

Rapport sur le mémoire
*Mitigating catastrophic forgetting via feature transfer and knowledge consolidation
for deep class-incremental learning*

présenté par
Quentin Ferdinand
en vue de l'obtention du grade de
Docteur de l'ENSTA Bretagne

Le mémoire de thèse de doctorat présenté par M. Quentin Ferdinand est un document de 119 pages composé de 4 chapitres accompagnés d'une introduction, d'une conclusion et d'une bibliographie d'environ 80 références. Il s'agit d'un manuscrit rédigé en anglais, agréable à lire, qui décrit avec pédagogie les travaux réalisés par M. Ferdinand dans le cadre d'une thèse CIFRE entre la société Naval Group (précédemment DCNS) et le laboratoire Lab-STICC, sous la direction de Benoit Clément (Professeur à l'ENSTA Bretagne) et le co-encadrement de Panagiotis Papadakis (IMT Atlantique), Gilles Le Chenadec (ENSTA Bretagne), et Quentin Oliverau (Naval Group).

Les travaux présentés par M. Ferdinand relèvent de l'apprentissage profond, un paradigme moderne de l'intelligence artificielle, et s'intéressent plus particulièrement à la problématique de l'apprentissage incrémental ou continu qui, dans le cadre d'une tâche de classification supervisée, vise à permettre à un modèle appris à partir d'un jeu de données étiquetées initial, d'évoluer en fonction de l'apparition de nouvelles classes ou de nouvelles données qui lui seraient fournies par la suite. Cette question est cruciale pour la société Naval Group, afin de permettre la mise à jour automatique et la configuration à la volée des systèmes embarqués de reconnaissance optique de navires. Une des principales difficultés auxquelles est confronté l'apprentissage incrémental réside dans le problème dit d'oubli catastrophique, où la prise en compte de nouvelles informations dans le modèle se fait au prix d'une perte des connaissances précédemment apprises. Les travaux de M. Quentin Ferdinand s'intéressent à cette question ouverte et de grande importance aujourd'hui en apprentissage profond, et les contributions proposées contribuent ainsi pleinement à l'état de l'art.

Le manuscrit débute par une très courte introduction qui pose brièvement le contexte de l'étude, ses objectifs, et décrit les trois contributions de l'auteur qui feront l'objet de chapitres dédiés : une analyse expérimentale des limites des solutions existantes pour l'apprentissage incrémental (chapitre 2), une méthode qui combine l'apprentissage contrastif et la distillation de connaissances incrémentale (chapitre 3, publication dans un workshop sur l'apprentissage continu à CVPR 2022), et enfin une méthode intitulée FECIL, basée sur les réseaux dynamiques et la compression de modèle qui vise à extraire de meilleures caractéristiques tout en réduisant le

coût calculatoire (chapitre 4). L'organisation du manuscrit suit donc logiquement le raisonnement scientifique mené par l'auteur dans le cadre de son aventure doctorale.

Le premier chapitre a pour objectif la présentation de l'état de l'art. Il débute par les approches d'apprentissage supervisé pour la classification d'image (notamment via les réseaux de neurones convolutifs), avant de se focaliser sur la problématique de l'apprentissage incrémental, décrivant notamment les difficultés auxquelles ce dernier fait face, et expliquant les scénarios relatifs à l'ajout incrémental de classes dans un modèle de classification. Sans que le choix soit clairement exprimé, on comprend que l'auteur s'intéressera au cas hors ligne, avec mémoire, et avec une architecture mono-tâche. Le chapitre se poursuit logiquement par une présentation des concepts existants en apprentissage incrémental, et l'auteur distingue trois façons d'aborder le problème : par révision (*rehearsal*), par régularisation, et par correction du biais favorisant les tâches (ou classes) les plus récentes. La présentation s'achève par la suggestion d'une méthode de référence, qui tire parti des avancées de l'état de l'art, et que l'auteur utilisera dans la suite de ses travaux.

