

## Article

# One-Shot Fault Diagnosis of Wind Turbines Based on Meta-Analogical Momentum Contrast Learning

Xiaobo Liu, Hantao Guo and Yibing Liu \* 

Key Laboratory of Power Station Energy Transfer Conversion and System, North China Electric Power University, Ministry of Education, Beijing 102206, China; liuxiaobo116@163.com (X.L.); g\_hantao@126.com (H.G.)

\* Correspondence: lyb@ncepu.edu.cn

**Abstract:** The rapid development of artificial intelligence offers more opportunities for intelligent mechanical diagnosis. Fault diagnosis of wind turbines is beneficial to improve the reliability of wind turbines. Due to various reasons, such as difficulty in obtaining fault data, random changes in operating conditions, or compound faults, many deep learning algorithms show poor performance. When fault samples are small, ordinary deep learning will fall into overfitting. Few-shot learning can effectively solve the problem of overfitting caused by fewer fault samples. A novel method based on meta-analogical momentum contrast learning (MA-MOCO) is proposed in this paper to solve the problem of the very few samples of wind turbine failures, especially one-shot. By improving the momentum contrast learning (MOCO) and using the training idea of meta-learning, the one-shot fault diagnosis of wind turbine drivetrain is analyzed. The proposed model shows a higher accuracy than other common models (e.g., model-agnostic meta-learning and Siamese net) in one-shot learning. The feature embedding is visualized by t-distributed stochastic neighbor embedding (t-SNE) in order to test the effectiveness of the proposed model.

**Keywords:** one-shot learning; momentum contrast learning; model-agnostic meta-learning; wind turbines



**Citation:** Liu, X.; Guo, H.; Liu, Y. One-Shot Fault Diagnosis of Wind Turbines Based on Meta-Analogical Momentum Contrast Learning. *Energies* **2022**, *15*, 3133. <https://doi.org/10.3390/en15093133>

Academic Editors: Qingshan Jiang, John (Junhu) Wang and Min Yang

Received: 21 March 2022

Accepted: 21 April 2022

Published: 25 April 2022

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



**Copyright:** © 2022 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

Fault diagnosis of wind turbines plays an important role in improving the reliability of wind turbines. However, the operating conditions of wind turbines change randomly, and multiple faults often occur simultaneously. In reality, wind turbines are not allowed to run with faults; when a fault occurs, the wind turbine will shut down. As a result, very few fault samples are available. Most deep learning will fall into overfitting when dealing with small sample problems, resulting in a low diagnostic accuracy. Few-shot learning (FSL) can train deep models with very limited data and solve the overfitting problem. The current few-shot learning includes generative modeling, meta-learning, and metric learning [1].

Generative adversarial networks (GAN) have been widely used to enhance the diversity of data [2–4]. Subsequently, there has been a proliferation of adversarial models presented by GAN, such as AnoGANs [5], GANormaly [6], etc.; auto-encoder (AE) is another way of generating samples. AE has now developed numerous variants, e.g., variational AE (VAE) [7], adversarial AE (AAE) [8], etc. However, generative modeling cannot solve the problem of very scarce data because of pattern collapse.

Meta-learning attempts to improve the network's ability to learn higher-level tasks by learning the feature representation of the task and generalizing on new tasks [9]. Wu et al. constructed seven few-shot transfer learning methods based on meta-learning [10]. Wang et al. proposed a meta-learning model based on a feature space metric for the fault diagnosis of bearings [11]. Feng et al. proposed a semi-supervised meta-learning network for low-probability fault diagnosis [12]. Su et al. proposed a novel method called data reconstruction hierarchical recurrent meta-learning for bearing fault diagnosis [13]. Wang et al.

proposed a metric-based meta-learning method named Reinforce Relation Network for bearing fault diagnosis [14].

Metric learning refers to learning a similarity metric under which a pair of similar samples can obtain a higher similarity score while non-similar pairs obtain a lower similarity score. Based on metric learning, Siamese neural networks [15,16], neural networks with external memories [17,18], relation networks [19], and graph neural networks [20] have been applied to few-shot learning.

The results of metric learning depend on the sampling strategy. The recent study of [21] suggests that the core of improving FSL also lies in improving the embedding learned. Self-supervised approaches based on contrast learning (CL) have recently become dominant. CL aims at grouping similar samples closer to and diverse samples far from each other [22], which is a kind of metric learning. CL learns the feature representation of samples by comparing data with positive samples and negative samples, respectively, in feature space. The model only needs to learn differentiation in feature space and does not pay too much attention to pixel details. This not only simplifies optimization, but also facilitates the establishment of a sampling strategy conducive to differentiation.

CL-based FSL has been widely used in few-shot image recognition [22–26]. At present, CL is seldom used in the field of fault diagnosis. Chen et al. proposed a new deep learning algorithm based on CL for the automated diagnosis of COVID-19 [27]. Ding et al. proposed a specific architecture for a self-supervised pre-training via contrast learning (SSPCL) deployment on bearing vibration signals [28].

At present, there are some limitations in the few-shot fault diagnosis: (1) The current few-shot fault diagnosis is mainly aimed at the testbed data under certain working conditions and realizing the data transfer learning from one fixed working condition to another working condition; (2) Currently, there are few analysis results when the sample size is very small (e.g., five samples); (3) At present, CL is seldom used in the few-shot fault diagnosis. Therefore, the innovations of this paper are as follows: (1) The actual operating conditions of wind turbines vary randomly, and the specific operating conditions at each moment are difficult to obtain. This paper proposes a stable few-shot fault diagnosis model of wind turbines that can adapt to various working conditions. (2) This paper analyzes the few-shot diagnosis results when the amount of data are very small. (3) This paper establishes a few-shot diagnosis model by combining contrastive learning with meta-learning.

Humans learn new concepts with little supervision; e.g., a child can generalize the concept of a giraffe from a picture in a book. In deep learning, the model can recognize the category from a single sample, which is called one-shot learning. How to learn transferable feature representations from a limited number of samples is a key challenge in few-shot learning. The data itself provide far more information than sparse labels. Contrast learning directly uses the data as supervised information to learn feature representations of sample data, and use them for downstream tasks. So, contrast learning is better able to learn generic knowledge to improve the characterization of the model and adapt it to the migration of different working conditions. Momentum contrast learning (MOCO) uses dictionary queuing and momentum updates to avoid losing consistency in features due to drastic changes in the encoder, while also keeping the encoder in a constant state of being updated. The stable updates of MOCO allow for better learning of the general knowledge of the data and, thus, better transfer of features to other tasks. Meta-learning equips models with certain prior knowledge or learning skills by learning multiple tasks; thus, they can better generalize and learn new tasks.

