

# Estimation de la couverture de configurations de numérisation 3D à l'aide de techniques d'apprentissage machine

Tingcheng LI <sup>a</sup>, Arnaud POLETTE <sup>a</sup>, Ruding LOU <sup>b</sup>, Jean-Philippe PERNOT <sup>a\*</sup>, Manon JUBERT <sup>c</sup>, Dominique NOZAIS <sup>c</sup>

<sup>a</sup> Arts et Métiers Institute of Technology, LISPEM, HESAM Université, F-13617 Aix-en-Provence, France

<sup>b</sup> Arts et Métiers Institute of Technology, LISPEM, HESAM Université, F-71100 Chalon-sur-Saône, France

<sup>c</sup> Innovative - Manufacturing and Control, I-MC, F-84120 Aix-en-Provence, France

\* e-mail : jean-philippe.pernot@ensam.eu

## 1. INTRODUCTION

Today, optical sensors are widely used in various measurement scenarios, e.g. on-board laser lidar for autonomous driving, structured light-based scanner for part measurement, and electronic total station for BIM applications. Compared with other techniques, the 3D reconstruction measurement system based on vision (e.g. structured light) have superiority on obtaining much greater number of surface points, efficiency and ease of use [1]. The existing research on scan view planning problem with the known model and vision measurement tools usually uses only the visibility analysis on the CAD model of the part to determine the possible scan surface [1] while the studies ignore the parameters that affect the scan, which makes the possible scanned area and the real reconstructed region different.

Indeed, lots of parameters affect the quality of the acquired point cloud, such as the choice of the scanner poses including its orientation with respect to the surfaces to be scanned, the roughness of the surfaces, the material and so on. Understanding correctly how these parameters influence the quality of one scan is a difficult task because the parameters are coupled, and their influence cannot be summarized concisely by an explicit formula. The research carried out in this paper focuses on the use of structured light-based scanning device, with the objective of identifying how a scanning configuration and its parameters affect the quality of the obtained point clouds.

To assess the quality of the scanning configuration including the pose of the scanner as well as the influencing parameters characterizing the way the part is scanned

(e.g. exposure, light distribution, roughness and materials), three problems should be solved: (1) how to measure the quality of a point cloud for a given configuration; (2) what are the parameters characterizing a configuration and considered as influencing parameters; (3) how to predict the quality of a configuration from its influencing parameters.

For the first question, researchers proposed several indicators to characterize the quality of the scanning results with respect to the adopted configuration. For instance, the measurement performance indicators and the statistical indicators have been proposed and compared on acquisitions from structured light device and from photogrammetry [2]. Based on this work, Li et al [3] summarizes the existing assessment metrics and explored new metrics for a single scan. It was notably found that some indicators (such as density [4] or dispersion [2]) do not vary a lot when the poses change, while the coverage indicator is more sensitive to the change and can thus be considered as a variable to be optimized when looking for an optimal scanning configuration. As a consequence, this work focuses on the a priori prediction of the coverage indicator from various scan configurations, in order to be able to estimate the coverage that will be obtained but before really scanning.

For the 2nd question, previous studies mainly analysed the influencing parameters in terms of the principle of reconstruction. In this paper, the influencing parameter are divided into three groups: (a) influencing factors brought about by the reconstruction method; (b) properties of the measured object; (c) parameters related to the scanning environment. Considering the group (a), two factors are

here considered: (I) the construction of the hardware, including the layout of the cameras, the lens, the resolution of the cameras and the projector for projecting the pattern series and so on; (II) the calibration error. It is noticeable that much effort is put on the later. It experienced the evolution of the camera model, and the development of the 3D calibration objects to 2D calibration objects. For group (b), the properties of measure object consist of the material of parts, processing methods, surface treatment process, and geometric features. The distribution and intensity of the environmental light, and the ambient temperature are considered in group (c).

