applied sciences

MDPI

Article

Classification of Intensity Distributions of Transmission Eigenchannels of Disordered Nanophotonic Structures Using Machine Learning

Raktim Sarma [1,2,*,†], Abigail Pribisova [1,†], Bjorn Sumner [2] and Jayson Briscoe [1]

[1] Sandia National Laboratories, Albuquerque, NM 87123, USA
[2] Center for Integrated Nanotechnologies, Sandia National Laboratories, Albuquerque, NM 87123, USA
* Correspondence: rsarma@sandia.gov
† These authors contributed equally to this work.

Abstract: Light-matter interaction optimization in complex nanophotonic structures is a critical step towards the tailored performance of photonic devices. The increasing complexity of such systems requires new optimization strategies beyond intuitive methods. For example, in disordered photonic structures, the spatial distribution of energy densities has large random fluctuations due to the interference of multiply scattered electromagnetic waves, even though the statistically averaged spatial profiles of the transmission eigenchannels are universal. Classification of these eigenchannels for a single configuration based on visualization of intensity distributions is difficult. However, successful classification could provide vital information about disordered nanophotonic structures. Emerging methods in machine learning have enabled new investigations into optimized photonic structures. In this work, we combine intensity distributions of the transmission eigenchannels and the transmitted speckle-like intensity patterns to classify the eigenchannels of a single configuration of disordered photonic structures using machine learning techniques. Specifically, we leverage supervised learning methods, such as decision trees and fully connected neural networks, to achieve classification of these transmission eigenchannels based on their intensity distributions with an accuracy greater than 99%, even with a dataset including photonic devices of various disorder strengths. Simultaneous classification of the transmission eigenchannels and the relative disorder strength of the nanophotonic structure is also possible. Our results open new directions for machine learning assisted speckle-based metrology and demonstrate a novel approach to classifying nanophotonic structures based on their electromagnetic field distributions. These insights can be of paramount importance for optimizing light-matter interactions at the nanoscale.

Keywords: machine learning; classification; disordered nanophotonics; open channels

check for updates

Citation: Sarma, R.; Pribisova, A.; Sumner, B.; Briscoe, J. Classification of Intensity Distributions of Transmission Eigenchannels of Disordered Nanophotonic Structures Using Machine Learning. *Appl. Sci.* **2022**, *12*, 6642. https://doi.org/10.3390/app12136642

Academic Editor: Saulius Juodkazis

Received: 25 May 2022
Accepted: 25 June 2022
Published: 30 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Light-matter interactions at the nanoscale are primarily dictated by the strength and spatial distribution of the electromagnetic fields inside nanophotonic structures. The ability to achieve a designable electromagnetic field distribution inside a photonic structure is vital when designing photonic devices for a wide variety of applications, ranging from sensing, optoelectronics, and nonlinear optics to quantum communications. The inverse design of a photonic structure with a desired field distribution requires physics-inspired and intuition-based approaches involving solving Maxwell equations [1]. Remarkable success has been demonstrated using these traditional techniques; however, given the increasing complexity of nanophotonic devices and advancements in nanofabrication, such trial-and-error approaches are now becoming computationally expensive and time consuming. Recently, a growing interest in exploring new optimization-based and data-driven-based approaches to inverse design nanophotonic structures has emerged [2].

The main optimization-based approaches that have been utilized so far in nanophotonics include evolutionary methods, such as genetic algorithms and particle swarming [3–5], simulated annealing [6], and gradient-based methods [7–10]. For data-driven approaches, machine learning techniques have been primarily used for the inverse design of nanophotonic structures. A general advantage of using data-driven methods over optimization-based approaches is that in data-driven approaches patterns can be learned even in very diverse datasets and a solution can be found extremely quickly after the learning phase. This behavior is advantageous for the simultaneous design of nanophotonic devices for different applications, such as designing a sensor for multiple wavelengths for which a common database can be used for training. Although not extensively investigated, another advantage of data-driven approaches is that they allow for the classification of different nanophotonic structures based on their electromagnetic field distributions or other properties. To date, machine learning has primarily been utilized to establish mappings between geometric parameters of nanophotonic structures and the reflection/transmission/absorption spectra [11–15]. However, the actual electromagnetic fields inside the structures are also of great importance, as they dictate the strength of light-matter interactions. For example, in nonlinear optical media, the nonlinear overlap integral, which dictates the nonlinear wave mixing, depends on the field distribution. Similarly, for sensing applications, in addition to the Purcell enhancement, the field distribution is crucial in determining the overall sensitivity of the device. The classification of nanophotonic structures based on their electromagnetic field distributions or intensity distributions is a vital preliminary task in the inverse design of nanophotonic structures. In this preliminary stage, classification allows for reduction of the investigated application-specific parameter space and can provide a new and efficient way of extracting information from complex datasets, especially those characteristics that may be difficult to extract purely from visualization-based analysis.

In this paper, we address this classification based on field distributions of nanophotonic structures and demonstrate, for the first time, the classification of high transmission eigenchannels of two-dimensional disordered systems using spatial intensity distributions. Disordered systems were chosen for this study because the speckle-like intensity distributions are extremely complex with large random spatial fluctuations that arise from the interference between the multiply scattered waves. Such systems are excellent candidates for this study, as a classification based purely on visualization is not possible. We demonstrate that by using machine learning techniques and by simultaneously looking at the intensity distribution of the transmission eigenchannels and transmitted speckle-like intensity distribution, it is possible to classify transmission eigenchannels.

Spatial profiles of the transmission eigenchannels of disordered systems depend on scattering strength, which dictates the energy density distribution inside the medium. The ability to classify spatial profiles is a viable technique to determine the strength and type of disorder in the system, which is of great importance for numerous applications ranging from random lasing [16] and biomedical sensing [17], white LEDs [18], to photovoltaics [19]. In addition, using transmitted speckle-like intensity distributions to perform classification of disorder strength using machine learning techniques can lead to advancements in speckle-based metrology. While efforts in this direction have been previously investigated, such as looking at speckle contrast and speckle statistics [20,21], this is, to the best of our knowledge, the first time that machine learning techniques have been utilized to perform classification and identification of disorder strength using the spatial distribution of energy densities.

Finally, we envision that our approach to machine learning-based classification could be generalized to other planar photonic systems. For example, metasurfaces are particularly promising, as the field distribution inside the meta-atoms that constitute the metasurface can also be expressed as a sum of multipoles, where the weights of these multipoles can be used as the input vector when performing classification [22].

