30th International Symposium on Space Flight Dynamics (ISSFD) – Uncertainty-Aware Visual Perception for Spacecraft Proximity Operations

Massimiliano Bussolino[(1)], Stefano Silvestrini[(1)], Michèle Lavagna[(1)]

[(1)]Politecnico di Milano, Via La Masa 32, Milan, Italy
Email: {massimiliano.bussolino; stefano.silvestrini; michelle.lavagna}@polimi.it

***Abstract* – Vision-based navigation is critical for autonomous spacecraft operations, and while standard deep learning methods are proving exceptional perception capabilities, their lack of interpretability and uncertainty quantification poses significant limitation to their utilization. To address this limitation, this paper presents a novel Deep Kernel Learning (DKL) architecture that merges the representational power of convolutional encoders with the rigorous statistical framework of Gaussian Processes. The proposed pipeline jointly estimates the relative Line-of-Sight (LOS) and monocular range to an uncooperative target (Envisat) from a single image, explicitly regressing both aleatoric and epistemic uncertainties. Experimental validation demonstrates that the network effectively quantifies epistemic predictive confidence, successfully conditioning its latent space to separate informative data from degraded inputs, while maintaining generalizability to real laboratory images. The effectiveness of the uncertainty quantification is validated through its integration into a navigation filter. The results demonstrate a robust capability to mitigate the impact of visual degradations, such as solar flares, while concurrently highlighting the critical need to transition toward a heteroscedastic noise model.**

I. INTRODUCTION

The rapid development of deep learning has profoundly transformed image analysis in various domains. In particular, convolutional neural networks (CNNs) allowed images to be encoded into an informative set of latent variables that capture hierarchical spatial and semantic features, enabling robust performance in tasks such as object recognition, image segmentation, and scene understanding. Despite their success, CNNs, and deep learning more broadly, share the limitation of providing only a point estimate without reliable measures of predictive uncertainty. This shortcoming is particularly critical in the domain of spacecraft proximity operations, a safety-critical scenario in an operational environment that cannot be faithfully reproduced on the ground for training or testing. In such settings, it is essential to quantify the epistemic uncertainty of the model to detect when it encounters input outside of its training distribution, allowing a safer and more reliable deployment [1]. Moreover, estimating the uncertainty of computer vision measurements is fundamental when such estimates are later fused into a navigation filter. In this context, the measurement noise covariance matrix transitions from being arbitrarily set to becoming adaptive, leading to more reliable state estimation.

For the interested reader, a review of uncertainty quantification methods in deep learning can be found in [2,3]. Uncertainty in deep learning is commonly quantified through Bayesian approaches. Bayesian Neural Networks (BNNs) provide a principled framework by placing distributions over network weights, though they are often computationally expensive in practice. To address this, approximate methods such as Monte Carlo (MC) dropout offer a scalable Bayesian alternative, while deep ensembles serve as a strong non-Bayesian approach.

Although deep ensembles are expected to provide more consistent results [4], both methods suffer the additional computational cost of multiple forward passes.

An alternative class of probabilistic models are Gaussian Processes (GPs), which provide a principled framework for capturing both predictive mean and uncertainty, specified by the mean function and kernel respectively [5]. These models demonstrated the ability to quantify uncertainty in intrinsically epistemic incomplete scenarios such as intermittent time series analysis [6]. However, a single-layer GP with a fixed kernel may lack the representational power to effectively model complex, hierarchical data structures. To increase this representational power, Deep Gaussian Processes (DGPs) were proposed, creating a hierarchical, deep structure inspired by deep neural networks. The limitation of GPs and DGPs is their cubic scaling, or $\mathcal{O}(n^3)$, with the number of training data points n, making them impractical for large-scale applications such as real time visual perception. This computational bottleneck motivates a different approach: Deep Kernel Learning (DKL) [7]. DKL is a hybrid model that combines the feature extraction power of a deep neural network with the probabilistic framework of a Gaussian Process. The neural network acts as a deep kernel to transform raw data into a compact, meaningful representation. A standard GP is then applied to these learned features.

GPs and DKL have shown promise in predicting system behaviour with quantified uncertainty. In the mobility systems domain, existing work has largely focused on guidance and control, including optimal control prediction [8], multi-agent safe planning [9], and vehicle

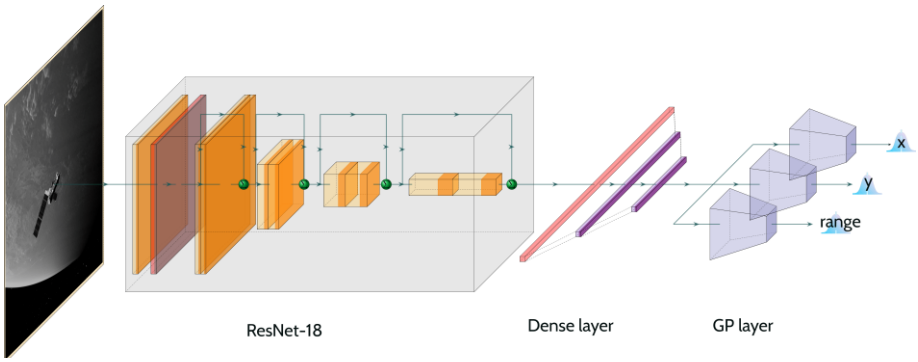

Fig. 1. *Proposed network architecture*

dynamics estimation [10]. To the best of the authors' knowledge, DKL has not yet been applied to spacecraft optical navigation. Relevant applications of DKL for uncertainty quantification in image-based tasks are in medical imaging [11,12].

