Easy-Ensemble Augmented-Shot-Y-Shaped Learning: State-of-The-Art Few-Shot Classification with Simple Components

Yassir Bendou, Yuqing Hu, Raphael Lafargue, Giulia Lioi, Bastien Pasdeloup, Stéphane Pateux, Vincent Gripon

▶ **To cite this version:**

Yassir Bendou, Yuqing Hu, Raphael Lafargue, Giulia Lioi, Bastien Pasdeloup, et al.. Easy-Ensemble Augmented-Shot-Y-Shaped Learning: State-of-The-Art Few-Shot Classification with Simple Components. Journal of Imaging, 2022, 8 (7), pp.179. 10.3390/jimaging8070179 . hal-03714237

HAL Id: hal-03714237
https://hal.science/hal-03714237

Submitted on 8 Jul 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Journal of *Imaging*

MDPI

Article

Easy—Ensemble Augmented-Shot-Y-Shaped Learning: State-of-the-Art Few-Shot Classification with Simple Components

Yassir Bendou [1,*], Yuqing Hu [1,2], Raphael Lafargue [1], Giulia Lioi [1], Bastien Pasdeloup [1], Stéphane Pateux [2] and Vincent Gripon [1,*]

[1] IMT Atlantique, Technopole Brest Iroise, 29238 Brest, France; yuqing.hu@imt-atlantique.fr (Y.H.); raphael.lafargue@imt-atlantique.fr (R.L.); giulia.lioi@imt-atlantique.fr (G.L.); bastien.pasdeloup@imt-atlantique.fr (B.P.)
[2] Orange Labs, 35510 Rennes, France; stephane.pateux@orange.com
* Correspondence: yassir.bendou@imt-atlantique.fr (Y.B.); vincent.gripon@imt-atlantique.fr (V.G.)

Abstract: Few-shot classification aims at leveraging knowledge learned in a deep learning model, in order to obtain good classification performance on new problems, where only a few labeled samples per class are available. Recent years have seen a fair number of works in the field, each one introducing their own methodology. A frequent problem, though, is the use of suboptimally trained models as a first building block, leading to doubts about whether proposed approaches bring gains if applied to more sophisticated pretrained models. In this work, we propose a simple way to train such models, with the aim of reaching top performance on multiple standardized benchmarks in the field. This methodology offers a new baseline on which to propose (and fairly compare) new techniques or adapt existing ones.

Keywords: few-shot learning; classification; deep learning; augmentations; self-supervision; ensembling; backbones; cropping; ambiguity

check for updates

Citation: Bendou, Y.; Hu, Y.; Lafargue, R.; Lioi, G.; Pasdeloup, B.; Pateux, S.; Gripon, V. Easy—Ensemble Augmented-Shot-Y-Shaped Learning: State-of-the-Art Few-Shot Classification with Simple Components. *J. Imaging* **2022**, *8*, 179. https://doi.org/10.3390/jimaging8070179

Academic Editors: Giuseppe Amato and Yudong Zhang

Received: 29 April 2022
Accepted: 15 June 2022
Published: 24 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Learning with few examples, or few-shot learning, is a domain of research that has become increasingly popular in the past few years. Reconciling the remarkable performances of deep learning (DL), which are generally obtained thanks to access to huge databases, with the constraint of having a very small number of examples may seem paradoxical. Yet, the answer lies in the ability of DL to transfer knowledge acquired when solving a previous task toward a different, new one.

The classical few-shot setting consists of two parts:

- A base dataset, which contains many examples of many classes. Since this dataset is large enough, it can be used to efficiently train DL architectures. Authors often use the base dataset alongside a validation dataset. As is usual in classification, the base dataset is used during training, and the validation dataset is then used as a proxy to measure generalization performance on unseen data and, therefore, can be leveraged to optimize the hyperparameters. However, contrary to common classification settings, in few-shot, the validation and base datasets usually contain distinct classes, so that the generalization performance is assessed on new classes [1]. Learning good feature representations from the base dataset can be performed with multiple strategies, as will be further discussed in Section 2;
- A novel dataset, which consists of classes that are distinct from those of the base and validation datasets. We are only given a few labeled examples for each class, resulting in a few-shot problem. The labeled samples are often called the support set and the remaining ones the query set. When benchmarking, it is common to use a large novel

dataset from which artificial few-shot tasks are sampled uniformly randomly, what we call a run. In that case, the number of classes n (named ways), the number of shots per class k, and the number of query samples per class q are given by the benchmark. This setting is referred to as n-way-k-shot learning. Reported performances are often averaged over a large number of runs.

In order to exploit knowledge previously learned by models on the base dataset, a common approach is to remove their final classification layer. The resulting models, now seen as feature extractors, are generally termed backbones and can be used to transform the support and query datasets into feature vectors. This is a form of transfer learning. In this work, we do not consider the use of additional data such as other datasets [2], nor semantic information [3]. Additional preprocessing steps may also be used on the samples and/or on the associated feature vectors, before the classification task. Another major approach uses meta-learning [4–9], as mentioned in Section 2.

It is important to distinguish two types of problems:

- In inductive few-shot classification, only the support dataset is available to the few-shot classifier, and prediction is performed on each sample of the query dataset independently of each other [9];
- In transductive few-shot classification, the few-shot classifier has access to both the support and the full query datasets when performing predictions [10].

Both problems have connections with real-world situations. In general, inductive few-shot corresponds to cases where data acquisition is expensive. This is the case for FMRI data, for example, where it is difficult to generalize from one patient to another and collect hours of training data on a patient could be harmful [11]. Alternatively, transductive few-shot corresponds to cases where data labeling is expensive. Such a situation can occur when experts must properly label data, but the data themselves are obtained cheaply, for instance in numerous medical applications [12,13].

In recent years, many contributions have introduced methodologies to cope with few-shot problems. There are many building blocks involved, including distillation [14], contrastive learning [15], episodic training [16], mixup [17], manifold mixup [1,18], and self-supervision [1]. As a consequence, it can appear quite opaque what the effective components are and whether their performance can be reproduced across different datasets or settings. Moreover, we noticed that many of these contributions report baseline performances that can be outperformed with a simpler training pipeline.

In this paper, we are interested in proposing a very simple method combining components commonly found in the literature and yet achieving competitive performance. We believe that this contribution will help have a clearer view on how to efficiently implement few-shot classification for real-world applications. Our main motivation is to define a new baseline with good hyperparameters and training routines to compare to and to start with, on which obtaining a performance boost will be much more challenging than starting from a poorly trained backbone. We also aim at showing that a simple approach reaches higher performance than increasingly complex methods proposed in the recent few-shot literature.

More precisely, in this paper:

- We introduce a very simple methodology, illustrated in Figure 1, for both inductive and transductive few-shot classification.
- We show the ability of the proposed methodology to reach or even beat state-of-the-art [9,19] performance on multiple standardized benchmarks of the field.
- All our models, obtained feature vectors, and training procedures are freely available online on our github: https://github.com/ybendou/easy accessed on 14 June 2022;
- We also propose a simple demonstration of our method using live video streaming to perform few-shot classification. The code is available at https://github.com/RafLaf/webcam accessed on 14 June 2022.

