Empowering ConvNeXt for Precision Classification of Industrial Surface Defects: A Comprehensive Approach with Multi-Scale Fusion

Jiaxin Cao1 and Yu Bai2,∗

1School of Computer Science and Engineering, Xi ’an Technological University

E–mail: caojiaxin@st.xatu.edu.cn

2School of Mechanical and Electrical Engineering, Xi ’an Technological University

Abstract: The industrial sector demands high precision in the classification of surface defects to ensure product quality. Traditional visual inspection methods, limited by their inconsistency and inability to scale, necessitate an advanced solution for defect detection and classification. This research introduces an enhanced ConvNeXt architecture that integrates deformable convolutions, attention mechanisms, and a multi-scale fusion approach to address the complex nature of defect imagery in industrial settings. Firstly, deformable convolutions are employed to provide the model with the flexibility to adapt to the varied and irregular shapes of surface defects. Unlike standard convolutions, these allow the network to modify its receptive field dynamically, enhancing its ability to capture crucial textural and geometric nuances. This adaptation significantly boosts the model’s accuracy in feature extraction from complex industrial surfaces. Secondly, to refine the focus within these enhanced feature maps, an attention mechanism is integrated. This mechanism prioritizes the most informative parts of the image, thus directing computational resources towards areas with potential defects. By doing so, it not only improves the model’s efficiency but also its effectiveness in recognizing subtle yet critical defect features that might otherwise be overlooked. Thirdly, the multi-scale fusion strategy is implemented to harmonize and leverage information across different scales and resolutions. This aspect of the model ensures comprehensive coverage and consistent performance across varying sizes and types of defects. It effectively aggregates the detailed local features captured by deformable convolutions and the prioritized global features enhanced by attention mechanisms, providing a robust classification output. Experimental results on diverse industrial datasets have demonstrated that the proposed model substantially outperforms existing methods in terms of both accuracy and reliability. The integration of these three advanced techniques—deformable convolutions, attention mechanisms, and multi-scale fusion—creates a synergistic effect that significantly elevates the capability of ConvNeXt for precise classification of industrial surface defects. This study not only proves the feasibility of enhancing a sophisticated architecture like ConvNeXt for industrial applications but also sets a new standard for automated defect classification systems, combining deep learning innovation with practical, impactful industrial use.

Key words. Editorials notices — Publications, bibliography — Miscellaneous

1. INTRODUCTION

In contemporary industrial production, maintaining stringent quality control standards, particularly in surface defect detection, is vital for safeguarding product integrity Ngan et al. (2011). The tradi- tional methodologies employed in detecting such de- fects typically oscillate between manual inspections and semi-automated techniques. Manual inspection, while thorough, is inherently limited by its subjec- tivity, potential for human error, and significant re- source demands. These factors contribute to substan- tial variability in product quality assessment, which can compromise the overall reliability of manufac- tured goods Chen et al. (2021).

Semi-automated methods, often based on basic image processing technologies, offer some relief from the labor intensity of manual inspections but intro-duce their own set of challenges. These systems generally lack the sophistication needed to handle the complexity and diversity of surface defects effectively Francesco et al. (2018). Industrial defects can manifest in a multitude of forms, varying greatly in size, shape, texture, and contrast. Such variability presents a formidable challenge for conventional image processing techniques, which are not inherently designed to adapt to the unpredictable nature of defect presentations.

The advent of deep learning has provided promis- ing new pathways for enhancing defect detection systemsRath et al. (2021), Mishra and Tyagi (2022), Blake and Michalikova (2021). Among these, Convolutional Neural Networks (CNNs) have shown sub- stantial potential due to their ability to learn and generalize from complex data inputs. However, despite the advancements facilitated by CNNs, their application in the industrial setting is often hampered by architectural rigidity. Standard CNNs typically oper- ate with fixed receptive fields, which are not optimal for capturing the highly irregular patterns that char- acterize many surface defects.

To address these limitations, this research pro- poses a strategic adaptation of the ConvNeXt architecture Liu et al. (2022), a model that mod- ernizes conventional convolutional neural networks (CNNs) by incorporating design principles inspired by the Vision Transformer (ViT) without directly using transformer mechanisms. Our adaptation is tailored specifically to meet the unique demands of surface defect detection in industrial environments. The research is focused on three main innovations designed to enhance the base ConvNeXt model:

- **Integration of Deformable Convolutions:** By incorporating deformable convolutions within the ConvNeXt model, this research ad- dresses the challenge of accurately capturing the complex and irregular forms of surface defects. Traditional convolutional layers, with their fixed geometric structure, are often inadequate for this task. Deformable convolutions allow the network's filters to adjust dynamically, aligning with the contours of each specific defect. This capability significantly enhances the accuracy of feature extraction across various defect types, providing a more detailed and precise analysis of each defect's characteristics.
- **Implementation of Attention Mechanisms:** The introduction of attention mechanisms within the ConvNeXt architecture marks a sub- stantial improvement in the model's efficiency and effectiveness. By focusing computational re- sources on image regions most likely to contain defects, these mechanisms enhance the model's ability to discern subtle yet critical defect fea- tures that conventional methods might overlook. This focus not only improves the detection accu- racy but also optimizes the overall resource usage

  of the model, leading to faster and more efficient defect identification.
- **Employment of Multi-Scale Fusion:** The multi-scale fusion strategy implemented in this research tackles the variability in defect sizes, a common issue in industrial settings. By integrating features extracted at multiple scales, the model is equipped to detect defects ranging from minute to prominent with consistent precision. This approach ensures a comprehensive analysis of each inspected item, enabling the model to provide a detailed assessment irrespective of the defect size or complexity.

