Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 2s (2025)
https://www.jneonatalsurg.com

Defect Detection Approach in Manufacturing Environments using Customizable Convolutional Neural Networks with Multi-Scale Attention Mechanisms

Abhishek Pandey[1]*, S. Mohanraj [2]

[1]*Assistant Professor, CDOE, Manipal University Jaipur,

Email ID : abhishek.pandey@jaipur.manipal.edu

[2]Assistant Professor, Peri College of Arts and Science, Chennai,

Email ID : mohankanchi5@gmail.com

.Cite this paper as Abhishek Pandey, S. Mohanraj , (2025) Defect Detection Approach in Manufacturing Environments using Customizable Convolutional Neural Networks with Multi-Scale Attention Mechanisms..*Journal of Neonatal Surgery*, 14, (2s) 828-839

ABSTRACT

Three primary issues have hindered the ability of quality control systems to identify superficial faults in industrial manufacturing processes: fuzzy edges, fault characteristics with several scales, and geometric elements that are hard to identify. This paper suggests a surface defect detection technique based on combining pixel-level semantic segmentation with multi-scale data. Industrial defect identification requires real-time and high accuracy in complex, dynamic environments. Conventional image processing & machine learning focused on handcrafted features struggle to meet these needs. The answer to this issue proposed by this research is AENet, a revolutionary immediate form of defect identification network based on an encoder-decoder architecture. In addition to having excellent detection accuracy and efficiency, AENet also shows strong convergence and generalization. The method known as Neural Architecture Search (NAS) allows networks that are driven by data to autonomously generate and adapt. Here, we provide a novel method for surface defect identification using network adaptive design, called NAS-ASDet. This method can be applied to industrial settings to build low-data-weight, high-performance defect detection networks. Based on four datasets, the experimental results demonstrate that the suggested approach achieves a smaller model size and better performance compared to competing methods like manual and NAS-based techniques..

Keywords: *Machine Learning, NAS-ASDet, Conventional Image Processing, AENet, Segmentation, Multi-Scale, Data-Driven, Experimental Results.*

1. INTRODUCTION

Inspection of surface defects is a crucial step in the industrial manufacturing process and is necessary for maintaining product quality [1]. Due to its superior defect inspection performance with faster speed & higher accuracy, computer vision based autonomous defect inspection technologies have gained popularity in industrial production as compared to manual defect inspection methods [1, 2]. Conventional methods for defect inspection often see surface defect inspection as a texture analysis problem and make use of a number of time-tested techniques, including texture filters, texture statistics, texture modeling, and texture structure. These techniques are effective when the flaws are straightforward and mostly rely on particular texture information. Accurate defect inspection is greatly hampered by the surface flaws that are typically complex, diverse, and varied in size in real industrial settings.

As industrialization has progressed, [2], rotating machinery has found widespread use across several industries. Critical parts of rotating machinery are vulnerable to deformation, cracks, and other damages when operating due to harsh environmental influences and uneven working conditions [2, 3]. If they fail, the average production will suffer enormous financial losses, and personal safety may even be in jeopardy. To increase the operational dependability and stability of rotating equipment, it is crucial to research a rotating machinery condition identification method.

Numerous details may be found in the sound produced by the vibration of rotating machinery, and the method of identifying the health status of equipment based on vibration signal has become a hot topic for research. The construction of feature engineering is a critical step in rotating machinery health status recognition. Methods based on deep learning, signal processing, and a combination of feature extraction or pattern recognition are the main categories of rotating machinery health status recognition techniques [3, 4]. The accuracy of recognizing a person's health status is directly influenced by the quality of the feature set. Related researchers have recently used signal processing techniques to investigate the extraction of statistical parameters in the time, frequency, and time-frequency domains, and they have produced certain findings.

It is not feasible to construct reference pictures using conditional generative models or use visual-semantic approaches employed in Optical Character Recognition (OCR) due to the random and unexpected geometries of metal surface imperfections. Real-time detection of metal imperfections in the surface is severely hampered by this. Currently, traditional image processing & machine learning are the mainstays of machine vision algorithms for surface defect identification [5]. The concept of local abnormality reflection is the basis for the detection and segmentation of flaws in conventional image processing techniques. Moreover, they can be divided into model-based, spectral, threshold, and structural approaches [5].

Furthermore, a few projects (FHENet, BV-YOLOv5S) concentrate on designing lightweight detection network to overcome the difficulties associated with deploying big models in industrial settings. Even though these techniques yield good results, they still need to be improved:

Typically, they rely on a pre-established network connection. As a result, speed is limited since features are extracted in a predetermined and limiting manner rather than being dictated by data features.

Although some approaches attempt to create from scratch, they are typically tailored for certain needs [6, 7]. No single network has demonstrated proficiency in every detection job. As a result, creating effective networks for particular purposes requires a significant investment of time and computing power.

Neural Architecture Search (NAS) NAS seeks to optimize performance with minimal computational resources by automatically designing the neural architecture.

By transforming a discrete search area into a continuous differentiated form, the differentiable search method enables the application of gradient descent to the search process, surpassing the efficiency of black-box optimization techniques like evolutionary algorithms and reinforcement learning. DARTS contained the first gradient-based concept ever put forth. However, two issues with gradient-based DARTS still exist:

The entire search procedure involves full candidate operations, which increases computing overhead and search durations;

There may be variations in performance as a result of the rough decoupling transfer [8].

Neuronal Topology Searching for Image Segmentation

Most people agree that Auto-Deep Lab was the first to demonstrate the use of NAS in image segmentation. Additionally, it introduced the segmentation job to the gradient optimization-based technique for the first time. By employing NAS, NAS-UNet outperformed the original U-net and showed better performance. A multibank design was employed by FasterSeg to get around model breakdown. Compared to Auto-Deep Lab, which expanded the search space with cross-layer connections, [8, 9], DCNAS created a more intricate super net. To make training easier, DNAS created a three-level disconnected search approach. Additionally, some strategies consider the combined impact of NAS and hardware awareness and search for lighter architectures in an effort to remove the deployment strain of huge networks.

