ACS Photonics

pubs.acs.org/journal/apchd5 Article

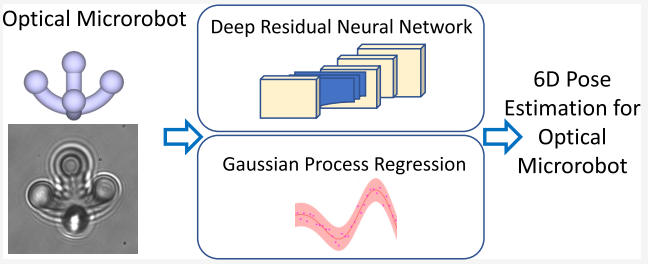
Data-Driven Microscopic Pose and Depth Estimation for Optical Microrobot Manipulation

Dandan Zhang,* Frank P.-W. Lo, Jian-Qing Zheng, Wenjia Bai, Guang-Zhong Yang,* and Benny Lo

Cite This: *ACS Photonics* 2020, 7, 3003−3014 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Optical microrobots have a wide range of applications in biomedical research for both in vitro and in vivo studies. In most microrobotic systems, the video captured by a monocular camera is the only way for visualizing the movements of microrobots, and only planar motion, in general, can be captured by a monocular camera system. Accurate depth estimation is essential for 3D reconstruction or autofocusing of microplatforms, while the pose and depth estimation are necessary to enhance the 3D perception of the microrobotic systems to enable dexterous micromanipulation and other tasks. In this paper, we propose a data-driven method for pose and depth estimation in an optically manipulated microrobotic system. Focus measurement is used to obtain features for Gaussian Process Regression (GPR), which enables precise depth estimation. For mobile microrobots with varying poses, a novel method is developed based on a deep residual neural network with the incorporation of prior domain knowledge about the optical microrobots encoded via GPR. The method can simultaneously track microrobots with complex shapes and estimate the pose and depth values of the optical microrobots. Cross-validation has been conducted to demonstrate the submicron accuracy of the proposed method and precise pose and depth perception for microrobots. We further demonstrate the generalizability of the method by adapting it to microrobots of different shapes using transfer learning with few-shot calibration. Intuitive visualization is provided to facilitate effective human-robot interaction during micromanipulation based on pose and depth estimation results.



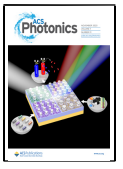
KEYWORDS: *optical microrobot, pose estimation, depth estimation, deep learning, image processing*

Optical microrobots represent a growing area of robotics research for biomedicine,[1] owing to advances in materials and microfabrication technologies. However, due to inherent challenges of microfabrication, assembly, and functionalization of sensors in microscales, ex vivo and in vitro microrobotic experimental setups normally rely on the camera view to observe and characterize the microrobots. Therefore, the vision system is essential for simultaneous, independent trapping and manipulation of different microobjects. The camera view can be used to implement computer vision algorithms for microrobot monitoring, which provides a form of feedback to improve user perception and enable the development of closed-loop control techniques.

For the study of microparticles that normally do not need pose estimation, position tracking has been widely explored. For example, position tracking of a microparticle has been developed for the digital video microscopy,[2] which has been used for optical tweezer calibration,[3] biomolecular forces measurement,[4] exploration of rheology of complex fluids,[5] and assessment of mechanical properties of soft tissues.[6] Some visual tracking algorithms have been translated and applied for tracking microparticles, especially for tracking fluorescent particles and molecules.[7,8] Machine-learning techniques have been applied for fast and accurate feature localization in holograms of colloidal particles,[9] while convolutional neural networks have shown their capabilities in particle tracking for microscopy applications.[10−12]

In addition to tracking microparticles, the robust vision-based techniques for tracking and 6D pose estimation of microrobots are necessary for micromanipulation and microassembly. Accurate depth and pose estimation plays an important role in closed-loop control of microrobots and improving the microscale 3D perception for the operator. It can support automatic microassembly tasks and indirect manipulation of objects such as cells.[13] For large-scale robotic systems, depth can be estimated by using a time-of-flight camera, structured illumination, pulsed Lidar, or stereo camera technologies. However, for microscale robotic systems, most of the available sensing and camera technologies are not able to estimate depth reliably. The transparency of optical micro-

Received: June 22, 2020
Published: September 25, 2020



robots and varying illumination levels set by the users bring additional challenges to depth estimation. Compared to less transparent or opaque objects, a transparent object can complicate pose estimation, since the motion of the microrobots along the optical axis may be unobservable due to parallel projection.[14]

Previous work in the area has proposed a number of methods for depth or pose estimation of microrobots. For example, the depth recovery of microgrippers and Carbon Nanotubes (CNT) for automation applications has been proposed.[15] In microrobotics research, feature-based depth estimation for microgrippers/microrobots with a Scanning Electron Microscope (SEM)[16−18] and 3D orientation estimation of the microrobot have been investigated.[19] Model-based tracking of a magnetic intraocular microrobot has also been proposed.[20] However, existing methods are often designed for a specific microgripper or microrobot, and they cannot be adapted to different microrobotic systems, which require the manipulation of mobile microrobots with a complex shape and various poses.

Compared to traditional approaches, machine learning based algorithms, such as neural networks, can provide more general solutions to assist with the processing of optical images for clinical and biomedical applications.[21−23] For optical microrobot depth estimation, convolutional neural networks (CNNs) combined with long short-term memory (LSTM) have been used[24] for sequential data regression based on gray scale monocular images generated in an Optical Tweezer (OT) setup during offline analysis, where trajectory reconstruction of microrobots was implemented following the periodic trajectory (such as the sinusoidal trajectory and the triangular trajectory). However, the movements of the microrobot are often random during micromanipulation. The model trained on one group of predefined periodic trajectories may not work well in others, since the microrobots can move with random motions. Moreover, the experimental results of prior works were only validated on one model, and it has not been tested for different microrobots with complex shapes.

In addition to depth estimation, 3D pose estimation in microrobotics could also benefit from recent advances in artificial intelligence. For example, CNNs have been employed for the estimation of the 3D pose of an object from a single image.[25] A CNN-based method for estimating the 3D pose and depth of optically transparent microrobots has been developed.[26] However, for a neural network to be able to estimate the 3D pose of different microrobots, a large database should be collected, which consists of microscopic images of different microrobots with the combination of different depth values and poses. The network for pose and depth estimation has to be trained from the raw data of the different robots, which is time-consuming. Moreover, the pose estimation was implemented with a relative orientation estimation mode, which means that accumulative errors may severely affect the accuracy of pose estimation. Therefore, a more generic method is needed for pose and depth estimation for other microrobots given that the available data for training the machine learning model is limited. It has been proved that the transfer learning framework can assist the discovery of the general underlying physical rules.[27] Therefore, to enable the method proposed in this paper to be adaptable to multiple microrobotic systems with various designs, transfer learning is used to reduce the need for collecting a large database for each microrobot.

To realize the target of precise pose and depth estimation mentioned above, an algorithm is designed in this paper. With an accurate pose and depth estimation of the optical microplatform/microrobots based on the video data, additional views can be used to enhance the 3D perception capabilities of the operators in the microscale. Since the camera can only capture the top-down view, an additional side view or 3D view can enable the operator to have a better sense of the microrobot's depth and pose within the targeted workspace. This paves a way for the automation in the 3D space for a microrobotic system, which has a high requirement on the sensing capabilities.

The main contributions of this paper are listed as follows: (1) Depth estimation for an optical microplatform is constructed based on a focus measurement and a GPR model; (2) A Hierarchical Semi Supervised ResNet architecture is developed in a multitask learning manner, which can be used for optical microrobots pose and depth estimation; (3) A GPR-ResNet Hybrid model is developed to further improve the accuracy of the depth estimation of a microrobot with various poses, which constructs a more precise depth estimation approach for optical microrobots; (4) Domain adaption is verified to demonstrate the generality of the proposed method for applications in different optical microrobotic systems; (5) A software for digital video microscopy is developed based on the pose and depth estimation results, with visualization of side views and 3D views of the microrobots, which is designed to assist the operator in conducting dexterous micromanipulation.

The remainder of the paper is structured as follows. First, the problem statement is presented, while the methodology is illustrated in Methodology for the depth estimation of microplatforms, as well as the pose and depth estimation of mobile microrobots. Experiments describes the experimental setup, while the results analyses are presented. Finally, the conclusions are drawn and the future prospects are described in Conclusions and Future Work.

METHODOLOGY

The proposed method aims to construct a precise and robust microscopic pose and depth estimation for optically transparent objects with complex shapes. Two types of objects are studied, namely, microplatforms (which are static objects) and mobile microrobots (which can be moved and have different out-of-plane poses during the operation). An illustration of the microplatforms and the mobile microrobots is shown in Figure 1, which includes micro Platform A, micro Platform B, micro mobile Robot A, and micro mobile Robot B.

2D Position Estimation. Given the microscopic video data, a bounding box is manually placed to identify the initial position of the microplatform/microrobot of interest. Each image frame of the video is cropped according to the bounding box. A Gaussian filter is applied to remove the noises from the images. Subsequently, a binary segmentation of the microplatform/microrobot is generated by applying a threshold operation to the video.

