ElecDaug: Electromagnetic Data Augmentation for Model Repair based on Metamorphic Relation

Jiawei He
State Key Laboratory for Novel Software Technology
Nanjing University, China

Zhida Bao
Harbin Engineering University
Harbin, China

Quanjun Zhang
State Key Laboratory for Novel Software Technology
Nanjing University, China

Weisong Sun
Jiawei Liu
State Key Laboratory for Novel Software Technology
Nanjing University, China

Chunrong Fang*
State Key Laboratory for Novel Software Technology
Nanjing University, China
fangchunrong@nju.edu.cn

Yun Lin*
Harbin Engineering University
Harbin, China
linyun@hrbeu.edu.cn

ABSTRACT

With the application of deep learning (DL) in signal detection, improving the robustness of classification models has received much attention, especially in automatic modulation classification (AMC) of electromagnetic signals. A large amount of electromagnetic signal data is required to obtain robust models in the training and testing process. However, the high cost of manual collection and the issue of low quality of automatically generated data contribute to the AMC model's defects. Therefore, it is essential to generate electromagnetic data by data augmentation. In this paper, we propose a novel electromagnetic data augmentation tool, namely ElecDaug, which directs the metamorphic process by electromagnetic signal characteristics to achieve automatic data augmentation. Based on electromagnetic data pre-processing, transmission or time-frequency domains characteristic metamorphic, ElecDaug can augment the data samples to build robust AMC models. Preliminary experiments show that ElecDaug can effectively augment available data samples for model repair. The video is at https://youtu.be/x5g6IVX_Q3s. Documentation and source code can be found here: https://github.com/ehhhhjw/tool_ElecDaug.git.

CCS CONCEPTS

• **Software and its engineering** → **Software maintenance tools**.

KEYWORDS

Model Repair, Data Augmentation, Metamorphic Relation, Automatic Modulation Classification

ACM Reference Format:

Jiawei He, Zhida Bao, Quanjun Zhang, Weisong Sun, Jiawei Liu, Chunrong Fang, and Yun Lin. 2022. ElecDaug: Electromagnetic Data Augmentation for Model Repair based on Metamorphic Relation. In *37th IEEE/ACM International Conference on Automated Software Engineering (ASE '22), October 10–14, 2022, Rochester, MI, USA*. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3551349.3559536

*Corresponding authors.

ASE '22, October 10–14, 2022, Rochester, MI, USA

ACM ISBN 978-1-4503-9475-8/22/10…$15.00
https://doi.org/10.1145/3551349.3559536

1 INTRODUCTION

Over the past few years, deep learning (DL) has been widely used in many security-critical application scenarios, such as self-driving cars and wireless communications [7]. When radar is used in driverless applications (e.g., Tesla), the onboard system will receive electromagnetic signals from GPS and perform automatic modulation classification (AMC) on them, the precision of which will directly relate to the vehicle security [4].

DL models with high robustness rely on a large amount of training data with diverse scenarios. However, existing electromagnetic data is hard to meet the needs of DL. The scarcity of data causes deficiencies in the boundary values, training data limitations, and sample diversity of DL-related AMC models, resulting in the trained models being prone to make erroneous predictions [3]. Existing electromagnetic data hardly supports DL in performing high-accuracy AMC models' training process as the high cost of manual collection and the low quality of automatic generation. Therefore, augmenting the existing electromagnetic data is a helpful solution to the issues caused by insufficient data and poor quality, which can repair the defects of AMC models in classification.

There are specific challenges in augmenting electromagnetic data. Since radio signals store information in electromagnetic waves, it is difficult for us to perceive the scenes expressed in the waves by human senses. Some existing electromagnetic data augmentation methods are performed by simulating the data generation process. For example, O'Shea et al. [6] augmented electromagnetic data by simulating transmitter parameters. The data augmented by this method does not consider the characteristics of the electromagnetic signal during the propagation of the physical channel, resulting in significant artifacts that cannot be used as training data.

In this paper, we propose a novel electromagnetic signal data augmentation tool, namely ElecDaug. By combining the characteristics of electromagnetic signals, we use metamorphic methods to pre-process the input data. A large amount of data with electromagnetic signal characteristics is augmented, compensating for the shortcoming caused by the lack of high-quality data while training electromagnetic correlation models. ElecDaug contains two components: the transmission metamorphic component and the time-frequency metamorphic component. The transmission metamorphic component takes into account external disturbances encountered during signal propagation (detailed in Section 3.1). The time-frequency metamorphic component considers the electromagnetic signal's transmission characteristics in the time and frequency domains (detailed in Section 3.2). We conduct a series of experiments to demonstrate the effectiveness and availability of ElecDaug and evaluate the classification effect of the AMC model trained with the augmented data. The experimental results show that data augmented by ElecDaug can repair the defects of the model, and the precision of the trained AMC model is improved by nearly 11% compared with the original one.

