applied sciences

MDPI

Article

DTR-GAN: An Unsupervised Bidirectional Translation Generative Adversarial Network for MRI-CT Registration

Aolin Yang [1], Tiejun Yang [2,3,4,*], Xiang Zhao [1], Xin Zhang [1], Yanghui Yan [1] and Chunxia Jiao [1]

[1] School of Information Science and Engineering, Henan University of Technology, Zhengzhou 450001, China; aolinyang_haut@126.com (A.Y.); learnerzx@gmail.com (X.Z.); learnerzx_haut@126.com (X.Z.); 15343931887@163.com (Y.Y.); jiaooo111@126.com (C.J.)
[2] School of Artificial Intelligence and Big Data, Henan University of Technology, Zhengzhou 450001, China
[3] Key Laboratory of Grain Information Processing and Control (HAUT), Ministry of Education, Zhengzhou 450001, China
[4] Henan Key Laboratory of Grain Photoelectric Detection and Control (HAUT), Zhengzhou 450001, China
* Correspondence: tjyanghlyu@126.com

Abstract: Medical image registration is a fundamental and indispensable element in medical image analysis, which can establish spatial consistency among corresponding anatomical structures across various medical images. Since images with different modalities exhibit different features, it remains a challenge to find their exact correspondence. Most of the current methods based on image-to-image translation cannot fully leverage the available information, which will affect the subsequent registration performance. To solve the problem, we develop an unsupervised multimodal image registration method named DTR-GAN. Firstly, we design a multimodal registration framework via a bidirectional translation network to transform the multimodal image registration into a unimodal registration, which can effectively use the complementary information of different modalities. Then, to enhance the quality of the transformed images in the translation network, we design a multiscale encoder–decoder network that effectively captures both local and global features in images. Finally, we propose a mixed similarity loss to encourage the warped image to be closer to the target image in deep features. We extensively evaluate methods for MRI-CT image registration tasks of the abdominal cavity with advanced unsupervised multimodal image registration approaches. The results indicate that DTR-GAN obtains a competitive performance compared to other methods in MRI-CT registration. Compared with DFR, DTR-GAN has not only obtained performance improvements of 2.35% and 2.08% in the dice similarity coefficient (DSC) of MRI-CT registration and CT-MRI registration on the Learn2Reg dataset but has also decreased the average symmetric surface distance (ASD) by 0.33 mm and 0.12 mm on the Learn2Reg dataset.

Keywords: multimodal image registration; image-to-image translation; unsupervised; deep learning

check for updates

Citation: Yang, A.; Yang, T.; Zhao, X.; Zhang, X.; Yan, Y.; Jiao, C. DTR-GAN: An Unsupervised Bidirectional Translation Generative Adversarial Network for MRI-CT Registration. *Appl. Sci.* **2024**, *14*, 95. https://doi.org/10.3390/app14010095

Academic Editors: Mizuho Nishio and Koji Fujimoto

Received: 9 November 2023
Revised: 14 December 2023
Accepted: 18 December 2023
Published: 21 December 2023

1. Introduction

Medical image registration establishes the precise alignment of medical images for subsequent medical analysis. In medical image analysis, multimodal image registration has a variety of applications, including image-guided interventions, diagnosis, and treatment planning. MRI is frequently used in conjunction with CT for MRI radiation therapy [1–3]. MRI has great soft tissue contrast and may be utilized to accurately portray artificial organs, and CT provides anatomical information along with the electron density for treatment planning and dose computation. Due to their complementary advantages, MRI-CT registration is often required to accurately characterize tumors and organs at risk (OAR) [4]. However, CT and MRI images are often acquired using different devices from the same patient, and the patient's position inevitably varies across modalities. Therefore, it is critical to align MRI images with CT images to facilitate diagnosis and treatment.

Compared with unimodal registration, the main challenge for multimodal image registration is the large differences in grayscale and texture between the different modalities, as well as the difficulty in finding the optimal transformation parameters. As abdominal images contain deformable organs and tissue caused by respiration and other physiological processes, abdomen registration is a challenging problem. The multimodal registration of the abdomen requires consideration of deformation, organ shape variation, and intensity discrepancy between imaging modes. For example, in the case of the liver, the grayscale values of CT and MRI images differ significantly, making it difficult to measure the similarity between the two modalities. Due to the presence of lower tissue contrast in CT images and the lack of obvious structured information in MRI, the image synthesis model from the CT to MRI modality has stronger nonlinearity.

With the development of deep learning-based methods, significant results have been obtained in medical image registration. Deep learning-based methods are mainly classified as supervised and unsupervised. Supervised registration requires the ground truth to guide more promising registration, whereby the dissimilarity between the deformation field and the ground truth is minimized during training. Sun et al. [5] proposed DVNet for CT-US registration, which used a patch-based CNN approach to estimate displacement vectors (DVs), but it had not been demonstrated on clinical CT and US data. Fan et al. [6] proposed BIRNet to achieve dual supervision registration, which used a dual-guided fully CNN network and employed gap filling to achieve promising registration. By enhancing the effectiveness of the label, the performance of the supervised method has been improved. However, supervised registration demands a great deal of accurately labeled data that are time-consuming and difficult to acquire.

The emergence of unsupervised methods can alleviate the lack of datasets that are labeled by experts. For unsupervised registration, the appropriate image similarity metrics need to be selected as the optimization objective. By training the DIRNet with a similarity metric, De Vos et al. [7] achieved the same precise performance as the traditional deformable image registration methods. Balakrishnan et al. [8] designed a CNN-based DIR approach called VoxelMorph for the MRI brain datasets, which achieved rapid registration. However, these methods are mainly utilized for unimodal registration. The intensity distribution of the different modal images is uncertain and complex, making it difficult to use these methods directly for multimodal registration.

The CNN was initially used to learn the mapping relationship for unsupervised registration. By using a pretrained CNN to directly extract features from a pair of images, Kori et al. [9] achieved zero-shot registration on the brain MRI dataset. Yu et al. [10] introduced the regularization term to constrain the anatomical structure deformation for PET-CT registration, in which the registration performance was improved. However, it is challenging for a CNN to map the features of the multimodal image to spatial relationships, as different modalities have significant gaps in appearance.

Subsequently, several deformable image registration (DIR) frameworks based on the GAN are proposed, which show promising performances by providing additional regularization. Yan et al. [11] presented the AIR framework, which utilized adversarial networks to distinguish good registration from poor registration. By connecting a registration network and a discrimination network, Fan et al. [12] created a general registration framework that can achieve monomodal and multimodal image registration. Mahapatra et al. [13,14] designed a cGAN-based [15] model that directly generated the warped image with the deformation field. Compared with the CNN, the GAN has an adversarial training process that makes the DVF more precise. Nonetheless, GANs also have some drawbacks, such as the long duration required for training and vanishing gradients.

Another major use of the GAN is to bridge the enormous appearance gap of different modalities in translation-based multimodal registration. By employing the enhanced CycleGAN [16] to transform a CT image into an MRI-like image, Xu et al. [17] estimated the final deformation field in a dual-stream method and achieved a better registration accuracy.

However, due to the lack of structural consistency across views, the cycle-consistent loss may generate deformed images, which is unfavorable for the subsequent registration performance.

In this paper, we design an unsupervised medical image registration method, DTR-GAN. The main idea is to simplify the complex multimodal image registration (images with an inconsistent grayscale relationship) to unimodal image registration (images with a consistent grayscale relationship) through the image-to-image translation algorithm. The main works are summarized below:

(1) We design an unsupervised image registration framework, DTR-GAN, via a bidirectional image-to-image translation network for MRI-CT registration. By using complementary anatomical structure information from CT and MRI simultaneously, DTR-GAN can overcome the grayscale differences in multimodal images and achieve end-to-end registration.

(2) We propose a multiscale encoder–decoder network to minimize loss of image detail during the image translation process, which can aggregate low-level and high-level features and ensure the generated coherent anatomical feature in the translation network.

(3) We design a mixed similarity loss to penalize the discrepancy in appearance between the target image and the warped image. By extracting the features common to both the source and target images in the deep feature space, DTR-GAN can avoid inconsistent structural deformations during training and enhance registration accuracy.

The rest of this paper is organized as follows. Section 2 reviews relevant studies. Section 3 details the architecture and training loss of our proposed method. Section 4 presents datasets, implementations, and experimental results. In Section 5, we discuss our proposed method. This work is summarized in Section 6.

2. Related Work

The GAN has been employed to transform multimodal image registration into unimodal registration. Tang et al. [18] employed the CycleGAN to synthesize the missing modalities to enhance the precision of multimodal registration algorithms. By employing the CycleGAN to synthesize 3D images and using two multimodal similarity measures, Tanner et al. [19] improved the registration performance effectively. Han et al. [20] used the CycleGAN to synthesize CT images from MRI based on inverse consistency networks for MRI-CT registration. Qin et al. [21] presented a translation-based image registration network for decomposing the image into shape and appearance spaces, which achieved a competitive performance to other methods. By using a translation-based registration method via disentangled representations, Wu et al. [22] designed similarity measures defined in the image space and content space to achieve a superior registration performance.

