A Generic Deep-Learning-Based Approach for Automated Surface Inspection: A Review

Ruoxu Ren, Terence Hung, and Kay Chen Tan, *Fellow, IEEE*

Dipesh Chatrola
Department of ICT
Marwadi University
Rajkot, India
dipesh.chatrola107612@marwadiuniversity.ac.in

***Abstract*—— In industry, automated surface inspection (ASI) is a tough process as obtaining training datasets is often expensive and are very dataset-dependent. In paper, propose a generic technique for ASI that uses little training data. First, it constructs a classifier based on the features of image patches, that are transferred from a pretrained deep learning network. The trained classifier is then convolved over the input image to get pixel-wise prediction. An experiment is done out on three public and one industrial datasets.**

Keywords— Automated surface inspection (ASI), Deep learning (DL), Convolutional neural network (CNN), Deep convolution activation feature (Decaf), Multinomial logistic regression (MLR), Feature transferring, Classification, Segmentation, Thresholding.

I. INTRODUCTION

The provided research paper[1] is all about Automated Surface Inspection using deep-learning approach. In earlier days, the surface inspection is handled manually as it's taking more time and the rate of detection hugely impacted by the subjectivity and expertise. So, to tackle it: automatic surface inspection was created to address the drawbacks of manual inspection. ASI methods can be categorized into four[2] parts:

1) Structural – it models the texture primitives and displacements and applied to repetitive patterns

2) Statistical – it's a statistical method, measure the distribution of pixel values and applied to stochastic textures

3) Filter-Based – it applies filter based on texture images and used as a pre-processing step

4) Model-based approaches – by modelling multiple properties of defects it constructs representations of images.

Deep Learning method is also discussed, it can achieve good performance on object recognition[3] but is rarely applied in field of ASI due to: large amount of data is required for DL method for avoiding over-fitting and it's usually for image classification or bounding box generation.

The research paper proposes a general strategy for ASI to address the two difficulties. The proposed technique extracts patch features using a pretrained DL network, produces a defect heat map based on the patch features, and then predicts the defect area by thresholding and segmenting the heat map.

Furthermore, the suggested technique identifies problem areas by generating a defect heat map depending on a trained patch-classifier. Experimental result suggested that the method works well on tiny datasets, which significantly decreases the cost of data collection, and that the proposed strategy is applicable to a variety of surface defect datasets.

II. SUMMARY

CNN is a type of feed-forward neural network and is applied in this paper, it consists of three major layers:

1) Convolution layer: responsible for recognizing features in pixels

2) Pooling layer: responsible for making these features more abstract

3) Fully-connected layer: responsible for using the acquired features for prediction.

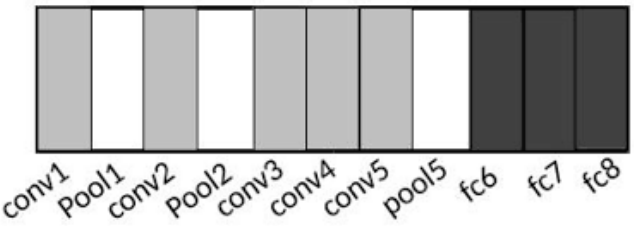


Fig. 1 Decaf layer structure. conv: convolution layers, pool: pooling layers and fc: fully connected layers.

Here in CNN, the convolution is applied to filter the information and produce a feature maps of input image. The convolution of small patch formula for the given filter is:

filters = activation function (weight of filter × small patch + bias of filter)

By looping the patch window in the whole image with particular patch size $x \times y$ and stride size s, in result we get convolution output as:

$$\mathrm{K} \times \lceil[(X - x + 1)/s]\rceil \times \lceil[(Y - y + 1)/s]\rceil$$

Where, K is the number of filters, $X \times Y$ is the size of input image and ceiling function denotes as $\lceil\cdot\rceil$.

After convolutional layer, pooling layer is applied because down sampling operation along height,width or length of input is done by this layer and to reduce the feature dimension and to avoid over-fitting problem. For example, max pooling with input size as $X \times Y$ and patch size as $x \times y$ gives output as size of:

$$\lceil[(X - 1)/\,x]\rceil \times \lceil[(Y - 1)/y]\rceil.$$

The fully connected layers are for compute the class score and they are typically the last few layers of a CNN. Decaf contains five convolutional, three pooling and three fully connected layers.

