

Learning without Real Data Annotations to Detect Hepatic Lesions in PET Images

Xinyi Yang, Bennett B. Chin, Michael Silosky, Jonathan Wehrend, Daniel V. Litwiller, Debashis Ghosh, and Fuyong Xing

Abstract—Objective: Deep neural networks have been recently applied to lesion identification in fluorodeoxyglucose (FDG) positron emission tomography (PET) images, but they typically rely on a large amount of well-annotated data for model training. This is extremely difficult to achieve for neuroendocrine tumors (NETs), because of low incidence of NETs and expensive lesion annotation in PET images. The objective of this study is to design a novel, adaptable deep learning method, which uses no real lesion annotations but instead low-cost, list mode-simulated data, for hepatic lesion detection in real-world clinical NET PET images. **Methods:** We first propose a region-guided generative adversarial network (RG-GAN) for lesion-preserved image-to-image translation. Then, we design a specific data augmentation module for our list-mode simulated data and incorporate this module into the RG-GAN to improve model training. Finally, we combine the RG-GAN, the data augmentation module and a lesion detection neural network into a unified framework for joint-task learning to adaptively identify lesions in real-world PET data. **Results:** The proposed method outperforms recent state-of-the-art lesion detection methods in real clinical ^{68}Ga -DOTATATE PET images, and produces very competitive performance with the target model that is trained with real lesion annotations. **Conclusion:** With RG-GAN modeling and specific data augmentation, we can obtain good lesion detection performance without using any real data annotations. **Significance:** This study introduces an adaptable deep learning method for hepatic lesion identification in NETs, which can significantly reduce human effort for data annotation and improve model generalizability for PET image quantification.

Index Terms—Lesion detection, PET images, list mode, generative adversarial network, domain adaptation

I. INTRODUCTION

NEUROENDOCRINE tumors (NETs) are rare tumors, which can present with a wide range of nonspecific symptoms such as abdominal pain or discomfort, making this

This work was supported in part by the University of Colorado SPARK Award.

Xinyi Yang, Debashis Ghosh, and Fuyong Xing are with the Department of Biostatistics and Informatics, University of Colorado Anschutz Medical Campus, Aurora, CO 80045, USA (e-mail: xinyi.2.yang@cuanschutz.edu).

Bennett B. Chin and Michael Silosky are with the Department of Radiology, University of Colorado Anschutz Medical Campus, Aurora, CO 80045, USA.

Jonathan Wehrend is with the Department of Radiology, Santa Clara Valley Medical Center, San Jose, CA 95128, USA.

Daniel V. Litwiller is with GE Healthcare, Denver CO, USA.

disease difficult to diagnose [1]. In clinical practice, it is important to detect NETs in the early stage so that patients can receive prompt and appropriate treatments. ^{68}Ga - and ^{64}Cu -DOTATATE positron emission tomography-computed tomography (DOTATATE PET/CT) are the standard-of-care imaging modalities for NET assessment [2], [3], and lesion identification and quantification are critical for subsequent treatment planning. Manual lesion identification is very time-consuming with potential high inter- and intra-observer variation. Thus, many automated methods have attempted to improve the efficiency and objectivity of lesion detection in PET images [4].

Deep neural networks have recently been adopted to identify glycolytically active lesions in fluorodeoxyglucose (FDG) PET imaging data, leading to improved performance compared with traditional image processing and statistical learning methods [5], [6]. However, these methods are designed and evaluated on specific well-annotated PET image datasets. In real world clinical practice, collection and annotation of large DOTATATE PET data sets are both time consuming and subject to human error. This is because NETs are rare tumors and accurate lesion annotation (e.g., delineating lesion boundaries) in complex PET images is very challenging and expensive. One could fine-tune a pre-trained model towards a target-specific PET image dataset, but it is probably still difficult to collect enough labeled target data for appropriate fine-tuning in real applications [7], [8]. In this study, we propose to learn a state-of-the-art deep lesion detection model without any real lesion annotations, but instead using computer simulated list mode PET images. List mode data is raw data acquired in real time, which can be reconstructed in various manners to simulate different imaging conditions [9]–[11]. In our list mode reconstructed PET data, the positions of lesion boundaries are known for each of the various reconstructions. For example, the lesion boundaries are the same for longer and shorter time reconstructed list mode data. Thus, we can eliminate the human effort for data annotation by using list mode data for model training.

The list mode-simulated PET image data may exhibit different distributions from real-world PET images used in routine clinical practice due to many factors such as differences in PET acquisition and processing protocols. For example, Fig. 1 shows a domain shift exists between the two types of data. Thus, a deep model trained with a specific list mode dataset may not be applicable to the desired actual clinical PET images

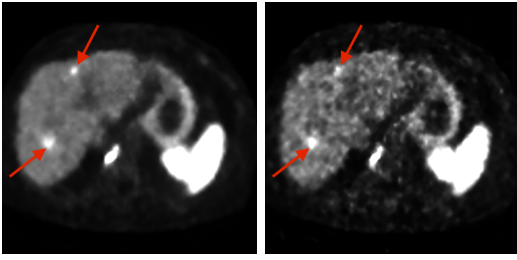


Fig. 1. Example real-world clinical and list mode-simulated PET images. The left represents a real clinical PET image from a NET patient, and the right denotes a list mode-simulated image using a different parameter. The red arrows represent hepatic lesions.

from another scanner or institution. Domain adaptation has recently been employed to address domain shifts in medical image computing, including X-ray, CT, magnetic resonance (MR), and microscopy imaging modalities [12]. In particular, domain adaptation based on generative adversarial networks (GANs) [13], [14] has attracted much attention [15], because it allows for pixel-space adaptation with image-to-image translation, which is beneficial to structured prediction tasks, such as object detection and image segmentation [8], [16], [17]. However, there is very limited exploration of domain adaptation in PET image analysis, especially for GAN-based domain adaptation. One potential reason is that the PET image characteristics, such as high image noise, low spatial resolution and small lesion size, often reduce the visibility of lesions [18], [19], such that lesion structures may not be preserved during image translation with GANs.

In this paper, we propose a novel region-guided GAN (RG-GAN) model to conduct image-to-image translation between list mode-simulated PET images and real-world clinical data. This RG-GAN model can preserve semantic content of interest, i.e., lesions, during image translation so that the lesions in the translated simulated PET images can be used as a supervision signal in domain adaptation for lesion detection. Specifically, we integrate the RG-GAN model and a lesion detection model into an end-to-end, unified framework for joint-task learning, such that the two models can benefit from each other. On one hand, the RG-GAN translates the list mode-simulated data into real world-style images, which appear as drawn from the real clinical PET image dataset, and feed the translated images into the lesion detection model for training. On the other hand, the lesion detection model sends feedback to the RG-GAN and assists with its training to produce desired translated images. In this way, we are able to train a state-of-the-art deep model to detect lesions in real clinical PET images without using any real data annotations. In order to deal with the limited diversity of list mode-simulated PET image data, we design a specific data augmentation module and incorporate it into the unified framework to improve model training. We evaluate the proposed lesion detection method on a real-world clinical ^{68}Ga -DOTATATE PET NET image dataset, and compare it with recent state-of-the-art domain adaptation approaches. In summary, the major contributions are three-fold:

- We propose to learn a hepatic lesion detection model

using list mode-simulated PET images and apply it to real-world PET data. With list mode reconstruction, we can simulate a large set of liver PET images to train a state-of-the-art deep model without any real target data annotations for hepatic lesion detection. To our knowledge, there are currently very few studies, if any, using list mode-simulated PET image data to assist with deep model training and improve model generalization for lesion identification in real-world clinical data.

