

Article

Vector Magnetic Anomaly Detection via an Attention Mechanism Deep-Learning Model

Xueshan Wu ^{1,2,3} , Song Huang ^{1,2,*}, Min Li ^{1,2,3} and Yufeng Deng ^{1,2,3}

¹ Key Laboratory of Petroleum Resources Research, Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing 100029, China; wuxueshan@mail.iggcas.ac.cn (X.W.); lim@mail.iggcas.ac.cn (M.L.); dengyufeng19@mails.ucas.ac.cn (Y.D.)

² Innovation Academy for Earth Science, Chinese Academy of Sciences, Beijing 100029, China

³ University of Chinese Academy of Sciences, Beijing 100049, China

* Correspondence: huangsong@mail.iggcas.ac.cn

Abstract: Magnetic anomaly detection (MAD) is used for detecting moving ferromagnetic targets. In this study, we present an end-to-end deep-learning model for magnetic anomaly detection on data recorded by a single static three-axis magnetometer. We incorporate an attention mechanism into our network to improve the detection capability of long time-series signals. Our model has good performance under the Gaussian colored noise with the power spectral density of $1/f^\alpha$, which is similar to the field magnetic noise. Our method does not require another magnetometer to eliminate the effects of the Earth's magnetic field or external interferences. We evaluate the network's performance through computer simulations and real-world experiments. The high detection performance and the single magnetometer implementation show great potential for real-time detection and edge computing.

Keywords: deep learning; Gaussian colored noise; magnetic anomaly detection (MAD); three-axis magnetometer



Citation: Wu, X.; Huang, S.; Li, M.; Deng, Y. Vector Magnetic Anomaly Detection via an Attention Mechanism Deep-Learning Model. *Appl. Sci.* **2021**, *11*, 11533. <https://doi.org/10.3390/app112311533>

Academic Editor: Valentina E. Balas

Received: 23 October 2021

Accepted: 2 December 2021

Published: 6 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Magnetic anomaly detection (MAD) is a passive method for detecting ferromagnetic objects. If the geomagnetic field is taken as the normal field, the magnetic anomaly is the relative magnetic field change produced by the appearance of ferromagnetic material, and the anomaly can be detected by the magnetometer. The MAD method is used to detect the appearance of ferromagnetic targets in the continuous magnetic records measured by the remote magnetometer [1].

Compared to magnetic anomalies caused by the ferromagnetic materials, factors such as biology, water, and almost soils are nearly transparent in the geomagnetic field. Moreover, the long-range passive observation device is easier to conceal from the target side. Therefore, MAD has advantages in intrusion alert and hidden target detection [2]. When the distance to the target body is relatively far compared to its size, the magnetic anomaly field, b , can be considered as a magnetic dipole field [3]:

Here, $\vec{m}$ is the target magnetic moment, and $\vec{r}$ denotes the vector from the target to magnetometer. μ_0 is the space's permeability.

$$\vec{B}(\vec{m}, \vec{r}) = \frac{\mu_0}{4\pi} \left[\frac{3(\vec{m} \cdot \vec{r})\vec{r}}{|\vec{r}|^5} - \frac{\vec{m}}{|\vec{r}|^3} \right] \quad (1)$$

There are two challenges in the MAD task: (i) the target signal is generally much weaker compared to the background noise. The target's magnetic signal decreases with the cube of distance, which makes MAD a method to extract weak signals. (ii) The background

noise is complex. The geomagnetic field as background noise is variable, colored, and similar to the target magnetic field. Its power spectral density is $1/f^\alpha$ ($0 < \alpha < 2$, and α varies with time and position) [4].

There were two major categories in MAD, including the target-based approach and noise-based approach, until now [5]. The orthogonal basis function (OBF) is the most representative target-based method [6]. The signal-to-noise ratio (SNR) is improved by matching multiple orthogonal basis functions with measured signals to the energy function. Several studies suggest that OBF is very efficient in Gaussian white noise, but it needs to be preprocessed in practical experiments, which undoubtedly increases the computation cost [7]. The noise-based methods, such as minimum entropy filtering [8], the high-order crossing magnetic anomaly method [9], and wavelet transform [10] suggested that the background field obeys certain distributions and is changed mainly due to the presence of ferromagnetic targets. Although these methods do not need prior information on targets, their performance is limited under low SNR.

Recently, many researchers have tried to use machine learning, particularly deep learning, to solve the MAD problem. For example, Fan extracted the target signal from the OBF energy function through support vector machines [11] and identified the target signal from the two-dimensional temporal frequency spectrum by using convolutional neural networks (CNN) [12]. The problems of high computational cost and reliance on Gaussian white noise remain to be solved. A researcher proposed a deep-learning model called DeepMAD that offers some new insights [13]. It is an end-to-end framework, which provides the possibility for real-time detection. However, it uses actual geomagnetic noise as the dataset to train a five-layer CNN model, which lacks generalization. Moreover, the CNN model cannot handle the relationship between local and global, and it can easily lose the information of long periods.

Such approaches, however, have additional requirements for the signal when dealing with a static magnetometer and a moving target. In order to eliminate the Earth's magnetism and reject distant external interferences, the reference magnetometer should be away from the target and under the same background noise as the measurement magnetometer. It requires data from both magnetometers before the signal processing. The above methods need the differential signal because they lack focus on the extraction of the overall features and attention to the target change areas. Rapid changing magnetic fields are easier to pick out, and the signals change slowly as the target enters the detection range.

