



Image Processing of Multispectral Data Project

A Study on Indian Pines And Pavia University

Team :

Balti Mohamed Aziz
Bzeouich Naoures
Cherif Ghassen
Silini Ahmed
Mejdi Omar

année universitaire:

2024 / 2025



1. Introduction:

1.1 Qu'est-ce que les images multi spectrales?

Les images multispectrales capturent des données à travers différentes longueurs d'onde de lumière, au-delà du spectre visible (rouge, vert, bleu). Cela permet de mieux comprendre les surfaces et les matériaux, offrant des informations précieuses dans des domaines tels que l'agriculture, la surveillance environnementale et l'analyse urbaine.

1.2 Jeux de données utilisés :

Dans ce projet, nous allons travailler avec deux images multispectrales bien connus, chacun accompagné de sa vérité terrain (ground truth) :

- **Indian Pines** : Cette scène couvre principalement des terres agricoles avec quelques forêts et structures humaines. C'est un excellent exemple pour analyser l'utilisation des terres dans les zones rurales et semi-rurales. La vérité terrain pour ce jeu de données nous permet d'identifier les classes d'occupation du sol associées à chaque pixel de l'image.
- **Pavia University** : Ce jeu de données représente un environnement urbain avec des routes, des bâtiments et des espaces verts, idéal pour étudier la couverture des sols en milieu urbain. De même, la vérité terrain associée nous permet de classifier les différents types de couverture du sol dans cette scène urbaine.

1.3 Objectifs du projet:

L'objectif de ce projet est d'explorer les images multispectrales en appliquant diverses techniques telles que la segmentation, les filtres spatiaux et l'analyse des textures. À travers ce projet, nous cherchons à comprendre comment les données multispectrales peuvent être utilisées pour classifier les types d'occupation du sol, extraire des informations spécifiques des images et améliorer la précision des analyses.



2. Exploration et PCA sur les données hyperspectrales:

Objectif:

Dans cette session, nous nous concentrerons sur l'exploration des images hyperspectraux Pavia University et Indian Pines

Étapes :

- **Explorer les données hyperspectrales:** Charger et décrire les images Pavia University et Indian Pines.
- **Visualisation des données:** Visualiser les images pour comprendre leur structure et en tirer des informations.
- **Appliquer l'Analyse en Composantes Principales (ACP) :** Effectuer une ACP pour réduire la dimensionnalité des images .
- **Analysier les résultats de l'ACP :** Interpréter les résultats de l'ACP pour identifier les principaux motifs et caractéristiques.

2.1 Explorer les données hyperspectrales:

Dans cette section, nous explorons deux images hyper spectraux : Indian Pines et Pavia University.

2.1.1 Analyse statistique des images

Nous commencerons par examiner les propriétés de base des images, telles que leurs dimensions, types de données et résumés statistiques comme les valeurs minimales, maximales, moyennes et les écarts-types. Cela nous donnera une compréhension des caractéristiques des données et de leur variabilité.

En utilisant Python, nous effectuons cette analyse en appliquant une fonction pour afficher les statistiques des images hyperspectrales et de leurs étiquettes de vérité terrain correspondantes.

Cette fonction fournit des informations importantes, ce qui nous aide à évaluer la qualité et le format des données.



INSTITUT NATIONAL DES SCIENCES APPLIQUÉES ET DE TECHNOLOGIE

Jeu de données	Dimensions	Type de données	Valeur min	Valeur max	Moyenne	Écart-type
indian_pines	145×145×220	uint16	955	9604	2503,44	1589,63
indian_pines_gt	145×145	uint8	0	16	4,22	5,28
paviaU	610×340×103	uint16	0	8000	1389,13	897,66
paviaU_gt	610×340	uint8	0	9	0,71	1,75

Interprétation

Indian Pines :

- Ce jeu de données contient **220 bandes spectrales**, ce qui signifie qu'il y a beaucoup d'informations par pixel,. Cela implique également une **haute dimensionnalité**, qu'il faudra traiter plus tard via une **ACP** comme une des techniques de réduction de dimension.
- L'**étendue des valeurs** (de 955 à 9604) avec une **valeur moyenne** d'environ **2503.44** et un **écart-type** de **1589.63** montrent une **grande variation des intensités**, ce qui peut refléter la diversité des types de sol et des matériaux présents dans l'image.

Jeu de données Pavia University :

- Le jeu de données Pavia contient **moins de bandes spectrales** (103), avec une **moyenne** de **1389.13** et un **écart-type** de **897.66**, mais reste riche en informations. La **valeur moyenne des pixels** est plus faible que celle d'Indian Pines, ce qui suggère des caractéristiques de données différentes.
- La **variabilité des valeurs de pixels** (de 0 à 8000) peut indiquer la présence de matériaux variés, tels que des zones urbaines, des plans d'eau, de la végétation, etc.

Vérités terrain (Ground Truth) :

- Le **GT de Indian Pines** comprend des valeurs allant de **0 à 16**, représentant **différentes classes d'occupation du sol**.
- Le **GT de Pavia University** est plus restreint avec des valeurs de **0 à 9**, ce qui indique un **nombre plus limité de catégories**.



Conclusion

- Ces premières informations statistiques nous donnent un **aperçu général des images**, en termes de structure, de distribution des valeurs de pixels, et de diversité des classes. **Elles sont utiles pour comprendre la richesse des données et anticiper les traitements à appliquer.**
- Cependant, **elles ne suffisent pas à elles seules pour interpréter pleinement le contenu spectral** de chaque image. En effet, nous ne pouvons pas simplement deviner quelles longueurs d'onde correspondent à quels matériaux ou types d'occupation du sol en se basant uniquement sur des statistiques comme la moyenne ou l'écart-type.
- Ainsi, pour **mieux comprendre la signification des bandes spectrales, la nature du lieu observé, ainsi que la correspondance des vérités terrain (ground truth)** avec les différentes classes, **nous allons devoir nous appuyer sur des ressources externes** : notamment les **descriptions techniques des images** (fournies dans les manuels officiels), qui détaillent les longueurs d'onde mesurées, le contexte géographique, et les annotations des classes. Cela nous permettra d'interpréter plus justement les résultats des futures analyses.

2.1.2 Complément d'analyse par la documentation :

Indian Pines

Nous avons consulté une ressource externe intitulée “**Spectral region identification versus individual channel selection in supervised dimensionality reduction of hyperspectral image data**”, qui fournit des détails précieux sur la scène Indian Pines.

Ce document précise que l'image Indian Pines a été capturée par le capteur **AVIRIS** au-dessus du site de test situé au nord-ouest de l'Indiana. Elle contient **145×145 pixels et 220 bandes spectrales** réparties sur la plage de longueurs d'onde **de 0.4 μm à 2.5 μm (micromètres)**, soit **400 à 2500 nanomètres**. Cette plage couvre la lumière visible ainsi que le proche infrarouge et l'infrarouge à ondes courtes.

Pour relier chaque bande à sa longueur d'onde approximative, on suppose une répartition **uniforme des longueurs d'onde** entre la première et la dernière bande, ce qui permet de faire une estimation par la formule suivante :

$$\lambda_i = \lambda_{min} + \left(\frac{i-1}{N-1} \right) (\lambda_{max} - \lambda_{min})$$



INSTITUT NATIONAL DES SCIENCES APPLIQUÉES ET DE TECHNOLOGIE

Inversement: pour retrouver le **numéro de bande i** à partir d'une **longueur d'onde** donnée, Voici la formule inverse

$$i = 1 + \left(\frac{\lambda - \lambda_{min}}{\lambda_{max} - \lambda_{min}} \right) \cdot (N - 1)$$

Avec

- λ : la longueur d'onde recherchée (en nanomètres)
- λ_i : est la longueur d'onde de la bande i
- $\lambda_{min} = 400 \text{ nm}$
- $\lambda_{max} = 2500 \text{ nm}$
- $N=220$ (nombre total de bandes)

Cela nous permet d'estimer qu'une bande comme la **bande 112** se situerait environ au centre du spectre à **1450 nm**, ce qui correspond à une zone où l'eau absorbe fortement. En effet, le document recommande de **supprimer les bandes 104 à 108, 150 à 163 et la bande 220** en raison de leur faible utilité due à **l'absorption de l'eau** dans ces régions spectrales.

Qu'est-ce que l'absorption de l'eau ?

Lorsque la lumière traverse ou est réfléchie par une surface contenant de l'eau (sol humide, végétation, atmosphère, etc.), certaines longueurs d'onde sont **fortement atténuées** car elles sont absorbées par les molécules d'eau. Ces longueurs d'onde correspondent à des "**bandes d'absorption**" naturelles.

Pavia University

Nous avons également consulté une ressource externe intitulée "**Locality Preserving Composite Kernel Feature Extraction for Multi-Source Geospatial Image Analysis**" pour Pavia University. Cette scène, acquise par le capteur **ROSIS** au-dessus de la ville de Pavia, contient 610×340 pixels et 103 bandes spectrales, avec une plage de longueurs d'onde allant de 430 nm à 860 nm. On peut utiliser la formule d'estimation des longueurs d'onde (mentionnée précédemment dans l'analyse d'Indian Pines) pour associer chaque bande à sa longueur d'onde correspondante avec la seule différence

- $\lambda_{min} = 430 \text{ nm}$
- $\lambda_{max} = 860 \text{ nm}$
- $N = 103$ (nombre total de bandes)



2.2 Visualisation des données:

Dans cette partie, nous traiterons séparément chaque image avec leurs vérités terrains respectifs. Pour chacun, nous détaillerons la méthode de visualisation en niveaux de gris et en fausses couleurs (RGB), ainsi que les étapes de normalisation des valeurs de pixels.

2.2.1 Normalisation des valeurs

Il est essentiel de normaliser les valeurs des pixels afin de les ramener dans une plage adaptée à l'affichage et à l'entraînement de modèles de deep learning. En effet, les valeurs initiales peuvent atteindre jusqu'à **9604 pour Indian Pines** et **8000 pour Pavia University**, ce qui dépasse largement l'échelle standard des images (généralement sur 8 bits, soit de 0 à 255). Pour cela, nous avons utilisé une fonction Python nommée `normalize_band(band)`, qui applique une normalisation linéaire à chaque bande selon la formule :

$$band_{norm} = \left(\frac{band}{band_{max}} \right) \cdot 255$$

Cette transformation permet de réduire dynamiquement l'échelle de chaque bande tout en conservant les proportions relatives des intensités. Cette étape est indispensable pour une visualisation correcte en niveaux de gris ou en RGB, ainsi que pour assurer une convergence stable lors de l'entraînement des réseaux de neurones.

2.2.2 Visualisation initiale en niveaux de gris :

Pour cela, on sélectionne une seule bande représentative, que l'on affiche comme une image 2D.

Indian Pines

Dans notre cas, nous avons choisi une bande située dans le **spectre bleu** du visible (environ 450 nm - bande 6), une zone généralement **peu affectée par l'absorption d'eau** et offrant une **bonne stabilité spectrale**.

Image en niveaux de gris (bande 6)



Indian Pines

Malgré la dominance de zones agricoles et végétales dans Indian Pines, cette bande fournit une **visualisation nette et contrastée**, en particulier des contours entre les différentes parcelles.



INSTITUT NATIONAL DES SCIENCES APPLIQUÉES ET DE TECHNOLOGIE

Pavia University

Pour le jeu de données **Pavia University**, qui représente un environnement **urbain** avec des bâtiments, routes, et espaces verts, un bon choix de bande pour une visualisation en niveaux de gris est la **bande 29** (environ 550 nm, soit le **vert** du spectre visible).

Elle est **au cœur du spectre visible**, dans une zone **relativement stable et moins bruitée**.

Image en niveaux de gris (bande 29)



Dans les environnements urbains, cette bande offre **un bon contraste visuel** entre les surfaces artificielles et naturelles.