To deal with the 3rd question with so many coupled influencing parameters, the use of neural network has been investigated in this paper to find a coverage estimator. Neural networks achieved much reputation on 2D tasks in classification and semantic segment, and it has been extended in various contexts now with its reputation that it can fit implicit functions to handle coupled multi-parameter tasks. Convolution neural network is widely used in structured data. LeNet-5 [8] was proposed in 1998 with class convolutional network structure: convolution-pooling-finally connected. AlexNet [9] followed the strategy was validated in the ImageNet competition. Then, VGGNet [10] and GoogleNet [11] were proposed with more convolutional layers but degradation problem was found that the accuracy of the network saturates or even decreases while the network depth increases. Kaiming He et al. [12] designed the residual block to solve this problem. Also, inception block was designed with extracting features in different scale space. Attention mechanism and transformer [13] have been proposed one after another. Our work makes contributions on threefold: (a) a novel 3D reconstruction prediction framework able to evaluate the coverage of the scanning configurations with a deep segmentation network; (b) several physical acquisition platforms built together with the digital twins, able to operate the scanning task with auto-registration; (c) one database including real scan point clouds of parts, scanner configurations, and scan data virtually generated from the CAD model, able to explore the methods of registration, analysis related to point cloud and so on. The work could be used for scan view plan problem on binocular stereo measurement to optimize the pose of the scanner.

The paper is organized as follows. Section 2 presents the overall framework of the work. The experimental validation is detailed in Section 3 and Section 4 ends this paper with conclusion and perspectives.

## 2. METHODOLOGY

This section introduces the acquisition platform for the creation of the database and the novel coverage prediction framework. The former offers the required data with scanning configurations for training the latter.

### 2.1. Acquisition platform

Several equipments have been set up and integrated within the acquisition platform, to perform the scan tasks

required to build the database used for the learning step. It consists of a CNC machine DMU 50 with an integrated probe, a structured light-based scanner GOCATOR 3210 by LMI, an UR5 robot, the PC to control the whole platform, a thermometer and a luminometer. The acquisition platform includes two parts: CNC platform (Figure 1) and robot platform (Figure 2). Each platform take part to the acquisition tasks, and the two platforms can be merged into one semi-automatic platform (Figure 3).

According to Li et al. [3], the coverage of a scan characterizes the way the surface of the geometric model of a part being digitized is covered. This metric is evaluated for each facet  $j$  of the triangle mesh associated to the CAD model of the scanned part. If the number of points  $N_j^{pts}$  of the scanned point cloud associated to the  $j$ -th facet is over a threshold, the facet is considered as *covered* (1). Otherwise, the  $j$ -th facet is not covered and it is considered as a zero (0) facet in this paper. Thus, the computation of this

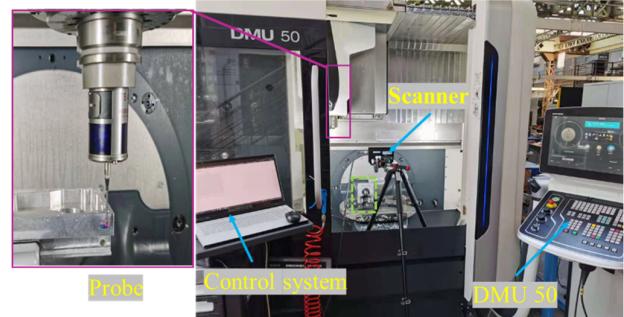


Figure 1 - CNC platform



Figure 2 - Robot platform

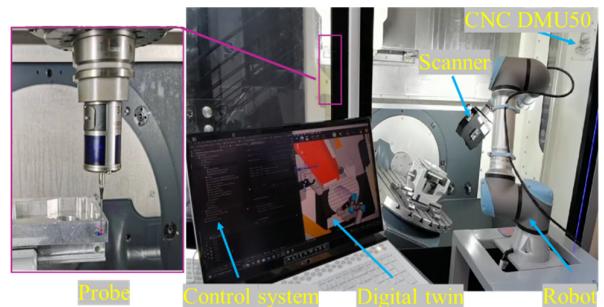


Figure 3 - Semi-automatic acquisition platform

indicator requires the point cloud to be aligned to the associated CAD model. To do so, the digital twin of the entire platform is established to record the position and orientation details, especially the **registration matrices** needed for the calculation of the coverage indicator.

