

Research article

Shiyin Du, Jie You, Jun Zhang, Zilong Tao, Hao Hao, Yuhua Tang, Xin Zheng and Tian Jiang*

Expedited circular dichroism prediction and engineering in two-dimensional diffractive chiral metamaterials leveraging a powerful model-agnostic data enhancement algorithm

<https://doi.org/10.1515/nanoph-2020-0570>

Received October 13, 2020; accepted December 4, 2020;
published online December 18, 2020

Abstract: A model-agnostic data enhancement (MADE) algorithm is proposed to comprehensively investigate the circular dichroism (CD) properties in the higher-order diffracted patterns of two-dimensional (2D) chiral metamaterials possessing different parameters. A remarkable feature of MADE algorithm is that it leverages substantially less data from a target problem and some training data from another already solved topic to generate a domain adaptation dataset, which is then used for model training at no expense of abundant computational resources. Specifically, nine differently shaped 2D chiral metamaterials with different unit period and one special sample containing multiple chiral parameters are both studied utilizing the MADE algorithm where three machine learning models (i.e., artificial neural network, random forest regression, support vector regression) are applied. The conventional rigorous coupled wave analysis approach is adopted to capture CD responses of these metamaterials and then assist the training of MADE, while the additional training data are obtained from our previous work. Significant evaluations regarding optical chirality in 2D

metamaterials possessing various shape, unit period, width, bridge length, and separation length are performed in a fast, accurate, and data-friendly manner. The MADE framework introduced in this work is extremely important for the large-scale, efficient design of 2D diffractive metamaterials and more advanced photonic devices.

Keywords: diffractive chiral metamaterials; machine learning methods; model-agnostic data enhancement algorithm; optical chirality; polarization-controllable devices.

1 Introduction

Optical chirality is viewed as one of the most promising and alluring aspects pertaining to chirality, a ubiquitous phenomenon in nature. Owing to its significant and striking characteristics, optical chirality has rapidly spread into the areas of integrated photonic communications [1], pharmaceuticals [2], chemistry [3], spectroscopy [4], sensitive detection [5], quantum computing [6], and nanoscience [7–9]. Understanding the chiroptical response not only assists one to explore the physical origins of such effect but also benefits its future perspectives. One crucial format of optical chirality, namely the circular dichroism (CD) [10], which describes the absorption variation when irradiated by the left circularly polarized (LCP) and right circularly polarized (RCP) light, should be acknowledged here. Particularly, the CD spectroscopy technique has been widely applied in performance characterization of the cutting-edge optical chirality-based nanodevices [11]. Two mirror images of chiral metamaterials that cannot be superimposed on each other, referred as enantiomers, behave differently to the LCP and RCP light, which is attributed to left or right handedness of samples. Importantly, optical chirality in artificial metallic structures can be engineered by changing the geometric parameters [12], demonstrating significant advantages over the natural materials owning weak chiroptical response [13].

Shiyin Du, Jie You and Jun Zhang contributed equally to this work.

*Corresponding author: Tian Jiang, Beijing Institute for Advanced Study, National University of Defense Technology, Beijing 100020, China; and College of Advanced Interdisciplinary Studies, National University of Defense Technology, Changsha 410073, China, E-mail: tjiang@nudt.edu.cn. <https://orcid.org/0000-0003-3343-5548>

Shiyin Du, Jun Zhang, Zilong Tao, Hao Hao and Yuhua Tang, State Key Laboratory of High Performance Computing, College of Computer, National University of Defense Technology, Changsha 410073, China. <https://orcid.org/0000-0002-4956-3379> (Y. Tang)

Jie You and Xin Zheng, National Innovation Institute of Defense Technology, Academy of Military Sciences PLA China, Beijing 100071, China

Recent years have witnessed growing academic interests in chiral metamaterials, considering their excellent potentials for applications in chiral sensing, filtering, switching, polarization manipulation, and biological detection [5, 14, 15]. Particularly, these metallic nanostructures can give rises to a plethora of remarkable physical phenomena when arranged into the periodic arrays, in which the adjacent modules will affect the overall near-field reaction. In contrast to the three-dimensional (3D) counterparts, two-dimensional (2D) chiral metamaterials have demonstrated distinguished and exceptional performances that can significantly promote the development of chirality-based photonic devices [16, 17], in terms of ultracompact footprint, limited losses, controllable optical dispersion, and excellent compatibility with CMOS fabrication [18–22]. One special but remarkable representative of such materials is 2D diffractive chiral metamaterial that possesses considerably large optical chirality in a higher-order diffraction beam than the zeroth order, which has emerged as a powerful platform for the exploration of tremendous unrevealed disciplines, including ultrafast detection, nonlinear chiroptical phenomena, polarization-selective communication devices, and chirality-relevant quantum optics [23]. While continuous efforts have been spent on the diverse diffractive metamaterials embracing the G-shaped [24], L-shaped [25], U-shaped [23] ones, etc., these investigation processes normally require complicated experimental setups or abundant computational resources. Therefore, a powerful and smart designing tool that can leverage previous research data is in great need, aiming at systematically exploring enormous 2D chiral metamaterials and other appealing photonic structures with the least resources, high accuracy, and ultrafast speed.

Deep learning networks or machine learning (ML) methods are significant for a wide range of scientific and industrial processes, covering the aspects of finance [26, 27], medicine [28–30], transportation [31–33], communication [34, 35], nanoscience [36–39], and sensors [40]. Particularly, these learning-based algorithms have emerged as one of the most successful research routes for computational physics and photonic device design [41–46]. In general, the training dataset created by either numerical simulations or experimental measurements is usually required in order to facilitate a predesigned artificial neural network (ANN) or other ML model to learn the underlying rules, followed by the trained model solving a target problem. Despite striking progress achieved in this area, the training process in most photonics-involved works relies heavily on optical response data from target devices [47–50], whose size is relatively large.

Unfortunately, it is not easy to acquire a large amount of data for many cases. Especially, for some topics needing to perform expensive experiments, this problem turns more prominent. Meanwhile, such abundant training data may not be recycled once the ML model is trained, leading to the serious waste of massive computing resources and time.

In this work, we introduce and employ a data enhancement algorithm associated with different ML methods to predict the optical chirality of various 2D diffractive chiral metamaterials, aiming at reducing the amount of training data with the assistance of a formerly addressed problem, whose key routine is recognized as the model-agnostic data enhancement (MADE) algorithm. Instead of consuming expensive computational resources to generate abundant training data for the target problem (target domain dataset), the MADE algorithm desires a relatively small target domain dataset computed by the rigorous coupled wave analysis (RCWA) method, as well as certain CD data from alternative already studied metamaterials (source domain dataset), to develop the domain adaptation dataset for model training. A vivid illustration of our proposed framework is shown in Figure 1. Specially, the source domain datasets are utilized to reduce the size of target domain datasets. Once the domain adaptation dataset and mapping dataset are available, the target issue turns to be the CD response study in the higher-order diffraction modes using deep learning network or other ML techniques. By means of the MADE algorithm, we first investigate the multiple shaped (e.g., T-like, U-like, and I-like) chiral metamaterials with diverse unit period and then look deeper into the impact of various geometric parameters on CDs for a specific chiral sample, during which the highly nonlinear dependence of CD response on the chiral parameters including shape, unit period, width, separation length, and bridge length is profoundly disclosed. Remarkable and significant superiorities of such algorithm over the traditional numerical methods or pure ML approaches are also revealed, in consideration of accuracy, computational speed, and resources, as well as the generalizable and flexible abilities.

2 Results and discussion

2.1 Principles of model-agnostic data enhancement

The model-agnostic data enhancement (MADE) algorithm makes use of a limited amount of data from a target issue together with certain samples from a previously studied

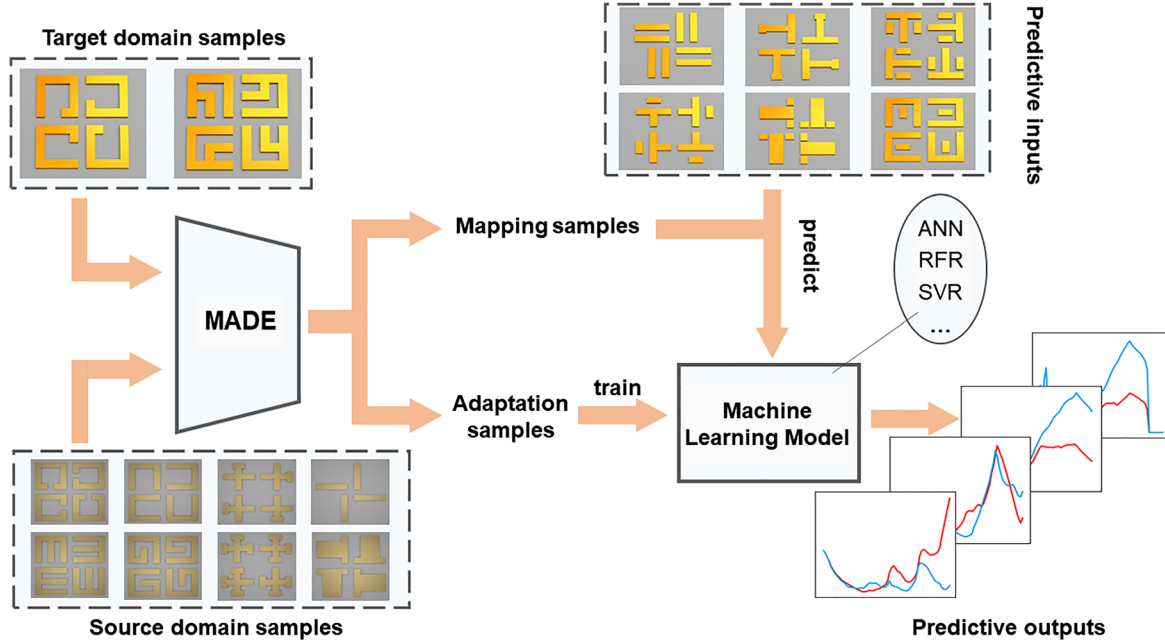


Figure 1: The framework of model-agnostic data enhancement (MADE) algorithm.