These contributions collectively enhance the Con- vNeXt model's utility for industrial applications, particularly in automating the detection and classifica- tion of surface defects. The strategic enhancements not only improve the model's adaptability to a va- riety of defect characteristics but also ensure that it can operate efficiently in a real-world manufacturing environment. This research thereby sets a new standard for defect detection systems, facilitating more reliable and precise quality control in industrial production processes.

2. Related Work

The rise of Industry 4.0 has fundamentally altered the manufacturing landscape by incorporating advanced technologies like the Internet of Things (IoT), big data, and artificial intelligence Xu et al. (2018), Hofmann and Ru…sch (2017), Ghobakhloo (2020). These technologies not only automate and optimize manufacturing processes but also provide the adaptability required for mass production to meet changing market demands.

Traditional techniques for detecting surface de- fects in manufacturing, which are mainly manual and error-prone, are progressively being replaced by auto- mated systems that utilize computer vision and machine learning. This transition aims to eliminate the inefficiencies and inaccuracies associated with manual inspections. Key historical perspectives on surface inspection techniques were outlined by Chin Chin (1988) in the late eighties and by Newman and Jain Newman and Jain (1995) in the mid-nineties. Li and Gu Li and Gu (2004) later reviewed advance- ments in free-form surface inspection. However, in the years that followed, there have been substantial improvements in the field of surface inspection with computer vision. Emerging areas such as tonality inspection and the growing implementation of color imaging technologies call for new algorithms capa- ble of efficiently processing vector-valued data. This paper delves into the latest developments in vision- based surface inspection, with a particular focus on techniques for analyzing textures.

Recent strides in deep learning have consider- ably advanced the capability for defect detection within industrial environments. A thorough anal- ysis of deep learning applications in surface defect detection across various industrial goods is provided by Saberironaghi et al. (2023), which delineates common issues such as the challenge of unbalanced data distributions prevalent in real-world datasets Saberironaghi et al. (2023). Their findings under- score an intensified effort to refine X-ray defect detection methodologies through deep learning, aim- ing to enhance both the precision and speed of these systems. Simultaneously, Li et al. (2023) have un- veiled an innovative automatic defect detection system crafted for Wire and Arc Additive Manufacturing (WAAM), employing a YOLO-attention mechanism. This system offers swift and dependable defect detec- tion, meeting the exigent requirements of dynamic manufacturing environments Li et al. (2023). This development is indicative of a broader movement to- wards implementing real-time, efficient detection sys- tems in manufacturing workflows. In a related en- deavor, Akhyar et al. (2023) present their work on a deep learning-driven surface defect inspection system engineered specifically for the steel industry. Dubbed the Forceful Steel Defect Detector (FDD), this system is designed to confront the unique challenges associ- ated with detecting defects on steel surfaces Akhyar et al. (2023). This breakthrough highlights the cus- tomization of deep learning solutions to meet specific industry needs, thereby improving the efficacy of au- tomated inspection systems. Additionally, Chen et al. (2023) examine defect detection techniques for 3D-printed ceramic parts with curved surfaces, which typically present low contrast. They propose a deep learning approach tailored to manage the intricacies of advanced ceramic materials, showcasing the adapt- ability of deep learning to a range of materials and complex geometric challenges Chen et al. (2023).

Despite significant advancements in leveraging deep learning for industrial defect detection, the potential of newer architectures like ConvNeXt has yet to be fully explored in this domain. While Con- vNeXt has shown promise in general image classifica- tion tasks due to its efficient balance of model depth, width, and resolution, its applications in specific in- dustrial settings remain underdocumented. Recent works have begun to tap into the capabilities of Con- vNeXt for complex image-related tasks. For instance, the research Chen et al. (2023) demonstrates the ap- plication of ConvNeXt in identifying manufacturing defects in nuclei segmentation and classification How- ever, these studies often do not address the unique challenges posed by the high variability of defects in different manufacturing processesgao **?**. This vari- ability can significantly affect the performance of deep learning models, which typically require large volumes of labeled data that represent the range of possible defect types and severities encountered in production environments. Furthermore, while Con- vNeXt's architecture is inherently suited for handling detailed and complex image data, its integration with technologies specifically tailored for industrial appli- cations, such as industrial imaging, has not been ade- quately studied. There is a gap in research regarding the optimal configuration of ConvNeXt's parameters to enhance its effectiveness in detecting subtle and non-uniform defects.

