# MATH6380P - Project 2: Anomaly Detection using Transfer Learning in Semiconductors

**Tuan-Anh Vu (ID: 20672378)**
Department of Computer Science and Engineering
The Hong Kong University of Science and Technology
Hong Kong SAR
tavu@connect.ust.hk

## Abstract

Convolutional Neural Network (CNN) techniques have proven to be very useful in image-based anomaly detection applications. CNN can be used as deep features extractor where other anomaly detection techniques are applied on these features. For this scenario, using transfer learning is common since pretrained models provide deep feature representations that are useful for anomaly detection tasks. Consequentially, anomaly can be detected by applying similarly measure between extracted features and a defined model of normality. In this study, we introduce a transfer learning framework for anomaly detection and show that with the proposed threshold settings, a significant performance improvement can be achieved.

## 1 Introduction

Machine learning-based approaches have been widely used for anomaly and outlier detection in several fields such as data networks [1–8]. wireless sensor networks [9–11] and manufacturing [12, 13]. Accurate industrial inspection is one of the main challenges in the manufacturing industry[14]; An efficient inspection process will help to reduce the overall manufacturing cost while maintain quality requirements. Computer-vision based approaches that use convolutional neural network (CNN) for anomaly detection in manufactured surface textures are commonly used in industrial quality inspection due to their higher accuracy. Besides anomaly detection accuracy, computing requirements in terms of time and memory are other important factors that should be considered as well. Anomaly detection can be generally defined as “the task of recognising that test data differ in some respect from the data that are available during training”[15]. In CNN-based anomaly detection, this definition assumes that training is conducted using normal images only to build a model that defines normality. This assumption meets practical requirements and matches real-world situations when anomalies samples are usually not available or insufficient to model the anomalous behaviors. Accordingly, anomaly detection techniques can be categorized into the following categories[15]: Probabilistic approaches, distance-based anomaly detection, reconstruction-based anomaly detection, domain-based detection approaches and information-theoretic techniques. Probabilistic approaches involve estimating the probability density function (pdf) of the training data. Hence, the resultant distribution models the normality, anomaly then can be detected by setting a threshold that defines the normal boundary. Distance-based approaches use distance metrics to measure the similarity between evaluation data and a model of normality. reconstruction-based techniques attempt to reconstruct a copy of the evaluation data according to learned model of normality. Hence, anomaly can be detected by comparing the input with its reconstruction. Domain-based approaches define normal domain by creating boundary of normality based on training data. Thus, evaluation data that fall outside the domain are classified as anomalies. Information-theoretic methods utilize the fact that anomaly presence in the data would alter its information-content and hence, model of normality is defined using computed information-content of training data. Consequentially, anomaly is detected by applying similarly measures between extracted information-contact of evaluation data and normality model. A

