buildings

MDPI

Article

YOLOv8s-SNC: An Improved Safety-Helmet-Wearing Detection Algorithm Based on YOLOv8

Daguang Han [1], Chunli Ying [2,*], Zhenhai Tian [3], Yanjie Dong [3], Liyuan Chen [4], Xuguang Wu [4] and Zhiwen Jiang [5]

1 School of Civil Engineering, Southeast University, Nanjing 211189, China; daguanghan@seu.edu.cn
2 School of Architecture and Civil Engineering, Loughborough University, Leicestershire LE113TU, UK
3 The Fourth Engineering Co., Ltd. of CCCC First Highway Engineering Co., Ltd., Tianjin 300350, China
4 Pingyang County Power Supply Company, State Grid Zhejiang Electric Power Co., Ltd., Wenzhou 325400, China
5 College of Civil Engineering, Nanjing Tech University, Nanjing 211816, China; 202261226070@njtech.edu.cn
* Correspondence: c.ying@lboro.ac.uk

Abstract: The use of safety helmets in industrial settings is crucial for preventing head injuries. However, traditional helmet detection methods often struggle with complex and dynamic environments. To address this challenge, we propose YOLOv8s-SNC, an improved YOLOv8 algorithm for robust helmet detection in industrial scenarios. The proposed method introduces the SPD-Conv module to preserve feature details, the SEResNeXt detection head to enhance feature representation, and the C2f-CA module to improve the model's ability to capture key information, particularly for small and dense targets. Additionally, a dedicated small object detection layer is integrated to improve detection accuracy for small targets. Experimental results demonstrate the effectiveness of YOLOv8s-SNC. When compared to the original YOLOv8, the enhanced algorithm shows a 2.6% improvement in precision (P), a 7.6% increase in recall (R), a 6.5% enhancement in mAP_0.5, and a 4.1% improvement in mean average precision (mAP). This study contributes a novel solution for industrial safety helmet detection, enhancing worker safety and efficiency.

Keywords: safety helmets; YOLOv8s; small object detection; SPD-Conv; attention mechanism; squeeze-and-excitation

check for updates

Citation: Han, D.; Ying, C.; Tian, Z.; Dong, Y.; Chen, L.; Wu, X.; Jiang, Z. YOLOv8s-SNC: An Improved Safety-Helmet-Wearing Detection Algorithm Based on YOLOv8. *Buildings* **2024**, *14*, 3883. https://doi.org/10.3390/buildings14123883

Academic Editor: Feng Guo

Received: 25 August 2024
Revised: 28 September 2024
Accepted: 4 October 2024
Published: 3 December 2024

1. Introduction

In the 21st century, the global industrial landscape has undergone a transformative phase, with many nations experiencing rapid industrial growth and technological advancements [1–3]. This expansion has elevated workplace safety to a critical priority across the globe, particularly within industrial and engineering sectors [4]. Safety helmets, as essential piece of personal protective equipment, play a vital role in mitigating head injuries and ensuring the well-being of workers across various industries, including construction and manufacturing [5,6]. According to a report by the International Labour Organization (ILO), falls from heights and struck-by accidents are among the leading causes of fatalities in the construction industry, highlighting the critical need for effective head protection [7].

The significance of safety monitoring on construction sites has prompted a surge in research for automatic helmet-wearing detection systems [8,9]. These systems are not only essential for regulatory compliance but also for fostering a proactive safety culture [10,11]. Manual supervision of helmet usage is not only labor-intensive but also prone to human error and inconsistency. Automated detection offers a reliable, consistent, and unbiased method of monitoring helmet usage, a task that is often unfeasible through manual supervision alone, especially on large or complex construction sites [12]. Furthermore, the cost-effectiveness of these systems, when considering the long-term savings from reduced accidents and improved safety compliance, makes them an attractive investment for construction firms.

However, the implementation of these systems is not without challenges. Among the various detection technologies, vision-based systems have become dominant due to their lower cost and easier implementation compared to sensor technology [13]. Despite their advantages, traditional vision-based helmet-wearing detection systems, which typically involve pedestrian detection, helmet positioning, and helmet recognition, still face a series of challenges. The complex and changeable environment of construction sites, with factors such as varying lighting conditions and obstructions, adds to the difficulty of accurate helmet detection. Additionally, the variability in helmet shape, color, and the presence of small, densely packed targets in crowded work scenes contribute to the risk of false detections and missed detections [14,15].

Therefore, although vision-based technology has made some progress in automatic helmet-wearing detection, methods of coping with complex and changing working environments and improving the accuracy and stability of detection still require further research and improvement. Existing detection algorithms often have difficulty identifying small objects or densely clustered objects, while factors such as occlusion and complex environments will further reduce the detection accuracy. In response to these challenges, we proposed an enhanced YOLOv8 algorithm to detect the wearing of helmets in complex environments.

The YOLOv8 model was selected for this helmet-wearing detection task due to its distinct advantages over other vision sensing models [16]. Firstly, YOLOv8 achieves an optimal balance between speed and accuracy, making it ideal for real-time applications [17]. Its high frames per second (FPS) rate ensures timely detection and response, crucial for preventing accidents in industrial settings. Secondly, the model's efficiency is remarkable, allowing for deployment on resource-constrained devices like drones or mobile devices, commonly used for industrial safety monitoring [18]. This efficiency is attributed to its anchor-free detection and efficient feature fusion techniques. Lastly, YOLOv8′s adaptability to various datasets and scenarios is unparalleled, thanks to its modular design and dynamic sample allocation strategy [19]. This adaptability enables easy customization and optimization for specific industrial safety helmet detection tasks. Compared to other models like Faster R-CNN and SSD, YOLOv8 demonstrates superior performance in terms of speed, accuracy, and efficiency, making it the most suitable choice for the research objectives [20].

Based on the original YOLOv8, our method introduces a sub-network specifically for detecting small objects, called Yolov8s-SNC, enhances the backbone and neck, and proposes a new detection head, which aims to enhance the detection performance of helmets in scenarios such as small targets, object occlusion, and complex environments, providing more reliable and effective solutions for safety monitoring of production sites including construction sites, and further improving the safety level of workers [21,22]. Furthermore, to ensure practical application, Yolov8s-SNC is designed to be deployable across various platforms, including unmanned aerial vehicles for aerial surveillance and integration with on-site cameras for real-time monitoring. This enables its use in a wide array of construction projects, from single-structure sites to complex multi-building developments.

2. Related Work

In the current research field of worker helmet-wearing detection, a variety of methods have emerged to solve this problem. These methods have their characteristics and are suitable for different scenarios and needs [23,24]. The following is an analysis of several major methods:

Methods based on traditional machine vision: This type of method mainly uses features such as the shape, color, and directional gradient histogram (HOG) of the helmet, as well as classifiers such as support vector machine (SVM) and random forest to locate the position of workers and helmets. This method may work well and be low-cost in simple scenarios but may have limited accuracy in complex scenarios. In addition, since these methods rely on manually designed features, they may be sensitive to lighting changes and occlusion [25]. Chiverton employed SVM classification on histogram features for helmet detection in motorcycle riders, integrating a tracking system for improved accuracy. However, it may struggle with varying lighting and complex backgrounds, and relies on manual feature extraction [26]. Waranusast et al. utilized KNN classification on extracted features for helmet detection, with a focus on motorcycle riders. It struggled with low-resolution images and overlapping objects, impacting accuracy [27]. Park et al. presented a 3D tracking method using stereo cameras for construction site monitoring. It lacks real-time processing and may struggle with occlusions and varying appearances of tracked entities [28].

Sensor-based methods: These methods use inertial sensors, infrared sensors, or other types of sensors to detect whether a worker is wearing a helmet. Sensor-based methods usually have high accuracy but may be costly and complex to deploy. In addition, sensors may be interfered with by the environment, which may affect the accuracy of detection. Lee et al. reviewed smart helmet technology, focusing on multimodal sensing for health and safety. It discussed various sensor types and their applications but lacks specific details on helmet detection accuracy and real-world implementation challenges [29]. Zhang et al. developed a smart hard-hat system utilizing Internet of Things (IoT) for real-time NHU detection and location, addressing limitations of previous methods. However, the system's reliance on RFID and potential interference from construction site equipment could affect accuracy and efficiency [30]. Wang et al. developed a smart helmet and smart shoes system using accelerometers and pressure sensors to detect near-falls during stair descent. However, the approach had its drawbacks, such as reliance on controlled experiments and the exclusion of varied stair conditions [31]. García et al. integrated AI and AR to enhance FR situational awareness, yet initial tests showed issues with Personal Protective Equipment (PPE) compatibility and system robustness [32].

