

FONDATION D'ENTREPRISE



**EASY-(DAI)2**  
**Economical & Automated SYstem**  
**for Distresses Detection on**  
**Airfield Infrastructure with**  
**Artificial Intelligence**

**Service Technique de l'Aviation Civile**  
**Université Gustave Eiffel**



**STAC**  
Service technique  
de l'Aviation civile

Lauréat de l'appel à projets 2020  
« Intelligence artificielle (IA) appliquée aux infrastructures en service »

Pilotes du projet EASY-(DAI)2 et auteurs du livrable :

- Michaël BROUTIN (DGAC/STAC)
- Fabien MENANT (Université Gustave Eiffel)

# Sommaire

<b>1</b>	<b>Résumé.....</b>	<b>4</b>
<b>2</b>	<b>Contexte et objectifs du projet .....</b>	<b>5</b>
2.1	Contexte .....	5
2.2	Objectifs du projet .....	5
<b>3</b>	<b>Les solutions techniques proposées.....</b>	<b>6</b>
3.1	Le mode de collecte de données .....	6
3.2	L'instrumentation .....	8
3.3	Le traitement et l'analyse des images par Deep Learning .....	10
<b>4</b>	<b>Applications expérimentales et évaluation des résultats.....</b>	<b>16</b>
4.1	Tests en laboratoire .....	16
4.2	Constitution de la vérité terrain.....	17
4.3	Critères utilisés pour caractériser le niveau de performance des résultats .....	18
4.4	Expérimentation sur le site de Lognes .....	19
4.5	Expérimentation sur le site de Roissy – Aéroport Charles de Gaulle .....	25
<b>5</b>	<b>Synthèse .....</b>	<b>33</b>
<b>6</b>	<b>Utilisation de l'aide de la Ferec dans le projet.....</b>	<b>34</b>
<b>7</b>	<b>Impact.....</b>	<b>34</b>
7.1	Mise en évidence du caractère collaboratif et d'intérêt général .....	34
7.2	Apports et retombées pour la collectivité .....	34
7.3	Publications et autres éléments de communication auxquels le projet et ses résultats ont donné lieu.....	35
<b>8</b>	<b>Perspectives ouvertes par le projet notamment en matière de recherche collaborative ...</b>	<b>35</b>
<b>9</b>	<b>Annexes (albums d'images en sortie du modèle DL) .....</b>	<b>36</b>

## 1 Résumé

Les services techniques de l'Etat s'orientent de plus en plus vers des solutions automatisées pour le relevé des dégradations de surface sur les chaussées aéronautiques. L'automatisation des relevés présente en effet plusieurs avantages tant par rapport aux contraintes d'exploitation que par rapport à la qualité des données collectées. Plusieurs méthodes issues de l'Intelligence Artificielle sont susceptibles de renforcer cette automatisation. Le projet EASY-(DAI)<sup>2</sup> vise à coupler ces méthodes avec des moyens de relevé peu onéreux, pouvant en conséquence être réalisés plus fréquemment, et à évaluer expérimentalement leur apport pour la détection et la classification des dégradations sur des chaussées aéronautiques.

## 2 Contexte et objectifs du projet

### 2.1 Contexte

Avec ses 160 aéroports (civils et militaires) situés en métropole et en Outre-mer, le patrimoine aéroportuaire français est l'un des plus grands et des plus utilisés en Europe. Il représente ainsi plus de 100 millions de mètres carrés de chaussées d'aérodromes à entretenir, répartis en chaussées souples et composites (75%) auquel il faut ajouter les chaussées rigides (25%). La gestion d'un tel patrimoine tient compte du vieillissement des infrastructures mais également de la croissance continue du trafic aérien qui intègre des flottes d'avions de plus en plus agressifs en termes de charges par roue sur les trains d'atterrissage et de pressions d'impact sur l'infrastructure. Cette gestion doit s'opérer dans un contexte de développement durable de plus en plus prégnant qui impose en particulier une optimisation des travaux d'entretien et de réhabilitation.

Une évaluation régulière de l'état des infrastructures est la clé d'une bonne gestion des actifs. En apportant des éléments de compréhension sur le comportement des chaussées, elle permet en effet d'anticiper et d'optimiser les tâches d'entretien. En France, la majorité des gestionnaires de chaussées aéronautiques s'appuient sur des relevés visuels de dégradations pour estimer cet état. Ces dégradations doivent être relevées selon la doctrine technique en vigueur (cf. catalogue des dégradations de chaussées aéronautiques, 2007). On dénombre ainsi environ 20 types de dégradations pour les chaussées souples et 11 pour les chaussées rigides, chacune devant faire l'objet d'une caractérisation en termes de nombre d'occurrences, de tailles ou de densités et enfin de sévérités. À l'instar de ce qui est fait pour la surveillance des réseaux routiers, l'évaluation de l'état de la chaussée aéronautique fait intervenir un indicateur global appelé Indice de Service (IS) qui est formulé depuis l'ensemble des dégradations. Celui-ci prend la forme d'un nombre compris entre 0 et 100 sachant qu'un IS égal à 100 correspond à une piste sans aucune dégradation tandis que pour un IS égal ou inférieur à 10 la piste doit être fermée. D'autres indicateurs peuvent venir compléter l'IS ; il s'agit de l'IS dit « structurel » (ISStruct) qui est uniquement basé sur les dégradations relevant d'un défaut de la structure dans son ensemble et l'IS dit « superficiel » (ISSurf) qui ne comprend que les défauts s'apparentant à des problématiques de surface.

Dans la plupart des cas, l'IS est renseigné depuis des relevés « manuels » de dégradations réalisés par des opérateurs lors d'inspections visuelles et pédestres sur le terrain. Pour ce faire, la zone étudiée est habituellement divisée en sections unitaires de 500m<sup>2</sup> pour les chaussées souples et/ou sur chaque dalle de béton pour les chaussées rigides. Cette méthode prend beaucoup de temps et implique de lourdes contraintes opérationnelles pour l'aéroport (ex : sécurisation) ; il en résulte des relevés peu fréquents ce qui n'est pas compatible avec un objectif de surveillance plus fine de l'infrastructure. Enfin, les informations recueillies par les opérateurs étant subjectives, les relevés sont susceptibles d'offrir des degrés de répétabilité et de reproductibilité restreints. Pour toutes ces raisons, les besoins en relevés automatisés de dégradations de chaussée (collecte et analyse) sont très forts dans le contexte aéronautique.

### 2.2 Objectifs du projet

Aujourd'hui, il existe des systèmes de relevés automatisés opérationnels performants (type LCMS) qui peuvent répondre aux besoins des gestionnaires. Ces matériels peuvent opérer à grand rendement (vitesse supérieure à 70km/h) tout en délivrant des relevés très précis et fiables. Cependant, de par leur coût et leur rareté, ils ne peuvent pas être utilisés sur l'ensemble du parc d'aéroports. Ainsi, plusieurs aéroports à travers la France n'ont pas les ressources financières ou l'accès à ces systèmes en particulier ceux qui sont hors de métropole où l'acheminement de tels équipements est peu envisageable. Par conséquent, des solutions de relevés automatisés moins coûteuses et plus faciles à mobiliser constitueraient des outils intéressants pour couvrir les aéroports qui ne peuvent bénéficier des

équipements haut de gamme. Dans cette perspective, l'Université Gustave Eiffel (UGE) et le Service Technique de l'Aviation Civile (STAC) ont décidé d'étudier conjointement la faisabilité d'une solution technique basée sur une instrumentation simplifiée et bas-coût pour la collecte de données ainsi que l'utilisation de l'Intelligence Artificielle pour leur analyse. Cette solution a été développée et testée dans le cadre du projet EASY-(DAI)2 suite à l'appel de la fondation FEREC.

Le recours à une instrumentation simplifiée vise à faciliter le déploiement de la méthode sur un maximum d'aéroports. En effet, un dispositif de collecte d'images de chaussée intégrable sur tout type de véhicule de service permettrait de procéder à la fois aux inspections quotidiennes des pistes pour la détection des débris / corps étrangers (FOD) et aux relevés de dégradations de chaussée sans coûts supplémentaires pour l'aéroport et sans compliquer davantage les contraintes opérationnelles actuelles.

Concernant l'étape de traitement et d'analyse des données, le projet EASY-(DAI)2 fait intervenir les techniques d'apprentissage profond (Deep Learning) pour la reconnaissance des dégradations de chaussée parmi les jeux d'images enregistrés. Ces techniques constituent des opportunités pour ce genre d'applications et sont donc de plus en plus rencontrées notamment dans le domaine routier. Par conséquent, ce projet s'intéresse également à la transférabilité de ces outils entre le contexte routier et le contexte aéronautique sachant que la plupart des dégradations à détecter sont similaires. Une application réalisée en vraie grandeur et une évaluation de la solution développée dans ce projet pour quelques types de dégradations constituent des éléments de sortie de l'étude.

### 3 Les solutions techniques proposées

Les relevés d'auscultation doivent pouvoir être réalisés par des opérateurs non-spécialistes (typiquement des agents de l'aéroport) au moyen de matériels facilement intégrables sur des véhicules « standard » et économiques pour une diffusion plus large. Pour satisfaire aux contraintes de simplicité, de rapidité de montage et de coût de l'instrumentation, nous nous sommes orientés vers un dispositif de vision mono-caméra pour la prise d'images de la chaussée couplé avec un récepteur GPS pour le recueil des coordonnées géographiques et le calcul de l'abscisse curviligne. Concernant le traitement des données, les techniques par IA ont été privilégiées car elles offrent un potentiel très intéressant en particulier quand la qualité des données entrantes n'est pas optimale. Les paragraphes suivants décrivent les instrumentations retenues pour ce projet incluant les parties matérielles et logicielles ainsi que les méthodes de traitement développées pour la détection des fissures.

#### 3.1 Le mode de collecte de données

Le mode prise de vue un élément important car il conditionne en particulier le choix de l'imageur. Pour ce projet, nous avons retenu une prise de vue « top-down », c'est-à-dire une caméra orientée perpendiculairement à la surface de la chaussée. Cette configuration permet d'enregistrer une image de surface de chaussée qui soit homogène en terme de résolution ce qui signifie que chaque pixel dans l'image code pour une même distance (sous réserve de ne pas avoir de distorsion géométrique significative).

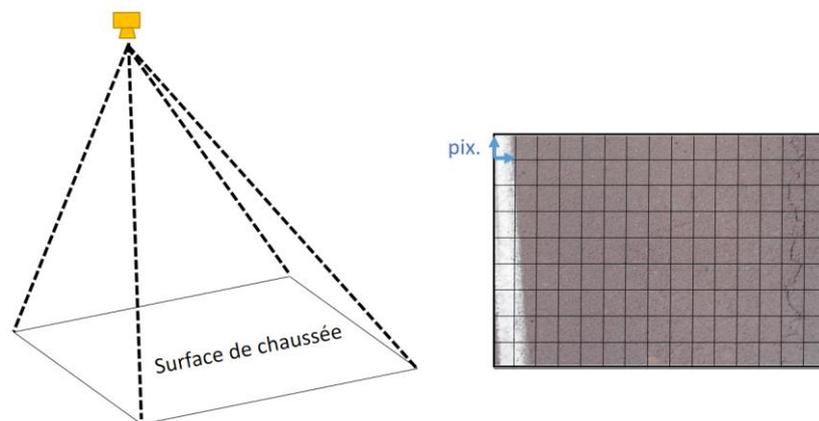


Figure 1. Schéma correspondant au mode de prise de vue

Ce mode de prise de vue conduit à l'obtention d'une surface au sol relativement faible et implique de procéder à un balayage « image par image » précis de la piste lors de la collecte. Dans l'idéal, un tel balayage devrait permettre d'enregistrer des images de sorte à couvrir l'intégralité de la piste. Dans la pratique, on se heurte à plusieurs difficultés :

- Si des images voisines « se chevauchent », une même dégradation sera sur-comptabilisée dans le relevé (sauf à disposer d'un outil permettant d'identifier les parties d'images en double et de les éliminer) ;
- Si des images voisines sont trop espacées alors le relevé sera incomplet ;
- Il n'y a que très peu d'amers visuels sur une piste (ex : peu de marquage au sol) ce qui rend difficile le maintien de trajectoire par le conducteur ;
- Le nombre d'images enregistrées peut devenir important (gestion du stockage des images).



