

Article

Not peer-reviewed version

Breast Cancer Bone Metastasis Lesion Segmentation on Bone Scintigraphy

Yi-You Chen , Po-Nien Yu , [Yung-Chi Lai](#) , [Te-Chun Hsieh](#) ^{*} , [Da-Chuan Cheng](#) ^{*}

Posted Date: 11 July 2023

doi: 10.20944/preprints202307.0674.v1

Keywords: bone metastasis segmentation; Double U-Net; pre-train; negative mining; transfer learning; deep learning



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Breast Cancer Bone Metastasis Lesion Segmentation on Bone Scintigraphy

Yi-You Chen ¹, Po-Nien Yu ¹, Yung-Chi Lai ², Te-Chun Hsieh ^{1,3,*} and Da-Chuan Cheng ^{1,*}

¹ Department of Biomedical Imaging and Radiological Science, China Medical University, Taichung 404, Taiwan

² Department of Nuclear Medicine, Feng Yuan Hospital, Ministry of Health and Welfare, Taichung 420210, Taiwan; daniellai999@hotmail.com

³ Department of Nuclear Medicine and PET Center, China Medical University Hospital, Taichung 404, Taiwan; d10119@mail.cmuh.org.tw

* Correspondence: d10119@mail.cmuh.org.tw (T.-C.H.); dccheng@mail.cmu.edu.tw (D.-C.C.); Tel.: +886-4-2205-3366 (D.-C.C.)

Abstract: Bone metastasis detection and quantification on bone scintigraphy is challenging and clinically important for treatment and patient's life quality. To develop a CNN based diagnostic system for automated segmentation on bone metastasis regions is non-trivial, especially for a small dataset. A dataset in house comprising 100 breast cancer patients and 100 prostate cancer patients is utilized for this research. The Double U-Net model is adapted through the integration of background removal, adding negative samples, and transfer learning methods for bone metastasis detection. The performance is investigated via 10-fold cross-validation and computed in pixel-wise scale. The best model we achieved has precision of 63.08%, sensitivity of 70.82%, and F1-score of 66.72%. The developed system has the potential to provide pre-diagnostic reports for physicians in final decisions and the calculation of the bone scan index (BSI) with the combination with bone skeleton segmentation.

Keywords: bone metastasis segmentation; Double U-Net; pre-train; negative mining; transfer learning; deep learning

1. Introduction

According to the gender statistics database published by the Gender Equality Committee of the Executive Yuan in Taiwan in 2023, breast cancer has the first rank among the top 10 cancer incidence rates in 2020 [1]. Breast cancer in Taiwan has the highest crude incidence rate and age-standardized incidence rate, which is classified into five stages, with stage 0 indicating non-invasive cancer and stages 1-3 indicating invasive cancer. Stage 4 represents metastatic cancer [2, 3]. The primary goal of treating non-metastatic breast cancer is to remove the tumor from the breast and regional lymph nodes and prevent metastatic recurrence. Local treatments for non-metastatic breast cancer may include surgical tumor removal, sampling or removal of underarm lymph nodes, and adjuvant radiation therapy. For metastatic breast cancer, the goals of treatment are to prolong life, alleviate symptoms [4] and better the patients' life quality.

Current methods for detecting breast cancer metastasis include clinical observation of distant organ involvement, organ biopsies, diagnostic imaging, and serum tumor markers. The skeletal system is the most frequent site of breast cancer metastasis, often presenting as osteolytic lesions [5]. Bone metastasis in breast cancer most commonly occurs in the spine, followed by the ribs and sternum [6]. While there are various types of bone lesions, metastatic bone disease is the primary clinical concern [7]. A study by Coleman and Rubens reported bone metastasis in 69% of breast cancer patients who died between 1979 and 1984, out of a total of 587 patients [8]. Breast cancer, prostate cancer, lung cancer, and other prevalent cancers account for more than 80% of cases of metastatic

bone disease. The clinical course of breast cancer is often prolonged, and the subsequent occurrence of bone involvement poses a significant burden on healthcare resources. Radiologically, bone metastases in breast cancer are predominantly osteolytic, leading to severe complications such as bone pain, pathological fractures, spinal cord compression, hypercalcemia, and bone marrow suppression [6]. Early detection of breast cancer bone metastasis is crucial for preventing and managing potential complications. Hormone therapy is the preferred initial treatment for breast cancer with bone metastasis, while chemotherapy is administered to patients who do not respond to or experience relapse after hormone therapy. For patients with newly diagnosed bone metastasis, this indicates disease recurrence or treatment failure. It is necessary to have a modification of the treatment strategy through adjustments in chemotherapy or hormone therapy based on the individual patient's disease condition. Therefore, a quantification on bone metastasis such as BSI computation before and after the treatment is clinically important for treatment observations.

One of the primary imaging techniques used in clinics for bone metastasis diagnosis is the whole-body bone scan (WBBS) with vein injection using the Tc-99m MDP tracer [9, 10]. WBBS offers the advantages of whole-body examination, cost-effectiveness, and high sensitivity, making it a preferred modality for bone metastasis screening [11]. Unlike X-radiography (XR) and computed tomography (CT) images, which can only detect changes in bone when there is approximately 40%-50% mineralization [12], bone scans exhibit higher sensitivity in detecting bone changes, capable of detecting alterations as low as 5% in osteoblast activity. The reported sensitivity and specificity of skeletal scintigraphy for bone metastasis detection are 78% and 48%, respectively [13]. Given the relatively low specificity of WBBS in image diagnosis, our goal is to develop a computer-aided diagnosis system utilizing deep learning to assist physicians in interpretation.

The bone scan index (BSI) is an image biomarker utilized in WBBS to evaluate the severity of bone metastasis in cancer patients. It enables a quantification on the degree of tumor involvement in the skeleton [14, 15]. BSI is used for observing disease progression or treatment response. The commercial software EXINI bone, developed by EXINI Diagnostics AB, incorporates aBSI (automated bone scan index) technology for comprehensive automated quantitative assessment of bone scan images [16]. In [16], there exists a strong correlation between manual and automated BSI assessment values ($\rho = 0.80$), which further strengthens ($\rho = 0.93$) when cases with BSI scores exceeding 10 (1.8%) are excluded. This indicates that automated BSI calculations can deliver clinical value comparable to manual calculations. Shimizu et al. has proposed an image interpretation system based on deep learning [17], using BtrflyNets for hotspot detection of bone metastasis and bone segmentation, followed by automatic BSI calculation. The aBSI technology has now become a clinically valuable tool. Nevertheless, there is still challenging on recognition performance (sensitivity and precision) in this technique.