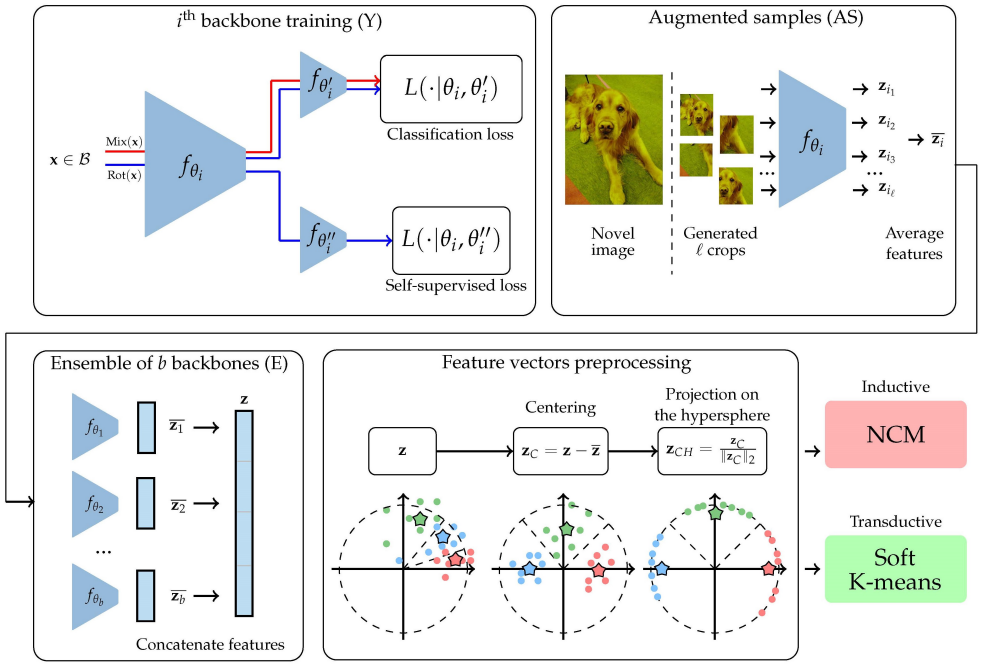

Figure 1. Illustration of our proposed method. Y: We first train multiple backbones using the base and validation datasets. We use two cross-entropy losses in parallel: one for the classification of base classes and the other for the self-supervised targets (rotations). We also use manifold mixup [18]. All the backbones are trained using the exact same routine, except that their initialization is different (random) and the order in which data batches are presented is also potentially different. AS: Then, for each image in the novel dataset and each backbone, we generate multiple crops, then compute their feature vectors, which we average. E: Each image becomes represented as the concatenation of the outputs of AS for each of the trained backbones. Preprocessing: We add a few classical preprocessing steps, including centering by removing the mean of the feature vectors of the base dataset in the inductive case, or the few-shot run feature vectors for the transductive case, and projecting on the hypersphere. Finally, we use a simple nearest class mean classifier (NCM) if in the inductive setting or a soft K-means algorithm in the transductive setting.

2. Related Work

There have been many approaches proposed recently in the field of few-shot classification. We introduce some of them following the classical pipeline. Note that our proposed methodology uses multiple building blocks from those presented hereafter.

2.1. Data Augmentation

First, data augmentation or augmented sampling are generally used on the base dataset to artificially produce additional samples, for example using rotations [1], crops [20], jitter, GANs [21,22], or other techniques [23]. Data augmentation on support and query sets, however, is less frequent. Approaches exploring this direction include [15], where the authors propose to select the foreground objects of images by identifying the right crops using a relatively complex mechanism, and [24], where the authors propose to mimic the neighboring base classes' distribution to create augmented latent space vectors.

In addition, mixup [17] and manifold mixup [18] are also used to address the challenging lack of data. Both can be seen as regularization methods through linear interpolations of samples and labels. Mixup creates linear interpolations at the sample level, while manifold mixup focuses on feature vectors.

2.2. Backbone Training

Mixup is often used in conjunction with self-supervision (S2) [1] to make backbones more robust. Most of the time, S2 is implemented as an auxiliary loss meant to train the backbone to recognize which transformation was applied to an image [25].

A well-known training strategy is episodic training. The idea behind it boils down to having the same training and test conditions. Thus, the backbone training strategy, often based on gradient descent, does not select random batches, but uses batches designed as few-shot problems [4,16,26,27].

Meta-Learning, or learning to learn, is a major line of research in the field. This method typically learns a good initialization or a good optimizer such that new classes can be learned in a few gradient steps [4–9]. In this regard, episodic training is often used, and recent work leveraged this concept to generate augmented tasks in the training of the backbone [28].

Contrastive learning aims to train a model to learn to maximize similarities between transformed instances of the same image and minimize agreement between transformed instances of different images [15,29–31]. Supervised contrastive learning is a variant that has been recently used in few-shot classification, where similarity is maximized between instances of a class instead of the same image [14,32].

2.3. Exploiting Multiple Backbones

Distillation has been recently used in the few-shot literature. The idea is to transfer knowledge from a teacher model to a student model by forcing the latter to match the joint probability distribution of the teacher [14,33].

Ensembling consists of the concatenation of features extracted by different backbones. It was used to improve performances in few-shot classification [28]. It can be seen as a more straightforward alternative to distillation. To limit the computationally expensive training of multiple backbones, some authors propose the use of snapshots [34].

2.4. Few-Shot Classification

Over the past few years, classification methods in the inductive setting have mostly relied on simple methods such as nearest class mean [35], cosine classifiers [36], and logistic regression [24].

More diverse methods can be implemented in the transductive setting. Clustering algorithms [15], embedding propagation [37], and optimal transport [38] were leveraged successfully to outrun performances in the inductive setting by a large margin.

3. Methodology

The proposed methodology consists of 5 steps, described hereafter and illustrated in Figure 1. In the experiments, we also report ablation results when omitting the optional steps.

3.1. Backbone Training (Y)

We used data augmentation with random resized crops, random color jitters, and random horizontal flips, which is standard in the field.

We used a cosine-annealing scheduler [39], where at each step, the learning rate is updated. During a cosine cycle, the learning rate evolves between η_0 and 0. At the end of the cycle, we warm restart the learning procedure and start over with a diminished η_0. We start with $\eta_0 = 0.1$ and reduce η_0 by 10% at each cycle. We use 5 cycles with 100 epochs each.

We trained our backbones using the methodology called S2M2R described in [1]. Basically, the principle is to take a standard classification architecture (e.g., ResNet12 [40]) and branch a new logistic regression classifier after the penultimate layer, in addition to the one used to identify the classes of input samples, thus forming a Y-shaped model (cf. Figure 1). This new classifier is meant to retrieve which one of four possible rotations

(quarters of 360° turns) has been applied to the input samples. We used a two-step forward–backward pass at each step, where a first batch of inputs is only fed to the first classifier, combined with manifold mixup [1,18]. A second batch of inputs is then has arbitrary rotations applied, and this is fed to both classifiers. After this training, the backbones are frozen.

We experimented using a standard ResNet12 as described in [40], where the feature vectors are of dimension 640. These feature vectors are obtained by computing a global average pooling over the output of the last convolution layer. Such a backbone contains $\sim$ 12 million trainable parameters. We also experimented with reduced-size ResNet12, denoted ResNet12$\left(\frac{1}{2}\right)$, where we divided each number of feature maps by 2, resulting in feature vectors of dimension 320, and ResNet12$\left(\frac{1}{\sqrt{2}}\right)$, where the number of feature maps are divided roughly by $\sqrt{2}$, resulting in feature vectors of dimension 450. The numbers of parameters are respectively $\sim$ 3 million and $\sim$ 6 million.

Using common notations of the field, if we denote $\mathbf{x}$ as an input sample and f as the mathematical function of the backbone, then $z = f(\mathbf{x})$ denotes the feature vector associated with $\mathbf{x}$.

From this point on, we used the frozen backbones to extract feature vectors from the base, validation, and novel datasets.

3.2. Augmented Samples

We propose to generate augmented feature vectors for each sample from the novel dataset. We did not perform this in the validation set as it is very computationally expensive. To this end, we used random resized crops from the corresponding images. We obtained multiple versions of each feature vector and averaged them. The literature has extensively studied the role of augmentations in deep learning [41]. Here, we assumed most crops would contain the object of interest. Therefore, the average feature vector can be used. On the other hand, color jitter might be an invalid augmentation since some classes rely extensively on their colors to be detected (e.g., birds or fruits).

In practice, we used $\ell = 30$ crops per image, as larger values do not benefit accuracy much. This step is optional.

3.3. Ensemble of Backbones

To boost performance even further, we propose to concatenate the feature vectors obtained from multiple backbones trained using the previously described method, but with different random seeds. To perform fair comparisons, when comparing a backbone with an ensemble of b backbones, we reduced the number of parameters per backbone such that the total number of parameters remains identical. We believe that this strategy is an alternative to performing distillation, with the interest of not requiring extra parameters and being a relatively straightforward approach. Again, this step is optional, and we perform ablation tests in the next section.

