

# Migrating Knowledge between Physical Scenarios Based on Artificial Neural Networks

Yurui Qu,<sup>\*,†,‡,§,||</sup> Li Jing,<sup>\*,†</sup> Yichen Shen,<sup>†</sup> Min Qiu,<sup>‡,§,||</sup> and Marin Soljačić<sup>†</sup>

<sup>†</sup>Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States

<sup>‡</sup>State Key Laboratory of Modern Optical Instrumentation, College of Optical Science and Engineering, Zhejiang University, Hangzhou 310027, China

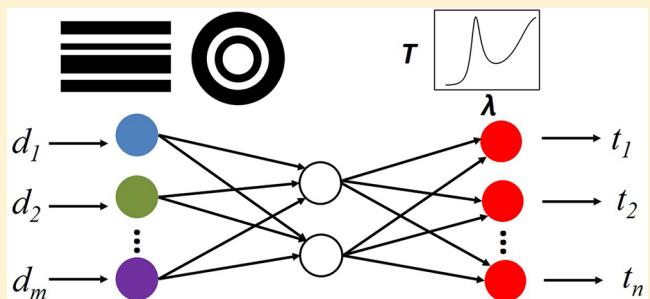
<sup>§</sup>School of Engineering, Westlake University, 18 Shilongshan Road, Hangzhou 310024, China

<sup>||</sup>Institute of Advanced Technology, Westlake Institute for Advanced Study, Westlake University, 18 Shilongshan Road, Hangzhou 310024, China

## Supporting Information

**ABSTRACT:** Deep learning is known to be data-hungry, which hinders its application in many areas of science when data sets are small. Here, we propose to use transfer learning methods to migrate knowledge between different physical scenarios and significantly improve the prediction accuracy of artificial neural networks trained on a small data set. This method can help reduce the demand for expensive data by making use of additional inexpensive data. First, we demonstrate that, in predicting the transmission from multilayer photonic film, the relative error rate is reduced by 50.5% (23.7%) when the source data comes from 10-layer (8-layer) films and the target data comes from 8-layer (10-layer) films. Second, we show that the relative error rate is decreased by 19.7% when knowledge is transferred between two very different physical scenarios: transmission from multilayer films and scattering from multilayer nanoparticles. Next, we propose a multitask learning method to improve the performance of different physical scenarios simultaneously in which each task only has a small data set. Finally, we demonstrate that the transfer learning framework truly discovers the common underlying physical rules instead of just performing a certain way of regularization.

**KEYWORDS:** *artificial neural networks, deep learning, transfer learning, physical scenarios, multilayer films, nanoparticles*



Deep learning is a powerful machine learning algorithm that discovers representations of data with multiple levels of abstraction based on multiple processing layers.<sup>1</sup> Recently, deep learning has received an explosion of interest because it continuously pushes the limit of traditional image recognition, machine translation, and decision-making, as well as many other applications.<sup>2–5</sup> Meanwhile, deep learning is also penetrating into other disciplines such as drug design,<sup>6,7</sup> genetics,<sup>8,9</sup> material science,<sup>10</sup> and physics, including classification of complex phases of matter,<sup>11,12</sup> electromagnetic inverse problem,<sup>13,14</sup> nanostructure design,<sup>15,16</sup> high-energy physics,<sup>17</sup> and quantum physics.<sup>18–20</sup> One drawback is that deep learning is a data-hungry method and can only work well if fed with massive data. However, collecting a large amount of data is slow and expensive for many numerical simulations, such as bands of three-dimensional (3D) photonic crystals,<sup>21</sup> and even much more difficult for experiments, since it might require fabricating tens of thousands of samples<sup>22</sup> or doing tens of thousands of measurements.<sup>23</sup> Similar situations are also common in other scientific areas in which collecting a large amount of simulated or experimental data is difficult. Direct learning from a small data set results in under-represented

features, which leads to poor performance. There has yet to emerge a solution to improve deep learning performance for scientific problems with a small data set.

Transfer learning has attracted growing interest in recent years because it can significantly improve the performance in the target task through the transfer of knowledge from the source task that has already been learned.<sup>24–27</sup> Transfer learning is a useful method when the source data set is large and inexpensive and the target data set is small and expensive. Jason Yosinski et al. demonstrated on ImageNet that transferred layers can improve classification accuracy by 2% on a new task after substantial fine-tuning.<sup>28</sup> Andrei A. Rusu et al. showed that learning from a simulation and transferring the knowledge to a real-world robot can solve the problem that training models on a real robot is too slow and expensive.<sup>29</sup> However, unlike classic transfer learning that only cares about doing well on one particular target task or domain, another method, called multitask learning, can do well on all related tasks that are trained simultaneously and benefit from each

