ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 978-1-6654-0540-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICASSP43922.2022.9746791

CROSS-DOMAIN FEW-SHOT LEARNING FOR RARE-DISEASE SKIN LESION SEGMENTATION

Yixin Wang[1] *Zhe Xu*[2] *Jiang Tian*[3] *Jie Luo*[4]
Zhongchao Shi[3] *Yang Zhang*[1,5] *Jianping Fan*[3,6] *Zhiqiang He*[1,5]

[1]Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China
[2]Department of Biomedical Engineering, The Chinese University of Hong Kong, Hong Kong, China
[3]AI Lab, Lenovo Research, Beijing, China
[4]Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA
[5]Lenovo Corporate Research & Development, Lenovo Ltd., Beijing, China
[6]Department of Computer Science, University of North Charlotte, NC, USA

ABSTRACT

Recently, deep learning (DL)-based skin lesion segmentation in dermoscopic images has advanced the efficient diagnosis of skin diseases. Commonly, most of the DL-based methods require a large amount of training data and can only perform accurate predictions on pre-defined classes. However, there exist some rare skin diseases with very limited labeled samples, which poses great challenges to typical DL-based methods. Few-shot learning (FSL) technique, which aims to train models with abundant seen classes and then generalizes to related unseen classes, is promising in addressing a similar problem. Unfortunately, simply borrowing the typical FSL is infeasible since collecting such abundant seen-class data (common skin diseases), is also difficult. In this paper, we propose a cross-domain few-shot segmentation (CD-FSS) framework, which enables the model to leverage the learning ability obtained from the natural domain, to facilitate rare-disease skin lesion segmentation with limited data of common diseases. Specifically, the framework consists of two processes, i.e., *specific learning* and *generic learning*, which are alternately optimized in a meta-training manner. A specific learner and a generic learner are tailored to build relationships between both processes. Experimental results demonstrate that our framework significantly improves the generalization ability from natural domain to unseen medical domain.

Index Terms— Few-shot segmentation, Cross-domain, Skin lesion

1. INTRODUCTION

Accurate segmentation and diagnosis of skin disease in skin imaging is critical for the early detection and treatment of skin cancer, which is by far the most common one of all cancers in the world [1]. Recently, deep learning methods have achieved remarkable success in medical image analysis. However, there exist some rare skin diseases that affect very few people (e.g., fewer than 1 in 2,000 people as defined in the European Union) but have a severe physical and intellectual impact on affected patients and their families [2, 3]. Due to the large shape variations, indistinct boundaries of skin lesions and extreme lack of rare-disease images and annotations, training a robust network for automatic rare-disease skin lesion segmentation is very challenging.

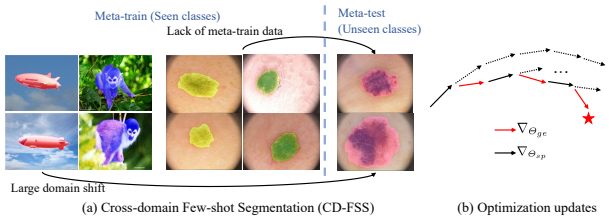

Fig. 1: Motivation. (a) We aim to improve the model's generalization ability to unseen rare skin diseases via few-shot segmentation methods. The key problems are: (1) inadequate skin lesion classes for training; (2) a large domain gap between natural and skin classes. (b) We expect an alternate optimization from these two domains to reach an optimal point.

Though some semi-supervised [4, 5, 6] or unsupervised methods [7, 8] alleviate these issues by utilizing unlabeled data, it is unrealistic to collect either abundant unlabeled data or expert annotations for rare disease. Instead of learning how to solve specific tasks, few-shot segmentation [9, 10, 11, 12, 13] was proposed to generalize the learned knowledge from abundant seen-class labeled data to predict unseen-class data. To achieve acceptable generalization to the unseen-class images (i.e., rare diseases), a considerable amount of annotated seen-class training data (i.e., common diseases) is required in a typical few-shot segmentation model. However, being a challenging task itself, collecting abundant seen-class data of common skin diseases is also infeasible in clinical practice.

