

PROCEEDINGS OF SPIE

SPIDigitalLibrary.org/conference-proceedings-of-spie

Multiple drone classification using millimeter-wave CW radar micro-Doppler data

Rahman, Samiur, Robertson, Duncan

Samiur Rahman, Duncan A. Robertson, "Multiple drone classification using millimeter-wave CW radar micro-Doppler data," Proc. SPIE 11408, Radar Sensor Technology XXIV, 1140809 (23 April 2020); doi: 10.1117/12.2558435

SPIE.

Event: SPIE Defense + Commercial Sensing, 2020, Online Only

Multiple drone type classification using millimetre-wave CW radar micro-Doppler data

Samiur Rahman^a, Duncan A. Robertson^a

^a University of St Andrews, SUPA School of Physics & Astronomy, St Andrews, Fife KY16 9SS, Scotland

ABSTRACT

This paper investigates the prospect of classifying different types of rotary wing drones using radar. The proposed method is based on the hypothesis that the rotor blades of different sizes and shapes will exhibit distinct Doppler features. When sampled unambiguously, these features can be properly extracted and then can be used for classification. We investigate various continuous wave (CW) spectrogram features of different drones obtained with a low phase noise, coherent radar operating at 94 GHz. Two quadcopters of different sizes (DJI Phantom Standard 3 and Joyance JT5L-404) and a hexacopter (DJI S900) have been used during the experimental trial for data collection. For classification training, we first show the limitation of the feature extraction based method. We then propose a convolutional neural network (CNN) based approach in which the classification training is done by using micro-Doppler spectrogram images. We have created an extensive dataset of spectrogram images for classification training, which have been fed to the existing GoogLeNet model. The trained model then has been tested with unseen and unlabelled data for performance verification. Validation accuracy of above 99% is achieved along with very accurate testing results, demonstrating the potential of using neural networks for multiple drone classification.

Keywords: Micro-Doppler, Radar, CW, millimeter-wave, classification, drones, deep learning, neural network, GoogLeNet

1. INTRODUCTION

Micro-Doppler is an intrinsic feature of the radar signature of commercially available rotary wing drones. The propeller blade rotations produce distinctive micro-Doppler signatures in their radar signal returns^{1,2}. A surveillance radar system can use this characteristic to identify the drones. A plethora of research work have been reported in recent years regarding the use of this feature to classify a drone, as well as for discriminating it from a confuser (i.e. bird)^{3,4,5,6}. Beyond classifying the target as a drone, the next logical question is to identify the drone type. In principle, discrimination between different models of drones is possible since different targets will produce different micro-Doppler signatures unique to their micro-motion properties. In practice, this is a complicated challenge when using a radar sensor in a dynamic target scenario. Commercial drones are targets with low radar cross section (RCS) compared to other airborne targets like planes or helicopters. Also, it is more difficult to do micro-Doppler based discrimination between two different types of drones than to distinguish between a drone and a bird.

Few preliminary works have been reported to address this problem using deep learning methods. In⁷, the concept of using a convolutional neural network (CNN) for classifying types of drones at different altitudes was demonstrated. In that work, spectrograms and corresponding cadence velocity diagrams (CVD) obtained from Ku-band radar were combined to form merged Doppler images as the dataset used to train the CNN. A widely used CNN structure, GoogLeNet⁸, was used for training, showing more than 90% accuracy. In⁹, a modified version of CNN, named fully-convolutional network, was used to classify between three different models of drones using simulated micro-Doppler data. Another CNN based multiple drone classification method was reported in¹⁰, again using simulated data, showing 93% classification accuracy.

In this paper, we have attempted to classify two quadcopters of different sizes and one hexacopter, using experimentally obtained micro-Doppler spectrogram images as the dataset. CNNs typically used to classify optical images so using radar spectrograms, which are not optical images and use false color scales, requires some caution. Our conjecture is that a high fidelity spectrogram detailing the distinct shapes of the propeller signatures will significantly improve the image based neural network training accuracy. Hence, we have used millimeter-wave radar data, obtained at 94 GHz, for better micro-Doppler sensitivity¹¹. A millimeter-wave radar has the advantage of higher Doppler sensitivity by producing spectrograms with better resolution in a shorter time. A CW spectrogram plot from a millimetre-wave radar without Doppler ambiguity

produces high quality radar signatures of the drone propeller blades. We have created a dataset of 1950 CW spectrogram images in total of three types of drone for training, where all the images are obtained from experimental flight trials. The classification training outcomes are discussed in this paper along with testing accuracy results. Also, a brief comparison between the CNN based classification method and feature extraction based method is made, showing the advantage of using CNN. It should be noted that all the radar data used here are obtained in staring mode for the simplicity of the experimental setup.

