BLS-GAN: A Deep Layer Separation Framework for Eliminating Bone Overlap in Conventional Radiographs

Haolin Wang[1], Yafei Ou[2*], Prasoon Ambalathankandy[3], Gen Ota[4], Pengyu Dai[2], Masayuki Ikebe[4], Kenji Suzuki[2], Tamotsu Kamishima[5]

[1]Graduate School of Health Sciences, Hokkaido University, Sapporo, Japan
[2]Institute of Integrated Research, Institute of Science Tokyo, Yokohama, Japan
[3]Processor Research Team, RIKEN Center for Computational Science, Kobe, Japan
[4]Research Center For Integrated Quantum Electronics, Hokkaido University, Sapporo, Japan
[5]Faculty of Health Sciences, Hokkaido University, Sapporo, Japan

Abstract

Conventional radiography is the widely used imaging technology in diagnosing, monitoring, and prognosticating musculoskeletal (MSK) diseases because of its easy availability, versatility, and cost-effectiveness. Bone overlaps are prevalent in conventional radiographs, and can impede the accurate assessment of bone characteristics by radiologists or algorithms, posing significant challenges to conventional clinical diagnosis and computer-aided diagnosis. This work initiated the study of a challenging scenario - bone layer separation in conventional radiographs, in which separate overlapped bone regions enable the independent assessment of the bone characteristics of each bone layer and lay the groundwork for MSK disease diagnosis and its automation. This work proposed a Bone Layer Separation GAN (BLS-GAN) framework that can produce high-quality bone layer images with reasonable bone characteristics and texture. This framework introduced a reconstructor based on conventional radiography imaging principles, which achieved efficient reconstruction and mitigates the recurrent calculations and training instability issues caused by soft tissue in the overlapped regions. Additionally, pre-training with synthetic images was implemented to enhance the stability of both the training process and the results. The generated images passed the visual Turing test, and improved performance in downstream tasks. This work affirms the feasibility of extracting bone layer images from conventional radiographs, which holds promise for leveraging layer separation technology to facilitate more comprehensive analytical research in MSK diagnosis, monitoring, and prognosis.

Code and Dataset —
https://github.com/pokeblow/BLS-GAN.git

Introduction

Conventional radiography (projection radiography) is a cost-effective and versatile diagnostic technology (Pasveer 1989; Ou et al. 2021), especially for the musculoskeletal (MSK) system (Grant and Wakefield 2018), due to its ability to produce bone images with high resolution and contrast. However, a significant limitation of this modality is the bone overlaps (Newton 2016; Low and Peh 2017), which is prevalent in MSK imaging. The overlaps introduce a textural mixture from both upper and lower tissues, which complicates the accurate localization and analysis of bone lesions, ultimately affecting the clinical diagnosis and management.

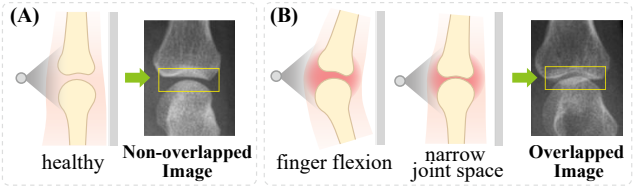

Figure 1: Explanation of bone overlap: Due to finger flexion and the excessively narrowed joint space, the joint can appear as bone overlap. Bone overlap can impede clinical imaging diagnosis and its automatic analysis in MSK diseases. (A) healthy joint without bone overlap, (B) joint with bone overlap.

This work explores rheumatoid arthritis (RA), one of the prevalent MSK diseases. RA is a chronic autoimmune inflammatory disease characterized by joint swelling and tenderness. Physicians typically diagnose, assess prognosis, and monitor RA by observing joint symptoms and imaging features, with joint space narrowing (JSN) and bone erosion in the fingers being crucial indicators of joint destruction (Aletaha and Smolen 2018; Platten et al. 2017). As the disease progresses, RA patients develop limited finger mobility, resulting from JSN, subluxation, and dislocation, which manifests in radiographic images as a transition from a clearly demarcated joint space without overlap to significant bone overlap, especially at the metacarpophalangeal (MCP) joints, as shown in Fig. 1 (A), (B). The texture mixture caused by the overlap poses a challenge to the imaging diagnosis of RA, especially the qualitative diagnosis and monitoring of JSN and bone erosion. This also presents a new challenge for automating qualitative and quantitative analyses in RA, particularly in images with extensive overlap. For instance, bone overlap in automated JSN progression quantification methods not only reduces the accuracy and robustness of registration-based methods (Ou et al.

*Corresponding Author: Yafei Ou (ou.y.ac@m.titech.ac.jp).

Scenario	Networks			Main Output
	G	D-like	R	
Bone Suppression	✓	△	×	Soft tissue w/o bone
Amodal Completion	✓	△	△	Occluded objects
Objects Removal	✓	△	✓	Image w/o objects
Bone Layer Separation	✓	✓	✓	Bone images w/o overlaps

Table 1: Comparison of different scenarios with the proposed bone layer separation scenario. G: Generator; D-like: Discriminator-like; R: Reconstruction. ✓: Mandatory, △: Partially includes, ×: Unnecessary.

2023; Wang et al. 2023) for some finger joint images, but also limits the scalability of these methods to other complex joints, such as the wrist, hip, and knee joints.

