

People's Democratic Republic of Algeria
Ministry of Higher Education and Scientific Research
Higher School of Computer Science
08 May 1945 – Sidi Bel Abbès



THESIS

In view of obtaining the **Master's** diploma

Field : **Computer Science**

Specialty : **Artificial Intelligence and Data Science (AIDS)**

Theme

**Multi-Expert Pipelines and Vision-Language Models for
Automated Telecom Infrastructure Inspection Verification**

Presented by :

- Mr. Saadi Ahmed

Submission date : **29/06/2026**

Before the jury composed of :

- Dr. KAZI TANI Mohammed Yassine
- Dr. TAOULI Amina
- Dr. DIF Nassima

President
Examiner
Supervisor

Academic Year : 2025/2026

Abstract

Automating the verification of telecommunications infrastructure inspections requires the use of computer vision and multimodal reasoning. Currently, in the maintenance process, field inspectors collect photographs, numerical readings, and compliance data from many locations, while a relatively small number of human specialists are responsible for verifying the submitted reports. This approach causes bottlenecks, especially as the number of sites being checked increases and when the verification process requires analysing heterogeneous visual and numerical evidence.

The thesis starts with the formalisation of the problem of inspection verification and the definition of the main obstacles to automation. These consist of the variety of checklists used for inspection, the lack of annotated data, the variability of outdoor photography, the necessity to detect tiny objects and analyse low-contrast display readings, and the known problems of general-purpose vision-language systems with counting, spatial reasoning, and OCR tasks. Overall, all discussed aspects show that telecom inspection verification involves more than an image classification task.

The study evaluates relevant datasets, task-specific computer vision models, vision-language models, and multi-agent inspection frameworks. The analysis confirms that specialised models can be efficient for specific subtasks like object detection, corrosion segmentation, and meter reading, but they require structured labelling and have limited generalisation capabilities when it comes to the full range of inspection checklist elements. Vision-language models allow increased flexibility via natural language commands but might produce unreliable outputs in domain-specific and spatially complex inspection situations. Thus, the study highlights the necessity of comparing task-decomposed specialist models with fine-tuned vision-language models for future telecom inspection systems.

The study lays down both the theoretical and technical groundwork for creating automated inspection verification systems in telecommunications infrastructure. This thesis argues that a reliable system should combine the efficiency of specialist computer vision models with the flexibility of multimodal reasoning while keeping human verification for cases of uncertainty and high risk.

Keywords: Telecommunications infrastructure inspection, computer vision, multimodal reasoning, vision-language models, multi-agent systems, inspection verification.

Résumé

L'automatisation de la vérification des inspections de l'infrastructure de télécommunications nécessite l'utilisation de la vision par ordinateur et du raisonnement multimodal. Dans le processus actuel de maintenance, les inspecteurs sur le terrain collectent des photographies, des mesures numériques et des données de conformité provenant de nombreux sites, tandis qu'un nombre relativement limité de spécialistes humains est chargé de vérifier les rapports soumis. Cette situation crée des goulots d'étranglement, surtout lorsque le nombre de sites à contrôler augmente et lorsque la vérification exige l'analyse de preuves visuelles et numériques hétérogènes.

Le mémoire commence par la formalisation du problème de vérification des inspections et par la définition des principaux obstacles à son automatisation. Ces obstacles comprennent la variété des listes de contrôle utilisées, le manque de données annotées, la variabilité des images prises en environnement extérieur, la nécessité de détecter de petits objets et d'analyser des affichages à faible contraste, ainsi que les limites connues des modèles vision-langage généralistes dans les tâches de comptage, de raisonnement spatial et d'OCR. Ces aspects montrent que la vérification des inspections télécoms va bien au-delà d'une simple tâche de classification d'images.

