A Vision-Based Approach for Ensuring Proper Use of Personal Protective Equipment (PPE) in Decommissioning of Fukushima Daiichi Nuclear Power Station

Shi Chen * and Kazuyuki Demachi

Department of Nuclear Engineering and Management, School of Engineering, The University of Tokyo, Bunkyo-ku, Tokyo 113-8656, Japan; demachi@m.t.u-tokyo.ac.jp
* Correspondence: shichen@g.ecc.u-tokyo.ac.jp

Received: 26 June 2020; Accepted: 24 July 2020; Published: 26 July 2020

Abstract: Decommissioning of the Fukushima Daiichi nuclear power station (NPS) is challenging due to industrial and chemical hazards as well as radiological ones. The decommissioning workers in these sites are instructed to wear proper Personal Protective Equipment (PPE) for radiation protection. However, workers may not be able to accurately comply with safety regulations at decommissioning sites, even with prior education and training. In response to the difficulties of on-site PPE management, this paper presents a vision-based automated monitoring approach to help to facilitate the occupational safety monitoring task of decommissioning workers to ensure proper use of PPE by the combination of deep learning-based individual detection and object detection using geometric relationships analysis. The performance of the proposed approach was experimentally evaluated, and the experimental results demonstrate that the proposed approach is capable of identifying decommissioning workers’ improper use of PPE with high precision and recall rate while ensuring real-time performance to meet the industrial requirements.

Keywords: decommissioning; Fukushima Daiichi nuclear power plant; personal protective equipment; nuclear safety; deep learning

1. Introduction

On 11 March 2011, the Great East Japan Earthquake damaged the electric power supply lines to the Fukushima Daiichi Nuclear Power Station (NPS). The tsunami that followed caused substantial destruction of the safety infrastructure on the site and led to the loss of both off-site and on-site electrical power, which resulted in the loss of cooling functions at the reactors and the spent fuel pools. Large amounts of radioactive materials were released, and the problem of radioactive contamination has severely affected the lives of people and shocked many countries throughout the world. Units 1 to 4 of the Fukushima Daiichi NPS were damaged during the disaster, and all reactors were brought to cold shutdown on December 2011. Units 5 and 6 were permanently shut down on December 2013. On April 2014, the Fukushima Daiichi Decontamination & Decommissioning Engineering Company was formed by the Tokyo Electric Power Company (TEPCO), in partnership with Japanese Government and principle contractors, to perform the decommissioning of Fukushima Daiichi NPS.

Decommissioning of Fukushima Daiichi NPS is challenging work that has never been done before, which presents industrial and chemical hazards as well as radiological ones, and indeed these hazards generally present high risk to workers [1]. At present, many TEPCO workers, nuclear reactors manufacturers, construction companies, and their contractors are engaged in the decommissioning project for the Fukushima Daiichi NPS and are consequently exposed to various health risks. At present,
the Fukushima Daiichi NPS site is divided into three zones (see Figure 1a), and workers are instructed
to wear proper Personal Protective Equipment (PPE) when working in different zones (as illustrated in
Figure 1b [2]). The designated zones (and the associated PPE) are listed below.

1. G Zone: general work uniforms are required, e.g., disposable dust masks.
2. Y Zone: coveralls are required, e.g., full-face dust masks or half-face masks.
3. R Zone: anorak and full-face masks are required.

(a) The Fukushima Daiichi Nuclear Power Station (NPS) site is divided into three zones: R zone, Y zone,
and G zone.

(b) PPE requirements for each zone in Fukushima Daiichi NPS.

**Figure 1.** Each zone of the Fukushima Daiichi NPS site has different Personal Protective Equipment
(PPE) requirements that should be adhered to.

TEPCO indicates that each zone has different PPE requirements that should be adhered to. Workers moving from a higher-level contamination zone to a lower-level contamination zone are required to remove their PPE in changing rooms. Nonetheless, workers do not precisely follow the
on-site safety regulations due to all kinds of reasons, even if they have been previously educated and trained. Moreover, even though the signs and partitions mark the R zone and Y zone areas in the Fukushima Daiichi site, intrusions into these areas may occur within only a few minutes of carelessness. In addition, TEPCO has contracted various tasks to more than 20 companies (primary contractors), each of which, in turn, outsources some of its tasks to multiple layers of subcontractors. This complex structure could hinder consistent enforcement of on-site regulatory rules [3]. Traditional on-site occupational safety monitoring is usually carried out by on-site/off-site observers and relies heavily on the “human eye”, which is not sufficient to protect workers because of human factors and human errors (e.g., errors and unintentional mistakes, poor judgment and bad decision-making, and disregard for procedures). Thus, an automated monitoring approach is highly desirable in performing the safety monitoring of decommissioning workers to ensure PPEs are appropriately used in different zones.

