

Improving the industrial defect recognition in radiographic testing by pre-training on medical radiographs

Han Yu, Xingjie Li^{*}, Huasheng Xie^{**}, Xinyue Li, Chunyu Hou

State Key Laboratory of Advanced Casting Technologies, China Academy of Machine Shenyang Research Institute of Foundry Co., Ltd, Liaoning, China



ARTICLE INFO

Keywords:

Transfer learning
Industrial defect recognition
Medical radiographs
Digital radiography

ABSTRACT

Deep learning methodologies have gained substantial traction for defect recognition in industrial radiographic testing including welds, castings and other fields. Regardless of the deep learning utilized, it has emerged as a standard configuration to use a model pre-trained from ImageNet to accelerate convergence and enhance recognition accuracy. However, there is a significant gap between the domain of natural images and industrial radiographs, raising the question of whether there might be a superior pre-training method than relying on ImageNet pre-training. Fortunately, medical radiographs are more similar to industrial radiographs than natural images because of the same imaging method. In this paper, we initially utilize numerous medical radiographic images from CheXpert dataset to train a pre-trained CNN model. Then, we apply this model to four distinct tasks within two radiographic testing scenarios to validate its advantages and generalization capabilities. Finally, experiments on multiple datasets indicate that our method brings more benefits than ImageNet pre-training or training from scratch, with a F1 score improvement of 3.41 %–13.72 % for defect classification and a mIoU improvement of 1.05 %–6.58 % for defect segmentation. It demonstrates that pre-training from medical radiographs provides a cost-free improvement for all kinds of tasks in industrial defect recognition.

1. Introduction

Radiographic testing (RT) is one of the five basic non-destructive testing (NDT) technologies. Because RT can intuitively display the size, distribution and shape of internal defects in mechanical components, it has been widely used in industrial scenes and can not be even replaced in some crucial fields. In recent years, digital radiography (DR) has gradually matured and become a research hot spot in RT [1–3]. The implementation process of DR technology includes three steps. 1) X-ray tube injects ray energy into the tested object. 2) Digital detector array (DDA) collects the results of the interaction between ray energy and the tested object. 3) Computer displays the interaction results in the form of images. From above, DR has high detection efficiency, low cost, little pollution, and creates conditions for the vision-based recognition of internal defects [4,5].

Advancements in machine vision technology have made deep learning algorithms the primary reliance for automatic defect recognition in DR. Mery et al. [6] established a public dataset called GDXray, which supports the development of computer vision algorithms for

defect recognition in radiographs. Chang et al. [7] proposed an improved segmentation model with cylindrical projection to strengthen the ability to recognize the minor size defects in welds. A GAN (Generative Adversarial Network) model named CECGAN [8] is proposed to solve the problem of data imbalance in welds radiographs. Tang et al. [9] utilize BCNN (Bilinear Convolutional Neural Networks) to improve the model's ability to classify defects in castings. Du et al. [10] proposed a dual-stream network for better segmentation result than UNet [11]. Yu et al. [12] proposed an adaptive selection network of depth and receptive field, which realized an end-to-end defect segmentation in casting X-rays. Zheng et al. [13] presented a U-shape saliency detection network for tire defect segmentation.

Typically, the methods mentioned above utilize the backbone network pre-trained in ImageNet [14] for feature extraction. Using weights pre-trained on ImageNet has been established as a beneficial method for conducting transfer learning across various image recognition applications. Numerous studies [15–18] have shown that a pre-trained convolutional neural network (CNN) not only accelerates the convergence of the model but also enhances its accuracy. When

* Corresponding author.

** Corresponding author.

E-mail address: lixj@chinasrif.com (X. Li).

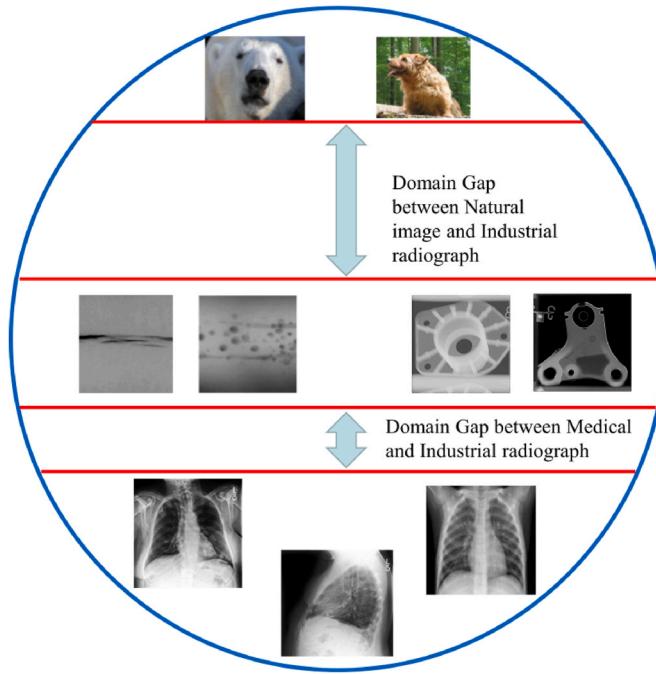


Fig. 1. The domain gap between industrial radiographs and natural images is significantly larger than that of medical radiographs.

Table 1

THE MAIN DISCREPANCY AMONG NATURAL IMAGES, MEDICAL RADIOPHGRAPHS, AND INDUSTRIAL RADIOPHGRAPHS.

	Natural Image	Medical radiograph	Industrial radiograph
Color depth	3	1	1
Perspective distortion	Significant	Minimal	Minimal
Semantic information	Rich	Less	Less
Discriminant information	Global	Local	Local
The number of images	Large	Medium	Less
The number of classes	≤ 1000	< 100	< 10

training data is scarce, the absence of pre-trained weights may result in the model failing to converge. The advantages of pre-training stem from the CNN's ability to derive generalized feature representations through a broad array of natural images. As shown in Fig. 1 and Table 1, the discrepancy between industrial radiographs and natural images is significantly larger than that of medical radiographs. The details can be summarized as follow:

Color depth: Due to the difference in imaging methods, natural images have color, while radiographs are single-channel gray images. It leads to the failure of the color features extracted by the pre-trained from ImageNet.

Perspective distortion: Perspective distortion is significant in natural images. However, the perspective distortion in radiography is typically minimal compared to natural images, as the objects being imaged are generally positioned at similar distances from the X-ray source, resulting in nearly parallel rays reaching the detector.

Semantic information: Each natural images appear different and have rich semantic information. By contrast, the similarity of radiographs is very high due to the high similarity of human tissue and mechanical components.

