Meta-Album: Supplementary Material

Ihsan Ullah[*], Dustin Carrión-Ojeda[*¶‖], Sergio Escalera[#%], Isabelle Guyon[*%], Mike Huisman[+], Felix Mohr[‡], Jan N. van Rijn[+], Haozhe Sun[*], Joaquin Vanschoren[§], Phan Anh Vu[*]
% ChaLearn, USA
‖ hessian.AI, Germany
Universitat de Barcelona, Spain
‡ Universidad de La Sabana, Colombia
¶ Technische Universität Darmstadt, Germany
∗ LISN/CNRS/INRIA, Université Paris-Saclay, France
§ TU/e Eindhoven University of Technology, The Netherlands
+ Leiden Institute of Advanced Computer Science (LIACS), Leiden University, the Netherlands
https://meta-album.github.io/

A Datasets and meta-data

Meta-Album datasets, their original sources and how the datasets are curated are explained separately in the following sections. All these datasets come from 10 domains:

- Large Animals
- Small Animals
- Plants
- Plant Diseases
- Microscopy
- Remote Sensing
- Vehicles
- Manufacturing
- Human Actions
- Optical Character Recognition

We have 3 versions of the datasets:

1. **Micro**: a minimal version with 20 randomly selected classes and 40 images per class;
2. **Mini**: a medium version with all classes having at least 40 images per class, including 40 randomly selected images per class;
3. **Extended**: a full version that consists of all classes and all images per class.

Sample images from Meta-Album are provided in Figure 1. The formatted Meta-Album datasets are referred to as “preprocessed" versions. Such preprocessed images are used in all Meta-Album versions: Micro, Mini, or Extended.

A.1 Large Animals *(LR_AM)*

BRD

When Meta-Album was created, the Birds dataset [26] contained images of 315 bird species, but now it has increased the number of species to 450. It has more than 49 000 images, each with a

36th Conference on Neural Information Processing Systems (NeurIPS 2022) Track on Datasets and Benchmarks.

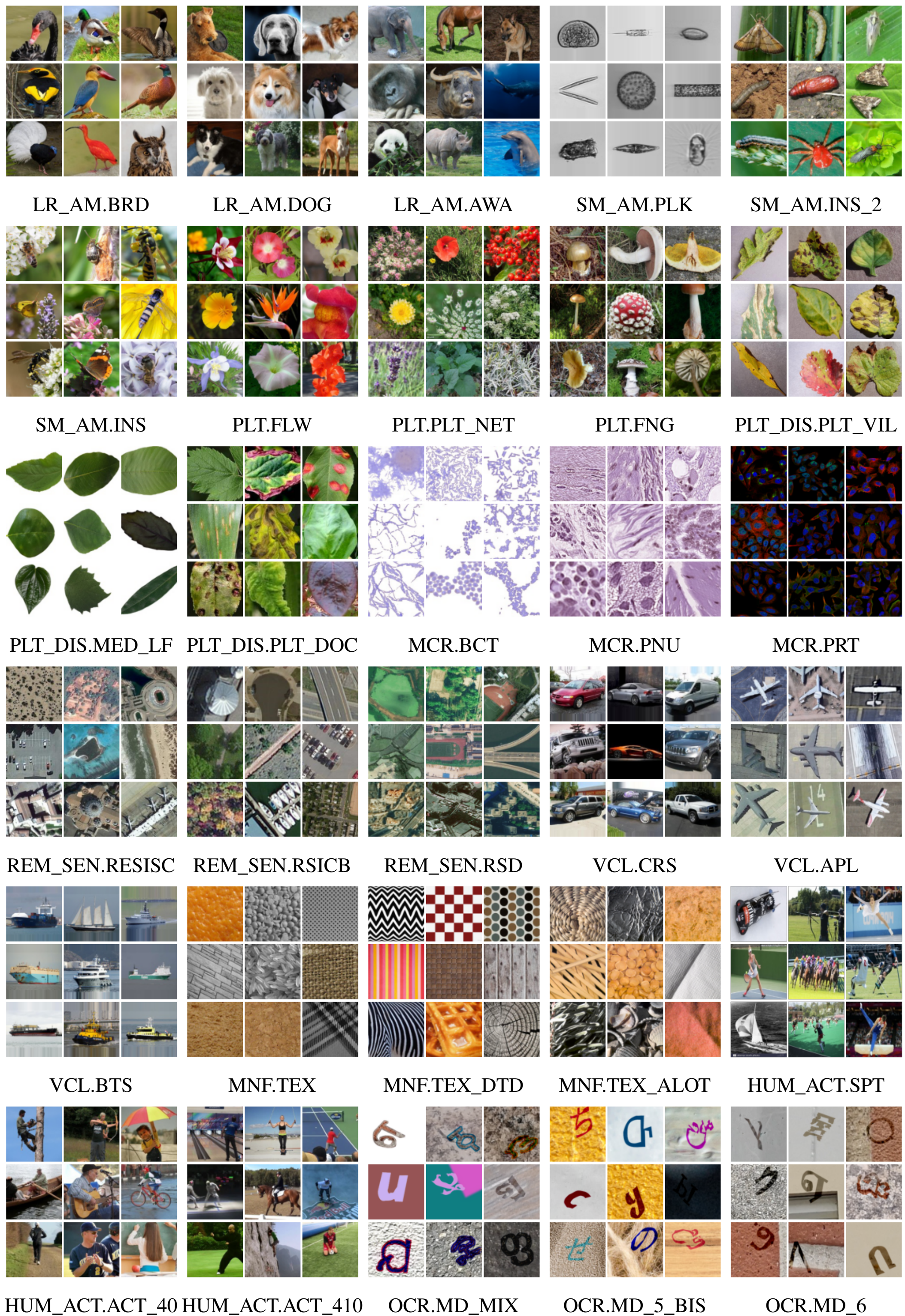


Figure 1: Sample images from Meta-Album datasets.

resolution of 224x224 px. All the images have their natural background, which can lead to bias since, for example, some birds are frequently found in water backgrounds. Additionally, the dataset is imbalanced regarding the ratio of male species images to female species images. The preprocessed version distributed in Meta-Album is made from the original dataset by resizing all the images to a resolution of 128x128 px using an anti-aliasing filter.

DOG

Researchers from Stanford University created the original Dogs dataset [14]. It contains more than 20 000 images belonging to 120 breeds of dogs worldwide. The images and annotations came from ImageNet for the task of fine-grained image categorization. The number of images per class and the resolution is not balanced. Each class can have 148 to 252 images with a resolution from 100x105 to 2 448x3 264 px. This dataset has a little inter-class variation and a large intra-class variation due to color, pose, and occlusion. Most of the images in this dataset are taken in man-made environments leading to a significant background variation. The preprocessed version of this dataset is prepared from the original dataset by cropping the images from either side to make squared images. In case an image has a resolution lower than 128 px, the squared images are done by either duplicating the top and bottom-most 3 rows or the left and right most 3 columns based on the orientation of the original image. These square images are then resized into 128x128 px using an anti-aliasing filter.

AWA

The original Animals with Attributes 2 (AWA) dataset [36] was designed to benchmark transfer-learning algorithms, in particular attribute base classification and zero-shot learning. It has more than 37 000 images from 50 animals, where each animal corresponds to a class. The images of this dataset were collected from public sources, such as Flickr, in 2016, considering only images licensed for free use and redistribution. Each class can have 100 to 1 645 images with a resolution from 100x100 to 1 893x1 920 px. To preprocess this dataset, we cropped the images from either side to make them square. In case an image has a resolution lower than 128 px, the squared images are done by either duplicating the top and bottom-most 3 rows or the left and right most 3 columns based on the orientation of the original image. Lastly, the square images are resized into 128x128 px using an anti-aliasing filter.

A.2 Small Animals *(SM_AM)*

PLK

The Plankton dataset is created by researchers at the Woods Hole Oceanographic Institution. Imaging FlowCytobot (IFCB) was used for the data collection. The Complete process and mechanism are described in the paper [31]. Each image in the dataset contains one or multiple planktons. The images are captured in a controlled environment and have different orientations based on the flow of the fluid in which the images are captured and the size and shape of the planktons. The preprocessed plankton dataset is prepared from the original WHOI Plankton dataset. The preprocessing of the images is done by creating a background squared image by either duplicating the top and bottom-most 3 rows or the left and right most 3 columns based on the orientation of the original image to match the width or height of the image respectively. A Gaussian kernal of size 29x29 is applied to the background image to blur the image. Finally, the original plankton image is pasted on the background image at the center of the image. The squared background image with the original plankton image on top of it as one image is then resized into 128x128 with anti-aliasing.

INS_2

The pest insects dataset [34] was originally created as a large scale benchmark dataset for Insect Pest Recognition. It contains more than 75 000 images belongs to 102 categories. It also has a hierarchical taxonomy and the insect pests which mainly affect one specific agricultural product are grouped into the same upper-level category. The preprocessed version is made from the original dataset by cropping the images in perfect squares and then resizing them into the required images size of 128x128.

INS

The original Insects dataset [29] is created by the National Museum of Natural History, Paris. It has more than 290 000 images in different sizes and orientations. The dataset has hierarchical classes which are listed from top to bottom as Order, Super-Family, Family, and Texa. Each image contains an insect in its natural environment or habitat, *i.e*, either on a flower or near to vegetation. The images are collected by the researchers and hundreds of volunteers from SPIPOLL Science project. The images are uploaded to a centralized server either by using the SPIPOLL website, Android application or IOS application. The preprocessed insect dataset is prepared from the original Insects dataset by carefully preprocessing the images, *i.e.,* cropping the images from either side to make squared images. These cropped images are then resized into 128x128 using Open-CV with an anti-aliasing filter.

A.3 Plants *(PLT)*

FLW

The Flowers dataset [21] consists of a variety of flowers gathered from different websites and some are photographed by the original creators. These flowers are commonly found in the UK. The images generally have large scale, pose and light variations. Some categories of flowers in the dataset has large variations of flowers while other have similar flowers in a category. The dataset was created back in 2008 at Oxford University by Nilsback, M-E. and Zisserman, A.[21]. The Flowers dataset in the Meta-Album meta-dataset is a preprocessed version of the original flowers dataset. The images are first cropped and made into squared images which are then resized into 128x128 with anti-aliasing filter.

PLT_NET

Meta-Album PlantNet dataset is created by sampling the Pl@ntNet-300k dataset, itself a sampling of the Pl@ntNet Project's repository. The images and labels which enter this database are sourced by citizen botanists from around the world, then confirmed using a weighted reliability score from others users, such that each image has been reviewed by 2.03 citizen botanists on average. Of the 1 081 classes in the original Pl@ntNet-300k dataset, PLT_NET retains the 25 most populous classes, belonging to 21 genera, for a total of 120 688 images total, with min 2 914, max 9 011 image distribution per class. Each image contains a colored 128x128 image of a plant or a piece or a plant from the corresponding class (or in some cases sketches of plants or plant cells on microscope slides), scaled from the initial variable width using the INTER_AREA anti-aliasing filter from Open-CV [2]. Almost all images were initially square; cropping by taking the largest possible square with center at the middle of the initial image was applied otherwise.

FNG

Meta-Album Fungi dataset is created by sampling the Danish Fungi 2020 dataset, itself a sampling of the Atlas of Danish Fungi's repository. The images and labels which enter this database are sourced by a group consisting of 3 300 citizen botanists, then verified by their peers using a ranking of each person's reliability, then finally verified by experts working at the Atlas. Of the 128 classes in the original Danish Fungi 2020 dataset, FNG retains the 25 most populous classes, belonging to six genera, for a total of 15 122 images total, with min 372, and max 1 221 images per class. Each image contains a colored 128x128 image of a fungus or a piece of a fungus from the corresponding class. Because the initial data were of widely varying sizes, we needed to crop a significant portion of the images, which we implemented by taking the largest possible square with center at the middle of the initial image. We then scaled each squared image to the 128x128 standard using the INTER_AREA anti-aliasing filter from Open-CV [2].

A.4 Plant Diseases *(PLT_DIS)*

PLT_VIL

The Plant Village dataset [13, 22] contains camera photos of 17 crop leaves. The original image resolution is 256x256 px. This collection covers 26 plant diseases and 12 healthy plants. The leaves are removed from the plant and placed on gray or black background, in various lighting conditions. All images are captured on a variety of gray backgrounds, except Corn Common rust which has a black background. For the curated version, we exclude the irrelevant Background and Corn Common Rust classes from the original collection. Plant Village has a 2-level label hierarchy, the supercategory is the crop type and the category is the disease type. We have preprocessed Plant Village for Meta-Album by resizing a subset from the original dataset to 128x128 image size.

MED_LF

The Medicinal Leaf Database [27] gathers 30 species of healthy and mature medicinal herbs. The leaves are plucked from different plants of the same species, then placed on a white uniform background. There are around 1 800 images in total, captured with a mobile phone camera. The original resolution is 1 600x1 200 px. We create Medleaf for Meta-Album by cropping them at the center and resize to 128x128 px.

PLT_DOC

The PlantDoc dataset [30] is made up of images of leaves of healthy and unhealthy plants. The images were downloaded from Google Images and Ecosia, and later cropped by the authors, so generally, one complete leaf fits in one image. The original, uncropped images are generally different in scale, light conditions, and pose. However, within one category, images of leaves that came from the same original image can be found. The images correspond to 27 classes, including plant disease names and plant species names, *e.g.*: Corn Leaf Blight and Cherry Leaf respectively. The dataset was created for a benchmarking classification model work, published in 2020 by Singh et al. The PlantDoc dataset in the Meta-Album benchmark is extracted from a preprocessed version of the original PlantDoc dataset. First, to get *i.i.d.* samples, only one leaf image per each original image is randomly picked. Then, leaves images are cropped and made into squared images which are then resized into 128x128 with anti-aliasing filter.

A.5 Microscopy *(MCR)*

BCT

The Digital Images of Bacteria Species dataset (DIBaS) [40] is a dataset of 33 bacterial species with around 20 images for each species. For the Meta-Album, since the images were large (2 048x1 532) with very few samples in each class, we decided to split each image into several smaller images before resizing them to 128x128. We then obtained a preprocessed dataset of 4 060 images with at least 108 images for each class. This dataset was also preprocessed with blob normalization techniques, which is quite unusual for this type of image. The goal of this transformation was to reduce the importance of color in decision-making for a bias-aware challenge.

PRT

This dataset is a subset of the Subcellular dataset [33] in the Protein Atlas project. The original dataset, which stems from the Human Protein Atlas Image Classification Kaggle competition, comprises 31 072 RGBY images of size 512x512 px, each of which belongs to one or more out of 28 classes. The labels correspond to protein organelle localizations. For Meta-Album, we performed two modifications: (1), to turn the dataset into a multi-class dataset, we dropped all images belonging to more than a single class and also those images that belong to classes with less than 40 members; (2) we converted the remaining images into RGB simply by dropping the yellow channel; this was also a common practice in the competition. Finally, and as for all datasets in Meta-Album, the images from the original dataset were resized to 128x128 image size.

PNU

The PanNuke dataset [8, 9] is a semi-automatically generated segmentation and classification task of nuclei. The dataset contains 7 753 images of 19 different tissue types. For the Meta-Album meta-dataset, even though this dataset was designed as a segmentation task, we were able to transform it into a tissue classification task since we had the tissue type for each sample in the dataset. We also resized the images to 128x128 pixels and applied stain normalization to avoid bias and remove some spurious features.