2. CNN TRAINING WITH MICRO-DOPPLER SPECTROGRAM IMAGES

CNN is a supervised learning method based on multilayer perceptron (MLP)¹². The training is done by using filters of various sizes to learn different features of an image. The low-level features correspond to basic shapes (lines, curves etc.), whereas the high-level features are more specific to the classifying objects (drone propeller rotation, bird wing flap etc.). Low-level features are prevalent in every image so we have selected a well developed and tested network, GoogLeNet, in this study. Using a pre-trained network has the advantage of providing a robust foundation for feature learning.

2.1 GoogLeNet

GoogLeNet is a very well developed network that pioneered the use of different sizes of convolutional layer filters in parallel, known as the inception module⁸, which considerably improves the training performance. This is because instead of randomly selecting the filter sizes for optimization, all the different combinations are used in GoogLeNet. This increases the computational complexity, hence GoogLeNet also uses 1x1 convolution for dimension reduction.

GoogLeNet consists of 144 layers in total in which every single element in an inception module is regarded as a separate layer. In this study, we have frozen the first 10 layers to use it as a pre-trained network. These layers correspond to the low-level generic features, hence we want those to remain the same. The final two layers are changed according to the number of classes for our purpose. The layer named 'loss3-classifier' is changed to a fully connected layer with three classes. The last layer ('output') is changed according to the class labels.

2.2 Training dataset

The training dataset has been created by separately acquiring radar data of three different drones, shown in Fig. 1. These are used as the three classes for the CNN training and their relevant details are as follows-

- *DJI Phantom Standard 3* (quadcopter, weight 1.216 kg, width 35 cm, blade length 13 cm)
- *DJI S900* (hexacopter, weight 3.3 kg, width 90 cm, blade length 16 cm)
- *Joyance JT5L-404* (quadcopter, weight 7 kg, width 85 cm, blade length 27 cm)

The different profiles and sizes of the propeller blades and different structures of the motor bodies contribute to different micro-Doppler features in the spectrograms which can be observed in Fig. 2.

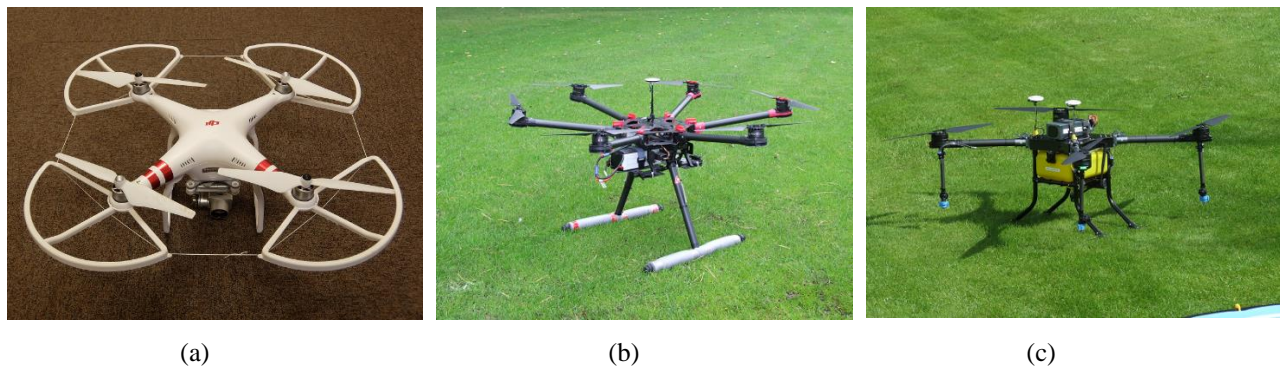


Figure 1. Drones used for classification training, a) DJI Phantom Standard 3, b) DJI S900, c) Joyance JT5L-404

