High-resolution imaging in acoustic microscopy using deep learning

MACHINE LEARNING
Science and Technology

OPEN ACCESS

RECEIVED
13 May 2023

REVISED
16 November 2023

ACCEPTED FOR PUBLICATION
8 January 2024

PUBLISHED
18 January 2024

PAPER

High-resolution imaging in acoustic microscopy using deep learning

Pragyan Banerjee[1], Shivam Milind Akarte[2], Prakhar Kumar[3], Muhammad Shamsuzzaman[4], Ankit Butola[4], Krishna Agarwal[4], Dilip K Prasad[5], Frank Melandsø[4] and Anowarul Habib[4,*]

[1] Department of Mathematics, Indian Institute of Technology Guwahati, Guwahati, Assam 781039, India
[2] Department of Mechanical Engineering, Birla Institute of Technology and Science, Pilani Hyderabad Campus, Hyderabad, Telangana 500078, India
[3] Department of Electronics Engineering, Indian Institute of Technology (ISM), 826004 Dhanbad, India
[4] Department of Physics and Technology, UiT The Arctic University of Norway, 9037 Tromsø, Norway
[5] Department of Computer Science, UiT The Arctic University of Norway, 9037 Tromsø, Norway
* Author to whom any correspondence should be addressed.

E-mail: anowarul.habib@uit.no

Keywords: acoustic imaging, scanning acoustic microscopy, high-resolution imaging, machine learning, transfer learning.

Abstract

Acoustic microscopy is a cutting-edge label-free imaging technology that allows us to see the surface and interior structure of industrial and biological materials. The acoustic image is created by focusing high-frequency acoustic waves on the object and then detecting reflected signals. On the other hand, the quality of the acoustic image's resolution is influenced by the signal-to-noise ratio, the scanning step size, and the frequency of the transducer. Deep learning-based high-resolution imaging in acoustic microscopy is proposed in this paper. To illustrate four times resolution improvement in acoustic images, five distinct models are used: SRGAN, ESRGAN, IMDN, DBPN-RES-MR64-3, and SwinIR. The trained model's performance is assessed by calculating the PSNR (Peak Signal to Noise Ratio) and SSIM (Structural Similarity Index) between the network-predicted and ground truth images. To avoid the model from over-fitting, transfer learning was incorporated during the procedure. SwinIR had average SSIM and PSNR values of 0.95 and 35, respectively. The model was also evaluated using a biological sample from Reindeer Antler, yielding an SSIM score of 0.88 and a PSNR score of 32.93. Our framework is relevant to a wide range of industrial applications, including electronic production, material micro-structure analysis, and other biological applications in general.

1. Introduction

High-resolution (HR) acoustic images can be used to facilitate biomedical or materials research, to investigate, measure, or determine the mechanical or bio-mechanical properties of the samples. Scanning acoustic microscope (SAM) also provides abundant and quantitative information about the objects under inspection. The capabilities of SAM include noninvasive micro-structural characterization of materials, characterization of surface and subsurface mechanical properties of piezoelectric materials, structural health monitoring of composite structures, surface defects on polymer circuits, and studies of anisotropic phonon propagations [1–7]. In HR SAM, 200 nm has been achieved by using a 4.4 GHz acoustic transducer [8]. The application of Acoustic Microscopy prevails in many areas, such as medical imaging (inspection of bones or internal structures like articular cartilage) and inspection of manufactured products like electronic chips and circuits [4, 9, 10]. Mainly, SAM utilizes a wide range of frequencies (10 MHz to 1.2 GHz) to produce visible images of the surface or sub-surfaces of an object without damaging the samples. Furthermore, by thoroughly studying the internal layers and the structures of these objects, we can analyze and detect defects efficiently. Microelectronics and semiconductor industries are a demanding and highly competitive market. SAM plays a vital role in the development of the improved molded for the flip chip packages. It is also capable of dealing with the complexity of miniaturized assemblies such as chip-scale packages and 3D IC stacks.

However, because acoustic microscopes scan things point-by-point (pixel-by-pixel), it takes a long time to conduct a thorough scan of even a tiny sample. A HR image needs a high-frequency transducer and a smaller step size during data acquisition. Because low-resolution (LR) ultrasound data needs scanning fewer spots on the object, the time required to execute such a scan will be dramatically reduced. Thus, we can obtain high-resolution (HR) ultrasound images with minimal effort by employing the super-resolution model. The patterns in naturally occurring images are quite perceptible to deep learning-based algorithms. Hence, using deep neural networks in computer vision is common and is also found to produce better results than shallow networks. However, when it comes to ultrasound imaging, there are not quite obvious patterns in the data. Hence, we believe very deep neural networks would not help us in our purpose.

With the rise of new technologies and the ability to generate high-quality images, old methods of generating images are becoming obsolete. Hence a large amount of research is being carried out on HR imaging in multiple domains [11–15]. In each of these works, authors tried to apply super-resolution techniques in solving real-world problems, specially in the bio-medical field. Another such example is brain magnetic resonance images (MRI) [16]. Because of restrictions such as patient comfort and extensive sample period, a typical MRI picture lacks appropriate resolution. As a result, a few researchers suggested super-resolution of brain magnetic resonance imaging using autoencoders, an unsupervised neural network that involves scanning fewer points on the object, substantially reducing the time required to execute such a scan [17, 18]. Another research in MRI proposed edge-enhanced super-resolution generative adversarial networks (EE-SRGAN) for MRI super-resolution in slice-select direction [19] because a LR MRI in slice direction leads to information loss and hence improper diagnosis.

This work aims to demonstrate HR imaging from LR acoustic images using deep learning. There has been a previous work on high resolution in SAM imaging using deep learning [20]. Its authors used a four-layered U-NET-inspired architecture. The authors were able to achieve a peak signal to noise ratio (PSNR) score of 28.4 and an NRMSE score of 0.05. On the other hand, our work achieves better results in the PSNR score, and we did not use NRMSE since it does not incorporate structural fidelity which structural similarity index (SSIM) incorporates. Also, instead of using smooth L_1 loss, a combination of pixel loss, GAN loss, and perceptual loss was used to improve the quality of the output images. Also, in the work [21], the authors used the SR-net architecture to achieve a two-fold image super-resolution. Their architecture, on the other hand, is primarily made up of CNN layers. Some of the models we investigated used CNNs, but it is clear that transformer-based approaches have recently outperformed CNNs in many cases. Our article explores the use of the recently popular transformer [22] networks that have the ability to learn to focus on important image regions by exploring the global interactions between different regions. Due to their better performance [23, 24], it can also be used for image restoration. Swin transformers [24], have the advantage of CNN in processing images of large size, because of the local attention mechanism. It also has the benefit of the transformer model's long-term dependency on the shifted window scheme. Figure 1 demonstrates the overall strategy applied in this paper.

The remainder of the article is organized as follows: the deep learning approach has been mentioned in section 2. Section 3 contains a brief description of the material and methods used in the experiments. The results of the approaches can be found in section 4. Section 5 concludes the article and section 6 provides a brief future direction of the presented work.

