

Article

Sppn-Rn101: Spatial Pyramid Pooling Network with Resnet101-Based Foreign Object Debris Detection in Airports

Abdulaziz Alshammari ^{1,*} and Rakan C. Chabaan ²¹ Information Systems Department, College of Computer Information and Sciences, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh 11432, Saudi Arabia² Hyundai American Technical Center, Inc., Superior Township, MI 48198, USA

* Correspondence: aashammari@imamu.edu.sa

Abstract: Over the past few years, aviation security has turned into a vital domain as foreign object debris (FOD) on the airport paved path possesses an enormous possible threat to airplanes at the time of takeoff and landing. Hence, FOD's precise identification remains significant for assuring airplane flight security. The material features of FOD remain the very critical criteria for comprehending the destruction rate endured by an airplane. Nevertheless, the most frequent identification systems miss an efficient methodology for automated material identification. This study proffers a new FOD technique centered on transfer learning and also a mainstream deep convolutional neural network. For object detection (OD), this embraces the spatial pyramid pooling network with ResNet101 (SPPN-RN101), which assists in concatenating the local features upon disparate scales within a similar convolution layer with fewer position errors while identifying little objects. Additionally, Softmax with Adam Optimizer in CNN enhances the training speed with greater identification accuracy. This study presents FOD's image dataset called FOD in Airports (FODA). In addition to the bounding boxes' principal annotations for OD, FODA gives labeled environmental scenarios. Consequently, every annotation instance has been additionally classified into three light-level classes (bright, dim, and dark) and two weather classes (dry and wet). The proffered SPPN-ResNet101 paradigm is correlated to the former methodologies, and the simulation outcomes exhibit that the proffered study executes an AP medium of 0.55 for the COCO metric, 0.97 AP for the pascal metric, and 0.83 MAP of pascal metric.

Keywords: foreign object debris (FOD); pre-processing; aircraft; classification; augmentation; annotation; resizing

MSC: 68T07

Citation: Alshammari, A.; Chabaan, R.C. Sppn-Rn101: Spatial Pyramid Pooling Network with Resnet101-Based Foreign Object Debris Detection in Airports. *Mathematics* **2023**, *11*, 841. <https://doi.org/10.3390/math11040841>

Academic Editors: Mohd Dilshad Ansari, Mohd Fauzi Bin Othman, Jawad Rasheed and Mazdak Zamani

Received: 21 December 2022

Revised: 27 January 2023

Accepted: 29 January 2023

Published: 7 February 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Foreign object debris (FOD) remains a chief issue in the aviation maintenance industry, which lessens the security level of the airplane. Fundamentally, FOD, which is called a foreign object, could lead to destruction in an airplane engine failure and human mortality [1]. An instance of such a disaster because of FOD remains the aircraft accident of French Flight 4590 in 2000 which took the life of 113 persons. The financial loss because of this crash has been approximated to be around three to four billion USD annually. FOD indicates that objects in the proximity of the airport, particularly on the runway, could destroy an aircraft [2]. FOD's instances include curved metallic bars, which led to the French Flight 4590 crash, parts separated out of airplanes or automobiles, concrete lumps out of runways, and plastic components [3].

Presently, for lessening FOD air mishaps, four runway scrutiny systems were established and modeled for APs such as Tarsier in the United Kingdom, FOD Finder in the United States of America, FODetect in Israel, and IFerret in Singapore [4]. Deep learning

(DL) technology is turned into a watchword since this remains capable of giving fine outcomes in image classification, object detection (OD), and natural language processing. It occurs because of the presence of huge datasets (DSs) and strong graphics processing units.

The deep learning methodology which possesses vital outcomes in image detection remains the convolutional neural network (CNN). This CNN methodology (CNNM) entered into several computer vision implementations such as IC, face authentication, semantic segmentation, OD, and image annotation. This CNN algorithm (CNNA) is exhibited to execute finer detection and identification since this possesses a finer resolution and strength for FOD detection (FODD) [5,6].

Studies on CNNM for predicting air mishaps because of FOD was performed experimentally, theoretically, and for the rest of the applicable intentions. The FOD materials' institution centered upon CNN was performed by Xu et al. In 2018 if this methodology were employed, it would have enhanced the material detection precision by 39.6% on FOD objects (FODO). Nevertheless, the definitive background (BG) could influence the classification between metal and plastic; thus, this needs radar or infrared to subdue it. Centered upon diverse publically attainable datasets such as ImageNet [7], Pascal VOC [8], and COCO [9], CNNAs were validated to execute finer detection and identification when compared with the conventional feature methodologies. Correlated with such physically modeled features, CNN-related features possess finer resolution and strength for FODD [10].

The FOD issue comprises two assignments: target location and object classification on the paved path. Focused on these two assignments, a new 2-phase structure has been modeled and presented in the present study. In the initial phase, the spatial pyramid pooling network with ResNet 101 (SPPN-RN101), as a complete convolutional network (NW), has been trained end-to-end to produce FOD position propositions. In the next phase, the CNN classifier (CNNC) centered upon Softmax with Adam Optimizer (SAO) was implemented to acquire the measurement criteria [11], rotation, and warping. Because of STN's execution, FODs can be rightly detected by produced features, regardless of image deformation. The emphasis and progresses in the present study can be encapsulated as the following ensues.

- A dense connection scheme has been embraced for enhancing the backbone NW's connection framework, and the identification accuracy has been optimized by reinforcing the feature distribution and assuring the maximal data flow within the NW.
- An enhanced spatial pyramid pooling (SPP) framework has been presented for pooling and concatenating the local features upon the different scales within a similar convolutional layer with fewer position mistakes while identifying little objects.
- SPPN-RN101 needs less memory and calculations, yet it exhibits finer execution—attains finer execution than the advanced NW paradigms while implementing the paradigm to the feature extraction (FE) NW of an object detector—and remains competent in the actual-time processing of a DNN-related to an object detector.
- A novel loss function (LF), comprising the MSE loss for position and the cross-entropy (CE) loss for classification, has been utilized for a quicker paradigm training speed and greater identification accuracy.

This study has been arranged as the following: Segment 2 highlights the associated studies for the survey on scanning methodologies (SMs) for FOD alongside the survey on neural network (NN)-related FOD. Segment 3 discusses the proffered method alongside pre-processing (PP), OD, and classification. Segment 4 presents the proffered paradigm's execution with the correlation of SPPN-RN101. Finally, Segment 5 sums up a comprehensive conclusion for the proffered paradigm.

Based on the convolution neural network, in order to improve the adaptability of the algorithm for the detection of foreign matters in aircraft, this paper puts forward the corresponding improved method. In view of the low image quality caused by dark and uneven illumination in the underground, firstly, the dataset is preprocessed. Aiming at the problem of the complex background of foreign objects and a large interference, the training data are correctly marked, the backbone network with good effect is adopted, and

the weighted feature fusion of feature layers with different scales is introduced, which speeds up the convergence speed and reduces the amount of parameter calculation. Under the condition of ensuring the detection speed, the detection accuracy has been greatly improved.

2. Related Works

For resolving the FODD issue, a few efficient algorithms have been proffered lately [12–16]. These efficient algorithms are centered upon disparate sensors, such as the dynamically scanning LiDAR system [12], mm-wave FMCW radar [14], and wideband 96 GHz Millimeter-Wave Radar [15], which can attain fine outcomes in disparate surroundings. A cosecant squared beam pattern in elevation and a pencil-beam pattern in azimuth, produced via folded reflectarray antenna by the phase-only control, has been assessed for identifying objects upon the ground [16]. A multiple-sensor system (MSS), centered upon FOD's innate feature, has been presented for detecting and identifying FOD. Centered upon a huge quantity of formal knowledge and a physically modeled feature extractor (FEr), such methodologies can transition an image's pel values (PVs) into an appropriate inward portrayal [17,18]. Hence, such methodologies remain efficient for identifying FODs with fewer noises yet remain inoperative for FODs with intricate definitive backgrounds and noise [19–22].

