European Journal of Radiology 154 (2022) 110433

Contents lists available at ScienceDirect

European Journal of Radiology

journal homepage: www.elsevier.com/locate/ejrad

Visual and quantitative evaluation of microcalcifications in mammograms with deep learning-based super-resolution

Takashi Honjo [a,b], Daiju Ueda [b,c,*], Yutaka Katayama [d], Akitoshi Shimazaki [b], Atsushi Jogo [b], Ken Kageyama [b], Kazuki Murai [b], Hiroyuki Tatekawa [b], Shinya Fukumoto [e], Akira Yamamoto [b], Yukio Miki [b]

[a] Department of Diagnostic and Interventional Radiology, Graduate School of Medicine, Osaka City University, Osaka, Japan
[b] Department of Diagnostic and Interventional Radiology, Graduate School of Medicine, Osaka Metropolitan University, Osaka, Japan
[c] Smart Life Science Lab, Center for Health Science Innovation, Osaka Metropolitan University, Osaka, Japan
[d] Department of Radiology, Osaka Metropolitan University Hospital, Osaka, Japan
[e] Department of Premier Preventive Medicine, Graduate School of Medicine, Osaka Metropolitan University, Osaka, Japan

ARTICLE INFO

Keywords:
Breast Cancer
Mammography
Microcalcification
Deep Learning
Artificial Intelligence
Super Resolution

ABSTRACT

Purpose: To evaluate visually and quantitatively the performance of a deep-learning-based super-resolution (SR) model for microcalcifications in digital mammography.

Method: Mammograms were consecutively collected from 5080 patients who underwent breast cancer screening from January 2015 to March 2017. Of these, 93 patients (136 breasts, mean age, 50 ± 7 years) had microcalcifications in their breasts on mammograms. We applied an artificial intelligence model known as a fast SR convolutional neural network to the mammograms. SR and original mammograms were visually evaluated by four breast radiologists using a 5-point scale (1: original mammograms are strongly preferred, 5: SR mammograms are strongly preferred) for the detection, diagnostic quality, contrast, sharpness, and noise of microcalcifications. Mammograms were quantitatively evaluated using a perception-based image-quality evaluator (PIQE).

Results: All radiologists rated the SR mammograms better than the original ones in terms of detection, diagnostic quality, contrast, and sharpness of microcalcifications. These ratings were significantly different according to the Wilcoxon signed-rank test ($p < .001$), while the noise score of the three radiologists was significantly lower ($p < .001$). According to PIQE, SR mammograms were rated better than the original mammograms, showing a significant difference by paired *t*-test ($p < .001$).

Conclusion: An SR model based on deep learning can improve the visibility of microcalcifications in mammography and help detect and diagnose them in mammograms.

1. Introduction

Breast cancer accounts for 24.5% of all cancer cases and 15.5% of cancer deaths in women globally [1]. Mammography is the most common breast screening test that can effectively reduce breast cancer-related mortality by aiding appropriate diagnosis and timely treatment [2]. One of the critical mammogram signs in breast cancer is the finding of microcalcifications. They can be observed on mammograms in 30%–50% of breast cancers but are often smaller than approximately 500 μm [3–5]. Microcalcifications are the most indistinct image findings that indicate breast cancer. Because detecting such subtle lesions is a big burden for radiologists, techniques improving visibility of microcalcifications could be helpful in clinical practice.

Recently, the application of convolutional neural networks (CNNs), a field of deep learning (DL) [6,7], has led to dramatic improvements in radiology [8,9]. A DL-based super-resolution (SR) technique can

Abbreviations: MLO, medio-lateral oblique; DL, deep learning; SR, super-resolution; CNN, convolutional neural network; FSRCNN, Fast Super-Resolution Convolutional Neural Network; SRCNN, Super-Resolution Convolutional Neural Network; SSIM, structural similarity; PSNR, peak signal-to-noise ratio; BI-RADS®, Breast Imaging Reporting and Data System; PIQE, perception-based image-quality evaluator.

* Corresponding author at: Smart Life Science Lab, Center for Health Science Innovation, Osaka Metropolitan University, Osaka, Japan.
E-mail address: ai.labo.ocu@gmail.com (D. Ueda).

https://doi.org/10.1016/j.ejrad.2022.110433
Received 29 March 2022; Received in revised form 30 May 2022; Accepted 30 June 2022
Available online 6 July 2022

estimate high-resolution output images from low-resolution input images. Most SR models increase resolution using acquired features between the original and low-resolution image. These images are then quantified by metrics such as structural similarity (SSIM) and peak signal-to-noise ratio (PSNR), and various studies have demonstrated the extent to which the resolution of the original images can be restored by applying the SR model [10,11].

In this study, we hypothesized that adapting the SR model to original mammograms may improve the visibility of microcalcifications. Full-reference image quality metrics such as SSIM and PSNR are not available because this SR mammogram has no ground truth image for reference. So we evaluated the SR mammogram by visual assessment of four breast radiologists and by perception-based image-quality evaluator (PIQE) [12], one of the non-reference image-quality metrics. PIQE has been used to assess medical images as a quantitative metric [13].

