IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 29 August 2022, accepted 11 September 2022, date of publication 14 September 2022,
date of current version 20 September 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3206528

RESEARCH ARTICLE

GMRNet: A Novel Geometric Mean Relation Network for Few-Shot Very Similar Object Classification

CANAN TASTIMUR[1] AND ERHAN AKIN[2], (Member, IEEE)
[1]Computer Engineering Department, Erzincan Binali Yildirim University, 24100 Erzincan, Turkey
[2]Computer Engineering Department, Firat University, 23190 Elâziğ, Turkey

Corresponding author: Canan Tastimur (ctastimur@erzincan.edu.tr)

ABSTRACT With the widespread use of Deep Learning (DL), the use of DL has increased to provide a solution to the problem of object recognition and classification. In addition to classifying many different types of objects, the Deep Metrics Learning(DML) technique is effective in classifying objects that are visually very similar to each other. In this study, a novel Relation Network (RN) based DML has been designed to classify objects in two different datasets we created. We distinguished groups of objects that had a high degree of similarity to each other. These objects have been categorized using few-shot learning(FSL) since they are quite similar to one another. The impact of changing the number of classes and samples in the database on the network's performance has been studied. It is shown how the network's accuracy varies depending on the N-way (number of classes) and K-shots (number of samples) combinations used in its design. Additionally, the performance of the network has improved by an average of 15% thanks to the contribution of the recently introduced geometric mean module to the RN in our study. The accuracy rate of our recommended RN in screw and spare parts datasets is 96.1% and 92.3%, respectively. The first dataset consists of 1800 screw images with 18 classes, while the second dataset consists of 4100 spare parts images with 20 classes. The effectiveness of our method is expressed by the two datasets that we have extensively experimentally studied.

INDEX TERMS Classification, deep metric learning, few shot learning, relation network.

I. INTRODUCTION

DL is used to classify objects in many areas such as health, industry, security, and social networking, and very good results are produced with DL. However, the high visual similarity of the classified objects to each other makes classification difficult. The general features of datasets consisting of objects that are highly similar to each other are as follows: the similarity between different classes is high; the difference within the same class is high, and the number of samples in the dataset is low. For these reasons, it is necessary to classify objects by taking the similarity criteria as criteria. The similarity between embedded features in the images is calculated with the distance learning function [1]. Some studies on object classification using DML are mentioned.

Image segmentation was performed with DML and few-shot learning [2]. The CNN network was used as a feature extractor of the VGG16 network and contrastive loss was used as a loss function. With 120,000 iterations and SGD optimizer parameters, 1-way 1-shot accuracy was 45.8%. In another study [3], deep feature mapping learning was proposed for low-dimensional person image embedding. The feature presentation and distance metric were combined in the proposed architecture. The GoogLeNet architecture was used. The significance of the proposed work, the addition of the global loss function to the objective function, and the comparison of positive and negative samples were assigned

The associate editor coordinating the review of this manuscript and approving it for publication was cheng Chin.

a weight value based on the positive double stiffness. The method was tested on three large datasets. The accuracy rates produced by the datasets were 35.67%, 54.76%, and 63.66%, respectively. The similarity criterion on two different datasets was calculated using deep adversarial metric learning, the Triplet network, and synthetic negative data [4]. For face and kinship verification, discriminative DML was proposed [5]. It was tested on five different datasets to extract general information from multiple features. A Mahalanobis distance loss function was used and 94.8% performance was achieved. In another application area, triplet networks and DML were used for ship recognition [6]. 60% success was achieved in the classification of 3965 ship imagery.

In [7], they aimed to label product description data with triplet loss. The dataset contained four images for each class. 81.2% success was achieved by using the parameters of the ResNet50 architecture. Two-way DML was proposed to find the relationship between image attributes and image tags [8]. After the image embeddings were calculated with the kNN method, the distance between the image embeddings and the label embedding was calculated. The proposed method, consists of three stages: two-way distance metric, reconstruction method and classification method. The first stage was learning the methods between images and labels; regularization of label embeddings was the second stage, and classification was the last stage. A success rate of 66.28% was achieved.

Classification was made in two different datasets with few-shot learning, zero-shot learning and relation network [9]. In another study [10], few-shot learning and RN classified four different datasets. The suggested method consisted of an embedding module, an attention module and a relation module. With the combination of RN and localized graph convolution, breast cancer was classified with a rate of 86.29% [11]. Person re-identification was implemented with the RN [12]. 88.9%, 75.6%, and 78.6% of the results were produced on three different datasets. The same method, hyperspectral image classification, was tried on three different datasets [13]. This study was also tested on three different datasets. 90.49%, 98.03%, and 89.20% were obtained.

The goal of this research is to classify object groups that are visually very similar to one another. DML aids in the calculation of object similarity. Designing a convolution-based RN with DML has resulted in the creation of a new architecture. The model's performance is assessed by using the proposed method on two different real-world datasets. In our study, we have created datasets with 1800 screw images with 18 classes and 4100 spare parts images with 20 classes. The proposed method is calculated by taking the geometric mean of the relationship between the image embeddings into account. With few-shot learning, a new geometric mean-based relation network (GMRNet) is designed in this study. The geometric mean module is added to the standard RN, which improves performance. The number of data samples available in FSL is limited. The designed network has been tested with combinations of N = 3,5,7,10 and K = 5,10 values.

The following are the study's main contributions:

- The GMRNet is intended to classify objects that have a high degree of similarity. The newly created network is innovative and effective. In terms of network performance, it is discovered to be more successful than [9], [10], [11], and 12].
- Two different real datasets have been created to test the performance of the designed new network. Both datasets were generated from real-world industry data. The majority of the other studies in the literature utilized preexisting datasets.
- It will be practical for industrial processes to distinguish between highly similar objects in the datasets produced by our study using computer vision algorithms. Finding the correct item without becoming confused with similar ones like broken, rusted, or worn-out spare parts, screws, or bolts would be useful. Additionally, this established method is easily adaptable to a wide variety of object types.
- Furthermore, we added a disposition module to the standard RN model. The architecture formed by the addition of this module is new. The performance of the new RN model, GMRNet, has increased from 86.5% to 96.1% and from 72.8% to 92.3% in screw and spare datasets, respectively. As a result, the accuracy rate has increased by 10% and 20% in two datasets when compared to standard RN model.

