

PAPER • OPEN ACCESS

## Safety and Compliance Management System Using Computer Vision and Deep Learning

To cite this article: Megha Nain *et al* 2021 *IOP Conf. Ser.: Mater. Sci. Eng.* **1099** 012013

View the [article online](#) for updates and enhancements.

You may also like

- [User Deep Energy Efficiency Evaluation via Load Perception and Clustering](#)  
Shuran Zhai and Zhaojie Zhang
- [Construction of Marginal Effect Model for Reliability Improvement of Urban Distribution Network](#)  
Ying Zhang, Ruibang Gong, Jiangbo Gu *et al.*
- [Reliability Engineering Design Method based on Generalized Margin Design theory](#)  
Wensheng Peng and Yonghong Li



**ECS**  
The  
Electrochemical  
Society  
Advancing solid state &  
electrochemical science & technology

**DISCOVER**  
how sustainability  
intersects with  
electrochemistry & solid  
state science research

# Safety and Compliance Management System Using Computer Vision and Deep Learning

Megha Nain<sup>1</sup>, Shilpa Sharma<sup>1</sup> and Sandeep Chaurasia<sup>2</sup>

<sup>1</sup>Department of Computer Applications, School of Basic Science, Manipal University Jaipur, Jaipur-Ajmer Express Highway, Dehmi Kalan, Near GVK Toll Plaza, Jaipur, Rajasthan 303007

<sup>2</sup>Department of Computer Science and Engineering, School of Information and Technology, Manipal University Jaipur, Jaipur-Ajmer Express Highway, Dehmi Kalan, Near GVK Toll Plaza, Jaipur, Rajasthan 303007

E-mail: shilpa\_1819@yahoo.com

**Abstract.** With the advancement in the Deep Learning and Computer Vision-based methodologies, the safety and productivity analysis of the construction entities (workers, equipment, and structures) on-site can potentially be improved. Construction industry is majorly the most dangerous sectors with respect to occupational safety and health because of the chaotic and dynamic environment at construction sites. The study of safety and compliance monitoring measures at construction by using the computer vision and deep learning-based approaches is the key focus of this paper. Safety and compliance are acclamatory to each other and are very essential for managing the safe conditions at construction. Manual monitoring of the compliant and unsafe conditions at the site is a challenging job and thus, a smart framework for automated monitoring is necessary for the assurance of a safe and healthy environment. A framework for safety and compliance management system is presented in this paper for enhancing the safer work culture along with the measures. This paper examines the prior literature on computer vision as per the safety and compliance constraints for (I) Understanding the existing SOTA (state-of-the-art) methods and their respective outcomes (II) Finding the challenges and limitations in the currently employed approaches and (III) Providing the potential directions for a future line of work. The purpose is to build a culture of safety using Computer Vision and Deep Learning.

## 1. Introduction

Safety is a prime requirement of an individual. Construction workers face the highest probability of death on job site because of the nature of work at the construction site. In fact, construction workers are almost 5 times more likely to die on the active project site than any other industry worker. As per the records from the International Labor Organization [1], around 318,000 work-related casualties occur every year and out of all the casualties, construction industry contributes a substantial part. Every day, 950 people dies and across 72,000 workers get hurt due to work-related accidents [2][3]. For India, construction is the biggest employer of the nation after the agriculture sector, it provides employment to more than 50 million individuals, and contributes towards national GDP by almost 9%, however, its workforce is extremely unsafe in comparison with any other industrial sectors of the country [4]. The main reason for these occupational accidents is the falls, person getting stuck in the equipment, noxious atmosphere, electrocution, and collisions. Numerous of these injuries can be avoided if workers always wear proper PPE (Personal Protective Equipment) during work, PPE includes helmets for avoiding head injuries, safety glasses and vests, hand gloves for hand protection, safety glasses for eye protection, and boots such as steel toe boots etc. To avoid these occupational accidents, compliance monitoring is essentially required, so that, the workers who are not using proper



Content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

protocols of safety, for example, Not wearing proper PPE (Vest, hardhats, harness etc.) and workers who are constantly violating norms, such workers need more attention for reducing accidents at construction. Safety must be the prime focus at every construction site and for complimenting the safety, compliance measures would further ensure the safe work culture. Safety is defined as liberty from any kind of danger, risk, injury, or loss which could harm an individual in a physical or psychological way and on contrast compliance is the act of obeying, accepting, and surrendering to laws that are essential for the safety of the individual [5]. With the introduction of smart safety framework using computer vision (CV) and deep learning (DL), safety at construction could be achieved in an enriched way. The present paper reviews the recent work on safety and compliance monitoring in the construction industry.

Most used practices involve site observations and inspections for finding the risk involved in the ongoing projects at construction [6], however, these manual inspections are not only ineffective in terms of time and cost but also less accurate [7][8]. These limitations make the construction environment more difficult to be managed as the worker's environment keeps on changing over time and the expert safety officer might not always be present at sites and consequently making the manual inspection processes unreliable for regular practice [9]. With the currently employed technologies in the construction industry, CV based detection and tracking methods have a huge role in the monitoring of complex large-scale sites. With the assimilation of CV and DL Models, continuous and automated monitoring of the unsafe acts/measures can be accomplished with better accuracy and reduced cost. Image features can automatically be extracted and applied to learn from the training data with the combination of CV and DL Methods. CNNs (Convolutional Neural Networks) is one of the most widely implemented DL methods, as it has been able to beat other DNNs (deep neural networks) very accurately and reliably in the areas such as classification, detection, and segmentation of the objects in a digital images and videos [10][11]. CV helps in providing a rich set of information about the chaotic scene at construction by capturing the digital data that expedite the understanding of the complex situations/tasks at the site, for instance, location, project entities behavior and the conditions at the site can be monitored in an effective and timely manner. These developments convey the technical as well as operational benefits above other sorts of sensing methods such as RFID [12][13], GPS [14] and UWB that involves the installation of sensors to all the construction entities which needs to be tracked and moreover, it provides very less information, such as, location [15][16][17]. Consequently, CV has been sourced to multiple areas in construction such as, progress monitoring [18][19], performance analysis [20], defect monitoring, and documentation in an automated way [21]. CV has an extensive scope in the monitoring of safety and compliance measures that addresses the limitation of the manual inspection practices and creating prospects for automating the tasks of risk identification and evaluation [21][22] by the extraction of useful information from the image and videos [12] and despite these advances in CV, these practices are less frequently applied to actual practice by the engineers. This suggests the requirement of further study about the more advanced and improved methods that could be adopted to make construction safer and healthier.

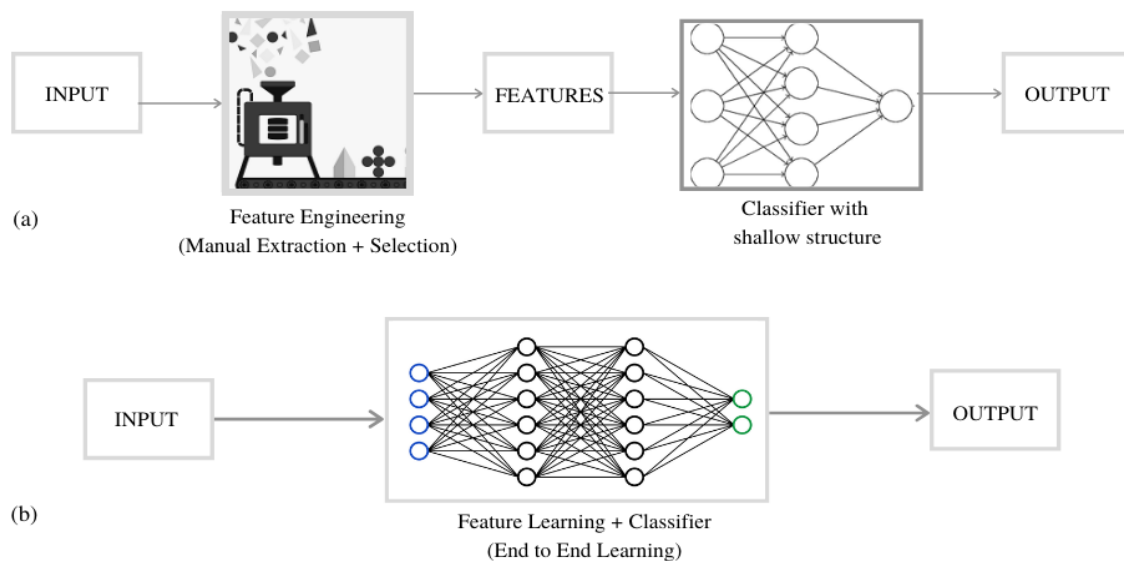
The paper is organized in 8 sections in a way that, section 2 briefly discusses an overview of CV techniques along with the DL. Then, Section 3 reviews safety monitoring using vision-based techniques. Section 4 focuses on monitoring of compliance measures, next, section 5 and 6 provides the details about collection of data set and SOTA methods employed based on the prior work in the construction sector using CV and DL and Section 7 outlines the brief discussion about the challenges and future scope of the computer vision research towards the goal of advanced automated construction. Lastly, section 8 delivers a crisp conclusion.

