

RESEARCH ARTICLE

Foreign Object Debris (FOD) Classification Through Material Recognition Using Deep Convolutional Neural Network With Focus on Metal

SYEDA MAHRUKH ZAINAB¹, KHURRAM KHAN^{1,2}, ADNAN FAZIL¹,
AND MUHAMMAD ZAKWAN¹

¹Department of Avionics Engineering, IAA, Air University, Islamabad 44000, Pakistan

²Faculty of Computer Science and Engineering, Ghulam Ishaq Khan Institute, Topi 23640, Pakistan

Corresponding author: Khurram Khan (khurram.jadoon@giki.edu.pk)

ABSTRACT Foreign object debris (FOD) is any undesired and unintended object placed or found in the specific vicinity of an aircraft (runway/ taxiway) that can cause damage to aircraft or harm personnel on board such as twisted metal strips, screws, nuts, and bolts, depleted concrete runway pieces, stones, pebbles and stationery items. To avoid FOD damages, all airport/ aviation organizations have deployed some sort of FOD prevention procedure. However, automatic FOD detection systems are still scarce owing to the inevitable reliance on human experts that lead to unavoidable human errors. Around 60% of FOD consists of metal which is the most deteriorating for an aircraft. Therefore, the implementation of material recognition techniques for FOD classification through Deep Convolutional Neural Networks (DCNN) is more important than FOD object detection as FOD could be of any shape, size or color. This paper developed a DCNN algorithm for FOD material classification with high accuracy for all included material classes (i.e., metal, concrete, plastic) in general and metal in particular. For this, a new dataset is introduced that consists of 2481 images taken on an operational airport runway in varying illumination and weather conditions. Through extensive testing, it was found that InceptionV3 is the best performing model with 18% improvement in metal recognition, and 11% improvement in average accuracy for all included classes.

INDEX TERMS Deep convolutional neural network (DCNN), FOD, material classification.

I. INTRODUCTION

Foreign object debris (FOD) is an undesired and unintended object placed or found in the specific vicinity of an aircraft (runway/ taxiway), which can cause damage to aircraft or harm personnel on board by getting sucked into the engine or hitting other parts of the aircraft. Therefore, FOD detection and prevention can play an important role in avoiding damage to aircraft or human lives. For effective prevention of FOD, airports use visual inspections and the use of sweepers, vacuums and magnet bars to collect debris. In the same context, some sophisticated equipment has been developed by certain companies for FOD detection. These include FODetect,

The associate editor coordinating the review of this manuscript and approving it for publication was Roberto Caldelli.

Tarsier and iFerret [1]. All these systems work on the principle of using cameras to capture the image for FOD detection, where the final verification is carried out by a human expert. These automatic

FOD detection systems can only be seen at a few airports around the globe. There are a few reasons for this scarce deployment, and the main reason is the last verification step which has two distinct disadvantages. First, a well-trained and experienced person is required which puts a burden on the airport authority to manage manpower overhead costs. The second disadvantage is the natural tendency of every human to cause an error, irrespective of his/her experience and expertise. Therefore, a better solution in such a scenario is the automatic recognition of FOD with sophisticated and cost-effective implementations through computer vision.

Deep Learning applications are vastly used in automatic recognition/ detection problems worldwide mostly in medical applications [2], [3], [4].

The field of computer vision can be explored in various ways to develop an efficient FOD recognition and classification system where object detection is a common problem in computer vision. Material recognition, on the other hand, is a relatively new but fundamental domain of computer vision. As FOD items could be of any type, color, or size, the object recognition approach for FOD classification would not be an effective choice. However, material recognition seems like a more promising domain for FOD classification [2] as segregating FOD items into material categories like metal, plastic, concrete, etc. is easier and covers almost all FOD types. Furthermore, in contrast to the decades of research on object recognition, material recognition is a flourishing and challenging field. The two main approaches followed by scientists for material recognition are handcrafted and automatic feature extraction as shown in Fig 1. Hand-crafted feature extraction further divides into surface reflectance [3], 3D texture [4], and feature fusion [5] approaches. All these approaches involve the collection of features from images through approaches like bidirectional reflectance distribution function Bidirectional Reflectance Distribution Function (BRDF) [6], Scale Invariant Feature Transform (SIFT) [7], Histogram of Gradient (HOG) [8], interest points [9], optical pyramids or optical flow [10]. These are computationally expensive and slow processes, especially for sensitive applications like airport security. Automatic feature extraction approaches, on the other hand, refer to those that involve acquiring image features using deep neural network techniques. These are desirable due to their fast recognition and higher precision.

Deep neural networks, used for automatic feature extraction in the material recognition domain, need huge datasets on the scale of millions of images to train the network for acceptable performance. Although huge datasets of 3 million images have also been used to train neural networks for material recognition [11], the images involved had materials photographed either indoors or in environments with light conditions significantly different from the actual FOD detection locations such as runways and taxiways.

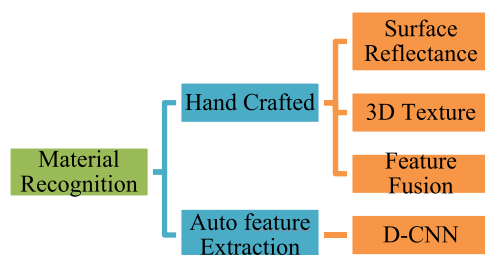


FIGURE 1. Material recognition approaches.

Moreover, metal recognition is important as more than 60% of FOD constitutes metals [12]. However, the results

of material classification in general and metal classification, in particular, were quite poor when these networks were used for material classification from FOD images [13]. Although material recognition for FOD detection is a new favorite yet its popularity in commercial airports is low due to the low accuracy of recognition [14].

