Group Counting Using Micro-Doppler Signatures From a 77GHz FMCW Radar

1st Dejvi Cakoni
OPERA - WCG
Université Libre de Bruxelles
Brussels, BE
dejvi.cakoni@ulb.be

2nd Laurent Storrer
OPERA - WCG
Université Libre de Bruxelles
Brussels, BE
laurent.storrer@ulb.be

3rd Philippe De Doncker
OPERA - WCG
Université Libre de Bruxelles
Brussels, BE
philippe.dedoncker@ulb.be

4th François Horlin
OPERA - WCG
Université Libre de Bruxelles
Brussels, BE
francois.horlin@ulb.be

***Abstract*—People counting and detection technologies have shown great versatility in various scenarios and have become an important tool for event organizers and city planners to optimize their operations. This paper presents a novel approach for people counting using Micro-Doppler Signatures (MDS) extracted from a Frequency-Modulated Continuous-Wave (FMCW) radar operating at 77GHz. The system utilizes the unique gait model of each individual , which results in a distinct instantaneous velocity over time, to generate the MDS that are later used to classify groups of different sizes with a Convolutional Neural Network (CNN). Those results are compared with using existing CNNs for image classification, in a transferred learning approach. The proposed system overcomes the limitations of existing camera-based people counting techniques such as the need for a clear line of sight and being affected by lighting conditions.**

***Index Terms*—group counting, radar signal processing, 77 GHz FMCW radar, CNN, micro-Doppler signature**

I. Introduction

In recent years, with the increased concern in public safety, there has been a growth in the demand for crowd surveillance and safety management systems. The estimation of crowd dynamics can help in preventing unanticipated accidents or issues in case of mass events or be of use for city planners to improve the daily commutes of its citizens. These systems can be implemented in various ways as, for example, image or video-based techniques. However, radar-based crowd monitoring systems are being considered due to their non-invasive properties and ability to work in low lighting conditions, which the previous systems are lacking.

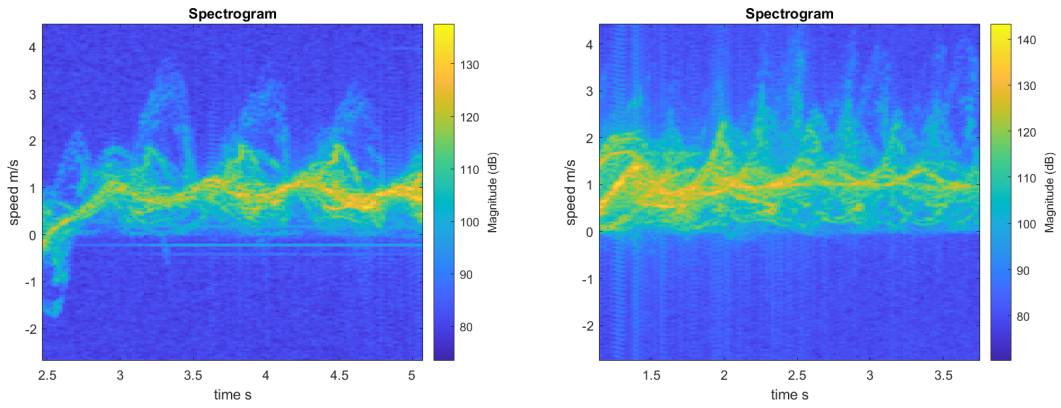
When it comes to radar-based people counting systems, several radar types and inputs have been used. Some of the existing research in literature consider an indoor, office-like environment where a few individuals (less than ten in practice) are mobile. These systems are based on the impulse-radio ultrawideband (IR-UWB) waveform, which compared to the Frequency Modulated Continuous waveform (FMCW) provides a much better range resolution but poor Doppler resolution. Since people in this environment move at very low speeds, the radar mostly relies on the range information to estimate the number of individuals in the room [1]. Another exploitable input can be the power spectral density (PSD) for applications with wider regions of interest (ROI) to improve people counting. [2]. The use of range-time maps obtained from a single-channel stepped-frequency continuous wave radar (SFCW) have also been explored for counting [3]. Passive radars have also been a growing area of radar research where features extracted from range-Doppler maps (RDM) have been used for counting [4] along with spectrograms [5]. However, these spectrogram estimates are built by observing WiFi signals at a frequency much lower than what we are considering here and considering an office-like scenario. Low-accuracy estimates achieved with a mm-wave FMCW radar can also be improved by using information coming from other devices like cameras [6] or by finely observing the vital signs like the heartbeat or the breathing rates with the radar. [7]. On the contrary to existing work focusing mainly on indoor environments, we will target an outdoor pedestrian street scenario where people are typically walking together in groups. Furthermore, distinctly in this work we will use the Micro-Doppler Signatures (MDSs) extracted from a FMCW Radar at 77GHz as input to a Convolutional Neural Network (CNN). We will compare these results with a transfer learning approach, using other pre-trained CNNs trained for image classification.

