| | |
|---|---|
| Title | Convolutional Networks for Voting-based Anomaly Classification in Metal Surface Inspection |
| Author(s) | Natarajan, Vidhya; Hung, Tzu-Yi; Vaikundam, Sriram; Chia, Liang-Tien |
| Citation | Natarajan, V., Hung, T.-Y., Vaikundam, S., & Chia, L.-T. (2017). Convolutional Networks for Voting-based Anomaly Classification in Metal Surface Inspection. 2017 IEEE International Conference on Industrial Technology (ICIT), 986-991. |
| Date | 2017 |
| URL | http://hdl.handle.net/10220/42367 |
| Rights | © 2017 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. The published version is available at: [https://doi.org/10.1109/ICIT.2017.7915495]. |

# Convolutional Networks for Voting-based Anomaly Classification in Metal Surface Inspection

Vidhya Natarajan, Tzu-Yi Hung, Sriram Vaikundam
Rolls-Royce@NTU Corporate Lab
Singapore

Liang-Tien Chia
School of Computer Science and Engineering
Nanyang Technological University
Singapore

***Abstract*—Automated Visual Inspection (AVI) systems for metal surface inspection is increasingly used in industries to aid human visual inspectors for classification of possible anomalies. For classification, the challenge lies in having a small and specific dataset that may easily result in over-fitting. As a solution, we propose to use deep Convolutional Neural Networks (ConvNets) learnt from the large ImageNet dataset [9] for image representations via transfer learning. Since a small dataset cannot be used to fine-tune a ConvNet due to overfitting, we also propose a Majority Voting Mechanism (MVM), which fuses the features extracted from the last three layers of ConvNets using Support Vector Machine (SVM) classifiers. This classification framework is effective where no prior knowledge of the best performing ConvNet layers is needed. This also allows flexibility in the choice of ConvNet used for feature extraction. The proposed method not only outperforms state-of-the-art traditional hand-crafted features in terms of classification but also obtains good results compared to other deep features with selected best layers on several anomaly and texture datasets.**

***Keywords—anomaly classification, convolutional neural network, transfer learning, voting, automated visual inspection, metal surface.***

## I. Introduction

Along with the ability to utilize computer vision technologies to consistently and systematically assist in task accomplishment, Automated Visual Inspection (AVI) systems are increasingly popular in industries. In general, AVI for metal surface consists of anomaly detection and classification to localize anomalies and recognize them respectively. While there are numerous techniques used for detection such as keypoint detection [3], phase Fourier reconstruction [15] and template matching [34], this paper aims to handle the challenging task of anomaly classification.

For real-world industrial datasets where anomalies are rare yet a necessity to be identified, these unique and industry-dependent anomaly datasets are typically small [33]. For example, AVI anomaly metal surface inspection dataset contains anomaly patches with a wide range of lighting conditions and different classes of anomalies also have a large intra-class variance and small inter-class variance as shown in Fig. 1. Hence, an effective anomaly classification framework is required as an apt solution for such a complex and challenging anomaly dataset.

Typical feature extraction techniques include Gabor filter bank [4], Scale Invariant Feature Transform (SIFT) [3] and Blur feature [5]. These popular traditional features have achieved comparable classification results in numerous domains besides anomaly classification [11], [17, 18, 19]. However, for small datasets, a discriminative feature extractor is necessary to avoid overfitting.

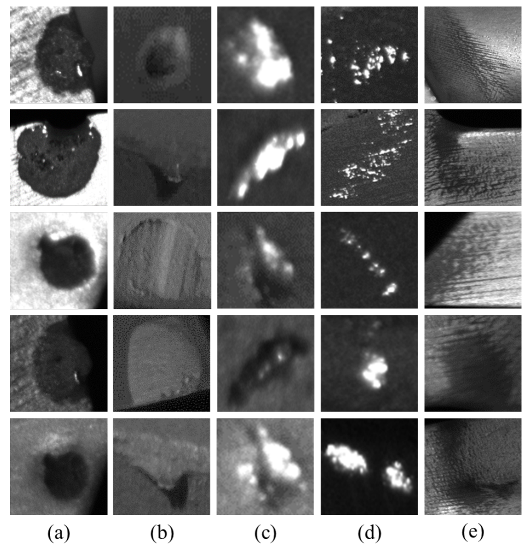



Fig. 1. Types of Anomalies in AVI Anomaly Dataset: (a) Melt, (b) Plusmetal (images have been adjusted for visualization), (c) Scratch, (d) Scuff and (e) Shadowing

While small datasets suffer from the need to have discriminative features, recently, large datasets such as ImageNet, have attained state-of-the-art performance by utilizing variations of Convolutional Neural Networks (ConvNets) for feature extraction and classification [7, 8], [16], [19], [27]. Literature has shown that ConvNets trained on these large datasets surpass the performance of traditional hand-crafted features in numerous classification, recognition and segmentation tasks [1, 2], [13, 14].