In related works, amodal completion with inpainting has been widely employed to reconstruct occluded regions in natural images, achieving notable advances (Ao, Ke, and Ehinger 2023; Zhang et al. 2023; Sargsyan et al. 2023; Ko and Kim 2023; Xu, Zhang, and Shi 2024). Objects removal can successfully remove impurities such as shadows and raindrops by reconstructing background textures (Elad, Kawar, and Vaksman 2023). In chest radiography, rib suppression leveraging deep learning models has effectively removed the ribs (Suzuki et al. 2006; Han et al. 2022), thereby enhancing the visibility of lung soft tissues and improving diagnostic efficiency for lesions. As illustrated in Table 1, there are substantial differences between our challenging scenario and others in terms of network structure design and output. According to the imaging principle of conventional radiography (Bushberg and Boone 2011), the images show the superposition of the X-ray absorption rates by different tissues, leading to significant differences between conventional radiographs and natural images. Although parallels can be drawn in terms of scenario descriptions and application contexts with amodal completion, the lack of a robust reconstruction mechanism results in discrepancies between the generated textures and authentic bone textures. Moreover, the complex texture characteristics of bones separate them from physical artifacts, such as shadows, rendering classical objects removal methods less effective in this challenging scenario. In the context of bone (rib) suppression, existing methods are limited by their limited ability to achieve distinct separation of the bone layer, and fail to produce the desired outcomes.

To address these challenges, we initiate a challenging research scenario - joint bone layer separation, and propose a multi-supervised framework named **B**one **L**ayer **S**eparation **G**enerative **A**dversarial **N**etwork (**BLS-GAN**), which implements extraction of separated bone layer images extraction from a single finger conventional radiographs and eliminates bone overlap in each bone layer image. This generative-based method provides a reliable image base for the independent evaluation of each feature of the bone layer and the study of automated analysis methods. Specifically, our contributions can be summarized as follows:

- **A Challenging Scenario for Amodal Completion**: The imaging principles of conventional radiography inherently result in bone overlap, which poses significant challenges for the clinical diagnosis and analysis of MSK system lesions, and for the development of automated qualitative and quantitative analysis methods. This issue is particularly problematic for diseases such as RA. The presence of overlaps can lead to substantial inaccuracies in the downstream JSN quantification task. Additionally, due to the requirement for strict adherence to the original texture of bones, classical amodal completion with inpainting is not feasible. This inspiration has spurred the exploration of a challenging research scenario.
- **Bone Layer Separation Framework**: This work designed and implemented a novel framework for the above challenging scenario. Compared with other methods, our framework offers the following two innovations. (1) introduced a radiography imaging principles-based reconstructor that leverages conventional radiography principles and includes a correction parameter to rectify overlapped regions in the reconstruction. (2) integrated a segmentation-based multi-channel supervisor network to distinguish between overlapped and non-overlapped regions, enhancing the authenticity and natural appearance of bone textures in the generated images.
- **Expert Assessments and Clinical Downstream Validation**: This work successfully passed the radiological technologist visual Turing test and can significantly enhance both the accuracy and stability in the clinical downstream task of JSN quantification.
- **Dataset for this Challenging Scenario**: We provide a dataset specifically designed for this challenging scenario, utilized in this paper. The dataset includes joint images and mask annotations for upper and lower bones.

Methodology

Conventional hand radiographs of RA patients often suffer from bone overlap in finger joints due to disease or positioning, leading to a mixture of texture information between the upper and lower bones. This poses several clinical and technical challenges. Thus, we explored a challenging research scenario: using conventional finger joint radiography as input, generating independent layer images of the upper and lower bone without overlap as output, as illustrated in Fig. 2. We define this process as *Bone Layer Separation*.

Bone Layer Separation Framework

This work proposed a GAN-based bone layer separation framework for finger joint radiographs to extract layer images and eliminate bone overlap in each layer. As shown in Fig. 2, the framework consists of three basic sub-networks: the layer image generator, the segmentation-based multi-channel supervisor, the correction parameter regression reconstructor, and synthetic images pre-training. We define that bone layer separation of the original image, layer images $1, ..., i, ..., n$ are generated. This study utilizes MCP joint images, therefore, n is set to 2.

Layer Image Generator In the original image, we defined and partitioned the upper and lower bones of the joint