Using Region of Interest Pooling (RoI) to normalize feature regions of various scales, this technique first eliminates potential region features from various layers of the convolutional neural network's structure. Afterward, it merges these multi-scale features to enhance the ability to express regional information [10]. The integration of the target's surrounding context into a deep neural network is a topic of numerous studies. Additionally, using created adversarial samples for training, Wang et al. presented an enhanced detection model for small target occlusion & deformation based on Fast R-CNN [10, 11]. A network that creates occlusion and deformation characteristics automatically is incorporated into the model to improve its robustness to these conditions. Through the use of regional feature occlusion & deformation processing, the detection model can obtain a greater number of adversarial samples, hence improving the capability of the trained model.

Multiple scales Feature Fusion Convolutional Neuronal Networks for Object Detection

The SSD algorithm is a single-stage target recognition technique with a quick detection speed that can finish tasks like target identification and localization in a single step. Additionally, by employing previous boxes of various sizes and numbers on feature maps of various scales, SSD network integrates the regression concept of YOLO with the anchor boxes mechanism of FtP-RCNN to forecast multi-scale target objects [12]. The Prior box is an anchor framework that creates various lengths, widths, & aspect ratios by navigating map features with sliding windows of varying sizes. This is the SSD network model in Figure 1.

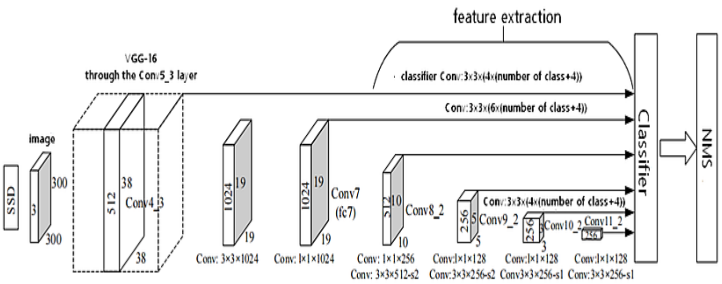

Fig. 1 SSD Network Architecture. [12, 13].

$S_k = S_{min} + \frac{S_{max}-S_{min}}{m-1}(k-1), k \in [1, m]$.....1

$\frac{[(S_{max}\times 100)-(S_{min}\times 100)]}{m-1} = 17$ 2

$$L(x, c, l, g) = \frac{1}{N}\left(L_{con\,f}(x, c) + \alpha L_{loc}(x, l, g)\right) \ldots..3$$

$L_{loc}(x, l, g) = \sum_{i\in Pos}^{N} \sum_{m\in\{cx,cy,w,h\}} x_{ij}^{k} smoth_{L1}\,(l_i^m - \widehat{g_j^m}$ 4

$L_{connf}(x, c) - \sum_{i\in Pos}^{N} x_{ij}^{p}\, Log\hat{c}_{i,}^{o}$ where 5

$\hat{c}_{i,}^{o} = \frac{\exp((\hat{c}_{i,}^{o})}{\sum_p \exp\,(\hat{c}_{i,}^{o})}$ 6

SSD Target Identification Algorithms Using Multiple Scales Feature Fusion

This research developed a multi-scale feature fusion approach that will improve the detection accuracy of SSD target identification algorithm in real complicated scenarios and encourage the use of detection of targets technology in service robots. For detection, the features of the neighboring layer and prediction layer were combined. In addition to fully utilizing multi-scale features, the upgraded SSD network also improves its complementarity of high- and low-level features, boosts the network's detection performance for multi-scale targets, and makes the model more useful in challenging situations [13]. Figure 2 displays the SSD network architecture based on multi-scale feature fusion.

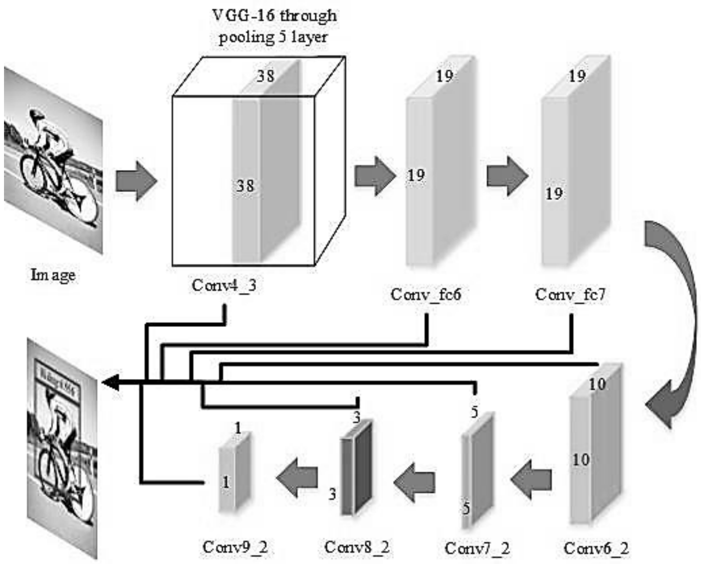

Fig. 2 SSD Network Architecture using multi-scale fusion of features. [14].

The target prediction layer is used as the benchmark, along with the distinctive characteristics of its adjacent layers—that is, features of different scales—are detected after fusion in order to ensure the target prediction layer's feature map remains constant in size and avoid issues with the feature's large spatial resolution after fusion, the high-level feature map's large disparity in information distribution, or the difficulty of learning the network later on. Down sampled feature maps are those below the prediction layer, and up sampled feature maps are those next to higher-level features.

The forecasting capability of late-stage variables is significantly impacted by the extraction of features and fusion techniques [15]. The extraction and fusing procedures for multiple-scale characteristics at all levels are created in this part to fully utilize them and lessen the impact of the network enhancement on late-stage data processing. Prior to merging features of various scales, the resolution consistency of mappings of features must be guaranteed. The processing scheme of neighboring characteristics of Conv3_3 and Conv5_3 is further examined using Conv4_3 as an example. Low-level features are those among those that are inferior to the targeted detecting layer, and high-level features are those that are lower than the goal identification layer [16].

