Feature Level Sensor Fusion for Passive RF and EO Information Integration

Asad Vakil
Department of Electrical and Computer Engineering
Oakland University
Rochester, MI
avakil@oakland.edu

Jenny Liu
Department of Electrical and Computer Engineering
Oakland University
Rochester, MI
jennyliu443@gmail.com

Peter Zulch
Information Directorate
Air Force Research Laboratory
Rome, NY
peter.zulch@us.af.mil

Erik Blasch
Air Force Office of Scientific Research
Arlington, VA
erik.blasch.1@us.af.mil

Robert Ewing
Sensors Directorate
Air Force Research Laboratory
Wright-Patterson Air Force Base
Dayton, OH
robert.ewing.2@us.af.mil

Jia Li
Department of Electrical and Computer Engineering
Oakland University
Rochester, MI
li4@oakland.edu

***Abstract* — Many different sensing modalities across the spectrum exist for collecting and processing data for the purposes of target detection, tracking and differentiation. However, each of these individual modalities from the electromagnetic spectrum contain benefits, limitations, and sources of uncertainty. While research has been conducted to integrate complementary data collected by electro-optical (EO) and radio frequency (RF) modalities, the processing of RF data usually applies traditional methods, such as Doppler. This paper explores the viability of using histogram of I/Q (in-phase and quadrature) data for the purposes of augmenting the detection accuracy that EO input alone is incapable of achieving. Processing the histogram of I/Q data via deep learning, enhances feature resolution for neural network fusion. Using the simulated data from the Digital Imaging and Remote Sensing Image Generation (DIRSIG) dataset, the resulting fusion of EO/RF neural network (FERNN) can achieve 95% accuracy in vehicle detection and scenario categorization, which is a 23% improvement over the accuracy achieved by a standalone EO sensor.**

Keywords—Heterogeneous Sensor Fusion, Deep Learning, Feature Level Fusion, Histogram of In-phase and Quadrature Components

I. Introduction

In the information age, there exists many ways of sensing, such as pressure, radar, acoustic, chemical, electromagnetic, thermal, proximity, and optical. Each of these independent modalities have their own individual strengths and weaknesses. Whether it is active or passive in nature, or vulnerable to different forms of interference the ability to utilize the information efficiently is invaluable for reliable, credible, and robust system performance. To enhance robustness, sensor fusion seeks to reduce uncertainty from multiple sources.

The field of information fusion includes multi-sensor data fusion and is applicable to a wide range of applications. From security, automotive systems, healthcare, weather forecasting, internet of things, navigation, to communication, the ability to process information and to ensure a level of reliability is of paramount importance. In some cases, the system of sensors might simply be redundant in nature, such as having a backup smoke detector in a room, while other systems might rely on complementary sensors such as a network of security cameras.

Not all forms of sensor fusion are necessarily homogeneous in nature. For instance, take any form of a smartphone. Even for something as simple as the screen reorienting itself for landscape mode is the combination of the gyroscope and accelerometer sensors inside the smartphone device. For any form of an augmented reality application, that uses a smartphone, optical data from the camera, audio information from the microphone, and perhaps even a magnetometer might all be sensor inputs for that application in addition to the gyroscope and the accelerometer. What had originally started as a simple auditory communication device can now accomplish feats that would have been undreamt of in the 1850s.

Sensor fusion between heterogeneous modalities shows promise in many domains [1]. But when dealing with modalities that involve radio frequency (RF) and electro-optical (EO) sensors, the focus has traditionally been on active RF sensors. Doppler radar and imaging radar (e.g., side-looking airborne radar) are well suited for actively tracking a moving target when used with any form of EO modality. The combined exploitation of the two sensor modalities can still be improved however [2]. RF-based sensors are not limited by visual interference from natural phenomenon such as fog, clouds, snow, or any other form of weather. In addition to this, RF based sensors can provide repetitive coverage over a wide geographical area, and in doing so can determine the precise distance and velocity of a target.

Both RF and EO modalities have their own respective shortcomings. Visual interference, such as weather from clouds, can obscure EO imagery collections. RF signals have challenges of maneuvering materials that are conductors. Additionally,

978-1-7821-2734-7/20/$31.00 ©2020 IEEE

electronic countermeasures such as jamming and spoofing can corrupt RF-based modalities and transmitted imagery. The relatively higher resolution of information from EO modalities and the reliability of RF based sensors promote research to combine these two heterogeneous modalities.

For this research, the goal was not to rely on active RF inputs, but to exploit the potentials of passive RF (P-RF) sensors. The approach is a low energy application of radar, and considerably more economical in value, and boasts of rapid updates compared to active RF modalities. Jamming and other countermeasures for traditional active RF modalities are also difficult to implement, making the modality significantly more reliable in terms of performance. With the modern computing power 21st century technology possesses and the sophisticated algorithms available, extracting meaningful information from background noise becomes a viable approach to gathering information.

Information fusion with EO and passive RF modalities requires collecting and experimenting with such data. In 2019, Air Force Research Laboratory (AFRL) and Michigan Tech Research Institute (MTRI), released their Experiments, Scenarios, Concept of Operations, and Prototype Engineering data set (ESCAPE) [3]. The dataset is a versatile toolkit of different sensor modalities and scenarios that include, infrared (IR), full motion video (FMV), passive RF data, acoustic, seismic, and active radar imagery data.

