Opto-Electronic Advances

Supplementary information

July 2024, Vol. 7, No. 7

DOI: 10.29026/oea.2024.240062

OptoGPT: A foundation model for inverse design in optical multilayer thin film structures

Taigao Ma[1], Haozhu Wang[2#] and L. Jay Guo[2*]

[1]Department of Physics, University of Michigan, Ann Arbor, Michigan 48109, USA; [2]Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, Michigan 48109, USA.

[#]Work done while at University of Michigan. Currently at Amazon.

*Correspondence: LJ Guo, E-mail: guo@umich.edu

This file includes:

Supplementary information for this paper is available at https://doi.org/10.29026/oea.2024.240062

Section 1: Model implementation

Section 1.1: Material refractive index

In our material database, we selected 18 different materials (Fig. 1(d)) that are widely accessible in many micro/nanofabrication facilities. We experimentally deposited a single layer on silicon substrate and measured their refractive index using ellipsometer. Their refractive index is given in Fig. S1.

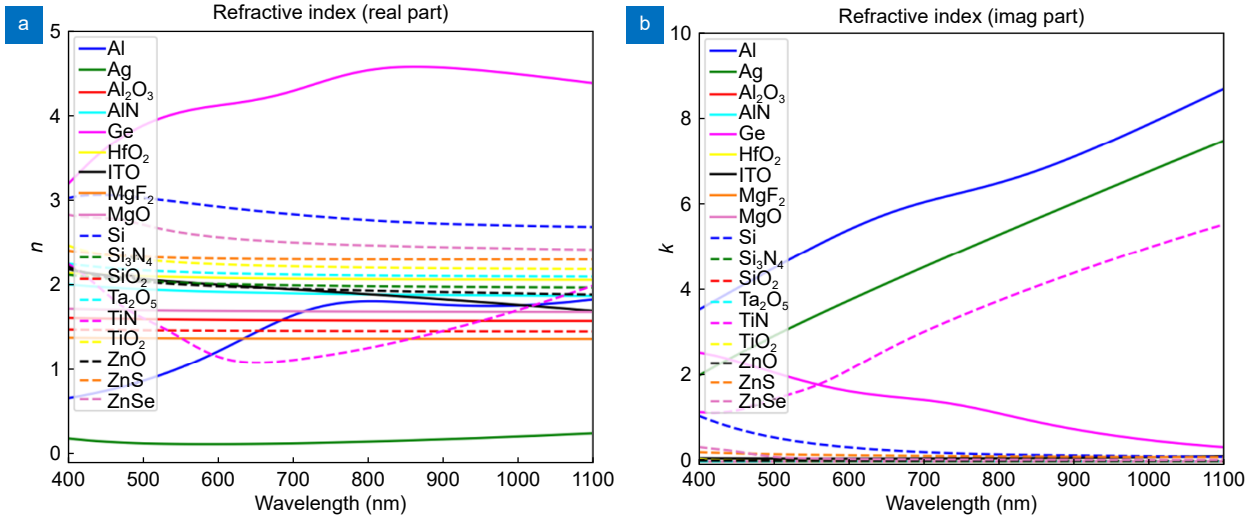


Fig. S1 | Measured refractive index real part (**a**) and imaginary part (**b**) for 18 different materials considered.

Section 1.2: Dataset generation

The measured refractive index is used for simulation during dataset generation. The training dataset and validation dataset consist of 10M and 1M randomly generated samples (we use glass substrate), respectively. Here, the randomness comes from three aspects: material, thickness and the total number of layers. Materials are uniformly sampled from material database and thickness is also uniformly sampled from 10 nm to 500 nm with 10 nm discretization. In addition, we make sure that nearby two layers have different materials. When sampling the total number of layers, considering that the number of possible structures increases exponentially as the total number of layers increases, we sample the number of layers with increasing ratios. We plot the histogram of our generated samples based on the number of layers in Fig. S2(a). Figure S2(b) gives a histogram of the total number of possible structures.

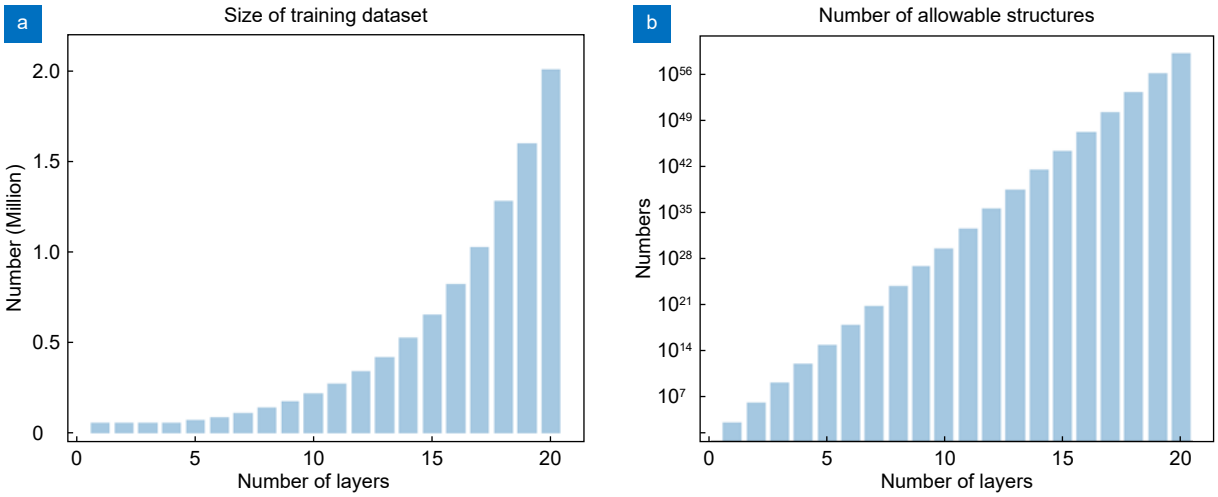


Fig. S2 | (**a**) Histogram of number of generated training data w.r.t. number of layers. (**b**) Histogram of number of allowable structures w.r.t. number of layers, which follows an exponential distribution. For the multilayer structure with twenty layers, the total number of allowable structures reaches 10^{59}.

After the structure is sampled, transfer matrix methods[S1] (TMM) is used to simulated the reflection and transmission spectrum. It took ~1200 h to simulate all 10 M structures on a single CPU and can be faster with parallel computing. It also took ~12 GB to store the generated dataset. We give four examples of generated structures and simulated spectra in Fig. S3.

a

Structure examples in the training datasets

Structure 1: ['MgO_490', 'HfO2_450', 'ITO_470', 'ZnO_20']
Structure 2: ['ZnS_270', 'Al2O3_260', 'Ta2O5_140', 'ZnSe_90', 'Ta2O5_220', 'Si3N4_160', 'ZnO_30']
Structure 3: ['HfO2_420', 'SiO2_230', 'AlN_310', 'MgF2_160', 'ITO_200', 'MgF2_430', 'Al_120', 'Ge_400', 'ZnS_370', 'TiN_90', 'Ta2O5_400', 'Al_180', 'TiN_380', 'TiO2_270', 'SiO2_290']
Structure 4: ['TiO2_360', 'Ta2O5_440', 'TiN_450', 'Si3N4_320', 'MgO_260', 'HfO2_300', 'Al2O3_460', 'ZnS_230', 'Ta2O5_140', 'ITO_140', 'Al_440', 'ZnS_130', 'ZnSe_110', 'ITO_310', 'MgF2_80', 'ITO_30', 'SiO2_30', 'MgO_80', 'Ge_120', 'ZnSe_10']

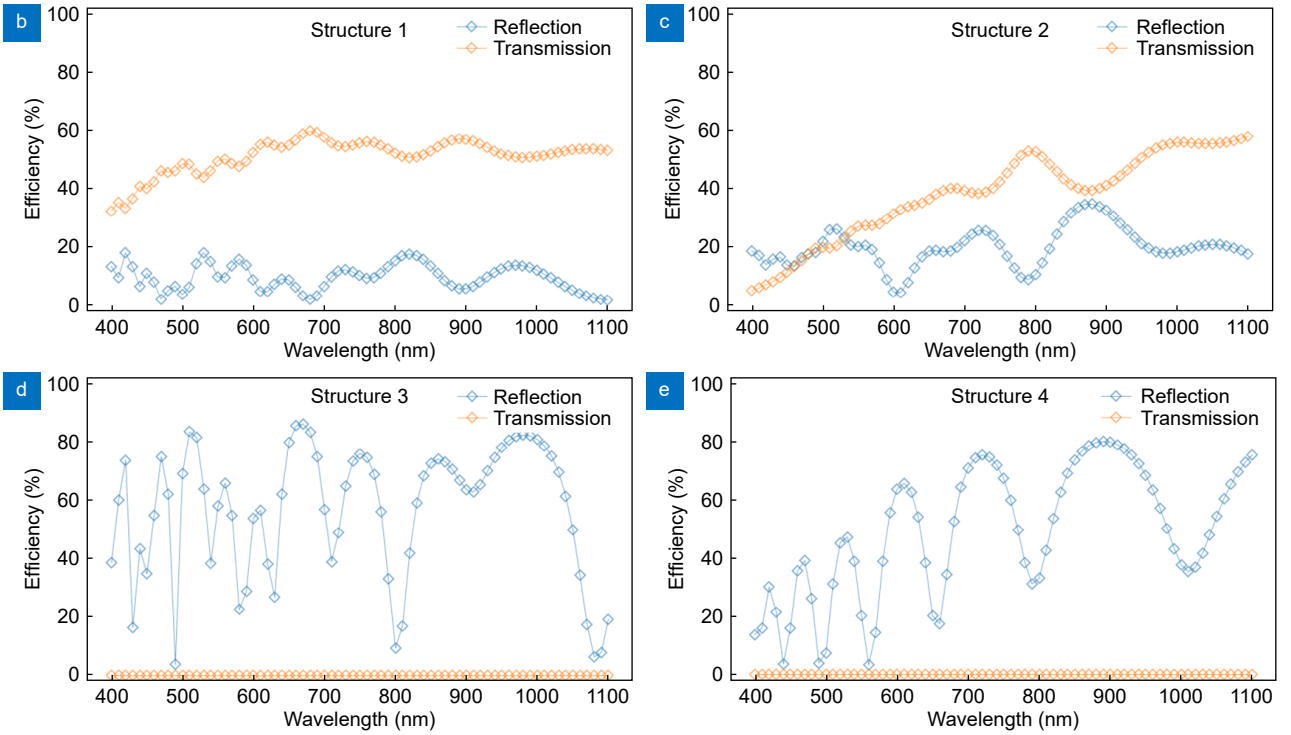


Fig. S3 | Four examples in the training dataset. Structures are given in (**a**) and their transmission and reflection spectrum are given in (**b**, **e**).

