



**HAL**  
open science

# Visual inspection of complex mechanical assemblies based on Siamese networks for 3D point clouds

Velibor Dosljak, Igor Jovančević, Jean-José Orteu

► **To cite this version:**

Velibor Dosljak, Igor Jovančević, Jean-José Orteu. Visual inspection of complex mechanical assemblies based on Siamese networks for 3D point clouds. QCAV'2023 - the 16th international conference quality control by artificial vision, Jun 2023, Albi, France. 5 p., 10.1117/12.2692751 . hal-04168217

**HAL Id: hal-04168217**

**<https://imt-mines-albi.hal.science/hal-04168217>**

Submitted on 21 Jul 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Visual inspection of complex mechanical assemblies based on Siamese networks for 3D point clouds

Velibor Došljak<sup>a</sup>, Igor Jovančević<sup>a</sup>, and Jean-José Orteu<sup>b</sup>

<sup>a</sup>Faculty of Natural Sciences and Mathematics, University of Montenegro, Cetinjska 2, 81000 Podgorica, Montenegro

<sup>b</sup>Institut Clément Ader (ICA) ; Université de Toulouse ; CNRS, IMT Mines Albi, INSA, UPS, ISAE ; Campus Jarlard, 81013 Albi, France

## ABSTRACT

This paper proposes a solution for the problem of visual mechanical assembly inspection by processing point cloud data acquired via a 3D scanner. The approach is based on deep Siamese neural networks for 3D point clouds. To overcome the requirement for a large amount of labeled training data, only synthetically generated data is used for training and validation. Real-acquired point clouds are used only in testing phase.

## 1. INTRODUCTION

Quality control is one of the essential parts of nowadays industry, both in production and maintenance. Much effort is being invested into automatizing this phase of the process. We developed a system for automatic visual conformity inspection of mechanical assemblies by processing data acquired by 3D sensors and also using Computer Aided Design (CAD) model of the assembly, as a reference. The goal is to detect if an expected mechanical part is present in the assembly, and also well mounted in a position predefined by the CAD specification. This decision is made by processing a 3D point cloud obtained by scanning the region of interest on the assembly.

We consider two types of detection problems: detecting when a mechanical part is not mounted (denoted absence/presence problem) and detecting when a wrong type of mechanical part is mounted (denoted multi-class problem). The multi-class problem was successfully solved in previous work<sup>1</sup> using the PointCNN classifier.<sup>2</sup> Therefore, in this work, the main goal is to detect if a part is mounted or not, i.e. to solve the absence/presence problem.

## 2. PREVIOUS WORK

In<sup>1</sup> authors have compared the results of different classification architectures in inspecting complex mechanical assemblies. The dataset used in<sup>1</sup> is the same as the one used in our experiments. There is a difference in that the authors in<sup>1</sup> have trained the model on 61 types (classes) of mechanical parts and then selected the best 20 types, while we select 9 types of mechanical parts and use only them for training, validation, and testing.

In<sup>3</sup> authors have proposed a method for defect segmentation on concrete surfaces based on the point cloud data. They used a network based on *PointNet++*<sup>4</sup> for segmentation. In addition to the point coordinates, they exploited colour and surface normal vectors. The main reason for adding normals and colour is the assumption that crack regions will have significant differences in normal vectors, and the colour will be darker.

In<sup>5</sup> authors have proposed a method based on a siamese neural network with an attention mechanism for defect detection in 2D images. As a loss metric, they used the Pair Balance Contrastive Loss function. The approach was tested on multiple datasets, made of images of capsules, carpets, screws, etc.. Obtained results are quite promising. This paper has served as an inspiration for the work presented in this paper.

---

Further author information: (Send correspondence to dosljakvelibor@gmail.com)

### 3. SYNTHETIC DATA GENERATION

All our conclusions are based on experiments with data from an industrial use case - a visual inspection of a mechanical assembly from the aerospace industry, during production. Examples of the CAD models of the mechanical parts of interest are given in figure 1. All the tested parts belong to the type "metallic support".

Training deep neural networks requires a large amount of labelled data, which is hard to obtain. There are augmentation approaches to increase the training dataset's size or apply different transfer learning techniques to gain satisfying results on the new case with a relatively small number of acquisitions in the training dataset. However, such techniques are not readily applicable in some cases. For example, annotating even a few point clouds for the segmentation task can be a tedious job. Also, performing many acquisition campaigns in large factories is not easily feasible, i.e. it is not practical to stop production to acquire a lot of data. For these reasons, our goal was to completely omit the necessity for collecting and annotating data and create a quality control system using just the CAD specification. Such synthetically generated point clouds can be automatically annotated without any human intervention. On the darker side, using synthetically generated data will inevitably increase the discrepancy between training and test data.