After preprocessing the image data, 2D positions are then computed from the center of mass of this binary microplatform/microrobot image after segmentation. The contrast of the images can be adjusted to make sure that the microplatform can be segmented by the thresholding.

Suppose that $p(x, y)$ represents the pixel of each image frame, $C(x, y)$ represents the gray scale level of that pixel. The

https://dx.doi.org/10.1021/acsphotonics.0c00997
ACS Photonics 2020, 7, 3003−3014

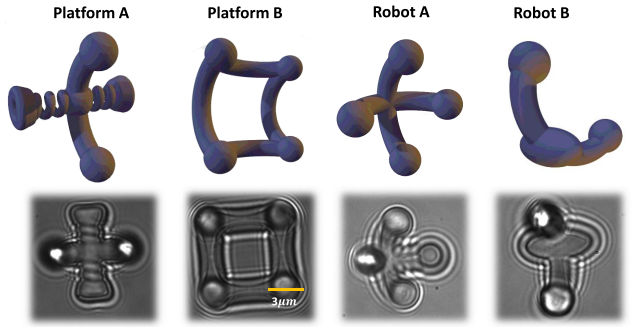


Figure 1. Illustration of the microplatforms and the mobile microrobots used for the validation of the proposed algorithm in this paper (top row images are CAD models and bottom row images are the indicative frames from the experimental data sets). Platforms A and B are microplatforms, and their poses will not change during experiments. Robots A and B represent a four-arm and a two-arm microrobot, respectively, which may have various poses during micromanipulation experiments.

centroid of the segmented area is $[x_c, y_c]$, which can be calculated as follows:

$$x_c = \frac{\sum_{x,y} C(x, y)x}{\sum_{x,y} C(x, y)}; \quad y_c = \frac{\sum_{x,y} C(x, y)y}{\sum_{x,y} C(x, y)} \tag{1}$$

In this way, the 2D planar position of the microplatforms/microrobots can be determined. A new video can be generated with a reduced dimension, whose center point is aligned with the centroid of the microrobot. Therefore, the resulting images will only contain the features of interest, which speeds up the calculation as the data size is reduced, without removing any necessary information. After determining the 2D position of the microrobot, depth and pose estimations can be conducted.

Focus Measurement. For a microplatform, the depth estimation is significant in visualizing the respective structure. Focus measurement information obtained from monocular microscope image sequences can be exploited to estimate the depth. The algorithm utilized to measure the focus level for every image pixel is usually referred to a focus measure (FM) operator. In order to estimate the translation along the z-axis (which is parallel to the optical axis) of the microplatform, the focus measurement information is utilized, which means that the mapping between the focus measurement for each image frame and the corresponding z-position of the microplatforms/microrobots should be established. The full image is used for the focus measurement calculations.

The approximation of a single focus measurement metric cannot be used directly to reconstruct the z-displacement, since two different depths can correspond to the same value of a specific focus measurement. Therefore, combining different metrics for depth regression is necessary. The metrics used for focus measurement are illustrated as follows.

For an image, entropy is related to the complexity contained of its content. Whenever the microplatform is displaced from the focal plane increase or decrease, the entropy value changes accordingly. For entropy based measurement method, prior to estimating the translation along the z-axis, a calibration routine is required to establish the mapping between entropy and the displacement of the microrobot along the z-axis. Since a focused image is expected to have a higher information content, the entropy computed can be used as one of the focus measures.[28]

Apart from the image entropy, the variance of the image Laplacian[29] can be used, which represents the second derivative of an image. The Laplacian highlights regions of an image containing rapid intensity changes and are often used for edge detection. If an image has a high variance, then the image is likely to be in-focus, since it contains many edge-like features. On the other hand, if it is of very low variance, then the image is likely to be blurred, with less edges; hence, the object may be out of focus.

The focus measure based on an alternative definition of the Laplacian is proposed,[30] which is defined based on the modified Laplacian. Suppose that $\Delta_m I(i, j)$ is the modified Laplacian of $p(x, y)$, computed as $\Delta_m I = |I^* L_X| + |I^* L_Y|$ ($*$ is the convolutional operator). The convolution masks used to compute the modified Laplacian are $L_X = [-1, 2, -1]$, and $L_Y = L_X^T$. Another popular focus measure based on the magnitude of image gradient is known as the Tenengrad focus measure.[28]

The four focus measurement metrics (M1: entropy; M2: Laplacian variance; M3: modified Laplacian; M4: Tenengrad operator) are defined as follows.

$$\mathrm{M1} = -\sum_{k=1}^{L} P_k \log(P_k) \tag{2}$$

$$\mathrm{M2} = \sum_{x,y} (I(x, y) - \Delta I)^2 \tag{3}$$

$$\mathrm{M3} = \sum_{x,y} \Delta_m I(x, y) \tag{4}$$

$$\mathrm{M4} = \sum_{x,y} ((G_x(i, j)^2 + G_y(i, j)^2)) \tag{5}$$

where P_k is the relative frequency of the kth gray scale level, L represents the total number of different gray scale levels, I is the image Laplacian obtained by convolving $p(x, y)$ with the Sobel operator, ΔI is the mean value of the image Laplacian, G_x and G_y are the image gradients in the x, y directions, computed by convolving the given image $p(x, y)$ with the Sobel operator.

The focus measurement can be affected by the presence of image noises from different noise sources. Considering that image noise is unavoidable, while the robustness of statistics-based operators may not be stable under various image contrast and image saturation, combining multiple focus measure is essential for accurate depth estimation. Therefore, a statistic regression model based on the focus measures should be explored.

Gaussian Process Regression. Gaussian process regression (GPR) is a nonparametric Bayesian approach that has many desirable properties and can be applied for depth estimation with digital video microscopy.[31] It is data-efficient and is easy to obtain through simple parametrization. A natural Bayesian interpretation is used to allow uncertainty during prediction. Therefore, GPR is used for the regression of the depth model by considering the multiple focus measurement metrics.

Let $\boldsymbol{M}(p_t)$ $(t = 1, 2, ..., T)$ denote the focus measure results of the image p collected at time step t using metric $\boldsymbol{M} = [M_1, M_2, ..., M_N]^T$ $(i = 1, 2, ..., N)$, where N is the total number of metrics ($N = 4$ is used in this paper), T is the total number of time steps. $f(.)$ is a function that maps the focus measures to a depth value. Considering that the training data may be noisy,

https://dx.doi.org/10.1021/acsphotonics.0c00997
ACS Photonics 2020, 7, 3003−3014

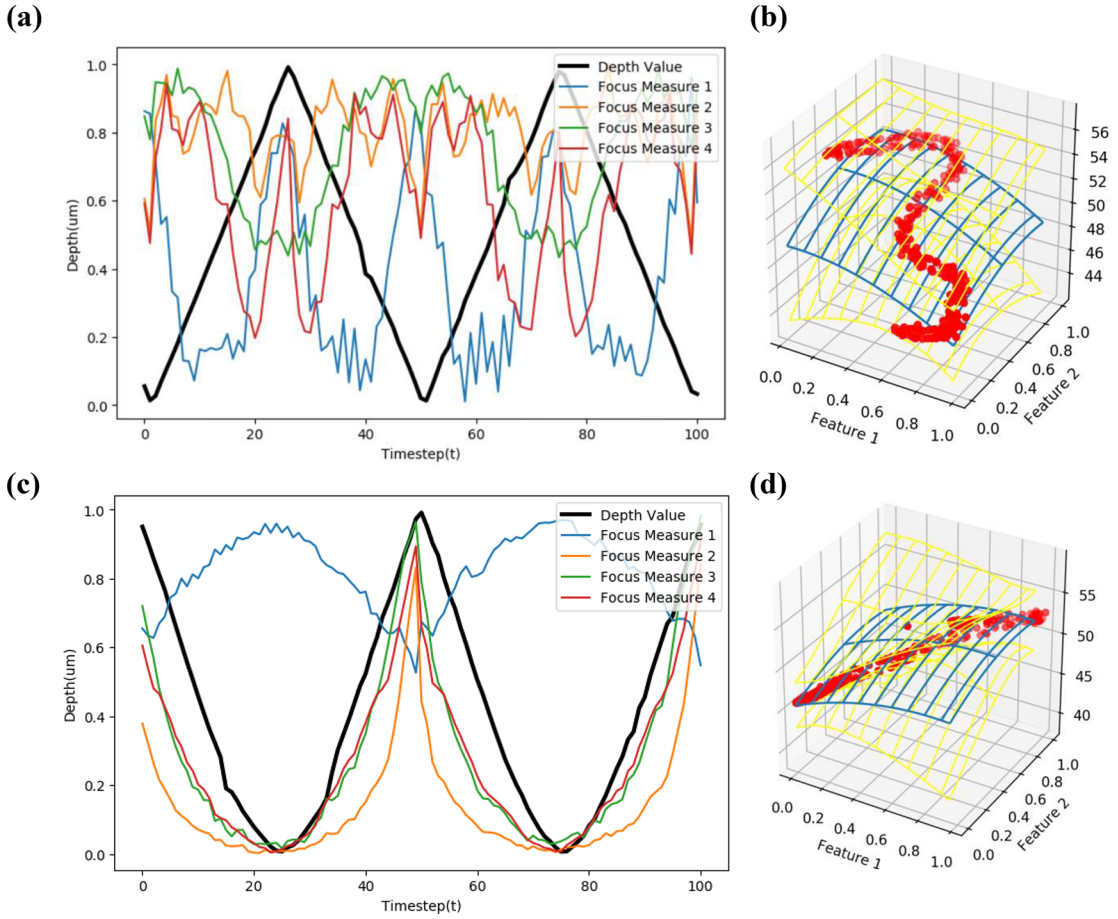


Figure 2. Overview of the focus measure results and the corresponding normalized depth level for microplatforms depth estimation calibration. (a) The normalized focus measurement metrics of different normalized depth values for microplatform A. (b) The visualization of the GPR results for microplatform A using two features. (c) The normalized focus measurement metrics of different normalized depth values for microplatform B. (d) The visualization of the GPR results for microplatform B using two features.