In addition to the data collected at the AFRL Stockbridge test site, MTRI provided Digital Imaging and Remote Sensing Image Generation (DIRSIG) simulation of EO imagery and P-RF sensors. The DIRSIG dataset contains 13 simulations that cover a variety of visual obscuration scenarios, while receiving RF signals at three different locations. Much like the measurement data from the ESCAPE data, the DIRSIG simulations are oriented around tracking one or more moving targets, cars, and provide various opportunities to test the uncertainty reduction of the chosen sensor fusion method.

For the multimodal heterogeneous sensor fusion research presented in this paper, the raw RF data are preprocessed to obtain I/Q histograms along the time. The histograms are aligned with the simulated EO data for the purposes of detecting and classifying the number of vehicles. When compared to the ground truth of the simulation, neither standalone modality was able to achieve accuracy above 90%. But with the fusion of EO and P-RF sensors, it could reach an accuracy of 95%.

This paper discusses the measures and architecture used to achieve sensor fusion using passive RF data and the EO frames generated by the MTRI DIRSIG simulation. Section II provides a background and Section III describes the technical approach. Section IV provides experimental results. Section V concludes the paper and discusses the future work.

II. Background

A. Methodology:

While there exist many methods and algorithms of sensor fusion, such as Dempster-Shafer theory, Bayesian networks, Kalman filters, and sum of Gaussians [4], the complex nature of raw RF data and the difficulties in establishing a correspondence between the EO and RF inputs indicates that a deep learning approach would be more suited for the research [5, 6, 7]. While methods such as Dempster-Shafer theory provide the opportunity to reason with uncertainty [8], the sheer size of the samples and RF data in raw format make finding a traditional correlation between EO and RF data very difficult. Given the many real world variables that can affect radio waves and cause attenuation, it would be a Sisyphean task to attempt to find a linear correlation over raw data.

Neural networks (NN) and deep learning boast of powerful classification ability, with some networks even capable of achieving accurate classification with simply end-to-end learning. Given that their attainments for applications in RF related functions, such as cognitive radio and radio signal processing [9], using a NN to classify passive RF data is a reasonable choice. Processing the input data, and then approximating a solution, is more feasible than developing a series of first-principle physics equations that approximate the problem.

B. Advantages of Fusion

Using both the EO and RF data to classify objects in the scenario has several advantages. The different phenomenon of the two modalities implies that the scope of the data they collect can better complement each other. While both EO and RF sensors are susceptible to interference, the cause of interference for one modality may not interfere with the reliability of the other. The EO data is limited by the scope and range of the data it can collect, but in turn is easier to exploit than the RF. For example, in the ESCAPE data, there are several situations in which the target(s) are visually obscured by the foliage, and as such, will cause the data to be incorrectly interpreted as being devoid of targets [10, 11].

In the DIRSIG dataset, there are similar scenarios where fusion of EO and P-RF can provide better accuracy over single modality. The EO sensor in the simulation is limited by range in detecting the number of vehicles. Even when the targets reach the scope of the EO sensor, there are several instances of optical obscuration that are caused by the simulated foliage. In order to overcome this difficulty, as any vehicle detection or tracking system would require, improving the accuracy via sensor fusion is necessary.

C. Fusion Architecture

There are many different methods of sensor fusion which include signals, feature, score, and decision-level fusion [12]. For the individual modalities, the sensors can be redundant, complementary, coordinated, and competitive. Sensor fusion between redundant sensor nodes entails using one of the nodes to calibrate the results of its peer node for spatial and temporal resolution. Complementary sensors work with the other sensors to make the best use out of diversity in position and the nature of the modality frequency. Sensors that are coordinated sequentially collect data as defined by the system over time. Competitive sensor nodes take independent measurements each other, unlike redundant sensors. The communication architectures for sensor fusion include decentralized, distributed, centralized, or hierarchical. Decentralized

architecture implies there is no communication between the sensor nodes, while centralized architecture provides measurements to a common unit. Distributed nodes interchange data at a given communication rate. Hierarchical methods are a hybrid of the other techniques [13].

The fusion of EO/RF neural network (FERNN) proposed in this paper seeks to accurately estimate the number of moving targets in the scene through EO/RF fusion. Four states were generated in order to make the best description of the scenario. The state is denoted by a one-hot vector $S = [s_1, s_2, s_3, s_4]$, whose elements are all 0's except a single 1 for the element corresponding to a specific state. The first state, i.e. $S = [1, 0, 0, 0]$, is defined as having no cars in the simulation, which is only used for the purposes of determining the accuracy of the EO architecture as the ground truth is that there is always a vehicle in all the simulations of DIRSIG dataset. The other three states are a single car, two cars, and then finally three or more cars in the simulation.

For data from EO sensors, each generated video was first processed into image frames, resized, and converted into grayscale, then available for NN training.

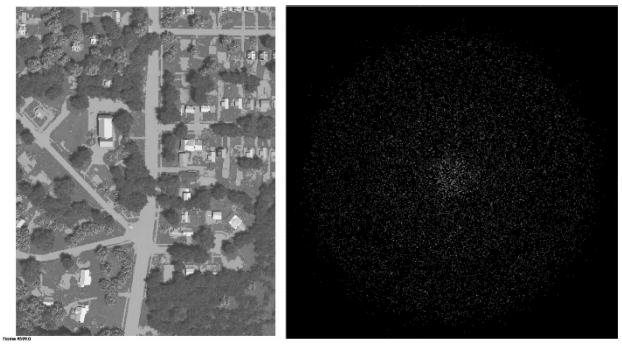

Figure 1: The extracted EO frame and the corresponding P-RF histogram (Simulation 9, NADIR RF reciever).

