Texture Classification for Object Detection in Aerial Navigation using Transfer Learning and Wavelet-based Features

J.M. Fortuna-Cervantes*[1], M.T. Ramírez-Torres[2], M. Mejía-Carlos[1], J. Martínez-Carranza[3,4] and J.S. Murguía-Ibarra[1]
[1]Universidad Autónoma de San Luis Potosí, Facultad de Ciencias-IICO, San Luis Potosí, México
[2]Universidad Autónoma de San Luis Potosí, Coordinación Académica Región Altiplano Oeste, San Luis Potosí, México
[3]Instituto Nacional de Astrofísica, Óptica, y Electrónica, Puebla, México
[4]University of Bristol, Bristol, UK

ABSTRACT

The use of Micro Aerial Vehicles (MAVs) has increased in engineering and civil applications to explore environments without previous information. In particular, in autonomous navigation, a fundamental part is that of detecting and locating targets of our interest. For this reason, computer vision has become an essential analysis tool. In this work, we focus on object classification in aerial navigation tasks, where texture is involved as a physical property of the object. We present a classification model using transfer learning and wavelet-based features as an additional feature extraction method. This model is trained with the Describable Textures Database (DTD), and a performance of 53% accuracy is obtained. Moreover, the images obtained from the environment show the generalization of learning for some database classes. Transfer learning fusion with wavelet analysis is recommended for small data sets of images with textures due to the limitation of learning about spectral information lost in conventional Convolutional Neural Networks (CNNs).

1 INTRODUCTION

Texture analysis is a traditional problem in computer vision because it involves obtaining information that describes the image content. In the robotics area, object detection is a problem for robots that perform tasks in real scenarios and in real-time, given the lighting conditions, indeterminate orientations, object identity, shape, color and texture. Furthermore, the information may differ in outdoor and indoor environments, which varies the target information [1]. Providing the resources to the robot by integrating sensors can improve object detection.

Micro Aerial Vehicles (MAVs) have been used in different environments due to their easy control and implementation of algorithms through computer vision. In tasks of the classification, detection and localization of the target. There are different vision methods in the area, such as optical flow, segmentation, edge detector, morphological operations and feature extractor for different tasks. These methods have been combined to improve detection performance while the MAV executes its aerial navigation (recognition) or autonomous flight. However, using these methods can be computationally expensive to perform real-time detection, affecting the overall system performance.

In image processing, texture can be defined from neighboring pixels and intensity distribution over the image [2]. Besides, there are some classification methods for texture analysis such as statistical, geometric, model and spectral. If we focus on spectral methods, these methods describe the texture in the frequency domain. They are based on the decomposition of a signal in terms of basis functions. Furthermore, they use the expansion coefficients as elements of the feature vector.

Deep learning has become a helpful tool for image classification, object detection, and segmentation. Especially if we talk about convolutional neural networks, these achieve learning multiple features to recognize targets without reference to their position, indeterminate orientations, scale, and target rotation. VGG16, VGG19, AlexNet, SSD and YOLO are architectures that have good performance in image classification and object detection tasks.

Therefore, we decide to merge these methodologies (deep learning and wavelet features) as a solution for texture classification. The objective is that the MAV performs the aerial navigation (inside the virtual environment) for the classification system to recognize the object, see figure 1. This work focuses on preview information (in data collected by MAV) and structural recognition of the object (with a particular texture) within a region of interest in the image plane.

The implementation of our system is developed with the fusion of two approaches. The first is in the spatial domain, using transfer learning. We take as a baseline the VGG16 architecture with the features of the ImageNet database. The second approach focuses on the spectral domain, applying the

*Email address(es): juan.manuel.fortuna@hotmail.com

Haar wavelet transform in two dimensions to obtain features at different scales [3]. The VGG16 network has been selected for its fast performance and implementation with transfer learning and adaptability with wavelet analysis. Internally, the system is divided into two stages: the first corresponds to feature extraction and the second one to the classification stage. We used the Describable Texture Database (DTD) to train our model, which contains 47 texture classes, with 120 images per class. On average, we have tested with some textures for the classification task in the virtual environment, and the prediction can be performed correctly, with an average processing speed of 2 fps.

