



Article

Transfer Learning for Modeling Plasmonic Nanowire Waveguides

Aoning Luo, Yuanjia Feng, Chunyan Zhu, Yipei Wang * and Xiaoqin Wu *

Key Laboratory of Optoelectronic Technology and Systems (Ministry of Education), College of Optoelectronic Engineering, Chongqing University, Chongqing 400044, China

* Correspondence: wangyp@cqu.edu.cn (Y.W.); xiaoqinwu@cqu.edu.cn (X.W.)

Abstract: Retrieving waveguiding properties of plasmonic metal nanowires (MNWs) through numerical simulations is time- and computational-resource-consuming, especially for those with abrupt geometric features and broken symmetries. Deep learning provides an alternative approach but is challenging to use due to inadequate generalization performance and the requirement of large sets of training data. Here, we overcome these constraints by proposing a transfer learning approach for modeling MNWs under the guidance of physics. We show that the basic knowledge of plasmon modes can first be learned from free-standing circular MNWs with computationally inexpensive data, and then reused to significantly improve performance in predicting waveguiding properties of MNWs with various complex configurations, enabling much smaller errors (~23–61% reduction), less trainable parameters (~42% reduction), and smaller sets of training data (~50–80% reduction) than direct learning. Compared to numerical simulations, our model reduces the computational time by five orders of magnitude. Compared to other non-deep learning methods, such as the circular-area-equivalence approach and the diagonal-circle approximation, our approach enables not only much higher accuracies, but also more comprehensive characterizations, offering an effective and efficient framework to investigate MNWs that may greatly facilitate the design of polaritonic components and devices.

Keywords: deep learning; transfer learning; plasmonics; nanowires; waveguides



Citation: Luo, A.; Feng, Y.; Zhu, C.; Wang, Y.; Wu, X. Transfer Learning for Modeling Plasmonic Nanowire Waveguides. *Nanomaterials* **2022**, *12*, 3624. <https://doi.org/10.3390/nano12203624>

Academic Editor: Onofrio M. Marago

Received: 28 September 2022

Accepted: 14 October 2022

Published: 16 October 2022

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



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

1. Introduction

As important building blocks for next-generation nanophotonic components and devices, metal nanowires (MNW) are able to guide surface plasmon polaritons (SPPs) with a tight confinement in transverse cross-sections, providing a promising platform to manipulate light and light–matter interactions at the deep-subwavelength scale [1–3]. With the development of fabrication techniques, MNWs can be readily prepared via both top-down and bottom-up approaches, offering different geometric sizes and cross-sectional shapes (e.g., square [4] and pentagonal cross-sections [5] through electron-beam lithography and chemical synthesis, respectively) with intriguing waveguiding properties [6] in the visible and near-infrared regime for various applications, including all-optical light routing [7,8], ultrasensitive sensing [9,10], and plasmon lasing [11,12].

Retrieving waveguiding properties lies at the heart of the investigation of these MNWs, not only in uncovering relationships between structural variables and plasmonic responses, but also in offering important physical insights for experiments and design guidelines for waveguiding plasmonic components and devices. One common approach for modeling MNWs is to utilize FEM or FDTD methods, in which waveguiding properties of these MNWs with different geometry configurations can be numerically obtained. However, the sharp corners and edges in polygonal cross-sections are subject to numerical errors induced by insufficient mesh resolution [13–15]. To resolve the abrupt geometric features, one has to resort to an extremely fine mesh with an element size of down to one-to-few or sub-nanometers [14], resulting in extra complexity and a tremendous consumption of computational resources, which cannot meet the urgent demand of large-scale designs

in plasmonic circuits and devices. To address this issue, an alternative approach is to approximate the waveguiding properties of polygonal-cross-section MNWs (refer to polygonal MNWs) to circular-cross-section MNWs (refer to circular MNWs) by utilizing the diagonal-circle approximation (DCA) method [16]. In this case, solutions of circular MNWs can be analytically obtained from Maxwell's equations or numerically obtained with much fewer mesh elements. Despite its inexpensiveness in computational time and resources, such an approach has a relatively low accuracy, leading to large errors in the calculation results (e.g., a deviation of 20% for semiconductor NWs, not to mention MNWs). In fact, as a mixture of plasmons (collective charge oscillations) and optical modes (solutions to Maxwell's equations at a given geometry), SPPs in MNWs inherit both of their properties and are thus extremely sensitive to the sharp geometric features and the overall symmetry of the system because they greatly modify the surface charge density and influence the hybridization of the optical modes, giving rise to distinct waveguiding properties of polygonal MNWs compared to circular ones.

Besides the aforementioned analytical and numerical methods, deep learning has recently emerged as a powerful data-driven tool for retrieving the photonic/plasmonic properties of nanostructures [17–29]. However, it is also known to be data hungry, and the task is usually accomplished with a large set of data for training [30–33], resulting in several challenges. For example, the preparation of the dataset still requires a time- and computational-resource-consuming process [34,35], and the generalization of the trained model of nanostructures with new configurations often has a poor performance [36]. On the other hand, by migrating the learned knowledge from the source task to the related target tasks of a similar problem, transfer learning offers a possible approach to addressing the above challenges [37], and it is especially useful for situations where the target dataset size is limited. Although transfer learning has been successfully utilized in predicting the properties of nanostructures, such as layered nanoparticles [33,38] and metamaterials [39,40], their application in the study of waveguiding properties of MNWs is still an underexploited territory.

To overcome the limitations of conventional deep learnings (e.g., the requirement of large sets of data and an inadequate generalization performance), as well as the drawbacks of non-deep learning methods (e.g., large errors for the approximation approach, and high time and resource consumption for the numerical simulation), we propose a transfer learning approach for modeling MNWs, offering a high performance with a small dataset size and strong generalization capability for various configurations (e.g., MNWs with different cross-sections, working environments, geometric sizes, and wavelengths). The basic idea of this method is to utilize the knowledge acquired in solving the source task of free-standing circular MNWs with source data that is computationally cheap to improve the performance of target tasks for MNWs in complex configurations with small target dataset sizes. We show that, compared to direct learning, our model can achieve a significantly improved performance with much smaller errors (~23–61% relative error reduction) and less trainable parameters (~42% relative size reduction). Moreover, our approach removes the need for a large set of training data, reducing the number of data instances by ~50–80% compared with direct learning. In addition, compared to numerical simulations, our approach enables a much faster computational time, reduced by five orders of magnitude. Meanwhile, compared to other non-deep learning methods, such as the circular-area-equivalence (CAE) approach and the DCA, our approach can offer not only a much higher accuracy, but also a more comprehensive characterization of the waveguiding properties. Benefitting from the advantages of transfer learning, our model provides a simple, lightweight but effective approach to retrieving the plasmonic properties of MNWs with high accuracy, which is much needed in the investigation of plasmonic nano-waveguides. It can greatly accelerate the building of structure–property libraries for plasmonic architectures, revealing hidden relationships between structural variables and plasmonic properties that may open new opportunities to meet the increasing demand for large-scale designs of next-generation nanophotonic circuits and devices.

