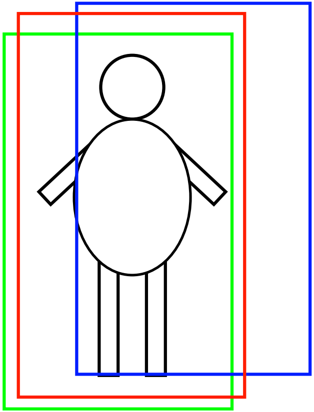
FeatureNMS: Non-Maximum Suppression by Learning Feature Embeddings

Niels Ole Salscheider

January 13th, 2021

Motivation

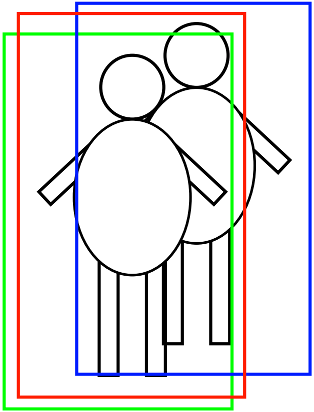
- State of the art object detectors: CNN designs
- They usually generate multiple detections per object
- Non-Maximum Suppression removes duplicates
  - Heuristic based on IoU



Motivation

- But assumptions of classical NMS do not always hold in crowded scenes

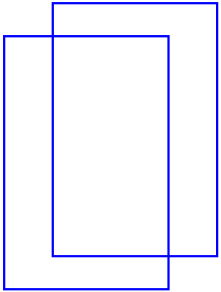
⇒ Idea: Rely on (visual) features in these cases!



Classical NMS

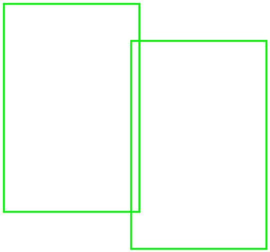
- Classical NMS compares IoU against threshold N

- IoU $\leq N \Rightarrow$ No duplicate
- IoU $> N \Rightarrow$ Duplicate



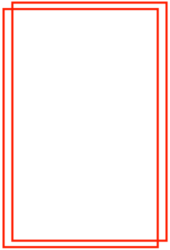
FeatureNMS

- FeatureNMS uses two IoU thresholds N_1 and N_2
- IoU $\leq N_1$ $\Rightarrow$ No duplicate



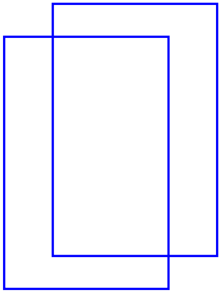
FeatureNMS

- FeatureNMS uses two IoU thresholds N_1 and N_2
- IoU $\leq N_1 \Rightarrow$ No duplicate
- IoU $> N_2 \Rightarrow$ Duplicate



FeatureNMS

- FeatureNMS uses two IoU thresholds N_1 and N_2
- IoU $\leq N_1 \Rightarrow$ No duplicate
- IoU $> N_2 \Rightarrow$ Duplicate
- $N_1 <$ IoU $\leq N_2 \Rightarrow$ Use similarity metric
  - Embedding distance $\geq$ T $\Rightarrow$ No duplicate
  - Embedding distance $< T \Rightarrow$ Duplicate



Feature Embedding

- Add network head to detection CNN

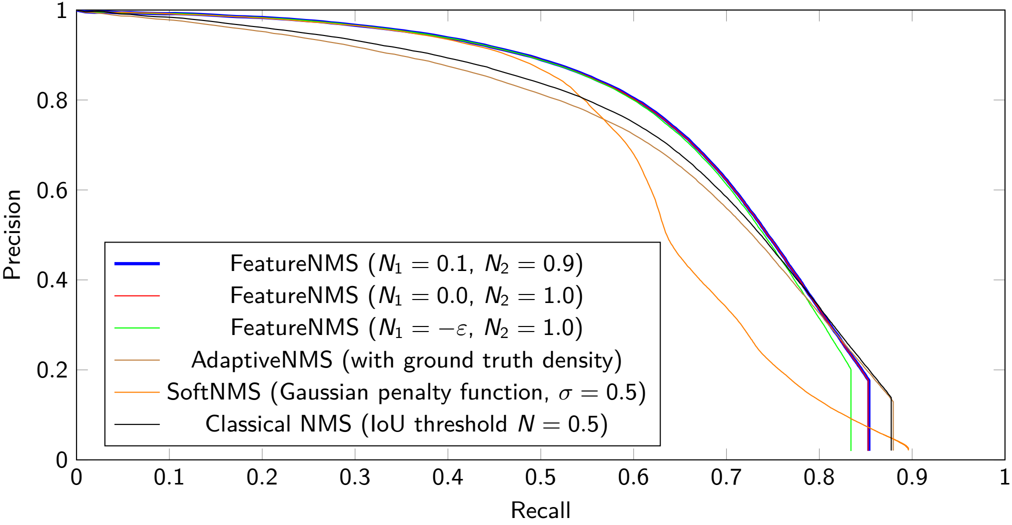
$\Rightarrow$ Predict embedding vector per detection

- Trained using Margin Loss
  - Same object: Distance below $\beta - \alpha$
  - Different objects: Distance above $\beta + \alpha$

Evaluation

- CrowdHuman dataset
- Train RetinaNet detector on training set, run on test set
- Post-process raw outputs with different NMS algorithms

Results



1
0.8
0.6
0.4
0.2
0
Precision
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Recall
FeatureNMS ($N_1 = 0.1$, $N_2 = 0.9$)
FeatureNMS ($N_1 = 0.0$, $N_2 = 1.0$)
FeatureNMS ($N_1 = -\varepsilon$, $N_2 = 1.0$)
AdaptiveNMS (with ground truth density)
SoftNMS (Gaussian penalty function, $\sigma = 0.5$)
Classical NMS (IoU threshold $N = 0.5$)

Conclusions

- FeatureNMS achieves state of the art performance
- It outperforms other approaches on the CrowdHuman dataset
- Learnt similarity metric is very discriminative

Thank you for your kind attention!

Feel free to contact me if you have any questions:

salscheider@fzi.de