



# A benchmark data set for aircraft type recognition from remote sensing images

Zhi-Ze Wu<sup>a</sup>, Shou-Hong Wan<sup>b</sup>, Xiao-Feng Wang<sup>c\*</sup>, Ming Tan<sup>c</sup>, Le Zou<sup>c</sup>, Xin-Lu Li<sup>a</sup>, Yan Chen<sup>a</sup>

<sup>a</sup> Institute of Applied Optimization, School of Artificial Intelligence and Big Data, Hefei University, Hefei, Anhui, 230601, China

<sup>b</sup> School of Computer Science and Technology, University of Science and Technology of China, Hefei, Anhui, 230027, China

<sup>c</sup> School of Artificial Intelligence and Big Data, Hefei University, Hefei, Anhui, 230601, China



## ARTICLE INFO

### Article history:

Received 19 May 2019

Received in revised form 16 January 2020

Accepted 23 January 2020

Available online 31 January 2020

### Keywords:

Benchmark

Airplane type recognition

Remote sensing images

Pattern recognition

## ABSTRACT

Aircraft type recognition from remote sensing images has many civil and military applications. In images obtained with modern technologies such as high spatial resolution remote sensing, even details of aircraft can become visible. With this, the identification of aircraft types from remote sensing images becomes possible. However, the existing methods for this purpose have mostly been evaluated on different data sets and under different experimental settings. This makes it hard to compare their results and judge the progress in the field. Moreover, the data sets used are often not publicly available, which brings difficulties to reproduce the works for fair comparison. This severely limits the progress of research and the state of the art is not entirely clear. To address this problem, we introduce a new benchmark data set for aircraft type recognition from remote sensing images. This data set is called Multi-Type Aircraft Remote Sensing Images (MTARSI), which contains 9'385 images of 20 aircraft types, with complex backgrounds, different spatial resolutions, and complicated variations in pose, spatial location, illumination, and time period. The publicly available MTARSI data set allows researchers to develop more accurate and robust methods for both remote sensing image processing and interpretation analysis of remote sensing object. We also provide a performance analysis of state-of-the-art aircraft type recognition and deep learning approaches on MTARSI, which serves as baseline result on this benchmark.

© 2020 Elsevier B.V. All rights reserved.

## 1. Introduction

With the rapid development of the technology, remote sensing images have become a data source of great significance, which is widely used in civil [1] and military fields [2]. Among the many possible objects to detect from remote sensing image, aircraft are a key concern. Aircraft are not only an important means of transportation in the civil sector, but also a strategic objective in the military field. Here, dynamic aircraft detection and type identification can provide information of great significance for a timely analysis of the battlefield situation and the formulation of military decisions [2,3]. Civil applications such as emergency aircraft search, airport surveillance, and aircraft detection are highly important as well [1,4].

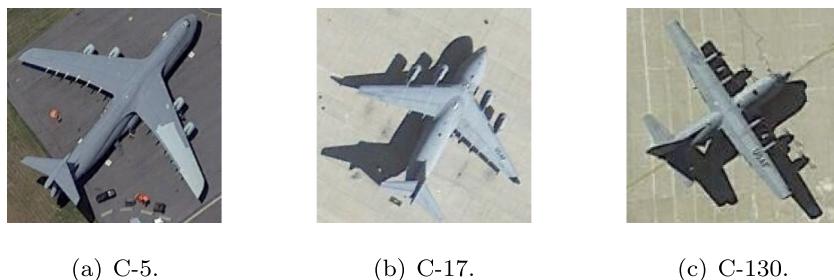
Due to the continuous improvement of the spatial resolution of remote sensing cameras, the information contained in single images is growing. The currently existing automated interpretation approaches for identifying aircraft from such images are unable to meet the needs of many real-world applications [5]. Therefore, the question how to accurately and quickly extract the position and type of aircraft has come into the focus of research.

In particular, aircraft type recognition has attracted significant attention from academia and industry. Aircraft differ in details such as model and type, as illustrated in Fig. 1, where three types of transport aircraft, C-1, C-17, and C-130, are shown in sequence. Even though these aircraft have different uses and functions, they are quite similar in appearance. In addition, complex backgrounds and the characteristics of the image sources, i.e., different satellites, cause additional difficulties for aircraft recognition. The apparent shape, shadow, and color of an object may further be influenced by the solar radiation angle, the radar radiation angle, and the surrounding environment [6]. Therefore, recognizing the aircraft type from remote sensing images remains a challenging problem.

\* Corresponding author.

E-mail addresses: [wuzhize@mail.ustc.edu.cn](mailto:wuzhize@mail.ustc.edu.cn) (Z.-Z. Wu), [wansh@ustc.edu.cn](mailto:wansh@ustc.edu.cn) (S.-H. Wan), [xfwang@hfuu.edu.cn](mailto:xfwang@hfuu.edu.cn) (X.-F. Wang), [tanming@hfuu.edu.cn](mailto:tanming@hfuu.edu.cn) (M. Tan), [zoule@hfuu.edu.cn](mailto:zoule@hfuu.edu.cn) (L. Zou), [xinlu.li@mydit.ie](mailto:xinlu.li@mydit.ie) (X.-L. Li), [chenyan@hfuu.edu.cn](mailto:chenyan@hfuu.edu.cn) (Y. Chen).

URL: <http://www.iao.hfuu.edu.cn> (Z.-Z. Wu).



**Fig. 1.** Three types of transport aircraft: C-5, C-17, and C-130.

In recent years, several data sets with images that can be used to train aircraft detection algorithms have been created, including UCMerced\_LandUse [7], NWPU-RESISC45 [8], PatternNet [9], and FGVC-Aircraft [10]. However, in UCMerced\_LandUse, NWPU-RESISC45, and PatternNet, aircraft are only a sub-category and appear only in a subset of the data. Also, there is no distinction between their types. While FGVC-Aircraft can be used for fine grained visual categorization of aircraft, it is not based on remote sensing images and instead contains natural scenes. In summary, the publicly available data sets are not suitable for the domain of aircraft type recognition from remote sensing images.

The currently existing approaches for aircraft type recognition can coarsely be divided into three categories, namely methods based on template matching [4,11,12], based on the shallow feature learning model [3,13,14], and based on deep neural networks [4,6,12,15,16]. Most of these works were evaluated under different experimental settings and on different data sets, which, moreover, are not publicly available. This makes it hard to judge the overall progress in the field and also reduces the reproducibility of results very significantly.

A benchmark data set is needed, which should be permanently publicly available and which would make it easy to perform fair algorithm comparisons. Such a benchmark would accelerate and promote the development of the research field [17]. The data set should demonstrate all the challenging aspects of the problem, for instance, the high diversity in the geometrical and spatial patterns of aircraft in remote sensing images. Along with the data set, a comprehensive set of results obtained with state-of-the-art approaches should be provided, so that new works can directly be compared to tested, published, and verified results. In this paper, we present such a benchmark data set, called the *Multi-type Aircraft Remote Sensing Images* (MTARSI) [18]. It contains 9'385 remote sensing images of 20 aircraft types, with complex backgrounds, different spatial resolutions, and complicated variations in pose, spatial location, illumination, and time period. As discussed in the benchmarking study [19], we describe the collection, simulation, labeling, and categorization of the samples for the data set in detail. This will allow researchers to extend the data set using the same approach as us, if the need for larger data set should arise due to the progress in technology, or to develop similar data sets for other domains. We then compare and evaluate the performance of several typical aircraft type recognition and deep learning approaches on MTARSI in order to provide proper baseline results.

The contributions of this paper are three-fold:

1. We construct, present, and publish MTARSI — the first public aircraft remote sensing image database. By doing so, we significantly aid the development of robust methods for aircraft type recognition from remote sensing images.
2. We describe in detail how this data set is created by applying image simulation methods supported by data gathered from online sources. This allows for applying our approach to generate high-quality benchmark image data sets for other application areas.

3. We provide the results obtained by state-of-the-art approaches in the field, to serve as baseline for comparison for researchers working on remote sensing computer vision. We show that MTARSI can indeed identify the strengths and weaknesses of different aircraft type recognition algorithms for remote sensing images.

The remainder of this paper is organized as follows. In Section 2, we review existing data sets for aircraft recognition and the state-of-the-art in aircraft type recognition from remote sensing images. Then, the design and contents of MTARSI are presented in Section 3. The performance of several state-of-the-art aircraft type recognition methods is evaluated on the MTARSI in Section 4. Finally, the paper closes with a conclusion and discussion in Section 5, where we also describe how to obtain the MTARSI data set from its public repository.

