Micro-Doppler Signal Representation for Drone Classification by Deep Learning

Julien Gérard, Joanna Tomasik, Christèle Morisseau, Arpad Rimmel, Gilles Vieillard

Micro-Doppler Signal Representation for Drone Classification by Deep Learning

Julien Gérard *, Joanna Tomasik *, Christèle Morisseau †, Arpad Rimmel * and Gilles Vieillard †
*LRI, CentraleSupélec, Université Paris-Saclay, F-91405, Orsay, France
Email: jgerard@lri.fr, joanna.tomasik@lri.fr, arpad.rimmel@lri.fr
†DEMR, ONERA, F-91123, Palaiseau, France
Email: christele.morisseau@onera.fr, gilles.vieillard@onera.fr

***Abstract*—There are numerous formats which represent the micro-Doppler signature. Our goal is to determine which one is the most adapted to classify small UAV (*Unmanned Aerial Vehicules*) with Deep Learning. To achieve this goal, we compare drone classification results with the different micro-Doppler signatures for a given neural network. This comparison has been performed on data obtained during a radar measurement campaign. We evaluate the classification performance in function of different use conditions we identified with a given neural network. According to the experiments conducted, the recommended format is a spectrum issued from long observations as its classification results are better for most criteria.**

I. Introduction

The increase of small drones in our everyday life raises security threats. To prevent any potential misuse, investments are made to detect and identify them. As explained in [1], methods based on vision or audibility do not permit reliable observations at distances longer than several hundred meters.

Another way to address this issue is the radar measurement. However, the low RCS (*Radar Cross Section*) of small drones makes them hard be differentiated from each other. Thus, the mainstream of research focuses on specific features to highlight micro-Doppler effects.

The main Doppler effect consists in a frequency shift due to the target radial speed. The micro-Doppler effect [2], [3] consists on Doppler modulation created by internal movement of the target such as drone rotors.

The classification is made by feature extraction from the radar signal. The micro-Doppler signal potential for classification with Bayes methods [4], [5] or SVM [5]–[7] has been discussed recently. The latest progress in machine learning, especially in CNN (*Convolutional Neural Network*), has revealed a new line to study this phenomenon as the neural network extracts features itself. Studies emphasize the advantages of neural networks in the case of micro-Doppler profiles for drone classification [8]–[10] or human movements [7], [11]–[13].

The micro-Doppler effect may be represented with different formats. The most common is a set of contiguous micro-Doppler profiles called spectrograms obtained thanks to STFT (*Short Time Fourier Transform*). However, other formats such as CP (*Cepstrum*), CVD (*Cadence Velocity Diagram*) [6], [14], [15] or even directly the time signal $x(t)$ are also investigated.

The spectrograms are mostly used as an input format for the CNN in the aforementioned studies. Certain studies tend to show that the combined use of two formats improves the CNN performance [9] and helps the network to extract specific features, others show differences in performances for some formats [10], despite the fact that all information is already present in the spectrogram.

The variety of formats used in these studies compromises the comparison of their results. Moreover, they are made under different conditions: simulated or real data [6], [8], sets of drones, various trajectories, classification goals, radars, etc.

Our main contribution is to compare the efficiency of different format inputs (described in Section II) for drone classification with a CNN using the criteria we propose to capture use conditions defined in Section III. Section IV, gives the experimental setup[1]. Eventually, we analyze the results obtained and indicate the format for which the classification preforms the best (Section V) and conclude (Section VI).

II. Formats — Time frequency analysis methods

We have selected the five most relevant transformations for the micro-Doppler analysis according to the literature: time signal after range compression $x(t)$, weighted spectrum of the signal *WSP*, cepstrum *CP*, spectrogram *SG* and cadence velocity diagram *CVD*. They are defined below with continuous equations. An example of these formats is shown in Fig 1.

All these formats but one are based upon FFT (*Fast Fourier Transform*). We did not observe any improvement with the use of zero-padding and thus did not apply it in the remainder. For all time windows h defined below, we chose Kaiser windows. We tested also rectangular windows and did not observe any significant difference.

