UNIVERSITÉ DU QUÉBEC À CHICOUTIMI

**MÉMOIRE PRÉSENTÉ À L'UNIVERSITÉ DU QUÉBEC À CHICOUTIMI
EN VUE DE L'OBTENTION DU GRADE DE
MAÎTRISE ÈS SCIENCES (M. SC.) EN INFORMATIQUE**

**PAR
BI TAH GUY-DANIEL DIALLO**

ANALYSE D'IMAGES POUR UN SYSTÈME INTERACTIF DE CAPTATION DE TABLEAUX PÉDAGOGIQUES EN SALLE DE CLASSE

JUILLET 2026

RÉSUMÉ

Le tableau demeure un support pédagogique central dans de nombreuses salles de classe, mais son contenu est généralement effacé à la fin du cours. Ce mémoire présente un système interactif visant à détecter, extraire, rectifier et conserver automatiquement la région d'un tableau pédagogique à partir d'images acquises dans des conditions réelles. La difficulté scientifique tient aux variations d'éclairage, aux reflets, aux prises de vue obliques, aux occlusions et à la diversité des salles. L'approche proposée combine des méthodes classiques de traitement d'images - filtrage, détection de contours, transformée de Hough, analyse géométrique et homographie - avec un modèle YOLOv8s-seg entraîné par transfert pour détecter et segmenter la surface du tableau. Le jeu de données comprend 416 images annotées par polygones : 336 images sont utilisées pour l'entraînement et 80 pour la validation. Aucun jeu de test indépendant n'a été constitué; les résultats doivent donc être interprétés comme une validation interne. Sur les 80 images de validation, le modèle obtient une précision de 0,987, un rappel de 1,000, une mAP@0,50 de 0,992 et une mAP@0,50-0,95 de 0,967. L'approche hybride porte l'Intersection over Union moyenne de 0,91 à 0,94 grâce à la correction géométrique. Le temps d'inférence mesuré sur processeur est de 476,7 ms par image, ce qui convient à une captation ponctuelle mais pas encore à une vidéo fluide. Les résultats montrent l'intérêt de combiner l'apprentissage profond, robuste aux variations visuelles, avec la géométrie projective, précise et interprétable. Les limites concernent principalement la taille du jeu de données, l'absence de test externe, l'absence de scènes négatives sans tableau et les scènes fortement occultées. La reconnaissance optique de caractères n'est pas implémentée dans la version évaluée et constitue une perspective de recherche.

Mots-clés : traitement d'images, tableau pédagogique, YOLOv8, segmentation, OpenCV, homographie.

REMERCIEMENTS

Je tiens tout d'abord à exprimer ma profonde gratitude à mon directeur de recherche, Yves Chiricota pour son encadrement, sa disponibilité et la qualité de ses conseils tout au long de ce travail de maîtrise. Ses orientations méthodologiques, ses remarques constructives et sa rigueur scientifique ont été déterminantes dans l'avancement et la structuration de ce mémoire.

Je remercie également les membres du corps professoral du département d'informatique pour la qualité de l'enseignement reçu durant mon parcours de maîtrise. Les connaissances théoriques et pratiques acquises au cours de cette formation ont constitué une base essentielle pour la réalisation de ce travail de recherche.

Mes remerciements s'adressent aussi à mes collègues et camarades de programme, pour les échanges enrichissants, les discussions techniques et le soutien mutuel qui ont contribué à maintenir un environnement de travail stimulant et collaboratif.

Je souhaite enfin remercier ma famille pour son soutien constant, sa patience et ses encouragements tout au long de mes études. Leur appui moral a été une source de motivation essentielle dans la réalisation de ce projet académique.

DÉCLARATION D'UTILISATION D'OUTILS NUMÉRIQUES

Une assistance par intelligence artificielle générative a été utilisée pour la révision linguistique, l'harmonisation de la structure et la mise en forme de certains schémas conceptuels. Les données expérimentales, les annotations, le code, les résultats, leur interprétation scientifique et les décisions méthodologiques demeurent sous la responsabilité de l'auteur. Les figures expérimentales proviennent du jeu de données et des traitements réalisés dans le cadre du projet; les schémas ont été redessinés afin d'éviter toute ambiguïté relative aux droits d'utilisation.

TABLE DES MATIÈRES

LISTE DES FIGURES

LISTE DES TABLEAUX

CHAPITRE 1
INTRODUCTION

1.1 Contexte et problématique

Dans l'enseignement en présentiel, le tableau demeure un support pédagogique important pour exposer progressivement un raisonnement, illustrer des concepts abstraits au moyen de schémas, d'équations ou de graphiques, et soutenir l'interaction entre l'enseignant et les étudiants. Dans ce mémoire, le tableau est envisagé comme un support d'écriture effaçable particulièrement utile lorsque la construction pas à pas du raisonnement occupe une place centrale.

Cependant, le contenu inscrit sur le tableau présente une caractéristique intrinsèque: son caractère éphémère. Une fois le cours terminé, les informations écrites sont effacées, entraînant la perte définitive de supports pédagogiques parfois essentiels à la compréhension du cours. Cette situation peut poser des difficultés aux étudiants, notamment lorsque le rythme du cours est soutenu, lorsque certaines explications sont complexes ou lorsque l'étudiant est absent. Du point de vue de l'enseignant, l'impossibilité de conserver une trace fidèle du tableau limite également les possibilités de réutilisation, d'archivage ou de diffusion du contenu pédagogique.

Face au caractère éphémère de ce support, diverses solutions technologiques ont été proposées, notamment les tableaux numériques interactifs et les systèmes de captation vidéo. Toutefois, ces solutions peuvent être coûteuses, dépendre d'une infrastructure dédiée ou demeurer peu adaptées aux salles de classe traditionnelles. Les principales familles de solutions existantes seront présentées et discutées au chapitre 2. Dans ce contexte, l'exploitation conjointe du traitement d'images et de la vision par ordinateur apparaît comme une piste pertinente pour concevoir un système automatisé, accessible et non intrusif.

La captation automatique du contenu d'un tableau pédagogique soulève plusieurs défis techniques et scientifiques. Contrairement à des objets rigides et bien définis, le tableau est observé dans des conditions variables: changements d'éclairage, présence de reflets, variations d'angle de prise de vue, occlusions partielles causées par l'enseignant ou par des

objets environnants, et diversité des environnements de salles de classe. Ces facteurs rendent la détection et l'extraction du tableau particulièrement complexes dans un contexte réel.

De plus, le tableau ne se réduit pas à une simple surface rectangulaire. Il constitue une région d'intérêt dont la détection doit être suffisamment précise pour permettre une extraction fidèle du contenu, tout en demeurant robuste face aux perturbations visuelles. En pratique, le système visé doit produire des résultats satisfaisants même lorsque le tableau est capté sous un angle, partiellement occulté par l'enseignant ou soumis à un éclairage non uniforme. Les approches fondées sur des règles trop simples se révèlent alors insuffisantes, tandis que les méthodes plus avancées doivent concilier précision, robustesse et temps de calcul raisonnable.

La problématique centrale de ce mémoire peut ainsi être formulée comme suit: comment concevoir un système automatique permettant de détecter, d'extraire et de conserver le contenu d'un tableau pédagogique à partir d'images ou de vidéos, tout en maintenant de bonnes performances lorsque la prise de vue est oblique, que certaines parties du tableau sont occultées ou que les conditions d'éclairage varient fortement?

La captation automatique du tableau s'inscrit dans la continuité des travaux de numérisation de tableaux blancs par caméra, qui combinent localisation des frontières, rectification projective et amélioration de l'image (Zhang et He, 2003; He et Zhang, 2004).

La figure 1.1 présente un exemple réel de tableau pédagogique utilisé dans le jeu de données.

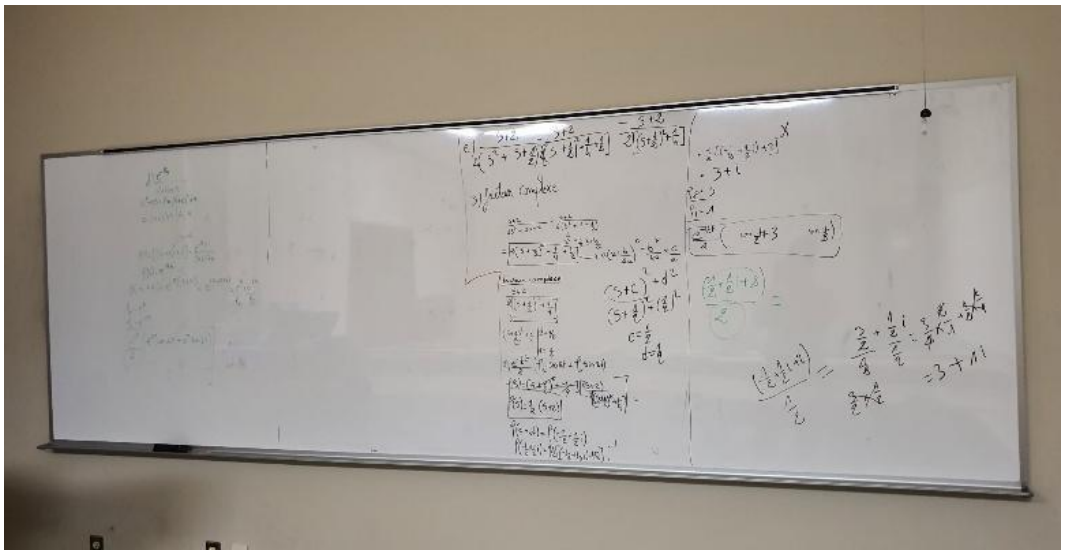

Figure 1.1 - Exemple de tableau pédagogique en salle de classe

La figure 1.2 met en évidence l'enjeu pédagogique: une information structurée pendant le cours disparaît après l'effacement si aucune procédure de captation n'est déclenchée.

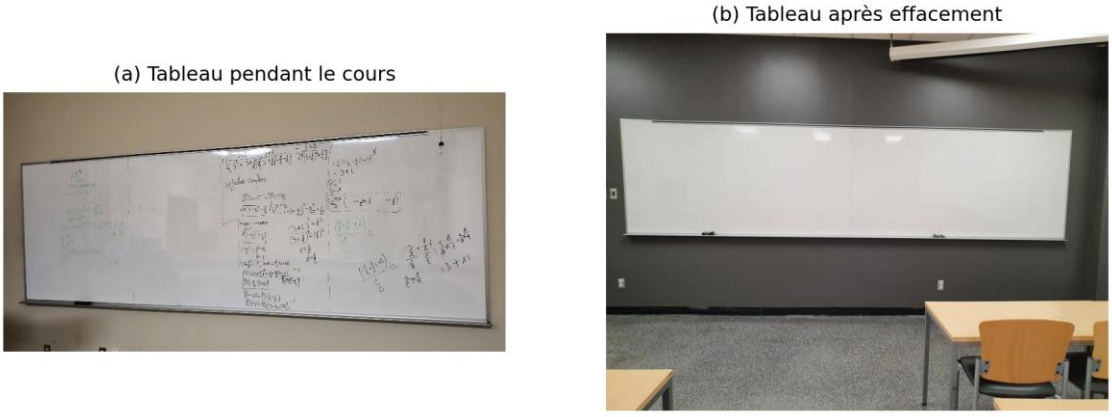

Figure 1.2 - Caractère éphémère du contenu inscrit au tableau : (a) pendant le cours; (b) après effacement

1.2 Objectifs et contributions

L'objectif général est de concevoir et d'évaluer un système interactif permettant de détecter, segmenter, rectifier et conserver la région d'un tableau pédagogique à partir d'une image ou d'un flux vidéo. Cet objectif général se décline en quatre sous-objectifs.

Les sous-objectifs consistent à étudier les méthodes classiques et profondes pertinentes pour la localisation d'un tableau; à constituer et annoter un jeu de données diversifié issu de salles de classe réelles; à intégrer YOLOv8s-seg (Jocher et al., 2023) et OpenCV dans un

pipeline hybride de segmentation et de rectification; à évaluer la précision, le rappel, la mAP, l'IoU et le temps d'inférence sur l'ensemble de validation.

La contribution associée au premier sous-objectif est une synthèse critique reliant traitement d'images, segmentation et géométrie projective. Le deuxième sous-objectif conduit à un jeu de 416 images annotées par polygones. Le troisième produit un pipeline modulaire combinant masque neuronal et correction géométrique. Le quatrième fournit une validation interne détaillée et une analyse explicite des limites.

1.3 Démarche générale

La démarche suit quatre phases. La première consolide les fondements théoriques et l'état de l'art. La deuxième constitue les données et définit les annotations. La troisième entraîne le modèle puis intègre sa sortie au traitement géométrique. La quatrième mesure les performances sur 80 images de validation et analyse les échecs. Les paramètres détaillés et les résultats sont présentés respectivement aux chapitres 4 et 5.

1.4 Organisation du mémoire

Le mémoire comporte six chapitres. Le chapitre 1 introduit le problème et les objectifs. Le chapitre 2 regroupe la revue de littérature et les fondements méthodologiques. Le chapitre 3 présente la conception du système hybride. Le chapitre 4 décrit l'implémentation et le protocole expérimental. Le chapitre 5 analyse les résultats et leurs limites. Le chapitre 6 conclut et présente les perspectives.

CHAPITRE 2
REVUE DE LITTÉRATURE ET FONDEMENTS MÉTHODOLOGIQUES

2.1 Contexte pédagogique et systèmes existants

Dans de nombreux contextes d'enseignement en présentiel, le tableau demeure un support pédagogique privilégié en raison de sa simplicité d'utilisation, de sa flexibilité et de son caractère interactif. L'enseignant y inscrit des formules, des schémas, des démonstrations ou des exemples, souvent en interaction directe avec les étudiants, en adaptant le contenu au rythme de la classe et aux questions posées.