In recent years, deep learning has developed rapidly in time-series processing, which provides new techniques for MAD. Recursive Neural Networks (RNNs) [14] can effectively capture features of temporal data, but they have problems in modeling long-term relationships. This problem is solved by introducing long and short-term memory networks, which are special kinds of RNNs. However, LSTM will lead to overfitting when the sequence contains linear relationships or noise. Bi-directional LSTM (BiLSTM) [15] further develops LSTM by combining forward and backward hidden layers, which can utilize information from both directions. Currently, LSTM and BiLSTM have been applied to text classification and other fields. In addition, the Attention Mechanism [16], which is based on the human attention mechanism, was first used in machine translation tasks and has now been extended to various research areas, such as image segmentation. It can obtain weight by calculating the magnitude of the influence of each time point of the input sequence on the output, and the obtained weight can be used in the attribution analysis of the data. Nowadays, it is also widely used with LSTM and CNN to process sequence data in a structure similar to U-Net [17]. This structure has been used in different anomaly detection demands, such as social media [18], marine science [19], and natural language processing [20].

In this study, we introduce a new end-to-end deep-learning model for MAD. An attention mechanism module was incorporated in the model to assist the CNN. This improved the ability to identify the target signal in the input time series. The attention mechanism in deep learning was inspired by human focus patterns. Humans will scan

the global environment and focus their attention on a certain area to analyze it when confronted with a complex image. Our model was simulated through two levels of attention mechanisms. One level globally detected the range of possible targets. Another one locally picked up the exact time point when the target appears. The training and detection processes are as follows. First, the network is trained by a training set of computer numerical simulations. Secondly, the model will output the probability of the presence of magnetic anomalies at each time point from the three-component magnetic signal recordings in the field experiment. Finally, the threshold value as needed is selected to help the detection. The main advantages can be summarized as follows:

1. The observation system is simple because there is no need for another magnetometer to calculate the differential signal.
2. There is no need to make assumptions about geomagnetic noise and record it in advance. Targets can be detected in colored noise with low SNR.
3. The computational cost is mainly concentrated in the model training phase. Due to low computational complexity for practical applications, it is suitable for edge computing and real-time detection.

2. Methods

2.1. Network Design and Architecture

Magnetic signals are sequential time series consisting of different local (targets presence) and more global (e.g., instrument background noise and daily variation of magnetic field) features. Hence, we can consider MAD as a classification problem at each time point. Traditionally, recurrent neural networks have been used for such sequence modeling. Relatively long-duration magnetic signals require downsampling prior to feeding to the recurrent layers to decrease the computational complexity. Here, we introduce several modules in our model. The general architecture is based on U-Net, which has presented an excellent performance in the segmentation task. The U-Net network structure is symmetrical and resembles the letter U. The most important advantage of U-Net for our network design is that it can build a very deep encoder by downsampling. Traditionally, recurrent neural networks (RNN) have been used for such sequence modeling; however, downsampling prior to the recurrent layers is imperative to manage the computational complexity when input is relatively long duration signals. Moreover, U-Net's skip layers can help keep detailed information.

Our deep-learning model for MAD consists of an encoder and decoder, including 1D convolutions, residual connections (Res CNN) [21], bi-directional long-short-term memories (BiLSTM), and a transformer (Figure 1).

Since excessively long sequences in the self-attentive mechanism cause computation to increase exponentially, we carried out downsampling through max-pooling layers. We added three skip connection layers to extract convolutional features at different levels. These features were accepted through the concatenate layers in the decoder for learning the picking task hierarchically. In this way, the magnetic field signal was transformed from a sequence in the time domain into a high-dimensional feature representation.

Next, we processed these expressions by two attention modules. The first global attention module guided the network attention to the part where the target appeared by assigning different weights. The second local attention module is the last part of the encoder. It mainly transmits the signal of the target appearance to the decoder. After passing through a very deep encoder, the signal reaches the bi-directional long short-term memory (BiLSTM) blocks. Long short-term memory (LSTM) is a specific type of RNN generally used in modeling longer sequences. The main element in the LSTM cell is a memory cell. At each time step, the LSTM cell receives an input, outputs a hidden state, and updates the memory cell according to the gate mechanism. Compared with the regular RNN, the LSTM could keep the long-period information, which is helpful for the attention mechanism. BiLSTM is used to assist the attention module in distinguishing long sequence signal features by incorporating time-positional information. As shown by the brown

squares in Figure 1, the double arrows indicate the recurrence direction of the BiLSTM. Decoders then use this information to map the high-level features by a sequence of target existence probabilities.

