

Projet « Deep Learning »

Teddy NKODIA

Sigma Clermont
Mastère spécialisé Expert en Science des données

Table des matières

1.	Introduction	2
2.	Synthèse de l'article principal (<i>[Cao and Choe, 2020]</i>)	2
2.1.	Présentation du contexte	2
2.2.	Explication des données utilisées et du prétraitement	2
2.3.	Description des modèles de deep learning utilisés	3
2.4.	Comparaison des performances entre différentes approches	3
3.	Synthèse des articles complémentaires sur le transfert learning	4
3.1.	Synthèse de l'article " <i>Transfer Learning for High Resolution Aerial Image Classification</i> " (<i>[Liang et al., 2016]</i>)	4
3.1.1.	Contexte et Objectifs	4
3.1.2.	Méthodologie	4
3.1.3.	Résultats et Conclusion	4
3.2.	Synthèse de l'article " <i>Transfer Learning for Classification of Optical Satellite Image</i> " (<i>[Zou et Zhong, 2018]</i>)	5
3.2.1.	Contexte et Objectifs	5
3.2.2.	Méthodologie	5
3.2.3.	Résultats et Conclusion	5
3.3.	Comparaison et Conclusions Générales	5
4.	Application dans l'étude de cas sur la classification des bâtiments endommagés	5
4.1.	Explication du principe du transfert learning	6
4.2.	Importance et utilité en deep learning	6
4.3.	Techniques principales : extraction de caractéristiques vs fine-tuning	6
4.4.	Avantages et inconvénients	7
5.	Implémentation en Python (PyTorch)	7
5.1.	Reproduction de l'expérience de <i>[Cao and Choe, 2020]</i> avec VGG-16	7
5.2.	Étude comparative des types de transfer learning (types A et B)	9
6.	Conclusion	12

1. Introduction

Ce rapport présente les expérimentations menées dans le cadre du projet, qui visent à explorer différentes approches de classification automatique des bâtiments endommagés à partir d'images satellites post-catastrophe. Le protocole expérimental s'appuie sur la reproduction d'une étude existante – celle de [Cao and Choe, 2020] – puis sur la mise en œuvre de plusieurs variantes modernes du transfert learning, selon les approches de type A (feature extraction) et type B (fine-tuning partiel).

Les modèles testés (VGG-16, ResNet18, DenseNet121) ont été sélectionnés pour leur architecture éprouvée en classification d'images. Chaque expérience a été conduite sur 10 époques, en raison des ressources limitées, tout en conservant des évaluations rigoureuses selon les métriques standards : **accuracy**, **AUC** et **F1-score**. L'objectif est de comparer ces approches en termes de performance, de stabilité et de coût computationnel, dans un contexte où les données sont spécifiques, déséquilibrées et potentiellement rares.

2. Synthèse de l'article principal (*[Cao and Choe, 2020]*)

2.1. Présentation du contexte

L'article de Cao et Choe (2020) s'inscrit dans le cadre de l'analyse des dégâts causés par l'ouragan Harvey en 2017 sur la région du Greater Houston, Texas. L'ouragan Harvey, classé en catégorie 4, a entraîné d'importants dommages, notamment des inondations massives et la destruction de nombreuses infrastructures. L'évaluation des dégâts après une catastrophe naturelle est une étape cruciale pour optimiser les efforts de secours et la gestion des ressources.

Traditionnellement, cette évaluation repose sur des inspections manuelles et des analyses d'images satellites par des experts, une approche longue et coûteuse. L'objectif de l'étude est de proposer une solution automatisée basée sur des modèles de deep learning pour classifier automatiquement les bâtiments comme endommagés ou non à partir d'images satellites.

2.2. Explication des données utilisées et du prétraitement

L'étude repose sur un ensemble de données comprenant des images satellites prises avant et après le passage de l'ouragan Harvey. Ces images proviennent de la métropole de Houston et ont été annotées manuellement par des volontaires afin de créer un jeu de données labellisées en deux catégories :

- Damage (endommagé) : bâtiments ayant subi des dégâts.
- No Damage (non endommagé) : bâtiments restés intacts.

Les ensembles de données sont organisés comme suit :

- 10 000 images pour l'apprentissage (réparties équitablement entre les deux classes).
- 2 000 images pour la validation.

- 9 493 images pour le test (non équilibrées, reflétant mieux la réalité du terrain).