This paper applies Deep Kernel Learning (DKL) to spacecraft computer vision, focusing on close-proximity operations with a known target. The proposed architecture estimates line-of-sight and range from single monocular images while explicitly quantifying uncertainty in unseen scenarios. The study highlights both the benefits of DKL for uncertainty-aware perception and the current limitations that remain to be addressed.

II. ARCHITECTURE

A DKL architecture is composed of a feature encoder, a convolution encoder for the computer vision case, followed by a GP layer, whose structure is reported in Eq. (1).

$$f(x) \sim \mathcal{GP}\big(m(x),\, k(x,x')\big) \tag{1}$$

In this formulation, $m(\cdot)$ is the mean function, $k(\cdot,\cdot)$ the kernel, and x the latent feature output of the encoder.

For exact GP models, the computational bottleneck is the computation of the inverse of the covariance matrix of the input, which scales cubically with the size of the training set n.

To overcome this limitation, the stochastic variational DKL formulation was introduced in [13], leveraging a set of learnable inducing points z of dimensionality $m \ll n$ to summarize the training set and reduce the computational complexity to $\mathcal{O}(nm^2)$.

The inducing points serve both to make the GP tractable and to regularize the model by constraining the variational posterior to remain close to the GP prior.

The objective function for training is the maximization of the Evidence Lower Bound (ELBO), defined in Eq.(2), where y are the training targets, f the latent function values, and u the inducing variables (i.e., $f(z)$).

$$\mathcal{L}_{ELBO} = \mathbb{E}_{q(f)}[\log p(y \mid f)] - \mathrm{KL}(q(u) \parallel p(u)) \tag{2}$$

In Eq. (2) the first term, i.e. the log-likelihood, is representative of the performance of the mean function in estimating the observable values, while the second term, the Kullback-Leibler (KL) divergence between the prior and posterior variational distribution, acts as a regularization term.

A major limitation of standard DKL models is their tendency to become overconfident when presented with out-of-distribution (OOD) inputs, leading to posterior variances that underestimate true uncertainty. This is partially due to the pathological behaviour of DKL itself [14]. This pathological behaviour is driven by the high representational power of the neural network encoder, which, if the KL regularization term is not effective enough, learns a degenerate feature map. This can manifest as either a latent space collapse, where the training data are projected into a very small, dense region, or as a sparse latent space, where many dimensions are near zero for a given input. Both scenarios result in a latent space that is no longer informative of the original image, making it useless for understanding OOD samples and leading to a complete breakdown of the model's uncertainty quantification. This problem was mitigated in the proposed work through the design of the network itself, through training and pre-training techniques and latent-space structuring losses.

To validate the robustness of this proposed methodology, a DKL network designed for the estimation of line-of-sight (LOS) and range of an artificial spacecraft is presented as a representative task for vision perception applied to spacecraft vision-based navigation.

The first component of the model is a convolutional encoder, specifically a ResNet-18 [15], chosen for its favourable trade-off between model size and performance.

Although prior work [16] has shown that Gaussian processes can handle high-dimensional inputs, their practical performance is typically limited to input dimensions on the order of a few hundred, significantly lower than the dimensionality of the last layer of ResNet-18. To address this, the output of the

