Visualizations of Fusion of Electro Optical (EO) and Passive Radio-Frequency (PRF) Data

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
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

NAECON 2021 - IEEE National Aerospace and Electronics Conference | 978-1-6654-4859-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/NAECON49338.2021.9696424

***Abstract*—In machine learning, the ability to reliably determine potential pitfalls in the decision making process of an algorithm is essential. In previous research, the fusion of passive radio frequency (P-RF) histograms generated from in-phase quadrature component (I/Q) data and enhanced electro-optical (EO) data are fused together in order to implement classification and tracking of different vehicle targets using the AFRL's ESCAPE dataset. In previous research, the impact of the P-RF data was confirmed to essential for achieving a higher performance when fused with EO data. This research, however, did not provide an intuitive means by which inferences and explanations for a human expert could interpret. In this paper, saliency maps are implemented in order to visualize the impact of P-RF data in the fusion model and thereby confirm the role the different modalities play in the heterogeneous sensor fusion in an intuitive manner. These visualizations provide context for which pixels activate neurons in the final layer of the model. Overlaying the two modality inputs with respect to time, the method presented in this paper is able to provide explainability for the fusion model while also achieving an F1 score of 0.9. The research in this paper uses the distribution and frequency of the appearance of different types of visualizations. Combined this with context from the scenarios with respect to the timeline of events, it becomes possible to draw inferences for how the two modalities are utilized by the fusion model.**

Keywords—Explainable AI, Heterogeneous Sensor Fusion, Electro-Optical, Passive Radio Frequency, Deep Learning, Feature Level Fusion, Histogram of In-phase and Quadrature Components, Overlayed Images, Visualization, Saliency Maps, Heat Maps

I. Introduction

The use of machine learning (ML) has considerably grown in both research and industrial application with the success of deep learning (DL) and neural networks (NN). Neural networks come with a variety of benefits, especially given the potential such blackbox algorithms have when it comes to end-to-end learning. Even with the potential such blackbox algorithms contain, it is critical to have an understanding or some level of transparency. When decision-making errors occur, it is imperative to have some understanding of how such mistakes were made, both for developmental and application reasons. A momentary mistake with a computer vision algorithm for an autonomous vehicle application could lead to a fatal accident. A decision-making error for a neural network that is used for a financial application could lead to a catastrophic loss of wealth. In the medical field, specifically in computer aided diagnosis, a false negative could lead to a fatal misdiagnosis.

For this reason, interpretability and explainability of ML algorithms become an increasingly pressing issue. It is important for accountability when such decision-making errors happen, and more importantly determine how and why the error occurred. Even if the model performs well, it is imperative to determine that it has been trained properly, as opposed to exploiting features that are only *exclusive* to the training set. It is therefore desirable to have some level of transparency even for development purposes. There are many methods to either recreate the framework to capture interpretability or to at least illuminate or provide inferences to the decision-making process of an AI model [1]. The field of explainable AI (XAI) has become a hotspot of the ML research community. There have been many attempts to define the notions of important XAI terms such as interpretability, explainability, reliability, and trustworthiness, but not necessarily a clear notion of how to incorporate them into the wide array of diverse applications that machine learning includes. For this reason, in this paper, we will use the terms "explainability" and "interpretability" in an interchangeable manner.

In previous research [2], it was determined that the impact of P-RF data in the fusion model was beneficial to its performance. However, compared to the more intuitive EO modality, the P-RF impact presented a major challenge. With the usage of dense optical flow (DOF) any potential targets detected with the EO modality are clearly shown when fed to the model. The P-RF data on the other hand is largely ambiguous due to the noise in the histograms. Even if the fusion model's performance is improved, determining why and how the P-RF data impacts the model is important.

For the purposes of achieving explainability with the fusion of P-RF and EO data, visualizations and overlayed inputs are used. Visualization methods provide more intuitive explanations from which we can infer the behavior of the fusion model. Using the frequency and distribution of this data, combined with expert knowledge of the events occurring with respect to time, inferences on how the P-RF data is utilized by the fusion model can be made. These inferences are backed by the model's performance and used as a means of confirming the impact by both modalities in the decision-making process.

II. Literature Review

A. EO/P-RF Sensor Fusion and the ESCAPE Dataset

For this research application, the goal was not to rely on active RF inputs, but to exploit the potentials of passive RF (P-RF) sensors in an explainable manner. 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.

For the multimodal heterogeneous sensor fusion research presented in this paper, the raw RF data are preprocessed to obtain I/Q histograms with respect to time. The histograms are aligned with the simulated EO data for the purposes of detecting and classifying the number of vehicles. From previous experiments that had compared the single input vs the fused output with respect to the ground truth of the simulation, neither standalone modality was able to achieve a performance greater than an F1 score of 0.9, with the P-RF histogram data being incapable of exceeding even an F1 score of 0.6 when given more than one potential choice. But with the fusion of both the DOF-EO image input and P-RF histogram inputs, the model could reach an F1 Score of 0.95.

