Instance contrastive learning with dynamic weighted variance for small sample steel defect recognition

Yanyang Liang,[1] Jiacong Chen,[1] Wenlve Zhou,[1]
Ying Xu,[1,✉] Yikui Zhai,[1] Ruggero Donida Labati,[2]
Vincenzo Piuri,[2] and Fabio Scotti[2]
[1]*Department of Intelligent Manufacturing, Wuyi University, Jiangmen, Guangdong, China*
[2]*Departimento di Information, Universita, Degli Studi di Milano, via Celoria 18, Milano, Italy*
✉Email: xuying117@163.com

As an essential element in industrial steel, automatic defect recognition can guarantee the surface quality through focused supervised learning with ample labelled samples. However, defect recognition inevitably features with data-limiting characteristic under the influence of costly expert labelling. To address this problem, a novel framework, Instance Contrast (InCo), is proposed with the inspiration of contrastive learning. This framework consists of two streams. One with instance labels attributed to the unlabelled data in each batch for classification, which is called Batch Instance Discrimination (BID). The other with different enhanced samples embedding of the same image aggregated by a new function named dynamic weighted variance loss (DWV loss). Therefore, better semantic features can be learned by model due to the moderation of embedding distance between similar steel defect images. Experimental results on the NEU-CLS database validate that the proposed method achieves 89.86% classification accuracy with only fine-tuning on the 1:32 training data, outperforming other general contrastive learning methods.

Introduction: Defect classification is the main function of automated visual inspection (AVI) instruments, it is crucial to ensure the quality of steel surface. In [1] and [2], a convolution neural network (CNN) was employed as a classifier to identify the surface defects of flat steel, resulting in a favourable performance. Despite great progress has been made in defect recognition by previous work, their performance is still highly dependent on labelled samples. And much time and labour are required to label samples, which is not in line with the efficiency standard in industry. Thus, limited data on steel defects draws the attention of researchers, and various methods have been proposed to overcome these defects [3–5]. Recently, breakthroughs have been witnessed in contrastive learning for small sample classification tasks via shortening the embedding distance among similar samples and lengthening the embedding distance among different samples. Chen et al. [6] brought up SimCLR that fine-tuned on 1% of labels, surpassing the framework AlexNet with labels less than 100 × . However, current contrastive learning is not fully applicable to the field of steel defect, since the Intra-class samples of the steel defect images possess high similarity. Furthermore, for representation learning of steel defect images, the case of large differences existing between different samples of the same category may even bring a negative impact (Figure 1a).

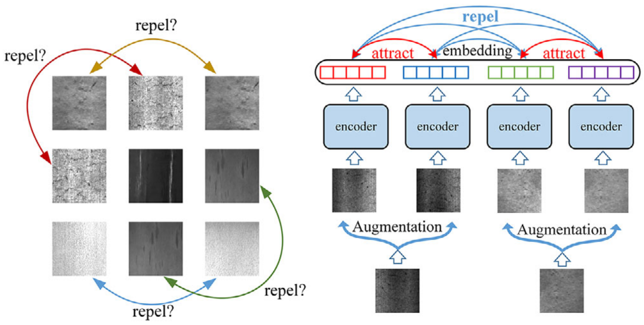

Fig. 1 *Visualisation of NEU-CLS dataset and strongly contrastive learning architecture. (a) sample from NEU-CLS database. (b) strong contrastive learning architecture*

In this Letter, to provide solution for the problems mentioned above, we proposed Instance Contrast (InCo), a new contrastive learning framework with the definition as weakly contrastive learning, whereas the current contrastive learning framework (Figure 1b) as strongly contrastive learning. The main contributions of this letter are four-fold: (1) A novel contrastive learning framework, InCo, is proposed, which can effectively surmount the effect of small sample in steel defect images. (2) The Batch Instance Discrimination (BID) is introduced to undermine the intensity of contrastive learning. (3) A new Dynamic-Weighted Variance loss (DWV loss) function is formulated for feature clustering to effectively adjust the distance between sample embedding. (4) Results certify that InCo performs better than strongly contrastive learning methods on the NEU-CLS [7] dataset.

Methodology: The method can be divided into two parts: Batch Instance Discrimination (BID) and the novel Dynamic-Weighted Variance loss (DWV loss). Figure 2 expounds the framework of this method.