**Received:** November 4, 2018

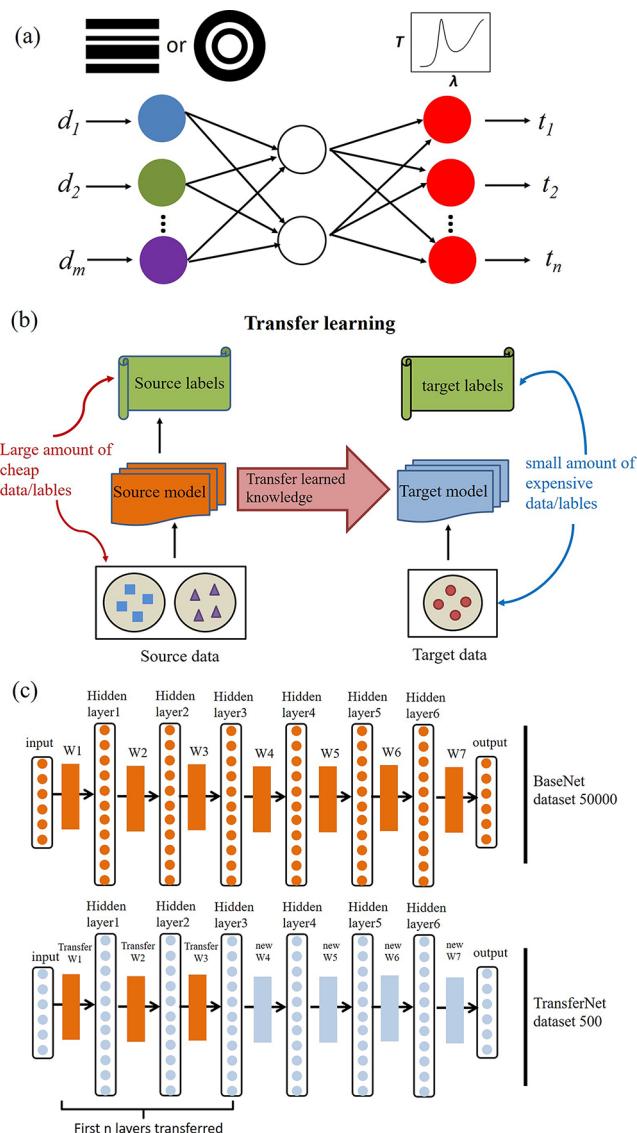
**Published:** May 2, 2019

other. Multitask learning has been used successfully across many applications such as natural language processing,<sup>30</sup> speech recognition,<sup>31,32</sup> and computer vision.<sup>33</sup>

In this Article, we propose a deep neural network architecture with transfer learning ability that can significantly improve the performance of physical problems even if their data sets are small. There are two types of physical problems: one is the forward prediction problem shown in this work (the goal is to predict some physical properties given a specific physical system) and the other is the inverse design problem (the goal is to design a specific physical system based on some desired physical properties, such as the inverse design of a thin film device with specific reflection spectrum proposed by Dianjing et al.<sup>14</sup>). In this paper, we focus on the forward prediction problems like transmission from multilayer films and scattering from nanoparticles. Although we focus here on certain particular photonic problems, the approach proposed can easily be generalized to many other scientific problems. The deep neural network with transfer learning is investigated in several cases: (1) The source and target data come from similar physical problems, for example, transmission from multilayer films with different numbers of geometric layers. The spectrum error continuously decreases as more layers of the neural network are transferred. Note that the former “layers” are used for photonic structures and the latter “layers” are used for weights and biases in neural networks. The relative error reduction is 50.5% (23.7%) when the source data comes from 10-layer (8-layer) films and the target data comes from 8-layer (10-layer) films. (2) The source and target data come from very different physical problems. The source data are scattering cross sections from 8-layer core–shell nanoparticles and the target data are transmissivities from 8-layer films. To our surprise, the relative error rate still decreases by 19.7% after transferring knowledge from the nanoparticle scattering problem to the multilayer film transmission problem. (3) Multiple tasks are 8-layer, 10-layer, 12-layer, and 14-layer films. The performance of multitask learning outperforms the direct learning trained only on a specific group of data. The neural network with transfer learning can significantly improve the performance of neural networks with only a small data set, which could benefit the applications of deep learning in many physical problems and other fields.

In classic deep learning, a model can be well trained for some task and domain if sufficient labeled data are provided. Let us assume that a task is the objective that our model aims to perform: for example, predict the transmission spectra of 10-layer films, as shown in Figure 1a. We can now train a model on such a data set and expect it to perform well on unseen data of a 10-layer film. However, this classic deep learning breaks down when we do not have sufficient labeled data for this task or domain. Training on a small data set will cause collapse in performance because of a severe overfitting problem. Transfer learning allows us to deal with this problem by leveraging the existing labeled data from some related task or domain, for example, transmissivities from 8-layer films, or even a very different task like scattering cross sections from core–shell nanoparticles. We try to store the knowledge gained in solving the source task in the source model and apply it to the target model to help the target task (see Figure 1b).

