



STRUCTURED SPATIAL REASONING FOR HUMAN POSE ESTIMATION

Ying Huang¹, Shanfeng Hu², Zi-ke Zhang³
¹Hangzhou Normal University, Hangzhou, China
²Northumbria University, Newcastle, UK
³Zhejiang University, Hangzhou, China

CHALLENGES

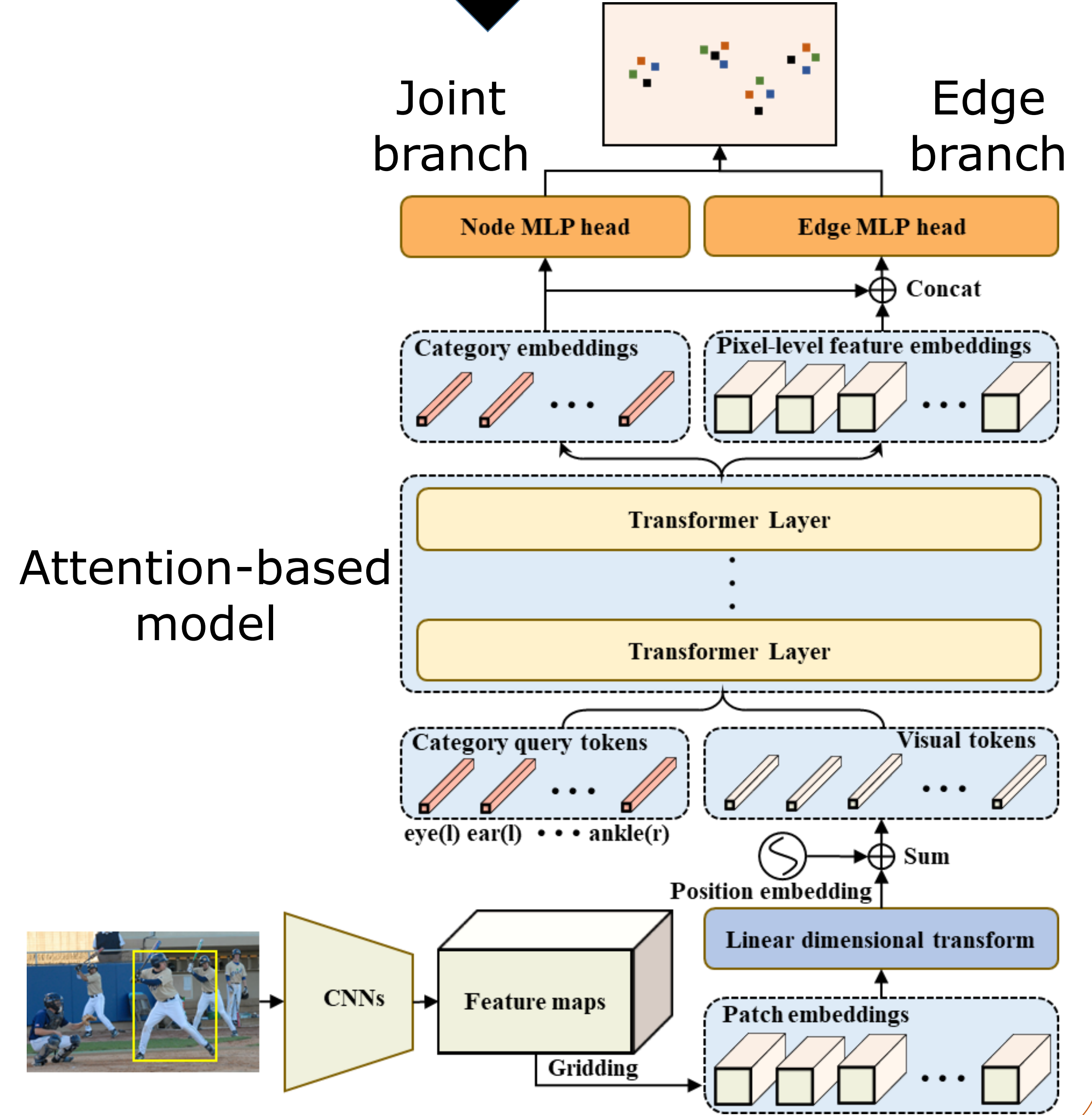
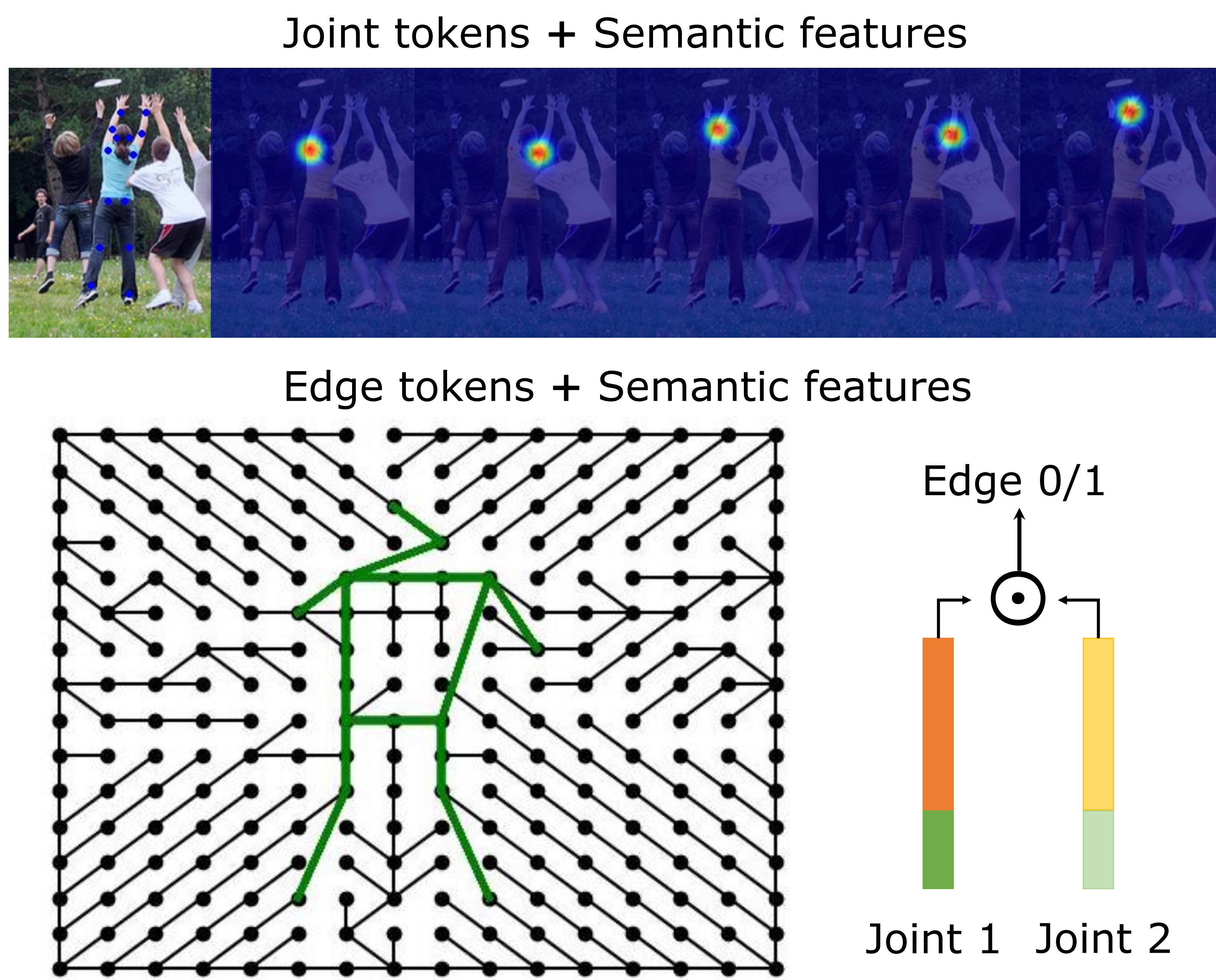