2 PRELIMINARIES

2.1 Metamorphic Relation

Metamorphic relations are central elements of metamorphic testing, which are necessary properties of the target function related to multiple inputs and their expected outputs [1]. Drawing on metamorphic testing, we use metamorphic relations to ensure that the augmented data is available. The metamorphic formula is as follows:

$$f_w(x) \neq f_w(x + r_x) \tag{1}$$

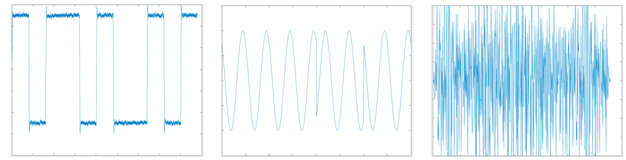
Where the disturb strength r_x should be small that $x + r_x$ from the unperturbed input data x remains the same. And the augmented data with perturbation is different from the original data, which has the perturbation obtained through the metamorphic relation.

The metamorphic relation utilizes an alternative mindset. When we cannot solve a problem from a particular perspective, we can corroborate it sideways by solving its equivalent. Due to the imperceptibility of electromagnetic signals, we cannot augment them intuitively like image processing methods; we do this by equivalently mapping their corresponding features to the data layer.

2.2 Signal Transfer Properties

During the transmission of the signal, conversion is usually performed by modulation to reduce external interference to the signal. As shown in Figure 1, the signal is transformed into different modulation types. We can not directly obtain valid information from the received wave. It is required to identify the signal's modulation type and demodulate the signal according to the modulation type. All the information is stored as plural numbers for modulated signals in this case. The signal $S_{echo}(t)$ received by the receiver is stored in the complex form of $I_{echo}(t)/Q_{echo}(t)$ channel data like: $S_{echo}(t) = \sigma I_{echo}(t) \cdot Q_{echo}(t)$. t denotes the time domain; σ is a constant and denotes the echo amplitude. $I_{echo}(t)$ and $Q_{echo}(t)$ store the time and frequency domain information of the signal, and both can be converted to each other.

In the signal propagation process, there is generally a specific loss in the received data, which makes the transmitter-generated data differ from the received data. To ensure the resemblance between the augmented and received data, we are required to model the signal loss due to external interference during transmission and the internal time-frequency domain. Besides, the internal time-frequency domain variation of the signal due to signal propagation also needs to be considered. We can simulate the signal variation in the channel and apply it to the data augmentation process, fixing the model's deficiencies due to the lack of sufficient high-quality data samples by deformation methods.



(a) BPSK (b) QPSK (c) QAM16

Figure 1: Different Modulation Signal

3 METHODOLOGY

In this section, we introduce the implementation of ElecDaug in detail. Figure 2 shows the workflow of ElecDaug, which consists of three key components: metamorphic methods, variable settings, and data augmentation. The metamorphic methods contain the transmission metamorphic components (Radio Noise, Signal Outage, Signal Disruptions) and the time-frequency metamorphic components (Channel Transformation, Power Zooming). After selecting the metamorphic methods, the label and parameters are set through the variable settings. The label of the augmented data can be modified, and the gap intensity also can be set. After that, we enter the data augmentation part, simulate the physical layer and data layer of the input data according to the selected metamorphic methods and the set parameters, augment the augmented data, and assign the label to the augmented data.

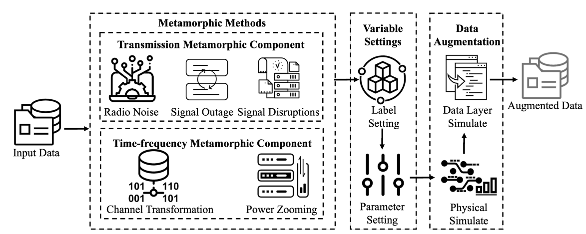


Figure 2: Overview of ElecDaug

3.1 Transmission Metamorphic Component

The transmission variation component generates scenario-specific data by simulating the external effects of electromagnetic signals during channel transmission. It contains three metamorphic operators: radio noise, signal outage, and signal disruptions.

