Data Efficient Learning: A comparison of Transfer Learning and Meta-Learning

Phan Anh Vu, Ihsan Ullah, Isabelle Guyon

Université Paris-Saclay
phan-anh.vu@universite-paris-saclay.fr

Abstract. Efficient learning from limited amount of data is crucial for machine learning. One way to tackle this is **transfer learning** by transferring knowledge learned from a large source dataset, to a smaller target dataset. However, this often leads to severe overfitting. Alternatively, **meta-leaning** algorithms are designed to learn efficiently in the low data regime. We present a technique to reduce overfitting while fine-tuning a pre-trained model, along with a comparison of transfer learning and meta-learning for few-shot image classification in multiple domains.

Keywords: few-shot learning, meta-learning, transfer learning, fine-tuning

1 Introduction

Many machine learning techniques require large amount of data to perform well. Meanwhile, collecting labeled data is difficult, tedious and costly. In application domains such as medical imaging, sufficient data for rare diseases is not available. In the low data regime, a common approach is to transfer knowledge learned from a large dataset.

Image classifiers often rely on models pre-trained on Imagenet [17], a large scale dataset, as a starting point. The pre-trained models can be used as fixed feature extractors, which tend to work better when the target dataset is more similar to the source dataset. Alternatively, in order to learn more domain-specific features, the pre-trained models can also be fine-tuned on the target dataset. Besides transfer learning, few-shot learning studies the problem of learning new task with a limited amount of data. Novel meta-learning algorithms have recently been proposed for this purpose.

Our work focuses on data-efficient learning on many real world applications such as bio-medicine, ecology, remote sensing, manufacturing and optical character recognition (OCR). While fine-tuning a pre-trained model, injecting noise via Dropout [20] layer and choosing an appropriate value for the weight decay help reduce overfitting. We also compare transfer learning and meta-learning approaches for tackling problems with small amount of data.

2 Transfer Learning by fine-tuning a pre-trained model

We use a pre-trained model as fixed feature extractor, and notice a poor performance. Fine-tuning the pre-trained model increases the performance, at the cost of severe overfitting. While fine-tuning a pre-trained network, using strong weight decay helps reduce overfitting. Another simple and effective technique is to inject Dropout [20] between higher convolution layers while retraining the whole network.
Different combination of weight decay and dropout are shown in Figure 1. Experiments are performed with a Resnet18 [7] architecture pre-trained on Imagenet [17]. We inject dropout layer between every convolution layer in the last 2 convolution blocks, and retrain the whole network. The best combinations is low dropout probability (0.1 between every convolution layer for the last 2 convolution blocks) and strong weight decay (10). This setting offers the best trade-off between validation score and overfitting.

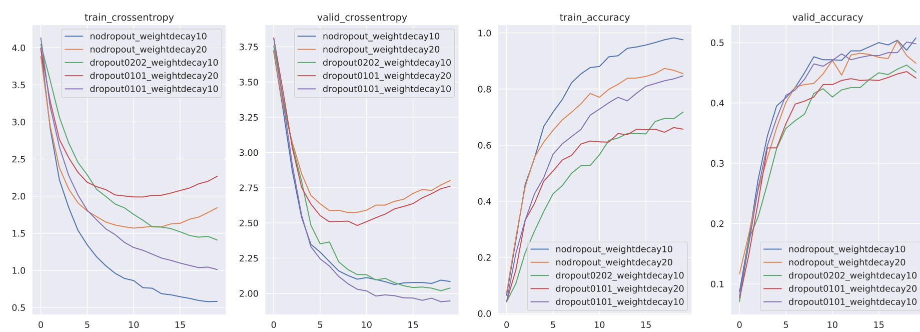


Fig. 1: Effects of weight decay and dropout on a skin disease dataset

3 Transfer learning vs meta-learning

In this section, we compare the results of transfer learning by fine-tuning a pre-trained model, with meta-learning techniques. We use the 5-way 5-shot settings, as explained in the Prototypical Network paper [19]. The Baseline model is trained on the subset of classes (5 classes) from one task. The Fine-tuned model is trained on all meta-train classes together. In the meta-test phase, we freeze the convolution body, and train only a linear layer as classifier head on every task. MAML [5] aims to find good parameter initialization which can adapt quickly to new task (with only a few update steps). Prototypical network [19] learns an embedding function, then predict based on the distance between embedding of an example and the embedding centroids (prototypes) of each class.
Prototypical Network is the best overall model. Surprisingly, Fine-tuning is quite competitive with this best approach. One remarkable exception is the

