SEMI-SUPERVISED MULTI-DOMAIN MULTI-TASK TRAINING FOR METASTATIC COLON LYMPH NODE DIAGNOSIS FROM ABDOMINAL CT

Saskia Glaser[†]   *Gabriel Maicas*[†]   *Sergei Bedrikovetski*[‖]   *Tarik Sammour*[‖‡]   *Gustavo Carneiro*[† *]

[†] Australian Institute for Machine Learning, The University of Adelaide, Australia
[‖] Faculty of Health and Medical Science, School of Medicine, The University of Adelaide, Australia
[‡] Colorectal Unit, Department of Surgery, Royal Adelaide Hospital, Australia

ABSTRACT

The diagnosis of the presence of metastatic lymph nodes from abdominal computed tomography (CT) scans is an essential task performed by radiologists to guide radiation and chemotherapy treatment. State-of-the-art deep learning classifiers trained for this task usually rely on a training set containing CT volumes and their respective image-level (i.e., global) annotation. However, the lack of annotations for the localisation of the regions of interest (ROIs) containing lymph nodes can limit classification accuracy due to the small size of the relevant ROIs in this problem. The use of lymph node ROIs together with global annotations in a multi-task training process has the potential to improve classification accuracy, but the high cost involved in obtaining the ROI annotation for the same samples that have global annotations is a roadblock for this alternative. We address this limitation by introducing a new training strategy from two data sets: one containing the global annotations, and another (publicly available) containing only the lymph node ROI localisation. We term our new strategy semi-supervised multi-domain multi-task training, where the goal is to improve the diagnosis accuracy on the globally annotated data set by incorporating the ROI annotations from a different domain. Using a private data set containing global annotations and a public data set containing lymph node ROI localisation, we show that our proposed training mechanism improves the area under the ROC curve for the classification task compared to several training method baselines.

Index Terms— semi-supervised, multi-task, multiple domain, ROI annotations, weak annotations, colon cancer, lymph node diagnosis, abdominal CT.

1. INTRODUCTION

Colon cancer is one of the most frequently diagnosed cancers in the world, with nearly 1 million new cases and 551,269 deaths in 2018 [1]. Patients with cancer have a high risk for nodal metastasis, therefore oncologically adequate surgery consists of segmental colectomy with lymph node dissection followed by adjuvant treatment [2]. However approximately 20% to 30% of patients develop another cancer months later, thus alternative treatment strategies including neoadjuvant therapy prior to surgery are currently being investigated [3]. Under these circumstances, accurate preoperative metastatic lymph node diagnosis is crucial in determining the eligibility of neoadjuvant treatment and to avoid the over treatment of patients. However, diagnosing lymph nodes is a challenging task that is prone to inter-observer variability [4]. As a result, computer-aided diagnosis systems are being designed to assist radiologists in staging lymph node metastasis on cross-sectional imaging.

State-of-the-art deep learning classification models [5] could in principle produce relatively accurate diagnosis of metastatic lymph nodes. However, these models are trained with weakly annotated data sets (i.e. image-level, or global, labels), and fail to produce precise diagnosis in situations where the regions of interest (ROIs) that explain such diagnosis occupy a small portion of the input image, which is the case for metastatic lymph node diagnosis from abdominal CT scans. The classification accuracy of models which perform diagnosis that depends on relatively small ROIs can be improved by incorporating localization information to the training process [6], [7], [8]. The main goal is that the model learns to focus on the relevant ROIs to perform classification. For these models, the training process becomes a multi-task training, where the aims are to classify the entire image and to localize the ROIs in image. However, such training process is based on a strongly annotated training set that contains local ROI and global diagnosis annotations, which are costly and time consuming to obtain.

Aiming to reduce the burden of obtaining such ROI and global annotations for the single-domain multi-task training mentioned above, we hypothesize that global classification accuracy can be improved if we use separate data sets (i.e., multi-domain), each one containing one particular type of annotation (i.e., multi-task). This setup introduces the new challenge of having a multi-domain multi-target training mechanism, where each domain (i.e., data set) is annotated for a different task – this motivates us to propose a new semi-supervised learning approach that transfers the multi-

task labels between multiple domains.

In this paper, we introduce a new semi-supervised training strategy to train multi-domain multi-task models where different domains have different types of annotations. Our strategy consists of jointly training a multi-task model with semi-supervised learning on the unlabelled task of the other domains. More precisely, we design a training strategy that improves image classification in one domain by including a ROI localization task from a different domain. We evaluate our training method on the problem of metastatic colon lymph node diagnosis from abdominal CT, using a private data set that only contains global labels (RAH data set) and a public data set that only has lymph node ROI annotations (CI data set) [9]. Results on the RAH data set show that our semi-supervised multi-domain multi-task training strategy outperforms other training strategies such as semi-supervised or multi-task learning without semi-supervision.