Radio Noise. The radio noise metamorphic operator simulates mechanical, thermal, and Gaussian white noise in the signal transmission process by injecting radio noise into the electromagnetic signal. The metamorphic formula is: $I_o^{'}(t) = I_o(t) + \rho \cdot f(\mu_I, \sigma_I^2)$, $Q_o^{'}(t) = Q_o(t) + \rho \cdot f(\mu_Q, \sigma_Q^2)$. Among that, ρ is the perturbation intensity, μ is the expected value, σ^2 is the variance. By changing the intensity of the injected noise, it can generate electromagnetic signal data under different noise environments.

Signal Outage. The signal outage metamorphic operator simulates the scenario of partial data missing due to transient transmission interruption during electromagnetic signal transmission. We set the threshold a, and a numbers are randomly selected in range of $[0, n)$ to develop the set $A = \{A_1, A_2, ...A_a\}$. Set $I_i = 0$ and $Q_i = 0$, $i \epsilon A$, n is the channel length of signal.

Signal Disruptions. The signal disruption metamorphic operator simulates the signal delay during signal transmission by selecting a continuous segment of signal data and changing the data to the I/Q average of this continuous data. We set the threshold a ,b and c, c is the length of block in range of $[1, n]$, and a numbers are randomly selected in $[0, n-c)$ to develop the set $A = \{A_1, A_2, ...A_a\}$, b is a number in range of $[0, n - c)$. n is the channel length of the signal and set $I_i = (\sum_{j=b}^{b+c-1} I_j)/c$ and $Q_i = (\sum_{j=b}^{b+c-1} Q_j)/c$, $i \epsilon A$.

In the transmission metamorphic component, for these three methods, we select the optimal parameter as the interference threshold for ElecDaug by variable setting. By adjusting the interference threshold, we can obtain electromagnetic data with different perturbation intensities during data augmentation.

3.2 Time-frequency Metamorphic Component

The time-frequency metamorphic component contains two kinds of metamorphic operators based on electromagnetic signals' time and frequency domains. In addition, these features are converted to the data layer by the interference between the signals present in the electromagnetic signal. The time and frequency domain information is as follows.

$$S(\tau, t) = \sigma \cdot r(\frac{t}{T_a})e\{-j\frac{4\pi R(t)}{\lambda}\} \cdot r(\frac{\tau - \frac{2R(t)}{c}}{\tau_p})e\{j\pi b(\tau - \frac{2R(t)}{c})^2\} \quad (2)$$

Where τ represents its characteristic of the echo wave in the frequency domain, $S(\tau, t)$ is derived from $S_t(t)$, which considers the electromagnetic data in the time domain and frequency domain. The time-frequency metamorphic relations are based on $S(\tau, t)$ transformation of the original data.

Channel Transformation. The channel transformation metamorphic operator mainly simulates the time domain interaction caused by the interference of adjacent electromagnetic signals in the timing sequence during signal propagation. We simulate this correlation by undirected interaction of partial I/Q data.

$$S(\tau + \phi, t + \phi), S(\tau, t) = S(\tau + \phi, t + \phi) \bigoplus S(\tau, t) \quad (3)$$

where, $S(\tau, t)$ is the received data at t moment, $S(\tau + \phi, t + \phi)$ is the received date at $t + \phi$ moment. $\bigoplus$ represents the exchange formula. τ is the interval between exchanged electromagnetic data. We control the metamorphosis intensity by setting threshold η, which controls the number of interacted electromagnetic data.

Power Zooming. The power zooming metamorphic operator considers the phenomenon of signal energy superposition due to the reflection of the signal. By directly varying the signal strength, the reflection of the signal when encountering a static obstacle is simulated. We simulate the strength of the received electromagnetic signal by controlling the deflation of $S(\tau, t)$. The metamorphic formula is $S^{'}(\tau, t) = \rho S(\tau, t)$. Where $S(\tau, t)$ and $S^{'}(\tau, t)$ is the original and deflated data. ρ is the strength of deflation.

In the time-frequency metamorphic component, we experiment with the variation approach on the time-frequency domain to get the optimal parameter value of ElecDaug. In augmenting process, the user inputs the original data, sets the variable parameter, and gets the augmented data of the corresponding intensity.

4 EVALUATION

To evaluate the effectiveness and availability of ElecDaug, we design the following research questions:

- **RQ1 (Effectiveness):** Whether ElecDaug augment valid data to repair the defects of the model?
- **RQ2 (Availability):** Whether the data augmented by ElecDaug always remain available?