2. Materials and Methods

Figure 1a shows an example of a two-dimensional disordered waveguide considered for this study. The disordered waveguides are assumed to be fabricated out of a high refractive index dielectric medium with effective refractive index n_e = 2.85. The disorder was introduced by randomly positioning air holes of radius (R = 150 nm) within the waveguide. To study disordered structures with different degrees of disorder, we changed the number (N_s) of the air holes within the waveguide; structures with N_s = 300, 400, and 500 were simulated. In these calculations, the values of refractive indices were chosen to emulate two-dimensional disordered silicon waveguides with air holes, which were investigated in recent experimental studies [23–27]. The vacuum optical wavelength (λ) was chosen to be 1.55 μm, as silicon is lossless at this wavelength. The width (W) of the waveguides was fixed to be 10 μm, such that the disordered waveguides supported N = 36 propagating modes, where N is given by $2W/(\lambda/n_e)$. The length (L) of the disordered region was fixed to be 20 μm. The length (L) was chosen such that the $L >> l_t$ for all the studied disordered waveguides, where l_t is the transport mean free path of the disordered system. This ensured that the wave propagation within the disordered medium was diffusive. Two empty buffer regions of 3 μm (~2λ), one at the beginning and one at the end of the disordered regions, were included in the simulation. These buffer regions were effectively silicon waveguides without scatterers and allowed us to also record the speckle-like intensity patterns of the transmitted and reflected light along with the intensity distributions of the transmission eigenchannels for our machine learning-based classification task.

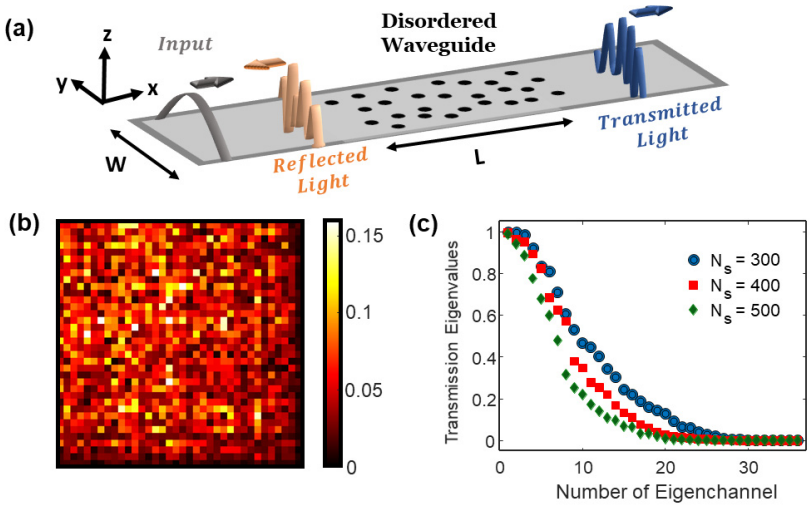


Figure 1. Schematic of the device and numerical simulations. (**a**) Schematic of the disordered waveguide studied in this paper. The grey region corresponds to the silicon waveguide and black circles correspond to air holes; (**b**) Absolute value of a transmission matrix calculated for a single random configuration with N_s = 300; (**c**) Transmission eigenvalues of disordered waveguides with three different scattering strengths (N_s = 300, 400, and 500 respectively). As N_s increases, due to enhanced backscattering, the values of the transmission eigenvalues decrease.

To calculate the transmission eigenchannels of the disordered waveguides, we used the recursive Green's function method, as described in Refs. [28–30]. First, we constructed the field transmission matrix t of the disordered waveguides, where t is a $N \times N$ complex matrix. The transmission matrix of the disordered waveguide gives the output field for any arbitrary input (Figure 1b). The transmission eigenchannels are eigenvectors of the matrix $t^{\dagger}t$ and the eigenvalues are the transmittance of the corresponding eigenchannels (Figure 1c). In the lossless diffusive regime, the density of the eigenvalues follows a bimodal distribution where some channels, called "open channels", have a transmittance/transmission eigenvalue ~1 and other channels have lower transmittance. The transmittance of the eigenchannels decreases exponentially with the index of the eigenchannel (Figure 1c). In the

diffusive regime, the diffusion process is dominated by the open channels, and the average transmittance is the mean of all the transmission eigenvalues, which is proportional to the ratio of the number of open channels to the total number of propagating channels [31]. As the number of scatterers, N_s, is increased, due to enhanced backscattering, the transmission eigenvalues decrease in magnitude, which results in an overall decrease in the average transmittance (Figure 1c).

To calculate the intensity distributions of the transmission eigenchannels, we again used the recursive Green's function method [28–30]. The transmission matrix gives the output field for any arbitrary input. We used the input and output fields as the boundary conditions to further compute the intensity distributions inside the disordered waveguide for different transmission eigenchannels. The field intensity was averaged over the waveguide cross-section along y to give the evolution of the energy density along the waveguide's propagating direction x. In this study, we focused on classification of the cross-section averaged intensity distributions of the first ten transmission eigenchannels. The transmission eigenchannels were chosen such that they span high, intermediate, and low transmittance. It is well known and experimentally demonstrated [26,32] that the cross-section averaged intensities of transmission eigenchannels with high, intermediate, and low transmittance have unique spatial profiles when averaged over a large statistical ensemble of random configurations. Figure 2a shows the cross-section averaged intensity, averaged over 10,000 random configurations, of the first ten transmission eigenchannels for a disordered waveguide with N_s = 300. Indeed, when averaged over a large ensemble, we see different spatial profiles for different transmission eigenchannels; however, the change in spatial profile is subtle for transmission eigenchannels with similar transmittance, such as the two eigenchannels T_1 and T_2 with highest and second-highest transmission eigenvalues. Furthermore, this subtle difference becomes impossible to decipher when only a single configuration is considered. Figure 2b shows the calculated two-dimensional intensity distribution of a transmission eigenchannel, T_1, with highest transmittance with N_s = 300. We can clearly see the very complex speckle-like intensity distribution both within the disordered region as well as in the buffer regions. Figure 2c plots the cross-section averaged intensity from the two-dimensional plot in Figure 2b overlaid on the statistically averaged spatial profiles of the transmission eigenchannels with highest and second-highest transmittance. Because of the large fluctuations that arise due to interference between the multiply scattered waves, it becomes impossible to visually determine whether the profile shown in Figure 2c corresponds to the transmission eigenchannel with highest or second-highest transmittance. Furthermore, even basic statistical analysis, such as distributions of root mean square deviation (RMSD), is not sufficient to classify and correlate the spatial profile of the intensity distribution to a specific transmission eigenchannel. In Figure 2d, we show the histogram of RMSD values of 5,000 configurations of the transmission eigenchannels with the highest (Channel 1) and second-highest transmittance (Channel 2). Both histograms have very similar spread and shape, making it difficult to correlate the intensity distribution to either of the transmission eigenchannels.