While there were some current attempts to apply NAS in industrial settings, the majority of them are task-specific or restricted to defect groupings. NAS is used to classify steel fracture defects. In order to analyze faults in photoelectric cells in electrical luminescence images, a NAS-based detection strategy was developed, along with an automated defect identification technique to identify industry wood veneer [15, 16]. These techniques fall short of what is needed to robustly design accurate defect localization in a variety of industrial contexts. Consequently, there is still much to learn about how to create a pixel-level defect recognition system for NAS that works well in business environments.

Objectives

Develop a system that can identify flaws in real-time or almost real-time to allow for quick action and reduce production downtime.

By targeting pertinent regions at various levels of abstraction, the CNN model—which combines multi-scale attention mechanisms with its capacity to learn hierarchical features—aims to increase accuracy.

2. LITERATURE REVIEW

(Sun, J., Wang, P., 2019) [17] In the manufacturing industry, surface fault inspection is crucial for quality control. This research proposes a unique convolutional neural network-based surface fault inspection system based on Adapted Multiscale Image Collection (AMIC). Initially, the ImageNet data set is used to pretrained the inspection networks. Secondly, the AMIC is defined, which is composed of with-contour local extraction and adaptive multiscale picture extraction using training images. The training set is significantly expanded using the AMIC, and autonomous image labelling is possible without the need for artificial consumption. After that, transfer learning is carried out using the AMIC determined from the training set of data.

(Hou, X., Liu, M., 2023) [18] Even though Automatic Visual Inspection (AVI) has made significant progress in the manufacturing sector, its limited attention span and lack of semantic information make it difficult to detect small-sized flaws with less pixel coverage. The majority of convolutional inspection techniques now in use ignore the context's long-range reliance and don't have adaptive fusion procedures to take advantage of diverse data. This research offers a unique Context Information & Spatial Awareness Based Network (CANet) for effective recognition and exploitation of minor defect features in AVI. It consists of two steps: CAblock and Laplacian. In particular, Cab lock uses a Spatial Attention Encoder (SAE) and a Context Block Decoder (CBD) to extract semantically significant data with extensive context by encoding spatial long-range dependence & decoding contextual information as channel-specific bias, respectively.

(Wan, Y., Yi, L., Jiang, 2024) [19] The real time and high accuracy requirements for industry defect identification in complex, sensitive, & dynamic situations are difficult for conventional image processing & machine learning based on handmade features to achieve. This research introduces AENet, a novel real-time defect identification network based on an encoder-decoder model, to handle this problem. It shows great convergence and generalization along with high detection efficiency and accuracy. First, the encoding network's spatial attention channel module uses a multi-head 3D self-attention system to utilize both channel and spatial attention. Parallelism and detection efficiency are enhanced as a result. Second, the cross-level attention fusion module, which combines input information from several levels, is incorporated into the AENet decoding network. When used in conjunction with multi-level up sample design, the decoder improves defect detail representation. Additionally, we incorporate a streamlined aggregators into the encoder-decoder network to efficiently extract feature information at various scales.

(Wang, P., Yang, Y., 2022) [20] In order to improve our comprehension of the intricate process dynamics that underpin metals Additive Manufacturing (AM) operations, a great deal of research has been done in the last several years on Machine Learning (ML) techniques. In-depth analysis and discussion of the most recent successful ML applications to one class of metal AM processes—laser powder bed fusion, or LPBF—are presented in this study. Process modelling, in-situ process evaluation, defect identification, off-line process optimization, & on-line process control are the three LPBF facets that will be the main topics of this study. Owing to the multi-physics mechanisms involved in LPBF and the related heterogeneous process sensing, many machine learning algorithms inherently contribute significantly to the identification of patterns within sensing data. Unsupervised component evaluation aids in the fusion of features taken from sensory data to improve modelling and data processing efficiency.

(Wang, Y., Hu, S., 2020) [21] In the fields of crowd safety, on-site management scheduling, and scene monitoring, an increasing variety of crowd density estimation techniques have been created. We presented Multi-scale Dilated The process of convolution of Convolutional Neural Network (Multi-scale-CNN), a convolutional neural network-based technique for estimating the density of a single static image. The suggested technique used a convolutional neural network to learn the mapping connection between a single image and density maps using the density maps regression approach. The two main parts of the network structure that has been chosen to accommodate character scale variations in crowd photos are a multi-scale dilated convolution to solve the scale change problem and a Convolutional Neural Network (CNN) for general feature extraction. The existing body of research on multi-column or multi-input convolutional brain networks' ability to solve multi-scale challenges is insufficient. In order to overcome the limitations of two networks, our approach combines multi-scale dilated convolution networks with a single-column network for feature extraction.

(Huang, C., Wu, X., 2019) [22] In e-commerce systems, online purchase forecasting plays a crucial role as the foundation for providing customers with individualized, engaging product lists. However, given the multitude of elements that influence it, such as (i) the complicated temporal pattern with hierarchical correlations and (ii) arbitrary category dependencies, predicting online purchases is far from simple. In order to take advantage of users' latent behavioural patterns with multi-scale temporal dynamics & arbitrary inter-dependencies within product categories, we build a framework for Graph Multi-scale Pyramid networks (GMP) to address these issues. The multi-scale pyramid modulation structure of networks that we first propose for GMP effortlessly maintains the underlying hierarchy temporal elements that influence consumers' purchasing decisions. Next, to encode the category temporal pattern at each scale, we use convolution recurrent neural networks.

(Qi, X., Li, K., 2020) [23] One of the more difficult tasks in remote sensing image processing is remote sensing image segmentation. The segmentation of remote sensing images is very important for industries like agricultural planting and urban planning that require a lot of information about the terrain. The huge shooting angle, feature complexity, and ultra-high resolution of the sensor photos provide technical challenges for this task. We suggest ATD-LinkNet, a deep learning-based networks with multiple configurable modules, as a solution to these problems. In particular, we suggest utilizing multi-scale convolution as well as attention mechanisms as the ATD-Link Net component to be used in an interchangeable module called AT block.

(Xiang, L., Wang, P., 2021) [24] The intricate and dynamic the working atmosphere of wind turbines frequently presents difficulties for defect detection and monitoring the condition. This research proposes a novel approach for wind turbine failure detection, whereby an Attention Mechanism (AM)-based Convolutional neural network, also known as CNN, cascades to a Long- and Short-term Memory Neural Network (LSTM). CNN architecture is built using Supervisory Control and Data Acquisition (SCADA) data from wind turbines as input variables to extract dynamic changes in data. AM is used to increase the importance of key information. By mappings weighed and parameter learning, AM may assign distinct weights to concentrate on the features of LSTM and improve learning accuracy.

