

Sujet de stage Master2/Ingénieur

Face swapping et deepfake vidéo

Environnement

Le laboratoire GREYC (UMR 6072) est un laboratoire en sciences du numérique basé à Caen en Normandie, composé de plus de 230 membres dont une centaine de chercheurs ou enseignants-chercheurs permanents, répartis dans 6 équipes de recherche couvrant de nombreux domaines de l'informatique et l'électronique.

L'équipe SAFE (Sécurité, Architecture, Forensique et Biométrie) du GREYC mène des activités de recherche dans le domaine de la sécurité informatique en Biométrie, en architecture et modèles de sécurité et en Science de l'investigation (Forensique) avec une continuité des aspects théoriques vers les applications, les trois s'enrichissant mutuellement.

Encadrement du stage : Christophe Charrier, Maxime Bérubé et Emmanuel Giguet

Mots clés : Deepfake, face swapping, apprentissage, temps réel

Introduction

Le face swapping est une technique de vision par ordinateur qui consiste à échanger le visage d'une personne avec celui d'une autre dans une image ou une vidéo. Cette technologie a gagné en popularité avec des applications comme les filtres de réalité augmentée (RA) dans les réseaux sociaux, mais elle a aussi des implications importantes dans le cinéma, les jeux vidéo et la sécurité numérique. Les progrès réalisés avec les réseaux génératifs adverses (GANs) ont permis de générer des visages échangés de manière de plus en plus réaliste, avec des détails précis et des ajustements d'expressions faciales.

Cependant, le défi actuel reste de réaliser ce face swapping en temps réel, avec des performances suffisantes pour une application en direct et une fidélité visuelle élevée, en prenant en compte des éléments tels que les expressions faciales, l'éclairage et la cohérence des mouvements.

Ce stage propose de développer un système de face swapping en temps réel, en se basant sur les techniques de GANs et en tenant compte des contraintes de performance requises pour un traitement en temps réel sur des dispositifs limités comme les smartphones ou les plateformes en ligne.

FaceShifter est l'une des méthodes les plus avancées en termes de qualité de génération de visages échangés. Il utilise un encodeur-décodeur couplé à un modèle discriminant pour préserver les détails faciaux tout en adaptant les expressions et l'éclairage du visage source. Li et al. [1] ont présenté ce modèle comme étant capable de générer des visages extrêmement réalistes même dans des scénarios complexes.

Le modèle FSGAN permet un face swapping en temps réel sans nécessiter de réentraînement. Il est capable de gérer des expressions faciales variées ainsi que des angles différents grâce à des sous-réseaux spécifiques qui traitent à la fois la reconstruction du visage et l'ajustement des expressions. Cette technique est très efficace dans les contextes où peu d'images de la personne cible sont disponibles. Nirkin et al. [2] ont introduit ce modèle comme une approche polyvalente capable de s'adapter à divers contextes avec peu de données.

Les techniques de few-shot face swapping permettent de réaliser des échanges de visages réalistes avec un nombre limité d'images de la personne cible. Ces modèles, tels que ceux proposés par Zhang et al. [3], reposent sur la création d'une représentation latente commune du visage source et cible, facilitant ainsi l'échange des visages tout en conservant des expressions et des textures fidèles avec un minimum d'exemples.

StarGAN v2 est un modèle basé sur le concept de style transfer entre différents domaines. Il permet de changer le visage tout en maintenant une haute qualité d'image. Le modèle est capable de manipuler des attributs faciaux comme l'âge, le genre ou l'expression tout en effectuant le swapping. Cette approche, proposée par Choi et al. [4], est particulièrement robuste pour les transformations multi-domaines, ce qui la rend idéale pour le face swapping dans des environnements variés.

Les techniques de reconstruction 3D jouent également un rôle important dans la génération de visages réalistes dans le face swapping. Elles permettent en effet de capturer la structure du visage et de la réappliquer à un autre visage, en tenant compte de l'éclairage et des angles de vue. Par exemple, SPARK [5] propose un modèle auto-supervisé capable de capturer des visages en temps réel avec une grande précision, en s'appuyant sur une seule caméra pour la reconstruction 3D.

La majorité des techniques précédemment citées repose sur l'utilisation des jeux de données massifs pour entraîner ces modèles. Pour répondre à cette limitation, le concept de few-shot learning a émergé. Cette approche permet aux modèles d'apprendre et de généraliser à partir d'un petit ensemble d'exemples. En combinant les techniques de GANs avec le few-shot learning, il devient possible de générer des visages réalistes avec peu d'images de la personne cible. Cette thèse propose d'explorer les dernières avancées en Few-shot Face Swapping avec GANs, avec l'objectif d'améliorer la qualité du swapping tout en minimisant les données d'entraînement nécessaires.