2. Deep learning approach

2.1. Original data-set and curation for supervised learning

The ultrasound images of 17 different coins were included in the data sets. Because a small data set with many similarities is unsuitable for training the DNN, coins from different countries were used to introduce variation into the data set, as coins from the same country have similar patterns. The dimension of each image is 400×400 pixel and the pixel size ($50\,\mu$m) is kept constant for both transducers (20 and 50 MHz). All coins have a circular shape. Each image is cropped into multiple images to create diversity and hence more robust training of the network. Cropping is done by starting from the top-left corner of the image and striding by 48 pixels in the x-direction and y-direction. The resulting cropped image is of the dimension 120×120. After cropping, the final coin data set is made of 850 images, where 800 images are used for the training of the network and 50 are reserved for validation.

2.2. Architecture and training parameters

SRGAN [25], ESRGAN [26], IMDN [27, 28], DBPN-RES-MR64-3 [29] and SwinIR [30] models are used for both training and testing. The following section contains information on the training hyperparameters. SwinIR outperforms the others when it comes to generating HR images and comparing them to the ground

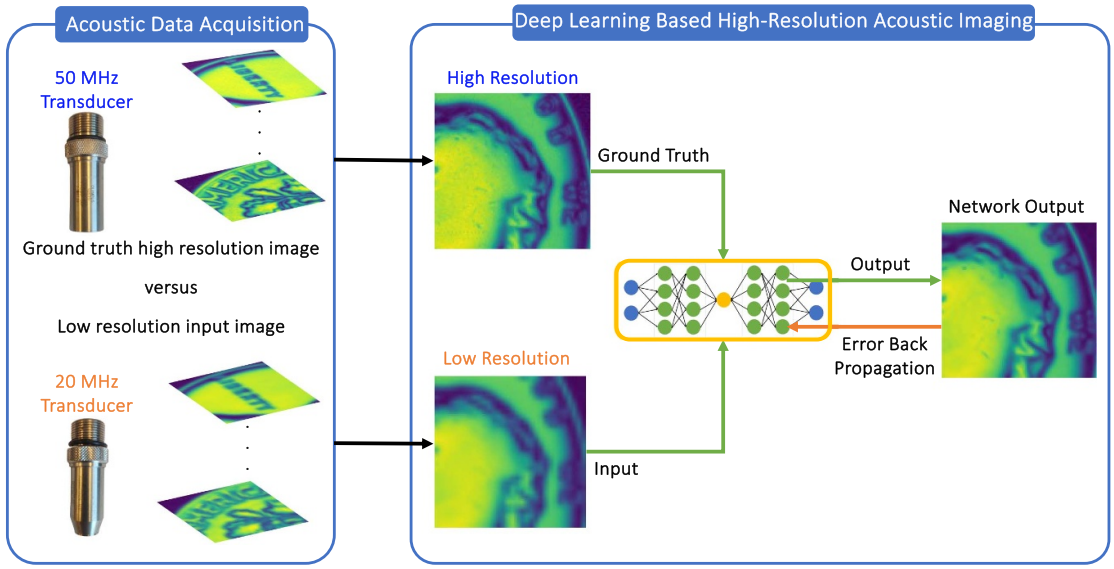

Figure 1. This figure depicts the overall strategy used in this paper. First, acoustic data is acquired for both high (50 MHz) and low resolution (20 MHz). This is followed by deep learning-based high-resolution imaging. High-resolution ground truth and corresponding low-resolution input images are fed into the network used to train. The network output is further used as error backpropagation to train the network.

truth. SwinIR consists of shallow feature extraction, deep feature extraction, and high-quality image reconstruction modules.

Shallow feature extraction

For a low quality input I_{LQ}, a 3×3 convolution layer $H_{SF}(.)$ is used for shallow feature extraction, F_0. This produces stable optimization and maps the input image space to a higher dimensional feature space [30].

Deep feature extraction

Followed by the shallow feature extraction, we have the deep feature extraction, F_{DF}, using K residual Swin transformer blocks (RSTB) and 3×3 convolutional layer. Here, intermediate features $F_1, F_2, \ldots, F_k$ and output deep feature F_{DK} are extracted block by block, given by;

$$F_i = H_{\mathrm{RSTB}_i}(F_{i-1}), i = 1, 2, \ldots, K \tag{1}$$

$$F_{\mathrm{DF}} = H_{\mathrm{CONV}}(F_K) \tag{2}$$

where the ith RSTB and last convolutional layer are denoted by $H_{\mathrm{RSTB}_i}(.)$ and $H_{\mathrm{CONV}}()$ respectively.

Each RSTB involves dividing images into patches, applying self-attention and feedforward layers, and adhering to core deep learning principles.

Each layer combines self-attention and feedforward neural networks. It starts with patch processing, resulting in X_i, representing patch embeddings at each layer. Multi-head self-attention (MSA), followed by residual connections and layer normalization, produces $Attention_i(X_i)$. Additionally, feedforward neural networks further process the output to capture intricate image features, yielding $FFN_i(LayerNorm_i(Attention_i(X_i) + X_i))$. Thus, we have,

$$H_{\mathrm{RSTB}_i}(X_i) = FFN_i(LayerNorm_i(Attention_i(X_i) + X_i)). \tag{3}$$

By stacking multiple Swin Transformer blocks, deeper and more complex features are extracted from images. The convolutional layer at the end of feature extraction brings the inductive bias of the convolution operation into the network.

Image reconstruction

The high-quality image I_{RHQ} is reconstructed by aggregating shallow and deep features as,

$$I_{\mathrm{RHQ}} = H_{\mathrm{REC}}(F_0 + F_{\mathrm{DF}}) \tag{4}$$

where $H_{\mathrm{REC}}(.)$ is the function of the high-quality image reconstruction module. While shallow features mainly focus on low frequencies, deep features recover lost high-frequencies. Due to the long skip

connection, the model can transmit low-frequency information directly to the high-quality image reconstruction module. To implement the reconstruction module, the sub-pixel convolutional layer is used to upsample the feature. Also, residual learning is used to reconstruct the residual between the low-quality and high-quality image instead of the high-quality image, given by;

$$I_{\mathrm{RHQ}} = H_{\mathrm{SwinIR}}\left(I_{\mathrm{LQ}}\right) + I_{\mathrm{LQ}} \tag{5}$$

where $H_{\mathrm{SwinIR}}(.)$ is the SwinIR function. The parameters of SwinIR are optimized by minimizing the L_1 pixel loss,

$$L = ||I_{\mathrm{RHQ}} - I_{\mathrm{HQ}}||_1 \tag{6}$$

where I_{RHQ} is obtained by taking I_{LQ} as an input of SwinIR, and I_{HQ} is the corresponding ground truth high quality image. The RSTB is a residual block with Swin transformer layers (STL) and convolutional layers. For input feature $F_{i,0}$ of the ith RSTB, the intermediate features $F_{i,1}, F_{i,2}, \ldots, F_{i,L}$ are extracted as,

