Detecting Humans in RGB-D Data with CNNs

Kaiyang Zhou[1], Adeline Paiement[2] and Majid Mirmehdi[1]

University of Bristol

1. University of Bristol, UK. 2. Swansea University, UK

Swansea University
Prifysgol Abertawe

1. Motivations

1. Deep learning is state-of-the-art but slow.
2. RGB-D sensor can enhance computer vision.

Goals: apply deep CNNs to RGB-D human detection; use depth images for region-of-interest selection.

2. Challenges

1. How to handle noisy depth data for ROI selection
2. How to use small data to train Depth-CNN?
3. How to effectively fuse RGB and depth detections?

3. Proposed Approach

A. Overview

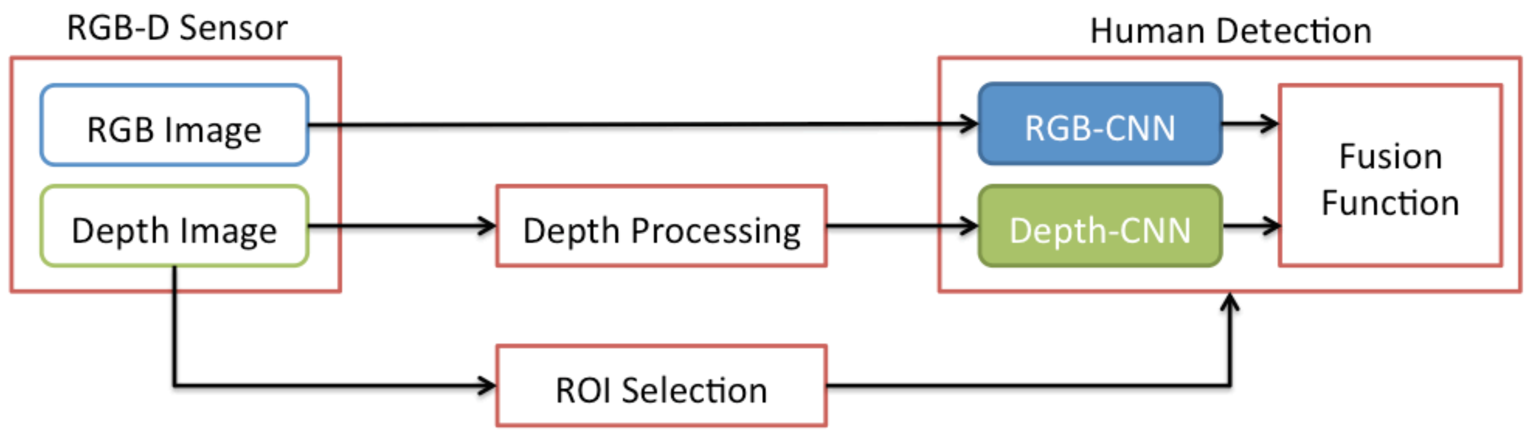
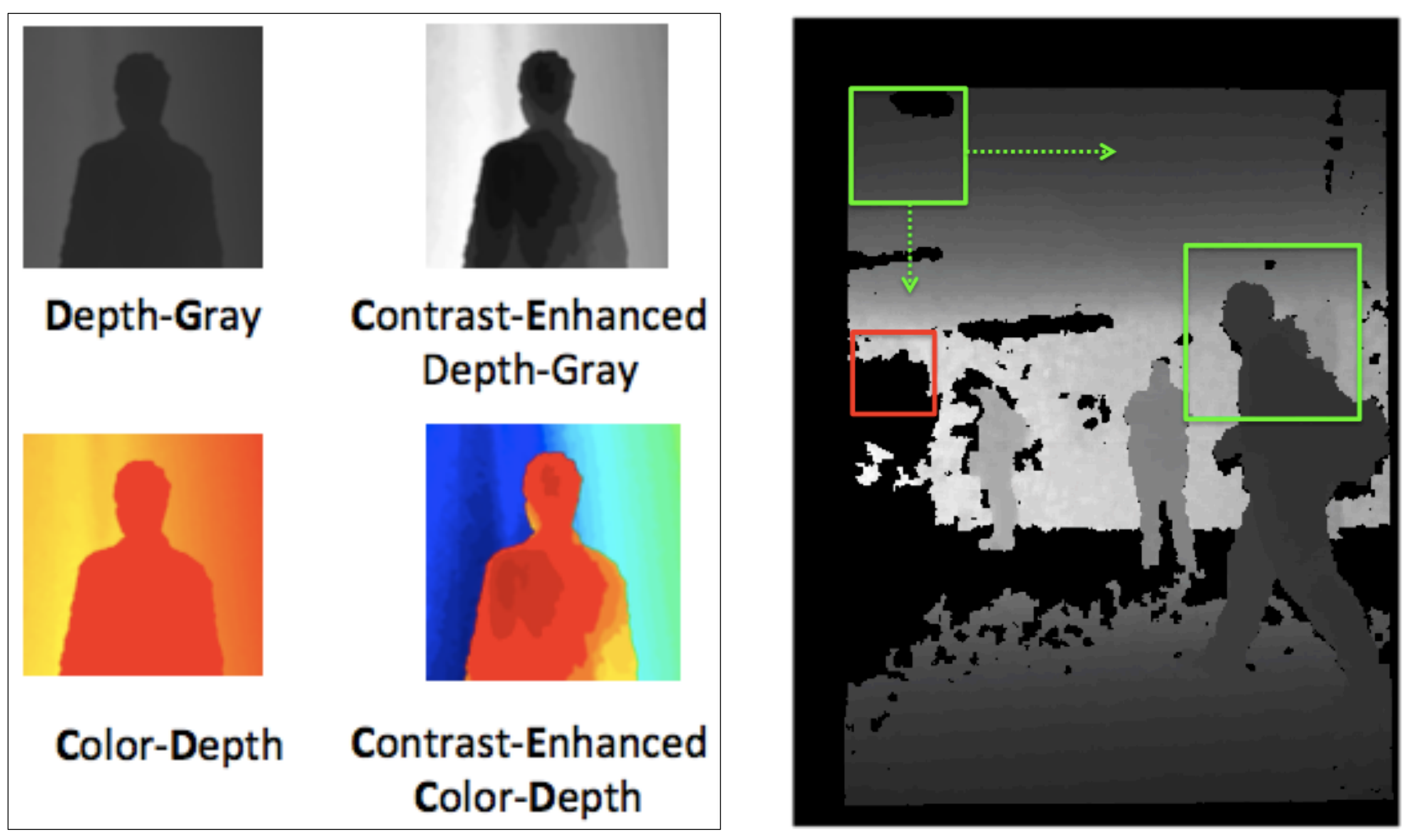


