applied sciences

MDPI

Article

A Super-Resolution Reconstruction Driven Helmet Detection Workflow

Yicheng Liu [1], Zhipeng Li [1], Bixiong Zhan [2], Ju Han [2] and Yan Liu [1,*]

[1] College of Electrical Engineering, Sichuan University, Chengdu 610065, China; liuyicheng@scu.edu.cn (Y.L.); 2020223035121@stu.scu.edu.cn (Z.L.)

[2] China Construction First Group Construction & Development Co., Ltd., Beijing 100102, China; Zhanbixiong@chinaonebuild.com (B.Z.); hanju@chinaonebuild.com (J.H.)

[*] Correspondence: debbie_ly77@126.com

Abstract: The degrading of input images due to the engineering environment decreases the performance of helmet detection models so as to prevent their application in practice. To overcome this problem, we propose an end-to-end helmet monitoring system, which implements a super-resolution (SR) reconstruction driven helmet detection workflow to detect helmets for monitoring tasks. The monitoring system consists of two modules, the super-resolution reconstruction module and the detection module. The former implements the SR algorithm to produce high-resolution images, the latter performs the helmet detection. Validations are performed on both a public dataset as well as the realistic dataset obtained from a practical construction site. The results show that the proposed system achieves a promising performance and surpasses the competing methods. It will be a promising tool for construction monitoring and is easy to be extended to corresponding tasks.

Keywords: helmet detection; super-resolution reconstruction; you only look once v5 (YOLOv5)

check for updates

Citation: Liu, Y.; Li, Z.; Zhan, B.; Han, J.; Liu, Y. A Super-Resolution Reconstruction Driven Helmet Detection Workflow. *Appl. Sci.* **2022**, *12*, 545. https://doi.org/10.3390/app12020545

Academic Editor: Vincent A. Cicirello

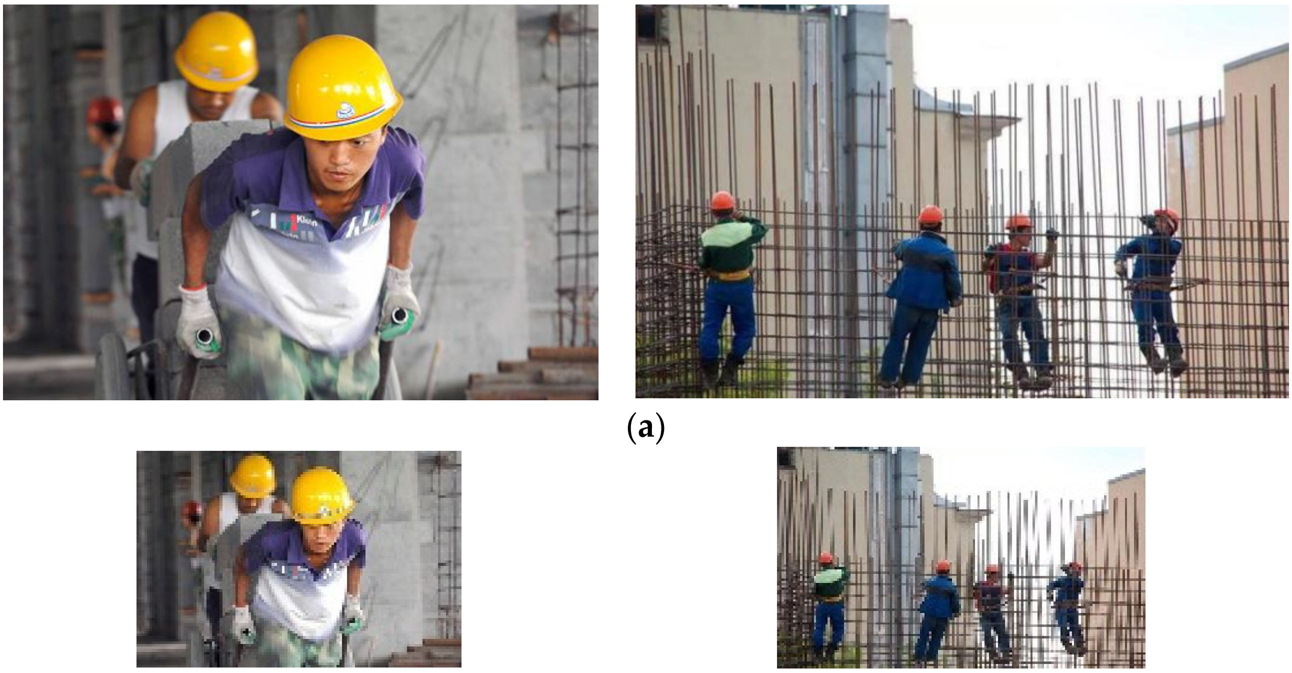
Received: 22 November 2021
Accepted: 29 December 2021
Published: 6 January 2022

1. Introduction

Given the continued rapid growth of modern society worldwide, the construction industry has developed increasingly fast. However, as well as the traditional hot issues such as design and construction technology, the safety of construction sites is becoming one of the hottest topics in the current construction industry. The construction industry is one of the most prone to safety accidents among all industries. Over the past 20 years, the construction industry has experienced a decline in accident rates, but the industry's accident rate is still about three times that of all industries [1]. Therefore, it is of great practical significance to study the safety guarantee in the construction industry.