Theoretical models of micro-Doppler effect [2], [3], [16], [17] give us expectations for each format depending on whether the drone variations (rotor speeds, orientation, etc.) are slow compared to the observation duration. We denote targets respecting this hypothesis as relatively stationary targets.

A. Time signal, $x(t)$

We denote as time signal $x(t)$, the complex signal corresponding to the target position, obtained with range compression. It is discretized at $\frac{1}{\text{PRF}}$ (*Pulse Repetition Frequency*). The other formats are produced by processing $x(t)$.

[1]We thank Jean-Paul Marcelin and Jean-François Petex for their work on the data collecting campaign with the ONERA measurement system Medycis.

B. Weighted Spectrum, WSP

The weighted spectrum, *WSP*, of a signal $x(t)$ is its FFT on a time window h: $\text{WSP}(x(t))(f) = \text{FFT}\left(h(t)x(t)\right)(f)$.

The WSP format offers a good frequency resolution ($\frac{\text{PRF}}{N}$ for an N point signal) but it has no time resolution. For relatively stationary drones, peaks corresponding to multiples of twice the rotor speed (two blades per rotor [16]) appeared.

To remove potential biases, the main Doppler effect due to target speed is set to zero. To equilibrate WSP, all frequency shifts lower/higher than $\pm\,4$ kHz are removed.

C. Cepstrum, CP

The cepstrum, *CP*, of a signal $x(t)$ is defined by the following equation (IFFT signifies *Inverse Fast Fourier Transform*):

$$\text{CP}(x(t))(\tau) = \text{IFFT}\left(\log\left(|\text{WSP}(x(t))(f)|^2\right)\right)(\tau).$$

Introduced in [18], CP is similar to WSP. The quefrencies τ, equivalent to frequencies f, highlight the signal echoes, especially when they are closed one to another $\left(\text{the quefrency resolution is } \frac{1}{\text{PRF}}\right)$. As for WSP, for relatively stationary drones, peaks due to the rotors may be observed.

This format, originally designed to highlight echoes for human visualization, is based on the modulus information of WSP. In human visualization, when we compare CP with WSP, we also use only its modulus (Fig. 1). Recent studies [19], using only the modulus for complex signals in the Deep Learning context, have shown that the loss of information due to the modulus may deteriorate the neural network accuracy.

The CP format transforms the convolution in the time domain into addition in the quefrency domain. Bogert *et al* [18] conclude that CP resists better to noise than WSP.

D. Spectrogram, SG

The spectrogram, *SG*, is obtained from $x(t)$ with STFT:

$$\text{SG}\left(x(t)\right)(\tau, f) = \int_t h(t-\tau)x(t)e^{-2i\pi ft}dt.$$

This format is based upon a trade-off between the time and frequency resolutions. The short time window h allows one to concatenate profiles to obtain a time resolution contrary to WSP. However, the frequency resolution is reduced. For an h window on M points, the time resolution is $\frac{M}{\text{PRF}}$ and the frequency resolution is $\frac{\text{PRF}}{M}$. We chose an h window of 2.5 ms.

As for WSP, the main Doppler effect is set to zero and all frequencies higher/lower than $\pm\,4$ kHz are suppressed.

For relatively stationary drones, we can expect to see blade flashes [8]. This effect can be put in light by window overlaps. For this reason, the overlap may improve the classifier quality despite the fact that it does not provide us with more information, increases the data size, and makes learning longer. After several experiments, we used an overlap of 60%, the lowest value giving the best classification results.

E. Cadence Velocity Diagram, CVD

The Cadence Velocity Diagram, *CVD*, is produced by performing FFT for all SG frequencies:

$$\begin{aligned}\text{CVD}\left(x(t)\right)(\nu, f) &= \text{WSP}\left(\text{SG}\left(x(t)\right)(\tau, f)\right)(\nu)\\ &= \int_\tau h(\tau)\text{SG}\left(x(t)\right)(\tau, f)e^{-2i\pi\nu\tau}d\tau.\end{aligned}$$

This format highlights the frequency periodicity better than SG on which it is based. This format is particularly adapted for substantially long observations.