(Ravindranath, M., 2020) [25] As sensory data becomes increasingly accessible, determining whether or not observations contain pertinent events is becoming an essential challenge for smart data service provision in programmes that depend on these data sources. However, present methods frequently fall short when trying to deduce uncommon events, such as seizures from Electroencephalogram (EEG) data. In this paper, we observe that multi-variate time series frequently carry robust localised multi-variate time-related characteristics that could help identify these events, at least in principle; however, neural architectures are unable to identify and utilise these features due to insufficient data to train for these events. In order to address this issue, we suggest the M2N N neural architecture, which is based on LSTMs and has an approach to attention that makes use of reliable multivariate temporal variables that are pre-extracted and supplied as side information to the NN.

(Srivastava, A., 2021) [26] Text independence writer verification is a difficult problem that determines the handwritten text's author by analysing various handwriting styles. In the past, methods for identifying writers focused on hand-crafted characteristics to highlight the variations between each writer. With the development of convolutional neural networks, deep learning techniques have become more sophisticated recently. Three distinct deep learning methods were presented in this study to efficiently capture the variations in each writer's handwriting: patch-based CNN, multi-scale feature fusion, and spatial attention mechanism. Our techniques are predicated on the ideas that handwritten visuals contain distinct spatial regions that are more particular to a writer's style, features with multiple scales propagate characteristics specific to a writer, and patch-based features provide greater durability and general representations that aid in distinguishing between different writers' handwriting.

(Kong, W., Liu, S., 2023) [27] Research on the identification of ships in SAR photos is important because ship target recognition technology has many applications in both the military and the civilian world. A lightweight ship recognition networks founded on the YOLOx-Tiny architecture is proposed, aiming at tackling the complicated and diversified backdrops, considerable disparities in ship dimensions, ranging and immediate identification challenges in the ship target identification task of SAR satellite images. The first step involves proposing a multiple scales ship extraction of features module. This module consists of a parallel multi-branch architecture coupled by a dilatation convolution layers with varying expansions rates, an asymmetrical convolutional layer, and a conventional combination layer. The recognition reliability of multi-scale ship objects is effectively improved and both local and global features are utilised more effectively; Second, we suggest a comprehensive SAR remote sensing image recognition technique based on an adjustable thresholds, which successfully suppresses false alarms brought on by background and increases detection speed, in order to guarantee detection performance and remove external interference.

(Xu, C., 2022) [28] The level set method is a crucial technology in the water extraction process that has been widely employed for picture segmentation. Finding the right starting surface variables, particularly will impact level set evolution's accuracy and speed, is one of the level-set technique's challenges. Deep learning-based segmentation of semantics has recently created interesting new research opportunities. Additionally, a good feature representation capability has been established by the Convolutional Neural Network as (CNN). With the goal to give deep a priori knowledge for the zero-level setup curve, which is which just has to depict the broad outline of a body of water rather than its precise borders, the CNN approach is utilised in this paper to produce the first SAR image segmentation map.

(Jiao, J., Zhao, M., 2020) [29] Machines troubleshooting has grown more important as the industrial sector develops quickly in order to guarantee the production and operation of safe equipment. As a result, a wide range of approaches have been investigated and created in recent years, with clever algorithms developing at a particularly quick pace. In the last five years, a great deal of research and application has been done on Convolutional Neural Networks (CNNs), which are a common example of intelligent diagnostic models. A substantial amount of this research has been published in academic publications and the proceedings of conferences. Nevertheless, no systematic review has been conducted to address these studies and provide a basis for future research.

3. METHODOLOGY

System Structure

We concentrate on surface defect detection or suggest an adaptive network design approach, termed NAS-ASDet, in light of the difficulties associated with manually creating detection networks or the difficulties brought on by the particular industrial needs for NAS. Surface defect identification is considered a segmentation task at the pixel level in this method [30]. With the help of NAS-ASDet, the network can ultimately create a specific detection network that is both highly effective and lightweight by adaptively adjusting the connection mode based on the characteristics of the data. Figure 3 illustrates the proposed method's adaptive network design process.

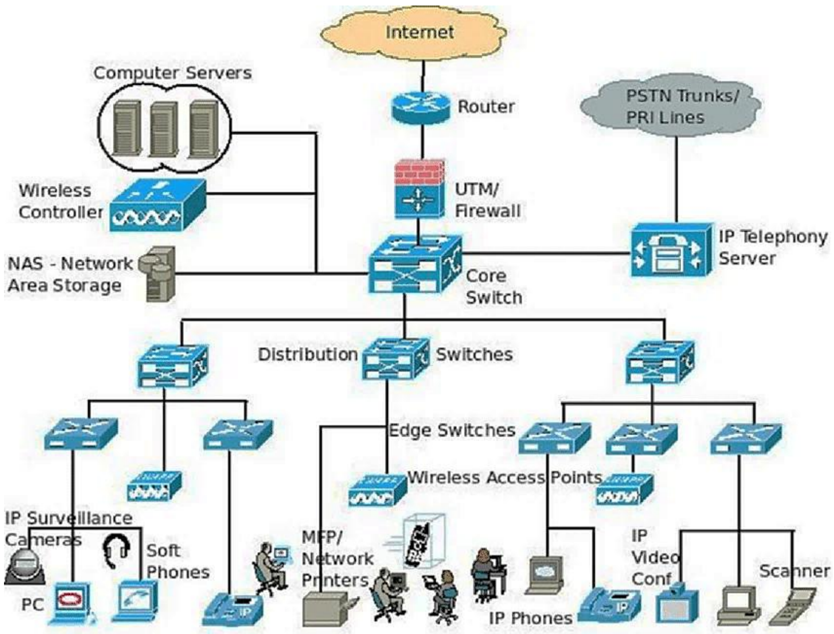

Fig. 3 The framework for networking design of architecture in NAS-ASDet. [30].

