2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)
Venice, Italy, April 8-11, 2019

TOWARDS CONTINUOUS DOMAIN ADAPTATION FOR MEDICAL IMAGING

Rahul Venkataramani *Hariharan Ravishankar* *Saihareesh Anamandra*

GE Global Research, Bangalore

ABSTRACT

Deep learning algorithms have demonstrated tremendous success on challenging medical imaging problems. However, post-deployment, these algorithms are susceptible to data distribution variations owing to *limited data issues* and *diversity* in medical images. In this paper, we propose *ContextNets*, a generic memory-augmented neural network framework for semantic segmentation to achieve continuous domain adaptation without the necessity of retraining. Unlike existing methods which require access to entire source and target domain images, our algorithm can adapt to a target domain with few similar images. We condition the inference on any new input with features computed on its support set of images (and masks, if available) through contextual embeddings to achieve site-specific adaptation. We demonstrate state-of-the-art domain adaptation performance on X-ray lung segmentation problem from three independent cohorts that differ in disease type, gender, contrast and intensity variations.

Index Terms— domain adaptation, continual learning, memory augmented neural networks, segmentation

1. INTRODUCTION

For successful translation of impressive lab results to acceptable healthcare products, deep learning algorithms have to overcome problems of domain adaptation [1] [2] and customization. Data domain shift is routinely encountered in clinical settings due to variations in equipment, protocol, demography and pathological conditions. Additionally, algorithms are expected to be customizable for a specific site/doctor to account for local preferences/guidelines.

A potential design solution to overcome above problems is to enable lifelong learning from incremental corrective annotations, post-deployment. The naive approach of fine-tuning the model on new domain data is infeasible without a large number of annotated samples. Further, this leads to *catastrophic forgetting*[3] - phenomenon of forgetting past knowledge when retrained on new data. To circumvent obtaining costly annotations, unsupervised domain adaptation approaches transform the image from the target domain to match images from the source domain, thereby resisting performance degradation. Early attempts included image preprocessing steps like contrast normalization, filtering etc. In [1], authors employed an adverserial loss to ensure alignment of feature embeddings from the source & target domains.

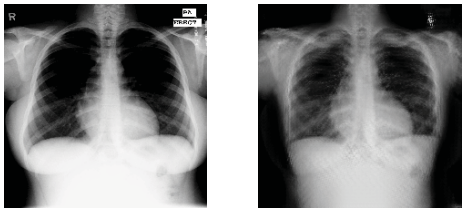
Recently, [4] used a variant of generative adverserial network (GAN), a CycleGAN to style transfer an image from target domain to source domain. However, there are practical challenges: 1) Training a GAN at each deployment site for target to source mapping. 2) Access to source domain images/masks at the deployment site. 3) Availability of a large target domain cohort. The assumption of mapping any new data to the source manifold, may not hold true while encountering rare events or unseen target distributions. From our experiments on a chest X-ray dataset, we observed loss of structural fidelity in cases with diseased lungs/ female scans as depicted in 1, leading to incorrect segmentation.

Memory augmented neural networks (MANN) have recently emerged as a popular framework for classification problems. By augmenting fuzzy matching abilities of neural networks with an external memory, MANNs demonstrate ability to learn from few samples, remember rare events [5] etc. Briefly, for every new sample, features are retrieved by querying stored memory and are combined with the primary neural network for final prediction. However, we are not aware of any attempt to use MANNs for continuous domain adaptation or for semantic segmentation.

In this work, we propose *ContextNets* - a lifelong learning methodology using MANNs for continuous domain adaptation of any deep learnt semantic segmentation algorithm. Compared to other methods, our approach: 1) Does not require any training post-deployment. 2) Does not require access to source domain data. 3) Is a single framework to perform domain adaptation in unsupervised/supervised variants.