Prior to being fed into FERNN, the raw P-RF data is processed in order to generate histograms from the I/Q data. The histograms are then fed into the neural network in conjunction with the corresponding EO frames, an example of which can be seen above in Figure 1. The histogram depicts the estimation of the probability distribution of the P-RF data for each of the individual states. The I and Q values are processed into bins to form the histogram. Figure 1 shows an extracted EO frame and the corresponding histogram of P-RF sensor that have been aligned in time.

With the two inputs, one as EO sensor output and the other as the P-RF histogram, the two samples are resized and fed into a deep neural network (DNN). In order to achieve heterogeneous feature-level fusion, the DNN is trained over the 2D matrices from the two modalities. The fusion neural network itself is a simple sequential model that compares the predicted states with the labels and then modifies its weights accordingly.

III. Technical Approach and Experiment Design

For the purposes of the experiment, six simulation scenarios were selected from the DIRSIG data based on the number of vehicles and the known waveforms being used in the simulation. From the six chosen scenarios, simulations 2 and 10 were used for the training and testing of the FERNN's accuracy in detecting one vehicle. For the scenario of two vehicles, simulations 4 and 9 were selected for testing and training, and for the scenario of three or more vehicles, simulations 11 and 12 were chosen.

Simulation 2 contains a single vehicle which transmits a tone, while simulation 10 has a single vehicle which transmits a 2G waveform. Simulations 4 and 9 share a common 2G waveform while simulations 11 and 12 share the 2G, 3G and 4G waveforms. The purpose of choosing these scenarios with different waveforms was to ensure that a level of robust change was added into the training in order to prevent the neural network from ignoring the EO input.

Table 1: Simulation Classifications

Simulation:	Number of Vehicles:	TX Waveform:
1	1	Tone
2	1	Tone
3	1	2G
4	2	2G, None
5	2	2G, 3G
6	3	2G,3G, None
7	3	2G, 3G, 4G
8	3	2G, 3G, 4G
9	2	2G, 3G
10	1	2G
11	3	2G, 3G, 4G
12	3	2G, 3G, 4G
13	3	2G, 3G, 4G

A. EO Neural Network

A separate convolutional neural network is trained for detection and estimation of the number of vehicles based on EO data. This network can achieve 91% accuracy upon using the modified labels that are exclusively just for image classification. For these modified labels, if a frame takes place during a period in which the vehicle is temporarily obscured by foliage, then the frame is labeled as having no vehicle detected, which is inaccurate when compared to the ground truth of the simulation. Likewise, simulation frames in which the vehicle has not entered the scope of the EO sensor are also labeled as not detecting a vehicle. In order to accomplish classification, both FERNN and the standalone EO network begins preprocessing, resizing and labeling the frames in order to read the grayscale values, after which the network then categorizes the image by the number of vehicles detected. Unlike the ground truth for the overall scenario simulation, the standalone EO neural network will

output that there are no vehicles detectable in the above two scenarios. For the purposes of comparison testing, the standalone EO neural network retains the original training it received to classify an image by the number of vehicles it detects, but upon being tested against the ground truth for each simulation, in terms the number of vehicle(s) are traveling in the area, the accuracy of the EO network decreased to an accuracy of only 72%. This result is expected, as the output of no vehicle being detected is erroneous when compared to the ground truth.

B. RF Feature Extraction and neural network

The most basic signal that can be collected for RF sensing are known as in-phase and quadrature components. These I/Q components are the basis of complex RF signal modulation and demodulations, used in hardware, software, and complex signal analysis, and the backbone of modern communication architectures. During previous experiments, our group had been able to successfully train a convolutional neural network to determine the occupancy of an enclosed indoor space based on the I/Q data of passive RF signals. For the DIRSIG dataset, however, the I/Q data in its raw format had proved to be ineffective for the purposes of vehicle detection.

To that end, the P-RF data was preprocessed in order to obtain new features for the neural network to read. The I/Q data was processed into a histogram to estimate the probability density function of the data. In the DIRSIG simulation, there are three SIGINT sensors to generate P-RF data. These sensors are placed orthogonally, north and west in the simulation, with one being placed in the nadir. The application of the nonparametric feature extraction in turn plots the data and the generated 2D matrices are then fed into the fusion networks in order to facilitate homogenous fusion between the three RF histogram sources.

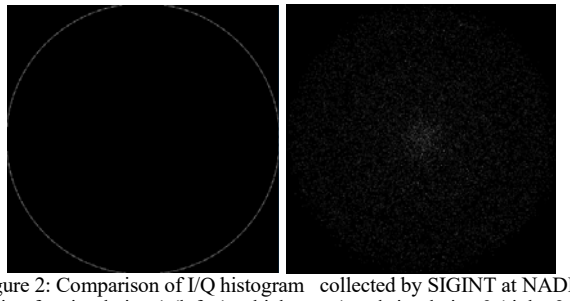

Figure 2: Comparison of I/Q histogram collected by SIGINT at NADIR location for simulation 1 (left, 1 vehicle, tone) and simulation 9 (right. 2 vehicles, 2G and 3G).

