The International Journal of Advanced Manufacturing Technology
Vision-Based Defect Detection for Mobile Phone Cover Glass using Deep Neural Networks
--Manuscript Draft--

Manuscript Number:	JAMT-D-17-03342
Full Title:	Vision-Based Defect Detection for Mobile Phone Cover Glass using Deep Neural Networks
Article Type:	Original Research
Keywords:	mobile phone cover glass; defect inspection; deep learning; sementic segmentation.
Corresponding Author:	Hu Su Institute of Automation Chinese Academy of Sciences Beijing, CHINA
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Institute of Automation Chinese Academy of Sciences
Corresponding Author's Secondary Institution:	
First Author:	Zhichao Yuan
First Author Secondary Information:	
Order of Authors:	Zhichao Yuan
	Zhengtao Zhang
	Hu Su
	Lei Zhang
	Fei Shen
	Feng Zhang
Order of Authors Secondary Information:	

Abstract:	The emergency of surface defect would significantly influence the quality of MPCG (Mobile Phone Cover Glass). Therefore, efficient defect detection is highly required in the manufacturing process. Focusing on the problem, an automatic detection system is developed in this paper. The system adopts backlight imaging technology to improve the signal to noise ration and imaging effect. Then, a modified segmentation method is presented for defect extraction and measurement based on deep neural networks. In the method, a novel data generation process is provided, with which the drawback that huge amount of data is required for training deep structured networks can be overcome. Finally, experiments are well conducted to verify that satisfactory performance is achieved with the proposed method.

Suggested Reviewers:	Haiyong Chen Hebei University of Technology

haiyong.chen@hebut.edu.cn

Zaojun Fang
Ningbo Institute of Industrial Technology, CAS
fangzaojun@nimte.ac.cn

Yuan Li
Beijing Institute of Technology
liyuan@bitedu.cn

Cover Letter

(1)   What is the main contribution to the field?
The emergency of surface defect would significantly influence the quality
of MPCG (Mobile Phone Cover Glass). Efficient defect detection is highly
required in the manufacturing process. In this paper, an automatic
detection system is developed.
The system adopts backlight imaging technology to improve the signal to
noise ration and imaging effect. Then, a modified segmentation method is
presented for defect extraction and measurement based on deep neural
networks, overcoming the difficulties encountered in existing methods,
such as imaging inconsistency and heavy requirement of training data. The
system we designed has been used in actual production process and in real
sense realized the application of computer vision in industry defect
detection.

(2)   What is novel? In theory, in experimental techniques, or a
combination of both?
The innovation we made is mainly in experimental techniques. We
originally applied the semantic segmentation methods into the field of
high-precision defect detection and measurement. And for addressing the
challenge, i.e., the requirement of huge labeled data, brought about by
deep learning methods, a data generation algorithm combined with an
augmentation process is proposed. With the data generation algorithm, the
variety of defects' morphologies and imaging inconsistency can be
simulated as well.

(3)   Does your paper have industrial applications? If yes, who are the
likely users?
Yes, we have industrial applications and have already been applied into
real industrial production. The users are mainly MPCG (Mobile Phone Cover
Glass) manufactories.

Vision-Based Defect Detection for Mobile Phone Cover Glass using Deep Neural Networks

Zhichao Yuan, Zhengtao Zhang, Hu Su (Corresponding author), Lei Zhang, Fei Shen and Feng Zhang

Abstract—**The emergency of surface defect would significantly influence the quality of MPCG (Mobile Phone Cover Glass). Therefore, efficient defect detection is highly required in the manufacturing process. Focusing on the problem, an automatic detection system is developed in this paper. The system adopts backlight imaging technology to improve the signal to noise ration and imaging effect. Then, a modified segmentation method is presented for defect extraction and measurement based on deep neural networks. In the method, a novel data generation process is provided, with which the drawback that huge amount of data is required for training deep structured networks can be overcome. Finally, experiments are well conducted to verify that satisfactory performance is achieved with the proposed method.**
Index Terms— **mobile phone cover glass, defect inspection, deep learning, semantic segmentation.**

I.   INTRODUCTION

DETECTION of the surface defect is a significant concern in different manufacture processes and is still mainly accomplished by human workers. Recently, this time-consuming task is increasingly becoming the bottleneck of productivity. With the development of image processing technology and artificial intelligence, the replacement of manual labor using automatic detection is becoming possible. Several automatic detection systems are reported and are applied in the productions of such as stainless steel strip [1,2], the large-scale optical device [3,4], TFT-LCD (Thin Film Transistor Liquid Crystal Display) panel [5], and wafer detection [6].

As a special case of the above, the automatic detection of defect for MPCG is active, not only to cut labor costs but also to improve the detection accuracy. In [7], PCA (Principal Components Analysis) was utilized to analyze image patches, in which way the features could be obtained and then used to classify the test patches into the classes of the equivalent training samples. The defects were recognized and their sizes were measured by simply counting the number of pixels in them. As can be seen, the extracted features are too primitive and the measured accuracy cannot satisfy the requirement of practical applications. In [8], the OMP algorithm (Orthogonal Matching Pursuit) [9] was suggested for selecting an optimal subset (smallest size and least representative error) of the prespecified training dictionary. Then, a sparsity ratio, which was defined as the quotient of the number of non-zero coefficients and the image size, was calculated to determine whether the testing image is qualified. Compared to [7], the method in [8] is more robust to environmental changes such as alignment and lighting for images. With the same consideration of robustness improvement, [10] adopted a different strategy, template

reference, to align the images. A subtraction and grayscale projection was then conducted on acquired images and template to extract defects. Additionally, a further IFCM (Improved Fuzzy c-means Cluster) algorithm was proposed to detect defects with ambiguous boundaries. [11] illustrated two problems encountered in MPCG defect detection, which respectively was high dimensionality of the defect feature and imbalanced defect examples. To address these two problems, an effective scheme was presented, demonstrating superior performance over other methods.

As can be concluded, significant progress has been made on this topic. But there remains much scope for improvement, especially for the above-mentioned methods. For example, the performance of the method proposed in [8] is highly dependent on the sparsity ratio and is sensitive to noises. For the method in [10], it is insufficient for image distortion rectifying to consider only rotations and shifts. And thus, the extraction after subtraction is not accurate. The methods in [8] and [11] cannot measure the defects' sizes, limiting the range of applications. Besides, in the methods, engineers commonly suffer from the heavy work of artificial feature design. There is no explicit guideline for choosing optimal features. As such, human experience is the key to the success.