Using Machine Learning, specifically Neural networks (NN) and deep learning, was an obvious decision as such methods boast of powerful classification ability. Some neural 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, 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-principal physics equations that approximate the problem.

B. Explainable AI

As explainable AI is an emerging field in research and industry, there has yet to be a widely adopted standard for explaining models. These types of explanations come in a wide variety of types, such as ante-hoc and post-hoc; local or global; or model agnostic and model specific. Such terms are derived from what level the explanations originate from to how the method can be applied in terms of other models. Even discussing methods of how to quantify such approaches is quite a task in of itself, as there are a number of different classifications for such types of interpretability. To even begin discussing the topic of explainable AI, the most important thing to do is to define interpretability.

There are a number of definitions of interpretability or explainability. For domains that deal heavily in images for example, interpretability might be defined as being able to map the predicted class into a domain that the human user might be able to make sense of. For non-image oriented inputs, being able to assign a weight to certain features might be a better way to provide transparency. In an ideal system, one might even define interpretability as a reasonable explanation as to why a collection of features contributed to the decision-making process, or at least determining how much weight the decision-making process gave to said features [4]. Whichever definition of interpretability one might subscribe to, given the lack of a widely adopted standard, so long as the method provides insights that can answer questions regarding how and why the model performs in the way that it does, that is a method that provides some level of transparency.

There are many means by which to divide different categories of XAI methods, such as the mechanism (sensitivity, decomposition, optimization, inversion [5], etc.), or procedure (ad-hoc, post-hoc, model specific, model agnostic [6], etc.). Regardless of the type of XAI method, the *end results* of such explanations traditionally can be broadly categorized into two types: analytical and intuitive explanations. Many approaches will provide analytical results, from the usage of LIME (local interpretable model-agnostic explanations) to SHAP (Shapely Additive exPlanations) [7]. Such methods can provide information such as determining the impact on the model output in a model-agnostic manner. As for intuitive explanations, while not as empirical in nature, with the knowledge of a human expert can be provide context for the decision making process. These include methods like visualization [8], such as Grad-CAM, which provides a post-hoc explanation for decision making process. Such methods typically work along with feature relevance techniques to provide the end result.

In previous research [9], we had determined the weighted impact of the P-RF data for the fusion model. The results indicated that the fusion of P-RF data with EO inputs was necessary for achieving a higher performance score. While it indicated the P-RF had an impact on the decision making, it was not clear as to when and how the P-RF aspect of the Fusion helped in decision making. As such, the next logical step was to begin researching the impact of the modalities in an intuitive manner.

C. Previous Research in XAI with respect to EO/P-RF Fusion

The ESCAPE dataset contains a number of scenarios and different sensor input modalities. Three of these scenarios that are used as the focus of this research for the purposes of this paper. Scenario 1, scenario 2C and scenario 2D, which are designated in this paper as scenarios 1, 2, and 3 respectively. Each of these scenarios uses a different number of vehicles. But for the purposes of collecting data, the sources of sensor input collection remain MTRI 11, 12, and 13 for P-RF data collection and one sole source of electro-optical data, MTR-EO-04. The usage of a single EO input is to ensure the P-RF input is necessary for fusion to achieve a higher performance.

AFOSR Grant FA9550-18-1-0287.

There are a total of five different types of ground vehicles used in the dataset, a gas motor Gator utility vehicle, a diesel motor Gator utility vehicle, a pickup truck, a panel van, and a stake rack truck. These five vehicles are the primary focus of the ESCAPE dataset, and by design are always the aforementioned targets of the dataset. As mentioned earlier, the scenarios are all made to "evade" detection, thereby challenging the models trained with this data set on their ability to differentiate between potential targets when engaging in tracking.

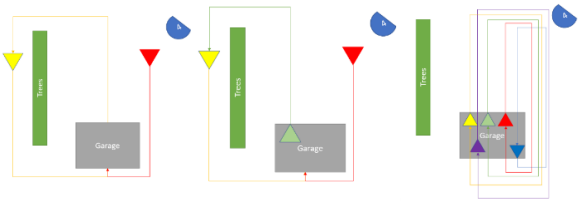

Figure 1: Comparison of Scenario's 1 (left), 2 (middle), 3(right)

Scenario 1 has two targets, with vehicle #1 traveling into the garage as seen via the EO input, while vehicle #2 travels into the garage unseen due to visual obscuration that prevents the DOF EO processed data to be unable to pick up the image. Both vehicles are of the same model, as the intention is to "deceive" any viewers into believing that the same vehicle is traveling in and out of a garage when in fact they have "switched" entirely. When vehicle #1 enters the garage, vehicle #2 exits the garage, and the objective of the first scenario is to successfully determine if the model can tell if and when the "switch" is made.

