Domain Generalization in Medical Image Segmentation via Meta-Learners

Hugo Oliveira and Roberto M. Cesar Jr.
Institute of Mathematics and Statistics
Universidade de São Paulo
São Paulo, Brazil
{oliveirahugo,rmcesar}@usp.br

Pedro H. T. Gama
Department of Computer Science
Universidade Federal de Minas Gerais
Belo Horizonte, Brazil
phtg@dcc.ufmg.br

Jefersson A. dos Santos
Computing Science and Mathematics
University of Stirling
Stirling, Scotland, UK
jefersson@ieee.org

***Abstract*—Automatic and semi-automatic radiological image segmentation can help physicians in the processing of real-world medical data for several tasks such as detection/diagnosis of diseases and surgery planning. Current segmentation methods based on neural networks are highly data-driven, often requiring hundreds of laborious annotations to properly converge. The generalization capabilities of traditional supervised deep learning are also limited by the insufficient variability present in the training dataset. One very proliferous research field that aims to alleviate this dependence on large numbers of labeled data is Meta-Learning. Meta-Learning aims to improve the generalization capabilities of traditional supervised learning by training models to learn in a label efficient manner. In this tutorial we present an overview of the literature and proposed ways of merging this body of knowledge with deep segmentation architectures to produce highly adaptable multi-task meta-models for few-shot weakly-supervised semantic segmentation. We introduce a taxonomy to categorize Meta-Learning methods for both classification and segmentation, while also discussing how to adapt potentially any few-shot meta-learner to a weakly-supervised segmentation task.**

***Index Terms*—meta-learning, few-shot learning, semantic segmentation, medical imaging, domain generalization**

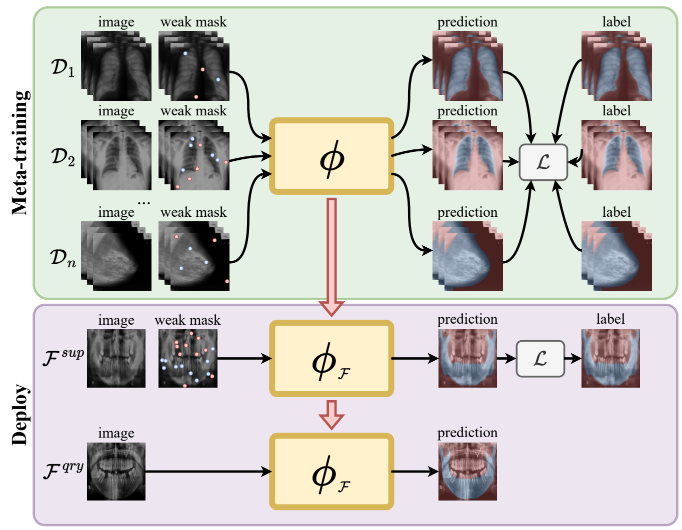

Fig. 1: Multi-task episodic training of ML algorithms for few-shot weakly-supervised segmentation in radiology. Multiple datasets $\mathcal{D}_1, \mathcal{D}_2, \dots \mathcal{D}_3$ are leveraged to train through a loss function $\mathcal{L}$ a model ϕ capable of generalizing to a target out-of-distribution domain $\mathcal{F}$ through its support set $\mathcal{F}^{sup}$.

I. Introduction

One of the most common and useful tasks in medical imaging is the segmentation of organs of interest or abnormalities such as tumors, nodules or fractures [24]. Automatic segmentation models can help in the detection of a myriad of illnesses, in prognosis/triage, and in the planning of surgery and medical treatment. This is further accentuated when the images are not flat 2D projections of the body (e.g. chest X-rays, mammograms or dental X-rays), but instead 3D data as in Computed Tomography (CT) and Magnetic Resonance (MR) volumes, or even 4D data as in functional Magnetic Resonance Imaging (fMRI). When dealing with high-dimensional data, a physician must visualize one slice at a time in a monitor, posing additional difficulties related to the spatial context and inter-slice coherence of annotations. The process of manually annotating in a voxelwise fashion one single medical volume can take hours, depending on the resolution of the image and the number of structures of interest, considerably slowing the analysis of such data.

Current state-of-the-art methods for medical image understanding rely mostly on Deep Neural Networks (DNNs) [15], [22]; however, Deep Learning methods for visual understanding [30], [38] are highly data-driven. While there is a plethora of large-scale RGB image datasets such as ImageNet [7], MS COCO [27] or Pascal VOC [9], multiple specific RGB domains (e.g. structural engineering, fine-grained animal classification or biometric applications) still lack large-scale datasets due to their intrinsically small-data characteristics. In medical imaging, apart from a few domains as mammography [32], [43] and chest X-rays [2], [51], ethical and privacy considerations hamper the compilation of truly large-scale public domain biomedical image datasets. Therefore, learning with a constrained amount of data is a highly sought after feature of modern visual understanding methods. To alleviate this problem, often called Few-Shot Learning (FSL), at

The authors would like to thank CAPES, CNPq (312102/2017-8, 424700/2018-2, and 311395/2018-0), FAPEMIG, FAPESP (grants #2020/06744-5 and #2015/22308-2), and Serrapilheira Institute (grant #R-2011-37776).

