



Deep Bayesian ICP Covariance Estimation

Andrea de Maio, Simon Lacroix

► To cite this version:

Andrea de Maio, Simon Lacroix. Deep Bayesian ICP Covariance Estimation. IEEE International Conference on Robotics and Automation (ICRA 2022), May 2022, Philadelphia, United States. pp.6519-6525, <10.1109/ICRA46639.2022.9811899>. <hal-03766231>

HAL Id: hal-03766231

<https://laas.hal.science/hal-03766231v1>

Submitted on 31 Aug 2022

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.



HAL Authorization

Deep Bayesian ICP Covariance Estimation

Andrea De Maio¹ and Simon Lacroix¹

Abstract—Covariance estimation for the Iterative Closest Point (ICP) point cloud registration algorithm is essential for state estimation and sensor fusion purposes. We argue that a major source of error for ICP is in the input data itself, from the sensor noise to the scene geometry. Benefiting from recent developments in deep learning for point clouds, we propose a data-driven approach to learn an error model for ICP. We estimate covariances modeling data-dependent heteroscedastic aleatoric uncertainty, and epistemic uncertainty using a variational Bayesian approach. The system evaluation is performed on LiDAR odometry on different datasets, highlighting good results in comparison to the state of the art.

I. INTRODUCTION

With the steady increase in 3D sensors availability, point clouds have seen a rapid diffusion and adoption. Their use benefits from the 3D information that they readily express, which allows to easily retrieve geometric properties and yield the possibility to extract complex geometric features to classify object shapes, perform scene segmentation, or assess complex situations. Point clouds are also used to solve motion estimation and self-localization problems in robotics. Scan registration algorithms provide an estimate of the 3D transformation between the positions at which two point clouds are acquired. Such processes are at the heart of LiDAR-based odometry [1] and SLAM frameworks [2]. Originally introduced in [3], the Iterative Closest Point (ICP) algorithm, in its many variants, gradually became the standard approach to the point-cloud registration problem [4].

In order to qualify the result of ICP, and especially to integrate it within a localization framework, it is essential to estimate its uncertainty. ICP errors stem from various error sources. Among those, the structure of the scene plays an important role: underconstrained situations, such as corridors or mostly flat environments, yield larger errors than geometrically more complex scenes (Fig. 1).

Building on the advances in deep learning techniques dealing with point clouds, we present an approach that estimates data-dependent error models for the ICP process in form of a covariance matrix. Our method learns heteroscedastic aleatoric uncertainty from ICP input data and uses Bayesian posterior approximation to capture the epistemic uncertainty.

The following section introduces the ICP process, reviews the various sources of errors and works relevant to ours: methods that estimate ICP covariance, data-driven approaches that have proven successful to estimate covariances for other estimation processes, and point-cloud based deep learning techniques. Section III formalizes the problem and depicts the learning process. Section IV evaluates our results

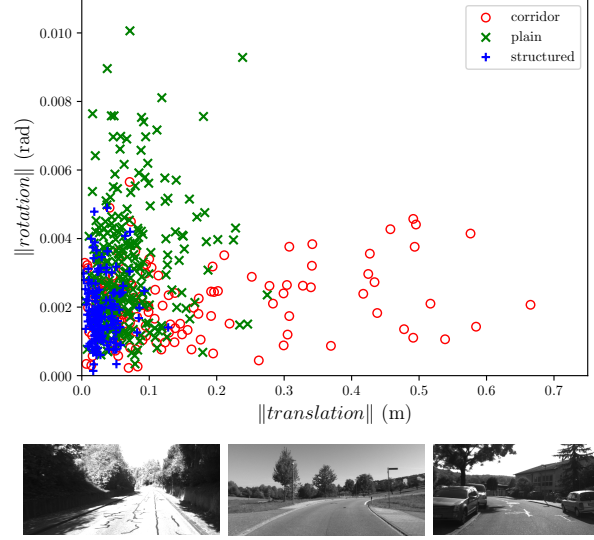


Fig. 1: Distribution of errors on a series of ICP estimates computed in three different kind of environments: corridor-like (left), wide plain (center), structured (right).

against the state-of-the-art, showing an accurate correlation between the real error of ICP and our predicted covariance.

II. BACKGROUND AND RELATED WORK

A. Iterative Closest Point process

ICP tackles the problem of aligning a *reading* point cloud $\mathcal{P} \in \mathbb{R}^{3 \times n}$ to a *reference* point cloud $\mathcal{Q} \in \mathbb{R}^{3 \times m}$, estimating the true rigid transformation between them. Using an initial estimation $\tilde{T} \in \text{SE}(3)$ of the transformation (*e.g.* as provided by a prior on the motion), the rigid transformation $\hat{T} \in \text{SE}(3)$ that estimates the true transformation is the one that minimizes the error function e :

$$\hat{T} = \arg \min_T (e(T(\mathcal{P}), \mathcal{Q})) \quad (1)$$

where $T(\mathcal{C})$ is the transformation T applied to the point cloud \mathcal{C} . The error function is the result of a weighted average distance between appropriately matched point pairs or point to plane associations. The optimization is an iterative process, and there is no guarantee to reach the optimal solution.