The MADE algorithm takes target domain samples and source domain samples as input to generate adaptation samples and mapping samples, which are thereby used for model training and testing, respectively.

topic (source domain) to improve prediction accuracy for the target problem, whose main procedure is illustrated in Algorithm 1. From the architecture of MADE in Figure 1, one

can see that two datasets, namely the source domain and target domain datasets, are needed at the input, which are extracted from a similar question that already exists and a

Algorithm 1 Model-Agnostic Data Enhancement

Input: $D_S^{n_S \times m_S}$ (The dataset of source domain, obtained from existing similar questions); $D_T^{n_T \times m_T}$ (The dataset of target domain, obtained from the problem to be solved);

Output: $D_A^{n_A \times m_A}$ (Adapted dataset, to be used for model training); D'_S (Mapped dataset, to be used for model prediction);

- 1: Call K-means algorithm to divide D_S into n_T clusters;
 - 2: if $m_S \neq m_T$ then
 - 3: Call PCA algorithm on cluster center and/or D_T with $\min\{m_T, m_S, n_T\}$ as feature extraction number (The transformed D_T is denoted as D'_T);
 - 4: end if
 - 5: Assign a cluster to each sample in D_T (one-to-one) by the lowest Euclidean distance between cluster center and $D_T(i)$ or $D'_T(i)$ ($i = 1, 2, \dots, n_T$);
 - 6: Calculate label $y_{D_S}(i)$ for sample $D_S(i)$ ($i = 1, 2, \dots, n_S$): $y_{D_S}(i) = \frac{\overline{D_T} + \alpha D_T(k)}{\alpha + 1}$, where α is a hyperparameter and $D_T(k)$ denotes the sample in D_T corresponding to the cluster where the sample $D_S(i)$ is located. Besides, $\overline{D_T}$ is the average of all samples of D_T ;
 - 7: Use (D_S, y_{D_S}) to train a fully connected neural network;
 - 8: Take D_S as input and exploit the trained neural network to calculate D'_S ;
 - 9: Select K samples from D'_S (denoted as φ_i) for $D_T(i)$ ($i = 1, 2, \dots, n_T$) by the lowest Euclidean distance, where K is a hyperparameter;
 - 10: Compute D_A based on K samples selected for $D_T(i)$ ($i = 1, 2, \dots, n_T$): $D_A(i \times K + j - K) = D_T(i) + \beta \varphi_i(j)$ ($j = 1, 2, \dots, K$), where β is a hyperparameter;
 - 11: return D_A, D'_S ;
-

new problem to be solved, respectively. While at the output of MADE, the domain adaptation dataset and mapping dataset are created, which are then used for model training and prediction. What follows is the application of ML models in predicting the behaviors of the target problem. It is worth noting that the MADE algorithm can be considered as the data preprocessing for neural networks or other ML models, which indicates that these two types of algorithms must be utilized (defined as MADE-ML) together to address a practical target problem. The main idea of MADE is to enhance the diversity of target domain data through source domain data, thereby expanding the training dataset. This is of vital significance since there has been worldwide recognition that ML methods (including deep learning networks) need to consume abundant data to learn features. In this case, a limited data size is very likely to cause failure in prediction [51, 52]. Here, the ML model is expressed as follows, with a function f representing the training model and the input of x :

$$f(x) = y + \text{loss} \quad (1)$$

where y is the label corresponding to x , and loss represents the error between the output and label. Normally, the model training section optimizes the parameters by gradient descent, enabling loss to approach zero. In this way, the final function f will be the mapping from the input x to the label y , denoted as $f: x \rightarrow y$. In our study, the source domain data is sufficient, whereas the size of target domain data is limited and not enough to support model training individually. Consequently, we propose a new function g to replace function f , which is given as follows:

$$g(x_S, \beta x_T) = y_S + \beta y_T + \text{loss} \quad (2)$$

where x_S and x_T represent the input of source domain data and target domain data, y_S and y_T are the label of the corresponding data, and β is a hyperparameter. After the

model training, loss is assumed to reach zero, leading to the simplification of Equation (2),

$$g(x_S, \beta x_T) = y_S + \beta y_T \quad (3)$$

In the prediction phase, x_S , y_S , and β are known, with $x_{\text{pred}} = x_T$. Then, the predicted value y_{pred} is equal to the following equation:

$$y_{\text{pred}} = y_T = \frac{g(x_S, \beta x_{\text{pred}}) - y_S}{\beta} \quad (4)$$

Obviously, the difference between these two methods is whether the source domain data is used. In our approach, different source domain samples are combined with certain target domain samples to acquire different training samples, thereby expanding the size of whole training dataset. If the number of samples in target domain and source domain is n_T and n_S , respectively, the total size of the training dataset can be up to $n_T \times n_S$ in our method. Although n_T is limited, the training dataset can still be augmented by the large n_S . In particular, each x_{pred} corresponds to multiple source domain samples in order to generate the input of function g , and thus, the final y_{pred} should be the average of multiple y_{pred} .

In other aspects, to avoid the potential data explosion caused by n_S , only partial source domain data are used to match the target domain data. Particularly, in the MADE model, we exploit the minimum Euclidean distance as a measurement standard to select K source domain samples for each target domain sample, leading to the total training dataset size of $K n_T$, where K is a hyperparameter (see step 9 in Algorithm 1). In fact, the Euclidean distance captures the similarity between two samples in our model. A new concept of superposition is introduced here. As shown in Figure 2, the sample y_S and sample y_T have similar distributions, so their superimposed results of $y_S + y_T$ are also similar to their distributions, indicating that the original

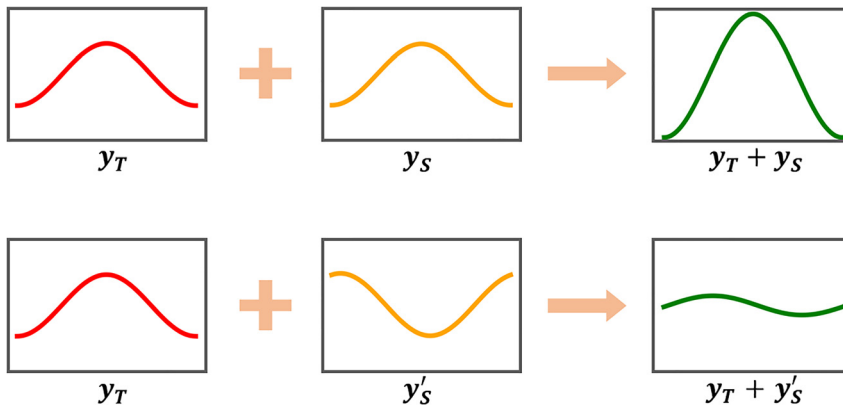


Figure 2: Superposition of two samples. The red and orange curves represent the distribution of samples in target domain and source domain, respectively, while the green lines correspond to the superposition of these two samples.

features of the data are retained to the greatest extent during the superimposition. Thus, when selecting K samples in the source domain that are similar to the target domain, the K results obtained by their superposition would follow similar distributions, which is conducive to model training. On the contrary, for the sample y'_S and sample y_T acquire different distributions, their superposition result $y_T + y'_S$ would be significantly different from y_T , suggesting that the characteristics of the original data are lost during the superposition. Therefore, it would be extremely beneficial to choose samples from source domain with high similarity to target domain.