By applying image-to-image translation to multimodal image registration, Arar et al. [23] used unimodal similarity metrics to train the network instead of using cross-modality similarity measures, which alleviated the drawbacks of a hand-crafted similarity measure. Cao et al. [24] implemented MRI-CT image registration using a patch-wise random forest to bridge the appearance gap in two modalities, which employed a bidirectional image synthesis network to generate two DVFs. Casamitjana et al. [25] proposed a method to train the synthetic model through the registration loss, which avoided unstable training when using GAN networks. Chen et al. [26] designed a discriminator-free image-to-image translation method to mitigate inconsistencies and artifacts that arose from the discriminator, thereby improving the performance of image registration.

To mitigate the coupling challenge between the registration and translation networks, Liu et al. [27] proposed a geometry-consistent adversarial training scheme, which guided the registration network to learn spatial deformation. Kong et al. [28] proposed a self-supervised IMSE method, which evaluated spatial differences in multimodal image registration to establish a novel image-to-image translation paradigm.

However, some multimodal image registration methods based on translation networks tend to synthesize complex anatomical details into simple image modalities, such as MRI to CT translation, which neglect the complex information in the other modality, leading to potential registration biases. Therefore, the quality of the transformed image in an image-to-image translation network plays a crucial role in the effectiveness of the subsequent image registration.

3. Methods

To better utilize the image anatomical information of CT and MRI modalities, we propose an image registration framework named DTR-GAN. The network structure is depicted in Figure 1. DTR-GAN comprises a registration network and a dual contrastive learning translation network based on the GAN, named DT-GAN, to achieve end-to-end registration.

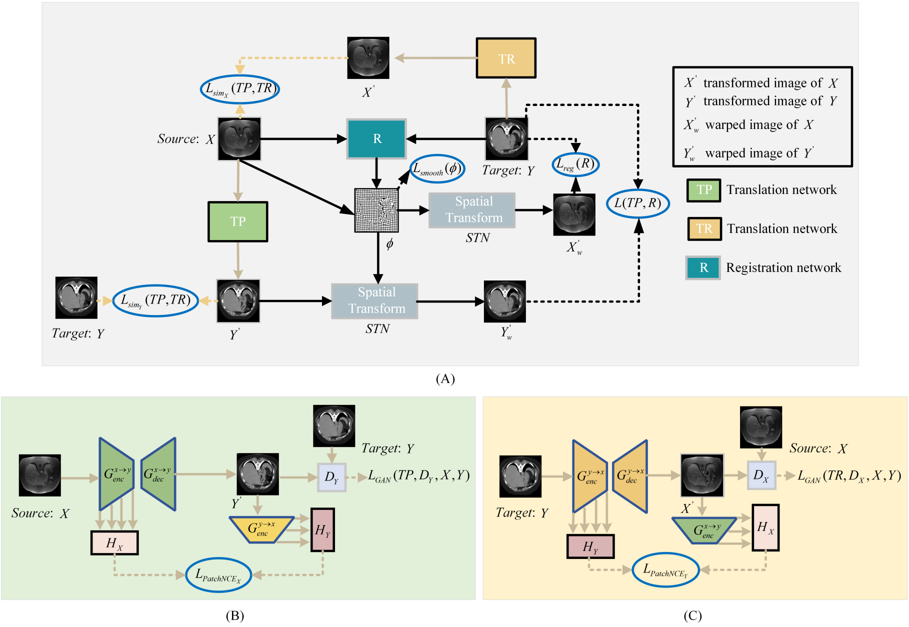

Figure 1. (**A**) shows the structure of the DTR-GAN framework. DTR-GAN consists of TP, R, and TR networks. Our proposed method learns two mappings between MRI and CT images in the translation network, which simplifies the complex multimodal image registration to unimodal image registration. (**B**) shows the structure of the translation network TP for $X \to Y$ transformation. (**C**) depicts the structure of the translation network TR for $Y \to X$ transformation.

The bidirectional translation network DT-GAN includes TP and TR networks, which represent the two mapping directions from the source image X to the target image Y and the target image Y to the source image X, respectively. Inspired by VoxelMorph, we employ a registration network that consists of U-net [29] and a Spatial Transformer Network (STN) [30] to warp the source image into the warped image. When the source image X and target image Y are used as input, the registration network R predicts an optimal mapping ϕ to achieve an excellent alignment of the warped image X'_w with the target image Y. Meanwhile, translation networks TP and TR take X and Y as inputs, respectively, to output the transformed images Y' and X'. The STN transforms the input images into warped images, where the source image X is warped into the registered image X'_w and the translated image Y' is warped into the registered image Y'_w. The entire process can be summarized as follows:

$$\begin{cases} \phi = R(X,Y) \\ Y' = TP(X) \\ X' = TR(Y) \\ Y'_w = STN(Y',\phi) = Y' \circ \phi = R(TP(X),R(X,Y)) \\ X'_w = STN(X,\phi) = X \circ \phi = R(X,R(X,Y)) \end{cases} \tag{1}$$

The loss of the translation network includes the adversarial loss, L_{GAN}; contrastive loss, $L_{PatchNCE_X}(TP, H_X, H_Y, X)$ and $L_{PatchNCE_X}(TR, H_X, H_Y, Y)$; and similarity loss, $L_{sim_X}(TP, TR)$ and $L_{sim_Y}(TP, TR)$. Adversarial loss L_{GAN} strives to minimize the disparity between the generated image and the corresponding target image, contrastive loss keeps the original structure of the image to retain the shape consistency in the image-to-image translation, and similarity loss avoids mode collapse in the translation process. The loss of the registration network includes the mixed similarity loss, $L_{reg}(R)$ and $L(TP, R)$, as well as the smooth loss $L_{smooth}(\phi)$, which maintains the appearance consistency between the warped image X'_w and the target image Y, making the warped transformed image Y'_w and target image Y consistent in appearance. The smooth loss $L_{smooth}(\phi)$ is utilized to prevent the discontinuity in the deformation field. The detailed calculation procedure for each loss function is shown in the subsequent section.

3.1. Registration Network

Figure 2 depicts the architecture of the registration network R, which is made up of U-net and the STN. U-net is utilized to capture the mapping between the input image pairs, while the STN transforms the source image into the warped image.

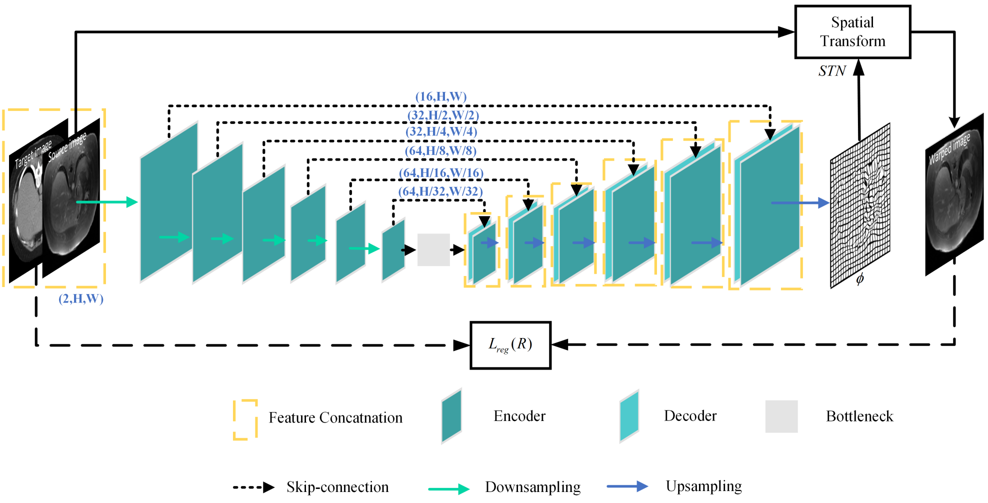

Figure 2. Structure of the registration network.

The registration network takes X and Y as input image pairs and produces the deformation field ϕ as output, in which $\phi = R(X, Y)$ and ϕ is a 2D vector. The STN resamples the source image X and the transformed image Y' into the warped images X'_w and Y'_w, respectively. The STN is defined as

$$\begin{aligned} X'_w[u] &= STN(X, \phi)[p] = X[p + \phi(p)] \\ &= \sum_{q \in N(p+\phi(p))} X(q) \prod_{d\{i,j\}} (1 - |p_d + \phi_d(p) - q_d|) \end{aligned} \tag{2}$$

$$\begin{aligned} Y'_w[u'] &= STN(Y', \phi)[p'] = Y'[p' + \phi(p')] \\ &= \sum_{q' \in N(p'+\phi(p'))} Y'(q') \prod_{d\{i,j\}} (1 - \left|p'_d + \phi_d(p') - q'_d\right|) \end{aligned} \tag{3}$$

where q and q' are pixel neighbors of $p + \phi(p)$ and $p' + \phi(p')$, respectively, and $d\{i, j\}$ represents the two dimensions of the image.

3.2. Dual Contrastive Translation Network

Since it is difficult to perform multimodal image registration with significant intensity differences using a single registration network, we use the bidirectional translation network to ensure a consistent shape in the image-to-image translation. It contributes to aligning two modalities well in the subsequent registration network.

