# NC2C-TransCycleGAN: Non-Contrast to Contrast-Enhanced CT Image Synthesis Using Transformer CycleGAN

Xiaoxue Hou[1] , Ruibo Liu[2] , Youzhi Zhang[1] , Xuerong Han[1] , Jiachuan He[2] , He Ma[1]*

[1] College of Medicine and Biological Information Engineering, Northeastern University, 110000 Shenyang, China
[2] Department of Radiology, The First Hospital of China Medical University, 110000 Shenyang, China

* Correspondence: mahe@bmie.neu.edu.cn

**Citation:** Hou, X. X., Liu, R. B., Zhang, Y. Z., Han, X. R., He, J. C., & Ma, H. (2023). NC2C-TransCycleGAN: Non-contrast to contrast-enhanced CT image synthesis using transformer CycleGAN. *Healthcraft Front.*, *2*(1), 34-45. https://doi.org/10.56578/hf020104.

**Abstract:** Background: Lung cancer poses a great threat to human life and health. Although the density differences between lesions and normal tissues shown on enhanced CT images is very helpful for doctors to characterize and detect lesions, contrast agents and radiation may cause harm to the health of patients with lung cancer. By learning the mapping relationship between plain CT image and enhanced CT image through deep learning methods, high quality synthetic CECT image results can be generated based on plain scan CT image. It has great potential to help save treatment time and cost of lung cancer patients, reduce radiation dose and expand the medical image dataset in the field of deep learning. Methods: In this study, plain and enhanced CT images of 71 lung cancer patients were retrospectively collected. The data from 58 lung cancer patients were randomly assigned to the training set, and the other 13 cases formed the test set. The Convolution Vison Transformer structure and PixelShuffle operation were combined with CycleGAN, respectively, to help generate clearer images. After random erasing, image scaling and flipping to enhance the training data, paired plain and enhanced CT slices of each patient are input into the network as input and labeled, respectively, for model training. Finally, the peak signal-to-noise ratio, structural similarity and mean square error are used to evaluate the image quality and similarity. Results: The performance of our proposed method is compared with CycleGAN and Pix2Pix on the test set, respectively. The results show that the SSIM value of the enhanced CT images generated by the proposed method improve by 2.00% and 1.39%, the PSNR values improve by 2.05% and 1.71%, and the MSE decreases by 12.50% and 8.53%, respectively, compared with Pix2Pix and CycleGAN. Conclusions: The experimental results show that the improved algorithm based on CylceGAN proposed in this paper can synthesize high-quality lung cancer synthetic enhanced CT images, which is helpful to expand the lung cancer image data set in the deep learning research. More importantly, this method can help lung cancer patients save medical treatment time and cost.

**Keywords:** Contrast-enhanced CT image synthesis; Cycle-Consistent adversarial networks improvement; Data augmentation; Deep learning

## 1. Introduction

Globally, cancer incidence and death are rising, with lung cancer being the most commonly diagnosed form of cancer (11.6% of the total cases) (Lahiri et al., 2023). The rapid development of computed tomography (CT) has been proven that it can significantly help the diagnosis of lung diseases (Bushara et al., 2023). CT can be divided into non-contrast CT(NCCT) and contrast-enhanced CT(CECT), and different types of CT have shown different advantages in various applications. CECT increases the density difference between lesions and normal tissues by injecting contrast media into blood vessels (Kojima et al., 2010). This helps doctors to understand the blood supply of the disease and the relationship between mediastinal lesions and cardiac macrovessels, thus improving the accuracy of differentiating benign from malignant lung diseases. However, the disadvantage of CECT is that it increases the scanning time for patients, the examination cost, and cannot be used in patients with contraindications to iodine contrast media. In addition, needle insertion into the human body during the use of contrast agents in

high-pressure syringes may cause discomfort (Wu et al., 2023) and may result in leakage of contrast agents or greater irritation to the patient's local skin. Instead, NCCT shows a low contrast between tumor areas and tissues, which is not conducive to the localization and qualitative diagnosis of lung lesions. The heart and surrounding tissues such as the chest wall, spine, and pulmonary vessels move several millimeters, producing a phenomenon of interleaved or staircase artifacts known as respiratory artifacts (Maier et al., 2021; Hertanto et al., 2012). Axial slices of NCCT from a lung cancer patient and CECT from an injected intravenous contrast agent are shown in Figure 1. Stripes, artifacts and motion artifacts are common phenomena in clinical work (Wang et al., 2023a). Post-processing features such as scan reconstruction parameters or ECG editing cannot eliminate this phenomenon, so the effect of artifacts on image quality cannot be ignored. In addition, the situation of blood supply in lung cancer patients is complicated on CT imaging, which makes the generation of CECT more difficult. With the wide application of Generative Adversarial Networks (GAN) in image generation (Skandarani et al., 2023), these problems are expected to be solved.
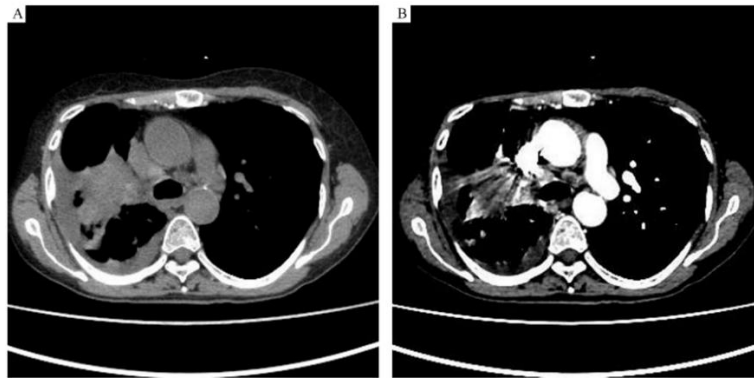


**Figure 1.** (A) Axial NCCT of lung cancer. (B) Axial CECT of lung cancer. During the image acquisition, due to the patient's autonomous or respiratory movement and other non autonomous movement, there are staggered layers between slice (A) and (B) and the generation of stripes and shadows in (B). NCCT, non-contrast computed tomography; CECT, contrast-enhanced CT computed tomography