Furthermore, both the calculation of Euclidean distance and the superposition between samples require the samples in different domains to have the same dimensions. It is thus necessary to perform domain adaptation on the source domain dataset to enhance the similarity between the two datasets (see steps 7–8 in Algorithm 1) and ensure their same dimensions (see steps 1–4 in Algorithm 1). In MADE, the domain adaptation is realized using a fully connected neural network, which can not only avoid the design of kernel functions but also exhibit the scalability [53], in contrast to ML methods. Additionally, if the source domain datasets are varied, the neural network does not need to retrain the domain adaptation model. The schematic illustration of the domain adaptive neural network is presented in Figure 3, which contains five key layers,

namely an input layer, an output layer, and three hidden layers. Notably, the number of neurons in the input layer depends on the sample size in source domain, while the neuron numbers in the output layer are dominant by the dimension of target domain datasets. Next, we denote the number of neurons in i th hidden layer as N_i ($i = 1, 2, 3$), with N_i to be a set of hyperparameters. We have to dynamically adjust N_i according to the samples in two domains. The activation layer for nonlinear transformation uses the ‘softplus’ function, which is defined as follows:

$$\sigma(x) = \ln(e^x + 1) \quad (5)$$

where x represents the output in hidden layers or the output layer. In order to incorporate the source domain with target domain, the domain adaptive neural network demands a large training dataset, with the mapping format varying among spectra–spectra, spectra–structure, structure–spectra, and structure–structure. Importantly, the domain adaptive neural network is a supervised learning model, in which each sample from source domain must have a label to guide training. Considering the big sample size gap between source domain and target domain, the one-to-one matching scheme is no longer suitable. To address this obstacle, we utilize the K-means algorithm to divide the source domain into n_T clusters, with the samples in each cluster corresponding to a target domain sample. Secondly, we obtain the one-to-one correspondence

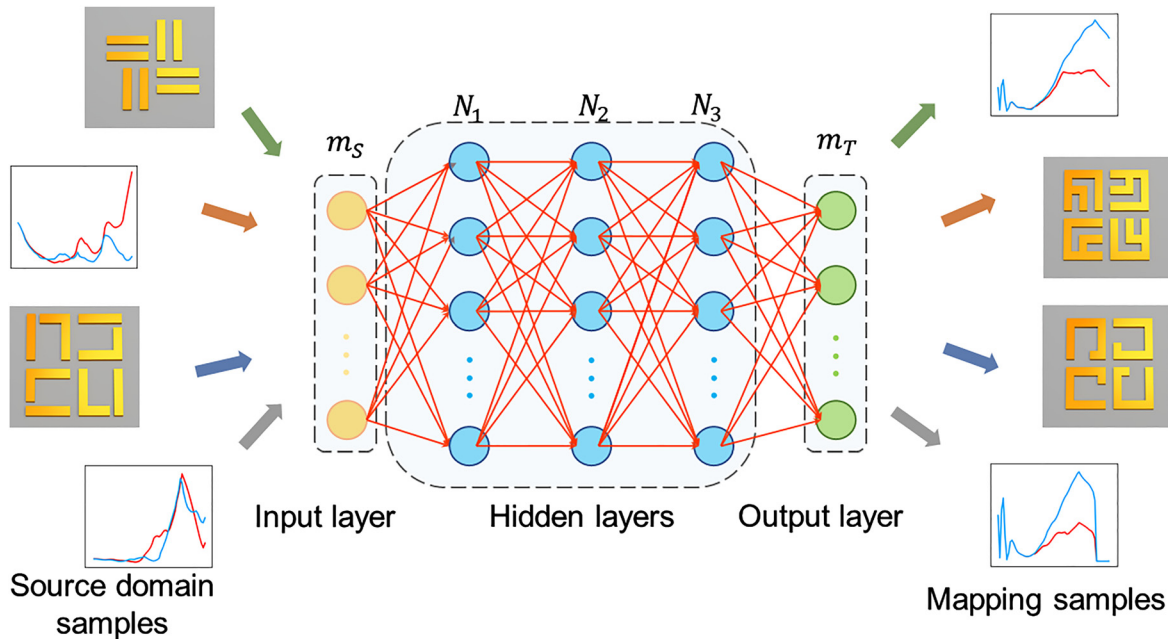


Figure 3: The architecture of the domain adaptive neural network.

The yellow, green, and blue circles stand for the neurons at the input layer, output layer, and hidden layers, respectively. The arrows on the left and right indicate data flow at the input and output, accordingly.

between clusters and target domain samples based on their lowest Euclidean distance (see step 5 of Algorithm 1). If the dimensions of source domain samples and target domain samples are inconsistent, the principal component analysis (PCA) algorithm is employed to reduce the relatively large dimension to make them match. The next step is to create the labels for source domain samples through the above correlation (step 6 of Algorithm 1). The simplest way is to directly use the target domain samples as labels. Nevertheless, this method would cause the neural network overfit, which is not able to learn the mapping from source domain to target domain well. To prevent such issues, a new label $y_{D_s}(i)$ for source domain sample $D_s(i)$ ($i = 1, 2, \dots, n_s$) is defined as follows:

$$y_{D_s}(i) = \frac{\overline{D_T} + \alpha D_T(k)}{\alpha + 1} \quad (6)$$

where α is a hyperparameter, and $D_T(k)$ denotes the sample in target domain corresponding to the cluster where the sample $D_s(i)$ is located. Additionally, $\overline{D_T}$ is the average of all samples (D_T). Compared to the simplest method mentioned above, the new method adds the global constraint $\overline{D_T}$ when calculating labels. This allows for dynamically adjusting the hyperparameter α to control the proportion of global constraints in the sample label, thereby avoiding the overfitting [54].

Lastly, the training of the neural network can be regarded as an optimization problem where the algorithm minimizes the discrepancy between the model output and the label for input data. Therefore, during the training process of the domain adaptive neural network, another loss function is applied to measure such discrepancy, which is mathematically expressed as follows:

$$\text{loss} = \frac{1}{m} \sum_{i=1}^m |y_{\text{pred}}(i) - y_{\text{real}}(i)| + \left| \overline{y_{\text{pred}}} - \overline{D_T} \right| \quad (7)$$

where y_{pred} and y_{real} represent the output of the neural network and the label for input data, respectively, and m is the size of batch data. Here, Adam optimization algorithm is used to train the domain adaptive neural network [55]. Note that the trained neural network is equivalent to a mapping from source domain to target domain, serving as the kernel function in ML algorithms. Hence, we take D_s as the input of the domain adaptive neural network to acquire the mapping result D'_s , which shares more similarities with D_T .

2.2 Model validation and performance comparison

The target theme in this context is the accurate and efficient prediction of the higher-order diffracted CD response in 2D

diffractive chiral metamaterials, whose vivid schematics are shown in Figure 4. Thus, two important categories of such metamaterials are investigated, namely multiple structures and multiple geometric parameters. Firstly, nine types of left-handed chiral metamaterials in Figure 4(a) are comprehensively studied, with one variable to be unit period. Particularly, these arrays are denoted as S1–S9, which possess a length of l and identical width and separation ($0.2l$), with the rest graphic parameters in a specific proportion to l , just like their shapes in Figure 4(a). This results in a unit period of $a = 2.4l$. These metamaterials share the same depth profiles: from top to bottom, they are the $h = 30$ nm gold arrays, a 10 nm Cr layer, a 200 nm SiO₂ film, and a Si substrate, respectively, as illustrated in Figure 4(b). On the other hand, we also focus on the influence of various geometric parameters on CDs in the higher-order diffraction beams. Hence, we select S1 metamaterials for such demonstration (see Figure 4(c)), whose unit cell consists of several key graphic parameters embracing a length of l , a bridge length of l_b , a width of w , a separation between two adjacent nanoparticles of s , and a gap length of $g = 0.5l - 1.5w$, determining the unit period of $a = 2l + 2w$. To show the common features of S1–S9 metamaterials, we employ the RCWA algorithm to compute the normalized intensities of $n = 1-4$ diffraction order beams under irradiation of both LCP and RCP light and illustrate one LCP result from S1 array ($a = 2.4 \mu\text{m}$, $l = 1 \mu\text{m}$, $w = s = 0.2 \mu\text{m}$, $l_b = 0.4 \mu\text{m}$) in Figure 4(d). It can be explicitly seen from this figure that light intensity in the third-order diffraction pattern is the weakest compared to other cases. Though not shown here, its CD response defined as $\text{CD} = (I_{\text{RCP}} - I_{\text{LCP}})/(I_{\text{LCP}} + I_{\text{RCP}})$ is however the largest, where I_{LCP} and I_{RCP} are the third-order diffracted beam intensities induced by LCP and RCP light, respectively. Similar phenomena are also observed in S2–S9 nanostructures.

