Human Activity Classification in Underwater using Sonar and Deep Learning

Hovannes Kulhandjian†, Narayanan Ramachandran†, Michel Kulhandjian‡, Claude D'Amours‡

†Department of Electrical and Computer Engineering, California State University, Fresno, CA 93740, U.S.A.

‡School of Electrical Engineering and Computer Science, University of Ottawa, ON, K1N 6N5, Canada

Email: {hkulhandjian@csufresno.edu, naranchs@mail.fresnostate.edu, mkk6@buffalo.edu, cdamours@uottawa.ca}

ABSTRACT

In this paper, we study the classification of human activity on the surface of a body of water using sonar. In particular, we investigate the classification of three different swimming styles; freestyle, butterfly, and backstroke. Experiments are conducted in a swimming pool to capture acoustic micro-Doppler signatures produced by the different swimming styles. Two acoustic hydrophones are used underwater; one to transmit a single tone signal in the direction of a swimmer and the other to receive the reflected waveform from the swimmer's body. We apply joint time-frequency analysis on the received acoustic signal to extract the micro-Doppler signatures present in the spectrogram. Each of these swimming style activities presents their own unique micro-Doppler signatures. To classify the acoustic micro-Doppler signatures, we explore a deep convolution neural network (DCNN) algorithm. Spectrogram can be considered as an image in which case applying DCNN can serve well for feature recognition purposes. We show that using the spectrogram images the DCNN algorithm can classify different swimming styles performed on the surface of the water with fairly high accuracy. Using the collected data set, we performed experiments where we used 80% of the data for training and the remaining 20% for validation purposes. The DCNN algorithm averaged 93.7% accuracy during training while it had a 90.8% average validation accuracy.

Keywords

Activity classification in underwater, swimming style classification, micro-Doppler sonar, deep convolution neural network (DCNN).

1. INTRODUCTION

Detection and classification of human activity on the water surface are essential in surveillance, border patrol, and search-and-rescue operations [1, 2]. Furthermore, it can be useful to detect a drowning person in a swimming pool. Some research has been conducted on human detection and activity classification on land using radar micro-Doppler signatures [3, 4], but underwater human detection and activity classification has not been extensively studied. When illuminated by sonar, humans produce micro-Doppler signatures due to moving limbs. The micro-Doppler signatures for different activities are unique and therefore can be used to detect and classify human activity [5]. A number of research works have studied classification of human activities and objects using different methods. In the following paragraphs, we briefly discuss some of the related works.

Balleri *et al.* in [6], use an ultrasound radar system to collect micro-Doppler signatures of personnel targets performing different actions. For training and recognizing the different human motions, k-nearest neighbor (kNN) and Naïve Bayesian classifiers are used. Experimental results show that acoustic radar can successfully be used over short ranges to collect micro-Doppler signatures of moving target and classify them.

Victor *et al.* in [7], explore convolution neural network (CNN) to automatically detect discrete events in continuous video such as swimming strokes. The CNN algorithm is employed to learn a mapping from a window of frames to a point on a smooth $1D$ target signal in which the peaks denote the location of a stroke. Experimental results show that with fairly high accuracy over 90% they were able to locate stroke spikes.

Williams in [8] uses deep convolution neural network (DCNN) for underwater target classification in synthetic aperture sonar (SAS) imagery. Objects such as dummy mine shapes, mine-like targets and man-made objects were deployed on the seafloor. Sonar surveys are conducted over the area with an autonomous underwater vehicle (AUV) to gather the data. The collected data was processed to produce scene-level SAS images. Within these images the differences between similar classes of object were learned by DCNN.

Einfalt *et al.* in [9], estimate human pose in real-world videos of swimmers. CCN algorithm is applied to infer the required pose information such as different swimming styles for detection and classification of different body joints. Evaluations of the proposed scheme demonstrated an average accuracy of 95.7% for classifying different swimming styles.

