Xu et al. EURASIP Journal on Image and Video Processing (2018) 2018:21
https://doi.org/10.1186/s13640-018-0261-2

EURASIP Journal on Image and Video Processing

RESEARCH Open Access

Foreign object debris material recognition based on convolutional neural networks

Haoyu Xu[1,2], Zhenqi Han[1,3*], Songlin Feng[1], Han Zhou[1] and Yuchun Fang[3]

Abstract

The material attributes of foreign object debris (FOD) are the most crucial factors to understand the level of damage sustained by an aircraft. However, the prevalent FOD detection systems lack an effective method for automatic material recognition. This paper proposes a novel FOD material recognition approach based on both transfer learning and a mainstream deep convolutional neural network (D-CNN) model. To this end, we create an FOD image dataset consisting of images from the runways of Shanghai Hongqiao International Airport and the campus of our research institute. We optimize the architecture of the D-CNN by considering the characteristics of the material distribution of the FOD. The results show that the proposed approach can improve the accuracy of material recognition by 39.6% over the state-of-the-art method. The work here will help enhance the intelligence capability of future FOD detection systems and encourage other practical applications of material recognition technology.

Keywords: Foreign object debris, Material recognition, Deep learning, Deep convolutional neural networks, Transfer learning

1 Introduction

Foreign object debris (FOD) refers to any object located in and around an airport (especially on the runway and the taxiway) that can damage the aircraft or harm air-carrier personnel [1]. Typical examples of FOD include twisted metal strips, components detached from aircraft or vehicles, concrete chunks from the runway, and plastic products. FOD poses a safety risk to an aircraft and a significant economic loss to airlines. The crash of Air France Flight 4590 that killed 113 personnel in 2000 was caused by a twisted metal strip [2], as shown in Fig. 1. Moreover, the direct economic loss due to FOD damage is conservatively estimated to be 3 ~ 4 billion USD per year [3].

To reduce or eliminate FOD damages, certain companies have developed FOD detection systems, such as the Tarsier system by QinetiQ, FODetect by Xsight, and iFerret by Stratech [4]. All these systems use a camera to take a photograph of suspicious FOD, and then, the photographs are verified by human experts. These systems have been commercially deployed in a few airports but have not achieved large-scale global usage. One main reason for this low-level deployment is that the final FOD verification step relies exclusively on recognition by a human expert, which has two disadvantages. The first disadvantage is that reliable verification requires a capable and experienced official, which incurs additional cost for the airport authority. For example, the Vancouver Airport filled this position with an employee from its FOD vendor. The second disadvantage is that people's recognition capability is not completely trustworthy because they are inevitably fatigued from time to time.

Han et al. [5, 6] worked on FOD object recognition using a support vector machine (SVM) and random forest. FOD object recognition is to identify what the FOD is. Unfortunately, the exact nature of FOD is varied because FOD can be composed of any object, any color and any size. Over 60% of the FOD items are made of metal. Therefore, recognition of the FOD material constitution has much greater practical significance than object recognition.

Material recognition is a fundamental problem in computer vision. In contrast with the several decades of object recognition research, material recognition has only begun receiving attention in recent years. It is a flourishing and challenging field. The approaches to

* Correspondence: hanzq@sari.ac.cn
[1]Shanghai Advanced Research Institute, Chinese Academy of Sciences, Shanghai 201210, China
[3]School of Computer Engineering and Sciences, Shanghai University, Shanghai 200444, China
Full list of author information is available at the end of the article

Fig. 1 The twisted metal strip that caused Air France Flight 4590 to crash

material recognition can be broadly categorized as hand-crafted or automatic feature extraction. Hand-crafted approaches can be further divided into surface reflectance [7–11], 3D texture [12–19], and feature fusion [20–22] approaches. Automatic feature extraction approaches refer to those that involve acquiring image features using a deep convolutional neural network (D-CNN) [23–26].

There are some correlations between the surface reflectance properties and the categories of materials. For example, wooden surfaces tend to be brown, whereas metallic ones tend to be shiny. However, different classes of materials may exhibit similar reflectance properties, such as the translucence of plastic, wax, and glass. Consequently, understanding an object's reflectance properties is insufficient to determine its material constitution. Similarly, different materials may have the same 3D texture pattern, as shown in Fig. 2. To overcome these challenges, researchers have attempted to combine different features or have attempted automatic feature extraction to perform material recognition tasks. Some remarkable results were obtained on some specific datasets in certain studies.

However, past research results remain inadequate to meet the demands of FOD material recognition. First, there is no specific FOD dataset for the task because of the unique airport environment. Although Bell et al. [25] used more than 300 million image patches for training, the images were acquired mainly in indoor environments where light conditions are quite different from the FOD emergence locations. The results were hence quite poor when these 300 million image patches were used for training while FOD images were used for testing (please refer to the "Section 4" for details). Second, a high-recognition ratio is necessary for metal recognition. Metallic objects are far more harmful than other materials. Meanwhile, 60% of FOD is constituted by metal [1]. However, according to prior results [19, 21], the recognition rate was quite low for metallic objects.

Fig. 2 Surfaces with similar textures may be composed of different materials. These objects are made of fabric, plastic, and paper, from left to right [21]

This paper proposes a novel FOD material recognition approach based on transfer learning and a mainstream deep convolutional neural network (D-CNN) model. This paper describes an FOD image dataset consisting of images taken on the runways of Shanghai Hongqiao International Airport and the campus of our research institute. The dataset consisted of 3470 images divided into three categories by material: metal, concrete, and plastic. The proposed approach is optimized to recognize metal because of the high risk that is due to its high-damage level to aircrafts and its high occurrence frequency in airports.

This research will help improve the intelligence capability, the ease of using, and the user experience of FOD detection systems. It will also encourage more applications of material recognition systems, especially in security and manufacturing, such as construction site management [27, 28].

The rest of this paper is organized as follows: Section 2 introduces related work, and our approach is described in Section 3. Section 4 presents a discussion of the experiment results, and Section 5 summarizes our conclusion and plan for future work.