2.2.2 Visualiser une fausse image RGB : **Indian Pines**

Une fois les bandes hyperspectrales normalisées, on peut créer une image RGB simulée pour la scène **Indian Pines** en choisissant les bandes correspondantes au spectre visible :

- **Plage spectrale de l'image** : $0.4 \text{ } \mu\text{m}$ à $2.5 \text{ } \mu\text{m}$
- **Nombre total de bandes** : 220
- **Largeur spectrale approximative par bande** : $\frac{2.5 - 0.4}{220} \cong 0.0095 \text{ } \mu\text{m}$

Pour trouver les bandes correspondant aux longueurs d'onde classiques du spectre visible : **Bleu (~0.45 μm)** :

$$\frac{0.45 - 0.4}{0.0095} \approx 5 \Rightarrow \text{bande } 5$$

Vert (~0.55 μm):

$$\frac{0.55 - 0.4}{0.0095} \approx 16 \Rightarrow \text{bande } 16$$

Rouge (~0.65 μm):

$$\frac{0.65 - 0.4}{0.0095} \approx 27 \Rightarrow \text{bande } 27$$

Image RGB simulée (bandes 27-16-5)



(a) image RGB simulée(bande 27-16-5)



(b) image réelle prise de la documentation

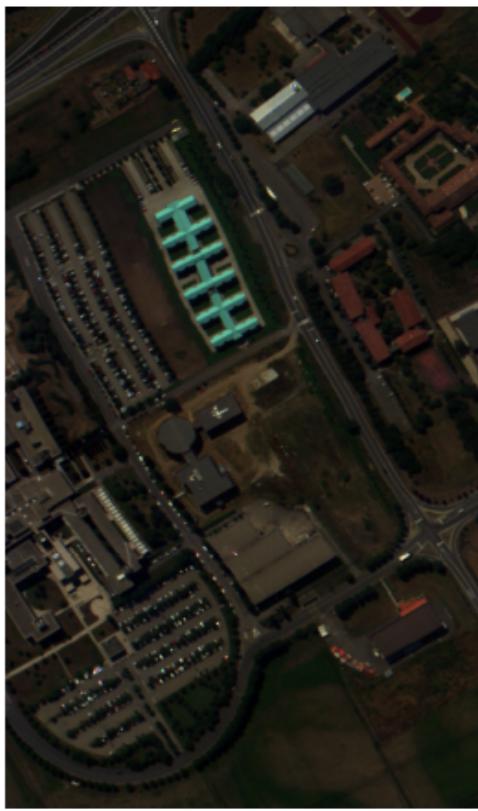
Pavia University

L'image comporte **103 bandes** réparties de manière uniforme entre **430 nm et 860 nm**, soit :

- **Largeur de bande** $\approx (860 - 430) / (103 - 1) \approx 4.215 \text{ nm}$

Correspondance des longueurs d'onde avec les bandes :

- **Bleu ($\sim 450 \text{ nm}$)** $\rightarrow (450 - 430) / 4.215 \approx \text{Bande } 5$
- **Vert ($\sim 550 \text{ nm}$)** $\rightarrow (550 - 430) / 4.215 \approx \text{Bande } 28$
- **Rouge ($\sim 650 \text{ nm}$)** $\rightarrow (650 - 430) / 4.215 \approx \text{Bande } 52$



(a) image RGB simulée(bande 52-28-5)



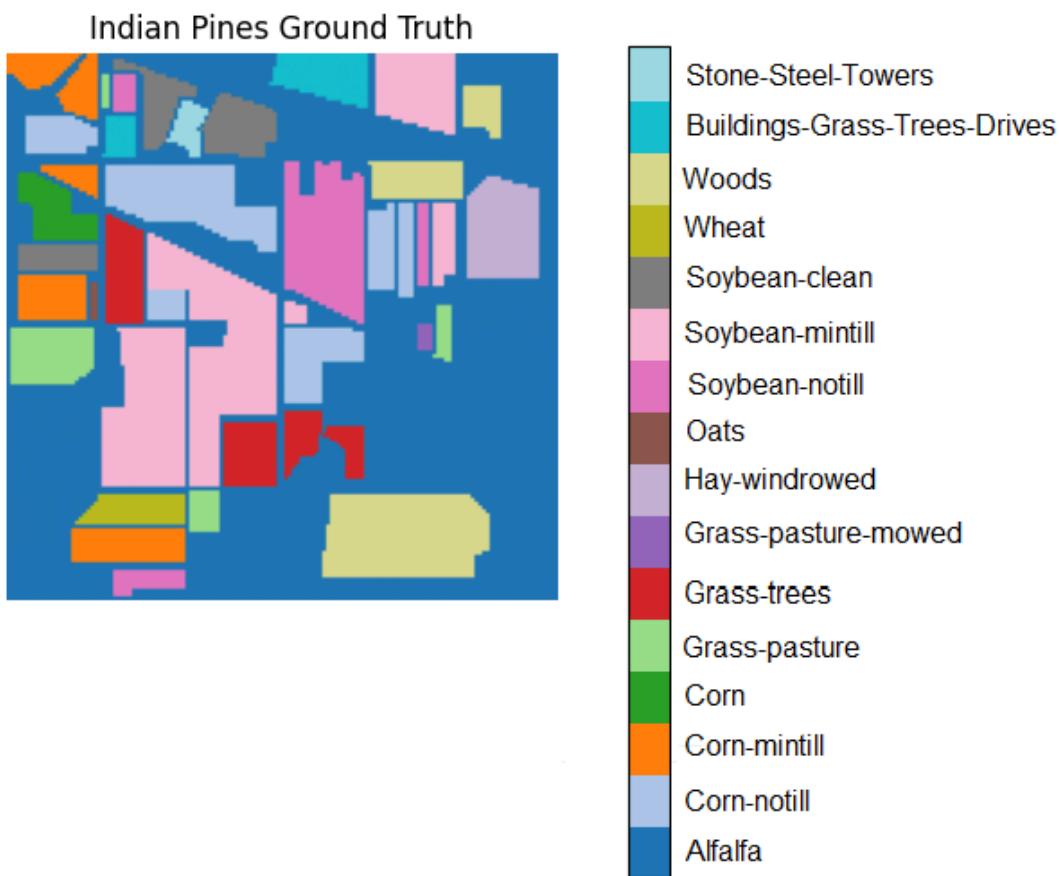
(b) image réelle prise de la documentation

2.2.3 Visualisation de Ground Truth :

La vérité terrain est essentielle dans les approches supervisées, car elle fournit des annotations fiables pour chaque pixel, permettant d'entraîner et d'évaluer des modèles de classification. Elle sert de référence pour analyser la répartition spatiale des différentes classes, mieux comprendre les zones dominantes dans une scène, et interpréter les résultats en les comparant à une réalité connue.

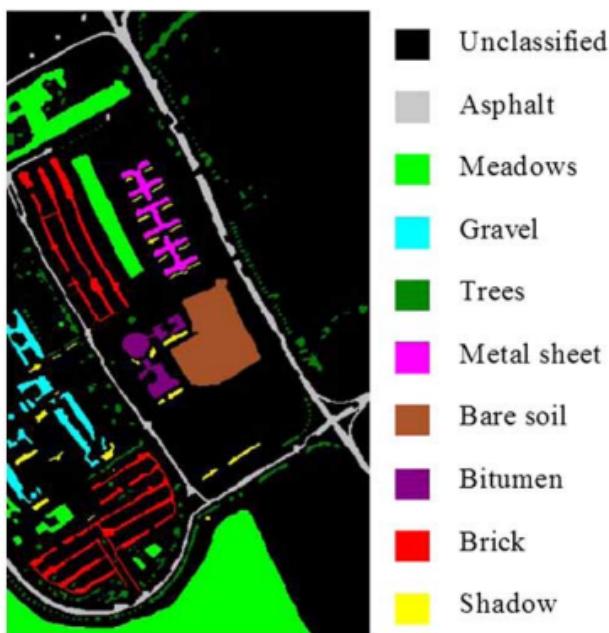
Indian Pines

Dans le cas du jeu de données **Indian Pines**, la vérité terrain révèle une scène majoritairement agricole, composée d'une grande variété de cultures telles que le maïs, le soja, le blé ou encore la luzerne. On y trouve aussi quelques zones de végétation naturelle (bois, herbes) ainsi que de rares surfaces artificielles, comme des routes ou des structures métalliques. Les classes sont assez déséquilibrées, certaines cultures étant bien plus représentées que d'autres.



Pavia University

Pour **Pavia University**, on est face à une scène urbaine dense. La vérité terrain distingue des surfaces artificielles variées comme l'asphalte, le gravier, les briques ou les toitures métalliques, ainsi que des éléments naturels comme les prairies et les arbres. Malgré la dominance du bâti, quelques zones de verdure et d'ombre sont également annotées. L'ensemble offre une grande diversité spectrale typique d'un environnement urbain mixte.



2.2.4 Visualisation ciblée de caractéristiques spectrales spécifiques :

Il est possible de visualiser des caractéristiques spécifiques en se concentrant sur certaines longueurs d'onde. L'idée est d'extraire des images en niveaux de gris, chacune représentative d'un type de matériau ou de propriété, en sélectionnant une bande unique liée à un comportement spectral connu.

Indian Pines

Dans cette approche, nous nous focalisons sur trois grandes familles de caractéristiques naturelles : la **végétation et l'eau**. Pour chacune, une bande a été sélectionnée selon sa position dans le spectre, permettant une interprétation visuelle rapide des zones dominantes dans la scène.

♦ Végétation

La végétation réfléchit fortement dans le **proche infrarouge (NIR)**. Pour **Indian Pines**, cela correspond à la **bande 65** (≈ 1000 nm), bien située dans la plage sensible au feuillage. L'image en niveaux de gris de cette bande devrait faire ressortir les zones agricoles (forte intensité) par opposition aux zones bâties ou non végétalisées (faible intensité).

Image en niveaux de gris (bande 65)



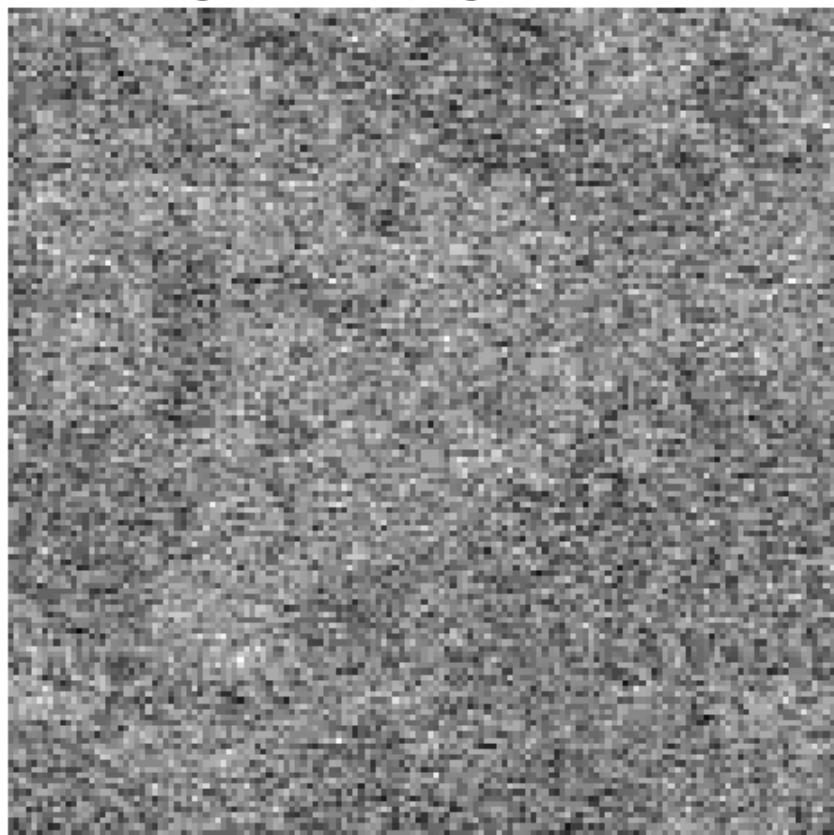
Interprétation et explication:

Lors de la visualisation de la végétation à l'aide d'une bande située dans la région du proche infrarouge (NIR), on remarque que **la majeure partie de l'image est lumineuse**. Cela s'explique par le fait que la scène Indian Pines est essentiellement **agricole**, donc largement couverte par une végétation saine, qui réfléchit fortement les longueurs d'onde du NIR. À l'inverse, les zones **non végétalisées** comme les **routes, les bâtiments ou autres surfaces artificielles** apparaissent **sombres**, car elles réfléchissent beaucoup moins dans cette région spectrale. Ces zones sombres sont peu nombreuses et de petite taille, ce qui confirme la prédominance de la végétation dans la scène.

