

Recherche d'architecture neuronale efficiente

Mots-clés : Deep learning, Neural Architecture Search, NAS benchmarks, CNN

Contexte

En raison du nombre important d'hyperparamètres utilisés pour définir les architectures d'apprentissage profond et, par conséquent, le nombre d'organisations possibles pour un réseau de neurones, la communauté de l'apprentissage profond a introduit la recherche d'architecture neuronale (*Neural Architecture Search - NAS*), des algorithmes capables d'automatiser la découverte d'architectures efficaces [2, 9, 10, 11].

Afin de comparer l'efficacité des NAS, la communauté a conçu des *benchmarks* spécifiques destinés à cet effet, à l'exemple du NAS-Bench-101 [12] et du NAS-Bench-201 [1]. Ces *benchmarks* fournissent un espace de recherche fini d'architectures, qui peut être utilisé pour comparer les algorithmes de NAS.

Les algorithmes NAS reposent généralement sur des méta-heuristiques qui utilisent la qualité des solutions (architectures de réseaux de neurones) comme critère de sélection (fonction *fitness*). Cette évaluation nécessite d'abord d'effectuer l'apprentissage du réseau, cependant, des travaux récents proposent d'approcher la qualité d'un modèle sans avoir à l'entraîner [4, 3]. Une approche NAS sans entraînement (IFA) a été mise en place au CERV [5], opérant sur les NAS-Bench-101 [12] et NAS-Bench-201 [1].

Les travaux de Strubell et al. [8] ont montré que l'IA (apprentissage et utilisation de réseaux neuronaux profonds) a un impact significatif sur l'environnement en raison de sa consommation d'énergie et de ses émissions de carbone. Selon Schwartz et al. [7], nous devrions évaluer l'IA sur la base de son efficacité en plus de sa performance. Dans cette optique, une métrique permettant d'inclure l'efficacité d'un modèle lors de son évaluation a été mise en place au CERV [6].

Stage

Dans le cadre de ce stage, nous nous intéresserons à la mise en place d'une méthode de recherche d'architecture, utilisant l'efficacité du modèle comme critère de sélection. Le stagiaire pourra se reposer sur les propositions faites au CERV [5, 6] pour atteindre cet objectif.

L'approche proposée sera ensuite comparée aux différentes techniques utilisées dans l'état de l'art.

Étapes principales du stage

- État de l'art sur la recherche d'architecture neuronale (NAS algorithmes) ainsi que la mesure de l'efficacité d'un réseau de neurones
- Proposition d'une (ou plusieurs) métrique pour l'évaluation de l'efficacité d'un réseau de neurones en plus de sa qualité.

- Familiarisation avec l’approche IFA et proposition d’une potentielle amélioration de l’approche (ou d’une nouvelle approche).
- Tests sur différentes bases de données de références (MNIST, CIFAR-10, etc..) et application de la proposition au domaine de la reconnaissance d’activités humaines.
- Comparaison avec l’état de l’art

Environnement de travail

- Langage : Python
- Keras ou Pytorch pour l’apprentissage automatique
- Matériel : Ordinateur portable

Contact

- **Encadrants** : Nassim Mokhtari et Alexis Nédélec
- **Contacts** : mokhtari@enib.fr nedelec@enib.fr
- **Lieu** : Centre Européen de Réalité Virtuelle (CERV)
- **Financement** : ENIB

References

- [1] X. Dong and Y. Yang. Nas-bench-201: Extending the scope of reproducible neural architecture search, 2020.
- [2] T. Elsken, J. H. Metzen, and F. Hutter. Neural architecture search: A survey. *Journal of Machine Learning Research*, 20(55):1–21, 2019. URL <http://jmlr.org/papers/v20/18-598.html>.
- [3] V. Lopes et al. Efficient Performance Estimation Without Training for Neural Architecture Search.” ICANN, “EPE-NAS, 2021.
- [4] J. Mellor, J. Turner, A. Storkey, and E. J. Crowley. Neural architecture search without training. In M. Meila and T. Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 7588–7598. PMLR, 18–24 Jul 2021. URL <https://proceedings.mlr.press/v139/mellor21a.html>.
- [5] N. Mokhtari, A. Nédélec, M. Gilles, and P. De Loor. Improving neural architecture search by mixing a firefly algorithm with a training free evaluation. In *2022 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, 2022. doi: 10.1109/IJCNN55064.2022.9892861.
- [6] N. Mokhtari, A. Nédélec, M. Gilles, and P. De Loor. Online human activity recognition using efficient neural architecture search with low environmental impact. In *Under review for the International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISAPP)*, 02 2024.

- [7] R. Schwartz, J. Dodge, N. Smith, and O. Etzioni. Green ai. *Communications of the ACM*, 63:54 – 63, 2019.
- [8] E. Strubell, A. Ganesh, and A. McCallum. Energy and policy considerations for deep learning in NLP. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3645–3650, Florence, Italy, July 2019. Association for Computational Linguistics. doi: 10.18653/v1/P19-1355. URL <https://aclanthology.org/P19-1355>.
- [9] Y. Sun, B. Xue, M. Zhang, G. G. Yen, and J. Lv. Automatically designing CNN architectures using the genetic algorithm for image classification. *IEEE Transactions on Cybernetics*, 50(9):3840–3854, sep 2020. doi: 10.1109/tcyb.2020.2983860. URL <https://doi.org/10.1109/tcyb.2020.2983860>.
- [10] X. Wang, M. He, L. Yang, H. Wang, and Y. Zhong. Human activity recognition based on an efficient neural architecture search framework using evolutionary multi-objective surrogate-assisted algorithms. *Electronics*, 12(1), 2023. ISSN 2079-9292. doi: 10.3390/electronics12010050. URL <https://www.mdpi.com/2079-9292/12/1/50>.
- [11] M. Wistuba, A. Rawat, and T. Pedapati. A survey on neural architecture search, 2019. URL <https://arxiv.org/abs/1905.01392>.
- [12] C. Ying, A. Klein, E. Christiansen, E. Real, K. Murphy, and F. Hutter. NAS-bench-101: Towards reproducible neural architecture search. In K. Chaudhuri and R. Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 7105–7114. PMLR, 09–15 Jun 2019. URL <https://proceedings.mlr.press/v97/ying19a.html>.