## 2. Overview of Computer Vision and Deep Learning

Computer Vision (CV) is an artificial Intelligence (AI) field having potential of extricating valuable information from the digital data such as images or videos. CV lies at the intersection of many other fields such as Computer Science, Mathematics, Engineering, Psychology and Physics. Computer Vision imitates the visual system of the human eye. Computer Vision along with Deep Learning produces outstanding accuracy and optimized results in visual recognition. DL lies within Machine Learning (ML) that has networks capable of learning from the unsupervised data.

For performing a computer vision task, the main process is to do the feature extraction in an image and for that, detection of the edges, corners, color, and object is the initial step. "For the feature extraction from a raw image, the conventionally used algorithms are, SIFT (Scale-Invariant Feature Transform),

SURF (Speeded-Up Robust Features) and BRIEF (Binary Robust Independent Elementary Features) [23], however, in case the image clarity is not so good and the number of classes for the classification goes higher, these conventional algorithms face difficulty in the feature extraction.” With the Deep Learning (DL) approaches, these challenges of the feature extraction can be overcome. In Figure 1, working of deep learning algorithms with the computer vision have been presented. In 2012, one of the most influential papers beat the earlier SOTA methods to image recognition in a famous ImageNet CV contest that proves the significance of applying DL approach with CV. The boom began with the CNNs and the various models of “ConvNets” and near to 100% accuracy in the tasks of image classification can be reached with the ConvNET architectures. Thus, DL is playing a major tool for the CV [24].



**Figure 1.** Computer Vision along with the Deep Learning (a) Standard flow of Computer Vision, (b) Deep Learning flow [23]

The power of the DL Methods to learn useful features by its own from the training data, which is annotated, makes it very useful in CV, especially the usages of CNNs for the tasks of image classification and object detection methods.

## 2.1 Techniques in Computer Vision for the Constructional Intelligence

Five Computer Vision techniques are discussed briefly.

**2.1.1 Image Classification.** In Image classification, a collection of a training dataset of labelled images is feed into the computer for processing of data. The computer studies images and learns about the visual appearance of the images. Extension of the image classification tasks to multiple frames leads to action recognition and then combining the prediction from each frame.

**2.1.2 Object Detection.** Object Detection is the task of identifying objects in an image, labelling them and outputting boundary boxes. This Technique varies with the methodology used. “Object detection when applied on each and every frame of a video turn into an object tracking task” [25].

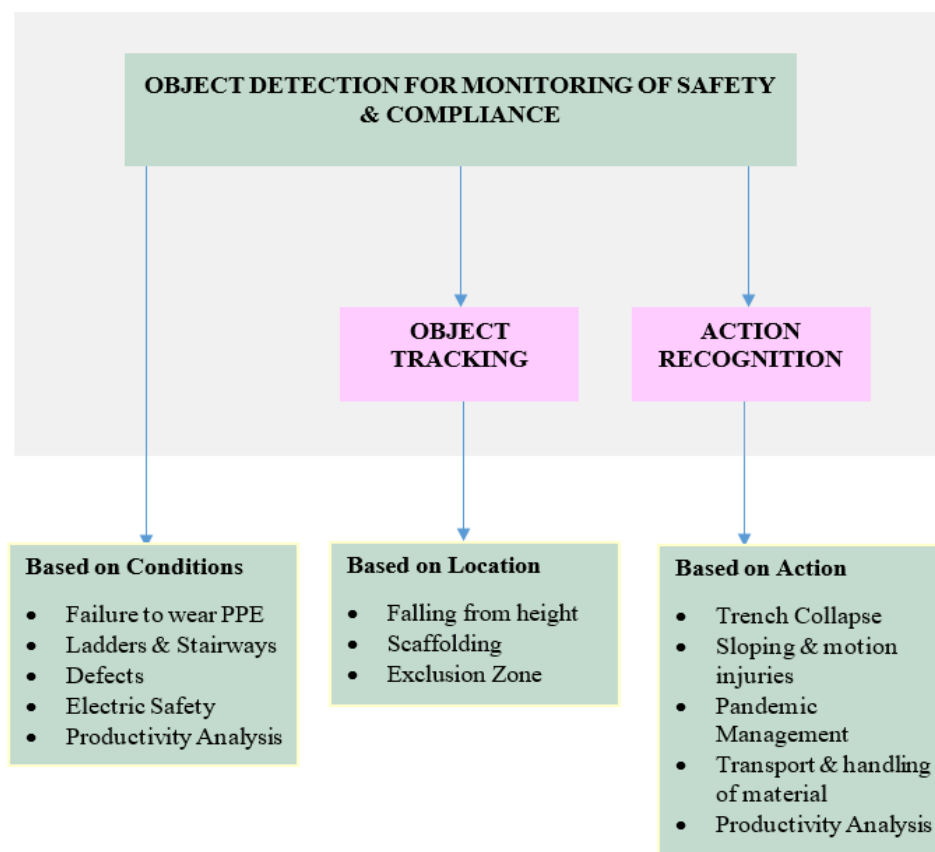
**2.1.3 Object Tracking.** This technique refers to tracking one or more moving objects in any given scene. This technique is traditionally been applied to monitor real-world interactions. Object tracking is applied to the frames of the video.

**2.1.4 Semantic Segmentation.** Segmentation is an essential part of the Computer Vision that divides the entire image into a group of pixels that can be labelled and classified. To be more specific, semantic segmentation attempts to understand the part that each pixel plays in each image.

**2.1.5 Instance Segmentation.** This technique categorizes all the various instance classes. For instance, a complex scene with many overlapping objects and backgrounds, then the classification of all the objects has been done, also the identification of their differences and how these objects are related to each other, these all need to be evaluated.

### 3. Computer Vision based framework for the monitoring of Safety at construction

Safety and compliance are complementary yet different from each other. As per the OSHA (Occupational Safety and Health Administration) which is United States Department of Labor, Compliance is a key for the for the prevention of injuries and accidents at the work [26]. For addressing the problem, previous research has been studied for getting the insight into reasons behind occupational accidents of the construction workers. J. Seo et al., 2015 [27] proposed a framework for ensuring the health and safety monitoring of the workers at construction. The earlier procedures and conditions that lead to these injuries suggests that more attention towards the safety and compliance area is required and with the help of proper safety and compliance measures, these accidents can potentially be avoided [28][29].



