Alexandria Engineering Journal (2023) 71, 161–172

HOSTED BY

ELSEVIER

Alexandria University

Alexandria Engineering Journal

www.elsevier.com/locate/aej
www.sciencedirect.com

ORIGINAL ARTICLE

Machine vision based damage detection for conveyor belt safety using Fusion knowledge distillation

Xiaoqiang Guo [a], Xinhua Liu [a,*], Paolo Gardoni [b], Adam Glowacz [c], Grzegorz Królczyk [d], Atilla Incecik [e], Zhixiong Li [d,*]

[a] School of Mechatronic Engineering, China University of Mining and Technology, Xuzhou 211006, China
[b] Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Champaign, IL 61820, United States of America
[c] Department of Electrical Engineering, Cracow University of Technology, Cracow 31-155, Poland
[d] Department of Manufacturing Engineering and Automation Products, Opole University of Technology, Poland
[e] Department of Naval Architecture, Ocean, and Marine Engineering, University of Strathclyde, Glasgow G11XQ, United Kingdom

Received 4 November 2022; revised 14 February 2023; accepted 7 March 2023

KEYWORDS

Reliability and Risk;
CenterNet;
Deep Learning;
Belt tear detection

Abstract A belt conveyor system is one of the essential equipment in coal mining. The damages to conveyor belts are hazardous because they would affect the stable operation of a belt conveyor system whilst impairing the coal mining efficiency. To address these problems, a novel conveyor belt damage detection method based on CenterNet is proposed in this paper. The fusion of feature-wise and response-wise knowledge distillation is proposed, which balances the performance and size of the proposed deep neural network. The Fused Channel-Spatial Attention is proposed to compress the latent feature maps efficiently, and the Kullback-Leibler divergence is introduced to minimize the distribution distance between student and teacher networks. Experimental results show that the proposed lightweight object detection model reaches 92.53% mAP and 65.8 FPS. The proposed belt damage detection system can detect conveyor belt damages efficiently and accurately, which indicates its high potential to deploy on end devices.

1. Introduction

A belt conveyor system is one of the most widely used and versatile equipment to transport bulk material, and they play an important role in carrying coals in the mining industry [1,2]. As the carrying medium of a belt conveyor system, the conveyor belt is vulnerable to damages caused by the metal foreign

* Corresponding authors at: School of Mechatronic Engineering, China University of Mining and Technology, Xuzhou 211006, China (X. Liu)
E-mail addresses: liuxinhua@cumt.edu.cn (X. Liu), zhixiong.li@yonsei.ac.kr (Z. Li).

https://doi.org/10.1016/j.aej.2023.03.034
1110-0168

Nomenclature

FCSA	Fused Channel-Spatial Attention
KL divergence	Kullback-Leibler divergence
FKD	Fusion Knowledge Distillation
C_{avg}	Weights of each channel
$dspan$	Span of two adjacent spatial pieces
Idx	Index list of all divided spatial pieces
s_i	Sum of each piece i
W_A	Weight list of channel attention
Spa_i	Spatial attention for each piece i
$\widehat{\mathrm{H}}_{gt}$	Modified probability
$\widehat{\mathrm{H}}_{xyc}$	Heatmap of class c in location of (x, y)
ω_i	Balanced weight for each class i
$\mathscr{L}_{feat}$	Feature-wise knowledge distillation loss
$\mathscr{L}_{resp}$	Response-wise knowledge distillation loss
$\mathscr{L}_{hw}$	Class balanced heatmap loss
$\mathscr{L}_{off}$	Object center offset loss
$\mathscr{L}_{total}$	Total loss of the improved CenterNet model
FPS	Frames Per Second
mAP	Mean Average Precision
LME	Logarithmic Model Efficiency
IPC	Industrial PC
CPU	Central Processing Unit
GPU	Graphics Processing Unit
PLC	Programmable Logic Controller

matters mixed in coals or the impact load of bulk material. Such damages to the conveyor belt may lead to the whole belt conveyor system working at a low efficiency or even be totally dysfunctional [3]. Hence, a stable and reliable conveyor belt damage detection system is urgently needed to detect and monitor the state of the conveyor belt so as to ensure the stable operation of the belt conveyor system [4].

The mainstream technologies of detecting conveyor belt damage are vision-based machine learning methods. Machine vision-based object detection methods have been studied and developed for many years with a complete theoretical system has formed, e.g., edge detection, feature extraction, semantic segmentation, Fourier transform and so on. Traditional machine vision-based methods depend on engineer experience and environmental conditions, which limits its usability and application. Traditional vision-based damage detection methods [5–9] consist of image acquisition, preprocessing, objection detection and postprocessing. To address these problems, auxiliary equipment is deployed. Considering the strong penetration characteristic of infrared light, one or more hyperspectral cameras were utilized in [10–13] to realize multi-images fusion and improve the algorithm performance of conveyor belt damage detection. Laser-assisted conveyor belt damage detection methods [14–16] that can convert indistinguishable damage features of the belt into distinct laser lines were also utilized. Besides, audio-visual combined detection methods based on sound signals and images of the conveyor belt were investigated in [17,18]. The fused audio-visual detection can identify different states of the conveyor belt.

With the advance of chips and algorithms, artificial neural network is widely used in basic science [19–22] and engineering applications [23–25]. Compared with traditional vision-based object detection methods, the deep learning-based object detection methods provide the advantages of excellent generalization and fast detection speed. The first network architecture that successfully applied the deep learning to object detection is R-CNNs. Then, one-stage object detection models, e.g., YOLO, SSD and CenterNet, are proposed to improve the detection accuracy. Besides the complicated detection procedures, two-stage methods, e.g. R-CNNs [26], have the advantage of excellent performance. However, their disadvantages are also obvious, such as complicated detection processes, slow detection speed and relatively big models. To overcome the above disadvantages, one-stage models which regard object detection as regression have been proposed and drawn much attention. The well-known one-stage object detection models are YOLO [27], SSD [28] and CenterNet [29] series. They have the characteristics of shared parameters, simplified models, fast detection speed and relatively poor performance than that of two-stage methods. Considering the velocity of belt (1.5 m/s or higher), the detection efficiency and compute capability of end devices, the belt damage detection model must balance the accuracy and reference speed. Either two-stage or one-stage models have suitable application fields; however, the one-stage models are more appropriate for conveyor belt damage detection, which needs to meet the requirements of real-time and stable detection.

As one of the representative one-stage anchor free object detection algorithms, CenterNet mainly consists of three modules, i.e., backbone, decoder and head. The backbone includes multiple convolutions and activation layers, and it is responsible for latent feature extractions. The latent feature map extracted from backbone is decoded to the form of prediction required, which is fed into the head part to output final predictions. CenterNet can simultaneously predict object classes and regresses object size and location, making the inference process fast and efficient. Researchers have proposed many novel object detection methods based on CenterNet. For instance, Guo et al. [30] put forward a one-stage object detection method based on CenterNet to classify and locate different ships in SSR images. The feature refinement module and feature pyramid fusion module were introduced to address the small object detection problem. A novel defect detection method based on CenterNet was revealed in [31] to inspect any generated defects during additive manufacturing. The density map branch of head output and loss count was introduced to realize comprehensive and accurate prediction. Dai et al. [32] proposed a CenterNet-based power line surveillance method to inspect specific regions and avoid accidents. The cascade guiding structure and improved loss function were introduced to improve the detection performance.