Among the safety accidents in the construction industry, the level of disability caused by head injury is the highest, thus, reducing head injuries is obviously the primary objective to ensure the safety of personnel [2]. Currently, the helmet is the most effective way to reduce head injuries because the impact resistance of the helmet can disperse the impact of weights when an accident occurs, so as to greatly reduce the head and neck injuries caused by impact. In order to protect the life safety of employees, the common rule is to force people to wear helmets before entering the construction area and starting working. Practically, it is difficult to implement this requirement since the workload of maintaining supervision is tedious and labor-consuming. Consequently, several kinds of monitoring systems have been developed including fixed camera-based systems and moving camera-based processes. These systems work in a closed loop, which captures information through cameras and performs analysis manually or automatically to produce alarms or any other control signals. It is obvious that moving cameras are more flexible and able to cover wider regions so these have attracted a larger amount of attention from users and researchers. Nonetheless, moving cameras, either cameras on-board an unmanned aerial vehicle (UAV) or web cameras, normally suffer from the problem of data transmission, according to the fact that monitoring image or video files always have a large size and ask for a wide bandwidth, which is a precious resource in wireless communication. To overcome this problem, compressing files by decreasing the resolution can be performed, but this often degrades the image quality, which is hard to be avoided during the trans

Mission, as shown in Figure 1. This ongoing chain reaction means that the difficulties of monitoring and analyzing images or videos are significantly increasing. In all honesty, this is not a big issue for manual analysis since human brains can adapt to this degrading in most cases, but it really affects the performance of most automatic analyzing methods. Taking into consideration the fact that automatic methods surpass manual approaches due to their advantages of being more effective and less labor-consuming, there is an urgent need for the improvement of automatic helmet detection performance based on low-resolution image or video files, which is the motivation of this paper.



(**a**)

(**b**)

Figure 1. The example of high-resolution images acquired from digital camera and the degraded low-resolution images obtained from the wireless channel due to the constraint of the available transmission bandwidth. (**a**) high-resolution images; (**b**) low-resolution images.

In recent years, there have been a large number of achievements focused on the automatic detection of helmets. Based on the employed research ideas, these methods can be divided into two kinds. Some of them use traditional solutions of object detection tasks, which consist of handcraft feature designing and machine learning models such as a support vector machine. The predesigned features are usually from general computer vision tasks, such as Haar-like features from face detection [3], and the deformable parts model (DPM), which could cascade different single features [4]. However, choosing or designing features is complicated, easy to be prone to poor accuracy and difficult to be extended from one scenario to another. The other methods use deep learning to solve object detection. The region-based convolutional neural network (RCNN) combines the selective search algorithm and convolutional neural network to detect targets, which makes for great improvements in accuracy and speed compared with traditional methods [5]. An RCNN is a two-stage detection network, the speed of which is difficult to meet in real construction engineering. The you only look once model (YOLO) realizes faster detection based on the improvement of the feature extractor backbone. Due to their significant improvement of detection performance, they replace traditional methods in a short time. Actually, most modern object detection architectures are developed and validated on prepared datasets, which consist of high-quality images. Although they achieve good accuracy on those well-captured images, their performance still decreases quickly as the input quality decreases. In other words, the design of detection architectures assumes that the input image quality meets the demand. However, in our application scenario, when these models are applied on detecting helmets from images obtained through moving cameras, which are degraded in quality, it is hard to prevent the performance crashing.

To solve this problem, we propose a super-resolution reconstruction driven helmet detection workflow to improve detection accuracy under poor image quality. The main contributions of the paper are as follows.

(1) We propose an end-to-end helmet monitoring system, which implements a super-resolution reconstruction driven helmet detection workflow. It works well with poor input image quality and is easy to collaborate with any kinds of image acquisition device, including a wireless web camera or UAV.

(2) We propose to train a super-resolution model with combination loss of l_1 and contextual loss, which enhance its accuracy. We train the super-resolution reconstruction model and the detection model iteratively from scratch to achieve final results.

(3) Validations are performed on both a public dataset as well as the realistic dataset obtained from a practical construction site. The results show the proposed workflow achieves a promising performance and surpasses the competing methods.

2. Related Work

2.1. Object Detection

As one of the fast-developing fields in recent years, deep learning-based object detection algorithms are becoming the leading methods to solve object detection tasks. Most successful methods can be divided into two main categories, two-stage detection and one-stage detection. The most representative methods following two-stage detection including the RCNN, Fast regions with convolutional neural networks (Fast-RCNN) and their variants [6,7]. The common idea of these methods is first to obtain region proposals, which might contain the objects, then change the task into a classification to attach each anchor box a label. It is almost the standard solution for long periods of time due to its relatively high accuracy. However, it has proven to be of less help in practical scenarios that ask for real-time monitoring because of their low detection speed. In contrast, one-stage detection methods try to solve the problem through regression. This first employs a feature extracting backbone, usually a convolutional neural network (CNN)-based one, to produce feature maps, then predicts the position, class and confidence of objects at the same time. Based on the improvement of the feature extractor backbone, YOLO evolves from YOLOv1 to YOLOv3 to achieve better accuracy and speed [8–10]. Adopting a cross stage partial network (CSP)-based darknet-53 as the backbone network and replacing feature pyramid networks (FPN) with a path aggregation network (PANet), YOLOv4 improves the detection accuracy of the model in advance [11]. Recently, the YOLOv5 network model has added a focus structure to the backbone network on the base of YOLOv4, and balanced the detection speed and accuracy. Currently, one-stage detection methods are widely used in engineering practice due to the good time efficiency. Nevertheless, in most cases, the accuracy of YOLO and its variants is not as high as that of two-stage methods, especially for small targets and the low-resolution input.

2.2. Super-Resolution Reconstruction

There have been a large number of attempts to improve the performance of super-resolution reconstruction as it is really a long story in the development of computer vision. The most widely used approaches are kinds of interpolation-based methods, such as bilinear interpolation or nearest-neighboring interpolation [12]. Since the process of interpolation always follows a fixed pattern to calculate the new-generated pixel values from existing ones in a low-resolution image, it is hard to adapt to an unknown image degrading protocol.

