Material Texture Recognition using Ultrasonic Images with Transformer Neural Networks

Xin Zhang and Jafar Saniie
Embedded Computing and Signal Processing (ECASP) Research Laboratory (http://ecasp.ece.iit.edu)
Department of Electrical and Computer Engineering
Illinois Institute of Technology, Chicago, Illinois, U.S.A.

Abstract— **Material texture recognition by estimating the grain size has been extensively used for the characterization of material structures. Ultrasonic inspection can approximate material grain size nondestructively with advantages of one-sided measurement, high penetration depth, and inspection accuracy. In ultrasonic testing, the energy of the signal attenuates as the ultrasonic signal propagates through the material. This attenuation is due to scattering and absorption, which are functions of the frequency and grain size distribution. Therefore, the attenuation and scattering of ultrasonic echoes can be used to evaluate grain size for microscopic texture. In this paper, we propose to use the transformer neural networks to learn grain scattering features for material texture recognition. The transformer neural network utilizes the multi-head attention mechanism to substantially reduce the computation cost. An ultrasonic testbed platform is assembled to acquire the 3D ultrasonic data cube to train the neural networks for texture analysis. The 3D data cube consists of a sequence of 2D ultrasonic C-scan images and is obtained from three different heat-treated steel blocks. Several state-of-the-art machine learning algorithms, the deep Convolutional Neural Networks (CNNs) and Support Vector Machine (SVM) were trained and compared to classify the grain scattering textures of three heat-treated steel blocks. To build a data-efficient automatic system for ultrasonic nondestructive evaluation (NDE) applications, a self-attention-based transformer neural network: Ultrasonic Texture Recognition Vision Transformer: UTRV Transformer, was proposed to classify material textures with high testing accuracy of 96.15%.**

Keywords — Material Texture Recognition, Grain Size Estimation, Transformer Neural Networks, Deep CNN, Self-Attention, Data-Efficient

I. INTRODUCTION

Material texture recognition nondestructively by using ultrasonic inspection to approximate the grain size is a promising method to characterize the structural and physical integrity of materials [1]. To estimate the grain size, the ultrasonic microstructure scattering signals and signal attenuation can be used [2]. However. to measure the signal attenuation is difficult as the attenuation shows variations of local grain size across the entire propagation path. Therefore, to estimate the grain size, the scattering signal is characterized by utilizing the statistical variation in the scattered energy as a function of depth [3]. In addition, the characterization of the backscattered signal is since the signal intensity is the non-explicit function of average grain size, random distribution of grains, and ultrasonic frequency [4]. Furthermore, for better characterization of the backscattered signals, the Rayleigh scattering regions where the ultrasonic wavelength is larger than the average grain size is applied to measure ultrasonic signals [5]. As the Rayleigh scattering region is most sensitive to frequency and grain size distribution. Traditionally, material texture recognition requires manual analysis, which is complicated and inaccurate. To enhance the performance for texture analysis, it is more reliable to deploy machine learning methods [6-8] to analyze ultrasonic images. In ultrasonic evaluation, neural networks have been used to estimate grain size by characterizing the backscattered signals [9][10][11][12]. And the deep Convolutional Neural Networks (deep-CNNs) is superior in quality on image recognition to learn image features automatically due to the use of multiple feature extraction stages [8]. However, deep CNNs have large amounts of parameters (hundreds of millions) which requires substantial computational resources for training and execution. In addition, the performance of deep CNNs is saturated with the growth of models and datasets [8]. Recently, the transformer neural networks [13] are emerging and become state-of-art models in Natural Language Processing (NLP). The transformer neural networks utilize a simple network architecture that replaces the convolutions and recurrence entirely with the attention mechanism. And the attention mechanism largely reduces training time [13]. To utilize these advantages in computer vision, the transformer neural networks are introduced to classify images and outperform the state-of-art deep CNNs using significantly fewer computational resources to train [14]. In this study, we aim to build a data-efficient automatic system for ultrasonic NDE applications to recognize material textures. The performance of using the transformer neural networks to learn grain scattering features and classify material textures is investigated. In addition, we proposed the Ultrasonic Texture Recognition Vision Transformer: UTRV Transformer, to recognize material textures using ultrasonic images. The ultrasonic C-scan images are constructed from acquired backscattered signals (3D data cube) from three heat-treated steel blocks with different grain sizes, then are used to train the UTRV Transformer neural network. Our objective is to build a data-efficient automatic system for ultrasonic NDE applications.