Cell-Level Searching Space

Since many handcrafted architectures are based on repeated fixed blocks as well as inspired, we design the search space for the lightweight network of things based on two distinct kinds of repeatedly stacked lightweight cells: normal cells as well as reduction cells. We do this because a small search space size should be used to reduce search difficulty. These two cells are built in the same way, with the regular cell's output feature map having the same dimensions as the input and the reduction cell's size being half that of the former.

Given the lightweight and low-computation needs of manufacturing environments, [30, 31], we first aggregate common detection network operations (such convolution, pooling, etc.) and employ separable convolutions rather than regular the convolutions to restrict cell light weighting. Table 1 summarizes the final complete candidate operation set O.

Table 1 Applicant operations set in NAS-ASDet. [31].

ID	Operation	Note
1	[Identify]	Skip Operation
2	[Zeros]	Skip connections
3	[Sep_conv_4×4]	3×3 separable convolutions
4	[Dil_conv_2×2]	3×3 dilate separable convolutions

5	[Sep_conc_8×8]	4×4 separable convolutions
6	[Sep_conv_5×5]	6×6 separable convolutions
7	[Max_pool_2×2]	3×3 max pooling's
8	[Ave_pool_3×3]	5×5 avg. pooling's
9	[Channel_att]	Channels attentions
10	[Spatial_att]	Spatials attentions

$$\bar{O}^{(i,j)}(x_i) \sum_{O \in O} \frac{e^{\alpha_O^{(i,j)}}}{\sum_{O' \epsilon O} e^{\alpha_{O'}^{(i,j)}}} O(x_i), (i < j) \text{ ....7}$$

$$n^{(j)} = \sum_{i<j} \bar{O}(i,j) \left(n^{(i)} + \sum_{m=k-2}^{k-1} \bar{O}(m,j)(c_m)\right) \text{ ......8}$$

$$C_k = \text{concat}(n^{(i)} + \text{res}(C_k - 1), i \in \{1, \ldots, N-3\} \text{ .....9}$$

$$A(Q) \cong \xrightarrow[\{C_{nor}, C_{reds}, \ldots, C_{nor}, C_{red}\}]{} \text{ ....10}$$

Network-Level Searching Area

Target prediction in the Auto-Deep Lab-invented segmentation NAS typically makes use of single-level characteristics. On the other hand, single-layer features only partially identify different scale target due to the variability of fault scales. We are inspired to incorporate this useful knowledge of multiscale feature fusion into the suggested NAS framework in order to improve the network's capacity to handle multi-scale challenges, given that the efficacy of bringing together multiscale features for enhancement of performance has been demonstrated in handmade architectures [31, 32].

Procedure for Searching

To explore a search space, we present a progressive search technique using a deep monitoring mechanism in this section. The original DARTS had a rough decoupling and additional search expenses. This search approach helps to lower the expenses associated with network construction while also improving detection performance and reducing optimization issues in intricate search procedures. Figure 4 displays the search diagram in its entirety.

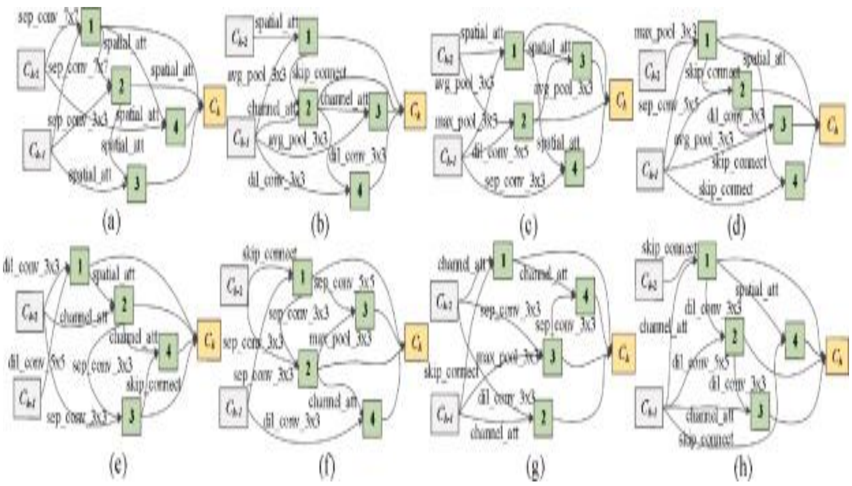

Fig. 4 The comprehensive supervisory combined with progressing search approach in NAS-ASDet. [32].

4. EXPERIMENTAL RESULTS

Collections of Data

We conducted experiments on four industry datasets with different obstacles (e.g., limited sample sizes, scale differences, lighting conditions) to assess the performance of NAS-ASdet.

MCSD-C datasets: This dataset is derived from several batches of surface flaws in motor commutator cylinders. Samples of representative surface defects that can be seen across the the commutator cylindrical surface are displayed in Figure 4. With 256×256, we chose 566 faulty samples [32, 33]. The remaining samples were used for testing, and 445 photos were used for training. Example defect images & accompanying ground truth in MCSD-C are displayed in Figure 5 rows.

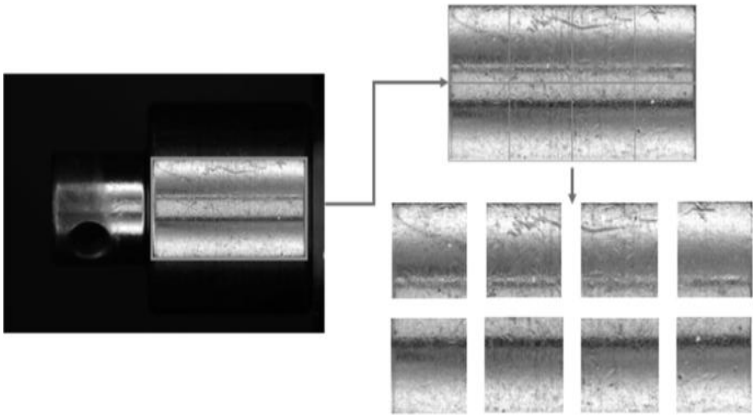

Fig. 5 Commutator cylinder surface and its surface defect samples. [33].