This paper proposes a vision-based approach to automatically identify the proper use of PPE in response to the limitations of the safety monitoring systems at the decommissioning sites. The current goal of this work is to detect the hard hats and full-face masks in each image captured by surveillance camera and to identify whether the individuals on the decommissioning sites are wearing PPE properly. This paper will describe how an image of an individual’s posture is characterized using OpenPose [4] to extract a body’s keypoints. Then, the location of a hard hat and full-face mask is determined using the YOLOv3 model [5]. Finally, the geometric relationship between the PPE and the body are analyzed to determine whether the PPE is used appropriately.

2. Related Works

At present, several approaches have been investigated for automatic identification of proper PPE use [6–11], which are divided into sensor-based approaches and vision-based approaches. Sensor-based approaches primarily rely on remote locating and tracking techniques, e.g., radio frequency identification (RFID). Kelm et al. [6] designed a RFID-based approach for PPE compliance checking. RFID tags were attached to PPE, and the RFID readers were positioned at the site access. However, only an individual entering the construction site could be checked. Dong et al. [7] developed a real-time location system for worker positioning to determine whether the worker is required to wear a hard hat in this area. A pressure sensor was attached to the hard hat to determine whether it was being worn, and if not, to transmit a warning. Generally, existing sensor-based approaches have difficulty identifying proper PPE use for individuals on the sites. The practical implementation of the tags or sensors may lead to high costs due to the large number of devices required.

Vision-based approaches are performed by processing images captured by an on-site surveillance camera. Vision-based approaches are nonintrusive and require fewer devices compared to sensor-based approaches because of the widespread application of on-site surveillance cameras. Shrestha et al. [8] proposed a vision-based approach to detect the edge of objects inside the region of the upper head, i.e., hard hats, using edge detection algorithms. However, this approach relies on the facial features. Therefore, individuals without their full face visible cannot be recognized. Park et al. [9] introduced a vision-based non-hardhat-use (NHU) detection approach that uses background subtraction and the histogram of oriented gradients (HOG) features to simultaneously detect humans and hard hats in images. The detected human body area and the hard hat area were then matched for NHU detection. However, the hard hat area could not be successfully identified in the case of individuals in various postures (e.g., sitting, bending, and crouching down) or occlusion. Additionally, the approach relied on background subtraction, which makes it unable to detect workers standing at the site without any movement. In general, these approaches rely heavily on hand-crafted features and may consequently fail under complicated conditions with different viewpoints, different individual postures, weather variability, and occlusions, which are very common in construction sites.

In recent years, deep learning-based object detection methods have shown remarkable performance in visual tasks in the architecture, engineering, and construction (AEC) industry. Fang et al. [10] proposed an automated approach to detect construction workers’ NHU based on
a two-stage object detection model, Faster R-CNN [12]. The bounding rectangles that surround workers in the image were annotated as the ground truth to train the Faster R-CNN model. The NHU workers were detected, and other regions in the image were identified as the background in the testing phase. Wu et al. [11] deployed Single Shot Multibox Detector (SSD) architecture with reverse progressive attention (RPA) for hard hat detection. A benchmark dataset GDUT-HWD was generated to train the SSD-RPA model. However, existing deep learning-based detection approaches are mainly focused on learning to localize only PPE(s) or NHU individual(s) in the obtained images, which may fail to identify PPE in cases of uncommon human gestures or appearance. Furthermore, almost no research studies have been conducted concerning proper use identifications for multiple PPE, i.e., more than hard hats. In response to these limitations, the scope of this paper is devoted to proposing a novel approach for the identification of proper PPE use and evaluating the effectiveness and robustness of the proposed approach for identifying the proper use of hard hats and full-face masks in the decommissioning of Fukushima Daiichi NPS.