The rest of the paper is organized as follows in Section 2 related work is shown, Section 3 introduces the methodology to approach the texture classification problem. Whereas Section 4 shows the results with the DTD dataset and the experimental part to test our model. Finally, Section 5 presents the conclusions.

Figure 1: We designed a system for texture classification in aerial navigation based on knowledge inference over the DTD database. See at `https://youtu.be/d41kgBw7Y_c`.

2 RELATED WORK

In recent years, MAVs applications for object detection tasks have been studied and developed [6][7]. Several approaches are using deep learning, giving excellent performance in applications. For example, in some tasks for autonomous navigation, we find in the literature a methodology for obstacle detection and avoidance using an architecture called AlexNet that allows classifying the images captured by the camera onboard the drone [8]. The learning of this architecture is transferred from the ImageNet database to improve the classification performance [9]. Moreover, to detect objects and autonomous landing, in [1], a detection system is presented to solve one of the missions included in the IMAV2019 indoor competition. They involve the implementation of the SSD7 onboard the MAV. This SSD7 network is chosen for its fast performance on low-budget microcomputers with no GPU. The method proposed in [10] is an architecture called YOLO, which presents an essential performance in real-time image detection and processing at 45 frames per second. Besides, in object detection tasks, in [11], the authors propose a deep learning approach to estimate the object's center in a robust way. Also, generating a line of sight as a guide proves to be a solution to avoid collision with other objects, due to complications such as varying illumination conditions, object geometry, and overlapping in the image plane.

On the other hand, many projects employ deep learning and wavelet analysis in visual processing. For example, on image classification, the method proposed in [12] converts images from the CIFAR-10 and KDEF database to the wavelet domain, thus obtaining temporal and frequency features. The different representations are added to multiple CNN architectures. This combination of information in the wavelet domain achieves higher detection efficiency and faster execution times than the spatial domain procedure. In this sense, the authors in [13] mention that although CNN is a universal extractor, in practice, it is not clear whether CNN can learn to perform spectral analysis. In [14], the authors propose an architecture called CNN Texture to have this approach within the CNN. Their idea focuses on the fact that the information extracted by convolutional layers is of minor importance in texture analysis. They use a statistical energy metric in the feature extraction stage. This information is concatenated with the classification stage, the fully connected layer. Specifically, the architecture shows an improvement in performance and a reduction in computational cost.

In terms of texture classification in image processing applications [15], the authors propose an architecture called wavelet CNN to generalize spectral information lost in conventional CNNs. This information is beneficial for texture classification as it usually contains details information about the object's shape. Furthermore, the model allows us to have fewer parameters than in traditional CNNs, so it is possible to train with less memory. In general, through a state-of-the-art review, we have observed that computational intelligence algorithms improve detection strategies in Micro Aerial Vehicle applications.

3 METHODOLOGY

This work proposes an approach based on transfer learning and wavelet features. This system allows to predict or classify the texture in images transmitted by the camera onboard the drone, whose objects are in an outdoor scenario (in Gazebo), a virtual simulation environment. We are only interested in texture recognition, mainly to know one of the characteristics of the object. So, we limit the image plane (640×360) to a region of interest (300×300 pixels). As a result, the system will have the image in RGB as well as a grayscale version. These two images are the inputs for our proposed classification system, see figure 2.