To assist the training of the MADE model, we first perform RCWA numerical simulations for the above two categories of metamaterials. Precisely, the target domain dataset (D_{T1}) for the first category contains 1899 pairs of the third-order diffraction I_{RCP} and I_{LCP} spectra for S1–S9 arrays in different gold lengths ($0.8-5 \mu\text{m}$), covering a wavelength range of $0.4-1.03 \mu\text{m}$. While for the second category, its target domain dataset (D_{T2}) is extracted from 2785 S1 samples, which consist of diverse gold length ($0.8-2 \mu\text{m}$), width ($0.1l-0.3l$), space ($0.2l-l$), and bridge length ($0.4l-l$). In addition, the source domain datasets are also needed for the MADE training process. In this work, we use the data from two published works [41, 42] that employ deep neural networks in CD response analysis for the source domain. Particularly, the data in work 1 contain chiroptical responses from nine different structures [41], while work 2

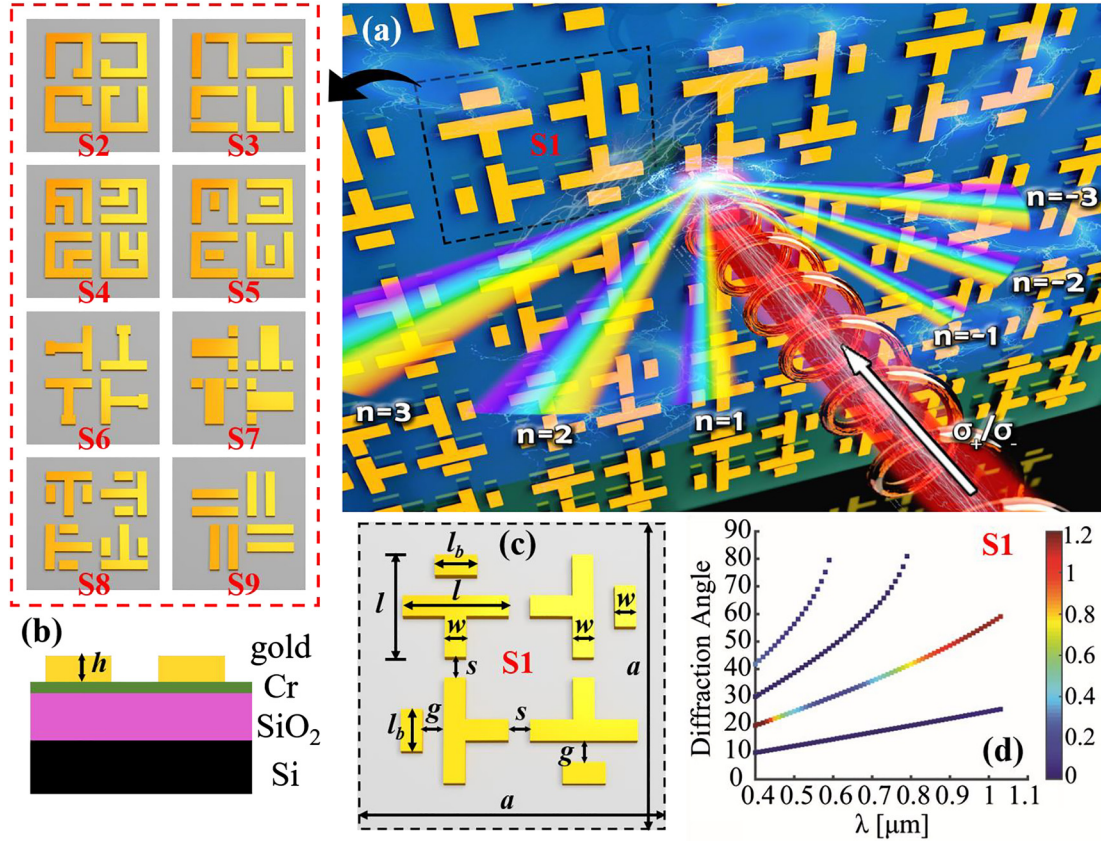


Figure 4: Optical description of various two-dimensional (2D) diffractive chiral metamaterials.

(a) The right panel shows the higher-order diffraction patterns in S1 metamaterial when irradiated by the circularly polarized light. Notably, shapes of the metallic array can be changed to other eight structures shown on the left, referred as S2–S9. (b) The depth profile for S1–S9 chiral metamaterials: the thickness of gold, Cr, and SiO₂ films is $h = 30, 10, 200$ nm, respectively, with Si layer to be the substrate. (c) The geometric illustration for S1 metamaterials, which contains a length of l , a width of w , a separation between two adjacent modules of s , a gap length of g , and a bridge length of l_b . (d) The rigorous coupled wave analysis (RCWA)-calculated normalized intensities for the $n = 1-4$ diffraction order beams in S1 metamaterial, under left circularly polarized (LCP) light excitation.

provides the different optical responses from a T-like array with four key graphic parameters [42].

Once the datasets of target domain and source domain are available, we start to train the MADE model incorporated with ANN, or simply MADE-ANN, aiming at predicting the third-order diffracted chiroptical responses. Specifically, regarding the first category, we randomly select 1200 samples from the target domain dataset D_{T1} as the input for MADE, with the remaining 699 samples for testing, while we acquire the nine-structure data (from work 1) as the source domain dataset. Hence, the comparisons of the normalized third-order LCP/RCP intensities calculated by MADE-ANN and RCWA are shown in Figure 5(a–i), in which the unit period of S1–S9 is picked randomly to guarantee the high precision of MADE-ANN. In case of the second category, we divide the target domain dataset D_{T2} into the training and verification parts at random, whose sample sizes are 2000 and 785, respectively, along with T-like structure data (from

work 2) as the source domain dataset. The resulting chiroptical responses predicted by MADE-ANN are shown in Figure 5(j–l), where the RCWA calculated curves are also provided. The most eye-catching conclusion from Figure 5 is that the MADE-ANN-predicted optical chirality (i.e., pred curves) matches extremely well with the RCWA simulations (i.e., label spectra) in all cases, indicating that the MADE algorithm is feasible to explore the relationship between chiral parameters and optical chirality.

In order to verify the model-agnostic feature of MADE, two more ML models in addition to the ANN, namely the random forest regression (RFR) [56] and support vector regression (SVR) [57], are implemented in the MADE framework, accompanied by two target domain datasets (D_{T1} and D_{T2}) and two source datasets (nine-structure dataset [41] and T-like structure dataset [42]). For each target domain dataset, we use two above source domain datasets to train these three ML models. Notably, before proceeding

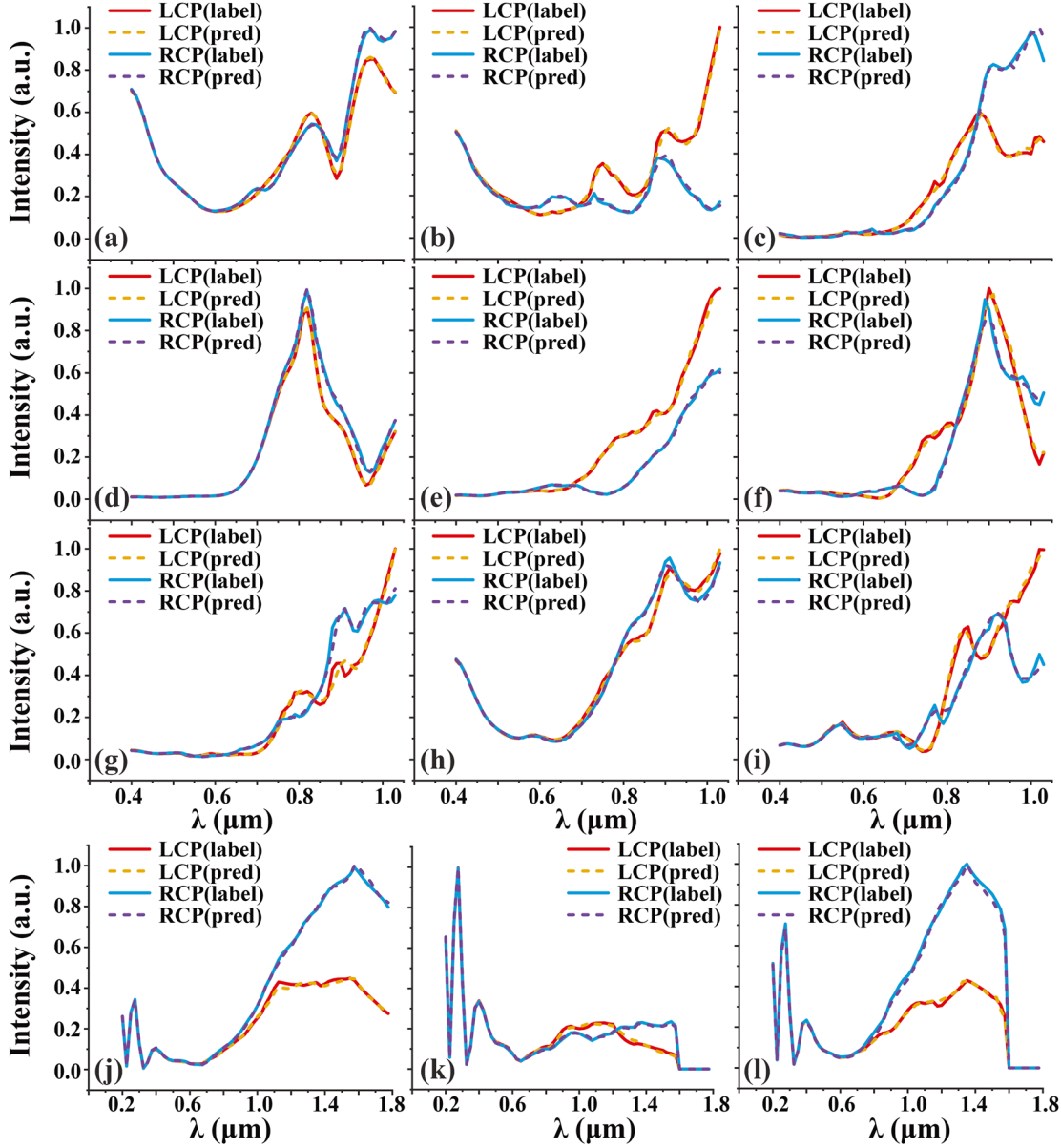


Figure 5: Comparison of the third-order diffracted chiroptical response calculated by rigorous coupled wave analysis (RCWA) (label) and a model-agnostic data enhancement (MADE) plus artificial neural network (ANN) model (pred).