To demonstrate the classification of energy density of ten transmission eigenchannels with the ten highest transmission eigenvalues, we used machine learning techniques that are known to be effective for the classification of complex data. Since transmission eigenchannels dictate not only the intensity distribution inside the disordered region but also the transmitted light through these structures, we looked at the cross-section averaged intensity of the disordered region as well as the intensity pattern both in the front and back empty buffer regions of the waveguides. All of the data for the different disordered waveguides with different numbers of scatterers were combined. The resulting dataset consisted of 750,000, 451-component vectors. Here 451-components correspond to 451 spatial points uniformly distributed along the length of the two-dimensional disordered waveguide (0–26 μm), and 750,000 corresponds to the data of cross-section averaged intensity distributions of ten transmission eigenchannels from disordered waveguides with three different disorder strengths (300, 400 and 500 scatterers, respectively), with 25,000 random config-

urations with ten transmission eigenchannels each for each disorder strength. For the classification, each intensity distribution from each of the transmission eigenchannels was assigned a port number from 0–9. Depending on the model, we represented the port number as either a single label encoding 0–9 or a one-hot vector encoding. The one-hot vector encoding was a 10-component vector representing the probability of the intensity distribution being produced by that port, where the correct port had a probability of 1.0 and all other ports had a probability of 0.0. As is typical when working with neural networks, we normalized the intensity distribution of the transmission eigenchannels of all the configurations by scaling them between 0.0 and 1.0. Figure 3a,b shows schematics of the architecture of the decision tree and neural network used in this work for the classification task.

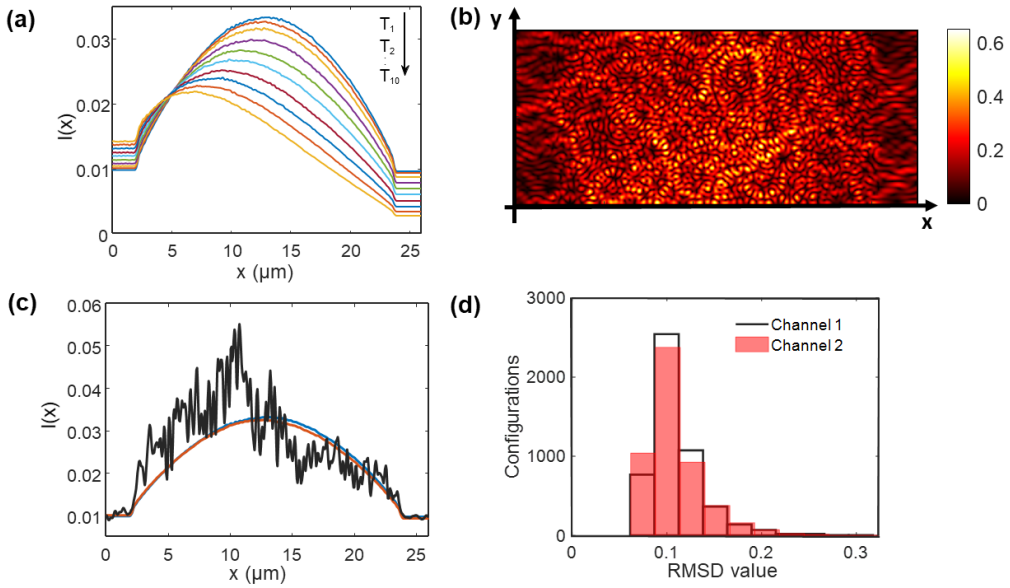


Figure 2. Numerical simulation of disordered photonic structures: (**a**) Numerically calculated spatial profiles of cross-section averaged intensity (proportional to energy density) of transmission eigenchannels with the ten highest transmission eigenvalues of disordered structures with N_s = 300. The different spatial profiles of the different transmission eigenchannels are shown in different colors. The intensities were averaged over 10,000 different random configurations. T_1 corresponds to the eigenchannel with the highest transmission eigenvalue; (**b**) Numerically calculated two-dimensional intensity distribution of the transmission eigenchannel with the highest transmission eigenvalue T_1 for a single random configuration with N_s = 300; (**c**) Cross-section averaged intensity distribution (corresponding to the two-dimensional intensity distribution shown in (**b**)) overlaid on the spatial profiles of transmission eigenchannels T_1 (blue line) and T_2 (red line) (**a**); (**d**) Histogram of the root mean square deviation (RMSD) values of 5000 configurations of the transmission eigenchannels with the highest transmittance (Channel 1) and second-highest transmittance (Channel 2). Both histograms have very similar spread and shape.

The dataset was split into 70% training data (525,000 intensity distribution-port pairs), 6% validation data (45,000 intensity distribution-port pairs), and 24% testing data (180,000 intensity distribution-port pairs). The training data were directly used to guide the improvement of the model over time. The validation data, which the model had not seen, gave an unbiased evaluation of the training; these data did not help improve the model but were used for selection of the best-performing model. After training, the testing data, which had not been previously run through the model, were used to evaluate the final model's accuracy. For each run, the training, validation, and test sets were randomly shuffled within their sets, ensuring that the model did not optimize training and testing for a particular subset of the data.