Describable Textures Dataset (DTD) was selected to be used. It contains 47 classes of 120 images in the wild. This means that the images have been acquired that in uncontrolled

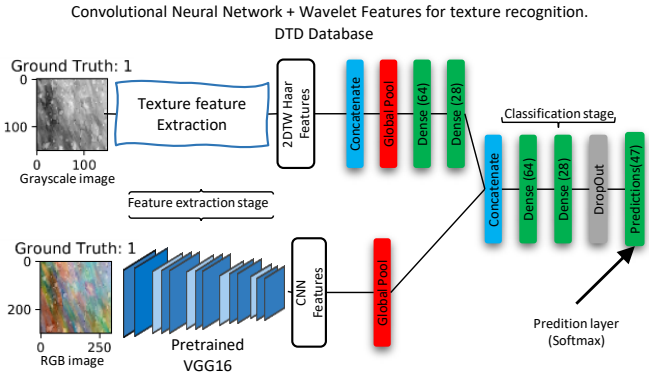


Figure 2: Texture classification system.

conditions [16]. This dataset includes ten divisions available with 40 training images, 40 validation images, and 40 test images for each class. Our experiment will create a new dataset, with the distribution of 70% for training, 15% for validation, and the remaining 15% for testing. Figure 3 shows some images from this set. One limitation of the dataset is the number of images per class, so it is decided to use the transfer learning method to improve the classification performance of our model. The synaptic weights are based on the ImageNet database, which will feed the feature extraction stage of the base architecture VGG16.

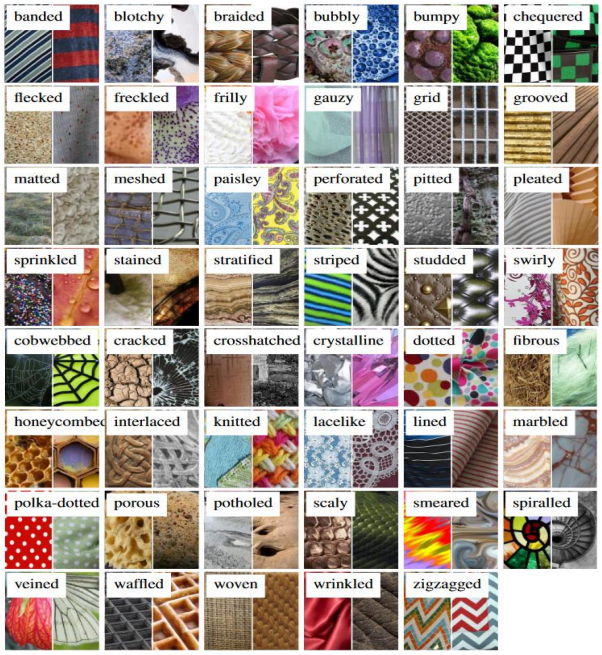


Figure 3: DTD dataset example images [16].

Before training, the Haar wavelet transforms in two dimensions is applied to one level, see figure 2. The factor of one represents the level of image decomposition. This new spectral information is essential for classification. Therefore, four sets (in the wavelet domain) are automatically generated to determine the characteristic attributes of each texture. This information can be combined with the spatial information of the VGG16 architecture. Also, it is essential to mention that this process is only applied to the image previously converted to grayscale, performing the decomposition for a single channel, ver figure 4 & 5.

For the test stage in MAV, a scenario is designed in the gazebo simulator. The virtual scenario is created with ten cubes with certain textures (figure 1). These textures are selected due to the performance achieved in the model testing stage. Therefore, the chosen classes have a performance above 70% accuracy (Table 3).

