CVPR
#48

CVPR
#48

CVPR 2023 Submission #48. CONFIDENTIAL REVIEW COPY. DO NOT DISTRIBUTE.

LSFSL: Leveraging Shape Information in Few-shot Learning

Anonymous CVPR submission

Paper ID 48

Abstract

Few-shot learning (FSL) techniques seek to learn the underlying patterns in data using fewer samples, analogous to how humans learn from limited experience. In this limited-data scenario, the challenges associated with deep neural networks, such as shortcut learning and texture bias behaviors, are further exacerbated. Moreover, the significance of addressing shortcut learning is not yet fully explored in the few-shot setup. To address these issues, we propose LS-FSL, which enforces the model to learn more generalizable features utilizing the implicit prior information present in the data. Through comprehensive analyses, we demonstrate that LSFSL-trained models are less vulnerable to alteration in color schemes, statistical correlations, and adversarial perturbations leveraging the global semantics in the data. Our findings highlight the potential of incorporating relevant priors in few-shot approaches to increase robustness and generalization.

1. Introduction

Intelligent systems based on Deep Neural Networks (DNN) have achieved impressive performance in perception tasks [2,26,51], yet DNNs remain limited in their ability to acquire new skills with little experience. Acquisition of novel skills with limited data is considered an important measure of intelligence [9], however, conventional supervised learning methods require significant amounts of labeled training data and computational resources. Therefore, in the quest for the building of intelligent agents, few-shot learning (FSL) [46,55] plays an imperative role. FSL is a learning paradigm that addresses these challenges by utilizing a small set of training samples to quickly learn the necessary skills and generalize to novel tasks with less direct supervision than conventional learning [25].

Neural networks in the conventional training setup inadvertently learn spurious correlations in the data as shortcuts [15] and are biased to learn local texture information [16] instead of focusing on global features. These challenges are further accentuated in the FSL setup due to the limited training data [40], which greatly affects the generalization and robustness of the FSL models. Furthermore, the over-parameterized few-shot models lead to overfitting on the limited data. The significance of addressing the challenges of shortcut learning and texture bias in the FSL domain has not been thoroughly studied. Consequently, tackling these issues will aid in learning robust features that enhance the generalization of few-shot models in diverse real-world scenarios.

On the other hand, humans excel at learning with limited data to quickly adapt to new situations and remain both precise and robust in a constantly changing environment with minimal experience [3,47]. This ability of humans to learn novel tasks with little to no prior experience or knowledge can be attributed to structural priors [27]. For instance, the brain encompasses a structural prior to language acquisition that provides a meaningful disposition toward language learning in infants. Furthermore, studies on developmental cognitive neuroscience [12,32] report the utility of semantic information by infants [11,28] and adults [16] to recognize objects. Therefore, we leverage the prior knowledge already implicitly available in the data as a supervisory signal to help few-shot models learn the intended solution.

Inspired by cognitive biases, we propose an approach called *Leveraging Shape Information in Few-shot Learning (LSFSL)*, which injects prior knowledge into the few-shot models in the form of shape awareness. This shape bias addresses the texture bias in the neural network and aids in learning robust global features. Consequently, the shape information might improve the generalization and offer a robust solution in FSL. The implicit shape information already presented in the visual data is used to train the few-shot models without the need for any additional dataset or generative techniques. During training, we learn a dual network architecture, where one model extracts and leverages shape information from standard RGB input. This shape information is distilled into the model trained with standard RGB image in synchrony through latent space and decision space alignment objectives. The alignment objectives regularize both networks and address overfitting to trivial solutions. Thus, LSFSL aims to address the vulnerability of few-shot

CVPR
#48

CVPR
#48

CVPR 2023 Submission #48. CONFIDENTIAL REVIEW COPY. DO NOT DISTRIBUTE.

models to shortcuts and texture bias by incorporating shape awareness.

With extensive experiments on multiple datasets and settings, we show that our proposed approach boosts the generalization and robustness of FSL models to a greater extent. LSFSL-trained models are less susceptible to spurious correlations, texture bias, and adversarial perturbations. The contributions of this work are as follows:

- We propose an approach to integrate FSL models with a shape-based inductive bias.

- We evaluate the susceptibility of LSFSL-trained models and counterpart baseline models to shortcut learning and adversarial attacks.

- Our approach offers flexibility to be plugged into meta [44] and non-meta [41] learning based few-shot approaches to improve their performance.

2. Related Work

Few-shot Learning: FSL methods are broadly divided into three categories. Optimization-based methods utilize meta-learning to enable the model to converge with a few training samples quickly. Methods identify a good model initialization [6,13], optimizers for effective weight update [36] or both [5]. Hallucination-based methods address data deficiency by extending the support set with similar samples using feature hallucination [20] or data augmentation techniques. This converts the low-shot problem approximately to a standard classification problem. Metric-based methods learn an embedding model to map each image to a latent feature space [21]. The query samples are classified as the closest support category in the feature space. Additionally, FSL methods incorporate meta-learning [13,44] or non-meta standard supervised learning based [41,53] training strategy to train a base learner. The resultant models from the later methods serve as an excellent embedding model and illustrate the ability to be used as a common architecture for different FSL settings.

Knowledge Distillation in FSL: With the aim of developing a good base embedding model to improve few-shot generalization, RFS-Distill [53] incorporate sequential self-distillation after pretraining the model on base classes following Born-again Networks (BAN) [14]. This Born-again distillation with feature normalization enhances the representations from the backbones until a certain generation. Similarly, Invariant and Equivariant Representations (IER-Distill) [41] utilize knowledge distillation to incorporate meaningful inductive biases into the base feature extractor. IER incorporates both invariance and equivariance inductive biases of certain geometric transformations. Unlike RFS, IER employs a multi-head distillation over standard supervised distillation. Liu *et al*. [30] learn a base embedding model by unifying the two-stage training in RFS-Distill using online self-distillation. We incorporate shape bias by distilling features in latent space and model decisions between two networks trained synchronously.

Overcoming Biases in FSL: The biases affecting the generalization of conventional DNNs create a larger impact on the FSL setup. Ringer *et al*., study the effects of texture bias illustrated by Convolutional Neural Networks (CNN) in FSL [40] and address the same by modifying the training data with a combination of non-texture and texture-based images. The image background also serves as a shortcut in the FSL scenario and impacts classification performance. Luo *et al*. [31] mitigates the problem by sampling the foreground and original images in the training and testing phases. Finally, even though the feature distribution of the base classes used in pretraining FSL models is well defined, the distribution is affected by the domain shift incurred by the novel classes. Tao *et al*. [52] propose Distribution Calibration Module (DCM) and Selected Sampling (SS) to mitigate class-agnostic and class-specific bias, respectively, in the meta-testing phase. However, LSFSL utilizes the implicit shape information in the training data with alignment objectives that aid in improving the generalization of few-shot models by learning global features.