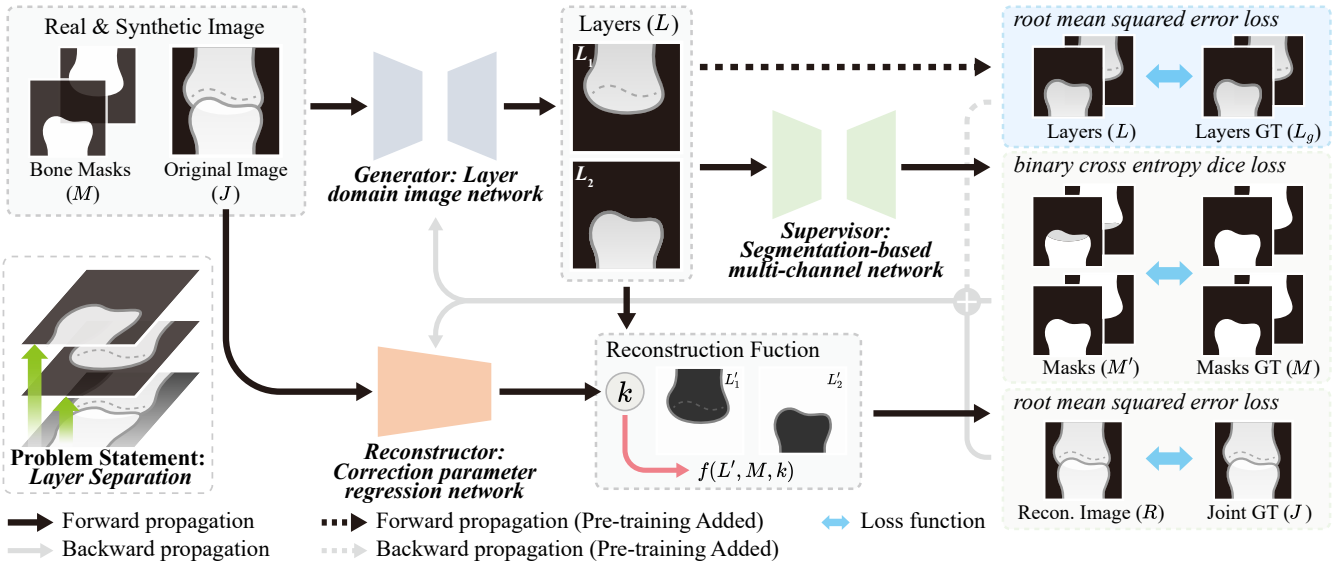

Figure 2: Explanation of bone layer separation: Layer-by-layer extraction of the upper and lower bones, followed by eliminating overlapped regions. The framework consists of three primary components: a generator, a supervisor, and a reconstructor. This process is performed as follows: (i) The generator produces bone layer images from the original joint image and bone masks. (ii) The layer images are discriminated by the supervisor and reconstructed through a reconstructor, yielding a discrimination mask and a reconstructed image. (iii) Discrepancies between the masks and ground truth (GT), and between the reconstructed and original image, are used to create a hybrid loss function that guides the generator and reconstructor during backpropagation. (iv) In the training pipeline, pre-training is performed in synthetic images, and the discrepancies of the real bone layer images are incorporated into the original loss function to facilitate the establishment of the initial model. Subsequently, the training is performed in real and synthetic images.

as two independent layers. However, in the presence of bone overlap, each layer contains regions that intersect with other bone layers. The goal of the generator is to eliminate these overlapped regions in each bone layer, achieving the separation defined in our study. The backbone network employed in this context can be any generation network.

The input is the original joint image J and its corresponding masks delineating the upper and lower bones of the joint, denoted as $M = \{M_1, ..., M_n\}$. The output of the generator network is the layer images of the upper and lower bones in the joint with masks, defined as $L = \{L_1, ..., L_n\}$. Assuming that the layer image generator is denoted as $\mathcal{G}$, the generation process can be defined as $L = \mathcal{G}(J) \cdot M$.

Segmentation-based Multi-channel Supervisor Unlike traditional discriminators in GANs, we integrated a segmentation-based multi-channel network to achieve pixel-level discrimination of layer images, named supervisor, which outputs two sets of four-channel masks. One set pertains to the segmentation of overlapped and non-overlapped regions, and the other set assesses the authenticity of the generated images. Therefore, our supervisor effectively identifies overlapped regions in images, thereby enhancing generator supervision. The backbone network employed in this context can be any segmentation network.

The layer images from generator L serve as the input to the supervisor. The output of the supervisor is defined as $M' = \{M'_1, ..., M'_n\}$. Suppose the supervisor network is denoted as $\mathcal{D}$. Therefore, the discrimination process can be defined as $M' = \mathcal{D}(L) \cdot M$, where M' represents the discrimination mask from the supervisor.

Radiography Imaging Principles based Reconstructor According to the principles of conventional radiography, different tissues exhibit varying absorption rates. Tissues with higher density demonstrate greater absorption, while those with lower density exhibit weaker absorption, resulting in radiographic representations (Bushberg and Boone 2011; Huda and Abrahams 2015). In the presence of overlap, the X-ray absorption of the upper layers of tissues influences the imaging of overlapped tissues, showing exponential decay.

In contrast to amodal completion in natural images, adherence to the imaging principles of radiography is crucial. Therefore, we introduce a reconstructor to thoroughly supervise layer image generation. In this reconstructor, the absorption rate image of the bone is calculated based on the layer image, and reconstructed based on the reconstruction function defined below. Our framework delineates the layer image of bones as a composite of bone texture and soft tissue texture. However, this amalgamation leads to the recurrent calculation of soft tissue texture within overlapped regions, thereby compromising reconstruction quality. To address this issue, we introduce a single correction parameter regression network using the VGG-18 network (Simonyan 2014), named correction parameter regression network, to derive a correction parameter to mitigate the impact of re-

dundant soft tissue calculations within overlapped regions.

The algorithm flow proceeds as follows: suppose $\mathcal{R}$ denoted as the reconstructor, with the original joint image as input and a single parameter k as output, which can be defined as $k = \mathcal{R}(I)$. Subsequently, the image is reconstructed according to the reconstruction function $f(L, M, k)$, as delineated in Eq. 1, where R denotes the reconstructed image and $M_{\cup} = \bigcup_{i=1}^{n} M_i$. In the mask regions corresponding to the upper and lower bones of the generated layer image, divide by the correction parameter k, followed by multiplying the layer superposition results in the mask regions by k.

$$R = f(L, M, k) = \left(1 - k \prod_{i=1}^{n} (1 - \left(1 - \frac{(1 - L_i)}{k}\right) \cdot M_i)\right) \cdot M_{\cup} \quad (1)$$

Framework Flowchart The image size processed by our framework is set to 256 $\times$ 256. We construct the loss function based on binary cross entropy (BCE) dice loss $\mathcal{L}_b(\hat{y}, y)$ (Yeung et al. 2022) and root mean squared error (RMSE) loss $\mathcal{L}_r(\hat{y}, y)$ (Chai and Draxler 2014), where y represents the GT, and $\hat{y}$ represents the predicted value.