Methods based on deep learning: With the development of deep learning technology, more and more studies have begun to use deep learning models for helmet-wearing detection. These methods usually use convolutional neural networks (CNN) or their variants, such as YOLO (You Only Look Once) or Faster R-CNN, to detect and identify helmets [33–35]. Methods based on deep learning usually have higher accuracy and robustness but require a lot of data and computing resources. In addition, the model complexity of these methods is high and may require specialized hardware support [36]. Yang et al. introduced a YOLOv3-based system for helmet detection in power construction scenes, achieving over 90% accuracy, yet it did not detail potential limitations such as variability in detection performance under different lighting conditions or helmet occlusions [37]. Wu et al. improved YOLO V3 by incorporating Densenet, enhancing detection accuracy under challenging conditions [38]. However, it did not address potential limitations such as the impact of varied lighting on detection performance. Long et al. presented a deep-learning-based solution for helmet detection that makes use of SSD and addresses small object problems [39]. Nevertheless, it did not address the system's robustness in different environmental circumstances or against occlusions. Wang et al. created a lightweight CNN for hardhat identification that uses MobileNet, a top-down module, and a residual block to provide real-time detection [40]. Unfortunately, it did not address the possibility of variable illumination conditions or dynamic situations affecting detection accuracy. Nipun et al. presented a deep learning model using YOLOv3 to detect PPE compliance, achieving high accuracy [41]. Despite this, there were challenges in detecting smaller PPE items and it was susceptible to occlusions, suggesting an enhanced algorithm is necessary to address these issues. Gao et al. presented DMS-YOLOv5 for small object detection, enhancing feature extraction and reducing background interference [42]. A potential limitation of the algorithm, however, is its sensitivity to occlusion and lighting conditions. Song et al. introduced an enhanced YOLOv8 model integrating DWR, ASPP, and NWD for helmet detection, achieving improved accuracy [35]. However, it did not explore the model's robustness in scenarios with variable lighting or dynamic backgrounds.

In summary, although the existing methods can meet the needs of workers' helmet-wearing detection to a certain extent, the existing detection algorithms have difficulty in detecting small targets and dense targets, and target occlusion and complex and changing construction environments will reduce the detection accuracy. To address these limitations, this paper proposes a helmet detection method based on improved YOLOv8, by introducing the SPD-Conv (spatial pyramid dilated convolution) module and constructing a new squeeze excitation enhanced precision head and small target detection head to improve the helmet detection performance for small targets, occluded targets, and complex scenes.

3. Methodology

3.1. YOLOv8 Network

The YOLO series has garnered acclaim for its adeptness in achieving a harmonious balance between speed and accuracy in the realm of target detection algorithms. Renowned for their capacity to swiftly and precisely identify targets, YOLO models are renowned for their ease of deployment across various mobile platforms, facilitating real-time applications [43]. Among these, YOLOv8 is a product of Ultralytics, integrating novel features and enhancements to augment both performance and adaptability [44]. Figure 1 delineates the architecture of the YOLOv8 network.

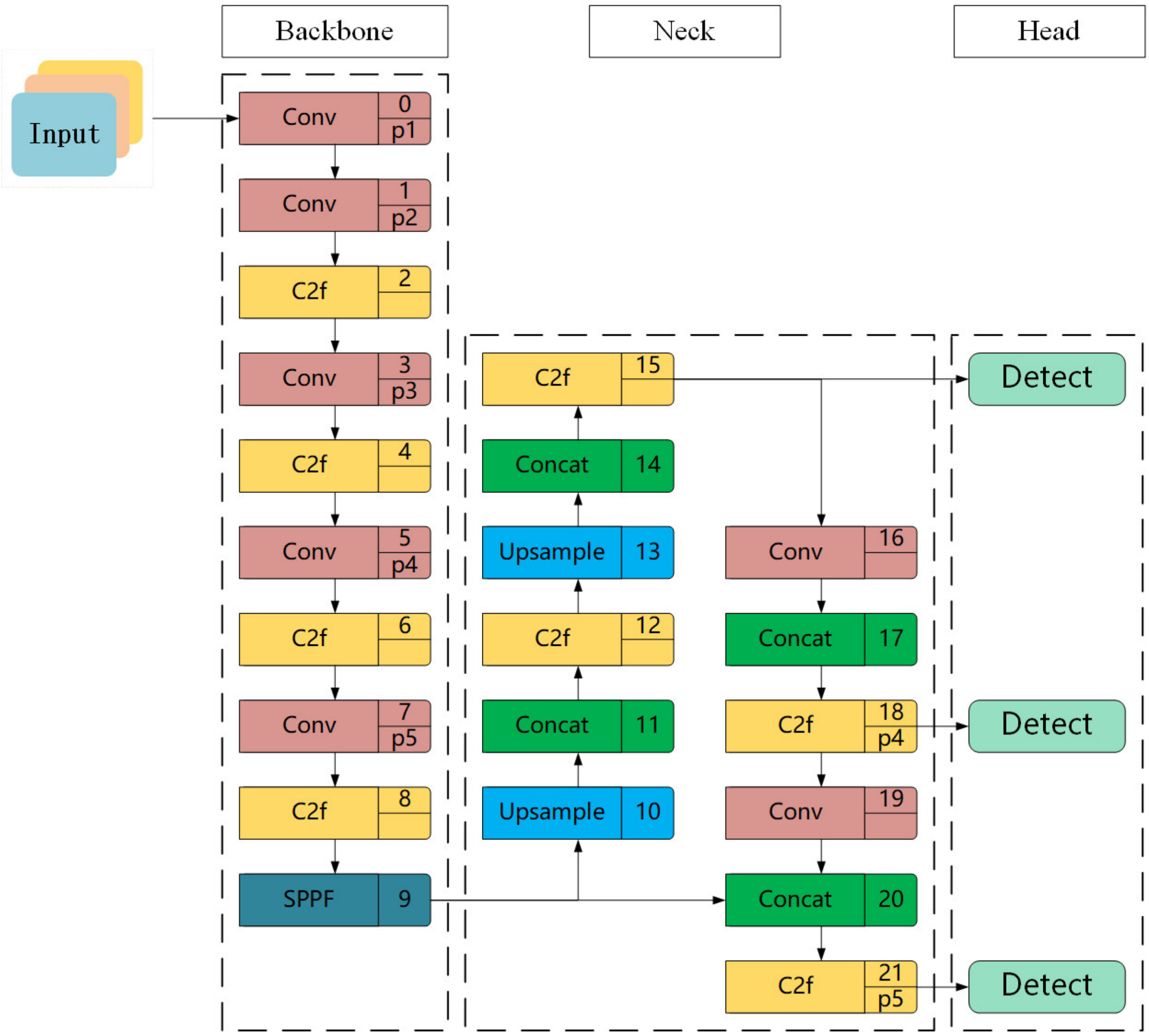


Figure 1. Architecture for YOLOv8 module.

The YOLOv8 architecture is delineated into four principal components: input, backbone, neck, and head, each fulfilling distinct roles in the process of object recognition and image segmentation [45].

(1) Input: Initially, input images undergo refinement via the Mosaic data augmentation technique, amplifying the model's generalization and robustness capabilities.
(2) Backbone: Central to the YOLOv8 model, the feature extraction network integrates multiple convolutional (Conv) layers, C2f modules, and spatial pyramid pooling with features (SPPF). The C2f module, drawing from the advancements of C3 and the efficient layer aggregation network (ELAN) from YOLOv7, intertwines across multiple branch layers, enriching gradient flow information while maintaining lightweight characteristics [46]. Concurrently, SPPF, an adaptation of spatial pyramid pooling (SPP), curtails network layers and redundancy, expediting feature fusion [47].
(3) Neck: This stage adopts the feature pyramid network (FPN) and path aggregation network (PAN) structures, bolstering multi-scale semantic expression and localization precision [48].
(4) Head: Here, prior features are leveraged for target category and location recognition, facilitating detection. The utilization of a decoupled head structure effectively trims down parameter count and computational complexity, while bolstering model generalization and robustness [49]. The traditional reliance on anchor nodes is supplanted with an anchor-free approach, enabling direct prediction of target center points and width-to-height ratios, thereby reducing the necessity for anchor frames [50].

Moreover, the loss computational aspect is governed by the task-aligned assigner dynamic sample allocation strategy, which adapts to varying datasets and model requirements, ensuring optimal alignment between classification and regression tasks [51]. The integration of distribution focal loss (DFL) alongside complete intersection over union loss (CIoU Loss) for the regression branch loss function, complemented via binary cross entropy (BCE) for classification loss, further fortifies alignment consistency, culminating in heightened model efficacy [52].

In essence, the YOLOv8 algorithm epitomizes a comprehensive fusion of cutting-edge methodologies and optimizations, manifesting in a robust framework capable of facilitating intricate object detection tasks with remarkable precision and efficiency [53].

3.2. YOLOv8s-SNC Network

In the YOLOv8s-SNC network, to enhance the detection of small and low-resolution objects, a small object detection head is integrated, which aims to capture fine details by leveraging high-resolution feature maps and establishing skip connections. In addition, an SPD-Conv module that employs spatial-to-depth conversion is adopted to reduce information loss and improve the network's sensitivity to small and dense targets. We also implement the SEResNeXt detection head, which combines the channel attention mechanism of SENet with the parallel path structure of ResNeXt to refine feature representation and improve detection accuracy in complex scenes. Besides, the C2f-CA module, an attention mechanism that emphasizes key information, especially small and dense targets, is embedded to enhance the feature capture ability of the model. Finally, to address the problem of detail loss due to multiple downsampling stages, we introduce a dedicated small object detection layer. The proposed network architecture is shown in Figure 2.