- We introduce a novel RG-GAN model for image-to-image translation and combine it with a lesion detection model into a unified framework for unsupervised domain adaptation. Compared with other GAN neural networks, our RG-GAN explicitly highlights the lesion regions so that the lesions are preserved during image translation. This is very important for PET images, where lesions are often difficult to detect due to high image noise, low spatial resolution and small object size. Thus, the RG-GAN facilitates learning of the lesion detection model in the framework and enables unsupervised domain adaptation between list mode-simulated PET images and unannotated real clinical data.
- We design a specific data augmentation module to enhance model training within the unified framework. Our list mode-simulated PET images have limited diversity, e.g., spherical-shaped lesions and similar image background. Therefore, we design and apply a series of specific geometric and visual image transformations to model training, such that our RG-GAN produces more diverse PET images for the lesion detection model.

II. RELATED WORK

Deep neural networks have exhibited great power in medical image computing, such as image classification, organ segmentation, object detection, and so on [20]–[23]. Recently, deep learning has been applied to PET image analysis including lesion or tumor quantification and have delivered encouraging performance [5], [6], [24], [25]. Wehrend *et al.* [19] and others [26], [27] have exploited a U-Net-like neural network [28] to locate NET lesions in DOTATATE PET images and has achieved promising lesion detection performance. Iantsen *et al.* [29] have used an enhanced U-Net with squeeze and excitation learning [30] for cervical tumor segmentation in FDG PET images and have showed better results than a standard U-Net [31]. Zhao *et al.* [32] have fed both CT and PET images into a U-Net for lesion localization in prostate cancers and have demonstrated that information fusion of different imaging modalities can help identify lesions in PET images. Other deep learning approaches [33]–[36] also apply pixel-to-pixel or voxel-to-voxel learning to lesion quantification in PET image data, mainly based on U-Nets [28], [31] or fully convolutional networks (FCNs) [37]. However, all these methods rely on a large amount of well-annotated PET data, which is very difficult to obtain in real practice. Additionally, when applying these methods to each new dataset, e.g., images from a different PET scanner, institution or imaging protocol, it is necessary to annotate sufficient data for model training. This is inefficient

and expensive due to the physician operator variability or a need for physician modification of lesion annotations in PET images [4].

Unsupervised domain adaptation (UDA) [38], [39], which transfers learned knowledge from existing annotated datasets to unlabeled target datasets (i.e., datasets of interest), has been employed to address domain shifts in medical image computing [12]. Ouyang *et al.* [40] have combined a variational autoencoder with domain adversarial training to learn a shared feature space for cross-modality cardiac image segmentation. Kamnitsas *et al.* [41] have introduced an adversarial UDA method for brain tumor segmentation in MR images, where a domain classifier is used to encourage domain-invariant feature learning. Ren *et al.* [42] have incorporated adversarial learning into a Siamese neural network to extract domain-agnostic features, which are used to classify prostate cancer histopathology whole-slide images. Dou *et al.* [43] have designed a plug-and-play domain adaptation module to align target features with the source domain for cardiac structure segmentation in MR images. These methods rely on feature matching or alignment, which may be difficult for structured prediction tasks [8], [16], [17], such as lesion detection in PET images.

Instead of conducting feature alignment, some recent UDA methods transform medical images from one style (or modality) to another in the image space. Huo *et al.* [44] have used a CycleGAN [14] for image translation and segmentation in abdominal CT and MR data, Chatsias *et al.* [45] have fed both CT images and labels to a CycleGAN for cardiac MR image synthesis and segmentation, Zhou *et al.* [46] have adopted a modified CycleGAN to convert conventional CT images to cone-beam CT and MR data for liver segmentation, Chen *et al.* [47] have applied diversity-sensitivity regularization [48] to GAN image translation for bone and cardiac CT and MRI image segmentation, and Chen *et al.* [49] have introduced another GAN-based UDA framework for both image transformation and feature alignment in heart and abdominal organ CT and MRI image segmentation. There are some other GAN-based UDA methods, which are applied to histopathological image classification [50], [51], depth-estimation from endoscopy [52], X-ray image segmentation [53], microscopic cell detection [8], [54], and so on. These approaches either do not explicitly enforce geometric structure content to be preserved during image translation or encourage semantic information preservation via implicit task model learning, which may provide insufficient supervision when the task models are not well trained. This will be particularly challenging for lesion identification in PET images, which have often high noise and low resolution so that the lesions are fuzzy and difficult to be separated from the background [4], [18], [19].

There is relatively less exploration of domain adaptation, especially GAN-based methods, for PET images compared with other modalities such as CT, MRI and microscopy. Shin *et al.* [55] have used a conditional GAN (cGAN) [56], [57] to adapt MR data to PET images for diagnosis of Alzheimer's disease, and Wei *et al.* [58] have applied another cGAN-based method to PET image synthesis from multimodal MR image data, showing effective prediction of myelin content

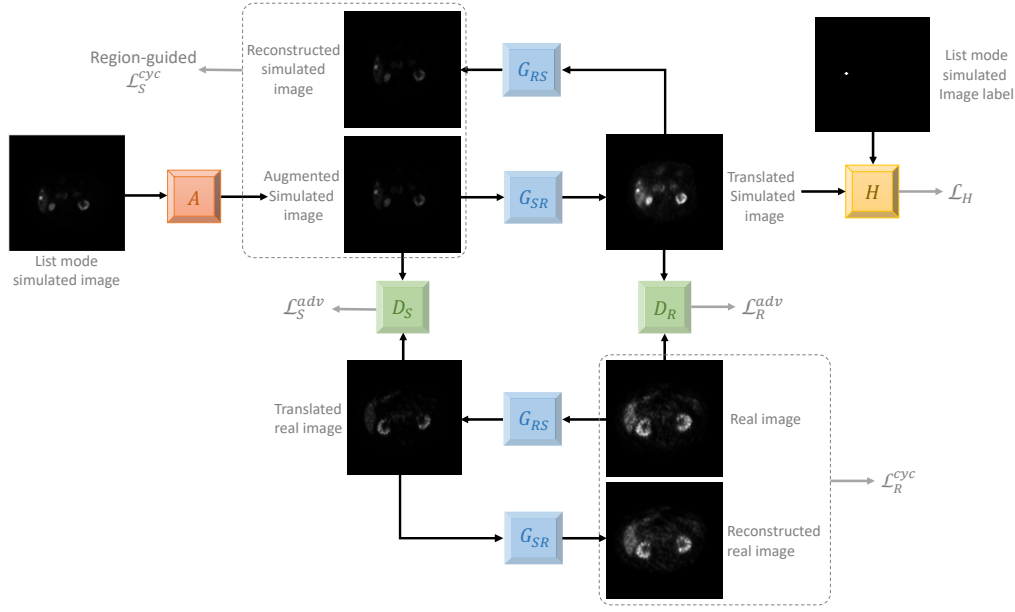
in multiple sclerosis. Ben-Cohen *et al.* [59] have combined a cGAN with an FCN to translate CT to PET images for liver lesion detection, and Bi *et al.* [60] have proposed a multi-channel GAN, which takes as input CT images and corresponding lesion labels, to synthesize PET images for lung tumor identification. By contrast, Armanious *et al.* [61] have augmented a cGAN with non-adversarial losses to generate brain CT images from PET data, and Choi *et al.* [62] have used a pixel-to-pixel cGAN [56] to synthesize structural MR images from amyloid PET. These approaches assume sufficiently diverse, well-annotated source training data available, but this may not be true in real applications. In addition, cGANs typically require paired source-target training images, which are not applicable to our application. Finally, none of these methods transform list mode-simulated, low-cost PET data to real world-style PET images for downstream tasks, such as lesion detection, tumor segmentation, or image quantification.

