Deep Learning Based Feature Discriminability Boosted Concurrent Metal Surface Defect Detection System Using YOLOv-5s-FRN

Reshma Vengaloor
Department of Electronics and Communication Engineering, Faculty of Engineering and Technology, SRM Institute of Science and Technology, Ramapuram, India
reshmav@srmist.edu.in

Roopa Muralidhar
Department of Electronics and Communication Engineering, Faculty of Engineering and Technology, SRM Institute of Science and Technology, Ramapuram, India
roopam@srmist.edu.in

Abstract: *Computer vision and deep learning techniques are the most emerging technologies in this era. Both of these can greatly raise the rate at which defects on metal surfaces are identified while performing industrial quality checks. The identification of faults over metal surfaces can be viewed as a significant challenge since they are easily impacted by ambient factors including illumination and light reflections. This paper proposes novel metal surface defect detection network called as YOLOv-5s-FRN in response to the problems of ineffective detection brought by the conventional manual inspection system. The proposed system is developed through the integration of a novel architectural module called as Feature Recalibration Network (FRN) to the You Only Look Once-version-5 small network (YOLOv-5s). In order to extract the global feature information from the provided image, FRN is able to evaluate the interdependencies between the channels. This improves the feature discrimination capability and prediction accuracy of the defect detection system. The incorporation of FRN structure makes YOLOv-5s architecture to selectively enhance the necessary features and discard the unwanted ones. Therefore, the proposed novel method will efficiently detect and classify the metal surface defects such as crazing, patches, inclusions, scratches, pitted surfaces and rolled in scale. North Eastern University surface defect database (NEU-DET) has been used to train and test the proposed architectural model. The suggested system has been compared with alternative models based on several performance matrices such as precision, recall and Mean Average Precision (mAP). It is observed that the proposed YOLOv-5s-FRN architecture provides significant performance improvement than state-of-the-art methods. The proposed system has been provided satisfactory results by means of improvement in mAP and time consumption. The proposed model has delivered value of mAP_0.5 as 98.05% and that of mAP_0.5:0.95 as 89.03%.*

Keywords: *Classification, defect detection, deep learning, FRN structure, yolov5.*

Received September 18, 2023; accepted December 4, 2023
https://doi.org/10.34028/iajit/21/1/9

1. Introduction

Examining the product's surface is a crucial stage in ensuring the quality and production effectiveness. It is necessary to identify and reject the defective one to maintain the overall quality. The traditional manual inspection will cause various negative effects such as reduction in accuracy, efficiency, lower sampling rate and high labor intensity. The undergoing trend to overcome these issues significantly is the automated defect detection. This technique can be based on machine vision [7, 21], which will be able to capture the image of the product surface and then apply an image analysis technique to identify the surface defect.

In industrial applications, it is very common to appear the damages on textile, metallic and glass surfaces. Yet environmental elements like illumination, light reflection, etc., may readily influence the metal surfaces. The primary objective of the proposed system is to analyze the metal surfaces and identify flaws in them. The proposed method is concentrated to detect and classify the various metal surface defects called as crazing, patches, inclusions, scratches, pitted surfaces and rolled in scale as shown in Figure 1. The manufacturing systems can be equipped with computer vision algorithms to boost the production rate and reduce the error rates. The defect detection system must be able to provide the location of the defect in pixel level as well as the type of the defect. Only then the subsequent actions can be taken to raise the manufacturing quality. As a result, industries place a high priority on finding metal surface defects [26].

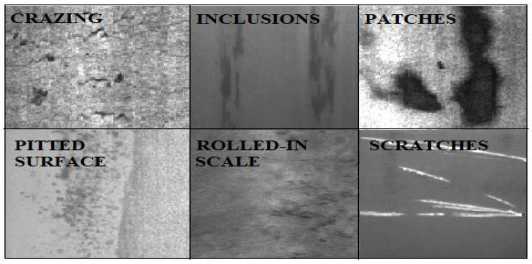


Figure 1. Various types of metal surface defects.

There are different image processing techniques such as thresholding-based method, segmentation-based method and edge detection-based method available. The feature-based methods along with the traditional machine learning techniques such as neural networks and Support Vector Machine (SVM) were used in the past decades for defect detection. Traditional image-processing techniques were describing the defects on the surfaces through hand crafted features. Local Binary Pattern (LBP), Histogram Oriented Gradient (HOG), Gray Level Co-Concurrence Matrix (GLCM) etc., are the commonly used hand-crafted features. It is crucial to recognize the exact features for the representation of the defects on the surface. Therefore, it is not advisable to use typical image processing algorithms for defect identification since they need complex threshold settings.

Deep learning techniques have recently gained popularity and it excels in a variety of applications. When compared to more traditional feature extraction techniques, they are effective at extracting features from the images. When it comes to recognizing defects or damages on metal surfaces, deep learning approaches are highly significant. You Only Look Once (YOLO) is one of the most prominent deep learning algorithms that is frequently utilized in industry. YOLO can be considered as a single stage target detection approach. It can be considered as one of the fastest algorithms, which exists today. YOLO is therefore suitable to situations where speed is a key factor without compromising accuracy and precision. Considering the advantages of the algorithm, the proposed system is developed on the basis of version 5 small network of YOLO (YOLOv-5s). YOLO will divide the input image into grid of size SxS. Instead of detecting the complete image, it is feasible to do so based on the grids. Then the vectors will be encoded for each grid to describe a cell. A cell can be represented as shown in Equation (1).

$$C_{11} = (P_c, B_x, B_y, B_w, B_h, C_1, C_2) \quad (1)$$

where, P_C stands for the probability of the object class, B_X and B_Y represent the center co-ordinates of the bounding box, B_w and B_h represent the width and height of the bounding box and C_1 and C_2 are the parameters that rely on the class of the object. The values of C_1 and C_2 can either be 0 or 1. Figure 2 shows an example of a cell, $C_{3,2}$.

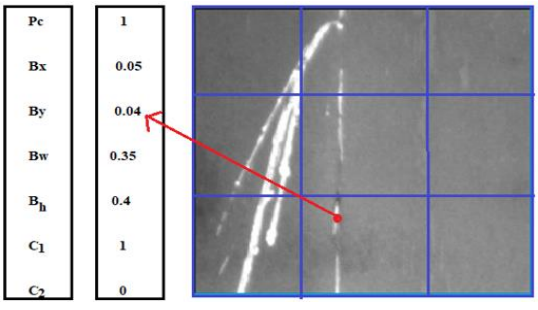


Figure 2. Representation of a cell.

In the proposed system, YOLOv-5s network is equipped with a novel architectural module called as called as Feature Recalibration Network (FRN) for enhancing the informative features and to suppress the unwanted ones. The FRN module will recalibrate the feature maps in order to get all of the global data from the image in order to do this. Therefore, the proposed YOLOv-5s-FRN can generate a noticeable performance boost with minimal computational expenses.

This paper is structured as follows: Section 2 describes problem statement and section 3 explains related works. Section 4 outlines the methodology of the proposed approach. Experimental results are described in section 5, where implementation details, performance analysis and comparison analysis of the proposed system has been explained. Following to this the conclusion as well as the future research scope have been explained.

2. Problem Statement

Though there are enormous researches and developments evolved in the domain of defect detection systems based on deep learning, achieving best value of accuracy and reduced time consumption is one of the challenges. Deep learning algorithms may not be able to provide satisfying performance parameters when the target defects are small and complex in nature. Developing a system which is able to extract important information from the given image can be a solution to the aforementioned issues. Such system has been proposed in this paper based on YOLOv-5s network with the development and integration of a new architectural module called as FRN.

3. Related Works

The implementation of deep learning techniques has yielded remarkable outcomes in industrial applications, and as a result, various innovative approaches for automated surface defect detection have been emerged. This section explains several existing techniques and architectures related to defect and object detection based on the application of deep learning.

Paper [30] proposes a method based on YOLOX for overcoming the issues in current detection methods like complex environment and defect morphology. The system has been provided mean average precision value as 78.45%. A Selective Prototype Network (SPNet) has been explained in [33] for improving the segmentation performance. It utilizes attention mechanism to train the model to understand novel surface classes with few labelled samples. An image processing-based technique has been presented in [25] for the development of smart farming system. In this, segmentation and feature extraction methods have been used to detect the diseased part of the plant and growth process of cultivations. The developed method is also capable to

identify the soil quality and atmospheric conditions.