(a)–(i) The $I_{\text{RCP}}/I_{\text{LCP}}$ results for the first category metamaterials, with the source domain dataset from nine similar structure [41] in work 1. Specially, the unit periods of S1–S9 arrays are randomly selected to ensure the high accuracy, which are (a) 2.94 μm , (b) 1.82 μm , (c) 1.32 μm , (d) 4.1 μm , (e) 2.6 μm , (f) 2.6 μm , (g) 2.2 μm , (h) 2.8 μm , (i) 2.28 μm , respectively. (j)–(l) The $I_{\text{RCP}}/I_{\text{LCP}}$ spectra for the second category metamaterials (S1) comprising different graphic parameters: (j) $l = 1.4 \mu\text{m}$, $w = 0.25l$, $s = l$, $l_b = 0.7l$; (k) $l = 2 \mu\text{m}$, $w = 0.15l$, $s = 0.2l$, $l_b = 0.9l$; (l) $l = 1.2 \mu\text{m}$, $w = 0.3l$, $s = l$, $l_b = l$. Here, the data of T-like nanostructures [42] in work 2 are utilized as the source domain dataset.

to the model training, we need to normalize the RCWA data to be in the range of [0,1], considering their relatively small absolute values, which is defined as follows:

$$y' = -\frac{\lg(y + 10^{-10})}{10} \quad (8)$$

where y represents original label and y' is normalized label. One important factor is mean absolute error (MAE), which

characterizes the difference between the MADE model output and the normalized RCWA label. Moreover, a baseline method is also defined, in which the model training only utilizes target domain dataset, without data from source domain. Thus, the quantitative results for the two categories of metamaterials are shown in Tables 1 and 2, respectively. From these two tables, we can find that MADE algorithm has reduced the MAEs of ANN, RFR, and

Table 1: Mean absolute error (MAE) of three machine learning (ML) models for the first category metamaterials.

Method and setup	420 samples ($\times 10^{-3}$)			810 samples ($\times 10^{-3}$)			1200 samples ($\times 10^{-3}$)		
	Baseline	MADE(9)	MADE(T)	Baseline	MADE(9)	MADE(T)	Baseline	MADE(9)	MADE(T)
ANN	9.014	3.03	3.866	5.597	1.792	2.46	4.429	1.399	1.727
RFR	6.341	3.025	4.04	4.842	2.125	3.417	4.082	1.694	2.695
SVR	30.972	15.035	17.87	29.031	14.545	17.068	28.249	14.392	16.735

“9” and “T” indicate nine-structure dataset and T-like structure dataset, used in the source domain, respectively.

Table 2: Mean absolute error (MAE) of artificial neural network (ANN), random forest regression (RFR), and support vector regression (SVR) regarding the second category metamaterials.

Method and setup	6000 samples ($\times 10^{-3}$)			1200 samples ($\times 10^{-3}$)			2000 samples ($\times 10^{-3}$)		
	Baseline	MADE(9)	MADE(T)	Baseline	MADE(9)	MADE(T)	Baseline	MADE(9)	MADE(T)
ANN	8.292	7.42	6.484	4.726	2.787	2.367	2.81	1.649	1.356
HFH	26.812	14.503	14.347	18.825	10.794	10.165	14.327	7.919	7.023
SVR	56.854	26.013	29.549	54.357	24.903	28.899	52.805	24.042	27.936

“9” and “T” indicate nine-structure dataset and T-like structure dataset, used in the source domain, respectively.

SVR in various degrees in both cases, provided that the source domain datasets are employed.

Moreover, it is easily discerned that the dataset types in the source domain would also affect the MAE performances. Precisely, the nine-structure data are more suitable to act as the source domain dataset for the first category metamaterials than the T-like structure data, which is the opposite case in the second category metamaterials. This indicates that the similarity of dataset between the source domain and target domain would largely contribute to high precision of MADE models, which in turn confirms the necessity of the domain adaptive neural network in the MADE framework. Also, it is undeniable that the ANN acquires the best performances regarding high predicting precision, while SVR suffers from the worst MAEs.

Additional insights into the features of the MADE algorithm are provided by the dependence of MAE on the amount of training data, with the corresponding curves being exhibited in Figure 6. Notably, the pure ML algorithms and MADE-ML methods are both considered for full comparison, with the latter algorithm possessing two source domain datasets. It can be intuitively seen from this figure that the improvement of predicting precision becomes more obvious with a larger amount of training data and then it turns steady after reaching a certain extent. This phenomenon means that the significant increase in the dimension of training data is required to boost the performance of the model. This trend highlights the advantages of the MADE algorithm, which can still enhance the effect of ML models even if the method of increasing the data

fails. From Y axis of all plots in Figure 6, one can find that the MADE models can significantly reduce the MAEs compared to the ANN, with the same size of training data. Meanwhile, if demanding the same level of MAE conditions, we can clearly discover that the amount of data for MADE algorithms is much smaller than the case without MADE. Interestingly, this trend becomes more prominent at a smaller MAE. Take Figure 6(a) for instance, to achieve the MAE of 6.678×10^{-3} , the pure ANN model needs 600 pairs of samples for training, while for the MADE-ANN algorithm with the nine-structure dataset (source domain), only 120 pairs of data are desired. When the MAE becomes even smaller (4.4×10^{-3}), the pure ANN uses 1200 pairs of training samples, whereas the MADE-ANN based on the nine-structure source dataset only adopts 270 pairs of samples. Similar trends are discovered in other ML algorithms and other MADE-ML models for two categories of metamaterials. These results undoubtedly prove that a MADE-ML algorithm is superior to a non-MADE algorithm in the prediction of the third-order diffracted chiroptical response.

Furthermore, the computational time is an alternative key factor that determines the performance of the MADE algorithm. Here, RCWA and three MADE-ML methods are employed for such evaluation, with the ML algorithm to be ANN, RFR, or SVR. To ensure a fair and efficient comparison, the number of training samples in target domain is fixed at 1000 for all three MADE-ML models. In addition, the nine-structure source domain dataset is employed in the first category metamaterials, while the T-like structure source domain dataset is applied in the second category.

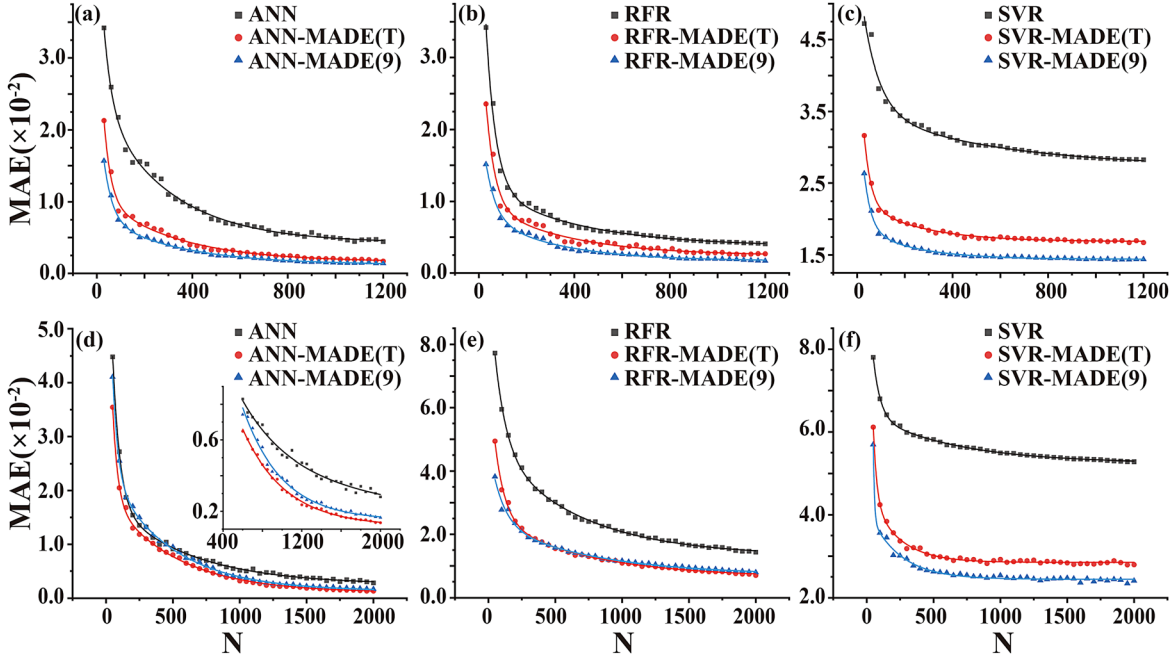


Figure 6: Comparison of MAE among different ML algorithms and the MADE models coupled with these algorithms, in cases of various numbers of training samples.