Another traditional idea is to treat the SR problem as image reconstruction [13,14]. Inspired by the learning method, recent super-resolution approaches directly learn the nonlinear relationship from the low–high-resolution images. Based on the learning process, it could be divided into supervised SR and unsupervised SR. The supervised SR requires aligned high- and low–high-resolution image pairs to train the CNN models to fit the mapping between images with different resolutions. Dong et al. propose the super-resolution convolutional neural networks (SRCNN), which effectively improves the effect and speed of image SR reconstruction compared with the traditional image SR algorithms [15]. Kim et al. propose a VDSR network, which increases the number of layers of CNN to 20 [16]. The algorithm combines residual structure and CNN with image SR reconstruction, and the image reconstruction effect has been significantly improved. Li et al. propose a multi-scale residual network (MSRN), which applies image multi-scale features to the residual structure to further improve the image reconstruction effect [17]. Zhang et al. propose a residual channel attention network RCAN, which applies a channel attention mechanism to the image super-resolution problem and achieves a better reconstruction effect than previous algorithms [18]. To apply these methods successfully, we need to prepare a large number of strictly aligned image pairs, which is not a simple task in practice. Thus, currently, simulated image pairs are used in research so as not to decrease its performance in real world applications. Unsupervised SR methods employ the GAN model and its variants to generate high-resolution images with the low-resolution input [19–21]. However, without paired training data, its accuracy is not as good as that of supervised methods.

2.3. Helmet Detection

Automatic helmet detection is urgently needed in construction engineering and safety driving monitoring. The traditional helmet detection methods focus on the design of the artificial features to lead the classification towards appropriately discriminating helmet from non-helmet targets. The well-known image descriptors such as the local binary pattern (LBP), local variance (LV) and histogram of oriented gradient (HOG) are used to enhance the feature extraction step, and they achieve promising accuracy through a supporting vector machine [22]. The circular Hough transform (CHT) accompanied with HOG descriptor are applied to extract the helmet attributes, and the multilayer perceptron (MLP) classifier is used to perform the final helmet classification [23]. The method combining multi-feature fusion and a support vector machine (SVM) is used to detect and track the helmet in a factory environment to keep an eye on safety production [24]. However, the choosing of manual features is labor-consuming and poor in generalization, which prevents their application. Nowadays, thanks to the development of deep learning and CNN, a large number of modern helmet detection approaches have been proposed. The CNN-based multi-task learning model has been designed for tracking individual motorcycles through identifying helmets [25]. The faster region-based convolutional neural network (Faster RCNN) is utilized to detect both motorcyclists and helmets [26]. The faster RCNN equipped with the multi-scale training and increasing anchors strategies has proved to be capable of detecting helmets on different scales [27]. Taking the processing speed into consideration, YOLO is even more popular. An improved YOLOv3 model has been applied to detect helmets and successfully increased the average accuracy [28]. Replacing the traditional YOLOv3 backbone of darknet-53 with a deep separable convolution structure, the performance of helmet detection has been further improved [29]. Nevertheless, all these models ask for quality stable input images, which is reachable in training set acquisition but difficult to reach in practical terms. In other words, if there is no assurance about input quality, the detection performance will decrease quickly. To the best of our knowledge, few methods have focused on solving the helmet detection problem with poor quality input images.

3. Method

We propose an end-to-end helmet monitoring system, which implements a super-resolution reconstruction driven helmet detection workflow as shown in Figure 2. There are two main modules in the workflow, the super-resolution reconstruction module and the detection module. The former implements the SR algorithm to produce high-resolution images, while the detection model performs the helmet detection. Based on the helmet detection results, we can perform semantic analysis of counting or wearing detection based on the specific monitoring task.

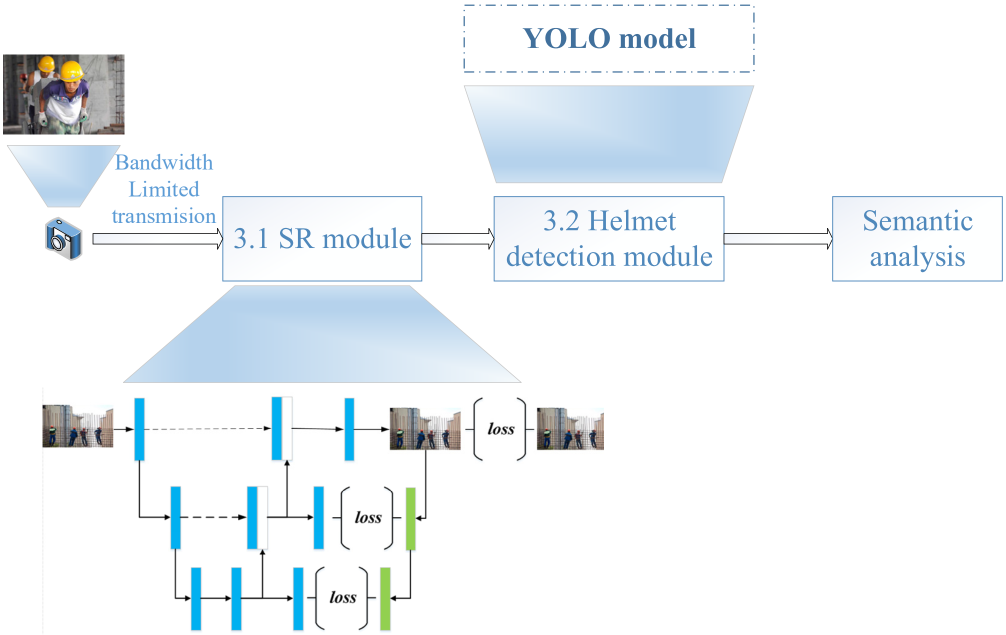


Figure 2. The workflow of the proposed end-to-end helmet monitoring system.