Object detection and image segmentation are active research areas in computer vision since decades. Cheng et al. applied a deep convolutional neural network (D-CNN) for the detection of bone metastasis from prostate cancer in bone scan images [18]. Their investigation specifically focused on the chest and pelvic regions, and the sensitivity and precision for detecting and classifying chest bone metastasis were determined by bounding boxes to be 0.82 ± 0.08 and 0.70 ± 0.11 , respectively. Regarding pelvic bone metastasis classification, the reported sensitivity and specificity were 0.87 ± 0.12 and 0.81 ± 0.11 , respectively. Cheng et al. conducted a more detailed study on chest bone metastasis in prostate cancer patients [19]. The average sensitivity and precision for detecting and classifying chest bone metastasis based on lesion locations are reported as 0.72 ± 0.04 and 0.90 ± 0.04 , respectively. For classifying chest bone metastasis based on patient-level outcomes, the average sensitivity and specificity are found to be 0.94 ± 0.09 and 0.92 ± 0.09 , respectively. Patents filed by Cheng et al. are referenced as [20, 21], which leverage deep learning for the identification of bone metastasis in prostate cancer bone scan images. Since they use bounding boxes, therefore, they are unable to calculate BSI.

Convolutional neural network (CNN) has demonstrated its ability in medical image segmentation [22]. For semantic segmentation tasks, early deep learning architectures include Fully Convolutional Networks (FCN). U-Net [23] is another widely used image segmentation architecture.

In a related study [24], a neural network (NN) model based on U-Net++ is proposed for automated segmentation of metastatic lesions in bone scan images. The anterior-posterior and posterior-anterior views are superimposed, and image segmentation is exclusively performed on the chest region of whole-body bone scan images. The achieved average F1-score is 65.56%.

Jha et al. has proposed Double U-Net [25], which combines an improved U-Net architecture with VGG-19 in the encoder part and utilizes two U-Net architectures within the network. In addition to using SE blocks to enhance the information in feature maps [26], Double U-Net also uses atrous spatial pyramid pooling (ASPP) with dilated convolutions to capture contextual information in the network [27]. These techniques contribute to achieving superior model performance compared to the U-Net architecture.

2. Materials and Methods

2.1. Materials

In this study, we have collected 200 bone scan images from the Department of Nuclear Medicine of China Medical University Hospital. Among these images, 100 images are obtained from breast cancer patients, including 90 images with bone metastasis and 10 images without bone metastasis. The remaining 100 images are from prostate cancer patients, with 50 images showing bone metastasis and 50 images without bone metastasis. This study has been approved by the Institutional Review Board (IRB) of China Medical University and Hospital Research Ethics Committee (CMUH106-REC2-130), approved in 27 September 2017.

The WBBS process can be described as follows. Patients undergo WBBS with a gamma camera (Millennium MG, Infinia Hawkeye 4, or Discovery NM/CT 670 system; GE Healthcare, Waukesha, WI, USA). Bone scans are acquired 2–4 hours after the intravenous injection of 740–925 MBq (20–25 mCi) of technetium-99m methylene diphosphonate (Tc-99m MDP) with an acquisition time of 10–15 cm/min. The collected WBBS images are saved in DICOM format. The raw images include anterior-posterior (AP) and posterior-anterior (PA) views, with a matrix size of 1024×256 pixels.

2.2. Image Labeling

To facilitate labeling the bone scan images, the Labelme software is used as the annotation tool. The manual annotation of bone metastasis images is carried out under the guidance and supervision of nuclear medicine physicians. This process is very time-consuming. The outputs generated by the Labelme software are saved in JSON format, and then converted to the PNG format.

2.3. Image Pre-Processing

The raw images possess a large memory size and the DICOM format is not directly suitable for neural network training. Moreover, the raw images exhibit variations in brightness and contrast levels. Thus, pre-processing of the raw images becomes imperative. The detection of the body range was accomplished using the projection profile, followed by the extraction of two views with dimensions of 950×256 pixels through cutting and centering. No scaling or other transformations were applied during this process. We utilized the brightness normalization method proposed in [19] for brightness pre-processing. This method uses a linear transformation to adjust the dynamic range of an image, with the objective of controlling the average intensity of each image within the range of (7, 14). The algorithm for the linear transformation is illustrated in Figure 1. The region below the knees, which is uncommon for bone metastasis, was excluded from the calculation of BSI. To obtain the region above the knees, pixels beyond row 640 are eliminated, resulting in two views with a spatial resolution of 640×256 pixels each. Finally, the pre-processed AP (anterior-posterior) and PA (posterior-anterior) view images were horizontally merged, generating images with a spatial resolution of 640×512 pixels.

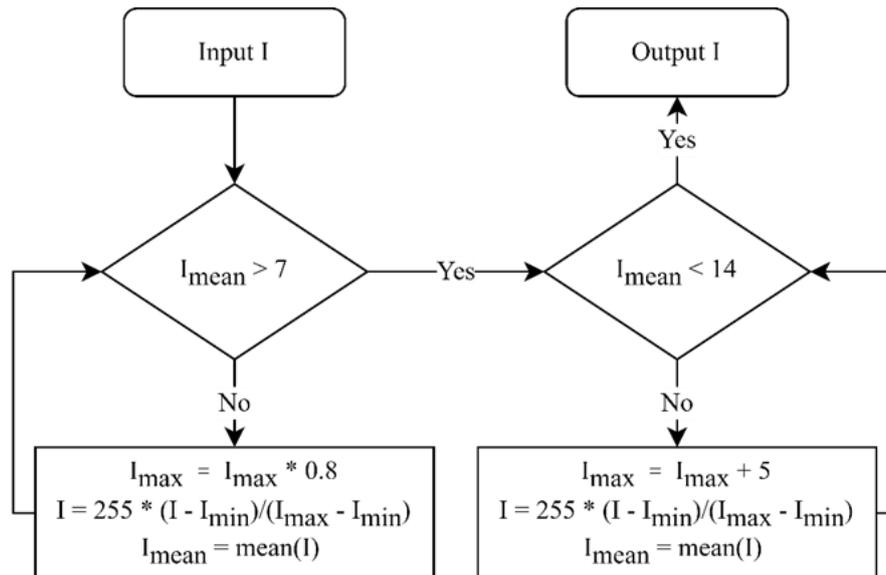


Figure 1. Flowchart of brightness normalization.