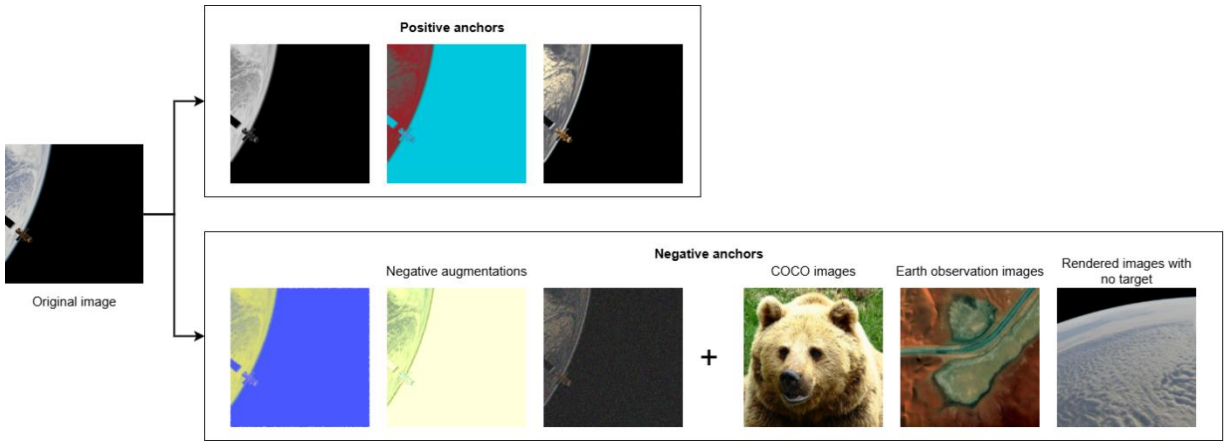

Fig. 2. *Example of training batches for positive and negative anchors*

convolutional encoder is first passed through to two fully connected layer to reduce the dimensionality to 16 features (see Fig.1). The choice of 16 dimensions represents a balance between preserving sufficient representational capacity and preventing the GP from operating on excessively high-dimensional inputs, which could degrade training or overfit the data. Finally, the last layer of the dense network serves as the input to three independent GPs. Spectral normalization is added to the intermediate dense layers for regularization.

Each GP employs a Gaussian likelihood, a constant mean function, and a Matérn kernel ($\nu = 5/2$) over 512 inducing points. Unlike the standard Radial Basis Function (RBF), the Matérn kernel features heavier tails. This property allows a smoother uncertainty gradient for out-of-distribution samples.

III. Training

The network training consists of two stages: first, a pre-training of the encoder where the GP head is temporarily replaced by an additional dense layer, and second, the training of the full DKL architecture utilizing the pre-trained encoder. This decoupled approach is implemented to ensure that the GP operates on an already structured embedding space, since the joint optimization of a highly parameterized convolutional encoder and the covariance hyperparameters of a Gaussian Process from scratch often leads to severe gradient instability. This section presents the dataset used for the training process, as well as the losses and most relevant hyperparameters.

A. Training dataset

The dataset used for both the pre-training and training of the network consists in 10.000 RGB images of the Envisat space debris, rendered with the Blender-based rendering environment available in the Astra research group [17]. The dataset covers a camera to target distance ranging from 40 to 200 m and a phase angle from 20 to 90 degrees. The images are rendered with a background realistic for a LEO (Low Earth Orbit) orbiting satellite, such as the Earth and a stellar field. While the original dataset consists of images rendered at 960×960 pixels with a focal length of 1182 pixels, all inputs were resized to a resolution of 448×448 pixels before being fed into the network to reduce computational overhead during training. Both to overcome the limited number of images in the dataset, and to avoid overfitting to the specific rendering output, limiting the generalization to real images, several augmentations were included during the training. Only geometry preserving augmentations where added, such as grayscale conversion, colour jittering, random inversion, solarization, and posterization (Fig. 2). Furthermore, optical degradations like Gaussian blur and sharpness adjustments were introduced, alongside the injection of standard Gaussian, spatially correlated pink, and salt-and-pepper noise.

B. Encoder pre-training

For the pretraining the GP head is temporarily substituted with a dense layer that maps the 16-dimensional latent space into the three outputs of the network: normalized centre of mass position on the image and range in kilometres, in order for the inputs to be in the range from 0 to 1. The loss during the pre-training is composed of three contributions. Firstly, the Mean Square Error (MSE) with respect to the ground truth LOS and range is considered, so to have the encoder parameters learn a representation apt to the task that it must learn.

More interestingly, two additional loss functions are employed to explicitly structure the latent representation. Specifically, contrastive learning is implemented, as proposed in [12], to enforce a clear separation between two distinct manifolds: one mapping in-distribution images, where the target geometry is structurally visible, and another mapping out-of-distribution (OOD) regions where the target is severely degraded or completely absent. This spatial separation is of paramount importance to allow the downstream GP head to flag images where the target is not clearly

identifiable, thereby increasing epistemic uncertainty and preventing latent space collapse.
However, the main risk with this method is inadvertently training the network to become a highly specific classifier for background features or visual artifacts, rather than a general geometric detector. To prevent the encoder from simply memorizing a specific type of negative anchor, the selection was designed to be as general as possible, focusing on the fundamental contrast between geometric presence and absence. For this reason, the negative dataset encompasses a wide structural variety, including inputs with extreme levels of noise and blur, elastic transformations, completely empty rendered scenes, real Earth observation imagery [18], and complex terrestrial backgrounds sampled from the COCO dataset [19], as shown in Fig. 2.
Since the prior of the GP head is a zero-mean Gaussian distribution, the encoder pre-training must provide an initial condition that approximates this topology. Failing to do so would result in an initial KL divergence penalty so severe that it would immediately destabilize the GP's covariance optimization, leading to the gradient instability issues previously discussed. For this reason, the third loss contribution is the SigReg loss [20], which, although not enforcing gaussianity, explicitly bounds and regularizes the variance of the latent embeddings, preventing the feature space from either collapsing into a single dense point or expanding unboundedly.
The pre-training phase is optimized using a learning rate of 1e-5, which undergoes a step decay by a factor of 0.5 every 30 epochs. To prevent the contrastive loss gradients from dominating the optimization process, a scaling factor of 0.2 is applied. The evolution of the training and validation losses is illustrated in Fig.3.
All objective components converge stably, with the contrastive penalty exhibiting the sharpest relative decline. Although the Mean Squared Error (MSE) plateaus near 0.1, indicating coarse line-of-sight and range estimations, this is considered acceptable. The objective of pretraining is strictly the geometric conditioning of the latent space, rather than high-fidelity spatial regression.