Figure 2. Schéma illustrant le balayage à réaliser pour la collecte d'images

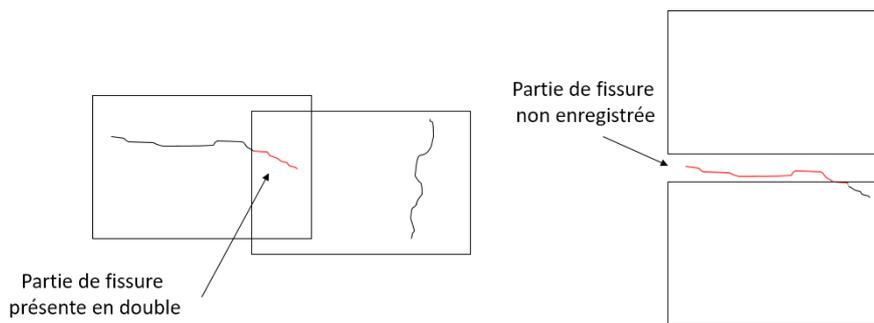


Figure 3. Schéma illustrant les problématiques de recellement des images voisines

Le balayage consécutif au relevé d'images constitue en soi une problématique d'aide au positionnement et au guidage du véhicule qui n'est pas l'objet du projet. Il est d'ailleurs également rencontré pour l'utilisation d'autres systèmes de mesures y compris les plus perfectionnés (ex : LCMS). Toutefois, l'UGE et le STAC ont tenté de limiter le plus possible les difficultés liées au balayage en travaillant sur les aspects suivants pendant le projet :

- Une fonctionnalité logicielle dans le programme d'acquisition qui permet de collecter les images selon un intervalle de distance paramétrable qui tient compte de la vitesse du véhicule en temps réel (ex : 1 image tous les 2m) ;
- L'interfaçage du logiciel avec un récepteur GNSS de plus haut niveau (mais de coût plus élevé) capable de disposer de corrections différentielles pour tendre vers l'obtention d'une position décimétrique.
- Le développement d'un dispositif d'aide à la conduite basé sur une centrale inertielle 9-axes qui affiche un cap à suivre en temps réel au conducteur.

Ce rapport ne traite pas de ces derniers points car ils ne constituent pas le cœur du sujet. Néanmoins, il est clair que des outils techniques et méthodologiques sont nécessaires à une collecte efficace et précise des images de chaussée sur les pistes aéronautiques. Le passage à une application pleinement opérationnelle impliquera donc des développements complémentaires.

Pour finir, il existe d'autres modes possibles de collecte comme par exemple celui reposant sur des images de type panoramiques. Dans ce dernier cas, des étapes d'étalonnages systématiques avant relevés sont bien souvent nécessaires et la gestion des zones de chevauchement entre images semble plus difficile (d'où le fait que cette hypothèse de travail n'ait pas été retenue dans ce projet).

### 3.2 L'instrumentation

L'instrumentation se compose d'une partie matérielle et d'une partie logicielle. Le composant principal de la partie matérielle est une caméra. Dans le cadre du projet EASY-(DAI)<sup>2</sup>, plusieurs catégories de caméras ont été considérées en début d'étude pour couvrir le maximum d'offres possibles :

- Une catégorie « mini-caméra » (webcam, caméras GoPro avec différents types d'objectifs, autres...) ;
- Une catégorie « caméra industrielle » ;

Une catégorie « smartphone » avait initialement été ciblée pour des relevés en mode panoramique. En raison d'un choix différent concernant la prise de vue et compte-tenu du fait qu'il aurait fallu adapter ou redévelopper des modèles de détection de dégradation à cette configuration, cette catégorie d'appareils n'a donc pas été retenue pour la suite du projet.

Après plusieurs essais en laboratoire (Cf. Section 3.1), la catégorie caméra industrielle a été retenue pour les expérimentations *in situ* car elle offre le meilleur compromis entre le coût (env. 2000 euros ce qui est acceptable au regard du cadre d'application), la facilité d'intégration, la robustesse et la qualité des images fournies.

Le second composant matériel est un récepteur GNSS. Ce dernier sert à la fois à la géolocalisation des images mais participe également à l'opération d'échantillonnage des images en se substituant à un codeur incrémental de distance (comme ceux qui doivent nécessiter un montage spécifique sur une des roues du véhicule). Ce matériel peut être de bas coût (mais offrant a minima un accès aux corrections SBAS) ou plus sophistiqué (avec corrections différentielles pour un gain en précision), l'essentiel étant qu'il délivre des trames NMEA lisibles par le logiciel d'acquisition.

La partie logicielle de l'instrumentation correspond au programme servant à l'acquisition et au stockage des images. Il s'agit d'un logiciel développé par l'UGE qui peut s'interfacer avec plusieurs types de caméras et de récepteurs GNSS. L'IHM de ce logiciel, réactive et volontairement épurée pour limiter les temps d'intervention, permet de procéder à quelques réglages de la caméra et d'activer / désactiver l'enregistrement des images.

L'ensemble est fixé sur un véhicule porteur tel qu'illustré sur la Figure 4. La hauteur du dispositif supportant la caméra est relativement grande (~ 2.5m) afin de couvrir une plus grande surface au sol et ainsi limiter les aller-retours lors du balayage de la piste.



Figure 4. Vue de l'instrumentation montée sur le véhicule porteur et utilisée pour la prise d'images

### 3.3 Le traitement et l'analyse des images par Deep Learning

#### 3.3.1 Démarche générale

Le laboratoire LAMES de l'Université Gustave Eiffel avait commencé à développer avant le projet EASY-(DAI)2 des modèles de reconnaissance et de classification d'images comportant des dégradations de chaussée à des fins d'auscultation des réseaux routiers secondaires. Certains de ces modèles étaient issus d'opérations traditionnelles du traitement d'images tandis que d'autres, plus récents, provenaient des techniques d'intelligence artificielle et plus précisément du Deep Learning (ou apprentissage profond). Dans ce contexte, un des objectifs poursuivis par le projet sur le plan de l'analyse des données est d'étudier la transférabilité des modèles de Deep Learning entre une application routière et une application aéronautique. En d'autres termes, on cherche à savoir si des modèles qui ont été entraînés sur des images de chaussée routière (relevant de réseaux secondaires) et de qualité très modeste (issues de webcam) peuvent être directement déployés pour le cas des chaussées aéronautiques souples. Un autre objectif est de savoir si ces mêmes modèles peuvent être renforcés ou adaptés par rapport à de nouveaux jeux de données d'entraînement (en particulier en cas de changement de caméra et/ou par rapport à de nouvelles dégradations) mais aussi par rapport à des besoins en précision et fiabilité plus importants. Le cas des chaussées aéronautiques rigides (en dalles de béton) est également pris en charge dans ce dernier objectif.

Pour traiter ces objectifs, le projet s'est organisé sous formes de plusieurs enchainements successifs de (ré)entraînement de modèles - expérimentations - évaluations de sorte à tendre peu à peu vers un ou plusieurs modèles suffisamment robuste(s) et adapté(s) au cas des chaussées aéronautiques qu'elles soient souples ou rigides.

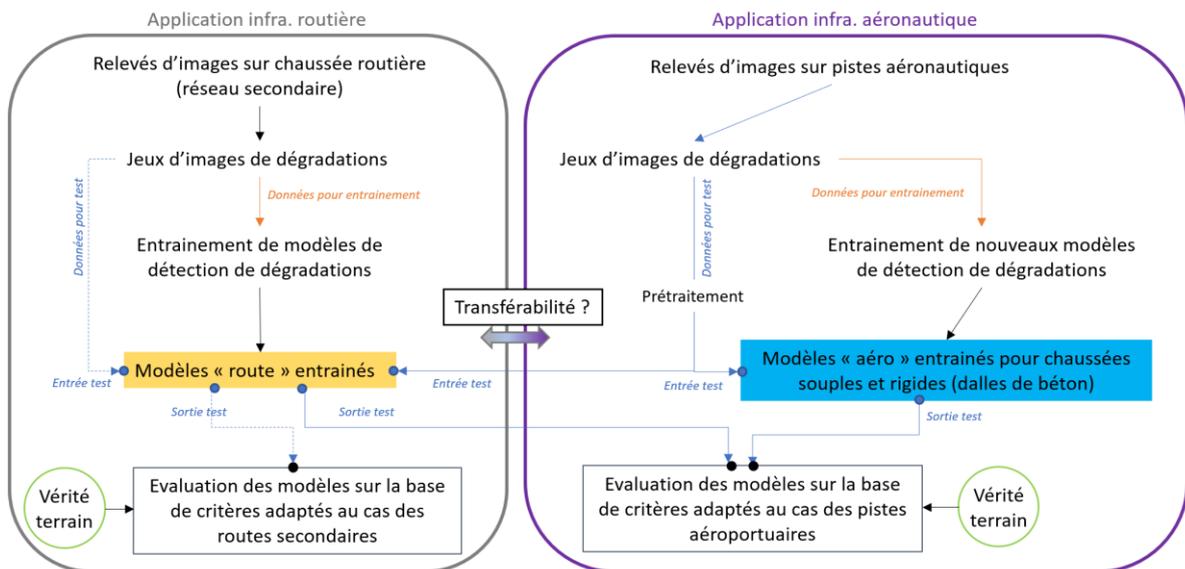


Figure 5. Interactions entre les applications routières et aéronautiques dans le projet EASY-DAI2 pour le traitement des données

### 3.3.2 Le prétraitement

Afin de ne pas « manquer » des dégradations lors de la collecte d'images, il a été décidé de générer volontairement des chevauchements entre les images pour ces expérimentations. Comme expliqué précédemment, ces chevauchements impliquent un risque de surestimation du nombre et de la longueur des fissures détectées. On distingue deux types de chevauchement :

- Le chevauchement longitudinal (dans le sens d'avancement du véhicule) dont l'ampleur est dépendante de la capacité du logiciel d'acquisition à prélever une image à intervalle de distance fixe ;
- Le chevauchement latéral qui peut être contrôlé en positionnant correctement le véhicule sur la ligne de début de relevé (véhicule à l'arrêt) mais dont le niveau peut varier par la suite en fonction de la capacité du véhicule à conserver sa trajectoire.

Pour estimer le chevauchement longitudinal, on calcule *a posteriori* la distance réelle entre deux images successives sur la base des coordonnées GPS correspondantes (durant la phase d'acquisition, seule une estimation de cette distance est considérée) ; cette distance inter-images est susceptible de varier plus ou moins fortement en particulier pendant les phases d'accélération/décélération du véhicule porteur. Le chevauchement latéral est quant à lui plus direct puisqu'on considère une zone de recouvrement fixe sur toute la progression du véhicule dans son couloir (hypothèse de travail retenue dans cette étude).

La phase de prétraitement consiste donc à rogner, pour chaque image et en fonction des distances évoquées précédemment, les régions qui sont communes à d'autres de sorte à éliminer les zones de chevauchement générées lors de l'acquisition.

À l'avenir, la gestion des chevauchements pourrait être améliorée en faisant évoluer l'instrumentation (ex : travail avec un codeur de roue et/ou avec un GPS plus précis avec une fréquence supérieure à 1Hz). Néanmoins, une telle évolution risque d'être pénalisante par rapport aux contraintes de simplification et de minimisation des coûts du dispositif.

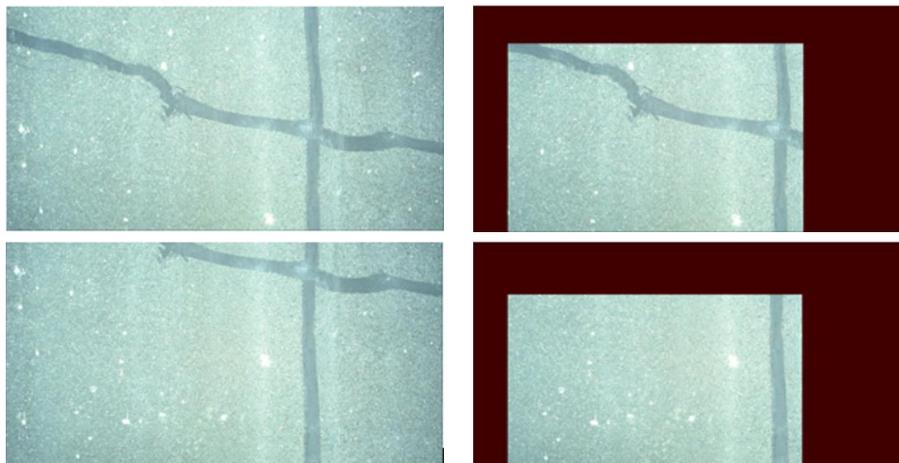


Figure 6. Illustration du découpage des images pour tenir compte des zones de chevauchement

### 3.3.3 Les classes d'objets à détecter

Le projet EASY-(DAI)2 s'intéresse à un nombre limité de classes d'objets qui rassemble plusieurs types de dégradations de surface. Les classes d'objet ont été choisies par rapport aux besoins principaux liés au contexte de l'étude, aux dégradations présentes sur les différents sites expérimentaux ainsi qu'au temps imparti pour réaliser ce projet. Un certain nombre de dégradations a donc été retenu sachant que l'après-projet pourra être l'occasion d'enrichir le nombre de classes.