3.4. Feature Vector Preprocessing

Finally, we applied two transformations as in [35] on feature vectors $\mathbf{z}$. Denote $\overline{\mathbf{z}}$ the average feature vector of the base dataset if in the inductive setting or of the few-shot problem if in transductive setting. The ideal $\overline{\mathbf{z}}$ would center the vectors of the few-shot runs around 0 and, therefore, would be the average vector of the combined support and query set. The number of samples being too small to compute a meaningful average vector in the inductive setting, we made use of the base dataset. In the transductive setting, queries are added to the support set for mean computation. The average vector is therefore less noisy and can be used to compute $\overline{\mathbf{z}}$. The first operation (*C*—centering of $\mathbf{z}$) consists of computing:

$$\mathbf{z}_C = \mathbf{z} - \overline{\mathbf{z}}\,. \tag{1}$$

The second operation (H—projection of $\mathbf{z}_C$ on the hypersphere) is then:

$$\mathbf{z}_{CH} = \frac{\mathbf{z}_C}{\|\mathbf{z}_C\|_2} \,. \tag{2}$$

3.5. Classification

Let us denote $\mathcal{S}_i$ ($i \in \{1, \ldots, n\}$) the set of feature vectors (preprocessed as $\mathbf{z}_{CH}$) corresponding to the support set for the i-th considered class and $\mathcal{Q}$ the set of (also preprocessed) query feature vectors.

In the case of inductive few-shot classification, we used a simple nearest class mean classifier (NCM). Predictions are obtained by first computing class barycenters from labeled samples:

$$\forall i : \overline{\mathbf{c}_i} = \frac{1}{|\mathcal{S}_i|} \sum_{\mathbf{z} \in \mathcal{S}_i} \mathbf{z} \,, \tag{3}$$

then associating with each query the closest barycenter:

$$\forall \mathbf{z} \in \mathcal{Q} : C_{ind}(\mathbf{z}, [\overline{\mathbf{c}_1}, \ldots, \overline{\mathbf{c}_n}]) = \arg\min_i \|\mathbf{z} - \overline{\mathbf{c}_i}\|_2 \,. \tag{4}$$

In the case of transductive learning, we used a soft K-means algorithm. We computed the following sequence indexed by t, where the initial $\overline{\mathbf{c}_i}$ are computed as in Equation (3):

$$\forall i, t : \begin{cases} \overline{\mathbf{c}_i}^0 & = & \overline{\mathbf{c}_i} \,, \\ \overline{\mathbf{c}_i}^{t+1} & = & \sum\limits_{\mathbf{z} \in \mathcal{S}_i \cup \mathcal{Q}} \frac{w(\mathbf{z}, \overline{\mathbf{c}_i}^t)}{\sum\limits_{\mathbf{z}' \in \mathcal{S}_i \cup \mathcal{Q}} w(\mathbf{z}', \overline{\mathbf{c}_i}^t)} \mathbf{z} \,, \end{cases} \tag{5}$$

where $w(\mathbf{z}, \overline{\mathbf{c}_i}^t)$ is a weighting function on $\mathbf{z}$, which gives it a probability of being associated with barycenter $\overline{\mathbf{c}_i}^t$:

$$w(\mathbf{z}, \overline{\mathbf{c}_i}^t) = \begin{cases} \frac{\exp\left(-\beta \left\|\mathbf{z} - \overline{\mathbf{c}_i}^t\right\|_2^2\right)}{\sum\limits_{j=1}^{n} \exp\left(-\beta \left\|\mathbf{z} - \overline{\mathbf{c}_j}^t\right\|_2^2\right)} & \text{if} \quad \mathbf{z} \in \mathcal{Q} \,, \\ 1 & \text{if} \quad \mathbf{z} \in \mathcal{S}_i \,. \end{cases} \tag{6}$$

Contrary to the simple K-means algorithm, we used a weighted average where weight values are calculated via a decreasing function of the L_2 distance between data points and class barycenters—here, a softmax adjusted by a temperature value β. In our experiments, we used $\beta = 5$, which led to consistent results across datasets and backbones. In practice, we use a finite number of steps. By denoting $\mathbf{c}_i^\infty$ the resulting vectors, the predictions are:

$$\forall \mathbf{z} \in \mathcal{Q} : C_{tra}(\mathbf{z}, [\overline{\mathbf{c}_1}^\infty, \ldots, \overline{\mathbf{c}_n}^\infty]) = \arg\min_i \|\mathbf{z} - \overline{\mathbf{c}_i}^\infty\|_2 \,. \tag{7}$$

4. Results

4.1. Ranking on Standard Benchmarks

We first report results comparing our method with the state-of-the-art using classical settings and datasets. We used the following datasets:

- MiniImagenet: A dataset extracted from ImageNet with 64 base classes, 16 validation classes, and 20 novel classes. Each class contains 600 images. The resolution is (84×84);
- TieredImagenet: Another subset of ImageNet with 351 base classes, 97 validation classes, and 160 novel classes. Classes contain a variable number of samples, usually about 1300. The resolution is (84×84);
- CUB-FS (Caltech-UCSD Birds-200-2011): This dataset is particularly challenging because it is only composed of pictures of birds. There are 100 base classes, 50 validation classes, and 50 novel classes. The number of images by class is not constant, close to 60. The resolution is (50×50);

- FC-100 (Fewshot-CIFAR-100): This is a subset of CIFAR 100 (Canadian Institute for Advanced Research 100). There are 60 base, 20 validation, and 20 novel classes containing 600 images. Images have a low resolution (32×32);
- CIFAR-FS (CIFAR-Fewshot): This is also a subset of CIFAR 100. There are 60 base, 16 validation, and 20 novel classes containing 600 images. Images have a low resolution (32×32).

For each method, we specified the number of trainable parameters and the accuracy of 1-shot or 5-shot runs. Experiments always used $q = 15$ query samples per class, and results were averaged over 10,000 runs. Results are presented in Tables 1– 5 for the inductive setting and Tables 6–10 for the transductive setting (the codes allowing for the reproduction of our experiments are available at https://github.com/ybendou/easy). Reported results for the existing methods are those specified by their respective papers. Some methods do not include their standard deviation over multiple runs.

Let us first emphasize that our proposed methodology shows a new state-of-the-art performance for MiniImageNet (inductive), TieredImageNet (inductive 1-shot setting) and FC100 (transductive), while showcasing competitive or overlapping results on other benchmarks. We believe that, combined with other more elaborate methods, these results could be improved by a fair margin, leading to a new standard of performance for few-shot benchmarks. In the transductive setting, the proposed methodology is less often ranked #1, but contrary to many alternatives, it does not use any prior on class balance in the generated few-shot problems. We provide such experiments in the Supplementary Material, where we show that the proposed method greatly outperforms existing techniques when considering imbalanced classes. Overall, our method has the benefit of being simpler while achieving competitive performance over multiple benchmarks.

Table 1. The 1-shot and 5-shot accuracy of state-of-the-art methods and the proposed solution on MiniImageNet in the inductive setting.

	Method	1-Shot	5-Shot
$\leq$ 12 M	SimpleShot [35]	62.85 ± 0.20	80.02 ± 0.14
	Baseline++ [36]	53.97 ± 0.79	75.90 ± 0.61
	TADAM [42]	58.50 ± 0.30	76.70 ± 0.30
	ProtoNet [16]	60.37 ± 0.83	78.02 ± 0.57
	R2-D2 (+ens) [28]	64.79 ± 0.45	81.08 ± 0.32
	FEAT [43]	66.78	82.05
	CNL [44]	67.96 ± 0.98	83.36 ± 0.51
	MELR [45]	67.40 ± 0.43	83.40 ± 0.28
	Deep EMD v2 [20]	68.77 ± 0.29	84.13 ± 0.53
	PAL [14]	69.37 ± 0.64	84.40 ± 0.44
	invariance-equivariance [46]	67.28 ± 0.80	84.78 ± 0.50
	CSEI [19]	68.94 ± 0.28	85.07 ± 0.50
	COSOC [15]	69.28 ± 0.49	85.16 ± 0.42
	EASY $2\times$ResNet12$\left(\frac{1}{\sqrt{2}}\right)$ (ours)	$\mathbf{70.63 \pm 0.20}$	$\mathbf{86.28 \pm 0.12}$
36 M	S2M2R [1]	64.93 ± 0.18	83.18 ± 0.11
	LR + DC [24]	68.55 ± 0.55	82.88 ± 0.42
	EASY $3\times$ResNet12 (ours)	$\mathbf{71.75 \pm 0.19}$	$\mathbf{87.15 \pm 0.12}$

Table 2. The 1-shot and 5-shot accuracy of state-of-the-art methods and the proposed solution on TieredImageNet in the inductive setting.