Un prétraitement est effectué pour améliorer la qualité des données et optimiser l'apprentissage des modèles :

- Augmentation des données : rotation, retournement horizontal, ajustement de la luminosité et du contraste.
- Redimensionnement des images : normalisation des tailles pour correspondre aux dimensions des réseaux de neurones pré-entraînés (150x150 pixels).
- Normalisation des pixels : mise à l'échelle des valeurs de pixel pour correspondre aux valeurs attendues par les architectures CNN.

2.3. Description des modèles de deep learning utilisés

L'article met en avant l'utilisation du transfert learning avec des réseaux de neurones convolutifs (CNN). Deux approches principales sont comparées :

- Extraction de caractéristiques : Utilisation d'un modèle CNN pré-entraîné (VGG-16) pour extraire les caractéristiques des images, suivie d'une classification avec un modèle plus simple comme la régression logistique ou une SVM.
- Fine-tuning du CNN : Adaptation du modèle pré-entraîné en ajustant les poids des couches supérieures tout en conservant les poids des premières couches fixes.

Le modèle principal utilisé est VGG-16, un réseau convolutif profond avec 16 couches d'apprentissage. L'intérêt du VGG-16 réside dans sa capacité à extraire des caractéristiques visuelles pertinentes, même pour des images différentes de celles présentes dans le jeu de données ImageNet sur lequel il a été initialement entraîné.

2.4. Comparaison des performances entre différentes approches

Les résultats de l'étude montrent que la meilleure approche est le fine-tuning du VGG-16 avec augmentation des données. Voici un résumé des performances obtenues :

- Modèle VGG-16 fine-tuné avec augmentation de données : 97,08% de précision sur l'ensemble de test.
- VGG-16 avec extraction de caractéristiques : Précision inférieure (autour de 95%) par rapport au fine-tuning.
- Modèles basiques (régression logistique, SVM) : Précision plus faible (à peine 91-92%).

Les conclusions de l'étude montrent que le fine-tuning des dernières couches d'un réseau CNN pré-entraîné est plus efficace que l'utilisation d'un classifieur sur des caractéristiques extraites. En outre, l'augmentation des données permet de réduire le sur-apprentissage et d'améliorer la généralisation du modèle.

L'étude de Cao et Choe (2020) démontre l'efficacité du transfert learning et du fine-tuning pour la classification d'images satellites en contexte de catastrophe naturelle. L'utilisation d'un CNN pré-entraîné comme VGG-16, combinée à un ajustement adaptatif des dernières

couches, permet d'obtenir des performances élevées et de surpasser les méthodes traditionnelles.

Cette approche pourrait être étendue à d'autres catastrophes naturelles et appliquée à des systèmes de surveillance automatisés pour réagir plus rapidement aux crises humanitaires.

3. Synthèse des articles complémentaires sur le transfert learning

Le Transfer Learning (TL) est une technique essentielle pour la classification d'images satellitaires, en particulier lorsque les ensembles de données labellisées sont limités. Les deux articles analysés ([Liang et al., 2016] et [Zou et Zhong, 2018]) explorent l'utilisation des réseaux de neurones convolutifs (CNNs) pré-entraînés sur de grands ensembles de données (comme ImageNet) pour améliorer la classification des images aériennes et optiques satellitaires.

3.1. Synthèse de l'article "*Transfer Learning for High Resolution Aerial Image Classification*" ([Liang et al., 2016])

3.1.1. Contexte et Objectifs

Cet article examine l'application du TL pour classifier des images aériennes à haute résolution. Les auteurs soulignent que le manque d'images labellisées rend difficile l'entraînement des CNNs en partant de zéro. Le TL permet d'exploiter des modèles pré-entraînés pour adapter la classification à des images satellites spécifiques.

3.1.2. Méthodologie

Les auteurs explorent deux approches :

- **Extraction de caractéristiques** : Utilisation de CNNs pré-entraînés pour extraire des descripteurs d'images, suivie d'un classifieur (ex: SVM).
- **Fine-tuning des CNNs** : Ajustement des couches supérieures du réseau sur les images cibles.

Les architectures AlexNet, VGG-F, GoogleNet et ResNet sont testées.

3.1.3. Résultats et Conclusion

- **ResNet-152** offre la meilleure précision (~96.24%) lorsqu'il est utilisé comme extracteur de caractéristiques.
- Le fine-tuning des dernières couches du CNN améliore les performances mais peut provoquer un sur-apprentissage si le jeu de données est trop petit.

Les auteurs concluent que le TL est une stratégie efficace pour la classification d'images aériennes, en particulier avec le fine-tuning des couches supérieures.