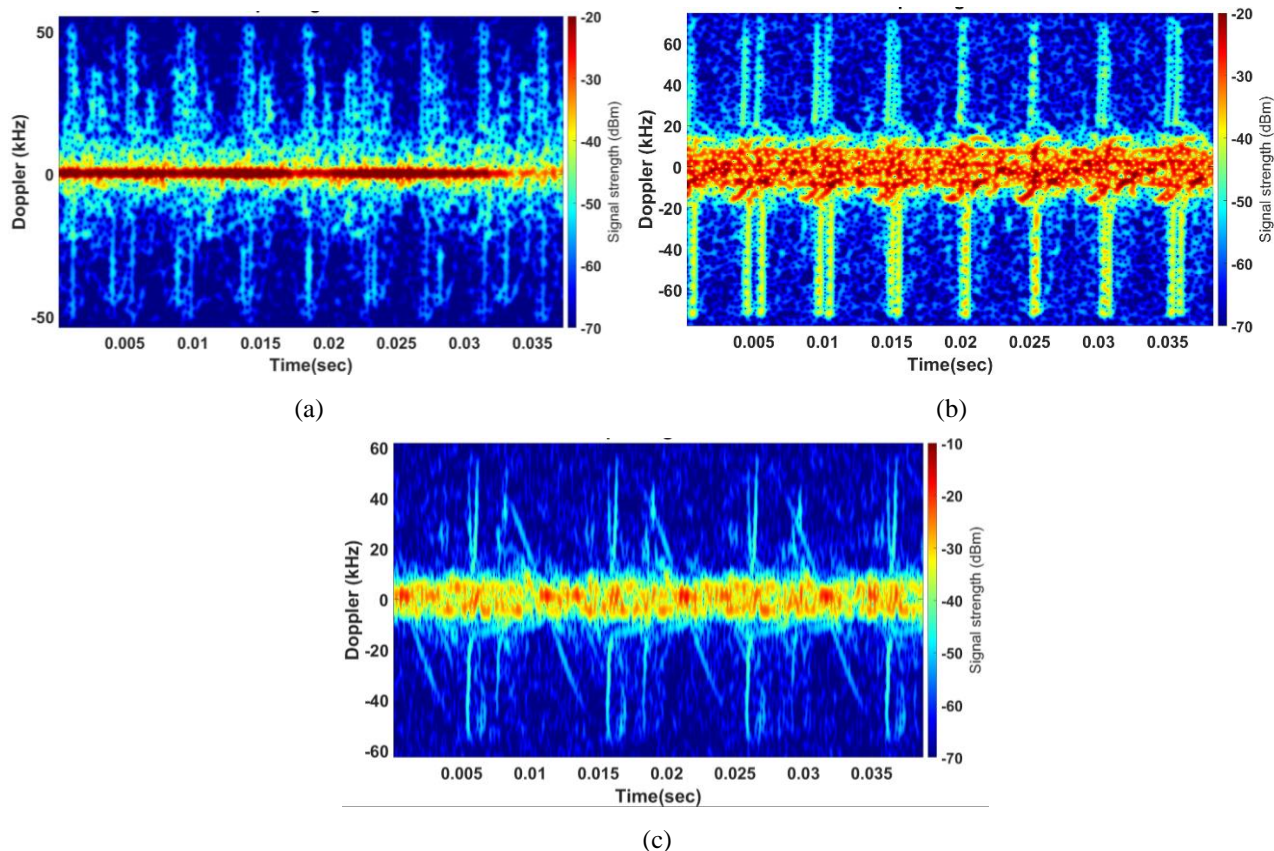


Figure 2. Example CW spectrograms of the three drones at ~ 30-50 m range, a) Phantom, b) S900, c) Joyance

Fig. 2 also reveals that the Doppler frequency values are quite similar for all three drones. The maximum Doppler of the Phantom is slightly lower than the other two, but not significant enough to use this as a classification feature. The same is true for the periodicity information which illustrates the fact that even though the Doppler features are clearly visible from the spectrogram and are very useful to discriminate a drone from other targets, they are not useful for distinguishing among drone types. On the other hand, as the blade flash patterns are visibly quite distinct, an image-based classification technique will potentially be a better solution. As it is very hard to quantify these more subtle differences in shapes, a neural network will be able to learn them and ultimately will have the ability to assign them to a particular class, eliminating the requirement for feature value calculation. This also shows that unambiguous Doppler measurements are required for the neural network to work. As there is a hardware constraint to operate an FMCW radar with sufficiently high Doppler sampling rate we have chosen to collect fully sampled CW Doppler data. Fig. 2 also justifies the use of highly Doppler sensitive millimeter-wave radar as the quality of the micro-Doppler signatures are far higher fidelity than those obtained by lower frequency systems as reported elsewhere. In the context of designing a radar to detect, track and classify drones, it may be necessary to combine FMCW data, for ranging and tracking, with CW data, for classification.

One of the main criteria for a robust neural network model is to have a diverse dataset. Simply having a large number of images for each class will not result in a high-fidelity classification model. For instance, a large dataset of a specific drone can be created from a single radar measurement where the drone hovers in a fixed spot for a certain period. The classification training may generate very high accuracy in this situation but the model will severely underperform in a dynamic scenario, which is to be expected in a practical situation. To generate diversity in the dataset, data has been taken at different ranges (15-60 m) and with different aspect angles ($\sim 5-12.5^\circ$), mostly with the drones hovering but also with them flying normally.. Various values regarding the parameters used for the radar data collection and the subsequent signal processing are provided in Table 1.