Kashyap *et al.* in [10], present the micro-Doppler signatures of two underwater unmanned vehicles (UUV) from simulated acoustic radar data. The UUV and the radar are assumed to be fully submerged in underwater. The radar scattered return signals are generated using primitive based modeling of the vehicles at 30 kHz. Two types of motion are considered; vehicles moving tangentially with respect to the radar and others moving towards the radar. Simulation results of a simple model show that different UUVs demonstrate unique micro-Doppler features that is a function of the length of the propeller blades, their rotation rate as well as the orientation of the target motion with respect to the radar.

Zhu *et al.* in [11], present an automatic target recognition for sonar on board of UUVs. Target features are extracted by the CNN algorithm operating on sonar images then support vector machine (SVM) is used to classify them. Matched filter is used in target recognition

WUWNet'19, October 23 - 25, 2019, Atlanta, GA, USA.

ACM ISBN 978-1-4503-7740-9/19/10 ... $15.00.

https://doi.org/10.1145/3366486.3366509

while the target classification is achieved with the trained SVM classifier based on feature extracted by the CNN algorithm.

Park *et al.* in [12], employ DCNN algorithm for human aquatic activity classification based on radar sensor. Human activity is collected in a form of micro-Doppler spectrograms, which are used in the DCNN for training and classification purposes. It is reported that the convolutional feature-based scheme has fairly low accuracy 45.1%. Applying transfer learning on pre-trained data and fine tuning the DCNN parameters, the accuracy was improved to 80.3%.

Unlike the previous studies, which mainly focus on detection and classification of human activity using radar, video images and sonar based on simulations only; in this work, we experiment the classification of human activity on the surface of a body of water based on acoustic micro-Doppler signatures and deep learning. After applying joint time-frequency analysis, we observe the micro-Doppler signatures present in a spectrogram. We investigate classification of three different swimming styles; freestyle, butterfly, and backstroke. Experiments are conducted in a swimming pool to collect a data set of spectrograms, which contain the micro-Doppler signatures present in different swimming styles. We develop a sonar system by deploying two Teledyne RESON TC4013 hydrophones [13] in the swimming pool used for collecting the reflected acoustic waveforms. The signals from the acoustic sensor are digitized by universal software radio peripheral (USRP) N210 [14] data acquisition board with the sampling rate of 195 kS/s. The experimental data are collected, preprocessed and analyzed in Matlab. Each of this swimming style activity presents their own micro-Doppler signatures. To recognize the micro-Doppler signatures, we explore DCNN technique, which is considered one of the most successful deep learning algorithms for image recognition [15]. Spectrogram can be considered as an image thus they can be used in the DCNN algorithm for the feature recognition. From the underwater measured data, we crop and collect the individual spectrograms of different swimming style strokes. 80% of data are used for training purposes and the remaining 20% are used for validation. We show that the DCNN algorithm can classify human activity on the surface of water based on spectrograms with high accuracy. Our experimental results based on the underwater collected data reveal that the training accuracy was 93.7%, and the validation accuracy was 90.8%.

The rest of the paper is organized as follows. In Section 2, we discuss the basic principle of micro-Doppler signatures generated by the different swimming styles, followed by the experimentation of underwater human activity classification with Doppler sonar in Sec. 3. The DCNN algorithm is presented in Sec. 4. After illustrating our experimental results in Sec. 5, we draw our main conclusions and provide future extensions to our proposed work in Sec. 6.

2. MICRO-DOPPLER MODELING

In this section, we discuss the basic mathematical model of the micro-Doppler phenomenon induced by vibrational motions.

Rotation can be considered as a special case of vibration. In coherent sonar, the variations in range results in a phase change in the returned signal from a target. Therefore, the Doppler frequency shift in the reflected signal, which represents phase variations with time, can be used to measure the vibrations and rotations of a reflecting surface [5].

Different swimming styles are as a result of unique underwater movement patterns of the swimmer's limbs. These motions can be captured when illuminated by a sonar. When a sonar device transmits a single tone acoustic signal from underwater at a carrier frequency f_c onto a swimming person, the reflected signal contains micro-Doppler effects centered around the f_c, due to micro-motion variations of the swimmer.

