2020 IEEE 2nd International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS) | 978-1-7281-6921-7/20/$31.00 ©2020 IEEE | DOI: 10.1109/ICECOCS50124.2020.9314611

Aircraft Type Classification in Remote Sensing Images using Deep Learning

Youssef Ben Youssef
Interdisciplinary Laboratory of Applied Sciences
National School of Applied Sciences,
Berrechid, Morocco
youssef.benyoussef@uhp.ac.ma

Mohamed Merrouchi
Lab computing, imaging and modeling of complex systems
FST of Settat, Hassan first University
Settat, Morocco
merrouchi37@gmail.com

Elhassane Abdelmounim
Lab of Signal Analysis and Information Processing
FST of Settat, Hassan first University
Settat, Morocco
hassan.abdelmounim@hotmail.fr

Taoufiq Gadi
Lab computing, imaging and modeling of complex systems
FST of Settat, Hassan first University
Settat, Morocco
gtaoufiq@yahoo.fr

***Abstract*—In aeronautic, redundancy of data is strongly desired to make optimal decisions. A large data source is available with the development of the technology used in remote sensing images. Remote Sensing Image Classification (RSIC) is widely exploited in military and civil fields. To improve the performance of multi-label classification, we addressed the problem of RSIC based on the Convolutional Neural Network (CNN) for remote sensing images of aircraft types. Previous studies have used intensive preprocessing which limits the rate of classification. We improved the network structure to make it more accurate and to limit underfitting or overfitting problems. A recent public dataset called Multi-Type Aircraft Remote Sensing Images (MTARSI) is used in this work to validate our method. Extensive experiments prove the effectiveness of the proposed method in terms of accuracy.**

***Index Terms*—Computer vision, Machine learning, Deep learning, Convolutional Neural Network, Classification.**

I. Introduction

Image classification is one component of computer vision and machine learning. Assigning automatically predefined labels to images is the aim of image classification. One of the important issues in remote sensing image processing is aircraft type classification, and it has been widely used in civil and military applications.

978-1-7281-6921-7/20/$31.00 ©2020 IEEE,

Machine learning (ML) has emerged as one of the most successful artificial intelligence techniques and has achieved impressive performance in the field of computer vision and image processing, with applications such as image medical classification [1]- [2] and remote sensing image scene classification [3]. All algorithms in ML are based on many handcrafted features available from images for doing the classification; those methods are named also handcraft features in classification. Recently in remote sensing image classification, DL is taking off most, and a growing number of relative papers are reported in the literature. DL has been widely applied in diverse study areas,including vegetated areas, urban areas, wetlands, and forest areas [4]. As a result, the details of ground objects, such as contour, structure, and texture information, can be obtained conveniently. Among DL algorithms used in classification, CNNs have gained popularity. Since 2012, CNN has attracted more attention because of the increasing computing power, availability of lower-cost hardware, open-source algorithms, and the rise of big data [5]. A typical trend is that the CNNs are getting deeper. By increasing depth, CNN can approximate the target function with increased nonlinearity and get better descriptor representations. However, the complexity of the network is increasing, which makes the network be more difficult to optimize and easier to get underfitting or overfitting.The main contributions of this work can be summarized as follows:

- CNN architecture is built in order to performed better accuracy;

- We optimize the parameters of the model to overcome the difficulties of underfitting and overfitting encountered in training and testing our model.

The remained of this work is organized as follows. In Section II, we briefly review the related work. We present Dataset MTARSI and CNN model of deep learning in Section III, whereas Section IV results and discussion are largely explained. The paper is end by conclusion in section V.

