remote sensing

MDPI

Article

An Effective Task Sampling Strategy Based on Category Generation for Fine-Grained Few-Shot Object Recognition

Shifan Liu, Ailong Ma *, Shaoming Pan and Yanfei Zhong

State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430072, China

* Correspondence: maailong007@whu.edu.cn

Abstract: The recognition of fine-grained objects is crucial for future remote sensing applications, but this task is faced with the few-shot problem due to limited labeled data. In addition, the existing few-shot learning methods do not consider the unique characteristics of remote sensing objects, i.e., the complex backgrounds and the difficulty of extracting fine-grained features, leading to suboptimal performance. In this study, we developed an improved task sampling strategy for few-shot learning that optimizes the target distribution. The proposed approach incorporates broad category information, where each sample is assigned both a broad and fine category label and converts the target task distribution into a fine-grained distribution. This ensures that the model focuses on extracting fine-grained features for the corresponding broad category. We also introduce a category generation method that ensures the same number of fine-grained categories in each task to improve the model accuracy. The experimental results demonstrate that the proposed strategy outperforms the existing object recognition methods. We believe that this strategy has the potential to be applied to fine-grained few-shot object recognition, thus contributing to the development of high-precision remote sensing applications.

Keywords: fine-grained; few-shot; object recognition; sampling strategy

Citation: Liu, S.; Ma, A.; Pan, S.; Zhong, Y. An Effective Task Sampling Strategy Based on Category Generation for Fine-Grained Few-Shot Object Recognition. *Remote Sens.* **2023**, *15*, 1552. https://doi.org/10.3390/rs15061552

Academic Editor: Gwanggil Jeon

Received: 3 February 2023
Revised: 9 March 2023
Accepted: 10 March 2023
Published: 12 March 2023

1. Introduction

The field of remote sensing has seen rapid development in the last few decades, with applications in areas, such as environmental resources, hydrology, and geology. Object recognition, which involves mapping visual features to object classes, is a crucial problem in remote sensing image analysis and has diverse civil applications, such as geographic information system (GIS) mapping, transportation planning, and navigation [1–7]. With the release of more remote sensing datasets and the increasing spatial and spectral resolutions, identifying fine-grained object classes, such as the detailed models of airplanes has now become possible. Fine-grained object recognition in computer vision requires the identification of the subtle differences in the local regions of an object. The current methods of fine-grained object recognition include deep convolutional neural network (DCNN)-based methods, localization/recognition-based methods, network integration based methods, and higher-order feature-coding-based methods. The part-based R-CNNs [8] are a representative approach that use a region-based convolutional neural network (R-CNN) to detect objects and their discriminative parts. Meanwhile, attention mechanisms have been applied to learn models without bounding box labels [9–11]. The network-integration-based methods learn multiple models and then integrate the predictions [12–14]. High-level feature encoding combines feature maps using outer products [15–17], with the bilinear model being the most representative. It is, however, unclear how these methods perform with remote sensing objects, but some studies have attempted to apply them. For example, Sumbul et al. [18] introduced multiple sources of data to address the low inter-class variance, small training set size for rare classes, and large class imbalance, while Aygünes et al. [19,20] used weakly supervised location-based recognition for fine-grained tree recognition.

The fine-grained object recognition task, which involves identifying the types of objects among a large number of closely related subcategories, differs from the traditional object identification tasks commonly studied in the remote sensing literature in at least two ways:

- The difficulty of accumulating a large number of similar subcategory samples can greatly limit the size of the training set for certain subcategories;
- The class imbalance can lead the traditional supervised formulations to prefer fitting to classes that occur frequently, while ignoring classes with a limited sample size.

In the data-driven models, these differences can create recognition difficulties, especially when there are limited labeled data, which is known as the few-shot learning (FSL) problem. This is a significant issue in the fine-grained object recognition task. The current FSL methods adapt models learned from auxiliary datasets with large amounts of data to new data with limited labeled data. During training, the concept of a "task" is introduced, where a small set of class-balanced samples are iteratively sampled from the auxiliary dataset, and the model learns the task distribution. New data with limited annotation can be considered a few-shot task and adapted using the task distribution learned from the auxiliary dataset.

Most of the current FSL methods can be divided into three main categories: (1) generation-based methods; (2) initialization-based methods; and (3) metric-based learning methods. The generation-based methods aim to learn a generator that can augment the data for new classes [21–23], while the initialization-based approaches [24] focus on learning a good initialization for the model parameters. The metric-based methods involve measuring the distance between a query set and a support set in the embedding space to determine the predicted categories [25,26]. For the FSL problem of remote sensing object recognition, there is not much current research. However, Fu et al. [27] did introduce an initialization-based few-shot approach to synthetic aperture radar (SAR) object recognition, while noting that the few-shot classification of remote sensing objects is more difficult than that of natural images. Some works have improved the effect of FSL on SAR data by modifying the network and metric functions [28,29]. There have also been some studies that have focused on addressing the few-shot transfer learning problem in SAR data [30,31].

However, the above FSL methods are not able to deal with remote sensing objects very effectively. The diverse sources of remote sensing data make the learned distribution inaccurate, and the complex background of remote sensing objects, the high image noise, and the small size of the objects make the features more difficult to detect with the model, which leads to a reduction of the FSL effect.

The current few-shot methods rely on a meta-learning training paradigm, i.e., learning the task distributions by learning them on sampled tasks. Factors, such as the source of the data and the way the samples are composed in the task, affect the task distribution, while the latter can be considered to be determined by the sampling strategy. Naturally, in order to make the learned task distribution closer to the real task distribution, we start from the perspective of improving the effectiveness of the sampling strategy. We therefore propose an improved sampling strategy for FSL tasks:

- We improve the sampling strategy for the fine-grained object recognition task. The task sampling strategy is improved so that each sampled task contains only subcategory samples belonging to the same broad category, e.g., fine-grained airplane samples belonging to the airplane category. This sampling strategy is used to keep each task-specific model focused on mining fine-grained features and solving the problem of the difficult extraction of remote sensing object features.
- The category synthesis method keeps the sampling task length consistent. For each task sampling operation, for the case where the number of fine-grained categories is less than the number of sampling operations, a category synthesis method is proposed to fill the gaps, to maintain the same task length for each sampling operation, and to ensure that an accurate task distribution is learned instead of a mixed distribution.