In this paper, we proposed a novel one-shot fault detection model named meta-analogical momentum contrast learning (MA-MOCO) for wind turbine drivetrain. By improving the momentum contrast learning (MOCO) and using the training idea of meta-learning, the one-shot fault diagnosis of wind turbines is analyzed. All the training data are used to train a basic classifier, which is used to construct the MA-MOCO model. Using the idea of meta-learning, some meta-learning tasks are constructed by randomly selecting samples from the training dataset to form support sets and query sets. The MA-MOCO is

further updated by comparing the query set with the positive and negative samples in the support set, and a simple and effectively differentiated feature space is established.

The rest of this paper is organized as follows: Section 2 introduces the basic concepts of meta-learning and MOCO. In Section 3, a one-shot fault diagnosis model named MA-MOCO for wind turbines is proposed. In Section 4, the field wind turbine datasets are input to the proposed model for training and testing; the results are analyzed. Section 5 presents the conclusion.

## 2. Momentum Contrast Learning and Meta-Learning

### 2.1. Momentum Contrast Learning

Self-supervised learning mainly uses auxiliary pretext to mine its own supervised information from large-scale unsupervised data; it uses this structured supervised information to train the network so that it can learn valuable representations for downstream tasks. CL is a kind of self-supervised learning, which constructs representations by learning to code the similarities or dissimilarities of two things and then measuring the distance between the positive and negative samples. The distance between the core idea sample and the positive sample is much larger than the distance between the sample and the negative sample:

$$\text{score}(f(x), f(x^+)) \gg \text{score}(f(x), f(x^-)) \quad (1)$$

where  $x^+$  and  $x^-$  are positive (similar) and negative (dissimilar) samples to  $x$ , respectively.  $f(\bullet)$  is the encoder model for CL.  $\text{score}$  is a function to quantify similarity, e.g., cosine similarity.

CL aims at grouping similar samples closer to and diverse samples far from each other [29]. To achieve this target, a similarity metric is used to measure how close the two embeddings are, e.g., noise contrastive estimation (NCE) [30], as follows:

$$L = E_{x, x^+, x^-} \left[ -\log \left( \frac{e^{f(x)^T f(x^+)}}{e^{f(x)^T f(x^+)} + e^{f(x)^T f(x^-)}} \right) \right] \quad (2)$$

where  $x^+$  and  $x^-$  are positive (similar) and negative (dissimilar) samples to  $x$ , respectively.

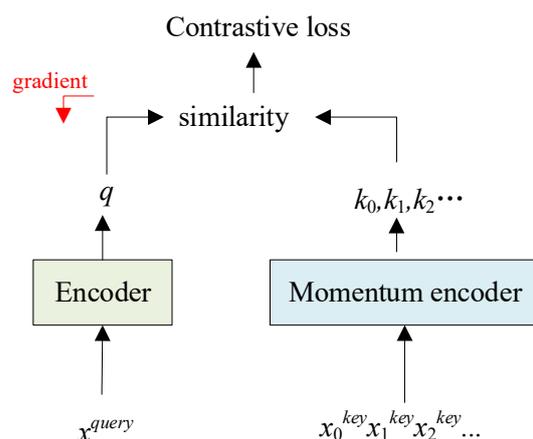
Currently, common CL models include MOCO [31], a simple framework for contrastive learning (SimCLR) [32], contrastive predictive coding (CPC) [33], and contrastive multiview coding (CMC) [21]. He et al. presented momentum contrast (MOCO) for unsupervised visual representation learning, and built a dynamic dictionary with a queue and a moving-averaged encoder [31]. Chen implemented SimCLR in the MOCO framework to establish stronger baselines that outperform SimCLR [34].

The structure diagram of MOCO is shown in Figure 1. MOCO trains a visual representation encoder by matching an encoded query  $q$  to a dictionary of encoded keys using a contrastive loss. The dictionary is built as a queue. The keys are encoded by a slowly progressing encoder, driven by a momentum update with the query encoder.

The query data are input to the encoder and the key data are input to the momentum encoder. Consider an encoded query  $q$  and a set of encoded samples  $\{k_0, k_1, k_2, \dots\}$  that are the keys of a dictionary. Assume that there is a single key (denoted as  $k^+$ ) in the dictionary that  $q$  matches. With similarity measured by a dot product, a form of a contrastive loss function, called InfoNCE [33], is as follows:

$$L_q = -\log \frac{\exp(q \cdot k_+ / \tau)}{\sum_{i=0}^K \exp(q \cdot k_i / \tau)} \quad (3)$$

where  $\tau$  is a temperature hyper-parameter. The sum is over one positive and  $K$  negative sample.



**Figure 1.** The update process of MOCO.

Define the encoder as  $f_q$  and the momentum encoder as  $f_k$ . Denote the parameter of  $f_k$  as  $\theta_k$  and those of  $f_q$  as  $\theta_q$ . Update the parameter  $\theta_q$  by back-propagation, and update the parameter  $\theta_k$  by:

$$\theta_k \leftarrow (m\theta_k + (1 - m)\theta_q) \quad (4)$$

where  $m \in [0, 1)$  is a momentum coefficient. In experiments, a relatively large momentum works much better than a smaller value, suggesting that a slowly evolving key encoder is core to making use of a queue.

## 2.2. Meta-Learning-Based Few-Shot Learning

Meta-learning is a general paradigm for few-shot learning, which involves two core processes: learning the transfer of knowledge across tasks; and rapidly adapting to new tasks. In the meta-learning training process,  $k$  data are selected from the training set as the support set  $S$ , and  $q$  data as the query set  $Q$ . If the training sample contains  $c$  categories and each category in the support set contains  $k$ -labeled data, the few-shot problem is called the  $c$ -way  $k$ -shot problem. When  $k = 1$ , this is one-shot learning.

The support set  $S$  and query set  $Q$  are extracted from the source domain data. The support set is used as a labeled sample to generate prototype features for the model, and the query set is used as a training sample to update the model; the support set and query set are to form a meta-task, and multiple meta-tasks form a training set. For a  $c$ -way  $k$ -shot problem, during the training phase,  $c$  categories are randomly selected in the training set, and  $k$  samples are selected from each category (a total of  $k \times c$  data) to construct a meta-task as the support set of the model ( $m = k \times c$ ); then, a batch of samples from the remaining data in these  $c$  categories are selected as the query set ( $n = q \times c$ ) to update the model.