II. Related Work

Initially, the modern framework of CNN is presented in [6]. The authors developed a multi-layer artificial neural network called LeNet-5 which could classify handwritten digits [7]. Later, many CNN models for classifying images have been proposed and considered in different issues for images. Previous research projects are studied for three typical CNN application cases in remote sensing image classification: scene classification, object detection and object segmentation are presented [8]. Due to state-of-the-art image classification using CNN such as VGG [9], GoogLeNet [10], ResNet [11], DenseNet [12], and E?cientNet [13] which were successfully applied to ImageNet dataset. Aircraft recognition from remote sensing images are focused on deciding whether an object is an aircraft or not based on reinforcement learning and convolutional neural networks [14].Wu and Prasad proposed convolutional recurrent neural network, in which a few convolution layers are followed by recurrent layers. Middle-level and locally invariant features are extracted from raw HSI and spectrally contextual features are then extracted from the features generated by convolution layers [15]. Fu et al. propose a fine-grained aircraft recognition method for remote sensing images. Their multi-class activation mapping uses two subnetworks, i.e., the target net and the object net, in order fully use the features of discriminative object parts [16]. Zhao et al. proposed the aircraft type recognition problem by detecting the landmark points of an aircraft using a vanilla network designed a keypoint detection model based on CNNs and a keypoint matching method to recognize aircraft, transforming the aircraft recognition problem into the landmark detection problem [17]. All the work cited above has been trained and tested using different dataset. The only common point is scene classification with intensive preprocessing which consequently affects the performance of the systems proposed. Zhi-Ze Wu et al. investigate the performance of five state-of-the-art CNN structures namely VGG, GoogLeNet, ResNet, DenseNet, and EfficientNet and obtained the result that EfficientNet is better than others in tern of accuracy [18]. Marmanis et al. used a pretrained CNN by the ImageNet challenge and exploited it to extract an initial set of representations for earth observation classification [19]. These methods for object type identification from remote sensing images have achieved significant results, but there are still many challenges. In the literature, there are numerous variants of CNN architectures with a huge parameters that must be adapted for each case studied.

III. Proposed Approach

In this section, information about the dataset MTARSI used in this work is given. Next, a detailed description of the evaluated network architecture is provided.

A. Dataset

Database existing in literature are categorized into three groups based on three kinds of classification tasks : scene classification, object detection and object segmentation. Among this Dataset, MTARASI dataset has the advantage that labeled images contain a single type of aircraft in different orientations. The MTARSI dataset was used for training and testing our proposed method. It is an open-source dataset for Aircraft Type Recognition from Remote Sensing [18]. MTARSI dataset has a total of 9,385 remote sensing images acquired from Google Earth satellite imagery and manually expanded. It contains 20 different types of aircraft and different sizes covering 36 airports. Each image contains exactly one complete labeled aircraft. The spatial resolution of the images varies in range 0.3 to 1 m and they contain various orientations, aspect ratios, and pixel sizes of the objects. In addition, the images vary according to the altitude, nadir-angles of the satellites, and the illumination. Some image patches have some cropped objects, and some examples are black and white panchromatic images. These variations in the MTARSI enable the trained aircraft classification architectures to achieve similar performance in different image conditions. The aircraft may differ on type and model as illustrated in Figure 1, where a sample of image aircraft types extracted from MTARSI dataset is depicted.

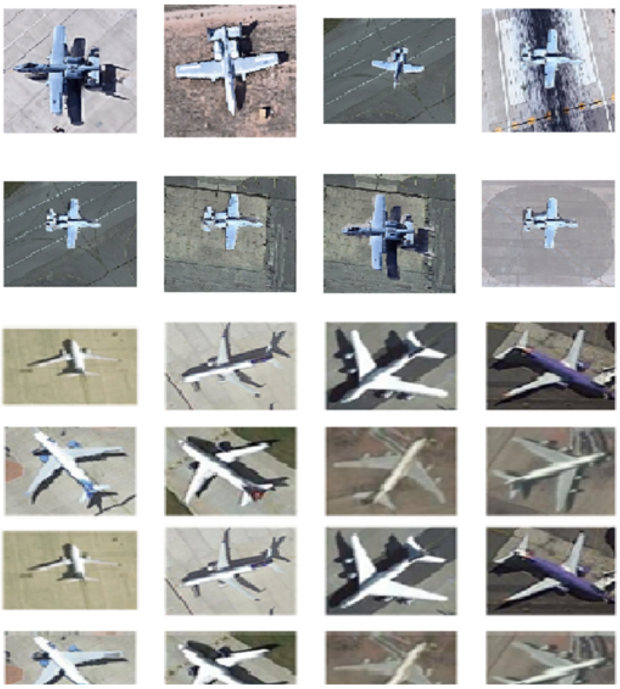


Fig. 1. Patches from the MTARSI dataset.

B. Convolutional Neural Network Architecture

Fig. 2. Network model input output.

