Semi-supervised Multi-domain Learning for Medical Image Classification

Thesis to be submitted in partial fulfillment of the requirements for the degree

of

Bachelor of Technology

by

Ruchika Chavhan
170260011
Department of Physics

Under the guidance of

Prof. Biplab Banerjee
Centre of Studies in Resources Engineering

Prof. Mohammed Aslam
Department of Physics

INDIAN INSTITUTE OF TECHNOLOGY BOMBAY

ACKNOWLEDGEMENTS

I would like to thank Prof. Biplab Banerjee for his guidance throughout this project. I am extremely grateful to him for the encouragement, motivation and the opportunities that he has provided in the span of two years that I have worked with him. His insightful feedback pushed me to sharpen my thinking and brought my work to a higher level.

I am also thankful to Prof. Nibaran Das for the several helpful discussion sessions.

In addition, I would like to thank my parents for their wise counsel and sympathetic ear. I am thankful to my sister Vishakha and my friend Abhishek for their constant support.

Ruchika Chavhan
IIT Bombay

ABSTRACT

We consider the task of performing semi-supervised image classification for multiple visual domains in medical data using a single integrated framework to alleviate two salient limitations: domain dependence of neural networks and data scarcity. Under this premise, we learn a universal parametric family of neural networks, which share a majority of their weights across domains by learning a few adaptive domain-specific parameters. We train these universal networks on a suitable *pretext task* that captures a meaningful representation for image classification and further finetune the networks using a small fraction of training data. We perform our experiments on five medical datasets spanning breast, cervical, and colorectal cancer. Extensive experiments on architectures of domain-adaptive parameters demonstrate that our data-deficient universal model performs equivalent to a fully supervised setup, rendering a semi-supervised multi-domain setting for medical data extremely feasible in the real world.

Keywords: Semi-supervised Learning, Multi-domain Learning, Medical image classification

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The adoption of machine learning techniques for medical data and image analysis has significantly expanded and gathered immense attraction from researchers in recent decades [24][22]. Consequently, several deep learning techniques are being rapidly utilized to automate and improve predictions in the fields of genomics research [11], cancer prognosis [4], and medical imaging [25]. The effect of this permeation can also be seen widely in the the field of medical disorder classification, tumor/lesion segmentation, abnormality detection in areas like neurology [2], ophthalmic [10] and thoracic imaging [38], and digital and microscopic pathology [16]. Observing research from the past few decades, it can be deduced that the proliferation of computer automation in medical research has led to an unprecedented expansion of algorithms and datasets catering to medical researchers.

1.1 Learning with data scarcity

However, owing to the data-centric nature of machine learning, a surge in research implies an escalated need for labeled data collection, which is both time consuming and laborious. The lack of labeled data has motivated myriad research in the field of unsupervised [32], self-supervised [27], and semi supervised [9] learning. These learning paradigms are discussed in detail below.

1.1.1 Self-supervised Learning

Self-supervised learning is an elegant subset of machine learning where a model can generate output labels intrinsically from unlabeled data. The *self-supervised*, also

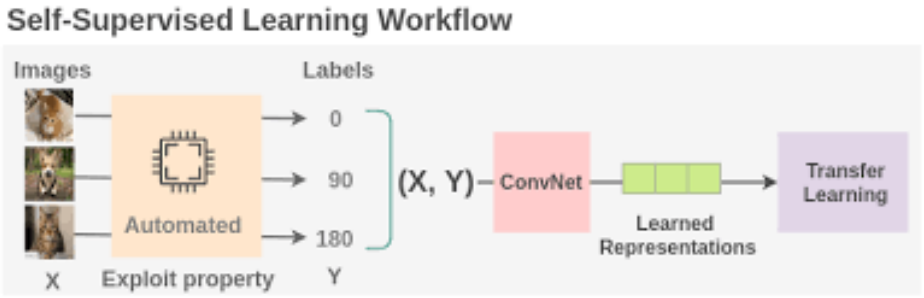

Figure 1.1: Self supervised learning. Credits: https://amitness.com/2020/02/illustrated-self-supervised-learning/

known as the *pretext task* guides a supervised loss function by learning inherent properties or semantic representations of the objects which are further used for related downstream tasks. A convolutional neural network (CNN) is trained to solve the pretext task and generate pseudo labels for the dataset based on the attributes learnt from the objective function of the pretext task. In some works, intrinsic latent representations learnt by solving a pretext task are utilized for semi-supervised learning with scarce annotations[31][21][36]. Parameters of the model trained on the self-supervised task are finetuned on a few annotated samples to perform a downstream visual task.

1.1.2 Semi-supervised Learning

Semi-supervised learning considers a few labeled samples available with a large amount of unlabeled data. The goal of a semi-supervised learning model is to make effective use of all of the available data, labeled and unlabeled. The semi-supervised setup holds immense utility in real world applications, especially for medical data on account of limited annotations, patients, and means of data collection. As a result, a rich body of literature exists for semi-supervised learning and self-supervised learning for medical data as well [1][19][14][15][5][3].

1.1.3 Unsupervised Learning

Unsupervised learning is a setting in which no labels are provided, where the algorithm is expected to learn patterns in the data based on similarity metrics. The most common clustering algorithms are k -means, hierarchical clustering, and Gaussian mixture models, etc. On higher dimensional data like images, the first step is to

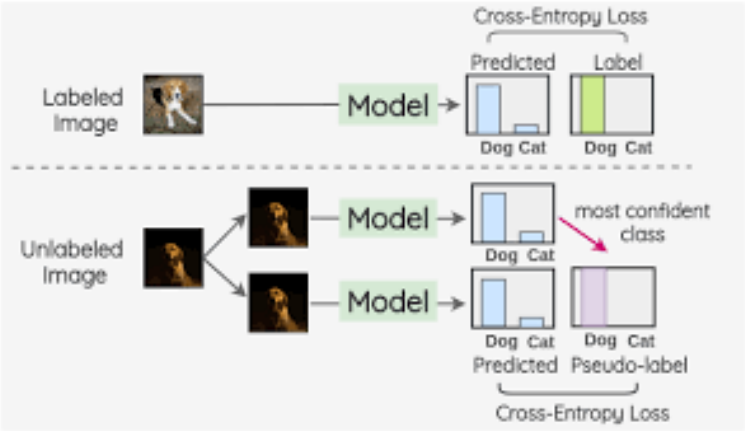

Figure 1.2: Semi supervised learning. Credits: https://amitness.com/2020/07/semi-supervised-learning/

project the data in a lower dimensional latent space. Clustering is then performed on this data.