Section 1.3: Model architecture and training details

We summarize the hyperparameters used in OptoGPT in Table S1. We use KL divergence as our training loss, with the goal of recovering the input structures from the probability distribution. During training, we use Adam optimizer and warmup procedure. The residual dropout and label smoothing are also used to provide regularization during training. The training loss curve is given in Fig. S4.

Table S1 | Hyperparameters of our OptoGPT.

Hyperparameters	Values
Spectrum dimension	142
Spectrum embedding	Fully connected layers: 142-142-1024
Dimension of hidden representation	1024
Number of decoder blocks (N)	6
Number of attention heads (H)	8
Context Length (K)	22
Dropout rate (p)	0.1
Batch size	1000
Label smoothing (ε_{ls})	0.1
Learning rate (lr)	0.0001
Learning rate decay	Linear warmup and cosine decay (see code for details)
Optimizer	Adam optimizer (β_1 = 0.9, β_2 = 0.98,ε = 10^{-9})
Size of model	58 M
Dataset size	10 M
Training epochs	~200
Training time	~2 week

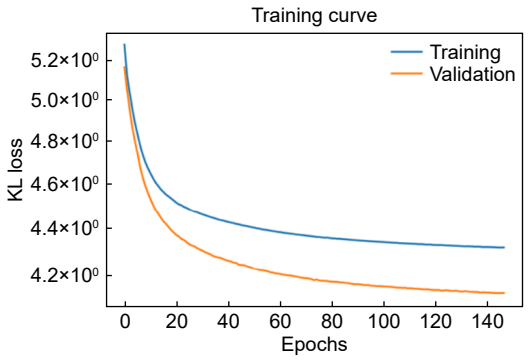


Fig. S4 | Training and validation loss curve. The validation loss is lower than training loss. This is because we are using regularization techniques (e.g., dropout) during training process.

Section 1.4: Visualization of multi-head self-attention

When human reads, usually we do not try to memorize all the words in a sentence. Instead, we selectively focus on these words that are important to form a basic understanding of this sentence. Self-attention is a mechanism that relates each single word with all the other words inside this sentence and selectively focus on several words that are important, similar to how human reads. Multi-head attention allows the model to focus on different aspects. To have a better understanding, in Fig. S5, we give a visualization of the attention map for the following structure:

['SiO2_240', 'Ta2O5_90', 'SiO2_130', 'TiO2_80', 'MgF2_140', 'HfO2_80', 'SiO2_130', 'Ta2O5_100', 'SiO2_130', 'Ta2O5_480', 'SiO2_160', 'Ta2O5_490']

This structure has twelve layers with alternating high-low refractive index profile, similar to the distributed brag reflector (DBR). The attention map is a matrix where each row corresponds to how much attention a single token should be put to other tokens in this sequence. The number of '000' corresponds to the token of 'BOS' (stands for beginning of sequence). It is a common token placed in front of the sequence of structure tokens and used in many other transformer models[S2]. The number of '013' corresponds to the token of 'EOS' (stands for end of sequence). Other numbers in front of each token specify its relative position in this multilayer structure. Usually, it is difficult to understand the physical meaning of each attention map because these machine learning models are black-box. Therefore, we only show attention maps of head 1, head 2 and head 4 in the first decoder block as they may have some meanings. We notice that attention map for head 2 and head 4 focus more in the layers right below and right above, while head 1 focuses more on the long-term alternating relationship (corresponding to the alternating high-low refractive index profile).

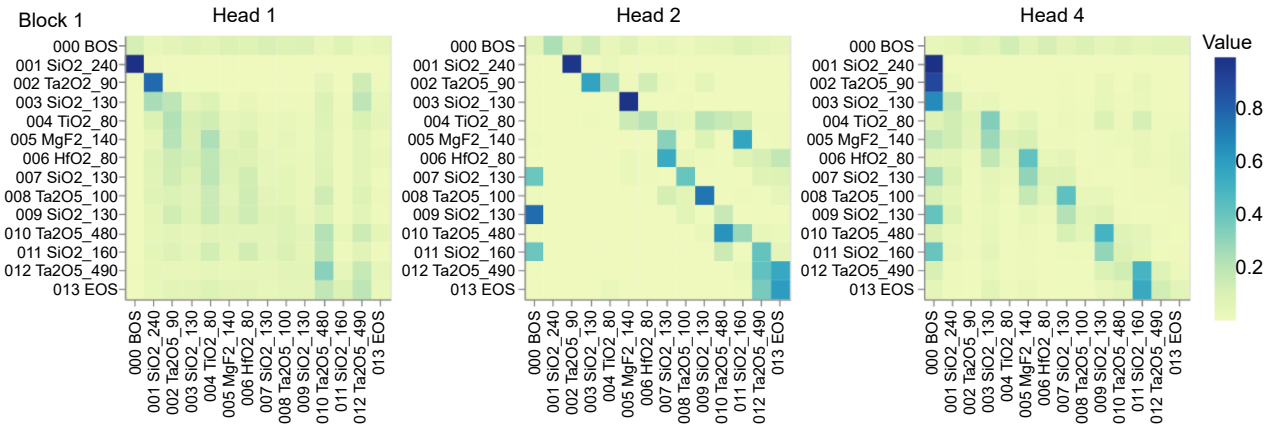


Fig. S5 | Multi-head attention map for the structure ['SiO2_240', 'Ta2O5_90', 'SiO2_130', 'TiO2_80', 'MgF2_140', 'HfO2_80', 'SiO2_130', 'Ta2O5_100', 'SiO2_130', 'Ta2O5_480', 'SiO2_160', 'Ta2O5_490'], which has a high-low index profile. The values at each matrix element are their attention w.r.t each other tokens (normalized to 1).

Section 2: Model performance

Section 2.1: Structure finetuning details

Since we discretize the thickness by 10 nm gap from 10 nm to 500 nm in our training data, the designed structures from our model will also have such discretization. 10 nm gap can be useful for some fabrication tools which cannot guarantee accurate deposition thickness, but may not be sufficient for other tools with high precision deposition, e.g., vacuum deposition. Therefore, we run a thickness finetuning (see Fig. S6(a)) by only optimizing the thickness with the goal of minimizing the Mean Absolute Error (MAE) of spectrum. For the structural color application, we minimize the visual color difference (denoted as ΔE, see Section 2.3). We use the Limited-memory BFGS method[S3] to finetune the thickness and use the designed thickness from our model as the optimization starting point.

In Fig. S6(b, c), we give one example of finetuning the structure 2 in Fig. 4(d) in the main text. The finetuning process (see solid line in Fig. S6(b)) is fast and quickly converges in less than 20 iterations because our model provides a good starting point for optimization. As a comparison, we run the same optimization algorithm but starting from some random points. We simplify the optimization task by using the same materials in structure 2 and only optimize the thickness. We run the optimization five times by starting from five different random points. Dashed lines in Fig. S6(b) show results of convergence and spectrum performance at each iteration. We can find that none of five optimizations show better spectrum performance than our finetuning.

Considering that Limited-memory BFGS method is a local optimization which depends on the starting points, we compare our finetuning with the particle swarm optimization (PSO), which is an advanced global optimization method. We also run five different optimizations starting from some random initial points. On average, it takes PSO 60 iterations to reach the same performance as our designed structure 3. None of these optimizations exceed the performance of finetuning. In other words, the design performance of our model before finetuning is equivalent to running the optimization algorithms for 60 iterations, which saves a lot of time and effort.