The rest of the paper is organized as follows : Section II describes the fundamentals of the FMCW radar. Next, Section III explains the human gait modelling with the experimental results and the simulation scenario used. Section IV presents the CNN architecture along with the transfer learning approach and the results achieved. Finally, we conclude this paper and discuss future directions in Section V.

II. System Architecture

A. FMCW Radar system

Frequency-Modulated Continuous-Wave (FMCW) radar is a type of radar that operates by transmitting a continuous wave



(a) Single person MDS

(b) Two people MDS

Fig. 1: Experimental results

signal that is modulated with a linear frequency ramp. This ramp causes the transmitted signal to continuously increase or decrease in frequency over time. This transmitted signal is called a chirp. The FMCW signal is composed of a finite series of K chirps, each with an instantaneous frequency linearly increasing with the time.

When the transmitted signal encounters a target object, some of the signal is reflected back to the radar receiver. The received signal is then mixed with the transmitted signal and low-pass filtered to cancel out replicas at twice the carrier frequency resulting from the mixing. The resulting frequency is proportional to the distance between the radar and the target object. By analyzing the resulting frequency signal, the FMCW radar can determine the range, speed and, in case of multiple antennas, Angle of Arrival (AoA) of target objects. Focusing on chirp k and denoting each chirp duration by T and the frequency bandwidth swept as B the time can be expressed as :

$$t = kT + t' \tag{1}$$

where $k = 0, \cdots, K-1$ and $t' \in [0, T]$. The instantaneous frequency of the transmitted signal is expressed as :

$$f_i(t) = \beta t' \tag{2}$$

where $\beta = \frac{B}{T}$ is called the frequency slope. The transmitted signal is then mathematically expressed as:

$$s(t) = \cos(2\pi f_c t + \phi_i(t)) \tag{3}$$

where f_c is the radar carrier frequency and $\phi_i(t)$ is the instantaneous phase resulting from the FMCW modulation, equal to :

$$\begin{aligned} \phi_i(t) &= 2\pi \int_{u=0}^{t} f(u)\, du \\ &= \pi k \beta T^2 + \pi \beta t'^2 \end{aligned} \tag{4}$$

At the receiver, after mixing, the resulting baseband signal caused by a single target reflection is :

$$x(t) \approx \kappa \exp(j2\pi f_B t') \exp(j2\pi f_D k T) \tag{5}$$

where κ is a complex factor that integrates the gain and all constant phase terms and f_B and f_D the so called beat and Doppler frequency respectively. By measuring f_D and f_B the targets speed and range can be resolved respectively since they are defined as :

$$f_D = 2\frac{v f_c}{c} \tag{6}$$

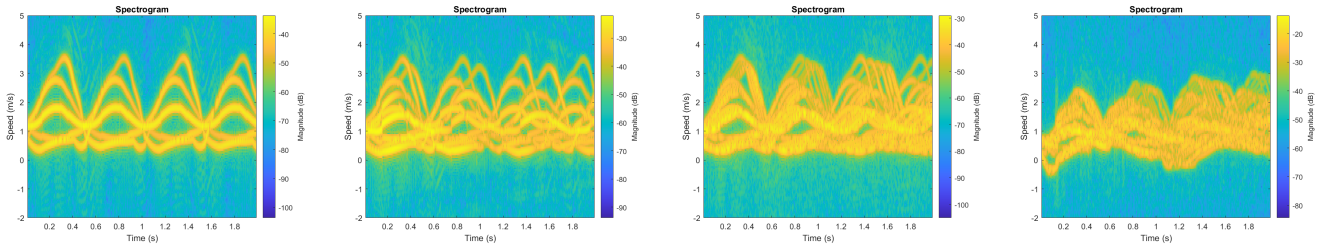
$$f_B = 2\frac{R_0 \beta}{c} \tag{7}$$

where v denotes the targets speed , R_0 the targets range and c the speed of light.