Recently deep networks have demonstrated remarkable ability to learn high-level features. The features would be certainly helpful to improve performance and are usually hard to design artificially. With the consideration, several attempts are made to use deep networks in surface defect detection. [12] applied a two-layer neural network to detect the defects on textured surfaces, whose inputs were statistical features extracted from randomly generated image patches. [13] designed a deep convolutional neural network, which performed well in the detection of rail defects, in part because of the beneficiary of the huge amount of image data. Another architecture of the deep convolutional neural network for surface defect detection was presented in [14], which was demonstrated to achieve higher performance than traditional detection methods. A generic method for ASI (Automated Surface Inspection) was proposed in [15]. In the method, the defect areas were predicted by thresholding the heat maps generated from the classified patches using a pre-trained deep learning network.

Focusing on the specific task of MPCG defect detection, several difficulties which limit existing detection methods should be emphasized: (1) great variety of defects' morphologies; (2) imaging inconsistency; (3) requirement of high-precision measurement of defects' sizes. Considering the problems, an automatic MPCG detection system is developed in this

paper. A data generation algorithm is firstly proposed to synthesize labeled training images. Subsequently, an augmentation process is conducted, in which the transformations, such as Gamma transformation, rotation, shift, zoom, etc., are involved. The generation step generates various morphologies of defects while the augmentation step simulates the imaging inconsistency. Using the two algorithms, the uncertainty caused by (1) and (2) could be appropriately dealt with. Finally, a pixel-level segmentation network is presented to measure the sizes of the defects with satisfactory precision, which solves (3).

The remainder of this paper is organized as follows. A simple description of the inspection instrument is presented in Section 2. A detailed analysis of traditional methods and the motivation of this paper are stated in Section 3. Section 4 provides a detailed illustration of the defect segmentation network and Section 5 describes the data set generation algorithm. Section 6 shows the experiments and results. Finally, this paper is concluded in Section 7.

II. Image Capture System

As illustrated in Fig. 1, we adopt backlight line-scanning imaging method. The reasons for that could be concluded as follows. Firstly, only one degree of freedom needs to be adjusted to ensure the imaging plane parallel to the surface of the cover glass. Secondly, in comparison with plane array camera at the same cost, line-scan camera would achieve higher resolution. In the detection process, the cover glass to be detected is perpendicular to the optical axis of the camera and parallel to the mask and backlight source. After illuminated by backlight source, the glass is transmitted in and all the areas of the glass will be successively exposed under the optical axis of the camera. The corresponding gray intensities are dependent on the amount of light penetrating through the glass.
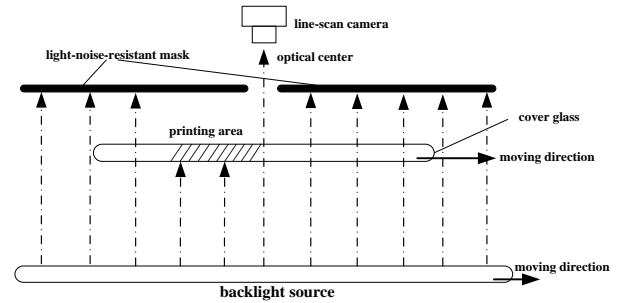

Figure 1: The principle of backlight line-scan imaging. The cover glass is parallel to the mask and backlight source and is perpendicular to the optical axis of the line-scan camera. The camera captures images line by line as the glass moves.
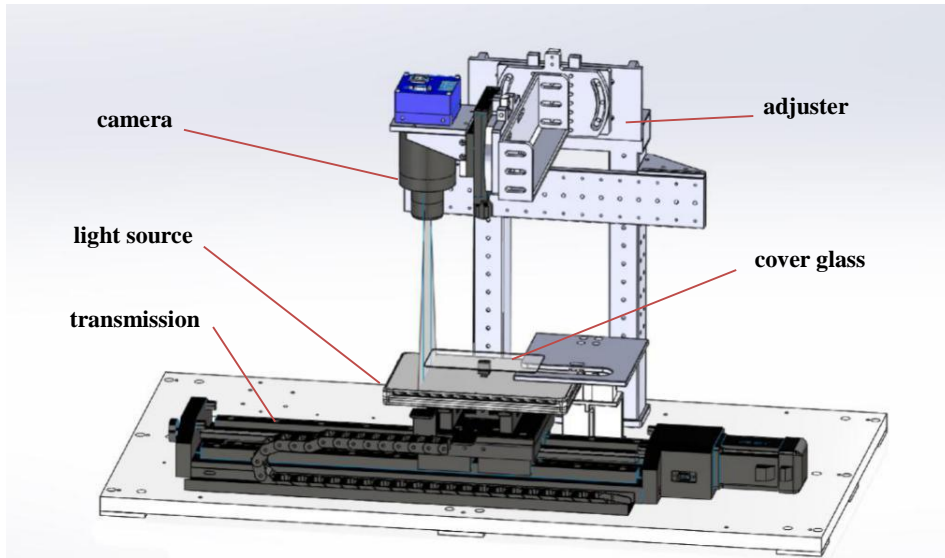

Figure 2: Mechanical structure of experimental platform.

According to the imaging principle, the mechanical structure of the experimental platform is designed and shown in Fig. 2. The platform primarily consists of transmission module and an imaging module. The transmission module is a motorized horizontal stage and the imaging module includes the line-scan camera and the backlight source. The camera is fixed on the platform and captures images when the glass, as well as the backlight source, is transferred by the transmission module through the view.

III. Analysis and Motivation

A. Typical Defects

An example of the upper half part of MPCG image is present in Fig. 3. As illustrated in the figure, an MPCG image can be divided into the printed region and view window region. The components, IR (Infrared Radiation) hole, Camera hole, and Receiver holes, are arranged inside of the printed region. Certainly, different defects would arise in different regions. Among them, the defects of IR hole are the most difficult to detect due to the ambiguous boundary and great variety of defects' morphologies, which are the major concern of the paper.