Researcher look for semi-supervised multi-task acquisition of deep convolutional representations in the paper, in which models are learned on a class of similar issues and then

applied to new tasks with limited training data to build a full deep model.

Although DeCAF is highly used but researchers first train a deep CNN on large-scale and then it applied to a small dataset. So, by removing few layers of the trained CNN anyone can performed it. For the feature extractor[4], [5] the weight of Decaf is generally reused. Here fc6 and fc7 are the two intermediate layers that are tested as an image feature[6], furthermore fc6 is better than fc7 when the quality is concerned.

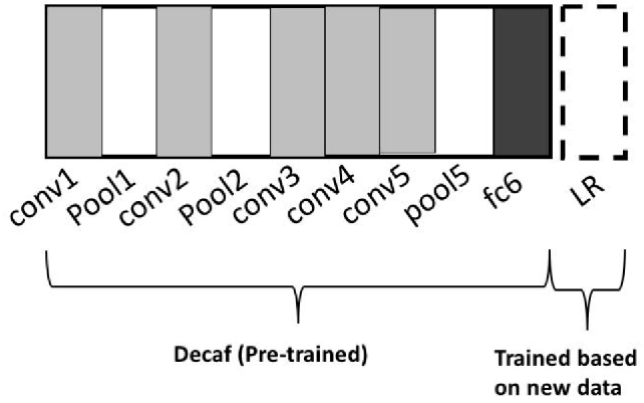


Fig. 2 Example of feature transferring from Decaf

III. PROPOSED GENERIC ASI METHOD

The method contains two phases: supervised training of patch classifier and to segment defect area given an input image based on trained classifier.

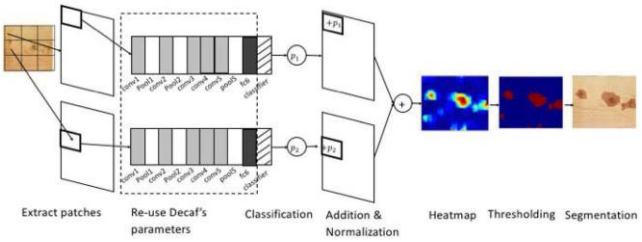


Fig. 3 Structure of proposed method

A. Training of Patch Classifier

In patch extraction the training patches are collected same as CNN. Suppose we have 6×6 as image size, 3×3 as patch size and 2×2 stride or kernel size then according to the formula $\lceil[(X-p+1)/s]\rceil \times \lceil[(Y-p+1)/s]\rceil$ where X, Y is image size and p is patch size and s is stride/kernel size, we get output equal to 4 patches that are extracted from an image size of 6×6.

Furthermore, by using feature transfer method the patch features are been extracted once the training patches are been collected. In computer vision[6], [7] most of the times fully connected layer i.e., fc6 is proven effective after selected as feature extractor.

Multinomial Logistic Regression (MLR) is chosen as a classifier as it close to softmax regression which generalize with two or more discrete outcomes(classes).

B. Segmentation Framework

The phase includes; patch extraction, feature extraction, classification and heatmap generation and prediction. As patch and feature extraction is same as above process that have done.

In classification, it classifies the outputs the probabilities of an input image that are belongs to different classes.

In heatmap they have used HM generation algorithm. And after obtaining heatmap, Otsu's method[8] is applied for binarization. The Otsu's method is all about choosing optimal threshold automatically by minimizes intraclass (within class) variance or by maximizes interclass (between class) variance. Otsu's method objective formula is:

$$\min_{T} \sigma^2_{intra}(T) = n_B(T)\,\sigma^2_B(T) + n_A(T)\,\sigma^2_A(T)$$

where

$\sigma^2_B(T)$ = variance of pixels below threshold and

$\sigma^2_A(T)$ = variance of pixels above threshold.