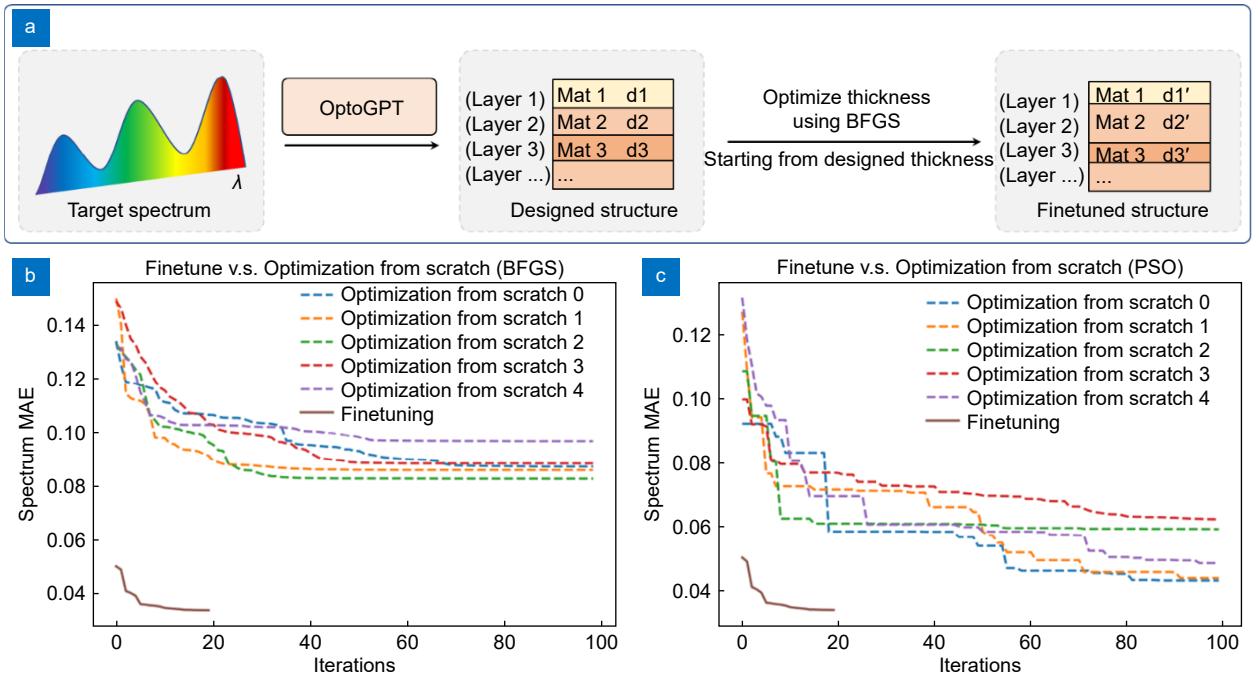
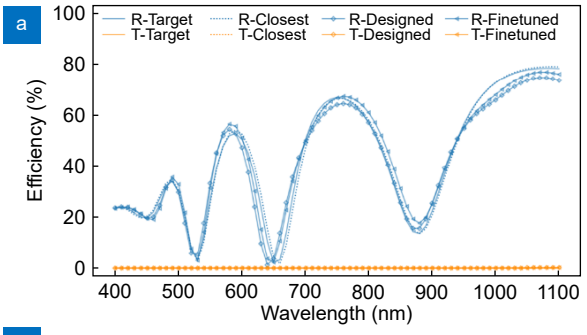
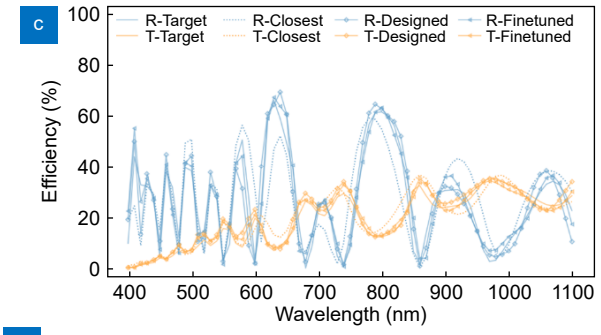


Fig. S6 | Details of finetuning the thickness. (**a**) The procedure of finetuning. We only optimize the thickness by setting the designed structure as the starting point for optimization. (**b**) Compares the converge rate of finetuning v.s. optimization from scratch using the same Limited-memory BFGS method. (**c**) Compares the converge rate of finetuning v.s. optimization from scratch using the particle swarm optimization (PSO). Finetuning gives a better performance and is faster (converges within 20 iterations) than designing from scratch.

Section 2.2: More examples of inverse design in the validation dataset

In Fig. S7, we give two more inverse design examples to visualize the design performance in the validation dataset. Similar to Fig. 4(d) in the main text, we show the target structure that corresponds to the designed spectrum in the validation dataset, the closest structure in the training dataset, five designed structures and the finetuned structure. The MAE in the last column denotes the spectrum performance.





b

	Structure	MAE
Target	['ZnSe_300', 'TiO2_140', 'TiN_200']	
Closest	['ZnSe_300', 'Ta2O5_150', 'TiN_280', 'Al_490', 'AlN_160', 'HfO2_490', 'Ta2O5_170', 'ZnS_60', 'Ta2O5_250', 'MgO_410', 'TiN_50', 'TiO2_470']	0.0063
Designed	**1: ['ZnSe_420', 'Ag_20', 'ZnS_420', 'Ge_360', 'MgF2_170', 'Ge_360']**	**0.0167**
	2: ['ZnSe_420', 'TiN_90']	0.0253
	3: ['ZnSe_440', 'Al_80', 'ZnS_480']	0.0298
	4: ['ZnSe_440', 'Al_30', 'ZnS_460', 'Ge_370', 'Si_280', 'ZnS_480', 'MgF2_70', 'Si_280']	0.0288
	5: ['ZnSe_400', 'ITO_40', 'TiN_70', 'ZnSe_440', 'Ge_40']	0.0200
Finetuned	From design 1: ['ZnSe_423', 'Ag_22', 'ZnS_419', 'Ge_360', 'MgF2_169', 'Ge_359']	**0.0149**

d

	Structure	MAE
Target	['TiO2_210', 'Si3N4_300', 'MgF2_70', 'SiO2_310', 'AlN_460', 'HfO2_250', 'SiO2_110', 'ZnS_300', 'Si3N4_70', 'TiO2_420', 'ZnS_60', 'AlN_480', 'TiO2_190', 'ITO_460']	
Closest	['Si3N4_100', 'HfO2_180', 'ZnO_40', 'Al2O3_340', 'HfO2_290', 'Al2O3_20', 'MgO_270', 'TiO2_340', 'MgF2_60', 'ITO_40', 'ZnSe_50', 'ZnS_290', 'Al2O3_10', 'ITO_450']	0.0579
Designed	**1: ['TiO2_230', 'MgO_410', 'TiO2_110', 'MgO_280', 'HfO2_340', 'Ta2O5_180', 'SiO2_100', 'ZnS_290', 'ITO_330', 'TiO2_330', 'ITO_150', 'ZnO_360', 'Si3N4_380', 'MgF2_30', 'Si3N4_410', 'MgO_390']**	**0.0352**
	2: ['Ta2O5_240', 'Al2O3_440', 'TiO2_250', 'Si3N4_460', 'HfO2_150', 'MgF2_110', 'ZnS_450', 'Al2O3_480', 'TiO2_390', 'ITO_280', 'MgO_440']	0.0422
	3: ['TiO2_220', 'AlN_290', 'MgF2_130', 'Al2O3_240', 'Ta2O5_420', 'TiO2_230', 'SiO2_100', 'Ta2O5_330', 'ITO_420', 'ZnS_370', 'TiO2_420']	0.0445
	4: ['TiO2_210', 'ZnO_310', 'MgF2_380', 'AlN_440', 'Ta2O5_260', 'SiO2_110', 'ZnS_300', 'ITO_360', 'MgO_240', 'TiO2_370', 'ITO_120', 'ZnO_470']	0.0461
	5: ['TiO2_230', 'Si3N4_280', 'MgF2_370', 'AlN_440', 'Ta2O5_260', 'SiO2_120', 'ZnS_300', 'MgO_300', 'ITO_350', 'ZnSe_30', 'TiO2_490', 'ITO_120']	0.0584
Finetuned	From design 1: ['TiO2_225', 'MgO_414', 'TiO2_118', 'MgO_277', 'HfO2_336', 'Ta2O5_181', 'SiO2_102', 'ZnS_301', 'ITO_337', 'TiO2_333', 'ITO_163', 'ZnO_367', 'Si3N4_386', 'MgF2_22', 'Si3N4_413', 'MgO_396']	**0.0266**

Fig. S7 | Two more examples for the inverse design from the validation dataset. (**a**, **c**) are the spectrum performance and (**b**, **d**) give the details of designed and finetuned structures.

Section 2.3: Designing structural color

Because our target spectrum (400 nm to 1100 nm) covers both visible region and near infrared region, it is possible to use our model to design structural color. Here, we select LAB as our color space. There are many other color spaces including RGB, xyY, XYZ. We do not use those color spaces because LAB space is uniform, which makes it convenient to define the color difference[S4]. We use the ΔE to determine the performance of color accuracy. A lower ΔE value indicates greater color accuracy, while a higher ΔE value means a significant color mismatch. Usually, when $\Delta E < 2$, it is visually difficult for human to distinguish the color difference. During inverse design, we first convert the LAB color to spectrum from 400 nm to 1100 nm using a proposed algorithm, then we modify this spectrum to fit for our model's design input. This method can be used to design for both transmissive and reflective structural color.

Section 2.3.1: Convert spectrum to color

For a given spectrum $S(\lambda)$, we first calculate the CIE 1931 XYZ color using:

$$X = \frac{1}{K}\int_{\lambda_1}^{\lambda_2} \bar{x}(\lambda)\, I(\lambda)\, S(\lambda)\, \mathrm{d}\lambda$$

$$Y = \frac{1}{K}\int_{\lambda_1}^{\lambda_2} \bar{y}(\lambda)\, I(\lambda)\, S(\lambda)\, \mathrm{d}\lambda$$

$$Z = \frac{1}{K}\int_{\lambda_1}^{\lambda_2} \bar{z}(\lambda)\, I(\lambda)\, S(\lambda)\, \mathrm{d}\lambda$$

where $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, $\bar{z}(\lambda)$ are the color matching functions (shown in Fig. S8), $I(\lambda)$ is the relative spectral power distribution of the illuminating light source (we use 'D65'). K is a normalizing factor. The $[\lambda_1, \lambda_2]$ is the visible spectrum range and we use 400 nm – 800 nm in our case. LAB is then calculated by doing a conversion from XYZ.

