Simposio CEA de Robótica, Bioingeniería, Visión por Computador y Automática Marina 2026

RBVM
Simposio CEA de Robótica,
Bioingeniería,
Visión por Computador
y Automática Marina

Sistema de visión para la identificación, el control y la gestión de existencias de subconjuntos simples en entornos de fabricación desorganizados de la industria naval.

Arcano-Bea, P.[a,*], Díaz-Labrador, A.[a], Vidal-Bralo, A.[a], Zayas-Gato, F.[a,*], Quintián, H.[a], Calvo-Rolle, J.L.[a]

[a] *University of A Coruña, Department of Industrial Engineering, CTC, CITIC, Ferrol, A Coruña, Spain.*

Resumen

El seguimiento eficiente del inventario es esencial en la fabricación industrial, donde superar las restricciones espaciales e identificar componentes no estructurados es fundamental para la programación dinámica de la producción. Este estudio investiga la aplicación del aprendizaje profundo en nubes de puntos, utilizando una arquitectura PointNet++ para clasificar diferentes variantes de bases de subconjuntos. Evaluamos el rendimiento del modelo utilizando cuatro resoluciones espaciales con el fin de valorar la relación entre la geometría y la eficiencia computacional. Utilizamos la prueba de Kruskal-Wallis para analizar los resultados y determinar el impacto del submuestreo de puntos en la clasificación. Nuestros hallazgos no mostraron diferencias estadísticamente significativas en el rendimiento entre las densidades evaluadas. Sugiriendo que una resolución de 1024 puntos basta para diferenciar piezas fiablemente, reduciendo considerablemente la carga computacional. Sugiriendo así que estos modelos de visión 3D podrían capturar datos de entornos desorganizados, aportando una base para el seguimiento de inventario en tiempo real y la secuenciación dinámica en fábricas modernas.

Palabras clave: Nubes de puntos 3D, Aprendizaje profundo, Sistemas de manufactura inteligente, Planificación y control de la producción, Robótica, Inteligencia Artificial.

Vision system for simple sub-assembly identification, control and stock management in unorganized shipbuilding manufacturing environments.

Abstract

Efficient inventory tracking is essential in industrial manufacturing, where overcoming severe spatial constraints and identifying unstructured components is essential to dynamic production scheduling.. This study investigates the application of deep learning applied to 3D point clouds, specifically using a tailored PointNet++ architecture, to classify different variants of sub-assembly base plates. We evaluated the performance of the network using four spatial resolutions, to assess the trade off between geometric detail and computational efficiency.We used the Kruskal–Wallis test to statistically analyse the results and determine the impact of point subsampling on classification accuracy. . Our findings showed no statistically significant differences in performance across the evaluated point densities. These results suggest that a resolution of 1,024 points is sufficient for reliable part differentiation, reducing computational overhead considerably. The results suggest that optimized 3D vision models could be used to capture data from unorganized manufacturing environments, providing a basis for real time automated inventory tracking and dynamic assembly sequencing in modern factories.

Keywords: 3D Point Clouds, Deep Learning, Intelligent Manufacturing Systems, Production Planning and Control, Robotics, Artificial Intelligence.

1. Introduction

The transition toward an Intelligent Industry has fundamentally altered the landscape of modern manufacturing, shifting factories from rigid assembly lines to highly responsive, data driven ecosystems. At the core of this transformation is the integration of cyber physical systems, widespread sensor deployment, and advanced analytics, which collectively empower facilities to monitor operations continuously and react swiftly to

*Autor para correspondencia: paula.arcano@udc.es, f.zayas.gato@udc.es

emerging disruptions (Zhong et al., 2017). Within these modernized plants, Manufacturing Execution Systems (MES) act as the central nervous system, bridging high level planning strategies with shop floor execution to maintain transparency and operational agility (Mantravadi and Møller, 2019).

Despite the widespread adoption of digital tools, physical constraints remain a significant challenge, particularly in high-mix manufacturing environments(Gan et al., 2023). Floor space is inherently limited, making the traditional approach of storing every specific part variant in dedicated, rigidly organized inventory buffers highly inefficient. To maximize their footprint and maintain operational flexibility, factories are increasingly turning to dynamic, shared storage solutions. In these setups, incoming components from internal suppliers or external logistics are temporarily pooled into mixed transit carts, pallets, or unstructured bulk bins(Chacko, 2024). While this approach effectively conserves physical space, it creates a severe data bottleneck: production managers lose real time, granular visibility over the exact quantities and types of parts currently available on the floor.