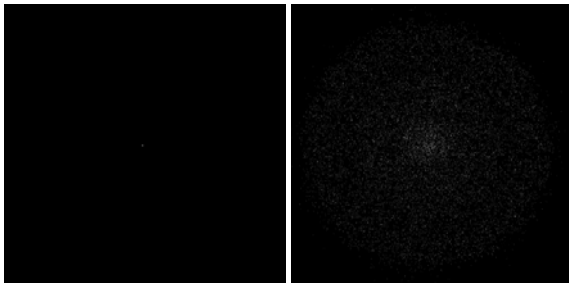

Figure 3: Comparison of I/Q histogram collected by SIGINT from simulation 4 (left, 2 vehicles, 2G) and simulation 11 (right, 3 vehicles, 2G, 3G, 4G)

For the DIRSIG simulation dataset, the selected simulations had different transmission waveforms that were collected by the three separate SIGINT sensors. Both simulations 1 and 2 had a transmission waveform of tone. Simulations 4, 9, 10, 11, and 12 all had a 2G waveform in the background, but only simulation 9, 11, and 12 had a 3G waveform, while only simulations 11 and 12 had 4G waveforms in the background.

As seen in Figures 2 and 3, the histograms generated by the preprocessing show visually different patterns based on the transmission waveforms being received by the simulation receiver. The passive RF signal produces different histograms, as simulation 4 (Figure 3, left) and simulation 9 (Figure 2, right) are noticeably different despite both waveforms being produced from simulation data with the same number of vehicles. Each of the simulation pairs were selected by the noticeable visual differences, as seen when comparing simulations 1 and 9 or 4 and 11, both show significant changes despite both being simulations that have the same number of vehicles in them. This was in order to prevent the heterogenous fusion network from being completely dependent on the RF histogram input, the selection of simulations was predominantly dictated by the how visually different the respective histograms of the processed RF SIGINT.

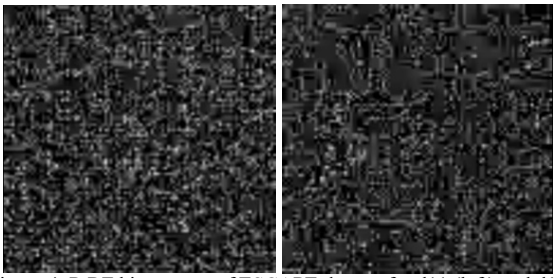

Figure 4: P-RF histograms of ESCAPE dataset for d11 (left) and d12 (right) at 0'01" in scenario s01r12.

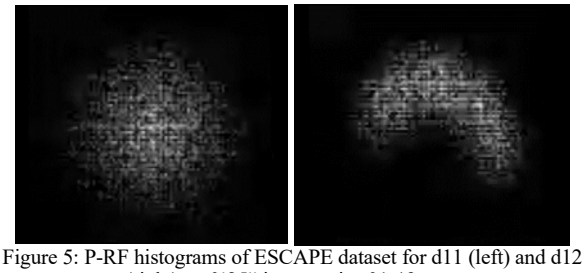

Figure 5: P-RF histograms of ESCAPE dataset for d11 (left) and d12 (right) at 0'25" in scenario s01r12.

For the ESCAPE dataset, the P-RF histograms were generated for feature extraction as well. Figure 4 shows typical histograms of P-RF sensor d11 and d12 in scenario s01r12 when a single vehicle is moving in the scene. Visual inspection does not detect any noticeable patterns in these two histograms. However, when the vehicle enters and stays inside the garage during the time period from 0'21" to 0'28", the histograms of d11 and d12 display strong patterns as shown in Figure 5. The pattern in d12 is different from that of d11 due to the location change. Although the cause of these patterns is unknown now, the correlation of the two histograms has discernible change in this time period as illustrated by Figure 6. This is an evidence

that I/Q histogram of P-RF sensors is an effective feature in vehicle detection. Please note the periodic peaks in Figure 6 is caused by a waveform that was transmitted every 4.8 seconds.

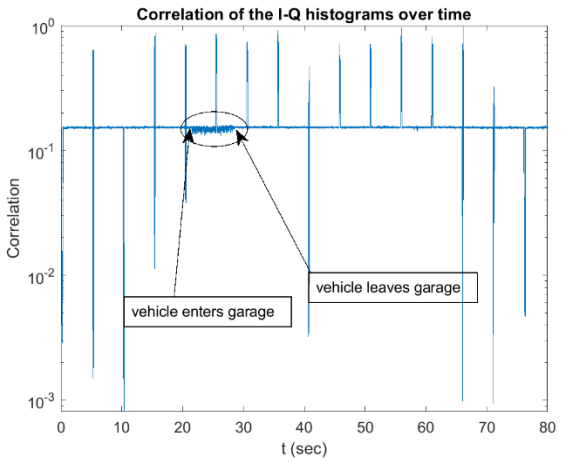

Figure 6: Correlation of P-RF histograms of d11 and d12 over time.

For the standalone RF neural network, the classification accuracy would rise to 100% accuracy within four epochs for differentiation the simulations in DIRSIG dataset. This, however, relies on the unique histograms formed by the different transmission waveforms. In order to balance training and testing data for the standalone RF and fusion networks, the simulations used for testing were limited to the ones that are as different of a transmission waveform as possible. Simulations 1 and 2 for example, are unique compared to other simulations in which only one car is detected because the waveform is a pure tone. Simulations 11, 12, and 13 all have a variety of signals, 2G, 3G, and 4G, while some simulations such as 10 or 4 are limited to a single waveform (2G). When trained under these constraints, the accuracy of the standalone P-RF neural network is 83%.