**Figure 2.** Framework for the Safety and Compliance Management System

A framework for the Safety and Compliance Management System to assure the safer conditions at construction has been presented in Figure 2. According to the formulated framework, firstly, an image data set is created using image sensing devices (shown in Figure 2), the data set formed comprised on 3 techniques of CV: (I) Object detection approach; (II) Object tracking; and (III) Action recognition technique in accordance with the unsafe and complaint conditions. Object detection is used for the identification of

conditions/situation-based risks and it is a prerequisite for the next methodologies that is object tracking and action recognition. Location can be tracked using the object tracking algorithms and thus location-based unsafe and complaint acts can be monitored. Action recognition techniques are applied with the help of sequential images to detect what the construction-related entities (e.g., worker and equipment) are doing to check the unsafe acts which could violate the safety and compliance protocols.

### 3.1 Safety conditions and actions

As per the OSHA [26], different conditions are acknowledged, wherein the safety is hampering in construction. Computer vision could absolutely help us in automated monitoring of such acts, thereby assuring the overall safety in construction.

**3.1.1 Personal Protective Equipment (PPE).** Personal protective equipment also termed to as "PPE", is an equipment for the protection against personal injury and illnesses at the workplace. These illnesses and injuries are the outcome of various workplace hazards, for example, radiological, chemical, electrical hazards etc. [30][31]. In Figure 3, various types of PPE are presented and a brief overview of the methods, data source and objectives for the safety monitoring are given in Table 1.



**Figure 3.** Types of PPE [30][31]

PPE includes hand protection gloves, safety glasses and work boots, ear protection, harness, helmets, safety vests, respirators, and complete bodysuits.

**Table 1.** An Overview of safety Monitoring using Computer Vision

| <b>Author Name and Year</b>          | <b>Data Set</b>                                                                                       | <b>Model Used</b>                                                                                                                                                                                                                                                                                                                                  | <b>Objective</b>                                                       | <b>Reference</b> |
|--------------------------------------|-------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------|------------------|
| <b>Qi Fang et al., 2018</b>          | 20,000 Plus Images: training Set including steeplejacks, workers, anchorages with and without webbing | <ul style="list-style-type: none"> <li>• Deep Learning Based occlusion mitigation Method</li> <li>• R-CNN model using localization and segmentation techniques</li> <li>• SSD for object detection at different scales</li> <li>• YOLO Method. (You Only Look Once) Version-3</li> <li>• Acoustic scene classification (ASC) classifier</li> </ul> | Falling prevention for Steeplejacks                                    | [32]             |
| <b>Daeho Kim et al., 2019</b>        | COCO benchmark Data Set<br>ImageNet Data Set                                                          | <ul style="list-style-type: none"> <li>• DNN, YOLO-V3</li> <li>• Image rectification Method – Measurement of distance using 2D Image from UAV (Unmanned Aerial Vehicle)</li> </ul>                                                                                                                                                                 | Struck-by accidents between the excavators and other heavy machineries | [33]             |
| <b>Yu Zhao et al., 2019</b>          | visual object classes (VOC) dataset and ImageNet dataset                                              | <ul style="list-style-type: none"> <li>• YOLO V3 Algorithm – Real Time detection</li> <li>• Kalman filter</li> <li>• Hungarian matching algorithms</li> </ul>                                                                                                                                                                                      | PPE - Helmets and colored vests detection to avoid accidents           | [34]             |
| <b>Clive Q.X. Poh et al., 2018</b>   | Image and Video Data                                                                                  | <ul style="list-style-type: none"> <li>• ML Techniques – Logistics regression (LR), K-Nearest Neighbors (K-NN), Boruta feature selection, decision tree (DT), Random forests (RF) and Support vector machines (SVM)</li> </ul>                                                                                                                     | Risky site classification for safety                                   | [35]             |
| <b>SangUk Han et al., 2013</b>       | Image and video using - VICON and Kinect                                                              | <ul style="list-style-type: none"> <li>• Vision framework for detection of unsafe behavior, motion templates, 3D skeleton extraction from videos</li> </ul>                                                                                                                                                                                        | behavior-based safety management                                       | [36]             |
| <b>Milad Memarzadeh et al., 2013</b> | 8000 annotated video data set                                                                         | <ul style="list-style-type: none"> <li>• (HOG+C) Histograms of Oriented Gradients and Colors</li> <li>• Sliding window at multiple scales</li> <li>• Binary SVM classifier</li> </ul>                                                                                                                                                              | Detection of equipment and worker                                      | [37]             |

|                           |              |                                                                         |                                                                                                                                                                                    |                            |      |
|---------------------------|--------------|-------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------|------|
| <b>Nipun Nath al,2020</b> | <b>D. et</b> | crowdsourcing and web-mining for data Set creation instances of workers | <ul style="list-style-type: none"> <li>• SSD (Single Shot Detector)</li> <li>• YOLO V3</li> <li>• R-FCN (Region-based fully convolutional network)</li> <li>• RetinaNet</li> </ul> | Real Time detection of PPE | [38] |
|---------------------------|--------------|-------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------|------|

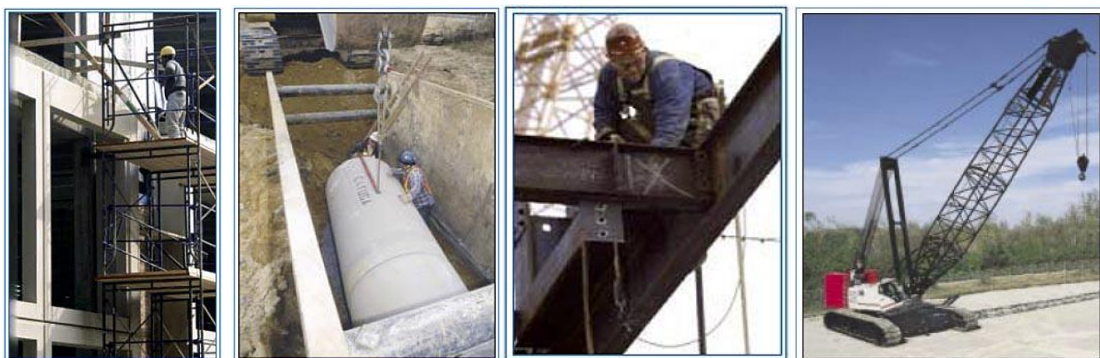
**3.1.2 Falling from heights.** Usage of Body harnesses or specialized body net for the avoidance of falling accidents. CV framework can be implemented to check for the harness.

**3.1.3 Ladders and Stairways.** Around 24,882 injuries and more than 36 fatalities every year due to falling on the stairways and ladders used in construction. To avoid an accident due to ladder and stairways, the placement of the ladder should be proper, safe, and long enough to reach the work area. Likewise, stairways ought to be free from dangerous objects and resources. Reese and Edison also identified unsafe conditions regarding the personal damage, property damage or equipment failure etc [39].

**3.1.4 Trench collapse.** Trench collapse leads to lots of fatalities every year. With the usage of protective equipment while entering any deep and big trenches could avoid accidents cases

**3.1.5 Collapse of Scaffolds.** Almost 2.3 million construction workers often work on scaffolds framework at the same time. If we protect these workers from scaffold-based accidents, then it would prevent an approximate 4,500 injuries and 50 fatalities every year. Scaffold must be strong enough to carry its own weight as well as 4 times the maximum projected load. Scaffold must be on solid footing to avoid any collapse situation.

**3.1.6 Sloping / Motion Injuries.** Based on the type of the soil type, allowable slopes are different. Sloping must be taken vigorously for the avoidance of accidents due to motion injuries. Figure 4 conveys the safety conditions.



**Figure 4.** Various actions for Safety (a) Scaffolding (b) Stairways (c) Harness for fall protection (d) Sloping [26]

Adhering to these safety requirements encourages self-assurance and faith and saves the lives of thousands of workers.

### 3.2 Defect monitoring

Defects must also be monitored on regular basis, which would not just enhance the safety of workers but also the safety of freshly built structures. In Table 2 below, we reviewed previous work on the defect monitoring for the safety of the site and workers.