In recent years GAN has been widely used in medical image tasks such as image segmentation (Beji et al., 2023; Dash et al., 2023; Skandarani et al., 2023; Zhong et al., 2023), lesion classification (Chen et al., 2023; Fan et al., 2023), and lesion detection (Esmaeili et al., 2023; Vyas & Rajendran, 2023). And the study of GAN in medical image synthesis tasks has dominated. The high-quality medical image data synthesised by GAN is now being validated by radiologists and can be used for radiological teaching or for big data deep network training (Kelkar et al., 2023). In addition, it has successfully remedied the problems of missing medical images and unbalanced data for classification and labeling due to difficulties in medical image data acquisition and involving patient privacy. CycleGAN is widely used for cross domain or cross-modality medical image synthesis due to its ability to process unpaired data. CycleGAN can be trained with unpaired data to simulate the implicit distribution of real data distribution to generate real images (Torbunov et al., 2023), which makes it very suitable for the interleaving problem caused by respiratory motion in this task. Chandrashekar et al. (2020) used CycleGAN to learn the relationship between soft tissue components to mimic contrast-enhanced CTA without contrast agents. It is also assumed that the raw data from non-contrast material CT contains enough information to distinguish blood and other soft tissue components. The accuracy of network output is assessed by comparing it with a contrast image. The test results show that the CTA generated from the non-contrast images are very similar to the ground truth. Agrawal et al. (2020) accomplished bidirectional exchange of content and style between two image modalities, CT and MR, on a pelvic dataset by means of an improved CycleGAN network. The validation results by radiologists showed that the subtle variations in MR and CT images generated by the improved network may not be identical to the real images, but can be used for medical purposes. In this work, it has been shown that cycleGAN can be used for medical image translation tasks.Many unpaired medical image synthesis tasks can achieve good results by using CycleGAN, but most of these methods are completed by CNN. Although CNN can well extract the local information in the feature map, such as image edge, textture and rich context semantic information, the receptive field of CNN is limited. In addition, the transpose convolution used in the up-sampling portion of the CycleGAN generator results in a checkerboard artifacts that greatly reduces the quality of the generated image.

The Convolution Vision Transformer structure merges the advantages of Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs). Convolutional Neural Networks (CNNs) are recognised for their efficiency in processing local features through their convolutional layers, while Vision Transformers (ViTs) excel at capturing global dependencies in an image through self-attention mechanisms (Maurício et al., 2023). The PixelShuffle operation, also known as sub-pixel convolution, is a technique mainly used for upscaling images in

super-resolution tasks (Wang et al., 2023b). The combination of Convolutional Vision Transformer structures and PixelShuffle operations in CycleGANs is a powerful tool for advanced image processing tasks. This approach leverages the strengths of convolutional operations, transformer models, and efficient upscaling to significantly enhance the quality and effectiveness of generated images.

In this paper, we propose the NC2C-TransCycleGAN, which integrates the CVT (Komorowski et al., 2023) structure that can perceive long distances range. In addition, the PixelShufflfle operation (Sun et al., 2023) is added into the CycleGAN to reduce the serious impact of the checkerboard artifacts caused by transposed convolution operation on the image generation quality. The advantage of NC2C-TransCycleGAN is that it can learn the complex information of lung cancer enhanced CT images in a specific way without manually annotating the lung cancer regions, generate more accurate enhanced regions, and then synthesize CECT images of lung cancer patients and avoid the appearance of checkerboard artifacts.

## 2. Materials and Methods

### 2.1 Materials

This study retrospectively collected CT images of lung cancer patients who underwent unenhanced and enhanced examinations between February 2017 and October 2020. The patient inclusion criteria were: (I) patients with lung cancer requiring lung CT enhancement imaging; (II) the CT sequence of lung cancer has its corresponding complete imaging data, clinical data and paired plain and enhanced chest images; (III) contraindication for iodinated contrast medium (CM); (IV) Imaging of patients with initial diagnosis of lung cancer on preoperative CT who had not undergone previous lung cancer surgery and other treatments. Exclusion criteria were as follows: (I) CT imaging sequences or clinical data of the lungs were incomplete, and paired plain and enhanced CT sequences were missing; (II) CT images with severe motion artifacts or other artifacts such as poor image quality affect the analysis of the model and experimental results; (III) known severe allergy to iodinated CM injection; (IV) patients who have undergone treatment such as tumor resection.

Table 1 shows that CT images in the dataset have different tube voltage (in kVp), tube current (in mAS), volume CT dose index (in mGy), and slice scan thickness parameters when acquired. The data set used in this study contained image data of 71 lung cancer patients, including 44 men and 27 women, accounting for 58% and 42% of the total data, respectively. The age of the patients ranged from 42 to 86 years with a mean age of 59.85 +8.67 years. The number of patients aged 42-50 accounted for 11% of the total data, of which 34% were 50-60 years old, 41% were 60-70 years old, and 14% were over 70 years old. Thirty-one percent of patients had metastasis, including pancreas, liver, kidney, brain, bone, chest wall, adrenal gland, lung, etc. Each patient in the dataset has two types of CT images, axial unenhanced CT and corresponding enhanced CT images. These images are all from two different CT manufacturers, Toshiba and Siemens. The data coverage of this experiment is comprehensive and diverse.

The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the First Hospital of China Medical University, and individual consent for this retrospective analysis was waived.

**Table 1.** Acquisition parameters of NCCT and CECT

| Parameters | NCCT | CECT |
|---|---|---|
| CTDIvol (mGy) | 15.27(3.87-22.40) | 16.01(3.75-25.31) |
| Voltage (kVp) | 90-120 | 90-120 |
| Tubecurrent (mA) | 111-671 | 111-743 |
| WindowCenter (HU) | 35-40 | 40-80 |
| WindowWidth (HU) | 180-400 | 180-400 |

Note: CTDIvol, volume CT dose index; NCCT, non-contrast computed tomography; CECT, contrast-enhanced CT computed tomography.

