

DANAE: a denoising autoencoder for underwater attitude estimation

Paolo Russo¹, Fabiana Di Ciaccio², Salvatore Troisi²

¹University of Rome La Sapienza, via Ariosto 25, Rome, paolo.russo@diag.uniroma1.it

²Parthenope University of Naples, Centro Direzionale Isola C4, Naples, ([fabiana.diciaccio](mailto:fabiana.diciaccio@uniparthenope.it), [salvatore.troisi](mailto:salvatore.troisi@uniparthenope.it))@uniparthenope.it

Abstract – One of the main issues for underwater robots navigation is their accurate positioning, which heavily depends on the orientation estimation phase. The systems employed to this scope are affected by different noise typologies, mainly related to the sensors and to the irregular noise of the underwater environment. Filtering algorithms can reduce their effect if opportunely configured, but this process usually requires fine techniques and time. In this paper we propose DANAE, a deep Denoising Autoencoder for Attitude Estimation which works on Kalman filter IMU/AHRS data integration with the aim of reducing any kind of noise, independently of its nature. This deep learning-based architecture showed to be robust and reliable, significantly improving the Kalman filter results. Further tests could make this method suitable for real-time applications on navigation tasks.

I INTRODUCTION

Localization is one of the most important tasks for unmanned robots, especially in underwater environments. Being a highly unstructured and GPS-denied environment, other than characterized by different noise sources and by the absence of man-made landmarks, the underwater setting provides more challenges for the orientation estimation.

In a typical configuration, the Euler angles representing the vehicle attitude (roll, pitch and yaw) are obtained through the integration of raw data acquired by the sensors embedded into an Inertial Measurement Unit (IMU), or in the more cost-effective Attitude and Heading Reference System (AHRS).

One of the most successful methods to perform this elaboration is based on the Kalman filter [1] (KF), in its linear and non linear versions [2]. Although known as the perfect estimator under some assumptions, the estimation provided by the KF strongly depends on a good knowledge of the error covariance matrices describing the noise affecting the system. Moreover, the on-line computation of these matrices is often required for anytime-varying or nonlinear system, squaring the number of necessary updating step at each time. Finally, the procedure employed to accurately fine tune the filter parameters is known to be unintuitive,

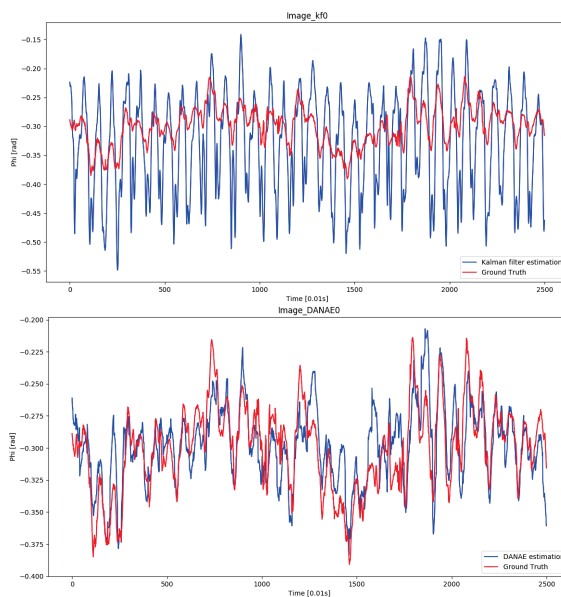


Fig. 1. KF (upside) and DANAE (downside) roll angle estimation compared to the ground truth. This experiment is made on a subsection of the slow walking set of OxIOD.

requiring specific settings for different scenarios [3].

In order to overcome these issues, we propose a deep learning-based approach which aims to attenuate any source of error from the attitude estimation of a KF. The strength of our proposed method stands in the ability of acting as a full-noise compensation model for both noise and bias errors, without the need to separately process each influencing factor.

Extensive tests performed on two different datasets to evaluate Euler angles show the power of our approach, with a sensible improvement of both mean squared error and max deviation w.r.t. ground truth data.