3. Methodology

In this section, the methodology to automatically identify the proper use of PPE is detailed, which follows the subsequent steps:

1. For each image captured by an on-site surveillance camera, individual(s) are detected, together with their keypoints coordinates, using an individual detection model.
2. PPE(s) are recognized and localized using an object detection model.
3. The proper use of PPE identification is performed by analyzing the geometric relationships of the individual’s keypoints and the detected PPE(s).

3.1. Individual Detection

The patterns to be estimated in conventional object detection-based approaches for the identification of proper PPE use [10,11] are the “Individual with PPE”, thus the image samples like “worker wearing hard hat” are required to train the object detection model, which are not as easily obtained as the information security of the construction sites. In contrast to these approaches, we employ an individual detection model in this work to specify individual features. This strategy makes the model of this work more robust with viewpoint changes and different individual postures based on the specified human pose by the individual detection model.

We characterize the decommissioning workers’ postures by extracting the body parts’ keypoints of the person detected in on-site images using OpenPose [4]. Instead of the conventional top-down approaches which first detect persons in the image and perform single-person pose estimation for each detected person, OpenPose provides a bottom-up approach by detecting all body parts in the image and associating them with a different person. As illustrated in Figure 2, OpenPose takes the image of size \(w \times h\) as input and processes images through a two-branch multi-stage convolutional neural network (CNN) to predict confidence maps for body part detection and part affinity fields (PAF) for body parts association. First, the backbone (VGG-19 [13] in the original paper) generates a set of feature maps \(F\) from the raw image. Subsequently, the pipeline is divided into multiple similar stages. There are two branches for each stage: (1) the branch to predict a set of confidence maps \(S^t\) for the candidates of joint position, and (2) the branch to produce a set of PAFs \(L^t\) for joint points relationships correlation. At the first stage, the feature map \(F\) is input to the CNNs to generate a set of confidence maps \(S^1\) and a set of PAFs \(L^1\). In each subsequent stage \(t (t > 1)\), the predictions of the previous stage, \(S^{t−1}\) and \(L^{t−1}\), are concatenated with the feature map \(F\) and fed into the current stage. The loss functions of stage \(t\) are

\[
\begin{align*}
    f^t_S &= \sum_{j=1}^{J} \sum_p W(p) \cdot \left\| S^t_j(p) - S^*_j(p) \right\|^2_2, \\
    f^t_L &= \sum_{c=1}^{C} \sum_p W(p) \cdot \left\| L^t_c(p) - L^*_c(p) \right\|^2_2. 
\end{align*}
\]
where $S^*_j$ and $L^*_j$ are the ground truth part confidence map and ground truth PAF of the $j$th joint point, respectively. $W$ is a binary mask with $W(p) = 0$ when the annotation is missing at an image location $p$. The overall loss function is

$$f = \sum_{t=1}^{T} \left( f^S_t + f^L_t \right)$$

(2)

**Figure 2.** Architecture of the OpenPose Model to extract the body parts’ keypoints of the person detected in images.

Finally, the confidence maps and the PAFs are parsed to output the 2D keypoints for all persons in the image by greedy inference (18 keypoints, pretrained using COCO 2016 keypoints challenge dataset [14], see Figure 3).

The choice of OpenPose is motivated by its functionality on an RGB image or video taken by on-site surveillance cameras in real-time. This provides a huge benefit in comparison with the skeletal tracking capability of RGB-D devices (e.g., Microsoft Kinect [15]) which depend on depth information.

To further improve the speed of estimation, we deploy a lightweight architecture, Mobilenetv2 [16], as the feature extractor instead of VGG-19. To achieve model acceleration, Mobilenetv2 employs a special convolutional filter called depthwise separable convolution as the replacement of the standard convolutional filter together with linear bottleneck ($1 \times 1$ convolutional layer without ReLU) to solve the problem of information loss due to nonlinear activation functions. First, a pointwise ($1 \times 1$) convolution is deployed to expand the low-dimensional input feature map to a higher-dimensional space suited to nonlinear activations. Next, a depthwise convolution is performed using $3 \times 3$ kernels to achieve spatial filtering of the higher-dimensional tensor. Finally, the spatially-filtered feature map is projected back to a low-dimensional subspace using another pointwise convolution.
3.2. PPE Detection

PPE(s) recognition and localization are carried out using an object detection model. Depending on the CNN architecture and the detection strategy, object detection approaches are broadly classified into single-stage and two-stage approaches. Two-stage approaches (e.g., the R-CNN family [12,17,18]) make predictions in two stages: first, a set of regions of interests are proposed by regional proposal network or select search, which are sparse as the candidates of potential bounding box can be infinite. Subsequently, a classifier is deployed to process the region candidates. On the other hand, one-stage approaches (e.g., YOLO and its variants [5,19,20]) make detection directly over a dense sampling of possible locations without the region proposal stage, which leads to a simpler architecture and makes it extremely fast in the inference phase.