3.1. SR Module

We take the dual regression network as our backbone architecture for super-resolution reconstruction [30]. The main idea is to add a dual regression task (from high-resolution image to low-resolution image) alongside the primal regression task (from low-resolution image to high-resolution image). Through the constraint of the reversible reconstruction, the mapping space between the low–high-resolution images is compressed and it is easier to fit to the real degrading relationship. The details of the SR model are shown in Figure 3.

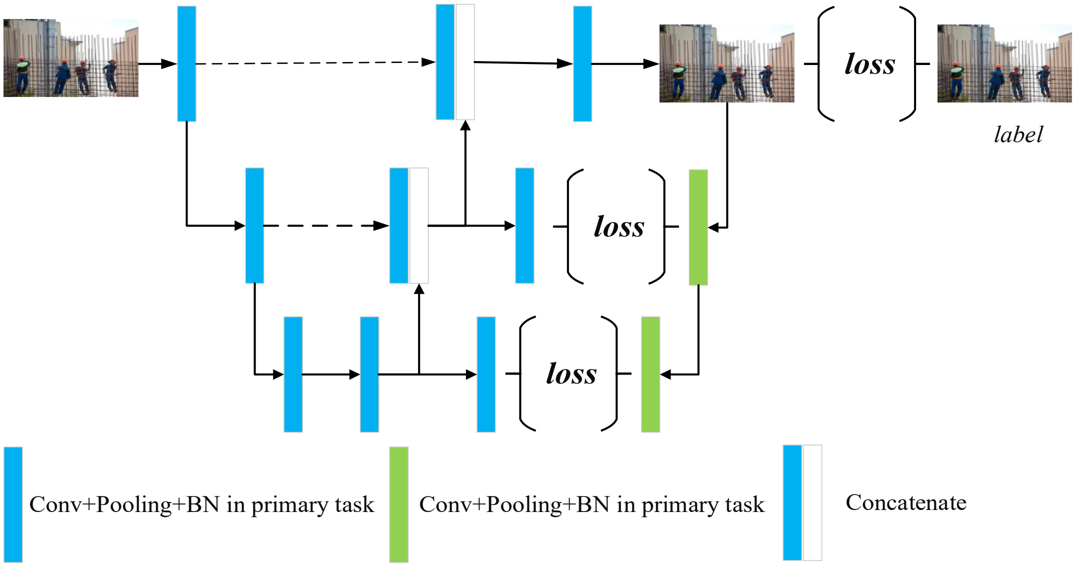

loss
label
loss
loss
Conv+Pooling+BN in primary task
Conv+Pooling+BN in primary task
Concatenate

Figure 3. The architecture details of the SR module.

We employ a simple U shape symmetric architecture to produce the primal task. The input low-resolution images are fed into feature extractor consisting of stacked convolution layers. Then, the abstracted features are mapped onto the target image space through trans-convolution. Alongside the easy-to-understand primal architecture, we attach a one-way path to map the resolution of image from high to low. Through improving the deep supervision of the mapping loop, we can achieve the advanced reconstruction performance.

Normally, the two regression tasks in SR are optimized following the classic l_1 loss, that is to say, the mean absolute error (MAE). However, taking into consideration that the main purpose of our SR module is to perform the preparation for the following object detection module, when comparing with pixel-wise similarity, we should pay more attention to the sematic information. Therefore, we employ a combination of l_1 loss and contextual loss as follows

$$l = \sum_{i=1}^{N} [l_1(P(x_i), y_i) + \alpha l_c(P(x_i), y_i) + l_1(D(P(x_i)), x_i)] \tag{1}$$

where (x_i, y_i) is the ith low-resolution and high-resolution image pair. $P(\cdot)$ and $D(\cdot)$ are the primal regression and dual regression, respectively. α is weighting coefficients of the contextual loss component. l_1 indicates the MAE and l_c denotes the contextual loss calculated from Equation (2).

$$l_c(x_i, y_i) = -\log(CX(F_P(x_i), F_D(y_i))) \tag{2}$$

where $F_P(\cdot)$ and $F_D(\cdot)$ are the feature maps obtained from the feature extractor during primal task and dual task, respectively. Based on our symmetric architecture shown in Figure 3, the two feature maps always have the same size N. In order to measure how similar the two feature maps are, we refer to the similarity in [31]. $CX(F_P(x_i), F_D(y_i)) = \frac{1}{N}\sum\limits_{j} \max\limits_{k} CX_{kj}$, where CX_{kj} calculates the similarity between the kth and the jth features from $F_P(\cdot)$ and $F_D(\cdot)$.

$$CX_{kj}(F_P(\cdot), F_D(\cdot)) = \frac{\exp\left(\frac{1-\widetilde{d_{kj}}}{h}\right)}{\sum_l \exp\left(\frac{1-\widetilde{d_{kl}}}{h}\right)} \tag{3}$$

where $\widetilde{d_{kj}} = d_{kj} / \left(\min\limits_{l} d_{kl} + \epsilon\right)$, and d_{kj} is the cosine distance between the kth and the jth features from $F_P(\cdot)$ and $F_D(\cdot)$. The parameters $h = 0.5$ and $\epsilon = 10^{-8}$. Due to the setting, the closer the two features, the smaller the d_{kj}. Consequently, the smoothed $\widetilde{d_{kj}}$ approaches 1 so as to produce large CX_{kj} as well as large $CX(F_P(x_i), F_D(y_i))$. Because the $F_P(\cdot)$ and $F_D(\cdot)$ are abstracted information obtained from backbone network, they are full of sematic information and could help the SR module focus more on high level similarity instead of pixel-wise alignment. This proved to be great help for our detection.

