Risk Factor Recognition for Automatic Safety Management in Construction Sites Using Fast Deep Convolutional Neural Networks

Jeongeun Park, Hyunjae Lee and Ha Young Kim *

Graduate School of Information, Yonsei University, Yonsei-ro 50, Seodaemun-gu, Seoul 03722, Korea; park.je@yonsei.ac.kr (J.P.); jaylee4274@yonsei.ac.kr (H.L.) * Correspondence: hayoung.kim@yonsei.ac.kr; Tel.: +82-2-2123-4194; Fax: +82-2-2123-8654

Abstract: Many industrial accidents occur at construction sites. Several countries are instituting safety management measures to reduce industrial accidents at construction sites. However, there are few technical measures relevant to this task, and there are safety blind spots related to differences in human resources’ capabilities. We propose a deep convolutional neural network that automatically recognizes possible material and human risk factors in the field regardless of individual management capabilities. The most suitable learning method and model for this study’s task and environment were experimentally identified, and visualization was performed to increase the interpretability of the model’s prediction results. The fine-tuned Safety-MobileNet model showed a high performance of 99.79% (30 ms), demonstrating its high potential to be applied in actual construction sites. In addition, via visualization, the cause of the model’s confusion of classes could be found in a dataset that the model did not predict correctly, and insights for result analysis could be presented. The material and human risk factor recognition model presented in this study can contribute to solving various practical problems, such as the absence of accident prevention systems, the limitations of human resources for safety management, and the difficulties in applying safety management systems to small construction companies.

Keywords: construction site; risk factors; safety management; deep learning; convolutional neural network; visualization

1. Introduction

Global economic growth over the years has created industries in various fields, and the increase in the number of industrial companies and workers has generally resulted in improved working conditions. However, more than 2 million people worldwide die every year from various industrial accidents and work-related diseases [1]. The influx of unskilled workers into industrial sites and insufficient safety education are causing industrial accidents and deaths [2–4]. These occupational accidents not only cause inconvenience and potential harm to workers, but also increase the costs that companies and the state have to bear owing to loss of labor and productivity disruptions. The cost of work-related injuries in 2019 totaled 171 billion dollars, including 53.9 billion dollars in lost wages and productivity, 35.5 billion dollars in medical expenses, and 59.7 billion dollars in administrative costs. Calculated as lost time, the total number of days lost because of injuries was 70,000,000. In particular, among various industries, construction sites account for a high proportion, i.e., approximately 17% of the total fatalities due to industrial accidents [5]. According to the International Labor Organization, at least 60,000 people die each year on construction sites, which equates to one death every 10 min. Many countries are trying to prepare countermeasures to address the risks and the high economic loss related to such industrial accidents. The United States enacted the Occupational Safety and Health Act in 1970, and established the Occupational Safety and Health Administration (OSHA), an...
affiliate of the Department of Labor. OSHA has set the goal of creating a safe and healthy work environment as a five-year strategic plan for 2018–2022. This includes strengthening penalties for employers who do not comply with safety and health regulations and implementing deterrence strategies against employers with a history of willful and repeated violations of the law [6]. The United Kingdom has established and implemented an HSE Business Plan every year. Most recently, the focus has been on the development of guidelines for risk management and control based on real-world cases as part of the guidance for improving the level of safety and health in the workplace [7]. In Germany, for effective implementation of policies related to occupational safety and health, the German joint occupational safety and health strategy was established. In particular, for 2019–2023, a single goal of risk assessment was set, and risk assessment commonly used throughout Germany was introduced to increase the consistency and efficiency of management [8]. In Korea, the construction sector was given priority to revise the legislation to reduce fatal accidents by 2022. On 8 January 2021, the National Assembly passed the Critical Disaster Company Penalty Act, which aims to secure workers’ safety rights and prevent serious accidents caused by insufficient safety management systems [9]. In accordance with the 13th Labor Accident Prevention Plan, Japan aimed to reduce fatal accidents by 15% and minor accidents by more than 5%. In particular, measures for preventing fall accidents and safety measures for dismantling work were reinforced in the construction industry [10]. As such, protecting the life and safety of workers and preventing occupational accidents are common goals of the state and industry, and it is very important to lay a foothold for reducing serious accidents through collaboration between the government and companies.