Distinct from skin images, the natural images possess plentiful classes. So why not leverage natural images to overcome the scarcity of seen-class skin images? In this work, we propose a cross-domain few-shot framework that empowers the model to effectively learn rare skin lesions segmentation using the 'learning to learn' ability obtained from the natural domain. In order to reduce the influence of large domain shift between the medical and natural domain, we design an alternate meta-training schema consisting of a *specific learning* and a *generic learning*. The former process produces specific object-related features using masked average pooling to incorporate contextual information from limited seen-class skin images, while the latter explores general low-level representations from adequate natural images. Meanwhile, a specific learner and a generic learner are adapted to mine various transferable knowledge between both processes for comparison. The two highly-interactive learning processes are synergistically optimized and alternately updated through a cross-domain meta-training algorithm for mutual promotion.

Despite some efforts are made on cross-domain few-shot classification [14, 15], few works have studied cross-domain few-shot segmentation (CD-FSS). Our work explores the 'learn to learn' ability obtained from natural domain to assist rare disease segmentation in medical domain. Experimental results on FSS-1000 (natural domain) and PH2 (medical domain) datasets demonstrate the superiority of our proposed cross-domain few-shot segmentation framework over SOTA in such a challenging area. More importantly, this work also helps drive further exploration to benefit the computer-aided diagnosis on rare diseases.

2. METHOD

In this section, we firstly present the problem settings and then detail the proposed cross-domain few-shot segmentation framework, including feature extractors, cross-domain prototype learning and a meta-training schema.

2.1. Problem Setting

Few-shot learning, referred to as 'learning to learn', aims to learn transferable knowledge that can be generalized to new classes with extremely scarce labeled data. Here, we define the few-shot segmentation problems as follows.

A large-scale dataset $\mathcal{D}_{train}$ including corresponding ground-truth masks for sets of classes $\mathcal{C}_{seen}$ is provided for training phase. We aim to build a segmentation model that learns quickly to segment sets of novel classes $\mathcal{C}_{unseen}$ in a dataset $\mathcal{D}_{test}$ with few annotations during the test phase, where $\mathcal{C}_{seen} \cap \mathcal{C}_{unseen} = \varnothing$. Once the model is trained, the parameters are fixed and can be directly tested on C_{unseen}. The training and testing phase are aligned with the following episodic paradigm. Intuitively, each of total N episodes is composed of a set of support images $\mathcal{S}$ and query images $\mathcal{Q}$,

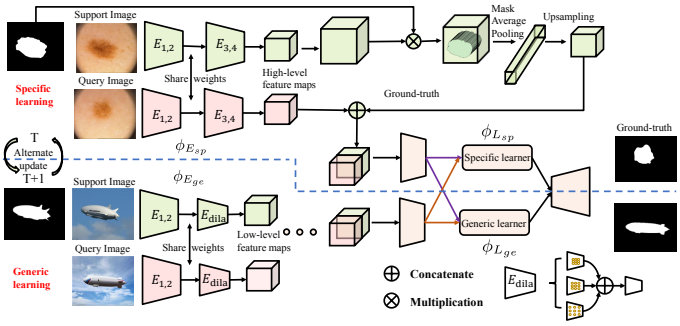

Fig. 2: Framework overview. It performs an alternate training (Algorithm 1) through specific learning and generic learning. Each process consists of two branches to extract representative features of support and query images and densely compare their similarity via prototype learning. A specific and a generic learner are integrated to exploit their mutual benefits following our meta-training schema. "Three dots" means the same prototype extraction architecture as the upper one.

where $\mathcal{D}_{\text{train}} = \{(\mathcal{S}_i, \mathcal{Q}_i)\}_{i=1}^{N_{\text{train}}}$ and $\mathcal{D}_{\text{test}} = \{(\mathcal{S}_j, \mathcal{Q}_j)\}_{j=1}^{N_{\text{test}}}$. Classes in different episodes are supposed to be different, so that the model obtains the ability to generalize to new classes on $\mathcal{Q}_j$ given few support images from $\mathcal{S}_j$ in D_{test}. Due to the extreme scarcity of rare skin disease images, we take only one annotated support image as supervision.