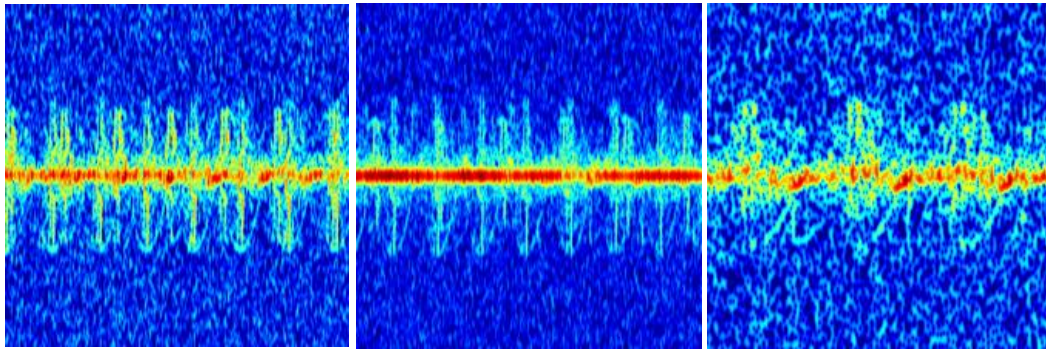
Table 1. Radar system and signal processing parameters

Operating frequency	94 GHz
Operating mode	CW, staring
Antenna beamwidth	Fan beam, 0.9° azimuth, 3° elevation
Polarization	Circular polarization (CP)
Sampling rate	100 kHz, 200 kHz
Maximum unambiguous velocity	80 ms ⁻¹ , 160 ms ⁻¹
Spectrogram image length	40.1 ms
Short Time Fourier Transform (STFT) length	512 samples
STFT window	Gaussian, width factor 16
STFT overlap	95%

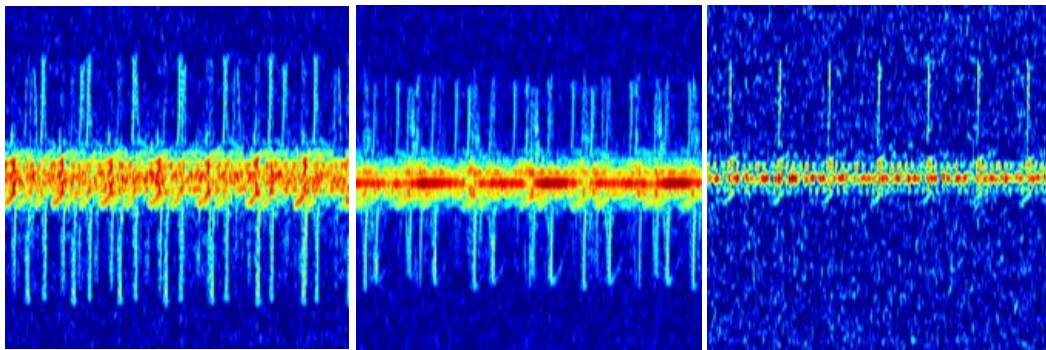
The radar system used here is a very low phase noise 94 GHz radar, named T-220¹³. During experimental trials, the drones were not always entirely within the beam, which produced spectrograms with lower signal strength. We have intentionally included quite a lot of such images in the dataset to ensure more generalized learning of the network, hence reducing the chance for overfitting of the model. We had multiple trials spanning several days and to maximize diversity, we have selected data evenly from multiple datasets instead of choosing heavily from a single measurement to yield a dataset of around 20,000 images. All the spectrograms have been produced in MATLAB[®] using the ‘jet 256’ colour scale, after being pre-processed for Doppler alignment. Eventually, we have selected 650 spectrogram images for each class, 1950 images in total, for training. Fig. 3 shows some example images from our dataset. The variation in the images can be easily observed which have occurred due to the differences in range, aspect angle and target position relative to the antenna beam. It can be seen that along with images with strong signal return, images where the micro-Doppler signals are barely visible are also chosen to prevent the model from memorizing.