Fig. 1: Detection pipeline.

B. Depth Encoding and ROI Selection



C. Human Detection

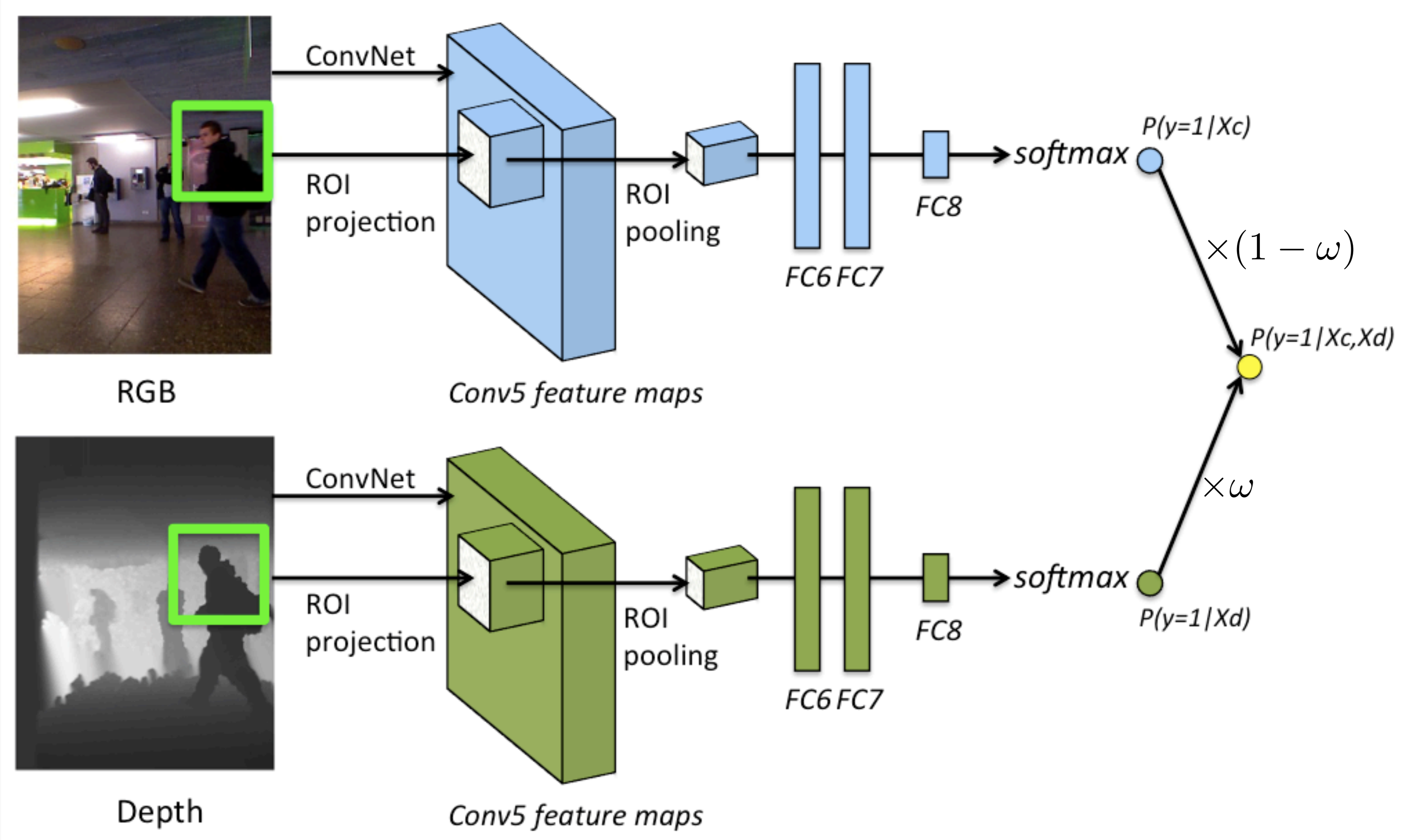


Fig. 3: CNN-based human detector.

Fig. 2: (Left) Different depth encodings: DG, CE, CD and our CECD. (Right) ROI selection, which has three steps: (a) Ground Plane Detection (GPD). (b) Scale-Informed ROI Search (SIS). (c) Candidate Proposals Filtering (CPF).

