PaintNet: Unstructured Multi-Path Learning from 3D Point Clouds for Robotic Spray Painting

Gabriele Tiboni[1,*], Raffaello Camoriano[1] and Tatiana Tommasi[1]

[1] *Politecnico di Torino, Corso F. Ferrucci 112, Torino, 10141, Italy*

Abstract

Popular industrial robotic problems such as spray painting require (i) adaptivity to free-shape 3D objects and (ii) planning multiple paths to solve the task. Yet, existing solutions make strong assumptions on input surfaces and output paths, limiting applicability in real-world scenarios. We introduce a novel 3D deep learning framework designed to target arbitrary 3D surfaces and generate a variable number of unordered paths. This is achieved by predicting short path segments and combining them to form long-horizon paths. We validate our approach on robotic spray painting tasks by releasing PaintNet, the first public dataset of real-world industry-grade painting paths on free-shape 3D objects. Our experiments demonstrate that the proposed approach predicts smooth output paths covering up to 95% of previously unseen object surface instances. The project page is available at https://gabrieletiboni.github.io/paintnet/.

Keywords

Computer Vision, Deep Learning, Robot Learning

1. Introduction

Conditioning robotic tasks on free-shape 3D objects is essential for many industrial applications and often these tasks unfold over long-time horizons, requiring significant amounts of computational resources for optimization and planning. Robotic spray painting is a typical example of this problem setting, where the robot must generate multiple trajectories for painting a surface, with each trajectory being a separate path through space. Even a simple planar surface becomes difficult when considering both its two sides, and the difficulties increase when facing an object composed of convex and concave parts with different samples showing significant variability in shape and size. The number and length of output paths will differ for every object instance. Despite its importance in product manufacturing, robotic spray painting remains a largely unsolved problem due to the lack of affordable and flexible solutions.

Existing research studies rely on decoupling the task into 3D object partitioning into convex surfaces, and offline trajectory optimization through either domain-specific heuristics [1, 2, 3, 4, 5, 6, 7, 8] or reinforcement learning-based policies [9]. Such approaches rely on simplified premises and are heavily tailored for specific shapes and convex surfaces only, which restricts their ability to generalize to novel objects. Additionally, approaches designed for autonomous 3D inspections and coverage path planning require expensive offline optimization routines, which hinder their practical applicability for industrial production lines [10, 11, 12, 13]. These limitations highlight the need for more suitable solutions that can operate on arbitrary 3D surfaces and efficiently handle complex multi-path planning problems.

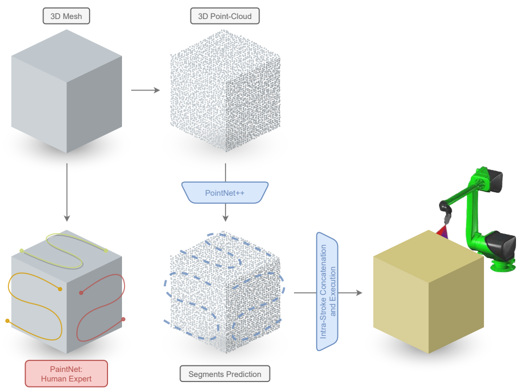


Figure 1: Overview of our method for multi-path prediction of 6D-pose spray painting paths given a raw 3D point-cloud in input.

In this work, we propose a novel method to address these challenges by designing a deep learning framework that can deal with unstructured high-dimensional input, such as 3D objects in the form of point clouds, and inherently cope with multiple output paths. Our approach learns a representation that captures geometric properties and surface details of a 3D object and consecutively predicts path segments that can be later concatenated to reconstruct long-horizon robotic paths (see Fig. 1). Un-

Ital-IA 2023: 3rd National Conference on Artificial Intelligence, organized by CINI, May 29–31, 2023, Pisa, Italy

*Corresponding author.
gabriele.tiboni@polito.it (G. Tiboni); raffaello.camoriano@polito.it (R. Camoriano); tatiana.tommasi@polito.it (T. Tommasi)
0000-0001-8722-918X (G. Tiboni); 0000-0002-8890-2732 (R. Camoriano); 0000-0001-8229-7159 (T. Tommasi)