To the best of our knowledge, no study has evaluated SR mammograms by visual rating or PIQE. This study aimed to assess visually and quantitatively SR mammograms based on deep learning.

2. Materials and methods

2.1. Datasets

Mammograms were collected retrospectively from April 2014 to March 2017 from patients who underwent mammography screening at MedCity21, an advanced medical center for preventive medicine established by Osaka City University Hospital. We extracted all mammograms with all types of microcalcifications reported by breast radiologists. As the study included patients who visited the institution for the first time, none of the datasets overlapped. Both left and right mediolateral oblique and craniocaudal images were collected if available. All examinations were performed using a single dedicated mammography system (Selenia Dimensions 2D/3D; Hologic Inc. Marlborough, MA, USA). All mammograms had the same pixel size. Fig. 1 shows a flowchart of patient selection. This study complies with the declaration of Helsinki. The ethics board at our institution comprehensively reviewed and approved the study protocol. Since the mammograms were acquired during daily clinical practice, the need for informed consent was waived by the ethics board.

2.2. Super resolution model

We used a fast SRCNN (FSRCNN), one of the primary SR models [14,15] and trained it by ImageNet dataset (non-medical images). The FSRCNN is a redesigned SRCNN that achieves a faster and better structure. It includes five major CNN steps, which reduce each layer's size compared to SRCNN. Fig. 2 shows the FSRCNN architecture. To reduce smoothing effects, we modified FSRCNN to apply 2 × 2 pixel binning to generate low-resolution images in training. The model was trained for 100 epochs, and the learning parameters were adopted at the lowest value of the validation loss function, the mean squared error. Adam [16], widely used as a learning rate optimizer, was run.

As above, FSRCNN is made of five parts: the feature extraction, shrinking, mapping, expanding, and deconvolution layers. First, FSRCNN performs feature extraction by 5 × 5 layer on the original low-resolution image without interpolation. Then, as a shrinking step, the transformation reduces the feature maps. Third, as a nonlinear mapping, this model replaces the 1 × 1 layer with multiple 3 × 3 layers. Next, as the expanding layer, a 1 × 1 layer is executed to increase the number of feature maps. The final part is a deconvolution layer that reconstructs the image using a 9 × 9 filter. From shrinking step to expanding step, narrow layers instead of a single wide layer can reduce the number of operations, achieving a faster model than SRCNN. Deconvolution is an upsampling technique by the inversed convolution, which is known to yield higher PSNR. We utilized the highest performance parameters in the original model [15]. The activation function after each convolution layer was the Parametric Rectified Linear Unit (PReLU). The detailed code is available in MATLAB version 9.6 (MATLAB; The MathWorks Inc, Natick, MA, USA).

2.3. Visual assessment

We applied the FSRCNN model to all mammograms included according to eligibility. These SR mammograms were evaluated independently by four board-certified breast radiologists (reader 1, 7 years of experience; reader 2, 15 years of experience; reader 3, 14 years of experience; and reader 4, 12 years of experience in breast radiology). The images were anonymized during extraction, and no patient information was displayed on the images. We defined the original mammogram as the reference and the SR image as the target mammogram. Then radiologists compared the target mammogram with the reference mammogram based on 14 criteria: detection of calcifications, diagnostic quality of calcifications, visibility of breast density, contrast between calcifications and surrounding tissue, contrast in general images, noise of calcifications, noise in general, sharpness of calcifications, sharpness in general, pectoral muscle visibility, nipple visibility, skin visibility, artifacts, and overall image quality.

The definitions of main criteria are as follows: contrast is defined as the spread of the distribution of each image's pixel values in computer vision. The image's contrast increases when the difference between the largest and smallest pixels is vast. Noise is a component introduced irregularly during the input or processing of images and appears as fluctuations in pixel values. When noise increases in an image, it produces more unwanted signals, such as random variations in color or

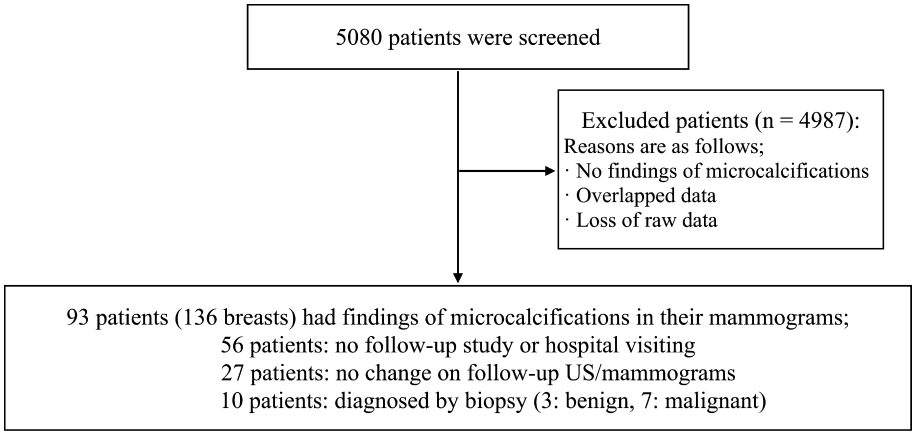

Fig. 1. Flowchart of patient selection and demographics. US – ultrasound.