The remainder of this study is organized as follows: Section II presents a brief literature review for Kalman-based algorithms and deep learning techniques applied on similar tasks. Section III introduces the concepts of KF and of deep denoising autoencoder, followed by a concise description of our method. In Section IV the datasets used for the experiments are described. Section V illustrates the details of our architecture and the hyper parameters chosen for the training. Section VI analyzes the experiment results

through the use of standard metrics, concluding with Section VII which contains some final considerations on the topic, including future possible improvements.

II RELATED WORKS

The use of KF in robotic applications is ubiquitous. [4] developed a GPS/IMU multisensor fusion algorithm to increase the reliability of the position information, while [5] and [6] proposed an effective Kalman Filter which is able to exploit low-cost AHRS for efficient attitude estimation under various dynamic conditions, with an interesting underwater application developed by [7] and [8].

Non-linear implementations of the filter include for example [9], which gets the robot pose by fusing camera and inertial data with an Extended Kalman Filter (EKF), and [10] where the same task is accomplished using an Unscented Kalman Filter (UKF). For a detailed comparison between different kalman filters, see [2] or [11].

Beside the critical task of sensor fusions, the estimation of both the sensors biases and noises is also crucial for an effective navigation system. Nonetheless, Kalman filtering-based techniques constitute a powerful approach even to this problem. For example [12] exploited a kalman filter for accurate biases estimation in a distributed-tracking systems, while [13] identified and successfully removed the noise using a Kalman filter in a real time application.

The rise of Deep Learning has radically changed fields like Computer Vision and Natural Language Processing. Since the spectacular success of [14], Convolutional Neural Networks (CNN) produce state of the art accuracy on classification [15], detection [16] and segmentation [17] tasks, with Recurrent Neural Networks (RNN) being the backbone models for speech recognition [18] and sequence generation [19]. Autoencoders are another successful deep architectures where the aim is to reconstruct a signal by learning a latent representation from a set of data. They have been used for realistic text [20] and images [21] generation; however, one of the most successful use of autoencoders is for noise removal. Since their introduction [22], denoising autoencoders (DAE) have been used for a broad number of tasks like medical images improvement [23], speech enhancement [24] and ECG signal boost [25]. To the best of our knowledge, DANAE is the first application of denoising autoencoder on attitude estimation.

III METHOD

In this section we provide some basic concepts on linear KF and autoencoders followed by a brief overview of our method. However it should be emphasized that DANAE is filter agnostic and can be used seamlessly on non-linear KF (EKF, UKF) as well as any other type of filter able to perform attitude estimation.

A. Kalman filter

KF is a widely used algorithm for state estimation of dynamic systems since it is able to minimize the related variance under some perfect model assumptions.

The system behaviour in a discrete time setting can be described by a state equation 1 and a measurement equation 2:

$$x_t = Ax_{t-1} + Bu_{t-1} + v_{t-1} \quad (1)$$

$$y_t = Cx_t + w_t \quad (2)$$

where x_t is the state vector to be predicted, x_{t-1} and u_{t-1} are the state and the input vectors at the previous time step and y_t represents the measurement vector. A and B are the system matrices and C is the measurement matrix. The vectors v_{t-1} and w_t are respectively associated with the additive process noise and the measurement noise, assumed to be zero mean Gaussian processes. The final estimate is obtained by a first prediction step (3, 4) followed by the update phase (5, 6, 7, 8):

$$x'_t = Ax_{t-1} + Bu_{t-1} \quad (3)$$

$$P'_t = AP_{t-1}A^T + Q \quad (4)$$

$$K_t = P'_t C^T (C P'_t C^T + R)^{-1} \quad (5)$$

$$x_t = x'_t + K_t (y_t - C x'_t) \quad (6)$$

$$P_t = (I - K_t C) P'_t \quad (7)$$

The a-posteriori state estimate x_t is obtained as a linear combination of the a-priori estimate x'_t and a weighted difference between the actual and the predicted measurements called residual (see equation 6); the Kalman gain (K in Eq. 5) minimizes the a-posteriori error covariance (P in Eq. 7), initially set by the user. Finally, Q and R are the covariance matrices of the process and of the measurement noise. Q models the dynamics uncertainty, while R represents the sensors internal noises. These matrices heavily affect the final filter performance, thus a tricky tuning process is necessary to correctly estimate noises statistics. A proper fine-tuning is also important for sensors biases estimation; however, even in this case traditional approaches based on the KF suffer from implementation complexity and require non-intuitive tuning procedures [26].