$$F_{i,j} = H_{\mathrm{STL}_{i,j}}\left(F_{i,j-1}\right), j = 1, 2, \ldots, L \tag{7}$$

where $H_{\mathrm{STL}_{i,j}}(.)$ is the jth STL in ith RSTB. STL is based on the standard MSA of the original transformer layer. Local attention and shifted window mechanism are the main differences. For a local window feature X, the *query*, *key*, and *value* matrices Q, K, and V are computed as,

$$Q = XP_Q \tag{8}$$
$$K = XP_K \tag{9}$$
$$V = XP_V \tag{10}$$

where P_Q, P_K, and P_V are projection matrices which are shared across windows. Using the self-attention mechanism, the attention matrix is thus computed as,

$$Attention\left(Q, K, V\right) = \mathrm{SoftMax}\left(QK^T/\sqrt{d} + B\right)V \tag{11}$$

where B is the learnable relative positional encoding. The attention function is performed h times in parallel and the results are concatenated for MSA. Also, a multi-layer perception (MLP) having two fully connected layers with GELU non-linearity between them is used for further feature transformations. Before MSA and MLP, the LayerNorm (LN) layer is added and the residual connection is employed for both modules. Hence,

$$X = MSA\left(LN\left(X\right)\right) + X \tag{12}$$
$$X = MLP\left(LN\left(X\right)\right) + X. \tag{13}$$

There is no connection across local windows when different layers have fixed partitions. Regular and shifted window partitioning are used alternatively for enabling cross-window connections. Here, shifted window partitioning refers to shifting the features by $(\lfloor M/2 \rfloor, \lfloor M/2 \rfloor)$ pixels before partitioning.

The model architecture is given in figure 2.

2.3. Training strategy

The transfer learning approach is adopted to generate HR images in acoustic microscopy. For SRGAN model [25], 2500 epochs are used for training purposes. The Adam optimizer [31] is used with the $learningrate = 10^{-4}$, $\beta_1 = 0.9$, and, $\beta_2 = 0.999$. For the ESRGAN model [26], 3000 epochs, Adam optimizer [31], $learningrate = 10^{-4}$, and $\beta_2 = 0.99$ is used. For IMDN, DBPN-RES-MR64-3, and SwinIR models, Adam optimizer [31], 3500 epochs, $learningrate = 10^{-4}$, $batchsize = 8$, and $\beta_2 = 0.99$ is used for the training purpose. However, a steep decline in SSIM and PSNR values on the validation set can be seen after the 420 epoch. The decline was more evident in SSIM than PSNR. These can be seen in figure 3. PSNR vs epoch is in figure 3, SSIM vs epoch is in figure 3.

A similar strategy was followed for the DBPN-RES-MR64-3 model, where the model was trained with $learningrate = 10^{-4}$, and Adam optimizer [31] with $batchsize = 16$, $\beta_1 = 0.9$, and $\beta_2 = 0.999$; and the IMDN model, where the model was trained with $learningrate = 10^{-4}$, and Adam optimizer [31] with $batchsize = 16$, $\beta_1 = 0.9$, and $\beta_2 = 0.999$.

Finally, the SwinIR model displayed the best results regarding both PSNR and SSIM scores. The G_loss of the SwinIR training has been displayed in figure 4.

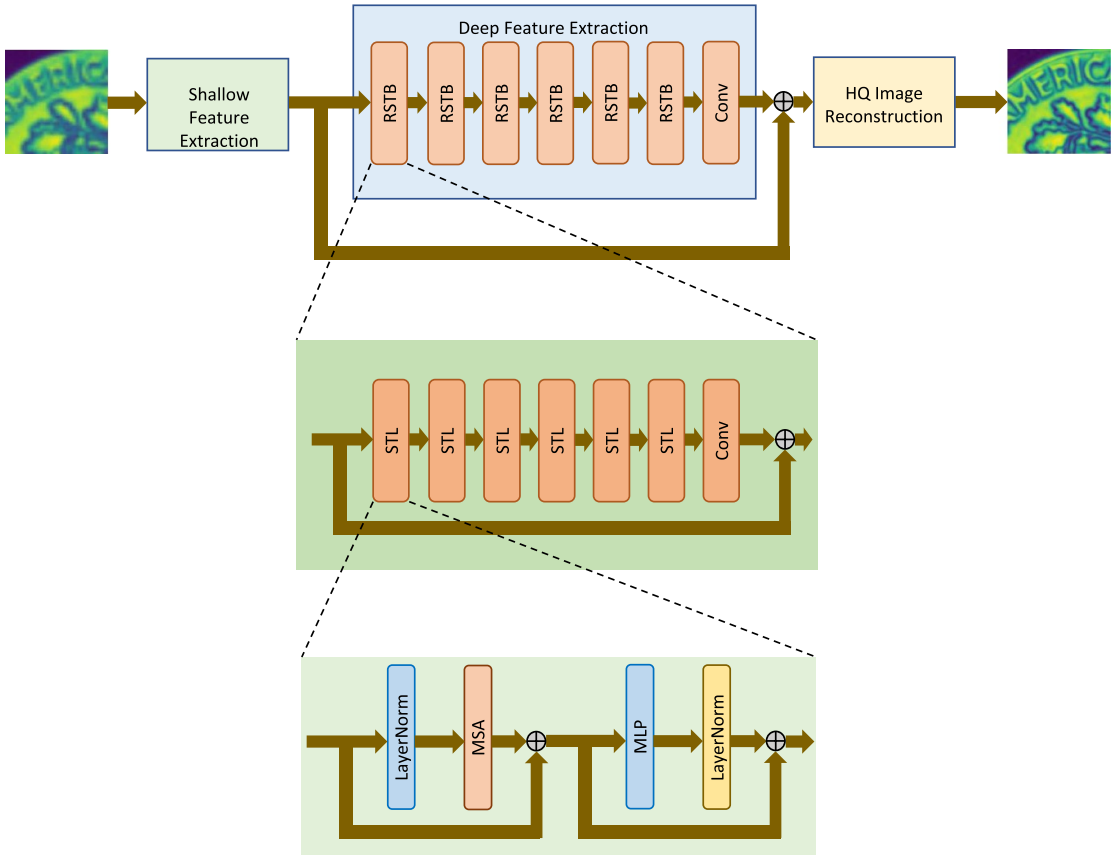

Figure 2. Diagram depicting the network architecture. This entire architecture represents the 'Network Under Training' in figure 1. The network consists of shallow feature extraction, deep feature extraction, and high-quality (HQ) image reconstruction modules. While the shallow feature extraction produces stable optimization and maps the input image space to a higher dimensional feature space, the model also has the deep feature extraction, consisting of K residual Swin transformer blocks (RSTB) and a 3×3 convolutional layer.

