



A Survey on Few-Shot Techniques in the Context of Computer Vision Applications Based on Deep Learning

Miguel G. San-Emeterio^(✉) 

Atos Research and Innovation, 28037 Madrid, Spain
mglez.sanemeterio@gmail.com

Abstract. This review article about Few-Shot Learning techniques is focused on Computer Vision Applications based on Deep Convolutional Neural Networks. A general discussion about Few-Shot Learning is given, featuring a context-constrained description, a short list of applications, a description of a couple of commonly used techniques and a discussion of the most used benchmarks for FSL computer vision applications. In addition, the paper features a few examples of recent publications in which FSL techniques are used for training models in the context of Human Behaviour Analysis and Smart City Environment Safety. These examples give some insight about the performance of state-of-the-art FSL algorithms, what metrics do they achieve, and how many samples are needed for accomplishing that.

Keywords: Few-Shot Learning · Deep Learning · Computer Vision · Human Behaviour Analysis · Smart City Environment Safety

1 Introduction

In September 2012 the Convolutional Neural Networks (CNNs) were in the spotlight of the Computer Vision community when a model developed by Alex Krizhevsky, Ilya Sutskever and Geoffrey Hinton [18] had an outstanding performance at the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [33]. This CNN model started the so-called “Deep Learning revolution”, which is still booming as of today. During this period, novel computer vision applications have arisen, allowing practitioners and researchers to tackle previously unsolvable problems and to create new out-of-the-box solutions. Although powerful, Deep Learning (DL) techniques have their own set of problems, being data and computer power their most limiting requirements. Since they are Machine Learning (ML) algorithms, they need large amounts of data for training, which in turn require high computational power to process.

Many researchers have focused their work on mitigating these problems, creating a variety of techniques that softens the workload needed to create labelled datasets for supervised training. Some of them consist of using low-complexity labels for training models for high-complexity tasks [49]; others try to augment

datasets by generating synthetic data [27,28], and others try to reuse previously learnt features [50]. Please note that there are other approaches beside the aforementioned three. Among the last of the three highlighted strategies there is a family of techniques called “Few-Shot Learning” (FSL), which includes “One-Shot Learning” and “Zero-Shot Learning”.

For image classification tasks, supervised training of ML models require large amounts of annotated data for each of the classes or categories. Few-Shot Learning comprises those techniques designed for enabling a trained model to learn a new category from few examples. Thus, “One-Shot” would ideally need a single example of the new category to be able to learn how to classify it, whereas “Zero-Shot” would only need its description. Although those techniques are feasible for some tasks and data-types (see [38] for One-Shot and [32] for Zero-Shot), most applications require more samples to perform accurately. For example, a previous work from the author shows how a model trained for classifying different plant species needed 80 samples to classify at 90% accuracy a new-learned class [3]. The true potential of Few-Shot Learning is its power to provide neural networks with adaptability and agility, which are features where CNNs do not traditionally shine on. This “learning speed” comes at a cost on accuracy and overall performance, although in many cases it is an acceptable trade-off. Also, while under regular conditions data scarcity hinders ML algorithms’ performance, potentially down to unacceptable levels, the application of FSL can enable the use of ML algorithms when data is scarce.

This review article discusses the latest developments in Few-Shot Learning techniques applied to Computer Vision models, focusing in the contexts of Human Behaviour Analysis and Smart City Environment Safety applications. In it, there are answers to some questions related to the number of samples needed for training accurate networks using FSL, the trade-off between number of samples and performance in this context, and others. This document is divided in 5 sections. The first one is the introduction. The second, called “Description of Few-Shot Learning”, discusses about Few-Shot Learning as a whole, and has 3 subsections titled “Applications of FSL”, “FSL Techniques” and “FSL Benchmarks for Computer Vision”. The third, “FSL for Human Behaviour Analysis applications”, discusses two recent papers featuring FSL in this thematic. The fourth is called “Few-Shot for Smart City Environment Safety applications” and, similarly to the third, includes the discussion of three recent papers on that matter. The fifth and last section is a brief conclusion.

