MVI-DCGAN Insights into Heterogenous EO and Passive RF Fusion

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

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

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

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

2022 9th International Conference on Soft Computing & Machine Intelligence (ISCMI) | 979-8-3503-2088-6/22/$31.00 ©2022 IEEE | DOI: 10.1109/ISCMI56532.2022.10068480

***Abstract*—As technology trends towards automation, deep neural network (DNN) based methods become more and more desirable from a technological, economical, and societal standpoint. However, owing to the way that these black box technologies operate, it can be difficult to troubleshoot potential errors, especially when dealing with data that the human mind cannot intuitively understand. For this reason, the use of explainable artificial intelligence (XAI) is integral to obtaining interpretability and understanding of these systems' techniques. The paper explores some of the known uses of XAI in Generative Adversarial Networks (GANs); i.e., in processing electro-optical (EO) and passive radiofrequency (Passive RF) data to achieve heterogenous sensor fusion. GANs are capable of generating realistic images, music text, and other forms of data, and the use of deep convolutional generative adversarial networks (DCGANs) to process such information provides "richer" corrective feedback from which the model can train from. Using the DCGAN approach, tone can provide visualizations from different types of neural networks and use them as a training source for the multiple visualizations input (MVI) DCGAN. The MVI-DCGAN uses these visualizations in order to track the vehicle target and further differentiate between other overlay visualization data and the generated overlay input visualizations. The paper demonstrates multiple sources of visualization input from different neural networks for the training of the MVI-DCGAN for a more robust training and directing the discriminator towards focusing on the P-RF aspects of the visualizations.**

Keywords—explainable artificial intelligence, heterogenous sensor fusion, GAN, EO, passive RF

I. Introduction

In the information age, there exists a large amount of data that can be used for any number of applications. How this data can be used is a question that entails factors such as determining the most appropriate approach, assessing if a traditional algorithm would be sufficient for the application, comparing if a neural network (NN)-based approach would be better, and deciding if the data requires preprocessing in order to meet the objective in question. For some applications that are less complex, traditional machine learning (ML) algorithms such as support vector machines (SVM), decision trees (DT), or k-nearest neighbors (KNN) would be more desirable than any class of neural networks. But for more complicated applications, such as automated driving, or data that could be better understood by unsupervised learning methods, the use of NN-based approaches from deep learning (DL) might be worth the tradeoffs in terms of training data requirements. The use of DL is especially beneficial for heterogeneous sensor fusion, as different modalities might provide radically different information, but the upstream fusion (e.g., processing data at sensor collection) of both modalities means a greater reduction in uncertainty.

When using DL-based approaches one of the major issues becomes the matter of transparency (e.g., of source data inputs), interpretability (e.g., of how algorithm computes), and explainability (e.g., of output results). Owing to the nature of how Deep NN learn, these models inherently become a black box algorithm. The lack of interpretability becomes an issue for both users and field experts, as unlike conventional algorithms, it's difficult to understand the decision-making process of the black box approach. Decision trees use the CART (Classification And Regression Trees) algorithm, random forest algorithms use bagging and feature randomness when building each individual tree to try and create an uncorrelated forest by prediction, and KNNs use clustering [1]. Hence, it is possible for a human user to see how these traditional algorithms have reached their decisions. For example, a decision tree can be plotted for a random forest algorithm, showing what factors are relevant to the decision-making process. Clustering via KNN can be visualized to look for outliers and demonstrates how the local nonlinear data is used for regression and classification.

The stakes of mistakes a NN can make are not high on an academic level, but when used in situations where the consequences can be grave, such as automated driving, medical applications, or even financial applications; having at least some level of understanding and interpretability is important. If not for any safety reason, then for the purposes of design and training it is important to have interpretability. On the design end, it is important to choose appropriate training data, the composition of which determines how the model learns. If designed incorrectly, the model might be trained on features irrelevant to the model's desired application but still perform well on the dataset in question. Having interpretability becomes even more important when the model does make errors, as it otherwise becomes difficult to troubleshoot a black box system's incorrect decision.

The issue of the black box problem, however, will still exist if one commits to using a DNN based approach. In order to provide a level of interpretability and understanding, the implementation of explainable AI (XAI) methods will be integral. This paper presents our findings in the use of EO and passive RF data for the purposes of heterogenous sensor fusion via MVI-DCGAN and XAI methods, with respect to generating a fused P-RF and EO overlay.