Algorithm: Defect Segmentation

1. Input ← *Im*, *σ*, and *k*
2. **procedure** SURFACE–INSPECTION
3. $\mathcal{HM}_{1,2\ldots Nd}$ = HM–GENERATION *(Im)*
4. **for** each $j \in$ *(1, 2, …, Nd)* **do**
5. ß*j* ← Otsu's *(HMj)*
6. *Regions* ← Felzenszwalb's *(ßj, σ, k)*
7. **for** each *R* ∈ *Regions* **do**
8. **if** *average(HM(R))* < $\frac{1}{Nd}$ **then**
9. Remove *R* from *Regions*
10. **end if**
11. **end for**
12. **end for**
13. **end procedure**
14. Output ← *Regions*

Although, the researchers have used minimizes intraclass variance. After getting the binarized image, the faulty areas are corrected using Felzenswalb's segmentation[9]. This algorithm uses a graph to categorize pixels based on the color patterns. The goal of Felzenswalb's segmentation is to integrate large potential defective regions and to remove small isolated segments from the thresholding stage. Because heat maps are generated for all types of defects, after the segmentation outcomes a threshold is set to eliminate the expected regions with lower defect likelihood.

C. Parameter Selection

Domain knowledge is required here for the selection of patch size.

Otsu's thresholding concept:

Step1: process the input image

Step 2: obtain image histogram (distribution of pixels)

Step 3: compute the threshold value

Step 4: replace image pixels into white in those regions where the image pixels are greater than the threshold value, else set it to the black color.

IV. Data Description

This experiment uses a broad range of surfaces having diverse textures to see if the proposed method (ASI) can be applied.

A. NEU Surface Defect Dataset

Data is available publically[10], it includes six types of defects of steel rode. Crazing, Inclusion, Patches, Pitted surface, Rolled-in scale and Scratches are the six various defects. The database comprises the labels for all of the images, but not the ground truth of defective regions.

B. Weld Defect Database

This is a public dataset from GDXray images[11], it includes five types of Xray images. Castings, Welds, Baggages, Nature, and Settings are the five types of image categories. Welds was chosen as the dataset for this study since it is connected to surface examination, whereas the others are not. Metal pipe radiographs are included in the welds dataset. The database comprises ten photos with pixel-by-pixel ground truth.

C. Wood Defect Database

This is public dataset[12] of wood defect; it includes two subsets. First is a set of defective samples that have been manually tagged and includes several forms of wood knots. This collection contains 438 photos representing seven different types of knot problems. Second subset is wood board images, where the ground truth is offered by frames. This collection contains 839 board photos representing eighteen different types of defects.

D. Micro-Structure Defect Dataset

This is an industrial dataset of Rolls-Royce; it includes training and testing set of images, images were taken using a microscope from titanium fan blades and have a dimension of 1944 × 2580 pixels. There are 49 defective images and 10 normal images in the training set. There are 10 defective photos and 10 normal images in the test set. Bounding boxes are used to label the defect regions.

V. Experimental Results

The outcomes of the experiments are presented in this section. There are two phases of evaluation. The phase 1 is to evaluate the transferred feature, which includes defect image categorization The proposed segmentation framework will be evaluated in the phase 2.

Here gradient boost classifier with loss of log-likelihood, logistic regression (LR), support vector machine (SVM), and Multi linear regression (MLR) classifiers is been used. For each test phase average accuracy is reported with 20 epochs. For classification, mainly two texture feature extraction approaches applied, multiresolution local binary patterns (MLBP)[13] and grey level co-occurrence matrix (GLCM)[14], are used among all public datasets for testing and comparison. The proposed method (MLR + Decaf) has achieved highest accuracy of 99.27% and that of best-performing benchmark method (MLR + GLCM) having 98.61% accuracy. The results are shown in Fig. 4.

Aside from generic texture characteristics, state-of-the-art handcrafted features for each dataset are chosen as benchmarks. Its purpose is to evaluate the performance of the transferred feature to algorithms that have been built specifically for it.

