

# AUS Repository

## Machine Learning-based X-Ray Projection Interpolation for Improved 4D-CBCT Reconstruction

Item Type	Article;Peer-Reviewed;Published version;Other
Authors	Ramesh, Jayroop;Sankalpa, Donthi;Mitra, Rohan;Dhou, Salam
Citation	J. Ramesh, D. Sankalpa, R. Mitra and S. Dhou, "Machine Learning-based X-Ray Projection Interpolation for Improved 4D-CBCT Reconstruction," in IEEE Open Journal of Engineering in Medicine and Biology, doi: 10.1109/OJEMB.2024.3459622.
DOI	<a href="https://doi.org/10.1109/OJEMB.2024.3459622">10.1109/OJEMB.2024.3459622</a>
Publisher	IEEE
Download date	2026-05-11 03:14:31
Link to Item	<a href="https://hdl.handle.net/11073/25591">https://hdl.handle.net/11073/25591</a>

# Machine Learning-based X-Ray Projection Interpolation for Improved 4D-CBCT Reconstruction

Jayroop Ramesh\*<sup>✉</sup>, Donthi Sankalpa\*, Rohan Mitra\*<sup>✉</sup>, and Salam Dhou<sup>✉</sup>, (Member, IEEE)

**Abstract—Goal:** Respiration-correlated cone-beam computed tomography (4D-CBCT) is an X-ray-based imaging modality that uses reconstruction algorithms to produce time-varying volumetric images of moving anatomy over a cycle of respiratory motion. The quality of the produced images is affected by the number of CBCT projections available for reconstruction. Interpolation techniques have been used to generate intermediary projections to be used, along with the original projections, for reconstruction. Transfer learning is a powerful approach that harnesses the ability to reuse pre-trained models in solving new problems. **Methods:** Several state-of-the-art pre-trained deep learning models, used for video frame interpolation, are utilized in this work to generate intermediary projections. Moreover, a novel regression predictive modeling approach is also proposed to achieve the same objective. Digital phantom and clinical datasets are used to evaluate the performance of the models. **Results:** The results show that the Real-Time Intermediate Flow Estimation (RIFE) algorithm outperforms the others in terms of the Structural Similarity Index Method (SSIM):  $0.986 \pm 0.010$ , Peak Signal to Noise Ratio (PSNR):  $44.13 \pm 2.76$ , and Mean Square Error (MSE):  $18.86 \pm 206.90$  across all datasets. Moreover, the interpolated projections were used along with the original ones to reconstruct a 4D-CBCT image that was compared to that reconstructed from the original projections only. **Conclusions:** The reconstructed image using the proposed approach was found to minimize the streaking artifacts, thereby enhancing the image quality. This work demonstrates the advantage of using general-purpose transfer learning algorithms in 4D-CBCT image enhancement.

**Index Terms—**4D-CBCT reconstruction, Deep Learning, Transfer learning, Intermediate Projection Interpolation, Multi-output Regression

**Impact Statement-** This study extends state-of-the-art deep learning-based video frame interpolation algorithms to minimize streaking artifacts in respiration-correlated CBCT images. This study has implications for enhancing the 4D-CBCT images without the need to increase radiation dose nor the scanning duration in clinical settings.

This paragraph of the first footnote will contain the date on which you submitted your paper for review. It will also contain support information, including sponsor and financial support acknowledgment. (\*Jayroop Ramesh, Donthi Sankalpa and Rohan Mitra contributed equally to this work.) (Corresponding authors: Jayroop Ramesh; Salam Dhou)

Jayroop Ramesh is with the Department of Computer Science and Engineering, American University of Sharjah, Sharjah 26666 UAE (e-mail: b00057412@aus.edu).

Donthi Sankalpa is with the Department of Computer Science and Engineering, American University of Sharjah, Sharjah 26666 UAE (e-mail: dsankalpa@aus.edu).

Rohan Mitra is with the Department of Computer Science and Engineering, American University of Sharjah, Sharjah 26666 UAE (e-mail: b00085023@aus.edu).

Salam Dhou is with the Department of Computer Science and Engineering, American University of Sharjah, Sharjah 26666 UAE (e-mail: sdhou@aus.edu).

## I. INTRODUCTION

CONE-BEAM Computed Tomography (CBCT) is a medical imaging technique that uses a scanning system rotating around the patient to capture projections via a cone-shaped X-ray Beam. CBCT has become increasingly important for patient positioning and verification in image-guided radiation therapy (IGRT) due to lower radiation dose compared to a CT scan while providing a relatively high resolution [1].

As CT scans can be harmful to patients who need to get the scan done frequently, utilizing 4D-CBCT has become a necessary alternative. To reconstruct a respiration-correlated (4D)-CBCT image, the CBCT projections are sorted into several respiratory phase bins (typically 10) based on the respiratory signal extracted from each projection [2]. Then, a 3D image is reconstructed for each phase from the corresponding projections using a reconstruction algorithm. A consequence of this approach is that fewer projections will be available for each respiratory phase resulting in streaking artifacts in the reconstructed images. Increasing the scanning duration can increase the number of captured projections but at the expense of exposing patients to more radiation which does not provide a feasible solution.

Generating intermediary projections via image interpolation has been suggested in the literature to increase the number of projections available for reconstruction. Traditional image interpolation methods [3][4][5] have been proposed in this domain. Although advanced methods improved image quality, they still faced challenges such as blurring and low fidelity, especially with moving anatomical structures. These limitations motivate the use of automated approaches for efficient generation of intermediary projections.

With the rise of machine learning techniques driven by advancements in state-of-the-art computer vision algorithms, the medical imaging discipline is being augmented on many fronts [6]. Video interpolation is a commonly employed video processing technique to generate intermediary frames between consecutive images, and such techniques are being enhanced drastically by deep-learning paradigms as well. Dynamic video processing networks can generate high-quality interpolated images without additional training, even when applied to different scanning settings.