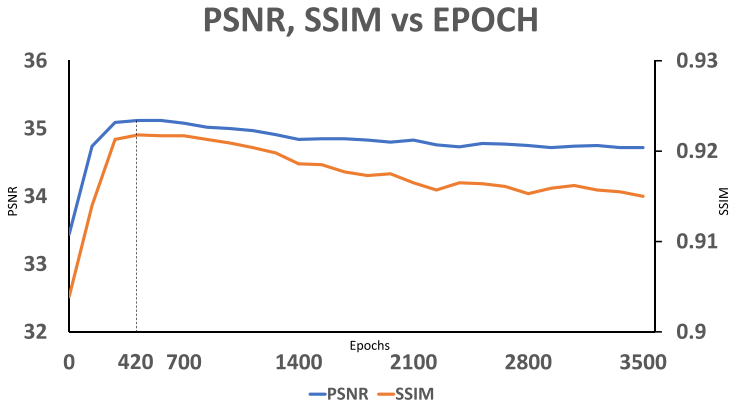

Figure 3. The effect of PSNR and SSIM in increasing the number of epochs can be seen through these figures for the SwinIR model. Both the SSIM and PSNR are seen to attain maxima at 420 epochs, after which they start to decrease. The SSIM is seen to decline more rapidly than the PSRN.

3. Materials and methods

3.1. Experimental setup

There are two operational modes in SAM namely; reflection and transmission mode. A detailed description of the working principles of both modes can be found elsewhere [32]. In this manuscript we consider reflection mode to scan the samples. To focus acoustic energy via a coupling medium (in this case, water), a concave spherical sapphire lens rod is generally utilized. Later on, ultrasound signals are generated from the signal generator and transmitted toward the sample and the reflected waves were recorded. The digitized

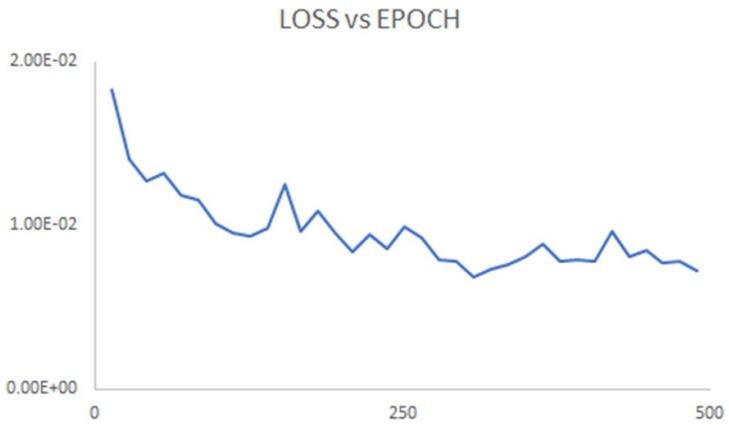

Figure 4. Loss vs epoch for SwinIR model is depicted here. As the training progresses with more and more epochs, the loss gradually decreases.

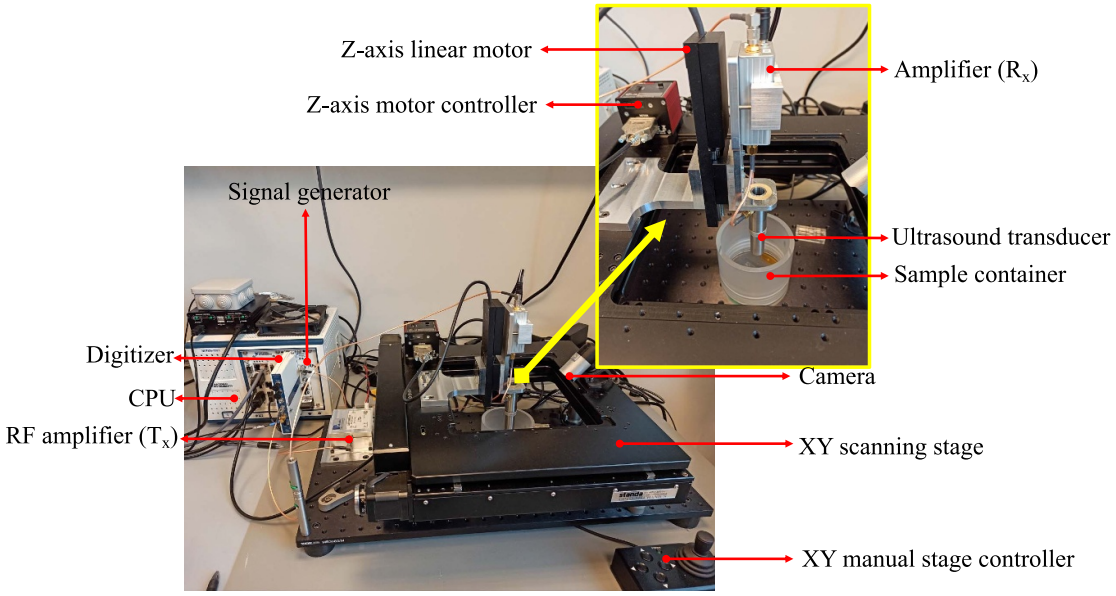

Figure 5. Figure depicts a labeled image of SAM used for image acquisition as discussed in this paper. The experimental setup demonstrates all the fundamental components that constitute a SAM.

signal from the samples is called an A-scan or amplitude scan. In order to acquire a C-scan of the sample, repeat this procedure at various points in the *XY* plane. In another way, a C-scan can be referred to as the summation of A-scans in two dimensions.

For this experiment the data acquisitions were performed on a custom-built SAM (figure 5), integrated with a Standa (8MTF-200-Motorized XY Microscope Stage) high-precision scanning stage controlled by LabVIEW [33] program. A similar experimental setup was employed earlier by our group to determine and correct the inclined sample [34, 35]. The acoustic microscopic features were implemented employing National Instruments' PXIe FPGA modules and FlexRIO hardware. It was enclosed in a PXIe chassis (PXIe-1082) which consists of an arbitrary waveform generator (AT-1212). For acoustic imaging, the transducer was excited with signals (Mexican hat) and delivered into an RF amplifier (AMP018032-T) for further amplification of the ultrasonic signals [36]. The acoustical reflections caused by the transmitted signal on the surface of the sample were picked up and fed into a custom-designed amplifier. The role of such an amplifier is to amplify the currents into an output potential. These signals were then amplified with a custom-designed pre-amplifier and digitized with a 12-bit high-speed (1.6 GS s^{-1}) digitizer (NI-5772).

For ground truth, an Olympus 50 MHz focused transducer having an aperture of 6.35 mm and a focal length of 12 mm was used to scan all the coins and also the biological specimen. By focusing the acoustic energy on the coin's top surface, the sample was scanned in the x and y directions with 50 μm steps. On the other-hand low resolution images were acquired with a 20 MHz transducer (focal length 50 mm). All

experiments were performed in distilled water and the room temperature was kept constant at around 22 °C during the experiments. For testing the models a biological sample was employed in SAM for imaging. The sample used for this experiment was a discarded reindeer antler that was collected from the jungle of Tromsø (Norway). The moss on the antler was first removed by cleaning it with lukewarm water and 96% ethanol. After cleaning, the sample was diced and boiled in distilled water at 100 °C in order to remove any unwanted biological substance from the antler. The sample was thereafter put on the sample holder and allowed to dry before being scanned.