In 2018, Redmon et al. proposed YOLOv3 [5] as an improvement of the YOLO family. As demonstrated in Figure 4, YOLOv3 uses a deep network architecture with residual blocks and a total of 53 convolutional layers, i.e., Darknet-53, for feature extraction, which has better performance and is 1.5× faster than ResNet-101 [21]. Drawing on the idea of feature pyramid networks [22], YOLOv3 makes predictions at three scales (“predictions across scales”) by downsampling the size of the input image by 32, 16, and 8. This means, with an input of 416 × 416, the YOLOv3 model provides feature maps of 13 × 13, 26 × 26, and 52 × 52, respectively, in three output layers.

The predictions of the YOLOv3 model are encoded as a 3-D tensor with bounding box candidate, objectness score, and class score (see the bottom right of Figure 4):

\[ N \times N \times (3 \times (4 + 1 + C)) \] (3)

where \( N \times N \) is the number of the grid cells of the model and \( C \) is the number of the classes to be classified.
Figure 4. Architecture of YOLOv3 Model to Detect PPE(s) in Images.

In contrast to the sum of squared errors for classification terms used in YOLO and YOLOv2, YOLOv3 uses logistic regression to predict a confidence score for each bounding box. The Softmax classifier deployed in the YOLOv2 assumes that a target belongs to only one class, and each box is assigned to the class with the largest score. However, in some complex scenarios, a target may belong to multiple classes (with overlapping class labels), thus YOLOv3 employs a multiple independent logistic classifier (using Sigmoid function) for each class instead of one softmax layer when predicting class confidence. In addition, YOLOv3 uses binary cross-entropy loss as the loss function to train class prediction (both YOLO and YOLOv2 use a loss function based on the sum of squares):

$$
\lambda_{\text{coord}} \sum_{i=0}^{N^2} \sum_{j=0}^{B} \mathbf{1}_{i,j} [\frac{(b_x - \hat{b}_x)^2 + (b_y - \hat{b}_y)^2 + (b_w - \hat{b}_w)^2 + (b_h - \hat{b}_h)^2}{2}]
$$

$$
+ \sum_{i=0}^{N^2} \sum_{j=0}^{B} \mathbf{1}_{i,j} \left[ -\log(p_c) + \sum_{i=1}^{n} \text{BCE}(\hat{C}_i, C_i) \right]
$$

$$
+ \lambda_{\text{noobj}} \sum_{i=0}^{N^2} \sum_{j=0}^{B} \mathbf{1}_{i,j} \left[ -\log(1 - p_c) \right]
$$

where $N^2$ is the number of the grid cells (13 $\times$ 13 grids, 26 $\times$ 26 grids, or 52 $\times$ 52 grids), and $B$ is the bounding boxes. $\mathbf{1}_{i,j}$ denotes that cell $i$ contains objects and the $j_{th}$ bounding box predictor in cell $i$ is “responsible” for that prediction. In contrast, $\mathbf{1}_{i,j}^{\text{noobj}}$ denotes that the $j_{th}$ bounding box predictor in cell $i$ that contains no objects. The parameters $\lambda_{\text{coord}}$ and $\lambda_{\text{noobj}}$ are employed to revise the overall loss by increasing the loss from bounding box coordinate predictions and decreasing the loss from confidence predictions for boxes that contain no objects. Additionally, binary cross-entropy loss $\text{BCE}(\hat{C}_i, C_i)$ is given by

$$
\text{BCE}(\hat{C}_i, C_i) = -\hat{C}_i \log(C_i) - (1 - \hat{C}_i) \log(1 - C_i)
$$

where $C_i$ is the objectness in cell $i$. 
In summary, the predictions of YOLOv3 are carried out by one single network that can be easily trained end-to-end to improve performance. High efficiency and speed make YOLOv3 a reasonable option for real-time processing for industrial purposes.

