Available online at www.sciencedirect.com

ScienceDirect

Procedia Manufacturing 00 (2019) 000–000

Procedia MANUFACTURING

www.elsevier.com/locate/procedia

10th Conference on Learning Factories, CLF2020

PPE Compliance Detection using Artificial Intelligence in Learning Factories

Balamurugan Balakreshnan[a], Grant Richards[b], Gaurav Nanda[b], Huachao Mao[b], Ragu Athinarayanan[b,]*, Joseph Zaccaria[c]

[a]*Microsoft Corporation, 1200 W Pine Creek Ct, Oak Creek, WI 53154, USA*
[b]*School of Engineering Technology, Purdue University, 401 N Grant St, West Lafayette, IN 47907, USA*
[c]*Rockwell Automation, 1201 S 2nd St, Milwaukee, WI 53204, USA*

Abstract

This project demonstrates the application of Artificial Intelligence (AI) and machine vision for the identification of Personal Protective Equipment (PPE), particularly safety glasses in zones of the Learning Factory, where safety risks exist. The objective is to design and implement an automated system for ensuring the safety of personnel when they are in the vicinity of machinery that presents potential risks to the eyes. Microsoft Azure Custom Vision AI and Intelligent AI Services, in conjunction with low-cost vision devices with lightweight onboard AI capability, provide a platform for a deep learning neural network model using publicly available images under the Creative Commons License. A combination of cloud-based and on-premises AI is used in this proof of concept system to provide a real-time vision-based safety system capable of detecting and recording potential safety breaches, promoting compliance, and ultimately preventing accidents before they happen. This system can be used to initiate different control actions in the event of safety violations and can detect multiple forms of protective wear. The flexibility of the system offers multiple benefits to learning factories and manufacturing organizations such as improved user safety, reduced insurance costs, and better detection and recording of safety violations. The hybrid AI architecture approach allows for flexibility in training and deployment based on the capability of local computing resources.

Keywords: Artificial Intelligence; Safety; Automation;

* Corresponding author. Tel.: +1-765-494-0448; fax: +1-765-494-6219.
E-mail address: rathinar@purdue.edu

1. Introduction

In the United States, the Occupational Safety and Health Administration (OSHA) enforces workplace standards to ensure safe and healthful working conditions. Liabilities associated with workplace safety in the US costs more than 360B USD annually. One of the many safety precautions enforced is the requirement for protective eyewear to reduce eye injuries in the workplace. In this work, we implement a compliance detection system for protective eyewear using Internet of Things (IoT) devices and Artificial Intelligence (AI) in laboratories and future applications in our Learning Factory. The use of IoT and AI has the potential for improving workplace safety by offering an automated approach to monitoring and mitigating injuries before they happen [1,2,3]. This system will use algorithmic decision-making and does not require human involvement in decision-making.

2. Artificial Intelligence & Azure Custom Vision

2.1. Hardware and software

The AI-based protective eyewear detection system uses commonly available hardware and software components to allow for ease of adoption with cloud computing support allowing for on-demand scalability. The primary hardware component in the system is an IoT camera with onboard AI computing capability, developed in a collaboration between Microsoft and Qualcomm [4]. The low-power eight-megapixel camera utilizes a Linux OS containing all software and security features required for edge deployment through a wireless link. A custom low-power graphics processor provides the capability of running AI object detection algorithms directly on the device. The camera utilizes an open-source development model with all specifications and supporting development details available on a GitHub repository [5].

The prototype platform is a hybrid system that utilizes a combination of on-premises and cloud-based computation. Microsoft's Azure platform provides cloud-based computing and storage for the system and provides a range of prebuilt tools designed for ease of deployment, integration, and modification. Cloud-based services allow for the creation and training of the detection model and also the capture and long-term storage of detection events. This platform can also communicate with edge devices to indicate when specific event conditions which warrant response are detected.

2.2. System architecture

The digital platform for this project leverages the benefits of a hybrid on-premise/cloud architecture to provide for maximum scalability (Fig. 1). The reference architecture allows for on-demand expansion of the system and supporting computing and storage capacities as well as allowing additional devices and zones to be configured and deployed with minimal additional programming. This platform flexibility allows operators to manage costs associated with infrastructure.