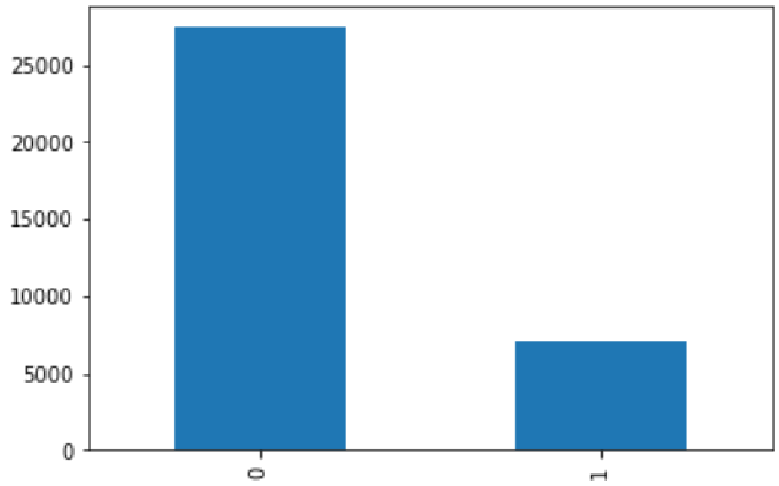


Figure 1: Class imbalanced

common theme among these categories is that they attempt to build a Model of Normality (MoN) using useful deep representations extracted from normal training data, anomaly then can be detected by measuring similarity between deep presentations of evaluation data and the MoN. Extracting useful feature presentations from images can be achieved by utilizing transfer learning[16–19] or by training a CNN from scratch. In contrast to transfer learning, training CNN model from scratch requires availability of training set consists of anomaly-free images. The size of the training set should be large enough to serve training purpose. In many real-world applications, such requirement is not possible or expensive to achieve. This fundamental problem of insufficient training data can be solved by using transfer learning to obtain deep feature representations from models pre-trained on large datasets without constraints on the relation between these training datasets and the target data[17] . Here, knowledge is transferred from the pre-trained model “source domain” to the subject model “target domain”[17]. Transfer learning has been widely used for different anomaly detection applications with very promising results [20–25].

The paper is outlined as follows. The next parts of this section provide literature review on related work and the dataset. Section 2 presents the anomaly detection approach. Section 3 presents experimental setup and discussion, respectively, the paper is finally concluded in Section 4.

**Related Works** Employing similarity measures with normality model for detecting anomaly in images has been widely adopted in industrial applications due to its effectiveness in detecting anomalous patterns. For example, P. Napoletano *et al.*, (2018) [26] and D. Carrera *et al.*, (2017) [27] use region-based methods to define normality, where a “dictionary” of normality is created from subregions “patches” taken from anomaly-free images. Anomalies subregions of a given image are detected through comparison with normal subregions of the dictionary. Since such methods are region-based, they allow for anomaly localization as well. Other approaches [28–36] use sparse representations of normal data to learn normality. In O. Rippel *et al.*, (2020) [20] and P. Christiansen *et al.*, (2016) [37], a Gaussian distribution was fitted to the deep feature representations of normal data to establish the normality model. Subsequently, the Mahalanobis distance [38] is used for anomaly detection. In B. Staar *et al.*, (2018) [39], a prototype of normal feature representations is created by averaging the learned deep representations of normal images. Here, instead of using normal images directly to define normality, “deep” normality is learned from deep feature representations of these normal images. Euclidean distance was used to measure the similarly between prototype and feature representation of evaluation images. J. Liu *et al.*, (2020) [40] propose an encoder-decoder-encoder CNN structure for anomaly detection in industrial product surface images. They introduced a dual prototype loss approach to encourage features vectors generated by the encoders to keep closer to their own prototype. Consequently, the mean square error between the features vectors is used as indicator of anomalies.