A contribution of this paper lies in leveraging the benefits of ConvNets which learn general image representations through large datasets for tasks on small datasets. Through transfer learning, these deep ConvNets can be used as strong feature extractors for small datasets. While benefitting from using pre-learnt ConvNet features, it is challenging to fine-tune the classification layer using small datasets due to overfitting. In addition, the discriminative power of the ConvNet layers is inconsistent for small datasets.

Hence, in this paper, we propose to use deep ConvNets for image representation via transfer learning. Therefore, the last three layers, excluding the softmax and last fully connected layers, are extracted as features. Furthermore, to have a stable classification solution, we propose a Majority Voting Mechanism (MVM) by fusing the features extracted from the last three layers before the classification layer using Support Vector Machine (SVM) classifiers. This framework can be used to attain robust classification without any prior knowledge of the best performing ConvNet and its layer. In addition, the flexibility of the proposed framework allows any ConvNet such as those of Visual Geometry Group (VGG) to be used for this purpose.

Our experimental results show that the proposed classification framework substantially improves the performance for small datasets and eliminates overfitting. Besides surpassing the performance of traditional hand-crafted features, the proposed method also performs well when compared to other deep features with selected best layers. The proposed approach shows promising performance when applied to metal surface anomaly datasets such as NEU [19] and AVI [15], [34] and other benchmark texture on material datasets such as CUReT [28] and KTH-TIPS2 [31].

## II. RELATED WORK

### A. Transfer Learning

Typically, traditional hand-crafted features such as Grayscale Histogram, Blur [5], Gabor filter bank [4] and DenseSIFT [17] are used to extract pixel intensity, blur regions, orientation and local descriptors respectively from images. In addition, these features can be represented with Spatial Pyramid Matching (SPM) [24], Sparse Coding (SC) [25] and Improved Fisher Vector (IFV) [18] to enhance performance. While Cimpoi et al. claims that combinations of hand-crafted features achieve comparable performance to deep ConvNets when tested on large datasets [11, 12], this cannot be ensured for the case with small datasets. Through transfer learning, small datasets utilize the advantages of the learnt ConvNet features by substituting them for traditional feature extractors [20, 21], [26].

In literature, numerous ConvNets trained on large datasets such as ImageNet have been discussed with different architectures [7, 8], [10]. The benefit of learning ConvNet features from a large dataset consisting of a wide range of classes is that general-purpose features can be learnt in the layers. Features such as orientation that are extracted from the layers of these ConvNets can be transferred to other datasets in different domains using techniques such as parameter tuning [13, 14], [22, 23]. For instance, Oquab et al. [22] proposed a transfer learning method which fine-tunes parameters of ConvNets so as to use the learnt models on other datasets. This method retrains the ConvNet softmax and last fully-connected layers to provide a model of better fit when applied to another new dataset that is relatively large.

Despite the rise in such fine-tuned ConvNets achieving increasing accuracy in classification tasks [7], fine-tuning is not suitable for small datasets due to overfitting of the ConvNet parameters. While the domains of these datasets are different from those used for training the ConvNet, it can still be used as feature extractors through transfer learning [7], [21], [23]. For instance, Cimpoi et al. proved that fine-tuning is not necessary for these features to be transferred to other datasets [27]. Furthermore, high classification accuracies have been noticed by Razavian et al. [21] when a ConvNet trained on the ImageNet dataset is used as feature extractors for scene datasets.

### B. Anomaly Classification

For anomaly classification of small datasets, most works in literature have been discussed on using traditional hand-crafted features for feature extraction. For the NEU dataset with defects on metal surfaces, Song et al. showed that using Scattering Convolutional Network (SCN), a type of ConvNet, can achieve comparable results [19]. However, Pan et al. later proved that traditional features such as DenseSIFT+IFV and Local SIFT Pattern (LSP) performed slightly better than SCN [18]. However, the performance of these traditional features only attributed to a marginal increase in performance. Similar to the small NEU defect dataset, Haddad et al. handled anomaly classification in semiconductor units using traditional hand-crafted features [33]. While traditional features are proposed for anomaly classification of small dataset, Patrick et al. claims that OverFeat, another ConvNet, performs well for the same NEU dataset [26]. Even though ConvNets have been explored for anomaly classification, only a single pre-selected ConvNet layer such as the first fully connected layer in OverFeat is always used for classification. In addition, these selected ConvNet layers are trained a single linear SVM classifier.

