Towards Automatic Optical Inspection of Soldering Defects

Wenting Dai
School of Computer Science & Engineering, , Nanyang Technological University
Singapore
daiw0004@e.ntu.edu.sg

Abdul Mujeeb
School of Electrical & Electronic Engineering, Nanyang Technological Univeisity
Singapore
amujeeb@ntu.edu.sg

Marius Erdt
Fraunhofer Research Center Nanyang Technological University
Singapore
marius.erdt@fraunhofer.sg

Alexei Sourin
School of Computer Science & Engineering, Nanyang Technological University
Singapore
assourin@ntu.edu.sg

***Abstract*—This paper proposes a method for automatic image-based classification of solder joint defects in the context of Automatic Optical Inspection (AOI) of Printed Circuit Boards (PCBs). Machine learning-based approaches are frequently used for image-based inspection. However, a main challenge is to manually create sufficiently large labeled training databases to allow for high accuracy of defect detection. Creating such large training databases is time-consuming, expensive, and often unfeasible in industrial production settings. In order to address this problem, an active learning framework is proposed which starts with only a small labeled subset of training data. The labeled dataset is then enlarged step-by-step by combining K-means clustering with active user input to provide representative samples for the training of an SVM classifier. Evaluations on two databases with insufficient and shifting solder joints samples have shown that the proposed method achieved high accuracy while requiring only minimal user input. The results also demonstrated that the proposed method outperforms random and representative sampling by ~ 3.2% and ~ 2.7%, respectively, and it outperforms the uncertainty sampling method by ~ 0.5%.**

Keywords—Classification of solder joint defects, Automatic Optical Inspection (AOI), active learning, K-means, SVM classifier.

I. INTRODUCTION

Surface Mount Technology (SMT) is widely used in Printed Circuit Boards (PCBs) assembly. The production pipeline usually includes three steps [1]: 1) Solder printing, where a layer of solder paste, which is made by fusible metal alloy used to adhere and connect the components with PCBs, is printed on the surface of the board; 2) Picking and placing, where each component is positioned on the board; 3) Reflowing, where the solder joints take shape by reflowing of the solder paste.

Different defects may occur at different stages. For example, during the picking and placing stage, defect cases of missing, wrong, doubled or damaged components may occur. In terms of possible soldering defects, most of them happen after the reflowing stage, such as defects at the chip components level (Solder bridge, Shifting, Insufficient solder, Excess solder, Pseudo joint, Raised components, Tombstoned components, Side termination) and defects at the level of IC package components (Solder bridge, Shifting, Insufficient solder, Excess solder, Pseudo joint).

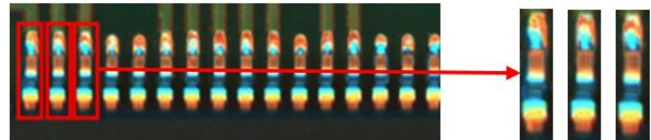


Figure 1. Individual solder joint segmentation.

Two types of defect detection methods are commonly used which are *basic electric tests* including the in-circuit test and the *functional board test*, as well as the *in-line inspection methods* such as the manual visual inspection [2]. However, when using these methods, the high number of potential defects and the small size of solder joints make the inspection challenging. In order to solve that, Automatic Optical Inspection (AOI) becomes a viable option.

The AOI system's pipeline includes three steps: 1) image capturing, 2) image segmentation, and 3) classification. During image capturing, PCBs images are captured by a color camera placed above the production station. For recording the 3D-shape information of solder joints, a three-color tiered illumination system was designed based on the reflection patterns [3]. Then, instead of directly classifying solder joints, image segmentation is performed as a pre-processing task. It segments an integrated image of PCB into individual solder joints, as it is shown in Figure 1. In our research project, these pre-processing works had been already performed. Thus the focus of our work is at the classification task.

Due to its excellent performance and distinguished ability of learning parameters on the provided dataset, machine learning concept has been widely adopted in AOI. However, the lack of labeled data becomes the biggest barrier. Annotating thousands of images is time-consuming and costly. Therefore, we propose a novel method integrating active learning and semi-supervised learning based on clustering analysis. It is not restricted to actively suggest relevant samples for manual annotation, but it also labels a large part of the unannotated samples automatically.

In this paper, Section II reviews the existing solder joint classification methods. The research motivation and hypothesis are stated in Section III. The theoretical background and the details of the proposed method are presented in Sections IV and V, followed by the validation experiments in Section VI. Finally, Section VII concludes the paper.

II. EXISTING SOLDER JOINTS CLASSIFICATION METHODS

According to the detailed survey of AOI of soldering in [2], 17 approaches about SMT solder joint inspection had been proposed before year 2013, meanwhile we also found new 9 methods proposed after year 2013. Here, we only list 15 typical approaches of them which show good performances. Based on the types of components, the algorithms are classified into two groups, which are solder joint inspection for non-Integrated Circuits (non-IC) package [4]-[13] and IC components [14]-[17] (such as chip components).