III. LESION DETECTION WITHOUT REAL DATA ANNOTATIONS

Fig. 2 presents the proposed unified framework for hepatic lesion detection without using real-world clinical PET image annotations. Our RG-GAN translates list mode-simulated PET images into real world-style images, which appear as drawn from real clinical PET data. The lesions are enforced to be preserved via a weighted image reconstruction loss during image translation. The translated simulated images, along with the lesion annotations in the original list mode-simulated data, are used to learn a lesion detection model, which sends feedback to the RG-GAN for training. In order to address the issue of no pairs of list mode-simulated and real-world PET images (because these types of data are acquired from different subjects), we leverage a cycle-consistency constraint [14] to perform image-to-image translation in both directions, i.e., simulated-to-real and real-to-simulated. In addition, we design a specific data augmentation to increase the diversity of list mode-simulated PET image data and improve the training of RG-GAN and lesion detection models.

A. RG-GAN for Lesion-Preserved Image Translation

General GAN-based image translation, such as CycleGAN [14] and other similar methods [63], [64], do not guarantee semantic content preservation such that objects of interest (e.g., lesions) may be removed after image transformation. Some recent approaches introduce an additional model, e.g., a supervised segmentation neural network, to encourage semantic structure preservation during GAN training [65]–[68]. We address this problem from a different perspective: directly apply a weighted matrix to image reconstruction and highlight the lesion regions during GAN training. This is very critical for PET image translation, because lesions often account for a much lower proportion of the image than the background, which may dominate the GAN training if the contributions from the lesion regions are not properly highlighted. In addition, PET images may have a low signal-to-noise ratio and limited spatial image resolution, which also pose additional challenges for segmentation model training.



S

Fig. 2. The proposed method for hepatic lesion detection without real PET image annotations. G_{SR} is the simulated-to-real generator to transform list mode-simulated images to real world-style PET images, and D_R is the corresponding discriminator to classify whether an input image is real. G_{RS} and D_S are the real-to-simulated generator and its associated discriminator, respectively. A denotes the specifically-designed data augmentation module for the list-mode simulated image data, and H represents the lesion detection model, which is trained with the translated simulated images and their associated image labels.

Let \mathbf{X}_S denote the list mode-simulated training PET image data and \mathbf{Y}_S represent the associated labels, each of which is a 3D binary image for one synthesized subject, with 1's for lesions and 0's for the other regions. Let \mathbf{X}_R denote the unannotated real-world clinical PET image data. We aim to construct a unified framework containing an RG-GAN model for robust image-to-image translation and a hepatic lesion detection model H for real clinical PET images. Specifically, the RG-GAN consists of two generator-discriminator pairs, G_{SR} - D_R and G_{RS} - D_S for simulated-to-real and real-to-simulated image translation, respectively. The generator G_{SR} learns to convert simulated images \mathbf{X}_S to real world-style data $G_{SR}(\mathbf{X}_S)$ such that the corresponding discriminator D_R cannot distinguish the translated simulated images $G_{SR}(\mathbf{X}_S)$ from the real ones \mathbf{X}_R . Similarly, the real-to-simulated generator G_{RS} conducts image translation in the reversed direction, trying to fool its associated discriminator D_S . We adopt an image reconstruction-based cycle-consistency constraint [14] to enable unpaired image translation between list mode-simulated and real-world data. More importantly, we design a lesion-specific, weighted matrix and apply it to the image reconstruction for preserving lesion regions during image translation. Formally, our RG-GAN for simulated-to-real image translation can be formulated as

$$\mathcal{L}_R^{adv}(D_R, G_{SR}) = \mathbb{E}_{\mathbf{x}_R \sim \mathbf{X}_R} [\log D_R(\mathbf{x}_R)] + \mathbb{E}_{\mathbf{x}_S \sim \mathbf{X}_S} [\log(1 - D_R(G_{SR}(\mathbf{x}_S)))], \quad (1)$$

$$\mathcal{L}_S^{cyc}(G_{RS}, G_{SR}) = \mathbb{E}_{\mathbf{x}_S \sim \mathbf{X}_S} [\|\mathbf{w}_S G_{RS}(G_{SR}(\mathbf{x}_S)) - \mathbf{w}_S \mathbf{x}_S\|_1], \quad (2)$$

$$\mathbf{w}_S = \xi F(\mathbf{y}_S) + \mathbf{1}, \quad (3)$$

where \mathbb{E} denotes the expectation operator and $\|\cdot\|_1$ represents the l_1 norm. The \mathbf{w}_S is a weighted matrix generated based on the label \mathbf{y}_S for simulated image \mathbf{x}_S , $F(\cdot)$ is a Gaussian filter with standard deviation of 1, $\xi > 0$ is used to emphasize lesions, and $\mathbf{1}$ is a matrix with all elements of 1 and with the same dimension as \mathbf{y}_S . In this way, the computation of the weighted matrix \mathbf{w}_S adapts to each individual subject and emphasizes different-sized lesions with higher weight values during model training, thus encouraging lesion preservation.

Because the real-world data does not have lesion annotations available, we model the real-to-simulated image translation as follows

$$\mathcal{L}_S^{adv}(D_S, G_{RS}) = \mathbb{E}_{\mathbf{x}_S \sim \mathbf{X}_S} [\log D_S(\mathbf{x}_S)] + \mathbb{E}_{\mathbf{x}_R \sim \mathbf{X}_R} [\log(1 - D_S(G_{RS}(\mathbf{x}_R)))], \quad (4)$$

$$\mathcal{L}_R^{cyc}(G_{SR}, G_{RS}) = \mathbb{E}_{\mathbf{x}_R \sim \mathbf{X}_R} [\|G_{SR}(G_{RS}(\mathbf{x}_R)) - \mathbf{x}_R\|_1]. \quad (5)$$

We implement the generators using an FCN architecture [69] with 9 residual learning blocks [70] and 2 transposed convolutional layers [71]. We use a 70×70 PatchGAN [56] for the discriminators, which are applied to image classification with a fully convolutional operation.