In our work, we propose to utilize the benefits of transfer learning from deep ConvNets for anomaly classification on small datasets. Differing from the standard approach for transfer learning [22], [27], we propose MVM to combine features from multiple layers of a ConvNet using SVM classifiers instead of a single pre-defined layer. In addition to providing stability in the discriminative power of the features, no prior knowledge of the best performing layer in any deep ConvNet is required.

## III. ANOMALY CLASSIFICATION USING DEEP CONVNETS

This paper proposes a classification framework using transfer learning from deep ConvNets for feature extraction in an anomaly classification task. Fig. 2 shows the proposed framework that extracts multiple deep features using ConvNets learnt from ImageNet, followed by classification using linear SVM classifiers. Eventually, a Majority Voting Mechanism (MVM) is also proposed as a generic classification framework to combine the extracted features using SVM classifiers.

### A. Deep ConvNets

With an influx in benefits and architectures of ConvNets trained from ImageNet, there is a wide range of ConvNets

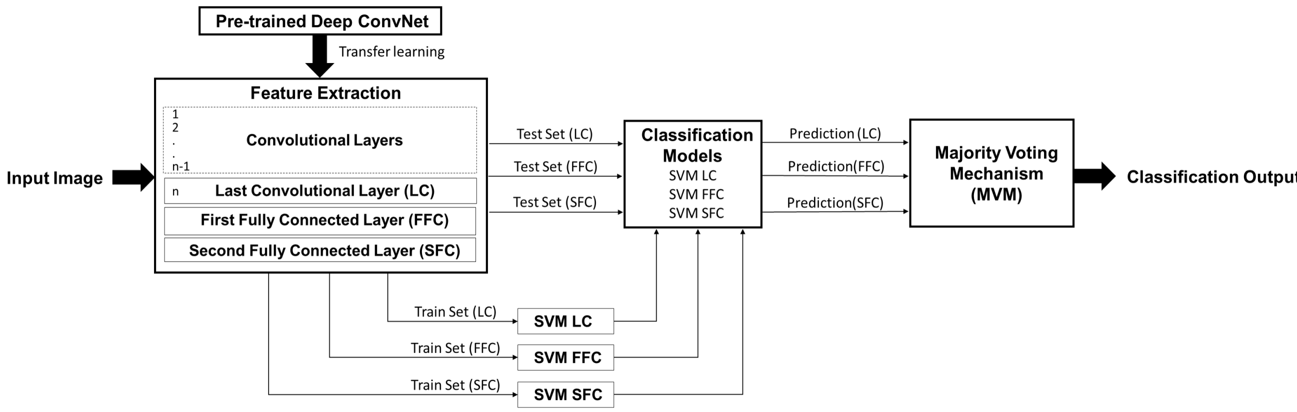



Fig. 2. Proposed Classification Framework using Deep ConvNets

that has generic features learnt in the layers. This provides numerous options for choosing a ConvNet to extract features for small datasets in our framework. From simple architectures such as LeNet, AlexNet and VGG models [6], [8], [23], to complex architectures such as GoogLeNet [10], this proposed framework allows any of these models to be used for feature extraction. Hence, a benefit from this classification framework lies in pre-training the ConvNet using a large dataset, where any ConvNet can be used for feature extraction.

While the framework allows any ConvNet to be used, we propose to use VGG models as a basis for comparison since classification is shown to increase using features extracted from deeper layers of a ConvNet [9], [10], [21]. These models consist of different architectures that follow traditional ConvNet architectures by LeNet and AlexNet [6], [8]. They include VGG-S, VGG-F, VGG-M, VGG-M-128, VGG-M-1024, VGG-M-2048, VGG-16 and VGG-19 as trained by Simonyan et al. [8] and Chatfield et al. [23]. Furthermore, VGG models perform better compared to other ConvNets such as DeCAF and AlexNet when tested on larger datasets [16], [23], [27]. Hence, in the experiments, we use VGG ConvNets as default.