Pour les chaussées « souples » aéronautiques (en enrobé bitumineux), les classes d'objets à détecter relevant des dégradations sont :

- La fissuration longitudinale (désignée par LC pour « Longitudinal Cracking » dans l'étude) ;
- La fissuration transversale (désignée par TC pour « Transversal Cracking » dans l'étude) ;
- Le faïençage (désigné par BC pour « Bloc Cracking » dans l'étude).

Pour les besoins d'une expérimentation, les classes suivantes ont été ajoutées :

- La fissuration longitudinale présente dans une réparation (désignée par RLC dans l'étude) ;
- La fissuration transversale présente dans une réparation (désignée par RTC dans l'étude).

Pour les chaussées « rigides » en dalles de béton, les classes d'objets correspondant aux dégradations sont :

- La fissuration longitudinale
- La fissuration transversale

Les classes d'objets suivantes ont été ajoutées de sorte à pouvoir localiser si besoin les dégradations par rapport à la géométrie de la dalle de béton

- Joint de dalle
- Coin de dalle

Dans tous les cas de figure, une détection implique de savoir classifier et localiser une ou plusieurs classes d'objet dans chacune des images du jeu de données.

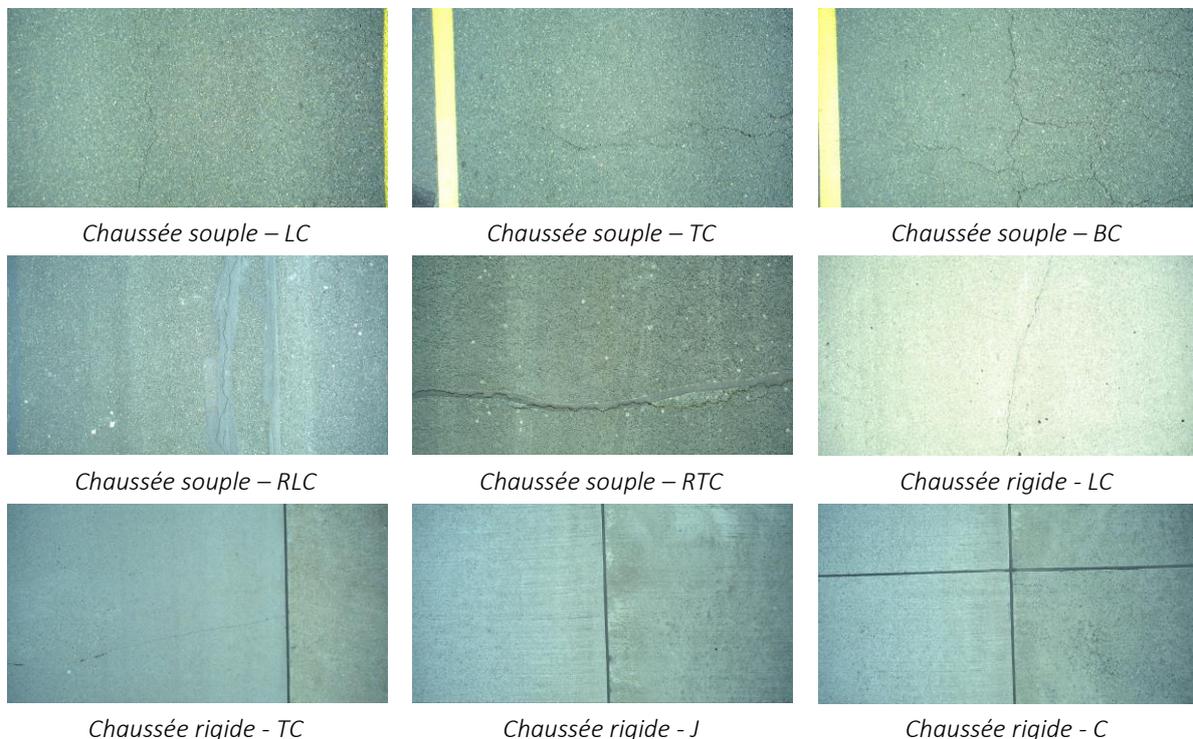


Figure 7. Illustrations des classes d'objets à détecter

### 3.3.4 Processus utilisé pour la génération des modèles de Deep Learning

Dans le cadre de travaux antérieurs, une première génération de modèles de détection et de reconnaissance de dégradations avait été développée par l'UGE pour le cas des réseaux routiers de faible trafic (routes secondaires). Ces modèles, construits en utilisant l'outil d'apprentissage TensorFlow, résultent d'une opération dite de « Transfert Learning » ; cette dernière consiste à utiliser des modèles existants « pré-entraînés » et en accès libre, puis à réorienter l'apprentissage vers de nouvelles cibles (*i.e.* des dégradations de chaussées dans le contexte de l'étude). L'architecture des modèles de Deep Learning se présentant sous la forme d'empilement de couches de neurones (ou réseau de neurones), le transfert de connaissance peut être perçu comme la réutilisation des réseaux des modèles pré-entraînés et l'ajout d'une ou plusieurs couches finales qui se « spécialisent » par rapport à l'extraction de nouvelles caractéristiques pertinentes.

Le développement des modèles de DL a été réalisé selon le processus illustré dans la Figure 8. Le prérequis au déroulement de ce processus est de disposer d'un jeu d'images suffisamment riche aussi bien sur le plan quantitatif (nombre d'images) que qualitatif (panel d'images contenant les cibles à identifier). Depuis le jeu de données, le processus se déroule alors selon deux grandes étapes :

- i) Une phase préparatoire
- ii) Une phase d'entraînement

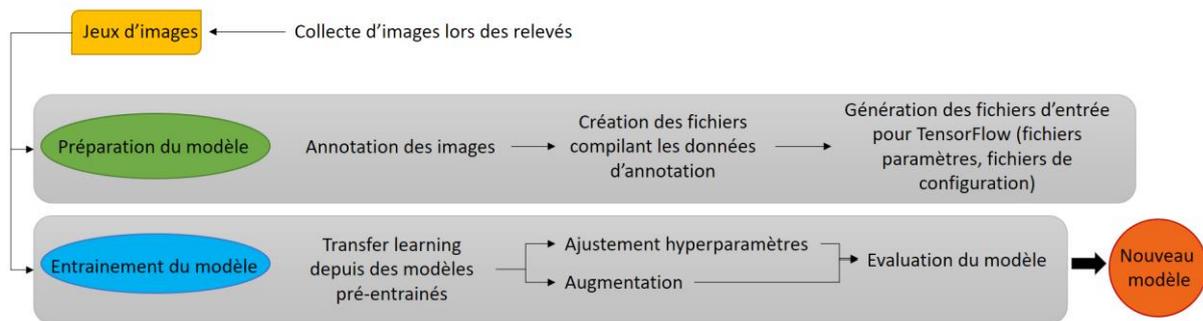


Figure 8. Schéma du processus suivi pour la création d'un modèle de deep learning

La phase préparatoire inclut le travail d'annotation (ou d'étiquetage) des images ; il consiste à identifier « manuellement » des dégradations parmi un ensemble d'images pour permettre aux modèles de retrouver, par eux-mêmes, des caractéristiques en vue de l'apprentissage. Cet exercice est également nécessaire pour le travail d'évaluation afin de savoir si la détection et/ou la classification proposée par un modèle entraîné est correcte ou pas. La phase préparatoire inclut également la définition d'un certain nombre de paramètres qui sont nécessaires au contrôle du processus d'entraînement (notamment les hyperparamètres).

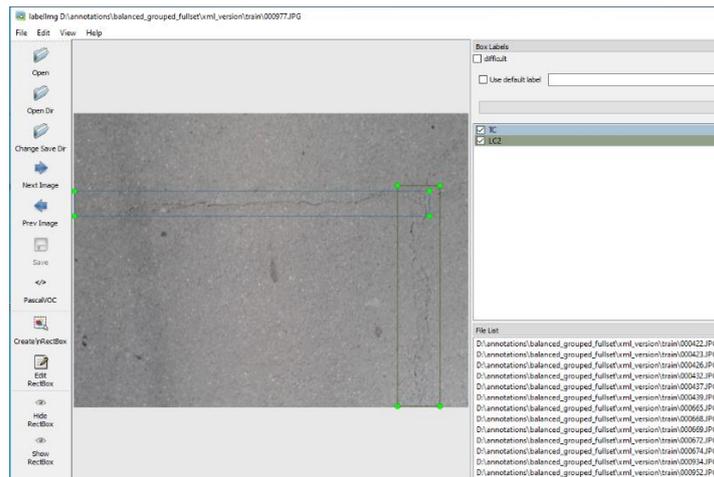


Figure 9. Interface utilisée pour l'annotation des images d'entraînement

Pendant la phase d'entraînement, les modèles vont chercher à mettre en place un raisonnement qui, à partir de données d'entrée annotées, va conduire à la reconnaissance et à la classification automatique des dégradations de chaussée. Ce travail passe par l'extraction de caractéristiques pouvant être plus ou moins abstraites. Dans le cas du transfert de connaissance, la phase d'entraînement se traduit par l'ajout d'une ou plusieurs couches dans le réseau de neurones, ce qui différencie le nouveau modèle du modèle initial.

Dans ce projet, la phase d'entraînement a été complétée avec deux étapes dans le but d'augmenter les performances des modèles : un ajustement des hyperparamètres et l'utilisation de la méthode d'augmentation. La première consiste à optimiser certains paramètres du modèle de sorte à ce que celui-ci s'améliore un peu plus dans ses prédictions par rapport aux données d'entrée (mais aussi à ce qu'il soit plus rapide) ; la seconde permet de générer de nouvelles images pour l'entraînement par des modifications aléatoires sur les images d'entrée (par rapport à la luminosité, au contraste, niveau de zoom, etc.).

À la fin de cette phase, on dispose de modèles entraînés qui doivent montrer de bonnes capacités de généralisation, c'est-à-dire être capables de faire des prédictions sur des images jamais rencontrées lors de l'entraînement. Pour cela, on découpe le jeu d'images issu de relevés *in situ* en plusieurs sous-ensembles dont l'un sert uniquement à apprécier le niveau de généralisation (sous-ensemble représentant environ 10% des images), le reste étant alloué à l'entraînement.

Dans le cadre du projet, le processus de traitement des données a été effectué au moyen d'un ordinateur « standard » interfacé avec Windows 10 et s'appuyant sur une carte graphique NVIDIA Quadro P4000 (RAM de 8 Go) et d'un microprocesseur 64 bits @ 3.7 GHz avec une mémoire de 32 Go. Avec cet équipement et à titre indicatif, il faut environ 1 jour pour entraîner un modèle depuis les données labellisées et moins de 2 secondes pour fournir une image associée à une boîte de détection.

### 3.3.5 Les modèles de Deep Learning utilisés

Le projet EASY-(DAI)2 s'est appuyé sur des travaux antérieurs réalisés par l'UGE qui avaient débouché sur la production de modèles de DL entraînés sur une base d'environ 6000 images en couleur. Ces modèles avaient été développés dans le cadre de relevés de dégradations sur les réseaux secondaires ; la collecte des images avait été réalisée par le biais de webcams montées sur des véhicules traceurs. Malgré la qualité relativement faible des images, une première évaluation de ces modèles avait montré des aptitudes intéressantes pour la détection et la classification d'un certain nombre de dégradations (celles qui sont les plus pertinentes à relever pour ce type de réseau).

