A Reinforcement Learning Method for Optical Thin-Film Design

Anqing JIANG¹, Nonmember and Osamu YOSHIE†², Member

SUMMARY  Machine learning, especially deep learning, is dramatically changing the methods associated with optical thin-film inverse design. The vast majority of research has focused on parameter optimization (layer thickness, and structure size) of optical thin-films. A challenging problem that arises is an automated material search. In this work, we propose a new end-to-end algorithm for optical thin-film inverse design. This method combines the ability of unsupervised learning, reinforcement learning and includes a genetic algorithm to design an optical thin-film without any human intervention. Furthermore, with several concrete examples, we have shown how one can use this technique to optimize the spectra of a multi-layer solar absorber device.

key words: optical thin-film design, reinforcement learning, neural combinatorial optimization

1. Introduction

In recent decades, significant fundamental advances combined with the spectacular progress of nano-scale fabrication methods have led to a broad range of innovations in the design of optical thin-film. Many applications, such as broadband filter [1]–[3], solar absorber [4]–[6], and radiative cooling device [7], [8], increasingly rely on the intricate nanostructure design for greater performance at target wavelengths. Most of researchers make such designs based on human intelligence to solve a fundamental photonic problem: choosing the best combination of materials and nano-structure of a layered optical thin-film. Human intelligence is often limited and the highest performance of selecting a layered thin-film cannot be achieved based solely on a researcher’s intuition. Human expert-based design is slow, and the performance of selecting an appropriate film is “un-perfect”, especially in cases when the design target is complicated.

With the development of computer-aided design (CAD) technology, the inverse design has gained significant attention as a powerful approach to design layered optical thin-film without human experience. Several computational algorithms have been proposed to solve the inverse design problem, such as the evolutionary algorithm [9], a genetic algorithm (GA) [10], [11], the needle algorithm [12], [13], and the particle swarm optimization (PSO) [14]. However, the design process based on computational algorithms is often time-consuming or computationally-intensive. For a complex nano-structure and board wavelength design target, these computational algorithms take a lot of computation time that is produced by the electromagnetic (EM) simulation, such as rigorous coupled-wave analysis (RCWA), the finite element method (FEM), the finite difference time domain (FDTD), and transfer matrix method (TMM). All of these methods are time and computationally expensive. In contrast, deep learning (DL) based algorithms, which are considered as being able to “learn” Maxwell’s equations, were proposed to solve the nonlinear relationships between the structural parameters and film’s performance by a large dataset [15]–[18]. Learning the design process by human experts, and reinforcement learning are other solutions to solve this problem, which train an agent to learn about the parameter space of a series by exploration [19]–[21]. While the previously described algorithms have worked well for structural parameters optimization (nano-structure size and layer thickness) for a in thin-film inverse design well, there is little research on the design of component materials for these processes.

Generally, the simultaneous optimization of structural and material parameters is a combinatorial optimization problem. A challenging problem that arises in this field is material parameters, such as the complex refractive index $R(\lambda) = n + ik$, where $\lambda$ is the wavelength. The direct processing of high-dimensional features leads to additional time and poorer optimizer performance, or even absolute failure of the inverse design. One intuitive method is to code the materials, which is applied by inverse design methods on a very small scale [22], [23]. These two methods address the following materials’ parameter optimization problem. The scope of these searches is complex and exponentially greater for a larger number of materials. A variety of component materials, such as metals, semiconductors, alloys, transparent material have been widely used in layered optical thin films to best match the design target. Simply encoding materials numerically in the form of key-value pairs, such as (Au, 1), (Ag, 2), (Cu, 3), is not a scalable approach. For supervised learning-based inverse design methods, such massive class categories result in an insufficient number of samples that have been used in the training. For the reinforcement learning based inverse design methods, substantial materials result in large discrete action spaces, that bring reinforcement learning to a larger class of problems. Unsupervised deep neural net-
works, e.g., autoencoder (AE) [24] and variation autoencoder (VAE) [25], work well for feature extraction. These methods achieve great success in generating abstract features with high dimensional data [26], [27]. Low-dimensional semantic space can be extracted by unsupervised-learning-based feature extraction methods from high-dimensional features by VAE-tSNE (variational autoencoder stochastic neighbor embedding) method [28]. In this technique, the model automatically learns a distribution of clusters and naturally creates a multi-scale representation. In this work, we use a VAE-tSNE feature reduction method to map the high dimensional complex refractive index of a material onto a two-dimensional semantic space. This method allows the environment space size of material not to increase with increasing of alternative materials.

