Cascade Mask Generation Framework for Fast Small Object Detection
Guangting Wang¹, Zhiwei Xiong¹, Dong Liu¹, Chong Luo²
¹University of Science and Technology of China ²Microsoft Research Asia

Motivation
- Performance of small object detection highly depends on input resolutions.
  - High-resolution input images incur high computational cost.
  - Low-resolution input images lead to low accuracy.
- Most background regions can be detected easily, even in a low resolution-image.
- Skipping convolutional operations in background regions can accelerate the detection processing.

Key Idea
- Multi-scale training and testing.
- Estimate background regions for the next scale
- Skip background regions in CNN.

MGM: Mask Generation Module
- The MGM estimates a coarse heatmap which indicates whether a region contains objects or not.
- In training phase, grids which have overlap with ground-truth boxes are set to be positive.

Discussion
- Discarding background regions can
  - accelerate forward-passing in CNN
  - reduce memory usage and allow larger input images
  - implicitly implement hard example mining

Experiments
- Detection performance on Tsinghua-Tencent 100K traffic sign dataset.
- Recall of the MGM
- Kept RoI percentage
- Visualization of results