FeatureNMS: Non-Maximum Suppression by Learning Feature Embeddings

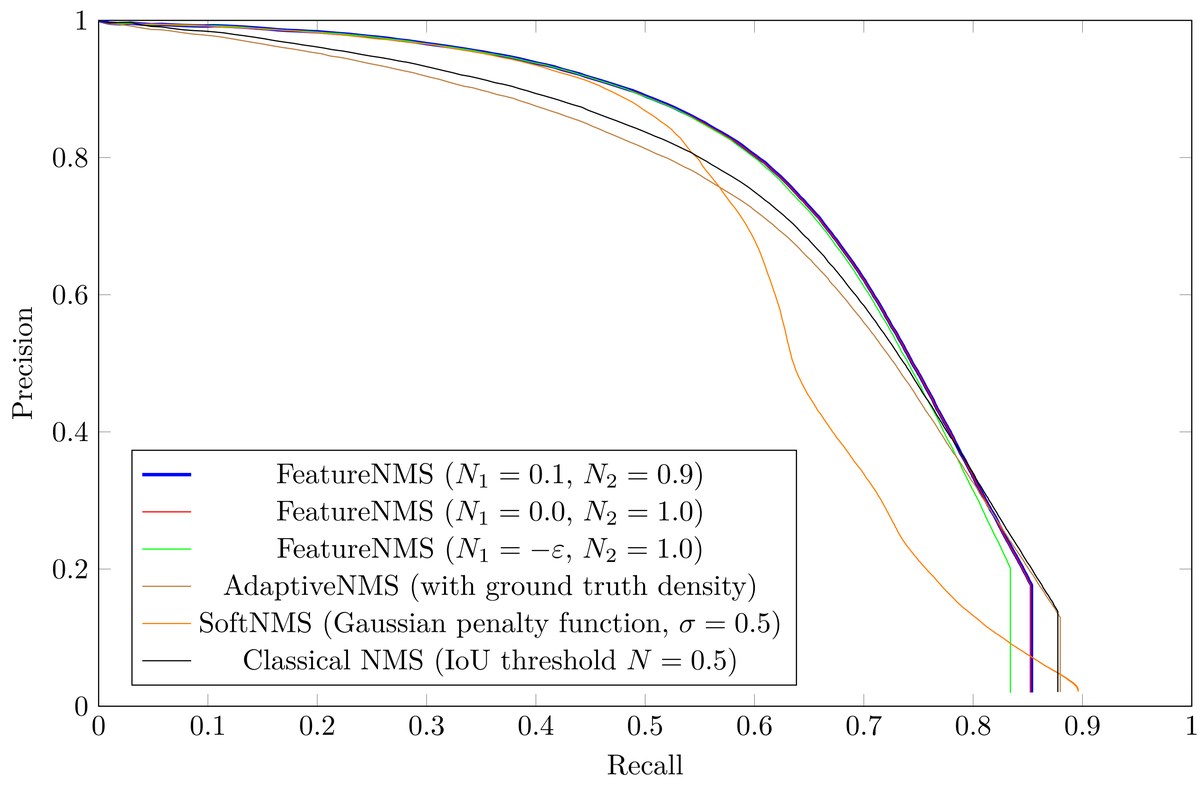
The classical Non-Maximum Suppression heuristic fails in scenes that contain objects with high overlap. FeatureNMS solves this problem by learning a similarity metric. This metric is used when the IoU does not allow to decide if a detection is a duplicate or not. It is computed on additional feature embedding vectors per anchor box. The object detector has to be modified to produce these feature vectors.

Our approach outperforms classical NMS and derived approaches and achieves state of the art performance.

Approach

- **Idea of classical NMS:** If Intersection over Union between two bounding boxes is over a threshold N, the second is a duplicate,
- otherwise not.
  Observation: There are cases in crowded scenes where the IoU does
- not allow to make a definite decision.
  Solution: Choose conservative thresholds N_1 and N_2 that allow to make a definite decision. If the IoU is however between N_1 and N_2, use a similarity metric between feature embeddings that describe the detected objects instead. We use the eucleadin distance as similarity
- metric.
  Feature embeddings:
  - Current state-of-the-art object detectors are based on CNN designs. Additional information can easily be predicted by adding another network head.