B. Radar Signal Processing

A 2D matrix of size $M \times N$ is formed by acquiring and sampling the mixed signal $x(t)$ across consecutive chirps for a single transmit antenna, with M being the number of transmitted chirps and N the number of samples per chirp. Next the Range-Doppler Map (RDM) is computed by first taking a Fast Fourier Transform (FFT) along the fast time for all chirps to obtain the so-called Range Profile (RP) containing the range information of the targets, followed by another FFT along the slow time to obtain Doppler information. Before performing the respective 1D FFTs, a mean subtraction is performed along the slow time to remove contributions from static objects.

However, in cases of groups walking together it is not possible in the RDM to distinguish and count the number of people as they appear as a single peak in the RDM. As discussed previously the frequency components of the targets will vary over time. In such way, the standard Fourier Transform is not suitable since it projects the signal on infinite sinusoids which are totally not localized in time and thus, it provides



(a) Single person MDS (b) Four people MDS (c) Seven people MDS (d) Eleven people MDS

Fig. 2: Examples of simulated MDS for different classes

the frequency information averaged over the whole signal time interval. In these cases, it is necessary to move from mono-dimensional functions to bi-dimensional functions (functions depending on both time and frequency) such as the Short Time Fourier Transform (STFT). Thus, our radar processing is as follows :

- Determine the RDM using a 2-D Fourier Transform.
- In the RMD, detect the group by finding the maximum power peak.
- Extract and concatenate the resulting peak index across all chirps in the RP.
- Perform STFT on the concatenated signal to extract the spectrograms i.e, MDS.

III. Human Gait Modelling

A. Experimental Results

This work is based on the Texas Instruments AWR1843 FMCW radar operating at 77GHz. For the purposes of this work, experimental data were collected to validate the system model used in the previous sections. Future work will include an extensive measurement campaign to validate our used techniques in real-world data. Experiments were done using the Texas Instruments AWR1843 FMCW radar operating at 77GHz [9]. A summary of the selected radar recording parameters can be seen in Table I.

Waveform parameters	Value
Carrier Frequency	77 GHz
Chirp bandwidth	3.5 GHz
Chirp duration	157 μs
Range resolution	0.04 m
Speed resolution	0.09 m/s
Maximum detectable range	21 m
Maximum detectable speed	6 m/s

TABLE I: Recording parameters for the FMCW radar

The experiments were performed in an indoor room with a single and two targets moving away from the radar. The resulting Micro-Doppler Signatures (MDS) can be seen in Fig. 1. It can be observed that in the measurement results for a single person and for two people, the maximum power in the measured MDS is observed in the torso, along with the lower parts of the arms and legs. We will leverage these results later to model the gait of our targets. Additionally, in cases of multiple people, it can be observed that their MDS differ in magnitude and phase, making them a useful tool for people counting.

B. Simulation

To study how the MDS evolves with an increasing number of targets in the scene, it is necessary to resort to simulations based on either mathematical or empirical models. One frequently used empirical model for generating micro-Doppler gait signatures is the global human walking model developed by Boulic, Magnenat-Thalman, and Thalman [10]. However, in this work, we do not consider all these points but instead use the experimental results to select the most significant points. As discussed previously, the most significant contributions to the MDS were from the torso, lower legs, and lower arms. Thus, only these points are used to model the gait of our targets, superimposed with the average speed of the groups. Examples of simulated MDSs for different group sizes are shown in Fig. 2. It should be noted that the simulated MDS represent targets moving away from the radar, which is considered as a positive frequency shift.

