

6th International Conference on Industry 4.0 and Smart Manufacturing

Adapting Vision Transformers for Cross-Product Defect Detection in Manufacturing

Nikolaos Nikolakis^{a,*}, Paolo Catti^a, Luca Fabbro^b, Kosmas Alexopoulos^a^a*Laboratory for Manufacturing Systems and Automation Department of Mechanical Engineering and Aeronautics, University of Patras, Rio Campus, Greece*^b*APTIV Connection System Service Italia S.p.A., Strada Del Francese 137, Turin, Italy*

Abstract

Advanced defect detection solutions that can easily adapt to different products and defect types are of high value for modern manufacturing companies. A significant challenge in developing and deploying such AI models is ensuring they generalize efficiently across diverse visual domains. This challenge is driven by limited data availability of high quality and the substantial effort required for labeling such datasets. This paper explores the adaptation of a Vision Transformer (ViT), originally trained to identify aesthetic defects in battery modules, for application in moulded plastic parts. By using transfer learning and generative AI techniques, this study evaluates fine-tuning and synthetic data augmentation strategies. The proposed approaches are assessed for their potential to enhance model adaptability and reduce dependency on extensive labelled datasets. A case study involving a battery manufacturing company with real-world data serves as the basis for this evaluation. Our preliminary findings suggest promising directions for enhancing the flexibility and efficacy of AI-driven defect detection systems in diverse manufacturing environments.

© 2025 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY-NC-ND license (<https://creativecommons.org/licenses/by-nc-nd/4.0>)

Peer-review under responsibility of the scientific committee of the 6th International Conference on Industry 4.0 and Smart Manufacturing

Keywords: Vision Transformers; Transfer Learning; Generative AI; Defect Detection.

1. Introduction

Quality control and defect detection are an integral part of manufacturing systems which aim to ensure that predefined quality standards of manufactured products are met [1]. While traditional manufacturing has relied on manual quality inspection to detect defective products [2], modern manufacturing is pivoting towards automated, non-

* Corresponding author. Tel.: +30-2610-910160

E-mail address: nikolakis@lms.mech.upatras.gr

contact and non-destructive inspection techniques [3]. This focus on early defect identification is driven by the manufacturer's desire to improve their environmental and economic sustainability [4].

Machine learning and deep learning approaches require large amounts of data to generate meaningful results [5]. This becomes increasingly evident in computer vision applications where the collection of images of defective products is challenging [6]. These challenges pose a significant barrier in training complex algorithms, such as ViTs capable of identifying manufacturing product defects.

To reduce the dependency on large datasets, approaches such as active learning [7], transfer learning and synthetic data generation have been explored [8]. However, these techniques are less focused on cross-product scenarios due to the uniqueness that usually accompanies manufacturing products [9].

In this context, the present study aims to adapt a ViT originally trained to identify aesthetic defects in battery modules, for defect detection in moulded plastic parts. By employing transfer learning and Generative Artificial Intelligence (GAI), the reliance on large image datasets is decreased while it increases the flexibility and effectiveness of Artificial Intelligence (AI)-driven defect detection systems. Thus, the key contribution of this study lies in combining transfer learning with GAI, enabling manufacturers to more easily adopt AI systems by minimizing data collection needs and reducing the time required for adapting existing AI models to different product domains.

2. Literature review

AI algorithms are becoming increasingly popular in modern manufacturing environments since AI-based approaches can provide solutions to a wide range of manufacturing problems such as predictive maintenance [10]. Additionally, an AI approach based on K-means clustering was used in [11] to enable proactive quality control through the early identification of defects on metal bars, while in [12] AI-driven models were of vital importance in improving the performance of physics-based models used to provide the behaviour of physical manufacturing assets.

Vision-based systems are becoming increasingly popular in manufacturing environments. Vision-based systems coupled with advanced AI algorithms are being used in modern manufacturing in different scenarios, such as the measurement of bin fill level to aid the scrap collection process [13] or to target the detection of defects through a non-contact and non-destructive defect detection approach that can enhance automation as well as a manufacturer's quality control process [1]. Vision-based systems are coupled with machine learning or deep learning algorithms capable of identifying defective products using 2D images that depict manufacturing products [14].

Traditionally variations of Convolutional Neural Network (CNN) algorithms have been used in vision-based setups [15]. In [16] CNNs were used for image classification in manufacturing targeting early defect identification. Furthermore, in [17] CNNs were utilized to online analyse and categorise images of manufactured products based on the presence or absence of defects on their surface. Apart from CNN algorithms, the use of long short-term memory (LSTM) networks has been explored as discussed in [18] where the LSTM was used to classify the texture of textile products and based on the classified texture detect the presence or absence of defects. Additionally, in [19] the applicability of KNNs is evaluated in defect identification through the classification of knot defect types.