This lack of precise, up to date inventory awareness severely limits the ability of the MES to sequence assembly tasks efficiently. Conventional tracking methods, such as manual stock counting or the use of fixed 2D barcode scanners, are too slow, labor intensive, and inherently unsuited for environments where parts are stored chaotically. Consequently, there is an urgent industrial demand for automated, contactless perception technologies that can digitize unstructured bins on the fly.

The deployment of 3D depth sensors provides a promising solution, as they capture rich geometric profiles that remain robust against poor lighting, visual clutter, and suspended particles (Guo et al., 2020). However, turning these raw depth captures into actionable inventory metrics requires highly dependable algorithms capable of distinguishing subtle structural differences between part variants, regardless of how they are physically positioned.

In recent years, the direct application of deep learning to 3D point clouds has emerged as a state of the art approach for complex industrial perception tasks (Liu et al., 2025). Unlike older methods that relied on manually engineered shape descriptors, modern neural networks learn to extract hierarchical geometric features directly from raw, unordered spatial data. In this particular domain, PointNet++ has emerged as a valuable instrument, employing a recursive architecture that processes nested partitions of point sets to capture both fine local details and broad structural topologies(Qi et al., 2017b). Its inherent invariance to point permutation and its capacity to handle complex, irregular geometries make it an ideal candidate for classifying mechanical parts inside cluttered industrial bins.

One industrial sector where these logistical bottlenecks are particularly problematic is shipbuilding. The construction of a vessel is a paradigm of high mix, low volume manufacturing, characterized by vast numbers of custom structural components that must be assembled in space constrained shipyard facilities(Wang et al., 2024). Among these components, simple sub-assemblies, specifically base plates, are manufactured in large quantities but in slightly differing geometric variations depending on their final structural purpose. These base plates are typically made by subcontractors who tend to accumulate them in shared transit bins on the shop floor. Without an automated system for verifying which plate variants are available, planners struggle to efficiently sequence subsequent welding and assembly operations, resulting in costly idle time and delays in the process.

To address this specific challenge, this work proposes a 3D approach for dynamic inventory management in shipyard environments. By combining a depth sensor with a customized PointNet++ architecture, we propose a system that can directly process unordered point clouds and classify mixed base plates, regardless of their physical pose. The selection of PointNet++ serves as a solid foundation rather than an attempt to set a new architectural standard. Its hierarchical feature learning is essential because, although the base plates share identical global dimensions, they differ in subtle local geometric features. While strict confidentiality agreements with our industrial partner prevent us from publishing the exact CAD models or photographs of these components, identifying these minor differences requires more than simple geometric descriptors.

Because this solution targets real time deployment on the shop floor, computational efficiency is as essential as classification accuracy(IYADUNNI et al., 2024). Therefore, our study systematically investigates the impact of spatial subsampling evaluating different point cloud resolutions to identify the optimal balance between geometric fidelity and processing speed required by modern Manufacturing Execution Systems.

This paper is organized as follows: Section 2 introduces the industrial case study and details the proposed methodological approach. Section 3 describes the methods and materials employed. Section 4 presents and analyzes the results, and Section 5 concludes the work while suggesting avenues for future research.

2. Case of Study and Approach

To evaluate the feasibility of deploying 3D deep learning for dynamic inventory management, we focus on an industrial case study developed in collaboration with a shipyard subcontractor. The study centers on the identification of base plates for simple sub-assemblies. In typical manufacturing workflows with constrained shop floor space, such as those found in modern shipbuilding environments, these structural components are often transported and temporarily stored in mixed batches rather than dedicated, perfectly aligned racks. The lack of a structured presentation makes traditional 2D vision systems and mechanical sorting mechanisms highly susceptible to error, as the parts appear in arbitrary orientations and varying distances from the sensor.

To address this unstructured environment, our proposed approach leverages direct 3D perception coupled with deep neural networks to achieve pose-invariant classification. The operational pipeline is designed to bridge the gap between physical stock accumulation and digital production scheduling.

First, we capture raw 3D point clouds of the visible surfaces of the objects within its field of view. Once these raw spatial data is acquired, it is fed into our pipeline. The first stage involves essential preprocessing, where individual objects are segmented from the background and their corresponding point clouds are isolated. To meet the real-time processing demands required for continuous inventory updates, we introduce a controlled subsampling step. Instead of processing the dense 10000 point captures, the geometries are downsampled to fixed smaller resolutions (ranging from 1024 to 8192 points). This spatial reduction ensures a predictable and minimal computational footprint while retaining the critical geometric features necessary for discrimination.