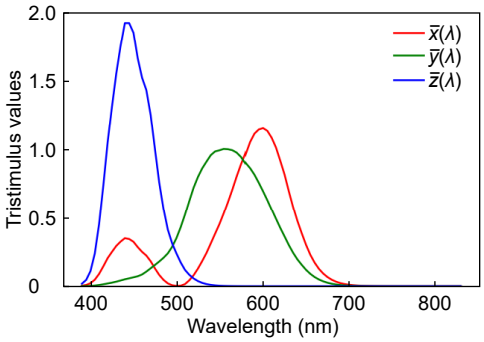


Fig. S8 | The color matching functions for 'CIE 1931 2 Degree Standard Observer'.

Section 2.3.2: Convert color to spectrum using optimization

Consider a three-dimensional color target $LAB_{\text{target}} = [L, a, b]$, we want to convert the color to the 71-dimensional spectrum $S = [s_{\lambda_1}, s_{\lambda_2}, \ldots, s_{\lambda_{71}}]$, where $\lambda_1, \lambda_2, \ldots, \lambda_{71} = 400\ \text{nm}, 410\ \text{nm}, \ldots, 1100\ \text{nm}$. We treat this as an optimization task with two different goals. The first goal is to make the color of the spectrum S as close to the target color as possible. The second goal is to make the spectrum smooth enough since abrupt jumps in spectrum is not physical. We use particle swarm optimization (PSO) to convert color to spectrum S by minimizing the loss:

$$Loss(S) = \Delta E(LAB_{\text{target}}, LAB_{\text{S}}) + \alpha * \sum_{i=1}^{71} \left(\frac{\mathrm{d}^2 S}{\mathrm{d}^2 \lambda_i}\right) ,$$

where LAB_{S} is the color of the converted spectrum. Notice that when calculating color, we only use the portion of visible wavelength from 400 nm to 800 nm. The second term is the second order derivative of spectrum w.r.t. the wavelength from 400 nm to 1100 nm, which quantifies the smoothness of spectrum. α is a factor that balances the loss of color accuracy and smoothness. In Fig. S9, we give one example of converting yellow color to spectrum under different α factors. Notice that the converted spectrum is less smooth when α is smaller.

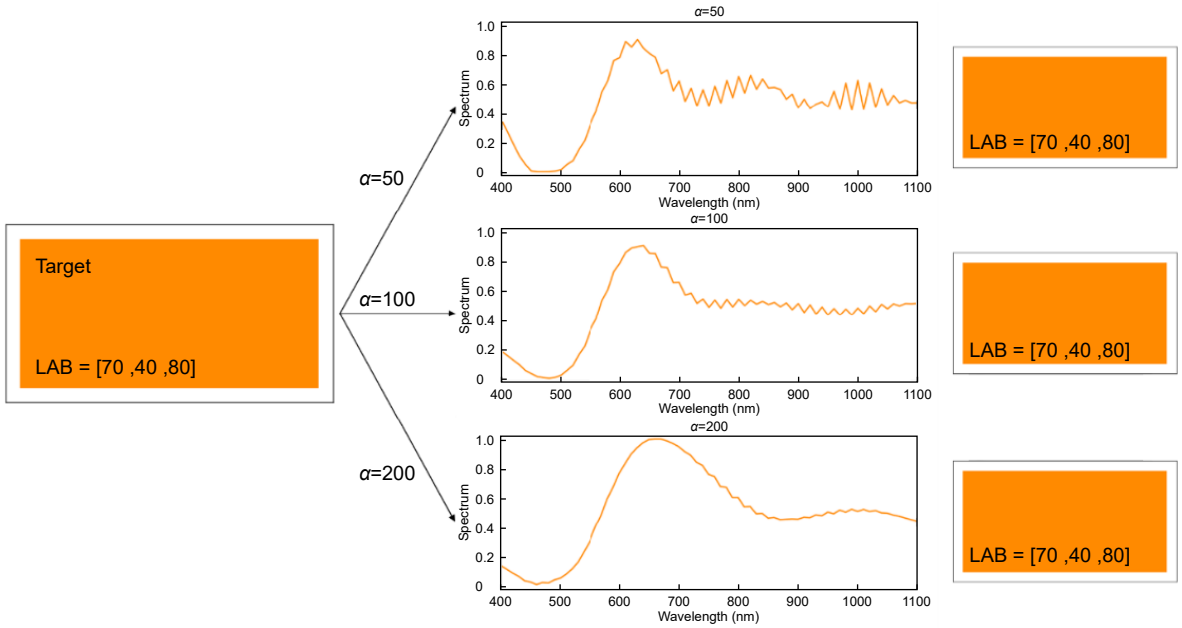


Fig. S9 | Illustration of converting LAB color to spectrum using different *α* factor. As *α* increases, the spectrum is smoother.

Section 2.3.3: More examples of structural color inverse design

In Fig. S10 (a–d), we detailed two more examples of designing the green structural color for both reflective type (a, b) and transmissive type (c, d). The LAB of the green color target is $[70, -80, 0]$. First, we obtain the converted spectrum from the LAB value using the optimization algorithm described above. Since our model takes in both reflection and transmission spectrum, we add extra modification to this converted spectrum. In detail, when designing the reflective type, we set the reflection spectrum to be the converted spectrum and set transmission spectrum to be 0 (see (a)). When designing the transmissive type, we set transmission to be the converted spectrum and set the reflection spectrum to be one minus the transmission spectrum (see (c)). Finally, these modified reflection and transmission spectrum are feed to

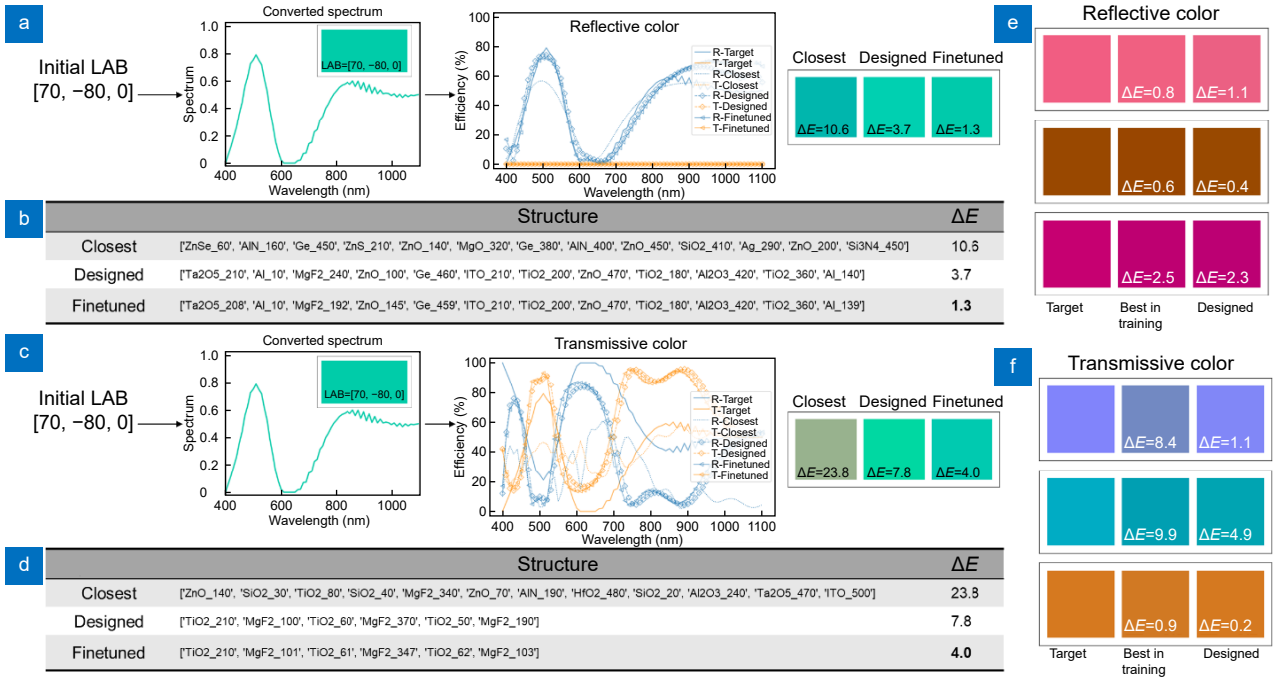
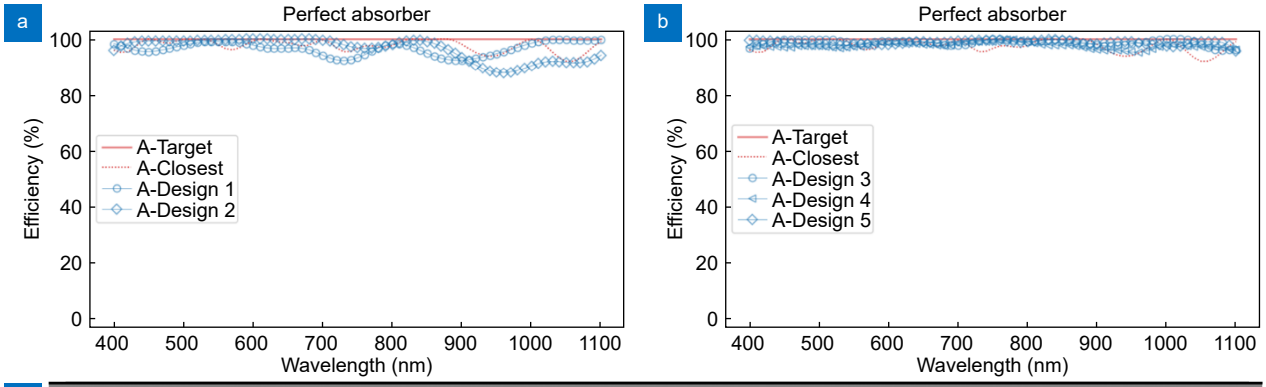


Fig. S10 | One design example with detailed color-to-spectrum conversion for reflective (**a**, **b**) and transmissive (**c**, **d**) type colors. More design examples without these details are shown in (**e**) and (**f**) for reflective and transmissive type color, respectively.

our model for inverse design. We also compare the spectrum performance and color difference in (a) and (c). The detailed structures of the closest in the dataset, the designed structure and finetuned structure are given in (b) and (d). More design examples given in (e–f) and we only give the color visualization results for simplicity.

