

# DASA: Domain Adaptation in Stacked Autoencoders using Systematic Dropout

Abhijit Guha Roy and Debdoot Sheet

Department of Electrical Engineering, Indian Institute of Technology Kharagpur, India

abhi4ssj@gmail.com, debdoot@ee.iitkgp.ernet.in

## Abstract

Domain adaptation deals with adapting behaviour of machine learning based systems trained using samples in source domain to their deployment in target domain where the statistics of samples in both domains are dissimilar. The task of directly training or adapting a learner in the target domain is challenged by lack of abundant labeled samples. In this paper we propose a technique for domain adaptation in stacked autoencoder (SAE) based deep neural networks (DNN) performed in two stages: (i) unsupervised weight adaptation using systematic dropouts in mini-batch training, (ii) supervised fine-tuning with limited number of labeled samples in target domain. We experimentally evaluate performance in the problem of retinal vessel segmentation where the SAE-DNN is trained using large number of labeled samples in the source domain (DRIVE dataset) and adapted using less number of labeled samples in target domain (STARE dataset). The performance of SAE-DNN measured using logloss in source domain is 0.19, without and with adaptation are 0.40 and 0.18, and 0.39 when trained exclusively with limited samples in target domain. The area under ROC curve is observed respectively as 0.90, 0.86, 0.92 and 0.87. The high efficiency of vessel segmentation with DASA strongly substantiates our claim.

## 1. Introduction

The under-performance of learning based systems during deployment stage can be attributed to dissimilarity in distribution of samples between the *source domain* on which the system is initially trained and the *target domain* on which it is deployed. Transfer learning is an active field of research which deals with transfer of knowledge between the *source* and *target domains* for addressing this challenge and enhancing performance of learning based systems [6], when it is challenging to train a system exclusively in the *target domain* due to unavailability of sufficient labeled samples. While domain adaptation (DA) have been primarily developed for simple reasoning and shallow network architectures, there exist few techniques for adapting deep networks

with complex reasoning [4]. In this paper we propose a systematic dropout based technique for adapting a stacked autoencoder (SAE) based deep neural network (DNN) [2] for the purpose of vessel segmentation in retinal images [1]. Here the SAE-DNN is initially trained using ample number of samples in the *source domain* (DRIVE dataset<sup>1</sup>) to evaluate efficacy of DA during deployment in the *target domain* (STARE dataset<sup>2</sup>) where an insufficient number of labeled samples are available for reliable training exclusively in the *target domain*.

**Related Work:** Autoencoder (AE) is a type of neural network which learns compressed representations inherent in the training samples without labels. Stacked AE (SAE) is created by hierarchically connecting hidden layers to learn hierarchical embedding in compressed representations. An SAE-DNN consists of encoding layers of an SAE followed by a target prediction layer for the purpose of regression or classification. With increase in demand for DA in SAE-DNNs different techniques have been proposed including marginalized training [3], via graph regularization [7] and structured dropouts [10], across applications including recognizing speech emotion [4] to fonts [9].

**Challenge:** The challenge of DA is to retain nodes common across *source* and *target domains*, while adapting the domain specific nodes using fewer number of labeled samples. Earlier methods [3, 7, 10] are primarily challenged by their inability to re-tune nodes specific to the *source domain* to nodes specific for *target domain* for achieving desired performance, while they are able to only retain nodes or a thinned network which encode domain invariant hierarchical embeddings.

**Approach:** Here we propose a method for DA in SAE (DASA) using systematic dropout. The two stage method adapts a SAE-DNN trained in the *source domain* following (i) unsupervised weight adaptation using systematic dropouts in mini-batch training with abundant unlabeled samples in *target domain*, and (ii) supervised fine-tuning with limited number of labeled samples in *target domain*. The systematic dropout per mini-batch is introduced only

