Multi-Tube-Voltage vBMD Measurement via Dual-Branch Frequency Balancing and Asymmetric Channel Attention

Mengze Zhang*[1], Yali Li*[2], HuiShu Yuan[2], and Zhen Qian[1]
*Authors contributed equally to this work.

[1] United Imaging Intelligence (Beijing) Co., Ltd.3rd Floor, Building 3, 9 Yongteng North Road,Haidian District,100094 Beijing, China
[2] Department of Radiology and Nuclear Medicine, Peking University Third Hospital, Peking University, 49 Huayuan North Rd., Haidian District, Beijing 100083, China

Abstract. Phantom-less volumetric bone mineral density (vBMD) measurement using computed tomography (CT) presents a cost-effective alternative to conventional phantom-based approaches, yet faces accuracy challenges across varying tube voltages. Current deep learning-based phantom-less solutions frequently overlook the critical role of frequency variance-a crucial factor for precise BMD measurement and cross-voltage generalization. We present a lightweight CT-based phantom-free vBMD measurement framework that addresses critical limitations in cross-voltage generalization. Core innovations include: (1) Frequency-balancing feature modulation with multi-band fusion, preserving spectral measurement cues; (2) A dual-branch architecture combining domain-specific convolutions with cross-frequency interaction; and (3) Asymmetric channel attention, which allocates attention weights based on frequency characteristics, enabling adaptive emphasis on critical low- and high-frequency components. Comprehensive evaluations across 80, 100, and 120 kVp tube voltages demonstrate the proposed method's superior measurement accuracy and reliability, achieving overall mean absolute errors of 5.990 mg/cm^3 and 7.175 mg/cm^3 on internal (1,614 images) and external (2,245 images) testing sets from two centers, respectively. These results suggest that our method offers a promising solution for clinical PL vBMD measurement across varying CT protocols.

Keywords: bone mineral density · quantitative computed tomography.

1 Introduction

Bone mineral density (BMD) is a critical indicator of bone strength, with volumetric BMD (vBMD) being vital for diagnosing osteoporosis. While traditional quantitative computed tomography (QCT) methods rely on expensive external phantoms and frequent recalibrations, phantom-less (PL) methods offer a cost-effective alternative by using internal tissues like muscle and fat as reference points [19]. Early PL approaches derived Hounsfield Unit (HU)-vBMD

conversion equations from reference tissues [18, 15], but the advent of artificial intelligence, particularly deep neural networks (DNNs), has enabled more sophisticated approaches. Some studies have used DNNs for volume-of-interest (VOI) segmentation [14, 1], while others have integrated radiomics and machine learning to enhance feature extraction [9]. However, many of these methods rely on heuristically selected features and complex workflows, limiting their efficiency in vBMD estimation. To address this, end-to-end DNN architectures, such as DenseNet and stacked convolutional layers, have been proposed to streamline vBMD prediction from 2D and 3D images [22, 6].

Despite recent progress, PL methods face significant challenges in clinical applications. Most approaches are optimized for 120 kVp scans, failing to account for the growing adoption of lower tube voltages (80-100 kVp), which reduce radiation exposure to patients [8]. Models trained on 120 kVp data show substantial performance degradation when applied to lower kVp scans due to variations in CT attenuation [20]. This limitation is particularly evident in internal-tissue-based PL methods, where vBMD estimation errors can reach up to 20 mg/cm^3 on non-120 kVp scans [20]. As a result, existing PL approaches struggle to adapt to the diverse imaging protocols used in modern clinical practice.

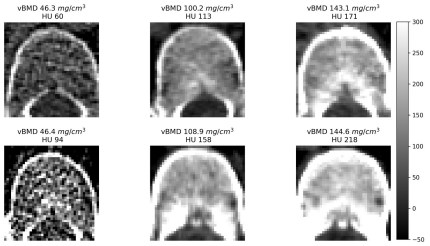

Fig. 1. Intuitive comparison of features in vBMD measurement. The first row shows vertebral bodies with varying bone densities at 120 kVp. The second row shows corresponding vertebral bodies at non-120 kVp, where vBMD texture remains similar, but HU values within the VOI differ significantly. Low vBMD vertebral bodies exhibit both reduced HU values and a sparser trabecular structure in the measurement area.