we adopt a regression model that contains an additive independent Gaussian noise term ϵ. Therefore, the regression model for the task with noise can be written as $z_t = f(\boldsymbol{M}(p_t)) + \epsilon$, where z_t represents the target depth values, $\epsilon \sim \mathcal{N}(0, \sigma_n^2)$ follows a Gaussian distribution.

Given a training data set $\mathbf{D} = \{(\boldsymbol{M}(p_t), z_t)|t = 1, 2, ..., T)\}$, a GPR model is trained to predict the value of a scalar output z_t, given new image data and the training data set D. Let $\boldsymbol{M}(\boldsymbol{p})$ denote an input vector with dimension $N \times T$ ($\boldsymbol{M}(\boldsymbol{p}) \in \mathbb{R}^{N\times T}$).

A Gaussian process can be regarded as a distribution over functions and is specified by a mean function and a covariance function. $f(.)$ is assumed to be distributed as a Gaussian process $\boldsymbol{GP}$, which can be denoted by eq 6.

$$f(\boldsymbol{M}(\boldsymbol{p})) \sim \boldsymbol{GP}(\mu(\boldsymbol{M}(\boldsymbol{p})),\ K(\boldsymbol{M}(\boldsymbol{p}),\ \boldsymbol{M}(\boldsymbol{p}'))) \tag{6}$$

where $\mu(.)$ represents the mean function, $K(.)$ denotes the covariance function, which models the dependence between the depth values when given different microscopic images with different focus measurement values ($\boldsymbol{M}(\boldsymbol{p})$ and $\boldsymbol{M}(\boldsymbol{p}')$) and can be calculated as follows.

$$K(\boldsymbol{M}(\boldsymbol{p}), \boldsymbol{M}(\boldsymbol{p}')) = \mathbb{E}[(f(\boldsymbol{M}(\boldsymbol{p})) - \mu(\boldsymbol{M}(\boldsymbol{p})))(f(\boldsymbol{M}(\boldsymbol{p}')) - \mu(\boldsymbol{M}(\boldsymbol{p}')))] \tag{7}$$

A Radial Basis Function (RBF) is used as the kernel to model the covariance function $K(.)$.

As an infinite-dimensional multivariate Gaussian distribution, all the data in the training set are jointly Gaussian distributed. The predictive distribution can be generated based on conditioning the joint Gaussian prior distribution on the observations.[32]

At test time, given an observation $\boldsymbol{M}(\boldsymbol{p}^*)$ ($\boldsymbol{p}^* = p_t(t = 1, 2, ..., T')$), the GPR model predicts the depth as $f(\boldsymbol{M}(\boldsymbol{p}^*))$. In the regression, the main target is to make inferences about the conditional distribution. Denote $X = \boldsymbol{M}(\boldsymbol{p})$, $X^* = \boldsymbol{M}(\boldsymbol{p}^*)$, the joint prior distribution of the training outputs $f(X)$ and the testing outputs $f(X^*)$ according to the prior can be modeled as follows.

$$\begin{bmatrix} f(X) \\ f(X^*) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu(X) \\ \mu(X^*) \end{bmatrix}, \begin{bmatrix} K(X, X) + \sigma_n^2\mathbf{I} & K(X^*, X)^T \\ K(X^*, X) & K(X^*, X^*) \end{bmatrix}\right) \tag{8}$$

After normalization to ensure prior mean to be zero, previous observations $f(X)$ and the target values $f(X^*)$ follow a multivariate normal distribution. The prediction of the GPR model can be obtained as shown in eq 9.

$$f(X^*)|X^*, X, f(X) \sim \boldsymbol{N}(K(X^*, X)[K(X, X) + \boldsymbol{I}\sigma_n^2]^{-1}f(X), K(X^*, X^*) - K(X^*, X)[K(X, X) + \boldsymbol{I}\sigma_n^2]^{-1}K(X, X^*)) \tag{9}$$

Based on the joint posterior distribution, the target value can be represented by the mean values, while the variances in the covariance matrix of $f(\boldsymbol{M}(p^*))$ indicate the uncertainties of the

https://dx.doi.org/10.1021/acsphotonics.0c00997
ACS Photonics 2020, 7, 3003−3014

target values predicted by the model. More details of the calculation for GPR can be obtained in ref 33.

Depth Estimation for Microplatform. For microplatforms that are static, we assume that the regression function mapping from the images (or the focus measurement metrics) of the target microplatforms to its relative depth levels follows a Gaussian process. Therefore, GPR can be used to regress the model. Since the GPR method is a data-efficient approach, a small data set for each microplatform is enough to regress the model. The model training process can also be known as a few-shot regression process, which maps the corresponding focus measurement metrics of the image sequence to the depth values.

During the depth calibration process, each microplatform is placed on the sample holder, while the focus plane is adjusted to obtain different images of the microplatform with different relatively depth values. The 2D position detection is used to obtain the region of interest of the microplatform. Subsequently, the focus measurement is conducted to generate the data for depth regression with GPR model. The depth values (ground truth data) and the obtained focus measurement metrics are normalized between 0 and 1 before calibration. Figure 2a and c show the normalized focus measurement metrics for different normalized depth values for microplatforms A and B, respectively.

For offline analysis, more metrics should be used to improve the accuracy of the regression. However, for online application, two features are sufficient for the depth regression. The two features with higher variance are chosen to train the GPR model. The visualization of the GPR training results is shown in Figure 2b and d for microplatforms A and B, respectively.

Though the GPR method can be extended to depth estimation for microrobots and the underlying regression process can be explained, the accuracy is not high enough. Moreover, for mobile microrobots, the calibration is required to be conducted with various out-of-plane poses. A robust approach for mobile microrobots depth estimation with various out-of-plane poses should be developed.

Pose and Depth Estimation for Microrobots. Depth and pose estimations are critical for monitoring and controlling microrobotic systems, but it is very challenging to obtain accurate estimations from a single image, due to the loss of 3D information during the image capture process.[34] Compared to handcrafted features, which are implemented based on focus measurements, the feature maps learned by neural networks can describe a more generalized mapping between the image and the specific depth value or pose. Therefore, we investigate the use of deep learning based methods to construct the digital video microscopy for microrobots with various poses.

The Deep Residual Network (ResNet) can achieve high classification accuracy by training with hundreds of layers of neurons with good performance,[35] and it can be used for depth and pose estimations. In a traditional CNN model, inputs are passed from nodes of each layer to the next, with adjacent layers connected by convolutional operators. ResNets extend CNNs by adding short-term memory to each layer of the network, which forces the new layer to learn something different from what the input has already learned. ResNet was chosen to be integrated in the digital video microscopy for microrobots with various poses.[35]

The coordinate definitions of microrobots A and B are shown in Figure 3a and c, respectively. The microrobots can rotate along the X, Y, and Z axes, and the rotation angles along

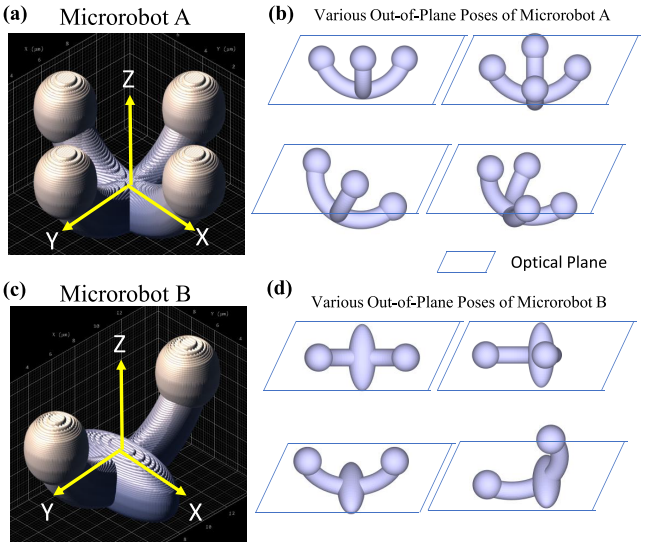


Figure 3. Coordinate definition and the illustration of out-of-plane poses of microrobots A and B. (a) The coordinate definition of microrobot A. (b) Various out-of-plane poses of microrobot A. (c) The coordinate definition of microrobot B. (d) Various out-of-plane poses of microrobot B.

the axes are known as pitch, roll, and yaw angles, respectively. The rotation along the Z axis can be easily observed, since Z axis is parallel to the optical axis. The rotation along X and Y axes is known as out-of-plane rotation, which is often difficult to be perceived by the operators.