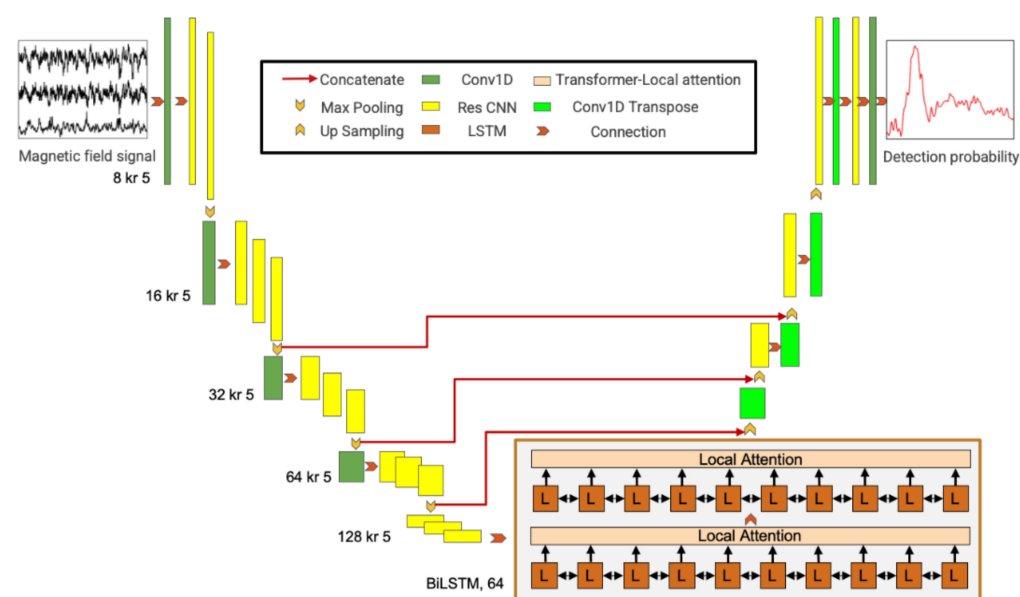


Figure 1. Our network architecture. The different types of network layers are shown in the illustration. The input is the magnetic field signal measured by the magnetometer and the output is the probability of magnetic anomaly detection at each sampling point. The details of each block are provided in the methods section. The convolutional layers read as: number of kernels, kr, kernel size.

2.2. Data Preparation

The data used in this paper are obtained by simulation. The establishment of the data is divided into two parts: one is the background noise data, and the other is the magnetic anomaly data generated by the moving target. By combining data from these two parts, we obtain the labeled training and testing data set.

2.2.1. Magnetic Anomaly Signal

According to (1), when the magnetometer is fixed, the simulation of the target magnetic anomaly is determined by several parameters, including magnetic dipole moment, distance vector and sampling rate.

We limited the magnetic dipole moment and speed to a specific range for simplicity without losing generality. The settings for simulation parameters are listed in Table 1. Due to the three-axis magnetometer, the signal of each axis differs in different directions. The magnetic anomaly signals are generated following Equation (1) and Table 1. We simulated a series of magnetic dipoles passing through the magnetometer straight from different directions. The data length for each simulated signal is 300 s. Figure 2 shows a typical signal waveform plot.

Table 1. Parameters used for signal simulation.

Parameters		Value
Magnetic moment $\vec{m}$	Speed (Km/h)	[40~60]
	X component (Am^2)	[4500]
	Y component (Am^2)	[0]
	Z component (Am^2)	[−2500]
Sampling rate (Hz)		128
Data length (s)		300

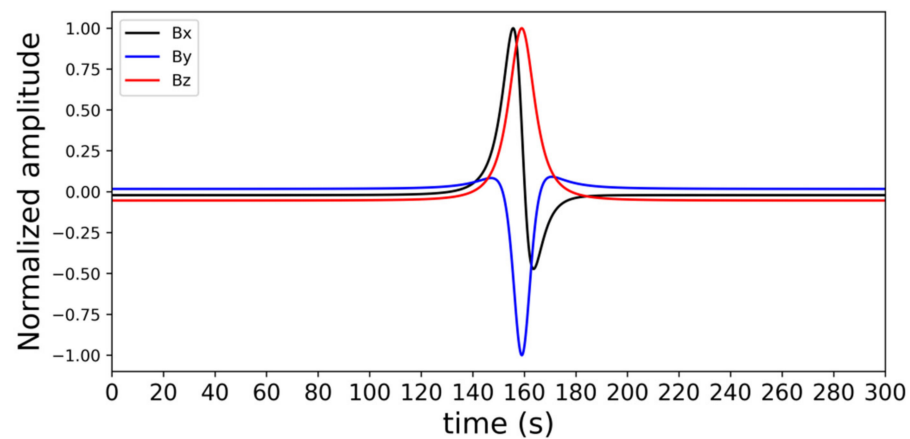


Figure 2. Simulated magnetic anomaly signal using dipole model.

2.2.2. Background Noise

Geomagnetic noise as background noise data does not provide good robustness of the model. The geomagnetic data that can be collected are recorded by different instruments for a limited period at specific locations. To improve the model's generalization, we obtained the colored background noise by transforming the Gaussian white noise in the frequency domain. We set its power spectral density as $1/f^\alpha$ ($0 < \alpha < 2$), then inverted it to the time domain. The typical background noise with the PSD of $1/f^{1.5}$ is shown in Figure 3.

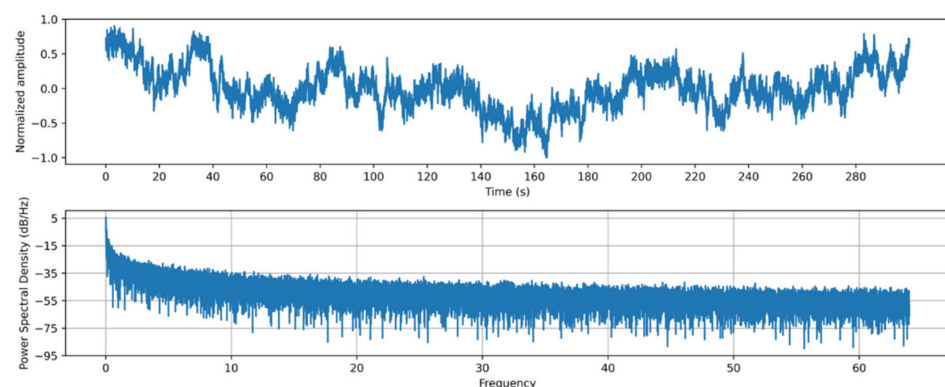


Figure 3. Simulated temporal plot background noises and PSD of $1/f^{1.5}$.