In this paper, a novel feature-wise knowledge distillation (KD) based on attention mechanism is introduced, and an improved lightweight CenterNet object detection algorithm is proposed. The main contributions of this paper are detailed as follows:

(1) Fusion Knowledge Distillation: a novel knowledge distillation algorithm combined with feature-wise and response-wise knowledge distillation is proposed in this paper. The feature-wise knowledge distillation is implemented based on the adaptive fusion of channel attention and spatial attention for latent feature maps, and the response-wise knowledge distillation aims at transferring knowledge of class heatmaps.
(2) Improved focal loss: to address the class-imbalance problem, an improved focal loss function is carried out. The improved focal loss function can weight hard samples bigger value than easy samples adaptively and adjust the proportion of different class samples in the loss function.
(3) Lightweight deep learning model: based on the proposed Fusion Knowledge Distillation and improved focal loss, a lightweight belt damage detection algorithm based on CenterNet is developed. The lightweight deep learning model can balance the model size and performance by adopting feature-fused knowledge distillation.

The rest paper is organized as follows. Section 2 introduces the improved conveyor belt damage detection method based on Fused Channel-Spatial Attention (FCSA) and Knowledge Distillation (KD). Experiments and corresponding results are presented in Section 3. Finally, Section 4 demonstrates the application of the detection system, and Section 5 concludes this paper and discusses future works.

2. The proposed conveyor belt damage detection method

2.1. The architecture of the conveyor belt damage detection method

The proposed novel damage detection method for conveyor belts is based on CenterNet and Fusion Knowledge Distillation (FKD). CenterNet is one of the anchor-free one-stage object detection deep neural networks. The outputs of CenterNet consist of class heatmaps, height and width regression and offsets of centers. Knowledge Distillation is inspired by the learning process of human beings. Fig. 1 shows the network architecture of the proposed conveyor belt damage detection algorithm, which mainly consists of a teacher network and a student network. The implementation of training for the deep neural network is usually the interactive process between the teacher network and the student network. The teacher network is a heavyweight, pretrained deep learning network, and it performs well in inferencing. Meanwhile, a lightweight deep learning network plays as a student learns from the teacher network in the training procedure. The teacher-student architecture plays an essential role in the knowledge distillation. In this paper, the ResNet50 is introduced as the backbone of the teacher network, and the ResNet18 is selected as the backbone of the student network. The feature-wise knowledge distillation is applied by optimizing the distance of latent feature maps from the student network to the teacher network. Furthermore, the response-wise knowledge distillation optimizes the distance of heatmaps between student and teacher networks.

CenterNet regards object detection as regression and has excellent inference performance of precision and speed. Both of teacher and student networks of the proposed network are based on CenterNet. The teacher network adopts ResNet50 as the backbone and is well-trained in advance. The ResNet18 is introduced as the backbone of the student network, which is trained with the strategy of FKD and FCSA. To efficiently train the lightweight student network, knowledge distillation is applied in latent feature maps and head-outputs. The backbone network contains a large number of high-level latent features, which are extremely useful to object detection. Therefore, the feature-wise Knowledge Distillation can compress essential information in latent feature maps of the teacher network and transfer it to the student network. The head of CenterNet consists of class heatmaps, height-width regression and center offsets, which are the model prediction and critical. Hence, the response-wise Knowledge Distillation is introduced to adjust the student model by the teacher model.

The backbone of ResNet18 is more compact than of ResNet50, but the compression of model is limited. Hence, other classical feature pyramid networks are introduced as

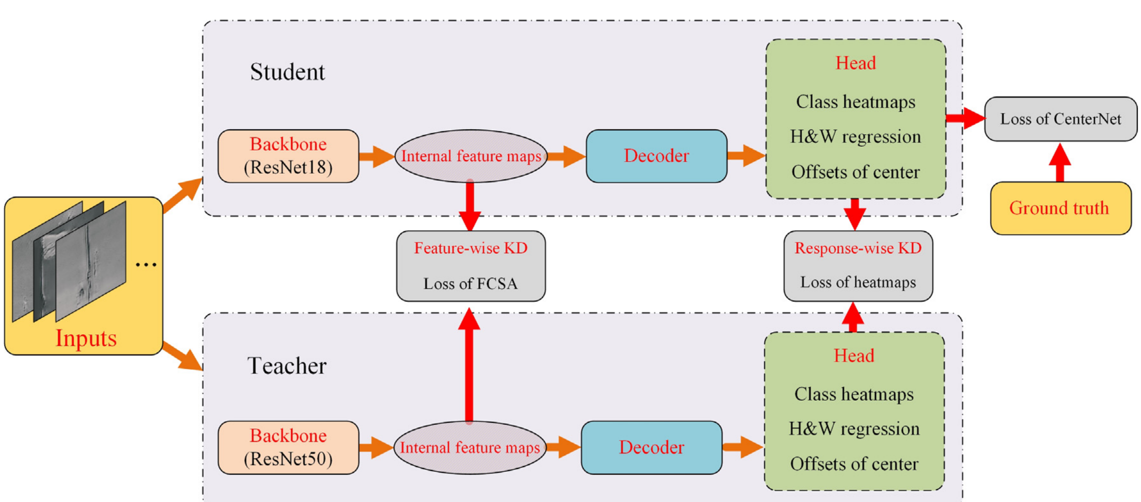

Fig. 1 The architecture of the proposed conveyor belt detection network. During the training process, input images are fed into Teacher and Student models, respectively. Teacher model provides soft labels for Student model.

the backbone of the student model and the performance is quite satisfactory. The experimental results are discussed in Section 3.

2.2. Fusion knowledge distillation

To achieve well-trained lightweight student model, the proposed knowledge distillation consists of feature-wise and response-wise Knowledge Distillation [33].

2.2.1. Feature-wise knowledge distillation

The proposed conveyor belt damage detection algorithm is based on an anchor-free CenterNet, which consists of the backbone network, decoder and head parts. The backbone network is responsible for feature extraction from input images and it contains high-dimensional features with large parameters. If the knowledge distillation is applied without feature compression, the student network training process would be high-cost and formidable. Hence, a feature-wise knowledge distillation based on the proposed Fused Channel-Spatial Attention (FCSA) is introduced. The schematic of the proposed FCSA is shown in Fig. 2.

(1) The channel attention is applied to obtain the segmentation coefficients for latent feature maps. Given the tensor of the backbone network $\boldsymbol{T}_b \in \mathbb{R}^{C\times H\times W}$ as the input, the output of the 2D average pooling along $H \times W$ dimensions, C_{avg}, can be described as:

$$C_{avg} = AvgPool(\boldsymbol{T}_b) \quad (1)$$

where $AvgPool$ denotes the operation of 2D average pooling, C, H and W are channel, height and width of latent feature maps, respectively. The C_{avg} is used to divide the latent feature maps of the backbone output into N pieces (N equals 10 in this paper). The span of two adjacent pieces is obtained by $dspan = C/N$. The C_{avg} is treated as the channel index and sorted in descending order. The top N weakest response points in C_{avg} is treated as the indexes for spatial attention, and the distance for each two divided indexes must be larger than d_{span}. All N divided indexes which satisfy the above requirements can be represented as an index set $Idx = \{0, c_1, c_2, ..., c_n, c_{\max}\}$, and c_{max} is the number of channels.