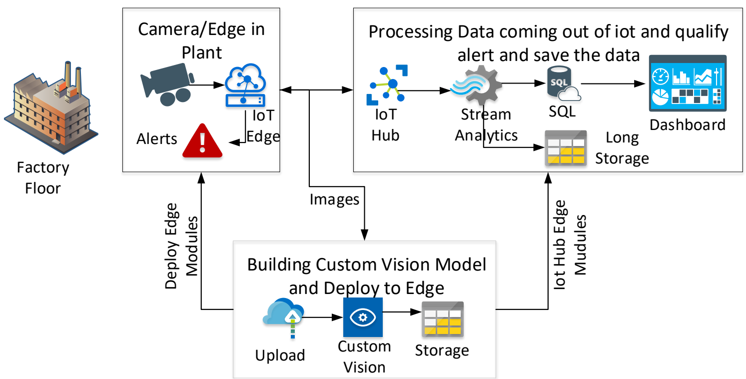


Fig. 1. Digital architecture of experimental system

2.2.1. Device (on-premise) services

The Azure AI Vision camera used in this project was oriented to capture an unobstructed field of view of the targeted area of a laboratory that supports the Learning Factory. The camera connects to a private local network over a Wi-Fi link. The local network has internet connectivity managed through a security appliance and gateway. Azure IoT Edge is a prebuilt on-device service that establishes a secure communications pathway between the camera and the cloud through Azure IoT Hub. The camera is paired with the Azure Hub using an on-board, browser-based provisioning tool. The camera's onboard vision AI uses a TensorFlow graph model (SqueezeNet [6]), a JSON configuration file for the Qualcomm Vision Intelligence GPU [7], and test files comprised of lists of labels and objects to identify. The camera also provides an HDMI port and an RTSP stream (TCP port 3000) to provide source video monitoring and capture capability. These video streams also include overlays of tagged detection boxed to indicate image elements used for object identification. The prototype system refreshes at 12-second intervals.

2.2.2. Azure (on-premise) services

Azure IoT Hub is a cloud-based service to coordinate and aggregate communication between edge devices and Azure. IoT Hub provides a dashboard for real-time monitoring of all connected devices and hosts a digital twin of each connected edge device, which allows for the migration of models between the cloud and locally deployed devices. This model updates through the transfer of a zip file containing the elements previously detailed in the camera discussion. Device data is aggregated at the IoT Hub and passes to the Stream Analytics service, which serves as the logic and decision processing center of the system. The Stream Analytics service provides real-time analysis of aggregated object detection alerts and simultaneously routes alerts to blob storage in raw format. The service also structures data and passes it to an SQL database in a structured format that can be accessed by a range of other services and software packages. A custom dashboard provides a real-time display of detected objects and supporting statistics.

2.2.3. Model training and development

The AI vision model is trained in the cloud using the Custom Vision AI service. This service provides a web-based interface for uploading and tagging images to be used for training. Model training is customizable for a range of scenarios (this system uses a compact model). Cloud resources are used for all training activities to minimize the requirement for extensive on-site computing capability. After training, a zip file containing all supporting information for the model is created and transferred to the local device through its cloud-hosted digital twin located on Azure IoT Hub.

2.2.4. AI vision model

The AI vision model in the test platform trains with a combination of images that feature persons wearing and not wearing safety glasses. Standard classifications (person, safety glasses) are used to tag specific regions of each of the training images in the Custom Vision dashboard. After completing all image tagging, the model trains using cloud computing resources which employ an open-source TensorFlow-based object detection algorithm to create a compact neural model that is optimized for performance on the Qualcomm CPU. This model and a supporting text file containing classification tags are deployed as a zip file to the vision device through its digital twin.

3. Experimental framework

The experimental setup used in the laboratory (Fig. 2) uses the AI vision enabled camera for two purposes; to evaluate how well the camera and model detect protective eyewear, and to collect image data for training the CustomVision AI model. The AI model trains with online images of people wearing safety glasses using the method detailed earlier previous section and deploys to the camera through its digital twin. The model is then evaluated for its ability to detect personnel with and without protective eyewear.

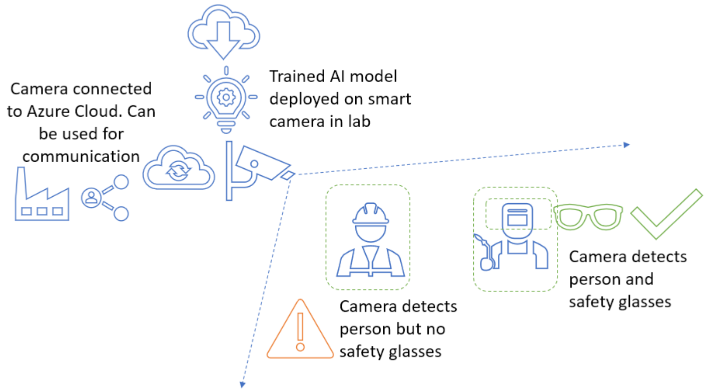


Fig. 2. Experimental framework of the AI vision system

The first stage of development is image data collection and tagging the images for training the custom vision model. For this purpose, three of the co-authors acted as lab workers and recorded their images with and without safety glasses. The camera's live stream (RTSP) is captured using VLC media player and parsed into images.