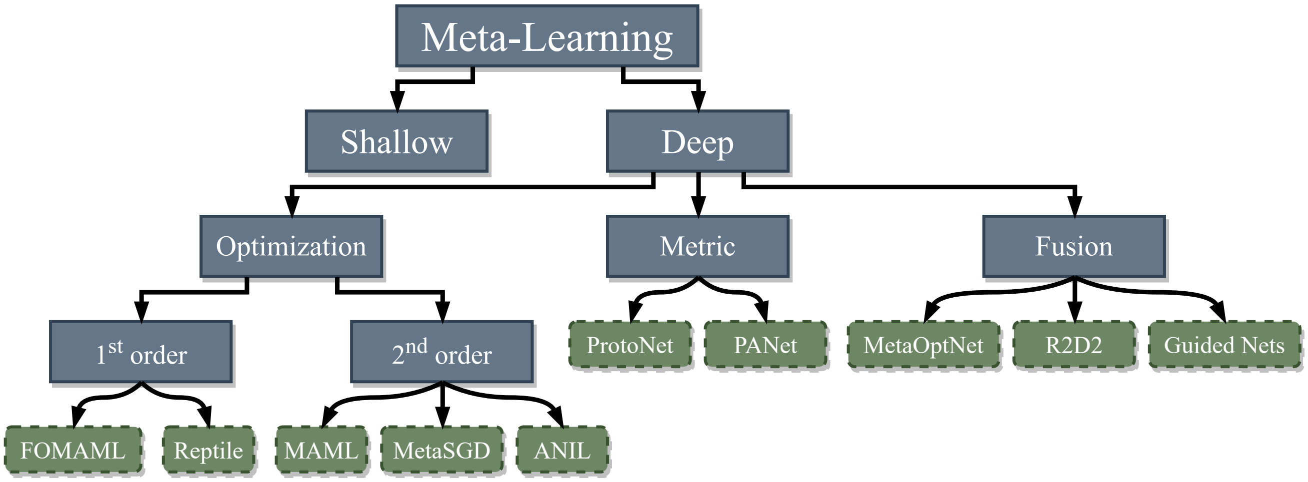

Fig. 2: Proposed deep ML taxonomy for image classification and segmentation. Three main branches are directly linked to deep ML: optimization- [10], [25], [33], [36], metric- [13], [41], [49] and fusion-based [1], [23], [37] strategies. Optimization-based techniques rely either on first- or second-order gradients. A non-exhaustive list of methods for each subfamily of ML methods is presented in green.

least three distinct bodies of knowledge emerged during the last decade: Domain Adaptation (DA) [50], Self-Supervised Learning (SSL) [21] and Meta-Learning (ML) [19].

DA leverage the knowledge obtained from related domains in order to ease the learning of a target task, usually by means of moment matching objectives [8], [44] or generative modeling [18], [34], [35]. SSL has achieved very promising performances in unsupervised pretraining during the last years, mainly when coupled with contrastive losses [4]–[6], [16]. Models pretrained with Contrastive SSL on ImageNet [7] have yielded state-of-the-art performances in other traditional visual tasks as object detection and semantic segmentation [4], [6] of RGB images, while the first SSL models pretrained on medical data [14], [42], [48] have gained attention.

In contrast to DA and SSL, ML encourages models to achieve good few-shot performance by explicitly enforcing a meta-model ϕ to have the ability of *learning to learn* rapidly from other datasets in related image domains. While shallow meta-knowledge extraction traditionally was achieved by combining shallow learning algorithms [45], [46] or selecting the best learner to a certain task [40] starting from handcrafted features, deep ML is achieved by training end-to-end Neural Networks in an episodic fashion [19], [20]. Deep ML has been adapted to classification, regression and even reinforcement learning [10], [11], [25], evidencing its versatility. Meta-Learning, similarly to DA, is capable of leveraging knowledge from related tasks in order to achieve good few-shot performances. However, instead of relying on generative models or moment matching, ML traditionally uses second order optimization [10], [25], [36], similarity learning [41] or fusion [1], [23], [47]. More recently, ML has been adapted to the task of semantic segmentation from few-shot weakly-annotated samples [12], [13], [28], [37], [49]. In this work we will focus on deep Meta-Learning methods trained in a episodic multi-task fashion as an alternative to achieve Few-Shot Weakly-Supervised (FSWS) segmentation tasks in radiology, as depicted in Figure 1.