Contrairement à des supports plus figés tels que les diapositives, le tableau permet une construction dynamique du savoir. Cette progression favorise la compréhension de raisonnements complexes, notamment dans les disciplines scientifiques et techniques. En contrepartie, le contenu du tableau demeure transitoire: il disparaît une fois le cours terminé, ce qui peut nuire à l'apprentissage lorsque les étudiants ne parviennent pas à recopier l'ensemble du contenu ou souhaitent revenir ultérieurement sur certaines explications.

Afin de pallier le caractère éphémère du tableau, plusieurs pratiques sont couramment utilisées. Certains enseignants prennent des photographies du tableau à la fin du cours, tandis que d'autres proposent des notes complémentaires ou des supports numériques. Cependant, ces solutions demeurent souvent informelles et dépendent fortement de l'initiative individuelle.

La prise de photographies manuelles, par exemple, peut entraîner des problèmes de cadrage, de qualité d'image ou de lisibilité, notamment en présence de reflets ou d'un éclairage insuffisant. De plus, cette pratique ne s'intègre pas toujours naturellement au déroulement du cours et peut représenter une contrainte supplémentaire pour l'enseignant.

Ces limites soulignent la nécessité de solutions automatisées permettant de capturer et d'archiver le contenu du tableau de manière fiable et systématique.

D'autres solutions reposent sur la captation vidéo de la salle de classe. Bien que cette approche permette de conserver l'intégralité du cours, elle génère des volumes importants

de données et ne facilite pas l'extraction ciblée du contenu du tableau. De plus, l'analyse manuelle des vidéos reste chronophage et peu adaptée à une exploitation pédagogique efficace.

Ces constats ont conduit la communauté scientifique à s'intéresser à des approches basées sur l'analyse automatique d'images, visant à détecter et à extraire directement la région du tableau à partir de données visuelles. Les principes de ces approches sont discutés dans les sections 2.4 et 2.5, avant d'être comparés de manière critique à la section 2.6.

Les systèmes existants se répartissent entre la photo manuelle, le tableau numérique interactif, la captation vidéo et les solutions de vision par ordinateur. Les travaux sur la capture de tableaux par caméra montrent qu'une solution non instrumentée peut localiser, rectifier et archiver la surface utile tout en conservant un matériel simple (Zhang et He, 2003; Bérard, 2003; Golovchinsky et al., 2009).

Tableau 2.1 - Comparaison des solutions de conservation du contenu d'un tableau

Solution	Type	Coût	Infrastructure requise	Conservation automatique	Avantages	Limites
Tableau traditionnel + photo manuelle	Méthode informelle	Très faible	Aucune	Non	Simplicité, accessibilité	Qualité variable, non systématique
Tableau numérique interactif (TNI)	Matériel spécialisé	Élevé	Installation fixe	Oui	Enregistrement intégré, interactivité	Coûtélevé, maintenance
Captation vidéo complète	Enregistrement audiovisuel	Moyen à élevé	Caméra+ stockage	Oui	Conservation intégraledu cours	Données volumineuses, extraction difficile
Système basé sur vision par ordinateur	Logiciel autonome	Faible à moyen	Caméra standard	Oui	Automatisation, faiblecoût matériel	Sensible aux conditions visuelles

Le tableau 2.1 montre que la vision par ordinateur réduit la dépendance à un matériel spécialisé, mais transfère la difficulté vers la robustesse algorithmique. Cette observation justifie l'étude d'une approche hybride.

2.2 Traitement d'images et géométrie projective

Les approches basées sur le traitement d'images exploitent des propriétés visuelles et géométriques du tableau, telles que la forme générale, les contrastes locaux, les contours et la cohérence spatiale. Elles mobilisent des outils classiques comme le seuillage, la morphologie mathématique, la détection de contours et l'analyse géométrique (Gonzalez et Woods, 2018; Serra, 1982; Canny, 1986; Duda et Hart, 1972).

Dans le cas des tableaux pédagogiques, ces méthodes sont pertinentes parce qu'elles permettent d'isoler une région d'intérêt généralement plane, de renforcer ses bords et de localiser ses coins en vue d'une correction de perspective. Elles demeurent interprétables et relativement peu coûteuses en calcul, ce qui les rend attractives pour un système interactif.

Leur limite principale tient toutefois à leur sensibilité aux conditions d'acquisition: éclairage non uniforme, reflets, occlusions partielles ou arrière-plan chargé peuvent perturber le seuillage, fragmenter les contours et dégrader la précision géométrique. Cette fragilité explique l'intérêt d'un complément par apprentissage profond.

Le traitement d'images numériques constitue un domaine fondamental de l'informatique visant à analyser, transformer et interpréter des images à l'aide d'algorithmes. Une image numérique peut être définie comme une fonction bidimensionnelle $f(x, y)$, où x et y représentent les coordonnées spatiales, et où la valeur de la fonction correspond à l'intensité lumineuse du pixel à cette position. Dans le cas des images en niveaux de gris, cette intensité est généralement codée sur une plage discrète, souvent comprise entre 0 et 255.

Dans un contexte de captation de tableaux pédagogiques, l'image constitue une projection du monde réel, influencée par de nombreux facteurs tels que l'éclairage ambiant, la qualité du capteur, l'angle de prise de vue et la présence de bruit. L'objectif du traitement d'images est alors de réduire l'impact de ces perturbations afin de mettre en évidence les structures pertinentes, notamment la surface du tableau et ses contours.

2.2.1 Seuillage et segmentation

Dans le cadre de ce mémoire, la segmentation a pour objectif d'isoler automatiquement la région correspondant au tableau pédagogique à partir de l'image acquise. Cette opération consiste à distinguer les pixels appartenant au tableau de ceux correspondant à l'arrière-

plan (mur, plafond, mobilier ou autres objets présents dans la salle). Dans l'approche proposée, la segmentation permet également d'identifier les frontières du tableau sous forme d'un contour fermé. Cette représentation est particulièrement importante puisqu'elle sert ensuite à la détection des coins, au redressement géométrique et à l'extraction de la région d'intérêt.

Parmi les techniques de segmentation les plus simples figurent les méthodes de seuillage. Le seuillage repose sur l'hypothèse que les pixels appartenant à des régions différentes présentent des niveaux d'intensité distincts. Un seuil T est alors utilisé pour séparer les pixels de l'image en deux classes; dans le cas d'un tableau, cette opération peut constituer une première approximation de la région d'intérêt lorsque le contraste entre le tableau et l'arrière-plan est suffisant (Gonzalez et Woods, 2018).

Dans les méthodes classiques de traitement d'images, le seuil T peut être déterminé de manière globale ou locale. Lorsque l'éclairage varie fortement dans l'image, un seuil unique devient souvent insuffisant. Le seuillage adaptatif constitue alors une alternative plus robuste, puisqu'il calcule automatiquement un seuil différent pour chaque région de l'image en fonction de son voisinage local. Cette stratégie permet de mieux préserver les contours des objets d'intérêt lorsque les conditions d'éclairage sont non uniformes.

Dans les expérimentations préliminaires fondées sur les méthodes classiques de traitement d'images, le seuillage adaptatif a été utilisé afin de mettre en évidence la région du tableau avant l'étape de détection des contours. Toutefois, les résultats obtenus se sont révélés sensibles aux variations d'éclairage et aux reflets présents sur la surface du tableau. Ces limitations ont motivé l'intégration ultérieure d'une approche basée sur l'apprentissage profond.

Le seuillage global constitue l'une des techniques les plus simples de segmentation d'images. Son principe consiste à comparer l'intensité de chaque pixel à une valeur seuil unique T. Les pixels dont l'intensité est supérieure au seuil sont affectés à une classe, tandis que les autres sont affectés à une seconde classe. Cette méthode est particulièrement efficace lorsque l'objet d'intérêt présente un contraste important avec son arrière-plan.

Contrairement au seuillage global, le seuillage adaptatif calcule un seuil local pour chaque région de l'image. Cette stratégie permet de mieux prendre en compte les variations

d'éclairage et les ombres présentes dans la scène. Elle est particulièrement utile dans les environnements réels où les conditions lumineuses sont rarement uniformes.

Les méthodes basées sur les contours reposent sur la détection des discontinuités d'intensité dans l'image. Les opérateurs tels que Sobel, Prewitt ou Canny permettent d'identifier les frontières entre les objets et leur arrière-plan. Dans le contexte de ce mémoire, ces méthodes peuvent être utilisées pour mettre en évidence les limites du tableau pédagogique.

Les approches fondées sur les régions connectées regroupent les pixels présentant des caractéristiques similaires en termes d'intensité ou de texture. Elles permettent d'identifier des zones homogènes dans l'image et sont souvent utilisées pour compléter les méthodes de seuillage ou de détection de contours.

Les méthodes classiques de segmentation reposent sur différentes hypothèses relatives à la distribution des intensités et à la continuité spatiale des régions. Selon Gonzalez et Woods (2018), le seuillage global constitue une solution simple et efficace lorsque l'histogramme de l'image présente une séparation nette entre l'objet et l'arrière-plan. Toutefois, sa performance diminue en présence d'éclairage non uniforme. Les méthodes adaptatives améliorent cette robustesse en calculant des seuils locaux (Gonzalez et Woods, 2018). De leur côté, les approches fondées sur les contours permettent une localisation plus précise des frontières des objets, mais restent sensibles au bruit et nécessitent généralement une étape de prétraitement (Sonka et al., 2015). Enfin, les méthodes basées sur les régions connectées dépendent fortement de la qualité de la segmentation initiale et du choix des critères de regroupement.

Le seuillage transforme une image en niveaux de gris f(x,y) en une image binaire g(x,y). Dans sa forme globale, un pixel est affecté au premier plan lorsque son intensité dépasse un seuil T; le seuil d'Otsu peut être estimé automatiquement en maximisant la séparation entre classes (Otsu, 1979; Gonzalez et Woods, 2018).

$$g(x,y) = 1 \text{ si } f(x,y) \geq T; 0 \text{ sinon} \qquad (2.1)$$

Dans cette expression, f(x,y) est l'intensité du pixel d'entrée, g(x,y) sa valeur binaire et T le seuil.

La figure 2.1 illustre l'effet du seuillage sur une image réelle du jeu de données.

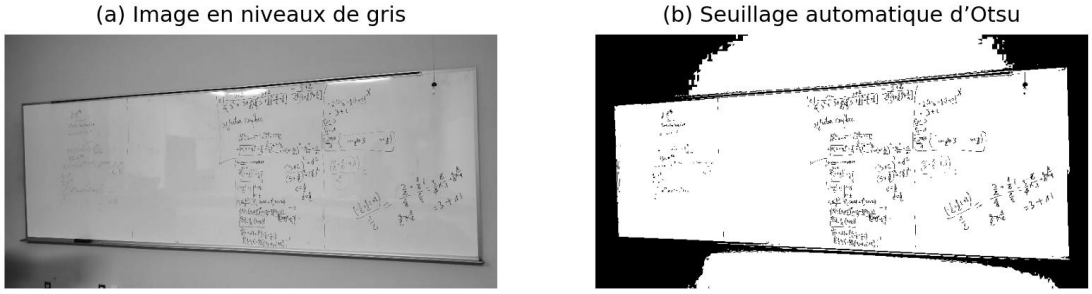

Figure 2.1 - Segmentation par seuillage : (a) niveaux de gris; (b) seuillage d'Otsu

Tableau 2.2 - Principales méthodes classiques de segmentation

Méthode	Principe	Avantages	Limites
Seuillage global	Seuil fixe	Simple, rapide	Sensible à l'éclairage
Seuillage adaptatif	Seuil local	Robuste aux variations	Paramétrage nécessaire
Segmentation par contours	Détection de gradients	Bonne localisation	Bruit possible
Régions connectées	Regroupement spatial	Identification d'objets	Dépend qualité initiale

Le tableau 2.2 met en relation chaque méthode avec ses avantages et ses limites en présence d'éclairage variable.

2.2.2 Morphologie mathématique

La morphologie mathématique est une branche du traitement d'images fondée sur la théorie des ensembles et sur l'utilisation d'un élément structurant, c'est-à-dire un petit motif géométrique servant à sonder localement la forme des objets (Serra, 1982). Elle est particulièrement adaptée à l'extraction et au renforcement des structures géométriques, ce qui en fait un outil pertinent pour la détection de tableaux pédagogiques.

Les opérations morphologiques fondamentales sont l'érosion et la dilatation. L'érosion retire les pixels situés sur la frontière d'un objet selon la forme de l'élément structurant; la dilatation ajoute au contraire des pixels autour de cet objet. À partir de ces deux opérations de base, on définit l'ouverture, utile pour éliminer de petites structures parasites, et la fermeture, utile pour combler des discontinuités dans les contours.

Dans le contexte de ce travail, les opérations morphologiques jouent un rôle essentiel dans le renforcement des bords du tableau et dans l'élimination des éléments parasites présents dans l'image, tels que les inscriptions environnantes ou les objets de la salle de classe.

La morphologie mathématique constitue un ensemble d'opérations non linéaires destinées à analyser la forme des objets présents dans une image. Ces opérations reposent sur l'utilisation d'un élément structurant B, déplacé sur l'image binaire A afin d'en modifier localement la géométrie. Les deux opérations fondamentales sont l'érosion et la dilatation, à partir desquelles sont construites les autres transformations morphologiques. L'érosion est définie par l'expression suivante:

$$A \ominus B = \{x \in \mathbb{Z}2 \mid Bx \subseteq A\}$$

Les éléments considérés sont les suivants: A représente l'image binaire; B l'élément structurant; Bx l'élément structurant centré au point x.

L'érosion réduit la taille des objets présents dans l'image en supprimant les pixels situés sur leurs frontières. Elle est principalement utilisée pour éliminer les petites structures parasites et le bruit.

$$A \oplus B = \{x \mid (B)x \cap A \neq \emptyset\}$$

La dilatation produit l'effet inverse de l'érosion. Elle agrandit les régions présentes dans l'image et permet de renforcer les contours ou de combler de petites lacunes.

$$A \circ B = (A \ominus B) \oplus B$$

L'ouverture correspond à une érosion suivie d'une dilatation. Elle permet d'éliminer les petits objets isolés tout en conservant la structure globale des régions importantes.