3. Survey on SMs for FOD

Su et al. [23] present a novel data compilation methodology for producing artificial data and assists in alleviating the data deficiency issue. Additionally, the authors proffer a novel identification methodology named Edge Proposal NW to lessen incorrect proposition positions and enhance identification execution. Finally, the authors performed multiple experimentations to authenticate the efficiency of the two methodologies and a few analysis experimentations to acquire an intense comprehension of the compiled data. Jing et al. [24] introduced a new methodology for identifying FOD centered upon random forest (RF). The data's intricacy in the airfield paved the way for images, and the FOD's versatility turned foreign object detection features were complicated for physical modeling. To subdue this adversity, the authors modeled the pixel visual feature (PVF), wherein weight and receptive field were discerned via learning and acquiring the optimal PVF (OPVF). Next, the RF's structure utilizing the OPVF for segmenting FOD was proffered. This proffered methodology's efficiency has been exhibited upon the FOD DS (FODDS). The outcomes exhibit that, when correlated with the initial RF and the DL methodology of Deeplabv3+, this proffered methodology remains dominant in precision and recall for FODD. Zhang et al. [25] modeled CNN with attention modules for precisely segmenting FOs out of an intricate definitive, and lively background. This proffered NW comprised an encoder and a decoder, and the attention mechanism was established in the decoder for capturing affluent semantic data. The visualization outcomes confirmed that the attention modules could concentrate on the vital areas' features and impede the impertinent definitive background that vitally enhanced the identification's accuracy. The outcomes exhibit that this proffered paradigm perfectly detects 97% of the FOs in the 1871 set of test images. Shukla et al. [26] established a procedure that employed CNNs for material detection. At first, this CNN paradigm trained itself with the features excerpted from the image samples. Lastly, the classification was performed with the CNN paradigm, which learned the classes acquired through the CNN of the materials' disparate classes. Experiential authentication has been performed by testing the CNNC's accuracy in opposing diverse DL classifiers. Ma et al. [27] implemented the CenterNet (CN) target identification AG for the foreign object identification of coal conveying belts (CCBs) in coal mines. Provided the CCBs' quick running speed and the impact of definitive background and light sources upon the objects to be examined, the CN's enhanced AG was proffered. Initially, the depth separable volume was presented. The goods substituted the conventional convolution that enhanced the identification effectiveness. Simultaneously, the normalization methodology

was enhanced to lessen the computer memory. Lastly, the weighted feature fusion (FF) methodology was included; thereby, every layer's features were completely employed, and the identification accuracy was enhanced. Son et al. [27] emphasized a DL-related foreign object identification AG (paradigm). The authors provided a synthetic methodology for effectually obtaining DL's training of DS, which could be employed for food quality assessment and food production procedures. Additionally, the authors carried out data optimization by employing a color jitter upon a synthetic DS and exhibiting that this technique remarkably enhanced the paradigm's illumination invariance features trained upon synthetic DSs. The paradigm's F-1 score (F1S), which trains almonds' synthetic DS at 360 lux illumination intensity, attained an execution of 0.82, the same as the paradigm's F1S, which trained the actual DS. Furthermore, the paradigm's F1S, which is trained with the actual DS when joined with the SDS, attained a finer execution when compared with the paradigm that was trained with the actual DS in the modification of illumination.

Alshammari [28] used the VGG16_CNN framework, which can be correlated with three advanced methodologies (moU-Net, DSNet, and U-Net) concerning diverse criteria. It was observed that the proffered VGG16_CNN attained 93.74% accuracy, 92% precision, 92.1% recall, and 67.08% F1 score. Tang et al. [29] proposed a novel method for smart foreign object detection and the generation of data automatically. A cascaded CNN was employed to detect foreign objects on the tobacco. The results showed notability for the proposed model. There are solutions and methods offered in research that have a lot of beautiful things [30–32].

The study introduces a new methodology for intelligent foreign object detection and automated data production. A cascaded CNN for identifying foreign objects on the tobacco pack's surface has been proffered in this study. This cascaded NW converts the examination into a 2-phase YOLO-related object detection comprising the tobacco pack's localization and the foreign object detection. For dealing with the image shortage with FOs, multiple data optimization methodologies were established to prevent over-fitting. Additionally, a data production method centered on homography transition and image fusion was established for producing synthetic images with FOs.

Centered upon the CNN to enhance the AG's adjustability to the identification of foreign objects in the airplane, the present study presents the correlating enhanced methodology. Considering the low image quality due to the dark and uneven illumination in the underground, initially, the dataset was pre-processed. Focusing on the issue of the FO's definitive background and big intrusion, the training data were rightly labeled, the backbone NW with a fine impact was embraced, and the weighted feature layers' FF having different scales were established, which accelerated the convergence speed and lessened the criteria computation quantity. Bound by the scenario of assuring the identification speed, the identification accuracy was highly enhanced. Table 1 shows the comparison between the existing and proposed methods.

Table 1. Comparison between existing and proposed methods.

Author/Year	Method	Advantage	Disadvantage
Noroozi et al. (2023) [16]	Augment method	It requires very few samples for training	High cost and time consuming
Zeitler et al. (2010) [17]	Arc-scanning synthetic aperture radar (ASSAR)	It shows robustness to irrelevant features.	Computational cost is higher
Wang et al. (2022) [18]	Prototype system	Easy and simple to use method	A larger tagged dataset is necessary for working
Zhong et al. (2022) [19]	Clutter-map constant FA rate (CMCFAR)	It is robust to overfitting	Slow prediction generator
Qin et al. (2021) [20]	Edge identification	Performs better than a single classifier	Need massive objects to obtain better results
Wong et al. (2022) [21]	Visual mapping-related system	It reduces variance	Difficult training of the network
Su et al. (2022) [23]	Edge proposal	It produces a more robust and accurate output which is resistant to overfitting.	Huge dependency on cluster centroids
Jing et al. (2022) [24]	Pixel visual feature	Can reduce the complexity in the data	Vanishing gradient and gradient exploding problems
Zhang et al. (2021) [25]	CNN	Less cost	Time consuming
Shukla et al. (2021) [26]	CNN	Less time consuming	Expensive
Ma et al. (2022) [27]	CenterNet	Easy and simple to use method	More time
Son et al. (2022) [27]	DL-related foreign object identification	It requires very few samples for training	Less speed

4. System Model

At first, the images out of the DS were classified as training data and testing data. The training data were pre-processed by rescaling and ensued by data optimization such as smoothening, zooming, rotation, and shear. The pre-processed image was provided to SPPN-RN101 for effectual OD. This placed the area proposition and correlation feature map (FM) via the ROI pooling procedure for acquiring the FM block in the fixed dimension. The identified object was classified by employing SAO. Figure 1 below illustrates FOD's comprehensive system framework.

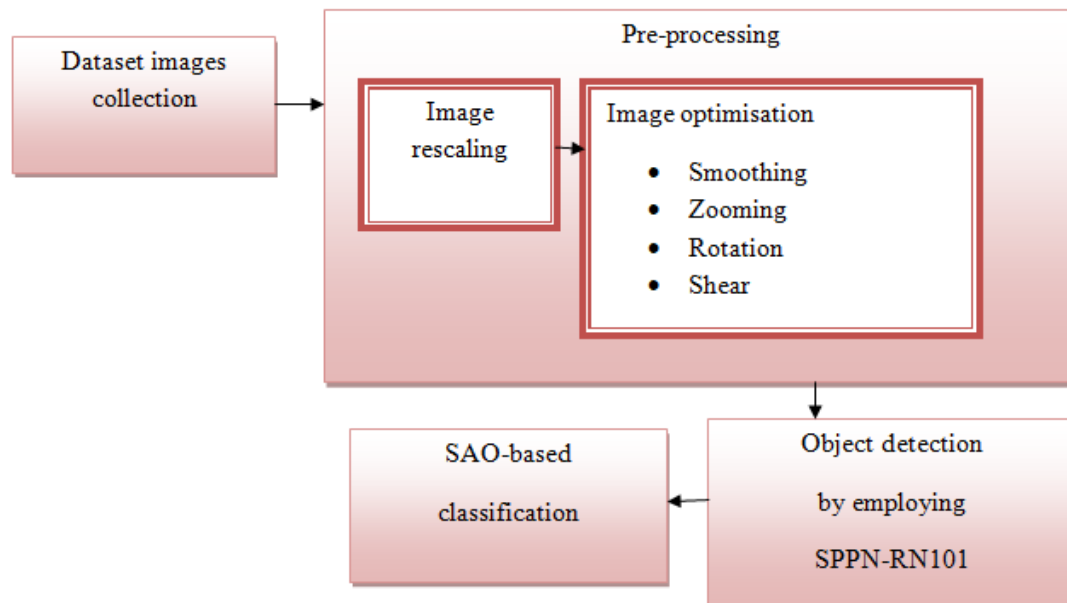


Figure 1. FOD's comprehensive system framework.