IV. CNN and Simulation Results

Deep learning is a branch of machine learning that focuses on automatically generating customized features using a series of nonlinear operations. Specifically, these algorithms consist of a sequence of functions that enable the learning of more intricate concepts by combining simpler functions in a stacked manner such that :

$$f^l(x) = \sigma(W^l x + b^l), \forall l \in [0, L] \tag{8}$$

where x denotes an input vector, σ a piecewise nonlinear function, W^l and b^l describe the layer-specific weights and biases respectively and L the number of layers in the network. Despite deep neural networks being researched on for several decades, they have gained significant attention and achieved impressive performance following a notable breakthrough in the ImageNet Large-Scale Visual Recognition Competition in 2012 [11]. The recent success of deep learning can be attributed to the availability of large datasets, affordable computational power and resources, algorithmic advancements, and a culture of open innovation. Convolutional Neural Networks (CNNs) are specialized neural networks which utilize locally connected neurons with shared weights, allowing

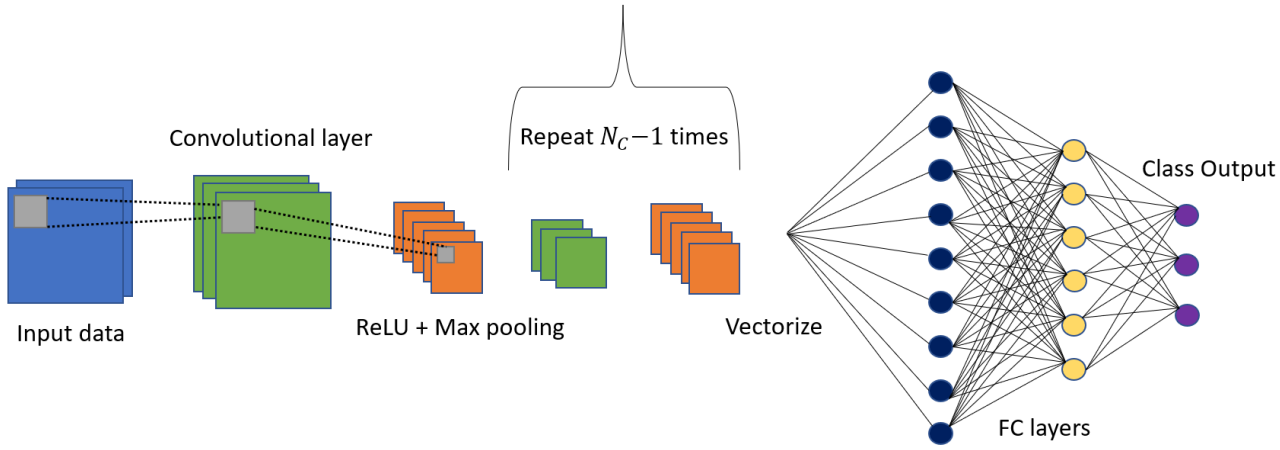


Fig. 3: CNN Architecture

convolutional filters to operate on small receptive fields of input data in a sliding-window manner. CNNs exhibits a grid-like topology, where different filters evolve to become specific feature detectors, starting from low-level color and edge detectors in early layers and progressing to high-level object detectors in later layers. The key distinction from a standard Feed-Forward Neural Network (FFNN) lies in the use of convolutions instead of plain matrix multiplications. To formalize, such convolutions are defined as :

$$\begin{aligned} D_{ij} &= (X * K)_{ij} \\ &= \sum_m \sum_n X_{i+m,j+n} K_{mn} \end{aligned} \tag{9}$$

with D representing the resulting feature map, X a 2-D input and K a filter $\in \mathbb{R}^{m \times n}$. Thus, (8) becomes :

$$f_j^l(X) = \sigma(X \times W_j^l + b_j^l), \forall l \in [0, L] \tag{10}$$

In addition to weight sharing, a technique called pooling is utilized to efficiently reduce the number of parameters and data size. Pooling involves averaging or maximizing the responses of cells arranged in an $n \times m$ grid, preserving essential information while reducing the overall data size [12]. This paper focuses on these types of neural networks, applied to MDSs extracted from a FMCW Radar.

