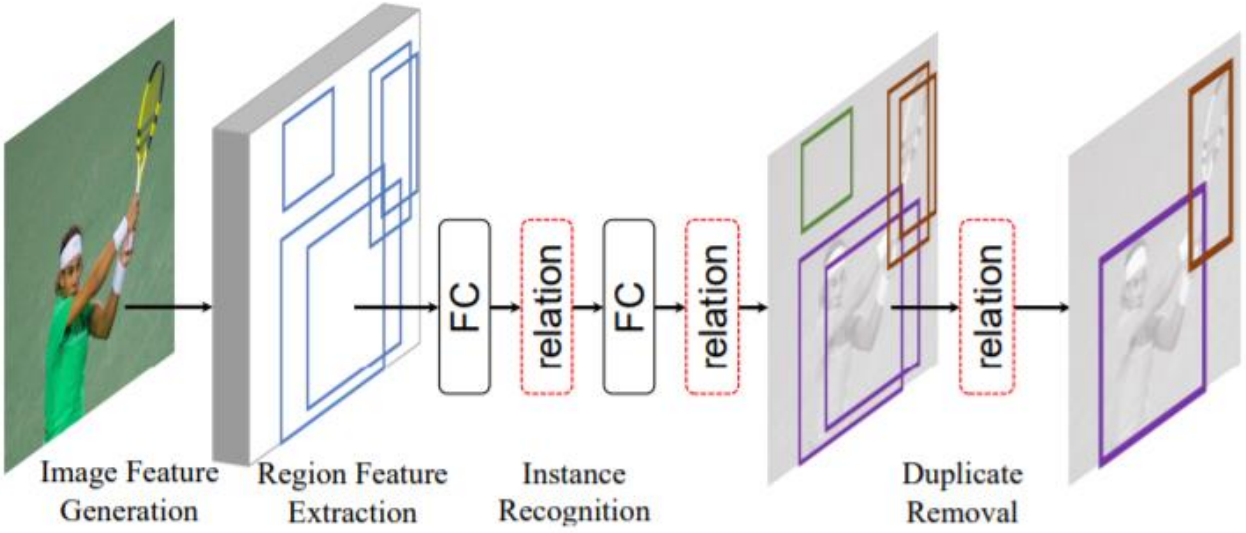


C*H*W->d*H*W->d*HW

FFN->center+bbox+class



Relation Networks for Object Detection



Geometry feature f_G+ Image feature f_A->Relation feature f_R

$$\mathbf{f}_R(n) = \sum_m \omega^{mn} \cdot (W_V \cdot \mathbf{f}_A^m) \qquad \omega^{mn} = \frac{\omega_G^{mn} \cdot \exp(\omega_A^{mn})}{\sum_k \omega_G^{kn} \cdot \exp(\omega_A^{kn})}$$

$$\omega_A^{mn} = \frac{dot(W_K \mathbf{f}_A^m, W_Q \mathbf{f}_A^n)}{\sqrt{d_k}} \qquad \omega_G^{mn} = \max\{0, W_G \cdot \mathcal{E}_G(\mathbf{f}_G^m, \mathbf{f}_G^n)\}$$

Relation Networks for Object Detection

backbone	test set	mAP	mAP_{50}	mAP_{75}	#. params	FLOPS
faster RCNN [38]	*minival*	32.2→34.7→**35.2**	52.9→55.3→**55.8**	34.2→37.2→**38.2**	58.3M→64.3M→64.6M	122.2B→124.6B→124.9B
	test-dev	32.7→35.2→**35.4**	53.6→**56.2**→56.1	34.7→37.8→**38.5**		
FPN [32]	*minival*	36.8→38.1→**38.8**	57.8→59.5→**60.3**	40.7→41.8→**42.9**	56.4M→62.4M→62.8M	145.8B→157.8B→158.2B
	test-dev	37.2→38.3→**38.9**	58.2→59.9→**60.5**	41.4→42.3→**43.3**		
DCN [10]	*minival*	37.5→38.1→**38.5**	57.3→57.8→**57.8**	41.0→41.3→**42.0**	60.5M→66.5M→66.8M	125.0B→127.4B→127.7B
	test-dev	38.1→38.8→**39.0**	58.1→**58.7**→58.6	41.6→42.4→**42.9**		

Table 5. Improvement (2fc head+SoftNMS [4], 2fc+RM head+SoftNMS and 2fc+RM head+e2e from left to right connected by →) in state-of-the-art systems on COCO *minival* and *test-dev*. Online hard example mining (OHEM) [40] is adopted. Also note that the strong SoftNMS method ($\sigma = 0.6$) is used for duplicate removal in non-e2e approaches.