### 2.2 Methods

In this paper, we improve CycleGAN and propose NC2C-TransCycleGAN for the virtual lung cancer enhanced CT image synthesis task. The overall network structure of NC2C-TransCycleGAN continues the dual structure of vanilla CycleGAN in Figure 2. The generator consists of an encoder, a transformer, and a decoder. Generator GCECT or FNCCT attempts to make the generated image as similar as possible to the samples in the underlying real domain of the image. The discriminators DCECT and DNCCT retain the PatchGAN structure from CycleGAN, which outputs a 30*30 image to discriminate whether the input image is from a fake image or a real image generated by the generator.
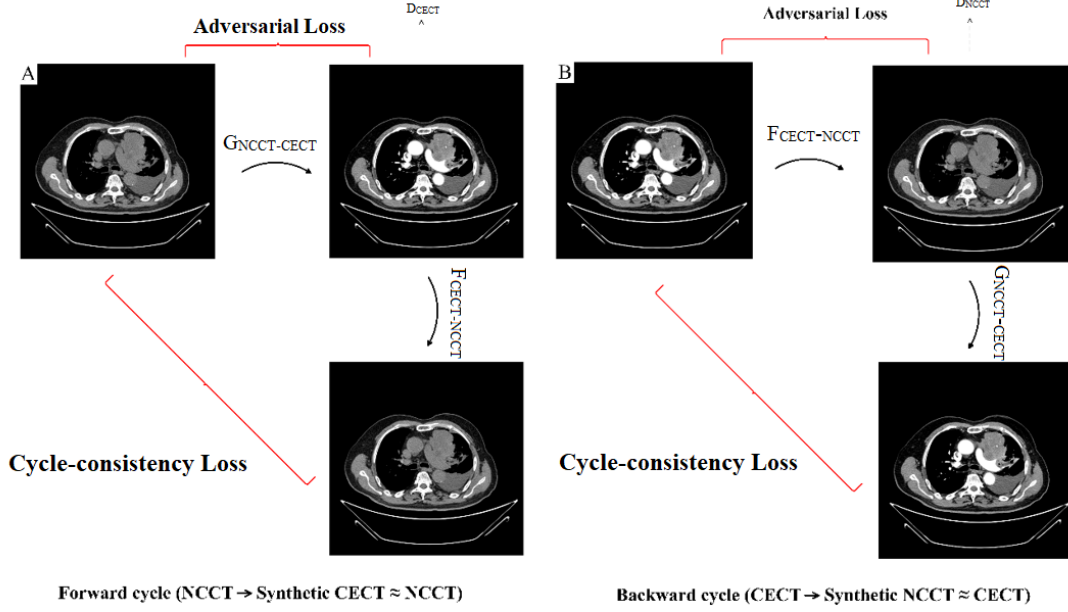
**Figure 2.** (A) forward cycle-consistency loss of CycleGAN in our task: NCCT to synthetic CECT to NCCT. (B) back cycle-consistency loss: CECT to synthetic NCCT to CECT

The NC2C-TransCycleGAN is still consistent with the cycle structure of CycleGAN, but its internal structure has been adjusted and updated to make it applicable to more complex medical image generation tasks.

2.1.1 Image preprocessing

Each patient in the dataset had CT images in both modalities. The two modalities were axial non-enhanced and corresponding enhanced CT images, respectively. Due to the influence of respiratory movement, the layers of these paired data will be separated from each other, and the location, size and shape of tumors are very different, which brings challenges to the task of medical image synthesis. We randomly selected 3253 pairs of data from 50 patients as the training set, 663 pairs of images from 8 patients as the validiation set and 813 pairs of data from 13 patients as the test set. The data of each patient were divided into multiple slices according to the axial scan direction. These two-dimensional slices were used to train the model. The 512*512 resolution downsampled from the original NCCT and CECT images in this experiment is 256*256 resolution. The CT images of lung cancer for each patient contained an average of more than 60 slices of 2-dimensional axial images. Based on Eq. (1) on the pixel values of NCCT and CECT, resampled to a uniform distribution in the 256 gray scale range. And data enhancement operations such as horizontal flipping are performed. After that, we normalize the original image and keep the information of the original image as much as possible. The normalization operation smoothes the distribution of the input layer, helps stochastic gradient descent, and helps the network converge quickly during the training process. In order to make the model training more generalized and robust, this experiment expands the training data by online data augmentation operation. Data augmentation methods are widely used in deep learning and provide an effective means to overcome the overfitting phenomenon caused by the scarcity of training data for the model. Random erasing and random flipping are selected in this experiment. Random erasing refers to randomly selecting a part of a rectangular region in an image and erasing the pixels in that region using random values. Random flipping means the operation of flipping the image by a certain angle along the horizontal or vertical direction. After preprocessing, the enhanced and non-enhanced CT images of lung cancer are input into the network in pairs for training. And real enhanced CT image of the same patient will be used as the ground truth data for network.

$$Image = \left( \frac{I - min(I)}{max(I) - min(I)} \right) * 255 \tag{1}$$

2.2.2 NC2C-TransCycleGAN architecture

The encoder of the generator in NC2C-TransCycleGAN is composed of three convolutional layers. The input lung cancer plain CT image is down sampled through these convolution layers, so that the number of channels of the feature map is expanded and the size is gradually reduced.

The encoder output is then passed sequentially to four residual blocks, two stage of CvT and three residual blocks in the converter module, which is used to convert the image domain between the two images, each of which

is composed of two convolutional layers. The purpose of the residual structure is to add part of the image features obtained from the input data after feature extraction to the output by means of residuals by introducing a constant mapping jump connection, which ensures that the information of the feature map of the previous layer can also be used in the later feature layers, thus better preserving the semantic information of the image and reducing the deviation between the output of the later layers of the network and the original input image content, which can both supplement information while avoiding gradient disappearance or gradient explosion when the network is deeper. After four residual blocks are extracted to the key features in the input CT image, two stage structure in CvT are connected to help obtain the global information.

A single stage of CvT is shown in (A) of Figure 3. The stage consists of two parts, the Convolutional Token Embedding layer and the Convolutional Transformer Block. The feature map with the shape of (N, C, H, W) is used as the input of the Convolutional Token Embedding layer. The details are as follows: the first step is to downsample the feature map after the input passes through a two-dimensional convolution layer, and the second step is to transform the two-dimensional feature map into a sequence shape (N, H*W, C) by reshape operation for layer normalization and input into the Transformer Token Embedding module, where the length of the sequence becomes shorter. Similar to convolution, without padding, the size of the feature map decreases after downsampling. The length of the sequence generated by the signature graph after the Convolutional Token Embedding module is reduced to simulate the convolution down-sampling operation. Unlike convolution, the Convolutional Token Embedding layer not only extracts local information from the image, but also long-range information with the participation of sequence. The sequence vector is reshaped to get a feature map, and the reshape after padding and convolution operations is the previous sequence shape, resulting in Q, K, V, respectively. Q is obtained when stride=1, K is obtained when step size is increased to 2, V is obtained separately. This can be formulated as:

$$x_i^{q/k/v} = Flatten\left(Conv2d\left(Reshape2D\left(x_i\right),s\right)\right) \qquad (2)$$