A.6 Remote Sensing *(REM_SEN)*

RESISC

RESISC45 dataset [4] gathers 700 RGB images of size 256x256 px for each of 45 scene categories. The data authors strive to provide a challenging dataset by increasing both within-class diversity and between-class similarity, as well as integrating many image variations. Even though RESISC45 does not propose a label hierarchy, it can be created from other common aerial image label organization scheme. We have preprocessed RESISC for Meta-Album by resizing the dataset to 128x128 px.

RSICB

RSICB128 dataset [18] covers 45 scene categories, assembling in total 36 000 images of resolution 128x128 px. The data authors select various locations around the world, and follow China's land-use classification standard. This collection has 2-level label hierarchy with 6 super-categories: agricultural land, construction land and facilities, transportation and facilities, water and water conservancy facilities, woodland, and other lands. The preprocessed version of RSICB is created by resizing the images into 128x128 px using an anti-aliasing filter.

RSD

RSD46 dataset [19, 37] is collected from Google Earth and Tianditu. The collection contains 46 scene categories, with a total of 117 000 images. Each scene category has between 500-3 000 images. The original resolution are 256x256 px or 512x512 px. We have created preprocessed version of RSD for Meta-Album by resizing the original dataset to 128x128 px.

A.7 Vehicles *(VCL)*

APL

The original Airplanes dataset [35] comprises more than 9 000 remote sensing images acquired from Google Earth satellite imagery, including 21 different types of aircraft from around 36 airports. All the images were carefully labeled by seven specialists in the field of remote sensing image interpretation. Each class can have 230 to 846 images, where each image contains only one complete aircraft, and they have variable resolutions. To preprocess this dataset, we cropped the images from either side to make them square, and then we resized them into 128x128 px using an anti-aliasing filter.

CRS

The original Cars dataset [15] was collected in 2013, and it contains more than 16 000 images from 196 classes of cars. Most images are on the road, but some have different backgrounds, and each image has only one car. Each class can have 48 to 136 images of variable resolutions. The preprocess version for this dataset was obtained by creating square images either duplicating the top and bottom-most 3 rows or the left and right most 3 columns based on the orientation of the original image. In this case, cropping was not applied to create the square images since following this technique results in losing too much information from the cars. Then, the square images were resized into 128x128 px using an anti-aliasing filter.

BTS

The original version of the Meta-Album boats dataset is called MARVEL dataset [11]. It has more than 138 000 images of 26 different maritime vessels in their natural background. Each class can have 1 802 to 8 930 images of variable resolutions. To preprocess this dataset, we either duplicate the top and bottom-most 3 rows or the left and right most 3 columns based on the orientation of the original image to create square images. No cropping was applied because the boats occupy most of the image, and applying this technique will lead to incomplete images. Finally, the square images were resized into 128x128 px using an anti-aliasing filter.

A.8 Manufacturing *(MNF)*

TEX

The original Textures dataset is a combination of 4 texture datasets: KTH-TIPS and KTH-TIPS 2 [7, 20], Kylberg Textures Dataset [16] and UIUC Textures Dataset [17]. The data in all four datasets is collected in laboratory conditions, *i.e.,* images were captured in a controlled environment with configurable brightness, luminosity, scale and angle. The *KTH-TIPS* dataset was collected by Mario Fritz and *KTH-TIPS 2* dataset was collected by P. Mallikarjuna and Alireza Tavakoli Targhi, created in 2004 and 2006 respectively. Both of these datasets were prepared under the supervision of Eric Hayman and Barbara Caputo. The data for *Kylberg Textures Dataset* and *UIUC Textures Dataset* data was collected by the original authors of these datasets in September 2010 and August 2005 respectively.

The Meta-Album Textures dataset is a preprocessed version of the original dataset (combination of 4 datasets). All the images are preprocessed by first cropping into perfect squared images and then resized into 128x128 with an anti-aliasing filter.

TEX_DTD

The Textures DTD dataset is a large textures dataset which consists of 5 640 images. The data is collected from Google and Flicker by the original authors of the paper "Describing Textures in the Wild"[5]. The data was annotated using Amazon Mechanical Turk. The data collection process is mentioned on the dataset overview page For Meta-Album meta-dataset, this dataset is preprocessed by cropping the images to square images and then resizing them to 128x128 using Open-CV with an anti-aliasing filter. This dataset has 47 class labels.

TEX_ALOT

Textures ALOT dataset [3] consists of 27 500 images from 250 categories. The images in the dataset are captured in controlled environment by the creators of the dataset. The images have different viewing angle, illumination angle, and illumination color for each material of texture.

A preprocessed version of Textures-ALOT is used in the Meta-Album meta-dataset. The images are first cropped into square images and then resized to 128x128 with anti-aliasing filter.

A.9 Human Actions *(HUM_ACT)*

SPT

The 100-Sports dataset [25] is a collection of sports images covering 73 different sports. Images are 224x224x3 in size and in .jpg format. Images were gathered from internet searches. The images were scanned with a duplicate image detector program and all duplicate images were removed. For Meta-Album, the dataset is preprocessed and images are resized into 128x128 pixels using Open-CV [2]

ACT_40

The Stanford 40 Actions dataset [38] contains images of humans performing 40 actions. There are 9 532 images in total with 180-300 images per action class. The dataset is designed for understanding human actions in still images. For Meta-Album, the dataset is preprocessed and images are resized into 128x128 pixels using Open-CV [2].

ACT_410

The MPII Human Pose dataset [1] is a state of the art benchmark for evaluation of articulated human pose estimation. It includes around 25 000 images containing over 40 000 people with annotated body joints. The images were systematically collected using an established taxonomy of every day human activities. Overall the dataset covers 410 human activities and each image is provided with an activity label. Each image was extracted from a YouTube video. Like other Meta-Album datasets, this dataset is preprocessed and all images are resized into 128x128 pixels.

A.10 Optical Character Recognition *(OCR)*

MD_MIX

OmniPrint-MD-mix dataset consists of 28 240 images (128x128, RGB) from 706 categories. The images are synthesized with OmniPrint [32], and no further processing was done. The OmniPrint synthesis parameters are stated as follows: font size is 192, image size is 128, the strength of random perspective transformation is 0.04, left/right/top/bottom margins are all 20% of the image size, the strength of pre-rasterization elastic transformation is 0.035, random translation is activated both horizontally and vertically, rotation is within -60 and 60 degrees, horizontal shear is within -0.5 and 0.5, brightness is within 0.8333 and 1.2, contrast is within 0.8333 and 1.2, color enhancement is within 0.8333 and 1.2. The other parameters vary between images. We designed 20 settings, each setting is used to synthesize 2 images. The 20 settings are described below:

1. plain white background, random color foreground, trivial image blending (pasting)
2. plain white background, random color foreground, trivial image blending (pasting), random color outline
3. plain white background, random color foreground, trivial image blending (pasting), morphological gradient operation with elliptical kernel (kernel size is 5)
4. plain white background, image/texture foreground, trivial image blending (pasting)
5. plain white background, image/texture foreground, image/texture outline, trivial image blending (pasting)
6. random color background, trivial image blending (pasting), random color foreground
7. random color background, trivial image blending (pasting), random color foreground, random color outline
8. random color background, trivial image blending (pasting), random color foreground, morphological gradient operation with elliptical kernel (kernel size is 5)
9. random color background, trivial image blending (pasting), image/texture foreground
10. random color background augmented with random polygons, trivial image blending (pasting), random color foreground
11. image/texture background, Poisson image blending [23], random color foreground
12. image/texture background, Poisson image blending [23], random color foreground, random color outline
13. image/texture background, Poisson image blending [23], random color foreground, morphological gradient operation with elliptical kernel (kernel size is 5)
14. image/texture background, Poisson image blending [23], image/texture foreground
15. image/texture background, Poisson image blending [23], image/texture foreground, morphological gradient operation with elliptical kernel (kernel size is 5)
16. image/texture background, Poisson image blending [23], image/texture foreground, random color outline
17. image/texture background, Poisson image blending [23], image/texture foreground, random color outline, morphological gradient operation with elliptical kernel (kernel size is 5)
18. image/texture background, Poisson image blending [23], image/texture foreground, image/texture outline
19. image/texture background, Poisson image blending [23], image/texture foreground, image/texture outline, morphological gradient operation with elliptical kernel (kernel size is 5)
20. image/texture background, Poisson image blending [23], image/texture foreground, image/texture outline, outline size is 10

All images/textures consists of photos taken by a personal mobile phone [32].

The 706 categories are characters from:

- Armenian
  - lowercase letters
- ASCII digits
- Balinese
  - consonants
  - digits
- Basic Latin
  - lowercase letters
- Devanagari
  - digits
- Georgian
- Gujarati
  - consonants
- Hebrew
- Hiragana
- khmer
  - consonants
- Mongolian
  - basic letters
- Myanmar
  - digits
- N Ko
  - letters
  - digits
- Oriya
  - consonants
- Russian
- Sinhala
  - independent vowels
- Tamil
  - consonants
  - digits
- Telugu
  - digits
- Thai
  - consonants
- Tibetan
  - digits

MD_5_BIS

OmniPrint-MD-5-bis dataset consists of 28 240 images (128x128, RGB) from 706 categories. The images are synthesized with OmniPrint [32], and no further processing was done. The OmniPrint synthesis parameters are stated as follows: font size is 192, image size is 128, the strength of random perspective transformation is 0.04, left/right/top/bottom margins are all 20% of the image size, the strength of pre-rasterization elastic transformation is 0.035, random translation is activated both horizontally and vertically, image blending method is Poisson Image Editing [23], rotation is within -60 and 60 degrees, horizontal shear is within -0.5 and 0.5, the foreground is filled with a random color, the background consists of images downloaded from Pexels.

The 706 categories are characters from:

- Armenian
  - lowercase letters
- ASCII digits
- Balinese
  - consonants
  - digits
- Basic Latin
  - lowercase letters
- Devanagari
  - digits
- Georgian
- Gujarati
  - consonants
- Hebrew
- Hiragana
- khmer
  - consonants
- Mongolian
  - basic letters
- Myanmar
  - digits
- N Ko
  - letters
  - digits
- Oriya
  - consonants
- Russian
- Sinhala
  - independent vowels
- Tamil
  - consonants
  - digits
- Telugu
  - digits
- Thai

  - consonants
- Tibetan
  - digits

MD_6

OmniPrint-MD-6 dataset consists of 28 120 images (128x128, RGB) from 703 categories. The images are synthesized with OmniPrint [32], no further processing was done. The OmniPrint synthesis parameters are stated as follows: font size is 192, image size is 128, the strength of random perspective transformation is 0.04, left/right/top/bottom margins are all 20% of the image size, the strength of pre-rasterization elastic transformation is 0.035, random translation is activated both horizontally and vertically, image blending method is Poisson Image Editing [23], rotation is within -60 and 60 degrees, horizontal shear is within -0.5 and 0.5, both foreground and background are images taken from a personal mobile phone [32].

The 703 categories are characters from:

- Arabic
- Armenian
  - uppercase letters
  - lowercase letters
- Balinese
  - independent vowels
- Basic Latin
  - uppercase letters
- Bengali
  - consonants
  - digits
  - independent vowels
- punctuation symbols
- Devanagari
  - consonants
  - independent vowels
- Ethiopic
  - digits
- Greek
- Gujarati
  - digits
  - independent vowels
- Katakana
- Khmer
  - digits
  - independent vowels
- Lao
  - consonants
  - digits
- Mongolian
  - digits
- Myanmar

    - consonants
    - independent vowels
- Oriya
    - digits
    - independent vowels
- Sinhala
    - astrological digits
    - consonants
- Tamil
    - independent vowels
- Telugu
    - consonants
    - independent vowels
- Thai
    - digits
- Tibetan
    - consonants

A.11 Meta-data

Meta-Album datasets have the following meta-data files:

- **labels.csv** : The meta-data in *labels.csv* consists of the *FILE_NAME*, *CATEGORY* and *SUPER_CATEGORY* (if any). In the case of OCR datasets, it contains some extra information about the images of characters in the data, *e.g.,* shear, stretch, rotation, font, etc.
- **info.json** : It consists of meta-data (dataset name, description, number of categories and super-categories, column names, etc.) which is useful in reading the data, specially for data quality control. It also gives important statistics about the dataset.
- **DATASET_info.json** : This file is a detailed overview of the dataset which includes:
    - Meta-Album ID
    - Domain ID
    - Domain name
    - Dataset ID
    - Dataset name
    - Dataset description
    - Data format
    - Image size
    - License name
    - License URL
    - Source
    - Source URL
    - Original author
    - Original contact
    - Citation
    - Meta-Album dataset creator
    - Created data
    - Contact details
    - Meta-Album website
    - Dataset download links (Exntended, Mini, Micro)
    - Data statistics (Exntended, Mini, Micro)

B License information of Meta-Album datasets

In this section, we provide details about the licenses of all Meta-Album datasets.

We researched the ownership of all datasets to determine the type of usage permission. We contacted original owners by email to clarify permissions if needed, when no explicit statement of license or permission was found.

Unless otherwise stated, all data are re-distributed by us under a non-commercial license. When no original license could be identified, we use CC BY NC 4.0 by default. For commercial use, users should contact the original owners, whose contact information is found in the datasheets, in Appendix A and on Meta-Album website (`https://meta-album.github.io/`).

Table 1: Meta-Album: Datasets license information

Domain Name	Set #	Dataset ID	License	Original source
Large Animals	0	*BRD*	CC0 Public Domain	Birds 400 [26]
	1	*DOG*	CC BY-NC 4.0	Stanford Dogs [14]
	2	*AWA*	Creative Commons	AWA [36]
Small Animals	0	*PLK*	MIT License	WHOI [31]
	1	*INS_2*	CC BY-NC 4.0	Pest Insects [34]
	2	*INS*	CC BY-NC 2.0	SPIPOLL [29]
Plants	0	*FLW*	GNU General Public License Version 2	Flowers [21]
	1	*PLT_NET*	Creative Commons Attribution 4.0 International	PlantNet [10]
	2	*FNG*	BSD-3-Clause License	Danish Fungi [24]
Plant Diseases	0	*PLT_VIL*	CC0 1.0	PlantVillage [13, 22]
	1	*MED_LF*	CC BY 4.0	Medicial Leaf [27]
	2	*PLT_DOC*	Creative Commons Attribution 4.0 International	Plant Doc [30]
Microscopy	0	*BCT*	CC BY-NC 4.0	DiBas [40]
	1	*PNU*	Attribution-NonCommercial-ShareAlike 4.0 International	PanNuke [8, 9]
	2	*PRT*	CC BY-SA 3.0	Protein Atlas [33]
Remote Sensing	0	*RESISC*	CC-BY-NC 4.0	RESISC45 [4]
	1	*RSICB*	CC BY-NC 4.0	RSICB128 [18]
	2	*RSD*	CC BY-NC 4.0	RSD46 [37, 19]
Vehicles	0	*CRS*	ImageNet License	Cars [15]
	1	*APL*	Creative Commons Attribution 4.0 International	Multi-type Aircraft [35]
	2	*BTS*	CC BY-NC 4.0	MARVEL [11]
Manufacturing	0	*TEX*	CC BY-NC 4.0	KTH-TIPS [7, 20] Kylberg [16] UIUC [17]
	1	*TEX_DTD*	CC BY-NC 4.0	Texture DTD [5]
	2	*TEX_ALOT*	CC BY-NC 4.0	Texture ALOT [3]
Human Actions	0	*SPT*	CC0 1.0 Public Domain	100 Sports [25]
	1	*ACT_40*	CC BY-NC 4.0	Stanford 40 Actions [38]
	2	*ACT_410*	Simplified BSD License	MPII Human Pose [1]
Optical Char. Recog.	0	*MD_MIX*	CC BY 4.0	
	1	*MD_5_BIS*	CC BY 4.0	OmniPrint [32]
	2	*MD_6*	CC BY 4.0	

C Data preparation

Data preparation process is done in many steps which include, depending on the dataset:

- Visual inspection and data cleaning
- Cropping images to a square around the region of interest (eventually background padding)
- Resizing squared images to 128x128 size (with anti-aliasing filter)
- Sub-sampling (to obtain the Mini and Micro versions).