The target of the 6D pose estimation of microrobots includes planar position estimation (2D), depth estimation (1D), planar rotation angle estimation (1D), and out-of-plane pose estimation (2D). The 2D planar position estimation has been addressed in the previous section, so the main focus of the construction of the deep learning model is to realize depth level (z_k, $k = 1, 2, ..., K$) classification, out-of-plane pose ($\alpha_i\beta_j$, $i = 1, 2, ..., I$, $j = 1, 2, ..., J$) classification, as well as planar rotation angle (γ_m, $m = 1, 2, ..., M$) classification, since all the data is in a discrete form during data collection. The basic model used for training is shown in Figure 4a. The ranges of the rotation angles and the depth level are determined by the specific application scenario.

A representative data set is built with ground-truth data, which spans across sufficient configuration spaces of the microrobot workspace. For each image collected, it has a corresponding label of $[\alpha_i, \beta_j, z_k]$, where α represents pitch angle, β represents roll angle, and z is the depth value. The minimal displacement between the rotation of the X and Y axes is 10°, while the displacement along the Z axis is 1 μm; $n = I \times J \times K$ number of unique classes for the database are collected, where $I = [\max(\alpha) - \min(\alpha)]/10 + 1$, $J = [\max(\beta) - \min(\beta)]/10 + 1$, and $K = \max(z) - \min(z) + 1$.

To reduce the data required to be collected, the database for the planar rotation angle estimation is self-generated.[36] In the original database, the default planar rotation angle for all the robot is 0°. That is to say, we did not collect the image data with different planar rotation angles. To obtain the ground-truth data for training, the recognition of different rotation angles, a set of predesigned geometric transformations (e.g., planar rotation for $\gamma_m (m = 1, 2, ..., M)$ degree) were applied to images with various out-of-plane poses and depth values in the database, where $M = [\max(\gamma_m) - \min(\gamma_m)]/5$ degree. To this

https://dx.doi.org/10.1021/acsphotonics.0c00997
ACS Photonics 2020, 7, 3003−3014

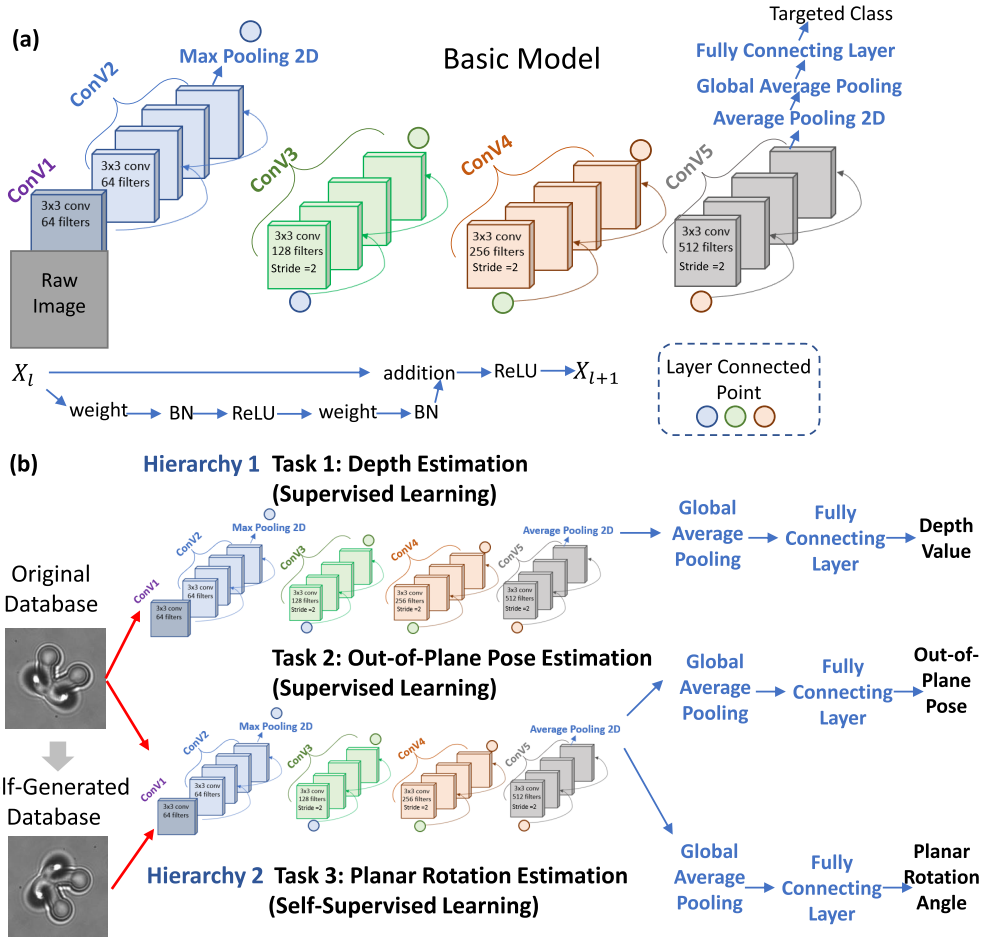


Figure 4. ResNet model and the overall framework. (a) The details of the basic model. (b) The overview of the Hierarchical Semi-Self Supervised ResNet architecture that is targeted for multitask learning of microrobots out-of-plane pose, depth, and planar rotation estimation.

end, the neural network is trained to recognize how many degrees each microrobot rotates in the Z axis by fixing all the parameters of the convolutional blocks and tuning the parameters of all the remaining parameters that are unfixed.

A Hierarchical Semi Supervised ResNet architecture is therefore constructed for the multitask classification, which is implemented by two parallel neural networks with the same structure using the original database for the depth and out-of-plane pose estimation, respectively, and the self-generated database is used for the planar rotation estimation, as shown in Figure 4b. In Hierarchy 1 of the neural network architecture, supervised learning is used for out-of-plane pose and depth estimation based on the labeled data. As for the planar rotation angle estimation, supervised learning is conducted based on the self-generated image data and the corresponding labels in Hierarchy 2. In this way, the Hierarchical Semi Supervised learning for multitask classification can be realized.

Hybrid Model with Few-Shot Calibration. For the microscopic pose and depth estimations of microrobots, an ideal training set would be the one that encompasses all possible poses and depth values. However, it is almost impossible to collect a complete database with all the targeted poses and depth values for every microrobot with a specific shape. Moreover, the data collected for the database is often in the form of snapshot data in a discrete form, which is nonsequential data. For micromanipulation tasks using optical microrobots, the data is sequential in a continuous form. This limits the accuracy for estimation when formulating the neural network model as a regression model. The classification model trained with data in a discrete form may not work for continuous depth values estimation. To enhance the precision of regression, a hybrid model based on the combination of GPR and ResNet is proposed.

One of the advantages of the GPR-based method is that it incorporates prior knowledge and is explainable. However, it is pose-dependent, and the root-mean-square-error (RMSE) of the regression is not highly compared to the deep learning method, and the results depend significantly on the accuracy of the pose estimation results. For the microrobot at the same depth level, if the pose of the microrobot changes, the focus measurement values will change accordingly, which may lead to another depth value using the same regression model. As for the ResNet model, higher accuracy for depth estimation can be achieved for microrobots with various poses. However, the black-box effects of the deep learning methods are difficult to be eliminated. Therefore, the combination of the two methods can complement each other and yield an explainable yet accurate pose estimation.

To further improve the accuracy of the depth estimation when the microrobots are switching between different poses, both GPR and the ResNet model are used. The GPR-based method is pose-dependent, which means that different GPR regression models have to be trained for the same microrobot with different poses. Few-shot calibration means that, for a

https://dx.doi.org/10.1021/acsphotonics.0c00997
ACS Photonics 2020, 7, 3003−3014

microrobot with a specific pose, a series of images with various depth levels are collected to generate the GPR model.

A Kalman filter is used to remove the noises for the real-time estimation results and fuse the information from the GPR and the ResNet models together. A process model is normally used to model the transformation of the process state, which can usually be represented as a linear stochastic difference equation. $x(t)$ represents the real depth value at time step t, and $v(t) = x(t+1) - x(t)$ represents the depth value increment. Therefore, the state-space function can be written as follows.

$$\begin{bmatrix} x(t) \\ v(t) \end{bmatrix} = A \begin{bmatrix} x(t-1) \\ v(t-1) \end{bmatrix} + \begin{bmatrix} \epsilon_g \\ \epsilon_d \end{bmatrix} \tag{10}$$

where A is the transition matrix for state space, $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, $X(t) = \begin{bmatrix} x(t) \\ v(t) \end{bmatrix}$ represents the state at time t, and ϵ_g and ϵ_d are random variables representing the process noise.

