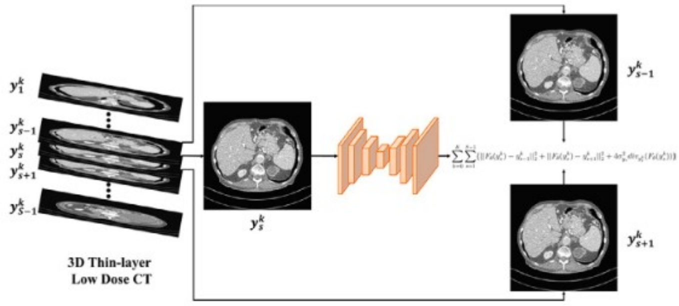
论文笔记 2022/03/21-03/27

1. Noise2Context: Context-assisted learning 3D thin-layer for low-dose CT
   https://aapm.onlinelibrary.wiley.com/doi/epdf/10.1002/mp.15119?saml_referrer
   作者：Zhicheng Zhang，Lei Xing
   期刊名：MP
   LDCT 无 clean data



They use the adjacent two slices as the supervision. Under certain assumptions, the unsupervised problem is equivalent to having the ground truth as the supervision.

For LDCT image denoising, the basic mathematical model can be formulated as: $y = x + n$, $y \in R^{M\times N}$ is the noisy LDCT image, $x \in R^{M\times N}$ is the clean CT image, NDCT, $n \in R^{M\times N}$ denotes the noise, where M and N denote the row and column, respectively. To obtain clean NDCT image from y, the regular solution is to train a neural network F with parameters θ according to the l_2 loss as Equation (1):

$$\mathcal{L}_\theta = \sum_{s=0}^{M\times N} \left\| F_\theta(y_s) - x_s \right\|_2^2 \quad (1)$$

To obtain a high-performance denoising neural network F according to Equation (1), numerous high-quality paired LDCT images and their NDCT counter-

2.2 | Noise2Context

To introduce N2C, some notations should be determined first. y_s^k, y_{s-1}^k and y_{s+1}^k are three adjacent LDCT images of the k^{th} patient, x_s^k, x_{s-1}^k and x_{s+1}^k are their clean counterparts, n_s^k, n_{s-1}^k and n_{s+1}^k are corresponding noise, s is the index of LDCT image. We first formulate the following Equation (2):

$$\theta = \operatorname{argmin}_\theta \sum_{k=0}^{K} \sum_{s=1}^{S-1} \left\{ \left\| F_\theta\left(y_s^k\right) - y_{s-1}^k \right\|_2^2 + \left\| F_\theta\left(y_s^k\right) - y_{s+1}^k \right\|_2^2 \right\} \quad (2)$$

By introducing auxiliary variables, x_s^k, and ignoring θ – irrelevant terms, the Equation (2) can be inferred to Equation (3).

$$\theta = \operatorname*{argmin}_\theta \sum_{k=0}^{K} \sum_{s=1}^{S-1} \left\{ 2\left\| F_\theta\left(y_s^k\right) - x_s^k \right\|_2^2 + 2(2x_s^k - y_{s-1}^k - y_{s+1}^k)^T F_\theta\left(y_s^k\right) \right\} \quad (3)$$

parts are necessary, which provide a powerful driving force to supervised learning. However, data acquisition is nontrivial to obtain, which limits the application scenario of neural network-based methods. In this

From Equation (3), we can see that the first term is the supervised l_2 loss. Since $y_s^k = x_s^k + n_s^k$, Equation (3) can be rewritten into Equation (4):

$$\theta = \operatorname*{argmin}_\theta \sum_{k=0}^{K} \sum_{s=1}^{S-1} \left\{ 2\left\| F_\theta\left(y_s^k\right) - x_s^k \right\|_2^2 + 2\left(2y_s^k - y_{s-1}^k - y_{s+1}^k\right)^T F_\theta\left(y_s^k\right) - 4(n_s^k)^T F_\theta\left(y_s^k\right) \right\} \quad (4)$$

work, without clean NDCT images as supervision, we focus on 3D thin-layer LDCT in an unsupervised manner by making full of the similarity between adjacent 3D LDCT slices. With some reasonable assumptions, we can employ an unsupervised loss function to train

There is a strong similarity between adjacent CT slices in 3D thin-layer LDCT as long as the thickness and spacing are small enough. So $2y_s^k$ can be closed to $y_{s-1}^k + y_{s+1}^k$, thus the second term in Equation (4) is about 0. Assume that the noise is subject to the zero-mean Gaussian distribution, the Equation (4) can be rewritten into[53]:

Here, σ is the standard deviation of the noise that can be estimated in advance.[54] As a consequence, we can obtain Equation (6). From Equation (6), we can see that if we use the adjacent two slices as the supervision. Under certain assumptions, the unsupervised problem is equivalent to having the ground truth as the supervision. Here, $div_{y_s^k}(F_\theta(y_s^k))$ can be estimated using Monte Carlo method.[53]

For evaluation, we employed two different data sites. The first is the Mayo testing dataset including 1 patient case of 526 2D CT images of 1 mm slice thickness due to the existed reference NDCT. To further test the generalizability and robustness in a real case, a pig head was scanned using our in-house CBCT system at both the low/normal dose levels. After obtaining the 2D

$$\mathcal{L}_{N2C} = \sum_{k=0}^{K} \sum_{s=1}^{S-1} \left\{ \left\| F_\theta\left(y_s^k\right) - y_{s-1}^k \right\|_2^2 + \left\| F_\theta\left(y_s^k\right) - y_{s+1}^k \right\|_2^2 + 4\sigma_{y_s}^2 div_{y_s^k}\left(F_\theta\left(y_s^k\right)\right) \right\} \quad (6)$$

2. Low-dose CT reconstruction by self-supervised learning in the projection domain
   https://arxiv.org/pdf/2203.06824.pdf
   和上一篇方法类似，只是在投影域做

应该是要投 MICCAI2022 的

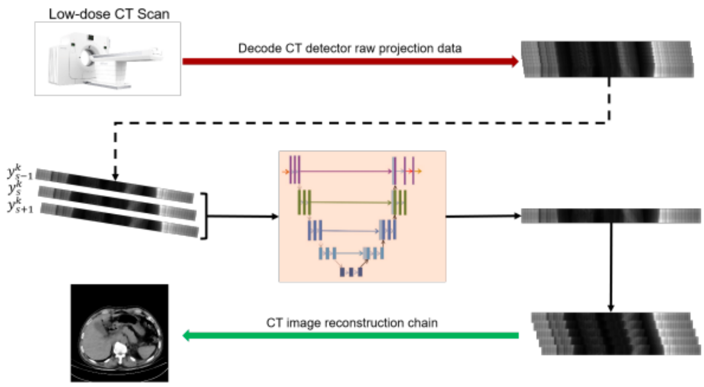


Fig. 1. The flowchart of the proposed method.To begin, we get the raw data from the clinical LDCT scan and decode it into projection data. The noise derived projection data is then obtained by randomly cropping three neighbouring angle projection data (y^k_{s-1},y^k_s and y^k_{s+1}) and feeding it into our model. Finally, CT images were reconstructed from the denoised projection data.

For LDCT denoising in the projection domain, the basic mathematical model can be formulated as: $y = x + n$, y is the noisy LDCT projection image, x is the corresponding clear reference, n denotes the noise. To obtain clean LDCT projection image from y, a neural network F_θ can be trained with object function as in Eq.1:

$$\theta = \arg\min_\theta \sum ||F_\theta(y) - x||_2^2 \tag{1}$$

In our model, we use the projection data of three adjacent angles (y^k_{s-1},y^k_s and y^k_{s+1}) as the input to the model and estimate the projection data of the intermediate angles. Since there is no corresponding clean projection images, we use a self-supervised learning method to train our model. The objective function of our network is Eq. 2.

$$\theta = \arg\min_\theta \sum_{k=0}^{K} \sum_{s=1}^{S-1} ||F_\theta(y^k_s) - y^k_{s-1}||_2^2 + ||F_\theta(y^k_s) - y^k_{s+1}||_2^2 + ||F_\theta(y^k_s) - y^k_s||_2^2 \tag{2}$$

Where y^k_s, y^k_{s-1}, and y^k_{s+1} are three adjacent projection data of the kth patient, s is the index of projection image.

Similar to the [22], we make the following assumptions: (a) the differences in projection data of adjacent angles are minor; (b) the noise in projection data of distinct projection angles is independent of each other and zero mean. Under those assumption, $||F_\theta(y^k_s) - y^k_{s-1}||_2^2 + ||F_\theta(y^k_s) - y^k_{s+1}||_2^2$ is equal to $2 * ||F_\theta(y^k_s) - x^k_s||_2^2$. As a consequence, Eq. 2 and Eq. 1 are equivalent.

We can observe from Eq.2 that if we employ the adjacent two slices and itself as the supervision, the unsupervised problem is equal to having the ground truth as the supervision under specific assumptions.

Because our method is a training strategy rather than an architectures per se, it may be used to any acceptable neural network backbone. In this work, to implement our Noise2Projection, we use standard U-net [17] with 32 basic feature-maps as the primary neural network, F, and Eq. 2 as the loss function.

假设：

(a) the differences in projection data of adjacent angles are minor;

(b) the noise in projection data of distinct projection angles is independent of each other and zero mean.