5. Image Collection (IgC)

As per the FAA, FOD generally incorporates the ensuing: 'airplane and engine fasteners (nuts, bolts, washers, safety wires, and so on), airplane parts (fuel caps, landing gear pieces, oil sticks, metal sheets, trapdoors, and tire pieces), mechanics' tools, food provisions, flight line articles (nails, personnel badges, pens, pencils, luggage tags, soda tins, and so on), apron objects (paper and plastic debris out of food and consignment pallets, baggage pieces, and detritus out of ramp devices), runway and taxiway items (concrete and asphalt blocks, rubber joint stuff, and paint fragments), building detritus (chunks of wood, stones, fasteners, and assorted metal parts), plastic and/or polyethylene items, natural objects (plant parts, wild animals, and volcanic ash), and contaminations out of winter scenarios (snow and ice).

For generating a practicable DS, which can be implemented in AP FOD administration, we gathered images in various scenarios. As weather and light situations in APs differ, the FODOs' DS should integrate this factor into incorporated data. Wet and dry surroundings give weather variance for FOD-A IgC. For light variance, the IgC procedure integrates bright, dim, and dark light scenarios. As every such environmental variance can be effortlessly excerpted to suit classification works, FOD-A incorporates classification labels for the weather (dry and wet) and light levels (bright, dim, and dark). As snow is instantaneously removed from of the AP surroundings, this remains inessential for integrating a snowy class. Whatsoever moisture stays subsequent to snow is removed and must even suit the wet class. FOD-A's dry and wet weather classes must include a big part of weather kinds that are actionable to APs. Substantially, the weather and light-level classification

annotations remain alongside FOD-A's concentration, that is, bounding box annotations (BBAs) for OD.

6. Image Rescaling

The input image (II) has been executed with rescaling and cropping from the middle, followed by decreasing average values (AVs) and scaling values by scale component. The present technique of seam carving (SC) rescaling remains to eliminate pels prudently. SC remains linear in the pels' quantity, and rescaling remains, hence, linear in the seams' quantity to be eliminated or included. On a mean, an image of dimension 300×300 to 100×100 was retargeted at around 2.2 s. Nevertheless, calculating tens or hundreds of seams lively remains arduous work. For dealing with this problem, this study gives a portrayal of the multiple-dimension image, which encodes, for an image of dimension s , a whole range of retargeting dimensions for 1×1 to $n \times m$, and still adds to $N' \times M'$, while $N' > M'$, $M' > n$. This data possesses much less of a memory footprint that could be calculated within a few seconds as a PP phase and enables the user to retarget an image that is consistently lively. An index map V of dimension $n \times m$ which encodes, for every pel, the seam index that eliminates it has been described, that is, $V(i, j) = t$ refers to that pel (i, j) which has been eliminated by the t -th seam elimination.

The coordinates of the object point M within the scene in the camera coordinate system (CS) $O_c - X_c Y_c Z_c$ remain x_c, y_c, z_c . The point M mapped to the image CS $O - XYZ$ remains M' , the coordinates remain (x, y) , and the focal length remains f . The mapping remains a 3D to the 2D procedure; the mapping association has been exhibited in Equation (1) and could be portrayed as a matrix format as in Equation (2).

$$\begin{cases} x = f \frac{x_c}{z_c} \\ y = f \frac{y_c}{z_c} \end{cases} \quad (1)$$

$$z_c m = \begin{cases} f x_c \\ f y_c \\ z_c \end{cases} \quad (2)$$

In the real scene, the image caught by the camera onsite remains a color image (CI); hence, it remains requisite for transforming the CI into a grayscale image (GI). Every pel within the CI is compiled of three color elements—red, blue, and green. CI's gray degree remains as the procedure of transforming three-element color values into a particular value as per a specific communication. The arithmetical equation for the same remains,

$$Gray = 0.299R(i, j) + 0.587G(i, j) + 0.114BR(i, j) \quad (3)$$

Amidst these, Gray portrays a gray value, $R(i, j)$ portrays a red element, $G(i, j)$ portrays a green element, and $BR(i, j)$ portrays a blue element. The CI is transformed into a GI. The finite variation computation gradient amplitude was acquired in the 2×2 neighborhood and remains susceptible to noise, which is effortless for identifying false edges, and the identification outcome remains coarse.

Lastly, the adaptable execution of artificially identifying the image's edge to fix the top and less thresholds remains bad. Centered upon this, the present work enhances the conventional canny edge identification (CEI) AG and acquires an enhanced edge identification methodology for the Canny operant. For an original image (OI) that is to be processed, this remains frequently similar to signal and noise. Hence, optimizing signal retention and removing noise remains the way to smooth an image. It needs a finer filtering methodology to smooth the image and remove noise and enable additional processing. The conventional CEI operant employs Gaussian filtering (GF) for smoothing the image. The arithmetical equation for the same remains,

$$G(x, y) = \frac{1}{2\pi\sigma^2} \exp\left[-\frac{x^2 + y^2}{2\sigma^2}\right] \quad (4)$$

GF remains a form of low-pass filtering, and the variance option remains crucial; its dimension portrays the band's narrowness and width. The bigger the variance, the narrower the frequency band (FB) remains that could repress the noise nicely; yet, this might lessen the image edge's (IE) sharpness because of the smooth transformation, and the IE particulars might be missed. The lesser the variance, the broader the FB remains, and additional edge particulars data could be sustained, yet the optimal noise decrement impact could not be attained. The image has been smoothed by employing acute median filtering. The noise within the image possesses its self-features. The acute value median filtering AG provides the parameters to assess the image pels' signal points and noise points as per their features and process these. The present study employs the AV median filtering to substitute the GF in the conventional CEI operant for smoothing the image. The arithmetical equation of the noise assessing parameters and filtering methodology rules of the filtering AG remain:

$$x_{ij} = \begin{cases} \text{noise}, & x_{ij} = \min(W[x_{ij}]), \max(W[x_{ij}]) \\ \text{signal}, & \min(W[x_{ij}]) < x_{ij} < \max(W[x_{ij}]) \end{cases} \quad (5)$$

where $[x_{ij}]$ portrays a digitalized image, *signal* portrays an SP within the image, *noise* portrays an NP within the image, $W[x_{ij}]$ portrays obtaining a window procedure upon the point x_{ij} within the image based upon the point (i, j) , $\min W[x_{ij}]$ portrays the minimal value for entire points within the window $W[x_{ij}]$, and $\max W[x_{ij}]$ portrays the maximal value for entire points within the window $W[x_{ij}]$. The filtering methodology could be depicted by:

$$y_{ij} = \begin{cases} \text{med}(W[x_{ij}]), & x_{ij} \in \text{noise} \\ x_{ij}, & x_{ij} \in \text{signal} \end{cases} \quad (6)$$

7. Image Optimization

Subsequently, data optimization has been executed subsequent to image PP. The proffered methodology's chief conception remains in producing alike images to construct a huge quantity of training data out of the OI data's little quantity. Since the training images' quantity impacts the IC's execution, which is centered upon a DL NW, an enormous DS should be made ready to train the DL NW. Even though the training data has been produced by employing the proffered methodology that remains similar to the initial data, these must remain non-linear to assure the diverseness needed to train the NW efficiently. The cause regarding relatedness remains that the training images' relatedness in a similar class must be ensured to some extent for the classifier's sufficient execution. Hence, this study proffers a data optimization methodology by producing an image that remains similar to the OI centered upon a similarity computation. Furthermore, this study aims at color perturbation since the color can be influenced by lighting and could be modified to another color. Thus, to produce novel images that are fairly centered upon the color perturbation to train the NW, this study tries to indicate the transition range. The form applying this conception, in this study regards the methodology of computing the relatedness betwixt the initial and the produced images. Hence, the PSNR conception for data optimization has been implemented.

$$10 \log_{10} \left(\frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n [I(i, j) - K(i, j)]^2 \right) = 20 \log_{10} (MAX_I^2) \quad (7)$$

$$-PSNR \sum_{i=1}^m \sum_{j=1}^n [I(i, j) - K(i, j)]^2 = mn * 10^{2 \log_{10} (MAX_I^2) 0.1 PSNR}$$

The relatedness betwixt two images could be computed. Hence, this study presents the methodology to produce the same images inversely regarding the PSNR equation (PSNRE). Initially, the PSNR range outcome value was fixed. Next, by inversely computing the PSNRE, the PVs' perturbation range has been discerned. Correspondingly, an NI DS comprising the training data has been produced. An image, which remains the same as the OI, must be produced to optimize the training data. Moreover, the PVs have been normally modified due to the lighting in the surroundings. Hence, the proffered methodology regards color perturbation. The inverse PSNRE has been employed to attain a rational transition. The inverse PSNR could be employed to discern the transition range concerning the color perturbation. For producing an NI, when regarding its relatedness with the OI, the color space's features have been definitively indicated. By employing the proffered methodology, the PVs have been calibrated into a single color space.