II. DEEP METRIC LEARNING

Machine learning algorithms use a set of rules or complex functions to find the output labels corresponding to the objective inputs in the data. Metric learning, on the other hand, seeks to learn the similarity function from data [14], [15], [16], [17]. Metric learning shortens the distance between feature vectors corresponding to different samples while increasing the distance between feature vectors corresponding to the same class. This is how it intends to learn data feature vectors. The distance metric is a new data format that uses sample similarity to provide a more meaningful and powerful discriminating model.

Traditional machine learning approaches are limited in their ability to process raw data. Deep learning is not required for preprocessing or feature extraction. The DL teaches itself high-level features. Euclid, Mahalanobis, Matusita, Bhattacharyya, and Kullback-Leibler distances are the most commonly used similarity measures for data classification [1]. DML uses deep architectures to learn from raw data and provides embedded feature similarity finding via nonlinear subspace learning [18], [19]. It is proportional to the distance between samples. As shown in Figure 1, it narrows the gap between similar samples while widening the gap between different objects. The deep learning metric loss function is used to accomplish this.

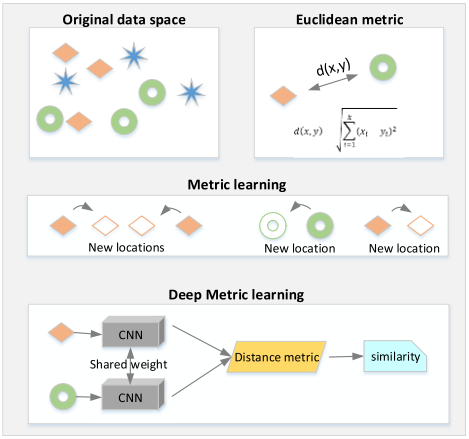


FIGURE 1. The framework of deep metric learning [1].

Although DML focuses on the metric loss function, informed sample selection also contributes significantly to classification. To contribute to the network's examples, distinct training examples should be supplied to the network. As a result, preprocessing and sample selection should be performed prior to applying DML to increase the network's success.

Let $K = \{(x_a, y_a)\}, a\epsilon[1, 2, \ldots, n]$ be a dataset, where the n parameter is number of images; assume that (x_a, y_a) is an a-th image, and the total number of images in the dataset is C, i.e., $y_a\epsilon[1, 2, \ldots, C]$. In Figure 2, $x_a \in \mathbb{R}^D$ image's related feature vector is $f(x_a)$, where $f : \mathbb{R}^D \rightarrow \mathbb{R}^d$ is a differentiable deep network with θ parameters. D represents image size and d refers to feature size. If we calculate the distance between two images with Euclidean distance, we get $D_{ab} = \|f(x_a) - f(x_b)\|^2$. Deep features that correlate to the x_a and x_b images are the $f(x_a)$ and $f(x_b)$ values. The Euclidean distance function can be substituted with other distance functions.

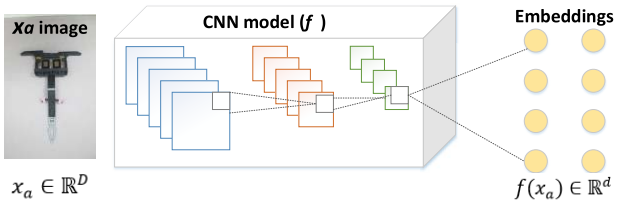


FIGURE 2. A training approach based on DML and CNN.

III. METHODOLOGY

A. PROBLEM DEFINITION

The train set and the test set in FSL-based classification are denoted by the D_{meta_train} and D_{meta_test}, respectively. Classes and examples are included in each dataset. FSL datasets contain $K-$ labeled data for each of the N classes [20], [21], [22]. The name of this issue is N-way K-shot. NxK is the maximum number of training examples. The training phase is challenging as a result. N classes from the D_{meta_train} are randomly chosen for each training iteration. Randomly chosen training samples are taken from each class [23]. Hence, the training set is expressed in (1). A query set is created by randomly selecting R from samples other than the samples selected for the training set [24], [25].

$$T = \left\{T_a^b; a = 1, \ldots, Nb = 1, \ldots, K\right\} \quad (1)$$

$$Q = \{q_m, l_m; m = 1, \ldots, R\} \quad (2)$$

T_a^b is the $b-th$ sample in the class a in (1). $l_m \in [1, \ldots, N]$ is the label of sample q_m in (2).

B. NOVEL RELATION NETWORK DESIGN

The aim of this study is to classify two different real datasets. A new GMRNet is designed with FSL and DML. GMRNet consists of an embedding module, a disposition module, and a relation module. It has been determined that the number of classes and the number of samples have an effect on network performance by combining them with different options. Figure 4 represents an overview of the proposed network. The relationship between the image pairs in the feature space is calculated with the CNN classifier. RNs basically have a working principle like regression.

1) THE EMBEDDING MODULE

This module's goal is to extract distinguishing features from the input data [26], [27]. It is made up of Convolution and Activation layers. The network in Figure 3 that we designed in our study produces the feature map $f_\varphi(x)$ of a given x image. As the images from the designed network are processed, $g_\varphi(x)$ is obtained.

$$f_\varphi(x) = w_f \times x \quad (3)$$

$$g_\varphi(x) = w_g \times f_\varphi(x) \quad (4)$$

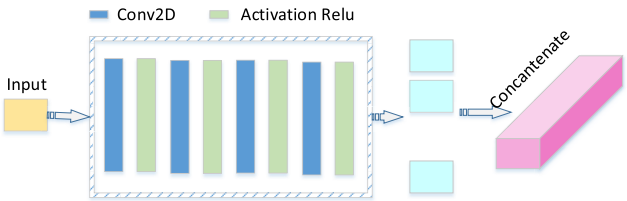


FIGURE 3. Schematic of the embedding module on GMRNet.