Typical defects of IR holes are provided in Fig. 4. In the figure, it is worth to note that images at left column are apparently darker than those at the right column. It is what we mentioned as imaging inconsistency caused by the environmental changes (such as lighting), fabrication error and instruments inconsistency.
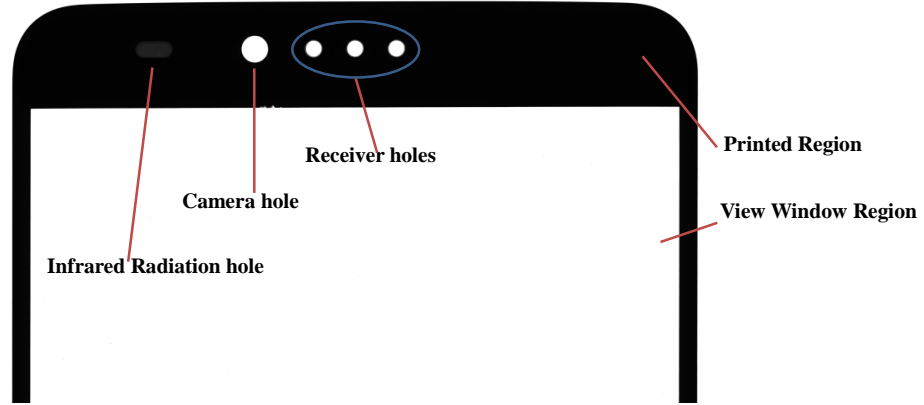

Figure 3: The upper half part of MPCG image. The figure illustrates different components of the image. The dark area is the printed region. The white area surrounded by the printed region is the view window region. On the top of the image, three components inside the printed region are shown, which are infrared radiation hole, camera hole, and receiver holes, respectively.
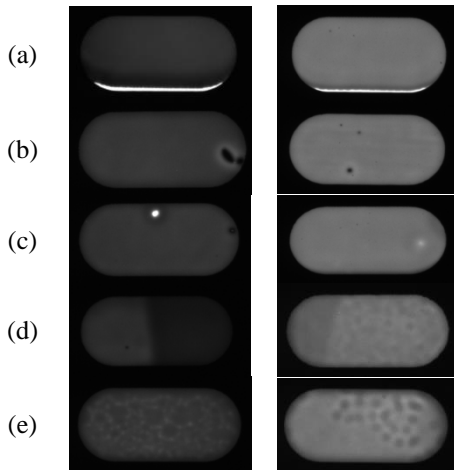

Figure 4: Typical defects of IR hole. The defects can be generally classified into (a) miss printing (b) black point (c) white point (d) contrast color and (e) peel-like appearance. The images illustrate the imaging inconsistency. The first column show the IR images with lower gray intensities while the second column are the ones with higher intensities.

B. Analysis of traditional methods

The IR defects are characterized by the differences of the gray value, i.e., they are usually darker or brighter than surrounded areas. We try to utilize traditional binarization methods and segmentation methods to solve the detection problem. The binarization methods could be classified into two categories, i.e., local methods and global methods. The difference between them is that global methods use only one threshold over the entire image while in local methods, multiple thresholds are involved. Local methods usually give more accurate results as compared to global methods especially in the conditions where the image is effected from bad shading, blurring, low resolution and non-uniform illumination. Both types of methods are employed to solve the problem. Two classical methods, Otsu's thresholding method [16] and Wellner's method [17], which respectively are the global method and local method, are first chosen. Besides, the segmentation methods provide another idea to solve the problem. The methods partition the image into disjoint regions which are homogeneous with respect to some properties, such as gray value or texture. Watershed method [18], which is well known to be a very powerful segmentation tool, is chosen as well to solve the detection problem. The results of the three chosen methods are shown in Fig. 5.

It can be seen from the figure that all of the methods performed poorly on the defect images. In the Otsu's method, the threshold was more apparently influenced by the background than the slight defects. Thus, what the threshold distinguished is actually the difference between foreground and background. Wellner's method could rule out the affection of the background but was sensitive to the noise. Watershed method could effectively detect obvious defects but still cannot exclude the affection of noise.

C. Motivation of the proposed method

Performances of traditional detection methods highly depend on how well the representations can model the properties of defects. However, there is no guideline for designing optimal representations, and thus, expertise is the key to the success of these methods, limiting their wide applications in industrial detections. In recent years, deep learning methods have been achieving good performances in various vision tasks but rarely applied to defect detection problems.

The difficulties encountered in defect detection, such as the great variety of defects' morphologies and imaging inconsistency can hardly be dealt with by the representations employed in traditional methods. Therefore, we try to apply deep learning methods to solve these difficulties. And for addressing the challenge, i.e., the requirement of huge labeled data, brought about by deep learning methods, a data generation algorithm combined with an augmentation process is proposed. The data generation algorithm is proposed to synthesize labeled training images. Subsequently, the augmentation process is conducted, in which the transformations, such as Gamma transformation, rotation, shift, zoom, etc., are involved. The generation step generates various morphologies of defects while the augmentation step

simulates the imaging inconsistency. And considering the requirement of high-precision measurement of defects' sizes, using pixel-level segmentation networks would be a good choice. Inspired by the idea of adversarial segmentation [19], we designed a segmentation network using adversarial training to obtain better robustness.
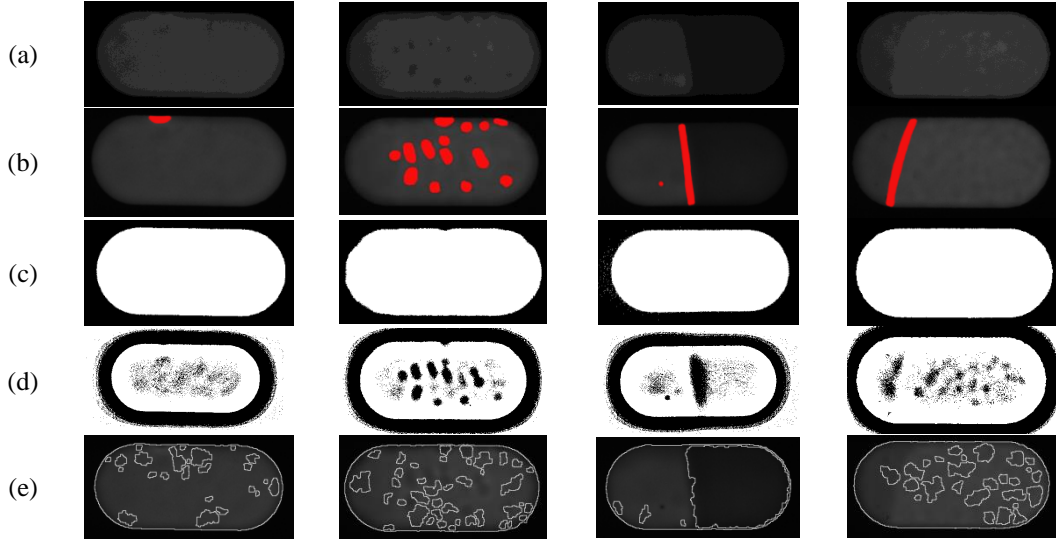

Figure 5: Experimental results of traditional methods. (a) Defect images. (b) Ground truth. (c) Otsu's method[16]. (d) Wellner's method[17]. (e) Watershed method[18].