**Table 2.** Overview on the Defect Monitoring using Computer Vision

| Author's Name and Year                    | Data Resource                                                                                                                                                           | Model Used                                                                                                                                                                                                                                 | Description                                                              | Reference |
|-------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------|-----------|
| <b>Husein Perez et al., 2019</b>          | Image data collected from Camera, mobile phone and copyright images                                                                                                     | <ul style="list-style-type: none"> <li>• Convolutional neural networks (CNN)</li> <li>• VGG-16</li> <li>• ResNet-50</li> <li>• Inception models</li> <li>• Transfer Learning – classifier</li> </ul>                                       | Detecting defects such as mould, deterioration, and stain                | [40]      |
| <b>Nhat-Duc Hoang et al., 2018</b>        | 500 Image Data Set Training and Testing ML classifier                                                                                                                   | Image Processing Algorithms SVM and least squares SVM                                                                                                                                                                                      | Walls Defects detection using Image Processing and ML Methodologies      | [41]      |
| <b>Mingxin Nie et al., 2019</b>           | 4000 + crack images 3800 Training Images 400 Testing Images                                                                                                             | <ul style="list-style-type: none"> <li>• YOLO -V3 based Method</li> <li>• Darknet 53 Network and binary cross entropy loss for class prediction</li> </ul>                                                                                 | Crack detection on pavements Accuracy 88 %                               | [42]      |
| <b>Stephanie German et al., 2012</b>      | Image Data Set                                                                                                                                                          | <ul style="list-style-type: none"> <li>• Spalled areas detection in length and depth form</li> <li>• Entropy-based thresholding algorithm</li> </ul>                                                                                       | concrete spalling measurement                                            | [43]      |
| <b>Billie F. Spencer Jr. et al., 2019</b> | Modified National Institute of Standards and Technology (MNIST Data set) - 99.5% accuracy ImageNet., AlexNet, GoogleNet Video Data Set – Using UAV SYNTHIA dataset [45] | <ul style="list-style-type: none"> <li>• Object Detection and Semantic Segmentation algorithms</li> <li>• Clustering algorithms (k-means algorithm)</li> <li>• Optical flow algorithms</li> <li>• Gaussian mixture model (GMM),</li> </ul> | Detection of structural damage<br><br>99.5% accuracy with MNIST Data set | [44]      |

Defect Monitoring is an essential element to keep the safe environment in construction for both the staff as well as the structure itself.

## 4. Computer Vision Based Framework for Compliance Monitoring

OSHA has also provided the guidelines for the control and prevention of COVID-19 spread at the workplace [46] that majorly includes (I) social distancing that involves limiting work in the occupied areas, limit

physical contact, break and lunchtimes in shifts. (II) Usage of face masks and shields, hand gloves as per the work requirement, and eye protectors are mandatory while at the site. (III) Proper sanitization and cleanliness must be promoted for a healthier environment.

#### 4.1 Compliance parameters and Measures

Some general actions and compliance measures to make construction safe as well as progressive by using computer vision and deep learning models have been presented.

##### 4.1.1 For Pandemic Management (Covid-19)

- *Social Distance Monitoring Using automated framework of Computer Vision*

Social Distance is the primary means to control the COVID-19 spread till we find the vaccine, thus, with the object detection and object tracking techniques, CV tool could be implemented for assuring the social distancing. Also, Camera Calibration is another aspect of measuring social distance. It is a universal approach that calculates the distance between 2 objects in an image. The approach converts the video into a birds-eye view or a top view for calculating the distance.

- *Face Mask Detection*

Training the face mask detector model with the dataset of construction workers with and without face mask could help in automated detection of violation of the face mask compliance at site. Further, Personal Protective Kit (PPE) for Face, head, eye, hand, and foot must always be mandatory at the construction site to avoid Covid-19 spread.

- *Daily health monitoring*

With the Thermal Imaging Cameras, daily checking of the temperature of the construction workers can be observed. Thermal Vision Cameras generates alarms automatically in case of any violations.

- *Sanitization using drones*

Drones are the most used tool during the corona crisis for the sanitization of the coronavirus on a massive scale.

**4.1.2 Failure to use of PPE (Personal Protective Equipment).** Wearing proper PPE would certainly avoid a huge number of accidental cases in construction [47]. Recently Nipun D. Nath et al., [38] presented 3 models based on DL for the real time verification of PPE, developed on YOLO architecture. Safety helmets and jackets are detected on real-time in their research. 3 approaches were used (I) Detection using ML model (e.g., DT and NNs) (II) Using Single CNN framework (III) Using classifiers based on CNN (ResNet-50, VGG-16, and Xception,)

For real-time detection, the result of the second approach achieved the best performance (mean average precision of 72.3%) and most suitable for running in lightweight gadgets.

**4.1.3 Exclusion Zones.** Around 32 million workers work in a hazardous condition such as exposure to chemicals. Workers must be aware of all the exposed chemically exposed zones. For keeping a healthy environment for the workers, compliance can play a greater role. Workers must not be allowed to enter these exclusion zones without proper PPE [48].

**4.1.4 Transport, earth-moving and handling of materials.** The transport equipment and the material handling equipment must be of proper design and in good form. These equipment's must be operated by only well-trained workers to avoid accidents [49].

**4.1.5 Electric Safety.** All the electrical tools must be checked on a regular basis and if there arises any fault condition that must be taken out and reported immediately. Also, Safety alarm System might be attached for receiving any such fault information at various electric spots.

#### 4.2 Performance monitoring

Performance of the workers could be taken as one of the compliance measures for the progress monitoring of the construction site. This would ensure the progress of the structure on time. In Table 3 and Table 4, the prior work and respective methodologies used for the performance and compliance monitoring in construction have been specified. Han and Lee [50] introduced a framework for computer vision that would track unsafe behaviors, focusing on four fragments comprising (I) Extraction of Unsafe behaviors from the safety documents and historical data (II) Using the experimental work for the collection of motion templates for unsafe actions (III) 3D skeleton extraction from video recordings; and (IV) Unsafe behaviors detection on real time using video data.

**Table 3.** Prior Work on the Performance Monitoring using Computer Vision

| Author's Name and Year              | Data Resource                                                                                       | Model Used                                                                                                                                                                                                                                                                                               | Description                                 | Reference |
|-------------------------------------|-----------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------|-----------|
| <b>Hesam Hamledari et al., 2017</b> | Video Data – Quadcopter (UAV)<br>Image Data - smartphone<br>Image Data -Publicly accessible sources | <ul style="list-style-type: none"> <li>•SFM (structure from motion) technique</li> <li>• HOG</li> <li>•Approaches based on color</li> <li>•Otsu Method</li> </ul>                                                                                                                                        | Progress Monitoring                         | [51]      |
| <b>Peddi et al., 2009</b>           | Image Data -<br>Time Lapse                                                                          | <ul style="list-style-type: none"> <li>•Object tracking techniques for BIM Elements</li> </ul>                                                                                                                                                                                                           | Appearance-based assessment                 | [52]      |
| <b>J. Gong et al., 2011</b>         | Video Data                                                                                          | <ul style="list-style-type: none"> <li>•Object detection and tracking</li> <li>•Activity identification algorithm- SVM</li> </ul>                                                                                                                                                                        | Analyze equipment for material handling     | [53]      |
| <b>Jie Gong et al., 2011</b>        | Image and Video Data Set                                                                            | <ul style="list-style-type: none"> <li>•Object detection and tracking techniques</li> <li>•Background subtraction</li> <li>•Color-based recognition method</li> <li>•cascade of simple features-based method</li> <li>•Mixture of Gaussian Models</li> <li>•Kalman Filter and Particle Filter</li> </ul> | Productivity analysis of workers            | [54]      |
| <b>J. Teizer et al., 2009</b>       | 2D Images                                                                                           | <ul style="list-style-type: none"> <li>•Tracking techniques for construction workers</li> </ul>                                                                                                                                                                                                          | Automated monitoring of project performance | [55]      |

**Table 4.** Prior Work on the Compliance Monitoring using Computer Vision

| Author Name | Data Sources | Model Used | Objective | Reference |
|-------------|--------------|------------|-----------|-----------|
|-------------|--------------|------------|-----------|-----------|

