Species Recognition of *Aspergillus* Conidia Using Convolutional Neural Networks in Scanning Electron Microscopy Imagery

Huaizhong Zhang[1] and Marta Filipa Simões[2,3]

[1]Dept. Computer Science, Edge Hill University, Ormskirk, United Kingdom
[2]State Key Laboratory of Lunar and Planetary Sciences, Macau University of Science and Technology, Taipa, Macau SAR, China.
[3]Macau Center for Space Exploration and Science, China National Space Administration (CNSA), Macau SAR, China.
Email: zhangh@edgehill.ac.uk, msimoes@must.edu.mo

***Abstract*—This paper presents a practical recognition method based on deep learning techniques, for fungal species, through Scanning Electron Microscopy (SEM) images. A small number of images was acquired. To circumvent the issue of not having many samples, a method of generating the training set is proposed to increase target signatures and optimize the baseline quality of inputs for object recognition. To tackle the challenge of detecting varied scale targets, a sophisticated and powerful Convolutional Neural Network (CNN) based on faster region R-CNN, with the prepared training dataset, was trained. In this study, the datasets for five different species of *Aspergillus* were previously collected via SEM. The proposed method is applied to identify the spore structures – conidia – in the images so as to recognize the species respectively. The initial experimental results show that the developed method can qualitatively and quantitatively identify the relevant species effectively, being of major importance for the development of easier diagnostic and identification tools in mycology.**

***Index Terms*—*Aspergillus* conidia, convolutional neural network, object recognition, scanning electron microscopy**

I. Introduction

In the study of fungi, the species of *Aspergillus* from section *Nigri* (known as black aspergilli) are of great interest in a variety of biotechnological applications, food science and clinical mycology [1], [2]. They consist of at least 27 different taxa in total [3], [4].

Some of the species within this section are associated with infections in animals, humans and plants. These later can have great impact on human health since they can affect food availability. Others have been reported to be opportunistic pathogens to humans (e.g., *A. tubingensis* [5]).

Within the medical context, *A. niger* is the most described species within the section and one of the causative agents of aspergillosis [6], a disease many times misdiagnosed as tuberculosis [7], [8]. Furthermore, this section of aspergilli has been reported as the most common etiological agent of aspergillosis, but systematic studies on its antifungal susceptibility profile are still missing. Within most of the reported cases, these fungi are not identified to species level. And, even though many novel species have been described, especially with the development of molecular biology, identification is complex and misidentification is a problem [9].

In microbiology, the identification of fungal species is a fundamental task which is time-consuming and very laborious because it implies the collection of large amounts of information about the fungal strains (e.g., morphology, physiology, biochemistry, and ecological roles). Over the years, efforts have been made to facilitate identification processes. Scanning Electron Microscopy (SEM), is one of the techniques applied to characterise fungal species and to analyse their ornamentations, sizes and shapes as shown in Fig. 1. The SEM images of fungal strains provide the geometric properties and structural information of each species and it has been found that each species has its own particular and characteristic spores' morphology. Since some species, like section *Nigri* here investigated, are phylogenetically close [5], making taxonomy and identification a very difficult task; it is imperative to find alternatives to circumvent this. Thus, pattern based methods in image processing can be naturally applied to automatically recognize these species by supervised learning.

Recently, alternative techniques to traditionally artificial neural networks have been proposed for a variety of applications. Particularly, Convolutional Neural Networks (CNNs) [10], [11] show unrivalled success in object recognition that is credited to some new localizing formulations proposed, instead of the obsolete sliding-window detector such as selective search [12], and region proposal network [13]. To date, a prevalent CNN framework [13] for object recognition is as follows: first, convolutional layers are employed to acquire region based features for detecting objects of interest; then a region-wise Multi-Layer Perceptron (MLP) classifier is followed to do the classification for recognizing objects. Therefore, this architecture provides a feasible way to

Manuscript received August 12, 2021; revised November 23, 2021.
Corresponding author: Huaizhong Zhang.

