Transformer-based INS/GNSS Fusion Architecture for Multi-sensor Navigation System

Julien Leblanc[1], Ivan Petrunin[2], Zhengjia Xu[3], Patrick Geragersian[4]
Cranfield University, Bedfordshire, MK43 0AL, United Kingdom

Sam Nardoni[5]
Spirent Communications PLC, West Sussex, RH10 1BD, United Kingdom

The increasing dependency of using Global Navigation Satellite System (GNSS) drives higher demands of providing robust, reliable, accurate, and resilient positioning data output in complex operational environments like urban scenarios. Multi-sensor navigation system envisions providing higher performance in terms of integrity, accuracy, reliability, and availability due to its redundancy in sensor number and types for compensating degradation of GNSS positioning quality, and providing reliable positioning output when satellites are unavailable. The fusion of heterogeneous sensor types also allows for leveraging the strengths of each sensor while compensating for their individual weaknesses. For instance, the fusion of Inertial Navigation System (INS) and GNSS facilitates compensating long-term error drifts in INS as well as improving GNSS performance existing multipath and signal blockages by generating high-order linear acceleration and acceleration rate data at a relatively fast rate from Inertial Measurement Unit (IMU) sensors. Nevertheless, one critical challenge in the multi-sensor navigation system relies on designing resilience fusion mechanisms applicable to diverse sensor types.

Regarding fusing distinguished sensor data together, analytics-based models are typically applied with filter designs like Kalman Filter (KF), Particle filter, Unscented or Extended KF, etc. However, it is found that typical filter designs do not show enough positioning accuracy on account of unrealistic estimation of measurement noise model and low adaption against nonlinear noise models. Advanced Kalman filters like Adaptive Kalman Filter (AKF) or Iterative Extended Kalman Filter (IEKF) were proposed to enhance its compatibility in processing more realistic data, but the practical deployment of those fusion algorithms to handle the real-world data still remains challenging.

For alleviation of challenges during modelling in the fusion design, data-driven approaches are widely investigated consisting of recognizing hidden patterns and updating models dynamically. Machine learning and deep learning models approve to be efficient in such purposes, especially in compensating and predicting measurement errors to establish robust fusion schemes by updating noise processes and measurements for positioning accuracy improvement. Due to processing strong capabilities of acquiring non-linearity in the machine learning approaches, the machine learning based fusion designs have shown great advantages against challenging environments. For instance, a Recurrent Neural Network (RNN) based fusion mechanism was studied in the INS/GNSS integration system that facilitates outputting reliable

[1] Graduate student, School of Aerospace Transport and Manufacturing (SATM)
[2] Senior Lecturer, Centre for Autonomous and Cyber-physical Systems
[3] Research Fellow, Centre for Autonomous and Cyber-physical Systems
[4] PhD candidate, School of Aerospace Transport and Manufacturing (SATM)
[5] Staff Systems Engineer, Spirent Communication Plc.

positioning in case of GNSS blockage. Nevertheless, those machine learning-based fusion designs present limitations of slow convergence and poor generalization of embracing distinguished sensor types and operating under new conditions.

Recently, the development of transformer networks originating as Natural Language Processing (NLP) tools has attracted growing attention when designing advanced navigation systems because of the adaptability and generalizability across various domains available within transformer mechanisms. Moreover, the transformer shows significant potential in positioning performance improvement. For instance, a previous study proposed a contextual transformer-based network for inertial navigation and proved that the new transformer design is more efficient compared to Long Short-Term Memory (LSTM) -based approaches for inertial navigation. The dominant reason causing better efficiency in the transformer compared to other models is the capability to capture spatial contextual information before fusing it with temporal knowledge, and also leveraging multi-head attention in the transformer decoder.