2) THE DISPOSITION MODULE

The samples x_q from the query set Q and x_{ij} from the training set T are obtained using $f_\varphi(x)$. Before passing the obtained x_q and x_{ij} to the relation module, the geometric mean values are calculated as in (5) and x_m is obtained. Geometric averages represent the central disposition in the set of points used as input. As a result, the relationship between the two inputs is determined. The relation module $g_\varphi(x)$ receives the x_m, x_q, and x_{ij} embeddings. A similarity scalar value in the range of 0-1 is generated based on the inputs. As in Figure 5, a relation

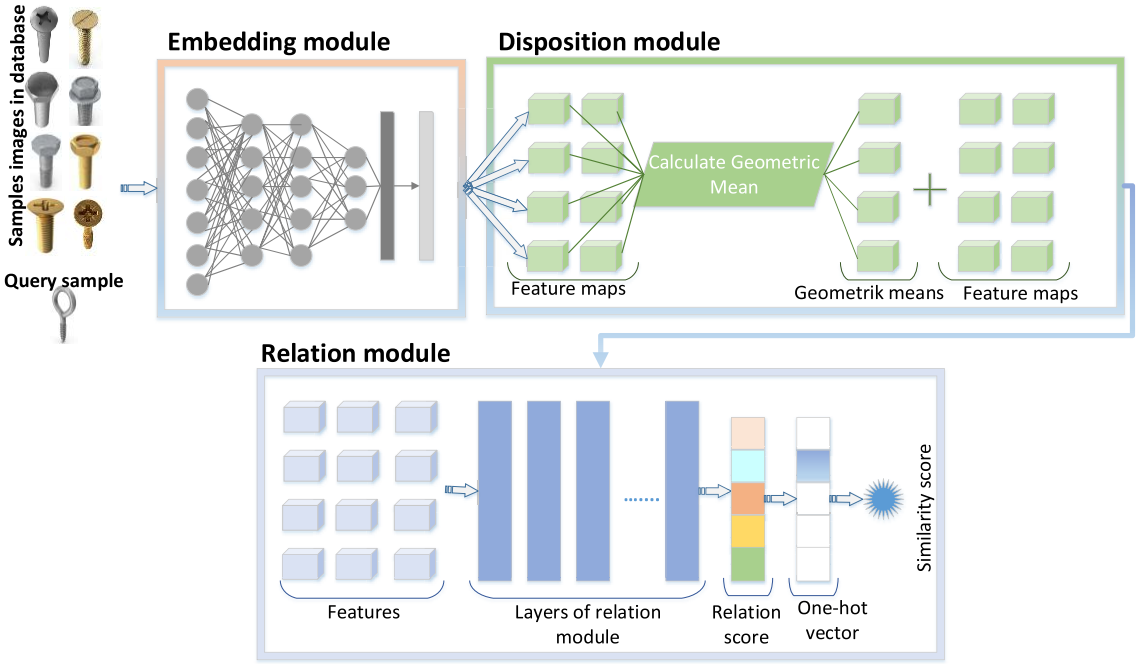


FIGURE 4. The overview of GMRNet.

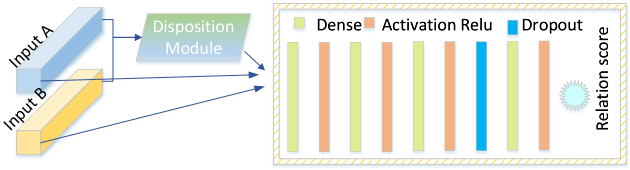


FIGURE 5. Schematic of the relation module on GMRNet.

score is obtained after the input values are passed through the generated CNN network.

$$m_\varphi(x) = \prod_{i=1}^{k} f_\varphi(x)^{1/k}, \quad k = \text{input size} \tag{5}$$

3) THE RELATION MODULE

The goal of this module is to calculate the degree of similarity between samples and to generate a correlation result [28]. Dense, Activation Relu, and Dropout layers were used in this module. It generates the r_w value in the relation module while calculating the similarity between the visual features and the geometric mean features. The higher the similarity score, the more similar the compared objects are; otherwise, the objects are different. Figure 5 depicts the relation module design used in our study. The $s_{g,m}$ similarity score is formed after concatenating $g_\varphi(x)$ and $m_\varphi(x)$ in the embedding module. The term $s_{g,m}$ is defined as follows. The $C(,)$ in (6) symbolizes the concatenation function.

$$s_{g,m} = r_w(C(g_\varphi(x_t), m_\varphi(x_t))), t = 1, 2, \ldots, k \tag{6}$$

TABLE 1. Algorithm of proposed GMRNet.

ALGORITHM 1: TRAINING PROCESS OF GMRNET

Input: E training iterations, γ train size, b batch size, lr learning rate, M geometric mean, $f_\varphi(x)$ feature embedding, w_g and w_f designed network parameters, r_w relation module, x input image, m training sample of x image

Output: $\acute{w}_g, \acute{w}_d, \acute{r}_w$ optimized training parameters for embedding, disposition, and relation modules

for $e = 0, 1, 2, \ldots, E-1$ **do**
  for $i = 0, 1, 2, \ldots, (\gamma/m)-1$ **do**
    $f_\varphi(x) \leftarrow w_f \times x$ # embedding function
    $M_\varphi(x) \leftarrow \prod_i^m f_\varphi(x)^{1/m}$ # disposition function
    $g_\varphi(x) \leftarrow w_g \times f_\varphi(x)$ # relation function
    Concatenate $g_\varphi(x)$ and $M_\varphi(x)$
    $S_{g,m} \leftarrow r_w(g_\varphi(x), M_\varphi(x))$ # similarity score
    $\mathcal{L} = CCE(S_{g,m}, x)$ # calculate CCE loss
    $\acute{w}_g, \acute{w}_d, \acute{r}_w = Adam(\nabla w_f, w_d, r_w[\mathcal{L}], lr)$
  end for
end for

The equation (7) refers to the loss function used in our study.

$$CCE = \sum_{i}^{c} t_i \, log \, \left(s_{g,m}(x)_i\right), t_i = \text{target vector} \tag{7}$$

IV. EXPERIMENTAL RESULTS

On the basis of DML and FSL, a new RN has been formed. We test the suggested network model using two different sets of our own data. The visual similarity of the objects in these datasets is their key characteristic. By collecting the similarity score between the objects, the built network has made it easier to classify these objects. In-depth comparisons are made

FIGURE 6. Sample images for the classes in the screw dataset used in this study.

