JOIV : Int. J. Inform. Visualization, 7(3) - September 2023 638-643

INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION

journal homepage : www.joiv.org/index.php/joiv

Closer Look at Image Classification for Indonesian Sign Language with Few-Shot Learning Using Matching Network Approach

Irma Permata Sari [a,*]

[a] *Information System and Technology Department, Universitas Negeri Jakarta, Rawamangun, East Jakarta, Indonesia*
Corresponding author: [*]irmapermatasari@unj.ac.id

***Abstract*— Huge datasets are important to build powerful pipelines and ground well to new images. In Computer Vision, the most basic problem is image classification. The classification of images may be a tedious job, especially when there are a lot of amounts. But CNN is known to be data-hungry while gathering. How can we build some models without much data? For example, in the case of Sign Language Recognition (SLR). One type of Sign Language Recognition system is vision-based. In Indonesian Sign Language dataset has a relatively small sample image. This research aims to classify sign language images using Computer Vision for Sign Language Recognition systems. We used a small dataset, Indonesian Sign Language. Our dataset is listed in 26 classes of alphabet, A-Z. It has loaded 12 images for each class. The methodology in this research is few-shot learning. Based on our experiment, the best accuracy for few-shot learning is Mnasnet1_0 (85.75%) convolutional network model for Matching Networks, and loss estimation is about 0,43. And the experiment indicates that the accuracy will be increased by increasing the number of shots. We can inform you that this model's matching network framework is unsuitable for the Inception V3 model because the kernel size cannot be greater than the actual input size. We can choose the best algorithm based on this research for the Indonesian Sign Language application we will develop further.**

***Keywords*— Few-shot learning; matching network; sign language recognition; computer vision.**

Manuscript received 20 Oct. 2022; revised 17 Jan. 2023; accepted 6 Jun. 2023. Date of publication 10 Sep. 2023.
International Journal on Informatics Visualization is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.

I. INTRODUCTION

Sign Language Recognition focuses on algorithms and methods to recognize the pattern of signs produced and set their interpretation in text or speech form [1], [2]. One of the main categories of Sign Language Recognition is vision-based [3]. Vision-based systems utilize images and videos of sign language. Sensor-based systems involve wearable devices set with sensors to extricate hand shapes and sign movements. Manual classification of images may be a tedious job [4], [5], especially when there are many amounts of it (for example, 10,000 images), and it would be very useful if we could automate the whole process using Computer Vision. In Computer Vision, the most basic problem is image classification [1]. An algorithm to classify the images into Computer Vision is few-shot learning.

Few-shot learning is an outstanding experimentation topic that has been spaciously inspected in the past few years [6] [7]–[14]. Few-shot learning for image classification (few-shot image classification) is a part of machine learning but still needs a lot of research and improvement [15], [16]. Few-shot learning aims to train a classifier and identify the invisible classes during training with some labeled examples [17], [18]. Few-shot learning is also classifying the new data even with some supervised learning training samples. The easiest approach to few-shot learning is the meta-learning approach. The meta-learning approach for image classification problems is divided into several methods: matching networks, prototypical networks, relation networks, and model-agnostic meta-learning (MAML). In this paper, we tried to classify the Indonesian Sign Language image with a few-shot learning image classification with a matching network approach.

II. MATERIALS AND METHODS

A. Few-Shot Learning

Few-shot learning (FSL) can work quite well even with a relatively small number of training samples [17], [19]–[26]. The common types of FSL are N-Shot Learning (NSL), Few-Shot Learning (FSL), One-Shot Learning (OSL), and Less than one or Zero-Shot Learning (ZSL)