Scenario 2 has three vehicles and follows the same pattern as scenario 1. 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, it is vehicle #3 that was parked in the garage the entire time that arrives to the location of vehicle #3. The third target, vehicle #3, is of a completely different model than #1 and #2 but provides a potential false target that is actively moving to increase the complexity of the scenario. The DOF EO will not be sufficient to determine the difference between the three vehicles as the source selected does not have access to all of the data.

Scenario 3 is the most complicated of the three and chosen due to the complexity of the five vehicle targets traveling at different speeds and with different models. Four targets (pickup truck, diesel motor, Gator motor vehicle, a gas motor utility vehicle) arrive out of the front of the garage while fifth vehicle (stake rack truck) 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 and appearance (diesel motor gator utility vehicle) will overtake the other at different points within the scenario, making tracking a challenging process for scenario 3.

Owing to the number of experiments with the ESCAPE dataset, prior research [10] indicated that there was some value with the P-RF histogram data with respect to correlation of certain events. While previous work with simulation data had proven to process the information cleanly, the amount of noise in the P-RF histograms made the correlation of sensor inputs difficult to understand. Further research on this phenomenon would occur with the usage of a greedy algorithm [9], but even at that point of interest there were clear changes that are apparent to any human observer when the vehicle leaves the garage.

The information provided by the EO modality, while highly efficient due to the application of dense optical flow, is not sufficient to accurately detect and categorize different vehicle targets by itself. The single view makes achieving a high level of performance impossible due to the periods of time in which the vehicle cannot be detected via the EO data input. Although there are more than one sources of EO data, only one input is used to promote the necessity of the P-RF histogram data. The P-RF data by itself, when used as the sole input for a model, is not sufficient for classification. That being the case, the P-RF histograms did host some key details that, while not always obvious to the human eye, did provide the neural network model with information needed to accurate train with respect to the ground truth of each of the scenarios tested. While the results and the weights of the P-RF data could be gauged, this still did not provide an intuitive insight into how and when the P-RF input was used. For this reason, it became imperative to conduct research into the use of visualizations with EO/P-RF sensor fusion.

III. Technical Approach and Experiment Design

A. *Experiment Overview*

While the approach to the fusion of P-RF and EO data was successful, there existed a number of questions regarding the effectiveness of the P-RF input. During research using a greedy algorithm, we had discovered that the P-RF data still played a relevant part in the decision-making process of a canonical correlation analysis based long short-term memory (LSTM) model. While the greedy algorithm recreated the model and did provide a variety of insights into the model, the local and global impact of different data frame attributes did not provide as intuitive of an explanation for the model's usage of the lowest weighted source of information, the P-RF histograms.

In order to obtain a more meaningful insight into the impact that the P-RF data was having on the model, the use of a post-hoc visualization method was implemented, specifically that of a Saliency Map. The issue of how to exploit the activated neurons with EO and P-RF data was initially something that had stumped our early research, as using multiple images within the same input made it difficult to implement available Saliency Map approaches. In most applications of visualization techniques that do include fusion, the melding of the two or more image sources being mapped together [11] but unfortunately such mapping would not be possible to implement with the P-RF data. Visualization methods require the use of a single image input, making it impossible to keep the two sources of data separate.

B. *Overlay Preprocessing*

For this reason, and in order to prevent a disparity between different modalities with respect to the activations used by the two modalities if placed side by side in the same image input, the decision to overlay the images over each other was made. The image inputs were already synchronized with respect to time and image size in order to handle fusion. Once overlaying

was implemented, the data was ready to be processed through a Convolutional Neural Network.

When attempts were made at processing the two inputs separately in earlier experiments, issues arose in the form of the P-RF data. While the saliency maps of the EO data would produce heatmaps that more or less matched the DOF input, the P-RF data would not produce any heatmaps. Owing to the performance of often less than an F1 score of 0.5, the model was essentially guessing and thereby did not rely on any features provided, producing a blank saliency map. As this made using the visualizations separately impossible, the next step was to determine how to better combine the two sources.

Traditionally, when using saliency map visualizations, the method of fusion typically maps one or more features on a shared map [12] [13] [14]. This is not possible to do with the ESCAPE dataset with the P-RF and EO data, however. There is a wide array of information that the dataset contains but GPS data is not available, as that would defeat the purpose of the sensor fusion tracking application. As mapping spatially onto the EO aspect of the overlay is not possible, the next step to consider is whether or not the fusion overlay will interfere with the performance of the model. Using the two sets of inputs synchronized with respect to time and comparing with the performance of the fusion model yielded a difference of less than 0.2 in F1 score. As the P-RF data cannot be mapped separately from the EO data, this left the only viable option to be overlaying the DOF EO and P-RF histogram inputs together. With the data overlayed with respect to time, and the histogram not interfering with the impact of the EO aspect of the overlay, the visualization maps could be produced.