II. Literature Review

A. EO/RF Sensor Fusion

A desired application for DL is to accurately detect and track vehicle targets using EO and P-RF sensor inputs. The fusion of EO and RF data for the purposes of tracking has been used extensively in similar applications [2] [3] [4], but the use of passive RF data is more challenging to implement without conventional methods such as Doppler radar. While the research focus has traditionally been on active RF sensors, the use of P-RF data comes with logistical and economic benefits, as it requires less power, is considerably harder to detect than active RF methods, and requires less hardware.

RF modalities excel in providing range, angular, and spectral resolution of collected information [5]. The benefits of combining RF data with higher spatial resolution of EO based sensors are extremely desirable for detection and tracking. There are several RF-based approaches that are used in applications such as tracking [6], proximity [7], localization [8], and detection [9]. Most EO modalities are intuitively easier to implement and for humans to understand owing to a human's reliance on sight, such as full motion video (FMV) and infrared (IR). RF-based sensors can also provide repetitive coverage over a wide geographical area, and in doing so, can determine the precise distance and velocity of a target.

B. Explainable AI

With the increase in deep learning and NNs in the research or industrial applications, the importance of having interpretability and understanding of these models cannot be understated. As explainable AI is an emerging concept, there has yet to be any uniform adoption of interpretability assessment criteria for XAI. There are several criteria used to describe different approaches for providing explainability. These include (1) post and ante-hoc methods, which describe *when* the method itself is implemented in the model, (2) local or global, which describe *what* level of interpretability is being provided, or (3) model agnostic or model specific, which describe *how* versatile the method is. Some examples include Bayesian Rule List, which is an ante-hoc explanation, LIME (Local Interpretable Model-Agnostic Explanations) and SHAP (SHapley Additive exPlanations) which are model-agnostic explanations, and partial dependence plots, which are global and model-agnostic in nature.

The type of explanations desired and what types can be applied are dependent on the model and desired application. For image processing, visualizations can be highly desirable. Visualizations are a post-hoc XAI methods that include gradients, activation maximizations, deconvolutions, and decompositions. These techniques use tools such as generative models or saliency maps in order to determine activations produced on the last layer of a deep convolutional neural network (DCNN). From these activations, DCNNs can form a pixel-by-pixel mapping that highlights what factors provided the highest level of confidence in that decision, which then can be overlaid on top of the original sample.

The explanation insights are different based on the application. In the case of visualization methods, heat maps and feature highlights provide a more user-oriented explanation that a human can understand. Visualization explanations are not restricted to users who have expert knowledge but are intuitive owing to the visual information being presented. In most cases, visualization insight is hard to quantify, though metrics such as Fréchet inception distance can be used to quantify similarities between different images. The other types of visual insights provided are more easily shown in empirical form, using algorithms such as LIME or SHAP, which include a comparison of what data sources are more relied upon on both local and global level, or to gain understanding as a human user as to what the model views.

C. DCGAN

In 2016, Bengio and LeCun [10] designed a deep convolutional neural networks Generative Adversarial Network (DCGAN) using two training techniques: feature matching and minibatch discrimination. In doing so, the DCGAN enhances its diversity of the discriminate network when discriminating samples. The idea behind DCGAN essentially aims at increasing the complexity of the generator network, in order to input it into a high dimensional tensor and add deconvolutional layers to go from the projected tensor to an output range. These deconvolutional layers will expand on the spatial dimensions, whereas a convolution layer will decrease the input spatial dimensions.

The DCGAN architecture presents a strategy for using convolutional layers in the GAN framework to produce higher resolution images. The development of DCGAN-based models therefore expands traditional GAN methods from multi-layer perceptron (MLP) into using a CNN structure; allowing the GAN to retain its ability to generate excellent data but incorporating the advantages of CNN feature extraction for image processing. The enhanced CNN provides a high-resolution image generation for applications in anything from the medical to the robotics fields, or for the creation of Deep Fakes.

There will always be tradeoffs when implementing new algorithm. DCGANs, much like the DCNNs they're based on, require large amounts of data [11] and can also run into issues with overfitting [12]. As the models are neural network based, it becomes possible to use XAI methods, such as LIME [13], SHAP [14], and saliency maps [15]. Combating the issue of overfitting and inefficiencies can take a number of different forms. Some models work to avoid mode collapse by using implicit variational learning or manifold-guided training [16]. For the problem of limited data, the generation of simulation data [17] can also be used as a solution. For the issue of overfitting, limiting the generator's samples to provide "richer" corrective feedback to prevent overfitting or counterfactual image generation [18].