In the “repel” part, the core of BID shares the similarity with supervised learning when assigning an instance label to each sample in each batch during the training stage of the model. Only by remembering the different in the label of the samples per epoch, can the model converge in a common training way. Therefore, K times augmentations are conducted to each sample in order to receive K batches of augmented data

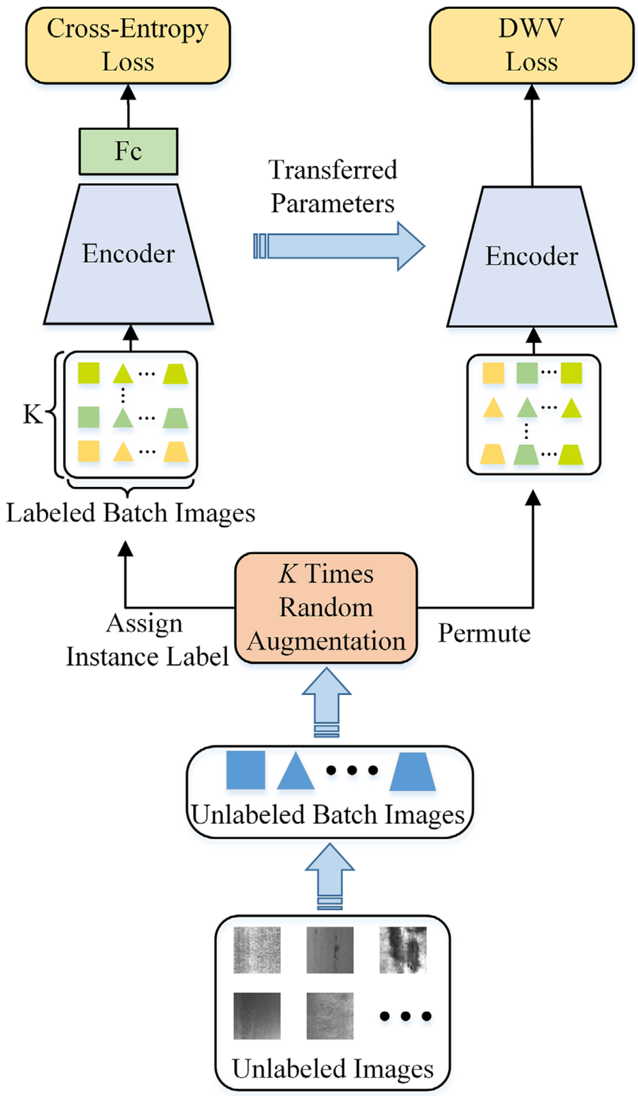

Fig. 2 *The pipeline of Instance Contrastive Learning approach*

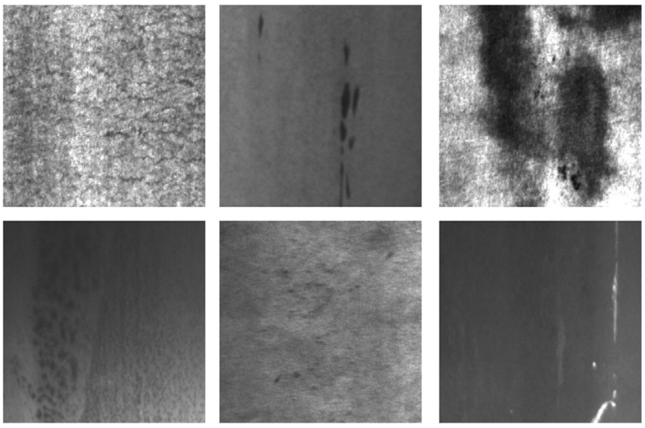

Fig. 3 *Examples of defect samples in the NEU-CLS. (a) crazing. (b) inclusion. (c) patches. (d) pitted surface. (e) rolled-in scale. (f) scratches*

for training. The Cross-Entropy loss is adopted to be an objective function guiding the BID implementation, expressed as:

$$L_{ce} = -\frac{1}{n}\sum_{i=1}^{n} \frac{\exp(w_i^T v/T)}{\sum_{j=1}^{n} \exp(w_j^T/T)} \tag{1}$$

where w_i refers to a weight for class i from the fully connected (FC) layer and T is a temperature coefficient that controls the distribution of representations.

In the "attract" part, general contrastive learning framework only conducts augmentation twice on per image with the resemblance calculation between two samples while in BID K-time calculation can be found. In order to aggregate K different enhanced samples of the same image, a new loss function called DWV loss is introduced and, written as:

$$L_{dwv} = \text{weight}(t)\frac{1}{nk}\sum_{i=1}^{n}\sum_{j=1}^{k}(z_i^k - \mu)^2 \tag{2}$$

where z_i^k represents an embedding from the global average pooling layer, μ is the mean value of z_i^k, t stands for the training epochs, and weight (t) denotes the hyperparameter used to adjust the DWV loss. The proper scheduling of the weight (t) plays an important role in the learning of the model. A large weight value may cause the model to collapse in the early stages of training. In contrast, the model can benefit a little from unlabelled data in a too small weight value. Drawing from [8], the weight (t) is defined as a gradually increasing function of time t, and the value of weight will change with T_1 and T_2, as followed:

$$\text{weight}(t) = \max\left\{a_f \min\left(\frac{t-T_1}{T_2-T_1}, 1\right), 0\right\} \tag{3}$$

