

UCLA

UCLA Previously Published Works

Title

An Efficient and Robust Method for Chest X-ray Rib Suppression That Improves Pulmonary Abnormality Diagnosis

Permalink

<https://escholarship.org/uc/item/6zq7w1j2>

Journal

Diagnostics, 13(9)

ISSN

2075-4418

Authors

Xu, Di
Xu, Qifan
Nhieu, Kevin
[et al.](#)

Publication Date

2023



DOI

10.3390/diagnostics13091652

Peer reviewed

Article

An Efficient and Robust Method for Chest X-ray Rib Suppression That Improves Pulmonary Abnormality Diagnosis

Di Xu ¹, Qifan Xu ¹ , Kevin Nhieu ¹, Dan Ruan ¹ and Ke Sheng ^{2,*} 

¹ Department of Radiation Oncology, University of California at Los Angeles, Los Angeles, CA 90095, USA; dixu@mednet.ucla.edu (D.X.); qifanxu@mednet.ucla.edu (Q.X.); kknhien7@g.ucla.edu (K.N.); druan@mednet.ucla.edu (D.R.)

² Department of Radiation Oncology, University of California at San Francisco, San Francisco, CA 94115, USA

* Correspondence: ke.sheng@ucsf.edu

Abstract: Background: Suppression of thoracic bone shadows on chest X-rays (CXRs) can improve the diagnosis of pulmonary disease. Previous approaches can be categorized as either unsupervised physical models or supervised deep learning models. Physical models can remove the entire ribcage and preserve the morphological lung details but are impractical due to the extremely long processing time. Machine learning (ML) methods are computationally efficient but are limited by the available ground truth (GT) for effective and robust training, resulting in suboptimal results. Purpose: To improve bone shadow suppression, we propose a generalizable yet efficient workflow for CXR rib suppression by combining physical and ML methods. Materials and Method: Our pipeline consists of two stages: (1) pair generation with GT bone shadows eliminated by a physical model in spatially transformed gradient fields; and (2) a fully supervised image denoising network trained on stage-one datasets for fast rib removal from incoming CXRs. For stage two, we designed a densely connected network called SADXNet, combined with a peak signal-to-noise ratio and a multi-scale structure similarity index measure as the loss function to suppress the bony structures. SADXNet organizes the spatial filters in a U shape and preserves the feature map dimension throughout the network flow. Results: Visually, SADXNet can suppress the rib edges near the lung wall/vertebra without compromising the vessel/abnormality conspicuity. Quantitatively, it achieves an RMSE of ~ 0 compared with the physical model generated GTs, during testing with one prediction in <1 s. Downstream tasks, including lung nodule detection as well as common lung disease classification and localization, are used to provide task-specific evaluations of our rib suppression mechanism. We observed a 3.23% and 6.62% AUC increase, as well as 203 (1273 to 1070) and 385 (3029 to 2644) absolute false positive decreases for lung nodule detection and common lung disease localization, respectively. Conclusion: Through learning from image pairs generated from the physical model, the proposed SADXNet can make a robust sub-second prediction without losing fidelity. Quantitative outcomes from downstream validation further underpin the superiority of SADXNet and the training ML-based rib suppression approaches from the physical model yielded dataset. The training images and SADXNet are provided in the manuscript.

Keywords: chest X-rays; deep learning; rib suppression; computer aided diagnosis



Citation: Xu, D.; Xu, Q.; Nhieu, K.; Ruan, D.; Sheng, K. An Efficient and Robust Method for Chest X-ray Rib Suppression That Improves Pulmonary Abnormality Diagnosis. *Diagnostics* **2023**, *13*, 1652. <https://doi.org/10.3390/diagnostics13091652>

Academic Editor: Philippe A. Grenier

Received: 10 April 2023

Revised: 2 May 2023

Accepted: 4 May 2023

Published: 8 May 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Respiratory diseases are among the major causes of morbidity and mortality globally, and the prevalence of pulmonary diseases has steadily increased [1,2]. Timely diagnosis is critical for effective intervention. Among all imaging tools, a chest X-ray (CXR) is the most widely used for pre-screening thoracic anomalies [3]. Compared with CT, the downside of a CXR is the overlapping anatomies in 2D projections. The high-contrast bony structures are one of the major interferences in CXRs, obscuring the underlying soft tissues. Therefore,

the suppression of bone shadows in CXRs is highly desired to improve the conspicuity of lung tissues and, subsequently, improve diagnoses.

Recently, clinical evidence indicates that rib-removed CXRs can improve the diagnosis of various pulmonary abnormalities [4,5]. In a post-processing setup, previous approaches can be summarized as: (1) a dual-energy CXR that exposes the patient twice to a high and low tube voltage for digital subtraction of the ribs [6]; (2) unsupervised physical models [3,7]; and (3) deep learning (DL) models with supervision generated from dual-energy (DE) CXRs or a domain-adapted digitally reconstructed radiograph (DRR) [8,9].

A dual-energy CXR exploits the energy-dependent X-ray tissue interactions in the ribs and soft tissues and, then, performs digital subtraction to suppress the bone shadow. The method requires a specialized instrument that is not widely available and exposes the patients to a higher imaging dose, due to the dual exposure. Physical rib suppression models utilize a single-energy (SE) CXR first to reconstruct the rib structures by empirically assuming a rib pixel intensity distribution and then subtracting the rib shadows from the original CXR. For instance, von Berg et al. proposed the transfer of the ribs into the “ST-space”, where the contour of the rib appears as a straight line. Partial derivative computation, smoothing, and reintegration were then performed along the rib contours up to the centerline [7]. These highly hand-crafted rib suppression methods can preserve the morphological detail but are not scalable for clinical use, due to the tedious manual annotation of each rib and the case-by-case hyperparameter (HP) fine tuning that demands an impractically long processing time, which diminishes the benefit of CXRs [4,7,10–12].

DL methods are a time efficient alternative to conventional operations, and their performance depends on the methods and the training data quality [13]. Except for diverse network schemes for superior representation learning, the core disparities in existing DL approaches for CXR rib suppression are in the modality of the training images and training pair generation. Most of the previous models were trained on either DECXR or DRR. Specifically, since DECXR is formed by two XR exposures at distinct energy levels, it can later be decomposed into “soft-tissue-only” and “bone-only” images [14,15]. Learning models trained on DECXR form image pairs with unprocessed images as the input and decomposed “soft-tissue-only” images as the ground truth (GT) [3,16]. However, DECXR-based frameworks are limited by data scarcity [17]. Since the ribcage is separable using computed tomography (CT), DRR-based methods attempt to project training pairs from the original and rib-suppressed CT [18]. Although DRRs simulate CXRs, their contrast and resolution differ substantially regarding the physical image formation details, detector signal processing, and post-processing [9]. The lack of paired images also hinders the domain adaptation training from a DRR to a CXR. Insufficiently trained DL approaches cannot learn the true mapping from raw to rib-removed CXRs. Consequently, the shadows of the rib edges are prevalent in DL predictions [3,9], adversely affecting the efficacy of downstream diagnostic tasks.

