ACTA IMEKO
ISSN: 2221-870X
September 2026, Volume 15, Number 3, 1 – 8

Less is more: Few-shot learning in 3D super resolution using synthetic CT images

Lukas Nepelius[1], Patrick Weinberger[1,2], Caroline Hastra[1], Miroslav Yosifov[1], Jonathan Glinz[1], Sascha Senck[1]

[1] *University of Applied Sciences Upper Austria, School of Engineering, Wels, Austria*
[2] *Johannes Kepler University, Linz, Austria*

ABSTRACT
In this work, we study data requirements and scaling behavior of super resolution (SR) machine learning networks on computed tomography (CT) scans of aluminum foam samples. The pipeline consists of pre-training a 3D U-Net model on purely synthetic data and fine-tuning this model on increasing quantities of real data. Synthetic data is produced by a shallow convolutional neural network (CNN), which learns real CT degradations. This degraded data is later used for training a 3D U-Net SR network in order to reduce manual alignment labor for training such models. Our results suggest that only a small fraction of the dataset is required for fine-tuning synthetic models and diminishing returns are observed after a few pairs. Furthermore, simply increasing the amount of real fine-tuning data does not consistently improve performance across the different data configurations. These findings indicate that dataset selection may be as important as dataset size when training SR models for CT enhancement.

Section: RESEARCH PAPER

Keywords: computed tomography; machine learning; super resolution; few-shot; degradation; data-scaling

Citation: L. Nepelius, P. Weinberger, C. Hastra, M. Yosifov, J. Glinz, S. Senck, Less is more: Few-shot learning in 3D super resolution using synthetic CT images, Acta IMEKO, vol. 15 (2026) no. 3, pp. 1–8. DOI: 10.21014/actaimeko.v15i3.2444

Section Editor: Emina Hadzalic, University of Rijeka, Croatia

Received: Juni 30, 2026; **In Final Form:** September 2, 2026; **Published:** September, 2026.

Funding: This work is supported by the IBW/EFRE & JTF 2021–2027 project 'HyperMAT', the project 'IMP4Zero-E: Innovative Monitoring- und Prüfverfahren für den Zero-Emission-Antriebsstrang' and the project 'VISION', funded by the Government of Upper Austria.

Corresponding Author: Lukas Nepelius, e-mail: lukas.nepelius@fh-wels.at

1. INTRODUCTION

CT is widely used in industry [1]–[3], archeology [4], cultural heritage [5], [6], and medicine [7], [8] to obtain detailed insights into the specimen of interest. CT, however, often suffers from a limited field of view (FOV) and a trade-off between resolution, scanning time, and image quality [9].

These limitations can be addressed by SR methods, which predict a high resolution (HR) image based on a low resolution (LR) input [10]. In literature, convolutional neural networks like SRCNN [11], adversarial approaches like SRGAN [12], vision transformers like SwinIR [13], and diffusion models like LDM [14] have been proposed over the years. The U-Net architecture [15] was introduced in 2015 for biomedical image applications and is still popular today. This is because of reduced data requirements and high predictive performance due to incorporation of different image resolutions related to the network. In supervised SR, paired data is required in order to train a model to learn the mapping between LR and HR [11]. However, the process of manual volume alignment on samples from different resolutions in CT is often expensive and time consuming.

Blind SR is defined as the task of producing a HR prediction without knowing the image degradation function, and therefore addresses this limitation [16]. More specifically, degradation models are used to create synthetic LR data for SR training. In literature, degradation models can be subdivided into the categories of explicit [17]–[21] and implicit [22]–[24] approaches. The former refer to algorithms that use fixed steps in the process like down-sampling and blurring, while the latter refer to models that learn the mapping between input and output implicitly, such as neural networks. In many works, authors suggest using a simple degradation process like bicubic down-sampling and Gaussian blurring followed by a final up-sampling operation [11], [17], [25]. Although efficient and easy to implement, we argue that the degradation in CT image acquisition has to be closely modeled and has potential to further enhance predictions in terms of image quality. Some authors use real scans at different resolutions and register them for SR [26], however, there still exists a lack of degradation methods specifically designed for CT images, which could have a high potential to further improve realism, efficiency, and automation. We have compared different degradation meth-

ods in previous work and found promising results with implicit degradation approaches [27], which is why we continue to work on this topic. Synthetic data can provide a realistic LR image for training SR models, but the image quality of predicted volumes can still be improved by fine-tuning on real pairs.

Transfer learning is a widely used strategy in machine learning to adapt a network trained on a source domain to a target domain [28]. A common strategy in this field is parameter-sharing or freezing parameters trained from the source domain. For instance, Yosinski et al. [29] proposed a fine-tuning strategy, in which the first few layers of the network are frozen while the deeper layers are reinitialized. The rational behind this approach is that the earlier layers in a network represent general information transferable to other domains and tasks. Amiri et al. [30] use a fine-tuning strategy where shallow layers in a U-Net are retrained and deeper ones are frozen for magnetic resonance imaging (MRI) data. Interestingly, they observe that chest x-ray scans perform better the other way around: freezing shallow weights and retraining deeper ones. Modern network architectures, such as the vision transformer (ViT) [31], employ a similar approach of retraining the last layer and freezing the rest of the transformer weights. Chen et al. [32] proposed a self-supervised learning paradigm, in which the pre-trained representations are frozen and the final multi-layer perceptron (MLP) layer is retrained in a supervised manner. Analogous to mentioned approaches, we apply a similar process of freezing earlier layers and retraining later ones.

