

Classification Automatisée des Microstructures Martensitiques et Perlites à Partir de Données Internet

Résumé du rapport disponible en fin de rapport si besoin.

Présentation du Projet

Ce projet vise à développer un système de vision par ordinateur capable de classifier automatiquement deux types de microstructures métalliques : la martensite et la perlite. Un aspect clé de ce projet réside dans l'utilisation de données extraites d'Internet, rendant le projet non seulement innovant mais aussi accessible, puisque les données sont disponibles pour tous. La qualité et l'uniformité de ces données jouent un rôle crucial, influençant directement la précision du modèle de classification. À travers l'expérimentation de diverses architectures de réseaux de neurones convolutifs, y compris l'adaptation des couches finales de modèles pré-entraînés pour correspondre à nos données spécifiques, nous visons à optimiser les performances de notre système.

État de l'Art

Notre projet s'inscrit dans le domaine en expansion rapide de la vision par ordinateur, qui applique les techniques de machine learning et de deep learning à l'interprétation des images capturées par caméra. L'intelligence artificielle, englobant le machine learning. Dans le machine learning existe le deep learning qui consiste à faire apprendre de manière automatique et hiérarchique les caractéristique des données. Il utilise des réseaux de neurones artificiels profonds, qui sont inspirés par le fonctionnement des neurones dans le cerveau humain. Ces modèles sont composés de multiples couches cachées, chacune capable de détecter des caractéristiques à différents niveaux d'abstraction

Objectif

Le but central de ce projet est de créer un modèle hautement fiable et précis pour la classification de microstructures martensitiques et perlites, en dépit des défis posés par l'utilisation de données extraites d'Internet. Ces données, souvent hétérogènes et de qualité inégale, peuvent induire un risque de surajustement et affecter la généralisation du modèle. Nous nous attelons donc à optimiser l'accuracy du modèle tout en perfectionnant les méthodes de nettoyage et de préparation du dataset. L'objectif est double : améliorer la robustesse du modèle face aux imperfections des données et, par extension, avancer les capacités de la vision par ordinateur dans le champ de la métallurgie. Ce projet ambitionne de démontrer qu'avec des techniques de traitement de données sophistiquées, il est possible d'extraire des connaissances précises à partir de sources diverses et imparfaites, ouvrant la voie à des applications novatrices en recherche et industrie.

Préparation du Dataset Custom

La création de notre dataset custom a nécessité une approche méthodique et rigoureuse, débutant par la collecte d'images à partir de sources en ligne via des techniques de scraping avancées. Ce processus initial a généré un vaste ensemble d'images brutes, parmi lesquelles nous avons dû identifier et extraire les microstructures pertinentes de martensite et de perlite.

Étapes clés de la préparation :