3.2. Detection Module

As one of the most advanced one-stage object detection models, YOLOv5 is chosen as our backbone model to detect helmets. Since there is a clear evolving track for YOLO models and the YOLOv5 is a combination of all the prior tricks and improvements, we do not recap the entire architecture in detail. Here, we only talk about how we use the model. Briefly, the three main components employed in YOLOv5 are backbone for feature extracting, neck for feature fusing and head for prediction. The cross stage partial network combined Darknet is used as backbone, which could abstract abundant information and the path aggregation network is utilized as neck to generate the feature pyramid so that it can enhance the capability of multi-scale detection. The head part follows the traditional YOLO head used in the prior version to obtain the prior box and classification result. In this paper, the detection module containing YOLOv5 model follows the SR module directly to implement the helmet detection. The loss function follows reference [8].

3.3. Dataset

We employ both public dataset as well as the realistic dataset to train our model. For SR task, the public data include the DIV2K and the Flickr2K, which contain 3550 paired images of high resolution, 2× and 4× low resolution [32,33]. There is no requirement with regard to the image content. The realistic dataset includes 5457 images randomly downloaded from the Internet. The chosen standard is that each image contains as least one person wearing helmet. We downsample these images through bicubic algorithm to generate 2× and 4× low-resolution images. All these paired images, 9007 in total, construct our training dataset. We obtain an individual test set, which includes 270 images via the same grabbing way of training set. The testing images are downsampled through randomly chosen methods available in the skimage toolbox of python. For helmet detection task, the aforementioned data excluding DIV2K and the Flickr2K are used as training and testing, respectively.

3.4. Metrics

The quantitative metrics employed in this paper are shown in Table 1. For SR task, we use the peak signal-to-noise ratio (PSNR) and the structural similarity (SSIM) values to measure the reconstructed image quality. The higher the value of PSNR is the better. The value of SSIM varies between 0 (worst) and 1 (best). For detection task, the precision, recall and average precision (AP) are calculated. Precision measures the capability that the model finds out targets which are real targets. Recall measures the capability that the model finds out real targets without missing. AP is calculated from the area under the precision–recall curve and values approaching 1 are the best.

Table 1. The definition of utilized quantitative metrics.

Task	Metrics	Definition
SR	PSNR	$\mathrm{PSNR} = 10\log_{10}\frac{255}{\frac{1}{3}\sum_{\mathrm{c}=\{\mathrm{R,G,B}\}}(\mathrm{MSE})_{\mathrm{c}}}$ where MSE indicates the mean square error of the image.
	SSIM	$\mathrm{SSIM}(I_0, I) = \frac{\left(2\mu_{I_0}\mu_I+c_1\right)\left(2\sigma_{I_0 I}+c_2\right)}{\left(\mu_{I_0}^2+\mu_I^2+c_1\right)\left(\sigma_{I_0}^2+\sigma_I^2+c_2\right)}$ where I_0 and I are the original and the reconstructed high-resolution images. μ and σ indicate mean and variance, respectively, and $c_1,\ c_2$ are constants
Detection	Precision	$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP+FP}}$ where $\mathrm{TP},\ \mathrm{FP}$ and FN indicate true positives, false positives and false negatives, respectively.
	Recall	$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP+FN}}$
	AP	Area under the precision–recall curve

3.5. Training

The entire method is implemented on the workstation equipped with two NVIDIA RTX3090 GPU and Intel i9 CPU. All coding work is based on Python 3.7 and PyTorch 1.8. The SR module and detection module are first trained separately from scratch. Then the two modules are finetuned together, alternately. Specifically, in each iteration, one module will be frozen while the other one is updating, then vice versa. The initial learning rate of the SR and detection are 0.0001 and 0.01, respectively. The Adam optimizer is applied with the momentum 0.9 and the batch size is 2.

4. Results

4.1. Performance of the Proposed SR Module

We compare the super-resolution reconstruction results of the proposed SR module with those of the other popular SR methods. According to the fact that our main purpose of this paper is helmet detection instead of pure super-resolution reconstruction, we choose the widely used interpolation methods for comparison since they are often utilized to adjust the network input resolution in object detection tasks. Our method starts from the DRN-S model designed for SR tasks so that it is necessary to compare our improvements with the original DRN-S model [30]. The achieved results are shown in Table 2. There is a gap between the PSNR values of SR-based methods and those of interpolation-based methods. Our SR module achieves a higher PSNR value while keeping its SSIM value consistent with that of DRN-S. This means that we can achieve better image quality so as to improve the accuracy of the coming detection module.

Table 2. The results achieved by different methods.

	Interpolation			**DRN-S**	**Our Method**
	Nearest Neighbor	**Bilinear**	**Bicubic**		
PSNR	23.716	25.277	25.343	27.964	27.991
SSIM	0.737	0.782	0.784	0.850	0.850

From Figure 4, we can review the super-resolution reconstructed images directly. Since our real targets are helmets, we focus on them more instead of on the background information. All helmet regions are zoomed in to visualize their details. It can be found that the helmets obtained from super-resolution reconstructed-based methods are clearer than those achieved by the interpolation-based method. To be specific, our SR module produces images that are less blurry but not so piecewise smooth to be able to obtain more median-frequency information.