♦ Absorption d'Eau

Les bandes autour de **1410 nm** sont connues pour être fortement absorbées par l'eau. D'après les recommandations documentées, la **bande 106** est idéale pour visualiser cette absorption..

Image en niveaux de gris (bande 106)



Interprétation et explication:

Lors de la visualisation des bandes situées dans les zones d'absorption de l'eau (notamment autour de 1350–1450 nm et 1800–1950 nm), l'image obtenue s'est révélée fortement bruitée et pixelisée. Visuellement, elle présentait une texture granuleuse avec peu de contrastes, rendant difficile toute interprétation spatiale ou spectrale. Cette apparence dégradée s'explique par le fait que dans ces plages de longueurs d'onde, la vapeur d'eau atmosphérique absorbe une grande partie du rayonnement, laissant très peu de signal utile atteindre le capteur. Le signal mesuré étant donc très faible, il est rapidement noyé dans le bruit, ce qui produit ces effets visuels de dégradation que nous avons observés.

Pavia University

Pour l'image Pavia University, les possibilités d'extraction de caractéristiques spectrales spécifiques sont limitées, car la majorité des 103 bandes disponibles sont concentrées dans la plage du spectre visible, entre 430 nm et 860 nm. Cette répartition restreint l'accès aux informations situées dans les longueurs d'onde plus éloignées, comme les régions SWIR, utiles pour l'analyse de l'eau ou des minéraux. Toutefois, la bande la plus longue (~860 nm) se situe dans le proche infrarouge (NIR), ce qui permet de distinguer les zones végétalisées.



(a) image gris simulée(bande 102)



(b) image réelle prise de la documentation

Interprétation et explication:

Étant donné que la scène est majoritairement urbaine, l'image en niveaux de gris correspondant à cette bande présente une luminosité globale moyenne, avec seulement quelques zones très brillantes. Ces zones correspondent vraisemblablement aux rares espaces verts présents dans l'image, comme les arbres ou petites pelouses, qui réfléchissent fortement dans le NIR, contrairement aux structures urbaines.



2.3 Appliquer l'Analyse en Composantes Principales (ACP):

2.3.1 Pourquoi ACP?

L'analyse en composantes principales (ACP) est une méthode de réduction de dimension qui transforme un jeu de données d'origine en un nouvel espace de variables (composantes principales), tout en conservant un maximum de variance.

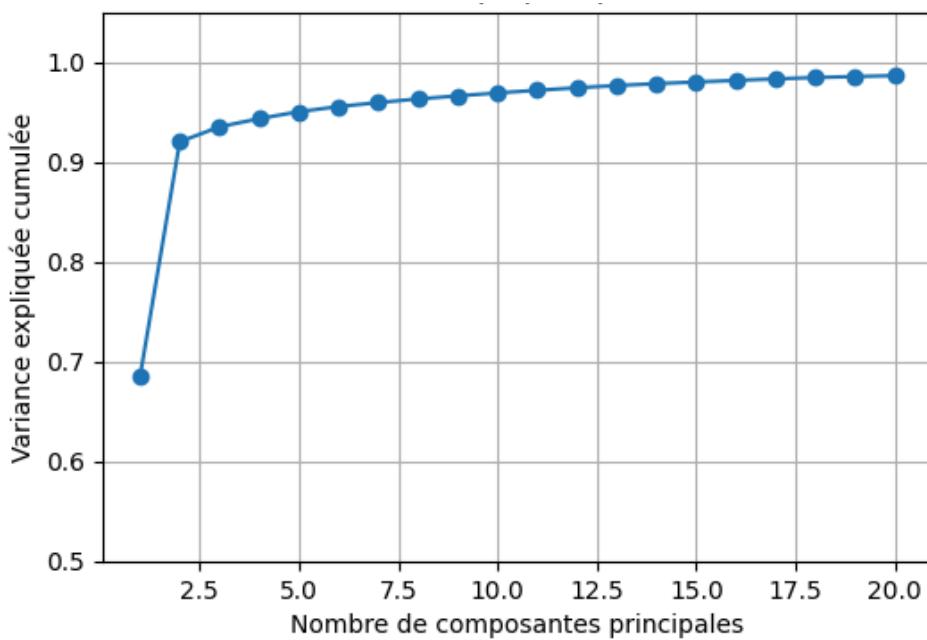
Son objectif est de réduire la redondance spectrale en compressant l'information contenue dans des centaines de bandes en un nombre limité de composantes significatives. Le critère principal pour choisir le nombre de composantes conservées repose généralement sur la proportion de variance expliquée : on cherche à capturer par exemple 98 % de la variance avec le moins de composantes possible.

2.3.2 Application du PCA sur les données:

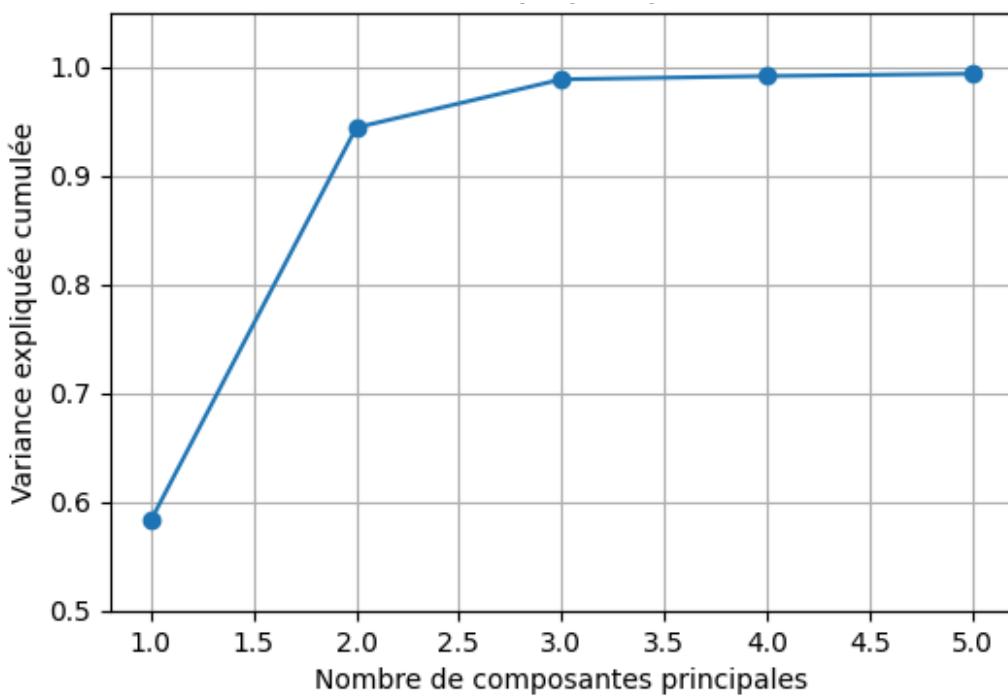
Dans notre cas, nous avons implémenté cette réduction à l'aide d'une fonction Python qui applique le PCA sur les données hyperspectrales de chaque image.

Après normalisation des pixels, la transformation est appliquée, et on choisit d'arrêter l'analyse lorsqu'on atteint une fraction de variance cumulée jugée 98%. Cela permet d'alléger les traitements à venir sans trop perdre d'information pertinente.

Dataset	#Composants	Variance Cumulative	Main Explained Variance Ratios (%)	Top Eigenvalues
Indian Pines	20	~98.65%	68.49%, 23.53%, 1.50%, 0.82%, 0.70%	26.8M, 9.2M, 585K, 321K, 271K
PaviaU	5	~99.37%	58.32%, 36.10%, 4.44%, 0.30%, 0.21%	31.3M, 19.4M, 2.38M, 161K, 112K



Variance expliquée par l'ACP d'Indian Pines



Variance expliquée par l'ACP de Pavia University



2.3.3 Analyser les résultats de l'ACP:

À partir des résultats obtenus via le script Python, voici l'interprétation des statistiques de l'Analyse en Composantes Principales (ACP) appliquée aux images hyperspectrales **Indian Pines** et **Pavia University** :

Indian Pines

Dans ce cas, la première composante principale capture à elle seule **68,49 %** de la variance totale. En ajoutant successivement les quatre suivantes, la variance cumulée atteint **98,01%** à la 15ème composante. Cela justifie le choix de s'arrêter à **15 composantes principales**, car **au-delà, le gain en information devient marginal**. On observe donc que **la majorité de la variabilité est concentrée dans un sous-espace réduit**, ce qui permet de réduire considérablement la dimension tout en conservant une très bonne représentation des données d'origine.

Pavia University

Pour l'image de Pavia, la première composante explique environ **58,32 %** de la variance, les deux premières en cumulent **94,41 %**, et les trois premières atteignent presque **98,86 %**. Cela traduit une décroissance plus progressive que pour Indian Pines, mais on remarque tout de même que **la majorité de l'information est concentrée dans les premières composantes**. Les quatrième et cinquième composantes n'ajoutent que très peu de variance (moins de 0,4 % chacune), confirmant l'intérêt d'une réduction de dimension en 3 dimensions.

2.3.2 Visualisation des composants du PCA en gris:

La visualisation des composantes principales en niveaux de gris permet d'observer **comment chaque composante résume une partie de l'information spectrale** de l'image. La **première composante** met généralement en évidence les **variations globales les plus marquées**, tandis que les composantes suivantes révèlent des **détails plus fins et localisés**. À mesure que l'on avance dans les composantes, l'intensité des détails diminue, ce qui reflète leur contribution plus faible à la variance totale de l'image hyperspectrale.

Indian Pines

Composante principale 1



Composante principale 2



Composante principale 3



Composante principale 4



Composante principale 5



Pavia University

Composante principale 1



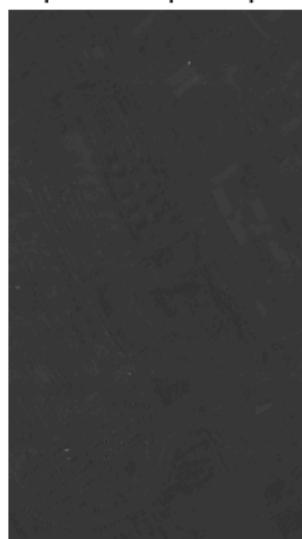
Composante principale 2



Composante principale 3



Composante principale 4



Composante principale 5



Ces observations confirment pleinement ce qui a été mentionné précédemment : **en visualisant directement les images des composantes principales, on retrouve bien les effets attendus du PCA.** La première composante restitue les structures dominantes avec clarté, tandis que les suivantes perdent progressivement en contenu informatif, illustrant visuellement la répartition décroissante de la variance capturée.



3. classification des images par un modèle de deep learning

Objectif:

Dans cette session, nous nous concentrerons sur les modèles de deep learning utilisés pour la classification des images.

Étapes :

- **Différents modèles d'apprentissage :** exploration des principaux modèles de deep learning pour la classification d'images, ainsi que les différences entre eux.
- **Modèle CNN appliqué à l'image Indian Pines :** mise en œuvre d'un CNN pour classifier les données hyperspectrales de la scène Indian Pines.
- **Modèle CNN pour l'image Pavia University :** mise en œuvre d'un CNN pour classifier les données hyperspectrales de la scène Pavia University.
- **Impact de l'exclusion du fond (background) sur la segmentation :** évaluer si le modèle entraîné sur Indian Pines sans fond détecte mieux les zones vertes dans Pavia University, grâce à la nature végétale de la scène d'origine.