Shape Bias in FSL: Stojanov *et al*. [48] illustrates the effectiveness of incorporating explicit shape bias for few-shot classification. The method learns a point cloud and image embedding on the base classes using 3D point cloud data and image data, respectively. However, the evaluation includes the nearest centroid classifier on the average of both embeddings of the query images. Unlike Stojanov *et al*. [48], LSFSL does not require any additional synthetic 3D object dataset for FSL. Therefore, LSFSL does not enforce any restrictions on the potential applications of incorporating shape bias.

3. Methodology

We formulate the FSL setup to train few-shot classifiers and then describe our proposed methodology to develop shape-aware few-shot classifiers.

3.1. Preliminaries

Generally, FSL for classification involves two steps: pretraining and meta-testing. In the pretraining step, model $\mathcal{M}$ comprising a backbone feature extractor f_Φ and fully-connected layers g_Θ is trained on base classes N_b from a base dataset $\mathcal{D}_b = \{(x_i, y_i)\}_{i=1}^{B}$. The classification loss function L_{CER} and the parameter regularization term $\Re$

CVPR
#48

CVPR
#48

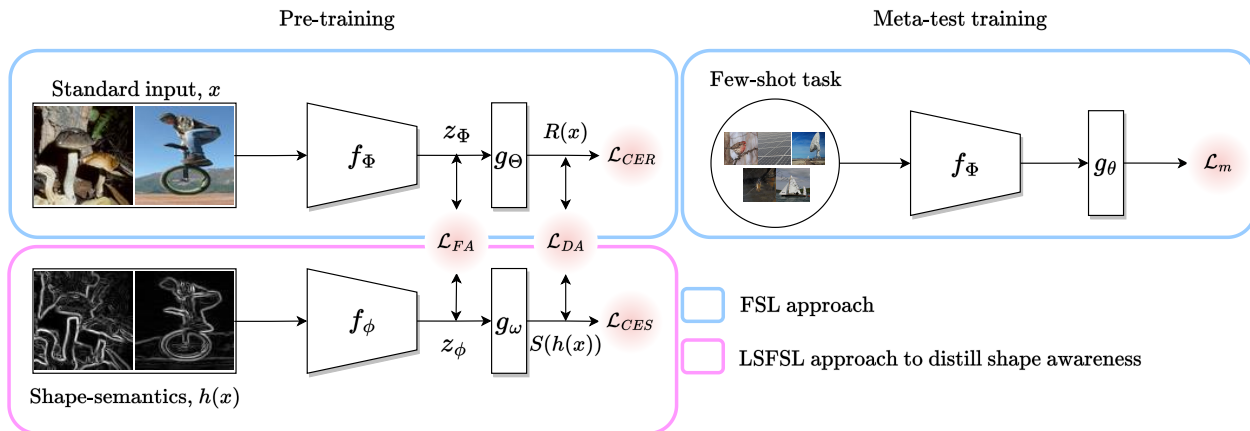CVPR 2023 Submission #48. CONFIDENTIAL REVIEW COPY. DO NOT DISTRIBUTE.

Figure 1. LSFSL pipeline illustrating the pretraining and meta-test training pipeline. The pretraining stage incorporates the shape bias to address the few-shot model bias to texture. The shape semantics, $h(x)$, are obtained by applying a Sobel edge operator on the RGB input image, x. The meta-testing train stage fits a logistic regression on the RGB image features from the RIN model.

used to pretrain the model is given by,

$$\mathcal{M} = \arg\min_{\Phi,\Theta} \mathbb{E}_{\mathcal{D}_b} \left[\mathcal{L}_{CER} \left(\mathcal{D}_b; \Phi, \Theta \right) \right] + \Re(\Phi, \Theta) \quad (1)$$

The pretraining strategy employs meta-learning [6,13,44] or standard supervised (non-meta) learning [41,53] training setup. The meta-testing phase utilizes a dataset $\mathcal{D}_m$ with unseen novel classes to the classes in $\mathcal{D}_b$. The meta-testing phase includes a series of few-shot tasks with data sampled from $\mathcal{D}_m$. Each task contains a meta-test training and meta-test testing step using the support set $S = (\mathcal{D}_i^{train})_{i=1}^N$ and the query set $Q = (\mathcal{D}_i^{test})_{i=1}^N$, respectively. N is the total number of tasks in the meta-testing phase. Each support set contains N_f unique novel classes and k_f examples per class. Therefore, each few-shot task is represented as a N_f-way, k_f-shot task.

In the meta-test train phase, the pretrained model $\mathcal{M}$ acts as a fixed feature extractor utilizing f_Φ or is fine-tuned by updating specific layers or the entire model in meta-testing. The embeddings of f_Φ are used to train a simple classifier g_θ such as a logistic regression or a support vector machine [53]. Additionally, in certain cases, the embeddings are classified using non-parametric classifiers like nearest neighbor by estimating class prototypes [44]. $\mathcal{L}_m$ is the loss function of the meta-test train phase with parameter regularization $\Re$.

$$\Psi = \arg\min_{\Psi} \mathbb{E}_S \left[\mathcal{L}_m \left(\mathcal{D}^{train}; \Psi \right) \right] + \Re(\Psi) \quad (2)$$

where Ψ is given by,

$$\Psi = \begin{cases} [\Phi^T, \theta^T]^T & \text{if both } \Phi \text{ and } \theta \text{ are trainable} \\ [\theta] & \text{if } \theta \text{ is trainable} \\ [\Phi] & \text{if } \Phi \text{ is trainable} \end{cases} \quad (3)$$

The overall objective is to minimize the average test error over the distribution of the meta-test test set. The query set is sampled from the same distribution as the corresponding support set.

$$\mathbb{E}_Q \left[\mathcal{L}_m \left(\mathcal{D}^{test}; \Psi \right) \right] \quad (4)$$

3.2. Incorporating Shape Awareness

With the use of CNNs in majority of classification tasks [46], the models are prone to learn local information [15]. This substantially affects generalization, even with the slightest perturbations, changes in background data, statistical irregularities, or color schemes. This effect is further exacerbated in few-shot settings, as the distribution shift between the training and testing class is more prevalent [4,40]. This calls for a more robust model for few-shot learning. Unlike CNNs, the human visual system is robust and recognizes objects under different conditions using fewer data. Studies have attributed this human behavior to cognitive biases of the brain or to gathered prior knowledge. The presence of cognitive biases helps to focus on the global discriminative shape features for recognition [16,28]. Motivated by this, we propose to impart an additional bias in the form of a shape on top of the generic inductive biases of CNNs at the learning stage.

We propose LSFSL, an approach to develop shape-aware FSL models leveraging the implicit shape information present in the RGB input image. Unlike [48], we do not utilize additional datasets to develop shape-aware models. We incorporate the shape information into the model during the pretraining stage. Our LSFSL model synchronously trains two networks: RIN (standard RGB Input Network) and SIN (Shape Input Network), by distilling shape knowledge between the networks [18]. Each network comprises a backbone feature extractor followed by fully-connected lay-

CVPR
#48

CVPR
#48

CVPR 2023 Submission #48. CONFIDENTIAL REVIEW COPY. DO NOT DISTRIBUTE.

