**PERSONAL PROTECTIVE EQUIPMENT (PPE) DETECTION**

19MIA1028

NITHYA SHARMA

19MIA1083

K NIHARIKA SAMYUKTHA

**ABSTRACT: Construction safety is a matter of great concern for practitioners and researchers worldwide. Even after risk assessments have been conducted and adequate controls have been implemented, workers are still subject to safety hazards in construction work environments. The need for personal protective equipment (PPE) is important in this context. Automatic and real-time detection of the non-compliance of workers in using PPE is an important concern. This study developed a framework to sense in real-time, the safety compliance of construction workers with respect to PPE, which is intended to be integrated into the safety workflow of an organization. The study makes use of the Convolutional Neural Networks model, which was developed by applying transfer learning to a base version of the YOLOv3 deep learning network.**

1. **INTRODUCTION:**

Construction usually involves high-risk activities requiring workers to operate at dangerous places and be exposed to risk. Based on the United States’ Bureau of Labor Statistics, the fatalities number increases gradually from 985 in 2015 to 1038 in 2018, with an increase of 2% every year. In China, 840 workers died during construction activities in 2018, and 52.2% of them were caused by falling from a high place. Similarly, according to the UK Health and Safety Executive (HSE), 147 workers suffered from fatal injuries in the UK in 2018/2019, where falling from a high place is the most significant kind of fatal accident. However, the majority of injuries, illness and fatalities could be avoided if workers wear suitable PPE like helmets, vest and so on. The helmet is an essential piece of PPE, which protects construction workers by resisting objects and absorbing shock from direct blows to the head by objects.

The main goal of PPE detection is to measure health and safety compliance to improve construction safety. Wearing helmets would reduce the injuries and even fatalities when meeting accidents. Meanwhile, another necessary PPE, the vest, is also required to be worn on construction sites for increasing visibility. The vest with flash lines would help others locate construction workers and avoid accidents, particularly in poor weather, like rainy and foggy days. Another aim is to understand the activities of works and optimize management. The helmet colors present different roles in different countries.

1. **LITERATURE SURVEY**

In computer vision, image classification is defined as the problem of assigning a single class (single-label classification) or multiple classes (multi-label classification) to an entire image. With the increase in quantity and quality of photos and videos taken from construction sites, more attention is being drawn to streamlining the process of automatically extracting content from digital imagery through image classification and object detection. For example, Zou and Kim (2007) utilized HSV (hue, saturation, and value) color space of images to identify excavators in construction photos. In particular, they used the threshold of saturation as a feature to distinguish a relatively colorful excavator object from the dark soil or white snow background. Brilakis et al. (2005), and Brilakis and Soibelman (2008) proposed a method to detect shapes in an image and identify corresponding material types (e.g., steel or concrete) within the texture of the detected shape region. Wu et al. (2009) employed Canny edge detection and watershed transformation methods to detect the edges of an object (e.g., columns in an image), and applied object reconstruction to locate and quantify objects (e.g., number of columns). Kim et al. (2016) used scene-parsing and label transfer to match a target image with a number of labeled images, find candidate images that match more closely, and transfer labels from candidate images to the target image.

Recent work has also utilized machine learning (ML) algorithms to automate the process of object recognition in construction site imagery. For example, Chi and Caldas (2011) used naïve Bayes (NB), and neural network (NN) classifiers to detect workers, loaders, and backhoes. Son et al. (2014) used a voting-based ensemble classifier combining several base classifiers such as support vector machine (SVM), NN, NB, decision tree, logistic regression, and k-nearest neighbor (KNN), to identify construction materials (e.g., concrete, steel, and wood) in an image. Dimitrov and Golparvar-Fard (2014), and Han and Golparvar-Fard (2015) used one-vs-all multi-class SVM to classify major construction materials (around 20 types).

