

Proposition de thèse

Contraintes structurelles pour les réseaux antagonistes génératifs (GANs - Generative Adversarial Networks)

Contexte

Les réseaux antagonistes génératifs (Generative Adversarial Networks, GANs [1]) sont des modèles non-supervisés utilisés pour générer des objets qui respectent les caractéristiques d'un ensemble d'apprentissage. Dans une première approximation les GANs cherchent à modéliser une distribution de probabilités à partir de données, sur un ensemble de caractéristiques / attributs détectés automatiquement. L'approche est celui d'un jeu à somme nulle implémenté comme une compétition entre deux composantes (le plus souvent des réseaux de neurones profonds) : un *générateur* et un *discriminateur*. Le générateur produit des données candidates (par ex. des images) et le discriminateur cherche à déterminer si l'échantillon proposé est réel ou bien s'il est le résultat du générateur. Les données *authentiques* de départ servent à entraîner le discriminateur avant le début de la compétition. Suit une étape d'apprentissage commune (à la fois pour le générateur et pour le discriminateur) durant laquelle le générateur cherche à tromper le discriminateur en produisant des candidats qui ressemblent de plus en plus aux données authentiques, alors que le discriminateur s'améliore aussi en essayant de détecter comme non authentiques tous les candidats produits par le générateur. Cette étape d'apprentissage étant connue pour être difficile et souvent instable, de nombreux travaux ont visé à améliorer la qualité de cet apprentissage, voir par exemple l'introduction des Wasserstein GAN [2].

Enjeux et objectifs

Depuis leur introduction en 2014, les GANs ont été utilisés avec beaucoup de succès dans plusieurs domaines applicatifs, par exemple la génération automatique des images photoréalistes (visualisation, design, mode, style artistique, scènes de jeux vidéo, imagerie astronomique), des patterns de mouvement dans les vidéos, et même la suggestion de nouvelles structures moléculaires pour la recherche pharmaceutique. Toutefois, malgré l'intérêt évident, il y a pour l'instant peu de travaux concernant la génération des objets/scènes 3D. Utilisés dans [3] pour la génération d'objets 3D, les GAN produisent des objets qui sont des volumes composés de voxels (éléments de volume) en utilisant des réseaux convolutionnels volumiques (le générateur établit une correspondance entre l'espace latent d'entrée et l'espace des représentations 3D). Ces travaux sont étendus dans [4] pour la reconstruction 3D à partir d'une seule projection 2D. Le problème principal des GANs dans ce contexte est qu'ils génèrent des candidats sans tenir compte de la structure interne, c'est à dire des objets présents et des relations entre ces objets, ce qui conduit à des scènes/objets souvent très peu réalistes, surtout pour les objets 3D pour qui l'aspect structurel est souvent essentiel.

Dans cette thèse on se propose d'investiguer ces limitations avec une vue applicative vers la génération des objets 3D, et l'accent mis sur la cohérence interne des objets générés par rapport aux composants. Un nombre de travaux existent pour des images 2D, mais à l'heure actuelle peu de travaux s'intéressent à la génération d'objets 3D composés. Pour aborder ce problème on envisage plusieurs pistes. D'abord, pour un approche multi-échelle, on utilise souvent des architectures hiérarchiques : cascades/pyramides de réseaux pour une génération de type *coarse-to-fine*, avec de niveaux de détails de plus en plus fins [5]. Une autre possibilité, inspirée par les GANs structurés [6], est l'utilisation de plusieurs générateurs conditionnés par la présence des sous-objets dans un jeu à n -participants dont le but est de collaborer pour battre le discriminateur. Dans [7], les auteurs utilisent des séquences de GANs (appelées Stacked GANs) pour générer des représentations bas niveau conditionnés par les représentations haut niveau fournies par l'inversion de la hiérarchie de couches du réseau convolutionnel discriminant. Ce type d'approche permet de conditionner la sortie à plusieurs niveaux de granularité. D'autres pistes prometteuses sont l'intégration des contraintes non-locales dans la fonction de perte du discriminateur

ou l'inclusion d'un module d'attention permettant de modéliser les relations entre différentes régions de l'image (comme dans les *Self-attention GAN* [8]). Suivant le temps disponible, il est aussi envisageable de s'intéresser aussi à la séparation entre *style* et *structure*, suivant les idées de [10].

