Naval Research Laboratory
Washington, DC 20375-5320

NRL/MR/5559--20-10,203

Multimodal Sensor Fusion with Deep Learning

Kristen L. Nock

Transmission Technology Branch
Information Technology Division

Elizabeth A. Gilmour
Dr. Samuel N. Blisard

Navy Center for Applied Research in Artificial Intelligence
Information Technology Division

November 20, 2020

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
20-11-2020

2. REPORT TYPE
NRL Memorandum Report

3. DATES COVERED *(From - To)*
Oct 2018 – Sep 2020

4. TITLE AND SUBTITLE
Multimodal Sensor Fusion with Deep Learning

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
61153N

6. AUTHOR(S)
Kristen L. Nock, Elizabeth A. Gilmour, and Dr. Samuel N. Blisard

5d. PROJECT NUMBER

5e. TASK NUMBER
IT015-09-47-1J24

5f. WORK UNIT NUMBER
1J24

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Research Laboratory
4555 Overlook Avenue, SW
Washington, DC 20375-5320

8. PERFORMING ORGANIZATION REPORT NUMBER
NRL/MR/5559--20-10,203

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Office of Naval Research
One Liberty Center
875 North Randolph Street
Arlington VA 22203-1995

10. SPONSOR / MONITOR'S ACRONYM(S)
ONR

11. SPONSOR / MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

DISTRIBUTION STATEMENT A: Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This report documents the efforts over the three year 6.1 base program titled Multimodal Sensor Fusion with Deep Learning. Herein a novel framework for fusing hyperspectral imagery (HSI) and LiDAR data for urban land use and land cover classification is detailed. In this approach multimodal sensor fusion was enhanced by utilizing deep learning and fuzzy logic to amalgamate information between spatial and spectral domains. Deep learning techniques enabled exploitation of mid and high level correlations between contrasting domains that have weak correlation between low level representations. The introduction of fuzzy logic further improved sensor fusion by providing advantages with difficult samples and the opportunity to gain network explainability.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON / 19b. TELEPHONE NUMBER *(include area code)*
Unclassified Unlimited	Unclassified Unlimited	Unclassified Unlimited	Unclassified Unlimited	16	Kristen L. Nock / (202) 767-0444

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

This page intentionally left blank.

CONTENTS

This page
intentionally
left blank

EXECUTIVE SUMMARY

This report documents the results of research into enhancing multimodal sensor fusion by utilizing deep learning (DL) and fuzzy logic to amalgamate information between spatial and spectral domains. Overall, this approach enriched information gain by fusion of disparate sensor data, which positively affects intelligence collection, data transmission, and the visualization of remotely sensed information. The overall approach was to implement DL architectures for concurrent multimodal sensor data leveraging state-of-the-art data fusion datasets, and then expand these DL capabilities by integrating fuzzy logic and fuzzy aggregation to extend the scope of ingestible information. The several advancements made in this research include:

- Implementing DL models into system-on-chip (SoC) hardware
- DL on hyperspectral imagery (HSI) data
    1. DL on HSI to obtain water properties and bottom depths
    2. Using open-set recognition approaches on HSI
- Ablation study of fusion methods within the framework
- Novel framework for multimodal sensor fusion of HSI and LiDAR using DL and fuzzy aggregation
- Exploring the role and usefulness of neuro-fuzzy logic in the context of automatically reasoning under uncertainty about complex scenes in remotely sensed data

The publications, [1, 2, 3, 4, 5], provide further detail on the advances made.

This page
intentionally
left blank

MULTIMODAL SENSOR FUSION WITH DEEP LEARNING

1. OBJECTIVE

We propose to enhance multimodal sensor fusion by utilizing deep learning techniques that enable exploitation of mid and high level correlations between sensing modalities. Our techniques will extend sensor fusion to include both spatial and spectral data improving classification and enriching the information obtained from these sensors.

2. BACKGROUND AND MOTIVATION

The underlying philosophy of sensor fusion is the idea that no single sensor can provide all the relevant information required to make informed decisions. There are many naval applications, including surveillance and mission planning, which benefit from having data from multiple modalities analyzed. However, the process for analyzing data from multiple modalities is manually intensive, often requiring specialized analysts to build relationships between information generated from each modality. We propose a method which will assist analysts by automatically fusing information using deep learning techniques. These techniques will enable fusion between multiple sensing modalities including spectral and spatial information, and will simultaneously increase the reliability of information generated and reduce the amount of time required to extract such information.