Notably, three pure ML methods encompassing ANN, RFR, and SVR are employed without using MADE framework, whereas in MADE relevant models the symbols of “9” and “T” indicate that the nine-structure data and T-like structure dataset are adopted as source domain datasets, respectively. The panels of (a)-(c) correspond to the first category metamaterials, while panels of (d)-(f) stand for the cases of the second category.

The results under the above circumstances are shown in Table 3. It is noticeable that each indicator in Table 3 represents the average central processing unit time consumed by an algorithm to predict 100 samples. One important finding is that the MADE-ML models take more computational time than the baseline methods, regardless of predicting accuracy. This is reasonable since the consumed time of MADE mainly comes from two aspects: on the one hand, the MADE algorithm needs to select K mapping dataset according to the minimum Euclidean distance, which means that computational time would be very large if the source domain contains too many samples; on the other hand, the overall predicting precision highly relies on

the average of multiple prediction results, which means multiple CD evaluations by MADE-ML are desired in this process. Although the prediction time of MADE-ML algorithms turns larger than that of the pure ML algorithm, this increment is negligible compared to the RCWA method. Equally important, the intuitive comparison of calculation accuracy between MADE-ML and RCWA is also presented in Table 3. Here, the mean absolute percentage error (MAPE) is the main evaluation factor, which is mathematically expressed as follows:

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i^{\text{pred}} - y_i^{\text{real}}}{y_i^{\text{real}}} \right| \times 100\% \quad (9)$$

Table 3: Average consumed central processing unit (CPU) time and mean absolute percentage error (MAPE) of different methods.

Method and setup	Average CPU time(ms)				Mean absolute percentage error (%)			
	First category		Second category		First category		Second category	
	Baseline	MADE(9)	Baseline	MADE(T)	Baseline	MADE(9)	Baseline	MADE(T)
RCWA	1.98×10^5	/	1.98×10^5	/	0	/	0	/
ANN	5.52×10^{-3}	1.06	6.79×10^{-2}	1.12	0.676	0.292	0.568	0.389
RFR	2.09	21.04	3.58	43.14	0.73	0.355	2.726	1.88
SVR	0.46	11.22	0.80	52.24	4.806	3.132	9.644	6.785

“9” and “T” indicate nine-structure dataset and T-like structure dataset, used in the source domain, respectively.

It represents the proportion of prediction error in the absolute value. Notably, all the training data for three ML models are furnished by RCWA, indicating that the MAPE of RCWA is zero. It is clearly observed that the ANN acquires the smallest MAPE compared to RFR and SVR, whose computational accuracy is larger than 99%. While the MAPE of SVR is the largest, the prediction precision of which is still over 90%. Although three ML methods lose a certain degree of accuracy, the MADE algorithm improves the prediction accuracy to some extent while keeping substantially small amount of training data.

2.3 Evaluating the third-order diffracted circular dichroism

Utilizing the MADE algorithm, we can explore the intricate and nonintuitive dependence of CD characteristics in the higher-order diffraction beams on geometrical parameters in an accurate and fast manner. In this context, the shape and unit period of 2D chiral metamaterials are two chiral parameters that most significantly affect the optical CD responses. Accounting for the analogy between the investigated diffractive metamaterials and a simple grating where the grating equation of $a \sin\theta = n \cdot \lambda$ is applied, the unit period is assumed to not only cause shifts in the resonance wavelengths but also influence distributions of higher-order diffraction beams in real space.

Notably, the unit period is selected to be proportional to the gold length of S1–S9 metamaterials, namely $a = 2.4l$. Thus, for simplicity, we directly investigate the third-order diffraction CD properties of S1–S9 nanostructures in case of various gold length l employing the MADE algorithm, with the results being presented in Figure 7. One significant finding from this figure is that the CD responses exhibit highly nonlinear dependences on both the wavelength λ and gold length l , in cases of nine differently shaped metamaterials. Moreover, the complicated bisignate characteristics of CD responses are observed at first glance for most l of S1–S9 metamaterials. One salient example is S4 array in which the triple or more sophisticated bisignate features of CD are discovered. Alternatively, the blue CD modes play a dominant role across the contour map of S5 nanostructures, while the red modes turn much stronger in S7 array. As for the other seven metamaterials, a quasi-balance seems to be reached between these two modes. On the other hand, the spatial locations of higher-order diffraction beams can be altered via the change of unit period, which enables the angle-resolved optical chirality detection with a larger flexibility. Equally important, it is apparent that the CD maps of S1–S9 metamaterials behave

quite distinctively, revealing that these metamaterials contain diverse electromagnetic modes and then different electric field distribution at near-field.

It is of vital importance at this stage to look deeper into the complex relationship between different chiral parameters and optical chirality. Thus, we study the third-order diffracted chiroptical response of S1 metamaterials possessing various gold width, separation length, bridge length, and gold length, by means of the MADE method. The corresponding results are illustrated in Figure 8. It is noticeable that to ensure the effectiveness of this investigation, only one factor is changed at a time with the other ones remaining the same. The most obvious finding to emerge from this figure is that the width, space length, bridge length, and gold length contribute diversely to optical chirality in the third-order diffracted beams. In general, the unit period follows a relation of $a = 2l + 2s$. Thus, the variations of space length and gold length would result in different unit periods. Additionally, diverse width and bridge length would cause changes in the patterns of metamaterials, generating a distinctive near-field distribution that determines the far-field chiroptical reaction [23, 58]. Here, the point sources are no longer suitable to describe our S1 metamaterials since their dimensions are not that small compared to the irradiation wavelength. This suggests that the CD properties of higher-order diffraction beams are highly dependent on the patterns and far-field reaction of a chiral module.

Moving on now to the detailed evaluation of each chiral parameter, from Figure 8(a), one can easily discern that the light intensity in the third-order diffraction beam increases with gold width until $w = 0.2l$, under irradiation by either LCP or RCP light. However, the LCP intensity exhibits a clear decay when the width turns larger than $0.2l$, unlike the case of RCP light. In the contrary, the strongest CD response is observed at $w = 0.3l$ of Figure 8(d), suggesting that to acquire a maximum CD, the relatively large LCP/RCP intensity is not a requirement. In terms of space, it is found from Figure 8(b, f) that the peaks of light intensity and chiroptical response both exhibit redshifts with space length, which can be explained by the enlarged unit period. Nevertheless, a larger space length would cause a weak coupling effect among adjacent nanoparticles, eventually affecting the CD performance. When it comes to the bridge length, a surprising observation from Figure 8(c) is that most the LCP and RCP resonant modes locate in the range of 1–1.35 μm , while the big optical CDs distribute across a much wider wavelength regime (see Figure 8(g)). Furthermore, the characterization of unit period's influence on optical chirality is performed in S1 metamaterial with graphic parameters of $w = 0.3l$, $s = 0.2l$,

