World Academy of Science, Engineering and Technology
International Journal of Biomedical and Biological Engineering
Vol:17, No:10, 2023

Heterogenous Dimensional Super Resolution of 3D CT Scans Using Transformers

Helen Zhang

***Abstract*—**Accurate segmentation of the airways from CT scans is crucial for early diagnosis of lung cancer. However, the existing airway segmentation algorithms often rely on thin-slice CT scans, which can be inconvenient and costly. This paper presents a set of machine learning-based 3D super-resolution algorithms along heterogenous dimensions to improve the resolution of thicker CT scans to reduce the reliance on thin-slice scans. To evaluate the efficacy of the super-resolution algorithms, quantitative assessments using PSNR (Peak Signal to Noise Ratio) and SSIM (Structural SIMilarity index) were performed. The impact of super-resolution on airway segmentation accuracy is also studied. The proposed approach has the potential to make airway segmentation more accessible and affordable, thereby facilitating early diagnosis and treatment of lung cancer.

***Keywords*—**3D super-resolution, airway segmentation, thin-slice CT scans, machine learning.

I. Introduction

SEGMENTATION of the airway tree using thoracic Computed Tomography (CT) is an important step for navigated bronchoscopy for lung nodule biopsy. Since a large percentage of the lung nodules are in the peripheral of the lung, it is important to be able to segment small airways. The bronchial tree is a complex 3D structure with various branches of different sizes and orientations, making accurate airway segmentation a challenging task.

Airway segmentation algorithms typically rely on "thin-slice" CT scans, with a slice thickness of 1 mm or less. However, in some cases, patients may not have a thin-slice CT scan due to non-ideal CT protocols or prior scans not intended for navigation. In such scenarios, ordering a new CT scan can be both inconvenient and costly and incurs additional radiation to the patient. This paper addresses this issue by reducing the requirement for thin-slice CT scans. As developing new airway segmentation models for thick-slice CT scans can be costly, the goal of this paper is to develop algorithms to improve the 3D CT scan resolution and investigate the feasibility of creating reasonable segmentations using CT scans with slice spacing of 2.5 mm or more. Specifically, a set of machine learning-based 3D super-resolution algorithms that can transform thick-slice CTs (e.g. 1.5 mm, 2 mm, 2.5 mm, 3 mm, 5 mm) into thin-slice CTs (e.g. 0.625 mm) are developed to virtually reduce the slice thickness and improve bronchoscopic airway segmentation.

Helen Zhang is with Thomas Jefferson High School for Science and Technology, Alexandria, VA 22312, USA. (e-mail: hzhang.1675@gmail.com).

II. Methodology

A. Literature Review

There are several approaches used for medical image super resolution (SR), including interpolation-based methods, reconstruction-based methods and machine learning-based methods.

The machine learning-based medical image SR approaches use various techniques, such as deep convolutional neural networks (CNNs), to learn the mapping between low-resolution and high-resolution images [2], [3], [5], [6]. Reference [2] presents a 3D CNN for reducing the slice thickness of CT images, and evaluates its impact on the reproducibility of radiomic features used in lung cancer diagnosis. Reference [5] proposed a Residual CNN-based method while [6] employs a 3D fully convolutional neural network (FCN) for improving image resolution using paired CT scans. These algorithms can effectively improve the Z-resolution of CT images and the reproducibility of radiomic features [4].

Another popular approach is using Generative Adversarial Networks (GANs) which consist of two networks, a generator and a discriminator [1], [7]. For example, a novel GAN-based approach called GAN-CIRCLE [7] is developed to enhance image quality by combining identical, residual, and cycle learning techniques.

Overall, machine learning-based medical image SR approaches have shown to be very effective in producing high-quality images, but they require a large dataset of high-resolution images for training. Additionally, the performance of these methods can be affected by the quality and size of the training dataset, as well as the specific architecture of the CNN or GAN used.

In this work, we adopted a transformer-based network, i.e., Swin Transformer for Image Restoration (Swin-IR) [8], for super-resolution of CT slices. The Swin-IR network models long-range context with better computational efficiency and has achieved state-of-the-art results in various 2D image restoration tasks. We adapted the Swin-IR network to perform SR in axial or coronal views of CT scans to improve the Z-resolution. Additionally, we extended the network architecture to utilize 3D convolution kernel, making it suitable for 3D CT scans and enabling 3D super-resolution. The effectiveness of our SR models was demonstrated through quantitative evaluations using PSNR and SSIM, as well as qualitative inspections on 3D airway segmentation.