2. RELATED WORK

Diagnosing metastatic lymph nodes from abdominal CT scans has traditionally been tackled with the design of hand-crafted features [10]. The main disadvantage of these methods lies in the sub-optimality of the features as there is no guarantee that these features are optimal to perform diagnosis, yielding relatively low accuracy scores. Feature sub-optimality has been addressed by the computer vision community by using deep learning classifiers that learn optimal features while being trained to perform diagnosis [5]. However, the performance of such classifiers decreases in problems where the regions of interest in the image to be classified are relatively small.

The medical image analysis community focused on including the localization of regions of interest in the training process [6], [7], [8] to increase classification performance, in a single-domain multi-task approach. The training process in such cases does not only optimise image classification, but also ROI localisation. Although successful, this increase in performance comes at the expense of strongly annotating training sets with ROIs and classification labels, a costly process for many medical image analysis problems.

Multi-task learning from multi-domain data [11] was proposed to learn from data sets that have been labelled for different but related tasks. The main aim is to improve the performance during inference on each of the labelled tasks in the data sets. However, these approaches require the multi-task annotations in all data sets to generalize well in each of the tasks across data sets, which is costly to be acquired. Moreover, multi-domain multi-task methods do not use all the information available in the data sets as they do not exploit, during training, un-labelled data for a given task.

Semi-supervised learning has been proposed to incorporate un-annotated data for a given task into the learning process. Initially proposed for incorporating unlabelled data from the same task and domain [12], [13], it has

recently been successfully explored in a multi-task set-up [14], [15]. Contrary to previous semi-supervised multi-task approaches [14], [15], we propose a training method that not only deals with tasks that have been labelled in the same domain, but also where each task is labelled in a different domain.

3. METHODS

3.1. Data Sets

Let $\mathcal{D}^{(1)} = \left\{ \left(\mathbf{v}_i^{(1)}, y_i^{(1)} \right) \right\}_{i \in \{1, \ldots, |\mathcal{D}^{(1)}|\}}$ be a weakly annotated data set, where $\mathbf{v}_i^{(1)} : \Omega \to \mathbb{R}$ represents the abdominal CT scan of the ith patient in the data set $\mathcal{D}^{(1)}$ and $\Omega \in \mathbb{R}^3$ is the volume lattice, and $y_i^{(1)} \in \{0, 1\}$ is the scan-level label indicating the presence ($y_i^{(1)} = 1$) or absence ($y_i^{(1)} = 0$) of any metastatic lymph node in the CT scan of patient i. Let $\mathcal{D}^{(2)} = \left\{ \left(\mathbf{v}_j^{(2)}, \mathbf{s}_j^{(2)} \right) \right\}_{j \in \{1, \ldots, |\mathcal{D}^{(2)}|\}}$ be a strongly annotated data set where $\mathbf{v}_j^{(2)} : \Omega \to \mathbb{R}$ represents the abdominal CT scan of the jth patient in the data set $\mathcal{D}^{(2)}$, and $\mathbf{s}_j^{(2)} : \Omega \to \{0, 1\}$ is the voxel-wise ROI annotation of the presence of a lymph node in the CT scan of patient j. Assume that each data set comes from a different domain, implying that the data set distributions $(\mathbf{v}, y) \sim P^{(1)}(\mathbf{v}, y)$ (for $(\mathbf{v}, y) \in \mathcal{D}^{(1)}$) and $(\mathbf{v}, \mathbf{s}) \sim P^{(2)}(\mathbf{v}, \mathbf{s})$ (for $(\mathbf{v}, \mathbf{s}) \in \mathcal{D}^{(2)}$) are different. Note that each data set is labelled for a different task: $\mathcal{D}^{(1)}$ for image classification and $\mathcal{D}^{(2)}$ for lymph node ROI detection. We split each data set into training, validation and testing sets in a patient-wise manner, forming the sets $\mathcal{T}_{train}^{(1)}$, $\mathcal{T}_{train}^{(2)}, \mathcal{T}_{val}^{(1)}, \mathcal{T}_{val}^{(2)}, \mathcal{T}_{test}^{(1)}$ and $\mathcal{T}_{test}^{(2)}$ for data sets $\mathcal{D}^{(1)}$ and $\mathcal{D}^{(2)}$.