For CVD on N points, produced from SG made with an h window of M points, the cadence resolution is $\frac{M\text{PRF}}{N}$ while its frequency resolution is $\frac{\text{PRF}}{M}$, identical to the SG one. This format is thus adapted to long observation durations (large N) as it requires a great number of profiles to execute the second FFT.

III. Sensitivity to use conditions

The performance of the five formats presented above are compared in Section V under different use conditions. Now, we describe the use conditions chosen. The first one (Subsection III-A) is considered as a reference case. The following ones are robustness to noise (Subsection III-B) and short observation duration (Subsection III-C). The last use condition (Subsection III-D) is the facility of training of the CNN.

We aim at observing the impact of the format chosen upon the quality of the CNN classification. We are interested not only in the classification rate under every condition but also in the gap between the reference and the current use condition (indicated by the Δ column in the tables of Section V).

A. Reference

The reference case consists in classifying drones into five classes. The reference observation is 300 ms. It corresponds to the minimum duration needed to differentiate two rotors within 100 RPM (*Revolution Per Minute*) for CP and WSP under the hypothesis of the relative stationarity of targets.

The reference SNR (*Signal Noise Ratio*) is between 30 and 50 dB, depending on the drone and the fragment of its trajectory (altitude, speed, orientation of the drone, etc.).

B. Robustness to noise

The reference case provides us with a signal with a good SNR for each target. Now, we assess which format allows a satisfactory classification regardless of the SNR deterioration.

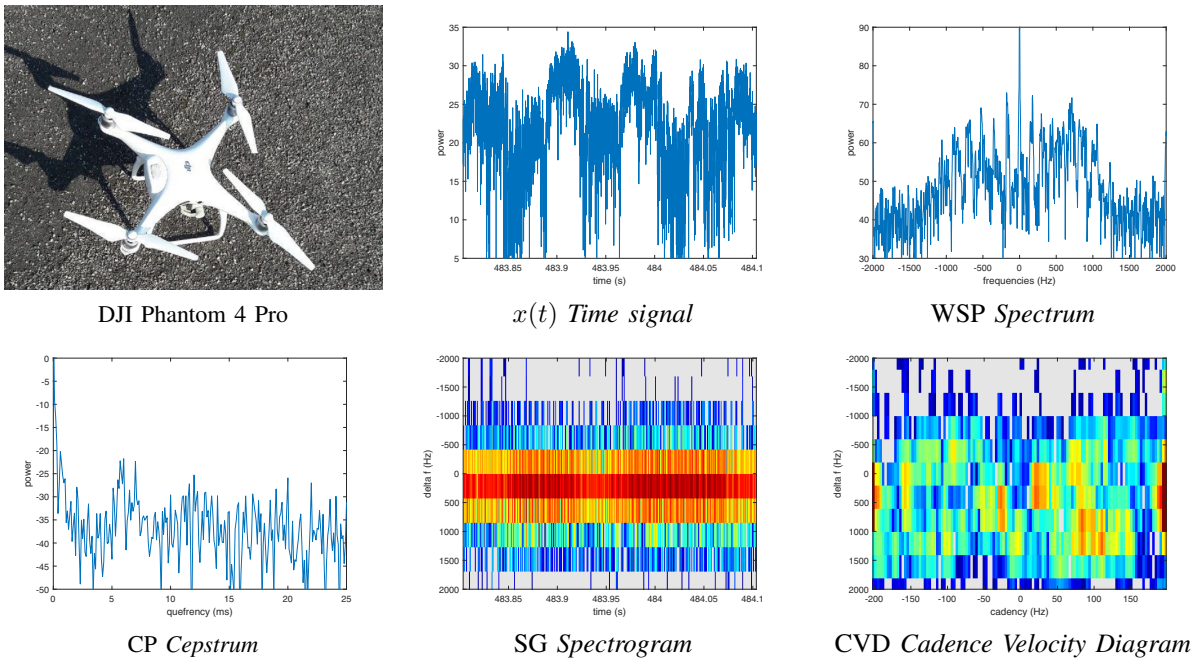
To achieve this goal, we added a white Gaussian noise to the reference data. The noise level is absolute and thus independent of the drone type. Consequently, it has a different impact on each signal section according to its SNR. The resulting database SNR is between 10 and 30 dB.