B. Data Augmentation for List Mode-Simulated Data

Our list mode-simulated PET images are obtained from an advanced PET scanner with limited normal subject data [11]. The lesions in the list mode data are known a priori, thus eliminating the need of data annotation for these images. However, the list mode-simulated PET image data exhibits very limited diversity such as spherical-shaped lesions and similar image background [11]. Using these simulated images

as input will significantly reduce the variation in output images from the RG-GAN, which learns a deterministic, one-to-one mapping. Thus, this would impair the generalization ability of the lesion detection model. To address this issue, we design a specific, stochastic data augmentation module for the list mode-simulated PET image data. This module consists of a series of geometric and visual image transformations as follows, each of which is applied independently with a certain probability, i.e., $p = 0.5$.

- **Random isotropic scaling** It scales the simulated image for all dimensions using an identical factor, which is randomly sampled from a uniform distribution from 0.9 to 1.1. This transformation adds more variation of lesion size to the data.
- **Random anisotropic scaling** It scales the simulated image with distinct factors for different spatial dimensions (e.g., image width and height), and each scaling factor is randomly sampled from a uniform distribution from 0.9 to 1.1. This technique can change the spherical lesions to more realistic, ellipsoidal lesions in the list mode-simulated data.
- **Random noise corruption** It corrupts the simulated PET images by adding random speckle noise, which is sampled from a Gaussian distribution with mean 0 and standard deviation 1. This aims to make simulated subjects have more divergent image background and mimic real-world data, which usually exhibit different levels of noise.
- **Intensity adjustment** It adjusts image contrast and brightness with two random factors (one for contrast and the other for brightness) sampled from a piecewise uniform distribution from 0.7 to 1.3. This can make the intensity distribution of the entire image look similar to that of real-world PET data.

Due to the unique characteristics of the list mode-simulated PET images, e.g., spherical lesions and similar image background, this data augmentation module is specifically designed and applied to only the list mode-simulated data for diversity enhancement during model training. It will help improve the model generalization ability during the testing stage.

C. Joint Model Learning for Image Translation and Lesion Detection

In addition to RG-GAN for image translation, our unified framework contains another deep neural network H for lesion detection. This neural network is built on a modified U-Net architecture [19], which has shown impressive results for lesion detection in PET images. It has four residual learning blocks [70] in the downsampling path and the upsampling path, and it adds two transposed convolutional layers [71] in the upsampling path for contextual information aggregation [72]. This neural network H takes as input the translated simulated images $G_{SR}(\mathbf{X}_S)$ and produces identical-dimensioned prediction maps for lesion identification. We optimize this network H using a linear combination of a binary cross-entropy loss and a Dice loss [73], which can handle the input data imbalance in our problem.

$$\mathcal{L}_H(G_{SR}, H) = \alpha \mathcal{L}_{BCE}(G_{SR}, H) + \mathcal{L}_{DICE}(G_{SR}, H), \quad (6)$$

$$\mathcal{L}_{BCE}(G_{SR}, H) = \mathbb{E}_{(\mathbf{x}_S, \mathbf{y}_S) \sim (\mathbf{x}_S, \mathbf{y}_S)} \left[\frac{-1}{N} \sum_{i=1}^N (\gamma y_S^i \log \hat{y}_S^i + (1 - y_S^i) \log(1 - \hat{y}_S^i)) \right], \quad (7)$$

$$\mathcal{L}_{DICE}(G_{SR}, H) = \mathbb{E}_{(\mathbf{x}_S, \mathbf{y}_S) \sim (\mathbf{x}_S, \mathbf{y}_S)} \left[1 - \frac{2 \sum_{i=1}^N y_S^i \hat{y}_S^i + 1}{\sum_{i=1}^N y_S^i + \sum_{i=1}^N \hat{y}_S^i + 1} \right], \quad (8)$$

where $y_S^i \in \mathbf{y}_S$ and $\hat{y}_S^i \in \hat{\mathbf{y}}_S$ are the i -th values of the gold-standard label \mathbf{y}_S and its corresponding prediction $\hat{\mathbf{y}}_S = H(G_{SR}(\mathbf{x}_S))$, respectively. The α controls the balance between the binary cross-entropy and Dice losses, and γ weights the foreground regions, i.e., lesions, in PET images. We add a smooth term, value of 1, to the numerator and the denominator in Eq. (8) to avoid division by zero.

We jointly train the RG-GAN and the lesion detection model by optimizing the following full objective

$$\arg \min_{\{G_{SR}, G_{RS}, H\}} \max_{\{D_S, D_R\}} \mathcal{L}_{joint} \quad (9)$$

where

$$\begin{aligned} \mathcal{L}_{joint} = & \mathcal{L}_S^{adv}(D_S, G_{RS}) + \mathcal{L}_R^{adv}(D_R, G_{SR}) \\ & + \lambda_{cyc}(\mathcal{L}_S^{cyc}(G_{RS}, G_{SR}) + \mathcal{L}_R^{cyc}(G_{SR}, G_{RS})) \\ & + \lambda_H \mathcal{L}_H(G_{SR}, H), \end{aligned} \quad (10)$$

where λ_{cyc} and λ_H are weighting parameters for the cycle-consistency loss and the lesion detection loss, respectively. With joint learning, the lesion detection model serves as an additional information resource to further improve RG-GAN training, such that the RG-GAN is encouraged to generate meaningful real world-like images with respect to the labels.

IV. EXPERIMENTS AND DISCUSSION

A. Datasets

A set of list mode-simulated PET/CT images (103 fake subjects with 485 simulated hepatic lesions) are generated from limited real-world normal patient data. The normal patient data is acquired on a state-of-the-art PET/CT scanner (Discovery MI, GE Healthcare) with silicon photodiode PET detectors, which have high sensitivity, high time-of-flight (TOF) resolution, and advanced corrections (e.g., depth dependent point spread function). The PET/CT images are acquired approximately 60 minutes after intravenous administration of approximately 5 mCi of ^{68}Ga -DOTATATE.

An older generation, photomultiplier tube-based PET/CT scanner (Discovery STE, GE Healthcare) is used to acquire a real-world target ^{68}Ga -DOTATATE PET/CT image dataset, which is of lower quality, using the same administered dose and 5 minutes per bed position. This real-world clinical dataset has 123 real subjects, which are different from the subjects used to simulate the aforementioned list mode data, and it includes 56 abnormal cases with ^{68}Ga -DOTATATE hepatic

lesions and 67 normal subjects without lesions. This real-world dataset is randomly divided into training: validation: testing in a ratio of 6:2:2, and the same ratio is applied to data split for the list mode-simulated dataset. This study is determined to be exempt from IRB review by the Colorado Multiple Institutional Review Board at University of Colorado Anschutz Medical Campus.

B. Implementation Details and Evaluation Metrics

We empirically set $\xi = 100$ in Eq. (3), $\alpha = 4$ in Eq. (6), $\gamma = 5$ in Eq. (7), and $\lambda_{cyc} = 10$, $\lambda_H = 15$ in Eq. (10). For the probability of data augmentation, we set the probability of intensity adjustment as 0.2, and the probability of the other data augmentation techniques as 0.5. Following CycleGAN [14], we optimize the RG-GAN using the Adam algorithm [74] with learning rate=0.0002 and batch size=1. We train the lesion detection neural network using stochastic gradient descent with Nesterov momentum with learning rate = 0.0005 and momentum = 0.99.