C. DKL training

In the second stage of training, the temporary dense layer is discarded, and the three independent GP heads (one per output) are attached to the 16-dimensional bottleneck of the pre-trained encoder. The entire architecture is then fine-tuned end-to-end by maximizing the ELBO. The ELBO objective inherently balances two critical terms: the expected log-likelihood, which minimizes the predictive regression error, and the KL divergence, which regularizes the variational distribution against the GP's exact prior. To ensure the network maintains its structural representation and preserves the capability to identify OOD inputs during this fine-tuning phase, the contrastive loss is maintained alongside the ELBO, with a reduced weighting factor.

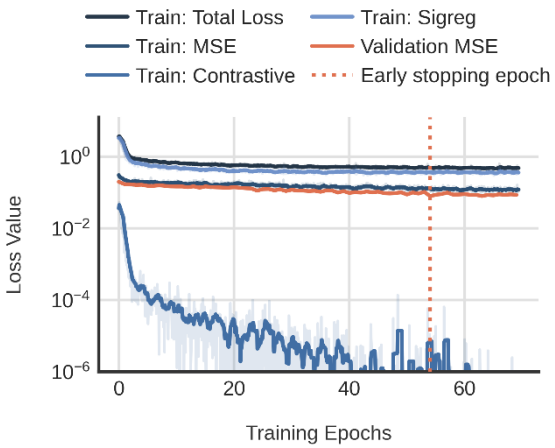

Fig. 3. *Pre-training losses*

To prevent this auxiliary loss from interfering with the GP covariance optimization, the repulsion mechanism is transitioned from a quadratic penalty to a linear ReLU activation. The spatial locations of the inducing points are initialized by sampling the latent feature distribution generated during the final epoch of pretraining. To optimize the end-to-end architecture, a differential learning rate strategy is implemented: the GP parameters are trained at a learning rate of 1e-3, the inducing point locations at 1e-2, and the pre-trained convolutional encoder is strictly constrained to 1e-5. This approach permits the rapid spatial repositioning of the inducing points across the evolving latent while preventing high gradients from disrupting the structural representations already learned by the encoder. Fig.4 illustrates the convergence profiles of both the ELBO objective and the Mean Absolute Error (MAE) across the three target outputs during training and validation. As detailed in Tab.1, the architecture achieves an average spatial regression error of 10 pixels across both image-plane dimensions, alongside a range estimation error of 7 meters. Towards the last training epochs, overfitting symptoms are observable, which are contained by the early stopping condition.

IV. Results and Discussion

To validate the performance of the proposed methodology, the experimental campaign is structured into three primary analyses. Initially, the geometric conditioning of the network's internal representation is investigated. Secondly, the challenge of the sim-to-real domain gap is addressed. This analysis is particularly crucial since if the constraints between positive and negative anchors are not correctly framed, the risk of under-confidence when the network encounters out-of-distribution real imagery is significant. Finally, the complete architecture is evaluated within a closed-loop navigation scenario.

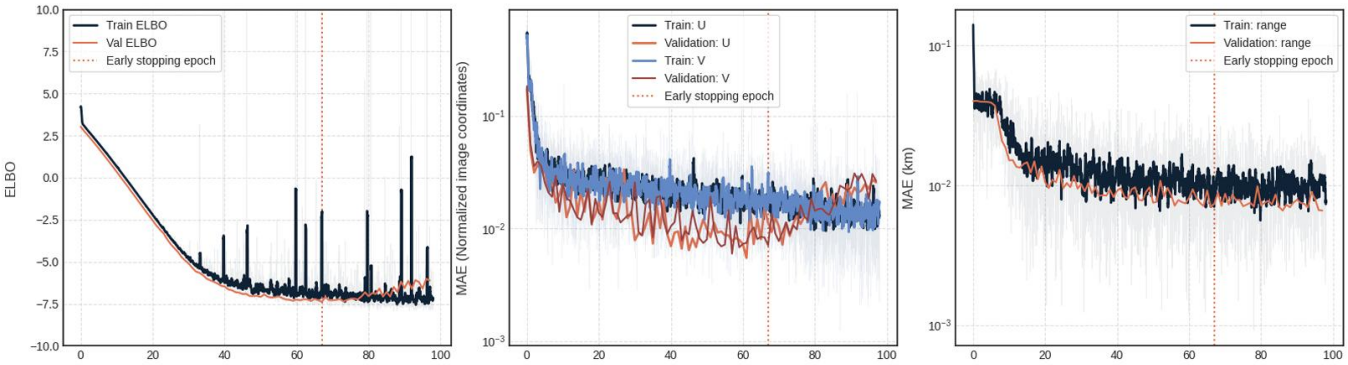

Fig. 4. *ELBO evolution (left) and Mean Absolute Error for the LOS (centre) and range (right) outputs during training*

A. Latent space dynamics

To investigate whether the internal representation of the DKL model respects the conditions imposed during training, the proposed method is compared against a baseline architecture trained only with the MSE loss. Specifically, a reference image is gradually degraded and passed through both trained encoders. Two types of degradation are analysed: the application of random Gaussian noise (with a standard deviation ranging from 0 to 0.5) and the simulation of a reflective flare of increasing intensity, shown in Fig. 5.