ViT, which was introduced in [20], have demonstrated improvements in terms of performance in comparison to traditional CNN models. The advantages of ViT were documented in [21] where a comparison was drawn between CNN and ViT. ViT can capture global semantic details in comparison to CNN models due to the locality of the convolution operation [21]. In [22] surface and structural defects identification was targeted using a ViT. Furthermore, in [23] defects such as clogs and voids are detected in additive manufacturing using ViTs. Nevertheless, ViT's training can be computationally expensive and time-consuming, can easily become overfitted with increased layers and their scalability in different environments can be challenging [24].

Transfer learning can easily scale up a machine or deep learning algorithm and adapt it to new datasets [7]. In [25] transfer learning is applied to support the training of defect detection algorithms. Similarly, in [26] a deep transfer learning model operating with scarce training data is proposed which targets the detection of defects in aeronautics composite materials.

To support transfer learning, domain adaptation techniques such as data augmentation and synthesis, feature alignment and ensemble methods are frequently used [27]. In [28] multiple variations of each input are generated thus the training data were augmented to improve a deep learning model's accuracy, ultimately resulting in a better classification performance. Similarly, data augmentation is applied in [29]. However, a distinct difference between

[28] and [29] is the application of data augmentation to embedding inputs in [29] rather than raw data inputs. Ultimately, an overall 5% improvement in model performance was achieved using the data augmentation techniques presented in [28] and [29].

Data synthesis or data generation is an approach that synthetically generates data based on existing datasets [30]. In [9] the use of large language models is explored, to produce contextually relevant data to create synthetic datasets targeting the enhancement of low-resource and long-tail problems in the context of transfer learning. Apart from large language models used for synthetic data generation, approaches such as generative adversarial networks (GANs) and variational autoencoders (VAEs) can also support the process of transfer learning [31]. GANs have been used to generate synthetic image data to enhance small and imbalanced datasets in [32]; thus, effectively avoiding overfitting during the training process. In addition, VAEs were used in [33] to generate synthetic images of railway defects to aid the defect identification process. As highlighted in [33], and [34] VAEs demonstrate an increased robustness when compared to GANs due to GANs' high sensitivity to noise and instability during training.

While significant advancements have been made in transfer learning, data augmentation, and synthesis techniques, an approach that combines these methods specifically for cross-product scenarios remains unexplored [35], [36]. Furthermore, based on the reviewed literature and as discussed in [37] it remains unclear how techniques such as transfer learning and synthetic data generation and augmentation can be integrated into a cohesive working pipeline to facilitate the reuse of already trained AI models and effectively enhance them for performing into different scenarios. In this study, an approach is presented to bridge this gap by coupling transfer learning and GAI to enhance the existing training dataset and retrain a ViT, capable of detecting minute defects, across product domains without requiring large datasets or the development of new models. Thus, the proposed approach may facilitate the practical deployment of AI solutions, particularly in detecting minute defects, by reducing computational costs, the need for high expertise, and the requirements for extensive and high-quality data.

3. Approach

This study employs an approach where three strategies are used to adapt a ViT in cross-product domains with a focus on defect identification. The approach aims to compare the suitability of applying a) transfer learning of ViT to a cross-product domain, b) retrained ViTs in cross-product domains where datasets are synthetically, and c) a combination of the previous two to check their complementarity. To facilitate this, high-resolution images are always considered to better capture minute details of small defects that certain products may suffer from. The images are 6000x4000 pixels with 3 colour channels, i.e. 24MP.

3.1. Transfer learning of ViT to cross-product domains

The identification of defects on the surface of manufacturing products can be challenging when the to-be-identified defects are of minute dimensions. Additionally, given that modern manufacturing aims to minimize the generation of defective products, an algorithm that focuses on the similarities between images is not optimal. In this context, a custom ViT is constructed which focuses on dissimilar features rather than similar ones. This behaviour of the model is enabled by embedding the dissimilarity attention mechanism within each transformer encoder block. The blocks also include normalization layers and feedforward networks.

The basis of the dissimilarity attention mechanism is the standard multi-head attention mechanism typically used in transformers, which is then adapted to emphasize differences between the embedded representations of input patches. The dissimilarity attention mechanism projects a given input X , with dimensions $B \times N \times D$ (B denotes the batch size, N is the number of patches, and D is the embedding dimension), into a query Q , key K and value V matrices using linear transformations. Using the query and key the dissimilarity attention mechanism computes the dot product of queries. To assess the ViT the performance metrics of accuracy, precision, recall, and f1-score are used [38].