1.2 Learning domain-independent Learning

Another hindrance in enhancing the scalability of machine learning techniques is that models understand multiple image datasets independently by learning separate models for every visual domain. Furthermore, these restrictions are amplified for medical datasets due to factors like scarcity, the rarity of the disease, the risk of data misuse, and lack of data-sharing incentives. Recently, research in the field of multi-domain learning [29][30] is proliferating guided by the aim of learning universal representations and feature extractors that can operate over several different visual domains. The primary goal is to develop models that can compactly represent multiple domains by leveraging associative knowledge between low and mid-level features of visually distinct domains. The underlying working principle behind these models is that multiple domains should share majority of their parameters except for certain weights, named adapters, that depend on the distribution of individual domains. Details of the multi-domain setup is described in detail in subsequent sections.

We jointly tackle the problem of scarce annotations and multi-domain training on medical datasets from distinct visual domains. We perform semi-supervised training on multiple domains by learning generalized representations using a vast majority of

shared parameters and a few domain-specific adapters. Essentially, this is equivalent to learning a single framework for multiple domains with the addition of a very few dependent parameters. By virtue of training only a singular universal model on multiple domains, a major merit of the proposed method is that we can use a single model for scarce datasets corresponding to distinct organs and tissues or obtained from different laboratories using disparate preparation methods. To demonstrate this, we perform our experiments on the task of image classification for five medical datasets spanning over breast, cervical, and colorectal cancer which are collected for either tissue (Histology) or cell (cytology) study. Our contributions in this thesis are three-fold:

- To the best of our knowledge, our approach is the first to perform medical image classification on a distinct variety of datasets under the multi-domain setting.
- Moreover, we challenge ourselves by introducing restrictions on annotated data collected from multiple domains, operating in a semi-supervised setting.
- We perform several experiments on different architectures of adapters to analyse and compare their behaviour in the presence of limited supervision.

The report is divided into 6 sections. Section 2 contains a brief description of previous works in the field of domain adaptation for medical data. Section 3 describes the proposed methodology of our proposed framework. Section 4 contains the Experimental Setup and Implementation details. In Section 5, we present the qualitative and quantitative results of our model. Finally, In section 6 we summarize the Self-supervised multi-domain learning framework.

Chapter 2

Related Work

In this section, we review prior works on multi-domain learning in medical data.

Although, there has been remarkable research in learning models for multiple domains in medical data [7][34][20][23], we argue that the methodology and goals of the proposed frameworks are significantly different from ours. We would like to indicate that learning models for multiple domains and a multi-domain setting are substantially different. For example, [6] introduced a framework for early Alzheimer's disease detection that utilize transfer learning to simultaneously learn the task and leverage information from multi-auxiliary domains to excel on the target domain. In the same spirit, [26] considers a self-supervised domain adaptation setting using multiple datasets for glaucoma detection. The above methods mainly focus on a transfer learning and domain adaptation pipeline that assume a certain degree of similarity between visual domains related to one body part. In contrast, our model is capable of operating without any assumptions about similarities between datasets as we consider domains from different body parts and sources to perform different image classification tasks. We would like to point out that the above works perform detection of only a single disease, implying that detection of different disease would require us to train a completely new model. Contrarily, our multi-domain universal models can perform detection of different diseases using a single universal model with only a few domain-specific adapters. Hence, this setup evinces itself to be extremely practical in a real life scenario where it is unfeasible to perpetually keep shifting between neural networks for different types of diagnosis.

Chapter 3

Proposed Methodology

Under the multi-domain setting, we aim to train neural networks that share a majority of their parameters across domains with exception of a few adaptive ones denoted by θ_d for $d = 1, 2, ...D$, where D is the total number of domains. We denote input images of each domain by $\mathcal{X}_d \in \mathbf{R}^{H \times W \times 3}$, projected onto a feature space $\mathcal{F}_d \in \mathbf{R}^{H_f \times W_f \times C_f}$ by a convolutional feature extractor $\mathcal{E}_\Theta : \mathbf{R}^{H \times W \times 3} \rightarrow \mathbf{R}^{H_f \times W_f \times C_f}$. Here, $\Theta = \{\theta_d \cup \psi\}$ denotes the complete set of domain-dependent θ_d and independent parameters ψ for the feature extractor, such that $\theta_d \cap \psi = \emptyset$. Subsequently, we obtain the labels for an image from domain d using a domain-specific linear classifier $\mathcal{L}_{\phi_d} : \mathbf{R}^{H_f \times W_f \times C_f} \rightarrow \mathbf{R}^{|C|_d}$, where $|C|_d$ is the number of categories in the domain. It is assumed that $\mathcal{L}_d$ consists of a softmax layer which returns a normalized probability distribution $\mathcal{P}_d = \mathcal{L}_d(\mathcal{E}(\mathcal{X}_d))$ over all classes in domain d. For notational convenience, we drop the parameters Θ_d and ϕ_d from the feature extractor and the linear classifier. To sum up, the domain-agnostic parameters are denoted by ψ while θ_d and ϕ_d are the domain-specific parameters.

3.1 Residual Adapters

A ResNet block [12] is a function $r_w : \mathbf{R}^{H \times W \times C} \rightarrow \mathbf{R}^{H \times W \times C}$ parametrised by weights w that performs the operation $r_w(x) = x + w \star x$. Here, the operation $\star$ consists of convolutional with batch normalization and ReLU function.

The primary idea behind residual adapters [29] is to modify the conventional residual network to contain domain-specific parameters. Naturally, to adapt a residual block $r_{i,w,d}$ for domain d, its parameters w need to be replaced by domain-specific

weights $\theta_{i,d}$, $i = 1, 2, ...|R|$, where $|R|$ is the number of residual adapters in the universal neural network. In order to restrict the number of domain-dependent parameters, the convolution layers in θ_d are implemented in the form of a filter bank of size 1×1. Apart from convolution filters, the scaling parameters of batch-normalization are also incorporated in the residual adapter modules by virtue of normalized outputs and stable training. It is worth noting that batch-normalization inherently consists of domain-specific scaling parameters that adds a certain degree of adaptation in the network. These residual adapters can be positioned in two ways [30] with respect to the domain-agnostic parameters: parallel and in series. We direct the readers to [30] for a thorough analysis of parallel and series residual adapters.