The remainder of the paper is organized as follows: the literature survey is presented in Section II whereas Section III explains in detail, the methodology followed for the research work. Section IV presents the experimental results and finally, Section V concludes this research work. The major contributions of this work have been summarized below:

- A robust feature-based approach is proposed to effectively classify material categories of FOD items with an average accuracy of 92%.
- The shortcomings of the most recent related approach presented in [13] are highlighted and resolved using simple but effective techniques (Ref Section III).
- FOD dataset of real FOD items (found by safety personnel on an operational runway) in three material categories has been introduced. A dataset of 2010 train images, 336 validation images and 126 test images has been introduced with an almost equal number of images in each category. Moreover, the images have been taken in the morning, afternoon, evening and night with an almost equal number of images for each scenario (Ref Section II Part C).
- The theoretical foundation behind the developed algorithm for improved classification accuracies is established in order to show coherence between theory and achieved results.
- Our proposed approach outperformed the existing state-of-the-art algorithm by achieving the highest classification accuracy achieved so far in the field of FOD classification through image processing using DCNNs.

II. LITERATURE SURVEY

This section discusses in detail the previous research related to material recognition, FOD detection/ recognition, and computer vision with deep neural networks to address the FOD classification problem taken up in this paper.

A. FOD DETECTION

FOD detection has been the focus of a lot of research due to the unavailability of an efficient and cost-effective solution. Besides FOD walks and other such observation mechanisms involving humans, detection systems for continuous monitoring on runways and other aircraft movement areas are now available for improved FOD detection, including capabilities to work in supplement with the airport staff. FAA has issued a summary of the detection system categories applying sensors and advanced methods to efficiently detect FOD as per the standard [12]. The involvement of human verification has always been a part of FOD prevention schemes. Even in today's technology-driven world, FOD detection procedures involving humans are the primary tool of any FOD prevention

routine. However, aviation safety setups worldwide have realized the need, effectiveness and efficiency of automated FOD prevention. Due to the rise of FOD-related incidents, manual methods have proven obsolete. Moreover, among the available techniques for FOD detection, millimeter wave radars are the most popular [15] yet most costly. However, the cost of optical camera systems is the lowest and therefore considered suitable for automatic FOD detection [16].

When optical systems are involved, image processing and computer vision are the most suitable approaches. For efficient and fast FOD recognition/ detection, the use of neural networks/Artificial Intelligence (AI) is the future as full real-time detection has gained much attention in recent years [17]. In computer vision, object detection is a well-established and constantly optimized task involving deep Convolutional Neural Networks (CNNs) [18], however, it is not quite suitable for efficient FOD prevention due to vast variations in object types. Even in FOD detection research, object detection has been used mostly for classifying FOD items [19]. However, the number of types of FOD that may be found on runways/ taxiways is almost infinite. On the other hand, it is well established that the most harmful FOD items are metal FOD. Hence, a more useful approach may be to use material recognition for FOD classification.

B. MATERIAL RECOGNITION FOR FOD DETECTION

Using material recognition for FOD classification becomes challenging as the features required for material properties of the items are not shape-dependent but depend on other properties like reflection properties, transparency, brightness, texture etc. The two main approaches to feature extraction for material recognition are handcrafted features and automatic feature extraction [5]. Most of the earlier works [20] involve handcrafted approaches but they are computationally expensive and slow processes, especially for sensitive applications like airport security. On the other hand, automatic feature extraction approaches involve acquiring image features using CNN techniques. These are desirable due to their fast recognition and higher precision. These CNN-based approaches attained state-of-the-art results on CURET, KTH-TIPS, and FMD datasets. However, these datasets were not suitable for training a CNN for FOD detection as these were mainly acquired from customized apparatus or in an indoor environment. The illumination conditions as well as environments of these datasets including materials in context (MINC) dataset, did not fit for FOD detection tasks.

C. RELEVANT IMAGE DATASET

The Columbia–Utrecht Reflectance and Texture Database (CURET) [21], the KTH-TIPS [22], the Flickr Material Database (FMD) [23], and the Material-in-context Database (MINC) [11] are the popular datasets widely used for material recognition. The CURET dataset caters to illuminations as well as angles on the scale of 205 varieties for 61 texture images. KTH-TIPS has four samples for each category with a

total of 11 material categories. Each sample is imaged under various conditions to add diversity. The Flickr Materials Dataset includes 10 material categories with 100 images of each category. MINC has approximately 3 million image patches of 23 material categories tailored from ImageNet. Though a lot of diversity is included in all these datasets, these are improper for FOD material recognition as these are mostly taken indoors and do not match the FOD emergence conditions or illuminations. Therefore, the main hurdle in achieving good accuracy for FOD detection is the absence of a suitable dataset under real illumination and weather diversity. In this context, Xu et al prepared the FOD dataset with three categories having 1000 images each of metal, plastic, and concrete [13]. Though this was a huge contribution to the non-availability of FOD-related datasets, still this dataset had the following drawbacks that became the reason for low recognition accuracy:

- Absence of real background environment due to security reasons and unapproachable airport runways.
- Images do not have FOD placed in the center hence the network cannot learn the edges and curves efficiently through scale invariance.
- Most images in the dataset (90%) consist of large items like wrenches, long bolts, plastic bottles, laptop chargers, electric sockets, large pieces of concrete, etc. Whereas almost all of the real FOD items found on the runway/ taxiway of considered airfield consist of small items like rusted nuts/ bolts, small screws, broken metal strips, bullet shells, plastic blanks, small pebbles, etc.