2 Description of Few-Shot Learning

In its broader sense, Few-Shot Learning is a category of Machine Learning techniques where the training dataset contains a limited number of supervised samples.

Since this is a broad description, common ML techniques such as data augmentation would be considered FSL as well, since this method enables accurate training with a limited sampling of annotated data. Nevertheless, in the context

of this article the definition of Few-Shot Learning is constrained to the following: a family of Machine Learning methods that reuse previous knowledge for training accurate models using very limited training samples.

It is noteworthy to say that in the context of Computer Vision FSL is usually referred to image classification tasks, although it can be used also for object detection, semantic segmentation and hybrid tasks such as image description with text, among others.

2.1 Applications of FSL

Few-Shot Learning allows to train models with limited training datasets, thus reducing the amount of samples needed for supervised learning processes. There are several scenarios where reduced datasets are needed, and consequently FSL techniques can be applied:

- Extreme data. Data is called “extreme” when it has at least one of these characteristics: great volume, great variations in values, great complexity or sparsity. Few-Shot Learning can be applied in those situations where acquiring sufficient amounts of labelled data is hard or impossible. It is the case of analyzing ancient Chinese documents [22] and hieroglyphs (oracle characters) [14]; or new drug discovery problems as discussed by Altae-Tran *et al.* [1] for example.
- Cut costs. Annotating large amounts of data is a time-consuming process. Also, processing large datasets costs more computational power than smaller ones. Since FSL can be used to reduce the amount of data needed for training, it can be applied to reduce costs in both dataset generation and computer utilization expenses.
- Improve learning processes. In some cases, Few-Shot can be applied to parts of datasets to increase the quality of training rounds. This is the case of imbalanced datasets where some classes are underrepresented in relation to others within the same dataset. It also can be used for *weakly-supervised learning* when the dataset contains mixtures of annotated and non-annotated data. In addition, FSL is capable of boosting some *transfer learning* methods such as *domain adaptation*.

Also, beside the general applications mentioned above, there are many possible fields of application. This document focuses on Computer Vision, but it is noteworthy to mention that FSL is widely applied in all sorts of Machine Learning related methods: Natural Language Processing (NLP), language translation, regression, reinforcement learning, data classification, sound recognition [8], sensor calibration [43], smart city insight transfer [39], etc.

2.2 FSL Techniques

Some techniques use the knowledge for reducing the complexity of the Neural Networks graphs themselves, while others use it for optimizing the weights and

parameters of the artificial neurons and the gradients. In this section the two most used FSL techniques are presented. Both intend to reuse prior knowledge to reduce the amount of data needed for successful learning.

Different taxonomies have been proposed for FSL techniques in several surveys and review articles, offering formal descriptions for transfer learning, meta learning, distance learning, embedding learning, etc. [2, 16, 25, 41]. It is noteworthy to say that many of these techniques share common traits and, depending on the context of application, some of them do not present sufficient unique characteristics to be distinguished from one another. For that reason, at this point in time it cannot be assumed that the categories themselves are standardized and fixed. In this regard, what it is called “multi-task learning” in one paper can be part of the “meta-learning” family in another. In this document there are short discussions about a couple of techniques but the category names are not intended to be taken as fixed, immutable references.

Metric Learning. In some papers it is also referred as “embedding learning” [41]. This technique consists on transforming the input data (the samples) into a lower-dimensional embedding located in a space where distances represent similarity between the different encoded samples. In the embedding space, two similar samples would lay at small distances whereas dissimilar samples will be far from each other. This sample discrimination can be applied to effectively reuse prior knowledge about the samples themselves, and thus drastically reduce the training samples needed for learning. This is possible mainly because the embedding space allows for fast and simple comparisons of samples, whereas in other approaches complex operations are needed to effectively process the embeddings into predictions. In this case the encoding functions are already pre-trained, and can be fine-tuned with few examples by applying a specific loss function that forces the neural network’s parameters to encode the samples in a given space in where the distances are related to the samples’ similarity. In this sense, metric learning can be used for other applications besides Few-Shot Learning [15]. It has proven to be effective for learning tasks that involve comparison like signature verification [6], product design [4] or face recognition [26], among others.