2.2.3. Data Set and Training

We added the simulated background noise with gradual SNR (from -10 dB to 10 dB) to the target signal to form the data set. Target signals greater than 0.2 in the normalized amplitude were marked as 1 to represent the presence of the target. The rest of the signals were marked as 0 to represent the absence of the target. We eventually generated a data set consisting of $100,000$ labeled examples. We randomly split the data into training (85%) set, validation (5%) set, and test (10%) set. We used the Adam optimizer [22] with an initial learning rate of 0.0003 . The batch size was set to be 64 , and the whole training consisted of 100 epochs.

3. Experiment

3.1. Simulation Experiment

We completed the simulation experiment with the test set. For a given α , 1000 testing instances were synthesized to obtain the average probability of each 0.5 SNR (-10 ~ 10 dB). We compared the predicted probability of each sampling point with the corresponding

ground truth. A typical test example is shown in Figure 4, with an SNR = 0 dB, and $\alpha = 1.5$. The SNR is defined as:

$$Power_{signal} = \frac{1}{N} \sum_{i=1}^N x[n]^2 \quad (2)$$

$$SNR = 10 \log_{10} \frac{P_{target}}{P_{noise}} dB \quad (3)$$

where $Power_{signal}$ is defined by N and $x[n]$, which are the length of the target signal and amplitude of each sampling point, respectively. Because we used only one magnetometer to measure the geomagnetic field, we calculated the deviation between observational and minimum geomagnetic data as the power of background noise.

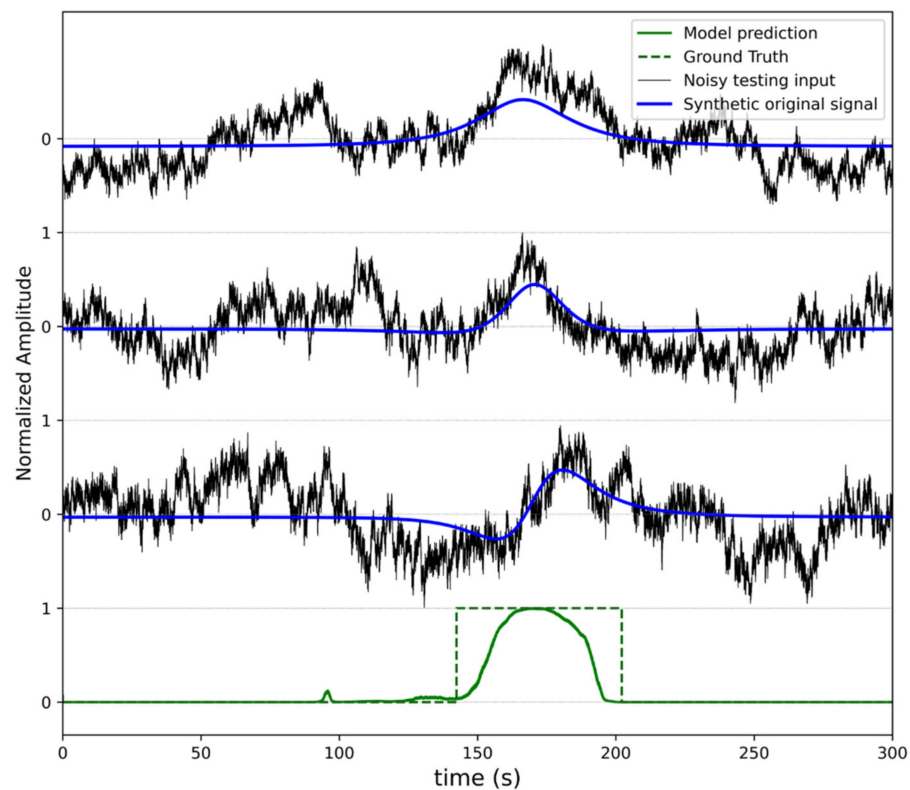


Figure 4. Typical test experiment. Three-axis simulation magnetic signals (the black lines) and the target signals (the blue lines). The probabilities are predicted by the model and the ground truth.

To better describe the model performance of each sampling point, we define the following two parameters: Precision (P) and Recall (R):

$$P = \frac{TP}{TP + FP} \quad (4)$$

$$R = \frac{TP}{TP + FN} \quad (5)$$

where TP is the number of true-positive points, which means correctly detected target numbers. FP is the number of false-positive points, which represents the falsely alarmed target numbers. FN is the number of false-negative points, which denotes the missing target numbers. The precision is the number of true positive results divided by the number of all positive results, including those not identified correctly, and the recall is the number of true-positive results divided by the number of all samples that should have been identified as positive. Precision is also known as positive predictive value, and recall is also known

as sensitivity in diagnostic binary classification. More examples of the tests are shown in Figure A1. The simulation results are shown in Figure 5.

