

# Towards Texture Accurate Slice Interpolation of Medical Images Using PixelMiner

**William Roger**

Maastricht University

**Philippe Lambin** (✉ [philippe.lambin@maastrichtuniversity.nl](mailto:philippe.lambin@maastrichtuniversity.nl))

Maastricht University

**Simon Keek**

Maastricht University

**Manon Beuque**

Maastricht University

**Sergey Primakov**

Maastricht University

**Elizaveta Lavrova**

Maastricht University

**Guangyao Wu**

Maastricht University

**Chengong Yan**

Maastricht University

**Sebastian Sanduleanu**

Maastricht University

**Hester Gietema**

Maastricht University Medical Centre+

**Roberto Casale**

Maastricht University

**Mariaelena Occhipinti**

Radiomics

**Arthur Jochems**

Maastricht University

**Henry Woodruff**

Maastricht University

---

**Research Article**

**Keywords:** generative modelling, PixelCNN, medical imaging, interpolation, CT

**Posted Date:** June 12th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-586453/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Towards texture accurate slice interpolation of medical images using PixelMiner

## Authors

W. Rogers<sup>1\*</sup>, S. A. Keek<sup>1</sup>, M. Beuque<sup>1</sup>, S. Primakov<sup>1</sup>, E. Lavrova<sup>1,2</sup>, G. Wu<sup>1</sup>, C. Yan<sup>1</sup>, S. Sanduleanu<sup>1</sup>, H. Gietema<sup>3</sup>, R. Casale<sup>1</sup>, M. Occhipinti<sup>4</sup>, A. Jochems<sup>1</sup>, H. C. Woodruff<sup>1,3</sup>, P. Lambin<sup>1,3</sup>

1 The D-Lab, Department of Precision Medicine, GROW – School for Oncology and Developmental Biology, Maastricht University, Maastricht, The Netherlands

2 GIGA Cyclotron Research Centre in vivo imaging, University of Liège, Liège, Belgium

3 Department of Radiology and Nuclear Medicine, GROW – School for Oncology and Developmental Biology, Maastricht University Medical Centre+, Maastricht, The Netherlands

4 Radiomics, Clos Chanmurly 13, 4000 Liege, Belgium

## Author contributions statements

All research was performed by W. Rogers. The main manuscript text was written by W. Rogers and S. Keek. All figures were created by W. Rogers. Source code and model architecture was created by W. Rogers, with additional source code provided by S. Primakov. All authors reviewed the manuscript and participated in a qualitative assessment trial. The project was supervised by P. Lambin.

## Address for correspondence

Philippe Lambin, MD, Ph.D.  
Department of Precision Medicine  
Maastricht University  
Minderbroedersberg 4-6,  
6211 LK Maastricht, The Netherlands  
Email: philippe.lambin@maastrichtuniversity.nl

## Keywords

generative modelling, PixelCNN, medical imaging, interpolation, CT

## **Acknowledgements**

Authors acknowledge financial support from ERC advanced grant (ERC-ADG-2015 n° 694812 - Hypoximmuno), ERC-2020-PoC: 957565-AUTO.DISTINCT, Authors also acknowledge financial support from EUROSTARS (DART, DECIDE), the European Union's Horizon 2020 research and innovation programme under grant agreement: ImmunoSABR n° 733008, MSCA-ITN-PREDICT n° 766276, CHAIMELEON n° 952172, EuCanImage n° 952103, TRANSCAN Joint Transnational Call 2016 (JTC2016 CLEARLY n° UM 2017-8295) and Interreg V-A Euregio Meuse-Rhine (EURADIOMICS n° EMR4). This work was supported by the Dutch Cancer Society (KWF Kankerbestrijding), Project number 12085/2018-2

# Abstract

Quantitative analysis models are used for various medical imaging tasks such as registration, classification, object detection, and segmentation which all benefit from high-quality imaging. We propose PixelMiner, a convolution-based deep-learning model which interpolates computed tomography (CT) imaging slices while better preserving quantitative imaging features than standard interpolation methods.

PixelMiner was designed to produce texture-accurate slice interpolations by trading off pixel accuracy for texture accuracy using a novel architecture. PixelMiner was trained on a large dataset consisting of 7,829 lung CT scans and validated using external lung CT datasets. We demonstrated the model's effectiveness by using edge preservation ratio (EPR), compared texture features commonly used in radiomics, and performed a qualitative assessment by human trial.

PixelMiner had the highest EPR 82% ( $p < .01$ ) of the time compared to the closest competing method. It had the lowest texture error, using a normalized root mean squared error (NRMSE) of 0.11 ( $p < .01$ ), with the highest reproducibility of concordance correlation coefficient (CCC)  $\geq 0.85$  ( $p < .01$ ). PixelMiner was chosen 72% of the time by human evaluation ( $p < .01$ ).

PixelMiner was not only demonstrated quantitatively to have improved structural and textural constructions but also shown to be preferable qualitatively.

# 1. Introduction

Radiological images are an important prognostic tool used to inform medical scientists and medical professionals alike. Radiological imaging is primarily used through qualitative assessment by physicians. However, there have been recent advancements in image acquisition, standardization, and analysis that has enabled the discovery of quantitative biomarkers, for example through radiomics [1]. Quantitative image analysis provides the means not only to diagnose oncological disorders [2-4], but also to personalize treatment [5] without the use of invasive procedures.

Quantitative image analysis is dependent on the quality of the image. Image quality is determined by an image's contrast and spatial resolution, as well as the level of noise and the presence of artifacts. The image quality affects the ability to discern different structures and low-frequency signals within the image [6]. A low image quality may therefore have a negative effect on quantitative image analysis [7]. Computed Tomography (CT) scans are acquired and reconstructed with a broad scope of imaging specific parameters which can affect image quality, including differences in reconstruction kernel [8] and slice spacing. Different CT reconstruction methods cause slice spacings to vary considerably creating challenges using predictive analytics. Differences in pixel size and slice spacing have for example been shown to contribute to the stability, robustness, and repeatability of radiomic features [9]. Reproducing features faithfully contributes to better performing object detection and classification models that rely on deep learning to make accurate predictions.

To overcome issues with heterogeneity, CT scans are often interpolated to a common slice spacing. Commonly used interpolation methods include nearest neighbors (NN), trilinear, and tricubic interpolation [10]. Medical scans can be thought of as continuous latent images, implying there is still much uncaptured information between slices that can be predicted. Progressing through the slices various structures move, grow, or rotate at different rates which a model needs to anticipate. Common interpolation models cannot anticipate these structural changes, which cause artifacts to appear on the interpolated slices, such as blurring or ghosting [11]. These artifacts have been shown to lead to errors in image registration [12]. Furthermore, during interpolation high frequency information can be lost [13], and signal reproductions can be too low which leads to distortions [14]. Texture information is represented in a wide range of spatial frequencies which are very important for image classification [15]. These studies suggest it is important to use an algorithm capable of capturing frequency signal information to generate accurate synthetic textures.

The aim of this research is to improve the quality of medical imaging using interpolation to allow for more homogeneous datasets through preprocessing. The proposed method is a deep

learning convolutional-based architecture called PixelMiner. PixelMiner miner is an autoregressive model that predicts a slice between two adjacent slices by taking in a three dimensional input including the target slice, where the prediction loss is determined by the difference between the target slice and a single interpolated slice. We hypothesize that our approach has major advantages over other methods as it is autoregressive, and as such it is able to include information about the interpolated slice in its predictions, in contrast to other models that rely solely on surrounding slices. We also hypothesize that because textures can be well summarized by their component spatial frequencies [15], a model designed specifically for processing signals will achieve improved performance. PixelMiner was compared against other existing interpolation methods by comparing the accuracy of texture generation using standard metrics and through radiomics feature analysis.

## 2. Methods

### 2.1 Data

The dataset used for training is a publicly available dataset from the Radiological society of North America (RSNA) Pulmonary Embolism Detection Challenge © RSNA 2020 [16]. This dataset consists of 7,829 chest CT scans, with pixel spacing and slice spacing of 1mm and 1.25mm, respectively. The dataset's terms and conditions allow the dataset to be shared or re-distributed in any form. For more information see <https://www.rsna.org/education/ai-resources-and-training/ai-image-challenge/rsna-pe-detection-challenge-2020>. There were five institutions contributing data, including AlfredHealth, Koç University Hospital, Center for Artificial Intelligence in Medicine & Imaging (AIMI) at Stanford, Unity Health Toronto, and Universidade Federal de São Paulo. All CT scans were retrospectively collected and anonymized at each institution and approved by the respective institutional review boards in accordance with relevant guidelines and regulations.

The dataset used for validation is the publicly available dataset from the Early Lung Cancer Action Program (ELCAP) Public Lung Image Database [17]. This database consists of 50 low-dose CT scans, with slice spacing of 1.25mm. These datasets' terms and conditions allow use for non-commercial purposes only, including academic research and education. All CT scans were collected by ELCAP in collaboration with Weill Cornell Medical College and approved by its institutional review board in accordance with relevant guidelines and regulations.