In distance metric learning, the main feature is the training loss function, as it is responsible for setting the characteristics of the embedding space. The most used loss functions in metric learning are contrastive loss [7], triplet loss [34, 42] and multi-class N-pair loss [36]. Recent research works explore novel loss functions for specific tasks, like Additive Angular Margin Loss (ArcLoss) for face recognition [10]. Other approaches use ensemble loss functions for better generalization of metric learning models [45]. The encoding space is usually Euclidean for simplicity reasons, but there are other research lines where both known and novel geometries are explored.

Multi-task Learning. In some research articles it is also called “meta-learning” [2], because “multi-task learning” is part of the that family of ML algorithms. This approach uses learning procedures where different learning tasks are trained

simultaneously, thus achieving to share parameters between them which gradually increases the generalization capabilities of the model. In order to learn different tasks, the model needs to make abstractions and learn more generic information. In multi-task learning there are two types of knowledge: task-agnostic and task-specific information. The first of them refers to all the preprocessing steps needed to aggregate and interpret features into embeddings. The second is about how to interpret such embeddings into actual predictions, which is specific for each given task.

The most common approach of multi-task learning is to directly share common layers (the encoder) and put different ones at the end in order to process each task’s output. In many cases, constraints would be applied to the artificial neurons so the Few-Shot task can only update parameters from the task-specific layers, while the encoder’s parameters (task-generic or task-agnostic) can be changed by all of the tasks. Another approach would be to have separate CNN encoders for each task and them concatenate the resulting feature vectors at the end of the feature extractor. In this case all encoders would contribute to the same loss function, and regularization terms would be applied to force the parameters of the different Neural Network layers to be similar, thus aligning the training process of each individual encoder.

2.3 FSL Benchmarks for Computer Vision

As of today, the most important benchmark for testing Few-Shot Learning methods applied to Computer Vision is Google Research’s Meta-Dataset [37], “a dataset of datasets for learning to learn from few examples”. More specifically, the most complete benchmark is its second version, named VTAB+MD [11], which is a combination of the Visual Task Adaptation Benchmark (VTAB) [47] and the Meta-Dataset. The Meta-Dataset is composed of a selection of images from many other datasets, namely ImageNet, Omniglot, FGVC-Aircraft, Birds-CUB-200-2011, DTD, Quick Draw, FGVCx Fungi, VGG Flower, Traffic Signs (GTSRB) and MSCOCO. It has a total of 4934 different classes. In addition to the collection of images, it has an unique protocol for evaluating Few-Shot algorithms, which improves the two previous FSL benchmark dataset, Omniglot [20] and Mini-ImageNet [38], in several ways:

1. The combination of several datasets of different domains results in a more realistic data heterogeneity, which allows for testing the model’s capacity to generalize unseen datasets. It also tests their ability for training with unbalanced datasets in terms of number of samples per class.
2. Its large scale and dataset mixture allows the evaluation protocol to take into account the ability of the different few-shot techniques to form relationships between classes. For example, if the model can do fine-grain classification for detecting different species of birds while being able to distinguish birds from common objects like dinner tables.
3. The data has been selected and structured in a such a way that it mimics realistic class unbalance, allowing the evaluation of the different models in different numbers of samples per class.

4. This benchmark has a clear evaluation guideline that combines different ML tasks, which in turn allows to evaluate the models' capacity of learning from different sources. It also has a set of baselines aimed to measure the benefit of meta-learning, i.e. whether the training process benefits from using more data, learning from different sources, reusing knowledge from pre-trained weights and meta-training parameters.