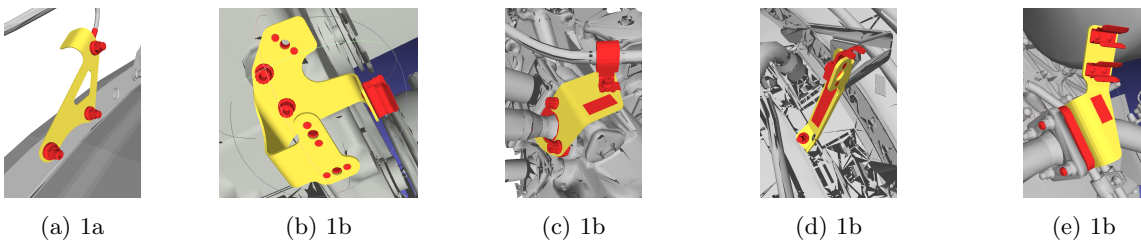


Figure 1: Examples of CAD models of the parts ("supports") whose presence we aim to verify.

This paper proposes a pipeline (figure 2) for generating realistic synthetic 3D data from CAD models. To generate synthetic point clouds, we uniformly sample the CAD models of supports. To make this synthetic cloud similar to one-shot cloud from the 3D sensor, we keep only the points visible from a fixed virtual camera position. To do that, we use the method<sup>6</sup> implemented in the Open3D library. Further, we crop the point cloud using an oriented bounding box (OOB) around the support. OOB is calculated from CAD and using a relative pose estimated by an available localization module. Finally, we downsample the cloud to 2048 points using *Farthest Point Sampling(FPS)* algorithm since this is the number of points required by the neural network architecture we used.

To augment the generated dataset, we employed different parameters for removing auto-occluded points, to mimic different sensor resolutions. Further, we added some random rotation and translation to the camera pose as well. We also added random noise to the bounding box position to emulate the localization algorithm error.

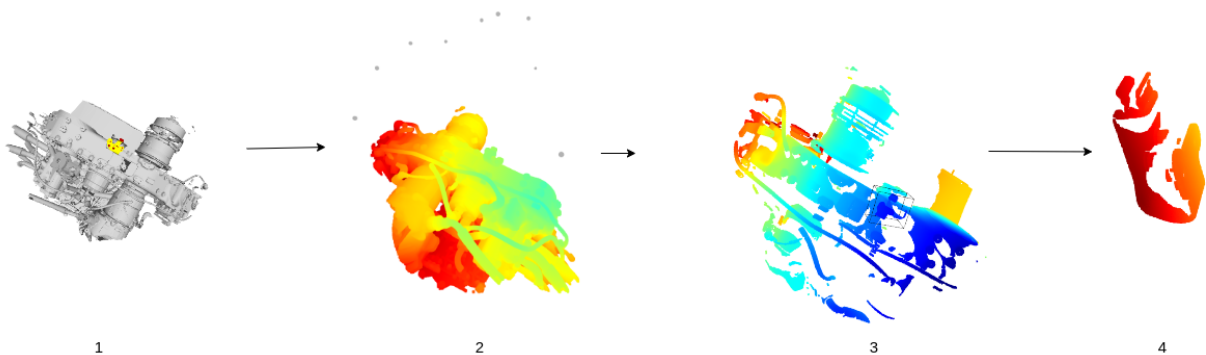


Figure 2: Synthetic data generation process: 1. CAD representation of the whole assembly 2. Point cloud sampled from the CAD model with predefined camera poses (grey points) 3. Point cloud after hidden points removal with marked crop area 4. Final synthetic cloud after cropping.

## 4. ABSENCE/PRESENCE PROBLEM

In our use case, not raising an alarm for a defective sample is more dangerous than declaring a non-defective sample as defective. Missing support is the most frequent type of problem.

In the following experiments, we will demonstrate our approach to the binary absence/presence problem by employing Siamese neural networks. As a handy feature of our solution, we enhance the ability to be applied on new mechanical assemblies, without any additional training.

We will further compare the results obtained using the classification approach with the results obtained using the Siamese architecture described in the following paragraphs.

We argue that it is easier to compare if two point clouds represent the same object, than to learn what each of them represents. Therefore, we compare the point cloud from the scanner with an expected point cloud generated from the CAD model. For each type of support, we will store a reference - one synthetically generated point cloud to be compared with many real point clouds. The set of these referencing point clouds will be called the ground truth set.