For the supervision of the generator and reconstructor, we employ the loss functions $\mathcal{L}_b$ and $\mathcal{L}_r$. Furthermore, we incorporate loss supervision additionally for the overlapped regions. Thus, the loss function of the networks can be defined as Eq. 2, where α_0 and β_0 are set to 0.5. As an additional weight of the overlapped regions, we introduced $M_{\cap} = \bigcap_{i=1}^{n} M_i$, $J_{\cap} = J \cdot M_{\cap}$, $R_{\cap} = R \cdot M_{\cap}$.

$$\mathcal{L} = \alpha_0 \mathcal{L}_b(M', M) + \beta_0(\mathcal{L}_r(R, J) + \mathcal{L}_r(R_{\cap}, J_{\cap})) \quad (2)$$

In addition, we train the supervisor simultaneously and independently. Regarding the input for the supervisor, the real sample comprises the original image with the mask, denoted as $J_r = J \cdot M$ where J represents the original joint image and M denotes the masks. The GT of real samples $M_r = \{\{M_1 - M_{\cap}, ..., M_n - M_{\cap}\}, \{M_1, ..., M_n\}\}$ is derived by eliminating the masked regions. Conversely, the fake sample consists of the layer image generated by the generator, expressed $J_f = \mathcal{G}(J, M)$. The GT for fake samples $M_f = \{\{M_1, ..., M_n\}, \{\mathbf{0}_1, ..., \mathbf{0}_n\}\}$, where $\mathbf{0}$ represents an all-zero matrix. We performed $\mathcal{L}_0$ loss for supervisor supervision. Thus, the loss function can be defined in Eq. 3, where α_1 and β_1 are set to 0.5.

$$\mathcal{L}_S = \alpha_1 \mathcal{L}_b(\mathcal{D}(J_r), M_r) + \beta_1 \mathcal{L}_b(\mathcal{D}(J_f), M_f) \quad (3)$$

Synthetic Images Pre-training We constructed synthetic images with overlap based on images without overlap. Specifically, we utilize the non-overlap real image as the foundation and randomly shift the upper and lower articular bones to create an overlapped region. Reconstruction is subsequently performed based on the upper and lower bones using reconstruction function Eq. 1 to generate the overlapped region. Regarding correction parameter k in synthetic images, we utilized a mathematical method for its determination. This process involves initially excluding the bone region from the image, subsequently solving the Laplace equation (Gong 2020), and ultimately calculating the mean value within the overlapping region.

We process pre-training with synthetic images S, specifically, since synthetic images are generated from non-overlap images, upper and lower bone GT L_g can be effectively obtained. Therefore, we introduce L_g into the loss function as defined in Eq. 4 and 5, to build the initial framework.

We performed pre-training of our framework with synthetic images and their corresponding masks to establish its foundational functionality. Specifically, since the synthetic images are synthesized from non-overlap images, we can effectively obtain the upper and lower bone GT (bone without overlap) L_g and its corresponding mask M. Consequently, L_g is incorporated into Eq. 4 as the GT for loss function calculation. Additionally, L_g is introduced as the additional real sample of the supervisor in training, thus the loss function is presented in Eq. 5, where $M_g = \{M, M\}$.

$$\mathcal{L}_p = \mathcal{L} + \mathcal{L}_r(L, L_g) \quad (4)$$

$$\mathcal{L}_{Sp} = \mathcal{L}_S + \mathcal{L}_b(\mathcal{D}(L_g), M_g) \quad (5)$$

In the main training, due to the absence of GT for the upper and lower bone in real images, we continue to apply the original loss function Eq. 2 and Eq. 3, and framework trained employed both synthetic and real images.

Implementation

The networks were implemented on a workstation with three GPUs (NVIDIA GeForce GTX 2080 Ti). The supervisor, generator, and reconstructor networks were trained using the Adam optimizer with an initial learning rate of $1e^{-5}$. In our practice, we commence by performing pre-training on synthetic images and their corresponding GT images, extending this preparatory phase across 250 epochs with a consistently maintained batch size of 12. Subsequently, we refine the loss function and GT, maintaining the same batch size, for an additional 50 epochs with both synthetic and real images to meticulously optimize the performance of our framework.

Experiments

To validate the robustness and reliability of the framework, we designed and conducted experiments to rigorously assess the fidelity of the generated layer images. We concentrated primarily on the MCP joint due to its critical significance in disease diagnosis and its higher susceptibility to overlap in practice.

We evaluated the reconstructed images with real overlap images in these four metrics: the mean squared error (MSE), the structural similarity (SSIM), the peak signal-to-noise ratio (PSNR), and the fréchet inception distance (FID).