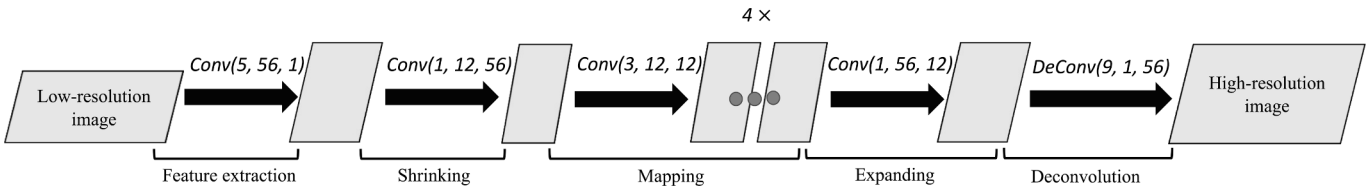

Fig. 2. Structure of the super-resolution model This figure demonstrates the network structures of FSRCNN. A convolutional layer is denoted as *Conv (f, n, c)*, where *f, n, and c* represent the filter size, number of filters, and number of channels, respectively. *Deconv* indicates the deconvolutional layer. In the shrinking step, the number of feature maps was reduced from *56* to *12*. In the expanding layer, the number of feature maps increased from *12* to *56*. There were four mapping layers. In the training of our model, low-resolution images were generated through 2 × 2 pixel binning.

brightness. Sharpness is the degree of change in luminance near the edge. When this change is abrupt, the image is sharp or appears blurry otherwise.

In addition, we defined the detection of microcalcifications is how easily the radiologists notice microcalcifications. The definition of the diagnostic quality of microcalcifications is how easily each reader diagnoses the microcalcifications. The items such as the visibility of breast density, pectoral muscles, and nipples mean whether each soft tissue has good visibility or not. Artifact is whether artifacts in the image processing are visible or not. Overall scores tell whether the whole image is preferable or not.

Each criterion was rated on a 5-point numerical rating scale (from 1 to 5), where 1 indicates that the conventional mammogram was strongly preferred, 5 that the SR mammogram was strongly preferred, and 3 that there was no difference between the images. The evaluation was performed using a 10-megapixel mammography display (RadiForce GX1030; EIZO, Tokyo, Japan). Readers could adjust the window and level settings, zoom, and pan. A comment field was available for each reader to report pseudo-lesions or imaging problems.

2.4. Quantitative assessment

We used perception-based image quality evaluator (PIQE) as a quantitative metric [12]. This index ranges from 0 to 100, where 100 is lowest worst score. PIQE is a no-reference image quality score by measuring the local variance of the perceived distorted blocks in an image. The PIQE algorithm works by estimating the local variance of in-block distortions in an image or pooling the estimated block-level distortions for overall image quality evaluation. The level of distortion is calculated using an average subtraction contrast normalization factor coefficient. Then a threshold is applied to the estimated distortion level to classify the block as distorted or undistorted. As the average of the scores from the distorted block, PIQE is obtained. It is an unsupervised method that does not require a trained model. We calculated the PIQE for both original and SR mammograms and analyzed them.

2.5. Statistical analysis

The Wilcoxon's signed-rank test was performed to assess the significance of the differences between visual ratings in reference to a similar study using a 5-point scale [17]. For the null hypothesis, a representative score was defined as 3, and the difference was significant the further away each reader's score was from it. A paired *t*-test was performed to compare the PIQE values between the SR and original mammograms. The null hypothesis is no difference in PIQE scores between each image. All the statistical tests were two-sided with a significance level of 0.05 and performed using R version 3.6.0 (R; R Foundation for Statistical Computing, Vienna, Austria).

3. Results

3.1. Patient and mammography characteristics

Among the 5080 patients who underwent breast cancer screening, we extracted 93 patients (136 breasts; mean age, 50 ± 7 years) with findings of microcalcifications and suspicious morphology in the reporting system (Table 1). Microcalcifications in the bilateral breasts were observed in 43 patients. Ten patients were biopsied, and 7 were diagnosed with breast cancer (Fig. 1). One patient had fatty, 17 had scattered, 73 had heterogeneously dense, and 2 had extremely dense breasts. Regarding the morphology of calcifications, 56 were amorphous, 31 were fine pleomorphic, 5 were coarse heterogeneous, and 1 was fine linear. Regarding distribution, 23 were categorized as diffused, 48 as regional, 19 as grouped, 2 as segmental, and 1 as linear calcification. Referring to the Breast Imaging Reporting and Data System (BI-RADS®) [18], none was classified as category 1, 49 as category 2, 42 as category 3, 1 as category 4, and 1 as category 5. According to the reporting system, we tabulated all findings of mammograms.