In this study, we investigate the impact of training SR models on purely synthetic data and fine-tuning them on different quantities of real pairs. The goal of our work is to empirically discover scaling behavior in SR for CT data, and to reduce the number of manually aligned real training pairs. Due to superior performance of a proposed shallow CNN degradation model from our previous work, we employ this network and a U-Net as the SR network. The focus of this study is on aspects of data-centric machine learning [33] rather than architectural model improvements, as this field is both relevant and underrepresented in literature. The main contributions of the paper are summarized:

- The effects of pre-training a SR model on synthetic data and fine-tuning it using real CT-volume pairs are investigated.
- Data-scaling patterns in the fine-tuning of SR models are empirically analyzed by comparing different dataset configurations.

2. METHOD

The proposed workflow consists of three consecutive stages: (1) degradation model training, (2) synthetic pre-training of the SR network, and (3) fine-tuning the SR model on real paired data. A visual summary of the study approach is given in Figure 1.

Firstly, we train a degradation model based on a shallow CNN network [27], which predicts LR images based on HR ones. The training set for degradation modeling is defined as $T_d = \{(H_{\downarrow}^{(i)}, L^{(i)})\}_{i=1}^{N_d}$, where $H_{\downarrow}^{(i)}$ denotes the bicubic down-sampled HR image and $L^{(i)}$ denotes the corresponding LR volume of the i-th training pair. In total, there are N_d data pairs for degradation training. In each pair, the first element serves as the model input, whereas the second element serves as the reference for loss computation. T_d is the training set and the network can be formulated as the function $d(H_{\downarrow}^{(i)}) = \hat{L}^{(i)}$, where d is the degradation network and the input for the model is the down-sampled HR image. $\hat{L}$ represents the predicted LR volume from the degradation model. As a loss function for the degradation model, the learned perceptual image patch similarity (LPIPS) score [34] is used, which compares similarities between feature maps from two image inputs. LPIPS is originally implemented using natural 2D RGB images, but can be adapted to grayscale 3D CT images as well. The loss function from the medical open network for AI (MONAI) framework [35] is employed, which calculates the LPIPS on extracted slices of the volume and averages the scores across all three spatial dimensions. The loss function is defined as

$$Loss_d = \text{LPIPS}(\hat{L}^{(i)}, L^{(i)}). \tag{1}$$

Secondly, an adapted 3D U-Net model [36] is used in combination with the synthesized LR and real HR volumes. A training set

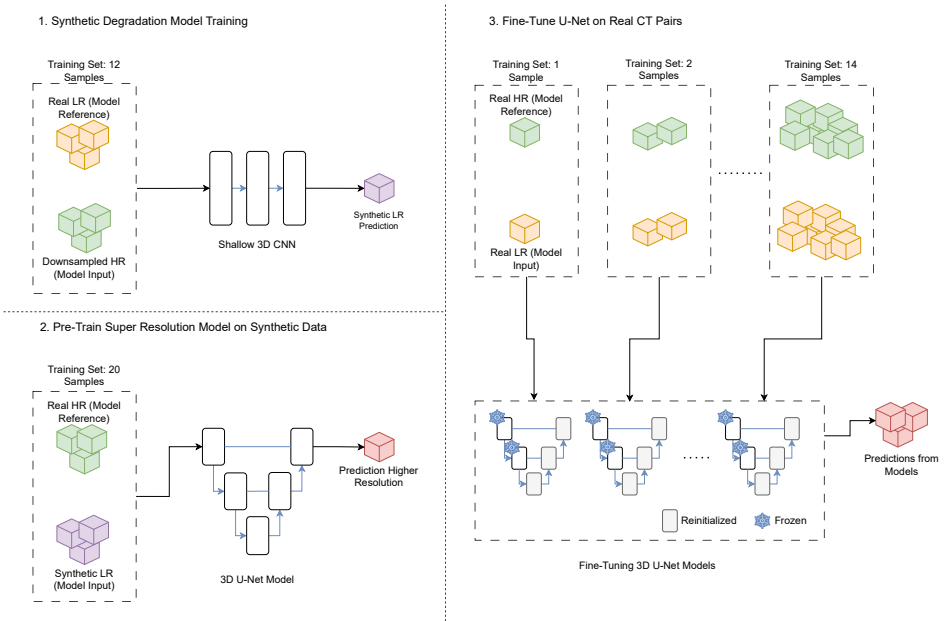


Figure 1. Overview of the three stages of machine learning experiments used in our work. First, the degradation model is trained using bicubic down-sampled HR and LR pairs. Secondly, the 3D U-Net is trained with synthetic LR and real HR pairs. Thirdly, the pre-trained network is fine-tuned with different quantities of manually aligned pairs.

$T_s = \{(\hat{L}_{\uparrow}^{(i)}, H^{(i)})\}_{i=1}^{N_s}$ is employed, where $\hat{L}_{\uparrow}^{(i)}$ is the bicubic up-sampled synthetic LR volume on entry i. The model training can be formulated as $s(\hat{L}_{\uparrow}^{(i)}) = \hat{H}^{(i)}$, where s is the SR network and $\hat{H}^{(i)}$ is the predicted sample in higher resolution. This time, the LR reconstruction serves as the input to the model and the HR sample as reference. The custom loss function in our method consists of weighted LPIPS and Huber [37] losses. A combined loss is constructed, because it captures both pixel-based and feature-based similarities in the volumes. The weights 0.1 and 0.9 were selected based on empirical experimentation with different parameters. The combined loss is formulated as

$$\begin{aligned} Loss_s =& 0.9 \times \text{LPIPS}(\hat{H}^{(i)}, H^{(i)}) + \\ & 0.1 \times \text{Huber}(\hat{H}^{(i)}, H^{(i)}). \end{aligned} \tag{2}$$

Thirdly, the trained U-Net model is fine-tuned using different quantities of real LR-HR pairs in order to observe data scaling patterns in SR networks. The main focus of our work is the identification of these patterns in order to minimize the requirements for manually aligned training pairs, as this process is time consuming and prone to alignment errors. For fine-tuning, we employ a strategy in which the weights of the encoder blocks of the U-Net are frozen and the weights from the decoder are retrained. Since encoder layers are expected to capture general image representations learned during synthetic pre-training, only decoder layers are updated. The training set sizes 1, 2, 4, 8, and 14 are evaluated, with the set sizes increasing exponentially. By increasing the dataset size exponentially, we emphasize on smaller quantities as they have the most value for reducing manual labor and computational costs. To enable a consistent and unbiased comparison of model performance, each SR experiment uses the same validation and test sets.

