Received: 4 August 2022 | Revised: 6 April 2023 | Accepted: 25 April 2023

DOI: 10.1002/mp.16482

RESEARCH ARTICLE

MEDICAL PHYSICS

# Bone-GAN: Generation of virtual bone microstructure of high resolution peripheral quantitative computed tomography

**Felix S. L. Thomsen**[1,2,3] | **Emmanuel Iarussi**[1,4] | **Jan Borggrefe**[2] | **Steven K. Boyd**[5] | **Yue Wang**[6] | **Michele C. Battié**[7]

[1]National Scientific and Technical Research Council (CONICET), Buenos Aires, Argentina

[2]Department of Radiology, Neuroradiology and Nuclear Medicine, Johannes Wesling University Hospital, Ruhr University Bochum, Bochum, Germany

[3]Department of Electrical and Computer Engineering, Institute for Computer Science and Engineering, National University of the South (DIEC-ICIC-UNS), Bahía Blanca, Argentina

[4]Laboratory of Artificial Intelligence, University Torcuato Di Tella, Buenos Aires, Argentina

[5]McCaig Institute for Bone and Joint Health, University of Calgary, Canada

[6]Spine lab, Department of Orthopedic Surgery, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China

[7]Common Spinal Disorders Research Group, Faculty of Rehabilitation Medicine, University of Alberta, Edmonton, Canada

**Correspondence**
Felix Thomsen, Instituto de Ciencias e Ingeniería de la Computación-ICIC-CONICET, San Andrés 800, Bahía Blanca 8000, Argentina.
Email: felix.thomsen@uns.edu.ar

**Funding information**
Salesforce, Grant/Award Number: Einstein AI Grant; Consejo Nacional de Investigaciones Científicas y Técnicas, Grant/Award Number: PIP2021-11220200102981CO

## Abstract

**Background:** Data-driven development of medical biomarkers of bone requires a large amount of image data but physical measurements are generally too restricted in size and quality to perform a robust training.

**Purpose:** This study aims to provide a reliable in silico method for the generation of realistic bone microstructure with defined microarchitectural properties. Synthetic bone samples may improve training of neural networks and serve for the development of new diagnostic parameters of bone architecture and mineralization.

**Methods:** One hundred-fifty cadaveric lumbar vertebrae from 48 different male human spines were scanned with a high resolution peripheral quantitative CT. After prepocessing the scans, we extracted 10,795 purely spongeous bone patches, each with a side length of 32 voxels (5 mm) and isotropic voxel size of 164 $\mu$m. We trained a volumetric generative adversarial network (GAN) in a progressive manner to create synthetic microstructural bone samples. We then added a style transfer technique to allow the generation of synthetic samples with defined microstructure and gestalt by simultaneously optimizing two entangled loss functions. Reliability testing was performed by comparing real and synthetic bone samples on 10 well-understood microstructural parameters.

**Results:** The method was able to create synthetic bone samples with visual and quantitative properties that effectively matched with the real samples. The GAN contained a well-formed latent space allowing to smoothly morph bone samples by their microstructural parameters, visual appearance or both. Optimum performance has been obtained for bone samples with voxel size $32 \times 32 \times 32$, but also samples of size $64 \times 64 \times 64$ could be synthesized.

**Conclusions:** Our two-step-approach combines a *parameter-agnostic* GAN with a *parameter-specific* style transfer technique. It allows to generate an unlimited anonymous database of microstructural bone samples with sufficient realism to be used for the development of new data-driven methods of bone-biomarkers. Particularly, the style transfer technique can generate datasets of bone samples with specific conditions to simulate certain bone pathologies.

**KEYWORDS**
bone microstructure, gestalt, progressive generative adversarial network, structural morphing, XtremeCT

## 1 | INTRODUCTION

Digital generation of microstuctural bone has the potential to create an anonymous and unlimited dataset for the development of new diagnostic microarchitectural parameters. Such a method may gain insight in the microstructure of osteopenic or osteoporotic bone, or help to analyze the performance of new microstructural parameters. It can also be used as a basis for an in silico simulation of clinical CT scans[1] to provide pairs of synthetic ground-truth and clinical CT that can be used then to train data-driven methods for CT noise suppression.[2] Additionally, when the latent space of the generative network is smooth and adequately well formed, one might be able to develop a simulation that transforms a bone sample into another just as it would degenerate under osteoporosis.

Recent high-quality artificial intelligence (AI) research was focused on solving difficult medical image problems for instance by using generative adversarial networks (GANs). GANs in medical imaging are often used to add non-existent image information to real patient scans, for instance for rib suppression from chest x-rays in order to improve the reliability of pulmonary disease diagnosis,[3] to synthesize contrast-enhanced magnetic resonance images[4] or to enhance the response of contrast agents in computed tomography imaging.[5] GANs are also used to reconstruct CT images from planar X-rays[6] or to synthesize high resolution diffusion tensor images from low resolution ones.[7] We focus here on the development of a robust and reliable GAN[8] to generate an artificial dataset of bone tissue with sufficient realism and variance but, in contrast to the concept of Cycle-GANs, without requiring to be able to copy a given real bone sample to perfection. Peña-Solórzano et al. proposed in this context a numerical non-AI model for the generation of short stacks of synthetic CT slices of bone microstructure.[9] They used a pipeline of random noise, Gaussian and Lorentzian distributions, and the Fourier transformation. Another approach combines GANs with a 3D tessellation mechanism that allows expansion of patch size from $64^3$ to $120^3$ voxels or more.[10] However, shortcomings of both methods are the lack of realism and inter-variability of the generated specimen. Thus, to our knowledge, a reliable method for the generation of realistic spongeous training data for neural networks or for the simulation of specific bone conditions was not proposed yet.

The development of bone microstructural parameters, such as bone volume by total volume, trabecular separation or plate-to-rod ratio, requires accurately modelled sample data with respect to the objective of investigation. For the development of parameters with non-AI techniques, that is based on physical and morphological properties, one could use simple 3D textures that are constructed by rods and plates of different orientations, positions and thicknesses.[11,12] However, training of complex data-driven approaches such as neural networks need realistic training and testing data that reflect the natural variation and entropy of real human bone and their unsuitability to simulate bone deterioration of the same specimen as assessed on long-term studies. While samples of real human bone could be used directly to prove a theoretical machine learning mechanism, they are often too limited in size to achieve clinically reliable results. Hence, the availability of specific GANs that can generate unlimited training data is a necessary requirement for data-driven AI. In comparison to conditional GANs, we propose a two-step approach with a *parameter-agnostic* and a *parameter-specific* component of the GAN network with the advantage to not require time consuming retraining of the generative model if the user desires to incorporate new control properties. Hence, our method can easily be adjusted to any specific research problem that is defined by a given constraint of bone microarchitecture. We hypothesize that our synthetic dataset will significantly improve performance and reliability of many existing deep-learning methods of bone microstructural analysis.[13,14]

## 2 | MATERIALS AND METHODS

We propose a model-entangled two-step approach. The first part is a *parameter-agnostic* GAN, trained on isotropic real high-resolution peripheral quantitative computed tomography (HRpQCT) patches of $32^3$ voxels (box of 5mm diameter) that generates in silico samples of the same kind (Figure 1). Synthetic samples are generated from a 32-dimensional latent code vector $\boldsymbol{z}$. The second part is an optimizer that provides *parameter-specific* control over the learned latent space to customize the morphological properties of the output volumes (Figure 2). We implemented 10 microstructural parameters that can be used in any combination to morph a given synthetic bone sample towards a reference microstructure. Equivalently, we are able to vary the gestalt or appearance of any synthetic bone sample and fixing the microstructural reference parameters.