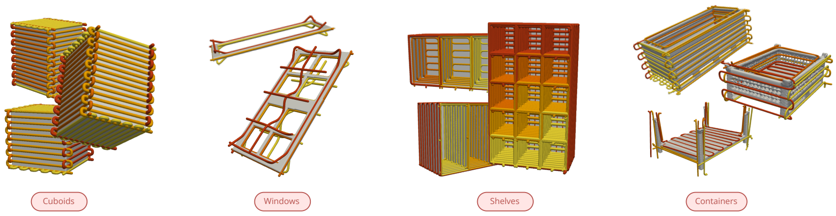


Figure 2: Overview of a few representative instances for each of the four categories included in the PaintNet dataset.

like heuristic techniques that need to be re-designed ad hoc for every task and object, our method can be applied to any 3D object-conditioned multi-path robotic task.

We ran an extensive experimental analysis on our newly designed dataset: PaintNet is the first collection of expert spray painting demonstrations on 3D objects in a multi-path setting. We plan to publicly release it to encourage future AI research on relevant industrial problems.

2. The PaintNet Dataset

The PaintNet dataset includes a total of 845 data pairs of 3D objects and corresponding expert spray painting demonstrations, collected in a real-world industrial scenario. It covers several shapes and task-specific trajectory patterns, spanning over four object categories of growing complexity: *cuboids*, *windows*, *shelves*, and *containers*. The data was generously provided by the EFORT group[1] and later preprocessed by the authors. In particular, all object meshes are released in a subdivided, smoothed watertight [14] version to avoid sharp edges and holes. Representative data samples are illustrated in Fig. 2.

3. Method

We approach multi-path learning for spray painting as a point cloud-based inference task, and present a tailored deep learning model to deal with unstructured output paths—*i.e.* unordered, and variable in number and length. The final objective of our pipeline consists in predicting path segments that are smoothly aligned and can be concatenated to resemble the original long-horizon strokes.

Segments Prediction. To guide the training of our model, expert paths are decomposed into unordered fixed-length segments. An optimal trade-off between the number (K) and length (λ) of predicted segments can therefore allow the model to inherently cope with the unstructureness of the original output space. We denote the set of ground truth path segments as $\mathbf{S} = \{\mathbf{s}^k\}_{k=1}^K$ composed of λ ordered poses—segment length—with $\mathbf{s}^k \in \mathbb{R}^{\lambda\times 6}$. Specifically, we consider an overlap of one pose among consecutive within-stroke segments to encourage contiguous predictions, resulting in a total number of $K = \sum_{i=1,\dots,I} \lfloor {}^{N_i-\lambda}/_{\lambda-1} \rfloor + 1$ ground-truth segments.

Our model takes as input the object point cloud $\mathbf{X}$ composed of unordered 3D points $x_{p=1,\dots,P} \in \mathbb{R}^3$, and provides as output a set of path segments $\mathbf{Y} = \{\boldsymbol{y}^k\}_{k=1}^K$, each of which contains λ ordered poses $y^k_{l=1,\dots,\lambda} \in \mathbb{R}^6$.

The learning objective is pursued by minimizing the following loss:

$$\mathscr{L}_{y2s} = \frac{1}{K}\sum_{\boldsymbol{y}\in\mathbf{Y}} \min_{\mathbf{s}\in\mathbf{S}} \|\boldsymbol{y} - \mathbf{s}\|_2^2 + \frac{1}{K}\sum_{\mathbf{s}\in\mathbf{S}} \min_{\boldsymbol{y}\in\mathbf{Y}} \|\mathbf{s} - \boldsymbol{y}\|_2^2 \,. \tag{1}$$

The symmetric Chamfer Distance [15]-based loss function above drives the prediction of unordered path segments close to the ones in the ground truth.

We furtherly bias the learning process to encourage output segments to be contiguous. To this end, we introduce the two sets of poses $\mathscr{B} = \{y_1^k\}_{k=1}^K$ and $\mathscr{E} = \{y_\lambda^k\}_{k=1}^K$, that respectively collect the beginning and ending poses of predicted segments. We then design an additional self-supervised Chamfer-based loss which guides close segments to have overlapping initial and ending poses:

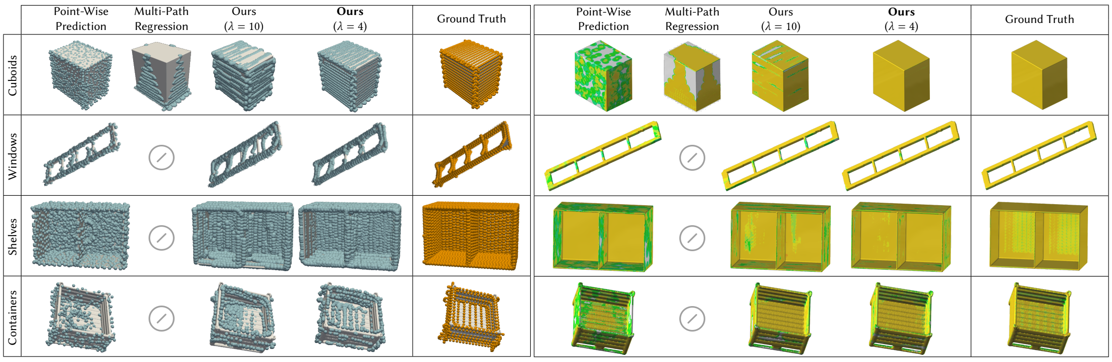
$$\mathscr{L}_{b2e} = \frac{1}{2K}\Big\{ \sum_{y_1^k\in\mathscr{B}} \min_{y_\lambda^j\in\mathscr{E}} \|y_1^k - y_\lambda^j\|_2^2 + \sum_{y_\lambda^k\in\mathscr{E}} \min_{y_1^j\in\mathscr{B}} \|y_\lambda^k - y_1^j\|_2^2 \Big\}\,, \tag{2}$$

with $j \neq k$. Overall, we train our model to optimize $\mathscr{L} = \mathscr{L}_{y2s} + \alpha\mathscr{L}_{b2e}$, with $\alpha \in \mathbb{R}^+$.

Intra-stroke Concatenation. We propose a simple technique based on segment proximity and alignment to demonstrate how predicted unordered path segments may be concatenated to reconstruct long-horizon paths executable on a real robot. Specifically, we interpret the segments as nodes of a graph and we aim at concatenating them such that each segment k has at most one outgoing $e_k^+ \leq 1$, and one incoming edge $e_k^- \leq 1$, where e is the signed edge degree. For each segment k, we evaluate the

[1]https://efort.com.cn/en/index.php/group

Table 1
Left: Predicted poses on representative PaintNet test instances (light blue) with corresponding ground-truth strokes (orange). **Right**: Spray painting coverage visualization when executing predicted and expert poses on a spray painting simulator. The colormap ranges from green (low paint thickness) to yellow (high paint thickness).



distance $d_k = min_j \|y^k_\lambda - y^j_1\|^2_2 + \|(y^k_\lambda - y^k_{\lambda-1}) - (y^j_2 - y^j_1)\|^2_2$ s.t. $j \neq k$ and $e^-_j = 0$, which considers proximity in space and orientation, as well as similarity in segment directions. Then, we connect two segments with a directed edge from k to j in case d_k falls below a predefined threshold τ, proceeding in ascending order of d_k.

Baselines. Due to the lack of generalizable data-driven solutions in current literature (see Sec. 1), we design two novel learning-based baselines for comparison with our model: *point-wise prediction* and *multi-path regression*. The former baseline uses a state-of-the-art shape completion pipeline to output unordered 6D poses rather than 3D points. The latter attempts to directly regress long-horizon output strokes, assuming that these are fixed in number and length across the training set—hence suitable for the *cuboids* category only.

Evaluation metrics. We introduce two evaluation axes to assess the performance of the considered baselines. One is the standard pose-wise Chamfer Distance (PCD) [15] to evaluate the goodness of the predicted disconnected poses w.r.t. the expert poses. The other is the percentage of surface covered by the predicted strokes—Paint Coverage (PC)—when executed on a spray painting simulator, relative to the ground truth.

4. Experiments

Implementation details. Our pipeline leverages an encoder architecture based on PointNet++ [16], that acts as a feature extractor from the input point cloud of 5120 down-sampled 3D points to a latent space of dimensionality 1024. A 3-layer MLP is then appended to generate output poses, with hidden size (1024,1024). Output 6D poses are encoded as the 3D Cartesian location plus the 3D unit vector indicating the 2-Dof orientation of the conic gun nozzle. Overall, we optimize our loss function $\mathscr{L}$ with $\alpha = 0.5$, orientation vectors weighted by 0.25, learning rate 10^{-3}, Adam optimizer, and 1200 epochs. Furthermore, we initialize our network with pre-trained weights from a shape classification task on ModelNet [17], and normalize input point-clouds by independently centering to zero mean and down-scaling by a category-specific factor. We carry out separate trainings for each PaintNet category of varying complexity while keeping the same hyperparameters.