## 2. Related work

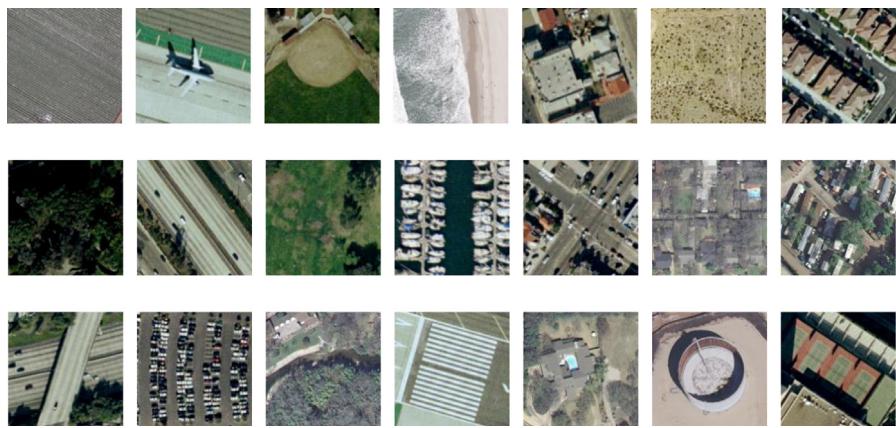
In this section, we first review several data sets commonly used for aircraft recognition and then present a comprehensive analysis of state-of-the-art for aircraft type recognition from remote sensing images.

### 2.1. Existing data sets for aircraft recognition

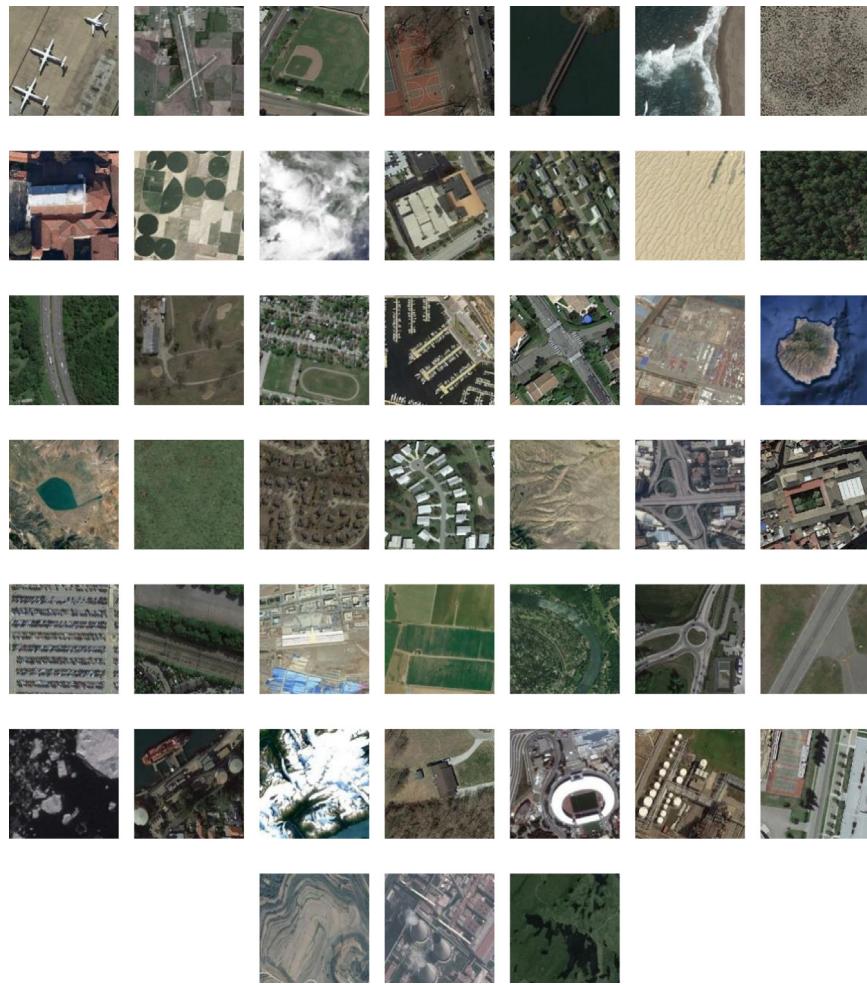
At present, there are four popular data sets for aircraft recognition, namely UCMerced\_LandUse [7], NWPU-RESISC45 [8], PatternNet [9], and FGVC-Aircraft [10].

The *UCMerced\_LandUse Data Set* [7] was established at the University of California, Merced, USA. It is one of the most widely used data sets in the field of remote sensing, especially for various remote sensing image classification tasks. The images in UCMerced\_LandUse were first selected from the United States Geological Survey National Map and then randomly cropped into sizes of  $256 \times 256$  pixels. The spatial resolution is 1 foot (about 0.3 m). UCMerced\_LandUse contains a total of 21 different scenario types: agricultural, airplane, baseball diamond, beach, buildings, chaparral, dense residential, forest, freeway, golf course, harbor, intersection, medium residential, mobile home park, overpass, parking lot, river, runway, sparse residential, storage tanks, and tennis court. The data set has a total of 2100 images, 100 in each category. Note that “aircraft” is only a sub-category of this data set, i.e., there are only 100 images relevant to our domain. Fig. 2 shows some examples of this data set.

The *NWPU-RESISC45 Data Set* [8] was created at the Northwestern Polytechnical University, Xi'an, China. The data set is by far the largest open source remote sensing imagery data sets. It contains a total of 31'500 images, which are based on an investigation of existing data sets and a selection of the most representative images of 45 remote sensing scene categories, including aircraft, airports, but also baseball fields, basketball courts, beaches, bridges, bushes, churches, circular farmland, clouds, commercial



**Fig. 2.** A subset of the images in the UCMerced\_LandUse data set [7].



**Fig. 3.** A subset of the images in the NWPU-RESISC45 data set [8].

areas, industrial areas, crossroads, railway, boats, snow mountains, and so on. Each category in the data set has 700 images with an image size of  $256 \times 256$  pixels and a spatial resolution of between 0.2 and 30 m. Aircraft are again a sub-category of this data set, this time including 700 images. Fig. 3 shows some examples of this data set.

The *PatternNet Data Set* [9] was established at the State Key Laboratory of Surveying and Mapping Remote Sensing Information Engineering of Wuhan University, Wuhan, China and the University of California, Merced, USA. The data set is currently

the largest high spatial resolution remote sensing image data set, with a total of 38 categories, such as airplanes, baseball field, basketball court, beach, bridges, cemeteries, shrubs, tanks, swimming pools, tennis courts, substations, and so on. Each category contains 800 images with a size of  $256 \times 256$  pixels and spatial resolutions ranging from 0.062 m to 4.693 m. The images of the PatternNet data set are obtained from the Google Earth Satellite Imagery and the Google Maps US Urban Interface. Fig. 4 shows some examples from this data set.



**Fig. 4.** A subset of the images in the PatternNet data set [9].



**Fig. 5.** A subset of the images in the FGVC-Aircraft data set [10].

The *Fine-Grained Visual Classification of Aircraft* (FGVC-Aircraft [10]) is a benchmark data set for the visual categorization of aircraft. The data set contains 10'200 images of aircraft, with 100 images for each of 102 different aircraft model variants, most of which are airplanes [10]. The (main) aircraft in each image is annotated with a tight bounding box and a hierarchical airplane model label. Aircraft models are organized in a hierarchy with four levels [10]. These are, from finer to coarser:

- **Model**, e.g., Boeing 737-76J: Since certain models are nearly visually indistinguishable, this level is usually not used in the evaluation of algorithms.
- **Variant**, e.g., Boeing 737-700: A variant collapses all the models that are visually indistinguishable into one class [10]. The data set comprises 102 different variants.

- **Family**, e.g., Boeing 737: The data set comprises 70 different families.
- **Manufacturer**, e.g., Boeing: The data set comprises 41 different manufacturers.

Note that this data has been used as part of ImageNet [20]. Some samples from this data set are displayed in Fig. 5, showing that the images contain front and side views of aircraft, while remote sensing images from satellites are usually top views. Thus, while FGVC-Aircraft can indeed be used for visual categorization of aircraft, it is not suitable for training or comparing algorithms for remote sensing.

To the best of our knowledge, no remote sensing data set for aircraft type recognition is publicly available. This is perhaps the primary reason for limiting the development in the area.

## 2.2. Review on aircraft type recognition of remote sensing images

Most of the recent methods for aircraft recognition from remote sensing images are focused on deciding whether an object is an aircraft or not [21–25]. Less attention has been awarded to the further fine-grained classification and recognition of the aircraft model and type. General image classification approaches can be divided into hand-crafted feature based methods [26–29] and deep learning based methods [30–33]. The research on object type identification from remote sensing images can roughly be divided into three categories, namely methods that based on template matching [4,11,12], based on the shallow feature learning model [3,13,14], and based on deep neural networks [4,6,12,15,16].