2. Background

2.1. Task Sampler

Meta-learning is a popular paradigm for few-shot learning methods. It involves sampling tasks $\mathcal{T}$ on auxiliary datasets $\mathcal{D}_{aux}$, adapting models on the tasks, and learning task distributions $p(\mathcal{T})$. The learned task knowledge $p(\mathcal{T})$ is then transferred to new tasks on few-shot datasets $\mathcal{D}_{new}$. To ensure that it is task knowledge that is transferred, the new data categories and the categories in the auxiliary dataset do not overlap. Tasks are sampled through category-balanced sampling of samples, where each task contains N categories, and each category has $K_{support} + K_{query}$ samples. Since $K_{support}$ and K_{query} are usually small values, each task $\mathcal{T}_i$ is a small-size set. The optimization objective for few-shot learning is shown in the following equation:

$$\arg\min_{\theta} \mathbb{E}_{\mathcal{T}\sim p(\mathcal{T})}\mathcal{L}(\theta, \mathcal{T}) \rightarrow \arg\min_{\theta} \sum_{i}^{n} \frac{1}{n}\mathcal{L}(\theta, \mathcal{T}_i)$$

where θ is model parameters, $\mathcal{L}$ is the loss function, and $p(\mathcal{T})$ is a task distribution.

To solve the few-shot problem using the meta-learning paradigm, the new task should have an identical distribution to the task distribution on the auxiliary data, i.e., $\mathcal{T}_{new} \sim p(\mathcal{T})$. In remote sensing, a task distribution is related to the data source region $\mathcal{R}$, the number of sampling categories N, and the number of samples per category $K_{support}$, K_{query}. Therefore, to ensure the effectiveness of the new task, at least the number of categories and the number of samples per category should be consistent with the auxiliary tasks.

2.2. Few-Shot Learning on Tasks

In FSL, the model learns by storing information about the task distribution among the meta-parameters. This stored information is called meta-information, and it can include the entire model parameters, a portion of the model parameters, or a specific hyperparameter. The optimization direction of the model is guided by the query set in the task during the training process. The overall FSL process is shown in Figure 1.

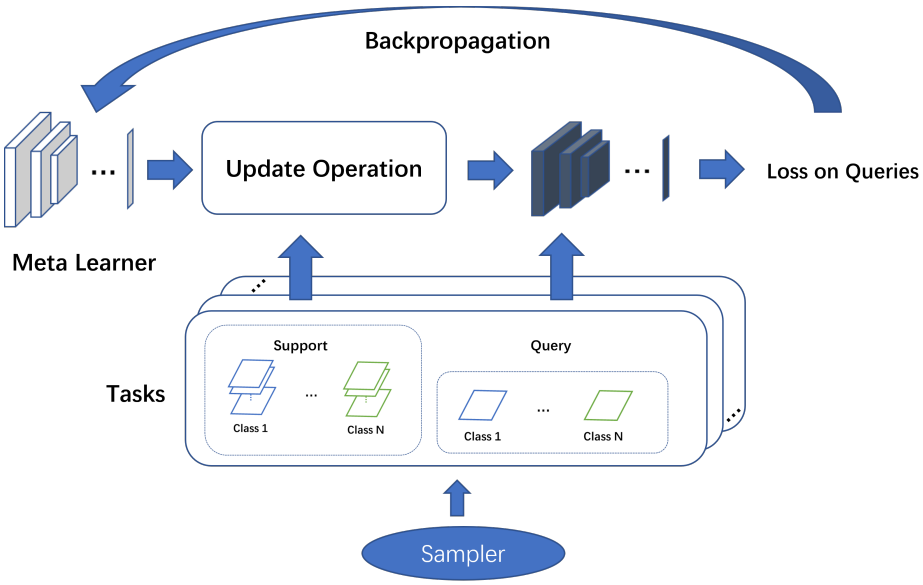

Figure 1. The main process of FSL involves adjusting the meta-parameters based on the loss on the query set using the task-specific parameters, which are fine-tuned from the meta-parameters through update operations on the support set generated by the task sampler. The update operations and loss calculation can vary depending on the specific FSL method being used.

3. Methodology

The overall system for fine-grained object recognition in remote sensing applications is illustrated in Figure 2, showing the input and output schematic. In the meta-learning step, the initialized model is transformed into a meta-model by learning on the auxiliary dataset through the tasks generated by the task sampler. The specific learning process is described in Figure 1. In the inference phase, we sample the few-shot task, fine-tune the meta-model on the support set, and finally output the prediction results for the query set.

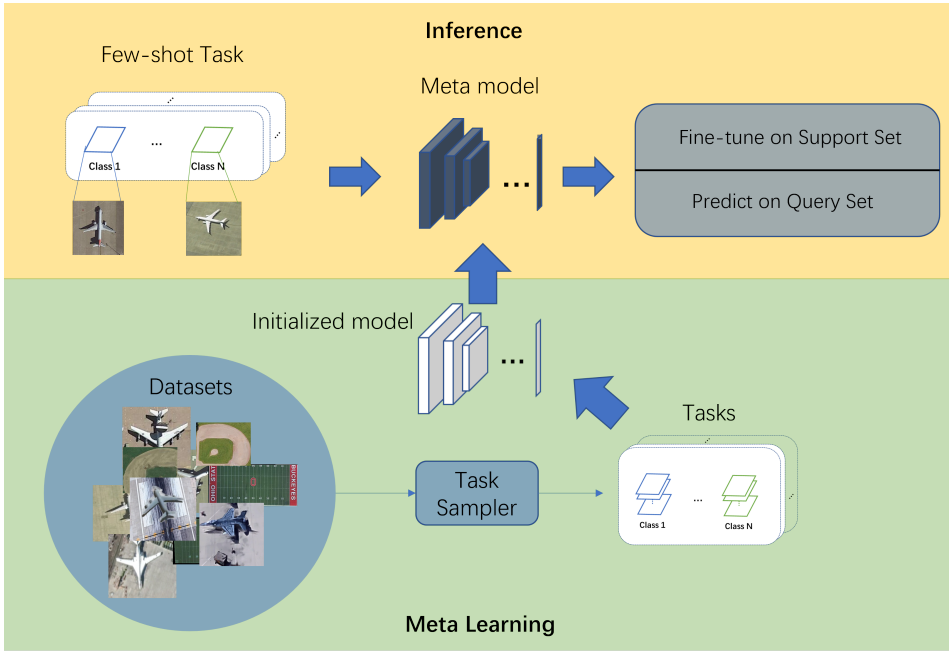

Figure 2. Schematic diagram of the input and output of the whole system. The generation of the meta-model and the subsequent output on the new data are inseparable from the task sampler.