Experiments: In the experiments, for the augmentation of InCo, crop with resize, brightness, contrast, random horizontal flip, and random vertical flip is performed with the augmentation times K set to 5. ResNet18, as the ConvNet, is trained at batch size 64 for 200 epochs, which also incorporates Cross-Entropy loss with a temperature coefficient T of 1.0 and DWV loss with $T_1 = 30$, $T_2 = 150$ and $a_f = 1.0$. Adam is adopted as the optimizer, setting the original learning rate to 0.3. All experiments are performed in an Ubuntu 18.04 operating system using Intel Core i7-9800X CPU. The electronic equipment is configured with NVIDIA GTX 2080 with 12GB memory and implemented by the Pytorch framework. InCo is tested on NEU-CLS, a surface defect benchmark dataset containing six common steel surface defects, namely crazing (Cr), inclusion (In), patches (Pa), pitted surface (PS), rolled-in scale (RS), and scratches (Sc). There are 300 defect images with a size of 200 $\times$ 200 in each class. 60% of all the samples are randomly selected as the training samples and 40% as the test samples, some of which can be referred to Figure 3.

To validate the effectiveness of InCo on a small sample classification, five small sample databases are created by randomly selecting proportionally (1:32, 1:16, 1:8, 1:4, 1:2) on each category of the original training set. Fine-tuned, as a transfer learning technology, and linear evaluation, as a recent unsupervised feature learning evaluation method, are both selected as the measurement methods in this work. In unsupervised learning, unlabelled data is used to learn visual representations and fine-tune the few-labelled data model, while linear evaluation devotes to controlling the weights of the feature extractors only at the FC layer. The Softmax classifier is utilized for fine-tuning or linear evaluation, the model is trained for 200 stages, setting the learning rate to 0.03.

Table 1. Comparison with contrastive learning methods on NEU-CLS small sample database

Method	1:32	1:16	1:8	1:4	1:2
Fine-tuned:					
SimCLR[6]	82.50	83.75	85.69	87.92	93.61
MoCov2[9]	63.75	73.47	78.61	84.17	91.81
SimSiam[10]	69.72	73.89	82.92	91.11	95.56
InCo-without-DWV	62.92	68.86	78.06	89.58	93.75
InCo-with-DWV	89.86	92.63	93.89	95.42	98.06
Linear evaluation:					
SimCLR[6]	65.56	67.64	72.78	84.86	88.75
MoCov2[9]	89.03	90.64	91.81	93.75	97.78
SimSiam[10]	75.83	77.91	83.47	90.56	94.72
InCo-without-DWV	60.69	65.56	77.08	87.08	89.31
InCo-with-DWV	87.78	91.25	93.75	94.41	97.92

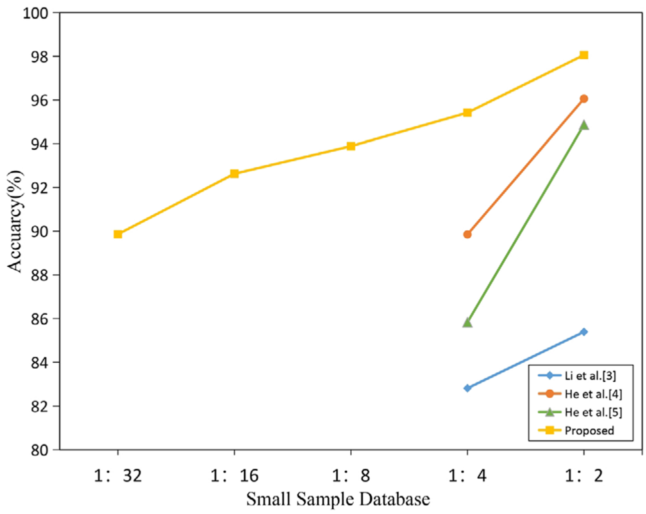

Fig. 4 *Classification accuracy on small sample database with various methods*

Comparisons have been conducted between the proposed one and other mainstream contrastive learning methods, where they will be trained with the augmentations raised in this letter at the unsupervised feature learning stage. Table 1 presents the result of the experiments. Based on NEU-CLS, it can be found that InCo outperforms the others [6, 9, 10], that is, other learning methods not absorb valuable semantic features, which certifies the correctness of our conjecture. Moreover, DWV loss features have a positive effect on InCo representation learning.