We contend that the answer to the rib suppression challenge lies in combining the two existing approaches. Inefficient physical models can maximally suppress the rib structures and preserve the lung tissue details, thus they are suitable to provide machine learning (ML) training input as a one-time investment. With such high-quality input, efficient DL models can be trained for wide clinical implementation.

We hereby introduce a benchmark dataset named FX-RRCXR for DL-based CXR rib suppression. FX-RRCXR is generated from the VinDr-RibCXR dataset [19], adapting the physical ST smoothing by von Berg et al. [7]. Next, we propose a supervised image denoising network, SADXNet, trained on FX-RRCXR for fast rib suppression in unseen CXRs. Lastly, we validate the impact of rib-removed CXRs in the detection of lung nodules using the NODE21 [20] dataset, and the classification and localization of benign lung disease using ChestX-ray14 [21] datasets. All the detection tasks are run on the Mask R-CNN framework [22].

2. Method

In this section, we will introduce the ST-smoothing method, FX-RRCXR dataset, SADXNet, and downstream validation of the NODE21 and ChestX-ray14 datasets.

2.1. ST Smoothing

The ST-smoothing algorithm assumes that the pixel intensities along one continuous contour of a rib are theoretically identical. As shown in Figure 1 S0, if the distances to the centerline of p_1 and p_2 are equal, these two points are deemed to be on the same contour and their corresponding rib intensities are equal.

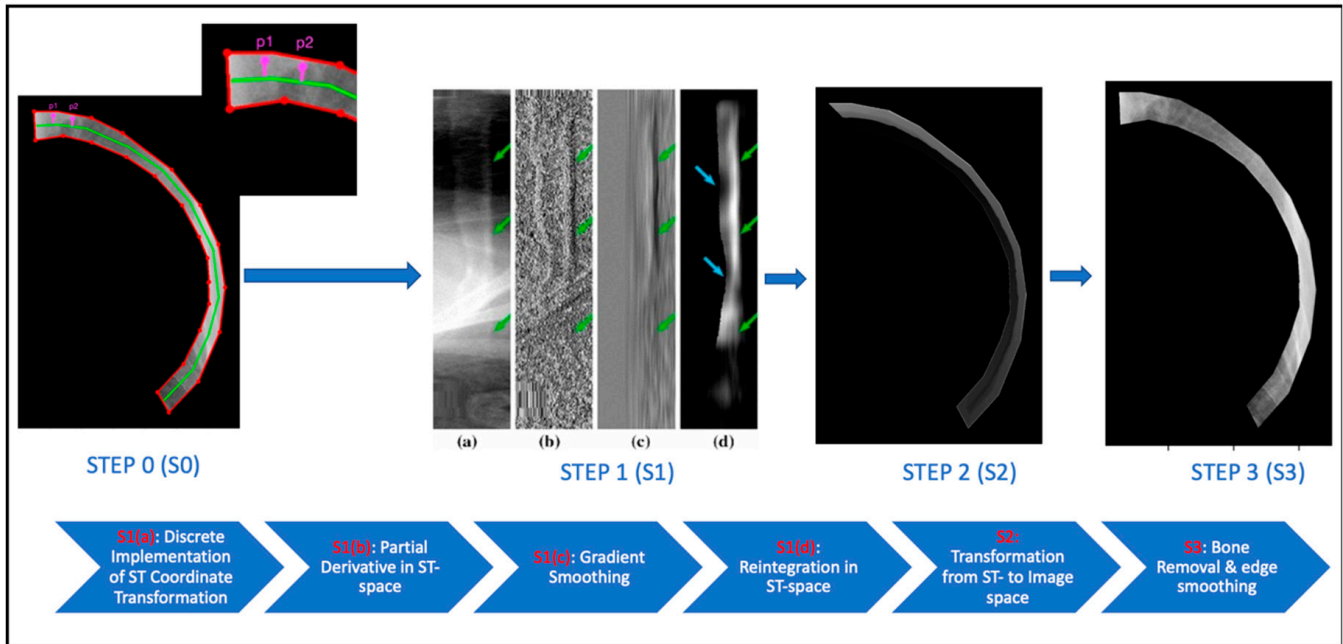


Figure 1. Major steps in the ST-smoothing algorithm applied to the fifth rib in the CXR with image visualization in the upper and pipeline description in the bottom section. The explanation for the purpose of each image is labeled in red on the flow chart.

2.1.1. From Image Space to ST-Coordinate System

We explain Figure 1 S0→S1(a) here with Equations (1) and (2).

$$T_C : (x, y) \mapsto (s, t) \tag{1}$$

$$T_C^{-1} : (s, t) \mapsto (x, y) \tag{2}$$

ST transformation T_C is a domain transformation used to generate a specific representation of a part of an image defined by the given closed cyclic contour $C : \gamma(t), t \in [0, C_{len}), \gamma(0) = \gamma(C_{len})$. T_C is defined by its inverse as follows:

$$T_C^{-1}(s, t) = \gamma(t) + s \cdot \frac{\gamma'(t)^\perp}{|\gamma'(t)|} \tag{3}$$

where $\frac{\gamma'(t)^\perp}{|\gamma'(t)|}$ is the contour norm at $\gamma(t)$. Figure 2a illustrates the transformation of contour C [7].

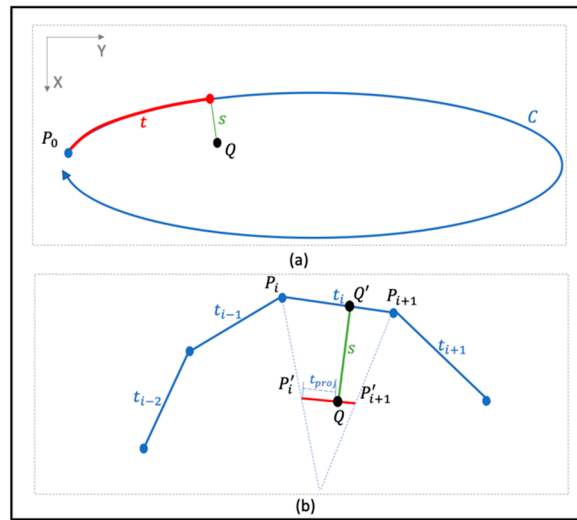


Figure 2. Visualization of the *ST*-coordinate system. (a) Point *Q* transformation from *XY*- into *ST* domain regarding contour *C*. (b) Discrete sampling in *ST*-coordinate system with respect to point *Q*.