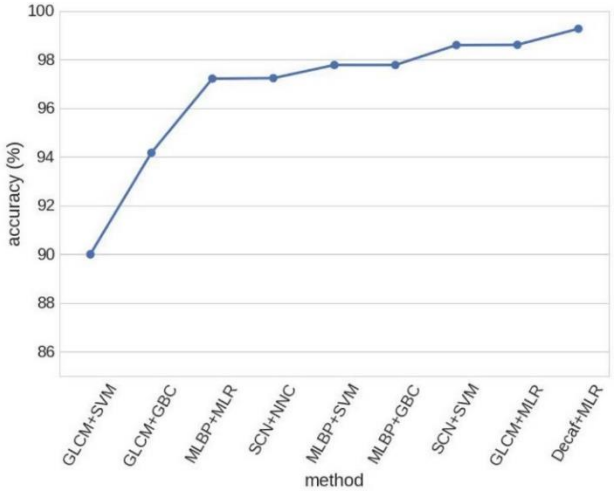


Fig. 4 Classification results of nine methods on NEU surface defect database.

The proposed method on patch sizes 25 × 25, 50 × 50, 75 × 75 has achieved highest accuracies of 90.50%, 88% and 88.50% respectively and that of best- performing benchmark method on patch sizes 25 × 25, 50 × 50, 75 × 75 having 65%, 85% and 80.50% accuracies respectively. The results are shown in Fig.5.

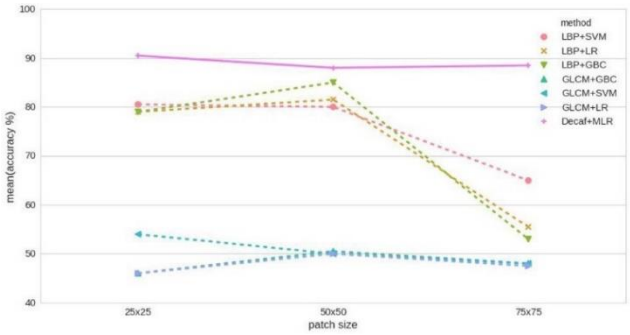


Fig. 5 Classification results of seven methods on weld defect database.

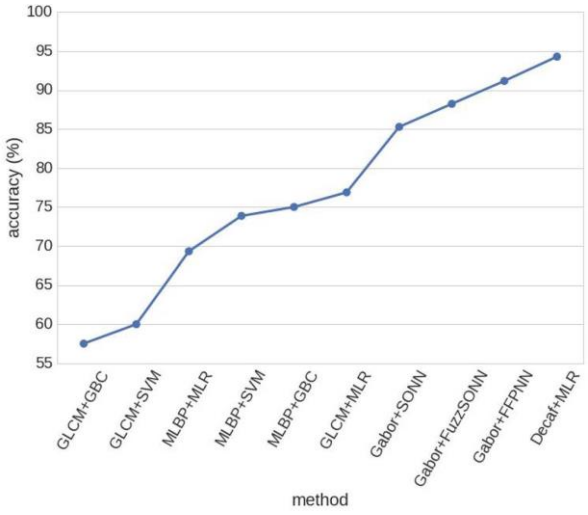


Fig. 6 Classification results of ten methods on wood defect database (subset 1).

The proposed method has achieved highest accuracy of 94.29% and that of best-performing benchmark method having 91.17% accuracy. The results are shown in Fig. 6.

The transferred feature is an elevated representation of the textures in this research, generated from the three classification tasks mentioned above. It outperforms both traditional texture features and hand-crafted features by a wide margin.

VI. Implementation Details

On a workstation with 24 cores and 32GB memory, the proposed framework is implemented. For patch size xx, the computational time of three different feature extraction algorithms. Because to its pixel-wise processing, the computational cost of the LBP-based approach rises exponentially as the patch size increases. Decaf always adjusts input images to predefined dimensions, so the computational cost of every feature is regardless of patch size. The computational cost of feature extraction is higher for small patches than LBP and GLCM-based approaches because the pretrained neural network is dense. The feature extraction step is performed in parallel in order to speed up the process. It takes on average 2 minutes to generate heat maps for defect segmentation of an image containing 2000 patches.

Python is used to implement the algorithms. Scikit-Learn, Numpy, Scipy, Scikit-Image, and Decaf are the packages used in this implementation. The Decaf pre-trained weights are publicly available.

VII. Discussion

For all texture images, the transferred feature surpasses hand-crafted features, according to the experimental results. Moreover, the proposed ASI approach is effective at defect segmentation.