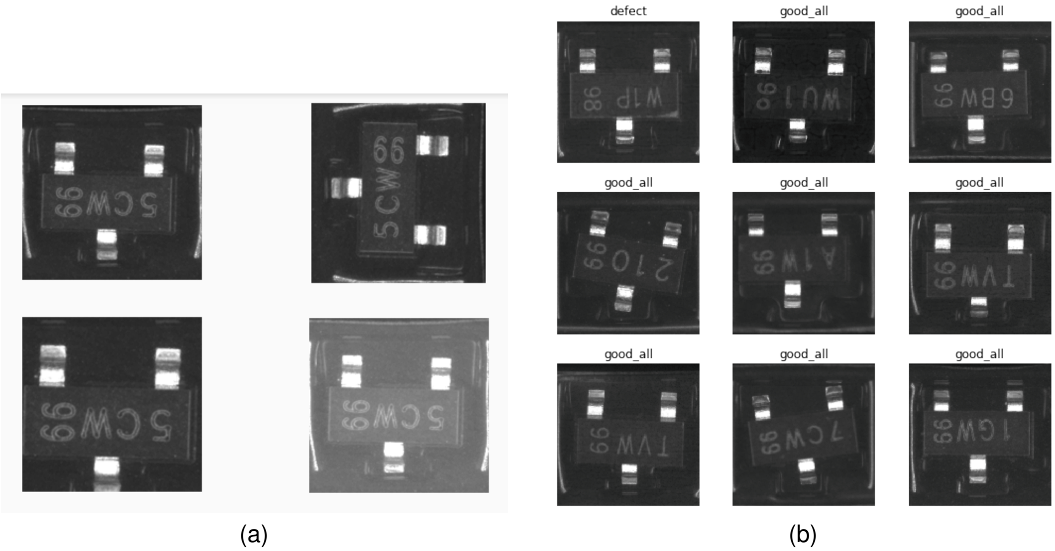
**Nexperia Dataset (In-class Kaggle Competition)** Nexperia is one of the biggest Semi-conductor manufacturers in the world. They produce billions of semi-conductors every year. Meanwhile, a lot of unqualified devices are mixed with the good ones. Mass production makes it difficult for human workers to examine all of the products. Therefore, we would like to use deep learning methods to help Nexperia pick out as many defect devices as possible while preserving the good ones, thus improving their yield rate. The Nexperia Dataset consists of 34,459 labeled images of semiconductors where 27,420 images in the dataset are “good” and only the remaining 7,039 are labeled as “defect”, thus, this problem dictates using techniques to tackle the inherent class imbalance, Fig. 1. To this end, we use image transforms to augment our dataset, oversample from the “defect” class and perform test time augmentation to produce predictions. The figure below shows the confusion matrix when a pre-trained model is finetuned on the dataset. A very high false positive rate is apparent.

## 2 Methodology

Taking inspiration from Project 1, I initially wanted to use a combination of Scattering Net and XGBoost, but ultimately my choice was dictated by the ability to do minibatch learning which is not possible with XGBoost. Thus, pre-trained deep CNNs were a natural choice. As we mentioned before, the dataset is imbalanced, therefore, we use image transforms to augment our dataset, oversample from the “defect” class and perform test time augmentation to produce predictions.

### 2.1 Oversampling and Data Augmentation

I over-sample from the underrepresented "defect" class to make the class distribution approximately uniform. I also noticed that not all of the images in the dataset were taken from the same angle. That necessitated in teaching the model the various orientations in which the images could occur. For this, I applied random flips, rotates, zooms, contrast and brightness changes etc. Fig 2 shows an example of the transforms applied and sample images. Some of the images are indeed very hard to classify as good or bad just by looking at the image itself and therein lies the biggest challenge.



(a) (b)

Figure 2: Visualization of image transformation and sample data of Nexperia Dataset

### 2.2 Transfer Learning

The success of this classification has relied heavily on characteristics of transfer learning. I used a pre-trained ResNet34 as the backbone model – first I trained the newly added deeper layers keeping shallow layers frozen. In the second pass, I unfroze the entire model and fine tuned all the 11 million parameters. Differential training was employed – different layer weights were updated using different learning rates, chosen from an interval.

## 3 Experimental Results and Discussion

Predictions were done using Test Time Augmentation – 8 transforms of each test image were generated and a prediction was generated on each transformed image. The final prediction was then computed as the majority vote of the 8 predictions. With some modifications on TTA transformations to match the transformations done on the training dataset, we can see that the Transfer Learning process achieve good results. Fig. 4 show the confusion matrix, learning rate and training process Transfer Learning approach.