2. Materials and Methods

2.1. Emerging Requirement of Computational Resources for MNWs with Sharp Corners and Asymmetric Configuration

We first start by demonstrating the emerging requirements of computational resources for MNWs with sharp corners and asymmetric configurations. The numerical simulations were performed with COMSOL Multiphysics (version 6.0, COMSOL AB, Stockholm, Sweden). As a typical case, the silica-substrate-supported pentagonal MNW was selected for demonstration. It can be clearly seen from Figure 1a that the mode profile in the pentagonal MNW was very different from that of circular MNWs, resulting in distinct mode characteristics and waveguiding properties (e.g., pentagonal MNW (bound mode) vs. circular MNW (leaky mode)). Therefore, polygonal MNWs cannot be directly modelled by an approximation to their circular counterparts and must revert to numerical simulations incorporating their sharp features. While in simulations, sharp edges and corners in polygonal MNWs are subject to numerical errors stemmed from meshing, leading to extra complexities and deteriorated accuracies in computation. Taking the calculation of the fundamental waveguiding property—the effective refractive index (n_{eff})—as an example (Figure 1b), n_{eff} of circular MNWs converges at a maximum mesh element size of ~ 55 nm with a negligible variation in the finer meshing. Meanwhile, for the convergence for pentagonal MNWs, one has to implement an extremely fine meshing with a maximum element size of 4 nm, even with regional refinement techniques (see Figure S1 in the Supplementary Materials for details). As a result of a much smaller meshing size, the total number of mesh elements multiply rapidly (Figure 1c), adding to the consumption of computational time and resources.

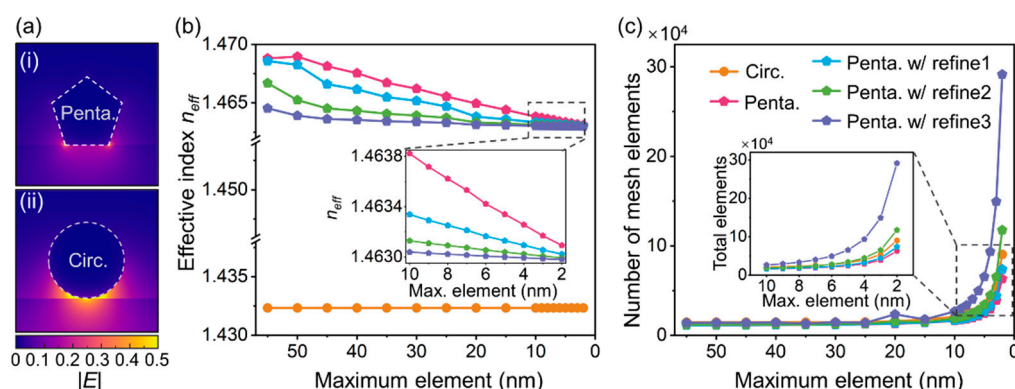


Figure 1. Emerging computational resource requirement for MNWs with abrupt geometric features and broken symmetries. (a) Mode profiles in terms of normalized electric field norm distributions in the typical (i) pentagonal MNW and (ii) circular MNW. For better visualization, they are plotted in a color bar ranging from 0 to 0.5 with saturation. (b) Calculated effective index n_{eff} vs. maximum mesh element size. (c) Total number of mesh elements vs. maximum mesh element size of MNWs. Insets in (b,c): enlarged views for maximum mesh size ranging from 2–10 nm. Orange dotted line: circular MNW. Red, blue, green, and purple lines with pentagonal symbols: pentagonal MNWs with 0–3 times regional meshing refinements (see Figure S1 in the Supplementary Materials for details). The MNW was placed on a silica substrate with a diameter of 300 nm working at 880 nm wavelength.

2.2. Model Architecture with Transfer Learning

In transfer learning, the learned knowledge from one problem can be transferred to multiple problems of the same type, offering opportunities to leverage common features to improve performance and reduce the dataset size [41–43]. As for our case, physically, plasmon modes in MNWs were solutions to the source-free Maxwell's equations in a given configuration (e.g., geometry, wavelengths, and working environments). By defining a

time-harmonic electric field E as $E(x, y, z) = E(x, y)e^{i(\beta z - \omega t)}$, it can be generally described as an eigenvalue problem [44,45]:

$$\begin{aligned} &\text{Find } \beta \in \mathbb{C} \text{ and } 0 \neq E \in H(\mathbf{curl}_\beta, \Omega) \\ &\text{s.t. } \mathbf{curl}_\beta(\mu_r^{-1} \mathbf{curl}_\beta E) = k_0^2 \varepsilon_r E \text{ in } \Omega \end{aligned} \quad (1)$$

under a boundary condition (e.g., $E \times \mathbf{n}|_{\partial\Omega} = 0$). Here, ε_r , μ_r , and k_0 are the permittivity, permeability, and wavenumber, respectively. β is the propagation constant (eigenvalue), Ω is an open set with a boundary $\partial\Omega$, H denotes the Hilbert space, and the operator \mathbf{curl}_β is defined as $\mathbf{curl}_\beta E(x, y) = \mathbf{curl}(E(x, y)e^{i\beta z})e^{-i\beta z}$. Therefore, for MNWs with arbitrary configurations, solving plasmon modes can be regarded as the same type of problem that yields learning transfer among them.

By applying the concept of transfer learning, our model was composed of a base net and a transfer net, enabling the migration of the knowledge acquired from a source task (T_s) through the base net to a target task (T_t) with the transfer net. As is schematically shown in Figure 2a, the base net was constructed with a framework of artificial neural networks (ANNs), consisting of the input, output, and 4 hidden layers to learn the physical properties of plasmon modes in free-standing circular MNWs. The associated dataset for training the base net (source data, $D_s = \{(X_s, Y_s) | X_s \in \mathbb{R}^{m_s \times 2}, Y_s \in \mathbb{R}^{m_s \times 4}\}$) can be readily obtained at an inexpensive computational cost through numerical simulation because the mesh elements of free-standing circular MNWs are far fewer than those of other configurations. Here, $X_s = \{(D^{(j)}, \lambda^{(j)}) | j = (1, m_s)\}$ represents the 2-dimensional configuration vector (diameter D , wavelength λ) of MNWs for m_s data instances. Additionally, $Y_s = \{(n_{eff}^{(j)}, A_m^{(j)}, L_m^{(j)}, \text{FOM}^{(j)}) | j = (1, m_s)\}$ represents the corresponding waveguiding properties derived from the physical quantities β and E , where n_{eff} , A_m , L_m , and FOM represent the effective refractive index, propagation length, mode area, and figure of merit, respectively. n_{eff} and L_m are derived from the real and imaginary part of β (Equations (S1) and (S2) in the Supplementary Materials), reflecting the mode characteristics and the spatial decay along the propagation direction (loss), respectively. A_m was calculated from the energy density integration (Equation (S3) in the Supplementary Materials), describing the capability of the energy confinement. In addition, the FOM provides an overall evaluation of the mode quality (Equation (S4) in the Supplementary Materials). A high mode quality indicates a combination of small loss and tight confinement. In this case, the goal of our T_s is to efficiently learn the basic knowledge of the plasmon modes via the learning objective:

$$\min. \sum_{p=1}^4 \alpha_p \text{Loss}_p(f_s(X_s), Y_s) \quad (2)$$

where $f_s(\cdot)$ is the source predictive function, Loss_p represents the loss per node in the output layer in terms of mean absolute percentage errors (MAPE), reflecting the difference between the prediction and the actual value, and α_p is the corresponding weight factor.

