PreAugNet: Improve Data Augmentation for Industrial Defect Classification with Small-scale Training Data

IsackFarady[1,2], Chih-Yang Lin[2*] and Ming-Ching Chang[3]

[1]Electrical Engineering, Universitas Mercubuana, Indonesia.
[2]Electrical Engineering, Yuan Ze University, Taiwan.
[3]Computer Science, University at Albany, USA.

*Corresponding author(s). E-mail(s): andrewlin@saturn.yzu.edu.tw;
Contributing authors: isack.farady@mercubuana.ac.id; mchang2@albany.edu;

Abstract

With the prevalence of deep learning and convolutional neural network (CNN), data augmentation is widely used for enriching training samples to gain model training improvement. Data augmentation is important when training samples are scarce. In this work, we focus on improving data augmentation for training an industrial steel surface defect classification network, where the performance is largely depending on the availability of high-quality training samples. It is very difficult to find a sufficiently large dataset for this application in real-world settings. When it comes to synthetic data augmentation, the performance is often degraded by incorrect class labels, and a large effort is required to generate high-quality samples. In this paper, we introduce a novel off-line pre-augmentation network (PreAugNet) which acts as a class boundary classifier that can effectively screen the quality of the augmented samples and improve image augmentation. This PreAugNet can generate augmented samples and update decision boundary via an independent classifier. New samples are automatically distributed and combined with the original data for training the target network. We show that these new augmentation samples can improve classification without changing the target network architecture. We evaluate this method for steel surface defect inspection using three real-world datasets: AOI steel defect dataset, MT, and NEU datasets.

The proposed method significantly increases the accuracy by 3.3% (AOI dataset), 6.25% (MT dataset) and 2.1% (NEU dataset), respectively

Keywords: data augmentation, synthetic sample generation, CNN, surface defect classification, decision boundary, PreAugNet

1 Introduction

In the era of deep learning, Convolutional Neural Networks (CNN) as typical feed forward neural networks have performed remarkably in the various computer vision systems including image classification [1–4] object detection [5–7] semantic segmentation [8, 9], etc. One of the most challenging supervised learning applications of CNN in the industry is the product defect recognition that have been extensively studied. In order to train a deep neural network model for real-world industrial use, probably the first immediate task is to collect sufficient labeled data. Without high-quality training data, overfitting will mostly occur, which causes the learned model to be highly biased to the seen samples but not be able to generalize against unseen data. It is well-known that regularization techniques can alleviate model overfitting, including the extended techniques of [10–12] and batch normalization [2]. Various heuristic techniques such as weight decay and early training stopping can reduce overfitting by penalizing parameter norms. Despite the practical values of these heuristics, the training of large network models for complex real-world industrial applications still demands a large amount of high-quality data.

Data augmentation is an effective approach to battle model overfitting [13]. Data augmentation is the process of supplementing and enriching available data for better generalization during training. For most computer vision problems, image transformations such as rotating, cropping, scaling, noise perturbation, or color adjusting [13] are popular means to substantially improve data amount [14]. When dealing with natural images, rotating, flipping, scaling transformations are *de facto* approaches used during training. Unfortunately, not all transformations are useful for every dataset or problem. For example, all categories in CIFAR10, CIFAR100 and ImageNet datasets [2] should be invariant to horizontal flips, since the mirror of an object is typically visually valid (e.g., a mirrored car is still a good training sample). However, not all image transformations are valid for problems such as character recognition [15], where non-existing symbols or symbol label change after transformation (e.g., a flipped '6' becomes '9') can harm model training.

In the process of augmentation paradigm, there are mainly three ways in which augmentation techniques can be applied, namely off-line augmentation, on-line augmentation, and hybrid methods. Off-line augmentation user has access to screening the augmented results but needs to concern about the quality of new samples. On other hand, on-line augmentation provides virtually infinite samples during training, however without ground truth for validation.

Numerous image recognition works apply off-line augmentation by producing synthetic images [16, 17] to effectively improve model training and alleviate over-fitting. However, the new synthetic samples are still generated from the modeling of existing samples, thus they are typically not sufficiently diverse. It is common that incorrectly augmented transformations can induce features far-away from the original sample, which harms model training. Since there is no easy way to find out but to check model effectiveness at the end of training, the evaluation of data augmentation can be very time consuming in the real-world setting. In this work, our major goal is to develop an efficient method that can provide insight on the selection process of off-line data augmentation sample generation, such that more diverse and representative samples can be generated to improve target model training.

Specifically, we develop an off-line data augmentation optimization approach that can effectively improve the screening of augmented samples to boost model training. We choose the industrial surface defect classification as the targeted application for evaluation. We construct an independent, lightweight data augmentation network named **Pre-Augmentation Network (PreAugNet)** that consists of a data augmentation generator, a feature extractor, and a data management module. Motivated by the idea of effective classification of Support Vector Machine (SVM) in determining decision boundary [18], We design a SVM classifier that predicts the label of a generated sample based on its extracted features. This process iterates in updating the new samples regarding the SVM decision boundary being modeled that are related to the data augmentation transformation process.

- We propose a lightweight PreAugNet that improve the off-line data augmentation for training a defect classification model. The pre-augmentation network learns to extract feature from input sample images and produce proper transformations to generate new sample images for data augmentation.
- We design a SVM decision boundary analysis to screen and iteratively update the samples produced from the PreAugNet to ensure the suitability of the transformed samples for target network training. We show how the iterative estimation and updating of class decision boundaries can be very effectively in screening and producing diverse augmentation samples that generalize better.
- Extensive experiments are performed to evaluate the performance of the PreAugNet against other state-of-the-art online and offline data augmentation methods. Specifically, we use ResNet [1] as the target network that is trained on three real-world datasets, namely AOI, MT and NEU datasets for steel surface defect inspection. PreAugNet significantly increases the classification accuracy by 3.3% (AOI), 6.25% (MT) and 2.1% (NEU), respectively.