The essence of the digital twin is more related to the coordinate system transformation. For the CNC platform, the matrix  $H_{gt}$  defines the transformation from the global coordinate system (CSYS)  $O_gXYZ$  to the one of the rotating table  $O_tXYZ$  (Figure 4) and its values can be found in the user manual or measured. Then, the transformation matrix  $H_{gp}$  between the workpiece and the global coordinate system can be calibrated by the probe of the CNC machine (Figure 5), and indirectly acquired  $H_{pt}$  by Equation (1). After scanning and doing registration with registration matrix  $H_{cp}$  (point cloud to part transformation), the scanner can be positioned in the digital twin with Equation (2). Finally, the digital twin is set up as in Figure 6. Once the calibration performed, the acquisitions can start, the table is successively rotated, and the acquisitions obtained for each position of the table, each acquisition being then located in the digital twin using the transformation matrices.

$$H_{pt} = H_{pg}H_{gt} \quad (1)$$

$$H_{ct} = H_{pt}H_{cp} \quad (2)$$

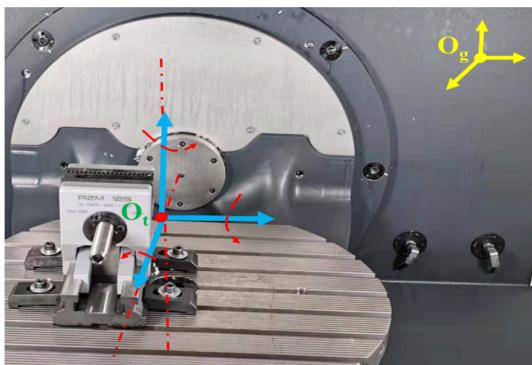


Figure 4 – The rotation table and global coordinate system

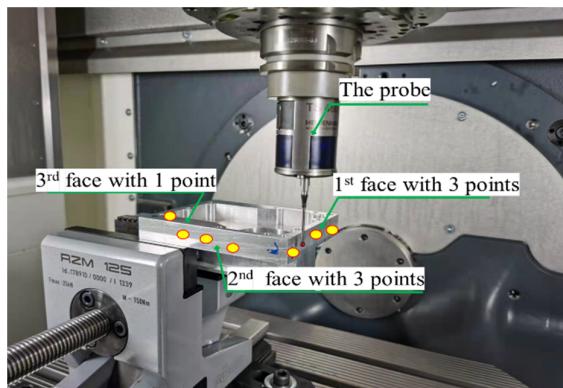


Figure 5 – Position the workpiece in the CNC CSYS

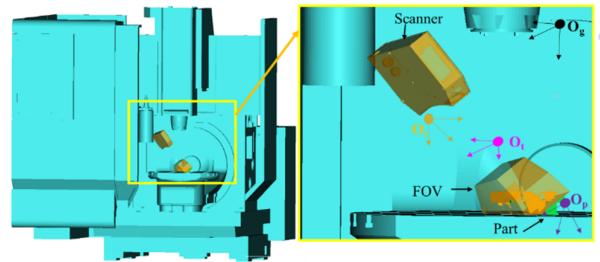


Figure 6 – Digital twin of the CNC acquisition platform

For the robot acquisition platform (Figure 7), it has 4 coordinate systems: robot base coordinate system  $O_bXYZ$ , end coordinate system  $O_eXYZ$ , scanner coordinate system  $O_cXYZ$ , and workpiece coordinate system  $O_pXYZ$ . Firstly, the transformation  $H_{ce}$  (scanner to robot end) can be obtained by eye-in-hand calibration [14]. Similar to the CNC platform, the position and orientation of the part ( $H_{pb}$ ) can be analyzed after one registration of the scan result  $H_{cp}^{(initial)}$  using Equation (3). For each robot pose, the pose of the robot end under  $O_bXYZ$  ( $H_{eb}^{(j)}$ ) is obtained from the robot control panel and the registration matrix for each pose acquisition  $H_{cp}^{(j)}$  is calculated by Equation (4).

$$H_{pb} = H_{eb}^{(initial)} H_{ce} (H_{pc}^{(initial)})^{-1} \quad (3)$$

$$H_{pc}^{(j)} = (H_{pb})^{-1} H_{eb}^{(j)} H_{ce} \quad (4)$$

At the end of the calibration step, all transformation matrices are known and can be used to reposition the numerous scanned point clouds onto the corresponding CAD models and thus be able to build the database automatically labelled with coverage information.