C. *Feature Level Fusion*

In order to achieve sensor fusion from heterogenous sensor modalities, both the resized and preprocessed grayscale EO frames and P-RF histograms are fed into a sequential FERNN. The data from both the RF and EO modalities is stacked into a sequence of arrays that acts as the training data for the neural network. After being standardized, normalized, and trained, the sequential model begins feature-level fusion.

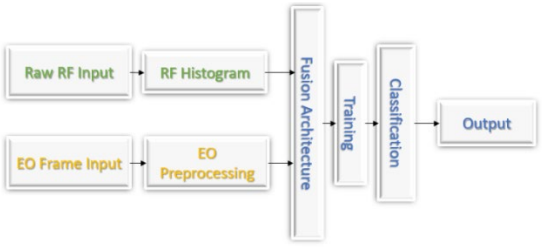

Figure 7: Fusion of EO and RF NN (FERNN) Architecture

FERNN takes the input arrays containing the values read off the preprocessed RF and EO data and then flattens them, using ReLu and Softmax before compiling the model with an Adam Optimizer. Unlike classical stochastic gradient descent, which maintains a single learning rate for all weight updates, Adam Optimizer utilizes individual adaptive learning rates for different parameters from estimates of the first and second moments of the gradients.

This approach combines the advantages of two other existing extensions of the stochastic gradient descent, Adaptive Gradient Algorithm and Root Mean Square Propagation. The chosen method for calculating loss for the model is Sparse Categorical Cross Entropy. Softmax was chosen in order to implement classification of different states, similar to the decision to implement Sparse Categorical Cross Entropy for multiclass classification. Compared to the results of the standalone RF and EO neural networks, the feature level fusion network can achieve 95% accuracy, with regards to the ground truth for the simulation.

D. *Decision Level Fusion*

Due to the RF features being extracted in a more linear format, more traditional methods of learning methods and probabilistic classifiers were implemented to compare FERNN with methods of decision level fusion. Logistic Regression (LR), Naïve Bayes (NB), Random Forest (RF), Gaussian Naïve Bayes(GNB), and Support-Vector Machine (SVM) were the primary focus for the comparative decision level fusion experiments. In addition to these methods, a modified version of FERNN was made to implement late fusion as well.

Naïve Bayes is a probabilistic classifier that applies Bayes' Theorem under the assumption that the data is independent of each other. Gaussian Naïve Bayes instead works under the assumption that the continuous values for each class are distributed to a Gaussian distribution. Random Forest is an ensemble learning method for classification that focuses on the generation of decision trees. These decision trees are used to avoid overfitting and use the mode of these classes to generate a prediction. Logistic Regression is a statistical model that is used to implement regression analysis to model the probability of a certain class or event. SVM is a supervised learning model that analyzes data for regression analysis and classification, however unlike Logistic Regression, Naïve Bayes, and Gaussian Naïve Bayes it is a non-probabilistic linear classifier.

In order to better use the available data, the approach for decision level fusion was ensemble learning methods, soft and hard voting, in addition to a neural network approach and late fusion via SVM. Soft voting uses the individual classifiers calculations for the probability of the outcomes, and averages out the resulting outputs. Hard voting uses majority vote in order to choose a model from the ensemble to make the final prediction with the available data. Besides soft and hard voting, the data from the feature level fusion experiments, the standalone EO and standalone RF neural networks, were used to implement late fusion. The predictions for each of the individual models were fed into an SVM model that used the concatenated values. Besides SVM, a neural network model used the same prediction values to classify the dataset.

IV. Results

The trained standalone EO neural network alone could reach 91% accuracy in determining the current number of detected vehicles based on the available image provided. However, when tested against the ground truth, with the knowledge of a vehicle moving outside of the camera's angle or obscured by local foliage, the overall accuracy of the neural network would decrease to only 72%. Similarly, the standalone P-RF neural network, could only reach an accuracy of 83% when taking into account all the histograms for the selected simulations, where scenario 1 describes a ground truth of one vehicle, scenario 2 describes a ground truth of two vehicles and scenario 3 describes a ground truth of three vehicles.

As seen in Table 2, the accuracy of the standalone P-RF neural network was significantly higher when trained with a small number of simulations. When comparing simulations 9 and 10, to differentiate between one vehicle and two vehicles, the neural network could perform at 95% accuracy. However, when the number of categories increases to 3 and the neural network was trained with 3 different simulations, the performance decreased to 92% accuracy. In this situation, histograms formed by the P-RF feature extraction are still unique enough to ensure an acceptable accuracy. In order to ensure the training data was robust enough for the purposes of vehicle detection, six simulations were combined in the training process, each of which containing visually different histograms. The addition of all these training and testing data reduced the accuracy of the neural network to 83%.

TABLE 2: RF ACCURACY COMPARISON

Situation	Accuracy
Comparison of simulations 9 and 10, differentiating between detecting one vehicle and two vehicles	95%
Comparison of simulations 2, 9, and 10, differentiating between detecting one vehicle, two vehicles, and three vehicles	92%
Combined data from simulations 2,4,9,10,12, and 13, differentiating between detecting one, two vehicles, and three vehicles	83%

As seen below in Table 3, the overall accuracy of different sensors on their own is unsatisfactory, with EO only managing to score a 72% accuracy against the ground truth and RF only reaching 83% accuracy when trained and tested with large number of different scenarios. With FERNN, the accuracy can reach a much higher level of 95%, when the EO frames, the data of P-RF sensors located at nadir, north, and west are all fed into the neural network.