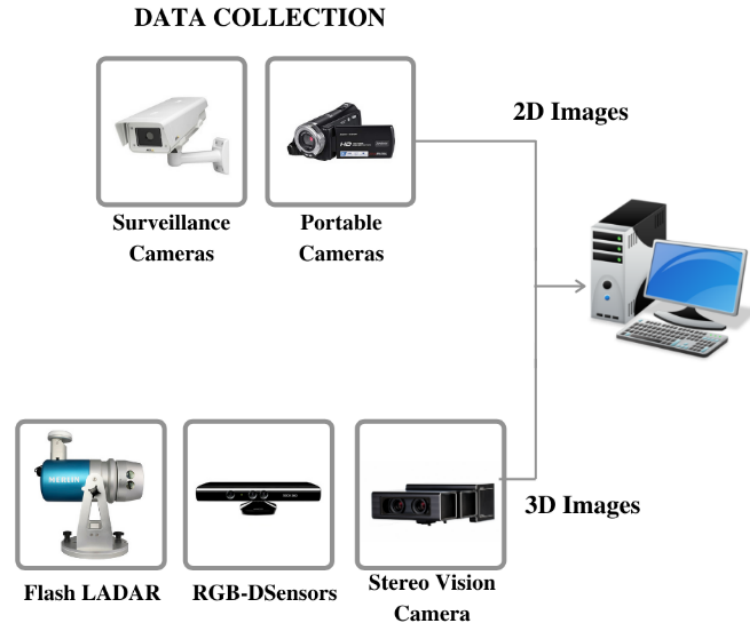
|                                   |                                        |                                                            |                                                                                         |      |
|-----------------------------------|----------------------------------------|------------------------------------------------------------|-----------------------------------------------------------------------------------------|------|
| <b>Nipun D. Nath et al., 2020</b> | Image and Video Data                   | YOLO V3 Model Object Detection techniques                  | Real Time Detection of PPE Violation                                                    | [38] |
| <b>S.Du et al., 2011</b>          | 2D Images                              | Object detection techniques for Hardhat                    | Failure to wear proper PPE                                                              | [56] |
| <b>J. Teizer et al., 2009</b>     | 2D Images                              | Tracking techniques for construction workers               | Automated monitoring of project performance                                             | [55] |
| <b>M.W. Park et al, 2011</b>      | 2D Images                              | Tracking Methods for workers and construction equipment    | Comparison of Vision tracking techniques                                                | [57] |
| <b>M. Memarzadeh et al., 2013</b> | 3D Images                              | Detection and Tracking Techniques of construction entities | Avoidance of accident due to close distance of Equipment and worker while in idle state | [58] |
| <b>S.J. Ray et al., 2012</b>      | 3D Images                              | Image Classification and action recognition                | Monitoring risky posture of worker                                                      | [59] |
| <b>S.Han et al, 2012</b>          | 3D Images and Videos                   | Image Classification                                       | Ladder climbing unsafe action                                                           | [60] |
| <b>S. chi et al., 2011</b>        | 2D Images                              | Detection techniques                                       | Performance monitoring                                                                  | [61] |
| <b>Qi Fang et al, 2018</b>        | 100,000 construction worker image Data | Faster R-CNN Method                                        | Hardhat compliance                                                                      | [62] |

## 5. Data Collection

Computer Vision research relies on 2 major factors (I) collection of data (digital data and time-lapse images) (II) analyzing the collected data. The principle of CV is “to analyze what you look over” [63]. The data collection is an extremely crucial phase in CV so that it can be analyzed properly and effectively. The image sensing devices for the data collection has been provided in Figure 5. Data Set “Pictor-v3” using web mining and crowdsourcing for the verification of PPE using YOLO architecture [38]. Depth based Cameras to examine unsafe acts and conditions for construction workers [62] Examples: Kinect and VICON [64].

There are several open access data sets available (as mentioned in the Table 1), For Example, MS-COCO, Open Images Dataset, Visual QA, ImageNet, Street View House Numbers (SVHN), CIFAR-10, MNIST, Fashion-MNIST etc.

The data collection is shown in figure 5.



**Figure 5.** Data Collection [27].

## 6. Methods

Most utilized methods for the classification and detection of the images are SVM (Support Vector Machines) and DL using CNNs and RNNs (Recurrent Neural Networks) to extract features automatically. Also, the best algorithms for the real-time detection, are YOLO [65], SSD [66], and RetinaNet and R-FCN [67] (region-based fully convolutional network) in terms of performance [38].

YOLO - V3 particularly is the fastest and accurate in comparison with other algorithms [68] such as SSD and R-FCN. The computational load of YOLO V3 is less (unlike R-CNN) which makes YOLO –V3 faster in detecting the construction entities. YOLO uses maximum suppression (NMS) technique [69][70][71] to avoid the redundancy in the output phase. With NMS Technique, bounding boxes having lower confidence level but having a higher percentage of overlapping will be removed and thereby, providing one bounding box for one object. The performance of an Object detection algorithm in CV can be measured using 3 metrics (I) Precision curves (II) DET (Detection Error Tradeoff curves) (III) MAP (Mean Average Precision).

In Equation (1), the formula for recall and precision is given.

Here, TP implies “True positive”, FN implies “false negative” and FP implies “false positive” [72].

$$\left\{ \begin{array}{l} Precision = \frac{TP}{TP + FP} \\ Recall = \frac{TP}{TP + FN} \end{array} \right. \quad (1) [72]$$

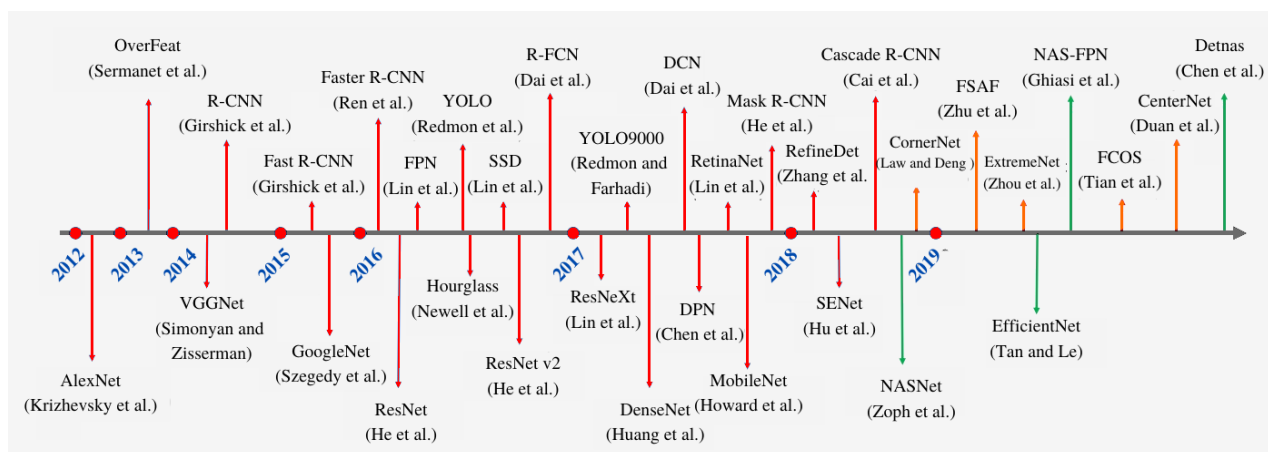
For Instance, IoU (Intersection over union) describes the overlap amongst ground truth (G) and the predicted box (P) given in Equation (2)

$$IoU = \frac{\text{intersection}}{\text{union}} = \frac{G \cap P}{G \cup P} \quad (2) [38]$$

Cameras for recording the images and videos on-site for the verification PPE compliance. Regular RGB is more suitable for practical usage rather than RGBD cameras due to limitation of view range [73] for detection. Prior methods include Video surveillance methods such as edge detection [74], motion detection, color feature and facial features [75] and Histogram of Oriented Gradient (HOG) [76] based ML algorithms such as k-NN and SVM [77].