### 2.2 Model

The proposed model, PixelMiner, is a causal autoregressive model based on PixelCNN++ [18, 19]. PixelMiner includes many of the features of PixelCNN++, including vertically and horizontally stacked masked convolutions, gated convolutions, and the logistic mixture likelihood loss. The model was trained using batches of three sequential slices as inputs to generate a single output slice. The middle input slice was used to evaluate the output slice so that the model learns to replicate the input slice. The practice of learning to replicate slices of high resolution scans can be used at generation time to interpolate slices in scans with larger slice spacings.

PixelCNN based models were designed to generate RGB images and cannot be adapted to perform slice interpolation due to stability issues, such as warping and artifacts. Its constraint in slice interpolation is its use of 2D masked convolutions. The PixelMiner architecture uses a custom mixed causal and non-causal convolutional block. PixelMiner also uses a combination of 2D and 3D convolutions, where the 3D convolutions are used to learn the relationships between the slices. This architecture creates a pathway for non-masked non-causal filters in the outer channels using only the external slices, while using masked causal filters in the middle. The causal channel receives information from both the middle and external slices. This allows the outer slices to provide complete information to the middle, and prevents the 2D convolutions tendency to ignore the outer slices. A schematic overview of this convolutional block is provided in figure 1A.

PixelMiner further utilizes residual blocks, each of which consists of six downsampling and upsampling layers which utilize long range skip connections. Feature maps are down-sampled using 2 stride PixelMiner convolutions from 64x64 to 32x32 to 16x16, and then upsampled again using transposed convolutions. The final model is made with the input convolutional layer,  $n$  residual blocks, and an output convolutional layer using an identity convolution to combine the final output. A schematic depiction of the architecture can be found in figure 1B.

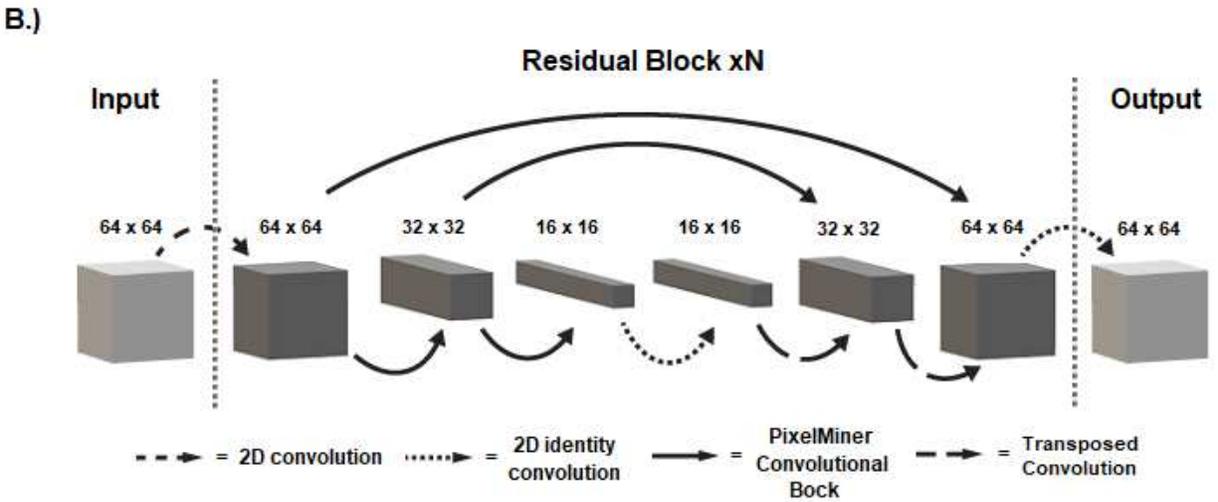
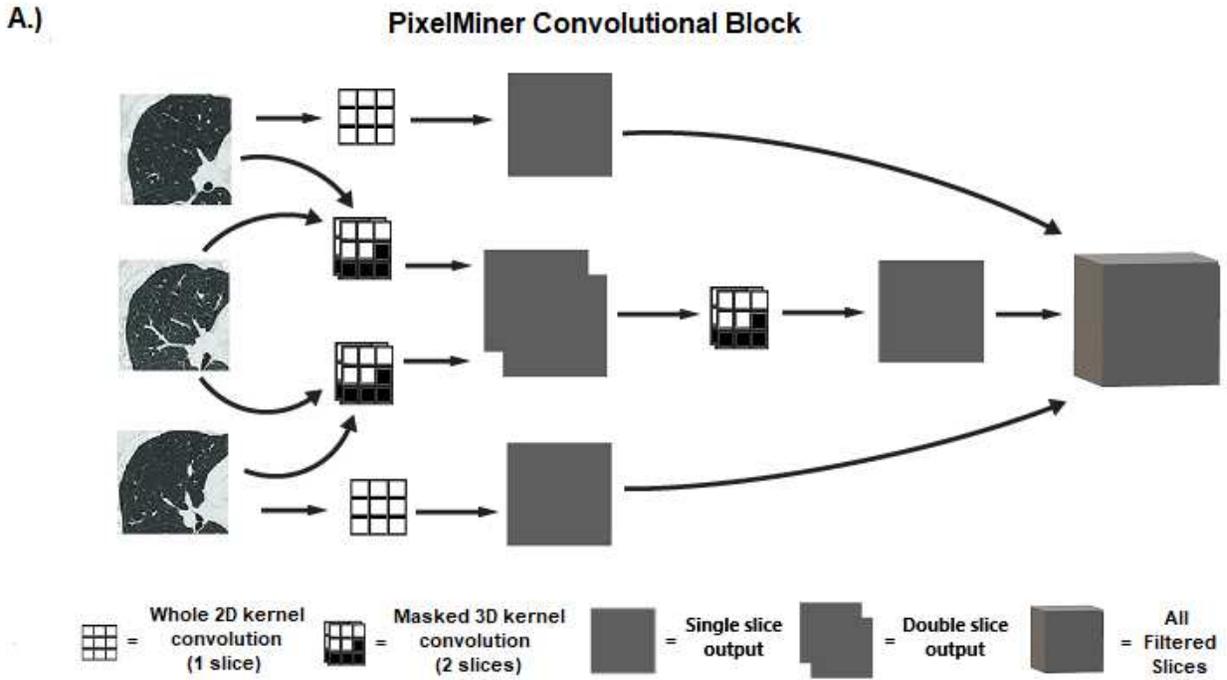


Figure 1: Diagram A is a schematic depiction of the PixelMiner convolutional block. Diagram B is the entire architecture where the input goes through a 2D convolution and the output goes through an identity convolution. The architecture is made up of N residual blocks with two down samples using two stride convolutions and two up samples using transposed convolutions .

## 2.3 Generation

The model upsamples a scan using two slices at a time. To begin, the first two sequential slices are selected, a top slice and a bottom slice. An empty slice is then inserted between the top and bottom slices. These three slices are then provided to the model. The model then generates an interpolated slice pixel by pixel until the entire slice is complete, and repeats the process until every slice of the scan is interpolated. Figure 2 provides a graphic example of the completed upsampling technique.

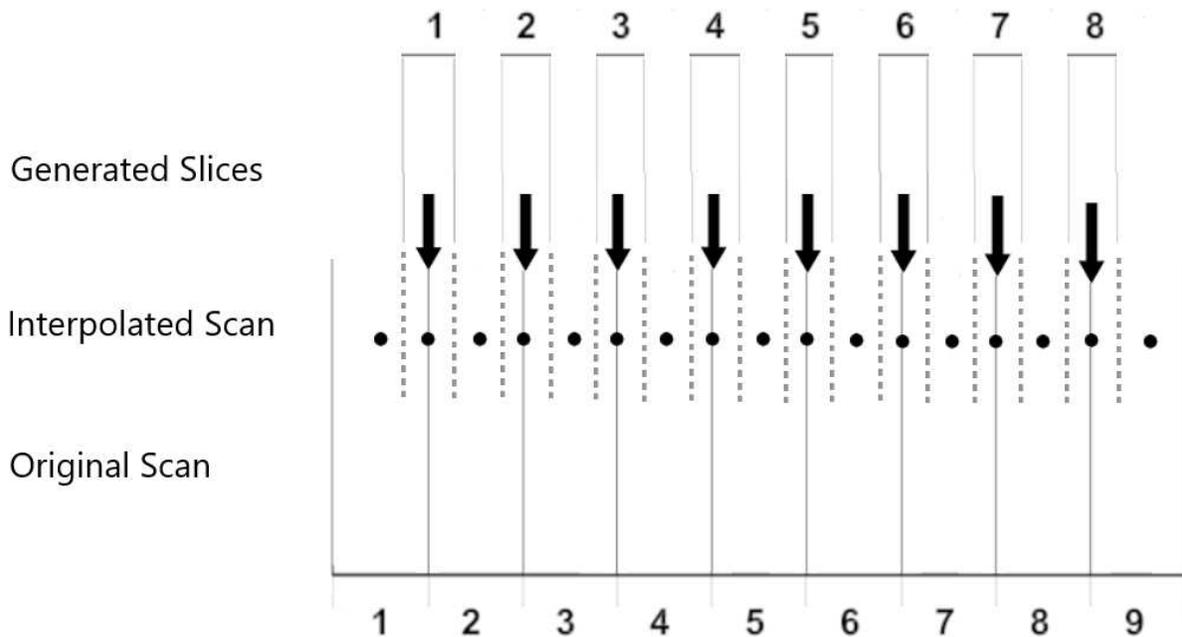


Figure 2: A graphic representation of the interpolation technique. In this example, a scan with nine slices with large slice spacings is upsampled with eight additional slices. The dots represent the center of each slice, and dotted lines represent new borders following interpolation, and the arrows represent the generated slices being inserted.

