

# Unsupervised Domain Adaptation for Automatic Estimation of Cardiothoracic Ratio

Nanqing Dong<sup>1,2</sup>, Michael Kampffmeyer<sup>3</sup>, Xiaodan Liang<sup>4</sup>, Zeya Wang<sup>1</sup>,  
Wei Dai<sup>1</sup>, and Eric Xing<sup>1</sup>

<sup>1</sup> Petuum, Inc., <sup>2</sup> Cornell University, <sup>3</sup> UiT The Arctic University of Norway,  
<sup>4</sup> Carnegie Mellon University

**Abstract.** The cardiothoracic ratio (CTR), a clinical metric of heart size in chest X-rays (CXRs), is a key indicator of cardiomegaly. Manual measurement of CTR is time-consuming and can be affected by human subjectivity, making it desirable to design computer-aided systems that assist clinicians in the diagnosis process. Automatic CTR estimation through chest organ segmentation, however, requires large amounts of pixel-level annotated data, which is often unavailable. To alleviate this problem, we propose an unsupervised domain adaptation framework based on adversarial networks. The framework learns domain invariant feature representations from openly available data sources to produce accurate chest organ segmentation for unlabeled datasets. Specifically, we propose a model that enforces our intuition that prediction masks should be domain independent. Hence, we introduce a discriminator that distinguishes segmentation predictions from ground truth masks. We evaluate our system’s prediction based on the assessment of radiologists and demonstrate the clinical practicability for the diagnosis of cardiomegaly. We finally illustrate on the JSRT dataset that the semi-supervised performance of our model is also very promising.

## 1 Introduction

Cardiomegaly, also referred to as heart enlargement, is ranked as the most frequent disease code among a public collection of radiology reports from the National Library of Medicine (NLM) according to a National Institutes of Health (NIH) study on medical information retrieval [4]. Cardiomegaly can result from other diseases or medical conditions, such as coronary artery disease and hypertension. It is suggested that cardiomegaly is associated with a high risk of sudden cardiac death [13]. The prevention of cardiomegaly starts from early detection and CTR measured from posterior-anterior (PA) CXR is an important indicator for cardiomegaly [5]. CTR is calculated as the ratio of maximal horizontal cardiac diameter to maximal horizontal thoracic diameter, and CTR greater than 0.5 is commonly considered as cardiomegaly [3, 5]. Manual measurement of CTR requires domain knowledge in radiology and extensive human labor in annotating CXRs, with results being error-prone due to observational error. This motivates the automation of CTR calculation and cardiomegaly detection. One common approach to estimating CTR is lung field segmentation [2].

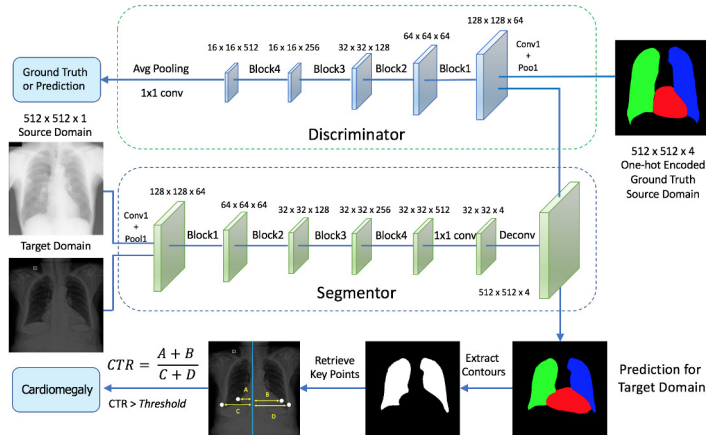


Fig. 1: Illustration of the architecture. In our proposed adversarial training procedure, the segmentor produces segmentations for the input images and the discriminator attempts to distinguish these predictions from ground truth annotations. A post-processing step (bottom part of figure) is used to predict cardiomegaly based on the predicted lung segmentation masks.

Recent advances in Convolutional Neural Networks (CNNs) have brought breakthroughs in the field of semantic segmentation, achieving state-of-the-art performance [1, 9]. Compared to traditional semantic segmentation, the annotated data for medical image segmentation is more difficult to be acquired, because of the limited available data and the tremendous cost of collecting and labeling it. Transfer learning is a common approach to solve tasks with data scarcity, utilizing the fact that CNNs generally learn feature representations that are robust across a variety of tasks [14]. However, as segmentation predictions based on these representations do not generalize very well to different datasets because of the dataset shift phenomena [7], it is commonly required to fine-tune the network based on a set of labels for the target domain. In particular, CXRs from different hospitals are often taken with different imaging protocols and commonly exhibit differences in noise levels, contrast and resolution. So it is impractical to directly use transfer learning techniques. See Figure 1 and Figure 3 for the differences between CXRs obtained at different hospitals.