AFOSR grant FA9550-18-1-0287.

III. Experiment Design

This Section covers a discussion of previous research and the implementation of the explainable DCGAN for tracking targets from multi-modal data. The model's objective in these experiments is the tracking and identification of different potential vehicle targets using EO and P-RF data. Finally, the steps taken for the development and experimentation of the modified DCGAN for fusion of data as FuDCGAN are established.

A. The Escape Dataset

In previous research, our research group had carried out heterogenous sensor fusion using a combination of P-RF and EO modalities using the ESCAPE dataset [19]. The ESCAPE dataset contains a large number of different scenarios and a variety of different sensors, including modalities such as acoustic and seismic.

For the purposes of the scenarios and sensor data chosen, the three that were used in this research are designated as Scenarios 1, 2, and 3, which correspond to the dataset's Scenarios 1, 2C, and 2D respectively. The number of vehicle targets between the three scenarios totals 10, and each scenario deals with a different number of targets. The overall purpose of the dataset is that all the targets are designed to "evade" detection, by employing a number of different tactics that incentivize the fusion model to use different modality data input. This approach comes with the benefit of thereby challenging any model or algorithm meant to differentiate between potential targets when engaging in tracking, as similar targets being moved in a manner that even human users might have difficulty in differentiating between with only visual information. The three scenarios all involve multiple vehicles entering and exiting a garage, with multiple vehicles of similar make and build as well as more visually different vehicle targets.

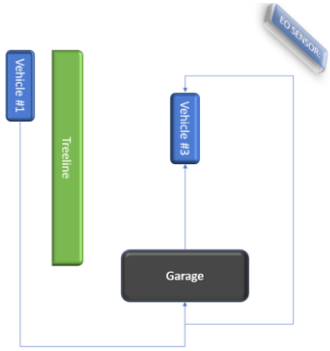

Fig. 1. Scenario 1 overview.

Scenario 1 has two total vehicle targets, both of which look visually similar to the other. For each of these scenarios, only one source of EO data was used to maximize the need for the model to utilize the RF data rather than ignoring the P-RF input. Vehicle #1 travels into the garage, which is easily observed via the EO input. As this happens, vehicle #2 travels into the garage from behind the tree line. In doing so, from the EO sensor's point of view it is "hidden" due to visual obscuration that prevents the model from detecting its movements most of the time. Once vehicle #1 enters the garage, vehicle #2 then exits the garage, and the objective of the first scenario is to successfully determine when the "switch" is made. If the model incorrectly determines that the vehicle exiting as #1, then that means the model has failed and the vehicle has successfully "evaded" detection.

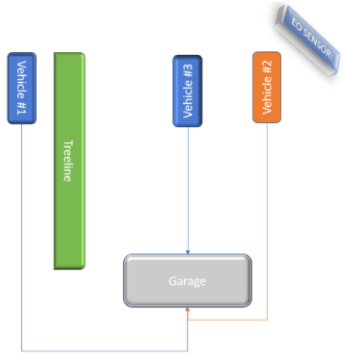

Fig. 2. Scenario 2 overview.

Scenario 2 is nominally more complicated. It has three vehicles and essentially follows the same pattern as Scenario 1, but only two of the three look visually similar. The difference is that rather than vehicle #1, which is visible, or vehicle #2, which is not possible to obtain at the video angle chosen switching in the garage, is that vehicle #3 that was parked in the garage the entire time. This makes it appear that vehicle #1 enters and exits when in fact it is hidden inside of the garage. The DOF EO input will not be sufficient on its own to determine the difference between the three vehicles, as the source selected does not have access to all of the different EO sensors.

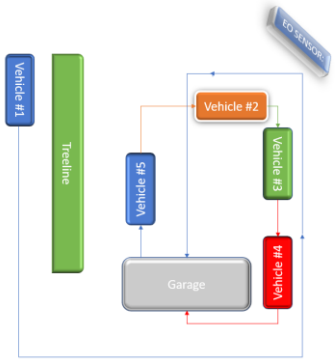

Fig. 3. Scenario 3 overview.