Consequently, this study proposes a transformer neural network-based architecture for INS/GNSS/Visual sensor fusion, where the transformer is applied as the key fusion component while the feature extraction from multiple INS outputs is performed using an ensemble of Convolutional Neural Networks (CNN). The encoder mechanism in the transformer enables learning of complex hidden patterns and relationships from GNSS and INS without relying on handcrafted features or heuristics in between. Moreover, the attention mechanism enables attention to relevant parts of the input sequence while performing fusion that makes the transformer suitable for generalization in GNSS applications covering diverse environments, receiver configurations, and data characteristics. Specifically, the attention mechanism allows the model to concentrate on more dominant measurements, identify and mitigate outliers, and effectively combine information from different sensors or time steps. With the utilization of multiple CNN per individual sensor before the transformer, the transformer can simultaneously process multiple elements of the input sequence in parallel, improving thus computational efficiency in the processing of large datasets that enables real-time or near real-time applications potentially.

A high-level architecture of the proposed transformer-based multi-sensor navigation solution is shown in Fig. 1.

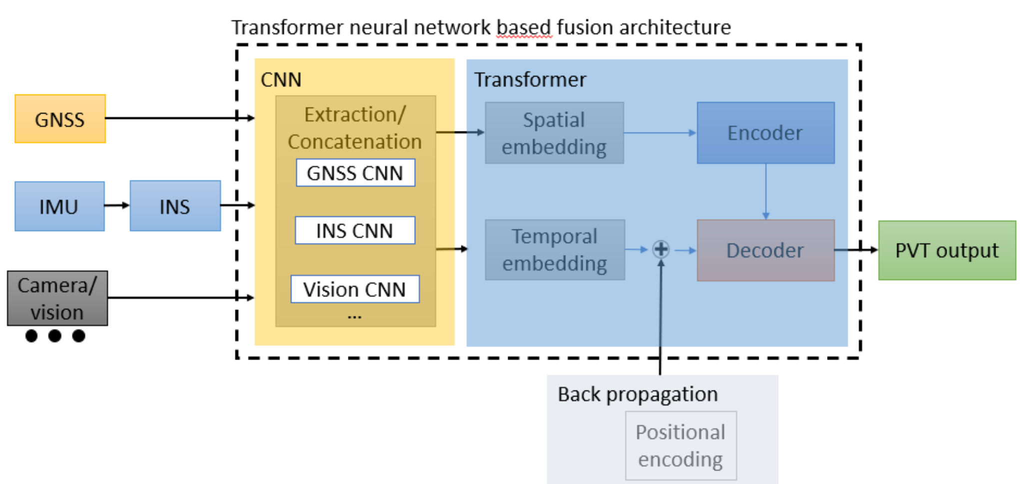

Fig 1: Architecture of the proposed model

In the proposition, CNN is added after each sensor as well as INS and GNSS for extracting representations from each sensor or system by applying the convolution. The transformer architecture is composed by three parts: the encoder, the (self) attention mechanism and the decoder. The encoder is composed of multiple self-attention layers along with feed-forward neural networks. It retrieves the input sequence and captures the contextual information to create representations for the extraction of the input sequence. The decoder composed of multiple layers of self-attention and feed-forward neural networks takes the output of the encoder and generates the fused PVT sequence. The self-attention mechanism applies the dot-product attention which computes a similarity score between a query vector and a set of pair “keys and values” in the input sequence. A demonstration of the transformer design derived from standard transformer architecture is illustrated in Fig. 2.