In this paper, we propose an unsupervised domain adaptation (UDA) framework based on adversarial networks, which allows us to learn domain invariant feature representations from openly available data sources in order to produce accurate chest organ segmentation for unlabeled datasets. Domain adaptation methods aim to reduce the problems of dataset shift, commonly, by aligning the learned source and target representation in a joint embedding space [12, 14]. Adversarial networks have become a popular choice to achieve this alignment, by introducing a discriminator that is trained to distinguish between the source and the target domain and by forcing the model to learn representa-

tions that can fool the discriminator. We propose an alternative training scheme where we utilize a discriminator that enforces our intuition that prediction masks should be domain independent by discriminating segmentation predictions from ground truth masks. We evaluate our system’s performance based on the assessment of radiologists on a CTR estimation dataset. Our approach outperforms the state-of-the-art UDA and shows the clinical practicability for the diagnosis of cardiomegaly. We finally illustrate that our approach can also be used for semi-supervised chest organ segmentation of the JSRT benchmark dataset.

## 2 Methodology

The complete pipeline is shown in Figure 1. The adversarial neural network consists of a discriminator and a segmentor. To demonstrate the generalization and simplicity of the methodology, we use ResNet18 as a backbone architecture [8]. The discriminator is a standard ResNet classifier and the segmentor is inspired by the Fully Convolutional Network (FCN) [9], but uses an output stride of 16, following the example of [1]. Provided the predicted labels for the two lungs, the CTR is calculated in a post-processing step.

### 2.1 Adversarial Training for Supervised Semantic Segmentation

Adversarial learning was first introduced in the Generative Adversarial Network (GAN) [6] as a two-model zero-sum game, in which one model generates candidates for the other network to evaluate. Inspired by [10], who used adversarial learning to improve semantic segmentation results, we let  $S$  be the segmentor and  $D$  be the discriminator.  $S$  is trained to produce realistic prediction masks in order to fool  $D$ , which in turn is attempting to discriminate these predictions from ground truth images in a binary classification.  $D$  is encouraged to learn a complex loss between the higher-order label statistics, which in practice cannot be explicitly formulated. Medical domain knowledge is being implicitly incorporated into this formulation as part of the annotated ground truth data.

An alternative training scheme is applied to train the segmentor and discriminator. Given  $D$ , the loss to be minimized for  $S$  is a multi-class cross-entropy loss for semantic segmentation, in addition to the binary cross-entropy loss for segmentation prediction  $S(\mathbf{x})$  being classified as ground truth by  $D$  [10].

$$J_{seg}(S(\mathbf{x}), \mathbf{y}) = -\frac{1}{B_S} \sum_s \frac{1}{HW} \sum_i \sum_c y_{s,i,c} \log S(x_{s,i,c}) \quad (1)$$

$$J_S(S(\mathbf{x}), \mathbf{y}) = J_{seg}(S(\mathbf{x}), \mathbf{y}) - \lambda_{adv} \frac{1}{B_S} \sum_s \log D(S(x_s)) \quad (2)$$

We use  $x_s$  and  $y_s$  to denote the input image and the ground truth, respectively, where  $x_s$  is of shape  $[H, W, 1]$  and  $y_s$  is of shape  $[H, W, C]$  for  $C$ -class one-hot encoded labels.  $B_S$  denotes the batch size for the segmentor training and  $i$

ranges over all the spatial positions. Given  $S$ ,  $D$  is optimized to maximize the probability of correctly distinguishing  $S(\mathbf{x})$  from  $\mathbf{y}$  as

$$J_D(S(\mathbf{x}), \mathbf{y}) = -\frac{1}{B_D} \sum_s [\log(D(y_s)) + \log(1 - D(S(x_s)))] , \quad (3)$$

where  $B_D$  is the batch size for the discriminator training.