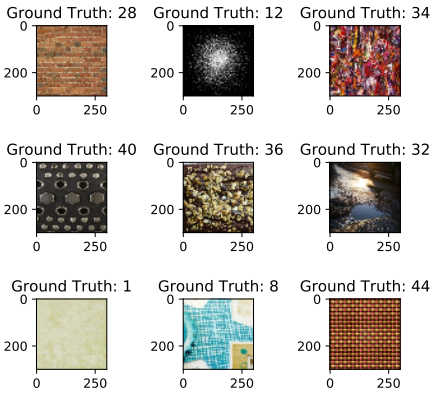


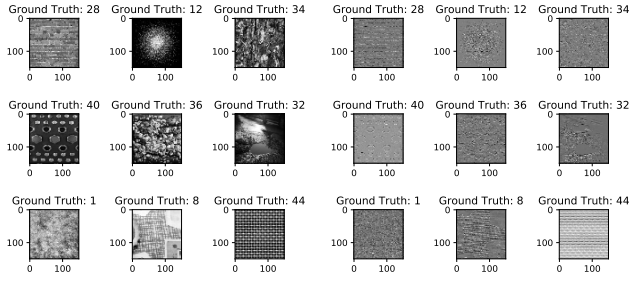
Figure 4: Images textures that have been decoded (Class) to train the classification model.

4 EXPERIMENTS AND RESULTS

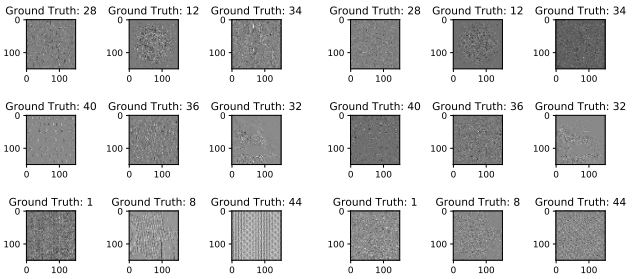
The experiment to train our learning model was carried out with the Keras API with Tensorflow as Backend [17]. Besides, the OpenCV libraries are used for image processing due to their ease of use and adaptability in programming. Also, we use the Pywt library [18] from which it was chosen the Haar wavelet transform as the feature extractor method. An aerial navigation experiment was performed using the ROS framework and Gazebo simulation environment to validate the classification system and its learning generalization. This section describes the results obtained in each experiment.

4.1 Model training

In the first instance, the VGG16 network was trained from scratch. As a result, it is not able to generalize its learning. Therefore, it is possible to use the transfer learning methodology. Table 1 shows the achieved performance of the pre-trained network and our proposal with the wavelet feature fusion. It shows the accuracy performance on the three sets to validate the model (training, validation, and test). In the case of the pre-trained network, slight overfitting is observed. The model will be adjusted to learn specific cases and will be



(a) Approximation. (b) Horizontal details.



(c) Vertical details. (d) Diagonal details.

Figure 5: Approximation and details set of wavelet features.

unable to recognize new textures. One way to improve the performance of the model is to integrate the wavelet features. In this case, we achieve the elimination of overfitting and homogeneity between the three sets. Besides, the value of the test set is highlighted because these are images that the model has never seen. In summary, the classification system has 14,778,735 synaptic learning weights. 64,047 are trainable parameters, of which 16,832 correspond to wavelet features. Table 2 summarizes the achieved performance of our classification system, as well as a comparison to AlexNet (trained from scratch), T-CNN, and Wavelet CNN [14][15].

	Training	Validation	Test
Pre-trained model	68.15	50.41	54.49
Our model	57.67	51.22	**53.19**

Table 1: Classification results for the pre-trained VGG16 network and our model indicated as accuracy (%).

	AlexNet	T-CNN	Wavelet CNN	Our model
DTD	22.7	55.8	59.8	**53.19**

Table 2: Classification results and comparison with other state-of-the-art pre-trained architectures with ImageNet, in terms of accuracy (%).

