

الجمهورية الجزائرية الديمقراطية الشعبية  
People's Democratic Republic of Algeria

وزارة التعليم العالي والبحث العلمي  
Ministry of Higher Education and Scientific Research

المدرسة العليا للإعلام الآلي ٨ ماي ١٩٤٥ - سيدي بلعباس  
Higher School of Computer Science  
8 Mai 1945 - Sidi Bel Abbès



## Master's Thesis

To obtain the diploma of **Master's Degree**  
Field of Study: **Computer Science**  
Specialization: **IASD**

## Theme

---

# Adapting Vision-Language Models for Remote Sensing-Based Precision Agriculture

---

Presented by  
**Meflah Yousra**

Defended on: **July 6, 2026**  
*In front of the jury composed of*

**Mr. [Jury Member Name]**  
**Mr. Larabi Mohammed ElAmine**  
**Mr. Iftene Meziane**  
**Mr. Chaib Souleyman**  
**Mr. [Jury Member Name]**

President of the Jury  
Thesis Supervisor  
Co-Supervisor  
Co-Supervisor  
Examiner

*Academic Year: 2025/2026*

## Abstract

Vision-Language Models (VLMs) have emerged as powerful multimodal foundation models capable of linking visual and textual information through a shared semantic space. While these models demonstrate strong performance on natural-image tasks, their direct application to remote sensing imagery remains challenging due to substantial differences in viewing geometry, spatial scale, spectral composition, and domain-specific semantics. These limitations are particularly critical in precision agriculture, where accurate interpretation of satellite and UAV imagery is required for tasks such as crop classification, disease detection, vegetation monitoring, and tree mapping.

This thesis presents a systematic review of adaptation strategies designed to bridge the gap between general-purpose VLMs and remote sensing imagery, with a specific focus on agricultural applications. Following a PRISMA-based methodology, 111 studies published between 2019 and 2026 were analyzed across five major scientific databases. A novel three-level taxonomy is proposed, categorizing adaptation approaches into data-level, model-level, and training/inference-level strategies. The review examines preprocessing techniques, multimodal data fusion, synthetic data generation, fine-tuning approaches, parameter-efficient adaptation methods, prompt learning techniques, and hybrid frameworks.

Beyond methodological analysis, the thesis evaluates adaptation strategies from an engineering perspective by considering computational requirements, annotation costs, deployment latency, hardware constraints, and operational readiness. Comparative analyses indicate that parameter-efficient approaches, particularly Low-Rank Adaptation (LoRA), provide the most favorable trade-off between performance and computational cost. The study further highlights the growing importance of multimodal fusion and remote sensing-specific foundation models while identifying persistent challenges related to spectral information utilization, geographic generalization, hallucination mitigation, and deployment in resource-constrained environments.

The findings culminate in a deployment-oriented decision framework and a research roadmap for future agricultural VLM systems. By synthesizing recent advances and organizing them within a unified taxonomy, this work provides both researchers and practitioners with practical guidance for selecting and deploying VLM adaptation strategies in precision agriculture and remote sensing applications.

**Keywords:** Vision-Language Models, Remote Sensing, Precision Agriculture, Foundation Models, CLIP, Multimodal Learning, Parameter-Efficient Fine-Tuning, LoRA, RemoteCLIP, Agricultural Monitoring.

## Résumé

Les modèles de vision-langage (VLMs) sont des modèles multimodaux puissants capables de relier les informations visuelles et textuelles à travers un espace sémantique partagé. Bien que ces modèles démontrent de bonnes performances sur les tâches d'images naturelles, leur application directe à l'imagerie de télédétection reste difficile en raison des différences substantielles en termes de géométrie de visualisation, d'échelle spatiale, de composition spectrale et de sémantique spécifique au domaine. Ces limitations sont particulièrement critiques dans l'agriculture de précision, où une interprétation précise des images satellitaires et des drones est nécessaire pour des tâches telles que la classification des cultures, la détection des maladies, la surveillance de la végétation et la cartographie des arbres.

Cette thèse présente une revue systématique des stratégies d'adaptation conçues pour combler le fossé entre les VLMs généralistes et l'imagerie de télédétection, avec un accent particulier sur les applications agricoles. Selon une méthodologie basée sur PRISMA, 111 études publiées entre 2019 et 2026 ont été analysées à travers cinq bases de données scientifiques majeures. Une nouvelle taxonomie à trois niveaux est proposée, catégorisant les approches d'adaptation en stratégies au niveau des données, au niveau du modèle et au niveau de l'inférence/entraînement. La revue examine les techniques de prétraitement, la fusion de données multimodales, la génération de données synthétiques, les approches de réglage fin, les méthodes d'adaptation efficaces en paramètres, les techniques d'apprentissage par incitation et les cadres hybrides.