A. *Non-IC Solder Joints Classification Methods*

With regards to the non-IC solder joint, in order to acquire more information, features are often extracted from different sub-areas, which assumes that particular region division methods are needed. Three approaches ([4][5][6]) adopt the *ensemble learning* concept, which makes the classification decision to be based on outputs of a series of learning algorithms rather than only on one result obtained from any of the constituent learning algorithms alone [18]. A Random Forest-based classification algorithm was presented in [4]. Before the ensemble processing, for keeping the executive efficiency, all decision trees are trained with only two simple features: the color and geometry. The inspection algorithm, which is using improved Adaptive Boosting (AdaBoost), and the decision tree were developed by Xie et al. [5]. The algorithm trains an AdaBoost [19] classifier in each sub-region divided from a solder joint image. The final decision is made by a classification and regression tree (CART) based on the outputs of all classifiers. Luo et al. [6] also developed a similar boosting inspection method, which consists of certain single-layer artificial neural networks. Feature extraction and reduction are also required, and Fisher criterion acts as a measure of feature evaluation. In addition, the authors were aware of the impact of imbalanced distribution of defective samples. Therefore, they weighted the influence of defective samples at the training stage and increased the loss cost of missing faults rather than false alert.

Besides ensemble learning methods, Support Vector Machine (SVM) [20] is also a popular choice for solder joint inspection. Thus, paper [7] shows that even a simple application of SVM in this area can achieve a promising performance. A two-stage classifier was proposed in [8] where five types of defects were detectable. A Bayes classifier firstly classifies a solder joint by its qualification after redundant features are filtered out based on the information gain. If the solder joint fails at the first stage, a multi-class SVM will detect its defects at the second stage. Another approach combining a neural network with genetic algorithm was proposed in [9]. Here, the genetic algorithm is introduced as a feature selector, since the simple two-layer neural network classifier can be easily over-fitted. However, the implementation of its encoding part is complicated, and it inevitably includes too many hyper-parameters.

Learning vector quantization (LVQ) [21]—a prototype-based supervised classification algorithm—was used in [10][11]. Instead of only feeding extracted features from orthogonal view direction images in LVQ like in [10], solder joint images from an oblique view are also captured through special equipment (mirror pyramid) in [11]. Without any heuristic feature extraction processing, pixel values of the acquired images from two viewing directions are directly submited into a standard LVQ neural network. Experiments revealed that multi-angle images are much more informative than images at only one orthogonal view. However, oblique view images cannot always be guaranteed due to the occlusion of adjacent components. Except for using 2D features in the mentioned methods, two approaches presented in [12][13] also compute 3D features from the images captured at tilt angles. At the first stage, the 2D features are fed into a neural network. If the result is not in the confidence interval, its 3D feature will be used in a Bayes classifier after extracting.

B. *IC Solder Joints Classification Methods*

Because detecting defects on the smaller IC solder joints is more difficult than on the non-IC solder joints, a relatively smaller number of papers focus on it. Won Ko et al. [14] proposed an LVQ and fuzzy rule-based classification method. It assumes that first, the three LVQ classifiers are independently trained on three different sub-regions of solder bin. Then, the quality of solder joint is decided based on the fitness value calculated from the results of LVQs based on fuzzy rules. A Multiple Neural Network System is presented in [15]. Two Multilayer perceptron neural networks (MLP) and an LVQ construct two learning levels. In the first level, the extracted features are independently fed into an LVQ and an MLP. If their outputs are inconsistent, all features together with the new extracted wavelet feature will be fed into a new MLP. Its advantages are obvious: there is no any special requirement for illumination system as that mentioned before, and it shows a recognition rate around 99.5%. However, extracting wavelet features is very time-consuming, and the images of all types of solder joints are required as references.

Besides neural network-based methods, model-based methods are also used. An appearance model-based method is presented in [16]. The appearance model is trained on qualified low-rank components after Robust Principle Component Analysis decomposes all qualified samples into low-rank components and error components. Eventually, a difference value between the testing sample and the model on a salient region defined by location prior knowledge is treated as the defect score. Although detection without feature engineering is more efficient, the rate of miss faults is high. Cai et al. [17] proposed a statistical model-based method for non-IC solder joints based on an improved visual background extraction (ViBe) algorithm. Inspired by object detection task, in which foreground objects are the changing parts between two adjacent frames, it is also possible to find defects by comparing the testing solder joint with a qualified model trained by several qualified samples. Though not including any prior

location knowledge, model-based methods still cannot work well if the samples with shift and rotation are not registered.

III. RESEARCH MOTIVATION AND HYPOTHESIS

Some of the mentioned approaches gained satisfactory performances, however they are not free from drawbacks. Even with sufficient dataset used, feature-based methods ([4]-[9] and [15]) cannot be generalized for detecting both solder joints of IC or non-IC components since these hand-crafted features are only designed for one type of solder joint, and these methods always include region division or location prior knowledge. Besides, handcrafted features should be carefully designed because even a small change may influence the performance of inspection method. It is not an easy task. With regards to model-based approaches, in order to increase the tolerant of variations some methods ([10]-[14], [16] and [17]) directly train models based on pixel values of images. They do not face problems caused by feature extraction, but they cannot work well if the samples are not perfectly registered.

Moreover, for all mentioned methods passive learning from the labeled dataset provided by users will cause a problem. In many real applications, such as biomedical image analysis, the majority of samples are unlabeled. For most types of classifiers, a satisfactory performance cannot be guaranteed when training samples are insufficient. The only solution is to request for more labeled samples. However, annotating unlabeled samples may not be an easy task, because special knowledge and skills are required and the annotation cost may also be high. In addition, since there can be redundant information among samples, the most commonly used method—randomly annotating samples from unlabeled dataset—is not an ideal solution. It will significantly reduce the annotating cost if the algorithm can automatically label parts of the samples allowing to make a confident decision and suggest parts of irredundant and informative samples to query annotators. However, the existing methods do not have this ability.