At the core of our approach is the deployment of a customized PointNet++ architecture for processing these subsampled point clouds. In this architecture, the network must correctly classify the specific base plate variant, regardless of its random physical pose.

3. Methods and materials

This section will describe all the methods and materials used in this work.

3.1. Dataset

Acquiring high-quality 3D data from industrial components is essential for developing robust machine learning models. To address this, this work relies on a custom built dataset of point clouds captured from three different physical replicas of base plates of simple sub-assemblies.

The creation of the dataset began with the fundamental design data. Technical blueprints provided by our industrial partner, which detailed the nominal geometry and dimensions of the original base plates, were carefully translated into high fidelity 3D CAD models. This step ensured that the essential structural features were accurately preserved before moving to physical manufacturing. The replicas were then fabricated via fused filament fabrication using a Bambu Lab P2S 3D printer. Polylactic acid (PLA) was chosen as the building material because it offers a reliable balance of ease of use, consistent results, and affordability.

Following fabrication, the 3D data acquisition was carried out using an Intel RealSense D455 depth camera. Instead of relying on static, uniform scans, we designed a dynamic acquisition protocol to better reflect real-world operational conditions. Each printed base plate was placed within the camera's active sensing volume and recorded multiple times. During this process, the position and orientation of the object relative to the sensor were altered. By varying viewpoints, standoff distances, and tilt angles, the dataset captures realistic changes in perspective and point density factors essential for training models intended for industrial inspection or robotic manipulation.

To guarantee that the geometric variations learned by the models are intrinsic to the objects, all extracted point clouds underwent a standardized pre-processing phase. This step involved normalizing the spatial coordinates of each sample to bring them into a common reference frame and scale, ensuring consistency across the entire dataset prior to model training and evaluation.

The raw captures generated dense point clouds, each consisting of approximately 10000 points. To evaluate the influence of point density on classification performance, a subsampling procedure was applied to the normalized data. Specifically, we created diverse subsets containing 1024, 2048, 4096, and 8192 points. Figure 1 illustrates an example of the 3D data extracted from one of the physical base plate replicas. The image compares the visual representation of the exact same object after the subsampling procedure, displaying two point clouds: a denser version with 8192 points on the left and a sparser version with 4096 points on the right.

Figure 1: Visual comparison of the 3D point clouds extracted from a physical base plate replica, showing subsampled resolutions of 8192 points (left) and 4096 points (right).

3.2. PointNet++

PointNet is a pioneering deep learning architecture for point-cloud understanding that operates directly on unordered sets of 3D points, avoiding voxelization or handcrafted geometric descriptors. It applies shared multilayer perceptrons (MLPs) to pointwise features and uses a symmetric aggregation function to obtain a permutation-invariant global descriptor(Qi et al., 2017a).

PointNet++ extends PointNet to better capture local geometric structure in point clouds by introducing a hierarchical feature learning scheme over progressively larger spatial neighborhoods(Qi et al., 2017b). Instead of processing all points independently, it builds a multiscale representation through repeated stages of sampling, grouping, and local feature extraction. At each stage, a subset of points is selected as centroids (typically via farthest point sampling), and nearby points are grouped within a neighborhood defined either by a radius (ball query) or by k-nearest neighbors. A shared PointNet-style network then aggregates features within each local region using symmetric pooling, producing descriptors that encode both fine-grained surface patterns and broader shape cues.

To handle variations in sampling density and to improve robustness across different object scales, PointNet++ can incorporate multi-scale grouping (MSG) or multi-resolution grouping, where features are computed over several neighborhood sizes and then combined. This hierarchical design generates gradually more abstract features as the network transitions to coarser point sets. Essentially, it simplifies the entire shape into a global representation, leading to effective point-cloud classification. Finally, feature propagation layers may be used (particularly in segmentation settings) to interpolate and refine features back to denser point sets, but for classification, the pipeline typically culminates in a global pooling and a lightweight MLP head that outputs class probabilities.

4. Experiments and Results

To evaluate the classification performance on our custom dataset, we implemented the PointNet++ architecture using the TensorFlow/Keras framework. The model is designed to flexibly accept input point clouds of varying densities, seamlessly accommodating the different subsampled resolutions discussed in the dataset section. At its core, the network hierarchically extracts local-to-global geometric features through a sequence of Set Abstraction (SA) modules. In these stages, spatial downsampling is governed by the Farthest Point Sampling (FPS) algorithm, which ensures a well-distributed and representative selection of centroids across the object's surface to robustly capture its underlying topology.

