IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom
Encadrants :

Projet PRAMeL

Sujet de Projet – TAF MCE

Projects on Recent Advances in Machine Learning

Protection de la propriété intellectuelle d'un modèle d'apprentissage profond par tatouage

Reda BELLAFQIRA
reda.bellafqira@imt-atlantique.fr
Gouenou COATRIUEX
gouenou.coatrieux@ imt-atlantique.fr
Département ITI

30 janvier 2023

Description du sujet

La technologie d'apprentissage profond ("Deep Learning") a permis des progrès importants et rapides dans les domaines de l'analyse du signal et d'application comme la reconnaissance faciale, vocale, du traitement automatisé du langage, de l'aide au diagnostic, etc. Cependant, établir un modèle d'apprentissage profond est une tâche non triviale qui nécessite : une grande quantité de données d'apprentissage, des ressources informatiques puissantes et des compétences humaines. De ce fait, la reproduction et la redistribution illégale d'un modèle constitue pour son auteur un préjudice économique non négligeable.

Plusieurs solutions originales ont été proposées afin de protéger la propriété intellectuelle des modèles. Elles s'appuient sur le tatouage numérique. Le principe du tatouage consiste à insérer un message (une marque) dans un document hôte en modifiant de manière imperceptible certaines de ses caractéristiques (pour une image les valeurs des pixels seront modifiées pour coder le message). Ce message peut ensuite aider à vérifier l'origine/la destination du document qu'il protège.

Les méthodes de tatouage pour les modèles d'apprentissage profond s'appuient sur les mêmes principes que celles pour les images ou les vidéos avec quelques contraintes spécifiques, cependant. Les attaques sur la marque sont plus sévères et des solutions sont à trouver pour y résister.

Mots-clés :
APPRENTISSAGE PROFOND, TATOUAGE

Travail à réaliser

Le travail consiste dans un premier temps à comprendre et implémenter un à plusieurs algorithmes d'apprentissage profond ("Deep learning") sur une base de données d'images et, dans un second temps, à appliquer le tatouage sur le modèle appris puis analyser l'influence de différentes attaques sur la détection de la marque et proposer quelques améliorations

Environnement de travail

Le travail à réaliser comporte une activité de recherche et nécessite des compétences en matière de traitement de l'information (machine learning), de sécurité (tatouage), comme aussi des connaissances d'outils de développements (Python, PyTorch).

Quelques références bibliographiques

[1] Uchida, Yusuke, et al. "Embedding watermarks into deep neural networks." *Proceedings of the 2017 ACM on international conference on multimedia retrieval*. 2017.

[2] Wang, Tianhao, and Florian Kerschbaum. "Riga: Covert and robust white-box watermarking of deep neural networks." *Proceedings of the Web Conference 2021*. 2021. --- https://www.youtube.com/watch?v=fe42B1MISMM ---