Beyond variations in scanning parameters such as lower kVp, PL methods also face challenges inherent to osteoporosis itself. The disease progressively reduces bone mineral content, leading to a sparser trabecular architecture [12]. While osteoporotic bone exhibits lower mean CT values (HU) [1], trabecular microarchitecture offers a more precise indicator of bone health [12], especially at lower tube voltages. At these voltages, global intensity measures become less reliable, making the extraction of subtle textural features crucial for accurate vBMD estimation (Fig.1).

Furthermore, existing DNN architectures have inherent limitations. Many models designed for 2D natural images prioritize low-frequency information [17], making them less effective at capturing the high-frequency textural details that

are critical for accurate bone density measurement. When applied to 3D medical images, these 2D-based DNN models struggle with depth-wise feature extraction, fail to leverage volumetric context effectively, and require excessive computational resources when extended to 3D architectures. This limitation is particularly evident under varying scanning parameters, where the ability to integrate both low- and high-frequency features is essential for maintaining accuracy.

To address these challenges, we introduce a novel DNN-based PL framework for vBMD measurement that leverages high-frequency feature extraction to enhance robustness against variations in scanning parameters. The proposed network adapts to diverse frequency domains, ensuring greater accuracy in real-world clinical settings. Our key contributions include:

- A novel network incorporating frequency domain modulation and a dual-branch architecture to capture both overall bone mineral density and fine-grained cancellous trabecular structures for vBMD measurement. The network enhances feature extraction through frequency decomposition, isolating low- and high-frequency components into distinct pathways for a more balanced representation.
- An asymmetric channel attention mechanism for low- and high-frequency modules optimizes attention weight allocation, enhancing the model's ability to focus on the most informative frequency components.
- Quantitative results confirm the superiority of our method on real-world clinical data, demonstrating robustness across varying tube voltages.

2 Methods

2.1 Related Work

Recent advances in computer vision have demonstrated the potential of frequency domain processing. In image restoration, Miao et al. decompose images into low- and high-frequency components using wavelet filtering, restoring each with diffusion models and U-shaped networks [10]. In segmentation tasks, Nam et al. apply convolution kernels of different sizes and fuse the multi-scale features using discrete cosine transforms [11]. Similarly, Chen et al. propose frequency-adaptive dilation convolution, adjusting both frequency and dilation rates at the feature map and convolution kernel levels, enhancing segmentation details [3]. For camouflaged object detection, Zhong et al. decompose features by frequency using offline discrete cosine transforms and then implement feature enhancement and high-order frequency channel selection [25].

These methods, though effective in leveraging high-frequency features, often involve repeated signal processing or additional convolutional layers, increasing complexity and computational cost, especially in 3D imaging. This limitation underscores the need for more efficient DNN designs that can preserve high-frequency features while balancing both low- and high-frequency components during feature extraction and convolution. Building on these insights, we propose a DNN architecture that emphasizes the preservation of high-frequency features while addressing the challenges posed by varying tube voltages (Fig. 2):