To transfer the ViT, whose architecture can be seen in Fig. 1 to another domain to identify defects on the surface of products, the aim is to retain the highest amount of acquired knowledge from the dataset of the first domain while adapting to the new domain's dataset. To achieve this, parts of the model have been frozen while specific layers of the model are unfrozen; thus, allowing them to train on new information. To retain the highest amount of generalized knowledge across different domains, the patch embedding layers, and positional encoding layers are kept frozen given

that they are typically responsible for converting image patches into embeddings as well as providing positional information, respectively. The early transformer encoder blocks are also kept frozen (blocks 1 to 6) given their ability to capture general low-level features from a dataset that can be useful across different domains. Following the frozen layers, the later transformer encoder blocks (blocks 7 to 12) are unfrozen due to their general ability to capture more

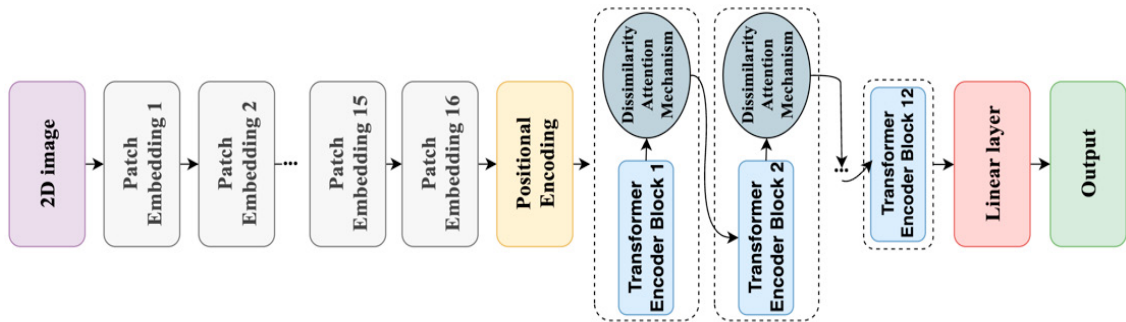


Fig. 1. The ViT's architecture.

high-level and specific features making them easily adaptable to new domains. Similarly, the dissimilarity attention mechanism that is embedded in the unfrozen blocks, becomes unfrozen to capture domain-specific dissimilarity information. Lastly, the linear layer is unfrozen since it is responsible for the final classification which is highly domain dependent.

3.2. ViT utilization in cross-product domains with synthetically enhanced datasets

As with all AI models, ViT can be utilized in cross-product domains by retraining it on new image datasets. To alleviate the dependency on collecting a high number of real-world images to retrain the ViT, generative AI is considered. Specifically given the limitations of GANs, identified in the literature review section, a VAE model is used for the generation of high-resolution images, based on a small sample of real-world ones. This is facilitated by the introduction of a data augmentation module that artificially increases the cross-product domain's dataset size using data augmentation techniques such as rotation, scaling, and flipping.

The architecture of the VAE can be seen in Fig. 2. Initially, both augmented and real-world images of the second domain are inputted to the VAE at their original high resolution of 24MP. The input images are provided to three two-dimensional convolutional layers which extract underlying features and reduce spatial dimensions. The output of the third convolutional layer is fed to a flatten layer which converts the 2D feature maps into a 1D vector which is fed to the dense layer to reduce the dimensionality and capture abstracted features. The output of the dense layer is fed to the mean and log variance layers. The mean layer outputs the mean of the latent space distribution for each inputted image as well as reduces the higher-dimensional feature representation of an image to a lower-dimensional latent space while providing the central point around which the latent space distribution is centred. Lastly, the log variance layer reduces the feature representation of the input image to match the latent space dimensionality while determining the speed of the latent space distribution. These two layers when combined define a Gaussian distribution in the latent space for each input image.

The decoder network is composed of two dense layers that expand the latent vector back to a higher-dimensional feature representation. They are followed by an unflatten layer that reshapes the 1D vector to a 2D feature map, which is then inputted to the five deconvolutional layers which unsample and reconstruct initial spatial dimensions and features. The output layer produces the final reconstructed image, matching the original input dimensions of 6000x4000x3. Lastly, the VAE's performance is quantified through the calculation of the binary classification loss (BCE), Kullback-Leibler Divergence (KLD), and beta (β) [39]. BCE signifies the reconstruction error of the generative AI model, KLD the difference between the latent space distribution and the prior distribution, while β is a weight parameter that balances the importance of KLD relative to BCE.