C.1 Visual inspection and data cleaning

The goal of this step is to remove non-suitable images.

Most Meta-Album datasets were downloaded from the Internet; a few were directly obtained from their original owners/creators. Cleaning steps included the removal of duplicate images, corrupted images, and images that included watermarks or identifiers. After cleaning, each dataset is saved in a folder including a sub-folder with the images, in a standard jpg format, and a meta-data file including file names, categories/classes/labels, super-categories, etc.

C.2 Image cropping

The goal of the step is to obtain square images.

Most of the datasets come with images of different sizes and resolutions. Usually, the object of interest (*e.g.,* bird, leaf, car, etc.) is already mostly centered in the middle of the image, with comfortable margins on all sides. In this case, the image is cropped to obtain a square by keeping the smallest dimension (width or height) and eliminating symmetrically spurious margins in the largest dimension.

Some datasets, *e.g.,* Plankton and Boats, are preprocessed in a different way because the rectangular images tightly fit the object of interest. In that case, the images are padded with extra background or either side in the smallest dimension, to obtain a square of the largest dimension. The padding is obtained from the 3 extreme rows (or columns) on either side, which is smoothed with a Gaussian kernel of size 29x29. We found that this technique minimally introduces image artifacts.

Some of the datasets, *e.g.,* RESISC and Textures datasets, are already square, hence the only preprocessing needed is resizing (see next sub-section). All OCR datasets are generated by the OmniPrint software [32]. The images are exactly in 128x128 size and hence no preprocessing is required at all.

Datasets in Human Actions domain are preprocessed in a different way as they contain humans performing some actions, *e.g.,* playing a sport, etc. For these datasets, a face detection software is used to detect a face in the image and then a reasonable area around the face is cropped in order to make sure that the human body along with the face is cropped. After cropping, like other datasets, the images are resized into 128x128 pixel sizes.

See Figure 2 for sample raw images from some of the datasets.

C.3 Image resizing

The goal of this step is to reduce the image size and obtain images of the same dimension for all datasets.

Except for the OCR datasets (images already in the correct size), all other dataset images are resized to 128x128 using Open-CV [2], with an anti-aliasing filter. This avoids the aliasing effect, often visible in the form of jagged edges, introduced by the loss of information during down-sampling.

The image resolution 128x128 is chosen based on preliminary experiments and analyses done for different resolutions of images on the insect dataset, presenting fine features necessary for recognition to the human eye: 64x64, 128x128 and 256x256. For this experiment (See Figure 3), a ResNet 152V2 architecture from Keras is used. It can be observed that 256x256 gives the highest scores among the observed resolutions but the dataset size also increases by a factor of 2. The images with resolution 128x128 are observed to be clearly recognizable to human eyes.

(a) Insects (b) Plankton

(c) Flowers (d) Texture DTD

Figure 2: Meta-Album datasets raw sample images.

Another experiment on the insect dataset, comparing 64x64 and 128x128 resolution, is also carried out. The number of images used for the experiment is 180 000, 135 000 of which are used for training, and the remaining for validation. The accuracy plot in Figure 4 shows clearly that 128x128 produces much better results. The experimental settings for this experiment include: ResNet-18 architecture, SGD optimizer, learning rate of 0.001 and 10 epochs. The better quality of 128x128 images is also clearly visually observable (See Figure 5).

C.4 Sub-sampling

The goal of this step is to keep only classes sufficiently populated (more than 40 examples per class) and to balance the dataset (even number of examples per class).

After the previous steps, we obtain what we call the "Extended version" of Meta-Album. The preprocessed datasets in this version conserve the proportions of samples of the original data, with different number of classes and images per class.

To facilitate experimentation and reduce the need for large computational resources, we have prepared the Mini version, including all classes with at least 40 images/class, but only 40 randomly chosen images from each class. We have also prepared the Micro version, including only 20 randomly selected classes from the Mini version, with 40 images per class.

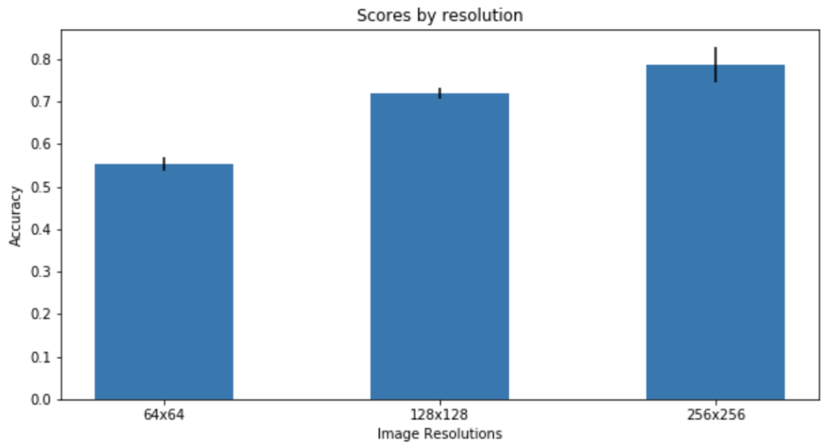


Figure 3: Image resolution comparison (64x64 vs 128x128 vs 256x256).

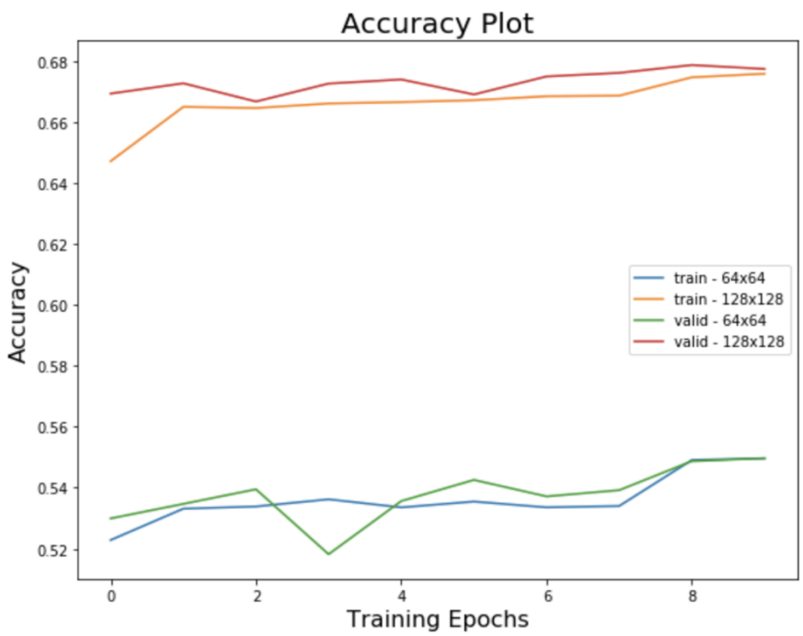


Figure 4: Image resolution comparison (64x64 vs 128x128 vs 256x256).

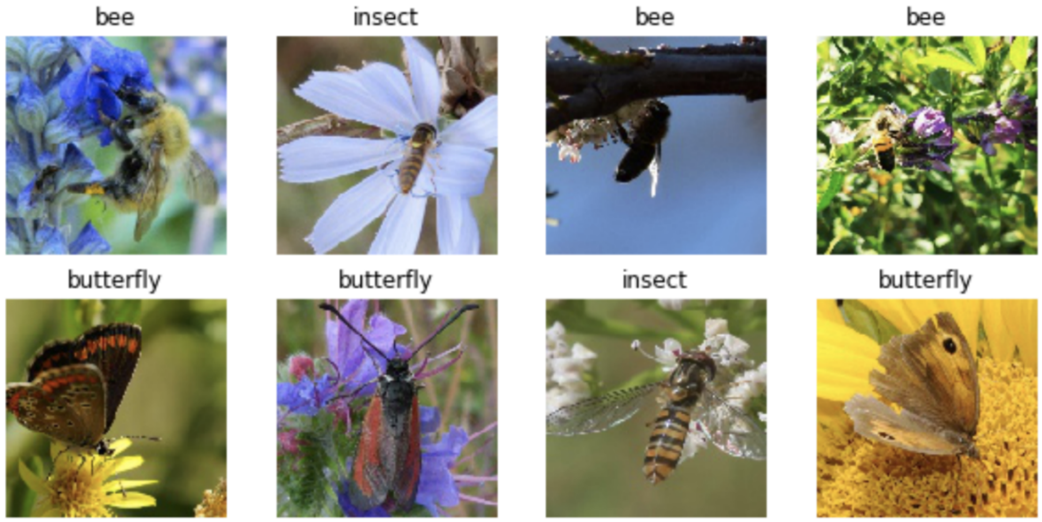
The software to prepare the datasets, *i.e.,* add backgrounds, crop images, resize images and extract a subset from a dataset, is available in the Meta-Album GitHub Repository (`https://github.com/ihsaan-ullah/meta-album`).

C.5 Quality control

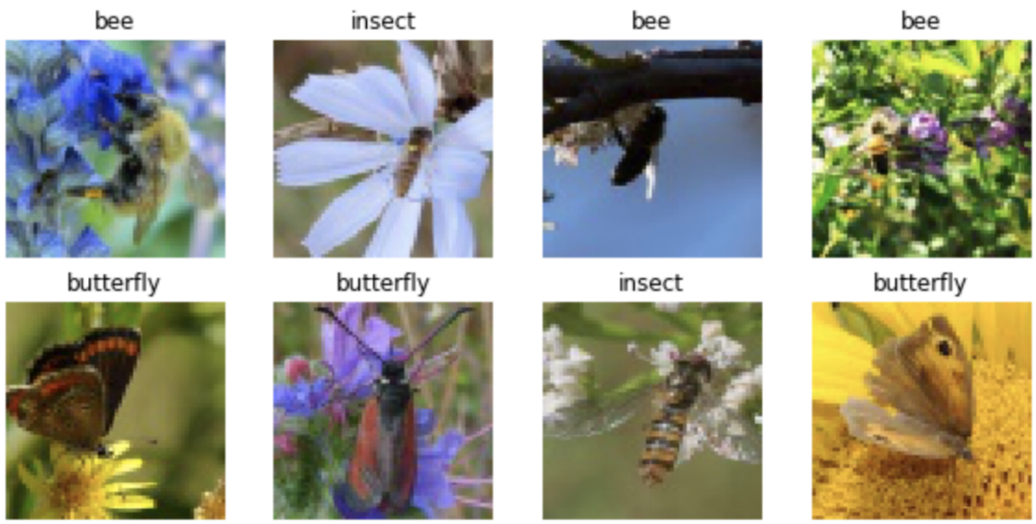
In order to ensure a high quality of datasets, we put a strong emphasis on quality control. We describe two ways to ensure this quality: factsheets and GradCam visualizations. We also define an easy to read Data format for all our datasets.

Data format

In order to generate Factsheets for quality control, the data is first formatted in the recommended Data format. This makes the data easy-to-read and understand. The data format has a root directory



(a) Sample 128x128 images



(b) Sample 64x64 images

Figure 5: Sample images in 128x128 and 64x64 resolutions.

(ideally with the dataset name and it consists of an *images* folder with all the required images for the fact-sheet generation, a *labels.csv* file with at least two columns, *i.e., FILE_NAME* and *CATEGORY* (category is synonym of class/label) which shows the name of the image instance and the label of that image respectively. One more important file is a *info.json* file in the root directory, which has important information about the dataset and how the data should be read for factsheets. The Data format repository has sample dataset for reference. To verify the data format, a python script is also provided and complete details of the formatting and testing are given in the README of the data format repository.

Factsheets

For each dataset, we generate a factsheet which is an automatically generated report showing basic information and experimental results on the dataset. These experiments are run on the formatted datasets (Mini version). The experiments are flexible and the configuration can be changed easily by changing the parameters of the software. For each dataset, we train a randomly initialized ResNet-18 architecture [12] on various 5-way 20-shot classification tasks, *i.e.,* the network is presented with training sets consisting of 20 images for each of the 5 classes. The remaining 20 images per class are used as query set to measure the generalization ability of the network. Note that the network is trained "from scratch" (randomly initialized weights), on every task. From the experiments, we construct ROC curves for every task for all classes.

Classes that are too hard to separate (indicating non-learnable tasks) can be identified easily with this experiment. AUC score below 0.7 indicates the difficulty in separation and hence could be studied further to construct a fair task.

In addition to this automatic quality check, we also generate Imagesheets for the datasets, which are PDF files with images of every class. This way, we can also check the data quality visually.

GradCam visualization

In order to detect potential bias stemming from the data and our curating process, we inspect where the baseline model focuses with GradCAM [28]. GradCAM is a post-hoc attention method, showing the contribution of each input element to the output.

By displaying the activation map on top of the original input image, we can observe the impact of each pixel on the prediction output. As such, we can verify whether the model focuses on the relevant subject. In this way, it can also be used to detect whether artifacts are introduced by the image processing in the picture or on the background and other artifacts. For example, we use GradCAM for detecting any artifacts introduced by preprocessing the backgrounds of plankton images to make squared images using padding (Figure 6). Other methods should also be considered to diagnose data [39].

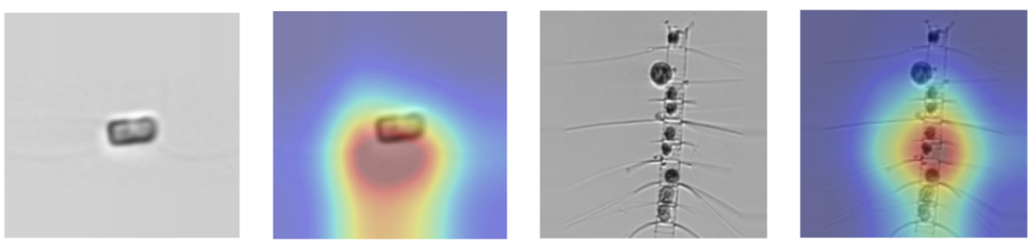
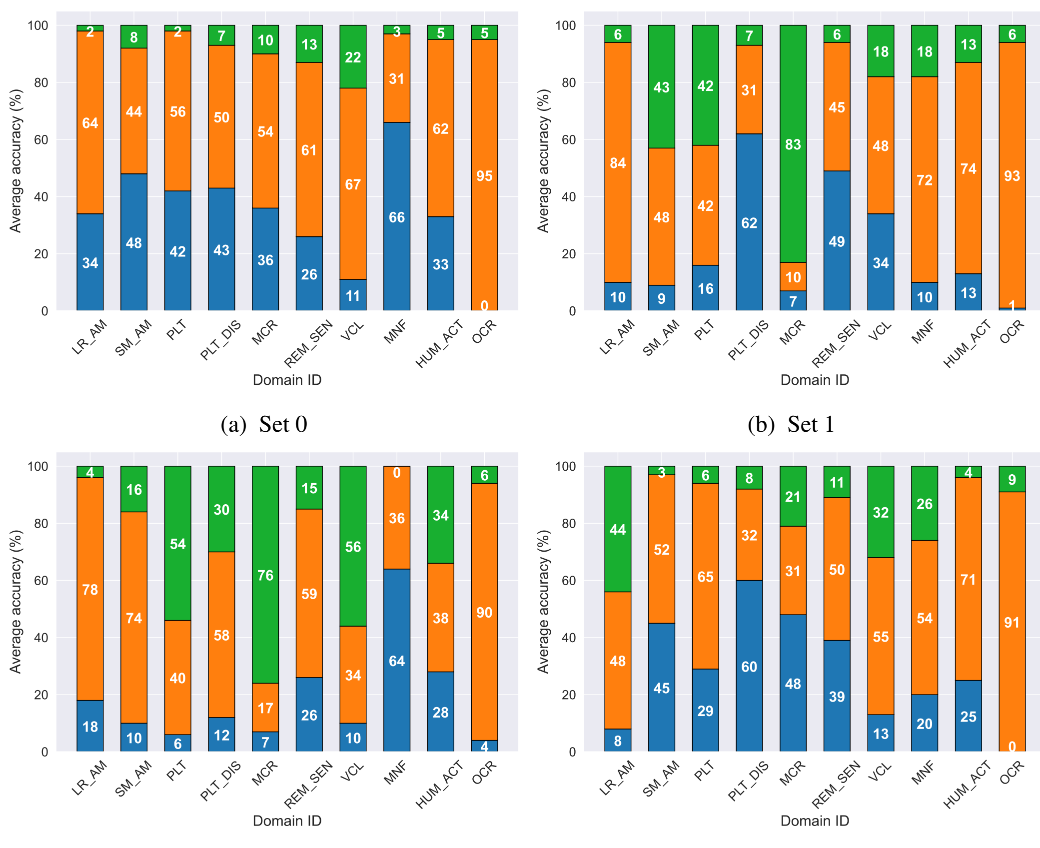


Figure 6: GradCAM activation maps to detect artifacts that could be introduced by image preprocessing.