However, at many actual construction sites in a large number of countries, there are no technical measures for an industrial accident prevention system for construction safety. In Korea, which is used as the reference country in this study, it is particularly difficult to address safety concerns at small construction sites, which are blind spots for safety management. Although more than half of industrial accident deaths at construction sites occur in small-scale construction projects costing less than 1.7 million dollars, there is no systematic operation-based environment that can be systematically and organically utilized. In addition, the aging of experts capable of safety management, the difficulties in supplying and receiving professional manpower, and the limitations in the accuracy of technical guidance are also causes for concern. Construction-related jobs are considered ‘blue-collar jobs’, so there are few applicants for these jobs, especially in Korea. Furthermore, the worker’s average age is constantly increasing, so there is concern about a considerable gap if a large number of skilled professionals retire [11,12]. Therefore, this study intends to develop an algorithm that automatically recognizes risk factors by applying a deep convolutional neural network (DCNN) to construction sites. Through this work, it is possible to change the paradigm from the existing experience-dependent industry to a knowledge-based and high-tech industry, and to prepare technical measures to overcome the limitations of manpower. The contributions of our study are as follows.

- We defined material and human factors that cause industrial accidents in the construction industry and proposed an algorithm that automatically recognizes them. As far as we know, this is the first study that covers both factors.
- Our proposed model achieves high performance (99.79%) on the dataset collected at the actual industrial sites. As it is a fast and lightweight model, it is suited for use in real-time at the actual industrial sites.
- Through visualization, the factors that most influence the final classification decision of the model were identified to increase the user’s understanding of the process or basis for which the result was derived, and the interpretability of the model was expanded.
2. Related Work

2.1. Studies to Prevent Industrial Accidents

To prevent industrial accidents, it is essential to be able to find and eliminate the causes of accidents. It is inappropriate to diagnose the cause of a disaster simply as an individual's careless behavior or negligence in safety management. A study was conducted to determine the root cause of industrial accidents. In 1959, Heinrich proposed Domino’s Theory, which states that disasters are the result of a chain reaction of accident factors [13]. As shown in Figure 1, the theory consists of five stages, and the root cause of accidents was considered to be genetic factors and the social environment. According to Domino’s Theory, most accidents and disasters can be prevented if the third stage, unsafe behavior (human factors) and unsafe conditions (material factors), can be controlled. Domino theory has been used around the world as the basis of safety management and education for over 30 years. However, Heinrich’s theory had limitations in that it gave excessive importance to human factors as the cause of the disasters. The Bird and Loftus modified Domino theory is similar to Heinrich’s perspective, i.e., that each disaster cause is a result of higher-level factors in a chain reaction [14]. The difference is that in addition to human factors, control and management aspects were emphasized as the root causes of accidents.

As mentioned above, given that an accident is a result caused by a specific cause, it is highly important to identify material and human risk factors that can cause accidents. Accordingly, many previous studies have focused on identifying risk factors. First, İssever et al. (2008) experimented on identifying the causes of accidents at construction sites by focusing on human risk factors [15]. A survey such as the Eysenck personality questionnaire and brief symptom inventory was carried out on 50 injured workers who had visited the infirmary from a construction site employing 1200 workers. In the control group, the same survey was conducted with 150 randomly selected workers within the same site. It was found that there were significant differences in the variables of visual cognition, neuroticism, and tiredness within 24 h between the accident occurrence group and the control group. Hamid et al. (2008) investigated the causes of accidents at construction sites by analyzing surveys, literature studies, and accident reports. As a result, material factors such as unsafe equipment, human factors such as negligence and non-wearing of personal protective equipment, and lack of safety management were pointed out as the main causes [16]. Based on previous studies that the risk–accident–injury relationship is fundamentally linked, Chi et al. (2013) showed that human factors can explain or link with each other, tools or equipment (material factors), working conditions and environments. In other words, unsafe worker behavior and unsafe working conditions are closely related and cause construction accidents [17]. Another study proposed a model to identify the root cause of construction site accidents. Umar and Egbru (2018) defined the four types of causes of accidents: equipment/materials, workers, environment, and management. As a result of investigating the frequency of accidents for each type, it was found that 41% of all accidents were worker-related factors accounted for the largest proportion, followed by management.
and equipment/material [18]. In addition, there have been many studies on the causes of disasters. However, these studies aimed to identify the causes of disasters theoretically, so it is difficult to apply them directly to construction sites and observe the empirical effect. Nevertheless, based on the findings of these studies that human and material factors have a great influence on the occurrence of accidents and have a close influence on each other, this study attempts to confirm whether the safety of the two factors that must be recognized to prevent accidents is secured.

