A defect detection method for topological phononic materials based on few-shot learning

Beini Zhang\textsuperscript{1,2}, Xiao Luo\textsuperscript{3}, Yetao Lyu\textsuperscript{4}, Xiaoxiao Wu\textsuperscript{3,∗} and Weijia Wen\textsuperscript{1,2,3,∗}

\textsuperscript{1} Advanced Materials Thrust, The Hongkong University of Science and Technology, Guangzhou, 511458, People’s Republic of China
\textsuperscript{2} HKUST Shenzhen-Hong Kong Collaborative Innovation Research Institute, Futian, Shenzhen, 518000, People’s Republic of China
\textsuperscript{3} Department of Physics, The Hong Kong University of Science and Technology, Hong Kong, 999077, People’s Republic of China
\textsuperscript{4} Robotics and Artificial Intelligence Division, Hong Kong Productivity Council (HKPC), Hong Kong, 999077, People’s Republic of China

∗ Authors to whom any correspondence should be addressed.
E-mail: phwen@ust.hk and xwuan@connect.ust.hk

Keywords: topological materials, few-shot learning, deep learning, defect detection, structure classification

Abstract
Topological phononic materials have been widely used in many fields, such as topological antennas, asymmetric waveguides, and noise insulation. However, due to the limitations of the manufacturing process, topological protection is vulnerable to some severe defects that may affect the application effect. Therefore, the quality inspection of topological materials is essential to ensure reliable results. Due to the low contrast and irregularity of defects and the similarity of topological phononics, they are difficult to recognize by traditional image processing algorithms, so manual detection is still mainstream at present. But manual detection requires experienced inspectors, which is expensive and time-consuming. In addition, topological materials are expensive to produce, and there is no large publicly available dataset, but deep learning usually relies on large datasets for training. To solve the above problems, we propose an automatic deep learning topology structure defect detection method (ADL TSDM) in this work, which could classify not only the structure of topological materials but also detect the defects of topological phononics based on a small dataset. ADL TSDM exploits the prior knowledge of the topological material structure and achieves an augmentation factor of more than 100 times through the random and fixed interval screenshot algorithm, thus enabling the training of deep neural networks with only two raw data. For defect detection, ADL TSDM has an accuracy of more than 97% and improves detection speed by more than 38% compared with manual detection. For structure classification, ADL TSDM can achieve an accuracy of over 99% and seven times faster speed compared with manual classification. Besides, the detection standard of ADL TSDM is unified, so the accuracy will not be affected by the experience of the inspectors, which has more potential in high-throughput industrial applications.

1. Introduction
In recent years, the discovery and development of topological phononic materials [1–5] have offered a soundly robust platform in acoustics for diverse applications, such as phononic antennas [3, 6–8], asymmetric waveguides [9–11], and noise insulation [12–14]. Because of the topological protection, their performance is protected against fabrication errors to a certain extent determined by their specific topological phases, such as valley-Hall phases and pseudo-spin phases [15–17]. Nevertheless, unlike the theoretical experiments where the edges tend to be infinitely sharp, the manufacturing process has technical limitations that make the edges stick smooth, so the topological protection is more fragile in the practical experimental environment [17, 18]. For example, 3D printing is one of the most common techniques to fabricate topological materials [19]. And this technique often results in unexpected defects such as deformation, loss of filament [20], and abnormal hole penetrations [21], which may break topological
Figure 1. Transport of topological edge states and defects. (a) Schematic of the metastructure comprised of the topological phase and trivial phase. A unit cell of the topological phase and trivial phase is highlighted by the dashed hexagon and magnified to show details, respectively. (b) Calculated projected band structure. Topological edge states of opposite pseudospins emerge in the bandgap. Shaded regions correspond to the projected bulk bands. (c) Pressure field map of excited topological edge state at 10.7 kHz. The star denotes where we place the point source to excite the metastructure. (d) Pressure field map but there is a defect in the interface between the topological phase and trivial phase, as denoted by the dashed rectangle. Inset shows an enlarged view of the region with the defect.