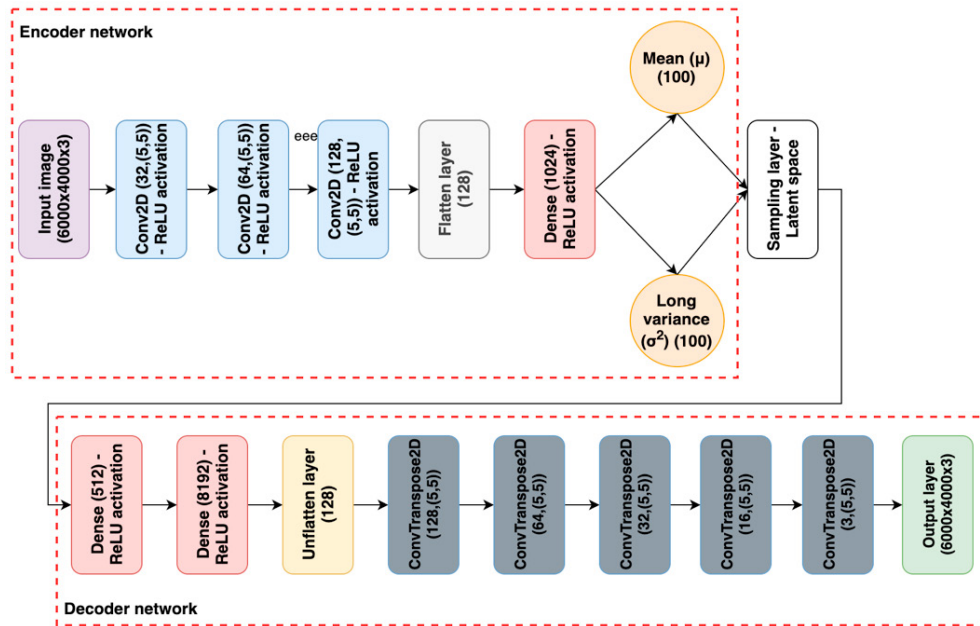


Fig. 2. VAE's architecture.

3.3. Coupling the transfer of ViT and synthetic data generation for cross-product defect identification

A hybrid approach is also considered coupling transfer learning with generative AI (Fig. 3). In this hybrid approach, a generative AI layer is considered which is positioned between the trained ViT model on the first domain and the transfer learning application in the cross-product second domain. This coupling uses the ViT presented in section 3.1. For the generative AI layer, the VAE presented in section 3.2 is utilized.

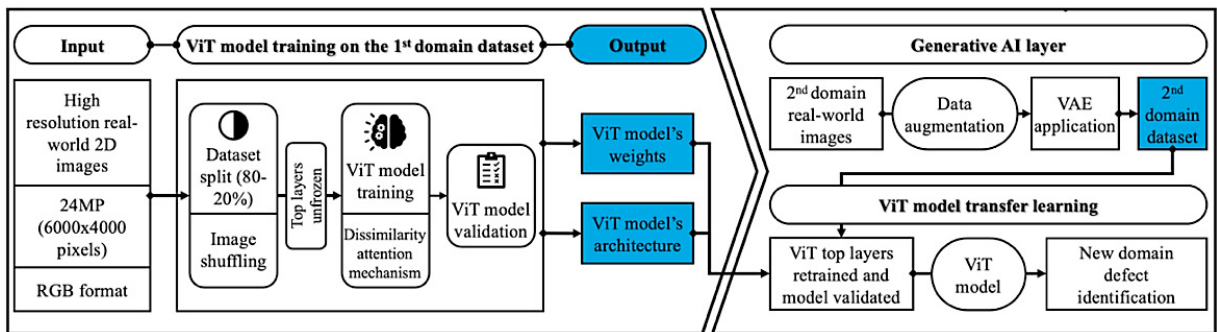


Fig. 3. Overview of the approach for cross-product defect identification using ViT and generative AI.

4. Experimentation

To evaluate the proposed approach in a cross-product scenario, two domains were selected from the manufacturing industry. The two domains serve as the basis for the experimentation where through it the best strategy in adapting a ViT in cross-product scenarios will be identified using the three strategies detailed in section 3. To evaluate the suitability of the two domains in the application of the approach, the experimentation began by quantifying and

comparing features from images of the two domains. The features used include the contrast of the texture of the image, the dissimilarity of the texture, the homogeneity of the texture, the energy of the texture, the correlation of the texture, the edge sharpness of the image and the surface smoothness. All features are normalized between 0 and 1 to make them comparable.

The first domain is focused on the production line of an electric vehicle manufacturer, specifically a laser welding process. The process was selected due to its criticality in defect generation facilitating the importance of introducing proactive quality control through early defect identification. The second domain belongs to the manufacturing of moulding plastic parts. This domain offers an industrial use case where defects are of minute size. This poses challenges in image collection due to the high-resolution images required to accurately capture information related to defects on the surface of the plastic parts.

Following the proposed approach, high-resolution, real-world images were collected of 24MP resolution from the first domain, which was labelled based on the presence or absence of defects on the surface of the battery cell. Images were collected using a vision-based system operating in the visible spectrum that has been installed in the line of the battery cell manufacturer. Defects that can be observed on the product's surface include scratches and spots on the surface of the battery cell. An example of a battery cell image can be seen in Fig. 4 (a), where a scratch on the cell's surface is visible. The dataset created to train the presented ViT was composed of approximately 1,800 images.