The rest of this paper is organized as follows: Section 2 discusses related works. Section 3 introduces the principle of pre-augmentation network, augmentation generator, and the SVM decision boundary update process.

Section 4 describes experimental results and performance analysis. Section 5 provides discussions and the conclusion.

2 Related Work

Data augmentation. Extensive works on real-time image augmentation offer massive efforts in image classification [13, 19–22]. For creating additional training samples from existing data, [23] shows the benefit of creating synthetic samples via combining the two approaches of data warping and synthetic over-sampling. Data warping methods generate samples through transformations that are applied in the data-space, while synthetic over-sampling creates samples according to the feature-space. The Generative Adversarial Networks (GAN) [24] are widely used in producing new realistic samples of certain data or class. By training using adversarial examples monitored by the discriminator, the generator of GAN can synthesize realistic-looking images that are sufficiently different from the original images [25, 26].

Industrial Defect Inspection. Steel surface defect inspection has received increased attention for ensuring quality control of industrial products. Surface defect detection is usually performed against complex industrial scenarios, which ends up as a challenging problem with hard usage constraints. Surface defects are the main cause of low-quality steel products. Steel surface defect recognition and classification approaches have improved significantly since the debut of deep learning with many advantages in the past decades [27–30]. In recent years, exploring the benefit of machine learning algorithms emerges when CNN features have been successfully integrated with basic superior classifier such as the Support Vector Machine (SVM) [18]. The works [31–33] can classify both linear and nonlinear problems with SVM kernel functions. In [27, 34], image data are successfully enhanced using CNN to produce better classification results. Many developments from these hybrid methods emerge in cases when the amount of available data is limited, which is particularly true for industrial defect inspection. As shown in [35–37], the CNN structure has been well-suited to deal with non-natural images with quality and scalability issues.

Data Augmentation for Defect Inspection. Many applications regarding industrial image processing face barriers of severe data scarcity. The works of [38, 39] overcome the shortage of defective samples by adopting GAN for effective data augmentation. In the discriminative training of GAN [40], the computational cost of generator increases, which tends to overfit to real data where data augmentation should be avoided. Regarding applying GAN data augmentation for industrial defect datasets, the GAN generator can learn a complex distribution from the limited available dataset. However, how useful the synthetic samples in regarding model training is questionable. GANs might not be able to cover the entire diversity of defect types, as the available defect samples in the first place can be already very scarce.

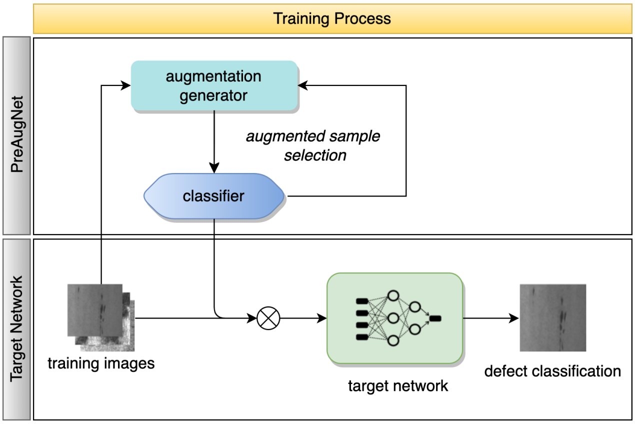

Fig. 1 Overview of the proposed off-line PreAugNet to improve data augmentation for the training of an industrial defect classification network.

In contrast to the previous approaches, in this work we designed an effective data augmentation approach based on feature enhancement that goes hand-in-hand with their class decision boundary that are crucial for training the target network. Our approach leverages CNN features that will be refined in a SVM decision boundary representation, such that new data augmentation samples can be created via data-driven learning and iteratively update. This way, our approach can create diverse but representative samples, that the target network has mostly not seen before to boost training its performance.

3 Methodology

We start with the main idea of introducing a Pre-Augmentation Network (PreAugNet) to perform off-line data augmentation that can improve the training of the target network model for industrial defect recognition. We first describe the PreAugNet with detailed module design and then explain the data flow regarding the splitting, sampling and colleting of the augmented samples inside the PreAugNet in Section 3.1. We describe how the PreAugNet updates the sample decision boundary during the sample search process in Section 3.2. Finally, we discuss the target network settings and how the augmented samples are added to improve its training in Section 3.3.

3.1 Off-line Pre-Augmentation Network

Figure 1 shows the architecture of the proposed PreAugNet, which are attached to a target network for training with data augmentation. PreAugNet produces new transformed samples via an augmentation generator, where the samples are screened to ensure that they gain representative and diverse features that are sufficiently different from the original samples. The generated samples are merged with the original samples for target network training.

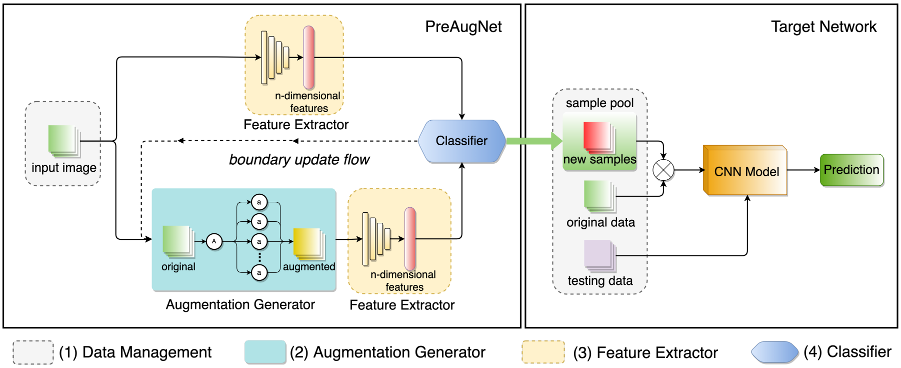

Fig. 2 The PreAugNet data augmentation as an integrated pipeline for target network training.