Let us first consider the template matching based methods. Here, Wu et al. [11] proposed a jigsaw reconstruction approach to extract aircraft shapes and then match them with standard templates. Zhao et al. [4] designed a keypoint detection model based on Convolutional Neuronal Networks (CNNs) [30] and a keypoint matching method to recognize aircraft, transforming the aircraft recognition problem into the landmark detection problem. Due to the use of CNNs, this method could also partially be classified as a deep learning approach. Furthermore, Zuo et al. [12] built a template matching method using both aircraft keypoints and segmentation results, which provided more detailed information to improve recognition accuracy.

An important shallow feature learning model based method was proposed by Hsieh et al. [13], who use a hierarchical classification algorithm based on four different features: wavelet transform, Zernike moment, distance transform, and bitmap. In [3], a coarse-to-fine method is built to integrate high-level information for aircraft type recognition. A method using the artificial bee colony algorithm with an edge potential function is proposed in [14] to solve this problem.

These methods for object type identification from remote sensing images have achieved significant results, but there are still many shortcomings. The methods based on template matching have a problem to adapt to complex scenarios [6]. The methods based on the shallow feature learning model have difficulties to extract high-level semantic features from objects and there is a bottleneck in the recognition performance [15].

In recent years, with the improvement of parallel GPU computation and the emergence of large-scale image data sets, deep neural networks [30–35] have experienced tremendous development and have proven to be efficient in a variety of vision tasks, such as the classification [36–39], detection [40–43], and segmentation [44–47] of images. Also benefiting from the existing large-scale data sets from different application domains [20, 48], deep neural networks can be initialized by using transfer learning [38,49]. As a result, the application of deep neural networks has led to breakthroughs in aircraft recognition from

**Table 1**  
Different aircraft classes and the number of images in each class of MTARSI.

Types	#images	Types	#images	Types	#images	Types	#images
B-1	513	C-130	763	F-16	372	A-10	345
B-2	619	C-135	526	F-22	846	T-6	248
B-29	321	C-17	480	KC-10	554	A-26	230
B-52	548	C-5	499	C-21	491	P-63	305
Boeing	605	E-3	452	U-2	362	T-43	306

remote sensing images. Diao et al. [15] used a pixel-wise learning method based on deep belief networks (DBNs) for object recognition in remote sensing images. In [4], Zhao et al. proposed a idea to address the aircraft type recognition problem by detecting the landmark points of an aircraft using a vanilla network. By combining semantic segmentation and aircraft type recognition, another CNN based method is used in [12]. Zhang et al. [16] developed an aircraft type recognition framework for learning representative features from images without type labels, based on conditional generative adversarial networks (GANs). Fu et al. [6] propose a fine-grained aircraft recognition method for remote sensing images. Their multi-class activation mapping (MultiCAM) utilizes two subnetworks, i.e., the target net and the object net, in order fully use the features of discriminative object parts.

Compared with handcrafted-feature based methods, neural-network based models can obtain better discriminative feature representations and achieve significant improvements in terms of the recognition accuracy. However, these models must be trained well on target domain data sets after the initialized stage.

One problem is that the aforementioned deep neural network based aircraft recognition methods [4,6,12,15,16] for remote sensing images were evaluated on different data sets under different experimental settings. This makes it hard to compare their results. Additionally, the data sets used in these works and in the research on the template matching based [4,11,12] and shallow feature learning based methods [3,13,14] are not publicly available. It is thus not easy to reproduce these works for a fair comparisons. Therefore, the state of the art of aircraft type recognition of remote images is not entirely clear.

To address this problem, we introduce a remote sensing image data set of diverse aircraft, which can be used to assess and evaluate the performance of aircraft type recognition algorithms. It also particularly benefits the development of both remote sensing image processing and interpretation analysis of remote sensing object in the future.

## 3. The MTARSI data set

To advance the state of the art in aircraft type recognition of remote sensing images, we construct MTARSI, a public large-scale aircraft image data set. In all, MTARSI has a total of 9'385 remote sensing images acquired from Google Earth satellite imagery and manually expanded, including 20 different types of aircraft covering 36 airports. The new data set is made up of the following 20 aircraft types: B-1, B-2, B-29, B-52, Boeing, C-130, C-135, C-17, C-5, E-3, F-16, F-22, KC-10, C-21, U-2, A-10, A-26, P-63, T-6, and T-43. All the sample images are carefully labeled by seven specialists in the field of remote sensing images interpretation. Each image contains exactly one complete aircraft. Each type of aircraft sample in the MTARSI data set is shown in Fig. 6. The numbers of sample images vary with different aircraft types (see Table 1) and range from 230 to 846.

The spatial resolution of the images is between 0.3 and 1.0 m. During the data set construction process, the sample images for each class in MTARSI are carefully chosen from different airports around the world, but mainly in the United States, England, Japan,



**Fig. 6.** Samples of the 20 aircraft types from MTARSI.

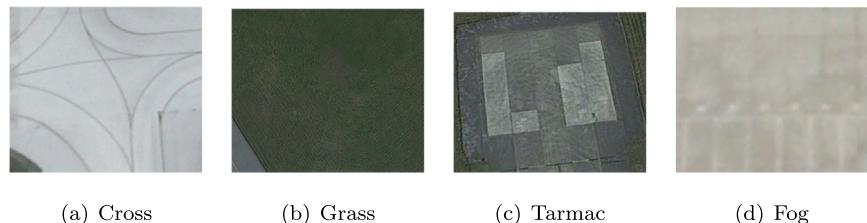
Russia, France, and China. They were extracted at different times and seasons under different imaging conditions, which increases the intraclass diversities of the data. Different colors of the body paint are chosen for the same type of aircraft and different backgrounds and poses are also taken into account. This way, the intra-class diversity of the data set is greatly increased. Moreover, for each type, aircraft with similar appearance but different models are selected, which gives the data set certain inter-class similarity. Therefore, this data set is challenging for methods that try to recognize the aircraft type.

In Table 2, we show some samples based on variations of resolution, pose, color, light, background, and model. Some aircraft, such as the B-2 bomber and the KC-10 tanker, are very rare and difficult to capture by satellite remote sensing. This situation hinders the process of collecting and building image data sets.

In order to mitigate this problem, we artificially expand the data by simulating images of aircraft that are rare and difficult to be observed. The specific process is shown in Fig. 8.

First, a Prewitt-operator based segmentation algorithm [50] is applied to an existing image of the aircraft. The Prewitt operator is often used in image processing, particularly within edge detection algorithms [51]. It is a discrete differentiation operator approximating the gradient of the image intensity function. With this operator, we can separate the aircraft from the background.

A different background image is then selected randomly from the relevant remote sensing images (such as airport runways and landing field, etc.) that do not contain any aircraft. Fig. 7 shows four background sample images, including cross runway, grass, tarmac and fog. Finally, the extracted aircraft image is rotated, mirrored, and then combined with the selected background to obtain a final simulated image. The extracted object and the selected



**Fig. 7.** Background samples for simulating rare aircraft.

**Table 2**  
Samples under the variation of resolution, pose, color, time/light, background, and model.

The figure displays a 5x3 grid of images illustrating the impact of various factors on a 3D airplane model. The columns represent different factors: Resolution, Pose, Color, Time/light, Background, and Model. Each row shows three variations of the model under that factor.

- Resolution:** The first column shows the effect of increasing resolution from low to high. The images transition from a blurry, low-resolution version on the left to a sharp, high-resolution version on the right.
- Pose:** The second column shows the effect of changing the camera pose from a front-on view to a top-down view. The images show the airplane from a side-on perspective on the left, transitioning to a top-down perspective in the middle, and back to a side-on perspective on the right.
- Color:** The third column shows the effect of changing the color scheme. The images show the airplane in white on the left, transitioning to a red and black color scheme in the middle, and back to white on the right.
- Time/light:** The fourth column shows the effect of varying lighting conditions over time. The images show the airplane in a dark, shadowed environment on the left, transitioning to a bright, well-lit environment in the middle, and back to a dark environment on the right.
- Background:** The fifth column shows the effect of different background environments. The images show the airplane on a tarmac with runway markings on the left, against a plain white background in the middle, and on a green grassy field on the right.
- Model:** The sixth column shows the effect of different 3D models of the same airplane. The images show a standard white airplane model on the left, a smaller, more compact model in the middle, and a larger, more detailed model on the right.

background image are always in the same spatial resolution and are obtained from the same remote sensing sensor.