Early development in sensor fusion enhance the reliability of detection systems through the application of Bayesian statistics to distributed sensors of the same type. Measurement from each sensor are processed independently and combined by weighing the detection results according to the reliability of each sensor [6]. However, data generated for multiple sensors is often not independent, thus these methods do not optimally exploit the collected data for maximum information exfiltration.

Recent research is begging to address this issue and is demonstrating how sensor fusion at the data level can further improve detection and classification [7]. These techniques exploit low-level features such as discontinuities due to edges, which are present in both HSI and LiDAR data, allowing the two modalities to be aligned. Once aligned, each modality is processed through independent classifiers and then fused together. While this method does allow for center fusion between multiple modalities it once again does not fully exploit all possible data correlations, limiting the information and intelligence that could be gathered.

Deep learning techniques allow for improved sensor fusion by extracting correlations between multiple modalities. These algorithms allow for data to be processed for multiple layers of abstraction and can bring information from disparate sensors to a common latent space. Here information from multiple sources can be fused together and then jointly exploited. Research into sensor fusion using deep learning techniques has demonstrated fusion between audio and imagery representing temporal and spatial domains respectively [8]. There is little to no correlation of the data at a low level; however, higher-level features show stronger

Manuscript approved November 18, 2020.

correlations. These higher-level correlations enable improved approaches to handling sensor fusion and classification [9]. In addition to improving classification result, the higher-level information extracted from deep learning methods can move the sensor data into an embedded space more readily understood by a human [10]. This process allows for faster retrieval of relevant data, reducing time and manual processing requirements needed to collect and analyze sensor data.

For this project, we proposed to develop sensor fusion techniques between spectral and spatial information obtained from HSI and LiDAR respectively. Our resulting approach uses deep learning techniques to allow mid and high-level sensor fusion between sensors enhancing the information content derived from collected data. This information can be readily used by an analyst to quickly and reliably make informed decisions. Building off of the previous 6.1 base effort, "Foundations for Complex Geospatial Uncertainty", our proposed approach also includes the use of fuzzy logic to enable the use of subjective or evidential information with objective data.

3. TECHNICAL APPROACH

The project is a collaborative effort between Information Technology and Marine Geosciences. These divisions provided expertise in deep learning algorithms, geospatial uncertainty, and sensor analysis. This strong collaboration enabled the development of deep learning algorithms which handle real sensor data, integrate uncertainty metrics and include qualitative information.

Our research required development in four major stages as described below. Research into the first three stages, data preprocessing, deep learning development, and the incorporation of fuzzy logic, began in parallel. The final stage, integration and validation, required a joint effort from all participants.

3.1 Data Preprocessing

We leveraged previously collected data containing both HSI and LiDAR data. Multimodal datasets can be extremely complex and benefit from pre-processing. For LiDAR data, this involves generating digital elevation maps from the point files which can be used as inputs to the deep learning algorithms. Additionally, due to the presence of large number of bands in the hyperspectral data, dimensionality reduction has become an important aspect in generating an information dense signal.

3.2 Deep Learning Development

In addition to generating the input, the team needed to explore approaches to feature extraction. The availability of several modalities reporting on the same phenomena introduces new degrees of freedom. State of the art single-sensor based classifiers were tested to give baseline results. In this way, any advancements made available through fusion networks, and networks using fuzzy logic, can be explored.

We developed deep learning algorithms to bring hyperspectral information and spatial information from the LiDAR data to a common latent space. The space was fused and further exploited through deeper layers for classification purposes. We also conducted research in architecture formation and training approaches independently from the data pre-processing stage.

3.3 Incorporation of Fuzzy Logic

The quality of the content within a given scene is a subjective measurement. By incorporating fuzzy logic into the research effort it is possible to use qualitative information to make a quantitative assessment. Further, the sensors also can provide ambiguous information due to granularity in the information or missing data. Use of fuzzy sets, belief structures, and fuzzy aggregation operations enables the mathematical modeling of these types of uncertainties and fusion of these data together.

3.4 Integration and Validation

Research was conducted into the feasibility of an integrated system. Such a system will at a minimum require preprocessed data, developed deep learning architectures, and the incorporation of subjective information through fuzzy logic. All members worked together to improve the understanding of the system strengths and limitations, quantify the uncertainty, and develop methods for potential transitions.

4. RESULTS

4.1 Accomplishments each Fiscal Year (FY)

4.1.1 FY18

In this year, multiple software packages were evaluated for their ability to train and deploy multimodal deep learning architectures, as well as their customizability for modifying layer types and integrating into fuzzy logic systems. We successfully trained single modality networks for handling imagery and separately hyperspectral datasets.