A neural network can be viewed as a mathematical mapping of input values x to output values y. In supervised classification, the function ψ assigns an input data to a given set of pre-defined classes in the output. Classification problem can be described slightly more formally, given a training set, our objective is to learn a function ψ: called a hypothesis, so that $\psi(x)$ is an optimal predictor for the corresponding value of y. In dataset MATRASI, we have extracted N = 8779 total images. Each image is 200×200 pixels, represented in the RGB color space (i.e,. three channels per image).We can represent each image as $D = 200 \times 200 \times 3 = 120.000$ distinct values. Finally, we know there are a total of C=20 class labels:

$$y = \psi(x) \tag{1}$$

where $y = \{c_1, c_2, \dots, c_{20}\}$ and $x \in \{0, 1, 2, \dots, 120.000\}$ The loss function is the cross entropy between the predicted probability and the true label y defined by

$$L(y, \hat{y}) = \sum_i y_i . \log \hat{y}_i \tag{2}$$

where $\hat{y}$ is the predicted probability vector (Softmax output), and y is the ground-truth vector. Loss function is used in the training process to find the parameter values for model proposed. The loss is returned on training and testing process and its interpretation is how well the model is doing for these two sets.

DL architectures have developed and have been applied in different fields such as audio recognition [20], natural language processing [21], and many classification tasks. In addition, as the most representative supervised DL model,CNNs have performed most algorithms in visual recognition. CNNs take advantage of the input image structure and define a network architecture in a more sensible way. The structure of CNNs allows the model to learn highly abstract feature detectors and to map the input descriptors into representations that can clearly outperform the performance of the subsequent layers. All convolutional layers contain filters with a spatial support of 3×3. While a larger filters is used in older deep models. For example, the AlexNet [5] architecture contains filters of 11×11, the recent trend is towards using smaller filters, e.g. the ResNet architecture does not contain filters larger than 3×3. The main advantage of the CNN is its flexibility to add or reduce the number of hidden layers in its structure for a given task. Furthermore, there are many optional techniques that can be used to train it. The CNN architecture is depicted in Figure 3. Generally, a CNN mainly consists of three parts: convolution layers, pooling layers, and fully connected layers.

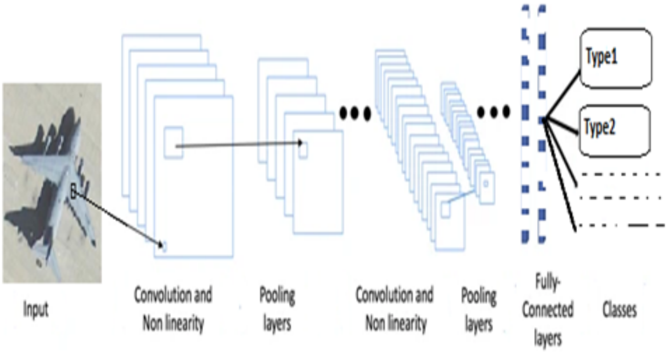


Fig. 3. CNN architecture.

- Convolutional layers: In convolution layers, the input maps are convolved with learnable kernels and are subsequently put through the activation function to form the output feature maps. After each convolution, operation the nonlinear layer is added. It has a nonlinear activation function. Without this property, a network would not be sufficiently intense and will not be able to model the response variable (as a class label). The convolutional layer computes the output feature map :

$$y^s = \sum_{i=1}^{q} W_i^s \star x^i + b \tag{3}$$

where $*$ is a two-dimensional discrete convolution operator, W is weight and b is a trainable bias parameter. To improve our classification accuracy, the parameters weight W or bias vector b are adjusting in training process.
- Nonlinear layer : in this layer, a nonlinear function applied to each component of a mapped component. The nonlinear layer is added after each convolution operation. It has an activation function, which brings nonlinear property. Rectified Linear Unit (ReLU) is commonly chosen and defined as:

$$ReLU(x) = max(0, x) \tag{4}$$

The output of ReLu is the maximum value between zero and the input value. The main advantage of using the ReLU function over other activation functions is that it does not activate all the neurons at the same time and ReLU train several times faster than their equivalents with other units.
- Pooling layer: it follows a convolutional layer and it is used to reduce the dimensionality of feature maps and improve the robustness of the extracted features. This allows us to reduce the number of parameters, which both shortens the training time and avoid overfitting. It is usually placed between two convolutional layers. There are two types of basic pooling operation which are the most commonly used: average pooling and max pooling, Detailed theoretical analysis can be found in [5]. Max Pooling returns the maximum value from the portion of the image covered by the Kernel
- Fully connected layer: the output maps of the last convolution layer or pooling layer are flattened into vectors, serving as the inputs to the first fully connected layer. The output of the final fully connected layer is the learned feature, forming the result of which is extracted from the input image by the convolutional network. In the training of the model, the flattened output is fed to a feed-forward neural network whose backpropagation algorithm is applied to each iteration. Over a fixed number of epochs, the model is able to distinguish between dominating and certain low-level features in images and classify them using the softmax classification technique. Softmax is usually used in the multi-classification tasks defined below :