These problems can be addressed if there would be an approach based on a general image representation method, active learning and semi-supervised learning. Therefore, we propose a clustering-based classification method, which uses deep ConvNet trained on a large dataset to extract general features through transfer learning. It combines concepts of active and semi-supervised learning to reduce annotating workload significantly. Furthermore, for providing a more helpful suggestion that takes data distribution into account, a clustering structure computed by the K-means method is also adopted in our proposed method.

IV. THEORETICAL BACKGROUND

A. *Transfer Learning*

Though a serious success in computer vision tasks has been achieved by Convolutional Neural Networks (CNNs) during the past ten years, this approach cannot be fully propagated to many real applications because insufficient labeled samples will overfit CNNs. Transfer learning, which transfers the knowledge gained in the source domain into the target domain, becomes then a possible solution.

In [22], Oquab et al. proposed a transfer convolutional neural network (TCNN), which transfers parts of weights from a pre-trained CNN on ImageNet classification task to the target task—Pascal VOC object detection. The experiments proved that mid-level features learned on large-scale datasets could be transferred to other tasks. Even in different tasks, there is no big gap between the types of objects occurring in their datasets. In our task, the solder joint images are very different from the ImageNet dataset. However, there are works ([23] [24]) which transfer the knowledge acquired from ImageNet dataset into different Automated Optical Inspection tasks, such as metal surface inspection and surface texture classification. It is possible because the general features can be learnt from a large dataset consisting of a wide range of classes [24]. It was experimentally proved that pre-trained CNNs could be used as a generic feature extractor in small datasets even without fine-tuning for AOI tasks.

B. *Active Learning*

Active learning is a sample-query-suggestion algorithm [25]. When the majority of data is unlabeled and the number of labeled samples is insufficient, it will assist classifiers to suggest parts of unlabeled samples for annotation actively. It firstly trains an initial classifier on the small labeled dataset. Based on the prediction of the current classifier, the algorithm selects a small part of "informative" samples followed by querying their labels and adding them to the labeled dataset. Further training the classifier on an enlarged dataset will significantly improve its performance. Therefore, training a classifier with high accuracy but only using a small number of labeled samples becomes possible.

There are three popular active learning sampling strategies: uncertainty sampling, diversity sampling, and representative sampling. However, instead of only selecting one sample at each iteration, batch-mode active learning that suggests multiple samples at each time is more efficient and desirable. Moreover, myopically querying the "best" queries according to any of these strategies often do not work well since it does not take into account the overlap in information content among the "best" instances. To solve this problem, multiple strategies are used in one batch mode sampling framework. Besides been introduced in some applications of structured datasets, such as text classification [26][27], this approach also shows good performances on non-structured datasets that are similar to what we face now, such as image database categorization [28], biomedical image analysis [29] and segmentation [30], object recognition [31] and face recognition [32].

C. *Semi-supervised learning*

Similarly to the considered active learning problems, due to the lack of labeled data, in semi-supervised learning a classifier can be learnt from both labeled data and unlabeled data. However, rather than requiring manual annotation, it can independently annotate unlabeled samples based on certain rules to improve the performance of the classifier. There are many approaches in this area including self-training [33], co-

training [34] generative models [35], and graph-based methods [36].

The simplest method—self-training—firstly trains an initial classifier on the insufficient labeled dataset. Then, the unlabeled data with the highest confidence values are automatically labeled as their predicted labels. For linear classifiers, the unlabeled data that are furthest away from the decision boundary are assigned the high confidence values. Cooperating with previously labeled samples, all of them are used to update the initial classifier. However, it will cause a problem that incorrectly labeled samples included in self-training will inevitably degrade the classification accuracy.

A clustering-based analysis of semi-supervised classifier used in [37] shows that the distribution of labeled data in each cluster will influence the accuracy of the classifier. In the unknown cluster that does not contain any label sample, the accuracy of the classifier is lower than in the others. Experiments show that adding label samples to the unknown cluster by active learning will improve the performance. In our approach, we adopt both active and semi-supervised learning to avoid the problem implicitly.

V. THE PROPOSED METHOD

We present a clustering-based active semi-supervised classification framework, which is illustrated in Figure 2. After data representation, the clustering structure of the whole dataset is analyzed, and an initial classifier is trained on the labeled dataset at the same time. After that, according to the rules of active learning and semi-supervised learning, two groups of unlabeled samples are manually and automatically annotated. Then, the initial classifier is retrained on the enlarged labeled dataset followed by new samples annotation iteration. Eventually, it will stop when meeting the stopping criterion.

A. Data Representation

Data representation includes feature extraction and dimension reduction. Transfer learning is used to transfer the generic image representation method from the CNN trained on the large-scale dataset into our AOI task. For small-scale datasets, fine-tuning strategies may lead the network to over-fit. Therefore, a pre-trained VGG-16 model [38], which contains 13 convolutional layers and three fully-connected layers, is directly adopted as the generic feature extractor after removing the last fully-connected layer. Through this way, the extracted feature vector for each sample has 4,096 dimensions.