Algorithm 1 LSFSL: Training Algorithm

Input: dataset $\mathcal{D}$, randomly initialized RIN model R with feature extractor f_Φ and classifier g_Θ, randomly initialized SIN model S with feature extractor f_ϕ and classifier g_ω, epochs E, softmax operator σ, stop-grad operator SG, cross-entropy loss CE, Kullback-Leibler divergence loss KLD, mean square error MSE, Sobel edge operator h, feature alignment loss factors (γ_r, γ_r) , decision alignment loss factors (λ_r, λ_s)

1: **for** epoch $e \in \{1, 2, .., E\}$ **do**
2:      sample a mini-batch $(x, y) \sim \mathcal{D}$
3:      $x_{shape} = h(x)$
4:      $z_\Phi = f_\Phi(x)$
5:      $z_\phi = f_\phi(x_{shape})$
6:      $R(x) = g_\Theta(f_\Phi(x))$
7:      $S(h(x)) = g_\omega(f_\phi(h(x_{shape})))$
8:      $\mathcal{L}_{CER} = CE(\sigma(R(x)), y)$
9:      $\mathcal{L}_{CES} = CE(\sigma(S(h(x))), y)$
10:     $\mathcal{L}_{FAR} = MSE(z_\Phi, SG(z_\phi))$ ▷ (Eq. 5)
11:     $\mathcal{L}_{FAS} = MSE(SG(z_\Phi), z_\phi)$ ▷ (Eq. 6)
12:     $\mathcal{L}_{\mathcal{FA}} = \gamma_r\mathcal{L}_{FAR} + \gamma_s\mathcal{L}_{FAS}$ ▷ (Eq. 7)
13:     $\mathcal{L}_{DAR} = KLD(\sigma(R(x))), SG(\sigma(S(h(x))))$ ▷ (Eq. 8)
14:     $\mathcal{L}_{DAS} = KLD(SG(\sigma(R(x))), \sigma(S(h(x))))$ ▷ (Eq. 9)
15:     $\mathcal{L}_{\mathcal{DA}} = \lambda_r\mathcal{L}_{DAR} + \lambda_s\mathcal{L}_{DAS}$ ▷ (Eq. 10)
16:     $\mathcal{L} = \mathcal{L}_{CER} + \mathcal{L}_{CES} + \mathcal{L}_{FA} + \mathcal{L}_{DA}$ ▷ (Eq. 11)
17:     Update parameters of R and S based on $\mathcal{L}$ using Stochastic Gradient Descent (SGD)
18: **end for**
19: **return** RIN model R

---

ers for classification. The SIN network is fed with an image with enhanced shape semantics to extract the shape information pertaining to the data. To extract the shape information, we employ the Sobel edge operator [45], which identifies the shape information through discrete differentiation from the RGB input image in a computationally inexpensive way. As shown in Figure 1, a standard input image x and an image with shape semantics $h(x)$ are passed to the RIN and SIN networks, respectively. LSFSL employs two bias alignments to distill shape information from SIN to RIN. The bias alignments: backbone feature alignment $\mathcal{L}_{FA}$ and decision alignment $\mathcal{L}_{DA}$ force RIN to focus on enhanced shape semantics. This approach of learning the shape information along with the standard biases aids in achieving robust representation in the RIN network by aggregating information from RGB and the edge input image.

The feature embeddings from RIN backbone f_Φ and the SIN backbone f_ϕ for a batch of training images x are represented as z_Φ and z_ϕ, respectively. The Sobel-operated shape image is represented as $h(x)$. Bias alignments are bidirectional, as noticed, SIN requires certain texture information to improve generalization (Equation 6) and vice versa distills

shape knowledge to RIN (Equation 5).

$$\mathcal{L}_{FAR} = \mathbb{E}[\|z_\Phi - \text{stopgrad}(z_\phi)\|_2^2] \tag{5}$$

$$\mathcal{L}_{FAS} = \mathbb{E}[\|\text{stopgrad}(z_\Phi) - z_\phi\|_2^2] \tag{6}$$

Therefore, putting Equation 5 and Equation 6 together, a strict alignment of the backbone features of RIN and SIN is accomplished using mean squared error (MSE) as follow,

$$\mathcal{L}_{FA} = \gamma_r\mathcal{L}_{FAR} + \gamma_s\mathcal{L}_{FAS} \tag{7}$$

where γ_r and γ_s control the influence of texture and shape on the final loss. This alignment of backbone features ensures that the earlier stages of RIN are more shape-aware in the representation space. Therefore, the feature alignment captures the generic feature representations unaffected by changes in the color schemes, perturbations, and backgrounds.

The decision alignment, $\mathcal{L}_{DA}$, is used to align the decision boundary of RIN and SIN. Enhancing the decision boundary of RIN with SIN forces RIN to utilize shape information for classification. Hence, $\mathcal{L}_{DA}$ forces the RIN to be less susceptible to learning from the shortcut cues in the data, such as color schemes and background information. The bidirectionality in decision alignment incorporates Kullback-Leibler divergence $\mathcal{D}_{KL}$ individually for each component, as given in Equation 8 and Equation 9.

$$\mathcal{L}_{DAR} = \mathcal{D}_{KL}(\sigma(R(x))\|\text{stopgrad}(\sigma(S(h(x)))) \tag{8}$$

$$\mathcal{L}_{DAS} = \mathcal{D}_{KL}(\text{stopgrad}(\sigma(R(x)))\|\sigma(S(h(x))) \tag{9}$$

$$\mathcal{L}_{DA} = \lambda_r\mathcal{L}_{DAR} + \lambda_s\mathcal{L}_{DAS} \tag{10}$$

where $R(x)$ and $S(h(x))$ are the output logits from RIN and SIN, respectively. σ is softmax operator. λ_r and λ_s control the influence of distilling shape to RIN and texture to SIN, respectively. Therefore, decision alignment reduces the vulnerability of RIN to learning from superficial cues.

$\mathcal{L}_{CER}$ and $\mathcal{L}_{CES}$ are the cross-entropy loss for classifying the input images by RIN and SIN, respectively. This standard supervision loss improves the generalization of both networks. The overall loss for training shape-aware FSL model using LSFSL is given by,

$$\mathcal{L} = \mathcal{L}_{CER} + \mathcal{L}_{CES} + \mathcal{L}_{DA} + \mathcal{L}_{FA} \tag{11}$$

The training algorithm for the proposed LSFSL approach is provided in Algorithm 1. The meta-testing phase utilizes only the shape-aware RIN model that aggregated shape information in addition to other inductive biases. The model trained by the aforementioned procedure can be enhanced using sequential knowledge distillation [53] (Algorithm 2) or online self-distillation [30] (Algorithm 3). This type of training is a generic procedure and extendable to both non-meta (Section 5) or meta-learning-based models (Section 9).

CVPR
#48

CVPR
#48

CVPR 2023 Submission #48. CONFIDENTIAL REVIEW COPY. DO NOT DISTRIBUTE.

