

Towards Robust Pedestrian Detection with Roadside Millimeter-Wave Infrastructure

Hem Regmi; Vansh Nagpal; Sanjib Sur

Computer Science and Engineering, University of South Carolina, Columbia, SC, USA

hregmi@email.sc.edu; vnagpal@email.sc.edu; sur@cse.sc.edu

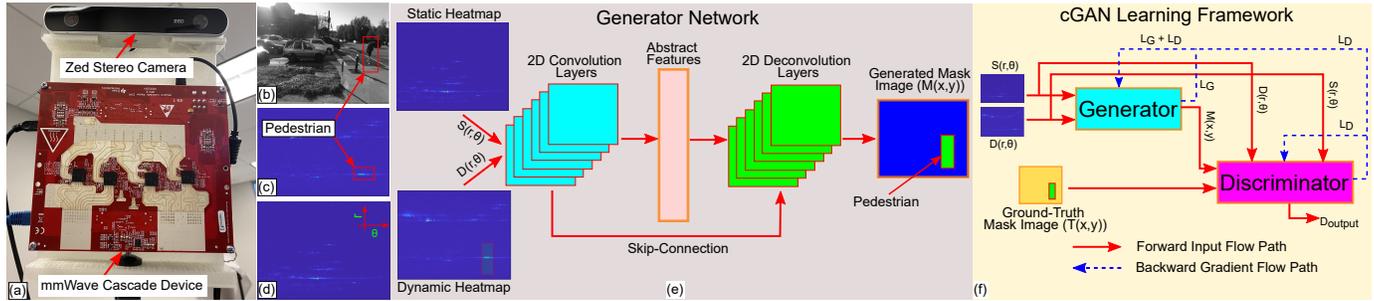


Fig. 1: (a) Experimental setup with a millimeter-wave (mmWave) cascade device and zed stereo camera. (b) Ground-truth image from zed stereo camera. (c–d) Range-Azimuth heatmaps for dynamic objects ($D(r, \theta)$) and static objects ($S(r, \theta)$) of sample mmWave reflection (w/ dynamic pedestrian marked). (e) Generator network architecture of cGAN to generate mask image with pedestrians. (f) Learning framework of cGAN with input and gradient flow paths.

Abstract—We present *MilliPED*, a system that uses a millimeter-wave device to identify pedestrians at traffic intersections and enhance road safety during inclement weather, such as low visibility and heavy rain, when vision cameras are ineffective. We evaluate it with 3000 millimeter-wave reflection samples of pedestrian crossing traffic intersections and show that accurate pedestrian detection is feasible with millimeter-wave devices.

Index Terms—Millimeter-Wave; Convolutional Neural Network; Object Detection.

I. INTRODUCTION

Vehicle-pedestrian collisions are one of the leading causes of fatalities worldwide. Pedestrian accidents frequently occur as a result of poor visibility or adverse weather conditions [1], making it challenging for drivers to spot pedestrians. Traditional methods for reducing pedestrian accidents involve a combination of measures, such as strict enforcement of traffic laws, equipping vehicles with range sensors, and educating pedestrians. Despite these efforts, pedestrian collisions remain a leading cause of traffic accidents. A solution to this problem would be actively detecting pedestrians near traffic intersections and providing this information to nearby vehicles, allowing drivers to take appropriate action and avoid accidents. Traffic enforcement cameras at intersections for monitoring road activity has the potential to improve pedestrian safety through object detection algorithms. However, this technology fails to work in harsh weather conditions such as heavy rain, snow, fog, hailstorms, *etc.* LiDAR-based systems, while more effective in such situations, do not function for close objects and require additional hardware such as mmWave for better detection accuracy.

In contrast, mmWave devices can work effectively under harsh weather conditions such as heavy rain, fog, snow, *etc.*,

and poor visibility because mmWave can penetrate through a fog, and rains effect on mmWave signal strength is insignificant for short range [2]. Additionally, the mmWave device can capture fine-grained details of the target object with the availability of broader bandwidth. However, pedestrian detection with a mmWave device is challenging because most reflections do not reach the receiver due to specularity [3]. Specularity happens because mmWave beamwidth is narrow, and the orientation of reflected signals does not align with the mmWave devices orientation, making essential properties of the target object lost. Moreover, the weak reflectivity of the target objects causes the reflected signals to be attenuated and noisy, making it difficult to detect the object.

We propose *MilliPED*, a system based on Conditional Generative Adversarial Networks (cGAN), for identifying pedestrians in a traffic intersections. *MilliPED* uses dynamic and static heatmaps of millimeter-wave reflections, which carry the wireless signature of moving and stationary objects in the field-of-view of the mmWave device, to detect and locate both moving and stationary pedestrians accurately. *MilliPED* achieves a median Intersection-over-Union (IoU) of 0.67 and a 90th percentile IoU of 0.83, indicating the accurate detection and localization of pedestrians.

II. *MilliPED* SYSTEM DESIGN

To detect and locate pedestrians, we formulate the problem as a generation of mask image of the pedestrians in the field-of-view rather than processing each pedestrian separately. Mask image allows *MilliPED* to handle single and multi-pedestrian cases and preserve the bounding box location of all pedestrians. We generate mask image $M(x, y)$ with pedestrians from static and dynamic millimeter-wave heatmaps

$[S(r, \theta), D(r, \theta)]$ (refer to Figure 1[e]), where $M(x, y)$ is 1 when (x, y) is a pedestrian, and 0 otherwise.