3.2. Visual and quantitative assessment

Table 2 summarizes the visual results for each evaluation criterion. All radiologists scored > 3 points for detection, diagnostic quality, contrast, and sharpness of microcalcifications. These criteria were

Table 1
Patient demographics

Characteristics	Value
No. patients	93
Mean age ± SD, years	50 ± 7.6
No. patients with only MLO view	11
No. breasts with microcalcifications	136
Morphology of calcification	
Amorphous	56
Fine pleomorphic	31
Coarse heterogeneous	5
Fine linear	1
Distribution of calcification	
Diffused	23
Grouped	19
Regional	48
Segmental	2
Linear	1
BI-RADS category	
1	0
2	49
3	42
4	1
5	1
Mammary glands density	
Almost entirely fatty	1
Scattered dense	17
Heterogeneously dense	73
Extremely dense	2

SD, standard deviation; MLO, mediolateral oblique; BI-RADS, Breast Imaging Reporting and Data System.

Table 2
Visual assessments for original and SR mammograms[a, b] (n = 136 breasts)

	1	2	3	4	5	mean	p value	
Detection of microcalcifications								
Reader1	0	0	31	55	50	4.140	<0.001	***
Reader2	0	0	1	135	0	3.993	<0.001	***
Reader3	0	0	1	135	0	3.993	<0.001	***
Reader4	0	0	63	73	0	3.537	<0.001	***
Diagnostic quality of microcalcifications								
Reader1	0	2	21	60	53	4.206	<0.001	***
Reader2	0	0	3	133	0	3.978	<0.001	***
Reader3	2	3	2	129	0	3.897	<0.001	***
Reader4	0	0	105	31	0	3.228	<0.001	***
Visibility of breast density								
Reader1	0	15	100	17	4	3.074	0.143	
Reader2	0	0	136	0	0	3.000	1.000	
Reader3	0	9	0	127	0	3.868	<0.001	***
Reader4	0	0	136	0	0	3.000	1.000	
Contrast of microcalcifications								
Reader1	0	3	13	51	69	4.368	<0.001	***
Reader2	0	0	2	0	134	4.971	<0.001	***
Reader3	2	3	2	129	0	3.897	<0.001	***
Reader4	0	0	58	78	0	3.574	<0.001	***
Contrast in general								
Reader1	0	0	30	76	30	4.000	<0.001	***
Reader2	0	1	0	135	0	3.985	<0.001	***
Reader3	0	0	0	136	0	4.000	<0.001	***
Reader4	0	0	135	1	0	3.007	1.000	
Noise of microcalcifications								
Reader1	1	53	75	7	0	2.647	<0.001	***
Reader2	0	136	0	0	0	2.000	<0.001	***
Reader3	0	131	0	1	4	2.103	<0.001	***
Reader4	0	2	133	1	0	2.993	0.773	
Noise in general								
Reader1	0	31	97	8	0	2.831	<0.001	***
Reader2	0	132	4	0	0	2.029	<0.001	***
Reader3	0	136	0	0	0	2.000	<0.001	***
Reader4	0	96	40	0	0	2.294	<0.001	***
Sharpness of microcalcifications								
Reader1	0	1	10	63	62	4.368	<0.001	***
Reader2	0	2	2	132	0	3.956	<0.001	***
Reader3	0	7	0	129	0	3.897	<0.001	***
Reader4	0	1	37	98	0	3.713	<0.001	***
Sharpness in general								
Reader1	0	0	20	85	31	4.081	<0.001	***
Reader2	0	0	136	0	0	3.000	1.000	
Reader3	0	0	0	136	0	4.000	<0.001	***
Reader4	0	0	134	2	0	3.015	0.346	
Visibility of pectoral muscles								
Reader1	14	63	46	12	1	2.434	<0.001	***
Reader2	0	0	128	0	0	3.000	1.000	
Reader3	0	136	0	0	0	2.000	<0.001	***
Reader4	0	0	130	0	0	3.000	1.000	
Visibility of nipples								
Reader1	32	91	10	3	0	1.882	<0.001	***
Reader2	0	5	131	0	0	2.963	<0.05	*
Reader3	0	136	0	0	0	2.000	<0.001	***
Reader4	0	0	136	0	0	3.000	1.000	
Visibility of skin								
Reader1	4	105	26	1	0	2.176	<0.001	***
Reader2	0	0	136	0	0	3.000	1.000	
Reader3	0	136	0	0	0	2.000	<0.001	***
Reader4	0	0	136	0	0	3.000	1.000	
Artifacts								
Reader1	0	7	129	0	0	2.949	<0.05	*
Reader2	0	0	136	0	0	3.000	1.000	
Reader3	0	0	136	0	0	3.000	1.000	
Reader4	0	0	136	0	0	3.000	1.000	
Overall scores								
Reader1	0	1	30	59	46	4.103	<0.001	***
Reader2	0	0	130	6	0	3.044	<0.05	*
Reader3	0	3	9	124	0	3.890	<0.001	***
Reader4	0	0	135	1	0	3.007	1.000	

[a] Score based on a 5-point scale (1 = original images strongly preferred, 2 = original images somewhat preferred, 3 = no preference, 4 = SR images somewhat preferred, 5 = SR images strongly preferred). Significance level: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