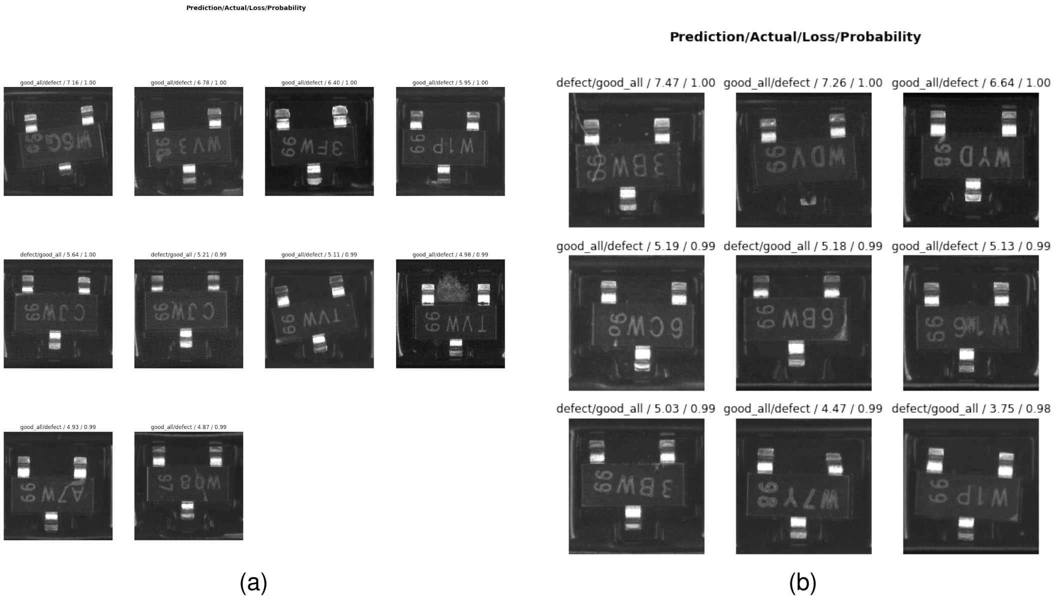

Prediction/Actual/Loss/Probability
good_all/defect / 7.16 / 1.00
good_all/defect / 6.78 / 1.00
good_all/defect / 6.40 / 1.00
good_all/defect / 5.95 / 1.00
defect/good_all / 5.64 / 1.00
defect/good_all / 5.21 / 0.99
good_all/defect / 5.11 / 0.99
good_all/defect / 4.98 / 0.99
good_all/defect / 4.93 / 0.99
good_all/defect / 4.87 / 0.99
(a)
Prediction/Actual/Loss/Probability
defect/good_all / 7.47 / 1.00
good_all/defect / 7.26 / 1.00
good_all/defect / 6.64 / 1.00
good_all/defect / 5.19 / 0.99
defect/good_all / 5.18 / 0.99
good_all/defect / 5.13 / 0.99
defect/good_all / 5.03 / 0.99
good_all/defect / 4.47 / 0.99
defect/good_all / 3.75 / 0.98
(b)

Figure 3: Prediction of (a) Pre-trained model and (b) after fine-tuned

## 4 Conclusion and Future Works

It is easily observable that even though without domain expertise, one might find it hard to differentiate between good and bad semiconductors just from images alone, deep CNNs do a commendable job at it. There are definitely patterns that hidden from the human eye which are exploited by the model. In this sense, this problem is very different from MNIST, CIFAR etc where humans are a formidable baseline.

It would be fascinating to compare the performance of the current model with:

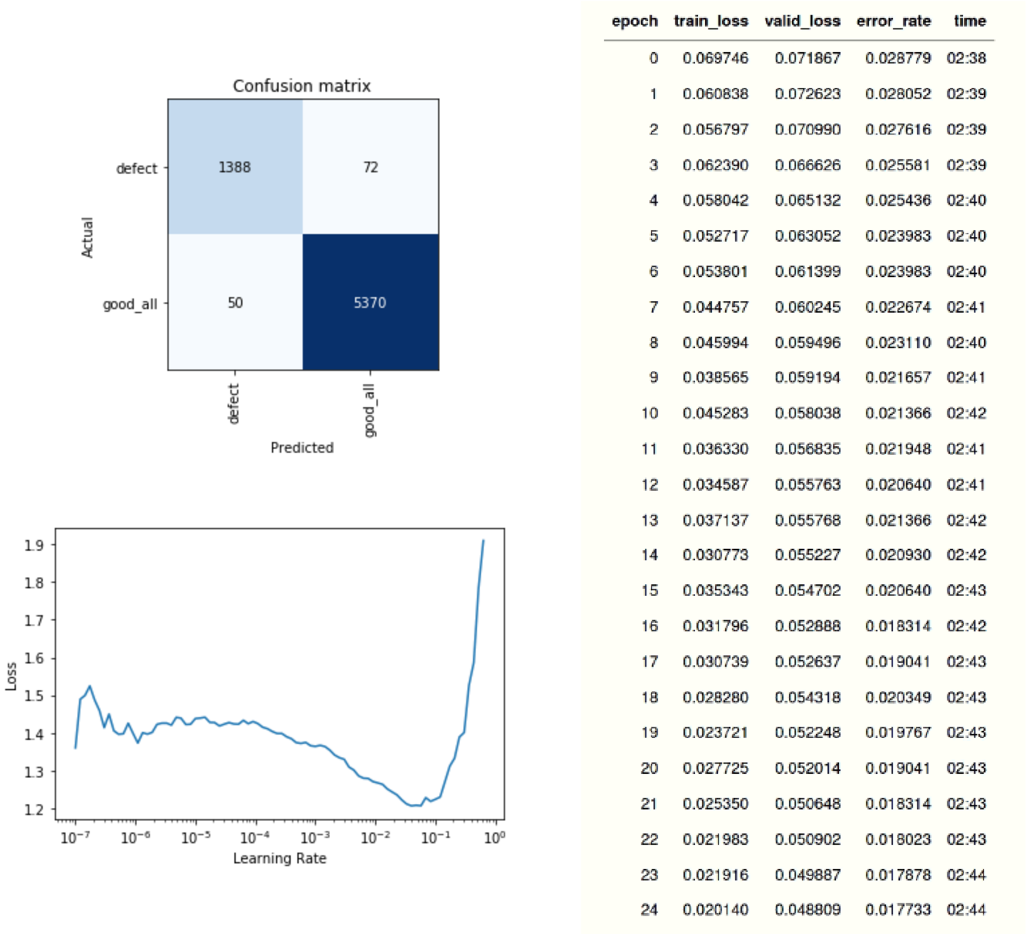
- A Generative Adversarial Network
- Variational AutoEncoder
- A combination of Scattering transform for feature extraction followed by gradient boosting (XGBoost) for classification

For the 3 in particular would be incredibly interesting, seeing as we established in Project 1, that scattering transforms are very good at extracting discriminative features.

## References

[1] M. Injadat, A. Moubayed, A. B. Nassif and A. Shami, "Multi-Stage Optimized Machine Learning Framework for Network Intrusion Detection," in *IEEE Transactions on Network and Service Management*, doi: 10.1109/TNSM.2020.3014929.

[2] T. Su, H. Sun, J. Zhu, S. Wang and Y. Li, "BAT: Deep Learning Methods on Network Intrusion Detection Using NSL-KDD Dataset," in *IEEE Access*, vol. 8, pp. 29575-29585, 2020, doi: 10.1109/ACCESS.2020.2972627.