3.2. Semi-supervised Multi-domain Multi-task training

Our model is composed of three modules: an encoder f_{θ_e}, a classification branch σ_{θ_c} and a detection branch g_{θ_d} with parameters $\theta_e, \theta_c, \theta_d$ respectively. The model receives as input a CT scan $\mathbf{v}$ and forms a feature embedding $\mathbf{o} = f_{\theta_e}(\mathbf{v})$ using the encoder. The embedding $\mathbf{o}$ is used as input by the classification and detection branches. The classification branch returns $\tilde{y} = \sigma_{\theta_c}(\mathbf{o})$ representing the binary classification of the input scan $\mathbf{v}$. The detection branch produces an ROI map $\tilde{\mathbf{s}} = g_{\theta_d}(\mathbf{o})$ that estimates the probability that each voxel represents a lymph node. A binary ROI mask $\tilde{\mathbf{s}}_\zeta$ is generated by thresholding $\tilde{\mathbf{s}}(\omega) > \zeta$, for $\omega \in \Omega$. See Fig. 1 for a summary of the architecture of our method.

The training consists of two alternating stages: 1) label propagation between the two data sets sets [16], and 2) multi-domain multi-task training using the real and propagated labels. For label propagation between the two data sets [16], we replace the original $\mathcal{D}^{(1)}$ by $\mathcal{D}_{new}^{(1)} = \mathcal{D}^{(1)} \bigcup \widetilde{\mathcal{D}^{(1)}}$, with $\widetilde{\mathcal{D}^{(1)}} = \{(\mathbf{v}, \tilde{y}) | \mathbf{v} \in \mathcal{D}^{(2)}, \tilde{y} = \sigma_{\theta_c}(f_{\theta_e}(\mathbf{v}))\}$ and the original $\mathcal{D}^{(2)}$ by $\mathcal{D}_{new}^{(2)} = \mathcal{D}^{(2)} \bigcup \widetilde{\mathcal{D}^{(2)}}$, with $\widetilde{\mathcal{D}^{(2)}} = \{(\mathbf{v}, \tilde{\mathbf{s}}) | \mathbf{v} \in \mathcal{D}^{(1)}, \tilde{\mathbf{s}} = g_{\theta_d}(f_{\theta_e}(\mathbf{v}))\}$.
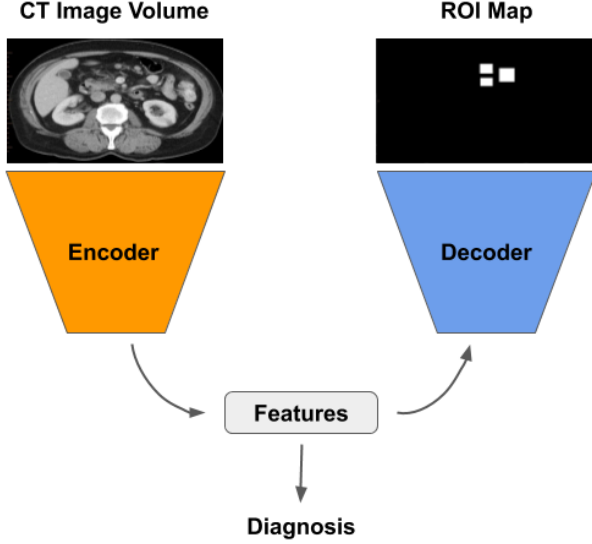
Fig. 1: Architecture of our method. The model receives a scan $\mathbf{v}$ as input and the encoder builds its feature representation $\mathbf{o} = f_{\theta_e}(\mathbf{v})$, which is used by the classification branch $\sigma_{\theta_c}(\mathbf{o})$ and the ROI detection branch $g_{\theta_d}(\mathbf{o})$. The ROI binary map is obtained by thresholding the output from the ROI detection branch.

For the multi-domain multi-task stage, we jointly minimise the classification loss with

$$\theta_c^*, \theta_e^* = \arg\max_{\theta_c, \theta_e} \mathbb{E}_{P^{(1)}(\mathbf{v}, y)}[\ell_C(\sigma_{\theta_c}(f_{\theta_e}(\mathbf{v})), y)], \quad (1)$$

where ℓ_C denotes the binary cross entropy loss, and the ROI detection loss with

$$\theta_d^*, \theta_e^* = \arg\max_{\theta_d, \theta_e} \mathbb{E}_{P^{(2)}(\mathbf{v}, \mathbf{s})}[\ell_D(g_{\theta_d}(f_{\theta_e}(\mathbf{v})), \mathbf{s})\}], \quad (2)$$

where ℓ_D denotes a loss that sums the voxel-wise cross entropy loss and the Dice loss [17]. We only use the subsets $\mathcal{T}_{train}^{(1)}$, $\mathcal{T}_{train}^{(2)}$, $\mathcal{T}_{val}^{(1)}$, $\mathcal{T}_{val}^{(2)}$ from $\mathcal{D}^{(1)}$ and $\mathcal{D}^{(2)}$ during the training process.