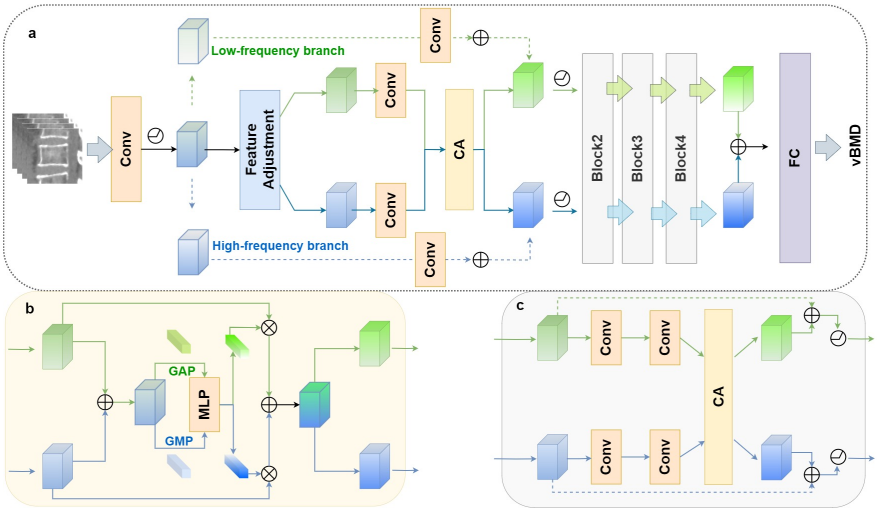

Fig. 2. The proposed network. The proposed network adopts a dual-branch architecture consisting of four distinct modules (a). The first module is responsible for spatial reallocation of feature maps in the frequency domain. The following modules incorporate convolutional layers designed to perform coupling and re-decoupling operations, guided by a channel attention mechanism (b and c). This design facilitates the effective fusion of frequency features, thereby enhancing the model's ability to dynamically process both low- and high-frequency information. *CA, channel attention; FC, fully connected.*

2.2 Decoupling Frequency Features

Generally, low-frequency features are associated with models' generalization, while high-frequency features contribute to improved performance [14, 1, 17]. In the context of vBMD measurement, low- and high-frequency features provide distinct yet complementary perspectives. Low-frequency features are essential for identifying the bone density measurement area and the macroscopic vertebral anatomy, enabling the model to quickly locate the candidate observation region. However, some low-frequency features are highly sensitive to tube voltage variations, which can compromise the model's stability [7]. In contrast, high-frequency features, which capture the trabecular structure of the bone, remain more stable across varying tube voltages, thereby enhancing the model's robustness [12]. Traditional methods for decomposing low- and high-frequency features often rely on computationally intensive techniques, such as wavelet transforms or convolutions with multi-scale kernels [11, 10]. Although effective, these approaches can be resource-intensive, particularly when applied to 3D medical imaging. To address this challenge while preserving effective feature decomposition, we propose a simple yet efficient alternative. Low-frequency components are extracted through average pooling, which downscales the original feature map [24, 4]. High-frequency components are then derived by calculating the residual

between the original feature map and the low-frequency components after performing nearest neighbor upsampling on the low-frequency features. This method not only simplifies the decomposition process but also ensures that the low- and high-frequency features remain complementary, thereby enhancing both computational efficiency and model performance.

2.3 Modulating Frequency Features

To modulate frequency features, existing research often applies the Fourier transform repeatedly within each convolutional module [3], leading to increased computational overhead. To address this limitation, we propose a frequency domain modulation module that integrates the Fourier transform with a spatial attention mechanism in the shallow convolutional layers. These shallow layers, which are critical for capturing local features, are particularly sensitive to high-frequency components [5]. By selectively enhancing high-frequency features, we can improve model performance while avoiding the redundancy of repeated frequency-dependent operations. This approach aligns with the practical requirements of vBMD measurement and leverages the distinct feature extraction characteristics of different layers in DNNs, optimizing both computational efficiency and model performance. Given an input feature map $X \in \mathbb{R}^{C,D,H,W}$, the process is outlined by the following equations:

$$\begin{aligned} Y &= \sum_{b \in B} \sigma(f(X_b; W_b)) \odot X_b \\ with \quad X_b &= \mathcal{F}^{-1}(M_b \odot \mathcal{F}(X)) \end{aligned} \tag{1}$$

Where $\mathcal{F}$ and $\mathcal{F}^{-1}$ are Fourier transform pairs, $M_b \in M_B$ is a binary frequency mask, $f(X; W)$ denotes convolution with parameters W, and $\odot$ represents Hadamard product.