We propose and test an implementation of an end-to-end reinforcement search system. This technique is the combination of an reinforcement learning and a genetic algorithm to design a layered optical thin-film from about 300 materials. The reinforcement learning environment space is based on a two-dimensional semantic space extracted by VAE-tSNE. In the following sections, we present the details of this algorithm’s formulation and search space. Furthermore, we use our proposed search system to design and optimize a layered solar absorber device.

2. Design Target

The spectrum of an optical thin-film $S(\lambda|R, d)$ = [Absorption($\lambda$), Transmission($\lambda$), Reflection($\lambda$)] is determined by its component materials’ refractive index $R = [R_1, \ldots, R_m]$, structural parameters (number of layers $m$, layer thickness $d = [d_1, \ldots, d_m]$), and fabrication errors $\epsilon(\lambda)$. Fabrication errors often result from detects of the manufacturing equipment and errors in the manufacturing processes. In practice, as the wavelength range is continuous, the optimization design problem of thin-film can be described in Eq. (1), which is used to minimize the nonlinear least square problem between a given structure $S(\lambda|R, d)$ and the target spectrum $S^*(\lambda)$.

$$\min_{R,d} \sum_\lambda (S(\lambda|R, d) - S^*(\lambda))^2.$$  

(1)

3. Method

Having defined the optimization problem in Eq. (1), we can now outline the use of the memetic algorithm in finding the optimal multi-layer structure. Similar to the human design process, where well-designed films always improve on previous experience, the algorithmic optimization of optical films is performed step-by-step and can be considered as a Markov decision process (MDP). Reinforcement learning, a branch of machine learning, has been proposed to solve the MDP problem through an exploration-reward. In this paper, we use a reinforcement learning algorithm, called asynchronous advantage actor-critic (A3C), to find the best optical thin-film structure, This method allows running multiple agents in parallel instead of using only one while updating the shared network periodically and asynchronously. Figure 1 shows the material selection system diagram.

3.1 Environment Space

Making a learner-friendly environment is one of the central challenges faced by reinforcement learning to solve a specific problem. With a common reinforcement learning environment, as Go [29], Atari [30] and Duckietown [31], the agent explores and learns in a 2D or 3D environment space. The refractive index of the materials of an optical thin-film is continuous high-dimensional data, especially in the case of broadband optimization, and therefore cannot be directly characterized in 2D or 3D space. In this manuscript, we present a novel application of VAE-tSNE to embed the high-dimensional material refractive index to semantically relevant 2D latent variables. Figure 2 shows the environment space. Specifically, we trained 2-input 1D-CNN VAE with 5, 10, 15 and 20 units in a hidden layer. The input of 2-input 1D-CNN VAE is corresponded optical parameters into discrete values at 1nm intervals in target wavelength range.
Table 1  Definition of actions used in A3C. \( \Delta x \) is the step in the environment \( x \) axis. \( \Delta y \) is the step in the environment \( y \) axis.

<table>
<thead>
<tr>
<th>Actions number</th>
<th># of layers</th>
<th>( \Delta x )</th>
<th>( \Delta y )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>0.01</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0.01</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>(-0.01)</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0</td>
<td>(-0.01)</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>0.01</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>0</td>
<td>0.01</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>(-0.01)</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>0</td>
<td>(-0.01)</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>(-0.01)</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>0</td>
<td>(-0.01)</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>0.01</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
<td>0</td>
<td>0.01</td>
</tr>
<tr>
<td>12</td>
<td>4</td>
<td>(-0.01)</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>4</td>
<td>0</td>
<td>(-0.01)</td>
</tr>
<tr>
<td>14</td>
<td>4</td>
<td>0</td>
<td>(-0.01)</td>
</tr>
<tr>
<td>15</td>
<td>4</td>
<td>0.01</td>
<td>0</td>
</tr>
</tbody>
</table>

(280nm-800nm). When training the VAE-tsNE method, we use over 800 materials. The latent space on the 20 units gets a minimum loss (0.003), after training 1,000 epochs. The tsNE embeds the 20 units latent vectors from the VAE generator for space building. By observing the environmental space, materials with similar properties are distributed in close proximity in this space after dimensionality reduction by VAE-tsNE. Considering that some of the materials have too similar optical performance and have the same application properties in practice. We deleted all glass materials except SiO2. Further, alloy materials can be mixed in arbitrary proportions, and we have streamlined the proportion of alloy material. Finally, we use 300 materials to build the environment. For evaluating the design performance of films, this environment of DQN use on one kind of multilayer optical film simulation algorithm called the transfer matrix method (TMM)[32].