3. TRAINING RESULTS

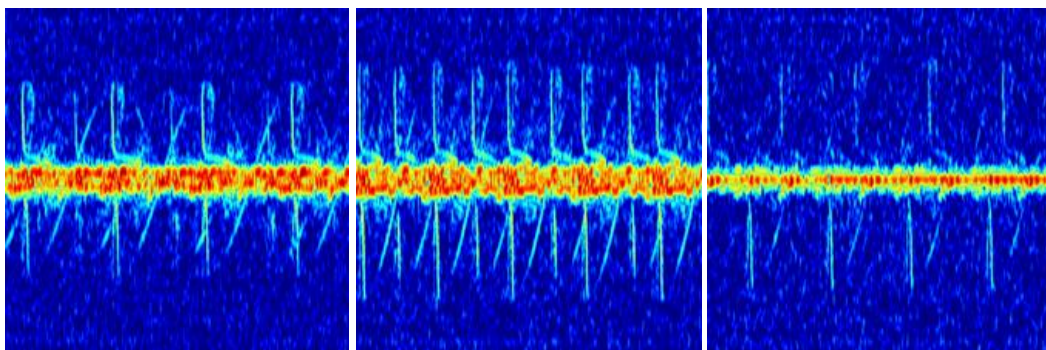
The training has been performed in MATLAB (version 2019b) by using a core i7-9700 CPU at 3.00 GHz and with 16 GB RAM. The GoogLeNet training is done by selecting 80% data for training and 20% data for validation, with a batch size of 50 images. We have also trained the datasets with 70-30% and 90-10% splits but the 80-20% split produced the best result in terms of validation accuracy. The validation data is used during training after every epoch where one epoch corresponds to all the iterations of all the batches, both forward and backward, once. This helps to adjust the training parameters as any large differences between the training and validation accuracy indicate overfitting. Overfitting is one of the major issues in neural network based classification models where the model has not learned the features comprehensively; instead it has remembered lots of features which are noise. Hence, special care has been taken to minimize this problem as much as possible. Usually for a very large dataset (e.g. GoogLeNet training with ~ 1 million images), a bigger portion of validation data is useful to test and resolve overfitting. For a small dataset, most of the data needs to be used for training, otherwise the model will not learn properly. Our dataset does not fall into either of these categories, it is a medium sized database so an 80-20% split is a good trade-off. The classification code has been instructed to randomly make this split, again to ensure maximum variation. One of the main parameters that affects overfitting is the learning rate. Very large values of learning rate often tends to underfitting, where the model has not learned the basic features properly. Meanwhile, a very small learning rate leads to overfitting. We have trained the model with five different learning rates (0.01, 0.005, 0.002, 0.001, 0.0001) to find the optimum accuracy. We found that training with a learning rate of 0.001 gives the best results. The lower values were quite underfitted whereas the learning rate of 0.0001 showed both the training and validation accuracy values to be 100%, indicating heavy overfitting. Fig. 4 shows the results of the training, where the achieved validation accuracy is 99.74%. A total number of four epochs have been used during training done by carefully monitoring when the training accuracy plot reaches a plateau. Once that the plateau is reached, running more epochs will make the network prone to more overfitting.



(a)



(b)



(c)

Figure 3. Example images from the training dataset, a) Phantom, b) S900, c) Joyance

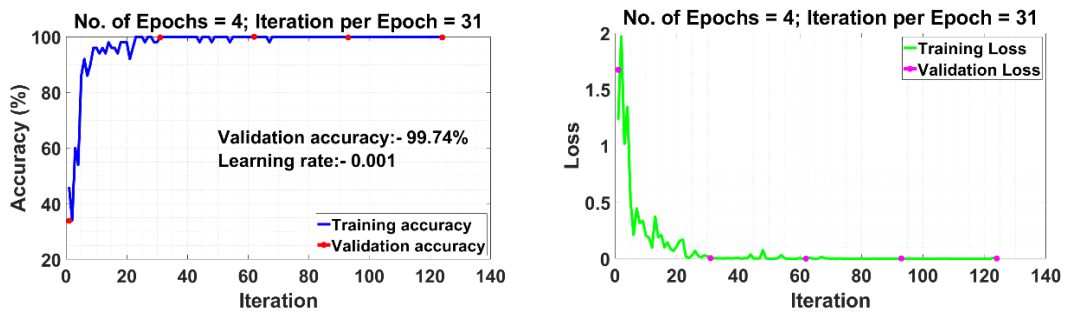


Figure 4. GoogLeNet training and validation accuracy plot (left) and corresponding loss function plot (right) for the 3-class training