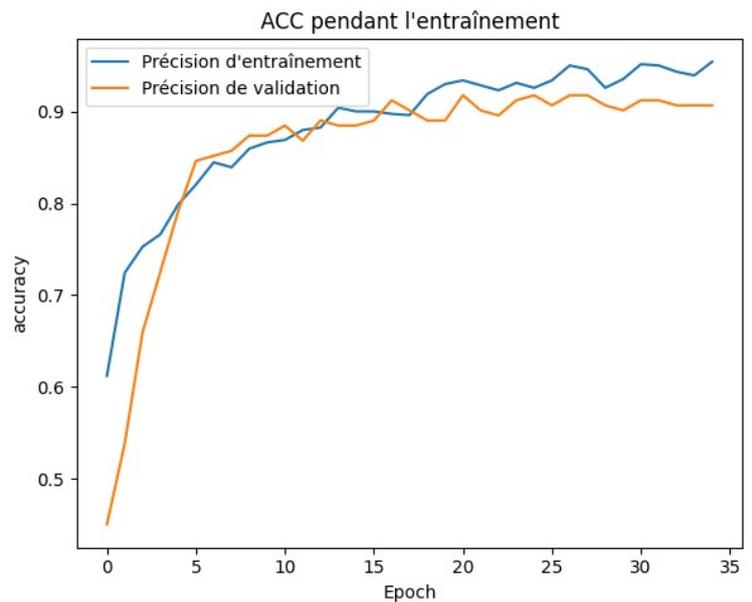
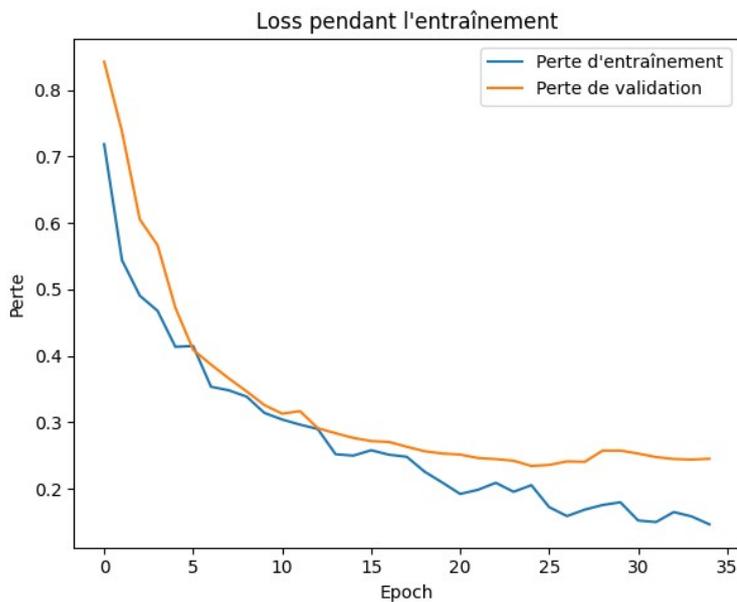
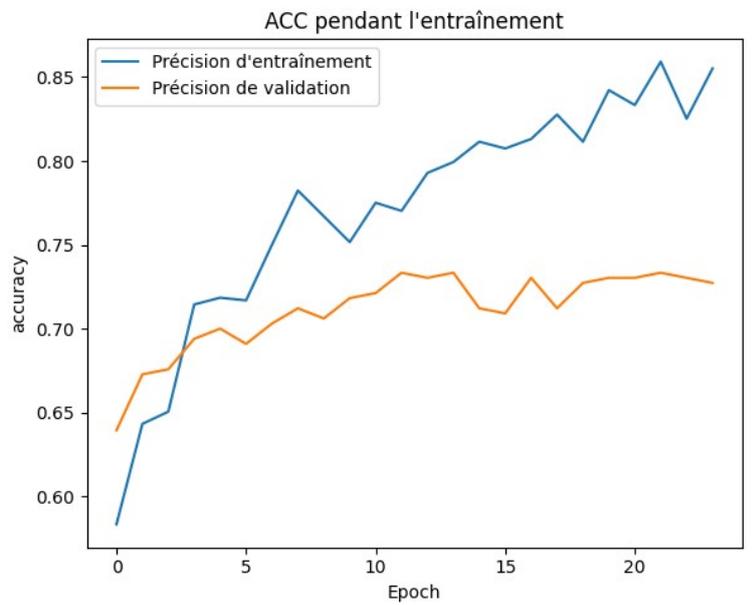
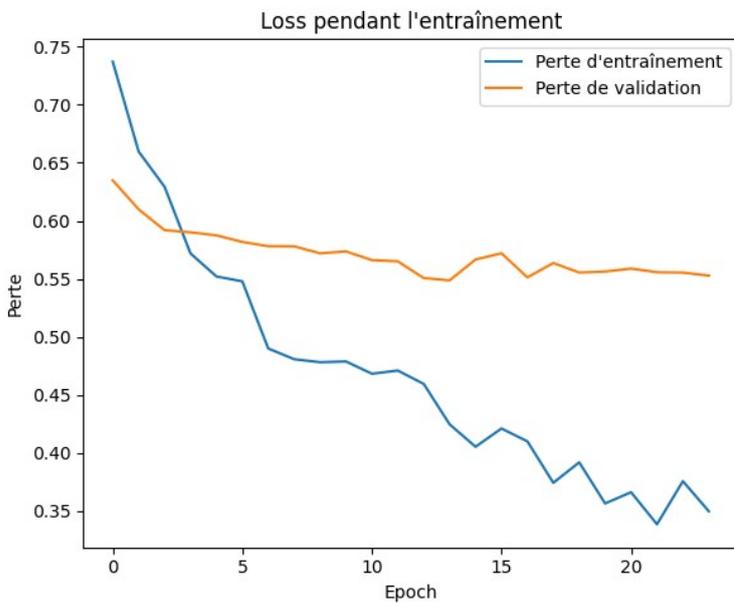
1. **Scraping des Images** : Mise en place d'outils de scraping pour collecter des images brutes directement depuis des sources Internet.
2. **Prétraitement** :
 - **Upscaling** : Amélioration de la résolution des images pour une meilleure analyse et identification des caractéristiques.
 - **Cropping Automatique** : Utilisation d'un modèle pré-entraîné sur un dataset plus restreint pour détecter et extraire automatiquement les zones d'intérêt (ROI) des images. Bien que ce modèle ait facilité l'identification des ROI, une annotation manuelle de 200 images a été nécessaire pour affiner notre approche.
3. **Validation du Dataset** : Développement d'un pré-modèle de classification entraîné sur un petit ensemble d'images de haute qualité vérifiées manuellement. Ce pré-modèle a été utilisé pour filtrer l'ensemble des images collectées, rejetant celles dont la classification présentait une faible confiance.
4. **Nettoyage et Étiquetage** : Le nettoyage approfondi et l'étiquetage précis des images restantes ont permis de constituer un dataset cohérent, essentiel pour l'entraînement efficace de nos modèles de classification avancés.
5. **Entraînement du Modèle** : Avec un dataset nettoyé et rigoureusement préparé, l'étape finale a été l'entraînement de notre modèle de classification, visant à atteindre une précision optimale dans la distinction entre les microstructures de martensite et de perlite.