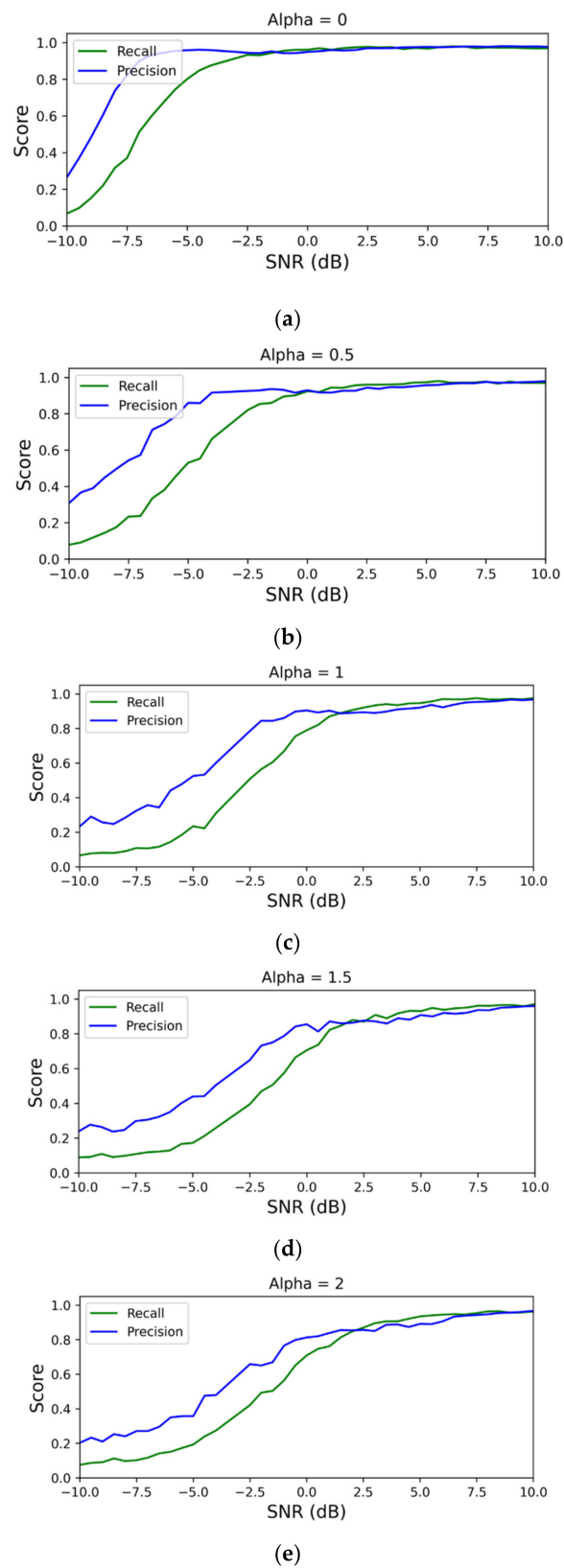


Figure 5. The recall and precision of the model under the Gaussian colored noise with the PSD of $1/f^\alpha$. (a) $\alpha = 0$. (b) $\alpha = 0.5$. (c) $\alpha = 1$. (d) $\alpha = 1.5$. (e) $\alpha = 2.0$.

We follow this rule to count the sampling points with probability greater than 0.5 represent true in each test experiment. Figure 5a shows excellent performance under the Gaussian white noise ($\alpha = 0$). It is apparent from this figure that the recall and precision are still up to 70% when the SNR is down to -5 dB. Figure 5b–e show the cases of several Gaussian colored noises ($\alpha = 0.5$, $\alpha = 1$, $\alpha = 1.5$ and $\alpha = 2.0$, respectively). The model has consistent results when the SNR is higher than 0 dB. However, model performance decreases slightly as the colored noise becomes stronger. A higher Precision means that the false alarm rate of magnetic anomaly detection is lower, and the results are more reliable. A higher Recall means that fewer magnetic anomalies are missed. Overall, it is a satisfactory result to maintain a good precision rate based on a high recall rate.

To further evaluate the network performance, we introduced another parameter *F-score*, which is calculated from the *Precision* and *Recall* of the test:

$$F - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (6)$$

In statistical analysis of binary classification, the *F-score* is a measure of a test's accuracy. The highest possible value of an *F-score* is 1.0, indicating perfect *Precision* and *Recall*, and the lowest possible value is 0 if either the *Precision* or the *Recall* is zero. We calculate the *F-score* based on the classification in Table 2. Each row is a different alpha, and each column represents a different SNR interval. The *F-score* in each bin is calculated from 5000 samples. From the table, each alpha has an SNR interval as a watershed. When the SNR is greater than this interval, the *F-score* increases rapidly to above 0.7. This certainly gives the confidence interval for the model. As the composition of colored noise becomes more and more (alpha), the model's reliance on SNR becomes stronger. When the SNR is higher than 2.5, the *F-score* is almost greater than 0.9, which indicates good performance of the model.

Table 2. The *F-score* of the model under the Gaussian colored noise.

<i>F-Score</i>	[−10, −7.5)	[−7.5, −5)	[−5, −2.5)	[−2.5, 0)	[0, 2.5)	[2.5, 5)	[5, 7.5)	[7.5, 10]
$\alpha = 0.0$	0.249	0.701	0.908	0.952	0.965	0.974	0.976	0.975
$\alpha = 0.5$	0.191	0.432	0.778	0.915	0.942	0.961	0.970	0.973
$\alpha = 1.0$	0.115	0.178	0.451	0.792	0.899	0.930	0.953	0.967
$\alpha = 1.5$	0.133	0.190	0.385	0.700	0.854	0.911	0.944	0.956
$\alpha = 2.0$	0.129	0.199	0.401	0.708	0.852	0.899	0.937	0.957

3.2. Field Experiment

In order to evaluate the model performance under an extremely high colored noise background in the field, we conducted a field experiment to detect the subway stop information in the Beijing Yuandadu Park, lasting 2 h (Figure A2). This park is surrounded by heavy traffic lines with many passengers and is close to the roads and subway stations.

The magnetometer we used is FGS-03, which is a low noise 3-axis fluxgate sensor. The sensor noise is less than $6 \text{ pT}/\sqrt{\text{Hz}}$, and the output sensitivity is 0.1 mV/nT . We chose the sample rate of 128 Hz for more information. A total of 120 min of three-component continuous magnetic records were obtained from the experiment.