  ⇒ Predict one additional embedding vector per detection
  - Train the output using Margin Loss: Detections belonging to the same ground truth object should have a distance below a threshold β with margin α. Detections belonging to different ground truth objects should have a distance above a threshold β with margin α.



Results on the CrowdHuman dataset.

Results

FeatureNMS outperforms all other NMS approaches that we compared it to on the CrowdHuman dataset. This dataset contains many crowsed scenes where the assumptions of classical NMS do not hold.

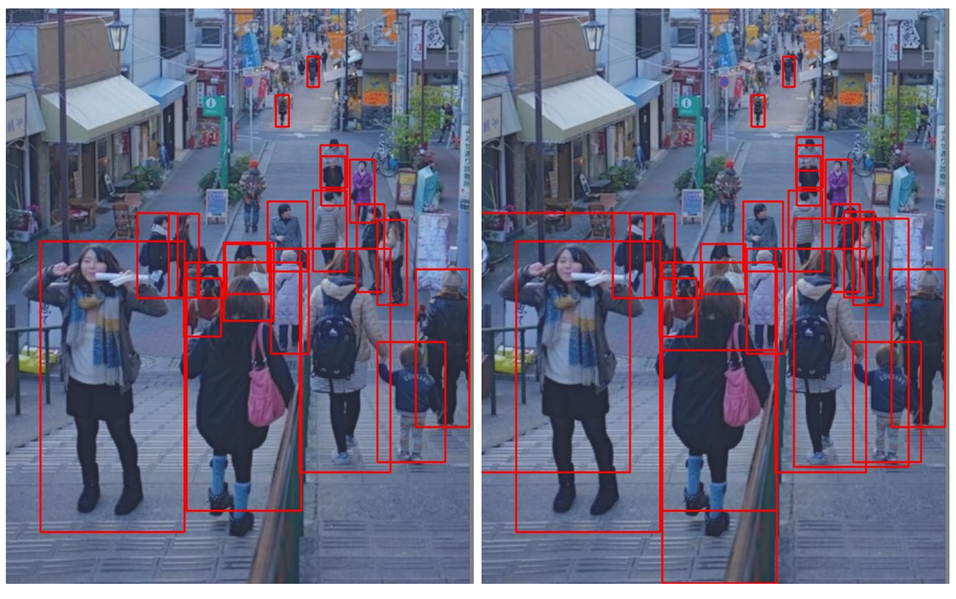
Our experiments show that FeatureNMS is not sensitive to the exact values of N_1 and N_2. Even when setting them to -ε and 1, the accuracy does not change notably. Doing so means that the heuristic solely depends on the similarity metric, and not on the classical heuristic. This highlights the discriminativeness of the learnt similarity metric.

$\mathcal{P} \leftarrow \text{GETPROPOSALS}(image)$
$\mathcal{P} \leftarrow \text{SORT}(\mathcal{P})$
$\mathcal{D} \leftarrow \emptyset$
while $\mathcal{P} \neq \emptyset$ **do**
  $p \leftarrow \text{POP}(\mathcal{P})$
  $isDuplicate \leftarrow \text{false}$
  for $d \in \mathcal{D}$ **do**
    $iou \leftarrow \text{GETIOU}(p, d)$
    if $iou > N_2$ **then**
      $isDuplicate \leftarrow \text{true}$
    else if $iou > N_1$ **then**
      $embeddingDist \leftarrow \text{GETEMBEDDINGDIST}(p, d)$
      if $embeddingDist < T$ **then**
        $isDuplicate \leftarrow \text{true}$
      end if
    end if
  end for
  if $\neg isDuplicate$ **then**
    $\text{PUSH}(p, \mathcal{D})$
  end if
end while

Proposed Algorithm. The FeatureNMS addition is highlighted.

Evaluation

We evaluate our approach on the CrowdHuman dataset.
We train the RetinaNet object detector on the training dataset and add an additional output per anchor box. This output predicts the embedding vectors and is trained using margin loss.
We then run the object detector on the test dataset and use different Non-Maximum suppression algorithms on the raw output. These include classical NMS, SoftNMS, AdaptiveNMS and FeatureNMS.



FeatureNMS Classical NMS
Example pictures.

FZI Research Center for Information Technology | Research Division ISPE-MPS
M.Sc. Niels Ole Salscheider | Email: salscheider@fzi.de