3.2. Synthèse de l'article "*Transfer Learning for Classification of Optical Satellite Image*" ([Zou et Zhong, 2018])

3.2.1. Contexte et Objectifs

Cet article se concentre sur l'application du TL à la classification d'images satellitaires optiques. L'objectif est d'examiner comment le fine-tuning d'un CNN pré-entraîné peut améliorer la classification en présence d'un jeu de données limité.

3.2.2. Méthodologie

- **Expérimentation avec AlexNet** : Le modèle est pré-entraîné sur ImageNet et fine-tuné couche par couche.
- **Impact du fine-tuning partiel** : Les auteurs testent différents degrés de fine-tuning pour déterminer l'approche optimale.

3.2.3. Résultats et Conclusion

- Un fine-tuning partiel (uniquement les dernières couches) permet d'atteindre une précision similaire à celle d'un fine-tuning complet, avec une convergence plus rapide.
- La stratégie évitant d'entraîner les premières couches réduit le risque de sur-apprentissage.

L'article valide que le TL est une solution efficace pour classifier des images satellitaires, tout en optimisant le processus d'entraînement.

3.3. Comparaison et Conclusions Générales

Critère	Liang et al. (2016)	Zou et Zhong (2018)
Domaine	Images aériennes haute résolution	Images satellites optiques
Modèles testés	AlexNet, VGG, GoogleNet, ResNet	AlexNet
Approches TL	Extraction de caractéristiques et fine-tuning	Fine-tuning couche par couche
Meilleure stratégie	Fine-tuning partiel des couches supérieures	Fine-tuning partiel des dernières couches
Meilleur modèle	ResNet-152	AlexNet

4. Application dans l'étude de cas sur la classification des bâtiments endommagés

Dans notre étude de cas, les images satellites diffèrent significativement des images d'ImageNet, tant en termes de contenu que de labels ("damage" et "no damage" au lieu de catégories d'objets). Toutefois, le TL reste pertinent, car il permet aux modèles de DL d'apprendre des filtres et opérateurs de convolution utiles à tout type d'image.

Nous allons donc explorer deux stratégies principales dans notre projet :

- **(A) Utilisation des features d'un DL avant la couche dense** : Extraction des caractéristiques des couches profondes d'un modèle CNN pré-entraîné, suivie d'un apprentissage avec un classifieur traditionnel (SVM, régression logistique).
- **(B) Fine-tuning d'un modèle existant** : Entraînement des dernières couches d'un CNN pré-entraîné sur les nouvelles données, tout en gardant les premières couches fixes.

Ces deux approches seront mises en œuvre et comparées afin d'optimiser la classification des bâtiments endommagés.

4.1. Explication du principe du transfert learning

Le transfert learning repose sur l'idée que les connaissances acquises lors de la résolution d'une tâche peuvent être appliquées à une autre tâche. En deep learning, cela signifie utiliser un modèle pré-entraîné sur une grande base de données pour une tâche spécifique et adapter ce modèle à une nouvelle tâche. Cette approche est particulièrement utile lorsque les données disponibles pour la nouvelle tâche sont limitées ou lorsque l'entraînement d'un modèle à partir de zéro serait trop coûteux en termes de temps et de ressources.

4.2. Importance et utilité en deep learning

Le transfert learning est devenu une technique essentielle en deep learning pour plusieurs raisons :

1. **Réduction du temps d'entraînement** : En utilisant un modèle pré-entraîné, le temps nécessaire pour entraîner un nouveau modèle est considérablement réduit. Par exemple, un modèle de classification d'images pré-entraîné sur ImageNet peut être utilisé pour classer des images médicales avec un temps d'entraînement beaucoup plus court.
2. **Amélioration des performances** : Les modèles pré-entraînés sur de grandes bases de données peuvent capturer des caractéristiques complexes qui améliorent les performances sur des tâches spécifiques. Par exemple, les modèles de reconnaissance d'objets pré-entraînés peuvent détecter des objets dans des images avec une grande précision.
3. **Utilisation efficace des ressources** : Le transfert learning permet de tirer parti des ressources computationnelles investies dans l'entraînement initial du modèle. Cela est particulièrement utile pour les organisations avec des ressources limitées.

4.3. Techniques principales : extraction de caractéristiques vs fine-tuning

Il existe deux principales techniques de transfert learning :

Extraction de caractéristiques : Cette technique consiste à utiliser les couches d'un modèle pré-entraîné pour extraire des caractéristiques des données. Ces caractéristiques sont ensuite utilisées pour entraîner un nouveau modèle. Par exemple, les couches convolutionnelles d'un modèle de classification d'images peuvent être utilisées pour extraire des caractéristiques visuelles, qui sont ensuite utilisées pour entraîner un classificateur simple.