The PreAugNet consists of four parts: (1) a data management module, (2) the augmentation generator, (3) the feature extractor and (4) the augmentation sample classifier. The data management module manages data distribution including data splitting, collection, sample dropping and merging for target network training. The data management module controls the data flow inside the PreAugNet and data pre-processing of the target network. It also plays an important role in maintaining a balance between the synthetic samples vs. the desired amount of total training samples. The augmentation generator produces a diverse set of new transformed images/samples, which will be screened and picked in the next step. The feature extractor consists of a standard CNN that extracts high-dimensional features from the original images. The augmentation sample classifier is a SVM that performs classification based on the extracted features and estimates the decision boundary to gauge the quality and suitability of the generated samples. Figure 2 illustrates the proposed PreAugNet data augmentation pipeline jointly with the training of the target model.

The augmentation generator $\mathbb{G}$ in Figure 2 performs image transformations to the original samples to generate images in different sizes and shapes. Both affine transformations in the spatial domain and color intensity adjustments in the pixel domain are incorporated for data augmentation selection. Those image transformations are effective means for data augmentation as they align with variations in the physical world. In reality, the same defect can occur at various sizes and locations on the steel surface with different illumination and viewing conditions. This way, the augmentation generator can effectively generate realistic new samples that are suitable for model training. Surface defect images often exhibit very few amounts of information as features. Thus, a robust feature extractor is essential for acquiring representative feature vectors from the defective samples. We use Inception-v3 [41]to extract features, with the same configuration for both the original and augmented images. The deep structure of Inception-v3 at last fully connected

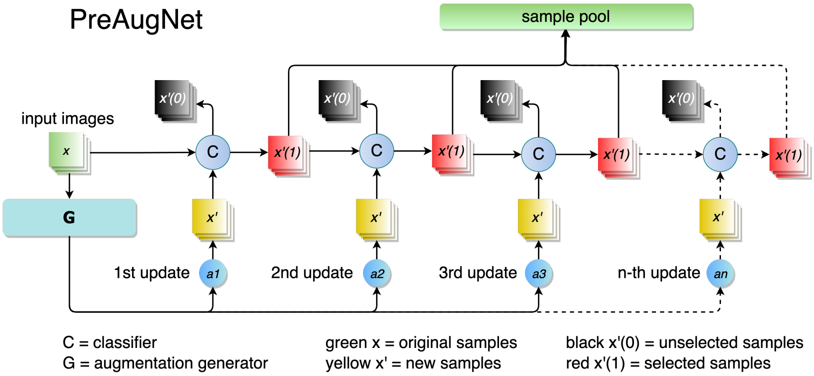

Fig. 3 The boundary update process in the pre-augmentation network. The original image x is transformed to x' via augmentation generator $\mathbb{G}$ which contains several augmentation methods {a1, a2, a3 ... an}. In the next iteration, classifier $\mathbb{C}$ updates the sample decision boundary according to new augmented images and performs the class prediction of the new input samples. The output of the $\mathbb{C}$ specifies a new sample x' to be collected into the sample pool or discarded for all classes. The red x' represents the correct class and black x' is the discarded sample. The update process inside PreAugNet continues until the predetermined α ratio of the target network matches.

layer retrieves 2048-dimensional features, which are fed into the classifier for boundary classification.

Figure 3 describes the augmented sample update process within the PreAugNet. Consider a dataset $X = \{x_i, y_i\}_{i=1}^{N}$, where x_i is an input image, y_i the associated class label, N the number of samples. Denote X' for the set of newly generated images, and α for the percentile of new samples w.r.t. the number of original samples. In this setup, the augmentation generator $\mathbb{G}$ takes input dataset X and produces augmented samples X', namely, $\mathbb{G} = X \to X'$. The augmentation process typically consists of multiple operations such as hue saturation adjustments, adding various noise types (random, multiplicative and additive Gaussian), removing high frequency component via jpeg compression, randomly drop channel of input image and image blurring, i.e., $\mathbb{G} = \{a_1, a_2, \ldots a_n\}$. These image transformations are important as they provide the source of variability for image augmentation.

As shown in Figure 3, the augmentation generator $\mathbb{G}$ is able to produce a large amount of images. However, not all the output of $\mathbb{G}$ are useful for training the target network. In order to control the distribution and variation of new generated samples, we apply a selection process to collect only "good" samples with the concept of the decision boundary of a SVM classifier $\mathbb{C}$. In the first iteration of the pre-augmentation network, data management module performs a "drop-select" function based on a ratio α. The ratio α controls the targeted number of new samples selected as correct results from the classifier $\mathbb{C}$. While the rest of mis-classified samples will be dropped, the data management module starts to recalculate the minimum number of new samples x' to fulfill the needs of the target network. In this phase, the next iteration starts and will automatically repeat the process until the specified condition is reached.

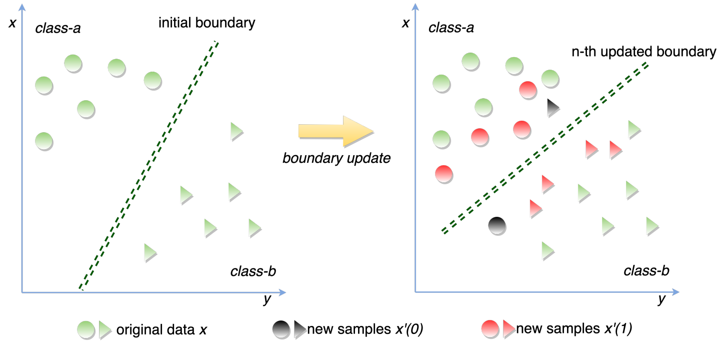

Fig. 4 Illustration of the boundary update of two class data observations for new generated samples x' (red and black) under 2-dimensional and linear condition. The dash line represents the boundary of the classifier $\mathbb{C}$ where the boundary line dynamically moves according to new samples input for each iteration updates.