We use a stage-wise training strategy [65], [66] to optimize the full objective function, Eq. (9). We first train only the RG-GAN for 35 epochs, then train the lesion detection network with RG-GAN-translated simulated images for 25 epochs, and finally jointly train both models for additional 20 epochs. After training, we discard the RG-GAN model, apply only the lesion detection neural network to new testing real-world PET data, and produce a prediction map for each input image. The lesions are identified by binarizing the prediction map with a threshold (i.e., 0.1) followed by connected component analysis.

We use precision, recall and F_1 score as model evaluation metrics [19], [32]. Given that a single subject can have multiple lesions, we adopt the Hungarian algorithm [75] to match gold standard annotations with automated detections. A detected lesion is considered true positive (TP) if the intersection over union (IoU) between this lesion and a gold-standard lesion annotation is greater than a certain threshold [19], [32], i.e., 5% in our experiments; otherwise, the detected lesion is viewed as false positive (FP). Any gold-standard lesion with no matched detection is defined as false negative (FN). With these definitions, we can calculate precision $P = TP/(TP + FP)$, recall $R = TP/(TP + FN)$, and F_1 score $F_1 = (2 \times P \times R)/(P + R)$, for the entire test set.

C. Comparison with State-of-The-Art UDA Methods

We compare our method with several recent state-of-the-art UDA segmentation approaches, including SynSeg-Net [44], anatomy-preserving domain adaptation to segmentation network (APA2Seg-Net) [46] and tumor-aware adversarial domain adaptation (TADA) [67]. We also employ a CycleGAN [14] to translate list mode-simulated PET images to real world-style data and then use these translated data to train a deep neural network with Eq. (6) for lesion detection. We denote this method as CycleGAN in the experiments. Table I presents the comparison results for all these methods in terms of precision, recall, and F_1 score. It is clear that our method outperforms all the others by a large margin, from 6.9% ~ 25.3% in the

TABLE I

COMPARISON WITH STATE-OF-THE-ART UDA METHODS IN TERMS OF PRECISION, RECALL, AND F_1 SCORE (%).

Models	Precision	Recall	F_1 score
CycleGAN [14]	76.1	55.9	64.4
SynSeg-Net [44]	62.1	51.4	56.2
APA2Seg-Net [46]	48.0	49.0	48.5
TADA [67]	78.6	58.3	66.9
Ours	80.7	67.9	73.8

F_1 score. Specifically, SynSeg-Net gives a low F_1 score probably because it does not explicitly enforce semantic content preservation before and after image translation. APA2Seg-Net provides a surprisingly low F_1 score, perhaps because the general-purpose MIND [76] and correlation coefficient [77] regularization may not be applicable to the challenging PET image data used in this study. CycleGAN with stage-wise training delivers a relatively better performance, and TADA further improves the F_1 score, perhaps because it incorporates the labels of source data into model training and help maintain geometric structure in input images. Our method produces the best performance, demonstrating its effectiveness to deal with PET image UDA. Fig. 3 shows some qualitative results of different methods for lesion detection.

D. Ablation Study

To evaluate each component of our method, we conduct the following ablation study shown in Table II: 1) *Simulated*: train a lesion detection model using only the list mode-simulated PET image data and directly apply this model to lesion detection in real-world data, 2) *Baseline*: the proposed method but with neither the weighted image reconstruction (i.e., $\mathbf{w}_S=1$ in Eq. (3)) nor the specific data augmentation module described in Section III-B, 3) *Data-Augment*: the proposed method without the weighted image reconstruction (i.e., $\mathbf{w}_S=1$) but with the proposed specific data augmentation, and 4) *RG-GAN*: the proposed method with the weighted image reconstruction but without the specific data augmentation. To evaluate the effectiveness of each type of image transformation in the specific stochastic data augmentation module, we also train some variants of the proposed method by sequentially adding one transformation: 5) using only random isotropic scaling (*RG-GAN+iso.*), 6) using random isotropic scaling and anisotropic scaling (*RG-GAN+iso.+aniso.*), 7) using random isotropic scaling, anisotropic scaling and noise corruption (*RG-GAN+iso.+aniso.+noise*), and 8) using all the image transformations (*Ours*). The *Real* means training a lesion detection model with real-world data annotations and testing on real-world data.

The *Simulated* model produces the lowest F_1 score, which is as expected, because it does not address the domain shift between the list mode-simulated and real-world data. The *Baseline* model improves the performance to an F_1 score of 55.6%, but it is much lower than the result of the *Data-Augment* or *RG-GAN* model. This demonstrates the effectiveness of the proposed list mode-specific data augmentation

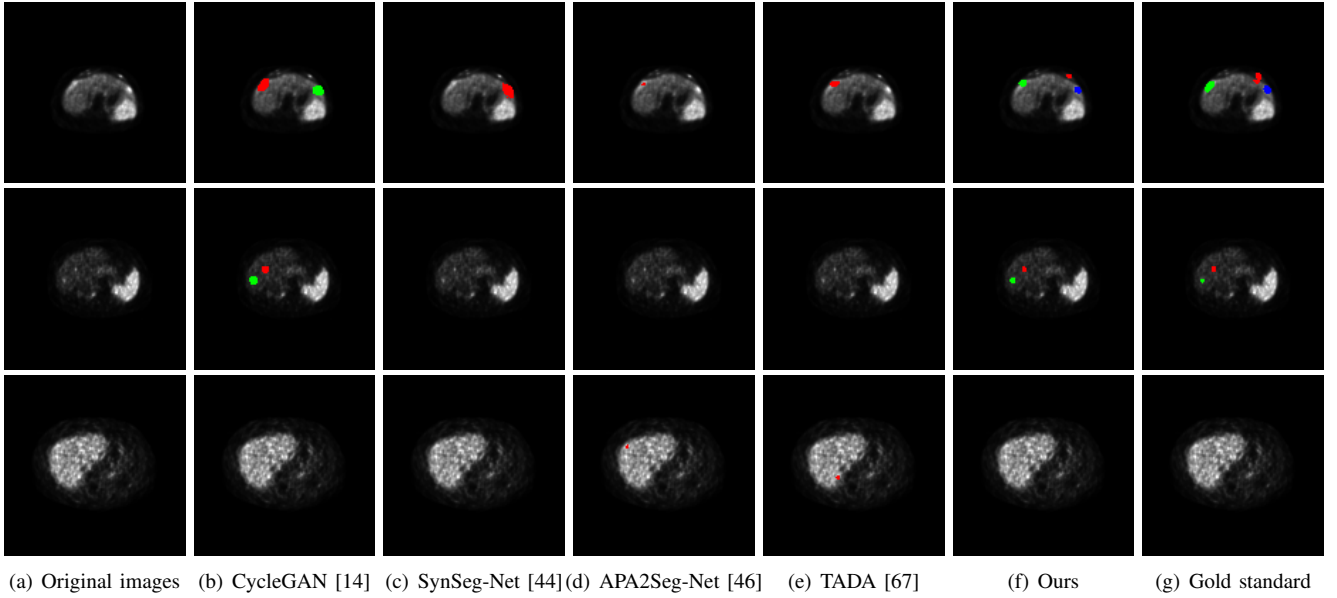


Fig. 3. Lesion detection of different methods. Rows 1 ~ 2 represent 2 different subjects with lesions (i.e., regions marked with different colors) and row 3 denotes a normal subject without lesions.