As training such a few-shot model requires abundant images of C_{seen} within that particular domain, it is much difficult to train a learner for medical images. Therefore, our focus in this work is to improve the generalization ability of few-shot segmentation models to unseen medical domains, with the help of obtaining learning ability from the natural domain. Specifically, as shown in Fig. 1, during training phase, we have a large set of natural data $\mathcal{D}_n \sim \mathcal{P}_n$ with plentiful classes $\mathcal{C}_n$ and a small set of skin disease images $\mathcal{D}_m \sim \mathcal{P}_m$ with only a few known classes $\mathcal{C}_m$, where $\mathcal{D}_n \cup \mathcal{D}_m = \mathcal{D}_{train}$, $\mathcal{C}_n \cup \mathcal{C}_m = \mathcal{C}_{seen}$ and $\mathcal{P}_n$, $\mathcal{P}_m$ denote independent natural and medical domain distributions. During the testing phase, the model is expected to be quickly adapted to new and rare disease classes $\mathcal{C}_{unseen}$ in D_{test}.

2.2. Cross-domain Meta-training Framework

2.2.1. Overview

We propose a novel framework to tackle the problem of CD-FSS, as shown in Fig. 2. The whole framework alternately optimizes the process of *specific learning* from skin lesions and *generic learning* from natural images via the proposed meta-training schema. Different-level representative vectors of support images are first extracted and then compared with query images through a cross-domain prototype learning.

Feature Extractor Given a support RGB image $I^{sup} \in \mathbb{R}^{3\times W\times H}$, with its binary mask $Y_c \in \{0,1\}^{W\times H}$ for class

c and a query image I^{que} with the same class c, where W and H are the width and height of the image, our model first embeds them into deep features by a shared backbone network. It has been proven in [16, 17, 11] that higher layers often relate to specific object concepts such as object categories while lower layers extract general low-level information such as fundamental colors and edges. Due to the large domain gap, high-level features corresponding to $\mathcal{C}_{unseen}$ is hard to be learned during training $\mathcal{D}_n$, but highly correlated to $\mathcal{D}_m$. Instead, we expect the model to obtain the learning ability of general features from few-shot learning on $\mathcal{C}_n$ and specific medical learning ability from $\mathcal{C}_m$. Based on these goals, we separately design the feature extractor E_{ge} and E_{sp} for generic learning and specific learning. We adopt the first two blocks of VGG-16 (denoted as $E_{1,2}$) during generic learning process, and the first four blocks (denoted as $\{E_{1,2}, E_{3,4}\}$) for the specific learning process. Following the prior few-shot classification works, we pre-train the feature encoder on the ImageNet dataset to enhance feature extraction ability. Besides, to better capture the multi-scale objects and context information of general features from natural images, we place multiple dilated convolution layers E_{dila} after $E_{1,2}$ to expand the receptive field without losing resolutions. Thus, $E_{ge} = \{E_{1,2}, E_{dila}\}$ can produce more robust and general features to improve generalized learning ability and $E_{sp} = \{E_{1,2}, E_{3,4}\}$ is supposed to extract more specific knowledge for enhancement of disease learning ability.

Cross-domain Prototype Learning We use mask average pooling [9, 10] over I^{sup} to learn prototypes for foreground and filter irrelevant areas. Intuitively, feature maps from feature extractors are interpolated to F with the same size as $W \times H$. We denote the prototype of class c by:

$$p_c = \frac{\sum_{x=1,y=1}^{W,H} Y_c^{(x,y)} * F^{(x,y)}}{\sum_{x=1,y=1}^{W,H} Y_c^{(x,y)}}, \tag{1}$$

where (x, y) indexes the pixel-level location. The obtained feature vectors of I^{sup} are upsampled and concatenated with extracted features from I^{que}. Then the concatenated feature maps are convoluted for dense target comparison, and pixels maximally similar to the class signature are activated. However, there's a large gap of the prototype learning features between medical and natural domain. Therefore, we design a generic learner L_{ge} and a specific learner L_{sp} to separately learn the specificity of features according to the current episode. Both learners consist of the same convolutional layers followed by a ReLU layer. They are optimized alternately via a meta-training schema to densely compare generic and specific features. The features are further concatenated and forwarded to a decoder for final segmentation prediction.