The following sections of this manuscript are described as follows. Section II presents our proposed taxonomy for characterizing ML methods into three distinct paradigms: gradient-based, similarity-based and fusion-based. Section III describes a general pipeline for each family of methods for few-shot image classification in the taxonomy, showing examples of ML strategies in each category, as well as their similarities/peculiarities. We describe the same three paradigms of ML methods for FSWS segmentation in Section IV, also generalizing how few-shot image classification methods in each clade of our taxonomy can be ported to conduct weakly-supervised segmentation. At last, we present our final remarks, conclusion and possible future research directions in ML in Section V.

II. Taxonomy

The simpler supervised learning problem definition tries to find the most likely set of parameters θ for a certain training dataset $\mathcal{D} = \{(\mathbf{x}^{(1)}, \mathbf{y}^{(1)}), (\mathbf{x}^{(2)}, \mathbf{y}^{(2)}), \ldots, (\mathbf{x}^{(k)}, \mathbf{y}^{(k)})\}$, where $\mathbf{x}^{(i)}$ is a data sample (e.g. an image) and $\mathbf{y}^{(i)}$ the associated decision (e.g. reference segmentation, class...), so that a function $f_\theta(\mathbf{x}_{new})$ yields accurate predictions $\hat{y}^{new} \sim y^{new}$ for novel unseen data points. The training of DNNs is highly data-driven, with the optimization computed by minimizing an objective function $\mathcal{L}(\hat{\mathbf{y}}^{(i)}, \mathbf{y}^{(i)})$ for a sample $(\mathbf{x}^{(i)}, \mathbf{y}^{(i)}) \in \mathcal{D}$ – or, more realistically, a batch of samples – in successive iterations. $\mathcal{L}$ is a loss function suited to the task at hand (e.g. Cross Entropy for classification/segmentation, L1 for regression/reconstruction, etc.). Supervised learning is, therefore, highly dependent on the variability present in $\mathcal{D}$, often failing to generalize to even small domain shifts in

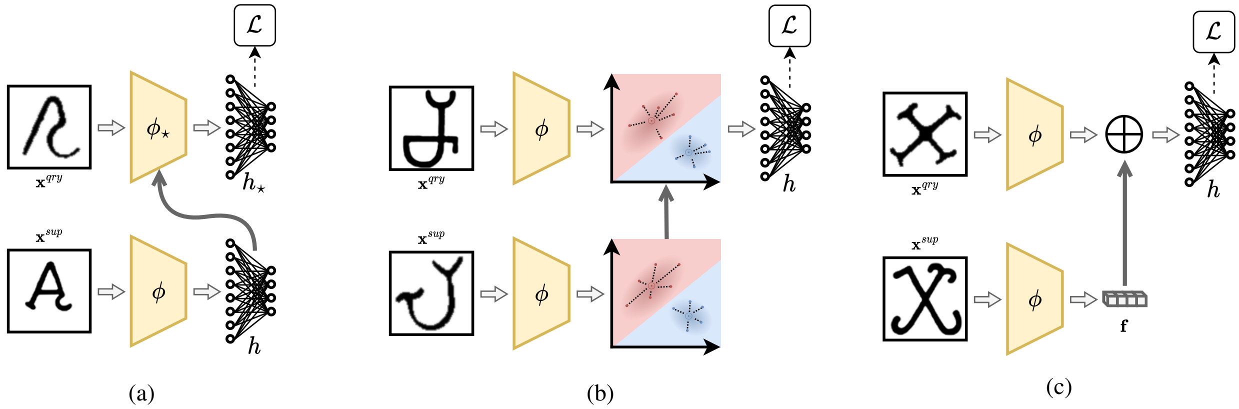

Fig. 3: Few-shot target digit classification scenario with meta-learners for few-shot classification of armorial digits [3]. Gradient-based ML algorithms (a) adapt ϕ and h during episodic training to better fit the target task $\tau_\star$ by adapting a temporary pair $\phi_\star$ and $h_\star$ and backpropagating through the gradients to update the meta-model (ϕ) and meta-head (h). In contrast to that, metric-based methods (b) maintain the same encoding ϕ function from the support to the query set, instead performing multi-task learning through the similarity embedding space e^{sup}, which is reused for the query set. At last, fusion-based strategies (c) also reuse the same ϕ for the support and query sets, while "guiding" the head h to predict the correct classes via the support features $\mathbf{f}$.

the distribution of novel samples (e.g. different acquisition equipment, acquisition parameters, samples in the tail of the distribution, etc.).

In this context, deep ML is an emerging research field that investigates highly generalizable algorithms for few-shot learning, improving on traditional supervised learning. There are at least five bodies of knowledge that serve as theoretical background for Meta-Learning algorithms: 1) early black-box approaches – also known as model-based ML – that output the parameters of a neural network specialized for target tasks [31], [39]; 2) metric-learning for similarity-based learning-to-learn in a highly efficient manner [41]; 3) the highly successful gradient-based ML paradigm [10], [36]; 4) data fusion approaches [1], [23], [37], [47] that "guide" the learning on the query set through the support features; and 5) the yet underdeveloped Bayesian methods [11]. Model-based approaches [31], [39] are often employed in highly specific settings, while bayesian approaches are still in their infancy; hence, in this proposal we will focus on gradient-, similarity- and fusion-based meta-learners, as depicted in Figure 2. Additionally, gradient-based methods can be divided according to their training strategies based on either first- or second-order optimization.