Chaque étape de ce processus a été cruciale pour assurer la qualité et la fiabilité de notre modèle, permettant une classification précise malgré les défis posés par la diversité et l'inconsistance des données initiales extraites d'Internet.

Entraînement

Au cœur de ce projet se trouve un processus d'entraînement rigoureux qui a consisté à évaluer une diversité d'architectures de réseaux de neurones convolutifs et à exploiter la puissance des modèles pré-entraînés. L'ajustement minutieux des dernières couches a été spécifiquement aligné avec les particularités de nos données. Un seuil de confiance défini a été appliqué pour filtrer le dataset, ce qui a permis d'éliminer les images douteuses et d'augmenter la fiabilité de notre entraînement. Nos expériences se sont appuyées sur des frameworks tels que TensorFlow et PyTorch, intégrant des modèles renommés comme ImageNet et YOLO pour tester leur efficacité.

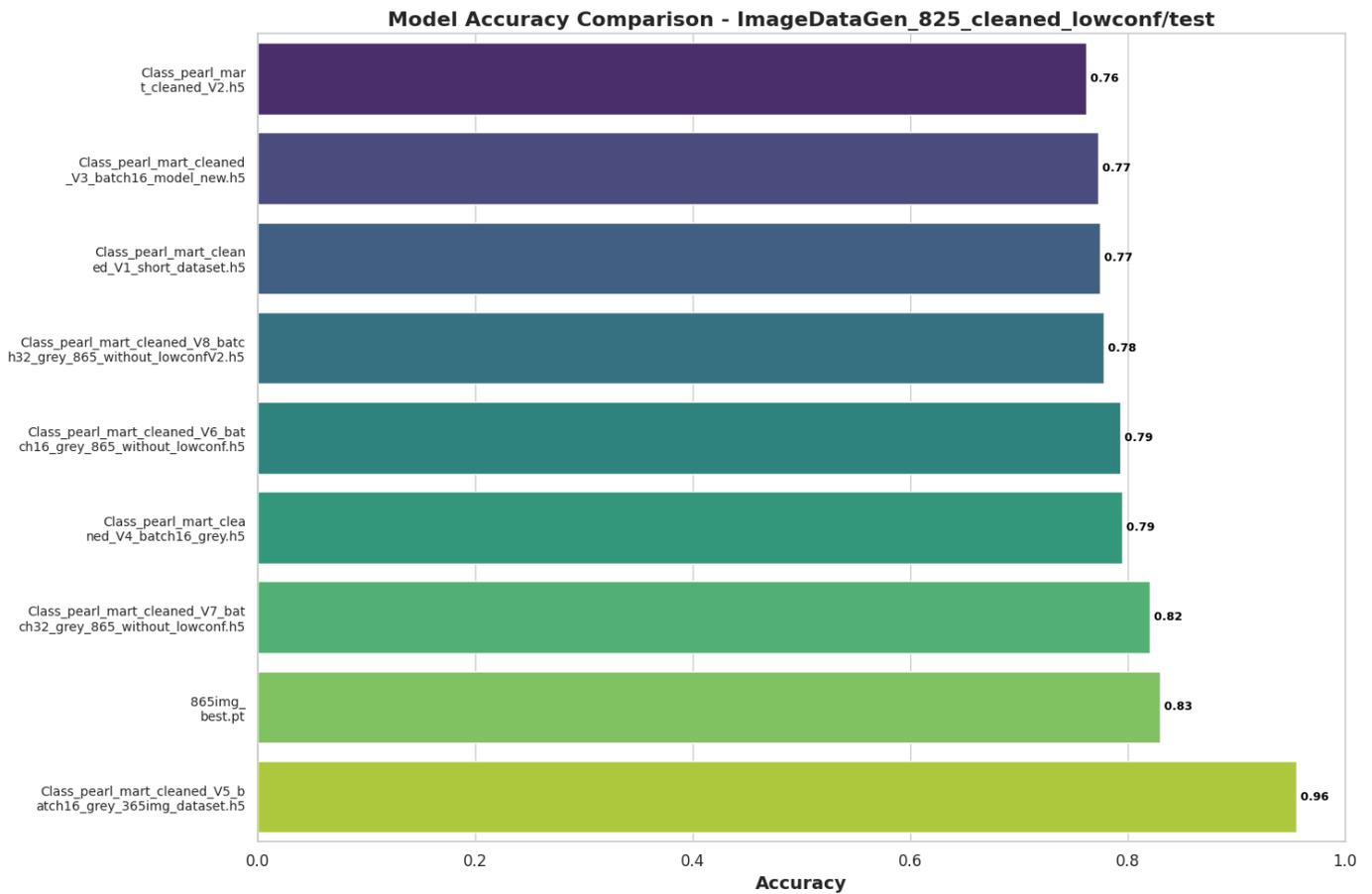
Les courbes d'apprentissage reflètent une convergence stable et prometteuse, suggérant que le modèle s'améliore de manière cohérente et fiable au fil du temps. La constance entre la perte d'entraînement et la précision de validation indique que le modèle n'est ni sous-ajusté ni surajusté



Résultats

Les résultats initiaux révèlent une augmentation significative de l'accuracy avec chaque itération, prouvant l'efficacité de notre approche méthodique. Les graphiques comparatifs offrent une visualisation concrète de l'évolution et de l'adaptabilité de nos modèles sur différents datasets. En particulier, les images comparatives mettent en exergue la supériorité des versions les plus récentes de nos modèles.

Ces avancées attestent de la faculté remarquable du deep learning à interpréter correctement et efficacement des données qui, bien que hétérogènes et issues de sources variées, peuvent être transformées en insights précis et fiables. Cela démontre le potentiel du deep learning à s'adapter et à exceller même lorsque les données initiales présentent des défis en termes de qualité et de cohérence.



Résumé du Projet de Classification des Microstructures Martensitiques et Perlites

Ce projet vise à classer précisément les microstructures martensitiques et perlites en exploitant des données accessibles sur Internet. En naviguant à travers les défis inhérents à la variabilité des données en ligne, ce travail exploite des réseaux de neurones convolutifs et des modèles pré-entraînés finement ajustés pour reconnaître et classer les deux structures. La méthodologie rigoureuse comprend le scraping des images, l'upscale, le cropping automatique, et une sélection méticuleuse des données pour l'entraînement. L'approche, testée sur des frameworks solides tels que TensorFlow et PyTorch, a donné des résultats encourageants, améliorant progressivement l'accuracy et prouvant la capacité du deep learning à fournir des insights significatifs à partir de données imparfaites. Ce travail illustre l'évolution et la flexibilité des modèles de deep learning face à la diversité des données, offrant ainsi de nouvelles perspectives pour les applications de vision par ordinateur dans la métallurgie sans avoir nécessairement énormément de data.