### 2.1 | Preparation of real bone patches

One hundred-fifty ex-vivo human lumbar vertebrae of 48 male donors were extracted and cleaned, removing all bone marrow, and scanned on a HRpQCT (XtremeCT, Scanco Medical AG, Brüttisellen, Switzerland) with isotropic voxel size of 82 $\mu$m and implicit calibration to density values. The mean age was 50 years (range 21–64) and the subjects were all employed

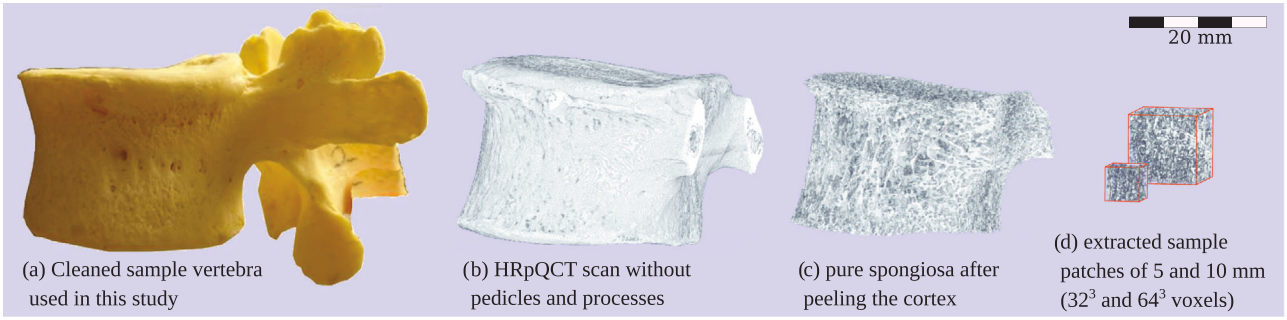


**FIGURE 1** (a) One of 150 prepared sample vertebrae used in this study. (b) Sample HRpQCT scan after cropping pedicles and processes. (c) The spongiosa of the vertebra obtained by peeling 2.5 mm (15 voxels) of the border. (d) Two sample patches of 5 mm ($32 \times 32 \times 32$ voxels) and 10 mm ($64 \times 64 \times 64$ voxels) used for training the proposed GAN. We extracted in total 10 795 patches of 5 mm and 3454 patches of 10 mm. GAN, generative adversarial network; HRpQCT, high-resolution peripheral quantitative computed tomography.

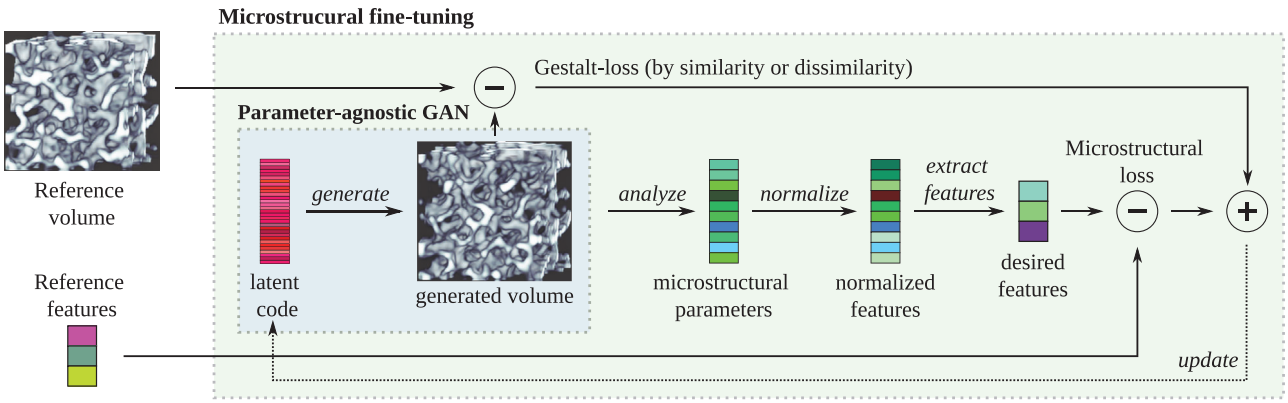


**FIGURE 2** Optimization of the latent code that represents a bone patch with high (or low) structural similarity to a reference patch and simultaneously microstructural features that fit a given set of reference features. The feature space is a subset of all normalized microstructural parameters and principal components. The parameter-agnostic GAN is a readily trained and static component during fine-tuning.

before death and with a short history of illness that was not necessarily related to bone degeneration. The vertebrae were sampled and prepared in Finland,[15] and later scanned in Canada[16] for research, being conducted by the Common Spinal Disorders Research Group, University of Alberta. There was no data exclusion applied and the study was approved by the local ethics commission.

We imported the native Scanco-files into Python and sampled them down by a factor of 2 to an isotropic voxel size of 164 $\mu$m, averaging each eight voxels of the original volume to one voxel of the new volume. This increased the signal-to-noise ratio and voxel-size while preserving all relevant microstructural information.[17] We removed pedicles and processes with a semi-automatic segmentation and patching procedure that was developed for this study. First, we cropped the vertebra body with a manually placed bounding box. By applying a combination of threshold segmentation, morphological closing and binary propagation, we selected then all voxels of the vertebral body. Peeling of the vertebral body was applied by removing voxels closer than 15 voxels (2.5 mm) to the background. Finally, we used a grid with an isotropic grid-size of 32 voxels (box of 5 mm) to extract a total of 10,795 purely spongeous and isotropic non-overlapping bone patches. This set was augmented to 172,720 patches by employing all 16 axis-aligned combinations of rotations and reflections preserving the vertical load direction of the samples. The specimens showed unnaturally low integral bone density, since they contained air in place of bone marrow. To reach microstructural parameters as computed from bone samples with bone marrow, we linearly mapped the raw scans in order to artificially replace air with bone marrow:

$$v = \frac{v_{\text{raw}} - d_{\text{air}}}{d_{\text{bone}} - d_{\text{air}}}(d_{\text{bone}} - d_{\text{marrow}}) + d_{\text{marrow}}, \qquad (1)$$

where is $v_{\text{raw}}$ the raw HRpQCT volume and $v$ the density-calibrated and with densities of 850 mg/cm$^3$ for bone ($d_{\text{bone}}$) and 0 mg/cm$^3$ for marrow ($d_{\text{marrow}}$). The apparent air density $d_{\text{air}} \approx -336.57$ mg/cm$^3$ was

computed from calibration values stored in the header of the Scanco aim-files. For training the GAN, patches were linearly normalized from [−350, 1100] mg/cm$^3$ to [0,1]. Only 0.001% of the entire voxels were thereby hard-clamped to 0 or 1. Accordingly, the patches generated by the GAN were again remapped from [0,1] to [−350, 1100] mg/cm$^3$ in production-mode.

## 2.2 | Generation of synthetic bone samples

GANs are trained to stochastically generate samples close to a distribution represented by the training set.[18] Despite the high success of these networks, training is still a very unstable process. Therefore, several algorithms have emerged to deal with training issues. The selection process of algorithms for the proposed context has been discussed previously.[19] We based our framework on Wasserstein GANs, consisting of a generator network $G : \boldsymbol{z} \mapsto \tilde{v}$, and a discriminator network, called *critic* in the context of WGANs, $C : v \mapsto \delta \in \mathbb{R}^+$ that returns a score of real appearance, used for training.[20] Both networks were simultaneously trained to try to fool each other: while $G$ learned to generate synthetic samples $\tilde{v}$ from an n-dimensional vector $\boldsymbol{z}$, the discriminator $C$ learned to distinguish synthetic from real samples. In order to improve convergence, we applied a gradient penalty mechanism (GP) at every iteration performing a critic parameter update (WGAN-GP). Formally:

$$\mathcal{L}_{\text{wgan}} = \mathbb{E}[C(v)] - \mathbb{E}[C(G(\boldsymbol{z}))] + \lambda \mathbb{E}\left[\left(\|\nabla_v C(v)\|_2 - 1\right)^2\right], \tag{2}$$

where $\mathbb{E}$ is the expectation value, $\lambda$ is the gradient penalty coefficient enforcing the Lipschitz constraint, $v$ is a real bone sample and $\boldsymbol{z}$ a latent code, sampled from the surface of an 32-dimensional hypersphere with unit radius.[21] Despite the aforementioned precautions to improve the training process, GANs are still prone to mode collapse. We therefore alleviate this problem with a mode seeking regularization term[22]:

$$\mathcal{L}_{\text{ms}} = \frac{\|G(\boldsymbol{z}_1) - G(\boldsymbol{z}_2)\|_2^2}{\arccos(\boldsymbol{z}_1 \cdot \boldsymbol{z}_2) + \epsilon}, \tag{3}$$

which minimizes the Euclidean distance between two samples $G(\boldsymbol{z}_1)$ and $G(\boldsymbol{z}_2)$ with respect to the angle between the latent codes $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$. The intuition behind this term is that the closer the latent codes, the greater the probability to collapse to the same mode. Minimizing Equation (3) encourages the generator to explore the image space, while forcing the critic to pay attention to minor mode samples. Finally, our training objective function optimizes:

$$\mathcal{L}_{\text{bone}} = \mathcal{L}_{\text{wgan}} + \gamma \mathcal{L}_{\text{ms}} \tag{4}$$