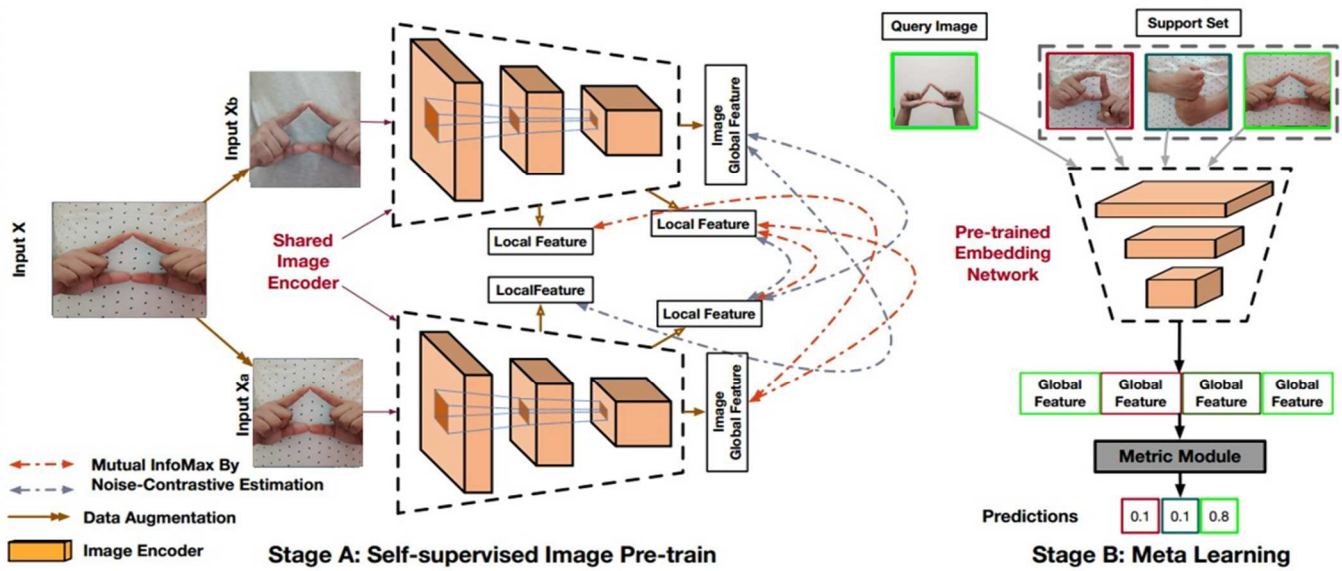

Fig. 1 The Embedding Network of Few-Shot Learning [27]

All the few-shot learning perspectives use advanced science of several forms (e.g., data, model, and algorithm) to bring down 'sample difficulty' described as the number of training samples required to ensure the loss of reduced empirical risk [28], [29]. Several studies have been carried out in research related to our work. Chen et al. [30] suggest a Self-Jig algorithm to enhance the input data in few-shot learning by combining new images that are either labeled or unlabeled. Zhao et al. [31] cut the features into three orthogonal parts to improve the few-short learning classification performance, assent to coincident feature selection, and dense estimation.

The approved scheme for few-shot learning is by way of meta-learning (learning-to-learn) with multi-additional tasks [15], [32]. Spyros Gidaris et al. [33] suggest the appealing composition of supervised and self-supervised loss to pre-train the embedding network. The constant with the assignment of the pre-trained embedding network in few-shot learning has been studied by Chen [27].

B. Matching Network

Our proposed research used few-shot learning with matching networks approach. A matching network is one of the meta-learning methods for image classification problems[34]. Matching networks embrace two neural embedding functions and an augmented memory [8], [35]–[37]. Some fixed jargon of the matching network [38]:

- Label set: Relate to a sample set of all potential categories.
- Support set: Related to some images of label set categories.
- Batch: Related to a support set.
- N-way, K-shot Method

N-way is a mount class in one task. K-shot is a mount support image for each class on a task. Another term is N-query is the amount of query images for each class in one task. N-task is a number of tasks to sample or iterate (epoch). Understanding data processing for Matching Networks, we can describe by Chen [27] in Fig. 1, and modified with the Indonesian Sign Language dataset.

A dataset is used to solve few-shot learning tasks. This dataset is divided into episodes. And then there are three following steps below [37]:

- The supported image and query image set is taken to a CNN that proceeds with embeddings.
- Each query image is characterized by applying the SoftMax of the cosine distance from its embeddings to support-set embeddings.
- The Cross-Entropy Loss on the resulting classification is backpropagated through CNN.

This approach allows a matching network to classify images by comparing different instances of the classes. Matching Networks differ from Prototypical Networks in the few-shot case with equivalence[39].

C. Research State Diagram and Novelty

We describe our research stage in Fig. 2. There are five stages: preparing the dataset, pre-processing, training model, testing, and evaluation. This research's novelty is classifying the Indonesian Sign Language image with few-shot learning using a matching network approach and testing it with some convolutional network models.

D. Dataset

The dataset used in this study is an image of the Indonesian Sign Language image. This dataset has a total of 312-image datasets and was labeled as 26 classes of the alphabet. The images from each class are given in Table I.