A metal surface defect detection model is described in [27], where a Deep Encoder Representation from Transformers (DERT) model has been modified by introducing Span-sensitive Texture Fusion (STF structure). This technique places more emphasis on jump points as well as loss data recovery. As a result, the performance of defect identification will improve. In [29], author has presented a model for metal surface damage identification using modified YOLOv4. Enhancement of YOLOv4 is done by incorporating a Self-dependent Attentive Fusion (SAF) block, a Component Randomized Mosaic Augmentation (CRMA) scheme and a Perturbation Agnostic (PA) label smoothing method. A comprehensive assessment score is proposed in [12] for saving the time spent for training deep networks using datasets. The given method will assess the defect detection capability of deep models like YOLO, Single Shot multibox Detector (SSD) and Faster Region-based Convolutional Neural Network (Faster R-CNN) without training them using datasets.

An augmentation method for the detection of defects over small-scale images is explained in [32]. This approach consists of two parts: the generation part and augmentation component. For this, two pairs of generators and discriminators are utilized. Defect areas and background area of image will be generated in first part. The generated defect images and real defect images get concatenated in augmentation part. A defect detection system based on YOLOv4 is explained in [28]. It does feature map fusion to produce a real-time mask map through Generative Adversarial Networks (GANs). The experimental results indicate that 1% improvement in recognition accuracy without compromising execution speed. A multilabel classification technique based on deep learning methods has been proposed in [9]. In this different deep learning models like DenseNet121, MobileNet-V2 and ResNet-50 are trained using various dataset. The weighted F1 score was initially determined to be 91%, and after utilizing the weighted loss approach, it was determined to be 92%. A deep learning technique for implementing defect detection in smart factory is explained in [20]. Open dataset of steel defects is used for the study. The training process was completed with a 96% accuracy rate.

A deep artificial neural network has been developed for the purpose of identifying metal surface defects and is described in [1]. The given method has combined several convolutional methods with GAN. It will produce an output indicating the defect's class and location with 5.8% increase in average precision. A unique detection method based on a compact Convolutional Neural Network (CNN) has been described in [35] to confirm the existence of defects in the target region. Comparing with the traditional image processing methods, the given method has achieved 8.57x speed up and the detection kernel finished in 7ms per picture. An unsupervised learning method has been explained in [23] for detecting cavities over the surfaces and cavity lookalikes. In the first step, this method will locate the Region Of Interest (ROI), and then the ROIs will be analyzed using a depth estimator. The developed methods are experimented using real world dataset.

Gyimah *et al*. [3] explains a method to extract resilient features by combining Non-Local (NL) means with wavelet thresholding and Robust Completed Local Binary Pattern (RCLBP) integration. These features will then be applied to the classification module to determine the damages on surfaces. The aforementioned approach has a very high speed. But the feature extraction capability is not so effective in case of inconspicuous textures. A defect detection approach using two stage method has been explained in [22]. It includes a segmentation network and a decision network. The model has been tested and trained using Kolektor Surface- Defect Dataset (KolektorSDD). A multi-scale cascade CNN called MobileNet-V2-dense has been proposed by Lin *et al*. [13]. The Siamese Neural Network using CNN has been developed by Kim *et al*. [8] for the identification of the damages over the small number of images.

YOLO, a one-stage target identification method, is one of the frequently utilized algorithms as a result of the considerable advancement of deep learning techniques [11]. With the changes in size of the input and batches, a YOLOv3-based model was proposed by Hetab *et al*. [4] for defect identification over the NEU-DET dataset. 70.06%, a low figure, was found for the mean average precision. An improved version of YOLOv3 was developed by Kou *et al*. [10] for target defect detection of NEU-DET dataset. By combining deep neural network and random forest classifier, an early cancer detection technique has been proposed in [16]. It utilizes various mammogram images from a predefined dataset. A lightweight defect recognition model for steel strips has been presented in [34]. It designs a shallow feature enhancement module for the backbone structure of YOLOv5 network. The given method achieves good performance by means of detection accuracy and speed.

The YOLOv3 architecture was enhanced by Ning and Mi [18] by adding a prediction box layer, which allowed the system to detect tiny damages. In this the defect labels were detected using K-means++ algorithm and provided 14.7% increment in mean average precision. The addition of the prediction layer resulted in an increase in the size of the system, which decreases the speed of detection. A Single Shot Multi-Box Detector has been developed by Lv *et al*. [14]. In order to train the newly built model, a new dataset with the name GC10-DET was created. A study on the identification of damages over steel strip surfaces was carried out by Fu *et al*. [2] utilizing transfer learning as the method of analysis. To do this, a pretrained VGG16 model was utilized as the basis for the feature method,

while CNN was used for the classification. The semantic segmentation method based on a modified U-shaped network was used to construct an automated system for metal surface defect detection, which has been described in [24].

In the realm of neural networks, the attention mechanism [19] is a useful module which can be incorporated into a variety of applications, including target detection, natural language processing, and semantic segmentation. The ultimate purpose of the algorithm is to extract the characteristics that are unique to the targets themselves. Because of the attention mechanism, the algorithm is able to ignore any irrelevant regions and concentrate only on the relevant regions that have been specified. Such a mechanism is explained in [31].

4. Proposed Methodology

In this paper, a single stage automated defect detection algorithm known as YOLOv-5s-FRN is proposed, and its overall functioning has been represented using a block diagram, as shown in Figure 3. Input images from a huge dataset will be pre-processed and passed through an augmentation operation before being divided for training and testing purposes. Each image from the dataset will be passed through discriminability boosted feature extraction processes at the time of model training. The efficient channel-oriented feature recalibration in the proposed system enables it to recognize and classify the images with greater accuracy. Model testing will be done using test images for visualization of test results and performance evaluation of the trained model.

The detailed architecture of the proposed system is shown in Figure 4. A newly featured backbone, neck, and head are the three primary structural elements of the proposed system. The A newly featured backbone is the one that is tasked with the responsibility of extracting informative details from the input image. Following the completion of the encoding and decoding processes, the neck structure will locate the pixels. The head module will identify and correctly interpret any defects or damages that are present on the surfaces.

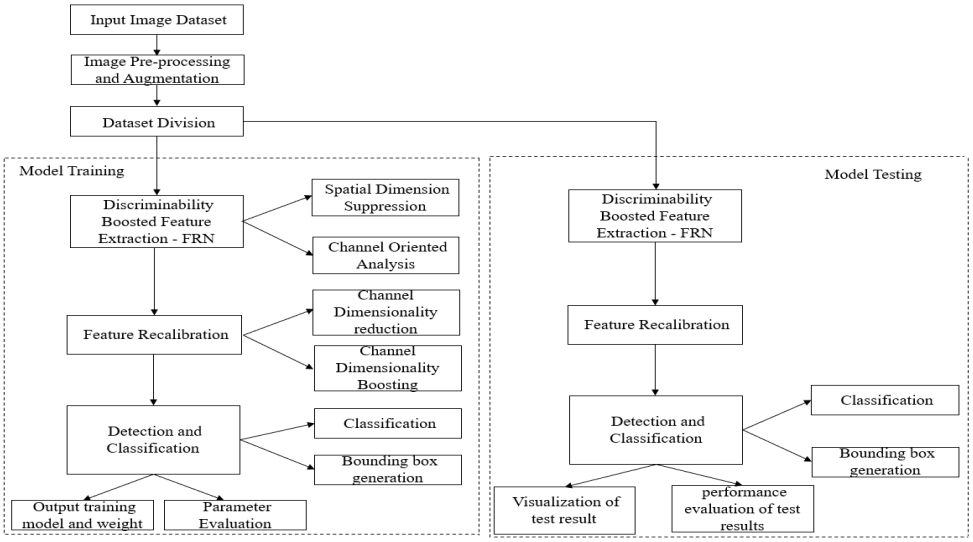


Figure 3. Functional block diagram of YOLOv-5s-FRN.