The received Doppler signal as a function of time is modeled as [5]

$$s(t) = Ae^{j(2\pi f_c t + \varphi(t))}, \tag{1}$$

where A is the reflectivity of the vibrating point scatterer and $\varphi(t)$ is the time-varying phase change of the vibrating scatterer. Assuming the vibrating scatterer is set to oscillate at a frequency of f_ν, the time varying phase can be expressed as

$$\varphi(t) = \beta \sin(2\pi f_\nu t), \tag{2}$$

where $\beta = 4\pi D_\nu/\lambda$, D_ν is the amplitude of the vibration and λ is the wavelength of the transmitted signal.

Since (1) is a periodic function, it can be expanded using Fourier series as

$$s(t) = A \sum_{n=-\infty}^{\infty} c_n e^{j2\pi(f_c + nf_\nu)t}, \tag{3}$$

where c_n is the Fourier series coefficient, which is expressed as

$$c_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{j\beta \sin(2\pi f_\nu t)} e^{-jn2\pi f_\nu t} dt = J_n(\beta), \tag{4}$$

where $J_n(\beta)$ is the nth-order Bessel function of the first kind.

Substituting (4) into (3) yields

$$s(t) = A \sum_{n=-\infty}^{\infty} J_n(\beta) e^{j2\pi(f_c + nf_\nu)t}. \tag{5}$$

Equation (5) represents a micro-Doppler frequency spectrum consisting of pairs of harmonic spectral components centered around the carrier frequency f_c. The spacing between the adjacent spectral lines is governed by f_ν. Since the phase of the reflected wave expressed in (2) is time-varying, the instantaneous frequency f_D, which represents the micro-Doppler frequency induced by the vibrations of the scatterer, can be expressed as

$$f_D = \frac{1}{2\pi}\frac{\varphi(t)}{dt} = \beta f_\nu \cos(2\pi f_\nu t) \tag{6}$$

$$= \frac{4\pi}{\lambda} D_\nu f_\nu \cos(2\pi f_\nu t). \tag{7}$$

The maximum micro-Doppler frequency change is $\frac{4\pi}{\lambda}D_\nu f_\nu$, which can be used to estimate the maximum displacement of a vibrating scatterer. The micro-Doppler caused by vibration is a sinusoidal function of time at the vibrating frequency f_ν. Those micro-Doppler variations reflected from a swimmer can be used for detection and classification of different swimming styles.

3. UNDERWATER EXPERIMENTATION OF HUMAN ACTIVITY DETECTION

In this section, we analyze the possibility of detecting human activity namely identifying three different swimming styles from underwater using a sonar system.

The different swimming styles are captured as acoustic micro-Doppler signal using two Teledyne RESON TC4013 hydrophones of 170 kHz bandwidth, which are used as acoustic transmitter (Tx) and receiver (Rx). The two hydrophones, separated by 20 cm from each other, are deployed 30 cm below the surface at the deeper end of the swimming pool. An open-source software development toolkit GNU Radio [16] is used to generate and transmit a sinusoidal signal at 40 kHz through the Tx hydrophone. The waveform generated by the GNU radio transmitter, shown in Fig. 1, is transmitted through the USRP N210 Tx using a special peripheral LFTX daughterboard operating at the frequency range $0 - 30$ MHz. A pre-amplifier (HP 467A Power Amplifier) is used to first amplify the signal generated by the USRP N210 Tx before transmitting the acoustic signal.

At the receiver side, the Rx hydrophone captures the transmitted 40 kHz waveform, which is first fed into a post-amplifier (Teledyne RESON VP2000 Voltage Preamplifier [17]) then the received signal is collected by the USRP N210 Rx using the LFRX daughterboard and it is controlled by the GNU Radio Rx software. The GNU Radio