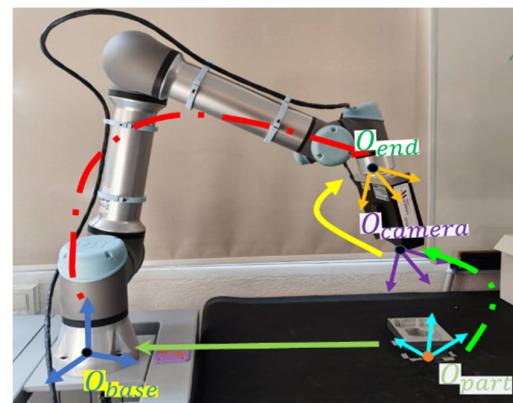


Figure 7 – Position the workpiece in the robot base CSYS

## 2.2. Prediction model construction

Using the previously acquired labelled database, the prediction model can be trained and tested. This subsection briefly details the structure of the model.

In this paper, only a subset of the influencing factors has been considered: parameters of the workpiece, manufacturing aspects through the texture and part of the configuration. Other aspects related to the material, to the

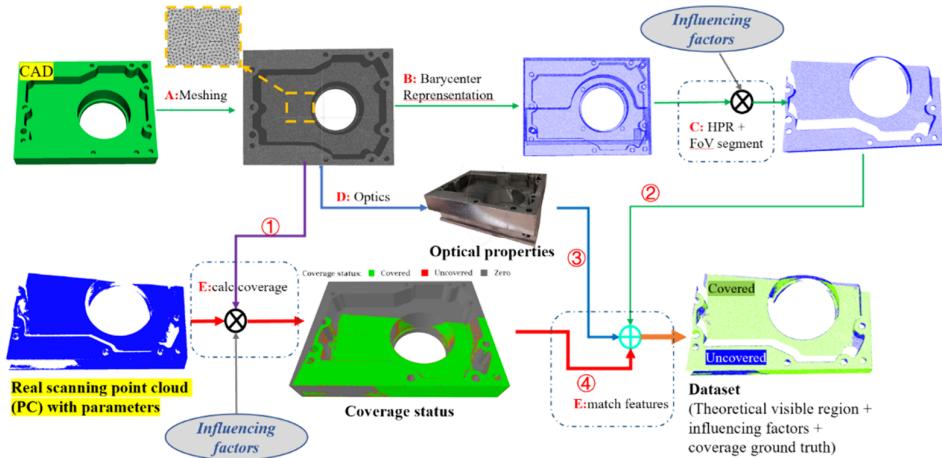


Figure 8 - Flowchart of generating dataset for training

lightning, to the ambient temperature and so on have not been considered in this study. The flowchart used to create the dataset is shown in Figure 8, wherein one labelled sample of the dataset (bottom right) is created from a real scan (bottom left) and from the CAD model (top left). The standardized mesh is obtained after CAD meshing (step A) while controlling the length of the edges considering the resolution of the scanner. In this work, the point cloud representation is chosen for its convenience, so barycenter is used instead of the facet (step B). After compressing the representation, hidden point removal algorithm and field of view (FOV) is used to segment redundant points and thus obtain the theoretical point cloud in Step C. The visual characteristics of the real workpiece are also considered using textures in Step D. In step E, the coverage is calculated with the real point cloud, transformation matrices, and the standardized mesh. It should be pointed out that the mesh is translated into the scanner coordinate system. Finally, all the features are merged into the descriptors and label for each facet/barycenter. The descriptors used in this paper are shown in Figure 9, where the count of values of each feature are 3, 3, 43 respectively.



Figure 9 - Elements of one descriptor for training

The architecture of the developed prediction model is based on U-Net and can be divided into 2 parts: encoder (block B of Figure 10) and decoder (block C of Figure 10). In the encoder, the input of each layer will successively be applied with a *channel attention mechanism*, and then a *convolution operation*, and next it will be merged with the original input into a new feature map. This process is described as the block A of Figure 10. Finally, the result of previous layer will take pooling operation to get a new feature as the input of the next layer. One layer of the model includes the one-layer encoder and one-layer decoder, shown as the block D. The input of the decoder consists of the up-sampling result of the previous layer and the output of the encoder in the same model layer. Each layer of the decoder does like encoder, applying the block

A and convolution. The up-sampling is done to prepare the input for the next layer.