## 2.2 Unsupervised Domain Adaption

Our approach to unsupervised domain adaptation is illustrated in Figure 1 and is based on the idea that prediction masks, unlike input images and intermediate feature representations, can be considered domain independent. Unlike in [10], we do not only make use of a discriminator to judge the quality of the segmentation mask, but also use it to align both source and target segmentation results with the domain-independent prediction mask. We propose an alternative training scheme, where we present the discriminator with real ground truth images from our source domain,  $y_s$ , and with segmentation mask predictions from both the source and the target domain,  $x_s$  and  $x_t$ , respectively. In order to learn domain invariant feature representations, we exploit the fact that we can train the segmentor using both the segmentation and the discriminator loss in the source domain to produce accurate segmentation prediction masks. However, simultaneously we enforce the fact that the segmentation masks for the target domain need to be of high quality. The updated losses are

$$J_{S-DA}(S(\mathbf{x}), \mathbf{y}) = J_S(S(\mathbf{x}), \mathbf{y}) - \lambda_{adv} \frac{1}{B_S} \sum_t \log D(S(x_t)), \quad (4)$$

$$J_{D-DA}(S(\mathbf{x}), \mathbf{y}) = J_D(S(\mathbf{x}), \mathbf{y}) - \frac{1}{B_D} \sum_t \log(1 - D(S(x_t))). \quad (5)$$

## 2.3 Estimation of CTR

CTR is the ratio of maximal horizontal cardiac diameter to maximal horizontal thoracic diameter, as formulated in the Danzer Method [3]. The diameters are the horizontal distance between horizontal coordinates of corresponding key points on the lung contours. As shown in Figure 2, the maximal horizontal cardiac diameter and maximal horizontal thoracic diameter can only be achieved by points above cardiophrenic angles and costophrenic angles, which can be retrieved by the use of a convex hull algorithm. With a hypothetical central line, the Danzer Method could be reinterpreted as  $\frac{A+B}{C+D}$ , while line segments A, B, C, D are all maximized independently. The constraints of maximizing  $A + B$  are that the points of intersection between lung contours and A and B must be above cardiophrenic angles. The points of intersection between the lung contours and the maximized A, B, C, and D are the key points. Provided the estimated CTR, cardiomegaly can be predicted under different thresholds for different age groups. Following [2], the threshold,  $T$ , is chosen to be 0.5.

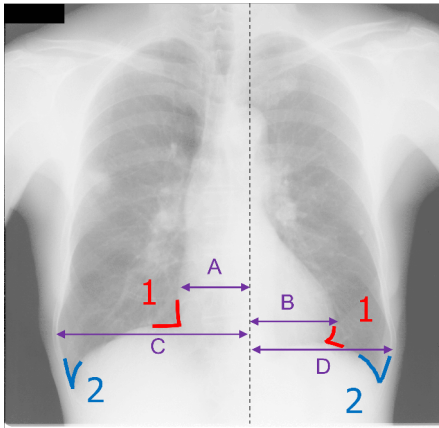


Fig. 2: Contour landmarks for lower lungs: cardiodiaphragmatic angles (1) and costophrenic angles (2).

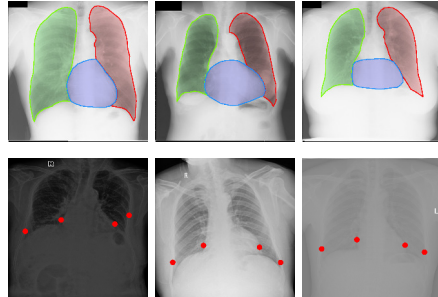


Fig. 3: Example images of the two datasets. The three images in the top row correspond to examples of the JSRT dataset, overlaid with the segmentation annotation. The three images in the second row originate from the Wingspan dataset overlaid with the key points for the CTR calculation.

#### 2.4 Semi-Supervised Semantic Segmentation

We further illustrate our model’s ability for the task of semi-supervised learning. As the annotated data are limited, it is common in medical image segmentation to have only a subset of training data labeled. Provided with a set of labeled and unlabeled datapoints  $\{(x_1, y_1), \dots, (x_l, y_l)\}, \{\tilde{x}_1, \dots, \tilde{x}_u\}$ , the task of semi-supervised learning aims to exploit the underlying data properties of the unlabeled data in addition to the labeled data.  $l$  and  $u$  correspond to the number of labeled and unlabeled examples, respectively. Similar to our unsupervised domain adaptation, we adopt an alternating training strategy, where the model is presented with both labeled and unlabeled data. We optimize  $S$  and  $D$  using Equation 4 and 5 and treat the labeled data as the source domain and the unlabeled data as the target domain. This lets us leverage the unlabeled data to align the distribution of segmentation predictions with the distribution of ground truth labels, effectively regularizing the model and improving overall performance.