The architecture of our model is given in figure 3. The network expects two point clouds as input. We obtain the feature vector from each point cloud using PointNet<sup>7</sup>-like architecture. After each convolution layer, we apply Batch normalization and the ReLU activation function. By applying *maxpool* function, we obtain a single global feature vector of length 1024 for both of input point clouds. This vector is further passed through a fully connected layer with 256-dimensional output. Then, we compute the distance vector as the squared distance of the corresponding coordinates. We pass obtained distance vector through two fully connected layers to obtain the probability that two input point clouds represent the same scene.

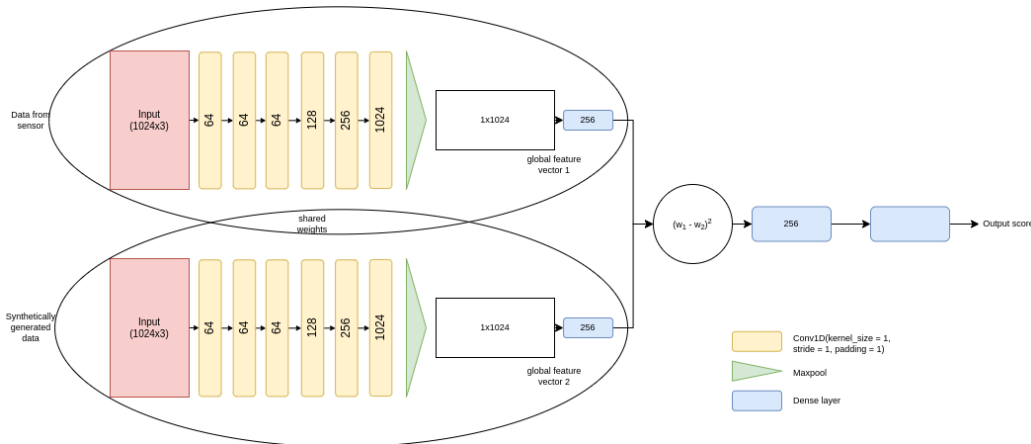


Figure 3: Architecture of the Siamese neural network that we use.

## 5. EXPERIMENTS AND RESULTS

It should be noted that all the evaluations in all the experiments are done purely on real data.

**Experiment 0 (baseline classification networks).** At first, we trained PointCNN classifier on synthetic data only and tested on real data. We did this in order to establish the baseline performance on the absence/presence problem, before testing our Siamese based approach in Experiment 1. We considered different setups for this problem just as did the authors in,<sup>1</sup> where we considered all absence samples as one class and all presence samples as the other class, both during the training and the testing phase (tab. 1 row 1). The second setup groups the absence samples of one support into one class during the training phase. Also, this setup groups the presence samples of one support into one class during the training phase. Further, this setup merges all absence samples into one class and all presence samples into another class at the testing phase, when interpreting the result of the classifier (tab. 1 row 2). These two models will be used as the baseline models for further comparisons.

**Experiment 1.** After testing Siamease architecture on the 9 types of supports we obtained results presented in the table 1 row 3. As we can see, our approach significantly improves results compared to the classification approach (table 1 rows 1, 2).

**Experiment 2.** In order to evaluate the generalization capabilities of the network, we added the models of two unknown types of supports to the ground truth set. We still train with the initial training set containing only 9 types of supports.

The results are shown in tab. 1 row 4. As it appears, the Siamese neural network can be successfully applied to the new use cases by adding new ground truth examples to the ground truth set without any additional training. This considerably simplifies its application and makes this technology suitable to meet industrial needs. It is important to note that supports we added are with relatively clear viewpoints, without high occlusions, which causes a smaller domain gap between real data and the synthetic data. Images of the CAD models of two new classes are given in figure 4.

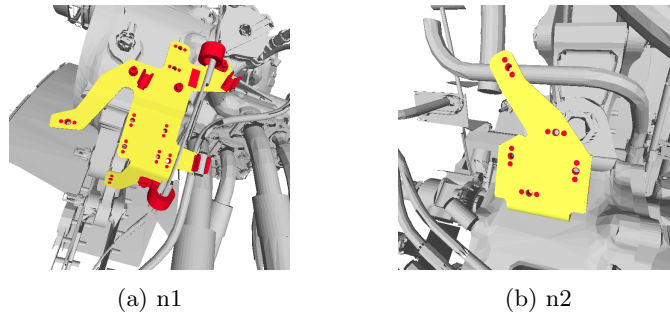


Figure 4: CAD models of two unknown supports used only for evaluating generalization capabilities of the network.