2.4. Background Removal

Given that the background of bone scan images predominantly consists of air, training on the air regions would generate meaningless results. Thus, a pre-process is used to eliminate the air regions from the images. This is achieved by applying Otsu thresholding [28] to the non-metastatic regions, allowing for the automatic determination of the optimal threshold to remove the air regions and generate a mask specifically for the non-metastatic class, as shown in Figure 2.

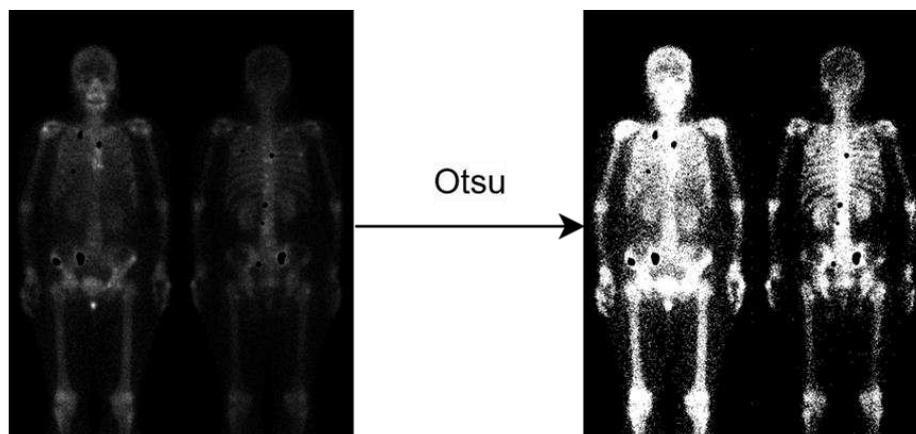


Figure 2. Illustration of background removal. Notably, the metastasis hotspots are eliminated (the black holes), if this image has metastasis.

2.5. Adding Negative Samples

To solve the challenge of distinguishing between lesions and normal bone uptake hotspots, it is important to include normal images in the training dataset. This inclusion helps reduce the false positive rate and enhance the precision of the model. The mask for normal images can be generated using two different techniques: Otsu thresholding and negative mining.

The first approach involved generating a mask for non-metastatic samples by directly applying Otsu thresholding. This process can be applied on both normal images without metastasis or metastatic images. The Otsu thresholding finds a threshold to distinguish uptake regions from air, thus creating the non-metastatic class. Notably, the metastatic regions must be excluded manually if we are processing an image with metastasis. The second approach, known as negative mining [19],

utilizes a two-step process. Initially, the model is trained with bone metastasis images. Subsequently, this trained model is used to make predictions on normal images. Any output predictions on the normal images are for sure to be false positives and collected as the non-metastatic class. Negative mining offers the advantage of automatically generating negative samples without the need for manual annotation by experts, thereby saving time and effort. Additionally, negative mining focuses on areas where the model is prone to making mistakes, enabling error review and subsequent model improvement.

2.6. Transfer Learning

Transfer learning is a widely used technique in neural networks to increase their performance. Before applying transfer learning, two crucial factors need to be considered: (1) the size of the target dataset and (2) the similarity between the target dataset and the pre-training dataset.

In this study, a relatively small dataset is utilized, and it is anticipated that pre-training with a dissimilar dataset would have limited effectiveness. Consequently, we select a pre-training dataset comprising highly similar bone scan images from the prostate cancer. Subsequently, fine-tuning is performed on the model using the target dataset, which consisted of breast cancer bone scan images. By leveraging transfer learning and selecting a pre-training dataset that closely aligned with the target dataset, we aim to capitalize on the shared characteristics between the two datasets, thereby improving the model's performance on the specific task at hand.

2.7. Data Augmentation

Neural networks (NNs) have demonstrated impressive performance in various computer vision tasks. However, to avoid overfitting, NNs typically require a substantial amount of training data. However, overfitting occurs while an NN excessively matches a specific dataset, whereas limiting its ability to generalize onto new test data. Medical images for annotations are often time-consuming so limited in training samples. Therefore, to solve the data scarcity and mitigate overfitting, data augmentation techniques are required.

In this study, we utilize two data augmentation techniques: brightness adjustment and horizontal flipping. For brightness adjustment, a linear transformation process like the brightness normalization shown in Figure 1 is applied. The average brightness value (m) of the original images served as a reference for adjusting the brightness. By setting the brightness range to $(m - 5, m + 5)$, six additional brightness values were generated within this range, resulting in a total of seven brightness levels when combined with m . Subsequently, the images corresponding to these seven brightness levels are horizontally flipped. Consequently, the dataset size is effectively augmented by a factor of 14.

2.8. 10-Fold Cross-Validation

To reduce the impact of a single random split and evaluate the model's generalization performance on independent test data, we conduct 10-fold cross-validation for performance evaluation. The positive samples are divided into ten parts, with eight parts used for training, one part for validation, and one part for test, respectively in each fold. Additionally, the negative samples are included in the training set of each fold. Consequently, the model is trained on these ten subsets of data, and the results are averaged to obtain a comprehensive assessment of the model's performance. This approach ensures a reliable evaluation and provides insights into the model's effectiveness across different subsets of the dataset.

2.9. Neural Network Model

In this study, we modify Double U-Net model. The original Double U-Net architecture is developed for binary segmentation tasks, requiring modifications to adapt it for multi-class segmentation purposes. The modified Double U-Net architecture, as illustrated in Figure 3, is utilized in this study. Specifically, the activation functions for OUTPUT 1 and OUTPUT 2 are changed to

softmax to facilitate multi-class segmentation. Additionally, the original input image is multiplied with each output class, resulting in separate images that served as inputs for NETWORK 2. This modification enables the model to handle multi-class segmentation efficiently.