Au-delà de l'analyse méthodologique, la thèse évalue les stratégies d'adaptation d'un point de vue ingénierie en considérant les exigences computationnelles, les coûts d'annotation, la latence de déploiement, les contraintes matérielles et la préparation opérationnelle. Les analyses comparatives indiquent que les approches efficaces en paramètres, en particulier l'adaptation de bas rang (LoRA), offrent le compromis le plus favorable entre performance et coût computationnel. L'étude souligne également l'importance croissante de la fusion multimodale et des modèles fondamentaux spécifiques à la télédétection, tout en identifiant les défis persistants liés à l'utilisation des informations spectrales, à la généralisation géographique, à l'atténuation des hallucinations et au déploiement dans des environnements à ressources limitées.

Les conclusions aboutissent à un cadre de décision orienté vers le déploiement et à une feuille de route de recherche pour les futurs systèmes VLM agricoles. En synthétisant les avancées récentes et en les organisant au sein d'une taxonomie unifiée, ce travail fournit aux chercheurs et aux praticiens des conseils pratiques pour sélectionner et déployer des stratégies d'adaptation des VLMs dans l'agriculture de précision et les applications de télédétection.

**Mots-clés :** Modèles de Vision-Langage, Télédétection, Agriculture de Précision, Modèles Fondamentaux, CLIP, Apprentissage Multimodal, Adaptation Efficace en Paramètres, LoRA, RemoteCLIP, Surveillance Agricole.

## ملخص

ظهرت نماذج الرؤية واللغة (VLMs) كنماذج أساسية متعددة الوسائط قوية قادرة على ربط المعلومات البصرية والنصية من خلال فضاء دلالي مشترك. ورغم أن هذه النماذج تُظهر أداءً قوياً في مهام الصور الطبيعية، إلا أن تطبيقها المباشر على صور الاستشعار عن بعد لا يزال يمثل تحدياً بسبب الاختلافات الجوهرية في هندسة الرؤية، والمقياس المكاني، والتركيب الطيفي، والدلالات الخاصة بالمجال. وتكتسي هذه القيود أهمية بالغة في الزراعة الدقيقة، حيث يتطلب التفسير الدقيق لصور الأقمار الصناعية والطائرات بدون طيار أداء مهام مثل تصنيف المحاصيل، والكشف عن الأمراض، ومراقبة الغطاء النباتي، ورسم خرائط الأشجار.

تقدم هذه الأطروحة مراجعة منهجية لاستراتيجيات التكيف المصممة لسد الفجوة بين نماذج الرؤية واللغة العامة وصور الاستشعار عن بعد، مع تركيز خاص على التطبيقات الزراعية. ووفقاً لمنهجية PRISMA، تم تحليل 111 دراسة منشورة بين عامي 2019 و2026 عبر خمس قواعد بيانات علمية رئيسية. ويُقترح تصنيف ثلاثي المستويات جديد، يصنف أساليب التكيف إلى استراتيجيات على مستوى البيانات، وعلى مستوى النموذج، وعلى مستوى الاستدلال/التدريب. تستعرض المراجعة تقنيات المعالجة المسبقة، ودمج البيانات متعددة الوسائط، وتوليد البيانات الاصطناعية، وأساليب الضبط الدقيق، وطرق التكيف الفعالة من حيث المعايير، وتقنيات التعلم بالمطالبات، والأطر الهجينة.

إلى جانب التحليل المنهجي، تقوم الأطروحة بتقييم استراتيجيات التكيف من منظور هندسي من خلال النظر في المتطلبات الحسابية، وتكاليف التعليق التوضيحي، وزمن انتقال النشر، وقيود الأجهزة، والجاهزية التشغيلية. تشير التحليلات المقارنة إلى أن الأساليب الفعالة من حيث المعايير، ولا سيما تكيف الرتبة المنخفضة (LoRA)، تقدم المقايضة الأكثر ملاءمة بين الأداء والتكلفة الحسابية. تسلط الدراسة الضوء أيضاً على الأهمية المتزايدة لدمج الوسائط المتعددة والنماذج الأساسية الخاصة بالاستشعار عن بعد، مع تحديد التحديات المستمرة المتعلقة باستخدام المعلومات الطيفية، والتعميم الجغرافي، والتخفيف من الهلوسة، والنشر في البيئات محدودة الموارد.

تتوج النتائج بإطار قرار موجه نحو النشر وخريطة طريق بحثية لأنظمة الرؤية واللغة الزراعية المستقبلية. من خلال تجميع التطورات الحديثة وتنظيمها ضمن تصنيف موحد، يقدم هذا العمل للباحثين والممارسين إرشادات عملية لاختيار ونشر استراتيجيات تكيف نماذج الرؤية واللغة في الزراعة الدقيقة وتطبيقات الاستشعار عن بعد.

**الكلمات المفتاحية:** نماذج الرؤية واللغة، الاستشعار عن بعد، الزراعة الدقيقة، النماذج الأساسية، CLIP، التعلم متعدد الوسائط، التكيف الفعال من حيث المعايير، LoRA، RemoteCLIP، المراقبة الزراعية.