In the second domain, which focuses on the manufacturing of moulding plastic parts, real-world image collection is challenging. Defects on the moulding plastic parts include burns on the surface of the part and the presence of excessive material inside the cavities of the parts. Due to the size of the plastic parts, these types of defects cannot be easily detected by human operators in the line. To address these challenges, these types of defects are being detected manually using specialized instruments, making the process highly sensitive to human expertise. This increases the need for an automated system for the identification of such defects using high-resolution images of the plastic parts. To facilitate the transfer learning process in this cross-product scenario, the generative AI layer was used to synthetically generate images of the parts based on collected real-world images. Approximately 50 real-world images were collected of the plastic parts.

To evaluate the first strategy, the ViT was transferred from the first to the second domain using the 50 real-world images collected. The second strategy was evaluated by using the VAE to synthetically generate approximately 200 images. Lastly, supporting the third strategy the ViT was transferred to the second domain whose dataset was enriched using the 200 images generated to facilitate the second strategy evaluation. An example of a defective plastic part can be seen in Fig. 4 (b), where the defect, specifically the burn on the plastic, is highlighted inside the red rectangle.

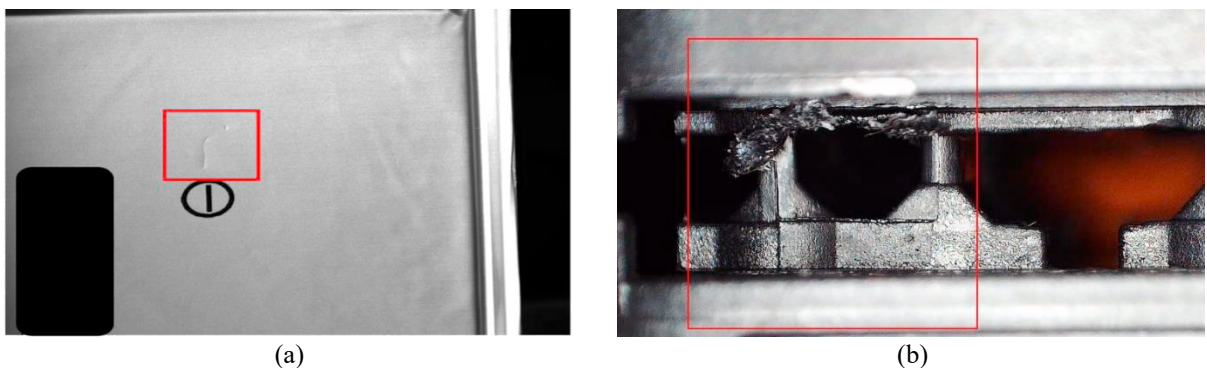


Fig. 4. (a) Battery cell image with a scratch on the cell's surface (highlighted inside the red rectangle), (b) Moulding plastic part with a burn (highlighted inside the red rectangle),

The construction, training and validation of the ViT in the first domain, the generation of synthetic images in the second domain, as well as the application of transfer learning of the ViT in the second domain, were implemented in a development environment where Python 3.9 was utilized, together with its accompanying libraries, such as PyTorch for model training and model transfer, and Torchvision 0.18.0 for the generation of synthetic images using the generative AI layer. The development environment consisted of a Windows PC, equipped with an Intel Core i9-

10850K processor, 64 GB of RAM, and an NVIDIA RTX 2070 SUPER GPU with 8GB GDDR6 VRAM, running Windows 10 Enterprise version 22H2.

To evaluate the ViT, the performance metrics of accuracy, precision, recall and f1-score were used. To evaluate the performance of the generative AI layer, BCE, KLD, and beta (β) are calculated. Furthermore, to validate the improvements resulting from the introduction of the generative AI layer in the proposed approach, the ViT is transferred to the second domain with and without the use of the synthetically enhanced second domain dataset.

5. Results and discussion

To begin with, the features of the real-world images of the two domains were calculated and normalized, as seen in Fig. 5. Based on the results presented in Fig. 5 the contrast of the texture of the images from both domains is significantly close, indicating similar variation in intensity between the two domains. Similar results can be seen for the edge sharpness of the images and the dissimilarity of the texture between images. In terms of homogeneity and energy of the texture in the two domain images, the second domain shows higher values in both metrics pointing to increased variation in the texture of the images of the second domain. In contrast, images from the first domain show higher texture correlation indicating more predictable textures and their surface smoothness is higher indicating less image complexity. Overall, given the relatively high commonalities between the features, the images from the two domains have a solid foundation for the application of transfer learning.