doi: 10.18178/ijpmbs.11.1.8-13

identify fungal species, which is many times a difficult and long task [14], by training an effective model. This paper presents a CNN based species recognition method for black aspergilli. In this initial study, we worked with images from five aspergilli species: *A. aculeatus*, *A. brasiliensis*, *A. carbonarius*, *A. ellipticus* and *A. lacticofeatus*. Fig. 1 illustrates two images collected for our experiment. Due to the experimental restrictions, a small number of images were collected for these five species (see Table I). To obtain the effective training set, we made use of the geometric conidium information, from the collected species image, so as to generate the ground truth and apply the affine transformation to augment the training set. Then, a learning model was trained by using the faster R-CNN techniques [13] and thus the species in *Aspergillus* images can be recognized by applying this obtained model. This paper shows the effort on how to use the CNN based techniques to identify fungal biological structures without a large dataset support, which the sample culturing and SEM imaging usually cost a lot both in time and expenses.

In the following section, we introduce the proposed method. Section III presents the relevant experimental results. The paper is concluded in Section IV.

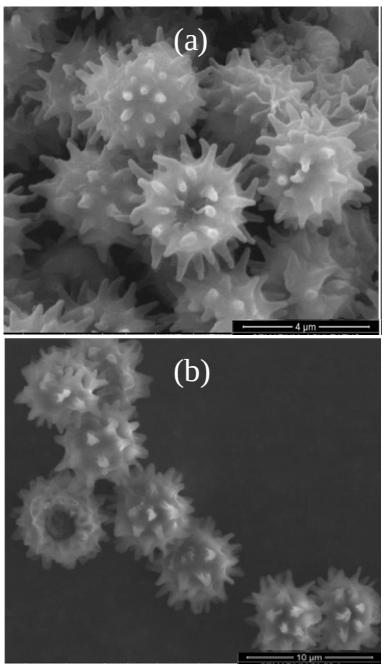


Figure 1. Examples of collected SEM images for *A. aculeatus* (a) and *A. carbonarius* (b) [3].

Your goal is to simulate the usual appearance of papers in the. We are requesting that you follow these guidelines as closely as possible.

II. Method

The proposed recognition method includes data collection, ground truth creation and augmentation, and model training using the faster R-CNN [13], [14]. These stages are unified into an overall CNN framework that is implemented based on the Caffe [15] development environment. The main contribution lies in the attempt that the CNN based techniques are able to be applied to detect the fungal biological structures via a small number of collected images (see Table I).

TABLE I. Number of the SGSs Generated for the Figures of Each Species

	A. ellipticus	*A. lacticofeatus*	*A. carbonarius*	*A. brasiliensis*	*A. aculeatus*
No. of images collected per species	4	6	6	4	3
No. of SGSs in the species' images	25	37	42	28	23
No. of SGSs transformed	325	481	546	364	299
No. of SGSs in total	350	518	588	392	322

A. Data Collection and Training Set Generation

Images from five strains belonging to *Aspergillus* section *Nigri* were selected. Briefly, for the collection of SEM data, the selected strains were inoculated on a drop of Malt Extract Agar (MEA), set on a metal SEM stub, and incubated for 7 days at 27 °C in the dark, as described in [3], [14]. The collected images are TIFF format and their size is 1024x943. In the acquired SEM data, each image presents a number of species conidia where each conidium represents one Species Geometric Structure (SGS). Thus, these SGSs are the fingerprint for recognizing the corresponding species. Table I lists the number of SGSs collected from the augmented datasets for each species. A variety of geometric transformations were used for the data augmentation, including flip, rotation, scaling, crop, translation, etc. Fig. 2(a) shows that we can obtain different transformed SGSs from an *A. aculeatus* image. These SGSs present the geometric properties of the conidia from *A. aculeatus* species. Thus, we used these SGSs of *A. aculeatus* to generate its ground truth via LabelImg [16]. The annotations were saved as XML files, i.e. Pascal VOC format [17]. Likewise, we created the ground truth for other four species of *Aspergillus*. Fig. 2(b) presents four generated examples from a SGS. Thus, the number of the ground truths generated for each species ranged from 350 to 588. The ratios, 7:2:1, were used to divide the total SGSs into the sets for training, testing, and validating, respectively. It is worth noting that we have simulated the background images in terms of the SEM experimental environment. Thus, the synthetic background is similar to the one in Fig. 1.