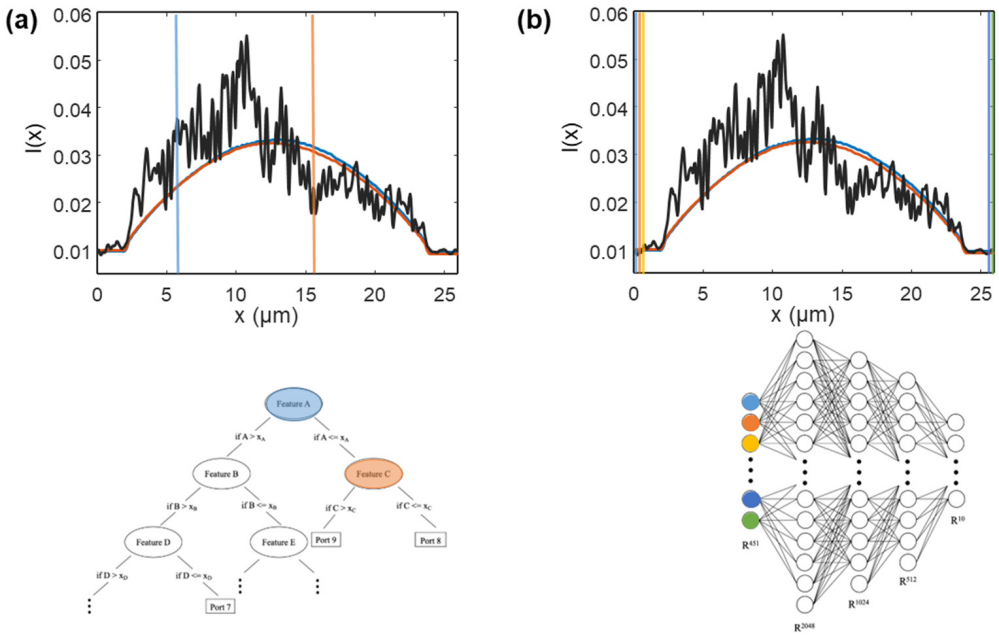


Figure 3. Diagrams explaining how the decision trees and neural networks were used for the classification task – the vertical-colored lines indicate the different intensity distribution values at different spatial points used for classification: (**a**) Diagram theoretically showing how the trained decision tree classifier can be used to classify an intensity distribution. Through training, the classifier creates the shown branching structure and learns which values of the distribution and which threshold value lead to the correct port number classification at the leaves of the tree; (**b**) Diagram theoretically showing how the trained neural network can be used to classify an intensity distribution. In the 451-neuron input layer, one value of the intensity distribution is assigned to one neuron in the order dictated by the distribution. Learned weights between the layers result in a 10-neuron one-hot vector with one port being marked as the predicted port.

Partitioning the data in this way allowed us to ensure that the model was not memorizing the intensity distribution-port parings but rather learning relevant characteristics of the intensity distributions of the transmission eigenchannels to match them to the correct ports. To analyze how much data is needed to train a high-accuracy model, we also tried training the model on balanced subsets of the whole dataset from 100% of the data (750,000 intensity distribution-port pairs) to 10% of the data (75,000 intensity distribution-port pairs). The results are shown in the Section 3.

We created different versions of the dataset with label encodings and one-hot vector encodings for comparison. We also created two multilabel datasets whose correct output encoded both the port number and the disorder strength of the disordered waveguide (i.e., the number of scatterers, N_s) that produced the intensity distribution of the transmission eigenchannels. For the first dataset, the one-hot vector encoding was a 30-component vector with the first ten positions corresponding to the ports of the disordered waveguides with 300 scatterers, the second ten positions to the disordered waveguides with 400 scatterers, and the last ten positions to the waveguides with 500 scatterers, with the ten positions corresponding to the ten ports on each configuration. For the second dataset, the one-hot vector was a 13-component vector with the first ten positions corresponding to the ports on the optical device and the last three positions corresponding to the number of scatterers (300, 400, and 500 respectively).

This is a supervised multi-class classification problem. It is supervised because the correct intensity distribution-port pairings are provided as labeled training data. It is a multi-class classification problem because there are ten different classes, corresponding to ten different ports, and each input, or intensity distribution of a transmission eigen-channel, can only belong to a single port.

Two machine learning algorithms were tested: a fully-connected neural network and a decision tree. The fully-connected neural network consisted of an input layer (451 neurons), three hidden layers (2048 neurons, 1024 neurons, and 512 neurons), and an output layer (10 neurons). For each iteration, a batch of 128 intensity distributions was passed through the model. In the input layer, one value of the intensity distribution was assigned to one neuron in the order dictated by the energy density distribution. In the hidden layers, all the neurons in the previous layer were connected with a certain weight value to each neuron in the next layer. The value that a neuron adopted in the next layer was the sum of all the neurons in the previous layer multiplied by their weights connecting them to that neuron. The ReLU activation function [33] dictated which neurons were activated and passed on their values to the next layer. In between each layer was a batch normalization layer that normalized these values based on the mean of the current batch and the added training stability. In the output layer, the softmax activation function converted these values into probabilities between 0.0 and 1.0 for each port number. Due to this setup, the one-hot vector encoding label representation was used for the fully-connected neural network. The position in the output vector with the highest probability corresponded to the predicted port number.

During training, the error between the predicted and actual probabilities was used to update the weights on the neuron connections between layers so that the most relevant features or characteristics of the intensity distributions of the transmission eigenchannels were used to make an accurate port number prediction. We implemented the fully-connected neural network using the Keras deep learning framework with TensorFlow and the weights between layers were optimized using Adam [34] and trained using categorical cross-entropy loss. We used Keras's EarlyStopping function to determine an early stop for training if the sum of errors over the entire validation set did not decrease for 50 passes of the entire training dataset through the model. Training took approximately three hours on a 32 GB GPU.

Decision tree classifiers [35] divide up the data into smaller and smaller subsets based on characteristics of the input data and the intensity distribution of the transmission eigenchannels, until they have a subset for each port with only intensity distributions that belong to that port. Generally, one-hot vector encodings are not used with decision trees, since they result in a long and narrow tree with many splits along the "trunk" of the tree. Because the one-hot 10-component vector is ten times longer than the label encoding one-component vector, this requires roughly ten times more splits to make the correct classification in all ten classes of the one-hot 10-component vector. Therefore, we utilized the label encoding. However, we also compared the accuracy between a decision tree trained with the label encodings and one trained with the one-hot vector encodings. Normalization has little effect on decision trees since there are no activation functions between branches; therefore, we used the same normalized dataset as with the neural network.

We implemented the decision tree with the DecisionTreeClassifier from the scikit-learn Python framework. We used the Gini impurity function to measure the quality of the divisions, required at least two samples to make a division, did not limit the depth or number of divisions, and considered all the features when making divisions. Training took approximately one hour on one CPU.

Since the decision tree classifier does not conduct validation during training, we verified that our model was not optimized for a particular subset of the data by using k-fold cross validation. The training and validation data, consisting of 76% of the overall data, was divided into five folds. We trained and tested the model five times with a different fold being chosen as the validation data each time, and the remaining folds were used to train the model. The testing data were then used for the final model's evaluation.

We created two versions of the multilabel dataset because the 13-component dataset would have required us to change the final activation function of the neural network to the sigmoid activation function. The softmax activation function marks only the highest

probability position in the vector as correct, while the sigmoid activation function can mark multiple positions as correct, given that they pass a given probability threshold. Therefore, to preserve the chosen neural network configuration, we tested the neural network on only the 30-component datasets and tested the decision tree on both the 13-component and 30-component datasets.