3. EXPERIMENTS

3.1. Datasets

An in-house dataset comprising 32 paired CT-scans from 10 aluminum foam specimen in 30 and 120 µm resolution was acquired using a Procon CT-Alpha system. This device is equipped with an X-Ray WorX XWT-240-CT source and a PerkinElmer 2048 detector. All scans were performed at a tube voltage of 230 kV using a 0.6 mm copper filter and reconstructed as 16-bit volumetric datasets. A detailed summary of scanning parameters is given in Table 1. The aluminum foam samples are chosen because of their porous microstructure and relatively homogeneous material properties. In previous work, we investigated more complex materials, such as human bones [8], and extend on this by analyzing the impact on more homogeneous structures with detailed pores.

For training the degradation model, 8 pairs of LR-HR scans, two pairs for validation and two pairs for the test set are used. For super resolution training, a data split with 14 training pairs, three validation and three test pairs are selected. To avoid information leakage between degradation learning and SR training, two disjoint subsets of aluminum foam samples with different pore structures were used. Across the experiments for SR, all of the trainings share the same validation set in order to compare the validation losses effectively across the models.

An overview of the different datasets is given in Table 2. Here, the *Base Real* setting represents a 3D U-Net trained only on real data and serves as a comparison against results from fine-tuned networks. *Base Synth* is the model trained purely on synthetic data generated from *Degradation* and *Fine 1s*, *Fine 2s*, *Fine 4s*, *Fine 8s*, and *Fine 14s* are the configurations for fine-tuning the *Base Synth* model.

Table 1. Scanning parameters for CT-acquisition and reconstruction of the aluminum foam dataset.

Parameter	High Resolution	Low Resolution
CT-System	Procon CT-Alpha	Procon CT-Alpha
Source	X-Ray WorX XWT-240-CT	X-Ray WorX XWT-240-CT
Detector	PerkinElmer 2048	PerkinElmer 2048
Tube voltage	230 kV	230 kV
Source-object dist. (SOD)	140 mm	390 mm
Source-detector dist. (SDD)	1200 mm	400 mm
Projections	1440	1040
Detector exposure time	800 ms	800 ms
Filter	0.6 mm Cu	0.6 mm Cu
Voxel size	29 µm	119 µm

Table 2. Data splits used in the study. The *Base Synth* model is pre-trained on synthetic data generated by the degradation model and fine-tuned on different quantities of real pairs, see *Fine 1s* to *Fine 14s*.

Model	Train Set	Validation Set	Test Set
Degradation	8	2	2
Base Synth	14	3	3
Fine 1s	1	3	3
Fine 2s	2	3	3
Fine 4s	4	3	3
Fine 8s	8	3	3
Fine 14s	14	3	3
Base Real	14	3	3

3.2. Training details

For degradation model training, the entire LR-HR samples are processed, which have an average dimension of $418 \times 414 \times 289$ voxels. Here, the HR images are bicubic down-sampled and serve as input to the model while the real LR samples serve as reference. The model is trained for 100 epochs with a batch size of one and 8 sample pairs as train set. The total amount of training iterations amounts to 800 steps. A learning rate of 0.001 and the AdamW [38] optimizer are used.

During training the SR models, a pre-processing pipeline is executed, where first the LR samples are bicubic up-sampled and a number of 512 sub-volumes of voxel size $128 \times 128 \times 128$ are extracted from random positions of the larger volume. These sub-volumes are augmented with a chance of 10 % per pair by rotation, flipping, brightness and contrast adaptions in 3D. The rotation and flipping operations are utilized from the MONAI framework, while the brightness and contrast adaptions are implemented in Python by element-wise multiplication and addition

$$A(I) = I \times c + b. \tag{3}$$

The intensity augmentation function A with input image I is defined as the element-wise multiplication with contrast factor c and addition with brightness constant b. We use random numbers for c in interval $[0.85; 1.15]$ and for b in interval $[-0.05; 0.05]$. The model is trained for 20 epochs with a batch size of three using all six GPUs in parallel. In total, the training consists of 2389 steps. The learning rate used is 0.001 and the AdamW optimizer is employed.

The training of the fine-tuning experiments stays almost same compared to the settings used for the synthetic SR model. We use

the pre-trained weights from the encoder of the SR U-Net and retrain the decoder weights. An early stopping strategy is employed from Pytorch Lightning [39], in which the training run is stopped if the validation loss does not decrease significantly for four epochs. A threshold of 0.0001 is defined for the validation loss to count as improvement. By utilizing early stopping and varying the number of training steps/epochs, we can easily compare the validation and test metrics of different models. This is because models with less data converge faster than those trained with more data, which often require more training steps.

3.3. Testing details

The testing process of the degradation model consists of using a down-sampled HR image, which was not seen by the model during training and applying the pre-trained degradation model.

Similarly for SR testing, a sample pair from the test set is used to evaluate the model. Due to larger volume dimensions, the pre-trained model is applied on $128 \times 128 \times 128$ sub-volumes, which are overlapping by 25 % in a sliding window fashion using the *SlidingWindowInferer* from MONAI. The predicted sub-volumes are combined afterwards by weighting the overlapping pixel values using a Gaussian distribution. Test metrics are calculated on a $640 \times 640 \times 640$ voxel center crop extracted from the pairs.