On the other hand, for the target task T_t with a predictive learner $f_t(\cdot)$, the associated dataset (target data, $D_t = \{(X_t, Y_t) | X_t \in \mathbb{R}^{m_t \times 2}, Y_t \in \mathbb{R}^{m_t \times 4}\}$) corresponds to the configuration vector and waveguiding properties of MNWs with broken symmetries or geometric sharp features. The goal of our transfer learning was to simultaneously improve the performance of $f_t(\cdot)$ and reduce the number of target data instances m_t via the knowledge acquired from T_s . To achieve this, a transfer net incorporating 6 hidden layers was deployed, in which the first 3 hidden layers with fixed weights and biases (transferred layers) were transferred from the base net containing the general features extracted in solving T_s , and the rest of the hidden layers were designed to learn the new features of the plasmon modes (Figure 2b). It is worth noting that such a design of the transfer net was also under the guidance of physics. From the physical point of view, the symmetry breaking or geometric sharpness actually lifts the degeneracy of the MNW system, leading to the hybridized mode generated by the coupling between the original symmetric modes [2]. Therefore, the

hybridized plasmon mode contains, but is not limited to, the features of the symmetric mode extracted from the circular MNWs.

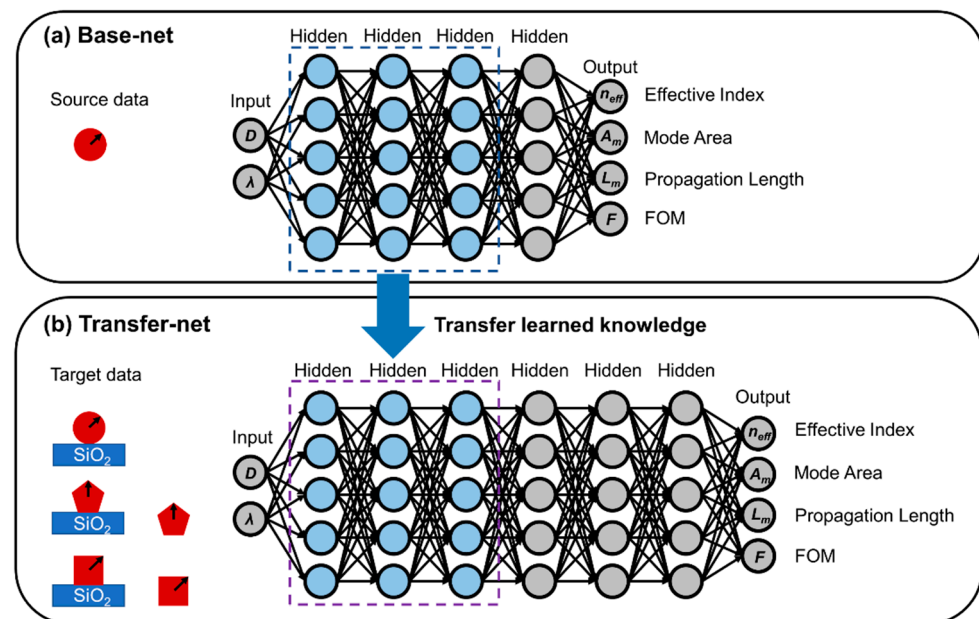


Figure 2. Schematic illustrations of the model architecture. (a) Base net and (b) transfer net based on ANN frameworks. The base net is used for gaining basic knowledge from free-standing circular MNWs, establishing the mapping from the input of diameters (D) and wavelengths (λ) to the output of effective index (n_{eff}), mode area (A_m), propagation length (L_m), and figure of merit (FOM). The transfer net is used to deal with MNWs in complex configurations with abrupt geometric features (free-standing pentagonal and square MNWs) and broken symmetries (silica-substrate-supported circular, pentagonal, and square MNWs). Transfer learning is enabled by migrating learned knowledge (blue neuron nodes within dashed boxes) from the trained base net to the transfer net. The blue rectangles (SiO_2) represent the silica substrates. The red circles and polygons represent MNWs with different cross-sectional shapes. The diameters D of circles/polygons are defined as twice the radii/circumradii (indicated by the black arrows).

The TensorFlow framework was used to construct the ANN. To train our model, the dataset we used contained 1680 groups of data obtained from the numerical simulation of MNWs with 6 types of configuration (280 groups for each), including free-standing and substrate-supported MNWs with circular, square, and pentagonal cross-sections. For each MNW configuration, the dataset was divided into the training dataset (~60%), the validation dataset (~20%), and the test dataset (~20%). Typical geometric sizes and operation wavelengths were considered to cover broad ranges of D (40–300 nm) and λ (520–900 nm) in the visible and near-infrared bands. To minimize the loss function in the training process, the Adam optimizer with an initial learning rate of 10^{-3} and a decaying rate of 0.99 was applied. The base net was firstly trained, and after training the base net, the transferred layers with fixed trainable parameters were transferred to the transfer net. The trainable parameters for the rest of the hidden layers in the transfer net were then initialized by random normal initialization for the training of the transfer net. For performance evaluation, due to the values of different waveguiding properties across several orders of magnitude, the individual performance of each of the four properties (n_{eff} , A_m , L_m , FOM) was evaluated from its MAPE on a given property P :

$$\sigma_P = \frac{100\%}{m_{test}} \sum_{i=1}^{m_{test}} \left| \frac{\hat{P}_i - P_i}{P_i} \right| \quad (3)$$

where \hat{P}_i and P_i represent the predicted and actual property, and m_{test} is the number of data instances in the test set. We also calculated the average of the four MAPEs ($\sum \text{MAPE}_p / 4$) to assess the average error σ_{avg} of our model.

3. Results and Discussion

3.1. Optimized Layout for Gaining the Basic Knowledge

We first discuss the optimized layout for the base net, which was used to gain basic knowledge of the plasmon modes by feeding a large set of training data D_s . Therefore, it is desirable to be lightweight and to have a small number of trainable parameters and an efficient training time without sacrificing the performance of $f_s(\cdot)$. For this purpose, the number of hidden layers (N_b) in the base net was determined by an overall evaluation of errors, trainable parameters, and training time. As is shown in Figure 3, the base net with four hidden layers was able to achieve the overall optimized performance, offering a minimum average error σ_{avg} of 1.94% (the corresponding individual errors σ_{neff} , σ_{Am} , σ_{Lm} , and σ_{FOM} are 0.16%, 2.04%, 3.00%, and 2.55%, respectively) compared to all other choices of N_b . Note that, although the base net with six hidden layers ($N_b = 6$) can achieve a smaller error of the FOM (1.94%) compared to the $N_b = 4$ case, it has a higher average error (2.54%) with more parameters that need to be trained (Figure 3b) and, consequently, a longer time needed for training (Figure 3b inset).