**TABLE 1** Generator network layers of G(z)

| | Kernel size | Activation | Output shape |
|---|---|---|---|
| **z** | – | – | 32 |
| 2×Conv3D | 4×4×4 | Leaky ReLU | 32×4×4×4 |
| 2×Conv3D | 3×3×3 | Leaky ReLU | 32×8×8×8 |
| 2×Conv3D | 3×3×3 | Leaky ReLU | 32×16×16×16 |
| 2×Conv3D | 3×3×3 | Leaky ReLU | 32×32×32×32 |
| 2×Conv3D | 1×1×1 | Leaky ReLU | 1×32×32×32 |
| tanh | – | – | 1×32×32×32 |

**TABLE 2** Critic network layers of C(v) used for training the generative adversarial network

| | Kernel size | Activation | Output shape |
|---|---|---|---|
| Conv3D | 1×1×1 | Leaky ReLU | 32×32×32×32 |
| 2×Conv3D | 3×3×3 | Leaky ReLU | 32×32×32×32 |
| 2×Conv3D | 3×3×3 | Leaky ReLU | 32×16×16×16 |
| 2×Conv3D | 3×3×3 | Leaky ReLU | 32×8×8×8 |
| Conv3D | 3×3×3 | Leaky ReLU | 32×4×4×4 |
| Conv3D | 1×1×1 | – | 1×1×1×1 |

The generator and critic are convolutional networks. We used 3D convolutional layers, Leaky ReLUs and pixel normalization after each 3D convolutional layer in the generator network (Tables 1 and 2). In total, each network was composed of ˜260,000 trainable parameters. We progressively trained the network for 150 epochs in four stages: 4 × 4 × 4, 8 × 8 × 8, 16 × 16 × 16 and 32 × 32 × 32. Each training stage consisted on five training epochs followed by five blending epochs to fade smoothly in the new layers. We used a batch size of 4 in all stages and minimized Equation (4) with $\gamma = 1.0$ using ADAM optimizer[23] ($\beta_1 = 0, \beta_2 = 0.99$) with learning rate = 0.001 for both, the generator and critic network. The gradient penalty coefficient was set to $\lambda = 10$. In order to further regularize the critic, we penalized outputs drifting away from zero by adding the average of the critic output squared to its loss with weight $\epsilon_{drift} = 0.001$. Additionally, we updated the generator-weights using Exponential Moving Average.[24] Training time was 14 h on a NVIDIA 3090 RTX GPU.

## 2.3 | Microstructural fine-tuning of synthetic bone samples

Microstructural fine-tuning was performed by combining the loss metrics of the visual appearance (gestalt-loss)

and 10 microstructural parameters (microstructural loss) which implement well-understood morphological and biological properties of trabecular bone. The gestalt loss $\mathcal{L}_{\text{gestalt}}$ was based on the structural similarity[25] but without luminance term:

$$\mathcal{L}_{\text{gestalt}}(v_1, v_2) = 1 - \underbrace{\mathbb{E}\left[\frac{2\sigma_1\sigma_2 + c_1}{\sigma_1^2 + \sigma_2^2 + c_1}\right]^{\alpha}}_{\text{Contrast term}} \cdot \underbrace{\mathbb{E}\left[\frac{\sigma_{12} + c_2}{\sigma_1\sigma_2 + c_2}\right]^{\beta}}_{\text{Structure term}} \tag{5}$$

with $v_1$ and $v_2$ two volumes to compare, $\mathbb{E}$ the expectation value, $\sigma_i^2$ the local variance of $v_i$, $\sigma_{12}$ the local covariance of both volumes and $\alpha = 1$ and $\beta = 4$ constants. We avoided the luminance-term of the structural similarity, since this property was directly adjusted by the bone mineral density (BMD). The variance and covariance terms were computed with a normalized Gaussian convolution mask of $7 \times 7 \times 7$ voxels.

The latent code $\boldsymbol{z}$ was sampled from the surface of a 32-dimensional hypersphere. By implementing this constraint, optimization was effectively performed in 31 dimensions on spherical coordinates of the hypersurface. The combined fine-tuning loss $\mathcal{L}_{\text{tuning}}$ was optimized with the limited-memory Broyden-Fletcher-Goldfarb-Shanno algorithm (L-BFGS), it reads formally:

$$\mathcal{L}_{\text{tuning}} = \gamma \underbrace{\mathcal{L}_{\text{gestalt}}(v, G(\boldsymbol{z}))}_{\text{Gestalt loss}} + (1-\gamma) \underbrace{\|\boldsymbol{s} - S(G(\boldsymbol{z}))\|_2^2}_{\text{Microstructural loss}}, \tag{6}$$

with $v$ a synthetic reference bone sample, $G(\boldsymbol{z})$ the generator network of the GAN, $\boldsymbol{s}$ the reference features, $S(G(\boldsymbol{z}))$ the function that computes microstructural features from a given latent code and $\gamma = 0.1$ a weighting constant. The design of the function $S(v)$ that computes microstructural features from a synthetic or real bone sample is described in the following sections.

## 2.4 | Differential microstructural parameters

Ten differentiable microstructural parameters were developed for the computation on a Graphics Processing Unit using the Pytorch package of the Python programming language. In contrast to conventional microstructural parameters,[26,27] our parameters are fuzzy in nature, very fast and allow backpropagation as required for microstructural fine-tuning and effective training of neural networks. Besides BMD and the standard deviation of the bone mineral density (BMD.SD), we used further differentiable parameters that can be divided in two groups: voxel-wise and convolution-based parameters.

### 2.4.1 | Voxel-wise parameters

We used the sigmoid and softplus function for the translation of microstructural parameters from crisp to fuzzy domain. The bone volume by total volume (BV/TV) was computed as the average partial bone volume per voxel obtained from a sigmoidal mapping:

$$\text{BV/TV} = \mathbb{E}\left[\left(1 + \exp\left(\frac{t - v}{\sigma}\right)\right)^{-1}\right] = \mathbb{E}[v_b] \tag{7}$$

with $v$ the input volume, $v_b$ the fuzzy binarized volume, threshold $t = 200$ mg/cm$^3$ and smoothness $\sigma = 45$ mg/cm$^3$. With those values, a voxel with integral density of 200 mg/cm$^3$ obtained a BV/TV of 50%, a voxel with 250 mg/cm$^3$ a BV/TV of 75% and a voxel with 406 mg/cm$^3$ a BV/TV of 99%. In contrast, the standard crisp definition would assign a local BV/TV of 100% to any density above 200 mg/cm$^3$. The marrow volume by total volume (MV/TV) reads accordingly

$$\text{MV/TV} = 1 - \text{BV/TV} = \mathbb{E}[1 - v_b]. \tag{8}$$

The softplus function $(x)_\epsilon^+$ is a fuzzy variant of $\max(x, 0)$:

$$(x)_\epsilon^+ = \text{softplus}(x, \epsilon) = \epsilon \ln\left(1 + \exp\left(\frac{x}{\epsilon}\right)\right), \tag{9}$$

with $\max(x, 0) \le (x)_\epsilon^+ \le \max(x, 0) + \ln(2)\epsilon$, for any $x$. We used the softplus function to define the fuzzy tissue mineral density (TMD):

$$\text{TMD} = \mathbb{E}\left[(v)_{\epsilon_1}^+ \, v_b\right] \,/\, (\mathbb{E}[v^\alpha])_{\epsilon_2}^+ \tag{10}$$

with $\epsilon_1 = 5$ mg/cm$^3$ and $\epsilon_2 = 0.015$ small constants to avoid division by zero and $\alpha \in [1, 2]$ ascribing the structural model. For $\alpha = 2$, the model assumes that all voxels contain a partial volume effect, thus a voxel with density 200 mg/cm$^3$ contained equal components of mineralized bone of 400 mg/cm$^3$ and marrow of 0 mg/cm$^3$. For $\alpha = 1$, the model assumes varying mineralization only, a voxel with density 200 mg/cm$^3$ is assigned then a TMD of 200 mg/cm$^3$. We set $\alpha = 1.4$ as a compromise between varying mineralization and partial volume effect. The term $(v)_{\epsilon_1}^+$ restricts the computation of the TMD to those voxels with a positive density. The mineral equivalent marrow density[28] (MeMD) reads

$$\text{MeMD} = 2t - \mathbb{E}\left[(2t - v)_{\epsilon_1}^+ \cdot (1 - v_b)\right] \,/\, (\mathbb{E}[(1 - v_b)^\alpha])_{\epsilon_2}^+ \tag{11}$$

with $\alpha, \epsilon_1$ and $\epsilon_2$ as defined above and $2t = 400$ mg/cm$^3$ the maximum MeMD.