4. Experiments

1. Comparison of depth encoding schemes.

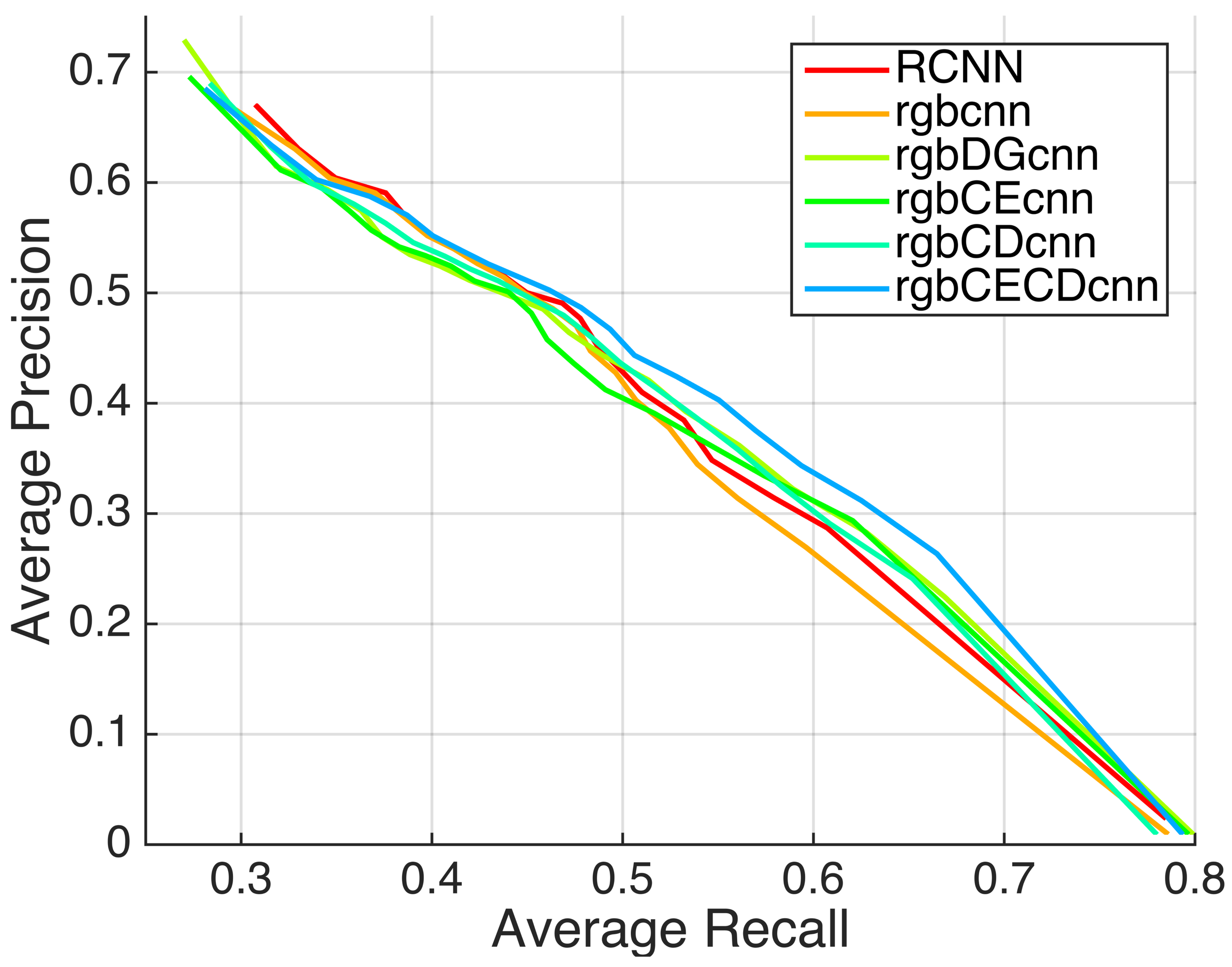


Fig. 4: Performances comparison between the baseline (RCNN and rgbcnn) and different RGB-Depth-CNNs. Our model with the proposed depth encoding (CECD) performs the best.