However, such a large number of dimensions will impact the performance of the K-means algorithm, which is suitable to relatively low-dimensional data. As a result, for reducing the feature dimensions, Principal Component Analysis (PCA) is used, which orthogonally transforms the data to a new coordinate system where all variables are uncorrelated [39].

Besides, the order of each principal component is determined by its variance, i.e. its position increases with the decrease of its variance. As a result, the principal components at the bottom of the list can be discarded because their

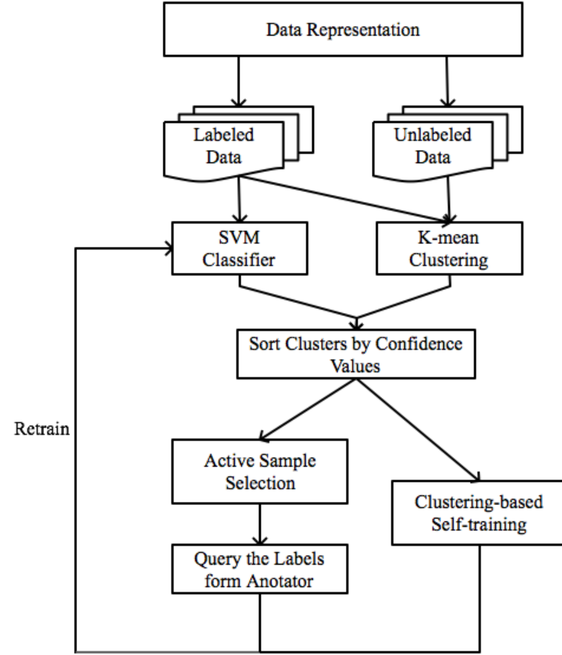


Figure 2. The framework of the clustering-based active semi-supervised classification method.

variances are small, which means these dimensions do not contain much information.

B. Clustering Analysis

For considering data distribution in both active learning and semi-supervised learning, the proposed method uses K-means algorithm to evaluate the clustering structure of the whole dataset including all labeled and unlabeled samples. This algorithm separates the dataset $X = \{x_1, x_2, \ldots, x_N\}$ into K clusters $S = \{S_1, S_2, \ldots, S_K\}$ where each sample belongs to the cluster S_* whose centroid C_* is closest to the sample. After that, an initial SVM classifier is trained on the labeled data followed by predicting the labels for all unlabeled samples. Based on these predictions, it evaluates the confidence to label the samples within each cluster. The used homogeneity value decided by the variance of the label distribution was firstly introduced in [32]. With reference to the binary classification problem, it is defined as:

$$\text{homogeneity}(S_i) = \frac{1}{N_i} \left| n_p^i - n_n^i \right| \tag{1}$$

where N_i is the size of the cluster S_i, and n_p^i and n_n^i are the numbers of the predicted positive and negative samples in this cluster, respectively. The range of this value is between 0 and 1. If it is closer to 1, it means the samples in this cluster are allowed to make confident predictions, and vice versa. Intuitively, if positive and negative labels are equally probable in a single cluster, in the data space it covers a more difficult part to classify than a cluster having a very skewed distribution, in which there is a clear predominant class value [32].

C. Active Sample Selection

We use *uncertainty sampling* as the active learning strategy, which assumes that getting the labels of the least certain samples is more helpful to improve the classifier's performance than getting the labels of others. Similarly, we initially select the r_a clusters with the least confidence values. Then, based on this strategy, active samples are suggested from each selected cluster. For probabilistic model, entropy is often used as the measure of uncertainty sampling. Regarding the non-probabilistic models, such as SVM classifier used in this paper, the sample having the smallest distance to the decision hyperplane is most uncertain to label [27].

D. Clustering-based Self-training

In contrast to active learning, self-training automatically labels the unlabeled data that are most certain to label. Inspired by this criterion, the most r_s confident clusters are selected for further self-training processing. Within each selected cluster, the unlabeled sample having the largest certainty values is automatically labeled as its predicted label from the current classifier. This measure is the same as what is used in active learning. After adding these two newly obtained datasets (D_A and D_S) to the labeled dataset, the current SVM classifier is retained on the enlarged labeled dataset.

E. Re-clustering

The evaluated clustering structure may not be as accurate as we expect since an accurate performance of the basic K-means algorithm cannot be guaranteed. It will possibly group the samples having different labels on the cluster edge. Thus, the proposed method will re-cluster the whole dataset after certain iterations to release the impact of the included errors.

VI. VALIDATION

In this section, we validate the proposed method on two different solder joint datasets acquired by an AOI system while focusing on insufficient solder (dataset_1) and shifting solder detection (dataset_2), respectively. Insufficient solder joints shown in Fig. 3 are caused by that the amount of solder paste printed on PCB is much less than the requirement. Fig. 4 presents six examples of shifting solder joints dataset.

The misalignment of solder joints happens because their components are not placed exactly where they are supposed to be. The numbers of positive and negative samples are 2,610 and 2,427 on dataset_1, whereas there are 1,276 and 1,267 on dataset_2 respectively. To be consistent with the architecture of the VGG-16 network, all data were resized to 224×224×3, followed by extracting 1×4,096-dimensional features.