2.4 Double-branch Convolution

The dual-branch convolution structure is a commonly used architecture in the frequency domain [4]. However, traditional dual-branch structures often achieve inter-domain feature interaction through additional convolutions, which can introduce unnecessary parameters and potentially lead to undesirable mixing of inter-domain information. To address this, we have designed a streamlined dual-branch convolution structure based on an asymmetric channel attention mechanism, making the model more lightweight. Furthermore, we ensure that each branch exclusively contains either low- or high-frequency features by effectively coupling and decoupling complementary frequency-domain components.

The convolution process consists of four blocks, each featuring residual connections, convolution layers, and channel attention. Residual connections are crucial for preserving important features and mitigating the vanishing gradient problem. A shallow network design is chosen to capture high-frequency features,

as shallower networks are more effective in this context. In contrast, deeper networks are better suited for capturing low-frequency features and increasing the number of parameters [23, 17]. The low- and high-frequency components $X = \{X_L, X_H\}$ are processed as:

$$\begin{aligned} Y_L &= f(X_L; W_L) + X_L \\ Y_H &= f(X_H; W_H) + X_H \end{aligned} \tag{2}$$

Here, W_L and W_H represents convolution kernels corresponding to X_L and X_H, respectively.

2.5 Feature Fusion and Re-splitting Utilizing Channel Attention

Many existing methods separate low- and high-frequency features, only fusing them at the final stage [10, 5]. However, in vBMD measurement, high-frequency features can benefit from integration with low-frequency features throughout the network. We propose introducing an interaction mechanism between these features at intermediate stages after each dual-branch module rather than only at the final fusion stage. This enables iterative refinement, preserving and enhancing their complementary nature.

The initial fusion of low- and high-frequency features is achieved through summation, given their complementary nature. We then apply global average pooling (GAP) and global max pooling (GMP) to distill channel-specific information, generating attention weights that highlight the most informative features from both domains. Using both GAP and GMP together enhances feature extraction compared to using a single pooling operation [16, 21].

The channel attention maps $A{=}\{A_L, A_H\}$ are generated by applying a shared MLP to the GAP and GMP results as follows:

$$\begin{aligned} \widetilde{X} &= upsample(X_L) + X_H \\ A_H &= \sigma(MLP(GMP(\widetilde{X})) \\ A_L &= \sigma(MLP(GAP(\widetilde{X})) \end{aligned} \tag{3}$$

Finally, we apply these attention weights to each feature map, enhancing the relevant features in both frequency domains. The coupled feature map is then decoupled into its low- and high-frequency components, preserving the distinct contributions of each domain:

$$\begin{aligned} \widetilde{X} &= A_H \odot X_H + A_L \odot X_L \\ Y_L &= AP(\widetilde{X}) \\ Y_H &= \widetilde{X} - upsample(Y_L) \end{aligned} \tag{4}$$

where $AP(x)$ is average pooling with kernel size $2 \times 2 \times 2$ and $upsample(x)$ indicates a nearest neighbor upsampling operation.

This approach ensures that both high- and low-frequency features are balanced and effectively integrated throughout the network, improving the robustness and accuracy of the vBMD measurement.

3 Experimental Results and Discussion

3.1 Datasets and evaluation metrics

Data were sourced from two medical centers, each using independent scanning equipment with standard or bone reconstruction kernels. One center's data (2,245 vertebral body images from 366 patients) served as the external testing set, while the other center's data was split into training, validation, and internal testing sets (1,614 vertebral body images from 248 patients) at a 6:2:2 ratio. All datasets included varying tube voltages and adopted similar reconstruction kernels. In both the internal test set and external test set, the proportion of patients with 120kVp, 100kVp, and 80kVp data is approximately 5:3:2 and 2:1:1, respectively.

