Research on AI Small Sample Image Recognition

Jixin Chen[1,a,*] and Songrong Lv[2,b]
[1]Wuhan Britain-China School, Wuhan, China
[2]NorthernYucai Foreign Language School, Heping Street, Shenyang, China
*

Keywords: AI Small Sample Image Recognition, Data Enhancement, Meta Learning.

Abstract: In computer vision and image recognition technology, as the input resolution changes, the recognition effect of convolutional neural network methods also varies. Meta learning allows computers to simulate the human brain and learn how to learn, which can achieve image classification more efficiently and flexibly. this paper will first focus on technical progress, application field expansion and challenge and response of AI image recognition. This article introduces image classification based on meta learning, followed by image classification based on data augmentation and image grading based on transfer learning, Finally, a summary and outlook on small sample image recognition technology were presented. With the continuous develop of deep learning, transfer learning, meta learning and other technology, AI small sample image recognition models will be more optimized and able to achieve higher recognition accuracy with less data. The technology will play an important role in more fields and bring more convenience and value to mankind.

1 INTRODUCTION

There is a significant difference between deep learning and human intelligence, where humans are good at identifying new classes of objects with very small numbers of samples, and deep learning can easily produce overfitting in this case. Therefore, the small sample problem has become one of the important research directions in the field of machine learning (Ge, Liu, Wang, et al. 2022). Small sample learning is quite important. Not only can it be used to learn rare cases. For example, for animals or plants with scarce information. For many reasons, accurate and comprehensive data (such as personal privacy or face data) requires only a small amount of prior information to accurately classify rare species of animals and plants. It also reduces the cost of data collection and calculation, because it requires only a small amount of data to train the model. Eliminate the high cost of collecting the data, model training at less cost. Its application to natural language processing. Medical applications, acoustic signals processing, image application. Character recognition and object recognition and other field, In the field of limited computing power, such as surveillance cameras and unmanned dring, the small sample image recognition task can well solve the problem because there is not enough computing power to process so much data. Nowadays, how to retain the powerful knowledge representation ability of deep learning while quickly learning useful knowledge from a small amount of data for image recognition has become a hot topic, at present, there are many methods such as meta learning, transfer learning and data enhancement. In the following paper is going to introduce several functions of the technology. Meta learning, also known as "Learning to learn", aims to train a "meta model", that can quickly adapt and learn new tasks. Unlike traditional machine learning algorithms, which focus on training models on specific tasks and data sets, meta learning focuses more on learning a general learning strategy or optimization method that can be applied to a number of different tasks. Optimization based meta learning focuses on the optimization process itself, rather than directly optimizing model parameters. It works by training an optimizer or learning strategy that enables the model to quickly find the model, so that it can quickly adjust and optimize its performance in the face of unknown territory. MAML is one of the representation

[a] https://orcid.org/0009-0001-2524-041X
[b] https://orcid.org/0000-0004-3143-9654

Chen, J. and Lv, S.
Research on AI Small Sample Image Recognition.
DOI: 10.5220/0013422200004558

In *Proceedings of the 1st International Conference on Modern Logistics and Supply Chain Management (MLSCM 2024)*, pages 474-479
ISBN: 978-989-758-738-2

algorithms in meta learning based on optimization (Gao, Han, Zhang, et al., 2023). MAML is designed to train a model that is able to adapt rapidly on small amounts of new task data without having to train from scratch. The single sample data enhancement method refers to the processing of a single sample to improve the generalization ability and performance of the model.

2 META-LEARNING BASED METHOD FOR FEW-SHOT IMAGE CLASSIFICATION

Few-shot image classification based on Meta-Learning which is the design of models that can adapt and improve their learning strategies, in other words, "learning how to learn." It focuses on how to design algorithms to efficiently learn new tasks with small amounts of data, thus toward significantly improving the applicability of the model. This type of learning method is commonly used by humans, and can also be used to improve the construction of models that have the capacity to quickly adapt their parameters to new data distributions when faced with a new task, to learn a new task while maintaining the memory of an already completed task, and to purchase the knowledge learned on the old task to apply it to the processing of the new task, thereby improving the efficiency. In this section, meta-learning methods for few-shot image classification are discussed through two meta-learning methods, Model-Agnostic Meta-Learning and visual meta-learning methods respectively.

2.1 Model-Agnostic Meta-Learning

That is, to find the best set of parameters which can be quickly adapted to different models, in order to achieve the effect of fast convergence with only a small amount of data references, the model needs a large amount of prior knowledge to correct this initialized parameter without interruption, so that it can be adapted to a large number of different models.