D Dataset difficulty analysis

In this section, we analyze the difficulty per domain of each set of Meta-Album datasets following the NeurIPS 2021 MetaDL challenge protocol (within domain setting). In Figure 7a, Figure 7b, Figure 7c, and Figure 7d, the top of the blue bar indicates the worst baseline performance (TrainFromScratch). The top of the orange bar indicates the performance of the winners of the NeurIPS 2021 challenge (MetaDelta++) [6]. The top of the green bar indicates the maximum achievable performance. The larger the green bar, the larger the *intrinsic difficulty*. The larger the orange bar, the larger the *modeling difficulty*.



(a) Set 0

(b) Set 1

(c) Set 2

(d) Set 3

Figure 7: Difficulty analysis for 4 sets (40 datasets) of Meta-Album. Each set consists of 10 datasets from 10 domains (one dataset per domain).

Experimental protocol. Each dataset was evaluated independently by splitting it into meta-training, meta-validation, and meta-testing sets with non-overlapping classes, as explained in Section 3.1. Then, for each dataset, the MetaDelta++ method was meta-trained for 30 minutes with batches of size 64. The trainers produced by MetaDelta++ were meta-validated every 50 batches on 50 5-way 5-shot tasks drawn from the corresponding meta-validation split. On the other hand, as explained in Section 3.2, the TrainFromScratch method has no meta-training and meta-validation phases; instead, it learns each task at meta-test time. Lastly, the learning algorithm with the best meta-validation performance was meta-tested on 100 5-way 5-shot tasks randomly sampled from the meta-testing split. The query set of each meta-test task contained 20 images per class. As described in Section 3.2, we averaged the results over 3 runs with different random seeds.

The experimental results show that in some domains like OCR, there is always a large modeling difficulty, as evidenced by the large orange bar, which makes them desirable to be used in challenges. On the other hand, the microscopy datasets (MCR) in Sets 1 and 2 have a large intrinsic difficulty (large green bar): even the winner's method does not get good results which may indicate that the data of these datasets are of insufficient quality. Moreover, some datasets are relatively easy even for the TrainFromScratch method, such as the manufacturing datasets (MNF) of Set 0 or 2. Finally, on average, Set 0 is the easiest while the hardest is Set 2.

E Within Domain Few-shot learning: additional results

In this section, we show additional results for the within domain few-shot learning experiments that were presented in Section 3.2. Table 2, Table 3, Table 4, and Table 5 display the average accuracy per technique and dataset in the [1, 5, 10, 20]-shot settings, respectively. Note that the 1-shot performance of matching networks and prototypical networks is not the same as they use different distance measures. More specifically, matching networks use cosine similarity while prototypical networks use squared euclidean distance.

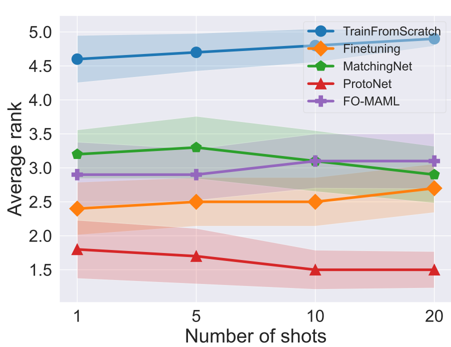
Table 6 displays the average ranks and running times per technique in different within domain settings. The rankings follow roughly the same pattern as in the right subplot of Figure 8. Note that the confidence intervals for the average ranks are larger than for the average accuracy as the former is computed over the 30 datasets, while the latter is computed at per-task level over all datasets. Figure 8 displays the average rank per method and average accuracy for within domain 5-way [1, 5, 10, 20]-shot classification.

Table 2: Full performance results for 5-way 1-shot image classification on all datasets. The 95% confidence intervals are computed at per-task level over 3 runs per dataset with 600 tasks per run (total tasks: $3 \times 600 = 1\,800$).

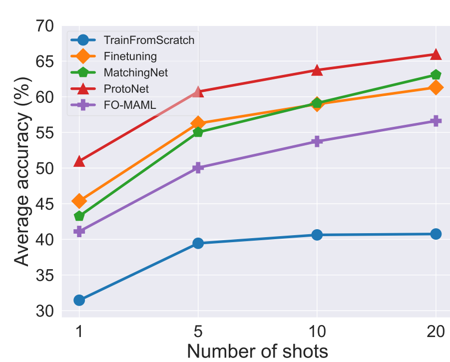
Domain	Dataset ID	TrainFromScratch	Finetuning	MatchingNet	ProtoNet	FO-MAML
Large animals	BRD	31.4 ± 0.37	62.7 ± 0.51	52.7 ± 0.59	67.7 ± 0.53	60.0 ± 0.58
	DOG	24.6 ± 0.27	31.7 ± 0.36	30.2 ± 0.42	33.6 ± 0.40	29.5 ± 0.38
	AWA	26.5 ± 0.31	32.5 ± 0.35	32.3 ± 0.42	31.7 ± 0.36	29.7 ± 0.42
Small animals	PLK	44.5 ± 0.53	60.0 ± 0.57	55.8 ± 0.60	66.7 ± 0.59	59.1 ± 0.58
	INS_2	23.8 ± 0.28	27.4 ± 0.31	26.8 ± 0.36	28.2 ± 0.35	24.7 ± 0.31
	INS	22.6 ± 0.25	34.5 ± 0.40	35.0 ± 0.49	36.1 ± 0.45	31.8 ± 0.51
Plants	FLW	35.4 ± 0.44	59.7 ± 0.51	48.3 ± 0.56	63.0 ± 0.49	55.0 ± 0.57
	PLT_NET	24.8 ± 0.29	31.5 ± 0.35	33.9 ± 0.49	30.5 ± 0.37	35.2 ± 0.41
	FNG	23.8 ± 0.25	23.8 ± 0.25	22.4 ± 0.24	23.7 ± 0.26	22.9 ± 0.26
Plant diseases	PLT_VIL	49.5 ± 0.47	69.0 ± 0.39	50.0 ± 0.46	64.1 ± 0.41	55.6 ± 0.42
	MED_LF	57.8 ± 0.46	63.0 ± 0.42	60.3 ± 0.39	68.4 ± 0.46	66.7 ± 0.43
	PLT_DOC	21.5 ± 0.23	23.1 ± 0.25	22.9 ± 0.26	24.1 ± 0.26	23.8 ± 0.27
Microscopy	BCT	35.3 ± 0.41	65.7 ± 0.45	70.0 ± 0.53	75.5 ± 0.43	58.7 ± 0.68
	PNU	24.2 ± 0.29	23.0 ± 0.24	23.3 ± 0.25	22.9 ± 0.25	24.2 ± 0.27
	PRT	22.0 ± 0.24	25.3 ± 0.27	27.9 ± 0.36	26.0 ± 0.30	28.0 ± 0.37
Remote sensing	RESISC	31.9 ± 0.39	44.0 ± 0.44	37.4 ± 0.46	42.2 ± 0.43	42.1 ± 0.50
	RSICB	47.0 ± 0.47	62.9 ± 0.46	50.2 ± 0.48	64.2 ± 0.46	60.2 ± 0.50
	RSD	33.7 ± 0.39	43.7 ± 0.44	45.1 ± 0.58	49.9 ± 0.46	51.6 ± 0.50
Vehicles	CRS	23.0 ± 0.25	48.9 ± 0.48	52.4 ± 0.61	63.5 ± 0.56	53.1 ± 0.64
	APL	29.9 ± 0.31	36.4 ± 0.39	34.0 ± 0.40	41.2 ± 0.43	33.1 ± 0.37
	BTS	23.6 ± 0.25	26.5 ± 0.28	23.8 ± 0.29	26.2 ± 0.29	23.7 ± 0.28
Manufacturing	TEX	59.8 ± 0.59	81.2 ± 0.46	72.8 ± 0.68	84.1 ± 0.50	84.5 ± 0.51
	TEX_DTD	24.4 ± 0.28	28.6 ± 0.34	26.3 ± 0.29	27.3 ± 0.32	27.6 ± 0.31
	TEX_ALOT	57.0 ± 0.59	89.8 ± 0.38	93.1 ± 0.38	96.0 ± 0.23	93.5 ± 0.35
Human actions	SPT	31.3 ± 0.41	45.4 ± 0.46	42.8 ± 0.57	48.2 ± 0.50	47.4 ± 0.55
	ACT_40	22.2 ± 0.23	24.8 ± 0.25	25.5 ± 0.30	24.7 ± 0.26	25.2 ± 0.30
	ACT_410	30.6 ± 0.36	33.2 ± 0.36	24.9 ± 0.29	27.6 ± 0.30	25.5 ± 0.30
OCR	MD_MIX	20.3 ± 0.19	40.8 ± 0.56	35.4 ± 0.78	89.0 ± 0.37	20.0 ± 0.15
	MD_5_BIS	20.2 ± 0.20	42.6 ± 0.44	69.6 ± 0.66	90.8 ± 0.35	20.2 ± 0.20
	MD_6	21.2 ± 0.22	79.7 ± 0.58	72.3 ± 0.65	92.1 ± 0.31	20.3 ± 0.21

Table 3: Full performance results for 5-way 5-shot image classification on all datasets. The 95% confidence intervals are computed at per-task level over 3 runs per dataset with 600 tasks per run (total tasks: $3 \times 600 = 1\,800$).

Domain	Dataset ID	TrainFromScratch	Finetuning	MatchingNet	ProtoNet	FO-MAML
Large animals	BRD	45.4 ± 0.43	80.5 ± 0.36	72.0 ± 0.41	82.4 ± 0.33	73.7 ± 0.42
	DOG	30.0 ± 0.31	40.1 ± 0.36	42.9 ± 0.39	41.8 ± 0.37	34.9 ± 0.38
	AWA	35.5 ± 0.32	42.6 ± 0.36	37.2 ± 0.33	45.2 ± 0.38	37.2 ± 0.36
Small animals	PLK	57.8 ± 0.50	73.9 ± 0.50	71.0 ± 0.50	76.7 ± 0.44	69.6 ± 0.49
	INS_2	28.3 ± 0.31	34.4 ± 0.32	29.1 ± 0.29	32.2 ± 0.31	31.4 ± 0.32
	INS	25.3 ± 0.25	43.6 ± 0.39	44.7 ± 0.42	45.0 ± 0.43	43.3 ± 0.43
Plants	FLW	50.8 ± 0.46	76.2 ± 0.42	68.0 ± 0.42	79.5 ± 0.33	72.8 ± 0.44
	PLT_NET	31.1 ± 0.30	40.5 ± 0.32	36.7 ± 0.29	41.1 ± 0.33	38.6 ± 0.39
	FNG	27.4 ± 0.26	27.6 ± 0.25	25.1 ± 0.25	29.1 ± 0.26	25.1 ± 0.25
Plant diseases	PLT_VIL	68.6 ± 0.41	85.8 ± 0.25	58.8 ± 0.33	80.9 ± 0.24	70.2 ± 0.35
	MED_LF	77.6 ± 0.35	79.9 ± 0.30	75.3 ± 0.30	83.6 ± 0.24	85.6 ± 0.22
	PLT_DOC	23.4 ± 0.23	27.5 ± 0.25	27.8 ± 0.26	26.3 ± 0.27	26.5 ± 0.26
Microscopy	BCT	51.1 ± 0.42	78.2 ± 0.36	78.6 ± 0.41	82.8 ± 0.34	80.4 ± 0.40
	PNU	28.5 ± 0.27	23.0 ± 0.21	26.5 ± 0.23	26.9 ± 0.24	31.0 ± 0.27
	PRT	25.1 ± 0.25	31.0 ± 0.25	31.5 ± 0.28	30.3 ± 0.28	33.3 ± 0.30
Remote sensing	RESISC	41.4 ± 0.38	57.5 ± 0.36	48.7 ± 0.33	62.7 ± 0.37	61.7 ± 0.35
	RSICB	60.5 ± 0.47	81.9 ± 0.33	77.2 ± 0.29	84.4 ± 0.26	79.9 ± 0.35
	RSD	39.9 ± 0.36	57.1 ± 0.41	56.2 ± 0.43	61.2 ± 0.39	49.4 ± 0.31
Vehicles	CRS	28.3 ± 0.28	65.3 ± 0.42	72.4 ± 0.45	70.2 ± 0.41	71.2 ± 0.48
	APL	40.1 ± 0.37	47.6 ± 0.37	42.7 ± 0.33	55.6 ± 0.38	45.0 ± 0.38
	BTS	27.8 ± 0.28	32.9 ± 0.25	29.2 ± 0.27	33.6 ± 0.27	28.9 ± 0.25
Manufacturing	TEX	71.4 ± 0.57	91.7 ± 0.27	91.5 ± 0.30	93.0 ± 0.28	92.5 ± 0.28
	TEX_DTD	27.2 ± 0.29	38.8 ± 0.34	36.4 ± 0.32	34.6 ± 0.32	35.6 ± 0.30
	TEX_ALOT	69.2 ± 0.57	96.9 ± 0.18	97.1 ± 0.18	98.2 ± 0.12	97.7 ± 0.16
Human actions	SPT	42.3 ± 0.43	60.8 ± 0.42	55.0 ± 0.43	62.0 ± 0.43	55.3 ± 0.44
	ACT_40	25.8 ± 0.24	29.6 ± 0.26	26.5 ± 0.27	32.3 ± 0.26	26.7 ± 0.25
	ACT_410	41.3 ± 0.34	43.6 ± 0.31	30.1 ± 0.27	40.6 ± 0.30	42.1 ± 0.48
OCR	MD_MIX	19.6 ± 0.18	53.1 ± 0.67	80.7 ± 0.42	95.7 ± 0.17	20.0 ± 0.18
	MD_5_BIS	20.3 ± 0.19	56.1 ± 0.49	90.0 ± 0.33	96.6 ± 0.17	20.5 ± 0.20
	MD_6	22.2 ± 0.23	90.2 ± 0.40	91.7 ± 0.27	97.0 ± 0.12	21.3 ± 0.21



(a) Average rankings



(b) Average accuracy

Figure 8: **Within domain few-shot learning.** Averages over 30 released Meta-Album datasets of: (a) algorithm rank (smaller is better). The 95% confidence intervals are computed at dataset level over 30 datasets. (b) classification accuracy. The 95% confidence intervals are computed at per-task level over 3 runs per dataset with 600 tasks per run (total tasks: $3 \times 600 \times 30 = 54\,000$).