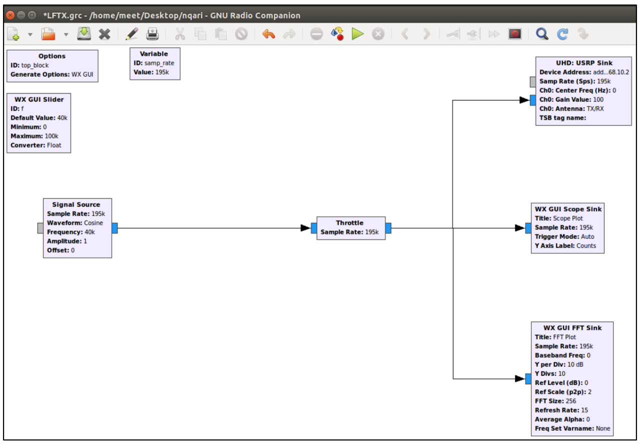


Figure 1: GNU radio transmitter.

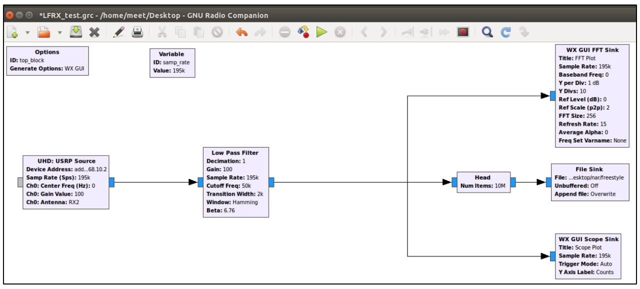


Figure 2: GNU radio receiver.

receiver diagram is shown in Fig. 2.

The experimental flowchart demonstrating the different procedures carried out in the transceiver side is shown in Fig. 3.

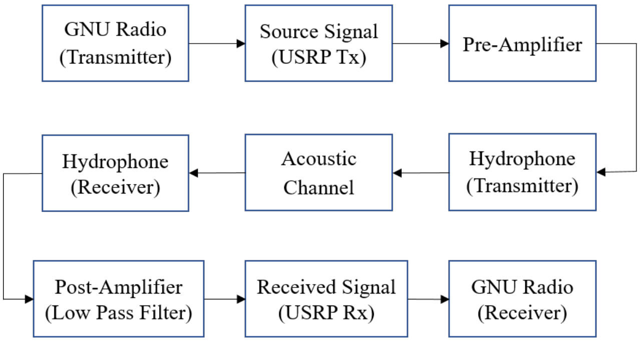
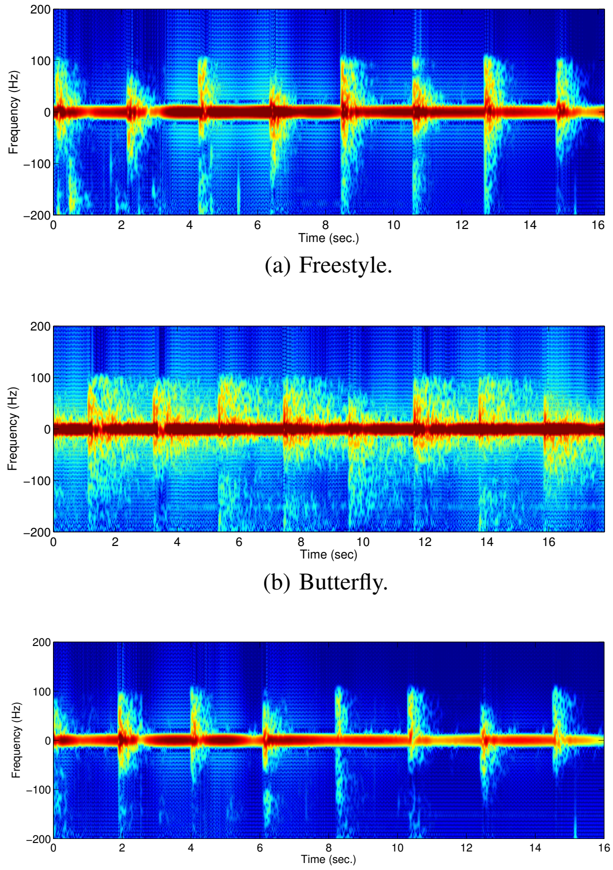


Figure 3: Experiment setup flowchart.