Transfer learning uses pre-trained deep learning models to overcome limitations of low-resource data [7]–[9]. In this work, we utilize pre-trained networks for video frame interpolation. The best-performing models across multiple natural

image/video datasets considered in this work are RIFE [10], DAIN[11], Softsplatting[12], SepConv[13], Super SloMo[14], BMBC[15] and CDFI[16]. These algorithms are applied in this work to the task of augmenting CBCT lung datasets for enhanced 4D-CBCT image quality in both synthetic phantom and clinical CBCT scans acquired for lung cancer patients. Moreover, a multi-output regression-based machine learning model is proposed for intermediary image interpolation by utilizing convolutions feature maps. We believe that this approach is highly applicable as similar approaches have been proposed in the literature and have shown promise in terms of enhancing image resolution [17], similar interpolation tasks [18], and improving 3D image generation [19]. Thus, the primary contributions of this work are:

- Leveraging the principle of transfer learning by using pre-trained deep learning video frame interpolation algorithms to perform intermediary CBCT projection interpolation.
- Developing a novel multi-output regression-based machine learning model utilizing convolutions feature maps for intermediary CBCT projection interpolation.
- Reconstruction of a 4D-CBCT image using the interpolated projections and the original ones and evaluating the reconstruction quality.

## II. MATERIALS AND METHODS

Our objective is to generate an intermediary (or interchangeably, interpolated) projection  $\hat{I}_t$ , given two input projections  $I_0$  and  $I_1$  where  $t \in (0, 1)$ . In the context of machine learning, the intermediary projection generation task is performed via prediction using trained machine learning models. The predicted projection  $\hat{I}_t$  is then evaluated with respect to its ground truth  $\hat{I}_t^{GT}$  using various loss functions.

### A. Datasets

In this work, phantom and clinical patient datasets are utilized. The characteristics of each of the datasets are described in Table I.

TABLE I. Phantom and Patient Datasets Specifications

Dataset	Imaging System	No. of Projs	Proj Size
Phantom 1	N/A	360	512 × 512
Phantom 2	N/A	720	512 × 512
Patient 1	Elekta XVI 3.5	701	512 × 512
Patient 2	Varian OBI	2396	1024 × 768
Patient 3	Varian OBI	1220 out of 2436	1024 × 768
4D-Lung (80 patients)	ADAC	180 per patient	512 × 512
LIDC (1018 patients)	QIICR	180 per patient	512 × 512
SPARE (5 patients)	Elekta	180 per patient	512 × 512

The two phantom datasets, *Phantom 1* and *Phantom 2*, were generated using the eXtended CArdio-Torso (XCAT) phantom [20] [21]. This phantom body depicts an average human male in shape, proportion, and composition. The datasets simulate the rotation of the X-ray source and the breathing motion of the patient. The motion of the heart has been stopped manually using the software [5] [22] [23]. Additionally, the three patient datasets, *Patient 1*, *Patient 2*, and *Patient 3*, consist of anonymous CBCT scans of patients shared by the Virginia Commonwealth University, Richmond, Virginia, USA,

under a Data Transfer and Use Agreement, and were retrospectively analyzed [2][24]. Moreover, the research protocol qualified for exempt approval from the Institutional Review Board (IRB) of the American University of Sharjah, United Arab Emirates (Approved IRB Protocol: 18-425). For more details on the device settings and data sources, please refer to the Supplementary material. Moreover, the generalizability of the proposed methods under different conditions was assessed using additional three publicly available databases involving multiple patient datasets: 4D-Lung [25], the Lung Image Database Consortium (LIDC) [26], and Sparse-view Reconstruction (SPARE) Elekta [27].

### B. ML Methods

1) *Regression-Based Algorithms*: The interpolation problem studied in this work can be posed as a regression problem where the original images are considered independent variables, while the predicted output image is considered a dependent variable. However, two main challenges had to be addressed in the context of our work, namely, dimensionality of data and multi-output regression (see Supplementary Materials).

Two different pre-processing techniques were applied to reduce the dimensionality of the datasets: i) *resizing and concatenating images*, as well as ii) deriving feature maps through convolutions (see Supplementary Materials).

Several regression algorithms were examined in this work, namely linear regression variants such as the Linear, Lasso and Ridge regression, K-Nearest Neighbors Regressor, Decision Tree Regressor, and the Support Vector Machine (SVM) Regressor with linear, Radial Basis Function (RBF), and polynomial kernels. Algorithms without multi-output support were used with a Multi-Output Regressor wrapper class that fits one model for each output.

2) *DL-based Algorithms*: In this section, the DL-based algorithms used for image interpolation are introduced. These algorithms are selected based on their differing approaches on handling motion in naturalistic videos, and their training process to ensure diversity in model representation. Details about the different loss functions used, handling motion, warping process and training for each algorithm can be found in the Supplementary Materials.

Fig. 1 depicts a typical composition of a deep learning algorithm used for video frame interpolation. Every model defined in this subsection need not have every component. However, they certainly make use of similar properties to produce the final synthesized image. In presenting these DL-based algorithms, we focus on the components and highlight the mechanisms that are relevant to our task. Each deep learning algorithm is adopted as a pre-trained model with frozen weights as initialized, and used for inference only. All parameters used are set according to the original source paper unless mentioned otherwise explicitly.

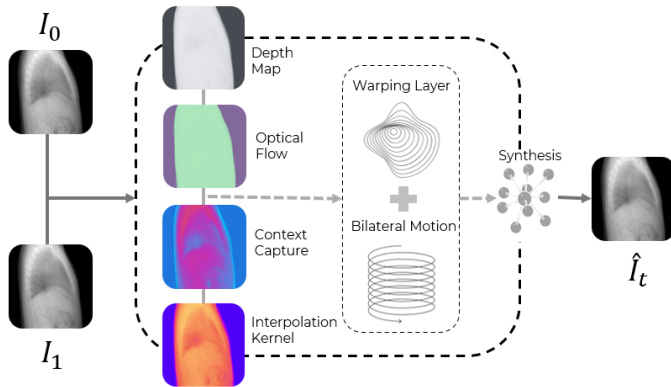


Fig. 1. Video intermediate frame interpolation pipeline using deep learning.