III. Meta-Learning for Classification

In a Meta-Learning setting, a task $\tau_i = \{\mathcal{D}^{sup}_{\tau_i}, \mathcal{D}^{qry}_{\tau_i}\}$ is defined by two sets of data points $\mathcal{D}^{sup}_{\tau_i}$ and $\mathcal{D}^{qry}_{\tau_i}$. Similarly to supervised learning, each one of these sets is comprised of image/label pairs $(\mathbf{x}_{\tau_i}, \mathbf{y}_{\tau_i})$. Most Meta-Learning algorithms can be conceptually described as having two nested loops: an outer and an inner loop. In the inner loop the model is trained to adapt to a task $\mathcal{T}_i$ sampled from a distribution of tasks $p(\mathcal{T})$, while the cumulative error of the tasks is used to pretrain the model in the outer loop. In other words, during an iteration of the inner loop, temporary parameters $\theta^*_{\tau_i}$ are computed using the support set ($\mathcal{D}^{sup}_{\tau_i}$) of each individual task, while the query set ($\mathcal{D}^{qry}_{\tau_i}$) is used to evaluate the performance of the model in novel data, enforcing generalization in each individual task. The base model θ, however, is only updated at the end of the iteration, considering the gradients computed in all inner loops, enforcing generalization across tasks. The same formulation applies to optimization-based [10], [36], metric-based [41], and fusion-based [1], [23], [47] algorithms, as shown in Figures 3a, 3b and 3c, respectively.

All aforementioned methods employ ML for few-shot image classification, not delving into FSWS segmentation on their initial experiments, although some recent works have explored this application of Meta-Learning [12], [13], [37], [49]. In Section IV we generalize the definition of meta-segmentation beyond these few proposed methods to adapt generic optimization-, metric- and fusion-based meta-learners originally designed for image classification to FSWS segmentation in medical imaging.

IV. Meta-Learning for Image Segmentation

Weakly-supervised **S**egmentation **L**earning (**WeaSeL**) [12] works toward the task of FSWS segmentation in medical imaging by adapting on MAML [10] for dense-labeling, with some additional properties that render it suitable for weak annotations. More specifically, instead of the simple Cross-Entropy loss used as default by the original MAML, a Selective Cross-Entropy was employed in WeaSeL in order to account for pixels with unknown labels. Additionally, during meta-training $\mathcal{D}^{sup}_{\tau_i}$ sets used in the inner loop optimization steps of the temporary models θ^*_i are composed of the weakly-annotated masks instead of the dense ones, while $\mathcal{D}^{qry}_{\tau_i}$ – used to directly optimize θ – is trained using the dense labels of the query set. This encourages the algorithm to predict the correct