Scenario 3 is the most complicated of the three and chosen due to the complexity of the five vehicle targets, all traveling at different speeds and with different makes. Four of these vehicles arrive out of the front of the garage, while the fifth vehicle arrives from out of view, thereby making the tracking at the end of the video input, linearly speaking, extremely difficult to conduct with only the EO input for that time frame. The variable speeds displayed by the five vehicle targets also presents an additional dimension of complexity with respect to tracking as the vehicles that are similar in design will overtake the other at different points within the scenario, making tracking a challenging process for Scenario 3.

While it is possible to implement tracking by including all EO sensor sources in the experiment design, the sensors chosen are only limited to one EO sensor and the three sources of P-RF sensors. By purposefully limiting the most intuitive sensor that provides the more reliable data, the model is incentivized to rely on the other source of information it possesses. From our findings with implementing explainable AI research via Saliency Maps, covered in the next section, we inferred several behaviors that the maps indicated of the model's behavior and decision-making process [20], which are detailed below.

B. *Saliency Map Activation Research*

The data is first preprocessed by combining dense optical flow for the EO data, and then creating histograms of the P-RF data [21], in the form of in-phase quadrature component data (IQ data). This method of preprocessing then combines the two sources of data into an overlayed view [20] for the purposes of implementing explainable AI via saliency maps. While it would be more ideal to split the two sources of data and then feed utilize them simultaneously, our research found that the fused view performed better, owing to the less impactful P-RF data. On its own, the P-RF data cannot be used effectively, performing too poorly for the saliency maps to show anything as the model is essentially guessing and therefore has nothing to show on the visualizations. The poor explainability of P-RF is to be expected, as the majority of the histograms are rather noisy with the exception of several events in which a visually noticeable change in the histograms occurs, with respect to entering and exiting the garage.

While it is possible to implement visualizations of the discriminator and use the resulting discriminator visualizations as a source of data for the generator, the resulting input would be radically different than the fused overlay input. Likewise, it is important to ensure the data being fed into the generator and discriminator are different in nature as they would otherwise not be useful for the discriminator to train on, by virtue of the generated images looking nothing like the generated visualizations. By feeding visualizations input from different deep learning models that use the same overlay input, a level of stability when feeding the insights of other models is ensured.

TABLE I. INPUT MODEL VISUALIZATIONS

Deep Learning Model:	***Samples per Scenario:***	***F1 Score:***
Convolutional Neural Network (CNN)	374	0.96
Feed Forward Neural Network (FFNN)	372	0.93
Deep Neural Network (DNN)	374	0.95

Each of the three scenarios has 1120 samples of fused overlay input for each target, and the visualizations from the three different deep learning models for each scenario were used as the input. The visualizations for each of the models used were shuffled randomly, to ensure the training data's integrity. While the model's individual performances were different, as seen in Table 1, the observations and overall heatmaps were similar with respect to the trends portrayed.

While the exact distribution of sample visualizations for the DCGAN isn't as ideal, the FFNN model's samples were chosen to be the smallest by two samples by virtue of the relatively higher performance of the CNN and the DNN. The added benefit of this approach is that it confirms a shared direction and coherence with respect to the features used. The visualization categories observed remained within the EO focused, RF Focused and Fusion Focused [20] variants, with no noticeable outliers. The visualization categories describe what types of information from the fused view demonstrates the most neuron activations. EO focused visualizations will focus solely on the target (vehicle), RF focused visualizations will focus on different parts of the histogram portions of the sample, while Fusion focused visualizations will tend towards both types of data. The trends to when the visualizations were used matched based on the target vehicle, focusing on EO focused visualizations primarily until it became impossible to visually locate the target (RF focused) or until similar targets were in range (Fusion focused).

C. *MVI-DCGAN*

First described in 2016, the implementation of a DCGAN [10] has led to the use of deep convolutional generative adversarial networks (DCGANs). DCGANs come with a variety of benefits design wise, having a relatively lower sampling cost, and having state-of-the-art performance in image generation. This, however, comes at the cost of being unable to calculate likelihood p model(s) as there are no modeled probability distributions, nor can latent variables be inferred from a sample. However, they come with a variety of different benefits, especially for the purposes of image processing.

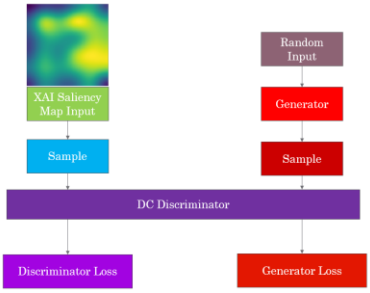

Fig. 4. MVI-DCGAN model overview.