FIGURE 7. Sample images for the classes in the spare part dataset used in this study.

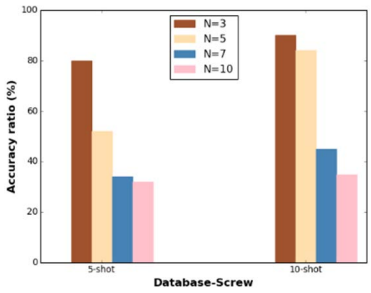


FIGURE 8. Accuracy GMRNet under different N-way and K-shot for the screw dataset.

between the novel relational network model we propose and the performance of the traditional RN model as described in the literature. When there are only a few training datasets available, FSL is helpful. FSL was used in our study when there wasn't a lot of data for each class at the start. In N-way K-shot classification, the N and K parameters were varied in a number of ways, and their impact on the model's effectiveness was discovered. Epochs 100, Adam as the optimizer algorithm, batch size 32, learning rate 0.001, and input image size 150 × 150 are the hyper-parameters of the recommended network.

TABLE 2. Comparisons of our study with DML-based studies.

Ref	Method	Applied problem	Acc(%)
[2]	VGG16	Image segmentation	45.80
[3]	GoogLenet	Person ID	63.66
[5]	CNN	Face and kinship verification	98.00
[6]	Triplet network	Maritime vessel identification	68.00
[7]	Triplet loss with ResNet50	Labeled description data	81.20
[8]	Deep Metric Net	Multi-label image classification	92.85
[9]	Relation network	Image classification	65.32
[10]	Relation network	Image classification	51.80
[11]	Relation network	Breast cancer classification	86.29
[12]	Relation network	Person identification	88.90
[13]	Relation network	Hyperspectral image classification	98.03
Our study	**Relation network**	**Similar objects classification**	**96.10 and 92.30**

TABLE 3. Performance results of GMRNet in N-way K-shot classification.

	Dataset-Screw Acc(%)		Dataset-Spare Acc(%)	
N	5-shot	10-shot	5-shot	10-shot
N=3	80	90	53	63
N=5	52	84	44	56
N=7	34	45	28	40
N=10	32	35	22	28

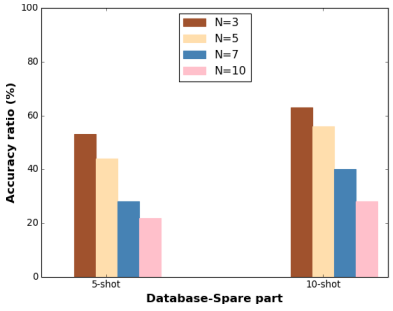
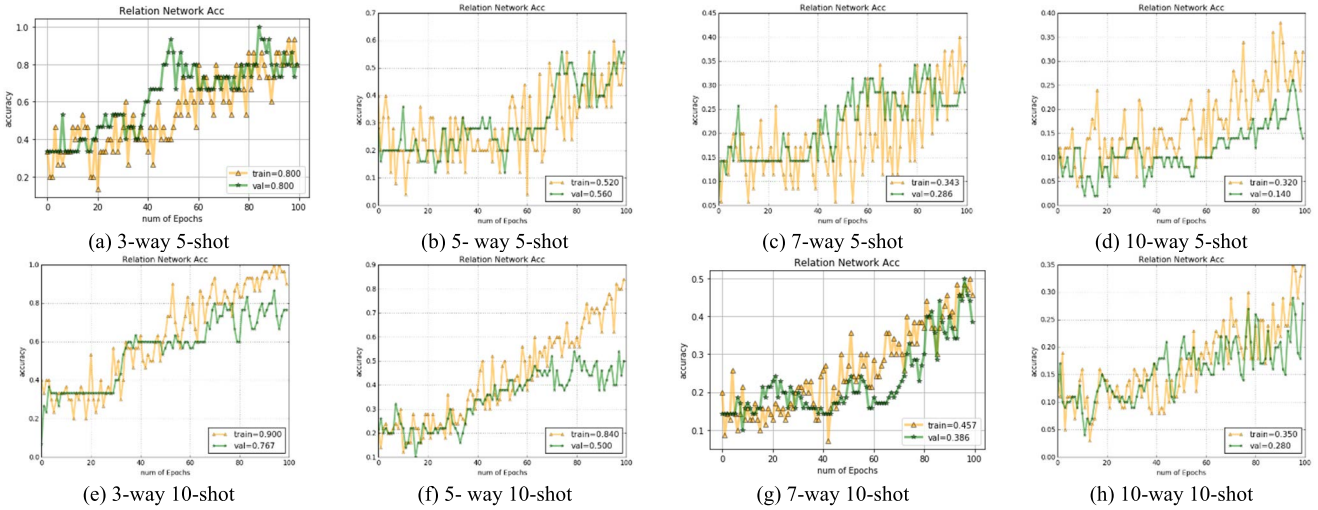


FIGURE 9. Accuracy of GMRNet under different N-way and K-shot for the spare part dataset.