Dans le second chapitre, l'auteur présente les résultats les plus saillants de l'analyse expérimentale qu'il a menée lors de sa première année de thèse pour comprendre le rôle joué par les différents composants d'un processus d'apprentissage incrémental, et pour identifier les pistes à suivre pour son amélioration. Cette démarche de compréhension par l'expérience est très intéressante et a certainement nécessité d'importants efforts de mise en œuvre et de synthèse. On pourra regretter que l'auteur ne positionne pas ses analyses face à l'état de l'art, ce qui nuit quelque peu à une pleine appréciation de l'impact des résultats obtenus (qui ne semblent d'ailleurs pas avoir fait l'objet d'une dissémination à la communauté). Le protocole expérimental, présenté en début de chapitre, s'appuie sur un unique jeu de données (CIFAR-100, alors qu'ImageNet-100/1000 seront utilisés dans les chapitres suivants) et une seule architecture (ResNet-32), et on peut légitimement s'interroger sur la validité des conclusions dans le cas général. Ces réserves mises à part, les analyses proposées par l'auteur sont particulièrement pertinentes et variées. Ainsi, M. Ferdinand se propose d'étudier les caractéristiques apprises par le réseau de neurones, en vérifiant notamment la pertinence des prototypes calculés à partir des caractéristiques moyennes de chaque classe, en mesurant les variances inter- et intra-classes, et en conduisant une analyse basée sur l'outil de visualisation t-SNE. Une conclusion inattendue concerne l'impact négatif que peut avoir le processus d'apprentissage incrémental sur les nouvelles classes, alors que l'oubli catastrophique concerne généralement les anciennes classes déjà vues par le modèle (on peut s'interroger dans ce contexte sur le rôle joué par certains composants tels que la réduction du biais, qui peut limiter en effet les performances sur les nouvelles classes). Éclairer les conclusions avec des mesures plus classiques de qualité de prédiction en sortie de réseau aurait sans doute permis de renforcer le poids des résultats obtenus. L'auteur s'intéresse ensuite à la stratégie de choix des prototypes à conserver dans la mémoire du réseau, et constate qu'elle a finalement peu d'incidence sur le résultat final. La dernière série d'expériences vise à étudier le mécanisme de distillation, et l'auteur, motivé par l'évolution de l'état de l'art en apprentissage par transfert, propose de s'intéresser non seulement à la distillation des sorties du réseau mais aussi à celle des caractéristiques intermédiaires. Les résultats obtenus montrent qu'il est possible d'éviter l'oubli catastrophique, au prix toutefois de performances de classification moindres sur les nouvelles classes. M. Ferdinand juge donc le compromis entre plasticité et stabilité non optimal, sans qu'une mesure formelle de ce compromis ne soit proposée. En outre, le scénario expérimental choisi dans cette section, qui consiste en un seul incrément de 50 classes (soit 50% du jeu de données) est peu réaliste, ce qui nuit à la portée des conclusions qui sont dressées. Les résultats présentés dans ce chapitre permettent néanmoins à l'auteur de justifier les orientations qu'il va suivre dans la suite de son manuscrit, motivé en cela par le problème inhérent à la dispersion des caractéristiques lors de l'ajout de nouvelles classes, et les promesses de la distillation de caractéristiques.

La première contribution méthodologique, présentée dans un workshop sur l'apprentissage incrémental à CVPR 2022, fait l'objet du chapitre 3, dont la plus grande partie (description de la méthode et des résultats) reprend cet

article dans son contenu original. L’auteur ajoute cependant une présentation de l’apprentissage contrastif non-supervisé et supervisé, et de leur usage dans des méthodes de distillation et d’apprentissage incrémental. On comprend que l’utilisation de l’apprentissage contrastif, qui consiste en l’ajout d’exemples et de contre-exemples créés artificiellement lors du processus d’entraînement d’un modèle, reste encore marginal dans le contexte de l’apprentissage incrémental, en particulier lorsque celui-ci s’appuie sur le principe de distillation de connaissances. S’inspirant de travaux récents (méthode CO2L publiée en 2021, déjà citée près de 200 fois), l’auteur propose un modèle où l’optimisation des poids du modèle n’est pas seulement basée sur les fonctions de coût usuelles d’entropie croisée et de distillation, mais également sur deux fonctions contrastives supplémentaires qui visent à améliorer respectivement les caractéristiques des nouvelles classes et les caractéristiques des anciennes classes conservées dans la mémoire. Ces choix sont bien motivés, très pertinents et exploitent judicieusement les avancées de l’état de l’art. Les expériences sont ici menées sur deux jeux de données de référence, et une comparaison est proposée avec quelques méthodes de l’état de l’art, dans une configuration (faible taille de lots) qui leur semble défavorable. On pourra ainsi s’interroger sur les raisons des très mauvaises performances rapportées pour la méthode de référence CO2L sur le jeu de données CIFAR-100, d’environ 20 à 30 points inférieures à celles indiquées dans l’article original. Dans le scénario choisi, la méthode proposée dépasse largement l’état de l’art, démontrant ainsi l’intérêt pour la communauté. L’étude expérimentale comporte également une analyse du rôle des différentes fonctions de coût, où l’auteur souligne leur apport systématique, bien que leur significativité reste à démontrer. En conclusion, la question du surcoût computationnel lié à l’ajout de deux vues contrastives est abordée comme motivation aux travaux présentés dans le chapitre suivant. Il aurait été intéressant d’intégrer ce paramètre dans l’étude comparative à l’état de l’art.