2.2. Studies Applying Deep Learning Algorithms to Safety Management

Deep learning is a sub-field of artificial intelligence, involving a model composed of multiple layers of neural networks and is characterized by a powerful learning performance for abstract representations of data. It has achieved the best performances in various fields such as object detection, natural language processing, speech recognition, and time series forecasting [19]. In particular, convolutional neural network (CNN)-based models are widely used in image classification and object detection. In 2012, a DCNN-based model in the ImageNet large-scale visual recognition challenge attracted considerable attention as it achieved tremendous performance improvement compared to the existing model [20]. Since 2015, when DCNN-based models such as ResNet and GoogLeNet were suggested, performances of the models have surpassed human cognitive abilities. Deep learning algorithms are gradually being applied to various fields based on their excellent performance.

Some studies applied deep learning in construction safety management. As a human factor-related study, Fang et al. (2018) used faster RCNN, a modified model of CNN, to recognize workers from construction site images and quickly recognize whether they were wearing a safety helmet or not. It achieved high performance with a misclassification rate of about 5% [21]. Nath et al. (2020) recorded high performance using the YOLO-v3 model, which is a real-time object detection model, for multi-label classification of workers, safety helmets, and vests [22]. There was a study combining CNN and other deep learning algorithms. Ding et al. (2018) recognized the unsafe behavior of construction site workers through a hybrid deep learning model that combined CNN and long short-term memory, and showed excellent model performance, surpassing the performance of the existing state-of-the-art models by 10% [23]. Luo et al. (2018) detected construction work objects with a CNN-based feature extractor and identified the relationship among the detected objects using network analysis [24]. It was successful at extracting various activity relation patterns that reflected the interactions between workers, materials, and equipment, but it did not focus on safety factor detection.

While the above studies focused on human factors, studies on material factors have also been conducted. Kolak, Chen, and Luo (2018) detected the unsafe behavior of construction site workers through a hybrid deep learning model that combined CNN and long short-term memory, and showed excellent model performance, surpassing the performance of the existing state-of-the-art models by 10% [23]. Luo et al. (2018) detected construction work objects with a CNN-based feature extractor and identified the relationship among the detected objects using network analysis [24]. It was successful at extracting various activity relation patterns that reflected the interactions between workers, materials, and equipment, but it did not focus on safety factor detection.

As mentioned above, deep learning models have been applied to safety management studies and showed good results. However, there are practical difficulties in applying these models to actual construction sites as a single framework because most studies have only focused on specific risk factors. Although Zhong’s (2020) study included relatively various risk factor labels, this methodology cannot conclusively reduce accidents when applied...
to construction sites because it mainly focused on classifying and identifying accident narratives by analyzing accident reports [31]. To overcome the limitations of existing studies, we propose an integrated recognition model that includes various risk factors as well as material and human factors.

3. Methodology
3.1. Dataset

Publicly released benchmark datasets related to existing construction sites are labeled only for equipment or are classified into large categories such as buildings and workers [22,32]. However, the dataset used in this study is labeled for material and human risk factors that can cause disasters directly. At this time, the human factors discussed in this paper refer to unsafe human behavior and the presence or absence of a device that is attributed to humans and guarantees safety. For example, in the former case, working near an edge with the risk of a fall is representative, and in the latter case, there is the wearing or not wearing of a safety belt or a helmet. The material factors refer to improper installation of a device that guarantees safety for work convenience. A typical example is not installing an outrigger or safety handrail, which increases the likelihood of a fall. Both factors are caused by human carelessness, but they are divided into human factors or material factors depending on whether the responsibility is attributable to a person or an object. Table 1 summarizes the data for risk factor classes at construction sites. The total number of data is 25,859, which consists of 4149 cases of material factors, 15,412 cases of human factors, and 6298 cases of factors defined as normal. As summarized in the table, there is an imbalance between classes in the collected dataset, and insufficient data per label to use deep learning models, which increase representation power by stacking many layers. Therefore, in this study, data augmentation techniques were used to increase the amount of data. Data augmentation is a technique that improves the performance of computer vision. It increases the size of the actual training dataset by applying perturbation (crop, jitter, shear, zoom, etc.) to the image. Data augmentation is conducted only on the training dataset and not on the validation and test sets.