B. Sources of error for ICP

There are five main sources of error for ICP described in the literature: wrong convergence, wrong initialization, sensor noise and bias, underconstrained situations, and intrinsic randomness of the process [5], [6].

¹LAAS-CNRS, Université de Toulouse, 7 Avenue du Colonel Roche, 31031, Toulouse, France, E-mail: andrea[dot]de[dash]maio[at]laas[dot]fr

a) *Convergence and initialization*: Wrong convergence comes from the iterative nature of ICP, which does not guarantee global convergence and can get stuck in local minima. This is closely related to wrong initialization, which can be a predominant source of error when the initial guess is far from the true solution [6].

b) *Sensor noise and biases*: Each sensor naturally suffers changes in its accuracy that depend on several environmental conditions (*e.g.* temperature) [7]. Also, sensor calibration can introduce distortion [8]. While these phenomena introduce a bias in the estimate [9], there is an additional source of sensor noise added to each point independently, which generally is a function of the distance, orientation and physical nature or texture of the perceived material.

c) *Underconstrained situations*: These occur when the environment does not offer enough information to fully estimate the transformation between the point clouds, mainly by precluding the establishment of good matches between points. While in the planar 2D case such situations can be exhaustively identified, namely the corridor and the circle, in the 3D world this becomes a more challenging task.

d) *Intrinsic ICP randomness*: Depending on the configuration and implementation, a certain degree of randomness is introduced by subprocesses of ICP, albeit the process itself is deterministic. For example, filtering clouds with outlier rejection and sub-sampling can yield different estimates, even with the same input data.

C. Estimating ICP covariance

Various methods for estimating ICP covariance have been proposed. Monte-Carlo algorithms [10] generate the covariance from a large number of samples of scan registration results. Different clouds and initializations are randomly subsampled to obtain a dispersion of the transformation that is used to compute the covariance. Closed-form methods [5], [6], [11], [12] provide algebraic solutions to the ICP covariance problem by linearizing the error functions defined in the ICP algorithm. These methods are potentially ill-founded if point reassociations happen at a scale smaller than the actual error [13]. Experiments on synthetic data have shown that this is often the case [14]. Early closed-form works, as the one in [5], have been demonstrated over optimistic [2]. Recently, [14] proposed a learning approach to the ICP covariance estimation problem. The work estimates a 3D covariance based on preprocessed feature descriptors that are used in the learning scheme. While such descriptors are generic, it is desirable to have a more flexible representation of features impacting on the error of ICP.

It is worth to notice that most of these work tackle different sources of error. The method presented in [6] shows the need for a proper initialization of ICP and proposes a novel approach to 3D uncertainty of ICP accounting for initialization errors. Censi's pioneering work [5] studies the error due to sensor noise and computes the covariance of this error as a function of the error metric minimized by any ICP implementation. The work in [14], closer to our approach, explores a data-driven approach to estimate the covariance focusing on sensor noise and aiming at detecting underconstrained situations.

D. Learning motion estimation error models

The literature in deep learning features a large amount of work in the estimation of uncertainty. The approaches are often tied to the process for which they predict error models. Interesting works in domains similar to ours have been proposed, mainly tackling the Visual Odometry (VO) case. As ICP, VO is a well-known family of techniques which estimate a rigid transformation between two consecutive frames for the monocular or stereovision cases. End-to-end methods perform frame to frame motion estimation [15] and covariance prediction [16]. Both estimates are in this case delegated to neural networks, and do not rely on any geometrical pipeline. The works in [17], [18] respectively show covariance and joint bias-covariance estimation to the geometric implementation of VO.

E. Deep learning for point clouds

Recently, deep learning methods applied on point clouds have proven successful for several tasks [19]–[22]. In particular, PointNet [19] paved the way for neural networks that directly use 3D point clouds for object classification and segmentation purposes, and has been extended towards learning point cloud features in metric space [20]. This extension increased the effectiveness of PointNet to learn features on different scales, in the manner of traditional convolutional networks. Other approaches rely on the spherical projection of point clouds, usually coined as range (or depth) image – which is, in most cases, the original structure of the sensor data. This is used in the LiDAR odometry techniques [23]–[25] which makes use of the entire scan to estimate the motion between two frames in an end-to-end manner.

III. DATA DRIVEN LEARNING OF ICP UNCERTAINTY

Excluding the process initialization, all error sources of ICP are inherited from the input data. Sensor noise is directly reflected in the data, as well as ICP subprocesses modifying its size, density and structure (*i.e.* randomness due to filtering). Additionally, the scene structure, with its inherent geometry and features, largely impacts the error of ICP (as illustrated in Fig. 1), as it defines the observability of the motion along the different dimensions.

Addressing all these sources in a unified framework is a challenging task. Leveraging recent progresses in deep learning for point clouds, we address the problem of estimating the covariance of ICP that stems from the last three error sources presented in Sec. II-B. Our hypothesis is that a learning scheme that processes the raw input data can capture these error sources and properly estimate ICP covariances.

In this section we describe our approach to estimate ICP covariances and discuss the nature of the learned uncertainties. The optimization problem is contextualized in a probabilistic framework and a description of the learning architecture incorporating different types of uncertainty is provided.