TABLE 3: CLASSIFICATION ACCURACY COMPARISON

Situation	Accuracy
Standalone EO	72%
Standalone RF	83%
Feature Level Fusion Architecture	95%

TABLE 4: EO AND NADIR SIGINT FUSION

Scenario	Precision	Recall	F1-Score
1	0.70	0.72	0.71
2	0.87	0.79	0.83
3	0.83	0.88	0.85
Accuracy			0.80
Macro AVG	0.80	0.80	0.80
Weighted AVG	0.80	0.80	0.80

TABLE 5: EO AND NORTH SIGINT FUSION

Scenario	Precision	Recall	F1-Score
1	0.87	0.65	0.74
2	0.67	0.92	0.78
3	0.82	0.86	0.84
Accuracy			0.78
Macro AVG	0.79	0.81	0.79
Weighted AVG	0.80	0.78	0.78

TABLE 6: EO AND WEST SIGINT FUSION

Scenario	Precision	Recall	F1-Score
1	0.79	0.75	0.77
2	0.92	0.82	0.86
3	0.75	0.92	0.83
Accuracy			0.82
Macro AVG	0.82	0.83	0.82
Weighted AVG	0.83	0.82	0.82

TABLE 7: EO AND RF SIGINT FUSION

Scenario	Precision	Recall	F1-Score
1	0.96	0.96	0.96
2	0.99	0.94	0.96
3	0.91	0.96	0.93
Accuracy			0.95
Macro AVG	0.95	0.95	0.95
Weighted AVG	0.95	0.95	0.95

Based on the results of the feature level fusion, a significant increase in accuracy was dependent on the number of feature sources for training available. As seen in Table 4, the F1 score for using only the EO and Nadir SIGINT is 80%. Similarly, the F1 scores for the EO and North and EO and West are at 78% and 82% respectively. Based on the results presented in Tables 4, 5, and 6, the accuracy of the neural network only reaches satisfactory values when all four sources of SIGINT (Table 7) and the corresponding frames are fed into the training.

$$F_1 \text{ Score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (1)$$

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (2)$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad (3)$$

In order to accurately evaluate the performance of FERNN, the F1 score, precision and recall were calculated based on Eqn. (1-3) for statistical analysis. The F1 score is the harmonic mean of the precision and recall, the measurements of positive predictive value and sensitivity for machine learning. Precision is the measurement of type I error, false positives, while recall is the measurement of type II error, false negatives. Figure 8 shows the F1 scores of all the neural networks we have trained for vehicle detection and scenario categorization. The fusion of EO and all 3 P-RF sensors yield the best result.

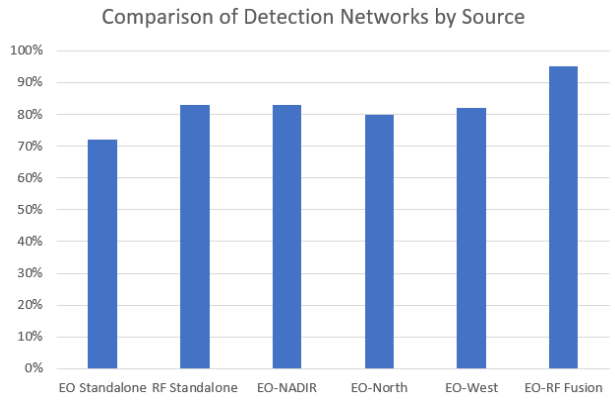

Figure 8: Comparison of F1 Scores for different neural networks implemented for vehicle detection.

Besides feature level fusion, more traditional methods were explored to analyze the effectiveness of using the RF features and EO input. Logistic Regression, Random Forest, Naïve Bayes, and Gaussian Naïve Bayes were all applied to the same datasets as the EO-RF fusion neural networks. As these ensemble methods are more traditional in nature and not based on neural networks, the evaluation of classification accuracy is conducted by k-fold Cross-Validation [needs reference here].

K-fold Cross-Validation is a procedure meant to estimate the skill of a machine learning model on unseen data. The limited samples are used to estimate how the model is expected to perform in general when used to make predictions on data that is not used during training. The process first shuffles the dataset randomly, splitting up the dataset into k groups. For each of these unique groups, one is designated as a test data set, while the others are treated as part of the training data set. The model is fit on the training data and then tested on that dataset, saving the evaluation score and discarding the model. After repeating the process, the skill of the model is summarized using the sample of the model evaluation scores.