4.1 Experimental Setup

We conduct experiments on the RML_2016.10a dataset [6]. The dataset contains 220, 000 input data, which are divided into 11 different modulations and consisting of 20 different signal-noise-ratio (SNR) ranging from −20dB to 18dB in 2dB steps. And the SNR is the ratio of signal to noise in an electronic device or system. Besides, the VT-CNN2 model [6] is selected as the primary neural network architecture to extract the characters from the RML_2016.10a dataset.

4.2 RQ1: Effectiveness of ElecDaug

We conducted several comparison experiments to demonstrate the effectiveness of augmented data in repairing model defects. In this process, we take the classification precision of the model on the test set as the reference for model repairing. The experiment processes are divided into the following steps: (1) divide the original data (RML_2016.10a) into the train set and test set in the ratio of 9 : 2; (2) choose metamorphic methods to augment data with the train set; (3) use the original and augmented training sets to train VT-CNN2 models and label them with its method; (4) compare the performance of the augmented data on VT-CNN2 models. For different metamorphic methods, we choose the optimal parameter for different metamorphic methods in our experiments, and the SNR is set to 0dB. In addition, since ML models may achieve different results in different rounds, we ran each model 10 times and de-averaged its performance. The experiment results are shown in Table 1.

Table 1: The Result of Model Classification Result

Augmentation Method	Augmentation Type	Accuracy(%)
Original	Original	71.97
Radio Noise	Transmission	73.35
Signal Outage	Transmission	74.43
Signal Disruptions	Transmission	75.51
Channel Transformation	Time-frequency	77.23
Power Zooming	Time-frequency	78.69
All Five Methods	All	**79.84**

As shown in Table 1, the data augmented by the five metamorphic methods in ElecDaug have significant improvement in improving the model for sample classification. In particular, the improvement effect of the time-frequency metamorphic methods is better than the transmission methods, and the models trained by combining the data augmented by the five metamorphic methods have the best classification performance. The results illustrate that all the electromagnetic data augmented by ElecDaug have positive results in repairing the model defects.

4.3 RQ2: Availability of ElecDaug

To verify the availability of ElecDaug, we compare the availability of the augmented data in different SNR environments. We perform incremental experiments in 20 SNR environments (between −20dB and 18dB with an interval of 2dB) to augment data to verify the availability of the metamorphic method in different SNR environments. We conduct incremental experiments and perform the same procedure as for RQ1 for electromagnetic data with different SNRs, and the experimental results are shown in Figure 3.

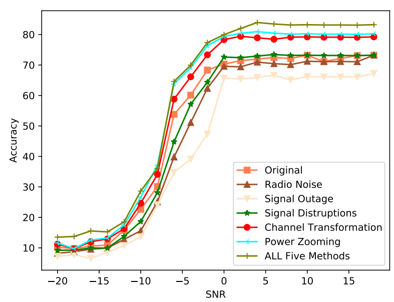


Figure 3: Data Effect under Different SNRs

As shown in Figure 3, all of these methods of ElecDaug have improved over the original in different SNRs. The result indicates that the data augmented by ElecDaug is available in different scenarios, which implies the availability of data augmentation methods based on the electromagnetic signal characteristics. In conclusion, we experimentally demonstrate that the ElecDaug tool works well to address the defects of the existing AMC model in classification. Data augmentation methods by the characteristics of the electromagnetic signal are available for repairing the model defects in different SNR environments.

5 USAGE

Figure 4 shows the page on the electromagnetic signal data augmentation page of ElecDaug. Through ElecDaug, users can upload the data to be augmented, select the corresponding metamorphic methods, set the metamorphic intensity, and complete the augmentation of electromagnetic data. After the augmenting process, users can enter the ElecDaug preview page and select a piece of augmented data. The selected data will be presented in the graph. The comparison data shows in Figure 5. Figure 5(a) shows the original data, and Figure 5(b) shows the augmented data, where the x-axis is the I-channel value, and the y-axis is the Q-channel value. When users click the "Reset" button, the existing task is cleared, and the user can continue with the next task.