2 Related work

Material recognition, a fundamental problem in computer vision, has a wide range of applications. For example, an autonomous vehicle or a mobile robot can make decisions on whether a forthcoming terrain is asphalt, gravel, ice, or grass. A cleaning robot can distinguish among wood, tile, or carpet. The approaches to material recognition are broadly divided into two categories according to feature extraction methods: hand-crafted features and automatic features. Hand-crafted approaches can be further divided into surface reflectance-based, 3D texture-based, and feature fusion-based approaches. Automatic feature extraction approaches refer to those acquiring image features through a D-CNN.

The most popular formalization for model surface reflectance is the bidirectional reflectance distribution function (BRDF). This function defines the amount of light reflected at a given point on a surface for any combination of incidence and reflection angles [21]. The BRDF has a parametric type [29, 30] and an empirical type [7–9, 11]. Parametric BRDF models cannot acquire a broad set of real-world reflectance properties. In

contrast, empirical BRDF models always require prior knowledge, such as illumination conditions, geometry, and surface material properties. Such prior knowledge cannot be expected to be available for real-world images. Zhang et al. [10] introduced an empirical model based on a reflectance disk and reflectance hashing. The reflectance disk, a measurement of the surface property, was built using a customized camera apparatus. Gaussian low-pass filters, Laplacian filters, and gradient filters were applied to the reflectance disk. Textons, referring to fundamental micro-structures in natural images, were computed by k-means clustering on the output of the filter banks. Following this approach, texton boosting and reflectance hashing were employed for feature selection and image classification. This approach is not feasible for real-world images, as reflectance disks are generated by a customized apparatus in a laboratory environment. Moreover, different surface materials may exhibit similar reflectance phenomena: for example, plastic, glass, and wax are translucent. Therefore, fulfilling the goal of material recognition only by using surface reflectance properties appears to be difficult.

Three-dimensional texture refers to surface roughness that can be resolved by the human eye or a camera. Such texture-based approaches follow the feature extraction-and-classification routine. Various researchers used a number of descriptors to extract the local features of an image. For example, some studies [12–17] used the maximum response, one study [18] used sorted random projections, and another study [19] applied a kernel descriptor for this purpose. These feature vectors were then fed into a classifier, usually SVM, latent Dirichlet allocation (LDA), or nearest neighbor. These approaches were designed to obtain salient results on CUReT [15–17], ETHTIPS [12–14], and FMD [20]. However, these datasets are inappropriate for FOD material recognition tasks. The images from CUReT and ETHTIPS datasets were captured using a customized apparatus in an ideal laboratory environment. These images not only had different appearance with real-world images but also were unobtainable in daily life. The FMD dataset is composed of real-world images from the website Flickr. However, the FMD dataset suffers three downsides with regard to FOD recognition: (1) Few samples are FOD alike. (2) The photos are barely taken outdoors. (3) There is a lack of intentional collection of images of metal, concrete, and plastic materials.

Sharan et al. [20–22] combined reflectance and 3D texture into new fused features as input to an LDA or an SVM classifier. They chose four groups of features, namely, color and texture (e.g., Color, Jet, and SIFT), micro-texture (e.g., Micro-Jet and Micro-SIFT), shape (e.g., curvature), and reflectance (Edge-Slice and Edge-Ribbon). As the previous work, this research was also performed on an FMD dataset that made it unfeasible for FOD material recognition tasks.

Since Hinton's monumental work [31] in 2006, deep learning has received considerable attention in both academia and industry because of its superior performance over other machine learning methods. He et al. [32] used a 152-layer D-CNN to obtain a 3.57% error rate on the ILSVRS2015 dataset. The result was better than the error of 5.1% incurred by humans [33]. Researchers have attempted to apply D-CNNs to automatically extract image features to achieve material recognition. Cimpoi et al. [23, 24] proposed the Fisher-vector CNN via amelioration of the pooling layer in the D-CNN. They reported a considerable amount of improvement over the work by Sharen et al. [21, 22]. Bell et al. [25] proposed a new dataset, the Material-in-context Database (MINC), for material recognition based on Imagenet [34]. They achieved an impressive recognition accuracy of 85.2% by utilizing Alexnet [35] and GoogLeNet models [36]. Zhang et al. [26] assumed that the features for object recognition could be helpful for material recognition to some extent and integrated features learned from the ILSVRC2012 [33] and the MINC [25]. The results were state-of-the-art, as expected. However, the MINC dataset was built using images taken from indoor environments, which are unsuitable for FOD material recognition.

3 Method

3.1 Dataset construction

The Columbia–Utrecht Reflectance and Texture Database (CUReT) [15, 17], the KTH-TIPS [13], the Flickr Material Database (FMD) [22], and the Material-in-context Database (MINC) [25] are open datasets for material recognition. CUReT consists of 61 textures imaged under 205 diverse illumination and angle conditions. KTH-TIPS has 11 material categories, each category with four samples imaged under various conditions. Images in the FMD dataset are from the Flicker website. This dataset has 1000 images of 10 material categories. MINC has approximately 300 million image patches tailored from ImageNet. As stated in Section 2, these four datasets are improper for FOD material recognition.

We choose metal, plastic, and concrete as three typical FOD materials to construct the dataset. According to FAA's AC 150/5220-24 [1], these materials appear most frequently on runways and taxiways. Furthermore, metallic FOD constitutes approximately 60% of all FOD. Metal and plastic may exhibit similarly intense reflectance phenomena under strong light in outdoor environments. To complicate things further, these images are taken on runways or taxiways, which are made of concrete. The concrete background of metal or plastic images poses a tremendous challenge to distinguishing these two materials from the background. It is similar with detecting Uyghur language text from complex backgrounds [37]. Therefore, careful treatment is imperative to recognize metal, plastic, and concrete correctly.