UiT The Arctic University of Norway has a focus on research on topics and concerns related to Arctic life such as reindeer. The local Arctic tribes consider that Reindeer's antlers indicate the health and well-being of reindeer. Therefore, investigating them is a long-term interest for us. Now, to study such biological samples, above mentioned preparation step is generally required which includes slicing the antler to view it under the microscope. Potentially, in the long term, our method can provide a more scientifically rooted study of this conjecture.

4. Results and discussion

The models correctly identified numerals, alphabets, and patterns in the coins. The SwinIR model produced the best results when compared to SRGAN, ESRGAN, DBPN-RES-MR64-3, and IMDN. The SwinIR model identified alphabets, digits, and patterns more effectively than the other models. The SwinIR model also has the data set's highest PSNR and SSIM scores. The SwinIR model has an average SSIM of 0.92 and a PSNR of 35.13. Figure 6 displays the outputs of the various models and the corresponding ground truth images.

The PSNR and SSIM scores of the pictures across multiple models are presented in the tables 1 and 2 for the 10 sample inputs displayed in figure 6. The SwinIR model has the highest PSNR and SSIM scores (in bold letters) for all ten inputs. Although the DBPN-RES-MR64-3 model comes in second place, the SwinIR model's SSIM and PSNR scores are superior in all circumstances. We can observe from figure 6 that the SwinIR model output is the most accurate and closest to the ground truth when compared to the other models.

4.1. Effect of transfer learning strategy

Initially, 17 coins from different countries and with different dimensions were scanned for 20 MHz (low resolution) and 50 MHz (ground truth). Later on, they were cropped and several images were created, and eventually, a training data set of 800 images were generated, the models were prone to overfitting and low PSNR and SSIM scores (7.92 and 0.064 respectively). Hence, transfer learning was used in training the models. This significantly increased the PSNR and SSIM scores and the overall results of the models. Hence, pre-trained models trained using data sets like celebA data-set [37] and DIV2K data-set [38–41] were used and they were fine-tuned using the coin data set, which improved the model performance to a large extend.

Figure 7 shows some more examples of outputs we got across models without using the transfer learning approach.

4.2. Comparison with conventional digital resolution enhancement techniques

Popular digital resolution enhancement techniques include the nearest neighbor interpolation algorithm, bi-linear interpolation, and cubic convolution interpolation. However, when applied to images, all of these algorithms have some drawbacks. Errors occur when the picture is overly expanded. The nearest neighbor interpolation algorithm results in significant image quality loss, as well as visible mosaic and jagged phenomena. Because of the poor design of the interpolation function, the output image of the bi-linear interpolation algorithm suffers from quality damage and low calculation accuracy. The cubic convolution interpolation algorithm requires a significant amount of calculation and is also complicated and time-consuming. Interpolation-based algorithms also have issues with computational complexity, noise amplification, and blurry images. Deep learning methods and techniques, on the other hand, have advanced in recent years, and thus deep learning-based SR models are used. These methods frequently achieve cutting-edge performance on various resolution enhancement benchmarks.

4.3. Testing on the unknown biological sample (Reindeer Antler)

A biological sample (Reindeer Antler) of unknown size, shape, and surface morphology was employed to validate the reproducibility of the suggested model. This experiment was conducted with a discarded reindeer antler gathered from the local jungle. To remove the moss, the antler was cleaned with lukewarm water and ethanol. Following cleaning, the antler was diced and boiled at 100 °C in distilled water for 30 minutes to eliminate any undesired biological substances. The sample was then dried and mounted on the sample holder in preparation for scanning.

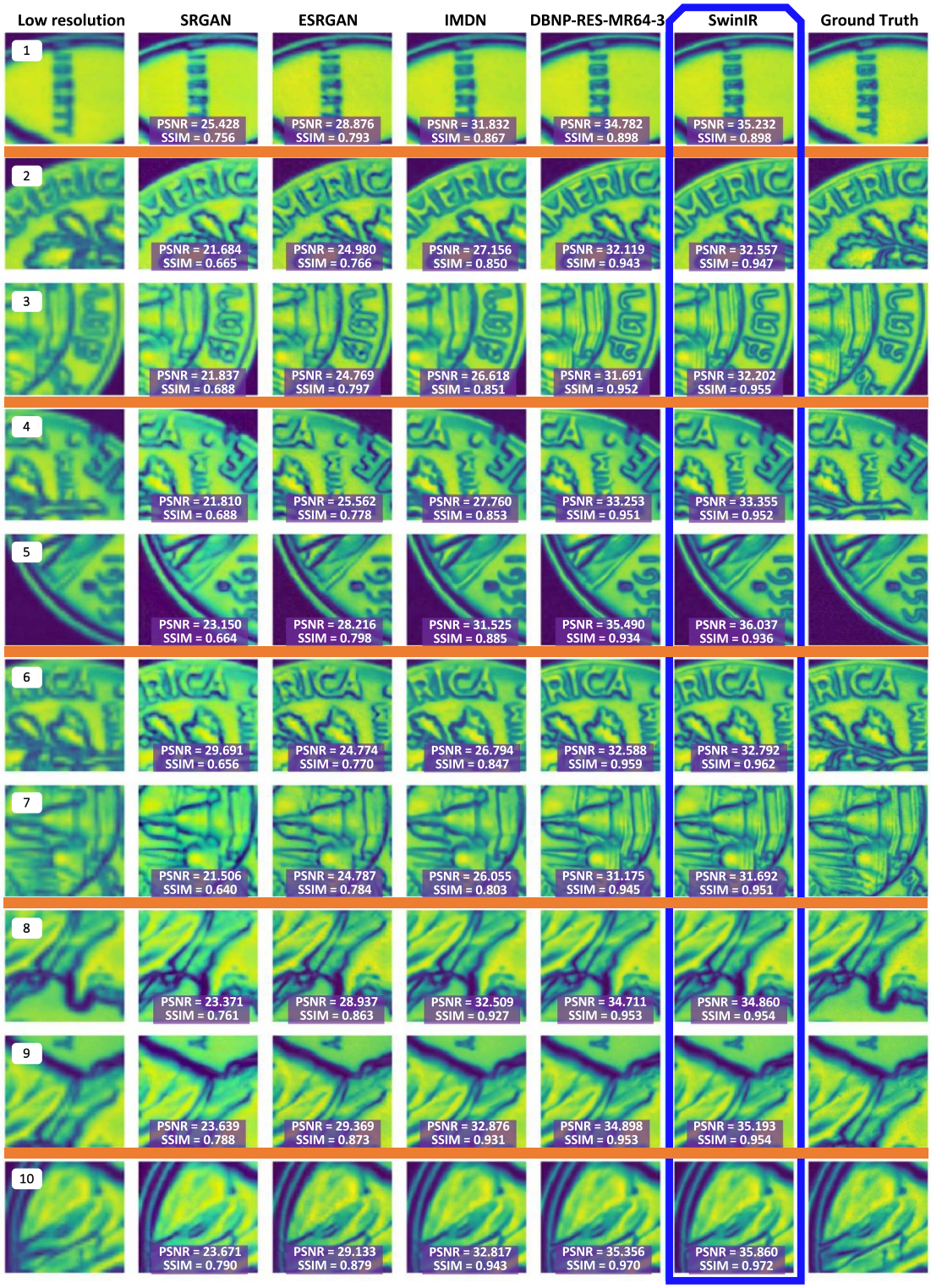

Figure 6. The figure depicts the results obtained after training different models for ten example input images. Visually, SwinIR (marked with a blue box) seems to provide the best results, they also have the best SSIM and PSNR scores. The input images have been chosen such that they contain text, digits, or patterns. All the images are 6.4 mm × 6.4 mm.