The measurement model is defined to describe the relationship between the process state and the measurements. Suppose that the depth level estimated by the neural network is z_d, while the depth value regressed by the GPR model is z_g. Kalman filter is used to fuse the information together. The measurement equation can be expressed as follows.

$$\begin{bmatrix} z_d(t) \\ z_g(t+1) - z_g(t) \end{bmatrix} = H \begin{bmatrix} x(t) \\ x(t+1) - x(t) \end{bmatrix} + \begin{bmatrix} \zeta_g \\ \zeta_d \end{bmatrix} \tag{11}$$

where H is an identity matrix, ζ_g and ζ_d represent measurement noise to model the inaccuracy caused by the ResNet or the GPR model. In this way, the results estimated by the two models can be regarded as noisy sensor measurements of the digital video microscopy.

The overview of the pose and depth estimations for microrobots based on the GPR-ResNet hybrid model with few-shot calibration is shown in Figure 5. After preprocessing

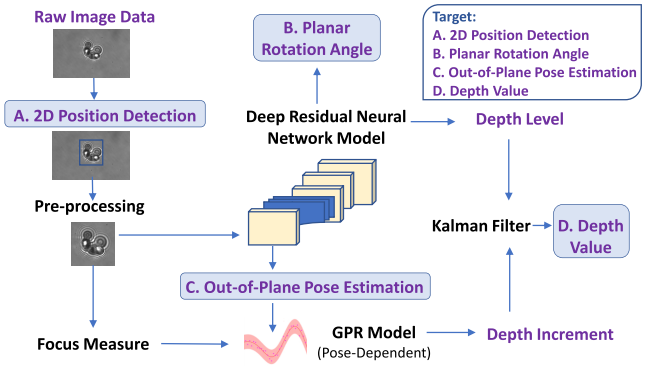


Figure 5. Overview of the pose and depth estimations for microrobots based on the GPR-ResNet hybrid model with few-shot calibration.

of the microrobot with 2D planar position estimation, the focus measurement is conducted, while the out-of-plane pose and the depth level are determined via the ResNet model. Based on the out-of-plane pose estimation results, the specific GPR model is used to generate the depth increment value based on the focus measurement results. The depth level and the depth increment are fed into the Kalman filter, while the final depth value can be generated.

Though deep learning based methods can achieve high classification accuracy for many applications, they need a large data set with diversified data for training. Moreover, the model trained on one microrobot will fail to estimate the pose and depth estimation of another microrobot. Therefore, domain adaption should be considered.

Domain Adaption. The major drawback of a deep network are the long training time and the enormous computational expenses. Therefore, to enable network adaptation, the residual network is trained with a big data set, and transfer learning is used to fine-tune the network for pose estimation for other mobile microrobots with complex shapes. To verify the effectiveness of the domain adaption of the proposed method, further experiments are conducted.

To standardize the training for the same microrobot, an array of microrobots printed at different poses are fabricated to generate the training data set for one microrobot. The raw image data of this microrobot are used for the end-to-end training for pose estimation. If the pose recognition is needed for another microrobot with different shapes, transfer learning can be applied to refine the network for the microrobot without the need to retrain the neural network model from scratch. Since the feature extraction processes are similar for pose estimation, the parameters for the convolutional layers are fixed, while the global average pooling layer and the fully connecting layer can be fine-tuned on other data sets with limited ground-truth data. In this way, the model trained based on the microrobot A can be adapted for microrobot B with relatively small refinements of the network. Similarly, the network model can be tuned to adapt to other microscopic systems with various point spread functions (PSF) using the same transfer learning technique.

EXPERIMENTS

In this section, the microrobot fabrication process and the experimental setup are introduced. The depth estimation results for microrobotic platforms are shown while the pose and depth estimation results for mobile microrobots are described.

Microrobot Fabrication. The Nanoscribe 3D-Printer (Nanoscribe, Germany) is used to fabricate the optical microrobots. Photoresist (Nanoscribe, IP-L 780) was selected as the material for microrobot fabrication. It is biocompatible and dielectric, with a refractive index of 1.52, which is higher than the surrounding medium (the refractive index for water is 1.33).

The two-photon polymerization (2PP) method was used for the manufacturing process,[37] where the resolution is set to 100 nm and realized by a 3D printing system (Nanoscribe GmbH, Germany). The microplatforms and microrobots for experimental validation were printed on the glass substrate and placed in deionized (DI) water within a spacer.[38]

Experimental Setup. In this paper, an Optical Tweezer (Elliot Scientific, U.K.) was used as an example to verity the effectiveness of the proposed method. The proposed method can be implemented in most of the optical systems that involve a microscope for optical microrobot monitoring and manipulation. For our experiments, the microplatforms and microrobots were imaged with a high-speed CCD camera (Basler AG, Germany) and an oil immersion lens microscope (Nikon Ti) with 100× magnification.

https://dx.doi.org/10.1021/acsphotonics.0c00997
ACS Photonics 2020, 7, 3003−3014

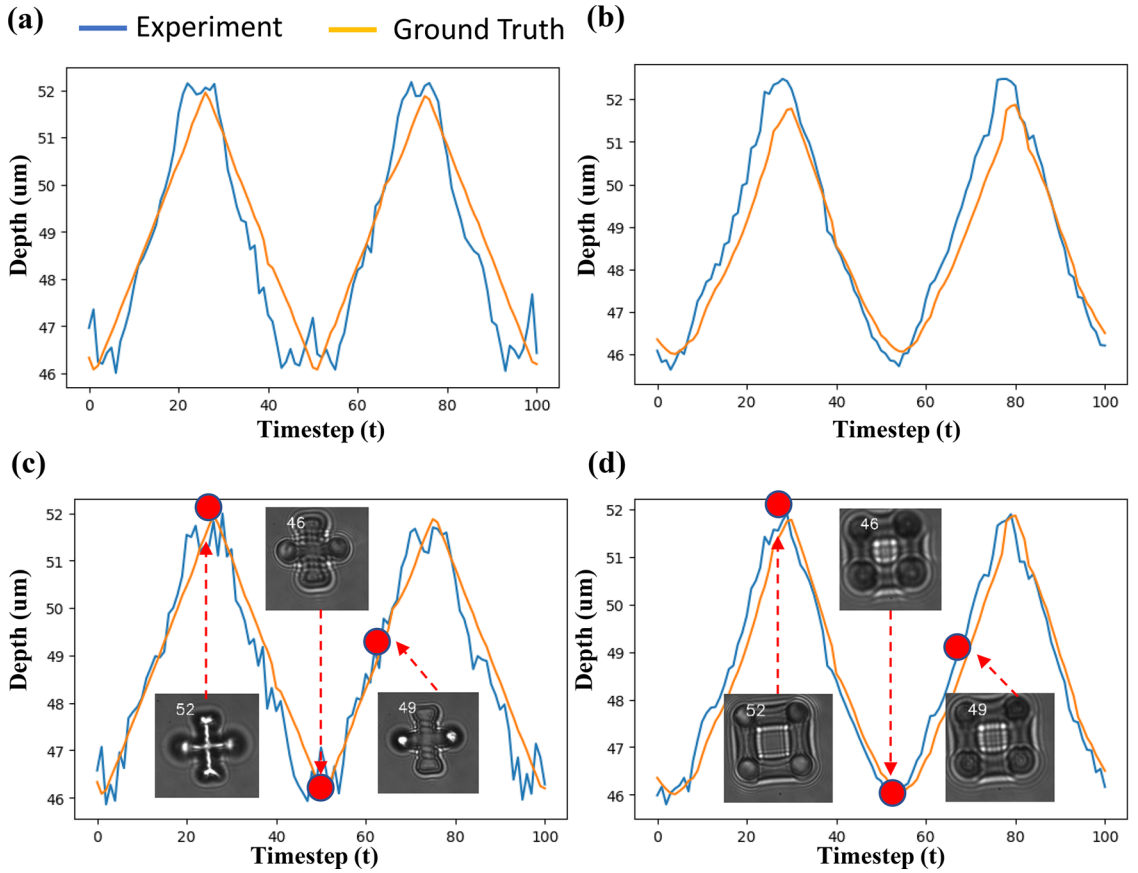


Figure 6. Validation of the GPR-based depth regression in Test 1. (a) Experimental results for depth estimation using two selected focus measurement metrics for microplatform A. (b) Depth estimation results using two selected focus measurement metrics for microplatform B. (c) Experimental results of the depth estimation using four focus measurement metrics for microplatform A. (d) Depth estimation results using four focus measurement metrics for microplatform B.