Conclusion

A generic DL-based ASI approach is provided in this research paper. Automated surface inspection method comprises feature transfer from a pretrained DL network and patch classifier convolution over the input image. In classification tasks, the transferred feature outperforms hand-crafted features markedly. For three datasets, the accuracy improvement ranged from 0.66% to 25.50%. In three defect types, the suggested method reduces error escape rate by 6% to 19%. In seven defect types, it increases the accuracy from 2.29% to 9.86% and achieves a 0% error escape rate for the micro structured dataset for segmentation job.

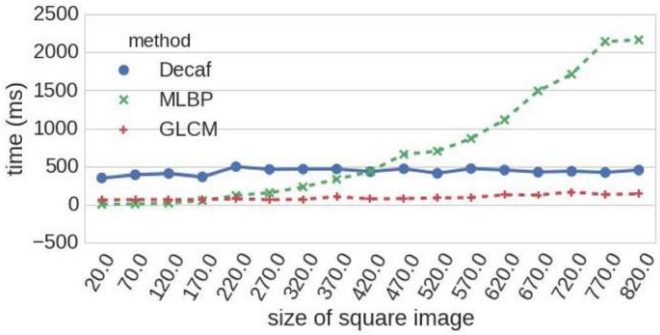
Furthermore, in future researchers will concentrate their efforts in two research areas. One approach is to accelerate the heat map creation process in order to enable real-time defect localization. The second approach is to use machine learning to automate the patch size selection procedure.

References

[1] R. Ren, T. Hung, and K. C. Tan, "A Generic Deep-Learning-Based Approach for Automated Surface Inspection," pp. 1–12, 2017.

[2] "A Review of Recent Advances in Surface Defect Detection using Texture analysis Techniques | ELCVIA: electronic letters on computer vision and image analysis." https://raco.cat/index.php/ELCVIA/article/view/150223 (accessed Apr. 16, 2022).

[3] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, Jan. 2015, doi: 10.1016/J.NEUNET.2014.09.003.



[4] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?"

[5] A. Sharif, R. H. Azizpour, J. Sullivan, and S. Carlsson, "CNN Features off-the-shelf: an Astounding Baseline for Recognition."

[6] "DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition." https://proceedings.mlr.press/v32/donahue14.html (accessed Apr. 16, 2022).

[7] M. Cimpoi, S. Maji, I. Kokkinosécole, S. Mohamed, and A. Vedaldi, "Describing Textures in the Wild," Accessed: Apr. 16, 2022. [Online]. Available: http://www.robots.ox.ac.uk/.

[8] "Otsu, N. (1975) A Threshold Selection Method from Gray-Level Histograms. Automatica, 11, 23-27. - References - Scientific Research Publishing." https://www.scirp.org/(S(lz5mqp453edsnp55rrgjct55))/reference/ReferencesPapers.aspx?ReferenceID=1244209 (accessed Apr. 16, 2022).

[9] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient Graph-Based Image Segmentation," *Int. J. Comput. Vis.*, vol. 59, no. 2, pp. 167–181, 2004, doi: 10.1023/B:VISI.0000022288.19776.77.

[10] K. Song and Y. Yan, "A noise robust method based on completed local binary patterns for hot-rolled steel strip surface defects," *Appl. Surf. Sci.*, vol. 285, pp. 858–864, 2013, doi: https://doi.org/10.1016/j.apsusc.2013.09.002.

[11] D. Mery *et al.*, "GDXray: The Database of X-ray Images for Nondestructive Testing," *J. Nondestruct. Eval.*, vol. 34, no. 4, p. 42, 2015, doi: 10.1007/s10921-015-0315-7.

[12] O. Silvén, M. Niskanen, and H. Kauppinen, "Wood inspection with non-supervised clustering," *Mach. Vis. Appl.*, vol. 13, no. 5, pp. 275–285, 2003, doi: 10.1007/s00138-002-0084-z.

[13] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, 2002, doi: 10.1109/TPAMI.2002.1017623.

[14] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification," *IEEE Trans. Syst. Man. Cybern.*, vol. SMC-3, no. 6, pp. 610–621, 1973, doi: 10.1109/TSMC.1973.4309314.