For the purposes of implementing explainable AI via MVI-DCGAN, the choice of DCGAN provides a DL-based approach that is capable of synthesizing data and differentiating between the produced data. Noting that the input of the MVI-DCGAN model is based in the randomly shuffled visualizations of a CNN, FFNN, and DNN. The distribution of the input frames from different visualization sources was conducted at random, and to ensure that what differences there are in the visualizations from different sources are better distributed. These visualizations are then fed as the samples for the discriminator, with the generator producing their own samples to input into the discriminator, as seen in Figure 4.

Once the MVI-DCGAN receives these visualizations, it begins the adversarial training. The input overlay visualization data are read and then resized into 96×96 images which are then fed into the discriminator. Both the discriminator and generator are CNNs. The discriminator and generator both have five layers, with the discriminator dense layers increasing

from 32 to 512. The generator uses the Tanh activation function, while the discriminator uses sigmoid. In order to gain visualizations from MVI-DCGAN, the models are saved and used to provide neuron activations.

The performance of the MVI-DCGAN was evaluated using the Fréchet Inception Distance (FID), which measures the difference of distributions of feature vectors for different image datasets. A larger FID score indicates larger differences and a poorer performance by the generator. Conversely, a smaller FID score indicates a greater performance by the generator with respect to generating similar images. When it comes to applications that are image based, such as saliency maps and generated images, measuring the quality can be difficult, not to mention can be inherently subjective [22]. For the purposes of these experiments, different FID scores were evaluated with respect to each of the 10 total targets across the three scenarios used, as detailed in Table 2.

TABLE II. MVI-DCGAN GENERATOR FRÉCHET INCEPTION DISTANCE (FID) SCORES

Scenario:	*Target:*	*FID Score:*
Scenario 1	S1 Vehicle #1	57.29
	S1 Vehicle #2	69.85
Scenario 2	S2 Vehicle #1	63.47
	S2 Vehicle #2	55.89
	S2 Vehicle #3	65.39
Scenario 3	S3 Vehicle #1	61.07
	S3 Vehicle #2	94.73
	S3 Vehicle #3	77.59
	S3 Vehicle #4	64.72
	S3 Vehicle #5	89.61

As seen in Table 2, the most difficult scenario for MVI-DCGAN to process was scenario 3. The scenario contains the most moving vehicles, with a total of five, and entails a large number of vehicle shuffling between the positions of the other vehicles. The similarity of the vehicles in terms of make and build are inherently removed by virtue of the preprocessing, as dense optical flow focuses on the target vehicle's movement more than its appearance. Given the movements for different vehicles, the poorest performance by the generator was with respect to vehicles 2, 3, and 5, while vehicles 1 and 4 performed comparable to the other vehicles in scenarios 1 and 2. Scenario 3 was also the only major outlier in which lower FID didn't correlate with a higher F1 score. Scenarios 2 and 1 are similar in terms of vehicles and movements, though the highest FID score and thus the poorest performance between the two would be Scenario 1's Vehicle #2. The visualizations for Scenario 1, Vehicle#2 are more focused on P-RF view visualizations and fused view visualizations but have less EO view visualizations.

From the scenarios and vehicle targets tested, Scenario 2, Vehicle #2 maintained the best FID score. As seen above in Figure 2, the generated visualizations display all three categories of visualizations (EO focused, RF focused, Fusion focused) for the scenario. While the quality of the generated image is only befitting of a generated image with an FID score of over 50, the visualizations remain focused on their corresponding frames. It is clear from all of the generated visualizations that the features learned were more focused on the P-RF and Fusion view than the EO views. The majority of the inputs and generated visualizations for scenario 2 shown also correspond with the different model's visualizations, focusing on the same types of neuron activation patterns.

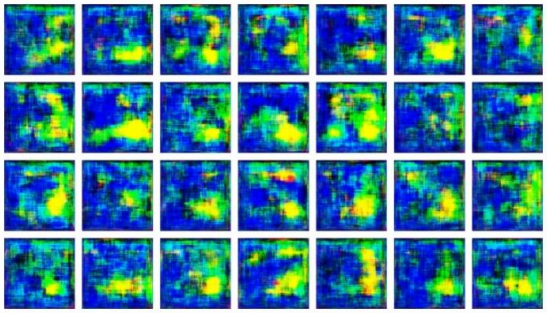

Fig. 5. Assorted generated overlay visualizations from the MVI-DCGAN Generator (Scenario 2, Vehicle #2) at different timeframes for Scenario 2.

