Hazard Detection in Supermarkets using Deep Learning on the Edge

M. G. Sarwar Murshed
Clarkson University

Edward Verenich
Air Force Research Laboratory
Clarkson University

Conrad Gende
Clarkson University

James J. Carroll
Clarkson University

Nazar Khan
Punjab University
College of Information Technology

Faraz Hussain
Clarkson University

Abstract

Supermarkets need to ensure a safe environment for shoppers and employees. Slips, trips, and falls can result in injuries that have a physical as well as financial cost. Timely detection of hazardous conditions such as spilled liquids or fallen items on supermarket floors can reduce the chances of serious injuries. Supermarket owners have to appoint permanent cleaners but accidents are common. Therefore, there is increasing interest from industry, especially from supermarkets, to do repetitive & dangerous work using robots instead of humans.

In recent years, deep learning (DL) techniques have been widely applied in a variety of domains (e.g. for mobile robots and autonomous vehicles) for making real-time decisions based on the surrounding environment [1]. However, building deep learning models not only needs a lot of computational power and memory but also dedicated hardware after deployment for reasonably fast inference. These constraints have traditionally inhibited the deployment of DL models on resource-scarce devices.

We present EdgeLite, a novel, lightweight deep learning model for easy deployment and inference on resource-constrained devices, allowing those devices to aid faster decision-making. EdgeLite can process supermarket image data on edge devices in order to detect potential floor hazards. The early detection of hazards such as spills and debris, can prevent potentially serious accidents. On a hazard detection dataset that we developed, EdgeLite, when deployed on two edge devices (viz. Raspberry Pi and the Coral Dev Board), outperformed six state-of-the art object detection models in terms of accuracy while having comparable memory usage and inference time.

EdgeLite has a CNN-based architecture with 19 layers, not counting the pooling layers. In order to extract features at different scales, we used filters with multiple sizes that operate on the same level. The different types of filters used were of size 1×1, 3×3 and 5×5. To make the network computationally cheaper, 1×1 convolutions were used to reduce the input channel depth and an extra 1×1 convolution was used before the 3×3 and 5×5 convolutions.

EdgeLite suffers from the vanishing gradient problem, which makes it difficult to train the the network. To address this problem, we added two auxiliary layers to the middle of the network which prevent the middle part of the network from dying out, and also have a regularizing effect. These layers are only used during training and discarded during inference. Thus, the deployed model is not burdened by these extra layers. Our CNN architecture consists of convolution, max-pooling, avg-pooling and EdgeLite layers. EdgeLite layers are incorporated into CNNs as a way of reducing computational expense through a dimensionality reduction with stacked 1×1 convolutions. Multiple kernel filter sizes are used in this layer and an extra 1 convolution is added whenever 3×3 and 5×5 layers are used. All the kernels are ordered to operate on the same level sequentially. A max-pooling is performed in this layer and the resulting outputs are concatenated, and then sent to the next layer. EdgeLite's architecture is inspired by InceptionNet [2] but the number of layers and size of the kernels is reduced to make it suitable for resource-constrained devices.

We built an original real-world dataset of images showing hazards in supermarket floors. This dataset contains supermarket images labeled either as having a hazardous floor or not. In addition to 1180 manually collected images, we also added synthetic images to enrich our dataset.

We trained EdgeLite on our hazard dataset and got the best model by training the network using the Adam optimizer with a momentum of 0.9 and batch size of 32. The learning rate was 0.002 with a decay of 0.00004 on the model weights. On our test set, EdgeLite outperformed other state-of-the-art models by achieving 92.37% accuracy.

References

[1] Antonio Brunetti, Domenico Buongiorno, Gianpaolo Francesco Trotta, and Vitoantonio Bevilacqua. Computer vision and deep learning techniques for pedestrian detection and tracking: A survey. *Neurocomputing*, 300:17 – 33, 2018.

[2] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–9, 2014.