2. Evaluation on out ROI method.

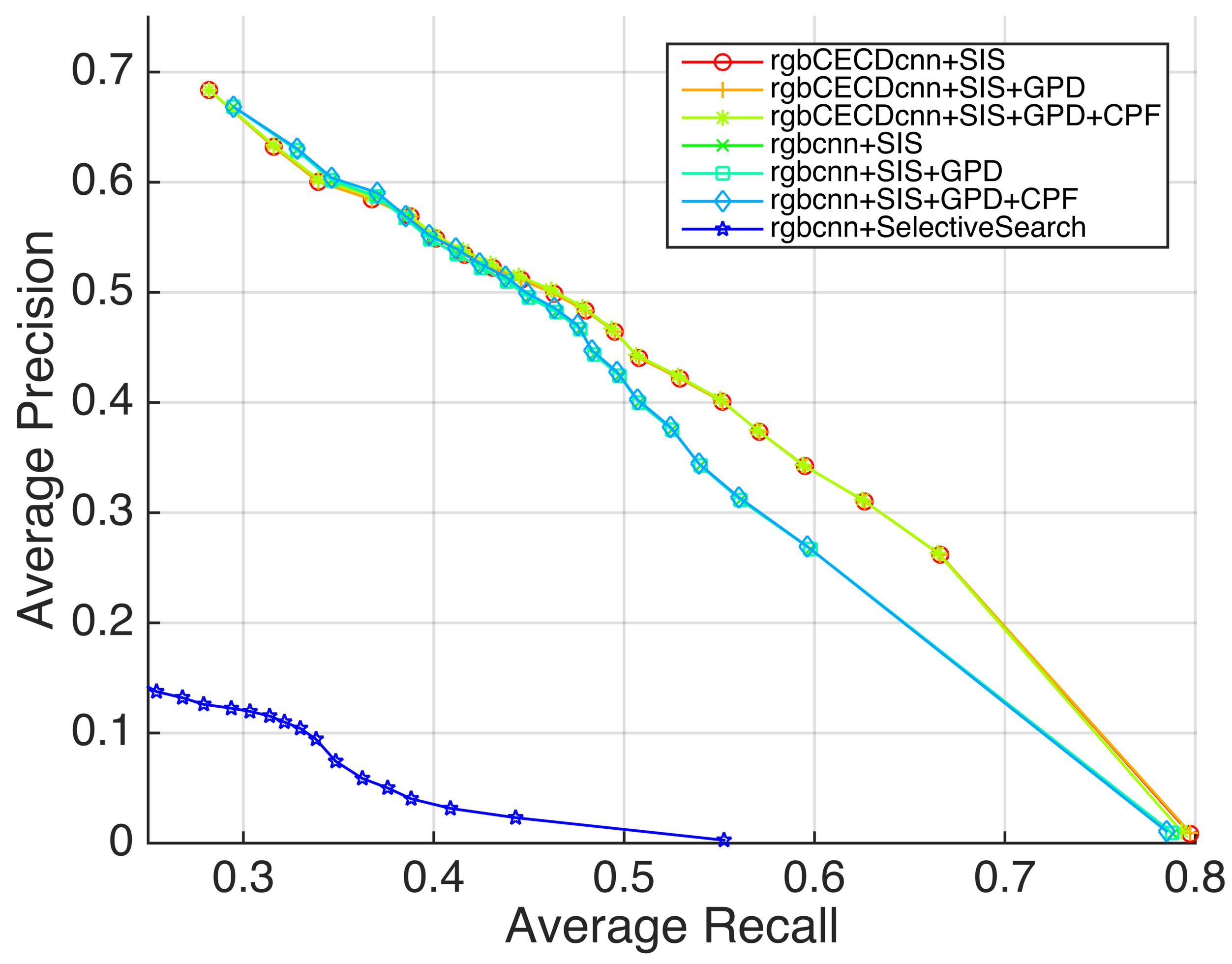


Fig. 5: Evaluation on our ROI method. It shows that using our ROI method does not affect the detection accuracy while the SelectiveSearch method exposes to the low-precision problem. (Our ROI method produces around 5000 windows per image)