3.1. Problem Formulation

3.1.1. Fine-Grained Problem

In Section 2.1, we discussed the task sampling paradigm for FSL training. However, this sampling strategy does not have a way to solve the problem of the difficult extraction of fine-grained remote sensing object features because the sampled tasks are not focused on mining fine-grained features, and in each task sampling operation, the sampled categories are:

$$\mathcal{C}^N = \{C_i^N | 1 \leq i \leq N, C_i^N \in \mathcal{C}_{aux}\}$$

We explicitly define each category C_i^N to belong to some broad category S_i^N and have $C_i^N \in S_i^N$. It is seen that $S_i^N = S_j^N$ does not always hold, and under extreme conditions there are even $S_i^N \cap S_j^N = \emptyset$ for any i, j, which means that the task is concerned with distinguishing differences between S_i^N, S_j^N rather than fine-grained C_i^N and C_j^N.

Further, such a sampling approach has a bias for some S_i^N, since

$$p(C_i^N \in S_i^N | \mathcal{T}) = \frac{\left\| \{\mathcal{C}_i | \mathcal{C}_i \in S_i^N\} \right\|}{\|\mathcal{C}_{aux}\|}$$

for each sampled category C_i^N from $\mathcal{C}_{aux}$, and the bias on S_i^N depends on the molecular $\left\| \{\mathcal{C}_i | \mathcal{C}_i \in S_i^N\} \right\|$.

Although the original intention of using this sampling paradigm is for the categories to be balanced in each task, each task will still have a bias toward certain broad categories, which is detrimental to the model learning of a generic large few-shot model.

3.1.2. Validity Issues of Task Distribution Learning

In Sections 2.1 and 2.2, we showed that the purpose of the model optimization is to eventually learn a task distribution. For the remote sensing recognition task, due to the diverse sample sources, the number of certain categories may not reach the required number of samples during the task sampling process, resulting in such a task being considered as coming from a different task distribution, and thus the learned distribution is a mixed distribution, which brings inaccuracy to the model learning.

3.2. Improved Task Sampler

In order to solve the problem of the difficult extraction of fine-grained features in the few samples of remote sensing imagery, we improve the task sampling strategy to make each task change from a few-shot task to a fine-grained few-shot task and let the model learn fine-grained few-shot features. Specifically, the sample categories in the unimproved task are from different broad categories, and such a task is called a recognition task. However, such a task cannot focus on fine-grained features because the two samples are from different broad categories which differ significantly, such as aircraft and ships(as discussed in Section 3.1.1).

We improve the sampling strategy by including in each task only samples sampled from a particular large class so that the model will focus on the intra-class differences within the classes in the task, i.e., the fine-grained features. Such improved tasks, which we call fine-grained tasks, end up with a fine-grained task distribution that the model learns. It is worth noting that introducing additional broad category information does not require additional annotation of the original dataset. In fact, for remote sensing data, although there are many fine-grained categories, the broad categories these fine-grained categories belong to are indeed very limited. We do not need additional annotation on the data. We only need to add appropriate judgments in the programming.

At the same time, a category generation technique is designed for the situation of this sampling strategy having an insufficient number of categories to be sampled. We use the idea of mix-up for category generation by sampling the beta distribution to obtain a factor λ, where, for any two category samples i, j considered to generate a new category, this new category sample is obtained by pixel-level fusion as $\lambda i + (1-\lambda)j$. Specifically, lambda belongs to the distribution:

$$f(\lambda) = \frac{\lambda^{-0.5}(1-\lambda)^{-0.5}}{\int_0^1 u^{-0.5}(1-u)^{-0.5}du} = Beta(\lambda; 0.5, 0.5)$$

We set the λ to a random variable instead of a static parameter value in order to avoid duplication of the resulting new category. For example, if there are only two categories and five categories are sampled, if lambda is a constant parameter, then the new categories generated are bound to be duplicated, so there is no guarantee that the length of the picked task is five. We chose the beta distribution because the distribution is symmetric about 0.5 when the two parameters of the beta distribution agree, and the range of the random variables is 0 to 1. Such a distribution is suitable for fusing pairs of samples.

3.3. Framework

The improvement methods discussed previously are integrated into a unified process. The overall flow is shown in Algorithm 1.

Algorithm 1 Fine-Grained Task Sampling Process

Input:

The auxiliary dataset $\mathcal{D}_{aux}$;
Collection of categories in the auxiliary dataset $\mathcal{C}_{aux}$;
Collection of broad categories in the auxiliary dataset $\mathcal{S}_{aux}$;
Number of categories in a task N;