The algorithms considered an expanded (see Supplementary Materials) in this work are:

- 1) Real-time Intermediate Flow Estimation algorithm (RIFE) [10].
- 2) Spatially-Adaptive Separable Convolution (SepConv) [28].
- 3) Depth-Aware Video Frame Interpolation (DAIN) [11].
- 4) Bilateral Motion Estimation with Bilateral Cost Volume (BMBC) [15].
- 5) Super SloMo [14].
- 6) Softmax Splatting [12]
- 7) Residue Refinement Video Frame Interpolation Network (RRIN) [29].
- 8) Compression-Driven Network Design for Frame Interpolation (CDFI) [16].

### C. Assessment Criteria

1) *Projection Evaluation*: Different evaluation metrics are considered to assess the quality of the resulting interpolated images and quantify information loss due to noise, namely Structural Similarity Index Measure (SSIM) [30], [31], Peak Signal to Noise Ratio (PSNR) [30], [32], and the Mean Squared Error (MSE) [30], [33].

2) *4D-CBCT Reconstruction Evaluation*: To assess the reconstruction quality, two image features were derived from a 4D-CBCT reconstructed image, namely organ contours and noise level in constant-intensity areas (see Supplemental Material). These metrics have been chosen based on the aim of the experiments; reduce the streaking artifacts while preserving the sharpness of the anatomy in the reconstructed images [34].

## III. RESULTS

In this section, the results of the projection interpolation methods as well as the image reconstruction experiments are presented.

### A. Experimental Setting

The evaluation metrics presented in this section were computed by applying the algorithms such that every frame served as part of both the input and output data, except for the first and last frames ( $I_0$  and  $I_{N-1}$ , respectively, where  $N$  is

the size of the dataset). Thus, each subsequent interpolation process ingests two consecutive projections, with overlap for non-edge projections ( $I_0$  and  $I_{N-1}$ ). The resulting interpolated projections were then compared with the corresponding ground truth frames.

### B. Evaluation of Machine Learning Regression-based Approach

The regression-based approach was trained using a 5-fold cross-validation technique with randomized contiguous triplet frames. Multi-output regressors perform well for small images but are resource-intensive for larger ones. As reported in Table II, linear regression models work well across all datasets, since small-angle rotations can be approximated as linear translation. Moreover, training time and memory requirements are minimal, enabling deployment on resource-constrained devices. Table II shows that resizing and convolutional methods yield comparable results overall.

Linear regression and Ridge regression achieved the best performance in the *Phantom 1* and *Phantom 2* datasets, respectively, while KNN regression showed better performance in the patient datasets. Overall, although the worst performance was observed in the *Phantom 1* and *Patient 2* datasets, the achieved SSIM was consistently above 99%, indicating high image similarity. Resizing and Ridge regression consistently yield high PSNR, effective even on larger images, indicating scalability. The KNN Regressor works by averaging the  $K$ -closest values in the dataset, which works well because the pixel values in the intermediary frame to be predicted are likely confined to a local region as its neighboring pixels across pair-wise images. On the other hand SVM with a polynomial kernel achieved similar performance, but with a discernible computational overhead.

We explored the generalizability of linear regression on three public, multi-patient databases as described in Section II. A. Table I. The linear regression model was trained on the first 80% of patients and tested on the remaining 20% of patients, with results shown in Table III.

Table III shows that linear regression performs better with more data, achieving higher PSNRs than using a single patient's scans. This is likely because training on multiple patients allows the model to learn a broader range of features across all 360° angles allowing it to better learn the structure of a CT scan. Linear regression achieves a high PSNR of 95dB on the LIDC dataset with an SSIM of 99.9%, showing the importance of more higher quality and diverse data in improving machine learning models for medical image generation. Real-life datasets expose the model to diverse anatomical variations, and imaging conditions improving generalization and performance.

### C. Evaluation of Deep learning-based Approach

The deep learning models considered in this work were implemented with the default parameters specified in the original works, and their respective pre-trained weights were utilized during inference. The criteria for choosing the "best" model is if it outperforms the others in at least two out of

TABLE II. Summary of results from the classical machine learning (regression) models per dataset.

Dataset	Best Regressor	Hyper-parameters	Best preprocess- ing method	PSNR	SSIM	Time (s)
Phantom 1	Linear	-	Resizing	39.91	0.99	0.69
Phantom 2	SVR & Ridge	SVR: Kernel= Poly, Deg=7; Ridge: $\alpha = 0.1$	SVR: Resizing; Ridge: Conv	SVR: 35.51; Ridge: 34.57	SVR: - ; Ridge: 0.99	SVR: 964.14; Ridge: 0.04
Patient 1	Ridge	$\alpha = 0.01$	Resizing	40.43	0.99	1.30
Patient 2	Ridge & KNN	Ridge: $\alpha = 0.1$ ; KNN: $K = 3$	Ridge: Resizing; KNN: Conv	Ridge: 35.79; KNN: 34.41	Ridge: 0.99; KNN: 0.99	Ridge: 2.56; KNN: 0.57
Patient 3	KNN	$K = 5$	Conv	43.07	0.99	0.0019

TABLE III. Quantitative comparisons of linear regression on additional datasets

Dataset	SSIM	PSNR	MSE
4D Lung	$0.93 \pm 0.032$	$85.58 \pm 2.52$	$0.00021 \pm 0.00016$
LIDC	$0.999 \pm 0.00078$	$95.42 \pm 1.89$	$0.000012 \pm 0.000027$
SPARE Elekta	$0.67 \pm 0.088$	$71.38 \pm 2.12$	$0.0054 \pm 0.0030$

three evaluation metrics. Table IV presents a summary of the averaged metrics.