<sup>1</sup><http://www.isi.uu.nl/Research/Databases/DRIVE>

<sup>2</sup><http://www.ces.clemson.edu/~ahoover/stare>

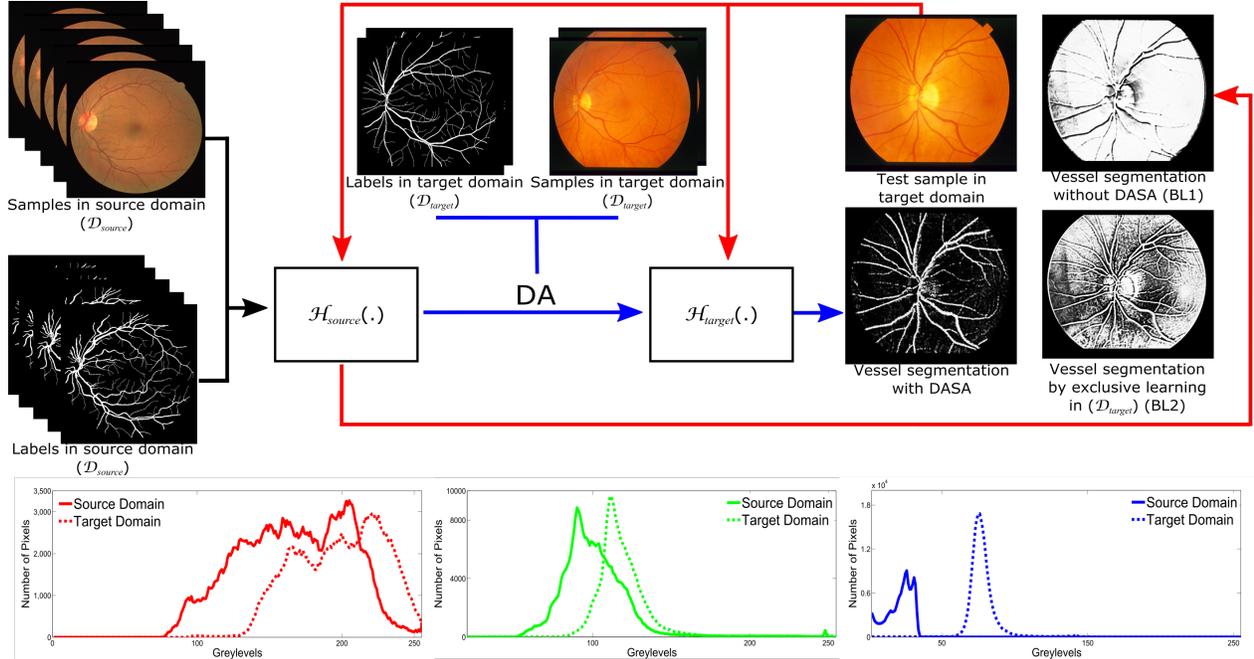


Figure 1. Overview of the process of DASA. It starts with learning a SAE-DNN model  $\mathcal{H}_{source}$  using ample labeled samples in  $\mathcal{D}_{source}$ . Limited number of labeled samples in  $\mathcal{D}_{target}$  are used to transform  $\mathcal{H}_{source} \xrightarrow{DA} \mathcal{H}_{target}$ . Results of vessel segmentation with domain adaptation are compared with (BL1) SAE-DNN trained in  $\mathcal{D}_{source}$  and deployed in  $\mathcal{D}_{target}$  without DASA and (BL2) SAE-DNN trained in  $\mathcal{D}_{target}$ . The shifts in distribution of color statistics across samples in  $\mathcal{D}_{source}$  and  $\mathcal{D}_{target}$  are also illustrated.

in the representation encoding (hidden) layers and is guided by a saliency map defined by response of the neurons in the mini-batch under consideration. Error backpropagation and weight updates are however across all nodes and not only restricted to the post dropout activated nodes, contrary to classical randomized dropout approaches [8]. Thus having different dropout nodes across different mini-batches and weight updates across all nodes in the network, ascertains refinement of domain specific hierarchical embeddings while preserving domain invariant ones.

The problem statement is formally introduced in Sec. 2. The methodology is explained in Sec. 3. The experiments are detailed in Sec. 4, results are presented and discussed in Sec. 5 with conclusion in Sec. 6.

## 2. Problem Statement

Let us consider a retinal image represented in the RGB color space as  $\mathcal{I}$ , such that the pixel location  $\mathbf{x} \in \mathcal{I}$  has the color vector  $\mathbf{c}(\mathbf{x}) = \{r(\mathbf{x}), g(\mathbf{x}), b(\mathbf{x})\}$ .  $N(\mathbf{x})$  is a neighborhood of pixels centered at  $\mathbf{x}$ . The task of retinal vessel segmentation can be formally defined as assigning a class label  $y \in \{\text{vessel}, \text{background}\}$  using a hypothesis model  $\mathcal{H}(\mathcal{I}, \mathbf{x}, N(\mathbf{x}); \{\mathcal{I}\}_{train})$ . When the statistics of samples in  $\mathcal{I}$  is significantly dissimilar from  $\mathcal{I}_{train}$ , the performance of  $\mathcal{H}(\cdot)$  is severely affected. Generally  $\{\mathcal{I}\}_{train}$  is referred to