The printed microrobots are kept fixed on the glass slide and placed on the piezo stage to generate the ground truth trajectories along the z-axis. The microrobots with various poses were translated using the piezo stage in the discrete mode or continuous mode. The sinusoidal, triangular, or random trajectory in the z-axis can be generated with different moving frequencies of the piezo stage, which are regarded as the continuous modes. For one specific pose, at least 1000 frames were collected for each subdata set. The live video of the microrobot was collected and processed offline. After the 2D position detection and estimation, the video data is cropped to the dimension of 256 × 256 pixels.

Results and Analysis. *Depth Estimation for a Microplatform.* The regression model is effective when the focus plane is within a reliable range (smaller than the overall dimension of the microplatform in the z-axis). To verify that the proposed method can estimate the relative depth of the microplatform compared to the focal plane, the translational stage is moved by following a predefined trajectory in the z-axis.

The depth of the 3D-printed microplatforms is estimated by using the GPR model obtained from prior calibration. The Nanopositioner (Mad City Laboratories Inc.) has a working range of 100 μm. During the experiment, the microrobots are moving between the depth level of 46 and 52 μm. Three tests were conducted in total, using three microrobots fabricated at different time but have the same structure and printing parameters. Figure 6 demonstrates the depth estimation results for the two microplatforms respectively during Test 1. The comparisons between the experimental curve and the ground truth data of the other tests can be viewed in Supplement 2 (Figures S2 and S3) of the SI.

The root-mean-square-error (RMSE) is used to measure the accuracy of the regression. Three tests were conducted to verify the effectiveness of implementing the GPR model for depth estimation. RMSE results for the regression of microplatform in all the tests are demonstrated in Table 1. It can be concluded that the regression accuracy can be significantly improved by using more features.

Pose and Depth Estimation for Microrobots. For 3D-printed microrobots of known geometries, a data set can be generated consisting of a set of microscope images and the corresponding 3D positions and orientations of the microrobot. The labeled data is used to train the model, where 20%

Table 1. RMSE Results for the Regression of Microplatforms

2 features	Test 1 (μm)	Test 2 (μm)	Test 3 (μm)	mean ± std
microplatform A	0.739	0.935	0.809	0.828 ± 0.099
microplatform B	0.649	0.555	0.645	0.616 ± 0.053
4 features	Test 1 (μm)	Test 2 (μm)	Test 3 (μm)	mean ± std
microplatform A	0.556	0.728	0.507	0.597 ± 0.116
microplatform B	0.432	0.390	0.462	0.428 ± 0.036

https://dx.doi.org/10.1021/acsphotonics.0c00997
ACS Photonics 2020, 7, 3003−3014

of the total images was used for testing. During the model training process for the other 80% of data, 70% was used for training, while 30% was used for validation to fine-tune the model. Detailed illustration of the 2D position detection process and the single image preprocessing for database construction can be found in Supplement 3 (Figure S5) in the SI.

To reduce the computational time, each image is resized to 50 × 50, while intensity of the image is normalized before feeding the images to the network. For both the training and testing data, the preprocessing of the image is conducted as follows.

$$p(x', y') = \left[p(x, y) - \frac{\sum_{x,y} p(x, y)}{2500} \right] / 128 \tag{12}$$

Data augumentation is used to avoid overfitting during the training process. The batch size is set to be 100. The cross entropy loss is computed while AdamOptimizer is used as the optimizer. After training 30 epoches, the accuracy of pose estimation for the targeted classes can reach 100% for the testing database. The results of the pose estimation are shown in Figure 7, including the top view obtained by the camera and

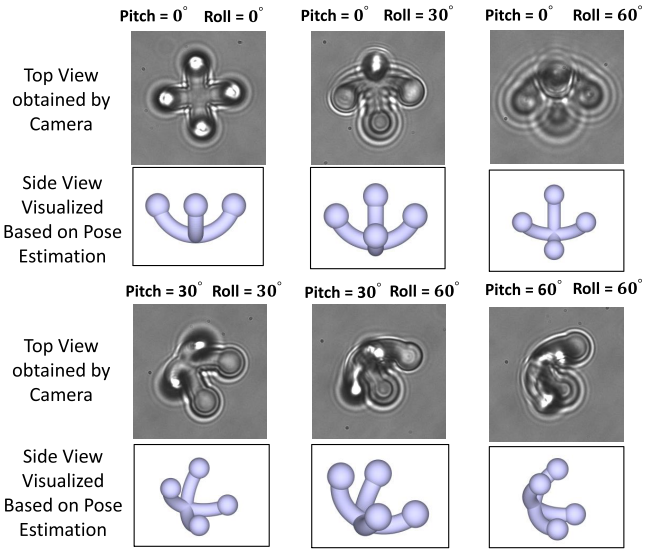


Figure 7. Overview of the results for pose estimation. The top view of the image obtained by microscopic camera is fed to the neural network; the out-of-plane pose can be estimated, while the additional side view can be visualized in the CAD model mode.

the side view augmented by a CAD model, which is generated with the pose estimation results. The same ResNet architecture can be used for the discrete depth level estimation for microrobots. The neural network can work simutaneously to generate the pose and depth estimation results.

The pose estimation accuracy for microrobot A is 99.93%. As for the depth estimation, the accuracy is 97.76% for microrobot A. The depth estimation results of microrobot A with various poses based on the DNN method can be viewed in Figure 8a, the model of which is trained using the discrete data set. The depth estimation results using the ResNet model for microrobot A and the comparisons with the ground-truth data is shown in Figure 8c for continuous depth value estimation.

As for the planar rotation angle estimation, the accuracy is 99.98%, which indicates that the Hierarchy 2 in the ResNet model is effective and the 6D pose estimation for the microrobot can be realized by using the self-generated data. The examples of planar rotation angle estimation are demonstrated in Figure 9. Two indicative sets of feature maps derived from the first two convolutional layers and max pooling layers are shown in Supplement 3 (Figure S6) in the SI to demonstrate the process of feature extractions by neural network.

Hybrid Model with Few-Shot Calibration and Domain Adaption. The model trained using the discrete data is adapted to the new data set with many unseen data. The few-shot calibration is performed at first. Several examples for the focus measurement calibration of microrobot A are demonstrated in Supplement 1 (Figure S1) in the SI, while the experimental results of the GPR model for the depth estimation of microrobot A with different poses can be viewed in Supplement 2 (Figure S4) in the SI. The root-mean-square-error results were calculated, as shown in Supplement 2 (Table S1) in the SI.

Figure 8b shows the results comparison between the ground-truth data and the depth estimation obtained based on GPR model for few-shot calibration of the small data set for microrobot A. Combining the depth estimation results from both the GPR and ResNet with Kalman filter, the depth estimation results using the hybrid model (GPR-ResNet model) for microrobot A and the comparisons with the ground-truth data are shown in Figure 8d.

The quantitative results for the comparisons between the GPR model, ResNet model, and the hybrid model are shown in Table 2. It can be clearly seen that, with the hybrid model, the RMSE for the depth estimation can be improved (0.381 μm), compared with the results generated by the GPR (0.669 μm) and ResNet (0.447 μm) models.

The database for microrobot A is constructed for training the initial network. To accelerate the training speed and enable domain adaptation, transfer learning is used to obtain the pose and depth estimation model for microrobot B. In order to test whether a small data set is enough for the neural network to adapt to the pose estimation for another microrobot, only 20% of the data is used for training. With the small data set, the pose estimation for microrobot B based on transfer learning is 99.93%, while the depth estimation accuracy is 95.47%, while the RMSE is 0.212 μm. The results demonstrate the adaptability of the proposed method for various types of microrobots.

Visualization. With precise pose and depth estimations, the monitoring of the microrobots with visualization can be realized to enable the operators to have a better perception of the microrobotics, as shown in Figure 10. The blue line represents the focal plane. The side view of the microrobot is shown based on the pose estimation results, while the depth value is demonstrated by adjusting the relative position between the model and the blue line. The 3D model can overlay the microscopic images during operation.