Omniprint datasets from OCR domain, where Prototypical Network outperforms all other methods by a large margin. Omniprint datasets [21] are synthetic datasets generated by software, and are very different from the source domain (Imagenet). Meta-learning algorithms seem to have an edge in this case. Figure 2 shows results for each dataset. More details about the experiments and datasets can be found in Appendix A and Appendix B respectively.

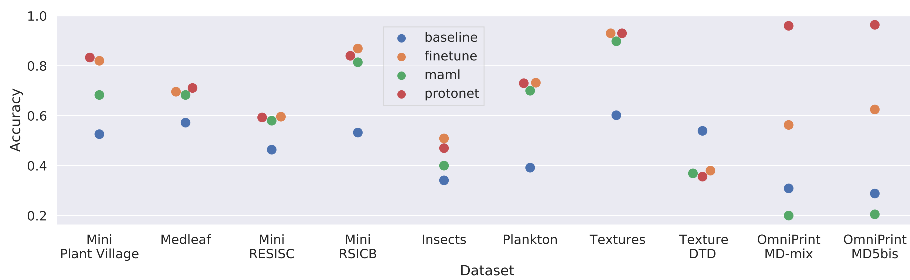


Fig. 2: Transfer learning and meta-learning result

4 Discussion and Conclusion

Fine-tuning a pre-trained model with Imagenet is quite competitive compared to meta-learning techniques for few-shot image classification. Moreover, meta-baseline authors [1] also reach the same conclusion. Meanwhile, when the target dataset is very different from the source dataset which the pre-trained model learns from (often Imagenet), meta-learning seems to work better than transfer learning. This finding echoes another publication comparing transfer learning and meta-learning [4]. While fine-tuning a pre-trained model, inserting Dropout and choosing appropriate weight decay help reduce overfitting significantly. For future work, a cross-domain comparison of performance with different algorithms is an interesting direction.

5 Acknowledgements and disclosure of funding

We gratefully acknowledge Haozhe Sun for providing the Omniprint datasets and all other data owners/creators. We received useful input from many members of the TAU team from the LISN laboratory, Wei Wei Tu from 4Paradigm Inc. China, the MetaDL technical crew: Mike Huisman, Felix Mohr, Jan N. van Rijn, Aske Plaat, Adrian El Baz, Zhengying Liu, Joaquin Vanschoren, Jennifer (Yuxuan) He and Benjia Zhou.
This work was supported by ChaLearn, the ANR (Agence Nationale de la Recherche, National Agency for Research) under AI chair of excellence HUMANIA, grant number ANR-19-CHIA-35400222 and Labex Digicosme project

ANR11LABEX0045DIGICOSME operated by ANR as part of the program Investissement d'Avenir Idex Paris Saclay (ANR11IDEX000302).

References

[1] Yinbo Chen et al. "A new meta-baseline for few-shot learning". In: *arXiv preprint arXiv:2003.04390* (2020).

[2] Gong Cheng, Junwei Han, and Xiaoqiang Lu. "Remote Sensing Image Scene Classification: Benchmark and State of the Art". In: *Proceedings of the IEEE* 105.10 (Oct. 2017), pp. 1865–1883. ISSN: 1558-2256. DOI: 10.1109/jproc.2017.2675998. URL: http://dx.doi.org/10.1109/JPROC.2017.2675998.

[3] Mircea Cimpoi et al. *Describing Textures in the Wild*. 2013. arXiv: 1311.3618 [cs.CV].