The CycleGAN learned both mappings simultaneously in image-to-image translation, utilizing anatomical information of two modalities, but it led to shape variations in the transformed images. CUT [31] employed contrastive learning to maintain content consistency but used the same embedding to capture information from two different image domains, which might not adequately capture the significant variability in different modalities.

To overcome these constraints, we design a bidirectional translation network DT-GAN. Specifically, we combine an improved CycleGAN with contrastive learning, which replaces the cycle-consistency loss with PacthNCE loss to avoid shape distortion in image-to-image translation. DT-GAN includes translation networks TP and TR, which comprise two generators $G^{x\to y}$ and $G^{y\to x}$ and two discriminators D_X and D_Y, the structure of the TP is shown in Figure 3. Generator $G^{x\to y}$ has an encoder $G_{enc}^{x\to y}$ and a decoder $G_{dec}^{x\to y}$. And generator $G^{y\to x}$ is made up of an encoder $G_{enc}^{y\to x}$ and a decoder $G_{dec}^{y\to x}$. When the image pairs X and Y are used as input, we obtain transformed images Y' and X' from generator $G^{x\to y}$ and generator $G^{y\to x}$ during the translation process, respectively.

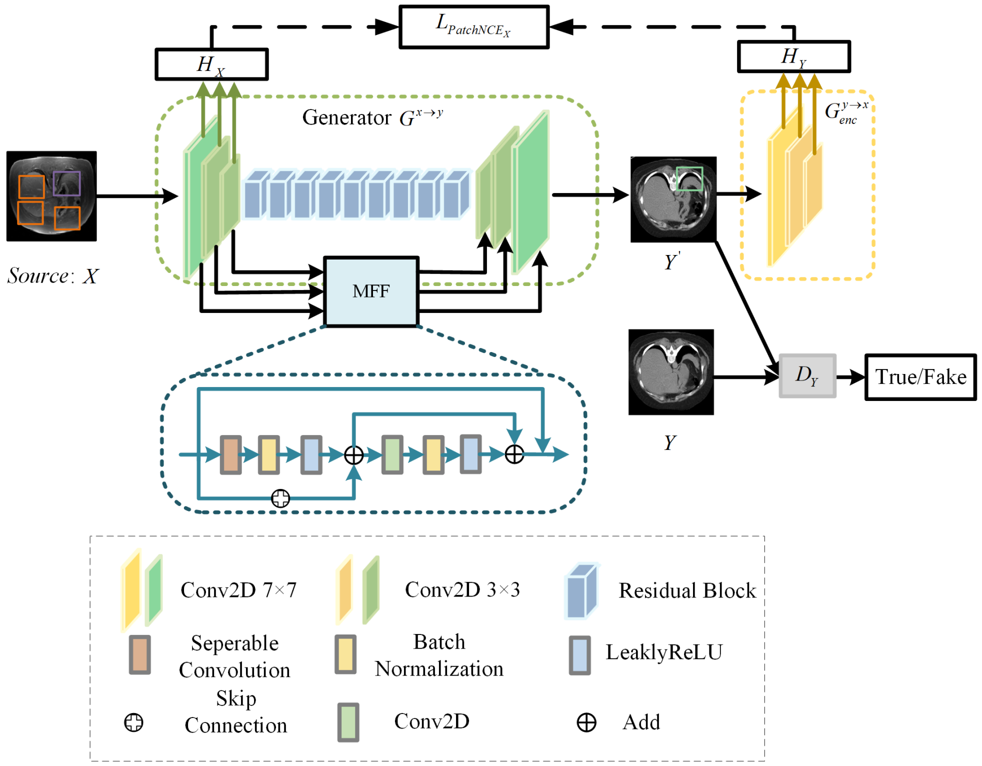

Figure 3. Structure of translation network TP.

3.2.1. Generator

Considering the existence of domain gaps between different domains, we use different embeddings (H_X and H_Y) to extract features of image X and image Y to capture variability in both two image domains. The generator module consists of $G^{x\to y}$ and $G^{y\to x}$ in translation networks TP and TR. Generator $G^{x\to y}$ and generator $G^{y\to x}$ use the same structure but represent translation in different directions, where $G^{x\to y}$ enables the mapping $X \to Y$ and $G^{y\to x}$ enables the mapping $Y \to X$. The generator $G^{x\to y}$ and generator $G^{y\to x}$ are both composed of the encoder, the decoder, nine residual blocks, and the multiscale feature fusion (MFF) module.

We randomly select the query sample, the positive sample, and the negative samples from the input image pairs at each time, which are represented by the green, purple, and orange image patches in Figure 3, respectively. We map the query, positive, and N negative samples into K-dimensional vectors with two-layer MLP, denoted as g, $g^+ \in \mathbb{R}^K$ and

$g^{-} \in \mathbb{R}^{N \times K}$. The distance between the computing query, positive, and negative samples is translated into a classification problem with $N + 1$ classes. The loss is defined as

$$\ell(g, g^{+}, g^{-}) = -log\left[\frac{exp(sim(g, g^{+})/\tau}{exp(sim(g, g^{+})/\tau + \sum_{n=1}^{N} exp(sim(g, g_n^{-})/\tau}\right] \tag{4}$$

where $sim(r, s) = r^T s/||r||||s||$ represents the cosine similarity. We set the value of τ to 0.07.

We use $G_{enc}^{x \rightarrow y}$ to extract the L-layer features of the generator $G^{x \rightarrow y}$ and employ H_X to map the image patches to a stack of features $\{z_l\}_L = \left\{H_{Xl}({G_{enc}^{x \rightarrow y}}_l(X))\right\}_L$, where ${G_{enc}^{x \rightarrow y}}_l(X)$ represents the output of the selected l-layer, $l \in 1, 2, 3, ..., L$, and H_X is a two-layer MLP. For the transformed images Y', a two-layer MLP H_Y is used to extract features from domain Y. Likewise, we obtain another stack of features $\hat{z}_l = \left\{H_{Yl}({G_{enc}^{y \rightarrow x}}_l(TP(X)))\right\}_L$. For the mapping direction of $Y \rightarrow X$, $\{m_l\}_L = \left\{H_{Yl}({G_{enc}^{y \rightarrow x}}_l(Y))\right\}_L$ and $\hat{m}_l = \left\{H_{Xl}({G_{enc}^{x \rightarrow y}}_l(TR(Y)))\right\}_L$, where m_l and $\hat{m}_l$ represent the stack of features.

MFF module. Compared with the original generator structure of DCL-GAN [32], the DT-GAN can capture local and global features by aggregating the spatial information of early feature maps. The structure of the MFF module is shown in Figure 3. MFF is employed to establish connections between the feature maps obtained during the downsampling path and the corresponding feature maps in the upsampling path. The MFF module is composed of two residual units, the separable convolution layer (SC), the convolution layer, the BatchNorm layer(BN), the Leaky-ReLU layer(LReLU), and the skip connection. The convolution layer kernel is 3 × 3. The structure is formulated as

$$z_i = LReLU(BN(Conv2D(LReLU(BN(SC(x_i))) + x_i))) + x_i \tag{5}$$

where x_i and z_i represent input and output feature maps.

3.2.2. Discriminator

The discriminator structure of DT-GAN is PatchGAN [33]. The network mainly comprises five layers: the first layer is a convolution with LReLU nonlinearity; the next three layers are all composed of convolution, instance normalization, and LReLU layers; and the last layer is convolution. And all convolutional kernels are 4 × 4.

In traditional GANs, the discriminator was employed to distinguish the entire generated image as real or fake. However, PatchGAN operates at a patch level instead of the entire image level. PatchGAN utilizes a size of 70 × 70 for the local patch, resulting in a final output of a matrix of size 30 × 30 by stacking convolutional layers, where each element represents a larger receptive domain in the source image, corresponding to a patch in the original image. By evaluating local patches, PatchGAN can capture and enforce precise details and textures in the generated images, achieving higher quality image translation.

3.3. Training Loss

3.3.1. Registration Loss

(1) Mixed similarity loss

The existing multimodal similarity measures utilized as loss functions in the network remain inefficient, thereby impacting the registration performance of the training process. For instance, the NMI is computationally difficult and unsuitable for gradient-based methods. Similarly, the NCC is dependent on the domain and cannot be generalized to all modalities. The translation-based registration framework can alleviate the problem of hand-crafted multimodal similarity measures. However, during the image-to-image translation process, there may be a distribution mismatch, especially when dealing with images that have a complex appearance. Therefore, it is unreliable to rely solely on translation algorithms to transform multimodal registration into unimodal registration. Hence, to reduce the impact of mismatched images generated during the translation process, we propose a mixed similarity loss that focuses on the structural information of the images, aiming to maximize the similarity between the registered image and the target image.

To penalize appearance discrepancies between target and warped images, we propose a similarity loss $L_{reg}(R)$. $L_{reg}(R)$ is defined as

$$L_{reg}(R) = \left\| H_{YP}(G_{enc}^{y\rightarrow x}(Y)) - H_{XP}(G_{enc}^{x\rightarrow y}(X_w^{'})) \right\|_1^{sum} \tag{6}$$

where *sum* represents summing them up together. We utilize separate embeddings H_{XP} and H_{YP} to extract the 256-dimensional feature stacks from the input image pairs and then project them onto 64-dimensional vectors. H_{XP} and H_{YP} consist of 2-layer MLP, alongside convolutional, ReLU, average pooling, linear transformation, ReLU, and final linear transformation layers. To explore the potential mapping relationships in the feature space, we extract features shared between the warped image and the target image to guide the accurate registration.