L'étude examine des ensembles de données pertinents, des modèles de vision par ordinateur spécialisés, des modèles vision-langage et des architectures multi-agents pour l'inspection. L'analyse montre que les modèles spécialisés peuvent être efficaces pour des sous-tâches précises, comme la détection d'objets, la segmentation de la corrosion et la lecture des compteurs, mais qu'ils nécessitent des annotations structurées et généralisent difficilement à l'ensemble des éléments des listes de contrôle. Les modèles vision-langage offrent plus de flexibilité grâce aux instructions en langage naturel, mais peuvent produire des résultats peu fiables dans des situations spécifiques au domaine ou spatialement complexes.

En général, cette étude fournit une base théorique et technique pour le développement de systèmes automatisés de vérification des inspections dans l'infrastructure de télécommunications. Elle soutient qu'un système fiable devrait combiner l'efficacité des modèles spécialistes de vision par ordinateur avec la flexibilité du raisonnement multimodal, tout en conservant la vérification humaine pour les cas d'incertitude et de risque élevé.

Mots-clés : infrastructure de télécommunications, vision par ordinateur, raisonnement multimodal, modèles vision-langage, systèmes multi-agents, vérification des inspections.

ملخص

تتطلب أتمتة التحقق من عمليات تفتيش البنية التحتية للاتصالات استخدام الرؤية الحاسوبية والاستدلال متعدد الوسائط. في عملية الصيانة الحالية، يقوم المفتشون الميدانيون بجمع الصور والقراءات الرقمية وبيانات المطابقة من عدد كبير من المواقع، بينما يتولى عدد محدود نسبيا من المختصين التحقق من التقارير المقدمة. يؤدي هذا الوضع إلى ظهور اختناقات في سير العمل، خاصة عند زيادة عدد المواقع التي يتم فحصها، وعندما تتطلب عملية التحقق تحليل أدلة بصرية ورقمية غير متجانسة.

تبدأ هذه المذكرة بصياغة مشكلة التحقق من عمليات التفتيش وتحديد أهم العوائق التي تواجه أتمتتها. تشمل هذه العوائق تنوع قوائم التحقق المستعملة في التفتيش، ونقص البيانات الموسومة، وتغير ظروف التقاط الصور في البيئة الخارجية، والحاجة إلى كشف الأجسام الصغيرة وتحليل الشاشات ذات التباين الضعيف، إضافة إلى محدودية نماذج الرؤية واللغة العامة في مهام العد، والاستدلال المكاني، والتعرف الضوئي على الحروف OCR. وتبين هذه الجوانب أن التحقق من تفتيش مواقع الاتصالات يتجاوز مهمة تصنيف الصور البسيطة.

تدرس هذه المذكرة مجموعات بيانات ونماذج رؤية حاسوبية متخصصة ونماذج رؤية ولغة وأطر تفتيش متعددة الوكلاء. يبين التحليل أن النماذج المتخصصة يمكن أن تكون فعالة في مهام محددة مثل كشف الأجسام، وتقسيم مناطق التآكل، وقراءة العدادات، لكنها تحتاج إلى تعليقات منظمة ولا تعمم بسهولة على جميع عناصر قوائم التحقق. في المقابل، توفر نماذج الرؤية واللغة مرونة أكبر من خلال التعليمات المكتوبة باللغة الطبيعية، لكنها قد تقدم نتائج غير موثوقة في الحالات الخاصة بالمجال أو الحالات التي تتطلب استدلالا مكانيا دقيقا.

بصفة عامة، توفر هذه الدراسة أساسا نظريا وتقنيا لتطوير أنظمة آلية للتحقق من عمليات تفتيش البنية التحتية للاتصالات. وتؤكد أن النظام الموثوق يجب أن يجمع بين كفاءة نماذج الرؤية الحاسوبية المتخصصة ومرونة الاستدلال متعدد الوسائط، مع الإبقاء على التحقق البشري في الحالات غير المؤكدة أو عالية الخطورة.

الكلمات المفتاحية: البنية التحتية للاتصالات، الرؤية الحاسوبية، الاستدلال متعدد الوسائط، نماذج الرؤية واللغة، الأنظمة متعددة الوكلاء، التحقق من التفتيش.