The majority of the aforementioned methodologies, however, requires the extraction of handcrafted image features that are particularly relevant to the given classes (Kolar et al., 2018). However, for content-rich imagery such as construction photos that contain a large number of highly diverse objects or cover a large visual field under a variety of environmental conditions (e.g., lighting, landscape, etc.), automatic feature extraction methods such as CNN and histogram of oriented gradients (HOG) are more advantageous because of their ability to self-learn features from a given dataset (Kolar et al., 2018). While HOG poorly performs when high-dimensional features are simultaneously considered for image classification, CNN achieves outstanding results in this task (Kolar et al., 2018) by overcoming the challenge of enormous computational power demanded by traditional NN (LeCun et al., 1998). A good example of CNN can be found in LeCun et al. (1998) which involves recognizing handwritten digits in an image. Other recent studies include but are not limited to classifying 1.2 million images (ImageNet dataset) into 1,000 different classes (various everyday objects and animals such as French fries, printer, umbrella, dog) (Krizhevsky et al., 2012, Simonyan and Zisserman, 2014).

Within the construction domain, there are several studies where CNN has been used for visual analysis of images and videos, mostly for construction safety. For example, Kolar et al. (2018) used CNN to detect safety guardrails in site photos. Siddula et al. (2016) combined the Gaussian mixture model (GMM) with CNN to detect objects of interest in images taken from roof construction sites. Ding et al. (2018) integrated the long short-term memory (LSTM) model with CNN to recognize unsafe behaviors of construction workers (e.g., climbing a ladder) in video frames. More recently, Luo et al. (2018) proposed a method that uses Region-based CNN (R-CNN) to detect 22 classes of construction-related objects and predict construction activities based on the spatial relevance between the detected objects. However, a majority of these object detection (i.e., classifying and localizing objects) algorithms are computationally intensive and require heavy processing power to perform analyses on high volumes of visual data. Moreover, the amount of collected visual data from construction sites is increasing as more contractors rely on reality capture technologies with mobile connectivity such as smartphones, tablet computer, and camera-equipped drones (Ham and Kamari, 2019). For example, a study by Han and Golparvar-Fard (2017) reported that more than 400,000 images were collected during the lifecycle of a 750,000-sf commercial construction project. Therefore, there is a dire need for fast and automated image filtering methods to use data transmission and storage capacities more efficiently. A recent example of existing studies in this direction by Ham and Kamari (2019) uses pixel-by-pixel semantic image segmentation (i.e., assigning a class to each pixel) to train a model to detect construction-related objects. However, manually annotating a large volume of images at pixel level is a tedious task, requiring substantial amount of time, cost, and human resources and, therefore, might daunt the usability of this method in real practice (Wei et al., 2016). In contrast, image-level annotation is a more practical approach as it requires assigning single or multiple classes to the entire image and, thus, reduces the time and effort to perform manual labeling (Wei et al., 2016). Therefore, considering the advantages of deep learning and imagelevel annotation, and informed by the need for faster algorithms to process and filter large volumes of visual data for rapid onsite documentation, this research aims at developing a CNN-based methodology to annotate construction site imagery with predefined labels (e.g., building, equipment, and worker). Compared to R-CNN algorithms, the proposed model can be applied in real-time on low-powered mobile devices, i.e., smartphones or drones.

1. **OBJECTIVE**

The goal of this project is to provide a scalable deep learning system for computer vision to identify PPE compliance. The study's goal of detecting hard helmets and safety jackets on building sites was accomplished. The related sub-objectives of this study included demonstrating how the Deep Learning based CV algorithm works to identify site-specific hard-hat and safety-jacket compliance. According to the goal of the site's safety monitoring, the study will also show how transfer learning may be used to expand the trained CV algorithm to include additional classes.

The model that was thus constructed and trained showed robustness in terms of detecting compliance in a variety of scenarios and carried out detection on video streams from building sites in close to real-time. Convolutional neural networks, a subset of deep learning computer vision methods, are used in the study to process the data, create, train, and evaluate the model in order to achieve this goal.

1. **METHODOLOGY**

For the dataset, we tried extracting the frames from a video. We extracted the frames and converted them into weights (.hdf5 file). Since, no video of construction sites were properly available, we have taken a crowd-sourced dataset to carry out the research. The data was collected from a dataset called Pictor-ppe. This section presents generic frameworks for verifying safety compliance for multiple PPE from visual data (image or video). In

particular, three different DL-based approaches are proposed to perform

the verification. Although the proposed techniques are

designed to work for any number (≥1) and type of PPE (e.g., hard hat,

safety vest, gloves, safety goggles, and steel toe shoes), for the scope of

this study, the technical discussions, data analysis, and validation are

particularly conducted for two types of PPE, i.e., hard hat (referred to as

hat) and safety vest (referred to as vest).