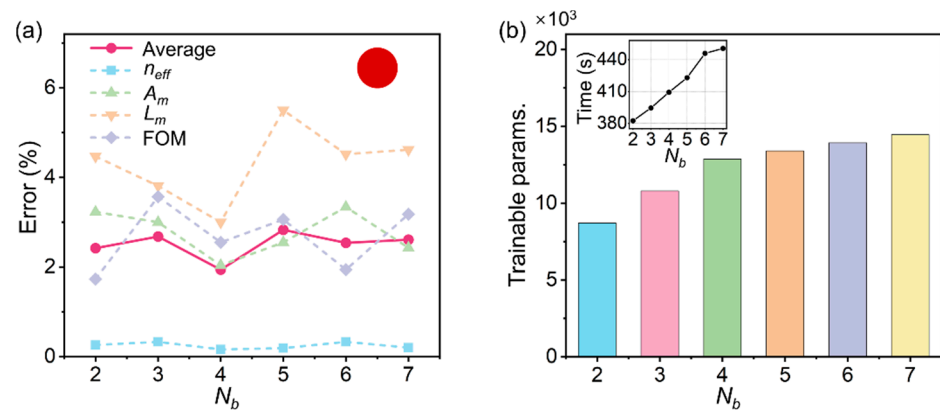


Figure 3. Optimized layout for the base net. (a) Dependence of errors on the number of hidden layers (N_b) in the base net. Dashed lines with squares, up triangles, down triangles, and diamonds represent the individual error of effective index (n_{eff}), mode area (A_m), propagation length (L_m), and figure of merit (FOM), respectively. Red solid line with dots: overall performance in terms of the average error. Inset: schematic illustration of the geometry of the circular MNW. (b) Number of trainable parameters vs. N_b . Inset: an increasing training time with the increase in N_b .

3.2. Performance Improvement in $f_t(\cdot)$ and Reduction in Training Parameters

After training the base net with the optimized layout, a certain number of hidden layers (N_t) in the base net containing the learned knowledge were copied to the transfer net to deal with MNWs of complex configurations. For comparison with the conventional direct learning (DL) approach without the learning transfer, the performance improvement in $f_t(\cdot)$ was evaluated by the change in the average error as:

$$\Delta(N_t) = 100\% \frac{\sigma_{avg}(0) - \sigma_{avg}(N_t)}{\sigma_{avg}(0)} \quad (4)$$

where $\sigma_{avg}(N_t)$ is the average error of the transfer net when N_t layers are transferred, and $\sigma_{avg}(0)$ represents the corresponding error via DL ($N_t = 0$, initializing and training all parameters in the transfer net). Under the above definition, $\Delta(N_t) > 0$ indicates a positive transfer, such that the knowledge stored in the transferred layers facilitates the learning of $f_t(\cdot)$. Meanwhile a negative $\Delta(N_t)$ indicates the opposite situation, and is

known as a negative transfer. Meanwhile, the larger the $\Delta(N_t)$, the greater the performance improvement.

To demonstrate the effectiveness and generality of our model, typical scenarios of MNWs with geometric abruptness or/and broken symmetries, including free-standing pentagonal MNWs (fp-MNWs), free-standing square MNWs (fs-MNWs), substrate-supported circular MNWs (sc-MNWs), substrate-supported pentagonal (sp-MNWs), and substrate-supported square MNWs (ss-MNWs) were investigated. As is shown in Figure 4, $\Delta(N_t)$ increases with more layers being transferred until reaching their minimum at $N_t = 3$, exhibiting an exceptional performance that is superior to the DL and generalization capability that is applicable for the MNW with every new configuration. For all cases at $N_t = 3$, the performance improvements of fp-, fs-, sc-, sp-, and ss-MNWs were 23.2%, 45.4%, 61.3%, 51.9%, and 46.5% compared to DL, yielding excellent average errors ($\sigma_{avg}(3)$) as small as 2.48%, 1.65%, 2.60%, 1.75%, and 2.60%, respectively. Compared to the improvements with learning transfers that were demonstrated in other nanostructures (e.g., among multi-layered films (~23–50%), from multi-layered nanoparticles to multi-layered films (~20%) [33]), a greater enhancement can be achieved for our MNWs, providing an effective way to migrate knowledge across varied configurations in the waveguiding system. It is also worth mentioning that, besides the performance improvement, transfer learning also enabled a reduced number of trainable parameters compared to DL (e.g., ~42% relative reduction in our model, Figure 4f), with a more efficient training process for every new task, consequently, because the transfer layers embedded in the transfer net were already pre-trained.

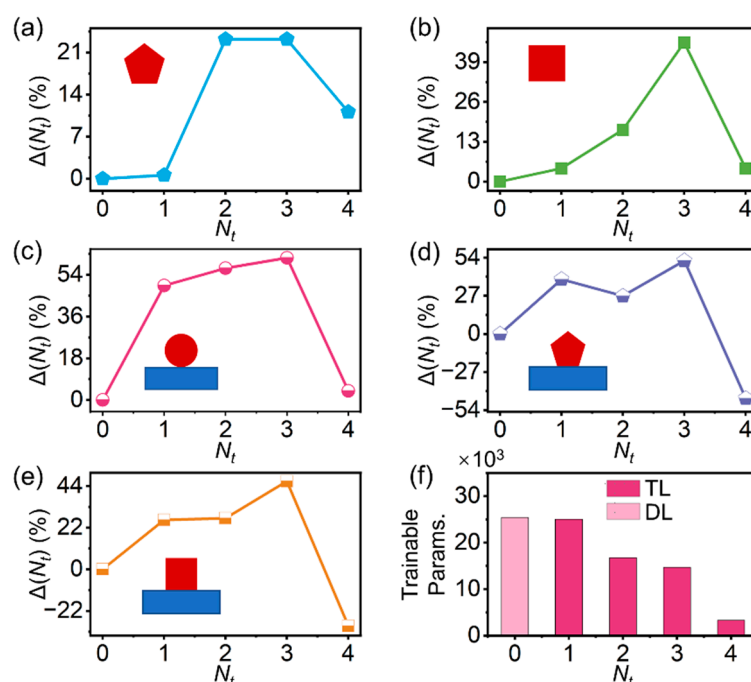


Figure 4. Significantly improved performance with reduced trainable parameters enabled by transfer learning. Dependence of the performance improvement on the number of transferred layers (N_t). (a) free-standing pentagonal MNWs, (b) free-standing square MNWs, (c) substrate-supported circular MNWs, (d) substrate-supported pentagonal MNWs, and (e) substrate-supported square MNWs. (f) Comparison of trainable parameters between direct learning (DL) and the transfer learning (TL). Insets: schematic illustrations of the geometries of MNWs with different configurations.