3.4. Image quality assessment

Classic image quality metrics such as peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) [40] are employed. PSNR calculates a pixel-wise similarity between prediction and reference image, and can be defined using the mean squared error (MSE):

$$\begin{aligned} MSE(Y,\hat{Y}) =& \frac{1}{n \times m \times o} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \sum_{k=0}^{o-1} (Y_{i,j,k} - \hat{Y}_{i,j,k})^2 \\ PSNR(Y,\hat{Y}) =& 20 \times \log_{10}(MAX_Y) - \\ & 10 \times \log_{10}(MSE(Y,\hat{Y})). \end{aligned} \tag{4}$$

Here, Y represents the reference image and $\hat{Y}$ the prediction with dimensions $n \times m \times o$. MAX_Y is the maximum pixel value of the reference image. The formulation is adapted from the *PSNRMetric* class from MONAI.

SSIM on the other hand takes into account the structural, brightness and contrast aspects of the two images. It can be defined as

$$SSIM(Y,\hat{Y}) = \frac{(2\mu_Y \mu_{\hat{Y}} + c_1)(2\sigma_{Y\hat{Y}} + c_2)}{(\mu_Y^2 + \mu_{\hat{Y}}^2 + c_1)(\sigma_Y^2 + \sigma_{\hat{Y}}^2 + c_2)}. \tag{5}$$

μ denotes the mean pixel and σ the sample variance value. c_1 and c_2 are small constant values to prevent zero-division issues.

Feature similarity (FSIM) [41] further extends the concepts of SSIM by focusing on low-level image features, as they represent the human visual system (HVS) more closely. The phase congruency (PC) and gradient magnitude (GM) from both images are extracted as main features for FSIM calculation. First, a similarity between the PC features S_{PC} and then a similarity between GM features S_G is calculated

Table 3. Average image quality metrics of the degradation model on the test set. Bicubic down-sampling of the HR volume serves as performance baseline. Best scores are highlighted **bold**.

Method	PSNR ↑	SSIM ↑	FSIM ↑	LPIPS ↓
Bicubic	30.00	0.878	0.976	0.398
Degradation	**32.88**	**0.945**	**0.984**	**0.047**

$$\begin{aligned} S_{\mathrm{PC}}(x) &= \frac{2PC_1(x) \times PC_2(x) + T_1}{PC_1^2(x) + PC_2^2(x) + T_1} \\ S_{\mathrm{G}}(x) &= \frac{2G_1(x) \times G_2(x) + T_2}{G_1^2(x) + G_2^2(x) + T_2} \\ S_{\mathrm{L}}(x) &= S_{\mathrm{PC}}(x) \times S_{\mathrm{G}}(x) \\ PC_{\mathrm{m}}(x) &= \max(PC_1(x), PC_2(x)) \\ FSIM &= \frac{\sum_{x \in \Omega} S_{\mathrm{L}}(x) \times PC_{\mathrm{m}}(x)}{\sum_{x \in \Omega} PC_{\mathrm{m}}(x)}. \end{aligned} \tag{6}$$

Here, PC_1, PC_2, G_1 and G_2 are the PC and GM values from both images. T_1 and T_2 represent constant small values for stabilizing the terms. Ω is defined as the image space and x is a pixel value.

LPIPS is defined as the L_2 distance of feature representations produced by a neural network between two images [34]. The network is pre-trained on the ImageNet [42] dataset and extracts abstract features to compare the similarity of two images. The metric is calculated by the equation

$$LPIPS = \sum_l \frac{1}{H_l W_l} \sum_{h,w} \| w_l \odot (\hat{y}^l_{hw} - \hat{y}^l_{0hw}) \|_2^2, \tag{7}$$

where $\hat{y}^l$ and $\hat{y}^l_0$ denote the feature activations extracted from a pre-trained network at layer l. Each feature map has spatial dimensions $H_l \times W_l$. The vector w_l represents learned channel-wise weights and $\odot$ is an element-wise multiplication. LPIPS is originally designed for 2D images, but can be adapted to 3D as well. We use the implementation from the MONAI framework, where 50 % of 2D slices in one axis are randomly selected, and the results from three axis are averaged afterwards to get the final score.

We report image quality metrics that focus on different image features in order to provide a robust evaluation of machine learning models.

4. RESULTS AND DISCUSSION

4.1. Degradation model

Table 3 summarizes the image quality metrics obtained on the test set from the degradation model. Here, we compare the prediction $\hat{L}$ from the degradation network with the real LR images. The learned degradation model produces LR images that more closely match the real LR distribution compared to simple bicubic down-sampling, as indicated by improved image similarity scores.

Figure 2 shows snapshots from slices of the model prediction, input and reference samples. It can be seen that the predicted LR samples from the degradation model include structures that are very similar to those in the real LR images. However, one limitation of this approach is that occasional duplication artifacts appear in fine structural regions.

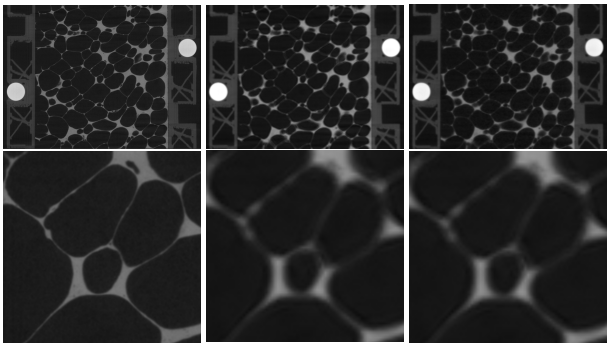


Figure 2. HR input and LR reference used for degradation model training. A zoomed region of the first row is shown in the second row of images.