[b] Readers 2 and 4 judged the visibility of pectoral muscles for eight and six breasts, respectively, to be unreadable (NA = 8, 6).

significantly different according to the Wilcoxon signed-rank test ($p < .001$). The overall score of three of the four readers was also significantly higher ($p < .05$). The score for microcalcification noise was significantly low ($p < .001$). Typical structures, such as the skin, pectoral muscles, and nipples, were assigned low scores by some readers. There was a slight difference in the artifacts' scores. No pseudo lesion was reported. There was no marked correlation between the years of experience of the readers and their evaluation tendencies. Fig. 3 shows examples of images rated higher overall, whereas Fig. 4 shows examples of images rated lower by some readers. Each figure suggests that while the SR images had higher contrast, the noise increased, especially in the enlarged images.

The mean PIQE of the original mammograms was 12.4 ± 5.0, and the PIQE of the SR mammograms was 3.2 ± 1.3, showing a significant difference by paired *t*-test ($p < .001$). Table 3 summarizes the PIQE results.

4. Discussion

In this study, we created SR mammograms by applying a DL-based SR model to original mammograms. Then we evaluated them visually and quantitatively. SR mammograms were visually superior to original mammograms in terms of primary criteria, except for noise. PIQE value of SR mammograms was significantly better than that of the original mammograms.

Notably, all readers scored ≥ 3 for the detection of microcalcifications on SR mammograms. It indicates that the method can increase the reader's detectability for microcalcifications in clinical practice. Diagnostic quality, contrast, and sharpness of microcalcifications were higher for SR mammograms than for original mammograms. It was probably because the SR model clearly defined the microcalcification edges, making it easier to distinguish them from the surrounding tissue. However, the noise rating was low for both microcalcifications and the entire image.

To the best of our knowledge, this is the first study to visually evaluate mammogram microcalcifications using a DL-based SR model. SR models on medical images have been previously reported [19,20]. PSNR and SSIM were often used as evaluation indices [21,22]. These were calculated using both reference images and images created by the SR models. Although the SSIM and PSNR are appropriate for overall image evaluation, they cannot be calculated without reference images. In addition, they are not suitable for local evaluation. To compensate for these shortcomings, we should rate microcalcification visually and calculate PIQE as a quantitative value. PIQE is an unsupervised method and non-referenced image quality metrics. SR mammograms showed lower PIQE, indicating the better quality of the image. It uses the mean subtracted contrast normalized coefficient to calculate scores of block-by-block distortions in images. Thus, PIQE may be consistent with visual image rating because it focuses on the local features, imitating human behavior [12].

Next, the SR model based on CNN inputs the low-frequency component image and estimates the high-frequency component image. The SRCNN uses the bicubic method to create low-resolution images [14]. However, this mechanism has room for improvement. The bicubic method can cause an undesirable smoothing effect. To suppress it, we used FSRCNN [15] without linear interpolation. In training, 2×2 binning was performed to generate low-resolution images. Binning is a