To reduce the computation burden, a continuous rib contour is considered as a piecewise linear contour (Figure 1 S0) for the discrete implementation of Equation (1). Figure 2b illustrates the implementation, with *s* and *t* formulated as:

$$s = \|\overrightarrow{QQ'}\| \tag{4}$$

$$t = t_{prev} + \|\overrightarrow{QP'_i}\| \cdot \frac{\|P_i P_{i+1}\|}{\|P'_i P'_{i+1}\|} \tag{5}$$

where t_{prev} is the sum of the length of all previous edges. The pixel intensities in the *ST* space are as follows:

$$I_{st_C}(s, t) = I(T_C^{-1}(s, t)) \tag{6}$$

$$(\hat{s}, \hat{t}) = T_C(T_C^{-1}(s, t)) \tag{7}$$

When it is assumed that a valid (s, t) always has a position (\hat{s}, \hat{t}) on the other side of the bone centerline $c(t)$, we can obtain $c(t)$ from:

$$c(t) = \max_S \forall (\hat{s}, \hat{t}) \neq (s, t) \tag{8}$$

2.1.2. Rib Extraction via Partial Derivatives Smoothing in *ST* Space

The current subsection explains steps S1(a)→S2 in Figure 1.

Discrete Partial Derivative in *ST* Space

As shown in Equation (9), the first-order partial derivative is calculated along the *s* axis in a discrete form to boost the overall computation. The definition of $I_{d_C}(s, t)$ represents the gradient orthogonal to the *t* axis of a rib, which means that any structure oriented along axis *t* does not contribute to the bone gradients.

$$I_{d_C}(s, t) = \partial_s I_{st_C}(s, t) = I(s, t) - I(s - 1, t) \tag{9}$$

Smoothing, Reintegration, and Transformation Back to the *XY* Domain

Improved from von Berg et al. [7], Gaussian smoothing ($\mathcal{G}_{\hat{t}_i}$) along the *t* axis at I_{d_C} and centerline smoothing (*C*) along the *s* axis at the reintegrated I_{r_C} are implemented. First,

since we hypothesize that signals along the t axis of I_{d_C} are independent from the ribs. A large Gaussian kernel ℓ_t was used to smooth out the signals from the soft tissue and leave only the signals from the ribs in $\mathcal{G}_{\ell_t}(I_{d_C})$. Note that ℓ_t is a HP. Next, after excluding the soft tissue signal via $\mathcal{G}_{\ell_t}(I_{d_C})$, we reintegrate towards the smoothed partial gradients to recover the bone signals $I_{r_C}(s, t)$ in the ST domain, as shown in Equation (10).

$$I_{r_C}(s, t) = \int_{s-1}^s \mathcal{G}_{\ell_t}(I_{d_C}(s, t)) + \mathcal{G}_{\ell_t}(I_{d_C}(s-1, t)) \quad (10)$$

Lastly, because T_C has a singularity at the centerline $c(t)$, an artificial edge was observed along $c(t)$, as shown in Figure 1 S2, after the reintegration of I_{r_C} . Therefore, a K-nearest neighbor (KNN) based centerline smoothing is applied along the s axis of I_{r_C} to smooth out the artificial edge, according to Equation (11).

$$\mathcal{C}(I_{r_C}(s_i, t_j)) = \begin{cases} I_{r_C}(s_i, t_j), & \text{if } \frac{I_{r_C}(s_i, t_j)}{I_{r_C}(s_{i-1}, t_j)} > \tau \\ \frac{\sum_{m=i-k+1}^i I_{r_C}(s_m, t_j)}{k}, & \text{otherwise} \end{cases} \quad (11)$$

where τ and k are two HPs and represent the threshold for conducting the KNN average and the number of neighbors used, respectively.

After the above steps, we transfer the rib intensity $\mathcal{C}(I_{r_C})$ from the ST back into the image space under Equation (12) to exclude possible negative values, which are uninterpretable in the image space.

$$I_{bone_C} = \max(\mathcal{C}(I_{r_C}), 0) \quad (12)$$

2.1.3. Rib Removal and Border Blending

We focus on Figure 1 S3 here. The initial rib-suppressed CXR I_{soft_C} is acquired by subtracting the I_{bone_C} from the raw CXR, as follows:

$$I_{soft_C}(x, y) = I(x, y) - I_{bone_C}(x, y) \quad (13)$$

To improve the continuity between the rib boundary r_b and its surrounding soft tissues, a KNN border smoothing function is applied to r_b :

$$r'_b(x, y) = \frac{\sum_{i=1}^k r_{b,i}(x, y)}{k} \quad (14)$$

where r_b is defined under the ST space and shown in Equation (15). s_b in Equation (15) and k in Equation (14) are two HPs.

$$\begin{aligned} I_{soft_C}(s, t) & \text{ in } r_b(s, t), \text{ if } s \leq s_b \\ r_b(x, y) & = T_C^{-1}(r_b(s, t)) \end{aligned} \quad (15)$$

Lastly, generating a complete soft tissue CXR requires iteratively repeating Sections 2.1.1–2.1.3 for each rib in a complete ribcage.

2.2. Data Cohort

2.2.1. VinDr-RibCXR Dataset

VinDr-RibCXR is selected for creating the FX-RRCXR dataset using our modified ST-smoothing algorithm. VinDr-RibCXR is a dataset for automatic rib segmentation and labeling of SECXR scans. It contains 245 images with corresponding rib masks annotated by an expert. Each CXR scan has 20 separate rib annotations in the left and right lungs, respectively. The dataset was pre-split into training and validation sets, with 196 scans in the training set and 49 in the validation set. We refer the readers to Nguyen et al. [19] for more details.

2.2.2. FX-RRCXR Dataset Preparation

We applied the ST-smoothing algorithm to 245 images from the VinDr-RibCXR dataset. The HPs required by ST smoothing are tuned for individual images using the random grid search algorithm. Excluding the process for HP searching, the ST smoothing took 40–70 min to process a rib-removed scan, depending on the number of pixels within the ribcage. We organized the original CXRs as input and their corresponding rib-suppressed scans as GT while preparing the image pairs and kept the same training and validation split as the VinDr-RibCXR.