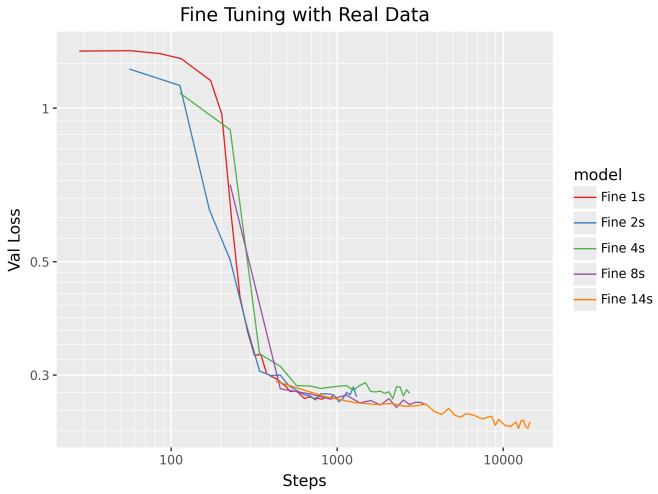


Figure 3. The validation loss curves from the fine-tuning runs. Smaller datasets tend to overfit faster while larger ones require more training steps and early stopping is triggered later. The validation loss used is our weighted combination of LPIPS and Huber loss.

4.2. Data efficiency study

In Figure 3, we plot the validation loss curves on double logarithmic scales for the fine-tuning experiments. As stated in the experiments section, we executed every run with the same early stopping strategy and the validation set stays identical across the runs. The early stopping strategy is chosen over a fixed number of epochs/steps because models trained with smaller datasets tend to overfit faster compared to larger datasets. With early stopping, every run terminates before overfitting occurs, which is visible in the plot. Additionally, a general trend can be observed that the validation loss values decline if the model is fine-tuned with more data. However, the *Fine 4s* model shows slightly worse loss values, which indicates that some pairs can also have a negative impact on fine-tuning.

Test similarity scores are calculated and visualized as boxplots in Figure 4. The first entry on the x-axis of the boxplot is a simple bicubic up-sampling of the real LR sample, which serves as a baseline for model performance comparison. Average test metric values are listed in the table on the right side.

The boxplots show that the bicubic baseline performance sometimes improves upon model predictions in terms of SSIM and LPIPS scores. This is likely due to small misalignments between the LR-HR pairs and the sensitivity on pixel-wise errors related to these metrics. However, the table with average test metrics show that *Fine 2s* generally yields the best image quality performance. Across the evaluation metrics for the fine-tuning experiments, the model trained with four pairs *Fine 4s* performs slightly worse compared to models with fewer data pairs, *Fine 1s* and *Fine 2s*. A possible explanation is that when using less samples overall, individual training pairs have a comparatively large influence on model adaptation. Some samples are less representative of the overall data distribution and may affect the optimization disproportionally, such that adding further training pairs does not necessarily improve model generalization. In our case, the image quality metrics indicate diminishing returns after two sample-pairs. Additionally, *Base Synth* scores similarly in image quality metrics to *Base Real*, which further suggests that synthetic LR data already can produce strong SR predictions in this setting.

Figure 5 shows a qualitative comparison of the SR model predictions on 2D slices extracted from the reconstructed volumes. The results highlight the effectiveness of the *Synth Base* model, as the model captures a high contrast and sharp edges trained purely with synthetic data. There are some slight structural differences and noise patterns introduced in *Fine 1s*, but the images remain relatively similar. The other model predictions observe minor differences, only slight adaptations to contrast or brightness. Compared to the input LR image, all model predictions show enhanced sharpness and contrast. The visual results further suggest that even a purely synthetic training-set or a fine-tuned model on a small number of real paired samples can significantly improve the image quality.

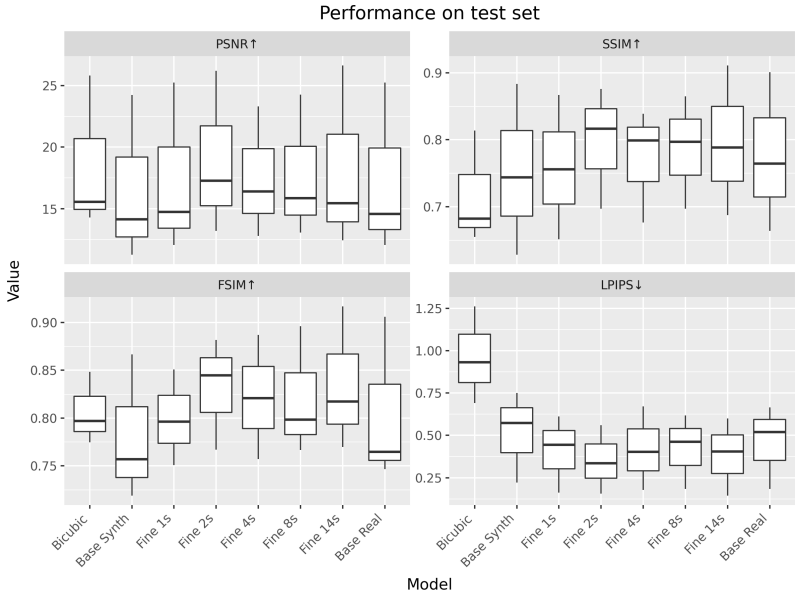
5. CONCLUSION AND OUTLOOK

In this work, a three-stage study is applied consisting of degradation model learning, pre-training a SR model with synthetic degraded data, and fine-tuning it on different amounts of real data. Within the scope of the investigated aluminum foam dataset, the results suggest that even without real LR samples, the model seems to produce a strong baseline for improving the image quality of CT-data. Additionally, a small number of real training pairs can further improve the synthetically pre-trained SR model, thereby reducing the need for manually aligned datasets. Experiments also revealed that increasing the amount of training data does not necessarily lead to consistent improvements in performance. This observation suggests that dataset selection may be as important as dataset size when adapting SR models to real CT acquisitions.