TABLE II

ABLATION STUDY. ISO., ANISO. AND NOISE REPRESENT RANDOM ISOTROPIC SCALING, ANISOTROPIC SCALING AND NOISE CORRUPTION, RESPECTIVELY.

Models	Precision	Recall	F ₁ score
Simulated	85.0	27.2	41.3
Baseline	50.3	62.1	55.6
Data-Augment	67.5	65.9	66.7
RG-GAN	79.1	61.4	69.1
RG-GAN + iso.	76.5	64.1	69.8
RG-GAN + iso. + aniso.	73.2	68.6	70.8
RG-GAN + iso. + aniso. + noise	81.9	64.1	72.0
Ours	80.7	67.9	73.8
Real	88.9	71.8	79.4

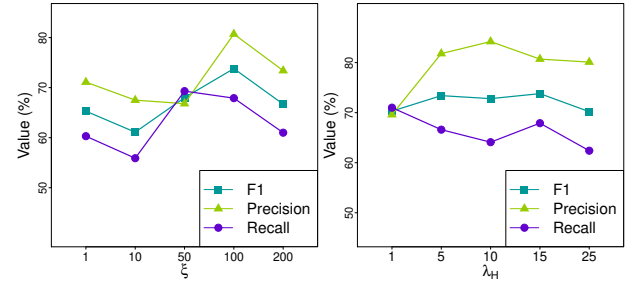


Fig. 4. The precision, recall, and F₁ score of the proposed method using different ξ values (left) for highlighting lesions in Eq. (3) and λ_H values (right) for weighting the lesion detection loss in Eq. (10).

and weighted image reconstruction. With a combination of the specific data augmentation and the *RG-GAN*, we can further boost the lesion detection performance. In particular, when using all the designed image transformations, we can further increase the F₁ score to 73.8% and greatly reduce the gap to the fully supervised *Real* model that is directly trained with real-world PET image annotations. This further suggests that the proposed data augmentation module is beneficial to UDA model training for PET images.

E. Effects of Parameters

We show the effect of ξ in Eq. (3), which is used to highlight lesions during image translation, as shown in the left panel of Fig. 4. For this experiment, we change only the value of ξ and keep all the other parameters fixed. We see that when $\xi < 50$, the F₁ score has a low value, and then it grows dramatically to around 74% at $\xi = 100$. The performance drops when ξ is too large, e.g., $\xi > 100$. This suggests that a small ξ value may not be able to preserve lesions during image translation, while a much higher ξ could be a too strong

emphasis on the lesion regions such that the contributions from other non-lesion regions are suppressed and leads to performance degradation. The right panel of Fig. 4 shows the effect of the weighting parameter λ_H in Eq. (10). We observe that our method can obtain stable performance in a range of $5 \leq \lambda_H \leq 15$ and shows a decreasing F₁ score when $\lambda_H > 15$. This indicates that it is important to achieve a balance between image translation and lesion detection in the joint learning.

We also demonstrate the effects of the probability p of applying each image transformation in the stochastic data augmentation module, as shown in Fig. 5. In this experiment, we change the probability value for only one image transformation at a time, while keeping the value fixed for the others. We see that the lesion detection performance, i.e., F₁ score, for all the image transformations drops dramatically when $p > 0.7$, and this is consistent with the observation in recent studies [78], [79]. Specifically, random isotropic scaling achieves its highest F₁ score at $p = 0.5$, around which it has slight fluctuation, and random anisotropic scaling has a similar trend but with moderate fluctuation. Random noise

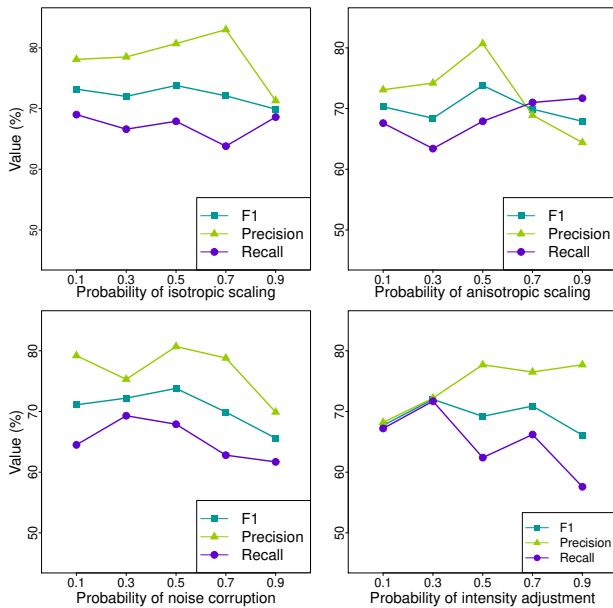


Fig. 5. The precision, recall, and F₁ score of the proposed method using different probability values for each image transformation in the stochastic data augmentation module.

corruption gradually improves the F₁ score from the beginning to $p = 0.5$, and intensity adjustment initially has an increasing F₁ score, which significantly decreases after $p > 0.7$. These results demonstrate that stochastic data augmentation with a proper probability is crucial to GAN-based domain adaptation in lesion detection.

V. CONCLUSION

In this work, we propose a novel end-to-end UDA framework that uses no real data annotations for hepatic lesion detection in real clinical ⁶⁸Ga-DOTATATE PET images. Within the framework, the newly designed RG-GAN model translates low-cost list mode-simulated PET images to real world-style data, and the lesion detection neural network learns an image-to-label mapping using these translated images and enables lesion identification in real PET images. A stochastic data augmentation module, which is specific for the list mode-simulated data, is incorporated into the framework to improve training of the RG-GAN and the lesion detection network, further boosting the performance. The proposed method outperforms several recent state-of-the-art UDA approaches in lesion detection and is very competitive with the fully supervised model that is trained with real data annotations. Our method eliminates the need of data annotation for automated hepatic lesion detection in PET images and can be extended to whole-body lesion identification. It holds great potential for generalizable PET image quantification in clinical research and practice.

REFERENCES

[1] B. Oronsky *et al.*, “Nothing but net: A review of neuroendocrine tumors and carcinomas,” *Neoplasia*, vol. 19, no. 12, pp. 991–1002, 2017.

[2] Y. Sanli *et al.*, “Neuroendocrine tumor diagnosis and management: 68ga-dotatate pet/ct,” *American Journal of Roentgenology*, vol. 211, no. 2, pp. 267–277, 2018.

[3] I. Kayani *et al.*, “A comparison of 68ga-dotatate and 18f-fdg pet/ct in pulmonary neuroendocrine tumors,” *Journal of Nuclear Medicine*, vol. 50, no. 12, pp. 1927–1932, 2009.