[4] Vincent Dumoulin et al. "Comparing Transfer and Meta Learning Approaches on a Unified Few-Shot Classification Benchmark". In: *arXiv preprint arXiv:2104.02638* (2021).

[5] Chelsea Finn, Pieter Abbeel, and Sergey Levine. "Model-agnostic meta-learning for fast adaptation of deep networks". In: *International Conference on Machine Learning*. PMLR. 2017, pp. 1126–1135.

[6] Mario Fritz et al. "THE KTH-TIPS database". In: 2004. URL: https://www.csc.kth.se/cvap/databases/kth-tips/index.html.

[7] Kaiming He et al. "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.

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

[9] David P. Hughes and Marcel Salath'e. "An open access repository of images on plant health to enable the development of mobile disease diagnostics through machine learning and crowdsourcing". In: *CoRR* abs/1511.08060 (2015). arXiv: 1511.08060. URL: http://arxiv.org/abs/1511.08060.

[10] Gustaf 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, Sept. 2011. URL: http://www.cb.uu.se/~gustaf/texture/.

[11] Svetlana Lazebnik, Cordelia Schmid, and Jean Ponce. "A sparse texture representation using local affine regions". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27.8 (Aug. 2005), pp. 1265–1278. DOI: 10.1109/TPAMI.2005.151. URL: https://hal.inria.fr/inria-00548530.

[12] Haifeng Li et al. "RSI-CB: A Large-Scale Remote Sensing Image Classification Benchmark Using Crowdsourced Data". In: *Sensors* 20.6 (2020), p. 1594. DOI: doi.org/10.3390/s20061594.

[13] P. Mallikarjuna et al. “THE KTH-TIPS 2 database”. In: 2006. URL: https://www.csc.kth.se/cvap/databases/kth-tips/index.html.

[14] J. Arun Pandian and G. Geetharamani. “Data for: Identification of Plant Leaf Diseases Using a 9-layer Deep Convolutional Neural Network”. In: *Mendeley Data, V1* (2019). DOI: 10.17632/tywbtsjrjv.1.

[15] Patrick Pérez, Michel Gangnet, and Andrew Blake. “Poisson Image Editing”. In: *ACM Trans. Graph.* 22.3 (July 2003), pp. 313–318. ISSN: 0730-0301. DOI: 10.1145/882262.882269. URL: https://doi.org/10.1145/882262.882269.

[16] S. Roopashree and J. Anitha (2020). *Medicinal Leaf Dataset*. DOI: 10.17632/nnytj2v3n5.1.

[17] Olga Russakovsky et al. “ImageNet Large Scale Visual Recognition Challenge”. In: *International Journal of Computer Vision (IJCV)* 115.3 (2015), pp. 211–252. DOI: 10.1007/s11263-015-0816-y.

[18] Hortense Serret et al. “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), p. 22.

[19] Jake Snell, Kevin Swersky, and Richard Zemel. “Prototypical Networks for Few-shot Learning”. In: *Advances in Neural Information Processing Systems*. 2017.

[20] Nitish Srivastava et al. “Dropout: a simple way to prevent neural networks from overfitting”. In: *The journal of machine learning research* 15.1 (2014), pp. 1929–1958.

[21] Haozhe Sun, Wei-Wei Tu, and Isabelle M Guyon. “OmniPrint: A Configurable Printed Character Synthesizer”. In: *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 1)*. 2021. URL: https://openreview.net/forum?id=R07XwJPmgpl.

A Experiment result

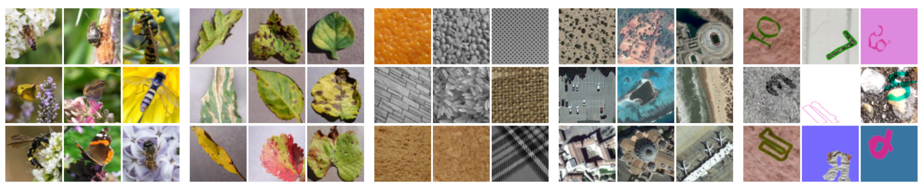
Table 1: **Experiment result.** Average meta-test accuracy (%) over many tasks and 0.95 confidence interval. We repeat experiments with 3 different random seeds. Bold entries denote the best performance per dataset.