#### **4. Evaluation of state-of-the-art algorithms on the MTARSI data set**

We now evaluate general identification algorithms and deep learning approaches for aircraft type recognition on our new data set. We first investigate the performance of the typical state-of-the-art (but still off-the-shelf) CNN AlexNet [30].

Transfer learning [38,49] is typically applied in object recognition. It prescribes two steps, namely pre-training on a different data set and then fine-tuning on the target data set (here: MTARSI). We thus conduct a series of experiments with AlexNet applying pre-training in Section 4.1.

The question how the fine-tuning parameters effects the final performance is then investigated in Section 4.2. Thus, we

establish a baseline for the performance of a very typical contemporary machine learning technique in our domain and on our data set. Any more sophisticated approach which is developed for aircraft type recognition from satellite images should be able to outperform this method.

We then conduct a larger study to compare ten aircraft recognition methods in Section 4.3. Six of them again apply deep learning based on CNNs while the other four methods are hand-crafted feature approaches. This way, we cover the state-of-the-art on aircraft recognition from remote sensing image data and can provide reproducible results that other researchers can compare with.

Finally, we then investigate which of the image features of MTARSI cause problems for which approach in Section 4.4. In summary, we answer questions such as:

- Which types of aircraft were recognized incorrectly by which method in our experiment?
  - Which aircraft were confused most often?
  - Which variations caused most trouble for which approach?

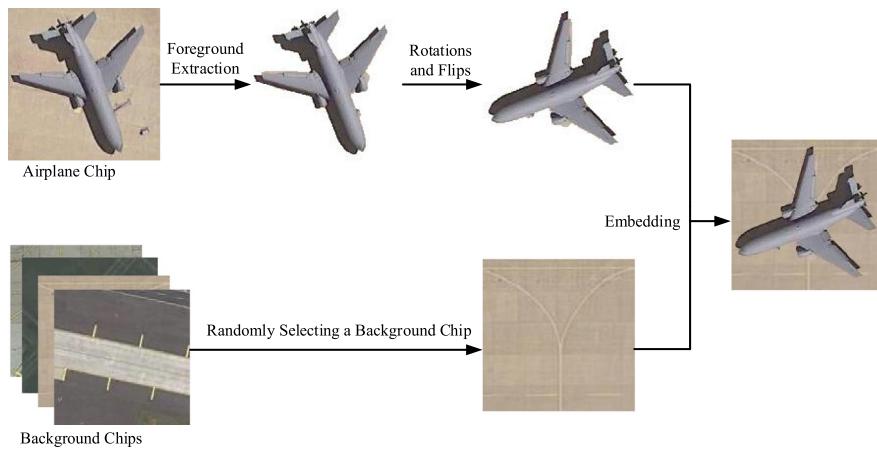
As environment for all experiments, we use a PC with Intel i7-7700 CPU with 3.6 GHz, 16 GB RAM, and an Nvidia GTX Titan X graphics card. The algorithms are implemented and executed based on the Caffe open source library [52].

#### *4.1. Experiments with pre-training AlexNet network using different source domain data sets*

We first want to analyze the impact of transferring knowledge from different source domains to our target domain, the aircraft recognition from remote sensing data. We also experimentally investigate which kind of source data set is appropriate for pre-training deep neural networks for aircraft type recognition from remote sensing images. This experiment therefore uses different source domain data sets, namely the seven data sets: UCMerched\_LandUse [7], NWPU-RESISC45 [8], PatternNet [9], ImageNet [20], AID [5], WHU-RS19 [53] and RSSCN7 [54]. AID, WHU-RS19, and RSSCN7 are the latest data sets used to classify remote sensing scenes. Based on the source data sets, a group of experiments that use MTARSI to fine-tune the AlexNet network model *without* using transfer learning is carried out for comparison.

When pre-training the AlexNet model with different data sets, the output of the eighth fully connected layer needs to be modified accordingly, that is, its output dimension should be equal to the number of categories used in the pre-trained data set. For example, when using the ImageNet data set for pre-training, the output dimension of the eighth fully connected layer in the AlexNet network model needs to be set to 1000.

As listed in [Table 3](#), AlexNet requires inputs of a fixed size  $227 \times 227 \times 3$ . After two alternations of convolutional and pooling layers, it produces as output 256 feature maps, each with the size  $13 \times 13$ . The third, fourth, and fifth convolutional layers are connected to each other without poolings layers. The



**Fig. 8.** Flowchart of the remote sensing aircraft image simulation strategy.

**Table 3**

Details of network configurations of AlexNet used in the experiments.

Layer	Input	Kernel	Stride	Pad	Output
Conv1	227*227*3	11*11	4	0	55*55*96
Pool1	55*55*96	3*3	2	0	27*27*96
Conv2	27*27*96	5*5	1	2	27*27*256
Pool2	27*27*256	3*3	2	0	13*13*256
Conv3	13*13*256	3*3	1	1	13*13*384
Conv4	13*13*384	3*3	1	1	13*13*384
Conv5	13*13*384	3*3	1	1	13*13*256
Pool5	13*13*256	3*3	2	0	6*6*256
FC6	6*6*256	6*6	1	0	4096*1
FC7	4096*1	1*1	1	0	4096*1
FC8	4096*1	1*1	1	0	1000*1

third convolutional layer (Conv3) takes the output of the second pooling layer (Pool2) as input and filters it with 384 kernels of size  $3 \times 3 \times 256$ . The fourth convolutional layer contains 384 kernels of size  $3 \times 3 \times 384$ , and the fifth convolutional layer has 256 kernels of size  $3 \times 3 \times 384$ . The series of convolutional layers is followed by a pooling layer (Pool5) and the three fully-connected layers F6, F7 and F8. The output of the last fully-connected layer (F8) is then a 1000-dimensional vector, i.e. a probability distribution over the 1000 class labels.

When the AlexNet network model is pre-trained to convergence, the pre-training parameters of all layers except the eighth fully-connected layer are migrated into the fine-tuning AlexNet network model. Then, the MTARSI data set is divided into a training and a test data set according to a ratio 4 : 1, that is, the training of AlexNet network model is fine-tuned using 80% of MTARSI.

The remaining 20% of the data are used to test and verify the network model after training and fine-tuning, and to evaluate the performance of the network model by identifying the correct rate. Based on the settings described in [30], the learning rate of each layer during the fine-tuning of the AlexNet model was set to 0.001, and it would be set to the previous 10% after 5000 iterations, and the momentum unit was set to 0.9. The weight decay is set to 0.0005, the batch size is set to 44, and the number of iterations depends on the convergence of the network model.

The experimental results are shown in Table 4. We note that when the MTARSI data set is used directly for training without transfer learning, the recognition accuracy of the AlexNet network model on MTARSI is 80.13%. However, when the data sets WHU-RS19, UCMerced\_LandUse, and RSSCN7 are used for pre-training, then even after parameter fine-tuning based on MTARSI, the recognition accuracy is lower. This is called the “negative transfer” phenomenon and means that the knowledge learned

**Table 4**

Pre-training results using different source domain data sets.

Pre-training data set	Scale	Categories	Accuracy
Without Pre-training	-	-	80.13%
WHU-RS19	1005	19	75.65%
UCMerced_LandUse	2100	21	74.47%
RSSCN7	2800	7	77.15%
AID	10'000	30	80.56%
NWPU-RESISC45	30'400	45	82.527%
PatternNet	31'500	38	81.79%
ImageNet	about 1.2 million	1000	84.61%

from the source domain has a negative impact on the learning on the target domain [49]. The reason for this phenomenon is that the three source domain data sets do not contain enough images and overfitting occurs during the pre-training. The knowledge in the source domain has not been learned well and thus cannot be transferred usefully to the target domain.