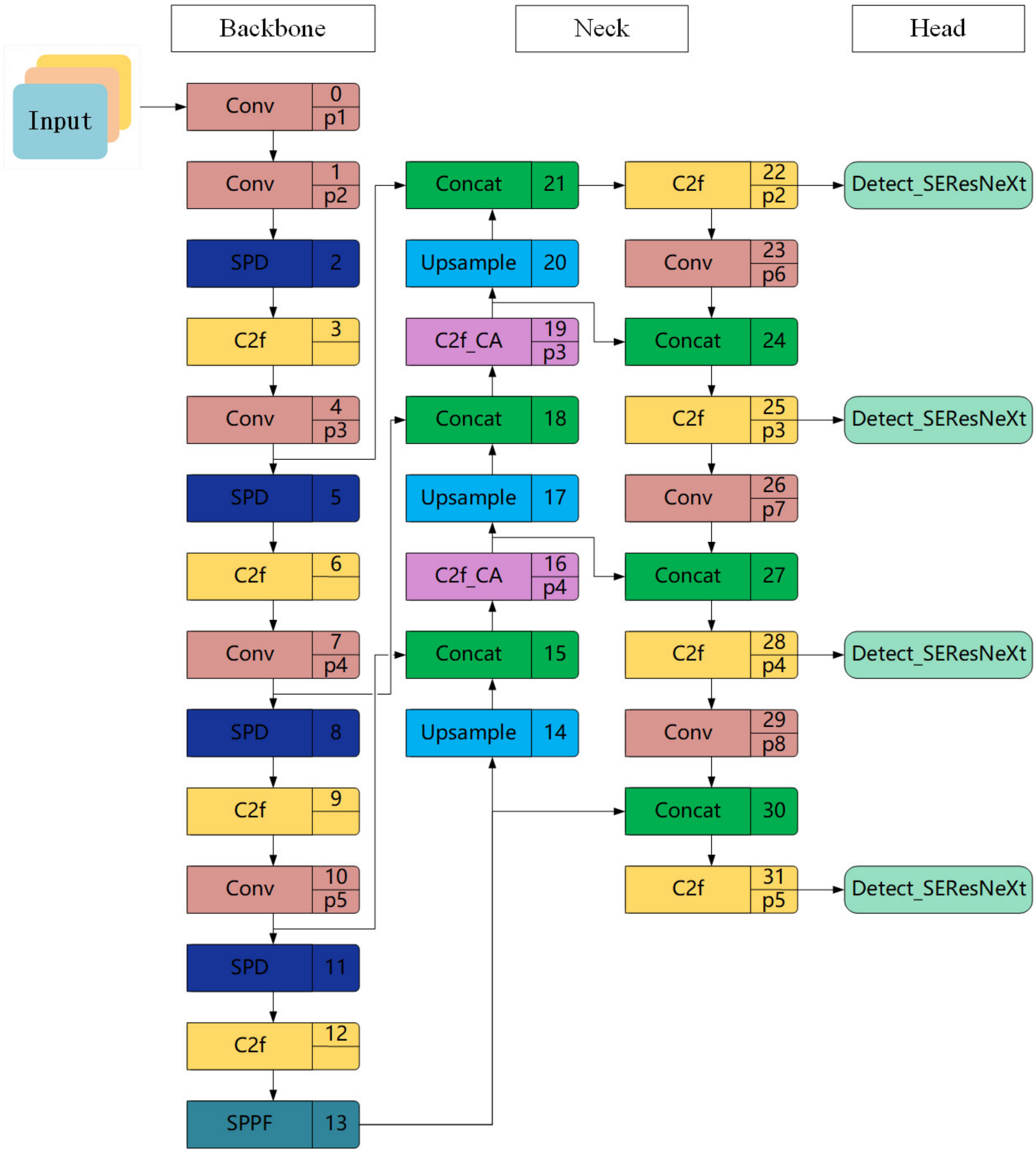


Figure 2. Architecture for YOLOv8s-SNC module.

3.2.1. Small Object Detection

In the context of YOLOv8-based safety helmet detection algorithms, the integration of a small target detection layer holds significant advantages for enhancing the model's capability in detecting diminutive targets. Conventionally, YOLOv8's backbone undergoes multiple downsampling operations, progressively extracting higher-level features from the input image. However, with each downsampling iteration, there is a concurrent reduction in feature map resolution and loss of low-level feature details. This cascade effect leads to a diminished ability to detect small targets as the model gradually loses substantial original information [54,55].

In the original YOLOv8 network structure, for instance, when the input image size is 640×640 pixels, the corresponding minimum detection scale is 80×80. Consequently, each detection head in the model has a receptive field of 8×8. When objects in the original image are smaller than 8 pixels in width or height, the existing model encounters difficulty in effectively recognizing these diminutive targets [56].

To mitigate this issue, this paper introduces an additional small target detection layer onto the original network architecture [57]. Leveraging features from the backbone's high-resolution shallow feature maps, obtained after two downsampling operations, and fusing them with the semantically enriched feature maps from the neck section, which have undergone three upsampling operations, facilitates the creation of skip connections. Subsequently, employing the C2f module enables the fusion of concatenated feature maps. Lastly, appending a small target detection head with a scale of 128×128 enriches the feature information pertinent to small targets, thus enhancing their detection efficacy.

In essence, the incorporation of the small target detection layer within the YOLOv8-based safety helmet detection algorithm presents a multifaceted advantage. By harnessing both low-level and high-level feature representations, facilitated through skip connections and feature fusion, the model gains a heightened sensitivity to small targets. This augmentation is instrumental in bolstering the algorithm's efficacy in accurately identifying and localizing safety helmets, particularly in scenarios where such targets may be comparatively diminutive in size within UAV images, thus enhancing the overall performance and robustness of the detection system.

3.2.2. SPD-Conv

Convolutional neural networks often perform poorly when dealing with complex tasks with low resolution or small object sizes. This is mainly because the existing convolutional neural network architecture has design flaws, especially the excessive use of stridden convolutional layers and pooling layers, which easily leads to the loss of fine information and also affects the effective learning of feature representation [58].

In order to improve the model's ability to detect small and dense objects, the SPD-Conv module is introduced to replace the convolution stride and pooling layer at the neck of the network structure [59]. This module converts the input image from space to depth, integrates the image's depth information into the model, reduces information loss, and thus enhances the model's perception of small and dense objects. SPD-Conv consists of two parts: one is the spatial depth conversion layer and the other is the non-strided convolution layer [60]. These two parts work together to convert and compress the image pixel information originally distributed in two-dimensional space into one-dimensional space, as shown in Figure 3. This design aims to overcome the challenges in low-resolution image processing and improve the network's performance in detail recognition.

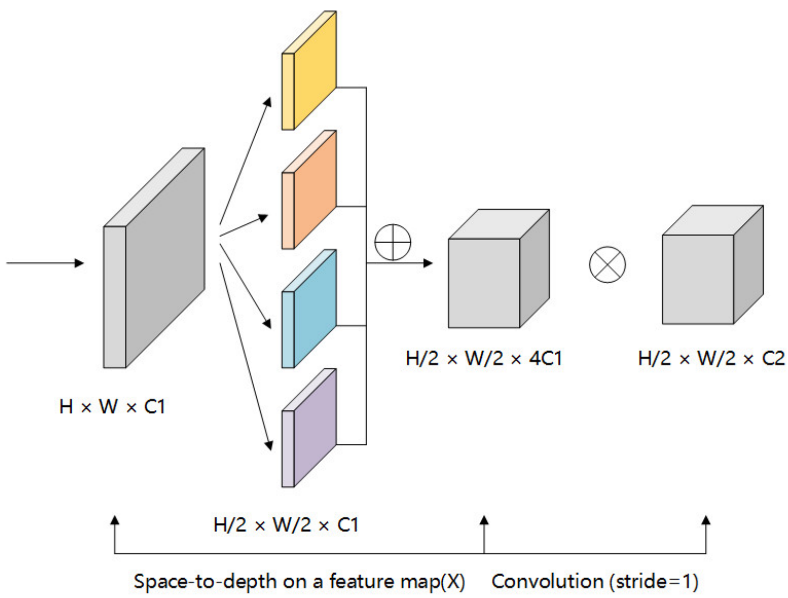
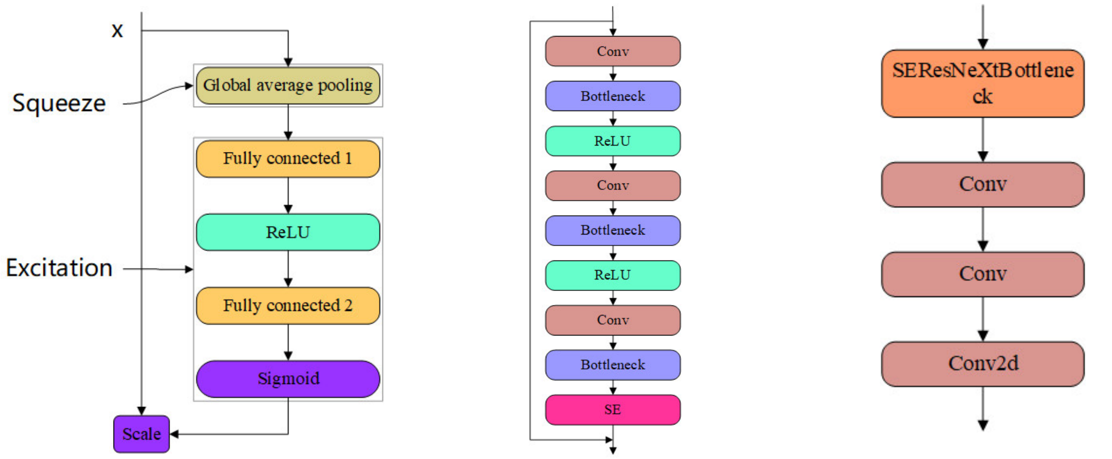


Figure 3. SPD-Conv module.