A. CREATING DATASETS

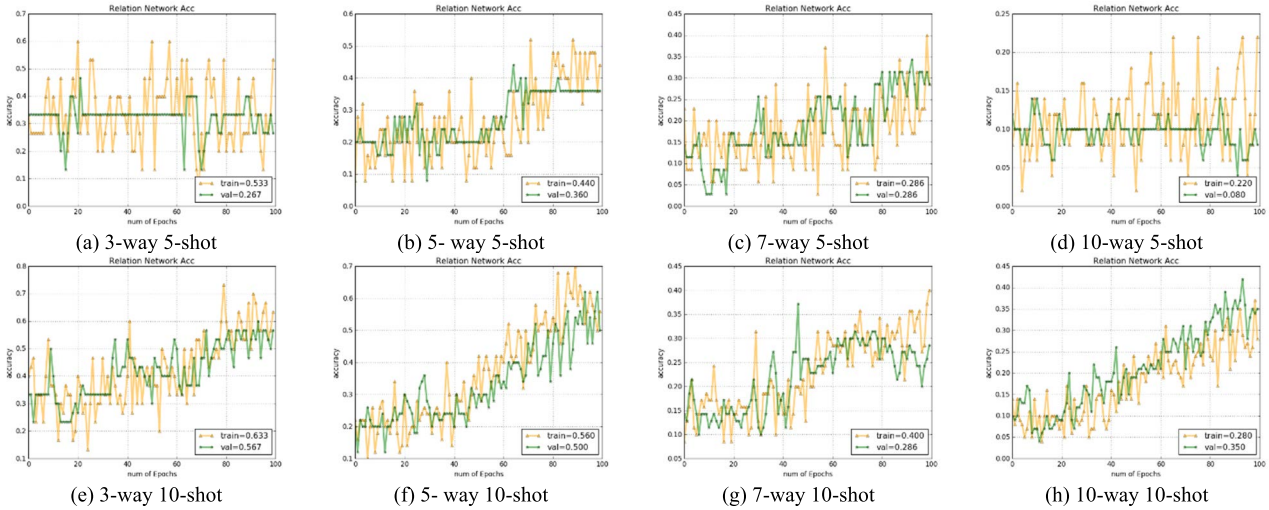
The dataset contains 150 × 150 pixel fastener images in RGB format. The fastener images, which contain 18 different classes, include six screw-type, five nut-type, and seven bolt-type classes. The images in the dataset were augmented with data augmentation techniques. For this purpose, each image was shifted by 0.2 on the x-axis and by 0.2 on the y-axis, as well as being rotated by a 30-degree angle, tilted by 0.2, and magnified by 0.2 percent. In addition, the images were



(a) 3-way 5-shot (b) 5- way 5-shot (c) 7-way 5-shot (d) 10-way 5-shot

(e) 3-way 10-shot (f) 5- way 10-shot (g) 7-way 10-shot (h) 10-way 10-shot

FIGURE 10. Accuracy ratios of GMRNet for N=3, 5, 7, and 10 way and K=5 and 10-shot on screw dataset.



(a) 3-way 5-shot (b) 5- way 5-shot (c) 7-way 5-shot (d) 10-way 5-shot

(e) 3-way 10-shot (f) 5- way 10-shot (g) 7-way 10-shot (h) 10-way 10-shot

FIGURE 11. Accuracy ratios of GMRNet for N=3, 5, 7, and 10 way and K=5 and 10-shot on spare parts dataset.

rotated on the horizontal axis. Each class includes approximately 100 images, with 1760 images for the training and 176 images for the testing. Each class in the spare part dataset comprises approximately 200 test images and 4100 train images. Let us refer to the new datasets that result from the augmentation of the number of images as 'screw ImAug dataset' and 'spare ImAug dataset' respectively.

When the existing studies in Table 2 are examined, the use of DML has been noticed through different methods for many different purposes. In this study, which is used to classify similar object groups, the DML-based network we developed is quite successful compared to many of the existing methods.

Considering Table 3, the accuracy rate decreased as the number of N-classes increased, while the accuracy increased as the number of K-samples increased. Since the objects in our dataset are highly similar and the number of samples is limited, the accuracy decreases as N increases. Figures 8 and 9 are diagrams designed to help you comprehend Table 3. Figure 10 and Figure 11 show the performance of GMRNet on the N-way K-shot problem with different N and K values. Factors such as the quality of the images in the dataset, the distribution of the images to the classes, the different variations of the images in each class, and the similarity of the images between the different classes greatly affect the success of the developed model. Our datasets have been expanded with image augmentation techniques in order to classify similar objects with high performance. After increasing the number of images in both datasets, it was tested on the RN model in the literature and the RN model we designed, and thus the effect of the dataset on the performance was observed.

Table 4 compares the existing network and GMRNet performance on three different datasets. The first two of these

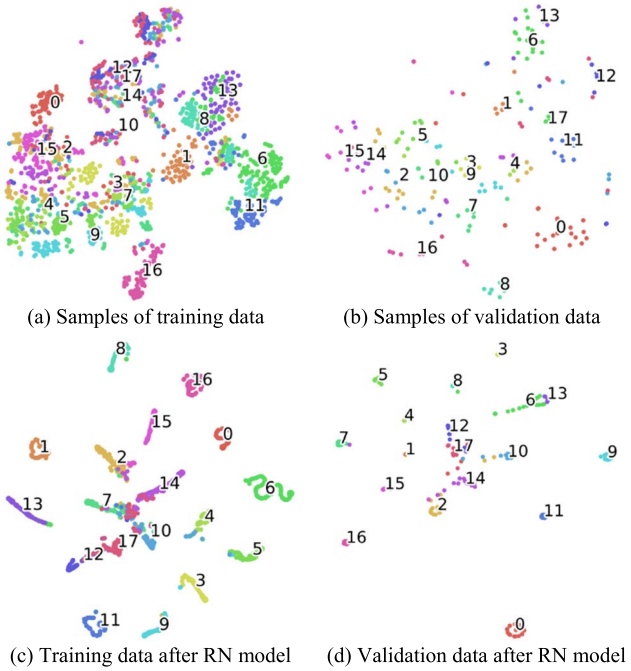


FIGURE 12. T-SNE images obtained by running the standard RN model in the literature on screw ImAug dataset.

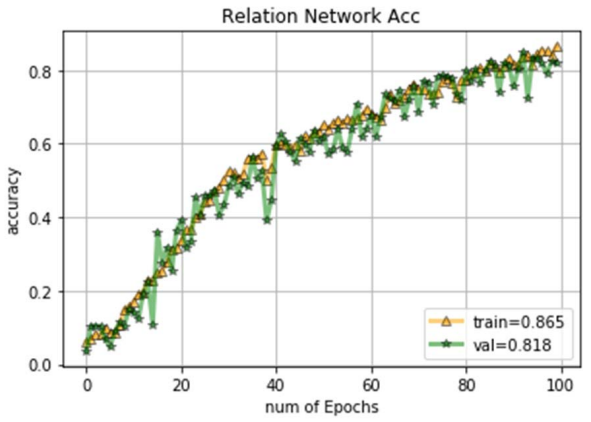


FIGURE 13. Accuracy rate of the standard RN model in the literature on screw ImAug dataset.