3.2 State

The state is a 2D array of the material parameters for structure. The total number of possible states for a 5 layers film is, therefore, 100^5 = 100^2 * 100^2 * 100^2 * 100^2 * 100^2 = 10^{10}. Manually searching all of these states is impossible. However, using A3C can produce desirable results in a reasonable time.

3.3 Actions

Actions determine the material changes that are needed to be applied to the optical thin-film. It is not feasible to set the position of material directly as action because the A3C agent is hard to train with a large number of discrete actions. To deal with this problem, the material actions are defined to change the material from the environment space. Table 1 shows a list of all the actions for a design example of a 5-layer optical thin-film design.

3.4 Rewards

Reward design enables the robustness of an RL system. To move our material design system in a certain and correct direction, we use both discrete-reward and continuous-reward. Considering that the environment space is sparse in the optimization process, we use this method of reward shaping to reduce the sparse payoff problem [33].

Observation reward is an RMSE (root mean squared error) loss type reward shaping function, which can be expressed as

\[
R_{\text{Obs}}(n, k, d) = \Sigma_{\lambda, \theta} W(\lambda) [S(\theta, \lambda; n, k, d) - S^*(\lambda, \theta)].
\]  

(2)

Reward 0 and Reward 1 can restrict the agent from choosing meaningless pairwise actions (x minus 0.1, x plus 0.1) to gain a meaningless reward. The observation reward (Reward 3) provides the agents with a decaying reward system, which is designed to perform more exploitation at the beginning of each agent’s action and more exploration at the end of each agent’s action.

3.5 Proximity Material Matching

We can see from Fig. 2 that our environment is sparse. This means that most of the environmental space does not correspond to a single material. To solve this problem, we define a rule that any point corresponding to a blank point is regarded as the closest material, as shown in Fig. 3. This operation maps every point in the environment space to corresponding states. As a result the A3C agents will learn the policy of reinforcement learning.

3.6 Thickness Optimization

In the thickness optimization process, to reduce the amount of computation in the overall material search, each material combination is optimized only once by the thickness optimization algorithm. Compared with several traditional
Table 2  Definition of rewards used in A3C.

<table>
<thead>
<tr>
<th>Reward Number</th>
<th>Situation</th>
<th>Reward value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Film performance is not improved in the threshold step</td>
<td>-1</td>
</tr>
<tr>
<td>2</td>
<td>Film performance is not improved</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>Film performance is improved</td>
<td>Observation reward</td>
</tr>
<tr>
<td>4</td>
<td>Film performance meets the target</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 3  Results of the thickness optimization by traditional methods. Six heuristic optimization algorithms are used. The column “Best” reports the largest value of absorption for each algorithm. Std represents the stability of each algorithm. Time is the average execution time to complete the optimization (in seconds).

<table>
<thead>
<tr>
<th>Type</th>
<th>Best</th>
<th>Std</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>8-layers solar absorber</td>
<td>0.88</td>
<td>0.02</td>
<td>35</td>
</tr>
<tr>
<td>GA</td>
<td>0.90</td>
<td>0.04</td>
<td>55</td>
</tr>
<tr>
<td>PSO</td>
<td>0.87</td>
<td>0.07</td>
<td>20</td>
</tr>
<tr>
<td>GASA</td>
<td>0.90</td>
<td>0.04</td>
<td>55</td>
</tr>
<tr>
<td>SA [34]</td>
<td>0.87</td>
<td>0.04</td>
<td>210</td>
</tr>
<tr>
<td>AFSA [35]</td>
<td>0.81</td>
<td>0.02</td>
<td>376</td>
</tr>
<tr>
<td>DQN</td>
<td>0.94</td>
<td>0.01</td>
<td>1180</td>
</tr>
</tbody>
</table>

Fig. 4  Thickness optimization by GA. a) Encoding of thickness of optical thin-film and mutation operation. b) The crossover operation of encoded thickness of optical thin-film

methods, as shown in Table 3, the genetic algorithm exhibited the best stability and best film performance in 50 attempts.