Les modèles ayant offert les meilleurs résultats proviennent de méthodes de détection basées sur des réseaux de neurones convolutifs et en particulier la méthode dite de « Faster R-CNN ». Chaque type de

modèle se distingue des autres par une architecture qui lui est propre ; cette architecture conditionne la façon dont les caractéristiques sont recherchées dans les images ainsi que la classification des objets détectés. Par exemple, pour la famille des « R-CNN » (Regional Convolutional Neural Network, le modèle va d'abord proposer un ensemble de régions dans l'image (ou zones d'intérêt) susceptibles de contenir la cible par l'utilisation d'un réseau de neurones convolutifs. Chacune de ces régions se voit pourvu par la suite d'un ensemble de « boites de détection » (ou « bounding box » en anglais) de taille et de ratio différents. L'ensemble est alors entré dans ce même réseau de neurones convolutifs qui délivre en sortie une information finale contenant la présence ou pas de l'objet (ou des objets) recherché(s), une bounding box associée (en cas de détection) et la classe de l'objet (ex : fissure, faïençage, etc.). Ce traitement de l'image par morceaux permet de détecter plusieurs objets d'une même classe ou de différentes classes sur une même image. Par rapport à d'autres architectures, la méthode des R-CNN peut s'avérer plus lente (même si la modélisation en Faster-RCNN offre un gain en temps de traitement) notamment vis-à-vis des méthodes « Single Shot multi-box Detector » (SSD) par exemple ; mais ce facteur n'est pas gênant car le traitement des données dans le projet EASY-(DAI)2 s'inscrit dans un travail en temps différé (pas de besoin en temps réel).

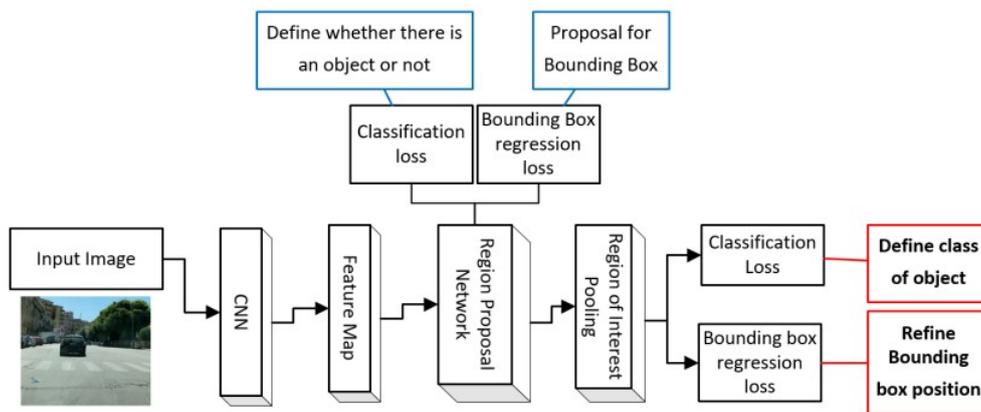


Figure 10. Schéma décrivant le fonctionnement d'un modèle de type R-CNN

## 4 Applications expérimentales et évaluation des résultats

### 4.1 Tests en laboratoire

Comme les expérimentations sur les pistes aéroportuaires présentées par la suite impliquent de travailler en un temps limité et sous contraintes de trafic aérien, les premiers essais se sont déroulés en « laboratoire ». Un tel cadre permet de travailler dans des conditions plus confortables pour tester un ensemble de dispositifs de prises de vue et de configurations de mesure (caméras, objectifs, hauteur d'installation, conditions de luminosité, etc.) en statique, à faible vitesse ou à vitesse plus élevée (sur route). Ces essais en laboratoire ont également permis de mettre à l'épreuve un certain nombre d'algorithmes de traitement des images dont plusieurs modèles de deep learning. À l'issue de ces tests, une instrumentation et quelques modèles de traitement de données ont ainsi pu être retenus pour la suite du projet (cf. sections 3.4 et 3.5). Le tableau ci-après récapitule les matériels de prises de vue testés pendant les essais en laboratoire.

Type de caméra	Avantages	Inconvénients
Caméra industrielle – obj. 12mm	Robustesse, possibilité d'interfaçage, nombreuses possibilités de réglages (temps d'exposition, netteté, etc.), facilité d'intégration sur véhicule.	Coût, temps de développement pour l'interfaçage informatique, poids importants des images brutes en sortie de la caméra.
Caméra GoPro	Robustesse, dispositif gyrostabilisé, grande fréquence d'enregistrement, facilité d'intégration sur véhicule.	Interfaçage informatique très limité.
Caméra GoPro / Mode grand angle	Robustesse, dispositif gyrostabilisé, facilité d'intégration sur véhicule, grande fréquence d'enregistrement, surface de l'image au sol.	Interfaçage informatique très limité, légère déformation de l'image (à corriger).
Caméra GoPro / Mode très grand angle	Robustesse, dispositif gyrostabilisé, facilité d'intégration sur véhicule, grande fréquence d'enregistrement, surface de l'image au sol.	Interfaçage informatique très limité, déformation de l'image (à corriger).
Caméra Osmo Pocket	Coût, dispositif gyrostabilisé, couplage avec smartphone.	Interfaçage informatique très limité, intégration sur véhicule.
Caméra Osmo x5 – obj. 15mm	Dispositif gyrostabilisé, couplage avec smartphone, compatibilité drone.	Coût, interfaçage informatique très limité, intégration sur véhicule.
Caméra fisheye (360°)	Surface de l'image au sol, facilité d'intégration sur véhicule.	Très grande déformation de l'image (nécessite une forte correction), recalage des images.
Webcam Logitech	Coût, facilité d'intégration sur véhicule.	Netteté des images (flou de bougé).

Tableau 1. Liste des caméras testées pendant le projet et description des atouts et inconvénients associés

En complément des tests portant sur le type de caméras, des prises de vues ont été réalisées à différentes hauteurs (de 2m à 4m) afin de trouver le meilleur compromis entre possibilité d'instrumentation, couverture de la zone à étudier pour faciliter le balayage sur la piste et résolution de l'image au sol. Ce travail a permis de simuler l'usage d'un drone si, dans le futur, ce dispositif est utilisé

pour la prise d'images (mais pour le moment trop d'incertitudes techniques et réglementaires entourent l'utilisation d'un tel moyen).

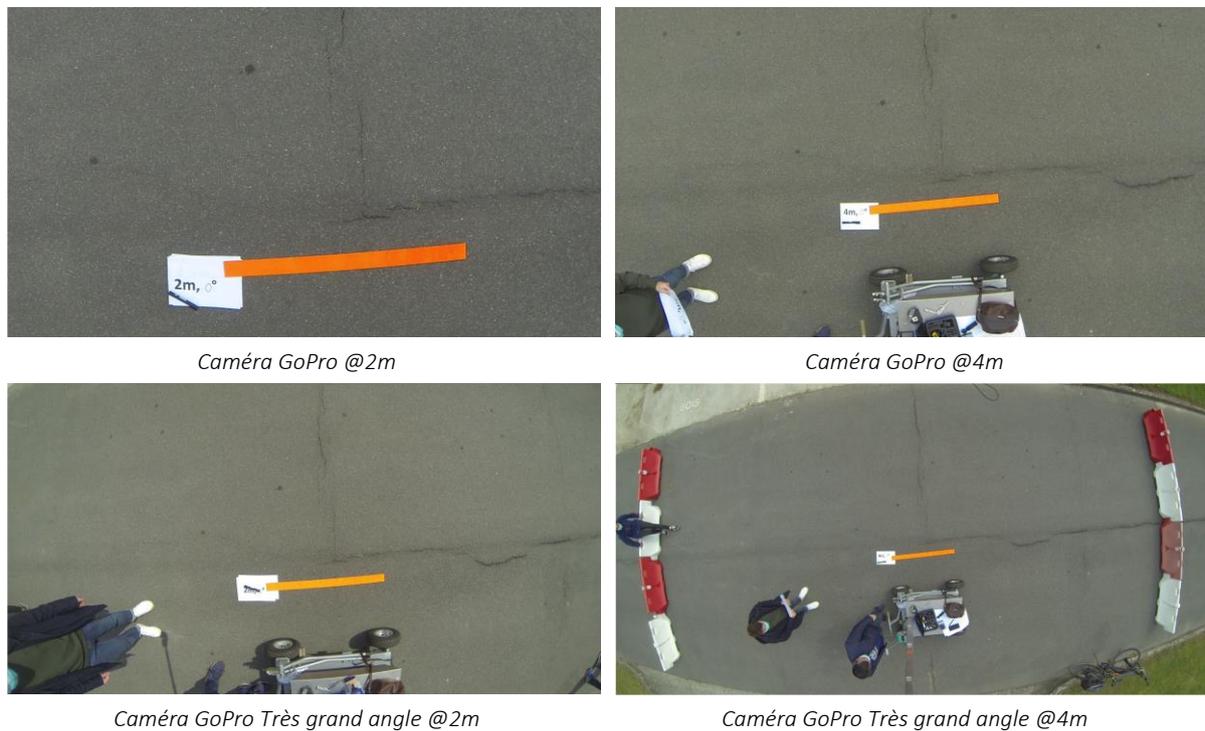


Figure 11. Exemples de prises de vues réalisées à différentes hauteurs et pour 2 types de caméras

L'instrumentation finale s'est arrêtée sur l'utilisation d'une caméra industrielle montée sur un véhicule d'essai à 2.5m de hauteur. Dans cette configuration, le véhicule peut évoluer jusqu'à 70 km/h tout en fournissant des images de bonne qualité, ces dernières correspondant à des images de chaussées d'environ 3m de large (direction transverse) et 1.6m de long.

## 4.2 Constitution de la vérité terrain

Afin d'estimer qualitativement et quantitativement les relevés de dégradations fournis par le dispositif expérimental associés aux modèles de deep learning, il est nécessaire de disposer d'une vérité terrain constituée d'informations de référence. Pour ce projet, ces dernières ont été obtenues en soumettant les images de chaussées « brutes » à plusieurs contributeurs familiers de ce genre de relevés de sorte à ce qu'un listing des dégradations soit établi pour chacune des images des jeux de données. Cette vérité terrain est donc basée sur l'examen d'images et non pas sur un relevé pédestre *in situ* ni sur un relevé issu d'un appareil de référence. En conséquence, il est possible par cette méthode de confirmer / infirmer la présence d'une dégradation ou de plusieurs dégradations dans l'image, de déterminer sa classe et d'estimer sa longueur (dans l'image et dans le système métrique). En revanche, cette vérité terrain n'intègre évidemment pas les dégradations qui n'auraient pas été filmées ou auraient été supprimées lors du prétraitement.

### 4.3 Critères utilisés pour caractériser le niveau de performance des résultats

L'évaluation de la qualité des résultats obtenus par l'utilisation conjointe du dispositif expérimental et des modèles de deep learning a principalement été effectuée selon l'examen de :

- La matrice de confusion ;
- L'indicateur « F1-score » ;
- Les longueurs estimées des dégradations.

La matrice de confusion met en relation la prédiction (ou détection) du modèle et la vérité terrain par le recensement de plusieurs sous-ensembles : les « vrais positifs », les « faux positifs », les « faux négatifs » et les « vrais négatifs ». Le F1-score résulte de l'exploitation de la matrice de confusion ; il est couramment utilisé pour représenter le niveau de performance des classifieurs. Le F1-score est souvent assorti des critères dits de précision et de rappel qui sont définis dans le tableau ci-après.

		Prédiction (détection)	
		Positif	Négatif
Vérité	Vrai	Vrai Positif (VP)	Vrai Négatif (VN)
	Faux	Faux Positif (FP)	Faux Négatif (FN)
Degré de Précision (Taux de détections positives exactes)		$P = TP/(TP+FP)$	
Degré de Rappel (Taux de « vrais » cas correctement détectés)		$R = TP/(TP+FN)$	
F1-score		$F1 = 2/(P^{-1} + R^{-1}) = TP/(TP+(FP+FN)/2)$	

Tableau 2. Matrice de confusion et critères utilisés pour le calcul du F1-score

Le F1-score est une sorte de compromis entre le degré de précision et le degré de rappel ; plus le modèle performe, plus le F1-score tend vers 1. Le F1-score s'intéresse particulièrement aux groupes correspondant à des bonnes ou des mauvaises détections ; il ne tient pas compte du groupe des « vrais négatifs » (pour lequel les modèles ne détectent pas de dégradations sur les images et qu'il n'y en a effectivement pas). L'exploitation de ce dernier groupe peut être réalisé à travers le degré de spécificité ( $VN/(VN+FP)$ ) qui caractérise le niveau de performance des modèles face à des images dépourvues de dégradations.