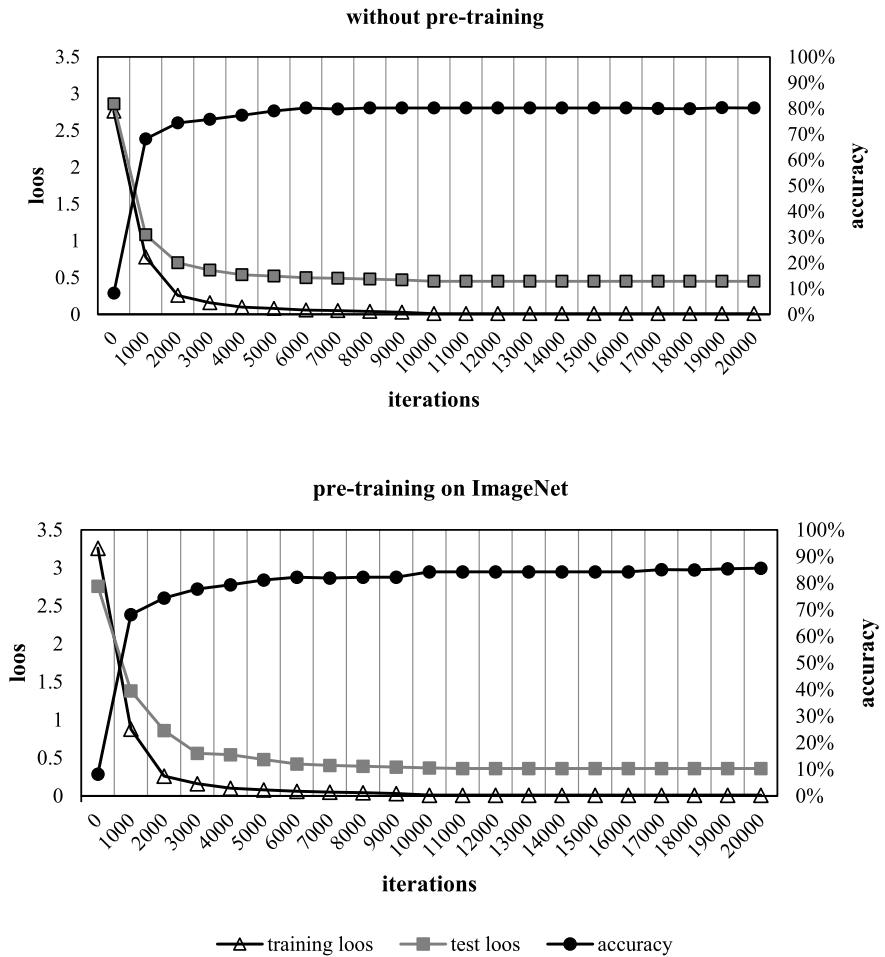
If we use the data sets AID, NWPU-RESISC45, PatternNet, or ImageNet as the source domain data set for pre-training, then AlexNet works well after the fine-tuning phase. The accuracy becomes 80.56%, 82.52%, 81.79%, and 84.61%, respectively, which is significantly higher than the 80.13% achieved without transfer learning. The results obtained with pre-training on ImageNet are the best.

When the size of the source domain data set used in the pre-training reaches a certain level and there is enough data to ensure that the knowledge in the source domain is grasped, transfer learning will promote the learning on the target domain MTARSI. In order to measure the value of the pre-training on the source domain data set, we additionally carry out a group of experiments where we directly train on MTARSI without using transfer learning. In Fig. 9, we illustrate the loss and recognition accuracy rate of the AlexNet network on MTARSI without pre-training and with pre-training on ImageNet, respectively.

It can clearly be seen that the AlexNet network model using transfer learning has better recognition accuracy and that the convergence speed and convergence effect on the training set and the test set are better than without transfer learning.

#### 4.2. Experiments of layers in AlexNet network under different learning rate settings

We now investigate the different layers of the AlexNet network model under different learning rates, in order to further study the inner characteristics of deep neural networks for processing the remote sensing images.



**Fig. 9.** Loss and recognition accuracy rate of the AlexNet network on MTARSI for no pre-training (top) and pre-training (bottom) on ImageNet.

**Table 5**  
Different learning rate settings the and corresponding experimental results.

Model	Conv-1	Conv-2	Conv-3	Conv-4	Conv-5	FC-6	FC-7	FC-8	Accuracy
Model-1	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	85.61%
Model-2	0	0.001	0.001	0.001	0.001	0.001	0.001	0.001	85.61%
Model-3	0	0	0.001	0.001	0.001	0.001	0.001	0.001	85.61%
Model-4	0	0	0	0.001	0.001	0.001	0.001	0.001	84.31%
Model-5	0	0	0	0	0.001	0.001	0.001	0.001	83.56%
Model-6	0	0	0	0	0	0.001	0.001	0.001	82.35%
Model-7	0	0	0	0	0	0	0.001	0.001	81.52%
Model-8	0	0	0	0	0	0	0	0.001	80.27%

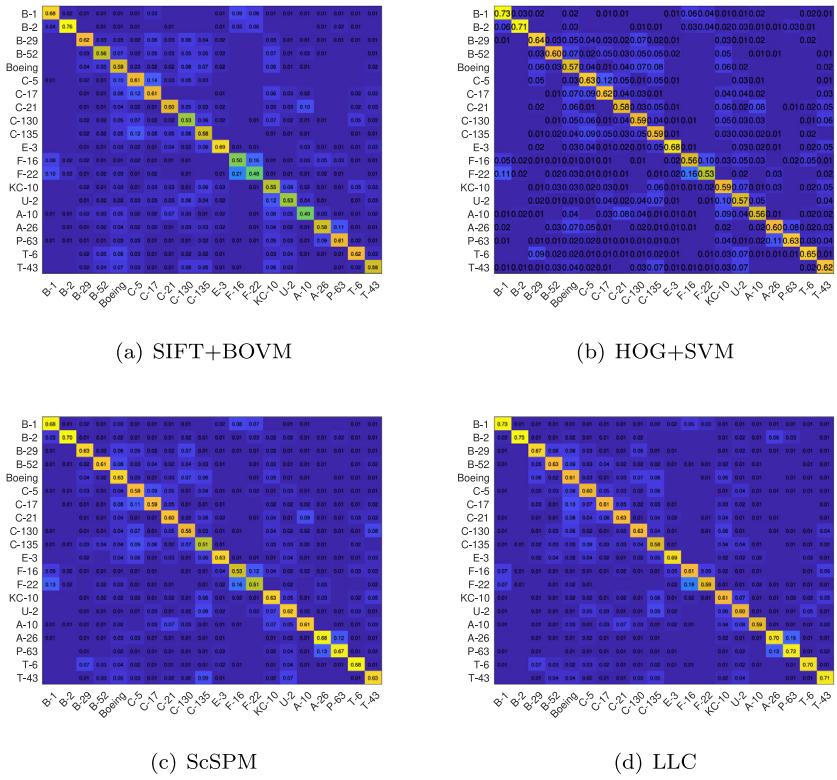
First, the ImageNet data set was used to pre-train the AlexNet network model, since we know that this has a very positive impact on the accuracy. Then, in the parameter transfer phase, eight different learning rate settings for the layers in the fine-tuning AlexNet network model were investigated. Each setting corresponds to a model, as shown in Table 5. These models were again fine-tuned using 80% of the MTARSI data set and tested on the remaining 20% of the data.

In Table 5, the learning rate “0” indicates that the layer parameters are transferred from the pre-training stage without fine-tuning on MTARSI. Each iteration consists of 5000 training cycles, the learning rate is set to the previous 10%, the impulse unit is set to 0.9, the weight decay is set to 0.0005, the batch size is set to 44, and the number of iterations depends on the convergence of the network model.

In the networks, the low-level convolutional layers generally extract common features such as edges and textures of the image, while the high-level convolutional layer and the all-connected

layer extract more task-related high-level features [55]. As can be seen from Table 5, Model-2 only uses the parameters of the Convolution Layer 1 transferred from the pre-trained AlexNet network and does not fine-tune the layer parameters of the target domain. It can achieve a recognition accuracy rate equivalent to Model-1. The same is true for Model-3. This also shows that the features extracted by the first few convolutional layers of the convolutional neural network have certain common semantics.

The experimental results of Model-4 to Model-8 show that the learning will be affected if no task-related parameter fine-tuning is performed for the high-level convolutional layer and the fully connected layer after the parameter migration. From this, we can conclude that the low-level parameters of the pre-trained CNNs can be directly used for the target task, while fine-tuning of the high-level parameters is necessary.



**Fig. 10.** Confusion matrices on MTARSI when using SIFT [26]+BOVW, HOG [27]+SVM, ScSPM [28], an LLC [29].

### 4.3. Experiments of different classification algorithms on MTARSI

We now investigate the performance of five state-of-the-art CNN structures on MTARSI, namely VGG [31], GoogLeNet [32], ResNet [33], DenseNet [34], and EfficientNet [35].

In the experiment, the five CNNs again use the ImageNet data set for pre-training. The MTARSI data set is also again divided into a training and a test set according to a ratio of 4 : 1 and the size of the images is fixed to 256 × 256 pixels. As in the experiment in Section 4.1, the learning rate of each layer of the network model in the parameter fine-tuning stage is uniformly set to 0.001, and the training rate is set to the previous 10%, the impulse unit is 0.9, the weight decay is 0.0005, the batch size is 44, and the number of iterations depends on the convergence of the network model.