Table 1. Data distribution by class.

| Factor         | Class                        | Training | Validation | Test  |
|----------------|------------------------------|----------|------------|-------|
| Material       | A-type ladder outrigger      | 390      | 56         | 112   |
|                | Movable scaffolding frame outrigger | 381      | 109        | 55    |
|                | Safety handrail              | 106      | 30         | 16    |
|                | Scissor-type aerial work platform | 2025    | 327        | 542   |
|                | Safety vest                  | 4173     | 838        | 951   |
| Human          | Safety helmet                | 4734     | 880        | 1150  |
|                | Working in pairs             | 622      | 178        | 89    |
|                | Working near edges           | 1257     | 301        | 239   |
| Material & Human| Normal                      | 4408     | 710        | 1180  |
|                | Total                        | 18,096   | 3429       | 4334  |

The image data were collected from devices such as cameras and mobile phones, and are of various sizes, such as 1920 × 1080, 789 × 927, 992 × 925, 854 × 929 or 597 × 923. When high-resolution data is used as input to the model, the number of parameters in the model increases notably, which increases computation costs and the possibility of overfitting to the training data, so the image size was adjusted. In addition, given that the sizes of images are very different, the work of unifying them into a single size was carried out. Although information loss in the image may occur during the resizing process, it was determined that sufficient information would be transmitted if the data was enough to identify risk factors by the naked eye, so the size of image was modified to 224 × 224. This size 224 × 224 is also the default input size of the ImageNet pre-trained model. Thereafter, the prepared dataset was divided into training, validation, and test
During the division of the dataset, it was adjusted so that a specific class could be evenly distributed among all sets without bias toward one set. The data of each class were randomly distributed among the training, validation, and test datasets at a ratio of approximately 7:1:2.

### 3.2. Risk Factor Recognition Model

In this study, pre-trained DCNNs (ResNet50, MobileNetV2, ShuffleNetV1) were selected as the backbone. The ResNet model won the 2015 ImageNet large-scale visual recognition challenge contest with an error rate of 3.57% [33]. When training a deep neural network, the gradient may gradually become lost as the network gets deeper. This is called the gradient vanishing problem. To address this for improved training performance, a residual connection was proposed using a new technique called skip connection. The MobileNet model having dramatically reduced parameters was developed by Google in 2017 to run a deep learning model in a mobile environment [34,35]. MobileNet is based on a structure in which the existing convolution operation is changed to depth-wise separable convolution, which is a combination of depth-wise convolution and point-wise convolution. Although the performance of MobileNet was not noticeably improved, its long-running speed and its memory usage are lower, which are improvements over existing models. MobileNet made it possible to utilize deep learning models in embedded devices such as mobile phones and tablet PCs. ShuffleNet is a model designed to focus on minimized parameters like MobileNet, and a new computational method combining point-wise group convolution and channel shuffle was proposed [36]. Point-wise group convolution is designed in a way that only operates on channels in a specific area, and the amount of computation is dramatically reduced. In addition, it was designed to efficiently extract features by allowing information between channels to be mixed well through the channel shuffle process.

In this study, these three backbone models were trained using the transfer-learning method. Transfer-learning has the advantage of achieving high accuracy in a relatively short time, so it was judged to be suitable for this research environment. At this time, fine-tuning was performed using two strategies, and the individual methods are as follows. First, only a part of the upper layer was retrained in the feature extractor of the pre-trained DCNN models. Given that the image of the construction site used in this study has little similarity to the ImageNet dataset used for pre-training, only the upper layer was partially retrained to extract information specific to the problem. Retraining the upper layer so that it can learn features that can help detect risk factors is important for the performance and training of algorithms. Second, all layers of feature extractors of pre-trained DCNN models were retrained. This method uses only the structure of the pre-training model and trains every parameter newly based on the prepared dataset. The overall model architecture used in this study for recognizing the risk factor is shown in Figure 2.
3.3. Performance Measurement Metric

The evaluation metric that can most intuitively indicate the performance of a model is accuracy. Accuracy represents the fraction of the total samples that was correctly classified by the classifier, i.e., the ratio of classifying true as true and false as false correctly. However, when the number of data per class is imbalanced as in this study, the accuracy is relatively high when predicting the classes of large numbers, but the performance for the classes with small numbers is low. Therefore, in this study, the F1 score is used as a performance metric. Before explaining the F1 score, it is necessary to understand precision and recall. Precision, also known as the true positive rate, refers to the proportion of positive classes actually predicted positive. Figure 3 is a confusion matrix and Equation (1) is the equation for precision. Here, TP stands for true positive, denoting the number of predictions where the classifier correctly predicts the positive class as positive. FP stands for false positive, denoting the number of predictions where the classifier incorrectly predicts the negative class as positive. Recall is also called the true positive rate and refers to the ratio of all negative samples correctly predicted as negative by the classifier. In the case of precision, what the classifier classifies as true is the criterion, whereas in the case of recall, the actual true is the criterion. Recall is calculated as in Equation (2), where FN is an abbreviation for false negative, which represents the number of predictions where the classifier incorrectly predicts the positive class as negative. The F1 score used as a performance measurement metric in this study is calculated as the harmonic mean of precision and recall. The formula is given in Equation (3).