Un autre critère permettant d'évaluer la performance des modèles de deep learning est l'estimation des longueurs des dégradations. Celle-ci résulte directement du rectangle de détection (appelé aussi boîte de détection ou « bounding box » en anglais) qui est créé par le modèle autour de la dégradation. En plus d'être capable de détecter (reconnaître et classer) une dégradation, la puissance des modèles de deep learning peut également s'apprécier par rapport à la délimitation géographique de la cible dans l'image qui est ensuite convertie en longueur de dégradations dans le système métrique. Ainsi, pour chaque détection de dégradation, une longueur (en mètre) a été calculée en l'assimilant au plus grand côté de la boîte de détection.

#### 4.4 Expérimentation sur le site de Lognes

En novembre 2021, la solution technique décrite dans le chapitre précédent a été expérimentée et testée sur l'aérodrome de Lognes-Emerainville en Seine-et-Marne. Cette infrastructure, gérée par Aéroport de Paris, est le premier aérodrome de France pour l'aviation civile légère privée et est ouvert au trafic international. La chaussée aéroportuaire sur ce site est souple (en enrobé bitumineux) et a subi de nombreuses réparations (pontages de fissures). Pour une première expérimentation *in situ* en conditions réelles de mesure, la piste a pu être fermée pendant plusieurs heures consécutives sans que cela ne pénalise significativement le trafic aérien. En outre, tout véhicule d'essai pouvait intervenir sur la piste ce qui a permis de travailler avec plusieurs véhicules.



Figure 12. Photos de l'expérience menée à Lognes

Sur le site de Lognes, les relevés ont été effectués sur 4 « bandes » (ou couloirs) d'une longueur de 550m chacune environ, une bande étant décalée de sa voisine d'une distance de 2m dans le sens transversal. Cela correspond au balayage de la moitié de la piste (celle qui contenait le plus de dégradations *a priori*) selon le sens 08/26 (orientation Ouest-Est). Les passages ont été réalisés à différentes vitesses (de 7 à 70 km/h) afin de tester le logiciel d'acquisition.



Figure 13. Vue du site de Lognes et représentation de la zone étudiée

Suite aux relevés, environ 1800 images de chaussée ont été collectées pour chaque gamme de vitesse. Ces images ont été soumises à différents modèles de deep learning en vue du traitement. Grâce à cette première expérimentation, plusieurs enseignements ont pu être tirés :

- Plusieurs modèles de deep learning ont été écartés suite à un trop faible nombre de détections par rapport à la vérité terrain.
- Il a été constaté des limites de la part du système d'acquisition vis-à-vis du séquençage des images à pas fixe pour les passages effectués à 50 et 70 km/h, (sur la base de l'exploitation des données GPS à 1Hz).
- La piste de Lognes contenant de nombreuses réparations, il a été décidé, après cette expérimentation, d'ajouter des classes relatives à la présence de dégradations sur des

réparations en place, ce cas pouvant être apprécié différemment selon les méthodes d'exploitation (ce travaillant impliquant un réentraînement des modèles).

Compte-tenu de ces remarques, les résultats ci-après portent sur le modèle de deep learning qui a le mieux performé (à savoir le Faster R-CNN après optimisation des hyperparamètres) sur des images issues de la caméra industrielle pour une vitesse de relevés de 30 km/h et en considérant les classes LC et TC sur chaussée aéronautique souple (la classe BC faisait également partie des objets à détecter mais comme la piste de Lognes ne présentait pas ce type de dégradation, elle n'apparaît pas dans les résultats).

		VP	FP	FN	Precision (%)	Rappel (%)	F1-score (%)
<b>Bande 1</b>	LC	236	59	99	80.00%	70.45%	75%
	TC	83	9	68	90.22%	54.97%	68%
	<b>Total Fissures</b>	319	68	167	82.43%	65.64%	73%
<b>Bande 2</b>	LC	93	18	47	83.78%	66.43%	74%
	TC	91	12	131	88.35%	40.99%	56%
	<b>Total Fissures</b>	184	30	178	85.98%	50.83%	64%
<b>Bande 3</b>	LC	215	41	104	83.98%	67.40%	75%
	TC	67	8	140	89.33%	32.37%	48%
	<b>Total Fissures</b>	282	49	244	85.20%	53.61%	66%
<b>Bande 4</b>	LC	11	11	20	50.00%	35.48%	42%
	TC	53	8	151	86.89%	25.98%	40%
	<b>Total Fissures</b>	64	19	171	77.11%	27.23%	40%
<b>Total toutes bandes</b>	LC	555	129	270	81.14%	67.27%	74%
	TC	294	37	490	88.82%	37.50%	53%
	<b>Total Fissures</b>	849	166	760	83.65%	52.77%	65%

Tableau 3. Performance du modèle de DL sur la métrique F1-score et autres critères (Lognes)

Le tableau précédent fait apparaître plusieurs observations qui sont commentées ci-après :

Le niveau de performance du modèle de DL pour la bande n°4 diffère significativement des autres selon le F1-score. Quand on connaît la configuration des lieux, la bande n°4 a en effet la particularité d'offrir un peu moins de dégradations que les autres d'une part et ces dégradations sont davantage situées dans des réparations (pontages) d'autre part. En conséquence, le modèle présente un F1-score faible (~ 40%) pour cette bande notamment à cause d'un nombre important de Faux Négatifs (FN) pour la classe TC ; le modèle a donc plutôt tendance à manquer des détections pour cette bande.

Pour les bandes n°1 à 3, le F1-score est bon pour la détection des fissures longitudinales (LC) puisqu'il est en général proche de 75%. Il est en revanche moins élevé pour la détection des fissures transversales (de 48% à 68%) sans qu'il n'y ait une explication particulière à avancer sachant que lors de l'entraînement des modèles, la précaution a été prise d'équilibrer les classes LC et TC pour l'apprentissage de sorte à ne pas renforcer / favoriser l'une ou l'autre de ces classes.

Par rapport à des résultats antérieurs obtenus sur route et avec le même modèle de DL, les F1-scores résultant de cette expérimentation sont moins bons. Cela peut s'expliquer par le caractère difficile des images à traiter par rapport à la présence très récurrente de pontages, ces derniers portant par ailleurs la plupart des fissures. Comme le modèle n'a pas été entraîné sur de tels cas, son niveau de performance est tout naturellement altéré. Afin d'illustrer les résultats obtenus, quelques images en sortie de modèle de DL sont présentées sur la Figure 14. D'autres illustrations sont fournies dans les annexes du livrable.

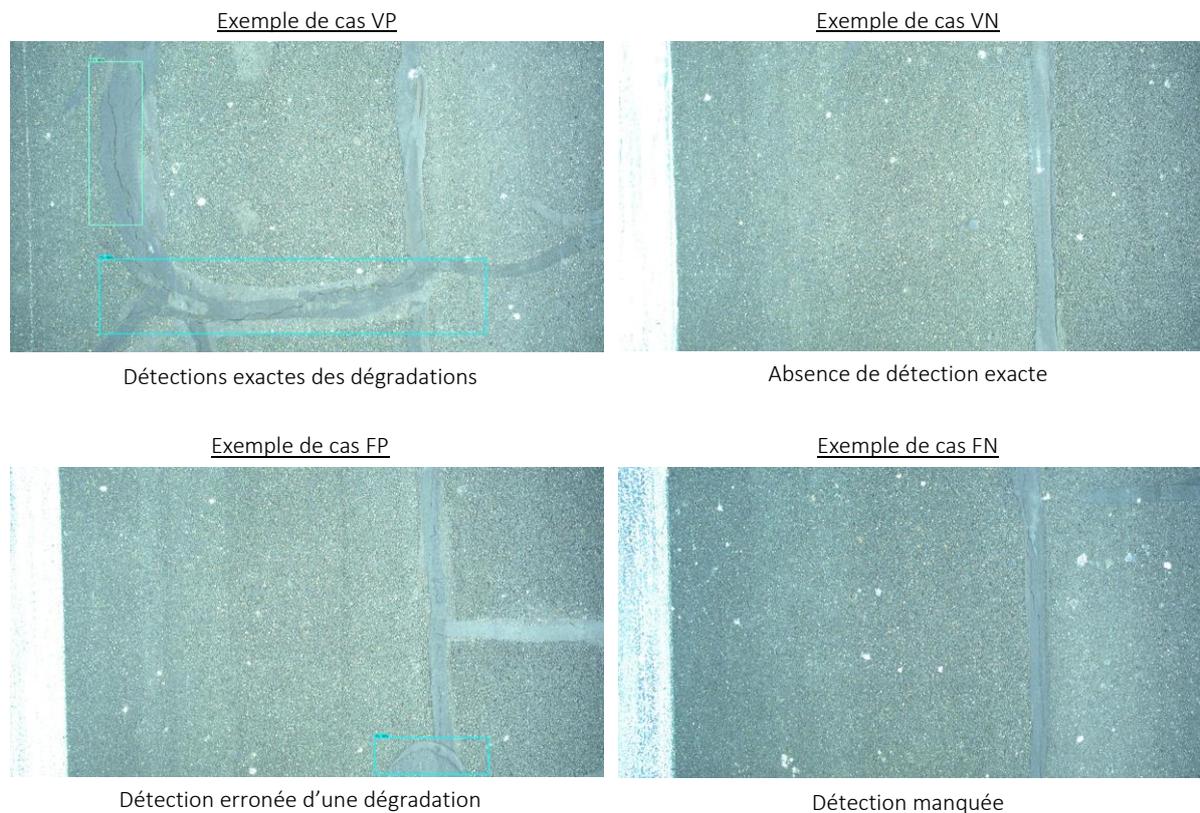


Figure 14. Illustrations d'images en sortie de modèle de DL pour les cas recensés dans la matrice de confusion

Le tableau suivant exprime la performance du modèle de deep learning sur la base des longueurs de dégradations détectées relativement à celles déterminées depuis la vérité terrain.

		Longueur dégradation Deep Learning (m)	Longueur dégradation Vérité Terrain (m)	DL / Vérité terrain (%)
<b>Bande 1</b>	LC	265.78	299.2	 89%
	TC	156.67	135.69	 115%
	Total Fissures	422.45	434.89	 97%
<b>Bande 2</b>	LC	113.94	139.49	 82%
	TC	179.89	173.23	 104%
	Total Fissures	293.83	312.72	 94%
<b>Bande 3</b>	LC	301.63	412.18	 73%
	TC	151.89	154.69	 98%
	Total Fissures	453.52	566.87	 80%
<b>Bande 4</b>	LC	63.4	39.84	 159%
	TC	144.31	146.7	 98%
	Total Fissures	207.71	186.54	 111%
<b>Total (toutes bandes)</b>	LC	744.75	890.71	 84%
	TC	632.76	610.31	 104%
	Total Fissures	1377.51	1501.02	 92%

Tableau 4. Comparaison des longueurs de détection estimées par le modèle de deep learning avec la vérité terrain

L'examen de ce tableau apporte un éclairage différent sur les résultats. On constate ainsi que les longueurs estimées des dégradations (toutes classes confondues) par le modèle DL sont, en proportion, relativement proches de celles que l'on devrait retrouver. On perçoit malgré tout deux tendances distinctes pour les classes étudiées :

- Pour la classe LC, une tendance à sous-estimer les longueurs des dégradations (en excluant les résultats de la bande n°4)
- Pour la classe TC, une tendance à surestimer les longueurs des dégradations

La sous-estimation systématique des longueurs peut s'expliquer en partie par l'effet des populations d'images avec des Faux Négatifs (détections manquées). Paradoxalement, les cas FN sont en proportion plus nombreux pour la classe TC (donc *a priori* on s'attendrait à une sous-estimation) tandis que les cas FP sont plus nombreux pour la classe LC (ce qui tendrait plutôt vers une surestimation). Si l'on fait le lien entre les résultats affichés à travers la matrice de confusion et les longueurs des dégradations, on peut donc supposer que le modèle produit des faux négatifs, donc des détections manquées, pour des dégradations de faibles longueurs (« bouts de fissures ») ; cela n'affecte donc que très peu la mesure de la longueur totale des dégradations. Dans quelques cas, la présence de faux négatifs peut également s'expliquer par une détection unique d'un ensemble de fissures « alignées » alors que dans la vérité terrain ces fissures ont été repérées séparément (Figure 15).