Domain	Dataset	Finetuning	MAML	ProtoNet	Baseline
Ecology	Insects	**50.9±0.7**	40.0±0.7	47.0±0.7	34.1±1.0
	Plankton	**73.2±0.7**	70.0±0.8	73.0±0.8	39.2±2.4
Bio-medicine	Mini Plant Village	82.0±0.5	68.3±0.6	**83.3±0.4**	52.6±4.8
	Medleaf	69.6±0.6	68.3±0.6	**71.1±0.5**	57.2±7.8
Manufacturing	Textures	**93.0±0.4**	89.8±0.6	**93.0±0.4**	60.2±3.9
	Texture DTD	**38.0±0.5**	36.9±0.5	35.6±0.5	53.9±3.3
Remote sensing	Mini RESISC	**59.6±0.7**	58.0±0.6	59.3±0.6	46.4±3.9
	Mini RSICB	**86.9±0.5**	81.4±0.6	84.0±0.5	53.2±4.7
OCR	OmniPrint-MD-mix	56.3±0.7	20.0±0.3	**96.0±0.3**	30.9±0.5
	OmniPrint-MD-5-bis	62.5±0.8	20.5±0.4	**96.4±0.3**	28.8±0.4

The 0.95 confidence interval is calculated as $t\frac{s}{\sqrt{n}}$. Here, s is the sample standard deviation of validation scores for all tasks from a dataset. n is the number of meta-test classification tasks for a dataset. t is the t-value at level 0.975 of the Student's t-distribution with $n-1$ degrees of freedom.

B Datasets

Details of the datasets used for these experiments, their original sources and how the datasets are curated are explained separately in the following sections. All datasets have exactly 40 images per category/class. The datasets come from 5 domains: ecology, bio-medicine, manufacturing, remote sensing and optical character recognition. All the datasets used for this experiment are part of the Meta-Album benchmark for few shot image classification (Pending NeurIPS submission). Figure 3 shows sample images from each dataset.



(a) Insects (c) Plant Village (e) Textures (g) RESISC (i) OmniPrint MD-mix

(b) Plankton (d) Medleaf (f) Texture DTD (h) RSICB (j) OmniPrint MD-5-bis

Fig. 3: Sample images for all datasets

B.1 Ecology

Insects The original Insects dataset [18] is createad by National Museum of Natural History, Paris[1]. 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 a vegetation. The images are collected by the researchers and hundreds volunteers of SPIPOLL Science project[2]. The images uploaded to a centralized server either by using the SPIPOLL website[2], Android[3] or IOS[4] mobile application.

[1] https://www.mnhn.fr/fr
[2] https://www.spipoll.org/
[3] https://play.google.com/store/apps/details?id=fr.eneo.spipoll&hl=fr
[4] https://apps.apple.com/fr/app/spipoll/id1495843067

The Meta-Album benchmark insect dataset is prepared from the original Insects dataset by carefully preprocessing the images i.e cropping the images from either sides to make squared images. These cropped images are then resized into 128x128 using open-cv with anti-aliasing filter. These preprocessed images are used in the final insects dataset which has 114 categories and total 4,560 images.

Plankton The Plankton dataset is created by the researchers at the Woods Hole Oceanographic Institution[5]. Imaging FlowCytobot (IFCB) was used for the data collection. Complete process and mechanism is described here [8]. 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 benchmark plankton dataset is prepared from the original WHOI Plankton's 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 on 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. The final plankton dataset is extracted from these preprocessed images and it has 91 categories and total 3,640 images.

B.2 Bio-medicine

Mini Plant Village The Plant Village dataset [6] [9, 14] 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 labels are captured on a variety of gray background, except Corn Common rust which has black background. For the curated version, we exclude the irrelevant Background and Corn Common Rust classes from the original collection. Plant Village has 2-level label hierarchy, the supercategory is the crop type and the category is the disease type. We create Mini Plant Village for Meta-Album by resizing a subset from the original dataset to 128x128 px.