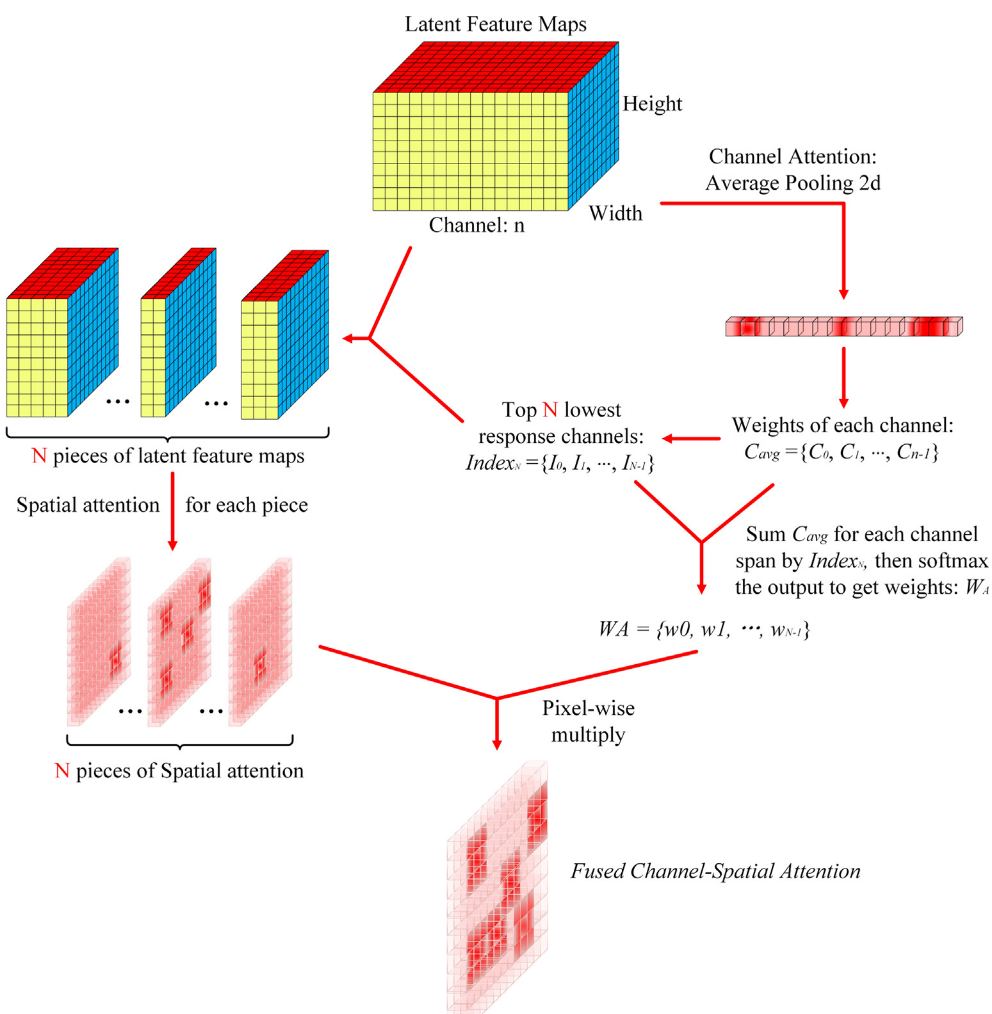

Fig. 2 The schematic of the Fused Channel-Spatial Attention. Latent feature maps are extracted along two paths: Channel and Spatial. Then two-path features are combined into Fused Channel-Spatial Attention.

(2) The weights for spatial attention represent the distribution of features along channel-wise. The channel attention C_{avg} has been divided into N pieces by the index set Idx sorted in descending order. The sum of each piece, s_i, can be obtained as follow:

$$s_i = \sum_{j=Idx_i}^{Idx_{i+1}} C_{avg}(j) \tag{2}$$

where Idx_i denotes the i-th member in Idx, and the set $sum_A = \{s_0, s_1, ..., s_n\}$. The N-weight of channel attention can be described as:

$$W_A = \mathrm{softmax}(sum_A) = \{\omega_0, \omega_1, ..., \omega_n\} \tag{3}$$

(3) For certain pieces feature maps in N pieces, F_i, the spatial attention is applied to compress the feature maps, which can be described as:

$$Spa_i = SpaAtt(F_i), i \in [0, N] \tag{4}$$

where $SpaAtt$ denotes the operation of the spatial channel, and the $SpaAtt$ process is shown in Fig. 3. It is seen that the $SpaAtt$ consists of average-pooling and max-pooling along the channel dimension. The outputs of two pooling layers are fed into a convolutional layer to integrate information. The fused attention for spatial and channel, $FCSA$, can be described as:

$$FCSA = \sum_i \omega_i \cdot Spa_i, i \in [0, N] \tag{5}$$

where $w_i \in W_A$, and the fused attention contains the distribution of spatial and channel information.

(4) The feature-wise knowledge distillation is applied by minimizing the Kullback-Leibler divergence of the latent feature maps from student network to teacher network. The loss function of feature-wise KD can be described as:

$$\mathscr{L}_{feat} = \frac{KL(FCSA_{student}||FCSA_{teacher})}{h' * w'} \tag{6}$$

where the h' and w' denote the height and width of the FCSA output, respectively. The latent feature maps are compressed by FCSA, which can guide the student network to focus on the essential information in latent feature maps and greatly reduce computation.

2.2.2. Response-wise knowledge distillation

CenterNet is one-stage anchor-free object detection model, whose output contains the prediction of classes and the objects locations. The response-wise knowledge distillation is applied by minimizing the Kullback-Leibler divergence:

$$\mathscr{L}_{resp} = \frac{\sum_{i=1}^{H' * W'} KL(H_i^s||H_i^t)}{H' * W'} \tag{7}$$

Where KL denotes the Kullback-Leibler divergence, H' and W' are the height and width of heatmaps, respectively. H_i^s and H_i^t are the pixel i response of heatmaps in student and teacher networks, respectively. Through the response-wise knowledge distillation, the student network can directly learn the class information from the teacher network.

2.3. The loss function of the proposed conveyor belt damage detection network

The loss function of CenterNet consists of class heatmap loss, height-width regression loss and center point offset loss. The class heatmap loss is based on focal loss and focuses on the hard samples which would contribute to back-propagation more than easy samples. The class heatmap loss in the original CenterNet can be described as:

$$\mathscr{L}_{hm} = -\frac{1}{N} \sum \alpha_t \left(1 - \widehat{\mathrm{H}}_{gt}\right)^{\gamma} \log \widehat{\mathrm{H}}_{gt} \tag{8}$$

where α_t and γ are hyperparameters useful to data-balance, and N is the number of keypoints in an image. And $\widehat{\mathrm{H}}_{gt}$ is the modified probability:

$$\widehat{\mathrm{H}}_{gt} = \begin{cases} \widehat{\mathrm{H}}_{xyc}, if y = 1 \\ 1 - \widehat{\mathrm{H}}_{xyc}, others \end{cases} \tag{9}$$

where $\widehat{\mathrm{H}}_{xyc}$ is the heatmap of class c in location of (x, y), which is represented in 2D Gaussian kernel:

$$\mathrm{H}_{xyc} = \exp\left(-\frac{(x - p_x)^2 + (y - p_y)^2}{2\sigma_p^2}\right) \tag{10}$$

where σ_p is an object size-adaptive standard deviation.