$x_i^{q/k/v}$ is the token input for Q/K/V matrices at layer i. $x_i$ is the original token before Convolution Projection. S is the size of convolution kernel size. Conv2d refers to a deep-wise separable convolution. (B) shows this process, which is called Convolutional Projection. Due to the change of step size, K and V are shortened after reshape. As shown in (C), when they enter the Multi-Head Self-Attention module and MLP, the number of neurons and the algorithm of network model will be greatly reduced. In summary, CVT not only guarantees the dynamic attention and global context of Transformer, but also preserves the invariance of translation, scaling and distortion in convolution neural networks. So we introduced CVT into our research work, taking out the individual stages in the network and fusing them into CycleGAN. And the head parameter of multi-head attention (MHA) is set to 2, which helps our model to focus on different aspects of information and capture abundant features. The number of perations of the Convolutional Transformer Block is also set to 2 in our model. Based on the global information of CvT, three residual blocks are used for further feature extraction. This part is called transformer module, which can transform the image from one domain to another.
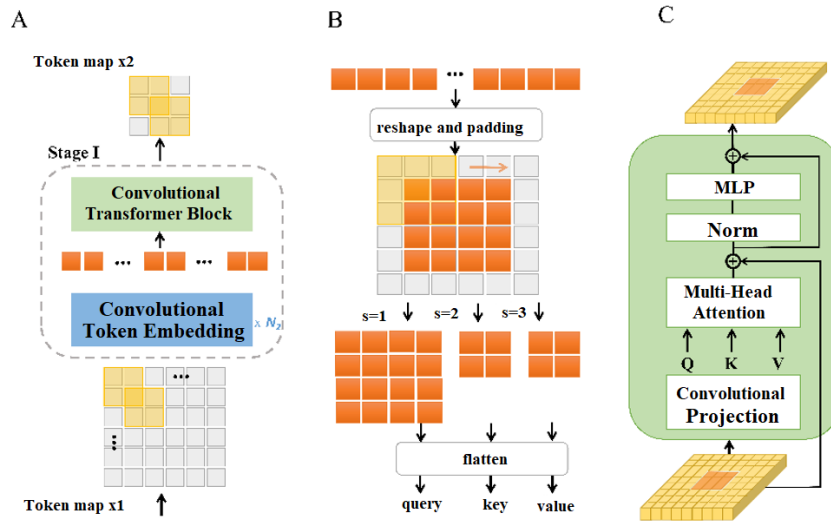


**Figure 3.** (A) one of the stages of the hierarchical multi-stagestructure facilitated by the Convolutional Token Embedding layer. (B) Convolutional projection with stride = 1 and stride = 2 to generate query(Q), key(K), value(V). (C) details of the Convolutional Transformer Block, where the first layer is Convolution Projection NC2C-TransCycleGAN optimizes the decoder of the vanilla CycleGAN

The decoder is similar to the inverse process of the encoder. The vanilla CycleGAN completes the upsampling operation using the transposed convolution operation shown in Figure 4, with the aim of expanding the feature map size while reducing the low-level features from the feature vector, and thus obtaining the synthetic CT generated by the network. However, since the size of the convolution kernel represented by the blue box is 3, it is not divisible by the step size of 2. Therefore, it can be seen in the feature map of the input of this figure that the region slid by the blue convolution kernel will have the uneven overlapping trajectories indicated in green in the figure, which is the checkerboard artifacts caused by transposed convolution. The actual performance of checkerboard artifacts on the synthetic enhanced CT task is shown in Figure 4. Although checkerboard artifacts can theoretically be avoided by learning the weights, they are still an insurmountable problem in the actual image reconstruction process. However, since transposed convolution has the property of learning parameters and makes a valuable contribution to the final result, NC2C-TransCycleGAN uses PixelShuffle operation in another branch parallel to transposed convolution. The PixShuffle operation, also known as sub-pixel convolution, is an up-sampling method suitable for image super-resolution tasks. It converts a low-resolution feature map into a high-resolution feature map by using ordinary convolution coupled with a pixel shuffle operation. To further enhance image recovery, the Channel Unit (CA) module and the Positional Unit (POS) module are often introduced when using the PixShuffle operation. Unlike the traditional direct interpolation method, the PixelShuffle operation is based on a low-resolution feature image of size H*W*C, convolved to obtain a feature image with r2 channels, where r denotes the magnification factor, and finally a high-resolution feature map of rH*rW*C is reconstructed by periodic shuffling. The output result generated by this branch of PixelShuffle in NC2C-TransCycleGAN performs concat operation with the result generated by the transposed convolution branch to generate the final synthetic CT image. Figure 5 shows the implementation flow of the NC2C-TransCycleGAN algorithm.
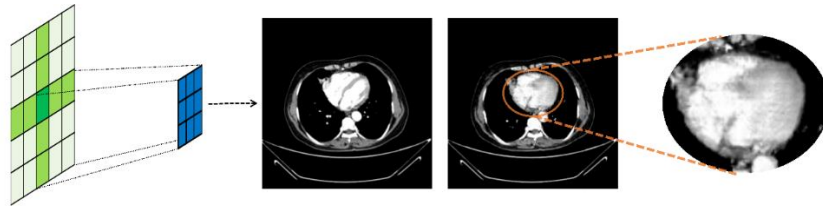


**Figure 4.** The principle of checkerboard artifacts in two-dimensional image
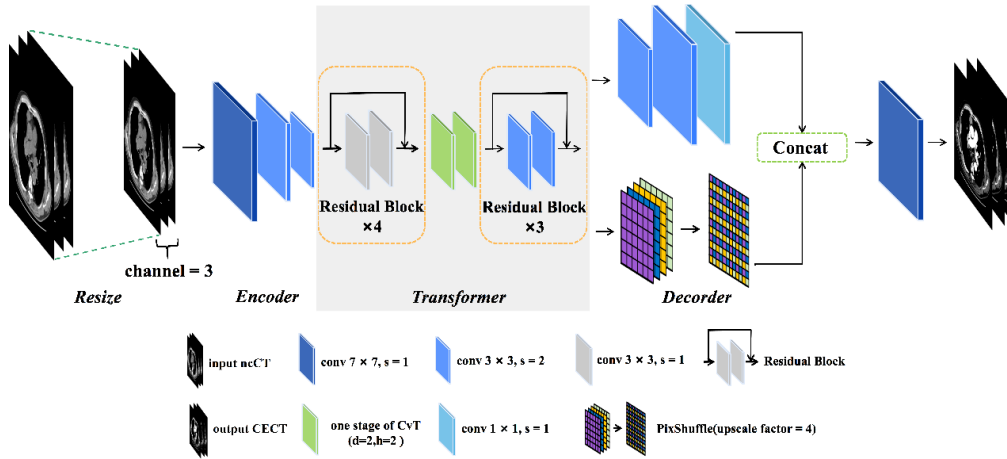


**Figure 5.** The pipeline of the Generator network architecture of proposed NC2C-TransCycleGAN
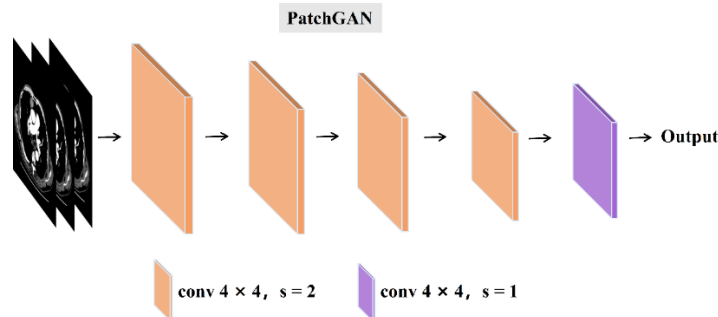