| epoch | train_loss | valid_loss | error_rate | time |
|---|---|---|---|---|
| 0 | 0.069746 | 0.071867 | 0.028779 | 02:38 |
| 1 | 0.060838 | 0.072623 | 0.028052 | 02:39 |
| 2 | 0.056797 | 0.070990 | 0.027616 | 02:39 |
| 3 | 0.062390 | 0.066626 | 0.025581 | 02:39 |
| 4 | 0.058042 | 0.065132 | 0.025436 | 02:40 |
| 5 | 0.052717 | 0.063052 | 0.023983 | 02:40 |
| 6 | 0.053801 | 0.061399 | 0.023983 | 02:40 |
| 7 | 0.044757 | 0.060245 | 0.022674 | 02:41 |
| 8 | 0.045994 | 0.059496 | 0.023110 | 02:40 |
| 9 | 0.038565 | 0.059194 | 0.021657 | 02:41 |
| 10 | 0.045283 | 0.058038 | 0.021366 | 02:42 |
| 11 | 0.036330 | 0.056835 | 0.021948 | 02:41 |
| 12 | 0.034587 | 0.055763 | 0.020640 | 02:41 |
| 13 | 0.037137 | 0.055768 | 0.021366 | 02:42 |
| 14 | 0.030773 | 0.055227 | 0.020930 | 02:42 |
| 15 | 0.035343 | 0.054702 | 0.020640 | 02:43 |
| 16 | 0.031796 | 0.052888 | 0.018314 | 02:42 |
| 17 | 0.030739 | 0.052637 | 0.019041 | 02:43 |
| 18 | 0.028280 | 0.054318 | 0.020349 | 02:43 |
| 19 | 0.023721 | 0.052248 | 0.019767 | 02:43 |
| 20 | 0.027725 | 0.052014 | 0.019041 | 02:43 |
| 21 | 0.025350 | 0.050648 | 0.018314 | 02:43 |
| 22 | 0.021983 | 0.050902 | 0.018023 | 02:43 |
| 23 | 0.021916 | 0.049887 | 0.017878 | 02:44 |
| 24 | 0.020140 | 0.048809 | 0.017733 | 02:44 |

(a) (b)

Figure 4: The confusion matrix and learning rate (a) and training process (b)

[3] A. Nagaraja, U. Boregowda, K. Khatatneh, R. Vangipuram, R. Nuvvusetty and V. Sravan Kiran, "Similarity Based Feature Transformation for Network Anomaly Detection," in *IEEE Access*, vol. 8, pp. 39184-39196, 2020, doi: 10.1109/ACCESS.2020.2975716.

[4] A. Moubayed, E. Aqeeli and A. Shami, “Ensemble-based Feature Selection and Classification Model for DNS Typo-squatting Detection,” *33rd IEEE Canadian Conference on Electrical and Computer Engineering (CCECE’20)*, 2020, pp. 1–6.

[5] F. Salo, M. Injadat, A. B. Nassif, A. Shami and A. Essex, "Data Mining Techniques in Intrusion Detection Systems: A Systematic Literature Review," in *IEEE Access*, vol. 6, pp. 56046-56058, 2018, doi: 10.1109/ACCESS.2018.2872784.

[6] M. Injadat, F. Salo, A. B. Nassif, A. Essex and A. Shami, "Bayesian Optimization with Machine Learning Algorithms Towards Anomaly Detection," *2018 IEEE Global Communications Conference (GLOBECOM)*, Abu Dhabi, United Arab Emirates, 2018, pp. 1-6, doi: 10.1109/GLOCOM.2018.8647714.

[7] A. Moubayed, M. Injadat, A. Shami and H. Lutfiyya, "DNS Typo-Squatting Domain Detection: A Data Analytics & Machine Learning Based Approach," *2018 IEEE Global Communications Conference (GLOBECOM)*, Abu Dhabi, United Arab Emirates, 2018, pp. 1-7, doi: 10.1109/GLOCOM.2018.8647679.

[8] D. J. Weller-Fahy, B. J. Borghetti and A. A. Sodemann, "A Survey of Distance and Similarity Measures Used Within Network Intrusion Anomaly Detection," in *IEEE Com-*

*munications Surveys & Tutorials*, vol. 17, no. 1, pp. 70-91, Firstquarter 2015, doi: 10.1109/COMST.2014.2336610.