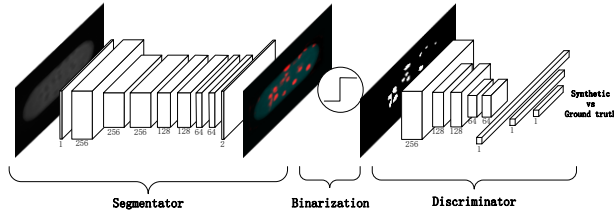## IV. DETECTION WITH SEGMENTATION NETWORK

Figure 6: The abridged general view of the defect segmentation model training with the adversarial network. The segmentor uses the standard segmentation method that gives out synthetic label maps. The adversarial network estimating the probability of the label map belonging to ground truth is a CNN-based discrimination network. The input of adversarial network is a collection of binary map acquired from both synthetic label maps and the ground truths.

A. Segmentation network

The goal of segmentation task is computing pixel-wise labels of target images. Generally, max-pooling layers that usually involved in CNNs (Convolutional Neural Network) would result in loss of accurate spatial information, limiting the applications in segmentation tasks. Considering this, several improvements on the original architecture are suggested, leading to new types of CNNs, such as FCN (Fully Convolutional Network) [20], deep deconvolution network [21] and dilated convolution network [22].

In contrast to standard CNNs, the FCN [20] interpreted the fully connected layers in the standard CNNs as convolutions with large receptive fields and also, the model employed fewer pooling layers. But the resolution reduction and small receptive field simultaneously restricted further improvement on accuracy. The deep deconvolution network [21] composed of deconvolution and unpooling layers is advantageous to handle object scale variations by eliminationg the limitation of fixed-size receptive field in the FCN. The network contributes to gaining accuracy a lot, especially in large objects phase. Nevertheless, accuracy loss still had not been completely solved essentially because of series of downsampled interlayers. Plugging the dilated module into existing segmentation architectures reliably improves the accuracy based on the utilization of receptive fields with increasing sizes. According to the above analysis, we prefer to reduce the number of pooling layers and choose the dilated convolution as our fundamental operator.
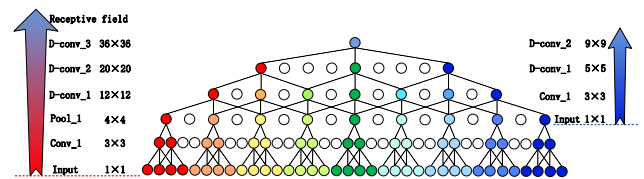
B. Dilated segmentation network

Figure 7: The first 5 layers of our network. The axis left computes receptive fields from the input layer while the right one only focuses on dilated layers. Conv, Pool, and D-conv respectively represent the convolutional layer, pooling layer and dilated convolutional layer. The right axis illustrates the receptive field size of the output layer w.r.t. Pool_1.

Let $f: \mathbb{z}^{+2} \to \mathbb{R}$ be a discrete function and $g: \mathbb{D} \to \mathbb{R}$ be a discrete two-dimensional filter, where $\mathbb{z}^+$ is positive integer field, $\mathbb{R}$ is the real field, and $\mathbb{D} = [-r, r]^2 \cap \mathbb{z}^{+2}$. The discrete convolution is defined as

$$(f * g)(n) = \sum_i f(i)g(n - 1) \qquad (1)$$

The dilated convolution equation is

$$(f *_d g)(n) = \sum_i f(i)g(\frac{n-1}{d}) \qquad (2)$$

where $*_d$ represents a dilated operator with size d, $d \in \mathbb{Z}^+$. Particularly, the origin discrete convolution is 1-dilated convolution with the respective field of size 3×3. Cascading the dilated convolution as

$$f_{i+1} = f_i *_{2^i} g_i \ , \text{for } i = 1,2, \dots + \infty \qquad (3)$$

where the domain of g_i is defined as $\mathbb{D} = [-1,1]^2 \cap \mathbb{z}^{+2}$, and the size of the receptive field of each layer de as f_i is $(2^i + 1) \times (2^i + 1)$. As expected, the receptive field is increasing exponentially.

The segmentation problem we are facing is a high-precision pixel-level labeling task, namely dense prediction. The loss function we displayed here is in common use by state-of-the-art segmentation models.

Since the labeling task aims at distinguishing defects from the background, we used the binary label for convenience. Therefore, the loss function is actually a binary cross-entropy

$$L(\hat{y}, y) = -\sum_{i=1}^{H \times W}[\hat{y}_i ln y_i + (1 - \hat{y}_i)\ln(1 - y_i)] \quad (4)$$

where $\hat{y}$ and y denote the predictions and ground truth, H and W are the height and width of images we predicted. The main task is to minimize the loss function w.r.t. the weights. We used a fully convolutional architecture with one polling layer combined with dilated convolutions to obtain high accuracy pixel-level predictions.

Layers	1	2	3	4	5	6	7	8
Name	Conv_1	Pool_1	D-conv_1	D-conv_2	D-conv_3	D-conv_4	D-conv_5	Conv_2
Dilation size	1		2	2	4	4	8	1
Kernel size	3×3		3×3	3×3	3×3	3×3	3×3	3×3
Kernel receptive field	3×3		5×5	5×5	9×9	9×9	17×17	3×3
Kernel receptive field w.r.t. Pool_1			3×3	3×3	5×5	5×5	9×9	2×2
Receptive field w.r.t. Pool_1			3×3	5×5	9×9	13×13	21×21	22×22
Receptive field	3×3	4×4	12×12	20×20	36×36	52×52	84×84	88×88

Table 1: Details of our dilation model. Considering the use of 2-dilated operator after layer Pool_1, the receptive field size of the output layer is the product of that w.r.t. Pool_1 (the right axis in Fig. 7) and receptive field size of Pool_1. Sizes of receptive fields are not gaining exponentially due to the alternative raising of convolution kernel size and dilation size.