Section 2.4: More examples of designing absorbers

Figure S11 gives more examples of designing perfect absorber in 400–1100 nm. To design a perfect absorber, we set both transmission and reflection in the target spectrum to be zero. This is because absorption = 100% – transmission – reflection, and a perfect absorber requires 100% absorption. We show five different designs in (c) as well as their absorption spectrum in (a-b). Specifically, structure 1 follows the material arrangement reported in Ref.[S5] while structure 2 follows the material arrangement considered in Ref.[S6]. Their spectra are given in (a). This is done by adding a design constraint on the material arrangement in the first several layers, similar to the constraint 4 in Fig. 6 in the main text. In ref.[S5], the designed structure for the perfect absorber is MgF_2 95.6 nm/SiO_2 14.7 nm/Al_2O_3 76.5 nm/TiO_2 48.0 nm/Si 14.2 nm/Ge 12.0 nm/Ti. In ref.[S6], the designed structure for the perfect absorber is MgF_2 118 nm/TiO_2 56 nm/Si 32 nm/Ge 33 nm/Cr 200 nm/Glass. Even though that 1) materials we are using may have different refractive index, 2) we do not have Ti or Cr in our material and 3) we only use spectrum inside 400–1100 nm, our model can still give similar thickness design for these layers with common materials. In (b), the design 3–5 actually exhibit better absorption performance (~98% absorption on average) than design 1 and 2 in 400–1100 nm. A future work with a close examination into these designs can possibly reveal how to design a perfect absorber.

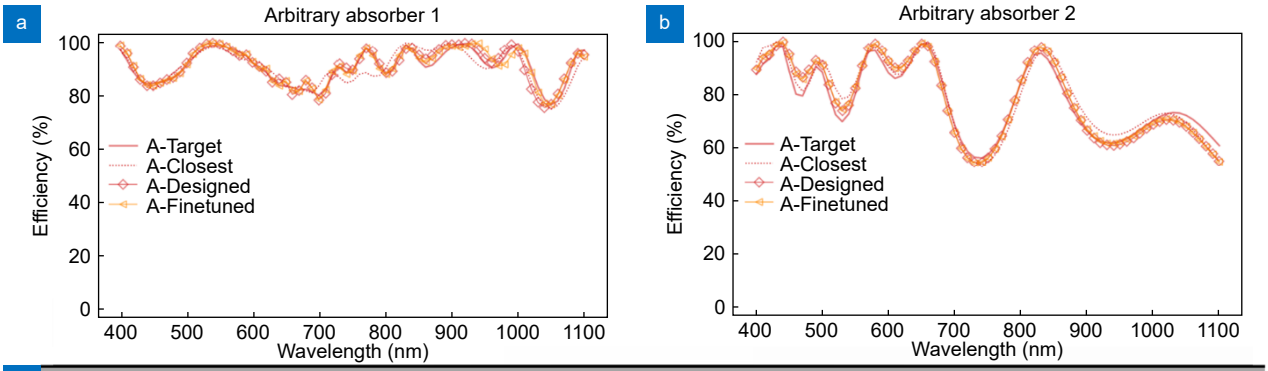


c

	Structure for "Absorbers"	MAE
Closest	['SiO2_90', 'ITO_90', 'TiO2_360', 'ZnSe_330', 'Si_440', 'TiO2_230', 'ZnS_290', 'Si_10', 'Ge_230', 'Al_220', 'MgF2_380', 'ZnS_430', 'ZnSe_260', 'ZnO_320', 'Ge_370', 'ZnSe_480', 'TiN_130']	0.0238
Designed	Design 1: ['MgF2_100', 'SiO2_10', 'Al2O3_90', 'TiO2_50', 'Si_490', 'Ge_420', 'TiN_410']	0.0344
	Design 2: ['MgF2_90', 'TiO2_50', 'Si_480', 'Ge_360', 'TiN_340']	0.0358
	Design 3: ['SiO2_90', 'ITO_70', 'ZnS_480', 'Si_370', 'Ge_220', 'Si_470', 'MgF2_340', 'ZnS_460', 'ZnSe_260', 'Si_260', 'ZnSe_490', 'Ag_420']	0.0120
	Design 4: ['Al2O3_80', 'ITO_50', 'ZnS_410', 'ITO_460', 'ITO_480', 'ZnS_490', 'Si_410', 'ZnS_500', 'Ge_90']	0.0247
	Design 5: ['SiO2_90', 'AlN_60', 'ZnS_500', 'ZnS_420', 'Si_440', 'Ge_310', 'ITO_220', 'ZnSe_370', 'MgF2_340', 'ZnSe_370', 'ZnS_500', 'TiO2_420', 'Ag_450']	0.0139

Fig. S11 | Examples of inverse design perfect absorbers in 400–1100 nm. We give five more designs, with spectrum comparisons shown in (**a**–**b**) and structures shown in (**c**).

In addition, our model can be used to design for arbitrary absorbers. The target arbitrary absorption spectrum is randomly selected from the validation dataset, by doing a calculation of absorption = 100% - transmission – reflection. When converting this arbitrary absorption spectrum to our model input, we set the target transmission spectrum to be zero and target reflection = 100% – target absorption. We give two examples of arbitrary absorber. Their spectra are illustrated in Fig. S12(a–b), and the designed and finetuned structures are given in Fig. S12(c–d).



c

	Structure for "Arbitrary absorber 1"	MAE
Closest	['MgF2_490', 'ZnS_460', 'Si_250', 'Al2O3_360', 'ITO_230', 'Si_380', 'ITO_90', 'MgF2_50', 'HfO2_340', 'Ag_430', 'Al_270', 'Si3N4_470', 'ITO_100', 'ZnS_20', 'Ge_60', 'ZnSe_180']	0.0252
Designed	['Al2O3_420', 'ZnS_500', 'MgO_170', 'ITO_370', 'ZnS_500', 'ZnSe_420', 'SiO2_170', 'ZnSe_260', 'Ag_260']	0.0164
Finetuned	['Al2O3_421', 'ZnS_498', 'MgO_174', 'ITO_371', 'ZnS_482', 'ZnSe_411', 'SiO2_175', 'ZnSe_284', 'Ag_259']	**0.0126**

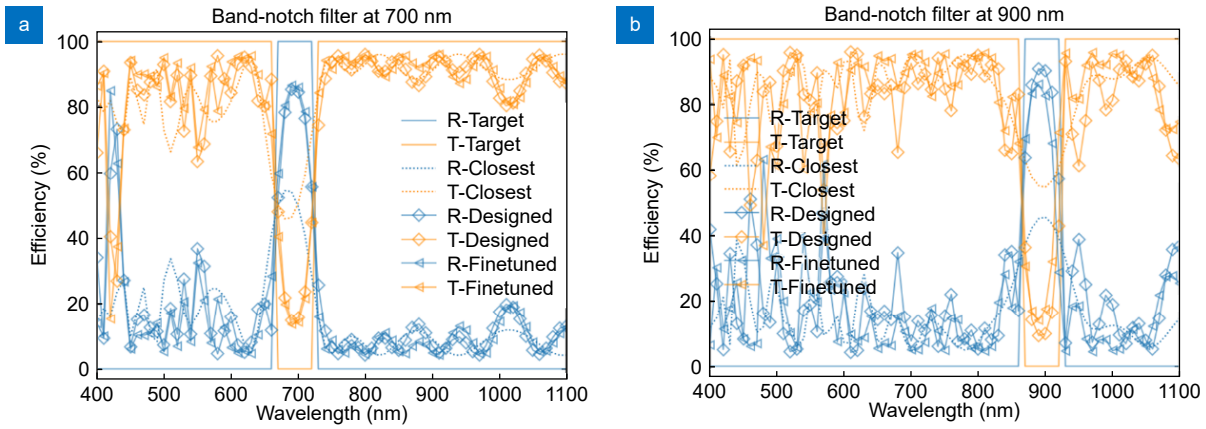
d

	Structure for "Arbitrary absorber 2"	MAE
Closest	['Al2O3_460', 'ZnS_460', 'Ag_190', 'Al_230', 'Ta2O5_450', 'Ge_450', 'MgF2_150', 'ZnS_500', 'Ag_240', 'ZnSe_110', 'SiO2_160', 'ZnSe_90', 'Al2O3_220', 'TiN_370', 'Ag_270', 'Ta2O5_250', 'ITO_130', 'HfO2_310', 'TiN_240', 'Si3N4_90']	0.0308
Designed	['Al2O3_470', 'ZnS_450', 'Ag_60', 'ITO_220', 'ZnS_250', 'Si_270']	0.0254
Finetuned	['Al2O3_468', 'ZnS_451', 'Ag_60', 'ITO_220', 'ZnS_249', 'Si_269']	**0.0250**

Fig. S12 | Two examples of design arbitrary absorbers in 400−1100 nm. The spectrum performance is given in (**a**−**b**), and structures are given in (**c**−**d**), respectively.