The Meta-Dataset protocol computes the model's rank by decreasing order of accuracy. As of today, the best performing model on Meta-Dataset benchmark 2021's model code-name "TSA", from Task-Specific Attention [23,24], achieved number 1 position on the Meta-Dataset with 1.65 mean rank across all datasets. The paper, titled "Cross-domain Few-shot Learning with Task-specific Adapters", proposes a novel multi-task learning methodology that combines task-agnostic with task-specific approaches. This is done by directly attaching task-specific adapters to pre-trained task-agnostic models. The article also features a new architecture for said adapters.

3 FSL for Human Behaviour Analysis Applications

In the context of Computer Vision, there are several tasks than can help Human Behaviour Analysis (HBA). All of them can be predicted using Deep Neural Networks, and for most of them there are recent research articles featuring FSL. Here are some examples of such tasks along with a Few-Shot citation: face detection, face recognition [29,40,48], facial expression recognition [9], person detection, person recognition, pose estimation, action recognition [12], person re-identification [13], person tracking [21], hand gesture recognition [30], motion prediction [46], etc. In this section two recent research articles on FSL are shortly discussed, each featuring a different HBA application.

Face Recognition. Few-Shot Learning is a natural match for face recognition. As explained in Sect. 2.2, metric learning is a technique that is not only used for FSL, but also other applications such as face recognition. In this regard, the use of the same loss functions such as triplet loss [34,42] or ArcFace loss [10] make face recognition to naturally be a Few-Shot Learning application as well.

A. Putra and S. Setumin published in 2021 an article in which they made a performance comparison of different activation functions on Siamese Networks for Face Recognition [29]. Their Few-Shot technique was to combine multi-task learning (parameter tying) with metric learning. In the article, the researchers trained a Siamese network using different activation functions for the final embedding prediction, which would in turn be used to measure distance (similarity) between the outputs of the two branches of the network. The tested activation functions were sigmoid, softmax, tanh, softplus and softsign. They tested each activation function by learning 1 to 19 classes with only 1 supervised sample (One-Shot Learning). In average, sigmoid activation function achieved 92% accuracy to recognize 19 new faces with only one sample for training using this FSL approach.

Action Recognition. In a recent publication (2021) by Mark Haddad [12], he explored a method for One-Shot and Few-Shot learning for action recognition tasks. The models had to learn classification of 10 different actions: bend, jack, jump, pump-jump, run, side, skip, walk, wave 1 and wave 2. For that matter, he encoded the movement of a few frames into an optical flow vectorized with KMeans method. Therefore, movements were parametrized as KMeans clusters, that can be treated as probability distributions for different sets of points that represent each action. Using Kullback-Leibler divergence loss, a model can be trained to perform metric learning with the aforementioned KMeans clusters. In his thesis, Haddad performed tests for both One-Shot Learning and for K-shot, being K a number between 1 and 8 samples. On the Weizmann Human Action dataset, he achieved an average of 89.4% classification accuracy for One-Shot Learning, increased to 98% with 8-Shot Learning (see Fig. 1).

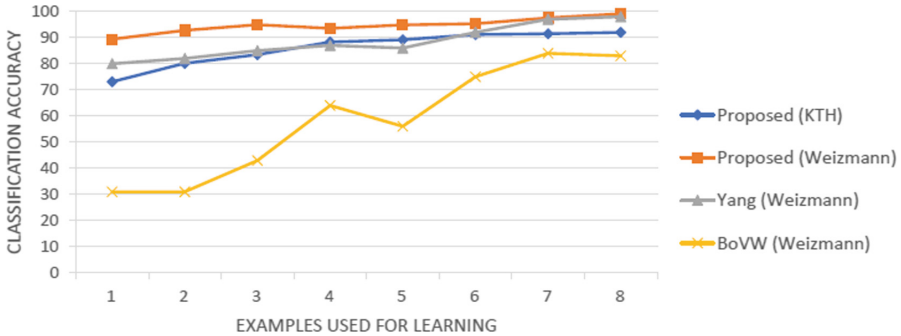


Fig. 1. Extracted Fig. 3.5 from Haddad’s thesis [12]. Quote: “Classification accuracy comparison between proposed method and others.” It shows results for two different datasets: KTH and Weizmann.