	Method	1-Shot	5-Shot
≤ 12 M	SimpleShot [35]	69.09 ± 0.22	84.58 ± 0.16
	ProtoNet [16]	65.65 ± 0.92	83.40 ± 0.65
	FEAT [43]	70.80 ± 0.23	84.79 ± 0.16
	PAL [14]	72.25 ± 0.72	86.95 ± 0.47
	DeepEMD v2 [20]	74.29 ± 0.32	86.98 ± 0.60
	MELR [45]	72.14 ± 0.51	87.01 ± 0.35
	COSOC [15]	73.57 ± 0.43	87.57 ± 0.10
	CNL [44]	73.42 ± 0.95	87.72 ± 0.75
	invariance-equivariance [46]	72.21 ± 0.90	87.08 ± 0.58
	CSEI [19]	73.76 ± 0.32	87.83 ± 0.59
	ASY ResNet12 (ours)	$\mathbf{74.31 \pm 0.22}$	$\mathbf{87.86 \pm 0.15}$
36 M	S2M2R [1]	73.71 ± 0.22	$\mathbf{88.52 \pm 0.14}$
	EASY 3×ResNet12 (ours)	$\mathbf{74.71 \pm 0.22}$	88.33 ± 0.14

Table 3. The 1-shot and 5-shot accuracy of state-of-the-art methods and the proposed solution on **CIFAR-FS** in the **inductive** setting.

	Method	1-Shot	5-Shot
≤ 12 M	S2M2R [1]	63.66 ± 0.17	76.07 ± 0.19
	R2-D2 (+ens) [28]	76.51 ± 0.47	87.63 ± 0.34
	invariance-equivariance [46]	$\mathbf{77.87 \pm 0.85}$	$\mathbf{89.74 \pm 0.57}$
	EASY 2×ResNet12$\left(\frac{1}{\sqrt{2}}\right)$ (ours)	75.24 ± 0.20	88.38 ± 0.14
36 M	S2M2R [1]	74.81 ± 0.19	87.47 ± 0.13
	EASY 3×ResNet12 (ours)	$\mathbf{76.20 \pm 0.20}$	$\mathbf{89.00 \pm 0.14}$

Table 4. The 1-shot and 5-shot accuracy of state-of-the-art methods and the proposed solution on CUB-FS in the inductive setting.

	Method	1-Shot	5-Shot
≤ 12 M	FEAT [43]	68.87 ± 0.22	82.90 ± 0.10
	ProtoNet [16]	66.09 ± 0.92	82.50 ± 0.58
	DeepEMD v2 [20]	$\mathbf{79.27 \pm 0.29}$	89.80 ± 0.51
	EASY 4×ResNet12$\left(\frac{1}{2}\right)$ (ours)	77.97 ± 0.20	$\mathbf{91.59 \pm 0.10}$
36 M	S2M2R [1]	$\mathbf{80.68 \pm 0.81}$	90.85 ± 0.44
	EASY 3×ResNet12 (ours)	78.56 ± 0.19	$\mathbf{91.93 \pm 0.10}$

Table 5. The 1-shot and 5-shot accuracy of state-of-the-art methods and the proposed solution on FC-100 in the inductive setting.

	Method	1-Shot	5-Shot
$\leq$ 12 M	DeepEMD v2 [20]	46.60 ± 0.26	63.22 ± 0.71
	TADAM [42]	40.10 ± 0.40	56.10 ± 0.40
	ProtoNet [16]	41.54 ± 0.76	57.08 ± 0.76
	invariance-equivariance [46]	47.76 ± 0.77	$\mathbf{65.30 \pm 0.76}$
	R2-D2 (+ens) [28]	44.75 ± 0.43	59.94 ± 0.41
	EASY 2×ResNet12$\left(\frac{1}{\sqrt{2}}\right)$ (ours)	$\mathbf{47.94 \pm 0.19}$	64.14 ± 0.19
36 M	EASY 3×ResNet12 (ours)	$\mathbf{48.07 \pm 0.19}$	64.74 ± 0.19

Table 6. The 1-shot and 5-shot accuracy of state-of-the-art methods and the proposed solution on MiniImageNet in the transductive setting.

	Method	1-Shot	5-Shot
$\leq$ 12 M	TIM-GD [47]	73.90	85.00
	ODC [48]	77.20 ± 0.36	87.11 ± 0.42
	PEM$_n$E-BMS* [38]	80.56 ± 0.27	87.98 ± 0.14
	SSR [49]	68.10 ± 0.60	76.90 ± 0.40
	iLPC [50]	69.79 ± 0.99	79.82 ± 0.55
	EPNet [37]	66.50 ± 0.89	81.60 ± 0.60
	DPGN [51]	67.77 ± 0.32	84.60 ± 0.43
	ECKPN [52]	70.48 ± 0.38	85.42 ± 0.46
	Rot + KD + POODLE [53]	77.56	85.81
	EASY 2×ResNet12$\left(\frac{1}{\sqrt{2}}\right)$ (ours)	$\mathbf{82.31 \pm 0.24}$	$\mathbf{88.57 \pm 0.12}$
36 M	SSR [49]	72.40 ± 0.60	80.20 ± 0.40
	fine-tuning(train+val) [54]	68.11 ± 0.69	80.36 ± 0.50
	SIB + E^{3}BM [55]	71.40	81.20
	LR + DC [24]	68.57 ± 0.55	82.88 ± 0.42
	EPNet [37]	70.74 ± 0.85	84.34 ± 0.53
	TIM-GD [47]	77.80	87.40
	PT+MAP [56]	82.92 ± 0.26	88.82 ± 0.13
	iLPC [50]	83.05 ± 0.79	88.82 ± 0.42
	ODC [48]	80.64 ± 0.34	89.39 ± 0.39
	PEM$_n$E-BMS* [38]	83.35 ± 0.25	$\mathbf{89.53 \pm 0.13}$
	EASY 3×ResNet12 (ours)	$\mathbf{84.04 \pm 0.23}$	89.14 ± 0.11

Table 7. The 1-shot and 5-shot accuracy of state-of-the-art methods and the proposed solution on CUB-FS in the transductive setting.

	Method	1-Shot	5-Shot
≤ 12 M	TIM-GD [47]	82.20	90.80
	ODC [48]	85.87	**94.97**
	DPGN [51]	75.71 ± 0.47	91.48 ± 0.33
	ECKPN [52]	77.43 ± 0.54	92.21 ± 0.41
	iLPC [50]	89.00 ± 0.70	92.74 ± 0.35
	Rot + KD + POODLE [53]	89.93	93.78
	EASY 4×ResNet12$\left(\frac{1}{2}\right)$ (ours)	**90.50 ± 0.19**	93.50 ± 0.09
36 M	LR + DC [24]	79.56 ± 0.87	90.67 ± 0.35
	PT+MAP [56]	**91.55 ± 0.19**	93.99 ± 0.10
	iLPC [50]	91.03 ± 0.63	**94.11 ± 0.30**
	EASY 3×ResNet12 (ours)	90.56 ± 0.19	93.79 ± 0.10

Table 8. The 1-shot and 5-shot accuracy of state-of-the-art methods and the proposed solution on **FC-100** in the **transductive** setting.

	Method	1-Shot	5-Shot
≤ 12 M	TADAM [42]	40.10 ± 0.40	56.10 ± 0.40
	EASY 2×ResNet12$\left(\frac{1}{\sqrt{2}}\right)$ (ours)	**54.47 ± 0.24**	**65.82 ± 0.19**
36 M	SIB + E^{3}BM [55]	46.00	57.10
	fine-tuning (train) [54]	43.16 ± 0.59	57.57 ± 0.55
	ODC [48]	47.18 ± 0.30	59.21 ± 0.56
	fine-tuning (train+val) [54]	50.44 ± 0.68	65.74 ± 0.60
	EASY 3×ResNet12 (ours)	**54.13 ± 0.24**	**66.86 ± 0.19**

Table 9. The 1-shot and 5-shot accuracy of state-of-the-art methods and the proposed solution on CIFAR-FS in the transductive setting.