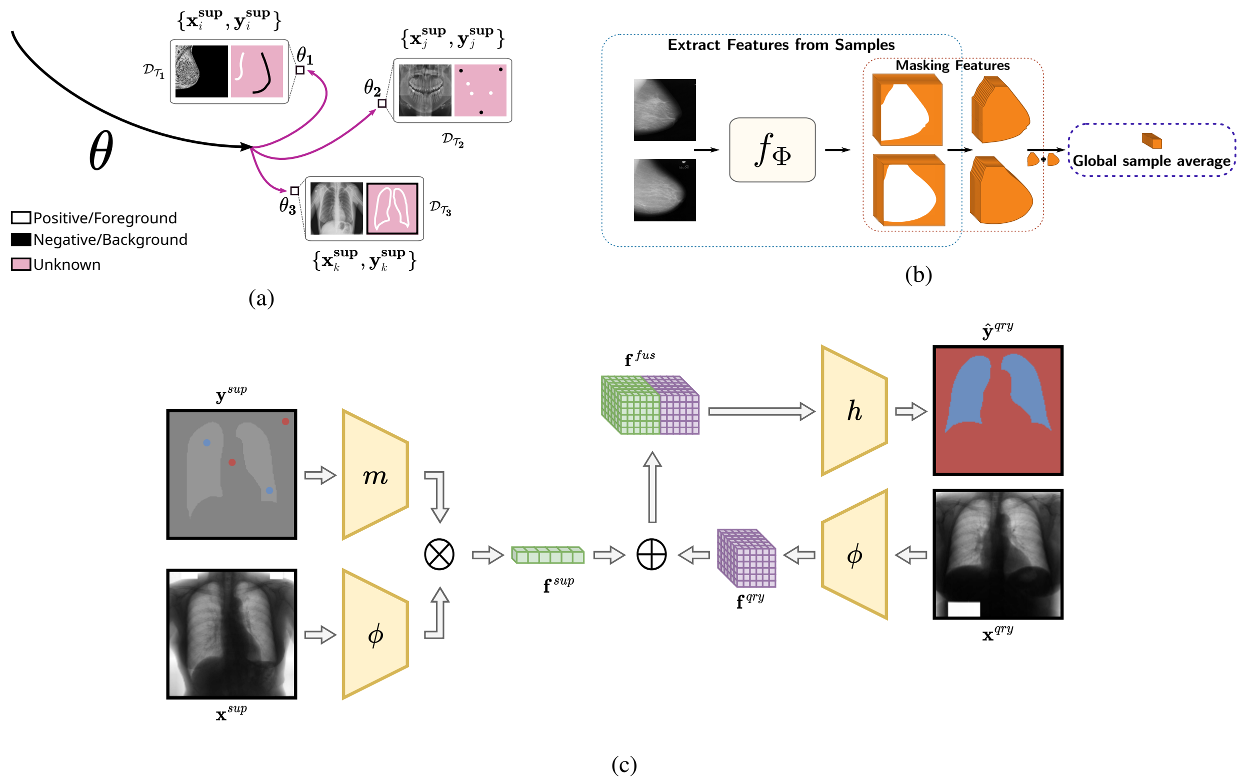

Fig. 4: Meta-Learning procedures for WeaSeL [12] (a), ProtoSeg [13] (b) and Guided Nets [37] in FSWS segmentation tasks. Adapted from Gama *et al.* [12], [13] and Rakelly *et al.* [37].

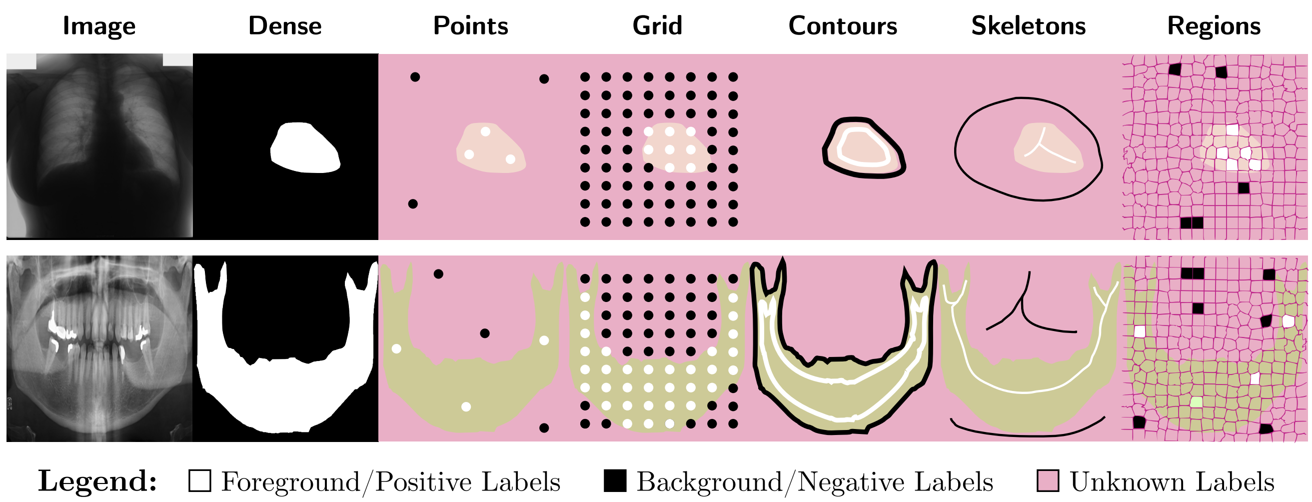

Fig. 5: Images, ground truths, and 5 weakly-annotated label modalities employed in both training and testing of FSWS segmentation meta-learners for two medical segmentation tasks: heart and lower mandible. Adapted from Gama *et al.* [13].

dense masks according to the sparse ones on a novel FSWS task $\mathcal{F}$. This idea can be ported to other gradient-based meta-learners, such as MetaSGD [25], ANIL [36], or Reptile [33] by simply replacing the first- or second-order ML algorithm used to train meta-model ϕ and meta-head h in an episodic fashion. A depiction of the meta-training phase of an optimization-based FSWS meta-learner can be seen in Figure 4a.

Aiming to adapt the metric-learning paradigm of ML algorithms, ProtoSeg [13] and PANets [49] adapt ProtoNets [41] to FSWS segmentation. During the meta-training phase, ProtoSeg minimizes the distances (e.g. euclidean, cosine or both) between the embedding representations $f_\phi(\mathcal{D}^{sup}_{\mathcal{T}_i})$ and $f_\phi(\mathcal{D}^{qry}_{\mathcal{T}_i})$ for positive (foreground) and negative (background) classes. These embeddings are used to generate the prototypical representations of the positive and negative classes by averaging the embeddings across all labeled pixels. As shown in Figure 4b, the computation of the prototypes for a certain class is similar to a masked average pooling considering only the pixels annotated as said class. For the testing phase on a target few-shot dataset $\mathcal{F}$, each pixel in the query set ($\mathcal{F}^{qry}$) is forwarded through f_ϕ and yields an embedded representation that can be compared to the prototypes computed on the support ($\mathcal{F}^{sup}$). Distances between each embedded query pixel and the prototypes are treated as logits for $\mathcal{F}^{qry}$. The ProtoSeg pipeline can be used for any choice of encoder ϕ (e.g. any FCN backbone [29], U-Nets [38], etc), distance metric (e.g. euclidean, cosine, manhattan, etc) or supervised classification loss based on logits (e.g. cross-entropy, dice, focal, etc).

