

Directives pour le projet de GLO-4030/7030 Apprentissage par réseaux de neurones profonds (23 août 2021)

Objectif L'objectif de ce projet est de vous permettre de pousser plus loin et d'appliquer des concepts vus en classe, sur un problème de votre choix. Pour les étudiants du 2^{ème} et 3^{ème} cycle, c'est aussi une manière d'explorer des approches de Deep Learning en lien avec votre sujet de recherche principal, si vous le désirez.

Choix du projet Un projet consistera à choisir un jeu de données compatible avec le Deep Learning (donc avec assez de données OU une architecture compatible pré-entraînée sur un grand jeu de données). Puis, avec un nombre suffisant d'entraînements et d'expérimentations, vous évalueriez la qualité des résultats obtenus, et testant quelques variantes soit au niveau de la méthodologie d'entraînement (data augmentation, semi-supervised learning, autoencodeurs, distillation, régularisation, diminuer le nombre de données utilisées en entraînement pour voir l'impact, etc) ou architecturale (comparer les résultats avec une deuxième architecture, si possible). Je poste des suggestions de projets ou de jeux de données ici : <https://ulaval-damas.github.io/glo4030/>

De surcroît pour les étudiants de GLO-7030, l'idéal serait en plus d'apporter une petite contribution, par exemple en explorant une nouveauté architecturale, en combinant deux approches ensembles, en exploitant des données non-étiquetées (semi-supervised learning) ou en appliquant les techniques de Deep Learning dans un domaine où il est peu utilisé pour l'instant (biologie, foresterie, transport, etc). Bien entendu, si le problème à résoudre lui-même est très difficile, il se peut que vous devriez plutôt mettre les énergies à le résoudre, plutôt qu'à tenter de faire des innovations. Il n'y a donc pas de formule unique. Si le temps le permet, faites une petite étude d'ablation pour pouvoir déterminer l'impact des composantes principales de votre approche. À ce sujet, relisez le paragraphe précédent pour vous inspirer. Dans tous les cas, prenez bien le soin de comprendre et d'analyser les résultats obtenus. Au final, vous devriez avoir plus d'un résultat quantitatif, pour vous permettre de voir des tendances.

Rapport à remettre Le rapport doit comporter au minimum les sections suivantes :

- Résumé (abstract)
- Introduction
- État de l'art (GLO-7030 seulement)
- Description de l'approche proposée, incluant l'architecture
- Méthodologie et expérimentation, incluant les détails d'entraînement
- Discussion des résultats, incluant ce qui n'a pas fonctionné, possibilités d'amélioration
- Conclusion

Vous pouvez ajouter plus de contenu, si vous le jugez pertinent. De plus, pour les étudiants de GLO-7030, le rapport doit être rédigé dans un format le plus proche possible d'un article scientifique. De ce fait, je recommande d'utiliser le gabarit *LaTex* des conférences internationales gérées par l'IEEE : <http://ras.papercept.net/conferences/support/tex.php>. Pour les étudiants de GLO-4030, le format typographique vous est laissé libre, mais il doit être convenable.

Évaluation Comme vous le constaterez sur la grille de correction, la majorité des points attribués au rapport ne le sont pas sur la réussite du projet, mais bien sur l'ensemble de la démarche (approche utilisée, méthodologie, expérimentation, analyse des résultats) et de sa bonne description. Je vous invite à consulter cette grille lors de l'élaboration de votre projet et des expérimentations qui vont s'y rattacher, pour maximiser l'efficacité de vos efforts. Notez aussi que l'évaluation pour GLO-7030 sera plus sévère, par exemple au niveau de la présentation des graphiques, de l'analyse des résultats, de la méthodologie, etc.

GLO-4030	
Critère	Score
Présentation du problème et du dataset	10 %
Présentation de l'approche et de l'architecture	13 %
Niveau de difficulté en fonction de la taille du groupe	15 %
Méthodologie et expérimentation, incluant les <i>training details</i>	25 %
Qualité et présentation des résultats obtenus	10 %
Discussion et conclusion (ce qui n'a pas marché, ce qui pourrait être ajouté, etc.)	12 %
Présentation générale du rapport	15 %
Total	100 %

GLO-7030	
Critère	Score
Présentation du problème et du dataset	10 %
Présentation de l'approche et de l'architecture, lien avec l'état de l'art	17 %
Originalité	6 %
Niveau de difficulté en fonction de la taille du groupe	12 %
Méthodologie et expérimentation, incluant les <i>training details</i>	20 %
Qualité et présentation des résultats obtenus	15 %
Discussion et conclusion (ce qui n'a pas marché, ce qui pourrait être ajouté, etc.)	10 %
Présentation générale du rapport	10 %
Total	100 %