We discuss these two architectural positions in detail in this section. To further our explanation, we define a diagonal convolution operator (Equation 3.1) as done in [30]. Here, an operator $diag(F(A)) \in \mathbf{R}^{F \times F \times C \times D}$ converts a matrix $A \in \mathbf{R}^{C \times D}$ to a bank of diagonal filters. This operator transforms the matrix A into a 1×1 filter bank embedded as the central element of a larger $L \times L$ filter bank by appending zeros around it.

$$diag(F(A))_{abcd} = A_{dc} a = b = \frac{(F-1)}{2} + 10 otherwise \tag{3.1}$$

Here, F, C, and D denotes the filter size, input channels and output channels respectively,

3.1.1 Parallel Adapters

In this configuration, the adapter modules are placed parallel to domain-independent modules as shown in Figure. The output of $\psi \in \mathbf{R}^{F \times F \times C \times D}$ and the adapter module are computed in parallel and further added. Following this process, the output of a parallel residual adapter is given by Equation 3.2.

$$y = \psi \star x + diag_1(\theta_d) \star x \tag{3.2}$$

$$y = (\psi + diag_1(\theta_d)) \star x \tag{3.3}$$

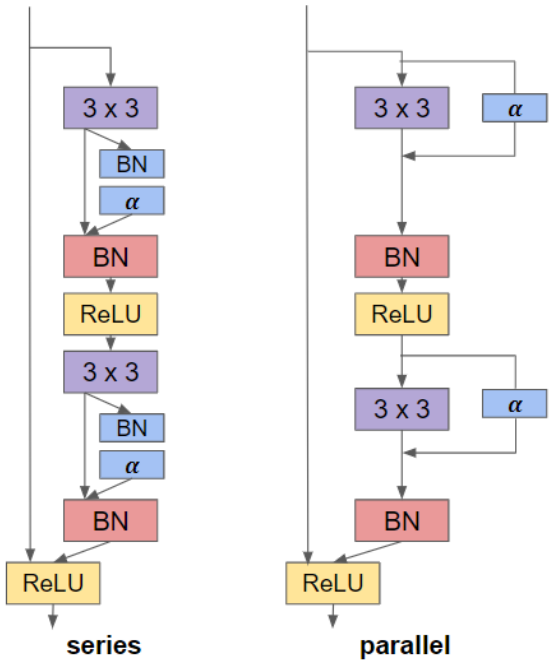

Figure 3.1: Series and parallel residual adapters. The blue colored blocks denote domain-dependent trainable parameters.

3.1.2 Series adapters

Residual adapters in the series configuration operate after domain-independent filter banks $\psi \in \mathbf{R}^{F \times F \times C \times D}$. The complete formulation of series adapters is given by Equation 3.4.

$$y = \rho(\mathbf{x}; \alpha) = \mathbf{x} + diag_1(\alpha) * \mathbf{x} \tag{3.4}$$

$$\rho(\mathbf{x}; \alpha) = diag_1(I + \alpha) * \mathbf{x} \tag{3.5}$$

3.2 Multi-domain semi-supervised training

Let us denote the pretext task/self-supervised task and the downstream task by $\mathcal{T}_p$ and $\mathcal{T}_m$ respectively. In this paper, we consider the downstream task to be multi-class image classification. Under this premise, a suitable pretext task performs auxiliary classification on d domains over $|C|_d^p$ categories obtained inherently from unlabeled data. We denote the datasets corresponding to pretext tasks and downstream tasks for

domain d by $D_{p,d} = \{\mathcal{X}_i^{p,d}, \mathcal{Y}_i^{p,d}\}_{i=1}^{N_{p,d}}$ and $D_{m,d} = \{\mathcal{X}_i^{m,d}, \mathcal{Y}_i^{m,d}\}_{i=1}^{N_{m,d}}$ respectively. Since, we operate in a semi-supervised setup, we assume that a few annotated samples are available to us for each domain such that $N_{p,d} >> N_{m,d} \forall d$. These annotated samples for task $\mathcal{T}_m$ are further used to finetune the parameters of the network trained on the pretext task $\mathcal{T}_p$.

We consider a neural network for image classification consisting of convolutional and fully-connected layers given by $\mathcal{N}(\cdot) = \mathcal{L}_d(\mathcal{E}(\cdot))$ such that $\mathcal{E} = \{\theta_d \cup \psi\}$. In multi-domain learning works [29][30], the domain-agnostic parameters ψ are obtained from a model pre-trained on a large dataset such as ImageNet while the domain-specific parameters are finetuned on corresponding domains. This ensures that a predominant number of parameters of the neural network are shared across domains. In the same spirit, we finetune only the domain-specific parameters θ_d, ϕ_d while solving both $\mathcal{T}_p$ and $\mathcal{T}_m$, while the domain-independent parameters ψ remain frozen and shared amongst all domains throughout the training process.

The entire training setup is demonstrated by Figure 3.2. Initially, we train the network $\mathcal{N}$ to solve the pretext task $\mathcal{T}_p$ using $D_{p,d}$. In order to perform the classification based pretext task, we introduce a different linear classifier $\hat{\mathcal{L}_{\phi_d\prime}} : \mathbf{R}^{H_f \times W_f \times C_f} \rightarrow \mathbf{R}^{|C|_d^p}$, which provides a probability distribution over auxiliary categories. The output of the network for the pretext task is obtained as $\hat{\mathcal{Y}_i^{p,d}} = \hat{\mathcal{L}}(\mathcal{E}(\mathcal{X}_i^{p,d}))$. Thereafter, the domain-specific parameters for the pretext task $\Theta_d\prime = \theta_d \cup \phi_d\prime$ are updated according to Equation 3.6, where $L_{p,d}(\cdot,\cdot)$ stands for cross-entropy loss for each domain and $B_{p,d}$ stands for the size of a minibatch.

$$\Theta_d\prime \leftarrow \Theta_d\prime - \eta \sum_{i=1}^{B_{p,d}} \frac{dL_{p,d}(\hat{\mathcal{Y}}_i^{p,d}, \mathcal{Y}_i^{p,d})}{d\Theta_d}, \forall d \tag{3.6}$$