Similarly to ProtoSeg, PANets [49] employ prototypes using a masked average pooling in the features extracted from support sets.For a distance function, they opted to use the cosine distance, instead of the euclidean distance metric. The main addition introduced by PANets was the proposal of a *Prototype Alignment Regularization* loss ($\mathcal{L}_{PAR}$) to more efficiently leverage both support and query labels in the meta-training phase. The idea is that during training, in addition to computing the cost of segmenting the query images based on the distance to the support prototypes, they also compute the cost of segmenting the support images using query prototypes, i.e. prototypes constructed from query extracted features. This idea was observed to considerably and reliably improve the performance of metric-based FSWS segmentation ML algorithms [49], possibly being adapted to any similarity-based FSWS method.

The meta-dataset $\mathcal{D}$ used for WeaSeL and ProtoSeg was composed of multiple chest, dental and mammographic X-ray datasets, totaling 13 distinct segmentation tasks. Both WeaSeL and ProtoSeg are designed to be agnostic to the weak annotation style and density, albeit with differing label efficiencies observed per labeling configuration [12], [13]. This is accomplished by randomly sampling the sparse annotation style of the simulated weakly-annotated masks along with the task $\tau_i \sim p(\mathcal{T})$ in each inner loop for $\mathcal{D}^{sup}_{\tau_i}$. The annotation styles used by Gama *et al.* can be seen in Figure 5, however we highlight that this is a non-exhaustive list of annotation modalities that can be further explored in future research. More details regarding the procedure can be seen in Gama *et al.* [13].

Guided Nets [37] are one of the most simple fusion-based method for FSWS segmentation. This method uses a single feature extractor ϕ that computes the features of support and query images and an additional embedding network m for the support masks. Support features are merged with mask embeddings via a function $\otimes$ – usually a simple multiplication – afterwards being collapsed to a 1D vector $\mathbf{f}^{sup}$ through averaging. $\mathbf{f}^{sup}$ is tiled in order to equal the spacial dimensions of the query features $\mathbf{f}^{qry}$. Then another function $\oplus$ (i.e. concatenation, addition or multiplication) is applied to merge $\mathbf{f}^{sup}$ and $\mathbf{f}^{qry}$, resulting in $\mathbf{f}^{fus}$. The fused tensor $\mathbf{f}^{fus}$ is then further processed by a segmentation head h, producing the final segmentation prediction $\hat{\mathbf{y}}^{qry}$ for the query. Guided Nets originally relied heavily on ImageNet pretraining [7], which limited their scope to RGB images. More recently, Gama *et al.* [13] adapted Guided Nets to be pretrained directly on the target domain, allowing for a greater domain versatility whenever a labeled meta-dataset can be gathered from public sources.

Following the schema presented in Figure 4c, a whole family of fusion-based meta-learners can be implemented, including methods traditionally proposed for classification, such as MetaOptNet [23] and R2D2 [1]. The general nature of that pipeline allows for multiple distinct mask embedding functions m, and fusing functions $\otimes$ and $\oplus$ to be used for FSWS segmentation.

V. Conclusion

In this tutorial we introduced a novel taxonomy for categorizing ML algorithms for few-shot image classification [1], [10], [23], [25], [33], [36], [41], [47] and FSWS segmentation [12], [12], [37], [49]. We present an overview of gradient-, metric- and fusion-based algorithms for image recognition, while also describing the rationale behind each ML paradigm. At last, this work discussed how each ML paradigm could be/has been ported to FSWS segmentation. We generalize the concept of FSWS using ML to potentially adapt any standard few-shot image classification meta-learner to dense labeling, instead of relying on the very few works previously published proposing to solve this particular task.

One major limitation of the approaches discussed in this work is that all ML algorithms – for both classification and FSWS segmentation – require multiple annotated datasets related to the target domain/task. This is a major hamper to domains outside of radiology, such as multi/hiperspectral remote sensing images, histopathology, biometric data, video segmentation and temporal signal analysis, which might not have related labeled tasks to be used in the meta-training phase. In those cases, SSL and Unsupervised/Semi-Supervised DA might be better alternatives to achieve few-shot learning on small-data tasks. Future works coupling meta-learners with SSL might be an alternative to mitigate the dependence on annotated data for meta-training. Apart from the application to other imaging domains, future directions for FSWS segmentation might include the use of more efficient segmentation architectures [17], other selective segmentation losses [26], [30] and interactive image segmentation using ML.