With the resolution of a single slice being 512 x 512 pixels, it is not feasible to use entire slices due to memory limitations. However, because PixelMiner is an autoregressive model, it is capable of generating slices in patches. Therefore, patches of 64 x 64 were used for training and slice interpolation. Patches are generated by beginning at the top left corner, and then expanding out based on the initial patch. Autoregressive models require enough observations to capture a signal to make optimal predictions [20]. With this in mind, PixelMiner uses patches of 64 x 64 with incomplete 32 x 32 sub-patches. An example of this process is given in figure 3.

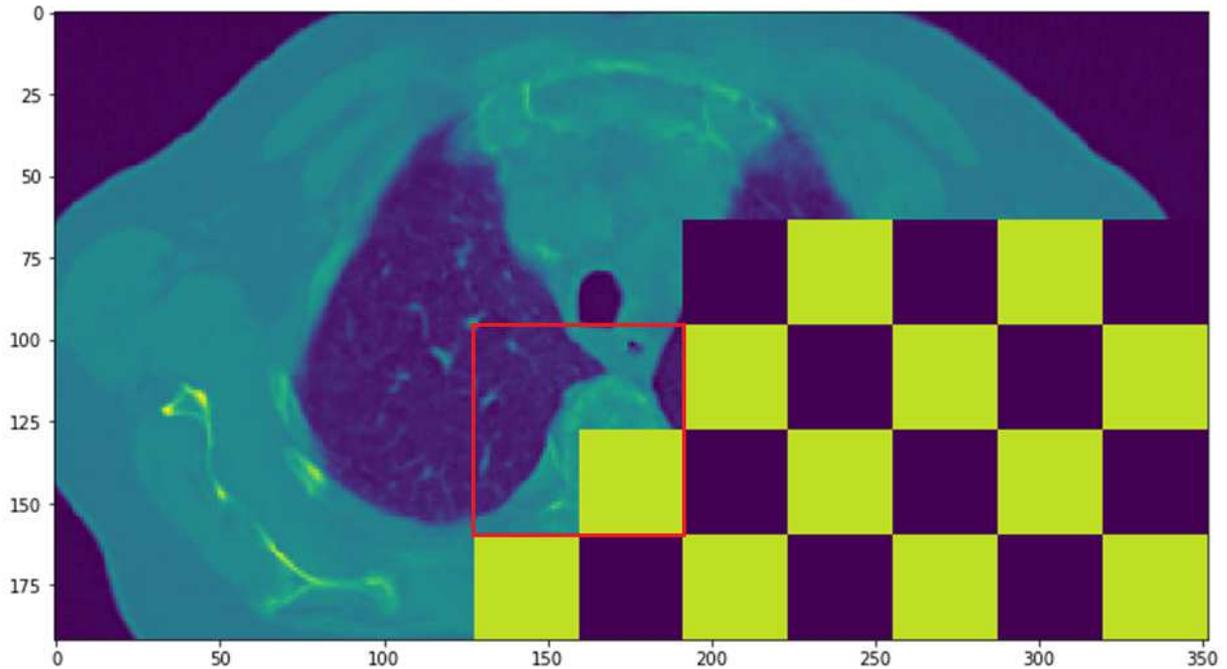


Figure 3: An example of an interpolated slice being generated in patches. The region contained in the red box indicates the next patch of 64 x 64 to be fed into the model. The yellow patch of 32 x 32 in the bottom right of the red box is the sub-patch of pixels that are to be generated next.

## 2.4 Training and Validation

Training is performed in an unsupervised fashion with the objective being of reproducing the middle input slice without the need for labels. The model receives three randomly selected slices in sequence and outputs a single slice, which is then used together with the middle input slice to calculate the loss. Figure 4 provides an overview of the model training, testing, and validation steps.

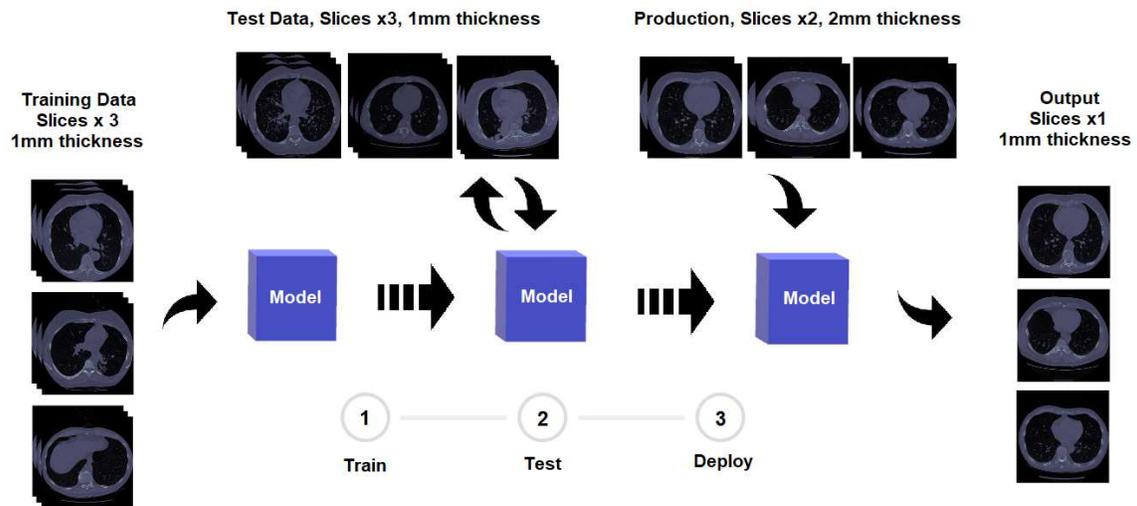


Figure 4: Overview of the model training, testing, and validation steps. As the model is unsupervised, there is no need for labels. It is trained with three inputs, where the middle slice is what the model learns to interpolate. At test time the model receives two inputs and produces a single output, where the output is compared to the ground truth. The production model takes in two inputs and outputs the interpolated slice. The output slices are combined together with the original slices to create the final upsampled scan.

At generation time for the proposed model, the interpolated synthetic slices and ground truth slices were compared for evaluation. In addition, synthetic slices were generated using other interpolation methods, including (tri-)linear [21, 22], (tri-)cubic [23, 24], window sinc (WS) [25], and NN [26] interpolation. Linear and BSpline (cubic) interpolation are methods commonly used in radiomic studies [27]. Windowed sinc and NN were done using the open source project the Insight Toolkit or ITK [28] with tools designed specifically for medical imaging. Figure 5 provides examples of interpolated slices.

