DA-VAE: A Semi-Supervised Approach for Anomaly Detection in Microstructure Imaging

Sungmin Lee
Smart Manufacturing Research Center
GERI(Gumi Electronics & Information Technology Research Institute)
Gumi, Korea
smlee@geri.re.kr

Dongyong Park
Root Technology Support Center
KITECH(Korea Institute of Industrial Technology
Daegu, Korea
dypark9606@kitech.re.kr

Abstract— The DA-VAE model is a semi-supervised anomaly detection framework designed to identify defects in microstructure images, especially in contexts where defective samples are scarce. In manufacturing environments such as acupuncture needle production, the lack of defective data hinders the training of traditional supervised models. To address this, we developed DA-VAE by combining data augmentation techniques with a Variational Autoencoder trained solely on normal samples. The model generates feature representations that capture the normal distribution and identifies anomalies through deviation detection. Comparative experiments show that DA-VAE outperforms conventional algorithms in both precision and recall, achieving robust performance even under extreme data imbalance. The model can be deployed in microstructure-based quality control systems across various precision manufacturing applications, improving reliability and efficiency while reducing reliance on manual inspection.

Keywords — DA-VAE, AI, Vision, Smart Manufacturing

I. INTRODUCTION

In recent years, microstructure-based defect detection has become increasingly important in precision manufacturing industries such as medical needle production, metal processing, and advanced materials. Ensuring the structural integrity of components based on surface or cross-sectional microstructures is essential for product safety and performance. However, the limited availability of defective samples in real manufacturing environments presents a major challenge for developing reliable supervised learning models.

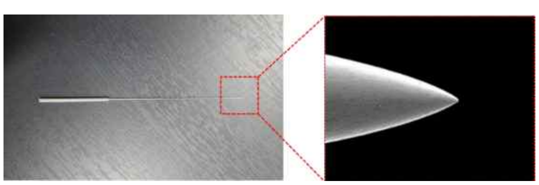

Fig 1. Microstructure image of an acupuncture needle used as training input.

To address this challenge, we propose a novel anomaly detection framework based on semi-supervised learning using Variational Autoencoders (VAE). Our approach, DA-VAE (Data-Augmented Variational Autoencoder), leverages only normal samples during training, making it well-suited for scenarios with extreme data imbalance. By augmenting the normal data and learning its latent distribution, the model can effectively identify anomalies that deviate from the learned representation.

This research aims to improve the robustness and accuracy of defect detection in microstructure imaging, enabling more scalable and automated quality control systems across various manufacturing settings.

II. RELATED WORK

Conventional defect detection techniques in manufacturing rely heavily on visual inspection or fully supervised machine learning methods. These approaches require large volumes of labeled defective samples, which are often unavailable in real industrial datasets. This has led to growing interest in unsupervised and semi-supervised learning methods for anomaly detection.

Autoencoder-based models, including traditional VAE, have shown promise in modeling normal data distributions. However, their effectiveness can be limited when dealing with highly imbalanced datasets or complex image structures such as microstructure patterns. Recent works have explored data augmentation and latent space manipulation to improve generalization, but few have integrated these strategies systematically.

Our work builds on the VAE framework by introducing a data augmentation pipeline optimized for microstructure imaging and designing a semi-supervised detection algorithm that improves sensitivity to subtle abnormalities. In doing so, we bridge the gap between anomaly detection theory and practical quality assurance in manufacturing**.**

III. METHODOLOGY

The proposed framework, DA-VAE (Data-Augmented Variational Autoencoder), is a semi-supervised anomaly detection model tailored for microstructure image analysis in manufacturing environments. In many real-world industrial settings—such as medical needle production or high-precision metal forming—defective data is scarce or unavailable. To overcome this limitation, DA-VAE is designed to be trained exclusively on normal samples, learning their intrinsic patterns and structures to detect deviations that signify potential anomalies.

The overall framework is composed of three main components: a data augmentation pipeline, a variational autoencoder model, and an anomaly scoring mechanism based on reconstruction and latent deviation. These components work together to enhance generalization, reduce overfitting, and accurately detect subtle irregularities in microstructure images**.**

1. Data Augmentation

The first component of the framework is the data augmentation module, which is used to expand the diversity of normal training data. Since deep learning models are sensitive to overfitting when trained on limited datasets, we apply a series of transformations to simulate real-world variability in microstructure appearances. These include geometric transformations such as rotations (e.g., 90°, 180°), horizontal and vertical flipping, and optical adjustments such as changes in brightness, contrast, and the injection of Gaussian noise. These operations do not introduce artificial defects but help the model become robust to variations that may occur due to lighting, orientation, or surface texture in imaging environments.