2.2 MAML++

The literature proposes an enhancement method for MAML, which presents five problems of MAML, namely, instability during training process, limitations in model generalization performance, reduction of flexibility of the framework, high computational cost, and the need for time-consuming and labor-intensive parameter tuning of the model before it can handle a new task, in order to ameliorate the above problems, the literature proposes MAML++(Edwards, 2018), which is an improved MAML-based variant, which provides better flexibility and more stable training, while improving efficiency and generalization performance. To solve the instability problem, the literature implements the method of optimizing multi-part loss to compute the target set loss of the minimized base network at each step towards the support set task, proposing that the minimized loss is the weighted sum of the target set loss after each support set loss update, and by using it to compute the weighted sum, the instability of the MAML due to backpropagation can be mitigated efficiently; furthermore, the literature addresses the issue of high overhead and high cost problem, the literature proposes annealing the derivative order during the training process, i.e., using the first-order gradient to compute the first stage of the training phase and using the second-order gradient for the rest of the training phase; to address the problem of the limitation of the generalization performance, the literature proposes to learn a different learning rate and direction for each adaptation process of the underlying network, which reduces the required memory and arithmetic while providing flexibility and generalization performance, and may help to mitigate overfitting. may help mitigate overfitting.

2.3 Context-Aware Meta-Learning

In recent years, the problem that purely visual domain models hardly have the ability to detect new objects when reasoning has gained much attention. To address this problem, a meta-learning algorithm has been proposed in the literature. That is, new visual concepts are learned during inference but not fine-tuned to simulate large-scale language models, and meta-learning is redefined as the process of modeling sequences of data points with known labels and data points with unknown labels by resorting to a pre-trained feature extractor.

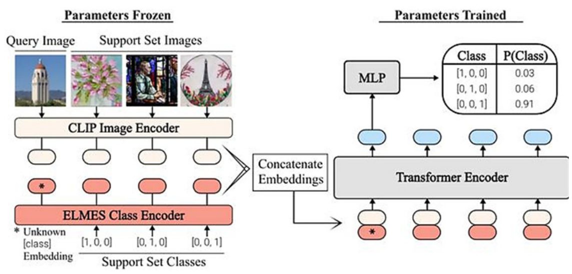

Figure 1: the process of CAML.

The general framework of the CAML method is shown in figure1 and consists of three components: a CLIP image encoder with frozen parameters, a Fixed Equal Length and Maximum Equal Angle Set (ELMES)-like encoder (and a non-causal sequence model).

(ELMES) class encoder (Fikus, 2018) and a non-causal sequence model. As shown in the figure 1, CAML encodes and supports the images using a pre-trained feature extractor and extracts them into a low-dimensional representation. The categories of the support set are encoded using the ELMES category encoder, but since the category of the query is unknown, the literature proposes a special learnable unknown marker embedding. Subsequently, CAML associates each image embedding with the corresponding query to form an input sequence that is significant after large-scale training. In order to accurately assess the performance of this sub-method for image recognition classification, the literature compares it to standard meta-learning evaluation benchmarks for generic object recognition, fine-grained image classification, inter-domain image classification, and unnatural image classification. In most of the evaluation environments, CAML performs quite competitively (Clark, 2021).

3 FEW-SHOT IMAGE CLASSIFICATION BASE ON DATA AUGMENTATION

The main problem with few-shot image classification tasks is that the lack of sufficient data, which means expanding the amount of data by manually processing the existing data or generating data through computation is a solution; in other words, just make some changes to the existing data, such as rotating or cropping, transposing, etc. to change the data, and then new data can be obtained. Even the simplest form of image processing can multiply the data by several times, and the effect is extraordinary. This way of making a limited amount of data produce a value equivalent to more data by not substantially enhancing the data is data enhancement. this paper will discuss the methods of single-sample data enhancement versus multi-sample data enhancement.

3.1 Methods Based on Single-Sample Data Enhancement

That is, the enhancement of the sample around the sample itself to carry out operations, including geometric transformation of the image such as rotation, cropping, zoom, etc., but also in the color change or insert noise blurring process.

The literature proposes a flexible data enhancement approach to vary the representation of an image without changing its semantics through the diffusion model DA-fusion, taking into account the following three conditions: applicability to all images, minimization of the tuning of a specific dataset, and the balance between reality and the virtual (Jonathan, Tim, 2022). The literature proposes two approaches to prevent Internet data leakage, model-centered leakage prevention and data-centered leakage prevention, and also, since standard data augmentation applies to all images, the literature tries to keep up with the flexibility of the images through diffusion-based augmentation techniques to enhance the credibility of the image generation. The pair diffuses the model by inserting a new embedding in the generated text encoder to get a new model that enhances the credibility of the new image concept.