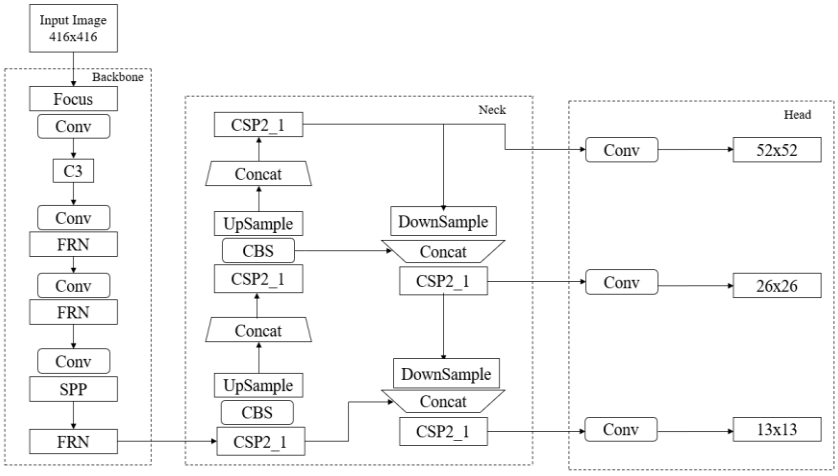


Figure 4. YOLOv-5s-FRN architecture.

The proposed system has been developed by modifying the backbone structure of basic YOLOv-5s model. In order to accomplish this, a novel architectural module named as FRN has been designed and integrated to the backbone module by replacing bottleneck and CSP1_X blocks. Embedding FRN structure into the YOLO network helps to achieve competitive accuracy through adaptive recalibration of channel-wise features. This innovative structure models the interdependencies of channels and convolutional characteristics in order to carry out channel-oriented analysis. It enables FRN to learn global information as well as capture hierarchical patterns with receptive fields. As a result, FRN is able to selectively highlight the features that are most relevant while simultaneously suppressing the features that are undesirable and less useful.

The architectural changes that have been done in the network and defining new structure of backbone with the incorporation of FRN structure is shown in Figure 5. The input image will go inside the backbone through focus structure. Following with the focus module in the architecture, there will be a convolution module named as C3. It is made up of three layers that are convolutional in nature. In between the various sets of convolutional layers is where the FRN has been inserted. After each iteration of the convolution operation, it will refine more useful characteristics.

```
# YOLOv5 FRN backbone
backbone:
  # [from, number, module, args]
  [ [ -1, 1, Focus, [ 64, 3 ] ], # 0-P1/2
    [ -1, 1, Conv, [ 128, 3, 2 ] ], # 1-P2/4
    [ -1, 3, C3, [ 128 ] ],
    [ -1, 1, Conv, [ 256, 3, 2 ] ], # 3-P3/8
    [ -1, 9, C3_FRN, [ 256, True ] ],
    [ -1, 1, Conv, [ 512, 3, 2 ] ], # 5-P4/16
    [ -1, 9, C3_FRN, [ 512, True ] ],
    [ -1, 1, Conv, [ 1024, 3, 2 ] ], # 7-P5/32
    [ -1, 1, SPP, [ 1024, [ 5, 9, 13 ] ] ],
    [ -1, 3, C3_FRN, [ 1024, False ] ], # 9
  ]
```

Figure 5. Defining network structure of backbone of YOLOv-5s-FRN.

In addition to FRN structure, the backbone mainly includes focus and Spatial Pyramid Pooling (SPP) structure. The SPP structure [6] is responsible for performing the operation of converting feature maps of arbitrary sizes into particularly defined sizes. Figure 6 shows the structure of focus module. The high-resolution feature map will be divided and stitched into several smaller resolution feature maps by the focus structure. This process will reduce the size of the input image and contributes to speeding up and deepening the feature extraction process. To facilitate this there are four slice blocks provided. Therefore, Both the width and height of the output image have been reduced to half of what they were in the input image, but the number of channels has been increased to four times what it was in the input image. The most significant benefit is that there is no loss of information. Then it performs CBS (Convolution, Batch normalization and Swish activation) operation after concatenating the output from slice blocks.

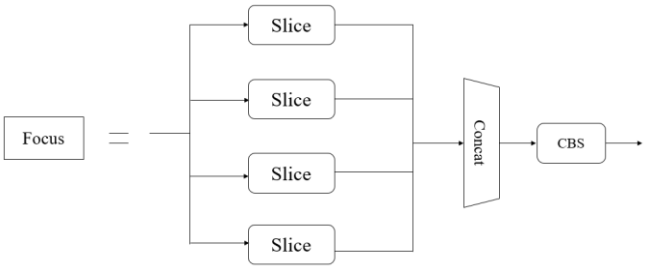


Figure 6. Structure of the focus module.

The neck structure includes CSP2_1 network, up sample and down sample blocks for the purpose of encoding, decoding and concatenation. The structure of CSP2_1 is shown in Figure 7, where CBS uses convolution, batch normalization and SiLu activation function for feature extraction. Neck structure transmits semantic information and location information to enable the network to combine feature data. The main purpose of using the CSP network is to reduce the computational cost by 10% or 20%. To do this, the base layer's feature map will be split into two sections, and they will be combined using a cross-stage hierarchy. This technique lowers the total number of parameters, which in turn speeds up the inference process and reduces the cost of computing.

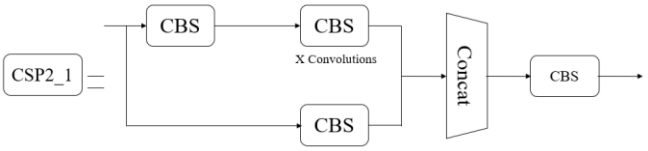


Figure 7. The structure of the CSP2_1 in neck module.

The head structure will be responsible for adjusting the number of channels and prediction. Three convolutional layers in the head structure will forecast the position of the bounding box, the prediction score, and the class of the item. The computation of the coordinates of bounding box by the head structure is based on the Equations (2), (3), (4), and (5).

$$b_x = ((2.\sigma(t_x) - 0.5) + C_x)) \quad (2)$$

$$b_x = ((2.\sigma(t_y) - 0.5) + C_y)) \quad (3)$$

$$b_w = P_w.(2.\sigma(t_w))^2 \quad (4)$$

$$b_h = P_h.(2.\sigma(t_h))^2 \quad (5)$$

where (b_x,b_y) denotes the centroid and (b_w,b_h) reflects the size of the bounding box that is predicted by the created model, respectively. The centroid of the bounding box of the feature map is given by the

coordinates (C_x,C_y), and the size of the bounding box is shown by the coordinates (P_w,P_h). The values (t_x,t_y) and (t_w,t_h) denote the center offset of the bounding box with respect to the network prediction and the appropriate scaling size respectively.

4.1. Feature Recalibration Network (FRN)

The basic building block of novel architectural module named as FRN is shown in Figure 8. The ultimate aim of developing such a structure is to improve the feature representational power of the network. Therefore, FRN will use feature map as input for performing channel-specific analysis. This helps to emphasis the informative features and to suppress the irrelevant ones. In order for the FRN to accomplish this, it will concentrate only on the channel-wise information rather than the spatial information. The main operational modules in FRN structure are transformation operation module, feature map generator, Spatial Dimension (SD) suppressor, Channel-Wise Attention (CWA) module and a rescaling module.

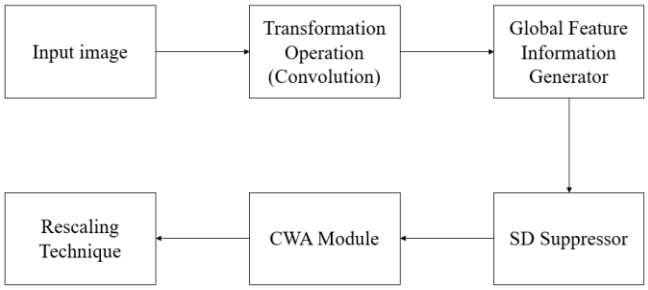


Figure 8. Block diagram of FRN.