Discriminat information: Natural image recognition is usually accomplished by identifying the clear and global subjects, while pathologies in medical radiographs and defects in industrial radiographs are recognized by detecting local abnormalities and texture variations.

The number of images: Natural images are easy to obtain and be labelled. However, relying heavily on expert experience, the datasets in medical and industrial have fewer images with annotations than in the natural image, which ranges from dozens of images to a couple hundred thousand. For example, ImageNet contains approximately 1.2 million training data. CheXpert [22] includes 220 thousand chest radiographs, and GDXray consists of 19 thousand radiographs.

The number of classes: The category of pathology and defect is limited, usually no more than 100. However, there are 1000 object classes in ImageNet.

Based on the above analysis, there is a significant gap between domain of natural images and industrial radiographs, which prompts a question: is there a better way than using the pre-trained model from natural images? Nevertheless, previous research mainly focuses on improving the structure of CNN, overlooking the possible influence of the pre-trained model. Therefore, we attempt to improve the industrial defect recognition in RT by obtaining a better pre-trained model. The core contribution of this paper can be outlined as follows.

- 1) We firstly explore the possibility of realizing a cost-free improvement of industrial defect recognition by using the pre-trained weight from medical radiographs.
- 2) We apply the pre-trained model, originally developed for medical radiograph analysis, to four specific tasks across two RT scenarios: defect classification and segmentation within both welds and casting environments. Several experiments demonstrate that training from medical radiographs brings more improvement than training from ImageNet or training from scratch.
- 3) We will release our pre-trained models, which will facilitate the community to apply this better pre-trained model to other applications in RT.

2. Related work

The deep neural network is naturally suitable for transfer learning. “Pre-training + fine-tune” becomes a standard configuration when using deep learning models. Yosinski et al. [15] found that “Pre-training + fine-tune” can improve generalization performance, and the transferability of features decreases as the distance between the pre-training task and downstream task increases. He et al. [16] concluded that ImageNet pre-training speeds up convergence early in training. Kornblith et al. [17] proved that pre-training from large datasets determines the lower limit of downstream tasks, and the improvement increases as the number of training data in downstream tasks decreases. Dan et al. [18] showed that pre-training would strengthen the robustness to noise label, imbalance, and cross-domain. However, given the marked differences between natural and medical images, there are numerous studies to explore whether pre-training directly from medical images yields more powerful improvement. Raghu et al. [19] explore the properties of ImageNet pre-training for medical images, which offers limited performance gains. Wen et al. [20] investigated the properties of medical pre-training and its transfer effectiveness on various medical tasks. The experiment showed that medical pre-training performed better for classification tasks but not for segmentation tasks. Multiple 3D medical datasets were aggregated [21] and utilized to train a 3D pre-trained CNN, which boosted the CT and MRI tasks. Inspired by the above research, we first questioned the necessity of transfer learning from ImageNet in industrial defect recognition and tried to explore a novel pre-training scheme.

3. Proposed method

This section provides a detailed description of our processing workflow, depicted in Fig. 2. Our approach consists of two primary phases. Initially, we train a backbone network using medical radiographs within a multi-label classification framework. Subsequently, the

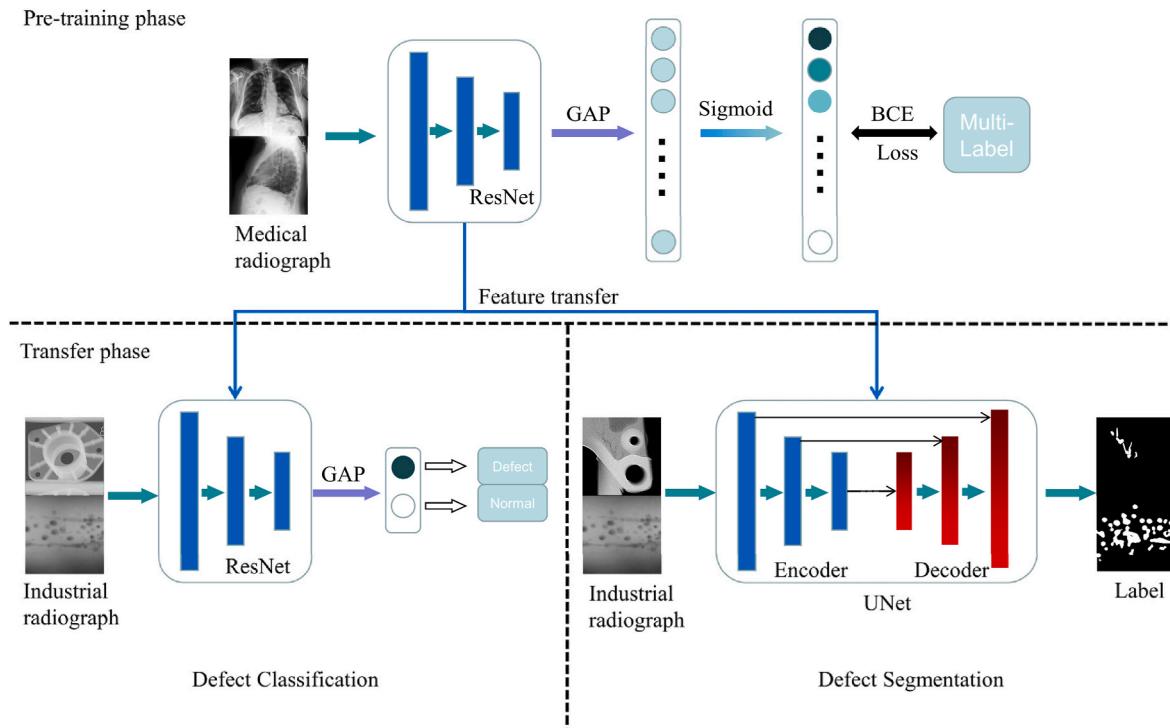


Fig. 2. The overall procedure of the proposed method.

Table 2

The details of Implementation in different tasks.

Dataset	Task	Batch Size	Initial learning rate	Final learning rate	Decline of learning rate	Epoch	Validation Method	Input Size	Number of samples
GDXray-weld	Defect Classification	32	1e-3	1e-6	cosine	100	5-Fold Cross Validation	256 × 256	316
	Defect Segmentation	16	5e-3	1e-6					
Zju-casting	Defect Classification	32	5e-3	1e-6	cosine	30	Hold-out	500 × 500	11,070
	Defect Segmentation	12	1e-3	1e-6					

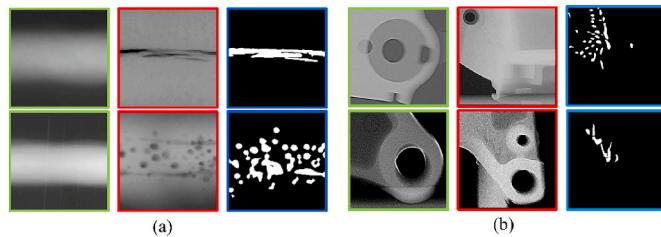


Fig. 3. The visualization of two datasets (a): GDXray-welds (b): Zju-castings. Images with the green boxes represent normal sample, images with the red boxes represent defective sample, and the blue are their annotation. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

pre-trained weights are applied to tasks involving industrial defect classification and segmentation to enhance their performance.