	Model	miniImageNet		tieredImageNet		CIFAR-FS		FC100	
		1-shot	5-shot	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot
Conv block	MAML[1] [13]	$48.70_{\pm 1.84}$	$63.11_{\pm 0.92}$	$51.67_{\pm 1.81}$	$70.30_{\pm 1.75}$	$58.90_{\pm 1.90}$	$71.50_{\pm 1.00}$	$38.10_{\pm 1.70}$	$50.4_{\pm 1.00}$
	PN[2] [44]	$49.42_{\pm 0.78}$	$68.20_{\pm 0.66}$	$53.31_{\pm 0.89}$	$72.69_{\pm 0.74}$	$55.50_{\pm 0.0.70}$	$72.00_{\pm 0.6}$	$35.30_{\pm 0.60}$	$48.60_{\pm 0.60}$
	RN[3] [50]	$50.44_{\pm 0.82}$	$65.32_{\pm 0.70}$	$54.48_{\pm 0.93}$	$71.32_{\pm 0.78}$	$55.00_{\pm 0.10}$	$69.30_{\pm 0.80}$	-	-
	R2D2[4] [7]	$51.20_{\pm 0.60}$	$68.80_{\pm 0.10}$	-	-	$65.30_{\pm 0.20}$	$79.40_{\pm 0.10}$	-	-
WRN-28-10	Boosting [17]	$63.77_{\pm 0.45}$	$80.70_{\pm 0.33}$	$70.53_{\pm 0.51}$	$84.98_{\pm 0.36}$	$73.62_{\pm 0.31}$	$86.05_{\pm 0.22}$	-	-
	Fine-tuning [10]	$57.73_{\pm 0.62}$	$78.17_{\pm 0.49}$	$66.58_{\pm 0.70}$	$85.55_{\pm 0.48}$	$76.58_{\pm 0.68}$	$85.79_{\pm 0.50}$	$43.16_{\pm 0.59}$	$57.57_{\pm 0.55}$
	LEO† [42]	$61.76_{\pm 0.08}$	$77.59_{\pm 0.12}$	$66.33_{\pm 0.05}$	$81.44_{\pm 0.09}$	-	-	-	-
	AWGIM [19]	$63.12_{\pm 0.08}$	$78.40_{\pm 0.11}$	$67.69_{\pm 0.11}$	$82.82_{\pm 0.13}$	-	-	-	-
ResNet-12	TEWAM [35]	$60.07_{\pm na}$	$75.90_{\pm na}$	-	-	$70.4_{\pm na}$	$81.30_{\pm na}$	-	-
	Shot-Free [37]	$59.04_{\pm na}$	$77.64_{\pm na}$	$63.52_{\pm na}$	$82.59_{\pm na}$	-	-	-	-
	MetaOptNet [29]	$62.64_{\pm 0.61}$	$78.63_{\pm 0.46}$	$65.99_{\pm 0.72}$	$81.56_{\pm 0.53}$	$72.60_{\pm 0.70}$	$84.30_{\pm 0.50}$	$41.10_{\pm 0.60}$	$55.50_{\pm 0.60}$
	DSN-MR [43]	$64.60_{\pm 0.72}$	$79.51_{\pm 0.50}$	$67.39_{\pm 0.82}$	$82.85_{\pm 0.56}$	-	-	-	-
	TADAM [33]	$58.50_{\pm 0.3}$	$76.70_{\pm 0.3}$	-	-	-	-	$40.10_{\pm 0.40}$	$56.10_{\pm 0.40}$
	PN [31]	$61.19_{\pm 0.40}$	$76.50_{\pm 0.45}$	-	-	-	-	-	-
	FSL [53]	$62.02_{\pm 0.63}$	$79.64_{\pm 0.44}$	$69.74_{\pm 0.72}$	$84.41_{\pm 0.55}$	$71.50_{\pm 0.8}$	$86.0_{\pm 0.5}$	$42.60_{\pm 0.7}$	$59.10_{\pm 0.6}$
	LSFSL (Ours)	$\mathbf{64.67}_{\pm 0.49}$	$\mathbf{81.79}_{\pm 0.18}$	$\mathbf{71.17}_{\pm 0.52}$	$\mathbf{86.23}_{\pm 22}$	$\mathbf{73.45}_{\pm 0.27}$	$\mathbf{87.07}_{\pm 0.17}$	$\mathbf{43.60}_{\pm 0.11}$	$\mathbf{60.12}_{\pm 0.17}$

Table 1. 5-way few-shot performance in FSL benchmark datasets. The backbone feature extractors are provided in the first column. The superscripts indicate models with convolutional blocks as feature extractors. [1]four Conv blocks with 32 filters each, [2]four Conv blocks with 64 filters each, [3]four Conv blocks with 64-96-128-256 filters, and [4]four Conv blocks with 96-192-384-512 filters. †is trained on train and validation splits. FSL is the baseline RFS model and LSFSL is the shape-distilled RFS model trained using our proposed approach.

4. Experimental Setup

We report results on the following FSL benchmark datasets: miniImageNet [54], tieredImageNet [39], CIFAR-FS [8], and FC100 [33].

miniImageNet incorporates 100 randomly sampled categories from ImageNet and 600 images per category. The train, validation, and test splits of the categories are 64, 16, and 20, respectively.

tieredImageNet encompasses 34 super-classes spanning 608 ImageNet classes, making it more challenging than miniImageNet. The train, validation, and test splits based on the super-class/class combinations are 20/351, 6/97, and 8/160

CIFAR-FS is derived by randomly splitting CIFAR-100 into 64, 16, and 20 classes for the train, validation, and test set, respectively.

FC100 splits CIFAR-100 based on super-classes similar to tieredImageNet. The train, validation, and test splits are 60, 20, and 20, respectively.

We demonstrate the utility of shape awareness instilled by LSFSL by training a prominent FSL method, RFS [53] in the LSFSL framework. In pretraining, the RGB images are passed to RIN after applying random crop, random horizontal flip with probability 0.5, and color jitter with 0.4 brightness factor, contrast factor, and saturation factor. The shape images are generated by applying the Sobel filter to the RGB image upsampled by 2. This is followed by applying a random crop and a random horizontal flip corresponding to the RGB image to the shape input. The crop size for the CIFAR-100 derivatives is 32×32, and that for the ImageNet derivatives is 84×84. For experiments on miniImageNet, CIFAR-FS, and FC100, we train for 65 epochs with 0.1 learning rate decay factor at epoch 60. For experiments on tieredImageNet, we train for 60 epochs with 0.1 learning rate decay after 30, 40, and 50 epochs. All experiments use an SGD optimizer and a ResNet-12 backbone. RFS baseline results are replicated using the implementation details from [53].

In the results and analyses sections followed, the FSL baseline model represents the RFS baseline results, and LS-FSL represents LSFSL enhanced RFS results unless stated otherwise. For LSFSL, the meta-testing train stage for each randomly sampled task involves training a logistic regression classifier using the support RGB image features from the pretrained and fixed backbone. The meta-testing test reports the classification accuracy on 15 query samples per class over 600 randomly sampled tasks. The few-shot classification 5-way accuracies and standard deviations are reported across 3 runs for all experiments unless stated otherwise.