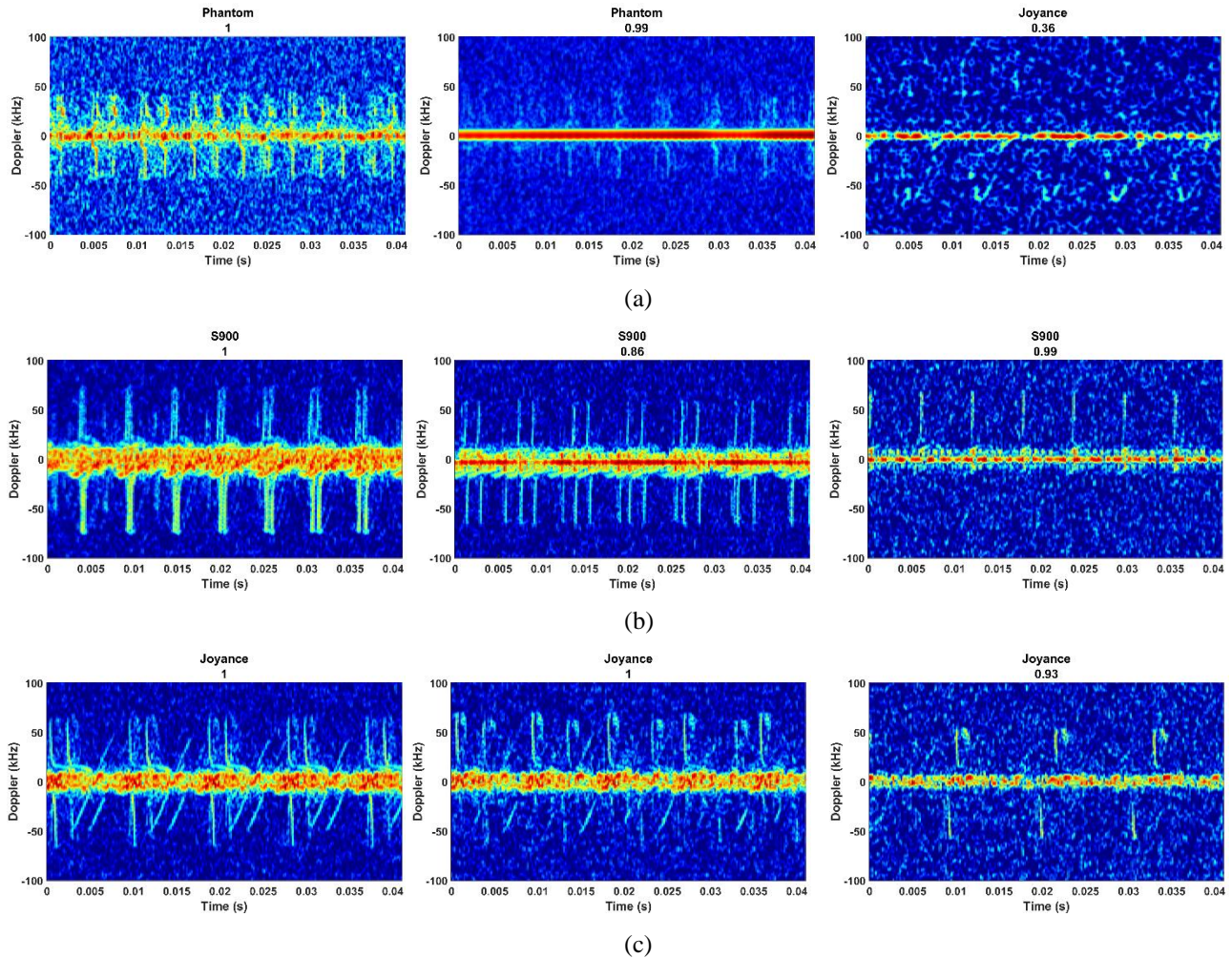


Figure 5. Example results of the drone type classification using unseen test data where the label on top of each image is the target classified by the network and the number below is the confidence level with maximum value of 1, a) Phantom spectrogram images, b) S900 spectrogram images, c) Joyance spectrogram images

We have selected 75 spectrogram images for each class, hence 225 in total, as the test dataset. These images are not part of the training dataset and so are entirely unseen by the classification model. We have then applied the model to classify all these images to verify the performance. As shown in Fig. 5, the process classifies the image, labels it according to the maximum score among the three classes, then provides that score under the classified label. The results are very encouraging. The model has accurately predicted all the 75 spectrogram images of the S900 and all 75 images of the Joyance. Examples are shown in Fig. 5 - note there are few images where there is very little micro-Doppler signature yet those are still correctly classified for both the S900 and Joyance. Among the 75 Phantom images, 9 have been classified incorrectly as either S900 or Joyance. In all those 9 cases, the images were of very low quality in terms of Doppler signatures, as can be seen in the last image of Fig. 5 (a). Here, the model has wrongly classified the Phantom as Joyance with 36% confidence. This means that when the model is used on an image which is essentially clutter, most of the time it gets classified as S900 or Joyance, not Phantom, showing that the training data was still lacking sufficient diversity in terms of low-fidelity Phantom images. Also, it should be accepted that any purely micro-Doppler based classification algorithm is always going to struggle when the radar has acquired negligible Doppler information due to the target not being properly within the beam.