## *B. Transfering Features from Deep ConvNets*

Traditional features and encoding methods such as DenseSIFT+IFV have shown promising results for classification of large texture and defect datasets [16], [18, 19], [27]. When there are many samples, highly discriminative features can be extracted and generic classifiers can be trained. However, for small datasets, it is challenging to learn generic features and classifiers using simple traditional features and encoders. Instead, strong features such as those in the layers of deep ConvNets trained using large datasets are required. However, ConvNets trained specifically using small datasets also suffer from overfitting.

A ConvNet requires a large dataset to ensure that the model is learnt to fit the dataset without over-fitting since numerous parameters need to be tuned. Instead of training ConvNets from scratch, those trained on large datasets such as ImageNet can be used to extract features of other datasets. Typically, for any trained ConvNet used in transfer learning, the classification layer is retrained using the dataset of the task. However for small datasets, fine-tuning causes overfitting. Therefore, we remove the last fully connected classification and softmax layers of the network and utilize the remaining ConvNet for feature extraction.

## *C. Multi-layered Deep Feature Extraction*

Deep features from various deep ConvNet layers are used as the discriminative power of layers vary for small datasets. Razavian et al. [21] proved that classification increases as features from deeper layers are extracted. On the other hand, Cimpoi et al. showed that features from the second fully connected (SFC) layer performed consistently on large texture on material datasets [27]. However, this is not the case for smaller datasets as the most discriminative feature alternates between the first fully connected (FFC) and SFC layers, both of whom represent the global image. Furthermore, the convolutional properties of the last convolutional (LC) layer in the local region of an image allows it to be interpreted as a local image descriptor [27]. To ensure that both the local and global image descriptors are represented in the feature extraction, deeper layers from the LC, FFC and SFC layers are used for feature extraction.

## *D. Multi-layered Deep Feature Classification*

ConvNet features complemented by SVM classifiers have shown to have improved the classification of datasets similar to ImageNet [21]. Besides allowing discriminative classifiers to be learnt without overfitting despite a small dataset, SVM also enables comparison with prior classification work [11], [18], [27]. Hence, these classifiers are used to train each of the extracted deep features [18].

Three linear SVM classifiers, namely SVM LC, SVM FFC and SVM SFC, are used to train the deep features from the LC, FFC and SFC layers respectively. These classifiers are used to classify the features extracted from the corresponding layers using the test data. The predicted labels from each of the layers, Predictions LC, FFC and SFC, are solely based on the features extracted and classifiers trained on that particular layer.

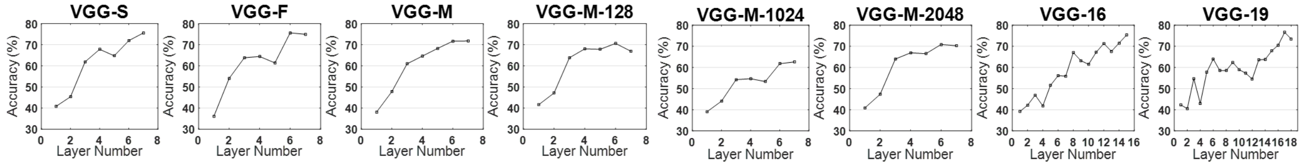



Fig. 3. Classification Accuracies across Layers of Deep ConvNets

Besides being discriminative in a small dataset, these SVM classifiers allow a controlled comparison with prior work for public defect and texture datasets.

### *E. Majority Voting Mechanism*

While the performance of SFC layer is consistent in large datasets [27], the best performing layer is entirely dependent on the dataset for small datasets. Despite this volatility in performance between the FFC and SFC, the SVM classifier has to be robust to handle this instability. Hence, MVM is also proposed by utilizing the features and classifiers from all the chosen layers to determine the classification output.

In this proposed mechanism, the majority of all predictions LC, FFC and SFC is assigned as the predicted label. A conditional priority is also assigned to a layer that has the best average performance across most of the ConvNets when all the Predictions are different and there is no majority. Based on experiments, we propose to assign prediction FFC as the conditional priority.

In addition, MVM is also used to enhance the classification results when there is no prior knowledge of the best performing layer in any deep ConvNet which is required in our framework that allows any ConvNet to be used. Since the best performing layer of a ConvNet is dependent on a dataset, especially for small datasets, this mechanism increases efficiency by mitigating the need to determine the best performing layer.