A. Minimization problem

Modeling errors for an estimation process can be tackled via a supervised approach. Consider an estimator, ICP in

our case, using sensory data $\mathcal{C}_i = \{\mathcal{P}_i, \mathcal{Q}_{i+1}\}$ with $\mathcal{P}_i \in \mathbb{R}^{n_i}$ and $\mathcal{Q}_{i+1} \in \mathbb{R}^{m_{i+1}}$. Its output ${}^i\hat{\mathbf{T}}_{i+1}$ is an estimate of the ground truth transformation ${}^i\mathbf{T}_{i+1}$. In practice, the point clouds used in the algorithm are the result of several filters [4]. The actual input data is a pair of decimated point clouds $\bar{\mathcal{C}}_i = \{\bar{\mathcal{P}}_i, \bar{\mathcal{Q}}_{i+1}\}$ with $\bar{\mathcal{P}}_i \in \mathbb{R}^{m_i}, m_i \ll n_i$ and $\bar{\mathcal{Q}}_{i+1} \in \mathbb{R}^{m_{i+1}}, m_{i+1} \ll n_{i+1}$. The error of ICP is computed as

$$\mathbf{e}_i = {}^i\hat{\mathbf{T}}_{i+1}^{-1} \cdot {}^i\mathbf{T}_{i+1} \quad (2)$$

We represent an error $\mathbf{e} \in \text{SE}(3)$ as a vector $\boldsymbol{\xi} \in \mathfrak{se}(3)$ member of the Lie algebra [26], [27]

$$\boldsymbol{\xi}^\wedge = \begin{bmatrix} \boldsymbol{\rho} \\ \phi \end{bmatrix}^\wedge = \begin{bmatrix} \phi^\wedge & \boldsymbol{\rho} \\ \mathbf{0}^\top & 0 \end{bmatrix} \quad (3)$$

where the *hat* operator \wedge turns the vector $\boldsymbol{\xi} \in \mathbb{R}^6$ into a matrix $\boldsymbol{\xi}^\wedge \in \mathbb{R}^{4 \times 4}$. The opposite linear map is represented as the *vee* operator \vee . The exponential map $\exp(\boldsymbol{\xi}^\wedge)$ allows to retract an element of the Lie algebra back to the group. Its inverse is the logarithmic map $\log(\mathbf{T})$. Once the error is defined as $\boldsymbol{\xi} = \log(\mathbf{e})^\vee$, the associated uncertainty $\boldsymbol{\Sigma} \in \mathbb{R}^{6 \times 6}$ can be expressed as

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{\rho\rho} & \boldsymbol{\Sigma}_{\rho\phi} \\ \boldsymbol{\Sigma}_{\phi\rho} & \boldsymbol{\Sigma}_{\phi\phi} \end{bmatrix} \quad (4)$$

This uncertainty representation is generally valid for small perturbations that are added to a given pose $\hat{\mathbf{T}} = \mathbf{T} \exp(\boldsymbol{\xi})$ where $\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ [27]. The equation above matches the error definition in Eq. 2, that is $\boldsymbol{\xi} \in \mathbb{R}^6$ can be viewed as the error between the real transformation and its estimate.

We define a dataset $\mathcal{D} = \{\mathcal{C}_i, \boldsymbol{\xi}_i | \forall i \in [1, d]\}$, where d is the size of the dataset. The goal is to predict the uncertainty of the error vector $\boldsymbol{\xi}_i$ along its dimensions. In the data, we noticed that the unbiased error assumption is valid for the vast majority of ICP estimates and the bias is very small otherwise. To keep the uncertainty assumption valid, we encode a bias term for the error vector in the distribution that ensures the error remains close to a zero-mean Gaussian. Therefore, it is possible to estimate the parameters of the Gaussian $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ as

$$\arg \max_{\boldsymbol{\mu}_{1:d}, \boldsymbol{\Sigma}_{1:d}} \sum_{i=1}^d p(\boldsymbol{\xi}_i | \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) \quad (5)$$

To maximize the probability of drawing the correct error given the estimated Gaussian parameters is equivalent to minimize the negative full log-likelihood function [18]. To ensure a positive definite covariance matrix we use the LDL decomposition as in [17].

B. Uncertainty estimation

ICP covariance can be decomposed into two categories: epistemic and aleatoric [28]. Both uncertainties are formulated as probability distributions over different entities.

Aleatoric uncertainty models the uncertainty over the input data to the process. It assumes an observation noise that can be either constant or vary with the input. As discussed in Sec.II-B, we observed that noise levels change with the scene (Fig. 1), hence we are interested in modeling the variation of

the noise in correlation with the input data. This is referred to as heteroscedastic aleatoric uncertainty (as opposed to homoscedastic uncertainty for the constant noise case).