$$f(x_i) = \frac{\exp(x_i)}{\sum_j^n \exp(x_j)} \tag{5}$$

Softmax function returns value in the range [0,1],it can be viewed as form of a probability distribution. It defines a flexible learning task with adjustable margin [22].

IV. Experiments Results and Discussions

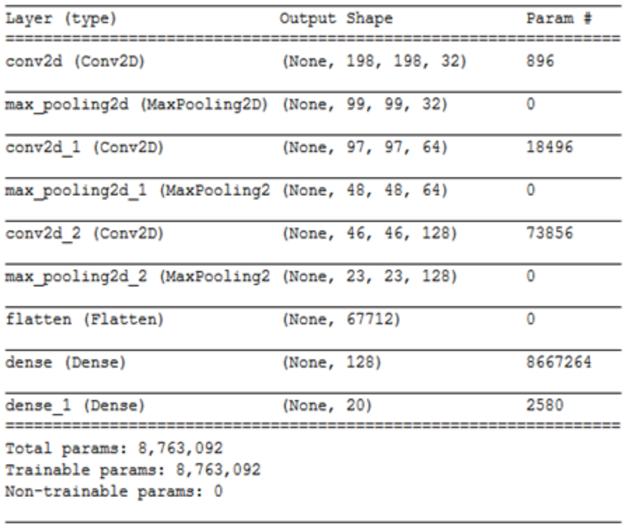
All the experiments were implemented on a PC with an HP i5 8th generation CPU processor 1.80GHz, 4 Go RAM, x64, using Python 3.7 for Windows 10 in Keras environment.

A. Data Preprocessing

In this study, 8779 images were extracted from the dataset MTARSI categorized in 20 types of aircraft. Before training our CNN, all image were resized to the size of 200×200 patches because all image are in different sizes; and normalized using ImageDataGenerator function, in which we divided all the pixels of the images by 255 to range min-max values between 0 and 1. Once we charged images from, we feed them into a CNN that does classification.

B. Building the model

After data preprocessing process,the model builded contains 3 convolution layers paired with a batch normalization, 3max-pooling layers and a fully connected layer with 2 hidden layers. Each three convolution layers is accompanied with a pooling layer. there are 20 final output nodes at the end of the CNN. Each node represents an aircraft type. The CNN model proposed in this work is shown in Figure 4. After several experiences as aim to improve accuracy and avoid overfit and underfit the CNN structure built has 3 convolution layers paired with a batch normalization and 3 max pool layers followed by 2 fully connected layers. The output is a score matrix



```
Layer (type)                 Output Shape              Param #
=================================================================
conv2d (Conv2D)              (None, 198, 198, 32)      896
_________________________________________________________________
max_pooling2d (MaxPooling2D) (None, 99, 99, 32)        0
_________________________________________________________________
conv2d_1 (Conv2D)            (None, 97, 97, 64)        18496
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 48, 48, 64)        0
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 46, 46, 128)       73856
_________________________________________________________________
max_pooling2d_2 (MaxPooling2 (None, 23, 23, 128)       0
_________________________________________________________________
flatten (Flatten)            (None, 67712)             0
_________________________________________________________________
dense (Dense)                (None, 128)               8667264
_________________________________________________________________
dense_1 (Dense)              (None, 20)                2580
=================================================================
Total params: 8,763,092
Trainable params: 8,763,092
Non-trainable params: 0
_________________________________________________________________
```

Fig. 4. The adopted CNN model.

for the weights for each class. Each filter has a kernel size of 3 × 3.