In Section II of this paper, we present the ultrasonic NDE system implementation to acquire backscattered signals. Section III presents the UTRV Transformer for texture recognition. Performance of the UTRV Transformer with deep CNNs and the kernel-SVM are analyzed and compared. Section IV concludes this paper.

II. System and Experiment Setup

Figure 1 shows the assembled ultrasonic NDE system we use to acquire backscattered signals from heat-treated steel blocks. In the ultrasonic system, a water tank is built to submerge the test specimen under the water. This guarantees good ultrasonic energy propagation. In addition, an ultrasonic pulser/receiver, Panametrics Model 5052PR, is used as the signal generator and echo receiver. To synchronize the digital signal from pulser/receiver to monitor the signals, an oscilloscope, Keysight MSOX2024A, is used as a high-frequency digitizer. Two stepper motors are mounted into the water tank to automate this system and allow the ultrasonic transducer to precisely move along x and y axes.

In Figure 2 the ultrasonic C-scan testing is shown to obtain backscattered signals. The C-scan was created with 200x200 measurement points and each measurement point contains 7680 backscattered signals. Therefore, the backscattered signals which are acquired from the ultrasonic C-scan form a 3D data cube. This 3D data cube consists of a sequence of 2D C-scan images with the size of 200x200 pixels. We use the sequence of C-scan images as inputs to train the UTRV Transformer neural network. In Figure 2, the volumetric C-scan testing area is represented as the red dashed line cube. The 2D C-scan images at four depths of three heat-treated steel blocks with grain sizes of 14, 24, and 50 microns are shown. As we can see that these C-scan images show the statistical variance of the microstructure scattering signals. In addition, the ultrasonic signals are acquired using a broadband piezoelectric transducer centered at 5 MHz. In this study, we use three types of heat-treated 1018 steel blocks for ultrasonic testing. These steel blocks have different grain sizes of 14, 24, and 50 microns, which will be referred to with notation *Grain14, Grain24,* and *Grain50* in this paper. We collected 4800 C-scan images for each of *Grain14*, *Grain24*, *Grain50* steel blocks to train the UTRV Transformer, deep-CNNs and kernel-SVM.