During the training process, different meta-tasks are sampled for each training; thus, overall, the training contains different combinations of categories, and this mechanism enables the model to learn common parts of different meta-tasks. The models that are learned through this learning mechanism will perform better in relation to classifying when faced with new unseen meta-tasks.

## 3. Meta-Analogical Momentum Contrast Learning

In this paper, we proposed a novel few-shot fault detection model named meta-analogical momentum contrast learning (MA-MOCO) for wind turbine drivetrain. By improving the momentum contrast learning (MOCO) and using the training idea of meta-learning, the one-shot fault diagnosis of wind turbines is analyzed. The model structure is shown in Figure 2.

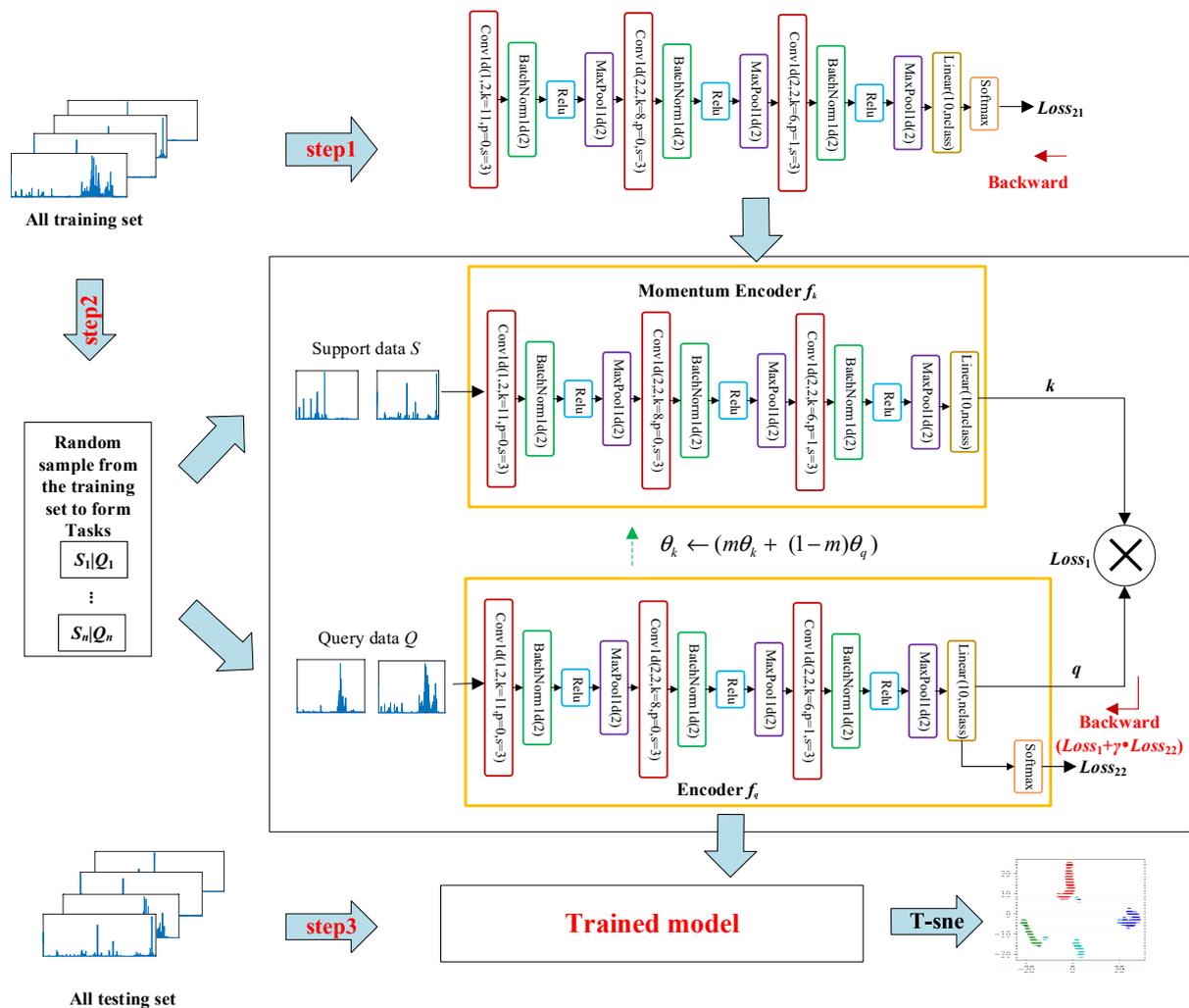


Figure 2. Model structure of MA-MOCO for wind turbines one-shot learning.

All the training data are used to train a basic classifier, which is used to construct the MA-MOCO model. Using the idea of meta-learning, some meta-learning tasks are constructed by randomly selecting samples from the training dataset to form support sets and query sets. The MA-MOCO is further updated by comparing the query set with the positive and negative samples in the support set, and a simple and effectively differentiated feature space is established.

The classifier model contains three convolution layers, three BatchNorm1d, three MaxPool1d, and one fully-connected layer. The number of neurons in each layer is marked in Figure 2. The activation functions of all the layers are rectified linear units (Relu), except for the last layer where the activation function is Softmax. All the data go through the fast Fourier transform and are then fed into the model. When a bearing or gear fails, the period  $T$  of the shock pulse is determined by the area where the defect is generated; the inverse  $f_i$  of its period reflects the fault characteristics as the fault characteristic frequency. When a bearing or gear fails, the period  $T$  of the shock pulse is determined by the area where the defect is generated; the inverse  $f_i$  of its period reflects the fault characteristics as the fault characteristic frequency. When a machine is faulty, the spectral peaks of the fault characteristic frequency will appear in its vibration spectrum; thus, the fault characteristic frequency can be used as a feature for fault diagnosis. Spectrum analysis is the most useful fault diagnosis method. The fast Fourier transform converts complex time-domain signals into easy-to-analyze frequency-domain signals, which can more clearly represent the fault characteristic frequencies. After converting the data into spectral data by the fast Fourier

transform, they can reflect the fault features more clearly and facilitate the neural network model to better learn and classify the features.

The one-shot fault diagnosis model is divided into the following steps:

In the first step, the baseline model is trained. All the training sets are input to the classifier model; the model is set as  $f$ . The base model parameters are updated with the learning rate  $lr_1$  as 0.01. Backpropagation updates are carried out according to Equation (5).

$$Loss_{21} = \sum_{x_i, y_i \sim T_i} [y_i \log f_\phi(x_i) + (1 - y_i) \log(1 - f_\phi(x_i))] \quad (5)$$

where  $y$  is the label corresponding to data  $x$ .