4 Few-Shot for Smart City Environment Safety Applications

In the context of Smart City Environment Security, there are some Computer Vision applications that are extensively applied, specially related to traffic and pedestrian surveillance. Yet another application is assessment for disasters and emergencies that threaten safety of cities, both for its population or its infrastructures. Emergency response is a good fit for the application of Few-Shot Learning methods since they are events where on-site data is usually scarce; and whose changing nature make the models require flexibility and adaptability. There are recent examples where FSL was effectively applied in emergencies such as the ongoing pandemic caused by Covid-19. As one of many examples, Lai *et al.* [19] used Deep Learning algorithms trained using Few-Shot techniques to classify lung lesions caused by COVID-19 using the small datasets available at the time. In this section three research articles where FSL was applied are discussed.

The first of them is related to Smart Cities Environment Safety, while the other two are focused on disaster damage assessment as part of emergency response applications.

Face Anti-spoofing. Recent developments in Deep Learning technology for face reenactment have made possible very realistic impersonating. This is a risk for governments and administrations, because it denies one of the traditional identity proofs, which is face identification in live video. For that reason, many researchers are working towards face anti-spoofing.

In 2021 Yang *et al.* published a paper on face anti-spoofing where they used FSL techniques for achieving One-Shot Learning of new unseen face reenactments [44]. They proposed a FSL technique called Few-Shot Domain Adaptation, which combines data augmentation using photorealistic style transfer and a complex compound loss function \mathcal{L}_{total} .

$$\mathcal{L}_{total} = \mathcal{L}_{CLS} + \lambda_1 \mathcal{L}_{Cont} + \lambda_2 \mathcal{L}_{Adv} + \lambda_3 \mathcal{L}_{Lfc} \quad (1)$$

\mathcal{L}_{CLS} is cross-entropy loss. \mathcal{L}_{Cont} is contrastive semantic alignment, which is a term used for reducing distances in the encoding space for positive pairs and increase them for the negatives (a form of metric learning). \mathcal{L}_{Adv} is the progressive adversarial learning, a term used for constraining the parameters of the different domain discriminators. \mathcal{L}_{Lfc} is the less-forgetting constrain, which is a mean square error penalization to prevent the parameters from shifting excessively during training. The λ s are trade-off parameters.

With this method they were able to detect face spoofing by training with a single sample and they managed to beat the previous best working model by 5% points in one of the tests (CMOS-ST benchmark when targeting HTER using protocol $C \rightarrow O$).

Disaster Damage Assessment. In this case there are two articles of interest that are noteworthy to discuss.

E. Koukouraki, L. Vanneschi and M. Painho published in 2021 a research article about urban damage detection after earthquake incidents from satellite images, trained using FSL techniques. In it, they tested four different methodologies: cost-sensitive learning, oversampling, undersampling and Prototypical Networks [35], each of them employing different data balancing methods. The best working model was Prototypical Networks (ProtoNets), which combine multi-task learning with metric learning by using distance loss functions to train Siamese neural networks (see Fig. 2). It achieved precision and recall superior to 50% in all four classes (undamaged, minor damage, major damage and destroyed) with average F-score of 64% over all classes.

In a similar work, J. Bowman and L. Yang studied further Few-Shot Learning methods for post-disaster damage detection from satellite images [5]. In this case, they use a FSL technique called “feature re-weighting” that is used for parameter tying (a type of multi-task learning approach) between classification and