We have also compared the CNN based training performance with a feature extraction based algorithm which we had proposed in our earlier work¹⁴. That method showed very promising results in accurately identifying a drone from a

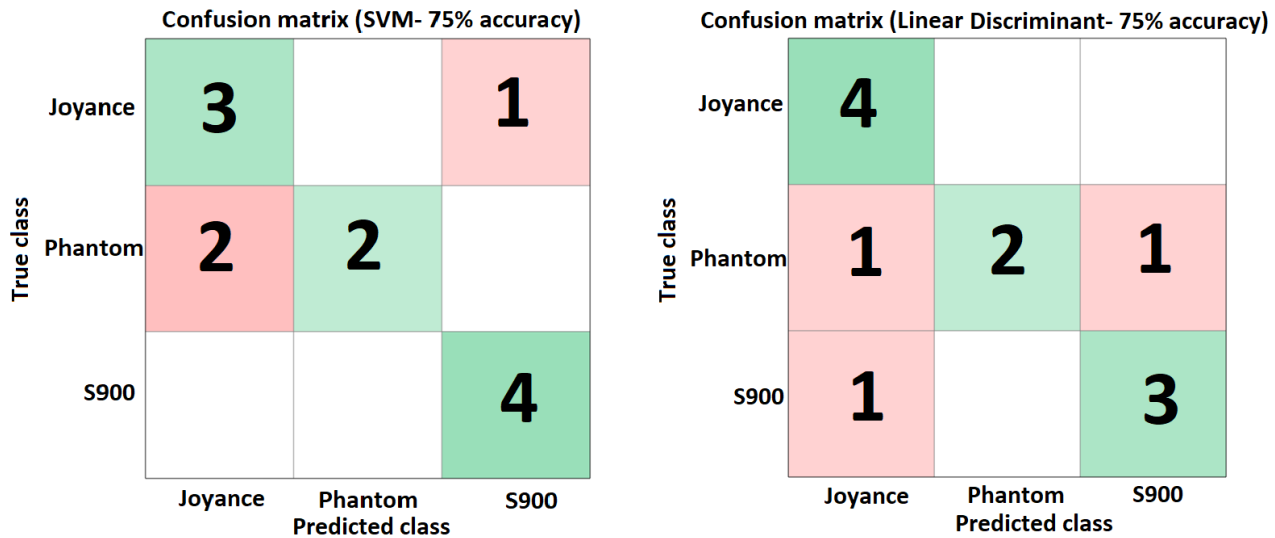


Figure 6. Confusion matrices after training by using the feature extraction method¹⁴, for SVM (left) and linear discriminant (right), showing poorer performance compared to the CNN model

confuser like a bird but was not able to discriminate between different drone types at all. The previous study was done using FMCW data which did not fully sample the propeller micro-Doppler, so that weakness was expected.

Here, we have tested the method using our CW datasets. Different Doppler feature values were obtained from the CW spectrogram plots of the three drones. We have then applied two commonly used classifiers, linear discriminant and support vector machine (SVM), for training. In both cases, the classification accuracy was 75%, much worse performance than the CNN based training – Fig. 6. Also, it is worth mentioning that the CNN based classification is going to be much faster in a real-time operation. This is because the step for calculating various feature values is eliminated in the case of CNN. In both cases the time required for spectrogram generation is going to be the same, but after that CNN only converts the spectrogram into an RGB image and then the trained model classifies the target. The RGB image conversion takes ~ 0.05 seconds and then GoogLeNet takes about 0.4 seconds to classify the image. Conversely, feature extraction can take about a second before the values are fed to the model for final classification. If one uses a more computationally heavy feature extraction method like singular value decomposition (SVD) or calculates more feature values, the classification time can increase further.