As the data flows through the network, the feature extraction pipeline progressively reduces the number of spatial points while simultaneously expanding the feature channels. At each SA level, local structures are encoded using shared multi-layer perceptrons (MLPs). This progressive downsampling and channel expansion allows the model to learn increasingly abstract representations, transitioning from fine-grained local surface details in the early layers to broader, higher level geometric patterns in the deeper layers. The feature extraction phase culminates in a final SA module that acts as a global aggregator, processing the remaining spatial information into a single, unified representation per point cloud.

Following this feature extraction, a global 1D max-pooling operation is applied to condense the learned features into a robust global shape descriptor. This distinctive signature vector is then fed into a classification head consisting of a series of fully connected layers. To stabilize the learning process during training and effectively mitigate overfitting, these dense layers are interleaved with Batch Normalization, ReLU activation functions, and Dropout regularization. Ultimately, a final dense layer with a softmax activation function provides the predicted class probabilities.

We defined the classification task as a multi-class problem corresponding to the three distinct types of base plates in our dataset in order to train the algorithm. For the weight updates, we employed the Adam optimizer configured with a constant learning rate of 0.0001. We employed 200 epochs of training with a batch size of 16 and dynamically shuffled the dataset to prevent ordering biases. To validate the reliability and reproducibility of our approach, we executed the entire training and evaluation pipeline twenty times, randomly varying the initialization seeds for each run. It is important to emphasize that all reported quantitative results correspond to the prediction phase. Once the models were trained, we assessed the network predictive performance exclusively on an independent and unseen test set. By using standard quantitative metrics including overall accuracy, precision, recall, and the F1 score on this separate data, we ensure that the evaluation accurately reflects the generalization capabilities of the model on new parts, rather than its performance on the training samples.

The quantitative performance of the implemented PointNet++ architecture on the custom base plate dataset is summarized in Table 1. To systematically evaluate the impact of spatial density on the network's predictive capabilities, the table reports the classification results across the four different input resolutions evaluated (1024, 2048, 4096, and 8192 points). The presented metrics represent the averaged performance across twenty independent runs of the evaluation pipeline with the different initialization seeds. Specifically, the table details the mean macro averaged precision, recall, and F1-score for each input size.

Table 1: Mean macro F1 score results for each model and resolution.

Point Clouds Resolution	Precision	Recall	F1-Score
1024 Points	0.9146	0.9138	0.9134
2048 Points	0.9160	0.9157	0.9153
4096 Points	0.9074	0.9065	0.9059
8192 Points	0.9041	0.9027	0.9021

As observed in Table 1, the mean classification performance remains highly consistent across all evaluated spatial densities, with macro F1-scores ranging between 0.9021 and 0.9153. To determine whether the variations in predictive performance among different input resolutions are statistically significant, we performed a statistical hypothesis test on the macro F1-scores obtained from twenty independent runs. First, we assessed the normality of the distributions using the Shapiro-Wilk test. While the results for the 2048, 4096, and 8192 resolution groups indicated normal distributions ($p > 0.05$), the 1024-point group violated the normality assumption ($p = 0.0016$). Concurrently, homoscedasticity was verified using Levene's test, confirming equal variances among the groups ($p = 0.3822$).

Because the assumption of normality was not met across all groups, we opted for the nonparametric Kruskal-Wallis H test to evaluate the global differences between the medians of the four resolutions. The test resulted in $H = 5.7697$ with a corresponding p-value of 0.1234. Since the p-value is greater than the standard significance level ($\alpha = 0.05$), the test indicates that there are no statistically significant differences in the classification performance across the different point cloud resolutions.

To be thorough, a post hoc Dunn's test with Bonferroni correction was used for pair-wise comparisons. The results confirmed that there were no significant differences between any pair of models.

From a practical and industrial standpoint, these statistical findings are highly relevant. The lack of significant difference in the F1-scores implies that increasing the point cloud density up to 8192 points does not provide any tangible improvement in the network's discriminative power for this specific case study. Therefore, the lowest evaluated resolution of 1024 points can be safely selected. This substantially reduces the computational overhead during the processing and inference stages, which aligns perfectly with the real-time processing constraints demanded by dynamic manufacturing execution systems.

5. Conclusions and future works

The combination of 3D computer vision and deep learning has become essential in modern manufacturing, especially when dealing with complex logistical challenges in confined, high-mix environments such as shipbuilding.

Our study focused on evaluating the feasibility of using a tailored PointNet++ architecture to automate the dynamic inventory management of mixed structural components. Specifically, we assessed how the subsampling of spatial data affects the network's ability to accurately classify different variants of base plates.