3.3. Identification of Proper PPE Use

To identify whether PPE(s) are appropriately used by decommissioning worker(s) in the image captured by an on-site surveillance camera, we analyze the geometric relationships between spatial structure characteristics of detected PPE(s) and individual(s).

We formulate the output of YOLOv3 as a set of PPE bounding boxes $B = \{B_1, B_2, ..., B_l\}$, where $l$ is the number of detected PPE in the image. Each bounding box $B_i = (x_i, y_i, w_i, h_i, c_i), i \in \{1, 2, ..., I\}$ contains five elements, where $(x_i, y_i)$ and $(w_i, h_i)$ are the bounding boxes’ position and size, respectively, and $c_i$ represents the class of PPE in the bounding box. Let $H = \{H_1, H_2, ..., H_I\}$ be the set of the detected individual(s) via OpenPose, where $J$ is the number of detected individual(s) in the obtained frame and $H_j = \{(x^{(0)}_j, y^{(0)}_j), (x^{(1)}_j, y^{(1)}_j), ..., (x^{(17)}_j, y^{(17)}_j)\}$ represents the detected body parts of the $j$th individual (see Figure 3). We associate a detected PPE $i^*$ to a specific individual $j^*$ by searching the minimum Euclidean distance between bounding boxes $B$ and detected neck keypoints $H^{(1)} = \{(x^{(1)}_0, y^{(1)}_0), (x^{(1)}_1, y^{(1)}_1), ..., (x^{(1)}_{j^*}, y^{(1)}_{j^*})\}$ (body part 1 in Figure 3) that satisfy the geometric constraints ($y_{i^*} < y^{(1)}_{j^*}$) to make sure the detected PPE $i^*$ is in the upper position of the detected individual $j^*$ to be associated:

$$i^*, j^* = \arg\min_{i \in \{1, 2, ..., I\}, j \in \{1, 2, ..., J\}} \sqrt{(x_i - x^{(1)}_{j})^2 + (y_i - y^{(1)}_{j})^2}$$

s.t. $y_{i^*} < y^{(1)}_{j^*}$

(6)

Subsequently, a distance is measured to determine whether each individual is using their associated PPE. We take advantage of Euclidean distance using the detected neck keypoints and hips keypoints (body parts 8 and 11 in Figure 3) of the detected individual $j^*$ as a dynamic reference threshold, which will keep changing synchronously when the distance between the individual and the camera changes:

$$\beta_{i^*j^*} = \max(\sqrt{(x^{(8)}_{i^*} - x^{(8)}_{j^*})^2 + (y^{(8)}_{i^*} - y^{(8)}_{j^*})^2}, \sqrt{(x^{(11)}_{i^*} - x^{(11)}_{j^*})^2 + (y^{(11)}_{i^*} - y^{(11)}_{j^*})^2}) \cdot \gamma$$

(7)

where $\gamma$ is the scaling coefficient to strike different PPE identification, which is set to 0.8 or 0.6 for hard hats or full-face masks, respectively.

If the Euclidean distance between the position $(x_{i^*}, y_{i^*})$ of the bounding box of $i^*$ and detected neck keypoint (body keypoints 1 in Figure 5a) of $j^*$ is smaller than the reference threshold $\beta_{i^*j^*}$, then the detected PPE $i^*$ is identified as appropriately used by the detected individual $j^*$ (Figure 5a); otherwise, even though PPE $i^*$ is associated with individual $j^*$, the condition of individual $j^*$ is identified as not using PPE properly (Figure 5b):

$$c_{i^*j^*} = \begin{cases} "\text{PPE proper use}" & \text{if } d(i^*, j^*) < \beta_{i^*j^*} \\ "\text{PPE improper use}" & \text{if } d(i^*, j^*) \geq \beta_{i^*j^*} \end{cases}$$

(8)

where

$$d(i^*, j^*) = \sqrt{(x_{i^*} - x^{(1)}_{j^*})^2 + (y_{i^*} - y^{(1)}_{j^*})^2}$$

(9)
Appl. Sci. 2020, 10, 5129 9 of 14

(a) Proper PPE use. 
(b) Improper PPE use.