An analytical software is used to provide more information to the users when constructing the digital microscopy. An example is shown in Figure 11, while the video for overall testing can be found in the Supporting Information, Video S1. The evaluation metrics are calculated online and scaled down. The value of each metric is visualized in digital form, and the radius of the circle indicates the value. Therefore, the variation

https://dx.doi.org/10.1021/acsphotonics.0c00997
ACS Photonics 2020, 7, 3003−3014

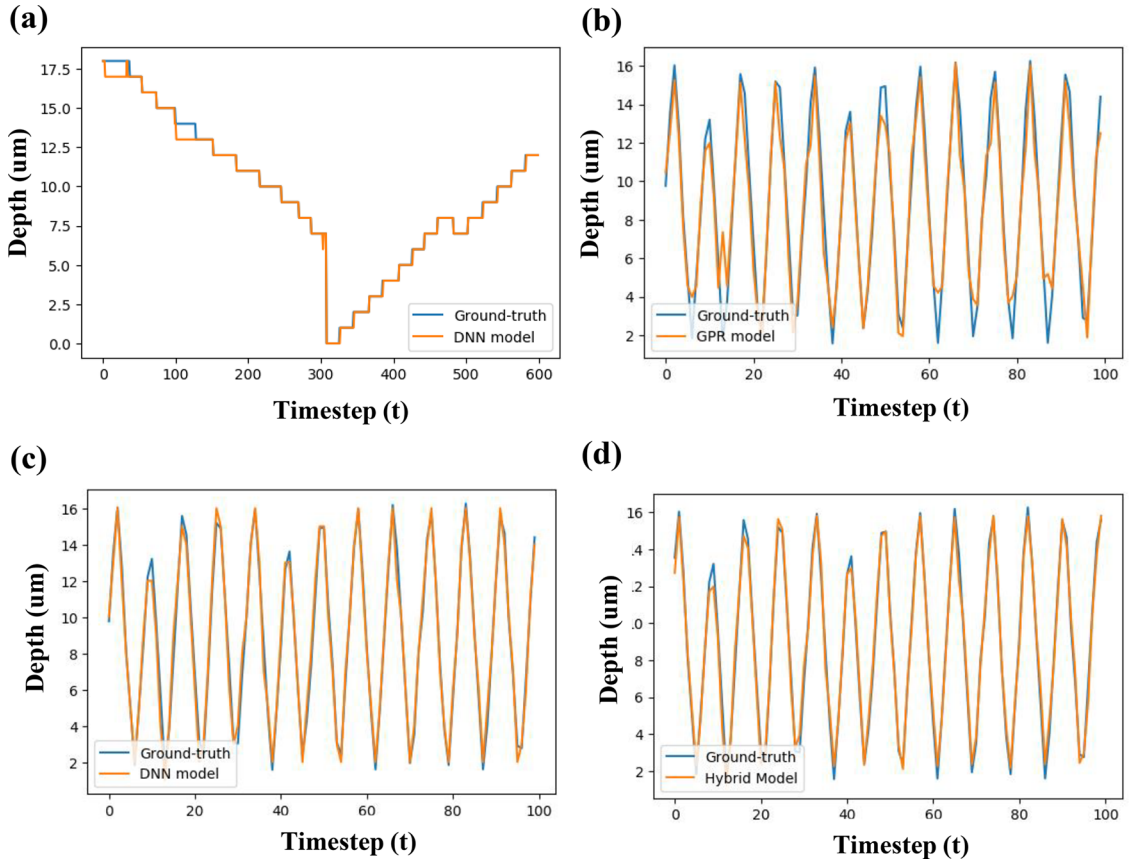


Figure 8. Overview of the pose and depth estimation experimental results for the microrobots. (a) The depth estimation results comparison between the ground-truth data and the ResNet model for the discrete data set. (b) The few-shot calibration results using the GPR model. (c) Depth estimation results using the modified ResNet-18 for microrobot A and the comparisons with the ground-truth data. (d) Depth estimation results using the hybrid model (GPR-ResNet model) for microrobot A and the comparisons with the ground-truth data.

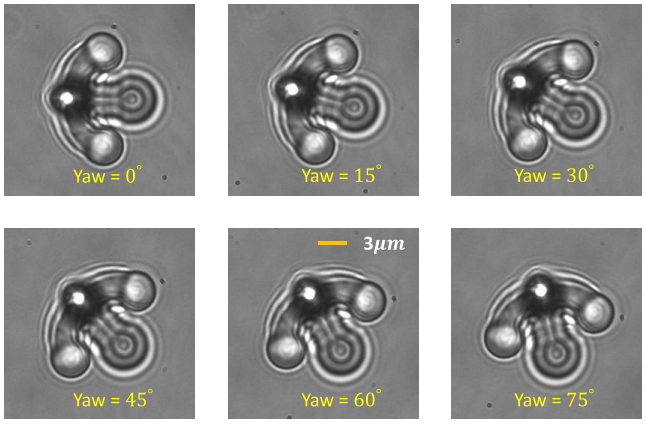


Figure 9. Overview of the planar rotation angle estimation via self-supervised learning based on the ResNet Model.

Table 2. Experimental Results of Domain Adaption for Verification of the Effectiveness of the Proposed Methods with Comparative Studies

	GPR-based model (μm)	ResNet model(μm)	proposed hybrid model (μm)
microrobot A	0.669	0.447	0.381
microrobot B	0.555	0.445	0.316

of the focus measurement can be monitored (see Figure 11b). The 2D entropy image is demonstrated to provide a more

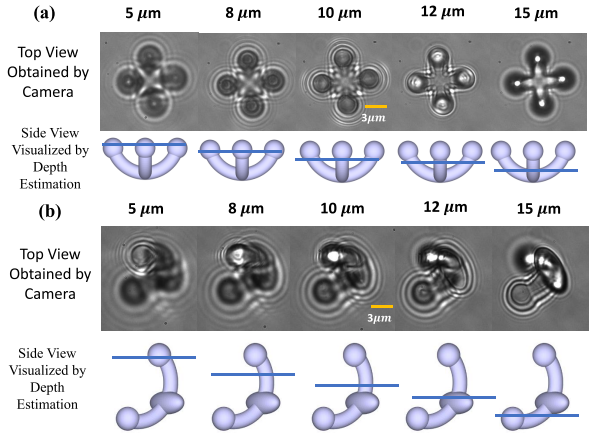


Figure 10. Visualization of microrobot monitoring with various poses and depth values, including the top view obtained by microscopic camera and the side view with CAD model based depth level demonstration.

intuitive visualization of the focus measurement. Figure 11d–f is the visualization of the depth and pose estimations mentioned above.

CONCLUSIONS AND FUTURE WORK

The aim of the proposed methodology is to develop deep learning method for the depth and pose estimation of microplatforms and microrobots, and to enhance the

https://dx.doi.org/10.1021/acsphotonics.0c00997
ACS Photonics 2020, 7, 3003−3014

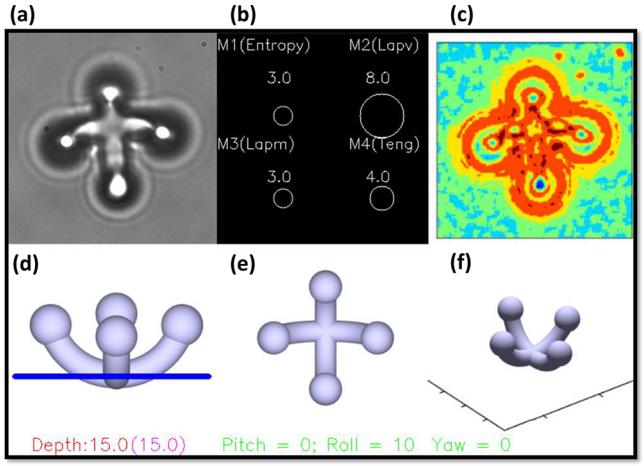


Figure 11. Analytical software is used to provide more information to the users when constructing the digital microscopy. (a) The original figure of the microrobot under microscope. (b) The visualization of focus measurement metrics after proper scaling. (c) The 2D entropy map visualization. (d) The side view of the microrobot with depth estimation results (the unit is μm). (e) The top view of the microrobot with out-of-plane pose estimation results (the unit is degree). (f) The 3D view of the microrobot with planar rotation angle estimation results (the unit is degree).

perception of the operator with visualization of additional views. The depth can be regressed by utilizing multiple focus measurement information via GPR for the microplatforms. A ground truth data set is used to generate the model for depth regression. The effectiveness of the model is tested on two microplatforms for depth estimation, which is significant for the construction of a digital microscope.

As for a microrobot, the microrobot's depth value and out-of-plane and planar rotation angles are estimated with a deep learning method, using a Hierarchical Semi Supervised ResNet architecture. The 6D pose estimation can therefore be implemented. The model is trained based on a database with discrete labels. To reach the goal of continuous depth value estimation, a hybrid model based on GPR and ResNet is developed, while a Kalman filter is used to estimate the depth values smoothly. Transfer learning is tested by fine-tuning the model to realize the pose and depth estimations for another microrobot as validation. Visualization is used to enhance the operator's 3D perception of the microrobots during micromanipulation.

Future work will include verifying the generalizability of the proposed method in different microscopic systems with various point spread functions (PSF) by using the transfer learning techniques for domain adaptation. Moreover, we will extend the method to pose and depth estimations of multiple microrobots simultaneously and demonstrate its usefulness in biomedical applications, such as indirect cell manipulation, which requires depth and pose information to enable multiple microrobots to work cooperatively. With the deep learning method, the 6D pose estimation can be addressed with a high estimation accuracy. However, the black-box effect has been known as one of the major drawbacks for deep learning methods. Explainable AI (XAI) is an emerging field that aims to eliminate the black-box effect and explain how the decisions of AI systems are made, which may be a future direction that is worth exploring.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsphotonics.0c00997.