Learning Architecture: In *MilliPED*, we use the generator network (G) (referring to Figure 1[e]) to complete missing regions of pedestrians that occur due to specular reflections, with the help of the ground-truth mask image $T(x, y)$; which is calculated as $M(x, y) = G[S(r, \theta), D(r, \theta)]$. G consists of convolution layers and skip-connection between layers to gain and retain useful abstract features from static and dynamic heatmaps and reconstruct the generated mask image, $M(x, y)$, with a series of deconvolutions [3]. On the other hand, the discriminator network (D) leads the learning process of the generator network by distinguishing between the generated mask image $M(x, y)$ and the ground-truth mask image $T(x, y)$. G attempts to deceive D while D tries to correct G. After going through several back-and-forth processes during training, G eventually learns to produce a mask image, $M(x, y)$, that D believes is the ground-truth. Figure 1(f) displays the learning framework for the cGAN network with the flow of input and gradient.

Data Processing: We build a custom setup with a millimeter-wave cascade device and zed stereo camera to acquire 10 data frames of reflection samples and ground-truth stereo images per second, respectively (see Figure 1[a–d]). Next, we calculate static and dynamic 2D heatmaps from reflection samples with a series of Fourier transforms since it is easier to learn pedestrian mask images from image-like inputs. Finally, we apply object detection algorithm [4] on stereo images to get ground-truth mask images $T(x, y)$, where $T(x, y)$ is 1 for pedestrians and 0 otherwise.

Loss Function & Training: *MilliPED*'s cGAN uses a combination of Mean Squared Error (MSE) and Binary Cross Entropy (BCE) as the loss function to guide the direction of the network and learning process. We calculate generator loss L_G as $\text{MSE}[M(x, y), T(x, y)]$, and the discriminator loss L_D as $\text{BCE}[D_{\text{output}}, 1 \text{ or } 0]$, respectively. D_{output} is the boolean output of D, and we use 1 for $T(x, y)$ and 0 for $M(x, y)$. To train G, we accumulate both L_G and L_D , while D only uses L_D . We update network parameters with an Adam optimizer, and the training continues until the model does not show any improvement for 30 consecutive epochs, with a maximum of 1200 epochs. Training the network with 15000 training samples on an NVIDIA RTX A6000 GPU on the host server requires around 24 hours.

III. PRELIMINARY RESULTS

Post-training, we use the trained G to generate mask images for 3000 test samples collected from various traffic scenarios during busy work hours, including pedestrians crossing the road. Figure 2(a)[i–ii] demonstrates the successful generation of mask images for pedestrians with one and two people in view. We notice that bounding box locations with two people are slightly off due to the close distance between them, making it difficult for the millimeter-wave device to distinguish. We

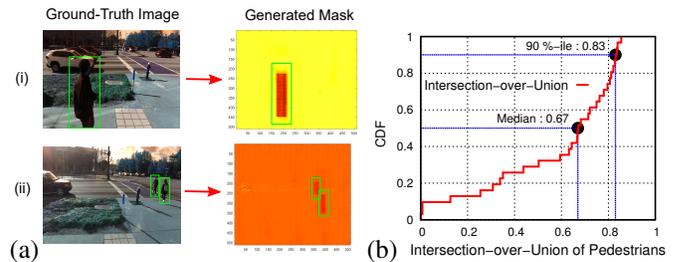


Fig. 2: (a) Generated mask images by *MilliPED* with for single and double pedestrians. (b) CDF of Intersection-over-Union (IoU) between ground-truth and generated mask images of pedestrians for 3000 test samples.

quantify *MilliPED*'s pedestrian detection and localization with Intersection-over-Union (IoU) that provides overlap between generated mask image $M(x, y)$ and ground-truth mask image $T(x, y)$. Figure 2(b) shows the cumulative distribution of the IoUs across all test samples and *MilliPED* achieve a median of 0.67 and a 90th percentile of 0.83, indicating accurate detection and localization of pedestrians.

IV. CONCLUSION & FUTURE WORKS

In this work, we present and assess a platform for detecting pedestrians in traffic intersections using millimeter-wave devices. Our results indicate that pedestrian detection at traffic intersections is possible and could help to reduce the chance of collisions caused by unfavorable weather and poor visibility. However, installing all these millimeter-wave devices are labor-intensive and expensive. We anticipate that picocells will be set up at the traffic intersections to enhance existing network infrastructures [5], so it would be beneficial to use picocells for network communication and pedestrian detection by sensing the environment to detect the surrounding objects when bandwidth is available, with limited interference to the picocell's communication duties. In the future, we plan to improve pedestrian detection accuracy by collecting more data, accounting for various scenarios, and developing the real-time system to use millimeter-wave devices for communication and pedestrian detection through context-aware deep learning models.

ACKNOWLEDGMENTS

We sincerely thank the reviewers for their comments. This work is partially supported by the NSF under grants CAREER-2144505, CNS-1910853, and MRI-2018966.

REFERENCES

- [1] Subasish Das, et al., "Factor association with multiple correspondence analysis in vehicle–pedestrian crashes," *Transportation Research Record*, vol. 2519, no. 1, 2015.
- [2] Yibo Zhu, et al., "Demystifying 60GHz Outdoor Picocells," in *Proceedings of the 20th Annual International Conference on Mobile Computing and Networking*, 2014.
- [3] H. Regmi, et al., "SquiggleMilli: Approximating SAR Imaging on Mobile Millimeter-Wave Devices," *Proc. of ACM IMWUT*, 2021.
- [4] Yang, Guanhao, et al., "Face mask recognition system with YOLOV5 based on image recognition," in *2020 IEEE 6th International Conference on Computer and Communications (ICCC)*, 2020.
- [5] Colin Blackman, et al., "5G Deployment: State of play in Europe, USA and Asia," *Policy Commons*, 2019.