2.3. SADXNet

Since rib signals can be treated as noise superimposed on the rib-suppressed CXRs, we proposed the design of an image-denoising network, named SADXNet, to be trained on the FX-RRCXR dataset. Inspired by the architecture of DenseNet [23], SADXNet is designed to be densely connected, as shown in Figure 3. For each layer in SADXNet, the feature maps of all the preceding layers and their own feature map are fed into all the subsequent layers. The advantage of dense connection is the better alleviation of the vanishing-gradient problem, strengthening feature propagation, and encouraging feature reuse [23]. The composition of SADXNet is detailed below.

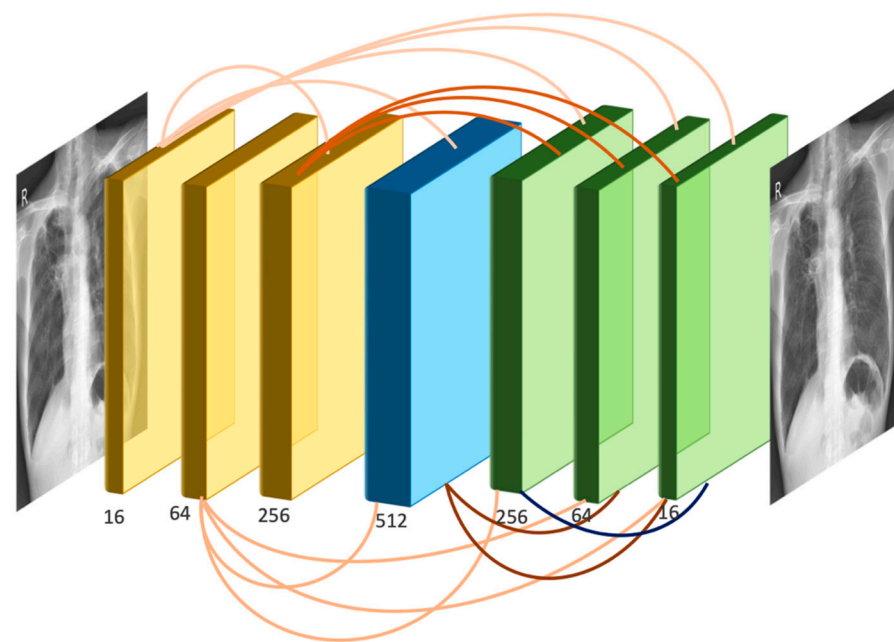


Figure 3. The network architecture of SADXNet.

Dense connectivity. Figure 3 shows the layout of the densely connected SADXNet schematically. Each layer can be described by Equation (16):

$$x_l = N_l([f_{c_l}(x_0), f_{c_l}(x_1), \dots, f_{c_l}(x_{l-2}), x_{l-1}]) \quad (16)$$

where $[x_0, x_1, \dots, x_{l-1}]$ represents the channel-wise concatenation of the feature maps produced from layer 0 to $l-1$ and $f_{c_l}(\cdot)$ is a 1×1 convolution (Conv) to unify the number of feature channels of x_0, \dots, x_{l-2} . The output channel of f_{c_l} is defined by $\frac{C(x_{l-1})}{L-1}$, where C represents the number of feature channels of x_{l-1} . The purpose of $f_{c_l}(\cdot)$ is to avoid the concatenated input feature maps, $N_l(\cdot)$ being overly large in channel dimension exceeding the GPU memory.

Composite function: We define $N_l(\cdot)$ as a composite function of three consecutive operations: batch normalization (BN), rectified linear unit (ReLU), and 3×3 convolution (Conv).

Pooling layers: Inspired by a pioneering study [8], we preserve the height (H) and width (W) dimensions in the feature maps as the shape of the input images throughout the Conv process without using down- or up-sampling layers.

Channel design: Overall, there are seven densely connected layers in the SADXNet. The channel of the convolutional kernel for each layer is designed in an increase-to-decrease setting to mimic the design of a fully convolutional network [24] in regard to channel dimension, so as to strike a balance between the model complexity (kernel with more channels) and training time (kernel with fewer channels).

Loss function: The cost function of the SADXNet is designed as a combination of the negative peak signal-to-noise ratio, the multi-scale structure similarity index measure [25], and the L_1 deviation measurement:

$$L = -\alpha \cdot L_{PSNR} + (1 - \alpha) \cdot [\beta \cdot L_{MS-SSIM} + (1 - \beta) \cdot L_1] \quad (17)$$

$$L_{PSNR} = \log_{10} \left(\frac{MAX_X^2}{1/mn \cdot \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [x_{ij} - y_{ij}]^2} \right) \quad (18)$$

$$L_{MS-SSIM} = 1 - \frac{(2\mu_x\mu_y + c_1)}{(\mu_x^2 + \mu_y^2 + c_1)} \cdot \prod_{j=1}^M \frac{(2\sigma_{x_j y_j} + c_2)}{(\sigma_{x_j}^2 + \sigma_{y_j}^2 + c_2)} \quad (19)$$

$$L_1 = \frac{1}{mn} \cdot \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \|x_{ij} - y_{ij}\|_1 \quad (20)$$

where α and β in Equation (17) are HPs set to 0.75 and 0.25, X and Y are the model input and target, MAX_X is the maximum possible input value, $[\sigma_{x_j y_j}, \dots, \sigma_{x_M y_M}]$ of Equation (18) set to [0.5, 1.0, 2.0, 4.0, 8.0], $c_1 = (k_1 S)^2$ and $c_2 = (k_2 S)^2$ of Equation (19) are two variables to stabilize the division with a weak denominator having S as the dynamic range of the pixel values (typically $2^{\# \text{ bits per pixel}} - 1$) and (k_1, k_2) as the constants, and $\|\cdot\|_1$ denotes the l_1 norm.

Model training: SADXNet was implemented in PyTorch, and the training was performed on a GPU cluster with $4 \times RTX A6000$. Per the SADXNet training, we set the maximum epoch number to 200 and observed that the model converged at around 100 epochs. The Adam optimizer with an initial learning rate (LR) of 0.001 and a batch size of 1×4 were applied.

Evaluation metrics: The root mean square error (RMSE) is used for evaluating the SADXNet performance with the corresponding GT generated from ST Smoothing.

2.4. Downstream Clinical Task Validation

We quantified the benefits of rib-suppressed images with two experiments, including lung nodule detection and a general pulmonary disease classification and localization task based on the NODE21 [20] and ChestXRy14 [21] datasets, respectively. The details are set out below.

Training input: We organized the input into three combinations: (1) solely raw CXRs; (2) solely soft-tissue CXRs; and (3) mixed raw and rib-suppressed CXRs. All the rib-suppressed images were predicted using SADXNet.