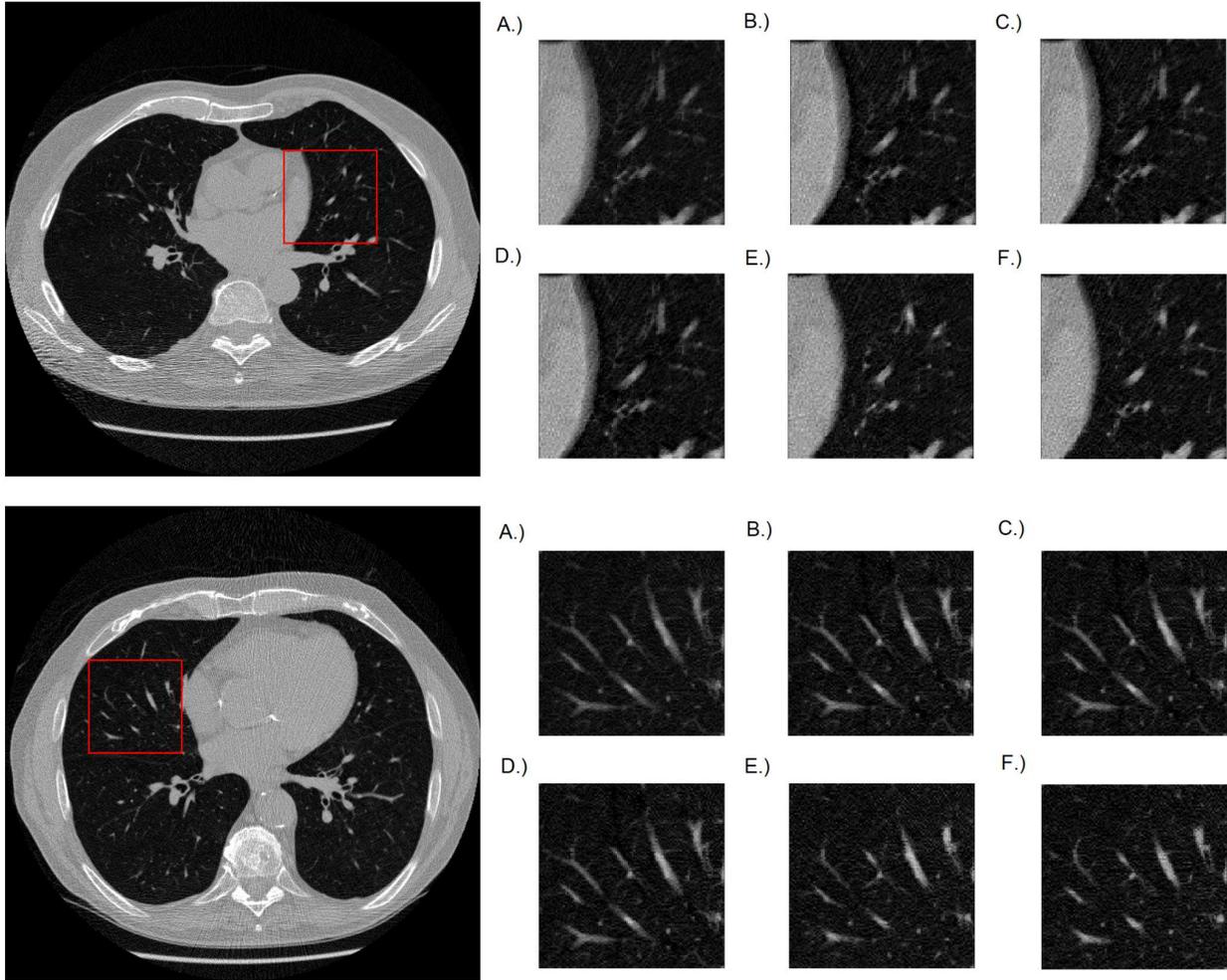


Figure 5: Five interpolation methods that are enlarged and compared side by side for two example slices. Scans to the left are the slices used and the red box indicates the region used for comparison. Methods include A.) linear, B.) windowed sinc, C.) nearest neighbor D.) BSpline and E.) PixelMiner (ours) with the corresponding F.) ground truth slice on the bottom right.

## 2.5 Evaluation

### 2.5.1 Edge Preservation Ratio

The edge preservation ratio (EPR) was used to evaluate the accuracy and sharpness of the shapes produced for all interpolation methods. EPR has been shown to outperform common metrics peak signal to noise ratio (PSNR) and structural similarity index (SSIM) in both gaussian blurring tests and human perception [29]. EPR uses a Canny edge detector to produce a binary map of an image and its reference. It then takes the union of pixels between the two images and divides by the total number pixels with highlighted edges from the referenced image. The EPR of each slice was determined and compared across interpolation methods. EPR results were

statistically tested using a binomial test to test frequency of higher ratios and a Wilcoxon signed rank test to test pairs of ratios for competing methods on the test dataset.

## 2.5.2 Texture Features

To test the interpolation methods' ability to retain texture information, gray-level co-occurrence matrix (GLCM), gray-level run length matrix (GLRLM), and gray-level size zone matrix (GLSZM) texture features were extracted. Feature extraction was performed in Python 3.7 with Pyradiomics 3.0.1 [30]. Features were extracted from the synthetic and ground truth slices to act as reference. Masks of regions of interest (ROI) were used for feature extraction to avoid subjectivity. A single image mask was created for the area around the lungs. A simple automated segmentation algorithm was created using thresholding and binary morphology operations to dilate or erode or close openings in the mask using the morphology module in the Scikit-image package [31]. Lungs were segmented using a threshold  $< -320$  Hounsfield units. Using the segmented ROI, Features were extracted and compared for synthetic slices and the ground truth using the NRMSE and concordance correlation coefficient (CCC). The CCC is concordance between paired ground truth values and test values. The statistic quantifies the agreement and reproducibility, shown as:

$$\mathbf{E}[(Y - X)^2] = (\mu_y - \mu_x)^2 + \sigma_x^2 + \sigma_y^2 - 2p\sigma_y\sigma_x$$

Where  $\mu_y$  and  $\mu_x$  are the means and  $\sigma_y^2$  and  $\sigma_x^2$  are the variances of the reference and tested slice features, respectively, and  $p$  is the Pearson correlation coefficient.

The mean squared mapped feature error (MSMFE) is a newly proposed metric. For each texture feature a feature map was generated using a sliding window with a size of 5x5 and step of 1. The RMSE is calculated on the reference ground truth feature map and the feature map of the evaluated scan. The final error is the NRMSE of all texture features. To test for statistical significance, the Wilcoxon signed rank test was used on all paired results. Furthermore, all lowest error rates were tracked and counted to be compared to PixelMiner, which was tested using a binomial test.

## 2.5.3 Qualitative Assessment: *an In Silico* Clinical trial

For qualitative assessment of the generated slices we set up an in silico clinical trial. This trial was performed using a newly created application that randomly gives a test subject three slices, of which one is the ground truth and the other two are corresponding synthetically generated slices. The two interpolated slices contained one generated by the PixelMiner model and one generated by another random interpolation method. Users were given 12 randomly selected slices from 50 different scans, which includes each competing interpolation method at least once. For each slice, the users had to choose which synthetic slice better represents the ground truth slice. To show the results were not a statistical anomaly, a binomial test was performed to test statistical significance.

## 3. Results

### 3.1 Edge Preservation Ratio

Table 1 provides an overview of the EPR values for each tested interpolation method. PixelMiner had higher EPR values for 82% ( $p < .01$ ) of the scans in the validation dataset compared to the next best performing interpolation method Windowed Sinc. In addition, PixelMiner had the highest mean EPR of 0.489 compared to other methods ( $p < .01$ ).

Table 1: Mean and standard deviation of the EPR values, and the percentage of scans an interpolation had a higher EPR (horizontally) and a lower EPR (vertically) compared to the other interpolation methods.

	EPR		Percent of scan with a higher EPR				
	Mean	SD	Linear (%)	Nearest Neighbor	BSpline	Windowed Sinc	PixelMiner
Linear	0.438	$\pm 0.027$	-	0	0	0	<b>0</b>
Nearest Neighbor	0.480	$\pm 0.013$	100	-	10	0	<b>10</b>
BSpline	0.481	$\pm 0.013$	100	90	-	12	<b>12</b>
Windowed Sinc	0.482	$\pm 0.012$	100	100	88	-	<b>18</b>
PixelMiner	<b>0.489</b>	<b><math>\pm 0.010</math></b>	<b>100</b>	<b>90</b>	<b>88</b>	<b>82</b>	-

### 3.2 Segmentation Extracted Texture Feature Outcomes

PixelMiner had the lowest average NRMSE at 0.109 ( $p < .01$ ). PixelMiner had a lower NRMSE value for 62.2% ( $p < .01$ ) of the scans in the validation dataset compared to the next best performing interpolation method Windowed Sinc. PixelMiner also had the highest amount of reproducible features (74.5% with  $CCC \geq 0.85$ ), and the highest mean CCC at 89.8. An overview of these metrics for the tested interpolation methods can be found in Table 2.

Table 2: Average RMSE values for the tested interpolation methods, and the percent of slices each interpolation method has lower a RMSE (horizontally) and a higher RMSE (vertically) in comparison to the other methods. Also displayed for each interpolation method are the percentage of features with  $CCC \geq .85$ , and the mean CCC.