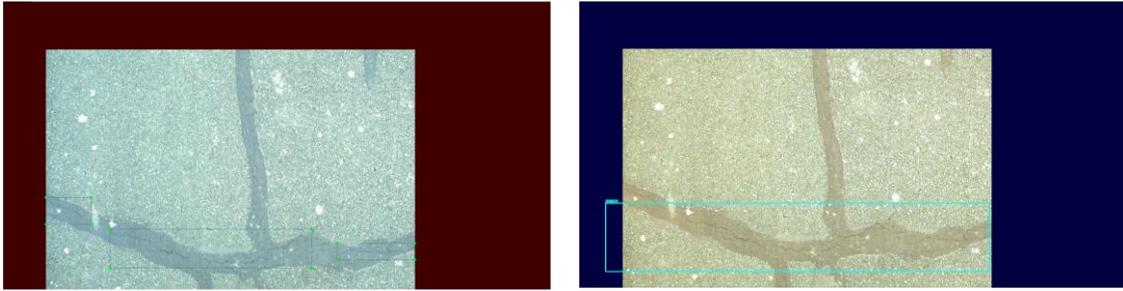


Figure 15. Vérité terrain faisant apparaître 3 TC distinctes (à gauche) et détection faisant apparaître une seule TC (à droite)

En agrégeant des fissures qui sont très proches les unes des autres dans la réalité terrain, on réduit significativement le nombre de faux négatifs. Cela conduit automatiquement à augmenter les degrés de rappel ainsi que le F1-score. Ce travail a été réalisé pour l'ensemble des bandes en considérant le total des classes LC et TC afin d'évaluer l'impact de ce genre de cas sur les résultats (Tableau 5).

	FN initial	FN après agrégation	Rappel initial	Rappel après agrégation	F1 score initial	F1 score après agrégation
Bande 1	167	157	65%	67%	73%	74%
Bande 2	178	102	51%	64%	64%	74%
Bande 3	244	158	54%	64%	66%	73%
Bande 4	171	87	27%	42%	40%	55%
Total	760	504	53%	63%	65%	72%

Tableau 5. Effet de l'agrégation des fissures dans la vérité terrain sur le niveau de performance du modèle DL

Concernant les cas de faux positifs (détections erronées) recensés dans le Tableau 3, il semblerait que le modèle de DL les produise principalement pour des dégradations de classe TC et de grande dimension; cette erreur est probablement due à une confusion avec les réparations de fissures qui occupent en général une place importante dans l'image. Dans le futur, il sera ainsi intéressant de calculer les distributions des composantes de la matrice de confusion en fonction des longueurs de fissuration pour valider l'hypothèse formulée précédemment. Ce travail n'a pas pu être réalisé dans ce projet faute de temps. En revanche, le modèle a été ré-entraîné de sorte à distinguer deux nouvelles classes que sont les fissurations dans des réparations dans le sens longitudinal (RLC) et dans le sens transversal (RTC). Un exemple de ce travail est illustré dans la Figure 16.



Figure 16. Exemple de détections des classes RLC et RTC (dégradations situées dans des pontages)

Ce nouveau modèle a été testé sur un faible échantillon d'images (~ 270 images, une image pouvant contenir plusieurs dégradations) prélevés sur le site de Lognes (images autres que celles utilisées pour l'entraînement bien entendu). Les résultats par rapport à la matrice de confusion sont présentés ci-après.

	<b>VP</b>	<b>FP</b>	<b>FN</b>	<b>Precision</b>	<b>Rappel</b>	<b>F1-score</b>
<b>TC</b>	74	20	20	78.72%	78.72%	79%
<b>LC</b>	154	40	38	79.38%	80.21%	80%
<b>RTC</b>	112	23	29	82.96%	79.43%	81%
<b>RLC</b>	190	42	47	81.90%	80.17%	81%

Tableau 6. Matrice de confusion et indicateurs servant au calcul du F1-score après ajout des classes RLC et RTC

Le fait d'avoir ajouté deux nouvelles classes qui prennent en charge des fissures présentes sur des réparations et d'avoir entraîné les modèles sur ces nouvelles cibles permet une amélioration sensible des résultats. Après ce travail, le critère F1-score est ainsi de l'ordre de 80% pour l'ensemble des classes de dégradations sur l'échantillon d'images considéré.

#### 4.5 Expérimentation sur le site de Roissy – Aéroport Charles de Gaulle

La deuxième expérimentation *in situ* du projet s'est déroulée sur l'aéroport Charles De Gaulle le 27 Avril 2022. Elle a concerné les aires suivantes qui sont désignées par leur indicatif opérationnel :

- Piste « Uniform » (chaussée rigide en dalles de béton)
- Piste « Charlie » (chaussée rigide en dalles de béton)
- Piste « Mike » (chaussée souple en enrobé bitumineux)

Pour cette expérimentation, un véhicule de la société ADP, seul à être autorisé à intervenir, a été instrumenté avec les mêmes équipements de mesure cités précédemment dans le rapport. Les relevés ont été réalisés à 30 km/h et sous contraintes du trafic aérien (interruptions des relevés par moment). Il faut également noter que, contrairement à Lognes, le nombre de dégradations et surtout de réparations sur ces pistes est bien plus faible ce qui réduit d'une certaine façon les possibilités statistiques dans l'analyse des résultats.

À l'heure où ce rapport a été rédigé (donc quelques semaines après l'expérimentation), l'ensemble des résultats n'est pas disponible mais des résultats partiels sont présentés par la suite.



Figure 17. Photo du véhicule instrumenté utilisé pour les essais à CDG.

##### 4.5.1 Relevés sur chaussée rigide

Cette expérimentation faisant intervenir des pistes en dalles de béton, l'introduction de nouvelles classes d'objets à détecter ainsi qu'un réentraînement des modèles ont été nécessaires. Ce dernier a consommé plus de 2000 images prélevées dans des bases de données libres mais également parmi les images de chaussée recueillies sur la totalité de la piste Uniform. En conséquence, l'évaluation du modèle de DL a uniquement été réalisé sur la piste Charlie pour le cas des chaussées rigides. Pour cette expérience, les classes d'objets suivants ont été considérés : fissuration longitudinale (LC), fissuration transversale (TC), joint de dalle (J) et coin de dalle (C). Les deux dernières classes ne sont pas des défauts de chaussée mais doivent être « appris » pour éviter la confusion avec les fissures d'une part et pour localiser les dégradations par rapport au coin de dalle d'autre part (cette information sera utile dans le futur pour identifier les cassures d'angle).



Figure 18. Vue des trajectoires du véhicule d'essais sur la piste Uniform (relevés ayant servi à l'entraînement)

Une fois l'entraînement réalisé, le modèle retenu pour traiter les images a été testé sur 9 bandes de la piste Charlie, calées sur des trajectoires parallèles du véhicule d'essais. Chaque bande mesure environ 330m de long et 3m de large.



Figure 19. Vue des trajectoires suivies par le véhicule d'essais sur la piste Charlie

Comme pour l'expérimentation de Lognes, une vérité terrain a été établie sur la base d'une analyse visuelle des images de chaussée enregistrées. C'est à partir de cette vérité terrain que le niveau de performance du modèle de DL a été établi dans le champ de l'indicateur F1-score et dans le champ de la longueur estimée de la dégradation.

		VP	FP	FN	Precision	Rappel	F1-score
Bande 1	LC	0	0	1	X	X	X
	TC	1	1	0	X	X	X
	Total Fissures	1	1	1	X	X	X
Bande 2	LC	2	1	1	X	X	X
	TC	2	0	0	X	X	X
	Total Fissures	4	1	1	X	X	X
Bande 3	LC	8	5	8	61.54%	50.00%	55%
	TC	7	3	9	70.00%	43.75%	54%
	Total Fissures	15	8	17	65.22%	46.88%	55%
Bande 4	LC	2	5	9	X	X	X
	TC	0	1	2	X	X	X
	Total Fissures	2	6	11	X	X	X
Bande 5	LC	1	2	3	X	X	X
	TC	2	4	0	X	X	X
	Total Fissures	3	6	3	X	X	X
Bande 6	LC	2	3	7	40.00%	22.22%	29%
	TC	1	0	11	100.00%	8.33%	15%
	Total Fissures	3	3	18	50.00%	14.29%	22%
Bande 7	LC	40	5	8	88.89%	83.33%	86%
	TC	1	2	2	33.33%	33.33%	33%
	Total Fissures	41	7	10	85.42%	80.39%	83%
Bande 8	LC	14	2	2	87.50%	87.50%	88%
	TC	0	1	1	X	X	X
	Total Fissures	14	3	3	82.35%	82.35%	82%
Bande 9	LC	3	3	0	X	X	X
	TC	0	3	0	X	X	X
	Total Fissures	3	6	0	X	X	X
Total toutes bandes	LC	72	26	39	73.47%	64.86%	69%
	TC	14	15	25	48.28%	35.90%	41%
	Total Fissures	86	41	64	67.72%	57.33%	62%

Tableau 7. Résultats du traitement par deep learning des images collectées sur la piste Charlie

Le Tableau 7 recense et caractérise les détections pour les fissurations longitudinales et transversales. Les critères de précision, de rappel et le F1-score sont calculés dans les bandes contenant suffisamment de défauts pour que cela ait un intérêt sachant que le nombre total de dégradations sur cette piste est assez faible. D'une façon générale, on recense un peu plus de faux positifs et de faux négatifs que pour les résultats obtenus à Lognes sur des chaussées souples. La conséquence est un F1-score moins performant. Par ailleurs, les quantités de faux positifs et de faux négatifs étant équilibrées, il y a donc à peu près autant de détections erronées que de détections manquées.

Les résultats, en termes de longueurs de fissurations sont présentés dans le Tableau 8.

		Longueur dégradation Deep Learning (m)	Longueur dégradation Vérité Terrain (m)
<b>Bande 1</b>	LC	0	0.2652
	TC	1.0913	0.2656
	Total Fissures	1.0913	0.5308
<b>Bande 2</b>	LC	2.2416	3.9983
	TC	0.8818	0.7479
	Total Fissures	3.1234	4.7462
<b>Bande 3</b>	LC	9.2165	13.4073
	TC	14.7974	10.3631
	Total Fissures	24.0139	23.7704
<b>Bande 4</b>	LC	9.4489	4.8168
	TC	1.1226	0.3339
	Total Fissures	10.5715	5.1507
<b>Bande 5</b>	LC	2.1075	2.793
	TC	7.1204	1.9767
	Total Fissures	9.2279	4.7697
<b>Bande 6</b>	LC	7.3706	6.4045
	TC	0.2073	3.6202
	Total Fissures	7.5779	10.0247
<b>Bande 7</b>	LC	49.029	50.7623
	TC	2.9676	1.4434
	Total Fissures	51.9966	52.2057
<b>Bande 8</b>	LC	17.3214	18.1841
	TC	2.6657	0.1826
	Total Fissures	19.9871	18.3667
<b>Bande 9</b>	LC	5.9376	1.5514
	TC	3.3569	0
	Total Fissures	9.2945	1.5514
<b>Total (toutes bandes)</b>	LC	102.6731	102.1829
	TC	34.211	18.9334
	Total Fissures	136.8841	121.1163

Tableau 8. Comparaison des longueurs de détection estimées par le modèle de deep learning avec la vérité terrain

Même si toutes les bandes apparaissent dans ce tableau, il est plus intéressant de regarder les résultats sur les bandes 3, 7 et 8 qui comportent un nombre plus significatif de dégradations. Pour celles-ci, on constate que les longueurs des dégradations estimées par le modèle sont proches de celles provenant de la vérité terrain. Ainsi, malgré le changement de type de surface (de l'enrobé au béton) le modèle conserve, de ce point de vue, un niveau de performance intéressant. En revanche, la prise en charge des joints de dalles est essentielle pour assurer une bonne détection. En effet, il peut y avoir confusion entre une fissure et un joint et ainsi générer une forte quantité de faux positifs. Pour cette raison, et comme dit précédemment, le modèle a été entraîné à détecter les joints et les coins de dalles de béton. Les résultats de cette détection sur une bande de la piste Charlie sont exposés dans le Tableau 9.

	<b>VP</b>	<b>FP</b>	<b>FN</b>	<b>Precision</b>	<b>Rappel</b>	<b>F1-score</b>
<b>Joints</b>	88	2	3	97.78%	96.70%	97%
<b>Coins</b>	21	0	1	100.00%	95.45%	98%

Tableau 9. Performance du modèle pour la détection des joints et des coins des dalles selon le F1-score

La prise en charge de la détection des joints de chaussée et des coins de dalle par le modèle est excellente comme l'indiquent les valeurs de F1-score. Malgré tout, il peut subsister quelques erreurs de détection et en particulier des faux négatifs pour les joints qui, pour les plus irréguliers d'entre eux, peuvent être apparentés à une fissure. Par conséquent, cette confusion mène à la production de faux positifs pour la détection des dégradations ce qui explique en partie les valeurs observées dans le Tableau 7.