The input image I has taken into consideration, which is being fed into the network. The input image has a dimension of HxWxC, where H, W and C are representing the height, width and channel respectively. Any kind of transformation applied to the input image I will generate output O. In this case, the transformation procedure is carried out by performing the convolution operation or set of convolutions. As O is the output of convolution operation, it can be considered as feature map with dimension of H'xW'xC'. Then the feature map O will traverse through the SD suppressor module. This module will aggregate the feature map across the spatial dimensions HxW, which leads to the creation of a channel descriptor. The distribution of the channel-wise feature responses will be included in the channel descriptors. The spatial dimension suppressor block will provide an output with a dimension of 1x1xC. Then, the CWA module will take the responsibility to produce channel excitations based on the channel dependencies. In order to make this process easier, CWA module includes two fully connected layers within it. The output from CWA module will then be directed to rescaling module. In rescaling module, it will first be rescaled to the original input dimension, and then it will be sent to the successive layers in the architecture.

The mathematical model of the FRN module is shown in Figure 9. Consider the convolution operation on input image I for generating the feature map O. Let K= [k_1, k_2, k_3,……k_c] be the set of kernels, where k_c refers to the parameter of c^{th} filter. Applying convolution on I using the kernel K will generate the output O=[O_1,O_2,O_3……O_C]. O can be expressed as shown in Equation (6).

$$Oc = Kc * I \sum_{S=1}^{C'} Kc^s * I^S \quad (6)$$

where * denotes the convolution operation, k_c^s= [k_c^1, k_c^2, k_c^3…… $k_c^{c'}$] and I^s = [I^1, I^2, I^3……$I^{c'}$]. In this k_c^S is a two-dimensional kernel and it will act on corresponding channel of *I*. The feature map can be obtained by doing a summing over each channel in this case.

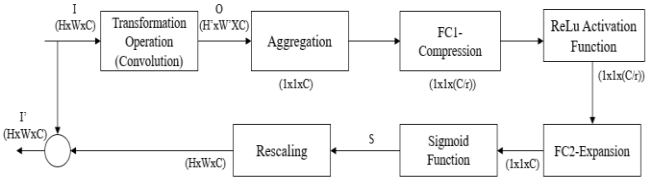


Figure 9. The mathematical model of the FRN.

The feature map will then be passed through SD suppression module for generating channel-wise statistics Y_C. To do this, each channel of the output feature O has taken into account and its each unit will be able to exploit the contextual information in and outside the receptive fields. This makes the network to perform better even in lower layers, where the size of receptive field is small. This is accomplished by the utilization of global average pooling, often known as a straightforward aggregation operation approach. Therefore, Y_C will be generated with a dimension of 1x1xC, which is on the basis of spatial dimensions of O, specifically H and W. i.e., Y_C has been generated by shrinking *O* through its spatial dimensions *HxW*. Equation (7) shows the calculation of the c^{th} element of *Y*.

$$Yc = F_{SD}(Oc) = \frac{1}{HxW} \sum_{i=1}^{H} \sum_{j=1}^{W} Oc(i,j) \quad (7)$$

where O_C is the feature map generated after convolution, $F_{SD}(O_C)$ is the global average pooling operation and *H* and *W* are the height and width of feature map. The statistics parameter obtained from SD suppressor module will then be passed through a CWA module in order to obtain the information related to channel-wise dependencies. It is possible for the CWA module to learn a nonlinear interaction between channels. In order to acquire the output of the CWA module, a straightforward gating mechanism with a sigmoid function has been implemented, as demonstrated in Equation (8).

$$S = F_{CWA}(Y, w) = \sigma(g(Y, w)) = \sigma(w_1 \delta(w_2, Y)) \quad (8)$$

where δ refers to Rectified Linear Unit (ReLu)

activation function, σ refers to the sigmoid activation function and w_1 and w_2 are the weights provided. As illustrated in Figure 10, the CWA module has been implemented with two Fully Connected (FC) networks in order to keep the complexity of the model to a minimum. The dimension of w_1 is C/r of C (i.e., ((C/r)xC), where r is the compression ratio. In first fully connected layer the channel dimension or the channel length is reduced to C/r. The dimension of w_2 is same as that of C (i.e., Cx(C/r)). Therefore, in second fully connected layer, the dimension of the channel is brought back to its original value, i.e., C itself. Therefore, the first fully connected layer can be considered as dimensionality-reduction layer and the second fully connected layer can be considered as a dimensionality increasing layer. Finally, the output obtained from CWA module will be fed into the rescaling module so that the initial dimension can be preserved. Equation (9) indicates the operation in rescaling module.

$$I' = F_{rescale}(Oc \cdot Sc) = Sc.Oc \quad (9)$$

where $I' = [I'_1, I'_2, I'_3 \ldots\ldots I'c]$ and $F_{rescale}(O_c.S_c)$ is the channel-wise multiplication of feature map Oc and the scalar *Sc*.

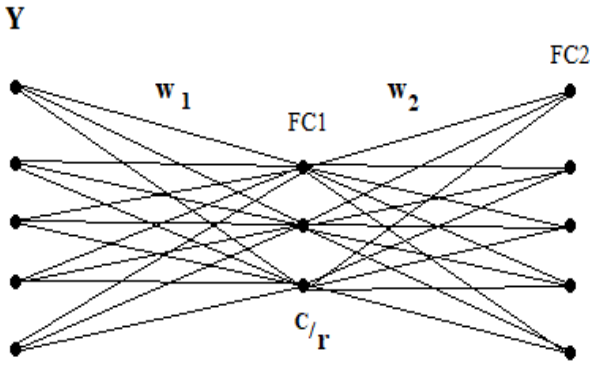


Figure 10. Fully Connected layers in CWA module.

5. Experimental Results

5.1. Implementation

The proposed YOLOv-5s-FRN architecture is developed and run on a computer with CPUi7-10900K@3.7GHz, GPU NVIDIA RTx3090 and 64GB running memory. The newly designed architecture has been trained and tested using a publicly available dataset called North Eastern University (NEU- DET) metal surface defect data set [17]. It consists of six various forms of surface defects called as rolled in scale, scratches, pitted surface, inclusion, patches and crazing. Rolled in scale defects are occurring during the rolling process of metal due to flaky mixture of iron oxides. Scratch is nothing but a mark over the surfaces, which happened due to sharp end of any object. Corrosion over the metallic objects causes pitted surfaces. Sometimes, the corrosion may occur on a specific portion of metal surface and it leads to the formation of cavity.

Inclusions are nothing but the occurrence of the flaws in shape of dots, blocks or lines. Patches are the irregular shapes appear on the metal surfaces. They may occur either in black or white color. It provides negative impact to the appearance of the metal surfaces. Crazing is a phenomenon which causes cracks over the metal surfaces. It occurs either due to the poor quality of the raw materials or high temperature in production environment. Their appearance will be like black spots with small or large areas.

The dataset consists of 1800 grey scale images and each category of defect is 300 in count. The pixel resolution of each image is 200x200. NEU-DET dataset is provided with the respective annotations of each image. The training and the test data are divided in 9:1 ratio. The training data set is again divided as training and validation data set in 9:1 ratio. The training process of the NEU-DET dataset has been lasted for around 8 hours to complete 1000 epochs. Training dataset includes 1440 images out of 1800 images in dataset.

	from	n	params	module	arguments
0	-1	1	3520	models.common.Focus	[3, 32, 3]
1	-1	1	18560	models.common.Conv	[32, 64, 3, 2]
2	-1	1	15808	models.experimental.C3	[64, 64, 1]
3	-1	1	73984	models.common.Conv	[64, 128, 3, 2]
4	-1	1	173761	models.common.C3_GC	[128, 128, 3, True]
5	-1	1	295424	models.common.Conv	[128, 256, 3, 2]
6	-1	1	691585	models.common.C3_GC	[256, 256, 3, True]
7	-1	1	1180672	models.common.Conv	[256, 512, 3, 2]
8	-1	1	656896	models.common.SPP	[512, 512, [5, 9, 13]]
9	-1	1	1446657	models.common.C3_GC	[512, 512, 1, False]
10	-1	1	131584	models.common.Conv	[512, 256, 1, 1]
11	-1	1	0	torch.nn.modules.upsampling.Upsample	[None, 2, 'nearest']
12	[-1, 6]	1	0	models.common.Concat	[1]
13	-1	1	313088	models.experimental.C3	[512, 256, 1, False]
14	-1	1	33024	models.common.Conv	[256, 128, 1, 1]
15	-1	1	0	torch.nn.modules.upsampling.Upsample	[None, 2, 'nearest']
16	[-1, 4]	1	0	models.common.Concat	[1]
17	-1	1	78720	models.experimental.C3	[256, 128, 1, False]
18	-1	1	147712	models.common.Conv	[128, 128, 3, 2]
19	[-1, 14]	1	0	models.common.Concat	[1]
20	-1	1	247552	models.experimental.C3	[256, 256, 1, False]
21	-1	1	590336	models.common.Conv	[256, 256, 3, 2]
22	[-1, 10]	1	0	models.common.Concat	[1]
23	-1	1	986624	models.experimental.C3	[512, 512, 1, False]
24	[17, 20, 23]	1	24273	models.yolo.Detect	[4, [[10, 13, 16, 30, 33, 23], [30, 61, 62, 45, 59, 119],

Model Summary: 212 layers, 7.10978e+06 parameters, 7.10978e+06 gradients

Figure 11. Number of layers and parameters obtained after training YOLOv-5s-FRN structure.