Since K-means algorithm does not work well on such high-dimensional datasets, the dimensions are further decreased to 1×500 by PCA. After that, we used the K-means implementation of Matlab and LIBSVM [40]. Because the K-means algorithm is used for computing the clustering structure of the whole dataset, K should be set to a larger number than the number of classes (15 is used in all experiments). Furthermore, both the numbers of the suggested active samples and the automatically labeled samples are also 15 in each iteration. However, the obtained clustering structure of the dataset possibly includes mistakes and is different every time. Hence, we re-cluster the dataset after every five iterations and average 20-run results of the proposed method.

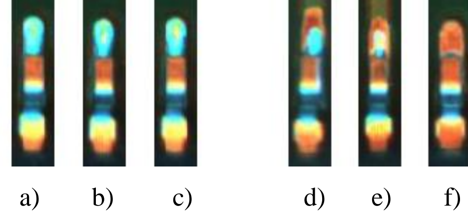


Figure 3. Qualified (a,b,c) and insufiicient (d,e, f) solder joint samples in dataset_1.

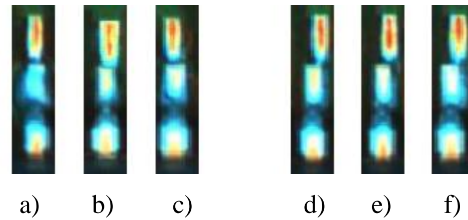
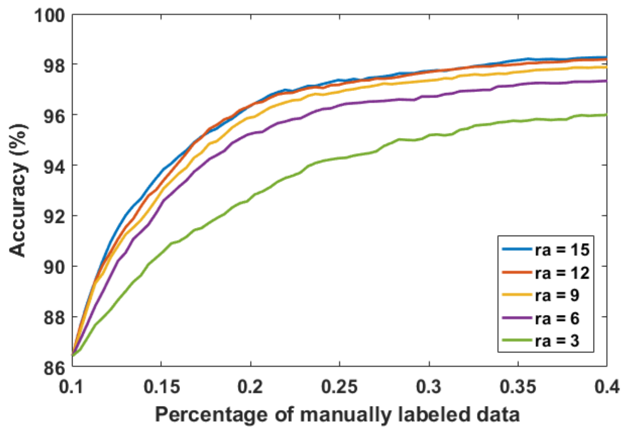


Figure 4. Qualified (a, b, c) and shifting (d,e, f) solder joint samples in dataset_2.

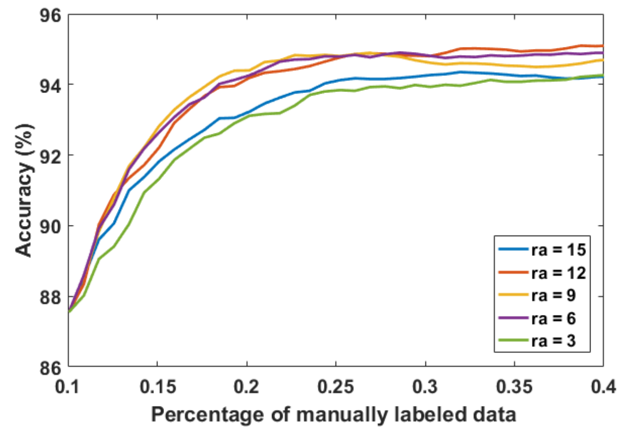
According to the current validation standard, 70% and 30% data are used for training and testing. Besides, in order to simulate the application scenario, in which the majority of samples are unlabeled, only 10% data in training set are labeled. For saving the computation cost, the algorithm will stop till the percentage of manually annotated data increased to 40%.

A. Influence of the Number of Selected Clusters in Active Learning

We compared the performances of choosing a fixed number of samples for manual annotation in different amounts of clusters (15 is used in all experiments). For example, when $r_a = 15$, it will choose only one sample in each selected cluster. The results on two datasets are shown in Fig. 5. The line chart shows the accuracy changes of the classifier trained on the continually enlarged labeled dataset. It is noteworthy that different optimal strategies are gained on two datasets. Regarding insufficient solder data, the largest r_a obtains the best performance, whereas the optimal choice is relatively small on the other dataset. It is probably caused by the different accuracies of the evaluated clustering structures of two datasets. Because the insufficient solder defect has various appearances, which is more irregular than the shifting solder joints, getting an accurate unsupervised clustering result in this dataset is harder. Consequently, informative samples may occur in every cluster rather than appear accurately in certain clusters.

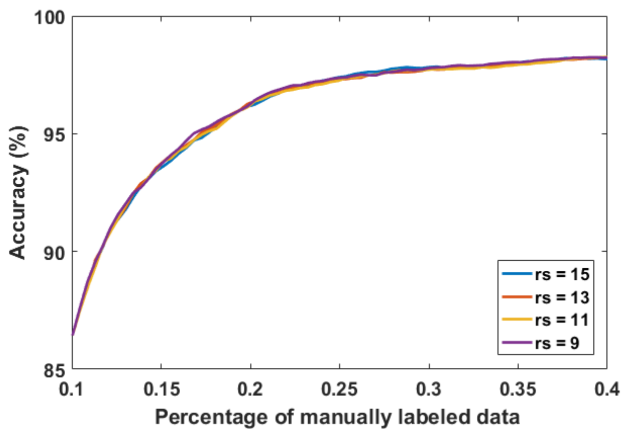


a)