	Method	1-Shot	5-Shot
≤ 12 M	SSR [49]	76.80 ± 0.60	83.70 ± 0.40
	iLPC [50]	77.14 ± 0.95	85.23 ± 0.55
	DPGN [51]	77.90 ± 0.50	90.02 ± 0.40
	ECKPN [52]	79.20 ± 0.40	**91.00 ± 0.50**
	EASY 2×ResNet12$\left(\frac{1}{\sqrt{2}}\right)$ (ours)	**86.99 ± 0.21**	90.20 ± 0.15
36 M	SSR [49]	81.60 ± 0.60	86.00 ± 0.40
	fine-tuning (train+val) [54]	78.36 ± 0.70	87.54 ± 0.49
	iLPC [50]	86.51 ± 0.75	90.60 ± 0.48
	PT+MAP [56]	**87.69 ± 0.23**	**90.68 ± 0.15**
	EASY 3×ResNet12 (ours)	87.16 ± 0.21	90.47 ± 0.15

Table 10. The 1-shot and 5-shot accuracy of state-of-the-art methods and the proposed solution on TieredImageNet in the transductive setting.

	Method	1-Shot	5-Shot
≤ 12 M	PT+MAP [56]	85.67 ± 0.26	90.45 ± 0.14
	TIM-GD [47]	79.90	88.50
	ODC [48]	83.73 ± 0.36	$\mathbf{90.46 \pm 0.46}$
	SSR [49]	81.20 ± 0.60	85.70 ± 0.40
	Rot + KD + POODLE [53]	79.67	86.96
	DPGN [51]	72.45 ± 0.51	87.24 ± 0.39
	EPNet [37]	76.53 ± 0.87	87.32 ± 0.64
	ECKPN [52]	73.59 ± 0.45	88.13 ± 0.28
	iLPC [50]	83.49 ± 0.88	89.48 ± 0.47
	ASY ResNet12 (ours)	$\mathbf{83.98 \pm 0.24}$	89.26 ± 0.14
36 M	SIB + E^3BM [55]	75.60	84.30
	SSR [49]	79.50 ± 0.60	84.80 ± 0.40
	fine-tuning (train+val) [54]	72.87 ± 0.71	86.15 ± 0.50
	TIM-GD [47]	82.10	89.80
	LR + DC [24]	78.19 ± 0.25	89.90 ± 0.41
	EPNet [37]	78.50 ± 0.91	88.36 ± 0.57
	ODC [48]	85.22 ± 0.34	91.35 ± 0.42
	iLPC [50]	$\mathbf{88.50 \pm 0.75}$	$\mathbf{92.46 \pm 0.42}$
	PEM_nE-BMS* [38]	86.07 ± 0.25	91.09 ± 0.14
	EASY 3×ResNet12 (ours)	84.29 ± 0.24	89.76 ± 0.14

4.2. Ablation Study

To better understand the relative contributions of components in the proposed method, we also compare, for each dataset, the performance of various combinations in Table 11 for the inductive setting and Table 12 for the transductive setting. Interestingly, the full proposed methodology (EASY) is not always the most efficient. We believe that for large datasets such as MiniImageNet and TieredImageNet, the considered ResNet12 backbones contain too few parameters. When reducing this number for ensemble solutions, the drop of performance due to the reduction in size is not compensated by the diversity of the multiple backbones. All things considered, only AS is consistently beneficial to the performance.

Table 11. Ablation study of the steps of proposed solution in inductive setting, for a fixed number of trainable parameters in the considered backbones. When using ensembles, we use $2\times$ResNet12$\left(\frac{1}{\sqrt{2}}\right)$ instead of a single ResNet12.

Dataset	E	AS	1-Shot	5-Shot
MiniImageNet			68.43 ± 0.19	83.78 ± 0.13
		✓	$\mathbf{70.84 \pm 0.19}$	85.70 ± 0.13
	✓		68.69 ± 0.20	84.84 ± 0.13
	✓	✓	70.63 ± 0.20	$\mathbf{86.28 \pm 0.12}$
CUB-FS			74.13 ± 0.20	89.08 ± 0.11
		✓	77.40 ± 0.20	$\mathbf{91.15 \pm 0.10}$
	✓		75.01 ± 0.20	89.38 ± 0.11
	✓	✓	$\mathbf{77.59 \pm 0.20}$	91.07 ± 0.11
CIFAR-FS			73.38 ± 0.21	87.42 ± 0.15
		✓	74.26 ± 0.21	88.16 ± 0.15
	✓		74.36 ± 0.21	87.82 ± 0.15
	✓	✓	$\mathbf{75.24 \pm 0.20}$	$\mathbf{88.38 \pm 0.14}$
FC-100			45.68 ± 0.19	62.78 ± 0.19
		✓	46.43 ± 0.19	64.16 ± 0.19
	✓		47.52 ± 0.19	63.92 ± 0.19
	✓	✓	$\mathbf{47.94 \pm 0.20}$	$\mathbf{64.14 \pm 0.19}$
TieredImageNet			72.52 ± 0.22	86.79 ± 0.15
		✓	$\mathbf{74.17 \pm 0.22}$	$\mathbf{87.81 \pm 0.14}$
	✓		72.14 ± 0.22	86.66 ± 0.15
	✓	✓	73.36 ± 0.22	87.37 ± 0.15

Table 12. Ablation study of the steps of proposed solution in **transductive** setting for a fixed number of trainable parameters in the considered backbones. When using ensembles, we use $2\times$ResNet12$\left(\frac{1}{\sqrt{2}}\right)$ instead of a single ResNet12.

Dataset	E	AS	1-Shot	5-Shot
MiniImageNet			80.42 ± 0.23	86.72 ± 0.13
		✓	$\mathbf{83.02 \pm 0.23}$	88.36 ± 0.12
	✓		80.27 ± 0.23	87.45 ± 0.12
	✓	✓	82.31 ± 0.24	$\mathbf{88.57 \pm 0.12}$
CUB-FS			86.93 ± 0.21	91.53 ± 0.11
		✓	89.80 ± 0.20	93.12 ± 0.10
	✓		87.28 ± 0.21	91.89 ± 0.10
	✓	✓	$\mathbf{90.05 \pm 0.19}$	$\mathbf{93.17 \pm 0.10}$
CIFAR-FS			84.18 ± 0.23	89.56 ± 0.15
		✓	85.55 ± 0.23	90.07 ± 0.15
	✓		84.89 ± 0.22	89.60 ± 0.15
	✓	✓	$\mathbf{86.99 \pm 0.21}$	$\mathbf{90.20 \pm 0.15}$
FC-100			51.74 ± 0.23	65.39 ± 0.19
		✓	52.93 ± 0.23	$\mathbf{66.51 \pm 0.19}$
	✓		53.39 ± 0.23	65.71 ± 0.19
	✓	✓	$\mathbf{54.47 \pm 0.24}$	65.82 ± 0.19
TieredImageNet			82.32 ± 0.24	88.45 ± 0.15
		✓	$\mathbf{83.98 \pm 0.24}$	$\mathbf{89.26 \pm 0.14}$
	✓		81.48 ± 0.25	88.40 ± 0.15
	✓	✓	83.20 ± 0.25	88.92 ± 0.14

4.3. Discussion

Regarding the inductive setting, the proposed method achieves state-of-the-art performance on MiniImagenet by a fair margin. On TieredImagenet, only S2M2R performs better in the five-shot setting. This can be explained by the fact that TieredImagenet is the largest of the considered datasets and it requires more parameters to be trained efficiently, reducing the effectiveness of the proposed ensemble approach. We also noticed subpar performance on CIFAR-FS, and we believe that this is due to the small resolution of images in the dataset, which cripples the augmented sample step. On the FC-100 dataset, our results in the inductive setting overlap with [46]; however, our method has the advantage of having lower confidence intervals compared to other methods on the same benchmark. Regarding the transductive setting, our method achieves competitive results without any prior on the number of classes. This is important since multiple methods tend to fail when the number of samples per class is different, which we show in the Supplementary Material. Our explanation is that multiple methods tend to overexploit this prior. This concern was first raised by [57]. Overall, our method is easy to implement and requires few hyperparameters to be tuned compared to other competitive methods.