The results of the saliency map visualization provided different insights that indicated a close relationship with the usage of the P-RF histogram data when overlayed with the EO data. While the normal P-RF histogram data does not display any notable activations (owing to rarely exceeding an F1 score of 0.56 when there are only two outputs that exist within the possible outcomes), the EO data when used by itself would similarly be focused on the available information. The application of dense optical flow had initially been adopted in order to prevent the “false positives” that would occur when attempts at enhancing the EO modality would be implemented, such as edge detection. For this reason, the fusion view of overlaying the two inputs together solves the issue with regards to producing meaningful saliency maps. While the performance of this model was slightly lower than previous models, as opposed to the 1.0 F1 score the CCA fusion model possesses, the advantage is that the generation of the saliency maps would provide better insights into how and when the P-RF data is being used to produce a decision, as opposed to simply citing weights and differences in model performance with and without the P-RF input.

C. Visualization Categorizations

The three scenarios each produce a total of 1120 overlayed inputs, which contain the fused P-RF histogram and DOF EO information. This data is implemented in the training, and then heatmap visualizations are produced for each of the corresponding frames. The results of the application of saliency maps and the heat maps of the activated neurons provide a few useful insights, which vary based on points within each of the respective scenarios for the ESCAPE dataset. There are three major groupings by which the maps provide information, as one way or another the information needed to conduct supervised training to correlate the input data. One to five outputs are the level of complexity that this dataset goes through for the purposes of this experiment.

As visualizations are an inherently subjective and not qualitative method of explanations, they are categorized in this paper based on the focus of the produced heatmaps. The overlay avoids confusion with DOF EO aspects such as the treeline by the distribution of the P-RF aspect on the fused image. Thereby making differentiating between the EO aspect (vehicle movements) and the P-RF Histogram aspect of the overlay easier.

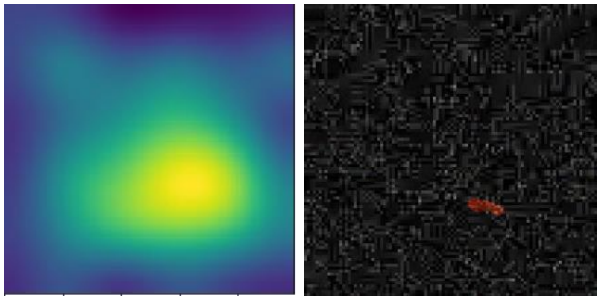

Figure 2: EO focused Overlayed Heterogeneous Input

The first of the three major categories are the activations that are inherently EO focused. These activations often occur in the simpler periods of data that can be found predominantly in all scenarios, due to the fact that there are large stretches of time in which the target vehicle can be seen in the overlayed frame. In scenarios 1 and 2, the switches between certain targets are considerably more linear, causing the focus to be on the EO aspect of the overlayed image rather than that of the P-RF histogram which has been overlayed on top of the DOF image whenever the appropriate vehicle in question is in plain view of the fusion model. When breaking down the movements of each of the three scenarios and training with respect to a particular vehicle, the overwhelming number of these EO focused activations were primarily when the vehicle being tracked occurred when the vehicle was visible on the DOF-EO data for the most part.

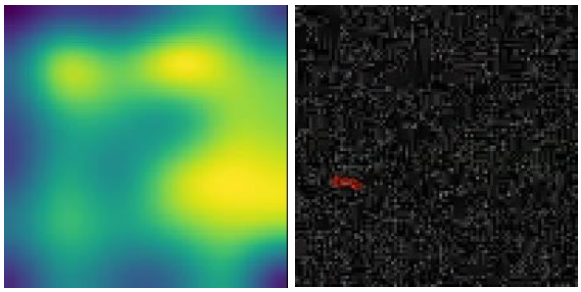

Figure 3: P-RF focused Overlayed Heterogeneous Input

The second of these three major categories are the activations that are inherently focused more on the P-RF activations. These activations are rarer, but only by virtue of the EO data predominantly capturing the majority of the vehicles’

"screentime" on the EO data. As seen in Figure 3, the focus on the vehicle [#2 in scenario 2] that can be spotted in the middle left of the overlayed image is nowhere near as focused on activation wise as the aspects of the histogram that have been overlayed on the right side of the image. Because this vehicle that can be seen was not the objective of that training run of the data [vehicle #1 in scenario 2], the activation of the vehicle that is visibly obvious as a potential target is nowhere near as focused on. From the heatmap we can infer that the neurons are not simply focused on the EO aspect of the overlayed image compared to the P-RF aspects.