An opportunity for collaboration with the Remote Sensing Division arose. Our collaborative efforts began regarding their large hyperspectral datasets that can be used not only for classification purposes, but also estimating physical environmental parameters including ocean depths and chemical composition of the water column. This difficult dataset enabled us to master handling HSI data, it also has direct transitional path for our research.

Further, we implemented deep learning into the mobile hardware shown in Figure A1. This was a valuable effort because the availability of onboard processing can lead to a reduction in data transmission, storage, and ultimately enhance information gain from autonomous systems. Specifically we successfully developed deep learning applications which can run on Android hardware, which includes many tablets and cell phones. We have also been successful in integrating models into system-on-chip (SoC) hardware which allows for the network outputs to run quickly and efficiently on customizable mobile hardware.

4.1.2 FY19

In this year, through our collaboration with the Remote Sensing Division, we published our research into the use of deep learning architectures to parameterize the water column in terms of depth, bottom type, and inherent optical properties (IOPs). In addition, we sought to classify depth categories of optically shallow (visible bottom) or optically deep waters. Overall, deep learning architectures performed best in optically shallow cases. This is true in deep/shallow classification, in bottom type classification of 114 types, depth regression, and in estimating the IOP levels of chlorophyll, colored dissolved organic matter, and total suspended sediments. In contrast, when the water is deep and the spectra contains less rich information other machine learning approaches performed best in parameterizing the IOPs.

Other efforts this year included working to replicate the current state of the art deep learning architectures that have demonstrated multimodal fusion for text, audio, and video datasets. We focused our efforts to datasets with overlapping spatial and spectral data. Our group obtained the 2018 IEEE GRSS Data Fusion Challenge data [11] acquired over the University of Houston campus and the neighboring area, representing 20 urban land-cover/land-use classes. Data from this collection is shown in Figure A2. This dataset contained multispectral-LiDAR point cloud data at a 50-cm ground sample distance (GSD), and HSI data covering a 380-1050 nm spectral range with 48 bands at a 1-m GSD.

4.1.3 FY20

In FY20, we incorporated fuzzy logic into our deep learning networks. We published our novel framework for multimodal sensor fusion of HSI and LiDAR [5]. In this publication, we test different fusion methods within the framework, demonstrating that sensor fusion improves classification accuracy over any single modality. This framework is shown in Figure A3. These fusion methods include established aggregation methods such as combination rules, unified neural networks, and fuzzy aggregation. Our methods of fuzzy aggregation are shown in Figure A4. The benefits of fuzzy aggregation extend beyond classification accuracy. A major challenge in the implementation of machine learning is the need for trusted and explainable machine learning models. Fuzzy aggregation provides a learned and transparent method of aggregation. This learned method can range from union, intersection, average, or other unique aggregation. Accordingly, this was used to explain the quality of the individual sources and their interaction characteristics. Additionally, because we understand how the network elements map to the learned fuzzy aggregation we can understand, validate, and do iterative development.

We also published our research on the role of possibilistic clustering to generate data point typicality degrees to improve challenges in neuro-fuzzy logic learning tools [3, 4]. Neuro-fuzzy is the application of fuzzy logic to neural networks. These fuzzy logic neurons/networks (FLN) hold the potential to help realize more explainable, interpretable, and ultimately trustworthy AI. There are a number of ways in which this can occur. For one, it is possible to insert human knowledge as rules into a FLN. Furthermore, a FLN can be derived from data then opened to study what variables, rules, and output combination strategies were learned. A visual representation of this can be found in Figure A5.

We also implemented a novel application of open-set recognition based on Objectosphere loss [12] to HSI. The goal of Objectosphere loss is to perform novel class detection and to classify the known samples with an accuracy similar to that of other models. It does this by pushing the learned feature vectors of novel samples towards the origin of the rendered spatial representation. Performing novelty detection becomes possible by assuming that all samples within a certain radius in the deep feature space are ‘unknowns.’ This is shown in Figure A6. Applying open-set recognition to HSI solves the real-world scenario of samples existing in the testing dataset that do not belong to classes the model was trained to classify. The trained model may need to used to classify data that contains anomalies that do not belong to any of the training classes.

Finally, we adapted a state of the art model for image segmentation, U-Net [14], for both our GRSS dataset and our other AeroRIT dataset [13]. The AeroRIT dataset is a large-scale aerial hyperspectral scene with pixel-wise annotations overlooking Rochester Institute, and in shown in Figure A7. In contrast to our initial approach, this application of segmentation (assigning class labels to specific pixels, and finding the boundaries of the objects within the image) allows us to build a better understanding regarding the regions contained throughout the captured scene. As a result of this research, we were able to add scene understanding and object identification via the task of dense semantic segmentation to our final suite of codes available for any future effort.