B. Denoising autoencoder

DAE is a deep convolutional model which is able to recover clean, undistorted output starting from partially corrupted data as input. In the original implementation, the input data is intentionally corrupted through a stochastic mapping:

$$\tilde{x} \sim q_D(\tilde{x}|x) \quad (8)$$

Then, the corrupted input is mapped into an hidden representation as in the case of a standard autoencoder:

$$h = f_\theta(\tilde{x}) = s(W\tilde{x} + b) \quad (9)$$

Finally, the hidden representation is mapped back to a reconstructed signal:

$$\hat{x} = g_{\theta'}(h) \quad (10)$$

During the training procedure, the output signal is compared with a reference signal in order to minimize the L2 reconstruction error:

$$\mathcal{L}(x - \hat{x}) = \|x - \hat{x}\|^2 = \|x - s(W\tilde{x} + b)\|^2 \quad (11)$$

In our implementation, DANAE takes as input the noisy angles prediction performed by the KF and as reference signal the ground truth angles provided by the dataset. For this reason, we underline that our method is able to remove both stochastic errors (e.g. electromagnetic- and thermo-mechanical- related ones), and systematic errors (due for example to sensors misalignment).

IV DATASETS

DANAE has been developed and tested on the Oxford Inertial Odometry Dataset (OxIOD) [27] and on the Underwater Caves Sonar dataset (to which we will refer as UCS) [28].

OxIOD has been chosen for its accurate ground truth measurements over big heterogeneous settings. Developed for deep learning-based inertial-odometry navigation, OxIOD provides 158 sequences (for a total of 42.587 km) of inertial and magnetic field data acquired from low-cost sensors. Five users made indoor and outdoor acquisitions while normally walking with phone in hand, pocket and handbag and slowly walking, running and performing mixed motion modes. Different smartphones were used to acquire data, but the major part was collected by an iPhone 7 plus equipped with an InvenSense ICM20600. A Vicon motion capture system was used to get the ground truth with a precision down to 0.5 mm.

The UCS dataset is collected by a Sparus AUV navigating in the underwater cave complex "Coves de Cala Viuda" in Spain. The vehicle explored two tunnels, closing a 500m-long path at a depth of approximately 20m. Among the equipped sensors (e.g. DVL, sonar, etc), a standard low cost Xsens MTi AHRS and an Analog Devices ADIS16480 are mounted. The latter is a 10 DOF MEMS which provides more accurate raw sensors measurements and dynamic orientation outputs (obtained by their EKF fusion). The elaboration of images containing six traffic cones placed on the seabed allowed the relative positioning of the vehicle. Unfortunately, the ground truth thus obtained is synchronized with the low-rate camera acquisitions, making the comparison with the high-rate IMU measurements inconsistent. For this reason, we assumed that the orientation directly provided by the AHRS could at first glance substitute the true ground truth. Despite not being a proper solution to the issue, this choice allowed

Table 1. OxIO Dataset: performance of KF and DANAE vs GT.

KF	ϕ	θ	ψ
Mean dev. [rad]	0.0661	0.0483	1.9518
Max dev. [rad]	0.2929	0.2134	9.0145
RMSE [rad]	0.0815	0.0600	2.4000
DANAE	ϕ	θ	ψ
Mean dev. [rad]	0.0224	0.0157	0.7392
Max dev. [rad]	0.1382	0.1082	5.7907
RMSE [rad]	0.0282	0.0196	1.3194

Table 2. UCS Dataset: performance of KF and DANAE vs GT. Since the GT value of ψ is not reliable, the corresponding results are not reported here.