A proper receptive field size of the output layer is the crux of a segmentation task. Meanwhile, designing a suitable network satisfying the desired receptive field size is important. However, due to the mixed-use of pooling layer and various dilated convolutional layers, it is not easy to calculate the receptive field sizes. The input images in our work are 400 pixels wide and 200 pixels high and the sizes of defects are usually smaller than 1/5 the width of the image, namely 80 pixels. Consequently, the ideal size of the receptive field should be no less than 80, in practice, we choose 88. The details of calculating the receptive field size are shown in Fig. 7 and Tab. 1.

C. Training with adversarial model

GANs (Generative Adversarial Networks) proposed by Goodfellow *et al.*[23] play an adversarial game between the generator G and the discriminator D. The discriminator is a neural network $D(x)$ that computes the probability sample x belongs to data distribution. Concurrently, the generator maps sample z from the prior $p(z)$ to data space using another neural network $G(z)$. $D(x)$ tells the samples generated by $G(z)$ from true data while $G(z)$ is trained to fit the data distribution and to confuse $D(x)$ into believing it. The optimization function to this game can be expressed as

$$\min_G \max_D E_{x \sim p_{data}}[log D(x)] + E_{z \sim p(z)}[log(1 - D(G(z)))] \qquad (5)$$

GANs have been shown to produce state-of-the-art results and have been widely applied to several other tasks, such as semi-supervised learning, 3D modeling, semantic segmentation and self-driving [19,24-26]. Training segmentation network with the adversarial model was proposed at [19] of which the experiment results showed that adversarial training approach leads to improved accuracy on the Stanford Background and PASCAL VOC 2012 datasets. In this paper, we follow the idea of adversarial segmentation to make the segmentation model S and the adversarial discriminator D play a minimax game. The goal of S is producing label maps to cheat D while the goal of D is distinguishing the predictions of S from the ground truth. The standard loss function of the segmentation model we described in (1) will be extended into a hybrid function that composed of two terms as

$$L = \{-\sum_{i=1}^{H \times W}[\hat{y}_i ln y_i + (1 - \hat{y}_i) \, ln(1 - y_i)]\} - \lambda\{-[\hat{z} ln z + (1 - \hat{z}) \, ln(1 - z)]\} \qquad (6)$$

where the first term is the original segmentation loss function and the latter one estimates the loss of the D. $\hat{z}$ and z are binary numbers denoting whether the input data is the prediction or ground truth. λ is a hyperparameter that limits the effect of D.

Training the adversarial model is equivalent to maximizing the loss function w.r.t. the second term, therefore, we train D only depending on the latter loss term in case of producing the adverse impact on S. The loss will dramatically increase while D discriminates the predictions of S from the ground truth accurately.
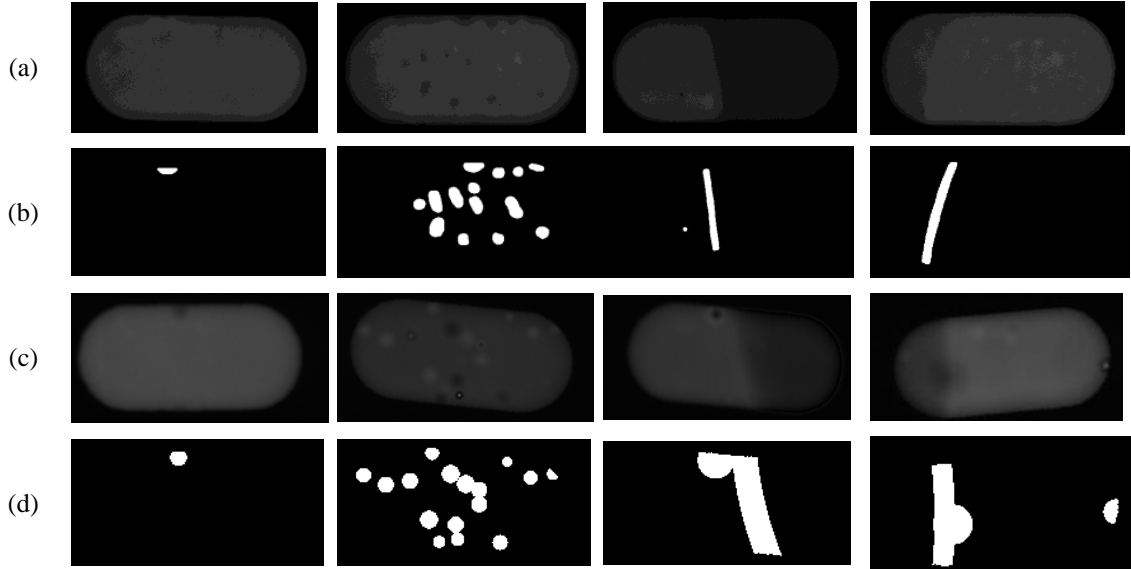
Figure 8: Examples of real and synthesized defects and the labels. (a) Real defects. (b) Real labels (ground truths). (c) Synthesized defects. (d) Synthesized labels.

V. DATA GENERATION

The traditional limitation of deep network is that a huge amount of labeled data is required in the training. The limitation has proven more serious in our work. Note that the common defect rate of MPCG is no higher than 10%. This means that to discover the expected number of defect samples, we have to manually inspect ten times the amount of MPCGs. Besides, with the defect samples, labeling them accurately would be another burden.