For comparison purposes, we also apply four hand-crafted feature-based approaches [26–29]. For the local patch descriptor, such as SIFT and HOG, we use a fixed size grid (16 × 16 pixels) with the spacing step of 8 pixels to extract all the descriptors in the image and adopt the average pooling method for each dimension of the descriptor to get the final image features. We fix the dictionary size to 4096 for BOVW and LLC, and to 256 for ScSPM.

Table 6 shows that the aircraft type recognition algorithms based on transfer learning are far superior to the traditional classification methods in aircraft target recognition performance and that the average accuracy rate is improved by nearly 26%. This also indicates that for aircraft type recognition from remote sensing images, methods based on CNNs can extract higher-level features with more generalization and image expression ability compared to the traditional feature-based methods. They also can more effectively distinguish aircraft of different types.

We further find that ResNet, DenseNet, and EfficientNet have similar performance, while AlexNet, GoogleNet, and VGG perform slightly worse. In terms of time performance of recognition,

**Table 6**

Experimental results of different classification algorithms on the MTARSI data set.

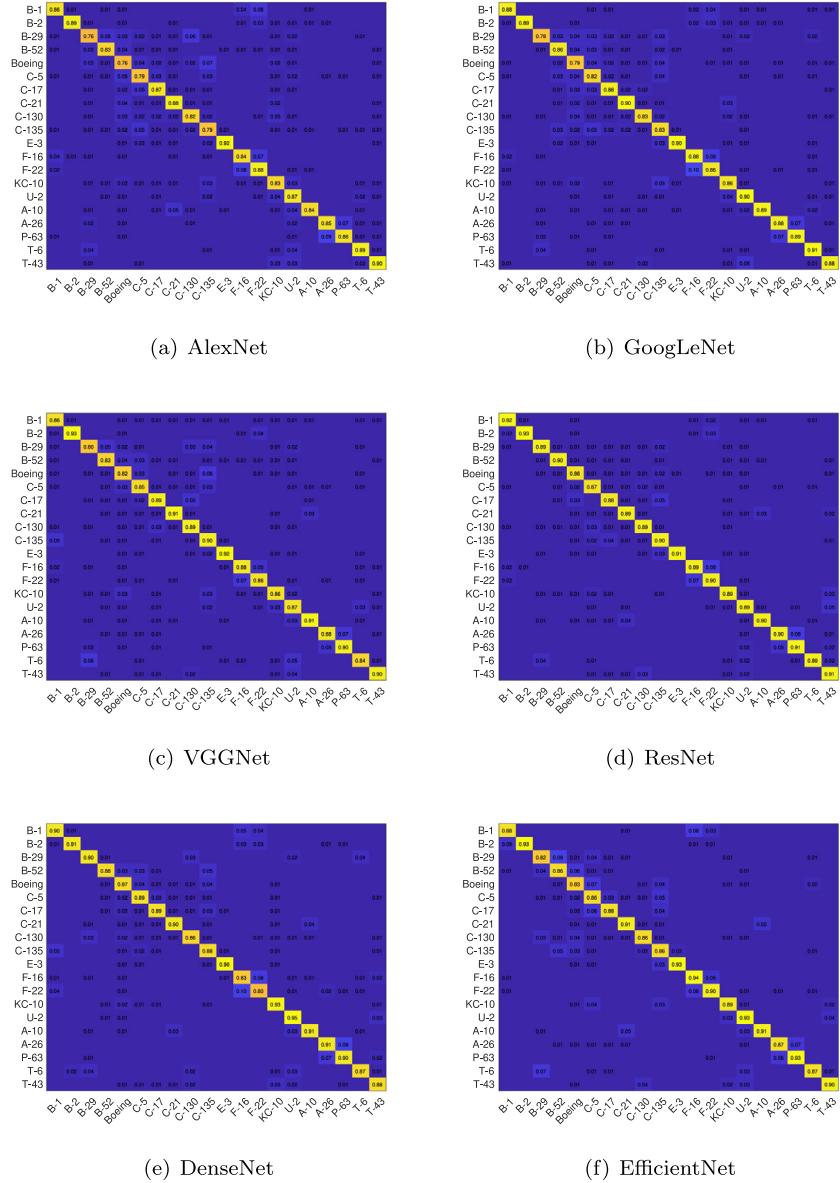
Method	Accuracy
SIFT [26]+BOVW	59.02%
HOG [27]+SVM	61.34%
ScSPM [28]	60.61%
LLC [29]	64.93%
AlexNet	85.61%
GoogleNet	86.53%
VGG	87.56%
ResNet	89.61%
DenseNet	89.15%
EfficientNet	<b>89.79%</b>

AlexNet and EfficientNet perform best, while the others are similar. Note that AlexNet has only eight layers, which is much shallower than the other deep models. For EfficientNet, a compound scaling method is used for balancing network width, depth, and resolution. With this, inference via EfficientNet is 6.1 times faster than with ResNet [35].

### 4.4. Analysis of image features of MTARSI

Based on our experiments with different classification algorithms, we now investigate which of the image features of MTARSI are most problematic. We first study the recognition rate for each type of aircraft over different classification methods and check which types were confused most often. In order to find which variations caused most trouble for which approach, we further conduct a new group of experiments on a validation set of size 600, by randomly selecting 100 images from the testing set of MTARSI based on their variations including resolution, pose, color, light, background, and aircraft model.

In Fig. 10, the confusion matrices on MTARSI for SIFT [26], HOG [27], ScSPM [28], LLC [29], are presented. Fig. 11 provides



**Fig. 11.** Confusion matrices on MTARSI when using AlexNet, GoogLeNet, VGG, ResNet, DenseNet, and EfficientNet.

the confusion matrices on MTARSI for AlexNet, GoogLeNet, VGG, ResNET, DenseNet, and EfficientNet. We find that the hand-crafted feature based methods do generalize, but suffer from some misclassifications on similar aircraft types or models, such as C-5 and Boeing or B-52 and C-5. The deep neural network based methods exhibit good performance except on some extremely similar aircraft. We also find that there are significant differences in the recognition performance of these two kinds of methods for each specific type.

We conclude that MTARSI is a challenging benchmark for most of the state-of-the-art methods. It is very likely that it can help to design, analyze, and improve practical applications of aircraft identification. For this purpose, deep learning models like ResNet, DenseNet, and EfficientNet lend themselves.

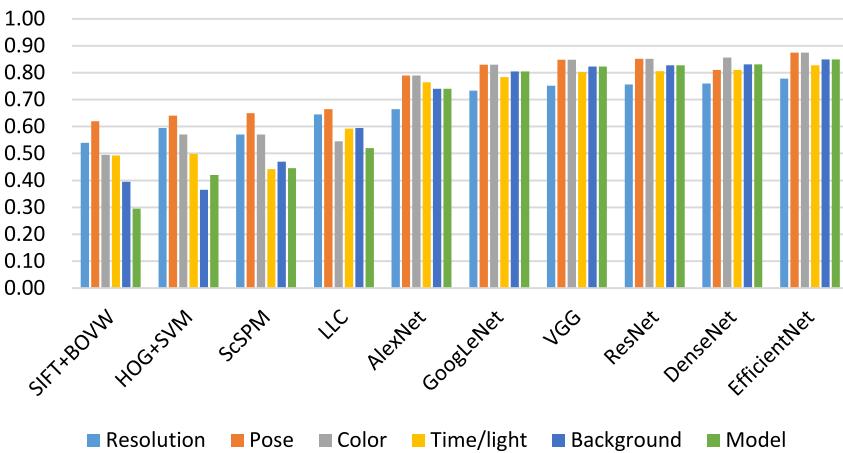
We illustrate the recognition rates of the state-of-the-art methods on the validation set including six variations, i.e., resolution, pose, color, light, background, and aircraft model, in Fig. 12. We can see that the aircraft model and background are the two biggest factors that influence the method based on hand-crafted feature, and the image resolution is the biggest variable that affects the performance of the deep learning method.

## 5. Conclusion and future work

The recognition of aircraft types from remote sensing images attracts much research interest. In this paper, we first review several data sets commonly used for aircraft recognition. We found that all of them either are not suitable for our domain or have significant shortcomings. This inspired us to develop and publish a new benchmark data set for aircraft type recognition of remote sensing images, MTARSI.

We then reviewed the state-of-the-art on this domain and found that, as a consequence from the lack of suitable benchmarks, most approaches were evaluated on different data sets under different experimental settings. It is thus hard to fairly compare the results published in literature.

In order to solve the problem, we make MTARSI available in the online repository [18]. MTARSI can be used to assess and evaluate the performance of aircraft type recognition algorithms for natural images. It also benefits the development of image processing, object recognition algorithms, and vision technology for remote sensing.