An artificial neural network structure of our transfer learning method is shown in Figure 1c. The input data are the thicknesses of each film or shell (the materials were fixed), and the output data are the transmissivities sampled at points



**Figure 1.** (a) Illustration of the neural network architecture. The input is the thickness of each film or each shell of the nanoparticle, and the output is the transmission or scattering spectrum. Only one hidden layer is drawn here for convenience. Our actual neural network has six hidden layers. (b) Learning process of transfer learning. Transfer learning techniques can transfer the learned knowledge from the source model to the target model to improve the performance of the target task. Generally, the source domain has a large amount of inexpensive data while the target domain only has a small amount of expensive data. (c) Neural network structure of transfer learning. Top row: The base network (BaseNet) learns from scratch on a large source data set of 50000 examples. Bottom row: The transfer network copies the first  $n$  layers from the BaseNet as the initialization weights and biases, and then the entire network is trained (fine-tuned) on the small target data set of 500 examples.

between 400 and 800 nm. In order to further explore these and many other potential applications, we have publicly released all the data used in this paper.<sup>34</sup> For a different number of films as input, we use a one-dimensional mask (unnecessary position set as 0) to stretch the input into the same length. The thicknesses are between 30 and 70 nm, and the materials are SiO<sub>2</sub> and TiO<sub>2</sub> for alternating layers of multilayer films and of core–shell nanoparticles. We train a fully connected neural network called BaseNet on the source domain with a data set

of 50000 examples. A total of 80% of the data are used for training the network and the other 20% are the validation data and the test data (10% each), which are the same for the target task. More details about the neural network architecture have been included in [Supporting Information](#).

TransferNet has the same network structure as BaseNet, while TransferNet copies the first  $n$  layers from the BaseNet as the initialization of weights and biases of the first  $n$  layers. The remaining higher layers of TransferNet are also initialized randomly with a normal distribution, and the entire TransferNet is fine-tuned simultaneously. TransferNet is trained on a target domain with a small data set of 500 examples. The spectrum error decreases as the amount of training data increases, as shown in [Figure S1](#). There is a breaking point at around 1000 examples. When the amount of data is less than 1000, the error increases and the performance deteriorates sharply. We choose 500 examples in our training process to make sure it is truly a small amount of data. Next, we compare the transfer learning with the direct learning on the same data set to demonstrate that transfer learning can truly improve the performance.

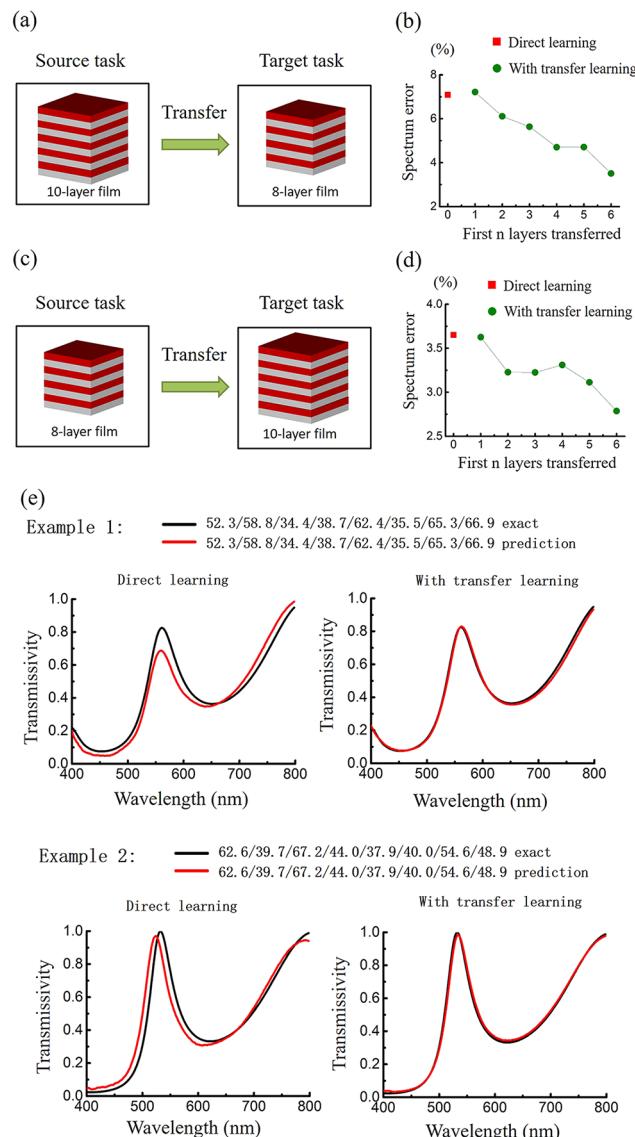
The most general method of calculating the transmittance of a multilayer film is based on a matrix formula<sup>35</sup> of the boundary conditions at the film surfaces derived from Maxwell's equations (see [Methods in the Supporting Information](#)). We use a neural network with direct learning and also with transfer learning, respectively, to approximate this transfer matrix formula, and we compare the performance in these two cases. We first explore the transfer learning between 8-layer films and 10-layer films and try to transfer knowledge in both directions, as shown in [Figure 2a,c](#). The spectrum error in this paper is defined as the average difference between the prediction and the exact result per spectrum point:

$$\text{Error} = \frac{1}{n} \sum_{i=1}^n \frac{|T_{\text{prediction}}(\lambda_i) - T_{\text{exact}}(\lambda_i)|}{T_{\text{exact}}(\lambda_i)} \quad (1)$$

where  $n$  is the number of the spectrum points. In our case, we sampled between 400 and 800 nm with a 2 nm step, so  $n = 200$  (not including 800 nm).