Medleaf The Medicinal Leaf Dataset [7] [16] gather 30 species of healthy and mature medicinal herbs. The leaves are plucked from different plants of the same species, then placed on white uniform background. There are around 1800 images in total, captured with a mobile phone camera. The original resolution is

[5] https://www.whoi.edu/
[6] https://data.mendeley.com/datasets/tywbtsjrjv/1
[7] https://data.mendeley.com/datasets/nnytj2v3n5/1

1600x1200 px. We create Medleaf for Meta-Album by selecting 27 classes having more than 40 images. Then we crop them at the center and resize to 128x128 px.

B.3 Remote sensing

Mini RESISC RESISC45 [8] [2] gathers 700 RGB images of size 256x256 px for each of 45 scene categories. The data authors strives to providing 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 create Mini RESISC for Meta-Album by resizing a subset from the original dataset to 128x128 px.

Mini RSICB RSICB128 [9] [12] covers 45 scene categories, assembling in total 36000 images of resolution 128x128 px. The data authors select various locations around the world, and follow China's land-use classification standard. This collections has 2-level label hierarchy with 6 supercategories: agricultural land, construction land and facilities, transportation and facilities, water and water conservancy facilities, woodland, and other lands. We create Mini RSICB for Meta-Album by resizing a subset from the original dataset to 128x128 px.

B.4 Manufacturing

Textures The original Textures dataset is a combination of4 texture datasets: *KTH-TIPS*[10][6], *KTH-TIPS 2*[10][13], *Kylberg Textures Dataset*[11][10] and *UIUC Tectures Dataset*[11]. The data in all four datasets is collected in laboratory conditions, i.e. images were captured incontrolled 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. Both of these datasets were prepared under the supervision of Eric Hayman and Barbara Caputo. The data for *Kylberg Textures Dataset* and *UIUC Tectures Dataset* data was collected by the original authors of these datasets. *KTH-TIPS* and *KTH-TIPS 2* datasets were created in 2004 and 2006 respectively. *Kylberg Textures Dataset* was created in September 2010 and *UIUC Tectures Dataset* in August 2005.
The Meta-Album Textures dataset is a subset of the original dataset(combination of 4 datasets). All the images are pre-processed by first cropping into perfect squared images and then resized into 128x128 with anti-aliasing filter. A subset from the original dataset is extracted with 40 images per class. The final dataset has 64 categories and total 2,560 images.

---

[8] https://gcheng-nwpu.github.io/
[9] https://github.com/lehaifeng/RSI-CB
[10] https://www.csc.kth.se/cvap/databases/kth-tips/index.html
[11] http://www.cb.uu.se/~gustaf/texture/

Texture DTD The Textures DTD dataset is a large textures dataset[12] which consists of 5,640 images. The data is collected from Google[13] and Flicker[14] by the original authors of the paper "Describing Textures in the Wild"[3]. The data was annotated using Amazon Mechanical Turk[15]. The data collection process is mentioned on the dataset overview page[12].
For the Meta-Album benchmark, a subset of this dataset is used after preproessing it i.e. cropping the images to square images and the resizing them to 128x128 using open-cv with anti-aliasing filter. This dataset has 47 categories and total 1,880 images.

B.5 Optical character recognition

OmniPrint-MD-mix dataset consists of 28,240 images (128x138, RGB) from 706 chacracter categories. The images are synthesized with OmniPrint [21], 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. All images/textures consists of photos taken by a personal mobile phone [21].
OmniPrint-MD-5-bis dataset consists of 28,240 images (128x138, RGB) from 706 character categories. The images are synthesized with OmniPrint [21], 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 [15], rotation is within -60 and 60 degrees, horizontal shear is within -0.5 and 0.5, foreground is filled with random color, background consists of images downloaded from Pexels.

[12] https://www.robots.ox.ac.uk/~vgg/data/dtd/index.html
[13] https://images.google.com/
[14] https://www.flickr.com/
[15] https://www.mturk.com/