A downside of neural networks is their opacity regarding the process of classification. While the initial layer corresponds to the values of the intensity distribution, the following hidden layers contain more neurons that account for features that relate the values to themselves, like the steepness of the curve they form. However, the precise features examined cannot be determined by analyzing the neural network, making this a black-box algorithm. On the other hand, the decision tree created by the decision tree classifier shows which exact value or linear combination of values led to a certain path being taken. However, with such complex data as ours, the tree can contain thousands of branches and be difficult to analyze visually. The confusion matrices do not provide justifications for their classifications, but they do give an idea of which intensity distributions were most difficult to distinguish. At the end of the Section 3, we theorize why the decision tree classifier had problems classifying certain ports.

3. Results

Over ten runs, the best decision tree classifier achieved an average accuracy of 99.43% (standard deviation: 0.0112%), and the best fully-connected neural network had an average accuracy of 98.07% (standard deviation: 0.727%). We note here that these accuracies are comparable to the accuracies obtained using the state-of-the-art classification techniques [36]. Figures 4 and 5 show the confusion matrices for the best neural network and best decision tree, respectively. The port labels are listed along the left side and the bottom. With i referring to the rows and j referring to the columns, the ij-th entry in the confusion matrix is the number of intensity distributions classified as port i whose correct label was port j. The diagonal from top left to bottom right entries have i = j, meaning that those intensity distributions were correctly classified. The neural network's confusion matrix shows that the network had the hardest time distinguishing between ports 0 and 1, ports 1 and 2, and ports 8 and 9 (Figure 4). Alternatively, the decision tree's confusion matrix shows that the decision tree had the hardest time distinguishing between ports 0 and 1, ports 7 and 8, and ports 8 and 9 (Figure 5).

We tested the best neural network and best decision tree classifier, trained on the full dataset, against the datasets of the individual optical devices. We did not see a significant change in performance. This is expected since we were testing the trained models on subsets of the original full dataset.

We compared a decision tree trained on one-hot vector encodings and a decision tree trained on label encodings of the port number. With label encodings and one-hot vector encodings, the decision tree achieved 99.45% and 99.58% accuracy, respectively. While the branching structures of the decision trees were different, the accuracy was not significantly affected by different encodings. However, the one-hot vector encoding tree took almost twice as long to train as the label encoding tree.

Generally, neural networks require much more data than decision tree classifiers to make accurate predictions. In Figure 6, we see that, on average, the decision tree maintained 2–9% higher accuracy than the neural network when trained on iteratively smaller subsets of the full dataset. Also, the accuracies of the trained neural networks had a higher standard deviation than the trained decision tree accuracies (Figure 6). The training of an accurate neural network is partially dependent on the initial, randomly-set weights. Therefore, even with the same hyperparameters, like shape and size of the hidden layers, the accuracy can be very variable (Figure 6). As well as being more accurate, this shows that decision trees also tend to train more dependably.

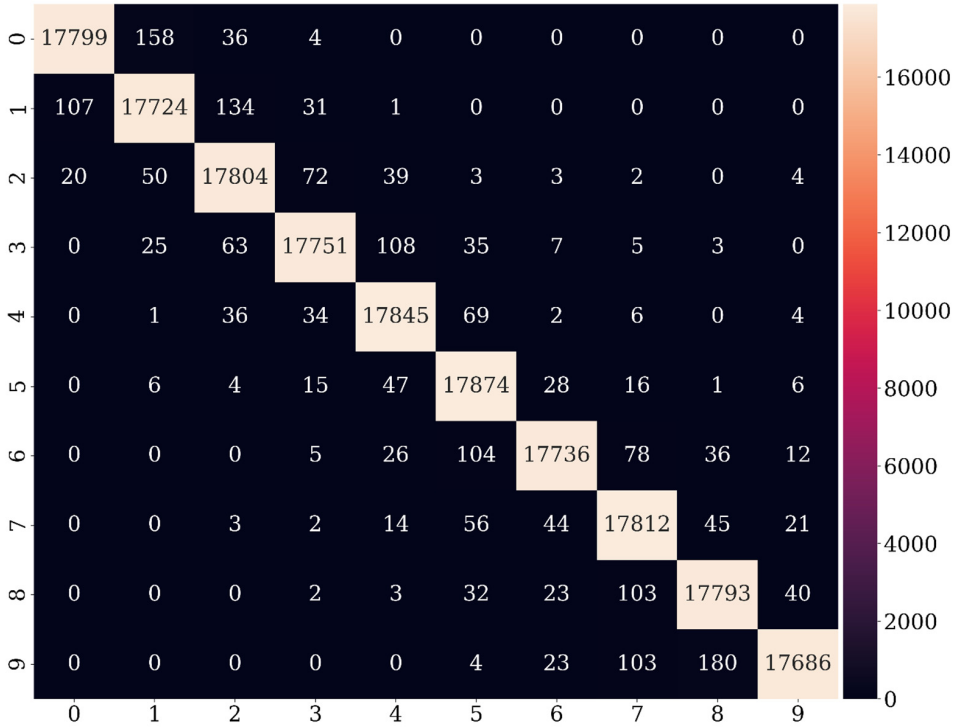


Figure 4. The confusion matrix corresponding to classification performed using a neural network trained on a dataset with 10-component output vectors. The data correspond to ten transmission eigenchannels of three devices with three different scattering strengths N_s. This classification allows for identification of the eigenchannel independent of the disorder strength. The maximum value in any cell can be 18,000.