In the backbone network of YOLOv8, the SPD-Conv module is introduced before each C2f (Cross-features fusion). Specifically, the SPD-Conv module is integrated at each key point of the feature extraction network to ensure that the network can fully capture and maintain the details of small targets before feature fusion. This improvement helps to perform more accurate classification and bounding box regression in the subsequent detection head network.

3.2.3. SEResNeXt Detection Head

This paper proposes a detection head called SEResNeXt, which combines the channel attention mechanism of SENet [61] and the parallel path structure of ResNeXt [62,63], aiming to improve the accuracy and robustness of the model for helmet detection in complex operating scenarios, see Figure 4.



(**a**) Structure of SE module (**b**) SEResNeXtBottleneck (**c**) Detect-SEResNeXt

Figure 4. SEResNeXt detection head structure.

The design principle of the SEResNeXt detection head is rooted in the bottleneck layer structure in deep learning. Features are gradually extracted and refined through a series of fine convolution operations to ensure that the feature representation capabilities of the model are improved while maintaining efficient calculations [64]. First, the structure uses a 1×1 convolutional layer to reduce the number of input channels to a smaller width. This operation significantly reduces the computational complexity because by reducing the number of input channels, the computational overhead of subsequent convolution operations is greatly reduced. Next, the structure utilizes 3×3 grouped convolutional layers to process features on different groups in parallel. The introduction of grouped convolution not only retains the expressive ability of the network but also enhances the model's ability to capture multi-scale and multi-directional features. This is because each group can learn different feature representations, thus enriching the feature space. After these two convolution operations, a 1×1 convolution layer is used to restore the number of channels to four times the original to ensure that there is enough feature dimension in the output stage to capture complex feature information. Each convolution operation is followed by a batch normalization layer to standardize feature distribution and stabilize the training process. The first two convolution operations are also connected with ReLU activation functions to introduce nonlinearity and improve the expressive ability of the network.

After these convolution operations, SEResNeXtBottleneck performs adaptive recalibration between channels through a squeeze-and-excitation (SE) module [65]. The SE module first obtains the global information of each channel through the global average pooling operation, then learns the dependency between channels through two fully connected layers, and finally scales and adjusts each channel through the sigmoid activation function. This process can enhance the expression of important features while suppressing unimportant features, thereby improving the quality of feature representation. Before performing the residual connection, if the input and output dimensions are inconsistent, a downsampling layer is used to resize the input. The introduction of the downsampling layer is to match the output dimension of the convolutional layer so that the element addition operation can be performed smoothly when performing residual connection.

The residual connection is an important part of the structure. It adds the input (or downsampled input) to the output processed by the SE module. This design can effectively alleviate the vanishing gradient problem in deep networks, allowing the model to be easier to train. Finally, the result of the residual connection is processed through a ReLU activation function to further enhance the nonlinear expression ability of the network. In general, SEResNeXtBottleneck combines the bottleneck design of ResNeXt and the advantages of the squeeze-and-excitation module and enhances feature extraction and expression capabilities through grouped convolution and adaptive recalibration. Grouped convolution captures multi-scale and multi-directional information by processing different feature groups in parallel, while the SE module strengthens the expression of important features and suppresses unimportant features through weight adjustment between channels. This design not only improves the performance of the model but also finds a balance between computational efficiency and feature expression capabilities, forming a powerful network module suitable for building deeper neural networks.

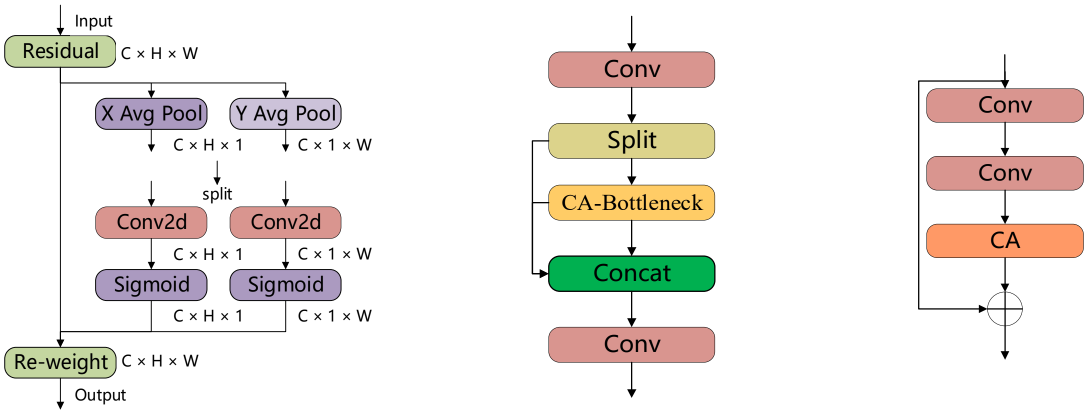
The advantages of this detection head are mainly reflected in the following aspects:

(1) Adaptive feature enhancement: The SE module captures global spatial information through global average pooling and learns the relationship between channels through the fully connected layer, achieving adaptive enhancement of useful features and suppression of redundant features.
(2) Multi-scale feature learning: The use of grouped convolution allows the model to learn features of different scales at the same time, which is particularly important for small target detection in complex scenes.
(3) Computing efficiency: Although SEResNeXt adds the SE module, its exquisite design avoids excessive consumption of computing resources and maintains the real-time nature of the algorithm.
(4) Modular design: The modular characteristics of the detection head enable it to be easily integrated into the existing YOLOv8 architecture without large-scale changes, making it easy to implement and maintain.
(5) Improve detection accuracy: Experimental results of the SEResNeXt detection head on multiple standard data sets show that it can effectively improve the detection accuracy of small targets and targets in complex backgrounds.
(6) Strong generalization ability: Thanks to the combination of parallel paths and channel attention mechanisms, the detection head shows good generalization ability and can adapt to changing industrial environments and different operating conditions.

In summary, the introduction of the SEResNeXt detection head provides an effective improvement solution for the YOLOv8 algorithm in the helmet-wearing detection task. It significantly enhances the accuracy and robustness of detection, offering a powerful solution for industrial safety production and technical support.

3.2.4. C2f-CA

In the field of industrial safety monitoring, especially for helmet-wearing detection in complex working environments, traditional target detection algorithms often have difficulty coping with the challenges of small target size and low resolution [66]. These challenges mainly stem from the design limitations of the existing CNN architecture, especially the loss of detailed information in the feature extraction stage [67]. To solve this problem, the C2f-CA module is introduced in the YOLOv8 backbone, which is an improved module based on coordinate attention (CA) and is specifically optimized for the detection performance of small and dense targets [68], as shown in Figure 5.



(**a**) Structure of SE module (**b**) SEResNeXtBottleneck (**c**) Detect-SEResNeXt

Figure 5. C2f-CA structure.

The core idea of the C2f-CA module is to enhance the model's ability to capture key information by introducing an attention mechanism in the feature extraction process. The module consists of two key parts: first, there are two convolutional layers. The purpose of the first convolutional layer is to double the number of channels of the input feature map without changing the size of the feature map. This can enrich the feature representation and provide richer information for subsequent detection tasks. Next, the second convolutional layer concatenates the output of the first convolutional layer with the output of the CA module. This step not only increases the dimension of the feature but also enhances the model's ability to express the target through feature fusion. The design of the CA module cleverly combines spatial position information with channel attention, allowing the network to locate the target object more accurately, especially for small targets in complex backgrounds, such as helmets, thereby significantly improving the detection accuracy.

In the YOLOv8 architecture, the C2f-CA module is integrated into the neck part, which is the key link in the model responsible for feature fusion. We choose to introduce the C2f-CA module in the first two C2f structures of the neck because this position can maximize the effect of the module. At this stage, the resolution of the feature map is moderate, which contains sufficient spatial details and integrates the high-level semantic information of the deep network. Through the processing of the C2f-CA module, the feature saliency of small targets can be effectively improved, the background interference can be reduced, and a more accurate and robust feature representation can be provided for the final detection head. In addition, the computational complexity of the C2f-CA module is relatively small and does not bring additional computational burden to the model, which makes it easy to integrate into lightweight networks such as YOLOv8 to meet real-time requirements.

4. Experiments and Results

4.1. Experimental Environment

All models were trained and tested on a graphics workstation equipped with the Windows 10 operating system. The core computing graphics card used was an NVIDIA RTX A5000 with 24 GB video memory, and the graphics driver was CUDA 11.3. The Python version was 3.9, and the Torch version was 1.10.2. The model training process was set to 100 training cycles, and the number of images per batch was 48. The initial learning rate was set to 0.03, with a final learning rate of 0.01. We utilized the SGD optimizer and set the weight decay to 0.0005.