Objectifs du stage

L'objectif principal de ce stage est de concevoir et implémenter un modèle capable de réaliser un face swapping en temps réel avec une haute qualité d'image et un faible temps de latence. Le projet couvrira plusieurs aspects du pipeline de face swapping, depuis la capture des visages jusqu'à la génération et le rendu en temps réel.

Les objectifs spécifiques sont :

1. Étude des GANs pour le face swapping : Comprendre les architectures existantes telles que FaceShifter ou FSGAN, et identifier les optimisations nécessaires pour le temps réel.
2. Conception d'un pipeline de traitement en temps réel : Développer un pipeline permettant de capturer des visages, d'effectuer le swapping et de générer les résultats sans latence perceptible, tout en tenant compte des contraintes matérielles.
3. Optimisation du modèle GAN : Adapter et optimiser un modèle GAN pour qu'il fonctionne efficacement en temps réel, notamment en réduisant les temps de calcul et en minimisant la consommation de ressources.
4. Évaluation des performances : Mesurer la qualité visuelle (réalisme, fluidité des transitions) et les performances en temps réel (temps de réponse, consommation de mémoire et processeur) du système, et identifier les points d'amélioration.

Les tâches à réaliser sont les suivantes :

- **Recherche bibliographique** : Revue des articles récents sur les GANs appliqués au face swapping, et étude des méthodes d'optimisation des réseaux neuronaux pour le temps réel.
- **Développement du modèle** : Implémentation d'un modèle de face swapping à partir des architectures existantes (par exemple FaceShifter, StarGAN v2, FSGAN). Optimisation des temps d'exécution à l'aide de techniques de compression de modèles (quantification, pruning) et d'accélération matérielle (GPU, TPU).
- **Conception d'un prototype d'application** : Intégration d'une interface utilisateur capable de capturer des visages en temps réel (via webcam ou caméra de smartphone). Développement du pipeline de face swapping avec traitement en temps réel et affichage des résultats instantanés.

- **Tests et validation** : Tester la robustesse du modèle dans différents scénarios (variations d'éclairage, angles de vue, expressions faciales). Évaluer les performances du modèle sur différentes plateformes matérielles (PC, smartphone) et optimiser l'efficacité du système.

Compétences requises

Le candidat devra avoir :

- de solides connaissances en apprentissage profond et réseaux neuronaux, notamment les GANs.
- une expérience en vision par ordinateur et traitement d'images.
- une appétence en optimisation de modèles et implémentation de systèmes en temps réel.
- des compétences en programmation (Python, TensorFlow ou PyTorch).
- un capacité à travailler sur des systèmes embarqués ou à intégrer des modèles dans des applications mobiles serait un plus.

Candidature

Pour postuler, envoyer par email aux encadrants un dossier avec CV, lettre de motivation, relevés de notes des deux dernières années de formation, ainsi que toute pièce susceptible de renforcer la candidature (lettre de recommandation, etc.).

Contact

Christophe Charrier, Université de Caen Normandie (christophe.charrier@unicaen.fr)

Maxime Bérubé, Université du Québec à Trois-Rivières (Maxime.Berube2@uqtr.ca)

Emmanuel Giguet, CNRS (emmanuel.giguet@unicaen.fr)

References

- [1] L. Li, W. Zhang, Y. Xu, and L. Zhang, "Faceshifter: Towards high fidelity and occlusion aware face swapping," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4851–4860, 2020.
- [2] Y. Nirkin, Y. Keller, and T. Hassner, "Fsgan: Subject agnostic face swapping and reenactment," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 7184–7193, 2019.
- [3] S. Zhang, X. Wang, S. Liu, J. Lin, and C. Wen, "Few-shot face swapping using gans," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, pp. 2455–2463, 2021.
- [4] Y. Choi, Y. Uh, J. Yoo, J.-W. Ha, and H. Choi, "Stargan v2: Diverse image synthesis for multiple domains," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8188–8197, 2020.
- [5] K. Baert, S. Bharadwaj, F. Castan, and B. Maujean, "Spark: Self-supervised personalized real-time monocular face capture," *arXiv preprint arXiv:2409.07984*, 2024.