The second stage is the creation of the Custom Vision model using the parsed images obtained from the video feed. Imaging tagging is performed by selecting a square border area around objects in the images and assigning tags. After all the images are tagged, the Custom Vision model trains with the tagged images and the model's predicted performance, which is calculated and recorded. The trained AI model is then downloaded from the Custom Vision tool and deployed to the on-premises camera by attaching the new model to the camera's cloud-based digital twin. Finally, the new AI model is evaluated to determine its ability to detect people, faces and protective eyewear accurately. Performance results appear in the next section.

The experimental AI models were evaluated on precision and recall [8]. Precision indicates the fraction of identified classifications that were correct. For example, if the model identified 100 images as dogs, and 99 of them were actually of dogs, then the precision would be 99%. Recall indicates the fraction of actual classifications correctly identified. For example, if there were 100 images of apples and the model identified 80 as apples, the recall would be 80%. Finally, mean Average Precision (mAP) indicates the overall detector performance across all tags.

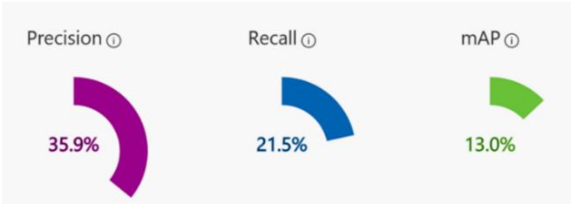
When interpreting prediction calls with a high probability threshold, the system tends to return results with high precision at the expense of recall—the detected classifications are correct, but many remain undetected. A low probability threshold produced the opposite result—most of the actual classifications are detected, but there are more false positives within that set. With this in mind, the probability threshold was according to the specific needs of the project.

4. Results

When receiving prediction results on the client-side, the same probability threshold values are used. Additionally, the minimum percentage of overlap between predicted bounding boxes and ground truth boxes to be considered as a correct prediction is set as 30%. The model output is then evaluated using mean average precision (mAP) with a higher value preferred. The model produces a bounding box on each detected object from the list of predefined classes and a recall curve is computed under the PR curve. We tested this model based on how it is capable of detecting safety glass in the real work environment.

4.1. Base model performance

One advantage of the proposed system is the ability to realize iterative performance improvement. This work explores two iterations of the supporting AI vision model. In the first iteration, model training is performed solely on images obtained through a Google search. The open-source informed AI vision model serves as a base reference. As expected, the performance of this model was poor and wholly unsuitable for practical application in a safety system. This base model has significant issues detecting safety glasses in general and is incapable of detecting glasses beyond an impractically small focal distance.



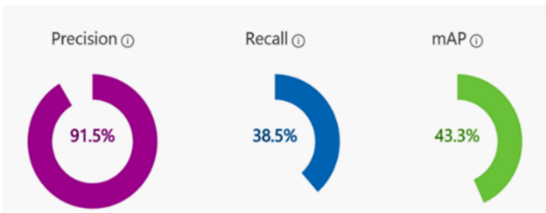
Performance Per Tag

Tag	Precision	Recall	A.P.	Image Count
safety_glasses	37.5%	9.7%	9.0%	117
face	50.0%	11.1%	11.0%	134
person	33.3%	37.0%	18.9%	140

Fig. 3. Performance of the original (open source) AI vision model

4.2. Improved model performance

The poor performance of the base model is likely due to considerable variations in the images obtained from Google search in terms of the image background, prominent colors, and relative size of the person in the images. Due to the poor prediction performance of the base model, the trained model deployed on the AI cameras in the lab environment was not able to detect safety glasses from its live feed. In order to address these shortcomings due to inconsistency between training and test images, an improved model was developed by training only on images obtained from the lab environment. New images were acquired directly with a camera mounted in the laboratory, tagged the model was retrained using the only locally acquired images. Results show a 155% improvement in overall model precision, 79.1% improvement in recall, and 233% improvement in the mean average precision compared to the base model. Although there is improvement in the overall performance, for applications involving PPE compliance detection, these numbers are still low.