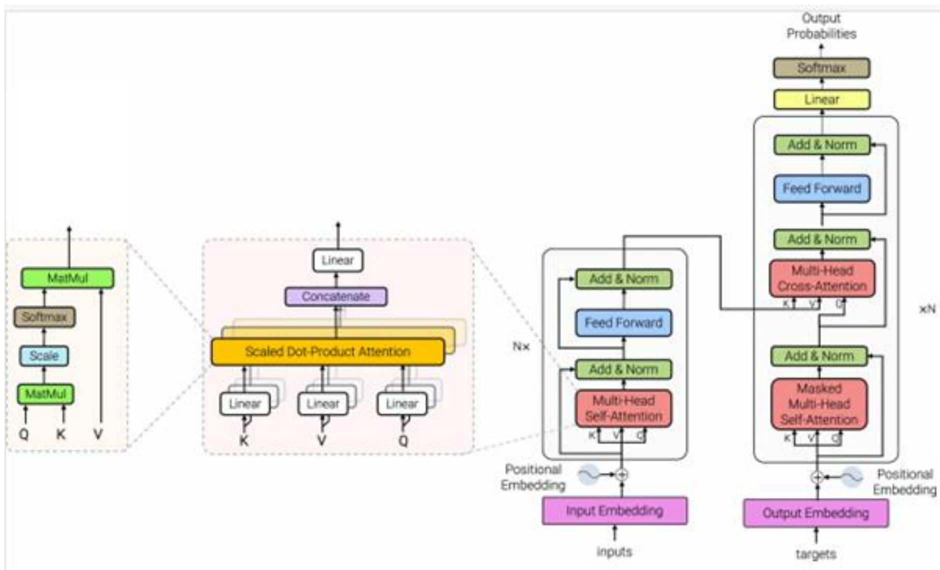

Figure 2: Standard architecture of transformer network

Experimental evaluations are performed with Hardware-in-the-loop (HIL) setup using Spirent GSS 7000 to generate GNSS signals passed then to a mass-market GNSS receiver (U-blox F9) and SimGEN platform to generate realistic IMU data. HIL simulation is commonly applied during the development and testing phases to validate and verify the functionality of a system by integrating real hardware components with a simulated environment.

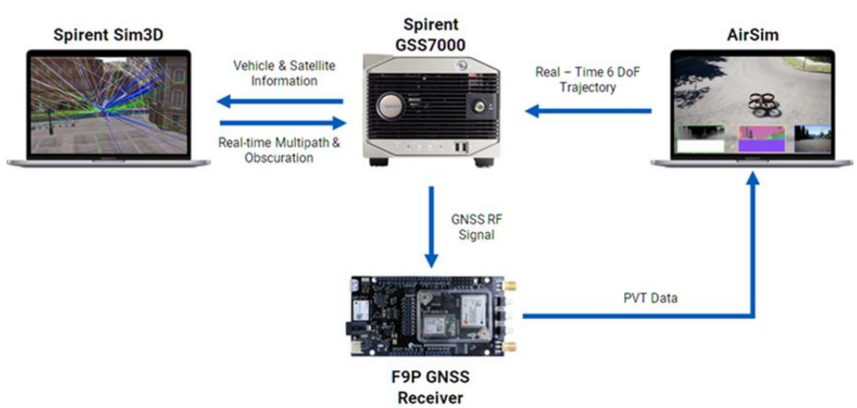

Figure 3: Integration of AirSim with Spirent GSS 7000 simulator/ Simulation setup

The proposed transformer-based fusion mechanism is evaluated by using Spirent’s SimGEN software and GSS 7000 hardware simulator which simulates GNSS signals and enables HIL functionality. Spirent’s

SimSENSOR software is implemented to simulate accelerometer and gyroscope stochastic and deterministic errors aiming to model the behaviour of a consumer-grade IMU. The simulation takes place in a dynamic environment, developed within the SimROUTE software aiming to capture the variability of open-sky and urban environments along with diversity in trajectory. OKTAL-SE Nav Sim3D software is used to realistically simulate the effects of multipath and GNSS blockages in an urban environment, as well as to get the training dataset with the ground truth. In the simulation setup, all the sensor inputs are synchronised. For analysing different shapes of transformer structures on fusion, the number of transformer layers is altered, as well as attention head and hidden dimensions in order to obtain the optimal configuration of the transformer architecture. For evaluating the transformer's performance in terms of generalization a variability in failure modes of sensor signals is explored by adding GNSS outages or blockages, drifts in IMU and weather conditions like fog, rain, or different daylight intensities in vision data, as shown in Fig. 5.