	CCC		NRMSE		Percent of features with a lower Error				
	Mean	$\geq .85$	Mean	SD	Linear (%)	Nearest Neighbor	BSpline	Windowed Sinc	PixelMiner
Linear	68.8	17.6%	0.342	$\pm 0.267$	-	18.6	17.2	14.7	<b>13.6</b>
Nearest Neighbor	89.0	64.6%	0.143	$\pm 0.133$	81.4	-	52.1	49.1	<b>32.4</b>
BSpline	87.9	64.7%	0.138	$\pm 0.131$	82.8	47.9	-	46.0	<b>32.8</b>
Windowed Sinc	86.7	64.7%	0.136	$\pm 0.133$	85.3	50.9	54.0	-	<b>37.8</b>
PixelMiner	<b>89.8</b>	<b>74.5%</b>	<b>0.109</b>	<b><math>\pm 0.123</math></b>	<b>86.4</b>	<b>67.6</b>	<b>67.2</b>	<b>62.2</b>	-

### 3.3 Mean Squared Mapped Feature Error

Table 3 provides an overview of the MSMFE for each texture feature for the different interpolation method. PixelMiner had a lower average MSMFE compared to the other interpolation methods at 0.619 ( $p < .01$ ). For 68.19% of the features, PixelMiner had the lowest MSMFE compared with the next best interpolation method ( $p < .01$ ).

Table 3: The average MSMFE for all GLCM features with  $p < .01$ . Alongside it is the reported percent of texture features for which the listed interpolation method had a lower MSMFE (horizontally) and higher MSMFE (vertically) in comparison to the other interpolation methods.

	MSMFE		Percent of features with the lowest average error				
	Mean	SD	Linear (%)	Windowed Sinc	Nearest Neighbor	BSpline	PixelMiner
Linear	0.507	$\pm 0.236$	-	36.3	36.5	36.4	<b>34.8</b>
Window Sinc	0.463	$\pm 0.217$	63.7	-	44.1	40.7	<b>31.5</b>
Nearest Neighbor	0.462	$\pm 0.217$	63.5	55.9	-	38.6	<b>31.2</b>
BSpline	0.462	$\pm 0.217$	63.6	59.3	61.4	-	<b>31.0</b>
PixelMiner	<b>0.434</b>	$\pm 0.222$	<b>65.2</b>	<b>68.5</b>	<b>68.8</b>	<b>69.0</b>	-

### 3.4 Qualitative Assessment

The application was presented to 16 participants. As some participants did not grade all 12 presented slices, a total of 140 comparisons were collected. The participants included medical professionals, radiologists, radiation oncologists, computer scientists, and students, who were tracked during the trial using a custom built application for performing comparison tests. PixelMiner was chosen on average 72% ( $p < 0.01$ ) of the time, and no other model was chosen as frequently. PixelMiner was chosen over other metrics Nearest Neighbor, BSpline, Windowed Sinc, and Linear 68.2%, 70%, 75%, and 96.3% of the time ( $p < 0.05$ ), respectively.

## 4. Discussion

The goal of this study was to create a model that can perform slice interpolation with a high degree of accuracy compared to other interpolation methods. We have shown that PixelMiner successfully outperforms linear, NN, BSpline, and Windowed Sinc interpolation methods by every metric used to evaluate the methods. PixelMiner has been shown to, on average, have lower error, higher reproducibility, and higher qualitative performance.

PixelMiner had a significantly lower NRMSE for compared to all other models, the lowest for 65.2% of the features compared to the next best method. Furthermore, the reproducibility of PixelMiner using a  $CCC \geq 0.85$  was shown to be 80% with a mean of 91.42, greater than all

other models. Furthermore, the average edge preservation ratio of 0.489 was found to be significantly higher than the other tested methods. If the features for a particular radiomics signature are known it would be worth doing some research beforehand to determine which interpolation method would complement the model being developed first. If the signature is unknown, or a practitioner is using deep learning, the results indicate PixelMiner to be the best choice for slice interpolation.

The approach is quite novel and it is difficult to compare to similar studies. PixelCNNs have been used to perform super resolution on 2d images [32], but to the best of our knowledge it has never been used on 3d images and it's the first time it has been used to generate high fidelity images. Deep learning based super resolution has been an active area of research and includes many models such as SRGAN [33], MFTV [34], FSRCNN [35], SRResNet-V54 [36], LapSRN [37], and multiple dense residual block based GANs [38]. These types of models are designed for 3d images by doing grid based upsampling using transposed convolutions. They are unable to be compared directly with PixelMiner since they would require downsampling in 3 dimensions to be evaluated, causing the models to lose valuable information that slice interpolation is able to retain. These super resolution models could also potentially be used in conjunction with PixelMiner, by upsampling after slices have already been interpolated.

Quantitative analysis in medical imaging relies on high quality images that are harmonious across many types of scanners. There is a potential for quantitative analysis of medical imaging to have a profound effect on prognosis, but it requires high quality data for training machine learning models. Medical imaging slice spacing is only one factor in the overall quality of images but it affects many different areas of quantitative analysis, such as image registration, detection, segmentation, and classification.

A disadvantage to using PixelMiner is that it is slower compared to other interpolation methods. Using a single 1080ti graphical processing unit (GPU), it could take a full 24 hours to interpolate all of the slices of a high resolution 2mm slice spacing scan. Future improvements could be made through better parallelization of GPUs at generation, but also through more efficient models using methods similar to MobileNet [39] or EfficientNet [40]. Furthermore, advancements in graphical processing units continue to grow exponentially, and modern GPUs could considerably bring down generation times.

PixelCNN++ is known for being affected by long range dependence, which makes pixels that are farther apart more difficult to predict. This could be partially alleviated by using self attention which showed modest improvements in the PixelSnail model [41]. Self attention was omitted from the current version of PixelMiner due to time constraints. With the development of the attention based model called a transformer [42], it could be possible to build an improved version of PixelMiner similar to the transformer based image generation model ImageGPT [43]. Transformers perform faster than RNNs but also address the issues with long range dependencies in PixelCNN and PixelRNN. Transformer based image models are a relatively

new development in computer vision models, but have the potential to push the capabilities of PixelMiner much further.

## 5. Conclusion

PixelMiner provides a new state of the art method for performing slice interpolation on medical imaging capable of improving imaging resolution while better preserving texture features, an important set of features in quantitative medical imaging analysis. PixelMiner performs well not only on the training dataset but also on the validated external datasets.

# References

1. Aerts, H.J.W.L., et al., Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nature Communications*, 2014. 5.
2. Kurland, B.F., et al., Promise and pitfalls of quantitative imaging in oncology clinical trials. *Magnetic Resonance Imaging*, 2012. 30(9): p. 1301-1312.
3. Buckler, A., et al., A Collaborative Enterprise for Multi-Stakeholder Participation in the Advancement of Quantitative Imaging. *Radiology*, 2011. 258: p. 906-14.
4. Buckler, A., et al., Quantitative Imaging Test Approval and Biomarker Qualification: Interrelated but Distinct Activities. *Radiology*, 2011. 259: p. 875-84.
5. Lambin, P., et al., Predicting outcomes in radiation oncology-multifactorial decision support systems. *Nature reviews. Clinical oncology*, 2012. 10.
6. Goldman, L.W., Principles of CT: Radiation Dose and Image Quality. *Journal of Nuclear Medicine Technology*, 2007. 35: p. 213-225.
7. Zhao, B., James, L. P., Moskowitz, C. S., Guo, P., Ginsberg, M. S., Lefkowitz, R. A., Qin, Y. Riely, G.J., Kris, M.G., Schwartz, L. H., Evaluating Variability in Tumor Measurements from Same-day Repeat CT Scans of Patients with Non-Small Cell Lung Cancer 1. *Radiology. Radiological Society of North America (RSNA)*. 2009.
8. Choe, J.A.-O., et al., Deep Learning-based Image Conversion of CT Reconstruction Kernels Improves Radiomics Reproducibility for Pulmonary Nodules or Masses. (1527-1315 (Electronic)).
9. Mackin, D., et al., Correction: Harmonizing the pixel size in retrospective computed tomography radiomics studies. *PLOS ONE*, 2018. 13: p. e0191597-e0191597.
10. Whybra, P., et al., Assessing radiomic feature robustness to interpolation in 18F-FDG PET imaging. *Scientific Reports*, 2019. 9.
11. Leng, J., G. Xu, and Y. Zhang, Medical image interpolation based on multi-resolution registration. *Computers & Mathematics with Applications*, 2013. 66(1): p. 1-18.
12. Josien, P.W.P., J.B.A. Maintz, and A.V. Max. Mutual information matching and interpolation artifacts. in *Proc.SPIE*. 1999.
13. Parker, J.A., R.V. Kenyon, and D.E. Troxel, Comparison of Interpolating Methods for Image Resampling. *IEEE Transactions on Medical Imaging*, 1983. 2(1): p. 31-39.
14. Crow, F.C., The aliasing problem in computer-generated shaded images. *Commun. ACM*, 1977. 20(11): p. 799-805.
15. Strand, J. and T. Taxt, Local frequency features for texture classification. *Pattern Recognition*, 1994. 27(10): p. 1397-1406.
16. Colak, E., et al., The RSNA Pulmonary Embolism CT Dataset. *Radiology. Artificial intelligence*, 2021. 3(2): p. e200254-e200254.
17. Henschke CI, M.D., Yankelevitz DF, et al, Early Lung Cancer Action Project. *Lancet*, 1999. 354: p. 1205.
18. van den Oord, A., et al., Conditional Image Generation with PixelCNN Decoders. *CoRR*, 2016. abs/1606.05328.