The inference consists of feeding the model with an input scan $\mathbf{v}$ that is encoded to obtain the feature embedding and forwarded through the classification branch $\sigma_{\theta_c^*}(f_{\theta_e^*}(\mathbf{v}))$ to yield the probability of metastatic lymph nodes.

4. EXPERIMENTS AND RESULTS

We assess our proposed method on the problem of diagnosing the presence of metastatic lymph nodes from abdominal CT scans. For the weakly labelled data set $\mathcal{D}^{(1)}$ we use a private data set from the Royal Adelaide Hospital (RAH data set) that contains 123 scans from 123 patients. Weak labels indicating the presence of any metastatic lymph node are obtained from pathology reports and clinical notes. There are 57 scans labelled with the presence of metastatic lymph nodes and 66 scans labelled with the absence of

metastatic lymph nodes. Note that no lymph node localization is provided in this data set. The data set $\mathcal{D}^{(2)}$ that provides lymph node localization information is publicly available [9] (CI data set). There are 595 ROI localisation annotations of lymph nodes from 85 scans of 85 patients. Note that no scan-level label about metastatic lymph node is provided with this data set. The RAH data set is randomly divided into training (90 patients, 43 of them with metastatic lymph nodes and 47 without), validation (10 patients, 4 with metastatic lymph nodes, 6 without), and testing (23 patients, 10 with metastatic lymph nodes, 13 without). The CI data set is randomly divided into training (62 patients with 439 annotated lymph nodes), validation (7 patients with 54 annotated lymph nodes), and testing (16 patients with 102 lymph nodes). We train our method with the training set from RAH and the entire CI data set. We utilize the validation set of the RAH data set to perform model selection and we report classification results on the testing set of the RAH data set. We pre-process each image in both data sets by subtracting the mean and dividing by the standard deviation of the training set of the corresponding data set.

The encoder $f_{\theta_e}(.)$ is represented by a 3-D Densenet [5] consisting of 5 dense blocks. The input volume $\mathbf{v}$ size is $512 \times 512 \times 256$ and the feature vector embedding $\mathbf{o} = f_{\theta_e}(\mathbf{v})$ is of size $16 \times 16 \times 8$. The classifier $\sigma_{\theta_c}(.)$ is composed of two fully connected layers and outputs the probability of presence of metastatic lymph nodes in $\mathbf{v}$. The decoder $g_{\theta_d}(.)$ is also composed of 5 dense blocks, and outputs an ROI map $\tilde{\mathbf{s}}$ of the same size as the input scan and is thresholded at $\zeta = 0.8$ to obtain $\tilde{\mathbf{s}}_\zeta$. The training process runs for 500 epochs and optimizes the parameters $\theta_e, \theta_c, \theta_d$ with Adam optimizer (learning rate of 0.05).

We evaluate our proposed semi-supervised multi-domain multi-task training method with the metastic lymph node classification performance, measured with the area under the ROC curve (AUC). We compare our proposed training strategy against other training procedures: 1) **supervised baseline:** a DenseNet [5] classifier trained on the RAH data set and composed of the encoder and classification branch; 2) **semi-supervised:** the same DenseNet classifier from (1), trained with the RAH training set and including the CI data set trained with propagated classification labels; and 3) **supervised multi-domain multi-task:** the proposed architecture trained with the RAH and CI data sets, where the encoder and classifier are trained with the RAH data set and the encoder and detector branches are trained with CI data set. We present quantitative classification results in terms of AUC on Table I.