C. Short observation

Some formats are more adapted to shorter/longer observations. For this use condition, the observation time is shortened in comparison to the reference case in order to assess observation duration robustness. The shorter observation time chosen is 36 ms corresponding to three turns for a rotor at 5 000 RPM.

D. Facility of training

Certain formats change data dimension or shape. These modifications impact substantially the time needed to execute each iteration in the CNN. They may also improve the contrast between data and thus reduce the number of iterations needed to reach the maximum accuracy value.



DJI Phantom 4 Pro | $x(t)$ *Time signal* | WSP *Spectrum*

CP *Cepstrum* | SG *Spectrogram* | CVD *Cadence Velocity Diagram*

Fig. 1: Signature of a DJI Phantom 4 Pro observed during 300 ms with the different formats. Data are displayed in dB, zooms are added for better visualization. Data depicted correspond to a random part of one trajectory (from 483.8 to 484.1 seconds)

To evaluate formats on this criterion, the time required for one retro-propagation is computed. We also determine the time needed to reach 95% of the maximum accuracy.

IV. Experimental setup

A. Measurement campaign

The measurements were performed during two weeks in April 2019. Five drones (DJI S1000-A2, grand Spyder, Gryphon, DJI Phantom 4 Pro, DJI Mavic Pro) were involved. The measurements were taken in the same conditions for all drones. Drones flew on trajectories differing by speed, altitude, pattern (up/down, rectilinear/circular movement, stationary/ rotation around its center, free fly) to have the largest variety of data. The measurements were influenced by meteorological conditions. Only classic commercial drones were used and no stealth material nor stealth shape were considered either. We did not equipped our drones with anything which would improve the signal quality.

The target-radar distance was $1.5\,\text{km} \pm 250\,\text{m}$ with only one target at a time. The measurements were performed at S band ($f_0 = 3\,\text{Gz}$) with PRF $= 10\,\text{kHz}$ on horizontal polarization. The Doppler ambiguity is $\Delta f_{max} = \frac{\text{PRF}}{2}$. If we consider a target at radial speed $v = 100\,\text{km/h}$, it implies a main Doppler shift of $\Delta f_D = \frac{2 f_0 v}{c} = 556\,\text{Hz}$ with c the light speed. In addition, the maximum micro-Doppler shift induced by a rotor of radius L is $\Delta f_{\mu\text{-}D} = \frac{8\pi L f_0}{c} \frac{\text{RPM}}{60}$ [16]. For typical RPM $=$ 5–6$\cdot 10^3$ round/min and $L = 10\,\text{cm}$, the shift $\Delta f_{\mu\text{-}D}$ is 2–3 kHz. Thus, our PRF allows us to observe the micro-Doppler effect produced by the drone rotors, $\Delta f_{max} > \Delta f_D + \Delta f_{\mu\text{-}D}$.

In Fig. 2, we present a long-time spectrogram (40 s observation) obtained by concatenation of spectrum such as the one presented in Fig. 1. The signal is compared to the theoretical impact of the different rotors obtained thanks to the drone log file. We point out that the rotor speeds do not vary much during the 300 ms period and thus the relatively stationarity hypothesis is respected.

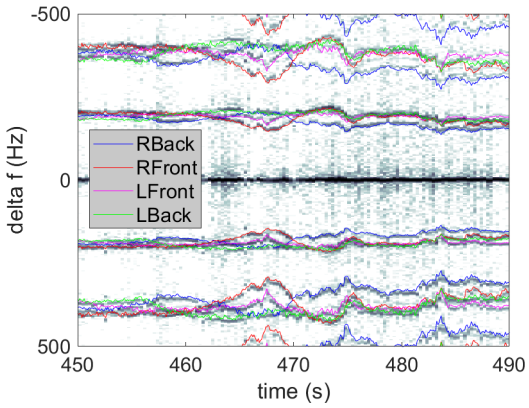


Fig. 2: Comparison between theorical data and real observation. The foreground lines correspond to theorical impact of rotors. The background is a concatenation of WSP to obtain a long-time spectrogram.