References

[1] Luca Bertinetto, Joao F Henriques, Philip HS Torr, and Andrea Vedaldi. Meta-Learning with Differentiable Closed-form Solvers. *arXiv preprint arXiv:1805.08136*, 2018.

[2] Aurelia Bustos, Antonio Pertusa, Jose-Maria Salinas, and Maria de la Iglesia-Vayá. PadChest: A Large Chest X-Ray Image Dataset with Multi-label Annotated Reports. *Medical Image Analysis*, 66:101797, 2020.

[3] Luciano Cardinali. A Tipografia Armorial: A Concepção de uma Identidade Visual Sertaneja. *DAT Journal*, 1(1):160–180, 2016.

[4] Mathilde Caron, Ishan Misra, Julien Mairal, Priya Goyal, Piotr Bojanowski, and Armand Joulin. Unsupervised Learning of Visual Features by Contrasting Cluster Assignments. In *NeurIPS*, volume 33, pages 9912–9924, 2020.

[5] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A Simple Framework for Contrastive Learning of Visual Representations. In *ICML*, pages 1597–1607. PMLR, 2020.

[6] Xinlei Chen and Kaiming He. Exploring Simple Siamese Representation Learning. In *CVPR*, pages 15750–15758, 2021.
[7] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. ImageNet: A Large-scale Hierarchical Image Database. In *CVPR*, pages 248–255. IEEE, 2009.
[8] Li Diao, Haoyue Guo, Yue Zhou, and Yayi He. Bridging the GAP Between Outputs: Domain Adaptation for Lung Cancer IHC Segmentation. In *ICIP*, pages 6–10. IEEE, 2021.
[9] Mark Everingham, SM Ali Eslami, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The PASCAL Visual Object Classes Challenge: A Retrospective. *IJCV*, 111(1):98–136, 2015.
[10] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks. In *ICML*, pages 1126–1135. PMLR, 2017.
[11] Chelsea Finn, Kelvin Xu, and Sergey Levine. Probabilistic Model-Agnostic Meta-Learning. In *NeurIPS*, 2018.
[12] Pedro Henrique Targino Gama, Hugo Oliveira, and Jefersson Alex dos Santos. Weakly Supervised Medical Image Segmentation. *SIBGRAPI*, 2021.
[13] Pedro Henrique Targino Gama, Hugo Neves Oliveira, Jose Marcato, and Jefersson Dos Santos. Weakly Supervised Few-Shot Segmentation Via Meta-Learning. *IEEE Transactions on Multimedia*, 2022.
[14] Matej Gazda, Jakub Gazda, Jan Plavka, and Peter Drotar. Self-Supervised Deep Convolutional Neural Network for Chest X-ray Classification. *arXiv preprint arXiv:2103.03055*, 2021.
[15] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT press, 2016.
[16] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum Contrast for Unsupervised Visual Representation Learning. In *CVPR*, pages 9729–9738, 2020.
[17] Sean M Hendryx, Andrew B Leach, Paul D Hein, and Clayton T Morrison. Meta-learning initializations for image segmentation. *arXiv preprint arXiv:1912.06290*, 2019.
[18] Judy Hoffman, Eric Tzeng, Taesung Park, Jun-Yan Zhu, Phillip Isola, Kate Saenko, Alexei Efros, and Trevor Darrell. CyCADA: Cycle-Consistent Adversarial Domain Adaptation. In *ICML*, pages 1989–1998. PMLR, 2018.
[19] Timothy Hospedales, Antreas Antoniou, Paul Micaelli, and Amos Storkey. Meta-Learning in Neural Networks: A Survey. *arXiv preprint arXiv:2004.05439*, 2020.
[20] Mike Huisman, Jan N Van Rijn, and Aske Plaat. A Survey of Deep Meta-Learning. *Artificial Intelligence Review*, 54(6):4483–4541, 2021.
[21] Longlong Jing and Yingli Tian. Self-Supervised Visual Feature Learning with Deep Neural Networks: A Survey. *IEEE TPAMI*, 2020.
[22] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. ImageNet Classification with Deep Convolutional Neural Networks. *NIPS*, 25:1097–1105, 2012.
[23] Kwonjoon Lee, Subhransu Maji, Avinash Ravichandran, and Stefano Soatto. Meta-Learning with Differentiable Convex Optimization. In *CVPR*, pages 10657–10665, 2019.
[24] Tao Lei, Risheng Wang, Yong Wan, Bingtao Zhang, Hongying Meng, and Asoke K Nandi. Medical Image Segmentation Using Deep Learning: A Survey. *arXiv preprint arXiv:2009.13120*, 2020.
[25] Zhenguo Li, Fengwei Zhou, Fei Chen, and Hang Li. Meta-SGD: Learning to Learn Quickly for Few-Shot Learning. *arXiv preprint arXiv:1707.09835*, 2017.
[26] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *ICCV*, pages 2980–2988, 2017.