Fig. 2 depicts our framework in action. For a test im-

Fig. 1: Image Transformation Using GANs



(a) Target domain image (b) Transformed image

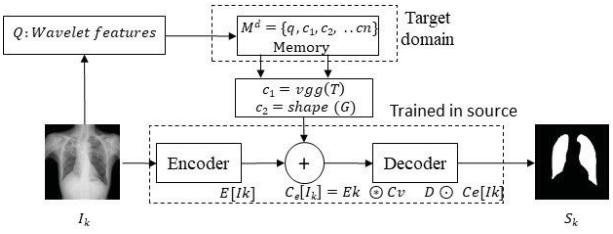


Fig. 2: *ContextNets* for continuous domain adaptation

age from target domain, we fetch similar images (Context set) and their descriptors (Context vectors) and embed it to model trained only from source domain. Section 2 and 3 describes our approach in detail. In Section 4, we demonstrate impressive domain adaptation performance of our algorithm on 3 independent cohorts for lung segmentation on chest X-rays.

2. METHODOLOGY

In this section we present the generic formulation and framework for continual domain adapation, while in Section 3, we present the practical realization of the same.

2.1. Continual Adaptation via Memory

The central part of the proposed approach is a memory module $\mathcal{M}^d$, for every target domain d, which aids in dynamically adapting to the domain distribution variations while observing new cases continually. Each row in $\mathcal{M}^d$ (Eq. 1) corresponds to one sample image observed so far and contains two sets of information - *Image Features* and *Context Features*.

Intuitively, the image features computed on image I_k, are used as keys to retrieve contextually similar images from domain d , while context features $c_1, c_2, ..c_n$ encapsulate the description of various properties of these images. q and c_i can be computed from respective query and context mapping functions as shown in Eq. 2.

$$\mathcal{M}^d = \{q, c_1, c_2, ..c_n\} \tag{1}$$

$$Q : I_k \rightarrow [q], C_i : I_k \rightarrow [c_i] \tag{2}$$

For every test image I_k, this memory is used to return a support set $C(I_k)$ by computing T nearest neighbors on the query features (Eq. 3). Further, we also collect the *context features* corresponding to these support set images (Eq. 4). We refer to support set of images $C(I_k)$ as *Context Set* and context features of the support set $C_v(I_k)$ as *Context Vector*.

$$C(I_k) = \{NN(q(I_k), \mathcal{M}[q])\}_T \tag{3}$$

$$C_v(I_k) = \mathcal{M}[c_1, c_2, ..c_n][C(I_k)] \tag{4}$$

Query functions can be designed based on image similarity, or subject information like gender, demography, age, disease diagnosis, etc. Similarly, context mapping functions can designed to model common properties like texture, contrast, shape properties of masks, etc. In section 3, we describe the design choices made for lung segmentation in X-ray.

2.2. Contextual Embedding

A fully convolutional (FCN) image-to-image network (U-Net [6]) is a mapping $D \circ E$, where the *encoder* E and *decoder* D are parameterized by a set of parameters θ. Given training pairs of images and segmentation masks $\{I_k, S_k\}$, $k = 1, 2, \ldots, N$, one learns the parameters θ by minimizing the training loss $J(\theta) = \sum_{k=1}^{N} \mathcal{L}(S_k, \hat{S}_k(\theta))$, e.g. RMSE loss, binary cross-entropy loss (BCE), where $\hat{S}_k(\theta) = (D \circ E)[I_k]$.

To test our hypothesis that utilising context images from target domain can improve domain adaptation, we propose *contextual embeddings*. We condition the inference on I_k by embedding $C_v(I_k)$ at the output of the encoder $E[I_k]$ as follows: $C_e[I_k] = E[I_k] \circledast C_v[I_k]$, where $\circledast$ is the embedding operator. The decoder works on the embedding output and produces output prediction $\hat{S}'_K = D \circ C_e[I_k]$. Finally, we optimize for loss between $\hat{S}'_k$ and ground truth shape S_k using $J(\theta)$. We denote the learnt model as θ^{cn} as *ContextNets* which maps an input image and its context vector to output predictions - $\theta^{cn} : \{I_k, C_v\} :\rightarrow S_k$.