Video S1 (AVI)

Supporting experimental details and figures (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Dandan Zhang − *Imperial College London, London, United Kingdom*; orcid.org/0000-0001-7649-7605; Phone: +44 07547525412; Email: d.zhang17@imperial.ac.uk

Guang-Zhong Yang − *Shanghai Jiao Tong University, Shanghai, China*; Email: gzyang@sjtu.edu.cn

Authors

Frank P.-W. Lo − *Imperial College London, London, United Kingdom*

Jian-Qing Zheng − *Oxford University, Oxford, United Kingdom*

Wenjia Bai − *Imperial College London, London, United Kingdom*

Benny Lo − *Imperial College London, London, United Kingdom*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsphotonics.0c00997

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors acknowledge funding from the UK Engineering and Physical Sciences Research Council (EPSRC) Program Grant EP/P012779/1 (Microrobotics for Surgery). The authors would like to acknowledge Drs. M. Power and F. Seichepine for the cleanroom training and Drs. M. Grammatikopoulou and A. Barbot for the optical tweezer system training.

■ REFERENCES

(1) Li, J.; de Ávila, B. E.-F.; Gao, W.; Zhang, L.; Wang, J. Micro/nanorobots for biomedicine: Delivery, surgery, sensing, and detoxification. *Science Robotics* **2017**, *2*, eaam6431.

(2) Crocker, J. C.; Grier, D. G. Methods of digital video microscopy for colloidal studies. *J. Colloid Interface Sci.* **1996**, *179*, 298−310.

(3) Jones, P. H.; Maragò, O. M.; Volpe, G. *Optical Tweezers: Principles and Applications*; Cambridge University Press, 2015.

(4) Neuman, K. C.; Nagy, A. Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. *Nat. Methods* **2008**, *5*, 491.

(5) Waigh, T. A. Advances in the microrheology of complex fluids. *Rep. Prog. Phys.* **2016**, *79*, 074601.

(6) Sahoo, S. K.; Misra, R.; Parveen, S. *Nanomedicine in Cancer*; Pan Stanford, 2017; pp 73−124.

(7) Thompson, R. E.; Larson, D. R.; Webb, W. W. Precise nanometer localization analysis for individual fluorescent probes. *Biophys. J.* **2002**, *82*, 2775−2783.

(8) Manley, S.; Gillette, J. M.; Lippincott-Schwartz, J. *Methods in Enzymology*; Elsevier, 2010; Vol. 475; pp 109−120.

(9) Hannel, M. D.; Abdulali, A.; O'Brien, M.; Grier, D. G. Machine-learning techniques for fast and accurate feature localization in holograms of colloidal particles. *Opt. Express* **2018**, *26*, 15221−15231.

(10) Litjens, G.; Kooi, T.; Bejnordi, B. E.; Setio, A. A. A.; Ciompi, F.; Ghafoorian, M.; Van Der Laak, J. A.; Van Ginneken, B.; Sánchez, C. I. A survey on deep learning in medical image analysis. *Med. Image Anal.* **2017**, *42*, 60−88.

(11) Newby, J. M.; Schaefer, A. M.; Lee, P. T.; Forest, M. G.; Lai, S. K. Convolutional neural networks automate detection for tracking of submicron-scale particles in 2D and 3D. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, 9026−9031.
(12) Helgadottir, S.; Argun, A.; Volpe, G. Digital video microscopy enhanced by deep learning. *Optica* **2019**, *6*, 506−513.
(13) Stavroulakis, P.; Leach, R. K. Invited Review Article: Review of post-process optical form metrology for industrial-grade metal additive manufactured parts. *Rev. Sci. Instrum.* **2016**, *87*, 041101.
(14) Grammatikopoulou, M.; Zhang, L.; Yang, G.-Z. Depth estimation of optically transparent laser-driven microrobots. *2017 IEEE/RSJ. International Conference on Intelligent Robots and Systems (IROS)* **2017**, 2994−2999.
(15) Marturi, N.; Tamadazte, B.; Dembélé, S.; Piat, N. Visual servoing-based depth-estimation technique for manipulation inside sem. *IEEE Trans. Instrum. Meas.* **2016**, *65*, 1847−1855.
(16) Cui, L.; Marchand, E.; Haliyo, S.; Régnier, S. Three-dimensional visual tracking and pose estimation in scanning electron microscopes. *2016 IEEE/RSJ. International Conference on Intelligent Robots and Systems (IROS)* **2016**, 5210−5215.
(17) Marturi, N.; Tamadazte, B.; Dembélé, S.; Piat, N. Image-guided nanopositioning scheme for sem. *IEEE Transactions on Automation Science and Engineering* **2018**, *15*, 45−56.
(18) Eichhorn, V.; Fatikow, S.; Wich, T.; Dahmen, C.; Sievers, T.; Andersen, K. N.; Carlson, K.; Bøggild, P. Depth-detection methods for microgripper based CNT manipulation in a scanning electron microscope. *Journal of Micro-Nano Mechatronics* **2008**, *4*, 27−36.
(19) Kudryavtsev, A. V.; Dembélé, S.; Piat, N. Full 3d rotation estimation in scanning electron microscope. *2017 IEEE/RSJ. International Conference on Intelligent Robots and Systems (IROS)* **2017**, 1134−1139.
(20) Bergeles, C.; Kratochvil, B. E.; Nelson, B. J. Visually servoing magnetic intraocular microdevices. *IEEE Transactions on Robotics* **2012**, *28*, 798−809.
(21) Rivenson, Y.; Ceylan Koydemir, H.; Wang, H.; Wei, Z.; Ren, Z.; Gunaydın, H.; Zhang, Y.; Gorocs, Z.; Liang, K.; Tseng, D.; et al. Deep learning enhanced mobile-phone microscopy. *ACS Photonics* **2018**, *5*, 2354−2364.
(22) Wu, Y.; Ray, A.; Wei, Q.; Feizi, A.; Tong, X.; Chen, E.; Luo, Y.; Ozcan, A. Deep learning enables high-throughput analysis of particle-aggregation-based biosensors imaged using holography. *ACS Photonics* **2019**, *6*, 294−301.
(23) Wu, Y.; Calis, A.; Luo, Y.; Chen, C.; Lutton, M.; Rivenson, Y.; Lin, X.; Koydemir, H. C.; Zhang, Y.; Wang, H.; et al. Label-free bioaerosol sensing using mobile microscopy and deep learning. *ACS Photonics* **2018**, *5*, 4617−4627.
(24) Grammatikopoulou, M.; Zhang, L.; Yang, G.-Z. Depth estimation of optically transparent microrobots using convolutional and recurrent neural networks. *2018 IEEE/RSJ. International Conference on Intelligent Robots and Systems (IROS)* **2018**, 4895−4900.
(25) Mahendran, S.; Lu, M. Y.; Ali, H.; Vidal, R. Monocular object orientation estimation using Riemannian regression and classification networks. *arXiv preprint arXiv:1807.07226* **2018**, na.
(26) Grammatikopoulou, M.; Yang, G.-Z. Three-dimensional pose estimation of optically transparent microrobots. *IEEE Robotics and Automation Letters* **2020**, *5*, 72−79.
(27) Qu, Y.; Jing, L.; Shen, Y.; Qiu, M.; Soljacic, M. Migrating knowledge between physical scenarios based on artificial neural networks. *ACS Photonics* **2019**, *6*, 1168−1174.
(28) Sun, Y.; Duthaler, S.; Nelson, B. J. Autofocusing in computer microscopy: Selecting the optimal focus algorithm. *Microsc. Res. Tech.* **2004**, *65*, 139−149.
(29) Pech-Pacheco, J. L.; Cristóbal, G.; Chamorro-Martinez, J.; Fernández-Valdivia, J. Diatom autofocusing in brightfield microscopy: a comparative study. Proceedings 15th International Conference on Pattern Recognition; ICPR-2000, Barcelona, Spain, Sept 3−7, 2000, IEEE, 2000; pp 314−317.
(30) Nayar, S.K.; Nakagawa, Y. Shape from focus. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **1994**, *16*, 824−831.
(31) Williams, C. K. I.; Rasmussen, C. E. Gaussian Processes for Regression. *Advances in Neural Information Processing Systems* **1996**, *8*, na.
(32) Rasmussen, C. E. *Gaussian processes in machine learning. Summer School on Machine Learning* **2004**, *3176*, 63−71.
(33) Schulz, E.; Speekenbrink, M.; Krause, A. A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions. *Journal of Mathematical Psychology* **2018**, *85*, 1−16.
(34) Chang, J.; Wetzstein, G. Deep optics for monocular depth estimation and 3d object detection. *Proceedings of the IEEE International Conference on Computer Vision*, Seoul, South Korea, Oct 27− Nov 3, 2019, IEEE, 2019; pp 10193−10202.
(35) He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition* **2016**, 770−778.
(36) Gidaris, S.; Singh, P.; Komodakis, N. Unsupervised representation learning by predicting image rotations. *arXiv preprint arXiv:1803.07728* **2018**, na.
(37) Kawata, S.; Sun, H.-B.; Tanaka, T.; Takada, K. Finer features for functional microdevices. *Nature* **2001**, *412*, 697.
(38) Zhang, D.; Barbot, A.; Lo, B.; Yang, G. Z. Distributed Force Control for Microrobot Manipulation via Planar Multi-Spot Optical Tweezer. *Adv. Opt. Mater.* **2020**, 2000543.

https://dx.doi.org/10.1021/acsphotonics.0c00997
ACS Photonics 2020, 7, 3003−3014