On the other hand, when we transferred all the hidden layers from the trained base net ($N_t = 4$), $\Delta(N_t)$ dramatically decreased, resulting in a deteriorated performance and even negative transfers for the sp-MNWs and ss-MNWs. Such behaviors correspond well to the characteristics of ANNs. Generally, the features on each layer of the ANN evolve

from general to specific along with the network, and the last layer is therefore very specific to a particular problem [46]. In our base-net case, the specialization of the fourth layer of the circular free-standing MNW for T_s made it inapplicable for T_t , and the transfer of such a layer will only lead to deterioration, rather than the enhancement of the performance for modeling MNWs with other new configurations.

3.3. Removing the Need for a Large Set of Training Data with Reduced m_t

In addition to the performance improvement with the reduction in trainable parameters, our transfer learning model also removes the requirement for a large set of training data D_t for MNWs with every new configuration. To demonstrate, using only a small portion of the training dataset ($\eta = 100\%m_t/m_{tot}$, where m_t is the number of data instances used for training and m_{tot} is the total number of data instances in the training dataset), we evaluated the η -dependent σ_{avg} for fp-, fs-, sc-, sp-, and ss-MNWs (Figure 5), and the corresponding performance of direct learning is also provided for comparison. As is shown, for a given η , the performance of transfer learning was much better than direct learning. Additionally, for a given σ_{avg} , transfer learning enables successful training with a much smaller η . For example, to maintain an acceptable σ_{avg} of $\sim 5\%$, only 20% ($m_t = 33$), 50% ($m_t = 83$), 50% ($m_t = 83$), 40% ($m_t = 66$), and 50% ($m_t = 83$) portions of training datasets are required for fp-, fs-, sc-, sp-, and ss-MNWs, respectively. Such an ability in achieving high performance with small datasets is able to greatly reduce the time of not only training, but also data preparation, that is computationally expensive through numerical simulations, significantly accelerating and facilitating the training and data acquisition process.

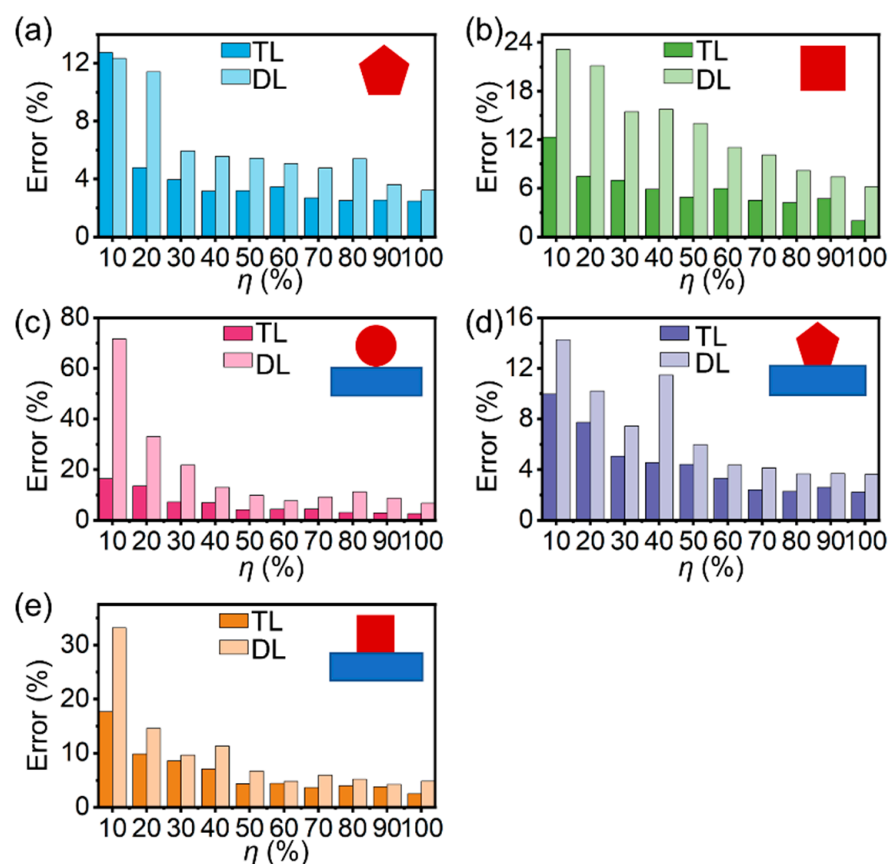


Figure 5. Reduced training dataset size enabled by transfer learning. Comparison of average errors between transfer learning and direct learning using different portions (η) of the training dataset for (a) free-standing pentagonal MNWs, (b) free-standing square MNWs, (c) substrate-supported circular MNWs, (d) substrate-supported pentagonal MNWs, and (e) substrate-supported square MNWs. Insets: schematic illustrations of the geometries of MNWs with different configurations.

3.4. Accurate, Effective and Comprehensive Mapping of Waveguiding Properties

With the merits of excellent performance and reduced dataset size, our transfer learning model circumvents the drawbacks of conventional deep learning, providing an accurate and efficient way to model MNWs with a high generalization capability. Moreover, our model also exhibits an overwhelming performance compared to non-deep learning methods (e.g., DCA and CAE [16]). For comparison, σ_{neff} calculated by our model was 0.33%, 0.18%, 0.12%, 0.23%, and 0.20% for the fp-, fs-, sc-, sp-, and ss-MNWs, which is one order of magnitude smaller than the ones of the DCA and CAE methods (~7% of DCA method and ~8% for the CAE method, see Figure S2 in the Supplementary Materials for details). Such a huge improvement is crucial for various situations where an accurate propagation constant is required (e.g., routers, couplers, and correlators) [47]. Besides the accurate prediction of n_{eff} , our model is able to obtain other waveguiding properties (L_m , A_m , and FOM) that have not been demonstrated by the DCA and CAE methods, yielding a comprehensive study of the plasmon modes in MNWs.

As an illustration, Figure 6 gives λ - D mappings of n_{eff} , A_m , L_m , and FOM for sc-MNWs (Figure 6a–d) and sp-MNWs (Figure 6e–h) using our model (see all configurations in Figure S3, Supplementary Materials). They exhibit no visual discrepancy over the broad ranges of λ and D compared to the results from numerical simulations (TL in Figure 6(ai)–(hi) vs. Sim. in Figure 6(aii)–(hii)). We further overlaid the results from the simulation and our model and provided the contour lines for better visualization (Figure 6(aiii)–(hiii)). As shown, the results obtained from our model (purple dotted lines) coincides very well with the ones using numerical simulations (light purple solid lines), while the time consumptions, by contrast, were reduced by five orders of magnitude (~560 ms for TL vs. ~12 h for Sim.). Therefore, our model offers an effective and effortless approach to systematically characterize the plasmonic waveguiding properties with varied configurations. Taking the mode characteristics as an example, with the increasing λ and D , the plasmon mode in the sc-MNW transits from the bound mode ($n_{eff} > 1.45$) to the leaky mode ($n_{eff} < 1.45$) in the region above the transition line (purple line with the label 1.45 in Figure 6(aiii)), while the plasmon mode in the sp-MNW is always the bound mode within the range of λ and D presented. Therefore, even at the same λ and D , the pentagonal and circular MNWs exhibit distinct A_m and L_m , with differences that can be as large as one order of magnitude (e.g., Figure 6b vs. Figure 6f), resulting in totally different application scenarios. In addition to revealing the mode characteristics, the generated plasmonic mappings also facilitate the optimization of the trade-off relation between confinement and loss, which lies at the heart of the design of plasmonic components and devices. The optimized trade-off can be achieved when the FOM reaches its maximum value. As shown in the FOM mappings generated by our model (Figure 6d,h), the local maximum FOM at a desirable D or λ can be found, as well as the global maximum FOM for all the D and λ combinations can be revealed. This result also indicates the ability to uncover relationships between the functional properties and design variables (e.g., cross-sectional shapes, working environments, geometric sizes, and wavelengths), which can offer a valuable reference and new opportunities for designing high-performance plasmonic components and devices.