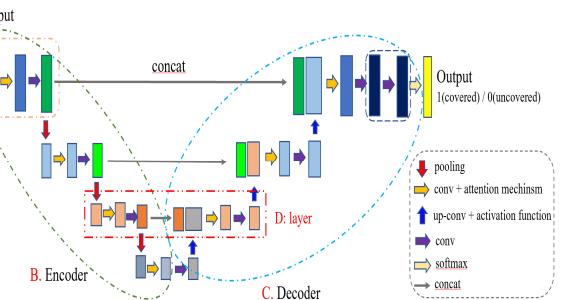


Figure 10 - Coverage prediction model

### 3. EXPERIMENTS AND RESULTS

To populate the database, several workpieces have been designed, manufactured and then scanned according to multiple poses and scanner settings. In this paper, two parts have been considered: pocket and stair-like (Figure 11) and the exposure is set to 6000 $\mu$ s. The multiple poses are visible in Figure 12 for the pocket workpiece.

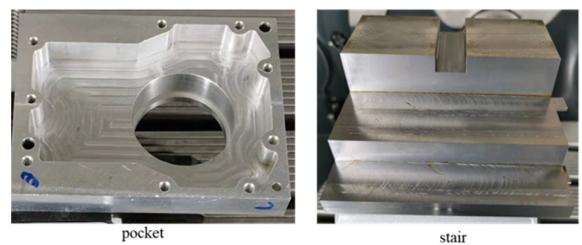


Figure 11 - Considered workpieces

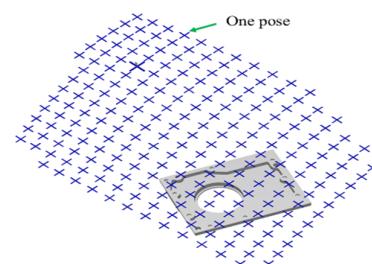
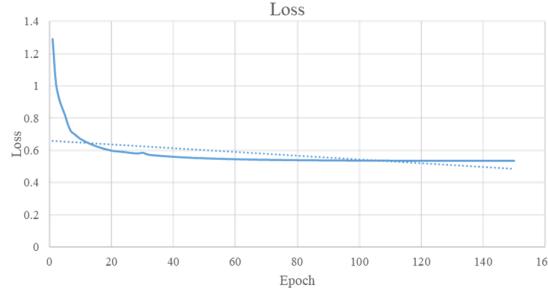


Figure 12 - Scanner poses for the pocket workpiece

The neural network inputs are prepared following the dataset generation steps in Section 2. To train the network, a subset of the acquisitions on **the pocket** are used and the remaining ones are used for testing. The stochastic gradient descent (SGD) with momentum is taken as optimization strategy to optimize the cross entropy as the loss function. The evolution of the loss function is shown in Figure 13.



**Figure 13 – Evolution of the loss function when training the network**

After 150 epochs, the loss tends to be stable and the training ends. The model is then tested on the remaining acquisitions from the pocket, as well as on the acquisitions from the stair-like part that have never been used for training. *F1-score* is used to evaluate the training of the model, defined as Equation (5). *Precision* is defined as the percentage of the real covered triangles among the whole predicted covered triangles while *Recall* is defined as the percentage of the real covered triangles among the whole scanning covered triangles.