5. FSL Performance

We evaluate the few-shot performance in 5-way 1-shot and 5-way 5-shot settings for all baseline methods and the

CVPR
#48

CVPR
#48

CVPR 2023 Submission #48. CONFIDENTIAL REVIEW COPY. DO NOT DISTRIBUTE.

Model	Shot	miniImageNet	Tinted-miniImageNet				
			Q	S	PT	PT+Q	PT+S
FSL	1	$62.02_{\pm0.63}$	$59.05_{\pm0.74}$	$55.50_{\pm0.67}$	$46.70_{\pm0.27}$	$40.18_{\pm1.03}$	$31.34_{\pm0.06}$
LSFSL		$\mathbf{64.67}_{\pm0.49}$	$\mathbf{62.15}_{\pm0.64}$	$\mathbf{62.53}_{\pm0.46}$	$\mathbf{48.79}_{\pm0.03}$	$\mathbf{45.94}_{\pm0.51}$	$\mathbf{33.39}_{\pm0.19}$
FSL	5	$79.64_{\pm0.44}$	$74.87_{\pm0.09}$	$68.78_{\pm0.23}$	$66.02_{\pm0.25}$	$50.47_{\pm1.13}$	$37.50_{\pm0.44}$
LSFSL		$\mathbf{81.79}_{\pm0.18}$	$\mathbf{77.91}_{\pm0.26}$	$\mathbf{79.24}_{\pm0.16}$	$\mathbf{66.34}_{\pm0.46}$	$\mathbf{58.51}_{\pm0.46}$	$\mathbf{44.99}_{\pm0.50}$

Table 2. Comparison of FSL baseline RFS model and LSFSL-trained counterparts on varying degrees of class-specific tinted-miniImageNet in the 5-way setting. The tints are applied on the query set (Q), support set (S), and only the base dataset in the pretraining stage (PT), pretraining and meta-test training stage (PT + S), pretraining and meta-test test set (PT+Q) in increasing order of presence of spurious correlations.
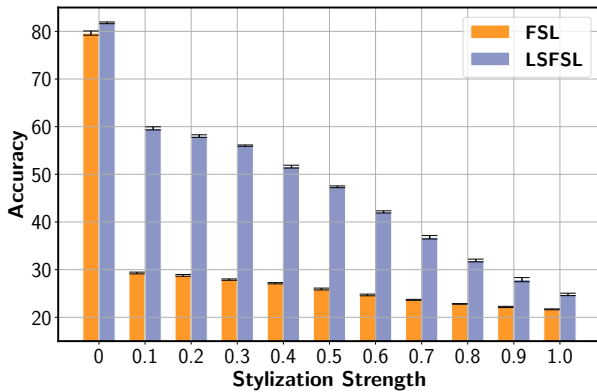

Figure 2. Comparative evaluation of susceptibility of models pretrained and meta-test trained on miniImageNet images to local texture information in the 5-way setting. FSL is the baseline RFS model and LSFSL is the shape-distilled RFS model trained using our proposed approach.

LSFSL model in Table 1. It can be seen that LSFSL, with shape semantics, demonstrates a performance gain of $1-2\%$ over the FSL baseline model as well as other state-of-the-art FSL methods across all datasets, including the more challenging FC100 and tieredImageNet datasets. FSL baseline model containing ResNet-12 backbone and trained by leveraging shape outperforms methods having WRN-28-10 backbone (3 times more parameters than ResNet-12). This indicates that distilling shape semantics to RIN by aligning with the SIN improves overall performance. The shape semantics act as an additional supervision to RIN and improve generalization.

6. Shortcut Learning

Shortcut learning refers to the tendency of DNNs to learn trivial patterns in the data. The decision rules learned in such a scenario substantially impede the generalization of DNNs to real-world scenarios [15]. These patterns can be of various forms: spurious correlations through tints [23], statistical irregularities introduced through Fourier transforms [24], and texture [40]. We look at the impact of learning shape semantics on reducing the susceptibility of DNNs to shortcut learning in FSL. All shortcut learning analyses are on miniImageNet unless mentioned otherwise.

6.1. Texture Bias

Unlike humans, DNNs strongly depend on local texture information to recognize objects [16]. To test the impact of learning shape semantics on texture bias in FSL, we evaluate the models pretrained and meta-test trained with miniImageNet images on texture-biased query images (stylized-miniImageNet), generated following the protocol in [22]. Specifically, we randomly select a texture from available texture pattern images and transfer the query image with texture transfer strength ranging between 0.1 and 1.0. Figure 2 demonstrates that the shape semantics learned by the LSFSL framework help the network look beyond the local texture features and consistently outperform the baseline in all texture transfer strengths.

6.2. Spurious Correlation Analysis

Training data collected in real-world scenarios include various inadvertent superficial cues. DNNs tend to learn the class correlation with these cues and therefore illustrate a limited generalization [15]. We test the ability of LSFSL models trained with shape awareness to resist learning spurious correlations by analyzing the generalization across 5 different scenarios of tints. Table 2 illustrates performance under increasing levels of spurious correlations introduced through class-specific tints, arranged from left to right in ascending order. We observe that the shape-aware LSFSL model significantly outperforms the corresponding baseline without shape awareness in all scenarios, even in the extreme scenarios of spurious correlations, i.e., PT+Q and PT+S. Thus, the shape-aware LSFSL-trained models are less susceptible to spurious correlations in the data.
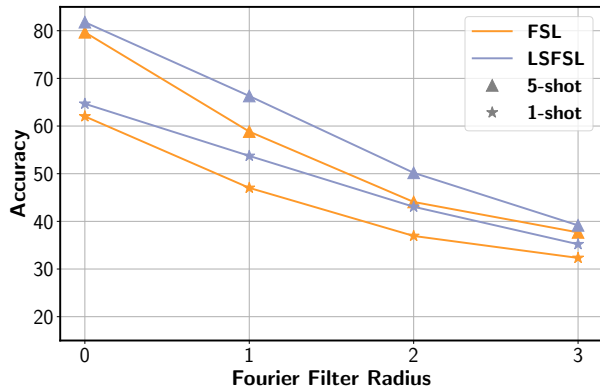
CVPR
#48

CVPR 2023 Submission #48. CONFIDENTIAL REVIEW COPY. DO NOT DISTRIBUTE.

CVPR
#48

Figure 3. Comparative evaluation of susceptibility to statistical regularities of models pretrained and meta-test trained on miniImageNet images in the 5-way setting. The meta-test testing is performed on the Fourier low-pass filtered query images of varying filter radius in the 5-way setting. FSL is the RFS model and LSFSL is the shape-distilled RFS model trained using our proposed approach.