La fermeture est l'opération inverse. Elle est utilisée pour combler les petites discontinuités présentes dans les contours et stabiliser la géométrie des objets détectés.

Les opérations morphologiques constituent un outil structurant dans le pipeline de traitement d'images. L'érosion et la dilatation permettent de manipuler la géométrie des objets détectés, tandis que l'ouverture et la fermeture améliorent la cohérence structurelle des régions d'intérêt. Dans le cadre de la détection de tableaux, ces opérations contribuent à

renforcer les contours dominants tout en réduisant l'impact du bruit et des artefacts présents dans l'image. Leur utilisation améliore ainsi la stabilité des étapes ultérieures de détection géométrique.

La morphologie mathématique transforme les régions binaires à l'aide d'un élément structurant. L'érosion réduit les régions, la dilatation les étend, l'ouverture retire les petits objets et la fermeture comble certaines discontinuités (Serra, 1982; Haralick et al., 1987).

La figure 2.2 présente ces quatre opérations sur un exemple synthétique reproductible.

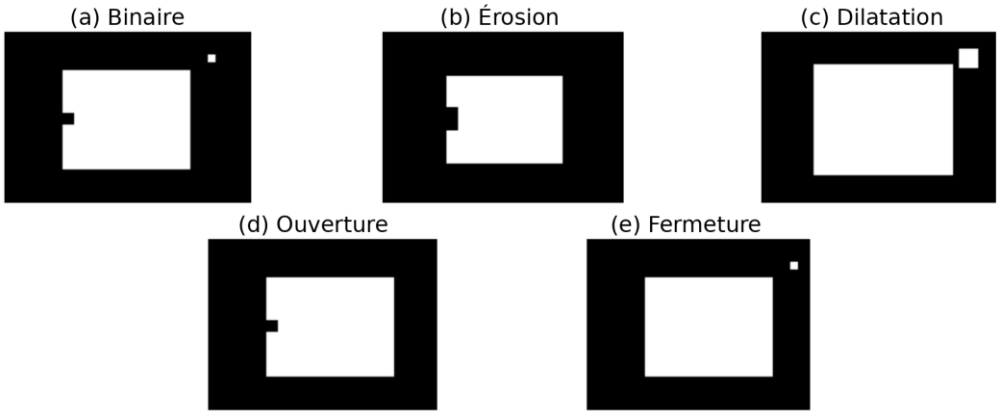

Figure 2.2 - Effets de l'érosion, de la dilatation, de l'ouverture et de la fermeture

Tableau 2.3 - Rôle des opérations morphologiques dans le pipeline

Opération	Définition	Effet principal	Application dans ce mémoire
Érosion	Réduction des régions selon un élément structurant	Supprime petits objets	Nettoyage du bruit
Dilatation	Expansion des régions	Renforce structures	Renforcement des bords
Ouverture	Érosion suivie de dilatation	Élimine petits artefacts	Prétraitement
Fermeture	Dilatation suivie d'érosion	Comble discontinuités	Stabilisation des contours

Le tableau 2.3 précise comment chaque opération intervient dans le nettoyage et la stabilisation des contours.

2.2.3 Détection de contours et de lignes

La détection de contours constitue une étape essentielle pour localiser les frontières entre différentes régions d'une image. Les contours correspondent à des discontinuités d'intensité, souvent associées aux limites physiques des objets présents dans la scène.

Parmi les méthodes de détection de contours, l'algorithme proposé par Canny (1986) demeure l'une des références les plus utilisées en vision par ordinateur. Son objectif est d'identifier les transitions importantes d'intensité dans une image, transitions généralement associées aux frontières physiques des objets présents dans la scène.

Le fonctionnement du détecteur de Canny repose sur quatre étapes principales.

La première étape est le lissage de l'image.

Dans un premier temps, l'image est filtrée à l'aide d'un filtre gaussien afin de réduire le bruit susceptible de provoquer la détection de faux contours.

$$I\_G = I * G \tag{2.2}$$

Dans l'équation (2.2), I_G est l'image filtrée; I, l'image originale; G, le noyau gaussien; * désigne la convolution.

Dans le contexte de ce mémoire, cette étape permet de limiter l'influence des imperfections présentes sur les surfaces du tableau ou sur le mur environnant.

La deuxième étape est le calcul du gradient.

Une fois le bruit réduit, l'algorithme calcule les variations locales d'intensité. Dans les équations (2.3) et (2.4), G_x et G_y sont les gradients selon x et y; I désigne l'image.

$$G_x = \partial I / \partial x \tag{2.3}$$

$$G_y = \partial I / \partial y \tag{2.4}$$

L'amplitude du gradient, $M(x, y)$, est donnée par l'équation (2.5):

$$M(x, y) = (G_x^{\,2} + G_y^{\,2})^{1/2} \tag{2.5}$$

Les fortes valeurs du gradient indiquent généralement la présence d'un contour.

La troisième étape est la suppression des non-maxima.

Après le calcul du gradient, plusieurs pixels voisins peuvent appartenir au même contour.

La suppression des non-maxima consiste à conserver uniquement les pixels présentant la plus forte intensité de gradient dans la direction du contour.

Cette opération produit des contours plus fins et mieux localisés.

La quatrième étape est le seuillage par hystérésis.

Enfin, deux seuils sont utilisés: TH est le seuil fort et TL le seuil faible.

Les pixels dont le gradient dépasse *TH* sont conservés comme contours certains.

Les pixels dont le gradient est compris entre *TL* et *TH* sont conservés uniquement s'ils sont connectés à un contour fort.

Cette stratégie réduit considérablement le nombre de faux contours.

L'ensemble du processus met ainsi en évidence les structures géométriques dominantes, notamment les bords du tableau, et prépare l'image aux traitements ultérieurs d'analyse de formes et de localisation.

Une caractéristique importante du tableau pédagogique est sa forme généralement rectangulaire. Les limites du tableau sont donc souvent représentées dans l'image par des segments de droite horizontaux et verticaux. La détection de ces lignes constitue une étape importante pour localiser précisément le tableau et identifier ses coins.

Parmi les méthodes les plus utilisées pour détecter des lignes dans une image, la transformée de Hough occupe une place importante. Introduite par Duda et Hart (1972), cette méthode permet d'identifier automatiquement des droites même lorsque leurs contours sont partiellement interrompus ou affectés par du bruit.

Le principe de la transformée de Hough consiste à représenter chaque point détecté dans l'image sous la forme d'un ensemble de droites possibles dans un espace de paramètres.

Une droite peut être écrite sous la forme $y = ax + b$, où a représente sa pente et b son ordonnée à l'origine. Cette forme devient instable pour les droites verticales; la transformée de Hough utilise donc la représentation polaire donnée à l'équation (2.6).

Chaque pixel appartenant à un contour vote alors pour plusieurs couples (ρ, θ). Lorsque plusieurs pixels sont alignés sur une même droite, leurs votes se concentrent dans une même région de l'espace paramétrique.

Les maxima observés dans cet espace correspondent ainsi aux droites les plus probables présentes dans l'image.

Considérons une image contenant un tableau rectangulaire. Après application du détecteur de contours de Canny, les bords supérieur, inférieur, gauche et droit du tableau apparaissent sous la forme de segments lumineux.

Chaque pixel appartenant à ces contours participe au vote dans l'espace de Hough. Les votes associés aux bords du tableau se renforcent mutuellement, ce qui produit quatre maxima dominants correspondant aux principales lignes du rectangle.

Ces lignes peuvent ensuite être utilisées pour déterminer les coins du tableau et calculer la transformation de perspective permettant de redresser l'image.

Cette étape est déterminante pour la localisation précise des limites du tableau et pour l'extraction ultérieure de ses quatre coins en vue de la correction de perspective.

Le détecteur de Canny combine lissage gaussien, calcul du gradient, suppression des non-maxima et seuillage par hystérésis (Canny, 1986). Les opérateurs de Sobel et de Prewitt constituent des alternatives plus simples pour approximer les dérivées de l'image (Sobel et Feldman, 1968; Prewitt, 1970).

La figure 2.3 montre que Canny conserve les structures dominantes du tableau tout en réduisant les variations de texture.

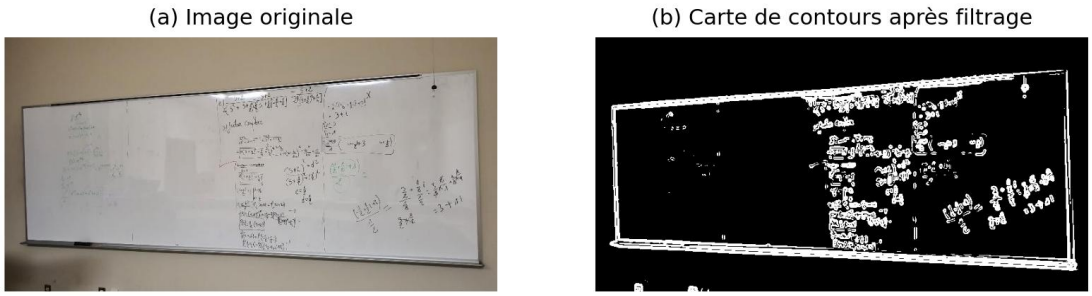

Figure 2.3 - Détection de contours : (a) image originale; (b) résultat de Canny

La transformée de Hough représente une droite par la paire (ρ,θ), où ρ est la distance de la droite à l'origine et θ l'angle de sa normale. Chaque pixel de contour vote pour les couples compatibles; les maxima de l'accumulateur correspondent aux lignes dominantes (Duda et Hart, 1972; Ballard, 1981).

$$\rho = x\cos(\theta) + y\sin(\theta) \qquad (2.6)$$

Les variables x et y désignent les coordonnées d'un pixel de contour; ρ est la distance perpendiculaire entre l'origine et la droite, tandis que θ est l'angle de la normale à cette droite par rapport à l'axe horizontal.

La figure 2.4 présente les segments dominants détectés sur l'image du tableau.

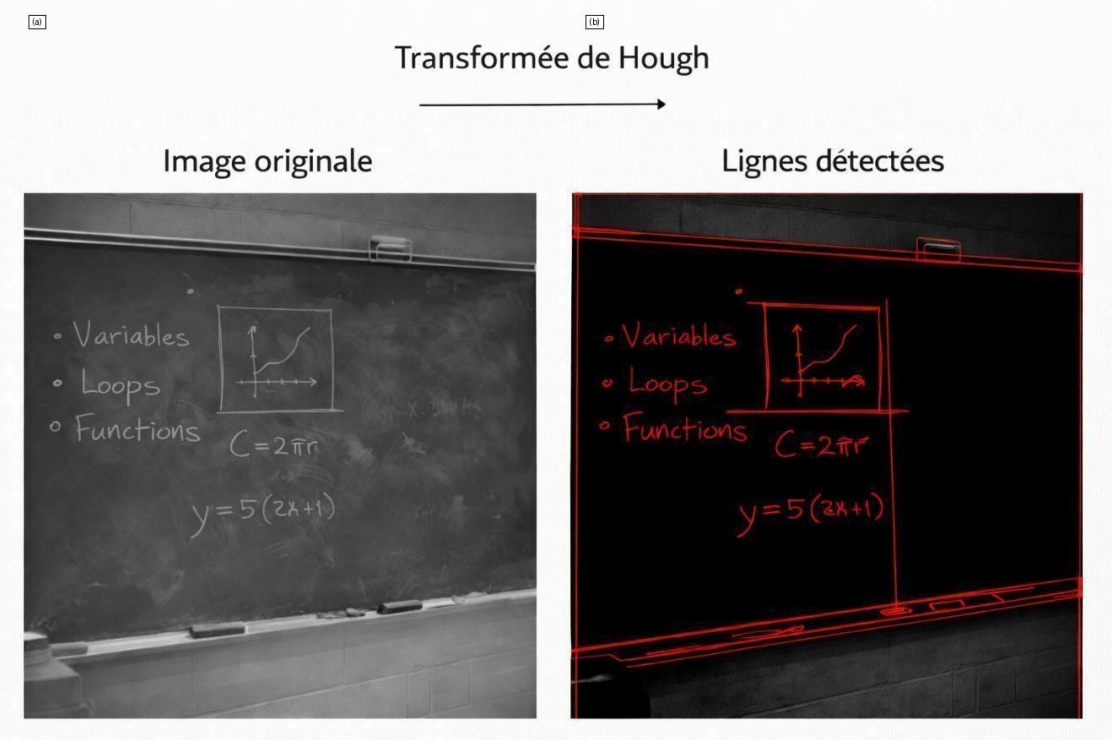

Figure 2.4 - Transformée de Hough : (a) image originale; (b) lignes détectées

2.2.4 Pipeline classique et correction de perspective

Les méthodes classiques de détection de tableaux reposent sur l'exploitation des propriétés visuelles et géométriques caractéristiques de cet objet. Dans un environnement pédagogique, le tableau se distingue généralement par une surface plane, de forme rectangulaire, présentant un contraste relatif avec l'arrière-plan. Ces caractéristiques permettent d'envisager une détection fondée sur l'analyse de l'intensité lumineuse, des contours et des structures géométriques.

Dans ce travail, l'approche classique vise à détecter automatiquement la région correspondant au tableau à partir d'une image acquise dans des conditions réelles de salle de classe. L'objectif n'est pas uniquement de localiser le tableau, mais également de préparer une extraction précise et exploitable de son contenu. L'objectif final du pipeline classique n'est pas uniquement de détecter la présence du tableau, mais également de

produire une image rectifiée de celui-ci. L'image obtenue à la sortie du traitement correspond à une vue frontale du tableau dans laquelle les distorsions de perspective ont été corrigées. Cette image peut ensuite être utilisée pour l'amélioration du contraste, l'extraction du contenu écrit ou l'archivage numérique des notes pédagogiques.