Table 4: Full performance results for 5-way 10-shot image classification on all datasets. The 95% confidence intervals are computed at per-task level over 3 runs per dataset with 600 tasks per run (total tasks: $3 \times 600 = 1\,800$).

Domain	Dataset ID	TrainFromScratch	Finetuning	MatchingNet	ProtoNet	FO-MAML
Large animals	BRD	47.5 ± 0.44	83.2 ± 0.34	78.8 ± 0.36	85.6 ± 0.29	78.7 ± 0.37
	DOG	31.4 ± 0.31	43.3 ± 0.35	46.5 ± 0.36	47.6 ± 0.37	38.8 ± 0.37
	AWA	37.8 ± 0.34	46.5 ± 0.35	40.9 ± 0.31	52.0 ± 0.32	42.2 ± 0.37
Small animals	PLK	59.6 ± 0.50	77.0 ± 0.47	73.4 ± 0.45	78.7 ± 0.42	71.7 ± 0.50
	INS_2	28.2 ± 0.31	37.0 ± 0.32	34.1 ± 0.30	36.1 ± 0.33	35.3 ± 0.33
	INS	26.1 ± 0.26	47.0 ± 0.38	47.4 ± 0.41	52.5 ± 0.40	44.0 ± 0.43
Plants	FLW	53.3 ± 0.48	78.6 ± 0.39	72.3 ± 0.38	82.8 ± 0.30	75.9 ± 0.43
	PLT_NET	32.7 ± 0.31	43.1 ± 0.31	41.3 ± 0.28	47.3 ± 0.28	38.0 ± 0.31
	FNG	27.9 ± 0.25	28.9 ± 0.24	28.7 ± 0.23	31.6 ± 0.25	29.2 ± 0.24
Plant diseases	PLT_VIL	71.3 ± 0.40	87.3 ± 0.24	70.6 ± 0.29	87.6 ± 0.19	76.9 ± 0.30
	MED_LF	81.3 ± 0.33	82.4 ± 0.27	74.9 ± 0.39	81.4 ± 0.23	88.6 ± 0.21
	PLT_DOC	23.7 ± 0.22	29.4 ± 0.25	31.6 ± 0.27	30.6 ± 0.24	27.4 ± 0.25
Microscopy	BCT	53.5 ± 0.41	80.5 ± 0.35	82.4 ± 0.37	83.9 ± 0.32	82.8 ± 0.38
	PNU	29.5 ± 0.27	22.5 ± 0.19	27.9 ± 0.22	29.3 ± 0.23	34.8 ± 0.25
	PRT	25.8 ± 0.26	33.0 ± 0.25	35.0 ± 0.27	33.4 ± 0.28	32.2 ± 0.27
Remote sensing	RESISC	42.4 ± 0.37	61.3 ± 0.36	54.4 ± 0.29	66.4 ± 0.35	68.3 ± 0.35
	RSICB	61.1 ± 0.46	86.9 ± 0.25	85.5 ± 0.26	85.8 ± 0.26	87.6 ± 0.26
	RSD	40.2 ± 0.35	61.2 ± 0.38	53.3 ± 0.57	64.7 ± 0.37	55.0 ± 0.29
Vehicles	CRS	29.9 ± 0.30	69.4 ± 0.40	76.7 ± 0.39	72.4 ± 0.39	74.0 ± 0.43
	APL	41.9 ± 0.39	50.1 ± 0.35	51.5 ± 0.41	55.5 ± 0.31	49.9 ± 0.39
	BTS	29.1 ± 0.29	35.9 ± 0.25	33.1 ± 0.26	38.3 ± 0.26	31.2 ± 0.25
Manufacturing	TEX	71.6 ± 0.58	93.8 ± 0.22	90.9 ± 0.29	93.4 ± 0.26	93.3 ± 0.25
	TEX_DTD	27.2 ± 0.28	42.4 ± 0.32	36.6 ± 0.33	37.5 ± 0.31	40.1 ± 0.30
	TEX_ALOT	70.7 ± 0.57	97.6 ± 0.15	98.5 ± 0.11	98.7 ± 0.09	98.3 ± 0.12
Human actions	SPT	44.0 ± 0.42	64.8 ± 0.39	62.2 ± 0.40	65.7 ± 0.42	62.3 ± 0.46
	ACT_40	26.7 ± 0.24	31.6 ± 0.24	31.3 ± 0.24	35.3 ± 0.25	28.1 ± 0.23
	ACT_410	42.7 ± 0.35	47.3 ± 0.31	31.8 ± 0.31	47.1 ± 0.28	40.8 ± 0.50
OCR	MD_MIX	19.5 ± 0.16	55.6 ± 0.70	92.3 ± 0.24	96.5 ± 0.16	20.3 ± 0.17
	MD_5_BIS	20.2 ± 0.18	60.0 ± 0.49	95.1 ± 0.21	96.9 ± 0.16	21.0 ± 0.21
	MD_6	22.1 ± 0.22	91.8 ± 0.36	94.5 ± 0.19	97.9 ± 0.10	45.5 ± 1.58

Table 5: Full performance results for 5-way 20-shot image classification on all data sets. The 95% confidence intervals are computed at per-task level over 3 runs per dataset with 600 tasks per run (total tasks: $3 \times 600 = 1\,800$).

Domain	Dataset ID	TrainFromScratch	Finetuning	MatchingNet	ProtoNet	FO-MAML
Large animals	BRD	47.8 ± 0.47	85.1 ± 0.32	82.2 ± 0.32	86.9 ± 0.27	81.0 ± 0.36
	DOG	31.7 ± 0.31	46.0 ± 0.36	48.1 ± 0.33	50.9 ± 0.36	41.7 ± 0.36
	AWA	37.8 ± 0.33	49.4 ± 0.34	46.7 ± 0.32	53.4 ± 0.33	48.9 ± 0.33
Small animals	PLK	59.9 ± 0.50	79.5 ± 0.44	77.3 ± 0.41	80.2 ± 0.40	75.8 ± 0.48
	INS_2	28.1 ± 0.31	39.3 ± 0.33	37.3 ± 0.32	41.3 ± 0.31	38.9 ± 0.33
	INS	25.8 ± 0.26	49.8 ± 0.39	51.2 ± 0.37	54.4 ± 0.36	47.2 ± 0.43
Plants	FLW	54.2 ± 0.50	80.5 ± 0.36	76.4 ± 0.33	85.7 ± 0.26	79.2 ± 0.35
	PLT_NET	32.6 ± 0.32	44.6 ± 0.37	47.7 ± 0.27	47.0 ± 0.32	41.6 ± 0.30
	FNG	27.8 ± 0.25	30.1 ± 0.24	32.2 ± 0.27	34.2 ± 0.24	31.3 ± 0.23
Plant diseases	PLT_VIL	72.0 ± 0.42	92.9 ± 0.17	80.7 ± 0.24	90.2 ± 0.17	85.3 ± 0.21
	MED_LF	82.6 ± 0.32	83.7 ± 0.30	81.5 ± 0.23	84.4 ± 0.25	93.1 ± 0.16
	PLT_DOC	23.5 ± 0.21	31.9 ± 0.25	31.8 ± 0.21	30.7 ± 0.23	32.2 ± 0.25
Microscopy	BCT	53.6 ± 0.42	80.8 ± 0.35	85.7 ± 0.28	87.0 ± 0.30	82.4 ± 0.40
	PNU	29.7 ± 0.26	23.5 ± 0.20	29.8 ± 0.21	27.2 ± 0.28	38.0 ± 0.26
	PRT	26.1 ± 0.25	34.2 ± 0.26	36.3 ± 0.26	36.8 ± 0.27	32.7 ± 0.25
Remote sensing	RESISC	42.8 ± 0.36	65.7 ± 0.34	64.7 ± 0.31	67.2 ± 0.31	65.0 ± 0.39
	RSICB	61.5 ± 0.49	89.2 ± 0.21	88.9 ± 0.22	90.9 ± 0.19	91.2 ± 0.22
	RSD	40.1 ± 0.35	65.9 ± 0.37	63.3 ± 0.36	67.6 ± 0.35	58.8 ± 0.28
Vehicles	CRS	30.1 ± 0.32	73.1 ± 0.39	76.4 ± 0.37	76.1 ± 0.35	75.8 ± 0.44
	APL	41.5 ± 0.38	52.7 ± 0.32	57.3 ± 0.27	57.9 ± 0.33	52.1 ± 0.32
	BTS	29.1 ± 0.30	36.9 ± 0.26	36.3 ± 0.24	40.1 ± 0.27	32.5 ± 0.24
Manufacturing	TEX	71.6 ± 0.59	95.4 ± 0.19	94.7 ± 0.21	95.1 ± 0.21	94.7 ± 0.22
	TEX_DTD	27.1 ± 0.28	46.2 ± 0.35	39.4 ± 0.31	41.7 ± 0.32	44.1 ± 0.29
	TEX_ALOT	70.4 ± 0.58	98.3 ± 0.11	98.9 ± 0.10	98.8 ± 0.08	98.6 ± 0.11
Human actions	SPT	44.4 ± 0.43	68.4 ± 0.39	66.1 ± 0.37	69.8 ± 0.38	63.5 ± 0.43
	ACT_40	26.8 ± 0.23	32.4 ± 0.23	33.6 ± 0.22	37.1 ± 0.24	29.5 ± 0.22
	ACT_410	42.5 ± 0.36	47.5 ± 0.30	40.7 ± 0.23	50.7 ± 0.29	55.9 ± 0.31
OCR	MD_MIX	19.5 ± 0.15	52.3 ± 0.77	94.6 ± 0.19	96.7 ± 0.15	20.4 ± 0.14
	MD_5_BIS	20.2 ± 0.17	71.8 ± 0.61	96.3 ± 0.17	96.7 ± 0.15	21.1 ± 0.21
	MD_6	21.4 ± 0.19	93.2 ± 0.34	96.5 ± 0.14	98.0 ± 0.09	45.9 ± 1.58

Table 6: The average rankings over all data sets and running times for all techniques on 5-way [1, 5, 10, 20]-shot image classification. The 95% confidence intervals are computed at dataset level over 30 datasets.

	1-shot		5-shot		10-shot		20-shot	
	AR	Time	AR	Time	AR	Time	AR	Time
TrainFromScratch	4.6 ± 0.35	1h29min	4.7 ± 0.28	1h29min	4.8 ± 0.25	1h29min	4.9 ± 0.16	1h29min
Finetuning	2.4 ± 0.39	3h31min	2.5 ± 0.36	3h37min	2.5 ± 0.36	3h41min	2.7 ± 0.36	4h18min
MatchingNet	3.2 ± 0.36	2h50min	3.3 ± 0.46	3h19min	3.1 ± 0.45	3h40min	2.8 ± 0.43	6h52min
ProtoNet	1.8 ± 0.43	2h52min	1.7 ± 0.41	3h15min	1.5 ± 0.29	3h40min	1.5 ± 0.32	6h58min
FO-MAML	2.9 ± 0.48	6h52min	2.9 ± 0.38	9h12min	3.1 ± 0.40	11h24min	3.1 ± 0.41	17h8min

F Cross-Domain Few-shot learning: additional results

Here, we show additional results for the cross-domain few-shot learning experiments from Section 3.2. Table 7, Table 8, Table 9, and Table 10 display the average accuracy per technique and dataset in the 5-way [1, 5, 10, 20]-shot settings. Moreover, Figure 9 displays the average rank per method and average accuracy for the same settings.

Table 11 displays the average accuracy per technique and dataset in the any-way any-shot setting. Table 12 and Table 13 display the results of the any-way any-shot setting but grouped by N-way and k-shot, respectively. Figure 10 shows a performance comparison of the different techniques when trained on each of the evaluated settings and tested on the novel any-way any-shot setting. From these results, we can observe that most of the techniques are not benefited from training with a variable number of shots; only MAML shows a performance improvement by this approach.

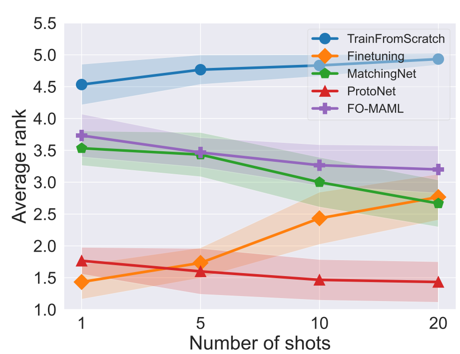
Table 14 displays the running times per technique in the different cross-domain settings. This time is separated in meta-training, meta-validation and meta-testing. Note that the meta-training and meta-validation time of TrainFromScratch is 0 because this baseline learns every task starting from a random initialization at meta-test time.

Table 7: Full performance results for cross-domain 5-way 1-shot image classification on all datasets. The 95% confidence intervals are computed at per-task level over 3 runs per dataset with 600 tasks per run (total tasks: $3 \times 600 = 1\,800$).