19. Salimans, T., et al., PixelCNN++: Improving the PixelCNN with Discretized Logistic Mixture Likelihood and Other Modifications. International Conference on Learning Representation, 2017.
20. Lehmann, E.L. and G. Casella, Theory of point estimation. 2nd ed. Springer texts in statistics. 1998, New York: Springer. xxvi, 589 p.
21. Myers, D.E., Kriging, cokriging, radial basis functions and the role of positive definiteness. Computers & Mathematics with Applications, 1992. 24(12): p. 139-148.
22. Stytz, M.R. and R.W. Parrott, Using kriging for 3d medical imaging. Computerized Medical Imaging and Graphics, 1993. 17(6): p. 421-442.
23. Keys, R., Cubic convolution interpolation for digital image processing. IEEE Trans Acoust Speech Signal Process. Acoustics, Speech and Signal Processing, IEEE Transactions on, 1982. 29: p. 1153-1160.
24. Hou, H. and H. Andrews, Cubic splines for image interpolation and digital filtering. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1978. 26(6): p. 508-517.
25. Harris, F.J., On the use of windows for harmonic analysis with the discrete Fourier transform. Proceedings of the IEEE, 1978. 66(1): p. 51-83.
26. Lehmann, T.M., C. Gonner, and K. Spitzer, Survey: interpolation methods in medical image processing. IEEE Transactions on Medical Imaging, 1999. 18(11): p. 1049-1075.
27. Whybra, P., et al., Assessing radiomic feature robustness to interpolation in 18F-FDG PET imaging. Scientific Reports, 2019. 9(1): p. 9649.
28. Johnson, H.J.a.M., Matthew M and Ibanez, Luis, The ITK Software Guide Book 1: Introduction and Development Guidelines-Volume 1. Kitware, Inc., 2015.
29. Yu, S., et al. Applications of edge preservation ratio in image processing. in 2014 12th International Conference on Signal Processing (ICSP). 2014.
30. van Griethuysen, J.J.M., et al., Computational Radiomics System to Decode the Radiographic Phenotype. Cancer research, 2017. 77(21): p. e104-e107.
31. van der Walt S, S.J., Nunez-Iglesias J, Boulogne F, Warner JD, Yager N, Gouillart E, Yu T, the scikit-image contributors, scikit-image: image processing in Python. PeerJ, 2014. 2: p. e453.
32. Dahl, R., M. Norouzi, and J. Shlens, Pixel Recursive Super Resolution. 2017.
33. Ledig, C., et al., Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. 2017.
34. Jian, X., et al., Image superresolution by midfrequency sparse representation and total variation regularization. Journal of Electronic Imaging, 2015. 24(1): p. 1-29.
35. Dong, C., C.C. Loy, and X. Tang. Accelerating the Super-Resolution Convolutional Neural Network. in Computer Vision – ECCV 2016. 2016. Cham: Springer International Publishing.
36. Rajan, D. and S. Chaudhuri, Generalized Interpolation for Super-Resolution. 2006. p. 45-72.
37. Lai, W., et al. Deep Laplacian Pyramid Networks for Fast and Accurate Super-Resolution. in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017.

38. Zhang, X., et al., CT super-resolution using multiple dense residual block based GAN. *Signal, Image and Video Processing*, 2021. 15(4): p. 725-733.
39. Howard, A.G., et al., MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *ArXiv*, 2017. abs/1704.04861.
40. Tan, M. and Q. Le, EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. 2019.
41. Chen, X., et al., PixelSNAIL: An Improved Autoregressive Generative Model. 2017.
42. Vaswani, A., et al., Attention Is All You Need. 2017.
43. Chen, M., et al., Generative Pretraining from Pixels. 2020.
44. Mohamed, S. and B. Lakshminarayanan, Learning in Implicit Generative Models. 2017.
45. Goodfellow, I., et al. Generative Adversarial Nets. in *Advances in Neural Information Processing Systems*. 2014. Curran Associates, Inc.
46. Goodfellow, I.J., NIPS 2016 Tutorial: Generative Adversarial Networks. *CoRR*, 2017. abs/1701.00160.
47. Diederik, P.K. and W. Max, An Introduction to Variational Autoencoders. 2019: now. 1.
48. van den Oord, A., N. Kalchbrenner, and K. Kavukcuoglu, Pixel Recurrent Neural Networks. *CoRR*, 2016. abs/1601.06759.
49. Horé, A. and D. Ziou. Image Quality Metrics: PSNR vs. SSIM. in *2010 20th International Conference on Pattern Recognition*. 2010.