\[
\text{Precision} = \frac{TP}{TP + FP} \tag{1}
\]

\[
\text{Recall} = \frac{TP}{TP + FN} \tag{2}
\]

\[
F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3}
\]
4. Results

The most suitable training method and model for this research problem was found through the following experiment. The pre-trained model was taken from the Keras application provided with pre-trained weights. The mini batch size, which is a hyper-parameter that determines subsets of the dataset over which the gradient is calculated and the weights updated, was fixed at 128, and Adam was used as an optimizer. The initial learning rate was set to 0.0001, which was reduced to 1/10 when there was no improvement during five epochs based on the validation loss of the model.

4.1. Experimental Results According to Training Methods and Models

In this study, the pre-trained backbone model was selected as ResNet50, MobileNetV2, and ShuffleNetV1 in consideration of the number of parameters and model size. The models modified to suit our study were named Safety-ResNet, Safety-MobileNet, and Safety-ShuffleNet, respectively. Later, we tried fine-tuning in two different ways to find the most suitable training method for the task of this study. Table 2 presents the experimental results. When only a part of the upper layer was retrained on new data (hereinafter referred to as method (1)), the F1 scores of Safety-ResNet, Safety-MobileNet, and Safety-ShuffleNet were 49.15%, 80.92%, and 96.59%, respectively. By contrast, when all layers were newly re-trained, (hereinafter referred to as method (2)), the F1 scores were 99.75, 99.79%, and 98.04%, respectively, indicating higher performance compared to method 1. Given that the size of the dataset used in this study is large enough to solve the task, it is judged that re-training the entire layer would have been more efficient. In particular, it seems that Safety-ResNet did not sufficiently learn the features of a new dataset when using method 1. Furthermore, when using method 1, the lighter the model, the better the performance.

Table 2. Comparison of performance according to training methods by model.

| Model               | Parameters     | Inference Time (ms) | Method 1 F1 score (%) | Method 2 F1 score (%) |
|---------------------|----------------|---------------------|-----------------------|-----------------------|
| Safety-ResNet       | 23,606,153     | 67                  | 49.15                 | 99.75                 |
| Safety-MobileNet    | 2,269,513      | 30                  | 80.92                 | 99.79                 |
| Safety-ShuffleNet   | 911,601        | 0.93                | 96.59                 | 98.04                 |

Based on the experimental results by model, Safety-MobileNet, Safety-ResNet, and Safety-ShuffleNet showed high performance, in that order. The performance of Safety-MobileNet and Safety-ResNet is similar, but Safety-MobileNet has 0.1 times fewer parameters and 2.23 times lower CPU inference time than Safety-ResNet (based on method 2). This is because Safety-MobileNet is a lightweight model designed for a mobile environment. Although Safety-ShuffleNet is a lightweight model like Safety-MobileNet, there is a difference in performance in terms of accuracy. Therefore, in this study, Safety-MobileNet was selected as the final backbone model. As mentioned above, Safety-MobileNet is designed to be able to use deep learning models even where computing performance is limited or battery performance is important. Considering actual construction sites where the risk recognition algorithm will be used, it is likely that mobile devices such as cell phones and tablet PCs
will be used to find risk factors. Therefore, it is very important to use a lightweight model as a backbone. Through experiments, it was confirmed that Safety-MobileNet was the most suitable for the purpose of this study.

Table 3 summarizes the experimental results of Safety-MobileNet when method 2 was used. F1 scores for each class are all between 99% and 100%, and high accuracy is obtained even though the dataset is imbalanced between classes. It was judged that this is because the characteristics between the classes are distinct. In particular, the material factor has a large difference in the equipment that appears in the image, so even people who do not have prior knowledge in the domain can distinguish each class to some extent. However, human factors do not cause a large difference in features between images, but it seems that model training was well performed and high performance was achieved.