3.1. Pre-training phase

The motivation of this paper is to train a high-performance CNN backbone with medical radiographs, which can serve as pre-trained weights to enhance defect recognition tasks in the industrial scene

Table 3

The quantitative comparison with other weight initialization in defect classification.

Dataset	Method	Accuracy↑	F1↑
GDXray-weld	Random	85.03	0.8529
	ImageNet	93.44	0.9379
	Ours	97.78	0.9699
Zju-casting	Random	71.31	0.7125
	ImageNet	89.56	0.8928
	Ours	91.69	0.9151

with insufficient training data. The dataset used for pre-training should include enough training data and categories to reach this target. More training data and categories can ensure the diversity of images and make the network learn discriminative features. Therefore, we select the CheXpert [22], a large dataset that contains 224,316 chest radiographs of 65,240 patients, as our dataset in the pre-training phase. The CheXpert also contains 14 categories: No Finding, Enlarged Cardiomediastinum, Cardiomegaly, Lung Opacity, Lung Lesion, Edema, Consolidation Pneumonia, Atelectasis Pneumothorax, Pleural Effusion, Pleural Other, Fracture, Support Devices. Each radiograph may belong to multiple categories at the same time. Thus, we approach it as a multi-label classification problem. ResNet50 [23] is chosen for the basic

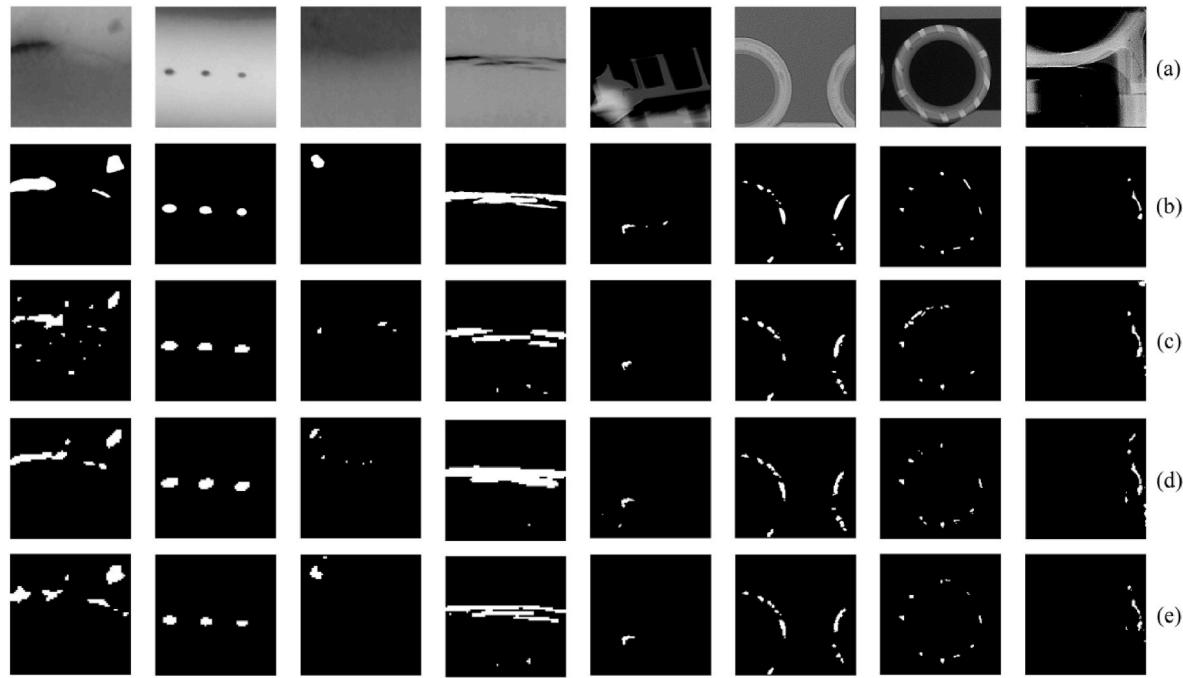


Fig. 4. Visual comparison of defect segmentation with other weight initialization methods. The first four columns are the results of the GDXray-weld dataset, and the last four columns are the results of the Zju-casting dataset. (a): Radiographs (b): Ground-truth (c): Random (d): ImageNet (e): Ours.

Table 4
The quantitative comparison with other weight initialization in defect segmentation.

Dataset	Method	Precision↑	Recall↑	Accuracy↑	mIoU↑
GDXray-weld	Random	69.20	48.35	73.64	68.00
	ImageNet	71.05	51.63	75.18	69.35
	Ours	74.28	61.93	80.45	73.92
Zju-casting	Random	50.61	50.69	75.16	66.58
	ImageNet	54.44	54.24	76.95	68.32
	Ours	56.79	54.93	77.31	69.04

framework of the backbone because it is the most popular CNN architecture. We firstly change the channel number of its first convolution layer from 3 to 1 due to the character of radiographs. The image is

transferred to its feature map by the backbone. The feature map is converted into the 14-dimensional vector through global average pooling (GAP) and a fully-connected layer. Then an element-wise sigmoid activation function is applied to the vector and obtain the final output. Every dimension of the final output is the predicted probability of the presence for each category. So we can employ a binary cross-entropy (BCE) loss to train this network:

$$L_{bce} = - \sum_{i=1}^{14} [y_i \log p(y_i) + (1 - y_i) \log(1 - p(y_i))] \quad (1)$$

where $p(y_i)$ means the predicted probability which the image belongs to category i and $y_i \in \{0, 1\}$ is the ground truth. $y_i = 1$ means a positive sample that contains the i_{th} pathology. Due to the limited image definition, some images in this dataset cannot be judged by experts whether