Un exemple de cette problématique est illustré au moyen des images de la Figure 20.

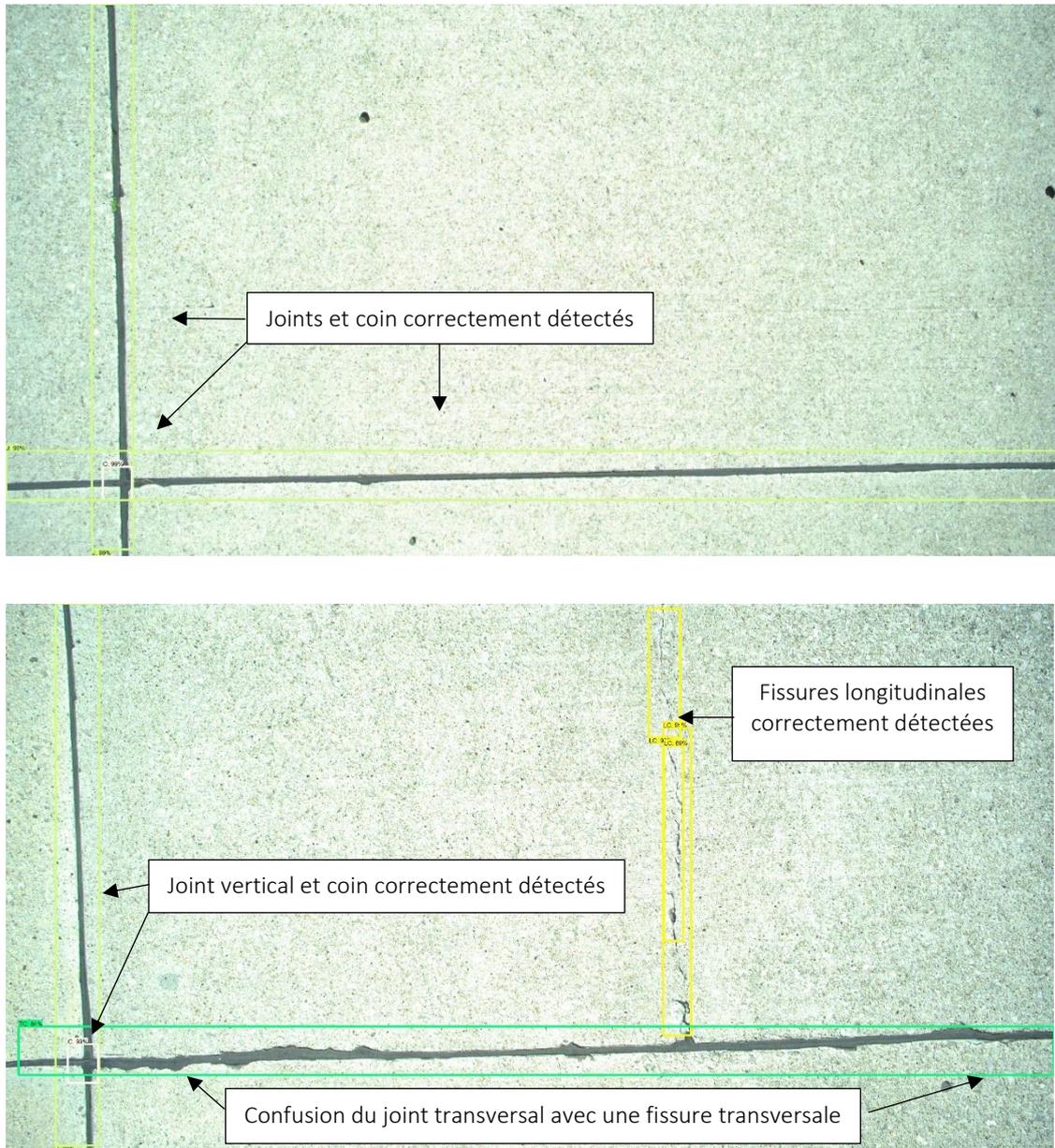


Figure 20. Illustration de la détection des joints et coins de dalles ainsi que les possibles confusions avec les dégradations

Pour l'après-projet, il est prévu d'utiliser la détection des coins de dalles pour évaluer les distances entre eux et d'éventuelles dégradations à proximité, le but étant d'identifier les cassures d'angle.



Figure 21. Détection d'une fissure à proximité d'un coin de dalle

#### 4.5.2 Relevés sur chaussée souple

Des relevés ont pu être réalisés sur une plateforme en enrobé bitumineux (piste Mike) contenant un nombre significatif de dégradations afin de compléter les résultats obtenus sur le site de Lognes. Malheureusement, à cause des contraintes d'exploitation aérienne de cette piste, le temps d'intervention a été extrêmement limité et il n'a pas été possible de réaliser un balayage précis de la piste avec le véhicule d'essais comme cela a été fait sur les pistes Charlie et Uniform. Par ailleurs, la vérité terrain pour cette piste n'a pas été établie en date de rédaction de ce rapport. On donne néanmoins les résultats suivants qui portent sur des longueurs de dégradations par bande ainsi que quelques illustrations de détection (d'autres sont également disponibles en annexes).



Figure 22. Vue des trajectoires suivies par le véhicule d'essais sur la piste Mike

	Longueurs de fissure Deep Learning (m)			Longueurs de fissure Vérité Terrain (m)		
	LC	TC	Total	LC	TC	Total
<b>Bande 6</b>	78.09	31.77	109.87	En cours de calcul		
<b>Bande 5</b>	67.02	61.5	128.52			
<b>Bande 1</b>	10.56	61.26	71.83			
<b>Bande 2</b>	83.01	35.02	118.03			
<b>Bande 3</b>	207.41	54.58	261.99			
<b>Bande 4</b>	469.47	69.72	539.18			
<b>Total</b>	<b>653.51</b>	<b>298.72</b>	<b>952.23</b>			

Tableau 10. Valeurs des longueurs de dégradations estimées par le modèle DL sur la piste Mike

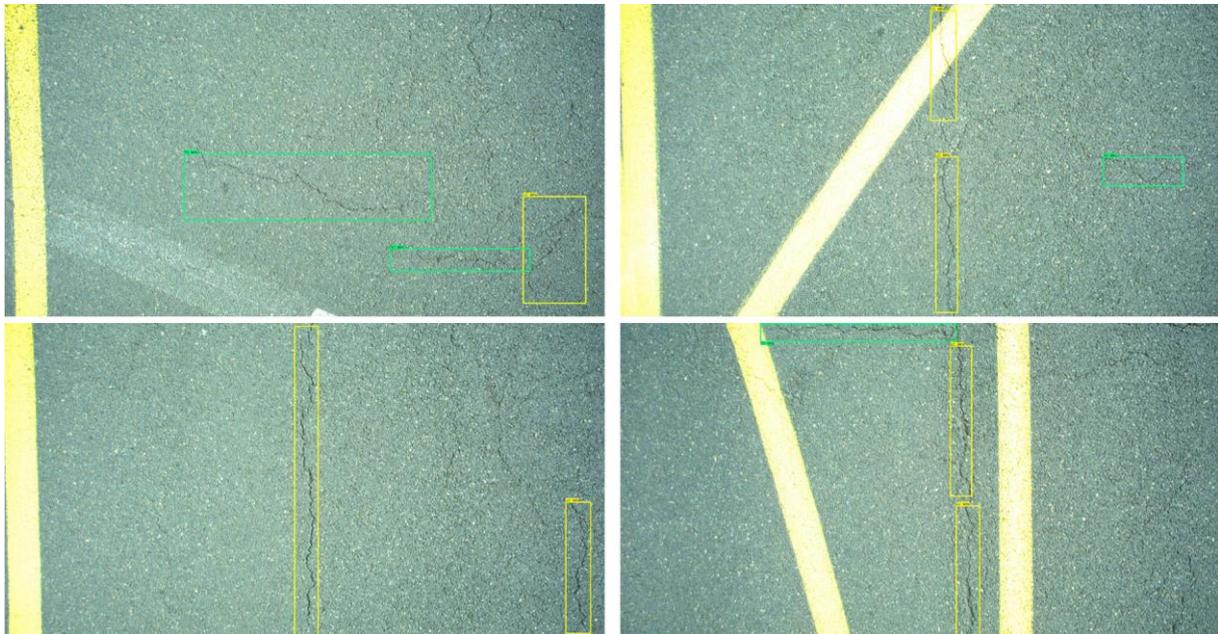


Figure 23. Illustrations de détections de dégradations par le modèle DL sur la piste Mike

Sur la base d'un examen visuel des images et d'après les impressions recueillies sur le terrain lors des relevés, la proportion de dégradations entre les 6 bandes de mesure semble cohérente (la bande 4 proche des hangars étant celle qui est manifestement la plus dégradée). En revanche, il semblerait que les longueurs de détections pour chaque bande soient sous-estimées car des faux négatifs apparaissent régulièrement sur les images. Ces impressions devront être confirmées en après-projet après confrontation avec la vérité terrain.

## 5 Synthèse

Grâce à des essais en laboratoire et à des expérimentations *in situ*, le projet EASY-(DAI)2 a permis de démontrer le potentiel de détection de dégradations de chaussées aéronautiques offert par le deep learning par une application sur des images issues d'un dispositif « allégé ». Un tel dispositif serait en effet indiqué pour l'auscultation de plateformes qui ne peuvent bénéficier de l'usage de moyens lourds soit pour des raisons de coût soit pour des raisons logistiques (disponibilité, acheminement du matériel, etc.). Par conséquent, un soin a été apporté au choix des composants les plus adéquats en début de projet (et tout particulièrement l'imageur) de sorte à définir une instrumentation « type » qui pourrait être dupliquée à l'avenir dans un cadre purement opérationnel. Une fois cette instrumentation figée et la qualité des données entrantes connue, la puissance de la méthode porte alors sur le traitement des images de chaussée pour l'identification et la classification des dégradations. Dans le projet EASY-(DAI)2, ce traitement a été assuré par un apprentissage profond (deep learning) qui est une technique provenant de l'intelligence artificielle.

Dans une première approche, les modèles de deep learning déployés pour ce projet avaient pour objectif de détecter une ou plusieurs dégradations dans une image de chaussée, d'identifier sa classe et de dimensionner sommairement chaque défaut au moyen d'un cadre de détection. Ces modèles ne permettaient cependant pas d'extraire le squelette des dégradations ni même d'estimer l'ouverture de la fissure. Ce projet a également été l'occasion de tester la transférabilité de modèles de deep learning entre une application routière et une application aéronautique. Les résultats expérimentaux ont montré qu'un tel transfert était possible mais qu'un réentraînement était nécessaire en particulier pour la prise en charge de défauts sur des dalles de béton. Il faut aussi noter que le niveau d'exigence pour la détection des dégradations sur chaussée aéronautique a volontairement été plus élevé pour ce projet par rapport à des projets portant sur le domaine routier où les longueurs de linéaires sont telles qu'un plus fort taux d'erreurs est en général mieux toléré.

À l'issue de plusieurs tests, un seul modèle de deep learning a été retenu et entraîné de sorte à ce qu'il puisse performer aussi bien sur chaussée aéronautique souple (bitumineuse) que rigide (béton). L'évaluation de ce modèle a été effectuée selon plusieurs critères : le F1-score pour mettre en avant la capacité à détecter et classifier la dégradation et le critère de longueur de fissuration pour estimer la capacité du modèle à dimensionner correctement cette même dégradation. Le niveau de performance du modèle de deep learning peut également être apprécié à travers d'autres critères. Ainsi, même s'ils n'ont pas fait l'objet d'une étude approfondie, les informations suivantes ont été obtenues :

- Le temps de traitement par image (d'environ 1s par image, ce qui est très correct par rapport aux besoins du gestionnaire) ;
- La capacité de détection par rapport à des techniques de traitement d'images traditionnelles de segmentation.

Sur ce dernier point et pour le cas des chaussées souples, nous avons constaté que le modèle de deep learning était moins gêné sur des images « difficiles » présentant par exemple des ombres portées au sol ou bien du marquage horizontal.

Suite à cette étude de faisabilité, un certain nombre de verrous doivent encore être levés pour tendre vers une solution opérationnelle qui puisse être une alternative aux relevés visuels pédestres. Ces verrous concernent :

- Une amélioration du modèle de deep learning pour réduire en particulier le nombre de faux négatifs ;
- Une aide à l'acquisition d'images de chaussée pour faciliter le balayage du véhicule d'essais sur la piste, augmenter la vitesse de passage et obtenir des images parfaitement séquencées (à pas fixe et bien juxtaposées).