Number of samples per category in a task $K_{support}$,K_{query}.
Output:
Tasks sampled from auxiliary datasets $\mathcal{T}$.
for iteration i = 1,2,3... **do**
Sample a broad category from $\mathcal{S}_{aux}$, denoted as S_i
Calculate $R_i = N \bmod \|\{C_i|C_i \in S_i\}\|$, $N_i = N \mid \|\{C_i|C_i \in S_i\}\|$
for j=1,2,3,..N_i **do**
Sample two category C_j^1, C_j^2 where $C_j^1, C_j^2 \in S_i$ and $C_j^1, C_j^2 \in \mathcal{C}_{aux}$
Sample $K_{support}$ samples belonging to C_j^1, C_j^2 from $\mathcal{D}_{aux}$, denoted as S_j^1, S_j^2
Sample K_{query} samples belonging to C_j^1, C_j^2 from $\mathcal{D}_{aux}$, denoted as Q_j^1, Q_j^2
Sample λ_j from Beta Distribution $\beta(.2, .2)$
For each image i, j in S_j^1, S_j^2 fused as $\lambda_j i + (1 - \lambda_j) j$, denoted as S_j
For each image i, j in Q_j^1, Q_j^2 fused as $\lambda_j i + (1 - \lambda_j) j$, denoted as Q_j
Define $\mathcal{T}_{ij} = (S_j, Q_j)$
end for
for $C_j \in S_i$ **do**
Sample $K_{support}$ samples belonging to C_j from $\mathcal{D}_{aux}$, denoted as S_j
Sample K_{query} samples belonging to C_j from $\mathcal{D}_{aux}$, denoted as Q_j
Define $\mathcal{T}_{ij}^r = (S_j, Q_j)$
end for
Define $\mathcal{T}_i = \{\mathcal{T}_{ij}|j \leq N\}$, $\mathcal{T}_i^r = \{\mathcal{T}_{ij}^r|C_j \in S_i\}$
end for
return $\mathcal{T} = \{\mathcal{T}_i, \mathcal{T}_i^r\}$

We consider the fusion sample $\lambda_j i + (1 - \lambda_j) j$ a brand new category sample. Since λ is a random variable, each fusion sample category is considered as different. As can be seen from the algorithm flow, the category sampled by each task belongs to the broad category S_i. Therefore, each task focuses on the small differences in the fine-grained categories of broad category S_i, forcing the model to mine fine-grained features. This sampling strategy, which further specifies the tasks as fine-grained few-shot tasks, provides an artificial prior to improve the mining ability for fine-grained features. Figure 3 briefly illustrates the difference between a task sampled by the proposed approach and a task sampled by the general policy.

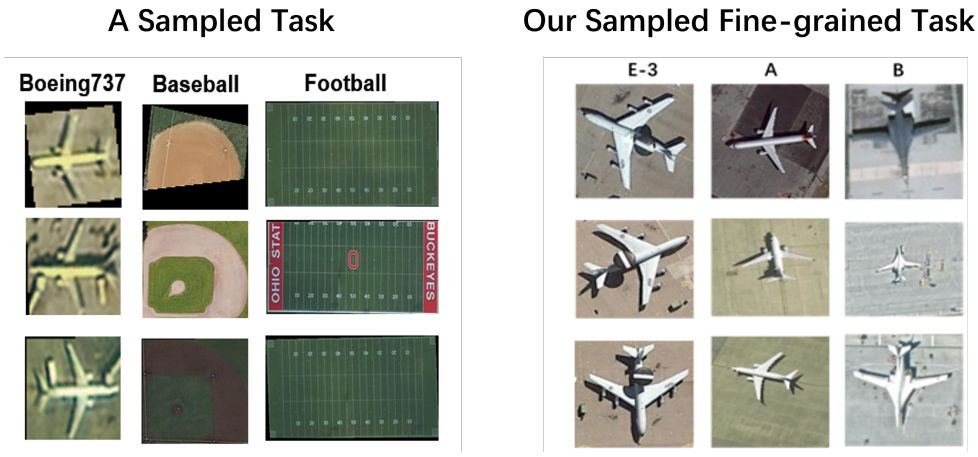

Figure 3. Difference between the task sampled by the proposed approach and the task sampled by the general policy. It can be seen that the tasks sampled by the proposed approach are all samples under the broad airplane category.

4. Datasets

4.1. FAIR1M Fine-Grained Dataset

The different models were trained and evaluated on the FAIR1M [32] dataset.There are three reasons for adopting this dataset:

- In terms of quantity, it is much larger than the other existing object datasets, and it can be divided into auxiliary datasets;
- It provides more rich fine-grained category information;
- It provides better image quality due to careful data cleaning procedures.

The FAIR1M dataset is unique in its focus on fine-grained object recognition in high-resolution remote sensing imagery. The dataset contains more than 15,000 images from over 100 locations worldwide, covering 37 fine-grained categories for object detection. The categories include ten types of airplanes, eight types of ships, nine types of vehicles, four types of courts, and three types of roads. The airplane categories consist of the most common types in civil aviation, such as *Boeing737* and *Airbus A320*. The ship categories are defined by their functions, such as *Passenger Ships* and *Engineering Ships*. The vehicle categories are also defined by their functions, such as *Small Cars* and *Excavators*. There are also categories for courts and roads. The dataset also includes categories for objects that do not belong to the specific types, labeled as *Other-Airplane*, *Other-Ship*, and *Other-Vehicle*. The distribution of instances per category reflects the authenticity and challenge of the dataset.

4.2. MTARSI Datasets

We also evaluated the models on the muti-type aircraft of remote sensing images (MTARSI) dataset [33]. There are two reasons for adopting this dataset:

- The dataset is an aircraft dataset containing 20 fine-grained aircraft categories, which is convenient for evaluating the effect of a model trained on FAIR1M;
- The objects are carefully selected from many airports around the world and are calibrated with correct and high-quality category labeling.

The MTARSI dataset is a valuable resource for researchers working on aircraft recognition in remote sensing images, as it offers a wide range of images with diverse aircraft types and environmental conditions. With careful labeling and data augmentation, the dataset provides an excellent opportunity for developing and evaluating algorithms for fine-grained object recognition in high-resolution remote sensing imagery.

4.3. Object Cropping and Subcategory Division

4.3.1. FAIR1M

Although the dataset is intended for object detection, we carefully cropped the objects according to the annotations, and some examples can be seen in Figure 4. The modified dataset has 5 categories—*Ship* , *Vehicle*, *Airplane*, *Court*, and *Road*—and 38 subcategories, which are shown in Figure 5 and Table 1 The samples of each subcategory show a long-tail distribution in Figure 5. and the cropped objects are suitable for forming a few-shot and fine-grained dataset. We formed the subcategories with a high number of samples into an auxiliary dataset for the meta-training phase, and the remaining samples were incorporated into a few-shot dataset for the testing phase. The detailed division is shown in Tables 2 and 3.