## 7. Discussion

Object Detection is the fundamental technique in CV and with the addition of DL, it becomes more prevailing. The evolution of CV techniques from 2012 to 2019 is shown in Figure 6.



**Figure 6.** Computer Vision Techniques evolution from 2012 to 2019 [78]

Alex-Net [78] Boosted Cascade [79][80], Histogram of oriented gradients (HOG) [81], CVPR [82][83][84], Regions with CNN features (RCNN) [85], Spatial Pyramid Pooling Network (SPP-Net) [86], Fast RCNN [87], Faster RCNN [88], You Only Look Once (YOLO) [89], Single Shot Detector (SSD) [90], Pyramid-Networks [91] and Retina-Net [92].

One of the major challenges in the detection and tracking approaches based on CV is their susceptibility towards poor light conditions, blurriness, and occlusion situations. Further computer hardware is a major factor while doing the testing i.e., the execution speed varies with respect to the computer hardware. Size of data set is also one of the challenging situations, researchers with limited data set are reliant on the usage of data augmentation techniques, chooses a small sample to undertake their trial work, that makes it hard for the comparison and metric evaluation such as precision and recall.

Privacy is also another issue, which makes compliance management challenging at sometimes. This paper involves the theoretical study of the safety and compliance measures, practical implementation of the same could be performed in future research for providing thorough results.

Ontology and CV techniques can be integrated for overcoming the semantic gap in the future line of research. “Ontology is an approach, popular for having computer-oriented and logic-based features. it is applied for modelling information which would help in providing the domain knowledge by the explicit definition of classes, instances, and relationships” [93][94][95]. The framework of ontology and CV techniques could be formed by developing spatial relationships between objects, which could automatically be detected using CV techniques [96].

Due to the capability of Deep Learning for automatically extracting features, DL is gaining immense popularity for methods of object classification and detection, however, DL works on the correlation between the input and output states and is unable to identify the causality. For instance, the monitoring of safety conditions demands

not just the necessity of identifying the individuals and operating conditions, but it also needs the interactions of the individuals which must be evaluated for the absolute safe workplace. Till this date, the interactions have not been analyzed, which could be again considered as a future area of research [97][98]. Moreover, new algorithms are required for the identification of several jobs concurrently.

## 8. Conclusion

Progress often takes the front seat, however, without a safe and healthy workforce, it is challenging to move forward. Computer Vision techniques in civil engineering have been perceived as a key segment for improved monitoring in terms of safety and compliance conditions. In this paper, various conducts using which safety and compliance monitoring could be achieved efficiently and progressively using computer vision framework along with deep learning have been discussed. A safety and compliance management framework with the featured conditions and measures to control the unsafe acts has been proposed. In Addition, with the compliance monitoring, not only the productivity but also, safety can be assured for both workers as well as new structures. Various techniques of computer vision and deep learning have been reviewed, furthermore, challenges and feasible areas for the future line of work are discussed. The notion is to form a healthier and progressive environment with the usage of appropriate safety and compliance techniques at construction.

## 9. References

1. Ahonen, international labor organization (ilo) 2010 *xix world congress on safety and health at work* [http://www.perosh.eu/wp-content/uploads/2013/06/perosh\\_osh-and\\_competitiveness\\_g\\_ahonen\\_fioh.pdf](http://www.perosh.eu/wp-content/uploads/2013/06/perosh_osh-and_competitiveness_g_ahonen_fioh.pdf) (last updated on 15-Oct-2020)
2. Patel D A K and Jha K N 2016 *An estimate of fatal accidents in indian construction, proc. of the 32nd annual arcom conf.* pp 5-7
3. Patel D A K and Jha K N 2015 *Neural network model for the prediction of safe work behavior in construction projects. Journal of construction engineering and management* 141(1). Doi: 141(1), 04014066.
4. Hämäläinen and P 2010 *global estimates of occupational accidents and fatal work-related diseases* (PhD thesis) (tampere university of technology)
5. Website: <https://www.mccrarencompliance.com/safety-versus-compliance> (last updated on 15-Oct-2020)
6. Hinze J and Godfrey R 2003 *An evaluation of safety performance measures for construction projects, j. Constr. Res.* 4 (01) 5–15 doi: 10.1016/j.ssci.2013.04.004
7. Levitt R E 1993 *Construction safety management, john wiley & sons*
8. Taneja S, Akinci B, Garrett J H, Soibelman, Ergen, Pradhan and Anil 2011 *Sensing and field data capture for construction and facility operations, j. Constr. Eng. Manage.* 137 (10) pp 870–881.
9. Laitinen H, Marjamäki M and Päiväranta K 1999 *The validity of the tr safety observation method on building construction, accid. Anal. Prevent.* 31 (5) pp 463–472.
10. Golparvar-fard M, Heydarian A and Niebles J C 2013 *Vision-based action recognition of earthmoving equipment using spatio-temporal features and support vector machine classifiers, adv. Eng. Inform.* 27 (4) pp 652–663, <https://doi.org/10.1016/j.aei.2013.09.001>.
11. Memarzadeh M, Golparvar-fard M and Niebles J 2013 *Automated 2d detection of construction equipment and workers from site video streams using histograms of oriented gradients and colors, autom. Constr.* 32 24–37, <https://doi.org/10.1016/j.autcon.2012.12.002>.
12. Brilakis I, Park M W and Jog G 2011 *Automated vision tracking of project related entities, adv. Eng. Inform.* pp 713–724.
13. Shahandashti S M, Razavi S N, Soibelman L, Berges M, Caldas C H, Brilakis I, Teizer J, Vela P A, Haas C T, Garrett J, Akinci B and Zhu Z 2011 *Data fusion approaches and applications for construction engineering, asce j. Constr. Eng. Manage. Reston virginia* 137 (10) pp 863–869.
14. Pradhananga N and Teizer J 2013 *Automatic spatio-temporal analysis of construction equipment operations using gps data, autom. Constr.* 29 pp107–122.
15. Teizer J, Allread B S, Fullerton C E and Hinze J 2010 *Autonomous pro-active real-time construction worker and equipment operator proximity safety alert system, automat. Constr.* pp 630–640.
16. Cheng T, Mantripragada U, Teizer J and Vela P A 2012 *Automated trajectory and path planning analysis based on ultra-wideband data, asce j. Comput. Civil eng. Reston virginia* pp 151–160.
17. Yang J, Cheng T, Teizer J, Vela P A and Shi Z K 2011 *A performance evaluation of vision and radio frequency tracking methods for interacting workforce, adv. Eng. Inform.* pp 736–747