D'autres travaux peuvent compléter cette étude en considérant par exemple un changement de méthode de prise de vue (vue panoramique, usage d'un drone) ou bien encore d'autres modèles de deep learning ou d'autres techniques en provenance de l'IA pour détecter les dégradations.

## 6 Utilisation de l'aide de la Ferec dans le projet

L'aide de la Ferec dans ce projet a principalement été mise à profit pour progresser efficacement et le plus vite possible sur l'utilisation des techniques de deep learning. Elle a ainsi contribué à l'embauche d'un post-doctorant pour une durée de 12 mois qui avait engagé des travaux sur cette thématique pendant sa thèse mais dans le contexte routier. Dans le projet EASY-(DAI)<sup>2</sup>, ce post-doctorant s'est focalisé sur la problématique de transférabilité des modèles de deep learning et de leur optimisation pour répondre au contexte du secteur aéronautique. Il a également fortement valorisé les travaux réalisés dans le projet par la rédaction d'un article pour une conférence internationale dédiée aux caractéristiques de surface ainsi qu'une communication pour un « user group » européen spécialisé dans les techniques d'auscultation de surface.

## 7 Impact

### 7.1 Mise en évidence du caractère collaboratif et d'intérêt général

Le projet EASY-(DAI)<sup>2</sup> a permis d'initier une collaboration entre le STAC et l'UGE sur la problématique de relevés de dégradations sur chaussées aéronautiques avec le soutien d'Aéroport De Paris (ADP). Grâce à ce projet, les équipes des différents organismes ont pu se découvrir et travailler ensemble notamment lors des expérimentations *in situ* menées sur les aéroports de Lognes-Emerainville et Roissy-Charles de Gaulle. Cette collaboration a réuni du personnel de recherche et des techniciens du STAC et de l'UGE pour les développements portant sur l'instrumentation embarquée ainsi que les outils de traitement des images. Grâce à l'intervention d'agents venant d'ADP, la collaboration s'est enrichie du point de vue du gestionnaire ce qui a facilité l'organisation des expérimentations et consolidé l'évaluation des résultats (contribution à la vérité terrain).

### 7.2 Apports et retombées pour la collectivité

Beaucoup d'acteurs de la profession s'intéressent à l'usage des techniques provenant de l'intelligence artificielle pour tenter de lever ou contourner des verrous pour la détection des dégradations depuis des relevés d'images. Contrairement aux techniques « traditionnelles » qui ont fait leurs preuves mais qui ont aussi montré leurs limites, les techniques issues de l'IA pour ce genre d'applications sont encore peu éprouvées et évaluées formellement ; il existe ainsi une forte attente de la part de la profession pour mieux cerner les possibilités offertes par ces techniques et en particulier le deep learning. Le livrable du projet EASY-(DAI)<sup>2</sup> présente des résultats quantifiés suite à l'application de ces techniques sur des chaussées aéronautiques. Il met également en avant la question de la transférabilité des modèles de deep learning initialement développés pour des applications routières (auscultation de grands linéaires) vers des applications aéronautiques (auscultation de plateformes) sachant que l'objectif global d'aide à la maintenance est commun aux deux cadres mais que les informations collectées ne desservent pas les mêmes indicateurs. Pour ces raisons, les retours d'expériences sont précieux et permettent d'alimenter en particulier les groupes de travail incluant les différents acteurs de la profession. Il existe actuellement plusieurs groupes qui discutent de l'actualisation ou de la refonte

des méthodologies d'analyse des dégradations aussi bien dans le secteur routier que dans le secteur aéronautique.

### 7.3 Publications et autres éléments de communication auxquels le projet et ses résultats ont donné lieu

Les travaux réalisés dans le cadre du projet EASY-(DAI)2 sont ou seront valorisés par les publications et communications suivantes :

- Un article intitulé « *Towards efficient airfield pavement surface condition monitoring using Deep Learning models* » accepté pour la conférence SURF 2022 (9th symposium on pavement surface characteristics) qui se déroulera à Milan du 12 au 14 septembre 2022 (conférence avec comité de lecture) ;
- Une communication orale qui a été acceptée pour la rencontre ERPUG 2022 (European Road Profile User's Group) et qui se déroulera à Edimbourg en octobre 2022 ;
- Une soumission d'article pour la revue RGRA qui éditera un dossier spécial sur les chaussées aéronautiques en septembre 2022.

## 8 Perspectives ouvertes par le projet notamment en matière de recherche collaborative

À l'issue de ce projet, les laboratoires du STAC, de l'UGE et d'ADP vont continuer à collaborer dans le cadre d'une convention de recherche. Après la finalisation des travaux engagés (finir l'analyse des relevés effectués sur les pistes CDG), les recherches vont porter sur le renforcement du traitement des images pour lequel il est prévu de travailler sur un modèle de deep learning qui permette d'extraire le squelette de la dégradation. Ce dernier apportera des informations plus précises sur les longueurs de fissuration et des informations « neuves » sur l'ouverture des fissures. De cette manière, on disposera de tous les éléments nécessaires pour détecter et caractériser les dégradations par rapport à une notion de gravité et calculer ainsi un Indice de Service (IS) tel qu'il est défini dans la doctrine technique. On vise également dans cet après-projet à recourir à une vérité terrain établie depuis un appareil de référence (de type LCMS) ; pour des raisons de calendrier, cela n'avait pas pu être réalisé lors des expérimentations à Lognes et Roissy-Charles de Gaulle.

## 9 Annexes (albums d'images en sortie du modèle DL)

### Expérimentation à Lognes (exemples de détections issues du modèle de deep learning)



*Détection exacte de dégradations (fissure prononcée)*



*Détections exactes de dégradations (fissures multiples)*



*Détections exactes des dégradations (fissures multiples)*



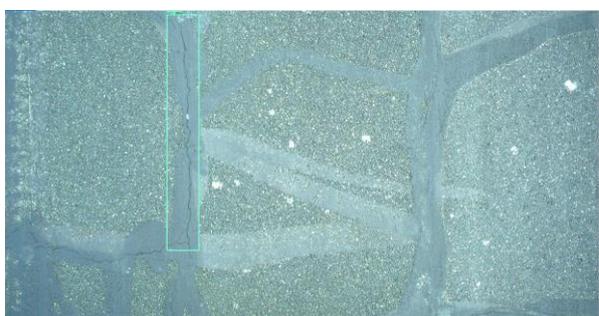
*Détection exacte de dégradations (fissure fine)*



*Détection erronée de dégradations (dans marquage)*



*Détection erronée de dégradations (à droite)*

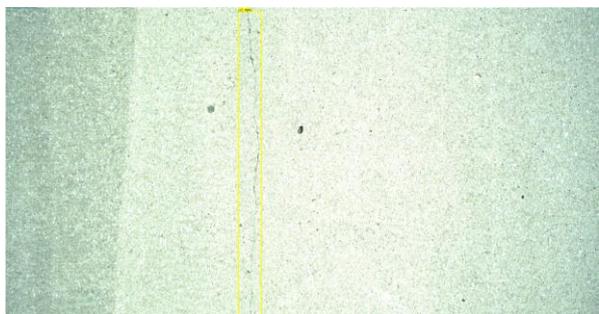


*Absence de détection (fissure transversale gauche)*



*Absence de détection*

Expérimentation à Roissy-Charles de Gaulle (piste Charlie - chaussée rigide)



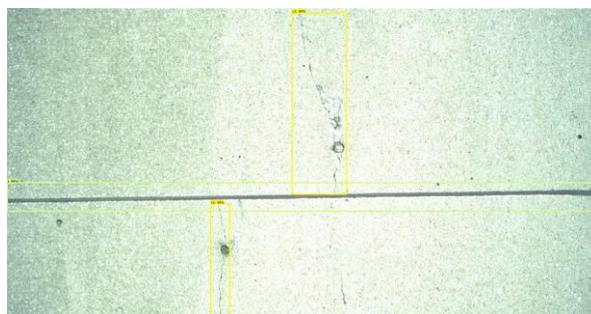
*Détection exacte de dégradations (fissure prononcée)*



*Détections exactes de dégradations (fissures multiples)*



*Détections exactes des dégradations et du joint*



*Détection exacte de dégradations (fissures multiples)*



*Détection erronée de dégradation (TC en bas d'image)*



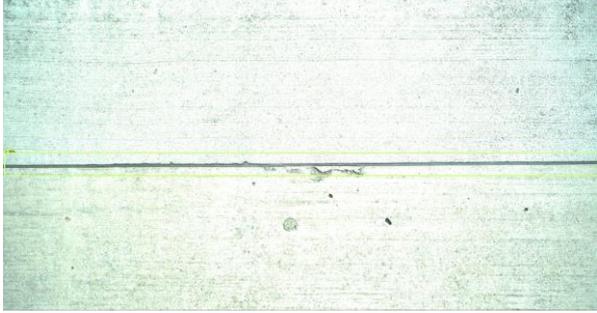
*Détection erronée de dégradation*



*Absence de détection (fissure en bas du joint)*



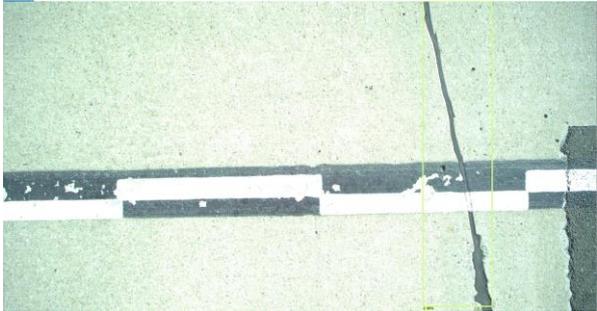
*Absence de détection (partie basse de la fissure)*



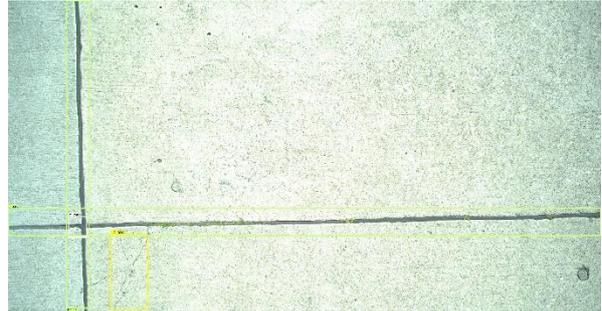
*Détection exacte de joint (horizontal)*



*Détection exacte de joint (vertical)*



*Détection exacte de joint (oblique)*



*Détection exacte du coin de dalle*



*Détection erronée de joint*



*Détection manquante du coin de dalle*

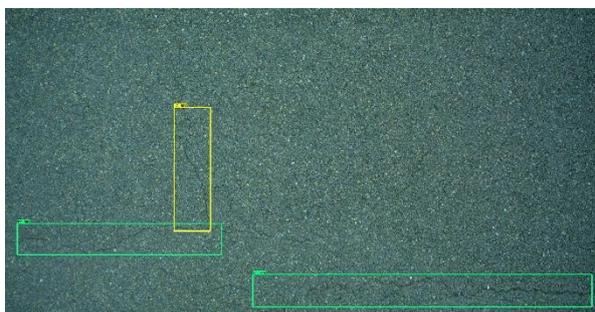
Expérimentation à Roissy-Charles de Gaulle (piste Mike - chaussée souple)



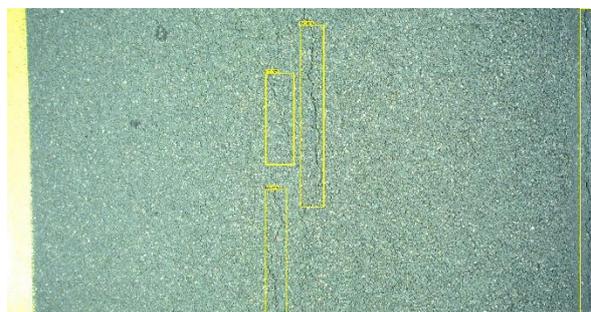
*Détection exacte de dégradations (LC)*



*Détection exacte de dégradations (TC)*



*Détection exacte de dégradations (fissures multiples)*



*Détection exacte de dégradations (fissures multiples)*



*Détection erronée de dégradation (confusion avec dispositif lumineux)*



*Détection erronée de dégradation (bordure de réparation)*



*Absence de détection*



*Absence de détection*