Our results showed that the PointNet++ model is highly effective for this task, maintaining a stable predictive performance regardless of the input density. The models achieved mean macro F1-scores ranging from 0.9021 to 0.9153 across the different resolutions. Moreover, rigorous statistical hypothesis testing using the Kruskal-Wallis test, followed by a Dunn's post hoc test, revealed no statistically significant differences in classification accuracy among the evaluated point densities.

From a practical perspective, this absence of significant variation is a highly favorable outcome. It indicates that the lowest evaluated resolution (1024 points) is entirely sufficient for accurate part differentiation. By intentionally reducing the point cloud density to this level, the computational burden associated with 3D data preprocessing and neural network inference can be minimized. This optimization is essential for real-time industrial applications, ensuring that the perception system can operate swiftly without becoming a bottleneck itself. In essence, this study suggests that using 3D deep learning pipelines can effectively address physical storage limitations and improve production flexibility.

As for future works, the findings of this study open several promising avenues for advancing 3D industrial perception. Although this research serves as an initial proof of concept conducted in a controlled environment, it lays the groundwork for real applications. First, while our current pipeline successfully classifies sub-assemblies replicas, subsequent research should focus on extending this approach to a wider range of components. Second, comparing the performance of different architectures, such as the Point Transformer and Graph Convolutional Networks, to the PointNet++ baseline could lead to improvements in accuracy and inference speed. Finally, a natural progression of this research involves physically integrating the classification model with a robotic manipulator, closing the loop between real time inventory recognition and the automated pickup of base plates for final assembly.

Acknowledgments

Paula Arcano-Bea's research was supported by the Xunta de Galicia (Regional Government of Galicia) through grants to Ph.D. (http://gain.xunta.gal), under the "Axudas á etapa predoutoral" grant with reference: ED481A-2025-088.

Grant PID2022-137152NB-I00 funded by MICIU/AEI/10.13039/501100011033 and by ERDF/EU.

CITIC, as a center accredited for excellence within the Galician University System and a member of the CIGUS Network, receives subsidies from the Department of Education, Science, Universities, and Vocational Training of the Xunta de Galicia. Additionally, it is co-financed by the EU through the FEDER Galicia 2021-27 operational program (Ref. ED431G 2023/01).

Anabel Díaz's research was supported by the Xunta de Galicia (Regional Government of Galicia) through grants to industrial Ph.D. (http://gain.xunta.gal), under the Doutoramento Industrial 2024 grant with reference: $02\_IN606D\_2024\_3100897$.

This research is co-financed by the Interreg Atlantic Area Programme through the European Regional Development Fund, $EAPA_0 019/2022$ SAtComm project.

References

Chacko, P., 2024. Accelerating the Integration of Low-Volume, High-Mix Production Organizations. Massachusetts Institute of Technology.

Gan, Z. L., Musa, S. N., Yap, H. J., 2023. A review of the high-mix, low-volume manufacturing industry. Applied Sciences 13 (3), 1687.

Guo, Y., Wang, H., Hu, Q., Liu, H., Liu, L., Bennamoun, M., 2020. Deep learning for 3d point clouds: A survey. IEEE transactions on pattern analysis and machine intelligence 43 (12), 4338–4364.

IYADUNNI, A. O., CHUKWUNWEIKE, M., AMARACHI, Q. O.-P., TITILOPE, T. A., 2024. Real-time inventory optimization in dynamic supply chains using advanced artificial intelligence. INTERNATIONAL JOURNAL OF MANAGEMENT & ENTREPRENEURSHIP 6 (12), 3830–3843.

Liu, Y., Zhang, C., Dong, X., Ning, J., 2025. Point cloud-based deep learning in industrial production: A survey. ACM Computing Surveys 57 (7), 1–36.

Mantravadi, S., Møller, C., 2019. An overview of next-generation manufacturing execution systems: How important is mes for industry 4.0? Procedia manufacturing 30, 588–595.

Qi, C. R., Su, H., Mo, K., Guibas, L. J., 2017a. Pointnet: Deep learning on point sets for 3d classification and segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 652–660.

Qi, C. R., Yi, L., Su, H., Guibas, L. J., 2017b. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. Advances in neural information processing systems 30.

Wang, T., Zhang, Y., Hu, X., 2024. A q-learning based hyper-heuristic scheduling algorithm with multi-rule selection for sub-assembly in shipbuilding. Computers & Industrial Engineering 197, 110567.

Zhong, R. Y., Xu, X., Klotz, E., Newman, S. T., 2017. Intelligent manufacturing in the context of industry 4.0: a review. Engineering 3 (5), 616–630.