While the results of the highest performing MVI-DCGAN generator's visualizations have a number of different variations, the discriminator's ability to differentiate between the generated and original samples proved to be more interesting. To determine if the MVI-DCGAN's training would be effective on the original overlayed EO/P-RF data, where the original overlay inputs were also used to test the discriminator's performance. By using the discriminator and saliency map visualization on its last DL layer, it becomes possible to gain insights on the input image and the neuron activations from that generated image.

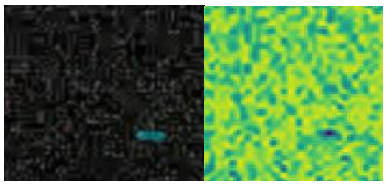

Fig. 6. Comparison of overlay input and visualization of the MVI-DCGAN Discriminator (Scenario 2, Vehicle #2) .

From Figure 6, it can be inferred that the discriminator focuses on the P-RF data for determining if an image is generated or a part of the training data. From the resulting visualizations, it appears that the discriminators neuron activations have an almost inverted insight into the overlay input, which even goes as far as to exclude the vehicle target #2, despite the vehicle in the overlay being its target. From the fusion neuron activations, it can be imitator has learned to focus on the P-RF features of the overlays and, for the purposes of this dataset, used that information in order to better differentiate from the generator's inputs. The other frames for tracking vehicle #2 of the scenario also show similar levels of inversion, as it appears that the EO aspect of the visualizations becomes less reliable for the discriminator, and thus aren't the focus features of these visualizations.

IV. CONCLUSION

In order to improve the fusion and use of P-RF and EO data, this paper introduces Multiple Visualizations Input (MVI)-DCGAN. The proposed MVI-DCGAN uses multiple sources of visualizations from other models that have been trained with fused overlay EO/PRF data, in order to utilize features to differentiate them from generated visualizations. The initial experimentation shows that while the model's F1 score corresponds to the FID performance of the trained model, it is not always the case. When used to determine what features the discriminator focuses on, there was an overwhelming focus on the P-RF aspects of the overlay input, going as far as to ignore the majority of the target vehicle. From these results, it can be inferred that the MVI-DCGAN generator's details with the edges of the vehicle target as well as the P-RF aspects of the data play the largest parts in terms of differentiating between generated image inputs as well as for differentiating between different targets.

While the discriminator's insights were of interest, the MVI-DCGAN's generated images could be improved in terms of FID score. In future research, a focus on increasing the sample size will be of importance, as will the number of models used for the input visualizations. Additional methods of providing explainability, such as LIME or SHAP, would be desirable as would additional metrics for evaluation of the generator's performance such as Inception Score (IS) or Rapid Scene Categorization. The results of the MVI-DCGAN were sufficient to train the discriminator on P-RF features of the overlay visualizations. The primary contributions of the MVI-DCGAN approach are the novel usage of other model's visualizations to "steer" the discriminator towards the usage of certain features and enforcing that focus with adversarial training. Our previous research has shown that for the purposes of vehicle tracking and classification, the overlay image inputs are sufficient for conventional and deep learning-based approaches, and MVI-DCGAN has demonstrated the possibility of learning to better rely on desired features with adversarial training.

ACKNOWLEDGMENT

This research is supported by AFOSR grant FA9550-21-1-0224.

REFERENCES

[1] M. Pulido, P. Melin and O. Castillo, "Particle swarm optimization of ensemble neural networks with fuzzy aggregation for time series prediction of the Mexican Stock Exchange," Information Sciences, vol. 280, pp. 188-204, 2014.

[2] B. Kahler and E. Blasch, "Sensor Management Fusion Using Operating Conditions," in IEEE National Aerospace and Electronics Conference, Dayton, Ohio, 2008 .

[3] A. Vakil, J. Liu, P. Zulch, E. Blasch, R. Ewing and J. Li, "A Survey of Multimodal Sensor Fusion for Passive RF and EO Information Integration," IEEE Aerospace and Electronic Systems Magazine, vol. 36, no. 7, pp. 44-61, 2021.

[4] E. Blasch, A. Vakil, J. Li and R. Ewing, "Multimodal Data Fusion Using Canonical Variates Analysis Confusion Matrix Fusion," in IEEE Aerospace Conference (50100), Dayton, OH, 2021.