To overcome this drawback, many ways have been presented, such as data augmentation, semi-supervised learning, and transfer learning. Data augmentation can be seen as a data preprocess skill. Concerning the topic of image processing, random rotations, shifts, zooms, and flips are commonly employed. Semi-supervised learning is a midway technique between supervised and unsupervised learning. Over and above the supervision information, the semi-supervised algorithm can produce a more competitive model when provided with unlabeled data. [27] extended ELMs (Extreme Learning Machines) with semi-supervised tasks based on manifold regularization. [28] exploited the power of generative models, which recognize the semi-supervised learning problem as a specialized missing data imputation task for the classification problem. [29] studied the more challenging problem of learning deep convolutional neural networks for semantic image segmentation from either (1) weakly annotated training data such as bounding boxes or image-level labels or (2) a combination of few strongly labeled and many weakly labeled images, sourced from one or multiple datasets. As described in those papers, semi-supervised techniques with significantly less annotation effort can learn models delivering competitive results on many kinds of machine learning tasks. Transferring learning has broken the assumption which may not hold in real-world applications that the distribution and feature space must be the same of training and future data in machine learning and data mining [30]. It is a technique that utilizes knowledge gained on solving a pre-existing problem with sufficient training data for another related problem and can greatly improve the performance of learning by avoiding much expensive data labeling efforts. In medical processing, obtaining big datasets with comprehensive annotations is particularly challenging. To avoid the difficulty while keeping satisfactory accuracy, [31]adopted the models which were firstly pre-trained on natural images and then fine-tuned on small annotated datasets. [32] developed a weakly supervised approach to detector training where the object location is not manually annotated but automatically determined based on binary (weak) labels indicating if a training image contains the object. [33] presented self-taught learning structure that can improve the performance on classification tasks based on transfer learning from unlabeled data randomly downloaded from the Internet.

In this section, we propose a data generation algorithm consisting of two steps, defect superposition, and data augmentation. The superposition step overlays defects on a chosen defectless image in randomly selected locations with stochastic sizes, shapes, and severities. The defect images generated from single one template would naturally have the same scale, shape, and background color, making them not accord with the actual situation. Therefore, the augmentation step is applied followed the superposition step to augment the variety of the generated defect images.

The pseudo-code is shown in Algorithm 1. In our algorithm, N_t is denoted as the number of training samples, $Q_s = \{q_1, q_2, \dots, q_n\}$ is the criteria set for defect detection, where q_i represents the ith criterion. For each defect in synthesized images, loc denotes the location, s denotes the size, SE denotes the severity and $S = \{s_1, s_2, \dots, s_n\}$ denotes the shape. For each training image Im_i, we initialize its corresponding label map $label_i$ to 0 and use an **RNG** (Random Number Generator) to stochastically select the number N_d of defects that will be superposed. The defect exported by function **DEFECT** depending on loc_j, s_j, SE_j, s_j will be labeled as 1 if any criteria in Q_s are violated.

Algorithm 1 Defect superposition

Input: $\mathcal{I}m$, N_t, and Q_s
PROCEDURE IMAGE SYNTHESIZING
    FOR each $i \in (1,2,\dots,N_t)$ **DO**
        $Im_i \leftarrow \mathcal{I}m,\ label_i \leftarrow \mathbf{0},\ N_d \leftarrow \textbf{RNG}$
        FOR each $j \in (1,2,\dots,N_d)$ **DO**
            $loc_j, s_j, SE_j, \mathcal{s}_j \leftarrow \textbf{RNG}$
            $Im_i \leftarrow Im_i + \textbf{DEFECT}(loc_j, s_j, SE_j, \mathcal{s}_j)$
            IF $loc_j, s_j, SE_j, \mathcal{s}_j \notin Q_s$ **DO**
                $label_i \leftarrow label_i + \mathbf{1}(loc_j, s_j, SE_j, \mathcal{s}_j)$
            ENDIF
        ENDFOR
    ENDFOR
END PROCEDURE
Output: $\{(Im_1, label_1);\ (Im_2, label_2);\ \dots;\ (Im_{N_t}, label_{N_t})\}$

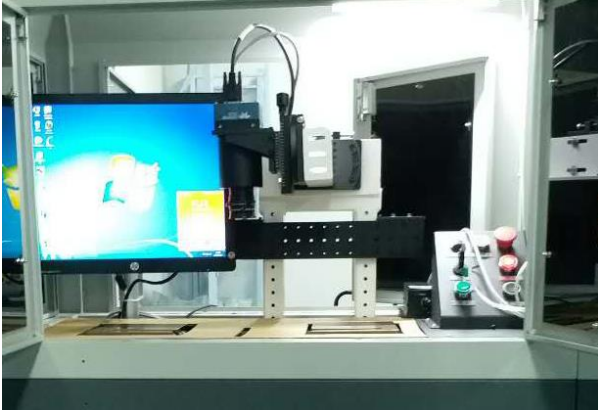## VI. EXPERIMENTS

A Experimental System

Figure 9: Experimental system.

    In accordance with the scheme provided in Section II, the established experimental system was illustrated in Fig. 9. As stated, the system consists of the motion module, the vision module, and the industrial control computer. The location accuracy of the motorized horizontal stage was 1 μm. The vision module included the camera and the light source. In the module, Dalsa line-scan camera was adopted for better imaging precision. Its resolution was 8192 pixels and the pixel size was $10 \times 10\ \mu m^2$. The size of MPCG was 140 *mm* long and 68 *mm* wide. To ensure integrity, 20000 lines were captured along the long side of MPCG during the detection process. The CPU of the computer was Intel Core i5 with a frequency of 3.2 GHz, and the GPU was NVIDIA GeForce GTX 1060.

B. Evaluation methodology

    To evaluate our algorithm, we borrowed the metrics that commonly used in semantic segmentation and scene parsing evaluation, which were pixel-level accuracy and IoU (region Intersection over Union). It was worth to note that accurate determination of whether or not an MPCG is acceptable was a major concern and was of particular importance in practical applications. Here, we adopted another metric to measure the accuracy of the binary classification. The definitions of the metrics are:

- IoU $= \dfrac{\Omega_{TP}}{\Omega_{FP} + \Omega_{TP} + \Omega_{FN}}$

- pixel accuracy $= \dfrac{\Omega_{TP} + \Omega_{TN}}{\Omega_{TP} + \Omega_{TN} + \Omega_{FP} + \Omega_{FN}}$

- binary classification accuracy $= \dfrac{\sum_i S_{ii}}{\sum_i \sum_j S_{ij}} = \dfrac{S_{00} + S_{11}}{S_{00} + S_{01} + S_{10} + S_{11}}$

where S_{ij} represents sample i classified to class j and other symbols are explained in Fig. 10.
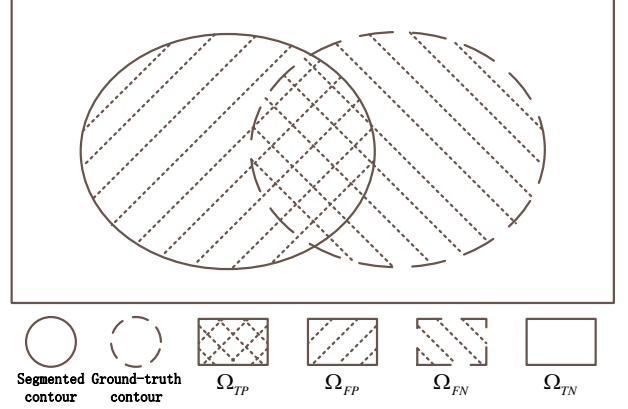

Figure 10: Schematic of segmentation result that shows the definition of crux elements for performance computing. The solid line circle stands for segmented contour and the dotted line denotes the ground-truth contour. Ω_{TP}, Ω_{FP}, Ω_{FN} and Ω_{TN} are true-positive, false-positive, false-negative and true-negative sets of the segmented points, respectively.