In order to avoid overfitting, image position augmentation method has been employed. It will flip, rotate and scale every image. Flipping make the images to flip into left, right and upside down. To generate the images in different angles and variety of orientations, rotation augmentation has been applied. Scaling operation will increase and decrease the size of the picture. After defining various parameters for developing the model, the network structure of the proposed system has been generated as shown in Figure 11. It can be seen from the structure of the network that the number of parameters created by the YOLOv-5s-FRN network is significantly more than the number generated by a typical YOLOv-5 network. This is because of the integration of FRN into the system. Figure 12 shows the structure definition of FRN module.

```
class C3_FRN (nn.Module):
    # C3 module with ContextBlock2d()
    def __init__(self, c1, c2, n=1, shortcut=True, g=1, e=0.5):  # ch_in, ch_out, number, shortcut, groups, expansion
        super(C3_GC, self).__init__()
        # super().__init__()
        c_ = int(c2 * e)  # hidden channels
        self.FRN = ContextBlock2d(c1, c_)
        self.cv1 = Conv(c1, c_, 1, 1)
        self.cv2 = Conv(c1, c_, 1, 1)
        self.cv3 = Conv(2 * c_, c2, 1)  # act=FReLU(c2)
        self.m = nn.Sequential(*[Bottleneck(c_, c_, shortcut, g, e=1.0) for _ in range(n)])

```

Figure 12. Structure definition of FRN module.

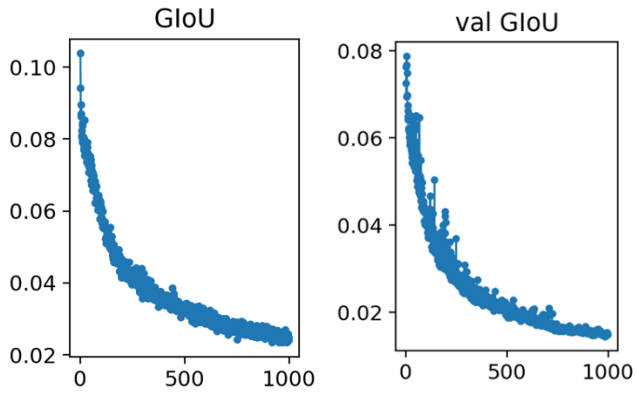
Regardless of the size of the image that is loaded into the YOLO, it will always convert the image to the same standard size. The YOLOv-5s-FRN has undergone training for a total of 1000 epochs with a batch size of 16. YOLOv-5s uses Sigmoid Linear Unit (SiLu) and sigmoid functions as activation function [5]. In addition to SiLu, FRN structure is used ReLu activation function [15]. While the sigmoid function is utilized for the convolution operation in the output layer, the SiLu function is employed for the convolution operation in the hidden layers. Pytorch was utilised in the development of the environmental setup for YOLOv-5s-FRN, so that a greater degree of flexibility could be achieved.

It is also possible to implement the defect detection system using X-Ray captured images. Such a system is required during the quality inspection of the products within concealed packages. Therefore, the primary stage of the defect detection system will be equipped with an image sensing devise using X-Ray source. X- rays will penetrate through the packaging material to capture the image of the product surface. X-Ray captured images will be saved as Digital Imaging and Communication in Medicine (DICOM) file. During the pre-processing stage, the DICOM images will be transformed into JPEG format and then the images will be rescaled into a predetermined size using code called cv2. The training module of the proposed YOLOv-5s-FRN will be modified in accordance with the new dataset of X-Ray images. After training process, the new network structure will be defined and it will generate suitable number of trainable parameters and network layers.

5.1. Performance Evaluation

The performance of the proposed YOLOv-5s-FRN has been analyzed and evaluated based on various matrices such as losses, precision, recall and mean Average Precision (mAP). Graphs of different losses, recall, mAP and precision for the proposed model have been obtained after defining and training the proposed system. The value of mean average precision is calculated based on different IoU thresholds. The ratio of overlap between the ground truth and the predicted value provides the Intersession-over-Union (IoU). Therefore, IoU plays a significant part in process of ensuring the accurate detection results. When the value of the IoU grows to a point where it is greater than 0.5, the model will then begin to assign positive labels. If the IoU value drops below 0.5, negative labels will be assigned to the observations.

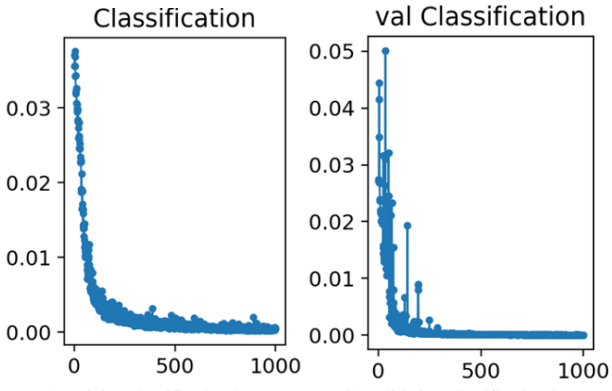
Different types of losses considered in this paper are Generalized Intersession Over Union (GIoU), objectness and the classification loss. Binary Cross Entropy (BCE) is used to compute the classification loss and objectness. The training and validation GIoU loss of the proposed system is shown in Figure 13. The GIoU is a loss matric which indicates how close the predicted bounding box to the ground truth. Ground truth is nothing but the labels on the image. The GIoU value is obtained to be a small value in the proposed system, i.e., less than 0.01. As it can be seen from the graphical representations, the error values are produced in the proposed system is less in value than that in the basic model as the epochs approach larger values. The lower values of the GIoU indicates that the model is good enough to generate a proper bounding box around the dataset.



a) GIoU loss during training. b) GIoU loss during validation.

Figure 13. GIoU loss obtained for YOLOv-5s-FRN model.

Another type of loss metric that is taken into consideration is known as classification loss. This loss metric evaluates how accurately all predicted bounding boxes are classified. Figure 14 is indicating the training and validation classification loss of YOLOv-5s-FRN structure. The classification loss of the suggested model decreases as the number of epochs rises, and its value keeps decreasing until final few epoch values.



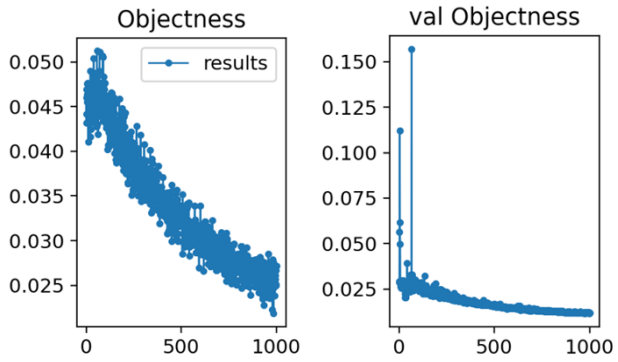
a) Training classification loss. b) Validation classification loss.

Figure 14. Classification loss obtained for YOLOv-5s-FRN model.

Objectness is nothing but the ability of the model to recognize the presence of an object in the analyzed image. Figure 15 indicates the graphical representation of objectness loss of YOLOv-5s with the FRN structure. According to the objectness graphs, it is obvious that the proposed model suffers notable losses in the earlier stages of the epochs. But the loss is getting reduced when the epoch number is tending towards 1000. Considering these three loss matrices, the final loss value can be calculated as shown in Equation (10).

$$Loss = \lambda_1 L_{cls} + \lambda_2 L_{obj} + \lambda_3 L_{GIoU} \quad (10)$$

where, λ_1, λ_2 and λ_3 are the constants and L_{cls}, L_{obj}, and L_{GIoU} are the classification loss, objectness loss and GIoU loss respectively.



a) Objectness loss during training. b) Objectness loss during validation.