B. Network Architectures and Approaches

Our goal is to develop methods that can take a thick-slice CT (such as 1.5 mm, 2 mm, 2.5 mm, 3 mm, or 5 mm) and convert it into a thin-slice CT (such as 0.625 mm) with the aim of improving airway segmentation. The module will take as input a CT volume and its metadata, as well as the target slice spacing. The output will be a CT volume with thinner slice spacing, without obvious hallucinated artifacts and with smoothness across slices.

1) Designs of 3D CT Super-Resolution

The 3D CT scan data can be processed and projected into different views, including the axial, coronal, and sagittal planes, as shown in Fig. 1. The task of CT super-resolution is unique in that only Z-dimension needs to be up-sampled.

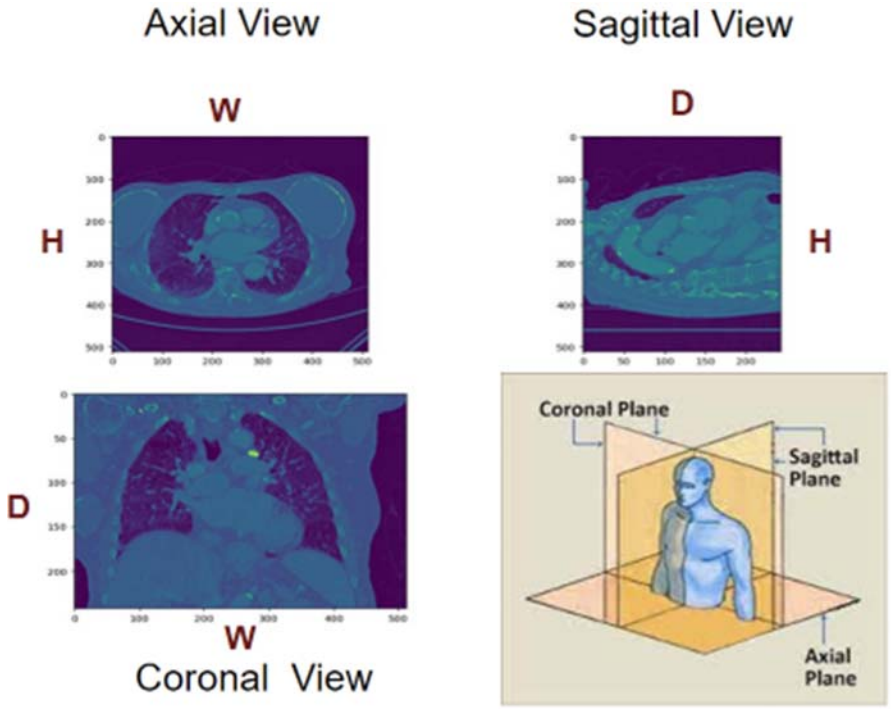

Fig. 1 The 3D CT data is sliced in three orthogonal directions to create 2D axial, sagittal, and coronal views respectively

In the context of airway segmentation, the most effective approach is to use the coronal view and perform up-sampling in the Z direction (which corresponds to the Y direction in the 2D image plane). As a result, there are multiple ways to approach 3D CT super-resolution (Fig. 2):

(a) 2D Approach: In 2D coronal (or sagittal) view, up-sample in both superior (X, left-right) and inferior (Y, up-down) dimensions.
(b) 1D Approach: In 2D coronal (or sagittal) view, only up-sample in inferior (Y) dimension.
(c) True-3D Approach: Re-scale directly in 3D volume data using 3D convolution.
(d) Pseudo-3D Approach: Re-scale directly in 3D volume data using 2D convolution.

The majority of the state-of-the-art super-resolution deep networks employ the up-sampling mechanism that increases both the width (X) and height (Y) dimensions (or superior and inferior axes) of a 2D image. While this 2D approach is straightforward and can be easily adapted for our application, it often results in unnecessary up-sampling along the width dimension, as depicted in Fig. 2 (a).

The 1D approach performs up-sampling in the appropriate inferior (Y) dimension, as shown in Fig. 2 (b), but implementing this approach requires modification of existing super-resolution networks.