C. Defect detection Experiment

    We manually labeled 170 defect samples and chose 60 defectless samples to evaluate the performance of the proposed method. In Fig.11, we give an illustration of the segmentation results from the network with and without adversarial training. The segmentation results are heat maps that denote the probabilities of pixels belonging to defects. We added the heat maps to the red channels of the images for better view. The results show that the adversarial training could better enforce spatial consistency among the labels and smoothen and strengthen the class probabilities over areas. In the training process, we firstly trained the adversarial discriminative network and the segmentation network independently, ensuring the meaningness of the adversarial loss. Then, we re-trained the two networks under the loss to fine-tune the weights for further improvement. However, these steps will make the training unstable after just a few epochs. We found that using an alternating scheme is more effective. We experimented with both fast and slow alternating schemes. At the fast scheme we alternate between updating the segmenting network's and the adversarial network's weights at every iteration of SGD (Stochastic Gradient Descent) and at the slow one, we alternate only after 100 iterations of each. We found the second scheme lead to the more stable training.

    We applied a binarization process on the predicted heat map to obtain pixel prediction accuracy, classify accuracy and mean IoU on the manual labeled sets. Compare the results of adversarial training with the dilated segmentation, we obtained a mean IoU of 68.68 over 55.94. And the performances of all three metrics on the validation data reflect the superiority of adversarial

training. The evaluations are listed in Tab. 2, and the classified results are provided in Tab. 3.
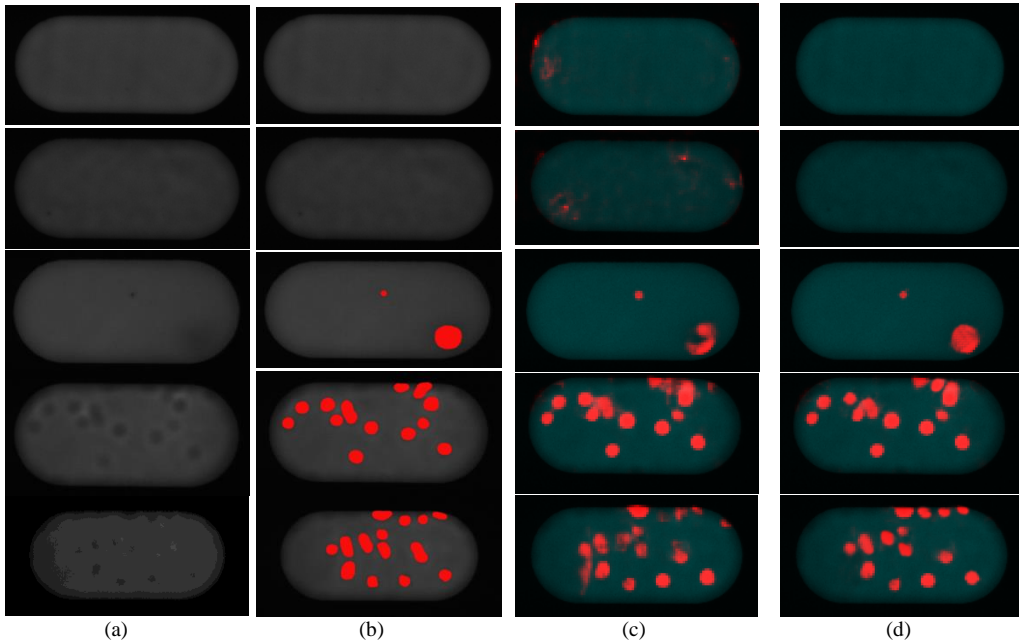

| (a) | (b) | (c) | (d) |

Figure 11: Results of the defect detection. (a) are the acquired samples; (b) are the ground truth; (c) and (d) are respectively segmentation results from the dilated network and the adversarial trained network.

	Mean IoU	Pixel acc.	Classify acc.
Dilated segmentation	55.94%	97.41%	96.52%
Adversarial segmentation	**68.68%**	**98.80%**	**98.26%**

Table 2: Segmentation accuracy on the manual labeled dataset.

		Determined class	
		Defect	Ok
True classes	Defect	169	1
	Ok	3	57

Table 3: Classification results on the manual labeled dataset.

D. Defect Size Measurement Experiment

Index	Dilated Segmentation (μm)		Adversarial Segmentation (μm)		Actual (μm)
	Size	Error	Size	Error	
1	30×20	-1×-13	30×30	-1×-3	31×33
2	40×40	-22×0	50×40	-12×0	62×40
3	130×70	8×-15	130×80	8×-5	122×85
4	200×230	-41×-43	210×250	-31×-23	241×273
5	380×50	-150×21	490×40	-40×11	530×29
6	460×330	-13×-11	460×340	-13×-1	473×341
7	2610×80	65×-7	2580×80	35×-7	2545×87
8	2790×40	-75×-13	2830×40	-35×-13	2865×53

Table 4: Results of the size measurement of defects.

Experiments were conducted as well to verify the effectiveness of the developed network in measuring the sizes of the defects. For this purpose, 8 typical defects at different scales were selected and measured. Considering the difficulties in measuring the sizes of defects with various shapes, we used the heights and widths of the corresponding MERs (Minimum Enclosing Rectangle) to represent the sizes for convenience in the experiment. The measuring results are shown in Tab. 4. In the table, the maximum error in our proposed method was about 40 μm while that in dilated segmentation method was 150 μm. As can be concluded, the proposed method achieved great improvement on measuring accuracy compared to the similar dilated segmentation method.