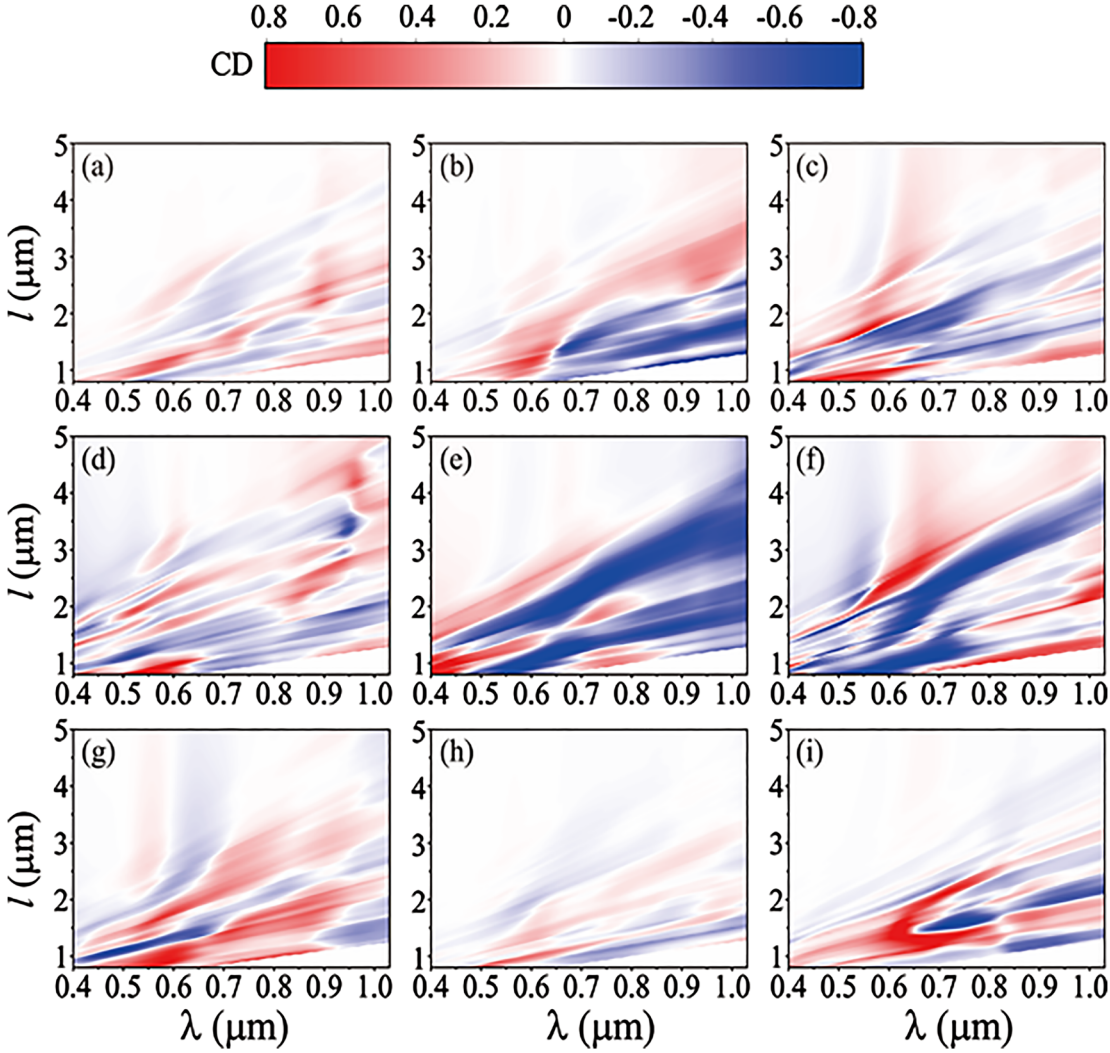


Figure 7: Contour maps of the third-order diffracted circular dichroism (CD) response in S1–S9 chiral metamaterials, accounting for different wavelength and gold length. All the presented results are computed using the model-agnostic data enhancement (MADE) method.

and $l_b = 0.6l$. The main results under the above circumstances are summarized in Figure 8(d, h). It is immediately discovered that the wavelength of LCP/RCP resonance becomes longer at a larger unit period, so does the wavelength of CD maximum. Anyhow, this dependence seems not to obey a fixed rule. Additionally, the strength of CD response in Figure 8(h) exhibits a highly nonlinear dependence on the unit period.

In conclusion, we have established a MADE framework associated with different ML networks to efficiently predict and evaluate the CD responses of various 2D chiral metamaterials, accounting for up to the fourth diffraction order beams. Precisely, the MADE algorithm requires only a small amount of CD spectra from a target metamaterial for the CD performance characterization, provided that additional data from previously explored nanostructures can be

utilized to assist the training process, which can address the data limitation of pure ML algorithms. Particularly, the MADE algorithm does not necessarily demand the high similarity of datasets between the source and target domains, and different ML techniques (e.g. ANN, RFR, SVR) can be perfectly applied in this framework, suggesting its excellent generalization ability. The traditional RCWA approach is utilized to furnish a limited amount of training data for MADE, along with some training data from a solved problem. Thus, the complicated and nonintuitive relationships between the higher-order diffracted optical chirality and the shape, unit period, width, space length, and bridge length of 2D diffractive chiral metamaterials are explicitly explored using the MADE framework together with various ML algorithms. This approach is confirmed to own remarkably fast computational speed and high

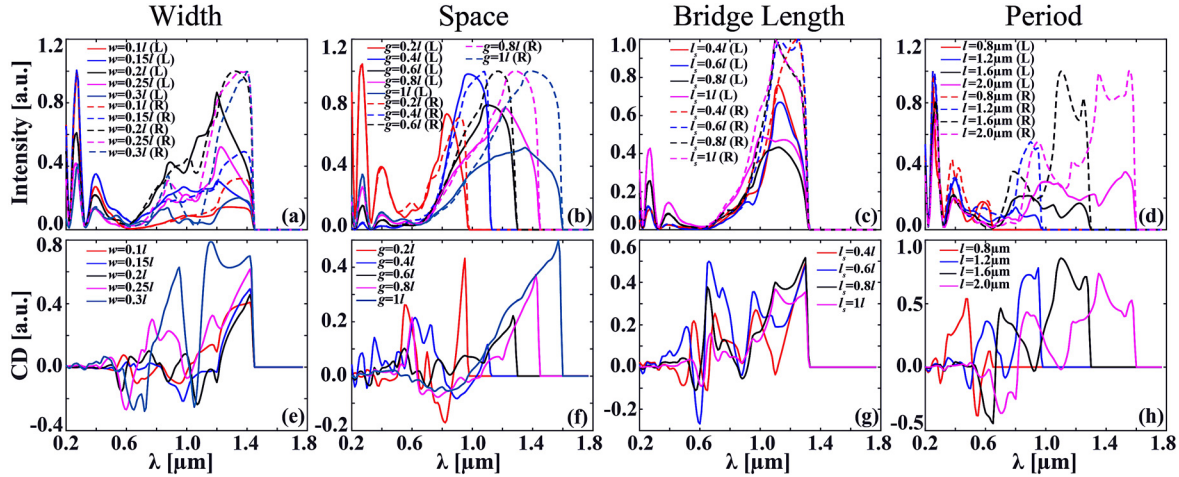


Figure 8: The model-agnostic data enhancement (MADE) model predicted chiroptical response spectra in S1 chiral metamaterials with different chiral parameters.

Specifically, (a)–(d) show the intensities in the third-order diffraction beams under the left/right circularly polarized (LCP/RCP) light excitation, while (e)–(h) correspond to their circular dichroism (CD) performances, considering different (a) (e) width, (b) (f) space length, (c) (g) bridge length, and (d) (h) gold length. Notably, if one geometric parameter is varied, the other ones are fixed at some values: (a)(e) $s = 0.2l$, $l_b = 0.4l$, $l = 1.8 \mu\text{m}$; (b)(f) $w = 0.2l$, $l_b = 0.7l$, $l = 1.2 \mu\text{m}$; (c)(g) $w = 0.25l$, $s = 0.4l$, $l = 1.4 \mu\text{m}$; (d)(h) $w = 0.3l$, $s = 0.2l$, $l_b = 0.6l$.

accuracy comparable to RCWA toward optical chirality manipulation in diffractive chiral metamaterials, transforming this rule-based problem into a small data-driven study. It is significant to mention that the MADE algorithm can be easily extended to the design of other functional metamaterials. For instance, our MADE approach can readily employ a limited amount of training data from 3D metamaterials, together with some results from an already solved nanostructure, to achieve the accelerated and efficient characterization of 3D samples. The generalization and flexibility of the MADE algorithm contribute in several ways to our understanding of transfer learning in the efficient prediction of chiroptical response inside diffracted chiral metamaterials and provide a basis for the large-scale, optimal design of complex photonic devices.

3 Methods

RCWA is viewed as an effective and efficient method to explore the electromagnetic properties of periodic nanostructures. By leveraging the RCWA method implemented in Synopsys RSoft DiffractMOD, we numerically calculate the intensities of higher-order diffraction beams for various 2D chiral metamaterials under circularly polarized light irradiation, in order to generate the target domain datasets for MADE training. Moreover, the most crucial MADE algorithm is implemented in Python. In particular, the K-means algorithm and PCA algorithm integrated in the MADE framework come from the ML library, namely scikit-learn [59]. The domain adaptive neural network is built by TensorFlow2.2-gpu, which is an open-source system for large-scale ML

[60]. Based on the target domain datasets provided by RCWA simulations and source domain datasets from some previously studied problems, we can generate domain adaptation datasets and mapping datasets. At this point, the target problem converts into a problem where the optical chirality of higher-order diffraction patterns is explored, by means of either a deep learning network or other ML techniques.

Author contributions: Shiyin Du, Jie You, and Jun Zhang contributed equally to this work.

Research funding: The authors are grateful for financial support from the National Natural Science Foundation of China (11804387, 11802339, 11805276, 11902358, 61805282, 61801498, 62075240); the Scientific Researches Foundation of National University of Defense Technology (ZK18-03-22, ZK18-01-03, ZK18-03-36); and the Science Fund for Distinguished Young Scholars of Hunan Province (2020JJ2036).

Conflict of interest statement: The authors declare no conflicts of interest regarding this article.

References

- [1] A. M. Vegni and V. Loscri, *Nano Commun. Netw.*, vol. 9, p. 28, 2016.
- [2] C. K. Savile, J. M. Janey, E. C. Mundorff, et al., *Science*, vol. 329, p. 305, 2010.
- [3] N. J. Greenfield, *Nat. Protoc.*, vol. 1, p. 2876, 2006.
- [4] K. Yao and Y. Zheng, *J. Phys. Chem. C*, vol. 123, p. 11814, 2019.