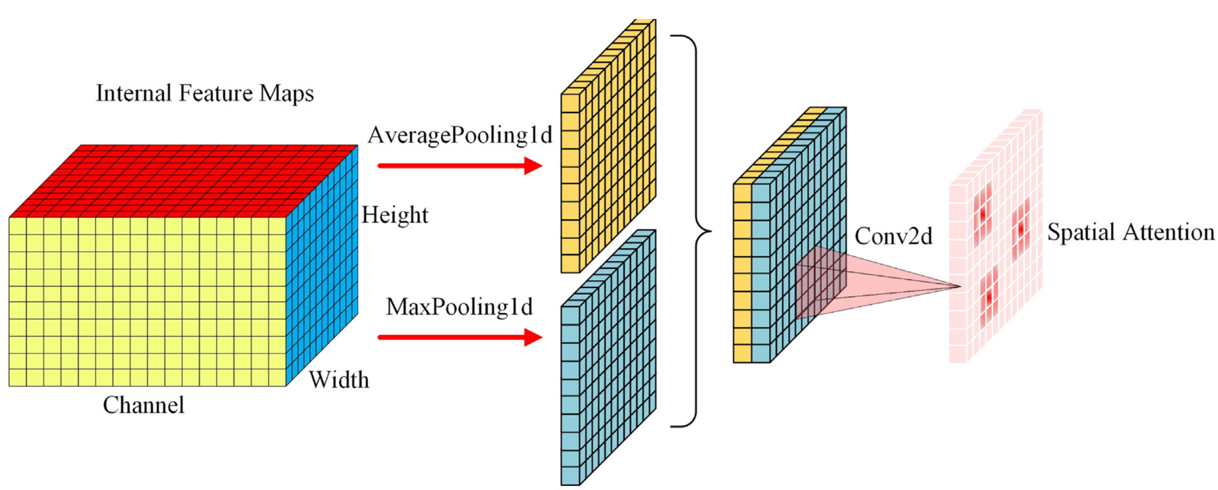

Fig. 3 The process of the spatial attention.

However, each type of conveyor belt damage has different probabilities of occurrence, and a certain type of belt damage is critical but scarce. Hence the experimental dataset for the conveyor belt contains three classes and has the typical characteristic of data-imbalance: the samples of scratch are much more than that of tear and crack. The imbalance dataset has the characteristic of long-tailed data distribution (few classes represent most data and most classes are not represented). The strategy of class-balance is introduced to address the class-imbalance problem. The class-balance weight is described as:

$$\omega_i = \frac{1}{E_{n_y}} = \frac{1-\beta}{1-\beta^{n_y}} \tag{11}$$

where β is hyperparameter and E_{n_y} denotes the effective number. n_y is the number of samples in ground-truth class y. The original focal loss contains a parameter of α_t, which serves as the class-balance weight. The weight w_i is introduced to replace α_t and solve the data-imbalance problem. So, the improved class heatmap loss with class-balance weight w_i can be described as:

$$\mathscr{L}_{hm} = -\frac{1}{N}\sum\nolimits_{xyc}\begin{cases} \omega_i\left(1-\widehat{\mathrm{H}}_{xyc}\right)^{\gamma}\log\widehat{\mathrm{H}}_{xyc}, if y=1 \\ \left(1-\widehat{\mathrm{H}}_{xyc}\right)^{\beta}\left(\widehat{\mathrm{H}}_{xyc}\right)^{\gamma}\log\left(1-\widehat{\mathrm{H}}_{xyc}\right), others \end{cases} \tag{12}$$

where β and γ are hyperparameters. The introduced weight, ω_i, can balance the loss of different classes for positive samples. The term of $\left(1-\widehat{\mathrm{H}}_{xyc}\right)^{\beta}$ weights the negative samples small value, which makes the model focus on the positive samples.

CenterNet does not contain the prior knowledge of anchor boxes and the size of objects are acquired by minimizing the height-width regression loss:

$$\mathscr{L}_{hw} = \frac{1}{N}\sum\nolimits_{i=1}^{N}\left|\widehat{S}_{p_k} - s_k\right| \tag{13}$$

where $\widehat{S}_{p_k}$ and s_k are predictions and ground-truth of each object k, respectively. Since the predicted object centers are offset to ground-truth. The offset loss is described as:

$$\mathscr{L}_{off} = \frac{1}{N}\sum_{p}\left|\widehat{O}_{\tilde{p}} - \left(\frac{p}{R} - \tilde{p}\right)\right| \tag{14}$$

The FKD is proposed to compress and transfer features from teacher to student. The total loss of the proposed conveyor belt damage detection network consists of object detection loss and FKD loss and it has:

$$\mathscr{L}_{total} = \mathscr{L}_{det} + \lambda_{feat}\mathscr{L}_{feat} + \lambda_{resp}\mathscr{L}_{resp} \tag{15}$$

$$\mathscr{L}_{det} = \mathscr{L}_{hm} + \mathscr{L}_{hw} + \mathscr{L}_{off} \tag{16}$$

where λ_{feat} and λ_{resp} are hyperparameters to control the proportion of $\mathscr{L}_{feat}$ and λ_{resp} in the total loss.

In order to present the proposed conveyor belt algorithm thoroughly, the pseudocode of the training process is described in Table 1.

Table 1 The pseudocode of the training process for the proposed conveyor belt damage model.

Algorithm 1 The training process for the proposed conveyor belt damage model

1: **Input:** hyperparameters (batch size b, epochs e_t and e_s, λ_{feat} and λ_{resp}), location of dataset A
2: Establish and initialize models: M_t and M_s, setup optimizer: Adam, Load training dataset A, samples $a \in A$
3: **Train Teacher** M_t**:**
  For *epoch* $= 1$ **to** e_t **do**
    Get $hm, h\&w, offsets = M_t(a)$, compute $\mathscr{L}_{det}$ based on eq. (16), then update parameters of model M_t
  end for
4: **Train Student** M_s**:**
  For *epoch* $= 1$ **to** e_s **do**
    Get $hm_t, h_t\&w_t, offsets_t = M_t(a)$, get $hm_s, h_s\&w_s, offsets_s = M_s(a)$, then compute $\mathscr{L}_{feat}$ based on eq. (6) and $\mathscr{L}_{resp}$ based on eq. (7), then compute $\mathscr{L}_{total}$ based on eq. (15), then update parameters of model M_s
  end for

3. Experiments and results

3.1. The implementation of the conveyor belt damage detection system

A conveyor belt damage detection system has been built to detect the damage on the surface of conveyor belt in real-time, and its schematic is shown in Fig. 4. The hardwire of the damage detection system comprises an image acquisition module, a transmission module, a data processing and an execution module. The image acquisition module consists of linear array industrial cameras, light sources and the corresponding controller. During the operation of this system, the acquired belt images are firstly transmitted to the data processing module through the transmission module, which consists of a gigabyte industrial switch and gigabyte ethernet cables. Then, an industrial PC (IPC) with high-performance graphics processing unit (GPU) processes the belt images and predicts the results of damage types, size and location. Finally, after receiving signals from a center server, the PLC in the execution module can raise the alarm or shut down the conveyor if specific conveyor belt damages are detected. Connecting one or more sets of acquisition, transmission, and data processing modules to the execution module, the proposed conveyor belt damage detection system can inspect multiple nodes of the conveyor belt simultaneously.