- Overlapping
- Similar appearance
- Similar action
- Small limbs
- Complex scene

SOLUTION

- Using both human joint and limb information to form a graph.
- Performing structured spatial reasoning on the graph.

FRAMEWORK



RESULTS

Approach	Input size	#Params	GFLOPs	AP	AP ^{SU}	AP ⁷⁵	AP ^M	AP ^L	AR
Simple-Res50	256×192	34.0M	8.9	70.4	88.6	78.3	67.1	77.2	76.3
Simple-Res101	256×192	53.0M	12.4	71.4	89.3	79.3	68.1	78.1	77.1
Simple-Res152	256×192	68.6M	15.7	72.0	89.3	79.8	68.7	78.9	77.8
HRNet-W32	256×192	28.5M	7.1	74.4	90.5	81.9	70.8	81.0	79.8
HRNet-W48	256×192	63.6M	14.6	75.1	90.6	82.2	71.5	81.8	80.4
PureTransformer	256×192	5.8M	1.3	65.6	86.4	73.0	63.1	71.5	72.1
TokenPose-L/D24	256×192	27.46M	10.98	75.0	89.7	81.9	71.7	81.8	80.3
TransPose-H-A6	256×192	17.5M	21.8	75.0	89.8	81.9	71.7	81.7	80.2
Ours	256×192	27.47M	10.99	75.3	90.6	82.6	72.3	79.5	80.1
Ours ⁺	256×192	27.47M	10.99	77.8	93.6	84.8	74.9	81.9	80.2

COCO-val dataset