By training our network on a befitting pretext task, we have ensured that our network comprises a meaningful semantic representation capable of supplementing knowledge for a downstream task. Consequently, the final step is to finetune only the parameters of the classification layer $\mathcal{L}_{\phi_d}$ on the downstream task using a few annotated samples and update its parameters using Equation 3.7. Here, $\mathcal{Y}_i^{m,d} = \mathcal{L}(\mathcal{E}(\mathcal{X}_i^{m,d}))$ denotes the output of the downstream classification task and $L_{m,d}(\cdot,\cdot)$ stands for cross-entropy loss for corresponding domains.

$$\phi_d \leftarrow \phi_d - \eta \sum_{i=1}^{B_{m,d}} \frac{dL_{m,d}(\hat{\mathcal{Y}}_i^{m,d}, \mathcal{Y}_i^{m,d})}{d\phi_d}, \forall d \tag{3.7}$$

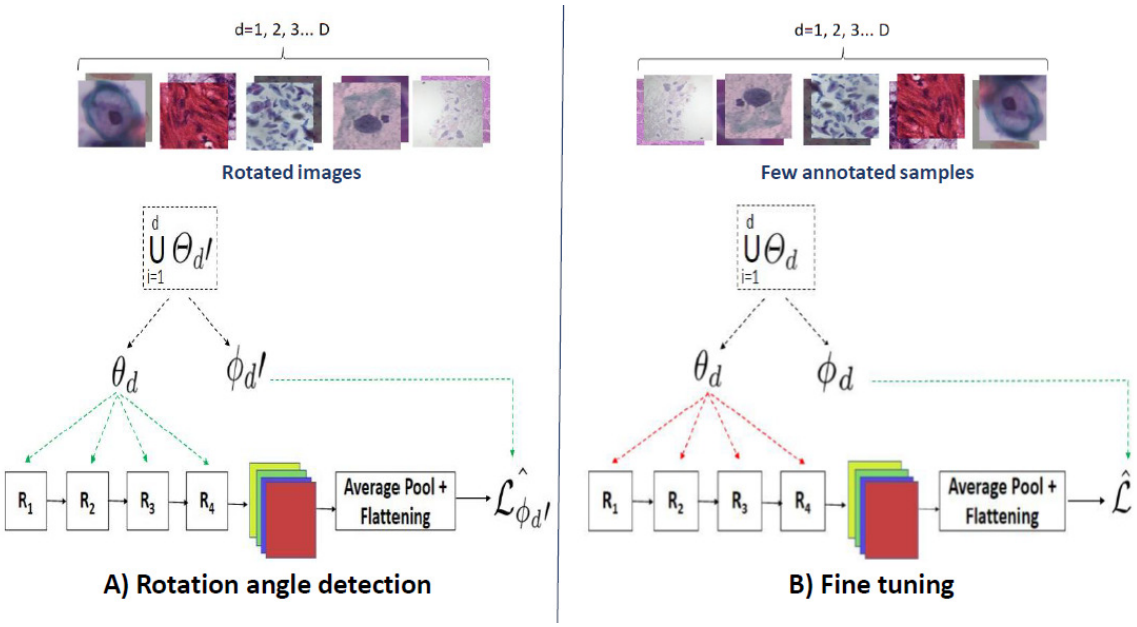

Figure 3.2: Overview of the Multi-domain semi-supervised setup. The domain-specific parameters for a domain are chosen from the union of all parameters given a domain index d. The green and red dashed arrows denote trainable and frozen domain-dependent parameters respectively.

During testing, domain-specific parameters $\theta_d \cup \phi_d$ corresponding to the domain d of the input image are retrieved from stored models to perform classification over $|\mathcal{C}_d|$ categories.

Algorithm 1 Training Algorithm

Stage 1: Self supervised training

Input: Pre-trained Resnet-26 model. Initialise parameters $\Theta_{d'}$.

1: **for** *epoch* $= 1$ to epochs **do**
2: Rotate $\mathcal{X}_i^{p,d}$ by an arbitrary angle from the set $Y_i^{p,d} = \{0, 90, 180, 270\}$.
3: $\hat{\mathcal{Y}}_i^{p,d} = \hat{\mathcal{L}}(\mathcal{E}(\mathcal{X}_i^{p,d}))$;
4: Update $\Theta_{d'}$ using Equation 3.6;
5: **end for**

Stage 2: Fine-tuning

Input: Frozen parameters θ_d. Initialise parameters ϕ_d.

6: **for** *epoch* $= 1$ to epochs **do**
7: $\mathcal{Y}_i^{m,d} = \mathcal{L}(\mathcal{E}(\mathcal{X}_i^{m,d}))$;
8: Update ϕ_d using Equation 3.7;
9: **end for**

Chapter 4

Experiments

4.1 Datasets

We perform experiments on five distinct datasets related to various types of cancers namely breast, colorectal and cervical. The data are collected by two techniques, i.e. Histology and Cytology.

- **Mendeley:** [13] This dataset consists of total 963 liquid based cytology pap smear images divided into four classes ($|C|_d = 4$) of pre-cancerous and cancerous lesions of cervical cancer namely High squamous intra-epithelial lesion, Low squamous intra-epithelial lesion, Negative for Intraepithelial malignancy, and Squamous cell carcinoma.

- **Herlev:** [17] This dataset contains 917 images of healthy and cancerous pap smears categorized into seven classes ($|C|_d = 7$) which are carcinoma, light dysplastic, moderate dysplastic, normal columnar, normal intermediate, normal superficiel, and severe dysplastic.

- **SIPaKMeD:** [28] This dataset is comprised of 4049 images pap smear slides divided into five categories ($|C|_d = 5$) containing normal, abnormal and benign cells specifically the superficial-intermediate, parabasal , koilocytotic, dyskeratotic, and metaplastic cells.

- **Kather:** [18] This dataset is a collection of histological images of human colorectal cancer sub-divided into eight classes ($|C|_d = 8$) of benign and malignant cancer.

- **BreakHis:** [35] This dataset contains 9,109 microscopic images of benign and malignant breast tumor tissue further subdivided into eight categories. However, for our experiments, we consider only two classes ($|C|_d = 2$) due to high inter-class similarity.