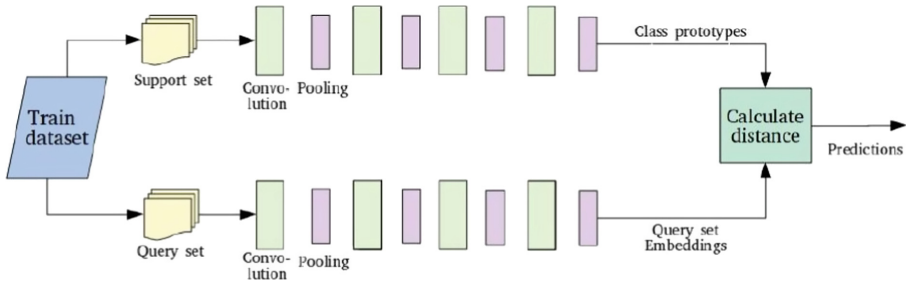


Fig. 2. Extracted Fig. 5 from Koukouraki *et al.* article [17]. It shows the network architecture used for the ProtoNets approach

object detection tasks. They used the encoder (feature extractor) from YOLO [31], and then added an additional CNN model, the re-weighting module, to generate re-weighting vectors for each class. These vectors are concatenated to the embeddings from the feature extractor in order to extend them before the final prediction takes place. In this manner, the re-weighting effectively fuses task-agnostic and task-specific functionalities thus providing FSL capabilities. With the addition of this module, they achieved to improve the mean average precision (mAP) of the baseline YOLO from 0.270 to 0.289 when training with 30 samples. This means that when applying feature re-weighting, the model gave a mAP almost 2% points better, when training with only 30 samples per class.

5 Conclusion

This review article about Few-Shot Learning techniques is different from others because it is heavily focused on Computer Vision Applications based on Deep Convolutional Neural Networks. Moreover, Sects. 3 and 4 feature commented examples of articles where FSL techniques were used for face recognition, action recognition, face anti-spoofing and post-disaster damage assessment from satellite. These examples show how FSL algorithms can be effectively used to learn from few supervised samples with competitive metrics in comparison to baseline models. They show how metric learning and multi-task learning are FSL techniques that combined offer current State of the Art results when training with few samples. As final though, it is noteworthy to mention that this paper emphasizes the importance of the loss function when it comes to *meta-learning* and *transfer learning*.

References

1. Altae-Tran, H., Ramsundar, B., Pappu, A.S., Pande, V.: Low data drug discovery with one-shot learning. *ACS Cent. Sci.* **3**(4), 283–293 (2017)
2. Antonelli, S., et al.: Few-shot object detection: a survey. *ACM Comput. Surv. (CSUR)*, 6–7 (2021)