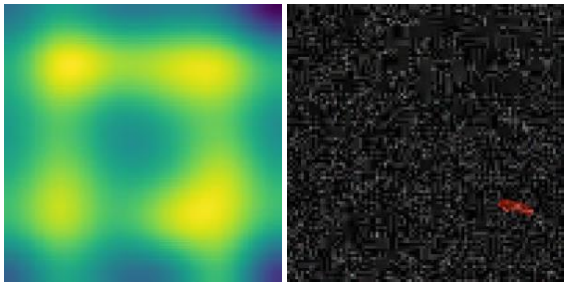

Figure 4: Fusion focused Overlayed Heterogeneous Input

The third of these three major categories are the activations that are focused on both the EO and the P-RF activations simultaneously. These were more common than the P-RF activations but not necessarily as common as the EO focused heatmaps within the three simulations. In situations where there is a potential target, but the actual target is not on screen, the model will use both sources of information on the overlayed input to attempt to determine the nature of the situation. scenario 1 and scenario 2 did possess a number of these moments, but within scenario 3 this occurred in the training for three of the five vehicle targets. While scenario 3 has more potential outcomes to train for, these moments seemed to heavily correlate with the start of the scenarios, linearly speaking. The example shown above in Figure 4 is from scenario 1, trained specifically for the vehicle designated as #2, which is out of sight from the EO source of the data, but a potential target is still in sight of the overlayed input.

IV. Results and Discussion

A. Scenario 1 Inferences

While the results of the saliency map neuron activation heatmaps don't provide anything that differs radically from the expected theoretical result, it does validate the expected outcomes based on the design of the supervised training. While the classification categories are considerably simple, it should be noted that the overlay network never peaked above a 0.9 F1 score, as the results of the training were only ever in the low to mid 0.8s. While some performance was sacrificed, the inferences as to what aspects of the data are aiding the training, are more obvious than it would be when using a blackbox system.

As is a major issue with any abstract XAI method, of course, the usage of terms such as fusion focused or P-RF focused or EO focused overlayed heterogeneous input. The classification of these saliency maps is subjective in nature and the classification of these saliency map overlays has to be done manually. Given the massive difference between a P-RF or fusion focused overlay compared to an EO focused overlay, as well as the rarity of P-RF focused overlays, that problem is mitigated to some degree. In order to attempt to evaluate the impact more objectively, and to avoid simply listing numbers without considering aspects such as time and the series of events that are occurring, the distribution of the scenario data is also graphed. To mitigate the effect of outliers and to better represent the data, the graph averages the values corresponding to their respective categories for each second. This comes with the advantage of viewing the changes with respect to time and to reduce the effect of outliers while still preserving the impact they have in terms of frequency.

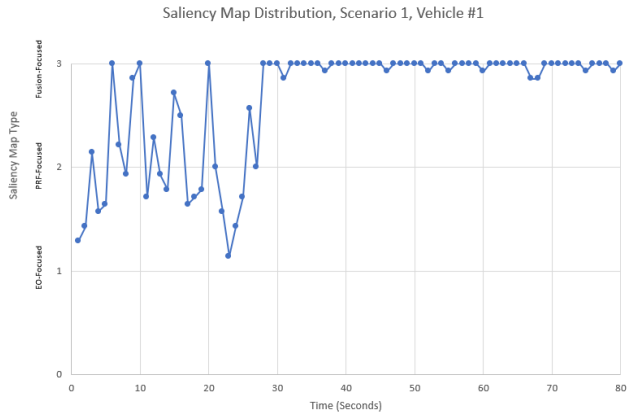

Figure 5: Scenario 1 Distribution of Saliency Maps for Vehicle #1

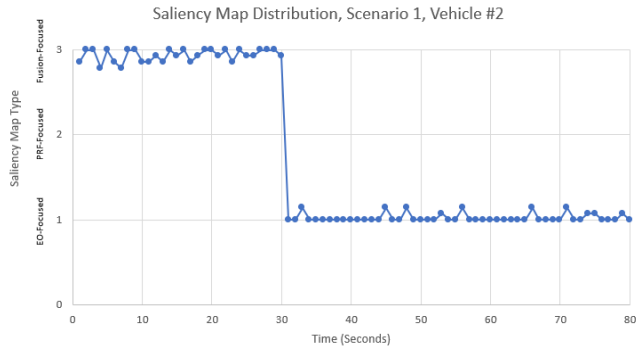

Figure 6: Scenario 1 Distribution of Saliency Maps with for Vehicle #2

During previous research with a greedy algorithm implementation, it was possible to determine that the EO input for vehicle #2 was relied upon more than it was for vehicle #1, which is now more intuitively explained with the saliency map results. With the visualizations we can confirm that vehicle #2, which spends more time in view of the scenario, does in fact have more EO focused activations than vehicle #1 does, and owing to the fact that vehicle #1 is not seen by the EO as much, the distribution of weights not being as reliant on the EO focused activations is something that is both supported by the distribution of saliency map activations and is an intuitive conclusion from the available data.

