Domain Agnostic Few-Shot Learning For Document Intelligence

Jaya Krishna Mandivarapu, Eric Bunch, Glenn Fung
American Family Insurance, Machine Learning Research Group
jmandivarapu1@student.gsu.edu , {ebunch, qyou, gfung}@amfam.com

ABSTRACT

Few-shot learning aims to generalize to novel classes with only a few samples with class labels. Research in few-shot learning has borrowed techniques from transfer learning, metric learning, meta-learning, and Bayesian methods. These methods also aim to train models from limited training samples, and while encouraging performance has been achieved, they often fail to generalize to novel domains. Many of the existing meta-learning methods rely on training data for which the base classes are sampled from the same domain as the novel classes used for meta-testing. However, in many applications in the industry, such as document classification, collecting large samples of data for meta-learning is infeasible or impossible. While research in the field of the cross-domain few-shot learning exists, it is mostly limited to computer vision. To our knowledge, no work yet exists that examines the use of few-shot learning for classification of semi-structured documents (scans of paper documents) generated as part of a business workflow (forms, letters, bills, etc.). Here the domain shift is significant, going from natural images to the semi-structured documents of interest. In this work, we address the problem of few-shot document image classification under domain shift. We evaluate our work by extensive comparisons with existing methods. Experimental results demonstrate that the proposed method shows consistent improvements on the few-shot classification performance under domain shift.

ACM Reference Format:
Jaya Krishna Mandivarapu, Eric Bunch, Glenn Fung, American Family Insurance, Machine Learning Research Group, jmandivarapu1@student.gsu.edu , {ebunch, qyou, gfung}@amfam.com, . 2022. Domain Agnostic Few-Shot Learning For Document Intelligence. *ACM Trans. Graph.* 37, 4, Article 111 (August 2022), 5 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

The challenges of document classification in an industry setting are many: scalability, accuracy and degree of automation, speed of delivery requirements, and limited time available by domain business experts. While a company may have a large collection of documents and high level metadata consisting of broad classes spanning this collection, there are inevitably use cases and workflows that need more granular and specialized document classes. The document classification within these specialized workflows is typically done by business and domain experts, whose time is valuable, and better spent on other tasks. This setup calls for an automation process that can be trained to classify document sub-classes with smaller amounts of training data. Few-shot learning methods offer exactly this benefit.

One possible approach is to pre-train a model on available large open-source document datasets. However, they tend to be significantly different from the internally generated workflow documents. Some reasons for this are that the open source document data sources may be scanned using lower quality scanners, the documents are of different types than those considered internal to the company, or the text content itself utilizes a different colloquial vocabulary. This paper proposes a few-shot meta-learning technique that utilizes both the visual and text components of a document, and is pre-trained on open source document datasets that are out of domain with respect to the internal company documents, which the model is evaluated on.

Author's address: Jaya Krishna Mandivarapu, Eric Bunch, Glenn Fung
American Family Insurance, Machine Learning Research Group
jmandivarapu1@student.gsu.edu , {ebunch, qyou, gfung}@amfam.com
.

2 RELATED WORK

This work explores a method which can be used for few-shot learning on multi-channel document data, in which the meta-training is done on a distinct domain of open source documents. Work has been done in this area addressing the separate problems of: a) Meta-training a few-shot model on document data. b) Combining visual and textual feature channels via canonical correlation c)Domain adaptation of models trained on image data.

This work proposes to address all of these issues with a single approach, driven by a distinct business need. Here we detail previous work done in each of these directions.

2.1 Meta-learning

Meta-learning has been a powerful tool to answer the challenge of the large data requirements that many deep learning models seem to face. So far, many of the applications of deep meta-learning have been in few-shot image classification [Finn et al. 2019; Ravi and Larochelle 2016; Snell et al. 2017]. Meta-learning for few shot image classification has often been evaluated on data sets such as ImageNet [Russakovsky et al. 2015], CIFAR-10 and CIFAR-100 [Krizhevsky 2009], and Omniglot [Lake et al. 2011]. However, there has been little done to apply the methods of meta-learning to industry level document images.