**Table 3. Performance by class in Safety-MobileNet.**

| Class                          | Precision | Recall | F1 Score |
|-------------------------------|-----------|--------|----------|
| A-type ladder outrigger       | 100       | 100    | 100      |
| Movable scaffolding frame outrigger | 100       | 100    | 100      |
| Safety handrail               | 100       | 100    | 100      |
| Scissor-type aerial work platform | 99.63     | 99.82  | 99.72    |
| Safety vest                   | 100       | 99.89  | 99.95    |
| Safety helmet                 | 99.74     | 99.65  | 99.70    |
| Working in pairs              | 100       | 100    | 100      |
| Working near edges            | 99.16     | 99.16  | 99.16    |
| Normal                        | 99.83     | 99.92  | 99.87    |

### 4.2. Visualization for Interpreting Prediction Results

Deep learning models have limitations in interpreting and reliability of the results because the process or basis for which the prediction is derived is opaque. To overcome this limitation, we determined which part of the image had the most influence on the final classification decision of the model through heatmap visualization. In this study, class activation map (CAM) was used, and the operation method of CAM is as follows. A total of $k$ outputs can be obtained by applying global average pooling (GAP) to the last $k$ feature maps obtained through the CNN layer. After linearly combining these values and applying softmax, the probability value is finally output. At this time, the activation map for class $c$ is the same as Equation (4).

$$A_c(i,j) = \sum_k w^c_k f_k(i,j)$$

(4)

$f_k(i,j)$ denotes the $k$-th feature image, and $w^c_k$ denotes the weight from the $k$-th feature image $f_k(i,j)$ to class $c$. If each feature map $f_k(i,j)$ is multiplied by the weight $w^c_k$ for each class, a heatmap can be obtained with as many as $k$ number of feature maps. By summing all these heatmap images pixel-wise, a single heatmap can be obtained, which is CAM [37]. The result of visualizing the part that most influences the model decision using CAM is shown in Figure 4. In Figure 4, it is observed that the parts that are important for predicting the class, such as a safety helmet or safety vest, are marked in red and have the most influence. Likewise, in the case of classes that indicate the risk of a fall accident, such as working near edges, the end of the facility is expressed in red. However, if the prediction is different from the actual class, it can be seen that areas that are completely unrelated to the class are displayed in red. Furthermore, a large portion of the image appears red or yellow. In other words, when compared to the correctly predicted image, a specific part did not affect the model’s prediction; rather, a wide range of areas affected the model, causing confusion. As a result of examining erroneously predicted images, a case was found that could be confusing even to the naked eye, or a case where a risk factor could belong to more than one class. Through this visualization work, it was possible to objectively present the classification basis of the trained model and to improve the interpretation power of the model.
5. Discussion

5.1. Implications

The implications of this study are as follows. First, the material and human risk factor recognition model presented in this study is meaningful in that it applies deep learning technology in the construction safety management field to address various real-world concerns, such as the absence of an accident prevention system at construction sites, the increase in the occupational accident fatality rate, and the difficulty in practice in applying the safety management system for small construction companies. These issues can be considered in future studies on the development of preventive disaster response systems or construction worker protection systems. Second, construction companies can use the results of this study to reduce safety accidents and improve work productivity by establishing a system for preventing safety accidents at construction sites and set up an efficient workforce plan. In particular, positive social and economic effects of reducing the risk of safety accidents by identifying the risk factors that cause accidents at construction sites in advance are expected. Finally, in this study, the model was interpretable from the analyst’s point of view through visualization work. In particular, the deep learning method has the disadvantage of being a black box model, making it difficult to interpret the process and basis for the results [38]. However, given that we made it possible to interpret the decision process of the model using a visualization tool, in order to further improve the performance of the model in the future, the model can be improved based on the decision basis for incorrectly predicted data.

Figure 4. Visualization results using CAM.
5.2. Limitations

In actual construction sites, there are multiple risk factors. Given that it is virtually impossible to detect infinite possible risk factors, we made it possible to selectively detect some risk factors to confirm the future applicability of the model. Therefore, it is difficult to recognize risk factors not corresponding to labels. In addition, as environmental factors specific to Korea were considered, additional training may be required to achieve the same performance by applying the proposed framework to a dataset collected in other countries. Furthermore, although individual risk factors were recognized in this study, in real environments, complex risk factors may arise, e.g., workers wearing neither safety vests nor helmets, or working alone without safety equipment in an environment that requires two-person work. Therefore, to apply the proposed framework effectively to various construction sites in the future, it is necessary to configure a dataset to recognize complex risk factors and design a model capable of multi-class classification.