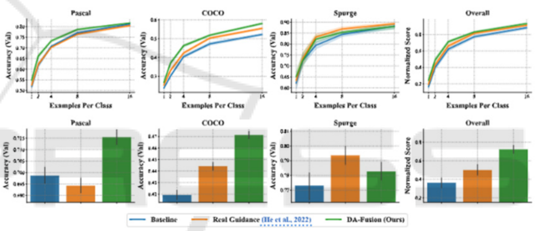

Figure 2: Comparison chart.

The results are shown in Figure 2, the DA-Fussion model greatly exceeds the baseline demand and occupies a competitive position in general, possessing better results. (Trabucco, 2023)

3.2 Methods for Multi-Sample Data Enhancement

Unlike single-sample data augmentation, multi-sample data augmentation utilizes multiple related or different samples to generate new samples through computation

The literature is based on the support set X = {(x, y)} K, N, where x represents an image, while y represents its label, K represents the number of images contained in each class, and N denotes the number of classes. The goal is to learn a function F θ (x) → y (which maps an image x to the corresponding label y). Following the idea of being similar to the pre-training method and restricting to small sample parameter training, the paper constructed a new method DISEF (Diversified In-domain Synthesis with Efficient Fine-tuning) based on the foundation of the pre-trained VLM as shown below in Figure 3:

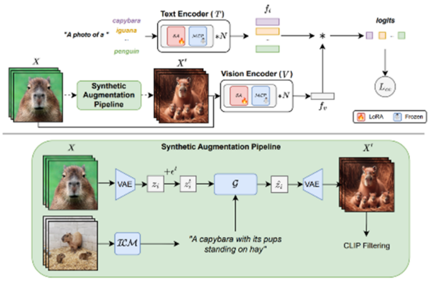

Figure: 3 SAP process (Victor, et al. 2023).

The Synthetic Augmentation Pipeline (SAP) is used for fine-tuning, which involves efficient tuning of the large model.

1. First, the input image (X) is trained by SAP to produce more training data (X').
2. Second, the images (X) and data (X') are passed into Vision Encoder to generate visual features (fv); meanwhile, the class labels are combined with predefined templates through Text Encoder (T) to generate text features (ft).
3. Then, logits = sim (fv, ft) (sim is a cosine-like function) and cross-entropy loss Lce are computed.
4. Finally, the initial model is changed by adding LoRA layers to the query function and embedding the value (V) of the self-attention (SA) layers in the text and vision encoder. (Wang, 2023)

The following figure illustrates the SAP process:

1. add labels to the image set (X) using the Image Description Model (ICM) while projecting them into the Stable Diffusion latent space.
2. Introduce noise into the latent vector and randomize the playback using each type of label to get a synthetic image (X')
3. filter the synthetic images (X') using CLIP to retain the synthetic images that meet their category expectations.

4 FEW-SHOT IMAGE CLASSIFICATION BASED ON MIGRATION LEARNING

Although machine learning is being applied in an increasing number of domains, existing supervised learning requires large amounts of labeled data, which is both time-consuming and costly, and migration learning is therefore receiving increasing attention in order to leverage previously labeled data to ensure model accuracy for new tasks, in other words, it is the conversion of knowledge gained from learning in a domain to learning from existing datasets in order to improve and enhance the learning efficiency of the model. Two important concepts are domain and task; a domain is a specific area, e.g., a mobile game and a computer game are two different domains, and a task is something to be done, e.g., a psychological test and a physical fitness test are two different tasks. The key points of transfer learning are to study the shared knowledge that can be transferred between different domains, to study how to write specific algorithms to provide the shared knowledge and enable the transfer, and to study in which cases transfer is more appropriate and whether the transfer method is suitable for a particular use case.

4.1 A Transfer Learning Approach Based on Feature Mapping

Which is focusing on how to find out the common features between the source and target domains and mapping them from the original feature space to the new feature space. The potential shared feature space found will go to build a bridge between the source and shared domains. If two domains are related and they share some data (e.g.the same parameters or common variables) some of which cause different distributions but others do not, parameters can be found that do not affect the distributions to make the distributions of the source and target domains similar. In this way, the source domain data in that space is the same as the target domain data, which leads to better utilization of the existing labeled data for training in the new space. In this paper, the paper will develop a discussion on the mapping based migration learning small sample image recognition method combined with meta-learning.

4.2 Sink-Horn Algorithm

The framework of the proposed scheme in the literature mainly follows the improvements made in the preprocessing of the latent space output of the backbone model β, so that the varying eigenvectors are differently normalized and the results will be more accurate due to the algorithm operated by the Sink horn mapping algorithm.

As in figure 4, the literature uses CNN to convert the image into potential space and transforms the algorithm to preprocess the vectors and subsequently processes one of the test data and maps it using the SINKHOEN algorithm to compare it with the central domains, which are projected into the potential space in the same way as the data that is not yet labeled, and the closest class will be assigned to be used as a prediction at the test samples, an algorithm that helps to conversion of distributions for individual classes to Gaussian classes.