2.2 Domain adaptation

In the traditional machine learning setting, the data samples used for training and testing an algorithm are assumed to come from the same distribution. In practical applications however, this is not always a valid assumption; the data available for training may fall into a different distribution than the data the model is expected to perform on in a live system. A typical example of this is a model which is trained on an open source data set is then desired to be used for inference in a smaller, proprietary data set, perhaps for a slightly different task. Domain adaptation is a subfield of machine learning that attempts to overcome this challenge. Typical

approaches include transfer learning [Pan and Yang 2010], semi-supervised learning [Pise and Kulkarni 2008; Zhu 2008], multi-task learning [Caruana 2004], and meta-learning [Huisman et al. 2021].

3 METHODOLOGY & PROBLEM DEFINITION

In this section, we briefly formally describe our proposed method, Cross Domain Few-Shot Learning using **D**eep **C**anonical **C**orrelation for **D**ocument **I**ntelligence dubbed as **DCCDI** in rest of the paper. Formally, a few-shot learning problem is denoted as $P = (\mathcal{D}_{\text{source}}, \mathcal{D}_{\text{target}})$; $\mathcal{D}_{\text{source}}$ is the meta-train set from where base classes as sampled for episodic training. Novel classes during meta-test are sampled from the target domain $\mathcal{D}_{\text{target}}$, such that $D_{\text{source}} \cap D_{\text{target}} = \emptyset$. For brevity, we will define the ***domain*** of the source dataset $\mathcal{D}_{source}$ to be

$$d_{source} = \{\mathcal{X}, \mathcal{Y}, P_{source}\} \tag{1}$$

where $\mathcal{X}$ is the feature space of all the inputs in d-dimensional space; $\mathcal{X} \subset \mathbb{R}^d$ and $\mathcal{Y}$ is the label space of all the labels; $\mathcal{Y} \subset \{1, \dots C\}$ where is the number of classes, and P_{source} is the joint probability distribution over the feature,label pairs of $\{\mathcal{X}, \mathcal{Y}\}$ denoted by $p(x, y)$. A similar definition can be made for the target dataset.

We focus on few-shot settings for document classification using a model f with parameters θ denoted as f_θ, updated via meta learning tasks using episodic training from the meta-train set $\mathcal{D}_{source}$ and aim to demonstrate generalization to novel classes present in the meta-test set $\mathcal{D}_{target}$.

During meta training the model f_θ is provided with a wide range of classification tasks $\mathcal{T}_i$ drawn from the dataset $\mathcal{D}_{\text{source}} = \{\mathcal{T}_1, \dots, \mathcal{T}_n\}$ where each episodic task is $\mathcal{T}_i = \{(x_1^i, y_1^i), \dots, (x_k^i, y_k^i)\}$, and where x_i represents image i and y_i its corresponding label. Each task $\mathcal{T}_i$ is further partitioned into a *support set* $\mathcal{S}_i$ used for training, and a *query set* $\mathcal{Q}_i$ used for testing. That is, $\mathcal{T}_i$ can be written as the disjoint union $\mathcal{T}_i = \mathcal{S}_i \cup \mathcal{Q}_i$. Overall, $\mathcal{D}_{source}$ can be written as $\mathcal{D}_{\text{source}} = \{(\mathcal{S}_1, \mathcal{Q}_1), \dots, (\mathcal{S}_n, \mathcal{Q}_n)\}$. We follow the conventional way of preparing the support and query sets for each task ($\mathcal{T}_i$), which is a C-way, N-shot classification problem in which C classes are randomly drawn from the entire set of classes from $\mathcal{D}$. Furthermore, for each of the sampled classes, N and M examples are sampled for the support and query set respectively such that each task $\mathcal{T}_i$ consists of $(\mathcal{S}_i, \mathcal{Q}_i)$ where $\mathcal{S}_i = \{(\mathbf{x}_i, y_i)\}_{i=1}^{C \times N}$ is a support set consisting of N labeled images for each of the C classes and the query set $\mathcal{Q}_i = \{\tilde{\mathbf{x}}_i, \tilde{y}_i\}_{i=1}^{C \times M}$ with M samples per class and $y, \tilde{y} \in \{1, \dots, C\}$ are the corresponding class labels.