RSDDs datasets: There are 128 fault samples of 55x1250 pixels in RSDDs. After adding a 55-pixel window that slides to the original photos, we have 191 training and 82 testing images. During training, images are downsized to 64 by 64. Figure. Example defective images and their accompanying ground truth are displayed in RSDDs in 5 rows.

KSDD datasets: It has only 54 fault samples and was taken in a controlled manufacturing setting with apparent surface cracks. A 500-pixel sliding window is added to the faulty samples, producing 191 training and 82 testing images. KSDD mainly uses a small sample size of photos to assess the capabilities for low-contrast faults. During training, images are downsized to 512 x 512.

DAGM datasets: Ellipses roughly cover the defect spots in the original DAGM dataset. Four different kinds of flaws are chosen for our experiment, and we redefine their labels at the pixel level. At last, we have 523 training photos & 255 test images with an initial resolution of 512 × 512. Exemplary defect images and accompanying ground truth in DAGM are displayed in Figure 5 rows.

Monitoring of Performance

Table 2 displays the outcomes of the NAS-ASDet quantitative analysis & visual defect prediction, respectively [33, 34]. To achieve the best detection performance, our suggested method performs better overall than previous methods in terms of model performance, complexity of computation, & visual outcomes.

Table 2 Comparing the performance of several methods on industrial datasets that include NAS and homemade networks. [34].

Dataset	Metrics	Designing architectures by design									Architectures based on NAS				Ours
		FCN	U-Net	PSPNet	DeepLabV63+	CSEPNet	TSERNet	PGANet	CSEPNet	LSANet	Auto DeepLab	NAS Unet	iNAS	DNAS	NAS-AS Det
MCSD-C	IoU (%)↑	67.52	65.12	67.0.9	68.65	70.80	72.0.5	71.30	70.05	74.26	66.14	69.14	64.15	69.55	**71.99**
	PA (%)↑	80.20	76.20	80.37	81.50	84.96	80.18	81.78	85.17	76.89	78.59	78.56	79.60	79.15	**83.40**
	F1 (%) ↑	81.11	81.98	82.55	82.65	81.60	84.70	82.44	84.75	84.46	81.36	71.49	80.04	80.90	**85.62**
	Params (M) ↓	31.95	6.80	47.70	0.15	17.75	190.5	26.45	187.5	26.40	2.01	2.00	5.31	8.15	**1.45**
	FLOPs (G) ↓	70.45	27.21	91.43	1.02	119.66	532.84	31.78	531.85	31.79	26.41	14.65	0.84	15.94	**2.30**

	Search time ↓	/	/	/	/	/	/	/	/	/	6:40’17”	4:12’0.2”	9:30’02”	5:14:18”	**1:39’28”**
RSDDs	**IoU (%)↑**	57.10	62.60	54.60	60.45	51.84	61.80	64.41	64.66	64.06	51.01	60.08	60.30	18.30	**66.81**
	PA (%)↑	70.0.2	74.02	70.61	73.77	70.78	74.15	78.95	74.30	75.14	66.14	74.80	75.20	55.31	**81.21**
	F1 (%) ↑	72.90	80.60	77.35	75.20	71.65	76.78	81.26	79.95	79.14	70.39	74.09	4.30	22.85	**80.74**
	Params (M) ↓	31.95	8.86	42.68	40.74	0.78	50.40	17.77	195.67	77.25	5.50	0.59	6.35	1.83	**1.25**
	FLOPs (G) ↓	4.30	2.81	5.80	1.89	1.15	53.50	7.40	33.26	1.95	0.75	0.54	0.05	2.59	**0.20**
	Search time ↓	/	/	/	/	/	/	/	/	/	3:13’12”	4:5’13”	1:50’31”	1:45’16”	**1:41’26”**
KSDD	**IoU (%)↑**	65.07	60.14	65.50	67.85	64.27	68.15	71.40	6702	70.65	70.34	74.90	66.87	66.05	**74.79**
	PA (%)↑	78.25	84.20	78.79	81.90	80.89	80.81	80.85	78.85	80.54	80.65	80.02	79.25	88.45	**82.54**
	F1 (%) ↑	81.90	83.20	81.98	80.90	81.40	76.84	17.80	80.98	84.45	86.65	84.55	80.9	7.65	**85.26**
	Params (M) ↓	31.95	7.86	45.70	38.77	0.70	50.46	15.36	195.67	25.40	4.12	0.75	5.010	99.05	**1.65**
	FLOPs (G) ↓	279.71	113.80	368.45	120.55	8.50	3299.13	475.70	2230.30	145.010	109.50	45.69	2.45	88.94	**7.98**
	Search time ↓	/	/	/	/	/	/	/	/	/	5:36’54”	8:24’50”	5:02’65”	6:05’50”	**2:36’29”**
DAGM	**IoU (%)↑**	69.43	77.45	79.80	80.35	81.010	78.82	80.80	88.85	88.65	80.29	78.99	78.08	78.15	**81.65**
	PA (%)↑	80.20	78.20	88.30	87.05	87.65	89.25	87.56	89.65	87.56	86.45	87.03	8589	85.75	**88.07**
	F1 (%) ↑	81.50	88.54	89.85	87.75	87.15	87.89	87.54	87.20	98.85	87.15	89.17	86.99	89.70	**89.50**
	Params (M) ↓	33.95	6.87	45.70	45.70	0.50	50.41	88.20	195.50	26.49	5.08	0.65	5.58	15.05	**2.27**
	FLOPs (G) ↓	280.75	120.88	380.45	130.56	7.100	3325.15	48170	2189.30	155.010	55.03	50.05	2.65	26.56	**7.60**
	Search time ↓	/	/	/	/	/	/	/	/	/	40:12’15”	20:06’25”	16:60’04”	11:30’15”	**9:27’35”**

Research on ablation

We design a set of evaluation ablation experiments to investigate the impact of designs in NAS-ASDet on performance [35].

Extraction of multiscale features driven by data

Repetitive cell searches are used in NAS-ASDet to extract features. We put this strategy up against popular classical feature extraction techniques [34, 35]. We keep the other parts of NAS-ASDet the same and replace the process of extracting features individually during the comparison. The limited performance and fixed features of the fixed connection method are evident in Table 3.