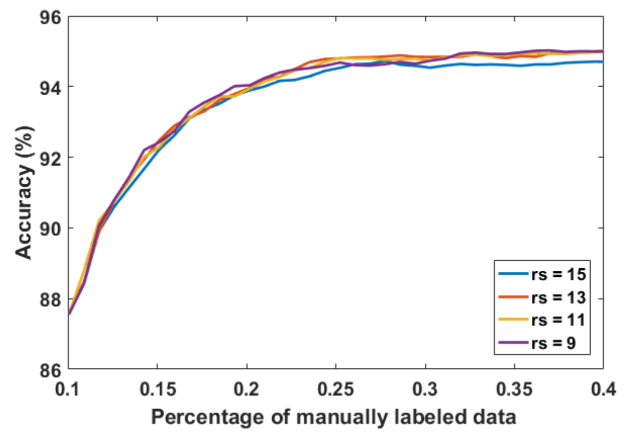


b)

Figure 5. Comparing the results of the proposed method when selecting different numbers of clusters in active learning on a) dataset_1 b) dataset_2.

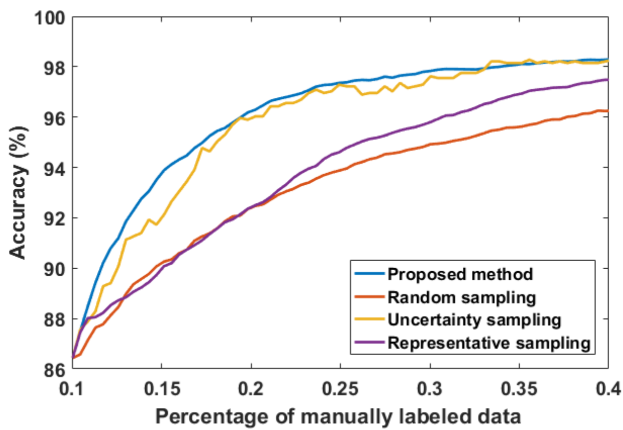


a)

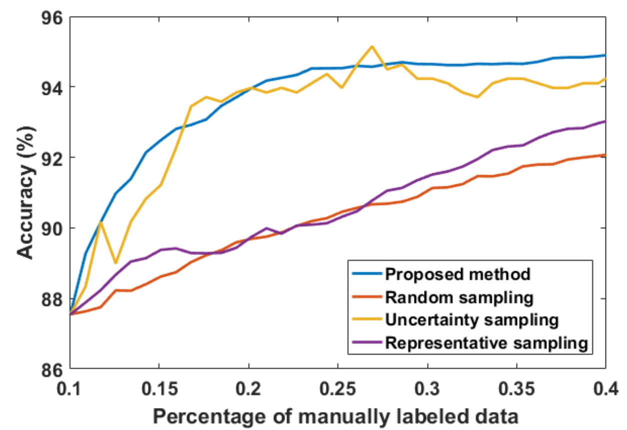


b)

Figure 6. Comparing the results of the proposed method when selecting different numbers of clusters in self-training on a) dataset_1 b) dataset_2.



a)



b)

Figure 7. Comparing the proposed method with 3 benchmarks on a) dataset_1 b) dataset_2.

B. *Influence of the Number of Selected Clusters in Self-training*

During clustering-based self-training processing, 15 samples selected from the highest confident r_s clusters will be automatically labeled without any requirement of manual action from human annotators. Similarly, we compare the influence of r_s —the number of the selected clusters, which is shown in Fig. 6. As we mentioned before, it helps to save the annotation cost but causes inevitable degradation by the incorrectly labeled samples. Therefore, we also compare the average error rate of automatically labeled samples over 20 runs in two datasets, which is presented in Table. 1. According to these two comparisons, we can conclude that the number of incorrectly labeled samples can be avoided when r_s is small.

C. *Comparisons with Benchmarks*

In the last two experiments, we compared the performance of the proposed method with three different benchmarks. Except for random selection that is the baseline for every active learning strategy, uncertainty sampling [27] and representative sampling [26] are chosen for fair comparisons, because they similarly use SVM as the classifier. Uncertainty sampling suggests samples that are closest to the decision boundary of the current classifier, while representative sampling queries the cluster medoids within the margin of the SVM classifier.

For comparison, the default numbers of the samples queried by annotators are the same as K for our method and the other three benchmarks. The accuracy changes are illustrated in Fig. 7. To quantitate their differences, we also computed the value of areas under curves, which is the approximate integral of accuracy. The results are presented in Table. 2.

It is evident that our method significantly outperforms random sampling and representative sampling at 2.94%, 2.33% on dataset1 and 3.4%, 3.13% on dataset2 respectively. At the same time, our approach achieves a slightly better result (94.87% and 91.18%) than the uncertainty sampling method (94.39% and 90.74%) on these two datasets.

TABLE I. THE AVERAGE ERROR RATE OF AUTOMATICALLY LABELED SAMPLES

	Dataset1 (%)	***Dataset2 (%)***
rs = 15	0.572	1.444
rs = 13	0.239	0.481
rs = 11	0.117	0.093
rs = 9	0.047	0.056

TABLE II. THE AVERAGE VALUES OF AREAS UNDER CURVES ON TWO DATASETS

	Dataset1 (%)	***Dataset2 (%)***
Proposed method	94.78	91.18
Random sampling	91.84	87.78
Uncertainty sampling	94.38	90.74
Representative sampling	92.44	88.05

VII. CONCLUSION

We presented a novel classification framework for AOI of solder defects based on active and semi-supervised learning concepts. Besides, taking data distribution into account, a clustering structure computed by the K-means method was also used. During offline training stage, it not only suggests parts of irredundant and informative samples to human annotators, but also automatically labels data with a small error rate (less than 1.5% of all automatically labeled samples). The experimental results proved that the performance of the classifier is continuously improved, but the annotation cost will be largely reduced compared with the other active learning strategies. After that, the trained classification model can be embedded in AOI station of actual manufacturing pipeline to classify individual solder joint, which assumes image segmentation has been done in advance. Therefore, the proposed approach has the potential to be applied in real-world industry AOI solder joint inspection settings. However, currently it is just a binary classifier, which labels samples as qualified and defective. In the future, we will generalize the algorithm for detecting more specific types of defects.