Le pipeline commence par un prétraitement visant à améliorer la qualité de l'image et à réduire le bruit. Il se poursuit par la détection de contours afin d'identifier les transitions significatives d'intensité correspondant aux frontières du tableau. La détection de lignes dominantes, notamment par la transformée de Hough, permet ensuite de localiser les structures géométriques principales. Enfin, une étape de finalisation assure l'extraction de la région d'intérêt et la correction de perspective. Ce pipeline constitue la base méthodologique des approches classiques utilisées avant l'intégration de modèles d'apprentissage profond.

Le pipeline classique présenté met en évidence une progression logique allant du prétraitement à la correction géométrique. Chaque étape contribue à la structuration progressive de l'information visuelle. Toutefois, la performance globale du pipeline dépend fortement de la qualité des étapes intermédiaires. Une erreur dans la détection de contours peut se propager jusqu'à la localisation des coins. Cette sensibilité justifie l'intégration ultérieure d'un module de validation par apprentissage profond.

Le prétraitement constitue une étape essentielle dans toute application de traitement d'images, en particulier lorsque les données sont acquises dans des environnements non contrôlés. Les images de tableaux pédagogiques peuvent être affectées par du bruit, des variations d'éclairage ou des artefacts visuels liés au capteur.

Dans les approches classiques de traitement d'images, une conversion en niveaux de gris est généralement réalisée afin de simplifier les calculs et de mettre en évidence les variations d'intensité utilisées par les algorithmes de détection de contours. Cette étape permet de réduire la complexité des traitements tout en conservant les informations structurelles essentielles. Cependant, dans l'approche basée sur l'apprentissage profond retenue dans ce mémoire, le modèle YOLOv8-seg utilise directement les images couleur acquises par la caméra. La conversion en niveaux de gris n'est donc utilisée que dans certaines étapes du pipeline classique étudié à titre comparatif.

Le prétraitement peut également inclure des opérations de normalisation ou d'amélioration du contraste, notamment lorsque l'éclairage est inhomogène. Ces opérations visent à rendre la surface du tableau plus homogène et à accentuer ses limites par rapport au reste de la scène.

Les différentes étapes du pipeline classique décrites précédemment nécessitent le réglage de plusieurs paramètres influençant directement la qualité de la détection du tableau. Afin d'obtenir un compromis satisfaisant entre réduction du bruit, préservation des contours et stabilité géométrique, plusieurs configurations ont été évaluées sur un sous-ensemble représentatif des images du jeu de données. Les paramètres finalement retenus sont présentés dans le tableau récapitulatif.

Les paramètres présentés dans le tableau récapitulatif ont été déterminés à partir d'expérimentations préliminaires réalisées sur un ensemble d'images représentatives du contexte de captation étudié. Le noyau gaussien de taille 5 × 5 a été retenu car il permet de réduire efficacement le bruit sans provoquer une perte excessive d'information sur les contours du tableau. Les seuils 50 et 150 utilisés par le détecteur de Canny offrent un compromis satisfaisant entre la détection des contours significatifs et l'élimination des faux contours liés aux écritures, aux reflets ou aux irrégularités du mur. Enfin, l'élément structurant 3 × 3 utilisé lors des opérations morphologiques permet de renforcer la continuité des contours tout en préservant la géométrie générale du tableau.

Ces paramètres correspondent directement aux valeurs utilisées dans l'implémentation OpenCV développée dans ce mémoire pour les étapes de prétraitement, de détection de contours et de traitement morphologique.

Une fois l'image prétraitée, la détection de contours constitue une étape clé pour identifier les limites du tableau. Les contours correspondent à des variations significatives d'intensité et sont généralement associés aux frontières physiques des objets présents dans la scène.

Des algorithmes de détection de contours, tels que le détecteur de Canny, sont utilisés pour extraire ces discontinuités. Le résultat de cette étape est une image binaire mettant en évidence les bords des objets. Dans le cas du tableau, ses contours apparaissent souvent

comme des lignes continues et relativement longues, ce qui permet de les distinguer des contours plus courts ou irréguliers correspondant à des éléments parasites.

Après la détection de contours, une analyse des composantes connexes peut être réalisée afin de filtrer les régions non pertinentes. Les contours de petite taille ou présentant une géométrie incohérente sont éliminés, tandis que les structures dominantes sont conservées pour l'analyse géométrique.

L'image originale est d'abord soumise au détecteur de contours de Canny afin d'extraire les transitions significatives d'intensité. Les contours obtenus servent ensuite d'entrée à la transformée de Hough, qui permet d'identifier les principales lignes présentes dans la scène. Les lignes détectées correspondent majoritairement aux bords du tableau et constituent la base de la localisation géométrique utilisée dans le pipeline classique.

La forme rectangulaire du tableau constitue un indice géométrique fort pouvant être exploité pour sa détection. À partir des contours extraits, une approximation polygonale permet de simplifier les formes détectées; les candidats pertinents sont ensuite retenus lorsqu'ils présentent quatre sommets, des côtés relativement longs et une cohérence géométrique avec un quadrilatère convexe occupant une surface significative de l'image.

L'analyse des segments permet alors d'identifier les lignes susceptibles de correspondre aux bords du tableau. Les lignes dominantes sont sélectionnées en fonction de leur longueur, de leur orientation et de la densité de points de contour qui les soutiennent. Les intersections entre les lignes horizontales et verticales les plus cohérentes fournissent une estimation des coins.

La transformée de Hough peut également être utilisée pour renforcer la détection des lignes droites, en particulier lorsque les contours sont partiellement discontinus. En combinant les résultats de ces différentes techniques, une estimation robuste des quatre coins du tableau peut être obtenue.

Dans un contexte réel, les images du tableau sont rarement acquises selon une vue parfaitement frontale. Les variations d'angle de prise de vue introduisent des distorsions de perspective qui peuvent affecter la lisibilité du contenu extrait.

Une fois les quatre coins du tableau identifiés, une transformation projective est appliquée afin de corriger ces distorsions. Cette opération repose sur le calcul d'une homographie H telle que x' ~ Hx, où x représente un point de l'image initiale et x' le point correspondant dans l'image redressée. On obtient ainsi une représentation normalisée du tableau, plus adaptée à l'archivage et aux traitements ultérieurs.

La correction de perspective constitue une étape déterminante pour garantir la qualité du système de captation, car elle permet de produire une image du tableau homogène et cohérente, indépendamment des conditions d'acquisition initiales.

Les méthodes classiques de détection de tableaux, telles que le prétraitement photométrique, la détection de contours, l'analyse géométrique et la correction de perspective, présentent plusieurs avantages. Elles reposent sur des principes bien établis, sont interprétables et relativement peu coûteuses en termes de calcul. Elles permettent également de comprendre précisément les mécanismes de détection et d'identifier les causes d'éventuelles erreurs.

Cependant, leur efficacité dépend fortement des conditions visuelles. Les variations d'éclairage, les reflets sur la surface du tableau, les occlusions partielles causées par l'enseignant et la complexité de l'arrière-plan peuvent dégrader significativement les performances du système. Dans certains cas, les contours du tableau peuvent être partiellement masqués ou confondus avec d'autres structures de la scène.

Ces limitations soulignent la nécessité d'enrichir l'approche classique par des techniques plus robustes, pouvant généraliser face à la variabilité des données. C'est dans cette perspective que l'intégration de méthodes basées sur l'apprentissage profond est envisagée dans la suite du mémoire.

La figure 2.5 synthétise le pipeline classique sans répéter les notions détaillées précédemment.

Figure 2.5 - Pipeline classique de détection et de rectification du tableau

Tableau 2.4 - Étapes du pipeline classique

Étape	Objectif	Outils utilisés
Conversion en niveaux de gris	Simplification	OpenCV
Filtrage gaussien	Réduction du bruit	cv2.GaussianBlur
Détection de contours	Extraction des bords	Canny
Détection de lignes	Identification géométrique	Hough
Approximation polygonale	Extraction quadrilatère	approxPolyDP
Correction de perspective	Normalisation	warpPerspective

Le tableau 2.4 associe chaque étape à son objectif et à l'opération OpenCV correspondante (Bradski, 2000).

Tableau 2.5 - Paramètres retenus pour le prétraitement classique

Paramètre	Valeur retenue	Unité ou interprétation
Noyau gaussien	5 × 5	pixels
Seuil Canny inférieur	50	niveau d'intensité
Seuil Canny supérieur	150	niveau d'intensité
Élément structurant	3 × 3	pixels

Le tableau 2.5 distingue explicitement les valeurs réellement retenues et leurs unités.

$$x' \sim Hx \tag{2.7}$$

La matrice d'homographie H relie un point homogène x de l'image originale au point rectifié x′, à un facteur d'échelle près (Hartley et Zisserman, 2004).

2.3 Apprentissage profond, segmentation et YOLO

Avec l'essor de l'apprentissage profond, les réseaux de neurones convolutifs se sont imposés comme des modèles performants pour l'analyse d'images, notamment en détection et en segmentation d'objets (Goodfellow, Bengio et Courville, 2016). Ces modèles apprennent automatiquement des représentations discriminantes à partir de grandes quantités de données, ce qui leur permet de mieux généraliser à des situations complexes.

Dans le contexte de la captation de tableaux, les approches profondes permettent d'apprendre des représentations plus robustes aux variations d'éclairage, de perspective et d'encombrement visuel, à condition de disposer d'un jeu de données annoté représentatif. Les modèles de la famille YOLO se distinguent en particulier par leur compromis entre rapidité et précision, ce qui les rend adaptés à une application interactive (Ali et Zhang, 2024).

En contrepartie, ces méthodes exigent une phase d'entraînement supervisé, une annotation rigoureuse des données et des ressources de calcul plus importantes. Leur interprétation reste également moins directe que celle d'un pipeline purement géométrique.

Au-delà des modèles de détection d'objets, plusieurs travaux récents se sont intéressés à l'amélioration de la segmentation, à la précision des contours et à l'intégration de contraintes de forme. Ces travaux sont particulièrement pertinents dans le contexte de la captation de tableaux pédagogiques, où la localisation précise des frontières du tableau conditionne la qualité de la correction de perspective et de l'extraction du contenu.

Des travaux récents en segmentation d'images montrent l'importance de la qualité des frontières pour l'extraction précise des objets. Kirillov et al. (2020) proposent PointRend, une approche qui traite la segmentation comme un problème de rendu et améliore la précision des contours en raffinant les prédictions aux zones incertaines. Cette idée est pertinente pour la captation de tableaux pédagogiques, car la correction de perspective dépend directement de la localisation précise des bords du tableau.

Kuo et al. (2019) introduisent ShapeMask, une méthode de segmentation qui exploite des connaissances de forme afin de raffiner progressivement les masques d'objets. Cette approche rejoint la problématique du présent mémoire, où le tableau possède une structure géométrique attendue, généralement rectangulaire.

Xie et al. (2021) présentent SegFormer, une architecture efficace de segmentation sémantique fondée sur les Transformers. Bien que cette approche ne soit pas spécifiquement conçue pour les tableaux pédagogiques, elle illustre l'évolution des méthodes modernes vers des modèles combinant précision de segmentation et efficacité computationnelle.

Les travaux sur les métriques centrées sur les frontières, tels que Boundary IoU, soulignent que l'évaluation classique par IoU peut être insuffisante lorsque la qualité des contours est déterminante. Cette observation est directement liée au système proposé, puisque la précision des frontières influence l'extraction des coins et la correction de perspective.

L'apprentissage profond a profondément transformé le domaine de la vision par ordinateur au cours des dernières années. Les réseaux de neurones convolutifs (Convolutional Neural Networks – CNN) se sont imposés comme des modèles particulièrement efficaces pour l'analyse d'images, grâce à leur capacité à apprendre automatiquement des représentations hiérarchiques à partir de données brutes.

Contrairement aux méthodes classiques, qui reposent sur des règles explicites et des transformations déterminées par l'expert, les CNN apprennent directement les caractéristiques pertinentes à partir d'exemples annotés. Un réseau convolutif est composé de couches successives de convolution, d'activation non linéaire et de sous-échantillonnage, permettant d'extraire progressivement des caractéristiques de plus en plus abstraites. Les premières couches détectent généralement des motifs simples tels que des bords ou des textures, tandis que les couches profondes identifient des structures complexes correspondant à des objets entiers.

Dans le cadre de la détection de tableaux pédagogiques, cette capacité d'apprentissage est particulièrement intéressante, car elle permet au modèle de s'adapter aux variations d'éclairage, de perspective et de contexte présentes dans les salles de classe.

La détection d'objets consiste à localiser et identifier des objets spécifiques au sein d'une image. Elle se distingue de la simple classification, qui attribue une étiquette à l'image entière, en fournissant également des informations spatiales sous forme de boîtes

englobantes (bounding boxes). La segmentation, quant à elle, vise à déterminer précisément les pixels appartenant à un objet donné.

Dans le cas de la captation de tableaux, la détection permet d'identifier la région approximative du tableau dans l'image, tandis que la segmentation offre une délimitation plus précise de sa surface. Ces deux tâches sont complémentaires et peuvent être combinées afin d'obtenir une localisation robuste et fine de la zone d'intérêt.

YOLO (You Only Look Once) est une famille de modèles d'apprentissage profond conçue pour localiser et reconnaître automatiquement des objets dans une image. Son nom, qui signifie littéralement « on ne regarde qu'une seule fois », indique que le modèle analyse l'image en une seule passe du réseau neuronal pour prédire à la fois la position des objets et leur classe.

Diwan et al. (2023) présentent une revue des modèles YOLO et soulignent leur intérêt pour les applications nécessitant un compromis entre rapidité d'exécution et précision de détection. Cette caractéristique justifie leur utilisation dans un système de captation automatique de tableaux, où le traitement doit rester suffisamment rapide pour être exploitable en contexte pédagogique.