Phantom-based QCT measurements at 120 kVp served as the gold standard. The conventional PL method utilizes linear regression at 120 kVp to convert vertebral HU values and internal reference tissues into vBMD. However, since HU values are highly dependent on tube voltage, directly applying this method in our study was not feasible. To address this limitation, we adapted the traditional PL method by incorporating a two-step approach inspired by Nakaura et al. Specifically, non-120 kVp HU values were first transformed into the 120 kVp HU space using a linear conversion formula [7]. This transformation was then followed by linear regression to estimate vBMD. This adaptation enables effective application of the PL method in multi-tube voltage scenarios. Additionally, we compared our method with ResNet-10 and OctResNet-10 architectures [2, 4], evaluating performance using mean absolute error (MAE), mean squared error (MSE), and the coefficient of determination (R^2).

3.2 Performance Comparisons of PL vBMD methods

Table 1: Comparison of Regression Performance Metrics.

kVp	Method	Internal Testing			External Testing		
		MAE ↓	MSE ↓	R^2 ↑	MAE ↓	MSE ↓	R^2 ↑
Overall	**Baseline**	7.327	93.808	0.949	8.332	119.742	**0.972**
	ResNet10	7.330	96.391	0.949	8.510	128.722	0.956
	OctResNet10	6.286	70.683	0.962	7.695	106.182	0.957
	Ours	**5.990**	**64.082**	**0.966**	**7.175**	**97.128**	0.963
120kVp	**Baseline**	6.834	83.241	0.954	7.148	87.486	0.975
	ResNet10	6.941	87.604	0.955	7.426	103.649	0.969
	OctResNet10	5.768	60.405	0.968	6.426	74.492	0.975
	Ours	**5.612**	**56.503**	**0.970**	**6.025**	**70.124**	**0.976**
100kVp	**Baseline**	8.467	116.274	0.941	14.020	274.810	**0.955**
	ResNet10	7.155	92.245	0.954	10.374	180.658	0.926
	OctResNet10	6.787	80.917	0.960	8.021	116.037	0.938

	Ours	**6.459**	**70.190**	**0.966**	**7.812**	**112.699**	0.945
80kVp	**Baseline**	6.639	82.911	0.949	**7.951**	**109.259**	**0.970**
	ResNet10	8.849	131.052	0.935	10.334	165.216	0.942
	OctResNet10	6.904	82.335	0.958	11.040	188.614	0.942
	Ours	**6.240**	**75.430**	**0.960**	9.982	162.702	0.937

We compared our method with the baseline PL vBMD, ResNet-10, and OctResNet-10 models. As shown in Table 1, our method outperformed the others across most metrics, especially at 120 kVp and 100 kVp. While it also surpassed the DNNs at 80 kVp, the DNNs underperformed relative to the baseline on the external 80 kVp dataset, likely due to significant image quality differences across centers under the extremely low tube voltage.

3.3 Ablation experiment

Table 2: Ablation Experiment of Balancing Features and Channel Attention in Frequency Domains.

		Internal Testing			**External Testing**		
		MAE ↓	MSE ↓	R^2 ↑	MAE ↓	MSE ↓	R^2 ↑
Balancing Features	**Channel Attention**						
×	×	6.616	78.980	0.957	8.053	115.373	0.959
✓	×	6.593	79.928	0.957	7.629	111.247	0.958
×	✓	6.502	75.652	0.959	7.783	112.943	0.957
✓	✓	**5.990**	**64.082**	**0.966**	**7.175**	**97.128**	**0.963**
Channel Attention							
(HF + LF) to (GMP + GAP)		6.637	75.897	0.959	7.800	110.783	0.956
HF to GAP + LF to GMP		6.985	85.647	0.953	8.018	118.075	0.957
HF to GMP + LF to GAP		**5.990**	**64.082**	**0.966**	**7.175**	**97.128**	**0.963**

HF, high-frequency; LF, low-frequency;

Ablation experiments demonstrated the synergistic effect of frequency balancing and channel attention. While feature balancing alone showed minimal impact on internal testing, combining it with channel attention produced the best results across all metrics. Furthermore, We evaluated three channel attention configurations for feature integration. Applying GMP to high-frequency features and GAP to low-frequency features yielded optimal results (Table 2), aligning with previous findings which suggest GAP captures low-frequency features and GMP is better suited for high-frequency ones [21, 13, 16].