3. CONTINUAL DOMAIN ADAPATION FOR LUNG SEGMENTATION FROM XRAY IMAGES

To study the effect of domain change on a trained segmentation model, we used 3 independent cohorts of chest X-ray images varying in disease type, intensity patterns & contrast:

1. **Montgomery dataset** [7] The Montgomery TB dataset is an open source dataset consisting of 138 posterior-anterior x-rays, of which 80 x-rays are normal and 58 x-rays are abnormal with manifestations of tuberculosis.
2. **JSRT dataset** [8] : The JSRT database consists of 274 chest X-rays with lung masks collected from typical clinical practice, but differing in image distributions.
3. **Pneumoconiosis dataset** : Pneumoconiosis is an occupational disease characterized by settlements of dust and other metal particles in the lungs. This internal dataset consists of 330 images acquired from a hospital site with high incidence of pneumoconioisis.

We use the Montgomery dataset as the source domain and evaluate the adaptation performance on the other 2 domains.

3.1. Image features

To design the query mapping function (Eq. 2) for retrieving relevant X-ray images, we explored multiple image similarity features from computer vision like wavelet features, histogram of gradients (HoG) among others. Similarly, we explored features extracted from popular pretrained models like

VGGNet, ResNet, etc. We evaluated the clustering similarity qualitatively and observed that the context set $C(I_k)$ derived using discrete Haar wavelet features was satisfactory.

3.2. Context Features

For the X-ray lung segmentation problem, we used two sets of context features - to model texture and shape respectively.

1. **Texture Features** We chose to use features extracted from fully connected layer (fc1) of the popular VGGNet architecture as our texture descriptors, as they have been shown effective on multiple limited data problems. The intuition is that, texture features from context set should help in handling cases with unexpected texture/contrast variations.

2. **Shape Features** Inspired by [9], we built a light-weight shape auto-encoder (SAE), which takes shape masks as input and projects it to the manifold of training shapes. We extract features at the ouput of encoder of SAE, with feature length = 256. Constraining the inference of a test image on the shape properties of the context set, can add robustness and regularization to the predictions. Additionally, on cases like pathology or pneumoconiosis or TB, where lungs might be partially affected, having shape descriptors from masks on images from context set help improve the performance.

3.3. *ContextNets*

We proceed to build *ContextNets* from source domain data using procedure described in Sec. 2.2. We explore both supervised and unsupervised domain adaptation variants as the availability of ground truth masks from target domain may not always be available. Equations 5 and 6 describe the training procedure for learning $ContextNet_1$ and $ContextNet_2$. q, t and g are wavelet features for querying, VGGNet features for texture and SAE features for shape, respectively. $ContextNet_1$ utilises the input image (I_k) along with texture features of the context set ($T_v(C(I_k))$) from source domain memory and learns θ_1^{cn} to determine segmentation masks S_k. $ContextNet_2$ additionally utilises shape features ($G_v(C(I_k))$) to learn θ_2^{cn}. Memory module $\mathcal{M}^s$ appropriately contains only the texture features for $ContextNet_1$ and both texture and shape features for $ContextNet_2$ -which corresponds to unsupervised & supervised settings respectively.

$$\theta_1^{cn} : \{I_k, T_v(C(I_k)\} \rightarrow S_k,\ \mathcal{M}^s = \{q, t\} \quad (5)$$

$$\theta_2^{cn} : \{I_k, T_v(C(I_k), G_v(C(I_k)\} \rightarrow S_k,\ \mathcal{M}^s = \{q, t, g\} \quad (6)$$

Finally, for every test image I_k^d from a target domain d, we fetch context set and context vectors from target domain memory $\mathcal{M}^d$ and infer the segmentation masks using θ_1^{cn} or θ_2^{cn} as shown in Fig. 2.