Table 3 Comparing several feature extraction techniques quantitatively using industrial datasets. [35, 36].

Methods	RSDDs		MCSD-C	
	Iou (%)	Params	Iou (%)	Params
VGG16	65.93	14.5M	54.96	54.95M
ResNet	64.98	26.3M	15.82	61.6M
Mobile_NetV3	63.64	7.6M	27.54	25.6M
Dense Net	68.96	9.0M	63.59	54.7M
GoogLeNet	62.98	2.6M	21.59	21.4M
Ours	98.89	2.9M	69.48	1.6M

The results of the trials demonstrate that adaptive fusion performs better than balanced fusion as well as single-level features [37, 38]. Figure 6 (a), (b) illustrates the effectiveness of adaptively adjusting the feature importance distributions.

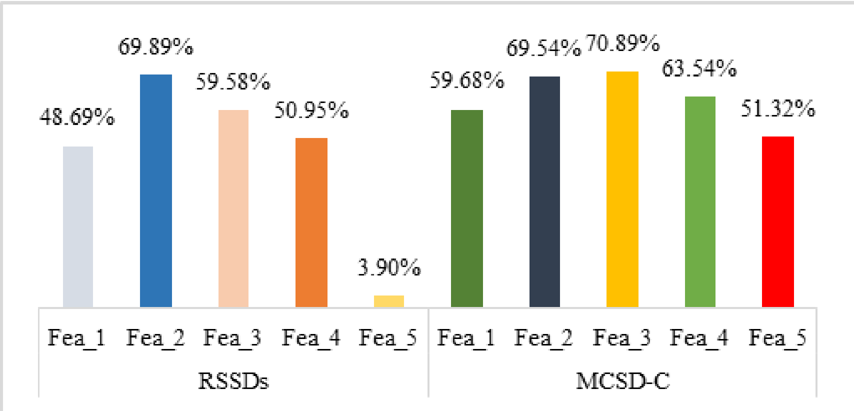

Fig. 6 (a) Comprehensive performance of various multiscale characteristics. [38].

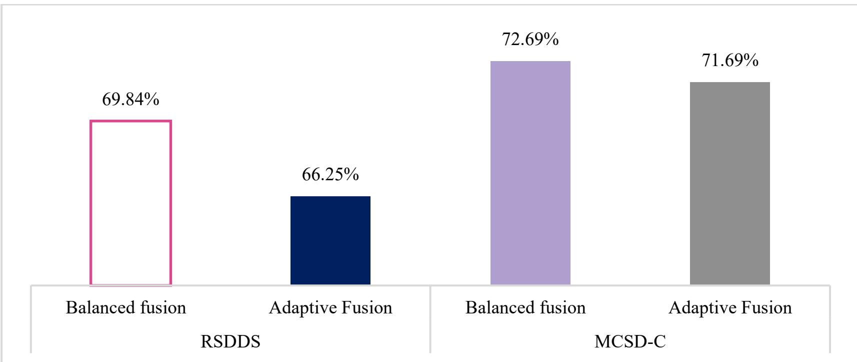

Fig. 6 (b) Comprehensive performance of various multiscale characteristics. [38].

5. CONCLUSION

In this paper, we offer NAS-ASDet, a NAS-based approach for generating adaptive architecture in surface defect detection. First, we leverage past manual architecture experience to design a streamlined & industry-relevant lightweight search space. We then provide a cell design that features searchable attention operations and is data-driven. Furthermore, we develop an adaptive multi-scale feature fusion system that may modify the distribution of features. Moreover, a deep supervision progressive search approach is intended for effectively investigating the search space. The outcomes of the experiments show that NAS-ASDet performs better than the manually created architectures and NAS ones.

Future work

In subsequent research, we intend to further lower the expressive limitation brought about by search space by adding more degrees of freedom in architectural search (e.g., scalable architecture). We will simultaneously attempt to use hardware-aware NAS to increase the inferences speed. Further research and development could be conducted using our technology in conjunction with an efficient data augmentation strategy

REFERENCES

[1] C. Xie, C. Xia, M. Ma, Z. Zhao, X. Chen, and J. Li, "Pyramid grafting network for one-stage high resolution saliency detection," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 11717–11726, 2022.

[2] Y. He, K. Song, Q. Meng, and Y. Yan, "An end-to-end steel surface defect detection approach via fusing multiple hierarchical features," IEEE Transactions on Instrumentation and Measurement, vol. 69, no. 4, pp. 1493–1504, 2019.

[3] B. Wei, K. Hao, L. Gao, and X.-s. Tang, "Detecting textile microdefects: A novel and efficient method based on visual gain mechanism," Information Sciences, vol. 541, pp. 60–74, 2020.

[4] L. Cui, X. Jiang, M. Xu, W. Li, P. Lv, and B. Zhou, "Sddnet: A fast and accurate network for surface defect detection," IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1–13, 2021.

[5] B. Su, H. Chen, K. Liu, and W. Liu, "Rcag-net: Residual channelwise attention gate network for hot spot defect detection of photovoltaic farms," IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1– 14, 2021.

[6] Y. Tu, Z. Ling, S. Guo, and H. Wen, "An accurate and real-time surface defects detection method for sawn lumber," IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1–11, 2020.

[7] Y. Huang, C. Qiu, X. Wang, S. Wang, and K. Yuan, "A compact convolutional neural network for surface defect inspection," Sensors, vol. 20, no. 7, p. 1974, 2020.

[8] D. Zhang, K. Song, J. Xu, Y. He, M. Niu, and Y. Yan, "Mcnet: Multiple context information segmentation network of no-service rail surface defects," IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1–9, 2021.

[9] Y. Li and J. Li, "An end-to-end defect detection method for mobile phone light guide plate via multitask learning," IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1–13, 2021.

[10] Gu, Z., Chen, H., Xu, Z., Lan, J., Meng, C., Wang, W.: DiffusionInst: Diffusion Model for Instance Segmentation (2022)

[11] Yuan, X.-c., Wu, L.-s., Peng, Q.: An improved otsu method using the weighted object variance for defect detection. Applied Surface Science 349, 472–484 (2015).