Moreover, the need for enhancements that specifi- cally address the intricacies of defect variability in industrial settings becomes apparent. Introducing at- tention mechanisms can significantly augment Con- vNeXt's ability to prioritize salient features in com- plex industrial images, potentially improving the pre- cision of defect detection. Alongside, the deployment of deformable convolutions and a multi-scale fusion strategy could further refine the model's adaptability and accuracy across different scales and types of de- fects, thus tailoring it more effectively to the diverse conditions encountered in industrial environments. Yang et al. (2023) explore innovative developments in the application of lightweight deep learning networks for diagnosing plant diseases, particularly focusing on rice disease. They employ a Squeeze-and-Excitation (SE) attention mechanism in their network, which en- hances both the accuracy and efficiency of the model. This approach demonstrates how effectively integrat- ing attention mechanisms with dynamic convolutions can enhance performance, especially in the field of ecological informatics.

Further building on these concepts, Zhu et al.

(2019) have carried out an in-depth empirical study that examines the role of spatial attention mechanisms within deep learning architectures. Their re- search explores the combined use of attention, deformable convolutions, and dynamic convolutions, in- vestigating how these components can be synergistically used to boost model performance Zhu et al. (2019). The findings underscore the significant role that various attention mechanisms play in enhanc- ing the accuracy of deep learning models tasked with analyzing complex visual data.

In a related vein, Wu et al. (2019) present a coun- terintuitive perspective in the domain of natural language processing (NLP). They propose that employ- ing less attention can sometimes lead to better outcomes. Their research introduces a novel approach using lightweight and dynamic convolutions designed to lessen the typically high computational demands of content-based self-attention mechanisms, without sacrificing cutting-edge results Wu et al. (2019). This work provides a scalable solution that balances high performance with reduced computational needs. Col- lectively, these studies highlight the flexibility and effectiveness of attention mechanisms and dynamic convolutions in diverse fields, ranging from ecological informatics to natural language processing. The ongoing advancements and applications of these technologies are crucial for the development of more effi- cient and potent deep learning systems.

3. Methodology

This research proposes an enhanced ConvNeXt ar- chitecture tailored for industrial surface defect detection. The model is designed to efficiently process im- ages by leveraging advanced convolutional features, while specifically addressing the challenges associated with the variability and complexity of defect appearances in industrial settings. The architecture inte- grates several key modifications to improve detection accuracy and processing efficiency: deformable con- volutions, attention mechanisms, and multi-scale feature fusion. Figure 1 provides a schematic overview of the adapted ConvNeXt model.

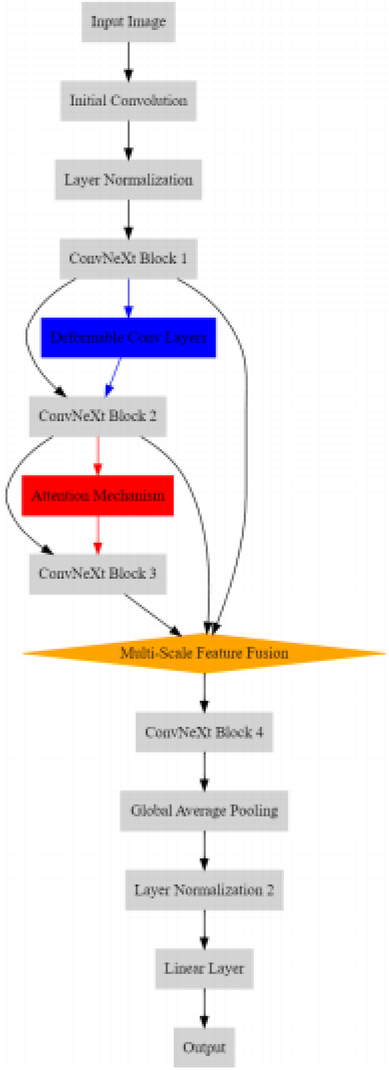

Fig. 1: Enhanced ConvNeXt Architecture incorporat- ing deformable convolutions, attention mechanisms, and multi-scale feature fusion.

The flow of data through the model starts with the input image, which undergoes initial preprocessing by a standard convolutional layer (Conv2d) to ex-tract basic features. These features are then normalized using layer normalization to stabilize the learn- ing process. Following this, the features enter the first of several ConvNeXt blocks, which have been modified to include deformable convolutional layers. These specialized layers allow the network to adjust its receptive field dynamically, enhancing its ability to accurately capture the irregular shapes of defects.

After the deformable convolution layers, an attention mechanism, specifically a Squeeze-and-Excitation (SE) block, is applied within each Con- vNeXt block. This mechanism focuses the network's processing capabilities on the most salient features, which is critical for detecting subtle and complex defect patterns.