3.2 The Decision Boundary Update Process for Augmentation Sample Screening

We construct a feature similarity measure by a classifier $\mathbb{C}$, where, in the case of given a high dimensional feature from X and X' samples. We aim to obtain an optimal boundary separation between X and X' for each class by learning the similarity degree and relationship among the features. We formulate the problem of searching the samples on the predicted result of the classifier $\mathbb{C} = \{x', y'_i\}_{i=1}^{N}$ with SVM. The basic idea is mapping the input feature vector into a high-dimensional space and generating a maximal distance of separation boundary.

As shown in Figure 4, in the process of searching the boundaries, the SVM classifier learns the n-dimensional features (x, y) from both the training set and new samples in feature-space and calculate the maximum boundary margin between the new observation samples of the class label in one iteration. When pre-augmentation network updates this process for n-iteration times, at the same time classifier automatically produces another correct sample with their respected labels. As illustrated in Figure 4, the boundary line of SVM will adjust to new generated samples and find the optimum with the input samples. In this condition, classifier simply marks the output samples x' as 1 (red) and 0 (black) for the correct samples and mis-classified samples respectively (as illustrated in Figure 3). In other words, the classifier $\mathbb{C}$ not only provides the class label of x' but also the mis-classified position of x' for the new sample in y'_i classes. Thus, our data management module performs the *"drop-select"* function on the classifier results $\mathbb{C}$ for next iteration where all the 1's (red) output will only be selected and stored into sample pool.

3.3 Target Network

The target network is the network module that demands data augmentation. Experiments are designed to evaluate how the various ways of data augmentations affect the training of this target network. For industrial defect recognition, we use the ResNet [1] model as our target network. Specifically, we use the same configuration as the original ResNet18 structure, including the loss function, batch normalization, and optimizer.

In the baseline method, target network is trained without using any augmentations at pre-processing and testing stage. The only change is in the adjustment of the input image size which is adjusted to the original model implementation. In order to control the number of augmentation samples, we define the α ratio where the ratio is cumulative augmented image in a single training process. We set the α ratio in percent of original image and limit the ratio to no more than the original data. This ratio concept is applied for both online and off-line pre-augmentations.

4 Experimental Results

4.1 Dataset and Training Details

We perform experiments on three challenging, real-world industrial steel surface defect datasets to evaluate the proposed method: (1) the Automatic Optical Inspection (AOI) steel defect dataset[1] , (2) the Magnetic Tile (MT) surface defect dataset [42], and (3) the NEU defect dataset [43]. The AOI dataset is a private dataset that contains five types of steel surface defect: *void defect, horizontal defect, vertical defect, edge defect*, and *particle defect*. The MT and NEU datasets are well-known public datasets and widely used for defect classification and detection. The MT dataset contains five type of defect: *blowhole, break, crack, fray* and *uneven*. The NEU dataset contains six defect classes: *crazing, inclusion, patches, pitted surface, rolled-in scale* and *scratches*. The red boxes of Figures 5 highlight steel defects in these datasets, which is visually quite similar to the steel background. Images are grayscale and the information provided by the defect samples is typically scarce.

Table 1 summarizes the number of defect images of the three datasets. Note that the number of defect samples are extremely low when compared with other image classification datasets such as ImageNet or COCO for different applications. The three datasets show two challenging conditions in the industrial use case. *First*, all datasets contain small-scale training data and lack of surface defect image representations. For instance, AOI defect dataset consists of five classes of defect with a total image for whole raw data is 1854 images and similarly in NEU dataset also consists of 1800 images. The MT dataset consists of 1344 images where only 392 defect images available. *Secondly*, the imbalance data of the three datasets cause additional challenging. Collecting a specific type of industrial defect sample is not an easy task, since the same

[1] https://aidea-web.tw/topic/701e1e79-84ff-49a5-86ee-a7f01c24c6f7

Table 1 Details and data statistics of the three industrial inspection datasets used for experimental evaluation.

	AOI	**MT**	**NEU**
Number of defect images	1854	392	1800
Number of defect classes	5	5	6
Defect image distribution	class-1 (492) class-2 (100) class-3 (378) class-4 (240) class-5 (644)	class-1 (115) class-2 (85) class-3 (57) class-4 (32) class-5 (103)	class-1 (300) class-2 (300) class-3 (300) class-4 (300) class-5 (300) class-6 (300)

type of defect does not frequently appear in a production line. It reflects in AOI and MT datasets where distribution data among the classes is not in the same average amount of images. In the case of AOI and MT dataset, imbalanced data distribution has obviously become a problem for defect recognition and detection. However, the NEU dataset shows another real problem that all classes have the same low number of images (300 defect images/class) and equally distributed.

We implement the proposed methods in PyTorch. Experiments are performed on a workstation with Linux and NVIDIA RTX 2080i GPU. Baseline experiments are performed with the original settings, where the AOI, MT and NEU datasets are directly processed by the baseline model without augmentation. In the next round, online augmentations take part for training based on α ratio to transform images in one single online augmentation method. The online augmentation performs random operation of transformation to original images during training process. On the other hand, off-line pre-augmentation network runs augmentation operation from the `Albumentation` [44] library to produce all samples inside the augmentation generator A that running separately from the target network.