TABLE I
INDONESIAN SIGN LANGUAGE DATASET

Category	Number
Training set	218
Validation set	94
Total dataset	312

In Table I, the dataset was cut into a training set (218 images), and a validation set (94 images).

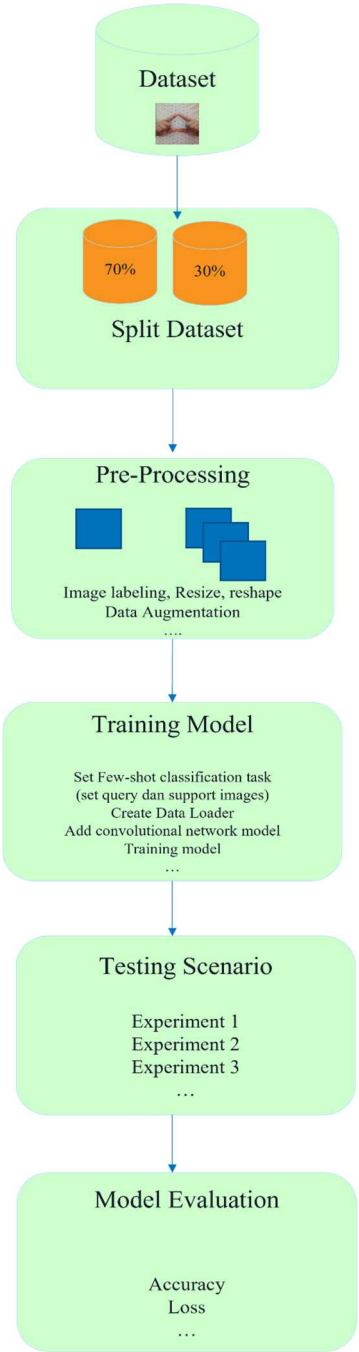

Fig. 2 Research stage diagram

Pre-Processing In training and validation, the dataset was resized to 84*84 pixels to fit in our model. We cut the whole dataset into a training set (18 classes), and a testing set (8 classes) or 70:30. Training the weight of the deep convolutional neural network is done by using the training set. On the other hand, verifying the model's generalization ability and the loss function value is done using a validation set [40].

III. Result and Discussion

A. Data Exploration

Our dataset can be visualized in Figure 3.

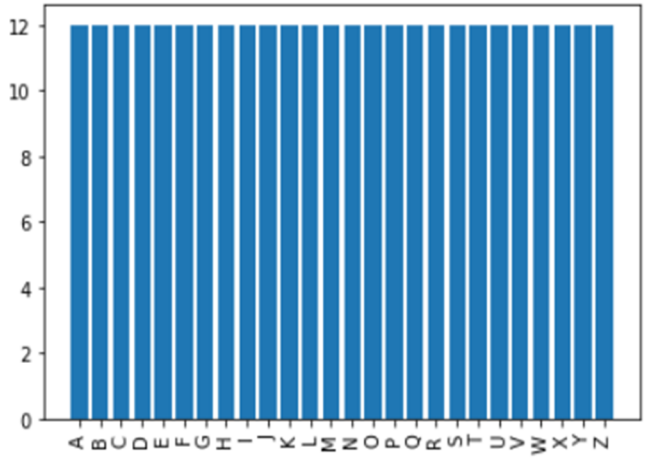

Fig. 3 Image Dataset of Indonesian Sign Language

Fig. 3 indicates that our dataset is listed in 26 classes of alphabets, A-Z. It has loaded 12 images for each class. Some sample images can be randomly shown from each class in Fig. 4.