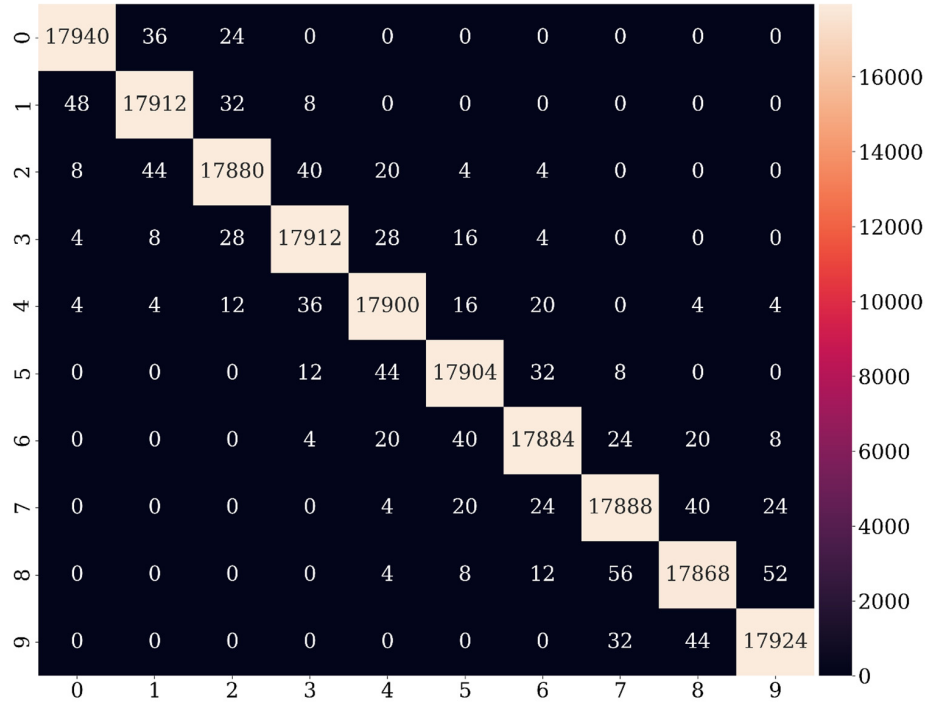


Figure 5. The confusion matrix corresponding to classification performed using a decision tree trained on a dataset with 10-component output vectors. The data correspond to ten transmission eigenchannels of three devices with three different scattering strengths N_s. This classification allows for identification of the eigenchannel independent of the disorder strength. The maximum value in any cell can be 18,000.

Percentage of data / number of data samples used for training and testing (% / samples)	Fully-connected neural network accuracy (%) over 10 runs	Decision tree classifier accuracy (%) over 10 runs	Fully-connected neural network standard deviation (%) over 10 runs	Decision tree classifier standard deviation (%) over 10 runs
100 / 750,000	98.07	99.43	0.727	0.0112
90 / 675,000	96.81	98.88	1.054	0.0137
80 / 600,000	96.25	98.29	0.644	0.0157
70 / 525,000	93.36	96.55	1.512	0.0194
60 / 450,000	89.96	94.32	2.809	0.0395
50 / 375,000	82.6	86.97	2.077	0.0564
40 / 300,000	73.61	81.04	2.171	0.0545
30 / 225,000	61.73	62.56	0.508	0.0781
20 / 150,000	39.73	37.65	1.41	0.1391
10 / 75,000	35.82	44.04	6.332	0.1682

Figure 6. A chart comparing the percentage accuracy and standard deviation of neural networks and decision trees trained on varying subsets of the testing and training datasets, ranging from 100% to 10% of the total data.

Finally, we tested the best neural network and decision tree classifier configurations on the first multilabel dataset with 30-component output vectors. The decision tree and neural network achieved 95.02% and 98.99% accuracy, respectively. Along with correctly identifying the port number, these model configurations allowed us to classify the disorder strength as well. The confusion matrices for the neural network and decision tree are shown in Figures 7 and 8, respectively. The decision tree took about 5 h to train, which is 2.5 times longer than the neural network, most likely due to longer output vectors requiring more training time to generate more divisions in the tree. The decrease in the decision tree performance might be accounted for by the increased complexity of the tree, as more output classifications, 30 classifications instead of 10 classifications, are required.

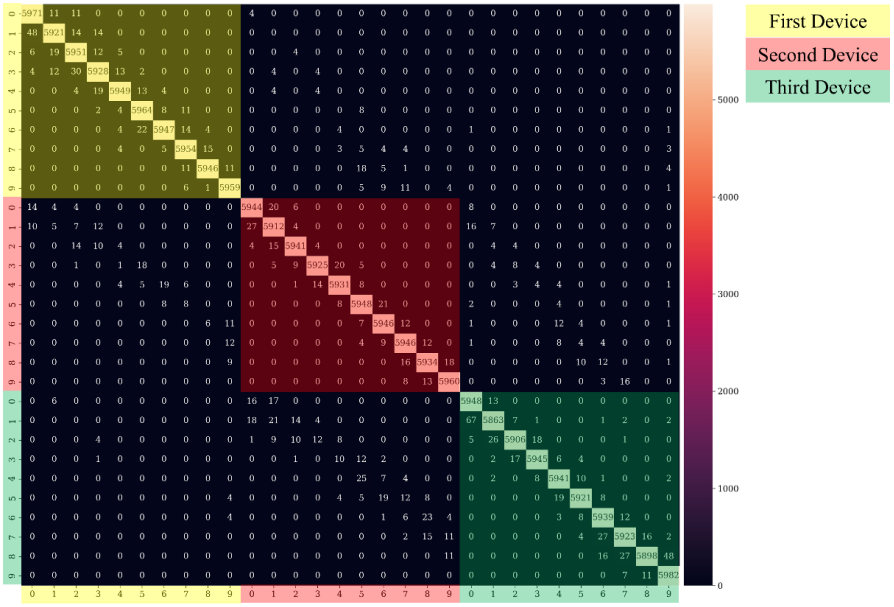


Figure 7. The confusion matrix corresponding to classification performed using a neural network trained on a dataset with 30-component output vectors. The data correspond to ten transmission eigenchannels of three devices with three different scattering strengths N_s. This classification allows for identification of both the eigenchannel as well as the disorder strength. The ports corresponding to the respective three devices are color coded. The maximum value in any cell can be 6000.