**Fig. 12.** Experimental results of the state-of-the-art methods on the validation set.

We then evaluate a set of representative aircraft type recognition approaches with various experimental protocols on the new data set. We found that our data set allows to clearly distinguish the performance of different methods. We also report which features of the images contained in the data set cause trouble for which approach. Researchers using MTARSI will therefore have a sound baseline of results to compare with, as well as pointers for where to look if their experimental results leave room for improvements.

In our future work, we will utilize MTARSI to develop better aircraft recognition methods. We also will further expand to collect more abundant data based on MTARSI and take into account other object types of remote sensing images.

### Declaration of competing interest

No author associated with this paper has disclosed any potential or pertinent conflicts which may be perceived to have impending conflict with this work. For full disclosure statements refer to <https://doi.org/10.1016/j.asoc.2020.106132>.

### Acknowledgments

We acknowledge support from the Youth Project of the Provincial Natural Science Foundation of Anhui, China 1908085QF285 and 1908085MF184, the Scientific Research and Development Fund of Hefei University, China 19ZR05ZDA, the Talent Research Fund Project of Hefei University, China 18-19RC26, the National Natural Science Foundation of China under grants 61673359 and 61672204, and the grant KJ2018A0555 of the key Scientific Research Foundation of Education Department of Anhui Province, China.

### References

- [1] J. Chen, B. Zhang, C. Wang, Backscattering feature analysis and recognition of civilian aircraft in TerraSAR-X images, *IEEE Geosci. Remote Sens. Lett.* 12 (4) (2015) 796–800, <http://dx.doi.org/10.1109/LGRS.2014.2362845>.
- [2] Y. Zhong, A. Ma, Y. soon Ong, Z. Zhu, L. Zhang, Computational intelligence in optical remote sensing image processing, *Appl. Soft Comput.* 64 (2018) 75–93, <http://dx.doi.org/10.1016/j.asoc.2017.11.045>.
- [3] G. Liu, X. Sun, K. Fu, H. Wang, Aircraft recognition in high-resolution satellite images using coarse-to-fine shape prior, *IEEE Geosci. Remote Sens. Lett.* 10 (3) (2013) 573–577, <http://dx.doi.org/10.1109/LGRS.2012.2214022>.
- [4] A. Zhao, K. Fu, S. Wang, J. Zuo, Y. Zhang, Y. Hu, H. Wang, Aircraft recognition based on landmark detection in remote sensing images, *IEEE Geosci. Remote Sens. Lett.* 14 (8) (2017) 1413–1417, <http://dx.doi.org/10.1109/LGRS.2017.2715858>.
- [5] G.-S. Xia, J. Hu, F. Hu, B. Shi, X. Bai, Y. Zhong, L. Zhang, X. Lu, Aid: A benchmark data set for performance evaluation of aerial scene classification, *IEEE Trans. Geosci. Remote Sens.* 55 (7) (2017) 3965–3981, <http://dx.doi.org/10.1109/TGRS.2017.2685945>.
- [6] K. Fu, W. Dai, Y. Zhang, Z. Wang, M. Yan, X. Sun, Multicam: Multiple class activation mapping for aircraft recognition in remote sensing images, *Remote Sens.* 11 (5) (2019) 544, <http://dx.doi.org/10.3390/rs11050544>.
- [7] Y. Yang, S.D. Newsam, Bag-of-visual-words and spatial extensions for land-use classification, in: D. Agrawal, P. Zhang, A. El Abbadi, M.F. Mokbel (Eds.), 18th ACM SIGSPATIAL International Symposium on Advances in Geographic Information Systems, ACM-GIS 2010, November 3–5, 2010, San Jose, CA, USA, Proceedings, ACM, 2010, pp. 270–279, <http://dx.doi.org/10.1145/1869790.1869829>.
- [8] G. Cheng, J. Han, X. Lu, Remote sensing image scene classification: Benchmark and state of the art, *Proc. IEEE* 105 (10) (2017) 1865–1883, <http://dx.doi.org/10.1109/JPROC.2017.2675998>.
- [9] W. Zhou, S. Newsam, C. Li, Z. Shao, PatternNet: A benchmark dataset for performance evaluation of remote sensing image retrieval, *ISPRS J. Photogramm. Remote Sens.* 145 (2018) 197–209.
- [10] S. Maji, E. Rahtu, J. Kannala, M.B. Blaschko, A. Vedaldi, Fine-grained visual classification of aircraft, 2013, CoRR <abs/1306.5151>. arXiv:1306.5151. URL <http://arxiv.org/abs/1306.5151>.
- [11] Q. Wu, H. Sun, X. Sun, D. Zhang, K. Fu, H. Wang, Aircraft recognition in high-resolution optical satellite remote sensing images, *IEEE Geosci. Remote Sens. Lett.* 12 (1) (2015) 112–116, <http://dx.doi.org/10.1109/LGRS.2014.2328358>.
- [12] J. Zuo, G. Xu, K. Fu, X. Sun, H. Sun, Aircraft type recognition based on segmentation with deep convolutional neural networks, *IEEE Geosci. Remote Sens. Lett.* 15 (2) (2018) 282–286, <http://dx.doi.org/10.1109/LGRS.2017.2786232>.
- [13] J.-W. Hsieh, J.-M. Chen, C.-H. Chuang, K.-C. Fan, Aircraft type recognition in satellite images, *IEE Proc., Vis. Image Signal Process.* 152 (3) (2005) 307–315.
- [14] C. Xu, H. Duan, Artificial bee colony (ABC) optimized edge potential function (EPF) approach to target recognition for low-altitude aircraft, *Pattern Recognit. Lett.* 31 (13) (2010) 1759–1772, <http://dx.doi.org/10.1016/j.patrec.2009.11.018>.
- [15] W. Diao, X. Sun, F. Dou, M. Yan, H. Wang, K. Fu, Object recognition in remote sensing images using sparse deep belief networks, *Remote Sens. Lett.* 6 (10) (2015) 745–754.
- [16] Y. Zhang, H. Sun, J. Zuo, H. Wang, G. Xu, X. Sun, Aircraft type recognition in remote sensing images based on feature learning with conditional generative adversarial networks, *Remote Sens. Lett.* 10 (7) (2018) 1123, <http://dx.doi.org/10.3390/rs10071123>.
- [17] T. Weise, X. Wang, Q. Qi, B. Li, K. Tang, Automatically discovering clusters of algorithm and problem instance behaviors as well as their causes from experimental data, algorithm setups, and instance features, *Appl. Soft Comput.* 73 (2018) 366–382, <http://dx.doi.org/10.1016/j.asoc.2018.08.030>.
- [18] Z. Wu, Multi-type Aircraft of Remote Sensing Images: MTARSI, Zenodo.org, 2019, <http://dx.doi.org/10.5281/zenodo.3464319>.
- [19] M. Jian, Q. Qi, H. Yu, J. Dong, C. Cui, X. Nie, H. Zhang, Y. Yin, K.-M. Lam, The extended marine underwater environment database and baseline evaluations, *Appl. Soft Comput.* (2019).
- [20] J. Deng, W. Dong, R. Socher, L. Li, K. Li, F. Li, ImageNet: A large-scale hierarchical image database, in: 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2009), 20–25 June 2009, Miami, Florida, USA, IEEE Computer Society, 2009, pp. 248–255, <http://dx.doi.org/10.1109/CVPRW.2009.5206848>.