Additionally, it is possible to model the variance intrinsic to the learned model, that is how certain of a given prediction the network is. Epistemic uncertainty can, in principle, be reduced if more data is used for training. In our case, it accounts for the network's poor predictions on out-of-data samples such as scene features that were rarely encountered during training. Epistemic uncertainty can be modeled by placing a distribution over the network's weights. Unfortunately, given a prior distribution $\mathcal{N}_{\mathbf{W}}(0, \sigma)$ over the weights, it is computationally difficult to analytically infer the posterior $p(\mathbf{W} | \mathbf{X}, \mathbf{Y})$, that is the probability of drawing a set of weights given the network input and output. There is a number of proposed methods to approximate this posterior [29], [30]. The work in [31] produces an approximation using dropout variational inference. This has proven to be an effective yet easily obtainable Bayesian approximation when dealing with deep complex models.

In our context, From Eq.5, we consider the mean as a noisy error estimate, and the covariance matrix as the aleatoric uncertainty due to the observation noise. To approximate the posterior for a Bayesian neural network estimating the ICP errors we resort to Monte-Carlo dropout. The epistemic uncertainty can be captured by keeping dropout activated at test time. Sampling N times the predicted ICP error vector and ICP covariance yields a distribution $[\hat{\boldsymbol{\mu}}_{1:N}, \hat{\boldsymbol{\Sigma}}_{1:N}] = f_{\mathbf{W}}(\bar{\mathcal{C}})$ where $f_{\mathbf{W}}(\bar{\mathcal{C}})$ is the output of a Bayesian neural network, function of the input point clouds.

The total covariance matrix, integrating aleatoric and epistemic uncertainty is computed as

$$\frac{1}{N} \sum_{n=1}^N (\hat{\boldsymbol{\mu}}_n - \bar{\boldsymbol{\mu}})(\hat{\boldsymbol{\mu}}_n - \bar{\boldsymbol{\mu}})^\top + \frac{1}{N} \sum_{n=1}^N \hat{\boldsymbol{\Sigma}}_n \quad (6)$$

where $\bar{\boldsymbol{\mu}} = \frac{1}{N} \sum_{n=1}^N \hat{\boldsymbol{\mu}}_n$ is the mean error vector over the sampled outputs. The first term in Eq. 6 computes the uncertainty over the predicted error vector, which presents a variance proportional to the network confidence over the estimate. The mean vector is not used to correct ICP estimates as the original error is generally noisy. It is rather evaluated in terms of dispersion after having approximated the model posterior.

C. Learning architecture

The improvements in point clouds based network involved a number of estimation processes. Recently, FlowNet3D [32] proposed an approach to tackle 3D scene flow expanding on the building blocks introduced in [20]. It provides the 3D flow for each point in a reference point cloud deducing its motion from the comparison with a second, successive, point cloud. Estimating the scene flow is a problem closely related to the scan registration. Intuitively, a network able to approximate such task results suitable for ICP related inferences. In fact, the main difference between the two tasks lies in the generalization needed to shift from a high dimensional flow to the estimate of a rigid transformation in the ICP case. We leverage the main structure of FlowNet3D to tackle the

Layer type	Radius	Sampling	MLP size
set conv	2.0	$1.0\times$	[32, 32, 64]
set conv	4.0	$0.25\times$	[64, 64, 128]
flow embedding	knn	$1.0\times$	[128, 128, 128]
set conv	8.0	$0.25\times$	[128, 128, 256]
set conv	16.0	$0.125\times$	[256, 256, 512]

TABLE I: Network architecture parameterization. Set upconv layers are the same as in the updated version of [32].

estimation of ICP errors, modifying few key aspects to adapt it to our problem. To complete the architecture we add a final MLP layer reducing the dimensionality to the number of parameters needed to predict the Gaussian parameters. First we follow the original regression layer from FlowNet3D reducing the feature space to $\mathbb{R}^3 \times n$. The MLP is extended to take as input the feature produced by the regression layer and to return the parameters generating the covariance matrix and predicted error vector. All the MLP (originally Linear-BatchNorm-ReLU) in FlowNet3D are stacked with Dropout modules to allow the estimation of the epistemic covariance matrix. Additionally, this proved to reduce overfitting as expected.

We initially assumed that including the *set upconv* layer would have not been needed. Intuitively, as in a classical convolutional scheme, the regression task we try to solve can be solved with direct dimensionality reduction, from sensory data to parameters describing uncertainty [17], [18]. Nonetheless, retaining upsampling layer produced better results in terms of predicted uncertainty and mean. We associate this behavior with the architecture using skip connections to use the same features extracted from the same point sets originally used ICP. This creates a 1:1 matching between the data used by the geometric estimator and the network, leaving less room for outliers. Our network parameterization, with sampling rates and radius, can be found in Table I.

IV. EVALUATION

A. ICP process and dataset

To learn errors produced by a geometric ICP algorithm we selected *libpointmatcher* [33]. It is a widely used open-source library offering a chain that filters point clouds, establishes points associations, removes outliers and estimates rigid transformations using an error minimizer. For our setup we parameterized the process similarly to [14]. Cloud random sampling is fixed at 2048 random points. We generate normals, used for the point-to-plane minimizer, using the 10 nearest points before subsampling.

To select the initial transform for ICP, we add an error to the ground truth T . In order not to draw errors from an overoptimistic distribution we sample guesses for the initial transforms from a non-zero mean normal distribution $\mathcal{N}(\mathcal{N}(T, a \cdot T), b \cdot T)$, with $a = 0.25, b = 0.2$. These figures yield more pessimistic estimates than commonly used aiding sensor (e.g odometry or inertial measurements) posing a challenge to our system.