A recent addition to the relevant datasets for FOD classification/ detection is the FOD-A dataset [24]. This dataset has 31 object categories with 30000 instances along with their bounding boxes. This dataset varies from the material classification dataset of this work and the previous work [13] in the following ways:

- The dataset has been prepared to target the object detection tasks for FOD categories. However, the list of FOD types/ objects is endless. On the contrary, many object categories like hammers, bolts, screws, nails, tools, metal strips, etc can be covered by just one category i.e. metal. Moreover, as some materials are more damaging due to their natural characteristics such as metal causing the most damage and consisting of more than 60 % FOD [12], it is more feasible and beneficial to work on material categories rather than object categories when it comes to FOD classification.
- In Machine Learning Computer Vision (MLCV) tasks, image variations of the slightest nature make a huge difference in classification/ detection accuracies. This introduced the image augmentation techniques widely used now to handle overfitting issues. FOD background is one such feature of FOD datasets that plays a key role in predicting the applicability concerning the practical implementation of an algorithm trained on relevant images in an airfield environment. Although the authors of the FOD-A dataset considered the varying light and

weather conditions, they have not mentioned the background in which the FOD objects of their dataset were placed. Both the material classification works for FOD (this work and [13]) prove that images with airport runway backgrounds are necessary for the practical implementation of FOD datasets in real airport environments.

- For any machine learning algorithm, the object categories should almost be equal in the number of instances/ images per category for the algorithm to be unbiased and more reliable in prediction. However, the number of instances per category of the FOD-A dataset varies from less than 400 instances to around 3200 instances. This is a huge variation for any algorithm as the algorithm would be inclined towards the category with greater instances. Therefore, both the material classification datasets (this work and [13]) have the same or almost the same number of images per category. This ensures the reliability/ robustness of the designed algorithm in practical applications for prediction.

Hence, from the above discussion, it can be inferred that the dataset introduced in this work has been carefully crafted to best match the real-world application of FOD classification through material recognition and the algorithms designed in this work are more robust and reliable for real airport applications. Moreover, transfer learning and data augmentation techniques have been used to address the smaller size of the dataset.

III. METHODOLOGY

This section explains the implementation of the existing FOD material classification techniques. Moreover, the section also explains the proposed methodology in terms of development, implementation and improvement in accuracy along with the details of the new dataset.

A. IMPLEMENTATION OF ALEXNET ARCHITECTURE

Xu et al. [13] used transfer learning in which Alex Net pre-trained on MINC was fine-tuned completely on the FOD dataset. Although this was the first work of its kind, the resulting accuracies were quite low. Details of the approach followed by Xu et al in comparison with the approach applied in this work are as follows:

- **Choice of the network:** Xu et al chose Alex Net as their network of choice for FOD recognition through material recognition. It was also assumed that the deeper the network the poorer the performance would be on FOD recognition accuracy. However, keeping in view the complexity of the material recognition task, the deeper the net the better should be the recognition accuracy [25] provided the other factors like correct choice of TL dataset, correct training methodology, and exploitation of techniques for handling overfitting are ensured. In this paper, InceptionV3 and ResNet architectures are selected as the evolution of CNNs and the work in [26] highlights that these two are the best performing networks for computer vision tasks in general

and material recognition tasks in particular. However, Alex Net was also implemented with different choices and much better results were achieved on previously available as well as newly developed FOD datasets.

TABLE 1. Confusion matrix of the algorithm proposed by Xu et al.

Label	Predicted Label		
	Metal (%)	Plastic (%)	Concrete (%)
Metal	66.67	11.43	21.90
Plastic	22.13	67.66	10.21
Concrete	1.00	0.00	99.00

- **Choice of TL weights dataset:** Xu et al used Alex Net pre-trained on MINC which is a dataset for material recognition. However, the weights of this pre-trained network were optimized for materials recognition and localization in the context of the image surroundings. In this paper, we used AlexNet pre-trained on ImageNet with thousand categories of object recognition. The network already optimized to classify the item in the center of the image was trained with the object category named on the material of the object.
- **Training methodology:** Xu et al used a pre-trained Alex Net and fine-tuned the complete network on the FOD dataset. The strategy for TL is used in combination with the training of a few last layers where complex features are learned by the network. On the other hand, in this research, only fully-connected layers of Alex Net have been trained on both datasets (Chinese [13] and newly developed in this work) and the results have shown remarkable improvement for both datasets.
- **Dataset type and size:** FOD dataset developed by Xu et al has all training images with the general items of the relevant material category. Dataset developed in this paper was completely collected with the FOD items found on the actual airfield's runway/ taxiway and was equally distributed in morning, afternoon, evening, and night timings for the diversity of illumination conditions. Furthermore, all the FOD items found by the flight safety set up at the airfield were small items like nuts, bolts, screws, broken metal strips, stationery items, plastic blanks, pebbles, etc.

B. DEVELOPMENT OF ALGORITHM FOR FOD CLASSIFICATION

Keras applications have pre-trained models of most state-of-the-art D-CNNs. The basic concept behind the algorithm development was to use pre-trained networks on ImageNet as these models performed outstandingly on object recognition tasks and also in Shang et al.'s work [26]. Also, all these

algorithms were optimized for recognition tasks and hence it was expected to work better on datasets having FOD items in the center of the image with a runway in the background. It was proven by results that the networks performed better with ImageNet as it is so far the largest data set available for feature learning and hence resulted in better training of the network. The prediction accuracy on all networks pre-trained on ImageNet was better than the results achieved by Xu et al in their work shown in TABLE 1. Pre-processing as per the model in use was used to make the input images as per the size and form expected by the model. ResNet expected images of size (224×224) or (299×299) while InceptionV3 expected input image size to be (299×299) . Stochastic gradient descent (SGD), Adam and RMS Prop were the three different optimizers used in this work. It was observed that on the newly developed dataset, the Adam optimizer performed best for ResNet50 and ResNet101. On the other hand, the best model optimization was achieved with the RMSprop optimizer for InceptionV3. Details of the results would be discussed in section IV.