[27] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft COCO: Common Objects in Context. In *ECCV*, pages 740–755. Springer, 2014.
[28] Quande Liu, Qi Dou, and Pheng-Ann Heng. Shape-Aware Meta-Learning for Generalizing Prostate MRI Segmentation to Unseen Domains. In *MICCAI*, pages 475–485. Springer, 2020.
[29] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully Convolutional Networks for Semantic Segmentation. In *CVPR*, pages 3431–3440, 2015.
[30] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. In *3DV*, pages 565–571. IEEE, 2016.
[31] Nikhil Mishra, Mostafa Rohaninejad, Xi Chen, and Pieter Abbeel. A Simple Neural Attentive Meta-Learner. In *ICLR*, 2018.
[32] Inês C Moreira, Igor Amaral, Inês Domingues, António Cardoso, Maria Joao Cardoso, and Jaime S Cardoso. INbreast: Toward a Full-Field Digital Mammographic Database. *Academic Radiology*, 19(2):236–248, 2012.
[33] Alex Nichol and John Schulman. Reptile: A Scalable Meta-Learning Algorithm. *arXiv preprint arXiv:1803.02999*, 2(3):4, 2018.
[34] Hugo Oliveira and Jefersson Alex dos Santos. Deep Transfer Learning for Segmentation of Anatomical Structures in Chest Radiographs. In *SIBGRAPI*, pages 204–211. IEEE, 2018.
[35] Hugo Oliveira, Edemir Ferreira, and Jefersson Alex dos Santos. Truly Generalizable Radiograph Segmentation with Conditional Domain Adaptation. *IEEE Access*, 8:84037–84062, 2020.
[36] Aniruddh Raghu, Maithra Raghu, Samy Bengio, and Oriol Vinyals. Rapid Learning or Feature Reuse? Towards Understanding the Effectiveness of MAML. *arXiv preprint arXiv:1909.09157*, 2019.
[37] Kate Rakelly, Evan Shelhamer, Trevor Darrell, Alexei A Efros, and Sergey Levine. Few-shot Segmentation Propagation with Guided Networks. *arXiv preprint arXiv:1806.07373*, 2018.
[38] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *MICCAI*, pages 234–241. Springer, 2015.
[39] Adam Santoro, Sergey Bartunov, Matthew Botvinick, Daan Wierstra, and Timothy Lillicrap. Meta-Learning with Memory-Augmented Neural Networks. In *ICML*, pages 1842–1850. PMLR, 2016.
[40] Kate A Smith-Miles. Cross-disciplinary perspectives on meta-learning for algorithm selection. *ACM Computing Surveys*, 41(1):1–25, 2009.
[41] Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical Networks for Few-Shot Learning. In *NIPS*, pages 4080–4090, 2017.
[42] Hari Sowrirajan, Jingbo Yang, Andrew Y Ng, and Pranav Rajpurkar. MoCo Pretraining Improves Representation and Transferability of Chest X-Ray Models. In *Medical Imaging with Deep Learning*, pages 728–744. PMLR, 2021.
[43] J Suckling et al. The Mammographic Image Analysis Society Digital Mammogram Database, 1994.
[44] Baochen Sun and Kate Saenko. Deep CORAL: Correlation Alignment for Deep Domain Adaptation. In *ECCV*, pages 443–450. Springer, 2016.
[45] Ricardo Vilalta and Youssef Drissi. A perspective view and survey of meta-learning. *Artificial Intelligence Review*, 18(2):77–95, 2002.
[46] Ricardo Vilalta, Christophe G Giraud-Carrier, Pavel Brazdil, and Carlos Soares. Using Meta-Learning to Support Data Mining. *Int. J. Comput. Sci. Appl.*, 1(1):31–45, 2004.
[47] Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Daan Wierstra, et al. Matching Networks for One Shot Learning. *NIPS*, 29:3630–3638, 2016.
[48] Yen Nhi Truong Vu, Trevor Tsue, Jason Su, and Sadanand Singh. An Improved Mammography Malignancy Model with Self-supervised Learning. In *Medical Imaging 2021: Computer-Aided Diagnosis*, volume 11597, page 115970W. International Society for Optics and Photonics, 2021.
[49] Kaixin Wang, Jun Hao Liew, Yingtian Zou, Daquan Zhou, and Jiashi Feng. PANet: Few-shot Image Semantic Segmentation with Prototype Alignment. In *CVPR*, pages 9197–9206, 2019.
[50] Mei Wang and Weihong Deng. Deep Visual Domain Adaptation: A Survey. *Neurocomputing*, 312:135–153, 2018.
[51] Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, and Ronald M Summers. ChestX-ray8: Hospital-scale Chest X-Ray Database and Benchmarks on Weakly-supervised Classification and Localization of Common Thorax Diseases. In *CVPR*, pages 2097–2106, 2017.