In the second step, the MA-MOCO model is trained. Assuming a  $c$ -way  $k$ -shot learning task,  $k$  pieces of data of each class are randomly selected from the training data as the support set  $S$ , and other  $q$  pieces of data are selected as the query set  $Q$ ; the support set and query set form a meta-learning task.  $N$  meta-learning tasks are constructed. The initial parameters of the model are selected from the trained baseline model, and each task is used to update the MA-MOCO parameters. The updated learning rate of each task is  $lr_2$  as 0.002.

The loss of the MA-MOCO model is represented in Equations (6) and (7). Equation (6) is the loss of the MOCO, and Equation (7) is the loss of classification.

$$Loss_1 = -\log \frac{\exp(q \cdot k_+ / \tau)}{\sum_{i=0}^K \exp(q \cdot k_i / \tau)} \quad (6)$$

$$Loss_{22} = \sum_{x_i, y_i \sim T_i} [y_i \log q_i + (1 - y_i) \log(1 - q_i)] \quad (7)$$

where  $q$  is the query set data,  $k$  is the support set data, and  $\tau$  is a temperature hyper-parameter.

Backpropagation updates are carried out according to Equation (8).

$$Loss_2 = Loss_1 + \gamma Loss_{22} \quad (8)$$

where  $\gamma$  is set to 0.2 because by comparison, loss  $Loss_1$  plays a more important role in updating the MA-MOCO model; the classification loss  $Loss_{22}$  only plays a fine-tuning role. In Equation (8),  $Loss_1$  is the contrast loss, and  $Loss_{22}$  is the cross-entropy loss of the classification. The two losses jointly promote the model to achieve better optimization results and are not directly related. Because of the small sample size of the known fault data and the severe sample imbalance problem, the classification loss  $Loss_{22}$  does not optimize the classification results well. Contrast loss  $Loss_1$  is a loss function with the negative sample self-discovery property, which is essential for learning high-quality self-supervised representations and effectively learning knowledge common to the data. Therefore, this paper mainly relies on the contrast loss  $Loss_1$  to update the model, and the classification loss  $Loss_{22}$  as a supplement to further improve the model accuracy; thus, the coefficient  $\gamma$  is taken as 0.2.

In the last step, the testing sets are fed into the trained model for classification and solved for accuracy. The feature embedding is visualized by t-distributed stochastic neighbor embedding (t-SNE) to test the effectiveness of the proposed model.

The complete algorithm flow is shown in Algorithm 1.

**Algorithm 1:** One-Shot Fault Diagnosis Based on MA-MOCO.

---

**Input:** Input training data  $Tr = \{(x_i, y_i)\}_{i=1}^M$ , testing data  $Te = \{x_i\}_{i=1}^I$ , classified model  $f_\phi$ , encoder  $f_q$  and momentum encoder  $f_k$ , updated learning rate  $lr_1$  of baseline and  $lr_2$  of MA-MOCO, the model parameters  $\theta$ , the momentum coefficient  $m$ , and the temperature hyper-parameter  $\tau$ .

##### (1) Pre-training baseline models #####

- 1: For each training epoch, do:
- 2: For each batch, do:
- 3:  $c_i = f(x_i)$
- 4: Backward propagation (with the learning rate as  $lr_1$ ) by Equation (5).
- 5: end

##### (2) train MA-MOCO models #####

- 6: Randomly draw data from  $Tr$  to form  $N$  tasks, each task containing  $k$  support sets and  $q$  query sets, to form  $\{(S_1, Q_1), (S_2, Q_2), \dots, (S_n, Q_n)\}$
- 7: For each training, do:
- 8: For each batch, do:
- 9:  $q_i = f_q(Q_i)$ ,  $k_i = f_k(S_i)$ ,  $c_i = \text{Softmax}(q_i)$
- 10: Update parameters of the encoder by:
 
$$\text{Loss}_2 = -\log \frac{\exp(q_i \cdot k_i / \tau)}{\sum_{i=0}^k \exp(q_i \cdot k_i / \tau)} + \gamma \sum_{x_i, y_i \sim T_i} [y_i \log c_i + (1 - y_i) \log(1 - c_i)]$$
- 11: Backward propagation of the encoder:  $\theta_q \leftarrow \theta_q - lr_2 \nabla_{\theta} \text{Loss}_2(\theta_q)$
- 12: Backward propagation of the momentum decoder:  $\theta_k \leftarrow m\theta_k + (1 - m)\theta_q$
- 13: end

##### (3) testing results and t-SNE #####

- 14: For the test set, calculate  $c_{T_i} = f(Te_i)$ , calculate the accuracy, and draw the t-SNE diagram.

**Output:** testing results.

---

The update function used within this model is Adam, with 100 training epochs for the pre-trained base model and 200 training epochs for the MA-MOCO update. The batch size is 32. Based on most CL experiences, the momentum coefficient  $m$  is set to 0.999; the temperature hyper-parameter  $\tau$  is set to 0.5.

#### 4. Case Analysis

In this section, one-shot learning cases are analyzed to verify the advantages of the proposed model, including the one-shot case of the wind turbine generator bearing and gearbox. To further validate the proposed model of MA-MOCO, we will compare it with some transfer learning algorithms, such as the Siamese net [16] and the model-agnostic meta-learning (MAML) net [34]. To make a fair comparison, we used the same datasets, the same data preprocessing methods (fast Fourier transform), the same classified model, the same epochs, and the same learning rates.

In general, the data of proper health operation are easy to obtain; however, failure data are difficult to obtain. This paper analyzes the classification results of the proposed model when there are only 5, 10, and 20 samples of fault data for training. Two case studies with one-shot settings are conducted. In order to assess the accuracy of all the proposed algorithms, data from 240 wind turbines operating under different conditions with different faults that are not involved in the training are fed into the model in this paper. The accuracy of the model is evaluated by comparing the predicted categories with the corresponding labels of the data and solving for the classification accuracy.

In this paper, the feature embeddings are visualized by t-SNE to test the validity of the proposed model. MAML and MA-MOCO have a classification layer; therefore, both use the output of the Softmax activation function to plot t-SNE. Siamese does not have a classification layer and relies on distance to determine the categories. Hence, Siamese plots t-SNE using the distance metric between the test data and each category of data.