For the target task of 8-layer films, direct learning has a test error of around 7.1% ([Figure 2b](#)). After transferring the first and second layers from trained BaseNet and retraining all the layers in TransferNet together, the spectrum error is reduced to 6.1%. With more layers of the BaseNet transferred, the spectra errors continuously decrease. When six layers of BaseNet are transferred, the spectrum error decreases to about 3.5%, which is 50.5% relative reduction compared to direct learning. Transfer learning also works well when knowledge is transferred from an 8-layer film to a 10-layer film ([Figure 2c](#)). Direct learning for 10-layer films has the spectrum error of about 3.7%. With six layers of BaseNet transferred, the spectrum error is reduced to around 2.8%, which is 23.7% relative error reduction (see [Figure 2d](#)).

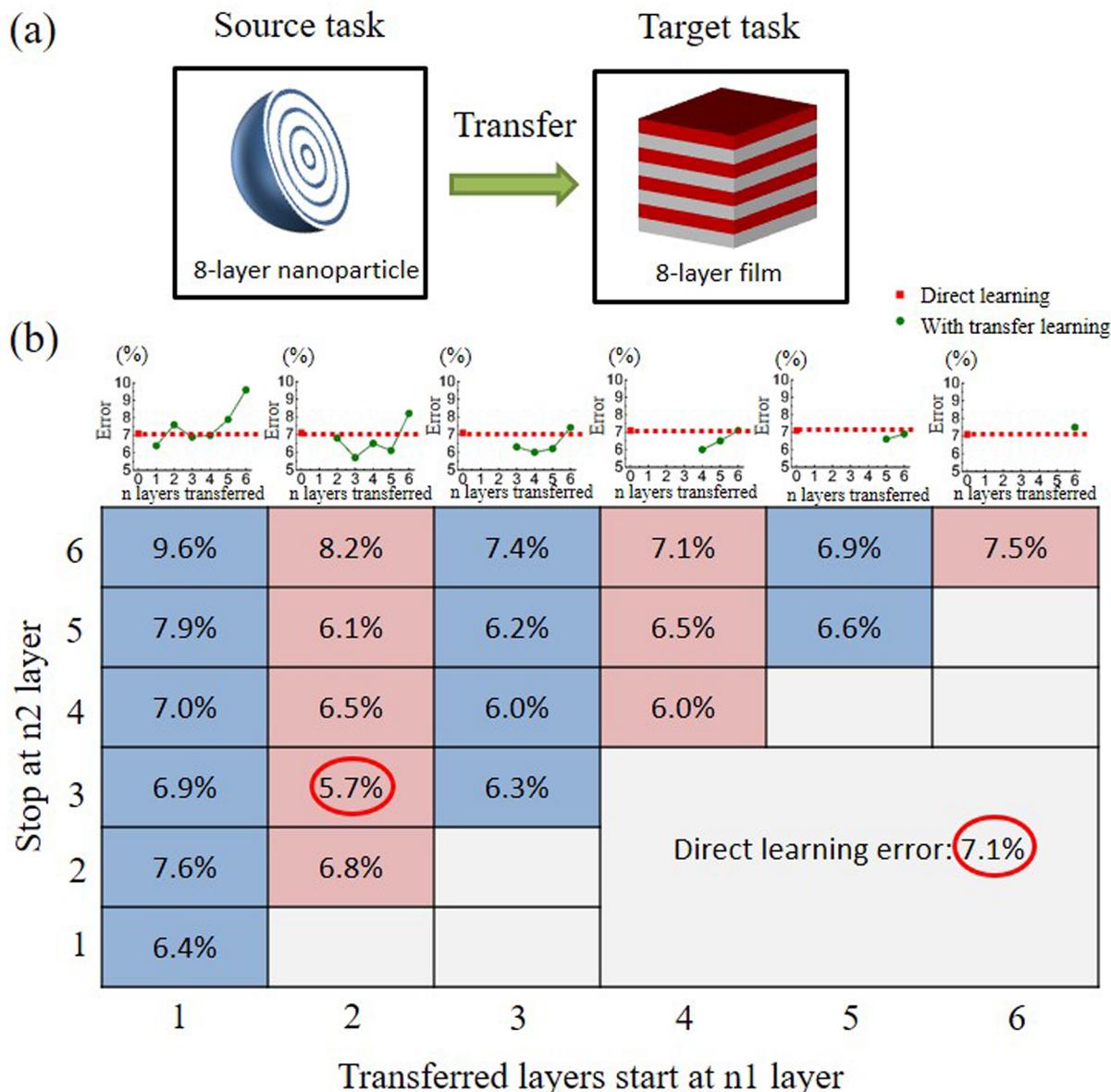
Two examples are presented in [Figure 2e](#) to demonstrate that transfer learning can give a better prediction of the transmission spectrum compared to direct learning. Two examples come from the case that the source domain is a 10-layer film and the target domain is an 8-layer film. Black lines and red lines are theoretical and predicted transmission spectra, respectively. For the first example, the predicted spectrum using direct learning has a lower peak transmissivity



**Figure 2.** (a) Illustration of the transfer learning process and (b) test errors for the first  $n$  layers of BaseNet transferred when the source domain are 10-layer films and the target domain are 8-layer films. (c, d) The case that the source domain are 8-layer films and the target domain are 10-layer films. (e) Two examples of transmission spectra for the case that the source domain is an 8-layer film and the target domain is a 10-layer film. Exact spectra are black lines and predicted spectra are red lines. Comparison of the direct learning to the transfer learning demonstrates that transfer learning can predict more accurate spectra than direct learning.

than the exact spectrum, and differences at other wavelengths are also obvious. For the second example, the entire predicted transmission spectrum based on direct learning shifts to a shorter wavelength compared to the exact spectrum. However, the predicted spectra using transfer learning for both cases are much more accurate. This result is surprising because the spectra are predicted by the neural network, which has only seen 400 training examples.