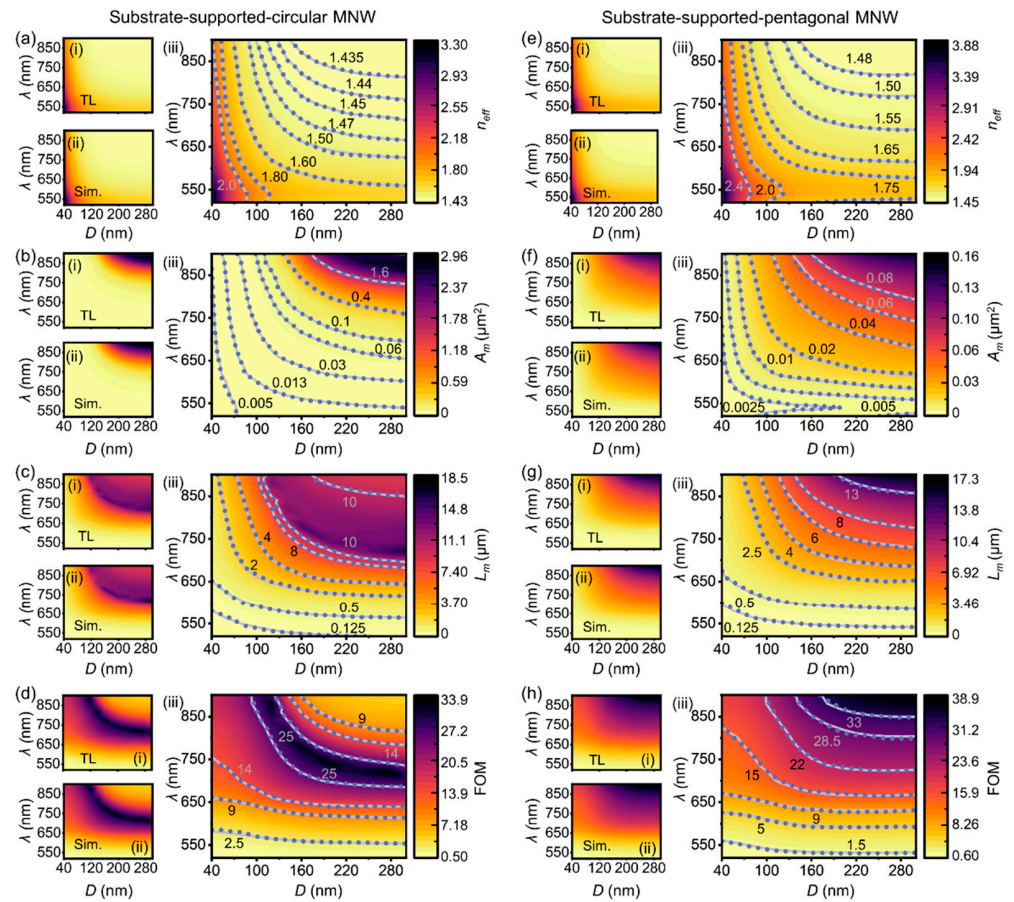


Figure 6. Accurate, effective, and comprehensive mappings of waveguiding properties of MNWs with different configurations enabled by our model. Waveguiding properties of (a–d) substrate-supported circular MNWs and (e–h) substrate-supported pentagonal MNWs over a broad range of diameters (D) and wavelengths (λ). (a,e) Effective index (n_{eff}), (b,f) mode area (A_m), (c,g) propagation length (L_m), (d,h) figure of merit (FOM). Predictions by our model (TL, (i) in (a–h)) coincide well with the numerical simulations (Sim., (ii) in (a–h)), which are further verified by overlaying the images of the TL and Sim. results ((iii) in (a–h)). For reference, contour lines are also provided in (iii), where the TL (purple dotted lines) exhibit excellent agreement with the Sim. (light purple solid lines).

4. Conclusions

In summary, based on a transfer learning approach, we have proposed a general model for predicting the waveguiding properties of MNWs of arbitrary cross-sectional shapes and working environments with varied geometric sizes and wavelengths. The model consists of a base net for learning the basic knowledge from the simple case of free-standing circular MNWs, and a transfer net for dealing with complex MNW configurations. The dependence of errors on the base-net layers and the transferred layers have been investigated to achieve the optimized performance. In addition, the conditions for the positive transfer and the negative transfer have been analyzed to give an insight into our neural network structure. We have showed that, by migrating the learned knowledge from the source task of the base net to the target tasks of the transfer net, the performance of the target tasks can be greatly improved, enabling much smaller errors with less trainable parameters than direct learning. Meanwhile, our model also works well with small datasets, saving ~50–80% of the number of data instances than direct learning, which greatly reduces the time of data preparation through numerical simulation. The generality and robustness of this approach has also been demonstrated by accurate predictions of various MNW configurations with broken symmetries or/and different cross-sectional shapes. Moreover, we have also demonstrated that, compared to other methods (DCA and CAE) that are only capable of retrieving the

effective index, our approach enables not only a much higher accuracy over a broad range of diameters and wavelengths, but also a more comprehensive characterization of the waveguiding properties, reflecting the trade-off between confinement and loss. Additionally, compared to numerical simulations, time consumption is reduced by five orders of magnitude. Benefitting from advantages, including a high performance with generalization capabilities, simple architecture with a small-scale neural network, and a lightweight dataset with a reduced size, our approach offers an effective route for accurately retrieving the plasmonic properties of MNWs without extensive training time and data, which may greatly facilitate the investigation and design of plasmonic/polaritonic components and devices.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/nano12203624/s1>, Figure S1: Illustration of the refinement method.; Figure S2: Mappings of waveguiding properties of MNWs with different configurations.; Figure S3: Dependences of effective indices and errors on normalized diameters. References [48–51] are cited in the supplementary materials.