### 2.4.2 | Convolution-based parameters

For the computation of the bone surface by total volume (BS/TV), trabecular number (Tb.N), separation (Tb.Sp) and thickness (Tb.Th) we combined a fuzzy version of the generalized mean intercept length tensor[29] and derived the remaining parameters with the parallel plate model.[30] The generalized mean intercept length tensor uses three directional gradient filters, and is mathematically equivalent to the directed secant method.[31] The bone surface is the area between marrow and bone. It can be computed from the fuzzy binarized bone volume ($v_b$) and three directed convolutions with finite differential kernels $K_x, K_y$ and $K_z$ of size $2 \times 2 \times 2$ voxels and four directional weights of 0.25 and $-0.25$ aligned in x-, y- and z-direction. The BS/TV reads

$$\mathrm{BS/TV} = a\,\Delta^{-1}\,\mathbb{E}\left[\sqrt{\sum_{\xi\in\{x,y,z\}}(v_b * K_\xi)^2}\right]. \tag{12}$$

with $a = 0.945$ an adjustment factor, $\Delta = 0.164$ mm the voxel spacing and $*$ the convolution operator. The remaining parameters read:

$$\mathrm{Tb.N} = (\mathrm{BS/TV}) \,/\, (2\,\mathrm{BV/TV})^{+}_{2\,\epsilon_2} \tag{13}$$

$$\mathrm{Tb.Th} = 2\,(\mathrm{BV/TV}) \,/\, (\mathrm{BS/TV})^{+}_{\epsilon_3} \tag{14}$$

$$\mathrm{Tb.Sp} = 2\,(\mathrm{MV/TV}) \,/\, (\mathrm{BS/TV})^{+}_{\epsilon_3} \tag{15}$$

with $\epsilon_2 = 0.015$ and $\epsilon_3 = 0.1$. Table 3 shows the robust minima and maxima (1%, 99% quantile) and the three center quartiles (25%, 50% and 75%) of the implemented differentiable microstructural parameters.

**TABLE 3** Statistical properties of the microstructural parameters computed on 10 795 real (bold) and synthetic patches (italic) of size $32 \times 32 \times 32$ voxels. The robust minimum (r.min) and maximum (r.max) are computed on the first and last percentile

| Percentile | 1 (r.min) | 25 | 50 (median) | 75 | 99 (r.max) |
|---|---|---|---|---|---|
| BMD (mg/cm$^3$) | **59.72** | **96.50** | **117.29** | **141.93** | **302.70** |
| | *58.50* | *98.06* | *115.32* | *137.84* | *258.54* |
| BMD.SD (mg/cm$^3$) | **74.65** | **97.24** | **110.09** | **126.38** | **265.87** |
| | *79.00* | *95.74* | *107.34* | *122.67* | *197.44* |
| TMD (mg/cm$^3$) | **277.14** | **301.85** | **317.17** | **337.59** | **487.91** |
| | *282 .08* | *300.16* | *315.03* | *333.78* | *426.35* |
| MeMD (mg/cm$^3$) | **27.53** | **46.52** | **54.52** | **62.74** | **87.96** |
| | *27.88* | *47.41* | *55.04* | *62.84* | *86.97* |
| BV/TV (%) | **11.05** | **19.06** | **24.27** | **30.46** | **60.52** |
| | *10.83* | *19.47* | *23.67* | *29.31* | *56.11* |
| MV/TV (%) | **39.48** | **69.54** | **75.73** | **80.94** | **88.95** |
| | *43.89* | *70.69* | *76.33* | *80.53* | *89.17* |
| BS/TV (1/mm) | **0.74** | **1.19** | **1.43** | **1.68** | **2.12** |
| | *0.70* | *1.22* | *1.43* | *1.65* | *2.16* |
| Tb.N (1/mm ) | **1.47** | **2.73** | **2.97** | **3.17** | **3.57** |
| | *1.65* | *2.79* | *3.01* | *3.17* | *3.47* |
| Tb.Th (mm) | **0.28** | **0.31** | **0.34** | **0.37** | **0.68** |
| | *0.29* | *0.31* | *0.33* | *0.36* | *0.60* |
| Tb.Sp (mm) | **0.42** | **0.82** | **1.05** | **1.35** | **2.39** |
| | *0.45* | *0.85* | *1.06* | *1.32* | *2.53* |

Abbreviations: BMD, bone mineral density; BMD.SD, standard deviation of the bone mineral density; BS/TV, bone surface by total volume; BV/TV, bone volume by total volume; MeMD, mineral equivalent marrow density; MV/TV, marrow volume by total volume; Tb.N, trabecular number; Tb.Sp, separation; Tb.Th, thickness; TMD, tissue mineral density.

## 2.5 | Normalization of microstructural parameters

To use any combination of microstructural parameters during optimization without favoring anyone in particular, we extracted the relative information of each parameter in relation to its distribution. For that purpose we designed a numerically robust power transform to normalize each parameter ($x$) to become asymptotically standard-normal distributed ($y$), and with additional robustness for values $x$ that do not belong to the original parameter distribution:

$$y = (\mathrm{H}(-\lambda) + \mathrm{sigmoid}(s_1 x + \Delta_1))^{\lambda}\,\mathrm{sign}(s_1\lambda)\,s_2 + \Delta_2 \tag{16}$$

with $x$ the input microstructural parameter, $\lambda \in \mathbb{R}^{\neq 0}$ and $s_1, s_2, \Delta_1$ and $\Delta_2 \in \mathbb{R}$ constants of the transform. The term $\mathrm{sigmoid}(s_1 x + \Delta_1)$ is the standard sigmoid function applied to a linear mapping of $x$, H the Heaviside function and sign the signum function. The constants $s_2$ and $\Delta_2$ normalize the signal to zero mean and unit standard deviation. The constants $\lambda, s_1$ and $\Delta_1$ were obtained by minimizing the logarithm of the Jarque-Bera statistic of $y$, that computes the deviation from normality based on its skewness and kurtosis. For numerical reasons we restricted $\lambda \in [-20, 20]$, $s_1 x \in [-20, 20]$, $\Delta_1 \in [-20.5, 19.5]$.

The individual normalization constants are shown in Table 4. While we used 10 microstructural parameters, their actual intrinsic dimension of information was much lower: some microstructural parameters (and their normalized pendants) were mathematically related, for example, Tb.N, Tb.Th and Tb.Sp were derived from BV/TV and BS/TV, other variables were correlated for physical or physiological reasons such as BMD and BV/TV. Thus, we applied a principal component analysis to identify and reduce the compound structural information to 2 orthogonal eigenvectors that were used for fine-tuning and to check reliability of the trained GAN. Figure 3 shows Pearson's correlations of the normalized microstructural properties and principal components of the real dataset. In particular, BMD, BV/TV, MV/TV and Tb.Sp were highly correlated with

**TABLE 4** Individual normalization constants for the 10 microstructural parameters

| | $\lambda$ | $s_1$ | $\Delta_1$ | $s_2$ | $\Delta_2$ |
|---|---|---|---|---|---|
| BMD | −0.03 | $1.12 \cdot 10^{-2}$ | 1.26 | $2.90 \cdot 10^{3}$ | 2850.90 |
| BMD.SD | 0.21 | $-8.79 \cdot 10^{-2}$ | −19.09 | $9.87 \cdot 10^{2}$ | 2.57 |
| TMD | −13.93 | $9.21 \cdot 10^{-3}$ | −1.30 | $1.90 \cdot 10^{4}$ | 4.00 |
| MeMD | 2.91 | $2.44 \cdot 10^{-2}$ | 0.17 | $1.15 \cdot 10^{1}$ | −6.37 |
| BV/TV | −13.17 | $2.57 \cdot 10^{0}$ | 1.10 | $1.73 \cdot 10^{4}$ | 5.29 |
| MV/TV | −13.18 | $-2.57 \cdot 10^{0}$ | 3.67 | $1.73 \cdot 10^{4}$ | −5.29 |
| BS/TV | −6.31 | $3.27 \cdot 10^{-2}$ | 1.40 | $7.36 \cdot 10^{3}$ | 173.69 |
| Tb.N | 6.56 | $3.26 \cdot 10^{-1}$ | −0.85 | $1.92 \cdot 10^{2}$ | −2.99 |
| Tb.Th | 0.27 | $-4.37 \cdot 10^{1}$ | −12.97 | $3.72 \cdot 10^{3}$ | 2.34 |
| Tb.Sp | −12.78 | $5.13 \cdot 10^{-1}$ | 1.22 | $1.61 \cdot 10^{4}$ | 6.08 |