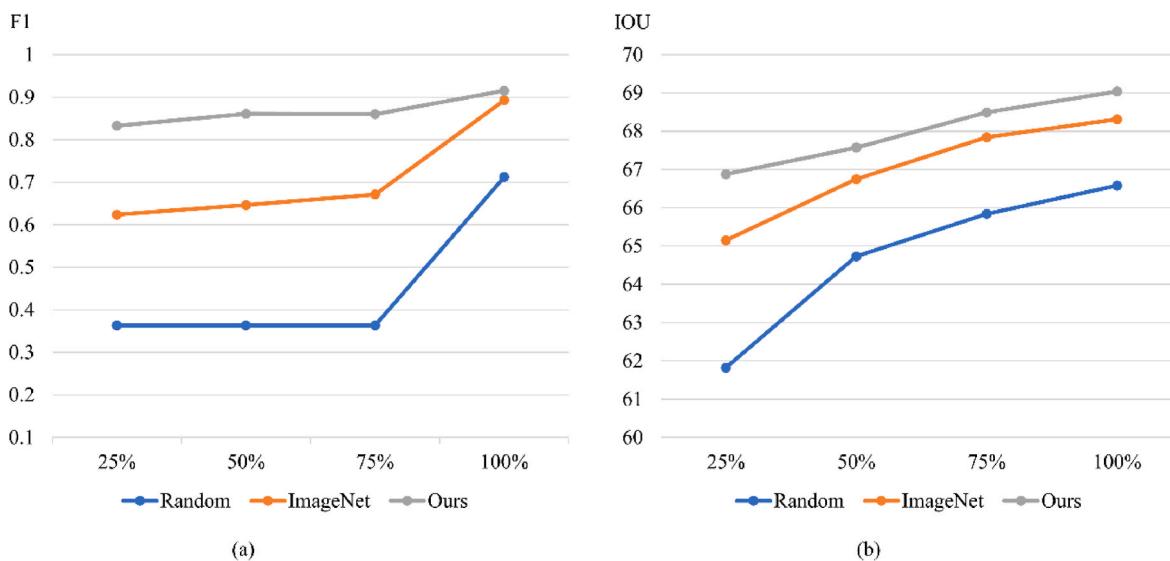


Fig. 5. The impact of training data magnitude.

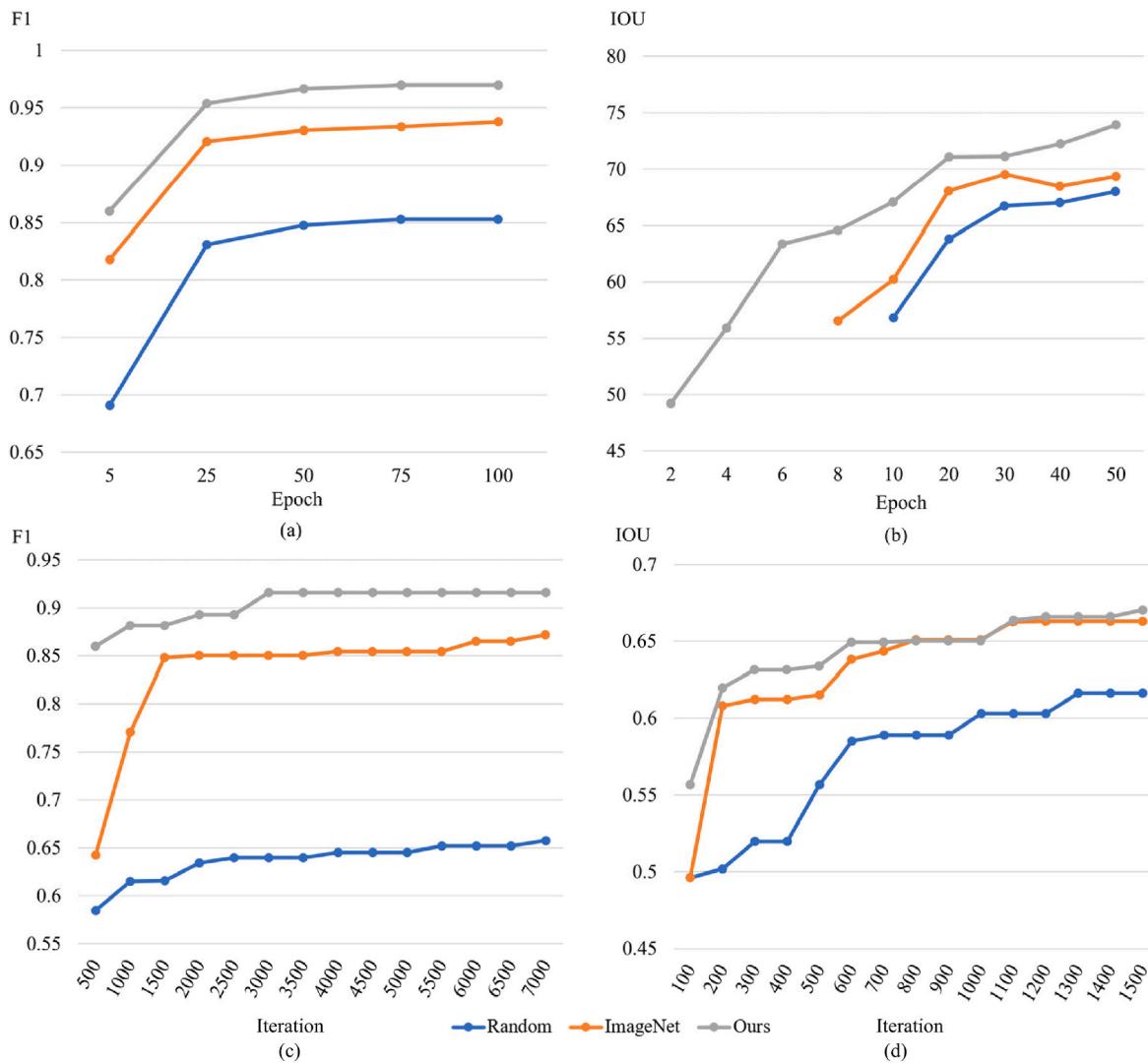


Fig. 6. The impact of training time magnitude (a): weld defect classification (b):weld defect segmentation (c):casting defect classification (d):casting defect segmentation.

it belongs to a category. These uncertain samples are regarded as the positive ones in this work.

Following the official division, we compute the loss in the validation set for every epoch. The network weights with the lowest validation loss are saved for downstream tasks.

3.2. Transfer phase

After obtaining the pre-trained weight, we will apply it to defect recognition tasks in radiographic testing. Defect classification and segmentation are the most common mode to achieve vision-based defect recognition. Defect classification aims to judge whether there are defects in the image. It is similar to the pre-training phase. We directly use the pre-trained weights as the initial weights of the backbone. Only a fully connected and a softmax layer are incorporated behind the backbone to reduce the interference. The feature maps extracted by pre-trained weights are transferred to the probability of the defect presence.

The target of defect segmentation is to achieve the accurate location, area measurement, and depth estimation of defects. UNet has become one of the most popular models in image segmentation with insufficient training data. Original UNet employs the VGG as its encoder, which extracts the multi-scale features of an input image. In this work, the pre-trained ResNet50 replaces VGG as the encoder, and random

initialization is used for the decoder. Pixel-wise BCE and Dice loss are jointly leveraged to fine-tune our network due to the imbalance data in defect segmentation task.