We acquired the probability of magnetic anomaly detection by our model using a 300 s-long sliding window on the data with 30% overlap per step. Figure 6 shows the magnetic records from the field experiment and the probability of MAD, respectively.

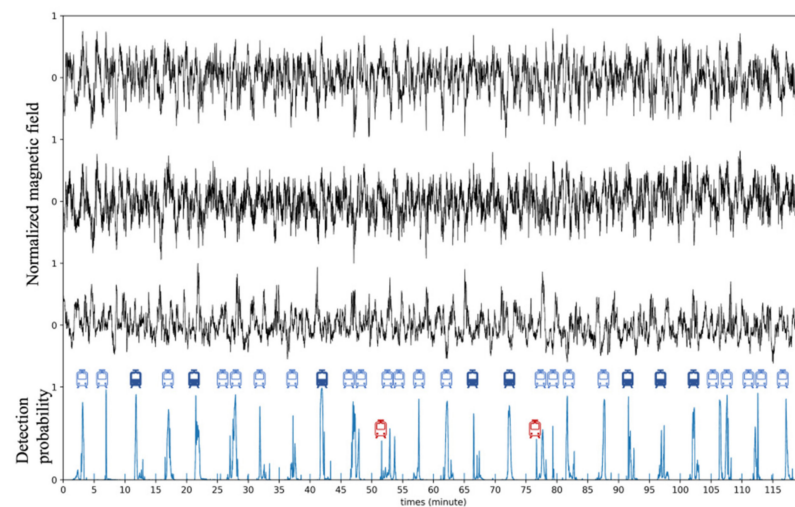


Figure 6. The data from the field experiment and the probability of MAD. The blue hollow icon represents a correct detection, the blue solid icon represents two trains, and the red hollow icon represents a false alarm.

The probability is almost higher than 0.5 or 0, shown in Figure 6, which indicates that the model can output direct magnetic anomaly detection results in the case of urban background noise. After checking the train schedules (Table A1), it was determined that 38 trains passed through the subway station near the magnetometer during the data recording period. The occurrence of these trains is consistent with the peak of the probability plot. There are only two false alarms with a probability of just reaching 50%, which can be avoided by increasing the screening threshold. Further analysis showed that when the time interval of two trains is small, it is very difficult to distinguish only by the peak. The eight blue solid icons show the following case: trains in both directions enter the station almost simultaneously. Figure 7 further demonstrates this case. Figure 7a is an enlargement of minutes 20 to 25, showing the two trains being detected almost simultaneously. The two peaks (red circles) show that the model's predictions are still distinguishable. Figure 7b shows a correct detection for minutes 85 to 90. Although there are still a few false alarms, they are easily filtered. Overall, the model completed the magnetic anomaly detection.

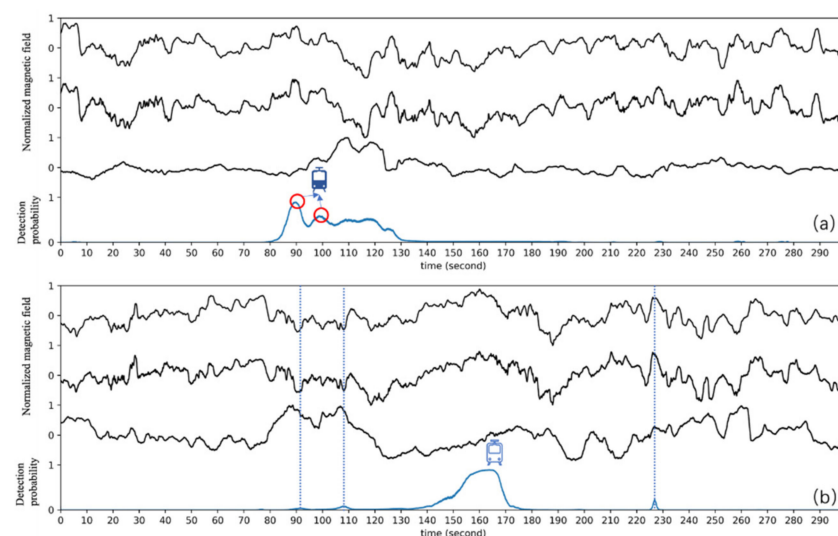


Figure 7. Two examples from the field experiment in Figure 6. The blue hollow icon represents a correct detection, the blue solid icon represents two trains, and the blue dotted line represents a false alarm. (a) minutes 20–25, (b) minutes 85–90.

4. Discussion

With the above experiments, the attention deep-learning model we proposed enables only one magnetometer to detect magnetic anomaly under Gaussian color background noise.

In addition to developing a powerful tool for magnetic anomaly detection, our study provides a novel perspective on the observation system. We suggest that two main factors control the performance of detecting moving targets: stations and generalizability. Because most methods need differential calculation by the reference magnetometer to eliminate interference, this is the reason why no comparison with other methods is made in this study. Moreover, different methods have different limitations; for example, OBF is not good at colored noise, MED is limited by SNR, and DeepMAD relies on datasets. We noted that deep learning could learn feature extraction from massive amounts of data to complete the classification task, and attention mechanisms can help the model process such sequential signals.

Our model carried out single station detection with the benefit of a very deep encoder and attention mechanism, and the generalizability was improved by training with large amounts of simulation data. The model is pioneering; however, it needs the following improvements. The first is multi-target detection: when multiple targets appear at the same time, it is difficult for the model to differentiate. The second is model implantation: the network structure is still too large to be implemented into mobile chips. The third is transfer learning: when going for specific target detection in a particular place, we suggest training the model with a dataset containing local geomagnetic and target anomalies.