It is difficult to detect defects by traditional algorithms because the topology structures usually have features of high densities, complex shapes, and large quantities [25, 26]. Therefore, it is essential but challenging to inspect the samples before the experiment thoroughly, and the inspection is mainly performed manually [27–29]. The phononic crystals in experiments usually have unit cells of hundreds to thousands, so manual inspection is slow and inefficient, which is difficult to meet the requirements of high-throughput assessment. With the development of deep learning, the defect detection based on convolution neural networks has been successfully implemented in the fields such as surface, manufacturing, and railway track [30–34]. However, traditional deep learning models are usually trained based on large datasets to complete effective training [35–37]. Many traditional fields, such as medical and materials, have relatively high data acquisition costs, so the importance of few-shot learning strategies increases daily [38, 39]. Therefore, we proposed an automatic deep learning topology structure defect detection method (ADL-TSDM) based on the Mask region-based convolutional network method (Mask R-CNN) and random and fixed interval screenshot algorithm (RFISA) in this work. For topological materials with high production costs, the RFISA data augmentation algorithm proposed in this paper can achieve augmentation over 100 times for a single image, thus significantly reducing the cost of data collection. And Mask R-CNN can precisely identify the contours of different phononics, which could accurately recognize topological phases and defects with high throughput.

Besides, to demonstrate the importance of the quality inspection, we consider the transport of the topological edge states in different situations with COMSOL Multiphysics, which is a commercial finite element method software package (https://comsol.com/). As shown in figure 1(a), the metastructure of the sonic crystal consists of triangular scatters with refractive index \( n \) periodically arranged in air, modelled with the density \( \rho_{\text{air}} = 1.2 \text{ kg m}^{-3} \) and speed of sound \( c_{\text{air}} = 343 \text{ m s}^{-1} \). The geometric parameters are \( A_0 = 12 \text{ mm} \) and \( l_0 = 5 \text{ mm} \). Depending on whether the distance of the holes in the unit cell is expanded or contracted, they can form topological or trivial phase due to the pseudospin-Hall effect [40, 41]. More specifically, for the distance \( r \), namely, the distance of a triangular scatter to the center of the unit cell, \( r > A_0/\sqrt{3} \) \((r < A_0/\sqrt{3})\) corresponds to the topological (trivial) phase. Regarding the metastructure of both phase with \( r_1 = 1.1A_0/\sqrt{3} \) and \( r_2 = 0.9A_0/\sqrt{3} \), we calculate their projected band structure along \( k_x \).
direction assuming a refractive index $n = 3$ for the triangular scatters. The calculated projected band structure in figure 1(b) reveals topological edge states in the bandgap arising from the pseudospin-Hall effect. For a perfect sample, the excited edge state shows a smooth transmission along with the interface as demonstrated in figure 1(c). However, when there are severe defects as considered in this work, even only in a small area, there can be very strong backscattering. To confirm this point, we consider here a simple case, that is, two scatters are fabricated with severe deformation. The calculated field maps are shown in figure 1(d), which shows a strong backscattering by the defect, and the transport of the topological edge states is significantly diminished. The above simulation results confirm the fact that the performance of the material can be greatly affected by defects, which illustrates the importance of quality inspection and structure analysis for topological materials.

2. Materials and methods

2.1. Dataset

Generally speaking, deep learning models require at least hundreds of images to complete training, and the preparation of high-precision topological materials is expensive, making it difficult to collect in large quantities during the design phase. After taking advantage of the high density and repeatability characteristics of the topological phononic material, RFISA can simulate a hybrid board design with different distribution states and take random screenshots, which can greatly reduce the data collection cost while ensuring the effectiveness of model training. The experimental dataset can be divided into training, validation, and testing, among which the training and validation datasets are used for feeding the model, and the testing dataset is used to evaluate the performance of the model. Both the training dataset and the validation dataset were augmented by RFISA from two original images of 1600 × 1376 (width × height) pixels, one with a pure topological design and the other with a pure trivial design. For the test dataset, we did not use any data augmentation methods, and the images were obtained from two batches of hybrid samples (trivial and topological) by random manual photography. For the defect recognition, the training dataset contains 220 images, the validation dataset contains 11 images, and the testing dataset contains 121 images. Besides, the image sizes of each dataset are 98 × 79 (width × height) pixels. For structure classification, the training dataset contains 270 images, the validation dataset contains 20 images, and the testing dataset contains 90 images. Besides, the image size of each dataset is 416 × 360 (width × height) pixels. The images of the training and validation datasets are all augmented by RFISA through two original images with pure topological and trivial phases, respectively. In order to ensure the training effect, the images in the testing dataset are sourced from the different experiments, which are independent of the training and validation datasets and adopt the hybrid design of the topological and trivial phase. Since each basic unit of the topological material is characterized by high structural repeatability and high density, the lateral and longitudinal distances between the basic units remain constant for a single-phase design. Therefore, RFISA takes advantage of the structural features of topological materials to perform random position interceptions for images of single-phase designs, which enables a small-sample-based neural network training by augmenting the original image with an augmentation factor of more than 100 times.