The cross-domain few-shot learning scheme matches closely with our real-world industry setting where the source domain $\mathcal{D}_{source}$ and the target domain $\mathcal{D}_{target}$ belong to different distributions. As in Eq. 1, the joint distribution of the source dataset is indicated as P_{source} and the target domain distribution can be denoted as P_{target}. Furthermore, as is the case in a cross-domain few-shot learning setting, $P_{source} \neq P_{target}$ and $\mathcal{Y}_s$ is disjoint from $\mathcal{Y}_t$. Also, similarly to a few-shot learning paradigm, during the episodic meta-training phase, the model is trained on a large number of tasks $\mathcal{T}_i$ sampled from the source domain $\mathcal{D}_{source}$. During the meta-testing phase, the model is presented with a support set $S = \{x_i, y_i\}_{i=1}^{K \times N}$ consisting of N examples from K novel classes and $\mathcal{Q} = \{x_i, y_i\}_{i=1}^{K \times M}$ consisting of M examples which are very different from the meta-training classes. After the meta-trained model $\hat{f}_\theta$ is adapted to the support set, a query set from novel classes is used to evaluate the model performance.

3.1 Canonical Correlation

Ideally, an effective document classification method needs to leverage both textual and image (including layout) information. Document image consists of two feature channels; a visual channel, and a text channel. Each has useful information that can be leveraged for document classification. When using deep convolutional neural networks for document image classification, the document is treated as an image to get the visual feature vector of the document images. On the other hand, all the text from the document image is extracted, converted into tokens and passed through a BERT-like pre-trained transformer-based language model to obtain textual features. Some of the ways of utilizing both the textual and visual features during the classification are to concatenate or average them before passing them through the final classification layer. Some of the disadvantages of doing this are a) More computational resources are required for model training for large dimensional features b) Difficult to maintain the synchronization between both the visual and textual modalities, which might impact model performance.

We address this dilemma by introducing the Deep Canonical Correlation for Document Intelligence Module (DCCDI) during meta-test phase to represent a document utilizing both the textual and visual features. By using the proposed DCCDI module, we produce highly correlated transformations of multiple modalities of data such as textual and visual using complex non-linear transformations. Canonical correlation was proposed by Hotelling [Hotelling 1992] which in a sense aligns two vectors via projections in such a way that the projections are maximally correlated. It is a widely used technique in the statistics community to measure the linear relationship between two multidimensional variables, used in finding linear projections of two multidimensional vectors that are maximally correlated. We use deep canonical co-relation method proposed by [Andrew et al. 2013] in our DCCM module with the goal of achieving fine-grained cross-modality alignment between the visual and textual modalities. As shown in RIGHT Figure 1; $V \in \mathbb{R}^{N \times d_1}$ is the multidimensional vector for the visual modality where d_1 is total number of dimensions and $T \in \mathbb{R}^{N \times d_2}$ is the multidimensional vector for the textual modality where d_2 is total number of dimensions. N is the total number of inputs available in each modality. The input multidimensional vectors in different modalities are transformed using two neural networks g with parameters ϕ_1, h with parameters ϕ_2

$$\mathcal{Z}_1 = g_{\phi_1}(V), \quad \mathcal{Z}_2 = h_{\phi_2}(T) \tag{2}$$

$\mathcal{Z}_1 \in \mathbb{R}^{N \times d}$ and $\mathcal{Z}_2 \in \mathbb{R}^{N \times d}$ are the d dimensional outputs of the neural networks. Both the neural networks g, h are optimized jointly with a goal of making the correlation between $\mathcal{Z}_1$ and $\mathcal{Z}_2$ as high as possible:,

$$(\phi_1^*, \phi_2^*) = \arg\max_{\phi_1, \phi_2} \text{corr}(g_{\phi_1}(V), h_{\phi_2}(T))$$