Fine-tuning : Cette technique implique de prendre un modèle pré-entraîné et de réentraîner certaines de ses couches (souvent les couches supérieures) sur les nouvelles données. Cela permet d'adapter le modèle aux spécificités de la nouvelle tâche. Par exemple, un modèle de classification d'images pré-entraîné peut être ajusté pour détecter des anomalies dans des images médicales en réentraînant les couches supérieures sur des images médicales.

4.4. Avantages et inconvénients

Avantages :

- **Efficacité** : Le transfert learning réduit le temps et les ressources nécessaires pour entraîner un modèle. Cela permet de développer des modèles performants rapidement, même avec des données limitées.
- **Flexibilité** : Le transfert learning peut être appliqué à diverses tâches avec des ajustements minimes. Par exemple, un modèle de reconnaissance d'objets peut être adapté pour détecter des objets spécifiques dans des images de surveillance.
- **Performance** : Les modèles pré-entraînés peuvent offrir des performances supérieures, surtout lorsque les données sont limitées. Par exemple, un modèle de classification d'images pré-entraîné peut atteindre une précision élevée même avec un petit ensemble de données.

Inconvénients :

- **Dépendance aux données initiales** : La qualité du transfert learning dépend fortement de la qualité et de la pertinence des données sur lesquelles le modèle initial a été entraîné. Si les données initiales sont trop différentes des nouvelles données, les performances peuvent être limitées.
- **Complexité** : Le fine-tuning peut être complexe et nécessite une compréhension approfondie du modèle pré-entraîné et de la nouvelle tâche. Par exemple, ajuster les hyperparamètres et les couches du modèle peut être difficile et nécessiter des essais et des erreurs.
- **Risques de surajustement** : Il y a un risque de surajustement si le modèle est trop ajusté aux nouvelles données, surtout si elles sont limitées. Cela peut entraîner une mauvaise généralisation du modèle sur des données non vues.

5. Implémentation en Python (PyTorch)

5.1. Reproduction de l'expérience de [Cao and Choe, 2020] avec VGG-16

Dans un premier temps, nous avons reproduit les expériences présentées dans l'article de [Cao and Choe, 2020], en utilisant le modèle pré-entraîné VGG-16. Ce modèle a été choisi car il est explicitement mentionné dans l'article comme base de l'approche proposée, combinant l'utilisation de réseaux de neurones convolutifs (CNN) et de transfer learning pour détecter les bâtiments endommagés après le passage de l'ouragan Harvey. L'architecture VGG-16 a été utilisée avec un fine-tuning partiel : seules les dernières couches fully connected du réseau ont

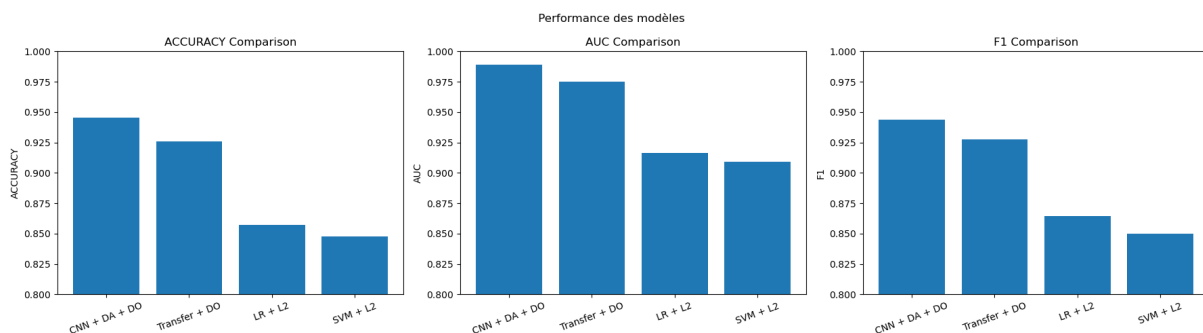
été réentraînées, tandis que les couches convolutionnelles initiales ont été gelées afin de conserver les représentations génériques apprises sur ImageNet. Cette approche a permis d'obtenir des résultats cohérents avec les observations de l'article, avec une précision de validation qui progresse régulièrement au fil des époques, indiquant une bonne capacité d'adaptation du modèle aux spécificités des images satellites post-catastrophe.