Overall, the best-performing model across all datasets and metrics consistently was RIFE, with SepConv-11 doing slightly better on *Phantom 2*. The marginal improvement in spatial resolution in some cases for RIFE can be due to its progressive refinement strategy exploiting the mechanisms IFNet block of its architecture, allowing for the retention of a robust global representation.

Figure 2 shows the sample original projections, the interpolated ones using the best performing model, and the corresponding difference images for *Phantom 2* and *Patient 1* datasets. The specific frames used for comparisons are no. 170 and no. 1809 from each dataset respectively, and were selected to show maximum anatomical details. The first frame and second frames obtained SSIM, PSNR and MSE scores of 0.99, 48.73, and 86.95, and 0.97, 42.14, and 3.96 when compared to the original projections respectively. From the difference images, it can be observed that the thickness and spread of some regions do vary slightly. However, to a greater extent, the motion artifacts have been accounted for and there is an overall structural fidelity in the interpolation quality.

TABLE IV. Quantitative comparisons of the deep learning algorithms across all datasets.

Dataset	Best Model	SSIM	PSNR	MSE
Phantom 1	RIFE	$0.99 \pm 0.01$	$43.50 \pm 3.06$	$10.49 \pm 99.08$
Phantom 2	SepConv-11	$0.99 \pm 0.03$	$48.84 \pm 4.51$	$45.74 \pm 509.21$
Patient 1	RIFE	$0.99 \pm 0.01$	$43.31 \pm 3.59$	$5.14 \pm 10.84$
Patient 2	RIFE	$0.97 \pm 0.01$	$41.53 \pm 2.35$	$5.23 \pm 2.75$
Patient 3	RIFE	$0.98 \pm 0.02$	$44.37 \pm 4.42$	$31.33 \pm 529.23$

#### D. Evaluation of 4D-CBCT Reconstruction

In this section, the 4D-CBCT reconstruction quality after considering the interpolated frames is evaluated. In this experiment, the *Phantom 1* dataset was considered. A total of 10 respiratory cycles were generated with each cycle comprising of 36 projections. The projections were binned to ten phase bins, each having 36 projections. Then a 4D-CBCT image was reconstructed with each phase was reconstructed from a phase bin.

Figure 3 represents the general structure of the experiments where projections belonging to the same phase are being pointed to by blue arrows. Applying interpolation to  $n$  projections belonging to the same phase would follow the calculation  $n + (n - 1)i$ , where  $i$  represents the order of the projections generated in between. For instance, if  $n = 36$  and  $i = 3$ , meaning 3 projections will be interpolated between each consecutive original projections, we get 105 interpolated projections and so on.

Figure 4 depicts the coronal views through the reconstructed anatomy at the end of the exhale phase after using the Feldkamp-Davis-Kress algorithm [35] with a *hann* filter of value 0.4 to cut-off some non-intrinsic high-frequency noise experienced during ramp filtering. The images correspond to reconstruction using (a) all 360 images, (b) only 36 original projections in the same respiratory phase, and (c) original 36 plus 35 interpolated projections between them for 71 total. The best-performing interpolation model RIFE was used to generate the interpolated projections. Table V reports their quantitative metric scores as well.

TABLE V. Quantitative metrics reporting performances after increasing levels of interpolation against the baseline original reconstruction with 360 images.

Projections	SSIM	PSNR	MSE	MAE
36 Original	0.471	20.11	0.0097	0.064
36 Original + 35 Interpolated	0.550	23.731	0.0042	0.042
36 Original + 105 Interpolated	0.672	26.430	0.0023	0.029
36 Original + 245 Interpolated	0.794	27.621	0.0017	0.022

Considering the original projections only, the reconstructed images in Figure 4 (a) naturally appear to be much clearer and have no streaking artifacts compared to (b). However, there is the matter of blurring in (a) induced by motion artifacts which arise when projections originate from multiple different respiratory phases. As such, the reconstructed images in (b) tend to have sharper edges, but this is overshadowed by levels of noise and streaking artifacts present, which is a consequence of an insufficient number of projections used in the reconstruction. As expected, the incorporation of synthetic interpolated projections in (c) decreases the severity of streaking artifacts while reducing blurriness.

Figure 5(a) compares the edge profiles to demonstrate the effect of using the interpolated projections along with the original ones on the quantitative spatial resolution of the region capturing the diaphragmatic movement. As shown in Figure 5(b), edges are sharper when more projections belonging to one phase are used for reconstruction, with improved pixel intensity in terms of magnitude and stability.

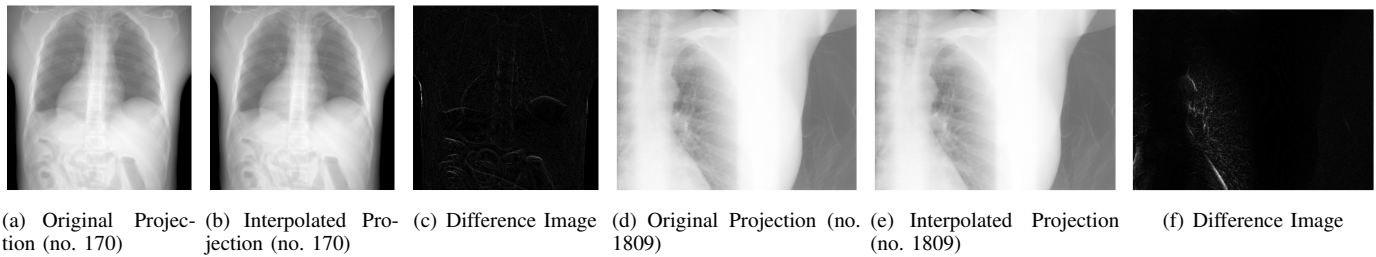


Fig. 2. Visual evaluation of the generated projections: (a) original sample projection from *Phantom 2* dataset, (b) corresponding interpolated projection and (c) their difference image, (d) original Sample projection from *Patient 1* dataset, (e) corresponding interpolated projection and (e) their difference image.