Datasets: NODE21 [20] encompasses 4882 CXR scans with the ratio of patient: volunteer = 1134:3748. Moreover, 5524 annotations were made to these images, with a maximum of three positive annotations for each scan. ChestX-ray14 [21] is a CXR set that has been text mined with fourteen lung diseases and bounding box annotations for 984 images in the pre-split test set. We extracted the 984 annotated images and then randomly sampled 3×984 healthy volunteers from the test set to construct a dataset for supervised detection training. For both datasets, we split the training and validation sets in the ratio of 7:3 and carefully balanced the proportion of positive and negative cases.

Detection network: Both tasks were performed in the Mask R-CNN pipeline and implemented in PyTorch [22], training the network separately on three different types of

input with data augmentation of random scaling, random cropping, and random Gaussian blur, an LR of 0.001, a stochastic gradient descent optimizer, and a batch size of 2×4 on the $4 \times RTX A6000$ GPU cluster. Altogether, we ran 30,000 and 50,000 training iterations for NODE21 [20] and ChestX-ray14 [21], respectively.

Evaluation metrics: The area under the curve (AUC), the true positive predictions (TP), the false positive predictions (FP), and the false negative predictions (FN) are used as evaluation metrics.

3. Results

3.1. Rib Suppression

3.1.1. ST Smoothing

Figure 4 demonstrates a patient with lung nodules in the left lung and compares the node visibility with and without rib removal using ST smoothing. Compared to previous rib-suppression methods with edge residuals or artifacts near the lung borders [4,9], ST smoothing carefully avoids those two drawbacks. Additionally, ST smoothing preserved the shape and morphological details of the lung tumors.

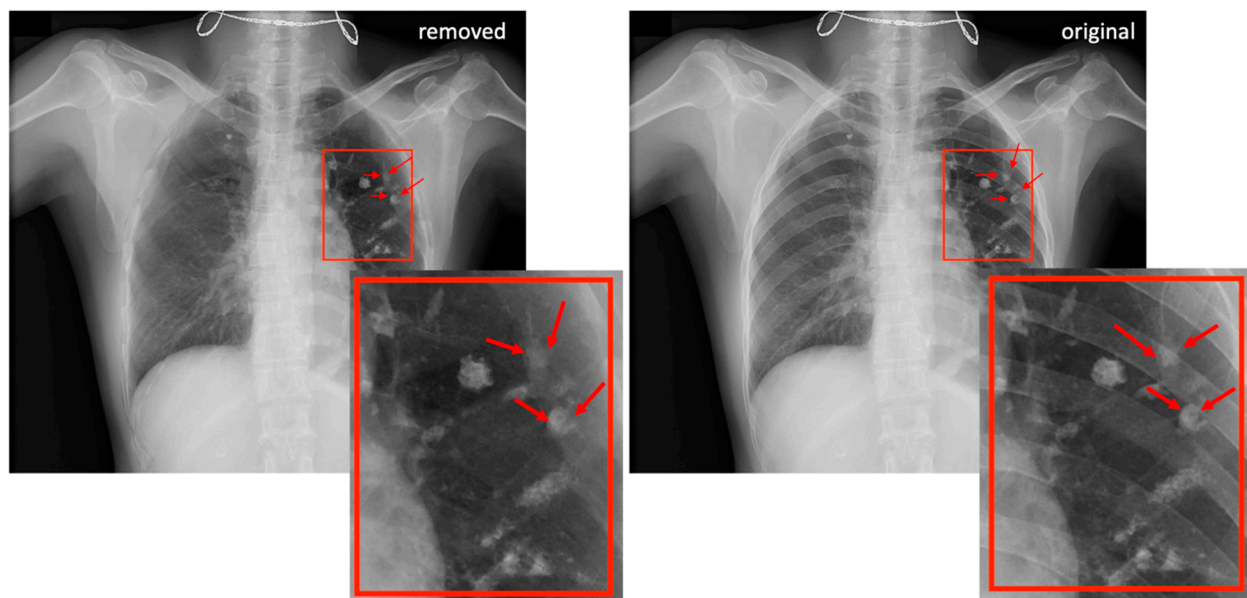


Figure 4. A sample case involving ST smoothing with several lung nodules in the left lung field. The left-hand side shows the rib-removed scan, and the right-hand side shows the original unprocessed CXR image. Red arrows denote radiologically confirmed nodules.

3.1.2. SADXNet

Figure 5 shows two subjects predicted by SADXNet. We found that the rib-suppressed scans predicted by SADXNet are visually indistinguishable from their corresponding GTs. Quantitatively, SADXNet achieves a $2.32 \pm 0.13 \times 10^{-5}$ test RMSE. Most notably, compared to the time consuming ST-smoothing algorithm, SADXNet suppresses one scan in <1 s.

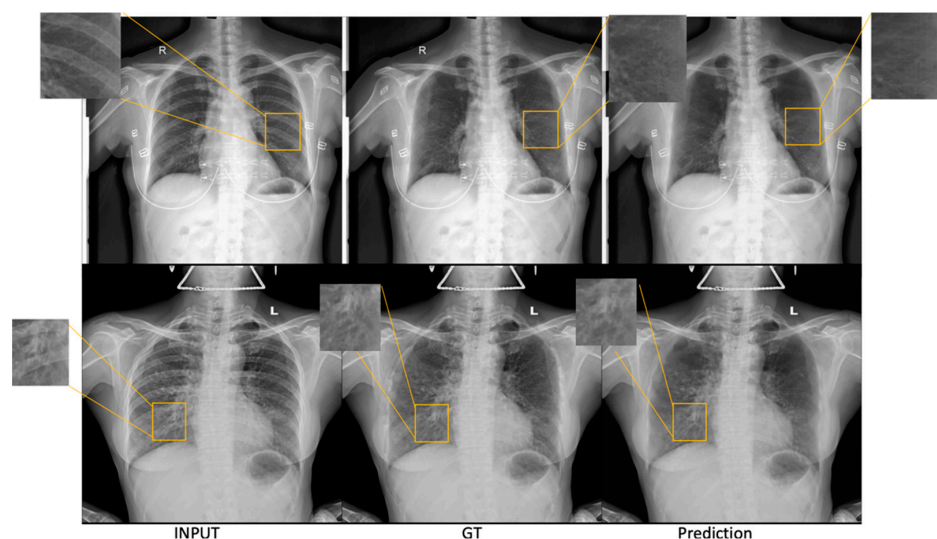


Figure 5. Two sample cases from the test set illustrate the results of SADXNet rib removal. Each row represents one patient. The columns from left to right are ordered as the first column showing the raw CXR, the second showing the rib-suppressed image generated from ST smoothing (GT for SADXNet training supervision), and the third showing the rib-suppressed CXRs predicted by the trained SADXNet model.