Fig. 4 Image dataset display

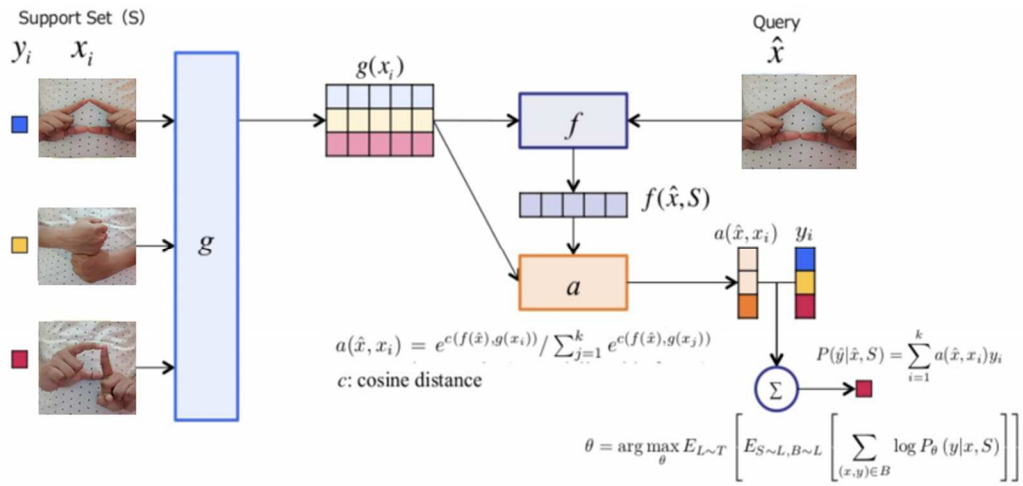

Fig. 5 Matching Networks Architecture (Figure adapted from [38])

We have five image displays for each class such as B, V, Z, A, and Q, in Fig. 4. The image dataset display can be changed if we run the program. We set it in some tasks in the few-shot learning algorithm, for example, n-way class, k-shot, query, and task image for each class. We can change its values for some experiments. To understand this term, we can see the architecture of the matching network can show in Fig. 5. In Fig. 5, we describe the support images and query images in a formulation. And then the display image can be seen in Fig. 6 and Fig. 7 below.

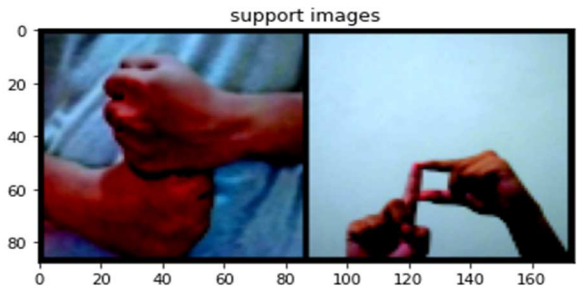

Fig. 6 Support Images

Fig. 6 shows two 'support images' because we set 2-way support image in each class. There are 26 classes in this research. And in Fig. 7 there are 10 images for each row (n-query). So, we obtained 20 query images for two classes. This image can be set according to our research tasks.

B. Experiment 1: Comparison with some convolutional network model

In this experiment, we check the accuracy and compare it with another few-shot learning it with different convolutional network models. We analyzed using a small task, two classes, and two images for each shot (2-way for each class and a 2-shot image for each class for training and validation).

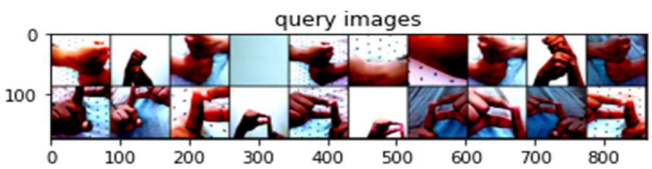

Fig. 7 Query Images

The outcome of the test is displayed in Table II as follows.

TABLE II
MATCHING NETWORKS WITH DIFFERENT MODEL

Exp. No	Convolutional Network Model	Accuracy	Loss
1	Inception_v3	Not suitable	-
2	Shufflenet_v2x1_0	69.33	0.16
3	Densenet161	49.58	0.69
4	Densenet201	69.17	0.59
5	Wide resnet50_2	46.50	0.69
6	Mobilenet_v2	46.83	0.83
7	Mnasnet1_0	85.75	3.05

Table II shows seven models in our experiment numbers 1-7. The best accuracy is experiment 7 (Mnasnet 1_0), 85.75%. Mnasnet is the best neural network model that achieves high accuracy and less inference time, using reinforcement learning search algorithm and factorized search space [41]. And then, we can inform that the matching network framework for this model is not suitable for the Inception V3 model because the kernel size can't be greater than the actual input size. The visualization of this experiment can be displayed in Fig. 8.