The collected received raw data is analyzed in Matlab. Fast Fourier transform (FFT) size of 256 with a step size of 10 was used in the experiment. After mixing the received data with the transmitted signal and passing it through a low pass filter (LPF) to remove the carrier frequency harmonics, we apply time-frequency analysis. Figure 4 shows three samples of different spectrograms of swimming styles for a) Freestyle, b) Butterfly and c) Backstroke.

The unique characteristics can be seen in Fig. 4. As an example, the spectrogram shown in Fig. 4(a) contains 8 samples of freestyle swimming patterns. The micro-Doppler variations are observed to be in the range from $0 - 100$ Hz. Each of these micro-Doppler patterns in the spectrogram plot are manually cropped into images of size 300×140 pixels for training and classification purposes in the deep learning algorithm, which is discussed next.



(a) Freestyle.

(b) Butterfly.

(c) Backstroke.

Figure 4: Swimming style spectrogram snapshots:

4. ACOUSTIC MICRO-DOPPLER CLASSIFICATION WITH DEEP LEARNING

A number of advanced deep learning algorithms have been developed for image classification [4, 3, 15, 18, 19]. A generalized CNN architecture is shown in Fig. 5, where it extracts features from the training images and then generates classifiers. The classifier weights are determined via the training process. The produced output y, shown in Fig. 5, is compared with the input data d and the error information e is fed back to the algorithm to update the weights and improve the classification process. In general, 80% of data is used for training purposes, and the remaining 20% for validating the CNN algorithm [20].

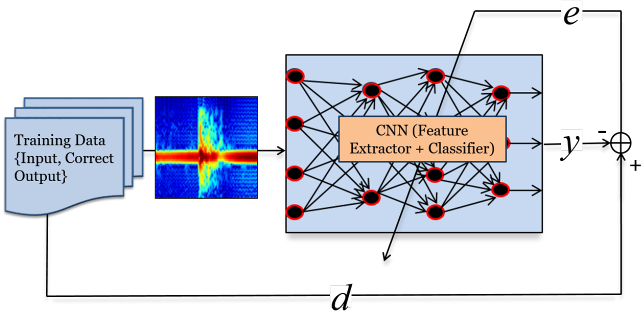


Figure 5: Generalized network of CNN.

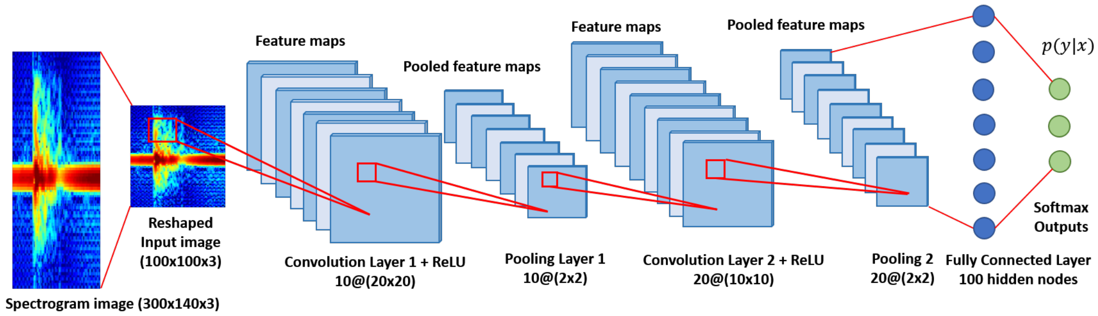


Figure 6: Architecture of the DCNN algorithm implemented in Matlab.

To classify the different swimming styles, we implement a DCNN algorithm in Matlab as follows. The captured spectrogram images are first re-sized from 300×140 pixel to 100×100 pixel RGB images with three classes, $1 - 3$. The input images undergo feature extraction network by first being processed by the convolution layer consisting of 10 convolution filters of size 20×20. The output from the convolution layer goes through the rectified linear unit (ReLU) function followed by the pooling layer, which employs max pooling process of 2×2 matrices. This process is repeated several times to create the output and train the machine with inherent features of the image. The output of the pooling layer is fed into a second convolution layer consisting of 20 convolution filters of size 10×10. Similarly, after passing the output through the ReLU function it undergoes the pooling layer with max pooling size of 2×2 matrices.