**Figure 6.** The pipeline of the Discriminator network architecture of proposed NC2C-TransCycleGAN

The discriminator is a PatchGAN network consisting of five convolution layers. The image of the input discriminator passes through multiple convolution layers to produce a 30*30 probability matrix output, where each convolution layer is composed of convolution, instanceNorm, and LeakyReLU operations in turn. each element of the probability matrix has a true probability for a 70*70 area of the original input NCCT image. This method is more efficient than trying to view the entire input. The discriminator network structure is shown in the Figure 6.

The experimental environment for this study is configured as follows: the graphics card model is NVIDIA Tesla P100; the video memory is 16 GB; CUDA version is 11.3; the operating system is Linux; the development language and version is Python 3.9; the development framework and version is PyTorch 1.8.1. In the training process of the experiment, we set the hyperparameters of the network as follows: for the learning rate scheduler, we use ReduceLROnPlateau in Pytorch to adaptively adjust the learning rate, which can detect the dynamics of the specified indicators within a specified number of times, and automatically decay the learning rate when the loss curve is no longer falling or the accuracy indicator is no longer rising, allowing the network to obtain better learning results, where the main parameters corresponding to the adaptive method. The initial learning rate of ReduceLROnPlateau is set to 0.005, and the patience is 7, the learning rate reduction coefficient is 0.01, batchsize is set to 1, the beta value of the Adam optimizer is set to 0.5, and the weighing parameter λ1 is set to 0.5. The loss function for network training uses the same loss function used for training in the vanilla CycleGAN.

### 2.2.3 Evaluation method

The CT image synthesis task can essentially be viewed as an image processing problem, so the quality of images generated by GAN networks can be measured using commonly used image quality evaluation criteria. In order to objectively compare the image synthesis effects of different networks, this study will introduce three evaluation metrics, PSNR (peak signal-to-noise ratio), SSIM (structural similarity), and MSE (mean squared error), which are commonly used internationally (Prodan et al., 2023) to quantitatively evaluate and compare the performance of three networks, NC2C-TransCycleGAN, original CycleGAN, and Pix2Pix. These indicators are widely used in the evaluation of medical image synthesis. The MSE formula is as follows, where sCECT represents the virtual image synthesized by the network:

$$MSE = \frac{1}{N} \sum_{i=1}^{N} \left( CECT(i) - sCECT(i) \right)^2 \tag{3}$$

where, $N$ denotes the total number of pixels in the image area, and i is the index of the aligned pixels in the image area.

PSNR is a commonly used index for evaluating image restoration. It is used to represent the ratio between the maximum signal of the image and the background noise. The larger the value, the better the image quality. The PSNR is defined as:

$$PSNR = 10 \cdot log_{10} \left( \frac{MAX^2}{MSE} \right) \tag{4}$$

MAX represents the maximum pixel value of ground truth CECT and sCECT images.

SSIM measures image similarity from three aspects: brightness, contrast, and structure. It is the most commonly used index to measure the performance of image reconstruction. The closer the SSIM value is to 1.0, the closer the generated enhanced CT is to the real enhanced CT and its formula can be expressed as:

$$SSIM = \frac{\left( 2\mu_{CECT}\mu_{sCECT} + c_1 \right)\left( 2\sigma_{CECT \cdot sCECT} + c_2 \right)}{\left( \mu_{CECT}^2 + \mu_{sCECT}^2 + c_1 \right)\left( \sigma_{CECT}^2 + \sigma_{sCECT}^2 + c_2 \right)} \tag{5}$$

## 3. Results

We compare the NC2C-TransCycleGAN with Pix2Pix and CycleGAN, which perform well in image generation. The average value of all test slice indexes is calculated, and the result of the average value is shown in Table 2. It can be seen from the results that our proposed method is higher than other methods in these three indicators. Among them, the SSIM value of the image generated by our method is 2.00% and 1.39% improved by the Pix2Pix and CycleGAN, and the PSNR value is 2.05% and 1.71% better than that of the Pix2Pix and CycleGAN, respectively. Compared with Pix2Pix and CycleGAN, our method improved MSE value by 12.50% and 8.53% respectively. This is due to the existence of cycle loss, which reduces CycleGAN's requirements for data pairing. Although each pair of data in our data set comes from the same patient, the layer-to-layer mismatch caused by the

breathing movement causes the layer-to-layer correspondence to be incomplete. This will have a great impact on the results of the Pix2Pix network. Because our method introduces the CvT structure, the network can also focus on more global information while extracting features, making the prediction of tumor blood supply more accurate than CycleGAN. Compared with CycleGAN using transposed convolution as upsampling, PixShuffle method makes a great contribution to the clarity of the generated image. Figure 7 is helpful to observe the improvement of checkerboard artifacts by our network, and we compared different lung locations in patients with lung cancer. We compare the effects of four groups of lung tomography, where each group contains tomographic images of the same positon from two patients with lung cancer. The first and second rows in (A) are CT images of the same location layer in two patients, as are (B), (C), and (D). The first column in the figure is NNCT, the second column is the real CECT, the third column is the result of synthetic CECT generated by Pix2Pix, the fourth column is the result generated by CycleGAN, and the last column is the result of NC2C-TransCycleGAN. The enlarged area and whole CT slice show that the better effect of our method is not accidental, but has good wide applicability. As can be seen in the images of the first patients of (A) and (C), the virtual CECT images generated by our network have an enhanced effect on the presence of a blood supply location that is closer to the ground truth. From the images of the first patient in (B) and (D), it can be seen that our proposed method can effectively reduce the influence of chessboard artifacts at cardiac tomography and other locations. The shape, image contrast and the course of the blood vessel section generated from the first patient of (A) and the second patient of (B), (C) and (D) are also closer to the reality, and pay great attention to the generation of details.