4.2. Data Set Introduction

For this research, we aggregated open-source safety helmet datasets, such as the Safe Helmet Wearing Dataset (SHWD) and the Electric Warehousing Helmet Detection (EWHD) dataset, accessible online. The collection totals 17,000 images, each meticulously captured under various conditions to depict distinct production scenarios, including construction operations, emergency power repairs, and industrial manufacturing settings. This number significantly surpasses the scale of datasets commonly used in prior research, which typically range from hundreds to a few thousand, with only a handful exceeding 10,000 images, as noted in References [69–71]. This extensive dataset includes a broad spectrum of positive and negative object samples, highlighting multiple varieties of safety helmets, ranging from small to obscured and densely packed, as well as those subject to potential interference from other types of headwear, as illustrated in Figure 6.

(**a**) occluded objects (**b**) small objects (**c**) fuzzy objects

(**d**) interference objects (**e**) dense objects (**f**) complex objects

Figure 6. Different kinds of samples in public datasets.

4.2.1. Data Collection and Annotation

To address the limitations of existing datasets in capturing the diverse scenarios of safety helmet usage in real-world construction environments, we have curated a comprehensive dataset by integrating multiple safety helmet datasets. This includes the Safety Helmet Wearing Dataset, Hardhat and Safety Vest Image for Object Detection dataset, Realistic Safety Clothing and Helmet Detection (SFCHD) dataset, and the Power Inspection Safety Helmet dataset.

Finally, our dataset consists of a total of 17,000 images captured from various construction sites, encompassing diverse scenarios such as road reconstruction, expansion, and significant/medium repair activities. These images depict workers in various postures—standing, squatting, bending—captured from multiple angles and distances, both indoors and outdoors. Additionally, the dataset includes images of workers wearing different-colored helmets, as well as instances of helmet removal and donning.

4.2.2. Data Set Labeling and Processing

The collected images underwent meticulous annotation and processing to ensure the accuracy and robustness of our dataset. Utilizing the Labellmg tool, each image was annotated with two categories of labels: “helmet” (0) and “no_helmet” (1), following the YOLO format. The annotation process yielded a total of 190,548 annotated objects, comprising 62,282 instances of safety helmets and 128,266 instances without safety helmets. To further enhance the robustness of our dataset, noise, random flip, and enhanced brightness were incorporated into the original images, thereby diversifying the training samples and improving the model’s generalizability.

4.2.3. Data Set Splitting

For training and evaluation purposes, the dataset was partitioned into the training set, test set, and validation set, maintaining an 8:1:1 ratio. This partitioning scheme ensures a balanced distribution of data for model training and evaluation, thereby facilitating effective model development and performance assessment.

4.3. Testing Model Evaluation Index

The experiment uses mean average precision (mAP), model parameters (Params), computational effort (GFLOPs), and frames per second (FPS) as evaluation indicators for small target detection methods. The mean average precision is related to precision (P) and recall (R) [72]. The calculation formulas for precision P and recall R are as follows, where TP is the number of successfully detected targets, FP is the number of incorrectly detected targets, and FN is the number of correct targets that were not detected.

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

Among them, TP (true positives) refers to correctly identified positive samples, FN (false negatives) refers to incorrectly identified positive samples, and FP (false positives) refers to incorrectly identified negative samples.

mAP@0.5 is the average detection accuracy of all categories when the IoU threshold is 0.5; mAP@0.5:0.95 is the average detection accuracy of all IoU thresholds between 0.5 and 0.95 with a step size of 0.05. IoU represents the average intersection-over-union ratio [73]. The formula for calculating mAP is:

$$mAP = \frac{\sum_{n=1}^{N} \mathrm{A}P_n}{N} \tag{3}$$

AP_n is the average precision of the nth category at a certain confidence threshold, and N is the total number of categories.

4.4. Ablation Experiment

To further investigate the degree of optimization of the detection performance of the original model through various improvement strategies, ablation experiments were conducted under the same parameters for comparative analysis. These experiments involved sequential integration modules, namely Yolov8s, small object detection head (4P), SPD-Conv module (Spd) for backbone enhancement, newly designed detection head (SEResNeXt), and C2f-CA module (C2f-CA) for neck enhancement. The results are summarized in Table 1, where the symbol "√" indicates that the corresponding method was adopted.

Table 1. Comparison of the ablation experiments.

Method	4P	Spd	SEResNeXt	C2f-CA	P	R	mAP_0.5	mAP_0.5:0.95	FLOPs/G	FPS/f·s^{-1}
Yolov8s					89.8	79.0	86.1	51.9	28.4	69
	√				91.4	84.6	90.8	54.3	36.6	62
	√	√			91.9	86.0	92.3	55.8	51.1	62
	√	√	√		92.2	86.4	92.3	55.8	51.1	45
	√	√		√	92.0	86.4	92.4	55.7	50.6	55
YOLOv8s-SNC	√	√	√	√	92.4	86.6	92.6	56.0	52.9	44

It can be seen from the data that with the addition of each module, the precision rate of the model increases by 2.6%, the recall rate increases by 7.6%, mAP_0.5 increases by 6.5%, and mAP_0.5:0.95 increases by 4.1%, which shows that gradually Each module integrated has a positive impact on model performance. In particular, the fully integrated and improved YOLOv8s-SNC model significantly improves the precision and recall rates, as well as the average accuracy under different IoU thresholds, while maintaining a high FPS. At the same time, although the FLOPs increased to 52.9 G, the computing cost increased and the FPS dropped to 44, it was still within the acceptable range. This shows that the proposed algorithm improvement is effective and can improve the performance of the model in target detection tasks while maintaining reasonable computational efficiency and meeting real-time requirements.

Figure 7 depicts the precision curve, recall curve, and mean average precision (mAP_0.5, mAP_0.5:0.95) curve of Yolov8s and YOLOv8s-SNC model training. Obviously, from the initial training stage, the four index curves of the enhanced algorithm show a steeper slope than the original algorithm. Therefore, the four indicators of YOLOv8s-SNC eventually greatly exceed the original algorithm.

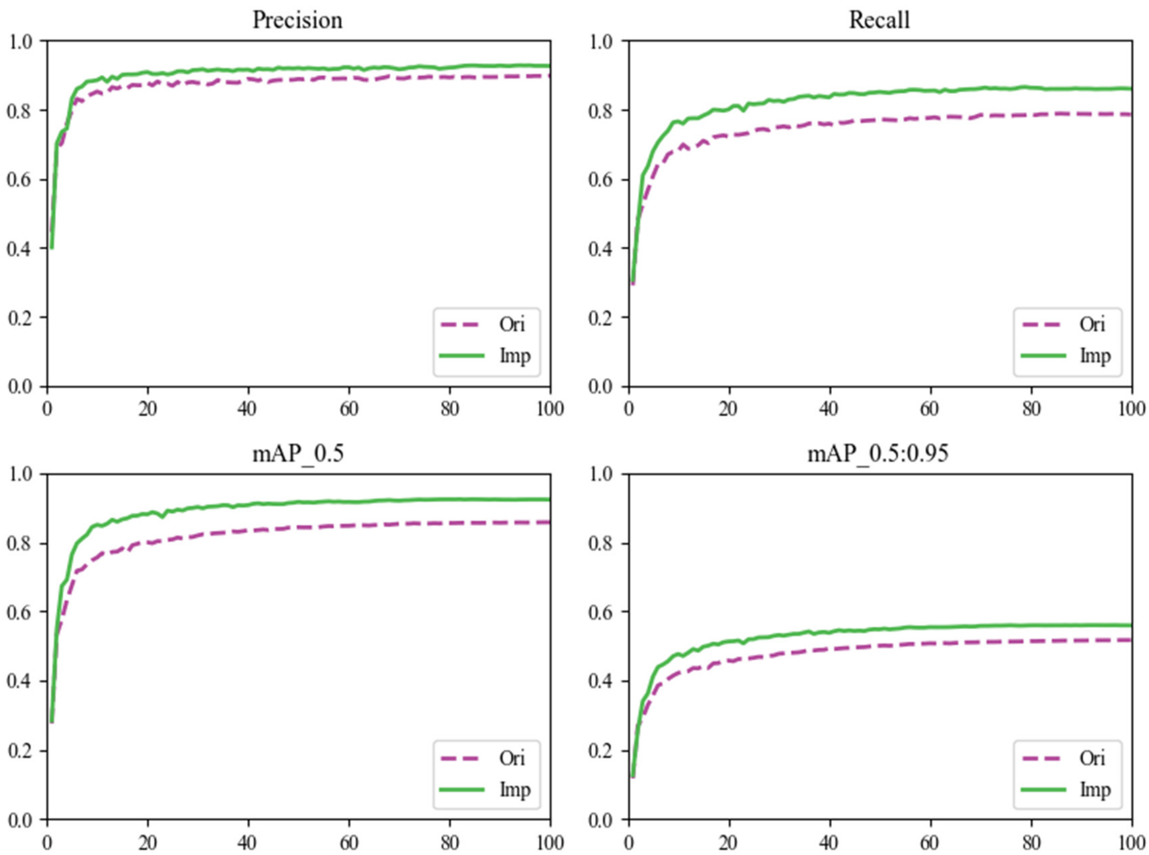


Figure 7. The R and AP curves of YOLOv8 and improved YOLOv8s. "Ori" and "Imp" denote YOLOv8 and improved YOLOv8s, respectively.