In Fig. 4 (a), we implement the genetic algorithm to optimize and improve the spectral performance of the layered optical thin-film with a population size of 100 and 500 generations. In the genetic thickness optimization algorithm, the chromosome represents the layer thickness of the thin-film. The variables of thickness are constructed from values that a boundary value (10nm – 200nm) is assigned. Another four key parameters of GA are the selection rate (0.3), mutation rate (0.1), crossover rate (0.5) and the elitist selection rate (0.1), to improve the solution of the GA. For the crossover operation in Fig. 4 (b), a random number is generated to determine the crossover position. For the mutation operation, the thickness is a random number between the boundary value. The elitist selection operation selects the best 10% chromosomes for the next generation in an iterative process, which is a slight variant of the general process of constructing a new population.

3.7 A3C

A3C is a parallel policy gradient algorithm in reinforcement learning to learn a policy $\pi(a_t | s_t; \theta)$ and to estimate the value function $V(s; \theta_v)$ by agents. The policy and value function is updated by a mix of n-step returns (states and actions) when a terminal state is reached. The update that is performed by the algorithm can be seen as $\nabla_{\theta} \log \pi(a_t | s_t; \theta') A(s_t, a_t; \theta, \theta_v)$, where $A(s_t, a_t; \theta, \theta_v)$ is an estimate of the advantage function. The expression is given by:

$$\gamma^k \sum_{i=0}^{k-1} \gamma^i r_{t+i} + \gamma^k V(s_{t+k}; \theta_v) - V(s_t; \theta_v).$$

The value function is learned by the critics in A3C, while multi-actors are updated by the parameters in the master model. These updates are performed in parallel. After several episodes, the agents get synced with the master model.
and the parameters are updated. The gradients are accumulated as part of training for stability like the parallelized stochastic gradient descent in the deep learning training process. In this parallel architecture, each agent in the same episode is sample from different experiences. Such a mechanism can have a better chance of improving the optimization results for combinatorial optimization problems such as material selection problem.

In our implementation, the actor-network consists of 4 layers fully connected neural network with (5, 32, 16, 1) units. The critic-network consists of 5 layers fully connected neural network, with (5, 32, 16, 16, 1) units. The activation function used in these two networks is the Relu. The Adam gradient descent for optimizing the networks.

4. Experiments and Results

4.1 Optimization of Solar Absorber Device

To demonstrate this material search system, we analyze a solar absorber device that can convert solar energy to thermal energy. This device uses multilayered thin-film structures consisting of alternating metal/alloys/semiconductor and dielectric layers. These layers have the advantage of excellent spectral properties in both broad solar spectral and wide incident angle regions, low thermal emittance, and high thermal stability.

We have performed a complete study of 4/6/8-layer SSR thin-film devices in our previous research [36]–[39]. The SSR with 5-layer structures have not been extensively studied. Figure 6(a) shows that solar energy is highest near the visible wavelengths. At an angle of normal light, the goal is 100% absorption at the wavelength range from 250 nm to 800 nm and no absorption at other wavelengths. We aim to achieve this performance with our studies.

From our search for 1000 epochs, we plot in Fig. 6(b) the best and cumulative absorption in each epoch. The agents develop the best material structure for a 5-layer structure with the material composition the thickness shown in Fig. 6(d). The algorithm selects materials in the following order [MgF$_2$, TiO$_2$, Si, Ge, Cu], respectively, thickness of each layer with the following order [35.3nm, 27.1nm, 112.5nm, 172.0nm, 200.0nm]. The designed absorber can achieve broadband absorption due to the characteristics of selected materials as shown in Fig. 6(c). The average absorption in the wavelength range, 250 nm to 800 nm, is above 91%. In contrast to other thin film optimization algorithms that can only handle thickness optimization problems, our algorithm selects effective constituent materials from a huge range of materials.

4.2 Method Comparison

On the solar absorber design task, we conducted an comparison study to compare our proposal with other method. We trained another two different models: 1) random search (RS): the action is totally by random agent, 2) deep Q learning (DQN): the action is selected from Q-value from DQN agent by $\epsilon$-policy. For each model, the search process is 1000 epoches. The maximum absorption values discovered by each model are reported in Fig. 7.