Abbreviations: BMD, bone mineral density; BMD.SD, standard deviation of the bone mineral density; BS/TV, bone surface by total volume; BV/TV, bone volume by total volume; MeMD, mineral equivalent marrow density; MV/TV, marrow volume by total volume; Tb.N, trabecular number; Tb.Sp, separation; Tb.Th, thickness; TMD, tissue mineral density.

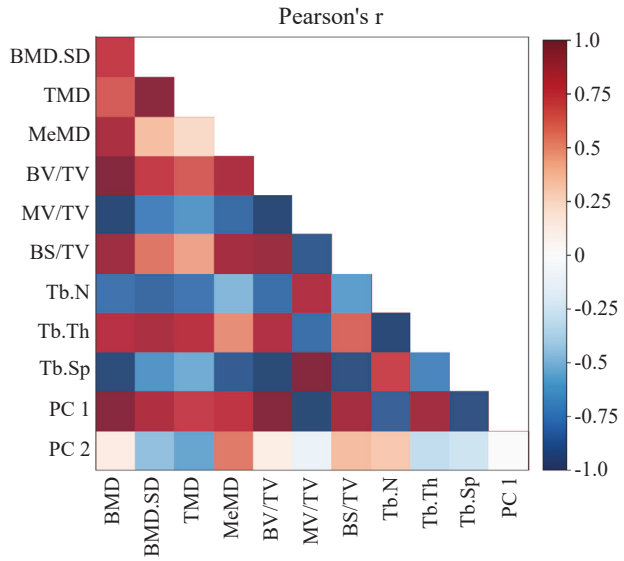


**FIGURE 3** Pearson's correlation coefficients computed on the normalized structural parameters (first 10 rows and columns) and the two principal components. PC 1 is highly correlated with BMD, BV/TV and MV/TV while PC 2 reflects remaining structural information in particular of BMD.SD, TMD and BS/TV. BMD, bone mineral density; BMD.SD, standard deviation of the bone mineral density; BS/TV, bone surface by total volume; BV/TV, bone volume by total volume; MV/TV, marrow volume by total volume; TMD, tissue mineral density.

the first principal component, and BMD.SD, TMD, MeMD and BS/TV with the second principal component.

## 2.6 | Evaluation

Reliability was accessed by analyzing the parameter-distribution of real and synthetic bone samples, and by employing 2-dimensional histograms of the principal components. We used 10,000 real bone patches from the not augmented training set, and 10,000 random synthetic bone samples. Qualitative tests of the GAN were conducted on real and synthetic samples of same microstructural properties. Performance of fine-tuning was demonstrated by varying either principal components or gestalt. Test of the gestalt-variability was performed on a set of samples with minimum structural similarity.

# 3 | RESULTS

Figure 4 shows the normalized density- and cumulative histograms of the 10 microstructural parameters. Synthetic distributions were similar to the real ones on all parameters, and particularly on the principal components. The normalized features of the real and synthesized data showed high accordance to the standard normal distribution. Figure 5 shows the joint-distribution of the two principal components which provides better insight into the differences between real and artificial data compared with the 1D histograms from Figure 4. The two principal components explained 95.73% of the total variance with 81.47% by the first and 14.26% by the second component. The first principal component explained nearly entirely BMD (density) and BV/TV but there was no clear parameter that was purely aligned with the second principal component. However, MeMD, TMD and BMD.SD were those parameters that differed most from BMD and therefore contained the highest amount of not densitometry-related information (Figure 5 right).

Figure 6 shows volume renderings of real and synthesized bone samples generated with a lower and higher threshold of 160 and 240mg/cm$^3$ and after bispline-upsampling to a voxel size of 20.5$\mu$m. Figure 6a shows 25 real bone patches that were selected as those that contained principal components closest to $[-2, -1, 0, 1, 2]$. The principal components of those patches did not deviate more than 0.025 from their nominal value except the patch on the right top with PC = (1.739, 1.631) instead of (2,2) since this value was outside the real distribution. By visually interpreting the figure, we found that the second principal component was related with the porosity and fine-graininess of the volume. Figure 6b shows synthetic bone samples that were obtained by fine-tuning 25 different synthetic bone samples to the same nominal principal components as used for the real bone patches. We stopped optimization at an accuracy of 0.001, except for the upper-right sample that obtained PC = (1.616, 1.315) instead of (2,2), due to the absence in the training set. In Figure 6c we first generated the center

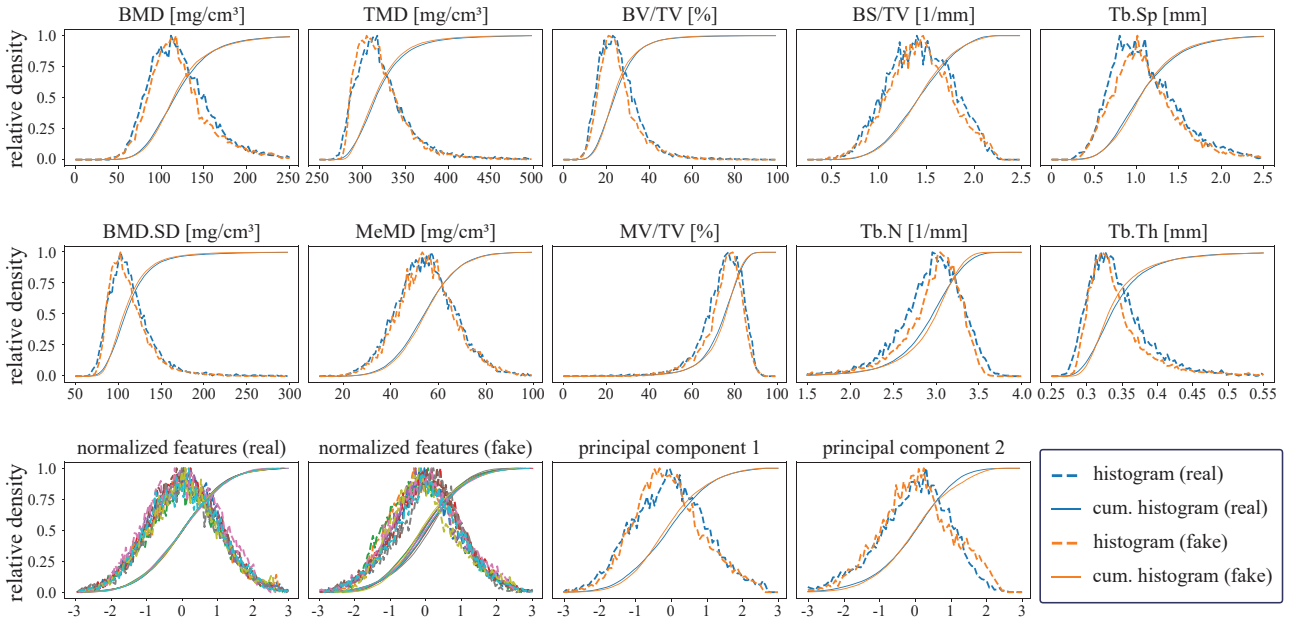


**FIGURE 4** First two rows: Normalized histograms and cumulative histograms of all microstructural parameters. The last row shows an overlay of all normalized parameters (two left boxes) and the principal components (two right boxes). The distributions were generated from 10,000 real and synthesized samples.