8. Object Detection

The OD NW embraces the traditional 2-phase identification procedure that has been developed out of the SPP NW and ResNet-101, such as the FE network (FEN) for applying the FE of high-resolution sensing images. Later, the region proposal network (RPN) was employed to perform the candidate region proposal (RP) and acquire the -confidence RPs and correlating coordinates on the FM attained by FEN. Finally, this places the RP and correlating FM via the ROI pooling procedure to attain a very precise object class and coordinates, and hence, this procedure can be called the classification.

9. ResNet-101

The ResNet-101 NW is composed of ninety-nine CLs that comprise four ResNet blocks (RNBs) and two fully connected layers. Because of its unique block framework, it can excerpt the deep feature within the image devoid of the vanishing gradient issue, which makes it immensely suitable for ship objects' FE in complex surroundings. For importing this NW into the OD NW, this study eliminates the final 2 fully connected layers and separates this into two portions: one has been employed as the FEN that incorporates the initial three RNBs, and the next (the fourth RNB) is employed subsequent to the ROI pooling procedure. The FEN configuration has been exhibited in the first part of Table 2.

Table 2. FEN configuration.

Block	Output Dimension	Layers	Layer Criteria
Conv1	$512 \times 512 \times 64$	Convolution	$7 \times 7, 64, \text{Stride} = 2$
Pool1	$256 \times 256 \times 64$	Max pooling	$3 \times 3, \text{Stride} = 2$
Block1	$128 \times 128 \times 256$	Convolution group	Stride = 2
Block2	$64 \times 64 \times 512$	Convolution group	Stride = 2
Block3	$64 \times 64 \times 1024$	Convolution group	Stride = 1
Block 4	$7 \times 7 \times 2048$	Convolution group	Stride = 2
Pool2	$7 \times 7 \times 2048$	Average pooling	$2 \times 2, \text{Stride} = 1$

Amidst these, the atrous convolution (AC) has a strong tool that could accurately and directly manage the FMs' filters. Consequently, the AC could excerpt multiple-scale features (MSFs) and generalize conventional convolution procedures. The ensuing Figure 2 depicts the ResNet-101 framework that comprises two layers. the ensuing equation σ portrays the non-linear function ReLu.

$$F = W_{2\sigma}(W_{1x}) \quad (8)$$

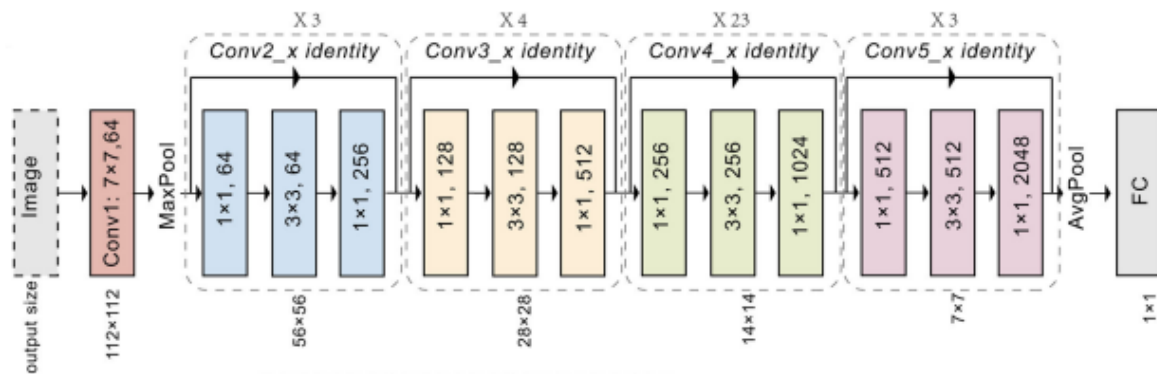


Figure 2. ResNet-101 framework.

Next, via a shortcut and second ReLU, the following output (OP) y is acquired:

$$y = F(x, \{W_i\}) + x \quad (9)$$

With regard to the stack or framework created by multiple stacks, if the IP is x , the feature this learns remains $H(x)$, and it is expected that this could learn the residual $F(x) = H(x) - x$. Its initial learning feature remains $F(x) + x$. If the residual remains zero, just the identity mapping is performed on the stack, and the NW execution does not drop. This facilitates the stack for learning novel features based on the IP features for possessing a finer execution. This can be exhibited as:

$$y = F_{(x[W_i])} + W_s x \quad (10)$$

Presuming a unidimensional signal, the OP FM ($y[a]$) correlating to every position a can be described as a function of the IP signal ($x[a]$) and a filter ($q[k]$) with length k :

$$y[a] = \sum_k x[a + r \cdot k] q[k] \quad (11)$$

in which r portrays the atrous rate, which explicitly impacts the sampling signal's stride. The filter's field of view remains additionally altered by modifying the r value. If r remains one, the process comprises a standard convolution. The atrous convolution (AtC) module could, hence, excerpt wider features scale if the atrous rate remains above one. In this, the convolutional procedure's receptive field rampantly enhances as the atrous rate remains optimized.

Subsequent to the AtC, the encoder module embraces the depthwise separable convolution. The convolution NW could lessen the computative intricacy and criteria quantity by decaying a standard convolution into a depthwise one, i.e., depthwise convolution (DWC) individually executes spatial convolution for every IP channel, and pointwise convolution has been implemented to join the DWC's OP. Subsequent to the EM, a decoder module has been utilized to enhance the edge segmentation's accuracy. Therefore, the last feature particulars, alongside the edge data, could be efficiently obtained and secured by remaining linked at lower-level and higher-level features excerpted out of the encoding-decoding procedure. In this, the excerpted paradigm LF has been provided as:

$$l_{overall}^s = l_{(j_i, k_i)}^s + l_R \quad (12)$$

in which $l_{(j_i, k_i)}^s$ and l_R indicate the softmax (SMx) CE LF and the regularization LF (RLF). These have been calculated in the ensuing equations:

$$l_{(j_i, k_i)}^s = - \sum_i^M j(x) \log k(x) \quad (13)$$

$$l_R = \mu \sum_i^M (x_i)^2 \quad (14)$$

in which j_i and k_i indicate the labeled PV and anticipated PV at i -th training image data, M indicates the images' quantity in the training DS and μ indicates the regularization criterion. The RLF could evade acute OF to serve the trained set, thus, optimizing the comprehensive ability for paradigm generalization.

10. SPP Network

For embracing the deep NW for images or random dimensions, the final pooling layer (PL) (for instance, pool5, subsequent to the final CL) has been substituted with an SPP layer (SPPL). Figure 3 exhibits this methodology. In every spatial bin, every filter's replies are pooled (max-pooling [MP] has been employed in this study). The SPP's Ops remain kM -dimensional vectors with the bins' quantity indicated as M (k remains the filters' quantity in the final CL).