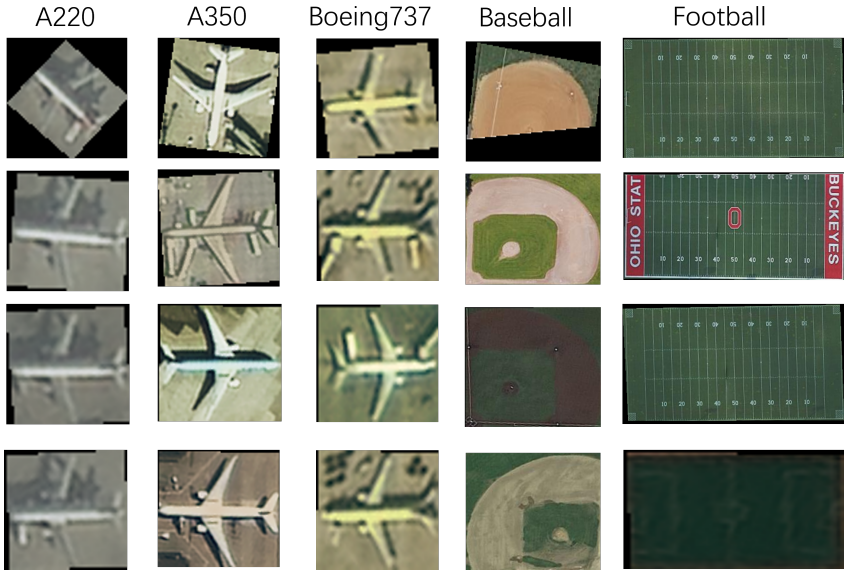

Figure 4. Cropped instances from the FAIR1M dataset.

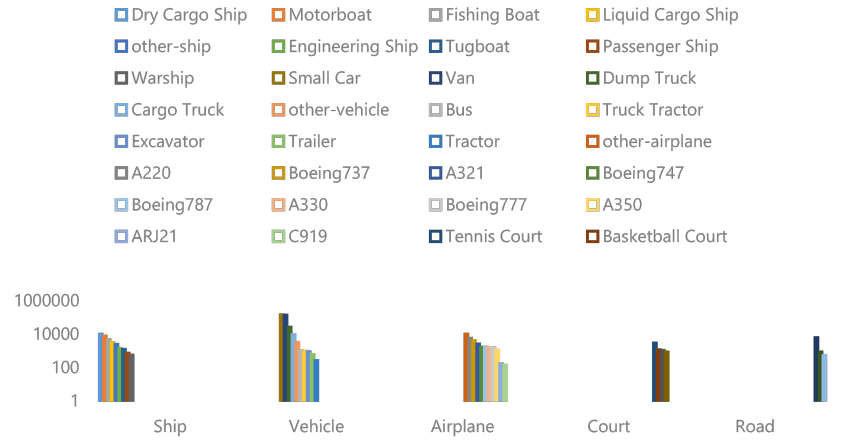

Figure 5. The distribution of the number of cropped subcategories per category.

Table 1. The 5 categories (*Ship*, *Vehicle*, *Airplane*, *Court*, and *Road*) and 38 subcategories in the FAIR1M dataset.

Category	Subcatogory
Ship	*Motorboat, Fishing Boat, Liquid Cargo Ship, Other-Ship, Dry Cargo Ship* *Passenger Ship, Tugboat, Warship, Engineering Ship*
Vehicle	*Small Car,Van* *Bus, Cargo Truck, Dump Truck, other-vehicle, Trailer, Tractor, Excavator, Truck Tractor*
Airplane	*Boeing737, A220, Other-Airplane* *Boeing747, Boeing777, Boeing787, ARJ21, C919, A321, A330, A350*
Court	*Baseball Field, Basketball Court, Tennis Court, Basketball Court* *Football Field*
Road	*Intersection* *Roundabout, Bridge*

Table 2. The subcategories in this table form the auxiliary dataset for training.

Category	Base Subcatogory
Ship	*Motorboat, Fishing Boat, Liquid Cargo Ship, other-ship, Dry Cargo Ship*
Vehicle	*Small Car,Van*
Airplane	*Boeing737, A220, Other-Airplane*
Court	*Baseball Field, Basketball Court, Tennis Court, Basketball Court*
Road	*Intersection*

Table 3. The subcategories in this table form the few-shot dataset for testing.

Category	Noval Subcatogory
Ship	*Passenger Ship, Tugboat, Warship, Engineering Ship*
Vehicle	*Bus, Cargo Truck, Dump Truck, other-vehicle, Trailer, Tractor, Excavator, Truck Tractor*
Airplane	*Boeing747, Boeing777, Boeing787, ARJ21, C919, A321, A330, A350*
Court	*Football Field*
Road	*Roundabout, Bridge*

4.3.2. MTARSI

Since the MTARSI dataset is released for object recognition, each sample is an image containing an object and a background, and no additional processing of the data is required. We apply this dataset directly in the experiments without any modification.

5. Experiments

5.1. General Overview of the Process

Broadly speaking, the experimental procedure was as follows:

- Determine base subcategories and novel subcategories (not overlapping);
- Use the FSL methods to learn from the base subcategories (#epoch: early stopping);
- Fine-tune the obtained model to the novel subcategories with the limited labeled data (only one task).

5.1.1. Input Data

The original images were of various sizes, so they were resized to a fixed size of 224 × 224 pixels for training.

5.1.2. Training

All the models were trained with the Adam optimizer in the inner gradient update and stochastic gradient descent (SGD) in the outer gradient update on a single GPU. The learning rate was set to 10^{-3} for the Adam optimizer and 1 for SGD. We used a meta-learning paradigm for the training, i.e., learning on a sampling task. Overall, the experiments were set to simulate the few sample case by setting the number of samples per class k in a task to 1, 5, and 10. The FSL methods were then applied to learn from the tasks.