The max pooling concept is demonstrated in Fig. 7. Stride is the sliding window operation, used in the convolution layer and in the max pooling operation in which case the stride is 2. Max pooling is a downsampling process in which it selects the maximum value from each view. Since the spectrogram images contain sharp edges max pooling instead of average pooling is used to extracts the most important features like edges. The classifier network consists of a fully connected layer comprised of 100 hidden nodes, which produce a Softmax output that is used for classifying the three different swimming styles. The architecture of the DCNN algorithm implemented in Matlab is shown in Fig. 6.

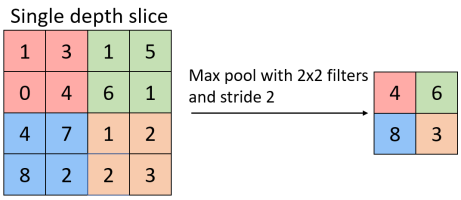


Figure 7: Max pooling principle.

5. EXPERIMENTAL RESULTS

In this section, we evaluate the performance of the proposed swimming styles classification scheme. The experimental setup is shown in Fig. 8.

Three different swimming styles are captured by the sonar system presented in Section 3. A swimmer of height 168 cm and weight 77 kg swims multiple times different swimming styles from the shallow water side of the pool towards the sonar system submerged in the deeper side of the pool. The distance from the starting position of the swimmer to the sonar system is about 20 m. The surface area of the pool was about 550 m^2. For each swimming style, we collect 100 samples of swimming style strokes, i.e., a total of 300 samples of spectrogram images. Out of these 300 images 240 are used for training and the remaining 60 for validation purposes.

The proposed DCNN algorithm discussed in Section 4 is implemented in Matlab to classify the three swimming styles. We used Dell Latitude E547 laptop with an 8th Generation Intel Core i7 processor for running the deep learning algorithm. The training algorithm is experimented with batch processing and stochastic gradient descent (SGD) algorithm with momentum. In the batch processing, each weight update is calculated for all errors of the training data, and the average of the weight updates is used for adjusting the weights. This method uses all of the training data and updates only once. The SGD algorithm on the other hand, calculates the error for each training data and immediately adjusts the weights. As an example, if we have 1000 training data points, the SGD adjusts the weights 1000 times.

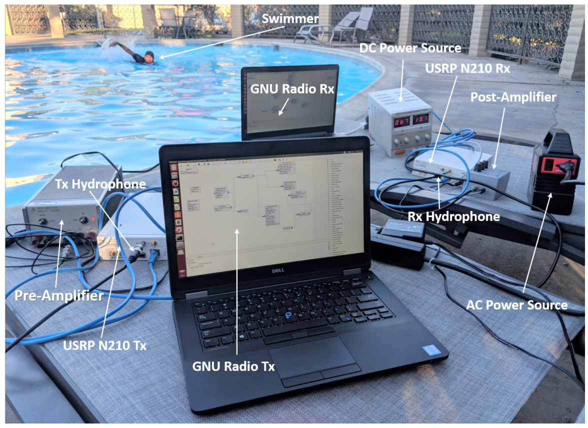


Figure 8: Experiment setup.

Figure 9 shows the average training error versus epoch. One epoch is when the entire training data set is passed forward and backward through the neural network once. Since the training data set is often limited, in practice, multiple epochs are utilized to allow the learning algorithm to run until the error from the model is sufficiently minimized. The SGD algorithm with momentum is observed to outperform the batch processing, as shown in Fig. 9. To reach an average training error of 0.08 batch processing requires 48 epochs while the SGD algorithm needs only 11 epochs. SGD algorithm with momentum helps to accelerate gradients vectors in the right directions and dampens oscillations. Thus, as we can see from Fig. 9, it leads to a faster convergence.