We select the KITTI odometry dataset [34], using LiDAR scans, which is made of urban and rural road scenes. We train and test using all ground truth tagged sequences but sequence 01, on which ICP fails to register most of the scans due to scenes lacking features. To compare against the state of the art in covariance estimation for ICP, we also present results on the challenging data sets of [35]. Since for the LiDAR odometry case this dataset is rather small, we refine our KITTI trained model with a few epochs of training, approximately 30 or less for early convergence, on all but one outdoor sequences and validate on the one left out. As expected, features learned on KITTI are ineffective to predict covariances on the *apartment* and *stairs* sequences, which are highly structured small indoor environments. Our system learns to model ICP errors with numerically small epistemic uncertainty. For reasons of space, we only show results containing the combined aleatoric+epistemic covariance when moving to unseen datasets, *i.e* applying models learned on KITTI to the ETH dataset.

B. Single-pair validation

We first assess the quality of the predicted covariances in a single-pair fashion, and then check their consistence on a full trajectory. The metrics we use are based on the relation between the observed errors and the predicted covariance. This is due to the absence of the true distribution which would allow for more classic figures of merit, such as the Kullback-Leibler divergence. While it would be possible to obtain a pseudo-true distribution by random sampling over the point cloud, this would affect the estimated covariance as well, because the network prediction is tied to the ICP configuration, and most importantly to the input data used by ICP.

As in [5], [6], we compute the *Normalized Norm Error (NNE)* to evaluate the scale of the predicted covariance w.r.t the actual error:

$$\text{NNE} = \frac{1}{N} \sum_{n=1}^N \sqrt{\frac{\|\xi_n\|_2^2}{\text{tr}(\hat{\Sigma}_n)}} \quad (7)$$

where $\text{tr}(\hat{\Sigma}_n)$ is the trace of the predicted covariance matrix. The optimal value for NNE is one. Values below and above one respectively represent pessimistic and optimistic uncertainty estimates. Table II shows NNE values computed over test KITTI sequences. The scale of the covariance is well predicted, and is only slightly pessimistic for two validation sequences (05 and 07).

NNE (KITTI)	05	06	07	09	10
Translation	0.70	0.61	0.58	0.94	0.90
Rotation	0.65	0.81	0.55	0.90	1.04

TABLE II: NNE translation and rotation values for the KITTI dataset.

The NNE is an immediate measure of the adequacy of the predicted variances, but it does not account for off-diagonal covariances. For this purpose, the Mahalanobis distance between ICP estimates and ground truth $\xi_n =$

<i>Single-Pair Mah. Dist (KITTI)</i>	05	06	07	09	10
Translation	0.80	0.83	0.74	0.88	1.03
Rotation	0.77	0.63	0.70	0.94	0.96

TABLE III: Single-pair Mahalanobis distance in terms of translation and rotation for the KITTI dataset.

$\log({}^n\hat{\mathbf{T}}_{n+1}^{-1} {}^n\mathbf{T}_{n+1})^\vee$, weighted by the predicted covariance Σ_n averaged over each consecutive pair in a sequence, is a better metric:

$$D_M = \frac{1}{N} \sum_{n=1}^N \sqrt{\frac{\xi_n^\top \Sigma_n^{-1} \xi_n}{\dim(\xi)}} \quad (8)$$

As in the NNE case, the optimal value for a normal distribution is one, with distances below and above one respectively denoting pessimistic and optimistic estimates.

Table III shows the average Mahalanobis distance for the validation sequences. All sequences present values positively close to 1, indicating that the system is able to correctly capture the uncertainty under the Gaussian assumption. For these sequences, the epistemic uncertainty is small and constant, which denotes confidence in the system prediction. Minor variations are encountered for seldom cases: an example is shown in Fig. 2 displaying the epistemic covariance before, during, and after an area corresponding to *wide plain* scenes shown in Fig. 1. Such spatial features are less frequently encountered during training.

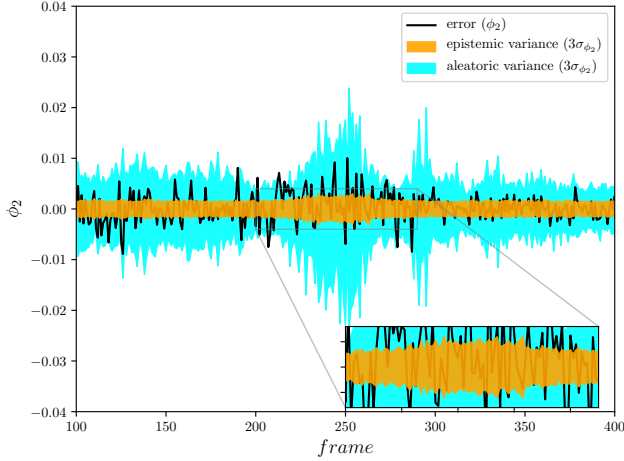


Fig. 2: Epistemic and aleatoric 3σ intervals in a rarely seen area lacking sharp geometrical features. The real error computed from the ground truth is represented in black. The zoomed-in area shows epistemic uncertainty increase, which corresponds to a *wide plain* environment.