Fig. 5. Comparison of features between the first and second domain images

Following the experiments defined in the section 4, the results of training the ViT in the first domain's dataset were collected. Its weights were then extracted to be transferred to the second domain in the context of evaluating the first strategy. The VAE was used to synthetically enrich the limited dataset of the second domain and then the ViT was retrained to evaluate the second strategy. Lastly, the third strategy was evaluated. To evaluate the third strategy, the ViT was transferred from the first domain to the second domain whose dataset was synthetically enhanced using the generative AI layer. These results have been grouped in Table 1.

Table 1. Experimentally extracted performance metrics.

Performance metric	ViT on first domain dataset	ViT transferred on the second domain dataset (strategy 1)	ViT retrained on the synthetically enhanced second domain (strategy 2)	ViT transferred on the synthetically enhanced second domain (strategy 3)
Accuracy	0.92	0.44	0.63	0.68
Precision	0.85	0.41	0.61	0.63
Recall	0.51	0.38	0.50	0.52
F1-score	0.64	0.39	0.54	0.56

The results in Table 1, validate the applicability of the ViT with its dissimilarity attention mechanism at accurately identifying true positive defects on the surface of products, given the high accuracy and precision the model achieved while training on the first domain. Additionally, the results signify a substantial average improvement across all performance metrics of approximately 46% of the transferred ViT to the synthetically enriched second domain dataset (strategy 3), when compared to the application of the transferred ViT on the original second domain dataset (strategy 1). Furthermore, an average improvement of 40% on all performance metrics was reached when transferring the ViT to the synthetically enhanced second domain dataset (strategy 2) when compared to the results obtained through the first strategy. These results point towards a high added value for the generative AI layer. To explore the ability of the VAE to accurately generate synthetic images, the BCE, KLD, and β were calculated and can be found in Table 2.

Table 2. Performance metrics of the VAE model.

Performance metrics of the VAE model	Score
BCE	122,622.77
KLD	4.43
beta (β)	0.99

The performance metrics' scores of the generative AI layer indicate that the VAE can generate the second domain's images with moderate overall effectiveness. Given the targeted high-resolution of the to-be-generated images, achieving a low BCE is challenging. In calculating the BCE, each pixel contributes to the total loss of the model and given the number of pixels in each image (24 million pixels), even a small error in each generated pixel could accumulate to a proportionally large BCE score. Nonetheless, the achieved BCE score is still relatively high, pointing to the model requiring further tuning and improvement. Moreover, the achieved KLD score of 4.43, indicates that the model's latent space is relatively well organized, and the encoded representations are close to the assumed prior distribution. Lastly, the beta (β) metric, is close to the desired 1, which indicates that the generative AI layer can balance the reconstruction and regularization losses.

Another factor that is affected by the application of the proposed approach is the computational resources and time used during the experimentation. Given the large number of high-resolution images that make up the first domain's dataset, ViT required more than 24 hours to train and validate. VAE is also computationally expensive. To generate the synthetic images of the second domain VAE required almost 3 hours. Nevertheless, this is also highly dependent on the architecture of the generative AI layer, where a deeper architecture can exponentially increase the computational time required for the enrichment of the second domain's dataset. However, with the introduction of transfer learning, the computational time is significantly reduced to adopt the ViT in the second domain. By taking into consideration also the time required to synthetically enrich the second domain dataset via the generative AI layer, the time required to transfer the ViT to the second domain required 7 hours; a significant reduction in time in comparison to the time needed to train the ViT on the first domain.

Ultimately, from the results presented in both Table 1 and Table 2, it is evident that the adaptability of ViT can be increased through the application of transfer learning in cross-product domains with minimal tuning in the model's architecture. Additionally, the introduction of the generative AI layer into the process of transfer learning for a ViT is of high value in cross-product scenarios, where product images of both domains share some similarities. The value of the generative AI layer is especially evident in the context of reducing the dependency in large real-world datasets

given that the generation of high-quality synthetic image datasets is possible. This is exponentially important in scenarios where the models are used for the early identification of defects, since accurate models can effectively determine the presence of defects, facilitating the overall sustainability increase of the manufacturer.

Nevertheless, the experimentation also signifies the importance of having a relatively high training dataset for training or transferring ViT. During the evaluation of the first strategy, the ViT had at its disposal only 50 images to train and validate upon, thus resulting in bad classification performance. The number of real-world second-domain images also affects scenarios 2 and 3. It is evident that for the generative AI layer to properly function and generate high-quality synthetic image datasets, a low number of images, such as 50, can pose significant barriers to achieving optimal generative performance.

Experimental results demonstrate the effectiveness of the proposed approach in cross-product defect detection. The training time for the ViT on the second domain was reduced by more than 30%, while the number of collected images needed by 98%. These reductions may have a significant impact for manufacturers, substantially lowering the costs and time required to scale AI solutions across different domains, thus their practical usability and adoption rate. Additionally, the results suggest that the approach mitigates one of ViT's known limitations, high training time. Another key finding was the successful reuse of the dissimilarity attention mechanism. However, further experimentation is required to validate these findings and explore whether customizing the mechanism could improve defect detection in other domains.