For both degradation processes, the L2 distance of the resulting embedding is plotted in Fig.5, alongside the boundary containing the in-distribution (ID) training inputs. When using only the MSE loss, the feature collapse described in Section II is clearly visible; the encoder projects all inputs into the same constrained region of the embedded space. Consequently, heavily degraded images remain tightly grouped with valid images, making it impossible for the covariance kernel to identify them as OOD.

In contrast, the proposed training architecture forces degraded images to move into distant regions of the embedding space, correctly triggering an increase in epistemic uncertainty. As shown in the figure, the noise degradation, which is a condition seen during training, is pushed to an L2 distance of 4 from the ID manifold, consistent with the contrastive margin imposed during training. The flare degradation, which represents a completely novel scenario for the network, is also successfully pushed outside the ID bounds, although with a lower magnitude. This demonstrates the capability of the proposed DKL framework to recognize unseen OOD scenarios by relying purely on the learned geometry of the target space. The same conclusions can be drawn from Fig.5, which visualizes the latent space projected into two dimensions via Principal Component Analysis (PCA). The plot clearly demonstrates that the ID and out-of-distribution OOD data occupy two distinct, separable regions. Furthermore, as the degradation is progressively applied, the embedded position of the reference image is driven outward from the ID boundaries directly into the OOD region.

B. Sim-to-Real generalization

To verify that the DKL model does not overestimate uncertainty when processing real images, which contain actual camera noise and optical distortions that complicate the sim-to-real domain gap in spacecraft vision-based navigation, the architecture was tested on images captured in the Argos laboratory of the Astra research group [21].

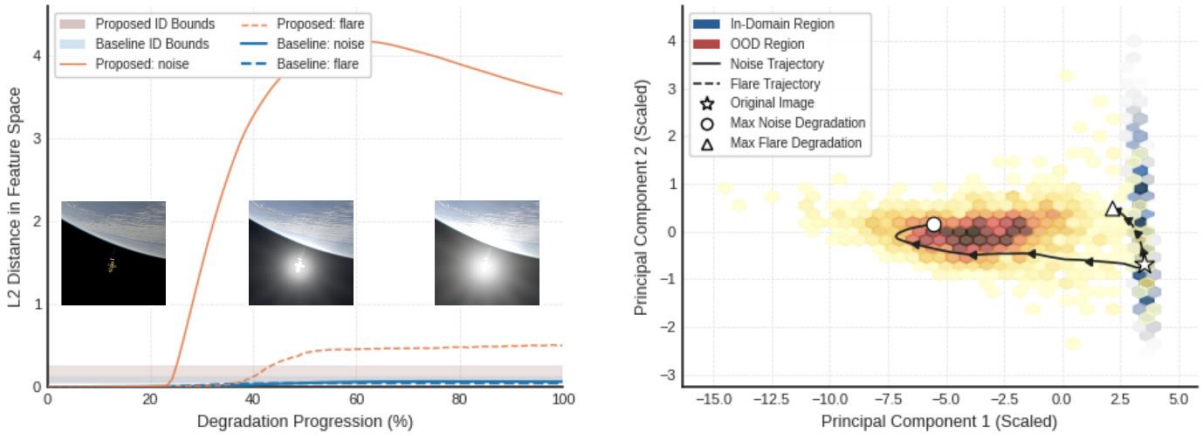

Fig. 5. *Latent space L2 distance of the degraded images (left), degraded images position in the PCA-reduced latent space (right)*

	MAE			Inferred Aleatoric Std. Dev.			Inferred Epistemic Std. Dev.		
	U (px)	V (px)	Range (m)	U (px)	V (px)	Range (m)	U (px)	V (px)	Range (m)
Validation	10.1	6.7	7.35	29.8	29.8	19.1	12.48	11.5	10.66
Lab images	4.6	4.7	17.9	29.8	29.8	19.1	11.5	11.5	10.45

Tab. 1. *MAE and inferred uncertainties for the validation set and the real laboratory images*

An initial set of 50 images of the Envisat mock-up was acquired using a FLIR Chameleon imager. Because range estimation strictly depends on camera geometry, these images were first modified to match the intrinsic camera matrix of the training set. They were then augmented through rotation, translation, and scaling to generate a final dataset of 200 images, covering the same 40 to 200-meter operational range as the training data.
An example of a laboratory image, along with its annotated estimations, is presented in Fig. 6. Finally, the performance in terms of Mean Absolute Error (MAE), alongside the estimated aleatoric and epistemic uncertainties, is compared between the synthetic validation set and the real images, with the normalized coordinates denormalized onto the original 960×960 images for the LOS and into meters for the range.