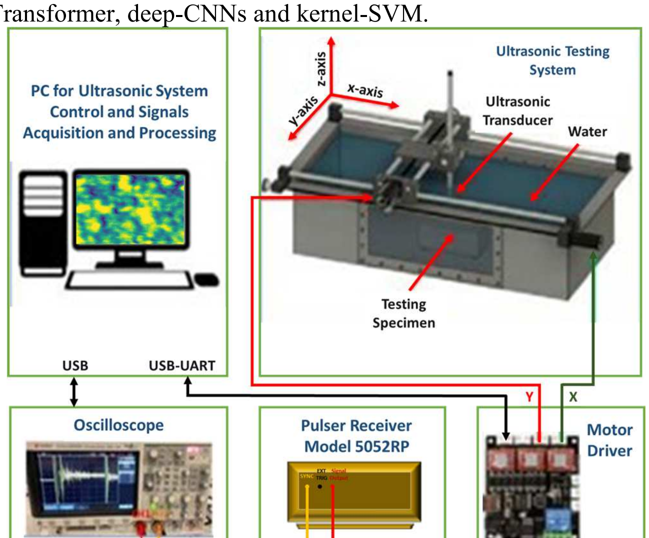

Figure 1. Schematics of Ultrasonic NDE System

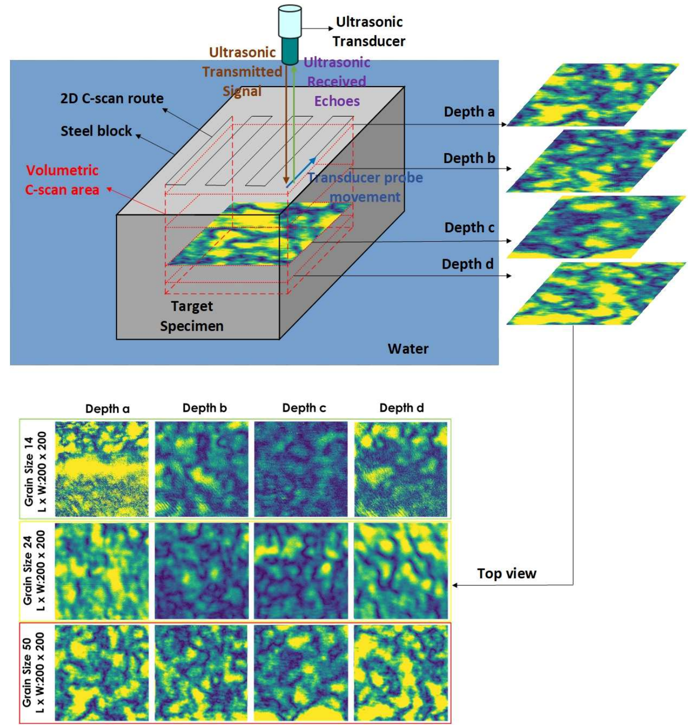

Figure 2. Ultrasonic C-scan Testing to Acquire Backscattered Signals and 2D Scattering Images

III. Material Texture Recognition

The backscattered signals are used to characterize the material microscopic texture of these three types of heat-treated steel blocks. These acquired C-scan images for *Grain14, Grain24, Grain50* steel blocks were constructed from acquired 3D data cubes, then been labeled to train the UTRV Transformer neural network.

In the following three subsections, the UTRV Transformer is introduced to classify three types of steel textures using ultrasonic C-scan images. To compare performance for material texture recognition, two different architectures of deep CNNs with two size levels, and the kernel-SVM were trained with C-scan images for texture classification. The UTRV Transformer consists of patch and position embeddings and eight transformer encoder blocks [13] to extract image features. Next, the features are fed into a multi-layer neural network to classify material textures. In addition, the deep CNNs were pre-trained using transfer learning (TL) [15] on large datasets (ImageNet) followed by further training with experimental ultrasonic images. To compare the UTRV Transformer and deep CNN architectures in parallel, the transformer encoders and convolutional bases were trained with the same two-layer fully connected neural network. This two-layer neural network consists of a hidden layer of 256 neurons and an output layer of 3 neurons. 'relu' activation is applied to the hidden layer with a dropout of 0.5 to improve the training performance [8]. The UTRV Transformer, deep CNNs, and kernel-SVM were trained with 50 epochs. Among all the acquired C-scan images, we used 80 percent to train the models and 20 percent to validate the recognition performance. The training was conducted on the Intel (R) Core (TM) i7-8750H, CPU@2.20GHz 2.21GHz computer with GTX 1070 GPU.