4.2 Pretext task

We choose rotation angle prediction (RotNet) [8] as the self-supervision task for our experiments. We rotate the image arbitarily by choosing one of the angles out of the set $\mathcal{A} = \{0, 90, 180, 270\}$. This implies that $|C|_d^p = 4$, $\forall d$. The parameters $\Theta_{d'}$ are trained to predict the angle by which the input image has been rotated. For all domains, we trained this proxy task for 200 epochs.

4.3 Architecture

We consider the baseline model to be a ResNet module in the 26-layer configuration as done in [30]. The network consists of 3 blocks of convolutional layers that output features containing 64, 128 and 256 channels respectively. Each block further consists of 4 residual blocks ($|R| = 4$) each, containing a domain-independent 3×3 convolutional layers followed by 1×1 domain-specific filter banks with a skip connection. The spatial resolution of the data is halved from a block to the next. The residual adapters are distributed throughout all the feature extractors modules which are followed by domain-specific classifiers. The model in Figure 3.2 depicts only one such convolutional block for ease of visualization of the training method.

4.4 Training details

We use the 80:20 training and validation set split for all the five datasets. We train the entire model using stochastic gradient optimization with a learning rate of 10^{-3}, momentum and weight decay. In contrast to [30], we do not use dropout in any of the residual adapters. We perform finetuning on the domain-specific classifier for 200 epochs for each domain. The training algorithm is mentioned in Algorithm 1.

Chapter 5

Results

5.1 Results and Discussion

In this section, we discuss the results obtained by our proposed methodology on the five datasets mentioned in Section 4.1. For every dataset we finetune the domain-specific parameters of the network with 10/25/50/100 labeled samples per class. In Fig 5.1, we present the performance of parallel residual adapters in the universal model on each dataset while varying the number of samples. Table 5.1 demonstrates the results of our universal model on all five medical datasets considered in this paper. We perform experiments on two architectural possibilities of the residual adapters: series and parallel and compare their performance. To demonstrate the generalization capability of our model under the constraints on labeled data, we also compare our results with the fully-supervised multi-modal training setup. We also provide an extensive qualitative visualisation of the performance of our models in the supplementary material.

Discussion: From Table 5.1, we observe that our model provides an accuracy that is almost equivalent to the performance of a fully-supervised model, inspite of severe data scarcity. As expected, the performance of all models increase with the introduction of more training samples. In some cases, addition of 15-25 samples provides a sharp boost in accuracies. We observe that in most cases the accuracy obtained with only 100 samples is appreciably close to model with full supervision. Interestingly, in many cases, the accuracy provided by some models with just 100 samples is better than full supervision.

Dataset	Model	Number of samples				
		10	25	50	100	Full
Mendeley	Parallel	76.52 ± 1.04	91.15 ± 0.52	97.40 ± 0.41	98.44 ± 0.52	$\mathbf{98.96 \pm 0.07}$
	Series	82.21 ± 1.12	91.67 ± 1.04	96.35 ± 0.52	$\mathbf{98.96 \pm 1.04}$	97.40 ± 1.86
Kather	Parallel	78.63 ± 0.61	90.63 ± 0.81	93.75 ± 1.41	$\mathbf{96.88 \pm 0.25}$	96.87 ± 2.02
	Series	74.60 ± 0.10	81.45 ± 0.20	88.21 ± 0.30	90.83 ± 0.20	$\mathbf{93.75 \pm 0.41}$
SipaKMed	Parallel	75.63 ± 0.38	90.63 ± 0.13	93.75 ± 0.28	94.13 ± 0.48	$\mathbf{94.63 \pm 1.63}$
	Series	78.13 ± 1.30	89.84 ± 0.25	92.19 ± 1.06	$\mathbf{96.88 \pm 0.88}$	95.25 ± 1.50
BreakHis	Parallel	43.75 ± 0.07	84.38 ± 0.52	87.50 ± 0.88	92.49 ± 0.30	$\mathbf{92.97 \pm 2.35}$
	Series	59.38 ± 0.02	84.64 ± 2.34	96.88 ± 0.21	$\mathbf{96.88 \pm 1.04}$	97.73 ± 0.85
Herlev	Parallel	50.63 ± 1.25	87.50 ± 2.13	93.75 ± 1.25	$\mathbf{98.86 \pm 0.63}$	98.66 ± 0.29
	Series	53.13 ± 1.48	75.00 ± 2.28	92.19 ± 0.63	96.36 ± 1.24	$\mathbf{97.32 \pm 0.85}$

Table 5.1: Results of the parallel and series adapters on five medical datasets by varying the number of samples per class. We report the accuracies in % of each model under different cases of limited supervision. "Full" denotes training with the complete training set. We also report the deviation of accuracy throughout the validation set. The greatest value in reach row is highlighted in **bold**.

5.2 Qualitative results

We provide a qualitative visualisation of results. For every dataset, we provide t-SNE plots and a a Grad-CAM analysis for images of each class. t-Distributed Stochastic Neighbor Embedding (t-SNE) [37] is an excellent tool to visualise high dimensional data and analyse similarities between data points. Gradient-weighted class activation mapping (Grad-CAM) [33] primarily uses the gradients of the target class at the final convolution layer to synthesize an intermediate localization map which highlights the most important regions in the image. The Grad-CAM plots effectively display the regions which contribute the most in prediction of a particular target-class. In Figure 5.2, 5.3, 5.4, **??**, 5.5, we present the t-SNE and Grad-CAM visualisation for all the datasets individually.