To minimize the appearance dissimilarity between the warped image of the translated image $Y_w^{'}$ and the target image Y, $L(TP, R)$ is defined as

$$L(TP, R) = \left\| Y_w^{'} - Y \right\|_1 \tag{7}$$

So, the mixed similarity loss L_{ms} is computed by

$$L_{ms} = L_{reg}(R) + L(TP, R) \tag{8}$$

(2) Smooth loss

To prevent discontinuity in the deformation field ϕ when making the warped image closer to the target image, the smooth loss L_{smooth}, which is simiar to the TV loss [34], is employed to constrain ϕ to avoid excessive distortion of the warped image.

$$L_{smooth}(\phi) = \sum_{p\in\Omega} \|\nabla\phi(p)\|^2 \tag{9}$$

$$\nabla\phi(p) = \left(\frac{\partial u(p)}{\partial x}, \frac{\partial u(p)}{\partial y}\right) \tag{10}$$

$$\frac{\partial u(p)}{\partial x} \approx \phi((p_x+1), p_y) - \phi(p_x, p_y) \tag{11}$$

$$\frac{\partial u(p)}{\partial y} \approx \phi(p_x, (p_y+1)) - \phi(p_x, p_y) \tag{12}$$

where Ω represents the positions of the neighboring pixels of p.

3.3.2. Translation Loss

(1) Adversarial loss

The adversarial loss, $L_{GAN}(TP, D_Y, X, Y)$ and $L_{GAN}(TR, D_X, X, Y)$, is utilized to match the distributions of the transformed image with the target image. Therefore, the overall adversarial loss L_{GAN} is

$$\begin{aligned} L_{GAN} &= L_{GAN}(TP, D_Y, X, Y) + L_{GAN}(TR, D_X, X, Y) \\ &= \mathbb{E}_{y\sim Y}[logD_Y(Y)] + \mathbb{E}_{x\sim X}[1 - logD_Y(TP(X))] \\ &+ \mathbb{E}_{x\sim X}[logD_X(X)] + \mathbb{E}_{y\sim Y}[1 - logD_X(TR(Y))] \end{aligned} \tag{13}$$

(2) PatchNCE loss

Inspired by CUT, we introduce PatchNCE loss to enforce content consistency during bidirectional image translation, which can ensure the maximum preservation of the shape of the input image. PatchNCE loss in the transformation directions of X to Y and Y to X can be estimated by

$$L_{PatchNCE_X}(TP, H_X, H_Y, X) = \mathbb{E}_{x \sim X} \sum_{l=1}^{L} \sum_{s=1}^{S_l} \ell(\hat{z}_l^s, z_l^s, z_l^{S \backslash s}) \tag{14}$$

$$L_{PatchNCE_Y}(TR, H_X, H_Y, Y) = \mathbb{E}_{y \sim Y} \sum_{l=1}^{L} \sum_{s=1}^{S_l} \ell(\hat{m}_l^s, m_l^s, m_l^{S \backslash s}) \tag{15}$$

where S_l represents the number of spatial locations in the generator encoder of each layer l, $z_l^s \in \mathbb{R}^{C_l}$, $m_l^s \in \mathbb{R}^{C_l}$, $z_l^{S \backslash s} \in \mathbb{R}^{(S_l-1)C_l}$, $m_l^{S \backslash s} \in \mathbb{R}^{(S_l-1)C_l}$, and C_l represents the number of channels in each layer l.

(3) Similarity loss

To prevent mode collapse during the generation of the transformed images into the translation network, the overall similarity loss $L_{sim}(TP, TR)$ is used to encourage the translated image's depth features to resemble the original image. The $L_{sim}(TP, TR)$ is defined as

$$\begin{aligned} L_{sim}(TP, TR) &= L_{sim_X}(TP, TR) + L_{sim_Y}(TP, TR) \\ &= \left[\left\| H_{xr}(H_X(G_{enc}^{x \to y}(X))) - H_{xf}(H_X(G_{enc}^{x \to y}(TR(Y)))) \right\|_1^{sum} \right] + \\ &\left[\left\| H_{yr}(H_Y(G_{enc}^{y \to x}(Y))) - H_{yf}(H_Y(G_{enc}^{y \to x}(TP(X)))) \right\|_1^{sum} \right] \end{aligned} \tag{16}$$

where H_{xr}, H_{xf}, H_{yr}, and H_{yf} are four light networks, which are used to project the embedded features to 64-dim vectors. Each light network is composed of a convolutional layer, which is subsequently followed by a ReLU activation function, average pooling, a linear transformation, another ReLU activation function, and, ultimately, a linear transformation.

3.4. Objection Function

The total loss function L_R of DTR-GAN is calculated by

$$\begin{aligned} L_R = {} & \lambda_R L_{reg}(R) + \lambda_T L(TP, R) + \lambda_S L_{smooth}(\phi) + \lambda_G L_{GAN} + \\ & \lambda_P (L_{PatchNCE_X}(TP, H_X, H_Y, X) + L_{PatchNCE_Y}(TR, H_X, H_Y, Y)) + \\ & \lambda_M L_{sim}(TP, TR) \end{aligned} \tag{17}$$

where we set λ_R = 10, λ_T = 1, λ_S = 0.2, λ_G = 0.15, λ_P = 0.3, and λ_M = 10.

4. Results

4.1. Dataset and Preprocessing

We utilize two abdominal datasets for verifying the effectiveness of our method, the Learn2Reg dataset [35] and the CHAOS dataset [36]. The Learn2Reg dataset consisted of 16 MRI and CT datasets and each scan included the liver, spleen, and left and right kidneys, four organs, which were labeled with manual and automatic organ segmentation. Each scan had a 3D volume of size 192 × 160 × 192 with a voxel spacing of 2 mm. For the experiment, we use Elastix [37] for coarse affine registration and set the train/validation/test datasets in 10/2/4 pairs randomly. The healthy abdominal organs were from 80 patients, where 40 of them went through CT scans and 40 of them went through MRI scans in the CHAOS dataset. We use 20 volumes of CT and T1-DUAL MRI images in the CHAOS dataset and randomly divide them into 14/2/4 pairs for train/validation/test datasets. For the preprocessing of both datasets, we select 70 central slices of each piece of data for our experiments. The input size of the slice is 256 × 256. Figure 4 illustrates examples from both the Learn2Reg dataset and the CHAOS dataset. Table 1 shows a concise description of the two datasets utilized in our method.

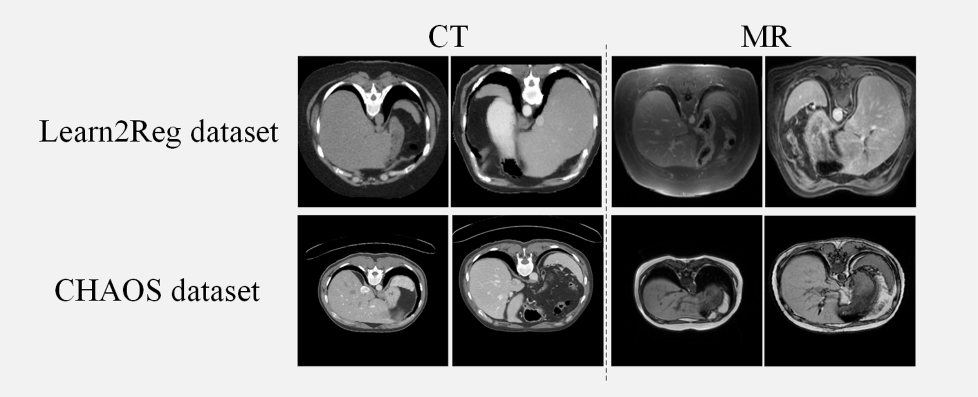

Figure 4. Instances from the Learn2Reg and CHAOS datasets.

Table 1. A concise description of the two datasets utilized in our method.

Source	Modality	Size	Train/Validation/Test	Resize
Learn2Reg	CT, MRI	192 ×160	10/2/4	✓
CHAOS	CT, MRI	-*	14/2/4	✓

* - represents that there are many different sizes and we cannot list them all.

4.2. Implementation Details

We utilize the Pytorch framework to implement DTR-GAN, which runs on a hardware environment equipped with a GeForce RTX 3080Ti GPU and 12 GB RAM. The network is trained for 300 epochs with a batch size of 1. The DTR-GAN employs the Adam optimizer, the initial learning rate is 2×10^{-4}, and there is linear decay in the learning rate after 200 epochs. The optimization parameters are β_1 = 0.5 and β_2 = 0.999.

4.3. Evaluation Metrics

DSC. The DSC [38] is primarily utilized to assess the overlapping degree of two images. The DSC is defined as

$$\mathrm{DSC} = \frac{2|\mathrm{W} \cap \mathrm{T}|}{|\mathrm{T}| + |\mathrm{W}|} \times 100\% \tag{18}$$

where T and W represent the labels of the target image and warped image, respectively. The DSC ranges from 0 to 1, with 0 indicating no overlap and 1 indicating complete overlap in the labels of the two images.