We constructed the FOD dataset using two approaches. The first approach involved taking the images in the backup runway of Shanghai Hongqiao International Airport. It appears to be impossible for us to collect images at a large scale because of the airport's strict entry requirements. Hundreds of images taken from Hongqiao Airport were used as test data. The other approach to construct the dataset was to collect images at the campus of our institute. We chose a campus road made of concrete to emulate as closely as possible the runway environment. We used a HikVision DS-2df8523iw-a camera with a maximum focal distance of 135 mm and a maximum resolution of 1280×768. The camera was mounted two meters from the ground, and the FOD was located approximately five meters from the camera. This setting was proportionally in accordance with the typical setup of a FOD detection system. The metallic FOD used includes wrenches, screws, nuts, metal strips, rusty nails, and iron sheets of various shapes and sizes. For plastic FOD, diverse shapes and sizes of polyethylene plastic pipes, bags, and straws were chosen. We chose different shapes and sizes of concrete blocks, stone blocks, and pebbles as samples of concrete. Images of each material category were taken from 9 a.m. to 5 p.m. on separate days to ensure different illumination circumstances. We also captured images from different angles. Figure 3 shows the typical image samples for the metallic, plastic, and concrete material categories from top to bottom. The original image sample was divided into $N \times N$ size patches by hand. The patches were further resized to 256×256 for convenient processing in Caffe.

Table 1 summarizes the statistics of the FOD dataset. The images taken in indoor environments were only used to compare our system with the prevalent D-CNN models to gauge their different performances for indoor and outdoor objects. These images were not used in either the training or the testing of our proposed D-CNN model.

The FOD dataset introduced in this paper was different in three aspects from previous datasets. First, there was a significant extent of intra-class variations for each material category. Images belonging to the same material category usually had completely different shapes or even identities. Second, all images had concrete as the background, emulating the circumstances in airports. Third, all images were captured in outdoor environments.

Fig. 3 Typical image samples of the FOD dataset

Table 1 Statistics of the FOD dataset

	Indoor	Campus road training	Airport runway and campus road testing
Metal	30	1000	105
Plastic	0	1000	235
Concrete	0	1000	100

3.2 Choice of the D-CNN model

A D-CNN, an extremely efficient and automatic feature-learning approach, transforms an original input to a higher-level and more abstract representation using non-linear models. A D-CNN is composed of multiple convolution layers, pooling layers, fully connected layers, and classification layers. The network parameters are optimized through the back-propagation algorithm.

D-CNNs have a broad set of applications in image classification, object recognition, and detection. Glasssix [38] trained a CNN with an improved ResNet34 layer and obtained 99.83% accuracy on the famous LFW face recognition database. Considering the scale, context, sampling, and deep combined convolutional networks, the BDTA team won the championship of the ILSVRC2017 object detection task. The Subbmission4 model provided by BDTA can detect 85 object categories and achieved a 0.73 mean average precision on DET task 1a (object detection with provided training data) [39]. Chen et al. provided an effective CNN named Dual Path Networks for object localization and object classification, which obtained a 6.2% localization error rate and a 3.4% classification error rate on the ILSVRC2017 object localization task [40]. Yan et al. provided a supervised hash coding with deep neural network for environment perception of intelligent vehicles, and the proposed method can obviously improve the search accuracy [41].

AlexNet [35], GoogLeNet [36], and VGG-16 [42] are widely established and open D-CNN models. They are de-facto candidate base models for researchers. AlexNet is composed of five convolutional layers and three fully connected layers, VGG-16 is composed of 13 convolutional layers and three fully connected layers, and GoogLeNet is composed of 21 convolutional layers and one fully connected layer. The main size of the convolutional kernel for AlexNet and VGG-16 is three by three, whereas that of GoogLeNet is the inception module, which is a two-layer convolutional network. Both AlexNet and VGG-16 use the maximum pooling mechanism. By contrast, GoogLeNet applies both the

maximum pooling and the average pooling schemes. We describe three D-CNN models in more detail in Table 2. The following choices were made with the help of experimental results. AlexNet was chosen as the base mode in our approach, as it yields the best performance for metallic FOD. Regardless of the model employed, metallic FOD is easily confused with plastic or concrete FOD. This observation verifies our choice of material categories, in addition to the reason that they occur most frequently on runways or taxiways. Please refer to Section 4 for detailed results.

With the increase of network depth, the recognition accuracies of VGG and GoogLeNet on outdoor metal images shown in Fig. 8 are reduced. We conjecture that ResNet [32] may have a low accuracy rate for FOD images from an outdoor environment. Thus, we did not perform experiments using ResNet on the FOD material dataset.

Table 2 Detailed descriptions of AlexNet, VGG-16, and GoogLeNet

	Alexnet	VGG16	GoogLeNet	Improved AlexNet
Input (RGB image)	227*227	224*224	224*224	227*227
Convolution (kernel size/stride)	11*11/4	3*3/1	7*7/2	11*11/4
		3*3/1		
Max. pooling	3*3/2	2*2/2	3*3/2	3*3/2
Convolution (kernel size/stride)	5*5/1	3*3/1	3*3/1	5*5/1
		3*3/1		
Max. pooling	3*3/2	2*2/2	3*3/2	3*3/2
Convolution (kernel size/stride)	3*3/1	3*3/1	Inception(3a)	3*3/1
	3*3/1	3*3/1	Inception(3b)	3*3/1
	3*3/1	1*1/1		3*3/1
Max. pooling	3*3/2	2*2/2	3*3/2	3*3/2
Convolution (kernel size/stride)		3*3/1	Inception(4a)	
		3*3/1	Inception(4b)	
		1*1/1	Inception(4c)	
			Inception(4d)	
			Inception(4e)	
Max. pooling		2*2/2	3*3/2	
Convolution (kernel size/stride)		3*3/1	Inception(5a)	
		3*3/1	Inception(5b)	
		1*1/1		
Pooling		Max. pool 2*2/2	Average pool 7*7/1	
Linear	FC-4096	FC-4096	FC-1000	FC-4096
	FC-4096	FC-4096		FC-4096
	FC-1000	FC-1000		FC-1000
				FC-3
Output	Softmax	Softmax	Softmax	Softmax

3.3 Transfer learning

The technique of transfer learning is applied in this paper to avoid the overfitting problem. Transfer learning is literally defined as the transfer of knowledge learned in one domain to another domain. The technique is especially useful for the D-CNN models because of their high demand in terms of the huge amount of human-labeled training data [43, 44]. Without sufficient training data, the D-CNN models tend to be over-fitted. It would be truly favorable to reduce the need and effort to collect, clean, and label a large amount of data with the help of transfer learning.