4.2 Texture classification DTD

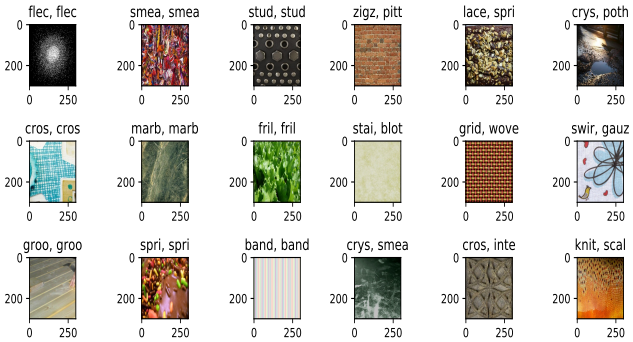
Other metrics evaluate the performance of the DTD dataset classes. The metrics such as precision, recall, and f1-score are given when applying the classification_report method, where it is necessary to involve the true labels and the prediction label of the model. Table 3 shows the classes that performed above 70% classification. Also, Table 4 shows three random classes that perform above 50% classification. This class selection analysis provides the basis for the design of the textured cubes of the Gazebo environment. On the other hand, we can observe the similarity and correlation between classes (about test set) by performing the prediction. Figure 6 shows the true label and the prediction label at the top of each texture.

Class	precision	recall	f1-score	support
bubb	0.73	0.61	0.67	18
cheq	1.00	0.78	0.88	18
fibr	0.73	0.61	0.67	18
fril	0.72	0.72	0.72	18
stri	0.77	0.94	0.85	18
stud	0.70	0.78	0.74	18
zigz	0.75	0.67	0.71	18

Table 3: Classes (test set) that results with precision above 70%.

Class	precision	recall	f1-score	support
hone	0.58	0.61	0.59	18
line	0.50	0.28	0.36	18
polk	0.65	0.61	0.63	18

Table 4: Classes (test set) that results with precision above 50%. They are chosen from the easy human visual perception of the texture.



(a) 450 Correctly classified. (b) 396 Incorrectly classified.

Figure 6: Classification of textures randomly (from a total of 846 images) using the DTD prediction model.

4.3 Texture classification in aerial navigation

Navigation and aerial recognition tested the classification model. We created a virtual environment with the Gazebo simulator, controlling and sending information from the camera onboard the drone using the ROS framework. In the world presented in figure 1, we positioned in a row the ten cubes with the selected textures. Therefore, the position of the cubes allows the evaluation of the prediction model during aerial exploration. The idea of the model is that it generalizes its learning to textured objects. In total, 1000 image captures were performed in a navigation recognition for each class. The proposed texture sets (bubbly, chequered, honey, striped, studded) obtain a high correlation with their original label above 60% accuracy, see figure 7.

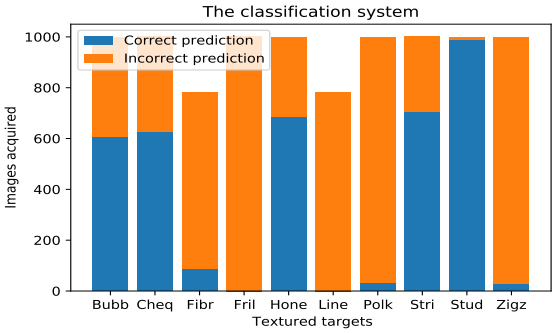


Figure 7: The number of images with textures obtained with the onboard camera while flying recognition.

Some images (figures 8 and 9) of the recognition set are shown, with its original label and its prediction label. However, we can observe that the five test images incorrectly predict the frilly, lined, polka-dotted, and zigzagged set. These five images relate to the whole recognition set, except for fibrous, polka-dotted, and zigzagged, which achieve at least 3% accuracy. This result allows us to understand the generalization of learning between the model and textured objects.