C. Training and Testing Process

After building the CNN, it was trained on 8,779 images for 20 epochs with batch size of 128, compiled with categorical crossentropy loss function and RMS optimizer with the learning rate 0.0001. RMSprop is a gradient-based optimization technique; it was developed as a stochastic technique for mini-batch learning[23]. The accuracy is one metric used to measure how often the algorithm classifies an image correctly. After training process,the CNN was tested on 2,122 images selected randomly in file image training dataset. The following graphs in figures 4 and 5 show the accuracy and loss vs the number of epochs in training and testing model. As we can see in Figure 5, the accuracy plots at each epoch shows that our model suffer little from overfitting. Beyond a epoch 10, the test accuracy is slightly lower than the training accuracy. This implies that our model proposed obtains better performance with less complexity when we choose . The highest test accuracy is reached from epoch 10 and does not increase. As figure 6 shows, beyond epoch 10, the loss test and train decrease. The highest test accuracy at all the epochs is reported as the best score 99.90%.

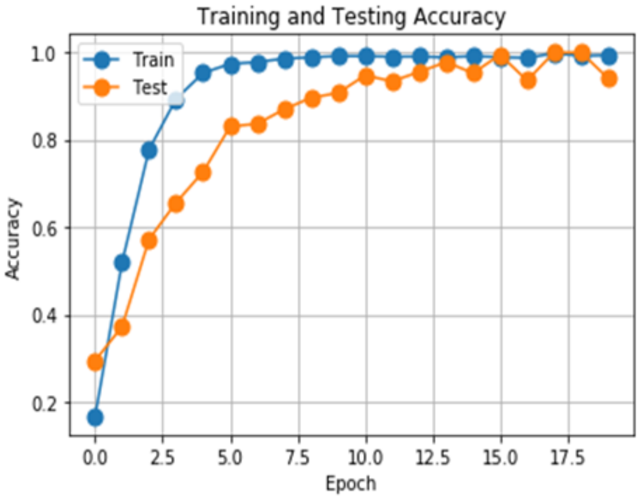


Fig. 5. Training and testing accuracy with epochs.

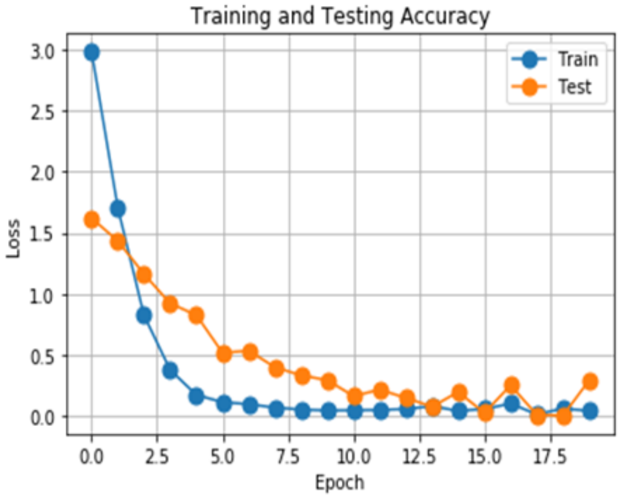


Fig. 6. Loss training and testing with epochs.

A comparison of the proposed approach with the other approaches developed by researchers using the same MTARSI dataset in [18] is presented in Table 3 in term of average accuracy related to all epoch.

TABLE I
COMPARISON RESULT OF PROPOSED SYSTEM

Method	*Accuracy(%)*
VGG	87.56
GoogLeNet	86.53
ResNet	89.61
DenseNet	89.15
EfficientNet	89.79
Our approach	90.66

The all average accuracy (90,66%) obtained is a good result on MTARSI dataset. Among all these methods, our approach is performance with less complexity architecture off CNN.

V. CONCLUSION

In this work, we have presented an approach based on convolutional neural networks for the multi-classification of aircraft type.The empirical results indicated that our approach provides superior results on MTARSI data sets. Accuracy obtained (90,66%) is a good result with only normalization data augmentation and without model regularisation. In our future work, we will utilize other techniques such as data augmentation to add more training images and develop better aircraft classification methods.

REFERENCES

[1] Y.BenYoussef, E.Abdelmounim and A.Belaguid, “Mammogram Classification using Support Vector Machine,” in handbook of researcher on advanced trends in microwave and communication theoretical and applied information technology, Ed. IGI Global, pp. 587-614,2017

[2] Y.BenYoussef,E.Abdelmounim,J.Zbitou,M.Elharoussi and M.N. Boujeda,“Comparison machine learning algorithms in abnormal mammograms classification,” in IJCSNS.Vol.17 No.5, pp.19-25, May 2017.