5. Conclusions

In conclusion, an advanced deep-learning model based on the attention mechanism is proposed to detect magnetic anomalies. The simulation and field experiment results show that the model is still applicable in the case of a single magnetometer, colored noise, and low SNR. The simple implementation and outstanding performance lead to a wide range of real-time application scenarios. This type of application is used for perimeter protection, such as underwater target detection, intrusion alert, or indoor ferromagnetic object location.

Author Contributions: Conceptualization, S.H.; methodology, X.W.; validation, X.W., M.L. and Y.D.; resources, S.H.; data curation, Y.D.; writing—original draft preparation, X.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China under grant 91858214 and the National Key R&D Program of China under grant 2018YFC0604004.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

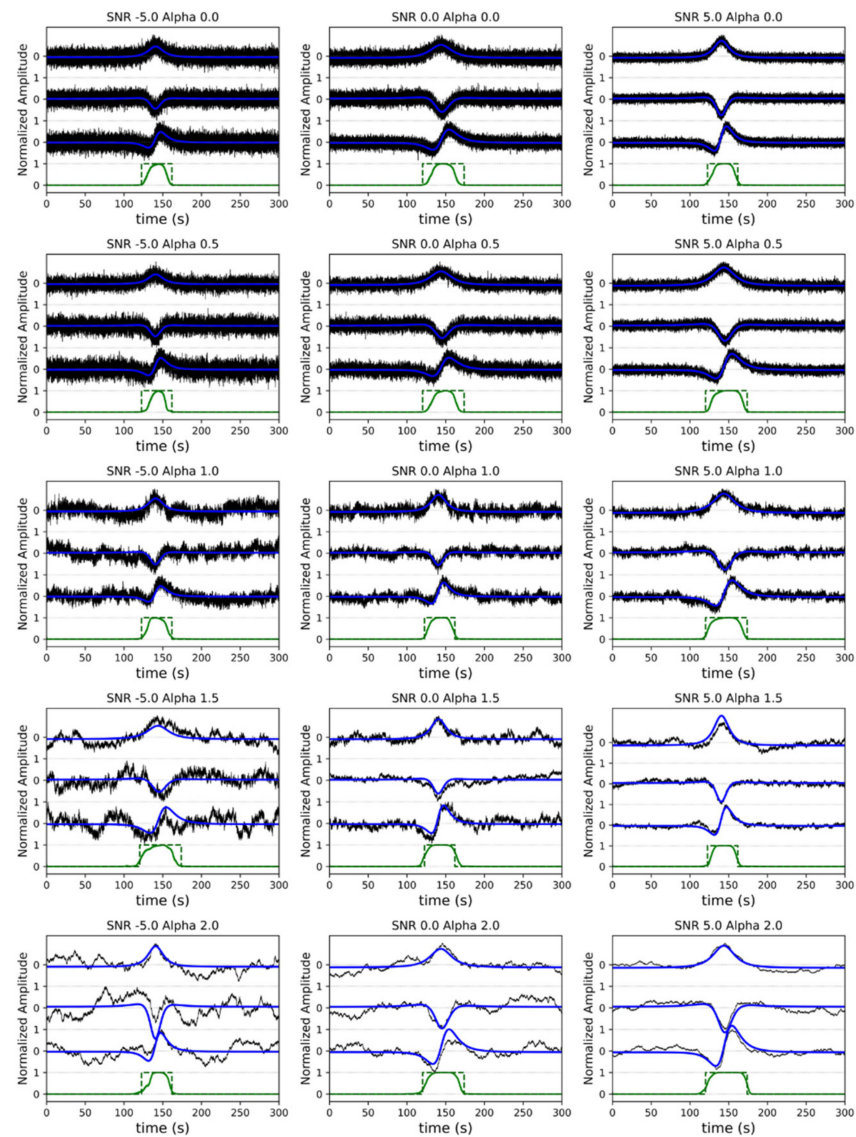


Figure A1. Typical test experiments of different SNR ($-5.0 \sim 5$) and Alpha ($0 \sim 2.0$). Three-axis simulation magnetic signals (the black lines) and the target signals (the blue lines). The probabilities are predicted by the model and the ground truth.

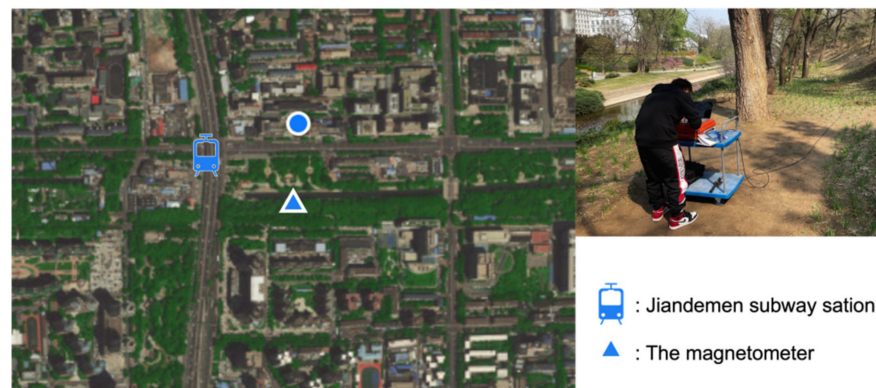


Figure A2. Location of field experiment and subway station.

Table A1. The timetable of Jiandemen station during the field experiment from 14:10 to 16:10.