3.2 DCCDI Model

Our primary focus in this work is to improve the generalization ability of few-shot classification models to unseen domains by

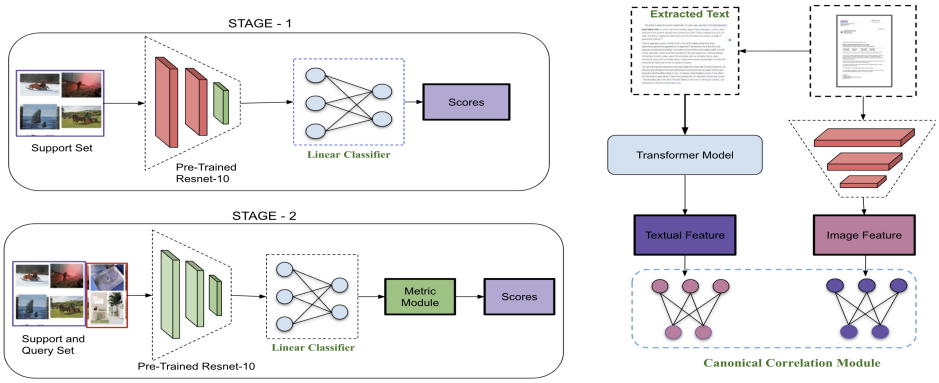

Figure 1: LEFT(Stage-1):Episodic training of last k layers of ResNet-10 (shown in green color) by support set. Left(Stage-2): Episodic training paradigm train all layers of ResNet-10 along with Metric-Learning Module using both support and Query sets. RIGHT: Canonical Correlation Block: Image Features and Textual features was trained using canonical correlation loss

Table 1: Few Classification accuracy on the INSR, miniRVL dataset when source domain is miniImagenet and tieredImagenet.

		miniImagenet dataset						**tieredImageNet**					
		INSR Dataset (5-way)			RVL Dataset (5-way)			INSR Dataset (5-way)			RVL Dataset (5-way)		
Baselines	**Model**	**5-shot**	**10-shot**	**20-shot**	**5-shot**	**10-shot**	**20-shot**	**5-shot**	**10-shot**	**20-shot**	**5-shot**	**10-shot**	**20-shot**
ProtoTypical Networks	Conv-4	57.97 %	62.21 %	65.4%	44.64%	48.85 %	53.42 %	61.09 %	61.41 %	65.18%	44.42%	46.94 %	54.50 %
ProtoTypical Networks	Resnet-10	50.04 %	53.33%	52.78%	49.50 %	52.58 %	53.25 %	52.21 %	52.52%	53.76%	46.86 %	46.18 %	52.30 %
Relational Networks	Conv-4	55.47 %	56.58 %	57.27%	41.28 %	46.93 %	49.10 %	50.28 %	56.82 %	60.29%	41.30 %	48.61 %	47.61 %
Relational Networks	ResNet-10	32.42 %	43.08%	46.53%	33.08 %	34.68 %	37.57%	29.36%	39.36 %	54.82%	26.58 %	29.81 %	44.29 %
Matching Networks	ResNet-10	48.53 %	50.21%	58.92%	40.57 %	44.76 %	52.21%	40.09 %	45.88%	48.10 %	33.68%	42.37 %	43.92%
DCCDI without textual	ResNet-10	**65.21** %	**70.93**%	**77.2**%	**60.85** %	**66.45** %	**70.32** %	**65.73** %	**71.75**%	**77.11**%	**61.32** %	**66.92** %	**71.04** %
DCCDI	ResNet-10	**67.82** %	**72.79** %	**79.78**%	**61.76** %	**67.40** %	**72.93** %	**66.68** %	**74.01** %	**78.85** %	**62.05** %	**67.55** %	**72.61**%

learning a prior on the model weights which is suitable for Meta-Fine-tuning during the meta-testing phase on document datasets. We have also proposed a canonical-correlation-based layer in the model to integrate effectively both the textual and visual modalities of the document images which can be seen as a fine-grained cross-modality alignment task.