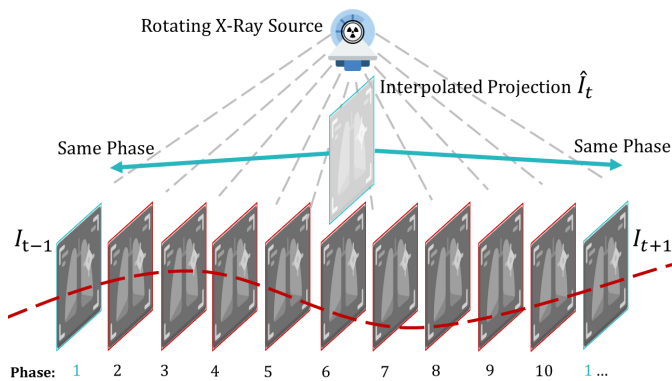


Fig. 3. Illustration of the 4D-CBCT interpolation process where dashed line illustrates respiratory motion.

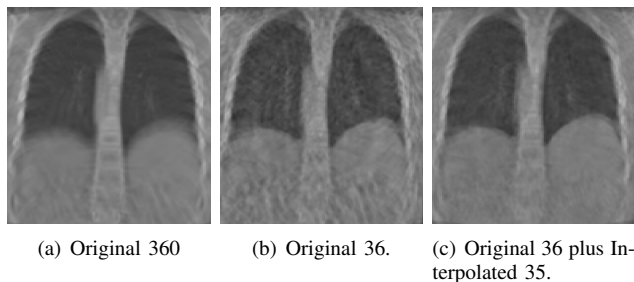


Fig. 4. Reconstructed 4D-CBCT phantom images using a combination of original and interpolated frames.

Table VI shows the corresponding noise levels, calculated as the standard deviations of the pixel intensities in the two ROIs of a reconstructed image selected from the background. The results are in concordance with the findings that using more projections in reconstruction yields lower noise levels and fewer streaks and lower noise. This is evident in the case of using 36 original plus 245 interpolated synthetic projections for reconstruction.

TABLE VI. Noise levels in two regions of interest (the standard deviation of the intensity in a ROI)

	360 Original	36 Original	36 Original + 35 Interpolated	36 Original + 105 Interpolated	36 Original + 245 Interpolated
ROI 1	0.0088	0.0474	0.0168	0.0100	0.0044
ROI 2	0.00071	0.0395	0.0268	0.0131	0.0072

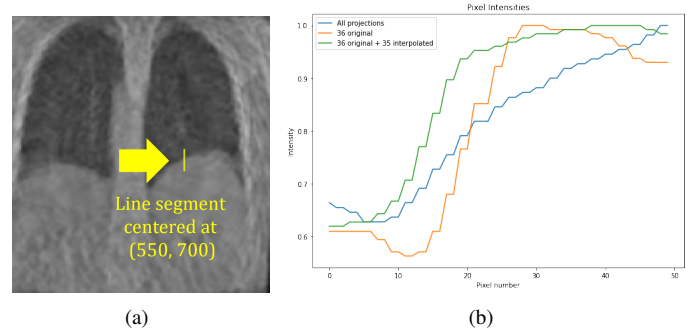


Fig. 5. Spatial resolution quality assessment of reconstructed images from coronal views. (a) Edge profile in the reconstructed phantom along the diaphragm. (b) Normalized pixel intensities of the edge profile in reconstructed images.

#### IV. DISCUSSION

The rationale for promoting 4D-CBCT is due to its inherent capability to faithfully represent the patient anatomy holistically for assessment by radiologists [36]. The experiments demonstrate the feasibility of pre-trained video frame interpolation algorithms in improving the quality of 4D-CBCT images.

Resizing and convolutions are equally effective as preprocessing methods, enabling linear and KNN regression models to perform well across datasets without the computational demands of methods like SVM. KNN's similar performance to Ridge regression likely stems from its output being an average of the  $K$ -nearest samples, akin to a linear combination—a concept central to linear regression. Additionally, KNN tends to select neighboring images, effectively averaging the  $K/2$  images before and after the current one.

Deep learning models used in this work achieved high image quality metrics  $SSIM > 90$ ,  $PSNR > 30$ , and  $MSE < 10$  without requiring re-training from scratch as with the baseline regression models. It was noted that RIFE has a significantly better image interpolation performance when compared to similar algorithms. Original and projections interpolated with RIFE were used to reconstruct a 4D-CBCT image to investigate practical utility. This yields the findings that using synthetic projections in reconstruction reduces both artifacts and motion-induced blurring.

Additionally, Table VII shows a similar experimental setup as described in section III, where LR and RIFE are applied

TABLE VII. Quantitative comparisons of RIFE and linear regression (LR) patient wise on SPARE Elekta

Patient No.	Model	SSIM	PSNR	MSE
1	<b>RIFE</b>	0.95 ± 0.029	38.44 ± 7.06	0.49 ± 1.11
1	<b>LR</b>	0.79 ± 0.057	73.61 ± 2.92	0.0035 ± 0.0022
2	<b>RIFE</b>	0.93 ± 0.024	33.75 ± 6.70	0.38 ± 1.38
2	<b>LR</b>	0.60 ± 0.10	65.49 ± 5.76	0.033 ± 0.026
3	<b>RIFE</b>	0.95 ± 0.72	36.28 ± 7.52	0.55 ± 0.72
3	<b>LR</b>	0.79 ± 0.061	70.35 ± 1.97	0.0066 ± 0.0030
4	<b>RIFE</b>	0.94 ± 0.033	31.98 ± 6.31	0.96 ± 0.98
4	<b>LR</b>	0.45 ± 0.085	62.56 ± 2.78	0.042 ± 0.019
5	<b>RIFE</b>	0.94 ± 0.044	34.80 ± 8.13	0.86 ± 1.19
5	<b>LR</b>	0.72 ± 0.093	71.73 ± 5.27	0.0068 ± 0.0043

independently for each patient selected from the SPARE Elekta dataset. LR is trained on the first 80% of projections of a single patient, and tested on the remaining 20% of projections, while RIFE is only used for inference on the same last 20% of projections. We observe that both RIFE and LR show perform well emphasizing the validity of the experimental setup and the generalizability of the proposed approaches to datasets collected under different conditions.