Dataset

This study was reviewed and approved by the Ethics Committee of the Hokkaido University Hospital (022-0336) and in accordance with the Declaration of Helsinki. The dataset utilized in this study comprises 168 posteroanterior (PA) radiographs of the hand sourced from 43 patients with RA. Of these patients, 88.5% are female. The average age in the dataset is 65.6 years, with a variance of 12.87 and an age range of 31-91 years. These images were prepared from the Faculty of Health Sciences, Hokkaido University, which employs its proprietary conventional radiography system and adheres to the Digital Imaging and Communications in

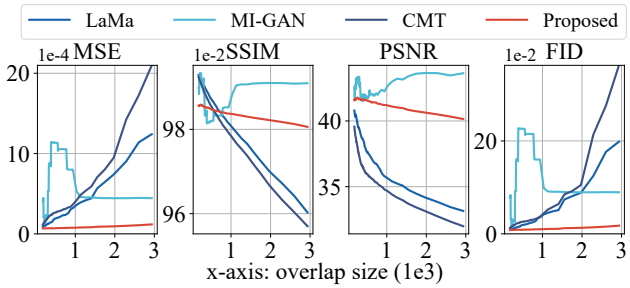

Figure 3: Comparison of the proposed framework with other methods in different metrics across overlap sizes.

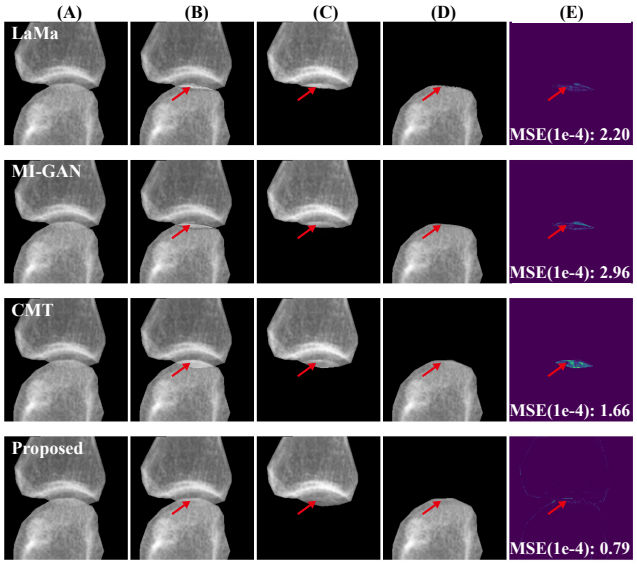

Figure 4: Comparison of the proposed framework with other methods. (A) Real Joint image; (B) Reconstructed Joint Image; (C) Upper Bone Layer; (D) Lower Bone Layer; (E) MSE Spectrum (A v.s. B).

Medicine (DICOM) standard for dataset management. Digital radiographs were acquired with the CALNEO smart C47 (Fujifilm, Tokyo, Japan) under the following conditions: tube voltage of 50 kV, tube current of 100 mA, exposure time of 0.02 milliseconds, source-to-image distance of 100 cm, resolution of 0.15 mm/pixel, image size of 1670×2010 pixels, and a bit depth of 16 bits. Given that MCP joints are more susceptible to bone overlap than other joints, our dataset exclusively utilized MCP joint images. These images were manually screened to remove those exhibiting severe bone erosion, resulting in 1,594 joint images, which included 672 overlapping images, with an average overlap size of 686.05 ± 1983.12 pixels. The dataset contains 430 MCP joints, which are divided into training and test sets by joint at a ratio of 3:1. An experienced radiological laboratory assistant annotated these images to label the upper and lower bones of the joint into two channels, further reviewed by radiologists.

Method	MSE (10^{-4})	SSIM (10^{-2})	PSNR	FID (10^{-2})
LaMa	5.47±16.15	97.98±3.29	37.45±4.77	8.72±33.55
MIN-GAN	5.63±23.72	**98.82±3.05**	**42.63±6.70**	10.70±50.78
CMT	9.26±32.27	97.83±3.28	35.94±4.83	15.83±70.78
Proposed	**0.88±0.70**	98.37±0.51	41.09±1.72	**1.20±1.33**

Table 2: Evaluation result of proposed framework and comparison with other methods in different metrics. Expressed as mean ± standard deviation.

Generate Image Evaluation

We conducted an experiment to evaluate the performance of our framework and compare it with other amodal completion methods with inpainting. The evaluation was performed on real images with overlaps. For the amodal completion network with inpainting, we implemented the model in (Sargsyan et al. 2023) and (Ko and Kim 2023). Utilizing the pre-training parameters, we subsequently trained on our dataset. The network independently predicts the upper and lower bones and reconstructed using the reconstruction function of synthetic images.

As shown in Table 2, our framework exhibit great performance in all four evaluation metrics. Specifically, our framework demonstrates higher accuracy and reliability compared to other methods, as evidenced by a significantly lower average MSE and FID. While SSIM and PSNR serve as indicators of generation quality, the realism of the generated bone layer holds greater significance. As demonstrated in Fig. 4, our method achieves highly realistic bone layer generation while preserving a level of quality comparable to that of existing approaches. Furthermore, in Fig. 3, as the size of the overlapped regions increases, the evaluation index decreases. This is because in cases with large overlap sizes, the joint generally suffers from severe bone erosion and extreme narrowing of the joint space, leading to notable changes in bone texture. Consequently, compared to cases with other overlap sizes, the task complexity of the framework increases significantly. However, our framework continues to significantly outperform other methods, particularly in demonstrating greater robustness at larger overlap sizes.

