Processes and EngIneering in Mechanics and Materials Laboratory (PIMM)

CNRS (UMR 8006) - Arts et Métiers
151 Boulevard de l'Hôpital, 75013 Paris – France
Tél : +33 144246463 - website : https://pimm.artsetmetiers.fr/en

PHD POSITION OPENNING

Title: *Multi-Scale and Multi-Source Data Fusion: Physics-Guided DL Models for Defect Detection in Railway Infrastructure*

Supervisors and contacts: Fakhreddine Ababsa (fakhreddine.ababsa@ensam.eu), Marc Rebillat (marc.rebillat@ensam.eu), Nicolas Hascoët (nicolas.hascoet@ensam.eu)

Funding : ANR FUSAR Project – Multi-Source Multi-Scale Data Fusion for Linear Transport Infrastructure

Starting date : Autumn 2024

Topic description

Context:

Today's rail infrastructure plays a major role in the mobility of people and goods, and its safety is critical. However, monitoring these huge rail networks poses complex challenges. The ambition of the FUSAR project is to develop an advanced, predictive alert system based on the fusion of multi-scale, multi-source data and potential risks, ensuring more effective management of linear transport infrastructures.

This warning system will make a significant contribution to the sustainability of these infrastructures by enabling proactive management of potential risks and hazards (e.g. the onset of a landslide, subsidence/swelling or sinkhole), thereby reducing maintenance costs, service interruptions and environmental impacts. To achieve this objective, we propose to combine information coming from multiple sources, in particular point data from IoT sensors installed in situ, linear data generated by LiDAR data captured daily by SNCF Réseau's track surveillance vehicles (ESV), surface data obtained by interferometric processing of satellite radar data (InSar) as well as satellite imagery in the visible range (Pleiade - RGB range). Combining and integrating these multimodal data, collected at various spatiotemporal scales, and interpreting them in relation to physical models will provide more robust and reliable information on the state of infrastructures. This data will enable us to carry out global monitoring and, by cross-referencing multisource data (with physical interpretation), to detect faults and warnings, thereby targeting risk areas more accurately and more quickly, particularly on earth structures.

This thesis will address some of the issues raised by the FUSAR project, in particular those relating to the fusion of heterogeneous data and physical models.

Objectives and research work:

During this thesis, the PhD candidate will develop two complementary approaches for multi-source data fusion and defect detection.

- **First Approach: Deep Learning for Data Fusion**
  The candidate will be in charge of developing multi-source data fusion solutions using various Deep Learning (DL) models. The goal is to identify the most effective combination for optimal interpretation of data such as LiDAR and InSAR [1,2]. We plan to leverage existing architectures like CNNs, RNNs, GANs, and Transformers, as well as innovative approaches incorporating attention mechanisms, autoencoders, and reinforcement learning techniques. For example, recurrent neural networks (RNNs) like LSTM or GRU will be used to capture temporal dependencies, which are crucial for evolving data

such as InSAR and IoT measurements [3,4,5]. Additionally, reinforcement learning will be explored to enhance the efficiency and accuracy of the fusion process by dynamically selecting relevant features from each data source and efficiently managing temporal sequences for early detection of sinkholes and landslides.

- **Second Approach: Physics-Guided Hybrid Models**

The second part of the thesis focuses on developing physics-informed neural networks (PINNs) for sinkhole detection [6,7]. This method will integrate geotechnical and monitoring data with fundamental physical principles. Key physical factors such as soil porosity and permeability, soil composition, underground erosion, and mechanical stresses will be identified first. This model will not only interpret existing data but also generate synthetic data to simulate various sinkhole scenarios. This hybrid approach will integrate historical and real-time data (such as satellite measurements, field observations, and geological data) with physical models of soil dynamics and erosion processes [8,9]. This will help predict where and when sinkholes or land subsidence are likely to occur.

In summary, the thesis will focus on developing and evaluating multi-source data fusion solutions based on DL and physics-guided hybrid approaches for the detection and prediction of defects in railway infrastructures.

References

[1] Qin, P., Huang, H., Tang, H., Wang, J., Liu, C. MUSTFN: A spatiotemporal fusion method for multi-scale and multisensory remote sensing images based on a convolutional neural network, Int. J. of Applied Earth Observation and Geoinformation, Vol. 115, 103113, **2022**.

[2] Tang, P., Chanussot, J., Guo, S., Zhang, W., Qie, L., Zhang, P., Fang, H., Du, P. Deep learning with multi-scale temporal hybrid structure for robust crop mapping, ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 209, **2024**.

[3] Toma, T. I., Choi, S. An End-to-End Convolutional Recurrent Neural Network with Multi-Source Data Fusion for Sleep Stage Classification. 2023 Int. Con. on Artificial Intelligence in Information and Communication (ICAIIC), pp. 564-569, **2023**.

[4] Xing, L., Liu, W. A Data Fusion Powered Bi-Directional Long Short Term Memory Model for Predicting Multi-Lane Short Term Traffic Flow. in IEEE Trans. on Intell. Transportation Systems, vol. 23, no. 9, pp. 16810-16819, **2022**.

[5] Hou, S., Li, W., Liu, T., Zhou, S., Guan, J., Qin, R., Wang, Z. MUST: A Multi-source Spatio-Temporal data fusion Model for short-term sea surface temperature prediction, Ocean Engineering, Volume 259, 111932, 2022.

[6] Karpatne, A., Atluri, G., Faghmous, J.H., Steinbach, M., Banerjee, A., Ganguly, A., Shekhar, S., Samatova, N.,Kumar, V. : Theory-guided data science : A new paradigm for scientific discovery from data. IEEE Transactions on knowledge and data engineering, Vol. 29, n° 10, p. 2318-2331, **2017**.

[7] Willard, J., Jia, X., Xu, S., Steinbach, M., Kumar, V. : Integrating scientific knowledge with machine learning for engineering and environmental systems. ACM Computing Surveys, Vol., 55, n°4, p. 1-37, **2022**.

[8] Husain, S., Rébillat, M., Ababsa, F. Image processing through deep learning after DI extraction for the SHM of aeronautic composite structures using Lamb waves. Proc. SPIE 12749, 16th Int. Conf. on Quality Control by Artificial Vision, 1274910, **2023**.

[9] Stanley, T.A., Kirschbaum, D.B., Sobieszczyk, S., Jasinski, M., Borak, J.,Slaughter, S., Building a landslide hazard indicator with machine learning and land surface models, Environmental Modelling and Software, **2020**.

Candidate Profile

The candidate must:

1. Have a Master 2 or equivalent in Machine Learning, AI, Applied mathematics, mathematics, computer science, or other field strongly related to Applied mathematics.
2. Have excellent analytical and communication skills in written and spoken English.
3. Be able to work independently and take responsibility for the progress and quality of the project.
4. Have experience in data collection, statistical data analysis and exploration, and geospatial data analysis.
5. Have excellent programming skills.

Interested candidates should send to **Mr. ABABSA** (fakhreddine.ababsa@ensam.eu) and **Mr. REBILLAT** (marc.rebillat@ensam.eu) an application containing:

1. a personal motivation letter (max. 1 A4 page) describing why you apply and how the position fits into your career plans,
2. a full CV showing how your profile fits the requirements (max 4 pages),
3. an electronic copy of your Master's thesis
4. recommendation letters
5. a list of referees we can contact.