In this paper, the parameters of the improved AlexNet model are initialized by those trained from MINC. This model continues to be trained by fine-tuning the weights of all layers based on the FOD dataset discussed in Section 4. It is observed that earlier layers' features of a D-CNN entail more generic features (e.g., edge detectors or color blob detectors) that are reusable for many tasks [45]. In addition, later layers of the D-CNN contain details more specific in the original dataset, e.g., MINC. The weights of later layers should be optimized more than the ones of earlier layers with the help of the new dataset, e.g., the FOD dataset. Therefore, the dedicated choice of weights' initialization is equivalent to shortening the distance from the starting point to the optimum, which helps avoid the overfitting problem.

Transfer learning has achieved a wide range of applications in many tasks. In the recognition task, Reyes et al. pre-trained a CNN using 1.8 million images and used a fine-tuning strategy to transfer learned recognition capabilities from the general domains to the specific challenge of the Plant Identification task [46]. Bell et al. trained all of their CNNs for material recognition by fine-tuning the network starting from the weights obtained on 1.2 million images from ImageNet (ILSVRC 2012) [25]. In object detection, OverFeat [47], the winner of the location task of ILSVRC2013, also used transfer learning. Google DeepMind used transfer learning to solve complex sequences of tasks [48].

3.4 Improved D-CNN based on AlexNet

Inspired by transfer learning, an improved D-CNN based on AlexNet is described in this section. The improved

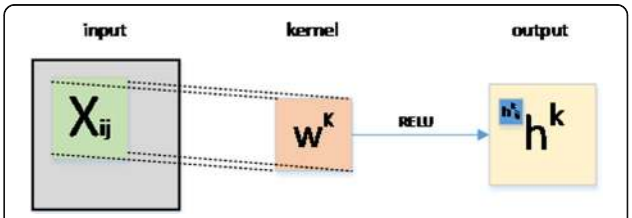

Fig. 4 Image convolution

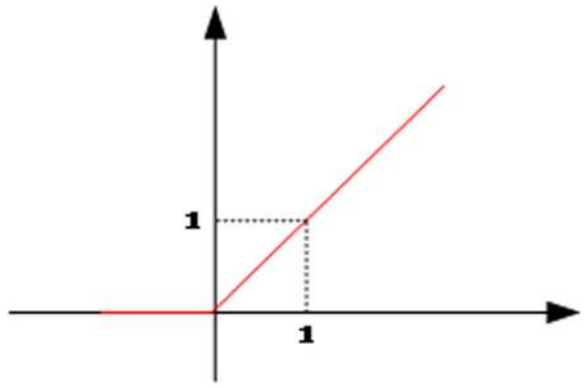

Fig. 5 RELU function

D-CNN model has an additional fully connected layer appended to the model as the last layer. It shares the first eight layers with AlexNet; hence, the model consists of five convolution layers and four fully connected layers. The ninth layer has three neuronal nodes, indicating that the network has three material tag outputs. We use softmax loss as the classifier. The detailed network structure is shown in Table 2. The experiments were conducted on the Caffe framework with Nvidia Tesla K20 GPU card. Caffe [49] is a highly effective framework for deep learning, such as Yan et al.'s framework [50–53] for HEVC coding unit. Using the FOD training dataset, we fine-tuned the improved D-CNN based on pre-training the weights in MINC. Our implementation for FOD material recognition follows the practice in Krizhevsky's and He's papers[32, 35]. During training, the inputs to the improved D-CNN were fixed-size 224×224 RGB images. The batch was set to 256, the momentum was set to 0.9, the weight decay (the L2 penalty multiplier) was set to 0.5, and the learning rate was set to 0.001. In total, the learning rate was decreased three times, and the learning was stopped after 20-K iterations. The FOD testing dataset was used for the FOD material recognition test after the fine-tuning stage. All of our experiment base above hyperparameters achieved state-of-the-art results.

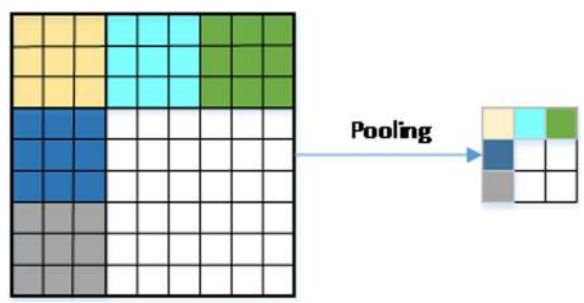

Fig. 6 Pooling

Fig. 7 Samples from the FOD verification dataset

The convolution layer learns input features through the convolution of the kernel and the input vectors, as shown in Fig. 4. The convolution function is given by Eq. (1):

$$h_{ij}^{k} = \mathrm{RELU}\left(\left(W^{k} * X\right)_{ij} + b^{k}\right) \tag{1}$$

In Eq. (1), W is the weight of the convolution kernel, and k indicates the number of convolution kernels. h_{ij}^{k} is the output of the convolution kernel k in the output layer (i, j), X is the input of the convolution layer, and b is the offset. RELU is an activation function $f(x) = \max(0, x)$, which is zero when $x < 0$ and is linear with slope 1 when $x > 0$, as shown in Fig. 5. Compared to the sigmoid and tanh functions, RELU can greatly accelerate the convergence of stochastic gradient descent [35].

The pooling layer is a down-sampling process that reduces the computational complexity and retains the rotational invariance of images—see Fig. 6. Mean pooling and max pooling are usually used in the pooling layer and involve averaging the image area and choosing the maximum value, respectively.