4.5. Comparative Experiments

To further quantitatively analyze the effect of the proposed method, in the same experimental environment, the objective indicators of classic target detection algorithms such as SSD, Faster-RCNN, YOLOv3, YOLOv5, YOLOX, and YOLOv8 are calculated and compared with the algorithm proposed in this paper on their corresponding test sets. The results are shown in Table 2.

Table 2. Results of ablation experiments.

Method	P/%	R/%	mAP@.5/%	mAP@.5:.95/%	FPS/f·s^{-1}	Params/MB
SSD	88.3	76.7	84.2	51.6	43	20.3
Faster-RCNN	90.5	80.5	87.4	51.4	23	43.5
RFBNet	83.6	73.4	75.6	41.6	51	98.9
YOLOv5s	89.7	82.4	88.8	53.6	68	8.7
YOLOX	92.0	81.6	88.4	54.8	56	22.3
YOLOv8s	89.8	79.0	86.1	51.9	69	10.6
Ours	92.4	86.6	92.6	56.0	44	14.5

By comparing the experimental results, it can be found that the P, R, and mAP of this algorithm are significantly better than other mainstream target detection algorithms, reaching 92.4%, 86.6%, and 92.6% respectively. The FPS of this algorithm is 44. Although it is slightly lower than other one-stage detection algorithms, it is still nearly twice as high as the two-stage target detection algorithm Faster R-CNN, which shows that while pursuing higher detection accuracy, detection is also considered. Balance of speed: Overall, the algorithm in this article is more suitable for running on mobile devices and more in line with the real-time and accuracy requirements of target detection based on drone images.

4.6. Detection Effect Analysis

In this section, we analyze the detection results of the proposed YOLOv8s-SNC algorithm and compare it with the original YOLOv8s algorithm. We select several representative images to demonstrate the detection performance of both algorithms.

Figure 8 presents the detection results of the YOLOv8s-SNC algorithm and the original YOLOv8s algorithm on the selected images. From Figure 8, we can observe that the YOLOv8s-SNC algorithm demonstrates superior performance in terms of detection accuracy and robustness compared to the original YOLOv8s algorithm.

Compared with a1 in Figure 8, in the scenario with dense small targets and interference, YOLOv8s-SNC can better detect the wearing of workers' helmets in the distant dense crowd, with a lower missed detection rate, and can better distinguish between empty helmets and worn helmets, while the original YOLOv8s has a high missed detection rate and also mistakenly detects empty helmets in the hands of workers.

Compared with a2 in Figure 8, YOLOv8s-SNC can detect workers wearing helmets of different colors more reliably in an environment with a dim background and strong light interference.

At the upward viewing angle, compared with a3 in Figure 8, YOLOv8s-SNC can accurately detect the maintenance workers wearing helmets on the electric poles, which shows that the Qi-YOLOv8s algorithm is more robust to occlusion caused by different shooting angles. Even if a large part of the object is occluded, it can accurately detect the target, which is crucial for accurately detecting the helmet in various scenarios.

In summary, the Qi-YOLOv8s algorithm outperforms the original YOLOv8s algorithm in detection accuracy and robustness. The proposed algorithm can accurately detect helmets in various challenging scenarios, such as dense small targets, complex backgrounds, and occlusions, demonstrating its superiority in the helmet detection task.

(**a1**) (**b1**)

(**a2**) (**b2**)

(**a3**) (**b3**)

Figure 8. Example detection results on the test dataset: (**a1**–**a3**) and (**b1**–**b3**) are the detection results of YOLOv8s and YOLOv8s-SNC respectively.

5. Discussion

YOLOv8s-SNC has shown promising results in enhancing the detection of safety helmets in industrial settings. However, the incorporation of several advanced modules has impacted the FPS, which is a critical parameter for real-time monitoring systems. To address this, model pruning could be considered as a method to optimize the FPS without significantly compromising the detection accuracy [74,75].

Striking a balance between accuracy and speed is essential for the deployment of monitoring systems that can operate efficiently in dynamic environments. Model pruning techniques, which involve the systematic removal of redundant or less important parameters, could help in achieving this balance. This optimization process could potentially lead to a lighter model that maintains high detection accuracy while improving the FPS, thus making the model more suitable for real-time applications [76].

The potential integration of YOLOv8s-SNC with the IoT and smart safety systems presents an opportunity for the automation of safety protocols. This integration could lead to a more proactive approach to safety management, where real-time data from helmet detection systems could trigger alerts or enforce safety measures automatically [77]. Simultaneously, we acknowledge that the algorithm may face challenges in detecting helmets in scenarios where workers are distributed across multiple floors or buildings. Future work will focus on enhancing the algorithm's capabilities in these complex environments, potentially through the integration of additional sensors or the use of building information modeling (BIM) data to improve detection accuracy. However, it is crucial to address ethical considerations and privacy concerns associated with surveillance technology [78]. Ensuring that such systems are designed and deployed with respect for workers' privacy and with clear guidelines on data usage is paramount [79,80].

In conclusion, the YOLOv8s-SNC algorithm holds significant potential to enhance worker safety by improving the detection of safety helmet usage. Future research should focus on expanding the model's capabilities, ensuring robustness across diverse conditions, and addressing ethical considerations to fully harness its potential in industrial safety applications. By doing so, we can move closer to a future where safety monitoring systems are not only accurate but also respectful of workers' rights and privacy.

6. Conclusions

This paper proposes an improved YOLOv8 helmet detection method for the automatic detection of helmet-wearing in complex working scenarios. In order to solve the problem of detail loss caused by multiple downsampling stages, a dedicated small object detection layer is introduced to ensure that the helmet can be accurately detected even in low image resolution; the space-to-depth conversion strategy is adopted, and the SPD-Conv module effectively reduces information loss and enhances the network's perception of small and dense targets, which is crucial for helmet detection in complex working environments; combined with the channel attention mechanism of SENet and the parallel path structure of ResNeXt, the designed SEResNeXt detection head can represent features more finely and improve the detection accuracy in complex scenes; the C2f-CA module is embedded in the specific area of the neck part to enhance the model's ability to capture key information, especially in the detection of small and dense targets, further improving the detection accuracy.

Experimental results show that compared with the original Yolov8s model, YOLOv8s-SNC improves precision by 2.6%, recall by 7.6%, mAP_0.5 by 6.5%, and mAP_0.5:0.95 by 4.1% while maintaining high real-time performance. These improvements demonstrate the effectiveness and application potential of our method in actual industrial safety monitoring; realize automatic detection of small targets, occluded targets, and helmet-wearing in complex working scenarios; improve supervision scope and timeliness; and reduce labor costs.

Moving forward, we aim to optimize the YOLOv8s-SNC model further, leveraging transfer learning and domain adaptation to adapt to diverse industrial environments. We will also explore model pruning to enhance computational efficiency without affecting detection performance. We plan to integrate YOLOv8s-SNC with IoT and smart safety systems for automated safety protocols and real-time monitoring. However, we recognize the ethical and privacy implications of surveillance technology and are committed to respecting workers' privacy and adhering to data usage guidelines.

In summary, the introduction of the YOLOv8s-SNC network not only provides an effective solution for industrial safety helmet detection but also contributes innovative ideas and technologies to the field of small target detection. We anticipate that this research outcome will play a significant role in enhancing worker safety and improving production efficiency, with the potential for broad application in future safety monitoring systems.