B. Convolutional Neural Networks

The findings on how the mechanisms in the visual cortex of the brain work, have successfully driven the CNN designing to deal with pattern based problems.

Similar to a traditional neural network architecture, a CNN is made up of layers that aim to obtain a set of locally connected neurons, between two layers, by learning data-specific kernels. Three main types of layers are employed to build a CNN model: convolutional layer, pooling layer and fully-connected layer. A typical CNN for object recognition is illustrated in Fig. 3. The input is an image and the output is a single vector of class scores. The role of each type of layer can be described as follows:

- Convolutional layer will generate a volume of feature maps, by computing a dot product between their weights used and the region connected to the input volume.
- Pooling layer will down-sample the feature map along the spatial dimensions.
- Fully-connected layer will create the class scores according to the given categories.

In addition, for making the CNN model robust, ReLu layer was used to apply an elementwise activation function. A dropout strategy was used, to perform the action of randomly ignoring neurons, for preventing inter-dependencies between neurons.

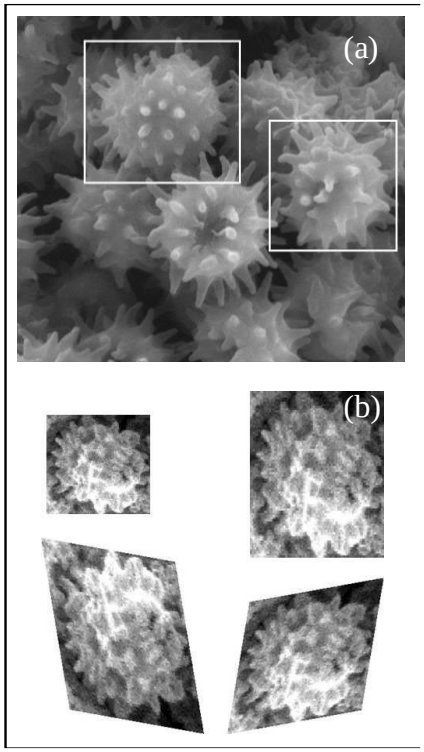


Figure 2. Generation of ground truth: cropped conidium structures (SGSs) for *A.* aculeatus (a) [3] and SGS examples obtained by the affine transformation (b).

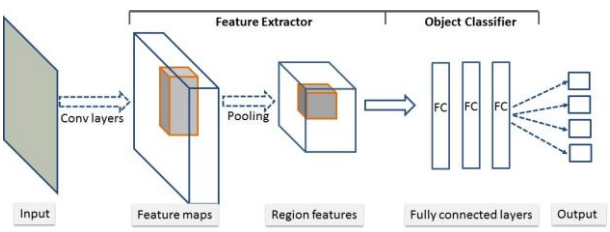


Figure 3. A typical CNN architecture.

C. *Object Detection and Recognition Using Faster-RCNN*

Faster R-CNN [13], a most recently developed object detection system, aims to integrate traditional region proposals and object detector/classifier into one CNN. It is composed of two main components. The first one is Region Proposal Networks (RPN), a fully convolutional network, that produces region proposals where the objects within are similar. The second one is Fast R-CNN [18] that uses the proposed regions to do classification and make a final decision on the existence of those objects. Our ATD/R system employs Faster-RCNN to carry out object detection and recognition. The following subsections introduce how Faster R-CNN works.

1) *RPN for generating region proposals*

This region proposal network is constructed as a Fully Convolutional Network (FCN) [19]. It produces region bounds and objectiveness scores simultaneously at each location. The RPN architecture (Fig. 4) is actually composed of a n×n convolutional layer (L1, n = 3 used), two sibling 1×1 convolutional layers for box regression (reg), and box classification (cls), respectively.