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (5)$$

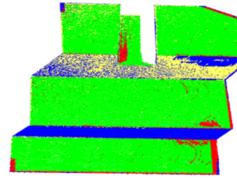
The test results are shown in Table 1 and one example of the pocket and stair-like are shown in Figure 14, where the *green* region means the region that model and ground truth both take it as *covered* (1), *blue* corresponding to the model and ground truth *zero* (0), *red* associated to one ground truth *zero* but model *covered*, and the *yellow* determining ground truth *covered* but model *zero*. For instance, on Test4, 93687 triangles that have been *covered* when scanning the stair-like are identified as *covered* by the network, and 23705 triangles that have not been *covered* are indeed identified as *zero* by the network, which is good. In yellow, the 18152 triangles that were *covered* but for which the network did not predict it properly, and in red the 6213 triangles that were not *covered* but which have been identified as *covered* by the system. This last case is certainly the worst case when considering the overall objective of optimizing the scanner position before scanning the manufactured part. Indeed, if the optimizer uses triangles estimated to be *covered* and which are not after the scan, there is a risk of not obtaining a good scan in reality. However, red triangles are few, and these results validate the model and its capacity to predict *a priori* the coverage, i.e. before scanning.

With the help of our work, the proposed large number of scanning configurations (including poses and exposure)

can be assessed in advance in the digital twin and optimized to obtain maximum coverage as few acquisitions with optimal configurations as possible.

**Tableau 1 – Results obtained on five test cases**

Pose	Accuracy	Precision	Recall	<i>F1-score</i>
Test1	0.8560	0.8167	0.8818	0.8480
Test2	0.9052	0.8979	0.8905	0.8942
Test3	0.8920	0.9001	0.9121	0.9060
Test4	0.8281	0.9378	0.8377	0.8849
Test5	0.8823	0.8157	0.9409	0.8738

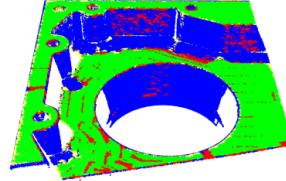


Test4	
Model	GT
Covered	Covered
93687	6213

Test5	
Model	GT
Covered	Covered
29005	6555

(a) Result of Test4 on the stair-like



(b) Result of Test5 on the pocket

**Figure 14 – Testing examples: stair-like (a) and pocket (b)**

#### 4. CONCLUSION

This work firstly analyses the influencing factors that have effects on scanning and the evaluation metrics that can be used to optimize the scanning. Based on this, the data collection platforms set up to prepare the data for the neural network have been presented. Then, the model for coverage prediction has been proposed and trained on two datasets. In the end, the method is validated, and it accurately predicts the coverage in advance of the scanning. The next steps concern the integration of this prediction model within and optimizing loop so as to identify the optimal poses and parameters of the scanner before scanning. The way light reflects on the object when scanning will also be studied, and specific features are to be identified through texture mapping.

#### ACKNOWLEDGMENTS

This work was supported by the China Scholarship Council (No. 202006830012) and the Innovative-Manufacturing and Control (I-MC) company.

#### REFERENCES

- [1] Peuzin-Jubert, M., Polette, A., Nozais, D., Mari, J. L., & Pernot, J. P. (2021). Survey on the view planning problem for reverse engineering and automated control applications. *Computer-Aided Design*, 141, 103094.
- [2] Catalucci, S., Senin, N., Sims-Waterhouse, D., Ziegelmeier, S., Piano, S., & Leach, R. (2020). Measurement of complex freeform additively manufactured parts by structured light and photogrammetry. *Measurement*, 164, 108081.
- [3] Li, T., Lou, R., Polette, A., Shao, Z., Nozais, D., & Pernot, J. P. (2022). On the Use of Quality Metrics to Characterize Structured Light-based Point Cloud Acquisitions. *Computer-Aided Design and Applications*, 20(6), 1190-1203.
- [4] Lartigue, C., Contreras, A., & Bourdet, P. (2002). Digitised point quality in relation with point exploitation. *Measurement*, 32(3), 193-203.
- [5] Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5), 359-366.
- [6] Kratsios, A., & Bilokopytov, I. (2020). Non-euclidean universal approximation. *Advances in Neural Information Processing Systems*, 33, 10635-10646.
- [7] Tabuada, P., & Gharesifard, B. (2022). Universal approximation power of deep residual neural networks through the lens of control. *IEEE Transactions on Automatic Control*.  
[8] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278-2324.
- [9] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84-90.
- [10] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- [11] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2818-2826.
- [12] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770-778.
- [13] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.
- [14] Shah, M. (2013). Solving the robot-world/hand-eye calibration problem using the Kronecker product. *Journal of Mechanisms and Robotics*, 5(3), 031007.