Subsequent to the attention-enhanced feature ex- traction, a multi-scale feature fusion process is employed. This process combines features from vari- ous depths within the network, enabling the model to capture and utilize information from different scales effectively. The fusion is achieved through a combination of upsampling and concatenation followed by a convolution operation, which integrates these features into a comprehensive representation that feeds into the final stages of the network.

The processed features are then passed through additional downsampling layers to reduce dimensionality and increase the receptive field, followed by global average pooling to summarize the features into a single vector. Another layer normalization step is applied before the final linear layer, which classifies the presence and type of defects based on the learned features.

3.1. Deformable Convolutions

Deformable convolutions are an innovative addition to the standard convolutional layers, which enable the network to better adapt to geometric variations in the input data. Within the context of the ConvNeXt architecture, they enhance the model's ability to capture the diverse and irregular forms that industrial surface defects may take.

Given an input feature map X, a standard convo- lution operation at a location p_0 on the output feature map Y can be defined as:

$$Y(p_0) = \sum_{p_n \in \mathcal{R}} w(p_n) \cdot X(p_0 + p_n) \qquad (1)$$

where R is the regular grid of points over which the convolution kernel w is applied, and p_n indexes positions in the kernel. In contrast, the deformable convolution introduces an offset Δp_n to the regular grid positions p_n:

$$Y(p_0) = \sum_{p_n \in \mathcal{R}} w(p_n) \cdot X(p_0 + p_n + \Delta p_n) \qquad (2)$$

The offsets Δp_n are learned parameters that al- low the convolutional operation to deform according to the input features, hence the name "deformable." This dynamic adjustment of the kernel's receptive field to the data allows for more flexible and precise extraction of features, particularly around the edges and contours of defects.

Incorporating deformable convolutions into the ConvNeXt blocks, as illustrated in the flowchart, in- volves modifying the standard convolutional layers to include the learnable offsets. This is performed im- mediately after the initial downsampling layers to en- sure that the network captures high-resolution defect features before they are condensed by further down- sampling operations.

$$\Delta p_n = f_\theta(X) \qquad (3)$$

where f_θ represents a function parameterized by θ, typically a small convolutional network, that outputs the offsets from the input feature maps X.

The integration of deformable convolutions within the ConvNeXt architecture allows for the model to better handle the varied appearance of defects in in- dustrial images, aligning with the overall flow of the network and contributing to the enhanced defect de- tection performance.

3.2. Attention Mechanism

The integration of attention mechanisms within the ConvNeXt architecture significantly enhances the model’s capability to focus on salient features within the image, particularly useful for highlighting areas with potential defects. One effective form of attention used in this context is the Squeeze-and-Excitation (SE) block, which recalibrates channel-wise feature responses by explicitly modelling interdependencies between channels.

The SE block operates in two main phases: squeeze and excitation. Given an input feature map U from the previous ConvNeXt block, the squeeze operation first global average pools each channel, resulting in a descriptor that encapsulates global dis- tributional information of the channel’s features:

$$z_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} u_c(i, j) \qquad (4)$$

where z_c is the channel-wise descriptor of channel c, u_c (i, j) denotes the value at position (i, j) in chan- nel c, and H and W are the height and width of the feature map, respectively.

Following the squeeze operation, the excitation phase employs a simple gating mechanism with a sigmoid activation to capture channel-wise dependen- cies. This is achieved by:

$$s = \sigma(\mathbf{W}_2\, \delta(\mathbf{W}_1 z)) \qquad (5)$$

where z is the squeezed channel-wise descriptor, σ denotes the sigmoid function, δ represents the ReLU activation, and $\mathbf{W}_1$ and $\mathbf{W}_2$ are the parameters of two fully connected layers that scale down and then scale up the dimensions, respectively. The output s is a collection of modulation weights that are applied to the feature map:

$$\tilde{x}_c = s_c \cdot u_c \qquad (6)$$

where $\tilde{x}_c$ is the recalibrated feature map for chan- nel c and s_c is the channel-specific weight derived from the SE block.

By applying the SE attention block after each ConvNeXt block and before multi-scale feature fusion, the architecture is refined to focus more on po- tentially defective regions within the image, making the detection process more accurate and efficient.

This method of attention not only boosts the rep- resentational power of the network by focusing on important features but also aligns with the overall objective of enhancing defect detection accuracy in industrial settings.

3.3. Multi-Scale Feature Fusion

The integration of multi-scale feature fusion within the ConvNeXt architecture is designed to ensure that the model effectively captures and utilizes informa- tion from various scales, enhancing its ability to detect defects of different sizes and complexities. This approach leverages the inherent hierarchical nature of convolutional networks, where lower layers capture fine details and higher layers capture more abstract representations.

The process of multi-scale feature fusion involves combining feature maps from different layers of the network. This is typically implemented before the final classification layers, ensuring that the fused features contribute directly to the detection perfor- mance. The fusion process can be mathematically described as follows:

$$F_{fused} = \gamma(F_1, F_2, \ldots, F_n) \qquad (7)$$

where F_{fused} represents the fused feature map, F_1, $F_2, \ldots, F_n$ are the feature maps from different lay- ers or blocks within the network, and γ is a fusion function. Commonly, γ can be a concatenation fol- lowed by a convolution layer, or more sophisticated operations like weighted averaging or feature pyra- mids, which are designed to retain critical informa- tion across scales.