4 Conclusion

We propose a lightweight frequency-domain network designed to address vBMD measurement across multi-tube voltage CT scans, achieving an optimal balance between precision and parameter efficiency. Precision is enhanced

through the adaptive rebalancing of cross-bandwidth frequency components, an asymmetric channel attention mechanism that allocates different weights based on frequency characteristics, and a double-branch convolutional architecture that preserves critical bone trabecular details while maintaining parameter efficiency. Additionally, the network, structured with shallow layers, one-pass frequency modulation, and double-branch convolution, ensures the retention of trabecular details while keeping the model lightweight. Through rigorous experimentation, we demonstrate that our method facilitates robust vBMD measurement across varying tube voltages. Future work will focus on addressing performance degradation due to image quality variations across different scanners.

5 Limitations

This study has several inherent limitations that should be addressed. First, the research indirectly assesses bone microstructure through the parameter of vBMD. While vBMD serves as a useful surrogate, it does not provide direct measurements of other microstructural indices, such as trabecular spacing, which require specialized scanning equipment and protocols not feasible for clinical populations. Second, the baseline approach employed—linear regression—while common, is susceptible to influences from various technical factors, including equipment variations and tube voltage settings, which may introduce bias or inaccuracies. Third, the scope of this study was limited to tube voltages in the range of 80/100/120 kVp, which are standard in national chest screening programs. This focus precluded an exploration of other voltage ranges, which could provide further insights into the relationship between tube voltage and vBMD measurements. Finally, the study was constrained by a relatively small sample size, limiting the ability to conduct more detailed subgroup analyses that could account for patient demographics and other potential confounding factors.

Despite these limitations, this study represents a pioneering effort in multi-tube voltage BMD measurement. With larger sample sizes, this methodology has significant potential for further optimization and prospective validation across diverse populations and tube voltage datasets.

6 Disclosure of Interests

The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Breit, H.C., Varga-Szemes, A., Schoepf, U.J., Emrich, T., Aldinger, J., Kressig, R.W., Beerli, N., Buser, T.A., Breil, D., Derani, I., et al.: Cnn-based evaluation of bone density improves diagnostic performance to detect osteopenia and osteoporosis in patients with non-contrast chest ct examinations. European Journal of Radiology **161**, 110728 (2023)