The training procedure for the improved conveyor belt damage detection model is achieved by using a data processing server, which is equipped with an intel i9-10900kf CPU, a 32G RAM and two RTX2080s graphics cards. The detailed description of data processing server is listed in Table 2. To speed up the training process, the proposed FKD-CenterNet model was trained on the high-performance server with two NVIDIA GPUs. The proposed model is established by a popular deep learning framework, PyTorch, which is based on the Python programming language. Its primary data type is single-precision floating point. It can be seen from Table 2, for single-precision floating point arithmetic, the performance of GPU is 66 times higher than of CPU. Hence, the data server with two GPUs used in this paper can provide powerful computing support. Three types of damage images, i.e., scratch, crack and tear, of the conveyor belt in the train and evaluation dataset were obtained from a laboratory simulation environment as shown in Fig. 5.

Considering memory size of two GPUs and model performance, the resolution of belt damage images, 416 × 416, is

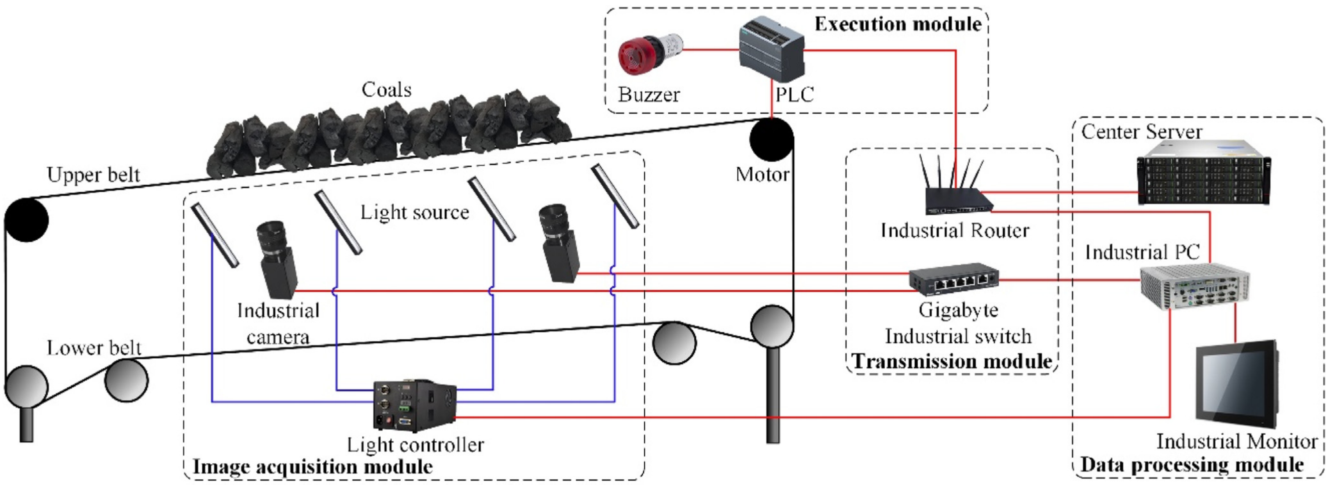

Fig. 4 The hardware framework of the conveyor belt damage detection system. It contains four modules: image acquisition module, transmission module, data processing module and execution module.

Table 2 The detailed description of data processing server.

No.	Device name	Type	Specification
1	CPU	Intel I9-10900kf	Frequency: 3.7 GHz Turbo Clock: 5.3 GHz TDP: 125 W FP32 (float): 0.169 TFLOPS
2	RAM	Kingston 16G DDR4 * 4	Dual channel, 4 slots 2666 MHz
3	Hard driver	TOSHIBA 512G SSD	Interface: NVMe Speed: 3100 M/s
4	GPU	NVIDIA RTX 2080 s * 2	Memory Size: 8 GB Bandwidth: 495.9 GB/s FP32 (float): 11.15 TFLOPS

appropriate for training the proposed FKD-CenterNet model based on RTX2080s. During the training process, the custom conveyor belt damage dataset has been augmented by horizontal and vertical flipping, 90% random cutting and $\pm 3^{\circ}$ rotation. The detailed description of the conveyor belt damage dataset is listed in Table 3. The samples of damaged belt images were acquired in the laboratory simulation environment and labeled carefully. Even though the data augmentation and cleaning were conducted, the data-imbalance of different classes of the conveyor belt damage dataset is still inevitable. To train the conveyor belt damage detection model well, the ratio of train, validation and test is set to 7:2:1. Each image may contain one or more detection objects. In order to balance the samples in the dataset, each number of classes n_y is set to 767, 548 and 193, respectively.

The proposed algorithm consists of two stages: teacher training and student training, and Adam optimizer with momentum of $\beta_1 = 0.93$, $\beta_2 = 0.999$ is applied in both training stages. The batch size is set as 6 and initial learning rate is 1×10^{-4} with decay rate 2.5×10^{-5} per 8 k iterations.

The optimizer plays an essential role to perform parameters update for a deep learning model. Hence, several different optimizers, such as SGD, Momentum, AdaGrad, RMSprop and Adam, have been tested. SGD (Stochastic Gradient Descent) is a basic optimizer, which tries to find the minimum in a random way. Momentum is a method that tries to accelerate optimization process and suppress oscillations in related directions. AdaGrad adapts the learning rate to the parameters and performs smaller updates for parameters. RMSprop optimizer is similar to Momentum and tries to dampen oscillations. Adam is a method that introduces adaptive policy for learning rate. The training results (Student model) of above optimizers are listed in Table 4. It can be seen that Adam optimizer per-

Fig. 5 The belt conveyor in the laboratory simulation environment.

Table 3 The detailed description of the conveyor belt damage dataset.

	Capacity	Resolution	Scratch	Crack	Tear	Augmented capacity
Train	874	–	511	364	143	–
Validation	250	–	155	121	36	–
Test	125	–	101	63	14	–
Total	1249	416 × 416	767	548	193	3498

forms excellent even though it takes more training time. Hence, Adam is selected as training optimizer.3.2 The experimental results and comparisons.

The well-trained teacher model is necessary for training the lightweight student network. The teacher model consists of the ResNet50 backbone, decoder and head module. The decoder parses the latent feature maps from the backbone to three branches, i.e., heatmaps, object sizes and center offsets, which are used to predict objects in head module comprehensively. The proposed FKD served as soft labels and guided the student model in the training process. After the training process for the improved CenterNet model, 345 (10%) testing images in the augmented dataset are verified. The average precisions of teacher model with backbone ResNet50 and student model with backbone ResNet18 are 94.55% and 92.53%, respectively. In the meanwhile, the recalls are 96.34% and 93.85%, respectively. The above results are obtained with 45.4 minimum fps, and Intersection over Union (IoU) has reached 86.4%. The prediction results of the teacher model are exhibited in Fig. 6. The score threshold is set as 0.5, and the most of belt damages can be detected. Although samples with scratch containing multiple small features are the typical hard ones in this dataset, it can be labeled out accurately.

The comparison of teacher (ResNet50) and student (ResNet18) models is shown in Fig. 7. During the training of the teacher model, the backbone of ResNet50 is frozen in the first 45 epochs to make the best use of the pretrained backbone of ResNet50 and focus on training the decoder and head module. Therefore, as shown in Fig. 7(b), a turning point appears around 45 epochs in the loss curve of the teacher model when the backbone of ResNet50 is unfrozen. At the end of the training process, both teacher and student models are stable. It is seen from Fig. 7(a) that the average precisions of the teacher model in terms of crack, tear and scratch are generally larger than that of the student model, denoting that the teacher model outperforms the student model. However, considering these three kinds of damages, the mean average precision (mAP) of the student model achieves as high as 92.53%, which is only a little smaller than that of the teacher model (94.55%), indicating the compact student model learned the excellent feature extraction ability based on the proposed Fusion Knowledge Distillation and has achieved almost comparable performance of the teacher model.