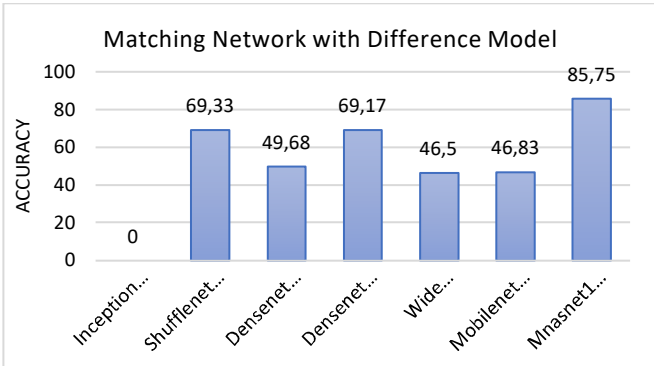

Fig. 8 Visualization of Matching Network with Different Model

C. Experiment 2: Comparison with different dataset

We compared our results with other datasets like Face Expressions, Fashion, and CUB datasets. We set 70% data for

training and 30% data for validation split. And then we set FSL tasks like:

1) Training set: In training process set 2-way for each class; n-shot = 1, 5, and 10 images each class; 2-query, and 500=tasks iterations for each batch.

2) *The validation set:* We set 2-way for each class, with 1, 5, and 10 images for n-shot, 2-query, and 300 iterations or epochs for the tasks.

The result of this experiment can be seen in Table III.

TABLE III
COMPARISON MATCHING NETWORK WITH DIFFERENT DATASET

No	Convolutional Network Model	Class	Accuracy		
			1-shot	5-shot	10-shot
1	Indonesian Sign Language	26	67.17	85.67	91.33
2	Face Expression	7	54.50	51.58	66.68
3	CUB	200	50.50	54.17	55.83
4	Fashion	13	75.58	79.17	82.83

Table III tells that the best accuracy is Indonesian Sign Language data, 2-way, and 10-shot. This experiment indicates that the accuracy will be increased by increasing the number of shots.

D. Experiment 3: Loss Comparison for Some Dataset

In the first experiment, we checked the accuracy and compared it with another few-shot learning with different convolutional network models. They have been available in Pytorch models. We analyzed that using a small task (2-way and 10-shots image for each class for training and validation). The results of the test are displayed in Table IV as follows.

TABLE IV
COMPARISON MATCHING NETWORKS FOR SOME DATASET

No	Dataset	Class	Loss
1	Indonesian Sign Language	26	0.43
2	Face Expression	7	0.70
3	CUB	200	0.64
4	Fashion	13	0.46

Based on Table IV, the smallest loss is Indonesian Sign Language data.

IV. CONCLUSION

We adopted a few-shot learning with matching network approach for Indonesian Sign Language Recognition. And we test some convolutional network models as the architecture for matching networks method. Based on our studied, we found the best accuracy is Mnasnet_1_0, 85.75%. This experiment also indicates that the accuracy will be increased by increasing the number of shots. It can be tested in future studies by increasing the number of tasks in few-shot learning.

ACKNOWLEDGMENT

We thank Universitas Negeri Jakarta-Engineering Faculty under the Basic Research Grant 2022 for funding this research.

REFERENCES

[1] C. M. Jin, Z. Omar, and M. H. Jaward, "A mobile application of American sign language translation via image processing algorithms," *Proc. - 2016 IEEE Reg. 10 Symp. TENSYMP 2016*, pp. 104–109, 2016, doi: 10.1109/TENCONSpring.2016.7519386.

[2] R. Alzohairi, R. Alghonaim, W. Alshehri, S. Aloqeely, M. Alzaidan, and O. Bchir, "Image based Arabic Sign Language recognition system," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 3, pp. 185–194, 2018, doi: 10.14569/IJACSA.2018.090327.

[3] D. Li, C. R. Opazo, X. Yu, and H. Li, "Word-level deep sign language recognition from video: A new large-scale dataset and methods comparison," *Proc. - 2020 IEEE Winter Conf. Appl. Comput. Vision, WACV 2020*, pp. 1448–1458, 2020, doi: 10.1109/WACV45572.2020.9093512.

[4] Woo, "乳鼠心肌提取 HHS Public Access," *Physiol. Behav.*, vol. 176, no. 1, pp. 139–148, 2017.

[5] F. Pahde, M. Puscas, T. Klein, and M. Nabi, "Multimodal prototypical networks for few-shot learning," *Proc. - 2021 IEEE Winter Conf. Appl. Comput. Vision, WACV 2021*, pp. 2643–2652, 2021, doi: 10.1109/WACV48630.2021.00269.