- (L1 layer): a n×n spatial window of the feature map of the last shared convolutional layer is input into this layer for generating region proposals. At each sliding position, k (= 9), region proposals (the green part in Fig. 4) are created by using 3 scales and 3 aspect ratios. These k proposals are mapped to a feature vector that is fed into the reg layer and cls layer.
- (reg layer): this layer performs the regression process for the input feature vector in terms of each sliding position. The outputs are the coordinates of k region proposals.
- (cls layer): this layer estimates the object probability for each region proposal. The outputs are the scores of each proposal.

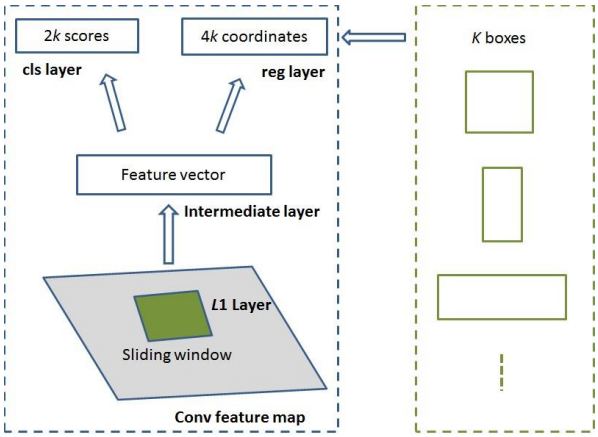


Figure 4. Region Proposal Network (RPN).

2) *Fast RCNN object detection network*

Fast R-CNN is an improved version of R-CNN [20] for accelerating the detection process. It uses bounding box proposal methods [12] to create bounding boxes. Then, Region of Interest (RoI) pooling is applied to generate a feature vector for each bonding box. Afterwards, the

feature vector is input to a 2-layer regression network and a classification network for fine-tuning the bounding boxes and obtaining class scores. Finally, Non-Maxima Suppression (NMS) is applied over all boxes to eliminate the redundant bounding boxes.

3) Model training

For creating region proposals, the RPN was trained end-to-end by back propagation and Stochastic Gradient Descent (SGD). For recognizing objects, fast R-CNN was adopted and was trained independently. In Faster R-CNN, a unified network is learnt from RPN and fast R-CNN by sharing convolutional layers. This was implemented by a pragmatic 4-step training algorithm as follows:

i) The RPN is trained as above, which is initialized with an ImageNet pre-trained model.
ii) A detection network is trained within fast R-CNN using the proposals generated in the trained RPN, which is also initialized by the ImageNet-pre-trained model.
iii) The initialization of RPN training is made through the detection network by fixing the shared convolutional layers and only fine-tuning the layers unique to RPN.
iv) The shared convolutional layers are kept fixed, and the layers unique to the fast R-CNN are fine-tuned.

Thus, a unified network is formed because the same convolutional layers are shared by both of the RPN and detection networks.

III. Results

To demonstrate the performance of our method, we applied it to different test images. Part of the test images (5 images, 45 SGSs) came from our collected data (5 images, 48 SGSs), extracted from public sources as cited in each image. Fig. 5 shows two examples, for the species *A. ellipticus* (a) and *A. lacticofeatus*. We can see that the conidia are recognized and matched with the right species with high confidence scores (see Table II). In Fig. 6, the species *A. carbonarius*, *A. brasiliensis* and *A. aculeatus* were successfully identified according to their species images.

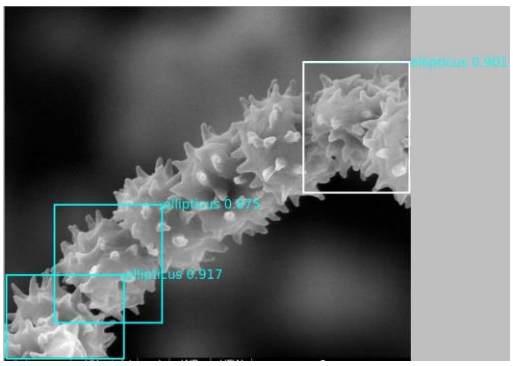


Figure 5. Resulting images for the species of *A. ellipticus* [3].

Due to the high resolution of Fig. 5 and Fig. 6, our method was able to recognize the relevant SGSs with high confidence scores.