Evaluation Metrics. The effectiveness of our proposed method on target network is examined in terms of final prediction accuracy for unseen test images. Compared with other classification methods, accuracy (%) is used to performance the evaluation of the prediction result. The accuracy is defined as a ratio of number of test images correctly classified to the number of all test images in target network.

4.2 Evaluation Results

We next present experimental results of the PreAugNet with ResNet-18 target network on the AOI, MT and NEU defect inspection datasets. Since the pre-augmentation network and target network are independent, we can separately train the baseline and pre-augmentation network. We prepare the pre-augmentation network to perform searching samples in parallel, where the data management automatically collects the samples. The number of the correct samples from the classifier will be added to target network before performing the off-line augmentation.

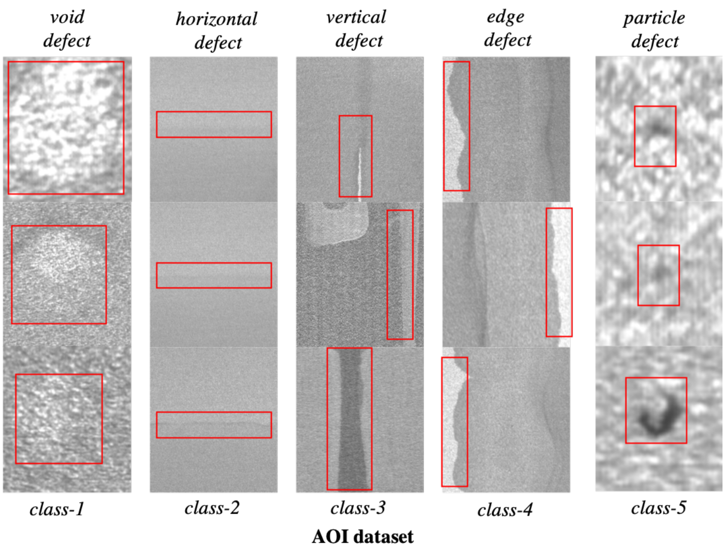

Fig. 5 Sample images of each class from the AOI industrial inspection dataset. Red boxes highlight steel defects.

4.2.1 Results on AOI Dataset

The AOI dataset consists of 1,854 images from five type of defects. As shown in Table 1, horizontal defect class-2 (100 images) and edge effect class-4 (240 images) have significant differences in terms of image distributions. In the case of AOI defect dataset, we conduct two scenarios to prove our off-line pre-augmentation network. The *First* scenario, we tackle the imbalance problem by adding more samples for the lowest class. At the initial stage of experiments, we adjust images from class-2 as the main source for generating new samples. Data management in the pre-augmentation network distributes only image from class-2 to the augmentation generator. Augmentation generator specifically generates new samples from class-2 with ratio $\alpha \leq$ N of original images. Since class-2 consists of 100 images, pre-augmentation network updates the searching process 7 times before reach the maximum number of new samples. In the next phase, we try to generate another sample from another lower class. For class-4 our pre-augmentation network needs 5 iterations update to produce similar number of samples. We present the result of pre-augmentation from the two classes in Table 2. We assume that the number of updates process heavily depends on the amount of the original images. The more resources we have the faster searching process will be. As we can see in Table 2 adding new samples for imbalance class data improves the accuracy of the target network.

The goal of the *second* scenario for AOI dataset is aiming to balance all the classes. This is the typical pre-augmentation scenario where the goal of this approach is to generalize data distribution among the classes. In this case, we primarily generate more samples with more augmentation methods to

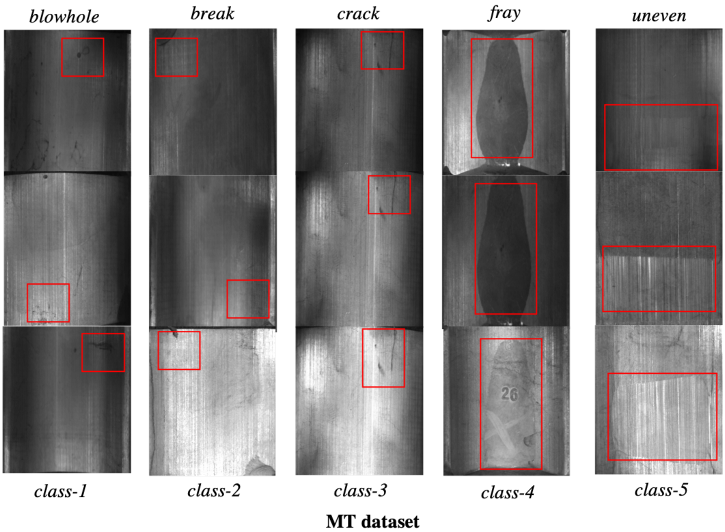

Fig. 6 Sample images of each class from the MT industrial inspection dataset. Red boxes highlight steel defects.

Table 2 Defect classification accuracy results (%) on AOI dataset with different approaches for data augmentation. All pre-augmentation methods boost the prediction accuracy including the method of adding sample to imbalance and low distribution data (class-2 and class-4).

Method	Accuracy (%)	Improvement (Δ)
Baseline	95.10	-
General on-line augmentation	95.60	0.5
PreAugNet (class-2)	97.10	2.0
PreAugNet (class-2 and class-4)	97.70	2.6
PreAugNet (all class)	98.20	3.1

the lower class and randomly set less transformation methods in higher class images. As the result, all classes share the same number of samples in the target network. The effect of this approach is the dataset will share the same average number of images. In practical, we carefully train pre-augmentation network for all classes and set the limit of α in data management module to match with the target sample distributions. With this scenario, the new samples from our pre-augmentation network successfully achieved better accuracy about 3.1% of AOI dataset compared to baseline.