# Figures

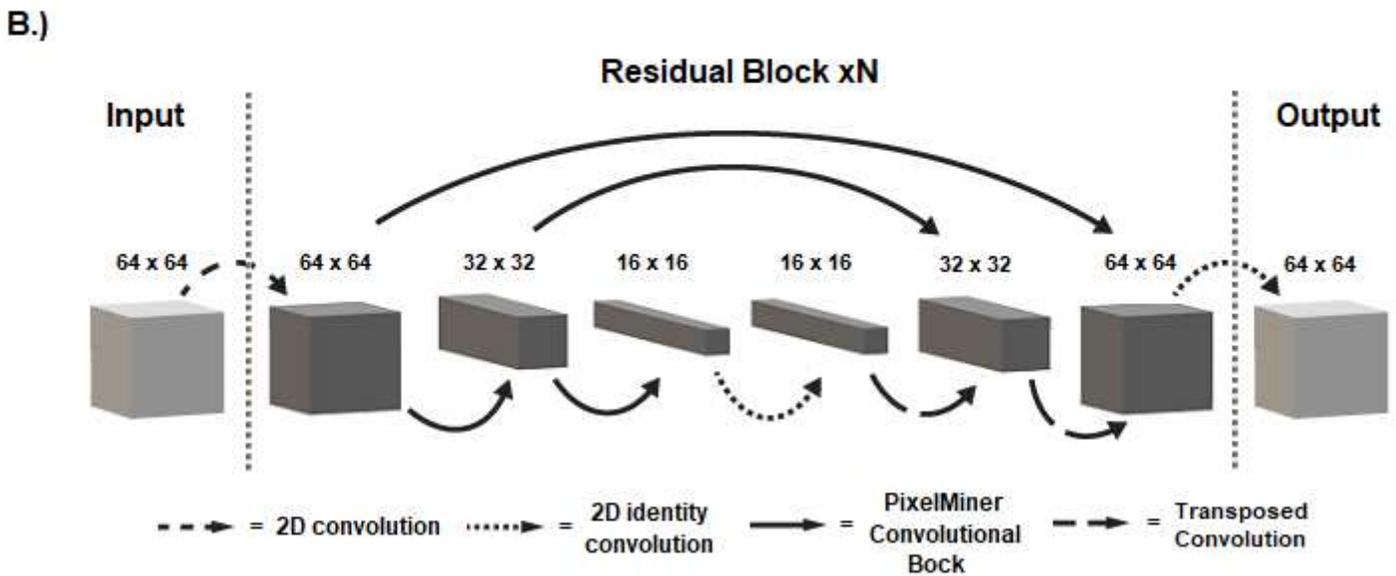
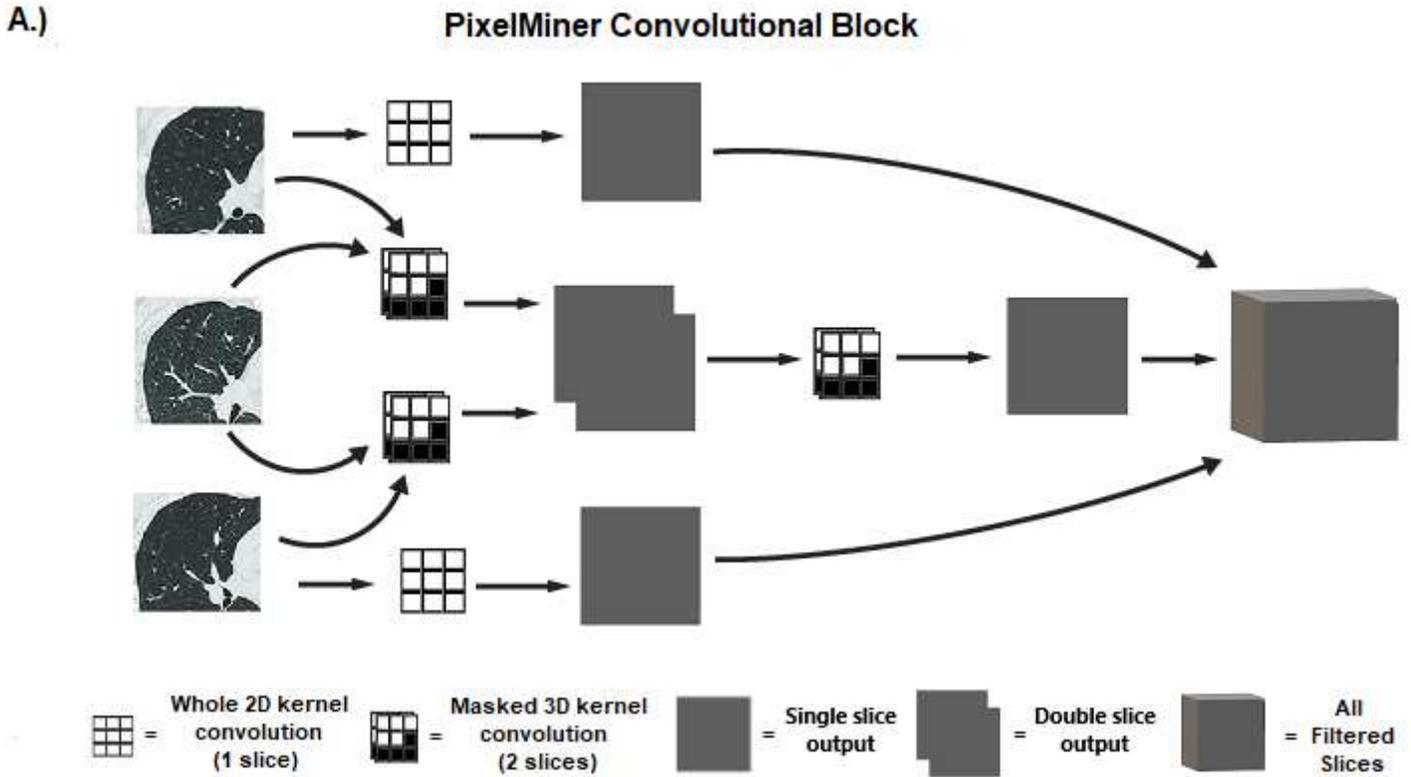
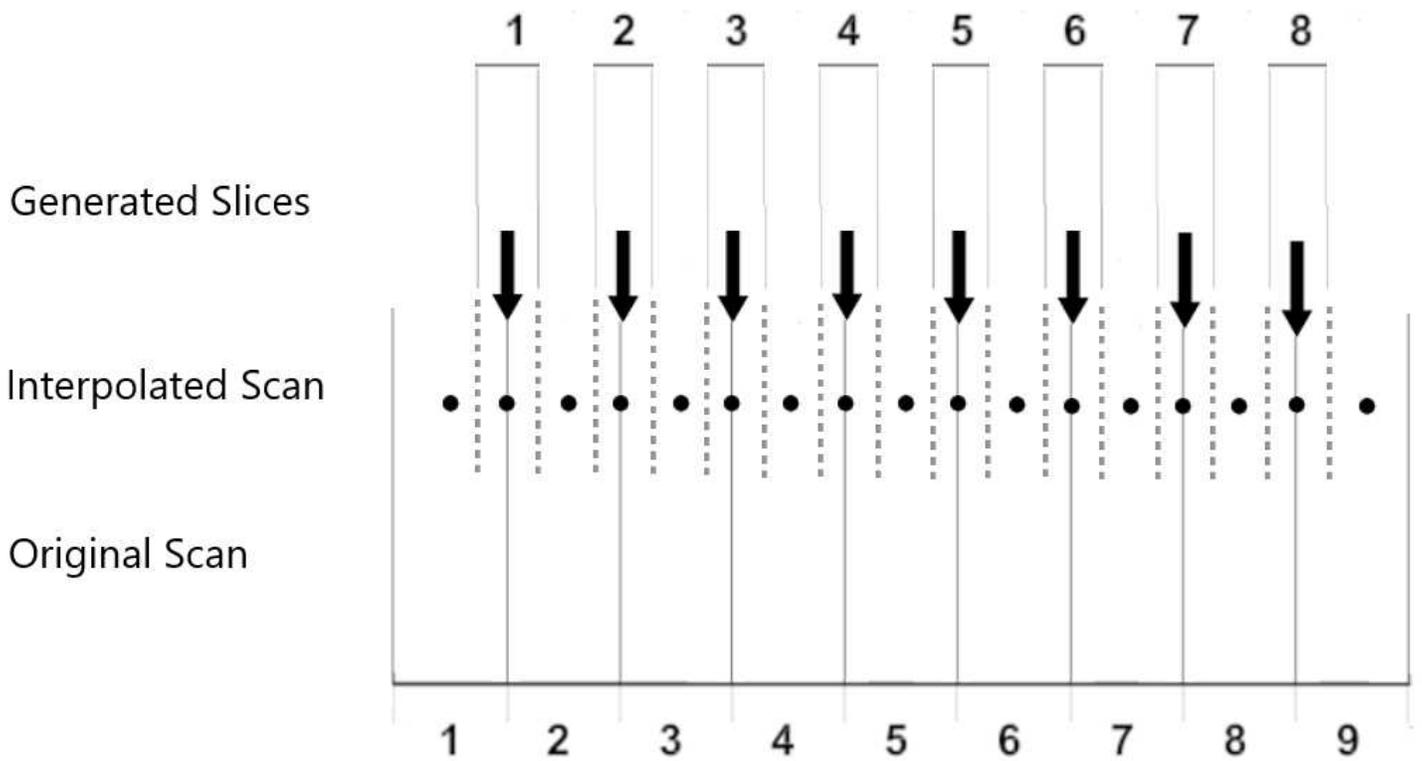


Figure 1

Diagram A is a schematic depiction of the PixelMiner convolutional block. Diagram B is the entire architecture where the input goes through a 2D convolution and the output goes through an identity convolution. The architecture is made up of N residual blocks with two down samples using two stride convolutions and two up samples using transposed convolutions.



**Figure 2**

A graphic representation of the interpolation technique. In this example, a scan with nine slices with large slice spacings is upsampled with eight additional slices. The dots represent the center of each slice, and dotted lines represent new borders following interpolation, and the arrows represent the generated slices being inserted.

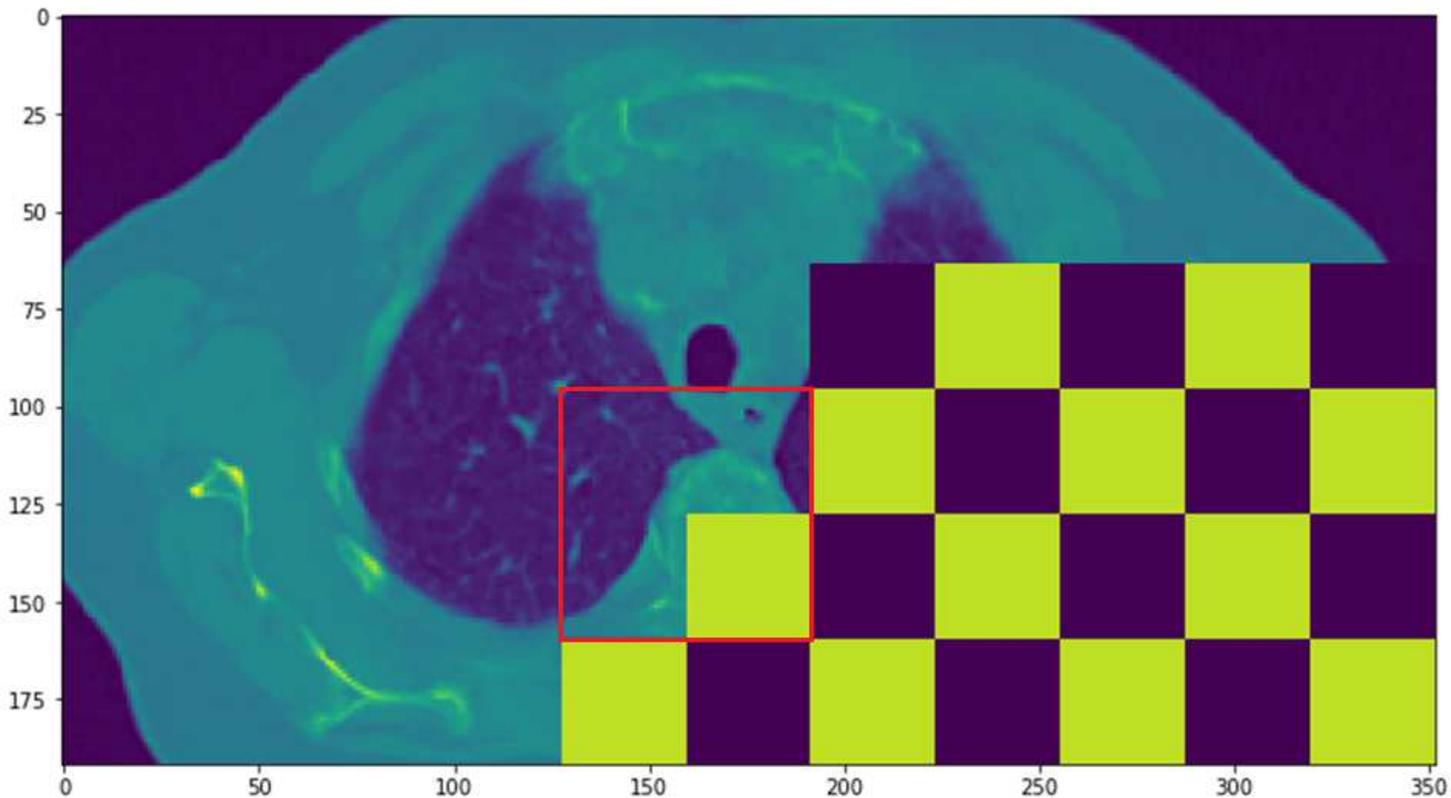


Figure 3

An example of an interpolated slice being generated in patches. The region contained in the red box indicates the next patch of 64 x 64 to be fed into the model. The yellow patch of 32 x 32 in the bottom right of the red box is the sub-patch of pixels that are to be generated next.