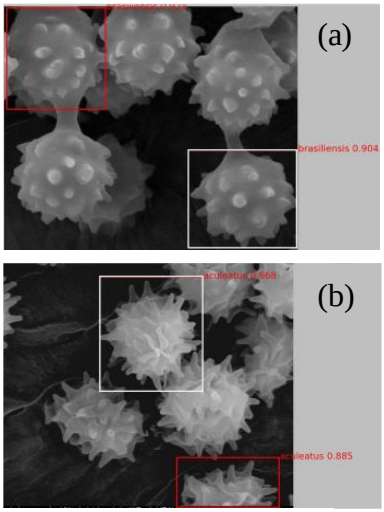


Figure 6. Resulting images for the species *A. brasiliensis* (a) [21] and *A.* aculeatus (b) [14].

Table II presents the obtained scores for the highlighted SGSs in these figures. We can see that our method recognized the species with the highest score, such as 0.901392 for *A. ellipticus* in Fig. 5, and 0.904234 for *A. brasiliensis* in Fig. 6(a).

TABLE II. The Scores Obtained in Our Recognition System for the SGSs Highlighted in Fig. 5 and Fig. 6. The Scores in Bold Are Correspondent to the Species Detected

	SGS with *A. ellipticus*	SGS with *A. lacticofeatus*	SGS with *A. carbonarius*	SGS with *A. brasiliensis*	SGS with *A. aculeatus*
A. ellipticus (Fig. 5)	**0.901392**	0.001268	0.001675	0.002184	0.004081
A. brasiliensis (Fig. 6(a))	0.007655	0.002187	0.001307	**0.904234**	0.001365
A. aculeatus (Fig. 6(b))	0.001344	0.000689	0.00096582	0.000825	**0.868145**

For the images with low resolution and poor contrast, as shown in Fig. 7, our method proved to be very promising since it still allowed us to obtain the correct species identification. Although, this may lead to a low confidence score such as 0.108 obtained for *A. carbonarius* in Fig. 7(b).

Our method can well use the conidial features and be trained to recognize the species in a new image. However, due to similar geometric properties, as seen for *A. brasiliensis* and *A. carbonarius*, our method may make incorrect judgement when applied to an image with low resolution and poor contrast. This is actually due to less geometric structure information provided. So, the performance of our method can be improved by significantly increasing the number of the ground truth images. Table III presents the confusion matrix with the confidence score 0.5 for the test dataset. The recognition accuracies for *A. brasiliensis* and *A. carbonarius* were 76.47% 86.21%, respectively. *A. carbonarius* was detected at 90.91%. Note that the score confidence was 0.5. To improve the identification performance, more

ground truth data needs to be provided so that the conidial geometric structure information of all the species can be enhanced.

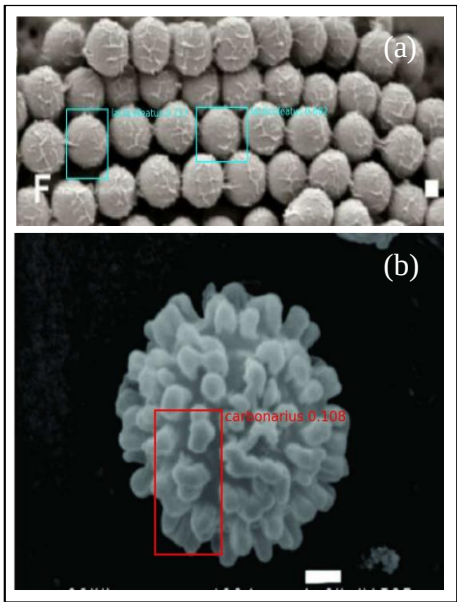


Figure 7. Resulting images obtained images from low resolution source images: *A.* lacticofeatus (a) [21] *A. carbonarius* (b) [22].

IV. Discussion and Conclusions

We developed a CNN based method to recognize aspergilli species. We managed to circumvent the issue of the small number of source images by employing a geometric transformation to augment the training data. A sophisticated training dataset was generated to train a CNN model in terms of Faster R-CNN. The initial results demonstrate that our method can effectively detect the species qualitatively and quantitatively.