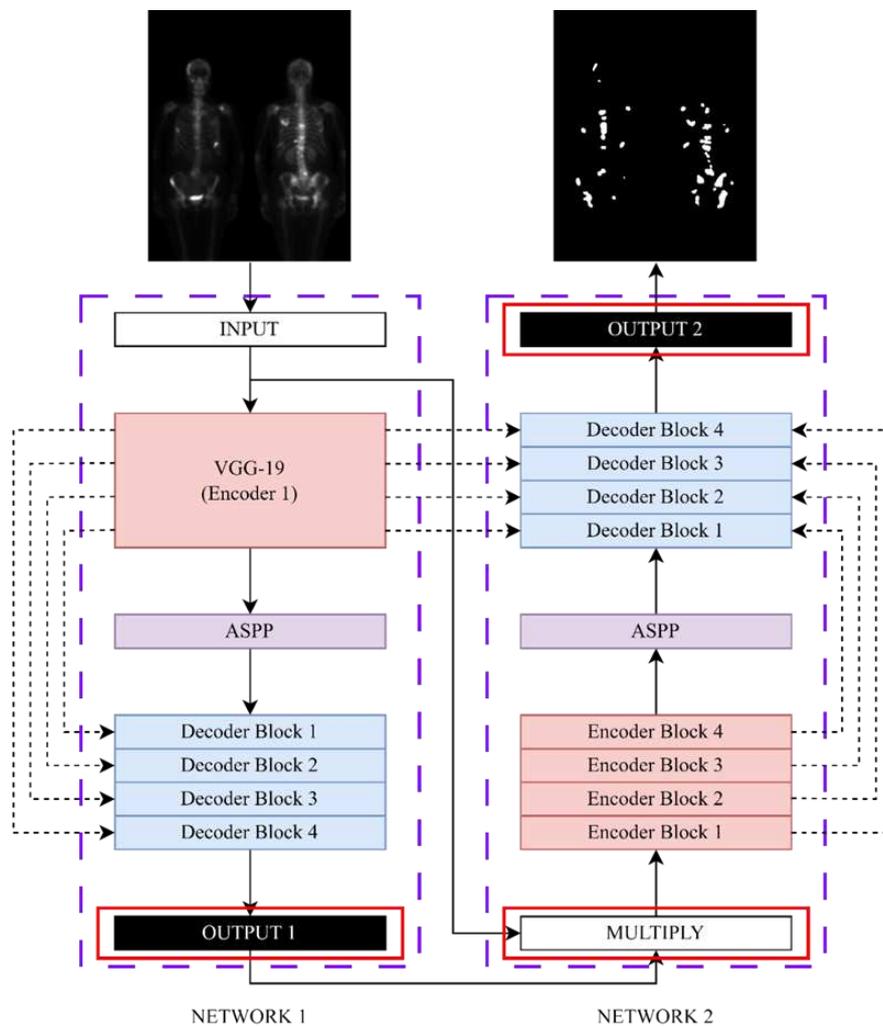


Figure 3. The modified architecture diagram of Double U-Net.

2.10. Loss Function

The selection of an appropriate loss function is a critical aspect in the design of deep learning architectures for image segmentation tasks, as it greatly impacts the learning dynamics of the algorithm. In our study, we consider two loss functions: the Dice loss (Equation 1), as originally proposed in [25], and the Focal Tversky loss (Equation 2). By comparing these loss functions, we aim to explore their respective influences on the model's performance in the context of our specific task.

The Dice coefficient is a widely adopted metric in computer vision for assessing the similarity between two images. In our study, we utilize a modified version of the Dice coefficient known as the Dice loss, which served as a loss function for our model.

$$DL(y, p) = 1 - \frac{2yp}{y + p} \quad (1)$$

where y is true value, and p is the predicted outcome.

The Focal Tversky loss is particularly well-suited for solving highly imbalanced class scenarios. It incorporates a γ coefficient that allows for the down-weighting of easy samples. Additionally, by adjusting the α and β coefficients, different weights can be assigned to false positives (FP) and false negatives (FN).

$$\text{FTL}(y, p) = \left(1 - \frac{yp}{yp + \alpha(1-y)p + \beta y(1-p)}\right)^\gamma \quad (2)$$

where $\gamma = 0.75$, $\alpha = 0.3$, and $\beta = 0.7$.

3. Results

3.1. Double U-Net Baseline Model

The original Double U-Net model is trained by 90 bone metastasis images of breast cancer patients. The Dice loss function is used, with a learning rate of 0.0001, a batch size of 4, and 500 epochs. All experimental results are obtained through 10-fold cross-validation, with a ratio of 8:1:1 for the training, validation, and testing sets, respectively. The goal of this experiment is to establish the baseline performance of the model.

To facilitate meaningful cross-comparisons, it is essential to evaluate the performance of deep learning models in each task using quantitative metrics. In this study, the precision, sensitivity, and F1-score are utilized for performance evaluation. Figure 4 shows the qualitative results, and the quantitative results are shown in Table 1.

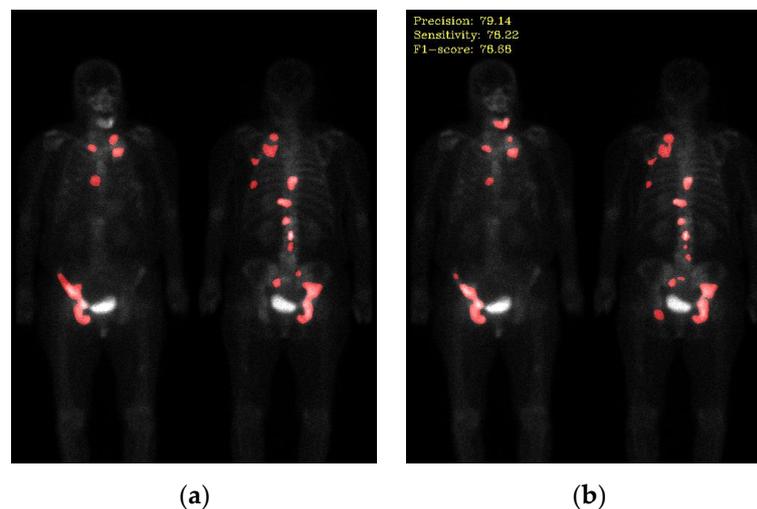


Figure 4. The qualitative result of the baseline model. (a) Ground truth; (b) segmentation results (precision: 79.14, sensitivity: 78.22, F1-score: 78.68).