For the above two tasks, it is worth mentioning that we will not freeze the pre-training weight but take it as the changeable initial value of the network weight in fine-tuning because of the remarkable discrepancy among image domains.

Subsequently, the pre-trained weights will be evaluated its advantage and generalization in the four tasks of two RT scenes.

4. Experimental setup

4.1. Datasets and preprocessing

RT is the most important means of non-destructive testing for welds and castings. We select the GDXray-welds [6] and Zju-castings [10] as our datasets for comparison with weight from natural images.

GDXray-welds dataset originally contains ten welds radiographs and their binary annotations for defects. The image size of these radiographs is from 366×3512 to 1091×4953 . Because the image size is too large, we crop the original radiographs and annotation images simultaneously and get 316 small patches. According to the annotation images, we can easily distinguish whether there are defects in some patches. Thus, we

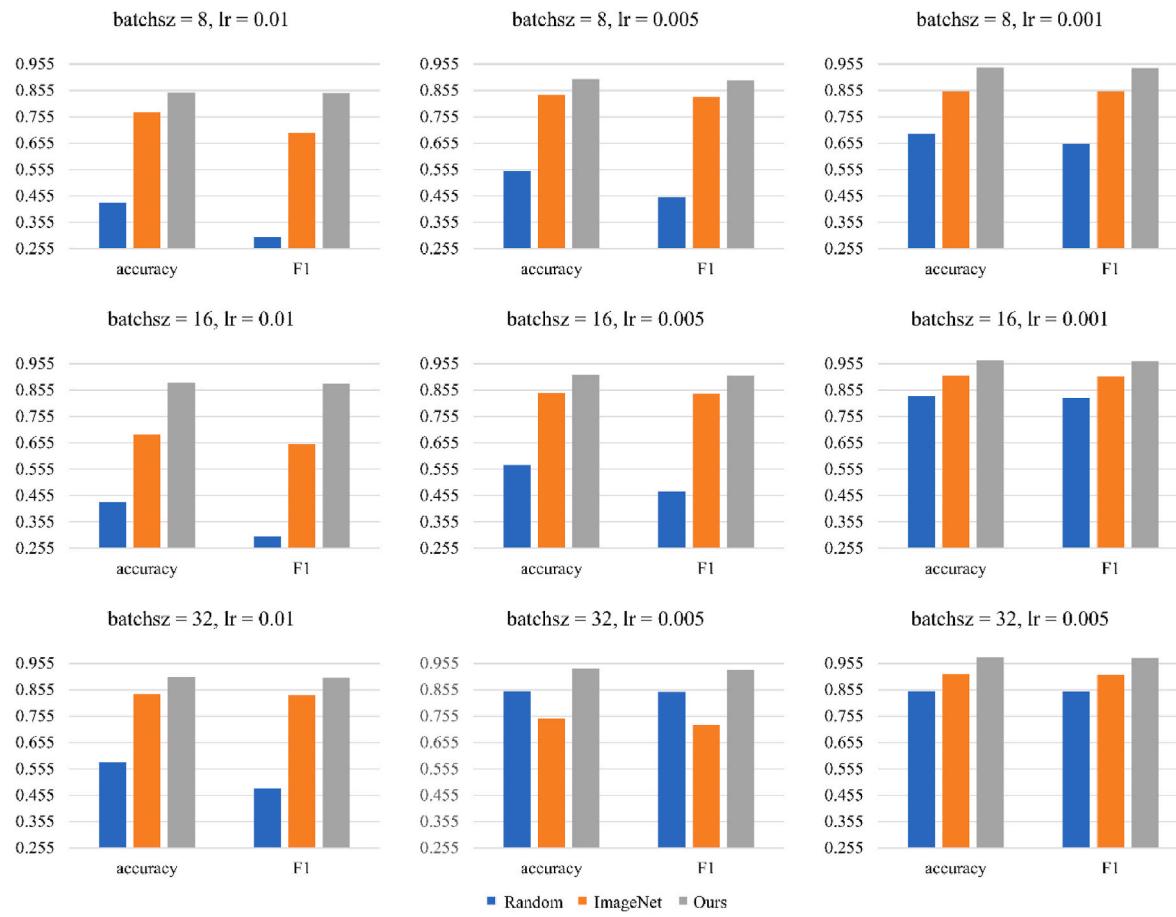


Fig. 7. The impact of hyper-parameters.

construct a defect classification task for distinguishing normal and defective patches. Naturally, the defective patches and their annotations can be used for defect segmentation tasks.

Zju-castings dataset collects 3148 castings radiographs and their labels of six automobile parts. All image size is 1000×1000 . We still utilize the above method to process this dataset. Fig. 3 and Table 2 shows the details of the two datasets.

The above two datasets and four tasks not only cover the typical scenes of RT but also validate the performance under a small-sample or relatively enough sample.

4.2. Implementation details

The computational setup for this study included an i9-9920X processor, 64 GB of RAM, and a TITAN RTX graphics card with 24 GB of GPU memory. All models were developed using the PyTorch framework. The weight trained by ImageNet is from the TorchVision package. We used horizontal and vertical flipping as the data augmentation methods for defect classification. For segmentation, transformations such as random scaling, and brightness adjustments, and rotation were applied. More important information is shown in Table 2 for each task.

4.3. Evaluation metric

For the defect classification task, we utilize Accuracy and F1 to evaluate. TP and FN denote the counts of defective images identified correctly or incorrectly, respectively. TN and FP indicate how many normal images are correctly or incorrectly recognized.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FN + FP} \quad (2)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (4)$$

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

$$mIoU = \frac{1}{2} \left(\frac{TP}{TP + FP + FN} + \frac{TN}{TN + FP + FN} \right) \quad (6)$$

As for the defect segmentation task, we select the precision, recall, and mIoU to evaluate the model performance. Here, TP and FN refer to the counts of defective pixels correctly predicted as defective and non-defective, respectively. Conversely, FP and TN denote the counts of non-defective pixels erroneously predicted as defective and correctly identified as non-defective.

5. Experiment

5.1. Comparison with other weight initialization

This section will explore the pre-trained backbone's performance on four tasks for industrial defect recognition to validate its transferable capability.