Additionally, as shown in Fig. 4, the bone layer separation framework achieves the extraction of bone layer images from a single image, effectively eliminating overlapped regions and preserving the complete bone texture. The reconstructed images closely resemble the original images, demonstrating excellent generation quality, which is corroborated by the loss spectrum diagram. In overlap situations, our framework effectively eliminates small overlaps, though the perceptual impact may be subtle. As overlap increases, our framework becomes more effective. Overlapped regions retain the bone textural characteristics, ensuring smooth continuity between overlapped and non-overlapped areas. However, in cases of large overlaps, such as in joints with severe bone erosion, the generation of overlapped regions can be unstable, with noticeable sharp edges. Nevertheless, we successfully extracted layer images and eliminated bone overlap. Compared to other methods, our framework ex-

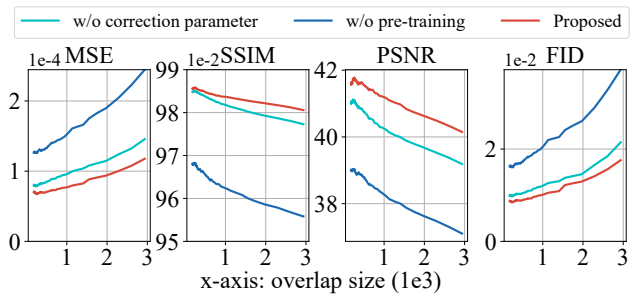

Figure 5: Ablation study results of our framework in different metrics across overlap sizes.

P	C	MSE (10^{-4})	SSIM (10^{-2})	PSNR	FID (10^{-2})
	✓	1.72±1.62	96.33±1.10	38.30±1.86	2.46±3.26
✓		1.05±0.88	98.20±0.70	40.34±1.76	1.44±1.86
✓	✓	**0.88±0.70**	**98.37±0.51**	**41.09±1 .72**	**1.20±1.33**

P: Pre-training, C: Correction Parameter.

Table 3: Comparison results in ablation study.

hibits significant superiority in both layer image generation and reconstruction, particularly in a large overlap, thereby showcasing outstanding accuracy and robustness.

In the amodal completion method with inpainting, the absence of a reconstruction process and non-adherence to conventional radiography principles result in an inability to accurately generate the texture of overlapped regions, leading to reduced accuracy and robustness. Conversely, our framework adheres strictly to imaging principles and incorporates supervision within the reconstruction, thereby enabling the precise generation of textures in overlapped regions. Particularly in large overlap, where the generation based on amodal completion, lacking real image supervision, exhibits limitation of generation. In contrast, our framework integrates the synthetic images pre-training, the segmentation-based supervisor, and the reconstruction structure, facilitating unsupervised network training on real images and extending the performance to accommodate a broader range of overlap.

Ablation Study

We conducted an ablation study centered on the integration of the correction parameter within the reconstructor, and synthetic image pre-training. We examined three distinct framework configurations: the reconstruction function without correction parameter; the training pipeline without the synthetic image pre-training; proposed framework.

As illustrated in Fig. 5, Fig. 6, and Table 3, The introduction of pre-training using synthetic images substantially enhances the quality of generated results. By establishing an initial model, the issues of erroneous generation due to network overfitting are markedly mitigated. Furthermore, the MSE distribution across overlap sizes has been significantly optimized, indicating that the framework performed more stably in different overlap sizes. Additionally, the implementation of correction parameters leads to a considerable improvement in the reconstruction. In comparison to direct reconstruction methods, it effectively reduces brightness amplification in overlapping areas and improves texture synthesis. This enhancement is particularly notable in scenes with extensive overlapped regions, where the continuity and clarity of textures are significantly better. Moreover, the MSE distribution for occluded area sizes shows substantial improvement when compared to direct reconstruction techniques. In conclusion, the integration of pre-training and correction parameters significantly enhances the stability and quality of the generated outputs, further underscoring the necessity of their introduction.

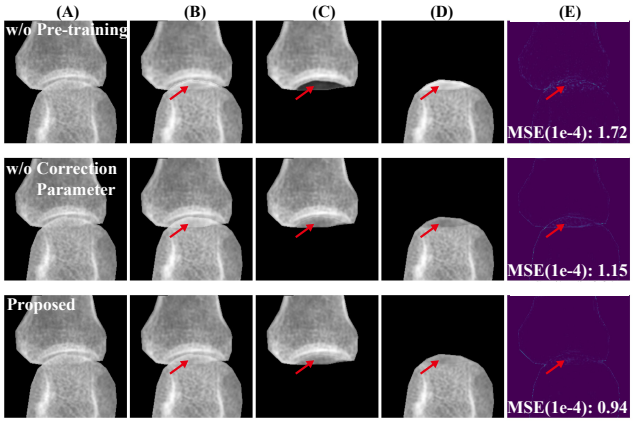

Figure 6: Visualization of ablation study. (A) Real Joint image; (B) Reconstructed Joint Image; (C) Upper Bone Layer; (D) Lower Bone Layer; (E) MSE Spectrum (A v.s. B).

Expert Assessments: Visual Turing Test

We conducted a Visual Turing Test on three sets of 50 images each, comprising joint, upper bone, and lower bone images, with a real-to-fake ratio of 1:1, which informed to subjects. For the joint images, real images were paired with synthetic images used for pre-training. The upper and lower bone sets contained real and generated layer images. Four subjects with 12, 17, 26, and 30 years experience as radiological technologists in the test, which lasted four hours.