To further verify the proposed Fused Channel-Spatial Attention and Fusion Knowledge Distillation, the backbones of MobileNet and Xception were introduced. And ResNet101 as another teacher model was introduced to verify the influence of different teacher models. Since the output sizes of different backbones are not compatible with the decoder of CenterNet, several convolutional and ReLU layers were appended to adjust the output size of the latent feature maps. To show the comparison more clearly, Logarithmic Model Efficiency (LME) is introduced as follows:

$$LME = -\log(mAP \cdot FPS / Paras) \tag{17}$$

where mAP denotes mean Average Precision, and FPS and $Paras$ represents Frames Per Second and Parameters, respectively. The LME considers the influences of model size and performance, and takes the negative logarithmic results to make it more readable. The experimental results of models with different backbones and other object detection algorithms are listed in Table 5.

The smaller LME represents more balanced performance among precision, speed and compute cost. It can be seen from Table 5 that the proposed conveyor belt damage detection algorithm with Fusion Knowledge Distillation and Fused Channel-Spatial Attention can achieve excellent performance and make a favorable balance between precision and speed. Although the Xception model barely show model compression because of its complicated structure, the mAP of the Xception model with FKD strategy has been largely improved than that of same model without FKD, which can be inferred from the decreased LME. The comparison between teacher models, student models with and without FKD strategy are listed in Table 5. It can be concluded that teacher models obtained better mAP scores but larger storage occupations, which are suitable for scenarios with high-performance requirements. The FKD strategy can improve compact model performance obvi-

Table 4 The training results based on different optimizers.

Optimizers	SGD	Momentum	AdaGrad	RMSprop	Adam
mAP (%)	87.15	88.97	91.24	90.61	92.53
Final loss	1.648	0.939	0.793	0.811	0.746
Time consumption (h)	8.1	9.1	9.8	8.7	9.5

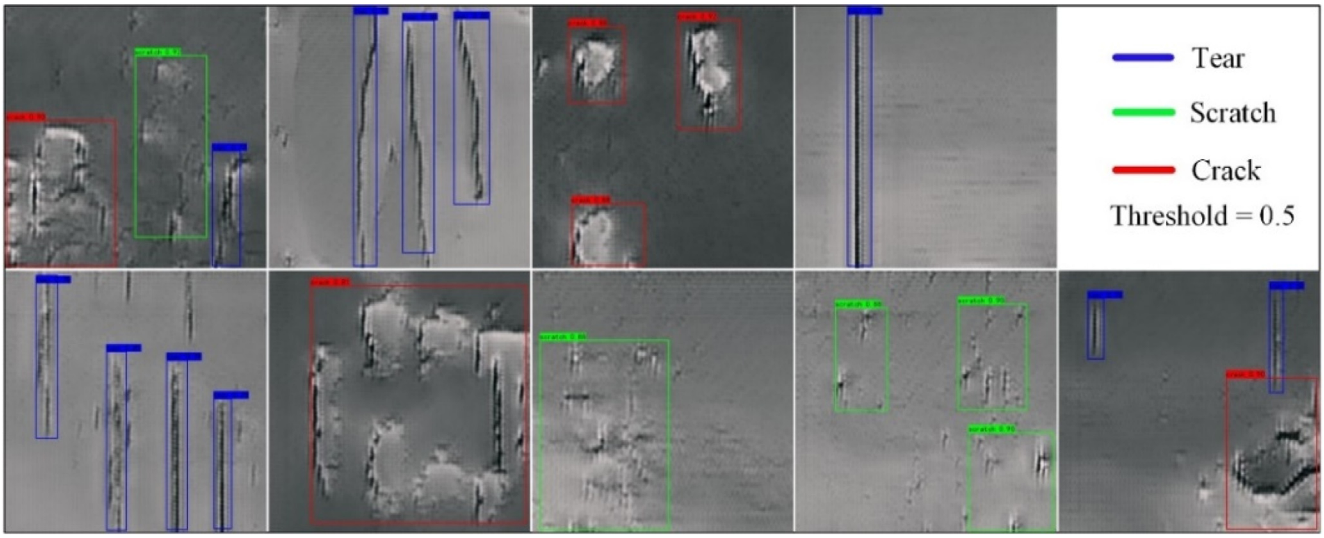

Fig. 6 The experimental results of the teacher model. The score threshold is set to 0.5 and different damages are displayed in blue, green and red colors. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

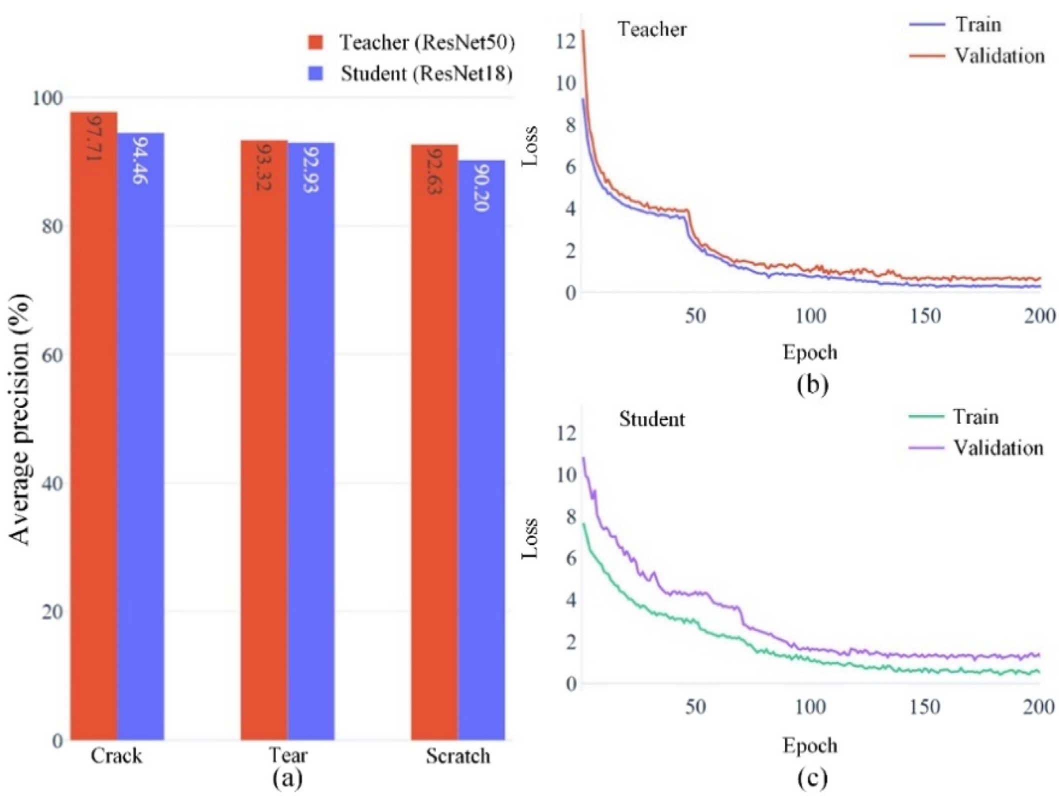

Fig. 7 The comparison of teacher (ResNet50) and student (ResNet18) models. (a) The AP of teacher and student models. (b) The loss curves of the teacher model. (c) The loss curve of the student model. Student model has a slight performance degradation compared with Teacher model, and it is within the acceptable range.

ously, which make it balanced in performance and resource consumptions. But different teacher models with small performance gaps hardly have influences on training effect of student models applied the FKD strategy. Therefore, the proposed FKD strategy can enhance compact model performance and is relatively easy to implement. In addition, YOLOX [34], DetectoRS [35] and Swin Transformer [36] are studied and compared with the proposed FKD-CenterNet model based on the custom conveyor belt dataset. These models reached excellent benchmark performance on public datasets, such as COCO and ImageNet, but required dedicated fine-tuning to custom dataset. The test results of above networks on custom

Table 5 The experimental results of the proposed CenterNet algorithm with different backbones and other models (ResNet · in bracket represents teacher model used for the proposed FKD strategy).