Table 1. The quantitative results of the baseline model. Use 10-fold cross-validation.

Fold Number	Precision	Sensitivity	F1-score
1	49.21	79.19	60.70
2	58.74	64.98	61.70
3	70.56	60.01	64.86
4	81.69	52.20	63.70
5	60.57	54.06	57.13
6	72.55	45.92	56.24
7	43.63	83.32	57.27
8	49.03	68.21	57.05
9	61.73	60.84	61.28
10	67.89	61.13	64.34
Mean	61.56	62.99	62.27

3.2. Double U-Net Architecture Modification and Optimization Methods

The modified Double U-Net model is trained by 90 bone metastasis images of breast cancer patients as well. Additionally, the Focal Tversky loss is used for comparison purposes. The use of background removal significantly enhances the model's performance. Thus, subsequent experiments utilize this approach.

Furthermore, 10 normal images of breast cancer patients are added into the training set with either Otsu thresholding or negative mining methods. Both methods resulted in an increase in precision but a decrease in sensitivity. However, adding negative samples does not result in a significant improvement in F1-score. Adding negative samples does increase the precision, which align with our expectations. In the case of the Dice loss model, the precision increases and is over 70%.

To better the performance, the model is pre-trained with 100 bone scan images of prostate cancer patients and then fine-tuned on the breast cancer dataset using transfer learning. Figure 5 shows the qualitative results, and the quantitative results are shown in Table 2. Adding negative samples improved the F1-score when using Otsu thresholding, while it decreased when using the negative mining method.

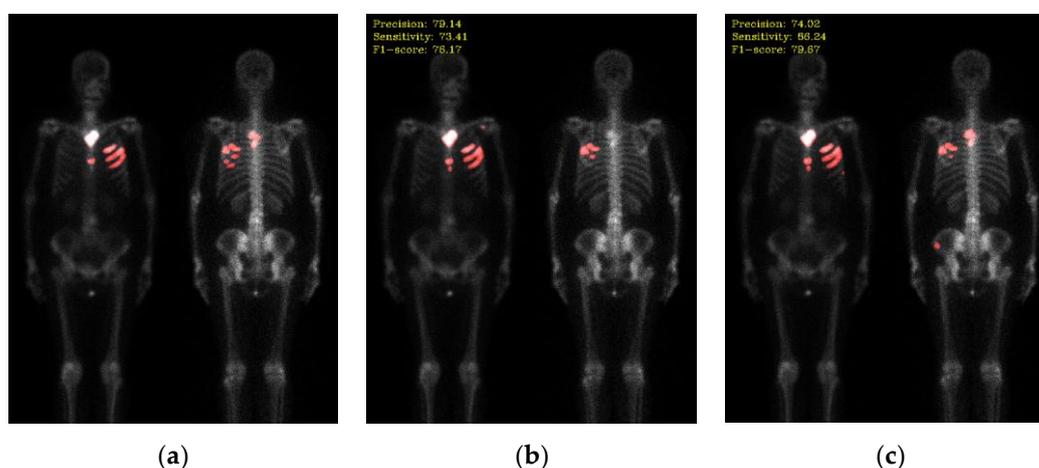


Figure 5. The qualitative results of the modified Double U-Net. (a) Ground truth; (b) segmentation results with Dice loss (precision: 79.14, sensitivity: 73.41, F1-score: 76.17); (c) segmentation results with Focal Tversky loss (precision: 74.02, sensitivity: 86.24, F1-score: 79.67).

Table 2. The quantitative results of the modified Double U-Net. Number is the average of 10-fold cross-validation in percentage.

Loss Function	Negative Samples	Transfer Learning	Precision	Sensitivity	F1-score
Dice	No	No	67.63	63.29	65.39
	Otsu Thresholding	No	70.24	61.80	65.75
		Yes	69.68	63.00	66.17
	Negative Mining	No	70.18	61.42	65.51
Yes		66.25	64.24	65.23	
Focal Tversky	No	No	67.87	65.04	66.43
	Otsu	No	69.96	63.55	66.60

Thresholding	Yes	63.08	70.82	66.72
Negative Mining	No	68.66	62.67	65.53
	Yes	61.49	69.28	65.15

Overall, the background elimination using Otsu thresholding shows the best improvement in performance, among the two methods: negative mining and Otsu thresholding. The same trend is observed in applying transfer learning.

3.3. Model Performance after Training Data Augmentation

In the previous experiment, the best optimization approach is found by adding negative samples using Otsu thresholding method, and applying transfer learning. The model trained with Focal Tversky loss achieves the best result. Thus, we conduct the further experiment on data augmentation for this model as shown in Table 3.

Table 3. The quantitative results of the modified Double U-Net with data augmentation.

Fold Number	Precision	Sensitivity	F1-score
1	65.69	72.10	68.75
2	67.51	66.55	67.03
3	69.10	69.82	69.46
4	82.94	61.02	70.31
5	60.18	61.69	60.93
6	77.22	53.48	63.19
7	58.39	70.78	63.99
8	51.06	71.14	59.45
9	61.93	65.56	63.69
10	59.15	80.25	68.10
Mean	65.32	67.24	66.26

4. Discussion

In this study, the vanilla Double U-Net architecture served as the baseline model for performance comparison. Subsequently, several optimization techniques are used to better the model. These techniques encompass adding negative samples (using Otsu thresholding), transfer learning, and data augmentation. Furthermore, the performance of different loss functions is compared. However, not all optimization methods result in improvement in performance. We discuss them in the following.