ASD is a metric employed to evaluate the surface similarity between two images, and it represents the average distance between two surfaces. The ASD considers the differences in edge positions after registration. The formula of the ASD is

$$\mathrm{ASD} = \frac{1}{|\mathrm{W}| + |\mathrm{J}|} \left(\sum_{\mathrm{w} \in \mathrm{W}} \mathrm{d}(\mathrm{w}, \mathrm{J}) + \sum_{\mathrm{j} \in \mathrm{J}} \mathrm{d}(\mathrm{j}, \mathrm{W}) \right) \tag{19}$$

where w and j are the points on the propagated surface W and the reference surface J, respectively. $d(\cdot)$ is expressed as the minimum Euclidean distance.

4.4. Results on the Learn2Reg and CHAOS Datasets

To validate the performance of DTR-GAN, we compare it with Affine [37], VoxelMorph [8], CR-GAN, SbR [25], RoT [23], IMSE [28], and DFR [26] on the Learn2Reg and CHAOS datasets. Affine represents the registration framework using the Elastix toolbox. VoxelMorph is a CNN-based registration network. CR-GAN is a VoxelMorph registration network based on the CycleGAN translation network with PatchNCE loss. SbR is a 2D synthesis-based registration network using contrastive learning to enforce geometric consistency. RoT is a multimodal registration network based on the geometric-preserving translation network.

IMSE is a self-supervised multimodal image registration network with a multimodal spatial evaluator to evaluate spatial differences. DFR is the translation-based multimodal registration network that removes the discriminator in translation work.

Results on the Lean2Reg dataset. Table 2 illustrates average DSC scores and ASD scores on the Lean2Reg dataset. Compared with DFR, the DSC scores of DTR-GAN are improved by 2.35% and 2.08% in MRI-CT registration and CT-MRI registration, respectively. In addition, compared with DFR, the ASD values of DTR-GAN are decreased by 0.33 mm and 0.12 mm, respectively. In addition, among the above methods, DTR-GAN achieves the best results for the DSC and ASD. Bidirectional translation networks TP and TR generate promising translated images in DTR-GAN, thereby improving the registration performance. Specifically, MFF aggregates multiscale information in the translation network DT-GAN, allowing DTR-GAN to better capture local and global features in images. By designing a mixed similarity loss, the anatomical details in the two modalities can be aligned well.

Table 2. Quantitative results of different methods on the Learn2Reg dataset. The best results are marked in bold for clarity. ↑ means a better registration performance should obtain higher metric values. ↓ means a better registration performance should obtain lower metric values.

	MRI-CT		**CT-MRI**	
	***DSC* (%) ↑**	***ASD* (mm) ↓**	***DSC* (%) ↑**	***ASD* (mm) ↓**
Affine [37]	65.51 ± 3.10	5.63 ± 1.77	64.91 ± 3.10	6.10 ± 1.77
VoxelMorph [8]	68.21 ± 2.42	6.10 ± 1.37	66.60 ± 3.94	6.32 ± 1.67
CR-GAN	77.81 ± 1.80	5.15 ± 1.57	76.19 ± 2.70	5.50 ± 1.51
SbR [25]	76.52 ± 2.62	4.55 ± 1.36	75.51 ± 1.60	4.55 ± 1.36
RoT [23]	74.81 ± 0.78	5.05 ± 1.79	73.30 ± 2.20	5.65 ± 1.79
IMSE [28]	74.29 ± 2.37	5.05 ± 2.12	73.79 ± 2.21	5.06 ± 1.62
DFR [26]	78.50 ± 2.50	4.54 ± 1.22	77.65 ± 2.50	4.74 ± 1.11
DTR-GAN	**80.85 ± 2.09**	**4.21 ± 1.65**	**79.73 ± 2.03**	**4.62 ± 1.76**

Figure 5 shows the registration results of some methods on the Learn2Reg dataset. Figure 5A,E show warped images, Figure 5B,F show deformation fields, and Figure 5C,D,G,H show checkerboard grids and overlapping images, respectively, which are visualizations of image alignment effects. Figure 5A,E show the warped images in DTR-GAN are closer to the target image. The first in column (A) denotes that the cycle-consistency loss would lead to shape distortion in CR-GAN. We can observe that DTR-GAN with similarity loss is more effective for appearance preservation in image registration. Figure 5D,H show the effects of overlaying the target image with the warped image, in which the target image is set to blue, the warped image is set to red, and the two images are superimposed together in purple. The purple in the image is more evident, indicating that the warped image and target image are better aligned. As shown in Figure 5, DTR-GAN shows an optimum performance.

Results on the CHAOS dataset. Table 3 shows the experimental results on the CHAOS dataset. Compared with DFR, the DSC scores of DTR-GAN increased by 1.98% and 1.51% on average in MRI-CT registration and CT-MRI registration, respectively. In addition, the ASD scores of DTR-GAN are decreased by 0.28mm and 0.26mm on average. Figure 6 illustrates the registration results of different methods. Figure 6A,E show warped images, Figure 6B,F show deformation fields, and Figure 6C,D,G,H depict checkerboard grids and overlapping images, respectively, which are visualizations of image alignment effects. As shown in Figure 6, we can observe that the warped image in DTR-GAN is closer to the target image. Figure 6C,G show that the alignment of DTR-GAN is more optimum than the other approaches. The registration results are shown in Figure 7, which shows that our DTR-GAN has a fine registration performance.

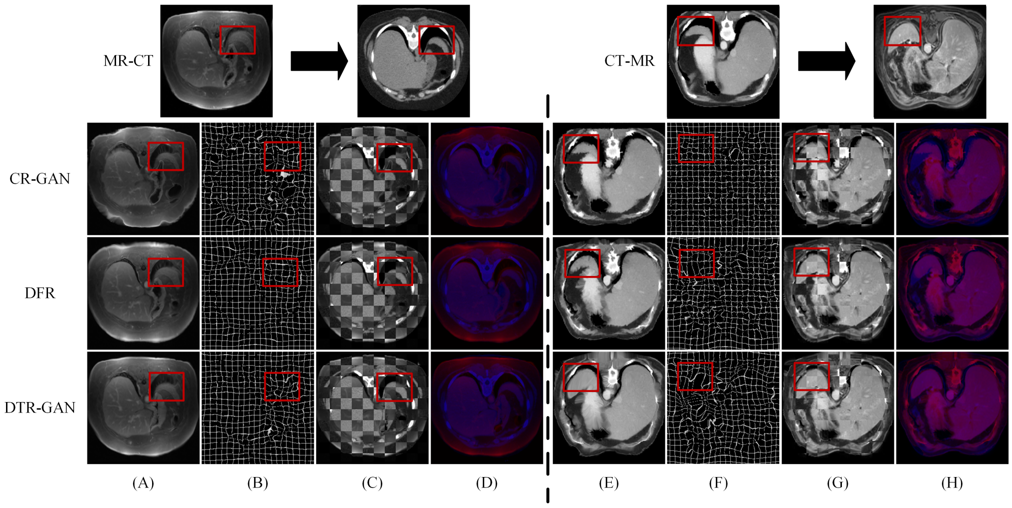

Figure 5. Experimental results on the Learn2Reg dataset: (**A**,**E**) are warped images, (**B**,**F**) are deformation fields, (**C**,**G**) are checkboard grids, and (**D**,**H**) are overlapping images.

Table 3. Quantitative results of different methods on the CHAOS dataset. The best results are marked in bold for clarity.

	T1-CT		**CT-T1**	
	***DSC* (%) ↑**	***ASD* (mm) ↓**	***DSC* (%) ↑**	***ASD* (mm) ↓**
Affine [37]	65.15 ± 6.20	7.54 ± 2.56	64.23 ± 6.40	7.66 ± 2.40
VoxelMorph [8]	69.59 ± 6.81	7.23 ± 2.41	68.83 ± 6.10	7.43 ± 2.40
CR-GAN	76.67 ± 4.20	7.08 ± 3.93	75.18 ± 4.81	7.26 ± 2.78
SbR [25]	75.49 ± 6.18	6.77 ± 2.65	74.63 ± 6.18	6.82 ± 2.66
RoT [23]	73.52 ± 5.62	7.55 ± 2.36	72.20 ± 5.29	7.62 ± 2.05
IMSE [28]	74.26 ± 2.07	7.17 ± 2.86	73.42 ± 2.30	7.40 ± 2.18
DFR [26]	78.19 ± 6.06	6.35 ± 2.41	77.53 ± 6.35	6.54 ± 2.08
DTR-GAN	**80.17 ± 1.77**	**6.07 ± 2.19**	**79.04 ± 2.40**	**6.28 ± 2.50**