##### 4.1. Case One: One-Shot Fault Diagnosis of the Generator Bearings for Wind Turbines

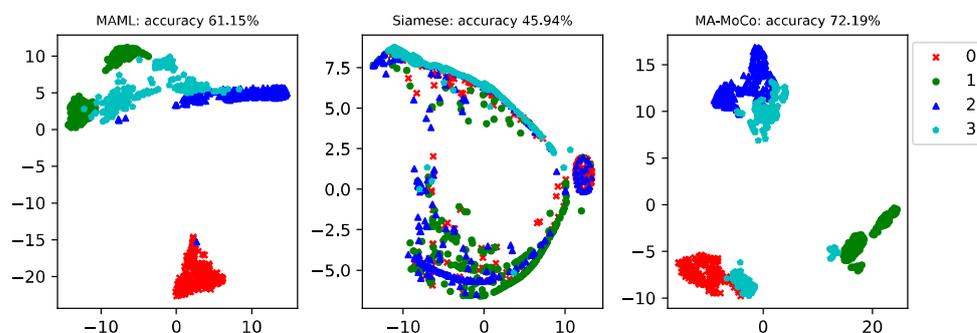
In this case, a one-shot fault diagnosis of the generator bearings for wind turbines is conducted. The available samples are shown in Table 1. The generator bearing data for the

wind turbine includes proper health operation data and three types of faults; each category contains 260 data. The latter two faults are compound faults. In the actual operating conditions, wind turbines have mostly compound faults; this paper studies the one-shot problem of compound faults, which has better engineering significance.

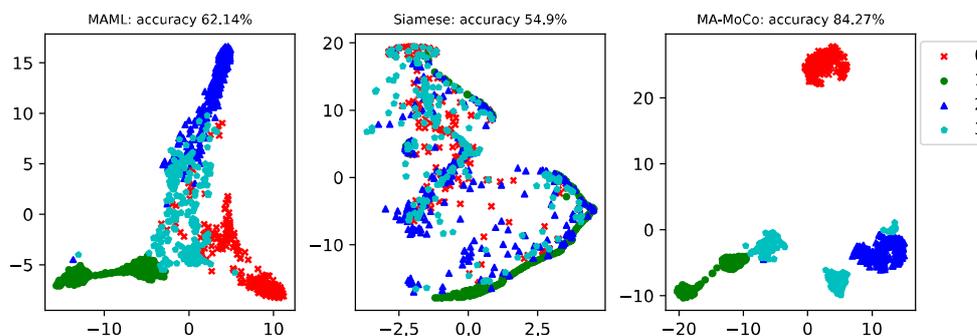
**Table 1.** Fault description of the generator bearings for wind turbines.

	Fault Type	Label	Number of Samples from the Training Set	Number of Samples from the Testing Set
Healthy	No faults	0	100	240
Fault 1	Outer ring failure	1	5 or 10 or 20	240
Fault 2	Inner ring failure + Outer ring failure	2	5 or 10 or 20	240
Fault 3	Inner ring failure + Rolling failure + Cage failure	3	5 or 10 or 20	240

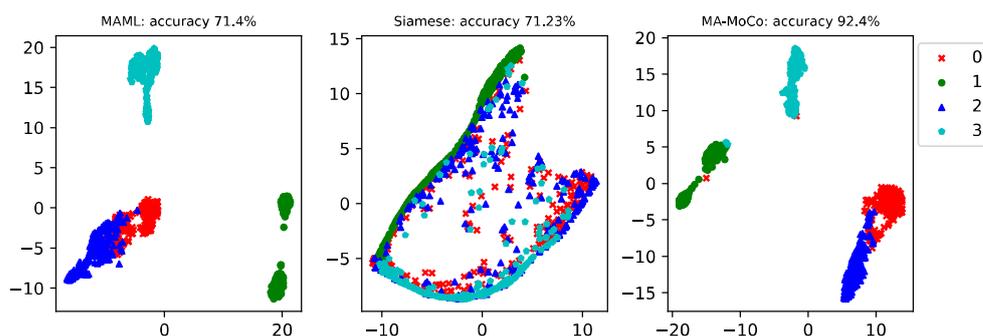
We analyzed the results when the failure data of the training set are only 5, 10, and 20, respectively. The latter 240 data in different operating conditions are the testing data. The model can also be further tested under different operating conditions. The Siamese net [16] and the MAML net [34] are used for comparison with the proposed model. The final t-SNE is shown in Figures 3–5. The classification accuracy of each algorithm is shown at the top of each graph. Figures 3–5 show the classification results and t-SNE plots of the model for the sample sizes of 5, 10, and 20 for each class of fault data in the training set, respectively. In the t-SNE plots in Figures 3–5, t-SNE reduces the classification result data to two dimensions, and the horizontal and vertical coordinates represent the two dimensions of the data after t-SNE reduction, respectively.



**Figure 3.** A 4-way 1-shot diagnosis of the generator bearing data when the training set of each class has only 5 fault samples.

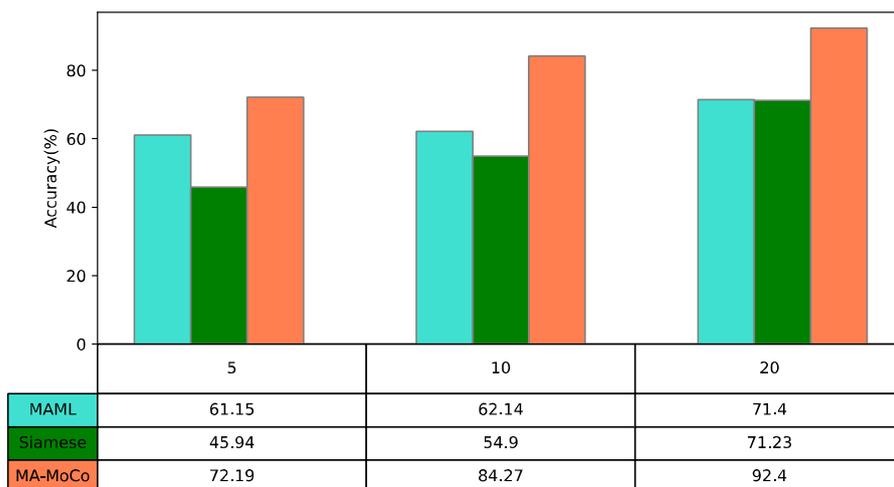


**Figure 4.** A 4-way 1-shot diagnosis of the generator bearing data when the training set of each class has only 10 fault samples.



**Figure 5.** A 4-way 1-shot diagnosis of the generator bearing data when the training set of each class has only 20 fault samples.

The fault classification accuracy of the generator bearings for wind turbines using different algorithms is shown in Figure 6. In order to analyze the influence of different sample sizes in the training set on the diagnostic accuracy of the model, this paper investigates the diagnostic accuracy of the model when the sample sizes of each type of fault in the training set are 5, 10, and 20, respectively. It can be seen that the higher the training sample size, the higher the classification accuracy of the model. The average accuracy of MA-MOCO is 18.05% higher than MAML and 25.59% higher than Siamese net.