TABLE 8: CLASSIFICATION ACCURACY OF TRADITIONAL PROBABILISTIC CLASSIFIERS AND LEARNING METHODS

Method and Input Data	Accuracy
Logistic Regression (EO and Nadir)	0.81(+/- 0.13)
Logistic Regression (EO, Nadir, and North)	0.75 (+/- 0.17)
Logistic Regression (EO, Nadir, North, and West)	0.73 (+/- 0.17)
Naïve Bayes (EO and Nadir)	0.64 (+/- 0.04)
Naïve Bayes (EO, Nadir, and North)	0.64 (+/- 0.08)
Naïve Bayes (EO, Nadir, North, and West)	0.63 (+/- 0.07)
Random Forest (EO, Nadir, North, and West)	0.64 (+/- 0.21)
Random Forest (EO, Nadir, and North)	0.69 (+/- 0.16)
Random Forest (EO, Nadir, North, and West)	0.67 (+/- 0.17)
Guassian Naïve Bayes (EO and Nadir)	0.73 (+/- 0.14)
Guassian Naïve Bayes (EO, Nadir, and North)	0.71 (+/- 0.15)
Guassian Naïve Bayes (EO, Nadir, North, and West)	0.70 (+/- 0.14)

As can be seen above in Table 8, Logistic Regression and Gaussian Naïve Bayes consistently performed better than Naïve Bayes and Random Forest in terms of accuracy. The accuracy for these classifiers improved when there were fewer inputs from the RF histograms. Considering the nature of the RF features and the EO inputs it is possible that because the data received as a 2-dimensional array that the classifiers that assume the data to be independent, Random Forest and Naïve Bayes, performed comparatively poorer than Logistic Regression and Gaussian Naïve Bayes.

TABLE 9: DECISION LEVEL FUSION COMPARISON

Method and Input Data	Accuracy
Hard Voting (LR, RF, NB, GNB) (EO and Nadir)	0.73 (+/- 0.14)
Hard Voting (LR, RF, NB, GNB) (EO, Nadir, and North)	0.71 (+/- 0.15)
Hard Voting (LR, RF, NB, GNB) (EO, Nadir, North, and West)	0.70 (+/- 0.12)
Soft Voting (LR, RF, NB, GNB) (EO and Nadir)	0.74 (+/- 0.13)
Soft Voting (LR, RF, NB, GNB) (EO, Nadir, and North)	0.72 (+/- 0.11)
Soft Voting (LR, RF, NB, GNB) (EO, Nadir, North, and West)	0.71 (+/- 0.13)

As seen above in Table 9, the results for hard and soft voting showed little difference in terms of total accuracy. Given the input methods being weighed against each other, and the accuracies they had individually, the result of the decision level fusion is dependent on the accuracies of the methods implemented. Soft voting performed marginally better than hard voting, as for the given inputs only one of the methods performed above 0.80 in accuracy.

TABLE 10: SVM FUSION FOR EO AND RF DATA

Scenario	Precision	Recall	F1-Score
1	0.76	0.84	0.81
2	0.93	0.96	0.94
3	1.00	0.73	0.85
Accuracy			0.88
Macro AVG	0.90	0.88	0.88
Weighted AVG	0.90	0.88	0.88

Besides implementing voting for decision level fusion, Support Vector Machine (SVM) learning was used in order to test the effectiveness of decision level fusion. For the SVM decision level fusion, the prediction values of the independently trained standalone EO and standalone RF neural networks were fed as a concatenated array of values. As seen above in Table 10, the late decision fusion implemented via SVM with the standalone EO and RF neural network classification weights could achieve an accuracy of 88%.

Out of the results of using feature and decision level fusion with the RF histograms and EO frames, the highest accuracy of all the methods tested was FERNN. Out of the decision level fusion methods, late fusion neural network (LFNN) was the closest in terms of accuracy, with an F1 score of 90.7%. From the results, it can be concluded that for this dataset of the RF and EO features, a neural network benefits more from feature level fusion over decision level fusion.

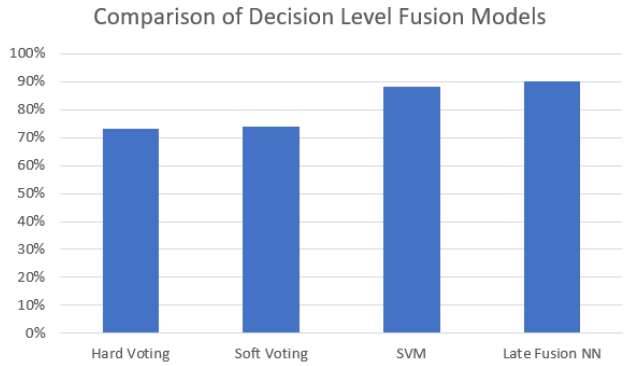

Figure 9: Comparison of accuracy for different neural networks implemented for decision level fusion.

Within the ensemble learning fusion, soft voting, which uses the individual outputted probabilities over simple majority voting, still performed as well if not better. Based on the approach and results, the association of the RF features with the corresponding EO frame produced more accurate results compared to the direct estimations of the output class in question. From the ensemble learning experiments and the standalone EO and RF network results, it can also be concluded that the association of changes in the RF histogram features was relied on closer than the EO, which was less accurate on its own compared to the ground truth.

When compared to the decision level fusion models that were SVM and late fusion neural network, the ensemble decision fusion significantly underperformed. The Late Fusion Neural Network (LFNN) was capable of achieving 90.7% accuracy with the same training set that the SVM decision level fusion model was able to achieve 88% accuracy with. In comparison, the soft and hard voting decision level fusion however, both failed to achieve even 80% accuracy. Compared to all the decision level fusion methods tested, FERNN, the proposed feature level fusion performed with the highest accuracy, achieving a 95% F1 score vs. the LFNN's 90.7% F1 score.

V. CONCLUSION

This paper proposed a feature level fusion network for integrating information from passive RF histograms and EO sensors and compared its performance to traditional methods of classifying linear input. From the results, it can be concluded that the application of P-RF histograms as a feature can significantly improve the accuracy of the neural network, particularly when fused at the feature level. The performance of the proposed EO and P-RF fusion network is superior to the performance of the neural networks of single sensor modality for both feature and decision level fusion, in regard to vehicle detection and scenario categorization. The next step in our research will be to train and test FERNN using the P-RF and EO modalities within the ESCAPE dataset.