[6] Y. Fu, Y. Xie, Y. Fu, and Y.-G. Jiang, "Meta Style Adversarial Training for Cross-Domain Few-Shot Learning," 2023.

[7] M. Kumar, V. Kumar, H. Glaude, C. De Lichy, A. Alok, and R. Gupta, "Protoda: Efficient Transfer Learning for Few-Shot Intent Classification," *2021 IEEE Spok. Lang. Technol. Work. SLT 2021 - Proc.*, pp. 966–972, 2021, doi: 10.1109/SLT48900.2021.9383495.

[8] V. Garcia and J. Bruna, "Few-shot learning with graph neural networks," *6th Int. Conf. Learn. Represent. ICLR 2018 - Conf. Track Proc.*, pp. 1–13, 2018.

[9] S. Ravi and H. Larochelle, "Optimization as a model for few-shot learning," *5th Int. Conf. Learn. Represent. ICLR 2017 - Conf. Track Proc.*, pp. 1–11, 2017.

[10] J. Snell, K. Swersky, and R. Zemel, "Prototypical networks for few-shot learning," *Adv. Neural Inf. Process. Syst.*, vol. 2017-Decem, pp. 4078–4088, 2017.

[11] Y. Liu, B. Schiele, and Q. Sun, "An Ensemble of Epoch-Wise Empirical Bayes for Few-Shot Learning," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 12361 LNCS, pp. 404–421, 2020, doi: 10.1007/978-3-030-58517-4_24.

[12] H. Tang, Z. Li, Z. Peng, and J. Tang, "BlockMix: Meta Regularization and Self-Calibrated Inference for Metric-Based Meta-Learning," *MM 2020 - Proc. 28th ACM Int. Conf. Multimed.*, pp. 610–618, 2020, doi: 10.1145/3394171.3413884.

[13] C. Xie, M. Tan, B. Gong, J. Wang, A. L. Yuille, and Q. V. Le, "Adversarial examples improve image recognition," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 816–825, 2020, doi: 10.1109/CVPR42600.2020.00090.

[14] Y. L. and H. T. S. J. Zhang, J. Song, L. Gao, "Progressive Meta-Learning With Curriculum," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 9, pp. 5916–5930, 2022, doi: 10.1109/TCSVT.2022.3164190.

[15] S. Qiao, C. Liu, W. Shen, and A. Yuille, "Few-Shot Image Recognition by Predicting Parameters from Activations," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 7229–7238, 2018, doi: 10.1109/CVPR.2018.00755.

[16] M. M. Rahman, M. S. Islam, M. H. Rahman, R. Sassi, M. W. Rivolta, and M. Aktaruzzaman, "A new benchmark on american sign language recognition using convolutional neural network," *2019 Int. Conf. Sustain. Technol. Ind. 4.0, STI 2019*, vol. 0, pp. 24–25, 2019, doi: 10.1109/STI47673.2019.9067974.

[17] W. Y. Chen, Y. C. F. Wang, Y. C. Liu, Z. Kira, and J. Bin Huang, "A closer look at few-shot classification," *7th Int. Conf. Learn. Represent. ICLR 2019*, no. 2018, pp. 1–17, 2019.

[18] Y. Ding *et al.*, *Multi-scale Relation Network for Few-Shot Learning Based on Meta-learning*, vol. 1. Springer International Publishing.

[19] W. Zheng, X. Tian, B. Yang, S. Liu, Y. Ding, and J. Tian, "applied sciences A Few Shot Classification Methods Based on Multiscale Relational Networks," 2022.

[20] D. Matching, "Double-View Matching Matching Network Network for for Few-Shot Learning to Classify Covid-19 in X-ray images Learning to Classify Covid-19 in X-ray images," vol. 1, no. 1, 2021, doi: 10.36244/ICJ.2021.1.4.

[21] R. Hou, H. Chang, B. Ma, S. Shan, and X. Chen, "Cross Attention Network for Few-shot Classification," vol. 1, no. c, pp. 1–12, 2019.

[22] H. Yang, X. He, and F. Porikli, "One-Shot Action Localization by Learning Sequence Matching Network," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 1450–1459, 2018, doi: 10.1109/CVPR.2018.00157.

[23] L. Gui, Y. Wang, D. Ramanan, and M. F. Moura, "Few-Shot Human Motion Prediction via," *Eur. Conf. Comput. Vis.*, pp. 1–19, 2020.