To compare our method even further, the related methods that use CNN networks for defect classification in small samples database are adopted. Records of comparing different methods are detailed in Figure 4. The results prove that the method possesses higher applicability for small sample defect classification, and the classification accuracy of training with 1:32 data is equivalent to the recognition rate of other methods training with 1:4 data.

Discussion: Verifiable by the experimental above, the performance of InCo is better than that of other mainstream contrastive learning approaches on steel defect data. The reason is that there exist similar intra-class samples in steel defect data, while mainstream contrastive learning requires the model to push each instance sample apart as much as possible, ignoring the potential relationships among samples and destroying the semantic consistency of similar intra-class samples. InCo weakens the discriminative strength among samples by using cross-entropy and DWV loss, which indirectly minimize the embedding distance of similar intra-class samples and improves the quality of representation learning.

Conclusion: In this Letter, a new contrastive learning framework namely Instance Contrast have presented for unsupervised visual representation learning of steel defect. This method can effectively address the problem that strongly contrastive learning damages the learned semantic features, leading to satisfactory results for unsupervised learning in NEU-CLS dataset. Through experiments, the proposed method can maintain accuracy even when provided with a small amount of training data. Therefore, it is believed that this framework can be deployed on automated visual inspection (AVI) instruments, which bears great significance in improving the efficiency of the industry. In the future, the efficiency improvement of InCo, especially at the time of algorithm training, will be the main focus of our work.

Acknowledgments: This work was supported by Key Research Projects for Universities of Guangdong Provincial Education Department (No. 2018KZDXM073, No. 2019KZDZX1017, No. 2020ZDZX3031), Guangdong Basic and Applied Basic Research Foundation (No. 2019A1515010716, No. 2021A1515011576) and Guangdong, Hong Kong, Macao and the Greater Bay Area International Science and Technology Innovation Cooperation Project (No. 2021A0505030080).

Conflict of interest statement: We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled, "Instance contrastive learning with dynamic weighted variance for small sample steel defect recognition".

Data availability statement: The raw/processed data required to reproduce these findings cannot be shared at this time as the data also forms part of an ongoing study.

Received: *8 July 2021* Accepted: *17 October 2021*
doi: 10.1049/ell2.12361

References

1 Ren, R., Hung, T., Tan, K.C., et al.: A generic deep-learning-based approach for automated surface inspection. *IEEE Trans. Cybern.* **48**(3), 929–940 (2018) https://doi.org/10.1109/TCYB.2017.2668395

2 Masci, J., Meier, U., Fricout, G., et al.: Multi-scale pyramidal pooling network for generic steel defect classification. In: International Joint Conference on Neural (IJCNN), Dallas, TX, USA, pp. 1–8 (2013) https://doi.org/10.1109/IJCNN.2013.6706920

3 Yi, L., Li, G., Jiang, M.: An end-to-end steel strip surface defects recognition system based on convolutional neural networks. *Steel Res. Int.* **88**(2), 176–187 (2017) https://doi.org/10.1002/srin.201600068

4 He, Y., Song, K., Dong, H., et al.: Semi-supervised defect classification of steel surface based on multi-training and generative adversarial network. *Opt. Laser Eng.* **122**(2), 294–302 (2019) https://doi.org/10.1016/j.optlaseng.2019.06.020

5 He, D., Xu, K., Zhou, P., et al.: Surface defect classification of steels with a new semi-supervised learning method. *Opt. Laser Eng.* **117**(2), 40–48 (2019) https://doi.org/10.1016/j.optlaseng.2019.01.-11

6 Chen, T., Kornblith, S., Norouzi, M., et al.: A simple framework for contrastive learning of visual representations. In: Proceedings of the International Conference on Machine Learning (ICML), Vienna, Austria, pp. 1597–1607 (2020)

7 Song, K., Yan, Y.: A noise robust method based on completed local binary patterns for hot-rolled steel strip surface defects. *Appl. Surf. Sci.* **285**(21), 858–864 (2013) https://doi.org/10.1016/j.apsusc.201-3.09002

8 Lee, D.: Pseudo-Label: the simple and efficient semi-supervised learning method for deep neural networks. In: Proceedings of the International Conference on Machine Learning (ICML), Atlanta, GA, USA, pp. 1–6 (2013)

9 Jain, R., Fan, H., Girshick, R., He, K.: Improved baselines with momentum contrastive learning. *arXiv preprint arXiv:2003.04297* (2020)

10 Chen, X, He, K.: Exploring simple Siamese representation learning. In: Proceedings of Computer Vision and Pattern Recognition (CVPR), Nashville, pp. 15750–15758 (2021)