Figure 5. Proper PPE use identification strategies. Reference threshold $\beta_{i \leftrightarrow j}^*$ is calculated from the Euclidean distance among detected neck keypoints (yellow dots) and hip keypoints (purple and blue dots). (a) Proper PPE use is identified if the Euclidean distance (red lines) between the detected PPE positions (red dots) and detected neck keypoints is smaller than the reference threshold $\beta_{i \leftrightarrow j}^*$; (b) otherwise, the condition is identified as improper PPE use.

4. Experimental Procedure

4.1. Experimental Dataset

To create the training dataset of PPE detection, we collected hard hat and full-face mask images from two sources: (1) Internet images retrieved using the web crawler, and (2) real-world images captured using the webcam. A total of 3808 images were collected (Table 1) and annotated to train a YOLOv3 model.

Table 1. Information of collected training dataset.

|                | Number of the Internet Images | Number of the Real-World Images | Total  |
|----------------|-------------------------------|--------------------------------|--------|
| Hard hat       | 933                          | 1209                           | 2142   |
| Full-face mask | 642                          | 1021                           | 1666   |

Regions of decommissioning worker could be captured at different resolutions in the images as the surveillance cameras are installed at different locations on the decommissioning site and the trajectory of workers is random. Thus, different distance conditions (3 m, 5 m, and 7 m) were considered in our experiments to validate the robustness of our proposed approach. The impact of individual posture was also taken into consideration in current experiments and three common worker postures—standing, bending, and squatting—were included in the testing dataset. To create the testing dataset in such a way that it could validate the performance of the trained model, three volunteers were instructed to perform different postures while wearing PPEs (or not) at different distances from the camera. The details are provided in Table 2, where positive samples refer to the individuals who are wearing PPE properly and negative samples refer to the individuals who are not wearing PPE. Finally, we randomly selected 500 images for each case from the collected image sequences and created a testing dataset.

4.2. Evaluation Metrics

We adopted precision and recall to evaluate the performance of the proposed approach:

$$\text{Precision} = \frac{TP}{TP + FP}$$ (10)
Recall = \frac{TP}{TP + FN} \quad (11)

where \( TP \) (true positive) is defined as the number of correct identification of individuals who are wearing PPE. \( FP \) (false positive) is the number of individuals who are not wearing PPE properly but are misidentified as wearing PPE properly, and \( FN \) (false negative) is the number of not detected ground truth of PPE proper use individuals, as defined in Table 3.

| Impacts     | No. | Value | Images | Hard Hat | Full-Face Mask |
|-------------|-----|-------|--------|----------|---------------|
|             |     |       |        | Positive Samples | Negative Samples | Positive Samples | Negative Samples |
| Distance    | 1   | 3 m   | 500    | 684      | 816           | 500             | 1000            |
|             | 2   | 5 m   | 500    | 770      | 730           | 500             | 1000            |
|             | 3   | 7 m   | 500    | 818      | 682           | 500             | 1000            |
| Individual posture | 1 | Standing | 500 | 1582 | 1813 | 972 | 2417 |
|             | 2   | Bending | 500    | 277      | 168           | 180             | 284             |
|             | 3   | Squatting | 500 | 413 | 247 | 348 | 299 |

Table 2. Information of collected testing dataset.

| Predicted | Ground Truth |
|-----------|--------------|
| TP        | Proper use   |
| FP        | Proper use   |
| FN        | Improper use |

Table 3. Definitions of true positive (TP), false positive (FP), and false negative (FN).

4.3. Implementation Details

We built the YOLOv3 model using TensorFlow [23] and initialized it based on pretrained weights on the ImageNet dataset [24]. Training of YOLOv3 was performed in two stages: (1) all convolutional layers were first frozen up to the last convolutional block in Darknet-53, and the model was trained with frozen layers to get a stable loss in 50 epochs; (2) all convolutional layers of Darknet-53 proceeded to unfreeze to perform fine-tuning in 50 epochs. The learning rate schedule is as follows. For the first stage, the model was trained with a learning rate of \( 1 \times 10^{-3} \); for the second stage, the model was trained with a learning rate started at \( 1 \times 10^{-4} \). An Adam optimizer [25] with a batch size of 8 was adopted throughout training.

5. Results and Discussion

Figure 6 illustrates the identification results of the examples on the testing dataset. Meanwhile, our model was able to run at \(~7.95\) FPS in a machine with a GeForce GTX 1080 Max-Q GPU, which is sufficient for real-time processing on the decommissioning site.