[4] M. Hatt *et al.*, “Classification and evaluation strategies of auto-segmentation approaches for pet: Report of aapm task group no. 211,” *Medical Physics*, vol. 44, no. 6, pp. e1–e42, 2017.

[5] M. Hatt *et al.*, “The first miccai challenge on pet tumor segmentation,” *Medical Image Analysis*, vol. 44, pp. 177–195, 2018.

[6] H. Zaidi and I. El Naqa, “Quantitative molecular positron emission tomography imaging using advanced deep learning techniques,” *Annual Review of Biomedical Engineering*, vol. 23, no. 1, pp. 249–276, 2021.

[7] E. Tzeng *et al.*, “Adversarial discriminative domain adaptation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 2962–2971.

[8] F. Xing *et al.*, “Adversarial domain adaptation and pseudo-labeling for cross-modality microscopy image quantification,” in *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention*, vol. 11764, 2019, pp. 740–749.

[9] Z. Zhang *et al.*, “Optimization-based image reconstruction from low-count, list-mode tof-pet data,” *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 4, pp. 936–946, 2018.

[10] J. Wielaard *et al.*, “Optimization of injected 68ga-psma activity based on list-mode phantom data and clinical validation,” *EJNMMI Physics*, vol. 7, p. 20, 2020.

[11] K. H. Leung *et al.*, “A physics-guided modular deep-learning based automated framework for tumor segmentation in pet,” *Physics in Medicine & Biology*, vol. 65, no. 24, p. 245032, 2020.

[12] H. Guan and M. Liu, “Domain adaptation for medical image analysis: A survey,” *IEEE Transactions on Biomedical Engineering*, vol. 69, no. 3, pp. 1173–1185, 2021.

[13] I. Goodfellow *et al.*, “Generative adversarial nets,” in *Proceedings of International Conference on Neural Information Processing Systems*, vol. 27, 2014, pp. 2672–2680.

[14] J.-Y. Zhu *et al.*, “Unpaired image-to-image translation using cycle-consistent adversarial networks,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2223–2232.

[15] X. Yi *et al.*, “Generative adversarial network in medical imaging: A review,” *Medical Image Analysis*, vol. 58, p. 101552, 2019.

[16] W. Hong *et al.*, “Conditional generative adversarial network for structured domain adaptation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1335–1344.

[17] N. Inoue *et al.*, “Cross-domain weakly-supervised object detection through progressive domain adaptation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 5001–5009.

[18] M. Silosky *et al.*, “Physical characteristics of (68)ga dotatate pet/ct affecting small lesion detectability,” *American Journal of Nuclear Medicine and Molecular Imaging*, vol. 11, no. 1, pp. 27–39, 2021.

[19] J. Wehrend *et al.*, “Automated liver lesion detection in 68ga dotatate pet/ct using a deep fully convolutional neural network,” *EJNMMI Research*, vol. 11, no. 1, pp. 1–11, 2021.

[20] H. Greenspan *et al.*, “Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique,” *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1153–1159, 2016.

[21] F. Xing *et al.*, “Deep learning in microscopy image analysis: A survey,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 20, no. 10, pp. 4550–4568, 2018.

[22] G. Litjens *et al.*, “A survey on deep learning in medical image analysis,” *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.

[23] D. Shen *et al.*, “Deep learning in medical image analysis,” *Annual Review of Biomedical Engineering*, vol. 19, no. 1, pp. 221–248, 2017.

[24] B. Cheng *et al.*, “Domain transfer learning for mci conversion prediction,” *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 7, pp. 1805–1817, 2015.

[25] Y. Pan *et al.*, “Synthesizing missing pet from mri with cycle-consistent generative adversarial networks for alzheimer’s disease diagnosis,” in *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2018, pp. 455–463.

[26] A. Weisman *et al.*, “Automated detection and quantification of neuroendocrine tumors on 68ga-dotatate pet/ct images using a u-net ensemble method,” pp. 3215–3215, 2022.

[27] E. A. Carlsen *et al.*, “A convolutional neural network for total tumor segmentation in [64cu] cu-dotatate pet/ct of patients with neuroendocrine neoplasms,” *EJNMMI research*, vol. 12, no. 1, pp. 1–10, 2022.