To implement this in the ConvNeXt framework, feature maps from selected ConvNeXt blocks are first adjusted to the same dimensionality, typically us- ing upsampling or downsampling techniques to match their spatial resolutions:

$$F_k' = \text{Resample}(F_k) \qquad (8)$$

where F_k' is the resampled feature map of F_k, and
Resample represents the resampling operation (either upsampling or downsampling) to match the size of the target feature map for fusion. After aligning the dimensions, the feature maps are combined using a fusion technique, which can in- volve learnable parameters to optimally blend fea- tures from different levels:

$$F_{\text{fused}} = \text{Conv}(\text{Concat}(F_1', F_2', \ldots, F_n')) \qquad (9)$$

where Concat denotes the concatenation of fea- ture maps, and Conv represents a convolution operation applied to the concatenated maps to produce a cohesive feature map that integrates multi-scale information effectively.

By employing multi-scale feature fusion, the Con- vNeXt model gains a more comprehensive understanding of the image, which is crucial for accurately detecting and classifying defects across various industrial scenarios. This method ensures that no detail, regardless of its scale, is overlooked, and enhances the model's robustness and accuracy.

4. Experiment and Results

4.1. Datasets

The performance of our method in accurately de- tecting and classifying surface defects is heavily influenced by the quality and scope of the datasets employed during its training and validation phases. This study employs two distinct industrial datasets, each designed to represent various defect types and industrial contexts, in order to thoroughly evaluate the efficacy of the proposed methodologies.

NEU Surface Defect Dataset The NEU Sur- face Defect Dataset Schlagenhauf and Landwehr (2021) comprises 1,800 images that specifically target defects appearing on hot-rolled steel strips. It classifies these defects into six unique categories, each sym- bolizing a different kind of surface anomaly: Crazing (Cr), Patches (Ps), Rolled-in Scale (Rs), Pitted Sur- face (Ps), Inclusion (In), and Scratches (Sc). The dataset ensures each category is equally represented with 300 images, subdividing them into 240 for training purposes and 60 for testing. This dataset offers a detailed portrayal of typical defects found in metal surfaces, including inclusions—both embedded and detachable, crazing that shows unavoidable surface cracks, patches that illuminate distinct metal char- acteristics, pitted surfaces marked by localized corrosion creating small cavities, and scratches indicative of surface abrasions. Rolled-in scale, another defect classification, refers to mill scale that becomes em- bedded in the metal during the production process. Figure 2 presents visual examples of each type of de- fect.

Ball Screw Drives Dataset Overview The Ball Screw Drives Dataset Schlagenhauf (2021) contains 21,835 high-resolution images in .png format. These images are divided into two distinct categories: P for pitting—representing localized failures on the surface, and N for no pitting, which includes im- ages free from surface defects. Specifically, the collec- tion has 11,075 images categorized as defect-free and 10,760 images marked with surface defects. For this research, 20% of the dataset is reserved for validation and testing, while the bulk of it, 80%, is designated for training. Examples of both defective and non- defective conditions from this dataset are showcased in Figures 3.

Both datasets are meticulously partitioned into training and testing groups to ensure comprehensive training and stringent evaluation under diverse con- ditions. Table 1 outlines the distribution and classification of the datasets, offering a detailed view of their structure and the particular challenges each defect type presents.

4.2. Experimental Setup

The experimental setup to assess the efficacy of In- dustNet is carefully crafted to ensure robustness and reproducibility in our findings. Here, we elaborate on the hardware and software configurations, parameter adjustments, and the metrics utilized to gauge the model's performance in identifying and categorizing surface defects in industrial goods.

4.2.1. Hardware and Software

Our experiments were executed on a computational platform featuring an NVIDIA GeForce RTX 3090 GPU, renowned for its adeptness in deep learning tasks owing to its high computational prowess and efficiency. Complementing this GPU is an Intel Core i9-10900K CPU and 32GB of RAM, facilitating seamless data management and processing capabilities.

In terms of software, we leveraged Python 3.8 alongside PyTorch 1.7.0, which provides a rich array of libraries and utilities for deep learning explo- ration. The dynamic computation graph of PyTorch facilitated effective model tuning and debugging. Ad- ditionally, CUDA 11.0 was employed to harness GPU acceleration, substantially reducing the time required for both training and testing phases.

4.2.2. Parameter Settings

IndustNet underwent training employing the Adam optimizer, lauded for its adaptive learning rate features that expedite model convergence. The initial learning rate was configured at 1e - 3, adjusted via a ReduceLROnPlateau schedule, which diminishes the learning rate by a factor of 0.1 in the absence of performance enhancements on the validation set over 20 consecutive epochs.