Both 1D and 2D approaches require projecting 3D volume CT data into 2D coronal (or sagittal) planes to generate training images. A single 3D scan can generate a multitude of 2D image frames. For example, 512 slices of 2D images can be generated for training purposes if the image resolution is 512 x 512 in the axial view. However, they are not true 3D SR approaches.

The majority of network architectures used for common image restoration tasks utilize 2D convolution kernels in their implementations, represented by a 4-dimensional tensor [B, C, H, W], where B is the batch size, C is the channel, and H and W are width and height respectively. To achieve super-resolution in all 3 dimensions, it is necessary to add the depth dimension and create a 5-D tensor [B, C, D, H, W] to perform 3D convolution. This is the true-3D approach, as shown in Fig. 2 (d). Understandably, the true-3D implementation is more complex with longer training times.

If we treat the channel dimension as the depth D dimension, we can simplify network adaptation for 3D convolution. For example, we can set the channel value of the input layer to C, and value for the output layer to 4C, we can achieve 4 times up-sampling in the D-dimension. This is the pseudo-3D approach.

However, in this approach, the convolution kernel is not really scanned in the d-dimension, it only process WxHxC blocks at a time, as shown in Fig. 2 (c). Although this approach may result in artifacts along the borders across blocks, it is still worth exploring since it can utilize most existing implementations of 2D SR networks and takes less training time.

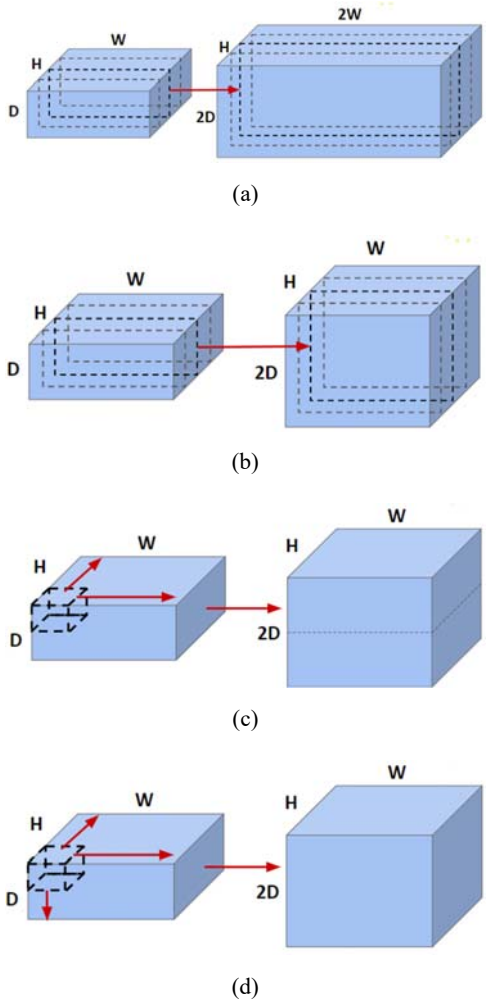

Fig. 2 Strategies for 3D CT SR: (a) 2D, (b) 1D, (c) Pseudo-3D, and (d) True 3D; a scale factor of 2 is used for illustration

In this paper, we adapted and extended the state-of-the-art super-resolution network architectures and built the pipeline for all the four approaches. Specifically, we trained and evaluated the 1D, pseudo-3D and true-3D approaches, measuring their performance using PSNR and SSIM metrics. Through this evaluation, we gained valuable insight into how each approach performs in improving airway segmentation.

2) Baseline Architecture: Swin-IR

Our super-resolution network architectures for the 1D, 2D, and pseudo-3D approaches are adapted from Swin-IR [8]. Swin-IR is a transformer-based network designed for image restoration and SR. Similar to traditional transformers, it is capable of modeling long-range context, but with improved computational efficiency. The network divides large images into non-overlapping patches, which allows for interactions across patches using shifted windows. The architecture is based on a coarse-to-fine approach to capture both low-level and high-level features and interactions. Swin-IR has achieved state-of-the-art results in various image restoration tasks.