A. *Texture Recognition using the Transformer Neural Networks*

The UTRV Transformer uses the transformer encoder as the main building architecture to learn grain scattering features. The transformer encoder includes the multi-head attention mechanism which reduces the computational cost to train [13]. Figure 3 shows the architecture of the UTRV Transformer. The acquired C-scans are fed as inputs to train this transformer neural network. Before training, each input ultrasonic image is divided into a sequence of image patches, which is similar to a sequence of word embeddings. Next, these image patches are fed into the UTRV Transformer to train. In training, each image patch is flattened into a single vector by concatenating channels of all pixels, then is linearly projected into lower dimensions with principal features. Because the transformer neural network is agnostic to the structure of input patches, we use the learnable position embedding [14] for each image patch. This enables the model to learn the spatial structure of grain scattering features in each C-scan image. Next, the embedded features are fed as inputs to train the transformer encoder and the multi-layer neural network. As shown in Figure 3, this transformer encoder includes eight identical encoder blocks, and each block aims to learn scattering features in ultrasonic images in a data-efficient method. In each transformer encoder block, it consists of alternating sub-layers of multi-head attention and position-wise fully connected feed-forward network [13]. And the layer normalization is applied before every sub-layer and the residual connection is applied after every sub-layer, which is shown in Figure 4. Next, these learned features are fed as inputs to the fully connected neural network to classify material textures. In this paper, the UTRV Transformer was trained to recognize material textures using C-scans and achieves 98.99% training accuracy and 96.15% testing accuracy to characterize *Grain14, Grain24, Grain50* steel blocks.

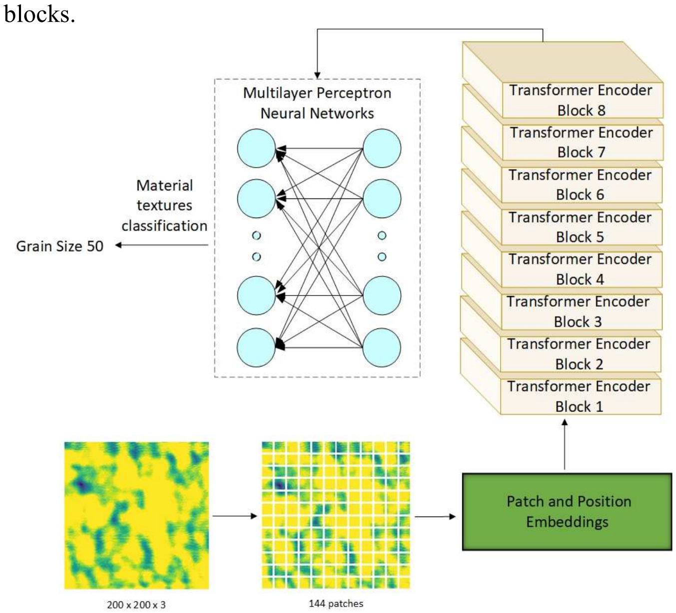

Figure 3. UTRV Transformer Architecture

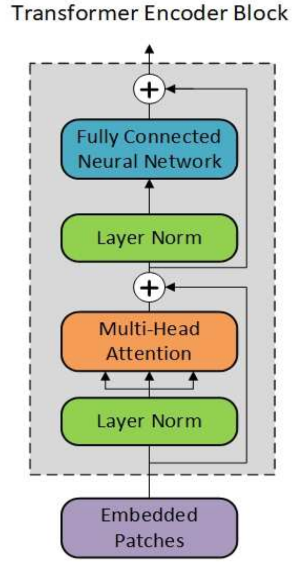

Figure 4. Transformer Encoder Block

As we can see, Figure 4 shows the structure of the transformer encoder block. The embedded patches are fed as inputs into the encoder, followed by layer normalization [13]. The layer normalization is used to reduce the training time by normalizing activities from neurons. Next, the multi-head attention mechanism is utilized to integrate information across the entire ultrasonic image and combine projected grain scattering features from different subspaces and positions. Then the joint features are processed in parallel. By applying this attention mechanism, the complexity per layer and sequential operations are reduced [13]. Next, the residual connection is used to enhance the training performance with fast training and higher accuracy. After processed by another layer normalization, the grain scattering features are fed as inputs into a two-layer fully connected neural network to further extract position features, followed by a residual connection for performance optimization. Therefore, by using the transformer encoder block, the grain scattering features are learned efficiently and parallelized in computation. This reduces the computation cost for fast training. In this study, each transformer encoder block is designed with four-head attention with 0.1 dropouts, followed by a two-layer fully connected neural network of 128 neurons and 64 neurons.