Next, we try to transfer knowledge between two very different tasks, the scattering from multilayer nanoparticles and the transmission from multilayer films, as shown in [Figure 3a](#). A scattering cross-section from a multilayer nanoparticle can be calculated using a transfer matrix method, but in the form of



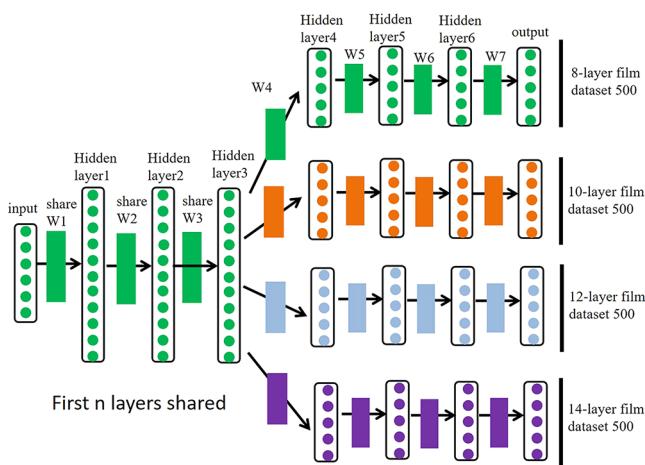
**Figure 3.** (a) Illustration of the transfer learning process when the source domain are 8-layer nanoparticles and the target domain are 8-layer films. (b) Test errors when transferred layers begin at the n<sub>1</sub> layer and stop at the n<sub>2</sub> layer of trained BaseNet. The insets are spectra errors for each column.

Bessel functions<sup>36</sup> (also see [Methods in the Supporting Information](#)). More similar structures with different materials have been studied.<sup>37,38</sup> In [Figure 3b](#), the first column of the table represents the spectrum error, with the first  $n$  layers of the BaseNet transferred. The error of transfer learning is 0.7% lower than direct learning (red dashed line) when only the first layer of the BaseNet is transferred before training the TransferNet. After transferring the first and second layers of the BaseNet, the error of transfer learning increases instead and surpasses the red dashed line, which is a negative transfer that is harmful for learning the target task. Transferring the first 3 or 4 layers of the BaseNet can help to reduce the spectrum error a little lower than the direct learning. However, if the first 5 or 6 layers of the BaseNet are transferred together, the final performance will deteriorate sharply. From the results we can tell that some layers transferred from the BaseNet are specific to the nanoparticle scattering problem. These layers of the

BaseNet will not help the target task learning at all, and can even be harmful for the final performance. Other layers of the BaseNet which are apparently general to both problems can be transferred and improve the target performance.

To find the layers of the BaseNet that contain the transferable physical knowledge, we utilize a grid search to study the performance of transferring layers starting from n<sub>1</sub> ending at n<sub>2</sub>, as shown in [Figure 3b](#). The test error decreases to 5.7% when 2<sup>nd</sup> and 3<sup>rd</sup> layers of the BaseNet are transferred, with around 19.7% relative error reduction compared to direct learning. The results are enlightening. Even for the case when there is not enough available data from a similar task, we can still utilize data from a different task, which largely expands the areas where this transfer learning method can be applied. Through this process, we are able to isolate the shared physical knowledge learned by the neural networks to some extent while keeping out the scenario-specific knowledge.

The transfer learning method described above requires a large amount of inexpensive data from related tasks. However, in some cases, there are several related tasks, but each of them has only a small amount of data. The classic transfer learning method cannot work well in this case. Here, we introduce another knowledge transferring method called multitask learning. As shown in Figures 4 and S2, multitask learning