The limitations of our study are the use of a relatively small in-house dataset consisting of a single material system and an identical scanning setup with reduced CT-artefacts. Consequently, the observed data efficiency and fine-tuning behavior may depend on the structural characteristics of the material, as well as the employed CT-acquisition setup. Future work will improve the generalizability of the findings by extending the dataset with different material systems, CT-artefacts, geometries, and imaging settings. In addition, data selection and sample importance methods are a promising direction for further reduction of data requirements and an improvement of image quality.

AUTHORS' CONTRIBUTION

Patrick Weinberger: Methodology, Conceptualization, Funding acquisition, Writing – review & editing. **Caroline Hastra**: Conceptualization, Writing – review & editing. **Miroslav Yosifov**: Writing – review & editing. **Jonathan Glinz**: Funding acquisition, Writing – review & editing. **Sascha Senck**: Data curation, Conceptualization, Supervision



	PSNR ↑	SSIM ↑	FSIM ↑	LPIPS ↓
Bicubic	18.56	0.717	0.807	0.901
Base Synth	16.55	0.752	0.781	0.516
Fine 1s	17.35	0.758	0.799	0.406
Fine 2s	**18.89**	**0.796**	0.831	**0.352**
Fine 4s	17.51	0.771	0.822	0.418
Fine 8s	17.74	0.786	0.821	0.421
Fine 14s	18.17	**0.796**	**0.835**	0.383
Base Real	17.29	0.777	0.806	0.456

Figure 4. Image quality assessment on the test set. A single boxplot shows the distribution of values across the three pairs and uses *Bicubic* and *Base Real* as comparison to SR predictions and synthetic data. On the right side, average values are shown for a more detailed numerical comparison of results. The best scores per metric are highlighted **bold**.

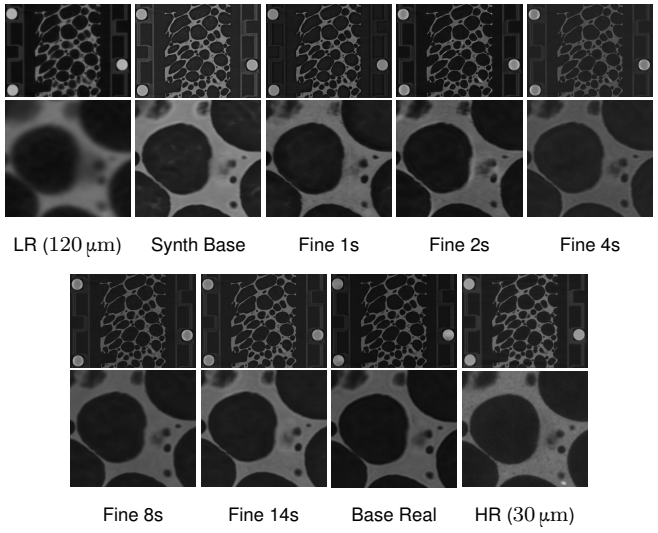


Figure 5. 2D slices extracted form the reconstructed scans and model predictions. Second rows depict a zoomed in area of the images in the first row. *Synth Base* is the model trained only on synthetic data and is used for fine-tuning. *Base Real* is completely trained with real data and serves to compare the fine-tuned models.

ACKNOWLEDGMENT

This work is supported by the IBW/EFRE & JTF 2021–2027 project 'HyperMAT', the project 'IMP4Zero-E: Innovative Monitoring- und Prüfverfahren für den Zero-Emission-Antriebsstrang', and the project 'VISION', funded by the Government of Upper Austria.

REFERENCES

[1] International Atomic Energy Agency, An Introduction to Practical Industrial Tomography Techniques for Non-destructive Testing (NDT), ser. IAEA-TECDOC, International Atomic Energy Agency, Vienna, 2020, no. 1931, ISBN: 978-92-0-120920-7.

[2] S. Senck, J. Glinz, S. Heupl, J. Kastner, K. Trieb, U. Scheithauer, S. S. Dahl, M. B. Jensen, Ceramic additive manufacturing and microstructural analysis of tricalcium phosphate implants using X-ray microcomputed tomography, Open Ceramics, vol. 19, 2024, p. 100628.
DOI: 10.1016/j.oceram.2024.100628

[3] J. Bennett, K. Gonzalez, J. Gruber, G. Mayr, J. Glinz, J. Kastner, T. Hötzer, B. Fischer, S. Senck, Multi-modal imaging and non-destructive evaluation of a carbon fiber overwrapped pressure vessel (COPV), e-Journal of Nondestructive Testing, vol. 29, 2024, no. 3.
DOI: 10.58286/29270

[4] S. Hughes, CT Scanning in Archaeology, Computed Tomography - Special Applications, L. Saba Ed., InTech, 2011, ISBN: 978-953-307-723-9.