5. Conclusions

In this paper, we introduced a simple backbone to perform few-shot classification in both inductive and transductive settings. Combined with augmented samples and ensembling, we showed its ability to reach state-of-the-art results when deployed using simple classifiers on multiple standardized benchmarks, even beating previous methods by a fair margin (>1%) in some cases. We expect this methodology to serve as a baseline for future work.

Author Contributions: Everyone participated in the writing of this article. Y.B. developed the cropping technique and participated in the optimization of hyperparameters. Y.H. developed the transductive part. R.L. participated in the optimization of the hyperparameters. G.L., B.P., S.P., and V.G. gave direction and supervision. V.G. was also responsible for the code basis on which the techniques were developed. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Available online: https://github.com/ybendou/easy. (accessed on 14 June 2022)

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

Abbreviations

The following abbreviations are used in this manuscript:

DL	Deep learning
S2	Self-supervision
S2M2R	Self-supervision manifold mixup rotations
AD	Augmented samples

Appendix A. Transductive Tests with Imbalanced Settings

Following the methodology recently proposed in [57], we also report performance in the transductive setting when the number of query vectors is varying for each class and is unknown. Results are presented in Tables A1–A3. The uncertainties of previously published methods were not reported by [57]. We note that the proposed methodology is able to outperform existing ones by a fair margin.

Table A1. The 1-shot and 5-shot accuracy of state-of-the-art methods and the proposed solution on MiniImageNet in the imbalanced transductive setting.

	Method	1-Shot	5-Shot
≤ 12 M	MAML [4]	47.6	64.5
	LR+ICI [58]	58.7	73.5
	PT+MAP [56]	60.1	67.1
	LaplacianShot [59]	65.4	81.6
	TIM [47]	67.3	79.8
	α-TIM [57]	67.4	82.5
	ASY ResNet12 (ours)	**75.65** $\pm$ **0.25**	**86.35** $\pm$ **0.14**
36 M	PT+MAP [56]	60.6	66.8
	SIB [60]	64.7	72.5
	LaplacianShot [59]	68.1	83.2
	TIM [47]	69.8	81.6
	α-TIM [57]	69.8	84.8
	EASY 3$\times$ResNet12 (ours)	**76.04** $\pm$ **0.27**	**87.23** $\pm$ **0.15**

Table A2. The 1-shot and 5-shot accuracy of state-of-the-art methods and the proposed solution on TieredImageNet in the imbalanced transductive setting.

	Method	1-Shot	5-Shot
≤ 12 M	Entropy-min [54]	61.2	75.5
	PT+MAP [56]	64.1	70.0
	LaplacianShot [59]	72.3	85.7
	TIM [47]	74.1	84.1
	LR+ICI [58]	74.6	85.1
	α-TIM [57]	74.4	86.6
	ASY ResNet12 (ours)	**78.15** $\pm$ **0.27**	**87.65** $\pm$ **0.17**
36 M	Entropy-min [54]	62.9	77.3
	PT+MAP [56]	65.1	71.0
	LaplacianShot [59]	73.5	86.8
	TIM [47]	75.8	85.4
	α-TIM [57]	76.0	87.8
	EASY 3$\times$ResNet12 (ours)	**78.46** $\pm$ **0.28**	**87.85** $\pm$ **0.13**

Table A3. The 1-shot and 5-shot accuracy of state-of-the-art methods and the proposed solution on CUB-FS in the imbalanced transductive setting.

	Method	1-Shot	5-Shot
≤ 12 M	PT+MAP [56]	65.1	71.3
	Entropy-min [54]	67.5	82.9
	LaplacianShot [59]	73.7	87.7
	TIM [47]	74.8	86.9
	α-TIM [57]	75.7	89.8
	ASY ResNet12 (ours)	**81.24** $\pm$ **0.27**	**87.27** $\pm$ **0.14**
36 M	EASY 3$\times$ResNet12 (ours)	**83.63** $\pm$ **0.25**	**92.35** $\pm$ **0.09**

Appendix B. Additional Ablation Studies

Appendix B.1. Influence of the Temperature in the Transductive Setting

In Figure A1, we show how different values of the temperature β of the soft K-means influence the performance of our model. We observed that $\beta = 5$ seems to lead to the best results on the two considered datasets, which is why we chose this value in our other experiments. Note that we used three ResNet12 with 30 augmented samples in this experiment.

Appendix B.2. Influence of the Number of Crops

In Figure A2, we show how the performance of our model is influenced by the number of crops ℓ used during augmented sampling (AS). When using $\ell = 1$, we report the performance of the method using no crops, but a global reshape instead. We observed that the performance keeps increasing as long as the number of crops used is increased, except for a small drop in performance when switching from a global reshape to crops—this drop can easily be explained as crops are likely to miss the object of interest. However, the computational time to generate the crops also increases linearly. Therefore, we used $\ell = 30$ as a trade-off between performance and time complexity. Here, we used a single ResNet12 for our experiments.

Appendix B.3. Influence of the Number of Backbones

In Figure A3, we show how the performance of our model is influenced by the number of backbones b used during the ensemble step (E). The performance increases steadily with a strong diminishing return. We used 30 augmented samples in this experiment.

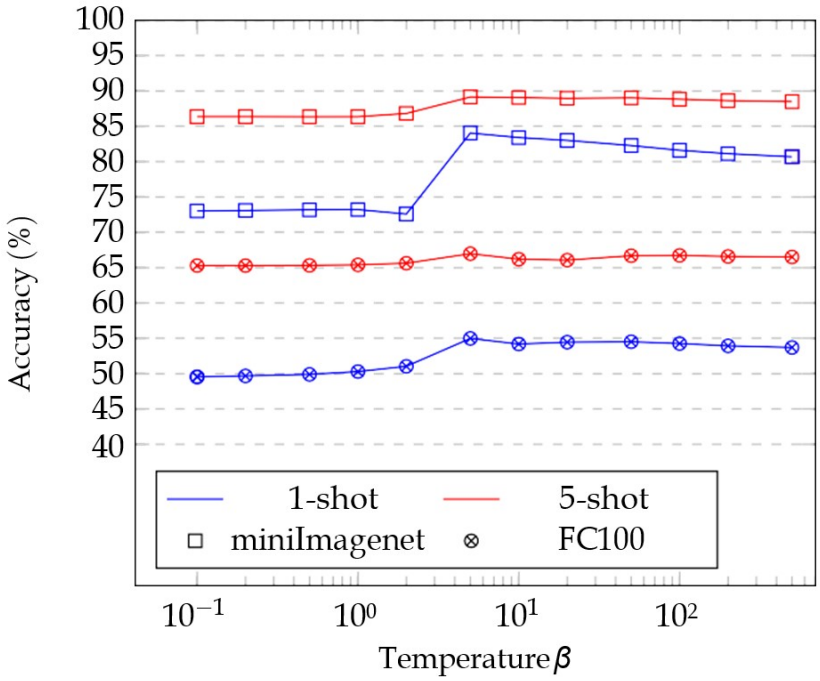

Figure A1. Ablation study of temperature of the soft K-means used in the transductive setting; we performed 10^5 runs for each value of β.

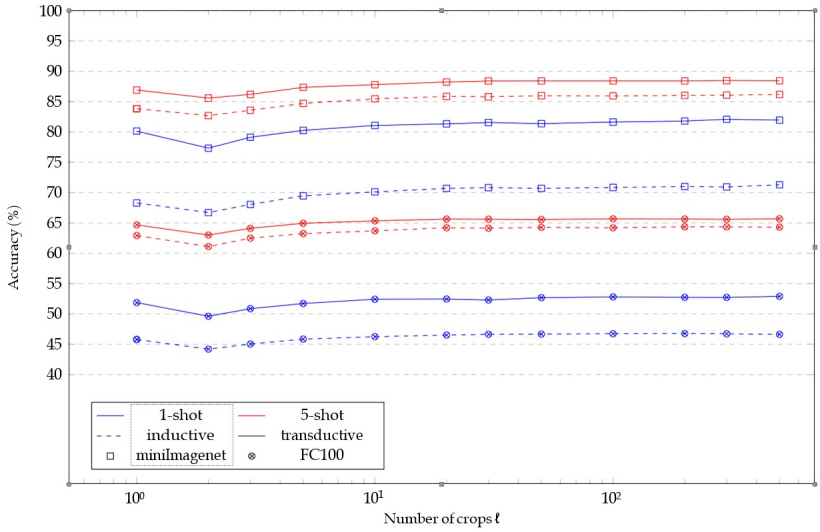

Figure A2. Ablation study of augmented samples; we performed 10^5 runs for each value of ℓ.