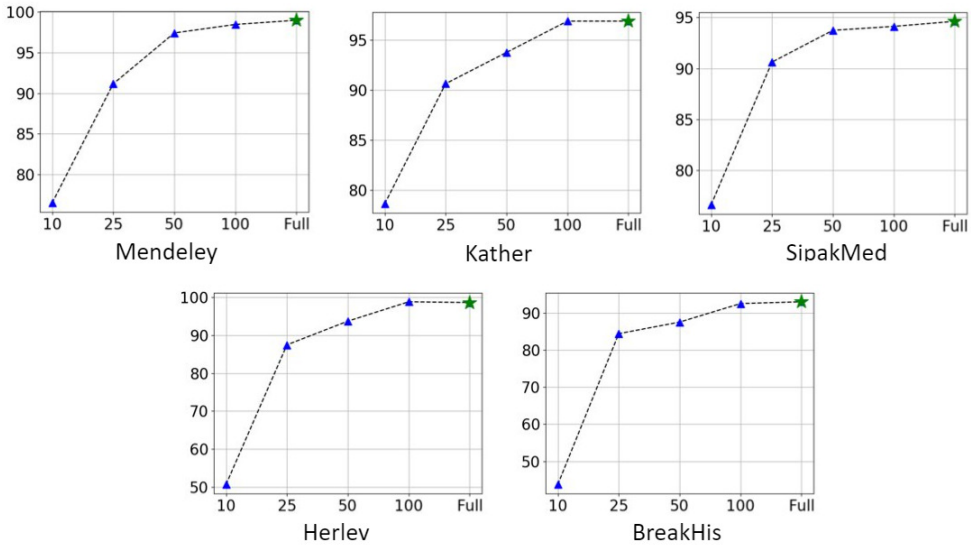

Figure 5.1: Comparison of accuracy of parallel adapters by varying the number of samples per class. x-axis denotes the number of samples and y-axis denotes accuracy of the model. The green star denotes the accuracy of the model under full supervision.