5. ASSOCIATIONS AND OUTPUTS

5.1 Associated Base Program Projects

Work developed for "Recursive Structured Learning" (6.1 IT) was leveraged for our design and implementation of deep learning architectures for single sensor data. This aided in our development of a baseline for comparison for each sensing modality independently against our fused architectures. Alternatively, our fused architectures focused on exploiting high-level correlations between the contrasting spectral and spatial domains.

Our integration of fuzzy logic builds off "Foundations for Complex Geospatial Uncertainty" (6.1 IT) which developed techniques for fuzzy aggregation between two sensors. Further, this research extends the techniques into the deep learning architectures via the extension principal. Fuzzy logic was brought into the physical layers of the network allowing fuzzy logic to be learned within the architecture. In our research, a data driven approach was taken and data containing epistemic uncertainty.

5.2 Publications

To date, the research from this effort has resulted in three conference papers [1, 2, 3], two journal publications [4, 5], and three posters. The publications detail our research successes in the following main categories:

- Deep learning on HSI data to obtain water properties and bottom depths
- Neuro-fuzzy logic
- Fuzzy aggregation for multimodal remote sensing classification

Two of our posters were delivered to Annual Workshops Naval Applications of Machine Learning (NAML), and the third was shown at International Conference on Fuzzy Systems (FUZZ-IEEE).

At NAML2019, the poster titled "Deep Learning on Hyperspectral Data to Obtain Water Properties and Bottom Depths" was presented. At NAML2020, we presented a poster titled "Open-set Recognition for Land Use Classification." Further, at FUZZ-IEEE2019 we presented a poster titled "An Approach to Multimodal Sensor Fusion Using Fuzzy Logic."

5.3 Transitions

Research from this 6.1 base project regarding the handling of HSI data was integrated into the successful NISE 6.1 program named Making Unlabeled Navy Data Useful: An Example with Hyperspectral/Multispectral Data.

Our research of fuzzy systems to model human reasoning and to provide a mechanism for biasing the system with feedback from an analyst from this project has been leveraged into a future base program currently under review. This proposal for Human Informed Machine Learning (HIML) aims to develop encodings and a general framework to inject human understanding, knowledge, and intuition into machine learning models.

5.4 Patents

Research from this project resulted in a method for fusion in mass spectral identification and quantitation of sea life (e.g. bacteria, plankton, algae). This method has resulted in a patent (currently under review).

ACKNOWLEDGMENTS

The authors wish to thank the NRL Research Advisory Committee for supporting this Base Research Program. Additionally, we would also like to thank Derek T. Anderson and other colleages at the Mizzou INformation and Data FUsion Laboratory (MINDFUL) at the University of Missouri for their help in the development, understanding and implementing the fuzzy systems used in our reseach efforts.

Bibliography

[1] David Bonanno ; Kristen Nock ; Leslie Smith ; Paul Elmore ; Fred Petry; “An approach to explainable deep learning using fuzzy inference,” Proc. SPIE 10207, Next-Generation Analyst V, 102070D (May 3, 2017); doi:10.1117/12.2268001

[2] Kristen Nock, Elizabeth Gilmour, Paul Elmore, Eric Leadbetter, Nina Sweeney, Frederick Petry, ”Deep learning on hyperspectral data to obtain water properties and bottom depths,” Proc. SPIE 11018, Signal Processing, Sensor/Information Fusion, and Target Recognition XXVIII, 110180Y (7 May 2019);

[3] Blake Ruprecht, Charlie Veal, Al Cannaday, Derek T. Anderson, Fred Petry, James Keller, Grant Scott, Curt Davis, Charles Norsworthy, Paul Elmore, Kristen Nock, Elizabeth Gilmour, ”Neuro-fuzzy logic for parts-based reasoning about complex scenes in remotely sensed data,” Proc. SPIE 11423, Signal Processing, Sensor/Information Fusion, and Target Recognition XXIX, 114230H (22 April 2020); https://doi.org/10.1117/12.2558094

[4] Blake Ruprecht, Wenlong Wu, Muhammad Aminul Islam, Derek Anderson, James Keller, Grant Scott, Curt Davis, Fred Petry, Paul Elmore, Kristen Nock, Elizabeth Gilmour, ”Possibilistic Clustering Enabled Neuro Fuzzy Logic,” 2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Glasgow, United Kingdom, 2020, pp. 1-8, doi: 10.1109/FUZZ48607.2020.9177593.