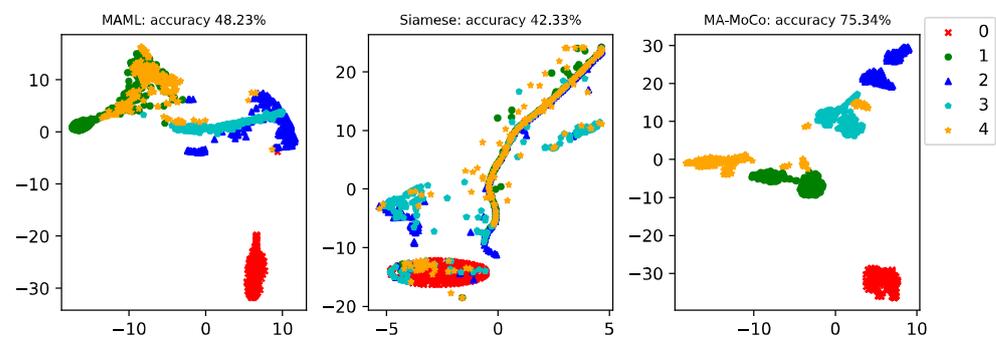
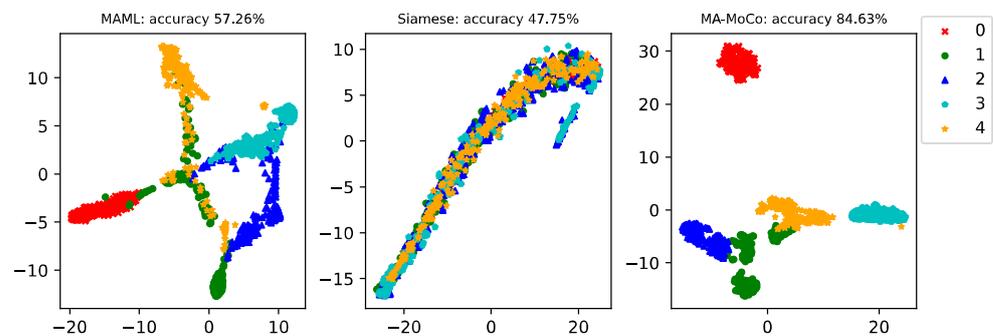
**Figure 6.** A 4-way 1-shot diagnosis of the accuracy of the generator bearing data when the training set of each class has 5, 10, and 20 fault samples.

#### 4.2. Case Two: One-Shot Fault Diagnosis of the Wind Turbine Gearbox

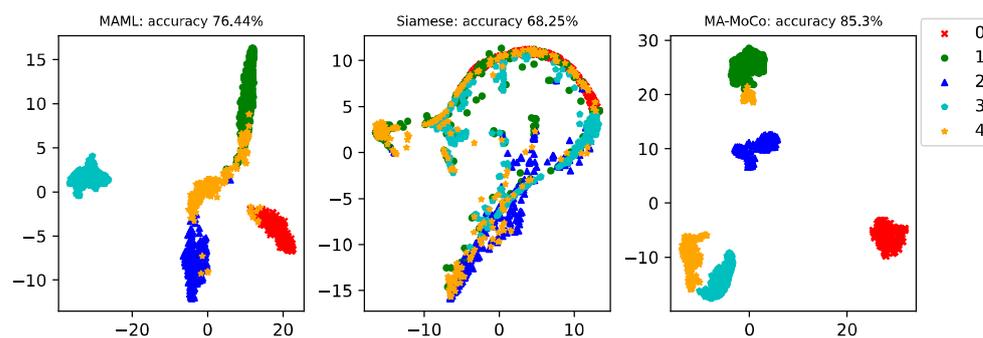
This case focuses on a one-shot fault diagnosis of the gearbox of wind turbines. The available gearbox samples are shown in Table 2. The samples contain one category of proper health operation data and four categories of fault data; each category contains 260 data. Fault 2 is a compound fault. We analyzed the results when the failure data of the training set are only 5, 10, and 20, respectively. The latter 240 data that are in different operating conditions are the testing data. The model can also be further tested for different operating conditions. The Siamese net [16] and the MAML net [34] are used for comparison with the proposed model. The final t-SNE is shown in Figures 7–9. The classification accuracy of each algorithm is shown at the top of each graph. Figures 7–9 show the classification results and t-SNE plots of the model for the sample sizes of 5, 10, and 20 for each class of fault data in the training set, respectively. In the t-SNE plots in Figures 7–9, t-SNE reduces the classification result data to two dimensions; the horizontal and vertical coordinates represent the two dimensions of the data after t-SNE reduction, respectively.

**Table 2.** Fault description of the wind turbine gearbox.

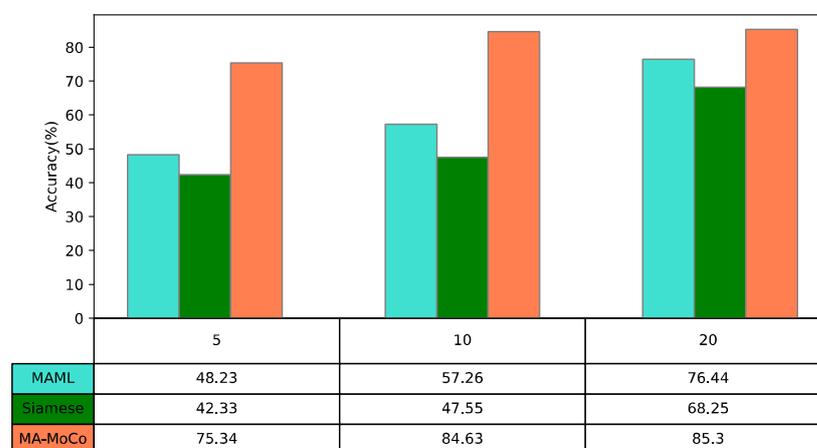
	Fault Type	Label	Number of Samples from the Training Set	Number of Samples from the Testing Set
Healthy	No faults	0	100	240
Fault 1	Spalling of the gears in the intermediate shaft	1	5 or 10 or 20	240
Fault 2	Broken teeth of the gears in the intermediate and high-speed shaft	2	5 or 10 or 20	240
Fault 3	Broken teeth of the gears in the high-speed shaft	3	5 or 10 or 20	240
Fault 4	Broken teeth of the gears in the intermediate shaft	4	5 or 10 or 20	240

**Figure 7.** A 5-way 1-shot diagnosis of the generator bearing data when the training set of each class has only 5 fault samples.**Figure 8.** A 5-way 1-shot diagnosis of the generator bearing data when the training set of each class has only 10 fault samples.