Training extended across 200 epochs, with early stopping mechanisms in place to avert overfitting. This halts the training process if the validation loss fails to improve over 20 consecutive epochs. To enhance model robustness against real-world varia- tions, data augmentation techniques such as random rotations, translations, and scaling (detailed in the Dataset subsection) were applied to the training im- ages.

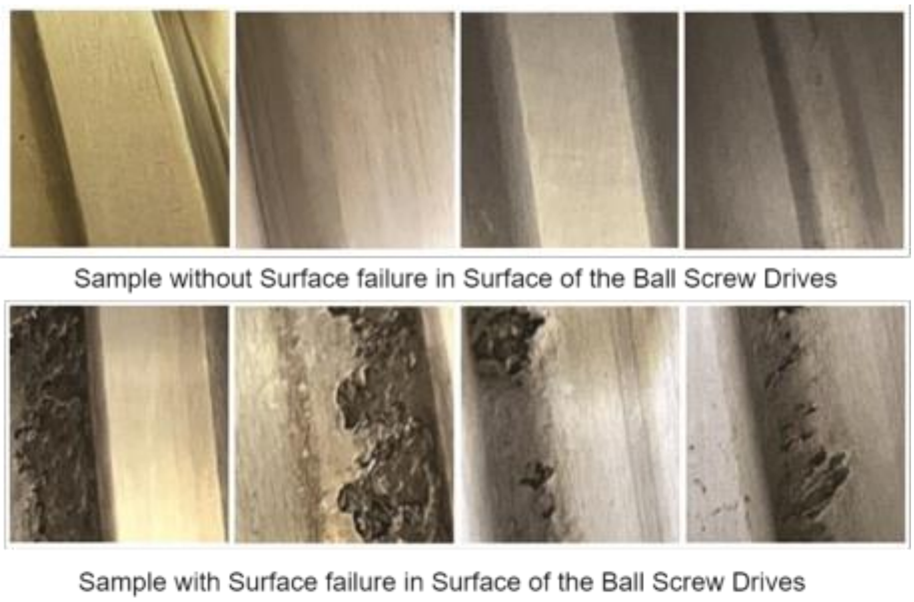
Sample without Surface failure in Surface of the Ball Screw Drives
Sample with Surface failure in Surface of the Ball Screw Drives

Fig. 2: Examples of NEU dataset.

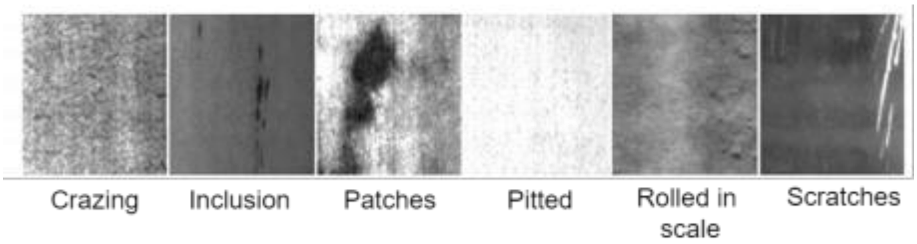

Fig. 3: Examples of Surface of the Ball Screw Drive.

4.2.3. Evaluation Metrics

The evaluation of IndustNet's performance incorpo- rates several pivotal metrics customary in defect detection studies:

- **Precision**: It denotes the ratio of accurately predicted positive observations to the total predicted positives, emphasizing a low false positive rate.
- **Recall (Sensitivity)**: This metric signifies the ratio of accurately predicted positive observations to all observations in the actual class.
- **F1-Score**: It represents the weighted average of Precision and Recall, offering significance when the costs of false positives and false negatives differ significantly.
- **Accuracy**: This metric gauges the ratio of cor- rectly predicted observations to the total observations, though its interpretation may be skewed in the presence of class imbalance.

5. Results Summary

The enhanced performance of our adapted ConvNeXt architecture was thoroughly evaluated using two specialized industrial datasets tailored to the detection and classification of surface defects. The results, as detailed in Table 2, underscore the significant im- provements achieved in precision, recall, F1-score, and overall accuracy across various defect categories.

For the NEU Surface Defect Dataset, our method demonstrated remarkable capabilities with precision values ranging from 92% to 99%, and recall rates from 91% to 99%. The model was particularly effective in identifying and classifying patches and inclusions, achieving a perfect F1-score of 99% for patches. The overall accuracy for this dataset stood at 95%, indi- cating robust performance across all types of defects.

Similarly, the evaluation on the Ball Screw Drives Dataset revealed an exceptional consistency in the detection of surface failures, with both precision and recall reaching as high as 99.5%. The accuracy for this dataset was impressive at 99%, highlighting the model’s effectiveness in environments with diverse de- fect characteristics.

These results validate the efficacy of integrating deformable convolutions, attention mechanisms, and multi-scale feature fusion within the ConvNeXt ar- chitecture. The modifications not only enhanced the model’s sensitivity to subtle and complex defect fea- tures but also improved its ability to generalize across different industrial contexts. Such advancements con- tribute substantially to the fields of automated visual inspection and quality control, promising significant reductions in manufacturing defects.