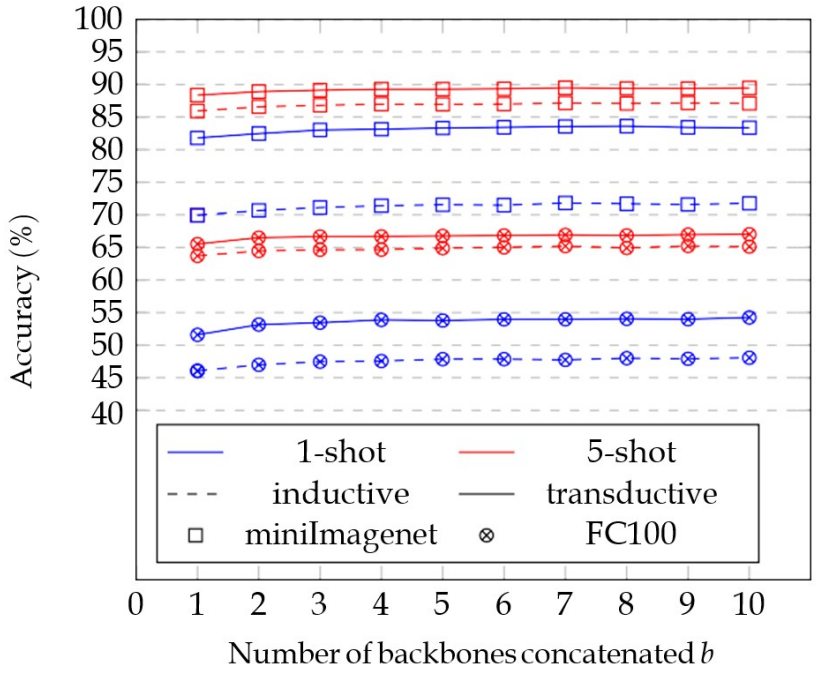

Figure A3. Ablation study of the number of backbones; we performed 10^5 runs for each value of b.

References

1. Mangla, P.; Kumari, N.; Sinha, A.; Singh, M.; Krishnamurthy, B.; Balasubramanian, V.N. Charting the right manifold: Manifold mixup for few-shot learning. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 4–8 January 2020; pp. 2218–2227.
2. Chen, D.; Chen, Y.; Li, Y.; Mao, F.; He, Y.; Xue, H. Self-supervised learning for few-shot image classification. In Proceedings of the ICASSP 2021—2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, USA, 6–11 June 2021; pp. 1745–1749.
3. Yan, S.; Zhang, S.; He, X. A Dual Attention Network with Semantic Embedding for Few-Shot Learning. In Proceedings of the AAAI, Honolulu, HI, USA, 27–28 January 2019; pp. 9079–9086.
4. 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, Singapore, 24–26 February 2017; pp. 1126–1135.
5. Munkhdalai, T.; Yuan, X.; Mehri, S.; Trischler, A. Rapid adaptation with conditionally shifted neurons. In Proceedings of the International Conference on Machine Learning, Stockholm, Sweden, 10–15 July 2018; pp. 3664–3673.

6. Lee, K.; Maji, S.; Ravichandran, A.; Soatto, S. Meta-learning with differentiable convex optimization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 10657–10665.
7. Munkhdalai, T.; Yu, H. Meta networks. In Proceedings of the International Conference on Machine Learning, Sydney, NSW, Australia, 6–11 August 2017; pp. 2554–2563.
8. Zhang, C.; Ding, H.; Lin, G.; Li, R.; Wang, C.; Shen, C. Meta navigator: Search for a good adaptation policy for few-shot learning. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 11–17 October 2021; pp. 9435–9444.
9. Scott, T.R.; Ridgeway, K.; Mozer, M.C. Adapted deep embeddings: A synthesis of methods for k-shot inductive transfer learning. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 3–8 December 2018; pp. 76–85.
10. Liu, Y.; Lee, J.; Park, M.; Kim, S.; Yang, E.; Hwang, S.J.; Yang, Y. Learning to propagate labels: Transductive propagation network for few-shot learning. *arXiv* **2018**, arXiv:1805.10002.
11. Bontonou, M.; Lioi, G.; Farrugia, N.; Gripon, V. Few-Shot Decoding of Brain Activation Maps. In Proceedings of the 2021 29th European Signal Processing Conference (EUSIPCO), Dublin, Ireland, 23–27 August 2021; pp. 1326–1330.
12. Henderson, R.D.; Yi, X.; Adams, S.J.; Babyn, P. Automatic Detection and Classification of Multiple Catheters in Neonatal Radiographs with Deep Learning. *J. Digit. Imaging* **2021**, *34*, 888–897. [CrossRef] [PubMed]
13. Konstantin, E.; Elena, S.; Manvel, A.; Alexander, T. Noise-resilient Automatic Interpretation of Holter ECG Recordings. In Proceedings of the BIOSIGNALS 2021-14th International Conference on Bio-Inspired Systems and Signal Processing. Part of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies, BIOSTEC 2021, Online, 11–13 February 2021; pp. 208–214.
14. Ma, J.; Xie, H.; Han, G.; Chang, S.F.; Galstyan, A.; Abd-Almageed, W. Partner-Assisted Learning for Few-Shot Image Classification. In Proceedings of the IEEE/CVF International Conference on Computer Vision 2021, Montreal, QC, Canada, 11–17 October 2021; pp. 10573–10582.
15. Luo, X.; Wei, L.; Wen, L.; Yang, J.; Xie, L.; Xu, Z.; Tian, Q. Rectifying the Shortcut Learning of Background for Few-Shot Learning. In Proceedings of the Advances in Neural Information Processing Systems, Online, 6–14 December 2021; Volume 34.
16. Snell, J.; Swersky, K.; Zemel, R. Prototypical networks for few-shot learning. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; Volume 30.
17. Zhang, H.; Cisse, M.; Dauphin, Y.N.; Lopez-Paz, D. MixUp: Beyond empirical risk minimization. In Proceedings of the 6th International Conference on Learning Representations, ICLR 2018—Conference Track Proceedings, Vancouver, BC, Canada, 30 April–3 May 2018.
18. Verma, V.; Lamb, A.; Beckham, C.; Najafi, A.; Mitliagkas, I.; Lopez-Paz, D.; Bengio, Y. Manifold mixup: Better representations by interpolating hidden states. In Proceedings of the 36th International Conference on Machine Learning ICML, Long Beach, CA, USA, 9–15 June 2019; pp. 11196–11205.
19. Li, J.; Wang, Z.; Hu, X. Learning Intact Features by Erasing-Inpainting for Few-shot Classification. In Proceedings of the AAAI Conference on Artificial Intelligence, Online, 2–9 February 2021; Volume 35, pp. 8401–8409.
20. Zhang, C.; Cai, Y.; Lin, G.; Shen, C. DeepEMD: Few-Shot Image Classification With Differentiable Earth Mover's Distance and Structured Classifiers. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Online, 14–19 June 2020; pp. 12203–12213.
21. Choe, J.; Park, S.; Kim, K.; Hyun Park, J.; Kim, D.; Shim, H. Face generation for low-shot learning using generative adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision Workshops, Venice, Italy, 22–29 October 2017; pp. 1940–1948.
22. Li, K.; Zhang, Y.; Li, K.; Fu, Y. Adversarial feature hallucination networks for few-shot learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Online, 14–19 June 2020; pp. 13470–13479.
23. Hariharan, B.; Girshick, R. Low-shot visual recognition by shrinking and hallucinating features. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 3018–3027.
24. Yang, S.; Liu, L.; Xu, M. Free lunch for few-shot learning: Distribution calibration. *arXiv* **2021**, arXiv:2101.06395.
25. Gidaris, S.; Bursuc, A.; Komodakis, N.; Pérez, P.; Cord, M. Boosting few-shot visual learning with self-supervision. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 8059–8068.
26. Ravi, S.; Larochelle, H. Optimization as a model for few-shot learning. In Proceedings of the International Conference on Learning Representations (ICLR), Toulon, France, 24–26 April 2017.
27. Vinyals, O.; Blundell, C.; Lillicrap, T.; Wierstra, D. Matching networks for one shot learning. In Proceedings of the Advances in Neural Information Processing Systems, Barcelona, Spain, 5–10 December 2016; Volume 29.
28. Liu, J.; Chao, F.; Lin, C.M. Task augmentation by rotating for meta-learning. *arXiv* **2020**, arXiv:2003.00804.
29. Luo, X.; Chen, Y.; Wen, L.; Pan, L.; Xu, Z. Boosting few-shot classification with view-learnable contrastive learning. In Proceedings of the 2021 IEEE International Conference on Multimedia and Expo (ICME), Shenzhen, China, 5–9 July 2021; pp. 1–6.
30. Liu, C.; Fu, Y.; Xu, C.; Yang, S.; Li, J.; Wang, C.; Zhang, L. Learning a Few-shot Embedding Model with Contrastive Learning. In Proceedings of the AAAI Conference on Artificial Intelligence, Online, 2–9 February 2021; Volume 35, pp. 8635–8643.
31. Majumder, O.; Ravichandran, A.; Maji, S.; Polito, M.; Bhotika, R.; Soatto, S. Revisiting contrastive learning for few-shot classification. *arXiv* **2021**, arXiv:2101.110058v1.