[5] Kristen Nock and Elizabeth Gilmour, ”Fuzzy aggregation for multimodal remote sensing classification,” 2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Glasgow, United Kingdom, 2020, pp. 1-7, doi: 10.1109/FUZZ48607.2020.9177691.

[6] Z. Chair, P. Varshney, “Optimal Data Fusion in Multiple Sensor Detection Systems,” IEEE Transaction in Aerospace and Electronics Systems, 644-653, September, 1987.

[7] T. Khuon, R. Rand, J. Greer, E. Truslow, “Distributed Adaptive Spectral and Spatial Sensor Fusion for Super-resolution Classification,” IEEE AIPR, Washington DC, October, 2012.

[8] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee and A. Y. Ng, “Multimodal deep learning.,” in ICML, Lise Getoor and Tobias Scheffer, Eds. 2011 pp-689-696, Omnipress.

[9] A. Merentitis and C. Debes, ”Automatic fusion and classification using random forests and features extracted with deep learning,” 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Milan, 2015, pp. 2943-2946.

[10] Wei Wan, Xiaoyan Yang, Beng Chin Ooi, Dongxiang Zhang, and Yueting Zhuang. 2016. Effective deep learning-based multi-modal retrieval. The VLDB Journal 25, 1

[11] Debes et al., 2014] Debes, C. ; Merentitis, A. ; Heremans, R. ; Hahn, J. ; Frangiadakis, N. ; van Kasteren, T. ; Liao, W. ; Bellens, R. ; Pizurica, A. ; Gautama, S. ; Philips, W. ; Prasad, S. ; Du, Q. ; Pacifici, F. : Hyperspectral and LiDAR Data Fusion: Outcome of the 2013 GRSS Data Fusion Contest. IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., 7 (6) pp. 2405-2418, June 2014.

[12] Dhamija, Akshay Raj, Manuel Günther, and Terrance Boult. ”Reducing network agnostophobia.” Advances in Neural Information Processing Systems. 2018.

[13] Rangnekar, A., Mokashi, N., Ientilucci, E. J., Kanan, C., and Hoffman, M. J. (2020). Aerorit: A new scene for hyperspectral image analysis. IEEE Transactions on Geoscience and Remote Sensing.

[14] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. 2015, pp. 234–241.

This page
intentionally
left blank

Appendix A

APPENDIX OF FIGURES

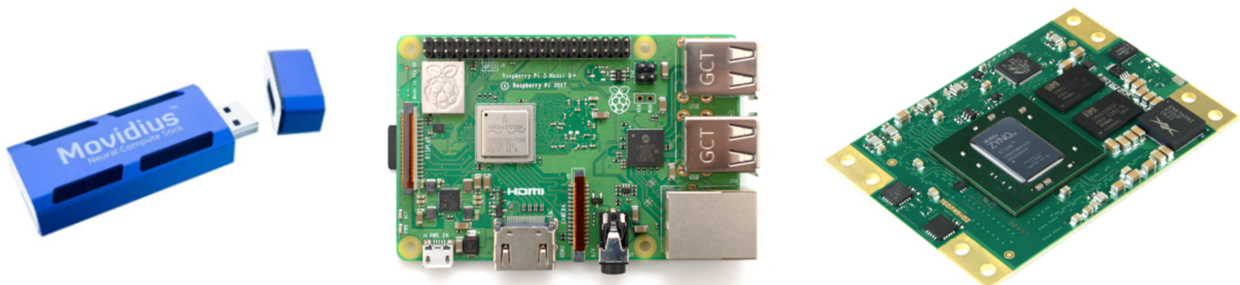

Fig. A1—The hardware we applied our deep learning models on: Neural Compute Stick, ARMv7 (Rasperry Pi), and Xilinx Zynq 7045 (SoC) FPGA respectively.