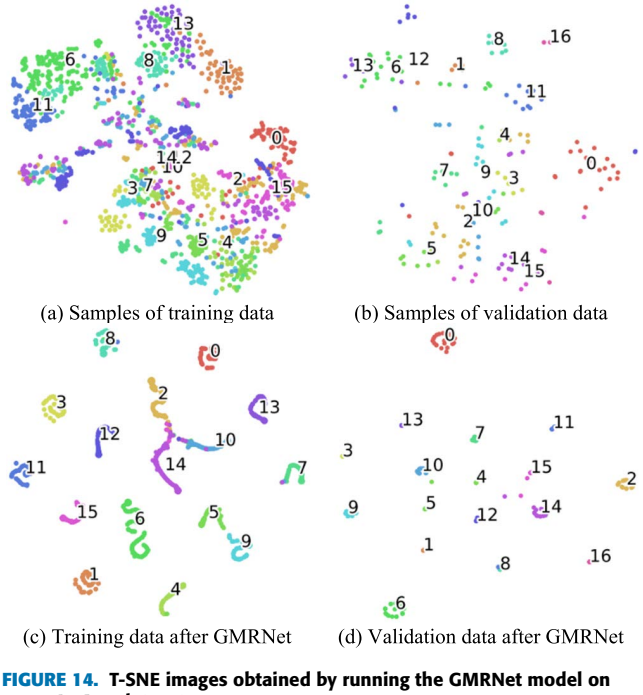


FIGURE 14. T-SNE images obtained by running the GMRNet model on screw ImAug dataset.

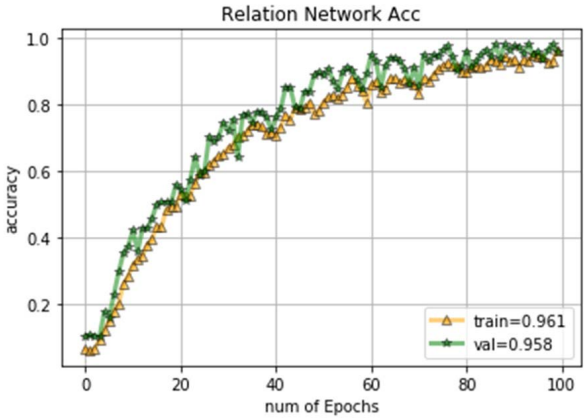


FIGURE 15. Accuracy rate of the GMRNet model on screw ImAug dataset.

TABLE 4. Comparisons of our study with RN model in the literature.

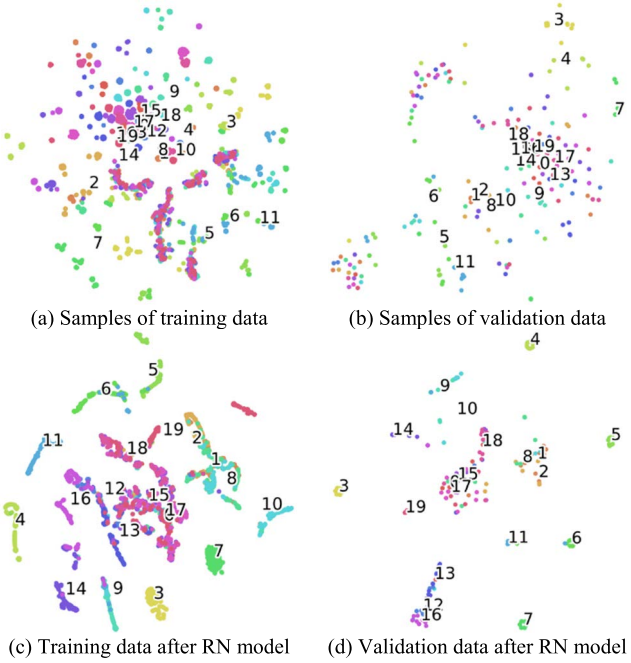
Dataset	RN model in the literature	GMRNet model
Screw ImAug	86.50	96.10
Spare ImAug	72.80	92.30
Omniglot	62.12	71.91

datasets are the datasets created in our study. On the other hand, Omniglot is a public dataset that is frequently used in studies on classification problems. Omniglot includes 1623 classes from 50 alphabets [29]. Each class has 20 samples drawn by various people. According to the results, the screw ImAug dataset has the best performance.

In Table 5, the GMRNet algorithm is compared with the existing transfer learning methods and meta-learning methods. Since the GMRNet method is based on meta-learning algorithms, meta-learning algorithms have been applied in this study for comparison purposes. Meta-learning algorithms consist of matching networks, prototypical networks, and relation networks. Using meta-learning algorithms, training has been done on a spare dataset and a screw dataset. In addition, transfer learning algorithms that are frequently used in the literature have been preferred and applied on two datasets. The obtained experimental results are added to Table 5. Thus, the proposed method in this study is compared with many existing methods. In Table 5, 'Acc', 'P', and 'R' represent accuracy, precision, and recall, respectively. When the table is examined, the fact that the accuracy rates and

TABLE 5. Performance comparisons of GMRNet and existing methods.

Method	Screw dataset				Spare dataset			
	Acc	P	R	Auc	Acc	P	R	Auc
VGG19	0.959	0.897	0.814	0.964	0.870	0.881	0.820	0.990
Inception	0.780	0.563	0.537	0.819	0.700	0.705	0.670	0.924
Xception	0.800	0.605	0.597	0.826	0.700	0.729	0.700	0.919
MobileNet	0.930	0.747	0.742	0.896	0.880	0.880	0.880	0.957
ResNet	0.930	0.498	0.456	0.741	0.840	0.840	0.840	0.931
Matching	0.660	0.750	0.590	0.870	0.630	0.740	0.600	0.910
Prototypical	0.600	0.490	0.530	0.850	0.580	0.640	0.640	0.880
Relation	0.865	0.871	0.701	0.989	0.728	0.895	0.657	0.970
GMRNet	0.961	0.842	0.817	0.983	0.923	0.969	0.947	0.998



(a) Samples of training data (b) Samples of validation data

(c) Training data after RN model (d) Validation data after RN model

FIGURE 16. T-SNE images obtained by running the standard RN model in the literature on spare ImAug dataset.