Assuming that the original CT image has dimensions W, H, and D along the X, Y, and Z directions, respectively, the architecture of the 2D approach we used for our work is the same as that of the traditional Swin-IR. For each coronal view, we create an image of size W x D, and the input tensor had dimensions [B, 1, D, W], where the number 1 represents the single-channel grayscale image for one slice. In our case, both the width W and height H of the image are set to 512 pixels. The number of slices, denoted by D, usually range from 50 to 200. B is the batch size. The output tensor for the 2D super-resolution process is [B, 1, s*D, s*W], where s is the scale factor. In our case, we set s = 4 to reduce the CT scan thickness from 2.5 mm to 0.625 mm.

In the 1D approach, the output tensor has dimensions [B, 1, s*D, W]. It is important to note that the up-sampling process only occurs in the D dimension. Therefore, adjustments must be made to the dimensions of the last few layers of the Swin-IR network to reflect this change.

In the pseudo-3D approach, we form an input tensor of [B, d, H, W] dimensions, where d is the number of slices determined by the available computer memory capacity. In our experiments, we set d = 16 and 64. The output tensor has dimensions of [B, s*d, H, W]. If s = 4, d = 64, then the dimension of the output data has dimensions of WxHx256.

3) Enhanced Solution: 3D Swin-IR

The Swin-IR network is based on the 2D convolutional kernel, which restricts its ability to effectively analyze 3D volume data. To fully utilize the 3D volume data, we need to incorporate the 3D convolutional kernel. Toward this end, we extended the base network using MONAI's Swin-UNETR [9]. While Swin-UNETR is designed for segmentation, we incorporated several layers from Swin-IR to enhance the model's capabilities for super-resolution. To maintain consistency and compatibility, we utilized the loss functions and optimization scheme of Swin-IR, resulting in our true-3D approach, as shown in Fig. 3.

III. Experimental Results

In this section, we present our experimental results and comparison study using our 1D, pseudo-3D, and true-3D approach. Note that the 2D approach does not generate the desired output dimension for virtual thickness reduction, therefore it is not included in the comparison study.

A. Data

Since the access to paired original thick slice CT and original thin slice CT data from scanner is limited, we generated synthetic thick slice CT from the original thin slice CT data by down-sampling. We used the synthesized thick (input) – thin (output) CT data pairs from the public datasets [10] to construct a training dataset consisting of approximately 220 CTs. The

dataset contains data of varying quality from multiple years. To evaluate and validate our models, we selected a set of 14 independent CT scans, which were obtained from different patients and from a different source. The image resolution in the axial view was 512x512. For training purposes, we selected only data with at least 256 slices for thin slice data. This ensures that the corresponding thick-slice data has at least 64 slices.

B. Evaluation Metrics

To evaluate the super-resolution performance, we used both PSNR and SSIM metrics.