C. Trajectory validation

In order to assess the quality of our predicted covariances on trajectories we take a multi-step approach. First, we compose ICP estimates to generate the final pose as ${}^0\hat{\mathbf{T}}_f = {}^0\hat{\mathbf{T}}_1 {}^1\hat{\mathbf{T}}_2 \dots {}^{f-1}\hat{\mathbf{T}}_f$ for a sequence with f poses. The ground truth final pose ${}^0\mathbf{T}_f$ is computed similarly. We propagate the

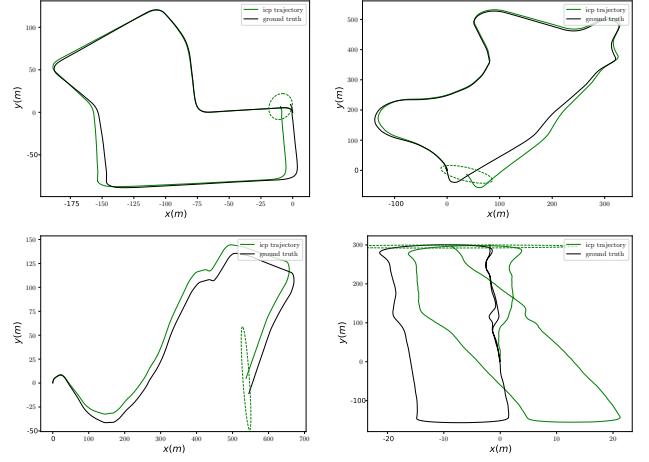


Fig. 3: Four propagated covariances in the 2σ interval (95%) on considered ICP trajectories compared to ground truth.

<i>Trajectory Mah. Dist (KITTI)</i>	05	06	07	09	10
Translation	2.31	1.51	1.26	1.36	4.86
Rotation	1.49	1.95	2.19	1.89	2.95

TABLE IV: Trajectory Mahalanobis distance in terms of translation and rotation for the KITTI dataset.

covariance associated to each estimate using the 4th order approximation, as in [27], to obtain an uncertainty estimate for the final ICP pose. This way, it is possible to evaluate the Mahalanobis distance (Eq. 8) in a spatially consistent manner. The results are shown in Table IV. They are only slightly optimistic, which is a remarkable outcome given the length of KITTI trajectories (up to more than 2km).

D. Fine tuning on other datasets

We also evaluate the results on the smaller dataset of [35] and compare to the state of the art. As our model requires a significant amount of data we initially learn on a large dataset. KITTI makes available a considerable number of LiDAR scans with different scenarios, but all correspond to outdoor road environments. To apply our approach to smaller datasets acquired in very different environments, a fine-tuning process is applied using the pre-trained model on the KITTI dataset. This allows to obtain much better results compared to re-training on such a small dataset and shows the importance of epistemic uncertainty as an indicator of the performances of a model on unseen data.

Table V shows the result of our approach on single-pair estimates with the dataset of [35]. While the environments are very different, the model performs remarkably well in outdoor scenes. Even architectural features, such as in *haupt-gebaude* can be leveraged to produce faithful covariance predictions as, intuitively, the scene presents a structure (flat ground plane, lack of indoor man-made elements) that is geometrically similar to the some scenes encountered in KITTI.

More importantly, Table V highlights the impact of the

Single-Pair Mahalanobis Distance (ETH)		<i>Gazebo winter</i>	<i>Gazebo summer</i>	<i>Mountain</i>	<i>Haupt-gebaude</i>	<i>Wood autumn</i>	<i>Wood summer</i>
Ours (Direct)	Trans (A)	2.05	2.32	4.15	2.68	4.01	4.03
	Trans (E+A)	2.04	2.12	3.88	2.49	4.19	3.78
Ours (Finetuned)	Trans (A)	0.64	1.12	1.27	1.04	1.86	1.03
	Trans (E+A)	0.59	1.06	1.21	0.98	1.70	0.90
Ours (Direct)	Rot (A)	1.33	1.93	3.06	2.61	2.10	2.33
	Rot (E+A)	1.30	1.89	3.03	2.47	2.07	2.19
Ours (Finetuned)	Rot (A)	0.48	0.53	0.65	1.02	0.54	0.82
	Rot (E+A)	0.44	0.49	0.65	0.96	0.51	0.71

TABLE V: Single-pair Mahalanobis distance in terms of translation and rotation for the ETH dataset using aleatoric only (A) and combined epistemic-aleatoric (E+A) uncertainty. A comparison between the pre-trained model on KITTI and after finetuning is presented. After finetuning, the epistemic uncertainty is less prominent and its (minor) impact on performances depends on whether the system is being pessimistic or optimistic.