1. **DATASET**

The dataset (named Pictor-v3) contains 774 crowd-sourced and 698 web-mined images. Crowd-sourced and web-mined images contain 2,496 and 2,230 instances of workers, respectively. A brief statistics of the dataset is shown in the following figure.

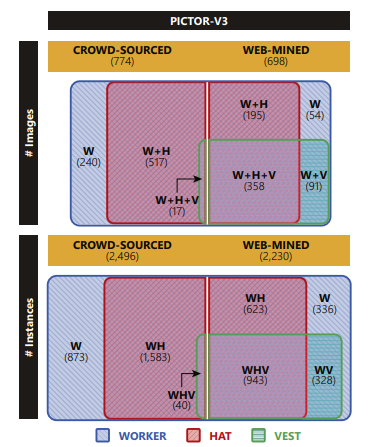


Figure 1: Statistics of the dataset

1. **IMPLEMENTATION**
   1. **FRAME EXTRACTION**

Frame extraction is done using Open-CV in python.

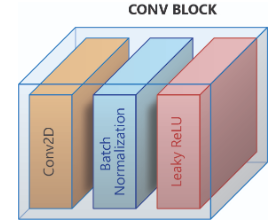
STEPS:

1. Opening the video file using cv2.VideoCapture().
2. Reading the video frame by frame.
3. Saving each frame in a folder using cv2.imwrite(). If there is no folder available, then we create a new folder and save the frames in it.
4. Release the VideoCapture and destroy all windows.
   1. **EXPLANATION OF MODEL**
      1. **YOLO Model**

YOLOv3 (You Only Look Once, Version 3) is a real-time object detection algorithm that identifies specific objects in videos, live feeds, or images. YOLO is a Convolutional Neural Network (CNN) for performing object detection in real-time. CNNs are classifier-based systems that can process input images as structured arrays of data and recognize patterns between them. YOLO has the advantage of being much faster than other networks and still maintains accuracy.

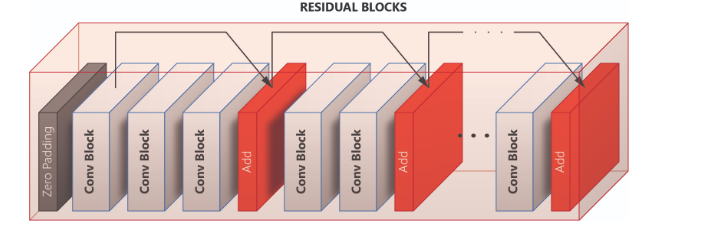
**YOLO Architecture**

* **CONVOLUTION BLOCK**



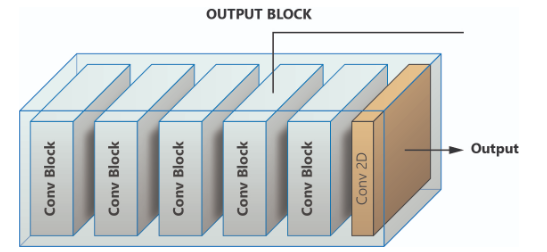
**Figure 2: Convolutional Block of YOLO Model**