Our work complements latest developments in the field using transfer learning [37]–[39] and exhibits the merits of augmenting sparse projections with general-purpose frame interpolation techniques that can contribute to 4D-CBCT reconstruction and overcome the need for fine-tuning (see Supplemental Material).

## V. CONCLUSION

In this work, a method leveraging general-purpose frame-interpolation algorithms is proposed to improve the quality of respiration-correlated CBCT reconstruction by reducing the undersampling-induced streaking artifacts. Various image interpolation algorithms were assessed using both digital phantom and clinical datasets. Among the algorithms considered, the RIFE algorithm demonstrated superior results in terms of SSIM ( $0.986 \pm 0.010$ ), PSNR ( $44.13 \pm 2.76$ ), and MSE ( $18.86 \pm 206.90$ ) across all datasets. Moreover, it was found that using the interpolated projections along with the original ones in reconstructing 4D-CBCT images decreases the severity of the streaking artifacts while preserving sharp edges in the image. The proposed approach allows for flexible adaptation into potentially resource-constrained, or low-data regimes.

## ACKNOWLEDGEMENT

The work in this paper was supported, in part, by the American University of Sharjah through the research grants (Grants #: EFRG18-BBR-CEN-04/EN0246 and FRG22-C-E28/EN4012) and the Open Access Program (Award #: OAPCEN-1410-E00325). This paper represents the opinions of the authors and does not mean to represent the position or opinions of the American University of Sharjah.

## CONFLICTS OF INTEREST

The authors have no conflicts of interest to declare.

## SUPPLEMENTARY MATERIALS

Supplementary Materials include descriptions of loss functions used, additional theoretical and implementation details for the DL-based algorithms, as well as further discussion contextualizing our results with related work.

## AUTHORS CONTRIBUTIONS

Conceptualization: SD; Methodology: JR, DS, RM, SD; Validation: JR, DS, RM; Formal Analysis: JR, DS, RM, SD; Investigation: JR, DS, RM; Resources: SD; Writing – Original Draft: JR, DS, RM; Writing – Review & Editing: JR, RM, SD; Visualization: JR, DS, RM; Supervision: SD.

## REFERENCES

- [1] D. C. Hatcher, "Operational Principles for Cone-Beam Computed Tomography," *The Journal of the American Dental Association*, vol. 141, 3S–6S, Oct. 1, 2010, ISSN: 0002-8177. DOI: 10.14219/jada.archive.2010.0359. PMID: 20884933. [Online]. Available: [https://jada.ada.org/article/S0002-8177\(14\)63738-7/fulltext](https://jada.ada.org/article/S0002-8177(14)63738-7/fulltext) (visited on 11/29/2023).
- [2] S. Dhou, Y. Motai, and G. D. Hugo, "Local intensity feature tracking and motion modeling for respiratory signal extraction in cone beam CT projections," *IEEE transactions on bio-medical engineering*, vol. 60, no. 2, Feb. 2013, ISSN: 1558-2531. DOI: 10.1109/TBME.2012.2226883. PMID: 23193225.
- [3] T. Lehmann, C. Gonner, and K. Spitzer, "Survey: Interpolation methods in medical image processing," *IEEE Transactions on Medical Imaging*, vol. 18, no. 11, pp. 1049–1075, Nov. 1999, ISSN: 1558-254X. DOI: 10.1109/42.816070. [Online]. Available: <https://ieeexplore.ieee.org/document/816070> (visited on 11/29/2023).
- [4] M. Bertram, J. Wiegert, D. Schafer, T. Aach, and G. Rose, "Directional View Interpolation for Compensation of Sparse Angular Sampling in Cone-Beam CT," *IEEE Transactions on Medical Imaging*, vol. 28, no. 7, pp. 1011–1022, Jul. 2009, ISSN: 1558-254X. DOI: 10.1109/TMI.2008.2011550. [Online]. Available: <https://ieeexplore.ieee.org/document/4738374> (visited on 11/29/2023).
- [5] S. Dhou, G. D. Hugo, and A. Docef, "Motion-based projection generation for 4D-CT reconstruction," in *2014 IEEE International Conference on Image Processing (ICIP)*, Oct. 2014, pp. 1698–1702. DOI: 10.1109/ICIP.2014.7025340.
- [6] M. Zhang, S. Gu, and Y. Shi, "The use of deep learning methods in low-dose computed tomography image reconstruction: A systematic review," *Complex & Intelligent Systems*, vol. 8, no. 6, pp. 5545–5561, Dec. 2022, ISSN: 2198-6053. DOI: 10.1007/s40747-022-00724-7.
- [7] M. Raghu, C. Zhang, J. Kleinberg, and S. Bengio, "Transfusion: Understanding Transfer Learning for Medical Imaging," in *Advances in Neural Information Processing Systems*, vol. 32, Curran Associates, Inc., 2019. [Online]. Available: [https://proceedings.neurips.cc/paper\\_files/paper/2019/hash/eb1e78328c46506b46a4ac4a1e378b91-Abstract.html](https://proceedings.neurips.cc/paper_files/paper/2019/hash/eb1e78328c46506b46a4ac4a1e378b91-Abstract.html) (visited on 11/29/2023).
- [8] C. Matsoukas, J. F. Haslum, M. Sorkhei, M. Soderberg, and K. Smith, "What Makes Transfer Learning Work for Medical Images: Feature Reuse & Other Factors," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, LA, USA: IEEE, Jun. 2022, pp. 9215–9224, ISBN: 978-1-66546-946-3. DOI: 10.1109/CVPR52688.2022.00901. [Online]. Available: <https://ieeexplore.ieee.org/document/9878482/> (visited on 11/29/2023).
- [9] J. Lu, V. Behbood, P. Hao, H. Zuo, S. Xue, and G. Zhang, "Transfer learning using computational intelligence: A survey," *Knowledge-Based Systems*, 25th Anniversary of Knowledge-Based Systems, vol. 80, pp. 14–23, May 2015, ISSN: 0950-7051. DOI: 10.1016/j.knsys.2015.01.010.
- [10] Z. Huang, T. Zhang, W. Heng, B. Shi, and S. Zhou, *Real-Time Intermediate Flow Estimation for Video Frame Interpolation*, Jul. 2022. DOI: 10.48550/arXiv.2011.06294. arXiv: 2011.06294. [Online]. Available: <http://arxiv.org/abs/2011.06294> (visited on 03/20/2023), preprint.
- [11] W. Bao, W.-S. Lai, C. Ma, X. Zhang, Z. Gao, and M.-H. Yang, *Depth-Aware Video Frame Interpolation*, Apr. 2019. DOI: 10.48550/arXiv.1904.00830. arXiv: 1904.00830. [Online]. Available: <http://arxiv.org/abs/1904.00830> (visited on 03/20/2023), preprint.