3.2. Downstream Clinical Task Validation

3.2.1. NOD21 Detection

According to Table 1, mixing raw and rib-suppressed images in training achieved the best detection scores than training on only a single source input. We also evaluated the mix-trained model on raw and rib-suppressed validation sets, respectively. Both scenarios achieved comparable performance, with slightly better outcomes for the raw CXR. Quantitatively, training using mixed images achieved approximately a 2–3% higher AUC, located more nodules (Figure 6 first row), and significantly reduced the FP (Figure 6 second row) than single-source images. Lastly, the performance of networks trained on a single image type is similar, with a slightly lower FP using rib-suppressed images.

Table 1. Evaluation results from the NODE21 dataset. FN, FP, and TP show the absolute number of predictions for a more straightforward comparison. The best performer is marked in bold.

Modality	Training Input	Validation Input	AUC	FN	FP	TP
Mask R-CNN	Raw	Raw	94.76%	48	1273	372
	Rib suppressed	Rib suppressed	95.32%	45	1193	375
	Raw + Rib	Raw	97.99%	32	1070	388
	suppressed	Rib suppressed	97.31%	33	1082	387

3.2.2. ChestX-ray14 Classification and Localization

As shown in Table 2, training with mixed scans achieved the best performance across the three input combinations, resulting in around a 6–7% higher AUC than single-source trained detectors and largely reducing the FP predictions. Single-source trained models roughly reach similar performance, except that a model trained with rib-suppressed images makes fewer FP classifications. Figure 6 visually confirms the quantitative results.

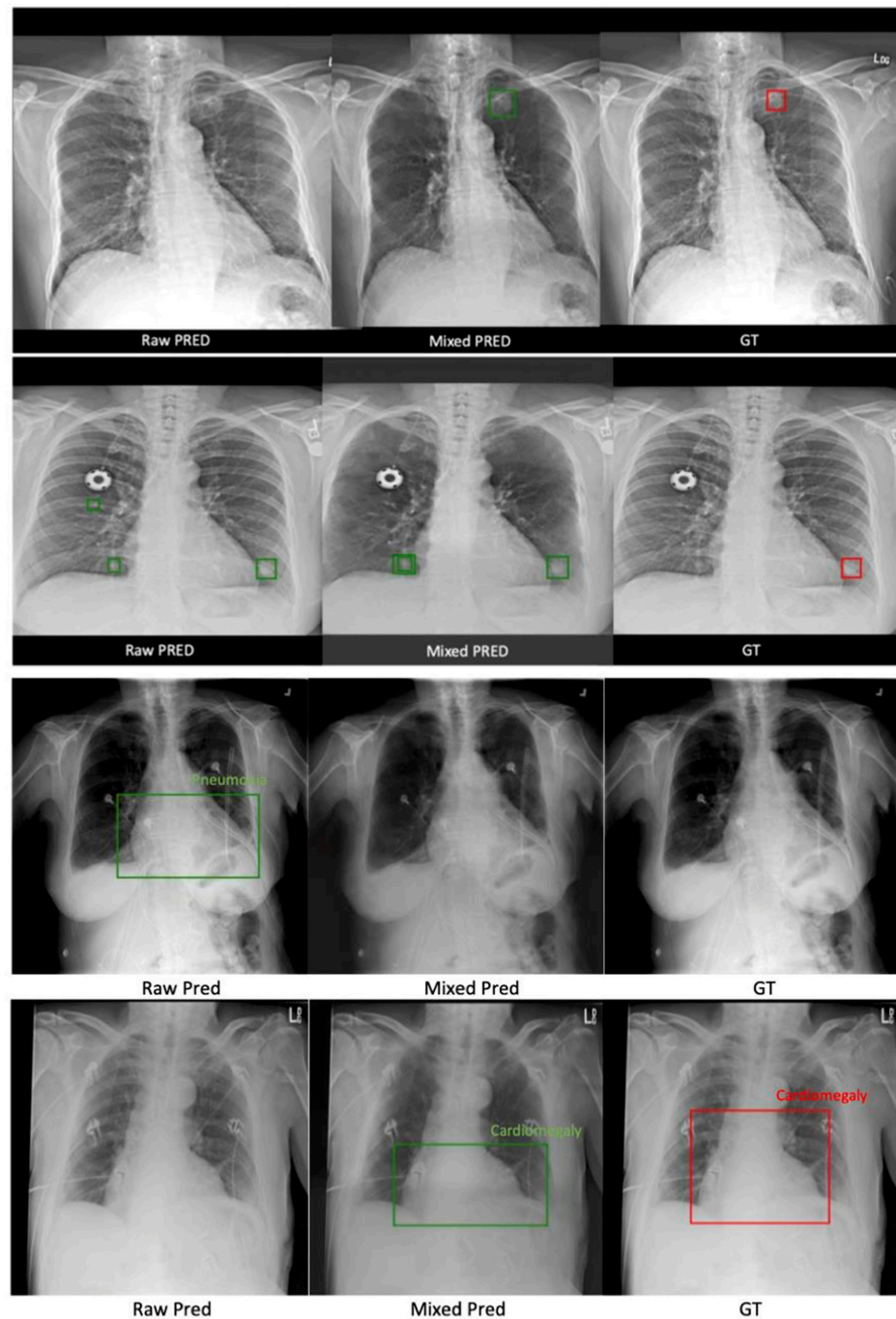


Figure 6. Two sample cases from the NODE21 test set in the first two rows, and two sample cases from the ChestX-ray14 test set in the third and fourth rows. The figure is organized into three columns: the first column shows predictions made using the model trained with raw CXRs only, the second column shows predictions made using the mix-trained model, and the last column shows the GTs. Red boxes denote manually labeled bounding boxes for the disease location. Green boxes are results of automated detection.

Table 2. Evaluation results from the ChestX-ray14 dataset. FN, FP, and TP show the absolute number of predictions for a more straightforward comparison. The best performer is marked in bold, and the lowest result is underlined.

Modality	Training Input	Validation Input	AUC	FN	FP	TP
Mask R-CNN	Raw	Raw	<u>80.54%</u>	137	3029	245
	Rib suppressed	Rib suppressed	81.55%	<u>138</u>	<u>2909</u>	<u>244</u>
	Rib suppressed	Raw	87.16%	116	2644	275
		Rib suppressed	86.89%	124	2701	267

Moreover, since ChestX-ray14 is for multiple lung disease localization, we also present the AUC for each disease in Table 3. Despite the varying disease-specific performance, the improvements benefiting from mixed training are consistent with Table 2.