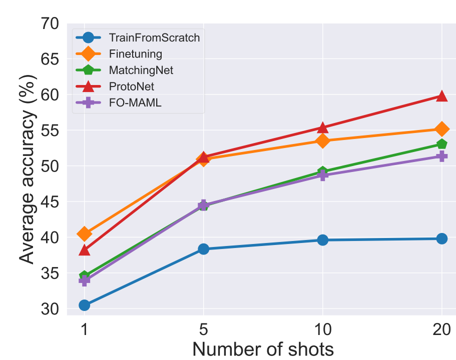
Domain	Dataset ID	TrainFromScratch	Finetuning	MatchingNet	ProtoNet	FO-MAML
Large animals	BRD	31.45 ± 0.38	42.93 ± 0.43	37.53 ± 0.48	40.07 ± 0.44	31.59 ± 0.35
	DOG	23.79 ± 0.27	26.38 ± 0.28	25.17 ± 0.29	26.15 ± 0.29	24.41 ± 0.29
	AWA	25.75 ± 0.31	36.69 ± 0.44	32.16 ± 0.43	36.83 ± 0.46	32.12 ± 0.44
Small animals	PLK	42.15 ± 0.49	50.72 ± 0.52	34.28 ± 0.47	44.78 ± 0.52	42.08 ± 0.48
	INS_2	23.81 ± 0.28	29.10 ± 0.32	27.03 ± 0.34	28.22 ± 0.33	25.48 ± 0.31
	INS	23.00 ± 0.25	31.01 ± 0.34	27.07 ± 0.35	30.57 ± 0.34	27.10 ± 0.34
Plants	FLW	38.47 ± 0.48	49.48 ± 0.51	50.23 ± 0.59	53.57 ± 0.57	42.91 ± 0.50
	PLT_NET	24.57 ± 0.28	32.80 ± 0.36	29.88 ± 0.37	30.78 ± 0.35	27.87 ± 0.34
	FNG	22.04 ± 0.24	26.92 ± 0.29	25.93 ± 0.33	27.11 ± 0.31	26.73 ± 0.34
Plant diseases	PLT_VIL	37.49 ± 0.45	52.47 ± 0.51	42.18 ± 0.48	50.03 ± 0.53	44.61 ± 0.45
	MED_LF	54.54 ± 0.56	68.19 ± 0.54	56.02 ± 0.59	64.76 ± 0.54	50.27 ± 0.61
	PLT_DOC	23.61 ± 0.28	30.54 ± 0.32	26.92 ± 0.33	33.77 ± 0.38	27.61 ± 0.34
Microscopy	BCT	32.26 ± 0.41	40.16 ± 0.44	34.64 ± 0.46	37.24 ± 0.43	35.40 ± 0.42
	PNU	22.61 ± 0.26	23.71 ± 0.24	23.01 ± 0.26	23.83 ± 0.25	22.58 ± 0.25
	PRT	21.78 ± 0.24	23.56 ± 0.24	24.36 ± 0.27	25.08 ± 0.28	25.10 ± 0.29
Remote sensing	RESISC	33.61 ± 0.44	42.84 ± 0.45	41.76 ± 0.52	44.37 ± 0.48	38.83 ± 0.47
	RSICB	45.48 ± 0.55	66.90 ± 0.51	59.01 ± 0.62	62.77 ± 0.57	60.44 ± 0.65
	RSD	31.74 ± 0.40	46.82 ± 0.49	37.97 ± 0.47	48.21 ± 0.50	38.64 ± 0.47
Vehicles	CRS	24.03 ± 0.26	27.10 ± 0.31	25.75 ± 0.35	26.70 ± 0.33	23.73 ± 0.26
	APL	30.76 ± 0.36	34.67 ± 0.37	28.29 ± 0.34	32.81 ± 0.36	26.90 ± 0.33
	BTS	22.77 ± 0.26	26.85 ± 0.29	25.30 ± 0.32	25.80 ± 0.28	23.30 ± 0.26
Manufacturing	TEX	53.41 ± 0.65	72.43 ± 0.58	71.29 ± 0.63	70.04 ± 0.65	62.54 ± 0.64
	TEX_DTD	24.37 ± 0.30	33.19 ± 0.37	26.70 ± 0.35	30.25 ± 0.38	26.95 ± 0.35
	TEX_ALOT	55.62 ± 0.59	83.82 ± 0.44	74.34 ± 0.60	84.43 ± 0.44	77.78 ± 0.62
Human actions	SPT	32.08 ± 0.40	38.68 ± 0.39	32.18 ± 0.39	36.02 ± 0.41	33.96 ± 0.39
	ACT_40	23.23 ± 0.26	28.83 ± 0.30	26.31 ± 0.31	27.19 ± 0.30	25.69 ± 0.32
	ACT_410	28.08 ± 0.35	34.61 ± 0.37	31.00 ± 0.40	35.52 ± 0.40	32.05 ± 0.39
OCR	MD_MIX	19.93 ± 0.19	35.68 ± 0.40	19.63 ± 0.20	19.79 ± 0.19	19.68 ± 0.20
	MD_5_BIS	20.46 ± 0.21	26.66 ± 0.28	20.21 ± 0.21	21.35 ± 0.22	20.06 ± 0.21
	MD_6	21.09 ± 0.22	50.13 ± 0.48	20.94 ± 0.23	28.75 ± 0.47	20.52 ± 0.22

Table 8: Full performance results for cross-domain 5-way 5-shot image classification on all datasets. The 95% confidence intervals are computed at per-task level over 3 runs per dataset with 600 tasks per run (total tasks: $3 \times 600 = 1\,800$).

Domain	Dataset ID	TrainFromScratch	Finetuning	MatchingNet	ProtoNet	FO-MAML
Large animals	BRD	45.83 ± 0.43	57.30 ± 0.44	50.10 ± 0.44	59.77 ± 0.41	46.19 ± 0.42
	DOG	28.70 ± 0.29	32.15 ± 0.30	32.16 ± 0.30	32.04 ± 0.30	31.55 ± 0.32
	AWA	32.75 ± 0.37	47.31 ± 0.48	44.43 ± 0.46	49.63 ± 0.48	44.06 ± 0.47
Small animals	PLK	55.01 ± 0.52	62.53 ± 0.49	50.65 ± 0.43	63.35 ± 0.49	53.75 ± 0.54
	INS_2	27.91 ± 0.30	37.71 ± 0.33	34.64 ± 0.34	35.32 ± 0.33	34.73 ± 0.35
	INS	27.15 ± 0.28	40.09 ± 0.38	36.29 ± 0.35	39.90 ± 0.38	35.62 ± 0.36
Plants	FLW	53.29 ± 0.51	64.07 ± 0.53	63.37 ± 0.49	70.72 ± 0.47	60.74 ± 0.49
	PLT_NET	29.74 ± 0.31	42.85 ± 0.37	40.03 ± 0.37	38.92 ± 0.34	40.89 ± 0.36
	FNG	25.41 ± 0.26	32.64 ± 0.32	32.28 ± 0.35	33.56 ± 0.34	32.43 ± 0.36
Plant diseases	PLT_VIL	55.92 ± 0.49	67.84 ± 0.49	58.64 ± 0.43	69.99 ± 0.47	61.85 ± 0.47
	MED_LF	72.79 ± 0.47	81.77 ± 0.42	70.24 ± 0.46	81.89 ± 0.39	76.26 ± 0.45
	PLT_DOC	27.52 ± 0.28	39.41 ± 0.35	35.70 ± 0.34	46.88 ± 0.38	36.01 ± 0.34
Microscopy	BCT	43.48 ± 0.47	49.64 ± 0.45	49.37 ± 0.48	54.46 ± 0.49	46.42 ± 0.44
	PNU	25.12 ± 0.26	25.75 ± 0.23	25.55 ± 0.25	25.34 ± 0.24	25.42 ± 0.26
	PRT	25.71 ± 0.26	26.99 ± 0.26	29.81 ± 0.28	29.89 ± 0.28	29.23 ± 0.29
Remote sensing	RESISC	41.74 ± 0.43	56.74 ± 0.44	54.84 ± 0.44	57.62 ± 0.45	52.88 ± 0.47
	RSICB	57.07 ± 0.51	82.30 ± 0.39	75.38 ± 0.43	77.57 ± 0.42	77.56 ± 0.46
	RSD	39.09 ± 0.40	62.31 ± 0.46	53.84 ± 0.43	67.28 ± 0.46	56.12 ± 0.46
Vehicles	CRS	29.54 ± 0.32	32.82 ± 0.33	31.28 ± 0.35	34.66 ± 0.37	29.70 ± 0.33
	APL	41.97 ± 0.47	48.65 ± 0.42	36.68 ± 0.37	49.66 ± 0.40	39.36 ± 0.39
	BTS	26.62 ± 0.28	33.48 ± 0.31	30.02 ± 0.28	35.56 ± 0.32	30.62 ± 0.28
Manufacturing	TEX	69.12 ± 0.60	81.83 ± 0.50	81.44 ± 0.46	85.84 ± 0.40	78.84 ± 0.52
	TEX_DTD	27.09 ± 0.29	45.20 ± 0.39	36.66 ± 0.36	43.37 ± 0.39	36.16 ± 0.36
	TEX_ALOT	68.29 ± 0.55	93.65 ± 0.26	89.34 ± 0.34	95.36 ± 0.20	89.77 ± 0.39
Human actions	SPT	43.78 ± 0.43	51.71 ± 0.43	45.05 ± 0.41	50.85 ± 0.41	45.40 ± 0.43
	ACT_40	28.33 ± 0.29	35.41 ± 0.33	34.48 ± 0.33	31.62 ± 0.31	33.23 ± 0.33
	ACT_410	38.91 ± 0.40	47.89 ± 0.40	44.65 ± 0.41	51.08 ± 0.40	45.87 ± 0.41
OCR	MD_MIX	19.78 ± 0.18	48.39 ± 0.41	19.65 ± 0.20	20.92 ± 0.21	19.25 ± 0.20
	MD_5_BIS	20.21 ± 0.19	33.26 ± 0.33	21.40 ± 0.21	51.36 ± 0.48	22.12 ± 0.23
	MD_6	22.28 ± 0.22	65.88 ± 0.45	23.84 ± 0.25	52.77 ± 0.57	22.96 ± 0.23



(a) Average rankings



(b) Average accuracy

Figure 9: **Cross-domain few-shot learning.** Averages over 30 released Meta-Album datasets of: (a) algorithm rank (smaller is better). The 95% confidence intervals are computed at dataset level over 30 datasets. (b) classification accuracy. The 95% confidence intervals are computed at per-task level over 3 runs per dataset with 600 tasks per run (total tasks: $3 \times 600 \times 30 = 54\,000$).

Table 9: Full performance results for cross-domain 5-way 10-shot image classification on all datasets. The 95% confidence intervals are computed at per-task level over 3 runs per dataset with 600 tasks per run (total tasks: $3 \times 600 = 1\,800$).

Domain	Dataset ID	TrainFromScratch	Finetuning	MatchingNet	ProtoNet	FO-MAML
Large animals	BRD	48.48 ± 0.46	61.79 ± 0.43	57.60 ± 0.40	62.38 ± 0.40	51.58 ± 0.40
	DOG	29.52 ± 0.29	34.00 ± 0.30	35.91 ± 0.32	36.10 ± 0.31	34.48 ± 0.32
	AWA	33.74 ± 0.38	50.16 ± 0.49	50.50 ± 0.45	54.60 ± 0.48	47.58 ± 0.48
Small animals	PLK	56.40 ± 0.52	64.48 ± 0.49	57.23 ± 0.43	70.63 ± 0.45	63.18 ± 0.44
	INS_2	28.03 ± 0.30	40.23 ± 0.34	38.74 ± 0.35	40.13 ± 0.35	38.89 ± 0.36
	INS	27.76 ± 0.30	41.61 ± 0.39	41.40 ± 0.35	45.36 ± 0.38	39.95 ± 0.38
Plants	FLW	55.70 ± 0.52	67.72 ± 0.50	68.24 ± 0.45	71.31 ± 0.46	66.97 ± 0.48
	PLT_NET	31.02 ± 0.32	45.71 ± 0.36	45.40 ± 0.37	43.16 ± 0.33	45.49 ± 0.35
	FNG	26.50 ± 0.27	34.09 ± 0.33	35.66 ± 0.35	37.73 ± 0.33	34.33 ± 0.33
Plant diseases	PLT_VIL	59.48 ± 0.51	70.68 ± 0.48	66.44 ± 0.44	74.39 ± 0.43	70.83 ± 0.43
	MED_LF	76.44 ± 0.47	83.83 ± 0.41	71.73 ± 0.40	86.76 ± 0.34	80.80 ± 0.40
	PLT_DOC	27.95 ± 0.29	41.36 ± 0.35	43.32 ± 0.34	50.87 ± 0.38	38.73 ± 0.34
Microscopy	BCT	46.23 ± 0.48	51.23 ± 0.46	57.60 ± 0.43	54.61 ± 0.45	51.40 ± 0.46
	PNU	26.17 ± 0.25	26.03 ± 0.23	28.53 ± 0.26	27.12 ± 0.25	28.64 ± 0.26
	PRT	26.65 ± 0.28	28.15 ± 0.26	31.40 ± 0.30	32.87 ± 0.27	32.55 ± 0.29
Remote sensing	RESISC	42.49 ± 0.42	60.34 ± 0.44	60.00 ± 0.41	61.27 ± 0.43	57.14 ± 0.46
	RSICB	57.65 ± 0.54	84.49 ± 0.36	79.37 ± 0.39	83.72 ± 0.36	81.53 ± 0.40
	RSD	39.90 ± 0.42	66.35 ± 0.46	64.29 ± 0.41	72.30 ± 0.42	61.07 ± 0.47
Vehicles	CRS	31.39 ± 0.35	34.76 ± 0.36	35.05 ± 0.36	37.47 ± 0.37	32.20 ± 0.31
	APL	43.64 ± 0.51	52.67 ± 0.43	41.73 ± 0.42	56.30 ± 0.39	42.43 ± 0.40
	BTS	27.44 ± 0.28	35.51 ± 0.31	31.82 ± 0.28	40.35 ± 0.33	33.93 ± 0.29
Manufacturing	TEX	71.33 ± 0.60	82.25 ± 0.50	85.79 ± 0.38	86.20 ± 0.41	84.93 ± 0.44
	TEX_DTD	27.09 ± 0.28	49.24 ± 0.37	40.48 ± 0.35	49.45 ± 0.38	39.60 ± 0.36
	TEX_ALOT	69.09 ± 0.56	94.96 ± 0.23	92.44 ± 0.26	97.04 ± 0.15	92.97 ± 0.32
Human actions	SPT	45.98 ± 0.44	55.94 ± 0.42	51.93 ± 0.39	54.67 ± 0.42	52.74 ± 0.42
	ACT_40	29.32 ± 0.28	37.45 ± 0.34	37.96 ± 0.34	35.27 ± 0.32	37.15 ± 0.35
	ACT_410	40.49 ± 0.40	51.72 ± 0.41	50.43 ± 0.39	58.72 ± 0.39	52.12 ± 0.41
OCR	MD_MIX	19.51 ± 0.17	50.45 ± 0.41	19.68 ± 0.21	35.54 ± 0.58	18.78 ± 0.20
	MD_5_BIS	20.37 ± 0.18	38.49 ± 0.34	22.84 ± 0.23	53.14 ± 0.39	22.75 ± 0.23
	MD_6	22.24 ± 0.22	69.41 ± 0.42	32.29 ± 0.31	51.61 ± 0.46	24.98 ± 0.25

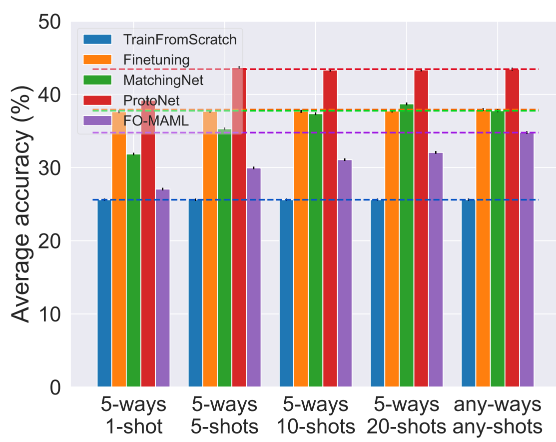


Figure 10: **Comparison of "cross-domain" few-shot learning trained on different settings but tested on the any-way any-shot setting.** We plot few-shot learning meta-test mean task accuracy, averaged over test tasks drawn from the 30 released Meta-Album datasets. The 95% confidence intervals are computed at per-task level over 3 runs per dataset with 600 tasks per run (total tasks: $3 \times 600 \times 30 = 54\,000$).

Table 10: Full performance results for cross-domain 5-way 20-shot image classification on all datasets. The 95% confidence intervals are computed at per-task level over 3 runs per dataset with 600 tasks per run (total tasks: $3 \times 600 = 1\,800$).