- [28] O. Ronneberger *et al.*, “U-net: Convolutional networks for biomedical image segmentation,” in *Proceedings of Medical Image Computing and Computer Assisted Intervention*, vol. 9351, 2015, pp. 234–241.
- [29] A. Iantsen *et al.*, “Convolutional neural networks for pet functional volume fully automatic segmentation: development and validation in a multi-center setting,” *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 48, p. 3444–3456, 2021.
- [30] J. Hu *et al.*, “Squeeze-and-excitation networks,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 8, pp. 2011–2023, 2020.
- [31] Ö. Çiçek *et al.*, “3d u-net: Learning dense volumetric segmentation from sparse annotation,” in *Proceedings of International Conference on Medical Image Computing and Computer Assisted Intervention*, 2016, pp. 424–432.
- [32] Y. Zhao *et al.*, “Deep neural network for automatic characterization of lesions on 68ga-psma-11 pet/ct,” *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 47, pp. 603–613, 2020.
- [33] V. Oreiller *et al.*, “Head and neck tumor segmentation in pet/ct: The hecktor challenge,” *Medical Image Analysis*, vol. 77, p. 102336, 2022.
- [34] P. Blanc-Durand *et al.*, “Automatic lesion detection and segmentation of 18f-fet pet in gliomas: A full 3d u-net convolutional neural network study,” *PLoS One*, vol. 13, no. 4, p. e0195798, 2018.
- [35] Z. Xue *et al.*, “Multi-modal co-learning for liver lesion segmentation on pet-ct images,” *IEEE Transactions on Medical Imaging*, vol. 40, no. 12, pp. 3531–3542, 2021.
- [36] D. Jin *et al.*, “Deeptarget: Gross tumor and clinical target volume segmentation in esophageal cancer radiotherapy,” *Medical Image Analysis*, vol. 68, p. 101909, 2021.
- [37] E. Shelhamer *et al.*, “Fully convolutional networks for semantic segmentation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 640–651, 2017.
- [38] G. Wilson and D. J. Cook, “A survey of unsupervised deep domain adaptation,” *ACM Transactions on Intelligent Systems and Technology*, vol. 11, no. 5, pp. 1–46, 2020.
- [39] M. Wang and W. Deng, “Deep visual domain adaptation: A survey,” *Neurocomputing*, vol. 312, pp. 135–153, 2018.
- [40] C. Ouyang *et al.*, “Data efficient unsupervised domain adaptation for cross-modality image segmentation,” in *Proceedings of International Conference on Medical Image Computing and Computer Assisted Intervention*, 2019, pp. 669–677.
- [41] K. Kamnitsas *et al.*, “Unsupervised domain adaptation in brain lesion segmentation with adversarial networks,” in *Proceedings of International Conference on Information Processing and Medical Imaging*, 2017, pp. 597–609.
- [42] J. Ren *et al.*, “Adversarial domain adaptation for classification of prostate histopathology whole-slide images,” in *Proceedings of International Conference on Medical Image Computing and Computer Assisted Intervention*, 2018, pp. 201–209.
- [43] Q. Dou *et al.*, “Unsupervised cross-modality domain adaptation of convnets for biomedical image segmentations with adversarial loss,” in *Proceedings of International Joint Conference on Artificial Intelligence*, 2018, pp. 691–697.
- [44] Y. Huo *et al.*, “Synseg-net: Synthetic segmentation without target modality ground truth,” *IEEE Transactions on Medical Imaging*, vol. 38, no. 4, pp. 1016–1025, 2019.
- [45] A. Chartsias *et al.*, “Adversarial image synthesis for unpaired multimodal cardiac data,” in *Proceedings of International Workshop on Simulation and Synthesis in Medical Imaging*, 2017, pp. 3–13.
- [46] B. Zhou *et al.*, “Anatomy-guided multimodal registration by learning segmentation without ground truth: Application to intraprocedural cbct/mr liver segmentation and registration,” *Medical Image Analysis*, vol. 71, p. 102041, 2021.
- [47] X. Chen *et al.*, “Diverse data augmentation for learning image segmentation with cross-modality annotations,” *Medical Image Analysis*, vol. 71, p. 102060, 2021.
- [48] D. Yang *et al.*, “Diversity-sensitive conditional generative adversarial networks,” in *Proceedings of the International Conference on Learning Representations*, 2019.
- [49] C. Chen *et al.*, “Unsupervised bidirectional cross-modality adaptation via deeply synergistic image and feature alignment for medical image segmentation,” *IEEE Transactions on Medical Imaging*, vol. 39, no. 7, pp. 2494–2505, 2020.
- [50] A. Bentaieb and G. Hamarneh, “Adversarial stain transfer for histopathology image analysis,” *IEEE Transactions on Medical Imaging*, vol. 37, no. 3, pp. 792–802, 2018.
- [51] M. T. Shaban *et al.*, “StainGAN: Stain style transfer for digital histological images,” in *Proceedings of the IEEE International Symposium on Biomedical Imaging*, 2019, pp. 953–956.
- [52] F. Mahmood *et al.*, “Unsupervised reverse domain adaptation for synthetic medical images via adversarial training,” *IEEE Transactions on Medical Imaging*, vol. 37, no. 12, pp. 2572–2581, 2018.
- [53] Y. Zhang *et al.*, “Task driven generative modeling for unsupervised domain adaptation: Application to x-ray image segmentation,” in *Proceedings of International Conference on Medical Image Computing and Computer Assisted Intervention*, 2018, pp. 599–607.
- [54] F. Xing *et al.*, “Bidirectional mapping-based domain adaptation for nucleus detection in cross-modality microscopy images,” *IEEE Transactions on Medical Imaging*, vol. 40, no. 10, pp. 2880–2896, 2021.
- [55] H.-C. Shin *et al.*, “Gandalf: generative adversarial networks with discriminator-adaptive loss fine-tuning for alzheimer’s disease diagnosis from mri,” in *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2020, pp. 688–697.
- [56] P. Isola *et al.*, “Image-to-image translation with conditional adversarial networks,” in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 5967–5976.
- [57] M. Mirza and S. Osindero, “Conditional generative adversarial nets,” *arXiv preprint arXiv:1411.1784*, 2014.
- [58] W. Wei *et al.*, “Learning myelin content in multiple sclerosis from multimodal mri through adversarial training,” in *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2018, pp. 514–522.
- [59] A. Ben-Cohen *et al.*, “Cross-modality synthesis from ct to pet using fcnet and gan networks for improved automated lesion detection,” *Engineering Applications of Artificial Intelligence*, vol. 78, pp. 186–194, 2019.
- [60] L. Bi *et al.*, “Synthesis of positron emission tomography (pet) images via multi-channel generative adversarial networks (gans),” in *Molecular Imaging, Reconstruction and Analysis of Moving Body Organs, and Stroke Imaging and Treatment*, 2017, pp. 43–51.
- [61] K. Armanious *et al.*, “Medgan: Medical image translation using gans,” *Computerized medical imaging and graphics*, vol. 79, p. 101684, 2020.
- [62] H. Choi and D. S. Lee, “Generation of structural mr images from amyloid pet: application to mr-less quantification,” *Journal of Nuclear Medicine*, vol. 59, no. 7, pp. 1111–1117, 2018.
- [63] Z. Yi *et al.*, “Dualgan: Unsupervised dual learning for image-to-image translation,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2849–2857.
- [64] T. Kim *et al.*, “Learning to discover cross-domain relations with generative adversarial networks,” in *Proceedings of International Conference on Machine Learning*, 2017, pp. 1857–1865.
- [65] J. Cai *et al.*, “Towards cross-modal organ translation and segmentation: A cycle-and shape-consistent generative adversarial network,” *Medical Image Analysis*, vol. 52, pp. 174–184, 2019.
- [66] Z. Zhang *et al.*, “Translating and segmenting multimodal medical volumes with cycle- and shape-consistency generative adversarial network,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 9242–9251.
- [67] J. Jiang *et al.*, “Tumor-aware, adversarial domain adaptation from ct to mri for lung cancer segmentation,” in *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2018, pp. 777–785.
- [68] C. Chen *et al.*, “Semantic-aware generative adversarial nets for unsupervised domain adaptation in chest x-ray segmentation,” in *Proceedings of International Workshop on Machine Learning Medical Imaging*, 2018, pp. 143–151.
- [69] J. Johnson *et al.*, “Perceptual losses for real-time style transfer and super-resolution,” in *Proceedings of European Conference on Computer Vision*, vol. 9906, 2016, pp. 694–711.
- [70] K. He *et al.*, “Deep residual learning for image recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.
- [71] V. Dumoulin and F. Visin, “A guide to convolution arithmetic for deep learning,” *arXiv:1603.07285 [stat.ML]*, pp. 1–31, 2016.
- [72] H. Chen *et al.*, “Dcan: Deep contour-aware networks for accurate gland segmentation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2487–2496.
- [73] S. A. Taghanaki *et al.*, “Combo loss: Handling input and output imbalance in multi-organ segmentation,” *Computerized Medical Imaging and Graphics*, vol. 75, pp. 24–33, 2019.
- [74] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” in *Proceedings of the International Conference on Learning Representations*, 2015.

- [75] H. W. Kuhn, "The hungarian method for the assignment problem," *Naval Research Logistics Quarterly*, vol. 2, pp. 83–97, 1955.
- [76] M. P. Heinrich *et al.*, "Mind: Modality independent neighbourhood descriptor for multi-modal deformable registration," *Medical Image Analysis*, vol. 16, no. 7, pp. 1423–1435, 2012.
- [77] Y. Ge *et al.*, "Unpaired whole-body mr to ct synthesis with correlation coefficient constrained adversarial learning," in *Proceedings of SPIE Medical Imaging 2019: Image Processing*, 2019, p. 1094905.
- [78] T. Karras *et al.*, "Training generative adversarial networks with limited data," in *Proceedings of the Neural Information Processing Systems*, vol. 33, 2020, pp. 12 104–12 114.
- [79] F. Xing and T. C. Cornish, "Low-resource adversarial domain adaptation for cross-modality nucleus detection," in *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention*, vol. 13437, 2022, pp. 639–649.