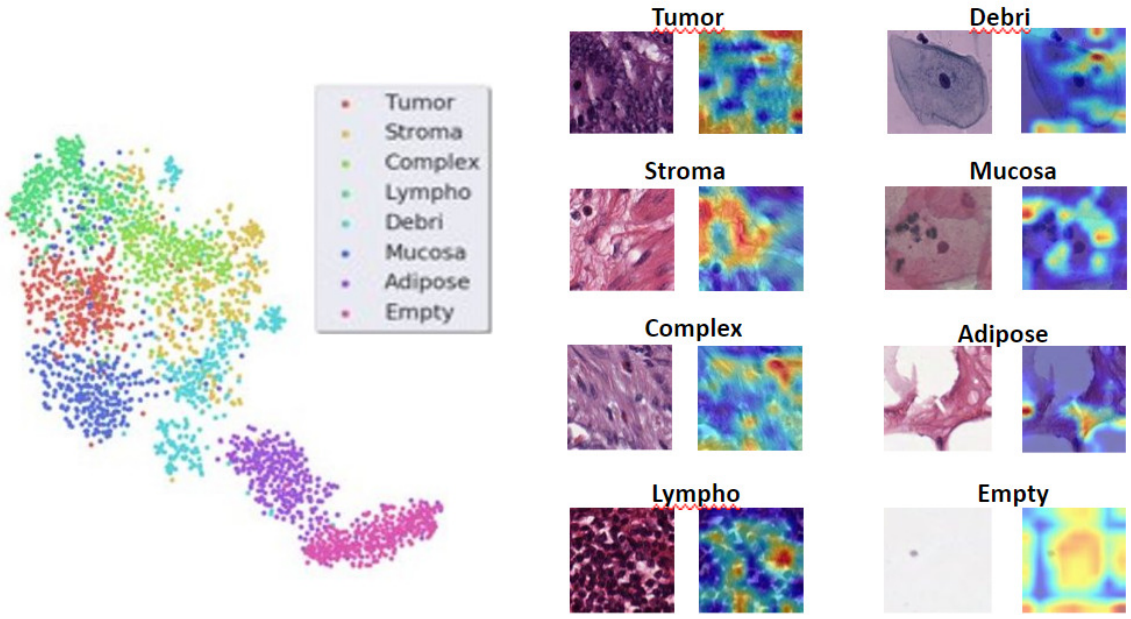

Figure 5.2: t-SNE and Grad-CAM visualisation for the Kather Dataset

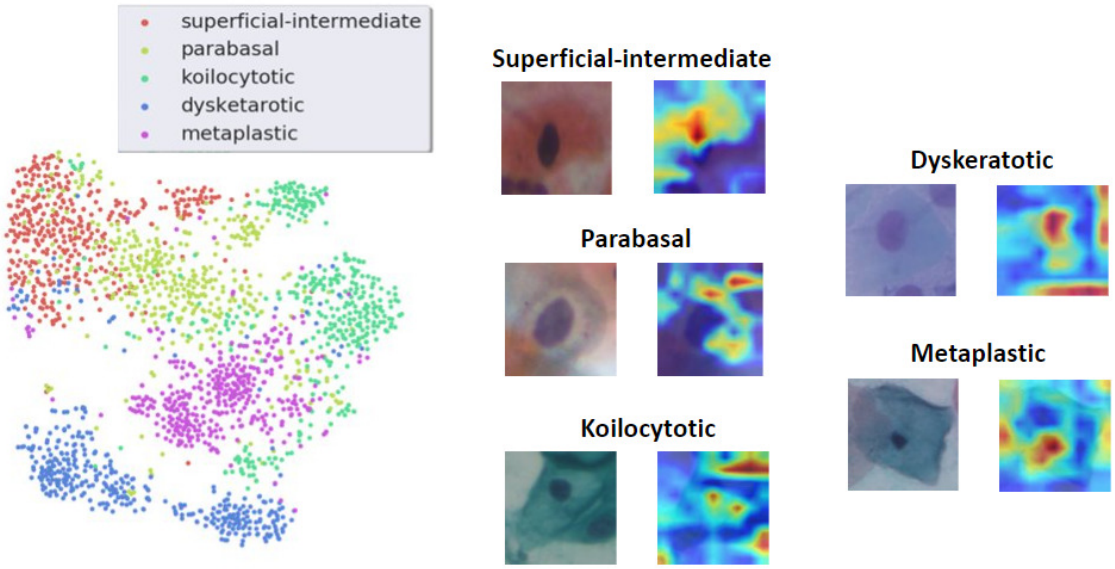

Figure 5.3: t-SNE and Grad-CAM visualisation for the SipakMed Dataset

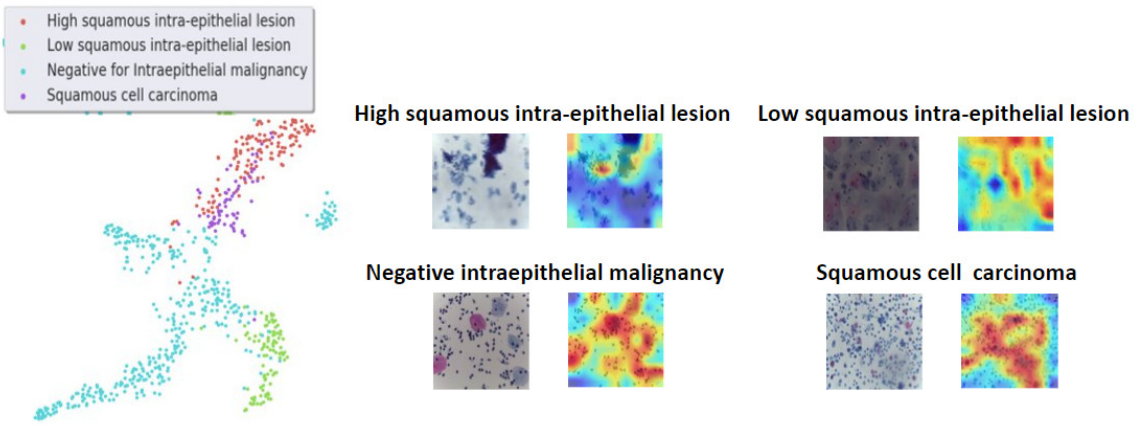

Figure 5.4: t-SNE and Grad-CAM visualisation for the Mendeley Dataset

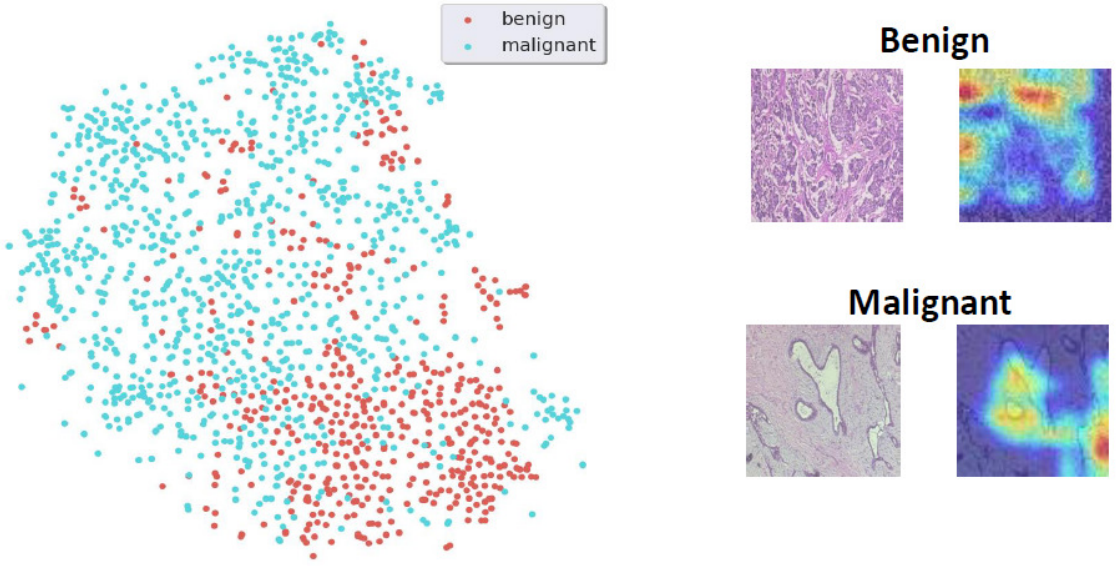

Figure 5.5: t-SNE and Grad-CAM visualisation for the Breakhis Dataset

Chapter 6

Conclusion

In this thesis, we proposed the concept of semi-supervised multi-domain learning in the domain of medical images. Our aim is to tackle two major restrictions that hinder the growth of machine learning in the medical domain, namely data scarcity and domain-dependence of models. To accomplish this, we introduce a universal family of models that share majority of their parameters except a few domain-specific parameters termed as adapters that leverage information from pretext tasks to perform image classification. We perform extensive experiments on five medical image datasets from different sub-domains namely, Medeley, Kather, SipakMed, BreakHis, and Herlev spanning breast, colorectal, and cervical cancer. Ultimately, we demonstrate that the performace of our models trained with as few as 100 samples is congruous with those trained under full supervision. This opens up riveting and exciting possibilities for a semi-supervised multi-domain setup for medical images under multiple scenarios. In the future, we wish to extend this setup to dense prediction tasks like semantic segmentation and object detection for tumor/lesion detection.

Bibliography

[1] Wenjia Bai, Chen Chen, G. Tarroni, Jinming Duan, F. Guitton, S. Petersen, Y. Guo, P. Matthews, and D. Rueckert. Self-supervised learning for cardiac mr image segmentation by anatomical position prediction. In *MICCAI*, 2019.

[2] Tom Brosch and Roger Tam. Manifold learning of brain mris by deep learning. 2013.

[3] Krishna Chaitanya, Ertunc Erdil, Neerav Karani, and Ender Konukoglu. Contrastive learning of global and local features for medical image segmentation with limited annotations. *Advances in Neural Information Processing Systems*, 2020.

[4] Anika Cheerla and Olivier Gevaert. Deep learning with multimodal representation for pancancer prognosis prediction. *Bioinformatics (Oxford, England)*, 07 2019.

[5] Liang Chen, Paul Bentley, Kensaku Mori, Kazunari Misawa, Michitaka Fujiwara, and Daniel Rueckert. Self-supervised learning for medical image analysis using image context restoration. *Medical Image Analysis*, 2019.

[6] Bo Cheng, Mingxia Liu, Zuoyong Li, and Daoqiang Zhang. Multi-domain transfer learning for early diagnosis of alzheimer's disease. *Neuroinformatics*, 2017.

[7] Yang Feng, Yubao Liu, and Jiebo Luo. Universal model for multi-domain medical image retrieval. 2020.

[8] Spyros Gidaris, Praveer Singh, and Nikos Komodakis. Unsupervised representation learning by predicting image rotations. 2018.

[9] Matthieu Guillaumin, Jakob Verbeek, and Cordelia Schmid. Multimodal semi-supervised learning for image classification. In *2010 IEEE Computer society*

conference on computer vision and pattern recognition, pages 902–909. IEEE, 2010.