Meta-training Schema For mutual benefits of the two learning processes, we propose a meta-training schema to conduct alternate optimization. We denote the weights of feature extractors as $\phi_{E_{ge}} = \{\phi_{E_{1,2}}, \phi_{E_{dila}}\}$ and $\phi_{E_{sp}} = \{\phi_{E_{1,2}}, \phi_{E_{3,4}}\}$, specific learner as $\phi_{L_{sp}}$, generic learner as $\phi_{L_{ge}}$ and other model weights ϕ_m. Specific learning and generic learning are properly trained in turn for each epoch. $\phi_{E_{1,2}}$ and ϕ_m are shared weights which are updated in both learning processes. $\phi_{L_{sp}}$ is updated only in specific learning while $\phi_{L_{ge}}$ is optimized only in generic learning. The whole training flow is detailed in Algorithm 1. The network predicts the category of each pixel of I^{que}, and the Dice loss is employed to optimize the network in an end-to-end manner during training. For the testing phase, only one labeled support image for each unseen-class disease is provided as guidance, and the query image of the same category is forwarded through the specific learning network without changing the parameters. With the obtained different learning abilities from the two learners, dense comparisons of both specific and general features are performed between the support and query images to generate a better prediction.

Algorithm 1: Cross-domain meta-training schema

Require: D_n, D_m, loss function $\mathcal{F}(.)$;
Initialize: $\phi_{E_{ge}}$, $\phi_{E_{sp}}$, $\phi_{L_{ge}}$, $\phi_{L_{sp}}$, ϕ_m and lr α, β;
Initialize the encoder from Pre-trained weights;
for *each epoch* **do**
  Specific learning:
  for *each iteration* **do**
    Sample support samples $\mathcal{S}$ and query samples $\mathcal{Q}$ from D_m;
    Freeze $\phi_{L_{ge}}$;
    Gradients $\nabla_{\Theta_{sp}} = \mathcal{F}'_{\Theta_{sp}}(\mathcal{S}, \mathcal{Q}; \phi_{E_{sp}}, \phi_{L_{sp}}, \phi_{L_{ge}}, \phi_m)$;
    Update parameters: $(\phi_{E_{sp}}, \phi_{L_{sp}}, \phi_m) = (\phi_{E_{sp}}, \phi_{L_{sp}}, \phi_m) - \alpha\nabla_{\Theta_{sp}}$
  end
  Generic learning:
  for *each iteration* **do**
    Sample $\mathcal{S}$ and $\mathcal{Q}$ from D_n;
    Freeze $\phi_{L_{sp}}$;
    $\nabla_{\Theta_{ge}} = \mathcal{F}'_{\Theta_{ge}}(\mathcal{S}, \mathcal{Q}; \phi_{E_{ge}}, \phi_{L_{sp}}, \phi_{L_{ge}}, \phi_m)$;
    $(\phi_{E_{ge}}, \phi_{L_{ge}}, \phi_m) = (\phi_{E_{ge}}, \phi_{L_{ge}}, \phi_m) - \alpha\nabla_{\Theta_{ge}}$
  end
end

3. EXPERIMENT

3.1. Implementation and Evaluation Metric

All the experiments are implemented in Keras with Tensorflow backend and trained on an NVIDIA Tesla V100 32GB GPU. The input images of FSS-100 and PH2 dataset are re-

Table 1: Quantitative results (DSC score) of different experimental settings and comparison with state-of-the-art methods on PH2 dataset with different unseen-class skin lesions.

Method	Melanomas	Common Nevus	Atypical Nevus	Average
Supervised training	89.50	92.23	90.12	90.62
w/o FSS	77.30	90.63	85.34	84.42
w/ FSS (FSS-1000)	61.38	60.11	56.83	59.44
w/ FSS (PH2)	84.60	92.33	90.14	89.02
w/ FSS (Fine-tune)	85.10	92.45	90.58	89.38
Feijie et al.[13]	88.03	91.42	92.28	90.58
LFT [15]	87.65	92.02	88.73	89.47
Ours	**91.35**	**94.48**	**93.26**	**93.03**

sized to 224×224. The Adam optimizer is adopted and the learning rate (lr) α of L_{sp} and β of L_{ge} are set as 0.0001 and 0.001, respectively. The batchsize and epoch are set as 1 and 40. We utilize Dice similarity coefficient (DSC) as the evaluation metric to measure the volumetric overlap between segmentation results and annotations.

3.2. Datasets

FSS-1000 We employ a latest few-shot object segmentation dataset FSS-1000 [18] for generic learning. This dataset consists of 1000 natural object classes, such as animals, fruits, and daily objects. Each class contains 10 images with corresponding pixel-wise ground-truth segmentation. All images are randomly sampled in the training phase.