VII. CONCLUSION

In this paper, we presented a novel MPCG defect detection instrument using backlight line-scanning

imaging method. Focusing on the detection of defects, we introduced a modified segmentation method based on deep learning and discussed the idea of training the networks with adversarial ways. Additionally, to overcome the drawback of the requirement of a huge amount of training data, a data generation process was provided. Finally, considerable manual labeled real images were collected and used to verify the proposed detection method. Experimental results show that the proposed method has high detection accuracy for the defects on MPCGs. In the future, we will focus on the research of applying unsupervised deep learning methods to synthesizing images with annotations.

REFERENCES

1. Ghorai S, Mukherjee A, Gangadaran M, Dutta PK (2013) Automatic defect detection on hot-rolled flat steel products. IEEE Transactions on Instrumentation and Measurement 62 (3):612-621

2. Luiz AM, Flávio LP, Paulo EA Automatic detection of surface defects on rolled steel using computer vision and artificial neural networks. In: IECON 2010-36th Annual Conference on IEEE Industrial Electronics Society, 2010. IEEE, pp 1081-1086

3. LI A-x, YANG T-y, ZHANG Y (2007) Preliminary research of surface defect recognition based on machine vision [J]. Journal of Chongqing University of Posts and Telecommunications (Natural Science) 4:012

4. Tao X, Zhang Z-t, Zhang F, Shi Y-L, Xu D Development of detection techniques of surface defects for large aperture optical elements based on machine vision. In: Control Conference (CCC), 2014 33rd Chinese, 2014. IEEE, pp 2935-2940

5. Torng J, Maung KK, Fan K-C (2013) Development of an automated optical inspection system for mobile phone panels. Journal of the Chinese Society of Mechanical Engineers Transactions of the Chinese Institute of Engineers, Series C/Chung-Kuo Chi Hsueh Kung Ch'eng Hsuebo Pao 34:103-108

6. Yoda H, Ohuchi Y, Taniguchi Y, Ejiri M (1988) An automatic wafer inspection system using pipelined image processing techniques. IEEE Transactions on Pattern Analysis and Machine Intelligence 10 (1):4-16

7. Li D, Liang L-Q, Zhang W-J (2014) Defect inspection and extraction of the mobile phone cover glass based on the principal components analysis. The International Journal of Advanced Manufacturing Technology 73 (9-12):1605-1614

8. Liang L-Q, Li D, Fu X, Zhang W-J (2016) Touch screen defect inspection based on sparse representation in low resolution images. Multimedia Tools and Applications 75 (5):2655-2666

9. Tropp JA, Gilbert AC (2007) Signal recovery from random measurements via orthogonal matching pursuit. IEEE Transactions on information theory 53 (12):4655-4666

10. Jian C, Gao J, Ao Y (2016) Automatic surface defect detection for mobile phone screen glass based on machine vision. Applied Soft Computing

11. Jian C, Gao J, Ao Y Imbalanced defect classification for mobile phone screen glass using multifractal features and a new sampling method. Multimedia Tools and Applications:1-22

12. Weimer D, Thamer H, Scholz-Reiter B (2013) Learning defect classifiers for textured surfaces using neural networks and statistical feature representations. Procedia CIRP 7:347-352

13. Faghih-Roohi S, Hajizadeh S, Núñez A, Babuska R, De Schutter B Deep convolutional neural networks for detection of rail surface defects. In: Neural Networks (IJCNN), 2016 International Joint Conference on, 2016. IEEE, pp 2584-2589

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

15. Ren R, Hung T, Tan KC (2017) A Generic Deep-Learning-Based Approach for Automated Surface Inspection. IEEE Transactions on Cybernetics

16. Otsu N (1979) A threshold selection method from gray-level histograms. IEEE transactions on systems, man, and cybernetics 9 (1):62-66

17. Wellner P (1993) Interacting with paper on the DigitalDesk. Communications of the ACM 36 (7):87-96

18. Beucher S, Lantuejoul C (1979) International Workshop on Image Processing: Real-time Edge and Motion Detection/Estimation.

19. Luc P, Couprie C, Chintala S, Verbeek J (2016) Semantic Segmentation using Adversarial Networks. arXiv preprint arXiv:161108408

20. Long J, Shelhamer E, Darrell T Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015. pp 3431-3440

21. Noh H, Hong S, Han B Learning deconvolution

network for semantic segmentation. In: Proceedings of the IEEE International Conference on Computer Vision, 2015. pp 1520-1528

22. Yu F, Koltun V (2015) Multi-scale context aggregation by dilated convolutions. arXiv preprint arXiv:151107122

23. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y Generative adversarial nets. In: Advances in neural information processing systems, 2014. pp 2672-2680

24. Denton E, Gross S, Fergus R (2016) Semi-supervised learning with context-conditional generative adversarial networks. arXiv preprint arXiv:161106430

25. Gadelha M, Maji S, Wang R (2016) 3d shape induction from 2d views of multiple objects. arXiv preprint arXiv:161205872

26. Santana E, Hotz G (2016) Learning a driving simulator. arXiv preprint arXiv:160801230

27. Huang G, Song S, Gupta JN, Wu C (2014) Semi-supervised and unsupervised extreme learning machines. IEEE transactions on cybernetics 44 (12):2405-2417

28. Kingma DP, Mohamed S, Rezende DJ, Welling M Semi-supervised learning with deep generative models. In: Advances in Neural Information Processing Systems, 2014. pp 3581-3589

29. Papandreou G, Chen L-C, Murphy K, Yuille AL (2015) Weakly-and semi-supervised learning of a DCNN for semantic image segmentation. arXiv preprint arXiv:150202734

30. Pan SJ, Yang Q (2010) A survey on transfer learning. IEEE Transactions on knowledge and data engineering 22 (10):1345-1359

31. Shin H-C, Roth HR, Gao M, Lu L, Xu Z, Nogues I, Yao J, Mollura D, Summers RM (2016) Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE transactions on medical imaging 35 (5):1285-1298

32. Shi Z, Siva P, Xiang T (2017) Transfer learning by ranking for weakly supervised object annotation. arXiv preprint arXiv:170500873

33. Raina R, Battle A, Lee H, Packer B, Ng AY Self-taught learning: transfer learning from unlabeled data. In: Proceedings of the 24th international conference on Machine learning, 2007. ACM, pp 759-766