ACKNOWLEDGEMENT

This work was conducted within the Delta-NTU Corporate Lab for Cyber-Physical Systems with funding support from Delta Electronics Inc and the National Research Foundation (NRF) Singapore under the Corp Lab@University Scheme.

REFERENCES

[1] D. Walsh, and J. Arena, “Seeing AOI for six-sigma PCB manufacturing,” Electronic Packaging & Production, 2001.

[2] M. Janóczki, Á. Becker, L. Jakab, R. Gróf, and T. Takács, “Automatic optical inspection of soldering, materials science - advanced topics, “Dr. Erika Hodulova (Ed.), InTech, DOI: 10.5772/51699, 2013.

[3] N.S.S. Mar, C. Fookes, P.K.D.V. Yarlagadda. “Design of Automatic vision Based Inspection System for Solder Joint Segementation,” Journal of AMME 34(2) (2009).

[4] H. Wu, “Solder joint defect classification based on ensemble learning,” Soldering & Surface Mount Technology, vol. 29, no. 3, pp. 164–170, 2017. [Online]. Available: https://doi.org/10.1108/SSMT-08-2016-0016.

[5] X. Hongwei, Z. Xianmin, K. Yongcong, and O. Gaofei, “Solder joint inspection method for chip component using improved adaboost and decision tree,” IEEE Transactions on Components, Packaging and Manufacturing Technology, vol. 1, no. 12, pp. 2018–2027, Dec 2011.

[6] B. Luo, Y. Zhang, G. Yu, and X. Zhou, “ANN ensembles based machine vision inspection for solder joints,” in 2007 IEEE International Conference on Control and Automation, pp. 3111–3115, May 2007.

[7] T. S. Yun, K. J. Sim, and H. J. Kim, “Support vector machine-based inspection of solder joints using circular illumination,” Electronics Letters, vol. 36, no. 11, pp. 949–951, May 2000.

[8] H. Wu, X. Zhang, H. Xie, Y. Kuang, and G. Ouyang, “Classification of solder joint using feature selection based on bayes and support vector machine,” IEEE Transactions on Components, Packaging and Manufacturing Technology, vol. 3, no. 3, pp. 516–522, Mar 2013.

[9] W. Hao, Z. Xianmin, K. Yongcong, O. Gaofei, and X. Hongwei, “Solder joint inspection based on neural network combined with genetic algorithm,” Optik - International Journal for Light and Electron Optics, vol. 124, no. 20, pp. 4110 – 4116, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0030402613000478.

[10] J. K. Kim and H. S. Cho, “Neural network-based inspection of solder joints using a circular illumination,” Image Vision Comp., vol. 13, pp. 479–490, 1995.

[11] T. Ong, Z. Samad, and M. Ratnam, “Solder joint inspection with multi-angle imaging and an artificial neural network,” The International Journal of Advanced Manufacturing Technology 2008; 38: 455– 62.

[12] T.-H. Kim, T.-H. Cho, Y. S. Moon, and S. H. Park, “Visual inspection system for the classification of solder joints,” Pattern Recognition, vol. 32, no. 4, pp. 565 – 575, 1999. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0031320398001034.

[13] T.H. Kim, T.H. Cho, Y.S. Moon, and S.H. Park, “Pattern classification of solder joint images using a correlation neural network,” Engineering Applications of Artificial Intelligence, vol. 9, no. 6, pp. 655 – 669, 1996. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0952197696000462.

[14] K. W. Ko and H. S. Cho, “Solder joints inspection using a neural network and fuzzy rule-based classification method,” IEEE Transactions on Electronics Packaging Manufacturing, vol. 23, no. 2, pp. 93–103, Apr 2000.

[15] G. Acciani, G. Brunetti, and G. Fornarelli, “A multiple neural network system to classify solder joints on integrated circuits,” International Journal of Computational Intelligence Research; 2006. 2: 337– 48.

[16] N. Cai, Y. Zhou, Q. Ye, G. Liu, H. Wang and X. Chen, "IC solder joint inspection via robust principle component analysis," IEEE Transactions on Components, Packaging and Manufacturing Technology, vol. 7, no. 2, pp. 300-309, Feb. 2017. doi: 10.1109/TCPMT.2016.2638503.

[17] N. Cai, J. Lin, Q. Ye, H. Wang, S. Weng, and B. W. K. Ling, "A new IC solder joint inspection method for an automatic optical inspection system based on an improved visual background extraction algorithm," in IEEE Transactions on Components, Packaging and Manufacturing Technology, *Manuf. Technol.*, vol. 6, no. 1, pp. 161-172, Jan. 2016.