6. Conclusions

In 2021, the number of human casualties in the construction industry in Korea was 30,659 cases. Among many risk factors such as falling, being hit by an object, accidental amputation, or electric shock, human damage due to falling is the most frequent [39]. Therefore, we used a dataset that focused on material factors related to falling and human factors that are most important in preventing accidents. In the existing studies using deep learning models in the field of safety management, there is no study on an integrated risk factor recognition model that covers material and human factors. Therefore, a framework for identifying material and human risk factors that cause disasters in construction sites was proposed. To find the optimal model and training method, the pre-trained deep learning models, i.e., Safety-ResNet, Safety-MobileNet, and Safety-ShuffleNet, were tested in two ways. Rather than retraining only a part of the upper layer of the pre-trained model, the performance when the entire model was newly retrained was better, and Safety-MobileNet showed the highest performance among the models. This seems to have resulted in better performance when many layers were newly trained because the characteristics of the data had little similarity to the dataset used for pre-training. In addition, the performance of the lightweight model with fewer parameters was better than the heavy model with many parameters in this study environment. Although the performances of Safety-MobileNet and Safety-ResNet were similar, it was determined that using Safety-MobileNet, which was developed to be used in an environment with limited computational power, would have a practical advantage when applying the model in practice in the future. Thus, it was selected as the final backbone model. Furthermore, visualization work was performed to understand the process of deriving the results of the model. Through this, we examined what factors most influenced the final prediction of the model, and improved the interpretability of the model. The framework proposed in this study is an initial study for the development of safety technologies such as disaster prevention, and it is meaningful in that it can be used as a source proprietary technology for the future development of preventive disaster response systems and construction worker protection systems.

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

Funding: This research was funded by the Korea Agency for Infrastructure Technology Advancement (KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (Grant 21CTAP-C163908-01).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data are not publicly available due to privacy.
Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. FACTS ON Safety at Work. Available online: https://www.ilo.org/wcmsp5/groups/public/---dgreports/---dcom/documents/publication/wcms_067574.pdf (accessed on 5 January 2022).
2. Kosoroski, M.D. Industrial injuries and the business cycle. Mon. Lab. Rev. 1938, 46, 579.
3. Curtis Breslin, F.; Polzer, J.; MacEachen, E.; Morrongiello, B.; Shannon, H. Workplace injury or “part of the job”: Towards a gendered understanding of injuries and complaints among young workers. Soc. Sci. Med. 2007, 64, 782–793. [CrossRef] [PubMed]
4. Davies, R.; Jones, P.; Núñez, I. The impact of the business cycle on occupational injuries in the UK. Soc. Sci. Med. 2009, 69, 178–182. [CrossRef] [PubMed]
5. Work Injury Costs. Available online: https://injuryfacts.nsc.org/work/costs/work-injury-costs/ (accessed on 5 January 2022).
6. Occupational Safety and Health Administration. Directive Number: CSP 02-21-01. Effective Date: 04/27/2021. Subject: Fiscal Year (FY) 2022 Instructions for 23(g) State Plan Grants. Available online: https://www.osha.gov/sites/default/files/enforcement/directives/CSP_02-21-01.pdf (accessed on 5 January 2022).
7. HSE Business Plan 2021/22. Available online: https://www.hse.gov.uk/aboutus/strategiesandplans/businessplans/plan2122.pdf (accessed on 5 January 2022).
8. Deutsche Gesetzliche Unfallversicherung, e. V DGUV—Prävention—GDA. Available online: https://www.dguv.de/de/prevention/gda/index.jsp (accessed on 5 January 2022).
9. Korea Occupational Safety and Health Agency (KOSHA). Available online: https://www.kosha.or.kr/kosha/info/getLaw.do (accessed on 5 January 2022).
10. JISHA: Japan Industrial Safety & Health Association. Available online: https://www.jisha.or.jp/english/ (accessed on 5 January 2022).
11. Korea Occupational Safety and Health. Available online: https://www.kosha.or.kr/kosha/data/IndustrialDisasterStatistics.do?mode=view&articleNo=421342&article.offset=0&articleLimit=10 (accessed on 5 January 2022).
12. Kim, S.; Lee, J.; Kang, C. Analysis of industrial accidents causing through jamming or crushing accidental deaths in the manufacturing industry in South Korea: Focus on non-routine work on machinery. Saf. Sci. 2021, 133, 104998. [CrossRef]
13. Heinrich, H.W.; Petersen, D. Industrial Accident Prevention, 5th ed.; McGraw-Hill: New York, NY, USA, 1980; ISBN 9780070280618.
14. Bird, F.E.; Loftus, R.G. Loss Control Management; Institute Press: Loganville, GA, USA, 1976.
15. İssever, H.; Özdilli, K.; Önen, L.; Tan, O.; Dişçi, R.; Yardımcı, O. Examination of personal factors in work accidents. Indoor Built Environ. 2008, 17, 562–566. [CrossRef]
16. Hamid, A.R.A.; Abd Majid, M.Z.; Singh, B. Causes of accidents at construction sites. Malays. J. Civ. Eng. 2008, 20, 2. [CrossRef]
17. Chi, S.; Han, S.; Kim, D.Y. Relationship between unsafe working conditions and workers’ behavior and impact of working conditions on injury severity in US construction industry. J. Constr. Eng. Manag. 2013, 139, 826–838. [CrossRef]
18. Egbu, C.; Umar, T. Causes of construction accidents in Oman. Middle East J. Manag. 2018, 5, 21. [CrossRef]
19. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. Nature 2015, 521, 436–444. [CrossRef]
20. Krichhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. Commun. ACM 2017, 60, 84–90. [CrossRef]
21. 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]
22. Nath, N.D.; Behzadani, A.H.; Paal, S.G. Deep learning for site safety: Real-time detection of personal protective equipment. Autom. Constr. 2020, 112, 103085. [CrossRef]
23. Ding, L.; Fang, W.; Luo, H.; Love, P.E.D.; Zhong, B.; Ouyang, X. A deep hybrid learning model to detect unsafe behavior: Integrating convolution neural networks and long short-term memory. Autom. Constr. 2018, 86, 118–124. [CrossRef]
24. Luo, X.; Li, H.; Cao, D.; Dai, F.; Seo, J.; Lee, S. Recognizing diverse construction activities in site images via relevance networks of construction-related objects detected by convolutional neural networks. J. Comput. Civ. Eng. 2018, 32, 04018012. [CrossRef]
25. Kolar, Z.; Chen, H.; Luo, X. Transfer learning and deep convolutional neural networks for safety guardrail detection in 2D images. Autom. Constr. 2018, 89, 58–70. [CrossRef]
26. Kim, H.; Kim, H.; Hong, Y.W.; Byun, H. Detecting construction equipment using a region-based fully convolutional network and transfer learning. J. Comput. Civ. Eng. 2018, 32, 04017082. [CrossRef]
27. Dai, J.; Li, Y.; He, K.; Sun, J. R-FCN: Object detection via region-based fully convolutional networks. In Proceedings of the Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016 (NIPS 2016), Barcelona, Spain, 5–10 December 2016.
28. Xiao, B.; Kang, S.-C. Vision-based method integrating deep learning detection for tracking multiple construction machines. J. Comput. Civ. Eng. 2021, 35, 04020071. [CrossRef]
29. Arabi, S.; Haghighat, A.; Sharma, A. A deep learning based solution for construction equipment detection: From development to deployment. arXiv 2019, arXiv:1904.09021.
30. Wang, M.; Wong, P.K.-Y.; Luo, H.; Kumar, S.; Delhi, V.-S.K.; Cheng, J.C.-P. Predicting safety hazards among construction workers and equipment using computer vision and deep learning techniques. In Proceedings of the 36th International Symposium on Automation and Robotics in Construction (ISARC), Banff, AB, Canada, 21–24 May 2019; International Association for Automation and Robotics in Construction (IAARC): Oulu, Finland, 2019.