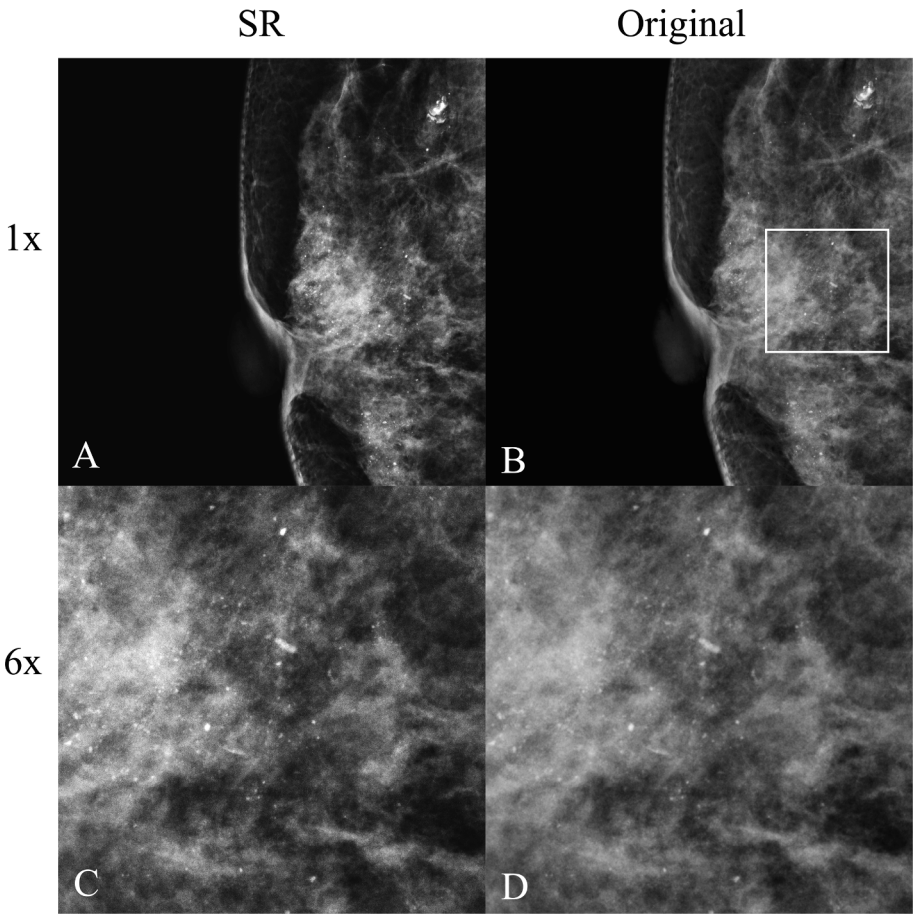

Fig. 3. Examples of highly evaluated mammograms Super-resolution mammograms (A, C) and original mammograms (B, D) of a 48-year-old woman. Diffuse punctate calcifications were classified as BI-RADS category 2. The area circumscribed by the square in B is magnified by six times in C and D. Original mammograms show diffuse calcifications, mainly amorphous or fine pleomorphic. SR mammograms show a more evident contrast between calcifications and surrounding tissue.

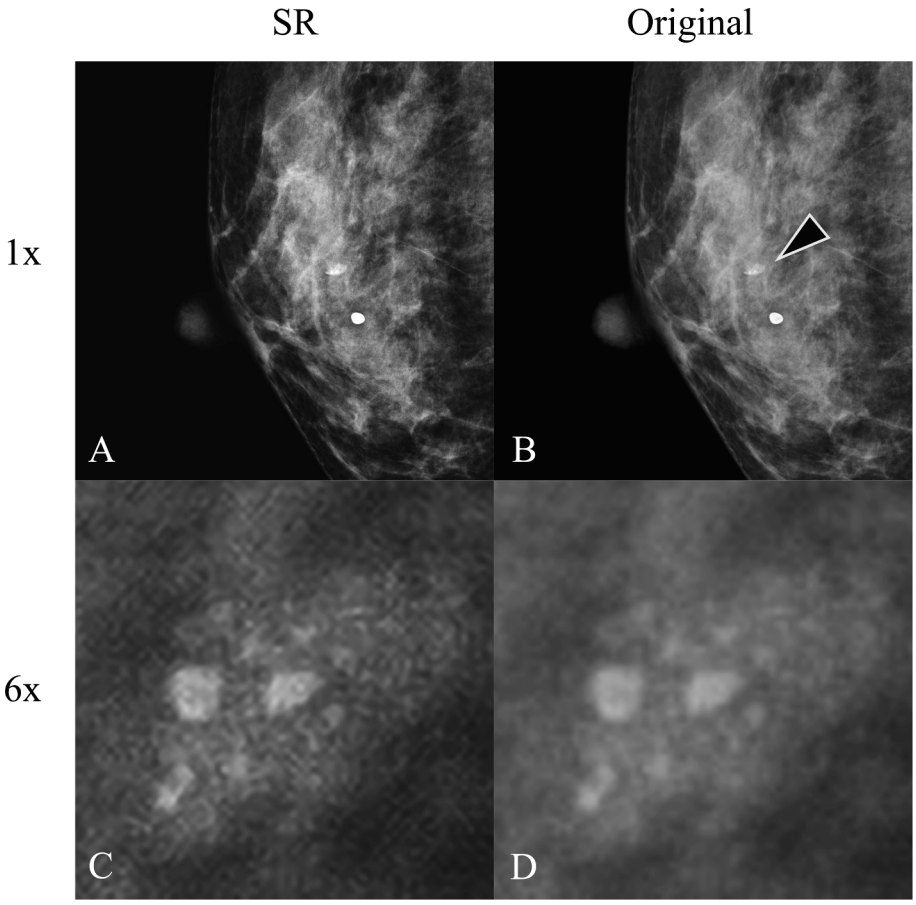

Fig. 4. Examples of low rated mammograms Super-resolution mammograms (A, C) and original mammograms (B, D) of a 44-year-old woman. Grouped and fine pleomorphic calcifications were classified as BI-RADS category 3. The area indicated by the arrowhead in B is magnified by six times in C and D. Contrast was more evident in the SR images, but the noise was noticeable in the post-processed images. The patient was followed up only with ultrasonography.

Table 3
PIQE assessments for the original and SR mammograms

	Original mammograms	SR mammograms
Total	12.36 ± 5.03	3.2 ± 1.33[a]
Mammary glands density		
Almost entirely fatty	12.93	6.25
Scattered dense	13.16 ± 4.3	2.68 ± 0.98
Heterogeneously dense	11.81 ± 4.79	3.33 ± 1.35
Extremely dense	22.81 ± 2.63	1.8 ± 0.33
Morphology of calcification		
Amorphous	13.07 ± 5.21	2.97 ± 1.21
Fine pleomorphic	11.52 ± 4.59	3.54 ± 1.47
Coarse heterogeneous	8.01 ± 2.68	4.4 ± 1.36
Fine linear	10.82 ± 0.76	2.55 ± 0.34
Distribution of calcification		
Diffused	12.18 ± 4.28	3.22 ± 1.43
Grouped	11.79 ± 4.8	3.42 ± 1.3
Regional	13.49 ± 5.84	2.86 ± 1.25
Segmental	10.57 ± 2.4	3.17 ± 1.03
Linear	8.27 ± 0.97	3.87 ± 1.43

PIQE, perception-based image-quality evaluator; SR, super-resolution.
[a] It showed a significant difference ($p < .001$).

non-linear technique for merging adjacent pixels into one larger pixel and can reduce the number of pixels and facilitate image processing. Besides, binning increases the signal-to-noise ratio to provide better contrast [23]. It may have increased the contrast score and decreased the noise rating.