Figure 15. Objectness loss obtained for YOLOv-5s- FRN model.

Precision and recall are the other evaluation indicators considered to check the performance of the proposed system. Precision is the measure which says about how precise the developed model to give the predictions. The precision value can be calculated by dividing the number of actual successes by the total number of guesses. Recall can be considered as the ability of model to identify the positive samples. Equations (11) and (12) indicate the calculation of precision.

$$Precision = \frac{TP}{TP+FP} \quad (11)$$

$$Recall = \frac{TP}{TP + FN} \quad (12)$$

where *TP*, *FP* and *FN* are true positive, false positive and false negatively respectively.

Precision obtained for YOLOv-5s-FRN structure is shown in Figure 16. When compared to the other already-existing base models, the value of accuracy converges more quickly using the proposed YOLOv-5s-FRN based defect detection system.

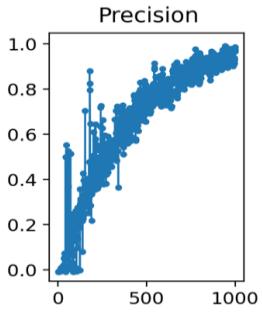


Figure 16. Precision obtained for YOLOv-5s-FRN structure.

Graphical representation of recall values obtained for YOLOv-5s-FRN structure is shown in Figure 17. Higher values of recall indicate that more samples are detected as positive. It can be computed as the proportion of number of positive samples classified correctly as positive to the total number of positive samples as shown in Equation (12). From the graphical representation in Figure 17, it is clear that the recall value obtained for the YOLOv-5s-FRN structure is 99.86%, which is a high value. The P-R curve obtained for YOLOv-5s-FRN architecture is shown in Figure 18.

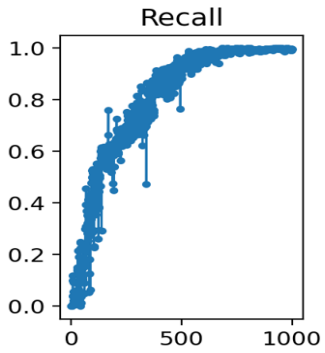


Figure 17. Recall values obtained for YOLO-v5-FRN structure.

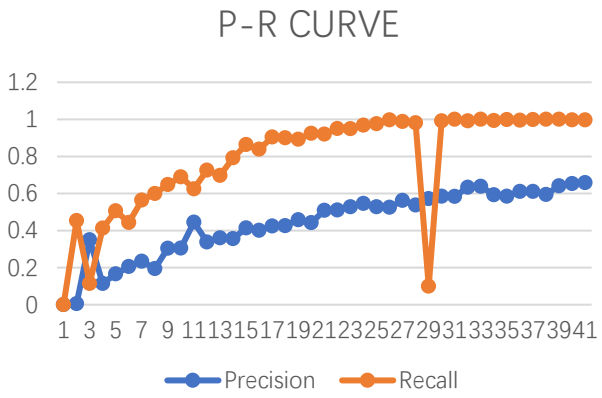


Figure 18. P-R curve obtained for the proposed system.

The determination of an average precision value for each class of defects based on the prediction of a model is termed as mean average precision. It is related to the area under Precision and Recall (P-R) curve. The value for the mean average precision will be determined by taking the mean of all averages of precision values for the distinct classes. Normally, mAP value will be calculated at an IoU value of 0.5 or 50% (mAP_0.5) and

at an IoU interval of 0.50 to 0.95 or 50% to 95% (mAP_0.5:0.95). The prediction of defect detection system with mean average precision value which is considered at a threshold value of 0.5 or greater that 0.5 will be considered as defect.

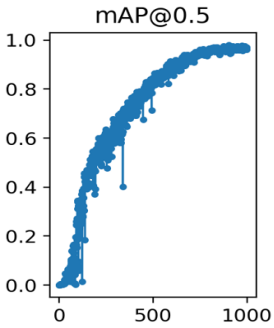


Figure 19. mAP_0.5 of YOLOv-5s-FRN structure.

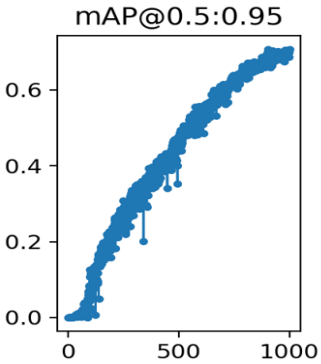


Figure 20. mAP_0.5:0.95 of YOLOv-5s-FRN structure.

The mAP_0.5 and mAP_0.5:0.95 obtained for YOLOv-5s-FRN structure is shown in Figure 19 and 20 respectively. The maximum value of mAP_0.5 and mAP_0.5:0.95 of proposed YOLOv-5s-FRN are obtained to be 98.05% and 89.03% respectively. It has been observed that the convergence of mAP occurs quicker in the proposed system. i.e., the value of mAP started to increase earlier than the other state-of-the-art approaches. The introduction of FRN structure makes the object detection system to obtain maximum value of mAP in minimal possible number of iterations, which contribute to enhance the performance of the proposed system.

Following with the execution completion of prediction module, the trained model will identify defects and form bounding box around each of them. Prediction module will generate a dictionary for representing the information related to defects and bounding boxes for each image. Each dictionary indicates class of the defect, coordinates of bounding box and confidence value of each defect. Confidence value indicates the percentage of probability that the identified defect was correctly labelled. Figure 21 indicates detection and classification of different types of defects done by the proposed system. Figure 22 indicates the dictionaries generated by predictor module for certain samples of images from the dataset during classification.

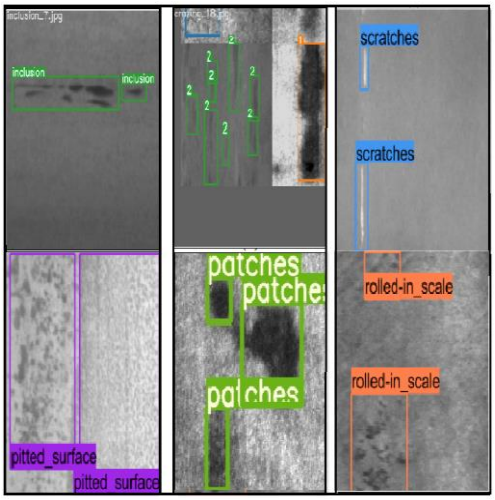


Figure 21. Detection and classification of different types of defects done by YOLOv-5s-FRN defect detection system.

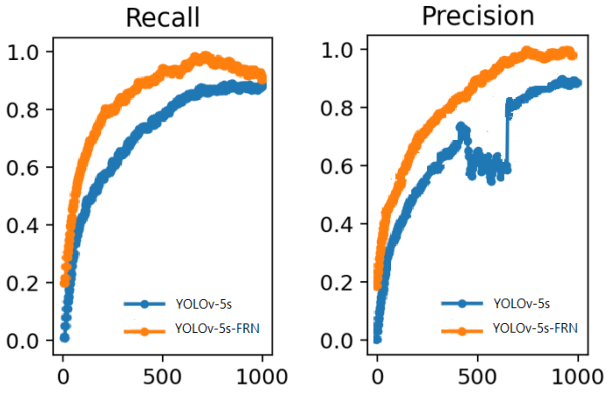
```
!python predictor.py

/usr/local/lib/python3.8/dist-packages/mmcv/__init__.py:20: UserWarning: On January 1
  warnings.warn(
device: cuda:0
Fusing layers...
3 object detected
patches_32.jpg {'patches': [{'box': [31, 83, 105, 148], 'confidence': 0.900756597518
device: cuda:0
Fusing layers...
3 object detected
inclusion_35.jpg {'inclusion': [{'box': [118, 2, 156, 93], 'confidence': 0.798150062
device: cuda:0
Fusing layers...
1 object detected
crazing_37.jpg {'crazing': [{'box': [73, 50, 200, 143], 'confidence': 0.745698630809
```

Figure 22. Bounding box axis values and corresponding confidence values obtained by predictor of the proposed system.