6.3. Statistical Regularity Analysis

Humans rely on high-level abstractions and structure in the data to recognize objects, whereas DNNs tend to learn statistical regularities that affect generalization in challenging testing scenarios [24]. We test the impact of distilling shape as an inductive bias on susceptibility to statistical regularities by evaluating query images after applying a radial low-pass Fourier filter at increasing severities. This filter preserves the visible high-level abstractions for human recognizability, but introduces a superficial statistical regularity by removing high-frequency components. Visualizations of the Fourier transformed images can be found in Appendix A.2. Figure 3 shows that shape-aware LSFSL models are less susceptible to statistical regularities than baseline models that lack shape awareness. We contend that the shape awareness introduced by LSFSL models confronts the vulnerability of DNNs to learn shortcuts and takes a positive step towards the human perception of processing high-level and structural information to recognize objects. Therefore, LSFSL leverages shape semantics to reduce the vulnerability of DNNs to shortcut cues such as texture, spurious correlations, and statistical regularities in FSL.

7. Robustness

Perturbations imperceptible to humans to input images result in adversarial examples that affect the generalization of DNNs and compromise the integrity of intelligent systems in safety-critical real-time applications. This complicates developing robust classifiers using only a few training samples, a scenario desired by many users [49]. In contrast to DNNs,
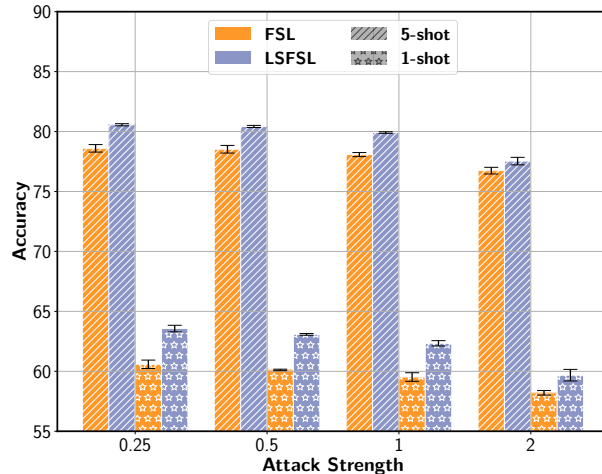

Figure 4. Comparative evaluation of few-shot performance in the 5-way setting on adversarial query images of miniImageNet generated using Square attack at multiple strengths. FSL and LSFSL are the baseline and shape-aware RFS models trained using our proposed approach, respectively.

humans are robust to adversarial examples [38] and also learn to adapt quickly to new tasks with limited samples. This is attributed to cognitive biases in the brain [34]. We study the utility of distilling shape semantics as an inductive bias in generalizing to adversarial examples. The query images are perturbed using a complex black box, Square [1] attack at multiple perturbation strengths. Figure 4 illustrates that LSFSL-trained models are more robust to adversarial query images. Therefore, the shape-aware counterparts of the baseline FSL methods are less prone to adversarial attacks due to learning high-level shape semantics information.

8. Ablation Study

Earlier, we saw that instilling shape bias into the FSL models brings discernible benefits. We investigate the impact of each component of the LSFSL framework individually. Table 3 reports the contribution of each alignment term in Equations (5), (6), (8) and (9) to FSL performance. Each alignment improves the accuracy by a considerable amount. However, the decision and backbone alignment provides a significant few-shot performance boost. Therefore, in addition to improving FSL performance, the alignment objectives help reduce susceptibility to shortcut cues such as texture bias, spurious correlations, and statistical regularities.

9. Generalization to Meta-learning

We illustrated earlier that incorporating shape information using LSFSL increases the generalization and robustness in various few-shot settings over RFS, a non-meta learning-

CVPR
#48

CVPR 2023 Submission #48. CONFIDENTIAL REVIEW COPY. DO NOT DISTRIBUTE.

CVPR
#48

$\mathcal{L}_{CER}$	$\mathcal{L}_{CES}$	$\mathcal{L}_{DAR}$	$\mathcal{L}_{DAS}$	$\mathcal{L}_{FAR}$	$\mathcal{L}_{FAS}$	1-shot	5-shot
✓						$62.02_{\pm 0.63}$	$79.64_{\pm 0.44}$
✓	✓	✓				$63.86_{\pm 0.54}$	$81.07_{\pm 0.29}$
✓	✓		✓			$62.56_{\pm 0.51}$	$80.72_{\pm 0.26}$
✓	✓	✓	✓			$63.90_{\pm 0.55}$	$81.25_{\pm 0.19}$
✓	✓			✓		$63.03_{\pm 0.48}$	$80.86_{\pm 0.29}$
✓	✓				✓	$62.89_{\pm 0.72}$	$81.00_{\pm 0.25}$
✓	✓			✓	✓	$63.44_{\pm 0.59}$	$81.24_{\pm 0.15}$
✓	✓	✓	✓	✓	✓	$\mathbf{64.67}_{\pm 0.49}$	$\mathbf{81.79}_{\pm 0.18}$

Table 3. Ablation study to illustrate the effect of each component of the decision and backbone feature alignments. The study is performed using the miniImageNet dataset in a 5-way setting. RFS baseline FSL model is trained using the proposed LSFSL approach.

Model	miniImageNet		Tinted-miniImageNet	
	1-shot	5-shot	1-shot	5-shot
FSL	$61.19_{\pm 0.40}$	$76.50_{\pm 0.45}$	$25.92_{\pm 0.24}$	$28.67_{\pm 0.14}$
LSFSL	$\mathbf{62.44}_{\pm 0.77}$	$\mathbf{77.62}_{\pm 0.48}$	$\mathbf{29.49}_{\pm 0.18}$	$\mathbf{32.35}_{\pm 0.25}$

Table 4. Comparative evaluation of few-shot performance between conventional PN (indicated as FSL) and PN with shape bias distilled using LSFSL. The results are reported on 5-way setting. tinted-miniImageNet provided below incorporates tints in the query samples whereas the support images are without tint.

based training setup. To evaluate LSFSL utility with meta-learning-based approaches, we considered adding LSFSL to the frequently reported Prototypical Networks (PN) [44] FSL approach. The model is trained for 120 epochs with a learning rate of 0.05 and 1000 tasks per training epoch. The meta-test train estimates the prototypes from the support set image features and classifies using the nearest class prototype in the meta-test testing step. The meta-test test accuracy is reported as an average across 5 runs. Each run incorporates 2000 tasks with 15 query images per class in a task. The FSL performance is reported in Table 4. It is clear that we improve performance and reduce the susceptibility to learning shortcuts in the form of color schemes and texture. Thus, it is evident that our approach can be extended to both meta-learning and non-meta-learning-based FSL settings.

10. Conclusion

We propose an approach, LSFSL, that incorporates one of the cognitive biases to address the susceptibility of few-shot models to spurious cues and texture information. This vulnerability is particularly detrimental in the FSL paradigm compared to conventional learning scenarios. Our shape-aware few-shot classifiers outperform the state-of-the-art methods across multiple FSL benchmarks in both 5-way 5-shot as well as 5-way 1-shot classification tasks. Additionally, the LSFSL approach demonstrates improved robustness against adversarial attacks and reduced sensitivity to shortcut cues, such as local texture information, spurious correlations, and statistical irregularities. Integrating shape semantics as an inductive bias leads to learning higher-level abstractions, thereby promoting both generalization and robustness. Lastly, we demonstrate the compatibility of our pipeline with both meta and non-meta-based FSL approaches. Our work aims to shed light on the untapped capabilities of DNN-based classifiers in low-data settings that are practical for real-world use, and we hope it opens up avenues for further research in this direction.