KF	ϕ	θ	ψ
Mean dev. [rad]	0.0326	0.0328	-
Max dev. [rad]	0.1476	0.1751	-
RMSE [rad]	0.0410	0.0412	-
DANAE	ϕ	θ	ψ
Mean dev. [rad]	0.0139	0.0147	-
Max dev. [rad]	0.0671	0.0769	-
RMSE [rad]	0.0177	0.0190	-

us to understand the ability of DANAE to work in a true underwater scenario with its unique features.

V EXPERIMENTAL SETUP

Some details on the experiments will be given in this section. Both datasets have been split in a training and test set: in the case of OxIOD, for each setting we used the run 1 as test set, leaving all the other sessions as training set. UCS provides instead a single file for each system containing all the measurements stored during the entire survey. We then decided to split the data, using the first 80% to train DANAE and the remaining 20% to test the performance.

Three main phases can be distinguished: during the first one, the inertial and magnetic field data are integrated with a linear KF, providing the estimation of the three Euler angles. In the second phase, these outputs are fed to the DANAE module for training, while on the third phase tests are performed using a pipeline of KF and DANAE. All the code has been developed in Python 3.6.9 running on Ubuntu 18.04, with the help of Pytorch framework.

A. KF initialization

We implemented the filter in its most basic formulation following the equations from 3 to 7. The covariance matrices P, Q and R have been initialized as identity matrix,

and no tuning has been done with relation to both the internal system and the measurements noises. The elaboration of accelerometer and magnetometer raw data provided the measurements vector (see Cx_t in Equation 2), while the gyroscope-derived angles have been set as external input (see Bu_{t-1} in Equation 1).

B. Autoencoder settings

DANAE can work with any input signal length; here, we present the version working with a length of 20. The encoder part of DANAE is made of four *dilated* 1D convolutions which bring the original 20-length signal to a hidden representation made of 128 features. The decoder part transforms this representation back to a 20-length vector by alternating three transposed-dilated 1D convolutions to four standard ones. While the transposed convolution is exploited to increase the input resolution back to the original size, the i_{th} standard convolution is working on the sum of the i_{th} encoder and the i_{th} decoder outputs. This approach, loosely inspired by the WaveNet architecture, is able to enforce additional constraints to the encoder/decoder pipeline, enabling a more faithful signal reconstruction. All these layers have 128 3×3 kernels with an appropriate dilation value depending on the layer depth, while stride and padding have been fixed to 1. The Adam optimizer chosen for the training has been set with a fixed learning rate of 0.002 with a batch size of 16. The number of epochs has been set to 100 for UCS and to 1 for each set of OxIOD. Additional experiments performed with different hyper-parameters values did not produce any sensible difference in the final accuracy, demonstrating the robustness of our approach.

VI RESULTS

To numerically evaluate the DANAE performances, simple estimator as mean deviation, maximum deviation and root mean square error have been calculated with respect to the ground truth values and compared with those of the KF. As can be seen from Table 1 and Table 2, DANAE is able to considerably improve the performance on all the estimators for both the datasets. Moreover, even though the strong noise affecting the KF predictions, DANAE exhibits a sensible lowering of the elongations for all the angles. In the case of OxIOD, DANAE provides a mean reduction of the RMSE equal to 63%; Figure 1 shows the comparison between the KF and DANAE estimations for the ϕ angle w.r.t. the ground truth.