C. IMPROVED METAL RECOGNITION ACCURACY

Along with the development of an algorithm with better test accuracy in general, the main focus of this work was to improve metal recognition accuracy in particular. It was already expected that the overfitting issue would be faced as the collected dataset was on the scale of a few thousand training images. For improvement of the metal accuracy, the following measures were taken which proved very effective.

1) SELECTION OF D-CNNs

Keeping in view the earlier work by Shang et al. [26], InceptionV3 was selected as the main D-CNN for achieving improved metal recognition accuracy. Among many available D-CNNs, InceptionV3 and InceptionV4 are the best-performing networks [26]. However, pre-trained inceptionV4 was not available in the Keras API applications, so InceptionV3 was implemented. As expected, the results of InceptionV3 were the best whereas ResNet50 and ResNet101 were used in this work to verify if the conjecture of a deeper network for FOD classification made by Xu et al. [13] in their work was correct. However, their conjecture was proven wrong as the results proved that with the right choices made, as explained further, ResNet performed better than AlexNet on the FOD dataset.

2) SELECTION OF PRE-TRAINED WEIGHTS

As explained earlier, the use of transfer learning becomes inevitable for most machine learning tasks as collecting datasets on the scale of millions of images is not always practically possible. Moreover, training a network from scratch is quite cumbersome and requires high-end processing hardware that ultimately makes the overall application of such an algorithm inefficient and questionable. In most cases, the demand is for lesser computation with faster results and higher accuracy. This calls for the new hallmark of deep

learning to be used for most applications, and transfer learning comes to the rescue in such cases. However, the selection of a source domain in light of the specific task domain is necessary. For example, image classification cannot be achieved from a network trained for Natural Language Processing. Hence, the first rule is to use a network trained for image classification. Where MINC is a compatible dataset when it comes to numbers, yet, ImageNet has the edge of being used in ILSVRC and the deep learning models have been optimized for ImageNet. Furthermore, MINC has only 23 categories while ImageNet has 1000 categories which make it more robust for related tasks as it trains the deepest layers for general features in a better way to adapt the later layers for any new dataset of computer vision. Hence, it was expected that the results of FOD classification even in the material categories would be better following this methodology. The results proved that with models pre-trained on ImageNet instead of MINC, all networks performed better than the published work [13].

3) SELECTION OF OPTIMIZERS

In both the works of Xu et al. [13] and Shang et al. [26], stochastic gradient descent is the optimizer used; however, Yaqub et al. in their work [27] found that among any tested optimizers, Adaptive momentum (Adam) optimizer performed best. Keeping this in view, SGD, as well as Adam optimizers were used for training all the models in this paper. The results proved that in all the models, Adam produced the best convergence and resulted in the best accuracies.

4) TECHNIQUES TO RESOLVE OVERFITTING

Owing to the small size of the collected FOD dataset in this work, it was evident that the algorithms would face overfitting issues. Overfitting is the term used to represent that the model fits too well on the training data and hence fails to generalize so the performance on unseen data is poorly affected. The following techniques were used to handle overfitting:

1. Reduced Complexity
2. Early Stopping [28]
3. Data Augmentation [29]
4. Dropout

D. NEWLY COLLECTED FOD DATASET

For effective training of CNNs, a relevant dataset is very crucial and plays a major role in defining the performance of the algorithm. For FOD classification, only one published dataset [13] is currently available. That too mostly has a concrete background of the university campus that does not properly emulate the runway background. Also, the types of FOD used in the dataset in each of the classification categories do not include the real FOD items found on runways by flight safety setups. In this context, a new dataset has been collected in this work where metal, plastic, and concrete were the three typical FOD materials constituting the dataset. According to Federal Aviation Administration (FAA), these materials appear most frequently on runways and taxiways.

For this dataset, a camera was held two meters from the ground and the FOD was placed approximately five meters from the camera. This setting had been selected according to the already available FOD dataset. The dataset includes an almost equal number of images for each category in the morning, afternoon, evening, and night illuminations and various weather conditions. A sample of the FOD dataset collected in this work is shown in Fig 2. The details of the images are shown in TABLE 2.

TABLE 2. New FOD dataset details.

Category	No. of Images		
	Training	Validation	Test
Metal	600	100	40
Concrete	760	112	40
Plastic	650	124	46
Total	2010	336	126

Hence, the dataset collected in this work has the following improvements as compared to the already published dataset:

1. Used real runway background of an operational airfield.
2. Used real FOD items are frequently found by the Flight Safety Team deployed at the airfield.
3. Diverse illumination conditions of morning, afternoon, evening, and night with equal proportions for each category.
4. Images have FOD placed in the center hence the network can learn the edges and curves efficiently through scale invariance (zooming due to convolution operations).

IV. EXPERIMENTAL RESULTS

This section presents the results from FOD classification algorithm predictions based on the methods developed in Section III.

A. ALEX NET ALGORITHM

AlexNet was implemented in two ways. First, AlexNet pre-trained on ImageNet was fine-tuned on the Chinese FOD dataset [13]. Second, AlexNet pre-trained on ImageNet was fine-tuned on the new FOD dataset of this work. This was done on the Pytorch framework as pre-trained AlexNet was not available in Keras applications.