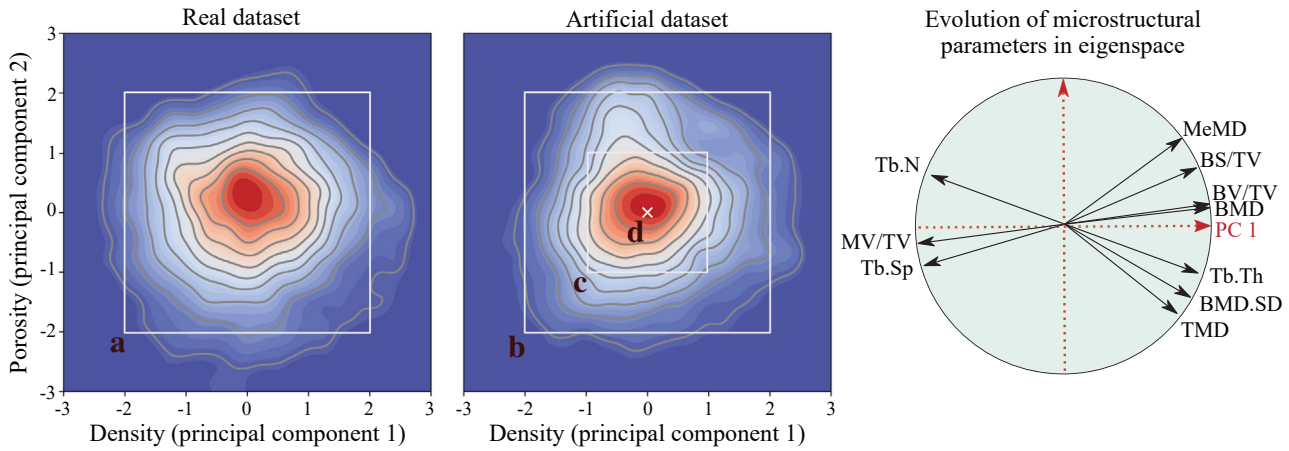


**FIGURE 5** Left: Distribution of the principal components of the real and artificial dataset. The boxes show the sample range of Figure 6a–d. Right: Association of normalized microstructural parameters with the principal components 1 and 2. Principal component 1 reflects mostly BMD (density) while the second principal component reflects properties most associated with TMD and MeMD. BMD, bone mineral density; TMD, tissue mineral density; MeMD, mineral equivalent marrow density.

synthetic bone sample with PC = (0, 0) and then morphed this sample by fixing gestalt but varying the principal components between −1 and 1. The morphed samples represent versions of the very same bone but with osteopenia, in a healthy state and with slightly increased density (PC 1) and meanwhile with varying porosity (PC 2). This is an example to simulate bone deterioration or healing by moving smoothly through the latent space without interrupting gestalt. Figure 6d shows a set of 25 different synthesized bone samples that contain the same microstructural parameters (PC = (0, 0)) with a tolerance of 0.001 but with varying appearance. This demonstrates the inter-variability of the GAN to produce different bone structures with the same structural properties and without collapsing modes.

## 4 | DISCUSSION

The GAN network was able to generate synthetic microstructural bone samples with a good visual

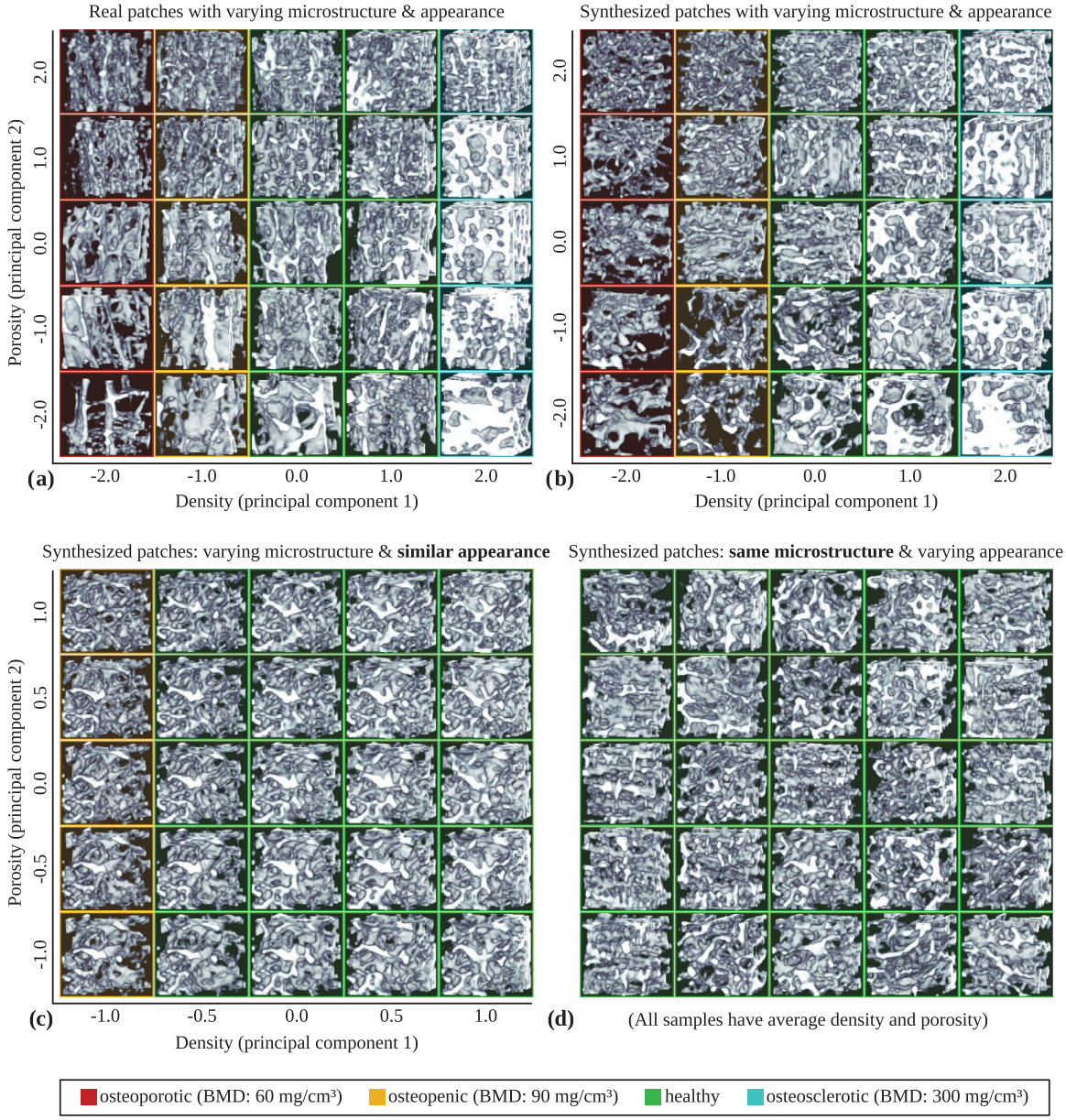


**FIGURE 6** Real and synthesized bone samples: (a) sampled real patches with given principal components between -2 and 2. (b) synthesized bone samples with same principal components as the real patches. (c) Variation of synthesized samples over the range of principal components -1 and 1 of the center seed. This shows the micro-architectural smoothness of the proposed method. (d) Synthesized samples with same microstructural properties but generated from different random seeds. This shows the level of entropy to vary visual appearance. Color-coding depicts the bone density, which is used as an indicator of osteopenia (borderline soft bone), osteoporosis (soft bone) or osteosclerosis (dense bone).

and statistical match compared with real microstructure obtained with HRpQCT. In contrast to existing approaches, our method is the first one that might reach clinical relevance and that is able to model to a certain extent pathological conditions. The size of the patches and voxels were not chosen arbitrarily but presented an optimum choice for the proposed purpose.