5.1.3. Fine-Tuning

Only one task was sampled from the few-shot dataset, and it had the same parameters (N and $K_{support}$) as the task sampled from the auxiliary dataset. The model parameters were fine-tuned on the support set for that task, and the query set was used to evaluate the performance. We sampled 20 times and evaluated the average performance of the model on the 20 tasks. Note that to reduce the variance, we increased the number of K_{query} to 50.

5.1.4. Evaluation Metrics

In this paper, to evaluate the performance of the fine-grained visual classification (FGVC) algorithms, we use the classification accuracy as the metric for this task. $N_{correct}$ and N_{all} denote the number of correctly predicted instances and total instances, respectively. The definition of the classification accuracy is as follows:

$$Acc = \frac{N_{correct}}{N_{all}}$$

5.2. Network Structure

We used the most common network structure for the FSL methods, where the network consisted of four blocks in series. Each block contained a 3 × 3 convolutional layer with 32 channels, a batch normalization layer, a maximum pooling layer of width 2, and a rectified linear unit (ReLU) activation layer.

5.3. Comparing Methods

We used first-order model-agnostic meta-learning(FOML) [24], Reptile [34], and Prototypical network(Proto.) [25] as comparison methods. The first two of these methods are initialization-based methods, and the last one is a metric-learning-based method. The proposed method is based on the Prototypical network with the replacement of the proposed task sampling strategy. All the methods used the same network structure described in the previous section.

5.4. Experiments Results

5.4.1. Experiments on FAIR1M

We performed FSL on the base category divisions in Table 2 and tested on the novel categories in Table 3 to obtain the following results, which are listed in Table 4.

Table 4. Only one task was sampled from the new classes. The model was then tuned on the support set and the precision was tested on the query set. The sampling was conducted 50 times, and the average performance accuracy on these tasks is reported.

ALL	5 Way 1 Shot	5 Way 5 Shot	5 Way 10 Shot
FOML	34.11% ± 9.32%	51.96% ± 10.11%	62.74% ± 10.08%
Reptile	37.24% ± 7.78%	54.32% ± 8.66%	63.36% ± 10.07%
Proto.	37.81% ± 7.67%	59.33% ± 9.45%	65.34% ± 9.13%
Ours	**38.30% ± 8.05%**	**61.41% ± 8.32%**	**66.73% ± 8.98%**

It can be seen from Table 4 that even if the base and novel categories do not overlap, a model trained on the base categories and fine-tuned on a small amount of labeled data from the novel categories can accurately predict the remaining data, demonstrating the effectiveness of the few-shot methods. Consistent with the academic research findings, the current FSL methods based on metric learning have a higher accuracy. Furthermore, the accuracy of each method improves as the number of shots increases. It is quite clear that as the number of samples increases, the distribution learned by the model becomes more accurate. The proposed method, which uses an improved fine-grained task sampling strategy, achieves a higher accuracy and larger performance improvement. Notably, the proposed task sampling strategy was consistent with the other methods during testing, suggesting that learning on the base categories with this strategy can lead to better few-shot results.

Furthermore, in the novel categories, we singled out broad categories for the model tuning and testing, and we picked out the *Aircraft*, *Ship*, and *Vehicle* categories. The results of these experiments are listed in Tables 5–7. As can be seen, compared to the experimental results in Table 4 obtained on all the categories, the proposed approach shows an

improvement, but not by much. Our analysis suggests that when a single broad category is picked for the testing, it is a simple case, such as the large category of aircraft, where the FSL model has a strong generalization ability due to the absence of interference from samples belonging to other broad categories, and the model is fine-tuned to have a higher accuracy, narrowing the gap with the proposed method. However, in most cases, the proposed method still obtains a higher accuracy, which is due to the better fine-grained task distribution learned on the auxiliary dataset. It can therefore be concluded that the key to the effectiveness of FSL lies in the efficiency of learning on the auxiliary dataset.

Table 5. The broad category of ship was picked from the base categories for the experiment, and the broad category of ship from the novel categories was used for the testing.

Ship	5 Way 1 Shot	5 Way 5 Shot	5 Way 10 Shot
FOML	40.34% ± 8.70%	52.26% ± 5.61%	58.49% ± 4.68%
Reptile	41.49% ± 5.98%	55.18% ± 4.96%	58.62% ± 3.45%
Proto.	43.64% ± 5.63%	56.63% ± 4.43%	60.28% ± 3.72%
Ours	**43.98% ± 6.54%**	**56.82% ± 4.18%**	**61.12% ± 3.70%**

Table 6. The broad category of vehicle was picked from the base categories for the experiment, and the broad category of vehicle from the novel categories was used for the testing.

Vehicle	5 Way 1 Shot	5 Way 5 Shot	5 Way 10 Shot
FOML	27.34% ± 5.52%	36.87% ± 5.44%	41.25% ± 5.07%
Reptile	26.45% ± 4.80%	35.54% ± 4.92%	42.38% ± 5.19%
Proto.	29.78% ± 6.06%	42.55% ± 5.62%	47.22% ± 4.99%
Ours	**30.02% ± 6.48%**	41.34% ± 5.57%	46.38% ± 5.29%

Table 7. The broad category of airplane was picked from the base category for the experiment, and the broad category of airplane from the novel categories was used for the testing.

Airplane	5 Way 1 Shot	5 Way 5 Shot	5 Way 10 Shot
FOML	22.33% ± 2.30%	35.68% ± 6.20%	45.39% ± 8.38%
Reptile	27.29% ± 3.83%	37.90% ± 5.21%	46.36% ± 8.23%
Proto.	27.75% ± 5.07%	40.68% ± 6.79%	47.02% ± 8.17%
Ours	**28.16% ± 5.23%**	**41.64% ± 6.53%**	**47.67% ± 7.55%**

5.4.2. Experiments on MTARSI

FSL focuses on learning the task distribution instead of the sample distribution, so it does not require the same classification system for the auxiliary and new datasets. To analyze FSL on a remote sensing dataset, we applied the model trained on the FAIR1M dataset to the MTARSI dataset. All the categories of the MTARSI dataset were used for the testing, with one task sampled for the model tuning and 50 tasks sampled for testing on the remaining data. The results are listed in Table 8.