- [21] Y. Li, K. Fu, H. Sun, X. Sun, An aircraft detection framework based on reinforcement learning and convolutional neural networks in remote sensing images, *Remote Sens.* 10 (2) (2018) 243, <http://dx.doi.org/10.3390/rs10020243>.
- [22] Y. Xu, M. Zhu, P. Xin, S. Li, M. Qi, S. Ma, Rapid airplane detection in remote sensing images based on multilayer feature fusion in fully convolutional neural networks, *Sensors* 18 (7) (2018) 2335, <http://dx.doi.org/10.3390/s18072335>.
- [23] G. Wang, X. Wang, B. Fan, C. Pan, Feature extraction by rotation-invariant matrix representation for object detection in aerial image, *IEEE Geosci. Remote Sens. Lett.* 14 (6) (2017) 851–855, <http://dx.doi.org/10.1109/LGRS.2017.2683495>.
- [24] F. Zhang, B. Du, L. Zhang, M. Xu, Weakly supervised learning based on coupled convolutional neural networks for aircraft detection, *IEEE Trans. Geosci. Remote Sens.* 54 (9) (2016) 5553–5563, <http://dx.doi.org/10.1109/TGRS.2016.2569141>.
- [25] Y. Tan, Q. Li, Y. Li, J. Tian, Aircraft detection in high-resolution SAR images based on a gradient textural saliency map, *Sensors* 15 (9) (2015) 23071–23094, <http://dx.doi.org/10.3390/s150923071>.
- [26] D.G. Lowe, Distinctive image features from scale-invariant keypoints, *Int. J. Comput. Vis.* 60 (2) (2004) 91–110, <http://dx.doi.org/10.1023/B:VISI.0000029664.99615.94>.
- [27] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), 20–26 June 2005, San Diego, CA, USA, IEEE Computer Society, 2005, pp. 886–893, <http://dx.doi.org/10.1109/CVPR.2005.177>.
- [28] J. Yang, K. Yu, Y. Gong, T.S. Huang, Linear spatial pyramid matching using sparse coding for image classification, in: 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2009), 20–25 June 2009, Miami, Florida, USA, IEEE Computer Society, 2009, pp. 1794–1801, <http://dx.doi.org/10.1109/CVPRW.2009.5206757>.
- [29] K. Yu, T. Zhang, Y. Gong, Nonlinear learning using local coordinate coding, in: Y. Bengio, D. Schuurmans, J.D. Lafferty, C.K.I. Williams, A. Culotta (Eds.), *Advances in Neural Information Processing Systems 22: 23rd Annual Conference on Neural Information Processing Systems 2009. Proceedings of a Meeting Held 7–10 December 2009, Vancouver, British Columbia, Canada*, Curran Associates, Inc., 2009, pp. 2223–2231.
- [30] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, in: P.L. Bartlett, F.C.N. Pereira, C.J.C. Burges, L. Bottou, K.Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems 25: 26th Annual Conference on Neural Information Processing Systems 2012. Proceedings of a Meeting Held December 3–6, 2012, Lake Tahoe, Nevada, United States*, 2012, pp. 1106–1114.
- [31] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, 2014, arXiv preprint <arXiv:1409.1556>.
- [32] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S.E. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7–12, 2015, IEEE Computer Society, 2015, pp. 1–9, <http://dx.doi.org/10.1109/CVPR.2015.7298594>.
- [33] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27–30, 2016, IEEE Computer Society, 2016, pp. 770–778, <http://dx.doi.org/10.1109/CVPR.2016.90>.
- [34] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4700–4708.
- [35] M. Tan, Q.V. Le, Efficientnet: Rethinking model scaling for convolutional neural networks, 2019, arXiv preprint <arXiv:1905.11946>.
- [36] L.M. Dang, S.I. Hassan, I. Suhyeon, A. kumar Sangaiah, I. Mahmood, S. Rho, S. Seo, H. Moon, Uav based wilt detection system via convolutional neural networks, *Sustain. Comput.: Inform. Syst.* (2018).
- [37] J.G. Ha, H. Moon, J.T. Kwak, S.I. Hassan, M. Dang, O.N. Lee, H.Y. Park, Deep convolutional neural network for classifying fusarium wilt of radish from unmanned aerial vehicles, *J. Appl. Remote Sens.* 11 (4) (2017) 042621.
- [38] S. Chakraborty, M. Roy, A neural approach under transfer learning for domain adaptation in land-cover classification using two-level cluster mapping, *Appl. Soft Comput.* 64 (2018) 508–525, <http://dx.doi.org/10.1016/j.asoc.2017.12.018>.
- [39] Z. Chai, W. Song, H. Wang, F. Liu, A semi-supervised auto-encoder using label and sparse regularizations for classification, *Appl. Soft Comput.* 77 (2019) 205–217, <http://dx.doi.org/10.1016/j.asoc.2019.01.021>.
- [40] X. Yang, H. Sun, K. Fu, J. Yang, X. Sun, M. Yan, Z. Guo, Automatic ship detection in remote sensing images from google earth of complex scenes based on multiscale rotation dense feature pyramid networks, *Remote Sens.* 10 (1) (2018) 132, <http://dx.doi.org/10.3390/rs10010132>.
- [41] K. Fu, Y. Li, H. Sun, X. Yang, G. Xu, Y. Li, X. Sun, A ship rotation detection model in remote sensing images based on feature fusion pyramid network and deep reinforcement learning, *Remote Sens.* 10 (12) (2018) 1922, <http://dx.doi.org/10.3390/rs10121922>.
- [42] J.C. Rangel, J. Martínez-Gómez, C. Romero-González, I. García-Varea, M. Cazorla, Semi-supervised 3d object recognition through CNN labeling, *Appl. Soft Comput.* 65 (2018) 603–613, <http://dx.doi.org/10.1016/j.asoc.2018.02.005>.
- [43] Y. Chen, X. Yang, B. Zhong, S. Pan, D. Chen, H. Zhang, Cnntracker: online discriminative object tracking via deep convolutional neural network, *Appl. Soft Comput.* 38 (2016) 1088–1098, <http://dx.doi.org/10.1016/j.asoc.2015.06.048>.
- [44] Z. Yan, M. Yan, H. Sun, K. Fu, J. Hong, J. Sun, Y. Zhang, X. Sun, Cloud and cloud shadow detection using multilevel feature fused segmentation network, *IEEE Geosci. Remote Sens. Lett.* 15 (10) (2018) 1600–1604, <http://dx.doi.org/10.1109/LGRS.2018.2846802>.
- [45] X. Gao, X. Sun, Y. Zhang, M. Yan, G. Xu, H. Sun, J. Jiao, K. Fu, An end-to-end neural network for road extraction from remote sensing imagery by multiple feature pyramid network, *IEEE Access* 6 (2018) 39401–39414, <http://dx.doi.org/10.1109/ACCESS.2018.2856088>.
- [46] M. Larsson, Y. Zhang, F. Kahl, Robust abdominal organ segmentation using regional convolutional neural networks, *Appl. Soft Comput.* 70 (2018) 465–471, <http://dx.doi.org/10.1016/j.asoc.2018.05.038>.
- [47] M. Mittal, L.M. Goyal, S. Kaur, I. Kaur, A. Verma, D.J. Hemanth, Deep learning based enhanced tumor segmentation approach for MR brain images, *Appl. Soft Comput.* 78 (2019) 346–354, <http://dx.doi.org/10.1016/j.asoc.2019.02.036>.
- [48] T. Lin, M. Maire, S.J. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C.L. Zitnick, Microsoft COCO: common objects in context, in: D.J. Fleet, T. Pajdla, B. Schiele, T. Tuytelaars (Eds.), *Computer Vision – ECCV 2014 – 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part V*, in: *Lecture Notes in Computer Science*, 8693, Springer, 2014, pp. 740–755, [http://dx.doi.org/10.1007/978-3-319-10602-1\\_48](http://dx.doi.org/10.1007/978-3-319-10602-1_48).
- [49] S.J. Pan, Q. Yang, A survey on transfer learning, *IEEE Trans. Knowl. Data Eng.* 22 (10) (2010) 1345–1359, <http://dx.doi.org/10.1109/TKDE.2009.191>.
- [50] J.R. Dim, T. Takamura, Alternative approach for satellite cloud classification: edge gradient application, *Adv. Meteorol.* 2013 (2013).
- [51] J.M. Prewitt, Object enhancement and extraction, *Pict. Process. Psychopictoris* 10 (1) (1970) 15–19.
- [52] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R.B. Girshick, S. Guadarrama, T. Darrell, Caffe: Convolutional architecture for fast feature embedding, in: K.A. Hua, Y. Rui, R. Steinmetz, A. Hanjalic, A. Natsev, W. Zhu (Eds.), *Proceedings of the ACM International Conference on Multimedia, MM ’14, Orlando, FL, USA, November 03 – 07, 2014*, ACM, 2014, pp. 675–678, <http://dx.doi.org/10.1145/2647868.2654889>.
- [53] G.-S. Xia, W. Yang, J. Delon, Y. Gousseau, H. Sun, H. Maître, Structural high-resolution satellite image indexing, in: *ISPRS TC VII Symposium-100 Years ISPRS*, Vol. 38, 2010, pp. 298–303.
- [54] Q. Zou, L. Ni, T. Zhang, Q. Wang, Deep learning based feature selection for remote sensing scene classification, *IEEE Geosci. Remote Sens. Lett.* 12 (11) (2015) 2321–2325, <http://dx.doi.org/10.1109/LGRS.2015.2475299>.
- [55] M.D. Zeiler, R. Fergus, Visualizing and understanding convolutional networks, in: D.J. Fleet, T. Pajdla, B. Schiele, T. Tuytelaars (Eds.), *Computer Vision – ECCV 2014 – 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part I*, in: *Lecture Notes in Computer Science*, 8693, Springer, 2014, pp. 818–833, [http://dx.doi.org/10.1007/978-3-319-10590-1\\_53](http://dx.doi.org/10.1007/978-3-319-10590-1_53).