Driving Direction	Time of Arrival
To MUDANYUAN (west)	14:15, 21, 27, 33, 39, 46, 52, 58
	15:04, 10, 16, 22, 28, 35, 41, 47, 53, 59
	16:05
To BEITUCHENG (east)	14:11, 17, 24, 30, 37, 43, 50, 56
	15:03, 09, 16, 22, 29, 35, 42, 48, 56
	16:01, 08

References

- Wiegert, R. Magnetic Anomaly Guidance System for Mine Countermeasures Using Autonomous Underwater Vehicles. In Proceedings of the Oceans 2003. Celebrating the Past Teaming Toward the Future (IEEE Cat. No.03CH37492), San Diego, CA, USA, 22–26 September 2003; Volume 4, pp. 2002–2010. [\[CrossRef\]](#)
- Paperno, E.; Sasada, I.; Leonovich, E. A New Method for Magnetic Position and Orientation Tracking. *IEEE Trans. Magn.* **2001**, *37*, 1938–1940. [\[CrossRef\]](#)
- Baldoni, J.A.; Yellen, B.B. Magnetic Tracking System: Monitoring Heart Valve Prostheses. *IEEE Trans. Magn.* **2007**, *43*, 2430–2432. [\[CrossRef\]](#)
- Sheinker, A.; Shkalim, A.; Salomonski, N.; Ginzburg, B.; Frumkis, L.; Kaplan, B.-Z. Processing of a Scalar Magnetometer Signal Contaminated by $1/f\alpha$ noise. *Sens. Actuators A Phys.* **2007**, *138*, 105–111. [\[CrossRef\]](#)
- Ginzburg, B.; Sheinker, A.; Salomonski, N.; SOREQ, N.; Frumkis, I.L.; Kaplan, B.Z. Application of Target-Based and Noise-Based Methods in Magnetic Anomaly Detection Systems. In Proceedings of the International Marine Electromagnetics Conference (MARELEC 2009), Stockholm, Sweden, 7–9 July 2009.
- Ginzburg, B.; Frumkis, L.; Kaplan, B.-Z. Processing of Magnetic Scalar Gradiometer Signals Using Orthonormalized Functions. *Sens. Actuators A Phys.* **2002**, *102*, 67–75. [\[CrossRef\]](#)
- Zhao, Y.; Zhang, J.H.; Li, J.H.; Liu, S.; Miao, P.X.; Shi, Y.C.; Zhao, E.M. A Brief Review of Magnetic Anomaly Detection. *Meas. Sci. Technol.* **2020**, *32*, 042002. [\[CrossRef\]](#)
- Sheinker, A.; Salomonski, N.; Ginzburg, B.; Frumkis, L.; Kaplan, B.-Z. Magnetic Anomaly Detection Using Entropy Filter. *Meas. Sci. Technol.* **2008**, *19*, 45205. [\[CrossRef\]](#)
- Sheinker, A.; Ginzburg, B.; Salomonski, N.; Dickstein, P.A.; Frumkis, L.; Kaplan, B. Magnetic Anomaly Detection Using High-Order Crossing Method. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 1095–1103. [\[CrossRef\]](#)
- Zhou, J.J.; Lin, C.S.; Huan, Y.C. Decreasing Noise in Magnetic Anomaly Detection Basing on Wavelet Denoising. *Appl. Mech. Mater.* **2013**, *368*, 1860. [\[CrossRef\]](#)
- Fan, L.; Kang, C.; Wang, H.; Hu, H.; Zhang, X.; Liu, X. Adaptive Magnetic Anomaly Detection Method Using Support Vector Machine. *IEEE Geosci. Remote Sens. Lett.* **2020**, *99*, 1–5. [\[CrossRef\]](#)
- Hu, M.; Jing, S.; Du, C.; Xia, M.; Peng, X.; Guo, H. Magnetic Dipole Target Signal Detection via Convolutional Neural Network. *IEEE Geosci. Remote Sens. Lett.* **2020**, *24*, 1–5. [\[CrossRef\]](#)
- Xu, X.; Huang, L.; Liu, X.; Fang, G. DeepMAD: Deep Learning for Magnetic Anomaly Detection and Denoising. *IEEE Access* **2020**, *8*, 121257–121266. [\[CrossRef\]](#)
- Williams, R.J. *Complexity of Exact Gradient Computation Algorithms for Recurrent Neural Networks*; Technical Report NU-CCS-89-27; Northeastern University, College of Computer Science: Boston, MA, USA, 1989.
- Huang, Z.; Xu, W.; Yu, K. Bidirectional LSTM-CRF Models for Sequence Tagging. *arXiv* **2015**, arXiv:1508.01991.
- Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. *arXiv* **2017**, arXiv:1706.03762.
- Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015, Munich, Germany, 5–9 October 2015; Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 234–241.
- Kim, T.-Y.; Cho, S.-B. Web Traffic Anomaly Detection Using C-LSTM Neural Networks. *Expert Syst. Appl.* **2018**, *106*, 66–76. [\[CrossRef\]](#)
- Oehmcke, S.; Zielinski, O.; Kramer, O. Input Quality Aware Convolutional LSTM Networks for Virtual Marine Sensors. *Neurocomputing* **2018**, *275*, 2603–2615. [\[CrossRef\]](#)
- Zhou, X.; Hu, B.; Chen, Q.; Wang, X. Recurrent Convolutional Neural Network for Answer Selection in Community Question Answering. *Neurocomputing* **2018**, *274*, 8–18. [\[CrossRef\]](#)
- He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778.
- Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980.