B. Deep Network configuration

The classification results we present are obtained with GoogLeNet [20]. The network has been chosen to compare our results with those from related works [9].

Each trajectory is split in chunks of duration equal to the observation time chosen. Each chunk is represented by all formats under study and sent as independent input to the network. The training and testing sets contain trajectories

collected on different days to assure data independence. There are 36 730 trajectory chunks in the learning set and 4 670 in the testing one. All chunks correspond to nearly three and half hours of accumulated measurement.

We stress that without day separation between the testing and training set, the classification accuracy reaches more than 98% in all configurations (even 100% for $x(t)$), much too optimistic. We thus strongly recommend that experiments should be conducted with the day separation.

We use a batch size of 256 chunks and fix a dropout of 0.6 on the first fully connected layer. In every configuration, we run 50 learnings of 25 000 iterations. The test accuracy is assessed every 200 iterations. The classification results presented correspond to the maximum of the mean accuracy and its associated confidence interval at 95%. The time/step values are measured on the same computer (GPU card: NVIDIA Tesla P100-PCIE-12GB).

For unidimensional (1-D) data, the network architecture is adapted to be 1-D (filters: $5 \times 5 \rightarrow 5 \times 1$, etc.). The real and imaginary components of each format are injected into the CNN as two color channels. Each data x is normalized by the formula $\frac{x-m}{M-m}$ with M, m the maximum, minimum modulus on the training database. All data has thus a modulus between zero and one without removing energy information.

V. Results

We present the classification results of each format under the different use conditions defined in Section III.

A. Reference

Format	Reference [%]
$x(t)$	75.8 $\pm$ 1.55
WSP	**98.1 $\pm$ 0.09**
CP	97.4 $\pm$ 0.11
SG	92.6 $\pm$ 0.32
CVD	94.5 $\pm$ 0.17

TABLE I: Reference use condition.

The network performs significantly better for the WSP and CP formats than the others (at least 3 points more). These two formats are based upon a good frequency / quefrency resolution despite the absence of time resolution. The CP results are 0.7 points lower than WSP. The differences might be explained by the loss of information while taking the modulus (c.f. Subsection II-C).

The network reaches 94.5% of accuracy for the CVD format which is 1.9 points above the SG performance despite the short observation. The second FFT made on SG has a positive impact on the classification performance.

Feeding the network with $x(t)$, produces the worst result (75.8%). Other formats project the information contained in $x(t)$ into another space. Such an operation makes the differences between drones be more contrasted.

B. Robustness to noise

Our CNN obtains the best results with noisy data when using the WSP format (92.7%). This result is more than 10 points higher than for any other format. The WSP format also resists the best to noise as its performance is only 5.4 points less than for the reference. In contrast, losses observed in classification with the use of other formats after noise injection are about 15 points.

Format	Reference [%]	Noise added [%]	Δ
$x(t)$	75.8 $\pm$ 1.55	72.6 $\pm$ 1.50	3.2
WSP	98.1 $\pm$ 0.09	**92.7 $\pm$ 0.15**	**5.4**
CP	97.4 $\pm$ 0.11	80.9 $\pm$ 0.30	16.5
SG	92.6 $\pm$ 0.32	73.4 $\pm$ 0.42	19.2
CVD	94.5 $\pm$ 0.17	79.2 $\pm$ 0.26	15.3

TABLE II: Robustness to noise.

The classification with the CP and CVD formats has similar results, around 80%. Surprisingly, the results with CP, based as WPS on long integration, are not better than the WSP one. CP was created to resist better than WSP to noise [18].

Similarly to the reference, the classification performance for SG is relatively low (73.4%) and worse than for CVD. The gap between these two formats have even increased from 1.9% to 5.8%. A possible explanation is the noise subduction by the second FFT in CVD.

The classification with the $x(t)$ format is poor.