A. Dataset Simulation and Class Lables

We simulate varying group sizes (1-12 people) in a pedestrian street. For each group size 100 MDS are simulated and generated leading to a dataset of 1200 MDS samples. These MDSs are then fed to a CNN in order to perform a classification task to estimate the group sizes. Some examples of the MDSs generated can be seen in Fig. 2. The goal is to count and classify different groups of people, thus we build our classes based on intervals of number of people. Considering 3 groups classes, the class labels decided are as follows :

- Class 1 : 1-4 people - Low sized group
- Class 2 : 5-8 people - Medium sized group
- Class 3 : 9-12 people - High sized group

B. CNN Architecture

A classical CNN architecture is implemented here, and displayed in Fig. 3. The network's structure was thoughtfully crafted through thorough experimentation with a wide range of hyperparameters, including the number of layers (convolutional, pooling, or fully connected), the size and quantity of filters, and other relevant factors. It consists of a feature extraction part with $N_C = 3$ convolutional blocks, and a classification part with $N_{FC} = 3$ fully connected (FC) layers followed by a softmax layer. Each MDS is scanned by the convolutional layers, followed by a rectifier linear unit (ReLU) layer and a max pooling. After each set of convolutions followed by the ReLU and the max pooling, the size of the convolutional filters is decreased and their number is increased. This is done to scan the MDS at each step with a finer resolution filter so that the CNN can extract different and finer features at each step. We start with 16 filters of size 7×7 in the first layer, to 32 filters of size 5×5 in the second layer and 64 filters of size 3×3 in the last convolutional stage. The ReLU activation function was chosen for its ability to handle the vanishing gradient problem [13]. The dataset is split 70%, 20% and 10% between training, validation and testing respectively.

C. Classification Results

As can be seen in Fig. 4 the proposed CNN architecture achieves an accuracy of 80% on average for the considered classes on the testing set. Especially for the low and high sized groups, the model achieves a better accuracy as the MDS are quite distinct compared to the medium sized group. Also it is worth noting that the miss classification errors do not exceed

(a) Custom CNN (b) AlexNet (c) GoogleNet

(d) ResNet-18 (e) ResNet-50

Fig. 4: Confusion matrices for the networks considered

more than one class i.e we never classify the low sized group and a high sized group and vice versa.

D. Transfer learning

Transfer learning is a machine learning technique that utilizes knowledge gained from one task to improve performance on a different but related task. Instead of starting from scratch, transfer learning allows models to benefit from pre-trained knowledge and adapt it to new tasks. It is particularly useful when there is limited labeled data available for the target task. In this paper, we compare the results achieved by our custom CNN to other networks widely used in image classification tasks. We will focus on networks pre-trained in the ImageNet database and fine tune them specifically for our dataset. Since these networks were trained to classify images into 1000 object categories, such as keyboard, mouse, pencil, and many animals, they have learned rich feature representations for a wide range of images. The goal is to leverage these networks for high-level feature extraction such as edges done in the early layers and re-train them in the later layers for feature extraction specific to our counting problem. In this paper, the networks considered for transfer learning are :

- AlexNet: AlexNet is a pioneering deep convolutional neural network architecture that revolutionized image recognition in 2012 by effectively using convolutional layers [11].
- GoogleNet: Also known as Inception, GoogleNet is a deep convolutional neural network architecture that introduced inception modules, enabling efficient information flow and achieving high accuracy with fewer parameters. [14]
- ResNet18: ResNet18 is a variation of the ResNet architecture, incorporating skip connections to address the challenge of vanishing gradients and allowing the training of deeper neural networks [15].
- ResNet50: ResNet50 is an extended version of the ResNet architecture, featuring 50 layers for increased depth and capacity, resulting in improved accuracy across various computer vision tasks [15].

A summary of the depth and number of parameters for these networks is shown in Table II.

Network	Depth	Number of parameters
AlexNet	8	61 million
GoogleNet	22	6.8 million
ResNet-18	18	11 million
ResNet-50	50	25 million

TABLE II: Networks Considered

Thus, in our proposed solution, we substitute the final convolutional, fully connected, softmax and classification layer to match and learn features related to our problem. We re-train the entire network on our dataset, but we give the newly added layers a 20 times higher weight and bias learning rate factor and decrease the initial learning rate significantly. This is done so that the network learns faster in the new layer and slow down learning in the transferred layers. In that way we preserve the high-layer feature extractor while learning new

features relevant to our dataset. A summary of the achieved results can be seen in Fig 4 and Table III.