## REFERENCES

- [12] S. Niklaus and F. Liu, *Softmax Splatting for Video Frame Interpolation*, Mar. 2020. DOI: 10.48550/arXiv.2003.05534. arXiv: 2003.05534. [Online]. Available: <http://arxiv.org/abs/2003.05534> (visited on 03/20/2023), preprint.
- [13] S. Niklaus, L. Mai, and F. Liu, *Video Frame Interpolation via Adaptive Separable Convolution*, Aug. 2017. DOI: 10.48550/arXiv.1708.01692. arXiv: 1708.01692. [Online]. Available: <http://arxiv.org/abs/1708.01692> (visited on 03/20/2023), preprint.
- [14] H. Jiang, D. Sun, V. Jampani, M.-H. Yang, E. Learned-Miller, and J. Kautz, *Super SloMo: High Quality Estimation of Multiple Intermediate Frames for Video Interpolation*, Jul. 2018. DOI: 10.48550/arXiv.1712.00080. arXiv: 1712.00080. [Online]. Available: <http://arxiv.org/abs/1712.00080> (visited on 03/20/2023), preprint.
- [15] J. Park, K. Ko, C. Lee, and C.-S. Kim, "BMBC: Bilateral Motion Estimation with Bilateral Cost Volume for Video Interpolation," *Computer Vision – ECCV 2020*, Lecture Notes in Computer Science, vol. 12359, A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, Eds., pp. 109–125, 2020. DOI: 10.1007/978-3-030-58568-6\_7.
- [16] T. Ding, L. Liang, Z. Zhu, and I. Zharkov, *CDFI: Compression-Driven Network Design for Frame Interpolation*, Mar. 2021. DOI: 10.48550/arXiv.2103.10559. arXiv: 2103.10559. [Online]. Available: <http://arxiv.org/abs/2103.10559> (visited on 03/20/2023), preprint.
- [17] X. Ying, H. Guo, K. Ma, J. Wu, Z. Weng, and Y. Zheng, "X2CT-GAN: Reconstructing CT from Biplanar X-Rays with Generative Adversarial Networks." arXiv: 1905.06902 [cs, eess]. (May 16, 2019), [Online]. Available: <http://arxiv.org/abs/1905.06902> (visited on 06/25/2024), preprint.
- [18] M. AlShurbaji and S. Dhou, "3D Image Generation from X-Ray Projections Using Generative Adversarial Networks," presented at the 2023 IEEE 23rd International Conference on Bioinformatics and Bioengineering (BIBE), IEEE Computer Society, Dec. 1, 2023, pp. 43–49, ISBN: 9798350393118. DOI: 10.1109/BIBE60311.2023.00015. [Online]. Available: <https://www.computer.org/csdl/proceedings-article/bibe/2023/931100a043/1UGTc0Oh9UQ> (visited on 06/24/2024).
- [19] "CycN-Net: A Convolutional Neural Network Specialized for 4D CBCT Images Refinement — IEEE Journals & Magazine — IEEE Xplore." (), [Online]. Available: <https://ieeexplore.ieee.org/document/9435315> (visited on 06/24/2024).
- [20] W. P. Segars, G. Sturgeon, S. Mendonca, J. Grimes, and B. M. Tsui, "4d xcat phantom for multimodality imaging research," *Medical physics*, vol. 37, no. 9, pp. 4902–4915, 2010.
- [21] M. E. Myronakis, W. Cai, S. Dhou, et al., "A graphical user interface for xcat phantom configuration, generation and processing," *Biomedical Physics & Engineering Express*, vol. 3, no. 1, p. 017003, 2017.
- [22] S. Dweek, S. Dhou, and T. Shanableh, "In-between projection interpolation in cone-beam CT imaging using convolutional neural networks," in *Medical Imaging 2022: Physics of Medical Imaging*, W. Zhao and L. Yu, Eds., San Diego, United States: SPIE, Apr. 2022, p. 164, ISBN: 978-1-5106-4937-8 978-1-5106-4938-5. DOI: 10.1117/12.2611474.
- [23] S. Sabah and S. Dhou, "Image-based extraction of breathing signal from cone-beam CT projections," in *Medical Imaging 2020: Image-Guided Procedures, Robotic Interventions, and Modeling*, B. Fei and C. A. Linte, Eds., International Society for Optics and Photonics, vol. 11315, SPIE, 2020, 113152S. DOI: 10.1117/12.2550462. [Online]. Available: <https://doi.org/10.1117/12.2550462>.
- [24] S. Dhou, A. Docef, and G. Hugo, "Image-based respiratory signal extraction using dimensionality reduction for phase sorting in cone-beam ct projections," in *Proceedings of the 2017 International Conference on Computational Biology and Bioinformatics*, ser. ICCBB '17, Newark, NJ, USA: Association for Computing Machinery, 2017, pp. 79–84, ISBN: 9781450353229. DOI: 10.1145/3155077.3155093. [Online]. Available: <https://doi.org/10.1145/3155077.3155093>.