PH2 Dataset We utilize PH2 dataset [19] with a total of 200 dermoscopic skin lesion images for specific learning and evaluation, including 80 common nevi, 80 atypical nevi and 40 melanomas. We select two diseases as seen classes in the training phase and the third one as the unseen rare class.

3.3. Quantitative and Qualitative Results

3.3.1. Comparison with Strong Baseline Methods

Table 1 demonstrates the strong generalization ability of few-shot segmentation to unseen classes. We first obtain the performance of fully supervised training as the upper bound and 'w/o FSS' as the lower bound, i.e., without applying any few-shot methods. The strong baseline models trained on seen common diseases are directly tested on unseen diseases. Here, we employ the method [20] ranked at the top of the PH2 segmentation task as a strong baseline and report its performance in the first line. 'w/ FSS (FSS-1000)' and 'w/ FSS (PH2)' report the results of few-shot segmentation methods, respectively, using only FSS-1000 data and PH2 seen classes during the training phase. We can observe that few-shot learning brings significant improvements over 'w/o FSS'. Especially, when generalizing to a much different disease Melanomas, 'w/ FSS (PH2)' yields nearly 8% improvement. Conversely, 'w/ FSS (FSS-1000)' leads to a large performance degradation, further indicating the large difficulties of cross-domain few-shot segmentation. Then, we found the strategy that pre-trains the few-shot segmentation model on FSS-1000 followed by a fine-tuning on PH2 seen classes (denoted as 'w/ FSS (Fine-tune)') is also beneficial. Despite the slight improvement, it also sheds light on the possibility of few-shot learning from natural to medical domain.

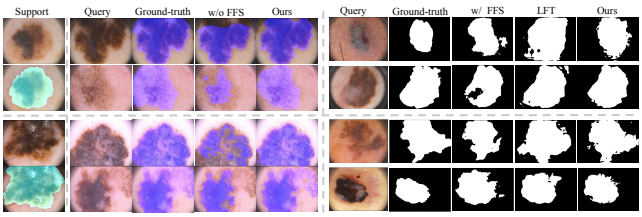

Fig. 3: Different query images with their predictions on unseen-class skin lesions with the guidance of the corresponding support image on the left.

3.3.2. Comparison with State-of-the-art Methods

The last three lines in Table 1 show the comparisons of our method with two State-of-the-art methods [15, 13] in few-shot learning area. We re-implement these methods with the same network for a fair comparison. First, we integrate the Learned Feature-Wise Transformation (LFT) module [15], which aims to solve the few-shot classification task under domain shift, into our network. Then, we choose a latest semi-supervised few-shot segmentation method [13] in medical area, referred to as Feijie et al., which adopts FSS-1000 data denoising as an unlabeled surrogate task during episodic training. Our framework performs the best and brings the most increment over 'w/ FSS (PH2)'. Significantly, our method even outperforms the upper bound supervised training on all the unseen lesions. This indicates that our framework can exploit learning ability from the natural domain to the medical domain more effectively than other existing methods without using the unseen-class data during training. As shown in Fig. 3, during testing on the unseen-class disease, our method can successfully segment the query images with appearance variations.

4. CONCLUSION

In this paper, we tackle the problem of rare-disease skin lesion segmentation using few-shot learning. We develop a cross-domain few-shot segmentation framework, which utilizes learning ability obtained from the natural domain to enhance the model's generalization to unseen-class diseases. Our work not only outperforms existing state-of-the-art methods, but also raises the awareness of researches on rare diseases and helps the community explore more about cross-domain few-shot learning in the medical domain.

5. REFERENCES

[1] Navid Razmjooy, Mohsen Ashourian, Maryam Karimifard, Vania V. Estrela, Hermes J. Loschi, Douglas do Nascimento, Reinaldo P. França, and Mikhail Vishnevski, "Computer-aided diagnosis of skin cancer: A review," *Current Medical Imaging Reviews*, vol. 16, no. 7, pp. 781–793, 2020.

[2] Xiaomeng Li, Lequan Yu, Yueming Jin, Chi-Wing Fu, Lei Xing, and Pheng-Ann Heng, "Difficulty-aware meta-learning for rare disease diagnosis," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, Cham, 2020, pp. 357–366.