We first compare the proposed method with training from random initialization and ImageNet in defect classification tasks. From Table 3, training from random initialization performs worst in both datasets as expected. The weight pre-trained from medical radiographs surpasses the other two initialization methods significantly. Compared with them,

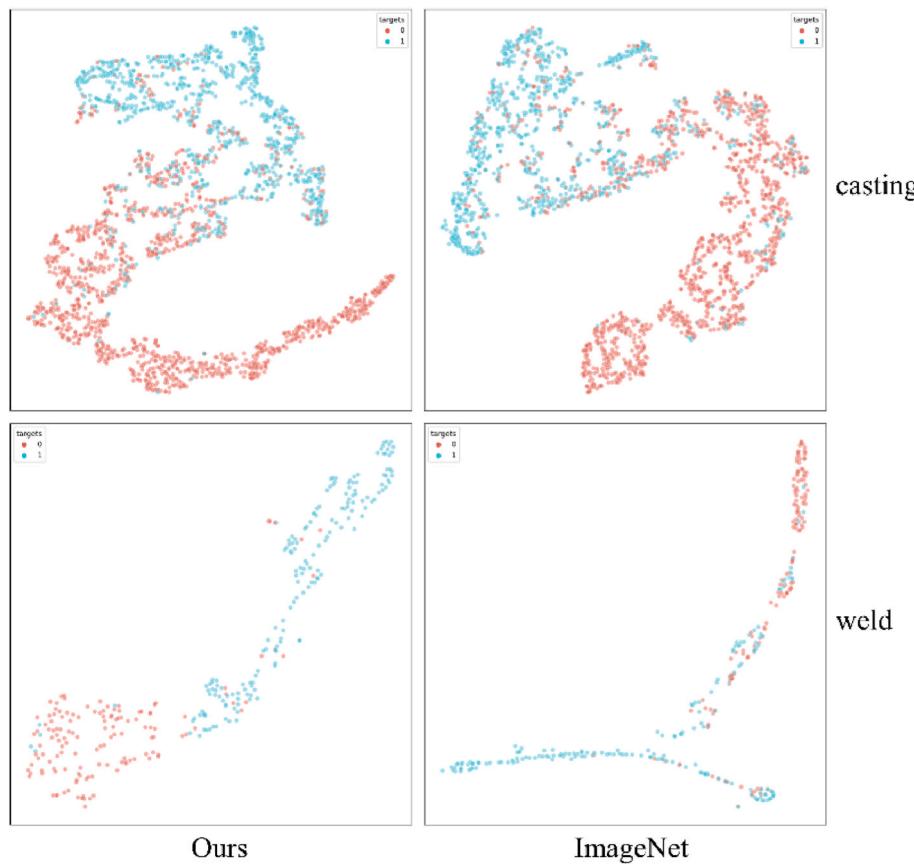


Fig. 8. The visualization of feature distribution based on t-SNE.

our method improves by 3.41 % and 13.72 % in the F1 score of GDXray-weld, respectively. Because there are more training samples in the Zju-casting dataset, our method's improvement is not as apparent as in GDXray-weld, but there is still an increase of about 2.49 % than training from ImageNet.

We visualize in Fig. 4 the representative results in two datasets. Intuitively, our segmentation results have more precise boundaries even in low contrast. From Table 4, pre-training from medical radiographs also benefits defect segmentation. Similar to the defect classification task, our method leads to a higher boost (6.58 %) in GDXray-weld than Zju-casting (1.05 %). Overall, Using pre-trained weights in classification tasks will obtain higher gain than segmentation. This occurs because the segmentation model has relatively few backbone parameters, and the impact of the decoder is somewhat restricted.

5.2. Impact of training data magnitude

The quantity of training data significantly impacts the performance of the deep learning model. In the defect recognition field, We often can not get enough training data, which requires the model to have strong feature extraction ability and represent the essential differences of samples. If our pre-trained model still has advantages in small samples, it will indicate that it has a better ability for feature extraction. To evaluate it, we train the three models using the different pre-trained weights with 25 %, 50 %, 75 %, and 100 % of the training data in the Zju-casting, validation and testing dataset remain unchanged for a fair comparison.

As shown in Fig. 5, the highest F1 and IOU scores are obtained, when all available training data is used for classification and segmentation tasks. When the quantity of training data is reduced, the model performance will drop because of the over-fitting. Using only 25 % data, our method achieves a 0.8327 F¹ score, which is 20.88 % higher than the

ImageNet in defect classification. Training from scratch even does not converge. When the number of training data grows, the performance disparity becomes smaller. Utilizing the 100 % data, our method still has a 2.49 % improvement. Our method improves more obviously in defect segmentation tasks when the amount of training data is insufficient as before. The above experiments demonstrate that pre-trained from medical radiographs has less dependence on the training data magnitude.

5.3. Impact of training time magnitude

In this part, we explore the impact of training time. The speed of convergence is also essential. If the pre-trained model can speed up the convergence, it means the network is trained from a better initial value. We utilize the GDXray-weld dataset and Zju-castings dataset for this testing and record the model performance when training different epochs or iterations in Fig. 6. Because of the different size of datasets, we use epochs and iterations as the time scales in the GDXray-weld dataset and Zju-castings dataset. From Fig. 6, pre-trained from medical radiographs have the fastest convergence speed among all tasks, followed by pre-trained from ImageNet. In the GDXray-weld dataset, our method achieves an F1 score of 0.86, 4.2 % higher than 0.8179 of ImageNet and 24.47 % higher than 0.6909 of random by training only five epochs. As the training process continues, our method still has the highest accuracy, but the gap is narrowing to 3.2 %–13.7 %. A similar phenomenon appears in the segmentation task. During the initial phase of the training process, the segmentation model has not converged, and their IOU values are NaN (not a number). However, our method converges from the second epoch, while pre-trained from ImageNet is from the 6th and random initiation is from the 8th. When all of the models begin to converge, our method's performance is also the best. In the Zju-castings dataset, we just demonstrate the early stage because the improvement