The figure 8 shows the result obtained after testing the unknown biological sample (Reindeer Antler). The original image was taken to be the ground truth and a LR image was generated from the original image (figure 8). The low-resolution image was given as input to the final model to get the generated output image using the SwinIR model. The result showed a PSNR value of 31.88 and an SSIM value of 0.8406. Although the PSNR and SSIM values are comparatively less than the average SSIM and PSNR scores obtained in the test data set, it is because a data set consisting of only coin images was used to train the model. Also, it is to be noted that the image is evaluated only quantitatively.

Table 1. The table contain the PSNR scores of the images across various models for the ten example inputs shown in figure 6. The SwinIR model has the best PSNR scores (written in bold letters) for all 10 inputs.

Image #	SRGAN	ESRGAN	IMDN	DBPN-RES-MR64-3	SwinIR
1	25.43	28.88	31.83	34.79	**35.23**
2	21.68	24.98	27.16	32.12	**32.56**
3	21.84	24.77	26.62	31.69	**32.20**
4	21.81	25.51	27.76	33.25	**33.35**
5	23.15	28.22	31.52	35.49	**36.04**
6	21.69	24.77	26.79	32.59	**32.79**
7	21.51	24.79	26.06	31.17	**31.69**
8	23.37	28.94	32.51	34.71	**34.86**
9	23.64	29.37	32.88	34.90	**35.19**
10	23.67	29.13	32.82	35.36	**35.86**

Table 2. The table represents the SSIM scores of the images across different models for the ten examples shown in figure 6. The SwinIR model has demonstrated the best SSIM scores (written in bold letters) for all ten inputs.

Image #	SRGAN	ESRGAN	IMDN	DBPN-RES-MR64-3	SwinIR
1	0.76	0.79	0.87	0.90	**0.91**
2	0.67	0.77	0.85	0.94	**0.96**
3	0.69	0.80	0.85	0.95	**0.96**
4	0.67	0.78	0.85	0.95	**0.95**
5	0.66	0.80	0.88	0.93	**0.94**
6	0.66	0.77	0.85	0.96	**0.96**
7	0.64	0.75	0.80	0.95	**0.95**
8	0.76	0.86	0.93	0.95	**0.95**
9	0.79	0.87	0.93	0.95	**0.95**
10	0.79	0.88	0.94	0.97	**0.97**

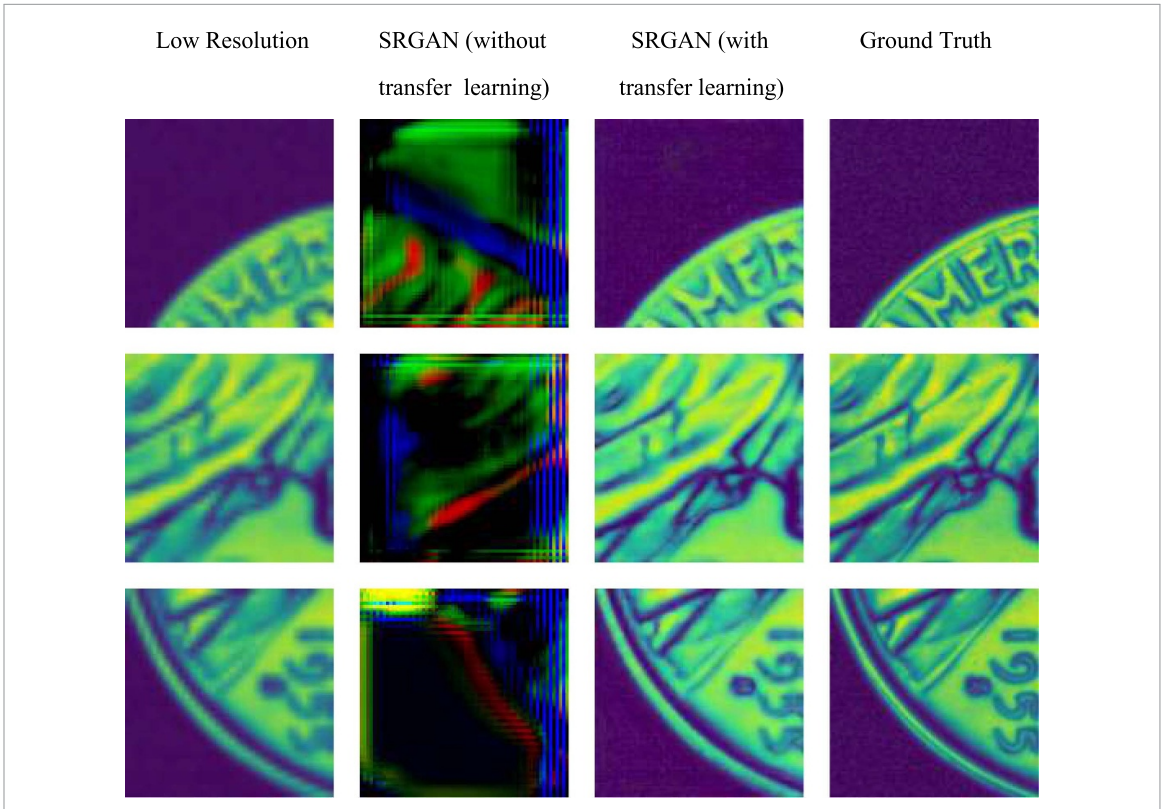

Figure 7. The figure depicts the results obtained after training SRGAN model without using transfer learning. The approach showed extremely poor results. The images had PSNR scores in the range of 7–8 and SSIM scores in the range of 0.03–0.07. All the individual images are 6.4 mm × 6.4 mm.