In the output layer of the network, the softmax classifier is applied to classify samples. The class tag of the

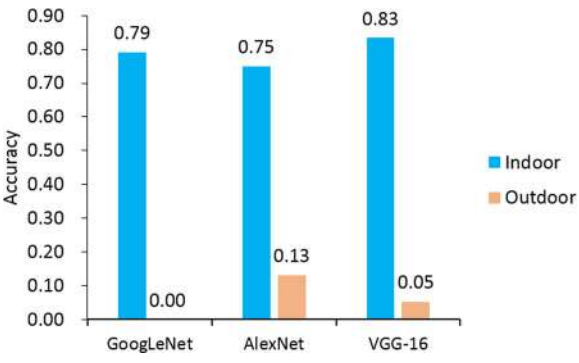

Fig. 8 Indoor and outdoor image testing results for GoogLeNet, AlexNet, and VGG-16 for metal

Table 3 Outdoor metal image test results for AlexNet, VGG-16, and GoogLeNet trained on the MINC dataset

	Carpet	Ceramic	Foliage	Leather	Metal	Paper	Plastic	Stone	Water
AlexNet	0.053	0.013	0	0.026	0.132	0.132	0.092	0.289	0.263
VGG-16	0	0	0	0	0.053	0.053	0.447	0.355	0.092
GoogLeNet	0	0	0.013	0	0	0.066	0.605	0.276	0.039

maximum possibility is chosen as the output result. The process of the output layer can be described as Eq. (2):

$$\hat{y} = \arg\max_i \left(\frac{e^{z_i}}{\sum_{j=1}^{N} e^{z_j}} \right) \tag{2}$$

where z is the activation value of last layer neurons. The activation value of neuron i is z_i. N is the number of categories and the number of last layer neurons. $\hat{y}$ is the prediction tag. For example, the FOD dataset has three class of images, so $N = 3$, in which 1 denotes metal, 2 denotes plastic, and 3 denotes concrete.

4 Experimental results and discussion

4.1 Improved D-CNN based on AlexNet

We chose the base model from AlexNet, VGG-16, and GoogLeNet. All of these models were trained on the MINC dataset. The model with the best recognition accuracy for metal was chosen as the base transfer learning model. The FOD verification dataset consisted of 24 indoor images (indoor images were included for performance comparison, although there was no indoor case for FOD detection) and 76 outdoor images, as shown in Fig. 7. The first row is the indoor images of the metal items, and the second row consists of the outdoor images.

Figure 8 shows the accuracy of FOD material recognition for the three models in the indoor and outdoor cases. All three convolution networks trained by the MINC dataset recorded an approximately 80% accuracy for indoor images but yielded poor performance for the outdoor case: AlexNet had an accuracy of 13%, VGG-16 had an accuracy of 5%, and GoogLeNet's accuracy was close to 0%! The unsatisfactory performance of the above three D-CNNs was mainly because the MINC dataset was built using images from Houzz, which collects interior decoration images. These pictures are usually taken under soft illumination and without strong reflectance. By contrast, FOD, on runways or taxiways, often experiences strong illumination, heavy rain, and dense fog. It is obvious that these three D-CNN models trained on the MINC dataset are not directly applicable to FOD material recognition. AlexNet has the best metallic material recognition.

Table 3 displays the material classification results for metal for GoogLeNet, AlexNet, and VGG-16 trained on the MINC dataset. The results show that GoogLeNet misclassified most metals as plastic or stone, AlexNet misclassified metal as stone, water, paper, and plastic, and VGG-16 misclassified metal as plastic and stone. This set of results show that metallic FOD tends to be easily misclassified as plastic and concrete. This observation justifies our choice of material categories.

4.2 Results of the improved model

In this section, we compared the performance of the three D-CNN models. The first was AlexNet with parameters trained by MINC, abbreviated as AM. The second was AlexNet with parameters trained by both the MINC and FOD datasets—this model was called

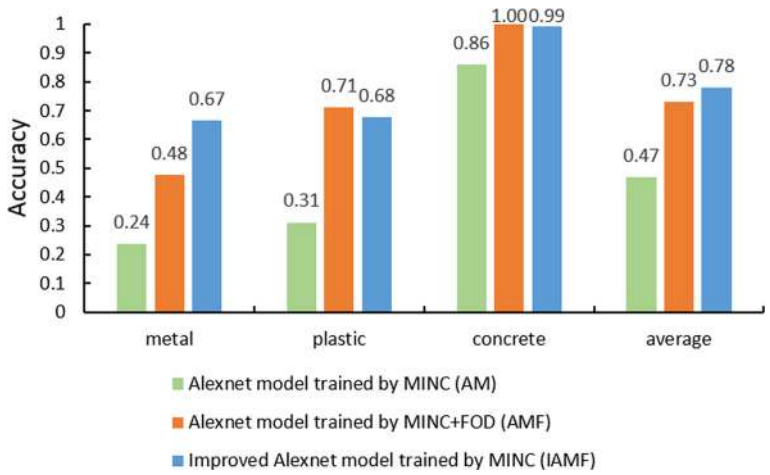

Fig. 9 FOD material recognition results

Table 4 Confusion matrix of the IAMF

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

AMF. The third model was the improved model shown in Table 2 with parameters trained by the MINC and FOD datasets, abbreviated as IAMF. All experiments were conducted within the framework of Caffe.

The dataset description is given in Table 1. To guarantee that the testing dataset was comparable to practical situations, the items in the FOD testing dataset must meet the following criteria: All samples in the FOD testing dataset should have been collected at the airport or the institutional campus. The testing samples did not overlap with those used for training. Furthermore, samples had various appearances within the same material category. Please refer to Fig. 3 for the testing samples.

Figure 9 shows that for the metal recognition. The IAMF model enhanced the recognition accuracy by 19% over AMF and by 42.86% over AM for metal recognition. The results prove the effectiveness of the IAMF model and the importance of the FOD dataset. However, we also noted that considerable room for improvement exists because the best accuracy value was only 67%. For plastic and concrete material recognition, the improved model yielded similar performance to that of AMF.