References

[1] Maksym Andriushchenko, Francesco Croce, Nicolas Flammarion, and Matthias Hein. Square attack: A query-efficient black-box adversarial attack via random search. In Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm, editors, *Computer Vision - ECCV 2020 - 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part XXIII*, volume 12368 of *Lecture Notes in Computer Science*, pages 484–501. Springer, 2020. 7

[2] Elahe Arani, Shruthi Gowda, Ratnajit Mukherjee, Omar Magdy, Senthilkumar Sockalingam Kathiresan, and Bahram Zonooz. A Comprehensive Study of Real-Time Object Detection Networks Across Multiple Domains: A Survey. *Transactions on Machine Learning Research*, 2022. Survey Certification. 1

[3] Vladislav Ayzenberg. Skeletal descriptions of shape provide unique perceptual information for object recognition. *Scientific Reports*, 9:9359, 06 2019. 1

[4] Reza Azad, Abdur R. Fayjie, Claude Kauffmann, Ismail Ben Ayed, Marco Pedersoli, and Jose Dolz. On the Texture Bias for Few-Shot CNN Segmentation. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 2673–2682. IEEE, 2021. 3

[5] Sungyong Baik, Myungsub Choi, Janghoon Choi, Heewon Kim, and Kyoung Mu Lee. Meta-Learning with Adaptive Hyperparameters. In Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin, editors, *Neural Information Processing Systems (NeurIPS)*, 2020. 2

[6] Shawn Beaulieu, Lapo Frati, Thomas Miconi, Joel Lehman, Kenneth O. Stanley, Jeff Clune, and Nick Cheney. Learning to Continually Learn. In Giuseppe De Giacomo, Alejandro Catalá, Bistra Dilkina, Michela Milano, Senén Barro, Alberto Bugarín, and Jérôme Lang, editors, *European Conference on Artificial Intelligence (ECAI 2020) - Conference on Prestigious Applications of Artificial Intelligence (PAIS 2020)*, volume 325 of *Frontiers in Artificial Intelligence and Applications*, pages 992–1001. IOS Press, 2020. 2, 3

[7] Luca Bertinetto, João F. Henriques, Philip H. S. Torr, and Andrea Vedaldi. Meta-learning with differentiable closed-form solvers. In *International Conference on Learning Representations (ICLR)*. OpenReview.net, 2019. 5

[8] Luca Bertinetto, João F. Henriques, Philip H. S. Torr, and Andrea Vedaldi. Meta-learning with differentiable closed-form solvers. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net, 2019. 5

[9] François Chollet. On the measure of intelligence. *CoRR*, abs/1911.01547, 2019. 1

[10] Guneet Singh Dhillon, Pratik Chaudhari, Avinash Ravichandran, and Stefano Soatto. A Baseline for Few-Shot Image Classification. In *International Conference on Learning Representations (ICLR)*. OpenReview.net, 2020. 5

[11] Gil Diesendruck and Paul Bloom. How specific is the shape bias? *Child development*, 74:168–78, 2003. 1

[12] James H. Elder. Shape from contour: Computation and representation. *Annual Review of Vision Science*, 4(1):423–450, 2018. PMID: 30222530. 1

[13] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks. In Doina Precup and Yee Whye Teh, editors, *International Conference on Machine Learning (ICML)*, volume 70 of *Proceedings of Machine Learning Research*, pages 1126–1135. PMLR, 2017. 2, 3, 5

[14] Tommaso Furlanello, Zachary Chase Lipton, Michael Tschannen, Laurent Itti, and Anima Anandkumar. Born-Again Neural Networks. In Jennifer G. Dy and Andreas Krause, editors, *International Conference on Machine Learning ICML*, volume 80 of *Proceedings of Machine Learning Research*, pages 1602–1611. PMLR, 2018. 2

[15] Robert Geirhos, Jörn-Henrik Jacobsen, Claudio Michaelis, Richard S. Zemel, Wieland Brendel, Matthias Bethge, and Felix A. Wichmann. Shortcut learning in deep neural networks. *Nature Machine Intelligence*, 2(11):665–673, 2020. 1, 3, 6

[16] Robert Geirhos, Patricia Rubisch, Claudio Michaelis, Matthias Bethge, Felix A. Wichmann, and Wieland Brendel. ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. In *International Conference on Learning Representations (ICLR)*. OpenReview.net, 2019. 1, 3, 6

[17] Spyros Gidaris, Andrei Bursuc, Nikos Komodakis, Patrick Pérez, and Matthieu Cord. Boosting Few-Shot Visual Learning With Self-Supervision. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 8058–8067. IEEE, 2019. 5

[18] Shruthi Gowda, Bahram Zonooz, and Elahe Arani. Inbiased: Inductive bias distillation to improve generalization and robustness through shape-awareness. *CoRR*, abs/2206.05846, 2022. 3

[19] Yiluan Guo and Ngai-Man Cheung. Attentive Weights Generation for Few Shot Learning via Information Maximization. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, (CVPR)*, pages 13496–13505. Computer Vision Foundation / IEEE, 2020. 5

[20] Bharath Hariharan and Ross B. Girshick. Low-Shot Visual Recognition by Shrinking and Hallucinating Features. In *IEEE International Conference on Computer Vision (ICCV)*, pages 3037–3046. IEEE Computer Society, 2017. 2

[21] Mingcheng Hou and Issei Sato. A closer look at prototype classifier for few-shot image classification. *CoRR*, abs/2110.05076, 2021. 2

[22] Xun Huang and Serge J. Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. *CoRR*, abs/1703.06868, 2017. 6

[23] Saachi Jain, Dimitris Tsipras, and Aleksander Madry. Combining diverse feature priors. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvári, Gang Niu, and Sivan Sabato, editors, *International Conference on Machine Learning, ICML 2022, 17-23 July 2022, Baltimore, Maryland, USA*, volume 162 of *Proceedings of Machine Learning Research*, pages 9802–9832. PMLR, 2022. 6

[24] Jason Jo and Yoshua Bengio. Measuring the tendency of cnns to learn surface statistical regularities. *CoRR*, abs/1711.11561, 2017. 6, 7

[25] Arthur Juliani, Kai Arulkumaran, Shuntaro Sasai, and Ryota Kanai. On the link between conscious function and general intelligence in humans and machines. *Transactions on Machine Learning Research*, 2022. Survey Certification. 1

[26] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012. 1

[27] Patricia K Kuhl. A new view of language acquisition. *Proceedings of the National Academy of Sciences*, 97(22):11850–11857, 2000. 1

[28] Barbara Landau, Linda B. Smith, and Susan S. Jones. The importance of shape in early lexical learning. *Cognitive Development*, 3(3):299–321, 1988. 1, 3

[29] Kwonjoon Lee, Subhransu Maji, Avinash Ravichandran, and Stefano Soatto. Meta-Learning With Differentiable Convex Optimization. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10657–10665. Computer Vision Foundation / IEEE, 2019. 5

[30] Sihan Liu and Yue Wang. Few-shot Learning with Online Self-Distillation. In *IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pages 1067–1070. IEEE, 2021. 2, 4

[31] Xu Luo, Longhui Wei, Liangjian Wen, Jinrong Yang, Lingxi Xie, Zenglin Xu, and Qi Tian. Rectifying the Shortcut Learning of Background: Shared Object Concentration for Few-Shot Image Recognition. *CoRR*, abs/2107.07746, 2021. 2, 5

[32] D. Marr, H. K. Nishihara, and Sydney Brenner. Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of*

CVPR
#48

CVPR 2023 Submission #48. CONFIDENTIAL REVIEW COPY. DO NOT DISTRIBUTE.