In Fig. 7, we plot the absorption and maximum absorption of the structures generated in each epoch over the entire searching trajectory. The result of A3C (91%) is more significant than other reinforcement learning agent as the random search (78%) and DQN (82.4%). We compare the searching trajectory of three models: 1) A3C find the best result around 200 epoch, 2) random search find the best result around 400 epoch, 3) DQN find the best result around 900 epoch. The A3C significantly achieves the best performance in our reinforcement learning environment.
5. Conclusion and Future Research

To conclude, this paper has created an RL based on A3C and GA to optimize the design of a multi-layer solar absorber. Using the VAE-tsNE method with more than 300 different materials, to reduce the continuous material optical parameters to the two-dimensional reinforcement learning environment space. This technique selected the most appropriate one by joint material search system. The material search system can automatically search material composition and structure to achieve a target optical spectrum based on A3C and GA. Compared with previous work, our approach can handle and select material from a significantly larger dataset without any human intervention. The VAE-tsNE model takes the optical material constant as the training samples and can generate the 2D features with material related. Optical films consist of materials whose features are distributed in the adjacent field also tend to have similar performance. Using the generated 2D material features, we propose a reinforcement learning environment can select materials and optimal thickness for each layer of a multi-layer structure. By comparing RS, DQN and A3C, we developed the A3C agent to achieve the best performance in solving our optical thin-film design problem. As a demonstration, we have used it to design a solar selective absorber. By using a diverse set of materials, the multi-layer structure has excellent solar absorption of 91% in visible light. Because of its versatility and effectiveness, our material search system proves to be a state of the art tool for multi-layer optical thin-film design. It can be not only used for thermal and energy applications, but also for other optical device designs and optimization.

References


als and alloys are well represented in the feature map, but important information by top-2 principal components. Met-shows extracted feature by PCA. We keep the most im-

In this paper, we compared multiple methods for 2d feature representation-feature reduction) VAE-tSNE approach.

In our implementation, shown in app.Fig. A-1, c, plasmonic matels (Ag, Au, Ag-Au) distributed in close field. The strong absorbing metals (Mo, Ti, Zn) are indicated in the upper right corner. Interestingly, insulators (visible transparent materials) and semiconductor materials are completely distinguished in different regions. Among them, the visible transparent material is distributed near the red dots, while the green area is dominated by the semiconductor material. Most importantly, semiconductor materials with different optical properties are well differentiated. Wide gap semiconductors (AlN) and compound semiconductors (GaAs, InAs) are significantly distinguished. Intuitively, the features generated by VAE-tSNE are more uniformly distributed throughout the feature space.

Appendix: Environment features comparison

In this paper, we compared multiple methods for 2d feature generation. In app.Fig. A-1, we choose three feature reduction methods (PCA, VAE, VAE-tSNE). The app.Fig. A-1.a shows extracted feature by PCA. We keep the most important information by top-2 principal components. Metals and alloys are well represented in the feature map, but the vast majority of the semiconductor materials are distributed in very close area. Further, we trained a VAE with 2 hidden units, Similarly with PCA result, the mutual positions of metals and alloys are related to their properties and some of the semiconductors correctly represent the relative relationships associated with the optical properties. To solve the problem that semiconductor materials cannot be well distinguished, we propose feature extraction by (feature representation-feature reduction) VAE-tSNE approach.

Anqing Jiang is pursuing the postgraduate degree in Waseda University, Japan. His research interests include neural combinational optimization and photoelectronic device design.

Osamu Yoshie is a professor at Waseda University and the director of IGSIAF, Institute for Global Strategies on Industry-Academia Fusion, at Waseda University, Japan. His research interest came from the experience of computer vision and augmented reality at FX Palo Alto Laboratory, USA and GMD (currently Fraunhofer), Germany and now gets extended to wide application of artificial intelligence, industrial IoT, community analysis and synthesis. He is a recipient of many awards from professional research societies, such as the Society of Instrument and Control Engineers (SICE), the Institute of Electrical Engineers (IEE) Japan, the Society of Plant Engineers Japan, Japan Industrial Management Association, etc. He is a fellow of IEE Japan.