In the DCNN algorithm, the convolution layers are defined with the batch normalization and the ReLU layers. The batch normalization layer helps to normalize the input layer by adjusting and scaling the activations, which can speed up the learning process. The ReLU layer captures interactions and non-linearities and can greatly accelerate the convergence of the SGD algorithm. Moreover, the convolution layers of the DCNN algorithm can be modified according to the needs of experimentation. The proposed DCNN algorithm is used to train those swimming styles depicted by the spectrograms. A data set of 300 spectrogram images are used to classify the 3 different swimming styles; freestyle, butterfly and backstroke.

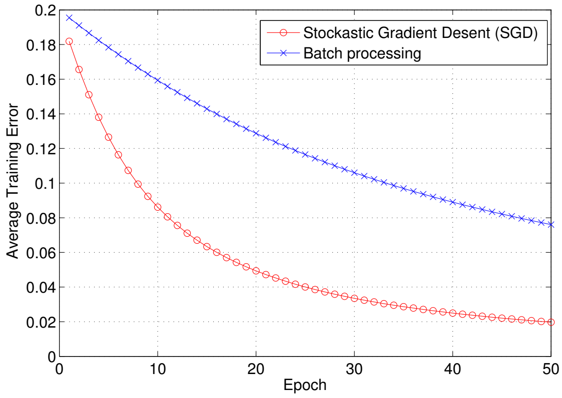


Figure 9: Average training error versus epoch.

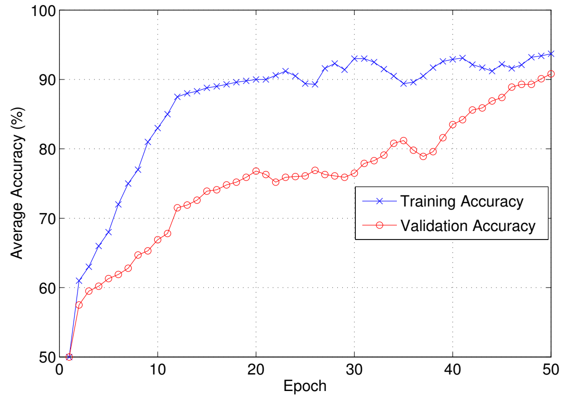


Figure 10: Training and validation accuracy versus epoch.

In Fig. 10, we plot the training and validation accuracy of the DCNN algorithm versus epoch. The batch size used for the training purposes is selected to be 10, which results in $300/10 = 30$ iterations per epoch. In other words, the weights of the neural network are updated 30 times after each epoch. A total of 50 epochs are used for data training, which results in a total of 1500 iterations. From the experimental results, we can see that an average training accuracy of 93.7% was achieved, and the average validation accuracy was 90.8%. To overcome the overfitting problem the validation accuracy and training accuracy graphs should be close to each other. As we can see in Fig. 10, at the final simulation round the average training and validation accuracy are fairly close to each other.

6. CONCLUSION AND FUTURE WORK

In this paper, we experimented classification of three swimming styles using acoustic sonar. Experiments were conducted in underwater to capture the micro-Doppler signatures present in the swimming styles using hydrophones and software defined radios. After collecting the raw data, time-frequency analysis is applied to extract the spectrogram images corresponding to the three different swimming styles. Deep convolution neural network (DCNN) algorithm is implemented in Matlab to classify the three different swimming styles using the micro-Doppler spectrogram images as input to the algorithm. Experimental results reveal that an average training accuracy of 93.7% is achieved, and the average validation accuracy was 90.8%. While we have focused on collecting data samples from a single swimmer, in the future, we will gather data from multiple swimmers. We will also experiment when the swimmer is swimming away from sonar not just towards the sonar. This work can also be extended for detection and classification of a drowning person, fish, marine mammals, submarines and AUVs.

7. REFERENCES

[1] J. A. Nanzer, *Microwave and millimeter-wave remote sensing for security applications*. Artech House, 2012.