### 3 Experimental Results

The **JSRT** dataset is released by the Japanese Society of Radiological Technology (JSRT) [11] and is a benchmark dataset for lung and heart segmentation. JSRT contains 247 grayscale CXRs with annotated lung and heart pixel-wise labels, where 154 have lung nodules and 93 don’t have lung nodules. Each CXR has a size of  $2048 \times 2048$  and the pixel spacing is 0.175mm. In this paper, JSRT is used as the source domain for the unsupervised domain adaption. See Figure 3 for examples from the dataset overlaid with the ground truth annotation.

Method	APE	MAE	RMSE
TL-SEG	16.0% $\pm$ 16.1%	8.9% $\pm$ 9.3%	0.13
TL-ADV	11.4% $\pm$ 11.2%	5.9% $\pm$ 5.9%	0.08
ADDA	9.2% $\pm$ 9.9%	5.1% $\pm$ 5.8	0.08
DA-ADV	5.8% $\pm$ 8.5%	3.3% $\pm$ 5.1%	0.06

Table 1: Results for the unsupervised domain adaptation of CTR estimation experiments. APE denotes average percentage error, MAE denotes mean absolute error, and RMSE denotes root mean square error.

The **Wingspan** dataset is provided by a private research institute, Wingspan Technology. The dataset contains 221 grayscale CXRs for adult patients with annotated key points for calculation of CTR. Each image was annotated by two licensed radiologists independently, and the annotations were accepted by both annotators and an independent reviewer. The de-identified data were collected from 6 hospitals, which have different imaging protocols. The image sizes, pixel spacing and clinical setup vary for each CXR. See Figure 3 for examples from the dataset with key point annotations and the differences to the JSRT dataset and Figure 4 for the large variety in the data modalities, which is not present in the available public benchmark datasets.

In our work, we use the Wingspan dataset as the target domain. We investigate the potential of our proposed approach for unsupervised domain adaptation for the task of CTR estimation. For this, we utilize the segmentation masks of the source domain (JSRT) to perform segmentation on our target domain (Wingspan) and use the predicted segmentation result to compute the CTR. We then show how our method can be easily adapted to semi-supervised semantic segmentation. We evaluate our approach on JSRT and illustrate that we can use the information encoded in our unlabeled data. The adversarial networks are trained using the Adam optimizer with a learning rate of  $10^{-3}$ . The discriminator is updated twice before the segmentor is updated, and  $\lambda_{adv}$  is  $10^{-4}$ . We use  $B_S = B_D = 8$ . JSRT is randomly split into 80% for training and 20% for testing. For all the experiments in this paper, no data augmentation is used, which further shows the robustness of our approach.

**Unsupervised Domain Adaptation:** To assess our performance for unsupervised domain adaptation, we compare our approach (DA-ADV) to three alternative approaches and present the quantitative results for the CTR estimation in Table 1. The baseline uses the segmentor trained on the source domain directly on the target domain. This corresponds to transfer learning without fine-tuning on the target domain (TL-SEG). The baseline segmentor can be improved by adding a discriminator with an adversarial training scheme (TL-ADV). Finally, we compare with one of the state-of-the-art approaches for domain adaptation, ADDA [14], which trains a segmentation network and then utilizes an adversarial loss to align the source and the target domain feature representations in order to minimize data shift. However, ADDA’s performance is highly dependent on the quality of the segmentation network, which is not robust. We observe

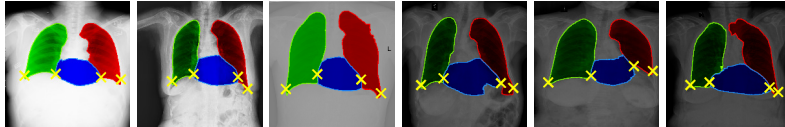


Fig. 4: Visualization of the segmentation and key point results for the Wingspan dataset for our proposed domain adaptation method.

Method	IoU (Lungs)	IoU (Heart)
Human Observer [15]	94.6% $\pm$ 1.8%	87.8% $\pm$ 5.4%
Supervised	95.5% $\pm$ 0.3%	90.2% $\pm$ 0.5%
Supervised (50%)	82.9% $\pm$ 3.5%	71.2% $\pm$ 7.6%
Supervised (25%)	75.4% $\pm$ 5.7%	62.4% $\pm$ 11.9%
Supervised (10%)	60.1% $\pm$ 9.6%	39.4% $\pm$ 14.7%
Semi-Supervised (50%)	90.4% $\pm$ 3.1%	81.2% $\pm$ 2.5%
Semi-Supervised (25%)	89.9% $\pm$ 3.3%	75.5% $\pm$ 5.4%
Semi-Supervised (10%)	81.7% $\pm$ 4.6%	69.4% $\pm$ 7.2%

Table 2: Results for the semi-supervised segmentation experiments. IoU denotes the Intersection over Union.

that our method outperforms the alternative approaches, providing considerable improvements for CTR estimation. Qualitative results for the predicted segmentation masks and the key points for images from the Wingspan dataset can be seen in Figure 4. Based on the threshold of 0.5, we predict cardiomegaly with our pipeline and achieve 87.78% in accuracy, 97.72% in precision, 84.21% in sensitivity and 95.57% in specificity.