Performance Per Tag

Tag	Precision	Recall	A.P.	Image Count
safety_glasses	50.0%	3.5%	11.7%	278
face	66.7%	1.9%	26.9%	488
person	93.6%	90.4%	91.2%	525

Fig. 4. Performance of the improved (local source) AI vision model

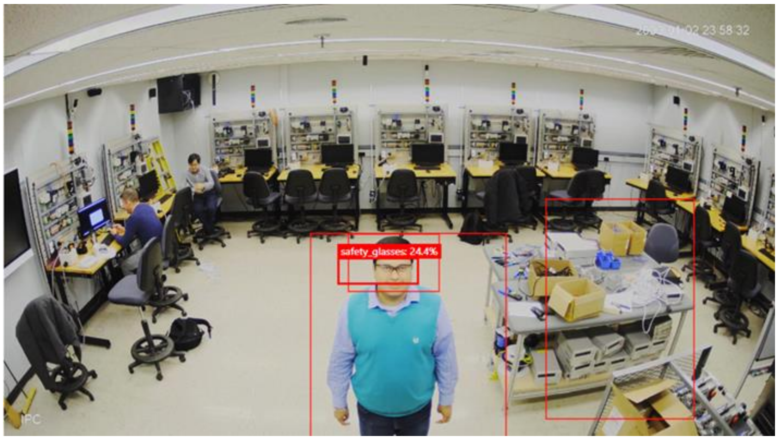


Fig. 5. Detection of safety glasses within a very complex scene

5. Results

The significant improvement between the base model and the improved model indicates that there is high likelihood to obtain performance improvements and expand operational capability through exploration of factors like background, light and depth of object and large training sets. The low-power AI enabled camera driving this experimental system holds promise in being capable of providing AI vision monitoring of protective equipment compliance. The affordability of the device and the accessibility and scalability of the supporting cloud-based computing architecture help reduce barriers to entry and adoption. It is highly expected that the limited on-board computing capability of the low-power platform will limit the ability to detect and process in multiple aspects with respect to more robust and powerful platforms with higher energy use and operational costs. Future efforts will focus on determining these operational limits and exploring appropriate applications and limits for deployment in the Learning Factory, such as distinguishing between safety glasses and regular glasses [9,10], which is relevant in a real-life work environment.

References

[1] D. Podgórski, K. Majchrzycka, A. Dąbrowska, G. Gralewicz, M. Okrasa, Towards a conceptual framework of OSH risk management in smart working environments based on smart PPE, ambient intelligence and the Internet of Things technologies, International Journal of Occupational Safety and Ergonomics 23 (1) (2017) 1–20. doi:10.1080/10803548.2016.1214431.

[2] P. V. Moore, OSH and the Future of Work: Benefits and Risks of Artificial Intelligence Tools in Workplaces, in: V. G. Duffy (Ed.), Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management. Human Body and Motion, Springer International Publishing, 2019, pp. 292–315.

[3] L. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu, M. Pietikäinen, Deep Learning for Generic Object Detection: A Survey, International Journal of Computer Vision (Oct 2019). doi:10.1007/s11263- 019-01247-4.

[4] A Smart Camera for the Intelligent Edge (2020). URL https://azure.github.io/Vision-AI-DevKit-Pages/

[5] Workplace Safety Identification (2020). URL https://azure.github.io/Vision-AI-DevKit-Pages/docs/community_project02/

[6] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, K. Keutzer, SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size, arXiv (2016).

[7] Vision Intelligence 400 Platform (2018). URL https://www.qualcomm.com/products/vision-intelligence-400-platform

[8] B. Karasulu, Review and Evaluation of Well-known Methods for Moving Object Detection and Tracking in Videos, Journal of Aeronautics and Space Technologies 4 (4) (2010) 11–22.

[9] B. Karasulu, Review and Evaluation of Well-known Methods for Moving Object Detection and Tracking in Videos, Journal of Aeronautics and Space Technologies 4 (4) (2010) 11–22.

[10] Z. Jing, R. Mariani, J. Wang, Glasses Detection for Face Recognition Using Bayes Rules, in: T. Tan,Y. Shi, W. Gao (Eds.), Advances in Multimodal Interfaces — ICMI 2000, Springer Berlin Heidelberg, Berlin, Heidelberg, 2000, pp. 127–134