3.2.1 Domain Agnostic Meta-Learning for Document Intelligence. We aim to train a model that can adapt swiftly to novel unseen classes. This problem setting is often formalized as cross domain few-shot learning. In this proposed approach, the model is meta-trained on a set of tasks generated using $\mathcal{D}_{source}$, such that the meta-trained model can quickly adapt to new unseen novel tasks using only a small number of examples or trials generated using $\mathcal{D}_{target}$. In this section, we formally state the problem and present the general form of our algorithm. Similar to Meta-Learning algorithms the proposed algorithm can be subdivided into following phases.

3.2.2 Meta-Training Phase. We used ResNet-10 as our visual feature extractor or encoder. It have been shown recently that this pre-training process significantly improves the generalization [Gidaris and Komodakis 2018; Lifchitz et al. 2019; Rusu et al. 2018]. We pre-train the visual feature encoder on a source dataset (miniImageNet or tiredImageNet) by incorporating a final linear layer.

After the pre-training stage, we start our meta-training process of few-shot classification training stage. First, we train and update the last k layers of the visual feature encoder E followed by a linear classifier layer. We minimize the standard cross-entropy loss on the meta-training dataset by using only the support set images as shown in the Stage-1 of Figure 1. After this Stage-1 training process, all the layers of the visual feature encoder block of the model f_θ will be unfrozen.

In the Stage-2 phase, we train the proposed model using the traditional episodic meta-learning paradigm. For each episode a new task $\mathcal{T}_i$ is sampled from $\mathcal{D}_{source}$, the model f_θ is trained with N samples present in the support set. The model is then tested on query samples from the same task. The prior parameters of the model f are then updated by considering the test error on the query set. Actually, the test error on sampled tasks $\mathcal{T}_i$ serves as the training error of the meta-learning process. All the parameters in the network are updated using the MAML [Finn et al. 2017] first order approach. For this stage-2, we proceed similarly to [Cai and Shen 2020; Chen et al. 2021; Guo et al. 2020] which successfully use a metric mapping module to project the final linear classifier scores onto a metric space which can be used to compare support and query samples, hence increasing the overall accuracy. A graph neural network is used for the Metric-Learning module which is similar to architecture used in few-shot graph neural networks [Garcia and Bruna 2018].

3.2.3 Meta-Testing or Meta-Deployment Phase. At the start of the meta-test phase the first l layers of the visual feature extractor was frozen and the last k layers are left unfrozen. With the main goal of adapting the meta-trained model for the business document domain, we introduce the CCDI module during our deployment phase. During Meta-Testing, a new task $\mathcal{T}_i$ is drawn from the $\mathcal{D}_{target}$. The input document images are resized to 224 × 224 then fed into the visual feature blocks. Visual features are extracted for each of the document image present in the support set using the visual feature encoder block from the meta-trained model $\hat{f}_\theta$. Similarly for each of the document images, text is extracted using Pytesseract and then fed into pre-trained longformer model [Beltagy et al. 2020] to extract textual embedding features. Both the textual and visual modalities are passed through its corresponding deep Canonical Correlation bock and jointly optimized. Training the canonical co-relation block results in representations that aligns both the modalities (Image and text). The resulting meta-trained model along with the metric module, which consists of an ensemble of a graph convolution neural network classifier and a linear classifier layer is then trained on this data. Finally both the scores are combined and treated as the final classification scores.