2.2. Methods

The main flow chart of the ADLTSMD proposed in this work is shown in figure 2. First, ADLTSMD will intercept the images of experiments through the RFISA, which can not only realize data augmentation but also improve the relative proportion of detection targets. Second, we need to annotate the data generated by RFISA using the software VIA (https://robots.ox.ac.uk/vgg/software/via/via_demo.html), and the ground truth of the annotation is the defect area or topological phononics. Third, ADLTSMD will conduct the training of Mask R-CNN by annotating files and inputting data. The successfully trained models could achieve two types of functions: (1) structure classification, which can effectively distinguish the topological phase from the trivial phase for the inputting data. (2) Defect recognition, which can recognize the defects of inputting data. Finally, ADLTSMD will output the number of each class for structure classification, and calculate the average defect rate for defect recognition. And the defect rate is calculated as $DR = \frac{DIN}{N}$, where $DR$ represents the defect rate, $DIN$ represents the number of images containing defects, and $N$ represents the total number of detected images. And the average accuracy (ACC) is calculated as $ACC = \frac{TN + TP}{TN + TP + FN + FP}$, where $TP$ represents the number of true positive, $FP$ represents the number of false positive, $TN$ represents the number of true negative, and $FN$ represents the number of false negative. Besides, ADLTSMD will automatically calculate the number of defects, the total number of defects, the total number of detected images, and the defect rate, then save the above results to files in the .xlsx table format according to the corresponding images.
2.3. Random and fixed interval screenshots

The RFISA proposed in this paper utilizes the repetitive structural features of topological materials, which could implement random cropping and hybrid structure simulation functions. For random cropping functions, we expected it to have a random starting position and a fixed image size. So, RFISA fixed the absolute interval with the `range()` function of `python3`, where it has range (start point, end point, step) three parameters. For example, `range (10, 1000, 10)` represents an array at regular intervals of 10 in the range 10 to 1000, which is [10, 20, 30, 40, ..., 1000]. Then used the `random.choice()` function of `python3` to realize the randomly selected values from the array produced by `range()` function, and the selected values will be used as the left-top point of the crop areas or the paste area. The detailed parameters of the start point and end point of the `range()` function could be measured according to the image size, and the step parameters could be calculated by following formulas:

\[
W_{\text{crop}} = \frac{(x_n - x_1)}{(n - 1)} \quad (1)
\]

\[
H_{\text{crop}} = \frac{(y_n - y_1)}{(n - 1)}. \quad (2)
\]

Besides, RFISA is able to change purely topological or purely trivial structures to hybrid structure images by replacing the smallest units with other structures. In the augmentation example of figure 3, two original images with pure topological design and pure trivial design were used in figures 3(a) and (c), respectively. And the smallest copied targets with pure topological structure (figure 3(b)) were obtained by random cropping from figure 3(a). The augmented results with hybrid structure (figures 3(d) and (e)) can be simulated by replacing the starting position with augmented results with a random position and fixed interval again to obtain the final crop results (figure 3(f)). For the hybrid simulation function, since the structure of phononic materials has different starting positions of the smallest units in odd and even rows (as shown in figure 3(c)), we need to determine the starting coordinate points by judging the parity of the
rows separately. The $y$ value of the left-top point of the paste area is shown in the formula (3), where $k$ is any integer greater than or equal to zero and less than the number of picture lines. And the $x$ value of the left-top point of the paste area is shown in the formula (4), where $t$ is any integer greater than or equal to zero and less than the number of picture columns. And the corresponding code of the algorithm is uploaded in the following link: https://github.com/boco927/ADLTSMD.git.