Based on Table 2, there is a significant improvement using our method compared with the original DRN-S. To evaluate the effort of our newly added contextual loss, we compared the performance of the SR module with different weight α, as shown in Figure 5. DRN-S refers to the original DRN-S model. C-0.001 refers to the SR model trained under the combined loss described in Equation (1) with $\alpha = 0.001$. C-0.0005 and C-0.0001 indicate $\alpha = 0.0005$ and $\alpha = 0.0001$, respectively.

4.2. Performance of the Proposed Helmet Detection Method

We show the detection results produced by different end-to-end workflows in Table 3. Each workflow employs the same YOLOv5 architecture but different input. The Interpolation+YOLOv5 workflow is exactly the same as with normal YOLOv5 since it uses pure interpolation to resize the input images to the standard resolution. The DRN+YOLOv5 workflow is also super-resolution reconstruction driven detection, which consists of the DRN model and YOLOv5. The proposed SR module+YOLOv5 indicates the workflow described in this paper. From Table 3, the combination of the proposed SR module and YOLOv5 achieves the best precision, which means that 88.4% predicted helmets are real helmets. Compared with the other two results, it has the fewest false predictions. However, since precision is normally in contradiction with recall, the recall of the proposed SR module+YOLOv5 lags a little behind that of the DRN+YOLOv5. The AP of the proposed SR module+YOLOv5 is the highest, which indicates it has the best overall performance. It can be seen that the input images produced by the SR module will improve all metrics due to the improvement in image quality. The proposed workflow surpasses the original DRN+YOLOv5 workflow on precision and AP.

Table 3. Detection performance of different workflows.

	Interpolation+YOLOv5	**DRN+YOLOv5**	**Proposed SR Module+YOLOv5**
Precision	0.853	0.878	**0.884**
Recall	0.632	**0.716**	0.715
AP (%)	0.435	0.500	**0.501**

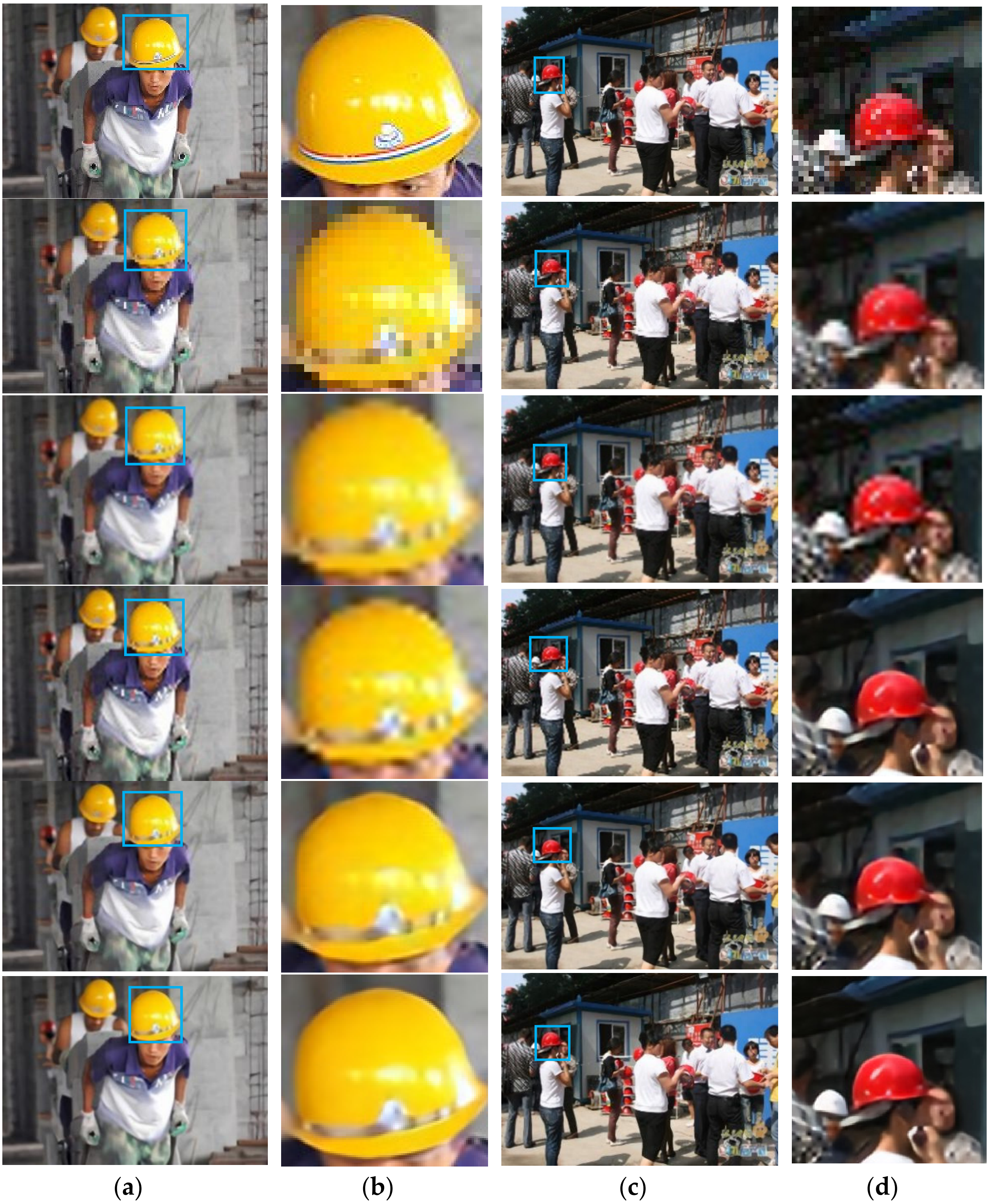


Figure 4. The comparison of the images produced by different SR method. From left to right, columns (**a**,**c**) indicate the complete images, columns (**b**,**d**) indicate the zoom-in regions defined by the blue bounding boxes. From the first row to the last row: original high-resolution image, images achieved via nearest neighbor interpolation, images achieved via bilinear interpolation, images achieved via bicubic interpolation, images achieved via DRN, images achieved via the proposed SR module.