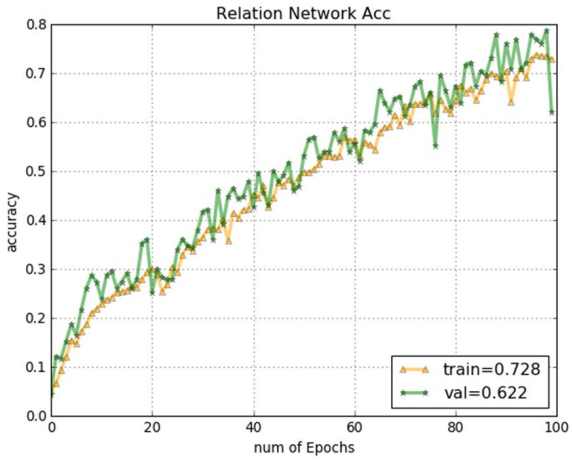
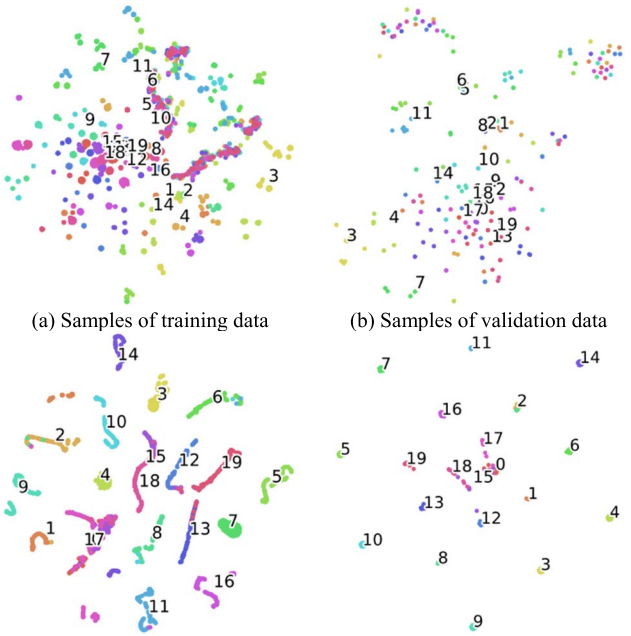


FIGURE 17. Accuracy rate of the standard RN model in the literature on spare ImAug dataset.

precision are high means that the results will be consistent no matter how many times the measurements are repeated. A high accuracy rate indicates that similar objects are correctly distinguished from each other. When the recall results in the table are examined, it is discovered that in some methods, this value is low. The ability to correctly guess the actual label is measured by the recall value. It states that GMRNet produces the best results among the meta-learning algorithms, allowing for accurate class distinctions. The high Auc values in the table show that the two classes are well distinguished from one another. In this case, it has been discovered that the spare dataset can be distinguished better than the screw dataset. When compared on the basis of a dataset, the accuracy rate of the methods on the spare dataset is slightly lower than screw dataset, due to the higher number of classes in the dataset. The higher the Auc value in the spare dataset, the better the classes are distinguished from each other and the less similarity between the classes in the spare dataset than the screw. The relation network generated the best results among the meta-learning methods, matching, prototypical, and relation networks, while the prototypical network obtained the worst results. The matching network compares image pairs, whereas the Relation network forms better results by establishing a relationship between image pairs. In cases where each image has different lighting and exposure, measuring the distance by comparing the images in the matching network and the prototypical network may result in incorrect labeling. Unlike the matching network, the prototype image produced for each class is the second of the images compared in the prototypical network. This image is frequently the mean of the images in the class. The prototypical network, on the other hand, yields less successful results. GMRNet, on the other hand, accomplishes better than the standard relation network and is detailed in Table 5.



(a) Samples of training data (b) Samples of validation data

(c) Training data after GMRNet (d) Validation data after GMRNet

FIGURE 18. T-SNE images obtained by running the GMRNet model on spare ImAug dataset.

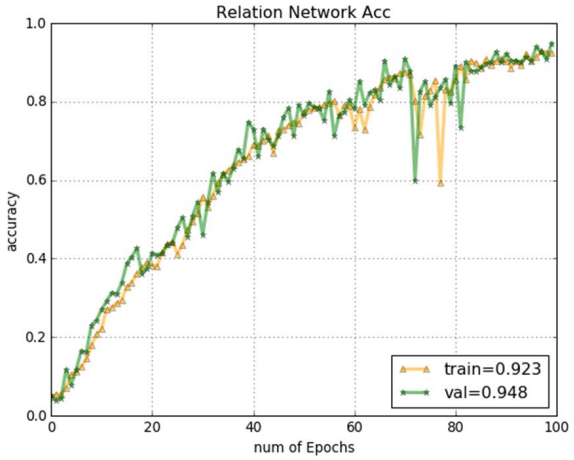


FIGURE 19. Accuracy rate of the GMRNet model on spare ImAug dataset.

V. CONCLUSION

Object classification and recognition are difficult due to the similarity of objects. Metric measurements are taken between objects using DML. If the compared objects are similar, the DML metric reduces the measurement, whereas if the objects are different, the measurement result increases. Two real data sets with similar object groups were used in this study. These datasets contain images of screws and spare parts. However, due to the limited number of images in our dataset, we resorted to FSL. With N-way K-shot learning, we discovered that the success rate decreases as the number of classes in our dataset grows. The model more easily classifies data with fewer classes. We also discovered that the large number of samples in our dataset improves performance by allowing the model to learn more effectively. We worked on N = 3, 5, 7, and 10 ways and K = 5 and 10 shots in this study. Thus, we obtained detailed experimental results. In this study, we developed a new geometric mean-based RN model. The addition of a new module, the Disposer mode, distinguishes this model from the standard RN model. The developed GMRNet was also tested on our datasets after data augmentation. As a result, in these datasets, GMRNet achieved results of 92.3% and 96.1%, respectively, compared to the standard RN model's 72.8% and 86.5%. This demonstrates how well the model we created performs in the categorization process. GMRNet was carried out for classification using t-SNE visualization. GMRNet isolates data points into distinct classes more successfully than regular RN when the generated images are inspected. The datasets we used for this study also include actual images of industrial processes. In this respect, our study outperforms other studies.