18. Taneja S, Akinci B, Garrett J H, Soibelman L, Ergen E, Pradhan A and Anil 2011 *Sensing and field data capture for construction and facility operations*, *j. Constr. Eng. Manage.* pp 137 870–881.
19. Golparvar-fard M and Peña-mora F 2007 *Application of visualization techniques for construction progress monitoring*, *comput. Civil eng.* 27.
20. Gong J and Caldas C H 2009 *Computer vision-based video interpretation model for automated productivity analysis of construction operations*, *j. Comput. Civil eng.* 24 (3) pp 252–263.
21. Chi S, Caldas and Kim D Y 2009 *A methodology for object identification and tracking in construction based on spatial modeling and image matching techniques*, *comput.-aid. Civil inf. Eng.* 24 (3) pp 199–211.
22. Golparvar-fard M, Bohn J, Teizer J, Savarese S and Peña-mora F 2011 *Evaluation of image-based modeling and laser scanning accuracy for emerging automated performance monitoring techniques*, *automat. Constr.* 20 (8) pp 1143– 1155.
23. Mahony N, Campbell S, Carvalho A, Harapanahalli S, Hernandez G V, Krpalkova L, Riordan D, Walsh J 2010 *Binary robust independent elementary features & dblp, , deep learning vs. Traditional computer vision in advances in computer vision proceedings of the 2019 computer vision conference (cvc). Springer nature switzerland ag*, pp. 128-144
24. Website: <http://www.image-net.org/challenges/lsrvrc/2012>. (last updated on 15-Oct-2020)
25. Website Source: <https://www.analyticsvidhya.com/blog/2020/05/social-distancing-detection-tool-deep-learning> (last updated on 15-Oct-2020)
26. Website: <https://www.osha.gov> (last updated on 15-Oct-2020)
27. Seo J et al. 2015 *Computer vision techniques for construction safety and health monitoring*, *adv. Eng. Informat.* [http:// dx.doi.org/10.1016/j.aei.2015.02.001](http://dx.doi.org/10.1016/j.aei.2015.02.001)
28. Ringen K, Seegal J and England A 1995 *Safety and health in the construction industry*, *ann. Rev. Public health* 16 (1) pp 165–188
29. Reese C D and Eidson J V 2006 *Handbook of osha construction safety and health*, *crc press*
30. Website: <https://www.osha.gov/personal-protective-equipment> (last updated on 15-Oct-2020)
31. Osha, *worker safety series: construction* Website: <https://www.osha.gov/publications/ osha3252/3252.html> (last updated on 15-Oct-2020)
32. Fang Q, Li H, Luo X, Ding L, Luo H and Li C 2018 *Computer vision aided inspection on falling prevention measures for steepjacks in an aerial environment*, doi:<https://doi.org/10.1016/j.autcon.2018.05.022>, *elsevier*
33. Kim D, Meiyingliu, Lee S H and Kamat V R 2019 *Remote proximity monitoring between mobile construction resources using camera-mounted uavs* doi : <https://doi.org/10.1016/j.autcon.2018.12.014>, *elsevier*
34. Zhao Y, Chen Q, Cao W, Yang J, Xiong J and Gui G 2019 *Deep learning for risk detection and trajectory tracking at construction sites*, *ieee access*, doi: 10.1109/access.2019.2902658
35. Poh C Q X, Ubeynarayana C U and Goh Y M 2018 *Safety leading indicators for construction sites: a machine learning approach*, *elsevier*, doi: <https://doi.org/10.1016/j.autcon.2018.03.022>
36. Han S and Lee S 2013 *A vision-based motion capture and recognition framework for behavior-based safety management*, *elsevier* doi:<http://dx.doi.org/10.1016/j.autcon.2013.05.001>
37. Memarzadeh M, Golparvar-fard M and Niebles J C 2012 *Automated 2d detection of construction equipment and workers from site video streams using histograms of oriented gradients and colors*, *elsevier*, doi : <http://dx.doi.org/10.1016/j.autcon.2012.12.002>
38. Nipun D. Nath et al. 2020 *Learning for site safety: real-time detection of personal protective Equipment* <https://doi.org/10.1016/j.autcon.2020.103085> *elsevier*
39. Reese C D and Eidson J V 2006 *Handbook of osha construction safety and health*, *crc press*, doi: <https://doi.org/10.1201/9781420006230>
40. Perez H, Joseph h, Tah M and Mosavi A 2019 *Deep learning for detecting building defects using convolutional neural networks*, *sensors*, 19, 3556; doi:10.3390/s19163556
41. Hoang N D 2018 *Image processing-based recognition of wall defects using machine learning approaches and steerable filters*, *hindawi computational intelligence and neuroscience volume*, article id 7913952, 18 pages <https://doi.org/10.1155/2018/7913952>
42. Nie M and Wang C 2019 *Pavement crack detection based on yolo v3*, 2019, *international conference on safety produce informatization (iicspi)*
43. German S, Brilakis I, Desroches R 2012 *Rapid entropy-based detection and properties measurement of concrete spalling with machine vision for post-earthquake safety assessments*, *elsevier*, doi: <http://dx.doi.org/10.1016/j.aei.2012.06.005>

44. Billie F, Spencer jr Et al. 2019 *Advances in computer vision-based civil infrastructure inspection and monitoring* - <https://doi.org/10.1016/j.eng.2018.11.030> published by elsevier ltd
45. Ros G, Sellart L, Materzynska J, Vazquez D, Lopez AM 2016 *The synthia dataset: a large collection of synthetic images for semantic segmentation of urban scenes*. In: *proc. of 2016 ieee conf. on computer vision and pattern recognition*; 27–30; las vegas, nv, usa. Piscataway: ieee. pp 3234–43.
46. Website source: <https://www.osha.gov/publications/osha3990.pdf> (last updated on 15-Oct-2020)
47. Website source: <https://www.osha.gov/publications/osha3252/3252.html> (last updated on 15-Oct-2020)
48. Website source: <https://www.osha.gov/publications/osha3084.pdf> (last updated on 15-Oct-2020)
49. Website: <https://www.ilo.org> (last updated on 15-Oct-2020)
50. Han S and Lee S 2013 *A vision-based motion capture and recognition framework for behavior-Based safety management, autom. Constr.* 35 pp 131–141, <https://doi.org/10.1016/j.autcon.2013.05.001>
51. Hamledari H , Brendamccabe and Shakibadavari 2017 *Automated computer vision-based detection of components of under-construction indoor partitions*, doi:<https://doi.org/10.1016/j.autcon.2016.11.009> , elsevier
52. Peddi A, Huan L, Bai Y and Kim S 2009 *Development of human pose analyzing algorithms for the determination of construction productivity in real-time, construction research congress* [https://doi.org/10.1061/41020\(339\)2](https://doi.org/10.1061/41020(339)2).
53. Gong J and Caldas C H 2011 *An object recognition, tracking, and contextual reasoning based video interpretation method for rapid productivity analysis of construction operations, autom. Constr.* 20 (8) pp 1211–1226, <https://doi.org/10.1016/j.autcon.2011.05.005>.
54. Gong J, Carlos H and Caldas 2011 *An object recognition, tracking, and contextual reasoning-based video interpretation method for rapid productivity analysis of construction operations, elsevier*, doi: 10.1016/j.autcon.2011.05.005
55. Teizer J and Vela P A 2009 *Personnel tracking on construction sites using video cameras, adv. Eng. Inform.* 23 (4) pp 452–462.
56. Du S, Shehata M, Badawy W 2011 *Hard hat detection in video sequences based on face features, motion, and color information, 2011 3rd international conference on computer research and development (iccrd), vol. 4, ieee*, pp. 25–29
57. Park M W, Makhmalbaf A and Brilakis I 2011 *Comparative study of vision tracking methods for tracking of construction site resources, automat. Constr.* 20 (7) pp 905–915.
58. Memarzadeh M, Golparvar-fard M and Niebles J C 2013 *Automated 2d detection of construction equipment and workers from site video streams using histograms of oriented gradients and colors, automat. Constr.* 32 pp 24– 37.
59. Ray S J and Teizer J 2012 *Real-time construction worker posture analysis for ergonomics training, adv. Eng. Inform.* 26 (2) pp 439–455.
60. Han S, Lee S and Peña-mora F 2012 *Vision-based detection of unsafe actions of a construction worker: case study of ladder climbing, j. Comput. Civil eng.* 27 (6) pp 635–644
61. Chi S and Caldas C H 2011 *Automated object identification using optical video cameras on construction sites, comput. Aided civ. Inf. Eng.* 26 (5) pp 368–380.
62. Fang Q, Li H, Luo X, Ding L, Luo H, Timothy M, Rosec and wangpeng A 2017 *Detecting non-hardhat-use by a deep learning method from far-field surveillance videos, elsevier*, doi :<http://dx.doi.org/10.1016/j.autcon.2017.09.018>
63. Fang W, Fang, Ding L, Zhong B, Love P E and Luo H 2018 *Automated detection of workers and heavy equipment on construction sites: a convolutional neural network approach, adv. Eng. Inform.* 37, <https://doi.org/10.1016/j.aei.2018.05.003> pp 139–149
64. Han S and Lee S 2013 *A vision-based motion capture and recognition framework for behavior- based safety management, autom. Constr.* 35, <https://doi.org/10.1016/j.autcon.2013.05.001> pp 131–141
65. Redmon J, Divvala S, Girshick R and Farhadi A 2016 *You only look once: unified, realtime object detection, proc. Ieee conference on computer vision and pattern recognition, las vegas, nv*, <https://doi.org/10.1109/cvpr.2016.91> pp 779–788
66. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C Y, Berg A C 2016 *Ssd: Single shot multibox detector, proc. European conference on computer vision, amsterdam, the netherlands*, [https://doi.org/10.1007/978-3-319-46448-0\\_2](https://doi.org/10.1007/978-3-319-46448-0_2) pp. 21–37