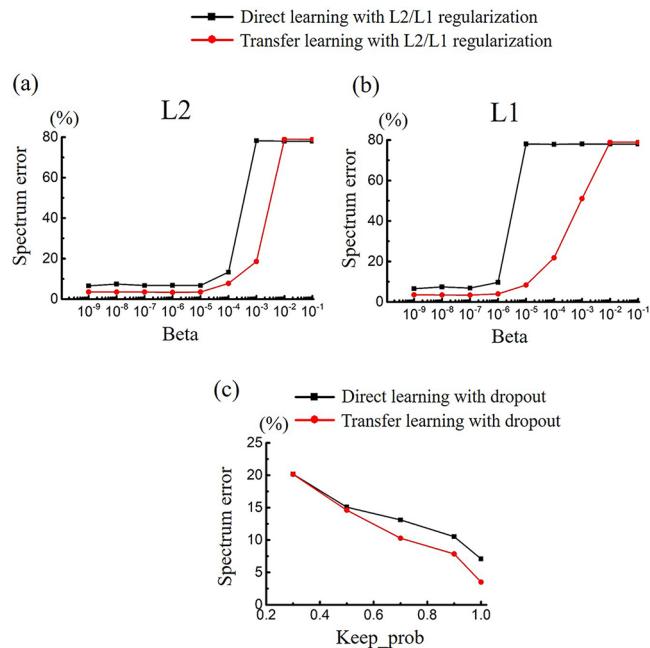
**Figure 4.** Neural network structure of multitask learning. The network shares the first  $n$  hidden layers and splits for the rest. Four target domains are 8-layer, 10-layer, 12-layer, and 14-layer films, and each has a small data set of 500 examples.

shares the first  $n$  hidden layers among all physical scenarios, while keeping several task-specific output layers. Here, four target data are from 8-layer, 10-layer, 12-layer, and 14-layer films, and each has a small data set of 500 examples. Each task can benefit from the knowledge learned in the other tasks, and the performance of each task can be improved compared to individual direct learning. This makes sense intuitively: the more tasks we are learning simultaneously, the more our model has to find a representation that captures all of the tasks and the less is our chance of overfitting the our original task. The key to this successful learning is to train the model for all the tasks simultaneously. All data are fed in each update of the model. The training algorithm needs to be adjusted slightly from the conventional backpropagation algorithm because of the split task-specific layers. When a training example is from an 8-layer film, only the shared layers and the specific layers belonging to the 8-layer film task are updated. Other task-specific layers are kept intact. The same operation is done to train all four tasks.

We compare the spectrum errors of direct learning to that of multitask learning in Table 1. Even if we only use four tasks learned together, each with a small data set of 500 examples, multitask learning has lower spectrum error than direct learning in each of four tasks. The relative reduction of the spectrum error is 9.9%, 16.2%, 1.7%, and 27.1% for 8-layer, 10-

layer, 12-layer, and 14-layer films, respectively. We expect that the performance can be better with more target tasks. The best neural network structure is different for each target task, as shown in Figure S3. For 8-layer and 10-layer films, the best performances are achieved when the first two hidden layers are shared. For 12-layer and 14-layer films, however, the best performances are achieved when the first three hidden layers are shared. We can also see that the performance deteriorates sharply if too many layers are shared. The reason is that the last several layers are specific for each task and cannot transfer knowledge among different tasks.

To demonstrate that the improved performance comes from transfer learning, not just from a certain way of regularization, we add L2 regularization, L1 regularization, and dropout, several most widely used regularization methods, to the neural network of the direct learning and the transfer learning. The L2/L1 regularization can help the direct learning to improve a little, from 7.1% error rate (without L2/L1 regularization) to 6.6% (with L2/L1 regularization). However, the transfer learning with L2/L1 regularization always performs better than the direct learning with L2/L1 regularization, as shown in Figure 5a,b. The lowest spectrum error of the direct learning



**Figure 5.** Spectrum errors for direct learning (black line) and transfer learning (red line) after adding (a) L2 regularization, (b) L1 regularization, and (c) dropout to the neural network. The source domain are 10-layer films and the target domain are 8-layer films.

with L2/L1 regularization is 6.6%; however, that of the transfer learning is 3.3% with L2 regularization and 3.4% with L1 regularization. This demonstrates that transfer leaning cannot be simply replaced by L2/L1 regularization.

Next, we add dropout to each hidden layer of the neural network. We control different keep\_prob to study the effect of the dropout on the neural network. When the keep\_prob equals 1.0, the case is the same as no dropout. As shown in Figure 5c, dropout does not help to improve the performance of both direct learning and transfer learning, but makes it worse. These results demonstrate that the transfer learning

**Table 1. Comparison of Direct Learning Error to Multitask Learning Error**

training data set	8 layers	10 layers	12 layers	14 layers
direct learning error	7.1%	3.7%	6.0%	12.9%
multitask learning error	6.4%	3.1%	5.9%	9.4%
relative error reduction	9.9%	16.2%	1.7%	27.1%

framework truly discovers the common underlying physical rules instead of just performing a certain way of regularization.

In conclusion, we present two transfer learning methods to help with the fact that deep learning methods cannot work well with small data sets in physical scenarios. We demonstrate that the neural network with transfer learning can give more accurate prediction compared to direct learning when trained on the same data set. The knowledge in the neural network can be transferred not only between similar physical scenarios, such as transmission from multilayer films with different number of films, but also between very different physical scenarios like scattering from core–shell nanoparticles and transmission from multilayer films. Multitask learning, on the other hand, can improve the performance of several related tasks simultaneously even if each task only has a small data set of 500 examples. The challenge of this transfer learning method is how to avoid negative transfer between two different tasks. Here we systematically select the general layers and specific layers in the neural network using grid search method, and it would be important to investigate if this process can be done automatically in the future. Looking forward, neural networks with transfer learning could not only benefit the development of deep learning in many physical problems of which data sets are expensive and small, but also in other areas of science such as biology, chemistry, and material science.