Table 1 - Characteristics of the Surface Defect Datasets

Dataset	Class label	Instances in Training	Instances in Testing	Total
NEU Surface Defect Dataset	Crazing (Cr)	240	60	300
	Patches (Ps)	240	60	300
	Rolled_in_Scale (Rs)	240	60	300
	Pitted surface (Ps)	240	60	300
	Inclusion (In)	240	60	300
	Scratches (Sc)	240	60	300
	Total	1440	360	1800
Surface of the Ball Screw Drives	Surface failure (P)	8608	2152	10760
	No Surface failure (N)	8860	2215	11075
	Total	17468	4367	21835

Table 1: This table summarizes the characteristics of the two major datasets used in this study, detailing their distribution across training and testing phases, along with total instances for each class label.

Table 2 Classification results

Dataset	Class Label	Precision (%)	Recall (%)	F1-Score (%)	Accuracy (%)
NEU Surface	Crazing (Cr)	96	95	95.5	
	Inclusion (In)	97	98	97.5	
	Patches (Ps)	99	99	99	

Defect dataset	Pitted_surface (Ps)	94	93	93.5	95
	Rolled in scale (Rs)	95	94	94.5	
	Scratches (Sc)	92	91	91.5	
Ball Screw Drives Dataset	Surface failure (P)	99.5	99.5	99.5	99
	No Surface failure (N)	99	99	99	

Table 2: Classification results showing enhanced performance metrics across two datasets for various defect types.

5.1. Ablation Study

The ablation study will be carried out by method- ically deactivating each critical component of our method one at a time, and then evaluating the ef- fects on performance metrics across both datasets.

The ablation study results, as summarized in Ta- ble 3, indicate that removing any of these components results in a measurable decrease in accuracy across both tested datasets. Specifically, the absence of deformable convolutions led to a reduction in ac- curacy by 1.5% on the NEU Surface Defect Dataset and 0.9% on the Ball Screw Drives Dataset. Simi- larly, omitting the attention mechanism and multi-scale feature fusion resulted in further performance declines. This underscores the synergistic effect of these components in enhancing the model’s capabil- ity to accurately identify and classify a wide range of defect types.

The full model, incorporating all proposed en- hancements, achieved an accuracy of 95.0% on the NEU Surface Defect Dataset and 99.0% on the Ball Screw Drives Dataset, setting new benchmarks for defect detection in industrial applications. These find- ings not only validate the proposed methodological enhancements but also highlight the potential of ad- vanced deep learning architectures in improving the quality control processes within manufacturing envi- ronments.

In conclusion, this research contributes signifi- cantly to the field of automated visual inspection by providing a robust, scalable, and highly accurate model for industrial defect detection. Future work will focus on further refining these techniques, ex- ploring additional datasets, and potentially integrating more advanced computational strategies to ex- tend the application scope and enhance real-world deployment effectiveness.

6. Conclusions

This research explored the development and enhance- ment of a ConvNeXt-based architecture specifically adapted for the challenging task of industrial surface defect detection. The study demonstrated the significant impact of incorporating deformable convo- lutions, attention mechanisms, and multi-scale feature fusion into the ConvNeXt framework. These enhancements collectively improved the model’s ability to accurately detect and classify a wide range of surface defects, reflecting a substantial advancement in automated visual inspection technology. The re- sults from extensive evaluations on two specialized datasets—the NEU Surface Defect Dataset and the Ball Screw Drives Dataset—highlight the effectiveness of the proposed model enhancements: The intro- duction of deformable convolutions allowed the model to adapt more flexibly to irregular defect shapes and sizes, enhancing detection accuracy. The integration of attention mechanisms helped to focus the model’s computational resources on the most salient features, significantly improving the precision of defect clas- sification. Multi-scale feature fusion enabled the model to effectively capture and utilize information across different scales, ensuring robust detection performance across various defect types. The enhanced model achieved high accuracy rates, with 95.0% on the NEU Surface Defect Dataset and 99.0% on the Ball Screw Drives Dataset, demonstrating its poten- tial to serve as a reliable tool in industrial quality control systems. These improvements suggest that the adoption of such advanced deep learning mod- els can lead to significant enhancements in manufacturing processes, potentially reducing costs asso- ciated with defects and ensuring higher quality standards. Future research will focus on expanding the capabilities of this model to include additional types of manufacturing materials and defect forms, explor- ing further integration with real-time manufacturing systems, and enhancing the model’s efficiency to fa- cilitate deployment in resource-constrained environments. Continuous improvement of the algorithms and exploration of new deep learning techniques will also be crucial to maintaining the relevance and effectiveness of this technology in rapidly evolving industrial scenarios.

In conclusion, the advancements presented in this research not only underscore the potential of deep learning in industrial applications but also pave the way for more sophisticated, accurate, and cost-effective solutions in automated manufacturing and quality control.