Author Contributions: Conceptualization, C.Y.; Methodology, D.H.; Software, Z.J.; Formal analysis, X.W.; Investigation, L.C.; Resources, Z.T., Y.D., L.C. and X.W.; Data curation, Z.T., Y.D. and X.W.; Writing—review & editing, D.H. and C.Y.; Supervision, L.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data used to support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest: Authors Zhenhai Tian and Yanjie Dong were employed by The Fourth Engineering Co., Ltd. of CCCC First Highway Engineering Co., Ltd. Authors Liyuan Chen and Xuguang Wu were employed by Pingyang County Power Supply Company, State Grid Zhejiang Electric Power Co., Ltd. The remaining authors declare that the re-search was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Wang, B.; Wu, C.; Huang, L.; Zhang, L.; Kang, L.; Gao, K. Prevention and control of major accidents (MAs) and particularly serious accidents (PSAs) in the industrial domain in China: Current status, recent efforts and future prospects. *Process Saf. Environ.* **2018**, *117*, 254–266. [CrossRef]
2. Maliha, M.N.; Abu Aisheh, Y.I.; Tayeh, B.A.; Almalki, A. Safety Barriers Identification, Classification, and Ways to Improve Safety Performance in the Architecture, Engineering, and Construction (AEC) Industry: Review Study. *Sustainability* **2021**, *13*, 3316. [CrossRef]
3. Dwyer, T. Industrial safety engineering—Challenges of the future. *Accid. Anal. Prev.* **1992**, *24*, 265–273. [CrossRef]
4. Helander, M.G. Safety hazards and motivation for safe work in the construction industry. *Int. J. Ind. Ergon.* **1991**, *8*, 205–223. [CrossRef]
5. Rasouli, S.; Alipouri, Y.; Chamanzad, S. Smart Personal Protective Equipment (PPE) for construction safety: A literature review. *Saf. Sci.* **2024**, *170*, 106368. [CrossRef]
6. Wong, T.K.M.; Man, S.S.; Chan, A.H.S. Critical factors for the use or non-use of personal protective equipment amongst construction workers. *Saf. Sci.* **2020**, *126*, 104663. [CrossRef]
7. Collie, A. Disparities in death at work: Reflections on occupational injury fatality data. *Occup. Environ. Med.* **2024**, *81*, 167–168. [CrossRef]
8. Lin, B. Safety Helmet Detection Based on Improved YOLOv8. *IEEE Access* **2024**, *12*, 28260–28272. [CrossRef]
9. Hayat, A.; Morgado-Dias, F. Deep Learning-Based Automatic Safety Helmet Detection System for Construction Safety. *Appl. Sci.* **2022**, *12*, 8268. [CrossRef]
10. Zhang, M.; Shi, R.; Yang, Z. A critical review of vision-based occupational health and safety monitoring of construction site workers. *Saf. Sci.* **2020**, *126*, 104658. [CrossRef]
11. Häikiö, J.; Kallio, J.; Mäkelä, S.-M.; Keränen, J. IoT-based safety monitoring from the perspective of construction site workers. *Int. J. Occup. Environ. Saf.* **2020**, *4*, 1–14. [CrossRef]
12. Fang, W.; Love, P.E.; Luo, H.; Ding, L. Computer vision for behaviour-based safety in construction: A review and future directions. *Adv. Eng. Inform.* **2020**, *43*, 100980. [CrossRef]
13. Liang, H.; Seo, S. Automatic Detection of Construction Workers' Helmet Wear Based on Lightweight Deep Learning. *Appl. Sci.* **2022**, *12*, 10369. [CrossRef]
14. Carbonari, A.; Giretti, A.; Naticchia, B. A proactive system for real-time safety management in construction sites. *Autom. Constr.* **2011**, *20*, 686–698. [CrossRef]
15. Papadopoulos, Y.; Mcdermid, J. Automated safety monitoring: A review and classification of methods. *Int. J. Cond. Monit. Diagn. Eng. Manag.* **2001**, *4*, 14–32.
16. Patel, K.; Patel, V.; Prajapati, V.; Chauhan, D.; Haji, A.; Degadwala, S. Safety Helmet Detection Using YOLOV8. In Proceedings of the 2023 3rd International Conference on Pervasive Computing and Social Networking (ICPCSN), Salem, India, 19–20 June 2023; pp. 22–26.
17. Shao, D.; Jiang, J.; Ma, L.; Lai, H.; Yi, S. Real-time medical lesion screening: Accurate and rapid detectors. *J. Real-Time Image Proc.* **2024**, *21*, 134. [CrossRef]
18. Elhanashi, A.; Dini, P.; Saponara, S.; Zheng, Q. TeleStroke: Real-time stroke detection with federated learning and YOLOv8 on edge devices. *J. Real-Time Image Proc.* **2024**, *21*, 121. [CrossRef]
19. Zhu, J.; Hu, T.; Zheng, L.; Zhou, N.; Ge, H.; Hong, Z. YOLOv8-C2f-Faster-EMA: An Improved Underwater Trash Detection Model Based on YOLOv8. *Sensors* **2024**, *24*, 2483. [CrossRef]
20. Farooq, J.; Muaz, M.; Khan Jadoon, K.; Aafaq, N.; Khan, M.K.A. An improved YOLOv8 for foreign object debris detection with optimized architecture for small objects. *Multimed. Tools Appl.* **2024**, *83*, 60921–60947. [CrossRef]
21. Savković, M.; Dasic, M.; Djapan, M.; Vukicevic, A.; Macuzic, I.; Stefanovic, M. *Improving Workplace Safety Using Advanced Industry 4.0 Technologies*; Springer: Cham, Switzerland, 2021; ISBN 978-86-7672-348-5.
22. Rathod, S.B.; Mahajan, R.A.; Khadkikar, P.A.; Vyawahare, H.R.; Patil, P.R. Improving Workplace Safety With AI-Powered Predictive Analytics: Enhancing Workplace Security. In *AI Tools and Applications for Women's Safety*; IGI Global: Hershey, PA, USA, 2024; pp. 232–249. ISBN 9798369314357.
23. Sanjana, S.; Sanjana, S.; Shriya, V.R.; Vaishnavi, G.; Ashwini, K. A review on various methodologies used for vehicle classification, helmet detection and number plate recognition. *Evol. Intel.* **2021**, *14*, 979–987. [CrossRef]
24. Park, M.-W.; Elsafty, N.; Zhu, Z. Hardhat-Wearing Detection for Enhancing On-Site Safety of Construction Workers. *J. Constr. Eng. Manage.* **2015**, *141*, 04015024. [CrossRef]
25. Wu, H.; Zhao, J. An intelligent vision-based approach for helmet identification for work safety. *Comput. Ind.* **2018**, *100*, 267–277. [CrossRef]

26. Chiverton, J. Helmet presence classification with motorcycle detection and tracking. *IET Intell. Transp. Syst.* **2012**, *6*, 259. [CrossRef]
27. Waranusast, R.; Bundon, N.; Timtong, V.; Tangnoi, C.; Pattanathaburt, P. Machine vision techniques for motorcycle safety helmet detection. In Proceedings of the 2013 28th International Conference on Image and Vision Computing New Zealand (IVCNZ 2013), Wellington, New Zealand, 27–29 November 2013; pp. 35–40.
28. Park, M.-W.; Koch, C.; Brilakis, I. Three-Dimensional Tracking of Construction Resources Using an On-Site Camera System. *J. Comput. Civ. Eng.* **2012**, *26*, 541–549. [CrossRef]
29. Lee, P.; Kim, H.; Zitouni, M.S.; Khandoker, A.; Jelinek, H.F.; Hadjileontiadis, L.; Lee, U.; Jeong, Y. Trends in Smart Helmets With Multimodal Sensing for Health and Safety: Scoping Review. *JMIR Mhealth Uhealth* **2022**, *10*, e40797. [CrossRef]
30. Zhang, H.; Yan, X.; Li, H.; Jin, R.; Fu, H. Real-Time Alarming, Monitoring, and Locating for Non-Hard-Hat Use in Construction. *J. Constr. Eng. Manag.* **2019**, *145*, 04019006. [CrossRef]
31. Wang, C.; Kim, Y.; Kim, D.G.; Lee, S.H.; Min, S.D. Smart Helmet and Insole Sensors for Near Fall Incidence Recognition during Descent of Stairs. *Appl. Sci.* **2020**, *10*, 2262. [CrossRef]
32. García, A.F.; Biain, X.O.; Lingos, K.; Konstantoudakis, K.; Hernández, A.B.; Iragorri, I.A.; Zarpalas, D. Smart Helmet: Combining Sensors, AI, Augmented Reality, and Personal Protection to Enhance First Responders' Situational Awareness. *IT Prof.* **2023**, *25*, 45–53. [CrossRef]
33. 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. *Automat. Constr.* **2019**, *106*, 102894. [CrossRef]
34. Chen, D.; Zhao, H.; Liu, W.; Du, D. Improved Safety Helmet-Wearing Detection Algorithm Based on YOLOv5. In *Proceedings of the 2023 Chinese Intelligent Systems Conference*; Springer: Singapore, 2023; pp. 433–442.
35. Song, X.; Zhang, T.; Yi, W. An Improved YOLOv8 Safety Helmet Wearing Detection Network 2024. *Sci. Rep.* **2024**, *14*, 17550. [CrossRef]
36. Wu, X.; Li, Y.; Long, J.; Zhang, S.; Wan, S.; Mei, S. A Remote-Vision-Based Safety Helmet and Harness Monitoring System Based on Attribute Knowledge Modeling. *Remote Sens.* **2023**, *15*, 347. [CrossRef]
37. Bo, Y.; Huan, Q.; Huan, X.; Rong, Z.; Hongbin, L.; Kebin, M.; Weizhong, Z.; Lei, Z. Helmet Detection under the Power Construction Scene Based on Image Analysis. In Proceedings of the 2019 IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT), Dalian, China, 19–20 October 2019; pp. 67–71.
38. Wu, F.; Jin, G.; Gao, M.; HE, Z.; Yang, Y. Helmet Detection Based On Improved YOLO V3 Deep Model. In Proceedings of the 2019 IEEE 16th International Conference on Networking, Sensing and Control (ICNSC), Banff, AB, Canada, 9–11 May 2019; pp. 363–368.
39. Huang, L.; Fu, Q.; He, M.; Jiang, D.; Hao, Z. Detection algorithm of safety helmet wearing based on deep learning. *Concurr. Comput.* **2021**, *33*, e6234. [CrossRef]
40. Wang, L.; Xie, L.; Yang, P.; Deng, Q.; Du, S.; Xu, L. Hardhat-Wearing Detection Based on a Lightweight Convolutional Neural Network with Multi-Scale Features and a Top-Down Module. *Sensors* **2020**, *20*, 1868. [CrossRef] [PubMed]
41. Nipun, D. Nath Deep learning for site safety: Real-time detection of personal protective equipment. *Autom. Constr.* **2020**, *112*, 103085. [CrossRef]
42. Gao, T.; Wushouer, M.; Tuerhong, G. DMS-YOLOv5: A Decoupled Multi-Scale YOLOv5 Method for Small Object Detection. *Appl. Sci.* **2023**, *13*, 6124. [CrossRef]
43. Vijayakumar, A.; Vairavasundaram, S. YOLO-based Object Detection Models: A Review and its Applications. *Multimed. Tools Appl.* **2024**, 1–40. [CrossRef]
44. Sohan, M.; Sai Ram, T.; Rami Reddy, C.V. A Review on YOLOv8 and Its Advancements. In *Proceedings of the Data Intelligence and Cognitive Informatics*; Jacob, I.J., Piramuthu, S., Falkowski-Gilski, P., Eds.; Springer Nature: Singapore, 2024; pp. 529–545.
45. Zuo, X.; Qi, C.; Chen, Y.; Shen, J.; Fan, H.; Yang, W. Learnable Cross-Scale Sparse Attention Guided Feature Fusion for UAV Object Detection. *IEEE Access* **2024**, *12*, 114212–114226. [CrossRef]
46. Wang, C.-Y.; Bochkovskiy, A.; Liao, H.-Y.M. YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. 2023, pp. 7464–7475. Available online: https://arxiv.org/abs/2207.02696 (accessed on 3 October 2024).
47. Li, Z.; He, Q.; Zhao, H.; Yang, W. Doublem-net: Multi-scale spatial pyramid pooling-fast and multi-path adaptive feature pyramid network for UAV detection. *Int. J. Mach. Learn. Cybern.* **2024**, 1–25. [CrossRef]
48. Shi, P.; He, Q.; Zhu, S.; Li, X.; Fan, X.; Xin, Y. Multi-scale fusion and efficient feature extraction for enhanced sonar image object detection. *Expert Syst. Appl.* **2024**, *256*, 124958. [CrossRef]
49. Zhang, N.; Donahue, J.; Girshick, R.; Darrell, T. Part-Based R-CNNs for Fine-Grained Category Detection. In Proceedings of the Computer Vision—ECCV 2014: 13th European Conference, Zurich, Switzerland, 6–12 September 2014; Proceedings, Part I 13. Springer International Publishing: Cham, Switzerland, 2014; pp. 834–849.
50. Li, Y.; Yu, W.; Guan, X. Current-Aided Multiple-AUV Cooperative Localization and Target Tracking in Anchor-Free Environments. *IEEE/CAA J. Autom. Sin.* **2023**, *10*, 792–806. [CrossRef]
51. Vo, X.-T.; Jo, K.-H. A review on anchor assignment and sampling heuristics in deep learning-based object detection. *Neurocomputing* **2022**, *506*, 96–116. [CrossRef]
52. Xiong, C.; Zayed, T.; Abdelkader, E.M. A novel YOLOv8-GAM-Wise-IoU model for automated detection of bridge surface cracks. *Constr. Build. Mater.* **2024**, *414*, 135025. [CrossRef]