3.1 Différents Modèles d'Apprentissage et Pourquoi Choisir les CNN pour la Classification d'Images

3.1.1 Différents Modèles d'Apprentissage

Le traitement des images hyperspectrales et multispectrales, comme celles de Pavia University ou Indian Pines, nécessite des approches capables de capturer à la fois les caractéristiques spatiales (textures, formes) et spectrales (signatures spectrales des matériaux). Voici un aperçu des principaux modèles d'apprentissage profond utilisés :

CNN (Convolutional Neural Networks) :

Les CNN sont les modèles les plus utilisés pour le traitement d'images. Ils extraient efficacement les caractéristiques **spatiales locales** grâce aux filtres convolutifs.

Exemples : AlexNet, VGGNet, ResNet, Inception.

Avantages : Bonne performance pour les images 2D, peu de prétraitement requis.



INSTITUT NATIONAL DES SCIENCES APPLIQUÉES ET DE TECHNOLOGIE

Applications en télédétection : Classification des cultures, détection des changements, segmentation d'objets.

Limite : Ne capturent pas bien les dépendances globales ni les informations spectrales par défaut.

3D-CNN (Convolution 3D) :

Ces modèles étendent les CNN classiques pour capturer à la fois les **dimensions spatiales et spectrales** des images hyperspectrales.

Avantages : Très efficaces pour les cubes hyperspectraux.

Applications : Classification pixel-par-pixel avec exploitation du cube 3D (largeur × hauteur × bandes spectrales).

Transfer Learning :

Technique consistant à réutiliser un modèle pré-entraîné sur une grande base d'images (ImageNet, etc.) puis à l'adapter à une tâche spécifique comme la classification d'images satellitaires.

Modèles utilisés : VGG16, ResNet50, MobileNet, EfficientNet.

Avantage : Moins de données nécessaires pour entraîner le modèle.

Limite : Les modèles sont souvent pré-entraînés sur des images RGB naturelles, pas sur des données hyperspectrales.

Vision Transformers (ViT) :

Ces modèles utilisent des **mécanismes d'attention** pour capturer les relations globales dans une image.

Avantage : Capturent les dépendances spatiales longues.

Limite : Nécessitent beaucoup de données pour être entraînés correctement.

Applications : De plus en plus adoptés pour les images satellitaires récentes, y compris la segmentation de scènes complexes.

Recurrent Neural Networks (RNN) et variantes (LSTM, GRU) :

Utilisés surtout pour modéliser des **séries temporelles** dans les images satellites multi-temporelles.

Applications : Suivi des cultures agricoles au fil du temps, détection du changement climatique.

LSTM-CNN hybrids : Permettent d'exploiter à la fois les informations spatiales et temporelles.

DenseNet :

Ce modèle connecte chaque couche à toutes les couches suivantes, favorisant une meilleure propagation de l'information.



Avantage : Moins de perte d'information, réduit le sur-apprentissage.

Utilisé dans : Tâches de classification et segmentation en télédétection.

3.1.2 Pourquoi Choisir les CNN pour la Classification d'Images

Les images satellitaires – en particulier les images hyperspectrales – sont très riches en informations. Chaque pixel ne contient pas seulement une couleur RGB, mais aussi des données spectrales complexes réparties sur des dizaines (Indian Pines: 220 et Paia University: 103) . Ce volume d'information rend les méthodes classiques de classification (comme les arbres de décision ou les SVM) rapidement limitées, car elles ne peuvent pas exploiter toute la richesse spatiale et spectrale disponible.

C'est dans ce contexte que les réseaux de neurones convolutifs (**CNN**) prennent tout leur sens. Ces modèles sont conçus pour analyser automatiquement les motifs dans les images : ils sont capables d'extraire des caractéristiques locales (formes, textures, contrastes) tout en tenant compte des relations spatiales entre les pixels. Cela les rend particulièrement adaptés à la reconnaissance de structures dans les images satellites, comme des champs agricoles, des routes ou des zones urbaines.

Les CNN offrent également une grande **flexibilité** : ils peuvent être adaptés en 1D, 2D ou 3D pour traiter des données purement spectrales, purement spatiales, ou une combinaison des deux – ce qui est idéal dans le cadre des images hyperspectrales. De plus, leur architecture permet d'ajouter des couches selon la complexité de la tâche, et d'optimiser l'apprentissage grâce à la rétropropagation.

Un autre avantage majeur est leur **capacité à apprendre automatiquement les bonnes caractéristiques**, sans avoir besoin de programmer manuellement des règles d'extraction. Cela permet un gain de temps important tout en obtenant de meilleures performances.

Enfin, l'**efficacité** des réseaux de neurones convolutifs (CNN) a été largement démontrée à travers des jeux de données de référence tels que Indian Pines et Pavia University. Dans ces contextes, les CNN surpassent régulièrement les méthodes traditionnelles comme les SVM ou les classificateurs basés sur des règles manuelles.



À titre d'exemple, une étude menée en 2018 a montré qu'une architecture CNN avancée pouvait atteindre une précision globale (Overall Accuracy) de **99,46 %** sur le jeu de données Indian Pines, contre 86,24 % pour un SVM classique et 97,81 % pour un modèle 3D-CNN. Ces chiffres témoignent clairement de la supériorité des CNN pour la classification d'images hyperspectrales, notamment grâce à leur capacité à exploiter efficacement à la fois l'information spectrale et spatiale.

3.2 model CNN pour l'image Indian Pines :

L'objectif de cette partie est de concevoir, entraîner et évaluer un modèle de réseau de neurones convolutif (CNN) capable de segmenter l'image hyperspectrale Indian Pines en identifiant correctement la classe de chaque pixel.

3.2.1 Prétraitement des données :

Avant d'entraîner le modèle, plusieurs étapes de prétraitement sont réalisées :

- Chargement des données : les données hyperspectrales et les étiquettes sont chargées depuis des fichiers .mat.
- Normalisation : un MinMax Scaler est utilisé pour ramener les valeurs des pixels entre 0 et 1, afin d'améliorer la convergence du modèle.
- Découpage en patches : Pour chaque pixel central, un patch local est extrait afin de capturer le contexte spatial dans son voisinage. La taille de ces patches peut varier (3×3 , 5×5 , 7×7 , 9×9 ou 11×11) en fonction des tests effectués, l'objectif étant de déterminer empiriquement la taille offrant les meilleures performances de classification.
- Réduction des bords : les pixels proches des bords de l'image sont ignorés, car un patch centré sur eux sortirait de l'image.
- Création d'un dataset PyTorch : chaque entrée devient un patch 3D (canaux spectraux \times hauteur \times largeur), associé à une étiquette de classe.

Après l'étape de prétraitement, chaque échantillon de données prend la forme suivante : `torch.Size([1, 1, 200, 7, 7])`

Voici ce que cela signifie :

- 1 : correspond à la taille de **batch** (ici 1, car on affiche un seul échantillon)
- 1 : dimension ajoutée pour compatibilité avec certains modèles 3D (channel dimension),
- 200 : nombre de **canaux spectraux** (dimensions spectrales de l'image hyperspectrale),
- 7 : **hauteur** du patch spatial autour du pixel central,
- 7 : **largeur** du patch spatial autour du pixel central.

3.2.2 Conception :

Le modèle est basé sur un réseau de neurones convolutif hybride, optimisé pour les images hyperspectrales. L'idée principale est d'utiliser à la fois des convolutions 3D (pour capturer les relations spatiales et spectrales) et des convolutions 2D (pour des opérations classiques de CNN sur des cartes de caractéristiques).

Pourquoi ce choix de modèle ?

Le choix d'un modèle hybride combinant des **convolutions 3D et 2D** est motivé par les caractéristiques spécifiques des images hyperspectrales, qui possèdent à la fois une **profondeur spectrale riche** et une **structure spatiale complexe**.

Voici pourquoi ce modèle représente une solution pertinente :

- **Exploitation conjointe de l'information spatiale et spectrale :**
Les convolutions 3D permettent d'extraire efficacement les **corrélations inter-bandes spectrales** tout en tenant compte de la structure spatiale locale.
- **Réduction progressive de la complexité :**
En utilisant d'abord des couches 3D, puis en compressant la dimension spectrale pour passer à des convolutions 2D, le modèle réduit le **nombre de paramètres** et la **charge computationnelle**, tout en conservant les caractéristiques essentielles.
- **Meilleure généralisation :**
L'architecture hiérarchique (3D → 2D → Fully Connected) permet de construire des représentations de plus en plus abstraites et discriminantes, ce qui améliore la **capacité de généralisation** du modèle sur les différentes classes.



- **Adapté aux variations locales :**
Grâce à l'utilisation de patches, le modèle est **invariant aux petites translations** locales et capable de s'adapter à différents contextes spatiaux autour d'un pixel.

3.2.3 Architecture :

Le modèle utilisé pour la segmentation de l'image Indian Pines repose sur une architecture CNN hybride appelée HybridSN (Hybrid Spectral-Spatial Network). Cette architecture est spécifiquement conçue pour exploiter à la fois l'information spectrale et spatiale contenue dans les images hyperspectrales.

Entrée du modèle :

L'entrée du réseau est un patch 3D de taille (1, 220, 7, 7) :
1 correspond au canal (valeur par défaut pour le traitement 3D),
220 représente le nombre total de bandes spectrales dans l'image Indian Pines,
la taille spatiale du patch choisi (3x3 , 5x5 , 7x7 , 9x9 ou 11x11).

Couches du modèle :

L'architecture HybridSN utilisée dans cette implémentation se compose des couches suivantes :

- **Convolution 3D :**
Type : Conv3d
Paramètres : in_channels=1, out_channels=8, kernel_size=(3, 3, 3), padding=(1, 1, 1)
Rôle : Extraire les caractéristiques conjointes spectrales et spatiales en traitant simultanément les trois dimensions.
Sortie : Un tenseur avec 8 cartes de caractéristiques 3D.
- **Batch Normalization 3D :**
Type : BatchNorm3d(8)
Rôle : Normaliser les activations issues de la couche précédente pour accélérer l'apprentissage et stabiliser l'entraînement.
- **Reshape (Réorganisation):**
La sortie du bloc 3D est aplatée et réorganisée pour devenir compatible avec les couches 2D.
Le résultat donne une dimension d'entrée de 1760 canaux pour la convolution 2D suivante.
Cela correspond à : 8 canaux × réduction spatiale/spectrale.

- **Convolution 2D:**

Type : Conv2d

Paramètres : in_channels=1760, out_channels=64, kernel_size=3, padding=1

Rôle : Exploiter davantage la structure spatiale des patches en appliquant des filtres 2D sur chaque plan de caractéristiques.