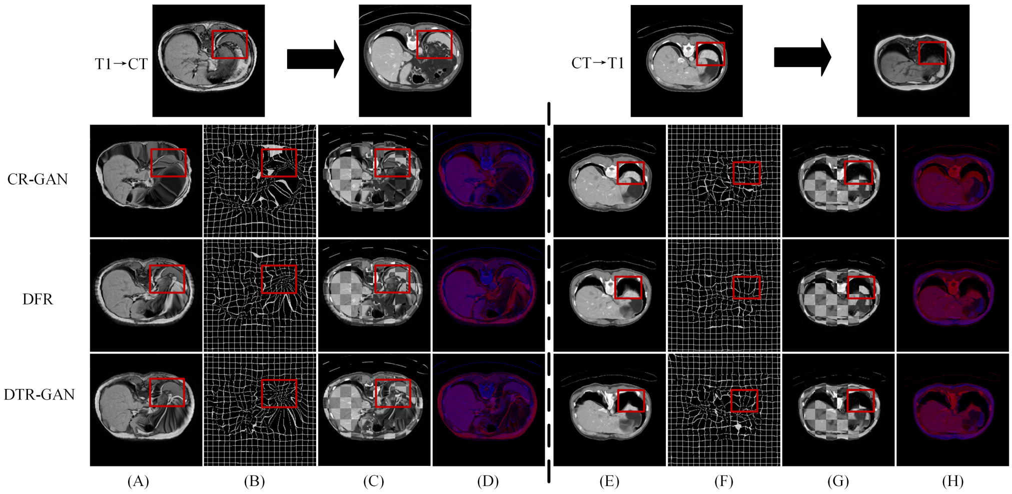

Figure 6. Experimental results on the CHAOS dataset. (**A**,**E**) are warped images, (**B**,**F**) are deformation fields, (**C**,**G**) are checkboard grids, and (**D**,**H**) are overlapping images.

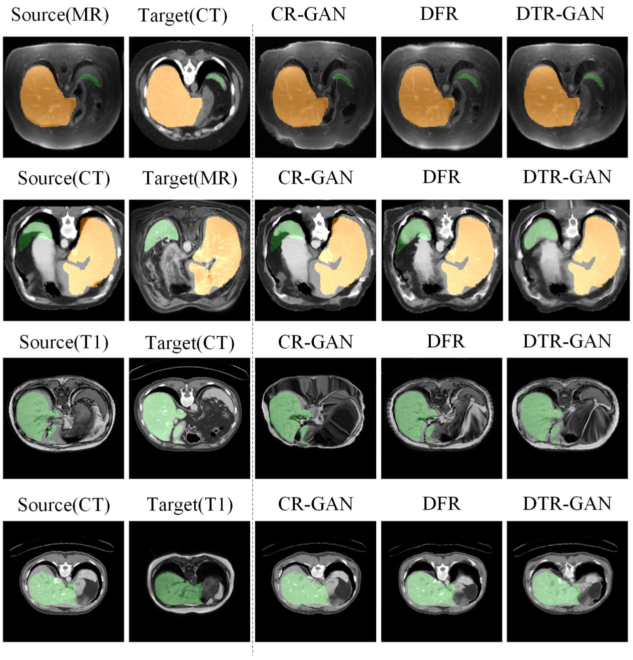

Figure 7. Visual comparison of registration accuracy using various methods is conducted on the Learn2Reg and CHAOS datasets. The first two columns display the original inputs, and all columns show vivid colors to visualize the label.

4.5. Ablation Experiment

To verify the effectiveness of network components, we perform detailed ablation studies, using a dual contrastive translation network and MFF module, as well as $L_{reg}(R)$ in different configurations. TR-GAN indicates the registration network, which contains the unidirectional translation network TP and utilizes the similarity loss $L_{reg}(R)$. DR-GAN represents the registration network, which contains bidirectional translation networks TP and TR with similarity loss $L_{reg}(R)$. DTR-GAN denotes the DR-GAN with the MFF module, and similarity loss $L_{reg}(R)$ is used. In addition, to verify the effectiveness of $L_{reg}(R)$, we replace $L_{reg}(R)$ in DTR-GAN with PatchNCE loss, which is used in DFR, named PDTR-GAN. Figures 8 and 9 show the average DSC and ASD of the MRI-CT registration on the Learn2Reg dataset and the CHAOS dataset, respectively. DTR-GAN achieves the best performance, which is comparable with other methods. The improvements with statistical significance represent that the bidirectional translated network can generate more refined translated images, improving the subsequent image registration performance.

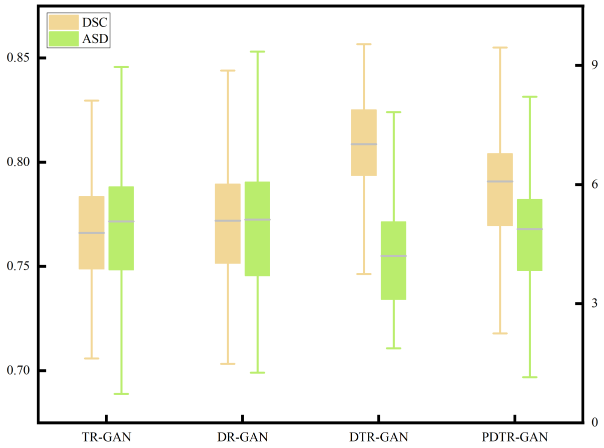

Figure 8. Boxplot of the distributions of the DSC and ASD of the MRI-CT registration on the Learn2Reg dataset.

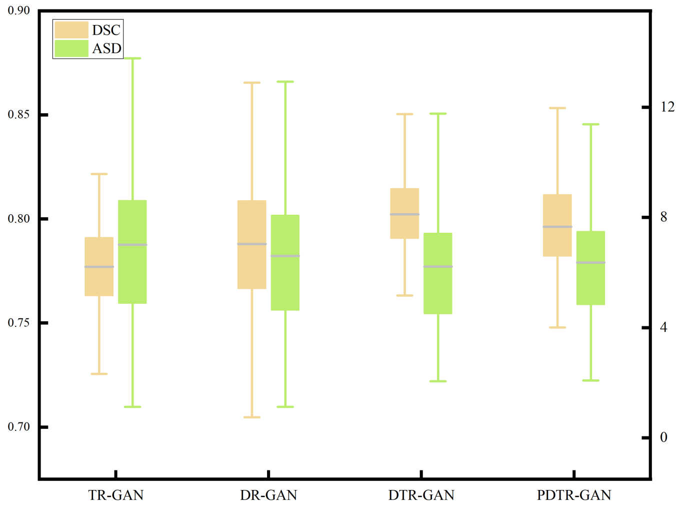

Figure 9. Boxplot of the distributions of the DSC and ASD of the T1-CT registration on the CHAOS dataset.

Tables 4 and 5 show the quantitative results in the enhancing registration performance on the Learn2Reg and CHAOS datasets, respectively. Compared with DR-GAN, DTR-GAN can capture more multiscale information and improve the registration performance, which is beneficial for generating consistent anatomical features during the image translation process. In addition, compared with PDTR-GAN, the DSC score of DTR-GAN is further improved, which demonstrates the effectiveness of maintaining appearance consistency in multimodal registration. Figures 10 and 11 show the visualization of the registration results of different methods. According to the results, our proposed method using a bidirectional translation network works better than using only a unidirectional network. The results in Figures 10 and 11 indicate that bidirectional translation networks can preserve a consistent shape to enhance the subsequent registration performance.

Table 4. Quantitative comparison of the proposed method on the Learn2Reg dataset. The best results are marked in bold for clarity.

	MRI-CT		**CT-MRI**	
	DSC **(%) ↑**	***ASD*** **(mm) ↓**	***DSC*** **(%) ↑**	***ASD*** **(mm) ↓**
TR-GAN	76.02 ± 2.68	4.91 ± 1.62	75.67 ± 2.64	5.10 ± 1.24
DR-GAN	77.03 ± 2.99	5.00 ± 1.88	76.30 ± 2.08	4.96 ± 1.94
DTR-GAN	**80.85 ± 2.09**	**4.21 ± 1.65**	**79.73 ± 2.03**	**4.62 ± 1.76**
PDTR-GAN	78.70 ± 2.70	4.79 ± 1.93	77.28 ± 2.07	4.87 ± 1.01

Table 5. Quantitative comparison of the proposed method on the CHAOS dataset.