[5] D. Shen, P. Zulch, M. Disasio, E. Blasch, G. Chen, Z. Wang, J. Lu and R. Niu, "Manifold learning algorithms for sensor fusion of image and radio-frequency data," in IEEE Aerospace Conference, Big Sky, Montana, 2018.

[6] M. A. Shukoor, S. S. Mukeshbhai and S. Dey, "12-Bit Multiresonator Based Chipless RFID System for Low-Cost Item Tracking," in IEEE International Conference on RFID Technology and Applications (RFID-TA), Delhi, India, 2021.

[7] Y. Nishida, "Proximity Motion Detection Using 802.11 for Mobile Devices," in IEEE International Conference on Portable Information Devices, Orlando, Florida, 2007.

[8] J. Liu, H. Mu, A. Vakil, R. Ewing, X. Shen, E. Blasch and J. Li, "Human Occupancy Detection via Passive Cognitive Radio," Sensors, vol. 20, no. 15, p. 4248, 2020.

[9] J. Liu, R. Ewing, E. Blasch and J. Li, "Synthesis of Passive Human Radio Frequency Signatures via Generative Adversarial Network," in IEEE Aerospace Conference (50100), Dayton, Ohio, 2021.

[10] Y. Bengio and Y. LeCun, "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks," in 4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, 2016.

[11] J. Kim and H. Park, "OA-GAN: Overfitting Avoidance Method of GAN oversampling based on xAI," in Twelfth International Conference on Ubiquitous and Future Networks (ICUFN), Barcelona, Spain, 2021.

[12] W. Fang, F. Zhang, V. S. Sheng and Y. Ding, "A Method for Improving CNN-Based Image Recognition Using DCGAN," CMC-Computers, Materials & Continua, vol. 57, no. 1, pp. 167-178, 2018.

[13] V. Nagisetty, L. Graves, J. Scott and V. Ganesh, "xAI-GAN: Enhancing Generative Adversarial Networks via Explainable AI Systems," arXiv: Learning, 2020.

[14] A. Ignatiev, N. Narodytska, N. Asher and J. Marques-Silva, "From Contrastive to Abductive Explanations and Back Again," in International Conference of the Italian Association for Artificial Intelligence (AIxIA Lecture Notes in Computer Science), Milan, Italy, 2020.

[15] V. Jain, O. Nankar, D. J. Jerrish, S. Gite, S. Patil and K. Kotecha, "A Novel AI-Based System for Detection and Severity Prediction of Dementia Using MRI," IEEE Access, vol. 9, pp. 154324-154346, 2021.

[16] D. Bang and H. Shim, "MGGAN: Solving Mode Collapse Using Manifold-Guided Training," in nternational Conference on Computer Vision Workshops (ICCV Workshops), 2021.

[17] Y. Xie and T. Zhang, "Imbalanced Learning for Fault Diagnosis Problem of Rotating Machinery Based on Generative Adversarial Networks," in Chinese Control Conference (CCC), Wuhan, China, 2018.

[18] M. Suzuki, Y. Kamcya, T. Kutsuna and N. Mitsumoto, "Understanding the Reason for Misclassification by Generating Counterfactual Images," in 2021 17th International Conference on Machine Vision and Applications (MVA), Aichi, Japan, 2021.

[19] P. Zulch, M. Distasio, T. Cushman, B. Wilson, B. Hart and E. Blasch, "ESCAPE Data Collection for Multi-Modal Data Fusion Research," in 2019 IEEE Aerospace Conference, Big Sky, Montana, 2019.

[20] A. Vakil, E. Blasch, R. Ewing and J. Li, "Visualizations of Fusion of Electro Optical (EO) and Passive Radio-Frequency (PRF) Data," in NAECON 2021 - IEEE National Aerospace and Electronics Conference, Dayton, OH, 2021.

[21] A. Vakil, J. Liu, P. Zulch, E. Blasch, R. Ewing and J. Li, "Feature Level Sensor Fusion for Passive RF and EO Information Integration," in 2020 IEEE Aerospace Conference, Big Sky, Montana, 2020.

[22] H. Boujut, A. Bugeau and J. Benois-Pineau, Semantic Multimedia Analysis and Processing (Visual Search for Objects in a Complex Visual Context: What We Wish to See), Boca Raton, FL: CRC Press, 2013, pp. 31-63.