3. Argüeso, D., et al.: Few-shot learning approach for plant disease classification using images taken in the field. *Comput. Electron. Agric.* **175**, 105542 (2020)
4. Bell, S., Bala, K.: Learning visual similarity for product design with convolutional neural networks. *ACM Trans. Graph. (TOG)* **34**(4), 1–10 (2015)
5. Bowman, J., Yang, L.: Few-shot learning for post-disaster structure damage assessment. In: *Proceedings of the 4th ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery*, pp. 27–32 (2021)
6. Bromley, J., Guyon, I., LeCun, Y., Säckinger, E., Shah, R.: Signature verification using a “Siamese” time delay neural network. In: *Advances in Neural Information Processing Systems*, vol. 6 (1993)
7. Chopra, S., Hadsell, R., LeCun, Y.: Learning a similarity metric discriminatively, with application to face verification. In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005)*, vol. 1, pp. 539–546. IEEE (2005)
8. Chou, S.Y., Cheng, K.H., Jang, J.S.R., Yang, Y.H.: Learning to match transient sound events using attentional similarity for few-shot sound recognition. In: *ICASSP 2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 26–30. IEEE (2019)
9. Ciubotaru, A.N., Devos, A., Bozorgtabar, B., Thiran, J.P., Gabrani, M.: Revisiting few-shot learning for facial expression recognition. *arXiv preprint [arXiv:1912.02751](https://arxiv.org/abs/1912.02751)* (2019)
10. Deng, J., Guo, J., Xue, N., Zafeiriou, S.: ArcFace: additive angular margin loss for deep face recognition. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4690–4699 (2019)
11. Dumoulin, V., et al.: Comparing transfer and meta learning approaches on a unified few-shot classification benchmark. *arXiv preprint [arXiv:2104.02638](https://arxiv.org/abs/2104.02638)* (2021)
12. Haddad, M.: An instance-based learning statistical framework for one-shot and few-shot human action recognition. Ph.D. thesis, Concordia University (2021)
13. Han, P., et al.: HMMN: online metric learning for human re-identification via hard sample mining memory network. *Eng. Appl. Artif. Intell.* **106**, 104489 (2021)
14. Han, W., Ren, X., Lin, H., Fu, Y., Xue, X.: Self-supervised learning of ORC-BERT augmentator for recognizing few-shot oracle characters. In: *Proceedings of the Asian Conference on Computer Vision* (2020)
15. Koch, G., Zemel, R., Salakhutdinov, R., et al.: Siamese neural networks for one-shot image recognition. In: *ICML Deep Learning Workshop*, vol. 2, Lille (2015)
16. Köhler, M., Eisenbach, M., Gross, H.M.: Few-shot object detection: a survey. *arXiv preprint [arXiv:2112.11699](https://arxiv.org/abs/2112.11699)* (2021)
17. Koukouraki, E., Vanneschi, L., Painho, M.: Few-shot learning for post-earthquake urban damage detection. *Remote Sens.* **14**(1), 40 (2021)
18. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: *Advances in Neural Information Processing Systems*, vol. 25 (2012)
19. Lai, Y., et al.: 2019 novel coronavirus-infected pneumonia on CT: a feasibility study of few-shot learning for computerized diagnosis of emergency diseases. *IEEE Access* **8**, 194158–194165 (2020)
20. Lake, B.M., Salakhutdinov, R., Tenenbaum, J.B.: Human-level concept learning through probabilistic program induction. *Science* **350**(6266), 1332–1338 (2015)
21. Lee, J., Ramanan, D., Girdhar, R.: MetaPix: few-shot video retargeting. *arXiv preprint [arXiv:1910.04742](https://arxiv.org/abs/1910.04742)* (2019)
22. Li, B., Wei, J., Liu, Y., Chen, Y., Fang, X., Jiang, B.: Few-shot relation extraction on ancient Chinese documents. *Appl. Sci.* **11**(24), 12060 (2021)