	T1-CT		**CT-T1**	
	DSC **(%) ↑**	***ASD*** **(mm) ↓**	***DSC*** **(%) ↑**	***ASD*** **(mm) ↓**
TR-GAN	77.47 ± 2.09	6.83 ± 2.82	76.68 ± 2.91	6.93 ± 2.90
DR-GAN	78.80 ± 2.97	6.45 ± 2.63	77.40 ± 2.33	6.76 ± 2.06
DTR-GAN	**80.17 ± 1.77**	**6.07 ± 2.19**	**79.04 ± 2.40**	**6.28 ± 2.50**
PDTR-GAN	79.92 ± 2.28	6.23 ± 2.09	78.36 ± 2.72	6.40 ± 2.67

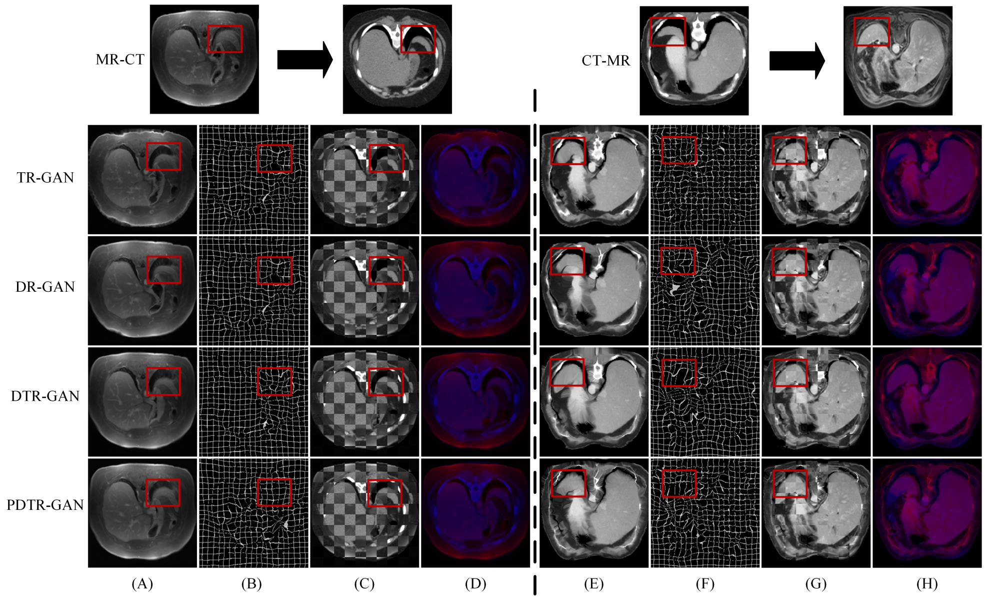

Figure 10. Comparison of our variant methods on the Learn2Reg dataset. (**A**,**E**) are warped images, (**B**,**F**) are deformation fields, (**C**,**G**) are checkboard grids, and (**D**,**H**) are overlapping images.

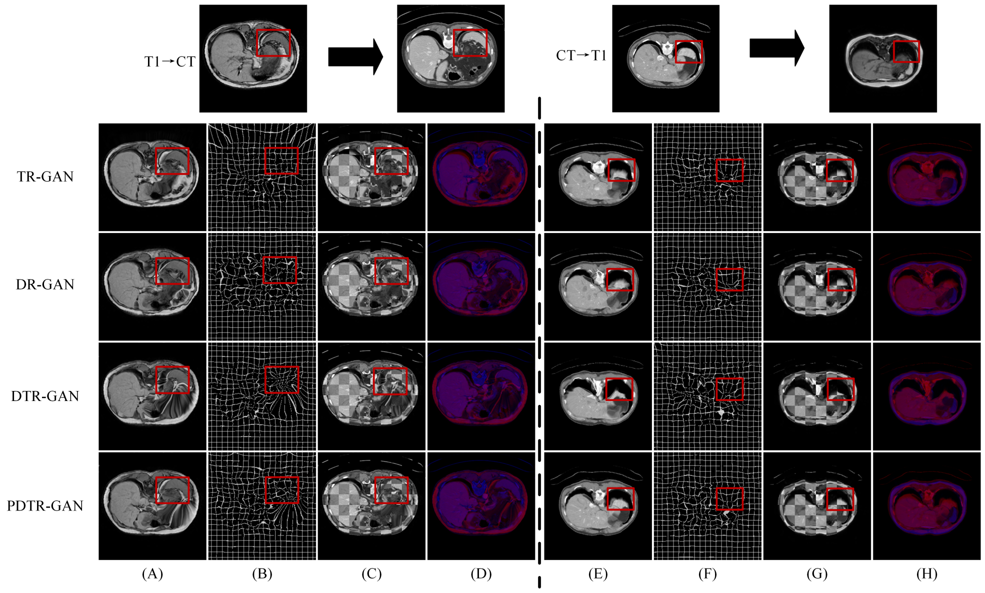

Figure 11. Comparison of our variant methods on the CHAOS dataset. (**A**,**E**) are warped images, (**B**,**F**) are deformation fields, (**C**,**G**) are checkboard grids, and (**D**,**H**) are overlapping images.

5. Discussion

During the process of image-to-image translation, it is common to lose detailed information and generate inconsistent features, which can potentially disrupt the performance of subsequent image registration. In this paper, we design DTR-GAN to achieve MRI-CT registration, which can generate consistent grayscale images and transform multimodal image registration into unimodal image registration.

To alleviate the cycle-consistency loss limitation of the CycleGAN, which would lead to the generation of distorted shapes, we improve the CycleGAN with PatchNCE loss to generate a shape-consistent transformed image, which enables the registration network to better complete alignment tasks. Specifically, the translation network is designed to implement bidirectional MRI-CT transformation, capturing and preserving more comprehensive anatomical information, which leads to more accurate translated images. During the encoding stage, the information obtained decreases layer by layer as the network layers deepen. Downsampling can result in the loss of low-level feature information that is extracted from the previous layers. To capture global and local features of the image in the translation network, we introduce the MFF module between the downsampling and upsampling layers in the translation network DT-GAN. By aggregating deep-level and top-level information, it leverages the quality of the translated image and improves the subsequent registration performance. In addition, we design a mixed similarity loss to calculate the similarity between warped and target images, which enhances the appearance of alignment in DTR-GAN.

One of the challenges in abdominal registration is the deformation at organ boundaries, which can be impacted by a variety of causes, such as large organ deformation, the filling and activity of adjacent organs, and respiratory movement. However, there are still some limitations in our approach. Our method is primarily designed for 2D registration and lacks the spatial structure information of the image. Hence, we will extend our network to 3D in the future. In addition, we will consider the decomposition of the target deformation field into several simpler ones in the next work, because it is hard to generate the accurate deformation field when the displacement between the images is large. Moreover, considering our proposed model has high computational complexity, we will focus on lightweight networks.

6. Conclusions

In this paper, an unsupervised multimodal image registration network, DTR-GAN, based on a bidirectional translation network, is designed for MRI-CT registration. The overall framework is based on two directional mapping translation networks to transform multimodal image registration into unimodal image registration. We incorporate the MFF module into the generator to effectively capture both global and local features during the image-to-image translation. By designing a mixed similarity loss, DTR-GAN is improved to focus on minimizing the dissimilarity in appearance between target and warped images and achieving high-quality MRI-CT registration. DTR-GAN shows its effectiveness and superiority in the experimental results compared with state-of-the-art methods on two abdominal MRI-CT datasets.

Author Contributions: T.Y.: Methodology, Supervision. A.Y.: Writing—original draft, Writing—review and editing. X.Z. (Xiang Zhao): Validation. X.Z. (Xin Zhang): Validation. Y.Y.: Formal analysis. C.J.: Visualization. All authors have read and agreed to the published version of this manuscript.

Funding: This paper was supported by the National Natural Science Foundation of China (Grant No. 62106067), the key specialized research and development program of Henan Province (Grant No. 202102210170), the Open Fund Project of the Key Laboratory of Grain Information Processing & Control (Grant No. KFJJ2021101), and the Innovative Funds Plan of the Henan University of Technology (Grant No. 2021ZKCJ14).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were used in this study. The data can be found at: https://learn2reg.grand-challenge.org/Learn2Reg2021/, https://chaos.grand-challenge.org/Combined_Healthy_Abdominal_Organ_Segmentation/.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Metcalfe, P.; Liney, G.P.; Holloway, L.; Walker, A.; Barton, M.; Delaney, G.P.; Vinod, S.; Tome, W. The Potential for an Enhanced Role for MRI in Radiation-Therapy Treatment Planning. *Technol. Cancer Res. Treat.* **2013**, *12*, 429–446. [CrossRef] [PubMed]
2. Johnstone, E.; Wyatt, J.J.; Henry, A.M.; Short, S.C.; Sebag-Montefiore, D.; Murray, L.; Kelly, C.G.; McCallum, H.M.; Speight, R. Systematic Review of Synthetic Computed Tomography Generation Methodologies for Use in Magnetic Resonance Imaging–Only Radiation Therapy. *Int. J. Radiat. Oncol. Biol. Phys.* **2018**, *100*, 199–217. [CrossRef] [PubMed]