Table 3. The AUC scores for 14 lung diseases in the ChestX-ray14 dataset. Tr/Val represents training/validation. The best results are marked in bold, and the lowest ones are underlined.

Disease	Input: Tr/Val	Raw/Raw	ST/ST	Raw + ST/Raw	Raw + ST/ST
	Atelectasis		79.21%	<u>77.35%</u>	85.27%
Cardiomegaly		83.47%	<u>82.03%</u>	93.84%	92.67%
Effusion		84.98%	<u>84.72%</u>	91.02%	91.02%
Infiltration		<u>69.84%</u>	70.00%	75.21%	74.32%
Mass		<u>77.23%</u>	77.38%	87.83%	85.79%
Nodule		76.32%	75.37%	82.33%	81.29%
Pneumonia		73.89%	<u>73.24%</u>	79.43%	78.33%
Pneumothorax		82.43%	<u>82.32%</u>	91.37%	90.62%
Consolidation		<u>74.57%</u>	75.23%	82.77%	81.98%
Edema		88.14%	<u>87.28%</u>	95.32%	95.33%
Emphysema		90.68%	<u>89.47%</u>	96.21%	95.79%
Fibrosis		<u>82.99%</u>	83.21%	87.45%	86.76%
Pleural Thickening		73.37%	<u>72.45%</u>	80.34%	79.85%
Hernia		<u>88.03%</u>	89.21%	97.24%	96.79%

4. Discussion

Chest X-rays (CXRs) are highly accessible and cost effective for point-of-care pulmonary disease screening. However, the interpretation of a CXR can be difficult because structures overlap with each other in 2D projection. Specifically, tissue details can be obscured by bright ribs in a CXR due to their higher atomic number and density. We hereby combined physical and ML methods' strengths to overcome their weaknesses. We first used ST smoothing and the VinDr-RibCXR [19] dataset to create a benchmark dataset, FX-RRCXR, with 245 paired original and rib-suppressed CXRs. We then trained a supervised denoising network, SADXNet, achieving high-quality sub-second rib suppression. Lastly, we evaluated the quality of the FX-RRCXR dataset and SADXNet using two downstream tasks, including lung nodule detections using the NODE21 [20] dataset and fourteen lung disease classifications and localization using the ChestXray14 [21] dataset.

Our contribution is thus twofold. First, we used a physical model to generate a qualitative dataset that supports deep learning, which is released to the public as part of the paper. Second, we trained a fully automated supervised deep network that achieved physical model quality 1×10^4 times faster. The enormous gain in efficiency benefits large-scale testing of rib-suppressed CXR for various clinical applications, two of which were exemplified here, including the testing efficacy of lung nodule detection and benign disease classification. In both end-to-end tests, our quantitative testing showed significant improvement in diagnosis with rib suppression. To the best of our knowledge, this is the first study to demonstrate the benefit of combining a physical method with ML in end-to-end diagnostic testing.

The current study also demonstrated the pipeline's robustness. Generally, distributional shift poses a major challenge in clinical deployment [26]. Here, we found that although SADXNet was trained on FX-RRCXR, it still robustly suppressed the rib structures

on scans from NODE21 and ChestX-ray14 (see Figure 6 middle column), yielding improved stability in clinically relevant tasks.

An interesting observation is that training a detection model on rib-suppressed CXRs alone does not improve model performance compared to training on original scans. In contrast, mixed training of two image sources can significantly reduce type I error while moderately increasing sensitivity. We attribute this fact to three considerations: (1) mixed training achieved better performance, namely a comparison in the region of interest (ROI) with and without the rib helped the learning model; (2) there was a significant decrease in type I error, namely assistance from the rib-removed scans makes the model more likely to avoid misidentifying noise in the rib structures, such as edges, as an ROI; and (3) there was a moderate reduction in type II error, namely the ribs only account for part of the superimposing anatomy in CXRs. Rib Suppression does not reduce interference from other anatomical and non-anatomical structures, including the heart, major vessels, mediastinum, and attached sensors. However, the current study may provide a roadmap for reducing additional interference for CXR-based diagnostic tasks.

5. Conclusions

We improved ST smoothing from von Berg et al. [7] and, based on that, further introduced a paired dataset, FX-RRCXR, that serves as a benchmark for supervised DL on CXR rib removal. Next, we proposed a denoising network, SADXNet, which learnt rib suppression by considering the ribs as noise on the lung tissues. Lastly, we validated the efficacy of rib suppression using two downstream tasks, including lung nodule detection and common lung anomaly classification and localization. The experimental results from the downstream tasks quantitatively substantiated the benefit of the FX-RRCXR dataset and SADXNet.

Author Contributions: Methodology, D.X., Q.X., D.R. and K.S.; Software, D.X.; Formal analysis, D.X.; Resources, K.S.; Data curation, D.X. and K.N.; Writing—original draft, D.X. and K.S.; Writing—review & editing, D.R. and K.S.; Supervision, K.S.; Funding acquisition, K.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by NIH grant number R01CA259008 and DOD grant number W81XWH2210044.

Institutional Review Board Statement: The study is based on publicly available data thus does not require IRB.

Informed Consent Statement: Not applicable for studies based on anonymized public data.

Data Availability Statement: The data will be available at <https://github.com/FluteXu/CXR-Rib-Suppression> (accessed on 9 April 2023).

Acknowledgments: The study is supported in part by NIH R01CA259008 and DOD W81XWH2210044.

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

References

1. The Lancet. GBD 2017: A fragile world. *Lancet* **2018**, *392*, 1683. [[CrossRef](#)] [[PubMed](#)]
2. Soriano, J.B.; Kendrick, P.J.; Paulson, K.R.; Gupta, V.; Abrams, E.M.; Adedoyin, R.A.; Adhikari, T.B.; Advani, S.M.; Agrawal, A.; Ahmadian, E.; et al. Prevalence and attributable health burden of chronic respiratory diseases, 1990–2017: A systematic analysis for the Global Burden of Disease Study 2017. *Lancet Respir. Med.* **2020**, *8*, 585–596. [[CrossRef](#)] [[PubMed](#)]
3. Chen, S.; Suzuki, K. Separation of bones from chest radiographs by means of anatomically specific multiple massive-training ANNs combined with total variation minimization smoothing. *IEEE Trans. Med. Imaging* **2014**, *33*, 246–257. [[CrossRef](#)]
4. Hogeweg, L.; Sanchez, C.I.; van Ginneken, B. Suppression of Translucent Elongated Structures: Applications in Chest Radiography. *IEEE Trans. Med. Imaging* **2013**, *32*, 2099–2113. [[CrossRef](#)]
5. Laskey, M.A. Dual-energy X-ray absorptiometry and body composition. *Nutrition* **1996**, *12*, 45–51. [[CrossRef](#)]
6. Kelcz, F.; Zink, F.E.; Peppler, W.W.; Kruger, D.G.; Ergun, D.L.; Mistretta, C.A. Conventional chest radiography vs dual-energy computed radiography in the detection and characterization of pulmonary nodules. *Am. J. Roentgenol.* **1994**, *162*, 271–278. [[CrossRef](#)]