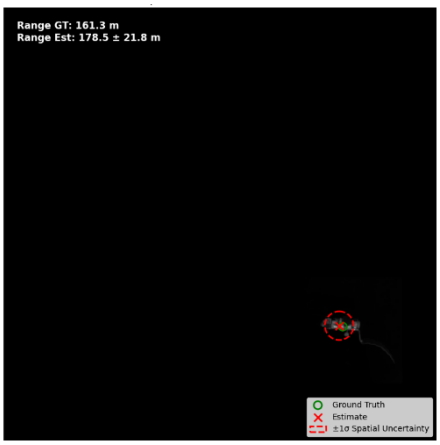

Fig. 6. *Real laboratory image with annotated estimations*

From an uncertainty perspective, the aleatoric variance remains identical across both the synthetic validation set and the real laboratory images. This is an expected as the current DKL framework employs a homoscedastic likelihood function that learns a global, input-independent observational noise parameter.
Regarding the epistemic uncertainty, the variations between the two domains are negligible. This indicates that the network does not treat the real-world sensor noise as an anomaly; rather, the encoder maps the physical laboratory images into the in-distribution ID region.
Interestingly, the MAE exhibits contrasting behavior across the state variables. The error for the Line-of-Sight (LOS) estimation (U, V) decreases on the real images, whereas the range estimation error significantly increases (from 7.35 m to 17.9 m). This divergence can be attributed to the specific conditions of the experimental setup. The improved LOS accuracy is likely a result of the controlled laboratory environment, where the uniform dark background and a restricted range of relative attitudes and illumination angles simplify feature extraction. The degradation in range accuracy highlights the sensitivity of monocular depth estimation. Because range relies entirely on apparent scale, even minor facility calibration errors, unmodeled optical distortions, or slight mismatches in the intrinsic camera matrix can degrade the depth estimation while leaving the 2D projected centroid (U, V) largely unaffected.

C. *Uncertainty aware navigation filtering*

Finally, the outputs of the neural network are evaluated within a navigation pipeline employing an Extended Kalman Filter (EKF). To establish an accurate ground-truth dynamic model, the relative trajectory was constructed by independently propagating the absolute orbits of the target and the chaser using Orekit [22], coupled with a custom attitude propagator. This reference trajectory dictated the relative camera poses used to render the synthetic 50000 image sequence at 0.1 Hz. To mitigate the domain gap inherent to perfectly rendered graphics, post-processing was applied to the images, including artificial blur and Gaussian noise ($\sigma = 0.05$) across all RGB channels. Furthermore, to rigorously test the filter's robustness, three specific adverse scenarios were injected into the trajectory: a period of amplified sensor noise from 1100 s to 1350 s, a simulated reflective flare from 3650 s to 3850 s, and a simulated loss of attitude tracking at 2500 s, which forces the target completely outside the camera's Field of View (FOV).
Fig. 7 illustrates the absolute error between the ground-truth measurements and the network's predictions, alongside the corresponding aleatoric and epistemic uncertainty contributions for both the Line-of-Sight (LOS) and range. Notably, the epistemic uncertainty correctly increases when the network encounters the three targeted degradation events along the trajectory. Furthermore, the aleatoric uncertainty demonstrates good calibration for the two LOS parameters.
On the other hand, the predictive uncertainty for the range is consistently underestimated. This discrepancy is partially due to the inherent difficulty of single-image depth estimation, but it primarily highlights a major limitation in the current architecture, i.e., the use of a homoscedastic likelihood model for a task with strictly input-dependent uncertainty. Because the spatial

sensitivity of monocular cameras naturally degrades at further distances, the true observational noise is heteroscedastic. By enforcing a globally constant noise parameter, the homoscedastic model inevitably fails to capture the growing uncertainty at larger ranges, while overestimating the uncertainty at lower distances.