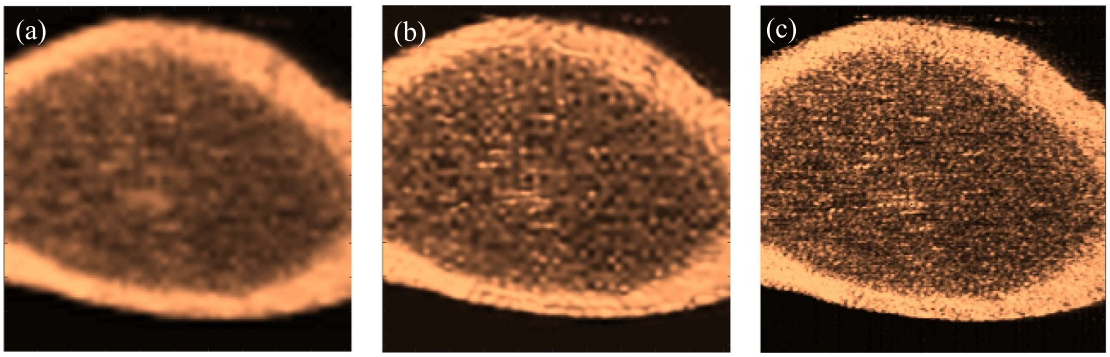

Figure 8. The figure depicts the results obtained after testing the unknown biological sample on SwinIR model. A PSNR score of 31.88 and SSIM score of 0.8406 was obtained on this sample. The comparatively low scores are due to the fact that the data set used consisted of only coin images which are different than this input image. All three images have the dimension of 12 mm × 12 mm.

5. Conclusion

In this paper, we developed an acoustic microscopy system that uses deep learning to improve the image resolution of industrial and biological samples by four times. Acoustic image acquisition was carried out on a custom-developed SAM, equipped with a high-precision scanning stage. Deep learning was used to improve the lateral resolution of the SAM images. SRGAN, ESRGAN, IMDN, SwinIR, and DBPN-RES-MR64-3 were the models compared in this study. All five models were trained and tested on 17 different coin images, and the results were reported in terms of PSNR and SSIM scores. The SwinIR model is made up of modules for shallow feature extraction, deep feature extraction, and high-quality image reconstruction. The model's long skip connections allow it to send low-frequency data directly to the high-quality image reconstruction module. The process took into account transfer learning. Because only 800 images were used for training, this was done to prevent the model from overfitting. Methods that did not use transfer learning were also implemented, but the results were demonstrated poor in terms of PSNR and SSIM. SwinIR model presented an average SSIM of 0.92 and a PSNR of 35.13. The SwinIR model was also used to test an unknown biological sample. The SSIM score was 0.8406 and the PSNR score was 31.88. Deep learning methods for resolution enhancement have many advantages over traditional digital resolution enhancement techniques, and the SwinIR model performed the best of the five deep learning techniques investigated in this paper. Deep learning-based models, specifically SwinIR, are found to closely approximate the ground truth image even with extremely limited training data.

6. Future directions

This work has established that the resolution enhancement can be used to improve the digital resolution of scanning acoustic microscopy, more work is needed in the future to further mature our technique. According to the usual signal processing theory, Nyquist sampling is mandatory in digitization, and on the other hand, the acoustic resolution sets the bandwidth of the measurement. While digital oversampling is routine for high-quality and high-definition imaging, it is reasonable to expect that performing barely Nyquist sampling and applying our technique should be able to support high-quality high-definition imaging according to the signal processing theory. However, the improvement in perceptive quality must be saturated, and performing digital resolution enhancement using Nyquist sampled images may present an optimum between the learnability of our approach and the extent of resolution enhancement sought. It is also likely that beyond a certain point, the learning becomes imprecise and the model introduces artifacts in the HR images. So, this aspect needs to be studied in an extensive manner and benchmarked. It is further interesting to consider that deep learning can learn features from the large data priors during training, which may not all be available in the testing when we actually present the data for resolution enhancement. This provides an opportunity to consider if we can perform sampling at a rate poorer than the Nyquist criterion and use the pre-learned data priors in the deep-learned model to compensate for the deficiency in measurement. This possibility however requires careful assessment and benchmarking and constitutes a future direction. We would also like to consider the effect of the sample material and inhomogeneity on the achievable resolution enhancement and the value of transfer learning our models on different types of samples or different scanning acoustic microscopy instruments.

It is interesting that the proposed method does not ask for any modification in the instrument, sample, or measurement protocol. Therefore, on the one hand, our technique can be directly applied to existing systems, on the other hand, it implies that the cost of a new instrument remains unchanged. Therefore, it is of interest to evaluate the value of using our technique in practical terms. Our technique can be of prime importance where time is critical and any advantage in scanning time translates to monetary or non-monetary value. Examples of such situations include material integrity or failure analysis during the process of implanting or checking medical implants or at a site of infrastructural failure for tactical decision-making. Moreover, our technique improves the throughput of existing scanning acoustical microscopes by orders of magnitude, which in commercial settings translates to more volume or lower operating costs of commerce.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://drive.google.com/file/d/1_d-R_OfCcE1E_mG259D7nToRHGduDZEh/view?usp=sharing. Data will be available from 7 June 2024.

Acknowledgments

A H would like to thank Arif S Ahmad for the initial discussions. This work was supported by the Research Council of Norway, Cristin Project, ID: 2061348. The publication charges for this article have been funded by a grant from the publication fund of UiT The Arctic University of Norway.

Author contributions statement

A H, A B, D K P, and K A have conceptualized the idea. A H designed the experiments. P B implemented the deep learning approach with initial help from S M A A H and P K developed the SAM. M S performed the SAM experiments with the help of A H. Funding was secured by F M and A H Formal analysis and experimental validation were performed by P B who also wrote the original draft, reviewed, and edited the manuscript with support from all co-authors.