- [5] Y. Y. Lee, R. M. Kim, S. W. Im, M. Balamurugan, and K. T. Nam, *Nanoscale*, vol. 12, p. 58, 2020.
- [6] C. Wagenknecht, C. M. Li, A. Reingruber, et al., *Nat. Photonics*, vol. 4, p. 549, 2010.
- [7] G. Lozano, T. Barten, G. Grzela, and J. G. Rivas, *New J. Phys.*, vol. 16, 2014, 013040.
- [8] G. Lozano, D. J. Louwers, S. R. Rodríguez, et al., *Light Sci. Appl.*, vol. 2, p. e66, 2013.
- [9] Y. Xu, Z. Shi, X. Shi, K. Zhang, and H. Zhang, *Nanoscale*, vol. 11, p. 14491, 2019.
- [10] R. Quidant and M. Kreuzer, *Nat. Nanotechnol.*, vol. 5, p. 762, 2010.
- [11] L. Torsi, G. M. Farinola, F. Marinelli, et al., *Nat. Mater.*, vol. 7, p. 412, 2008.
- [12] V. K. Valev, J. J. Baumberg, C. Sibilia, and T. Verbiest, *Adv. Mater.*, vol. 25, p. 2517, 2013.
- [13] S. Zu, T. Han, M. Jiang, F. Lin, X. Zhu, and Z. Fang, *ACS Nano*, vol. 12, p. 3908, 2018.
- [14] Y. Luo, C. Chi, M. Jiang, et al., *Adv. Opt. Mater.*, vol. 5, p. 1700040, 2017.
- [15] I. De Leon, M. J. Horton, S. A. Schulz, J. Upham, P. Banzer, and R. W. Boyd, *Sci. Rep.*, vol. 5, p. 13034, 2015.
- [16] Z. Shi, X. Ren, H. Qiao, et al., *J. Photochem. Photobiol. C Photochem. Rev.*, p. 100354, 2020, <https://doi.org/10.1016/j.jphotochemrev.2020.100354>.
- [17] Z. Shi, R. Cao, K. Khan, et al., *Nano-Micro Lett.*, vol. 12, p. 1, 2020.
- [18] Y. Hu, J. You, M. Tong, et al., *Adv. Sci.*, n.d., p. 2000799, <https://doi.org/10.1002/advs.202000799>.
- [19] Y. Hu, T. Jiang, J. Zhou, et al., *Nano Energy*, vol. 68, p. 104280, 2020.
- [20] G. Li, S. Zhang, and T. Zentgraf, *Nat. Rev. Mater.*, vol. 2, p. 1, 2017.
- [21] C. Lan, Z. Shi, R. Cao, C. Li, and H. Zhang, *Nanoscale*, 2020, <https://doi.org/10.1039/D0NR02574G>.
- [22] H. Hu, Z. Shi, K. Khan, et al., *J. Mater. Chem. A*, vol. 8, p. 5421, 2020.
- [23] C. Kuppe, C. Williams, J. You, et al., *Adv. Opt. Mater.*, vol. 6, p. 1800098, 2018.
- [24] K. E. Chong, I. Staude, A. James, et al., *Nano Lett.*, vol. 15, p. 5369, 2015.
- [25] C. Kuppe, X. Zheng, C. Williams, et al., *Nanoscale Horizons*, vol. 4, p. 1056, 2019.
- [26] J. B. Heaton, N. G. Polson, and J. H. Witte, *Appl. Stoch Model Bus. Ind.*, vol. 33, p. 3, 2017.
- [27] R. Culkun and S. R. Das, *J. Invest. Manag.*, vol. 15, p. 92, 2017.
- [28] A. Rajkomar, J. Dean, and I. Kohane, *N. Engl. J. Med.*, vol. 380, p. 1347, 2019.
- [29] F. Wang, L. P. Casalino, and D. Khullar, *JAMA Intern. Med.*, vol. 179, p. 293, 2019.
- [30] T. Ching, D. S. Himmelstein, B. K. Beaulieu Jones, et al., *J. R. Soc. Interface*, vol. 15, p. 20170387, 2018.
- [31] F. Zantalis, G. Koulouras, S. Karabetsos, and D. Kandris, *Future Internet*, vol. 11, p. 94, 2019.
- [32] H. Nguyen, L. M. Kieu, T. Wen, and C. Cai, *IET Intell. Transp. Syst.*, vol. 12, p. 998, 2018.
- [33] S. H. Fang, Y. X. Fei, Z. Xu, and Y. Tsao, *IEEE Sensor. J.*, vol. 17, p. 6111, 2017.
- [34] O. Simeone, *IEEE Trans. Cogn. Commun. Netw.*, vol. 4, p. 648, 2018.
- [35] D. Zibar, M. Piels, R. Jones, and C. G. Schäffer, *J. Lightwave Technol.*, vol. 34, p. 1442, 2015.
- [36] Y. Kiarashinejad, S. Abdollahramezani, and A. Adibi, *Npj Comput. Mater.*, vol. 6, p. 1, 2020.
- [37] X. Lin, Z. Si, W. Fu, et al., *Nano Res.*, vol. 11, p. 6316, 2018.
- [38] S. Molesky, Z. Lin, A. Y. Piggott, W. Jin, J. Vucković, and A. W. Rodriguez, *Nat. Photonics*, vol. 12, p. 659, 2018.
- [39] J. Yang, F. Luo, T. S. Kao, et al., *Light Sci. Appl.*, vol. 3, p. e185, 2014.
- [40] Y. Jin, B. Wen, Z. Gu, et al., *Adv. Mater. Technol.*, vol. 5, p. 2000262, 2020.
- [41] Z. Tao, J. You, J. Zhang, X. Zheng, H. Liu, and T. Jiang, *Optic Lett.*, vol. 45, p. 1403, 2020.
- [42] Z. Tao, J. Zhang, J. You, et al., *Nanophotonics*, vol. 9, p. 2945, 2020.
- [43] T. Ueno, H. Hino, A. Hashimoto, Y. Takeichi, M. Sawada, and K. Ono, *Npj Comput. Mater.*, vol. 4, p. 1, 2018.
- [44] W. Ma, F. Cheng, and Y. Liu, *ACS Nano*, vol. 12, p. 6326, 2018.
- [45] Y. Kiarashinejad, M. Zandehshahvar, S. Abdollahramezani, O. Hemmatyar, R. Pourabolghasem, and A. Adibi, *Adv. Intell. Syst.*, vol. 2, p. 1900132, 2020.
- [46] L. Chen, Y. Yin, Y. Li, and M. Hong, *Opto-Electronic Adv.*, vol. 2, p. 190019, 2019.
- [47] C. Sathyaseelan, V. Vijayakumar, and T. Rathinavelan, *J. Mol. Biol.*, 2020, <https://doi.org/10.1016/j.jmb.2020.08.014>.
- [48] Y. Li, Y. Xu, M. Jiang, et al., *Phys. Rev. Lett.*, vol. 123, p. 213902, 2019.
- [49] C. Sathyaseelan, V. Vinothini, and T. Rathinavelan, *bioRxiv*, 2020, <https://doi.org/10.1101/2020.03.16.993352>.
- [50] B. S. Rem, N. Käming, M. Tarnowski, et al., *Nat. Phys.*, vol. 15, p. 917, 2019.
- [51] Y. LeCun, Y. Bengio, and G. Hinton, *Nature*, vol. 521, p. 436, 2015.
- [52] J. Schmidhuber, *Neural Netw.*, vol. 61, p. 85, 2015.
- [53] T. Hofmann, B. Schölkopf, and A. J. Smola, *Ann. Stat.*, p. 1171, 2008, <https://doi.org/10.1214/009053607000000677>.
- [54] I. V. Tetko, D. J. Livingstone, and A. I. Luik, *J. Chem. Inf. Comput. Sci.*, vol. 35, p. 826, 1995.
- [55] D. P. Kingma and J. Ba, arXiv preprint arXiv:1412.6980, 2014.
- [56] A. Liaw and M. Wiener, *R. News*, vol. 2, p. 18, 2002.
- [57] H. Drucker, C. J. Burges, L. Kaufman, A. J. Smola, and V. Vapnik, Eds., “Advances in neural information processing systems,” in *Proceedings of the 9th International Conference on Neural Information Processing Systems*, Princeton, New Jersey, U.S., NIPS’96, 1996, pp. 155–161.
- [58] V. V. Klimov, I. V. Zabkov, A. A. Pavlov, R. C. Shiu, H. C. Chan, and G. Y. Guo, *Optic Express*, vol. 24, p. 6172, 2016.
- [59] F. Pedregosa, G. Varoquaux, A. Gramfort, et al., *J. Mach. Learn. Res.* vol. 12, p. 2825, 2011.
- [60] M. Abadi, A. Agarwal, P. Barham, et al., arXiv preprint arXiv:1603.04467, 2016.