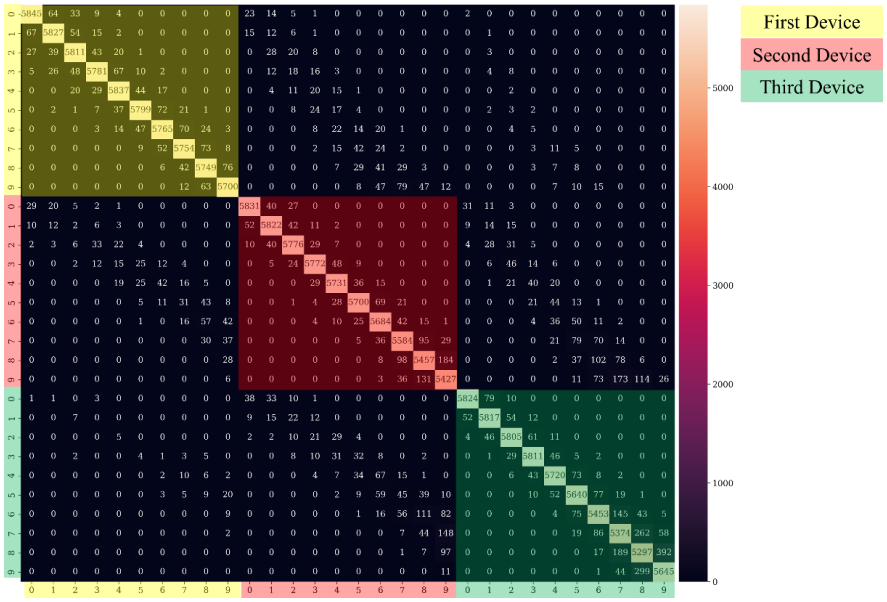


Figure 8. The confusion matrix corresponding to classification performed using a decision tree trained on a dataset with 30-component output vectors. The data correspond to ten transmission eigenchannels of three devices with three different scattering strengths N_s. This classification allows for identification of both the eigenchannel as well as the disorder strength. The ports corresponding to the respective three devices are color coded. The maximum value in any cell can be 6000.

We also tested the decision tree on the second multilabel dataset with 13-component output vectors where it achieved an accuracy of 99.44%. This decision tree took approximately 1.5 h to train. The higher accuracy and shorter training time mostly likely correspond to the smaller output vectors. Multilabel confusion matrices require a class-wise comparison with one-versus-rest binarization. This means that for three devices with ten vectors each it produces 30 two-by-two confusion matrices corresponding to the number of true negatives (not classifying an erroneous port), true positives (classifying the correct port), false negatives (not classifying a correct port), and false positives (classifying an erroneous port) for each port of each device. We chose not to include all of these confusion matrices.

Finally, to assess the importance of including the transmitted speckle-like intensity distribution in the classification process, we utilized scikit-learn's permutation importance function to identify which features, or values of the intensity distribution, were most vital to the decision tree's accurate performance. The permutation feature importance is defined as the decrease in the model's accuracy when a single feature value is randomly shuffled and indicates how much the model depends on that feature. We evaluated the permutation feature importance on three separate runs with the same configuration. Of the top 20 most important features out of 451, 12 of the 20 were the points near the end of the intensity distribution, between positions 409 and 450, corresponding to $L > 23$ μm. These points correspond to the transmitted intensity pattern in the buffer region at the end. When we calculated the mean value and standard deviation of each intensity distribution point separately for each port, we noticed that the points near the end of intensity distribution, between positions 412 and 450, consistently had the lowest standard deviation, even while their values differed significantly between ports. The features with the smallest standard deviation between samples of the same port were most important in the classification process and in distinguishing between ports. This gives some insight into the decision tree's accurate performance.

4. Conclusions and Summary

In summary, we have demonstrated, for the first time, the classification of transmission eigenchannels of individual disordered nanophotonic structures using supervised machine

learning techniques. We show that along with classification of transmission eigenchannels it is also possible to classify the photonic device based on its disorder strength for a single configuration utilizing the spatial distribution of the energy density of the transmission eigenchannels. To perform this classification, we used supervised learning methods, namely decision trees and fully connected neural networks. In both cases, we achieved an accuracy of classification greater than 99%.

While we performed the classification using disordered photonic structures, our approach is general and should be applicable to other photonic systems, such as metasurfaces. Although in this work we did not perform inverse design, our proof-of-concept results can open new directions for the inverse design of photonic devices. The classification of different structures based on different desired properties can be utilized as a first step in performing optimizations that will lead to the desired photonics devices. This can be especially important for applications in nonlinear optics, where simultaneous optimization of multiple quantities, such as field enhancements, field overlaps, and emission directionality is required [37], and where optimizing one of these quantities does not necessarily optimize the others. In such multi-parameter-based optimizations, performing the classification followed by an optimization on a smaller parameter space can be a more efficient technique for achieving an optimum design. This will be investigated in future studies.

Finally, there has recently been a growing interest in two-dimensional disordered silicon-based photonics on-chip devices for numerous applications ranging from spectroscopy and lasing to sensing. Such photonic devices are also excellent platforms for fundamental studies related to wave interference effects, such as probing Anderson localization [23,24], and mesoscopic correlations [25–27]. Since it is relatively easy to probe the energy densities of single devices in such systems by measuring the out-of-plane scattering, combining such measurements with machine learning algorithms could provide additional degrees of freedom for optimizing the light-matter interactions. This could be of paramount importance for applications such as random lasing and sensing.

Author Contributions: Conceptualization, R.S. and J.B.; methodology, R.S and A.P.; software, R.S. and A.P.; validation, R.S., A.P., B.S. and J.B.; formal analysis, R.S., A.P.; writing—original draft preparation, R.S., A.P.; writing—review and editing, R.S., A.P., B.S., J.B.; supervision, R.S., J.B.; project administration, J.B. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the U.S. Department of Energy, Office of Basic Energy Sciences, Division of Material Sciences and Engineering and performed, in part, at the Center for Integrated Nanotechnologies, an Office of Science User Facility operated for the U.S. Department of Energy (DOE) Office of Science. Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

Data Availability Statement: The data generated during the current study are available from the corresponding authors on reasonable request.

Conflicts of Interest: The authors declare that they have no conflict of interest.

References

1. Shelby, R.A.; Smith, D.R.; Schultz, S. Experimental verification of a negative index of refraction. *Science* **2001**, *292*, 77–79. [CrossRef] [PubMed]
2. Yao, K.; Unni, R.; Zheng, Y. Intelligent nanophotonics: Merging photonics and artificial intelligence at the nanoscale. *Nanophotonics* **2019**, *8*, 339–366. [CrossRef] [PubMed]
3. Goldberg, D.E.; Holland, J.H. Genetic algorithms and machine learning. *Mach. Learn.* **1988**, *3*, 95–99. [CrossRef]
4. Robinson, J.; Rahmat-Samii, Y. Particle swarm optimization in electromagnetics. *IEEE Trans. Antennas Propag.* **2004**, *52*, 397–407. [CrossRef]
5. Kennedy, J.; Eberhart, R. Particle swarm optimization. In Proceedings of the ICNN'95—International Conference on Neural Networks, Perth, Australia, 27 November–1 December 1995.