Table 4 shows the confusion matrix of the IAMF model. It is obvious that metallic FOD was most easily misclassified as concrete objects. We inferred that the concrete runway and taxiway might act as recognition noise because they were the backgrounds in the images. For the case of plastic FOD, metal and plastic were easily confused for each other. We inferred that metals and plastics have similar characteristics under the condition of strong light illumination.

We further examined the deep features of these three D-CNN models to understand the reasons for the performance differences. Deep feature, the degree of a neuron's excitation, is defined as the input value to the

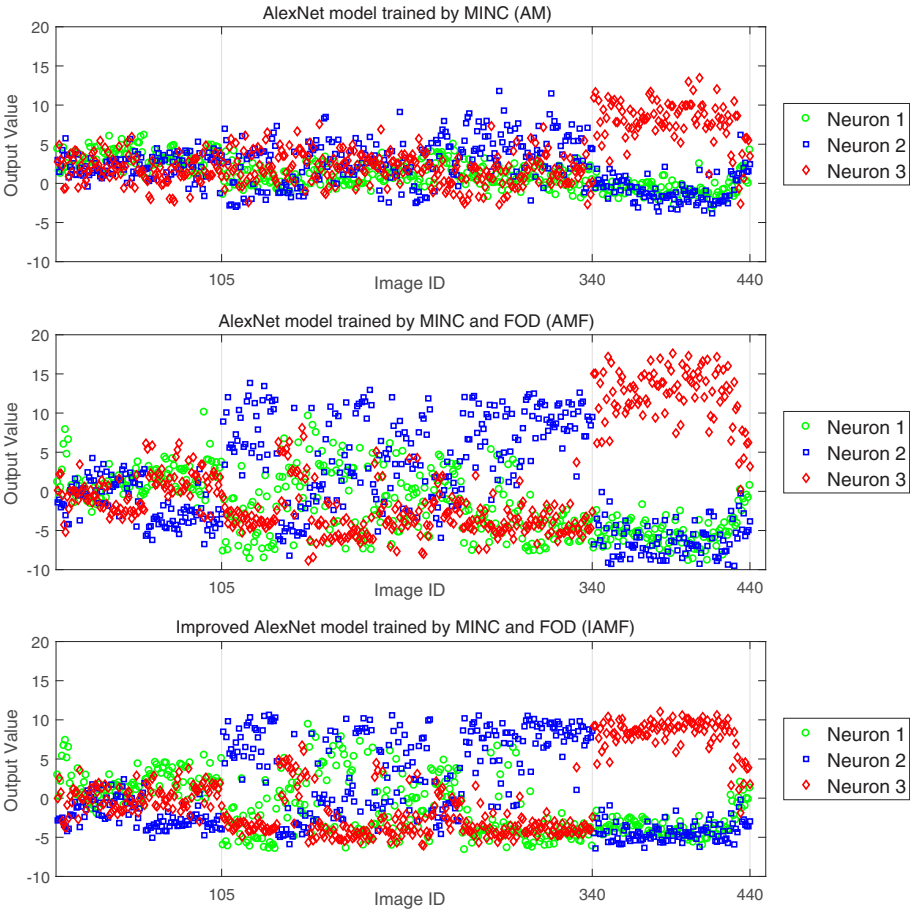

Fig. 10 The neural output value distributions of three convolutional neural networks

last softmax layer. Neuronal excitation values were taken from layer 8 of the AM, layer 8 of the AMF, and layer 9 of the IAMF over the test images. Figure 10 compares the deep features of the three D-CNN models over the test images, where the horizontal axis means their image IDs, and the vertical axis denotes the value of deep features. From top to bottom, three rows of Fig. 10 demonstrate the results for AM, AMF, and IAMF, respectively. The metal image IDs are from No. 1 to No. 105. No 106 to 340 indicates the plastic image IDs and others are for the concrete image IDs.

We found that the ability to discriminate material was based on the degree of neuronal excitation. For example, a neuron might have been excited for metal but not for plastic. To judge the discrimination ability of the D-CNN model, we observed the distributions of the different neurons' excitations. The higher the value of a certain neuron's excitation compared to others, the stronger the ability to discriminate a certain material. For example, according to the red circles in Fig. 10, the values of Neuron 3 were better than the values of Neuron 1 and Neuron 2 for concrete images (image ID 341–440). Thus, the AFM model had stronger ability to discriminate concrete. According to the green circles in Fig. 10, compared with the AM model and the AMF model, Neuron 1 of the IAMF model had better excitation values than the other neurons for metal images (image ID 1–105). As a result, the IAMF model had stronger discrimination ability for metal than other models. Besides, the IAMF model had a more concentrated neuron excitation distribution, indicating that the IAMF model had a more stable discrimination ability for FOD material. Therefore, the discrimination abilities of the three D-CNN models for FOD gradually increased from the AM model to the IAMF model. The result also confirmed the effectiveness of the IAMF model.

5 Conclusions

FOD material recognition is a challenging and significant task that must be performed to ensure airport safety. The general material recognition dataset is not applicable to FOD material recognition. Therefore, a new FOD dataset was constructed in this study. The FOD dataset was different from previous material recognition datasets in that all training and testing samples were collected in outdoor environments, e.g., on a runway, on a taxiway, or on campus. We compared the performances of three well-known D-CNN models on the new dataset. The results were far from acceptable, especially for the recognition of metal, which accounts for 60% of all FOD. An improved D-CNN model was then introduced and compared with AlexNet. The new model achieved a 38.6% improvement over AlexNet in terms of the recognition of metal FOD.

We also inferred that concrete backgrounds can adversely affect the FOD material recognition performance, leading to the misclassification of metal or plastic as concrete. Therefore, our future work will investigate possible approaches to introduce image segmentation to distinguish metal and plastic from concrete. Other technologies, such as radar or infrared imaging, may be required for better recognition results.