Frankly speaking, although the developed method is able to effectively identify fungal species with a small number of SEM images, identification can be extensively enhanced with a higher amount of data. As shown in Table III, the detection accuracy of the model could be improved with some extra efforts.

TABLE III. The Confusion Matrix for the Confidence Score 0.5

Actual / Predicted	*A. ellipticus*	*A. lacticofeatus*	*A. carbonarius*	*A. brasiliensis*	*A. aculeatus*
A. ellipticus	86.96%	8.70%	0	0	4.34%
A. lacticofeatus	6.06%	90.91%	0	0	3.03%
A. carbonarius	0	0	86.21%	13.79%	0
A. brasiliensis	5.88%	0	17.65%	76.47%	0
A. aculeatus	0	6.67%	6.67%	0	86.67%

We currently adapted the geometrical transformations to augment the data. With the advanced techniques in synthetic data generation like Deep Convolutional Generative Adversarial Networks (DCGAN) [23], the more synthetic data could be generated to enrich the training data profile that should help improve the performance of the method. Another point is worth noting that the current method aims to detect the fungal species class. However, the biologists are keen to know the details of the species structure. Thus, the instance based object recognition will be promising to do the investigation such as Mask R-CNN [24].

For the purpose of future evaluation of our method, regarding fungal imagery, we will collect more images to generate more species conidial signatures. This will enhance the training set and improve the identification performance. In addition, we will extend our method to all 27 species of *Aspergillus* from section *Nigri*, by collecting SEM images of conidia for all the species within this section.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Huaizhong Zhang and Marta Filipa Simões conducted the research; Huaizhong Zhang developed and analyzed the data; Huaizhong Zhang drafted the paper; Huaizhong Zhang and MFS wrote, reviewed and edited the manuscript. All authors have approved the final version.

References

[1] M. L. Abarca, F. Accensi, J. Cano, and F. J. Cabanes, "Taxonomy and significance of black aspergilli," *Antonie van Leeuwenhoek*, vol. 86, pp. 33-49, 2004.

[2] E. D'Hooge, P. Becker, D. Stubbe, A. C. Normand, R. Piarroux, and M. Hendrickx, "Black aspergilli: A remaining challenge in fungal taxonomy?" *Medical Mycology*, vol. 57, no. 6, pp. 773-780, 2019.

[3] M. F. Simões, C. Santos, and N. Lima, "Structural diversity of *Aspergillus* (Section *Nigri*) spores," *Microscopy and Microanalysis*, vol. 19, no. 5, pp. 1151-1158, 2013.

[4] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 6, no. 6, pp. 84-90, 2017.

[5] E. Bathoorn, N. E. Salazar, S. Sepehrkhouy, M. Meijer, H. D. Cock, and P. J. Haas, "Involvement of the opportunistic pathogen *Aspergillus tubingensis* in osteomyelitis of the maxillary bone: A case report," *BMC Infectious Diseases*, vol. 13, no. 1, p. 59, 2013.

[6] C. Paulussen, J. E. Hallsworth, S. Álvarez-Pérez, W. C. Nierman, P. G. Hamill, D. Blain, *et al.*, "Ecology of aspergillosis: Insights into the pathogenic potency of *Aspergillus fumigatus* and some other *Aspergillus* species," *Microbial Biotechnology*, vol. 10, no. 2, pp. 296-322, 2017.

[7] C. Qiu and D. N. Su, "Pulmonary aspergillosis resembling to pulmonary tuberculosis," in *Pulmonary Aspergillosis*, C. Qiu, P. X. Lu, and S. P. Wu, Eds., Singapore: Springer, 2019, pp. 27-79.

[8] D. N. Su, C. Qiu, P. X. Lu, and S. P. Wu, "Epidemiology of pulmonary aspergillosis," in *Pulmonary Aspergillosis*, Singapore: Springer, 2019, p. 7.