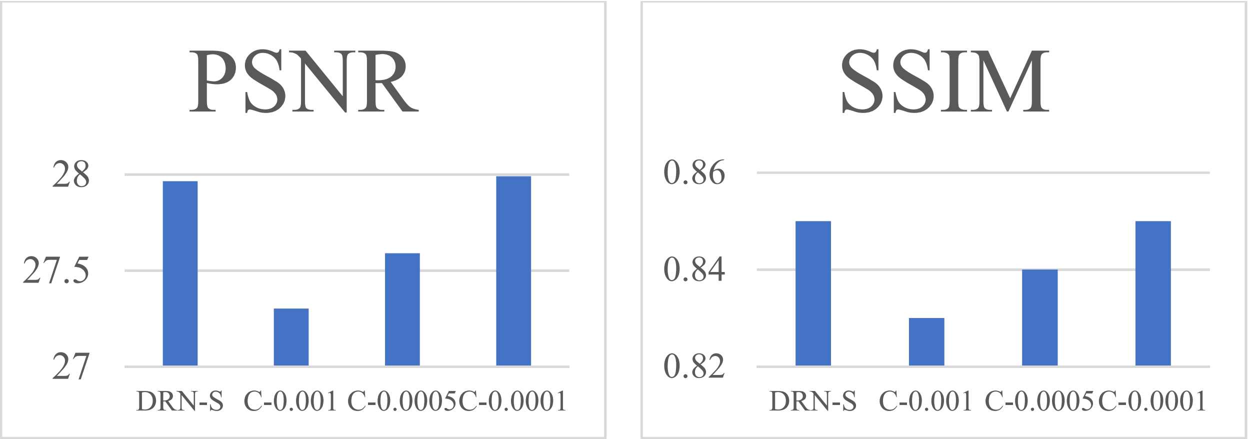


Figure 5. The PSNR and SSIM varied from the weighting of contextual loss.

5. Discussion and Conclusions

We propose an end-to-end helmet monitoring system, which implements a super-resolution reconstruction driven helmet detection workflow. It is designed for the scenario where the input image quality is limited, which is easily faced in engineering practice. For example, input images are acquired from a moving camera and transmitted through a bandwidth-limited wireless channel. Because of the limited bandwidth, images are always compressed and so have poor resolution and quality. The degrading of the input images will consequently decrease the helmet detection precision. To overcome this problem, the proposed SR driven helmet detection workflow consists of two sequential steps in the entire workflow. First, we use a super-resolution reconstruction module to improve the image resolution and quality instead of direct interpolation. Then, the processed images are fed into the detection module consisting of YOLOv5 to perform helmet detection. The two modules are trained separately from scratch and finetuned together, alternately. This is a typical multi-task learning strategy to help increase task specific accuracy by utilizing other tasks as constraints. Validation shows the effectiveness of our workflow. The comparison of the performance of different SR reconstruction methods shows that the proposed SR module could increase the PSNR value while maintaining a consistent SSIM value. The comparison of the performance of different detection workflows shows that the proposed SR module is effective at guiding the YOLOv5 and detection precision and AP are both increased. Generally speaking, based on current results, this will be a promising tool for helmet detection, which can be easily used in construction monitoring or traffic safety monitoring. Moreover, SR driven detection is a general workflow that is easy to be extended to other similar object detection tasks to solve the problem of performance degrading caused by poor inference input quality when the training input quality is good. Currently, our main idea is to use the individual model on specific tasks and combine tasks together. The model will be redundant if there are a large number of tasks. In the future, we will keep working on identifying a semantic subspace to attempt to remove the influence of image quality on detection performance.

Author Contributions: Conceptualization, Y.L. (Yicheng Liu) and Y.L. (Yan Liu); methodology, Y.L. (Yicheng Liu); software, Z.L.; validation, Y.L. (Yicheng Liu), Z.L. and Y.L. (Yan Liu); data curation, B.Z.; writing—original draft preparation, Z.L.; writing—review and editing, Y.L. (Yan Liu); visualization, B.Z.; supervision, J.H.; project administration, B.Z. and J.H.; funding acquisition, J.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Research on automatic and intelligent safety management technology at construction site, grant number CSCEC-2020Z-10 and Research on intelligent construction site management based on Internet of things and image recognition technology, grant number KJYF-2019-12.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The open access dataset DIV2K can be found in the website https://data.vision.ee.ethz.ch/cvl/DIV2K/ (accessed on 24 December 2021). The open access dataset Flickr2K can be found in the website https://yingqianwang.github.io/Flickr1024/ (accessed on 24 December 2021).