The results in Table 4 demonstrated that for the joint image sets, the scores of the four radiological technologists were considerable. But the combined metrics of accuracy, sensitivity, and specificity for the upper and lower image sets were around 0.5. This variability suggests that the synthesized joint images are not completely identical to the real ones, but the aggregated accuracy indicates their suitability as pre-training data. Additionally, the near-random ability of observers to distinguish real from generated images in the upper and lower sets suggests our method has effectively passed the visual Turing test.

Clinical Validation: JSN Quantification

We conducted experiments to clinical performance on downstream tasks, JSN progression quantification, comparing the MSE differences between results with and without the introduction of our bone layer separation framework.

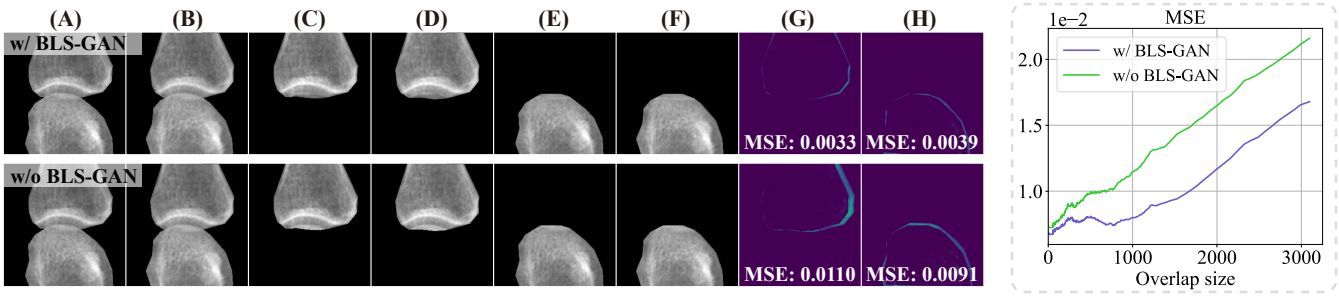

Figure 7: Comparison of JSN progression quantification performance with and without bone layer separation. The **left** panel illustrates the visualization results, while the **right** panel depicts the curve of MSE across the overlap size. (A) Moving Joint Image; (B) Fixed Joint Image; (C) Moving Upper Bone; (D) Fixed Upper Bone; (E) Moving Lower Bone; (F) Fixed Lower Bone; (G) MSE Upper Spectrum (C v.s. D); (H) MSE Lower Spectrum (E v.s. F).

		R1	R2	R3	R4	Overall
Joint Image (Real & Synthetic)	sensitivity	0.76	0.36	0.24	0.40	0.44
	specificity	0.52	0.40	0.20	0.44	0.39
	accuracy	0.64	0.38	0.22	0.42	0.41
Upper bone image (Real & Generated)	sensitivity	0.56	0.48	0.52	0.60	0.54
	specificity	0.44	0.64	0.48	0.32	0.47
	accuracy	0.50	0.56	0.50	0.46	0.51
Lower bone image (Real & Generated)	sensitivity	0.64	0.60	0.44	0.36	0.51
	specificity	0.60	0.44	0.64	0.36	0.51
	accuracy	0.62	0.52	0.54	0.36	0.51

Table 4: Visual Turing Test evaluation results over three image groups. Radiological technologists were tasked to label each set of images as real or fake.

JSN is a crucial indicator for MSK diagnosis, especially for RA progression. (Wang et al. 2023) demonstrated that JSN can be quantified using deep registration to analyze changes between fixed (baseline) and moving (follow-up) images of finger joints. This method utilizes images and joint masks as inputs to produce registration parameters for JSN calculation. The MSE between registered and fixed images validates the registration results.

The MSE results show that the pipeline incorporating BLS-GAN achieves an MSE of 0.0088 ± 0.0118, which is notably lower than the MSE of 0.0103 ± 0.0133 observed in the pipeline without BLS-GAN (P value < 0.0001, 95% confidence interval: 0.001197 to 0.001828, Paired T-test). As shown in Fig. 7, the experimental outcomes further indicate that the introduction of the bone layer separation framework substantially improves both the accuracy and stability of deep registration in JSN quantification, particularly when overlap sizes are less than 1000 pixels. Although the MSE of our framework increases linearly with larger overlap sizes, it consistently remains lower than that of the deep registration method alone, thereby demonstrating the efficacy of this work in managing varying degrees of image overlap.

Conclusion and Limitation

This work initiated a challenging amodal completion scenario for medical images called bone layer separation, which aims to address the impact of MSK joint bone overlap in conventional radiography. We implemented a GAN-based framework named BLS-GAN, which can provide a high-quality image with reasonable bone characteristics and texture. This framework is expected to eliminate the bone overlap in complex joints such as the wrist, hip, and knee, extending the application of automated quantitative methods to a broader range in conventional radiography.

This framework uses a unique reconstructor based on absorption-based imaging principles, reducing recurrent calculations in soft tissue, and achieving high-quality reconstruction. The segmentation-based multi-channel supervisor network accurately distinguishes between overlapped and non-overlapped regions and verifies the authenticity of generated images. Additionally, synthetic images pre-training enhances the stability of the training process and generation.

The expert assessments and clinical validation demonstrated that the framework is capable of generating bone layer images with high clarity, exceptional stability, and a remarkable resemblance to real images. Additionally, the framework significantly enhances the accuracy and stability of downstream JSN quantification tasks. The introduction of our framework addresses the challenges of misalignment and instability caused by overlap in deep registration methods, thereby promoting broader adoption of JSN quantification using deep registration. Additionally, this advancement establishes a practical foundation for extending the method to more complex joints with intricate overlap and provides a solid technical basis for comprehensive, high-precision JSN quantification analysis. To the best of our knowledge, this study is the first application and exploration of amodal completion in conventional radiographs, enabling new developments in amodal completion in medical imaging.