ACKNOWLEDGMENTS

This research is supported by AFOSR grant FA9550-18-1-0287.

REFERENCES

[1] L. Snidaro, J. Garcia, J. Llinas, E. Blasch (eds.), *Context-Enhanced Information Fusion*: *Boosting Real-World Performance with Domain Knowledge*, Springer, 2016.

[2] D. L. Hall. J. Llinas, "An Introduction to multisensory data fusion," Proceedings of the IEEE 85(1): 1997, pp. 6-23.

[3] P. Zulch, M. Distasio, T. Cushman, B. Wilson, B. Hart and E. Blasch, "ESCAPE Data Collection for Multi-Modal Data Fusion Research," *2019 IEEE Aerospace Conference*, Big Sky, MT, USA, 2019, pp.1-10. doi: 10.1109/AERO.2019.8742124

[4] E. Blasch, S. Ravela, A. Aved (eds.), *Handbook of Dynamic Data Driven Applications Systems*, Springer, 2018.

[5] S. Riyaz, K. Sankhe, S. Ioannidis, K. Chowdhury, "Deep Learning Convolutional Neural Networks for Radio Identification," IEEE Communications Magazine, 56(9): 2018, pp. 146-152.

[6] T. Mukherjee, P. Kumar, D. Pati, et al., "LoSI: Large Scale Location Inference through FM Signal Integration and Estimation," *IEEE Big Data Mining and Analytics*, 2019.

[7] E. Blasch, R. Cruise, U. Majumder, T. Rovito, "Methods of AI for Multi-modal Sensing and Action for Complex Situations," *AI Magazine*, 2019.

[8] E. Blasch, J. Dezert, B Pannetier, "Overview of Dempster-Shafer and Belief Function Tracking Methods," *Proc. SPIE*, Vol. 8745, 2013.

[9] V. Miraftab and M. Yu, "Innovative Combline RF/Microwave Filter EM Synthesis and Design Using Neural Networks," *2007 International Symposium on Signals, Systems and Electronics*, Montreal, QC, 2007, pp. 1-4. doi: 10.1109/ISSSE.2007.4294399

[10] Xiang He, Daniel N. Aloi and Jia Li, "Probabilistic Multi-Sensor Fusion Based Indoor Positioning System on a Mobile Device," 2015 Sensors 15: pp. 31464–31481; doi:10.3390/s151229867

[11] D. Shen, E. Blasch, P. Zulch, M. Distasio, R. Niu, J. Lu, et al., "A Joint Manifold Leaning-Based Framework for Heterogeneous Upstream Data Fusion," *Journal of Algorithms and Computational Technology* (JACT), Vol. 12, Issue 4, 2018, pp. 311-332.

[12] Y. Zheng. E. Blasch, Z. Liu, *Multispectral Image Fusion and Colorization*, SPIE Press, 2018.

[13] M. Liggins II, D. Hall, J. Llinas, Handbook of Multisensor Data Fusion: Theory and Practice, CRC Press, 2017.

[14] E. Blasch, "NAECON08 Grand Challenge Entry Using the Belief Filter in Audio-Video Track and ID Fusion," *Proc. IEEE Nat. Aerospace Electronics Conf (NAECON)*, 2009.

[15] S. Chambon, M. N. Galtier, P. J. Arnal, G. Wainrib and A. Gramfort, "A Deep Learning Architecture for Temporal Sleep Stage Classification Using Multivariate and Multimodal Time Series," in *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 4, 2018, pp. 758-769. doi: 10.1109/TNSRE.2018.2813138I.

[16] W. Guo, J. Wang and S. Wang, "Deep Multimodal Representation Learning: A Survey," in *IEEE Access*, vol. 7, 2019, pp. 63373-63394, . doi: 10.1109/ACCESS.2019.2916887

[17] J. Shi, X. Zheng, Y. Li, Q. Zhang and S. Ying, "Multimodal Neuroimaging Feature Learning With Multimodal Stacked Deep Polynomial Networks for Diagnosis of Alzheimer's Disease," in *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 1, 2018, pp. 173-183. doi: 10.1109/JBHI.2017.2655720R.

[18] D. Wu *et al*., "Deep Dynamic Neural Networks for Multimodal Gesture Segmentation and Recognition," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 8, pp. 1583-1597, 1 Aug. 2016. doi: 10.1109/TPAMI.2016.2537340Y.

[19] Y. Zhu, J. Chen, C. Chen and H. Wang, "An algorithm of heterogeneous sensors track fusion for target tracking," *2010 Chinese Control and Decision Conference*, Xuzhou, 2010, pp. 1383-1387. doi: 10.1109/CCDC.2010.5498194M.

[20] L. Snidaro, J. Garcia, J. Llinas, E. Blasch (eds.), *Context-Enhanced Information Fusion*: *Boosting Real-World Performance with Domain Knowledge*, Springer, 2016.

[21] B. Duraisamy, M. Gabb, A. Vijayamohnan Nair, T. Schwarz and T. Yuan, "Track level fusion of extended objects from heterogeneous sensors," *2016 19th International Conference on Information Fusion (FUSION)*, Heidelberg, 2016, pp.876-885