## IV. Experimental Results

This section compares the effectiveness of the proposed classification framework across datasets from three domains, namely industrial AVI anomaly dataset, public defect and benchmark texture on material datasets. All the datasets, dominate with texture characteristics on materials. Every input image is resized to 224x224x3 to comply with the architecture of the selected ConvNets. For most ConvNets, the feature vectors are of dimensions 14x14x512, 1x4096 and 1x4096 respectively for the LC, FFC and SFC layers. On the other hand, for VGG-M-128, VGG-M-1024 and VGG-M-2048, the features from their respective SFC layers are 1x128, 1x1024 and 1x2048. In order to train a strong classifier for the features extracted from small datasets, input data is split into train and test sets based on the experimental settings of each dataset. The average multiclass classification accuracy is used as the metric to compare the effectiveness of the proposed framework.

### *A. AVI Anomaly Dataset*

AVI dataset consists of localized anomaly patches from five unique classes namely Melt, Plusmetal, Scratch, Scuff and Shadowing as shown Fig. 1[1]. There are 30, 40, 20, 180 and 50 samples of anomalies in each of the respective classes. This small dataset has an uneven number of anomaly samples in each class, which is a challenge for the classification task. Furthermore, the inter-class variance between classes of anomalies such as Scuff and Scratch is low. On the other hand, the intra-class variance in each anomaly class is very high such as the samples in Plusmetal that are very different from each other. The experiments are evaluated for AVI anomaly dataset based on the average classification accuracy over 10 runs. For each run, the dataset is split randomly into train and test data, where half the dataset is used for training.

#### *1) Classification across Layers of Deep ConvNets*

The performance of the features extracted at each of the ConvNet layers is compared for the eight selected deep networks as shown in Fig. 3. A linear SVM classifier is used to depict the discriminative capabilities of each ConvNet layer on this small dataset. An upward trend in the performance is noted across all the ConvNets as deep features are extracted from higher layers. It can be observed that the best performance are typically from the last three layers of each ConvNet, which correspond to the LC, FFC and SFC layers. Similar to the conclusion by Razavian et al. [21], features from higher layers also perform well in the classification of the challenging AVI anomaly dataset. Hence, it is effective to use these three layers to extract deep features in the proposed framework.

#### *2) Comparison of Deep Features and MVM*

The performance of the deep features extracted from the LC, FFC and SFC layers and MVM using each of these layers as priority are compared in Table I. The best performing layer mostly alternates between FFC and SFC for these ConvNets in this dataset. However, this instability in performance can be tackled using the proposed MVM as it requires no prior knowledge on the best performing layer. For most ConvNets, MVM performs as the best or a close second to its respective best layer. For instance, the best deep feature, FFC of VGG-19,

TABLE I. Classification Accuracy of Deep Features and Corresponding MVM in Percentage

| Deep ConvNet | LC | FFC | SFC | MVM (LC) | MVM (FFC) | MVM (SFC) |
|---|---|---|---|---|---|---|
| VGG-S | 64.73 | 71.92 | **75.56** | 73.29 | 73.19 | 74.09 |
| VGG-F | 61.38 | **75.60** | 74.92 | 73.90 | 75.12 | 74.20 |
| VGG-M | 68.23 | 71.66 | 70.54 | 72.34 | **72.54** | 72.44 |
| VGG-M-128 | 67.83 | 70.61 | 66.89 | 70.26 | **70.86** | 70.46 |
| VGG-M-1024 | 53.32 | 61.78 | **62.60** | 59.88 | 59.38 | 59.12 |
| VGG-M-2048 | 66.49 | 70.87 | 70.27 | **71.16** | 70.93 | 71.03 |
| VGG-16 | 67.61 | 71.52 | **75.37** | 73.09 | 73.59 | 73.19 |
| VGG-19 | 70.50 | **76.59** | 73.46 | 74.80 | 75.26 | 74.56 |
| Average | 65.01 | 71.32 | 71.20 | 71.09 | **71.36** | 71.14 |