Le spectre des applications ou ces travaux peuvent être appliquées est très large : design (exploration des formes alternatives), mode, jeux vidéo (génération de décors, personnages, modèles de visage, générateurs de terrain et paysages), design industriel (structures, meubles, design automobile), architecture (modèles d'intérieur, bâtiments, distributions de bâtiments, planification urbaine), reconstruction 3D. Pour chacun de ces domaines il existe déjà de nombreux outils pour l'automatisation des tâches, outils qui peuvent bénéficier grandement par l'arrivée des générateurs 3D avec des contraintes structurelles.

Profil recherché

Le (la) candidat(e) doit avoir une bonne maîtrise du domaine de l'apprentissage statistique et de l'apprentissage profond, avec à la base de bonnes connaissances mathématiques. De bonnes capacités à programmer en Python sont également nécessaires, l'emploi de modèles d'apprentissage profond nécessitant l'utilisation de bibliothèques comme PyTorch, TensorFlow etc. avec portage du code sur GPU. Envoyez vos candidatures (avec CV, lettre de motivation, notes obtenues) à Michel Crucianu et Marin Ferecatu (courriels : prénom.nom@cnam.fr).

Organisation

La thèse débutera en septembre 2019 et durera 3 ans, sous la direction de Michel Crucianu, co-encadrement Marin Ferecatu (équipe Vertigo). La thèse se déroulera au CNAM Paris (<http://www.cnam.fr/>) dans l'équipe de recherche Vertigo du laboratoire CEDRIC (<http://cedric.cnam.fr/>). Plusieurs enseignants-chercheurs, doctorants et post-doctorants du CEDRIC travaillent sur l'apprentissage profond et certains sur les GANs.

Références

- [1] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In Proc. of International Conference on Neural Information Processing Systems NIPS 2014.
- [2] M. Arjovsky, S. Chintala, L. Bottou. Wasserstein generative adversarial networks. In Proc. of International Conference on Machine Learning ICML 2017
- [3] J. Wu, C. Zhang, T. Xue, W. T. Freeman, J. B. Tenenbaum. Learning a probabilistic latent space of object shapes via 3D generative-adversarial modeling. In Proc. of International Conference on Neural Information Processing Systems NIPS 2016
- [4] J. Wu, C. Zhang, X. Zhang, Z. Zhang, W. T. Freeman, J. B. Tenenbaum, Learning Shape Priors for Single-View 3D Completion and Reconstruction, Proc. Of European Conference on Computer Vision ECCV 2018
- [5] E. L. Denton, S. Chintala, A. Szlam, Rob Fergus, Deep Generative Image Models using a Laplacian Pyramid of Adversarial Networks, Proc. of International Conference on Neural Information Processing Systems NIPS 2015
- [6] Z. Deng, H. Zhang, X. Liang, L. Yang, S. Xu, J. Zhu, E.P. Xing, Structured Generative Adversarial Networks, Proc. of International Conference on Neural Information Processing Systems NIPS 2017
- [7] X. Huang, Y. Li, O. Poursaeed, J. Hopcroft, S. Belongie, Stacked Generative Adversarial Networks, Proc. of International Conference on Computer Vision and Pattern Recognition CVPR 2017

- [8] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena. Self-attention generative adversarial networks. arXiv preprint arXiv:1805.08318, 2018.
- [9] R. Zhang, S. Tang, Y. Li, J. Guo, Y. Zhang, J. Li, and S. Yan. Style separation and synthesis via generative adversarial networks. In Proceedings of the 26th ACM Multimedia, 2018