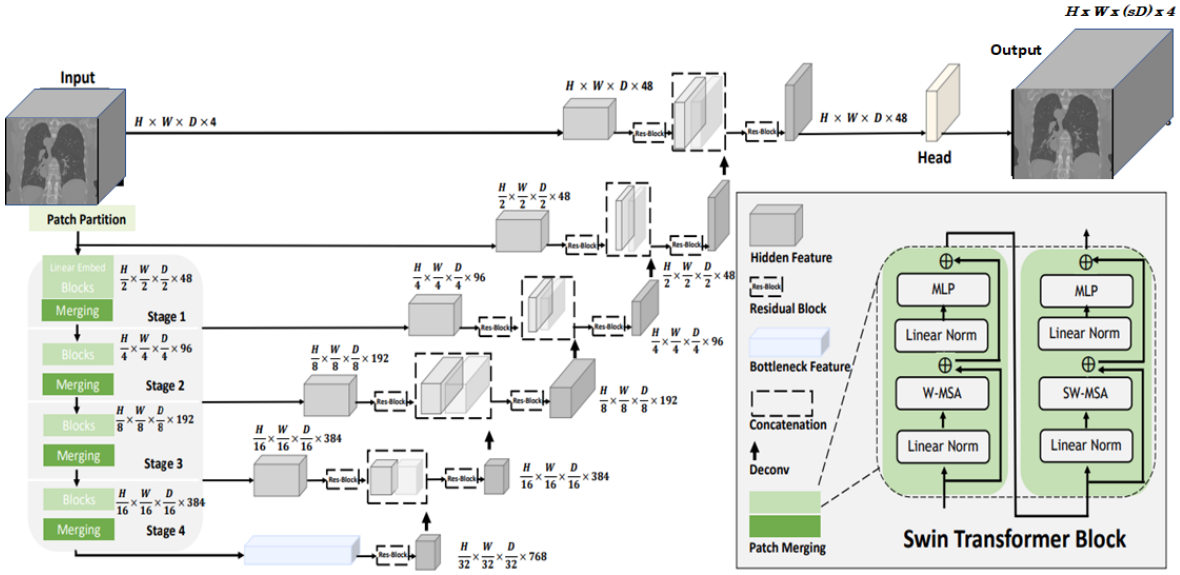

Fig. 3 The architecture for the true-3D super-resolution, which is built based on Swin-UNETR [9]

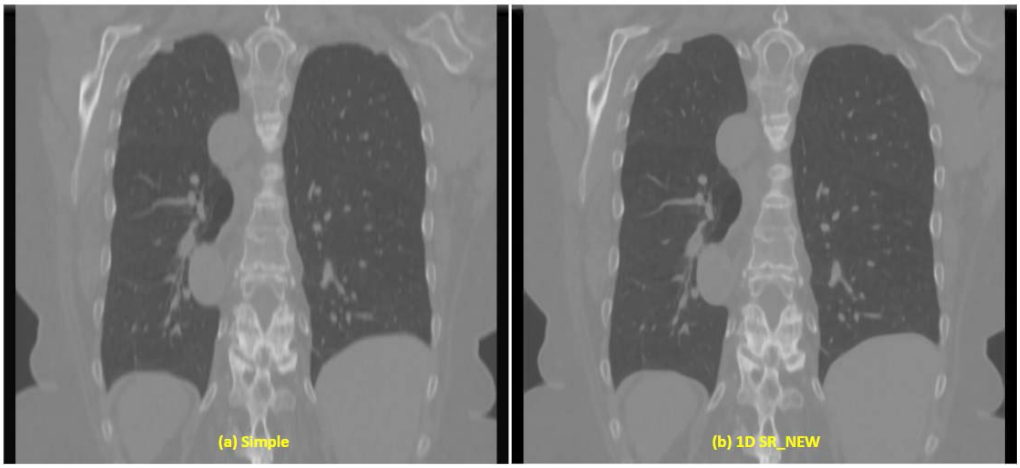

Fig. 4 Comparison of super-resolution (x4) using 1D approach (b) vs. simple up-sampling approach (a)

PSNR is a widely used image quality assessment metric that is based on the mean squared error (MSE) between a reference image and a compressed or constructed image. It provides a measure of the level of compression or noise in the image. Higher PSNR values indicate better image quality, while lower values suggest that the image is more distorted or noisy.

SSIM is based on the perceived changes in structural information, luminance, and contrast of an image compared to a reference image. It compares local patterns of pixel intensities between the two images, rather than just comparing the overall brightness and color. SSIM values range from -1 to 1, with 1 indicating identical images and values closer to 0 indicating greater dissimilarity. Note that in our case, the original thin-slice data are available and serve as the reference image. We also qualitatively compare the airway segmentation performance to compare the usefulness of the super-resolution.

C. Experimental Results and Compassion Study

We applied our learned models with 1D, peudo-3D, and true-3D approaches to the test data. We observe that all three methods are able to improve the quality of the original thick slice data. Two examples are shown in Figs. 4 and 5.

The quantitative evaluation using PSNR and SSIM metrics is shown in Fig. 6. We calculated both PSNR and SSIM metrics

by comparing the original thin slice data to the super-resolved data, using their projections in all three views (axial, coronal, and sagittal).