[9] T. Yu, X. Wang and A. Shami, ”Recursive Principal Component Analysis-Based Data Outlier Detection and Sensor Data Aggregation in IoT Systems,” in *IEEE Internet of Things Journal*, vol. 4, no. 6, pp. 2207-2216, Dec. 2017, doi: 10.1109/JIOT.2017.2756025.

[10] C. Yin, S. Zhang, J. Wang and N. N. Xiong, "Anomaly Detection Based on Convolutional Recurrent Autoencoder for IoT Time Series," in *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, doi: 10.1109/TSMC.2020.2968516.

[11] L. Yang, A. Moubayed, I. Hamieh and A. Shami, "Tree-Based Intelligent Intrusion Detection System in Internet of Vehicles," *2019 IEEE Global Communications Conference (GLOBECOM)*, Waikoloa, HI, USA, 2019, pp. 1-6, doi: 10.1109/GLOBECOM38437.2019.9013892.

[12] T. Tayeh, S. Aburakhia, R. Myers and A. Shami, ”Distance-Based Anomaly Detection for Industrial Surfaces using Triplet Networks,” in *2020 IEEE 11th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, Vancouver, Canada, Nov. 2020, to be published.

[13] J. Liu et al., "Anomaly Detection in Manufacturing Systems Using Structured Neural Networks," *2018 13th World Congress on Intelligent Control and Automation (WCICA)*, Changsha, China, 2018, pp. 175-180, doi: 10.1109/WCICA.2018.8630692.

[14] J. Villalba-Diez, D. Schmidt, R. Gevers, J. Ordieres-Meré, M. Buchwitz and W. Wellbroc, “Deep Learning for Industrial Computer Vision Quality Control in the Printing Industry 4.0,” *Sensors (Basel, Switzerland)*, 19(18), 3987, 2019, doi:10.3390/s19183987.

[15] M. A. F. Pimentel, D. A. Clifton, L. Clifton and L. Tarassenko, “A review of novelty detection,” *Signal Processing*, vol. 99, pp. 215–249, 2014.

[16] S. J. Pan and Q. Yang, “A survey on transfer learning,” *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[17] R. Chalapathy and Sanjay Chawla, "Deep learning for anomaly detection: A survey," *arXiv:1901.03407*, 2019.

[18] J. Hu, J. Lu and Y. Tan, "Deep transfer metric learning," *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, 2015, pp. 325-333, doi: 10.1109/CVPR.2015.7298629.

[19] S. Kornblith, J. Shlens and Q. V. Le, "Do Better ImageNet Models Transfer Better?," *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, 2019, pp. 2656-2666, doi: 10.1109/CVPR.2019.00277.

[20] O. Rippel, P. Mertens and D. Merhof, “Modeling the Distribution of Normal Data in Pre-Trained Deep Features for Anomaly Detection,” *arXiv:2005.14140*, 2020.

[21] J. Andrews, T. Thomas, E. Morton and L. Griffin, “Transfer Representation-Learning for Anomaly Detection,“ *Proceedings of the 33rd International Conference on Machine Learning ICML 2016*, New York, USA, 2016.

[22] A. Kumar and V. Vaidehi, "A transfer learning framework for traffic video using neuro-fuzzy approach," *Sādhanā 42*, pp. 1431–1442, 2017, doi: 10.1007/s12046-017-0705-x.

[23] P. Liang, H. Yang, W. Chen, S. Xiao and Z. Lan, “Transfer learning for aluminium extrusion electricity consumption anomaly detection via deep neural networks,” *International Journal of Computer Integrated Manufacturing*, 31(4-5), pp. 396–405, 2018.

[24] K. Li, N. Du and A. Zhang, “Detecting ecg abnormalities via transductive transfer learning,” *In Proceedings of the ACM Conference on Bioinformatics, Computational Biology and Biomedicine*, PP. 210–217, 2012.

[25] I. Almajai, F. Yan, T. de Campos, A. Khan, W. Christmas, D. Windridge and J. Kittler, “Anomaly detection and knowledge transfer in automatic sports video annotation,“ in *Detection and identification of rare audiovisual cues*, pp. 109–117, Springer, 2012.