## ■ ASSOCIATED CONTENT

### S Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: [10.1021/acsphtnics.8b01526](https://doi.org/10.1021/acsphtnics.8b01526).

Methods of analytically solving transmission from a multilayer film and scattering from a multilayer nanosphere via the transfer matrix method; Technical details about the neural network architecture; Spectrum errors vs the amount of data; Learning process of multitask learning; Spectrum errors for multitask learning; The training and validation errors over epochs; Spectra errors when the target task is 10-, 16-, and 24-layer films ([PDF](#))

## ■ AUTHOR INFORMATION

### Corresponding Authors

\*E-mail: [yuruiqu@zju.edu.cn](mailto:yuruiqu@zju.edu.cn).

\*E-mail: [ljing@mit.edu](mailto:ljing@mit.edu).

### ORCID

Yurui Qu: [0000-0002-8837-2663](https://orcid.org/0000-0002-8837-2663)

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

We acknowledge discussions with John Peurifoy. Research was sponsored in part by the Army Research Office and was accomplished under Cooperative Agreement Number W911NF-18-2-0048. This material is based on work supported in part by the National Science Foundation under Grant No. CCF-1640012. This material is based on work supported in part by the Semiconductor Research Corporation under Grant No. 2016-EP-2693-B. This material is also based in part upon work supported by the Defense Advanced Research Projects Agency (DARPA) under Agreement No. HR00111890042, as well as in part by the MIT-SenseTime Alliance on Artificial

Intelligence. This project was also supported by the National Key Research and Development Program of China (Grant No. 2017YFA0205700) and the National Natural Science Foundation of China (Grant No. 61425023). Y.Q. was supported by Chinese Scholarship Council (CSC No. 201706320254).

## ■ REFERENCES

- (1) LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436.
- (2) Hinton, G.; Deng, L.; Yu, D.; Dahl, G. E.; Mohamed, A.-r.; Jaitly, N.; Senior, A.; Vanhoucke, V.; Nguyen, P.; Sainath, T.; Kingsbury, B. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal processing magazine* **2012**, *29*, 82–97.
- (3) Krizhevsky, A.; Sutskever, I.; Hinton, G. E. Imagenet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90.
- (4) Mnih, V.; Kavukcuoglu, K.; Silver, D.; Rusu, A. A.; Veness, J.; Bellemare, M. G.; Graves, A.; Riedmiller, M.; Fidjeland, A. K.; Ostrovski, G.; Petersen, S.; Beattie, C.; Sadik, A.; Antonoglou, I.; King, H.; Kumaran, D.; Wierstra, D.; Legg, S.; Hassabis, D. Human-level control through deep reinforcement learning. *Nature* **2015**, *518*, 529.
- (5) Silver, D.; Huang, A.; Maddison, C. J.; Guez, A.; Sifre, L.; Van Den Driessche, G.; Schrittwieser, J.; Antonoglou, I.; Panneershelvam, V.; Lanctot, M.; Dieleman, S.; Grewe, D.; Nham, J.; Kalchbrenner, N.; Sutskever, I.; Lillicrap, T.; Leach, M.; Kavukcuoglu, K.; Graepel, T.; Hassabis, D. Mastering the game of Go with deep neural networks and tree search. *Nature* **2016**, *529*, 484.
- (6) Ma, J.; Sheridan, R. P.; Liaw, A.; Dahl, G. E.; Svetnik, V. Deep neural nets as a method for quantitative structure–activity relationships. *J. Chem. Inf. Model.* **2015**, *55*, 263–274.
- (7) Gawehn, E.; Hiss, J. A.; Schneider, G. Deep learning in drug discovery. *Mol. Inf.* **2016**, *35*, 3–14.
- (8) Leung, M. K.; Xiong, H. Y.; Lee, L. J.; Frey, B. J. Deep learning of the tissue-regulated splicing code. *Bioinformatics* **2014**, *30*, i121–i129.
- (9) Xiong, H. Y.; Alipanahi, B.; Lee, L. J.; Bretschneider, H.; Merico, D.; Yuen, R. K.; Hua, Y.; Gueroussov, S.; Najafabadi, H. S.; Hughes, T. R.; Morr, Q.; Scherer, S. W.; Blencowe, B. J.; Frey, B. J.; et al. The human splicing code reveals new insights into the genetic determinants of disease. *Science* **2015**, *347*, 1254806.
- (10) Ramprasad, R.; Batra, R.; Pilania, G.; Mannodi-Kanakkithodi, A.; Kim, C. Machine learning in materials informatics: recent applications and prospects. *npj Computational Materials* **2017**, *3*, 54.
- (11) Schoenholz, S. S.; Cubuk, E. D.; Sussman, D. M.; Kaxiras, E.; Liu, A. J. A structural approach to relaxation in glassy liquids. *Nat. Phys.* **2016**, *12*, 469.
- (12) Wang, L. Discovering phase transitions with unsupervised learning. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2016**, *94*, 195105.
- (13) Peurifoy, J.; Shen, Y.; Jing, L.; Yang, Y.; Cano-Renteria, F.; DeLacy, B. G.; Joannopoulos, J. D.; Tegmark, M.; Soljačić, M. Nanophotonic particle simulation and inverse design using artificial neural networks. *Science Advances* **2018**, *4*, No. eaar4206.
- (14) Liu, D.; Tan, Y.; Khoram, E.; Yu, Z. Training deep neural networks for the inverse design of nanophotonic structures. *ACS Photonics* **2018**, *5*, 1365–1369.
- (15) Malkiel, I.; Nagler, A.; Mrejen, M.; Arieli, U.; Wolf, L.; Suchowski, H. Deep learning for design and retrieval of nanophotonic structures. *arXiv preprint arXiv:1702.07949* **2018**, na.
- (16) Ma, W.; Cheng, F.; Liu, Y. Deep-Learning Enabled On-Demand Design of Chiral Metamaterials. *ACS Nano* **2018**, *12*, 6326.
- (17) Baldi, P.; Sadowski, P.; Whiteson, D. Searching for exotic particles in high-energy physics with deep learning. *Nat. Commun.* **2014**, *5*, 4308.
- (18) Biamonte, J.; Wittek, P.; Pancotti, N.; Rebentrost, P.; Wiebe, N.; Lloyd, S. Quantum machine learning. *Nature* **2017**, *549*, 195.