**Table 2.** Comparison of test metrics results of three models

| Evaluation Criteria | Pix2Pix | CycleGAN | NC2C-TransCycleGAN |
|---|---|---|---|
| PSNR(dB) | 19.9046 | 19.9718 | 20.3130 |
| SSIM | 0.8378 | 0.8429 | 0.8546 |
| MSE | 734.9404 | 703.0354 | 643.0944 |

Note: CycleGAN: Cycle-Consistent Adversarial Networks. NC2C-TransCycleGAN: non-contrast to contrast-enhanced CT image synthesis using transformer CycleGAN. PSNR: peak signal-to-noise ratio. SSIM: structural similarity. MSE: mean squared error.

In addition, the data related to the experiment were statistically analyzed using SPSS (https://www.ibm.com/cn-zh/analytics/spss-statistics-software) statistical software (Windows version 22.0) in this paper. As appropriate, chi-square test or univariate analysis of variance was used to compare the age, sex, smoking, cancer type, metastasis, lesion location, and image thickness of lung cancer patients between the training set and the test set. No significant difference was found in patient feature between the training and test datasets (p=0.081-0.598, p >0.05). Detailed results are given in Table 3. The Kruskal-Wallis test was used to compare the evaluation indexes of the three groups model CVT, CycleGAN, Pix2Pix, and the independent samples Ttest was used to compare the results of model indexes between each two groups. The statistical results in Table 4 show that the synthetic CECT results generated by the three models are very significantly and statistically different (<0.001).

**Table 3.** Patient characteristics and statistics

| Variable | $\chi^2$ | Male Number |
|---|---|---|
| Age(years) | / | 0.551 |
| Sex | 0.356 | 0.213 |
| Type | 1.386 | 0.500 |
| Metastatic | 0.495 | 0.482 |
| Position | 0.278 | 0.598 |
| Thickness | 3.054 | 0.081 |

Note: $\chi^2$: chi-square.

**Table 4.** Statistical differences about TransCycleGAN, Cyclegan and Pix2Pix method

| Model | PSNR *p*-value | MSE *p*-value | SSIM *p*-value |
|---|---|---|---|
| NC2C-TransCycleGAN CycleGAN Pix2Pix | 2.46E - 43 | 2.46E - 43 | 8.84E - 48 |
| NC2C-TransCycleGAN CycleGAN | 5.31E - 16 | 9.57E - 14 | 5.64E - 26 |
| NC2C-TransCycleGAN Pix2Pix | 1.34E - 32 | 6.10E - 10 | 3.84E - 11 |

Note: CycleGAN: Cycle-Consistent Adversarial Networks. NC2C-TransCycleGAN: non-contrast to contrast-enhanced CT image synthesis using transformer CycleGAN. PSNR: peak signal-to-noise ratio. SSIM: structural similarity. MSE: mean squared error.
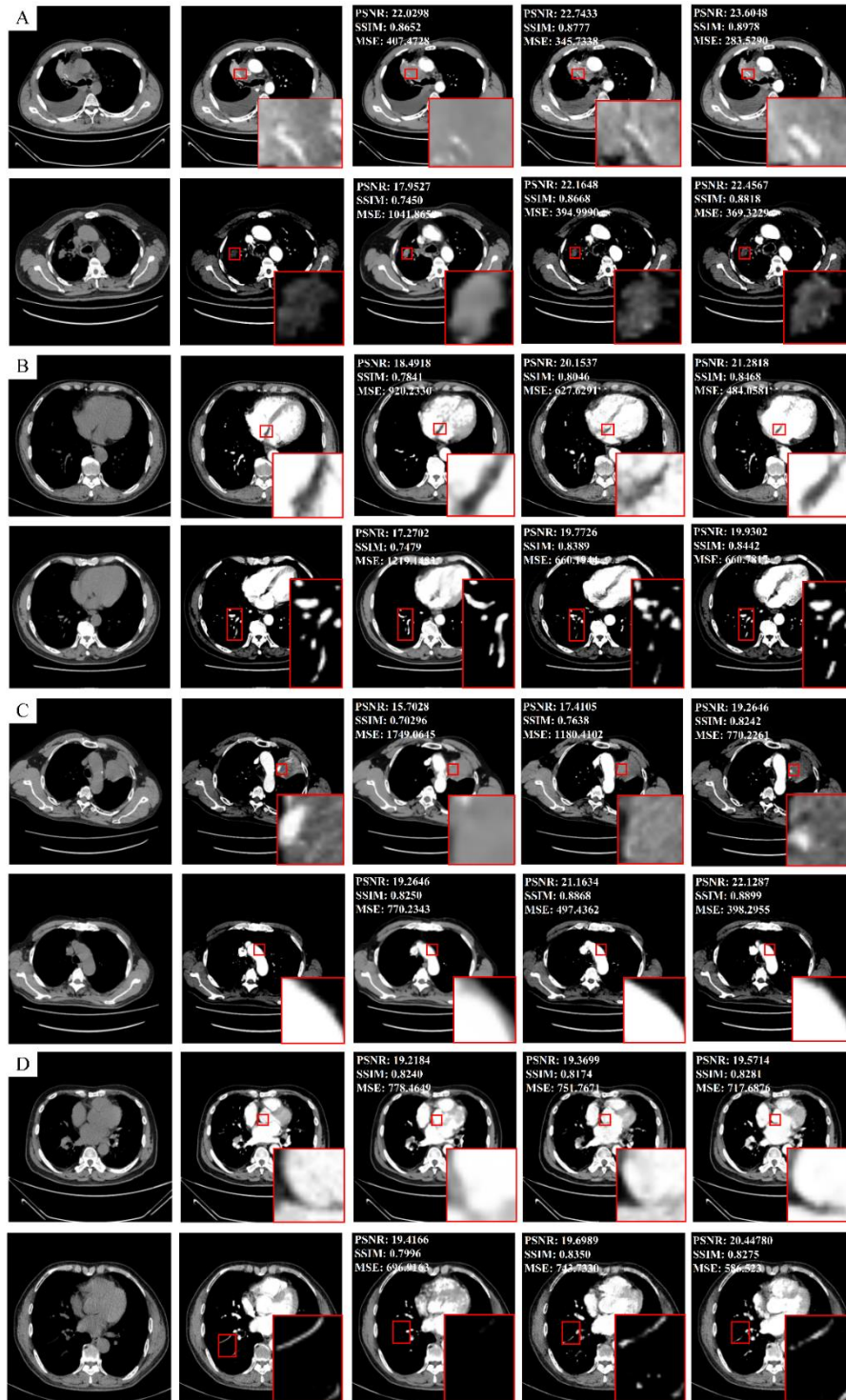
**Figure 7.** The synthetic CT results at four different positions produced by the three models