5 CONCLUSION

The localization and object detection tasks using visual information are challenging, particularly when objects exhibit repetitive texture. However, these tasks open the opportunity for various applications using Micro Aerial Vehicles equipped with onboard cameras to be used for object detection and recognition, for instance, for parcel pick-up, place recognition, landing zone detection and many more. Seeking to improve the detection and recognition stage, in this work we have investigated the use of spectral analysis in combination with deep neural networks. In particular, in this proposal, we merged the (additionally created) spectral feature maps to CNN learning. Also, it is shown that the model used achieves to eliminate overfitting and better accuracy in the classification of textures with a significant increase in the number of

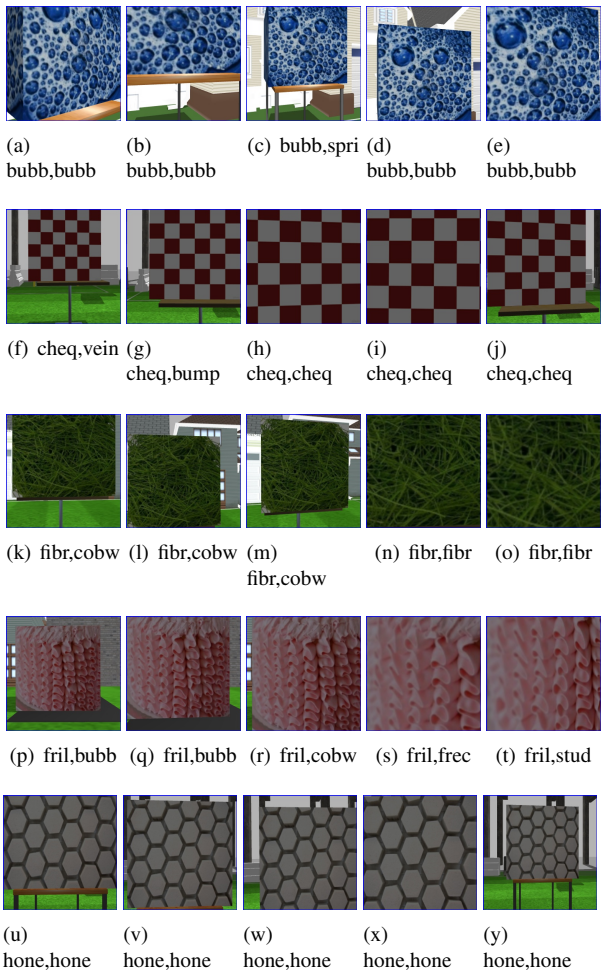


Figure 8: Image sequence acquired by the camera onboard the drone. The classification system has a good inference on the texture in the first, second, and fifth rows.

parameters to be trained. The tests performed in the simulation show some interesting results. The prediction model shows the creation of a widespread understanding of the texture attached to the objects. Furthermore, despite having a low classification rate, it is shown that the model correctly classifies most of the test classes.

As future work, this will test with other texture features, also seeking to conduct tests in real-world scenarios.

ACKNOWLEDGEMENTS

J. M. Fortuna-Cervantes is a doctoral fellow of CONACYT (México) in the program of "Ciencias Aplicadas" at UASLP-IICO.

REFERENCES

[1] Aldrich A Cabrera-Ponce and José Martínez-Carranza. Onboard cnn-based processing for target detection and autonomous landing for mavs. In *Mexican Conference on Pattern Recognition*, pages 195–208. Springer, 2020.