3. Sharafudeen, M.; Chandra, S.S.V. Leveraging Vision Attention Transformers for Detection of Artificially Synthesized Dermoscopic Lesion Deepfakes Using Derm-CGAN. *Diagnostics* **2023**, *13*, 825. [CrossRef] [PubMed]
4. Commandeur, F.; Simon, A.; Mathieu, R.; Nassef, M.; Arango, J.D.; Roll, Y.; Haigron, P.; De Crevoisier, R.; Acosta, O. MRI to CT Prostate Registration for Improved Targeting in Cancer External Beam Radiotherapy. *IEEE J. Biomed. Health Inform.* **2016**, *21*, 1015–1026. [CrossRef]
5. Sun, Y.; Moelker, A.; Niessen, W.J.; van Walsum, T. Towards Robust CT-Ultrasound Registration Using Deep Learning Methods. In *Understanding and Interpreting Machine Learning in Medical Image Computing Applications, Proceedings of the First International Workshops, MLCN 2018, DLF 2018, and iMIMIC 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, 16–20 September 2018*; Springer International Publishing: Cham, Switzerland, 2018; pp. 43–51._5. [CrossRef]
6. Fan, J.; Cao, X.; Yap, P.T.; Shen, D. BIRNet: Brain image registration using dual-supervised fully convolutional networks. *Med. Image Anal.* **2019**, *54*, 193–206. [CrossRef] [PubMed]
7. De Vos, B.D.; Berendsen, F.F.; Viergever, M.A.; Staring, M.; Išgum, I. End-to-End Unsupervised Deformable Image Registration with a Convolutional Neural Network. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Proceedings of the Third International Workshop, DLMIA 2017, and 7th International Workshop, ML-CDS 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, 14 September 2017*; Springer International Publishing: Cham, Switzerland, 2017; pp. 204–212._24. [CrossRef]
8. Balakrishnan, G.; Zhao, A.; Sabuncu, M.R.; Guttag, J.; Dalca, A.V. VoxelMorph: A learning framework for deformable medical image registration. *IEEE Trans. Med. Imaging* **2019**, *38*, 1788–1800. [CrossRef]
9. Kori, A.; Krishnamurthi, G. Zero Shot Learning for Multi-Modal Real Time Image Registration. *arXiv* **2019**, arXiv: 1908.06213.
10. Yu, H.; Zhou, X.; Jiang, H.; Kang, H.; Wang, Z.; Hara, T.; Fujita, H. Learning 3D non-rigid deformation based on an unsupervised deep learning for PET/CT image registration. In *Medical Imaging 2019: Biomedical Applications in Molecular, Structural, and Functional Imaging*; SPIE: Bellingham, WA, USA, 2019; Volume 10953, pp. 439–444. [CrossRef]
11. Yan, P.; Xu, S.; Rastinehad, A.R.; Wood, B.J. Adversarial Image Registration with Application for MR and TRUS Image Fusion. In *Machine Learning in Medical Imaging, Proceedings of the 9th International Workshop, MLMI 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, 16 September 2018*; Springer International Publishing: Cham, Switzerland, 2018; pp. 197–204._23. [CrossRef]
12. Fan, J.; Cao, X.; Wang, Q.; Yap, P.T.; Shen, D. Adversarial learning for mono-or multi-modal registration. *Med. Image Anal.* **2019**, *58*, 101545. [CrossRef]
13. Mahapatra, D. GAN Based Medical Image Registration. *arXiv* **2018**, arXiv: 1805.02369.
14. Mahapatra, D.; Antony, B.; Sedai, S.; Garnavi, R. Deformable medical image registration using generative adversarial networks. In Proceedings of the 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), Washington, DC, USA, 4–7 April 2018; pp. 1449–1453 . [CrossRef]
15. Mirza, M.; Osindero, S. Conditional generative adversarial nets. *arXiv* **2014**, arXiv: 1411.1784.
16. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. In Proceedings of the IEEE International Conference on Computer vision, Venice, Italy, 22–29 October 2017; pp. 2223–2232. [CrossRef]
17. Xu, Z.; Luo, J.; Yan, J.; Pulya, R.; Li, X.; Wells, W.; Jagadeesan, J. Adversarial Uni- and Multi-modal Stream Networks for Multimodal Image Registration. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020, Proceedings of the 23rd International Conference, Lima, Peru, 4–8 October 2020*; Springer International Publishing: Cham, Switzerland, 2020 ; pp. 222–232._22. [CrossRef]
18. Tang, Z.; Yap, P.T.; Shen, D. A New Multi-Atlas Registration Framework for Multimodal Pathological Images Using Conventional Monomodal Normal Atlases. *IEEE Trans. Image Process.* **2018**, *28*, 2293–2304. [CrossRef] [PubMed]
19. Tanner, C.; Ozdemir, F.; Profanter, R.; Vishnevsky, V.; Konukoglu, E.; Goksel, O. Generative Adversarial Networks for MR-CT Deformable Image Registration. *arXiv* **2018**, arXiv: 1807.07349.
20. Han, R.; Jones, C.K.; Lee, J.; Wu, P.; Vagdargi, P.; Uneri, A.; Helm, P.A.; Luciano, M.; Anderson, W.S.; Siewerdsen, J.H. Deformable MR-CT image registration using an unsupervised, dual-channel network for neurosurgical guidance. *Med. Image Anal.* **2022**, *75*, 102292. [CrossRef]
21. Qin, C.; Shi, B.; Liao, R.; Mansi, T.; Rueckert, D.; Kamen, A. Unsupervised Deformable Registration for Multi-modal Images via Disentangled Representations. In Proceedings of the International Conference on Information Processing in Medical Imaging, Davis, CA, USA, 22 May 2019; Springer International Publishing: Cham, Switzerland, 2019; pp. 249–261._19. [CrossRef]
22. Wu, J.; Zhou, S. A Disentangled Representations based Unsupervised Deformable Framework for Cross-modality Image Registration. In Proceedings of the 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Mexico City, Mexico, 1 November 2021; pp. 3531–3534. [CrossRef]
23. Arar, M.; Ginger, Y.; Danon, D.; Bermano, A.H.; Cohen-Or, D. Unsupervised Multi-Modal Image Registration via Geometry Preserving Image-to-Image Translation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 13410–13419. [CrossRef]
24. Cao, X.; Yang, J.; Gao, Y.; Guo, Y.; Wu, G.; Shen, D. Dual-core steered non-rigid registration for multi-modal images via bi-directional image synthesis. *Med. Image Anal.* **2017**, *41*, 18–31. [CrossRef] [PubMed]
25. Casamitjana, A.; Mancini, M.; Iglesias, J.E. Synth-by-Reg (SbR): Contrastive Learning for Synthesis-Based Registration of Paired Images. In *Simulation and Synthesis in Medical Imaging, Proceedings of the 6th International Workshop, SASHIMI 2021, Held in*

Conjunction with MICCAI 2021, Strasbourg, France, 27 September 2021; Springer International Publishing: Cham, Switzerland, 2021; pp. 44–54._5. [CrossRef]

26. Chen, Z.; Wei, J.; Li, R. Unsupervised Multi-Modal Medical Image Registration via Discriminator-Free Image-to-Image Translation. *arXiv* **2022**, arXiv: 2204.13656. https://doi.org/10.24963/ijcai.2022/117.
27. Liu, Y.; Wang, W.; Li, Y.; Lai, H.; Huang, S.; Yang, X. Geometry-Consistent Adversarial Registration Model for Unsupervised Multi-Modal Medical Image Registration. *IEEE J. Biomed. Health Inform.* **2023**, *27*, 3455–3466 . [CrossRef]
28. Kong, L.; Qi, X.S.; Shen, Q.; Wang, J.; Zhang, J.; Hu, Y.; Zhou, Q. Indescribable Multi-modal Spatial Evaluator. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 17–21 June 2023; pp. 9853–9862.
29. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015, Proceedings of the 18th International Conference, Munich, Germany, 5–9 October 2015*; Springer International Publishing: Cham, Switzerland, 2015 ; pp. 234–241._28. [CrossRef]
30. Jaderberg, M.; Simonyan, K.; Zisserman, A. Spatial transformer networks. *Adv. Neural Inf. Process. Syst.* **2015**, *28*. [CrossRef]
31. Park, T.; Efros, A.A.; Zhang, R.; Zhu, J.Y. Contrastive Learning for Unpaired Image-to-Image Translation. In *Computer Vision–ECCV 2020, Proceedings of the 16th European Conference, Glasgow, UK, 23–28 August 2020;* Springer International Publishing: Cham, Switzerland, 2020 ; pp. 319–345._19. [CrossRef]
32. Han, J.; Shoeiby, M.; Petersson, L.; Armin, M.A. Dual Contrastive Learning for Unsupervised Image-to-Image Translation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 746–755. [CrossRef]
33. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-Image Translation with Conditional Adversarial Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1125–1134. [CrossRef]
34. Rudin, L.I.; Osher, S.; Fatemi, E. Nonlinear total variation based noise removal algorithms. *Phys. D Nonlinear Phenom.* **1992**, *60*, 259–268. [CrossRef]
35. Hering, A.; Hansen, L.; Mok, T.C.; Chung, A.C.; Siebert, H.; Häger, S.; Lange, A.; Kuckertz, S.; Heldmann, S.; Shao, W.; et al. Learn2Reg: Comprehensive Multi-Task Medical Image Registration Challenge, Dataset and Evaluation in the Era of Deep Learning. *IEEE Trans. Med. Imaging* **2022**, *42*, 697–712. [CrossRef]
36. Kavur, A.E.; Gezer, N.S.; Barış, M.; Aslan, S.; Conze, P.H.; Groza, V.; Pham, D.D.; Chatterjee, S.; Ernst, P.; Özkan, S.; Baydar, B. CHAOS Challenge-combined (CT-MR) healthy abdominal organ segmentation. *Med. Image Anal.* **2021**, *69*, 101950. [CrossRef]
37. Marstal, K.; Berendsen, F.; Staring, M.; Klein, S. SimpleElastix: A User-Friendly, Multi-lingual Library for Medical Image Registration. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Las Vegas, NV, USA, 27–30 June 2016; pp. 134–142. [CrossRef]
38. Dice, L.R. Measures of the Amount of Ecologic Association Between Species. *Ecology* **1945**, *26*, 297–302. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.