6. Conclusions

This study examined the effectiveness of coupling generative AI with transfer learning in computer vision applications for product defect detection using a ViT, specifically designed for this purpose. The inclusion of a dissimilarity attention mechanism in the ViT further supports its suitability for the detection of minute defects on the surface of manufacturing products.

Focusing on cross-product defect detection, the proposed approach successfully managed to transfer a ViT trained on real-world data from one domain to another. A generative AI layer, through a VAE model, facilitated this transfer by significantly improving the performance of the model in the second domain. Nevertheless, this approach reduces the reliance on large real-world datasets, an adequate number of initial image samples remains necessary for the generative AI layer to generate accurate and high-quality synthetic dataset.

Future work will refine the proposed method, focusing on the identification of a baseline number of image samples required for the GAI layer to be highly effective in synthetic image generation. This will highly increase the robustness of the approach while ensuring a consistent processing time. Additionally, improvements to the VAE architecture will be explored aiming to simplify the model and reduce the computational resources and time required. Moreover, the ViT's architecture will be finetuned by experimenting with the number of unfrozen layers to improve the transferability of the model. Finally, the proposed approach will be tested in real-world environments and across different domains to evaluate its reproducibility in different applications.

Acknowledgement

This work was partially supported by the HORIZON-CL4-2021-TWIN-TRANSITION-01 openZDM project, under Grant Agreement No. 101058673.

References

- [1] Chrysosolouris G, Alexopoulos K, Arkouli Z (2023) A Perspective on Artificial Intelligence in Manufacturing.
- [2] Colledani M, Tolio T, Fischer A, Iung B, Lanza G, Schmitt R, Váncza J (2014) Design and management of manufacturing systems for production quality. *CIRP Annals* 63:773–796
- [3] Medici V, Martarelli M, Paone N, Pandarese G, van de Kamp W, Verhoef B, Sipsas K, Broechler R, Besada L R, Alexopoulos K, Nikolakis N (2023) Integration of Non-Destructive Inspection (NDI) systems for Zero-Defect Manufacturing in the Industry 4.0 era. In: 2023 IEEE International Workshop on Metrology for Industry 4.0 & IoT (MetroInd4.0&IoT). pp 439–444
- [4] Nikolakis N, Catti P, Chaloulos A, Van De Kamp W, Coy MP, Alexopoulos K (2024) A methodology to assess circular economy strategies for sustainable manufacturing using process eco-efficiency. *Journal of Cleaner Production* 445:141289
- [5] Shinde P P, Shah S (2018) A Review of Machine Learning and Deep Learning Applications. In: 2018 Fourth International Conference on