4. CONCLUSION AND FUTURE WORKS

The main objective of this study was to prove the concept of using CNN to classify multiple drone types in a realistic scenario. Three different drones have been used in this study as target classes. Millimeter-wave CW radar data has been used to generate micro-Doppler spectrogram images for the corresponding drones. This has enabled the production of high quality images for training. A large dataset was then created which was used to train GoogLeNet. Over 99% validation accuracy has been achieved. The validation of the method is then demonstrated by feeding images unknown to the model, showing 100% classification accuracy in the case of S900 and Joyance drones and 88% accuracy in the case of a Phantom drone. These are very promising results illustrating that a large and diverse experimental radar dataset can not only be used for classifying drones from other targets, but for classifying different models of drones as well. The advantage of the proposed method over a feature extraction based method is also discussed. It should be emphasized that there are still a lot of opportunities to upgrade the proposed method. The inaccurate classification of the low-quality Phantom images definitely suggests the necessity of further improvement. The more diverse the dataset is, the more robust it will be for general purpose use. In this study, only a single radar system has been used with a fixed polarisation for data collection. More variation in the dataset can be achieved by utilizing radar systems with different operating frequencies, polarisation, antenna beamwidth, sampling rate etc. Also, even though the dataset used here is obtained from various ranges and aspect angles, that variation can be made larger. Data taken at different places will also create changes in the surrounding clutter. Variation in signal processing can also be attained, but careful consideration should be taken while doing so. It should be kept in mind that, fundamentally, radar spectrograms are not optical images, so changing parameters like noise floor

threshold, STFT length and color scale will alter the spectrograms significantly and may affect the CNN based classification process. When developing a CNN based model, careful decisions should be made regarding the extent of the versatility of the model being used in practice.

5. ACKNOWLEDGEMENTS

The authors acknowledge the funding received by the Army Research Laboratory under Cooperative Agreement Number: W911NF-19-2-0075. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government.

REFERENCES

- [1] Chen, V. C., [The micro-doppler effect in radar], Artech House (2011).
- [2] Chen, V. C., Li, F., Ho, S.-S. and Wechsler, H., "Analysis of micro-Doppler signatures," *IEE Proc. - Radar, Sonar Navig.* **150**(4), 271 (2003).
- [3] Molchanov, P., Harmanny, R. I. A., de Wit, J. J. M., Egiazarian, K. and Astola, J., "Classification of small UAVs and birds by micro-Doppler signatures," *Int. J. Microw. Wirel. Technol.* **6**(3-4), 435-444 (2014).
- [4] Tahmouh, D., "Detection of small UAV helicopters using micro-Doppler," *Proc. SPIE 9077, Radar Sens. Technol. XVIII 9077*, K. I. Ranney and A. Doerry, Eds., 907717, International Society for Optics and Photonics (2014).
- [5] Ritchie, M., Fioranelli, F., Borrion, H. and Griffiths, H., "Classification of loaded/unloaded micro-drones using multistatic radar," *Electron. Lett.* **51**(22), 1813-1815 (2015).
- [6] Hoffmann, F., Ritchie, M., Fioranelli, F., Charlish, A. and Griffiths, H., "Micro-Doppler based detection and tracking of UAVs with multistatic radar," 2016 IEEE Radar Conf. RadarConf 2016, 1-6, IEEE (2016).
- [7] Kim, B. K., Kang, H.-S. and Park, S.-O., "Drone Classification Using Convolutional Neural Networks With Merged Doppler Images," *IEEE Geosci. Remote Sens. Lett.* **14**(1), 38-42 (2017).
- [8] Szegedy, C., Wei Liu, Yangqing Jia, Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. and Rabinovich, A., "Going deeper with convolutions," 2015 IEEE Conf. Comput. Vis. Pattern Recognit., 1-9, IEEE (2015).
- [9] Brooks, D. A., Schwander, O., Barbaresco, F., Schneider, J.-Y. and Cord, M., "Temporal Deep Learning for Drone Micro-Doppler Classification," 2018 19th Int. Radar Symp., 1-10, IEEE (2018).
- [10] Choi, Byunggil; Oh, D., "Classification of Drone Type Using Deep Convolutional Neural Networks Based on Micro-Doppler Simulation," ISAP 2018 2018 Int. Symp. Antennas Propag. Oct. 23-26, 2018, Parad. Hotel Busan, Busan, Korea (2018).
- [11] Rahman, S. and Robertson, D. A., "Millimeter-wave micro-Doppler measurements of small UAVs," *Proc. SPIE 10188, Radar Sens. Technol. XXI 10188*, K. I. Ranney and A. Doerry, Eds., 101880T, International Society for Optics and Photonics (2017).
- [12] Haykin, S. S., [Neural networks : a comprehensive foundation], Prentice Hall (1999).
- [13] Robertson, D. A., Brooker, G. M. and Beasley, P. D. L., "Very low-phase noise, coherent 94GHz radar for micro-Doppler and vibrometry studies," *Proc. SPIE 9077, Radar Sens. Technol. XVIII 9077*, K. I. Ranney and A. Doerry, Eds., 907719, International Society for Optics and Photonics (2014).
- [14] Rahman, S. and Robertson, D. A., "Millimeter-wave radar micro-Doppler feature extraction of consumer drones and birds for target discrimination," *Radar Sens. Technol. XXIII*, K. I. Ranney and A. Doerry, Eds., 28, SPIE (2019).