Model	Backbone	Storage (MB)	mAP (%)	FPS	LME
Teacher	ResNet50	127.8	94.55	45.4	1.0910
	ResNet101	184.3	95.21	38.4	1.6176
Student	ResNet18 (ResNet50)	89.7	92.53	65.8	0.3875
	MobileNet v3 (ResNet50)	71.3	89.82	46.1	0.5434
	Xception (ResNet50)	125.1	91.47	42.6	1.1664
	ResNet18 (ResNet101)	89.7	92.85	65.5	0.3886
	MobileNet v3 (ResNet101)	71.3	88.37	44.9	0.5861
	Xception (ResNet101)	125.1	91.63	42.7	1.1623
Same model without FKD strategy	ResNet18	89.7	63.8	66.2	0.7532
	MobileNet v3	71.3	67.4	45.9	0.8350
	Xception	125.1	72.9	42.3	1.4004
YOLOv3	DarkNet53	235.9	85.47	43.7	1.8431
YOLOX	CSPDarkNet	193.0	87.91	73.6	1.0929
SSD300	VGG16	101.6	81.73	59.1	0.7436
Fast R-CNN	ResNet50	108.3	87.27	8.4	2.6928
DetectoRS	ResNet50	515.0	89.4	9.2	4.1370
Swin Transformer	Swin-S	170.2	83.2	14.3	2.6606

conveyor belt damage dataset are very competitive in a certain aspect, e.g., mAP or FPS, but these models, even compact version, contains a large number of parameters. Hence, the LME results of above three popular networks are not attractive and these models are not appropriate for end devices. In general, the proposed lightweight models with the backbone of ResNet18 or MobileNet are suitable for end devices. Meanwhile, the proposed Fusion Knowledge Distillation can also be performed in other networks to compress CNN models.

4. Application of the proposed FKD-CenterNet

After the verification via experiments, the FKD-CenterNet based conveyor belt damage detection system was deployed in a practical mining seam to inspect and detect the state of the belt conveyor surface in real-time, as demonstrated in Fig. 8. The inference is processed on industrial PCs, and the specification is listed in Table 6. The model deployed in the production environment is based on ResNet50 and trained by FKD strategy. The belt images are captured by CMOS industrial camera at fixed 25 fps in the actual environment. It has been verified that the proposed FKD-CenterNet model with the deployed industrial PC can reach the maximum 43 fps while detecting belt damage. The theoretical computing power of NVIDIA GTX 1660 s is 5.027 TFLOPs. The GPU resource occupancy rate is around 45%, when the proposed model is inferencing and deployed on the 1660 s card. Therefore, any supported GPU computing card with more than 5 TFLOPs computing power can meet the requirements. In the meanwhile, the speed of data access in memory is fast enough and would not restrict GPU inference speed. The performances of CPU, RAM and SSD are not very important in this case. It is capable of handling 130% normal computation load with the configuration in Table 6. Of course, a high-performance IPC would be fantastic. However, a higher FPS (greater than 43) does not improve the detection accuracy of the proposed model, and it is necessary to achieve a balance between cost and speed. Therefore, the configuration in Table 6 is ideal for this case. Each captured image is preprocessed (denoise and smooth) and fed into the well-trained CenterNet model. Due to the model inference does not require a high-performance GPU, the industrial PC is much more compact than data server used for training. Since conveyor belt damages tend to be found near driving roller or joint section, two detecting nodes are set near front and end conveyor machine, and one is set in the middle of 800-meter conveyor belt.

The lightweight model trained with FKD is found to suit the end device very well as it reduced data transmission delays and improved the stability of the damage detection system. During the industrial test, the proposed FKD-CenterNet model detect various conveyor belt damages with 97.9% mAP, in particular, 100% tear, 98.7% crack and 94.9% scratch. Considering the severity of different damage types and undesirable false detection of tear damage, the trial results are quite acceptable. The application of the belt damage detection system has greatly improved the level of coal mining automation and reduced the labor cost of the enterprise.

5. Conclusions and future works

In this paper, an improved conveyor belt damage detection method based on the FKD-CenterNet algorithm is proposed to address the damage detection problem for conveyor belt images. Firstly, The Fused Channel-Spatial Attention is proposed and the Fusion Knowledge Distillation is introduced. It makes model focus on essential features and strengthens the extraction capability of latent feature maps. Then, the improved loss function combined with Fusion Knowledge Distillation is introduced, which balances different classes and transfer compressed knowledge from the teacher network. Hence, the proposed loss function makes the training process effective. Finally, plenty of comparative experiments show that the proposed lightweight model reaches 92.53% mAP and 65.8

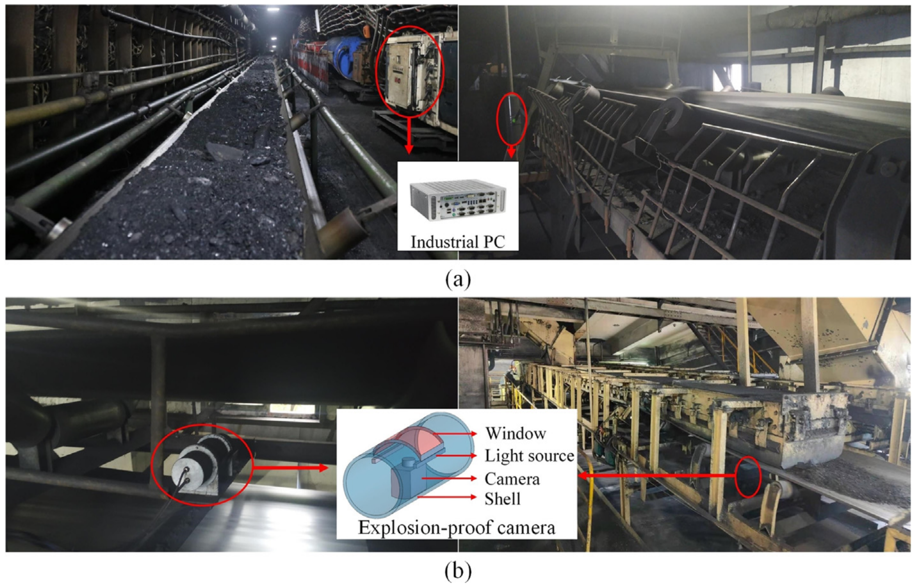

Fig. 8 The application of the belt damage detection system. (a) and (b) show the deployed industrial PC and explosion-proof camera, respectively.

Table 6 The description of the industrial PC for the model inference.