In detail, due to the difference in the starting positions of odd-numbered lines and even-numbered lines of topological phononic materials, we need to prejudge the number of lines where the paste area is located first: in which the mod represents the remainder operation, $I_w$ represents the interval of the width, $I_h$ represents the interval of the height, and the calculation method of $I_w$ and $I_h$ is similar with $W_{\text{crop}}$ and $H_{\text{crop}}$, respectively. If $k \mod 2 = 0$ (then $k = 2, 4, 6, \ldots$), and the paste area $(y_n = y_3, y_5, y_7, \ldots)$ will belong to the odd line, so the start value will be the $x_2$. If $k \mod 2 \neq 0$ (then $k = 1, 3, 5, \ldots$), and the paste area $(y_n = y_2, y_4, y_6, \ldots)$ will belong to the even line, so the start value will be the $x_1$. The values of $t$ are similar to those of $k$, and both belong to the positive integer interval.

$$y_n = y_1 + k \cdot I_h \quad (k = 1, 2, 3, \ldots) \quad (3)$$

$$x_m = \begin{cases} x_2 + t \cdot I_w & \text{if } k \mod 2 = 0 \\ x_1 + t \cdot I_w & \text{if } k \mod 2 \neq 0. \end{cases} \quad (4)$$

![Figure 3. An augmentation example of the RFISA. (a) Image with purely topological structure. (b) Copied targets. (c) Original image with purely trivial structure. (d) Augmented results of odd rows. (e) Augmented results of even rows. (f) Crop results.](image-url)
Figure 4. The structure of Mask R-CNN. Mask R-CNN extracts features through the backbone network residual network 101 (ResNet101), followed by target region proposal through RPN based on feature maps, and then outputs target location, mask, and category through RoI, softmax and convolutional layers.

Table 1. The training parameters.

| Parameters                      | Setting                        |
|---------------------------------|--------------------------------|
| IMAGES_PER_GPU                  | 1                              |
| STEPS_PER_EPOCH                 | 100                            |
| DETECTION_MIN_CONFIDENCE        | 0.3                            |
| VALIDATION_STEPS                | 50                             |
| RPN_ANCHOR_STRIDE               | 1                              |
| RPN_ANCHOR_RATIOS               | [0.5, 1, 2]                    |
| RPN_NMS_THRESHOLD               | 0.7                            |
| WEIGHT_DECAY                    | 0.0001                         |
| LEARNING_MOMENTUM               | 0.9                            |
| LEARNING_RATE                   | 0.005                          |
| Epoch                           | 180 (structure classification), 150 (defect detection) |

With the combination usage of the random cropping and the hybrid structure simulation, RFISA can augment the original data by more than 200 times, which can effectively improve the training efficiency of the model and hugely decrease the data collection costs.

2.4. Mask R-CNN

Mask R-CNN is a two-stage instance segmentation network capable of classification and segmentation for object detection [42]. The simple structure of Mask R-CNN is shown in figure 4, which is composed of the feature extraction network, region proposal network (RPN) [43, 44], region of interest (RoI) [44–46], and fully convolution nets [47]. Among them, the feature extraction network adopts residual network (ResNet) 101 [48] as the backbone, and its main function is to extract features from the whole image. The main role of RPN is to propose candidate object bounding boxes based on the feature maps. For each candidate box, RoI will further apply classification and bounding-box regression for them to produce the final category and coordinate prediction. The Mask branch is implemented by fully convolution nets, and its function is to generate masks with detailed contours for each detection target. Besides, Mask R-CNN constructs a top–down ResNet-FPN structure for feature transfer between different scales as shown in the Mask R-CNN of figure 2, which aims at improving the detection rate of small targets. Compared with Faster region-based convolutional (Faster R-CNN), Mask R-CNN increases the mask branch to output accurate contour masks for each measurement target, and Resnet-FPN improves the accuracy of small target detection by passing the complex features extracted from the deep convolutional layer back to the shallow network. Since the topological material defects are relatively small and sensitive to size and contour shape. Compared with
Figure 5. The results of structure classification. (a) Original image. (b) Human detection result. (c) ADL TSDM detection result. (d) Category distribution image. This example is a hybrid design, which contains five topological phases and 13 trivial phases. Both the manual detection result and ADL TSDM result have accurately identified the corresponding categories.