Table 2
Chamfer Distance averaged over each category's test set, upscaled by 10^4. The lower the better.

	Cuboids	Windows	Shelves	Containers
Point-Wise Prediction	959.29	950.72	455.74	1073.15
Multi-Path Regression	8.32×10^5	-	-	-
Ours ($\lambda = 10$)	37.98	118.50	56.06	364.54
Ours ($\lambda = 4$)	**18.25**	**57.17**	**36.65**	**274.84**

Table 3
Spray painting coverage: % of covered mesh vertices with respect to ground-truth trajectories. Results are averaged over the test set. The higher the better.

	Cuboids	Windows	Shelves	Containers
Point-Wise Prediction	5.42%	39.90%	26.40%	71.99%
Multi-Path Regression	79.41%	-	-	-
Ours ($\lambda = 10$)	79.64%	68.84%	70.88%	82.88%
Ours ($\lambda = 4$)	**95.30%**	**84.05%**	**73.03%**	**89.32%**

Segments Prediction We report qualitative results on a subset of test instances in Tab. 1 (Left), and the full quantitative results on the test set in terms of PCD in Tab. 2. Despite optimizing for the PCD evaluation metric explic-

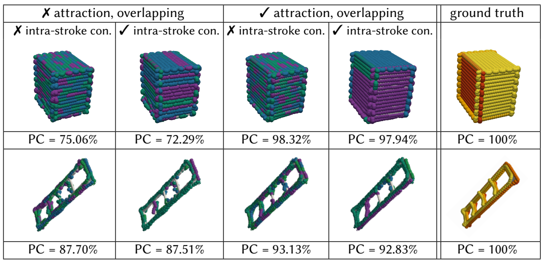


Figure 3: Intra-stroke concatenation post-processing step ($\tau = 0.15$) on cuboids and windows, from our approach ($\lambda = 4$).

itly, the point-wise prediction baseline applied to path generation leads to highly sparse poses, failing to preserve structure across all object categories. Interestingly enough, directly regressing a known number of 6 strokes for the cuboids category (multi-path regression baseline) also turns out to be problematic due to compounding errors on euclidean distances among high-dimensional sequences. On the other hand, with $\lambda = 4$ we observe the capability of our model to predict output path segments that closely resemble the ground truth and maintain a contiguous structure across all categories. Intuitively, the network is biased towards learning local spray painting patterns, which drastically simplifies the task and does not require learning implicit high-level planning. At the same time, the attraction loss term $\mathscr{L}_{b2e}$ assures aligned and contiguous predictions with nearby segments.

Spray Painting Coverage. We perform a thorough paint coverage analysis by executing the disconnected predictions—segments, single poses or long-horizon strokes—in a spray painting simulation in a random permutation. Ground-truth paint thickness references are obtained through the execution of the known long-horizon trajectory. Qualitative results describing deposited paint thickness on a few instances of PaintNet are depicted in Table 1 (Right). The complete quantitative paint coverage values are reported in Table 3 Overall, we draw similar conclusions as for the inference analysis: uniformly sparse poses predicted by the point-wise prediction model lead to poor coverage results, while the contiguous nature of predicted path segments with $\lambda = 4$ allows for up to 95.30% surface coverage. These results importantly demonstrate that supervised learning is a promising approach for learning the downstream task without directly optimizing for spray painting coverage.

Intra-stroke Concatenation. We inspect the capability of our proposed post-processing step to reconstruct long-horizon strokes for practical execution on robotic systems. We demonstrate the effectiveness of the intra-stroke concatenation step in Fig. 3, highlighting the contribution of both the attraction loss $\mathscr{L}_{b2e}$ and overlapping component to obtain optimal qualitative and quantitative results. In particular, we note that paint coverage results are preserved after the concatenation step, albeit not exactly the same: an effect likely due to the merging of overlapping poses during the concatenation.