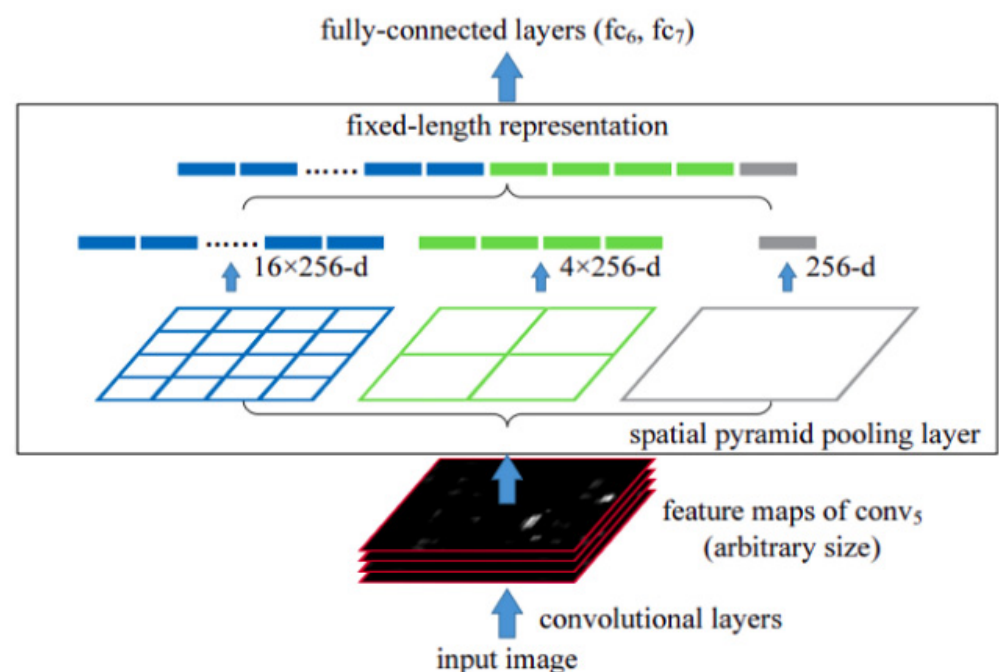


Figure 3. SPP NW Framework.

The fixed-dimensional vectors remain the IP to the FCL. Alongside SPP, the II could remain of whatsoever dimension. It does not merely facilitate a random aspect proportion but facilitates random scales. The II could be rescaled to whatsoever scale (for instance, $\min(w, h) = 180, 224, \dots$) and implement a similar deep NW. If the II remains at different scales, the NW (having similar filter dimensions) would excerpt features at different scales. For dealing with the problem of differing image dimensions in training, an array of predetermined dimensions is taken into account. We regarded two dimensions: 180×180 alongside 224×224 . Instead of clipping a little 180×180 area, the above-mentioned 224×224 area was rescaled 180×180 . Thus, the areas at both scales vary just in the resolution and not in content/layout.

For the NW to agree to 180×180 IPs, one more fixed-dimension-IP (180×180) NW was applied. The FM dimension subsequent to conv5 remained $a \times a = 10 \times 10$ here. Next, heretofore $win = a/n$ and $str = a/n$ were employed to apply every pyramid PL. The SPPL's OP of the 180-NW possesses a similar fixed length to the 224-NW. Consequently, this 180-NW possesses precisely similar criteria as the 224-NW within every layer. That is to say, in the course of training, the differing-IP-dimension SPP-net was applied by two fixed-dimension NWs, which exchanged criteria.

The global and LMFs were implemented jointly to enhance the OD's accuracy. The traditional SPP splits the IP FM into $a_i = n_i \times n_i$ bins (in which a_i portrays the bins' quantity in the feature pyramid's [FP] i th layer) as per the scales, which portray FP's disparate layers. The FMs were pooled by the sliding windows (SWs), whose dimensions remained similar to the bins. The $a_i.d$ dimensional feature vector, with d as the filter's quantity, was acquired to remain the FCL's IP. The novel SPP block has 3 MP layers (MPLs) presented between the deep coupling block and the OD layer within the NW. The IP FMs' quantity was lessened from 1024 to 512 by employing the 1×1 convolution. Subsequent to this, the FMs were pooled in different scales. In Equation (15), $\text{size}_{\text{pool}} \times \text{size}_{\text{pool}}$ indicates the SWs dimension, and $\text{size}_{\text{fmap}} \times \text{size}_{\text{fmap}}$ indicates the FMs' dimension.

$$\text{size}_{\text{pool}} = \left\lceil \frac{\text{size}_{\text{fmap}}}{n_i} \right\rceil \quad (15)$$

Consider $n_i = 1, 2, 3$ and pool the FMs by the disparate SWs whose dimensions remain $\left\lceil \frac{\text{size}_{\text{fmap}}}{3} \right\rceil * \left\lceil \frac{\text{size}_{\text{fmap}}}{3} \right\rceil$, $\left\lceil \frac{\text{size}_{\text{fmap}}}{2} \right\rceil * \left\lceil \frac{\text{size}_{\text{fmap}}}{2} \right\rceil$ and $\left\lceil \frac{\text{size}_{\text{fmap}}}{1} \right\rceil * \left\lceil \frac{\text{size}_{\text{fmap}}}{1} \right\rceil$ accordingly; thus, the pooling stride remains entirely one. The padding was employed for maintaining OP FMs' consistent dimension; thus, 3 FMs with $\frac{\text{size}_{\text{fmap}}}{1} * \frac{\text{size}_{\text{fmap}}}{1} * 512$ could be acquired. The NW's final portion remains the OD block, where the DC block's OP FMs with a high resolution were rebuilt and concatenated with the SPP block's OP FMs with a low resolution.

Next, these FMs were convoluted by the $1 \times 1 \times (K \times (5 \times C))$ convolution for acquiring the FMs of $S \times S \times (K \times (5 \times C))$ for OD. The SPPN-RN101's anticipations for every bounding box (BB) could be indicated by $b = [b_x, b_y, b_w, b_h, b_c]^T$, in which $[b_x, b_y]$ represented the box's middle coordinates, b_w and b_h represented the box's breadth and length, and b_c represented confidence. The offsets t_x, t_y from the image's upper left corner to the grid middle in b_x, b_y , and b_c , were constricted to $[0, 1]$ by the sigmoid function. Likewise, BB's ground truth (GT) could be indicated as $g = [g_x, g_y, g_w, g_h, g_c]^T$. Every BB's classification outcome remained $\text{class} = [\text{class}_1, \text{class}_2, \dots, \text{class}_c]^T$, the classification's GT remained $\text{Pr}(\text{Class})l \in C$, and the anticipated probability which the object appertained to the l class remained $\text{Pr}(\text{Class})l \in C$.

11. SAO-Related Classification

For speeding up the conventional Adam optimizer (AO) within the softmax layer optimization procedure, an AO with a power-exponential learning rate (PELR) was employed to train the CNN paradigm, in which the iteration trajectory and phase dimension were managed by the PELR for hitting the optima. The PELR could be modified adaptably as per the former phase's learning rate (LR) and the gradient association between the former phase and the present phase. The former gradient value (GV) was employed to modify the correction factor to reach the adaptable modification's requisites. It assists in adjusting the LR within a little range, every iteration's criteria remain comparably steady, and the learning phase was chosen as per the suitable GV for modifying the NW paradigm's convergence performance (CvP) and assuring the NW paradigm's steadiness and efficiency. AO remains an AG, which executes a stage-by-stage optimization on a haphazard objective function (ObF). The gradient's updated rules for the criteria remain:

$$\theta_{t+1} = \theta_t - \frac{P}{\sqrt{v'_{t+\varepsilon}}} \cdot m't \quad (16)$$

$$m't = \beta_1 m_{t-1} + (1 - \beta_1) g_t \quad (17)$$

$$v_t = \beta_1 v_{t-1} + (1 - \beta_2) g_t^2 \quad (18)$$

in which $\theta_t = \theta_{t+1}$ portrays the obF, t portrays the time criterion, $\beta_1, \beta_2 \in [0, 1] \in [0, 1]$ portrays the shifting average index's decomposition rate, η portrays the LR, ε portrays a

constant criterion $\varepsilon = 0.9999$, and $m't$ and $v't$ portray the first-order moment (FOM) and second-order moment (SOM) assessment subsequent to the gradient alteration accordingly. When mt and vt were initialized to the 0 vector, these were offset to 0. It remains requisite for correcting the deviation where $m't$ and $v't$ would be rectified by:

$$m't = \frac{1^{mt}}{1 - \beta_1^t}, v't = \frac{vt}{1 - \beta_2^t} \quad (19)$$

In the initial AO AG, the FOM to the non-middle SOM assessment was altered, and the offset was lessened. Nevertheless, in the procedure of rolling bearing fault detection and classification, the AG possessed a weak impact in fitting the paradigm's convergence state. A correction feature was included in the LR for dealing with the initial AO AG's weaknesses. The downward trends' PELR was employed as the fundamental, and the former phase's GV was employed to adjust this to reach the adaptable modification requisites to alter the NW paradigm's CvP. The paradigm for PELR remains:

$$\mu = \mu_0 m^{-k} \quad (20)$$

in which μ_0 denotes original LR $\mu_0 = 0.1$, k denotes the hyper-criterion, and m denotes the iterative intermediate where m has been discerned by the iterations' quantity. The maximal iterations' quantity has been described by:

$$m = 1 + \frac{t}{R} \quad (21)$$

in which t indicates the iteration quantity and R indicates the iterations' maximal quantity. If Equation (21) is joined with Equation (22), the LR update format remains:

$$\mu(t) = \mu_0 \left[1 + \frac{t}{R} \right]^{-k} \quad (22)$$

The CNN paradigm centered upon identity mapping was constructed. The paradigm comprises ten CLs, one MPL, and one FCL. Subsequent to the FCL, the enhanced AO was employed for updating and computing the NW criteria that impacted the paradigm's training and OP for turning them approximately or hitting the optimum value. Lastly, the data were sent via the SMx classifier, and the correlating classification outcomes remained the OP.