Table 3: Ablation Study Results on Accuracy

Model Configuration	Accuracy on NEU Surface Defect Dataset (%)	Accuracy on Ball Screw Drives Dataset (%)
w/o Deformable Convolutions	93.5	97.8
w/o Attention Mechanism	93.2	97.5
w/o Multi-Scale Fusion	93.0	97.2
Full Model	**95.0**	**99.0**

Table 3: Ablation study results demonstrating the impact of each component on the overall accuracy of the adapted ConvNeXt model across two datasets.

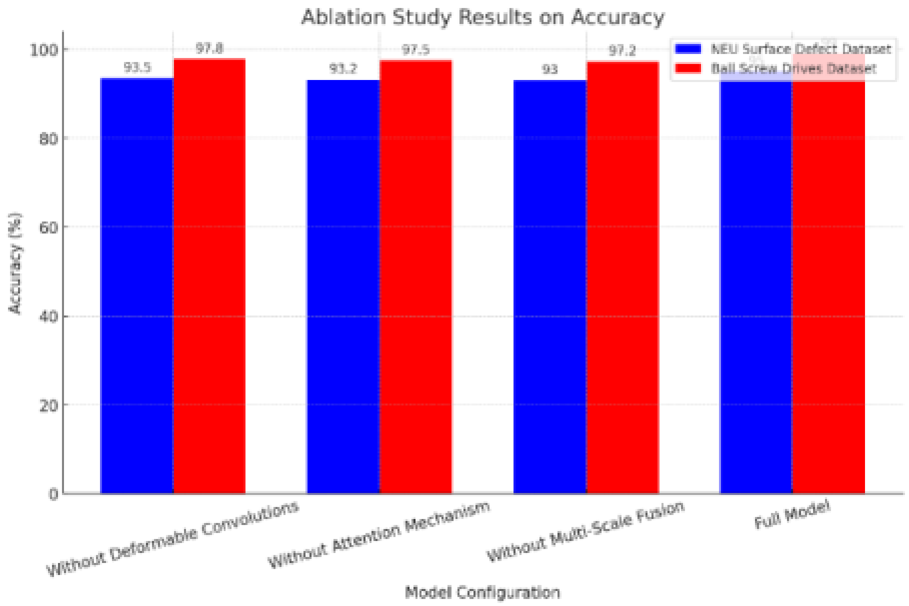

Fig. 4: Ablation study results.

REFERENCES

1. Akhyar, F., Liu, Y., Hsu, C.-Y., Shih, T. K., and Lin, C.- Y. 2023, The International Journal of Advanced Man- ufacturing Technology, 126, 1093
2. Blake, R. and Michalikova, K. F. 2021, Analysis and Metaphysics, 20, 159
3. Chen, W., Zou, B., Huang, C., et al. 2023, Ceramics In- ternational, 49, 2881
4. Chen, Y., Ding, Y., Zhao, F., et al. 2021, Applied Sci- ences, 11, 7657
5. Chin, R. T. 1988, Computer Vision, Graphics, and Image Processing, 41, 346
6. Francesco, G., Sonia, L., and Alessandro, N. 2018, in 2018 IEEE Power & Energy Society General Meeting (PESGM), IEEE, 1–5
7. Ghobakhloo, M. 2020, Journal of cleaner production, 252, 119869 Hofmann, E. and R¨usch, M. 2017, Computers in industry, 89, 23Li, W., Zhang, H., Wang, G., et al. 2023, Robotics andComputer-Integrated Manufacturing, 80, 102470
8. Li, Y. and Gu, P. 2004, Computer-Aided Design, 36, 1395

9. Liu, Z., Mao, H., Wu, C.-Y., et al. 2022, in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 11976–11986
10. Mishra, S. and Tyagi, A. K. 2022, Artificial intelligence- based internet of things systems,), 105
11. Newman, T. S. and Jain, A. K. 1995, Computer vision and image understanding, 61, 231
12. Ngan, H. Y., Pang, G. K., and Yung, N. H. 2011, Image and vision computing, 29, 442
13. Rath, M., Satpathy, J., and Oreku, G. S. 2021, in Arti- ficial intelligence to solve pervasive internet of things issues (Elsevier), 103–123
14. Saberironaghi, A., Ren, J., and El-Gindy, M. 2023, Algo- rithms, 16, 95Schlagenhauf, T. 2021,
15. Schlagenhauf, T. and Landwehr, M. 2021, Data in Brief, 39, 107643
16. Wu, F., Fan, A., Baevski, A., Dauphin, Y. N., and Auli, M. 2019, arXiv preprint arXiv:1901.10430,)
17. Xu, L. D., Xu, E. L., and Li, L. 2018, International journal of production research, 56, 2941
18. Yang, Y., Jiao, G., Liu, J., Zhao, W., and Zheng, J. 2023, Ecological Informatics, 78, 102320
19. Zhu, X., Cheng, D., Zhang, Z., Lin, S., and Dai, J. 2019, in Proceedings of the IEEE/CVF international confer- ence on computer vision, 6688–6697