[3] G. Cheng, C. Yang, X. Yao, L. Guo and J. Han, “When Deep Learning Meets Metric Learning: Remote Sensing Image Scene Classification via Learning Discriminative CNNs,” in IEEE Transactions on Geoscience and Remote Sensing, vol. 56, no. 5, pp. 2811-2821, May 2018, doi: 10.1109/TGRS.2017.2783902..

[4] L.Maa,Y.Liuc,X.Zhang,Y.Ye,G.Yind, and B.A.Johnson, “Deep learning in remote sensing applications:Ameta-analysis and review,” ISPRS Journal of Photogrammetry and Remote Sensing 152, pp.166-177,2019

[5] A.Krizhevsky, I.Sutskever,GE. Hinton, “Imagenet classification with deep convolutional neural networks,”In Advances in neural information processing systems 25, NIPS,2012.

[6] Y.Le Cun,B.Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, L. D. Jackel, “Handwritten digit recognition with a back-propagation network,” in Proceedings of the Advances in Neural Information Processing Systems (NIPS), pp.396-404,1989.

[7] Y.LeCun,L.Bottou,Y.Bengio,P.Haffner, “Gradient-based learning applied to document recognition,” Proceedings of IEEE 86 (11),1998.

[8] L. Zhang, L.Zhang, and B.Du,“Deep Learning for Remote Sensing Data : A technical tutorial on the state of the art,” IEE Geoscience and remote sensinG magazine,pp.22-40,2016.

[9] K. Simonyan, A. Zisserman,“Very deep convolutional networks for large-scale image recognition,” in Proceedings of the International Conference on Learning Representations (ICLR), 2015. arXiv preprint arXiv:1409.1556, 2014b.

[10] C. Szegedy et al., “Going deeper with convolutions,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, 2015, pp. 1-9, doi: 10.1109/CVPR.2015.7298594.

[11] K. He, X. Zhang, S. Ren, J. Sun, “Deep residual learning for image recognition,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770-778.

[12] 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.

[13] M. Tan, Q. V. Le,“ Efficientnet: Rethinking model scaling for convolutional neural networks,” arXiv preprint arXiv:1905.11946.

[14] Y. Li, K. Fu, H. Sun and X. Sun, “An aircraft detection framework based on reinforcement learning and convolutional neural networks in remote sensing images,” Remote Sensing 10 (2) (2018) 243. doi:10.3390/rs10020243.

[15] H.Wu, S.Prasad,“Convolutional Recurrent Neural Networks for Hyperspectral Data Classification”. Remote Sens. 2017, 9, 298. doi.org/10.3390/rs9030298.

[16] K.Fu, W.Dai,Y.Zhang, Z.Wang,M.Yan and X.Sun,“ MultiCAM: Multiple Class Activation Mapping for Aircraft Recognition in Remote Sensing Images,” Remote Sens. 2019, 11, 544. doi:10.3390/rs11050544

[17] 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 Geoscience and Remote Sensing Letters 14 (8) 2017.pp.1413-1417.doi:10.1109/LGRS.2017.2715858.

[18] Z.Z.Wu, S.H..Wan, X.Fang et al. “A benchmark data set for aircraft type recognition from remote sensing images,” Applied Soft Computing Journal (2020) .doi.org/10.1016/j.asoc.2020.106132.

[19] D. Marmanis, M. Datcu, T. Esch, and U. Stilla, “Deep learning earth observation classification using ImageNet pretrained networks,” IEEE Geosci. Remote Sens. Lett., vol. 13, no. 1, 2016.pp. 105–109.

[20] A.R. Mohamed, T. N. Sainath, G. Dahl, B. Ramabhadran, G. E. Hinton, and M. A. Picheny, “Deep belief networks using discriminative features for phone recognition,” in Proc. IEEE Int. Conf. Acoust. Speech, and Signal Processing, Prague, Czech Republic, 2011, pp. 5060–5063.

[21] R. Collobert and J. Weston, “A unified architecture for natural language processing: Deep neural networks with multitask learning,” in Proc. Int. Conf. Mach. Learning, Helsinki, Finland, 2008, pp. 160–167.

[22] Liu, W., Wen, Y., Yu, Z., and Yang, M.“Large-margin softmax loss for convolutional neural networks,” ICML, 2(3), 7.

[23] J.Xu Z. Zhang,Triedman,Y.Liang and G.Van den Broeck,“A Semantic Loss Function for Deep Learning with Symbolic Knowledge,”2016, available in https://arxiv.org/pdf/1609.04747.