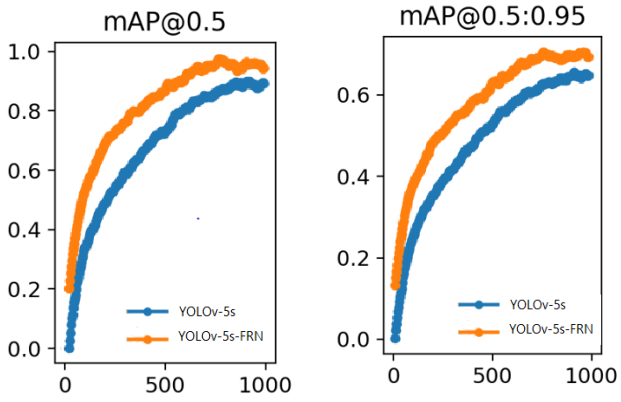
5.2. Comparison of Various Deep Methods

The proposed method has been compared with basic YOLOv-5 model in order to evaluate the effectiveness of new architectural module called as FRN. Graphical representations obtained on the basis of recall and precision has been shown in Figure 23-a) and (b). Comparison graphs of basic YOLOv-5s and YOLOv-5s-FRN on the basis of mean average precision has been shown in Figure 24-a) and (b). From the graphical representations it is clear that the performance of YOLOv-5s-FRN is superior to that of basic YOLOv-5s model.



a) Comparison based on recall. b) Comparison based on precision.

Figure 23. Comparison of YOLOv-5s-FRN model with basic YOLOv-5s model.



a) Comparison based on mAP_0.5. b) Comparison based on mAP_0.5::0.95.

Figure 24. Comparison of YOLOv-5s-FRN model with basic YOLOv-5s model based on mean average precision.

In order to ensure the effectiveness of the proposed method, it has been compared with the other existing deep learning based algorithms. Table 1 shows the comparison between different deep methods for defect detection based on different parameters.

Table 1. Comparison of different deep methods.

Method	Accuracy	Recall	mAP@0.5	mAP@0.5:0.95	Computational time
SSD	0.971	0.956	0.957	0.6981	29ms 7s
FASTER-RCNN	0.977	0.963	0.899	0.7102	37ms 7s
YOLOV-2	0.98	0.966	0.819	0.7712	24.91ms 4.03s
YOLOV-3	0.984	0.974	0.855	0.7856	21.75ms 8.46s
YOLOV-5	0.9912	0.9901	0.9683	0.8139	22.18ms 4s
Proposed method	0.9989	0.9986	0.9805	0.8903	21ms 9s

The proposed method has been compared with Single Shot Defect Detection (SSD), Faster Region-Convolutional Neural Network (Faster R-CNN), YOLOV-2, YOLOV-3, and basic YOLOV-5 networks. SSD is a detection algorithm, which adopts a feature extraction structure with the aim of improving accuracy value. Faster R-CNN is a defect detection system which will identify the region of interest and then executes detection operation. YOLOV-2, YOLOV-3, and YOLOV-5 are the subsequent improved versions of YOLO, which is capable to do the defect detection in single stage.

From Table 1, it is observed that the proposed model has an accuracy value of 0.9989 and a recall value of 0.9986, which are highest values when comparing with the other deep methods taken into consideration. On the other hand, it is obtained that the proposed method is capable of producing a mean-average precision values (98.05% of mAP@0.5 and of mAP@0.5::0.95) that are superior to those delivered by existing approaches. Computational time of the proposed system is obtained to be 21ms and 9s, while the computational time of the YOLOv-3 is nearly same, i.e., 21.75ms and 8.46s. However, the design has not seen any computational delays even with the inclusion of the new network known as FRN. Graphical representation of comparative analysis of proposed system has been shown in Figure 25. Therefore, it is obtained that the incorporation of FRN structure offers higher performance in terms of accuracy, recall, precision, mean average precision, and computing time. Comparing with the other techniques, the cost effectiveness of the proposed method is obtained to be moderate as it incorporates a new module into network structure.

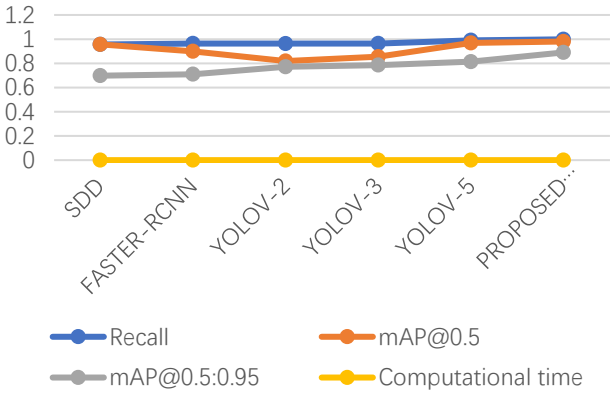


Figure 25. Comparative analysis of various deep methods.

6. Conclusions

An automated defect detection system named as YOLOv-5s-FRN has been proposed as a method that can detect and classify the metal surface defects with high levels of accuracy and efficiency. The proposed system has been developed with the integration of a new architectural module called as FRN. FRN module performs channel-oriented analysis and understanding of each image, which helps the network to filter the most important features and neglect other ones. The ability of FRN structure to extract relevant deep features help to boost the feature discrimination capability of the proposed system. From the experimentations executed using NEU-DET dataset for 1000 epochs, it is observed that the convergence of precision, recall and mAP values are occurring more quickly in the proposed system. Therefore, YOLOv-5s-FRN provide more accurate interpretations and deliver performance matrices that are superior to those of the basic YOLOv-5s network and other deep methods. The maximum value of mean average precision has been achieved with minimum number of iterations, which leads to the performance enhancement of the proposed system by means of speed. For the purpose of performance evaluation, the proposed technique has been compared with the other defect detection methods, which are already in use, on the basis of various indicators.

The future research can focus on the detection of defects on the metal surfaces in concealed packages in industries. For doing this, the network can be trained using X-Ray captured images. Additionally, the development of a production software can be done to

make such an automated defect detection system, which is accessible to a variety of applications.

References

[1] Bayraktar E., Tosun B., Altintas B., and Celebi N., "Combined GANs and Classical Methods for Surface Defect Detection," *in Proceedings of the Signal Processing and Communications Applications Conference*, Safranbolu, pp. 1-4, 2022. Doi: 10.1109/SIU55565.2022.9864705.

[2] Fu J., Zhu X., and Li Y., "Recognition of Surface Defects on Steel Sheet Using Transfer Learning," *arXiv Preprint*, *arXiv:1909.03258v2*, 2019. https://doi.org/10.48550/arXiv.1909.03258

[3] Gyimah N., Girma A., Mahmoud M., Nateghi S., Homaifar A., and Opoku D., "Robust Completed Local Binary Pattern (RCLBP) for Surface Defect Detection," *in Proceedings of the IEEE International Conference on Systems*, *Man, and Cybernetics*, Melbourne, pp. 1927-1934, 2021. DOI: 10.1109/SMC52423.2021.9659140

[4] Hatab M., Malekmohamadi H., and Amira A., "Surface defect detection using YOLO network," *in Proceedings of the SAI Intelligent Systems Conference*, London, pp. 505-515, 2020. https://doi.org/10.1007/978-3-030-55180-3_37

[5] He K., Zhang X., Ren S., and Sun J., "Deep Residual Learning for Image Recognition," *in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, USA, pp. 770-778, 2016. https://doi.org/10.48550/arXiv.1512.03385

[6] He K., Zhang X., Ren S., and Sun J., "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, pp. 1904-1916. 2015. Doi:https://doi.org/10.48550/arXiv.1406.4729

[7] He Y., Song K., Meng Q., and Yan Y., "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. Doi: 10.1109/TIM.2019.2915404

[8] Kim M., Park T., and Park P., "Classification of Steel Surface Defect Using Convolutional Neural Network with Few Images," *in Proceedings of the 12th Asian Control Conference*, Kitakyushu, pp. 1398–1401, 2019.