Domain	Dataset ID	TrainFromScratch	Finetuning	MatchingNet	ProtoNet	FO-MAML
Large animals	BRD	48.94 ± 0.48	63.79 ± 0.43	62.04 ± 0.37	67.99 ± 0.36	58.02 ± 0.41
	DOG	29.73 ± 0.31	35.63 ± 0.32	40.19 ± 0.32	40.05 ± 0.32	36.24 ± 0.31
	AWA	33.87 ± 0.38	51.84 ± 0.49	54.40 ± 0.43	58.49 ± 0.46	50.15 ± 0.48
Small animals	PLK	56.86 ± 0.54	65.72 ± 0.49	61.97 ± 0.41	76.19 ± 0.41	63.71 ± 0.52
	INS_2	27.98 ± 0.31	42.31 ± 0.35	43.40 ± 0.35	44.28 ± 0.36	40.70 ± 0.36
	INS	27.75 ± 0.30	43.22 ± 0.39	46.51 ± 0.37	49.40 ± 0.40	43.69 ± 0.38
Plants	FLW	56.41 ± 0.52	69.22 ± 0.51	68.95 ± 0.45	75.35 ± 0.43	71.06 ± 0.46
	PLT_NET	31.26 ± 0.31	47.83 ± 0.38	49.95 ± 0.34	47.25 ± 0.34	48.10 ± 0.35
	FNG	26.39 ± 0.27	35.44 ± 0.34	39.99 ± 0.34	42.02 ± 0.34	36.88 ± 0.34
Plant diseases	PLT_VIL	60.83 ± 0.52	72.18 ± 0.48	70.51 ± 0.39	78.77 ± 0.39	73.50 ± 0.43
	MED_LF	77.76 ± 0.46	85.50 ± 0.38	77.28 ± 0.38	88.53 ± 0.29	83.06 ± 0.38
	PLT_DOC	28.05 ± 0.28	43.92 ± 0.36	48.06 ± 0.34	54.50 ± 0.38	42.12 ± 0.33
Microscopy	BCT	46.74 ± 0.47	52.04 ± 0.46	58.27 ± 0.44	58.56 ± 0.47	58.92 ± 0.54
	PNU	26.19 ± 0.25	26.61 ± 0.24	29.93 ± 0.27	29.88 ± 0.25	30.87 ± 0.26
	PRT	26.93 ± 0.28	28.71 ± 0.26	34.81 ± 0.27	35.98 ± 0.28	33.15 ± 0.27
Remote sensing	RESISC	42.93 ± 0.44	62.09 ± 0.44	61.86 ± 0.42	64.37 ± 0.40	60.75 ± 0.44
	RSICB	57.97 ± 0.54	86.17 ± 0.35	82.72 ± 0.34	86.73 ± 0.32	83.35 ± 0.40
	RSD	39.29 ± 0.40	68.37 ± 0.45	68.91 ± 0.41	76.32 ± 0.40	65.66 ± 0.43
Vehicles	CRS	31.53 ± 0.36	35.88 ± 0.37	36.38 ± 0.35	42.31 ± 0.37	35.64 ± 0.36
	APL	43.64 ± 0.53	55.19 ± 0.45	48.82 ± 0.37	62.64 ± 0.40	45.55 ± 0.39
	BTS	27.31 ± 0.29	36.60 ± 0.32	36.96 ± 0.30	46.13 ± 0.33	35.53 ± 0.28
Manufacturing	TEX	70.89 ± 0.60	82.78 ± 0.51	86.10 ± 0.37	89.39 ± 0.31	88.69 ± 0.37
	TEX_DTD	26.78 ± 0.29	51.60 ± 0.38	44.25 ± 0.35	50.70 ± 0.37	42.53 ± 0.36
	TEX_ALOT	69.64 ± 0.56	95.71 ± 0.22	94.65 ± 0.22	97.73 ± 0.13	94.94 ± 0.24
Human actions	SPT	46.76 ± 0.44	58.44 ± 0.42	55.69 ± 0.37	59.35 ± 0.39	55.40 ± 0.41
	ACT_40	29.43 ± 0.30	39.43 ± 0.35	41.35 ± 0.36	39.34 ± 0.35	39.66 ± 0.35
	ACT_410	40.34 ± 0.42	54.19 ± 0.42	56.65 ± 0.38	64.87 ± 0.37	55.20 ± 0.41
OCR	MD_MIX	19.54 ± 0.16	52.72 ± 0.41	20.39 ± 0.22	46.29 ± 0.37	18.30 ± 0.19
	MD_5_BIS	20.32 ± 0.17	40.75 ± 0.37	25.81 ± 0.29	56.00 ± 0.37	22.34 ± 0.23
	MD_6	21.78 ± 0.20	70.69 ± 0.40	44.25 ± 0.37	64.42 ± 0.39	26.73 ± 0.26

Table 11: Full performance results for cross-domain any-way any-shot image classification on all datasets. The 95% confidence intervals are computed at per-task level over 3 runs per dataset with 600 tasks per run (total tasks: $3 \times 600 = 1\,800$).

Domain	Dataset ID	TrainFromScratch	Finetuning	MatchingNet	ProtoNet	FO-MAML
Large animals	BRD	29.95 ± 0.92	37.96 ± 0.91	43.08 ± 0.76	51.93 ± 0.72	34.40 ± 0.82
	DOG	19.58 ± 0.73	22.95 ± 0.76	25.76 ± 0.71	27.05 ± 0.72	22.69 ± 0.70
	AWA	21.53 ± 0.74	33.64 ± 0.83	38.01 ± 0.73	42.67 ± 0.76	35.19 ± 0.73
Small animals	PLK	36.18 ± 0.97	47.26 ± 0.90	42.89 ± 0.80	52.86 ± 0.73	45.29 ± 0.84
	INS_2	18.01 ± 0.70	27.07 ± 0.77	28.12 ± 0.71	29.37 ± 0.70	25.22 ± 0.73
	INS	17.88 ± 0.69	25.93 ± 0.78	29.86 ± 0.71	33.81 ± 0.72	27.91 ± 0.71
Plants	FLW	34.11 ± 0.98	42.86 ± 0.90	51.66 ± 0.75	60.11 ± 0.72	46.61 ± 0.85
	PLT_NET	20.30 ± 0.74	30.94 ± 0.82	33.53 ± 0.73	32.49 ± 0.74	30.26 ± 0.75
	FNG	17.58 ± 0.67	23.26 ± 0.75	26.94 ± 0.70	28.65 ± 0.71	25.05 ± 0.69
Plant diseases	PLT_VIL	38.25 ± 1.02	49.60 ± 0.97	51.78 ± 0.77	61.95 ± 0.72	50.69 ± 0.92
	MED_LF	54.99 ± 0.93	63.73 ± 0.84	62.36 ± 0.63	75.24 ± 0.55	58.56 ± 0.86
	PLT_DOC	18.06 ± 0.67	27.55 ± 0.77	30.39 ± 0.70	39.09 ± 0.73	28.13 ± 0.70
Microscopy	BCT	29.68 ± 0.95	35.47 ± 0.90	41.98 ± 0.80	42.60 ± 0.78	34.13 ± 0.87
	PNU	17.70 ± 0.64	18.50 ± 0.64	21.52 ± 0.62	19.42 ± 0.60	18.47 ± 0.61
	PRT	17.80 ± 0.70	18.10 ± 0.66	22.03 ± 0.67	22.10 ± 0.68	21.40 ± 0.67
Remote sensing	RESISC	27.05 ± 0.87	39.70 ± 0.86	43.55 ± 0.73	49.34 ± 0.71	39.34 ± 0.80
	RSICB	36.19 ± 1.03	67.64 ± 0.74	66.59 ± 0.65	70.96 ± 0.62	65.20 ± 0.76
	RSD	25.61 ± 0.88	46.74 ± 0.90	50.28 ± 0.75	59.66 ± 0.73	47.44 ± 0.77
Vehicles	CRS	20.10 ± 0.72	22.36 ± 0.74	24.12 ± 0.70	27.77 ± 0.70	21.02 ± 0.69
	APL	26.67 ± 0.91	37.08 ± 0.86	31.05 ± 0.70	40.28 ± 0.73	27.91 ± 0.78
	BTS	17.96 ± 0.68	24.09 ± 0.76	24.28 ± 0.69	30.36 ± 0.71	21.75 ± 0.67
Manufacturing	TEX	45.63 ± 1.11	66.63 ± 0.82	73.21 ± 0.60	79.37 ± 0.53	66.78 ± 0.82
	TEX_DTD	17.84 ± 0.65	34.59 ± 0.78	30.79 ± 0.69	32.95 ± 0.70	28.23 ± 0.68
	TEX_ALOT	44.15 ± 1.11	83.10 ± 0.56	85.45 ± 0.43	91.05 ± 0.39	81.62 ± 0.74
Human actions	SPT	28.81 ± 0.89	35.82 ± 0.89	37.01 ± 0.75	43.60 ± 0.74	35.15 ± 0.82
	ACT_40	19.11 ± 0.70	24.74 ± 0.77	27.81 ± 0.70	27.96 ± 0.70	23.55 ± 0.68
	ACT_410	25.03 ± 0.80	33.79 ± 0.83	40.50 ± 0.70	46.14 ± 0.70	36.83 ± 0.73
OCR	MD_MIX	13.42 ± 0.54	37.76 ± 0.79	12.99 ± 0.55	14.00 ± 0.56	12.88 ± 0.55
	MD_5_BIS	13.73 ± 0.54	28.78 ± 0.77	15.34 ± 0.58	24.55 ± 0.77	14.73 ± 0.57
	MD_6	15.16 ± 0.59	50.02 ± 0.92	20.00 ± 0.65	46.12 ± 0.82	16.85 ± 0.60

Table 12: Full performance results for cross-domain any-way any-shot image classification per number of ways. The 95% confidence intervals are computed at per-task level using the number of task shown in the table.

N-way	Evaluated Tasks	TrainFromScratch	Finetuning	MatchingNet	ProtoNet	FO-MAML
2	2 925	0.70 ± 0.01	0.78 ± 0.01	0.73 ± 0.01	0.77 ± 0.01	0.72 ± 0.01
3	2 708	0.55 ± 0.01	0.66 ± 0.01	0.61 ± 0.01	0.65 ± 0.01	0.58 ± 0.01
4	2 910	0.46 ± 0.01	0.58 ± 0.01	0.54 ± 0.01	0.59 ± 0.01	0.52 ± 0.01
5	2 704	0.38 ± 0.01	0.53 ± 0.01	0.48 ± 0.01	0.54 ± 0.01	0.46 ± 0.01
6	2 856	0.33 ± 0.01	0.48 ± 0.01	0.44 ± 0.01	0.50 ± 0.01	0.43 ± 0.01
7	2 683	0.30 ± 0.01	0.44 ± 0.01	0.42 ± 0.01	0.47 ± 0.01	0.40 ± 0.01
8	3 025	0.26 ± 0.00	0.40 ± 0.01	0.39 ± 0.01	0.45 ± 0.01	0.37 ± 0.01
9	2 864	0.24 ± 0.00	0.37 ± 0.01	0.37 ± 0.01	0.43 ± 0.01	0.35 ± 0.01
10	2 943	0.22 ± 0.00	0.35 ± 0.01	0.35 ± 0.01	0.41 ± 0.01	0.33 ± 0.01
11	2 892	0.19 ± 0.00	0.33 ± 0.01	0.33 ± 0.01	0.40 ± 0.01	0.31 ± 0.01
12	2 983	0.18 ± 0.00	0.31 ± 0.01	0.32 ± 0.01	0.38 ± 0.01	0.29 ± 0.01
13	3 076	0.17 ± 0.00	0.29 ± 0.01	0.31 ± 0.01	0.37 ± 0.01	0.28 ± 0.01
14	2 952	0.15 ± 0.00	0.28 ± 0.01	0.30 ± 0.01	0.36 ± 0.01	0.26 ± 0.01
15	2 811	0.14 ± 0.00	0.26 ± 0.01	0.29 ± 0.01	0.35 ± 0.01	0.25 ± 0.01
16	2 682	0.13 ± 0.00	0.24 ± 0.01	0.27 ± 0.01	0.33 ± 0.01	0.23 ± 0.01
17	2 842	0.12 ± 0.00	0.24 ± 0.01	0.27 ± 0.01	0.33 ± 0.01	0.22 ± 0.01
18	2 779	0.12 ± 0.00	0.23 ± 0.01	0.26 ± 0.01	0.32 ± 0.01	0.21 ± 0.01
19	2 711	0.11 ± 0.00	0.21 ± 0.01	0.25 ± 0.01	0.31 ± 0.01	0.20 ± 0.01
20	2 654	0.10 ± 0.00	0.21 ± 0.01	0.25 ± 0.01	0.31 ± 0.01	0.19 ± 0.01

Table 13: Full performance results for cross-domain any-way any-shot image classification per number of shots. The 95% confidence intervals are computed at per-task level using the number of task shown in the table.

k-shot	Evaluated Tasks	TrainFromScratch	Finetuning	MatchingNet	ProtoNet	FO-MAML
1	2 635	0.21 ± 0.01	0.30 ± 0.01	0.29 ± 0.01	0.30 ± 0.01	0.25 ± 0.01
2	2 577	0.24 ± 0.01	0.35 ± 0.01	0.33 ± 0.01	0.35 ± 0.01	0.29 ± 0.01
3	2 571	0.25 ± 0.01	0.36 ± 0.01	0.34 ± 0.01	0.38 ± 0.01	0.31 ± 0.01
4	2 803	0.26 ± 0.01	0.37 ± 0.01	0.36 ± 0.01	0.40 ± 0.01	0.33 ± 0.01
5	2 737	0.26 ± 0.01	0.38 ± 0.01	0.37 ± 0.01	0.41 ± 0.01	0.34 ± 0.01
6	2 748	0.25 ± 0.01	0.37 ± 0.01	0.36 ± 0.01	0.41 ± 0.01	0.33 ± 0.01
7	2 761	0.26 ± 0.01	0.38 ± 0.01	0.37 ± 0.01	0.43 ± 0.01	0.34 ± 0.01
8	2 761	0.25 ± 0.01	0.38 ± 0.01	0.37 ± 0.01	0.43 ± 0.01	0.34 ± 0.01
9	2 718	0.26 ± 0.01	0.39 ± 0.01	0.38 ± 0.01	0.44 ± 0.01	0.35 ± 0.01
10	2 791	0.27 ± 0.01	0.40 ± 0.01	0.40 ± 0.01	0.46 ± 0.01	0.37 ± 0.01
11	2 850	0.26 ± 0.01	0.39 ± 0.01	0.39 ± 0.01	0.45 ± 0.01	0.36 ± 0.01
12	2 747	0.26 ± 0.01	0.39 ± 0.01	0.39 ± 0.01	0.46 ± 0.01	0.36 ± 0.01
13	2 580	0.26 ± 0.01	0.39 ± 0.01	0.40 ± 0.01	0.46 ± 0.01	0.37 ± 0.01
14	2 607	0.26 ± 0.01	0.39 ± 0.01	0.40 ± 0.01	0.46 ± 0.01	0.37 ± 0.01
15	2 683	0.26 ± 0.01	0.39 ± 0.01	0.40 ± 0.01	0.47 ± 0.01	0.37 ± 0.01
16	2 616	0.27 ± 0.01	0.40 ± 0.01	0.40 ± 0.01	0.47 ± 0.01	0.37 ± 0.01
17	2 632	0.27 ± 0.01	0.41 ± 0.01	0.41 ± 0.01	0.48 ± 0.01	0.38 ± 0.01
18	2 703	0.25 ± 0.01	0.38 ± 0.01	0.40 ± 0.01	0.47 ± 0.01	0.36 ± 0.01
19	2 753	0.25 ± 0.01	0.39 ± 0.01	0.40 ± 0.01	0.47 ± 0.01	0.37 ± 0.01
20	2 727	0.25 ± 0.01	0.39 ± 0.01	0.40 ± 0.01	0.47 ± 0.01	0.37 ± 0.01

Table 14: The running times for all techniques on all cross-domain settings over all datasets. The number of tasks in each phase is: 540 000 during meta-training, 129 600 during meta-validation, and 54 000 during meta-testing.