Section 2.5: More examples of designing filters

Examples are given for band-notch filters centered at different wavelength.



c

	Structure for "Band-notch filter at 700 nm"	MAE
Closest	['Al2O3_420', 'Si3N4_20', 'TiO2_20', 'Ta2O5_390', 'MgF2_330', 'HfO2_410', 'SiO2_80', 'MgF2_330']	0.1622
Designed	['SiO2_490', 'Al2O3_220', 'Ta2O5_230', 'SiO2_360', 'Ta2O5_250', 'Al2O3_310', 'Ta2O5_270', 'Al2O3_310', 'Ta2O5_430', 'MgO_280', 'HfO2_410']	0.1483
Finetuned	['SiO2_473', 'Al2O3_226', 'Ta2O5_237', 'SiO2_351', 'Ta2O5_249', 'Al2O3_323', 'Ta2O5_247', 'Al2O3_317', 'Ta2O5_428', 'MgO_285', 'HfO2_436']	**0.1368**

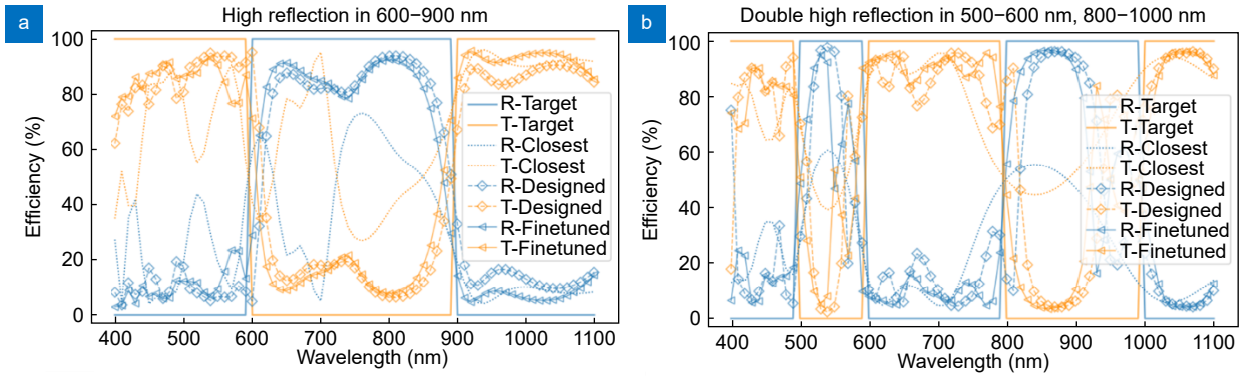
d

	Structure for "Band-notch filter at 900 nm"	MAE
Closest	['Al2O3_460', 'MgF2_420', 'AlN_70', 'Ta2O5_80', 'MgO_330', 'MgF2_30', 'Ta2O5_140', 'HfO2_270', 'Ta2O5_240', 'HfO2_220', 'Ta2O5_100', 'MgO_350', 'SiO2_360']	0.1306
Designed	['Al2O3_70', 'HfO2_110', 'AlN_460', 'MgO_350', 'Ta2O5_140', 'MgO_390', 'AlN_460', 'Ta2O5_330', 'Al2O3_460', 'HfO2_500', 'MgO_410', 'Ta2O5_120', 'MgO_350', 'Ta2O5_340', 'SiO2_190', 'MgO_480', 'Ta2O5_130', 'MgF2_480']	0.1845
Finetuned	['Al2O3_96', 'HfO2_114', 'AlN_423', 'MgO_372', 'Ta2O5_132', 'MgO_388', 'AlN_483', 'Ta2O5_305', 'Al2O3_458', 'HfO2_500', 'MgO_411', 'Ta2O5_135', 'MgO_351', 'Ta2O5_308', 'SiO2_210', 'MgO_415', 'Ta2O5_176', 'MgF2_482']	**0.1667**

Fig. S13 | More examples of inverse design band-notch filters at 700 nm (**a**) and 900 nm (**b**) and structure are given in (**c**−**d**), respectively.

Section 2.6: More examples of designing broadband reflector

Examples are given for broadband reflections centered at different wavelength and different number of bands.



c

	Structure for “High reflection in 600−900 nm”	MAE
Closest	['MgF2_250', 'TiO2_240', 'HfO2_50', 'MgF2_140', 'TiO2_100', 'Ta2O5_150', 'MgF2_500', 'AlN_80', 'SiO2_340', 'HfO2_240']	0.3248
Designed	['MgF2_250', 'TiO2_90', 'MgF2_140', 'HfO2_80', 'MgF2_130', 'TiO2_90', 'MgF2_140', 'Ta2O5_490', 'MgF2_170', 'HfO2_500']	0.1641
Finetuned	['MgF2_252', 'TiO2_87', 'MgF2_131', 'HfO2_79', 'MgF2_132', 'TiO2_77', 'MgF2_147', 'Ta2O5_495', 'MgF2_167', 'HfO2_495']	**0.1445**

d

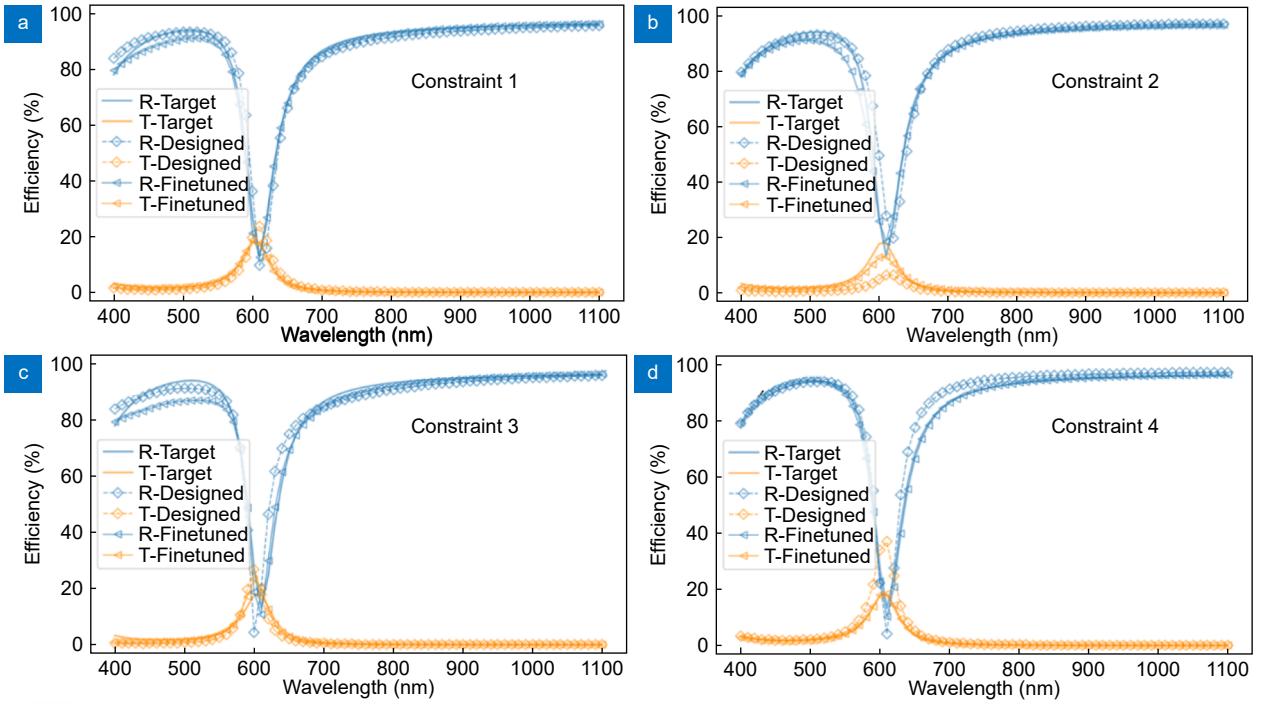
	Structure for “Double high reflection in 500−600 nm, 800−1000 nm”	MAE
Closest	['MgO_20', 'AlN_340', 'SiO2_110', 'ZnO_70', 'SiO2_230', 'Ta2O5_80', 'SiO2_110']	0.3294
Designed	['Al2O3_330', 'Ta2O5_70', 'MgF2_490', 'TiO2_60', 'Al2O3_250', 'Ta2O5_60', 'MgF2_490', 'HfO2_320', 'MgF2_470', 'Ta2O5_90', 'SiO2_460', 'HfO2_310', 'MgF2_490']	0.2428
Finetuned	['Al2O3_318', 'Ta2O5_81', 'MgF2_477', 'TiO2_76', 'Al2O3_212', 'Ta2O5_77', 'MgF2_461', 'HfO2_299', 'MgF2_467', 'Ta2O5_92', 'SiO2_454', 'HfO2_330', 'MgF2_500']	**0.1970**

Fig. S14 | More examples of inverse design broadband reflector. We give the details for the task “High Reflection in 600−900 nm”, “Double High Reflection in 500−600 nm, 800−1000 nm” here. Their structures are given in (**c**−**d**), respectively.

