

Supplementary Material

Apple recognition and picking sequence planning for harvesting robot in a complex environment

Wei Ji, Tong Zhang, Bo Xu, Guozhi He

Jiangsu University Ringgold standard institution, Zhenjiang, China

Input : $N_t, \mathcal{B} = \{b_1, \dots, b_N\}, \mathcal{S} = \{s_1, \dots, s_N\}$
 N_t is the NMS threshold
 \mathcal{B} is the list of initial detection boxes
 \mathcal{S} contains corresponding detection scores

```

begin
   $\mathcal{D} \leftarrow \{\}$ 
  while  $\mathcal{B} \neq \text{empty}$  do
     $m \leftarrow \text{argmax } \mathcal{S}$ 
     $\mathcal{M} \leftarrow b_m$ 
     $\mathcal{D} \leftarrow \mathcal{D} \cup \mathcal{M}; \mathcal{B} \leftarrow \mathcal{B} - \mathcal{M}$ 
    for  $b_i$  in  $\mathcal{B}$  do
      if  $iou(\mathcal{M}, b_i) \geq N_t$  then
         $\mathcal{B} \leftarrow \mathcal{B} - b_i; \mathcal{S} \leftarrow \mathcal{S} - b_i$ 
      end
    end
  end
  return  $\mathcal{D}, \mathcal{S}$ 
end

```

NMS

Soft-NMS

Figure S1. The pseudo-code in green is replaced with the one in red in soft non maximum suppression. We propose to revise the detection scores by scaling them as a linear or Gaussian function of overlap.

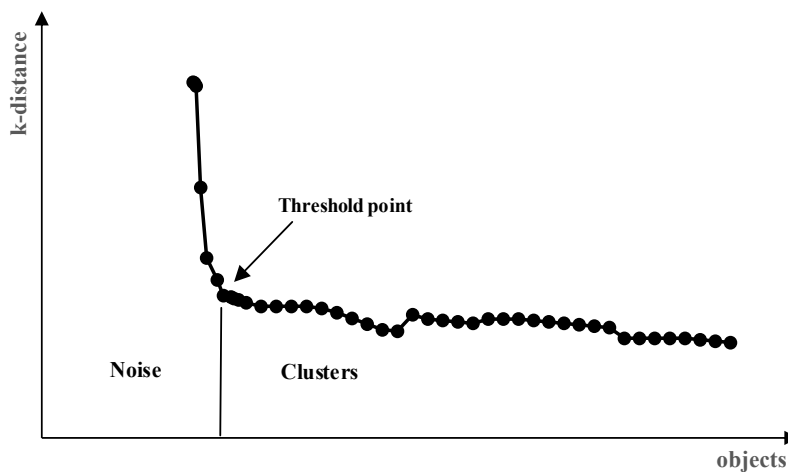


Figure S2. K-distance diagram.

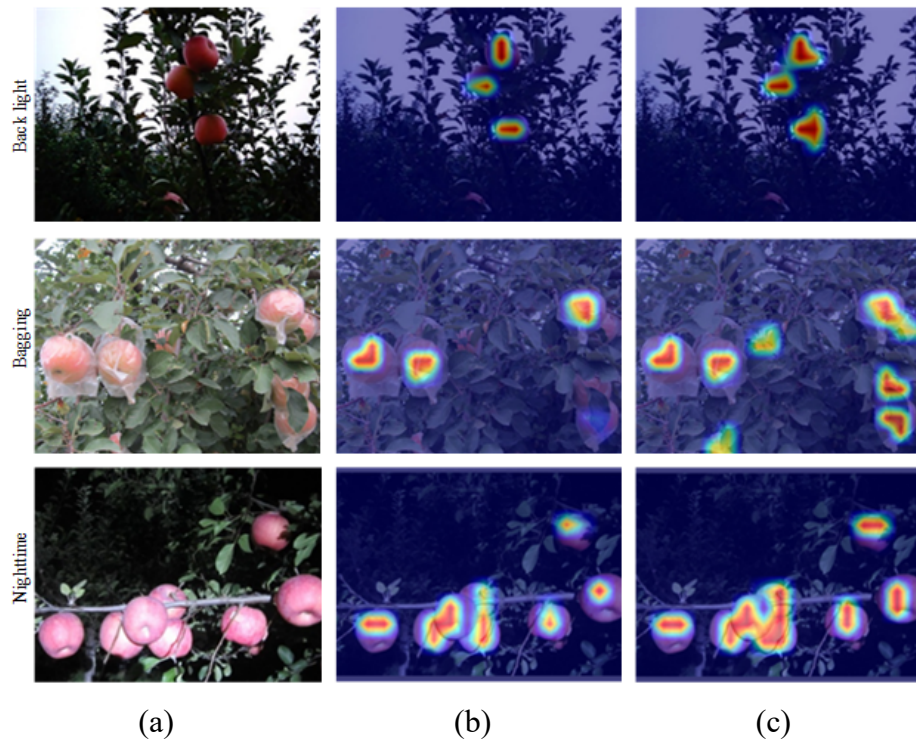


Figure S3. Feature extraction heat map distribution in different environments. a) Original images; b) YOLOV5s; c) EF-YOLOV5s.