B. *Texture Recognition using the Deep Convolutional Neural Networks*

To evaluate the performance of the UTRV Transformer on material texture recognition, we trained two state-of-art deep-CNNs: VGG16 (VGGNet) and InceptionResNetV2 [8], to recognize material textures using experimental C-scan images. The VGG16 is a large size deep-CNN with 138,357,544 parameters and a CNN base size of 528 MB. In the ImageNet Large Scale Visual Recognition Challenge (ILSVRC), VGG16 achieves 92.7% top-5 test accuracy on the ImageNet dataset [8]. It includes 16 weight layers and follows the architecture pattern of 3x3 convolution kernel and 2x2 max pooling. In addition, the InceptionResNetV2 is a middle size deep-CNN with 55,873,736 parameters and a CNN base size of 215 MB.

In the ILSVRC, the InceptionResNetV2 achieves 95.3% top-5 test accuracy on the ImageNet dataset [8]. The InceptionResNetV2 consists of 164 layers deep and is optimized using pointwise convolution, batch normalization, convolution factorization, and residual connections [8]. In this study, VGG16 and InceptionResNetV2 were trained using the transfer learning with the same 2 dense layers as the UTRV Transformer shown in Figure 3 using C-scan images. These two deep CNNs achieve 100% training accuracy and 92.98%, 94.40% testing accuracy separately to characterize *Grain14, Grain24, Grain50* steel blocks.

C. Texture Recognition using the Support Vector Machine

For further performance analysis, we trained a non-linear Support Vector Machine (SVM) [16] using the kernel trick to classify the material textures. The kernel-SVM generally scales well in classifying high-dimensional data such as images. In this study, we use the radial basis function kernel [16] which is mostly used in non-linear SVM classification. The kernel-SVM was trained using the experimental C-scan images and achieves 93.46% training accuracy and 91.25% testing accuracy to characterize *Grain14, Grain24, Grain50* steel blocks.

TABLE 1. PERFORMANCE COMPARISON FOR MATERIAL TEXTURES CLASSIFICATION USING UTRV TRANSFORMER

Model	Training Accuracy	Validation Accuracy
UTRV Transformer	98.99%	96.15%
InceptionResNetV2	100%	94.40%
VGG16	100%	92.98%
Kernel-SVM	93.46%	91.25%

Table I shows material textures classification results to characterize *Grain14, Grain24, Grain50* steel blocks using the UTRV Transformer, deep CNNs and kernel-SVM. As we can see that the UTRV Transformer achieves 98.99% training accuracy and 96.15% testing accuracy, which outperforms InceptionResNetV2, VGG16 in testing accuracy and kernel-SVM in both training accuracy and testing accuracy. Although these two deep-CNNs have higher training accuracy than the UTRV Transformer, the deep-CNNs were pre-trained using transfer learning on large datasets (ImageNet) followed by further training with the experimental ultrasonic images. The transfer learning enhances the classification accuracy while reducing the computational cost. However, the pre-trained deep-CNN models perform texture recognition with lower accuracy in testing and are heavy on execution due to the large number of parameters that need to be trained. In addition, it is time-consuming and more prone to overfitting to training these two deep-CNN models without using the transfer learning.