[18] L. Rokach, "Ensemble-based classifiers". Artificial Intelligence Review. 33 (1-2): 1–39. doi:10.1007/s10462-009-9124-7, 2010.

[19] Y. Freund and R. E. Schapire, “A decision-theoretic generalization of on-line learning and an application to boosting,” Lecture Notes Computer and System Sciences; vol. 904, pp. 23–27, 1995.

[20] C. Cortes, and V. Vapnik, “Support-vector networks,” Machine Learning; 20 (3): 273–297. doi:10.1007/BF00994018, 1995.

[21] T. Kohonen, and M.A. Arbib “Learning vector quantization,” The Handbook of Brain Theory and Neural Networks, Cambridge; MA: MIT Press, pp. 537–540, 1995.

[22] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, “Learning and transferring mid-level image representations using convolutional neural networks.” IEEE International Conference on Computer Vision and Pattern Recognition, 2014.

[23] S. Kim, W. Kim,Y. Noh, and F.C. Park, “Transfer learning for automated optical inspection,” Neural Networks (IJCNN); doi: 10.1109/IJCNN.2017.7966162, 2017.

[24] V. Natarajan, T. Hung, S. Vaikundam, and L. Chia, “Convolutional networks for voting-based anomaly classification in metal surface inspection,” IEEE International Conference on Industrial Technology; doi: 10.1109/IJCNN.2016.7727522, 2017.

[25] B. Settles, “Active learning literature survey,” University of Wisconsin, Madison, 52(55-66):11. 2, 4, Jan 2010.

[26] Z. Xu, K. Yu, V. Tresp, X. Xu, and J. Wang. “Representative sampling for text classification using support vector machines.” In: Sebastiani F. (eds) Advances in Information Retrieval. ECIR 2003. Lecture Notes in Computer Science, vol 2633. Springer, Berlin, Heidelberg, 2003.

[27] S. Tong and D. Koller. “Support vector machine active learning with applications to text classification,” In Proceedings of the International Conference on Machine Learning (ICML), pp. 999–1006. Morgan Kaufmann, 2000.

[28] N. Grira, M. Crucianu, and N. Boujemaa, “Active semi-supervised clustering for image database categorization,” Content-Based Multimedia Indexing, 2005.

[29] Z. Zhou, J. Shin, L. Zhang, S. Gurudu, M. Gotway, and J. Liang “Fine-tuning convolutional neural networks for biomedical image analysis: actively and incrementally,” In IEEE conference on computer vision and pattern recognition, Hawaii, pp 7340–7349, 2017.

[30] L. Yang, Y. Zhang, J. Chen, S. Zhang, and D.Z. Chen. “Suggestive annotation: A deep active learning framework for biomedical image segmentation,”In Descoteaux M., Maier-Hein L., Franz A., Jannin P., Collins D., Duchesne S. (eds) Medical Image Computing and Computer-Assisted Intervention − MICCAI 2017. Lecture Notes in Computer Science, vol 10435. Springer, Cham, 2017.

[31] F. Gao, Z. Yue, J. Wang, J. Sun, E. Yang, and H. Zhou, “A novel active semi-supervised convolutional neural network algorithm for SAR image recognition,” Computational Intelligence and Neuroscience, vol. 2017, Article ID 3105053. doi:10.1155/2017/3105053, 2017.

[32] D. Ienco, A. Bifet, I. Žliobaitė, and B. Pfahringer. “ Clustering based active learning for evolving data streams,” In Fürnkranz J., Hüllermeier E., Higuchi T. (eds) Discovery Science. DS 2013. Lecture Notes in Computer Science, vol 8140. Springer, Berlin, Heidelberg, 2013.

[33] H. J. Scudder. “Probability of error of some adaptive pattern-recognition machines,” IEEE Trans. Inf. Theory, pp. 363-371, 1965.

[34] A. Blum, and T. Mitchell. “Combining labeled and unlabeled data with co-training,” in: Proceedings of the Eleventh Annual Conference on Computational Learning Theory, in: COLT’ 98, ACM, New York, NY, USA, pp. 92–100, doi:10.1145/ 279943.279962, 1998.

[35] K. Nigam, A.K. McCallum, S. Thrun, and T. Mitchell, “Text classification from labeled and unlabeled documents using EM”, Mach. Learn. 39 (2–3) pp. 103–134, doi:10.1023/A:1007692713085, 2000.

[36] M. Belkin, P. Niyogi, and V. Sindhwani, “Manifold regularization: A geometric framework for learning from labeled and unlabeled examples,” J. Mach. Learn. Res., vol. 7, pp. 2399–2434, Dec. 2006. [Online]. Available: http://dl.acm.org/citation.cfm?id=1248547.1248632

[37] N. Piroonsup, and S. Sinthupinyo. “Analysis of training data using clustering to improve semi-supervised self-training,” Knowledge-Based Systems, Volume 143, pp. 65-80, ISSN 0950-7051, 2018.

[38] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” arXiv preprint arXiv:1409.1556, 2014.

[39] I. T. Jolliffe, “Principal component analysis, series: Springer series in statistics,” 2nd ed., Springer, NY, XXIX, 487 p. 28 illus. ISBN 978-0-387-95442-4, 2002.

[40] C. Chang, and C. Lin, LIBSVM: A library for support vector machines, 2016. Available at https://www.csie.ntu.edu.tw/~cjlin/libsvm/.