As can be seen from the results, the effect of the aircraft recognition performance on the FAIR1M dataset (Table 7) is similar to the performance on the MTARSI dataset. Furthermore, the accuracy is about the same, implying that the final performance effect difference does not come from the sample difference between the sampled MTARSI and FAIR1M datasets but depends on the learned task distribution on the auxiliary dataset. The good news is that this implies that the model trained on the auxiliary dataset is highly generalizable, giving the application a high migration adaptation capability.

Table 8. Only one task was sampled from the new classes. The model was then tuned on the support set, and the precision was tested on the query set. The sampling was conducted 50 times, and the average performance accuracy on these tasks is reported.

ALL(Airplane)	5 Way 1 Shot	5 Way 5 Shot	5 Way 10 Shot
FOML	25.30% ± 4.88%	37.22% ± 5.63%	46.06% ± 5.90%
Reptile	27.54% ± 4.25%	40.86% ± 6.35%	48.23% ± 6.36%
Proto.	28.97% ± 5.10%	42.34% ± 6.16%	46.56% ± 6.70%
Ours	**29.06% ± 5.27%**	**44.95% ± 6.77%**	**48.69% ± 6.70%**

6. Conclusions

The remote sensing fine-grained object recognition task has a few-shot problem that is difficult to get around, i.e., the lack of fine-grained object data. Although some few-shot methods have been proposed to address learning from a small number of data samples, these methods have not been adapted to remote sensing objects, where the object context is complex and fine-grained features are difficult to extract. In this study, we first investigated the FSL methods and found that only a small amount of work has been conducted in the field of remote sensing. This inspired us to design a few-shot approach for remote sensing tasks. In this paper, we reviewed the general paradigm of training FSL methods on sampled tasks. We then affirmed the paradigm's usefulness for FSL, but pointed out its shortcomings for remote sensing data. The problem is that the task has a preference for certain broad categories, the task may not be fine-grained, and the model has difficulty extracting the already hard-to-extract fine-grained features of remote sensing imagery, which can adversely affect the model learning. To solve this problem, we proposed an improved task sampling strategy: the fine-grained task sampling strategy. The fine-grained task sampling strategy ensures that the samples in each sampled task come from a certain broad category, which ensures that the model focuses on the small differences in the fine-grained features on that task. This reinforces the model to learn fine-grained features. We tested the effectiveness of this sampling strategy on the FAIR1M and MTARSI datasets. We found that the FSL method can achieve a large accuracy improvement using the improved sampling strategy. In addition, from the performance effect on the two datasets, we pointed out that the performance of FSL depends on the learning effect on the auxiliary datasets. In our future work, we will attempt to further improve the existing network structure to make it more suitable for remote sensing tasks. We will also bring the few-shot approach to other tasks in remote sensing, such as remote sensing object detection.

Author Contributions: S.L.: methodology, software, validation, formal analysis, investigation, data curation, writing—original draft, visualization; A.M.: conceptualization, writing—review and editing, project administration, supervision; S.P.: supervision; Y.Z.: supervision, funding acquisition. All authors have read and agreed to the published version of the manuscript

Funding: This research received no external funding.

Data Availability Statement: All the data come from public datasets, and the data used in the paper can be downloaded from https://www.gaofen-challenge.com/benchmark accessed on 22 January 2022 and https://zenodo.org/record/3464319#.Y9pg6nBBxEY accessed on 25 January 2022.

Acknowledgments: The authors would like to thank the Editor, Associate Editor, and anonymous reviewers for their helpful comments and suggestions that improved this article. This work was supported in part by the National Key Research and Development Program of China under Grant No. 2022YFB3903404, in part by the National Natural Science Foundation of China under Grant No. 42171336