5. DISCUSSION AND CONCLUSION

The experimental results presented in Table I show that our proposed training method outperforms several baseline training strategies for metastatic lymph node diagnosis from abdominal CT scans. As explained in Sec. 1, the baseline

Training Method for the classifier	Data Sets	AUC
Supervised Baseline	RAH	0.81
Semi-Supervised	RAH + CI	0.84
Supervised Multi-domain Multi-Task	RAH + CI	0.82
Semi-Supervised Multi-domain Multi-Task	RAH + CI	**0.86**

Table I: Classification AUC on the RAH Test Set obtained by the classifier trained with our proposed training strategy and baseline methods.
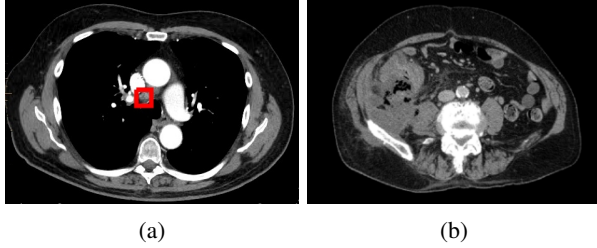

(a)   (b)

Fig. 2: Visual results produced by our proposed method on the RAH data set. Image 2a shows the positive classification of an scan containing a metastatic lymph node (marked in red). Image 2b contains the negative classification of an image with non-metastatic lymph nodes.

classifier achieves the lowest AUC score, which can be in part due to the lack of lymph node localization. Interestingly, propagating the classification labels from RAH to CI data set (without including any localization in the training process) yielded better results for classification than the supervised multi-domain multi-task baseline. We believe this is due to difficulty of integrating multiple tasks from different domains into the training process without any semi-supervision. Finally, our proposed semi-supervised multi-domain multi-task training outperformed all baseline methods. As we hypothesized in Sec. 1, jointly semi-supervising each of the labelled tasks from a different domain results in a more accurate model, probably due to the extra-supervision that can facilitate the addition of data from a different domain.

In conclusion, we proposed a new semi-supervised multi-domain multi-task training, where we semi-supervised the detection task of the data set containing global annotations and the classification task on the data set containing ROI annotations. Results on diagnosing the presence of metastatic lymph nodes from CT scans showed that our method successfully incorporates lymph node localization information from a different domain to improve classification results in the original domain. We leave for future work the evaluation of the lymph node ROI localization task in each data set.

6. REFERENCES

[1] Freddie Bray, Jacques Ferlay, and et al., "Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a cancer journal for clinicians*, 2018.

[2] Laura M Fernandez, Albert Parlade, and et al., "How reliable is ct scan in staging right colon cancer?," *Diseases of the Colon & Rectum*, 2019.

[3] Hyungsik Park, Seung Hyun Cho, and et al., "Potential image-based criteria of neoadjuvant chemotherapy for colon cancer: multireaders diagnostic performance," *Abdominal Radiology*, 2019.

[4] Audrey H Choi, Rebecca A Nelson, and et al., "Accuracy of computed tomography in nodal staging of colon cancer patients," *WJGS*, 2015.

[5] Gao Huang, Zhuang Liu, and et al., "Densely connected convolutional networks," in *CVPR*, 2017.

[6] Yun Lu, Qiyue Yu, and et al., "Identification of metastatic lymph nodes in mr imaging with faster region-based convolutional neural networks," *Cancer research*, 2018.

[7] Zhe Li, Chong Wang, and et al., "Thoracic disease identification and localization with limited supervision," in *CVPR*, 2018.

[8] Gerard Snaauw, Dong Gong, and et al., "End-to-end diagnosis and segmentation learning from cardiac magnetic resonance imaging," in *ISBI*, 2019.

[9] Holger R Roth, Le Lu, and et al., "A new 2.5 d representation for lymph node detection using random sets of deep convolutional neural network observations," in *MICCAI*, 2014.

[10] YQ Huang, CH Liang, and et al., "Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer.," *Journal of clinical oncology*, 2016.

[11] Damien Fourure, Rémi Emonet, and et al., "Multi-task, multi-domain learning: application to semantic segmentation and pose regression," *Neurocomputing*, 2017.

[12] Jiahui Li, Shuang Yang, and et al., "Signet ring cell detection with a semi-supervised learning framework," in *IPMI*, 2019.

[13] Jiamin Liu, Jianhua Yao, and et al., "A semi-supervised cnn learning method with pseudo-class labels for atherosclerotic vascular calcification detection," in *ISBI*, 2019.

[14] Abdullah-Al-Zubaer Imran and Demetri Terzopoulos, "Semi-supervised multi-task learning with chest x-ray images," in *MICCAIw - MLMI*, 2019.

[15] Shuai Chen, Gerda Bortsova, and et al., "Multi-task attention-based semi-supervised learning for medical image segmentation," in *MICCAI*, 2019.

[16] Xiaojin Zhu and Zoubin Ghahramani, "Learning from labeled and unlabeled data with label propagation," .

[17] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *3DV*, 2016.