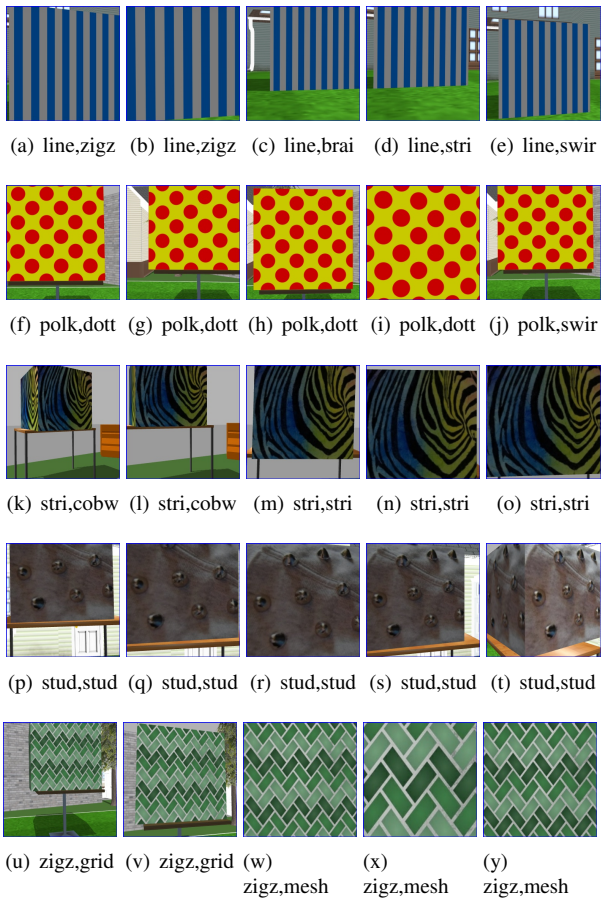


Figure 9: Image sequence acquired by the camera onboard the drone. In the second, third, and fourth rows, the classification system gets good classification performance.

[2] Natalia S Vassilieva. Content-based image retrieval methods. *Programming and Computer Software*, 35(3):158–180, 2009.

[3] LC Yan, B Yoshua, and H Geoffrey. Deep learning. *nature*, 521(7553):436–444, 2015.

[4] Yoshua Bengio, Ian Goodfellow, and Aaron Courville. *Deep learning*, volume 1. MIT press Massachusetts, USA:, 2017.

[5] Stephane G Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE transactions on pattern analysis and machine intelligence*, 11(7):674–693, 1989.

[6] Yakoub Bazi and Farid Melgani. Convolutional svm networks for object detection in uav imagery. *Ieee transactions on geoscience and remote sensing*, 56(6):3107–3118, 2018.

[7] Yalong Pi, Nipun D Nath, and Amir H Behzadan. Convolutional neural networks for object detection in aerial imagery for disaster response and recovery. *Advanced Engineering Informatics*, 43:101009, 2020.

[8] Sinahi Dionisio-Ortega, L Oyuki Rojas-Perez, Jose Martinez-Carranza, and Israel Cruz-Vega. A deep learning approach towards autonomous flight in forest environments. In *2018 International Conference on Electronics, Communications and Computers (CONIELECOMP)*, pages 139–144. IEEE, 2018.

[9] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25:1097–1105, 2012.

[10] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.

[11] Sunggoo Jung, Sunyou Hwang, Heemin Shin, and David Hyunchul Shim. Perception, guidance, and navigation for indoor autonomous drone racing using deep learning. *IEEE Robotics and Automation Letters*, 3(3):2539–2544, 2018.

[12] Travis Williams, Robert Li, et al. An ensemble of convolutional neural networks using wavelets for image classification. *Journal of Software Engineering and Applications*, 11(02):69, 2018.

[13] Travis Williams, Robert Li, et al. Wavelet pooling for convolutional neural networks. *Proc. Int. Conf. on Learning Representations*, 2018.

[14] Vincent Andrearczyk and Paul F Whelan. Using filter banks in convolutional neural networks for texture classification. *Pattern Recognition Letters*, 84:63–69, 2016.

[15] Shin Fujieda, Kohei Takayama, and Toshiya Hachisuka. Wavelet convolutional neural networks. *arXiv preprint arXiv:1805.08620*, 2018.

[16] Mircea Cimpoi, Subhransu Maji, Iasonas Kokkinos, Sammy Mohamed, and Andrea Vedaldi. Describing textures in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3606–3613, 2014.

[17] Francois Chollet et al. *Deep learning with Python*, volume 361. Manning New York, 2018.

[18] G. Lee, K. Wohlfahrt R. Gommers, F. Waselewski, and A. O'Leary. Pywavelets: A python package for wavelet analysis. In *Journal of Open Source Software*, volume 4, page 1237, 2019.