CVPR
#48

London. Series B. Biological Sciences, 200(1140):269–294, 1978. 1

[33] Boris N. Oreshkin, Pau Rodríguez López, and Alexandre Lacoste. TADAM: Task dependent adaptive metric for improved few-shot learning. In Samy Bengio, Hanna M. Wallach, Hugo Larochelle, Kristen Grauman, Nicolò Cesa-Bianchi, and Roman Garnett, editors, *Neural Information Processing Systems (NeurIPS)*, pages 719–729, 2018. 5

[34] A. Emin Orhan and Brenden M. Lake. Improving the robustness of ImageNet classifiers using elements of human visual cognition. *CoRR*, abs/1906.08416, 2019. 7

[35] Limeng Qiao, Yemin Shi, Jia Li, Yonghong Tian, Tiejun Huang, and Yaowei Wang. Transductive Episodic-Wise Adaptive Metric for Few-Shot Learning. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 3602–3611. IEEE, 2019. 5

[36] Sachin Ravi and Hugo Larochelle. Optimization as a Model for Few-Shot Learning. In *International Conference on Learning Representations (ICLR)*. OpenReview.net, 2017. 2

[37] Avinash Ravichandran, Rahul Bhotika, and Stefano Soatto. Few-Shot Learning With Embedded Class Models and Shot-Free Meta Training. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 331–339. IEEE, 2019. 5

[38] Manish V. Reddy, Andrzej Banburski, Nishka Pant, and Tomaso A. Poggio. Biologically Inspired Mechanisms for Adversarial Robustness. In Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin, editors, *Neural Information Processing Systems (NeurIPS)*, 2020. 7

[39] Mengye Ren, Eleni Triantafillou, Sachin Ravi, Jake Snell, Kevin Swersky, Joshua B. Tenenbaum, Hugo Larochelle, and Richard S. Zemel. Meta-Learning for Semi-Supervised Few-Shot Classification. *CoRR*, abs/1803.00676, 2018. 5

[40] Sam Ringer, Will Williams, Tom Ash, Remi Francis, and David MacLeod. Texture Bias Of CNNs Limits Few-Shot Classification Performance. *CoRR*, abs/1910.08519, 2019. 1, 2, 3, 6

[41] Mamshad Nayeem Rizve, Salman H. Khan, Fahad Shahbaz Khan, and Mubarak Shah. Exploring Complementary Strengths of Invariant and Equivariant Representations for Few-Shot Learning. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10836–10846. Computer Vision Foundation / IEEE, 2021. 2, 3

[42] Andrei A. Rusu, Dushyant Rao, Jakub Sygnowski, Oriol Vinyals, Razvan Pascanu, Simon Osindero, and Raia Hadsell. Meta-Learning with Latent Embedding Optimization. In *International Conference on Learning Representations (ICLR)*. OpenReview.net, 2019. 5

[43] Christian Simon, Piotr Koniusz, Richard Nock, and Mehrtash Harandi. Adaptive Subspaces for Few-Shot Learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4135–4144. Computer Vision Foundation / IEEE, 2020. 5

[44] Jake Snell, Kevin Swersky, and Richard S. Zemel. Prototypical Networks for Few-shot Learning. In Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett, editors,

Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA, pages 4077–4087, 2017. 2, 3, 5, 8

[45] Irwin Sobel and Gary Feldman. A 3×3 isotropic gradient operator for image processing. *Pattern Classification and Scene Analysis*, pages 271–272, 1973. 4

[46] Yisheng Song, Ting Wang, Subrota K. Mondal, and Jyoti Prakash Sahoo. A Comprehensive Survey of Few-shot Learning: Evolution, Applications, Challenges, and Opportunities. *CoRR*, abs/2205.06743, 2022. 1, 3

[47] Lynn K. A. Sörensen, Sander M. Bohté, Dorina de Jong, Heleen A. Slagter, and H. Steven Scholte. Mechanisms of human dynamic object recognition revealed by sequential deep neural networks. *bioRxiv*, 2022. 1

[48] Stefan Stojanov, Anh Thai, and James M. Rehg. Using Shape To Categorize: Low-Shot Learning With an Explicit Shape Bias. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1798–1808. Computer Vision Foundation / IEEE, 2021. 2, 3

[49] Akshayvarun Subramanya and Hamed Pirsiavash. A simple approach to adversarial robustness in few-shot image classification. *CoRR*, abs/2204.05432, 2022. 7

[50] Flood Sung, Yongxin Yang, Li Zhang, Tao Xiang, Philip H. S. Torr, and Timothy M. Hospedales. Learning to Compare: Relation Network for Few-Shot Learning. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1199–1208. Computer Vision Foundation / IEEE Computer Society, 2018. 5

[51] Saeid Asgari Taghanaki, Kumar Abhishek, Joseph Paul Cohen, Julien Cohen-Adad, and Ghassan Hamarneh. Deep semantic segmentation of natural and medical images: a review. *Artificial Intelligence Review*, 54(1):137–178, 2021. 1

[52] Ran Tao, Han Zhang, Yutong Zheng, and Marios Savvides. Powering Finetuning for Few-shot Learning: Domain-Agnostic Bias Reduction with Selected Sampling. *Association for the Advancement of Artificial Intelligence (AAAI)*, 2022. 2

[53] Yonglong Tian, Yue Wang, Dilip Krishnan, Joshua B. Tenenbaum, and Phillip Isola. Rethinking Few-Shot Image Classification: A Good Embedding is All You Need? In Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm, editors, *European Conference in Computer Vision (ECCV)*, volume 12359 of *Lecture Notes in Computer Science*, pages 266–282. Springer, 2020. 2, 3, 4, 5

[54] Oriol Vinyals, Charles Blundell, Tim Lillicrap, Koray Kavukcuoglu, and Daan Wierstra. Matching Networks for One Shot Learning. In Daniel D. Lee, Masashi Sugiyama, Ulrike von Luxburg, Isabelle Guyon, and Roman Garnett, editors, *Advances in Neural Information Processing Systems 29*, pages 3630–3638, 2016. 5

[55] Yaqing Wang, Quanming Yao, James T. Kwok, and Lionel M. Ni. Generalizing from a Few Examples: A Survey on Few-shot Learning. *ACM Comput. Surv.*, 53(3):63:1–63:34, 2020. 1