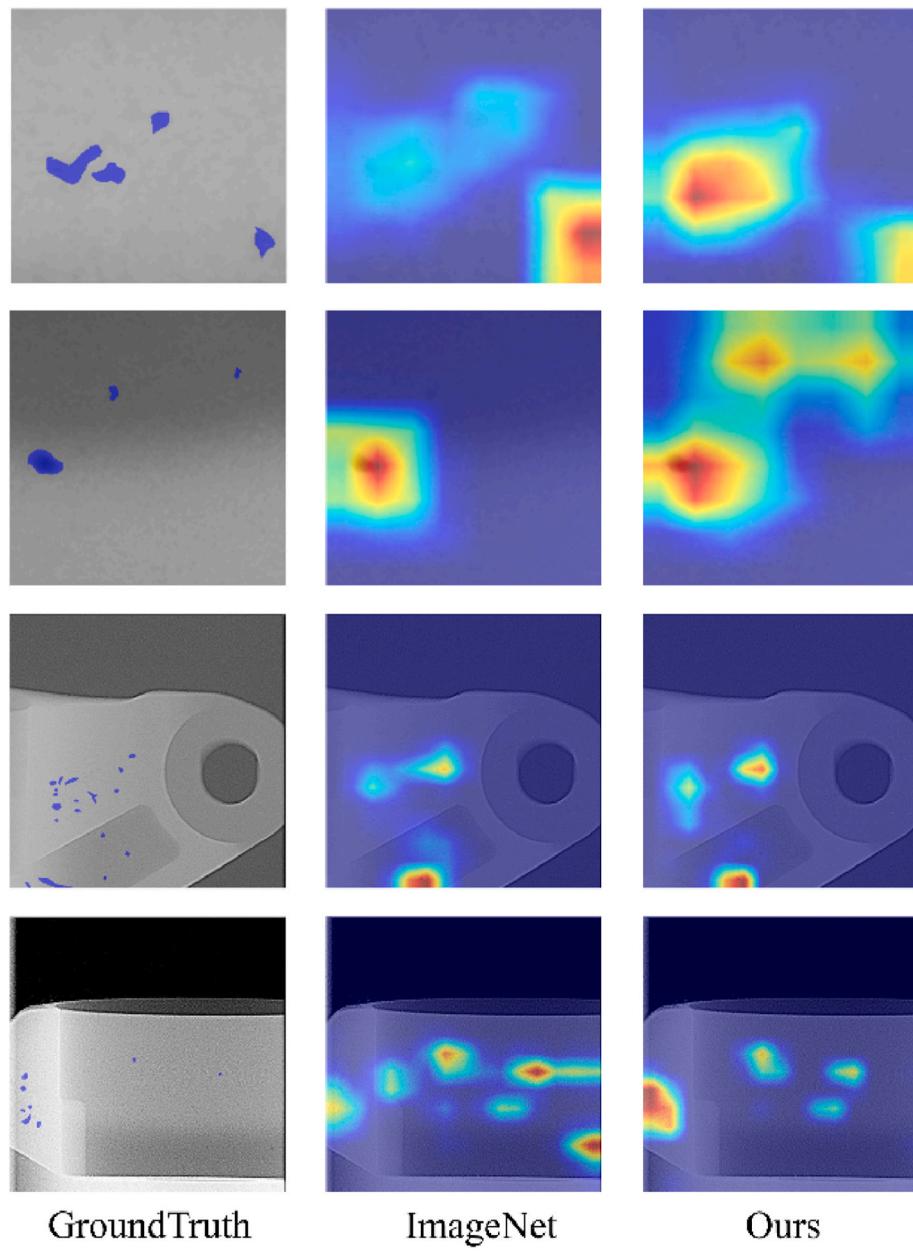


Fig. 9. The Grad-CAM visualization for different pre-trained model.

Table 5
The quantitative comparison with different backbone in defect classification.

Backbone	Method	Accuracy↑	F1↑
VGG16	ImageNet	48.79	0.3259
	Ours	56.33	0.3596
ResNet50	ImageNet	93.44	0.9379
	Ours	97.78	0.9699
ViT	ImageNet	57.30	0.4531
	Ours	72.03	0.7132

process is significant. From Fig. 6(c) and (d), pre-trained from medical radiographs also lead to a faster convergence speed, which proves that our method has excellent generalization in the field of industrial radiographic testing.

Table 6
The quantitative comparison with other different segmentation model and backbone in mIoU.

Segmentation Model	Backbone	ImageNet	Ours
UNet	VGG16	59.38	64.41
	ResNet50	68.32	69.04
	ViT	63.46	68.24
Deeplabv3+	VGG16	47.54	48.87
	ResNet50	62.13	62.84
	ViT	59.24	61.31

5.4. Sensitivity to hyper-parameters

Batch size and learning rate are crucial hyper-parameters in neural network training. To evaluate the sensitivity of pre-trained weights to these hyper-parameters, various combinations of batch size and learning rate were tested during the training process for a weld defect

classification task. As illustrated in Fig. 7, our method surpasses the other two initialization methods and achieves high stability for hyper-parameters.

5.5. Feature visualization

We visualize in Fig. 8 the distribution of features extracted by model pre-trained from ImageNet or medical radiographs based on the t-SNE method [24]. Each point means a sample of GDXray-weld or Zju-casting. Reds are the defective samples, and greens are the normal. We can see that all features extracted by our method are mapped more closely in the embedding space than ImageNet. Intuitively, the Aggregation effect of our method looks better in GDXray-weld, which aligns with the findings presented in Table 3.

Moreover, we also illustrate class activation maps of the different models with the Grad-CAM method [25]. Grad-CAM is an explainability method that can help understand the deep learning model. It calculates the gradient of a differentiable output and then spatially pools these gradients to ascertain neuron importance weights. These weights are then used to merge the activation maps, identifying the most critical features. As shown in Fig. 9, pre-trained from medical radiographs focuses more on the defective area.

5.6. Robustness across different backbone networks

To further demonstrate the robustness of our proposed method with various backbone networks, we extended our experiments to include two additional architectures: VGG16 [26] and ViT [27]. The selection of these models was deliberate, representing distinct categories of deep learning approaches. VGG16, as a representative of straightforward CNN, provides insights into how traditional architectures perform with our method. In contrast, the ViT represents the latest advancements in deep learning, leveraging self-attention mechanisms to model long-range dependencies. For the segmentation model, we further introduced another widely used model, Deeplabv3+ [28], which is an ideal complement to the UNet model in validating the adaptability of our proposed approach.

For the defect classification tasks, Table 5 presents a quantitative comparison of different backbone architectures with the ImageNet pre-trained weights and the proposed method in GDXray-welds dataset. The results show that the proposed method consistently outperforms the ImageNet pre-trained models across all architectures. Similar conclusions also occur in defect segmentation tasks. Table 6 compares the performance of different segmentation models and backbone architectures in terms of the mIoU in Zju-casting dataset. The results indicate that the proposed method enhances the segmentation performance for all models and backbones. For instance, the ViT backbone shows improved mIoU scores, increasing from 63.46 to 68.24 with UNet and from 59.24 to 61.31 with Deeplabv3+. These results highlight the robustness of our proposed method across different types of backbone networks, demonstrating its effectiveness in enhancing defect recognition capabilities.

6. Limitation

While our study contributes to the field of industrial defect recognition in RT, there are several limitations that should be considered. First, our study was limited to two RT scenes, which may not fully represent the wide variety of industrial radiographs. Additionally, while pre-training from medical radiographs proved to be effective, it may not be the optimal pre-training way for industrial radiograph applications. Finally, there may be potential biases in our study, such as selection bias in the choice of datasets and models. In light of these limitations, future research could focus on exploring other pre-training methods to achieve even better results in industrial defect recognition.