[9] Komijani A., Vafaeinezhad F., Khoramdel J., Borhani Y., and Najafi E., "Multi-Label Classification of Steel Surface Defects Using Transfer Learning and Vision Transformer," *in Proceedings of the International Conference on Information and Knowledge Technology*, Karaj, pp. 1-5, 2022. doi: 10.1109/IKT57960.2022.10039038

[10] Kou X., Liu S., Cheng K., and Qian Y., "Development of a YOLO-V3-based Model for Detecting Defects on Steel Strip Surface," *Measurement*, vol. 182, pp. 109454, 2021. https://doi.org/10.1016/j.measurement.2021.109454.

[11] Liang F., Zhou Y., Chen X., Liu F., and Zhang C., "Review of Target Detection Technology Based on Deep Learning," *in Proceedings of the 5th International Conference on Control Engineering and Artificial Intelligence*, Sanya, pp. 132-135, 2021. https://doi.org/10.1145/3448218.3448234.

[12] Lin H. and Wibowo F., "Image Data Assessment Approach for Deep Learning-Based Metal Surface Defect-Detection Systems," *IEEE Access*, vol. 9, pp. 47621-47638, 2021. doi: 10.1109/ACCESS.2021.3068256.

[13] Lin Z., Ye H., Zhan B., and Huang X., "An Efficient Network for Surface Defect Detection," *journal of Applied Sciences*, vol. 10, no. 17, pp. 6085, 2020. https://doi.org/10.3390/app10176085

[14] Lv X., Duan F., Jiang J., Fu X., and Gan L., "Deep Metallic Surface Defect Detection: The New Benchmark and Detection Network," *Sensors Journal*, vol. 20, no. 6, pp. 1562, 2020. https://doi.org/10.3390/s20061562

[15] Nair V. and Hinton G., "Rectified Linear Units Improve Restricted Boltzmann Machines," *in Proceedings of the 27th International Conference on International Conference on Machine Learning*,
Haifa, pp. 807-814, 2010. https://dl.acm.org/doi/10.5555/3104322.3104425

[16] Narayanan L., Krishnan S., and Robinson H., "A Hybrid Deep Learning Based Assist System for Detection and Classification of Breast Cancer from Mammogram Images," *The International Arab Journal of Information Technology*, vol. 19, no. 6, pp. 965-974, 2022. https://doi.org/10.34028/iajit/19/6/15

[17] Neu Surface Defect Database, Available: http://faculty.neu.edu.cn/yunhyan/NEU_surface_defect_database.html, Last Visited, 2023.

[18] Ning Z. and Mi Z., "Research on Surface Defect Detection Algorithm of Strip Steel Based on Improved YOLOV3," *in Proceedings of the International Conference on Electronic Materials and Information Engineering*, Xain, pp. 1-6, 2021. 10.1088/1742-6596/1907/1/012015

[19] Niu Z., Zhong G., and Yu H., "A Review on the Attention Mechanism of Deep Learning," *Neurocomputing Journal*, vol. 452, pp. 48–62, 2021.
https://doi.org/10.1016/j.neucom.2021.03.091

[20] Prihatno A., Utama I., Kim J., and Jang Y., "Metal Defect Classification Using Deep Learning," *in Proceedings of the International Conference on Ubiquitous and Future Networks*, Jeju Island, pp.

389-393, 2021. Doi: 10.1109/ICUFN49451.2021.9528702

[21] Qing Y., Fing I., Tang I., Xu W., and Zho X., "Development of an Automatic Monitoring System for Rice Light-Trap Pests Based on Machine Vision," *Journal of Integrative Agriculture*, vol. 19, no. 10, pp. 2500-2513, 2020. https://doi.org/10.1016/S2095-3119(20)63168-9

[22] Tabernik D., Sela S., Skvarc J., and Skocaj D., "Segmentation-Based Deep-Learning Approach for Surface-Defect Detection," *Journal of Intelligent Manufacturing*, vol. 31, pp. 759-776, 2019. https://doi.org/10.1007/s10845-019-01476-x

[23] Truong Duy T. and Natori N., "Efficient Defect Detection from Consecutive Monocular Images by Deep Learning," *in Proceedings of the IEEE International Conference on Machine Learning and Applications*, Miami, pp. 473-478, 2020. Doi: 10.1109/ICMLA51294.2020.00080.

[24] Vengaloor P. and Muralidhar R., "Deep Learning Based Semantic Segmentation Technique for Anomaly Detection on Metal Surfaces using High Calibre U-Shaped Network," *International Information and Engineering Technology Association*, vol. 39, no. 6, pp. 2023-2031, 2022. https://doi.org/10.18280/ts.390614

[25] Vengaloor R. and Muralidhar R., "Smart Farming Using IOT: The Surveillance and the Regulation of Agricultural Area," *Bulletin of Environment, Pharmacology and Life Sciences,* vol. 10, no. 11, pp. 206-212, 2021.

[26] Wang C., Yu Y., Yu J., Zhang Y., and Zhao Y., "Microstructure Evolution and Corrosion Behavior of Dissimilar 304/430 Stainless Steel Welded Joints," *Journal of Manufacturing Process*, vol. 50, pp. 183-191, 2020. https://doi.org/10.1016/j.jmapro.2019.12.015.

[27] Wang C. and Xie H., "MeDERT: A Metal Surface Defect Detection Model," *IEEE Access*, vol. 11, pp. 35469-35478, 2023. Doi: 10.1109/ACCESS.2023.3262264.

[28] Wang C., Xu J., Liang X., and Yin D., "Metal Surface Defect Detection Based on Weighted Fusion," *in Proceedings of the International Conference on Virtual Reality and Visualization*, Recife, pp. 179-184, 2020. doi: 10.1109/ICVRV51359.2020.00044

[29] Wang C., Zhou Z., and Chen Z., "An Enhanced YOLOv4 Model with Self-Dependent Attentive Fusion and Component Randomized Mosaic Augmentation for Metal Surface Defect Detection," *IEEE Access*, vol. 10, pp. 97758-97766, 2022. Doi: 10.1109/ACCESS.2022.3203198.

[30] Wang H., Li M., Gao P., and Zhang S., "YOLOXD: A New Network for Metal Surface Defect Detection," *in Proceedings of the IEEE 2nd International Conference on Computer Systems*, Qingdao, pp. 75-79, 2022. Doi: 10.1109/ICCS56273.2022.9988327.

[31] Woo S., Park J., Lee J., and Kweon I., "CBAM: Convolutional Block Attention Module," *in Proceedings of the 15th European Conference on Computer Vision*, Munich, pp. 3-19, 2018. https://doi.org/10.1007/978-3-030-01234-2_1

[32] Ye Z., Liu M., Zhang S., and Wei P., "Dual-Path GAN: A Method for Enhancing Small-Scale Defect Detection on Metal Images," *in Proceedings of the 41st Chinese Control Conference*, Hefei, pp. 6292-6297, 2022. Doi: 10.23919/CCC55666.2022.9902599

[33] Yu R., Guo B., and Yang K., "Selective Prototype Network for Few-Shot Metal Surface Defect Segmentation," *in Proceedings of the IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1-10, 2022. Doi: 10.1109/TIM.2022.3196447.

[34] Zhao H., Wan F., Lei G., Xiong Y., and Xu L., "LSD-YOLOv5: A Steel Strip Surface Defect Detection Algorithm Based on Lightweight Network and Enhanced Feature Fusion Mode," *Sensors Journal*, vol. 23, pp. 6558, 2023. https://doi.org/10.3390/s23146558.

[35] Zhou X., Nie Y., Wang Y., Pingguo C., and Ye M., "A Real-time and High-Efficiency Surface Defect Detection Method for Metal Sheets Based on Compact CNN," *in Proceedings of the International Symposium on Computational Intelligence and Design*, Hangzhou, pp. 259-264, 2020. Doi: 10.1109/ISCID51228.2020.00064.

Reshma Vengaloor received her MTech Degree from TKM Institute of Science and Technology, Kerala, India. She is currently pursuing Ph.D. in SRM Institute of Science and Technology, Ramapuram, Chennai, India. Her research area of interest includes Digital Image processing and Deep Learning.

Roopa Muralidhar received her Ph.D. from Sathyabama Institute of Science and Technology, India. She is currently working as Associate Professor in SRM Institute of Science and Technology, India. She is Fellow member of IETE, New Delhi. Her research area of interest includes Wireless Communication, Mobile Ad Hoc Network, and Wireless Networks. She has published several papers in high impact journals.