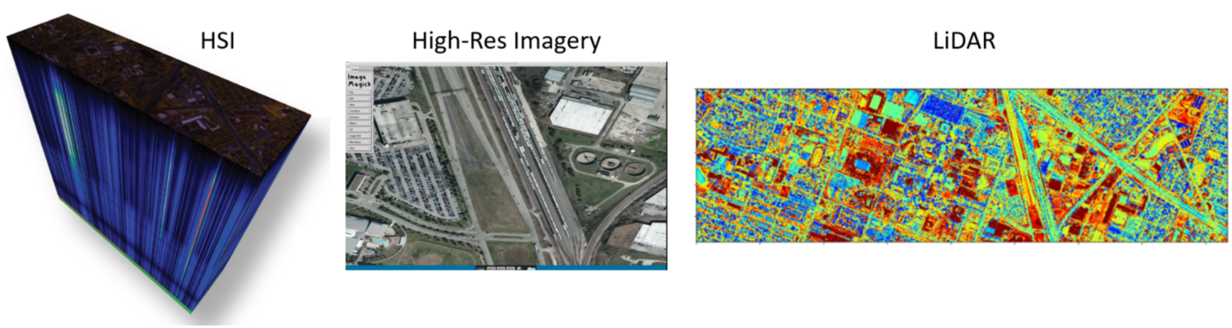

Fig. A2—The Geoscience and Remote Sensing Society (GRSS) provided co-registered HSI and LiDAR that was collected over the University of Houston campus. The dataset, grss_dfc_2018, includes co-registered LiDAR and HSI, as well as high-resolution imagery. The dense dataset has 20 classes related to urban objects and land use.

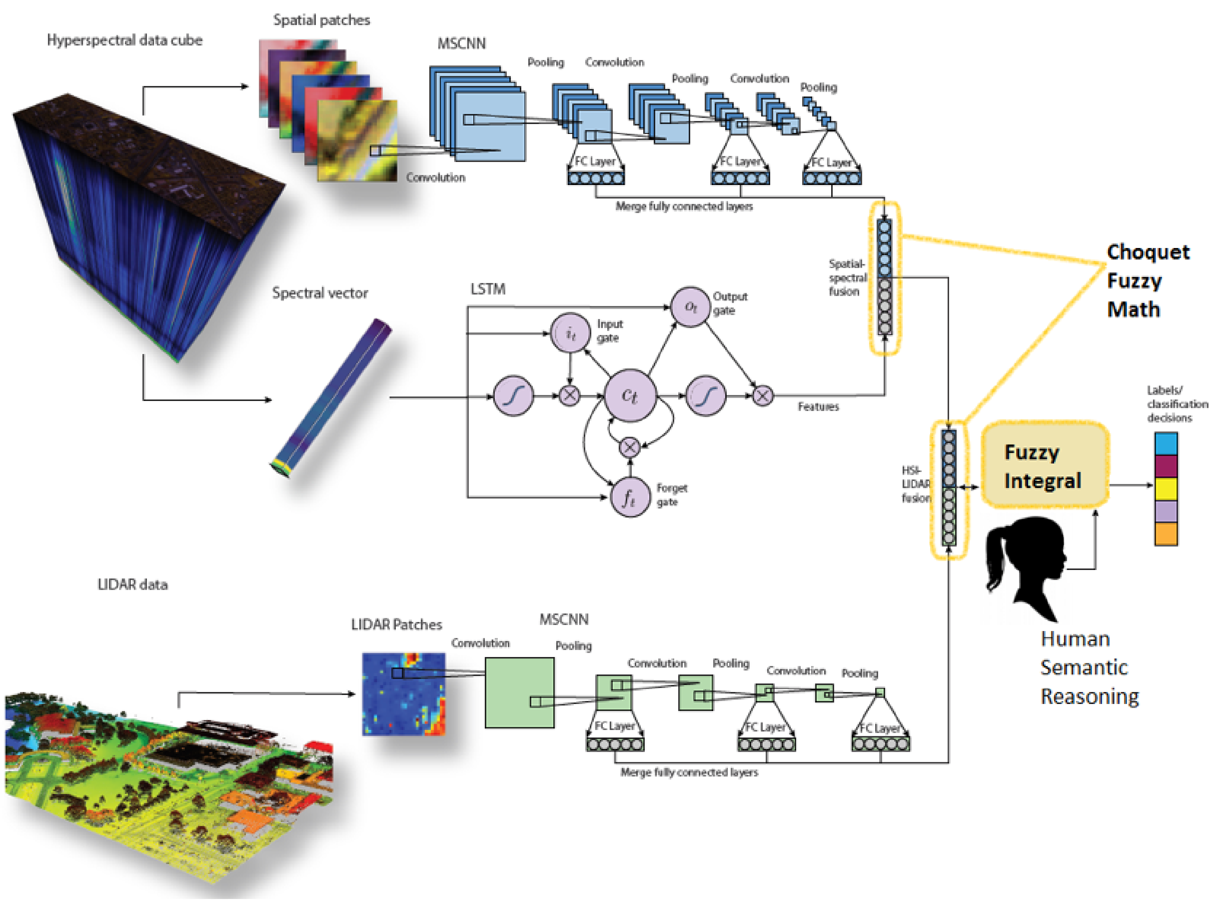

Fig. A3—Diagram of our novel framework for multimodal sensor fusion. Three separate neural network feeds are fused to make predictions. The spatial feature network (top) is multi-scale convolutional neural network (MSCNN) that is trained on nine pixel by nine pixel patches of HSI. The spectral feature network (middle) is an LSTM that is trained on single pixel vectors of HSI. The LiDAR network (bottom) is a CNN that is trained on nine pixel by nine pixel patches of LiDAR. These three neural networks are combined using different fusion methods to predict the class of training samples.