5.1. Impact of Distance

The identification results under different distances are reported in Table 4. The resolution of the individual regions in the image gradually became smaller as the distance between the camera and individuals increased. The precision and recall rate for the identification of proper hard hat use gradually decreased with increasing distance from the camera, but the precision was above 95% for all measured distances, while the recall rate was above 90% except for measurements at 7 m. The overall performance of the identification of proper hard hat use is acceptable (precision rate: 97.59%; recall rate: 89.21%). For the identification of proper full-face mask use, the performance of our approach declined only slightly as distance increased; the precision and recall rates remained higher than 97%.
Figure 6. Identification examples on the testing dataset.

Table 4. Identification results under different distance.

| Categories       | Distance | TP  | FP  | FN  | Precision (%) | Recall (%) |
|------------------|----------|-----|-----|-----|---------------|------------|
| Hard hat         | 3 m      | 666 | 27  | 18  | 96.10         | 97.37      |
|                  | 5 m      | 695 | 36  | 75  | 95.08         | 90.26      |
|                  | 7 m      | 665 | 15  | 153 | 97.79         | 81.30      |
| Full-face mask   | 3 m      | 499 | 0   | 1   | 100           | 99.80      |
|                  | 5 m      | 500 | 7   | 0   | 98.62         | 100        |
|                  | 7 m      | 487 | 0   | 13  | 100           | 97.40      |

5.2. Impact of Individual Posture

The identification results for different individual postures are shown in Table 5. The high overall precision showed in the results represents the excellent performance of our model for various individual postures. For the identification of proper hard hat use, the recall rate for the squatting position is lower than the others because of failures of the body parts detection via OpenPose. However, the recall rate is still above 81%. The identification results indicate that the impact of individual posture has little effect on the identification of proper full-face mask use performance, as the precision and recall rate remaining robust in different individual postures.

The overall precision and recall rates are 97.64% and 93.11%, respectively, which demonstrates the robustness of the proposed approach in the identification of proper PPE use at different distances and individual postures.
Table 5. Identification results under different individual postures.

| Categories       | Individual’s Posture | TP  | FP  | FN  | Precision (%) | Recall (%) |
|------------------|----------------------|-----|-----|-----|---------------|------------|
| Hard hat         | Standing             | 1435| 78  | 147 | 94.84         | 90.71      |
|                  | Bending              | 247 | 0   | 30  | 100           | 89.17      |
|                  | Squatting            | 344 | 0   | 69  | 100           | 81.29      |
| Full-face mask   | Standing             | 968 | 7   | 4   | 99.28         | 99.59      |
|                  | Bending              | 179 | 0   | 1   | 100           | 99.44      |
|                  | Squatting            | 339 | 8   | 9   | 97.69         | 97.41      |

6. Conclusions

This paper has presented a novel vision-based approach to address the difficulties of proper PPE use management in the decommissioning of the Fukushima Daiichi NPS. First, we created a dataset using Internet images and real-world images to train the YOLOv3 model to recognize hard hats and full-face masks. Subsequently, we conducted the identification of proper PPE use using geometric relationships of the outputs of OpenPose and YOLOv3. The performance of the proposed approach was experimentally evaluated under various distance and individual posture conditions. The experimental results indicate that the proposed approach was capable of identifying the decommissioning workers who are not wearing PPE properly with high precision (97.64%) and recall rate (93.11%), while ensuring real-time performance (7.95 FPS on average). The scope for on-site occupational safety monitoring was determined to be preliminary. This work has great prospects with a wide range of applications, e.g., proper PPE use management in a COVID-19 hospital facility. Future studies are to be investigated following the consideration of the augmentation of the training dataset to increase the safety of the monitored targets, e.g., safety gloves and anoraks. Furthermore, the on-site system development and implementation are suggested to be performed considering the deployment IoT platform, e.g., the Microsoft Azure IoT [26], to perform the inference of our identification model.

Author Contributions: Conceptualization, S.C. and K.D.; methodology, S.C.; software, S.C.; validation, S.C.; investigation, S.C.; resources, K.D.; data curation, S.C.; writing—original draft preparation, S.C.; writing—review and editing, S.C. and K.D.; visualization, S.C.; supervision, K.D.; project administration, K.D.; funding acquisition, K.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Japan Society for the Promotion of Science (JSPS) Grants-in-Aid for Scientific Research Grant No. 19K05324.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript.