C. Short observation duration

Format	Reference [%]	Short signal [%]	Δ
$x(t)$	75.8 $\pm$ 1.50	67.1 $\pm$ 1.01	8.7
WSP	98.1 $\pm$ 0.09	93.7 $\pm$ 0.12	4.4
CP	97.4 $\pm$ 0.11	**94.0 $\pm$ 0.12**	**3.4**
SG	92.6 $\pm$ 0.32	87.3 $\pm$ 0.23	5.3
CVD	94.5 $\pm$ 0.17	79.3 $\pm$ 0.26	15.2

TABLE III: Short observation

Although CP and WSP have comparable classification values (around 94%), CP withstands better than WSP to the reduction of observation time (3.4% instead of 4.4% in Δ column). Shortening the observation time diminishes the number of peaks produced by the rotors. The remaining peaks, however, are sufficient for the CNN to discriminate the drones.

Once again, the classification with SG is low compared to WSP and CP with only 87.3% of success rate and a degradation of 5.3 points due to the short observation.

The CVD classification performance decreases drastically (3 times as much as for any other format). It confirms the statement given in Subsection II-E: this format is made for long observations. For short ones, the second FFT is performed on too few profiles and produces thus a poor resolution in terms of frequency cadence. For observations longer than the reference case, we expect that its performance might increase.

For each format, shortening the observation time decreases the network performance ($x(t)$ too low to be considered). Despite that the short observation duration is sufficient to capture several rotor speeds, not all the characteristics needed to identify the target are present. We believe that the dynamic evolution of the drone is used to classify. In a real scenario, the observation time depends in particular on the environment (if the drone can be hidden by obstacles). Thus, the latter has a strong impact on the results.

D. Facility of training

Format	Dimension	Time/Step [s]	Time [s]
$x(t)$	$1 \times 3000 \times 2$	0.39	79
WSP	$1 \times 2400 \times 2$	0.32	449
CP	$1 \times 3000 \times 2$	0.39	394
SG	$20 \times 300 \times 2$	0.18	425
CVD	$20 \times 120 \times 2$	0.10	163

TABLE IV: Facility of training.

The time per step and the time to reach 95% of the maximum accuracy value are given in seconds.

The time per step for 1-D data is longer than for 2-D data of same size because the 2×2 max-pool is replaced by a 2×1 one and, consequently, 1-D data is less reduced on deeper levels of the network. However, the CNN does not necessarily need more iterations to reach 95% of its maximum accuracy. The time to reach this maximum may be twenty times longer. Thus each training takes between several minutes and three hours.

Except for the fast training with CVD, the other formats have similar training speed. WSP is the slowest format. For $x(t)$, only few steps are needed. The time is therefore significantly shorter than the other formats.

VI. Conclusions

Table V summarizes our results. This qualitative evaluation is based on a trade-off between absolute classification and the Δ column. For each case, we assess one of the grade —, -, 0, +, ++ (from very bad to very good, respectively).

Format	Use conditions			
	A	**B**	**C**	**D**
$x(t)$	—	—	—	++
WSP	**++**	**++**	**+**	**-**
CP	++	-	++	0
SG	+	—	0	0
CVD	+	-	—	+

TABLE V: Qualitative analysis of the CNN performance
A: Reference, B: Robustness to noise,
C: Short observation, D: Facility of training.

Using WSP input, produces better classification results than all other formats for the reference case. Moreover, the classification performance decreases significantly less with a weak SNR than for other formats. WSP is also robust to shortening of the observation time, only CP being more resilient. We therefore recommend using the WSP format to classify drones according to micro-Doppler characteristics, particularly for data with a poor SNR.

The network results with the CP format are close to those obtained with WSP for the reference case while resisting considerably less than it to SNR degradation. CP is, however, more robust to shorter observations. When doing classification with data with a good SNR and a short observation, we recommend comparing the results obtained with CP and WSP.

The networks with CVD, SG or $x(t)$ inputs are always outperformed by the one with WSP, we thus do not recommend them.

We remind that the training and testing sets should be collected on different days as the random data repartition might improve the network performance artificially.

The different formats contain at best as much information as the $x(t)$ format while giving better performance. Input formats have thus an influence on the performance despite the extractions made by a CNN. It might be due to the small amount of data and the environmental bias inherent to radar measurements.