4.2.2 Results on the MT Dataset

MT dataset is steel dataset with 5 defect classes, which the defect types are very close to the background. Since MT dataset has very small number of defect images for all classes , then pre-augmentation network will be directed to produce more samples from all class distributions. The augmentation generator produces more samples of color transformations method from the class

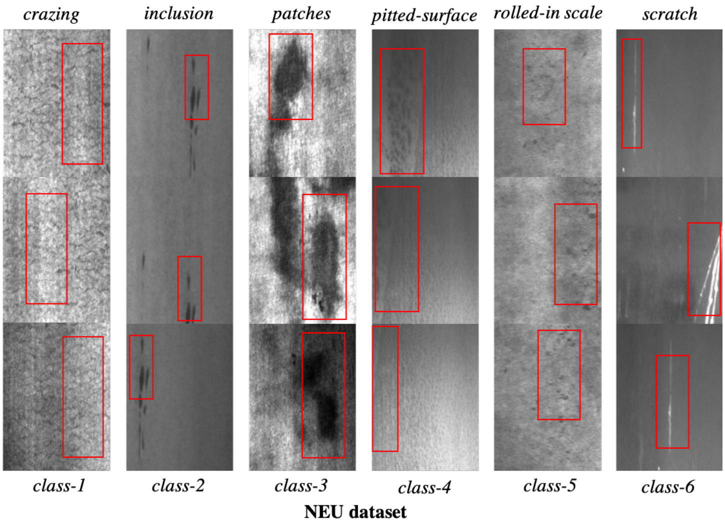

Fig. 7 Sample images of each class from the NEU industrial inspection dataset. Red boxes highlight steel defects.

Table 3 Comparison test accuracy (%) on MT dataset across different data augmentation methods.

Method	**Accuracy (%)**	**Improvement (Δ)**
Baseline	92.50	-
General on-line augmentation	96.20	3.70
PreAugNet	97.50	5.0

with lower number to class with higher number of images. We train the pre-augmentation network according to α ratio for all class distributions where we set α from 0.1 to 0.9 of the original image distributions.

In the target model prediction results as presented in Table 3, after adding samples from pre-augmentation network the final accuracy increased about 5% at the maximum α ratio (0.7). Despite that on-line augmentation yielded the highest accuracy about 3.7% of the baseline, Pre-augmentation consistently matches or outperforms baseline with alteration to all list of α ratios employed. In the scenario with same α ratio for all class in Figure 5, we also found that the new samples from pre-augmentation network surpassed the baseline after $\alpha = 0.2$ and achieve better accuracy at higher ratio. It seems that relatively small number of samples in the MT training dataset are not generalized well but with larger ratio and more samples added, pre-augmentation is become more robust across general on-line augmentation.

Table 4 Comparison test accuracy (%) on NEU dataset across different data augmentation methods. Our pre-augmentation network is trained across all classes and combined with on-line augmentation surprisingly managed to outperform synthetic data augmentation with GAN.

Method	Accuracy (%)	Improvement (Δ)
Baseline	97.40	-
General on-line augmentation	98.40	1.0
PreAugNet	**99.20**	1.80
Synthetic data augmentation [39]	99.11	-

4.2.3 Results on the NEU Dataset

Because NEU dataset is composed of equally distributed images for all classes, the off-line Pre-augmentation strategy for NEU dataset may differ substantially from AOI and MT dataset. In NEU dataset, our pre-augmentation network is focused on generating new samples by increasing the number of samples for all classes equally where we set the same initial α ratio for all classes during the pre-augmentation training. The pre-augmentation network simply generates new samples for all classes in the same manner for all α ratios. In details, the augmentation generator produces new samples in several stages by determining the types and number of transformations accordingly that it requires several updates in the process of collecting new samples in sample pool. That means, even though during the process of updating the new samples in classifier the number of iterations required is not the same for each class, but at the end of searching process all classes will get the same number of new samples.

Our testing result are shown in Table 4. As can be seen from the results, PreAugNet improves the overall accuracy over the general augmentation and synthetic GAN method [39]. Secondly, applying Pre-augmentation with α ratio 0.8 achieves the highest accuracy of NEU dataset about 1.8% compare to baseline.

4.2.4 Experiment on the Combination of Augmentation Methods

This section evaluates whether the off-line pre-augmentation network and on-line augmentation combined improves the prediction result. In combination augmentation mechanism, we randomly perform augmentation on the new samples and original samples during training. We re-train the target network for every 10% additional samples and capture the highest accuracy. We present several different α ratios for two different augmentation approaches: off-line pre-augmentation and general on-line augmentation method. Figure 8 demonstrates how the increasing number of samples affect the accuracy of target network for all datasets. The accuracy of target network constantly matches or improves from baseline and general augmentation. The performance of off-line pre-augmentation and combination method are particularly good on the AOI dataset, the improvement occurred in the addition of new samples starting from the small ratio. In other words, this results show that our method can indeed enlarge the defect samples in target network to produce better accuracy. The new samples from pre-augmentation network are more robust than default random on-line augmentation. This phenomenon is in line with our assumption that the pre-augmentation network only distributes samples that have been correctly selected based on the boundary in the classifier so that the new generated samples are more robust and useful for the target network.

Table 5 Comparison test accuracy (%) across all datasets with combination of general augmentation and PreAugNet methods.

	AOI		**MT**		**NEU**	
Method	Acc. (%)	Improv. (Δ)	Acc. (%)	Improv. (Δ)	Acc. (%)	Improv. (Δ)
Baseline	95.10	-	92.50	-	97.40	-
General augmentation	95.60	0.5	96.20	3.70	98.40	1.0
PreAugNet (combination)	**98.40**	**3.3**	**98.75**	**6.25**	**99.50**	**2.10**

Note: Acc. : Accuracy ; Improv. : Improvement

In Figure 8, it can be seen that not all pre-augmentation combination produces better results compare to single on-line augmentation. Result on MT dataset, the combination methods produce unstable accuracy at higher α ratio. We found that combination method produces lower than baseline at α ratio $(0.3, 0.4)$. We assume this phenomenon occurs due to lack of data in sample pool so that some samples forwarded from pre-augmentation are identical. We also found that if the original class data was too small, the pre-augmentation network required more updates to reach the expected ratio than other class. These multiple updates affect the searching time on our pre-augmentation network.