The fault classification accuracy of the gearboxes for the turbines using different algorithms is shown in Figure 10. In order to analyze the influence of different sample sizes in the training set on the diagnostic accuracy of the model, this paper investigates the diagnostic accuracy of the model when the sample sizes of each type of fault in the training set are 5, 10, and 20, respectively. It can be seen that the higher the training sample size, the higher the classification accuracy of the model. The average accuracy of MA-MOCO is 21.12% higher than MAML and 29.05% higher than Siamese net.



**Figure 9.** A 5-way 1-shot diagnosis of the generator bearing data when the training set of each class has only 20 fault samples.



**Figure 10.** A 5-way 1-shot diagnosis of the accuracy of the wind turbine gearboxes data when the training set of each class has 5, 10, and 20 fault samples.

#### 4.3. Discussion of the Results

Supervised learning models trained by labels often learn only some task-specific knowledge, but not generic knowledge, making the feature representations difficult to transfer to other tasks. The data provide far more information than sparse labels, and self-supervised learning directly uses the data themselves as supervised information to learn feature representations of the sample data and use them for downstream tasks. So, self-supervised learning is better able to learn generic knowledge to improve the characterization of the model and adapt it to the migration of different working conditions. Contrast learning, as a self-supervised method, uses contrast loss, which makes representations from the same data (positive examples) closer and representations from different data (negative examples) more dispersed. Momentum contrast learning (MOCO) uses dictionary queuing and momentum updates to avoid losing consistency in features due to drastic changes in the encoder, while also keeping the encoder in a constant state of being updated. Meta-learning, also known as ‘learning to learn’, equips models with certain prior knowledge or learning skills by learning multiple tasks so that they can better generalize and learn new tasks.

How to learn transferable feature representations from a limited number of samples is a key challenge in few-shot learning. In this paper, MOCO is added to few-shot learning to improve the representational power by learning generic knowledge of different categories. At the same time, the original few-shot classification further optimizes the model to achieve a complementary role. The few-shot model is trained through the task-based training idea in meta-learning. Support sets are input to the model, and the query sets are used to update the model to improve its ability to adapt to different working conditions. By combining

MOCO and meta-learning, the model achieves better results and is able to adapt to the different operating conditions of wind turbines.

## 5. Conclusions

Fault diagnosis of wind turbines plays an important role in improving the reliability of wind turbines. However, wind turbines are not allowed to operate with faults. If an abnormality is detected during operation, the turbine will immediately shut down; thus, it is difficult to obtain a sample of faults. When the fault samples are small, ordinary deep learning can fall into overfitting, resulting in a low diagnostic accuracy. Therefore, few-shot fault diagnosis of wind turbines is of great practical importance.

In reality, the operating conditions of wind turbines change randomly; often, multiple faults occur simultaneously. Sample data for different fault types under different operating conditions are difficult to obtain. Therefore, it is necessary to study fault diagnosis when the fault samples are very few, i.e., one-shot learning.

A novel method based on meta-analogical momentum contrast learning (MA-MOCO) is proposed in this paper to solve the problem of very few samples of wind turbine failures, especially one-shot. By improving the momentum contrast learning (MOCO) and using the training idea of meta-learning, the one-shot fault diagnosis of wind turbine drivetrain is analyzed.

This paper analyzes the diagnosis of one-shot tasks of single and compound faults in wind turbine generator bearings and wind turbine gearboxes. It is also compared with other algorithms to verify the accuracy and stability of the proposed method. The results are also presented by t-SNE. All the training data are used to train a basic classifier, which is used to construct the MA-MOCO model. Using the idea of meta-learning, some meta-learning tasks are constructed by randomly selecting samples from the training dataset to form support sets and query sets. The MA-MOCO is further updated by comparing the query set with the positive and negative samples in the support set; thus, a simple and effectively differentiated feature space is established.

The proposed model MA-MOCO combines the advantages of CL in the establishment of a differentiated feature space and meta-learning in learning new tasks. In order to assess the accuracy of all the proposed algorithms, data from 240 wind turbines operating under different operating conditions with different faults that are not involved in the training are fed into the model in this paper. The accuracy of the model is evaluated by comparing the predicted categories with the corresponding labels of the data and solving for the classification accuracy. The results show that the proposed model MA-MOCO is superior to the Siamese net and model-agnostic meta-learning (MAML) net in the classification accuracy of wind turbine datasets. The proposed model can better solve the problems of the variable operating conditions and composite diagnosis of wind turbines.

**Author Contributions:** Methodology, X.L.; writing—original draft preparation, X.L.; writing—review and editing, H.G. and Y.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

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

## References

1. Wang, Y.; Yao, Q.; Kwok, J.T.; Lionel, M.N. Generalizing from a few examples: A survey on few-shot learning. *ACM Comput. Surv.* **2020**, *53*, 1–34. [[CrossRef](#)]
2. Hu, T.; Tang, T.; Lin, R.; Chen, M.; Han, S.; Wu, J. A simple data augmentation algorithm and a self-adaptive convolutional architecture for few-shot fault diagnosis under different working conditions. *Measurement* **2020**, *156*, 107539. [[CrossRef](#)]
3. Zheng, T.; Song, L.; Wang, J.; Teng, W.; Xu, X.; Ma, C. Data synthesis using dual discriminator conditional generative adversarial networks for imbalanced fault diagnosis of rolling bearings. *Measurement* **2020**, *158*, 107741. [[CrossRef](#)]