23. Li, W.H., Liu, X., Bilen, H.: Cross-domain few-shot learning with task-specific adapters. arXiv preprint [arXiv:2107.00358](https://arxiv.org/abs/2107.00358) (2021)
24. Li, W.H., Liu, X., Bilen, H.: Improving task adaptation for cross-domain few-shot learning. arXiv preprint [arXiv:2107.00358](https://arxiv.org/abs/2107.00358) (2021)
25. Li, X., Yang, X., Ma, Z., Xue, J.H.: Deep metric learning for few-shot image classification: a selective review. arXiv preprint [arXiv:2105.08149](https://arxiv.org/abs/2105.08149) (2021)
26. Liu, W., Wen, Y., Yu, Z., Li, M., Raj, B., Song, L.: SphereFace: deep hypersphere embedding for face recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 212–220 (2017)
27. Nikolenko, S.I.: Synthetic Data for Deep Learning. SOIA, vol. 174. Springer, Cham (2021). <https://doi.org/10.1007/978-3-030-75178-4>
28. Picon, A., San-Emeterio, M.G., Bereciartua-Perez, A., Klukas, C., Eggers, T., Navarra-Mestre, R.: Deep learning-based segmentation of multiple species of weeds and corn crop using synthetic and real image datasets. *Comput. Electron. Agric.* **194**, 106719 (2022)
29. Putra, A.A.R., Setumin, S.: The performance of Siamese neural network for face recognition using different activation functions. In: 2021 International Conference of Technology, Science and Administration (ICTSA), pp. 1–5. IEEE (2021)
30. Rahimian, E., Zabihi, S., Asif, A., Atashzar, S.F., Mohammadi, A.: Trustworthy adaptation with few-shot learning for hand gesture recognition. In: 2021 IEEE International Conference on Autonomous Systems (ICAS), pp. 1–5. IEEE (2021)
31. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 779–788 (2016)
32. Romera-Paredes, B., Torr, P.: An embarrassingly simple approach to zero-shot learning. In: International Conference on Machine Learning, pp. 2152–2161. PMLR (2015)
33. Russakovsky, O., et al.: ImageNet large scale visual recognition challenge. *Int. J. Comput. Vision* **115**(3), 211–252 (2015). <https://doi.org/10.1007/s11263-015-0816-y>
34. Schroff, F., Kalenichenko, D., Philbin, J.: FaceNet: a unified embedding for face recognition and clustering. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 815–823 (2015)
35. Snell, J., Swersky, K., Zemel, R.: Prototypical networks for few-shot learning. In: Advances in Neural Information Processing Systems, vol. 30 (2017)
36. Sohn, K.: Improved deep metric learning with multi-class n-pair loss objective. In: Advances in Neural Information Processing Systems, vol. 29 (2016)
37. Triantafillou, E., et al.: Meta-dataset: a dataset of datasets for learning to learn from few examples. arXiv preprint [arXiv:1903.03096](https://arxiv.org/abs/1903.03096) (2019)
38. Vinyals, O., Blundell, C., Lillicrap, T., Wierstra, D., et al.: Matching networks for one shot learning. In: Advances in Neural Information Processing Systems, vol. 29 (2016)
39. Wang, J., Li, W., Qi, X., Ren, Y.: Transfer knowledge between cities by incremental few-shot learning. In: Gao, H., Wang, X. (eds.) CollaborateCom 2021. LNICST, vol. 407, pp. 241–257. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-92638-0_15
40. Wang, L., Li, Y., Wang, S.: Feature learning for one-shot face recognition. In: 2018 25th IEEE International Conference on Image Processing (ICIP), pp. 2386–2390. IEEE (2018)
41. Wang, Y., Yao, Q., Kwok, J.T., Ni, L.M.: Generalizing from a few examples: a survey on few-shot learning. *ACM Comput. Surv. (CSUR)* **53**(3), 1–34 (2020)

42. Weinberger, K.Q., Saul, L.K.: Distance metric learning for large margin nearest neighbor classification. *J. Mach. Learn. Res.* **10**(2), 207–244 (2009)
43. Yadav, K., Arora, V., Jha, S.K., Kumar, M., Tripathi, S.N.: Few-shot calibration of low-cost air pollution (PM2.5) sensors using meta-learning. arXiv preprint [arXiv:2108.00640](https://arxiv.org/abs/2108.00640) (2021)
44. Yang, B., Zhang, J., Yin, Z., Shao, J.: Few-shot domain expansion for face anti-spoofing. arXiv preprint [arXiv:2106.14162](https://arxiv.org/abs/2106.14162) (2021)
45. Zabihzadeh, D.: Ensemble of loss functions to improve generalizability of deep metric learning methods. arXiv preprint [arXiv:2107.01130](https://arxiv.org/abs/2107.01130) (2021)
46. Zang, C., Pei, M., Kong, Y.: Few-shot human motion prediction via learning novel motion dynamics. In: *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, pp. 846–852 (2021)
47. Zhai, X., et al.: A large-scale study of representation learning with the visual task adaptation benchmark. arXiv preprint [arXiv:1910.04867](https://arxiv.org/abs/1910.04867) (2019)
48. Zheng, W., Gou, C., Wang, F.Y.: A novel approach inspired by optic nerve characteristics for few-shot occluded face recognition. *Neurocomputing* **376**, 25–41 (2020)
49. Zhou, X., Girdhar, R., Joulin, A., Krähenbühl, P., Misra, I.: Detecting twenty-thousand classes using image-level supervision. arXiv preprint [arXiv:2201.02605](https://arxiv.org/abs/2201.02605) (2021)
50. Zhuang, F., et al.: A comprehensive survey on transfer learning. *Proc. IEEE* **109**(1), 43–76 (2020)