4.3 Limitations

We next discuss limitations regarding our method. First, we note that the arbitrary image transformation is not preferred in the augmentation searching process, since the augmentation generator need to produce a lot more samples before an acceptable sample can be found. Likewise, the time consuming for searching boundary is heavily depending on total input sample to classifier. As a result, when the number of transformed samples is larger than original image, we split the input samples in batches to be fed into the SVM classifier. Furthermore, due to various transformation in augmentation generator, the preparation of new samples can be very challenging to achieve in small iterations. Since the position boundary heavily depend on the quality of features, low-quality samples can weaken the searching process or even failed to improve the decision boundary. If this happens, re-running the process for another iteration can typically resolve the issue.

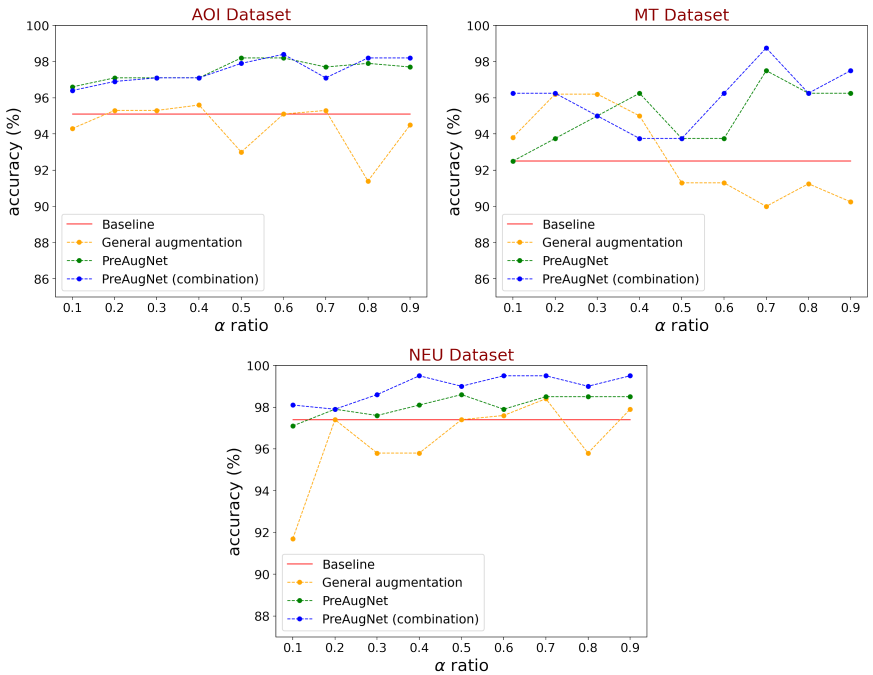

Fig. 8 Test accuracy (%) on AOI, MT and NEU defect datasets. Comparisons across baseline, default on-line augmentation, Pre-augmentation and combination method with different α ratios.

5 Conclusion

In this work, we design the Pre-Augmentation Network (PreAugNet) for generating and screening augmented samples to improve data augmentation in training a target network. The PreAugNet iteratively retrieves CNN features from the raw samples to improve the generated samples, where the updating process is governed by a SVM classifier with decision boundary analysis. This way, the new samples produced from the PreAugNet are much diverse and suitable for effective data augmentation. The effectiveness of this approach is evaluated on the industrial defect recognition problem over three real-world datasets. We compare our PreAugNet with multiple data augmentation approaches, and we also compare our end-to-end pipeline with multiple state-of-the-art surface defect classification methods. Extensive experiments show that the PreAugNet with a standard ResNet-18 target network can achieve 3.3% accuracy improvement on the AOI dataset, 6.25% on the MT dataset, and 2.1% on the NEU dataset. Results demonstrate the effectiveness of the PreAugNet data augmentation on improving the training of a defect classification network.

Future Work. Joint training solutions have great potential in reducing training cost and time for real-world applications. Future work includes tighter integration of the proposed pre-augmentation network with the target network, such that better training performance might be obtained. Also, our approach can be deployed to in other real-world applications where data scarcity remains the bottleneck.

References

[1] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)

[2] Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems **25** (2012)

[3] Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)

[4] Tan, M., Le, Q.: Efficientnet: Rethinking model scaling for convolutional neural networks. In: International Conference on Machine Learning, pp. 6105–6114 (2019). PMLR

[5] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., Berg, A.C.: SSD: Single shot multibox detector. In: European Conference on Computer Vision, pp. 21–37 (2016). Springer

[6] Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: Unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 779–788 (2016)

[7] Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: Towards real-time object detection with region proposal networks. Advances in neural information processing systems **28** (2015)

[8] He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask R-CNN. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2961–2969 (2017)

[9] Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-assisted Intervention, pp. 234–241 (2015). Springer

[10] Choe, J., Shim, H.: Attention-based dropout layer for weakly supervised object localization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2219–2228 (2019)

[11] Golnaz Ghiasi, Q.V.L. Tsung-Yi Lin: Dropblock: A regularization method

for convolutional networks. Advances in neural information processing systems **31** (2018)

[12] Singh, K.K., Lee, Y.J.: Hide-and-seek: Forcing a network to be meticulous for weakly-supervised object and action localization. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 3524–3533 (2017)

[13] Shorten, C., Khoshgoftaar, T.M.: A survey on image data augmentation for deep learning. Journal of big data **6**(1), 1–48 (2019)