Parmi les architectures de détection d'objets, le modèle YOLO se distingue par son efficacité et sa rapidité. Contrairement à des approches plus anciennes qui traitent la détection en plusieurs étapes, YOLO reformule le problème comme une tâche de régression unique prédite en une seule passe du réseau (Redmon et al., 2016).

Le principe de YOLO repose sur la division de l'image en une grille. Pour chaque cellule de la grille, le modèle prédit un ensemble de boîtes englobantes potentielles ainsi que des scores de confiance associés aux classes d'objets. Cette approche permet une détection en temps réel, ce qui est particulièrement pertinent dans le contexte d'un système interactif de captation de tableaux.

Dans ce travail, la variante YOLOv8s-seg est utilisée afin de détecter automatiquement la présence du tableau dans l'image tout en produisant un masque de segmentation plus précis. Ce choix vise à conserver un compromis entre coût de calcul, rapidité d'inférence et précision géométrique.

Sapkota et al. (2024) montrent que YOLOv8 peut constituer une alternative efficace à des modèles de segmentation plus lourds, notamment lorsque la rapidité d'inférence et la simplicité de déploiement sont importantes. Ce constat rejoint l'objectif du présent mémoire, qui consiste à proposer une solution de captation automatique combinant précision de segmentation et coût computationnel raisonnable.

Le modèle YOLOv8s-seg utilisé dans ce travail représente un compromis pertinent entre précision et efficacité computationnelle. Son architecture légère autorise une détection rapide tout en conservant une capacité de segmentation suffisante pour le tableau. Le recours à l'apprentissage par transfert réduit en outre le besoin en données massives et accélère la convergence du modèle.

L'efficacité d'un modèle d'apprentissage profond dépend fortement de la qualité et de la diversité des données d'entraînement. Dans le cadre de ce mémoire, un ensemble d'images de tableaux pédagogiques a été constitué et annoté afin de permettre l'apprentissage supervisé du modèle. Chaque image est associée à une annotation indiquant la position du tableau sous forme de boîte englobante ou de masque de segmentation.

Le processus d'entraînement consiste à optimiser les paramètres du réseau en minimisant une fonction de perte mesurant l'écart entre les prédictions du modèle et les annotations de référence. Cette optimisation est réalisée à l'aide d'algorithmes de descente de gradient et de ses variantes.

Plusieurs paramètres influencent les performances du modèle, notamment la taille des images d'entrée, le nombre d'époques d'entraînement, la taille du lot (batch size) et le taux d'apprentissage. Un ajustement adéquat de ces paramètres est nécessaire afin d'obtenir un compromis satisfaisant entre précision, stabilité de convergence et temps de calcul.

Les paramètres d'entraînement ont été sélectionnés afin d'équilibrer stabilité de convergence et capacité de généralisation. Une taille d'image de 640 × 640 pixels conserve une résolution spatiale suffisante pour localiser correctement le tableau sans provoquer une surcharge de calcul excessive. Le choix de 100 époques permet d'atteindre une convergence satisfaisante tout en limitant le risque de surapprentissage, tandis qu'un batch size de 8 reste compatible avec les ressources matérielles disponibles. L'emploi de

YOLOv8s-seg répond enfin à un compromis explicite entre légèreté du modèle, rapidité d'inférence et précision de segmentation.

Les approches basées sur l'apprentissage profond présentent plusieurs avantages majeurs. En apprenant leurs caractéristiques directement à partir d'exemples annotés variés, elles deviennent moins sensibles que les méthodes purement géométriques aux variations d'éclairage, aux changements d'angle de prise de vue et à la complexité de l'arrière-plan. Elles peuvent ainsi généraliser à des situations nouvelles lorsque le jeu de données d'entraînement est représentatif.

Cependant, ces méthodes ne sont pas exemptes de limitations. Elles nécessitent des jeux de données annotés, parfois coûteux à constituer. Leur complexité computationnelle est plus élevée que celle des méthodes classiques, ce qui peut poser des contraintes dans des environnements à ressources limitées. Enfin, leur fonctionnement interne est souvent perçu comme moins interprétable, ce qui peut être un inconvénient dans un contexte académique où la compréhension des mécanismes est essentielle.

Ces considérations justifient l'approche adoptée dans ce mémoire, qui consiste à combiner les forces des méthodes classiques et des modèles d'apprentissage profond afin de concevoir un système hybride à la fois robuste et explicable.

Un réseau de neurones convolutif apprend des représentations hiérarchiques au moyen de convolutions, de fonctions d'activation et d'opérations de sous-échantillonnage (LeCun et al., 1998; Goodfellow et al., 2016). La figure 2.6 présente cette chaîne sans reprendre une illustration externe.

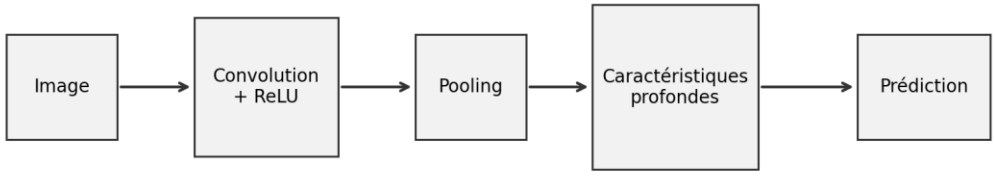

Figure 2.6 - Architecture fonctionnelle simplifiée d'un réseau de neurones convolutif

Les réseaux entièrement convolutifs, U-Net et Mask R-CNN ont établi des stratégies majeures pour la segmentation sémantique ou d'instances (Long et al., 2015; Ronneberger et al., 2015; He et al., 2017). YOLO reformule la détection comme une prédiction effectuée en une seule passe, ce qui favorise un compromis entre rapidité et précision (Redmon et al., 2016; Redmon et Farhadi, 2018).

La figure 2.7 illustre le principe de localisation appliqué à une image du jeu de données.

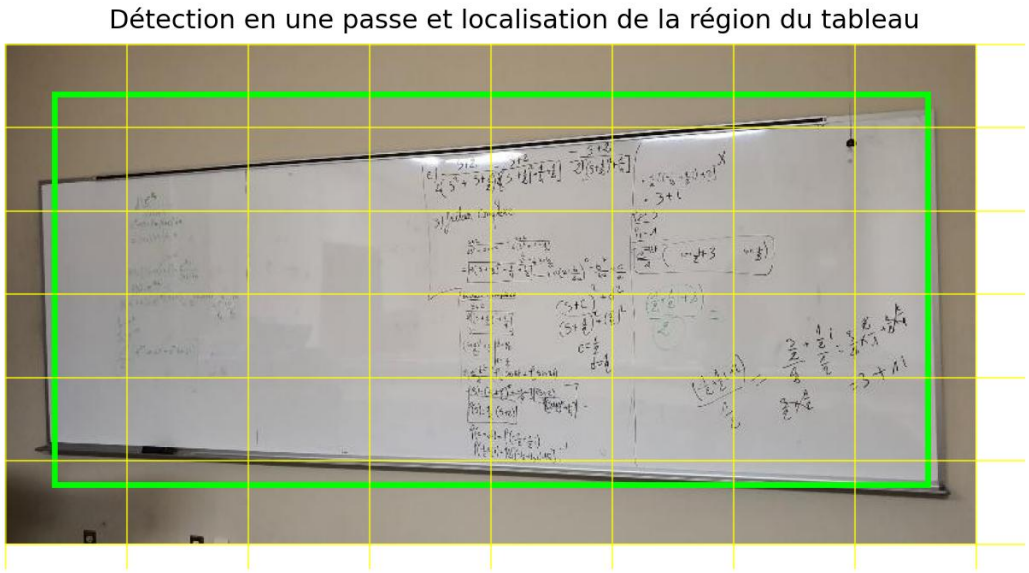

Figure 2.7 - Principe de localisation de YOLO appliqué au tableau pédagogique

Tableau 2.6 - Caractéristiques du modèle profond utilisé

Caractéristique	Choix retenu
Architecture	YOLOv8s-seg
Tâche	Détection et segmentation d'instance
Initialisation	Poids préentraînés
Cadre logiciel	PyTorch et Ultralytics 8.3.153
Classe	board (tableau)

Le tableau 2.6 résume le modèle retenu; les hyperparamètres effectivement utilisés sont reportés au chapitre 4.

2.3.1 Évolution des représentations visuelles

Avant la généralisation des réseaux profonds, la reconnaissance visuelle reposait largement sur des descripteurs conçus manuellement. SIFT recherche des points d'intérêt stables sous des changements d'échelle et d'orientation, tandis que SURF propose une approximation plus rapide fondée sur des réponses de filtres (Lowe, 2004; Bay et al., 2008). FAST vise pour sa part une détection de coins peu coûteuse, et HOG décrit la distribution locale des orientations du gradient (Rosten et Drummond, 2006; Dalal et Triggs, 2005). Ces méthodes demeurent utiles pour comprendre la géométrie d'une scène, mais elles nécessitent des règles de sélection et d'appariement qui deviennent fragiles lorsque le tableau est peu contrasté ou partiellement masqué.

L'apprentissage profond déplace une partie de cette conception vers les données. L'amélioration obtenue par AlexNet sur ImageNet a montré qu'un réseau convolutif profond pouvait apprendre des représentations discriminantes à grande échelle (Krizhevsky et al., 2012). Les architectures régionales, puis Fast R-CNN, ont ensuite amélioré la localisation d'objets en combinant propositions de régions et classification neuronale (Girshick, 2015). Cette évolution explique le choix d'un détecteur moderne pour localiser le tableau, tout en conservant des opérations géométriques explicites pour la rectification.

La famille YOLO a progressivement amélioré son compromis entre vitesse et précision. YOLOv4 a systématisé plusieurs choix d'entraînement et d'architecture permettant d'augmenter les performances sans multiplier les passes de détection (Bochkovskiy et al., 2020). YOLOv7 a poursuivi cette recherche d'efficacité au moyen de stratégies d'entraînement et d'agrégation adaptées aux détecteurs en une passe (Wang et al., 2023). La version utilisée dans ce mémoire est obtenue par l'interface Ultralytics, dont le logiciel et la configuration sont documentés par Jocher et al. (2023).

2.3.2 Jeux de référence et évaluation

Les protocoles de détection contemporains héritent de jeux de référence comme PASCAL VOC et Microsoft COCO. PASCAL VOC a contribué à formaliser l'évaluation par précision moyenne, tandis que COCO a généralisé l'analyse à plusieurs seuils d'IoU et à la segmentation d'instances (Everingham et al., 2010; Lin et al., 2014). L'IoU classique peut cependant être peu sensible à certaines erreurs de frontière; Boundary IoU complète cette lecture en accordant davantage de poids à la qualité des limites segmentées (Cheng et

al., 2021). Le présent mémoire conserve les métriques disponibles dans les journaux d'expérimentation et précise leur portée au chapitre 5.

La capture de tableaux constitue un domaine plus spécialisé que les grands jeux génériques. Des travaux consacrés à l'apprentissage en ligne ont déjà combiné localisation du tableau, amélioration de l'image et diffusion du contenu pédagogique (Gor et al., 2012). Ils confirment que la valeur d'un système ne dépend pas seulement de la détection: la rectification, la lisibilité et l'intégration au scénario de cours sont également déterminantes. Les principes généraux de formation de l'image, de calibration et de géométrie projective utilisés ici s'inscrivent dans la synthèse de Szeliski (2022).

2.4 Reconnaissance optique de caractères

La reconnaissance optique de caractères (OCR) convertit une image de texte en une représentation symbolique exploitable. Les systèmes classiques, comme Tesseract, combinent segmentation, classification et modèles linguistiques, tandis que les approches récentes utilisent des transformeurs préentraînés (Smith, 2007; Li et al., 2023). Dans le cas d'un tableau pédagogique, l'OCR doit traiter l'écriture manuscrite, les symboles mathématiques, les contrastes irréguliers et les chevauchements.

L'OCR n'a pas été implémenté ni évalué dans la version du système présentée dans ce mémoire. Il est étudié ici pour situer son rôle potentiel après la rectification du tableau et il est conservé comme perspective de recherche. Cette précision élimine l'ambiguïté entre les fonctionnalités réalisées et les fonctionnalités envisagées.

2.5 Synthèse critique et positionnement

L'analyse de l'état de l'art montre qu'aucune famille de méthodes ne répond seule à l'ensemble des contraintes du problème. Les approches classiques basées sur le traitement d'images offrent une bonne explicabilité et un coût de calcul modéré, mais leur robustesse décroît lorsque les contours sont dégradés ou lorsque le contraste est faible. Les approches basées sur l'apprentissage profond, au contraire, résistent mieux aux perturbations visuelles, mais elles dépendent d'un apprentissage supervisé, d'un jeu de données annoté et d'une phase de réglage plus coûteuse.

Une approche hybride apparaît dès lors justifiée: les méthodes classiques peuvent structurer l'analyse de la géométrie du tableau et fournir des indices interprétables, tandis que la détection profonde sert de mécanisme de validation et de renforcement lorsque les conditions d'acquisition deviennent difficiles.

C'est sur cette complémentarité, explicitée dans les chapitres suivants, que se fonde la solution proposée dans ce mémoire.

Un tableau numérique interactif (TNI) est un dispositif pédagogique composé d'un écran ou d'une surface interactive reliée à un ordinateur, permettant d'afficher, d'annoter, de manipuler et d'enregistrer du contenu numérique pendant un cours. Contrairement au tableau traditionnel, dont le contenu disparaît lorsqu'il est effacé, le TNI permet de conserver les annotations et de les partager avec les apprenants.