7. Conclusion

In this paper, we intend to challenge the validity and rationality of using pre-trained weight from ImageNet for defect recognition tasks of industrial RT. Through a comprehensive analysis of the differences among natural images, medical radiographs and industrial radiographs, we explore the possibility of realizing a cost-free improvement of industrial defect recognition using the pre-trained weight from medical radiographs. Under the multi-label classification framework in the CheXpert dataset, a pre-trained backbone network is obtained and transferred into defect classification and segmentation tasks. Multiple experiments demonstrate that pre-training from medical radiographs brings more improvement than from ImageNet or scratch, especially when there is less training data. In addition, using our pre-trained weights in classification tasks will obtain a higher gain than segmentation.

Moving forward, we plan to investigate self-supervised learning methods based on industrial radiographs to further improve representation learning.

CRediT authorship contribution statement

Han Yu: Software, Methodology, Conceptualization. **Xingjie Li:** Project administration, Funding acquisition. **Huasheng Xie:** Writing – review & editing, Visualization, Investigation, Data curation. **Xinyue Li:** Software. **Chunyu Hou:** Validation, Conceptualization.

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.

Acknowledgment

This research was funded by the Natural Science Joint Foundation of Liaoning Province (2023-MSLH-330), Science and Technology Joint Program of Liaoning Province (2023JH2/101700038).

Data availability

Data will be made available on request.

References

- [1] DeMaio DN, Herrmann T, Noble LB, et al. Best practices in digital radiography. *Radiol Technol* 2019;91(2):198–201.
- [2] Körner M, Weber CH, Wirth S, et al. Advances in digital radiography: physical principles and system overview. *Radiographics* 2007;27(3):675–86.
- [3] Sharma K, Sharma K, Sharma J, et al. Evaluation and new innovations in digital radiography for NDT purposes. *Ion Exch Adsorpt* 2023.
- [4] Gao Y, Song G, Li S, Chen F, Chen D, Song A. LineSpyX: a power line inspection robot based on digital radiography. *IEEE Rob Autom Lett* 2020;5(3):4759–65.
- [5] Chen Lei, et al. 3D positioning of defects for gas turbine blades based on digital radiographic projective imaging. *NDT E Int* 2023. Art no. 102751.
- [6] Mery D, Riffo V, Zschepel U, et al. GDXray: the database of X-ray images for non-destructive testing. *J Nondestr Eval* 2015;34(4):1–12.
- [7] Chang Y, Wang W. A deep learning-based weld defect classification method using radiographic images with a cylindrical projection. *IEEE Trans Instrum Meas Oct.2021;70:1–11*. Art no. 5018911.
- [8] Guo R, Liu H, Xie G, Zhang Y. Weld defect detection from imbalanced radiographic images based on contrast enhancement conditional generative adversarial network and transfer learning. *IEEE Sensor J May.2021;21(9):10844–53*.
- [9] Tang Z, Tian E, Wang Y, et al. Non-destructive defect detection in castings by using spatial attention bilinear convolutional neural network. *IEEE Trans Ind Inf Jan. 2021;17(1):82–9*.
- [10] Du W, Shen H, Fu J. Automatic defect segmentation in X-ray images based on deep learning. *IEEE Trans Ind Electron Dec. 2021;68(12):12912–20*.
- [11] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Proc. Int. Conf. Med. Image comput. Comput.-Assist (MICCAI). Intervention; 2015. p. 234–41.

- [12] Yu H, Li X, Song K, et al. Adaptive depth and receptive field selection network for defect semantic segmentation on castings X-rays. *NDT E Int Dec.* 2020;116(6):102345.
- [13] Zheng Z, Yang H, Zhou L, Yu B, Zhang Y. HLU2-Net: a residual U-structure embedded U-net with hybrid loss for tire defect inspection. *IEEE Trans Instrum Meas Nov.* 2021;70:1–11. Art no. 3527511.
- [14] Deng J, Dong W, Socher R, Li L-J, Li Kai, Fei-Fei Li. ImageNet: a large-scale hierarchical image database. In: Proc. IEEE conf. Comput. Vis. Pattern recognit (CVPR); 2009. p. 248–55.
- [15] Yosinski J, Clune J, Bengio Y, et al. How transferable are features in deep neural networks? In: Proc. Adv. Neural inf. Process. Syst.; 2014. p. 3320–8.
- [16] He Kaiming, et al. Rethinking imagenet pre-training. In: Proc. IEEE int. Conf. Comput. Vis. (ICCV); 2019. p. 4917–26.
- [17] Kornblith, et al. Do better imagenet models transfer better? In: Proc. IEEE conf. Comput. Vis. Pattern recognit (CVPR); 2019. p. 2661–71.
- [18] Dan H, Lee K, Mazeika M. Using pre-training can improve model robustness and uncertainty. In: arXiv: 1901.09960; 2019 [Online] Available: <https://arxiv.org/abs/1901.09960>.
- [19] Raghu M, Zhang C, Kleinberg J, et al. Transfusion: understanding transfer learning for medical imaging. In: Proc. Adv. Neural inf. Process. Syst; 2019. p. 32.
- [20] Wen Y, Chen L, Deng Y, et al. Rethinking pre-training on medical imaging. *J Vis Commun Image Represent Jul.* 2021;78(5):103145.
- [21] Chen S, Ma K, Zheng Y. Med3d: transfer learning for 3d medical image analysis. arXiv: 1904.00625 2019 [Online] Available: <https://arxiv.org/abs/1904.00625>.
- [22] Irvin J, et al. Chexpert: a large chest radiograph dataset with uncertainty labels and expert comparison. In: arXiv:1901.07031; 2019 [online] Available: <https://arxiv.org/abs/1901.07031>.
- [23] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proc. IEEE conf. Comput. Vis. Pattern recognit(CVPR); 2016. p. 770–8.
- [24] Laurens VDM, Hinton G. Visualizing high-dimensional data using t-SNE. *J Mach Learn Res* 2008;9:2579–605.
- [25] Selvaraju RR, Cogswell M, Das A, et al. Grad-cam: visual explanations from deep networks via gradient-baassisted localization. In: Proc. IEEE int. Conf. Comput. Vis. (ICCV); 2017. p. 618–26.
- [26] Simonyan Karen, Zisserman Andrew. Very deep convolutional networks for large-scale image recognition. 2014. arXiv preprint arXiv:1409.1556.
- [27] Alexey Dosovitskiy. An image is worth 16x16 words: transformers for image recognition at scale. 2020. arXiv preprint arXiv: 2010.11929.
- [28] Chen Liang-Chieh, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation. Proceedings of the European conference on computer vision (ECCV). 2018.