ORCID iD

Anowarul Habib https://orcid.org/0000-0001-6515-3145

References

[1] Briggs A, Briggs G and Kolosov O 2010 *Acoustic Microscopy* vol 67 (Oxford University Press)
[2] Brand S, Raum K and Czurratis P 2008 Scanning acoustic microscopy an application for evaulating varnish layer conditions non-destructively *2008 IEEE Ultrasonics Symp.* (IEEE) pp 615–8
[3] Wolfe J P 2005 *Imaging Phonons: Acoustic Wave Propagation in Solids* (Cambridge University Press)
[4] Wagle S, Habib A and Melandsø F 2017 Ultrasonic measurements of surface defects on flexible circuits using high-frequency focused polymer transducers *Jpn. J. Appl. Phys.* **56** 07JC05
[5] Hofmann M, Pflanzer R, Habib A, Shelke A, Bereiter-Hahn J, Bernd A, Kaufmann R, Sader R and Kippenberger S 2016 Scanning acoustic microscopy-a novel noninvasive method to determine tumor interstitial fluid pressure in a xenograft tumor model *Trans. Oncol.* **9** 179–83
[6] Habib A, Shelke A, Vogel M, Brand S, Jiang X, Pietsch U, Banerjee S and Kundu T 2015 Quantitative ultrasonic characterization of c-axis oriented polycrystalline aln thin film for smart device application *Acta Acust. U. Acust.* **101** 675–83
[7] Habib A, Shelke A, Vogel M, Pietsch U, Jiang X and Kundu T 2012 Mechanical characterization of sintered piezo-electric ceramic material using scanning acoustic microscope *Ultrasonics* **52** 989–95
[8] Hadimioglu B and Quate C 1983 Water acoustic microscopy at suboptical wavelengths *Appl. Phys. Lett.* **43** 1006–7
[9] Habib A, Vierinen J, Islam A, Martinez I Z and Melandsø F 2018 In vitro volume imaging of articular cartilage using chirp-coded high frequency ultrasound *2018 IEEE Int. Ultrasonics Symp. (IUS)* (IEEE) pp 1–4
[10] Pflanzer R, Hofmann M, Shelke A, Habib A, Derwich W, Schmitz-Rixen T, Bernd A, Kaufmann R and Bereiter-Hahn J 2014 Advanced 3D-sonographic imaging as a precise technique to evaluate tumor volume *Trans. Oncol.* **7** 681–6
[11] Yuan X, Cui X, Gu H, Wang M, Dong Y, Cai S, Feng X and Wang X 2020 Evaluating cervical artery dissections in young adults: a comparison study between high-resolution MRI and CT angiography *Int. J. Cardiovascular Imag.* **36** 1113–9
[12] Pahlevaninezhad H *et al* 2018 Nano-optic endoscope for high-resolution optical coherence tomography, *in vivo Nat. Photon.* **12** 540–7
[13] Wu C, Gleysteen J, Teraphongphom N T, Li Y and Rosenthal E 2018 *In-vivo* optical imaging in head and neck oncology: basic principles, clinical applications and future directions *Int. J. Oral Sci.* **10** 1–13
[14] Tang Y *et al* 2021 High-resolution 3D abdominal segmentation with random patch network fusion *Med. Image Anal.* **69** 101894
[15] Mancuso J J, Chen Y, Li X, Xue Z and Wong S T 2013 Methods of dendritic spine detection: from golgi to high-resolution optical imaging *Neuroscience* **251** 129–40
[16] Cao Y *et al* 2021 Automatic detection and segmentation of multiple brain metastases on magnetic resonance image using asymmetric UNet architecture *Phys. Med. Biol.* **66** 015003

[17] Park S, Gach H, Kim S, Lee S J and Motai Y 2021 Autoencoder-inspired convolutional network-based super-resolution method in MRI *IEEE J. Trans. Eng. Health Med.* **PP** 1–1
[18] Myronenko A 2018 3D MRI brain tumor segmentation using autoencoder regularization *Int. MICCAI Brainlesion Workshop* (Springer) pp 311–20
[19] Liu J, Chen F, Wang X and Liao H 2019 An edge enhanced SRGAN for MRI super resolution in slice-selection direction *MBIA 2019, and 7th Int. Workshop, MFCA 2019* (*Shenzhen, China*, *October 17*) pp 12–20
[20] Makra A, Bost W, Kalló I, Horváth A, Fournelle M and Gyöngy M 2020 Enhancement of acoustic microscopy lateral resolution: a comparison between deep learning and two deconvolution methods *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **67** 136–45
[21] de Leeuw den Bouter M L, Ippolito G, O'Reilly T P, Remis R F, van Gijzen M B and Webb A G 2022 Deep learning-based single image super-resolution for low-field MR brain images *Sci. Rep.* **12** 6362
[22] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser L and Polosukhin I 2017 Attention is all you need *CoRR* (arXiv:1706.03762)
[23] Carion N, Massa F, Synnaeve G, Usunier N, Kirillov A and Zagoruyko S 2020 End-to-end object detection with transformers *European conf. on computer vision* (*August 2020*) pp 213–29
[24] Dosovitskiy A *et al* 2020 An image is worth 16 × 16 words: transformers for image recognition at scale *Preprint* (arXiv:2010.11929)
[25] Ledig C *et al* 2017 Photo-realistic single image super-resolution using a generative adversarial network *2017 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)* pp 105–14
[26] Wang X, Yu K, Wu S, Gu J, Liu Y, Dong C, Qiao Y and Change Loy C 2018 Esrgan: enhanced super-resolution generative adversarial networks *Proc. European Conf. on Computer Vision (ECCV)* pp 0–0
[27] Hui Z, Gao X, Yang Y and Wang X 2019 Lightweight image super-resolution with information multi-distillation network *Proc. 27th ACM Int. Conf. on Multimedia (ACM MM)* pp 2024–32
[28] Zhang K, Gu S and Timofte R 2019 Aim 2019 challenge on constrained super-resolution: methods and results *The IEEE Int. Conf. on Computer Vision (ICCV) Workshops*
[29] Haris M, Shakhnarovich G and Ukita N 2019 Deep back-projection networks for single image super-resolution *IEEE Trans. Pattern Anal. Mach. Intell.* **43** 4323–37
[30] Liang J, Cao J, Sun G, Zhang K, Van Gool L and Timofte R 2021 Swinir: image restoration using swin transformer (arXiv:2108.10257)
[31] Kingma D and Ba J 2014 Adam: a method for stochastic optimization *Int. Conf. on Learning Representations*
[32] Gupta S K, Pal R, Ahmad A, Melandsø F and Habib A 2023 Image denoising in acoustic microscopy using block-matching and 4D filter *Sci. Rep.* **13** 13212
[33] Standa 2020 Motorized xy microscope stage-motorized positioners & controllers-catalog-opto-mechanical products-standa (available at: www.standa.lt/products/catalog/motorised_positioners?item=609&prod=motorized_xy_microscope_stage) (accessed 16 March 2022)
[34] Kumar P, Yadav N, Shamsuzzaman M, Agarwal K, Melandsø F and Habib A 2022 Numerical method for tilt compensation in scanning acoustic microscopy *Measurement* **187** 110306
[35] Gupta S K, Habib A, Kumar P, Melandsø F and Ahmad A 2023 Automated tilt compensation in acoustic microscopy *J. Microsc.* **292** 90–102
[36] BNT 2019 Becker nachrichtentechnik gmbh (bnt), "1 w wideband amplifier" (available at: www.becker-rf.com/files_db/1608213582_2017__17.pdf) (Accessed 16 March 2022)
[37] Liu Z, Luo P, Wang X and Tang X 2015 Deep learning face attributes in the wild *Proc. Int. Conf. on Computer Vision (ICCV)*
[38] Agustsson E and Timofte R 2017 Ntire 2017 challenge on single image super-resolution: dataset and study *The IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) Workshops*
[39] Timofte R, Agustsson E, Van Gool L, Yang M-H, Zhang L and Lim B 2017 Ntire 2017 challenge on single image super-resolution: methods and results *The IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) Workshops*
[40] Timofte R, Gu S, Wu J, Van Gool L, Zhang L, Yang M-H and Haris M 2018 Ntire 2018 challenge on single image super-resolution: methods and results *The IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) Workshops*
[41] Ignatov A and Timofte R 2019 Pirm challenge on perceptual image enhancement on smartphones: report *European Conf. on Computer Vision (ECCV) Workshops*