31. Zhong, B.; Pan, X.; Love, P.E.D.; Ding, L.; Fang, W. Deep learning and network analysis: Classifying and visualizing accident narratives in construction. Autom. Constr. 2020, 113, 103089. [CrossRef]

32. Xiao, B.; Kang, S.-C. Development of an image data set of construction machines for deep learning object detection. J. Comput. Civ. Eng. 2021, 35, 05020005. [CrossRef]

33. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; IEEE: Piscataway, NJ, USA, 2016.

34. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv 2017, arXiv:1704.04861.

35. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.-C. MobileNetV2: Inverted residuals and linear bottlenecks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; IEEE: Piscataway, NJ, USA, 2018.

36. Zhang, X.; Zhou, X.; Lin, M.; Sun, J. ShuffleNet: An extremely efficient convolutional neural network for mobile devices. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; IEEE: Piscataway, NJ, USA, 2018.

37. Zhou, B.; Khosla, A.; Lapedriza, A.; Oliva, A.; Torralba, A. Learning deep features for discriminative localization. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; IEEE: Piscataway, NJ, USA, 2016.

38. Shwartz-Ziv, R.; Tishby, N. Opening the black box of Deep Neural Networks via information. arXiv 2017, arXiv:1703.00810.

39. Construction Safety Management Integrated Information. Ministry of Land, Infrastructure and Transport. Available online: https://www.csi.go.kr/hrp/stats/hrpStatsByAccdList.do (accessed on 5 January 2022).