[14] Hernández-García, A., König, P.: Data augmentation instead of explicit regularization. arXiv preprint arXiv:1806.03852 (2018)

[15] LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proceedings of the IEEE **86**(11), 2278–2324 (1998)

[16] Bowles, C., Chen, L., Guerrero, R., Bentley, P., Gunn, R., Hammers, A., Dickie, D.A., Hernández, M.V., Wardlaw, J., Rueckert, D.: GAN augmentation: Augmenting training data using generative adversarial networks. In: arXiv:1810.10863 (2021)

[17] Frid-Adar, M., Diamant, I., Klang, E., Amitai, M., Goldberger, J., Greenspan, H.: GAN-based synthetic medical image augmentation for increased cnn performance in liver lesion classification. Neurocomputing (2018)

[18] Vapnik, V.N.: An overview of statistical learning theory. IEEE transactions on neural networks **10**(5), 988–999 (1999)

[19] DeVries, T., Taylor, G.W.: Improved regularization of convolutional neural networks with cutout. arXiv preprint arXiv:1708.04552 (2017)

[20] Yun, S., Han, D., Oh, S.J., Chun, S., Choe, J., Yoo, Y.: Cutmix: Regularization strategy to train strong classifiers with localizable features. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 6023–6032 (2019)

[21] Zhang, H., Cisse, M., Dauphin, Y.N., Lopez-Paz, D.: mixup: Beyond empirical risk minimization. arXiv preprint arXiv:1710.09412 (2017)

[22] Zhong, Z., Zheng, L., Kang, G., Li, S., Yang, Y.: Random erasing data augmentation. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, pp. 13001–13008 (2020)

[23] Wong, S.C., Gatt, A., Stamatescu, V., McDonnell, M.D.: Understanding data augmentation for classification: When to warp? In: 2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA), pp. 1–6 (2016). IEEE

[24] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. Advances in neural information processing systems **27** (2014)

[25] Liu, M.-Y., Tuzel, O.: Coupled generative adversarial networks. Advances in neural information processing systems **29** (2016)

[26] Zhu, J.-Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2223–2232 (2017)

[27] Chen, H., Pang, Y., Hu, Q., Liu, K.: Solar cell surface defect inspection based on multispectral convolutional neural network. Journal of Intelligent Manufacturing **31**(2), 453–468 (2020)

[28] Cheon, S., Lee, H., Kim, C.O., Lee, S.H.: Convolutional neural network for wafer surface defect classification and the detection of unknown defect class. IEEE Transactions on Semiconductor Manufacturing **32**(2), 163–170 (2019)

[29] Huang, Y., Qiu, C., Wang, X., Wang, S., Yuan, K.: A compact convolutional neural network for surface defect inspection. Sensors **20**(7), 1974 (2020)

[30] Park, J.-K., Kwon, B.-K., Park, J.-H., Kang, D.-J.: Machine learning-based imaging system for surface defect inspection. International Journal of Precision Engineering and Manufacturing-Green Technology **3**(3), 303–310 (2016)

[31] Elleuch, M., Maalej, R., Kherallah, M.: A new design based-SVM of the CNN classifier architecture with dropout for offline arabic handwritten recognition. Procedia Computer Science **80**, 1712–1723 (2016)

[32] Kang, J., Park, Y.-J., Lee, J., Wang, S.-H., Eom, D.-S.: Novel leakage detection by ensemble CNN-SVM and graph-based localization in water distribution systems. IEEE Transactions on Industrial Electronics **65**(5), 4279–4289 (2017)

[33] Niu, X.-X., Y.Suen, C.: A novel hybrid CNN-SVM classifier for recognizing handwritten digits. Pattern Recognition **45**(4), 1318–1325 (2012)

[34] Li, C., Xu, T., Zhu, J., Zhang, B.: Triple generative adversarial nets. Advances in neural information processing systems **30** (2017)

[35] Joshi, K.D., Chauhan, V., Surgenor, B.: A flexible machine vision system for small part inspection based on a hybrid SVM/ANN approach. Journal of Intelligent Manufacturing **31**(1), 103–125 (2020)

[36] Sun, X., Liu, L., Li, C., Yin, J., Zhao, J., Si, W.: Classification for remote sensing data with improved CNN-SVM method. IEEE Access **7**, 164507–164516 (2019)

[37] Xue, D.-X., Zhang, R., Feng, H., Wang, Y.-L.: CNN-SVM for microvascular morphological type recognition with data augmentation. Journal of medical and biological engineering **36**(6), 755–764 (2016)

[38] He, Y., Song, K., Dong, H., Yan, Y.: Semi-supervised defect classification of steel surface based on multi-training and generative adversarial network. Optics and Lasers in Engineering **122**, 294–302 (2019)

[39] Jain, S., Seth, G., Paruthi, A., Soni, U., Kumar, G.: Synthetic data augmentation for surface defect detection and classification using deep learning. Journal of Intelligent Manufacturing, 1–14 (2020)

[40] Zhong, J., Liu, X., Hsieh, C.-J.: Improving the speed and quality of GAN by adversarial training. arXiv preprint arXiv:2008.03364 (2020)

[41] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2818–2826 (2016)

[42] Huang, Y., Qiu, C., Yuan, K.: Surface defect saliency of magnetic tile. The Visual Computer **36**(1), 85–96 (2020)

[43] He, Y., Song, K., Meng, Q., Yan, Y.: An end-to-end steel surface defect detection approach via fusing multiple hierarchical features. IEEE Transactions on Instrumentation and Measurement **69**(4), 1493–1504 (2019)

[44] Buslaev, A., Iglovikov, V.I., Khvedchenya, E., Parinov, A., Druzhinin, M., Kalinin, A.A.: Albumentations: Fast and flexible image augmentations. Information **11**(2), 125 (2020)