L'analyse comparative des solutions existantes met en évidence un compromis structurel entre coût, infrastructure et niveau d'automatisation. Les tableaux numériques interactifs offrent une solution intégrée et performante, mais leur coût élevé et leur dépendance à une infrastructure spécifique limitent leur accessibilité. À l'inverse, les méthodes informelles telles que la photographie manuelle sont accessibles mais non systématiques et fortement dépendantes de l'utilisateur. Les approches basées sur la vision par ordinateur apparaissent ainsi comme une alternative équilibrée, combinant automatisation et faible coût matériel. Cette observation justifie l'orientation adoptée dans ce mémoire vers une solution logicielle reposant sur l'analyse d'images.

La synthèse des approches issues de l'état de l'art confirme que les méthodes classiques et les approches profondes présentent des avantages complémentaires. Les premières sont interprétables et peu coûteuses, mais sensibles aux conditions réelles. Les secondes offrent une robustesse accrue au prix d'une complexité computationnelle plus élevée. L'approche hybride proposée vise précisément à exploiter cette complémentarité afin de maximiser la robustesse tout en conservant une certaine explicabilité du processus de détection.

Tableau 2.7 - Synthèse critique des familles de méthodes

Approche	Forces	Limites	Rôle dans ce mémoire
Classique	Interprétable; coût réduit	Sensible aux reflets et aux	Analyse géométrique et

Approche	Forces	Limites	Rôle dans ce mémoire
		contours incomplets	rectification
Profonde	Robuste aux variations visuelles	Dépend des données annotées	Détection et segmentation
Hybride	Robustesse et précision spatiale	Pipeline plus complexe	Solution proposée

Le tableau 2.7 montre qu'aucune famille ne satisfait seule toutes les contraintes; l'approche hybride est donc retenue.

CHAPITRE 3
CONCEPTION DU SYSTÈME HYBRIDE

3.1 Principes de conception

La conception du système interactif de captation de tableaux repose sur une idée centrale: combiner la rigueur des méthodes classiques de traitement d'images avec la robustesse des techniques d'apprentissage profond, afin d'obtenir une solution à la fois explicable, performante et adaptée aux environnements pédagogiques réels.

Le système proposé vise à fonctionner à partir d'images fixes ou de flux vidéo capturés à l'aide d'une caméra standard, sans nécessiter d'infrastructure spécifique ni de matériel spécialisé. Cette contrainte influence fortement les choix méthodologiques adoptés. Le système doit être destiné à traiter des images présentant des variations d'éclairage, des changements de perspective et des éléments perturbateurs présents dans la scène.

La conception repose sur quatre principes directeurs étroitement liés: la modularité du système, qui permet de faire évoluer indépendamment chaque composant; la robustesse face aux conditions réelles de captation; le potentiel de traitement interactif; et la possibilité d'étendre ultérieurement la plateforme vers des fonctionnalités telles que la reconnaissance de texte ou l'archivage intelligent.

La conception vise la modularité, la traçabilité des étapes et la compatibilité avec une caméra standard. La sortie du réseau sert à isoler la région probable du tableau; les opérations géométriques déterminent ensuite les coins, calculent l'homographie et produisent une vue rectifiée.

3.1.1 Besoins fonctionnels et critères d'acceptation

Le premier besoin fonctionnel consiste à recevoir une image exploitable sans imposer un dispositif de captation propriétaire. L'entrée peut provenir d'une image enregistrée ou d'une trame extraite d'un flux vidéo. Le système doit préserver la résolution utile, vérifier que le fichier est lisible et transmettre une représentation cohérente au module de segmentation. Cette étape paraît élémentaire, mais elle conditionne toutes les opérations suivantes: une orientation incorrecte, un canal alpha inattendu ou une conversion de couleur incohérente peut modifier la prédiction et rendre la comparaison des essais difficile.

Le deuxième besoin est la localisation de la surface du tableau. Le résultat attendu n'est pas seulement une boîte englobante: un masque doit isoler la région utile afin que les limites géométriques puissent être estimées. Le critère d'acceptation associé est la présence d'un masque suffisamment continu pour permettre l'extraction d'un contour principal. La confiance de détection ne doit donc pas être interprétée seule; elle doit être accompagnée d'un contrôle de cohérence spatiale portant sur l'aire, la position et la forme générale de la région prédite.

Le troisième besoin est la production d'une vue frontale. Le système doit transformer une région observée en perspective en une image dont les bords opposés sont approximativement parallèles. Cette sortie doit conserver le contenu écrit sans introduire une déformation arbitraire. Le quatrième besoin est l'archivage: l'image rectifiée doit pouvoir être enregistrée avec un nom distinct et des informations minimales de traçabilité. Dans ce mémoire, l'archivage porte sur l'image; la génération d'un texte OCR et la diffusion dans une plateforme pédagogique demeurent hors du périmètre évalué.

Tableau 3.1 - Traçabilité des besoins fonctionnels du système

Besoin	Critère vérifiable	État dans le prototype
Acquisition	Image lisible et dimensions connues	Implémenté
Segmentation	Masque de la classe board	Implémenté et évalué
Localisation géométrique	Contour principal et quatre coins ordonnés	Implémenté
Rectification	Vue frontale obtenue par homographie	Implémenté
Archivage	Sauvegarde de l'image rectifiée	Implémenté
OCR	Texte reconnu et métriques dédiées	Non implémenté

Le tableau 3.1 distingue les fonctions réellement observables dans le prototype de celles qui relèvent des perspectives. Cette séparation évite d'attribuer à la version évaluée des capacités qui n'ont pas fait l'objet d'une expérimentation.

3.1.2 Principes de modularité et de traçabilité

La modularité permet de remplacer une étape sans réécrire l'ensemble du système. Le modèle de segmentation peut, par exemple, être remplacé par une version plus légère sans modifier la fonction d'homographie, à condition de conserver le même contrat de sortie. De

la même manière, l'algorithme d'approximation polygonale peut évoluer tout en réutilisant le masque produit par le réseau. Cette organisation réduit le couplage entre apprentissage et géométrie et facilite les essais comparatifs.

La traçabilité impose de pouvoir relier chaque résultat à une entrée, à une version de modèle et à un ensemble de paramètres. Les journaux disponibles n'ont pas conservé toutes les informations matérielles de l'entraînement; cette lacune est explicitée au chapitre 4. En revanche, le modèle, la taille d'image, le nombre d'époques, la taille du lot et la répartition des données sont consignés. Une version ultérieure devrait ajouter un identifiant d'exécution, la graine aléatoire, la version exacte des poids et l'empreinte du jeu de données.

Le système est enfin conçu pour échouer de manière explicite. Lorsqu'aucun masque satisfaisant n'est produit ou lorsque quatre coins ne peuvent pas être estimés, il est préférable de retourner un statut d'échec que de fabriquer une rectification trompeuse. Ce principe est particulièrement important dans un contexte pédagogique, car une image mal cadrée peut supprimer une partie des notes et donner l'impression que le contenu a été correctement conservé.

3.2 Architecture du pipeline hybride

L'architecture du système est organisée en modules successifs formant une chaîne de traitement cohérente. Chaque module remplit une fonction spécifique et peut être amélioré indépendamment des autres.

L'image ou le flux vidéo est d'abord acquis, puis prétraité afin d'améliorer sa qualité visuelle. La méthode classique propose une première localisation du tableau, que YOLO valide ou renforce. Le système estime ensuite les coins, extrait la région d'intérêt, corrige la perspective, puis sauvegarde ou affiche le tableau rectifié.

Cette organisation modulaire permet d'assurer une séparation claire entre les différentes étapes, facilitant l'analyse des performances et l'amélioration progressive du système.

Le choix d'une approche hybride repose sur l'analyse critique développée dans les chapitres précédents. Les méthodes classiques exploitent efficacement la géométrie du

tableau, notamment sa forme rectangulaire, la présence de contours marqués et la possibilité d'estimer une homographie. Elles sont rapides et interprétables, mais se dégradent lorsque les contours deviennent ambigus ou lorsque le contraste diminue.

À l'inverse, la détection profonde par YOLO offre une meilleure capacité de généralisation face aux perturbations visuelles, à condition de disposer d'un modèle entraîné sur des données représentatives. L'approche hybride retenue consiste donc à utiliser les méthodes classiques pour structurer l'analyse géométrique et YOLO comme mécanisme de validation et de renforcement lorsque la scène devient plus difficile à interpréter.

Cette combinaison permet de réduire les faux positifs, d'améliorer la précision de localisation et de conserver un temps de traitement raisonnable, ce qui justifie son adoption dans ce mémoire.

La figure 3.1 distingue les composants réellement implémentés de l'extension OCR non réalisée.

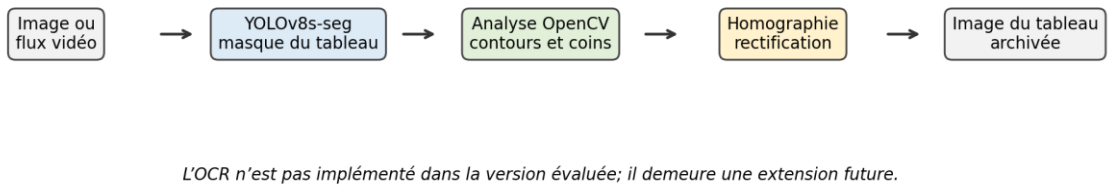

Figure 3.1 - Architecture du système hybride implémenté et position de l'extension OCR

3.2.1 Responsabilités des modules

Le module d'acquisition normalise la représentation d'entrée et transmet l'image au réseau. Le module neuronal retourne un ou plusieurs masques, des scores de confiance et des informations de classe. Le module de sélection conserve la prédiction associée au tableau et vérifie sa cohérence. Le module géométrique convertit ensuite le masque en contour, simplifie ce contour et détermine les sommets susceptibles de représenter les coins. Enfin, le module de rectification calcule la transformation projective et génère l'image frontale.

Cette décomposition rend explicite la frontière entre résultat appris et traitement déterministe. Le réseau apprend à reconnaître l'apparence du tableau à partir des annotations, alors que l'ordre des coins et l'homographie suivent des règles géométriques. Une erreur de segmentation peut donc être distinguée d'une erreur de rectification. Cette distinction est utile pour analyser les échecs: un masque absent n'appelle pas la même correction qu'un masque correct dont le contour possède trop de sommets.

L'extension OCR apparaît dans l'architecture uniquement comme consommateur potentiel de l'image rectifiée. Elle n'intervient ni dans la détection, ni dans le calcul des métriques présentées. Cette position est cohérente avec une conception en couches: l'amélioration de la qualité géométrique peut bénéficier ultérieurement à l'OCR sans que la reconnaissance de caractères soit nécessaire au fonctionnement du pipeline de captation.

3.2.2 Contrats d'entrée et de sortie

Chaque interface possède un contrat simple. L'acquisition fournit une matrice d'image et ses dimensions. La segmentation fournit un masque défini dans le repère de l'image d'entrée. La géométrie retourne quatre points ordonnés dans ce même repère. La rectification reçoit ces points et produit une nouvelle matrice dont les dimensions sont déterminées par les distances maximales entre les coins opposés. L'archivage reçoit l'image finale et un identifiant de capture.

La conservation d'un repère commun jusqu'au calcul de l'homographie réduit les erreurs de conversion. Lorsqu'une image est redimensionnée à 640 × 640 pixels pour l'inférence, les coordonnées du masque doivent être ramenées aux dimensions de l'image originale avant la rectification. Une confusion entre ces deux repères peut déplacer les coins et dégrader la qualité malgré une segmentation visuellement correcte.

Le contrat d'erreur est tout aussi important que le contrat de succès. Une sortie vide, un masque trop petit, un quadrilatère dégénéré ou une homographie non calculable doivent être signalés. Dans un prototype interactif, ce signal peut déclencher une nouvelle capture ou demander à l'utilisateur de modifier le cadrage. Il ne doit pas être transformé silencieusement en image d'apparence valide.

3.3 Étapes de traitement

Les éléments considérés sont les suivants: Acquisition et prétraitement.

L'image est capturée à partir d'une caméra positionnée dans la salle de classe. Une conversion en niveaux de gris est effectuée, suivie d'un filtrage gaussien afin de réduire le bruit. Une normalisation du contraste peut être appliquée si nécessaire.

La détection de contours est réalisée à l'aide d'un algorithme de type Canny. Les contours dominants sont ensuite analysés pour identifier des structures rectangulaires potentielles. Une approximation polygonale permet d'extraire les formes présentant quatre sommets distincts.

Le modèle YOLO est appliqué à l'image afin de détecter la présence du tableau. Si la détection est confirmée et cohérente avec les résultats de l'analyse classique, la région identifiée est retenue. En cas de divergence, des critères de confiance sont utilisés pour sélectionner la détection la plus fiable.

Les quatre coins du tableau sont estimés avec précision. Une transformation projective est ensuite appliquée pour redresser l'image et obtenir une vue frontale normalisée.

L'image redressée est affichée à l'utilisateur (enseignant), qui peut choisir de l'enregistrer ou de la rejeter. Cette dimension interactive permet de garder un contrôle humain tout en automatisant la majorité du processus.

Tableau 3.2 - Entrées, traitements et sorties du pipeline hybride

Étape	Entrée	Traitement	Sortie
Acquisition	Image ou flux vidéo	Lecture et contrôle du format	Image brute
Segmentation	Image brute	YOLOv8s-seg	Masque du tableau
Géométrie	Masque	Contours, quadrilatère, ordre des coins	Quatre coins
Rectification	Image et coins	Homographie OpenCV	Tableau frontal
Archivage	Image rectifiée	Sauvegarde horodatée	Fichier image

Le tableau 3.2 précise les responsabilités de chaque module et facilite la reproductibilité de l'implémentation.

3.3.1 Acquisition et préparation de l'image

La préparation commence par la lecture de l'image et la vérification de ses dimensions. Une image vide ou trop petite ne doit pas être transmise au modèle. La représentation des couleurs est ensuite harmonisée avec celle attendue par l'interface d'inférence. Cette conversion est distincte des opérations classiques présentées au chapitre 2: dans le pipeline profond, il n'est pas nécessaire de produire une image binaire avant la segmentation, car le réseau reçoit l'information visuelle complète.