- **Batch Normalization 2D :**

Type : BatchNorm2d(64)

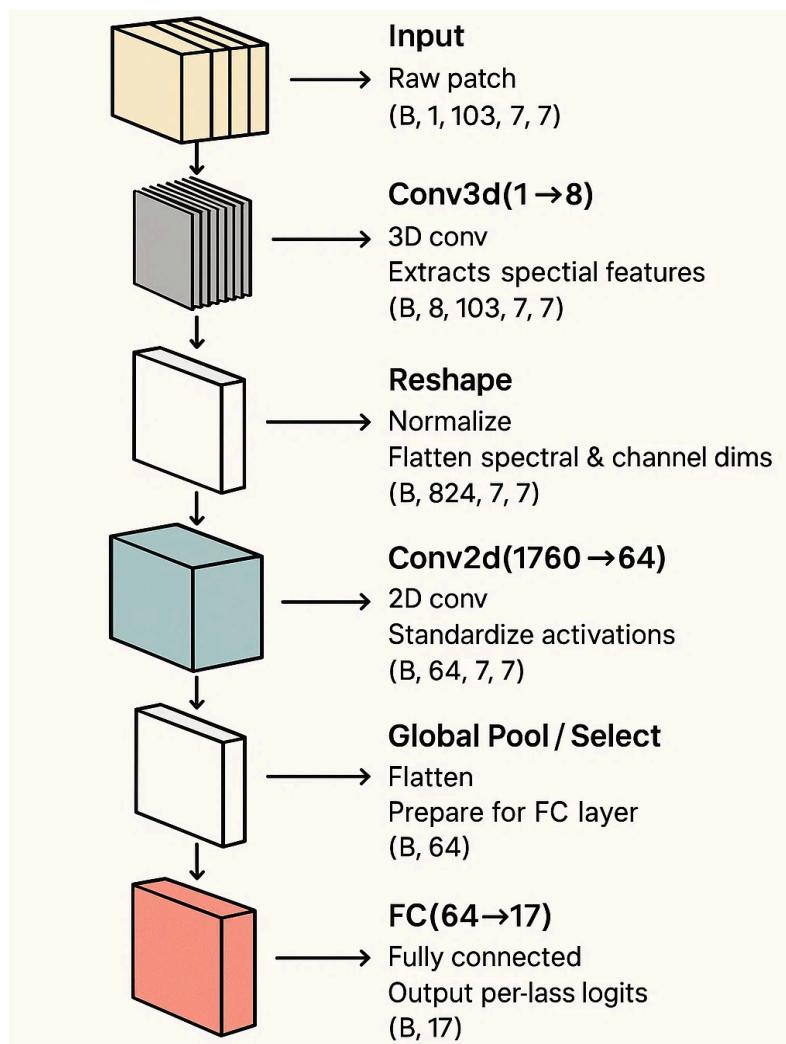
Rôle : Normaliser les cartes de caractéristiques produites par la convolution 2D.

- **Couche Fully Connected (FC) :**

Type : Linear

Paramètres : in_features=64, out_features=17

Rôle : Produire une probabilité d'appartenance à l'une des 17 classes présentes dans l'image Indian Pines (16 classes réelles + 1 classe vide).



3.2.4 Entraînement du modèle :

L'entraînement du modèle de classification est une étape cruciale qui influe directement sur la capacité du réseau à apprendre des représentations discriminantes à partir des images hyperspectrales.

- **Fonction de perte : CrossEntropyLoss pondérée**

Cette fonction est bien adaptée aux tâches de classification multi-classes. La pondération des classes est utilisée pour compenser le déséquilibre important dans les données (certaines classes étant sous-représentées), évitant ainsi que le modèle privilégie les classes dominantes.

- **Optimiseur : Adam**

L'optimiseur Adam est choisi pour sa capacité d'adaptation dynamique des taux d'apprentissage pour chaque paramètre, ce qui favorise une convergence rapide et stable.

- **Learning rate : 0.0006**

choisi empiriquement pour permettre un apprentissage progressif sans divergence.

- **Weight decay : 1e-4,**

pour appliquer une régularisation L2 qui limite les sur-apprentissages (overfitting) en pénalisant les poids trop grands.

- **Stratégie d'entraînement :**

Le modèle est entraîné avec différentes durées d'entraînement : 15, 20 et 25 époques, afin de trouver un bon compromis entre performance et temps de calcul. Les résultats obtenus pour chaque configuration seront comparés, dans le but d'identifier le nombre optimal d'époques permettant une convergence efficace du modèle.

Tous les échantillons disponibles sont utilisés pour l'entraînement, sans séparation explicite en validation ou test à ce stade. La **validation** interviendra dans la section suivante (**3.2.5**) lors de la comparaison entre différentes tailles de patchs.

- **Matériel d'entraînement :**

L'entraînement est réalisé sur GPU (CUDA) si disponible, ce qui accélère considérablement les calculs, en particulier pour les opérations de convolution 3D. Sinon, l'exécution se fait automatiquement sur CPU.

- Batch size : 32

Ce choix permet un bon équilibre entre utilisation mémoire et stabilité des gradients. Des batchs plus petits augmenteraient le bruit, tandis que des batchs plus grands consommeraient trop de mémoire.

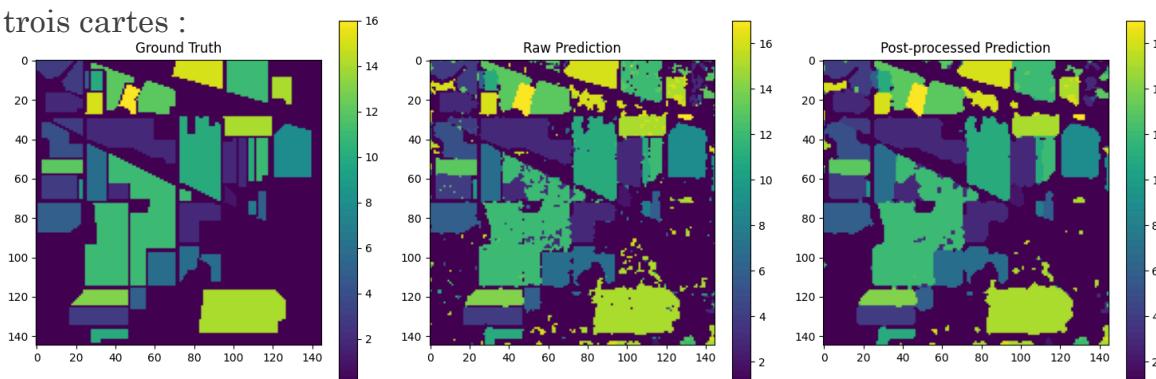
3.2.5 Post-traitement :

Après l'obtention de la carte de segmentation brute produite par le modèle HybridSN, un post-traitement est appliqué afin d'améliorer la qualité des prédictions. Nous avons utilisé un **CRF (Conditional Random Field)**, une méthode probabiliste efficace pour affiner les résultats de segmentation.

Le CRF prend en entrée la carte des probabilités générée par le modèle ainsi que l'image originale. Il permet de **corriger les erreurs de classification isolées** en tenant compte de la **cohérence spatiale et spectrale** des pixels voisins. Concrètement, il favorise les régions homogènes tout en respectant les contours réels des objets, ce qui améliore considérablement la précision de la segmentation, en particulier dans les zones de transition entre différentes classes.

Ce traitement postérieur se traduit visuellement par une carte de segmentation plus lisse, moins bruitée, et plus fidèle à la réalité, comparée à la prédiction brute (Raw Prediction). Cette étape est donc essentielle pour obtenir un résultat final exploitable dans un contexte d'analyse d'image hyperspectrale.

L'image ci-dessous illustre l'impact du post-traitement par CRF. On y trouve trois cartes :



À gauche, l'image Ground Truth, représentant les vraies classes du jeu de données .

Au centre, la carte de segmentation brute produite par le modèle sans traitement supplémentaire.

À droite, la carte après post-traitement, montrant une meilleure délimitation des régions et une réduction du bruit.

3.2.6 Comparaison des performances selon différents hyperparamètres :

comparaison entre différents patchs :

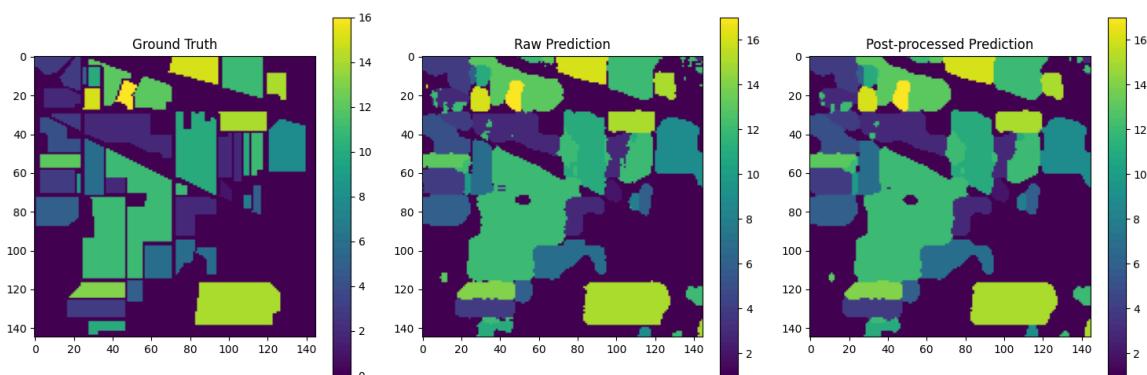
L'objectif ici est de **former un modèle** en utilisant des patches de différentes tailles et d'analyser les résultats pour identifier la taille de patch qui donne les meilleurs résultats en termes de performance.

Comparaison qualitative:

Présentons d'abord l'**image de référence réelle** de la scène *Indian Pines*, qui servira de **base de comparaison visuelle**



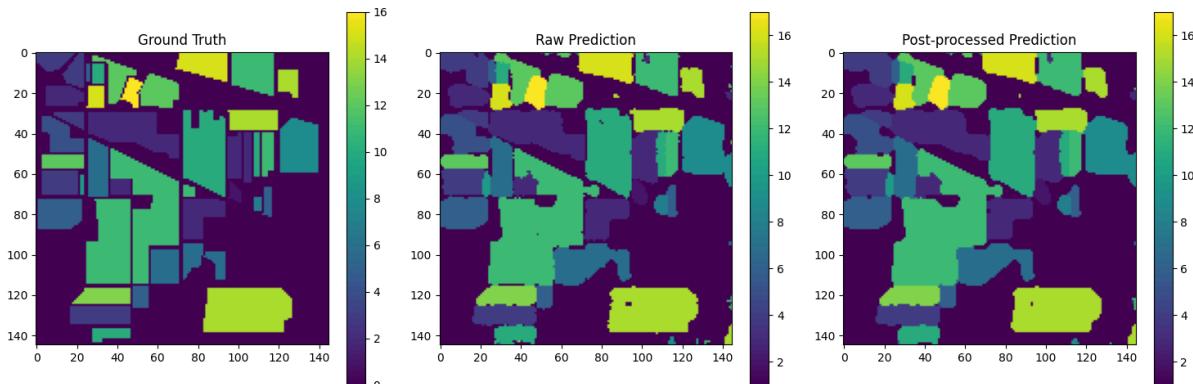
- Résultats d'entraînement sur patch size = 9:



Avec un patch de taille 9, l'image prédite est plus lisse, mais les détails aux bords sont moins visibles. Cela limite la précision sur les contours des objets.

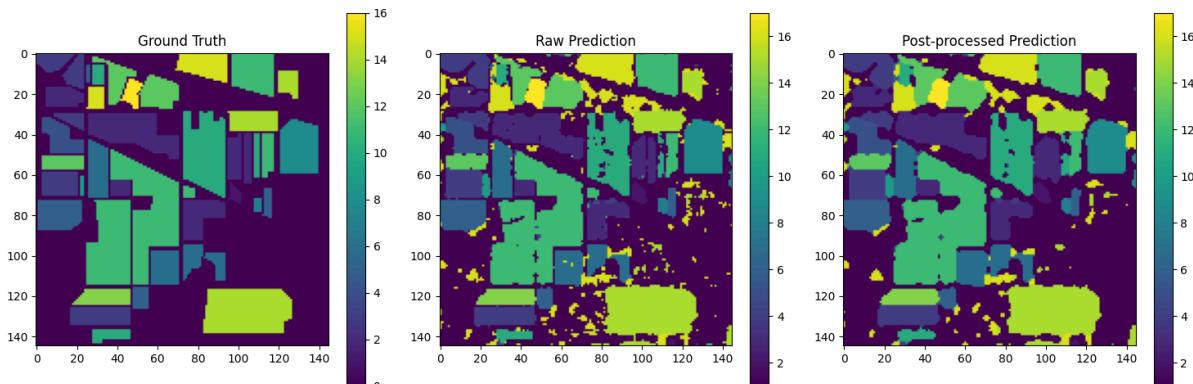
Le post-traitement CRF réduit le bruit et affine les frontières, rendant l'image plus nette et homogène.