[5] F. G. Bossema, W. J. Palenstijn, A. Heginbotham, M. Corona, T. Van Leeuwen, R. Van Liere, J. Dorscheid, D. O'Flynn, J. Dyer, E. Hermens, K. J. Batenburg, Enabling 3D CT-scanning of cultural heritage objects using only in-house 2D X-ray equipment in museums, Nature Communications, vol. 15, 2024, no. 1, p. 3939.
DOI: 10.1038/s41467-024-48102-w

[6] M. Vopálenský, L. Machačko, Twinned Orthogonal Adjustable Tomograph Device in Conservation Services: Case Studies of a Recent X-ray Investigation into Baroque Oil Paintings, ChemPlusChem, vol. 90, 2025, no. 7.
DOI: 10.1002/cplu.202500063

[7] H. Xiao, Z. Yang, T. Liu, S. Liu, X. Huang, J. Dai, Deep learning for medical imaging super-resolution: A comprehensive review, Neurocomputing, vol. 630, 2025, p. 129667.
DOI: https://doi.org/10.1016/j.neucom.2025.129667

[8] S. Senck, P. Weinberger, L. Nepelius, A. Haghofer, B. Woegerer, J. Glinz, M. Yosifov, L. Behammer, J. Kastner, K. Trieb, E. Kranioti, S. Winkler, Optimizing µCT resolution in tarsal bones: A comparative study of super-resolution models for trabecular bone analysis, Tomography of Materials and Structures, vol. 8, 2025, p. 100063.
DOI: 10.1016/j.tmater.2025.100063

[9] L. L. Frazer, N. Louis, W. Zbijewski, J. Vaishnav, K. Clark, D. P. Nicolella, Super-resolution of clinical CT: Revealing microarchitecture in whole bone clinical CT image data, Bone, vol. 185, 2024.
DOI: 10.1016/j.bone.2024.117115

[10] P. Purkait, B. Chanda, Super Resolution Image Reconstruction Through Bregman Iteration Using Morphologic Regularization, IEEE Transactions on Image Processing, vol. 21, 2012, no. 9, pp. 4029–4039.
DOI: 10.1109/TIP.2012.2201492

[11] C. Dong, C. C. Loy, K. He, X. Tang, Image Super-Resolution Using Deep Convolutional Networks, Transactions on Pattern Analysis and Machine Intelligence, vol. 38, 2016, no. 2, pp. 295–307.
DOI: 10.1109/TPAMI.2015.2439281

[12] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, W. Shi, Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network, Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Honolulu, HI, 2017, pp. 105–114.
DOI: 10.1109/CVPR.2017.19

[13] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, R. Timofte, SwinIR: Image Restoration Using Swin Transformer, International Conference on Computer Vision Workshops (ICCVW), IEEE, Montreal, BC, Canada, 2021, pp. 1833–1844.
DOI: 10.1109/ICCVW54120.2021.00210

[14] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, B. Ommer, High-Resolution Image Synthesis with Latent Diffusion Models, Conference on Computer Vision and Pattern Recognition (CVPR), arXiv, 2022, pp. 10 674–10 685.
DOI: 10.1109/CVPR52688.2022.01042

[15] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation, Medical Image Computing and Computer-Assisted Intervention (MICCAI), Springer International Publishing, 2015, N. Navab, J. Hornegger, W. M. Wells, A. F. Frangi (editors), vol. 9351, pp. 234–241.
DOI: 10.1007/978-3-319-24574-4_28

[16] A. Liu, Y. Liu, J. Gu, Y. Qiao, C. Dong, Blind Image Super-Resolution: A Survey and Beyond, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, pp. 1–19.
DOI: 10.1109/TPAMI.2022.3203009

[17] K. Zhang, J. Liang, L. Van Gool, R. Timofte, Designing a Practical Degradation Model for Deep Blind Image Super-Resolution, International Conference on Computer Vision (ICCV), IEEE, 2021, pp. 4771–4780.
DOI: 10.1109/ICCV48922.2021.00475

[18] X. Liu, S. Su, W. Gu, T. Yao, J. Shen, Y. Mo, Super-Resolution Reconstruction of CT Images Based on Multi-scale Information Fused Generative Adversarial Networks, Annals of Biomedical Engineering, vol. 52, 2024, no. 1, pp. 57–70.
DOI: 10.1007/s10439-023-03412-w

[19] J. Xiao, H. Yong, L. Zhang, Degradation Model Learning for Real-World Single Image Super-Resolution, Computer Vision - ACCV, Springer International Publishing, 2021, vol. 12623, pp. 84–101.
DOI: 10.1007/978-3-030-69532-3_6

[20] X. Wang, L. Xie, C. Dong, Y. Shan, Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data, International Conference on Computer Vision Workshops (ICCVW), IEEE, Montreal, BC, Canada, 2021, pp. 1905–1914.
DOI: 10.1109/ICCVW54120.2021.00217

[21] K. Zhang, L. Van Gool, R. Timofte, Deep Unfolding Network for Image Super-Resolution, Conference on Computer Vision and Pattern Recognition (CVPR), arXiv, 2020.
DOI: 10.48550/ARXIV.2003.10428

[22] Y. Tian, L. Fan, P. Isola, H. Chang, D. Krishnan, StableRep: Synthetic Images from Text-to-Image Models Make Strong Visual Representation Learners, Conference on Neural Information Processing Systems (NeurIPS), arXiv, 2023, vol. 36, pp. 48 382–48 402.
DOI: 10.48550/ARXIV.2306.00984

[23] A. Ferreira, K. Xie, C. Wilpert, G. Correia, F. B. Ordonez, T. G. Oliveira, M. Bode, R. Siepmann, F. Hölzle, R. Röhrig, J. Kleesiek, D. Truhn, J. Egger, V. Alves, B. Puladi, Enhancing Privacy: The Utility of Stand-Alone Synthetic CT and MRI for Tumor and Bone Segmentation, arXiv, 2025.
DOI: 10.48550/ARXIV.2506.12106

[24] T. Zhao, W. Ren, C. Zhang, D. Ren, Q. Hu, Unsupervised Degradation Learning for Single Image Super-Resolution, arXiv, 2018.
DOI: 10.48550/ARXIV.1812.04240