Author Contributions: Conceptualization, A.L. and Y.W.; methodology, A.L. and Y.W.; formal analysis, A.L., C.Z., Y.F. and X.W.; data curation, C.Z. and Y.F.; writing—original draft preparation, A.L. and Y.W.; writing—review and editing, X.W.; supervision, Y.W. and X.W.; funding acquisition, Y.W. and X.W. All authors have read and agreed to the published version of the manuscript.

Funding: National Natural Science Foundation of China (Nos. 62005031 and 62005032), Fundamental Research Funds for the Central Universities (Nos. 2021CDJQY-046 and 2022CDJXY-018), and Innovation Support Plan for Returned Overseas Scholars (No. cx2021058).

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: The authors thank Lili Zeng for helpful discussion.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Guo, X.; Ma, Y.; Wang, Y.; Tong, L. Nanowire Plasmonic Waveguides, Circuits and Devices. *Laser Photonics Rev.* **2013**, *7*, 855–881. [CrossRef]
- Wei, H.; Pan, D.; Zhang, S.; Li, Z.; Li, Q.; Liu, N.; Wang, W.; Xu, H. Plasmon Waveguiding in Nanowires. *Chem. Rev.* **2018**, *118*, 2882–2926. [CrossRef] [PubMed]
- Wei, H.; Yan, X.; Niu, Y.; Li, Q.; Jia, Z.; Xu, H. Plasmon–Exciton Interactions: Spontaneous Emission and Strong Coupling. *Adv. Funct. Mater.* **2021**, *31*, 2100889. [CrossRef]
- Bellido, E.P.; Bicket, I.C.; Botton, G.A. The Effects of Bending on Plasmonic Modes in Nanowires and Planar Structures. *Nanophotonics* **2022**, *11*, 305–314. [CrossRef]
- Tao, A.; Kim, F.; Hess, C.; Goldberger, J.; He, R.; Sun, Y.; Xia, Y.; Yang, P. Langmuir–Blodgett Silver Nanowire Monolayers for Molecular Sensing Using Surface-Enhanced Raman Spectroscopy. *Nano Lett.* **2003**, *3*, 1229–1233. [CrossRef]
- Nauert, S.; Paul, A.; Zhen, Y.-R.; Solis, D.; Vigderman, L.; Chang, W.-S.; Zubarev, E.R.; Nordlander, P.; Link, S. Influence of Cross Sectional Geometry on Surface Plasmon Polariton Propagation in Gold Nanowires. *ACS Nano* **2014**, *8*, 572–580. [CrossRef] [PubMed]
- Guo, X.; Qiu, M.; Bao, J.; Wiley, B.J.; Yang, Q.; Zhang, X.; Ma, Y.; Yu, H.; Tong, L. Direct Coupling of Plasmonic and Photonic Nanowires for Hybrid Nanophotonic Components and Circuits. *Nano Lett.* **2009**, *9*, 4515–4519. [CrossRef]
- Wang, Y.; Wu, X.; Wang, P. Asymmetric Cavity Mode Engineering in a Single Plasmonic Nanowire. *J. Light. Technol.* **2021**, *39*, 5855–5863. [CrossRef]
- Gu, F.; Zeng, H.; Zhu, Y.B.; Yang, Q.; Ang, L.K.; Zhuang, S. Single-Crystal Pd and Its Alloy Nanowires for Plasmon Propagation and Highly Sensitive Hydrogen Detection. *Adv. Opt. Mater.* **2014**, *2*, 189–196. [CrossRef]
- Wang, Y.; Guo, X.; Tong, L.; Lou, J. Modeling of Au-Nanowire Waveguide for Plasmonic Sensing in Liquids. *J. Light. Technol.* **2014**, *32*, 4233–4238. [CrossRef]
- Wu, X.; Xiao, Y.; Meng, C.; Zhang, X.; Yu, S.; Wang, Y.; Yang, C.; Guo, X.; Ning, C.Z.; Tong, L. Hybrid Photon-Plasmon Nanowire Lasers. *Nano Lett.* **2013**, *13*, 5654–5659. [CrossRef] [PubMed]
- Sidiropoulos, T.P.H.; Röder, R.; Geburt, S.; Hess, O.; Maier, S.A.; Ronning, C.; Oulton, R.F. Ultrafast Plasmonic Nanowire Lasers near the Surface Plasmon Frequency. *Nat. Phys.* **2014**, *10*, 870–876. [CrossRef]