The patch size of 5 mm (32 × 32 × 32 voxels) might seem small but is sufficiently large to train local neural networks that deal with the filtering of bone microstructure.[2] Besides being a well manageable size for the given GAN architecture and the chosen hardware with a relative small latent code of 32 dimensions it has an important statistical advantage: the 48 donors were healthy subjects without osteoporotic or osteosclerotic conditions. That means, when extracting big patches of for example, 2 × 2 × 2 cm ($128^3$ voxels), we would not expect to get bone patches that indicate a pathologic condition. In turn, by choosing small patches of 5 mm, the inter-variability of microstructural properties

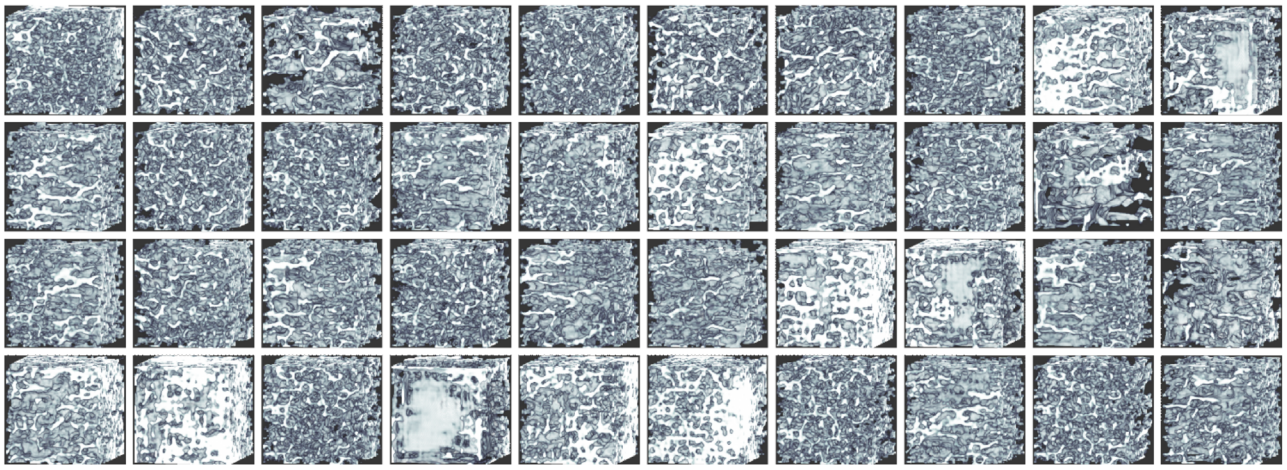

**FIGURE 7** Fourty synthesized patches with size 64 × 64 × 64 voxels that are boxes with edge length of 1 cm and generated by adding an additional stage to the GAN. Patches were generated by maximizing the gestalt-entropy (difference in appearance). GAN, generative adversarial network.

(Figure 4 and 6) increases in such a way that we get not only patches of healthy bone but also those that have microstructural parameters of pathologic conditions, such as BMD smaller 80 mg/cm$^3$ or Tb.Sp larger than 1.5 mm.[32] It is important to note that we assume that an osteoporotic-appearing bone patch could origin indistinguishably from a healthy or osteoporotic patient. The reasoning behind this is the degenerative nature of osteoporosis, that forms a smooth continuum between healthy and pathological conditions, and the assumption that our microstructural parameters are suited to determine those bone pathologies. HRpQCT data was scanned with an isotropic voxel size of 82 $\mu$m and an effective spatial resolution that could resolve a minimum structure of 142.2 $\mu$m.[33] By downsampling the HRpQCT scans to 164 $\mu$m - and surpassing 142.2 $\mu$m - we completely removed blur related to the effective spatial resolution and lowered the noise to a level that practically rendered them perfect artifact-free representations of the bones.

We could also increase the sample size to 64 × 64 × 64 voxels (Figure 7). However, since this implied a reduction of the training set to 3454 samples and an increase of intrinsic patch-complexity by factor 8, the performance was not strong enough to allow microstructural fine-tuning. In any case, the larger samples still serve for the generation of an unlimited dataset that can be used for the development of bone biomarkers, and they also contain more realism than bone samples generated with competing methods.[9,10] The robust generation of even larger samples than 64 × 64 × 64 voxels would require significant more training data, time and computing power. Employing tessellation techniques might be a practical solution to obtain samples of bigger sizes without conducting additional scans, but such a method requires the development of a good modeling of the inter-variability between tiles.[34]

The principal motivation for the implementation of our method was the generation of synthetic ground-truth data as a basis for an in silico simulation of clinical CT scans. Clinical CT scanners still contain significant lower effective spatial resolutions than the one used here.[17] Simulation of clinical CT scans from real or synthetic HRpQCT data can be done by employing an image-deterioration method. This method must have knowledge of the optical transfer function, the noise power spectrum and some specific noise anisotropy patterns, we recently developed a prototype of such a method.[1] It is worth to note that the mentioned approach did not use CT projection sinograms for the clinical CT deterioration but a scheme of noise- and blur-imputation on already reconstructed but very sharp and low-noise HRpQCT baseline data. Similar methods have been used elsewhere for downgrading CT images.[35,36] The pairs of synthetic ground-truth and blurry clinical CT bone samples provide then an infinite synthetic training dataset with feasibly still enough realism to train data-driven methods of clinical applications, such as a convolutional neural network for tissue-specific noise removal.[2] Since the synthetic bone pairs are not expected to be perfectly realistic, the method requires still some additional pairs of real HRpQCT and clinical CT scans to be used for fine-tuning or validation.

Another application is the development and testing of microstructural parameters such as Tb.Sp.[12,29,37] It is known that accuracy of different Tb.Sp - methods depends on the local microstructure which might easily be controlled with the microstructural fine-tuning. Similarly, new microstructural parameters could directly be developed on synthetic bone samples using AI. To study the evolution or degeneration of bone, one might

try to make sense of the topology of the latent space. Briefly, having two synthetic bone samples that are close in latent space and similar in bone gestalt but different on microstructural parameters, we can find a path on the hypersurface in latent space that generates synthetic bone samples without interrupting gestalt. This path represents a possible evolution from the first to the second sample, and if the GAN is adequately well-formed and we find the correct path, this approach might allow to simulate a certain kind of bone degeneration or the effect of drug treatment. A reasonable choice of the path is the shortest or direct one which we evaluated in Figure 6c. For the prototype version of our method we showed an example of possible microstructural changes of Residronate and Teriparatide treatments.[19] However, this method requires further investigation since it is only based on a mathematical model rather than a model based on biological first principles.

Robustness and reliability were provided by training with 10,795 individual bone samples, augmented by factor 16, and by using a new set of smooth derivable microstructural parameters. An ablation analysis showed that the distribution of microstructural parameters became only statistically different when using a fifth or less of the donors (less than 10 donors or 2000 not augmented bone patches). Hence, training data based on 48 donors could be considered stable for our purpose. Compared with a former approach,[19] we achieved in this version a better match of microstructural parameters between real and synthetic bone samples (Figure 4 and 5). In particular the new implementations of Tb.Sp and BS/BV represent an extended quality that allows for the first time a sufficiently fast computation, as required for machine learning. The parameter MeMD was originally intended to differentiate between blood and fat in the bone marrow but our raw data did not contain any bone marrow, thus MeMD served here just as an additional feature to compare real and synthesized samples but without any specific physical meaning. The remaining parameters were not affected by the absence of real bone marrow. Besides the new microstuctural parameters, we also improved the GAN, the normalization technique that is applicable to any microstructural parameter, and defined a new gestalt-loss that optimally entangles microstructure and appearance. It is worth to note that the exact reproduction of a real bone sample was not met by our method. However, this property makes the dataset anonymous and independent from physical data. In future studies, the set of microstructural parameters could be extended by metrics that express the porosity or the plate-to-rod ratio, such as the structural model index or the Hölder exponent[11,38] that might express the second principal component (vertical dimension in Figure 6). Those parameters must be however modified first to become derivable and sufficiently fast for the application in network training.

Synthetic samples generated with a GAN are always limited to the conditions of the training set. For instance, our GAN would not necessarily succeed to generate cervical bone microstructure or samples of the sacrum, since it was trained only on lumbar samples. Following the same principle it would not be able to synthesize degenerated bone with a level of osteoporosis that was not observable in the training set. Thus, including bone patches from osteoporotic patients to the training set would extend the distribution of microstructural parameters and could furthermore improve the synthesis of osteoporotic bone samples.

## 5 | CONCLUSION

We presented a method to generate unlimited realistic training data for 3D bone microstructural analysis. The proposed algorithm can be easily extended to further microstructural parameters. In particular, we were able not only to generate healthy samples but also those with specific pathologies, for example, to mimic to a certain level a cohort with osteoporosis (respecting the limits of the training samples). To achieve our algorithm, we defined a set of new derivable and GPU-ready bone microstructural parameters, a robust normalization method based on the Box-Cox transform, and novel loss terms for the GAN and microstructural fine-tuning. Training was performed on a real bone dataset of 150 human vertebrae, scanned with HRpQCT. We discussed possible applications of the method for the development of AI-based bone texture classification and noise suppression, testing of microstructural parameters and the possible simulation of degeneration or evolution of bone under osteoporosis.