Among all the optimization methods, background removal is the most significant technique in improving the model's performance. In bone scan images, the background mostly consists of air, which is unnecessary and irrelevant information. Extracting features from the air region would lower the model's performance and waste computational resources. Applying Otsu thresholding can easily remove the air background. This method generates significant improvement without manual labeling, which is time-consuming.

In comparison from F1-score in Table 2, the Otsu thresholding method shows slightly better than the negative mining method. We think that this might be due to the insufficient number of negative samples, compare to the area made by background removal as shown in Figure 2. In the future, we plan to add more negative samples and further investigate the negative mining method to validate

our assumption. The combination of background removing and negative mining would also a possible way to try.

In reference [19], the task involves lesion detection and classification of three classes: metastatic, indeterminate, and normal hotspots. They use bounding boxes to define the region. The background removing we propose here cannot be applied to their application, since our task is semantic segmentation on bone metastatic lesions. In general, the negative mining technique reduces false positives, with the sacrifice of slight sensitivity. The difference in tasks between the two studies may have contributed to the variation in the effectiveness of negative mining.

Data augmentation is a commonly used technique to improve the robustness of deep learning models by generating new data samples from existing ones. It can solve data scarcity and alleviate overfitting issues. In our study, we utilize two data augmentation methods: brightness adjustment and horizontal flipping. However, the model performance is not better. Data augmentation has a smaller impact on more complex network models [29]. In addition to the brightness adjustment and horizontal flipping used in our study, there are other methods such as geometric transformations, color space augmentation, random erasing, feature space augmentation, and generative adversarial networks [30]. However, not all augmentation method is suitable to our study.

Apart from data augmentation, regularization methods specific to the model architecture, such as dropout, batch normalization, one-shot, and zero-shot learning algorithms, can effectively alleviate overfitting issues. In the Detectron2 and Scaled-YOLOv4 models [31], using the same data and data augmentation methods, the two models exhibited significant differences in performance. The former achieves better results after data augmentation, while the latter obtains worse results. This indicates that the effectiveness of data augmentation is greatly influenced by the model architecture. Due to computational cost considerations, our study only uses brightness adjustment and horizontal flipping. In the future, we need to explore different data augmentation methods to find the optimal combination for the target dataset and network model.

In this study, we have tried Dice loss and Focal Tversky loss, and the latter generates better results. The selection of loss function plays a crucial role in model performance. For complex tasks like segmentation, there is no universally applicable loss function. It largely depends on the properties of the training dataset, such as distribution, skewness, boundaries, etc. For segmentation tasks with extreme class imbalance, Focal-related loss functions are more appropriate [32]. Additionally, since the vanilla Double U-Net model has a higher precision than sensitivity, we are keen to use Tversky-related loss functions to balance the false positives (FP) and false negatives (FN) rates. Therefore, we adopt Focal Tversky loss as the compared loss function. In the future, further exploration and research can be conducted on the selection of optimizer.

Not all hotspots in bone scan images represent bone metastases; normal bone absorption, renal metabolism, inflammation, and injuries can also cause hotspots in the images, leading to false positives in segmentation. In addition to the inherent imaging principles of bone scan images that make training the model challenging, the presence of artifacts in the images is also a crucial factor leading to misclassification. Examples of such artifacts include high-activity areas like the kidneys and bladder, the injection site of the radioactive isotope, and motion artifacts, as shown in Figure 6. Apart from artifacts in breast cancer images, prostate cancer bone scan images also exhibit high-activity artifacts from catheters, urine bags, and diapers, as shown in Figure 7. In the future, appropriate removal can be applied to minimize the impact from artifacts, or additional classes such as benign lesions and artifacts can be introduced to train the model more accurately.