Setting	Phase	TrainFromScratch	Finetuning	MatchingNet	ProtoNet	FO-MAML
5-way 1-shot	Meta-train	0h0m	3h23m	6h33m	6h47m	17h40m
	Meta-validation	0h0m	24h53m	2h48m	2h45m	8h52m
	Meta-test	29h29m	9h21m	0h18m	0h19m	2h48m
5-way 5-shot	Meta-train	0h0m	3h30m	7h13m	7h0m	18h2m
	Meta-validation	0h0m	25h41m	3h24m	3h16m	9h32m
	Meta-test	29h24m	9h22m	0h19m	0h20m	2h58m
5-way 10-shot	Meta-train	0h0m	3h25m	7h25m	7h10m	19h28m
	Meta-validation	0h0m	25h45m	4h2m	3h55m	10h30m
	Meta-test	30h7m	9h26m	0h20m	0h21m	3h16m
5-way 20-shot	Meta-train	0h0m	3h24m	8h3m	8h20m	22h10m
	Meta-validation	0h0m	27h11m	5h21m	5h23m	13h18m
	Meta-test	29h56m	9h22m	0h24m	0h25m	3h39m
any-way any-shot	Meta-train	0h0m	3h20m	7h4m	7h59m	19h37m
	Meta-validation	0h0m	30h11m	8h27m	8h58m	17h21m
	Meta-test	30h2m	9h32m	0h32m	0h34m	4h19m

G Benchmarks/Datasets comparison

We compare the **Meta-Album meta-dataset** with other (meta-)datasets in Table 15. Each row shows a (meta-)dataset while each column shows features. The last three rows show versions of the newly proposed meta-dataset, *i.e.,* Meta-Album. The table describes the following features, divided in quantitative and qualitative features:

Quantitative features

- number of domains
- number of datasets
- number of total images
- min/max number of classes per domain
- min/max number of images per class
- size on disk (in GB)

Qualitative features

- Whether the (meta-)datasets has datasets from multiple domains
- Whether the (meta-)dataset is lightweight (easy to store on most computers, *i.e.,* < 20 GB)
- Whether the (meta-)dataset has a uniform number of images per class
- Whether the (meta-)dataset has images of uniform size
- Whether the (meta-)dataset is designed to be extended with more datasets or domains.

We discuss the number of domains and extensibility in more detail.

Dataset domains

We could identify the following domains in Meta-Dataset, VTAB, and CIFAR-100. For other datasets, we have not been able to find an adequate description of the domains, and as such we refrain from this.

Meta-Dataset The meta-dataset is divided into total 7 domains as listed below:

1. Organisms, *e.g.,* birds, fungi, flowers.
2. Common objects, *e.g.,* from ImageNet, ILSVRC, COCO.
3. Characters, *i.e.,* Omniglot.

Table 15: Feature comparison between Meta-Album and other large-scale or (meta-)datasets

Dataset/ Meta-Dataset	# of domains	# of datasets	# of images	min/max classes per domain	min/max images per class	size on disk	multi-domain	lightweight (<20GB)	uniform # of images per class	uniform image size	extendable
Meta-Dataset	7	10	53 068 000	43/1 696	3/140 000	210 GB	✓	✗	✗	✗	✗
VTAB	3	19	2 244 000	2/397	40/1 000	100 GB	✓	✗	✗	✗	✗
MS-COCO	1	1	328 000	80/80	9/10 777	44 GB	✗	✗	✗	✗	✗
Mini Imagenet	1	1	60 000	100/100	600/600	1 GB	✗	✓	✓	✓	✗
Omniglot	1	1	32 000	1 623/1 623	20/20	148 MB	✗	✓	✓	✓	✗
CUB-200	1	1	6 000	200/200	20/39	647 MB	✗	✓	✗	✗	✗
CIFAR-100	3	1	60 000	15/50	600/600	161 MB	✗	✓	✓	✓	✗
Meta-Album ***Micro***	**10**	**40**	**32 000**	**19/20**	**40/40**	**380 MB**	✓	✓	✓	✓	✓
Meta-Album ***Mini***	**10**	**40**	**220 950**	**19/706**	**40/40**	**3.9 GB**	✓	✓	✓	✓	✓
Meta-Album ***Extended***	**10**	**40**	**1 583 624**	**19/706**	**1/187 384**	**15 GB**	✓	✓	✗	✓	✓

4. Textures, *i.e.,* DTD.
5. Vehicles, *i.e.,* FGVC-Aircraft.
6. Drawings, *i.e.,* Quick, Draw!
7. Traffic signs, *i.e.,* the German Traffic Sign Recognition Benchmark.

CIFAR-100 The CIFAR-100 is divided into 3 domains:

1. people, reptiles, small mammals, aquatic mammals, fish, insects, large carnivores, large omnivores and herbivores, medium-sized mammals, non-insect invertebrates
2. trees, flowers, fruit and vegetables
3. vehicles 1, vehicles 2, large natural outdoor scenes, large man-made outdoor things, household furniture, household electrical devices, food containers

VTAB VTAB is divided into 3 domains:

1. *Natural* image tasks include images of the natural world captured through standard cameras, representing generic objects, fine-grained classes, or abstract concepts.
2. *Specialized* tasks utilize images captured using specialist equipment, such as medical images or remote sensing.
3. *Structured* tasks often derive from artificial environments that target understanding of specific changes between images, such as predicting the distance to an object in a 3D scene, counting objects, or detecting orientation.

Extensibility

Meta-Album is the only one among these datasets that has been designed explicitly to be extended. We are inviting the community to contribute more domains and datasets. We provide open-source code to prepare more datasets following the Meta-Album format, and we have a review process so that newly added datasets are checked for quality and subsequently included.

Important Links

Meta-Album Website :
https://meta-album.github.io/

Meta-Album GitHub repo :
https://github.com/ihsanullah2131/meta-album

NeurIPS Cross-Domain MetaDL Competition 2022 :
https://github.com/DustinCarrion/cd-metadl

NeurIPS MetaDL Challenge 2021 :
https://autodl.lri.fr/competitions/210

MetaDL Self Service :
https://competitions.codalab.org/competitions/31280

Factsheets repo :
https://github.com/ihsanullah2131/meta-album/tree/master/Factsheets

Data Format :
https://github.com/ihsanullah2131/meta-album/tree/master/DataFormat

Data Format Check :
https://github.com/ihsanullah2131/meta-album/blob/master/DataFormat/check_data_format.py

Factsheet Report Script :
https://github.com/ihsanullah2131/meta-album/blob/master/Factsheets/generate_pdf_report.py

Factsheet Report Template :
https://github.com/ihsanullah2131/meta-album/blob/master/Factsheets/template.html

Contact

For any query about the Meta-Album meta-dataset, reach us out by email meta-album@chalearn.org.

References

[1] M. Andriluka, L. Pishchulin, P. Gehler, and B. Schiele. "2D Human Pose Estimation: New Benchmark and State of the Art Analysis". In: *2014 IEEE Conference on Computer Vision and Pattern Recognition*. 2014, pp. 3686–3693.

[2] G. Bradski. "The openCV library." In: *Dr. Dobb's Journal: Software Tools for the Professional Programmer* 25.11 (2000), pp. 120–123.

[3] G. J. Burghouts and J.-M. Geusebroek. "Material-Specific Adaptation of Color Invariant Features". In: *Pattern Recognition Letters* 30.3 (2009), pp. 306–313.

[4] G. Cheng, J. Han, and X. Lu. "Remote Sensing Image Scene Classification: Benchmark and State of the Art". In: *Proceedings of the IEEE* 105.10 (2017), pp. 1865–1883.

[5] M. Cimpoi, S. Maji, I. Kokkinos, S. Mohamed, and A. Vedaldi. "Describing Textures in the Wild". In: *2014 IEEE Conference on Computer Vision and Pattern Recognition*. 2014, pp. 3606–3613.

[6] A. El Baz et al. "Lessons learned from the NeurIPS 2021 MetaDL challenge: Backbone fine-tuning without episodic meta-learning dominates for few-shot learning image classification". In: *NeurIPS 2021 Competitions and Demonstrations Track*. Vol. 176. Proceedings of Machine Learning Research. 2022, pp. 80–96.

[7] M. Fritz, E. Hayman, B. Caputo, and J. Eklundh. *THE KTH-TIPS database*. https://www.csc.kth.se/cvap/databases/kth-tips/index.html. 2004.

[8] J. Gamper, N. A. Koohbanani, K. Benet, A. Khuram, and N. Rajpoot. "PanNuke: an open pan-cancer histology dataset for nuclei instance segmentation and classification". In: *European Congress on Digital Pathology*. Springer. 2019, pp. 11–19.

[9] J. Gamper, N. A. Koohbanani, S. Graham, M. Jahanifar, S. A. Khurram, A. Azam, K. Hewitt, and N. M. Rajpoot. *PanNuke Dataset Extension, Insights and Baselines*. 2020. arXiv: 2003.10778.

[10] C. Garcin, alexis joly, P. Bonnet, A. Affouard, J.-C. Lombardo, M. Chouet, M. Servajean, T. Lorieul, and J. Salmon. "Pl@ntNet-300K: a plant image dataset with high label ambiguity and a long-tailed distribution". In: *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*. 2021.

[11] E. Gundogdu, B. Solmaz, V. Yücesoy, and A. Koç. "MARVEL: A Large-Scale Image Dataset for Maritime Vessels". In: *Computer Vision – ACCV 2016*. Vol. 10115. Lecture Notes in Computer Science. 2017, pp. 165–180.

[12] K. He, X. Zhang, S. Ren, and J. Sun. "Deep Residual Learning for Image Recognition". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition*. 2016, pp. 770–778.

[13] D. P. Hughes and M. Salathé. *An open access repository of images on plant health to enable the development of mobile disease diagnostics through machine learning and crowdsourcing*. 2015. arXiv: 1511.08060.

[14] A. Khosla, N. Jayadevaprakash, B. Yao, and L. Fei-Fei. "Novel Dataset for Fine-Grained Image Categorization". In: *First Workshop on Fine-Grained Visual Categorization, IEEE Conference on Computer Vision and Pattern Recognition*. Colorado Springs, CO, 2011.

[15] J. Krause, M. Stark, J. Deng, and L. Fei-Fei. "3D Object Representations for Fine-Grained Categorization". In: *2013 IEEE International Conference on Computer Vision Workshops*. 2013, pp. 554–561.

[16] G. Kylberg. *The Kylberg Texture Dataset v. 1.0*. External report (Blue series) 35. Centre for Image Analysis, Swedish University of Agricultural Sciences and Uppsala University, Uppsala, Sweden, 2011.

[17] S. Lazebnik, C. Schmid, and J. Ponce. "A sparse texture representation using local affine regions". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27.8 (2005), pp. 1265–1278.

[18] H. Li, X. Dou, C. Tao, Z. Wu, J. Chen, J. Peng, M. Deng, and L. Zhao. "RSI-CB: A Large-Scale Remote Sensing Image Classification Benchmark Using Crowdsourced Data". In: *Sensors* 20.6 (2020).

[19] Y. Long, Y. Gong, Z. Xiao, and Q. Liu. "Accurate object localization in remote sensing images based on convolutional neural networks". In: *IEEE Transactions on Geoscience and Remote Sensing* 55.5 (2017), pp. 2486–2498.

[20] P. Mallikarjuna, A. T. Targhi, M. Fritz, E. Hayman, B. Caputo, and J. Eklundh. *THE KTH-TIPS 2 database*. https://www.csc.kth.se/cvap/databases/kth-tips/index.html. 2006.

[21] M. Nilsback and A. Zisserman. "Automated Flower Classification over a Large Number of Classes". In: *Sixth Indian Conference on Computer Vision, Graphics & Image Processing*. 2008, pp. 722–729.

[22] J. A. Pandian and G. Geetharamani. "Data for: Identification of Plant Leaf Diseases Using a 9-layer Deep Convolutional Neural Network". In: *Mendeley Data, V1* (2019).

[23] P. Pérez, M. Gangnet, and A. Blake. "Poisson Image Editing". In: *ACM Transactions on Graphics* 22.3 (2003), pp. 313–318.

[24] L. Picek, M. Šulc, J. Matas, T. S. Jeppesen, J. Heilmann-Clausen, T. Læssøe, and T. Frøslev. "Danish Fungi 2020 - Not Just Another Image Recognition Dataset". In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 1525–1535.

[25] G. Piosenka. *100 Sports Image Classification*. https://www.kaggle.com/datasets/gpiosenka/sports-classification.

[26] G. Piosenka. *BIRDS 450 - SPECIES IMAGE CLASSIFICATION*. https://www.kaggle.com/datasets/gpiosenka/100-bird-species.

[27] S. Roopashree and J. Anitha. *Medicinal Leaf Dataset*. https://data.mendeley.com/datasets/nnytj2v3n5/1. Version 1. 2020.

[28] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra. "Grad-cam: Visual explanations from deep networks via gradient-based localization". In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 618–626.

[29] H. Serret, N. Deguines, Y. Jang, G. Lois, and R. Julliard. "Data quality and participant engagement in citizen science: comparing two approaches for monitoring pollinators in France and South Korea". In: *Citizen Science: Theory and Practice* 4.1 (2019).

[30] D. Singh, N. Jain, P. Jain, P. Kayal, S. Kumawat, and N. Batra. "PlantDoc: A Dataset for Visual Plant Disease Detection". In: *Proceedings of the 7th ACM IKDD CoDS and 25th COMAD*. 2020, pp. 249–253.

[31] H. M. Sosik, E. E. Peacock, and E. F. Brownlee. *Annotated Plankton Images - Data Set for Developing and Evaluating Classification Methods*. https://hdl.handle.net/10.1575/1912/7341. 2015.

[32] H. Sun, W.-W. Tu, and I. M. Guyon. "OmniPrint: A Configurable Printed Character Synthesizer". In: *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*. 2021.

[33] P. J. Thul, L. Åkesson, M. Wiking, D. Mahdessian, A. Geladaki, H. Ait Blal, T. Alm, A. Asplund, L. Björk, L. M. Breckels, et al. "A subcellular map of the human proteome". In: *Science* 356.6340 (2017).

[34] X. Wu, C. Zhan, Y. Lai, M.-M. Cheng, and J. Yang. "IP102: A Large-Scale Benchmark Dataset for Insect Pest Recognition". In: *2019 IEEE Conference on Computer Vision and Pattern Recognition*. 2019, pp. 8787–8796.

[35] Z. Wu. *Multi-type Aircraft of Remote Sensing Images: MTARSI*. https://zenodo.org/record/3464319#.YOFAAlJByrc. 2019.

[36] Y. Xian, C. H. Lampert, B. Schiele, and Z. Akata. "Zero-Shot Learning—A Comprehensive Evaluation of the Good, the Bad and the Ugly". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41.9 (2019), pp. 2251–2265.

[37] Z. Xiao, Y. Long, D. Li, C. Wei, G. Tang, and J. Liu. "High-resolution remote sensing image retrieval based on CNNs from a dimensional perspective". In: *Remote Sensing* 9.7 (2017).

[38] B. Yao, X. Jiang, A. Khosla, A. L. Lin, L. Guibas, and L. Fei-Fei. "Human action recognition by learning bases of action attributes and parts". In: *Proceedings of the IEEE international conference on computer vision*. 2011, pp. 1331–1338.

[39] Q.-s. Zhang and S.-C. Zhu. "Visual interpretability for deep learning: a survey". In: *Frontiers of Information Technology & Electronic Engineering* 19.1 (2018), pp. 27–39.

[40] B. Zieliński, A. Plichta, K. Misztal, P. Spurek, M. Brzychczy-Włoch, and D. Ochońska. "Deep learning approach to bacterial colony classification". In: *PLOS ONE* 12.9 (2017), pp. 1–14.