- Computing Communication Control and Automation (ICCUBE). pp 1–6
- [6] Talaei Khoei T, Ould Slimane H, Kaabouch N (2023) Deep learning: systematic review, models, challenges, and research directions. *Neural Comput & Applic* 35:23103–23124
 - [7] Papacharalampopoulos A, Alexopoulos K, Catti P, Stavropoulos P, Chrysosolouris G (2024) Learning More with Less Data in Manufacturing: The Case of Turning Tool Wear Assessment through Active and Transfer Learning. *Processes* 12:1262
 - [8] Yang L, Hanneke S, Carbonell J (2013) A theory of transfer learning with applications to active learning. *Mach Learn* 90:161–189
 - [9] Guo X, Chen Y (2024) Generative AI for Synthetic Data Generation: Methods, Challenges and the Future.
 - [10] Cerquitelli T, Nikolakis N, O'Mahony N, Macii E, Ippolito M, Makris S (eds) (2021) *Predictive Maintenance in Smart Factories: Architectures, Methodologies, and Use-cases*.
 - [11] Ntoulmperis M, Catti P, Discepolo S, Kamp WVD, Castellini P, Nikolakis N, Alexopoulos K (2024) 3D point cloud analysis for surface quality inspection: A steel parts use case. *Procedia CIRP* 122:509–514
 - [12] Catti P, Nikolakis N, Sipsas K, Picco N, Alexopoulos K (2024) A hybrid digital twin approach for proactive quality control in manufacturing. *Procedia Computer Science* 232:3083–3091
 - [13] Alexopoulos K, Catti P, Kanellopoulos G, Nikolakis N, Blatsiotis A, Christodouloupoulos K, Kaimenopoulos A, Ziata E (2023) Deep Learning for Estimating the Fill-Level of Industrial Waste Containers of Metal Scrap: A Case Study of a Copper Tube Plant. *Applied Sciences* 13:2575
 - [14] Schlosser T, Friedrich M, Beuth F, Kowerko D (2022) Improving automated visual fault inspection for semiconductor manufacturing using a hybrid multistage system of deep neural networks. *J Intell Manuf* 33:1099–1123
 - [15] Sun Y, Xue B, Zhang M, Yen GG (2020) Automatically designing CNN architectures using genetic algorithm for image classification. *IEEE Trans Cybern* 50:3840–3854
 - [16] Frick J, Grudowski P (2023) Quality 5.0: A Paradigm Shift Towards Proactive Quality Control in Industry 5.0. *International Journal of Business Administration* 14:51
 - [17] Zhang B, Jaiswal P, Rai R, Guerrier P, Baggs G (2019) Convolutional neural network-based inspection of metal additive manufacturing parts. *Rapid Prototyping Journal* 25:530–540
 - [18] Kumar KS, Bai MR (2023) LSTM based texture classification and defect detection in a fabric. *Measurement: Sensors* 26:100603
 - [19] Cetiner I, Ali Var A, Cetiner H (2016) Classification of Knot Defect Types Using Wavelets and KNN. *EIAEE* 22:67–72
 - [20] Dosovitskiy A, Beyer L, Kolesnikov A, et al (2021) An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale.
 - [21] Üzen H, Turkoglu M, Ozturk D, Hanbay D (2024) A novel hybrid attention gate based on vision transformer for the detection of surface defects. *SIViP*.
 - [22] Wang H (2023) Investigating Lightweight Transformer Models for Defect Detection. *AJST* 7:10–16
 - [23] Bimrose MV, Hu T, McGregor DJ, Wang J, Tawfick S, Shao C, Liu Z, King WP (2024) Detecting and classifying hidden defects in additively manufactured parts using deep learning and X-ray computed tomography. *J Intell Manuf*.
 - [24] Shang H, Sun C, Liu J, Chen X, Yan R (2023) Defect-aware transformer network for intelligent visual surface defect detection. *Advanced Engineering Informatics* 55:101882
 - [25] Chen H, Lin H, Xu Q, Li Y, Zheng Y, Fei J, Yang K, Fan W, Nie Z (2023) Cross-Domain Transfer Learning for Galvanized Steel Strips Defect Detection and Recognition. *Journal of Computing and Information Science in Engineering*.
 - [26] Gong Y, Shao H, Luo J, Li Z (2020) A deep transfer learning model for inclusion defect detection of aeronautics composite materials. *Composite Structures* 252:112681
 - [27] Zheng Z, Li R, Liu C (2024) Learning robust features alignment for cross-domain medical image analysis. *Complex Intell Syst* 10:2717–2731
 - [28] Boukli Hacene G, Gripon V, Farrugia N, Arzel M, Jezequel M (2018) Transfer Incremental Learning Using Data Augmentation. *Applied Sciences* 8:2512
 - [29] Wolfe C, Lundgaard K (2019) Data Augmentation for Deep Transfer Learning.
 - [30] Alexopoulos K, Nikolakis N, Chrysosolouris G (2020) Digital twin-driven supervised machine learning for the development of artificial intelligence applications in manufacturing. *International Journal of Computer Integrated Manufacturing* 33:429–439
 - [31] Lu Y, Shen M, Wang H, Wang X, van Rechem C, Fu T, Wei W (2024) Machine Learning for Synthetic Data Generation: A Review.
 - [32] Chatterjee S, Hazra D, Byun Y-C, Kim Y-W (2022) Enhancement of Image Classification Using Transfer Learning and GAN-Based Synthetic Data Augmentation. *Mathematics* 10:1541
 - [33] Ferdousi R, Yang C, Hossain MA, Laamarti F, Hossain MS, Saddik AE (2024) Generative Model-Driven Synthetic Training Image Generation: An Approach to Cognition in Railway Defect Detection. *Cogn Comput*.
 - [34] Bond-Taylor S, Leach A, Long Y, Willcocks CG (2022) Deep Generative Modelling: A Comparative Review of VAEs, GANs, Normalizing Flows, Energy-Based and Autoregressive Models. *IEEE Trans Pattern Anal Mach Intell* 44:7327–7347
 - [35] Iman M, Arabnia HR, Rasheed K (2023) A Review of Deep Transfer Learning and Recent Advancements. *Technologies* 11:40
 - [36] Gupta V, Choudhary K, Tavazza F, Campbell C, Liao W, Choudhary A, Agrawal A (2021) Cross-property deep transfer learning framework for enhanced predictive analytics on small materials data. *Nat Commun* 12:6595
 - [37] Wang R, Hoppe S, Monari E, Huber MF (2023) Defect Transfer GAN: Diverse Defect Synthesis for Data Augmentation.
 - [38] Powers D, Ailab (2011) Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation. *J Mach Learn Technol* 2:2229–3981
 - [39] Heuver R/ PR (2020) Generating facial morphs through PCA and VAE.