[2] T. Melodia, H. Kulhandjian, L. Kuo, and E. Demirors, "Advances in Underwater Acoustic Networking," in *Mobile Ad Hoc Networking: Cutting Edge Directions*, 2nd ed., S. Basagni, M. Conti, S. Giordano, and I. Stojmenovic, Eds. Inc., Hoboken, NJ: John Wiley and Sons, 2013, pp. 804–852.

[3] X. Li, Y. He, and X. Jing, "A Survey of Deep Learning-Based Human Activity Recognition in Radar," *Remote Sensing*, vol. 11, no. 9, p. 1068, May 2019.

[4] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, "Deep learning for sensor-based activity recognition: A survey," *Pattern Recognition Letters*, vol. 119, pp. 3–11, Mar. 2019.

[5] V. C. Chen and H. Ling, *Time-frequency transforms for radar imaging and signal analysis*. Artech house, 2002.

[6] A. Balleri, K. Chetty, and K. Woodbridge, "Classification of personnel targets by acoustic micro-doppler signatures," *IET radar, sonar & navig.*, vol. 5, no. 9, pp. 943–951, Dec. 2011.

[7] B. Victor, Z. He, S. Morgan, and D. Miniutti, "Continuous video to simple signals for swimming stroke detection with convolutional neural networks," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Honolulu, Hawaii, USA, Jul. 2017, pp. 66–75.

[8] D. P. Williams, "Underwater target classification in synthetic aperture sonar imagery using deep convolutional neural networks," in *Proc. of the IEEE international conference on pattern recognition (ICPR)*, Cancún, Mexico, Dec. 2016, pp. 2497–2502.

[9] M. Einfalt, D. Zecha, and R. Lienhart, "Activity-conditioned continuous human pose estimation for performance analysis of athletes using the example of swimming," in *Proc. of the IEEE Winter Conference on Applications of Computer Vision (WACV)*, Lake Tahoe, NV, USA, Mar. 2018, pp. 446–455.

[10] R. Kashyap, I. Singh, and S. S. Ram, "Micro-doppler signatures of underwater vehicles using acoustic radar," in *Proc. of the IEEE Radar Conference (RadarCon)*, Arlington, VA, USA, May 2015, pp. 1222–1227.

[11] P. Zhu, J. Isaacs, B. Fu, and S. Ferrari, "Deep learning feature extraction for target recognition and classification in underwater sonar images," in *Proc. of the IEEE Annual Conference on Decision and Control (CDC)*, Arlington, VA, USA, May 2017, pp. 2724–2731.

[12] J. Park, R. Javier, T. Moon, and Y. Kim, "Micro-doppler based classification of human aquatic activities via transfer learning of convolutional neural networks," *Sensors*, vol. 16, no. 12, p. 1990, Nov. 2016.

[13] "Teledyne RESON hydrophones," http://www.teledynemarine.com/reson-tc4013, 2019, [Online; accessed 16-July-2019].

[14] "USRP N210 Software Defined Radio (SDR) - Ettus Research," https://www.ettus.com/all-products/un210-kit/, 2019, [Online; accessed 16-July-2019].

[15] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, http://www.deeplearningbook.org.

[16] "GNU Radio," https://www.gnuradio.org/, 2019, [Online; accessed 16-July-2019].

[17] "Teledyne RESON VP2000 Voltage Preamplifier," http://www.teledynemarine.com/reson-ec6081, 2019, [Online; accessed 16-July-2019].

[18] V. Dumoulin and F. Visin, "A guide to convolution arithmetic for deep learning," *arXiv:1603.07285 [stat.ML]*, Mar. 2016. [Online]. Available: https://arxiv.org/abs/1603.07285

[19] H. Kulhandjian, P. Sharma, M. Kulhandjian, and C. D'Amours, "Sign language gesture recognition using doppler radar and deep learning," in *Proc. of the IEEE Global Communications Conference (GLOBECOM) Workshop*, Waikoloa, Hawaii, USA, Dec. 2019, pp. 1–6.

[20] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, "Convolutional neural networks for medical image analysis: Full training or fine tuning?" *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1299–1312, May 2016.