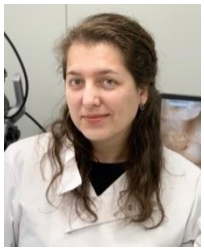
[9] H. Badali, H. Fakhim, F. Zarei, M. Nabili, A. Vaezi, N. Poorzad, *et al.*, "In vitro activities of five antifungal drugs against opportunistic agents of *Aspergillus Nigri* complex," *Mycopathologia*, vol. 181, no. 3-4, pp. 235-240, 2016.

[10] M. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in Computer Vision – ECCV 2014, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds., vol. 8689, Switzerland: Springer, Cham, 2014, pp. 818-833.

[11] K. Simonyan and A. Zisserman. (2014). Very deep convolutional networks for large-scale image recognition. [Online]. Available: https://arxiv.org/abs/1409.1556

[12] J. R. R. Uijlings, K. E. A. Sande, T. Gevers, and A. W. M. Smeulders, "Selective search for object recognition," *International Journal of Computer Vision*, vol. 104, no. 2, pp. 154-171, 2013.
[13] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 91-99, 2015.
[14] M. F. Simões, "Quality parameters in a culture collection-Micoteca da Universidade do Minho," Ph.D. dissertation, Centre of Biological Engineering, Minho University, Braga, Portugal, 2013.
[15] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, *et al.*, "Caffe: Convolutional architecture for fast feature embedding," in *Proc. 22nd ACM International Conference on Multimedia*, 2014, pp. 675-678.
[16] Tzutalin. (2015). Git code. [Online]. Available: https://github.com/tzutalin/labelImg
[17] M. Everingham, S. A. Eslami, L. V. Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes challenge: A retrospective," *International Journal of Computer Vision*, vol. 111, no. 1, pp. 98-136, 2015.
[18] R. Girshick, "Fast R-CNN," in *Proc. IEEE International Conference on Computer Vision (ICCV)*, Santiago, 2015, pp. 1440-1448.
[19] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 3431-3440.
[20] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-based convolutional networks for accurate object detection and segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 1, pp. 142-158, 2016.
[21] M. F. Simões, C. Santos, and N. Lima, "The importance of structural diversity of spores in the taxonomy of *Aspergillus* (section *Nigri*)," *Microscopy and Microanalysis*, vol. 19, suppl. S4, pp. 77-78, 2013.
[22] R. A. Samson, J. A. M. P. Houbraken, A. F. Kuijpers, J. M. Frank, and J. C. Frisvad, "New ochratoxin A or sclerotium producing species in *Asperqillus* section *Nigri*," *Stud. Mycol.*, vol. 50, pp. 45-61, 2004.
[23] A. Radford, L. Metz, and S. Chintala. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. [Online]. Available: https://arxiv.org/abs/1511.06434
[24] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 42, no. 2, pp. 386-397, 2018.

Huaizhong Zhang is a Senior Lecturer in the Department of Computer Science, Edge Hill University, Ormskirk, UK.
His research interests lie mostly in the areas of deep learning and big data, image analysis, object tracking and recognition. In his past research experiences, he has carried out his studies with a variety of image/video modalities like CT, MRI, OCT, Thermal video, Hyperspectral remote sensing, etc. A broad of the cutting-edge techniques have been involved into the relevant investigation. Among these, the main methods involved include CNN based techniques, Bayesian methods, level sets, MCMC, PDE, statistical learning, and graph theory amongst others. In addition, he has conducted various software developments, e.g., Proteomics/open source pipelines, video surveillance system, and computer assisted instruction.



Marta Filipa Simões currently works in astrobiology at the State Key Laboratory of Lunar and Planetary Science (SKLPlanets), Macau University of Science and Technology (MUST), China.
She has previously done research in UK, Saudi Arabia, and Portugal. She has a PhD in Chemical and Biological Engineering, a MSc in Clinical Microbiology, and a degree in Biotechnological Engineering. She is a microbiologist with a special interest in astrobiology (astromycology), applied microbiology, and mycology.
Her current research focuses are fungal ecology and biodiversity in environmental analogues to outer-space conditions, bioprospection and application of filamentous fungi, and fungal growth containment and exploitation to our needs in outer-space similar conditions.