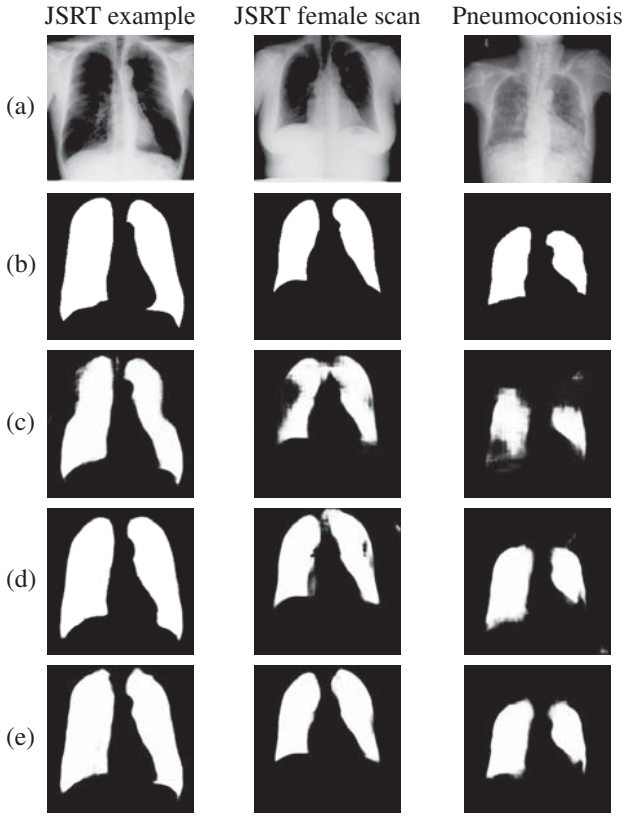


Fig. 3: Visualization of *ContextNet* results and comparisons on three exemplar target images. Row (b) displays ground-truth masks. Row(c) depicts UNet results. Rows (d) and (e) correspond to $ContextNet_1$ and $ContextNet_2$ respectively

4. EXPERIMENTS AND RESULTS

Our FCN architecture contains 8 convolutional layers shared equally between the enconder and decoder. We experimented with different combinations like concatenation, average and sum for our context embedding function. The SAE was another fully-convolutional neural network with very few parameters $\sim$ 10k. We also varied the size of the context from 3 to 5 and obtained similar results. In rest of the paper, we will share results for context size of 5 and average of context vectors as our embedding. We optimized for BCE loss and all the models were trained for 100 epochs with batch_size = 5. Further, we perform standard preprocessing steps like histogram normalization to ensure reasonable baseline comparison, without which the baseline results were very poor.

We use Dice coefficient betwen predicted and ground truth masks as the performance evaluation metric. We benchmark $ContextNet_{1,2}$ with following approaches:

- *NoDA*: We build U-Net on the source domain data, and test on target domains without any adaptation. This scenario establishes the lower baseline for comparisons.

Category	Algorithm	JSRT	Pneumoconiosis
Baseline	*NoDA*	0.896 ± 0.075	0.882 ± 0.186
	TransferLearnt	0.97 ± 0.075	0.97 ± 0.186
SoTA	*SeUDA*	0.945	$-$
Proposed Approach	$ContextNet_1$	0.958 ± 0.033	0.933 ± 0.024
	$ContextNet_2$	$\mathbf{0.965 \pm 0.013}$	$\mathbf{0.949 \pm 0.017}$

Table 1: Quantitative Evaluation of *ContextNets*

- *TransferLearnt*: If the entire set of target domain images and masks are available, the obvious strategy is to retrain/finetune the model from source domain. This scenario represents the gold-standard performance. However, this impractical scenario is included, only to establish the upper limit of performance.
- *SeUDA*: We compare our algorithm with the state-of-the-art results for unsupervised domain adaptation, a task-aware generative adverserial network[10]. We however note that these methods require access to source domain dataset during deployment setting. We report their results on JSRT from [10] after averaging the right and left lung performance.