1) ALEX NET ON CHINESE DATASET

Chinese dataset [13] was first used for training, validation and predictions on AlexNet. All images were transformed in the following ways:

1. Resize to 256×256 .
2. Centre crop to 224×224 .
3. Random horizontal flip.
4. Random vertical flip.
5. Random rotation of 10° .

6. To tensor.

7. Normalization

After the above-mentioned transforms were applied to the images, Alex Net was prepared and trained in the following steps:

TABLE 3. AlexNet training steps.

STEP #	ACTION
1	Download AlexNet pre-trained on ImageNet (Five convolution layers and three fully connected layers).
2	Freeze the whole network except fully connected layers.
3	Adjust the last softmax layer for three outputs (metal, concrete, plastic) instead of 1000 (ImageNet classes).
4	Instantiate Cuda device, move input and Alex Net to GPU for speed.
5	Train Alex Net over 20 epochs with SGD optimizer (learning rate 0.001, momentum 0.9, and weight decay 0.5).
6	Save model with best validation accuracy in *.pth file.
7	Load saved model to predict test images of the Chinese dataset.

The result, as shown in TABLE 4, proved that the results of the ImageNet pre-trained model with the same parameters were far better than the model trained on MINC. However, it was observed that most FOD items in the Chinese dataset were large objects whereas those found on a real runway are small objects like nuts, bolts, and small metal strips. It was also noted that although metal and concrete accuracies improved drastically, plastic recognition had shown a considerable decline. Hence, it was decided to train AlexNet on the New dataset and compare performance for cross-validation of both datasets. Furthermore, the developed algorithm should be robust as it would be expected to perform well in real applications for real FOD items.

TABLE 4. Confusion matrix of chinese dataset with AlexNet pre-trained on imagenet.

True Class	Predicted Class			Average (%)
	Metal (%)	Concrete (%)	Plastic (%)	
Metal	94	0	0	81
Concrete	0	100	0	
Plastic	52	0	48	

2) ALEX NET ON NEW DATASET

After appropriate labelling of the new dataset into train and validation folders, it was uploaded to Google Drive and accessed in Colab for algorithm training, validation, and

prediction. The same transformations were applied to the images and the same steps for preparation and training of AlexNet were performed as mentioned in section IV subsection I. However, it was observed throughout the research and experiments on all models (except InceptionV3) that the new dataset performed much better with Adam optimizer as compared to SGD.



FIGURE 2. Sample images of new fod dataset.

It was inferred that the better performance of the Adam optimizer was due to the self-reducing capability of its learning rate. Due to small FOD items, the network required greater iterations and a systematic reduction in learning rate. Hence, the model achieved optimized validation accuracy on the Adam optimizer with a learning rate of 0.01, beta_1 0.9, beta_2 0.999, epsilon 0.6, and 128 epochs. The optimized model with the best validation accuracy was saved in .pth format and the same saved model was uploaded to predict test images of the new dataset. The prediction results are shown in TABLE 5.

TABLE 5. Confusion matrix of the new dataset with AlexNet pre-trained on imagenet.

True Class	Predicted Class			Average (%)
	Metal (%)	Concrete (%)	Plastic (%)	
Metal	70	20	10	82
Concrete	0	100	0	
Plastic	8	16	76	

It was observed that though the performance of metal recognition was not as high as AlexNet on the Chinese dataset, the model seems more reasonable as all classes have comparative recognition accuracy with average true positive accuracy almost the same as that of the Chinese dataset. Still, these results were better than the state-of-the-art results for all classes.

3) CROSS MODEL VALIDATION

To check the robustness of both the models trained on the Chinese dataset and the new dataset, test images of the new dataset were predicted by the AlexNet model trained and validated on the Chinese dataset, and test images of the Chinese dataset were predicted using AlexNet model trained and validated on the new dataset. The result as shown in TABLE 6 validates that Alex Net trained on the new dataset seems to be more robust to unseen images as compared to the one trained on the Chinese dataset.

The matrix in TABLE 6 clearly shows that the model trained on the new dataset performed much better on Chinese test images. Furthermore, the model trained on the Chinese dataset predicted all test images of the new dataset as metal which itself is quite erratic behavior. In contrast, the model trained on the new dataset had the representation of all classes in predictions but with lower accuracy. Hence, the new dataset is more robust and reliable for further deep learning of different architectures.

TABLE 6. Cross model validation matrix for AlexNet.

Train	Test	
	Chinese Dataset (%)	New Dataset (%)
Chinese dataset	81	33
New dataset	44.7	82

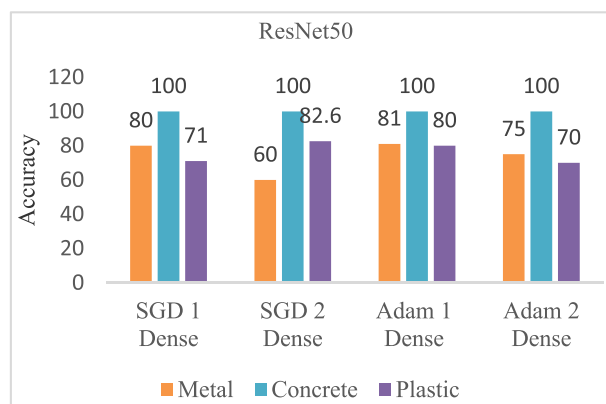


FIGURE 3. Graph showing recognition accuracy of each class with SGD and Adam optimizers with one or two dense layers on top of base ResNet50 model.

B. RES NET ALGORITHM

The first step in the development of the new algorithm was to check the new FOD dataset on ResNet50. The residual blocks in ResNet architecture help avoid blown-out gradients. Hence, its performance was much better even though being deeper than previous architectures. Xu et al. [13] did not work on ResNet assuming that deeper architectures would not work well on FOD classification. However, the evolution of deep nets [30] shows that ResNet architectures are the best after

InceptionV3 and V4. It was hence planned to test ResNet50 and ResNet101 architectures on the new dataset.