as the *source domain* and  $\mathcal{I}$  or the set of samples used during deployment belong to the *target domain*. The hypothesis  $\mathcal{H}(\cdot)$  which optimally defines *source* and *target domains* are also referred to as  $\mathcal{H}_{source}$  and  $\mathcal{H}_{target}$ . DA is formally defined as a transformation  $\mathcal{H}_{source} \xrightarrow{DA} \mathcal{H}_{target}$  as detailed in Fig. 1.

## 3. Exposition to the Solution

Let us consider the source domain as  $\mathcal{D}_{source}$  with abundant labeled samples to train an SAE-DNN ( $\mathcal{H}_{source}$ ) for the task of retinal vessel segmentation, and a target domain  $\mathcal{D}_{target}$  with limited number of labeled samples and ample unlabeled samples, insufficient to learn  $\mathcal{H}_{target}$  reliably as illustrated in Fig. 1.  $\mathcal{D}_{source}$  and  $\mathcal{D}_{target}$  are closely related, but exhibiting distribution shifts between samples of the *source* and *target domains*, thus resulting in under-performance of  $\mathcal{H}_{source}$  in  $\mathcal{D}_{target}$  as also illustrated in Fig. 1. The technique of generating  $\mathcal{H}_{source}$  using  $\mathcal{D}_{source}$ , and subsequently adapting  $\mathcal{H}_{source}$  to  $\mathcal{H}_{target}$  via systematic dropout using  $\mathcal{D}_{target}$  is explained in the following sections.

### 3.1. SAE-DNN learning in the source domain

AE is a single layer neural network that encodes the cardinal representations of a pattern  $\mathbf{p} = \{p_k\}$  onto a trans-



## 4. Experiments

**SAE-DNN architecture:** We have a two-layered architecture with  $L = 2$  where  $AE_1$  consists of 400 nodes and  $AE_2$  consists of 100 nodes. The number of nodes at input is  $15 \times 15 \times 3$  corresponding to the input with patch size of  $15 \times 15$  in the color retinal images in RGB space. AEs are unsupervised pre-trained with learning rate of 0.3, over 50 epochs,  $\beta = 0.1$  and  $\rho = 0.04$ . Supervised weight refinement of the SAE-DNN is performed with a learning rate of 0.1 over 200 epochs. The training configuration of learning rate and epochs were same in the *source* and *target* domains, with  $\tau = 0.1$ .

**Source and target domains:** The SAE-DNN is trained in  $\mathcal{D}_{source}$  using 4% of the available patches from the 20 images in the training set in DRIVE dataset. DA is performed in  $\mathcal{D}_{target}$  using (i) 4% of the available patches in 10 unlabeled images for unsupervised adaptation using systematic dropout and (ii) 4% of the available patches in 3 labeled images for fine tuning.

**Baselines and comparison:** We have experimented with the following SAE-DNN baseline (BL) configurations and training mechanisms for comparatively evaluating efficacy of DA: **BL1:** SAE-DNN trained in *source domain* and deployed in *target domain* without DA; **BL2:** SAE-DNN trained in *target domain* with limited samples and deployed in *target domain*.

## 5. Results and Discussion

The results comparing performance of the SAE-DNN are reported in terms of *logloss* and area under ROC curve as presented in Table 1, and DA aspects in Fig. 3.

	<i>logloss</i>	Area under ROC
Source domain	$0.19 \pm 0.05$	$0.90 \pm 0.02$
BL1	$0.40 \pm 0.31$	$0.86 \pm 0.03$
BL2	$0.39 \pm 0.68$	$0.87 \pm 0.01$
<b>DASA</b>	$0.18 \pm 0.02$	$0.92 \pm 0.02$

Table 1. Comparison of Performance with the baselines

**Hierarchical embedding in representations learned across domains:** AEs are typically characteristic of learning hierarchical embedded representations. The first level of embedding represented in terms of  $w_1$  in Fig. 3(g) is over-complete in nature, exhibiting substantial similarity between multiple sets of weights which promotes sparsity in the nature of  $w_2$  in Fig. 3(h). Some of these weight kernels are domain invariant, and as such remain preserved after DA as observed for  $w_1$  in Fig. 3(i) and for  $w_2$  in Fig. 3(j). Some of the kernels which are domain specific, exhibit significant dissimilarity in  $w_1$  and  $w_2$  between *source domain* in Figs. 3(g) and 3(h) vs. *target domain* in Figs. 3(i) and 3(j). These are on account of dissimilarity of sample statistics in

the domains as illustrated earlier in Fig. 1 and substantiates DASA of being able to retain nodes common across *source* and *target domains*, while re-tuning domain specific nodes.