6 RELATED WORK

While there are many groups working to improve the capacity of AMCs. O'Shea et al. [6] proposed the VT-CNN2 model and achieved good effects in AMC. However, the DL model is limited by the scene

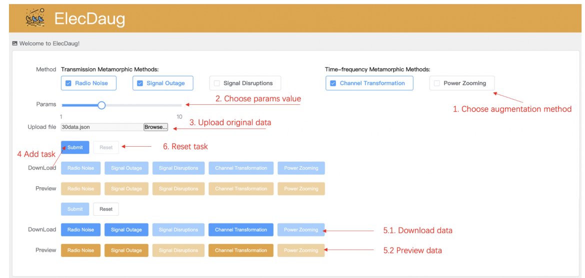
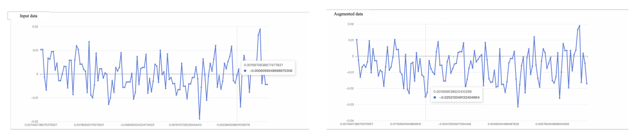


Figure 4: The User Guidance of ElecDaug System



(a) The Input Data (b) The Augmented Data

Figure 5: The Data Comparison Page

features. Liang et al. [3] proposed that the deficiency in the quantity of data and the unbalanced rate of data quality can lead to the poor performance of existing classification models in real life, and proposed an ensemble learning model to improve the classification ability of the model under different data. Hou et al. [2] proposed the TauMed tool to augment the dataset and carried out that data augmentation by metamorphic relationships can effectively improve the classification ability of deep learning models. It was not until Moin et al. [5] explored the feature differences between the inverse and original samples. Most researchers focus on the relationship of features between the enhanced data and the original data. Differing from previous research, we investigate the properties of the electromagnetic signal in the time domain, frequency domain, and signal transmission process. We use these properties to propose some metamorphic methods that form our prototype tool ElecDaug to augment more data consistent with real scenarios.

7 CONCLUSION

In this paper, we propose a novel electromagnetic signal data augmentation tool, namely ElecDaug, to solve the issue of low efficiency of modulation identification for electromagnetic application tools. Based on metamorphic relations, we obtain three transmission metamorphic methods and two time-frequency metamorphic methods according to the characteristics of wireless signals in transmission and time-frequency domains. To the best of our knowledge, it is the first attempt to augment data through metamorphic relations, and ElecDaug provides an effective instruction of data augmentation for repairing model defects. To verify the effectiveness and availability of the data augmented by ElecDaug, we conducted comprehensive experiments on the VT-CNN2 model with the RML_2016.10a dataset. Experimental results show that the data augmented by ElecDaug has a usable effect on repairing model defects due to insufficient data and poor diversity.

ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers for insightful comments. This research is partially supported by the National Natural Science Foundation of China (61932012, 62141215) and Science, Technology and Innovation Commission of Shenzhen Municipality (CJGJZD20200617103001003).

REFERENCES

[1] Tsong Yueh Chen, Fei-Ching Kuo, Huai Liu, Pak-Lok Poon, Dave Towey, TH Tse, and Zhi Quan Zhou. 2018. Metamorphic testing: A review of challenges and opportunities. *ACM Computing Surveys (CSUR)* 51, 1 (2018), 1–27.

[2] Yunhan Hou, Jiawei Liu, Daiwei Wang, Jiawei He, Chunrong Fang, and Zhenyu Chen. 2021. TauMed: Test Augmentation of Deep Learning in Medical Diagnosis. In *Proceedings of the 30th ACM SIGSOFT International Symposium on Software Testing and Analysis (ISSTA 2021)*. 674–677.

[3] Xiayu Liang, Ying Gao, and Shanrong Xu. 2022. ASE: Anomaly Scoring Based Ensemble Learning for Imbalanced Datasets. *arXiv preprint arXiv:2203.10769* (2022).

[4] Yixing Luo, Xiao-Yi Zhang, Paolo Arcaini, Zhi Jin, Haiyan Zhao, Fuyuki Ishikawa, Rongxin Wu, and Tao Xie. 2021. Targeting Requirements Violations of Autonomous Driving Systems by Dynamic Evolutionary Search. In *2021 36th IEEE/ACM International Conference on Automated Software Engineering (ASE)*. IEEE, 279–291.

[5] Armin Moin. 2021. Data Analytics and Machine Learning Methods, Techniques and Tool for Model-Driven Engineering of Smart IoT Services. *2021 IEEE/ACM 43rd International Conference on Software Engineering: Companion Proceedings (ICSE-Companion)* (2021), 287–292.

[6] Timothy J. O'Shea, Johnathan Corgan, and T. Charles Clancy. 2016. Convolutional radio modulation recognition networks. In *Proceedings of the 17th EANN*. Springer, Aberdeen, UK, 213–226.

[7] Yun Tang, Yuan Zhou, Tianwei Zhang, Fenghua Wu, Yang Liu, and Gang Wang. 2021. Systematic testing of autonomous driving systems using map topology-based scenario classification. In *2021 36th IEEE/ACM International Conference on Automated Software Engineering (ASE)*. IEEE, 1342–1346.