No.	Device name	Type	Specification
1	CPU	Intel I5-8500	Frequency: 3.0 GHz Turbo Clock: 4.1 GHz TDP: 65 W
2	RAM	8G DDR4 * 2	Dual channel, 2 slots 2133 MHz
3	Hard driver	256G SSD + 1 T HDD	Interface: NVMe + SATA Speed: 510 M/s
4	GPU	NVIDIA GTX 1660 s	Memory Size: 6 GB Bandwidth: 336.0 GB/s FP32 (float): 5.027 TFLOPS

FPS. Based on the proposed evaluation indicator balanced accuracy and inference speed, the proposed lightweight model obtains best performance, 0.3875. Therefore, the comparisons between the proposed FKD-CenterNet and other popular models and industrial test proved that the proposed model can detect various belt damages and be deployed on end devices effectively. The application in the coal mining industry shows the effectiveness of the proposed conveyor belt damage detection method.

Harsh underground environment for coal mining leads to poor image quality, which affects the performance of conveyor belt damage detection based on machine vision. In the future, we will focus on algorithm development for image enhancement and make efforts to obtain high-resolution and clear belt images. And we will try to establish a cascade model of denoising and detection, which makes the training and detection processes more efficient and accurate.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work is supported by the Key Research and Development Plan (Industrial Foresight and Common Key Technology) for Promoting Scientific and Technological Innovation Project Xuzhou City, Jiangsu Province, China (no. KC19074), the Priority Academic Program Development of Jiangsu Higher Education Institution (PAPD). The research leading to these results has received funding from the Norwegian Financial Mechanism 2014-2021 under Project Contract No 2020/37/K/ST8/02748.

References

[1] K. Chaabane et al, Integrated imperfect multimission selective maintenance and repairpersons assignment problem, Reliab. Eng. Syst. Saf. 199 (2020).

[2] N. Khakzad, F. Khan, P. Amyotte, Dynamic risk analysis using bow-tie approach, Reliab. Eng. Syst. Saf. 104 (2012) 36–44.

[3] S. Jocelyn et al, Application of logical analysis of data to machinery-related accident prevention based on scarce data, Reliab. Eng. Syst. Saf. 159 (2017) 223–236.

[4] X.W. Liu et al, Integrated decision making for predictive maintenance of belt conveyor systems, Reliab. Eng. Syst. Saf. 188 (2019) 347–351.

[5] S.F. Jin et al, Practical speed measurement for an intelligent vehicle based on double Radon transform in urban traffic scenarios, Meas. Sci. Technol. 32 (2) (2021).

[6] X.Q. Guo et al, Machine vision-based intelligent manufacturing using a novel dual-template matching: a case study for lithium battery positioning, Int. J. Adv. Manuf. Technol. 116 (7–8) (2021) 2531–2551.

[7] X.Y. Liu et al, Measuring Shape Parameters of Pearls in Batches Using Machine Vision: A Case Study, Micromachines 13 (4) (2022).
[8] J. Li, C. Miao, The conveyor belt longitudinal tear on-line detection based on improved SSR algorithm, Optik – Int. J. Light Electron Opt. 127 (2016).
[9] X.-L. Hao, H. Liang, A multi-class support vector machine real-time detection system for surface damage of conveyor belts based on visual saliency, Measurement 146 (2019) 125–132.
[10] R. Yang et al, Infrared spectrum analysis method for detection and early warning of longitudinal tear of mine conveyor belt, Measurement 165 (2020) 107856.
[11] C. Hou et al, Multispectral visual detection method for conveyor belt longitudinal tear, Measurement 143 (2019) 246–257.
[12] T. Qiao et al, Integrative binocular vision detection method based on infrared and visible light fusion for conveyor belts longitudinal tear, Measurement 110 (2017) 192–201.
[13] B. Yu et al, Dual band infrared detection method based on mid-infrared and long infrared vision for conveyor belts longitudinal tear, Measurement 120 (2018) 140–149.
[14] F. Zeng et al, Mini-Crack Detection of Conveyor Belt Based on Laser Excited Thermography, Appl. Sci. 11 (22) (2021).
[15] Z. Lv et al, Visual detection method based on line lasers for the detection of longitudinal tears in conveyor belts, Measurement 183 (2021) 109800.
[16] W. Li, C. Li, F. Yan, Research on belt tear detection algorithm based on multiple sets of laser line assistance, Measurement 174 (2021).
[17] J. Che et al, Longitudinal tear detection method of conveyor belt based on audio-visual fusion, Measurement 176 (2021) 109152.
[18] C. Hou et al, Research on Audio-Visual Detection Method for Conveyor Belt Longitudinal Tear, IEEE Access 7 (2019) 120202–120213.
[19] H. Jafarzadeh et al, Hydrogen production via sodium borohydride hydrolysis catalyzed by cobalt ferrite anchored nitrogen-and sulfur co-doped graphene hybrid nanocatalyst: Artificial neural network modeling approach, Chem. Eng. Res. Des. 183 (2022) 557–566.
[20] R.N.E. Tiri et al, Improving hydrogen generation from dehydrogenation of dimethylamine borane using polyvinylpyrrolidone stabilized platinum-rhodium nanoclusters as highly efficient and reusable catalysts: Development of ANN model, Chem. Eng. Res. Des. 182 (2022) 305–311.
[21] A.A. Mehrizi et al, Artificial neural networks modeling ethanol oxidation reaction kinetics catalyzed by polyaniline-manganese ferrite supported platinum-ruthenium nanohybrid electrocatalyst, Chem. Eng. Res. Des. 184 (2022) 72–78.
[22] V.R. Hosseini et al, Application of a physics-informed neural network to solve the steady-state Bratu equation arising from solid biofuel combustion theory, Fuel 332 (2023) 125908.
[23] X. Guo et al, Damage Detection for Conveyor Belt Surface Based on Conditional Cycle Generative Adversarial Network, Sensors 22 (9) (2022) 3485.
[24] G. Wang et al, A belt tearing detection method of YOLOv4-BELT for multi-source interference environment, Measurement 189 (2022) 110469.
[25] M. Zhang et al, Deep learning-based damage detection of mining conveyor belt, Measurement 175 (2021) 109130.
[26] G. Ross, et al., *Rich feature hierarchies for accurate object detection and semantic segmentation*. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014: p. 580-587.
[27] J. Redmon, et al., *You Only Look Once: Unified, Real-Time Object Detection*. 2016. 779-788.
[28] Liu, W., et al., *SSD: Single Shot MultiBox Detector*. Vol. 9905. 2016. 21-37.
[29] X. Zhou, D. Wang, and P, Krähenbühl, Objects as Points, 2019.
[30] H. Guo et al, A CenterNet + + model for ship detection in SAR images, Pattern Recogn. 112 (2021) 107787.
[31] R. Wang, C.F. Cheung, CenterNet-based defect detection for additive manufacturing, Expert Syst. Appl. 188 (2022) 116000.
[32] Z. Dai et al, Cascade CenterNet: Robust Object Detection for Power Line Surveillance, IEEE Access 9 (2021) 60244–60257.
[33] J. Gou et al, Knowledge Distillation: A Survey, Int. J. Comput. Vis. 129 (6) (2021) 1789–1819.
[34] Z. Ge, et al., *YOLOX: Exceeding YOLO Series in 2021*. 2021.
[35] S. Qiao, L.C. Chen, A. Yuille. *DetectoRS: Detecting Objects with Recursive Feature Pyramid and Switchable Atrous Convolution*. in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2021.
[36] Z. Liu, et al. *Swin Transformer: Hierarchical Vision Transformer using Shifted Windows*. in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. 2021.