Multispectral Deep Neural Networks for Pedestrian Detection

Jingjing Liu[1]
jl1322@cs.rutgers.edu
Shaoting Zhang[2]
szhang16@uncc.edu
Shu Wang[1]
sw498@cs.rutgers.edu
Dimitris N. Metaxas[1]
dnm@cs.rutgers.edu

[1] Department of Computer Science
Rutgers University
Piscataway, NJ, USA

[2] Department of Computer Science
UNC Charlotte
Charlotte, NC, USA

Multispectral pedestrian detection is essential for around-the-clock applications, *e.g.*, surveillance and autonomous driving. In some sense, color and thermal images provide complementary visual information. As shown in Figure 1, thermal images usually present clear silhouettes of human objects [1], but losing fine visual details of human objects (*e.g.* clothing) which can be captured by RGB cameras (depending on external illumination), Nevertheless, except very recent efforts (*e.g.*,[2]), most of previous studies concentrated on detecting pedestrians with color or thermal images only. It is still unknown how color and thermal image channels can be properly fused in DNNs to achieve the best pedestrian detection synergy.

Figure 1: Yellow bounding boxes indicate detection failures with one image channel.

In this paper, we focus on how to make the most of multispectral images (color and thermal) for pedestrian detection. With the recent success of DNNs on generic object detection, it becomes very natural and interesting to exploit the effectiveness of DNNs for multispectral pedestrian detection. We deeply analyze Faster R-CNN [3] for this task and then model it into a convolutional network (ConvNet) fusion problem. We carefully design four distinct ConvNet fusion architectures that integrate two-branch ConvNets on different DNNs stages, *i.e.*, convolutional stages, fully-connected stages, and decision stage, corresponding to information fusion on low level, middle level, high level, and confidence level. All these models outperform the strong baseline detector Faster-RCNN on KAIST multispectral pedestrian dataset (KAIST) [4].

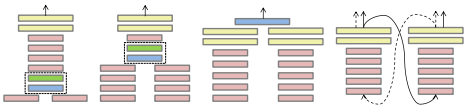

Figure 2: ConvNet fusion models for color and thermal images. From left to right are fusions at low level (Early Fusion), middle level (Halfway Fusion), and high level (Late Fusion), confidence level (Score Fusion), respectively.

We reveal that our Halfway Fusion model – fusion of middle-level convolutional features, provides the best performance on multispectral pedestrian detection. Our Halfway Fusion model significantly reduces the missing rate of baseline method Faster R-CNN by 11%, yielding a 37% overall missing rate on KAIST, which is also 3.5% lower than the other proposed fusion models. We speculate that middle-level convolutional features from color and thermal branches are more compatible in fusion: they contain some semantic meanings and meanwhile do not completely throw all fine visual details.

[1] Y. Socarrás, S Ramos, D. Vázquez, A.M. López, and T. Gevers. Adapting pedestrian detection from synthetic to far infrared images. In *ICCVW*, 2011.

[2] J. Wagner, V Fischer, M. Herman, and S. Behnke. Multispectral pedestrian detection using deep fusion convolutional neural networks. 2016.

[3] S. Ren, K. He, R. Girshick, and J. Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. In *NIPs*, 2015.

[4] S. Hwang, J. Park, N. Kim, Y. Choi, and I.S. Kweon. Multispectral pedestrian detection: Benchmark dataset and baseline. In *CVPR*, 2015.