Project Title: Face Deepfake Detection (FDD)

Host Laboratory: *Laboratoire d'Informatique Gaspard-Monge (LIGM)*

International Host Laboratory: Melbourne Centre for Data Science, The University of Melbourne**,** Australia

Principal Supervisors:

- Dr. Chaohui WANG (Associate Professor with *HDR*, *Université Gustave Eiffel,* France)

- Dr. Mingming Gong (Assistant Professor, The University of Melbourne, Australia)

Specialties: *Informatique*; *Data Science*; *Intelligence Artificielle*

Subject:

- **Research Background:**

Due to the rapid development of the generative model of deep neural networks, it has been able to perform general image re-rendering with a high quality (e.g., [1]). Accordingly, very real fake facial images and videos (Deepfake) have been produced [2,3]. These generated face images are widely used in various fields, but some criminals release false information by modifying images or videos, disrupting social order and also posing challenges to the face recognition system. At the end of 2017, a Reddit user by the name of 'deepfakes' was using deep learning to swap faces of celebrities into pornographic videos, and was posting them online. The discovery caused a media frenzy and a large number of new deepfake videos began to emerge thereafter. Therefore, it is imperative to add Deepfake detection technology to the future face recognition systems.

Since the current deepfake is mainly generated by GAN [4,5], the technical maturity of GAN determines the difficulty of Deepfake detection. If the GAN model is close to perfect, it is impossible to detect Deepfake. Fortunately, the current high-quality GAN model cannot make the generated image consistent with the actual image distribution. The main reason is that strong constraints are usually added to the discriminator, such as spectral normalization and additional supervision loss, etc..

- **Main Goals:**

In the aforementioned context, this project aims to develop an efficient Deepfake detection algorithm using time frequency analysis and deep learning techniques. Two particular directions to study are: (1) how to use normal neural networks without special constraints together with some of the basic features of GAN to design efficient algorithms; (2) explore differences in statistical characteristics between real images and synthetic images, and incorporate them into the whole detection framework to boost the performance.

- For more details, please send emails to: chaohui.wang@univ-eiffel.fr

References:

[1] Wang, Chaoyue, Chaohui Wang, Chang Xu, and Dacheng Tao. "Tag disentangled generative adversarial networks for object image re-rendering." In International Joint Conference on Artificial Intelligence (IJCAI). 2017.

[2] Tolosana, Ruben, et al. "Deepfakes and beyond: A survey of face manipulation and fake detection." Information Fusion 64 (2020): 131-148.

[3] https://www.vice.com/en/article/gydydm/gal-gadot-fake-ai-porn

[4] Goodfellow, Ian, et al. "Generative adversarial nets." Advances in neural information processing systems 27 (2014).

[5] Gong, Mingming, et al. "Twin auxiliary classifiers gan." Advances in neural information processing systems 32 (2019): 1328.