4 EXPERIMENTS AND RESULTS

4.1 Datasets and Document Pre-processing

Robustness of the proposed approach has been tested using standardized few-shot classification datasets **miniImageNet** [Ravi and Larochelle 2016], **tieredImageNet** as the single source domain, and evaluate the trained model on two document domain datasets as target domain. MiniImageNet consists of 60,000 images from 100 classes,in which 64 classes for training, 16 classes for validation, 20 classes for the test. The TieredImageNet consists 608 classes in total 351 train classes, 97 validation classes and 160 test classes respectively. We use the following two datasets to evaluate our proposed model. The Insurance company dataset dubbed as **INSR** for anonymity contains 5772 document images which spans across 11 categories. Some Categories from the INSR dataset include Medical Bills, Medical Authorizations, Medical Records etc. The second dataset is a few-shot learning dataset dubbed as The **miniRVL** dataset [Harley et al. 2015] that has been generated from a larger RVL dataset which consists of 400,000 images which spans across 16 categories and 1000 images per class is taken for miniRVL dataset. The data relates to document classification and include Advertisements, Emails among other document types.Textual features for each document image was extracted using PyTesseract and passed through pre-trained longformer model *longformer-base-4096* [Beltagy et al. 2020] which can handle documents of thousands of tokens and the textual feature vector is then collected for each image.

4.2 Implementation Details:

. We compared the proposed method using visual and textual features to the proposed method using only visual features along with standard baselines. Prior to our meta training phase we also pretrained our image feature extractor by minimizing the standard supervised cross-entropy loss on the source domain dataset such as miniImageNet or TieredImageNet. This is similar to several recent works [Gidaris and Komodakis 2018; Lifchitz et al. 2019; Rusu et al. 2018] that have shown significant improvement in classification accuracy via this method. In all our experiments. We used the canonical correlation block containing two neural networks consisting of 3 layers. Both the network parameters were optimized together using canonical loss using the RMSprop optimizer with a learning rate of 0.001. Performance is measured by computing the average accuracy across 3 independent runs.

4.3 Comparing classification accuracy

4.3.1 MiniImageNet. Results comparing the baselines to our model on meta-trained on miniImagenet and deployed on document image datasets are shown in Table 1. For 5-shot 5-way, 10-shot 5-way, and 20-shot 5-way, our proposed model outperforms all the existing baselines. As shown in the Table 1 all the baseline models, Prototypical Networks, Matching Networks and Relational Networks work well when the embedded model is Conv-4 and the performance degrades rapidly when a Resnet-10 block was used as an embedding model for each metric-based baseline method. As shown, the baseline method which closest performance to our proposed approach Prototypical network which achieves 61.09% accuracy for (5-shot, 5-way). However, performance doesn't improve much for both the (10-shot, 5-way) and the (20-shot, 5-way) classification. The proposed CCDI Model without the canonical co-relation block achieved an accuracy of 65.73% and the one with the canonical co-relation block has achieved an state-of-the-art accuracy of 66.68% which shows the significance of proposed method during the meta-testing phase. Our experiments followed the same setup described above for meta-testing on the open source RVL dataset. Results are shown in Table 1. However the miniRVL dataset classification task is more challenging as it contains many documents that doesn't follow specific layout structures within classes. As shown Table 1 (b), the proposed method also outperforms the rest of the baseline methods and achieves state-of-the-art performance on this dataset by a large margin of up to 12% when compared to its closer competitor: Relational Networks.

4.3.2 TieredImageNet. To test the effectiveness of the proposed approach when meta-trained on a bigger domain, we repeated the experiments of the previous section. Results are shown in Table 1.

In comparison, our proposed method achieved an accuracy of 66.68%. PrototTypical Networks (Con-4) came in second, with an accuracy of only 61.09% , followed by Relational Networks Networks(Con-4) with an accuracy of 50.28%. Finally, we found that our proposed model outperform those existing baselines in all the scenarios of (5-way, 5-shot), (10-way, 5-shot) , (20-way, 5-shot) and results can be found in the Table 1.

5 CONCLUSION AND FUTURE WORK

In many companies, millions of unlabeled documents containing information relevant to many business-related workflows have to be processed to be classified or / and to extract key information. Unfortunately, a large percentage of these documents consists of unstructured formats in the form of images and PDF documents. Examples of these types of documents include: medical bills, attorney letters ..etc. However, the labeled data needed by traditional learning approaches maybe too expensive and taxing on business

experts and hence not practical in real-world industry settings. Hence, a few-shot learning pipeline is highly desired.