1) RESNET50 ON NEW DATASET

The already uploaded new dataset on Google Drive was accessed through Colab. The preparation and training of ResNet50 were done in the following steps:

STEP #	ACTION
1	Resize images to 224 x 224
2	Import ResNet50 (pre-trained on ImageNet) from Keras applications
3	Add new dense layer(s) on top of the base ResNet50 model. (Include top layer set to false)
4	Freeze the base model to train only the top fully connected layer(s)
5	Optimize the model with both SGD (lr 0.0001, decay 1e-6, momentum 0.9, nesterov True) and Adam (learning rate 0.01, betas (0.9, 0.999), epsilon 0.6) optimizers.
6	Save model with best validation accuracy in pth file
7	Load saved model to predict test images of the Chinese dataset

The results with one (added) dense layer optimized with Adam optimizer were state-of-the-art. The results (see Fig 3) with two dense layers were also better than the published test accuracies for all classes but slightly lower than those with one dense layer.

2) RESNET101 ON NEW DATASET

All the steps followed for ResNet50 were also followed for ResNet101. However, the following changes were made keeping in view the results achieved for ResNet50:

1. ResNet101 was only optimized with Adam optimizer, owing to its poor performance with ResNet50.
2. ResNet101 was only optimized with one dense layer added, as two layers did not perform well for ResNet50.
3. The results achieved with ResNet101 were also state-of-the-art as shown in TABLE 7.

Hence, it was proven that with the methods developed in this work, AlexNet and ResNet architectures had comparable average performance. However, as the main focus of this work is metal recognition accuracy, the second to best performing network was ResNet50 with almost 81% accuracy.

C. INCEPTIONV3 ALGORITHM

As explained in the previous sections, the selection of InceptionV3 was due to the following facts:

1. Inception architectures are proven to be among the best-performing architectures in terms of recognition accuracy.
2. Inception performed the best for material recognition tasks in Shang et al.'s works [26] in comparison with ResNet-18, ResNet-34, Google Net, and VGG16.

TABLE 7. Confusion matrix of the new dataset with resnet101 with ADAM optimizer and one dense layer added on top.

True Class	Predicted Class			Average (%)
	Metal (%)	Concrete (%)	Plastic (%)	
Metal	75	10	15	82
Concrete	0	100	0	
Plastic	20	8	72	

Hence, the same steps as followed for ResNet50 and ResNet101 were taken for the development of the InceptionV3 algorithm. However, the following changes were made to this algorithm to achieve optimization:

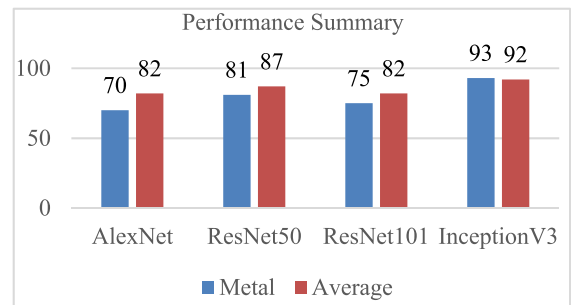


FIGURE 4. Graph showing metal and average recognition accuracies of each D-CNN.

1. All the images were resized to 299 x 299.
2. InceptionV3 demonstrated poor performance on SGD as well as Adam optimizers hence RMS Prop (lr 2e-5) was used and gave state-of-the-art results for the new dataset.
3. It was inferred that InceptionV3 was pre-trained on ImageNet using RMS Prop [31], hence, it would give better results with the same optimizer used to fine-tune fully connected layers.
4. Two dense layers were added on top of the model after excluding the top of the pre-trained architecture.
5. Prediction results of InceptionV3 achieved on the new dataset test images are shown in TABLE 8 and along with confusion matrix shown in TABLE 9.

The performance parameters of InceptionV3 were noted as follows:

The proposed method outperforms the existing schemes for FOD classification. However, the limitations of the proposed method are as follows:

- A comparatively small number of training images (i.e. 2010 images) limits the performance. A larger dataset can provide even better performance.
- The consideration of only one runway background limits the generalization of the algorithm.
- Inference time of 16ms/image may also be reduced by using MobileNet or SqueezeNet in future work.

TABLE 8. Performance parameters for inceptionv3.

Performance Measure	InceptionV3
Average Accuracy	92
Precision	0.91
Recall	0.91
F-Score	0.92
Train Time	13.44 min
Inference Time	16ms/image
Epoch Time	6.3s (avg)

TABLE 9. Confusion matrix of the new dataset for inceptionv3 with RMSPROP optimizer and two dense layers added on top.

True Class	Predicted Class			Average (%)
	Metal (%)	Concrete (%)	Plastic (%)	
Metal	93	3	4	92
Concrete	0	100	0	
Plastic	10	7	83	

D. SUMMARY OF RESULTS

Hence, a new FOD dataset was developed and used to optimize AlexNet, ResNet50, ResNet101, and InceptionV3 algorithms. All architectures gave state-of-the-art results with InceptionV3 proved to be the best for metal accuracy as well as average accuracy. The summary of results of all architectures for metal and average recognition is shown in Fig 4.