Acknowledgements
Not applicable

Funding
This work was supported by the National Natural Science Foundation of China (No. 61170155).
This work was also supported by the Science & Technology Commission of Shanghai Municipality (No. 15DZ1100502).

Availability of data and materials
The FOD material dataset can be downloaded from Google Drive.
Link: https://drive.google.com/file/d/1UTxZQUipkX6_rC9AoCaweeeeOPrOG1_B/view?usp=sharing

Authors' contributions
HYX proposed the framework of this work and drafted the manuscript. ZQH designed the proposed algorithm and performed all the experiments. SLF and YCF offered useful suggestions and helped in modifying the manuscript. ZH helped in drafting the manuscript. All authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Publisher's Note

Author details
[1]Shanghai Advanced Research Institute, Chinese Academy of Sciences, Shanghai 201210, China. [2]University of Chinese Academy of Sciences, Beijing 100049, China. [3]School of Computer Engineering and Sciences, Shanghai University, Shanghai 200444, China.

Received: 6 April 2017 Accepted: 15 March 2018
Published online: 03 April 2018

References
1. M. J. O'Donnell, "Airport foreign object debris (FOD) detection equipment," FAA, AC (150/5220) vol. 24 (2009).
2. WIKI, Foreign object damage. https://en.wikipedia.org/wiki/Foreign_object_damage. Accessed Nov 2016.
3. CAAC Airport Division, et al., "FOD Prevention Manual," (2009)
4. H Zhang et al., The current status and inspiration of FOD industry in international civil aviation. Civ. Aviat. Manage. **295**, 58–61 (2015)
5. Z Han et al., *A novel FOD classification system based on visual features*, In International Conference on Image and Graphics, *LNCS* Vol. 9217 (2015), pp. 288–296. https://doi.org/10.1007/978-3-319-21978-3_26
6. Z Han, Y Fang, H Xu, *Fusion of low-level feature for FOD classification*, In 10th International Conference on Communications and Networking in China (2015), pp. 465–469. https://doi.org/10.1109/CHINACOM.2015.7497985
7. M Jehle, C Sommer, B Jähne, *Learning of optimal illumination for material classification*, In 32nd Annual Symposium of the German Association for Pattern Recognition, *LNCS* Vol. 6376 (2010), pp. 563–572. https://doi.org/10.1007/978-3-642-15986-2_57
8. C Liu, J Gu, Discriminative illumination: per-pixel classification of raw materials based on optimal projections of spectral BRDF. IEEE Trans. Pattern Anal. Mach. Intell. **36**(1), 86–98 (2014). https://doi.org/10.1109/TPAMI.2013.110