## ACKNOWLEDGMENTS

Felix S. L. Thomsen and Emmanuel Iarussi were supported by grants from Salesforce, USA (Einstein AI) and National Scientific and Technical Research Council, Argentina (PIP 2021). We thank Dr. Tapio Videman for sampling and preparing the human vertebrae.

Open access funding enabled and organized by Projekt DEAL.

## CONFLICT OF INTEREST STATEMENT

The authors have no conflicts to disclose.

## REFERENCES

1. Tomsen F, Borggrefe J. XCT2PCCT: accurate simulation of in-vivo photon-counting computed tomography of the human spongiosa from high resolution peripheral computed tomography. *Wissenschaftstag der Mühlenkreiskliniken*. 2022.
2. Thomsen F, Delrieux C, Pisula J, et al. Noise reduction using novel loss functions to compute tissue mineral density and trabecular bone volume fraction on low resolution QCT. *Comput Med Imaging Graph*. 2020;86:101816.

3. Han L, Lyu Y, Peng C, Zhou SK. GAN-based disentanglement learning for chest X-ray rib suppression. *Med Image Anal*. 2022:102369.
4. Zhao J, Li D, Kassam Z, et al. Tripartite-GAN: synthesizing liver contrast-enhanced MRI to improve tumor detection. *Med Image Anal*. 2020;63:101667.
5. Haubold J, Hosch R, Umutlu L, et al. Contrast agent dose reduction in computed tomography with deep learning using a conditional generative adversarial network. *Eur Radiol*. 2021;31:6087-6095.
6. Ying X, Guo H, Ma K, Wu J, Weng Z, Zheng Y. X2CT-GAN: reconstructing CT from biplanar X-rays with generative adversarial networks. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2019;10619-10628.
7. Anctil-Robitaille B, Desrosiers C, Lombaert H. Manifold-aware CycleGAN for high-resolution structural-to-DTI synthesis. In: *Computational Diffusion MRI*. Springer; 2021:213-224.
8. Ratti E, Graves M. Explainable machine learning practices: opening another black box for reliable medical AI. *AI and Ethics*. 2022;2:1-14.
9. Peña-Solórzano C, Albrecht D, Paganin D, et al. Development of a simple numerical model for trabecular bone structures. *Med Phys*. 2019;46:1766–1776.
10. Zhang H, Yang L, Li C, Wu B, Wang W. Scaffoldgan: synthesis of scaffold materials based on generative adversarial networks. *Comput Aided Des*. 2021;138:103041.
11. Thomsen F, Peña J, Lu Y, et al. A new algorithm for estimating the rod volume fraction and the trabecular thickness from in vivo computed tomography. *Med Phys*. 2016;43:6598-6607.
12. Thomsen F, Delrieux C, Muñoz J. Evaluation of directed secant methods in 2D. In: *2019 XVIII Workshop on Information Processing and Control (RPIC)*. 2019:277-282.
13. Wang T, Lei Y, Fu Y, et al. A review on medical imaging synthesis using deep learning and its clinical applications. *J Appl Clin Med Phys*. 2021;22:11-36.
14. AlAmir M, AlGhamdi M. The Role of Generative Adversarial Network in Medical Image Analysis: An in-depth survey. *ACM Comput Surv*. 2022;55(5), 1–36.
15. Videman T, Nurminen M, Troup J. Lumbar spinal pathology in cadaveric material in relation to history of back pain, occupation, and physical loading. *Spine*. 1990;15:114-119.
16. Wang Y, Battié MC, Boyd SK, Videman T. The osseous endplates in lumbar vertebrae: thickness, bone mineral density and their associations with age and disk degeneration. *Bone*. 2011;48:804-809.
17. Thomsen F, Horstmeier S, Niehoff J, Peña J, Borggrefe J. Effective spatial resolution of Photon Counting CT for imaging of trabecular structures is superior to conventional clinical CT and similar to High Resolution Peripheral CT. *Investig Radiol*. 2022;57:620-626.
18. Goodfellow I, Bengio Y, Courville A. *Deep Learning*. MIT press; 2016.
19. Iarussi E, Thomsen F, Delrieux C. Generative modelling of 3D in-silico spongiosa with controllable micro-structural parameters. In: *Int, Conf on Med. Image Computing and Computer-Assisted Intervention*. Springer; 2020:785-794.
20. Arjovsky M, Chintala S, Bottou L. Wasserstein generative adversarial networks. In *International conference on machine learning*. PMLR; 2017; pp. 214–223.
21. Gulrajani I, Ahmed F, Arjovsky M, Dumoulin V, Courville AC. Improved training of Wasserstein GANs. In: *Advances In Neural Information Processing Systems*. 2017:5767-5777.
22. Mao Q, Lee HY, Tseng HY, Ma S, Yang MH. Mode seeking generative adversarial networks for diverse image synthesis. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019:1429-1437.
23. Kingma D, Ba J. Adam: A method for stochastic optimization. In: *International Conference on Learning Representations*. 2015:1-15.
24. Yazıcı Y, Foo CS, Winkler S, Yap KH, Piliouras G, Chandrasekhar V. The unusual effectiveness of averaging in GAN training. 2018:arXiv:1806.04498. 2018.
25. Wang Z, Bovik AC, Sheikh HR, et al. Image quality assessment: from error visibility to structural similarity. *IEEE Trans Med Imag*. 2004;13:600-612.
26. Graeff C. *Bone Strength Surrogate Markers*. PhD thesis, Technische Universität Hamburg-Harburg; 2010.
27. Thomsen F. *Medical 3D Image Processing Applied to Computed tomography and Magnetic Resonance Imaging*. PhD thesis, Universidad Nacional del Sur, Bahía Blanca, Argentina; 2017.
28. Peña J, Thomsen F, Damm T, et al. Bone-marrow densitometry: assessment of marrow space of human vertebrae by single energy high resolution-quantitative computed tomography. *Med Phys*. 2016;43:4174-4183.
29. Moreno R, Borga M, Smedby Ö. Generalizing the mean intercept length tensor for gray-level images. *Med Phys*. 2012;39:4599.
30. Haidekker MA, Dougherty G. Medical imaging in the diagnosis of osteoporosis and estimation of the individual bone fracture risk. In: *Medical Image Processing*. Springer; 2011:193-225.
31. Whitehouse W. The quantitative morphology of anisotropic trabecular bone. *J Microsc*. 1974;101:153-168.
32. Issever AS, Link TM, Kentenich M, et al. Trabecular bone structure analysis in the osteoporotic spine using a clinical in vivo setup for 64-slice MDCT Imaging: comparison to $\mu$CT imaging and $\mu$FE modeling. *J Bone Miner Res*. 2009;24:1628-1637.
33. Manske SL, Zhu Y, Sandino C, Boyd SK. Human trabecular bone microarchitecture can be assessed independently of density with second generation HR-pQCT. *Bone*. 2015;79:213-221.
34. Frühstück A, Alhashim I, Wonka P. Tilegan: synthesis of large-scale non-homogeneous textures. *ACM Trans Graph*. 2019;38(4):1-11.
35. Divel S, Pelc N. Accurate image domain noise insertion in CT images. *IEEE Trans Med Imaging*. 2019;39:1906-1916.
36. Kim B, Divel SE, Pelc NJ, Baek J. A methodology to train a convolutional neural network-based low-dose CT denoiser with an accurate image domain noise insertion technique. *IEEE Access*. 2022;10:86395-86407.
37. Hildebrand T, Rüegsegger P. A new method for the model-independent assessment of thickness in three-dimensional images. *J Microsc*. 1997;185:67-75.
38. Hildebrand T, Rüegsegger P. Quantification of bone microarchitecture with the structure model index. *Comput Methods Biomech Biomed Eng*. 1997;1:15-23.


**How to cite this article:** Thomsen FSL, Iarussi E, Borggrefe J, Boyd SK, Wang Y, Battié MC. Bone-GAN: Generation of virtual bone microstructure of high resolution peripheral quantitative computed tomography. *Med Phys*. 2023;1-12. https://doi.org/10.1002/mp.16482