Le redimensionnement utilisé par YOLO préserve l'image dans une fenêtre de 640 pixels. Selon l'implémentation, un remplissage peut être ajouté pour conserver le rapport d'aspect. Les coordonnées retournées doivent alors être interprétées en tenant compte de ce redimensionnement. La procédure retenue s'appuie sur les fonctions de post-traitement d'Ultralytics pour récupérer le masque dans le repère approprié, puis sur OpenCV pour les opérations géométriques.

Aucune normalisation destinée à embellir le contenu écrit n'est appliquée avant la mesure des performances. Ce choix sépare l'évaluation de la localisation du tableau de l'amélioration visuelle qui pourrait être ajoutée après rectification. Il évite aussi qu'un filtre choisi pour un type d'éclairage particulier modifie artificiellement la comparaison entre les méthodes.

3.3.2 Sélection et stabilisation du masque

Lorsque plusieurs prédictions sont présentes, le système recherche la classe board et conserve la région la plus cohérente avec la surface attendue. Le score de confiance indique

la certitude du modèle, mais l'aire relative du masque et sa position apportent une vérification complémentaire. Une petite région rectangulaire située sur un mur ne doit pas être assimilée automatiquement au tableau principal.

Le masque est converti en image binaire afin d'extraire un contour externe. Une fermeture morphologique légère peut réunir des segments proches lorsque la frontière présente de petites discontinuités. Ce traitement doit rester conservateur: une dilatation excessive pourrait inclure le mur ou le cadre voisin et déplacer les coins. L'objectif du post-traitement n'est donc pas de redessiner le tableau, mais de rendre le contour suffisamment stable pour l'approximation polygonale.

Une vérification d'aire élimine les contours négligeables. La région retenue doit aussi posséder une largeur et une hauteur non nulles. Ces contrôles sont simples, mais ils préviennent des divisions par zéro et des homographies dégénérées. Ils constituent une couche de sûreté déterministe autour de la prédiction neuronale.

3.3.3 Estimation des coins et rectification

L'approximation polygonale réduit le nombre de points du contour en fonction de son périmètre. Si quatre sommets sont obtenus, ils sont interprétés comme les coins candidats. Dans le cas contraire, le rectangle orienté minimal fournit une solution de repli. Cette seconde stratégie est moins fidèle à une frontière irrégulière, mais elle garantit quatre points et demeure adaptée à la forme générale d'un tableau.

Les quatre points doivent être ordonnés de manière stable: supérieur gauche, supérieur droit, inférieur droit et inférieur gauche. L'ordre est déterminé à partir des sommes et différences de coordonnées, puis vérifié par l'orientation du quadrilatère. Une permutation incorrecte peut retourner l'image, croiser les côtés ou produire une transformation incohérente. Cette étape est donc contrôlée avant le calcul matriciel.

La largeur de sortie est estimée par la plus grande distance entre les paires de coins supérieurs et inférieurs; la hauteur utilise les côtés gauche et droit. Les points d'arrivée forment un rectangle de ces dimensions. OpenCV calcule alors la matrice projective et rééchantillonne l'image. Le résultat est une vue frontale destinée à l'archivage et, éventuellement, à des traitements de contraste ou d'OCR.

Tableau 3.3 - Gestion des situations d'exception dans le pipeline

Situation	Détection	Réponse prévue
Aucun masque board	Sortie neuronale vide	Signaler l'échec et demander une nouvelle capture
Masque très petit	Aire sous le seuil de cohérence	Rejeter la région
Contour irrégulier	Plus de quatre sommets	Utiliser le rectangle orienté minimal
Quadrilatère dégénéré	Largeur ou hauteur quasi nulle	Ne pas calculer l'homographie
Plusieurs tableaux	Plusieurs masques valides	Conserver la région principale ou demander un choix

Le tableau 3.3 formalise les situations d'exception. Certaines réponses, comme le choix manuel entre plusieurs tableaux, relèvent d'une évolution de l'interface; le principe de ne pas produire silencieusement une image incorrecte s'applique néanmoins à la version actuelle.

3.4 Scénario d'utilisation

Le système peut être utilisé selon plusieurs scénarios complémentaires: une captation automatique à intervalles réguliers pendant le cours, un déclenchement manuel par l'enseignant lorsqu'un contenu mérite d'être conservé, ou encore une intégration dans un système de gestion pédagogique. Dans chacun de ces cas, l'objectif est de réduire la charge cognitive de l'enseignant tout en garantissant la conservation fidèle du contenu du tableau.

Dans un scénario d'usage envisagé, une caméra pourrait capturer le contenu du tableau pendant un cours en présentiel. Le prototype actuel traite les images pour détecter et rectifier le tableau, puis produit des fichiers destinés à l'archivage. La diffusion par réseau, le partage automatique et l'interaction avec des étudiants à distance ne sont pas implémentés; ils constituent des extensions futures.

Ces extensions pourraient soutenir l'accessibilité et la continuité pédagogique, mais leur faisabilité et leur utilité devront être évaluées dans un travail ultérieur.

Le système conçu dans ce travail se distingue par une architecture hybride structurée, par un équilibre entre interprétabilité et robustesse, par sa capacité à fonctionner avec du matériel standard et par une modularité qui facilite les extensions futures. Il ne s'agit donc pas simplement d'une juxtaposition de YOLO et d'OpenCV, mais d'une intégration cohérente de plusieurs paradigmes au service d'une problématique pédagogique précise.

La figure 3.2 présente la chaîne d'utilisation prévue, depuis la captation jusqu'à la consultation.

Figure 3.2 - Scénario d'utilisation du système en contexte pédagogique

3.4.1 Interaction pendant le cours

Le scénario de référence suppose une caméra positionnée de façon à voir l'ensemble du tableau. L'enseignant conserve ses pratiques d'écriture habituelles. Une capture peut être déclenchée à un moment pertinent, par exemple avant l'effacement d'une partie du contenu. Le système localise alors le tableau, rectifie la région et enregistre l'image. Cette interaction limite la charge cognitive: elle ne demande pas de redessiner les notes dans un outil spécialisé.

Une captation périodique est également envisageable, mais le temps d'inférence mesuré sur CPU ne permet pas d'assimiler le prototype à un système vidéo fluide. Le scénario retenu privilégie donc une capture ponctuelle ou à faible fréquence. Cette formulation est cohérente avec la mesure de 476,7 ms par image et évite de promettre une cadence qui n'a pas été démontrée.

Après le cours, les images rectifiées peuvent être classées avec les autres ressources pédagogiques. Le mémoire ne met pas en œuvre l'authentification, la synchronisation avec un environnement numérique d'apprentissage ni la reconnaissance du texte. Ces services constituent une couche applicative ultérieure. La contribution évaluée s'arrête à la production d'une image exploitable et traçable.

3.4.2 Exigences non fonctionnelles

La robustesse est la première exigence non fonctionnelle. Le système doit tolérer des variations raisonnables de perspective, de luminosité et de contenu écrit. Elle ne signifie

pas que toutes les scènes sont résolues; elle se mesure par la stabilité des métriques et par l'analyse des cas d'échec. La deuxième exigence est l'interprétabilité des étapes géométriques: il doit être possible d'inspecter le masque, le contour et les coins afin de localiser la source d'une erreur.

La reproductibilité impose de consigner les paramètres logiciels et les données utilisées. Le chapitre 4 fournit les éléments conservés et signale ceux qui ne le sont pas. La maintenabilité repose sur les contrats modulaires décrits précédemment. Enfin, la portabilité est favorisée par l'utilisation de Python, OpenCV et Ultralytics, mais elle doit être vérifiée sur d'autres systèmes avant d'être considérée comme acquise.

La protection des informations pédagogiques constitue une exigence future importante. Une image de tableau peut contenir des noms, des résultats ou des informations réservées au groupe. Un déploiement réel devrait définir une durée de conservation, des droits d'accès et une procédure de suppression. Ces aspects ne modifient pas les métriques de vision, mais ils conditionnent l'acceptabilité du système dans un établissement.

CHAPITRE 4
IMPLÉMENTATION ET PROTOCOLE EXPÉRIMENTAL

4.1 Environnement logiciel et matériel

L'implémentation a été réalisée en Python. OpenCV assure la lecture des images, l'extraction des contours, l'approximation polygonale et la transformation projective; Ultralytics fournit l'interface d'entraînement et d'inférence de YOLOv8s-seg. La configuration ci-dessous ne retient que les éléments vérifiables.

Tableau 4.1 - Environnement logiciel et matériel reproductible

Élément	Configuration documentée
Système	Windows 11
Processeur	AMD Ryzen 5 PRO 3500U
Python	3.11.3
Ultralytics	8.3.153
OpenCV	4.x
Modèle initial	yolov8s-seg.pt
Mesure d'inférence	CPU; 476,7 ms par image
Matériel d'entraînement	Non conservé de manière suffisamment précise dans les journaux

Le tableau 4.1 sépare les éléments vérifiés des informations non conservées. Le mémoire ne revendique donc pas une configuration GPU précise. Seul le temps d'inférence CPU est utilisé pour l'analyse de performance. Python, NumPy, PyTorch et OpenCV assurent respectivement l'orchestration, le calcul matriciel, l'apprentissage et le traitement d'images (Harris et al., 2020; Paszke et al., 2019; Bradski, 2000).

4.1.1 Organisation de l'environnement d'exécution

L'environnement est organisé autour d'un interpréteur Python et d'un ensemble de bibliothèques spécialisées. Ultralytics charge les poids, prépare les entrées et expose les masques de segmentation. PyTorch réalise les opérations tensorielles nécessaires au réseau. NumPy assure les conversions entre tableaux numériques, et OpenCV traite les images, les contours et les transformations. La séparation de ces responsabilités réduit les conversions implicites et rend plus facile l'identification d'une erreur de type ou de dimension.

Les versions de Python et d'Ultralytics sont consignées parce qu'une évolution de l'interface peut modifier le format des résultats ou la manière de charger un modèle. La mention OpenCV 4.x demeure moins précise; une reproduction stricte devrait conserver la version complète obtenue au moment de l'expérience. Le même principe s'applique à PyTorch, NumPy et aux pilotes matériels. L'absence de ces détails n'invalide pas les résultats observés, mais elle limite la reproduction bit à bit.

Le processeur AMD Ryzen 5 PRO 3500U est l'environnement documenté pour la mesure d'inférence. Cette mesure ne peut être extrapolée directement à un autre processeur, à un GPU ou à un appareil mobile. Le temps dépend du nombre de cœurs, de la fréquence, des optimisations de la bibliothèque et des opérations de lecture. La valeur rapportée doit donc être comprise comme un repère expérimental local.

4.1.2 Mesures de reproductibilité

Une exécution reproductible devrait débuter par la création d'un environnement isolé, l'installation de versions figées et la vérification de l'accès aux poids. Le fichier de configuration du jeu de données doit définir la classe board et les chemins d'entraînement et de validation. Avant l'entraînement, un contrôle doit confirmer le même nombre d'images et d'annotations, ainsi que l'absence de fichiers d'annotation vides ou illisibles.

L'entraînement devrait également enregistrer la graine aléatoire, l'empreinte des fichiers, la configuration complète et les courbes par époque. Ces éléments n'ont pas tous été conservés dans les journaux disponibles. Le présent document ne reconstruit pas des valeurs manquantes: il décrit les paramètres vérifiés et transforme les lacunes en exigences explicites pour une réplication future.

Après l'entraînement, la reproduction exige de distinguer les poids finaux des meilleurs poids sélectionnés pendant l'optimisation. Les métriques doivent être recalculées sur le même sous-ensemble de validation avec la même taille d'image. Une modification du seuil de confiance ou de l'IoU de suppression peut changer le nombre de faux positifs et doit être déclarée lorsque des résultats sont comparés.

4.2 Jeu de données et annotation

Afin d'entraîner le modèle de segmentation, un jeu de données spécifique a été constitué à partir d'images réelles de tableaux pédagogiques capturées dans différents environnements de salle de classe. Ces images présentent des variations significatives en termes d'éclairage, d'angle de prise de vue et de contexte visuel.

L'annotation des images a été réalisée à l'aide de la plateforme makesense.ai, qui permet de définir manuellement des polygones correspondant à la région du tableau. Chaque image a été annotée en traçant précisément le contour du tableau, afin d'obtenir un masque de segmentation fidèle à la réalité.

L'outil makesense.ai permet d'exporter les annotations au format JSON. Ces annotations ont ensuite été converties au format requis par YOLOv8 pour la segmentation, dans lequel chaque objet est représenté par une classe suivie des coordonnées normalisées des points du polygone.

Cette étape d'annotation constitue une phase essentielle du projet, car la qualité des annotations influence directement la performance du modèle entraîné.

Le jeu de données contient 416 images positives de tableaux capturées dans des salles réelles. Chaque image comporte au moins un tableau; aucune scène entièrement dépourvue de tableau n'a été incluse. Les images présentent des variations de perspective, d'éclairage, de distance et de contenu. Une seule classe, nommée board, est annotée par un polygone décrivant la surface du tableau.

La figure 4.1 illustre quatre environnements du jeu de données.