[1]Limited disclosure due to confidentiality reasons

TABLE II. CLASSIFICATION ACCURACY OF TRADITIONAL FEATURES AND BASELINE OF PROPOSED CLASSIFICATION FRAMEWORK IN PERCENTAGE

| Feature | Accuracy |
|---|---|
| Blur | 35.63 |
| Gabor Filter Bank | 34.98 |
| Gabor Filter Bank + IFV | 33.99 |
| Grayscale Histogram | 39.50 |
| Grayscale Histogram + SPM | 50.76 |
| DenseSIFT | 35.16 |
| DenseSIFT + SC | 39.18 |
| DenseSIFT + IFV | 45.18 |
| DenseSIFT + SC + SPM | 51.42 |
| MVM-VGG-19 | **75.26** |

achieves an accuracy of 76.59%, while its corresponding best MVM with conditional priority for prediction FFC, MVM (FFC), has an accuracy of 75.26%. Furthermore, MVM (FFC) performs the best when compared to other single layered deep features and the other two MVM with different conditional priorities in terms of average accuracy across all ConvNets. Therefore, we use MVM (FFC) with VGG-19 which achieves the best MVM (FFC) performance across all the ConvNets, denoted by MVM-VGG-19, as the baseline for comparison in the following experiments.

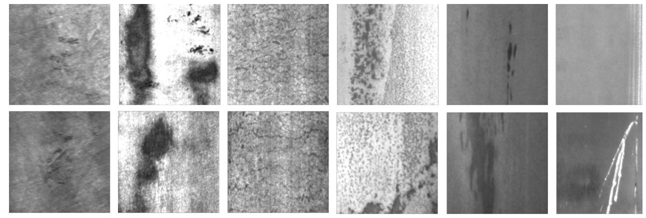
*3) Comparison with Traditional Hand-Crafted Features*

Traditional features such as Grayscale Histogram, Blur [5], Gabor filter bank [4] and DenseSIFT [17], with techniques such as SPM [24], SC [25] and IFV [18] are compared with the baseline as shown in Table II. The performance of these hand-crafted features using a linear SVM classifier is less than 52% in terms of accuracy. However, the baseline MVM-VGG-19 achieves 75.26% which surpasses those of traditional features by at least 23%. It is worth mentioning that using deep features across all the ConvNets outperform the best hand-crafted feature. These experiments show that for small datasets that have anomalies with different inter and intra class variances, the proposed framework is more discriminative than the traditional features.

The baseline from our framework, MVM-VGG-19 with a conditional priority to prediction FFC, is further analyzed for classification on other public datasets of defects on metal surface and texture on materials.

## *B. NEU Defect Dataset*

The NEU dataset [19] is also a small texture-like defect dataset consisting of 300 localized defect images for each of the six defect classes on steel surfaces as shown in Fig. 4. It can be noted that this public dataset has more training samples for each class than those from the most dominant class in AVI



(a) (b) (c) (d) (e) (f)

Fig. 4. Types of Defect in NEU Defect Dataset: (a) Rolled-in-Scales, (b) Patches, (c) Crazing, (d) Pitted Surface, (e) Inclusion and (f) Scratches

anomaly dataset. While DenseSIFT+IFV has a high accuracy on this dataset, this is not so for AVI dataset [18]. Hence, the baseline MVM-VGG-19 is compared with the state-of-the-art results using the experimental settings as that of Pan et al. [18] as shown in Table III. In addition to surpassing the performance of other ConvNet such as OverFeat [26] and SCN [19], the MVM-VGG-19 also performed comparably to other state-of-the-art traditional features like Local Binary Pattern (LBP), LSP and DenseSIFT+IFV. It can be observed that for this defect dataset, both the traditional features and the proposed baseline achieved a high classification. Despite the lack of prior information on the best performing ConvNet layer, the proposed mechanism performs with stability when applied to this dataset.

TABLE III. CLASSIFICATION ACCURACIES FOR NEU DEFECT DATASET IN PERCENTAGE

| Feature | Accuracy |
|---|---|
| SCN [19] | 98.60 |
| LBP [26] | 97.93 |
| OverFeat [26] | 98.70 |
| DenseSIFT+IFV [18] | 99.86 |
| LSP [18] | 99.87 |
| MVM-VGG-19 | **99.50** |

## *C. Texture on Material Datasets*

CUReT [28], UMD [32], UIUC [29], KTH-TIPS [30], KTH-TIPS2 [31] are some large benchmark datasets used for texture classification on materials. While these public datasets have texture properties similar to that on metal surface in the anomaly and defect dataset, they are also much larger and have illumination invariance. Using the respective experimental settings for each dataset, the performance of MVM-VGG-19 is compared against state-of-the-art ConvNet features as shown in Table IV [11], [27]. MVM-VGG-19 performs reasonably well when compared to the prior methods for almost all the datasets. Even for KTH-TIPS-2b, the performance of MVM-VGG-19 is comparable to its best performance. While pre-selected single layers are used for feature extraction in previous methods, the proposed method requires no prior knowledge on the performance of single layers. Unlike the features from the pre-selected single layer that uses global image descriptors, our proposed method fuses features from three layers which consist of both global and local image descriptors. Hence, the proposed framework is effective for classification on these datasets.