En complément, nous avons également reproduit plusieurs autres variantes expérimentales issues de l'article, à savoir :

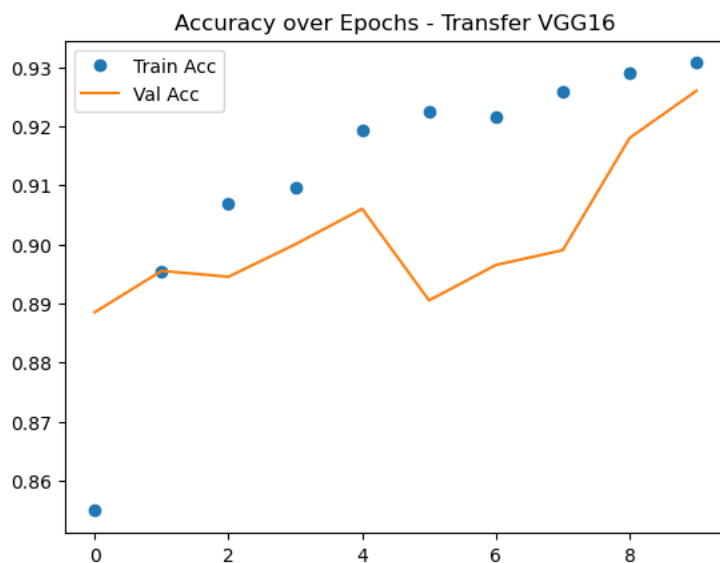
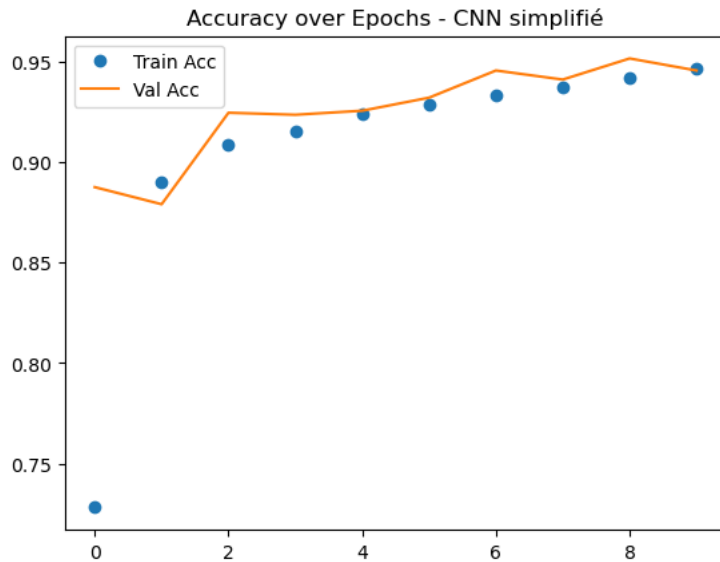
- Un CNN entraîné from scratch, combiné à de la data augmentation (DA) et du dropout (DO) pour limiter le surapprentissage,
- Une version Transfer Learning + DO utilisant VGG-16 pré-entraîné,
- Un modèle de régression logistique (LR) avec régularisation L2,
- et un SVM (Support Vector Machine) également avec régularisation L2.

Toutes ces expériences ont été conduites sur 10 époques uniquement, en raison des limitations matérielles de la machine utilisée. Malgré ce cadre réduit, les résultats obtenus restent très satisfaisants et comparables à ceux rapportés dans l'article. Le CNN + DA + DO a obtenu les meilleurs résultats, avec une accuracy de 94.55 %, un AUC de 0.989 et un score F1 de 0.9435. Le modèle Transfer + DO a également bien performé (accuracy 92.60 %), suivi des approches plus classiques de Logistic Regression (85.70 %) et SVM (84.75 %). Ces résultats confirment l'intérêt du deep learning, et plus particulièrement du transfer learning, dans des contextes de données limitées mais bien annotées comme c'est le cas ici.

	accuracy	auc	f1
CNN + DA + DO	0.9455	0.989132	0.943553
Transfer + DO	0.9260	0.975040	0.927522
LR + L2	0.8570	0.916254	0.864198
SVM + L2	0.8475	0.909402	0.849975