9. C Liu, G Yang, J Gu, *Learning discriminative illumination and filters for raw material classification with optimal projections of bidirectional texture functions*, In Computer Vision and Pattern Recognition (2013), pp. 1430–1437. https://doi.org/10.1109/CVPR.2013.188
10. H Zhang, K Dana, K Nishino, *Reflectance hashing for material recognition*, In Computer Vision and Pattern Recognition (2015), pp. 3071–3080. https://doi.org/10.1109/CVPR.2015.7298926
11. J Filip, P Somol, *Materials classification using sparse gray-scale bidirectional reflectance measurements*, International Conference on Computer Analysis of Images and Patterns, *LNCS*. Vol. 9257 (2015), pp. 289–299. https://doi.org/10.1007/978-3-319-23117-4
12. E Hayman, B Caputo, M Fritz, JO Eklundh, *On trhe significance of real-world conditions for material classification*, In European Conference on Computer Vision, *LNCS*. Vol. 3024 (2004), pp. 253–266. https://doi.org/10.1007/978-3-540-24673-2_21
13. B Caputo, E Hayman, M Fritz, JO Eklundh, Classifying materials in the real world. Image Vis. Comput. **28**(1), 150–163 (2010). https://doi.org/10.1016/j.imavis.2009.05.005
14. B Caputo, E Hayman, P Mallikarjuna, *Class-specific material categorization*, In 10th IEEE International Conference on Computer Vision, Proc. IEEE Int. Conf. Comput. Vision II (2005), pp. 1597–1604. https://doi.org/10.1109/ICCV.2005.54
15. M Varma, A Zisserman, *Classifying images of materials: achieving viewpoint and illumination independence*, In European Conference on Computer Vision, *LNCS* Vol. 2352 (2002), pp. 255–271. https://doi.org/10.1007/3-540-47977-5_17
16. M Varma, A Zisserman, A statistical approach to material classification using image patch exemplars. IEEE Trans. Pattern Anal. Mach. Intell. **31**(11), 2032–2047 (2009). https://doi.org/10.1109/TPAMI.2008.182
17. M Varma, A Zisserman, A statistical approach to texture classification from single images. Int. J. Comput. Vis. **62**(1-2), 61–81 (2005). https://doi.org/10.1023/B:VISI.0000046589.39864.ee
18. L Liu, PW Fieguth, D Hu, Y Wei, G Kuang, Fusing sorted random projections for robust texture and material classification. IEEE Trans. Circuits. Syst. Vid. Technol. **25**(3), 482–496 (2015). https://doi.org/10.1109/TCSVT.2014.2359098
19. D Hu, L Bo, *Toward robust material recognition for everyday objects*, In British Machine Vision Conference, Proc. BMVC (2011), pp. 1–11. https://doi.org/10.5244/C.25.48
20. C Liu, L Sharan, EH Adelson, R Rosenholtz, *Exploring features in a Bayesian framework for material recognition*, In 2010 IEEE Conference on Computer Vision and Pattern Recognition, Proc. CVPR (2010), pp. 239–246. https://doi.org/10.1109/CVPR.2010.5540207
21. L Sharan, C Liu, R Rosenholtz, EH Adelson, Recognizing materials using perceptually inspired features. Int. J. Comput. Vis. **103**(3), 348–371 (2013). https://doi.org/10.1007/s11263-013-0609-0
22. L Sharan, R Rosenholtz, EH Adelson, Accuracy and speed of material categorization in real-world images. J. Vis. **14**(9), 1–24 (2014). https://doi.org/10.1167/14.9.12
23. M Cimpoi et al., *Describing textures in the wild*, In 2014 IEEE Conference on Computer Vision and Pattern Recognition, *Proc. CVPR* (2014), pp. 3606–3613. https://doi.org/10.1109/CVPR.2014.461
24. M Cimpoi, S Maji, A Vedaldi, *Deep filter banks for texture recognition and segmentation*, In 2015 IEEE Conference on Computer Vision and Pattern Recognition, Proc. CVPR (2015), pp. 3828–3836. https://doi.org/10.1109/CVPR.2015.7299007
25. S Bell, P Upchurch, N Snavely, K Bala, *Material recognition in the wild with the materials in context database*, In 2015 IEEE Conference on Computer Vision and Pattern Recognition, Proc. CVPR (2015), pp. 3479–3487. https://doi.org/10.1109/CVPR.2015.7298970
26. Y. Zhang et al., "Integrating deep features for material recognition," (2015) [arXiv:1511.06522].
27. H Son, C Kim, N Hwang, C Kim, Y Kang, Classification of major construction materials in construction environments using ensemble classifiers. Adv. Eng. Inform. **28**(1), 1–10 (2014). https://doi.org/10.1016/j.aei.2013.10.001
28. A Dimitrov, M Golparvar-Fard, Vision-based material recognition for automated monitoring of construction progress and generating building information modeling from unordered site image collections. Adv. Eng. Inform. **28**(1), 37–49 (2014). https://doi.org/10.1016/j.aei.2013.11.002
29. JJ Koenderink et al., Bidirectional reflection distribution function of thoroughly pitted surfaces. Int. J. Comput. Vis. **31**(2-3), 129–144 (1999). https://doi.org/10.1023/A:1008061730969
30. M Oren, Generalization of the Lambertian model and implications for machine vision. Int. J. Comput. Vis. **14**(3), 227–251 (1995). https://doi.org/10.1007/BF01679684
31. GE Hinton et al., A fast learning algorithm for deep belief nets. Neural Comput. **18**(7), 1527–1554 (2006). https://doi.org/10.1162/neco.2006.18.7.1527
32. K. He et al., "Deep residual learning for image recognition," (2015) [arXiv:1512.03385].
33. O Russakovsky et al., ImageNet large scale visual recognition challenge. Int. J. Comput. Vis. **115**(3), 211–252 (2015). https://doi.org/10.1007/s11263-015-0816-y
34. J Deng et al., *ImageNet: a large-scale hierarchical image database*, In 2009 IEEE Conference on Computer Vision and Pattern Recognition, Proc. CVPR (2009), pp. 248–255. https://doi.org/10.1109/CVPR.2009.5206848].
35. A Krizhevsky, I Sutskever, GE Hinton, *ImageNet classification with deep convolutional neural networks*, In 26th Annual Conference on Neural Information Processing Systems, Adv. Neural Inf. Proces. Syst (2012), pp. 1097–1105
36. C Szegedy et al., *Going deeper with convolutions*, In 2015 IEEE Conference on Computer Vision and Pattern Recognition, Proc. CVPR (2015), pp. 1–9. https://doi.org/10.1109/CVPR.2015.7298594].
37. C Yan et al., Effective Uyghur language text detection in complex background images for traffic prompt identification. IEEE Trans. Intell. Transport. Syst. **19**(1), 220–229 (2018). https://doi.org/10.1109/TITS.2017.2749977
38. Labeled Faces in the Wild. http://vis-www.cs.umass.edu/lfw/results.html#glasssix. Accessed Sept 2017.
39. Large Scale Visual Recognition Challenge 2017 (ILSVRC2017), Task 1a: Object detection with provided training data. http://image-net.org/challenges/LSVRC/2017/results. Accessed Sept 2017.
40. Y. Chen, J. Li, H. Xiao et al., "Dual path networks," (2017) [arXiv:1707.01629].
41. C Yan et al., *Supervised hash coding with deep neural network for environment perception of intelligent vehicles*, IEEE Transactions on Intelligent Transportation Systems (2018). https://doi.org/10.1109/TITS.2017.2749965
42. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," (2014) [arXiv:1409.1556].
43. L Shao, F Zhu, X Li, Transfer learning for visual categorization: a survey. IEEE Trans. Neural Netw. Learn. Syst. **26**(5), 1019–1034 (2015)
44. F Zhuang et al., Survey on transfer learning research. J. Software **26**(1), 26–39 (2015). https://doi.org/10.13328/j.cnki.jos.004631
45. Y Sun, X Wang, X Tang, *Deeply learned face representations are sparse, selective, and robust*, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Proc. CVPR (2015), pp. 2892–2900
46. A. K. Reyes, et al. "Fine-tuning deep convolutional networks for plant recognition," CLEF (Working Notes), 2015.
47. P. Sermanet, et al. "Overfeat: Integrated recognition, localization and detection using convolutional networks," [arXiv:1312.6229] (2013).
48. V Mnih et al., Human-level control through deep reinforcement learning. Nature **518**(7540), 529–533 (2015)
49. Y Jia et al., *Caffe: convolutional architecture for fast feature embedding*, In 2014 ACM Conference on Multimedia, Proc. ACM Conf. Multimedia (2014), pp. 675–678. https://doi.org/10.1145/2647868.2654889
50. C Yan, Y Zhang, J Xu, et al., A highly parallel framework for HEVC coding unit partitioning tree decision on many-core processors. IEEE Signal. Process. Letters **21**(5), 573–576 (2014)
51. C Yan et al., Efficient parallel framework for HEVC motion estimation on many-core processors. IEEE Trans. Circuits. Syst. Vid. Technol. **24**(12), 2077–2089 (2014)
52. C Yan et al., Parallel deblocking filter for HEVC on many-core processor. Electron. Lett. **50**(5), 367–368 (2014)
53. C Yan et al., Efficient parallel HEVC intra-prediction on many-core processor. Electron. Lett. **50**(11), 805–806 (2014)