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

References

1. Kurien, M.; Kim, M.K.; Kopsida, M.; Brilakis, I. Real-time simulation of construction workers using combined human body and hand tracking for robotic construction worker system. *Autom. Constr.* **2017**, *86*, 125–137. [CrossRef]
2. Zhong, H.; Yanxiao, W. 448 cases of construction standard statistical characteristic analysis of inductrial injury accident. *Stand. China* **2017**, *2*, 245–247.
3. Viola, P.; Jones, M.J. Robust Real-Time Face Detection. *Int. J. Comput. Vis.* **2004**, *57*, 137–154. [CrossRef]
4. Felzenszwalb, P.F.; Mcallester, D.A.; Ramanan, D. A discriminatively trained, multiscale, deformable part model. In Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 23–28 June 2008.
5. Tang, T.; Zhou, S.; Deng, Z.; Zou, H.; Lei, L. Vehicle Detection in Aerial Images Based on Region Convolutional Neural Networks and Hard Negative Example Mining. *Sensors* **2017**, *17*, 336. [CrossRef] [PubMed]
6. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
7. He, K.; Zhang, X.; Ren, S.; Sun, J. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1904–1916. [CrossRef] [PubMed]
8. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788. [CrossRef]
9. Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger. In Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 6517–6525. [CrossRef]
10. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767.
11. Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-J.M. YOLOv4 Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934.
12. Kirkland, E.J. *Bilinear Interpolation*; Springer: Manhattan, NY, USA, 2010.
13. Liu, Y. An Improved Feedback Network Superresolution on Camera Lens Images for Blind Superresolution. *J. Electr. Comput. Eng.* **2021**, *2021*, 5583620. [CrossRef]
14. Chen, Y.; Liu, L.; Phonevilay, V.; Gu, K.; Xia, R.; Xie, J.; Zhang, Q.; Yang, K. Image super-resolution reconstruction based on feature map attention mechanism. *Appl. Intell.* **2021**, *51*, 4367–4380. [CrossRef]
15. Dong, C.; Loy, C.C.; He, K.; Tang, X. Learning a Deep Convolutional Network for Image Super-Resolution. In *Computer Vision—ECCV 2014*; Springer: Cham, Switzerland, 2014; Volume 8692, pp. 184–199.
16. Kim, J.; Lee, J.K.; Lee, K.M. Accurate Image Super-Resolution Using Very Deep Convolutional Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 1646–1654.
17. Li, J.; Fang, F.; Mei, K.; Zhang, G. Multi-scale Residual Network for Image Super-Resolution. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; Volume 11212, pp. 527–542.
18. Zhang, Y.; Li, K.; Li, K.; Wang, L.; Zhong, B.; Fu, Y. Image Super-Resolution Using Very Deep Residual Channel Attention Networks. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; Volume 11211, pp. 294–310.
19. López-Tapia, S.; Lucas, A.; Molina, R.; Katsaggelos, A.K. A single video super-resolution GAN for multiple downsampling operators based on pseudo-inverse image formation models. *Digital Signal Process.* **2020**, *104*, 102801. [CrossRef]
20. Majdabadi, M.M.; Ko, S.B. Capsule GAN for robust face super resolution. *Multimedia Tools Appl.* **2020**, *79*, 31205–31218. [CrossRef]
21. Bulat, A.; Yang, J.; Tzimiropoulos, G. To Learn Image Super-Resolution, Use a GAN to Learn How to Do Image Degradation First. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; Volume 11210, pp. 187–202.
22. Badaghei, R.; Hassanpour, H.; Askari, T. Detection of Bikers without Helmet Using Image Texture and Shape Analysis. *Int. J. Eng.* **2021**, *34*, 650–655. [CrossRef]
23. E Silva, R.R.V.; Aires, K.R.T.; Veras, R. Detection of helmets on motorcyclists. *Multimed. Tools Appl.* **2018**, *77*, 5659–5683. [CrossRef]
24. Sun, X.; Xu, K.; Wang, S.; Wu, C.; Zhang, W.; Wu, H. Detection and Tracking of Safety Helmet in factory environment. *Meas. Sci. Technol.* **2021**, *32*, 105406. [CrossRef]

25. Lin, H.; Deng, J.D.; Albers, D.; Siebert, F.W. Helmet Use Detection of Tracked Motorcycles Using CNN-Based Multi-Task Learning. *IEEE Access* **2020**, *8*, 162073–162084. [CrossRef]
26. Yogameena, B.; Menaka, K.; Perumaal, S.S. Deep learning-based helmet wear analysis of a motorcycle rider for intelligent surveillance system. *IET Intell. Transp. Syst.* **2019**, *13*, 1190–1198. [CrossRef]
27. Gu, Y.; Xu, S.; Wang, Y.; Shi, L. An Advanced Deep Learning Approach for Safety Helmet Wearing Detection. In Proceedings of the 2019 International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), Atlanta, GA, USA, 14–17 July 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 669–674.
28. Xu, K.; Deng, C. Research on Helmet Wear Identification Based on Improved YOLOv(3). *Laser Optoelectron. Progr.* **2021**, *58*, 0615002.
29. Xiao, T. Improved YOLOv3 Helmet Wearing Detection Method. *Comput. Eng. Appl.* **2021**, *57*, 216–223.
30. Guo, Y.; Chen, J.; Wang, J.; Chen, Q.; Cao, J.; Deng, Z.; Xu, Y.; Tan, M. Closed-Loop Matters: Dual Regression Networks for Single Image Super-Resolution. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 14–19 June 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 5406–5415.
31. Mechrez, R.; Talmi, I.; Zelnik-Manor, L. The Contextual Loss for Image Transformation with Non-aligned Data. In Proceedings of the Computer Vision—ECCV 2018, Munich, Germany, 8–14 September 2018; Springer: Berlin/Heidelberg, Germany, 2018; pp. 800–815.
32. Agustsson, E.; Timofte, R. NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 21–26 July 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 1122–1131.
33. Lim, B.; Son, S.; Kim, H.; Nah, S.; Lee, K.M. Enhanced Deep Residual Networks for Single Image Super-Resolution. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 21–26 July 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 1132–1140.