## 4. Discussion

The complexity of the blood supply around the tumor requires the generation of models with strong learning ability. To explore the capability of NC2C-TransCycleGAN in generating image details, we used the difference images as a reference in Figure 8 and selected a more diverse and complex CECT layer of lung cancer patients. In Figure 8, the first row of (A) shows the results of CECT, Pix2Pix-generated images, CycleGAN, and NC2C-TransCycleGAN, respectively. The first column of the second row of (A)shows the NCCT images, and the remaining columns show the synthetic CECT generated by Pix2Pix, CycleGAN, and NC2C-TransCycleGAN, respectively and the difference images between the real CECT. From the visualization of the generated results and the results marked by the red magnified area, it can be seen that for the same locations of the CT images, our

network control generates details and judgments about the presence or absence of the enhanced area are more accurate than Pix2Pix and CycleGAN. As in difference image of (A), our network accurately generates the burr signs of the lesions, which shows that our network has a strong ability to extract the features of the lesions during training. In difference image of (B) and (C), even the liquefied face, our network can generate an result closer to that of a true image liquefied surface and more similar contrast, and is more adept at generating details of the image. In difference image of (D), the shape of the bronchial bifurcation and its surroundings, the generation results of our network are closer to the real images than those of other networks. In summary, NC2C-TransCycleGAN can produce higher quality mediastinum, blood vessels, lesions, liquefaction surfaces and other areas, which shows that our method has greater potential to help doctors diagnose patients or medical dataset augmentation.
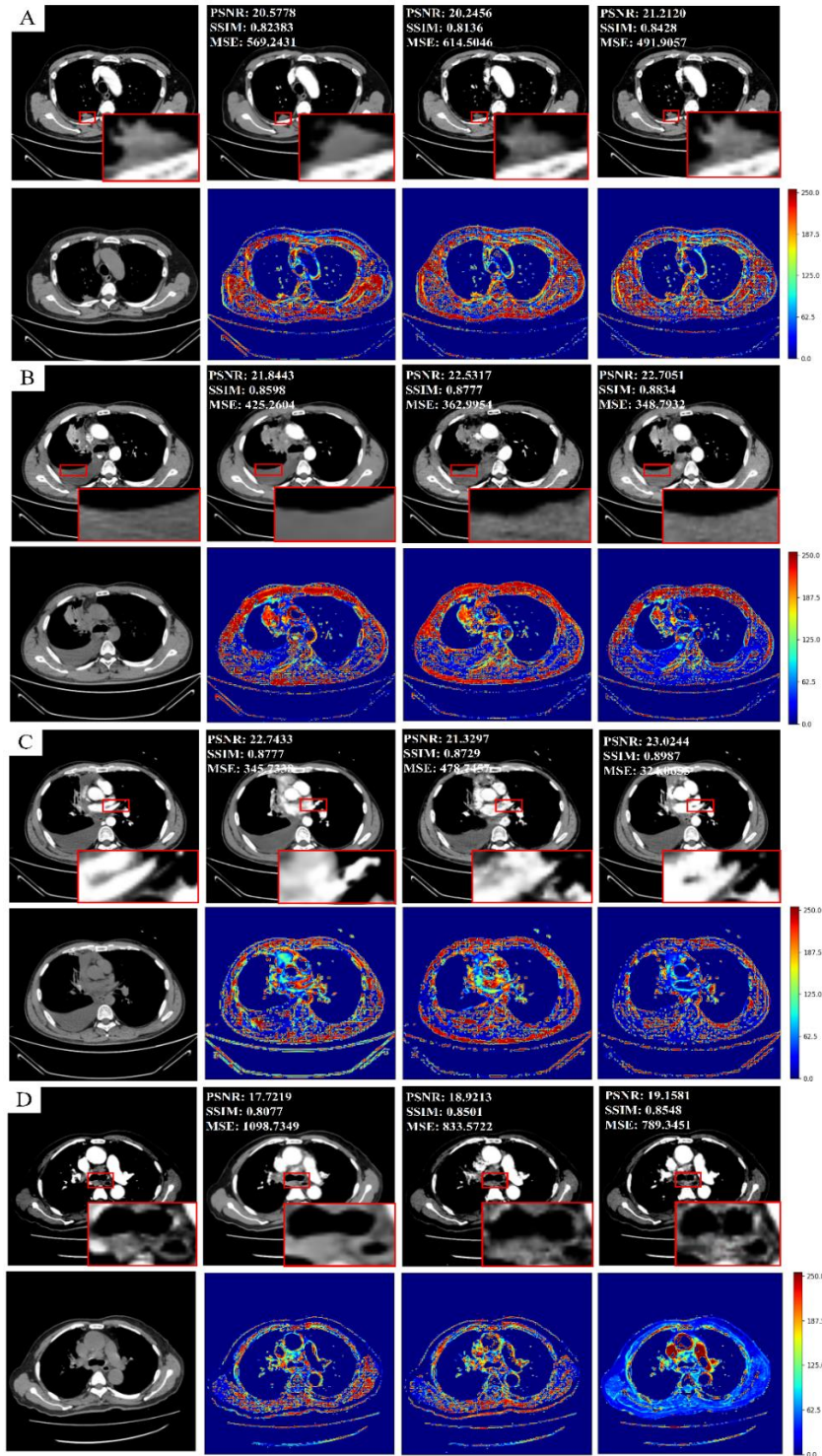


**Figure 8.** The synthetic CT results at four different positions produced by the three models

In conclusion, although the resulting enhanced CT images show differences in subtle details from the ground truth enhanced CT images, the comparison results show that by improving CycleGAN, our proposed NC2C-TransCycleGAN network has better image quality and potential to accomplish this task. The results of the models in this study can assist in the training of deep learning models, such as pretraining or data enhancement techniques. In the future, our method will expand to generate enhanced 3D CT images and be used as an extended synthetic reality training dataset for lung cancer detection to compensate for the lack of data in the real image distribution. The conversion of low-dose CT to normal-dose flat-scan CT can also be accomplished with the help of deep learning methods, thus helping to reduce the radiation dose ingested into the patient's body.

## Ethical Approval

The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the First Hospital of China Medical University, and individual consent for this retrospective analysis was waived.

## Data Availability

The data used to support the research findings are available from the corresponding author upon request.

## Conflicts of Interest

All authors have completed the ICMJE uniform disclosure form. The authors have no other conflicts of interest to declare.

## References

Agrawal, V., Kori, A., Anand, V. K., & Krishnamurthi, G. (2020). Structurally aware bidirectional unpaired image to image translation between CT and MR. *arXiv preprint*, arXiv:2006.03374. https://doi.org/10.48550/arXiv.2006.03374.