In this work, we proposed a novel method for few-shot document image classification under domain shift for semi-structured business documents, using the canonical correlation block to align extracted text and image feature vectors. We evaluate our work by extensive comparisons with existing methods on two datasets. We rigorously benchmarked our method against the state-of-the-art few-shot computer vision models on both an insurance process derived dataset and the miniRVL dataset. The results showed our method consistently performed better than existing baselines on few-shot classification tasks.

REFERENCES

Galen Andrew, Raman Arora, Jeff Bilmes, and Karen Livescu. 2013. Deep canonical correlation analysis. In *International conference on machine learning*. PMLR, 1247–1255.

Iz Beltagy, Matthew E. Peters, and Arman Cohan. 2020. Longformer: The Long-Document Transformer. *arXiv:2004.05150* (2020).

John Cai and Sheng Mei Shen. 2020. Cross-domain few-shot learning with meta fine-tuning. *arXiv preprint arXiv:2005.10544* (2020).

R. Caruana. 2004. Multitask Learning. *Machine Learning* 28 (2004), 41–75.

Da Chen, Yuefeng Chen, Yuhong Li, Feng Mao, Yuan He, and Hui Xue. 2021. Self-supervised learning for few-shot image classification. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 1745–1749.

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

Chelsea Finn, Aravind Rajeswaran, Sham Kakade, and Sergey Levine. 2019. Online Meta-Learning. arXiv:1902.08438 [cs.LG]

Victor Garcia and Joan Bruna. 2018. Few-Shot Learning with Graph Neural Networks. arXiv:1711.04043 [stat.ML]

Spyros Gidaris and Nikos Komodakis. 2018. Dynamic few-shot visual learning without forgetting. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 4367–4375.

Yunhui Guo, Noel C Codella, Leonid Karlinsky, James V Codella, John R Smith, Kate Saenko, Tajana Rosing, and Rogerio Feris. 2020. A broader study of cross-domain few-shot learning. In *European Conference on Computer Vision*. Springer, 124–141.

Adam W Harley, Alex Ufkes, and Konstantinos G Derpanis. 2015. Evaluation of deep convolutional nets for document image classification and retrieval. In *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 991–995.

Harold Hotelling. 1992. Relations between two sets of variates. In *Breakthroughs in statistics*. Springer, 162–190.

Mike Huisman, Jan N. van Rijn, and Aske Plaat. 2021. A survey of deep meta-learning. *Artificial Intelligence Review* 54, 6 (Apr 2021), 4483–4541. https://doi.org/10.1007/s10462-021-10004-4

A. Krizhevsky. 2009. Learning Multiple Layers of Features from Tiny Images.

B. Lake, R. Salakhutdinov, Jason Gross, and J. Tenenbaum. 2011. One shot learning of simple visual concepts. *Cognitive Science* 33 (2011).

Yann Lifchitz, Yannis Avrithis, Sylvaine Picard, and Andrei Bursuc. 2019. Dense classification and implanting for few-shot learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 9258–9267.

Sinno Jialin Pan and Qiang Yang. 2010. A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering* 22 (2010), 1345–1359.

Nitin Namdeo Pise and Parag Kulkarni. 2008. A Survey of Semi-Supervised Learning Methods. In *2008 International Conference on Computational Intelligence and Security*, Vol. 2. 30–34. https://doi.org/10.1109/CIS.2008.204

Sachin Ravi and Hugo Larochelle. 2016. Optimization as a model for few-shot learning. (2016).

Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. 2015. Imagenet large scale visual recognition challenge. *International journal of computer vision* 115, 3 (2015), 211–252.

Andrei A Rusu, Dushyant Rao, Jakub Sygnowski, Oriol Vinyals, Razvan Pascanu, Simon Osindero, and Raia Hadsell. 2018. Meta-learning with latent embedding optimization. *arXiv preprint arXiv:1807.05960* (2018).

Jake Snell, Kevin Swersky, and Richard S Zemel. 2017. Prototypical networks for few-shot learning. *arXiv preprint arXiv:1703.05175* (2017).

Xiaojin Zhu. 2008. Semi-Supervised Learning Literature Survey. *Comput Sci, University of Wisconsin-Madison* 2 (07 2008).