53. Pham, D.-A.; Han, S.-H. Deploying a Computer Vision Model Based on YOLOv8 Suitable for Drones in the Tuna Fishing and Aquaculture Industry. *J. Mar. Sci. Eng.* **2024**, *12*, 828. [CrossRef]
54. Zhang, R.; Zhu, F.; Liu, J.; Liu, G. Depth-Wise Separable Convolutions and Multi-Level Pooling for an Efficient Spatial CNN-Based Steganalysis. *IEEE Trans. Inf. Forensics Secur.* **2020**, *15*, 1138–1150. [CrossRef]
55. Chen, F.; Zhang, L.; Kang, S.; Chen, L.; Dong, H.; Li, D.; Wu, X. Soft-NMS-Enabled YOLOv5 with SIOU for Small Water Surface Floater Detection in UAV-Captured Images. *Sustainability* **2023**, *15*, 10751. [CrossRef]
56. Wang, G.; Chen, Y.; An, P.; Hong, H.; Hu, J.; Huang, T. UAV-YOLOv8: A Small-Object-Detection Model Based on Improved YOLOv8 for UAV Aerial Photography Scenarios. *Sensors* **2023**, *23*, 7190. [CrossRef]
57. Xu, W.; Cui, C.; Ji, Y.; Li, X.; Li, S. YOLOv8-MPEB small target detection algorithm based on UAV images. *Heliyon* **2024**, *10*, e29501. [CrossRef]
58. Zhang, Z.; Qu, Y.; Wang, T.; Rao, Y.; Jiang, D.; Li, S.; Wang, Y. An Improved YOLOv8n Used for Fish Detection in Natural Water Environments. *Animals* **2024**, *14*, 2022. [CrossRef]
59. Sunkara, R.; Tie, L. No more strided convolutions or pooling: A new CNN building block for low-resolution images and small objects. In *Proceedings of the Machine Learning and Knowledge Discovery in Databases*; Amini, M.-R., Canu, S., Fischer, A., Guns, T., Novak, P.K., Tsoumakas, G., Eds.; Springer Nature: Cham, Switzerland, 2023; pp. 443–459.
60. Liu, S.; Cao, L.; Li, Y. Lightweight Pedestrian Detection Network for UAV Remote Sensing Images Based on Strideless Pooling. *Remote Sens.* **2024**, *16*, 2331. [CrossRef]
61. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141. [CrossRef]
62. Hitawala, S. Evaluating ResNeXt Model Architecture for Image Classification. *arXiv* **2018**, arXiv:1805.08700. [CrossRef]
63. Gao, J.; Chen, Y.; Wei, Y.; Li, J. Detection of Specific Building in Remote Sensing Images Using a Novel YOLO-S-CIOU Model. Case: Gas Station Identification. *Sensors* **2021**, *21*, 1375. [CrossRef] [PubMed]
64. Li, H.; Wang, Z.; Yue, X.; Wang, W.; Tomiyama, H.; Meng, L. An architecture-level analysis on deep learning models for low-impact computations. *Artif. Intell. Rev.* **2023**, *56*, 1971–2010. [CrossRef]
65. Zhu, H.J.; Gu, W.; Wang, L.M.; Xu, Z.C.; Sheng, V.S. Android malware detection based on multi-head squeeze-and-excitation residual network. *Expert Syst. Appl.* **2023**, *212*, 118705. [CrossRef]
66. Hollands, J.G.; Terhaar, P.; Pavlovic, N.J. Effects of Resolution, Range, and Image Contrast on Target Acquisition Performance. *Hum. Factors* **2018**, *60*, 363–383. [CrossRef]
67. Zhang, Z.; Tan, L.; Robert, T.L.K. An Improved Fire and Smoke Detection Method Based on YOLOv8n for Smart Factories. *Sensors* **2024**, *24*, 4786. [CrossRef]
68. Hou, Q.; Zhou, D.; Feng, J. Coordinate Attention for Efficient Mobile Network Design. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 13713–13722. [CrossRef]
69. Fu, Z.; Ling, J.; Yuan, X.; Li, H.; Li, H.; Li, Y. Yolov8n-FADS: A Study for Enhancing Miners' Helmet Detection Accuracy in Complex Underground Environments. *Sensors* **2024**, *24*, 3767. [CrossRef]
70. Xu, H.; Wu, Z. MCX-YOLOv5: Efficient helmet detection in complex power warehouse scenarios. *J. Real-Time Image Pract.* **2024**, *21*, 27. [CrossRef]
71. Zhang, Y.; Qiu, Y.; Bai, H. FEFD-YOLOV5: A Helmet Detection Algorithm Combined with Feature Enhancement and Feature Denoising. *Electronics* **2023**, *12*, 2902. [CrossRef]
72. Padilla, R.; Netto, S.L.; da Silva, E.A.B. A Survey on Performance Metrics for Object-Detection Algorithms. In Proceedings of the 2020 International Conference on Systems, Signals and Image Processing (IWSSIP), Niteroi, Brazil, 1–3 July 2020; pp. 237–242.
73. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef]
74. Menghani, G. Efficient Deep Learning: A Survey on Making Deep Learning Models Smaller, Faster, and Better. *ACM Comput. Surv.* **2023**, *55*, 259:1–259:37. [CrossRef]
75. Tmamna, J.; Ayed, E.B.; Fourati, R.; Gogate, M.; Arslan, T.; Hussain, A.; Ayed, M.B. Pruning Deep Neural Networks for Green Energy-Efficient Models: A Survey. *Cogn. Comput.* **2024**, *2024*, 1–12. [CrossRef]
76. Gupta, M.; Agrawal, P. Compression of Deep Learning Models for Text: A Survey. *ACM Trans. Knowl. Discov. Data* **2022**, *16*, 61:1–61:55. [CrossRef]
77. Jayasree, V.; Kumari, M.N. IOT Based Smart Helmet for Construction Workers. In Proceedings of the 2020 7th International Conference on Smart Structures and Systems (ICSSS), Chennai, India, 23–24 July 2020; pp. 1–5.
78. Henriksen-Bulmer, J.; Faily, S.; Jeary, S. Privacy risk assessment in context: A meta-model based on contextual integrity. *Comput. Secur.* **2019**, *82*, 270–283. [CrossRef]
79. Zwitter, A. Big Data ethics. *Big Data Soc.* **2014**, *1*, 2053951714559253. [CrossRef]
80. Markham, A.N.; Tiidenberg, K.; Herman, A. Ethics as Methods: Doing Ethics in the Era of Big Data Research—Introduction. *Soc. Media + Soc.* **2018**, *4*, 2056305118784502. [CrossRef]