There is some confusion of vehicle #2, as there are brief instances where the focus shifts onto the true vehicle #1.

Occurring because the vehicle very briefly does become visible, as well as some outliers in which the model appears to mistake vehicle #2 for vehicle #1. The fusion oriented view occurs for the activations for tracking vehicle #1 and vehicle #2 whenever the two are not in sight. As the two are visually similar even after the application of DOF. As such, from the results and from the distribution of weights, we can infer on how the model relies on the P-RF data with the large number of fusion focused heatmaps as there is a visually similar vehicle but also indications that the target is inside the garage and not outside from the P-RF features.

B. Scenario 2 Inferences

As seen with scenario 2, the addition of a third vehicle, drastically changed some of the distributions of the types of saliency map activations. The distributions of figures 8 and 9 differ drastically, from which we can infer that due to the differences in the appearance on the overlay, the "red herring" that is vehicle #1 in scenario 2, is almost exclusively tracked using the EO aspect of the overlay. The data also indicates that there are moments, similar to scenario 1, in which the two visually similar vehicles cause some confusion and force the model to focus on the fusion view of the overlay as opposed to relying on the EO aspects of the overlay. While there was considerable number of EO-focused visualizations, for the most part the vehicle #2 saliency maps were fusion-focused.

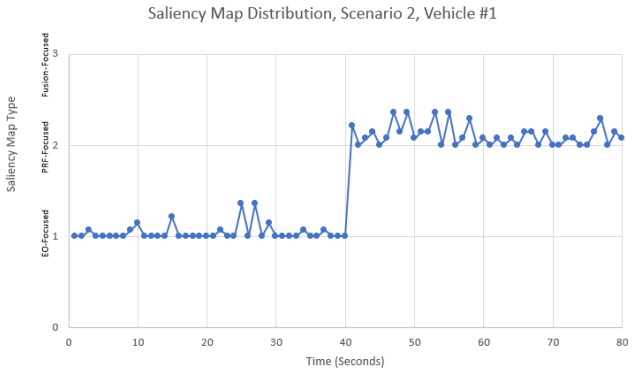

Figure 7: Scenario 2 Distribution of Saliency Maps for Vehicle #1

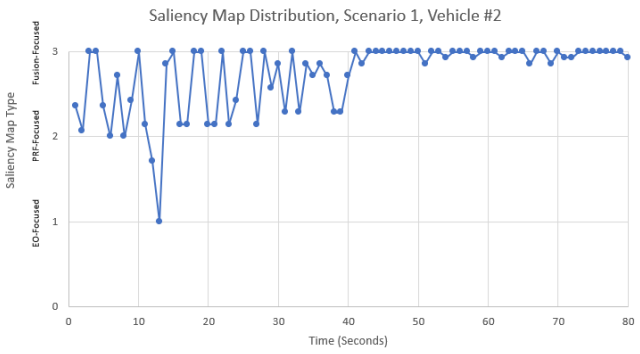

Figure 8: Scenario 2 Distribution of Saliency Maps with for Vehicle #2

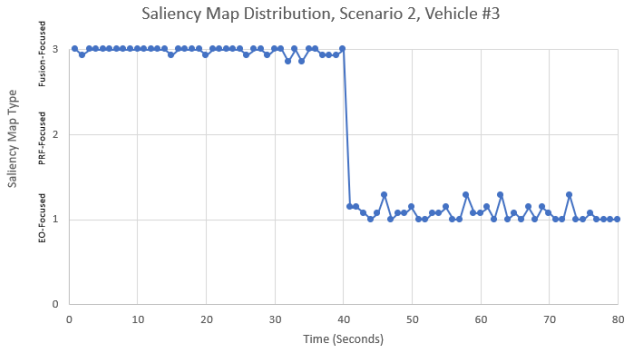

Figure 9: Scenario 2 Distribution of Saliency Maps with for Vehicle #3

C. Scenario 3 Inferences

As seen in figure 10, the distribution for scenario 3, vehicle #1 is predominantly focused on the EO focused saliency maps. The distribution of the P-RF to EO weights in previous research involving this vehicle in scenario 3 showed a distinctly higher dependency on the EO input, and that is reflected in the following figure. As vehicle #1 spends most of its time outside it is not surprising that the saliency map overlays produced are primarily EO based. There are a lot of P-RF and Fusion focused outputs once the vehicle enters the garage, with a focus on a Fusion saliency maps predominantly.