Dans le chapitre 4, l’auteur présente sa seconde méthode intitulée FECIL (Feature Expansion and enhanced Compression for Incremental Learning). Motivé par l’essor des réseaux dynamiques, et notamment par le succès de l’approche DER (publiée en 2021, déjà citée 300 fois) qui aborde le problème de l’oubli catastrophique en conservant les capacités mémorielles existantes auxquelles sont ajoutés des neurones supplémentaires pour apprendre les nouvelles classes, M. Ferdinand explore donc une stratégie radicalement différente de celle suivie lors des deux chapitres précédents. Pour autant, il n’oublie pas que l’oubli catastrophique en apprentissage incrémental ne peut être raisonnablement appréhendé avec une mémoire qui ne cesse de croître, et propose donc une solution qui exploite les propriétés des réseaux dynamiques tout en évitant une complexité croissante du modèle, grâce à une étape supplémentaire de compression de réseau. Ainsi, l’approche, qui suppose une forte parcimonie du réseau, procède en deux étapes : extension du modèle pour apprendre à reconnaître les nouvelles classes, puis simplification du modèle pour conserver la complexité initiale. En outre, M. Ferdinand, s’inspirant de résultats très récents (2022) ayant souligné l’intérêt de l’augmentation de données par mélange artificiel (*mixup*) pour améliorer les performances de l’apprentissage incrémental, suggère de ne pas limiter le mélange aux nouvelles classes, mais de l’effectuer entre les nouveaux exemples et ceux contenus dans la mémoire associée aux anciennes classes. Cette stratégie, particulièrement originale et pertinente puisqu’elle permet de limiter le biais favorisant les nouvelles classes, est mise en œuvre dans la phase de compression du réseau par distillation. Le chapitre débute logiquement par une présentation des réseaux dynamiques et du mélange de données, avant de présenter la méthode proposée par l’auteur, avec une formalisation adéquate de ses différents composants. Les expériences sont menées comme précédemment sur le jeu de données CIFAR-100, accompagné cette fois-ci du jeu ImageNet-1000, sans que le changement par rapport au chapitre précédent soit justifié. Les résultats obtenus démontrent non seulement les bonnes performances de la solution proposée par rapport à des approches concurrentes à taille de modèle fixe, mais aussi, et de façon plus surprenante, un niveau de compétitivité significatif face à l’approche DER pourtant favorisée par un nombre de paramètres croissant avec les incréments. Ce résultat important est brièvement discuté, et aurait probablement mérité d’être étudié plus en détail. En outre, la précision moyenne obtenue par l’auteur au cours des différents incréments est du même niveau que celle rapportée par une approche supervisée entraînée avec l’ensemble des données (la précision observée lors du dernier incrément restant cependant d’environ 10 points inférieure, illustrant que le problème

reste ouvert). Une représentation des classes par la méthode t-SNE est proposée, mais limitée à un seul incrément, ce qui ne permet pas d'apprécier le comportement des différentes stratégies d'apprentissage incrémental au fur et à mesure de l'apparition de nouvelles classes. A l'aide d'une étude systématique des différents composants de la méthode, M. Ferdinand illustre la pertinence de ses choix et notamment l'intérêt de la stratégie de mélange avec la mémoire qu'il propose. Celle-ci est d'ailleurs évaluée en complément d'autres approches de l'état de l'art, et les résultats obtenus montrent son intérêt systématique. Au vu des résultats expérimentaux très significatifs, les différentes contributions proposées dans ce chapitre mériteraient d'être largement diffusés (le manuscrit mentionne un article en cours de rédaction). La complexité du modèle aurait également pu être comparée à celle de la solution proposée au chapitre précédent, puisqu'il s'agit là de l'argument qui avait été mis en avant pour justifier cette nouvelle contribution.

En conclusion, M. Ferdinand rappelle tout d'abord, de façon très synthétique, les points saillants de son raisonnement qui l'ont amené à proposer plusieurs contributions originales en apprentissage incrémental. Il propose ensuite rapidement quelques perspectives de recherche, en lien avec le processus de distillation, l'apprentissage contrastif, les approches pour l'apprentissage incrémental, ou encore l'utilisation de méthodes génératives de dernière génération comme celles utilisées dans le cadre auto-supervisé. On pourra regretter que ne soit pas envisagé la comparaison ou le couplage entre les deux approches proposées dans les chapitres 3 et 4, ni discuté le scénario applicatif (reconnaissance de navires) mis en avant en introduction du manuscrit. La bibliographie d'environ 80 références, dont 15 publiées après le début de thèse, clôture le document.

Pour conclure, le mémoire de thèse de M. Quentin Ferdinand décrit une étude bien construite et présentée sur la problématique de l'apprentissage incrémental. Les trois chapitres principaux retracent le déroulement chronologique de la thèse, et illustrent la pertinence de la démarche scientifique, qui a débuté par une compréhension des limites de l'état de l'art à l'aide d'analyses expérimentales, et s'est poursuivie par la présentation de deux méthodes complémentaires inspirées des avancées récentes de l'état de l'art. Les résultats obtenus sur des jeux de données publics, et comparés à des méthodes de référence, illustrent l'importance des travaux, qui mériteraient de bénéficier d'une dissémination accrue pour accroître leur impact dans la communauté. Sous réserve du respect des contraintes imposées par le contexte industriel de la thèse, celui-ci serait également favorisé par une mise à disposition des codes logiciels, non évoquée dans le manuscrit.

Au vu de la pertinence scientifique des travaux menés par l'auteur et de leur restitution de qualité dans le manuscrit de thèse, j'émet un **avis favorable** à la présentation orale des travaux de M. Quentin Ferdinand, en vue de l'obtention du grade de Docteur en STIC – Informatique de l'ENSTA Bretagne.

Vannes, le 19 novembre 2023

Sébastien Lefèvre