**Accelerated learning with DA:** The advantage with DA is the ability to transfer knowledge from *source domain* to learn with fewer number of labeled samples and ample number of unlabeled samples available in the *target domain* when directly learning in the *target domain* does not yield desired performance. Figs. 3(k) and 3(l) compare the learning of  $w_1$  and  $w_2$  using ample unlabeled data in *source* and *target domain* exclusively vs. DA. Fig. 3(m) presents the acceleration of learning with DA in *target domain* vs. learning exclusively with insufficient number of labeled samples.

**Importance of transfer coefficient:** The transfer coefficient  $\tau$  drives quantum of knowledge transfer from the *source* to *target domains* by deciding on the amount of nodes to be dropped while adapting with ample unlabeled samples. This makes it a critical parameter to be set in DASA to avoid over-fitting and negative transfers as illustrated in Table. 2 where optimal  $\tau = 0.1$ . Generally  $\tau \in [0, 1]$  with  $\tau \rightarrow 0$  being associated with large margin transfer between domains when they are not very dissimilar, and  $\tau \rightarrow 1$  being associated otherwise.

$\tau$	0	0.05	0.1	0.15	0.2
<i>logloss</i>	0.39	0.24	0.18	0.21	0.32

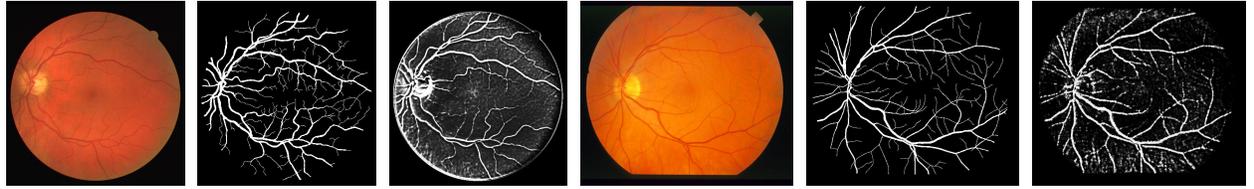
Table 2. Variation of *logloss* in DA with variation of  $\tau$

## 6. Conclusion

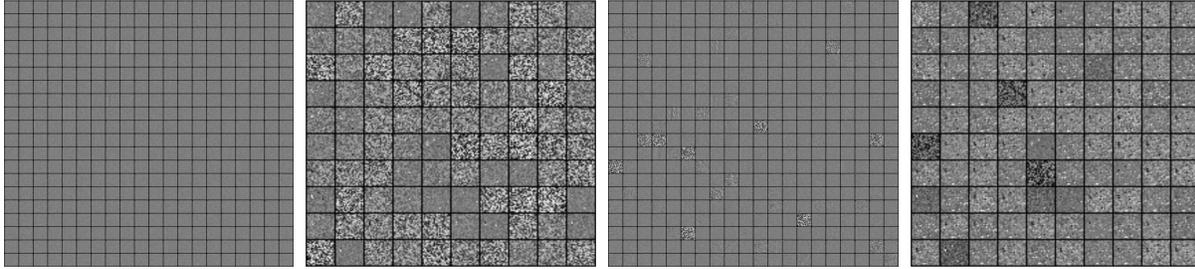
We have presented DASA, a method for knowledge transfer in an SAE-DNN trained with ample labeled samples in *source domain* for application in *target domain* with less number of labeled samples insufficient to directly train to solve the task in hand. DASA is based on systematic dropout for adaptation being able to utilize (i) ample unlabeled samples and (ii) limited amount of labeled samples in *target domain*. We experimentally provide its efficacy to solve the problem of vessel segmentation when trained with DRIVE dataset (source domain) and adapted to deploy on STARE dataset (target domain). It is observed that DASA outperforms the different baselines and also exhibits accelerated learning due to knowledge transfer. While systematic dropout is demonstrated on an SAE-DNN in DASA, it can be extended to other deep architectures as well.