[10] Varun Gulshan, Lily Peng, Marc Coram, Martin Stumpe, Derek Wu, Arunachalam Narayanaswamy, Subhashini Venugopalan, Kasumi Widner, Tom Madams, Jorge Cuadros, Ramasamy Kim, Rajiv Raman, Philip Nelson, Jessica Mega, and Dale Webster. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*, 316, 11 2016.

[11] Ehsan Hajiramezanali, S. Z. Dadaneh, Alireza Karbalayghareh, M. Zhou, and X. Qian. Bayesian multi-domain learning for cancer subtype discovery from next-generation sequencing count data. In *NeurIPS*, 2018.

[12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity mappings in deep residual networks. In Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, editors, *Computer Vision – ECCV 2016*, pages 630–645, Cham, 2016. Springer International Publishing.

[13] Elima Hussain, Lipi B. Mahanta, Himakshi Borah, and Chandana Ray Das. Liquid based-cytology pap smear dataset for automated multi-class diagnosis of pre-cancerous and cervical cancer lesions. *Data in Brief*, 2020.

[14] Abdullah-Al-Zubaer Imran, Chao Huang, Hui Tang, Wei Fan, Yuan Xiao, Dingjun Hao, Zhen Qian, and Demetri Terzopoulos. Self-supervised, semi-supervised, multi-context learning for the combined classification and segmentation of medical images (student abstract). *Proceedings of the AAAI Conference on Artificial Intelligence*, 34:13815–13816, 04 2020.

[15] Amir Jamaludin, Timor Kadir, and Andrew Zisserman. Self-supervised learning for spinal mris. *Workshop on Deep Learning in Medical Image Analysis*, 2017.

[16] Andrew Janowczyk, Ajay Basavanhally, and Anant Madabhushi. Stain normalization using sparse autoencoders (stanosa): Application to digital pathology. *Computerized Medical Imaging and Graphics*, 2017. Recent Developments in Machine Learning for Medical Imaging Applications.

[17] Jan Jantzen, Jonas Norup, George Dounias, and Beth Bjerregaard. Pap-smear benchmark data for pattern classification. *Nature Inspired Smart Information Systems (NiSIS)*, 01 2005.

[18] Jakob Nikolas Kather, Frank Gerrit Zöllner, Francesco Bianconi, Susanne M Melchers, Lothar R Schad, Timo Gaiser, Alexander Marx, and Cleo-Aron Weis. Collection of textures in colorectal cancer histology, 2016.

[19] Naji Khosravan. Semi-supervised multi-task learning for lung cancer diagnosis. volume 2018, pages 710–713, 07 2018.

[20] Zeljko Kraljevic, Thomas Searle, Anthony Shek, Lukasz Roguski, Kawsar Noor, Daniel Bean, Aurelie Mascio, Leilei Zhu, Amos A Folarin, Angus Roberts, Rebecca Bendayan, Mark P Richardson, Robert Stewart, Anoop D Shah, Wai Keong Wong, Zina Ibrahim, James T Teo, and Richard JB Dobson. Multi-domain clinical natural language processing with medcat: the medical concept annotation toolkit. 2020.

[21] Samuli Laine and Timo Aila. Temporal ensembling for semi-supervised learning. 10 2016.

[22] Tao Lei, Risheng Wang, Yong Wan, Xiaogang Du, Hongying Meng, and Asoke Nandi. Medical image segmentation using deep learning: A survey. 2020.

[23] Haoliang Li, YuFei Wang, Renjie Wan, Shiqi Wang, Tie-Qiang Li, and Alex C. Kot. Domain generalization for medical imaging classification with linear-dependency regularization. *Advances in Neural Information Processing Systems*, 2020.

[24] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen A.W.M. van der Laak, Bram van Ginneken, and Clara I. Sánchez. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 2017.

[25] Alexander Selvikvåg Lundervold and Arvid Lundervold. An overview of deep learning in medical imaging focusing on mri. *Zeitschrift für Medizinische Physik*, 2019.

[26] Nooshin Mojab, Vahid Noroozi, Darvin Yi, Manoj Prabhakar Nallabothula, Abdullah Aleem, Phillip Yu, and Joelle Hallak. Real-world multi-domain data applications for generalizations to clinical settings. 2020.

[27] Abass Olaode, Golshah Naghdy, and Catherine Todd. Unsupervised classification of images: A review. *International Journal of Image Processing (IJIP)*, 2014.

[28] M. E. Plissiti, P. Dimitrakopoulos, G. Sfikas, C. Nikou, O. Krikoni, and A. Charchanti. Sipakmed: A new dataset for feature and image based classification of normal and pathological cervical cells in pap smear images. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 3144–3148, 2018.

[29] S-A Rebuffi, H. Bilen, and A. Vedaldi. Learning multiple visual domains with residual adapters. In *Advances in Neural Information Processing Systems*, 2017.

[30] Sylvestre-Alvise Rebuffi, Hakan Bilen, and Andrea Vedaldi. Efficient parametrization of multi-domain deep neural networks. In *CVPR*, 2018.

[31] Sylvestre-Alvise Rebuffi, Sebastien Ehrhardt, Kai Han, Andrea Vedaldi, and Andrew Zisserman. Semi-supervised learning with scarce annotations. *Technical report*, 2019.

[32] Lars Schmarje, Monty Santarossa, Simon-Martin Schröder, and Reinhard Koch. A survey on semi-, self-and unsupervised learning for image classification. *arXiv:2002.08721*, 2020.

[33] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. *International Journal of Computer Vision*, 2019.

[34] L. Shen, W. Zhu, X. Wang, L. Xing, J. M. Pauly, B. Turkbey, S. A. Harmon, T. H. Sanford, S. Mehralivand, P. Choyke, B. J. Wood, and D. Xu. Multi-domain image completion for random missing input data. *IEEE Transactions on Medical Imaging*, 2020.

[35] Fabio Spanhol, Luiz Soares de Oliveira, Caroline Petitjean, and Laurent Heutte. Breast cancer histopathological image classification using convolutional neural networks. 07 2016.

[36] Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *Advances in Neural Information Processing Systems*, 2018.

[37] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 2008.

[38] X. Zhou, T. Ito, R. Takayama, S. Wang, T. Hara, and H. Fujita. Three-dimensional ct image segmentation by combining 2 d fully convolutional network with 3 d majority voting. 2016.