* **RESIDUAL BLOCK**



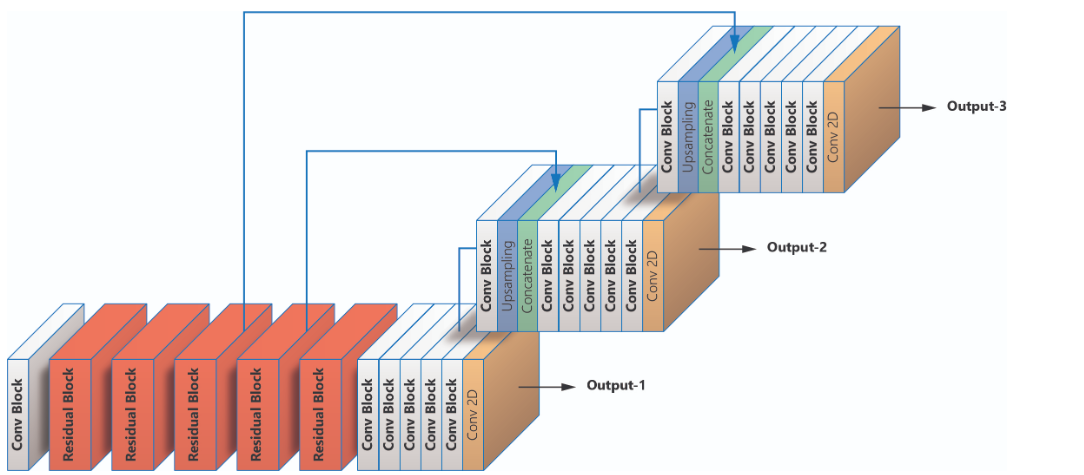
**Figure 3: Residual Block of YOLO Model**

* **OUTPUT BLOCK**



**Figure 4: Output Block of YOLO Model**

* **THE COMPLETE MODEL**



**Figure 5: The complete YOLO Model**

* + 1. **Overview of the three different approaches**

In Approach-1, a YOLO-v3-based model is trained and tested to detect worker (W) and different PPE types, e.g., hat (H), and vest (V), individually. Next, a verification algorithm is applied to check if each worker is wearing the PPE properly. This algorithm, particularly, checks all the workers and classifies each worker into one of the four classes: worker wearing no hat and vest (W), worker wearing only hat (WH), worker wearing only vest (WV), and worker wearing both hat and vest (WHV). In contrast, in Approach-2, a YOLO-v3-based model is designed to localize workers and directly classify them based on their PPE attire, i.e., W, WH, WV, and WHV. Finally, Approach-3 first uses a YOLO-v3-based model to detect only workers (regardless of the PPE attire), then crops parts of the image that contain workers, and finally, applies a CNN-based classifier to each cropped image to classify it into one of the possible classes, i.e., W, WH, WV, and WHV.

1. **CONCLUSION & FUTURE SCOPE**

In this project we have used deep learning-based computer vision algorithms in the automated detection of the key processes that sustain construction safety and on-site management. Using YOLOv3, a state of art object detection algorithm, this project demonstrates how safety compliance can be automatically detected by using a trained model to examine data from sites. We used 3 approaches in this project.

Approach 1 – Detect only workers.

Approach 2 - Detect Hat, Vest and workers separately.

Approach 3 - Detect worker, worker with hat, worker with vest, worker with hat and vest.

In future we try to detect PPE kit using live videos, webcam etc. We will also try to build an app for this.

1. **REFERENCES:**
   1. Alippi C., Disabato S. and Roveri M. (2018). Moving convolutional neural networks to embedded systems: the Alexnet and VGG-16 case. Proceedings of 17th ACM/IEEE International Conference on Information.Processing in Sensor Networks. IEEE Press. 212-223.
   2. Bottou L. (2010). Large-scale machine learning with stochastic gradient descent. Proceedings of COMPSTAT,Springer, 177-186.
   3. Brilakis I., Soibelman L. and Shinagawa Y. (2005). Material-based construction site image retrieval. Journal of Computing in Civil engineering, Vol. 19, No. 4, 341-355.
   4. Brilakis I. and Soibelman L. (2008). Shape-based retrieval of construction site photographs. Journal of Computing in Civil engineering, Vol. 22, No. 1, 14-20.
   5. Buja A., Stuetzle W. and Shen Y. (2005). Loss functions for binary class probability estimation and classification:Structure and applications. Working draft, 1-49.
   6. Chi S. and Caldas C. H. (2011). Automated object identification using optical video cameras on construction sites.Computer‐Aided Civil and Infrastructure Engineering, Vol. 26, No. 5, 368-380.
   7. Deng J., Dong W., Socher R., Li L.-J., Li K. and Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. IEEE conference on computer vision and pattern recognition, 248-255.
   8. Dimitrov A., and Golparvar-Fard M. (2014). Vision-based material recognition for automated monitoring of construction progress and generating building information modeling from unordered site image collections. Advanced Engineering Informatics, Vol. 28, No. 1, 37-49.