[24] P. Sirinam, M. S. Rahman, N. Mathews, and M. Wright, "Triplet fingerprinting: More practical and portable website fingerprinting with N-shot learning," *Proc. ACM Conf. Comput. Commun. Secur.*, pp. 1131–1148, 2019, doi: 10.1145/3319535.3354217.

[25] C. Liu *et al.*, "Learning a Few-shot Embedding Model with Contrastive Learning," *35th AAAI Conf. Artif. Intell. AAAI 2021*, vol. 10A, pp. 8635–8643, 2021, doi: 10.1609/aaai.v35i10.17047.

[26] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," *34th Int. Conf. Mach. Learn. ICML 2017*, vol. 3, pp. 1856–1868, 2017.

[27] D. Chen, Y. Chen, Y. Li, F. Mao, Y. Hey, and H. Xue, "Self-Supervised Learning for Few-Shot Image Classification," *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, vol. 2021-June, pp. 1745–1749, 2021, doi: 10.1109/ICASSP39728.2021.9413783.

[28] T. Banerjee, N. R. Thurlapati, V. Pavithra, S. Mahalakshmi, D. Eledath, and V. Ramasubramanian, "Few-shot learning for frame-wise phoneme recognition: Adaptation of matching networks," *Eur. Signal Process. Conf.*, vol. 2021-Augus, pp. 516–520, 2021, doi: 10.23919/EUSIPCO54536.2021.9616234.

[29] H. Li, W. Dong, X. Mei, C. Ma, and F. Huang, "LGM-Net : Learning to Generate Matching Networks for Few-Shot Learning," 2019.

[30] Z. Chen, Y. Fu, K. Chen, and Y. G. Jiang, "Image block augmentation for one-shot learning," *33rd AAAI Conf. Artif. Intell. AAAI 2019, 31st Innov. Appl. Artif. Intell. Conf. IAAI 2019 9th AAAI Symp. Educ. Adv. Artif. Intell. EAAI 2019*, vol. 1, pp. 3379–3386, 2019, doi: 10.1609/aaai.v33i01.33013379.

[31] B. Zhao, X. Sun, Y. Fu, Y. Yao, and Y. Wang, "MSplit LBI: Realizing feature selection and dense estimation simultaneously in few-shot and zero-shot learning," *35th Int. Conf. Mach. Learn. ICML 2018*, vol. 13, pp. 9421–9432, 2018.

[32] F. Sung, Y. Yang, and L. Zhang, "Relation Network for Few-Shot Learning," *Cvpr*, pp. 1199–1208, 2018.

[33] S. Gidaris, A. Bursuc, N. Komodakis, P. P. Perez, and M. Cord, "Boosting few-shot visual learning with self-supervision," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2019-Octob, pp. 8058–8067, 2019, doi: 10.1109/ICCV.2019.00815.

[34] A. Parnami and M. Lee, "Learning from Few Examples: A Summary of Approaches to Few-Shot Learning," no. January 2020, pp. 1–32, 2022.

[35] Y. Cheng, M. Yu, X. Guo, and B. Zhou, "Few-shot Learning with Meta Metric Learners," no. Nips, pp. 1–9, 2019.

[36] D. Lee, J. Yoon, J. Song, S. Lee, and S. Yoon, "One-Shot Learning for Text-to-SQL Generation," 2019.

[37] O. Vinyals, C. Blundell, T. Lillicrap, K. Kavukcuoglu, and D. Wierstra, "Matching networks for one shot learning," *Adv. Neural Inf. Process. Syst.*, pp. 3637–3645, 2016.

[38] S. Jadon, "An Overview of Deep Learning Architectures in Few-Shot Learning Domain," 2020, doi: 10.13140/RG.2.2.31573.24803/1.

[39] J. Snell, K. Swersky, and R. Zemel, "Prototypical Networks for Few-shot Learning," no. Nips, 2017.

[40] I. P. Sari, Widodo, M. Nugraheni, and P. Wanda, "A Basic Concept of Image Classification for Covid-19 Patients Using Chest CT Scan and Convolutional Neural Network," *Proceeding - 1st Int. Conf. Inf. Technol. Adv. Mech. Electr. Eng. ICITAMEE 2020*, pp. 175–178, 2020, doi: 10.1109/ICITAMEE50454.2020.9398462.

[41] C. Michael, "Neural Architecture Search For Smart Phones," 2019.