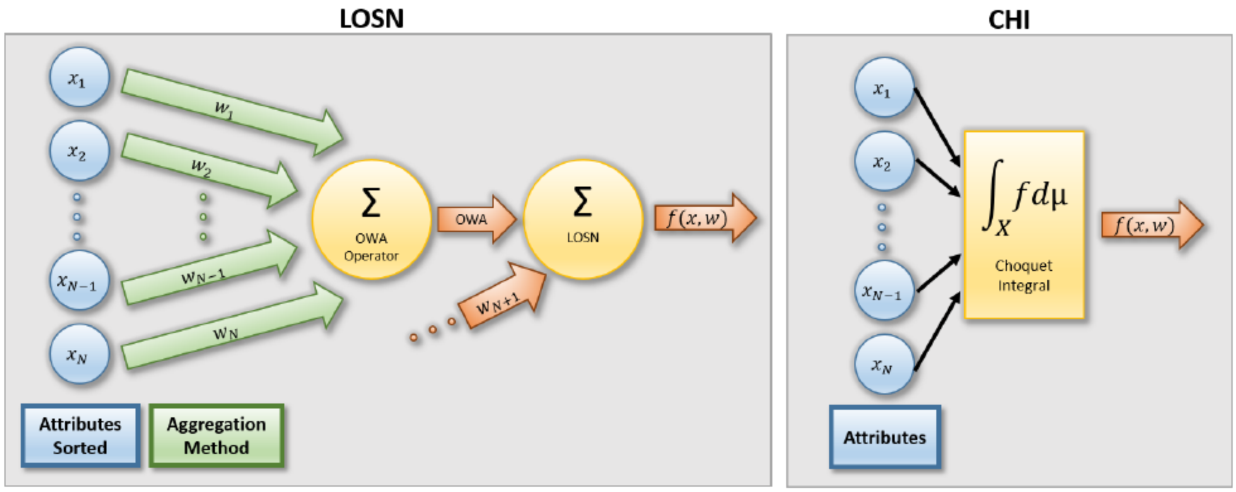

Fig. A4—This diagram shows the fuzzy aggregation operators used. The linear order statistic neuron (LOSN) is a generalization of the ordered weighted average (OWA) operator. Mathematically, the LOSN is the sum of a bias, and dot product of a sorted input and weights selected based on a aggregation method. The Choquet integral (ChI) acts as a generalized expectation operator. There will be N number of non-monotonic fuzzy measures μ used in this parametric non-linear aggregation function.