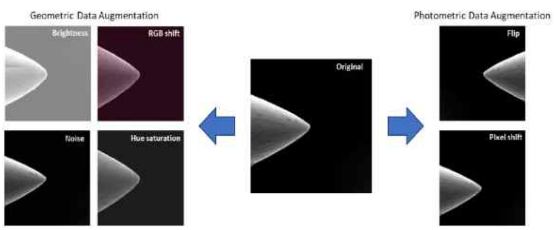

Fig.2 Data Augmentation Methodology

2. Variational Autoencoder (VAE) Training

At the core of the DA-VAE framework is a variational autoencoder, trained solely on the augmented normal data. The VAE consists of an encoder network that compresses the input image into a latent vector and a decoder network that reconstructs the original image from this representation. During training, the model minimizes a composite loss function consisting of the reconstruction error and a regularization term—the Kullback-Leibler (KL) divergence—which ensures that the learned latent space conforms to a standard Gaussian distribution. This helps the model capture the underlying structure of normal samples while also regularizing the latent space for anomaly detection.

Formally, the loss function is defined as:

$$\mathcal{L}_{VAE} = \mathbb{E}_{q(z|x)}[\log p(x|z)] - D_{KL}(q(z|x)||p(z))$$

where x is the input image, z is the latent representation, and p(z) is the prior distribution.

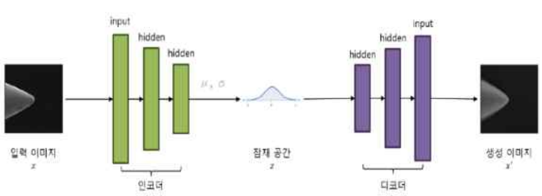

Fig. 3 Architecture of Variational Autoencoder

3. Anomaly Detection

Once training is complete, the VAE can be used for inference and anomaly detection. When a test image is passed through the trained model, two types of scores are computed:

Reconstruction Error– A high reconstruction error suggests that the input image does not conform to the learned distribution of normal data.

Latent Distance– The distance between the latent vector of the test image and the cluster of latent vectors representing the training (normal) data. This can be measured using Euclidean or Mahalanobis distance.

If either of these scores exceeds a predefined threshold, the image is flagged as anomalous. This dual-scoring approach improves robustness and sensitivity, particularly in cases where anomalies are visually subtle.

4. System Integration

To facilitate deployment in practical environments, the DA-VAE framework includes a custom-built GUI-based auto-labeling system. This interface allows users to visualize input images, reconstructed outputs, anomaly scores, and detection results. It supports batch processing, log tracking, and the generation of defect heatmaps. The system is designed to integrate easily into existing quality inspection workflows in precision manufacturing lines.

IV. CONCLUSION AND FUTURE RESEARCH DIRECTIONS

In this study, we propose an anomaly detection method that combines data augmentation with generative models, improving upon traditional anomaly detection algorithms. Acupuncture needles, as manufactured products, reflect common challenges in industrial settings—namely, data imbalance and limited sample availability. To address these issues, our approach leverages semi-supervised learning using generative models trained solely on normal data, supplemented by data augmentation to expand the dataset.

We also conduct a comparative analysis of various data augmentation strategies to identify those most effective for anomaly detection and examine how the number of augmentation iterations affects accuracy, ultimately proposing an optimized augmentation method.

Through comparisons with existing anomaly detection techniques, we demonstrate the superior performance of the VAE-based algorithm. Our proposed method offers a practical AI-driven solution that can be trained and deployed using real-world industrial data. Furthermore, the method shows great potential for application in modern manufacturing environments, particularly in automated process systems requiring robust quality control.

REFERENCES

[1] S. M. Erfani, S. Rajasegarar, S. Karunasekera, and C. Leckie, High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning, 2016.

[2] L. Breiman, Random forests, 2001.

[3] F. T. Liu, K. M. Ting, and Z. H. Zhou, Isolation forest, 2008.

[4] D. P. Kingma and M. Welling, Auto-encoding variational bayes, 2013.

This research was supported by the Ministry of Trade, Industry and Energy (MOTIE) and the Korea Institute for Advancement of Technology (KIAT) through the “DX-based Platform for Root Process Manufacturing Innovation (RS-2023-KI002779)” program.