4. Ren, Z.; Zhu, Y.; Yan, K.; Chen, K.; Kang, W.; Yue, Y.; Gao, D. A novel model with the ability of few-shot learning and quick updating for intelligent fault diagnosis. *Mech. Syst. Signal Process.* **2020**, *138*, 106608. [[CrossRef](#)]
5. Schlegl, T.; Seebeck, P.; Waldstein, S.M.; Ursula, S.; Langs, G. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In Proceedings of the 2017 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; Lecture Notes in Computer Science. Volume 10265, pp. 146–157.
6. Akcay, S.; Atapour-Abarghouei, A.; Breckon, T.P. GANomaly: Semi-supervised anomaly detection via adversarial training. In Proceedings of the 2018 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 July 2018; Lecture Notes in Computer Science. Volume 11363, pp. 622–637. [[CrossRef](#)]
7. Kingma, D.P.; Welling, M. Auto-encoding variational bayes. *arXiv* **2014**, arXiv:1312.6114.
8. Makhzani, A.; Shlens, J.; Jaitly, N.; Goodfellow, L.; Frey, B. Adversarial autoencoders. *arXiv* **2016**, arXiv:1511.05644.
9. Hospedales, T.M.; Antoniou, A.; Micaelli, P.; Storkey, A. Meta-learning in neural networks: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, 3079209. [[CrossRef](#)]
10. Wu, J.; Zhao, Z.; Sun, C.; Yan, R.; Chen, X. Few-shot transfer learning for intelligent fault diagnosis of machine. *Measurement* **2020**, *166*, 108202. [[CrossRef](#)]
11. Wang, D.; Zhang, M.; Xu, Y.; Lu, W.; Yang, J.; Zhang, T. Metric-based meta-learning model for few-shot fault diagnosis under multiple limited data conditions. *Mech. Syst. Signal Process.* **2021**, *155*, 107510. [[CrossRef](#)]
12. Feng, Y.; Chen, J.; Zhang, T.; He, S.; Xu, E.; Zhou, Z. Semi-supervised meta-learning networks with squeeze-and-excitation attention for few-shot fault diagnosis. *ISA Trans.* **2021**, *120*, 383–401. [[CrossRef](#)]
13. Su, H.; Xiang, L.; Hu, A.; Xu, Y.; Yang, X. A novel method based on meta-learning for bearing fault diagnosis with small sample learning under different working conditions. *Mech. Syst. Signal Process.* **2022**, *169*, 108765. [[CrossRef](#)]
14. Wang, S.; Wang, D.; Kong, D.; Wang, J.; Li, W.; Zhou, S. Few-shot rolling bearing fault diagnosis with metric-based meta learning. *Sensors* **2020**, *20*, 6437. [[CrossRef](#)] [[PubMed](#)]
15. Koch, G.; Zemel, R.; Salakhutdinov, R. Siamese neural networks for one-shot image recognition. In Proceedings of the 2015 International Conference on Machine Learning, Lille, France, 6–11 July 2015; Volume 1, pp. 1–8.
16. Zhang, A.; Li, S.; Cui, Y.; Yang, W.; Dong, R.; Hu, J. Limited data rolling bearing fault diagnosis with few-shot learning. *IEEE Access* **2019**, *7*, 110895–110904. [[CrossRef](#)]
17. Cai, Q.; Pan, Y.; Yao, T.; Yan, C.; Mei, T. Memory matching networks for one-shot image recognition. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4080–4088. [[CrossRef](#)]
18. Tran, K.; Sato, H.; Kubo, M. Memory Augmented Matching Networks for Few-Shot Learnings. *Int. J. Mach. Learn. Comput.* **2019**, *9*, 743–748. [[CrossRef](#)]
19. Sung, F.; Yang, Y.; Zhang, L.; Xiang, T.; Philip, H.S.T.; Timothy, M.H. Learning to compare: Relation network for few-shot learning. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 1199–1208. [[CrossRef](#)]
20. Garcia, V.; Bruna, J. Few-shot learning with graph neural networks. In Proceedings of the 2017 International Conference on Machine Learning, Sydney, Australia, 6–11 August 2017; Volume 1, pp. 1–13. [[CrossRef](#)]
21. Tian, Y.; Krishnan, D.; Isola, P. Contrastive Multiview Coding. In *European Conference on Computer Vision*; Springer: Cham, Switzerland, 2020; Lecture Notes in Computer Science; Volume 12356, pp. 776–794.
22. Yang, J.; Chen, H.; Yan, J.; Chen, X.; Yao, J. Towards better understanding and better generalization of few-shot classification in histology images with contrastive learning. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 19–24 June 2022. [[CrossRef](#)]
23. Majumder, O.; Ravichandran, A.; Maji, S.; Polito, M.; Soatto, S. Revisiting Contrastive Learning for Few-Shot Classification. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021. [[CrossRef](#)]
24. Liu, C.; Fu, Y.; Xu, C.; Yang, S.; Li, J.; Wang, C.; Zhang, L. Learning a Few-shot Embedding Model with Contrastive Learning. *Natl. Conf. Assoc. Adv. Artif. Intell.* **2021**, *35*, 8635–8643.
25. Kim, Y.; Shin, J.; Yang, E.; Hwang, S.J. Few-shot Visual Reasoning with Meta-analogical Contrastive Learning. *Adv. Neural Inf. Processing Syst.* **2020**, *33*, 16846–16856.
26. Liu, H.; Zhang, F.; Zhang, X.; Zhan, S.; Zhang, X. An Explicit-Joint and Supervised-Contrastive Learning Framework for Few-Shot Intent Classification and Slot Filling. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 1945–1955. [[CrossRef](#)]
27. Chen, X.; Yao, L.; Zhou, T.; Dong, J.; Zhang, Y. Momentum Contrastive Learning for Few-Shot COVID-19 Diagnosis from Chest CT Images. *Pattern Recognit.* **2021**, *113*, 107826. [[CrossRef](#)]
28. Ding, Y.; Zhuang, J.; Ding, P.; Jia, M. Self-supervised pretraining via contrast learning for intelligent incipient fault detection of bearings. *Reliab. Eng. Syst. Saf.* **2022**, *218*, 108126. [[CrossRef](#)]
29. Jaiswal, A.; Babu, A.R.; Zadeh, M.Z.; Banerjee, D.; Makedon, F. A survey on contrastive self-supervised learning. *Technologies* **2021**, *9*, 2. [[CrossRef](#)]
30. Gutmann, M.; Hyvärinen, A. Noise-contrastive estimation: A new estimation principle for unnormalized statistical models. *J. Mach. Learn. Res.* **2010**, *9*, 297–304.

31. He, K.; Fan, H.; Wu, Y.; Xie, S.; Girshick, R. Momentum Contrast for Unsupervised Visual Representation Learning. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 9729–9738. [\[CrossRef\]](#)
32. Chen, T.; Kornblith, S.; Norouzi, M.; Hinton, G. A Simple Framework for Contrastive Learning of Visual Representations. In Proceedings of the 2020 International Conference on Machine Learning, Virtual, 12–18 July 2020; pp. 1597–1607.
33. Oord, A.; Li, Y.; Vinyals, O. Representation Learning with Contrastive Predictive Coding. *arXiv* **2018**, arXiv:1807.037482018.
34. Finn, C.; Abbeel, P.; Levine, S. Model-agnostic meta-learning for fast adaptation of deep networks. In Proceedings of the 2017 International Conference on Machine Learning, Sydney, Australia, 6–11 August 2017; PMLR 70. pp. 1126–1135.