12. Performance Analysis

DS description—Images confined within the material recognition (MR) DS were in a zoomed-in form. This remains similar to the images gathered in the course of implementing FODD works that would not be zoomed into objects; hence, FOD-A presents images in a zoomed-out form with BBs. Moreover, the MR DS comprises around three thousand object instances, whereas the FOD-A DS comprises above thirty thousand object instances. Briefly, the FOD-A DS remains very suitable for FODD works since this comprises annotation formats finely matched to the AP's surroundings (that is, BB annotations alongside weather and light classification annotations), many other object instances, and depictive object classes. Figure 4 shows a sample dataset provided by Everingham et al. [8].



Figure 4. Sample dataset.

13. Comparative Analysis

In Figure 5a, for the adjustable clamp, the AP is 93.01%. In Figure 5b for the class AdjustableWrench, the AP is 77.7%. In Figure 5c, for the class battery, the AP is 95.34%. In Figure 5d for class Bolt, the AP is 98.22%. In Figure 5e for class BoltNutset, the AP is 87.74%. In Figure 5f for class Boltwashed, the AP is 89.99%. When analyzing Figure 5g, for the class Clamppart, the AP is 86.59%. In Figure 5h, for the class cutter, the AP is 98.75%. In Figure 5i, for the class Fuel Cap, the AP is 97.99%, and for the class hammer, the AP is 89.92%. As shown in Figure 5j and, for the hose, the AP is 64.29%, as shown in Figure 5k. For the class label, the AP is 90.23%, as shown in Figure 5l. The precision and recall curve is shown in Figure 6.

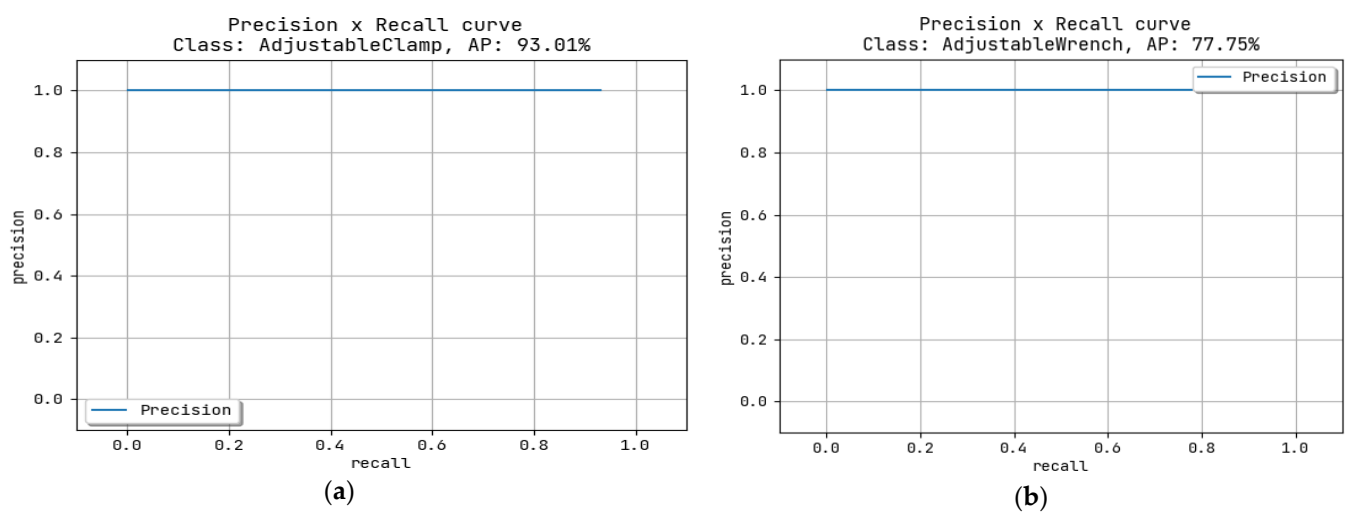


Figure 5. Cont.

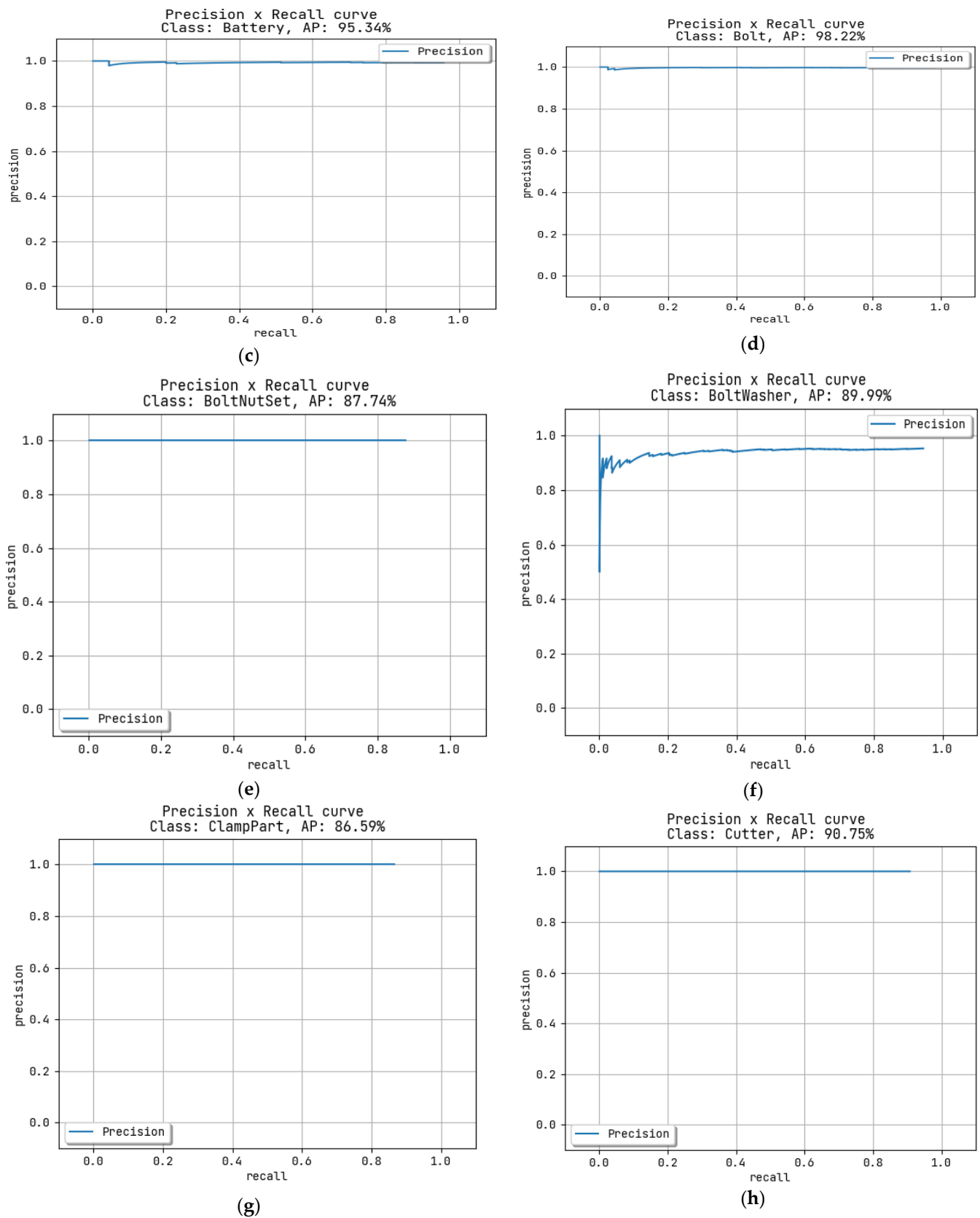


Figure 5. Cont.