Network	Input Size	Average testing accuracy
AlexNet	224×224×3	≈ 82%
GoogleNet	224×224×3	≈ 86%
ResNet-18	224×224×3	≈ 87%
ResNet-50	224×224×3	≈ 86%

TABLE III: Networks Considered

V. Conclusion and Future Work

In conclusion, we investigated the problem of radar based group counting using micro-Doppler signatures. We proposed a simulator based on the Boulic, Magenat-Thalman and Thalman model to generate micro-Dopper signatures for varying group sizes and used experimental measurements of the human gait with a radar to select the relevant body parts. We tackled counting as a classification problem, and applied a CNN on the generated MDS. This approach achieved high accuracy results for counting. We finally used fine-tuning transfer learning on CNNs used in the ImageNet challenge and achieved better accuracies than what was previously possible with our custom CNN.

Future work includes an extensive measurement campaign and dataset collection, comparing the CNN architecture proposed to other Machine Learning methods and tackling larger group sizes.

References

[1] J. -H. Choi, J. -E. Kim, N. -H. Jeong, K. -T. Kim and S. -H. Jin, "Accurate People Counting Based on Radar: Deep Learning Approach," 2020 IEEE Radar Conference (RadarConf20), 2020, pp. 1-5, doi: 10.1109/RadarConf2043947.2020.9266496

[2] Jae-Ho Choi, Ji-Eun Kim, and Kyung-Tae Kim. People Counting Using IR-UWB Radar Sensor in a Wide Area. IEEE Internet of Things Journal, 8(7):5806–5821, April 2021. IEEE Internet of Things Journal.

[3] Y. Jia et al., "ResNet-Based Counting Algorithm for Moving Targets in Through-the-Wall Radar," in IEEE Geoscience and Remote Sensing Letters, vol. 18, no. 6, pp. 1034-1038, June 2021, doi: 10.1109/LGRS.2020.2990742.

[4] Ali El Amine and Valery Guillet. Device-Free People Counting Using 5 GHz Wi-Fi Radar in Indoor Environment with Deep Learning. In 2020 IEEE Globecom Workshops (GC Wkshps, pages 1–6, December 2020.

[5] C. Tang, W. Li, S. Vishwakarma, K. Chetty, S. Julier and K. Woodbridge, "Occupancy Detection and People Counting Using WiFi Passive Radar," 2020 IEEE Radar Conference (RadarConf20), Florence, Italy, 2020, pp. 1-6, doi: 10.1109/RadarConf2043947.2020.9266493

[6] M. Stephan, S. Hazra, A. Santra, R. Weigel and G. Fischer, "People Counting Solution Using an FMCW Radar with Knowledge Distillation From Camera Data," 2021 IEEE Sensors, 2021, pp. 1-4, doi: 10.1109/SENSORS47087.2021.9639798.

[7] J. Weiß, R. Pérez and E. Biebl, "Improved People Counting Algorithm for Indoor Environments using 60 GHz FMCW Radar," 2020 IEEE Radar Conference (RadarConf20), 2020, pp. 1-6, doi: 10.1109/RadarConf2043947.2020.9266607.

[8] L. Servadei, H. Sun, J. Ott, M. Stephan, S. Hazra, T. Stadelmayer, D. S. Lopera, R. Wille, and A. Santra, "Label-aware ranked loss for robust people counting using automotive in-cabin radar," 2021. [Online]. Available: https://arxiv.org/abs/2110.05876

[9] Texas Instruments, "AWR1843 Single-Chip 77- to 79-GHz FMCW Radar Sensor datasheet (Rev. C)" , https://www.ti.com/lit/ds/symlink/awr1843.pdf?ts=1682628530350.

[10] Boulic, R., Thalmann, N.M. & Thalmann, D. "A global human walking model with real-time kinematic personification". The Visual Computer 6, 344-358 (1990). https://doi.org/10.1007/BF01901021

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in Proc. Adv. Neural Inf. Process. Syst., 2012, pp. 1097–1105.

[12] B. Vandersmissen et al., "Indoor Person Identification Using a Low-Power FMCW Radar," in IEEE Transactions on Geoscience and Remote Sensing, vol. 56, no. 7, pp. 3941-3952, July 2018, doi: 10.1109/TGRS.2018.2816812.

[13] M. S. Seyfioğlu, A. M. Özbayoğlu and S. Z. Gürbüz, "Deep convolutional autoencoder for radar-based classification of similar aided and unaided human activities," in IEEE Transactions on Aerospace and Electronic Systems, vol. 54, no. 4, pp. 1709-1723, Aug. 2018, doi: 10.1109/TAES.2018.2799758.

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

[15] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.