Outlier and novelty detection in classification or segmentation models

Supervisor's names:

Alexandre Baussard (Full professor - alexandre.baussard@utt.fr)

1/ short description

Recent models for image classification and segmentation have achieved excellent performance. However, to ensure the reliability of the results produced when the data comes from a real uncontrolled environment, it is necessary to be able to decide whether a new observation belongs to the distribution of observations used for training, or not. It is also useful to be able to determine whether these outliers do not come from a new class of data.

The project proposes to tackle these problems using different approaches, with the aim of keeping these models unchanged to guarantee their initial performance. The considered methods could be based on statistical methods or machine learning approaches.

2/ Laboratory

This thesis project will be carried out within the Laboratoire Informatique et Société Numérique (LIST3N), a Research Unit of the Université de Technologie de Troyes (UTT). Research at LIST3N is structured around four axes: data processing, optimization, networks, technologies and practices. LIST3N also relies on several UTT institutes, industrial chairs and technology platforms.

This project will be developed in the data processing team, which research topics are decision and estimation, process modeling and diagnostics, based on shared expertise in statistics. In recent years, the members of this team have also developed expertise in the fields of machine learning and artificial intelligence for processing data from different observation systems.

3/ Project

Introduction / context:

Anomaly detection is an important topic in data analysis. The definition of an anomaly may vary according to the field of application or the objective pursued, but it is generally considered to be an observation that deviates considerably from the rest of the observations, as if generated by a different process. So, in some cases, anomaly detection can help improve data quality by removing or replacing abnormal data. In other cases, the anomalies reflect an event and provide useful new knowledge. For example, anomaly detection can prevent material damage and thus encourage preventive maintenance in the industrial sector. They can also be a sign of the presence of a new phenomenon that we need to try and take into account, such as the detection of a new class linked, for example, to the appearance of a new model of vehicle, aircraft or other in surveillance applications.

In recent years, models for image classification and segmentation have achieved excellent performance. However, the deployment of these models in operational contexts requires that they are not only efficient, but also robust in the face of any new situation varying more or less strongly from those encountered during the learning process. It has been shown that, unfortunately, these models can provide a very high level of confidence in a result when it is not correct. It is therefore necessary to develop methods to reduce the risk of error. To do this, it is necessary to better understand and explain the decisions made, to be able to detect any decision errors potentially linked to the presence of an anomaly. The development of these methods, which will be the subject of this project, will be of interest in a several ways: detecting and processing anomalies, gaining a better understanding of decision-making, providing an operator with a level of confidence in the decision, etc.

Work plan:

Based on the results of previous work such as [ACRE_20] and the results of master's degree internships, we will be looking at anomaly detection methods that are independent of the classifier/segmentation models implemented to avoid modifying their intrinsic performance. In fact, some solutions involve modifying the model at the risk of altering its performance. Moreover, these solutions inevitably require new learning stages. Our preferred approach, which consists in adding detection to the model, seems to us to be much better and less impactful for the models. Among the methods, the generalized ODIN approach [HSU_20] to lead to accurate results. In brief, it consists of using a trained classifier model, a statistical criterion and input pre-processing to construct a statistic for each input, enabling anomalies to be detected. The generalized version also makes it possible to construct a criterion without the explicit use of outliers, which also seems to us to be a good way of proceeding given the lack of available information on the distribution of data to be excluded. There exist approaches based on statistical tests such as [HAR_22]. It's this type of method that we want to put in place as a priority.

Another area of interest is the use of 'prototypes' combined with other elements such as auto-encoders [LAI_21]. Data prototypes are elements that reflect the characteristics of class data or groups of class data. They are used in certain approaches, notably for learning with little data. The idea here would be to exploit a more generic representation of the spaces linked to known classes defined using these prototypes and to anomalies that deviate from them.

In all cases, our aim will be to propose one or more anomaly detection methods that can be deployed on existing models without modifying them, in order to avoid any loss of performance and any need to re-learn the models; an operation that can be very costly in terms of computing power in certain cases (often several million parameters to optimize). From this point on, another objective of this work will be to use this detector to highlight the possible existence of a hierarchy of relevant information characteristic of the classes to be identified in the data (not used during the learning phase). This novelty detection is also a problem that remains open, especially when we're trying to achieve a certain genericity in the method.

So, the aim of the project is to exploit these different elements to achieve a better understanding of decision-making in a model.

The project can be developed in three phases:

- Phase one: bibliographical study and implementation of existing solutions to assess their performance, limitations... Based on this initial work, proposal and evaluation of variants to perfect understanding and mastery of these tools and methods.

- Phase two: proposal of new approaches with the aim of remaining as generic as possible (i.e. application-independent). We can consider different approaches: statistical, learning, ...

- Phase three: evaluation for different applications (anomaly detection, novelty detection) and databases.

References:

[ACRE_20] A. d'Acremont, G. Quin, A. Baussard, R. Fablet. Detection of outliers for deep neural networks trained from synthetic data. Conference on Articial Intelligence for Defense (CAID), November 2020.

[BALA_21] L. Balasubramanian, F. Kruber, M. Botsch and K. Deng. Open-set Recognition based on the Combination of Deep Learning and Ensemble Method for Detecting Unknown Traffic Scenarios. arXiv:2105.07635v1, May 2021.

[CHEN_21] J. Chen, Y. Li, X. Wu, Y. Liang and S. Jha. Robust. Out-of-distribution Detection for Neural Networks. arXiv:2003.09711v6, December 2021.

[HAR_22] M. Haroush, T. Frostig, R. Heller, D. Soudry. A Statistical Framework for Efficient Out of Distribution Detection in Deep Neural Networks. International Conference on Learning Representations (ICLR), 2022.

[HSU_20] Y.-C. Hsu, Y. Shen, H. Jin and Z. Kira, Generalized ODIN: Detecting Out-of-distribution Image without Learning from Out-of-distribution Data. CVPR 2020.

[LAI_21] Y. Lai, Y. Han, Y. Wang. Anomaly Detection with Prototype-Guided Discriminative Latent Embeddings. arXiv:2104.14945v3, September 2021.

[PANG_20] G. Pang, C. Shen, L. Cao, A. Van Den Hengel. Deep Learning for Anomaly Detection: A Review. arXiv:2007.02500v3, December 2020.

[RAMI_22] A. Ramírez Rivera, A. Khan, I.E.I. Bekkouch, and T. Shakeel Sheikh. Anomaly Detection Based on Zero-Shot Outlier Synthesis and Hierarchical Feature Distillation. IEEE Trans Neural Networks and Learning Systems, vol. 33, no. 1, 2022.

[YANG_21] J. Yang, K. Zhou, Y. Li, and Z. Liu. Generalized Out-of-Distribution Detection: A Survey. arXiv:2110.11334v1, October 2021.