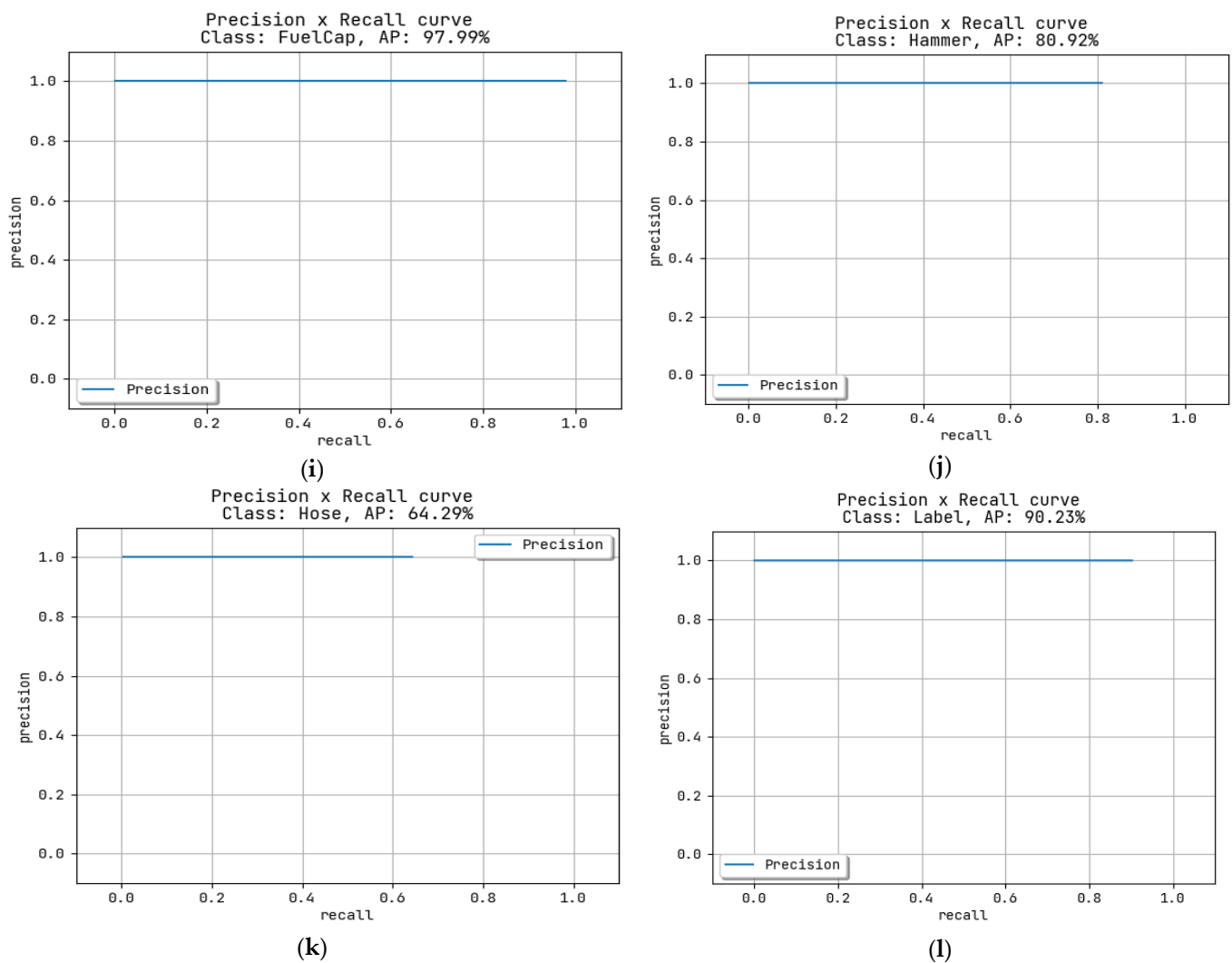


Figure 5. Analysis of precision (vs.) recall for various classes (a) AdjustableClump (b) Adjustable Wrench (c) Battery (d) Bolt (e) BoltNutSet (f) BoltWasher (g) Clamppart (h) Cutter (i) FuelCamp (j) Hammer (k) Hose (l) Labe.

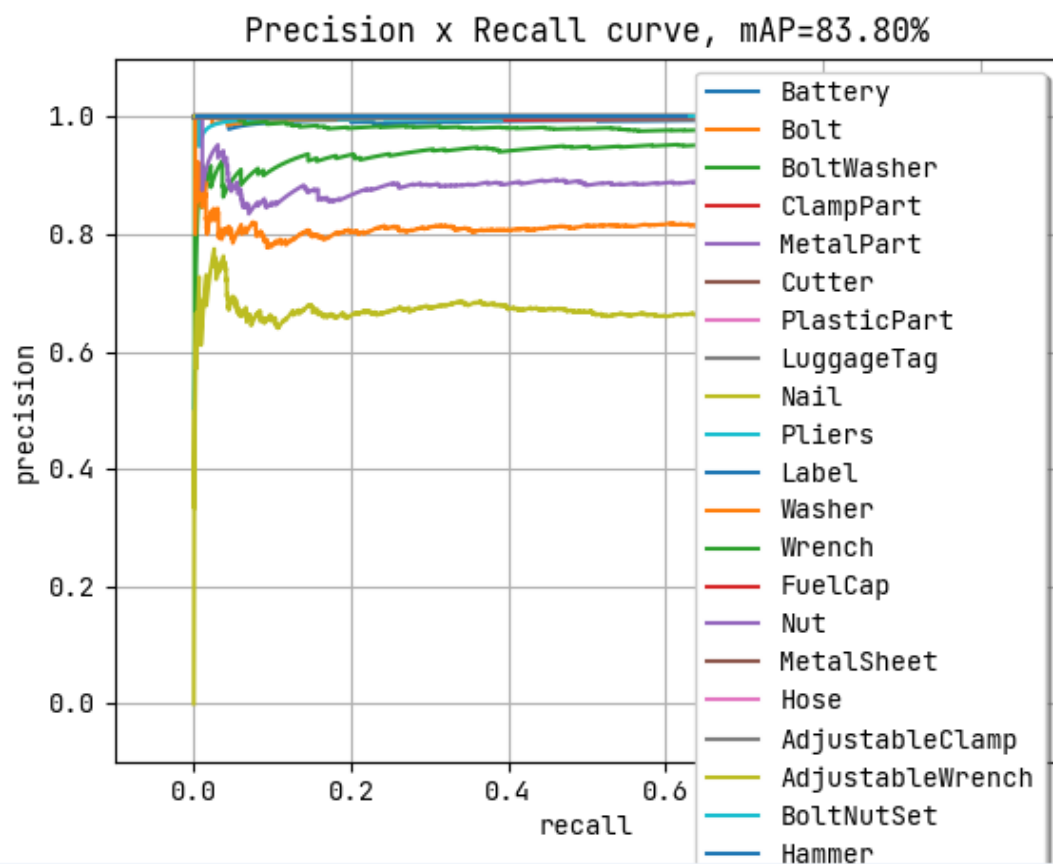


Figure 6. Precision vs. recall curve.

14. Conclusions

FOD MR remains an arduous and vital job, which should be executed to assure the AP's security. The normal MR DS was not implemented to FOD MR. Thus, a novel FOD DS was built in this research work. The FOD DS remains disparate from the former MR DSs, wherein the entire training and testing samples were gathered in outdoor surroundings, for instance, on a runway, taxiway, or campus. This study correlates the two renowned paradigms' execution with the novel DS. It has been observed that the SPPN-RN101 accomplishes this. A prospective study would examine feasible techniques for enhancing image segmentation and differentiating the FOs. The limitation of this work is moderate accuracy. The rest of the technologies, such as radar or infrared imaging, might be needed for finer detection outcomes.

Author Contributions: Methodology, R.C.C.; Software, A.A. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, Saudi Arabia, under grant number RP-21-07-10.

Data Availability Statement: Publicly available dataset was analyzed in this study. This data can be found here: <https://github.com/FOD-UNOmaha/FOD-data> (accessed on 10 August 2022).