6. Kirkpatrick, S.; Gelatt, C.D.; Vecchi, M.P. Optimization by simulated annealing. *Science* **1983**, *220*, 671–680. [CrossRef]
7. Bendsoe, M.P.; Kikuchi, N. Generating optimal topologies in structural design using a homogenization method. *Comput. Methods Appl. Mech. Eng.* **1988**, *71*, 197–224. [CrossRef]
8. Bendsoe, M.P.; Sigmund, O. *Topology Optimization: Theory, Methods, and Applications*; Springer: Berlin, Germany, 2004.
9. Lalau-Keraly, C.M.; Bhargava, S.; Miller, O.D.; Yablonovitch, E. Adjoint shape optimization applied to electromagnetic design. *Opt. Exp.* **2013**, *21*, 21693–21701. [CrossRef]
10. Lin, Z.; Groever, B.; Capasso, F.; Rodriguez, A.W.; Loncar, M. Topology-optimized multilayered metaoptics. *Phys. Rev. Appl.* **2018**, *9*, 044030. [CrossRef]
11. Malkiel, I.; Mrejen, M.; Nagler, A.; Arieli, U.; Wolf, L.; Suchowski, H. Plasmonic nanostructure design and characterization via Deep Learning. *Light. Sci. Appl.* **2018**, *7*, 60. [CrossRef]
12. Nadell, C.; Huang, B.; Malof, J.; Padilla, W. Deep learning for accelerated all-dielectric metasurface design. *Opt. Express* **2019**, *27*, 27523–27535. [CrossRef]
13. Wiecha, P.R.; Arbouet, A.; Girard, C.; Muskens, O.L. Deep learning in nanophotonics: Inverse design and beyond. *Photonics Res.* **2021**, *9*, B182–B200. [CrossRef]
14. Qiu, C.; Wu, X.; Luo, Z.; Yang, H.; He, G.; Huang, B. Nanophotonic inverse design with deep neural networks based on knowledge transfer using imbalanced datasets. *Opt. Express* **2021**, *29*, 28406–28415. [CrossRef] [PubMed]
15. So, S.; Badloe, T.; Noh, J.; Bravo-Abad, J.; Rho, J. Deep learning enabled inverse design in nanophotonics. *Nanophotonics* **2020**, *9*, 1041–1057. [CrossRef]
16. Cheng, X.; Genack, A. Focusing and energy deposition inside random media. *Opt. Lett.* **2014**, *39*, 6324–6327. [CrossRef]
17. Song, Q.; Xiao, S.; Xu, Z.; Shalaev, V.; Kim, Y. Random laser spectroscopy for nanoscale perturbation. *Opt. Lett.* **2010**, *35*, 2624–2626. [CrossRef]
18. Leung, V.Y.F.; Lagendijk, A.; Tukker, T.W.; Mosk, A.P.; Ijzerman, W.J.; Vos, W.L. Interplay between multiple scattering, emission, and absorption of light in the phosphor of a white light-emitting diode. *Opt. Exp.* **2014**, *22*, 8190–8204. [CrossRef]
19. Wiersma, D. Disordered photonics. *Nat. Photonics* **2013**, *7*, 188–196. [CrossRef]
20. Curry, N.; Bondareff, P.; Leclercq, M.; Hulst, N.F.; Sapienza, R.; Gigan, S.; Gresillon, S. Direct determination of diffusion properties of random media from speckle contrast. *Opt. Lett.* **2011**, *36*, 3332–3334. [CrossRef]
21. Thompson, C.A.; Webb, K.J.; Weiner, A.M. Diffusive media characterization with laser speckle. *Appl. Opt.* **1997**, *36*, 3726–3734. [CrossRef]
22. Butakov, N.A.; Schuller, J.A. Designing multipolar resonances in dielectric metamaterials. *Sci. Rep.* **2016**, *6*, 38487. [CrossRef]
23. Yamilov, A.G.; Sarma, R.; Redding, B.; Payne, B.; Noh, H.; Cao, H. Position-dependent diffusion of light in disordered waveguides. *Phys. Rev. Lett.* **2014**, *112*, 023904. [CrossRef] [PubMed]
24. Sarma, R.; Golubev, T.; Yamilov, A.; Cao, H. Control of light diffusion in a disordered photonic waveguide. *Appl. Phys. Lett.* **2014**, *105*, 041104. [CrossRef]
25. Sarma, R.; Yamilov, A.; Neupane, P.; Cao, H. Using geometry to manipulate long-range correlation of light inside disordered media. *Phys. Rev. B* **2015**, *92*, 180203(R). [CrossRef]
26. Sarma, R.; Yamilov, A.; Petrenko, S.; Bromberg, Y.; Cao, H. Control of energy density inside a disordered medium by coupling to open or closed channels. *Phys. Rev. Lett.* **2016**, *117*, 086803. [CrossRef] [PubMed]
27. Sarma, R.; Yamilov, A.; Cao, H. Enhancing light transmission through a disordered waveguide with inhomogeneous scattering and loss. *Appl Phys. Lett.* **2017**, *110*, 021103. [CrossRef]
28. Liew, S.F.; Cao, H. Modification of light transmission channels by inhomogeneous absorption in random media. *Optics. Exp.* **2015**, *23*, 11043–11053. [CrossRef] [PubMed]
29. Chong, Y.D.; Stone, A.D. Hidden black: Coherent enhancement of absorption in strongly scattering media. *Phys. Rev. Lett.* **2011**, *107*, 163901. [CrossRef]
30. Lee, P.A.; Fisher, D.S. Anderson localization in two dimensions. *Phys. Rev. Lett.* **1981**, *47*, 882–885. [CrossRef]
31. Dorokhov, O.N. Localization and transmission coefficient for two coupled metal chains with disorder. *Solid State Commun.* **1982**, *44*, 915–919. [CrossRef]
32. Davy, M.; Shi, Z.; Park, J.; Tian, C.; Genack, A.Z. Universal structure of transmission eigenchannels inside opaque media. *Nat. Commun.* **2015**, *6*, 6893. [CrossRef]
33. Agarap, A.F. Deep learning using Rectified Linear Units (ReLU). *arXiv* **2018**, arXiv:1803.08375.
34. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
35. Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J. *Classification and Regression Trees*, 1st ed.; Taylor & Francis Group: New York, NY, USA, 2017.
36. Aloysius, N.; Geetha, M. A review on deep convolutional neural networks. In Proceedings of the 2017 International Conference on Communication and Signal Processing (ICCSP), Tamilnadu, India, 6–8 April 2017; pp. 588–592.
37. Sarma, R.; Xu, J.; de Ceglia, D.; Carletti, L.; Campione, S.; Klem, J.; Sinclair, M.B.; Belkin, M.A.; Brener, I. An all-dielectric polaritonic metasurface with a giant nonlinear optical response. *Nano Lett.* **2022**, *22*, 896–903. [CrossRef] [PubMed]