7. Von Berg, J.; Young, S.; Carolus, H.; Wolz, R.; Saalbach, A.; Hidalgo, A.; Gimenez, A.; Franquet, T. A novel bone suppression method that improves lung nodule detection: Suppressing dedicated bone shadows in radiographs while preserving the remaining signal. *Int. J. CARS* **2016**, *11*, 641–655. [[CrossRef](#)] [[PubMed](#)]
8. Rajaraman, S.; Zamzmi, G.; Folio, L.; Alderson, P.; Antani, S. Chest X-ray Bone Suppression for Improving Classification of Tuberculosis-Consistent Findings. *Diagnostics* **2021**, *11*, 840. [[CrossRef](#)]
9. Han, L.; Lyu, Y.; Peng, C.; Zhou, S.K. GAN-based disentanglement learning for chest X-ray rib suppression. *Med. Image Anal.* **2022**, *77*, 102369. [[CrossRef](#)]
10. Wu, D.; Liu, D.; Puskás, Z.; Lu, C.; Wimmer, A.; Tietjen, C.; Soza, G.; Zhou, S.K. A learning based deformable template matching method for automatic rib centerline extraction and labeling in CT images. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 980–987. [[CrossRef](#)]
11. Hogeweg, L.; Sánchez, C.I.; de Jong, P.A.; Maduskar, P.; van Ginneken, B. Clavicle segmentation in chest radiographs. *Med. Image Anal.* **2012**, *16*, 1490–1502. [[CrossRef](#)]
12. Simkó, G.; Orbán, G.; Máday, P.; Horváth, G. Elimination of clavicle shadows to help automatic lung nodule detection on chest radiographs. In Proceedings of the 4th European Conference of the International Federation for Medical and Biological Engineering, Antwerpen, Belgium, 23–27 November 2008; Sloten, J.V., Verdonck, P., Nyssen, M., Hauelsen, J., Eds.; Springer: Berlin/Heidelberg, Germany, 2009; Volume 22, pp. 488–491. [[CrossRef](#)]
13. Dargan, S.; Kumar, M.; Ayyagari, M.R.; Kumar, G. A Survey of Deep Learning and Its Applications: A New Paradigm to Machine Learning. *Arch. Computat. Methods Eng.* **2020**, *27*, 1071–1092. [[CrossRef](#)]
14. Kido, S.; Kuriyama, K.; Kuroda, C.; Nakamura, H.; Ito, W.; Shimura, K.; Kato, H. Detection of simulated pulmonary nodules by single-exposure dual-energy computed radiography of the chest: Effect of a computer-aided diagnosis system (Part 2). *Eur. J. Radiol.* **2002**, *44*, 205–209. [[CrossRef](#)]
15. Kido, S.; Nakamura, H.; Ito, W.; Shimura, K.; Kato, H. Computerized detection of pulmonary nodules by single-exposure dual-energy computed radiography of the chest (part 1). *Eur. J. Radiol.* **2002**, *44*, 198–204. [[CrossRef](#)]
16. Suzuki, K.; Abe, H.; MacMahon, H.; Doi, K. Image-processing technique for suppressing ribs in chest radiographs by means of massive training artificial neural network (MTANN). *IEEE Trans. Med. Imaging* **2006**, *25*, 406–416. [[CrossRef](#)] [[PubMed](#)]
17. Manji, F.; Wang, J.; Norman, G.; Wang, Z.; Koff, D. Comparison of dual energy subtraction chest radiography and traditional chest X-rays in the detection of pulmonary nodules. *Quant. Imaging Med. Surg.* **2016**, *6*, 1–5. [[CrossRef](#)]
18. Li, Z.; Li, H.; Han, H.; Shi, G.; Wang, J.; Zhou, S.K. Encoding CT Anatomy Knowledge for Unpaired Chest X-ray Image Decomposition. In *Lecture Notes in Computer Science, Proceedings of the Medical Image Computing and Computer Assisted Intervention—MICCAI 2019, Shenzhen, China, 13–17 October 2019*; Shen, D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P.-T., Khan, A., Eds.; Springer International Publishing: Berlin/Heidelberg, Germany, 2019; Volume 11769, pp. 275–283. [[CrossRef](#)]
19. Nguyen, H.C.; Le, T.T.; Pham, H.H.; Nguyen, H.Q. VinDr-RibCXR: A Benchmark Dataset for Automatic Segmentation and Labeling of Individual Ribs on Chest X-rays. *arXiv* **2021**. [[CrossRef](#)]
20. Sogancioglu, E.; Murphy, K.; Van Ginneken, B. NODE21. Zenodo. 2021. Available online: <https://zenodo.org/record/5548363#.ZFV71OzML0o> (accessed on 9 April 2023).
21. Wang, X.; Peng, Y.; Lu, L.; Lu, Z.; Bagheri, M.; Summers, R.M. ChestX-Ray8: Hospital-Scale Chest X-Ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 3462–3471. [[CrossRef](#)]
22. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. *arXiv* **2018**. [[CrossRef](#)]
23. Huang, G.; Liu, Z.; van der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. *arXiv* **2016**. [[CrossRef](#)]
24. Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. *arXiv* **2014**. [[CrossRef](#)]
25. Zhao, H.; Gallo, O.; Frosio, I.; Kautz, J. Loss Functions for Image Restoration with Neural Networks. *IEEE Trans. Comput. Imaging* **2017**, *3*, 47–57. [[CrossRef](#)]
26. Betancur, J.; Commandeur, F.; Motlagh, M.; Sharir, T.; Einstein, A.J.; Bokhari, S.; Fish, M.B.; Ruddy, T.D.; Kaufmann, P.; Sinusas, A.J.; et al. Deep Learning for Prediction of Obstructive Disease from Fast Myocardial Perfusion SPECT. *JACC Cardiovasc. Imaging* **2018**, *11*, 1654–1663. [[CrossRef](#)] [[PubMed](#)]

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.