**Semi-Supervised Semantic Segmentation:** As a baseline we train the model respectively on 10%, 25% and 50% of annotated data in a supervised manner. As a comparison, we train the model on the whole dataset in a semi-supervised manner, while only portions of the data used in the supervised setting are provided with the labels. Table 2 provides the results of our semi-supervised experiments. Our approach clearly makes use of the unlabeled data, achieving large performance gains. To put our results into perspective and to illustrate the performance that can be achieved when all training labels are available, we also train the model on the fully labeled training dataset.

## 4 Conclusions

In this paper, we present an approach to unsupervised domain adaptation for the task of CTR estimation that is based on the intuition that prediction masks should be domain independent. Using an adversarial training approach, we show that we can predict cardiomegaly from a dataset without segmentation annotations. We further illustrate how our approach can be adapted for semi-supervised learning.

**Acknowledgements.** We thank Wingspan Technology for collecting and annotating the data for this study.

## References

1. Chen, L., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.: Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE TPAMI* (2017)
2. Dallal, A.H., Agarwal, C., Arbabshirani, M.R., Patel, A., Moore, G.: Automatic estimation of heart boundaries and cardiothoracic ratio from chest x-ray images. *Proceedings of the SPIE, Volume 10134, id. 101340K* 10 pp.(2017). 134 (2017)
3. Danzer, C.S.: The cardiothoracic ratio. *The American Journal of the Medical Sciences* 157, 513–554 (1919)
4. Demner-Fushman, D., Kohli, M.D., Rosenman, M.B., Shooshan, S.E., Rodriguez, L., Antani, S.K., Thoma, G.R., McDonald, C.J.: Preparing a collection of radiology examinations for distribution and retrieval. *JAMIA* 23 2, 304–10 (2016)
5. Dimopoulos, K., Giannakoulas, G., Bendayan, I., Liodakis, E., Petraco, R., Diller, G.P., Piepoli, M., Swan, L., Mullen, M., Best, N., A Poole-Wilson, P., Francis, D., Rubens, M., A Gatzoulis, M.: Cardiothoracic ratio from postero-anterior chest radiographs: A simple, reproducible and independent marker of disease severity and outcome in adults with congenital heart disease. *International Journal of Cardiology* 166 (2011)
6. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: *NIPS*. pp. 2672–2680 (2014)
7. Gretton, A., Smola, A., Huang, J., Schmittfull, M., Borgwardt, K., Schölkopf, B.: Covariate shift and local learning by distribution matching, pp. 131–160. MIT Press, Cambridge, MA, USA (2009)
8. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *CVPR*. pp. 770–778 (2016)
9. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: *CVPR*. pp. 3431–3440 (2015)
10. Luc, P., Couprie, C., Chintala, S., Verbeek, J.: Semantic segmentation using adversarial networks. In: *NIPS Workshop on Adversarial Training* (2016)
11. Shiraishi, J., Katsuragawa, S., Ikezoe, J., Matsumoto, T., Kobayashi, T., Komatsu, K.i., Matsui, M., Fujita, H., Kodera, Y., Doi, K.: 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* 174(1), 71–74 (2000)
12. Shu, R., Bui, H., Narui, H., Ermon, S.: A DIRT-t approach to unsupervised domain adaptation. In: *ICLR* (2018)
13. Tavora, F., Zhang, Y., Zhang, M., Li, L., Ripple, M., Fowler, D., Burke, A.: Cardiomegaly is a common arrhythmogenic substrate in adult sudden cardiac deaths, and is associated with obesity. *Pathology* 44, 187–91 (03 2012)
14. Tzeng, E., Hoffman, J., Saenko, K., Darrell, T.: Adversarial discriminative domain adaptation. In: *CVPR*. pp. 2962–2971. *IEEE* (2017)
15. Van Ginneken, B., Stegmann, M.B., Loog, M.: Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database. *Medical Image Analysis* 10(1), 19–40 (2006)