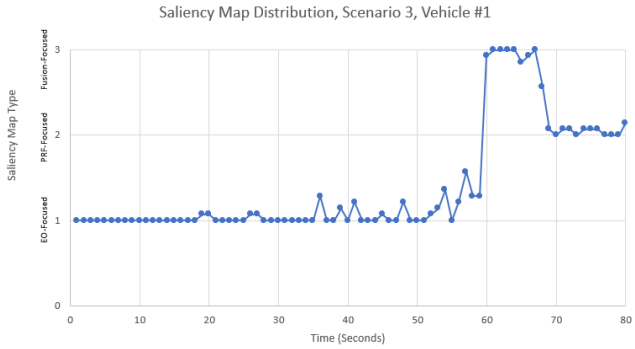

Figure 10: Scenario 3 Distribution of Saliency Maps with for Vehicle #1

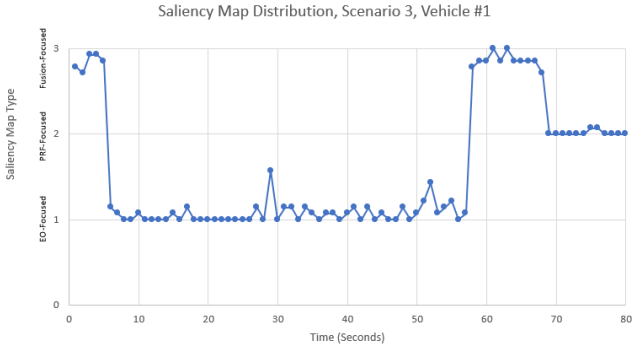

Figure 11: Scenario 3 Distribution of Saliency Maps with for Vehicle #2

Vehicle #2 does not provide any particularly interesting results, other than some scattered aspects that focus on P-RF and Fusion views. These come in brief frame instances when the vehicles are swapping but the majority of which are outliers.

One thing that remains constant between the distributions for scenario 3 is that the distributions tend towards P-RF differences once vehicle #4 enters the garage, most of the methods default from fused view to P-RF view.

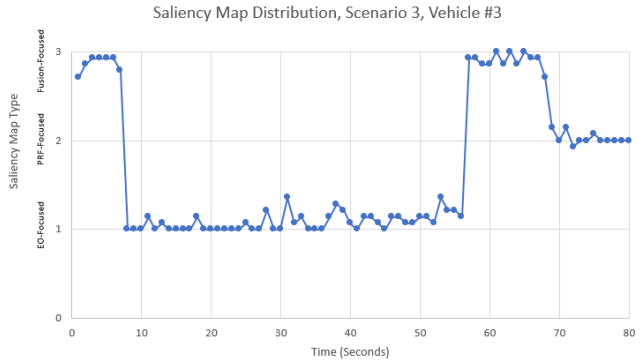

Figure 12: Scenario 3 Distribution of Saliency Maps with for Vehicle #3

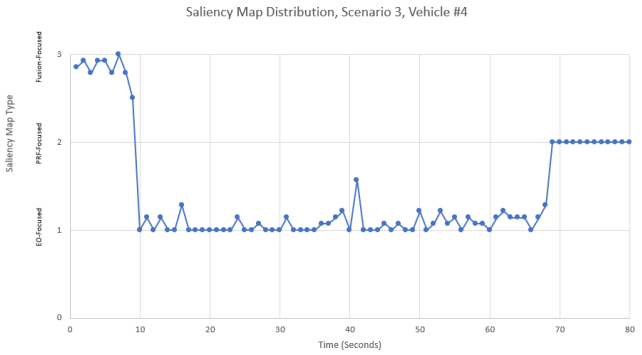

Figure 13: Scenario 3 Distribution of Saliency Maps with for Vehicle #4

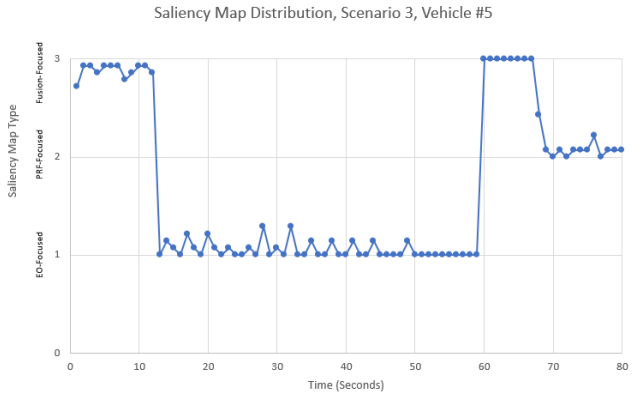

Figure 14: Scenario 3 Distribution of Saliency Maps with for Vehicle #5

While the model was able to achieve an F1 score of 0.9, it should be noted as mentioned earlier that the classification of the fusion, EO, and P-RF focused heatmaps is subjective in nature and was accomplished manually. That being said, the usage of the visualizations appears to demonstrate an intuitive pattern for the utilization of the overlays provided. This, in turn, provides for us inferences for how the fusion model might prioritize or weigh different sources of information in the tracking application. Combined with the knowledge of the scenarios, when certain targets are visible to the sensor, and when similar targets are also in sight, this provides some inferences for how the model operates.