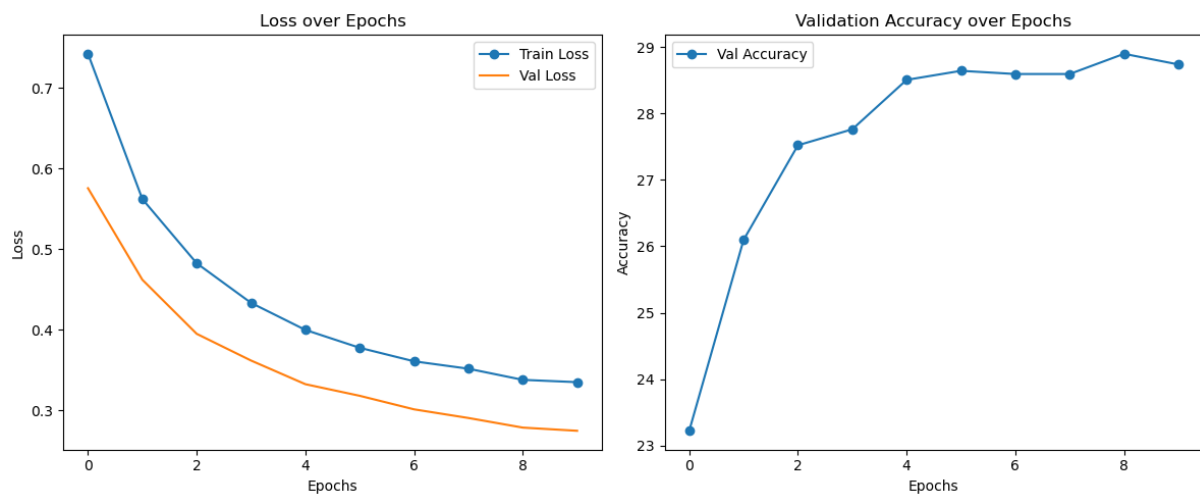
En comparant les courbes de performance, on observe que le CNN entraîné from scratch atteint une accuracy équivalente à VGG-16 (~95 %) en fin d'entraînement, mais sa courbe démarre plus bas et évolue de manière moins régulière, indiquant une période de stabilisation plus longue. En revanche, le modèle VGG-16 affiche une progression plus stable dès les premières époques, ce qui souligne l'apport du transfert de connaissances initiales.



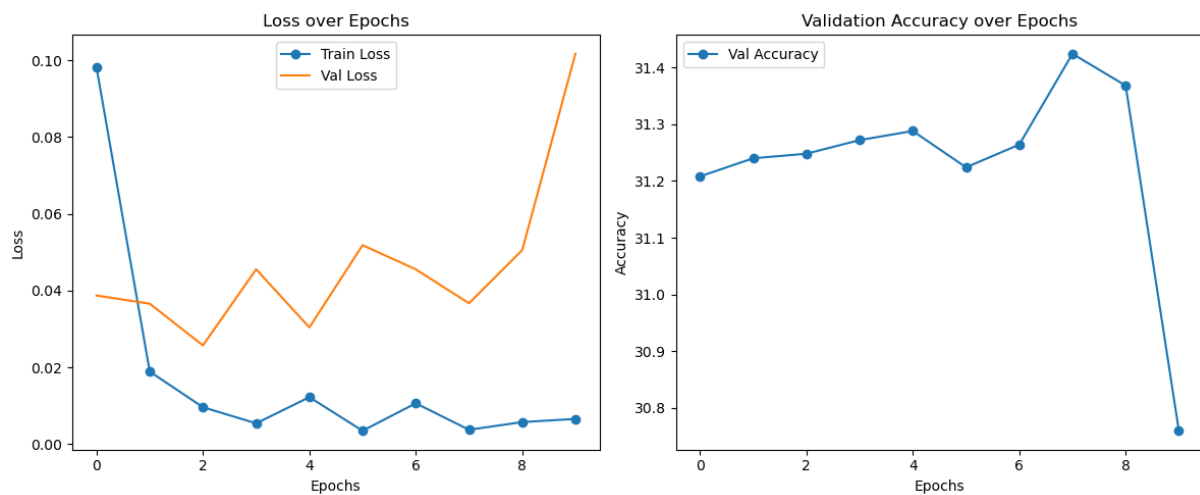
5.2. Étude comparative des types de transfer learning (types A et B)

Dans un second temps, une étude comparative des deux approches classiques de transfer learning a été menée à l'aide des modèles ResNet18 et DenseNet121, conformément à la typologie demandée dans le sujet.