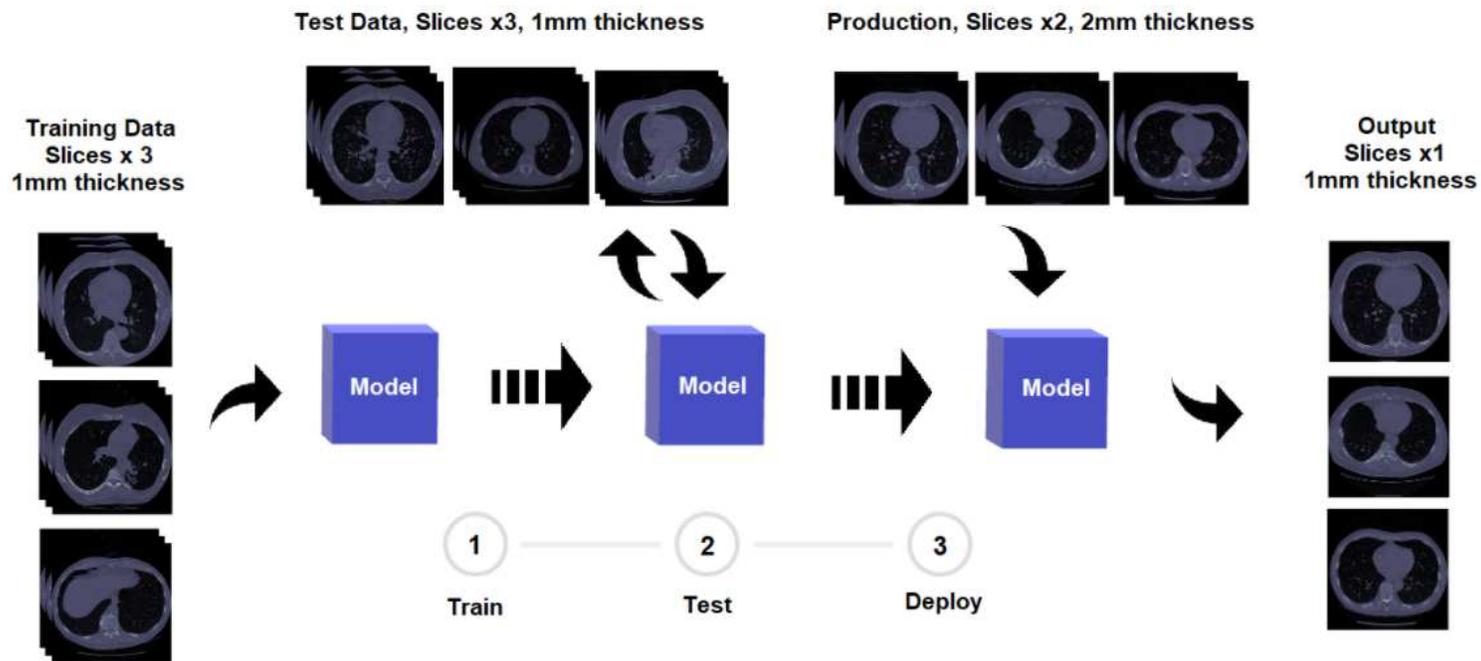
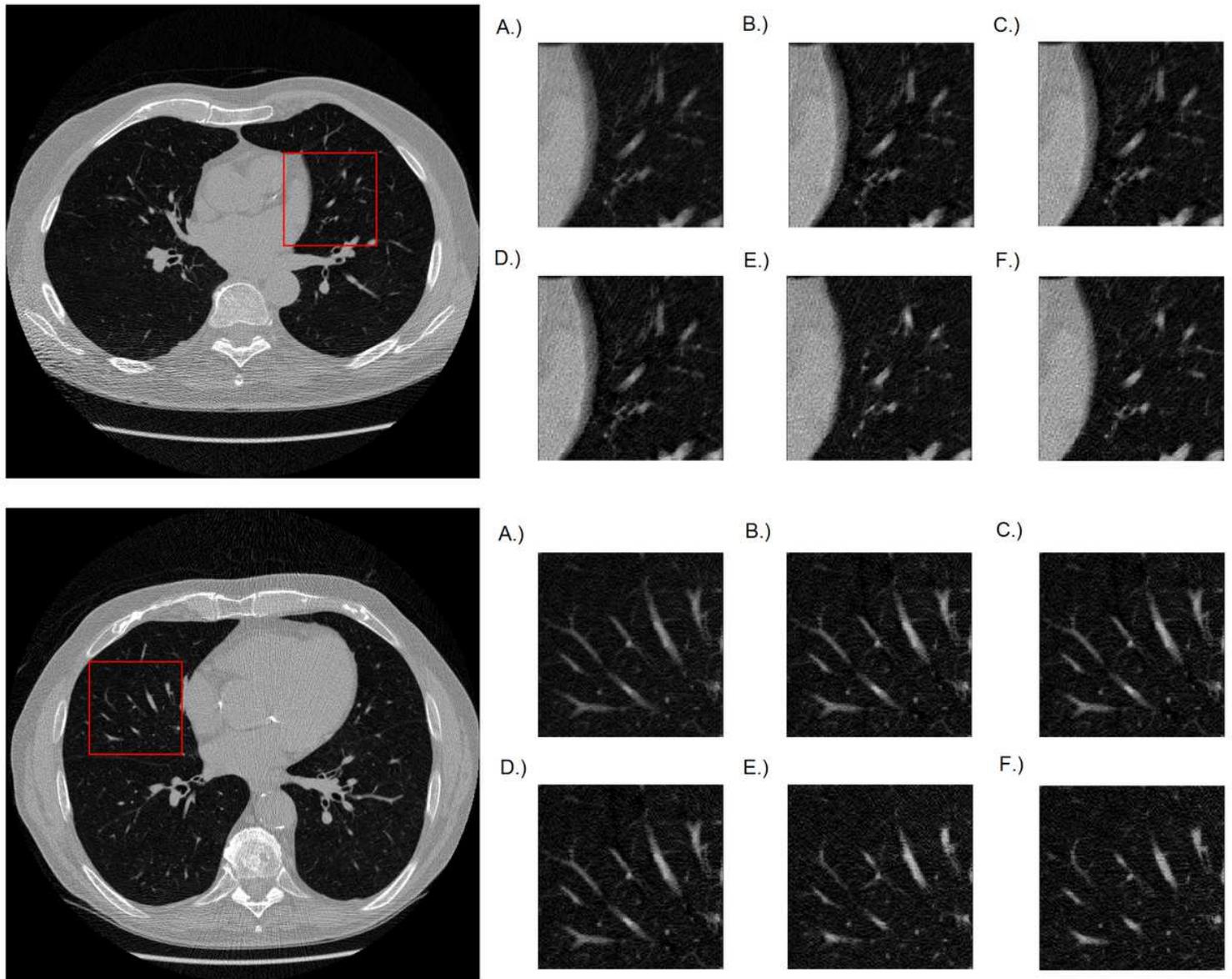


Figure 4

Overview of the model training, testing, and validation steps. As the model is unsupervised, there is no need for labels. It is trained with three inputs, where the middle slice is what the model learns to interpolate. At test time the model receives two inputs and produces a single output, where the output is compared to the ground truth. The production model takes in two inputs and outputs the interpolated slice. The output slices are combined together with the original slices to create the final upsampled scan.



**Figure 5**

Five interpolation methods that are enlarged and compared side by side for two example slices. Scans to the left are the slices used and the red box indicates the region used for comparison. Methods include A.) linear, B.) windowed sinc, C.) nearest neighbor D.) BSpline and E.) PixelMiner (ours) with the corresponding F.) ground truth slice on the bottom right.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryMaterials.pdf](#)