Our current framework is designed to extract bone structures from raw joint radiographs, neglecting essential soft tissue information. While a correction parameter reduces recurrent soft tissue calculations in overlapped regions by adjusting brightness, it does not completely address soft tissue texture interference, which compromises image quality in both overlapped and non-overlapped regions. Future work will focus on accurately generating and differentiating soft tissue regions from bone layers, which is challenging due to the lack of positive samples (without bones) for supervision.

Acknowledgments

We express our deepest gratitude to Hiroyuki Takashima, MD, PhD, Hiroyuki Sugimori, MD, PhD, Kaori Tsutsumi, MD, PhD, and Takaaki Yoshimura, MD, PhD, Faculty of Health Sciences, Hokkaido University, Sapporo, Japan, for their invaluable support and guidance throughout the course of our research. Their expertise and assistance, particularly during the visual Turing test, were instrumental in ensuring the success of this study.

References

Aletaha, D.; and Smolen, J. S. 2018. Diagnosis and management of rheumatoid arthritis: a review. *Jama*, 320(13): 1360–1372.

Ao, J.; Ke, Q.; and Ehinger, K. A. 2023. Image amodal completion: A survey. *Computer Vision and Image Understanding*, 229: 103661.

Bushberg, J. T.; and Boone, J. M. 2011. *The essential physics of medical imaging*. Lippincott Williams & Wilkins.

Chai, T.; and Draxler, R. R. 2014. Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature. *Geoscientific model development*, 7(3): 1247–1250.

Elad, M.; Kawar, B.; and Vaksman, G. 2023. Image denoising: The deep learning revolution and beyond—a survey paper. *SIAM Journal on Imaging Sciences*, 16(3): 1594–1654.

Gong, Y. 2020. Decompose X-ray Images for Bone and Soft Tissue. *arXiv preprint arXiv:2007.14510*.

Grant, W. R.; and Wakefield, R. J. 2018. Musculoskeletal radiology. *ABC of Rheumatology*, 177.

Han, L.; Lyu, Y.; Peng, C.; and Zhou, S. K. 2022. GAN-based disentanglement learning for chest X-ray rib suppression. *Medical Image Analysis*, 77: 102369.

Huda, W.; and Abrahams, R. B. 2015. Radiographic techniques, contrast, and noise in x-ray imaging. *American Journal of Roentgenology*, 204(2): W126–W131.

Ko, K.; and Kim, C.-S. 2023. Continuously masked transformer for image inpainting. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 13169–13178.

Low, K. T.; and Peh, W. C. 2017. Radiography limitations and pitfalls. *Pitfalls in Musculoskeletal Radiology*, 3–32.

Newton, H. B. 2016. *Handbook of neuro-oncology neuroimaging*. Academic Press.

Ou, X.; Chen, X.; Xu, X.; Xie, L.; Chen, X.; Hong, Z.; Bai, H.; Liu, X.; Chen, Q.; Li, L.; et al. 2021. Recent development in x-ray imaging technology: Future and challenges. *Research*.

Ou, Y.; Ambalathankandy, P.; Furuya, R.; Kawada, S.; Zeng, T.; An, Y.; Kamishima, T.; Tamura, K.; and Ikebe, M. 2023. A sub-pixel accurate quantification of joint space narrowing progression in rheumatoid arthritis. *IEEE Journal of Biomedical and Health Informatics*, 27(1): 53–64.

Pasveer, B. 1989. Knowledge of shadows: the introduction of X-ray images in medicine. *Sociology of Health & Illness*, 11(4): 360–381.

Platten, M.; Kisten, Y.; Kälvesten, J.; Arnaud, L.; Forslind, K.; and van Vollenhoven, R. 2017. Fully automated joint space width measurement and digital X-ray radiogrammetry in early RA. *RMD open*, 3(1): e000369.

Sargsyan, A.; Navasardyan, S.; Xu, X.; and Shi, H. 2023. MI-GAN: A Simple Baseline for Image Inpainting on Mobile Devices. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 7335–7345.

Simonyan, K. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Suzuki, K.; Abe, H.; MacMahon, H.; and Doi, K. 2006. Image-processing technique for suppressing ribs in chest radiographs by means of massive training artificial neural network (MTANN). *IEEE Transactions on medical imaging*, 25(4): 406–416.

Wang, H.; Ou, Y.; Fang, W.; Ambalathankandy, P.; Goto, N.; Ota, G.; Okino, T.; Fukae, J.; Sutherland, K.; Ikebe, M.; et al. 2023. A deep registration method for accurate quantification of joint space narrowing progression in rheumatoid arthritis. *Computerized Medical Imaging and Graphics*, 108: 102273.

Xu, K.; Zhang, L.; and Shi, J. 2024. Amodal completion via progressive mixed context diffusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9099–9109.

Yeung, M.; Sala, E.; Schönlieb, C.-B.; and Rundo, L. 2022. Unified focal loss: Generalising dice and cross entropy-based losses to handle class imbalanced medical image segmentation. *Computerized Medical Imaging and Graphics*, 95: 102026.

Zhang, X.; Zhai, D.; Li, T.; Zhou, Y.; and Lin, Y. 2023. Image inpainting based on deep learning: A review. *Information Fusion*, 90: 74–94.