- Résultats d'entraînement sur patch size = 7:



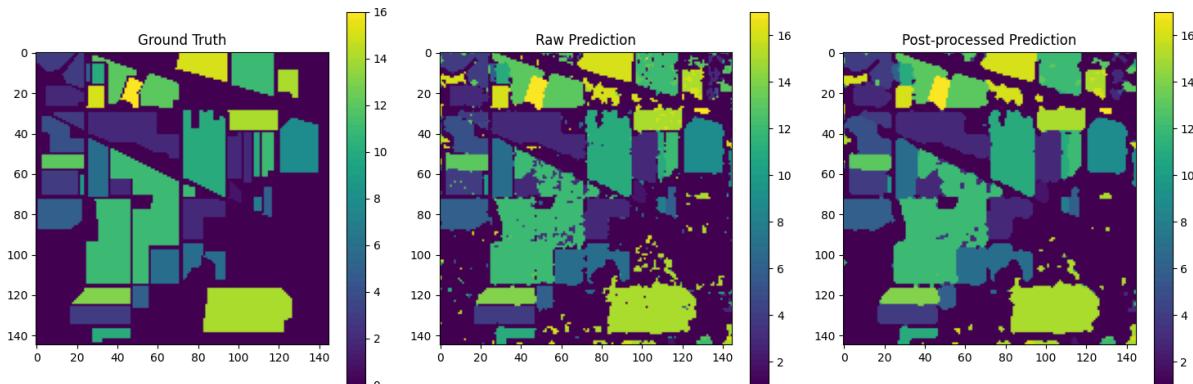
Avec un patch de taille 7, l'image montre plus de détails et une meilleure netteté. Cependant, certaines informations fines restent absentes, ce qui peut affecter la précision globale.

- Résultats d'entraînement sur patch size = 5:



Pour un patch de taille 5, l'image prédite présente un certain niveau de bruit, avec des détails parfois exagérés, ce qui peut donner un aspect moins homogène à l'ensemble. Cependant, cette taille de patch permet de mieux faire ressortir les contours et les limites entre les différentes régions. Les frontières entre classes deviennent plus marquées, ce qui est particulièrement utile pour distinguer les zones adjacentes de manière plus précise, malgré une apparence globale légèrement plus irrégulière.

- Résultats d'entraînement sur patch size = 3:



Avec une taille de patch de 3×3 , les résultats de prédiction ne diffèrent pas énormément de ceux obtenus avec les patches de 5×5 . Toutefois, l'image paraît un peu plus stable, avec une légère diminution du bruit visuel. Cette taille de patch plus petite atténue les détails trop prononcés, tout en conservant les contours essentiels des régions, ce qui aboutit à une segmentation plus homogène et plus cohérente.

Comparaison **quantitative**:

Voici la comparaison des performances selon la taille du patch, sous forme de tableau clair :

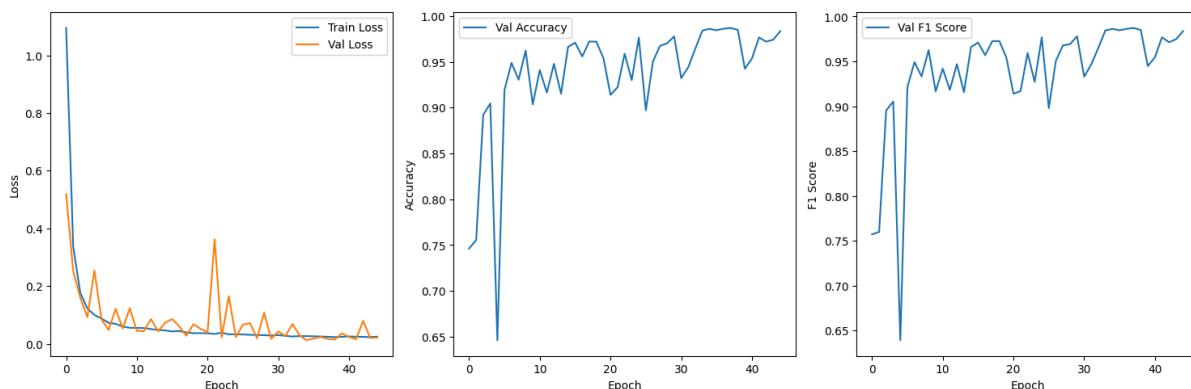
Patch Size	Train Loss	Val Loss	Val Acc	Val F1	Best Val Acc	Best F1 Score
3×3	0.1034	0.0848	0.9015	0.9049	0.9127	0.9133
5×5	0.0556	0.1236	0.9036	0.9166	0.9623	0.9626
7×7	0.0630	0.0460	0.9268	0.9273	0.9368	0.9377
9×9	0.0790	0.1353	0.8809	0.8835	0.9164	0.9162

Conclusion :

Parmi toutes les tailles de patchs testées, **le patch 5×5 s'est révélé être la solution optimale**, offrant le meilleur compromis entre précision, stabilité et F1 Score. Il permet de capturer efficacement l'information spatiale et spectrale sans introduire trop de bruit, contrairement aux patches trop petits (3×3) ou trop larges (9×9). Ce choix constitue donc la base la plus pertinente pour la suite du traitement et de la classification.

comparaison entre différents époque:

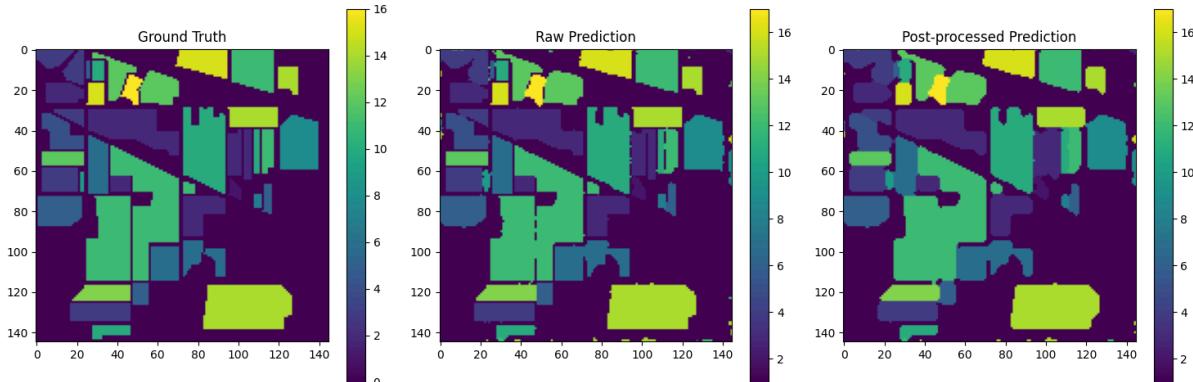
Afin de déterminer le nombre optimal d'époques pour l'entraînement du modèle, nous avons lancé une exécution complète sur **45 époques**. À chaque époque, les **valeurs de la fonction de perte (loss)**, l'**exactitude (accuracy)** et le **F1 Score** ont été enregistrées. Cela nous a permis de tracer des **courbes d'évolution** et d'**observer la convergence du modèle**. L'objectif est d'identifier le **point de stabilité**, où le modèle cesse de s'améliorer significativement, afin de **trouver un bon compromis entre performance et temps de calcul**.



Analyse des résultats :

Bien que des pics de précision supérieurs aient été observés **au-delà de l'époque 35**, nous avons choisi de **retenir l'époque 35** comme point d'arrêt optimal. Cette décision vise à **éviter le surapprentissage (overfitting)**, tout en conservant une **excellente performance** et une bonne généralisation.

Nous visualisons donc les résultats obtenus à l'époque 35, considérée comme le **meilleur compromis entre performance maximale et robustesse du modèle.**



Avec une taille de patch de 5 et un entraînement sur 35 époques, nous avons obtenu les résultats les plus précis et les plus fidèles jusqu'à présent. L'image prédictive se rapproche visuellement beaucoup de la vérité terrain, avec des détails fins et des limites bien définies. Fait intéressant, dans ce cas, la prédiction brute semble plus fidèle que celle post-traitée par CRF. En effet, bien que le CRF réduise habituellement le bruit, ici il a tendance à lisser excessivement l'image, entraînant une perte de certains détails importants capturés par le modèle.

3.2.7 Choix final du modèle :

Après une série d'expérimentations comparant différentes tailles de patchs et différentes durées d'entraînement, nous retenons la configuration suivante comme **la plus performante et la plus équilibrée** :

- Un **patch de taille 5×5**, qui a fourni les meilleurs résultats en termes de F1 Score et de précision.
- Une **durée d'entraînement de 35 époques**, identifiée comme point optimal avant sur-apprentissage.
- Tous les paramètres du modèle ont été ajustés pour **favoriser une convergence stable**, en tenant compte de la complexité des données hyperspectrales et du déséquilibre des classes.

Cette configuration permet d'assurer **une généralisation robuste** tout en maintenant un **temps de calcul raisonnable**.

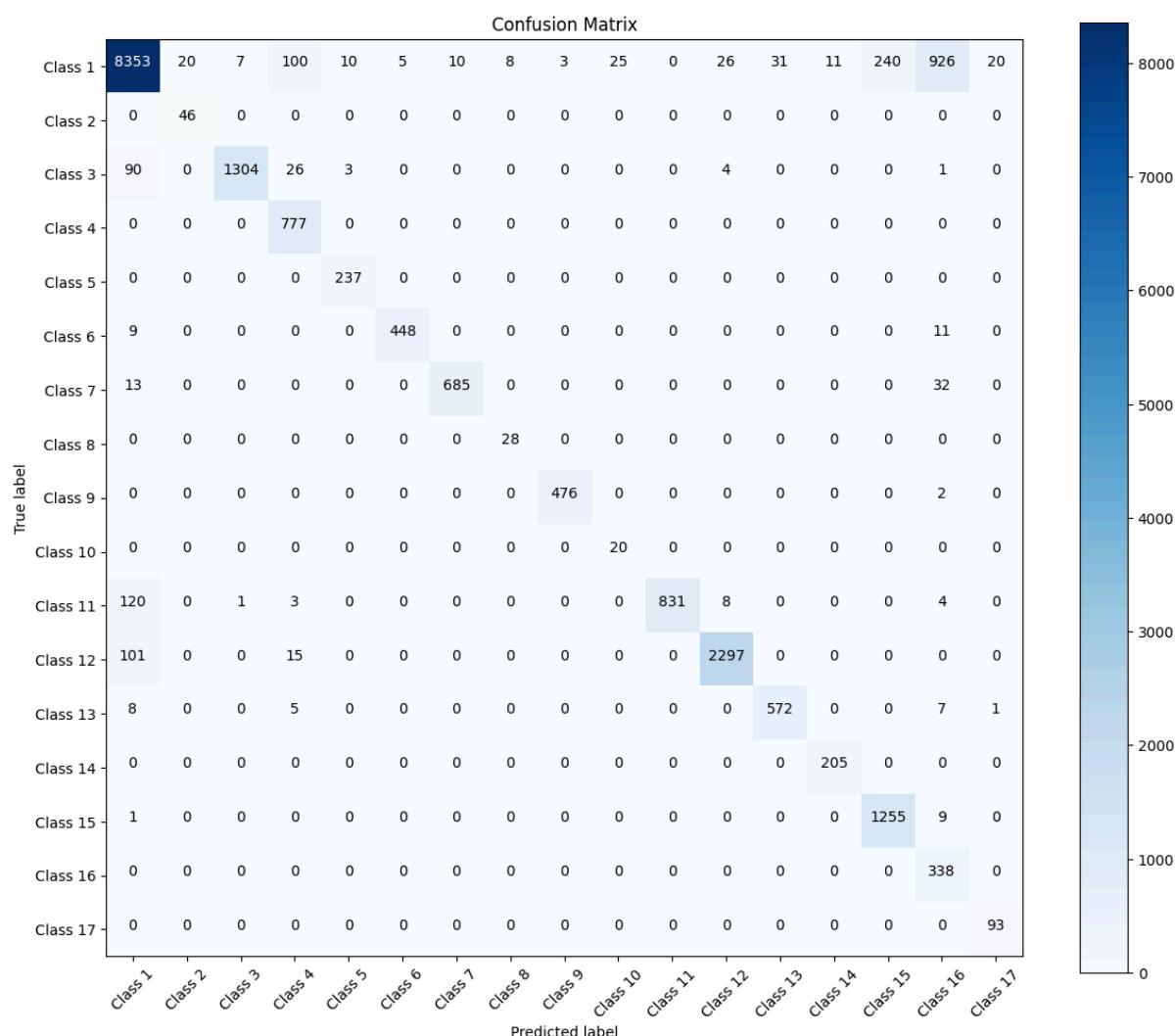


Tableau récapitulatif des paramètres du modèle final :

Taille du patch	5 × 5
Nombre d'époques	35
Batch size	32
Optimiseur	Adam
Learning rate	0.0006
Weight decay	1e-4 (régularisation L2)
Fonction de perte	CrossEntropyLoss (avec pondération)
Matériel	GPU (CUDA) si disponible, sinon CPU

3.2.8 Évaluation finale du modèle :

Dans la suite de notre analyse, nous allons générer une matrice de confusion afin d'évaluer plus précisément la performance du modèle. Celle-ci nous permettra de visualiser dans quelle mesure les pixels ont été correctement classés par rapport à leurs vraies classes, et d'identifier les éventuelles confusions entre certaines catégories.