REFERENCES

[1] M. Kaya and H. S. Bilge, "Deep metric learning: A survey," *Symmetry*, vol. 11, no. 9, p. 1066, 2019.
[2] Y. Li, "Supervised few-shot image segmentation with deep metric learning," in *Proc. Int. Conf. Electron. Inf. Technol. Smart Agricult. (ICEITSA)*, Dec. 2021, pp. 431–434.
[3] X. Yang, P. Zhou, and M. Wang, "Person reidentification via structural deep metric learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 10, pp. 2987–2998, Oct. 2019.
[4] Y. Duan, J. Lu, W. Zheng, and J. Zhou, "Deep adversarial metric learning," *IEEE Trans. Image Process.*, vol. 29, pp. 2037–2051, 2020.
[5] J. Lu, J. Hu, and Y.-P. Tan, "Discriminative deep metric learning for face and kinship verification," *IEEE Trans. Image Process.*, vol. 26, no. 9, pp. 4269–4282, Sep. 2017.
[6] E. Gundogdu, B. Solmaz, A. Koç, V. Yücesoy, and A. A. Alatan, "Deep distance metric learning for maritime vessel identification," in *Proc. 25th Signal Process. Commun. Appl. Conf. (SIU)*, 2017, pp. 1–4.
[7] X. Zhao, H. Qi, R. Luo, and L. Davis, "A weakly supervised adaptive triplet loss for deep metric learning," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 3177–3180.
[8] C. Li, C. Liu, L. Duan, P. Gao, and K. Zheng, "Reconstruction regularized deep metric learning for multi-label image classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 7, pp. 2294–2303, Jul. 2019.
[9] F. Sung, Y. Yang, L. Zhang, T. Xiang, P. H. S. Torr, and T. M. Hospedales, "Learning to compare: Relation network for few-shot learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1199–1208.
[10] B. Hui, P. Zhu, Q. Hu, and Q. Wang, "Self-attention relation network for few-shot learning," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, Jul. 2019, pp. 198–203.
[11] S. Rhee, S. Seo, and S. Kim, "Hybrid approach of relation network and localized graph convolutional filtering for breast cancer subtype classification," 2017, *arXiv:1711.05859*.
[12] H. Park and B. Ham, "Relation network for person re-identification," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 7, pp. 11839–11847.
[13] K. Gao, B. Liu, X. Yu, J. Qin, P. Zhang, and X. Tan, "Deep relation network for hyperspectral image few-shot classification," *Remote Sens.*, vol. 12, no. 6, p. 923, 2020.
[14] F. Cakir, K. He, X. Xia, B. Kulis, and S. Sclaroff, "Deep metric learning to rank," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1861–1870.
[15] M. Huai, H. Xue, C. Miao, L. Yao, L. Su, C. Chen, and A. Zhang, "Deep metric learning: The generalization analysis and an adaptive algorithm," in *Proc. IJCAI*, Aug. 2019, pp. 2535–2541.
[16] W. Zheng, B. Zhang, J. Lu, and J. Zhou, "Deep relational metric learning," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 12065–12074.

[17] J. Wang, F. Zhou, S. Wen, X. Liu, and Y. Lin, "Deep metric learning with angular loss," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2593–2601.
[18] W. Ge, "Deep metric learning with hierarchical triplet loss," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 269–285.
[19] A. Khan, E. Fleming, N. Schofield, M. Bishop, and N. Andrews, "A deep metric learning approach to account linking," 2021, *arXiv:2105.07263*.
[20] D. Kang, H. Kwon, J. Min, and M. Cho, "Relational embedding for few-shot classification," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 8822–8833.
[21] S. Baik, J. Choi, H. Kim, D. Cho, J. Min, and K. M. Lee, "Meta-learning with task-adaptive loss function for few-shot learning," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 9465–9474.
[22] X. Sun, B. Wang, Z. Wang, H. Li, H. Li, and K. Fu, "Research progress on few-shot learning for remote sensing image interpretation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 2387–2402, 2021.
[23] Y. Wang, Q. Yao, J. T. Kwok, and L. M. Ni, "Generalizing from a few examples: A survey on few-shot learning," *ACM Comput. Surv.*, vol. 53, no. 3, pp. 1–34, May 2021.
[24] Z. Yue, H. Zhang, Q. Sun, and X.-S. Hua, "Interventional few-shot learning," 2020, *arXiv:2009.13000*.
[25] A. Ravichandran, R. Bhotika, and S. Soatto, "Few-shot learning with embedded class models and shot-free meta training," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 331–339.
[26] J. He, R. Hong, X. Liu, M. Xu, Z.-J. Zha, and M. Wang, "Memory-augmented relation network for few-shot learning," in *Proc. 28th ACM Int. Conf. Multimedia*, Oct. 2020, pp. 1236–1244.
[27] X. Zhou, J. Li, Z. Wang, R. He, and T. Tan, "Image inpainting with contrastive relation network," in *Proc. 25th Int. Conf. Pattern Recognit. (ICPR)*, Jan. 2021, pp. 4420–4427.
[28] J. Liu, C. Shi, D. Tu, Z. Shi, and Y. Liu, "Zero-shot image classification based on a learnable deep metric," *Sensors*, vol. 21, no. 9, p. 3241, 2021.
[29] B. Lake, R. Salakhutdinov, J. Gross, and J. Tenenbaum, "One shot learning of simple visual concepts," in *Proc. CogSci*, 2011, pp. 1–7.

CANAN TASTIMUR was born in Elâziğ, Turkey, in 1991. She received the B.S. and M.S. degrees in computer engineering from Firat University, Elâziğ, in 2013 and 2017, respectively, where she is currently pursuing the Ph.D. degree. She is currently a Research Assistant with the Computer Engineering Department, Erzincan Binali Yildirim University. Her research interests include fault diagnosis, image processing, and railway inspection.

ERHAN AKIN (Member, IEEE) was born in Erzincan, Turkey, in 1963. He received the B.S. and M.S. degrees in electrical engineering and the Ph.D. degree in ac drives from Firat University, Elâziğ, Turkey, in 1984, 1987, and 1994, respectively. He was an Assistant Professor with the Department of Electrical Engineering, Firat University, where he was an Associate Professor of electrical machines. He is currently a Full Professor of computer engineering with Firat University. His main research interests include power electronics, digital control of variable-speed ac drives, fuzzy control and soft-computing techniques, railway systems, and image processing.