A similar result is also found in the UCS experiments: DANAE output faithfully resembles the reference signal for the estimated angles, with a final mean reduction of the RMSE equal to 55%. Figure 2 reports the corresponding results for the θ angle. Unfortunately, ψ exhibits a perturbed behaviour both in the estimated and ground truth values, reason why numerical values are here omitted. This

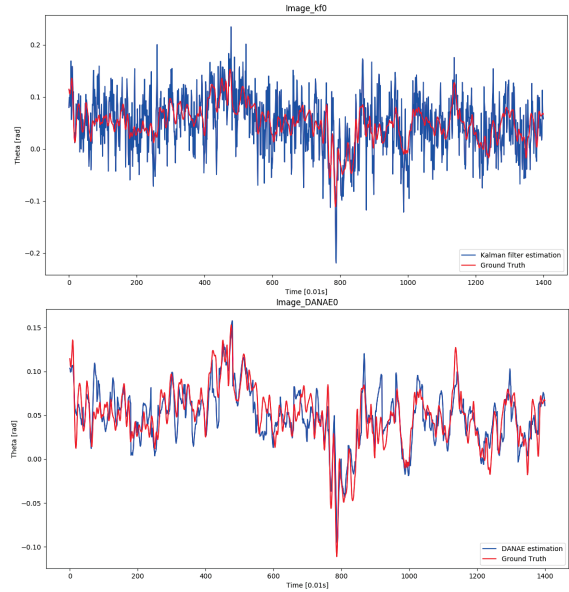


Fig. 2. KF (upside) and DANAE (downside) pitch angle estimation compared to the ground truth for the UCS dataset.

can be probably related to erroneous sensors calibrations or to magnetometer effects, whose non-linearity results in a scale factor error. Moreover electromagnetic-produced deviations can considerably alter the estimations of this angle [29].

VII CONCLUSIONS

This paper proposes DANAE, the first denoising autoencoder model for attitude estimation. Our noise and filter agnostic architecture is able to compensate any kind of error with respect to a reference angle, boosting the KF prediction for both the considered datasets. Further experiments will be performed on non-linear filters (EKF and UKF), and deployments for on-line underwater applications will be investigated.

REFERENCES

- [1] R. E. Kalman, “A new approach to linear filtering and prediction problems,” *Transactions of the ASME—Journal of Basic Engineering*, vol. 82, no. Series D, pp. 35–45, 1960.
- [2] M. St-Pierre and D. Gingras, “Comparison between the unscented kalman filter and the extended kalman filter for the position estimation module of an integrated navigation information system,” in *IEEE Intelligent Vehicles Symposium, 2004*, pp. 831–835, IEEE, 2004.
- [3] V. M. Tereshkov, “An intuitive approach to inertial sensor bias estimation,” *International Journal of Navigation and Observation*, vol. 2013, 2013.
- [4] F. Caron, E. Duflos, D. Pomorski, and P. Vanheeghe, “Gps/imu data fusion using multisensor kalman filtering: introduction of contextual aspects,” *Informa-*