L'approche de type A consiste à utiliser un modèle pré-entraîné comme extracteur de caractéristiques fixes : toutes les couches du backbone sont gelées, et seule une nouvelle tête de classification est ajoutée et entraînée sur les données satellites. Cette approche, testée avec ResNet18, a donné des résultats corrects mais limités en termes de précision brute ($\approx 29\%$). Toutefois, les métriques globales restent très satisfaisantes avec un AUC de 0.9654 et un F1 Score de 0.9040. Les courbes d'apprentissage montrent une diminution progressive des pertes d'entraînement et de validation, ainsi qu'une montée rapide de l'accuracy jusqu'à saturation, indiquant une convergence maîtrisée mais un plafond de performance atteint précocement.

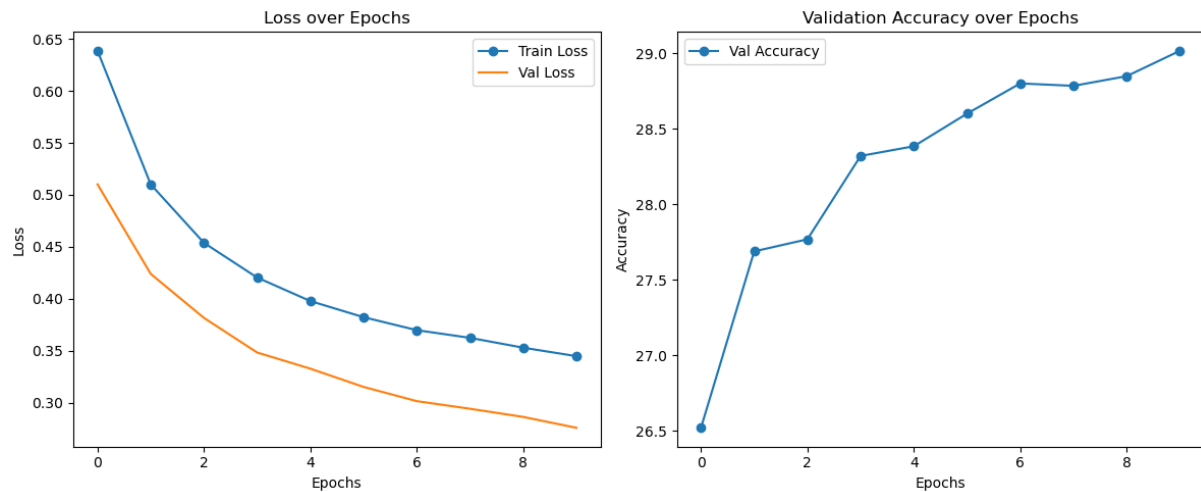


ResNet18 Type A

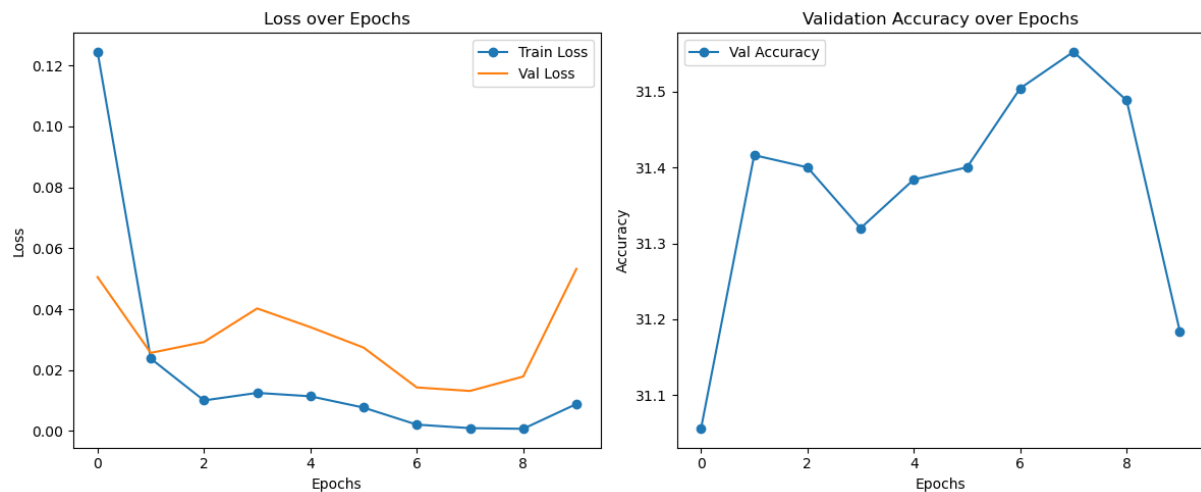


ResNet18 Type B

Avec DenseNet121, les résultats sont légèrement meilleurs pour la même approche : AUC de 0.9679 et F1 Score de 0.9120, avec des courbes similaires mais une progression plus fluide de l'accuracy. Cela confirme la robustesse intrinsèque de DenseNet121, même sans réentraînement des couches profondes.