IV. CONCLUSION

The transformer neural networks utilize a simple network architecture that replaces the convolutions and recurrence entirely with the attention mechanism. And the attention mechanism largely reduces training time for data-efficient applications. In this study, we aim to build a data-efficient automatic system for ultrasonic NDE applications to recognize material textures accurately. The performance of using the transformer neural networks to learn grain scattering features and classify material textures is investigated. In addition, we proposed the Ultrasonic Texture Recognition Vision Transformer: UTRV Transformer, to recognize material textures using ultrasonic images. To examine experimental results, an ultrasonic testbed platform is assembled, and C-scan ultrasonic testing is created to obtain the backscattered signals from three heat-treated steel blocks with different grain sizes. The 2D C-scan images are constructed from acquired backscattered signals (3D data cube), then are used to train the UTRV Transformer neural network. To evaluate the performance of the UTRV Transformer on material texture recognition, we trained state-of-art machine learning models: two deep convolutional neural networks and the kernel support vector machine using the ultrasonic C-scan images. A series of experiments were conducted to collect ultrasonic images to train the machine learning models. Our results show that the UTRV Transformer achieves 98.99% training accuracy and 96.15% testing accuracy, which outperforms InceptionResNetV2, VGG16 in testing accuracy and kernel-SVM in both training accuracy and testing accuracy. To improve the texture recognition performance using the UTRV Transformer, future work will involve utilizing the data augmentation, deploying more transformer encoder blocks, and further fine-tuning hyperparameters in this transformer neural network.

REFERENCES

[1] J. Saniie and N. M. Bilgutay, "Quantitative grain size evaluation using ultrasonic backscattered echoes," *The Journal of the Acoustical Society of America 80, 1816,* vol. 80, no. 6, pp. 1816-1824, 1986.

[2] J. Saniie, T. Wang and N. Bilgutay, "Statistical Evaluation of Backscattered Ultrasonic Grain Signals," *The Journal of the Acoustical Society of America,* vol. 84, p. 400, 1998.

[3] T. Wang, J. Saniie and X. Jin, "Analysis of low-order autoregressive models for ultrasonic grain signal characterization," *IEEE Transactions. on Ultrasonics, Ferroelectrics, and Frequency Control,* vol. 38, no. 2, 1991.

[4] J. Saniie, T. Wang and N. M. Bilgutay, "Analysis of homomorphic processing for ultrasonic grain signal characterization," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control,* vol. 36, no. 3, 1989.

[5] J. Saniie and N. M. Bilgutay, "Grain Size Evaluation Through Segmentation and Digital Processing of Ultrasonic Backscattered Echoes," in *IEEE 1984 Ultrasonics Symposium*, Dallas, 1984.

[6] X. Zhang, T. Gonnot, J. Saniie. "Real-Time Face Detection and Recognition in Complex Background," *Journal of Signal and Information Processing*, Volume 8, 2017.

[7] X. Zhang, J. Saniie, A. Heifetz, "Neural Learning Based Blind Source Separation for Detection of Material Defects in Pulsed Thermography Images," *IEEE International Conference on Electro-Information Conference (EIT)*, 2020.

[8] W. Rawat and Z. Wang. "Deep convolutional neural networks for classification: A comprehensive review," *Neural computation*, 2017.

[9] M. Unluturk and J. Saniie, "Neural networks for ultrasonic grain size discrimination," in *IEEE Ultrasonics Symposium*, San Antonio, 1996.

[10] M. S. Unluturk, P. Simko and J. Saniie, "3D-2 Frequency Discrimination of Ultrasonic Signal Using Neural Networks for Grain Size Estimation," in *IEEE Ultrasonics Symposium Proceedings*, New York, 2007.

[11] J. Saniie, M. Unluturk, T. Chu, "Frequency discrimination using neural networks with applications in ultrasonics microstructure characterization," in *IEEE 1992 Ultrasonics Symposium Proceedings*, Tucson, 1992.

[12] M. S, Unluturk and J. Saniie, "Neural Network Based Order Statistic Processing Engines," *Journal of Signal and Information Processing*, vol. 3, no. 1, pp. 30-34, 2012.

[13] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. "Attention is all you need," *In NIPS*, 2017.

[14] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, N. Houlsby. "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," *ICLR* 2021.

[15] PM. Cheng, HS. Malhi. "Transfer learning with convolutional neural networks for classification of abdominal ultrasound images*," Journal of Digital Imaging*. 30:234–243, 2017.

[16] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, A. Lopez. "A comprehensive survey on support vector machine classification: Applications, challenges and trends," *Neurocomputing*, 2020.