L'analyse de la matrice de confusion révèle que la majorité des pixels de chaque classe ont été correctement prédits, se concentrant principalement sur la diagonale, ce qui indique une bonne capacité du modèle à distinguer les différentes classes. Pour certaines classes, on observe même une correspondance presque parfaite entre les prédictions et les étiquettes réelles, sans confusion notable avec d'autres classes. D'autres présentent de légères dispersions, avec quelques pixels prédits à tort comme appartenant à des classes voisines, souvent similaires sur le plan spectral. Ces erreurs restent cependant marginales et n'affectent pas significativement la performance globale, ce qui confirme la robustesse et la précision de notre modèle sur les données analysées.



3.2.9 Conclusion :

Dans ce travail, nous avons conçu, entraîné et évalué un modèle CNN hybride combinant des convolutions 3D et 2D pour la classification d'images hyperspectrales sur le dataset *Indian Pines*. Ce choix d'architecture permet de capturer efficacement les relations spatiales et spectrales complexes propres à ce type de données.

Grâce à une stratégie d'entraînement progressive, incluant la pondération des classes et une régularisation adaptée (L2), nous avons pu obtenir une excellente performance globale. Après expérimentation, le meilleur compromis a été atteint avec un patch de taille **5×5** et **35 époques** d'entraînement, évitant à la fois le sous-apprentissage et le sur-apprentissage.

L'évaluation finale démontre une très bonne capacité de généralisation du modèle, avec des scores élevés en précision et F1-score, même pour des classes initialement sous-représentées.

Ce modèle constitue donc une solution robuste et performante pour la classification supervisée d'images hyperspectrales, tout en restant relativement simple à entraîner et à adapter à d'autres scènes similaires.

3.3 model CNN pour l'image Pavia University :

3.3.1 Réutilisation du modèle HybridSN :

Pour le jeu de données Pavia University, nous avons appliqué le même modèle HybridSN utilisé précédemment sur Indian Pines. Étant donné la résolution spatiale plus élevée de cette image, nous avons choisi d'expérimenter deux tailles de patchs : **5×5** et **7×7**, dans le but de mieux capter les détails locaux.

L'architecture du réseau est conservée, avec une adaptation du nombre de canaux d'entrée aux 103 bandes spectrales propres à Pavia. En revanche, le temps d'exécution est notablement plus élevé, principalement dû à la taille plus importante de l'image et à sa richesse spatiale.

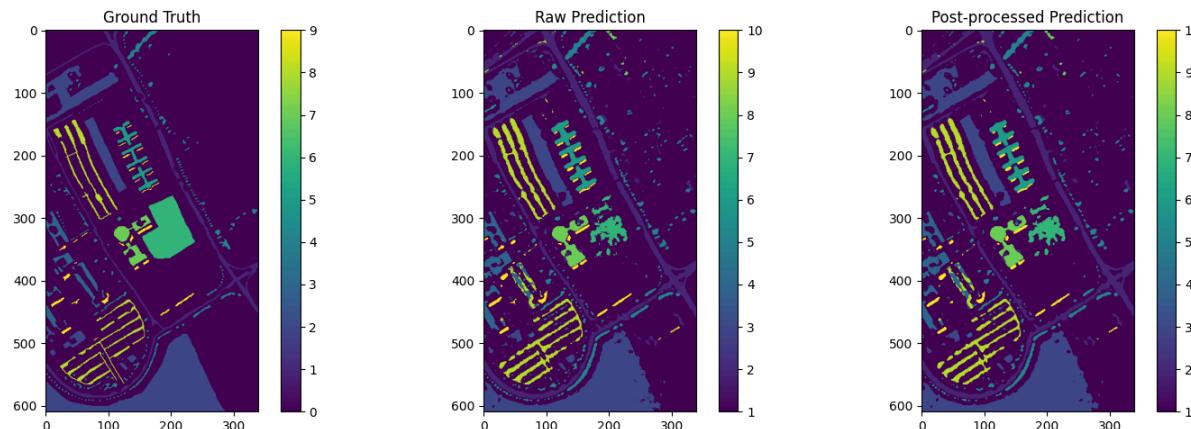
3.3.2 Résultats de classification

Les cartes de prédiction montrent des différences notables selon la taille de patch utilisée.

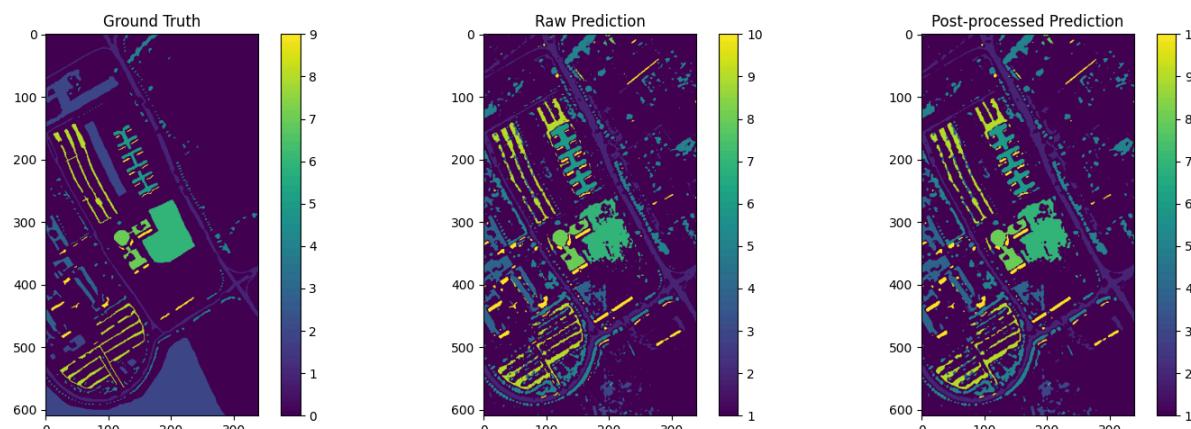
- Avec un patch 5×5 , les bordures sont plus fines, mais on observe un bruit important et une certaine instabilité dans les zones homogènes.
- Avec un patch 7×7 , les contours apparaissent plus lissés, et l'image finale est globalement plus cohérente, même si elle tend à épaisser les contours.

3.3.3 Comparaison qualitative entre patchs 5×5 et 7×7 :

Voici les résultats de visualisation obtenus pour Pavia University obtenu avec un patch 7×7 .



Voici les résultats de visualisation obtenus pour Pavia University obtenu avec un patch 5×5 .



interprétation :

Le **patch 7×7** permet une meilleure **cohérence visuelle** sur l'ensemble de l'image.

Les **zones homogènes sont mieux délimitées**, et les **contours sont plus nets et continus**.

Le **patch 5×5**, bien que plus précis dans certaines zones fines, engendre **plus de bruit et moins de stabilité** sur les régions étendues.

3.3.4 Comparaison quantitative :

Patch Size	Accuracy	F1 Score	Observations
5×5	0.7756	0.7680	Précision locale, mais bruit élevé
7×7	0.9104	0.9092	Meilleure adaptation à la résolution spatiale de Pavia

3.3.5 Conclusion sur Pavia University

Le **patch 7×7 est mieux adapté à Pavia University** grâce à sa **Résolution spatiale plus élevée**, permettant de mieux capturer les structures urbaines complexes et les textures plus fines.

En revanche, la **performance globale du modèle sur Pavia reste inférieure** à celle obtenue sur *Indian Pines*. Cela peut s'expliquer par la **nature urbaine** de l'image, caractérisée par une plus grande variabilité spatiale et spectrale, rendant la classification **plus difficile** avec le même modèle.

3.4. Impact de l'exclusion du fond (background) sur la segmentation :

Dans cette section, nous allons explorer l'impact de l'exclusion du fond sur la segmentation des images satellites des ensembles de données Indian Pines et Pavia University. Nous avons formé un modèle en utilisant l'image Indian Pines avec fond, et nous allons maintenant tester la même architecture de modèle sur des images où le fond a été exclu. Cette comparaison nous permettra d'évaluer comment l'absence de fond affecte la performance du modèle de segmentation.

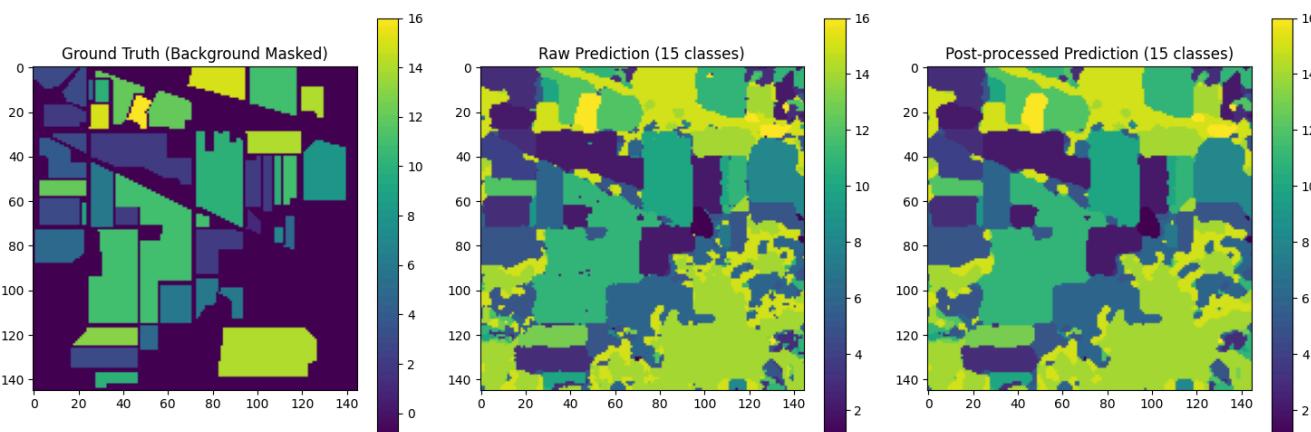
3.4.1 Entraînement du modèle :

Nous avons d'abord utilisé un modèle de segmentation existant qui a été formé sur l'image avec le fond. Ce modèle a été entraîné à l'aide de l'image spectrale Indian Pines. Les données ont été prétraitées et normalisées avant d'être introduites dans le modèle.

L'étape suivante a consisté à utiliser le même modèle pour entraîner les images sans fond. Pour ce faire, nous avons éliminé la zone de fond des images, en nous concentrant uniquement sur les régions d'intérêt. Cela a impliqué de définir un masque qui excluait le fond, en ne conservant que les pixels correspondant aux classes spécifiques présentes dans les scènes.