**Experiment 3.** We only trained the network on the public *ModelNet40* dataset and generated samples for the ground truth set from our CAD models without any further training on our data (tab. 1 row 5). It is important to note that there is a significant domain gap between our dataset and *ModelNet40* dataset. While our data are industrial metallic parts with diverse shapes, *ModelNet40* dataset contains everyday objects such as chairs, tables, lamps, etc. Nevertheless, our model manages to learn some useful information.

Data	Model		Precision	Recall	F1	No.cases
Train 9 classes synthetic data. Test 9 classes real data.	PointCNN	Absence	0.89	<b>0.61</b>	0.72	28
	binary	Presence	0.96	0.99	0.97	241
	PointCNN	Absence	0.88	<b>0.75</b>	0.81	28
	multiclass	Presence	0.97	0.99	0.98	241
	Siamease	Absence	0.93	<b>0.89</b>	0.91	28
	nn	Presence	0.99	0.99	0.99	241
Train 9 classes synthetic data. Test 11 classes real data.	Siamease nn	Absence	0.91	<b>0.88</b>	0.89	33
		Presence	0.99	0.99	0.99	404
Train ModelNet40 dataset. Test 9 classes real data.	Siamease nn	Absence	1.00	0.36	0.53	28
		Presence	0.93	1.00	0.96	269

Table 1: Results obtained in our experiments on the presence/absence problem.

## 6. CONCLUSION AND FUTURE WORK

Our work is oriented towards applying deep Siamese neural networks to an industrial inspection problem named conformity verification. The goal is to check that a mechanical assembly conforms to its predefined CAD model. We define the mentioned task as a binary absence/presence problem.

Furthermore, we propose a method for generating realistic synthetic data. We use only the synthetic data for training the models and we evaluate them on real data. The proposed method for solving the absence/presence problem appears more robust to the domain gap than previously used classification approaches.

We plan to further experiment with models that pay more attention to the features shared between domains (such as the geometry of the supports in our case) by using classification and attention-based architectures for feature extraction, to reduce the domain gap further.

## ACKNOWLEDGMENTS

The authors express their gratitude to the DIOTA company, for providing the data and their industrial insight, both valuable for this research work. DIOTA has its Head Office at 201 Pierre and Marie Curie Street, 31670 Labège, France. Authors would like to thank the French “Région Occitanie” for its financial support to the joint research laboratory ”Inspection 4.0” between IMT Mines Albi/ICA and the DIOTA company.

## REFERENCES

- [1] Boughrara, A., Jovancevic, I., Abdallah, H. B., Dolives, B., Belloc, M., and Orteu, J.-J., “Inspection of mechanical assemblies based on 3D deep learning approaches,” in [*Fifteenth International Conference on Quality Control by Artificial Vision*], Komuro, T. and Shimizu, T., eds., **11794**, 1179407, International Society for Optics and Photonics, SPIE (2021).
- [2] Li, Y., Bu, R., Sun, M., Wu, W., Di, X., and Chen, B., “Pointcnn: Convolution on  $\mathcal{X}$ -transformed points,” (2018).
- [3] Bolourian, N., Nasrollahi, M., Bahreini, F., and Hammad, A., “Point cloud-based concrete surface defect semantic segmentation using modified pointnet++,” *Proceedings of the 29th EG-ICE International Workshop on Intelligent Computing in Engineering* (2022).
- [4] Qi, C. R., Yi, L., Su, H., and Guibas, L. J., “Pointnet++: Deep hierarchical feature learning on point sets in a metric space,” (2017).
- [5] Takimoto, H., Seki, J., Situju, S., and Kanagawa, A., “Anomaly detection using siamese network with attention mechanism for few-shot learning,” *Applied Artificial Intelligence* **36** (12 2022).
- [6] Katz, S., Tal, A., and Basri, R., “Direct visibility of point sets,” *ACM Trans. Graph.* *26*, 3, Article 24 (July 2007). [http://www.weizmann.ac.il/math/ronen/sites/math.ronen/files/uploads/katz\\_tal\\_basri\\_-\\_direct\\_visibility\\_of\\_point\\_sets.pdf](http://www.weizmann.ac.il/math/ronen/sites/math.ronen/files/uploads/katz_tal_basri_-_direct_visibility_of_point_sets.pdf).
- [7] Qi, C. R., Su, H., Mo, K., and Guibas, L. J., “Pointnet: Deep learning on point sets for 3d classification and segmentation,” (2017).