- tion fusion, vol. 7, no. 2, pp. 221–230, 2006.
- [5] W. Li and J. Wang, “Effective adaptive kalman filter for mems-imu/magnetometers integrated attitude and heading reference systems,” *The Journal of Navigation*, vol. 66, no. 1, pp. 99–113, 2013.
 - [6] F. Di Ciaccio, S. Gaglione, and S. Troisi, “A preliminary study on attitude measurement systems based on low cost sensors,” *Proceedings of R3 in Geomatics Workshop (R3GEO), Communications in Computer and Information Science (CCIS), Springer*, in press.
 - [7] B. Allotta, A. Caiti, R. Costanzi, F. Fanelli, D. Fenucci, E. Meli, and A. Ridolfi, “A new auv navigation system exploiting unscented kalman filter,” *Ocean Engineering*, vol. 113, pp. 121–132, 2016.
 - [8] X. Zhang, X. Mu, H. Liu, B. He, and T. Yan, “Application of modified ekf based on intelligent data fusion in auv navigation,” in *2019 IEEE Underwater Technology (UT)*, pp. 1–4, IEEE, 2019.
 - [9] M. B. Alatise and G. P. Hancke, “Pose estimation of a mobile robot based on fusion of imu data and vision data using an extended kalman filter,” *Sensors*, vol. 17, no. 10, p. 2164, 2017.
 - [10] H. G. De Marina, F. J. Pereda, J. M. Giron-Sierra, and F. Espinosa, “Uav attitude estimation using unscented kalman filter and triad,” *IEEE Transactions on Industrial Electronics*, vol. 59, no. 11, pp. 4465–4474, 2011.
 - [11] B. Allotta, L. Chisci, R. Costanzi, F. Fanelli, C. Fantacci, E. Meli, A. Ridolfi, A. Caiti, F. Di Corato, and D. Fenucci, “A comparison between ekf-based and ukf-based navigation algorithms for auvs localization,” in *OCEANS 2015-Genova*, pp. 1–5, IEEE, 2015.
 - [12] E. Taghavi, R. Tharmarasa, T. Kirubarajan, Y. Bar-Shalom, and M. McDonald, “A practical bias estimation algorithm for multisensor-multitarget tracking,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 52, no. 1, pp. 2–19, 2016.
 - [13] K. Nirmal, A. Sreejith, J. Mathew, M. Sarpotdar, A. Suresh, A. Prakash, M. Safonova, and J. Murthy, “Noise modeling and analysis of an imu-based attitude sensor: improvement of performance by filtering and sensor fusion,” in *Advances in Optical and Mechanical Technologies for Telescopes and Instrumentation II*, vol. 9912, p. 99126W, International Society for Optics and Photonics, 2016.
 - [14] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in neural information processing systems*, pp. 1097–1105, 2012.
 - [15] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.
 - [16] K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask r-cnn,” in *Proceedings of the IEEE international conference on computer vision*, pp. 2961–2969, 2017.
 - [17] P. Russo, T. Tommasi, and B. Caputo, “Towards multi-source adaptive semantic segmentation,” in *International Conference on Image Analysis and Processing*, pp. 292–301, Springer, 2019.
 - [18] A. Graves, A.-r. Mohamed, and G. Hinton, “Speech recognition with deep recurrent neural networks,” in *2013 IEEE international conference on acoustics, speech and signal processing*, pp. 6645–6649, IEEE, 2013.
 - [19] A. Graves, “Generating sequences with recurrent neural networks,” *arXiv preprint arXiv:1308.0850*, 2013.
 - [20] R. Baynazarov and I. Piontkovskaya, “Binary autoencoder for text modeling,” in *Conference on Artificial Intelligence and Natural Language*, pp. 139–150, Springer, 2019.
 - [21] K. Sohn, H. Lee, and X. Yan, “Learning structured output representation using deep conditional generative models,” in *Advances in neural information processing systems*, pp. 3483–3491, 2015.
 - [22] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, “Extracting and composing robust features with denoising autoencoders,” in *Proceedings of the 25th international conference on Machine learning*, pp. 1096–1103, 2008.
 - [23] L. Gondara, “Medical image denoising using convolutional denoising autoencoders,” in *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*, pp. 241–246, IEEE, 2016.
 - [24] X. Lu, Y. Tsao, S. Matsuda, and C. Hori, “Speech enhancement based on deep denoising autoencoder,” in *Interspeech*, pp. 436–440, 2013.
 - [25] P. Xiong, H. Wang, M. Liu, S. Zhou, Z. Hou, and X. Liu, “Ecg signal enhancement based on improved denoising auto-encoder,” *Engineering Applications of Artificial Intelligence*, vol. 52, pp. 194–202, 2016.
 - [26] V. M. Tereshkov, “A simple observer for gyro and accelerometer biases in land navigation systems,” *The Journal of Navigation*, vol. 68, no. 4, pp. 635–645, 2015.
 - [27] C. Chen, P. Zhao, C. X. Lu, W. Wang, A. Markham, and N. Trigoni, “Oxiod: The dataset for deep inertial odometry,” *arXiv preprint arXiv:1809.07491*, 2018.
 - [28] A. Mallios, E. Vidal, R. Campos, and M. Carreras, “Underwater caves sonar data set,” *The International Journal of Robotics Research*, vol. 36, no. 12, pp. 1247–1251, 2017.
 - [29] V. Renaudin, M. H. Afzal, and G. Lachapelle, “Complete triaxis magnetometer calibration in the magnetic domain,” *Journal of sensors*, vol. 2010, 2010.