67. Lin TY, Goyal P, Girshick R, He K and Dollár P 2017 *Focal loss for dense object detection*, *proc. Ieee international conference on computer vision, venice, italy*, <https://doi.org/10.1109/iccv.2017.324> pp. 2980–2988
68. Redmon J, Farhadi A 2018 *yolov3: An incremental improvement*, *arxiv preprint*, [arxiv:1804.02767](https://arxiv.org/abs/1804.02767)
69. Redmon J, Divvala S, Girshick R and Farhadi A 2016 *You only look once: unified, realtime object detection*, *proc. Ieee conference on computer vision and pattern recognition, las vegas, nv*, <https://doi.org/10.1109/cvpr.2016.91>. pp. 779–788
70. Redmon J and Farhadi A 2017 *Yolo9000: better, faster, stronger*, *proc. Ieee conference on computer vision and pattern recognition, honolulu, hi*, <https://doi.org/10.1109/cvpr.2017.690> pp. 7263–7271
71. Girshick R, Donahue J, Darrell T and Malik J 2014 *Rich feature hierarchies for accurate object detection and semantic segmentation*, *proc. Ieee conference on computer vision and pattern recognition, columbus, oh* <https://doi.org/10.1109/cvpr.2014.81> pp. 580–587
72. Xie Z, Liu H, Li Z and He Y 2018 *A convolutional neural network based approach towards real-time hard hat detection*, *proc. Ieee international conference on progress in informatics and computing (pic)*, *ieee*, <https://doi.org/10.1109/pic.2018.8706269> pp. 430–434,
73. Fang Q, Li H, Luo X, Ding L, Luo H, Rose T M, An W 2018 *Detecting non-hardhat-use by a deep learning method from far-field surveillance videos*, *autom. Constr.* 85 <https://doi.org/10.1016/j.autcon.2017.09.018> pp 1–9
74. Du S, Shehata M and Badawy W 2011 *Hard hat detection in video sequences based on face features, motion and color information*, *proc. 3rd international conference on computer research and development (iccrd), shanghai, china 4* <https://doi.org/10.1109/iccrd.2011.5763846> pp. 25–29
75. Shrestha K, Shrestha P P, Bajracharya D and Yfantis E A 2015 *Hard-hat detection for construction safety visualization*, *journal of construction engineering –8*, <https://doi.org/10.1155/2015/721380>.
76. Park M W, Elsafty N and Zhu Z 2015 *Hardhat-wearing detection for enhancing on-site safety of construction workers*, *j. Constr. Eng. Manag.* 141 (9) 04015024, , [https://doi.org/10.1061/\(asce\)co.1943-7862.0000974](https://doi.org/10.1061/(asce)co.1943-7862.0000974).
77. Weston J and Watkins C 1998 *Technical report csd-tr-98-04*, department of computer science, university of london
78. Krizhevsky, Sutskever I, and Hinton G 2012 *Imagenet classification with deep convolutional neural networks*. In *nips* pp. 1097–1105
79. Viola P and Jones M 2001 *Rapid object detection using a boosted cascade of simple features*, "in *computer vision and pattern recognition Cvpr 2001. Proceedings of the 2001 ieee computer society conference on*, vol. 1. *Ieee*, 2001, pp. I–i.
80. P. Viola and m. J. Jones, "robust real-time face detection," *international journal of computer vision*, vol. 57, no. 2, pp. 137–154, 2004.
81. Dalal N and Triggs B 2005 *Histograms of oriented gradients for human detection*, "in *computer vision and pattern recognition Cvpr 2005. Ieee computer society conference on*, vol. 1. *Ieee* pp. 886–893.
82. Felzenszwalb P, Mcallester D, and Ramanan D, "a discriminatively trained, multiscale, deformable part model," in *computer vision and pattern recognition*, 2008. *Cvpr 2008. Ieee conference on*. *Ieee*, 2008, pp. 1–8.
83. P. F. Felzenszwalb, Girshick R B, and Mcallester D 2010 *Cascade object detection with deformable part models*, "in *computer vision and pattern recognition (cvpr) ieee conference on*. *Ieee* pp. 2241–2248.
84. Felzenszwalb P F, Girshick R B, Mcallester D, and Ramanan D 2010 *Object detection with discriminatively trained part-based models*, " *ieee transactions on pattern analysis and machine intelligence*, vol. 32, no. 9 pp. 1627– 1645
85. Girshick R, Donahue J, Darrell T and Malik J 2014 *Rich feature hierarchies for accurate object detection and semantic segmentation*, "in *proceedings of the ieee conference on computer vision and pattern recognition* pp 580– 587.
86. He K, Zhang X, Ren S, and Sun J 2014 *Spatial pyramid pooling in deep convolutional networks for visual 29 recognition*, "in *euclidean conference on computer vision*. Springer pp 346–361.
87. Girshick R 2015 *Fast r-cnn* in *proceedings of the ieee international conference on computer vision* pp. 1440–1448.
88. Ren S, He K, Girshick R, and Sun J 2015 *Faster r-cnn: Towards real-time object detection with region proposal networks*, "in *advances in neural information processing systems*, 2015 pp 91–99

89. Redmon J, Divvala S, Girshick R, and Farhadi A 2016 *You only look once: unified, real-time object detection*, " in *proceedings of the ieee conference on computer vision and pattern recognition* pp 779–788.
90. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C Y and Berg A C 2016 *Ssd: Single shot multibox detector*, " in *euclidean conference on computer vision*. Springer pp. 21–37
91. Lin T Y, Doll'ar P, Girshick R B, He K, Hariharan B, and Belongie S J 2017 *Feature pyramid networks for object detection*. " In *cvpr*, vol. 1, no. 2 p. 4.
92. Lin T Y, Goyal P, Girshick R, He K, and Doll'ar P 2018 *Focal loss for dense object detection*, " *ieee transactions on pattern analysis and machine intelligence*
93. Website Source: <https://arxiv.org/abs/1905.05055> (last updated on 15-Oct-2020)
94. Zhong B, Wu H, Li H, Sepasgozar S, Luo H and He L 2019 *A scientometric analysis and critical review of construction related ontology research*, *autom. Constr.* 101, <https://doi.org/10.1016/j.autcon.2017.04.002>. pp 17–31
95. Xing X, Zhong B, Luo H, Li H, Wu H 2019 *Ontology for safety risk identification in metro construction*, *comput. Ind.* 109, <https://doi.org/10.1016/j.compind.2019.04.001> pp 14–30
96. Fang W, Zhong B, Zhao N, Love P E, Luo H, Xue J, Xu S 2019 *A deep learning-based approach for mitigating falls from height with computer vision: convolutional neural network*, *adv. Eng. Inform.* 39 <https://doi.org/10.1016/j.aei.2018.12.005> pp 170–177
97. Fang Q, Li H, Luo X C, Ding L Y, Luo H B and Rose T M *Detecting nonhardhat-use by a deep learning method from far-field surveillance videos*, *autom. constr.* 85 <https://doi.org/10.1016/j.autcon.2017.09.018> pp 1–9,
98. Dung C V and Anh L D 2019 *Autonomous concrete crack detection using deep fully convolutional neural network*, *autom. Constr.* 99 <https://doi.org/10.1016/j.autcon.2018.11.028> pp 52–58