V. CONCLUSION

From this research, it was concluded that FOD classification through material recognition can be achieved with high recognition accuracy with different architectures provided the models are pre-trained on ImageNet and a suitable dataset of real FOD with real runway background is used. In this context, a new FOD dataset of almost 2500 images of real FOD found on the considered airfield was collected with a real Runway as background. Furthermore, AlexNet, ResNet50, ResNet101, and InceptionV3 algorithms pre-trained on ImageNet were optimized on the new dataset. All the models demonstrated more than 80% accuracy outperforming state-of-the-art work for metal recognition. The results also demonstrated that the InceptionV3 is the best performing network with 93% metal recognition and 92% average recognition for all three classes i.e. metal, concrete, and plastic. Moreover, ResNet50 was the second best with 81% metal recognition and 87% average recognition. The use of real FOD items and real Runway background in the dataset makes the algorithm more reliable for subsequent deployment for real-time applications on operational Runways.

The current research work was done on a few selected models due to availability and probability of good performance.

However, the results prove that almost all networks had comparable average accuracies. Hence, the following recommendations are made for future work:

1. Dataset extension through the inclusion of new images taken on various Runways may be studied to improve robustness and generalization.
2. A combination of Chinese and the new datasets may be studied for performance and more generalization of the models.
3. Mobile Net may be used as a base model for its application in cost-effective solutions. Furthermore, other models like Xception, InceptionV4, Squeeze Net, etc may be used to get even better results for FOD classification.

Availability of data and material

The FOD material dataset can be downloaded from Google Drive.

https://drive.google.com/drive/folders/1XA73Gr4d6kYnCc2_PZrr6GHWzzSPpZ_S?usp=sharing

Competing interests

The authors declare that they have no conflict of interest.

REFERENCES

- [1] H. Zhang, "The current status and inspiration of FOD industry in international civil aviation," *Civil Aviation Manage.*, vol. 295, pp. 58–61, Jan. 2015.
- [2] Y. H. Bhosale and K. S. Patnaik, "Application of deep learning techniques in diagnosis of COVID-19 (Coronavirus): A systematic review," *Neural Process. Lett.*, to be published.
- [3] Y. H. Bhosale and K. Sridhar Patnaik, "IoT deployable lightweight deep learning application for COVID-19 detection with lung diseases using RaspberryPi," in *Proc. Int. Conf. IoT Blockchain Technol. (ICIBT)*, May 2022, pp. 1–6.
- [4] Y. H. Bhosale, S. Zanwar, Z. Ahmed, M. Nakrani, D. Bhuyar, and U. Shinde, "Deep convolutional neural network based covid-19 classification from radiology X-ray images for IoT enabled devices," in *Proc. 8th Int. Conf. Adv. Comput. Commun. Syst. (ICACCS)*, Mar. 2022, pp. 1398–1402.
- [5] A. Shukla, G. Kalnoor, A. Kumar, N. Yuvaraj, R. Manikandan, and M. Ramkumar, "Improved recognition rate of different material category using convolutional neural networks," *Mater. Today, Proc.*, to be published.
- [6] M. Jehle, C. Sommer, and B. Jähne, "Learning of optimal illumination for material classification," in *Proc. Joint Pattern Recognit. Symp.* Cham, Switzerland: Springer, 2010, pp. 563–572.
- [7] E. Hayman, B. Caputo, M. Fritz, and J.-O. Eklundh, "On the significance of real-world conditions for material classification," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2004, pp. 253–266.
- [8] C. Liu, L. Sharan, E. H. Adelson, and R. Rosenholtz, "Exploring features in a Bayesian framework for material recognition," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 239–246.
- [9] J. J. Koenderink, A. J. Van Doorn, K. J. Dana, and S. Nayar, "Bidirectional reflection distribution function of thoroughly pitted surfaces," *Int. J. Comput. Vis.*, vol. 31, no. 2, pp. 129–144, 1999.
- [10] P. C. Ng and S. Henikoff, "Sift: Predicting amino acid changes that affect protein function," *Nucl. Acids Res.*, vol. 31, no. 13, pp. 3812–3814, 2003.
- [11] S. K. G. Manikonda and D. N. Gaonkar, "Islanding detection method based on image classification technique using histogram of oriented gradient features," *IET Gener., Transm. Distrib.*, vol. 14, no. 14, pp. 2790–2799, 2020.
- [12] M. Ghahremani, Y. Zhao, B. Tiddeman, and Y. Liu, "Interwoven texture-based description of interest points in images," *Pattern Recognit.*, vol. 113, May 2021, Art. no. 107821.
- [13] R. Szeliski, *Computer Vision: Algorithms and Applications*. Cham, Switzerland: Springer, 2010.