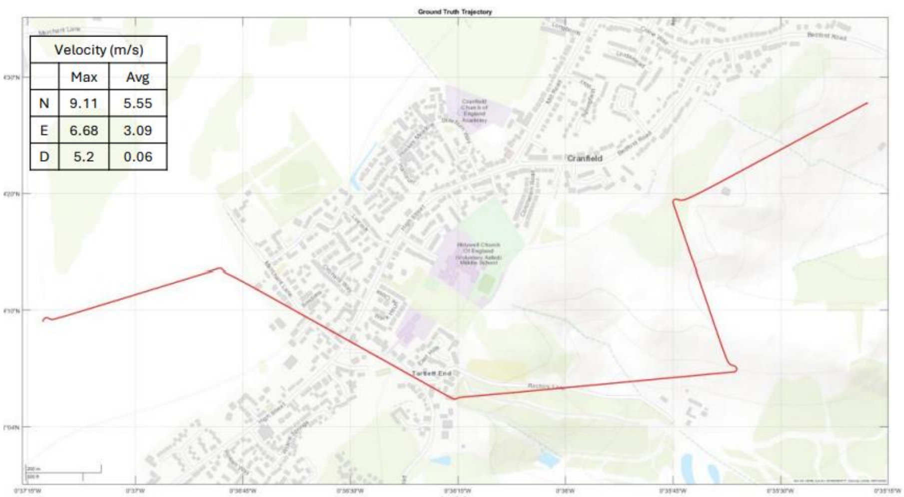

Figure 4: SimROUTE interface

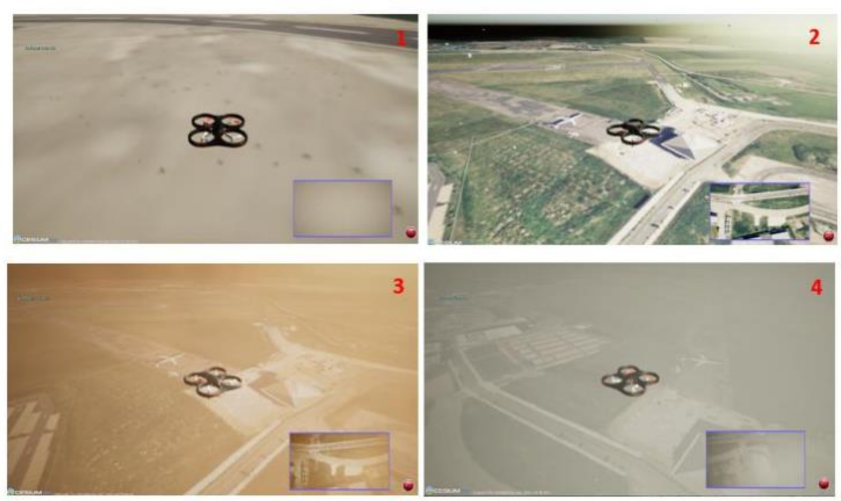

Figure 5: Examples of different weather conditions to test the algorithm from AirSim

The conventional and the state of art techniques such as the Extended Kalman Filter, LSTM-based model, and Gated Recurrent Unit (GRU) network based-model are used as benchmarks to reveal the performance improvement by having transformer in terms of positioning accuracy and navigation solution robustness in similar scenarios. Sensor data collected from both open sky and urban environments are used for validating the positioning accuracy improvement and resilience of the proposed architecture to the failures. Error distribution metrics such as Root Mean Square Error (RMSE) and 95^{th} percentile are used for evaluation and comparison purposes. To evaluate the proposed system's computational efficiency, a comparative study of training and inference times is performed against the previously mentioned similar fusion architectures.

The results demonstrate that the transformer-based fusion architecture consistently outperforms existing methods in terms of accuracy and robustness. Better performance could be expected, as transformer networks already showed more accurate results over already existing convolutional architectures CNNs (such as ResNet) and recurrent networks such as LSTM for inertial navigation only. The computational complexity, however, is found inferior to other approaches, such as GRU, due to the presence of multiple convolutional operations and overall complexity. Overall, our findings highlight the potential of the transformer neural network in fusing GNSS and IMU data, offering a data-driven approach that outperforms traditional fusion techniques and can generalize easier, providing this type of ML-based fusion with a strong practical advantage.