2. Cardoso, M.J., Li, W., Brown, R., Ma, N., Kerfoot, E., Wang, Y., Murrey, B., Myronenko, A., Zhao, C., Yang, D., et al.: Monai: An open-source framework for deep learning in healthcare. arXiv e-prints pp. arXiv–2211 (2022)
3. Chen, L., Gu, L., Zheng, D., Fu, Y.: Frequency-adaptive dilated convolution for semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3414–3425 (2024)
4. Chen, Y., Fan, H., Xu, B., Yan, Z., Kalantidis, Y., Rohrbach, M., Yan, S., Feng, J.: Drop an octave: Reducing spatial redundancy in convolutional neural networks with octave convolution. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 3435–3444 (2019)
5. Fang, M., Damer, N., Kirchbuchner, F., Kuijper, A.: Learnable multi-level frequency decomposition and hierarchical attention mechanism for generalized face presentation attack detection. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 3722–3731 (2022)
6. Fang, Y., Li, W., Chen, X., Chen, K., Kang, H., Yu, P., Zhang, R., Liao, J., Hong, G., Li, S.: Opportunistic osteoporosis screening in multi-detector ct images using deep convolutional neural networks. European Radiology **31**, 1831–1842 (2021)
7. Garner, H.W., Paturzo, M.M., Gaudier, G., Pickhardt, P.J., Wessell, D.E.: Variation in attenuation in l1 trabecular bone at different tube voltages: caution is warranted when screening for osteoporosis with the use of opportunistic ct. American Journal of Roentgenology **208**(1), 165–170 (2017)
8. Ha, H.I., Hong, S.S., Kim, M.J., Lee, K.: 100 kvp low-tube voltage abdominal ct in adults: radiation dose reduction and image quality comparison of 120 kvp abdominal ct. Journal of the Korean Society of Radiology **75**(4), 285–295 (2016)
9. Jiang, Y.W., Xu, X.J., Wang, R., Chen, C.M.: Radiomics analysis based on lumbar spine ct to detect osteoporosis. European radiology **32**(11), 8019–8026 (2022)
10. Miao, Y., Deng, J., Han, J.: Waveface: Authentic face restoration with efficient frequency recovery. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6583–6592 (2024)
11. Nam, J.H., Syazwany, N.S., Kim, S.J., Lee, S.C.: Modality-agnostic domain generalizable medical image segmentation by multi-frequency in multi-scale attention. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11480–11491 (2024)
12. Nguyen, T.P., Chae, D.S., Park, S.J., Yoon, J.: A novel approach for evaluating bone mineral density of hips based on sobel gradient-based map of radiographs utilizing convolutional neural network. Computers in Biology and Medicine **132**, 104298 (2021)
13. Nirthika, R., Manivannan, S., Ramanan, A., Wang, R.: Pooling in convolutional neural networks for medical image analysis: a survey and an empirical study. Neural Computing and Applications **34**(7), 5321–5347 (2022)
14. Pickhardt, P., Nguyen, T., Perez, A., Graffy, P., Jang, S., Summers, R., et al.: Improved ct-based osteoporosis assessment with a fully automated deep learning tool. radiol artif intell. 2022; 4 (5): e220042
15. Prado, M., Khosla, S., Chaput, C., Giambini, H.: Opportunistic application of phantom-less calibration methods for fracture risk prediction using qct/fea. European radiology **31**(12), 9428–9435 (2021)
16. Qin, Z., Zhang, P., Wu, F., Li, X.: Fcanet: Frequency channel attention networks. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 783–792 (2021)

17. Wang, H., Wu, X., Huang, Z., Xing, E.P.: High-frequency component helps explain the generalization of convolutional neural networks. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 8684–8694 (2020)
18. Weaver, A.A., Ronning, I.N., Armstrong, W., Miller, A.N., Kiani, B., Martin, R.S., Beavers, K.M., Stitzel, J.D.: Computed tomography assessment of pelvic bone density: Associations with age and pelvic fracture in motor vehicle crashes. Accident Analysis & Prevention **193**, 107291 (2023)
19. Winsor, C., Li, X., Qasim, M., Henak, C., Pickhardt, P., Ploeg, H., Viceconti, M.: Evaluation of patient tissue selection methods for deriving equivalent density calibration for femoral bone quantitative ct analyses. Bone **143**, 115759 (2021)
20. Woisetschläger, M., Hägg, M., Spångeus, A.: Computed tomography-based opportunistic osteoporosis assessment: a comparison of two software applications for lumbar vertebral volumetric bone mineral density measurements. Quantitative Imaging in Medicine and Surgery **11**(4), 1333 (2021)
21. Woo, S., Park, J., Lee, J.Y., Kweon, I.S.: Cbam: Convolutional block attention module. In: Proceedings of the European conference on computer vision (ECCV). pp. 3–19 (2018)
22. Wu, Y., Yang, X., Wang, M., Lian, Y., Hou, P., Chai, X., Dai, Q., Qian, B., Jiang, Y., Gao, J.: Artificial intelligence assisted automatic screening of opportunistic osteoporosis in computed tomography images from different scanners. European Radiology pp. 1–9 (2024)
23. Xu, Z.J., Zhou, H.: Deep frequency principle towards understanding why deeper learning is faster. In: Proceedings of the AAAI conference on artificial intelligence. vol. 35, pp. 10541–10550 (2021)
24. Yi, Z., Tang, Q., Azizi, S., Jang, D., Xu, Z.: Contextual residual aggregation for ultra high-resolution image inpainting. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 7508–7517 (2020)
25. Zhong, Y., Li, B., Tang, L., Kuang, S., Wu, S., Ding, S.: Detecting camouflaged object in frequency domain. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 4504–4513 (2022)