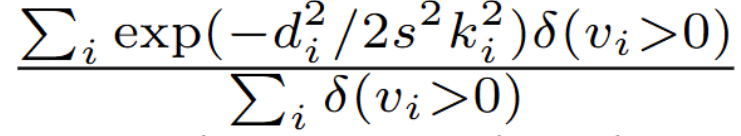
1. **Dataset and Evaluation metric**  
   **1.1. COCO Keypoint Detection**  
   (1)introduction  
   he COCO dataset contains over 200, 000 images and 250, 000 person instances labeled with 17 key- points. We train our model on COCO train2017 dataset,including 57K images and 150K person instances. We evaluate our approach on the val2017 set and test-dev2017 set,containing 5000 images and 20K images, respectively  
   (2)Evaluation metric  
   OKS(Object Keypoint Similarity): uesd to determine whether the detection is correct

  
di is the Euclidean distance between the detected keypoint and the correspond-  
ing ground truth, vi is the visibility flag of the ground truth, s is the object scale, and ki is a per-keypoint constant that controls falloff.  
  
**1.2. MPII Human Pose Estimation**  
(1)introduction  
The MPII Human Pose dataset [2] consists of images taken from a wide-range of real-world activities with full-body pose annotations. There are around 25K images  
with 40K subjects, where there are 12K subjects for testing and the remaining subjects for the training set.  
(2)Evaluaion metric  
PCKh(head-normalized probability of correct keypoint) score  
A joint is correct if it falls within αl pixels of the groundtruth position, where α is a constant and l is the head size that corresponds to 60% of the diagonal length of the  
ground-truth head bounding box  
**1.3. Pose Tracking**(1)introduction  
PoseTrack [28] is a large-scale benchmark for human pose estimation and articulated tracking in video. Thedataset, based on the raw videos provided by the popular.MPII Human Pose dataset, contains 550 video sequences with 66, 374 frames.The video sequences are split into 292, 50, 208 videos for training, validation, and testing, re-  
spectively. The length of the training videos ranges between 41 − 151 frames, and 30 frames from the center of the video are densely annotated. The number of frames in the validation/testing videos ranges between 65 − 298 frames. The 30 frames around the keyframe from the MPII Pose dataset are densely annotated, and afterwards every fourth frame is annotated. In total, this constitutes roughly 23, 000 labeled frames and 153, 615 pose annotations.  
(2)Evaluation metric  
mAP(single) MOTA(multi)