[25] J. Chi, Z. Sun, H. Wang, P. Lyu, X. Yu, C. Wu, CT image super-resolution reconstruction based on global hybrid attention, Computers in Biology and Medicine, vol. 150, 2022, p. 106112.
DOI: 10.1016/j.compbiomed.2022.106112

[26] A. Klos, L. Salvo, P. Lhuissier, Super-resolution X-ray tomography using deep learning applied to the 3D quantification of defects in lattice structures, Scientific Reports, vol. 15, 2025, no. 1.
DOI: 10.1038/s41598-025-20372-4

[27] N. Lukas, W. Patrick, M. Yosifov, L. Behammer, U. Bodenhofer, S. Sascha, Impact of Synthetic and Real CT Training Datasets for Deep Learning Super Resolution Models, e-Journal of Nondestructive Testing, vol. 31, 2026, no. 3.
DOI: 10.58286/32559

[28] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, Q. He, A Comprehensive Survey on Transfer Learning, Proceedings of the IEEE, vol. 109, 2021, no. 1, pp. 43–76.
DOI: 10.1109/JPROC.2020.3004555

[29] J. Yosinski, J. Clune, Y. Bengio, H. Lipson, How transferable are features in deep neural networks? Advances in Neural Information Processing Systems (NIPS), Curran Associates, Inc., 2014, vol. 27.
DOI: 10.48550/ARXIV.1411.1792

[30] M. Amiri, R. Brooks, H. Rivaz, Fine tuning U-Net for ultrasound image segmentation: which layers? arXiv, 2020.
DOI: 10.48550/ARXIV.2002.08438

[31] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, N. Houlsby, An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale, International Conference on Learning Representations (ICLR), arXiv, 2020.
DOI: 10.48550/ARXIV.2010.11929

[32] X. Chen, S. Xie, K. He, An Empirical Study of Training Self-Supervised Vision Transformers, International Conference on Computer Vision (ICCV), 2021.
DOI: 10.48550/ARXIV.2104.02057

[33] D. Zha, Z. P. Bhat, K.-H. Lai, F. Yang, Z. Jiang, S. Zhong, X. Hu, Data-centric Artificial Intelligence: A Survey, arXiv, 2023.
DOI: 10.48550/ARXIV.2303.10158

[34] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, O. Wang, The Unreasonable Effectiveness of Deep Features as a Perceptual Metric, Conference on Computer Vision and Pattern Recognition (CVPR), arXiv, 2018.
DOI: 10.48550/ARXIV.1801.03924

[35] M. J. Cardoso, W. Li, R. Brown, N. Ma, E. Kerfoot, Y. Wang, B. Murrey, A. Myronenko, C. Zhao, D. Yang, V. Nath, Y. He, Z. Xu, A. Hatamizadeh, W. Zhu, Y. Liu, M. Zheng, Y. Tang, I. Yang, M. Zephyr, B. Hashemian, S. Alle, M. Z. Darestani, C. Budd, M. Modat, T. Vercauteren, G. Wang, Y. Li, Y. Hu, Y. Fu, B. Gorman, H. Johnson, B. Genereaux, B. S. Erdal, V. Gupta, A. Diaz-Pinto, A. Dourson, L. Maier-Hein, P. F. Jaeger, M. Baumgartner, J. Kalpathy-Cramer, M. Flores, J. Kirby, L. A. D. Cooper, H. R. Roth, D. Xu, D. Bericat, R. Floca, S. K. Zhou, H. Shuaib, K. Farahani, K. H. Maier-Hein, S. Aylward, P. Dogra, S. Ourselin, A. Feng, MONAI: An open-source framework for deep learning in healthcare, arXiv, 2022.
DOI: 10.48550/ARXIV.2211.02701

[36] T. Falk, D. Mai, R. Bensch, O. Çiçek, A. Abdulkadir, Y. Marrakchi, A. Böhm, J. Deubner, Z. Jäckel, K. Seiwald, A. Dovzhenko, O. Tietz, C. Dal Bosco, S. Walsh, D. Saltukoglu, T. L. Tay, M. Prinz, K. Palme, M. Simons, I. Diester, T. Brox, O. Ronneberger, U-Net: deep learning for cell counting, detection, and morphometry, Nature Methods, vol. 16, 2019, no. 1, pp. 67–70.
DOI: 10.1038/s41592-018-0261-2

[37] P. J. Huber, Robust Estimation of a Location Parameter, The Annals of Mathematical Statistics, vol. 35, 1964, no. 1, pp. 73–101.
DOI: 10.1214/aoms/1177703732

[38] I. Loshchilov, F. Hutter, Decoupled Weight Decay Regularization, International Conference on Learning Representations (ICLR), arXiv, 2017.
DOI: 10.48550/ARXIV.1711.05101

[39] W. Falcon, PyTorch Lightning, 2020.
https://github.com/Lightning-AI/lightning

[40] Zhou Wang, A. Bovik, H. Sheikh, E. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE

Transactions on Image Processing, vol. 13, 2004, no. 4, pp. 600–612.
DOI: 10.1109/TIP.2003.819861

[41] Lin Zhang, Lei Zhang, Xuanqin Mou, D. Zhang, FSIM: A Feature Similarity Index for Image Quality Assessment, IEEE Transactions on Image Processing, vol. 20, 2011, no. 8, pp. 2378–2386.
DOI: 10.1109/TIP.2011.2109730

[42] J. Deng, W. Dong, R. Socher, L.-J. Li, Kai Li, Li Fei-Fei, ImageNet: A large-scale hierarchical image database, Conference on Computer Vision and Pattern Recognition, IEEE, Miami, FL, 2009, pp. 248–255.
DOI: 10.1109/CVPR.2009.5206848