- [14] S. Bell, P. Upchurch, N. Snaveley, and K. Bala, "Material recognition in the wild with the materials in context database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3479–3487.
- [15] FAA. (2010). *Airport Foreign Object Debris (FOD) Management*. [Online]. Available: https://www.faa.gov/documentLibrary/media/Advisory_Circular/AC_150_5210-24.pdf
- [16] H. Xu, Z. Han, S. Feng, H. Zhou, and Y. Fang, "Foreign object debris material recognition based on convolutional neural networks," *EURASIP J. Image Video Process.*, vol. 2018, no. 1, pp. 1–10, Dec. 2018.
- [17] Z.-D. Yuan, J.-Q. Li, Z.-N. Qiu, and Y. Zhang, "Research on FOD detection system of airport runway based on artificial intelligence," *J. Phys., Conf. Ser.*, vol. 1635, no. 1, Nov. 2020, Art. no. 012065.
- [18] X. Wang, Z. Lan, Z. Li, Y. Lin, and Q. Xiao, "A practical application of runway foreign object debris detection system at the airport," in *Proc. Int. Conf. Microw. Millim. Wave Technol. (ICMMT)*, May 2018, pp. 1–3.
- [19] W. Chen, "Runway foreign object detection using RGB," *Aeronaut. J.*, vol. 119, no. 1212, pp. 229–243, Feb. 2015.
- [20] B. Vogel, "An object lesson in finding FOD," *Jane's Airport Rev.*, vol. 20, no. 8, p. 2, 2007.
- [21] K. W. Adi, C. E. Widodo, A. Puji, and U. Sri, "Detection of foreign object debris (Fod) using convolutional neural network (CNN)," *J. Theor. Appl. Inf. Technol.*, vol. 100, no. 1, pp. 184–191, 2022.
- [22] Y. Liu, Y. Li, J. Liu, X. Peng, Y. Zhou, and Y. L. Murphey, "FOD detection using DenseNet with focal loss of object samples for airport runway," in *Proc. IEEE Symp. Ser. Comput. Intell. (SSCI)*, Nov. 2018, pp. 547–554.
- [23] H. Zhang, Z. Jiang, Q. Xiong, J. Wu, T. Yuan, G. Li, Y. Huang, and D. Ji, "Gathering effective information for real-time material recognition," *IEEE Access*, vol. 8, pp. 159511–159529, 2020.
- [24] M. Varma and A. Zisserman, "Classifying images of materials: Achieving viewpoint and illumination independence," in *Proc. Eur. Conf. Comput. Vis.*, Cham, Switzerland: Springer, 2002, pp. 255–271.
- [25] B. Caputo, E. Hayman, M. Fritz, and J.-O. Eklundh, "Classifying materials in the real world," *Image Vis. Comput.*, vol. 28, no. 1, pp. 150–163, 2010.
- [26] L. Sharan, R. Rosenholtz, and E. H. Adelson, "Accuracy and speed of material categorization in real-world images," *J. Vis.*, vol. 14, no. 9, p. 12, 2014.
- [27] T. Munyer, P.-C. Huang, C. Huang, and X. Zhong, "FOD—A: A dataset for foreign object debris in airports," 2022, *arXiv:2110.03072*.
- [28] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, "A survey of the recent architectures of deep convolutional neural networks," *Artif. Intell. Rev.*, vol. 53, pp. 5455–5516, Apr. 2020.
- [29] X. Shang, Y. Xu, L. Qi, A. Hirpa Madessa, and J. Dong, "An evaluation of convolutional neural networks on material recognition," in *Proc. IEEE SmartWorld, Ubiquitous Intell. Comput., Adv. Trusted Comput., Scalable Comput. Commun., Cloud Big Data Comput., Internet People Smart City Innov. (SmartWorld/SCALCOM/UIC/ATC/CBDCOM/IOP/SCI)*, Aug. 2017, pp. 1–6.
- [30] M. Yaqub, J. Feng, M. Zia, K. Arshid, K. Jia, Z. Rehman, and A. Mehmood, "State-of-the-art CNN optimizer for brain tumor segmentation in magnetic resonance images," *Brain Sci.*, vol. 10, no. 7, p. 427, Jul. 2020.
- [31] A. K. Tushar, A. Ashiqzaman, and M. R. Islam, "Faster convergence and reduction of overfitting in numerical hand sign recognition using DCNN," in *Proc. IEEE Region Humanitarian Technol. Conf. (R10-HTC)*, Dec. 2017, pp. 638–641.
- [32] S. K. Noon, M. Amjad, M. A. Qureshi, and A. Mannan, "Overfitting mitigation analysis in deep learning models for plant leaf disease recognition," in *Proc. IEEE 23rd Int. Multitopic Conf. (INMIC)*, Nov. 2020, pp. 1–5.
- [33] A. Canziani, A. Paszke, and E. Culurciello, "An analysis of deep neural network models for practical applications," 2016, *arXiv:1605.07678*.
- [34] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 1–9.



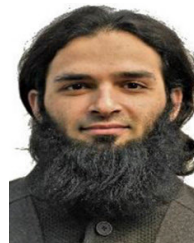
SYEDA MAHRUKH ZAINAB was born in Rawalpindi, Pakistan, in July 1984. She received the bachelor's degree in avionics engineering from the College of Aeronautical Engineering, NUST, Risalpur, and the master's degree in avionics engineering, with a focus on image processing from Air University, Islamabad. Her interests include computer vision, artificial intelligence, neural networks, wireless communications, and aviation related domains.



KHURRAM KHAN received the B.S. degree in computer engineering from COMSATS University Pakistan, in 2007, and the M.S. and Ph.D. degrees in electronic and communication engineering from Hanyang University, South Korea, in 2010 and 2019, respectively. Currently, he is with the Department of Avionics Engineering, Air University, Islamabad, Pakistan. His research interests include computer vision, pattern recognition, and machine/deep learning.



ADNAN FAZIL received the B.E. degree in avionic engineering (Hons.) from the College of Aeronautical Engineering (CAE), National University of Sciences and Technology (NUST), Pakistan, the M.S. degree in avionics engineering, with a specialization in communications, and the Ph.D. degree in wireless communications from Air University, Islamabad, Pakistan. He has 14 years of experience in academics and research. Currently, he is an Assistant Professor with the Department of Avionics Engineering, Air University. His research interests include wireless Ad Hoc networks, transmission scheduling in wireless networks, signal processing, and applications of machine/deep learning.



MUHAMMAD ZAKWAN received the B.S. degree in electrical engineering from Air University, Pakistan, in 2007, the M.S. degree in communication and information systems from Northwestern Polytechnic University, China, in 2010, and the Ph.D. degree in electrical and electronics engineering from Koç University, Turkey, in 2017. Since 2017, he has been an Assistant Professor with Air University. His research interests include silicon photonic integrated circuits (Si-PIC), communication systems, artificial intelligence, mobile ad-hoc networks, and sensor networks.

• • •