Trajectory Mahalanobis Distance (ETH)		<i>Gazebo winter</i>	<i>Gazebo summer</i>	<i>Mountain</i>	<i>Haupt-gebaude</i>	<i>Wood autumn</i>	<i>Wood summer</i>
CELLO-3D [14]	Trans	0.1	0.2	-	0.3	0.1	0.1
<i>Brossard et al.</i> [6]	Trans	1.8	1.0	1.2	1.8	1.2	1.5
Ours (Finetuned)	Trans (A)	0.83	1.28	0.87	1.83	0.93	0.69
CELLO-3D [14]	Rot	0.2	0.2	-	0.2	0.3	0.3
<i>Brossard et al.</i> [6]	Rot	3.7	2.3	1.2	2.9	4.2	4.7
Ours (Finetuned)	Rot (A)	0.71	1.03	0.81	1.48	1.69	1.65

TABLE VI: Trajectory Mahalanobis distance in terms of translation and rotation for the ETH dataset using aleatoric only (A) uncertainty, compared to two state-of-the-art approaches. Epistemic uncertainty is not showed after finetuning.

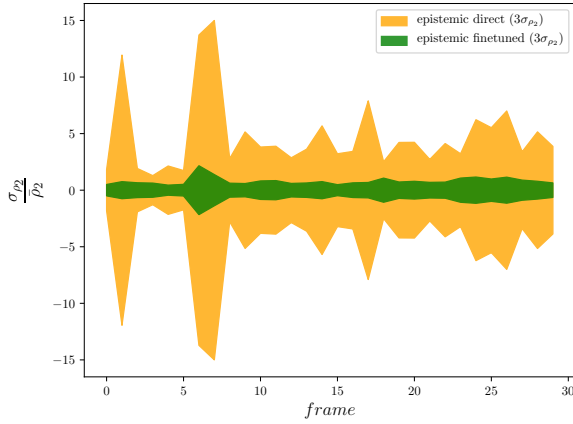


Fig. 4: Epistemic uncertainties normalized by the mean in the direct and finetuned case.

finetuning, not only in the expected improved results, but also in the magnitude of the epistemic uncertainty. The improvement brought by considering the model uncertainty is evident when testing the model trained on KITTI directly on an unseen dataset, where, as expected, the network has little confidence in its prediction for new scenarios. In fact, while some outdoor features may recur in the ETH dataset, motion type and scene size are remarkably different when compared to the KITTI dataset. After finetuning, not only the weight of the epistemic covariance is greatly reduced as shown in Fig. 4, but the overall results are closer to the

optimal value.

Finally, Table VI compares our results against CELLO-3D [14] and *Brossard et al.* [6] on aggregated transforms on the same dataset. It is worth noticing that given the strong focus on ICP initialization from *Brossard et al.*, it is more relevant to compare our approach to CELLO-3D, as our work solely focus on the errors introduced directly and indirectly by the data on a purely data-driven approach. CELLO-3D is often over pessimistic while *Brossard et al.* is generally optimistic. Especially on rotational uncertainty, our method outperforms the two approaches by a significant margin.

V. CONCLUSION

We presented an approach to estimate faithful covariances for the ICP scan registration process. Using a data-driven paradigm we have shown that it is possible to learn uncertainty for the selected algorithm, and to generalize to a certain degree on multiple datasets. Leveraging a neural network architecture conceived to directly process point clouds, we estimate covariances with a close connection to the data used in the ICP algorithm. Standard metrics show the reliability of our approach, and that it generally outperforms the state-of-the-art. Future improvements will concern the network architecture, possibly tailoring it to the scan registration problem. A possible approach is to have the flow embedding layer (and the successive layers) doubled to account for both point clouds motions, whereas in the scene flow the problem is cast on the reference cloud only. Additionally, the role of initialization and the use of the initial guess in the learning process can also be considered.