[12] Boukdir, A., Benaddy, M., Meslouhi, O.E., Kardouchi, M., Akhloufi, M.: Character-level arabic text generation from sign language video using encoderdecoder model. Displays 76, 102340 (2022).

[13] Lin, Y.: Wafer pattern counting, detection and classification based on encoderdecoder cnn structure. In: 2022 Intermountain Engineering, Technology and Computing (IETC), pp. 1–5 (2022).

[14] Wang, S., Wang, H., She, S., Zhang, Y., Qiu, Q., Xiao, Z.: Swin-t-nfc crfs: An encoder–decoder neural model for high-precision uav positioning via point cloud super resolution and image semantic segmentation. Comput. Commun. 197(C), 52–60 (2023).

[15] Jie, F., Nie, Q., Li, M., Yin, M., Jin, T.: Atrous spatial pyramid convolution for object detection with encoder-decoder. Neurocomputing 464, 107–118 (2021).

[16] Wang, C., Wang, H.: Cascaded feature fusion with multi-level self-attention mechanism for object detection. Pattern Recognition 138, 109377 (2023).

[17] Sun, J., Wang, P., Luo, Y. K., & Li, W. (2019). Surface defects detection based on adaptive multiscale image collection and convolutional neural networks. IEEE Transactions on Instrumentation and Measurement, 68(12),

4787-4797.

[18] Hou, X., Liu, M., Zhang, S., Wei, P., & Chen, B. (2023). CANet: Contextual Information and Spatial Attention Based Network for Detecting Small Defects in Manufacturing Industry. Pattern Recognition, 140, 109558.

[19] Wan, Y., Yi, L., Jiang, B., Chen, J., Jiang, Y., & Xie, X. (2024). AENet: attention enhancement network for industrial defect detection in complex and sensitive scenarios. The Journal of Supercomputing, 1-24.

[20] Wang, P., Yang, Y., & Moghaddam, N. S. (2022). Process modeling in laser powder bed fusion towards defect detection and quality control via machine learning: The state-of-the-art and research challenges. Journal of Manufacturing Processes, 73, 961-984.

[21] Wang, Y., Hu, S., Wang, G., Chen, C., & Pan, Z. (2020). Multi-scale dilated convolution of convolutional neural network for crowd counting. Multimedia Tools and Applications, 79, 1057-1073.

[22] Huang, C., Wu, X., Zhang, X., Zhang, C., Zhao, J., Yin, D., & Chawla, N. V. (2019, July). Online purchase prediction via multi-scale modeling of behavior dynamics. In Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining (pp. 2613-2622).

[23] Qi, X., Li, K., Liu, P., Zhou, X., & Sun, M. (2020). Deep attention and multi-scale networks for accurate remote sensing image segmentation. IEEE Access, 8, 146627-146639.

[24] Xiang, L., Wang, P., Yang, X., Hu, A., & Su, H. (2021). Fault detection of wind turbine based on SCADA data analysis using CNN and LSTM with attention mechanism. Measurement, 175, 109094.

[25] Ravindranath, M., Candan, K. S., & Sapino, M. L. (2020, October). M2NN: Rare event inference through multi-variate multi-scale attention. In 2020 IEEE International Conference on Smart Data Services (SMDS) (pp. 53-62). IEEE.

[26] Srivastava, A., Chanda, S., & Pal, U. (2021, November). Exploiting multi-scale fusion, spatial attention and patch interaction techniques for text-independent writer identification. In Asian Conference on Pattern Recognition (pp. 203-217). Cham: Springer International Publishing.

[27] Kong, W., Liu, S., Xu, M., Yasir, M., Wang, D., & Liu, W. (2023). Lightweight algorithm for multi-scale ship detection based on high-resolution SAR images. International Journal of Remote Sensing, 44(4), 1390-1415.

[28] Xu, C., Zhang, S., Zhao, B., Liu, C., Sui, H., Yang, W., & Mei, L. (2022). SAR image water extraction using the attention U-net and multi-scale level set method: flood monitoring in South China in 2020 as a test case. Geo-Spatial Information Science, 25(2), 155-168.

[29] Jiao, J., Zhao, M., Lin, J., & Liang, K. (2020). A comprehensive review on convolutional neural network in machine fault diagnosis. Neurocomputing, 417, 36-63.

[30] Guo, R.; Wang, T.; Li, Y.T. Health Status Recognition of Gear Pump Based on Bayesian-LSTM. J/OL. J. Mech. Eng. 2023, 59, 1–9.

[31] Rana, A.S.; Thomas, M.S.; Senroy, N. Reliability evaluation of WAMS using Markov-based graph theory approach. Iet Gener. Transm. Distrib. 2017, 11, 2930–2937.

[32] Zhao, R.; Yan, R.; Chen, Z. Deep learning and its applications to machine health monitoring. Mech. Syst. Signal Process. 2019, 115, 213–237.

[33] Chen, R.X.; Zhang, Y.; Hu, X.L. Health state recognition of harmonic reducer based on depth feature learning of voltage signal. Chin. J. Sci. Instrum. 2021, 42, 234–241.

[34] Chen, Z.Q.; Chen, X.D.; José, V.D.O. Application of deep learning in equipment prognostics and health management. Chin. J. Sci. Instrum. 2019, 40, 206–226.

[35] Lei, Y.G.; Jia, F.; Zhou, X. Health monitoring method for big data of mechanical equipment based on deep learning theory. J. Mech. Eng. 2015, 51, 49–56.

[36] Chen, H.P.; Hu, N.Q.; Cheng, Z. A deep convolutional neural network based fusion method of two-direction vibration signal data for health state identification of planetary gearboxes. Measurement 2019, 146, 268–278.

[37] Zhang, Y.; Feng, Q.Q.; Huang, Q.Q. Rolling bearing state recognition under variable condition using part-based representation of no negativity constrained auto encoder networks. Chin. J. Sci. Instrum. 2020, 41, 77–85.

[38] Chen, R.X.; Chen, S.Y.; Hu, X.L. Life status identification of rolling bearings under different working conditions by multi-source integrated GFK. J. Vib. Eng. 2020, 33, 614–621..