Beji, A., Blaiech, A. G., Said, M., Abdallah, A. B., & Bedoui, M. H. (2023). An innovative medical image synthesis based on dual GAN deep neural networks for improved segmentation quality. *Appl. Intell., 53*(3), 3381-3397. https://doi.org/10.1007/s10489-022-03612-6.

Bushara, A. R., Vinod Kumar, R. S., & Kumar, S. S. (2023). LCD-capsule network for the detection and classification of lung cancer on computed tomography images. *Multimed. Tools Appl., 82*(24), 37573-37592. https://doi.org/10.1007/s11042-023-14893-1.

Chandrashekar, A., Shivakumar, N., Lapolla, P., Handa, A., Grau, V., & Lee, R. (2020). A deep learning approach to generate contrast-enhanced computerised tomography angiograms without the use of intravenous contrast agents. *Eur. Heart J., 41*(Supplement_2), ehaa946-0156. https://doi.org/10.1093/ehjci/ehaa946.0156.

Chen, Y., Lin, Y., Xu, X., et al. (2023). Multi-domain medical image translation generation for lung image classification based on generative adversarial networks. *Comput. Methods Programs Biomed., 229,* 107200. https://doi.org/10.1016/j.cmpb.2023.107200.

Dash, A., Ye, J., & Wang, G. (2023). A review of Generative Adversarial Networks (GANs) and its applications in a wide variety of disciplines: From medical to remote sensing. *IEEE Access, 12,* 18330-18357. https://doi.org/10.1109/ACCESS.2023.3346273.

Esmaeili, M., Toosi, A., Roshanpoor, A., Changizi, V., Ghazisaeedi, M., Rahmim, A., & Sabokrou, M. (2023). Generative adversarial networks for anomaly detection in biomedical imaging: A study on seven medical image datasets. *IEEE Access, 11*, 17906-17921. https://doi.org/10.1109/ACCESS.2023.3244741.

Fan, M., Huang, G., Lou, J., Gao, X., Zeng, T., & Li, L. (2023). Cross-parametric generative adversarial network-based magnetic resonance image feature synthesis for breast lesion classification. *IEEE J. Biomed. Health Inform., 27*(11), 5495–5505. https://doi.org/10.1109/JBHI.2023.3311021.

Hertanto, A., Zhang, Q., Hu, Y. C., Dzyubak, O., Rimner, A., & Mageras, G. S. (2012). Reduction of irregular breathing artifacts in respiration-correlated CT images using a respiratory motion model. *Med. Phys., 39*(6Part1), 3070-3079. https://doi.org/10.1118/1.4711802.

Kelkar, V. A., Gotsis, D. S., Brooks, F. J., Prabhat, K. C., Myers, K. J., Zeng, R., & Anastasio, M. A. (2023). Assessing the ability of generative adversarial networks to learn canonical medical image statistics. *IEEE Trans. Med. Imaging., 42*(6), 1799–1808. https://doi.org/10.1109/TMI.2023.3241454.

Kojima, C., Umeda, Y., Ogawa, M., Harada, A., Magata, Y., & Kono, K. (2010). X-ray computed tomography contrast agents prepared by seeded growth of gold nanoparticles in PEGylated dendrimer. *Nanotechnology, 21*(24), 245104. https://doi.org/10.1088/0957-4484/21/24/245104.

Komorowski, P., Baniecki, H., & Biecek, P. (2023). Towards evaluating explanations of vision transformers for medical imaging. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, *Vancouver, BC, Canada*, 3725-3731.

Lahiri, A., Maji, A., & Potdar, P. D. (2023). Lung cancer immunotherapy: Progress, pitfalls, and promises. *Mol. Cancer, 22*(1), 40. https://doi.org/10.1186/s12943-023-01740-y.

Maier, J., Lebedev, S., Erath, J., Eulig, E., Sawall, S., Fournié, E., Stierstorfer, K., Lell, M., & Kachelrieß, M. (2021). Deep learning-based coronary artery motion estimation and compensation for short-scan cardiac CT. *Med. Phys., 48*(7), 3559-3571. https://doi.org/10.1002/mp.14927.

Maurício, J., Domingues, I., & Bernardino, J. (2023). Comparing vision transformers and convolutional neural networks for image classification: A literature review. *Appl. Sci., 13*(9), 5521. https://doi.org/10.3390/app13095521.

Prodan, M., Vlăsceanu, G. V., & Boiangiu, C. A. (2023). Comprehensive evaluation of metrics for image resemblance. *J. Inf. Syst. Oper. Manag., 17*(1), 161-185.

Skandarani, Y., Jodoin, P. M., & Lalande, A. (2023). Gans for medical image synthesis: An empirical study. *J. Imaging, 9*(3), 69. https://doi.org/10.3390/jimaging9030069.

Sun, B., Zhang, Y., Jiang, S., & Fu, Y. (2023). Hybrid pixel-unshuffled network for lightweight image super-resolution. *In Proceedings of the AAAI Conference on Artificial Intelligence, Washington, DC, USA, 37*(2), 2375–2383. https://doi.org/10.1609/aaai.v37i2.25333.

Torbunov, D., Huang, Y., Yu, H., et al. (2023). Uvcgan: Unet vision transformer cycle-consistent gan for unpaired image-to-image translation. *In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, Hawaii, USA,* 702-712.

Vyas, B. & Rajendran, R. M. (2023). Generative adversarial networks for anomaly detection in medical images. *Int. J. Multidiscipl. Innov. Res. Methodol., 2*(4), 52-58.

Wang, C. J., Rost, N. S., & Golland, P. (2023a). Spatial-intensity transforms for medical image-to-image translation. *IEEE Trans. Med. Imaging., 42*(11), 3362–3373. https://doi.org/10.1109/TMI.2023.3283948.

Wang, D., Zhuang, L., Gao, L., Sun, X., Huang, M., & Plaza, A. J. (2023b). PDBSNet: Pixel-shuffle down-sampling blind-spot reconstruction network for hyperspectral anomaly detection. *IEEE Trans. Geosci. Remote Sens. 61*, 1-14. https://doi.org/10.1109/TGRS.2023.3276175.

Wu, L., Tang, S., Chen, W., Chen, X., Zhang, L., & He, L. (2023). Study of low kV, low contrast agent dosage, and low contrast agent flow rate scan in computed tomography angiography of children's liver. *IJ Radiology, 20*(4), e138586. https://doi.org/10.5812/ijradiol-138586.

Zhong, G., Ding, W., Chen, L., Wang, Y., & Yu, Y. F. (2023). Multi-scale attention generative adversarial network for medical image enhancement. *IEEE Trans. Emerg. Top. Comput. Intell., 7*(4), 1113–1125. https://doi.org/10.1109/TETCI.2023.3243920.