3. Effectiveness of adding depth detections

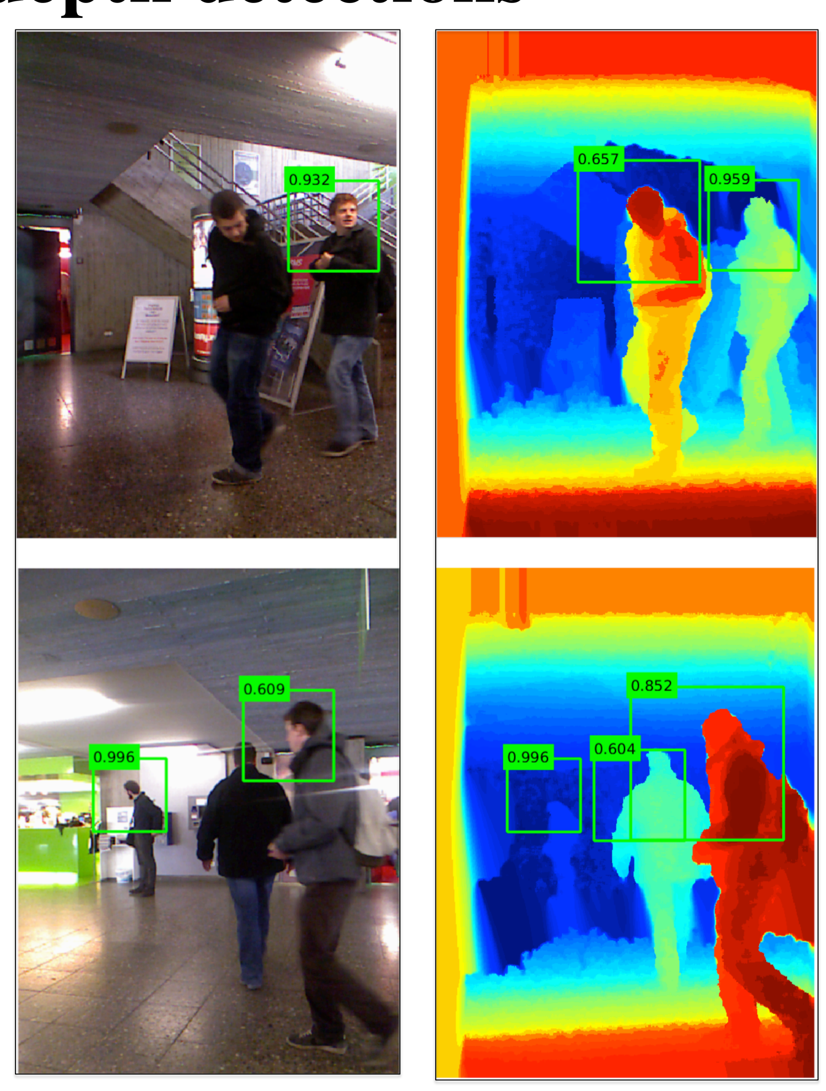


Fig. 6: Detections of rgbcnn (left column) and rgbCECDcnn (right column). Top row shows Depth-CNN is more robust to pose deformations. Bottom row shows depth detections can compensate RGB detections.