5. Conclusions

In this paper, we tackle the core robotic problem of long-horizon, multiple path generation for tasks involving free-form 3D objects. In this context, we introduce PaintNet, the first industry-grade supervised dataset for robotic spray painting. We then present a novel method for learning the underlying task by building on 3D deep learning architectures and predicting path segments. We validate our method on the PaintNet dataset and evaluate its performance in simulation. Future work enabled by PaintNet will include the evaluation of more painting quality metrics beyond coverage, such as thickness accuracy deviation. In addition, investigating learning methods incorporating painting quality feedback could also lead to improved performance. Finally, we believe the proposed approach can also pave the way for research on other object-centric multi-path tasks in robotics, such as sanding, welding, or cleaning.

Acknowledgments

This study was carried out within the FAIR - Future Artificial Intelligence Research and received funding from the European Union Next-GenerationEU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR) – MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.3 – D.D. 1555 11/10/2022, PE00000013). This manuscript reflects only the authors' views and opinions, neither the European Union nor the European Commission can be considered responsible for them. We also acknowledge the support of EFORT Europe (www.efort.com) and the EU H2020 Elise project (www.elise-ai.eu).

References

[1] W. Sheng, N. Xi, M. Song, Y. Chen, P. MacNeille, Automated cad-guided robot path planning for spray painting of compound surfaces, in: IEEE/RSJ IROS, 2000.

[2] H. Chen, N. Xi, Automated tool trajectory planning of industrial robots for painting composite surfaces, The International Journal of Advanced Manufacturing Technology 35 (2008) 680–696.

[3] X. Li, O. A. Landsnes, H. Chen, M.-V. Sudarshan, T. A. Fuhlbrigge, M.-A. Rege, Automatic trajec-

tory generation for robotic painting application, in: ROBOTIK, 2010.
[4] M. Andulkar, S. Chiddarwar, Incremental approach for trajectory generation of spray painting robot, Industrial Robot: An International Journal 42 (2015) 228–241.
[5] P. N. Atkar, A. Greenfield, D. C. Conner, H. Choset, A. A. Rizzi, Uniform coverage of automotive surface patches, The International Journal of Robotics Research 24 (2005) 883–898.
[6] D. Gleeson, S. Jakobsson, R. Salman, F. Ekstedt, N. Sandgren, F. Edelvik, J. S. Carlson, B. Lennartson, Generating optimized trajectories for robotic spray painting, IEEE Transactions on Automation Science and Engineering (2022).
[7] W. Chen, X. Li, H. Ge, L. Wang, Y. Zhang, Trajectory planning for spray painting robot based on point cloud slicing technique, Electronics 9 (2020).
[8] G. Biegelbauer, A. Pichler, M. Vincze, C. Nielsen, H. Andersen, K. Haeusler, The inverse approach of flexpaint [robotic spray painting], IEEE RAM 12 (2005) 24–34.
[9] J. Kiemel, P. Yang, P. Meißner, T. Kröger, Paintrl: Coverage path planning for industrial spray painting with reinforcement learning, in: RSS Workshop, 2019.
[10] B. Englot, F. Hover, Sampling-based coverage path planning for inspection of complex structures, in: ICAPS, 2012.
[11] A. Bircher, M. Kamel, K. Alexis, M. Burri, P. Oettershagen, S. Omari, T. Mantel, R. Y. Siegwart, Three-dimensional coverage path planning via viewpoint resampling and tour optimization for aerial robots, Autonomous Robots 40 (2016) 1059–1078.
[12] W. Jing, D. Deng, Y. Wu, K. Shimada, Multi-uav coverage path planning for the inspection of large and complex structures, in: IEEE/RSJ IROS, 2020.
[13] S. Ivić, B. Crnković, L. Grbčić, L. Matleković, Multi-uav trajectory planning for 3d visual inspection of complex structures, Automation in Construction 147 (2023) 104709.
[14] J. Huang, H. Su, L. Guibas, Robust watertight manifold surface generation method for shapenet models, arXiv preprint arXiv:1802.01698 (2018).
[15] H. Fan, H. Su, L. J. Guibas, A point set generation network for 3d object reconstruction from a single image, in: IEEE CVPR, 2017.
[16] C. R. Qi, L. Yi, H. Su, L. J. Guibas, Pointnet++: Deep hierarchical feature learning on point sets in a metric space, NeurIPS (2017).
[17] Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang, J. Xiao, 3d shapenets: A deep representation for volumetric shapes, arXiv preprint arXiv:1406.5670 (2014).