Moreover, this SR model can be applied to other radiological images, such as chest radiographs, CT images, and MR images, because it was trained with general images and not medical ones. The SR model described here does not require any changes in the mammography equipment. Therefore, another advantage is that this model can be implemented directly in the picture archiving and communication systems. Although it was possible to utilize down-sampled mammograms in training, we chose merits that we could prepare for sufficient data and adapt our model to other modalities avoiding overfitting. This technology may ease the strain on radiologists because interpreting tiny microcalcifications is a heavy task for radiologists. This burden can cause errors with $\geq$ 10% of detectable cancers missed [24], and misdiagnosis of the mammography is the most significant cause of lawsuits in radiology in the US [25]. Our SR model can process a mammogram within a minute, minimizing the reader's burden with dense breasts where calcifications are challenging to detect. This model is also applicable to non-calcified lesions. While mammograms with masses, asymmetries, and distortions can show better resolution by the SR model, it may be the most effective for indistinct microcalcifications.

This study has several limitations. First, whether the SR model can help radiologists determine a diagnosis closer to the pathological diagnosis has not yet been investigated. Furthermore, whether calcifications are more easily detected in a prospective clinical breast cancer screening trial remains unknown. Additionally, overdiagnosis and overtreatment may occur if microcalcifications are easily detected. We believe a study that more closely mimics a clinical setting should be undertaken. Moreover, although each reader was not informed about the differences between SR and original mammograms, they may have distinguished them by resolution, which may have introduced bias that affected the results. In addition, mammograms were obtained at a single institution. Our SR model trained by general images is adjustable regardless of

manufacturer, filter, X-ray tube, or dose. However, future multi-center studies are warranted. Regarding patient demographics, the most common calcification morphologies were amorphous and fine pleomorphic. Only a few patients had segmental calcification. Thus, further research with a larger patient population is necessary.

5. Conclusions

We applied the FSRCNN model to mammography with calcifications. Then we visually and quantitatively compare pre -and post-processed images. Results suggest that our model can help detect and diagnose microcalcifications in mammograms.

Funding

This work was supported by the Japan Society for the Promotion of Science KAKENHI Grant No. JP20K16734.

Acknowledgments

We thank MedCity21 for joining our study.

References

[1] H. Sung, J. Ferlay, R.L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, F. Bray, GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA Cancer J. Clin. 71 (3) (2021) 209–249.

[2] S.W. Duffy, L. Tabár, A.M.F. Yen, P.B. Dean, R.A. Smith, H. Jonsson, S. Törnberg, S. L.S. Chen, S.Y.H. Chiu, J.C.Y. Fann, M.M.S. Ku, W.Y.Y. Wu, C.Y. Hsu, Y.C. Chen, G. Svane, E. Azavedo, H. Grundström, P. Sundén, K. Leifland, E. Frodis, J. Ramos, B. Epstein, A. Åkerlund, A. Sundbom, P. Bordás, H. Wallin, L. Starck, A. Björkgren, S. Carlson, I. Fredriksson, J. Ahlgren, D. Öhman, L. Holmberg, T.H.H. Chen, Mammography screening reduces rates of advanced and fatal breast cancers: results in 549,091 women, Cancer. 126 (2020) 2971–2979, https://doi.org/10.1002/cncr.32859.

[3] R.R. Millis, R. Davis, A.J. Stacey, The detection and significance of calcifications in the breast: a radiological and pathological study, Br. J. Radiol. 49 (1976) 12–26, https://doi.org/10.1259/0007-1285-49-577-12.

[4] J.N. Wolfe, Analysis of 462 breast carcinomas, Am. J. Roentgenol. Radium Ther. Nucl. Med. 121 (1974) 846–853, https://doi.org/10.2214/ajr.121.4.846.

[5] W.A. Murphy, K. DeSchryver-Kecskemeti, Isolated clustered microcalcifications in the breast: Radiologic-pathologic correlation, Radiology. 127 (1978) 335–341, https://doi.org/10.1148/127.2.335.

[6] G. Hinton, Deep learning-a technology with the potential to transform health care, JAMA. 320 (2018) 1101–1102, https://doi.org/10.1001/jama.2018.11100.

[7] Y. Lecun, Y. Bengio, G. Hinton, Deep learning, Nature. 521 (2015) 436–444, https://doi.org/10.1038/nature14539.