From Table 1, it is apparent that *NoDA* suffers immensely when subjected to domain changes. On JSRT dataset, both the unsupervised and supervised versions of *ContextNet* show remarkable improvements of 6% and 7%, respectively over *NoDA*. Interestingly, our algorithm $ContextNet_2$ even outperforms state-of-the art method - *SeUDA* which requires access to the source dataset, along with additional computational burden of building GAN models per target domain.

Table 1 also demonstrates the efficacy of proposed approach on the challenging pneumoconiosis dataset which contains lungs that are partially/fully affected by the disease. With no retraining or requirement of source/target domain data, we demonstrate remarkable performance and almost match the impractical upper baseline - *TransferLearnt*.

Fig 3 presents visualization of results. The columns correspond to 3 exemplar images and rows display results from different methods. These images are chosen in increasing order of domain adapation complexity. The first image suffers from texture variations and second image depicts structural variations due to female breasts. The final column shows an extreme case of domain adaptation challenge due to manifestation of stage-3 pneumoconiosis. In all the cases, UNet's performance degrades significantly. Both $ContextNet_1$ and $ContextNet_2$ perform similarly on the first case, as the challenge is only due to texture varations. However, on cases 2 and 3, effect of modeling shape plus texture is apparent, with $ContextNet_2$ producing smoother and accurate shape masks compared to $ContextNet_1$. Finally, by limiting the target memory size to just 50, we tested the robustness of the methods, and obtained similar performance of ~ 0.96 for $ContextNet_2$ on JSRT dataset.

5. CONCLUSION

In this paper, we propose a practical method for domain adaptation for image segmentation. We emulate expert clinicians' process of drawing experience from similar cases through the use of an external memory. Our approach achieves continuous domain adaptation with extremely few annotated samples from the target domain. We also alleviate privacy issues by eliminating the need to store source domain data. Further, our method is a lifelong learning paradigm suited to dynamic changes in data distributions. Future research would be focussed on automatic design of various components of the system and a memory management module.

6. REFERENCES

[1] Konstantinos Kamnitsas et al., "Unsupervised domain adaptation in brain lesion segmentation with adversarial networks," in *IPMI*. Springer, 2017, pp. 597–609.

[2] Mohsen Ghafoorian et al., "Transfer learning for domain adaptation in mri: Application in brain lesion segmentation," in *MICCAI*. Springer, 2017, pp. 516–524.

[3] James Kirkpatrick et al., "Overcoming catastrophic forgetting in neural networks," *Proceedings of the National Academy of Sciences*, vol. 114, pp. 3521–3526, 2017.

[4] Yue Zhang, Shun Miao, Tommaso Mansi, and Rui Liao, "Task driven generative modeling for unsupervised domain adaptation: Application to x-ray image segmentation," *arXiv preprint arXiv:1806.07201*, 2018.

[5] Lukasz Kaiser, Ofir Nachum, Aurko Roy, and Samy Bengio, "Learning to remember rare events," *arXiv preprint arXiv:1703.03129*, 2017.

[6] Olaf Ronneberger et al., "U-net: Convolutional networks for biomedical image segmentation," in *MICCAI*. Springer, 2015, pp. 234–241.

[7] Stefan Jaeger et al., "Two public chest x-ray datasets for computer-aided screening of pulmonary diseases," *Quantitative imaging in medicine and surgery*, 2014.

[8] Junji Shiraishi et al., "Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists' detection of pulmonary nodules," *American Journal of Roentgenology*, vol. 174, pp. 71–74, 2000.

[9] Hariharan Ravishankar et al., "Learning and incorporating shape models for semantic segmentation," in *MICCAI*. Springer, 2017, pp. 203–211.

[10] Cheng Chen et al., "Semantic-aware generative adversarial nets for unsupervised domain adaptation in chest x-ray segmentation," *arXiv:1806.00600*, 2018.