PPE  Personal Protective Equipment
NPS  Nuclear Power Station
TEPCO  Tokyo Electric Power Company
RFID  radio frequency identification
NHU  non-hardhat-use
HOG  histogram of oriented gradients
AEC  architecture, engineering and construction
WOI  worker-of-interest
SSD  Shot Multibox Detector
RPA  reverse progressive attention
CNN  convolutional neural network
PAF  part affinity fields
References

1. Safety Assessment for Decommissioning; Number 77 in Safety Reports Series; INTERNATIONAL ATOMIC ENERGY AGENCY: Vienna, Austria, 2013.
2. Efforts to Improve Working Environment and Reduce Radiation Exposure at Fukushima Daiichi Nuclear Power Station; Technical Report; Tokyo Electric Power Company: Tokyo, Japan, 2016.
3. Mori, K.; Tateishi, S.; Hiraoka, K. Health issues of workers engaged in operations related to the accident at the Fukushima Daiichi Nuclear Power Plant. In Psychosocial Factors at Work in the Asia Pacific; Springer: Cham, Switzerland, 2016; pp. 307–324.
4. Cao, Z.; Hidalgo, G.; Simon, T.; Wei, S.E.; Sheikh, Y. OpenPose: Realtime multi-person 2D pose estimation using Part Affinity Fields. arXiv 2018, arXiv:1812.08008.
5. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. arXiv 2018, arXiv:1804.02767.
6. Kelm, A.; Laußat, L.; Meins-Becker, A.; Platz, D.; Khazaee, M.J.; Costin, A.M.; Helmus, M.; Teizer, J. Mobile passive Radio Frequency Identification (RFID) portal for automated and rapid control of Personal Protective Equipment (PPE) on construction sites. Autom. Constr. 2013, 36, 38–52. [CrossRef]
7. Dong, S.; He, Q.; Li, H.; Yin, Q. Automated PPE misuse identification and assessment for safety performance enhancement. In Proceedings of the ICCREM 2015, Lulea, Sweden, 11–12 August 2015; pp. 204–214.
8. Shrestha, K.; Shrestha, P.P.; Bajracharya, D.; Yfantis, E.A. Hard-hat detection for construction safety visualization. J. Constr. Eng. 2015, 141, 1–8. [CrossRef]
9. Park, M.W.; Elsafty, N.; Zhu, Z. Hardhat-wearing detection for enhancing on-site safety of construction workers. J. Constr. Eng. Manag. 2015, 141, 04015024. [CrossRef]
10. Fang, Q.; Li, H.; Luo, X.; Ding, L.; Luo, H.; Rose, T.M.; An, W. Detecting non-hardhat-use by a deep learning method from far-field surveillance videos. Autom. Constr. 2018, 85, 1–9. [CrossRef]
11. Wu, J.; Cai, N.; Chen, W.; Wang, H.; Wang, G. Automatic detection of hardhats worn by construction personnel: A deep learning approach and benchmark dataset. Autom. Constr. 2019, 106, 102894. [CrossRef]
12. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 7–12 December 2015; pp. 91–99.
13. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. arXiv 2014, arXiv:1409.1556.
14. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft coco: Common objects in context. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; Springer: Cham, Switzerland, 2014; pp. 740–755.
15. Kinect for Windows. Available online: https://developer.microsoft.com/en-us/windows/kinect/ (accessed on 22 July 2020).
16. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520.
17. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
18. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Las Condes, Chile, 11–18 December 2015; pp. 1440–1448.
19. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 779–788.
20. Redmon, J.; Farhadi, A. YOLO9000: Better, faster, stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 7263–7271.
21. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778.
22. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.

23. Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M.; et al. Tensorflow: A system for large-scale machine learning. In Proceedings of the 12th USENIX symposium on operating systems design and implementation (OSDI 16), Savannah, GA, USA, 2–4 November 2016; pp. 265–283.

24. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. Imagenet large scale visual recognition challenge. Int. J. Comput. Vis. 2015, 115, 211–252. [CrossRef]

25. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. arXiv 2014, arXiv:1412.6980.

26. Microsoft Azure IoT. Available online: https://azure.microsoft.com/en-us/overview/iot/ (accessed on 6 June 2020).