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

References

1. Zhang, L.; Zhang, L.; Du, B.J.I.G.; Magazine, R.S. Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geosci. Remote. Sens. Mag.* **2016**, *4*, 22–40.
2. Fatima, S.A.; Kumar, A.; Pratap, A.; Raoof, S.S. Object recognition and detection in remote sensing images: A comparative study. In Proceedings of the 2020 International Conference on Artificial Intelligence and Signal Processing (AISP), Amaravati, India, 10–12 January 2020; pp. 1–5.
3. Jiang, B.; Li, X.; Yin, L.; Yue, W.; Wang, S. Object recognition in remote sensing images using combined deep features. In Proceedings of the 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Chengdu, China, 15–17 March 2019; pp. 606–610.
4. Yang, C.; Dong, Y.; Du, B.; Zhang, L. Attention-Based Dynamic Alignment and Dynamic Distribution Adaptation for Remote Sensing Cross-Domain Scene Classification. *IEEE Trans. Geosci. Remot. Sens.* **2022**, *60*, 1–13.
5. Chen, W.; Ouyang, S.; Yang, J.; Li, X.; Zhou, G.; Wang, L. JAGAN: A framework for complex land cover classification using Gaofen-5 AHSI images. *IEEE J. Sel. Top. Appl. Earth Obs. Remot. Sens.* **2022**, *15*, 1591–1603. [CrossRef]
6. Chen, W.; Ouyang, S.; Tong, W.; Li, X.; Zheng, X.; Wang, L. GCSANet: A global context spatial attention deep learning network for remote sensing scene classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remot. Sens.* **2022**, *15*, 1150–1162. [CrossRef]
7. Liu, Q.; Dong, Y.; Zhang, Y.; Luo, H. A Fast Dynamic Graph Convolutional Network and CNN Parallel Network for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remot. Sens.* **2022**, *60*, 1–15. [CrossRef]
8. Zhang, N.; Donahue, J.; Girshick, R.; Darrell, T. Part-based R-CNNs for fine-grained category detection. In *Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 834–849.
9. Sermanet, P.; Frome, A.; Real, E. Attention for fine-grained categorization. *arXiv* **2014**. [CrossRef]
10. Sun, M.; Yuan, Y.; Zhou, F.; Ding, E. Multi-attention multi-class constraint for fine-grained image recognition. In Proceedings of the European Conference on Computer Vision (ECCV), Glasgow, UK, 23–28 August 2020; pp. 805–821.
11. Oliveau, Q.; Sahbi, H. Learning attribute representations for remote sensing ship category classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remot. Sens.* **2017**, *10*, 2830–2840. [CrossRef]
12. Ge, Z.; Bewley, A.; McCool, C.; Corke, P.; Upcroft, B.; Sanderson, C. Fine-grained classification via mixture of deep convolutional neural networks. In Proceedings of the 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Placid, NY, USA, 7–10 March 2016; pp. 1–6.
13. Ge, Z.; McCool, C.; Sanderson, C.; Corke, P. Subset feature learning for fine-grained category classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Virtual, 19–25 June 2021; pp. 46–52.
14. Wang, D.; Shen, Z.; Shao, J.; Zhang, W.; Xue, X.; Zhang, Z. Multiple granularity descriptors for fine-grained categorization. In Proceedings of the IEEE International Conference on Computer Vision, Cambridge, MA, USA, 20–23 June 1995; pp. 2399–2406.
15. Zhang, X.; Zhou, F.; Lin, Y.; Zhang, S. Embedding label structures for fine-grained feature representation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 1114–1123.
16. Kong, S.; Fowlkes, C. Low-rank bilinear pooling for fine-grained classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 365–374.
17. Yu, C.; Zhao, X.; Zheng, Q.; Zhang, P.; You, X. Hierarchical bilinear pooling for fine-grained visual recognition. In Proceedings of the European Conference on Computer Vision (ECCV), Glasgow, UK, 23–28 August 2020; pp. 574–589.
18. Sumbul, G.; Cinbis, R.G.; Aksoy, S.; Sensing, R. Multisource region attention network for fine-grained object recognition in remote sensing imagery. *IEEE Trans. Geosci. Remot. Sens.* **2019**, *57*, 4929–4937. [CrossRef]
19. Aygüneş, B.; Aksoy, S.; Cinbiş, R.G. Weakly supervised deep convolutional networks for fine-grained object recognition in multispectral images. In Proceedings of the IGARSS 2019–2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 1478–1481.
20. Aygunes, B.; Cinbis, R.G.; Aksoy, S.J.I.J.O.P.; Sensing, R. Weakly supervised instance attention for multisource fine-grained object recognition with an application to tree species classification. *Isprs J. Photogramm. Remot. Sens.* **2021**, *176*, 262–274. [CrossRef]
21. Schwartz, E.; Karlinsky, L.; Shtok, J.; Harary, S.; Marder, M.; Feris, R.; Kumar, A.; Giryes, R.; Bronstein, A.M. Delta-encoder: An effective sample synthesis method for few-shot object recognition. *arXiv* **2018**. . [CrossRef]
22. Gao, H.; Shou, Z.; Zareian, A.; Zhang, H.; Chang, S.F. Low-shot learning via covariance-preserving adversarial augmentation networks. In Proceedings of the NIPS'18: Proceedings of the 32nd International Conference on Neural Information Processing Systems, Red Hook, NY, USA, 3–8 December 2018.
23. Pfister, T.; Charles, J.; Zisserman, A. Domain-adaptive discriminative one-shot learning of gestures. In *Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 814–829.
24. Finn, C.; Abbeel, P.; Levine, S. Model-agnostic meta-learning for fast adaptation of deep networks. In Proceedings of the International Conference on Machine Learning, PMLR, Sydney, Australia, 6–11 August 2017; pp. 1126–1135.
25. Snell, J.; Swersky, K.; Zemel, R. Prototypical networks for few-shot learning. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017.
26. Vinyals, O.; Blundell, C.; Lillicrap, T.; Wierstra, D. Matching networks for one shot learning. In Proceedings of the 30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain, 5–10 December 2016.

27. Fu, K.; Zhang, T.; Zhang, Y.; Wang, Z.; Sun, X.; Sensing, R. Few-shot SAR target classification via metalearning. *IEEE Trans. Geosci. Remot. Sens.* **2021**, *60*, 1–14. [CrossRef]
28. Tang, J.; Zhang, F.; Zhou, Y.; Yin, Q.; Hu, W. A fast inference networks for SAR target few-shot learning based on improved siamese networks. In Proceedings of the IGARSS 2019–2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 1212–1215.
29. Gao, F.; Xu, J.; Lang, R.; Wang, J.; Hussain, A.; Zhou, H. A Few-Shot Learning Method for SAR Images Based on Weighted Distance and Feature Fusion. *Remot. Sens.* **2022**, *14*, 4583. [CrossRef]
30. Tai, Y.; Tan, Y.; Xiong, S.; Sun, Z.; Tian, J. Few-shot transfer learning for sar image classification without extra sar samples. *IEEE J. Sel. Top. Appl. Earth Obs. Remot. Sens.* **2022**, *15*, 2240–2253. [CrossRef]
31. Rostami, M.; Kolouri, S.; Eaton, E.; Kim, K. Deep transfer learning for few-shot SAR image classification. *Remot. Sens.* **2019**, *11*, 1374. [CrossRef]
32. Sun, X.; Wang, P.; Yan, Z.; Xu, F.; Wang, R.; Diao, W.; Chen, J.; Li, J.; Feng, Y.; Xu, T.; et al. FAIR1M: A benchmark dataset for fine-grained object recognition in high-resolution remote sensing imagery. *Isprs J. Photogramm. Remot. Sens.* **2022**, *184*, 116–130. [CrossRef]
33. Wu, Z.Z.; Wan, S.H.; Wang, X.F.; Tan, M.; Zou, L.; Li, X.L.; Chen, Y.J.A.S.C. A benchmark data set for aircraft type recognition from remote sensing images. *Appl. Soft Comput.* **2020**, *89*, 106132. [CrossRef]
34. Nichol, A.; Schulman, J. Reptile: A scalable metalearning algorithm. *OpenAI* **2018**, *2*, 4.