[26] P. Napoletano, F. Piccoli and R. Schettini, “Anomaly detection in nanofibrous materials by CNN-based self-similarity,” *Sensors (Basel)*, 18(1), 2018, doi: 10.3390/s18010209.

[27] D. Carrera, F. Manganini, G. Boracchi and E. Lanzarone, "Defect Detection in SEM Images of Nanofibrous Materials," in *IEEE Transactions on Industrial Informatics*, vol. 13, no. 2, pp. 551-561, April 2017, doi: 10.1109/TII.2016.2641472.

[28] Z. Zhang, Y. Xu, J. Yang, X. Li and D. Zhang, "A Survey of Sparse Representation: Algorithms and Applications," in *IEEE Access*, vol. 3, pp. 490-530, 2015, doi: 10.1109/ACCESS.2015.2430359.

[29] G. Boracchi, D. Carrera and B. Wohlberg, "Novelty detection in images by sparse representations," *2014 IEEE Symposium on Intelligent Embedded Systems (IES)*, Orlando, FL, 2014, pp. 47-54, doi: 10.1109/INTELES.2014.7008985.

[30] Bernhard Schölkopf; John Platt; Thomas Hofmann, "Efficient sparse coding algorithms," in *Advances in Neural Information Processing Systems 19: Proceedings of the 2006 Conference* , MITP, 2007, pp.801-808.

[31] A. Adler, M. AElad, Y. Hel-Or and E. Rivlin, “Sparse coding with anomaly detection,” *J. Signal Process. Syst.*, 79, pp. 179–188, 2015.

[32] Q. Zhao and F. Karray, "Anomaly Detection for Images using Auto-Encoder based Sparse Representation," in *the 17th International Conference on Image Analysis and Recognition*, 2020.

[33] J. Sun, X. Wang, N. Xiong and J. Shao, "Learning Sparse Representation With Variational Auto-Encoder for Anomaly Detection," in *IEEE Access*, vol. 6, pp. 33353-33361, 2018, doi: 10.1109/ACCESS.2018.2848210.

[34] B. Zhao, L. Fei-Fei and E. P. Xing, "Online detection of unusual events in videos via dynamic sparse coding," *CVPR 2011*, Providence, RI, 2011, pp. 3313-3320, doi: 10.1109/CVPR.2011.5995524.

[35] Y. Cong, J. Yuan and J. Liu, "Sparse reconstruction cost for abnormal event detection," *CVPR 2011*, Providence, RI, 2011, pp. 3449-3456, doi: 10.1109/CVPR.2011.5995434.

[36] B. Pilastre, L. Boussouf, S. D’Escrivan and J. Tourneret, "Anomaly Detection in Mixed Telemetry Data Using a Sparse Representation and Dictionary Learning," *Signal Processing*, Vol. 168, 2020.

[37] P. Christiansen, L. N. Nielsen, K. A. Steen, R. N. Jørgensen and H. Karstoft, “Deepanomaly: Combining background subtraction and deep learning for detecting obstacles and anomalies in an agricultural field,” *Sensors*, vol. 16, no. 11, p. 1904, 2016.

[38] P. C. Mahalanobis, “On the generalized distance in statistics,” *Proceedings National Institute of Science of India*, 1936.

[39] B. Staar, M. Lütjen and M. Freitag, "Anomaly detection with convolutional neural networks for industrial surface inspection," in *Proc. CIRP*, Gulf Naples, Italy, pp. 484–489, 2018.

[40] J. Liu, K. Song, M. Feng, Y. Yan, Z. Tu and L. Zhu, "Semi-supervised anomaly detection with dual prototypes autoencoder for industrial surface inspection," *Optics and Lasers in Engineering*, Vol. 136, 2021, doi: 10.1016/j.optlaseng.2020.106324.