Section 3: Design flexibility

Section 3.1: Results of thickness finetuning when designing FP resonator

Finetuned results for the designed structures in Fig. 6 in main text are given here. The peak position of the designed structures may deviate from the design target due to discretized thickens. The thickness finetuning can remove such limitations and make the inverse design more accurate.



e

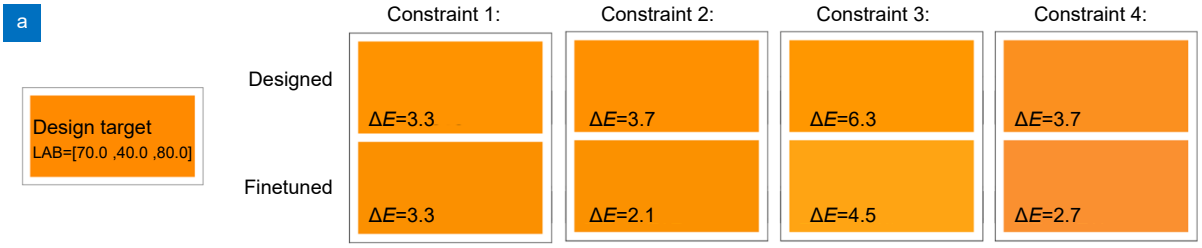
	Structure for FP Resonator		MAE	MAE (Finetuned)
Constraint 1	['SiO2_100', 'Ag_30', 'AlN_110', 'Ag_50']	Finetuned: ['SiO2_71', 'Ag_29', 'AlN_107', 'Ag_52']	0.0129	0.0071
Constraint 2	['Ag_30', 'Al2O3_150', 'Al_30', 'MgF2_400']	Finetuned: ['Ag_28', 'Al2O3_146', 'Al_23', 'MgF2_441']	0.0175	0.0105
Constraint 3	['Si3N4_60', 'Ag_40', 'Al2O3_140', 'Ag_50']	Finetuned: ['Si3N4_60', 'Ag_40', 'Al2O3_140', 'Ag_50']	0.0186	0.0138
Constraint 4	['Ag_30', 'Si3N4_100', 'Ag_40']	Finetuned: ['Ag_21', 'Si3N4_97', 'Ag_51']	0.0175	0.0049

Fig. S15 | Detailed results of finetuning designs with constraints in Fig. 6 in the main text. (**a**–**d**) shows the spectrum of finetuned structures with constraints 1–4, respectively. (**e**) shows the detailed structures and spectrum performance.

Section 3.2: Another example of design flexibility: designing structural color

We use another example of designing a transmissive-type orange structural color to demonstrate the design flexibility. Details of designed structures and finetuned structures are also given.

a



b

Design restriction	Structures	ΔE
Target	An orange reflective structural color with LAB = [70, 40, 80]	
Closest in data	['ZnO_150', 'TiN_50', 'Ta2O5_12C', 'Al_70', 'Ag_30', 'MgF2_30', 'Si_170', 'ZnO_190', 'ITO_420', 'HfO2_500', 'Al2O3_40', 'ZnS_260', 'Ag_270', 'Si_340', 'TiN_480', 'ZnSe_10']	4.3
Constraint 1: Fix the first layer to be 70 nm ZnO	Designed: ['ZnO_70', 'SiO2_60', 'ZnSe_40', 'SiO2_80', 'ZnSe_40', 'SiO2_60', 'TiO2_60', 'MgF2_50'] Finetuned: ['ZnO_50', 'SiO2_82', 'ZnSe_31', 'SiO2_106', 'ZnSe_30', 'SiO2_82', 'TiO2_45', 'MgF2_91']	3.3 3.3
Constraint 2: Limit the first layer in [10, 200] nm	Designed: ['ZnSe_30', 'MgF2_60', 'ZnSe_40', 'MgF2_110', 'ZnSe_40', 'MgF2_240', 'HfO2_80'] Finetuned: ['ZnSe_25', 'MgF2_92', 'ZnSe_24', 'MgF2_111', 'ZnSe_41', 'MgF2_267', 'HfO2_60']	3.7 2.1
Constraint 3: Only use SiO_2 and TiO_2	Designed: ['TiO2_60', 'SiO2_50', 'TiO2_60', 'SiO2_60', 'TiO2_50', 'SiO2_100', 'TiO2_50', 'SiO2_80', 'TiO2_160'] Finetuned: ['TiO2_49', 'SiO2_80', 'TiO2_49', 'SiO2_81', 'TiO2_49', 'SiO2_81', 'TiO2_49', 'SiO2_79', 'TiO2_137']	6.3 4.5
Constraint 4: Specify the material at each layer to be TiO_2/MgF_2/ZnSe/SiO_2/ZnSe and design the thickness only	Designed: ['TiO2_60', 'MgF2_250', 'ZnSe_40', 'SiO2_100', 'ZnSe_130'] Finetuned: ['TiO2_48', 'MgF2_262', 'ZnSe_45', 'SiO2_79', 'ZnSe_134']	3.7 2.7

Fig. S16 | Another example of design flexibility for transmissive orange structural color. Here we consider four different constraints "1: Fix the first layer to be 70 nm ZnO", "2: Limit the first layer in [10, 200] nm", "3: Only use SiO_2 and TiO_2", "4: Specify the material at each layer to be TiO_2/MgF_2/ZnSe/SiO_2/ZnSe and design the thickness only". The results of color impression are given in (**a**). The designed structures are given in (**b**).

Section 4: Generalization ability

Section 4.1: Finetuning model for different angles and polarization using smaller training set

The inverse design performance of pretrained model and finetuned model on the angled-resolved spectrum with different polarizations are compared here. Since the pretrained model is trained on 0° incident spectrum, it does not perform well when designing for angled spectrum. The distribution of MAE moves to the right side as the incident angle of designed spectrum increases.

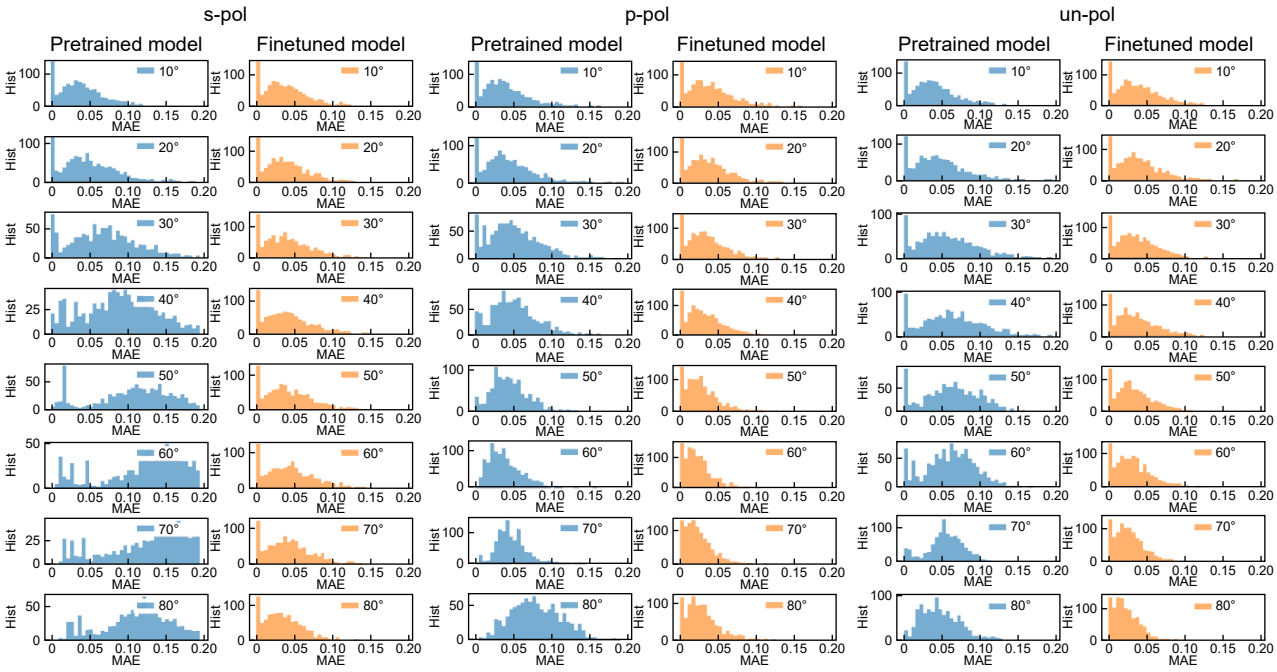


Fig. S17 | Histogram of inverse design performance on different angles (10°–80°) and polarization states (s-pol, p-pol, and un-pol) for pretrained model and finetuned model. The pretrained model is the original model trained on a large 10M dataset with normal incident spectrum, while the finetuned model is finetuned on a smaller 1M dataset with specified angle and polarization state. Based on the high MAE, we can see that the pretrained model is bad for angled/polarized spectrum inverse design. However, after finetuning, they exhibit much better design performance, which can be verified by the small MAE.

Section 4.2: More examples for mixed sampling

Examples are given for simultaneously designing multiple spectrums at different incident angles and polarization states.