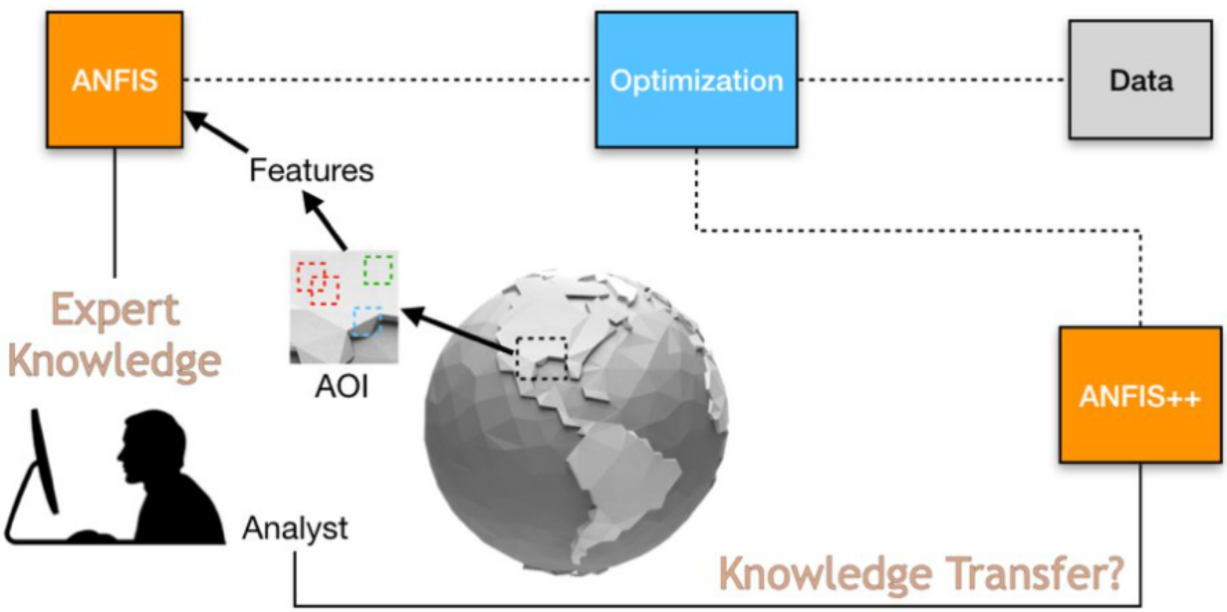

Fig. A5—Our research involved exploring the role and usefulness of neuro-fuzzy logic in the context of automatically reasoning under uncertainty about complex scenes in remotely sensed data. This diagram is a high-level illustration is adaptive neuro-fuzzy inference system (ANFIS). First, expert knowledge is transferred into an adaptive neuro-fuzzy inference system (ANFIS) for sake of automating some process, e.g., object detection or land classification in remote sensing. Next, data is used and the solution is optimized to produce an augmented ANFIS, "ANFIS++". The ANFIS++ is used in place of the expert and it is analyzed to determine differences for the sake of discovering new domain specific logic that might be of interest to the expert and/or analyzed for validation of the machine learned model.

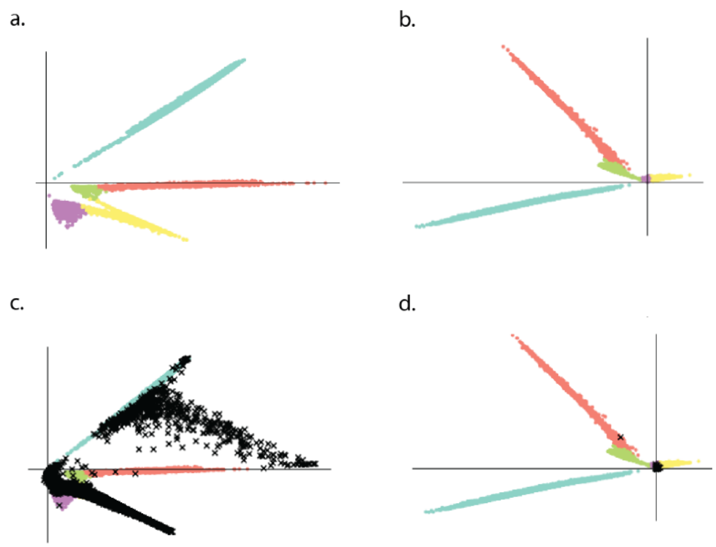

Fig. A6—This figure is a visual representation of using Objectosphere loss to identify novel classes in an HSI dataset. In all subplots, 'known' classes are plotted as a different color while the novel/unknown classes are plotted in black. On the left, subplots a and c, standard cross entropy loss is used. On the right, subplots b and d, Objectsphere loss is used. Subplots c and d clearly show that when the unknowns are plotted into the deep feature space, the model trained with the Objectosphere loss are clustered at the origin, and can easily be separated from the knowns.

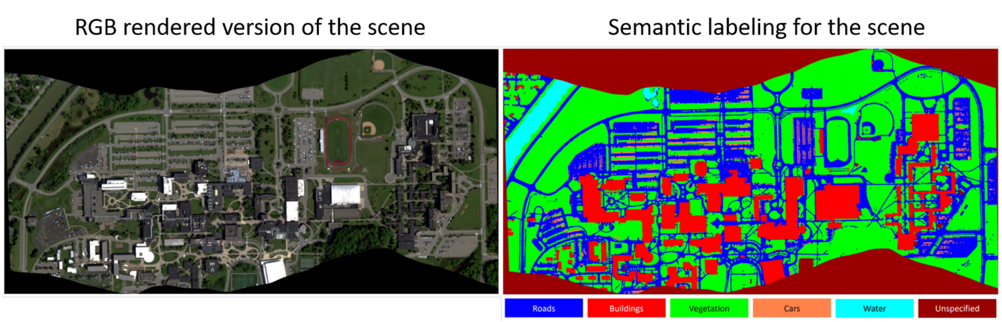

Fig. A7—The AeroRIT scene overlooking Rochester Institute of Technology's university campus. The spatial resolution is 1973 × 3975 pixels and covers the spectral range of 397 nm - 1003 nm in 1 nm steps. We used this dataset for our state of the art image segmentation research.