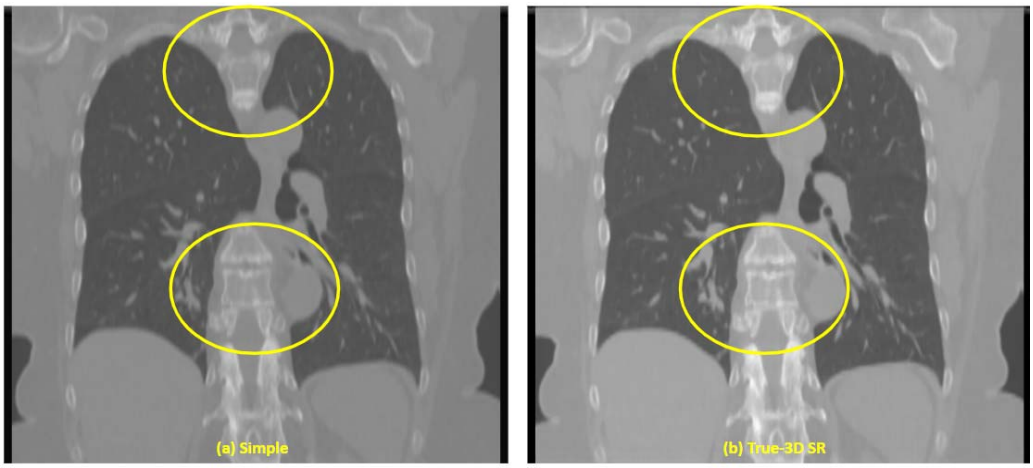

Fig. 5 Comparison of super-resolution (x4) using true-3D approach (b) vs. simple up-sampling approach (a)

	Simple Up-Sample vs Original High-Res			1D Super-Res vs Original High-Res			Pseudo-3D Super-Res vs Original High-Res			True-3D Super-Res vs Original High-Res		
	Axial	Coronal	Sagittal	Axial	Coronal	Sagittal	Axial	Coronal	Sagittal	Axial	Coronal	Sagittal
PSNR	78.0094	78.0741	78.0746	77.0283	77.1711	77.1721	78.0962	78.1851	78.1858	69.9332	70.0840	70.0852
SSIM	0.9269	0.9195	0.9194	0.9792	0.9774	0.9772	0.9698	0.9656	0.9654	0.9597	0.9596	0.9598

Fig. 6 Quantitative evaluation of 1-D, pseudo-3D, and true-3D super-resolution methods, in comparison with the simple up-sampling method; both PSNR and SSIM metrics are used

According to the PSNR metric, both the 1-D and pseudo-3D approaches exhibit similar performance to the simple up-sampling method. However, when evaluated using the SSIM metric, all three approaches outperform the simple up-sampling method. The SSIM metric aligns closely with human visual assessment and is the more appropriate metric for the super-resolution evaluation. In addition, we observe that the 1D approach performs surprisingly well according to SSIM metric. This may be due to the fact we evaluated in the three projected views rather than in the original 3D space.

IV. Conclusion

This paper introduces a set of heterogeneous dimensional super-resolution methods aimed at improving the thickness of CT scans. In particular, we have adapted and enhanced the state-of-the-art transformer architectures to effectively address the problem of super-resolution. We plan to explore more comprehensive evaluation methods, including an assessment of the super-resolution quality using airway segmentation performance.

References

[1] A. Kudo, et. al. Virtual Thin Slice: 3D Conditional GAN-based Super-resolution for CT Slice Interval. https://arxiv.org/pdf/1908.11506.pdf
[2] S. Park, et. al. Deep Learning Algorithm for Reducing CT Slice Thickness: Effect on Reproducibility of Radiomic Features in Lung Cancer. Korean J Radiol. 2019 Oct; 20(10): 1431–1440.
[3] U. Agrawal, et. al. Enhancing Z-resolution in CT volumes with deep residual learning. SPIE Medical Imaging 2021.
[4] H. Xie, et. al. High through-plane resolution CT imaging with self-supervised deep learning. 2021 *Phys. Med. Biol.* 66 145013.
[5] W. Bae, et. al. Residual CNN-based Image Super-Resolution for CT Slice Thickness Reduction using Paired CT Scans: Preliminary Validation Study. https://openreview.net/forum?id=S1RzBW2oz.
[6] M. Kiss, et. al. Z-Super Resolution CT-Image Generation with A Deep 3D Fully Convolutional Neural Network. https://doi.org/10.1016/j.ijrobp.2020.07.249.
[7] C. You, et. al. CT Super-resolution GAN Constrained by the Identical, Residual, and Cycle Learning Ensemble (GAN-CIRCLE) https://arxiv.org/pdf/1808.04256.pdf.
[8] J. Liang, et. al. SwinIR: Image Restoration Using Swin Transformer. https://arxiv.org/abs/2108.10257.
[9] https://github.com/Project-MONAI
[10] NLST dataset in the Cancer Image Archive. https://cdas.cancer.gov/datasets/nlst/