Voici le résultat de cet entraînement :

- une accuracy en validation de 0.9901,
- une perte d'entraînement (Train Loss) de 0.0176,
- une perte en validation (Val Loss) de 0.0255.





3.4.2 Observation des résultats :

Le modèle a montré un bon comportement avec l'image sans fond, obtenant une accuracy en validation de 0.9901, une perte d'entraînement (Train Loss) de 0.0176, et une perte en validation (Val Loss) de 0.0255. Toutefois, l'absence de fond a conduit le modèle à remplir ces zones vides en assignant à ces régions la classe la plus proche, souvent celle des pixels voisins. Cela a entraîné des résultats où le modèle tente de "reconstituer" le fond par les classes les plus susceptibles d'être présentes, mais la segmentation obtenue n'est pas aussi similaire à celle obtenue avec l'image d'origine contenant un fond. Ce phénomène est dû à l'absence de contextes ou d'informations supplémentaires apportées par les pixels de fond, ce qui peut influencer l'exactitude des segments dans des zones non couvertes.

3.4.3 Évaluation :

L'évaluation du modèle a été réalisée en se basant sur l'accuracy, la perte d'entraînement (Train Loss) et la perte en validation (Val Loss). Les résultats obtenus montrent une accuracy en validation de 0.9901, une perte d'entraînement de 0.0176, et une perte en validation de 0.0255, indiquant que le modèle a bien appris à classifier les régions d'intérêt tout en conservant une faible erreur de généralisation sur les données de validation. Ces métriques suggèrent que le modèle a une bonne capacité à segmenter les images sans fond, tout en maintenant une performance compétitive par rapport aux images avec fond.

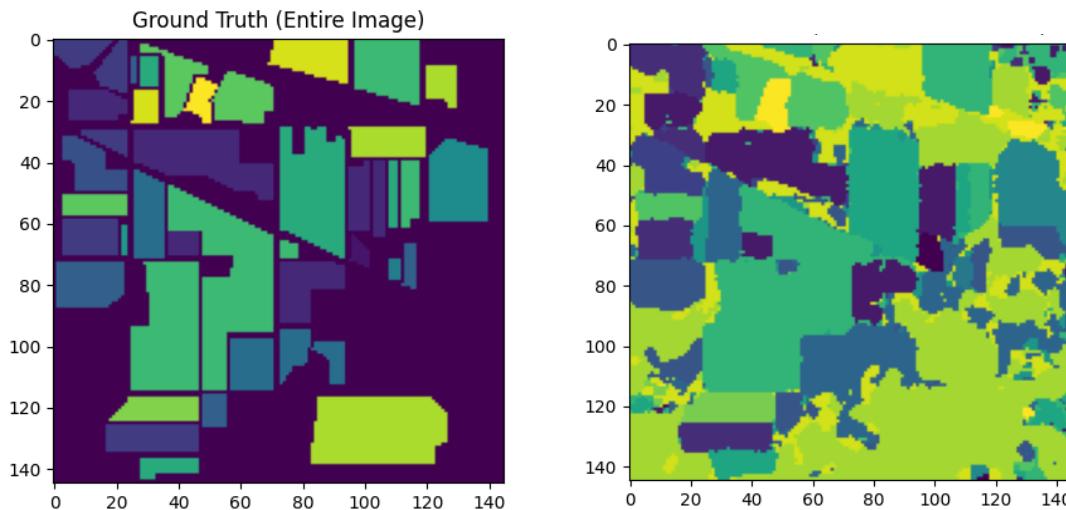
3.4.4 Tester le modèle sur l'image Indian Pines (avec fond) :

Dans cette étape, nous avons évalué le modèle préalablement entraîné sans la classe de fond en le testant sur l'image Indian Pines complète, incluant le fond.

Voici les résultats obtenus lors de cette évaluation :

- Accuracy (exactitude) sur l'image complète : 0.4838
- F1-score sur l'image complète : 0.7355

Voici ci-dessous la visualisation de la prédiction obtenue



Interprétation des résultats :

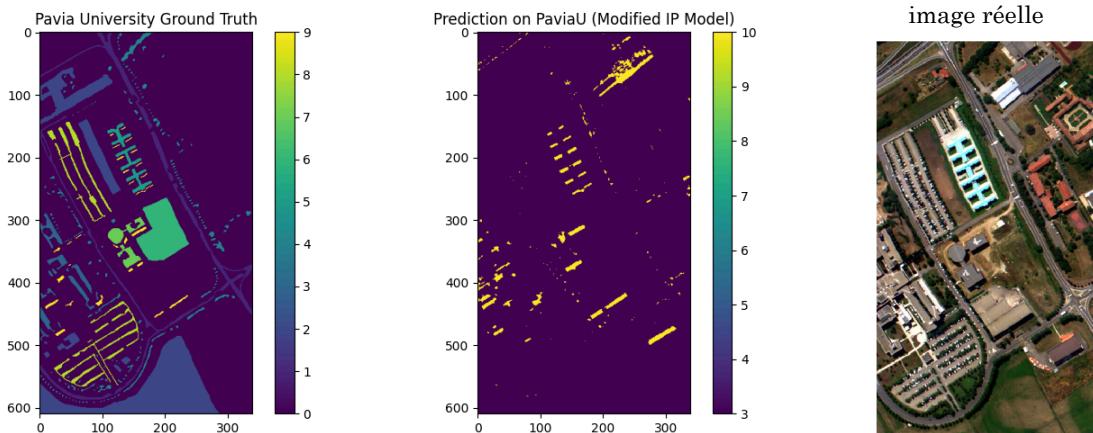
L'analyse visuelle de la segmentation montre que le modèle **ne parvient pas à bien reconnaître les régions correspondant à la classe 0 (le fond)**. En effet, cette classe est largement ignorée dans la prédiction, car elle n'a jamais été présentée au modèle pendant l'entraînement. On observe également une certaine confusion dans les contours des objets et une segmentation moins précise dans les zones mixtes. Cela entraîne une **précision globale relativement faible (48.38%)**, malgré un **F1-score plus acceptable (73.55%)**, ce qui reflète que le modèle reste capable d'identifier correctement plusieurs classes **non-fond** tout en étant perturbé par les pixels de fond inconnus.

Cette expérience confirme qu'**exclure la classe de fond lors de l'entraînement rend le modèle moins adapté à des images contenant cette classe**, surtout si on souhaite l'utiliser sur l'image entière sans pré-traitement. Pour des applications sur l'image complète, il serait donc préférable de conserver la classe de fond ou d'utiliser un masque de segmentation préalable pour éviter son influence.

3.4.5 Tester le modèle sur l'image Pavia University :

Dans la continuité de l'expérience précédente, nous avons appliqué le même modèle, entraîné sans la classe de fond, sur l'image Pavia University complète. L'objectif est de vérifier si l'exclusion du fond pendant l'apprentissage affecte également les performances du modèle sur cet autre jeu de données, ou si l'impact varie selon la nature de l'image.

Voici ci-dessous la visualisation de la prédiction obtenue :



Interprétation :

En comparant l'image réelle de Pavia University à la prédiction produite par le modèle entraîné sans la classe de fond, nous observons un comportement très spécifique : le modèle prédit uniquement les zones végétales, c'est-à-dire principalement les espaces verts et les arbres situés dans l'environnement universitaire. Ces zones sont effectivement bien localisées et cohérentes avec la réalité, ce qui montre que le modèle a bien appris à reconnaître ce type de texture ou de signature spectrale.

En revanche, aucune autre classe n'est prédite : ni les routes, ni les bâtiments, ni les zones urbaines ne sont identifiées. Cela s'explique par le fait que, lors de l'entraînement, le modèle n'a été exposé qu'à des exemples végétaux (issus d'Indian Pines sans fond), et n'a donc jamais appris à distinguer ou classifier les autres types de surfaces. Par conséquent, lorsqu'il est confronté à une image plus variée comme celle de Pavia University, il projette uniquement les connaissances acquises, à savoir la reconnaissance de la végétation.

Cette prédiction sélective met en lumière les limites de la généralisation d'un modèle lorsqu'on restreint son apprentissage à un sous-ensemble de classes. Bien qu'il excelle sur les classes connues (ici, la végétation), il échoue totalement à identifier toute autre entité qui ne figure pas dans sa base d'apprentissage initiale. Cela souligne l'importance d'un jeu d'entraînement représentatif et diversifié si l'on souhaite obtenir une segmentation complète et réaliste sur des images complexes.



3.4.6 Conclusion :

Les expériences ont montré que l'exclusion du fond peut, dans certains cas, améliorer la qualité de la segmentation. Le modèle entraîné uniquement sur des zones végétales (à partir d'Indian Pines sans fond) a échoué à prédire correctement les zones non végétales lorsqu'elles ont été réintroduites, d'où une performance dégradée.

En revanche, lorsqu'il a été appliqué à l'image Pavia University, il a réussi à identifier précisément les zones végétales (arbres, pelouses), tout en ignorant totalement les zones urbaines. Cela montre que le modèle a bien appris à segmenter la végétation, mais reste limité aux classes vues pendant l'entraînement.

Ces résultats soulignent qu'un apprentissage ciblé peut être efficace, à condition que les classes à prédire soient similaires à celles vues pendant l'apprentissage. Toutefois, cela réduit la capacité de généralisation du modèle à des scènes plus variées.

4. Conclusion Générale:

Ce projet nous a permis d'explorer en profondeur les spécificités des images hyperspectrales, à travers plusieurs étapes d'analyse, de traitement et de classification.

L'étude des bandes spectrales nous a montré qu'une image hyperspectrale est bien plus qu'une image classique : elle contient une richesse d'informations réparties sur des centaines de longueurs d'onde. Cette diversité permet de détecter des détails invisibles à l'œil nu, mais nécessite une sélection rigoureuse des bandes pertinentes selon le contexte.

La visualisation des vérités terrain nous a révélé la complexité des scènes réelles, marquées par des classes déséquilibrées ou spatialement proches, rendant la classification plus difficile. Elle met en évidence le rôle crucial d'annotations de qualité pour entraîner des modèles fiables.



INSTITUT NATIONAL DES SCIENCES APPLIQUÉES ET DE TECHNOLOGIE

L'application de la réduction de dimension par PCA a mis en lumière l'importance de simplifier les données sans perdre l'essentiel. Les premières composantes capturent l'essentiel de la variance, ce qui facilite le traitement tout en conservant l'information utile.

L'architecture HybridSN, combinant réduction de dimension, apprentissage profond (3D/2D CNN) et post-traitement (CRF), a démontré de bonnes performances sur les jeux de données Indian Pines et Pavia University. Le choix des hyperparamètres (taille des patches, nombre d'époques, etc.) s'est révélé déterminant pour adapter le modèle aux caractéristiques des images.

Une expérimentation particulière, consistant à entraîner le modèle sans la classe de fond, a permis de mieux comprendre son comportement face à des classes absentes. Lors du test sur Pavia University, bien que le modèle n'ait appris que sur des zones végétales, il a su détecter efficacement les zones vertes (arbres, espaces naturels), mais a échoué à reconnaître les routes ou zones urbaines. Cela montre une bonne capacité de généralisation sur des classes apprises, mais aussi une limitation importante face aux classes non vues.

En somme, ce travail nous a permis de développer une véritable intuition sur les données hyperspectrales et les méthodes adaptées à leur traitement. Il ouvre des perspectives intéressantes pour des applications ciblées (ex. détection de végétation), tout en soulignant l'importance de la diversité des classes lors de l'apprentissage pour garantir une classification complète et fiable.

Code Source :

Le code source de ce projet est disponible [\[ici\]](https://github.com/NaouresBzeouich/Image_Processing_of_Multispectral_Data_Project) (https://github.com/NaouresBzeouich/Image_Processing_of_Multispectral_Data_Project).