DenseNet 121 Type A



DenseNet 121 Type B

L'approche de type B, quant à elle, repose sur un fine-tuning partiel du modèle pré-entraîné : seules les dernières couches convolutionnelles et la tête de classification sont réentraînées, ce qui permet au réseau d'adapter ses filtres aux spécificités visuelles des images satellites post-catastrophe. Cette stratégie, appliquée à ResNet18, a permis une nette amélioration des performances, avec un AUC de 0.9988 et un F1 Score de 0.9702. Cependant, les courbes de loss révèlent un surapprentissage marqué, avec un écart croissant entre la loss d'entraînement (proche de zéro) et la loss de validation qui augmente à partir de la 5e époque. L'accuracy, bien que stable et élevée (~31.4 %), chute brutalement en fin d'entraînement.

En revanche, le DenseNet121 en type B affiche une performance encore supérieure avec un AUC de 0.9985 et un F1 Score de 0.9844, tout en montrant une meilleure stabilité sur la validation. Malgré une légère oscillation dans la courbe d'accuracy, les performances se maintiennent sur la durée, ce qui suggère un meilleur compromis entre capacité d'apprentissage et généralisation.

Ces résultats confirment l'intérêt du fine-tuning (type B) dans des contextes de données spécifiques et complexes comme les images satellites post-catastrophe. Cependant, ils soulignent également l'importance d'un encadrement rigoureux de l'entraînement pour éviter le surajustement, notamment avec des architectures plus simples comme ResNet18. En parallèle, l'approche de type A, bien que plus simple et moins coûteuse en calcul, permet déjà d'atteindre des résultats robustes, en particulier lorsqu'elle est appliquée à des architectures performantes telles que DenseNet121.

Modèle	Type de transfert	Accuracy (%)	AUC	F1 Score	Commentaire
ResNet18	Type A	~29.0	0.9654	0.9040	Courbe stable, bonne convergence mais plafonnement rapide.
ResNet18	Type B	~31.4	0.9988	0.9702	Très bon score, mais surapprentissage marqué (loss val ↑).
DenseNet121	Type A	~29.0	0.9679	0.9120	Performances solides, bonne stabilité, généralisation correcte.
DenseNet121	Type B	~31.6	0.9985	0.9844	Meilleure performance globale, très bon équilibre entraînement/validation.

- Les valeurs d'accuracy sont celles observées sur les courbes, légèrement plafonnées par le nombre réduit d'époques (10).
- Les scores AUC et F1 confirment cependant une très bonne qualité de séparation, notamment pour les modèles fine-tunés.
- Le DenseNet121 en type B est globalement le modèle le plus performant dans cette étude.

6. Conclusion

Ce projet a permis d'explorer de manière approfondie l'apport du deep learning et du transfer learning pour la classification d'images satellites post-catastrophe, dans le but d'identifier automatiquement les bâtiments endommagés. Dans un premier temps, la reproduction de l'étude de *[Cao and Choe, 2020]* à l'aide du modèle VGG-16 pré-entraîné a confirmé l'efficacité du fine-tuning partiel associé à des techniques de régularisation telles que le dropout et la data augmentation. Les performances obtenues, à la fois en accuracy, AUC et F1-score, sont restées élevées malgré une contrainte de 10 époques d'entraînement.

Dans un second temps, l'étude comparative des approches de transfert type A (feature extractor) et type B (fine-tuning) avec les modèles ResNet18 et DenseNet121 a mis en lumière plusieurs enseignements. D'une part, le type A offre une solution simple, stable et peu coûteuse, capable d'atteindre des résultats corrects sans réentraînement massif. D'autre part, le type B permet une adaptation plus fine aux données spécifiques, et atteint les meilleurs

scores globaux, en particulier avec DenseNet121 (F1-score de 0.9844, AUC de 0.9985), tout en maintenant une bonne stabilité.

L'analyse fine des courbes d'apprentissage a également permis d'identifier les risques de surapprentissage liés à un fine-tuning mal contrôlé, soulignant ainsi l'importance d'un encadrement méthodologique rigoureux.

En conclusion, ce projet illustre concrètement la puissance et la flexibilité du transfer learning pour des cas d'usage critiques comme la détection de dommages post-catastrophe, tout en apportant des recommandations pratiques pour choisir la bonne approche en fonction des contraintes de données et de calcul.