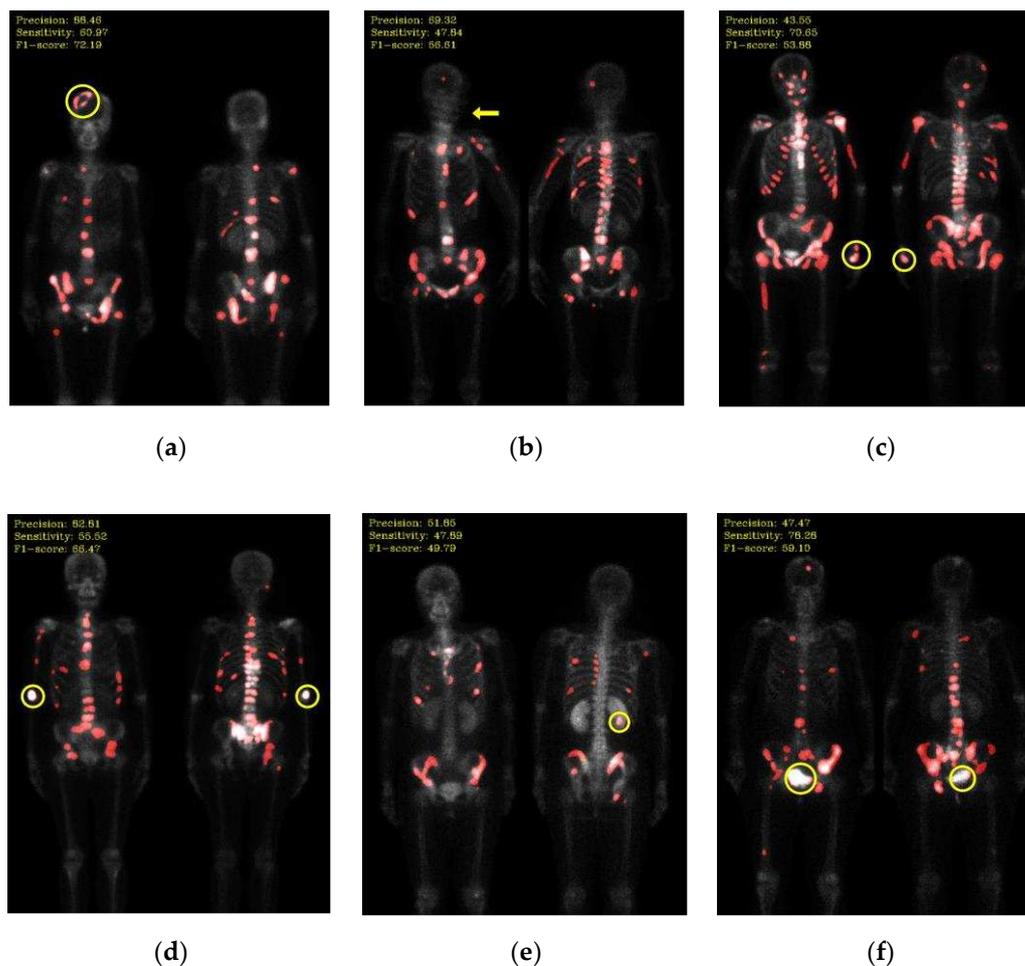


Figure 6. Mis-segmentation of non-metastatic lesions. (a) Bone fracture (head region) (precision: 88.46, sensitivity: 60.97, F1-score: 72.19); (b) motion artifact (head region) (precision: 69.32, sensitivity: 47.84, F1-score: 56.61); (c) injection site (wrist) (precision: 43.55, sensitivity: 70.65, F1-score: 53.88); (d) injection site (elbow) (precision: 82.81, sensitivity: 55.52, F1-score: 66.47); (e) kidney (precision: 51.85, sensitivity: 47.89, F1-score: 49.79); (f) bladder (precision: 47.47, sensitivity: 78.28, F1-score: 59.10).

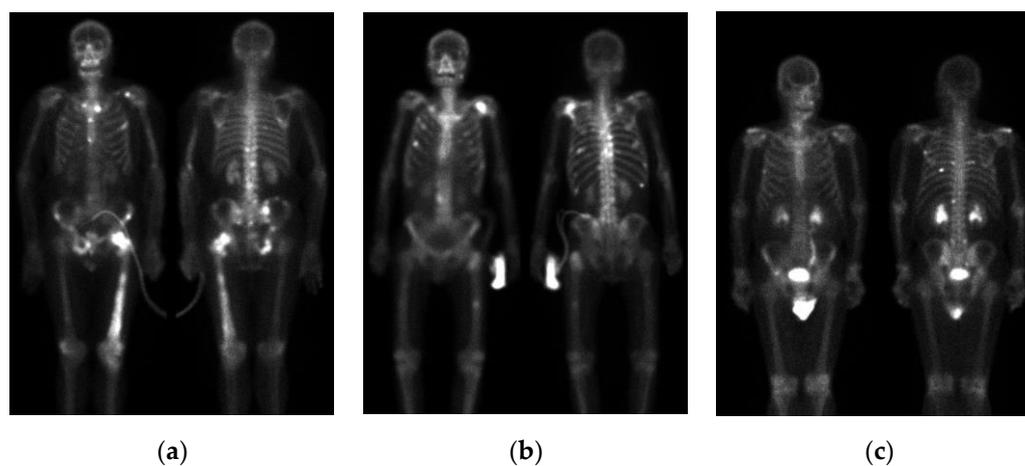


Figure 7. Artifacts in bone scan images of prostate cancer. (a) Catheter; (b) urinary bag; (c) diaper.

The image pre-processing methods used in this study have significantly improved the model. Cheng et al. proposes a pre-processing method where the original images are combined into a 3D image to alleviate the issue of spatial connectivity loss [19]. View aggregation, an operation applied to bone scan images, are used to enhance areas of high absorption [24]. This method enhances lesions

that appear in both anterior and posterior view images and maps lesions that only appear in either anterior or posterior view images. In the future, we can explore additional image pre-processing methods.

Due to the limited number of the dataset in house in this study, the model generalization is limited, and the training process may suffer from biases and limitations. The collaboration with other medical centers can be established to obtain more diverse and abundant data, thereby enhancing the model performance and generalization capability.

5. Conclusions

In this study, we modify the Double U-Net model to perform multiclass segmentation of bone metastatic lesions in breast cancer bone scan images. The model is optimized using background removal and transfer learning. The precision, sensitivity, and F1-score of the segmentation of bone metastatic lesions are calculated in pixel-wise scale and they reach 63.08%, 70.82%, and 66.72%, respectively.

Author Contributions: Conceptualization, DCC; methodology, DCC; software, PNY and YYC; validation, PNY and YYC; formal analysis DCC; investigation, YCL and TCH; resources, TCH and DCC; data curation, PNY and YYC; writing—original draft preparation, YYC; writing—review and editing, DCC; visualization, DCC; supervision, DCC; project administration, DCC; funding acquisition, DCC. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Science and Technology Council (NSTC), Taiwan, grant number MOST 111-2314-B-039-040.

Institutional Review Board Statement: The study was approved by the Institutional Review Board (IRB) and the Hospital Research Ethics Committee (CMUH106-REC2-130) of China Medical University.

Informed Consent Statement: Patient consent was waived by IRB due to this is a retrospective study and only images were used without patient's identification.

Acknowledgments: We thank to National Center for High-performance Computing (NCHC) for providing computational and storage resources.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Gender Equality Committee of the Executive Yuan. Available online: https://www.gender.gov.tw/gecdb/Stat_Statistics_DetailData.aspx?sn=nLF9GdMD%2B%2Bv41SsobdVgKw%3D%3D (accessed on 24 February 2023).
2. Breastcancer.org. Available online: <https://www.breastcancer.org/treatment/planning/options-by-stage> (accessed on 25 February 2023).
3. Simos, D.; Catley, C.; van Walraven, C.; Arnaout, A.; Booth, C.M.; McInnes, M.; et al. Imaging for distant metastases in women with early-stage breast cancer: a population-based cohort study. *Cmaj* **2015**, *187*, E387-E397.
4. Waks, A.G.; Winer, E.P. Breast cancer treatment: a review. *Jama* **2019**, *321*, 288-300.
5. Akhtari, M.; Mansuri, J.; Newman, K.A.; Guise, T.M.; Seth, P. Biology of breast cancer bone metastasis. *Cancer biology & therapy* **2008**, *7*, 3-9.
6. Kakhki, V.R.D.; Anvari, K.; Sadeghi, R.; Mahmoudian, A.S.; Torabian-Kakhki, M. Pattern and distribution of bone metastases in common malignant tumors. *Nucl. Med. Rev.* **2013**, *16*, 66-69.
7. Rubens, R.D. Bone metastases—the clinical problem. *European Journal of Cancer* **1998**, *34*, 210-213.
8. Coleman, R.E.; Rubens, R.D. The clinical course of bone metastases from breast cancer. *British journal of cancer* **1987**, *55*, 61-66.
9. Hamaoka, T.; Madewell, J.E.; Podoloff, D.A.; Hortobagyi, G.N.; Ueno, N.T. Bone imaging in metastatic breast cancer. *J. Clin. Oncol.* **2004**, *22*, 2942-2953.
10. Even-Sapir, E.; Metser, U.; Mishani, E.; Lievshitz, G.; Lerman, H.; Leibovitch, I. The detection of bone metastases in patients with high-risk prostate cancer: 99mTc-MDP Planar bone scintigraphy, single- and multifield-of-viewSPECT, 18F-fluoride PET, and 18F-fluoride PET/CT. *J. Nucl. Med.* **2006**, *47*, 287-297.
11. Costelloe, C.M.; Rohren, E.M.; Madewell, J.E.; Hamaoka, T.; Theriault, R.L.; Yu, T.K.; A.; et al. Imaging bone metastases in breast cancer: techniques and recommendations for diagnosis. *The lancet oncology* **2009**, *10*, 606-614.