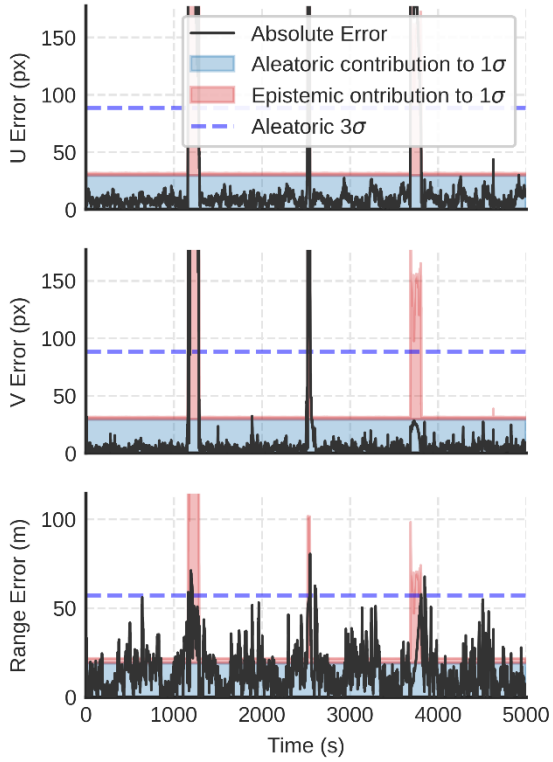

Fig. 7. *DKL measurements with estimated uncertainty*

The spatial measurements extracted by the DKL network are subsequently fused within the EKF. The filter's internal propagation relies on the classical Clohessy-Wiltshire (C-W) relative motion equations [23]. The process noise covariance matrix is parameterized as diagonal, with constant variances of 1e-4 and 1e-6 for the position and velocity states, respectively. The measurement noise covariance matrix is dynamically updated at each temporal step utilizing the predictive variances output by the DKL model.

To account for the underconfidence of the monocular range estimation detailed previously, the DKL-derived range variance is conservatively inflated by a factor of 3. Furthermore, a gating mechanism is introduced: incoming measurements are entirely discarded if the predicted epistemic uncertainty exceeds the 3σ bound of the corresponding aleatoric uncertainty. This specific condition serves as a threshold for identifying and isolating severely OOD or visually degraded inputs. The resulting absolute state errors, alongside the filter's estimated 3σ covariance bounds for the relative position in the Hill frame, are presented in Fig. 8.

The filter successfully converges, although it exhibits a significant periodic error caused by range estimation inaccuracies and the elliptical shape of the relative orbit. Because the range is unobservable in the ballistic CW equations when relying solely on LOS measurements, the filter cannot damp the intrinsic range measurement error, functioning primarily as a low-pass filter. However, the positive impact of the epistemic covariance is demonstrated: during periods of visual degradation, there is a corresponding spike in the measurement noise covariance. This allows the filter to rely heavily on the dynamic model when the visual measurements are untrustworthy, significantly reducing the impact of faulty data on the final state estimation.

The computational time for the processing of DKL remains bounded below 8 milliseconds on an RTX 4060 laptop GPU.

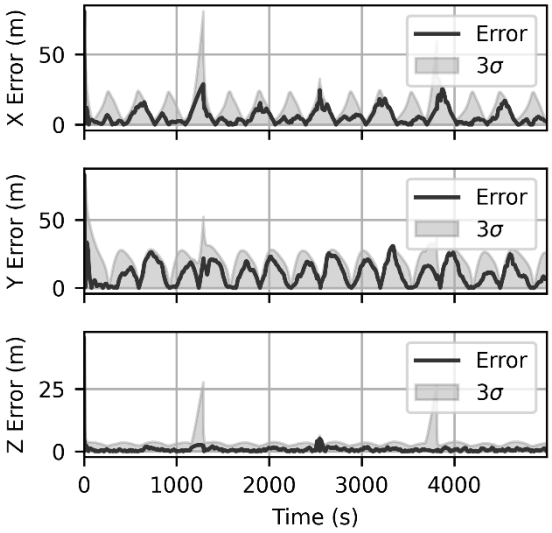

Fig. 8. *Filter position states with estimated covariance*

V. Conclusions

This paper presents a Deep Kernel Learning architecture that leverages the principled statistical framework of Gaussian Processes, coupled with the representational power of a convolutional encoder, to provide aleatoric and epistemic uncertainty quantification for vision-based perception in the context of spacecraft relative navigation. The proposed solution explores different training techniques that successfully condition the latent embedding of the input images, separating informative data from degraded inputs. The architecture is evaluated on the problem of estimating the Line-of-Sight (LOS) and range to the Envisat spacecraft from a single image. The conducted tests demonstrated the network's capability to estimate epistemic uncertainty when faced with unseen and degraded visual conditions, improving the overall robustness of the navigation pipeline while maintaining good generalization to non-synthetic images unseen during training. Nevertheless, the current architecture is limited by its use of a homoscedastic

likelihood model, which significantly affects the aleatoric uncertainty estimation of the range regression task. Future work will primarily focus on transitioning toward a heteroscedastic noise model, as well as expanding the test setup and scenarios to assess the robustness of the proposed method.

VI. References

[1] A. Kendall and Y. Gal, 'What uncertainties do we need in bayesian deep learning for computer vision?', *Advances in neural information processing systems*, vol. 30, 2017.

[2] M. Abdar, F. Pourpanah, S. Hussain, D. Rezazadegan, L. Liu, M. Ghavamzadeh, P. Fieguth, X. Cao, A. Khosravi, U. R. Acharya, V. Makarenkov, and S. Nahavandi, 'A Review of Uncertainty Quantification in Deep Learning: Techniques, Applications and Challenges', *Information Fusion*, vol. 76, pp. 243–297, Dec. 2021, doi: 10.1016/j.inffus.2021.05.008.

[3] J. Caldeira and B. Nord, 'Deeply Uncertain: Comparing Methods of Uncertainty Quantification in Deep Learning Algorithms', *Mach. Learn.: Sci. Technol.*, vol. 2, no. 1, p. 015002, Dec. 2020, doi: 10.1088/2632-2153/aba6f3.