Table 2. The result analysis of structure classification for test dataset.

| Method     | Topology | Trivial | TP    | FP    | TN    | FN    | ACC  | Time/s |
|------------|----------|---------|-------|-------|-------|-------|------|--------|
| Human      | 1093     | 528     | 1621  | 0     | 0     | 0     | 100% | 60.3   |
| ADL TSDM   | 1086     | 530     | 1612  | 4     | 0     | 9     | 99.2%| 7.6    |

other deep learning models such as Faster R-CNN and R-CNN, Mask R-CNN is not only more accurate for small targets, and its mask branching makes the contour features of each detected target available for output, which is potentially meaningful for future acoustic research scholars to accurately measure the size of defects, so we choose Mask R-CNN as our main framework. The training parameters of the Mask R-CNN model in this paper are shown in table 1.

3. Results and discussion

3.1. Structure classification

The manual classification is difficult for the hybrid designs (figure 5), which owns the design with a mixture of topological and trivial. Since the trivial and topological structures are highly similar in morphological characteristics, differing only in the center distance. Therefore, only experienced inspectors can correctly distinguish the different topological phononics in a hybrid design. However, ADL TSDM provides a new solution to identify and segment experimental materials by deep neural networks, which perform structural
detection for different topological phononics. For human classification results, we annotated the test dataset by three specialists and measured the average processing time as shown in table 2. Compared with human inspection methods for classifying material structures, ADLTSMD has an accuracy of 99.2%, which is comparable to manual inspection results. Besides, the accuracy of our model for the training dataset and validation dataset are all as high as 99%. At the same time, ADLTSMD is seven times faster than manual inspection ($\frac{Time_{human}}{Time_{ADLTSMD}}$) [49], which is more adaptable to the need for high-throughput industrialization. The detailed values for the topological and trivial phases classification are shown in figure 6(a). For example, 528.00 ($\text{Matrix}[2, 2]$) represents the number of labels that manually annotated as trivial and also predicted by ADLTSMD to be trivial, while the 32.49% represents the percentage of 528.00 after normalization ($\frac{\text{Matrix}[2, 2]}{\text{\text{Matrix}[1, 1] + \text{Matrix}[1, 2] + \cdots + \text{Matrix}[3, 3]}}$).

3.2. Defects detection
The defect-recognition results of the model are shown in figure 7 and table 3. As for human inspection results of defects detection, we invited three specialists to annotate the test dataset and measure the average processing time. According to the experimental results, ADLTSMD has an ACC rate of over 97% that is comparable to human detection in the defect detection, while the average inspection time is greatly reduced by over 38% (($\frac{Time_{human} - Time_{ADLTSMD}}{Time_{human}}$)) [50]. Besides, the accuracy of our model for the training dataset and validation dataset are all around 97%. The confusion matrix of defects detection is shown in figure 6(b). For example, 14.00 ($\text{Matrix}[2, 1]$) represents the number of labels that manually annotated as defects but wrongly predicted by ADLTSMD to be background, while the 1.67% represents the percentage of 14.00 after normalization ($\frac{\text{Matrix}[2, 1]}{\text{\text{Matrix}[1, 1] + \text{Matrix}[1, 2] + \cdots + \text{Matrix}[3, 3]}}$).
Table 3. The result analysis of defect recognition for test dataset.

| Method  | TP  | FP  | TN  | FN  | DIN | N  | DR  | ACC  | Time/s |
|---------|-----|-----|-----|-----|-----|----|-----|------|--------|
| Human   | 300 | 0   | 538 | 0   | 116 | 121| 95.9%| 100% | 11.18  |
| ADLTSMD | 286 | 10  | 528 | 14  | 116 | 121| 95.9%| 97.1%| 6.97   |

In addition, we supplemented the cross-validation experiment by randomly dividing and training the training dataset and validation dataset with 90% and 10% ratios, and then randomly selecting ten images from the test dataset for accuracy testing. After repeating the above steps five times, the final accuracy of structure classification was in the range of 99.34% ± 0.41%, and the accuracy of defect detection was in the range of 97.66% ± 1.5%. The above values are similar to the accuracy in tables 2 and 3 that no cross-validation method is used, thus effectively demonstrating the stability of the ADLTSMD.