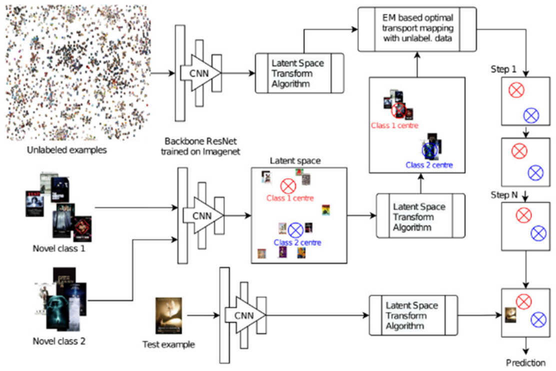

Figure 4: Mapping Process (Tomáš, Daniel, Pavel, 2021).

The performance of the method was tested through small sample datasets CIFAR-FS and CUB, and the method proposed in the literature was extremely accurate under one to five tests, outperforming other forms of feature preprocessing, and achieved the second place in the METADI competition for the year 2020.

5 CONCLUSION

The transferable knowledge in small sample learning is highly dependent on the distribution of training data. If the distribution of training data is significantly different from that of the new class, the model may gain biased transmission ability. Therefore, studying the impact of data bias on model performance is of great significance. In addition, after summarizing and analyzing different methods, it was found that better transferable knowledge can be learned from training data related to the content. Case intensive related data and diverse categories can further enrich the diversity and relevance of the data. With the continuous development of artificial intelligence deep learning, small sample recognition will be significantly improved. Deep learning models, especially convolutional neural networks and converters, will be able to better learn and extract effective feature representations from limited data. The key to small sample recognition lies in the model's generalization ability, that is, whether the model can accurately recognize unseen images when only a small number of samples are seen. Future research will focus on developing models with stronger generalization ability. By introducing techniques such as meta learning and data augmentation, the model's ability to adapt to new samples will be improved. The core challenge of small sample learning lies in the scarcity of data. To address this challenge.

Researchers need to develop more effective mechanisms for data collection, tagging, and sharing. Meanwhile, transfer learning and other techniques can also be used to transfer models trained on large datasets to small sample tasks.

In the future, artificial intelligence small sample recognition technology will play an important role in many fields, promoting the intelligent transformation of related industries. With the continuous expansion of application scenarios, the paper believe that small sample image recognition technology will achieve even more brilliant achievements. At the same time, researchers also need to pay attention to issues such as data and privacy protection to ensure the healthy development of technology. In summary, artificial intelligence small sample image recognition technology has broad development prospects and enormous application potential. Future research will be dedicated to solving technical challenges and application scenarios, and better serving human society.

AUTHORS CONTRIBUTION

All the authors contributed equally and their names were listed in alphabetical order.

REFERENCES

Alec, R., Jong, W. K., Chris, H., Aditya, R., Gabriel, G., Sandhini, A., Girish, S., Amanda, A., Pamela, M., Jack, C., Gretchen, K., Ilya, S., 2021. Learning Transferable Visual Models From Natural Language Supervision. arXiv:2103.00020.

Antreas, A., Harrison, E., Amos, S.,2018, How to train your MAML. arXiv:1810.09502.

Arnav, C., Zhuang, L., Deepak, G. et al. 2023. One-for-all: Generalized lora for parameter. ArXiv:2306.07967.

Brandon, T., Kyle, D., Max, G., Ruslan, S., 2023. Effective Data Augmentation with Diffusion Models. arXiv:2302.07944.

Gao, P., Han, J. M., Zhang, R. R. et al., 2023. Llama-adapter v2: Parameter-efficient visual instruction model. ArXiv:2304.15010.

Ge, Y. Z., Liu, H., Wang, Y. et al. 2022. Survey on Deep Learning Image Recognition in Dilemma of Small Samples. Software.

Jonathan, H., Tim, S., 2022. Classifier-free diffusion guidance. ArXiv:2207.12598.
Matthew, F., John, J., Emily, J.K., Dustin, G.M., 2018. Equiangular tight frames that contain regular simplices. Linear Algebra and Its Applications, 98-138.
Tomáš, C., Daniel, V., Pavel, K., 2021, Transfer learning based few-shot classification using optimal transport mapping from preprocessed latent space of backbone neural network. arXiv:2102.05176.
Victor, G. T. C., Nicola, D'A., Yiming, W., Nicu, S., Elisa, R., 2023, Diversified in-domain synthesis with efficient fine-tuning for few-shot classification. arXiv:2312.03046.