- (19) Deng, D.-L.; Li, X.; DasSarma, S. Quantum entanglement in neural network states. *Phys. Rev. X* **2017**, *7*, 021021.
- (20) Carleo, G.; Troyer, M. Solving the quantum many-body problem with artificial neural networks. *Science* **2017**, *355*, 602–606.
- (21) Joannopoulos, J. D.; Johnson, S. G.; Winn, J. N.; Meade, R. D. *Photonic Crystals: Molding the Flow of Light*; Princeton University Press, 2011.
- (22) Noda, S.; Tomoda, K.; Yamamoto, N.; Chutinan, A. Full three-dimensional photonic bandgap crystals at near-infrared wavelengths. *Science* **2000**, *289*, 604–606.
- (23) Zahavy, T.; Dikopoltsev, A.; Moss, D.; Haham, G. I.; Cohen, O.; Mannor, S.; Segev, M. Deep learning reconstruction of ultrashort pulses. *Optica* **2018**, *5*, 666–673.
- (24) Pan, S. J.; Yang, Q. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering* **2010**, *22*, 1345–1359.
- (25) Bengio, Y. Deep learning of representations for unsupervised and transfer learning. *Proceedings of ICML Workshop on Unsupervised and Transfer Learning*; 2012; pp 17–36.
- (26) Thrun, S.; Pratt, L. *Learning to Learn*; Springer Science & Business Media, 2012.
- (27) Torrey, L.; Shavlik, J. *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques*; IGI Global, 2010; pp 242–264.
- (28) Yosinski, J.; Clune, J.; Bengio, Y.; Lipson, H. How transferable are features in deep neural networks? *Advances in neural information processing systems* **2014**, 3320–3328.
- (29) Rusu, A. A.; Vecerik, M.; Rothörl, T.; Heess, N.; Pascanu, R.; Hadsell, R. Sim-to-real robot learning from pixels with progressive nets. *arXiv preprint arXiv:1610.04286* **2016**, na.
- (30) Collobert, R.; Weston, J. A unified architecture for natural language processing: Deep neural networks with multitask learning. *Proceedings of the 25th International Conference on Machine Learning*; ACM, 2008; pp 160–167.
- (31) Huang, J.-T.; Li, J.; Yu, D.; Deng, L.; Gong, Y. Cross-language knowledge transfer using multilingual deep neural network with shared hidden layers. *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference*; IEEE, 2013; pp 7304–7308.
- (32) Deng, L.; Hinton, G.; Kingsbury, B. New types of deep neural network learning for speech recognition and related applications: An overview. *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference*; IEEE, 2013; pp 8599–8603.
- (33) Girshick, R. Fast r-cnn. *Proceedings of the IEEE international conference on computer vision*; IEEE, 2015; pp 1440–1448.
- (34) Qu, Y. Data of transmission from multilayer films and scattering from multilayer nanoparticles. [https://github.com/yuruiqu/Transfer-learning\\_Data](https://github.com/yuruiqu/Transfer-learning_Data).
- (35) Bass, M.; DeCusatis, C.; Li, G.; Mahajan, V.; Stryland, E. Optical Properties of Materials, Nonlinear Optics, Quantum Optics. *Handbook of Optics*; McGraw-Hill, 2010; Vol. IV.
- (36) Qiu, W.; DeLacy, B. G.; Johnson, S. G.; Joannopoulos, J. D.; Soljačić, M. Optimization of broadband optical response of multilayer nanospheres. *Opt. Express* **2012**, *20*, 18494–18504.
- (37) Abrashuly, A.; Valagianopoulos, C. Limits for Absorption and Scattering by Core-Shell Nanowires in the Visible Spectrum. *Phys. Rev. Appl.* **2019**, *11*, 014051.
- (38) Sheverdin, A.; Valagianopoulos, C. Core-shell nanospheres under visible light: Optimal absorption, scattering, and cloaking. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2019**, *99*, 075305.