[4] F. K. Gustafsson, M. Danelljan, and T. B. Schon, 'Evaluating scalable bayesian deep learning methods for robust computer vision', presented at the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops, 2020, pp. 318–319.

[5] C. E. Rasmussen, 'Gaussian Processes in Machine Learning', in *Advanced Lectures on Machine Learning*, vol. 3176, in Lecture Notes in Computer Science, vol. 3176. , Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 63–71. doi: 10.1007/978-3-540-28650-9_4.

[6] S. Damato, D. Azzimonti, and G. Corani, 'Forecasting intermittent time series with Gaussian Processes and Tweedie likelihood', *International Journal of Forecasting*, p. S0169207025000974, Nov. 2025, doi: 10.1016/j.ijforecast.2025.10.001.

[7] A. G. Wilson, Z. Hu, R. Salakhutdinov, and E. P. Xing, 'Deep kernel learning', presented at the Artificial intelligence and statistics, PMLR, 2016, pp. 370–378.

[8] H. Li, W. Yao, Y. Dong, Q. Gao, and Y. Deng, 'Deep Kernel-Based Optimal Control Prediction in Aerospace Missions', *IEEE Trans. Aerosp. Electron. Syst.*, vol. 58, no. 3, pp. 1621–1633, Jun. 2022, doi: 10.1109/TAES.2021.3133225.

[9] Z. Zhu, E. Biyik, and D. Sadigh, 'Multi-Agent Safe Planning with Gaussian Processes', in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Las Vegas, NV, USA: IEEE, Oct. 2020, doi: 10.1109/IROS45743.2020.9341169.

[10] J. Ning and M. Behl, 'Scalable Deep Kernel Gaussian Process for Vehicle Dynamics in Autonomous Racing'. In: *Proceedings of The 7th Conference on Robot Learning*

[11] M. Siebert, J. Graßhoff, and P. Rostalski, 'Uncertainty Analysis of Deep Kernel Learning Methods on Diabetic Retinopathy Grading', *IEEE Access*, vol. 11, pp. 146173–146184, 2023, doi: 10.1109/ACCESS.2023.3343642.

[12] Z. Wu, Y. Yang, J. Gu, and V. Tresp, 'Quantifying predictive uncertainty in medical image analysis with deep kernel learning', presented at the 2021 IEEE 9th international conference on healthcare informatics (ICHI), IEEE, 2021, pp. 63–72.

[13] A. G. Wilson, Z. Hu, R. R. Salakhutdinov, and E. Xing, 'Stochastic Variational Deep Kernel Learning', in *Advances in Neural Information Processing Systems*, vol. 29, 2016.

[14] S. W. Ober, C. E. Rasmussen, and M. van der Wilk, 'The promises and pitfalls of deep kernel learning', presented at the Uncertainty in Artificial Intelligence, PMLR, 2021, pp. 1206–1216.

[15] K. He, X. Zhang, S. Ren, and J. Sun, 'Deep residual learning for image recognition', Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[16] Z. Xu, H. Wang, J. Phillips, and S. Zhe, 'Standard gaussian process is all you need for high-dimensional bayesian optimization', presented at the International Conference on Learning Representations, 2025, pp. 94842–94862.

[17] M. Bechini, M. Lavagna, and P. Lunghi, 'Dataset generation and validation for spacecraft pose estimation via monocular images processing', *Acta Astronautica*, vol. 204, pp. 358–369, Mar. 2023, doi: 10.1016/j.actaastro.2023.01.012.

[18] D. Velazquez, P. R. López, S. Alonso, J. M. Gonfaus, J. Gonzalez, G. Richarte, J. Marin, Y. Bengio, and A. Lacoste, 'EarthView: A Large Scale Remote Sensing Dataset for Self-Supervision', Jan. 14, 2025, doi: 10.48550/arXiv.2501.08111.

[19] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, 'Microsoft coco: Common objects in context', presented at the European conference on computer vision, Springer, 2014, pp. 740–755.

[20] R. Balestriero and Y. LeCun, 'LeJEPA: Provable and Scalable Self-Supervised Learning Without the Heuristics', Nov. 14, 2025, doi: 10.48550/arXiv.2511.08544.

[21] M. Piccinin, S. Silvestrini, G. Zanotti, and A. Brandonisio, 'ARGOS: calibrated facility for Image based Relative Navigation technologies on ground verification and testing', IAC 2021.

[22] L. Maisonobe, B. Cazabonne, R. Serra, E. Ward, and M. Journot, *CS-SI/Orekit: 13.1.5*, Zenodo, May 2026.. doi: 10.5281/ZENODO.7249096.

[23] W. H. Clohessy and R. S. Wiltshire, 'Terminal Guidance System for Satellite Rendezvous', *Journal of the Aerospace Sciences*, vol. 27, no. 9, pp. 653–658, Sep. 1960, doi: 10.2514/8.8704.