We also observe that all formats we studied are not specifically designed for classification with a CNN. An interesting research line would be to produce a micro-Doppler signature format dedicated for CNN, outperforming all-purpose formats.

References

[1] J. Farlik, M. Kratky, J. Casar, and V. Stary, "Radar cross section and detection of small unmanned aerial vehicles," in *Proc. IEEE ME*, 2016.

[2] V. C. Chen, "Analysis of radar micro-Doppler with time-frequency transform," in *Proc. IEEE SSAP*, 2000.

[3] V. C. Chen, F. Li, S.-S. Ho, and H. Wechsler, "Micro-Doppler effect in radar: phenomenon, model, and simulation study," *Proc. IEEE AES*, vol. 42, 2006.

[4] F. Fioranelli, M. Ritchie, H. Griffiths, and H. Borrion, "Classification of loaded/unloaded micro-drones using multistatic radar," *Electronics Letters*, vol. 51, no. 22, 2015.

[5] P. Molchanov, R. I. Harmanny, J. J. de Wit, K. Egiazarian, and J. Astola, "Classification of small UAVs and birds by micro-Doppler signatures," *IJMWT*, vol. 6, no. 3-4, 2014.

[6] L. Fuhrmann *et al.*, "Micro-Doppler analysis and classification of UAVs at ka band," in *Proc. IEEE IRS*, 2017.

[7] M. S. Seyfioğlu, S. Z. Gürbüz, A. M. Özbayoğlu, and M. Yüksel, "Deep learning of micro-Doppler features for aided and unaided gait recognition," in *Proc. IEEE RadarConf*, 2017.

[8] D. A. Brooks, O. Schwander, F. Barbaresco, J.-Y. Schneider, and M. Cord, "Temporal deep learning for drone micro-Doppler classification," in *Proc. IEEE IRS*, 2018.

[9] B. Kim, H. Kang, and S.-O. Park, "Drone classification using convolutional neural networks with merged Doppler images." *IEEE GRSL*, vol. 14, no. 1, 2017.

[10] J. S. Patel, C. Al-Ameri, F. Fioranelli, and D. Anderson, "Multi-time frequency analysis and classification of a micro-drone carrying payloads using multistatic radar," *The Journal of Engineering*, vol. 2019, no. 20, pp. 7047–7051, 2019.

[11] Y. Kim and T. Moon, "Human detection and activity classification based on micro-Doppler signatures using deep convolutional neural networks," *IEEE GRSL*, vol. 13, no. 1, 2016.

[12] K. Ishak, N. Appenrodt, J. Dickmann, and C. Waldschmidt, "Human motion training data generation for radar based deep learning applications," in *Proc. IEEE ICMIM*, 2018.

[13] R. P. Trommel, R. Harmanny, L. Cifola, and J. Driessen, "Multi-target human gait classification using deep convolutional neural networks on micro-Doppler spectrograms," in *Proc. IEEE EuRAD*, 2016.

[14] Y. Kim, M. Nazaroff, and D. Oh, "Extraction of micro-Doppler characteristics of drones using high-resolution time-frequency transforms," *MOTL*, vol. 60, no. 12, 2018.

[15] R. Harmanny, J. De Wit, and G. P. Cabic, "Radar micro-Doppler feature extraction using the spectrogram and the cepstrogram," in *Proc. IEEE EuRAD*, 2014.

[16] J. Martin and B. Mulgrew, "Analysis of the theoretical radar return signal form aircraft propeller blades," in *Proc. IEEE RadarConf*, 1990.

[17] T. H. Melino R., Bourne C., *Modelling Helicopter Radar Backscatter*. ADF, DST, Electronic Warfare and Radar Division, 2011.

[18] B. P. Bogert, M. Healy, and J. Tukey, "The quefrency alanysis of time series for echoes," *Time series analysis*, 1963.

[19] D. S. Williamson, Y. Wang, and D. Wand, "Complex ratio masking for monaural speech separation," *IEEE TASLP*, vol. 24, no. 3, 2016.

[20] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE CVPR*, 2015.