[8] K. Yasaka, H. Akai, A. Kunimatsu, S. Kiryu, O. Abe, Deep learning with convolutional neural network in radiology, Jpn. J. Radiol. 36 (2018) 257–272, https://doi.org/10.1007/s11604-018-0726-3.

[9] D. Ueda, A. Shimazaki, Y. Miki, Technical and clinical overview of deep learning in radiology, Jpn. J. Radiol. 37 (2019) 15–33, https://doi.org/10.1007/s11604-018-0795-3.

[10] W. Yang, X. Zhang, Y. Tian, W. Wang, J.H. Xue, Q. Liao, Deep learning for single image super-resolution: a brief review, IEEE Trans. Multimedia. 21 (2019) 3106–3121, https://doi.org/10.1109/TMM.2019.2919431.

[11] S. Park, H.M. Gach, S. Kim, S.J. Lee, Y. Motai, Autoencoder-inspired convolutional network-based super-resolution method in MRI, IEEE J. Transl. Eng. Health Med. 9 (2021) 1800113, https://doi.org/10.1109/JTEHM.2021.3076152.

[12] N. Venkatanath, D. Praneeth, B.h. Maruthi Chandrasekhar, S.S. Channappayya, S. S. Medasani, Blind image quality evaluation using perception based features, in: 2015 Conference on Communications (NCC), First National, 2015, pp. 1–6.

[13] S. Higashiyama, Y. Katayama, A. Yoshida, N. Inoue, T. Yamanaga, H. Kawahata, T. Ichida, Y. Miki, J. Kawabe, Usefulness of a No-Reference Metric for Evaluation of Images in Nuclear Medicine -A Comparative Study with Visual Assessment. https://doi.org/10.21203/rs.3.rs-372416/v1, 2021.

[14] C. Dong, C.C. Loy, K. He, X. Tang, Learning a deep convolutional network for image super-resolution, computer vision – ECCV, Springer International Publishing 2014 (2014) 184–199.

[15] C. Dong, C.C. Loy, X. Tang, in: Accelerating the Super-Resolution Convolutional Neural Network, in: Eur, Springer, Cham, 2016, pp. 391–407.

[16] D.P. Kingma, J. Ba, Adam: A Method for Stochastic Optimization, arXiv Preprint ArXiv:1412.6980. 2014 Dec 22.

[17] S.Y. Ahn, K.J. Chae, J.M. Goo, The potential role of grid-like software in bedside chest radiography in improving image quality and dose reduction: an observer preference study, Korean J. Radiol. 19 (2018) 526–533, https://doi.org/10.3348/kjr.2018.19.3.526.

[18] E. Sickles, C.J. D'Orsi, L.W. Bassett, et al., ACR BI-RADS® mammography. ACR Bi-RADS®Atlas, Breast Imaging Reporting and Data System, American College of Radiology, Reston, Virginia, 2013.

[19] A.S. Chaudhari, K.J. Stevens, J.P. Wood, A.K. Chakraborty, E.K. Gibbons, Z. Fang, A.D. Desai, J.H. Lee, G.E. Gold, B.A. Hargreaves, Utility of deep learning super-resolution in the context of osteoarthritis MRI biomarkers, J. Magn. Reson. Imaging. 51 (2020) 768–779, https://doi.org/10.1002/jmri.26872.

[20] W. Tan, P. Liu, X. Li, Y. Liu, Q. Zhou, C. Chen, Z. Gong, X. Yin, Y. Zhang, Classification of COVID-19 pneumonia from chest CT images based on reconstructed super-resolution images and VGG neural network, Health Inf. Sci. Syst. 9 (2021) 10, https://doi.org/10.1007/s13755-021-00140-0.

[21] K. Umehara, J. Ota, T. Ishida, Super-resolution imaging of mammograms based on the super-resolution convolutional neural network,, open J. Med. Imaging. 07 (2017) 180–195, https://doi.org/10.4236/ojmi.2017.74018.

[22] K. Umehara, J. Ota, T. Ishida, Application of super-resolution convolutional neural network for enhancing image resolution in chest CT, J. Digit. Imaging. 31 (2018) 441–450, https://doi.org/10.1007/s10278-017-0033-z.

[23] X. Jin, K. Hirakawa, Analysis and processing of pixel binning for color image sensor, EURASIP J. Adv. Signal Process. 2012 (2012) 125, https://doi.org/10.1186/1687-6180-2012-125.

[24] E. Banks, G. Reeves, V. Beral, D. Bull, B. Crossley, M. Simmonds, E. Hilton, S. Bailey, N. Barrett, P. Briers, R. English, A. Jackson, E. Kutt, J. Lavelle, L. Rockall, M.G. Wallis, M. Wilson, J. Patnick, Influence of personal characteristics of individual women on sensitivity and specificity of mammography in the million women study: cohort study, BMJ. 329 (2004) 477, https://doi.org/10.1136/bmj.329.7464.477.

[25] J.S. Whang, S.R. Baker, R. Patel, L. Luk, A. Castro, The causes of medical malpractice suits against radiologists in the United States, Radiology. 266 (2) (2013) 548–554.