- [25] G. D. Hugo, E. Weiss, W. C. Sleeman, et al., *Data from 4d lung imaging of nscl patients*, 2016.
- [26] O. Ozdemir, R. L. Russell, and A. A. Berlin, "A 3D Probabilistic Deep Learning System for Detection and Diagnosis of Lung Cancer Using Low-Dose CT Scans." arXiv: 1902.03233 [cs]. (Jan. 20, 2020), [Online]. Available: <http://arxiv.org/abs/1902.03233> (visited on 06/25/2024), preprint.
- [27] C.-C. Shieh, Y. Gonzalez, B. Li, et al., "Spare: Sparse-view reconstruction challenge for 4d cone-beam ct from a 1-min scan," *Medical physics*, vol. 46, no. 9, pp. 3799–3811, 2019.
- [28] S. Niklaus, L. Mai, and F. Liu, *Video Frame Interpolation via Adaptive Convolution*, Mar. 2017. DOI: 10.48550/arXiv.1703.07514. arXiv: 1703.07514. [Online]. Available: <http://arxiv.org/abs/1703.07514> (visited on 03/20/2023), preprint.
- [29] H. Li, Y. Yuan, and Q. Wang, "Video Frame Interpolation Via Residue Refinement," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2020, pp. 2613–2617. DOI: 10.1109/ICASSP40776.2020.9053987.
- [30] U. Sara, M. Akter, and M. S. Uddin, "Image Quality Assessment through FSIM, SSIM, MSE and PSNR—A Comparative Study," *Journal of Computer and Communications*, vol. 07, no. 03, pp. 8–18, 2019, ISSN: 2327-5219, 2327-5227. DOI: 10.4236/jcc.2019.73002.
- [31] Z. Zhang, M. Huang, Z. Jiang, et al., "Patient-specific deep learning model to enhance 4D-CBCT image for radiomics analysis," *Physics in Medicine and Biology*, vol. 67, no. 8, Apr. 1, 2022, ISSN: 1361-6560. DOI: 10.1088/1361-6560/ac5f6e. PMID: 35313293.
- [32] T. Reynolds, P. Lim, P. J. Keall, and R. O'Brien, "Minimizing 4DCBCT imaging dose and scan time with Respiratory Motion Guided 4DCBCT: A pre-clinical investigation," *Biomedical Physics & Engineering Express*, vol. 7, no. 2, Jan. 28, 2021, ISSN: 2057-1976. DOI: 10.1088/2057-1976/abdc82. PMID: 33455950.
- [33] R. T. O'Brien, O. Dillon, B. Lau, et al., "The first-in-human implementation of adaptive 4D cone beam CT for lung cancer radiotherapy: 4DCBCT in less time with less dose," *Radiotherapy and Oncology*, vol. 161, pp. 29–34, Aug. 1, 2021, ISSN: 0167-8140. DOI: 10.1016/j.radonc.2021.05.021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0167814021062630> (visited on 06/24/2024).
- [34] V. Mudeng, M. Kim, and S.-w. Choe, "Prospects of Structural Similarity Index for Medical Image Analysis," *Applied Sciences*, vol. 12, no. 8, p. 3754, 8 Jan. 2022, ISSN: 2076-3417. DOI: 10.3390/app12083754. [Online]. Available: <https://www.mdpi.com/2076-3417/12/8/3754> (visited on 06/25/2024).
- [35] L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical cone-beam algorithm," *Journal of the Optical Society of America A*, vol. 1, no. 6, p. 612, Jun. 1, 1984, ISSN: 1084-7529, 1520-8532. DOI: 10.1364/JOSAA.1.000612. [Online]. Available: <https://opg.optica.org/abstract.cfm?URI=josaa-1-6-612> (visited on 11/29/2023).
- [36] W. Cai, M. H. Hurwitz, C. L. Williams, et al., "3D delivered dose assessment using a 4DCT-based motion model," *Medical Physics*, vol. 42, pp. 2897–2907, 6Part1 2015, ISSN: 2473-4209. DOI: 10.1118/1.4921041.
- [37] Y. Mao, Z. Deng, H. Yao, T. Ye, K. Kawaguchi, and J. Zou, *Last-Layer Fairness Fine-tuning is Simple and Effective for Neural Networks*, Apr. 2023. DOI: 10.48550/arXiv.2304.03935. arXiv: 2304.03935 [cs]. [Online]. Available: <http://arxiv.org/abs/2304.03935> (visited on 05/09/2023), preprint.
- [38] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, *How transferable are features in deep neural networks?* Nov. 2014. DOI: 10.48550/arXiv.1411.1792. arXiv: 1411.1792 [cs]. [Online]. Available: <http://arxiv.org/abs/1411.1792> (visited on 05/09/2023), preprint.
- [39] G. Dong, C. Zhang, L. Deng, et al., "A deep unsupervised learning framework for the 4d cbct artifact correction," *Physics in Medicine & Biology*, vol. 67, no. 5, p. 055012, 2022.