## Acknowledgement

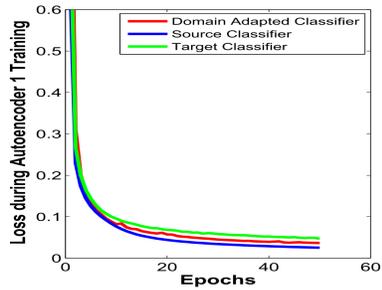
We acknowledge NVIDIA for partially supporting this work through GPU Education Center at IIT Kharagpur.



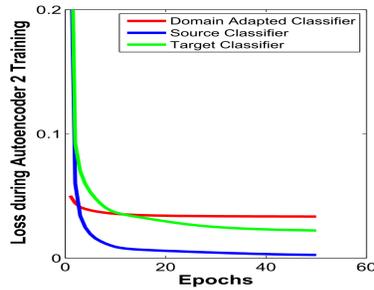
(a) Source domain sample. (b) Source domain labels. (c) Source domain prediction. (d) Target domain sample. (e) Target domain labels. (f) Target domain prediction with DA.



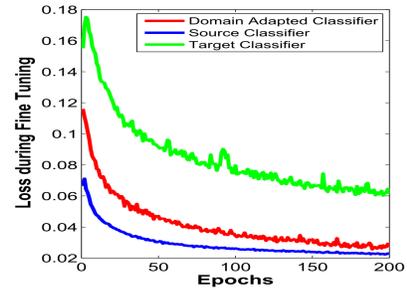
(g) Over complete representation in  $w_1$  in source domain. (h) Sparse representation in  $w_2$  in source domain. (i) DA representation in  $w_1$ . (j) DA representation in  $w_2$ .



(k)  $J(\mathbf{W})$  vs. epochs in training AE1



(l)  $J(\mathbf{W})$  vs. epochs in training AE2



(m)  $J(\mathbf{W})$  vs. epochs in training SAE-DNN

Figure 3. Performance of the vessel segmentation with (a-c) SAE-DNN on sample 01 (Test) in  $\mathcal{D}_{source}$  (DRIVE), (d-f) DASA on sample 0163 in  $\mathcal{D}_{source}$  (STARE), (g, h) representation learned by the SAE-DNN in  $\mathcal{D}_{source}$ , (i, j) DA representations using  $\mathcal{D}_{target}$ , learning dynamics vs. epochs in (k, l) AEs and (m) SAE-DNN indicating the higher efficacy of DASA compared to direct learning with limited samples in  $\mathcal{D}_{target}$ .

## References

- [1] M. D. Abràmoff, M. K. Garvin, and M. Sonka. Retinal imaging and image analysis. *IEEE Rev. Biomed. Engg.*, 3:169–208, 2010. 1
- [2] Y. Bengio. Learning deep architectures for AI. *Found., Trends, Mach. Learn.*, 2(1):1–127, 2009. 1
- [3] M. Chen, Z. Xu, K. Weinberger, and F. Sha. Marginalized denoising autoencoders for domain adaptation. In *Proc. Int. Conf. Mach. Learn.*, pages 767–774, 2012. 1
- [4] J. Deng, Z. Zhang, F. Eyben, and B. Schuller. Autoencoder-based unsupervised domain adaptation for speech emotion recognition. *IEEE Signal Process. Lett.*, 21(9):1068–1072, 2014. 1
- [5] S. Haykin. *Neural Networks and Learning Machines*. Pearson Education, 2011. 3
- [6] S. J. Pan and Q. Yang. A survey on transfer learning. *IEEE Trans. Knowledge., Data Engg.*, 22(10):1345–1359, 2010. 1
- [7] Y. Peng, S. Wang, and B.-L. Lu. Marginalized denoising autoencoder via graph regularization for domain adaptation. In *Proc. Neural Inf. Process. Sys.*, pages 156–163, 2013. 1
- [8] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.*, 15(1):1929–1958, 2014. 2, 3
- [9] Z. Wang, J. Yang, H. Jin, E. Shechtman, A. Agarwala, J. Brandt, and T. S. Huang. Real-world font recognition using deep network and domain adaptation. In *Proc. Int. Conf. Learning Representations*, page arXiv:1504.00028, 2015. 1
- [10] Y. Yang and J. Eisenstein. Fast easy unsupervised domain adaptation with marginalized structured dropout. *Proc. Assoc., Comput. Linguistics*, 2014. 1