TABLE IV. CLASSIFICATION ACCURACIES FOR TEXTURE ON MATERIAL DATASETS IN PERCENTAGE

| Dataset | DeCAF [11] | AlexNet [27] | VGGM [27] | VGG19 [27] | MVM-VGG-19 |
|---|---|---|---|---|---|
| CUReT | 97.9 ± 0.4 | 94.4 ± 0.4 | 94.2 ± 0.3 | 94.5 ± 0.4 | **98.7 ± 0.6** |
| UMD | 96.4 ± 0.7 | 95.9 ± 0.9 | 97.2 ± 0.9 | 97.7 ± 0.7 | **99.1 ± 1.3** |
| UIUC | 94.2 ± 1.1 | 91.1 ± 1.7 | 94.5 ± 1.4 | 97.0 ± 0.7 | **98.0 ± 1.6** |
| KT | 96.9 ± 0.9 | 95.5 ± 1.3 | 96.1 ± 0.9 | 97.9 ± 0.9 | **98.6 ± 2.8** |
| KTH-2a | 78.4 ± 2.0 | - | - | - | **84.6 ± 3.3** |
| KTH-2b | 70.7 ± 1.6 | 71.5 ± 1.3 | 71.0 ± 2.3 | **75.4 ± 1.5** | 73.4 ± 4.1 |

## V. Conclusion

In this paper, we have proposed a flexible multi-layered deep feature extraction from ConvNets via transfer learning as part of a classification framework for small datasets. The flexibility of the proposed method allows any ConvNet to be used for feature extraction. In addition, a Majority Voting Mechanism (MVM) is also proposed to overcome the problems of overfitting from small datasets by fusing multi-layered deep features using linear SVM classifiers. Furthermore, this method requires no prior knowledge of the best performing ConvNet layer. Remarkable performance on challenging anomaly and texture datasets can be observed when compared to traditional hand-crafted features and other ConvNets with pre-selected layers.

## Acknowledgment

This work was conducted within the Rolls-Royce@NTU Corporate Lab with support from the National Research Foundation (NRF) Singapore under the Corp Lab@University Scheme.

## References

[1] Z. Sun, G. Bebis and R. Miller, "On-road vehicle detection using Gabor filters and support vector machines," International Conference on Digital Signal Processing, 2002.

[2] A. K. Jain and F. Farrokhnia, "Unsupervised texture segmentation using Gabor filters," IEEE International Conference on Systems, Man and Cybernetics, 1990.

[3] D.G. Lowe, "Distinctive image features from scale-invariant keypoints," International Journal of Computer Vision, vol. 60, no. 2, pp. 91-110, 2004.

[4] D.C. Choi, Y.J.Jeon, J.P. Yun and S.W.Kim, "Pinhole detection in steel slab images using Gabor filter and morphological features", Applied Optics, vol. 5, no. 26, pp. 5122-5129, 2011.

[5] J. Shi, L. Xu and J. Jia, "Discriminative Blur Detection Features," IEEE International Conference on Computer Vision and Pattern Recognition, 2014.

[6] Y. LeCun, L. Bottou, Y. BenSgio and P. Haffner, "Gradient-Based Learning Applied to Document Recognition," Proc. IEEE, vol. 86, no. 11, pp. 2278-2324, 1998.

[7] A. Krizhevsky, I. Sutskever and G. E. Hinton, "ImageNet classification with deep convolutional neural networks." Advances in Neural Information Processing Systems 25, 2012.

[8] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," IEEE International Conference on Learning Representations, 2014.

[9] J. Deng, W. Dong, R. Socher, L. J. Li, Kai Li and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," IEEE International Conference on Computer Vision and Pattern Recognition, 2009.

[10] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke and A. Rabinovich, "Going deeper with convolutions," IEEE International Conference on Computer Vision and Pattern Recognition, 2014.

[11] M. Cimpoi, S. Maji, I. Kokkinos, S. Mohamed and A. Vedaldi, "Describing Textures in the Wild," IEEE International Conference on Computer Vision and Pattern Recognition, 2014.