[3] M Khoury and R Valdez, "Rare diseases, genomics and public health: an expanding intersection," *Genomics and Health Impact Blog*, 2016.

[4] X. Li, L. Yu, H. Chen, C. W. Fu, L. Xing, and P. A. Heng, "Transformation-consistent self-ensembling model for semisupervised medical image segmentation," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–12, 2020.

[5] Wenjia Bai, Ozan Oktay, Matthew Sinclair, Hideaki Suzuki, Martin Rajchl, Giacomo Tarroni, Ben Glocker, Andrew King, Paul M. Matthews, and Daniel Rueckert, "Semi-supervised learning for network-based cardiac mr image segmentation," in *Medical Image Computing and Computer-Assisted Intervention MICCAI 2017*, Cham, 2017, pp. 253–260.

[6] S. M. Jaisakthi, Aravindan Chandrabose, and P. Mirunalini, "Automatic skin lesion segmentation using semi-supervised learning technique," 2017.

[7] B. Bozorgtabar, S. Sedai, P. K. Roy, and R. Garnavi, "Skin lesion segmentation using deep convolution networks guided by local unsupervised learning," *IBM Journal of Research and Development*, vol. 61, no. 4/5, pp. 6:1–6:8, 2017.

[8] C. Chen, Q. Dou, H. Chen, J. Qin, and P. A. Heng, "Unsupervised bidirectional cross-modality adaptation via deeply synergistic image and feature alignment for medical image segmentation," *IEEE Transactions on Medical Imaging*, vol. 39, no. 7, pp. 2494–2505, 2020.

[9] Xiaolin Zhang, Yunchao Wei, Yi Yang, and Thomas S. Huang, "Sg-one: Similarity guidance network for one-shot semantic segmentation," *IEEE Transactions on Cybernetics*, vol. PP, pp. 1–11, 06 2020.

[10] K. Wang, J. H. Liew, Y. Zou, D. Zhou, and J. Feng, "Panet: Few-shot image semantic segmentation with prototype alignment," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 9196–9205.

[11] C. Zhang, G. Lin, F. Liu, R. Yao, and C. Shen, "Canet: Class-agnostic segmentation networks with iterative refinement and attentive few-shot learning," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 5212–5221.

[12] Cheng Ouyang, Carlo Biffi, Chen Chen, Turkay Kart, Huaqi Qiu, and Daniel Rueckert, "Self-supervision with superpixels: Training few-shot medical image segmentation without annotation," in *Computer Vision – ECCV 2020*, Cham, 2020, pp. 762–780.

[13] Abdur R Feyjie, Reza Azad, Marco Pedersoli, Claude Kauffman, Ismail Ben Ayed, and Jose Dolz, "Semi-supervised few-shot learning for medical image segmentation," 2020.

[14] Yunhui Guo, Noel C. Codella, Leonid Karlinsky, James V. Codella, John R. Smith, Kate Saenko, Tajana Rosing, and Rogerio Feris, "A broader study of cross-domain few-shot learning," in *Computer Vision – ECCV 2020*, Cham, 2020, pp. 124–141.

[15] Hung-Yu Tseng, Hsin-Ying Lee, Jia-Bin Huang, and Ming-Hsuan Yang, "Cross-domain few-shot classification via learned feature-wise transformation," in *International Conference on Learning Representations*, 2020.

[16] Jason Yosinski, Jeff Clune, Anh Nguyen, Thomas Fuchs, and Hod Lipson, "Understanding neural networks through deep visualization," 2015.

[17] Matthew D. Zeiler and Rob Fergus, "Visualizing and understanding convolutional networks," in *Computer Vision – ECCV 2014*, Cham, 2014, pp. 818–833, Springer International Publishing.

[18] Xiang Li, Tianhan Wei, Yau Pun Chen, Yu-Wing Tai, and Chi-Keung Tang, "Fss-1000: A 1000-class dataset for few-shot segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[19] Teresa Mendonça, Pedro Ferreira, André Marçal, Catarina Barata, Jorge Marques, Joana Rocha, and Jorge Rozeira, *PH2: A Public Database for the Analysis of Dermoscopic Images*, pp. 419–439, 09 2015.

[20] Prashant Brahmbhatt1 and Siddhi Nath Rajan2, "Skin lesion segmentation using segnet with binary cross-entropy," 07 2019.