It appears the fusion model utilizes the P-RF data sparingly, as after training the EO data is clearly the more reliable source of information. While DOF does occasionally pick up on very minor aspects of the EO data, such as trees moving, the clear indicator of a moving vehicle is much easier to pick up on each of the generated frames than the changes in the histogram. Once the situation becomes more complicated, the model tends towards relying on the fusion view. We can also infer that in the absence of reliable EO indicators of the vehicle's presence, seen prevalently in scenario 2, that it then defaulting entirely to P-RF inputs. With occasional instances of attempting to discern if the vehicle is moving behind the treeline. in the case of the three scenarios chosen.

V. Conclusion

In this paper, we present our research using visualization methods to provide intuitive sources to infer fusion model behavior. Keeping in mind the context for these visualizations, with respect to decision making and the use of features from different modalities. Normally, the processing of I/Q data in the form of histograms may not provide the obvious inferences when used with the corresponding DOF EO input. However, from the saliency maps and neuron activation heat maps, it can be inferred that the model still relies on the P-RF data depending on the available information. As these insights are less quantitative and more intuitive, the use of a visualization technique combined with expert knowledge of the training set provided information that helps to better understand the relationship between the two sources of data in the fusion model. While the P-RF histogram data may not be as useful as the DOF-EO image inputs, can be decisively used when the DOF-EO image inputs, which have by and far been shown to be a preference for the fusion model, are rendered less effective. The primary contribution of this paper is the combination of XAI visualizations to provide inferences on the P-RF/EO fusion model. Using the frequency in which these different types of visualizations appear to better understand the model as a whole.

In future research, it is our goal to potentially implement event-based processing techniques. In order to assess and evaluate the effectiveness of the XAI results more objectively rather than rely only on post-hoc visualization. For the purposes of tracking, it would be beneficial to implement methods that can determine the selective relevance of input features. Another aspect of the evaluation that might be improved is the usage of an independent means of grouping visualizations. Such as crowd voting, to enable a less subjective labeling process for the saliency maps. Artificial training with a neural network approach is unlikely to be as beneficial given the temporal relationship between the different points in the scenario, and for that reason the other next logical step is returning to a model that can exploit the temporal aspect and implementing the visualization techniques there.

Acknowledgment

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

VI. References

[1] F. K. Došilović, M. Brčić and N. Hlupić, "Explainable artificial intelligence: A survey," *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO,* pp. 0210-0215, 2018.

[2] 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.

[3] 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.

[4] S.-M. Moosavi-Dezfooli, A. Fawzi and P. Frossard, "DeepFool: a simple and accurate method to fool deep neural networks.," *Seyed-Mohsen,* pp. 2574-2582, 2016.

[5] E. Tjoa and C. Guan, "A Survey on Explainable Artificial Intelligence (XAI): Towards Medical XAI," *IEEE transactions on neural networks and learning systems,* pp. 1-21, 2020.

[6] Z. Lipton, "The Mythos of Model Interpretability: In Machine Learning, the Concept of Interpretability is Both Important and Slippery," *Association for Computing Machinery,* vol. 16, no. 3, pp. 1542-7730, 2018.

[7] K. Främling, M. Westberg, J. M. Madhikermi and M. , "Comparison of Contextual Importance and Utility with LIME and Shapley Values," in *Lecture Notes in Computer Science*, 2021.

[8] A. Barredo Arrieta, N. Diaz Rodriguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado González, S. Garcia, S. Gil-Lopez, D. Molina, V. R. Benjamins, R. Chatila and F. Herrera, "Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI," *Information Fusion,* vol. 58, pp. 82-115, 2019.

[9] A. Vakil, "Heterogeneous Multimodal Sensor Fusion Via Canonical Correlation Analysis and Explainable AI," Oakland University, Rochester, Michigan, 2020.

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

[11] J. Yosinski, J. Clune, A. Nguyen, T. Fuchs and H. Lipson, "Understanding Neural Networks Through Deep Visualization," *ArXiv,* p. abs/1506.06579, 2015.

[12] J. Han, E. J. Pauwels and P. d. Zeeuw, "Fast saliency-aware multi-modality image fusion," *Neurocomputing,* vol. 111, pp. 70-80, 2013.

[13] P. Linardatos, V. Papastefanopoulos and S. Kotsiantis, "Explainable AI: A Review of Machine Learning Interpretability Methods," *Entropy,* vol. 23, no. 18, p. 1, 2020.

[14] H. A. Khan, M. M. Khan, K. Khurshid and J. Chanussot, "Saliency based visualization of hyper-spectral images," in *IEEE International Symposium on Geoscience and Remote Sensing (IGARSS)*, Milan, Italy, 2015.