13. Kim, S.-W.; Turng, L.-S. Three-Dimensional Numerical Simulation of Injection Molding Filling of Optical Lens and Multiscale Geometry Using Finite Element Method. *Polym. Eng. Sci.* **2006**, *46*, 1263–1274. [[CrossRef](#)]
14. Lesina, A.C.; Vaccari, A.; Berini, P.; Ramunno, L. On the Convergence and Accuracy of the FDTD Method for Nanoplasmonics. *Opt. Express* **2015**, *23*, 10481–10497. [[CrossRef](#)] [[PubMed](#)]
15. Nayak, J.K.; Jha, R. Numerical Simulation on the Performance Analysis of a Graphene-Coated Optical Fiber Plasmonic Sensor at Anti-Crossing. *Appl. Opt.* **2017**, *56*, 3510–3517. [[CrossRef](#)]
16. Bao, Q.; Wu, H.; Yang, L.; Wang, P.; Guo, X.; Tong, L. Circular-Area-Equivalence Approach for Determining Propagation Constants of a Single-Mode Polygonal Nanowire. *J. Opt. Soc. Am. B* **2022**, *39*, 795. [[CrossRef](#)]
17. Molesky, S.; Lin, Z.; Piggott, A.Y.; Jin, W.; Vucković, J.; Rodriguez, A.W. Inverse Design in Nanophotonics. *Nat. Photonics* **2018**, *12*, 659–670. [[CrossRef](#)]
18. Malkiel, I.; Mrejen, M.; Nagler, A.; Arieli, U.; Wolf, L.; Suchowski, H. Plasmonic Nanostructure Design and Characterization via Deep Learning. *Light Sci. Appl.* **2018**, *7*, 60. [[CrossRef](#)]
19. Chugh, S.; Ghosh, S.; Gulistan, A.; Rahman, B.M.A. Machine Learning Regression Approach to the Nanophotonic Waveguide Analyses. *J. Light. Technol.* **2019**, *37*, 6080–6089. [[CrossRef](#)]
20. Chugh, S.; Gulistan, A.; Ghosh, S.; Rahman, B.M.A. Machine Learning Approach for Computing Optical Properties of a Photonic Crystal Fiber. *Opt. Express* **2019**, *27*, 36414–36425. [[CrossRef](#)]
21. Yao, K.; Unni, R.; Zheng, Y. Intelligent Nanophotonics: Merging Photonics and Artificial Intelligence at the Nanoscale. *Nanophotonics* **2019**, *8*, 339–366. [[CrossRef](#)] [[PubMed](#)]
22. Hegde, R.S. Deep Learning: A New Tool for Photonic Nanostructure Design. *Nanoscale Adv.* **2020**, *2*, 1007–1023. [[CrossRef](#)] [[PubMed](#)]
23. Xu, L.; Rahmani, M.; Ma, Y.; Smirnova, D.A.; Kamali, K.Z.; Deng, F.; Chiang, Y.K.; Huang, L.; Zhang, H.; Gould, S.; et al. Enhanced Light–Matter Interactions in Dielectric Nanostructures via Machine-Learning Approach. *Adv. Photonics* **2020**, *2*, 026003. [[CrossRef](#)]
24. So, S.; Badloe, T.; Noh, J.; Rho, J.; Bravo-Abad, J. Deep Learning Enabled Inverse Design in Nanophotonics. *Nanophotonics* **2020**, *9*, 1041–1057. [[CrossRef](#)]
25. Ma, W.; Liu, Z.; Kudyshev, Z.A.; Boltasseva, A.; Cai, W.; Liu, Y. Deep Learning for the Design of Photonic Structures. *Nat. Photonics* **2021**, *15*, 77–90. [[CrossRef](#)]
26. Anika, N.J.; Mia, M.B. Design and Analysis of Guided Modes in Photonic Waveguides Using Optical Neural Network. *Optik* **2021**, *228*, 165785. [[CrossRef](#)]
27. Wiecha, P.R.; Arbouet, A.; Girard, C.; Muskens, O.L. Deep Learning in Nano-Photonics: Inverse Design and Beyond. *Photon. Res.* **2021**, *9*, B182. [[CrossRef](#)]
28. Jiang, J.; Chen, M.; Fan, J.A. Deep Neural Networks for the Evaluation and Design of Photonic Devices. *Nat. Rev. Mater.* **2021**, *6*, 679–700. [[CrossRef](#)]
29. Wu, X.; Wang, Y. A Physics-Based Machine Learning Approach for Modeling the Complex Reflection Coefficients of Metal Nanowires. *Nanotechnology* **2022**, *33*, 205701. [[CrossRef](#)]
30. Bhadeshia, H.K.D.H. Neural Networks and Information in Materials Science. *Stat. Anal. and Data Min.* **2009**, *1*, 296–305. [[CrossRef](#)]
31. LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)] [[PubMed](#)]
32. Sun, C.; Shrivastava, A.; Singh, S.; Gupta, A. Revisiting Unreasonable Effectiveness of Data in Deep Learning Era. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 843–852.
33. Qu, Y.; Jing, L.; Shen, Y.; Qiu, M.; Soljačić, M. Migrating Knowledge between Physical Scenarios Based on Artificial Neural Networks. *ACS Photonics* **2019**, *6*, 1168–1174. [[CrossRef](#)]
34. Sung, N.; Kim, M.; Jo, H.; Yang, Y.; Kim, J.; Lausen, L.; Kim, Y.; Lee, G.; Kwak, D.; Ha, J.-W.; et al. NSML: A Machine Learning Platform That Enables You to Focus on Your Models. *arXiv*, 2017; preprint. arXiv:1712.05902.
35. Sudharsan, B.; Patel, P.; Breslin, J.; Ali, M.I.; Mitra, K.; Dustdar, S.; Rana, O.; Jayaraman, P.P.; Ranjan, R. Toward Distributed, Global, Deep Learning Using IoT Devices. *IEEE Internet Comput.* **2021**, *25*, 6–12. [[CrossRef](#)]
36. Huang, L.; Yang, X.; Liu, T.; Ozcan, A. Few-Shot Transfer Learning for Holographic Image Reconstruction Using a Recurrent Neural Network. *APL Photonics* **2022**, *7*, 070801. [[CrossRef](#)]
37. Pan, S.J.; Yang, Q. A Survey on Transfer Learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [[CrossRef](#)]
38. Qiu, C.; Wu, X.; Luo, Z.; Yang, H.; He, G.; Huang, B. Nanophotonic Inverse Design with Deep Neural Networks Based on Knowledge Transfer Using Imbalanced Datasets. *Opt. Express* **2021**, *29*, 28406. [[CrossRef](#)]
39. Ma, W.; Cheng, F.; Xu, Y.; Wen, Q.; Liu, Y. Probabilistic Representation and Inverse Design of Metamaterials Based on a Deep Generative Model with Semi-Supervised Learning Strategy. *Adv. Mater.* **2019**, *31*, 1901111. [[CrossRef](#)]
40. Xu, D.; Luo, Y.; Luo, J.; Pu, M.; Zhang, Y.; Ha, Y.; Luo, X. Efficient Design of a Dielectric Metasurface with Transfer Learning and Genetic Algorithm. *Opt. Mater. Express* **2021**, *11*, 1852. [[CrossRef](#)]
41. Day, O.; Khoshgoftaar, T.M. A Survey on Heterogeneous Transfer Learning. *J. Big Data* **2017**, *4*, 1–42. [[CrossRef](#)]
42. Lin, Y.; Li, M.; Watanabe, Y.; Kimura, T.; Matsunawa, T.; Nojima, S.; Pan, D.Z. Data Efficient Lithography Modeling With Transfer Learning and Active Data Selection. *IEEE Trans. Computer-Aided Design Integr. Circuits Syst.* **2019**, *38*, 1900–1913. [[CrossRef](#)]
43. Evci, U.; Dumoulin, V.; Larochelle, H.; Mozer, M.C. Head2Toe: Utilizing Intermediate Representations for Better Transfer Learning. In Proceedings of the International Conference on Machine Learning, PMLR, Baltimore, MD, USA, 28 June 2022; pp. 6009–6033.

-
44. Zolla, F.; Renversez, G.; Nicolet, A.; Kuhlmei, B.; Guenneau, S.; Felbacq, D. *Foundations of Photonic Crystal Fibres*; Imperial College Press: London, UK, 2005.
 45. Nicolet, A.; Geuzaine, C. Waveguide Propagation Modes and Quadratic Eigenvalue Problems. In Proceedings of the 6th International Conference on Computational Electromagnetics, Aachen, Germany, 4–6 April 2006.
 46. Yosinski, J.; Clune, J.; Bengio, Y.; Lipson, H. How Transferable Are Features in Deep Neural Networks? In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014; pp. 3320–3328.
 47. Wang, Y.; Feng, Y.; Zeng, L.; Wu, X. Versatile and High-Quality Manipulation of Asymmetric Modes in Bent Metal Nanowires. *Opt. Mater. Express* **2022**, *12*, 2782.
 48. Johnson, P.B.; Christy, R.W. Optical Constants of the Noble Metals. *Phys. Rev. B* **1972**, *6*, 4370–4379.
 49. Maier, S.A. *Plasmonics: Fundamentals and Applications*; Springer: New York, NY, USA, 2007.
 50. Oulton, R.F.; Bartal, G.; Pile, D.F.P.; Zhang, X. Confinement and propagation characteristics of subwavelength plasmonic modes. *New J. Phys.* **2008**, *10*, 105018. [[CrossRef](#)]
 51. Zhang, S.; Xu, H. Optimizing Substrate-Mediated Plasmon Coupling toward High-Performance Plasmonic Nanowire Waveguides. *ACS Nano* **2012**, *6*, 8128–8135. [[CrossRef](#)] [[PubMed](#)]