32. Khosla, P.; Teterwak, P.; Wang, C.; Sarna, A.; Tian, Y.; Isola, P.; Maschinot, A.; Liu, C.; Krishnan, D. Supervised contrastive learning. In Proceedings of the Advances in Neural Information Processing Systems, Online, 6–12 December 2020; Volume 33, pp. 18661–18673.
33. Tian, Y.; Wang, Y.; Krishnan, D.; Tenenbaum, J.B.; Isola, P. Rethinking few-shot image classification: A good embedding is all you need? In Proceedings of the European Conference on Computer Vision, Glasgow, Scotland, 23–28 August 2020; pp. 266–282.
34. Huang, G.; Li, Y.; Pleiss, G.; Liu, Z.; Hopcroft, J.E.; Weinberger, K.Q. Snapshot ensembles: Train 1, get m for free. *arXiv* **2017**, arXiv:1704.00109.
35. Wang, Y.; Chao, W.L.; Weinberger, K.Q.; van der Maaten, L. SimpleShot: Revisiting Nearest-Neighbor Classification for Few-Shot Learning. *arXiv* **2019**, arXiv:1911.04623.
36. Chen, W.Y.; Wang, Y.C.F.; Liu, Y.C.; Kira, Z.; Huang, J.B. A closer look at few-shot classification. In Proceedings of the 7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, 6–9 May 2019; pp. 1–17.
37. Rodríguez, P.; Laradji, I.; Drouin, A.; Lacoste, A. Embedding propagation: Smoother manifold for few-shot classification. In Proceedings of the European Conference on Computer Vision, Glasgow, Scotland, 23–28 August 2020; pp. 121–138.
38. Hu, Y.; Gripon, V.; Pateux, S. Squeezing Backbone Feature Distributions to the Max for Efficient Few-Shot Learning. *Algorithms* **2022**, *15*, 147.
39. Loshchilov, I.; Hutter, F. Sgdr: Stochastic gradient descent with warm restarts. *arXiv* **2016**, arXiv:1608.03983.
40. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
41. Shorten, C.; Khoshgoftaar, T.M. A survey on image data augmentation for deep learning. *J. Big Data* **2019**, *6*, 1–48. [CrossRef]
42. Oreshkin, B.N.; Rodriguez, P.; Lacoste, A. Tadam: Task dependent adaptive metric for improved few-shot learning. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 3–8 December 2018; pp. 721–731.
43. Ye, H.J.; Hu, H.; Zhan, D.C.; Sha, F. Few-shot learning via embedding adaptation with set-to-set functions. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 8808–8817.
44. Zhao, J.; Yang, Y.; Lin, X.; Yang, J.; He, L. Looking Wider for Better Adaptive Representation in Few-Shot Learning. In Proceedings of the AAAI Conference on Artificial Intelligence, Online, 2–9 February 2021; Volume 35, pp. 10981–10989.
45. Fei, N.; Lu, Z.; Xiang, T.; Huang, S. MELR: Meta-learning via modeling episode-level relationships for few-shot learning. In Proceedings of the International Conference on Learning Representations, Online, 26 April–1 May 2020.
46. Rizve, M.N.; Khan, S.; Khan, F.S.; Shah, M. Exploring Complementary Strengths of Invariant and Equivariant Representations for Few-Shot Learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 10836–10846.
47. Boudiaf, M.; Ziko, I.; Rony, J.; Dolz, J.; Piantanida, P.; Ben Ayed, I. Information maximization for few-shot learning. In Proceedings of the Advances in Neural Information Processing Systems, Online, 6–12 December 2020; Volume 33, pp. 2445–2457.
48. Qi, G.; Yu, H.; Lu, Z.; Li, S. Transductive few-shot classification on the oblique manifold. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Online, 11-17 October 2021; pp. 8412-8422.
49. Shen, X.; Xiao, Y.; Hu, S.X.; Sbai, O.; Aubry, M. Re-ranking for image retrieval and transductive few-shot classification. In Proceedings of the Advances in Neural Information Processing Systems 2021, Online, 6–14 December 2021; Volume 34, pp. 25932–25943.
50. Lazarou, M.; Stathaki, T.; Avrithis, Y. Iterative label cleaning for transductive and semi-supervised few-shot learning. In Proceedings of the IEEE/CVF International Conference on Computer Vision 2021, Online, 11-17 October 2021; pp. 8751–8760.
51. Yang, L.; Li, L.; Zhang, Z.; Zhou, X.; Zhou, E.; Liu, Y. Dpgn: Distribution propagation graph network for few-shot learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Online, 14–19 June 2020; pp. 13390–13399.
52. Chen, C.; Yang, X.; Xu, C.; Huang, X.; Ma, Z. ECKPN: Explicit Class Knowledge Propagation Network for Transductive Few-shot Learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 19–25 June 2021; pp. 6596–6605
53. Le, D.; Nguyen, K.D.; Nguyen, K.; Tran, Q.H.; Nguyen, R.; Hua, B.S. POODLE: Improving Few-shot Learning via Penalizing Out-of-Distribution Samples. In Proceedings of the Advances in Neural Information Processing Systems 2021, Online, 6–14 December 2021; Volume 34, pp. 23942–23955.
54. Dhillon, G.S.; Chaudhari, P.; Ravichandran, A.; Soatto, S. A baseline for few-shot image classification. *arXiv* **2019**, arXiv:1909.02729.
55. Liu, Y.; Schiele, B.; Sun, Q. An ensemble of epoch-wise empirical bayes for few-shot learning. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; pp. 404–421.
56. Hu, Y.; Gripon, V.; Pateux, S. Leveraging the feature distribution in transfer-based few-shot learning. In Proceedings of the International Conference on Artificial Neural Networks, Bratislava, Slovakia, 14–17 September 2021; pp. 487–499.
57. Veilleux, O.; Boudiaf, M.; Piantanida, P.; Ben Ayed, I. Realistic evaluation of transductive few-shot learning. In Proceedings of the Advances in Neural Information Processing Systems, Online, 6–14 December 2021; Volume 34.
58. Wang, Y.; Xu, C.; Liu, C.; Zhang, L.; Fu, Y. Instance credibility inference for few-shot learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 12836–12845.

59. Ziko, I.; Dolz, J.; Granger, E.; Ayed, I.B. Laplacian regularized few-shot learning. In Proceedings of the International Conference on Machine Learning, Online, 13–18 July 2020; pp. 11660–11670.
60. Hu, S.X.; Moreno, P.G.; Xiao, Y.; Shen, X.; Obozinski, G.; Lawrence, N.D.; Damianou, A. Empirical bayes transductive meta-learning with synthetic gradients. *arXiv* **2020**, arXiv:2004.12696.