REFERENCES

- [1] J. Zhang and S. Singh, "LOAM: Lidar Odometry and Mapping in Real-time,," in *Robotics: Science and Systems*, 2014.
- [2] E. Mendes, P. Koch, and S. Lacroix, "ICP-based pose-graph SLAM," in *2016 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, pp. 195–200, IEEE, 2016.
- [3] P. Besl and N. McKay, "A Method for Registration of 3-D Shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 129–256, 1992.
- [4] F. Pomerleau, F. Colas, and R. Siegwart, "A review of point cloud registration algorithms for mobile robotics," *Foundations and Trends® in Robotics*, vol. 4, no. 1, pp. 1–104, 2015.
- [5] A. Censi, "An accurate closed-form estimate of ICP's covariance," in *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3167–3172, IEEE, 2007.
- [6] M. Brossard, S. Bonnabel, and A. Barrau, "A new approach to 3D ICP covariance estimation," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 744–751, 2020.
- [7] F. Pomerleau, A. Breitenmoser, M. Liu, F. Colas, and R. Siegwart, "Noise characterization of depth sensors for surface inspections," in *2012 2nd International Conference on Applied Robotics for the Power Industry (CARPI)*, pp. 16–21, IEEE, 2012.
- [8] J.-E. Deschaud, "IMLS-SLAM: scan-to-model matching based on 3D data," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2480–2485, IEEE, 2018.
- [9] J. Laconte, S.-P. Deschênes, M. Labussière, and F. Pomerleau, "Lidar measurement bias estimation via return waveform modelling in a context of 3D mapping," in *2019 International Conference on Robotics and Automation (ICRA)*, pp. 8100–8106, IEEE, 2019.
- [10] A. G. Buch, D. Kraft, *et al.*, "Prediction of ICP pose uncertainties using Monte Carlo simulation with synthetic depth images," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 4640–4647, IEEE, 2017.
- [11] S. M. Prakhya, L. Bingbing, Y. Rui, and W. Lin, "A closed-form estimate of 3D ICP covariance," in *2015 14th IAPR International Conference on Machine Vision Applications (MVA)*, pp. 526–529, IEEE, 2015.
- [12] N. Gelfand, L. Ikemoto, S. Rusinkiewicz, and M. Levoy, "Geometrically stable sampling for the icp algorithm," in *Fourth International Conference on 3-D Digital Imaging and Modeling, 2003. 3DIM 2003. Proceedings.*, pp. 260–267, IEEE, 2003.
- [13] S. Bonnabel, M. Barczyk, and F. Goulette, "On the covariance of ICP-based scan-matching techniques," in *2016 American Control Conference (ACC)*, pp. 5498–5503, IEEE, 2016.
- [14] D. Landry, F. Pomerleau, and P. Giguere, "CELLO-3D: Estimating the Covariance of ICP in the Real World," in *2019 International Conference on Robotics and Automation (ICRA)*, pp. 8190–8196, IEEE, 2019.
- [15] S. Wang, R. Clark, H. Wen, and N. Trigoni, "DeepVO: Towards end-to-end visual odometry with deep recurrent convolutional neural networks," in *Robotics and Automation (ICRA), 2017 IEEE International Conference on*, pp. 2043–2050, IEEE, 2017.
- [16] S. Wang, R. Clark, H. Wen, and N. Trigoni, "End-to-end, sequence-to-sequence probabilistic visual odometry through deep neural networks," *The International Journal of Robotics Research*, vol. 37, no. 4-5, pp. 513–542, 2018.
- [17] K. Liu, K. Ok, W. Vega-Brown, and N. Roy, "Deep inference for covariance estimation: Learning Gaussian noise models for state estimation," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1436–1443, IEEE, 2018.
- [18] A. De Maio and S. Lacroix, "Simultaneously Learning Corrections and Error Models for Geometry-based Visual Odometry Methods," *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 6536–6543, 2020.
- [19] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "Pointnet: Deep learning on point sets for 3d classification and segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 652–660, 2017.
- [20] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "Pointnet++: Deep hierarchical feature learning on point sets in a metric space," *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017.
- [21] H. Thomas, C. R. Qi, J.-E. Deschaud, B. Marcotegui, F. Goulette, and L. J. Guibas, "Kpconv: Flexible and deformable convolution for point clouds," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 6411–6420, 2019.
- [22] C. R. Qi, O. Litany, K. He, and L. J. Guibas, "Deep Hough voting for 3D object detection in point clouds," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 9277–9286, 2019.
- [23] Y. Cho, G. Kim, and A. Kim, "Deeplo: Geometry-aware deep lidar odometry," *arXiv preprint arXiv:1902.10562*, 2019.
- [24] Z. Li and N. Wang, "Dmlo: Deep matching lidar odometry," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 6010–6017, IEEE, 2020.
- [25] Q. Li, S. Chen, C. Wang, X. Li, C. Wen, M. Cheng, and J. Li, "Lo-net: Deep real-time lidar odometry," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8473–8482, 2019.
- [26] J. Sola, J. Deray, and D. Atchuthan, "A micro Lie theory for state estimation in robotics," *arXiv preprint arXiv:1812.01537*, 2018.
- [27] T. D. Barfoot and P. T. Furgale, "Associating uncertainty with three-dimensional poses for use in estimation problems," *IEEE Transactions on Robotics*, vol. 30, no. 3, pp. 679–693, 2014.
- [28] A. Kendall and Y. Gal, "What uncertainties do we need in bayesian deep learning for computer vision?," in *Advances in neural information processing systems*, pp. 5574–5584, 2017.
- [29] M. Stephan, M. D. Hoffman, D. M. Blei, *et al.*, "Stochastic Gradient Descent as Approximate Bayesian Inference," *Journal of Machine Learning Research*, vol. 18, no. 134, pp. 1–35, 2017.
- [30] J. Postels, F. Ferroni, H. Coskun, N. Navab, and F. Tombari, "Sampling-free epistemic uncertainty estimation using approximated variance propagation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 2931–2940, 2019.
- [31] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning," in *Proceedings of the 33rd International Conference on Machine Learning (ICML)*, 2016.
- [32] X. Liu, C. R. Qi, and L. J. Guibas, "FlowNet3D: Learning scene flow in 3D point clouds," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 529–537, 2019.
- [33] F. Pomerleau, S. Magnenat, F. Colas, M. Liu, and R. Siegwart, "Tracking a Depth Camera: Parameter Exploration for Fast ICP," in *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 3824–3829, IEEE Press, 2011.
- [34] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3354–3361, IEEE, 2012.
- [35] F. Pomerleau, M. Liu, F. Colas, and R. Siegwart, "Challenging data sets for point cloud registration algorithms," *The International Journal of Robotics Research*, vol. 31, no. 14, pp. 1705–1711, 2012.