Acknowledgments: The authors extend their appreciation to the Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University (IMSIU) for funding and supporting this work through Research Partnership Program no. RP-21-07-10.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Cao, X.; Wang, P.; Meng, C.; Bai, X.; Gong, G.; Liu, M.; Qi, J. Region based CNN for foreign object debris detection on airfield pavement. *Sensors* **2018**, *18*, 737. [\[CrossRef\]](#)
2. Yuan, Z.D.; Li, J.Q.; Qiu, Z.N.; Zhang, Y. Research on FOD detection system of airport runway based on artificial intelligence. *J. Phys. Conf. Ser.* **2020**, *1635*, 012065. [\[CrossRef\]](#)
3. Hussin, R.; Ismail, N.; Mustapa, S. A study of foreign object damage (FOD) and prevention method at the airport and aircraft maintenance area. *IOP Conf. Ser. Mater. Sci. Eng.* **2016**, *152*, 012038. [\[CrossRef\]](#)
4. Luo, M.; Wu, C.; Sun, H.; Xie, X.; Wu, X. Demonstration of Airport Runway FOD Detection System Based on Vehicle SAR. *IOP Conf. Ser. Mater. Sci. Eng.* **2018**, *452*, 042204.
5. Mehling, B. Artificial neural networks. *Stud. Syst. Decis. Control* **2019**, *131*, 11–35.
6. Xu, H.; Han, Z.; Feng, S.; Zhou, H.; Fang, Y. Foreign object debris material recognition based on convolutional neural networks. *Eurasip J. Image Video Process.* **2018**, *2018*, 21. [\[CrossRef\]](#)
7. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M. Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [\[CrossRef\]](#)
8. Everingham, M.; Gool, L.; Williams, C.K.; Winn, J.; Zisserman, A. The pascal visual object classes (VOC) challenge. *Int. J. Comput. Vis.* **2010**, *88*, 303–338. [\[CrossRef\]](#)
9. Lin, T.; Maire, M.; Belongie, S.J.; Bourdev, L.; Girshick, R.; Hays, J.; Perona, P.; Ramanan, D.; Zitnick, C.L.; Dollár, P. Microsoft COCO: Common Objects in Context. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; pp. 740–755.
10. Ren, S.; He, K.; Girshick, R.B.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 11–12 December 2015; pp. 91–99.
11. Li, X.; Li, L.; Flohr, F.; Wang, J.; Xiong, H.; Bernhard, M.; Pan, S.; Gavrilu, D.M.; Li, K. A unified framework for concurrent pedestrian and cyclist detection. *IEEE Trans. Intell. Transp. Syst.* **2017**, *18*, 269–281. [\[CrossRef\]](#)
12. Mund, J.; Zouhar, A.; Meyer, L.; Fricke, H.; Rother, C. Performance evaluation of LiDAR point clouds towards automated FOD detection on airport aprons. In Proceedings of the 5th International Conference on Application and Theory of Automation in Command and Control Systems, Toulouse, France, 30 September–2 October 2015; pp. 85–94.
13. Li, Y.; Xiao, G. A new FOD recognition algorithm based on multi-source information fusion and experiment analysis. *Proc. SPIE* **2011**, *8193*, 769–778.
14. Li, J.; Deng, G.; Luo, C.; Lin, Q.; Yan, Q.; Ming, Z. A Hybrid Path Planning Method in Unmanned Air/Ground Vehicle (UAV/UGV) Cooperative Systems. *IEEE Trans. Veh. Technol.* **2016**, *65*, 9585–9596. [\[CrossRef\]](#)
15. Özen, B.; Baykut, S.; Tulgar, O.; Belgül, A.U.; Yalçın, İ.K.; Şahinkaya, D.S.A. Foreign object detection on airport runways by mm-wave FMCW radar. In Proceedings of the 25th IEEE Signal Processing and Communications Applications Conference, Antalya, Turkey, 15–18 May 2017; pp. 1–4.
16. Futatsumori, S.; Morioka, K.; Kohmura, A.; Okada, K.; Yonemoto, N. Detection characteristic evaluations of optically-connected wideband 96 GHz millimeter-wave radar for airport surface foreign object debris detection. In Proceedings of the 41st International Conference on Infrared, Millimeter, and Terahertz Waves, Copenhagen, Denmark, 25–30 September 2016; pp. 1–2.
17. Zeitler, A.; Lanteri, J.; Pichot, C.; Migliaccio, C.; Feil, P.; Menzel, W. Folded reflectarrays with shaped beam pattern for foreign object debris detection on runways. *IEEE Trans. Antennas Propag.* **2010**, *58*, 3065–3068. [\[CrossRef\]](#)
18. Wang, Y.; Song, Q.; Wang, J.; Yu, H. Airport Runway Foreign Object Debris Detection System Based on Arc-Scanning SAR Technology. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–16. [\[CrossRef\]](#)
19. Zhong, J.; Zhang, K.; Zeng, Q.; Liu, X. A False Alarm Elimination Algorithm of Foreign Objects Debris Detection Based on Duffing Oscillator. *IEEE Access* **2022**, *10*, 7588–7597. [\[CrossRef\]](#)
20. Qin, F.; Bu, X.; Liu, Y.; Liang, X.; Xin, J. Foreign object debris automatic target detection for millimeter-wave surveillance radar. *Sensors* **2021**, *21*, 3853. [\[CrossRef\]](#)
21. Wong, B.; Marquette, W.; Bykov, N.; Paine, T.M.; Banerjee, A.G. Human-Assisted Robotic Detection of Foreign Object Debris Inside Confined Spaces of Marine Vessels Using Probabilistic Mapping. *arXiv* **2022**, arXiv:2207.00681. [\[CrossRef\]](#)
22. Su, J.; Su, Y.; Zhang, Y.; Yang, W.; Huang, H.; Wu, Q. EpNet: Power lines foreign object detection with Edge Proposal Network and data composition. *Knowl.-Based Syst.* **2022**, *249*, 108857. [\[CrossRef\]](#)
23. Jing, Y.; Zheng, H.; Lin, C.; Zheng, W.; Dong, K.; Li, X. Foreign Object Debris Detection for Optical Imaging Sensors Based on Random Forest. *Sensors* **2022**, *22*, 2463. [\[CrossRef\]](#)
24. Zhang, K.; Wang, W.; Lv, Z.; Fan, Y.; Song, Y. Computer vision detection of foreign objects in coal processing using attention CNN. *Eng. Appl. Artif. Intell.* **2021**, *102*, 104242. [\[CrossRef\]](#)
25. Shukla, A.; Kalnoor, G.; Kumar, A.; Yuvaraj, N.; Manikandan, R.; Ramkumar, M. Improved recognition rate of different material category using convolutional neural networks. *Mater. Today Proc.* **2021**. [\[CrossRef\]](#)
26. Ma, G.; Wang, X.; Liu, J.; Chen, W.; Niu, Q.; Liu, Y.; Gao, X. Intelligent Detection of Foreign Matter in Coal Mine Transportation Belt Based on Convolution Neural Network. *Sci. Program.* **2022**, *2022*, 9740622. [\[CrossRef\]](#)
27. Son, G.J.; Kwak, D.H.; Park, M.K.; Kim, Y.D.; Jung, H.C. U-Net-Based Foreign Object Detection Method Using Effective Image Acquisition System: A Case of Almond and Green Onion Flake Food Process. *Sustainability* **2021**, *13*, 13834. [\[CrossRef\]](#)

28. Alshammari, A. Construction of VGG16 Convolution Neural Network (VGG16_CNN) Classifier with NestNet-Based Segmentation Paradigm for Brain Metastasis Classification. *Sensors* **2022**, *22*, 8076. [[CrossRef](#)] [[PubMed](#)]
29. Tang, J.; Zhou, H.; Wang, T.; Jin, Z.; Wang, Y.; Wang, X. Cascaded foreign object detection in manufacturing processes using convolutional neural networks and synthetic data generation methodology. *J. Intell. Manuf.* **2022**, 1–17. [[CrossRef](#)]
30. Krichen, M. Anomalies Detection through Smartphone Sensors: A Review. *IEEE Sens. J.* **2021**, *21*, 7207–7217. [[CrossRef](#)]
31. Everingham, M.; Eslami, S.; Van Gool, L.; Williams, C.; Winn, J.; Zisserman, A. The pascal visual object classes challenge: A retrospective. *Int. J. Comput. Vis.* **2015**, *111*, 98–136. [[CrossRef](#)]
32. Krichen, M.; Mihoub, A.; Alzahrani, M.; Adoni, W.; Nahhal, T. Are Formal Methods Applicable To Machine Learning And Artificial Intelligence? In Proceedings of the 2022 2nd International Conference of Smart Systems and Emerging Technologies (SMARTTECH), Riyadh, Saudi Arabia, 9–11 May 2022; pp. 48–53. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.