4. Conclusion

In acoustic experiments, the quality of materials will directly affect the experimental effect. However, topological materials mainly rely on manual detection at present, which requires a high degree of professional testing personnel. To improve the efficiency of defects detection, we propose a new deep learning-based defect detection method, ADLTSMD, in this work. ADLTSMD has comparable accuracy with manual detection and improves detection time by more than 38% for defects detection. In addition, ADLTSMD can classify structural designs of materials with an accuracy of over 99%, and the classification speed is seven times faster than manual detection. With the adoption of ADLTSMD, acoustic material quality inspection efficiency can be significantly improved. ADLTSMD can reduce the dependence on expert experience, which will open more possibilities for fully intelligent acoustic industrial processes. Besides, ADLTSMD can perform fast augmentation of the original image with an augmentation factor of over 100 times through the RFISA, thus successfully challenging effective training based on less than five original images in this paper. The few-shot learning strategy adopted by ADLTSMD effectively reduces the amount of training data required for deep neural networks, thus opening more possibilities for deep learning applications in traditional fields that possess a high cost of data collection.

Acknowledgments

This research was funded by '2019 Shenzhen-Hong Kong Innovation Circle (Category D) (SZST120SC15) (SGDX2019081623160973)'，'Zhongshan-HKUST Research Program (ZSST20SC01)', 'Zhuhai Innovation and Entrepreneurship Team Project (ZHO1110406180043PWC)', and 'the Project of Hetao Shenzhen-Hong Kong Science and Technology Innovation Cooperation Zone (HZQB-KCZYB-2020083)', and supported by AIR@InnoHK Centre of Hong Kong Quantum AI Lab.

Data availability statement

The data generated and/or analysed during the current study are not publicly available for legal/ethical reasons but are available from the corresponding author on reasonable request.

ORCID iDs

Xiao Luo © https://orcid.org/0000-0002-0431-7924
Weijia Wen © https://orcid.org/0000-0003-3784-7494

References

[1] Ma G and Sheng P 2016 Acoustic metamaterials from local resonances to broad horizons Sci. Adv. 2 e1501595
[2] Ma G, Xiao M and Chan C T 2019 Topological phases in acoustic and mechanical systems Nat. Rev. Phys. 1 281–94
[3] Liu Y, Chen X and Xu Y 2020 Topological phononics: from fundamental models to real materials Adv. Funct. Mater. 30 1904784
[4] Wu X et al 2020 Deterministic scheme for two-dimensional type-II Dirac points and experimental realization in acoustics Phys. Rev. Lett. 124 075501
[5] Wu X, Meng Y, Hao Y, Zhang R-Y, Li J and Zhang X 2021 Topological corner modes induced by Dirac vortices in arbitrary geometry Phys. Rev. Lett. 126 226802
[44] He K, Gkioxari G, Dollár P and Girshick R 2017 Mask R-CNN Proc. IEEE Int. Conf. Computer Vision pp 2961–9
[45] Girshick R 2015 Fast R-CNN Proc. IEEE Int. Conf. Computer Vision pp 1440–8
[46] Dai J, He K and Sun J 2016 Instance-aware semantic segmentation via multi-task network cascades Proc. IEEE Conf. Computer Vision and Pattern Recognition pp 3150–8
[47] Long J, Shelhamer E and Darrell T 2015 Fully convolutional networks for semantic segmentation Proc. IEEE Conf. Computer Vision and Pattern Recognition pp 3431–40
[48] Xie S, Girshick R, Dollár P, Tu Z and He K 2017 Aggregated residual transformations for deep neural networks Proc. IEEE Conf. Computer Vision and Pattern Recognition pp 1492–500
[49] Riedmiller M and Braun H 1993 A direct adaptive method for faster backpropagation learning: the RPROP algorithm IEEE Int. Conf. Neural Networks (IEEE) pp 586–91
[50] Tang Z, Jiang L, Zhou J, Li K and Li K 2015 A self-adaptive scheduling algorithm for reduce start time Future Gener. Comput. Syst. 43–44 51–60