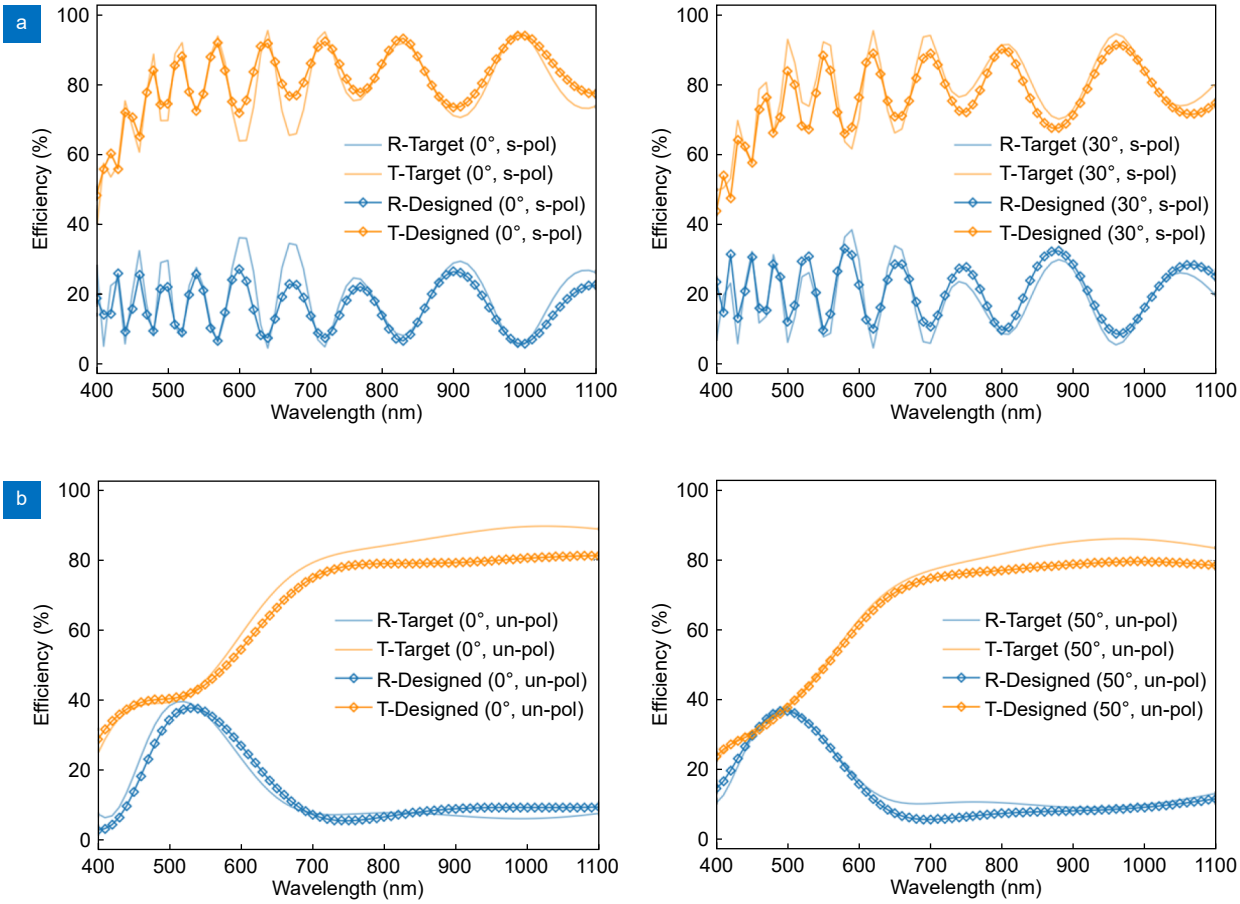


Fig. S18 | Two more examples for designing structures that can satisfy multiple angles and polarization states simultaneously. (**a**) Designing for both 0° s-polarized spectrum and 30° s-polarized spectrum. (**b**) Designing for both 0° un-polarized spectrum and 50° un-polarized spectrum.

Section 5: More discussions

Section 5.1: Data and code availability

We believe our model can facilitate researchers and engineers working in multilayer thin structures in many aspects. Upon publication, we will post our training data for others to explore other designs of multilayer structures. Due to patent filing, we are restricted to share our code at this point; but could become available at a later time.

Section 5.2: Limitations of our model

As mentioned in the discussion in the main text, our model is trained on a limited dataset, where the spectrum region and types, material, thickness, number of layers are restricted. Therefore, at this moment, there are some applications that our model cannot deal with. For example, designing for radiative cooling (requires spectrum coverage to 2 μm). In addition, our dataset (10^7) is also small compared to the enormous size of possible structure (10^{59}). Therefore, we cannot expect the model can always perform well when designing for different spectra. In Fig. S17, we give one such example for designing "band-pass filter at 750 nm" where our model completely fails. The big difference of the spectrum from the closest structure in the training dataset confirms that our training dataset does not contain sufficient data for our model to learn.

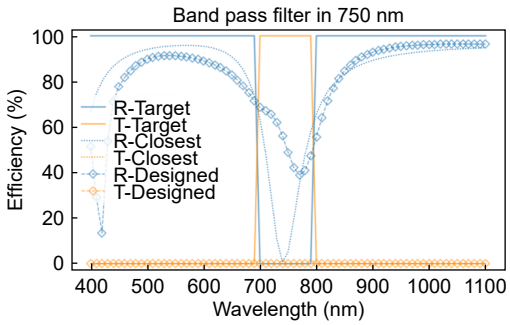


Fig. S19 | One example when our model fails to give a structure for the "band-pass filter in 750 nm". The target spectrum requires a transmission at 750 nm while the designed structure does not have any transmission. The closest structure in the dataset also does not have transmission, which means this type of spectrum target is outside of our training dataset. This is because our training dataset is still small (10^7) compared to the total possible structures (10^{59}).

Section 5.3: Comparison of our model with existing methods

In Table S2, we compare our model with existing methods from five aspects:

1. Versatile for different types of structures: design the total number of layers, material arrangements and thickness simultaneously.

2. Versatile for design targets: quickly adapt to different design targets without restarting the design process.

3. Generalized to angled incidence and polarizations: to suit for a wider range of application situations with different angled incidence and polarization states.

4. Multiple design: In principle, we can always get different designs if we restart the optimization process or retrain the model from different random points. It is desirable that the method can intuitively give multiple designs without doing so.

5. Flexible design: to incorporate different design constraints without restarting the optimization process or retraining the model.

From the table, we can see that our model demonstrates promising performance outperforming existing methods.

Table S2 | Comparison of our method with existing methods. [a]: GLOnet is based on optimization and OML-PPO is based on reinforcement learning so they do not need training data. [b]: The model size (number of parameter) is not provided in the original paper or code repository, so we estimate it based on their model architecture.

Methods	ref.	Model size (M)	Training dataset size (M)	Demonstrated design targets	Versatile for structures?	Versatile for design targets?	Generalized to angles and pol?	Multiple design?	Flexible design?
PSO	ref.[S7]	---	---	Filters	No	No	No	No	No
Genetic algorith	ref.[S8]	---	---	Antireflection coatings	No	No	No	No	No
Needle optimization	ref.[S9]	---	---	Optical coatings	No	No	No	No	No
Memetic optimization	ref.[S10]	---	---	Radiative cooling	Yes	No	No	No	No
GLOnet	ref.[S11,S12]	~0.2	---[a]	Optical transfer function	Yes	No	No	No	No
OML-PPO	ref.[S13]	~0.5	---[a]	Absorber	Yes	No	No	Yes	No
Tandem network	ref.[S14]	~0.2[b]	~0.5	Transmission	No	Yes	No	No	No
GAN	ref.[S15]	~1.1[b]	~0.05	Structural color	No	Yes	No	Yes	No
MDN	ref.[S16,S17]	~3.6[b]	~0.1	Transmission	No	Yes	No	Yes	No
MST	ref.[S5]	~14	~0.14	Absorber	No	Yes	No	No	No
OptoGPT	(Ours)	~56	~10	Multiple	Yes	Yes	Yes	Yes	Yes

References

S1. Byrnes SJ. Multilayer optical calculations. arXiv: 1603.02720 [physics] (2020). https://doi.org/10.48550/arXiv.1603.02720

S2. Devlin J, Chang MW, Lee K et al. BERT: pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* 4171–4186 (Association for Computational Linguistics, 2019); http://doi.org/10.18653/v1/N19-1423.

S3. Zhu CY, Byrd RH, Lu PH et al. Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. *ACM Trans Math Softw* **23**, 550–560 (1997).

S4. Wang DY, Liu ZY, Wang HZ et al. Structural color generation: from layered thin films to optical metasurfaces. *Nanophotonics* **12**, 1019–1081 (2023),

S5. Chen W, Gao Y, Li YY et al. Broadband solar metamaterial absorbers empowered by transformer-based deep learning. *Adv Sci* **10**, 2206718 (2023).

S6. Yang CY, Ji CG, Shen WD et al. Compact multilayer film structures for ultrabroadband, omnidirectional, and efficient absorption. *ACS Photonics* **3**, 590–596 (2016).

S7. Rabady RI, Ababneh A. Global optimal design of optical multilayer thin-film filters using particle swarm optimization. *Optik* **125**, 548–553 (2014).

S8. Schubert MF, Mont FW, Chhajed S et al. Design of multilayer antireflection coatings made from co-sputtered and low-refractive-index materials by genetic algorithm. *Opt Express* **16**, 5290–5298 (2008).

S9. Tikhonravov AV, Trubetskov MK, DeBell GW. Application of the needle optimization technique to the design of optical coatings. *Appl Opt* **35**, 5493–5508 (1996).

S10. Shi Y, Li W, Raman A et al. Optimization of multilayer optical films with a memetic algorithm and mixed integer programming. *ACS Photonics* **5**, 684–691 (2018).

S11. Zhang XM, Bai BF, Sun HB et al. Incoherent optoelectronic differentiation based on optimized multilayer films. *Laser Photonics Rev* **16**, 2200038 (2022),

S12. Jiang JQ, Fan JA. Global optimization of dielectric metasurfaces using a physics-driven neural network. *Nano Lett* **19**, 5366–5372 (2019).

S13. Wang HZ, Zheng ZY, Ji CG et al. Automated multi-layer optical design via deep reinforcement learning. *Mach Learn Sci Technol* **2**, 025013 (2021).

S14. Liu DJ, Tan YX, Khoram E et al. Training deep neural networks for the inverse design of nanophotonic structures. *ACS Photonics* **5**, 1365–1369 (2018).

S15. Dai P, Sun K, Yan XZ et al. Inverse design of structural color: finding multiple solutions *via* conditional generative adversarial networks. *Nanophotonics* **11**, 3057–3069 (2022).

S16. Unni R, Yao K, Zheng YB. Deep convolutional mixture density network for inverse design of layered photonic structures. *ACS Photonics* **7**, 2703–2712 (2020).

S17. Unni R, Yao K, Han XZW et al. A mixture-density-based tandem optimization network for on-demand inverse design of thin-film high reflectors. *Nanophotonics* **10**, 4057–4065 (2021).