[12] F. Perronnin, J. Sa ́nchez, and T. Mensink, "Improving the Fisher kernel for large-scale image classification." European Conference on Computer Vision, 2010.

[13] R. Girshick, J. Donahue, T. Darrell and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," IEEE International Conference on Computer Vision and Pattern Recognition, 2014.

[14] B. Hariharan, P. Arbel ́aez, R. Girshick and J. Malik, "Simultaneous detection and segmentation," European Conference on Computer Vision, 2014.

[15] T.Y. Hung, S. Vaikundam, V. Natarajan and L.T. Chiat, "Phase Fourier Reconstruction for Anomaly Detection on Metal Surface Using Salient Irregularity", MultiMedia Modeling, 2017.

[16] M Cimpoi, S Maji and A Vedaldi, "Deep filter banks for texture recognition and segmentation", IEEE Conference on Computer Vision and Pattern Recognition, 2015.

[17] L. Sharan, R. Rosenholtz and E. H. Adelson, "Material perceprion: What can you see in a brief glance?" Journal of Vision, 9:784(8), 2009.

[18] S.Y. Pan, T.Y Hung and L.T. Chia, "Using Material Classification Methods for Steel Surface Defect Inspection", International Symposium on Industrial Electronics, 2016.

[19] K.C. Song, S.P. Hu and Y.H. Yan, "Automatic Recognition of Surface Defects on Hot-Rolled Steel Strip Using Scattering Convolution Network." Journal of Computational Information Systems, vol. 10, no.7, pp. 3049-3055, 2014.

[20] J. Yosinski, J. Clune, Y. Bengio, H. Lipson, Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence and K. Weinberger, "How transferable are features in deep neural networks?" Advances in Neural Information Processing Systems 2, 2014.

[21] A. S. Razavin, H. Azizpour, J. Sullivan and S. Carlsson, "CNN Features off-the-shelf: an Astounding Baseline for Recognition.", DeepVision workshop, 2014.

[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] K. Chatfield, K. Simonyan, A. Vedaldi and A. Zisserman, "A Return of the devil in the details: Delving deep into convolutional nets." British Machine Vision Conference, 2014.

[24] S. Lazebnik, C. Schmid and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories." IEEE International Conference on Computer Vision and Pattern Recognition, vol. 2, pp 2169–2178, 2006.

[25] J. Yang, K. Yu, Y. Gong and T. Huang, "Linear Spatial Pyramid Matching Using Sparse Coding for Image Classification," IEEE International Conference on Computer Vision and Pattern Recognition, 2009.

[26] P. H. Chen and S. S. Ho, "Is overfeat useful for image-based surface defect classification tasks?," IEEE International Conference on Image Processing, 2016.

[27] M. Cimpoi, S. Maji, I. Kokkinos and A. Vedaldi, "Deep filter banks for texture recognition, description, and segmentation." International Journal of Computer Vision, vol. 118, no. 1, pp. 65–94, 2016.

[28] K. J. Dana, B. van Ginneken, S. K. Nayar and J. J. Koenderink, "Reflectance and texture of real world surfaces," ACM Transactions on Graphics, vol. 18, no. 1, pp. 1–34, 1999.

[29] S. Lazebnik, C. Schmid and J. Ponce, "A sparse texture representation using local affine regions," IEEE Transactions on Pattern Analysis and Machine Learning, vol. 28, no. 8, pp. 2169–2178, 2005.

[30] E. Hayman, B. Caputo, M. Fritz and J.-O. Eklundh, "On the significance of real-world conditions for material classification," European Conference on Computer Vision, 2004.

[31] B. Caputo, E. Hayman and P. Mallikarjuna, "Class-specific material categorisation," International Conference on Computer Vision, 2005.

[32] Y. Xu, H. Ji and C. Fermuller, "Viewpoint invariant texture description using fractal analysis," International Journal of Computer Vision, vol. 83, no. 1, pp. 85–100, 2009.

[33] B. M. Haddad, S. Yang, L. J. Karam, J. Ye, N. S. Patel and M. W. Braun, "Multifeature, Sparse-Based Approach for Defects Detection and Classification in Semiconductor Units," IEEE Transactions on Automation Science and Engineering , vol.PP, no.99, pp.1-15, 2016.

[34] S. Vaikundam, T.Y. Hung and L.T. Chia, "Anomaly region detection and localizationin metal surface inspection," IEEE International Conference on Image Processing, 2016.