12. Vijayanathan, S.; Butt, S.; Gnanasegaran, G.; Groves, A.M. Advantages and limitations of imaging the musculoskeletal system by conventional radiological, radionuclide, and hybrid modalities. *Seminars in Nuclear Medicine* **2009**, *39*, 357-368.
13. O'Sullivan, G.J.; Carty, F.L.; Cronin, C.G. Imaging of bone metastasis: an update. *World journal of radiology* **2015**, *7*, 202.
14. Imbriaco, M.; Larson, S.M.; Yeung, H.W.; Mawlawi, O.R.; Erdi, Y.; Venkatraman, E.S.; Scher, H.I. A new parameter for measuring metastatic bone involvement by prostate cancer: the Bone Scan Index. *Clin. Cancer Res.* **1998**, *4*, 1765-1772.
15. Erdi, Y.E.; Humm, J.L.; Imbriaco, M.; Yeung, H.; Larson, S.M. Quantitative bone metastases analysis based on image segmentation. *Journal of Nuclear Medicine* **1997**, *38*, 1401-1406.
16. Ulmert, D.; Kaboteh, R.; Fox, J.J.; Savage, C.; Evans, M.J.; Lilja, H.; et al. A novel automated platform for quantifying the extent of skeletal tumour involvement in prostate cancer patients using the Bone Scan Index. *European urology* **2012**, *62*, 78-84.
17. Shimizu, A.; Wakabayashi, H.; Kanamori, T.; Saito, A.; Nishikawa, K.; Daisaki, H.; et al. Automated measurement of bone scan index from a whole-body bone scintigram. *Int. J. Comput. Assist. Radiol. Surg.* **2020**, *15*, 389-400.
18. Cheng, D.C.; Liu, C.C.; Hsieh, T.C.; Yen, K.Y.; Kao, C.H. Bone metastasis detection in the chest and pelvis from a whole-body bone scan using deep learning and a small dataset. *Electronics* **2021**, *10*, 1201.
19. Cheng, D.C.; Hsieh, T.C.; Yen, K.Y.; Kao, C.H. Lesion-based bone metastasis detection in chest bone scintigraphy images of prostate cancer patients using pre-train, negative mining, and deep learning. *Diagnostics* **2021**, *11*, 518.
20. Cheng, D.C.; Liu, C.C.; Kao, C.H.; Hsieh, T.C. System of deep learning neural network in prostate cancer bone metastasis identification based on whole body bone scan images. U.S. Patent US11488303B2, 1 November 2022.
21. Cheng, D.C.; Liu, C.C.; Kao, C.H.; Hsieh, T.C. System of deep learning neural network in prostate cancer bone metastasis identification based on whole body bone scan images. Taiwan Patent 202117742 / I709147, 1 November 2020.
22. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; et al. A survey on deep learning in medical image analysis. *Medical image analysis* **2017**, *42*, 60-88.
23. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. *arXiv* **2015**, arXiv:1505.04597v1.
24. Cao, Y.; Liu, L.; Chen, X.; Man, Z.; Lin, Q.; Zeng, X.; Huang, X. Segmentation of lung cancer-caused metastatic lesions in bone scan images using self-defined model with deep supervision. *Biomedical Signal Processing and Control* **2023**, *79*, 104068.
25. Jha, D.; Riegler, M.A.; Johansen, D.; Halvorsen, P.; Johansen, H.D. Doubleu-net: A deep convolutional neural network for medical image segmentation. In Proceedings of the IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS), Rochester, MN, USA, 28-30 July 2020; pp. 558-564.
26. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. *arXiv* **2019**, arXiv:1709.01507v4.
27. Chen, L.C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking atrous convolution for semantic image segmentation. *arXiv* **2017**, arXiv:1706.05587v3.
28. Otsu, N. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern.* **1979**, *9*, 62-66.
29. Bhuse, P.; Singh, B.; Raut, P. Effect of Data Augmentation on the Accuracy of Convolutional Neural Networks. In Proceedings of the Information and Communication Technology for Competitive Strategies (ICTCS) ICT: Applications and Social Interfaces, Jaipur, Rajasthan, India, 11-12 December 2020; pp. 337-348.
30. Shorten, C.; Khoshgoftaar, T.M. A survey on image data augmentation for deep learning. *Journal of big data* **2019**, *6*, 1-48.
31. Moustakidis, S.; Siouras, A.; Papandrianos, N.; Ntakolia, C.; Papageorgiou, E. Deep learning for bone metastasis localisation in nuclear imaging data of breast cancer patients. In Proceedings of the 12th International Conference on Information, Intelligence, Systems & Applications (IISA), Chania Crete, Greece, 12-14 July 2021; pp. 1-8.
32. Jadon, S. A survey of loss functions for semantic segmentation. In Proceedings of the IEEE conference on computational intelligence in bioinformatics and computational biology (CIBCB), Via del Mar, Chile, 27-29 October 2020; pp. 1-7.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.