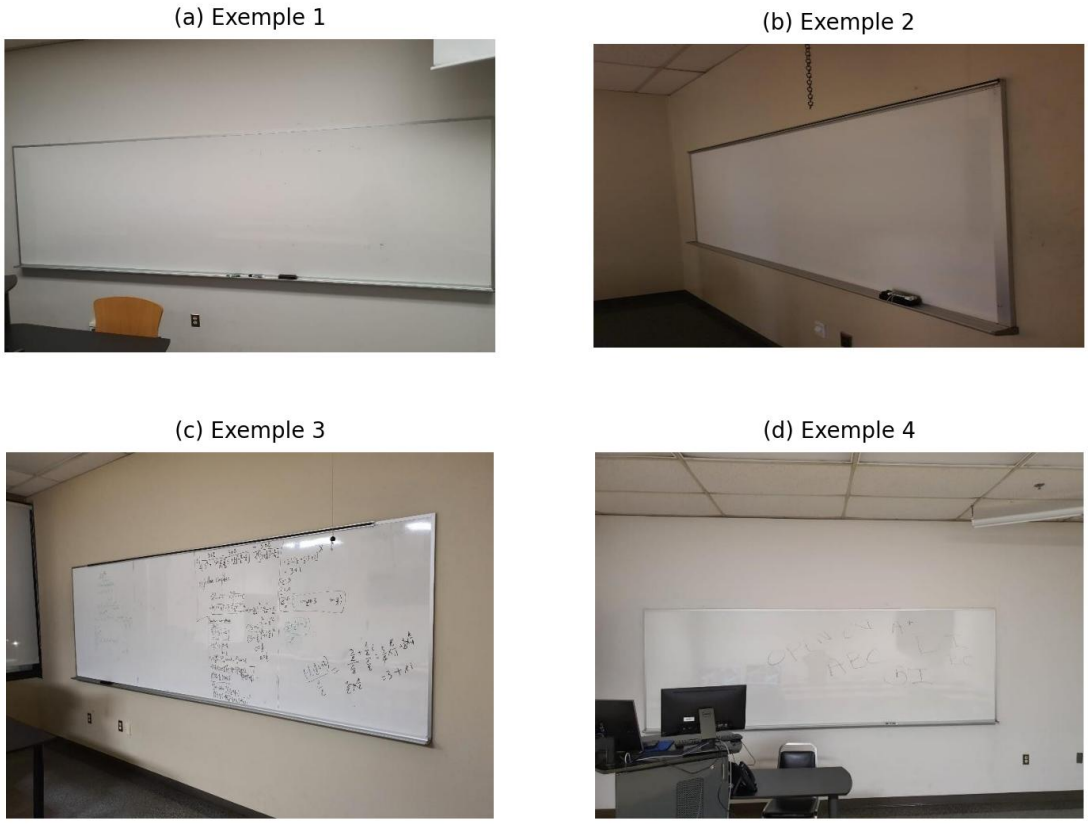

Figure 4.1 - Exemples du jeu de données : (a) vue frontale; (b) perspective oblique et faible éclairage; (c) tableau avec contenu et reflets; (d) vue éloignée avec mobilier au premier plan

4.2.1 Principe et format de l'annotation

Chaque image est associée à un polygone entourant la surface visible du tableau. Les coordonnées sont normalisées relativement à la largeur et à la hauteur afin de rester compatibles avec le format de segmentation YOLO. La classe 0 correspond à board. Une annotation valide doit comporter un nombre suffisant de points, demeurer dans les limites de l'image et suivre la frontière utile plutôt que le seul contenu écrit.

L'annotation de la surface complète est importante pour la rectification. Si le polygone entoure uniquement les lignes de texte, le réseau apprend une région variable dont la géométrie ne correspond pas nécessairement au cadre du tableau. À l'inverse, si l'annotation inclut une grande partie du mur, les coins extraits seront déplacés. La consigne retenue consiste donc à suivre le contour extérieur visible de la surface pédagogique.

Les situations d'occlusion exigent une convention constante. Lorsque le tableau est partiellement masqué par une personne ou un objet, le polygone doit représenter la surface

attendue en s'appuyant sur les bords visibles, sans entourer l'obstacle comme une partie du tableau. Cette convention favorise l'apprentissage de la forme globale, mais elle introduit une part d'interprétation humaine qui devrait être contrôlée par une seconde lecture dans une étude ultérieure.

4.2.2 Contrôle de qualité des annotations

Le premier contrôle est syntaxique: chaque fichier doit contenir un identifiant de classe et une suite paire de coordonnées normalisées. Le deuxième est géométrique: l'aire du polygone doit être positive et ses points ne doivent pas produire une figure auto-intersectée. Le troisième est visuel: le polygone est superposé à l'image afin de vérifier qu'il suit les bords du tableau.

Une inspection visuelle est particulièrement nécessaire pour les images où le bord du tableau se confond avec le mur. Dans ces cas, une différence de quelques pixels peut modifier l'IoU à des seuils élevés. La qualité de l'annotation devient donc une source d'incertitude, au même titre que le modèle. Le mémoire ne dispose pas d'une mesure d'accord inter-annotateurs; cette absence est signalée comme une limite méthodologique.

Pour une réplication, un échantillon aléatoire devrait être vérifié par une seconde personne, et les divergences devraient être résolues selon une règle documentée. On pourrait mesurer l'IoU entre annotateurs avant la résolution afin d'estimer la part d'incertitude introduite par la vérité terrain. Cette information serait utile pour interpréter la différence relativement faible entre YOLO seul et l'approche hybride.

Tableau 4.2 - État de documentation des conditions d'acquisition

Condition souhaitée	Nombre documenté	Information disponible
Éclairage normal	Non consigné	Présence observée dans le corpus
Faible luminosité	Non consigné	Présence observée, sans étiquette dédiée
Reflets	Non consigné	Présence observée, sans étiquette dédiée
Perspective oblique	Non consigné	Présence observée, sans étiquette dédiée
Occlusion	Non consigné	Présence observée, sans étiquette dédiée
Distance de capture	Non consigné	Variation présente, mesure non conservée

Le tableau 4.2 répond explicitement à la demande de ventilation par condition. Les catégories pertinentes sont identifiées, mais leurs effectifs n'ont pas été consignés pendant

la collecte. Les quatre images de la figure 4.1 sont illustratives et ne permettent pas d'inférer la distribution des 416 images. Inventer des effectifs a posteriori rendrait le protocole moins rigoureux; ces limites, y compris l'absence d'images négatives, sont donc documentées et un schéma d'étiquetage est proposé pour les travaux futurs.

4.3 Séparation des données

Tableau 4.3 - Répartition du jeu de données

Sous-ensemble	Nombre d'images	Proportion	Utilisation
Entraînement	336	80,8 %	Optimisation des poids
Validation	80	19,2 %	Calcul des métriques présentées
Test indépendant	0	0 %	Non constitué
Total	416	100 %	Jeu complet

Le tableau 4.3 fournit la répartition exacte. Les 80 images de validation n'ont pas servi à l'optimisation directe des poids, mais elles appartiennent au même jeu de collecte que les images d'entraînement. L'absence de jeu de test indépendant limite l'estimation de la généralisation externe et est explicitement prise en compte dans la discussion.

4.3.1 Rôle des sous-ensembles

L'ensemble d'entraînement fournit les exemples utilisés pour ajuster les poids du réseau. Les augmentations éventuellement appliquées par la bibliothèque créent des variantes pendant l'apprentissage, mais elles ne constituent pas de nouvelles images indépendantes. L'ensemble de validation mesure le comportement des poids sur des images qui ne participent pas directement au calcul du gradient. Il sert aussi à sélectionner le meilleur état du modèle pendant l'exécution.

Cette utilisation de la validation signifie que les métriques finales ne sont pas une estimation entièrement indépendante. Même si les 80 images ne mettent pas à jour les poids, leur performance peut influencer le choix de l'époque ou du fichier de poids. Un véritable ensemble de test devrait rester inutilisé jusqu'à la fin de toutes les décisions. Aucun sous-ensemble de ce type n'a été constitué, et les résultats sont présentés comme une validation interne.

Le découpage 336/80 correspond à 80,8 % et 19,2 %, valeurs proches d'un partage 80/20. Cette précision remplace toute formulation approximative. Les effectifs, plutôt qu'un pourcentage arrondi, permettent de vérifier le total de 416 images et d'éviter l'ambiguïté entre validation et test.

4.3.2 Prévention des fuites de données

Une fuite de données survient lorsque des informations très proches apparaissent dans les sous-ensembles d'entraînement et de validation. Dans un corpus issu de séquences vidéo, deux trames consécutives peuvent être presque identiques et conduire à une estimation trop optimiste. Le protocole idéal doit donc regrouper les images par séance, salle ou séquence avant le découpage, plutôt que de répartir aléatoirement chaque image.

Les métadonnées nécessaires à une vérification complète par séance n'ont pas été conservées dans le document expérimental. Il n'est donc pas possible d'affirmer que toute proximité entre scènes a été éliminée. Cette menace est intégrée à l'analyse de validité. Pour une nouvelle collecte, chaque capture devrait recevoir un identifiant de salle, de session, de caméra et de séquence afin de réaliser un découpage par groupe.

La déduplication visuelle pourrait compléter ce contrôle. Des empreintes perceptuelles ou des descripteurs d'image permettraient de repérer des paires quasi identiques entre les sous-ensembles. Une inspection manuelle des plus proches voisins confirmerait les cas suspects. Cette procédure n'a pas été appliquée rétrospectivement et ne doit pas être confondue avec les résultats effectivement obtenus.

4.4 Paramètres d'entraînement

Après l'annotation, les données ont été organisées selon la structure attendue par YOLOv8.

Le fichier data.yaml contient les informations relatives aux chemins des données et au nombre de classes, ici une seule classe correspondant au tableau.

Les coordonnées des polygones ont été normalisées par rapport aux dimensions de l'image, conformément aux exigences du format YOLO segmentation.

L'entraînement du modèle a été réalisé à partir du modèle pré-entraîné yolov8s-seg.pt, ce qui permet de bénéficier d'un apprentissage par transfert. La commande d'entraînement

renseigne notamment la taille des images d'entrée, le nombre d'époques et la taille du lot utilisés pendant l'apprentissage.

Les principaux paramètres retenus sont les suivants: imgsz = 640 pour le redimensionnement des images d'entrée, epochs = 100 pour le nombre de cycles d'apprentissage, batch = 8 pour la taille du lot, et la variante légère yolov8s-seg afin de conserver un compromis pertinent entre précision et rapidité (Ultralytics, 2023).

L'apprentissage par transfert permet d'adapter un modèle déjà entraîné sur de grandes bases de données générales à la tâche spécifique de détection de tableaux pédagogiques.

Cette section décrit uniquement la configuration d'entraînement et ne tire aucune conclusion des scores. Les métriques objectives, les comparaisons et l'analyse de la robustesse sont regroupées au chapitre 5, après la définition du protocole d'évaluation.

Tableau 4.4 - Hyperparamètres utilisés pour l'entraînement

Paramètre	Valeur retenue	Justification
Modèle	yolov8s-seg.pt	Compromis entre précision et coût
Taille d'image	640 × 640 pixels	Résolution standard Ultralytics
Époques	100	Convergence du transfert d'apprentissage
Taille du lot	8	Compatibilité avec les ressources
Classe	1 (board)	Une seule région d'intérêt
Optimiseur	Sélection automatique	Aucune valeur explicite conservée
Taux initial	0,01	Valeur documentée dans la configuration

Le tableau 4.4 distingue les paramètres explicitement fixés de ceux délégués à la configuration automatique d'Ultralytics. Cette présentation évite d'attribuer au protocole un optimiseur qui n'a pas été consigné.

4.4.1 Transfert d'apprentissage

Le modèle est initialisé à partir de poids préentraînés plutôt que de valeurs aléatoires. Le transfert d'apprentissage réutilise des représentations visuelles générales, puis adapte les couches au domaine des tableaux. Cette stratégie est appropriée à un corpus de 416 images, dont la taille serait limitée pour entraîner un réseau profond depuis zéro. Elle réduit le temps nécessaire à la convergence et stabilise les premières époques.

Le transfert ne garantit toutefois pas une généralisation automatique. Les données d'origine des poids représentent de nombreux objets et environnements, mais la classe board possède des caractéristiques propres: grande surface plane, cadre rectangulaire, reflets et contenu écrit variable. L'apprentissage doit ajuster la tête de segmentation et les représentations de haut niveau à ces caractéristiques sans dégrader les informations générales utiles.

La taille de lot 8 détermine le nombre d'images utilisées avant une mise à jour des poids. Un lot plus grand pourrait lisser le gradient, mais demander davantage de mémoire. Le choix retenu correspond aux ressources disponibles. Le nombre de 100 époques fixe une limite d'entraînement; il ne signifie pas que la dernière époque est nécessairement le meilleur modèle, d'où l'importance de conserver les poids associés à la meilleure validation.

4.4.2 Augmentation et variabilité

Les augmentations visent à présenter au réseau des variantes compatibles avec la scène réelle. Des modifications modérées de luminosité, de couleur, d'échelle ou de perspective peuvent améliorer la robustesse. Elles doivent cependant préserver le sens de l'annotation. Une transformation trop forte peut rendre le tableau méconnaissable ou produire une géométrie qui n'existe pas dans les conditions d'utilisation.

La configuration détaillée de chaque augmentation n'a pas été conservée dans les éléments disponibles. Le mémoire ne lui attribue donc pas un effet quantitatif. Une réplication devrait exporter la configuration complète d'Ultralytics, car les valeurs par défaut peuvent évoluer d'une version à l'autre. Un essai d'ablation, avec et sans augmentation, permettrait ensuite de mesurer leur contribution.

L'absence d'effectifs par condition empêche également de vérifier si les augmentations compensent un déséquilibre réel ou simulé. Une future base devrait associer à chaque image des étiquettes de luminosité, perspective, réflexion, occlusion et distance. Il deviendrait alors possible de comparer la performance brute, la performance après augmentation et la performance par sous-groupe.

4.5 Procédure d'inférence et de rectification

Pour chaque image, l'inférence produit un masque de segmentation associé à la classe board. Le masque est binarisé puis converti en contour externe. Une approximation polygonale recherche quatre sommets; lorsqu'elle n'en fournit pas directement, le rectangle orienté minimal sert de solution de repli. Les sommets sont ensuite ordonnés selon leur position relative avant le calcul de l'homographie.

La figure 4.2 montre des sorties de segmentation sur quatre images distinctes.

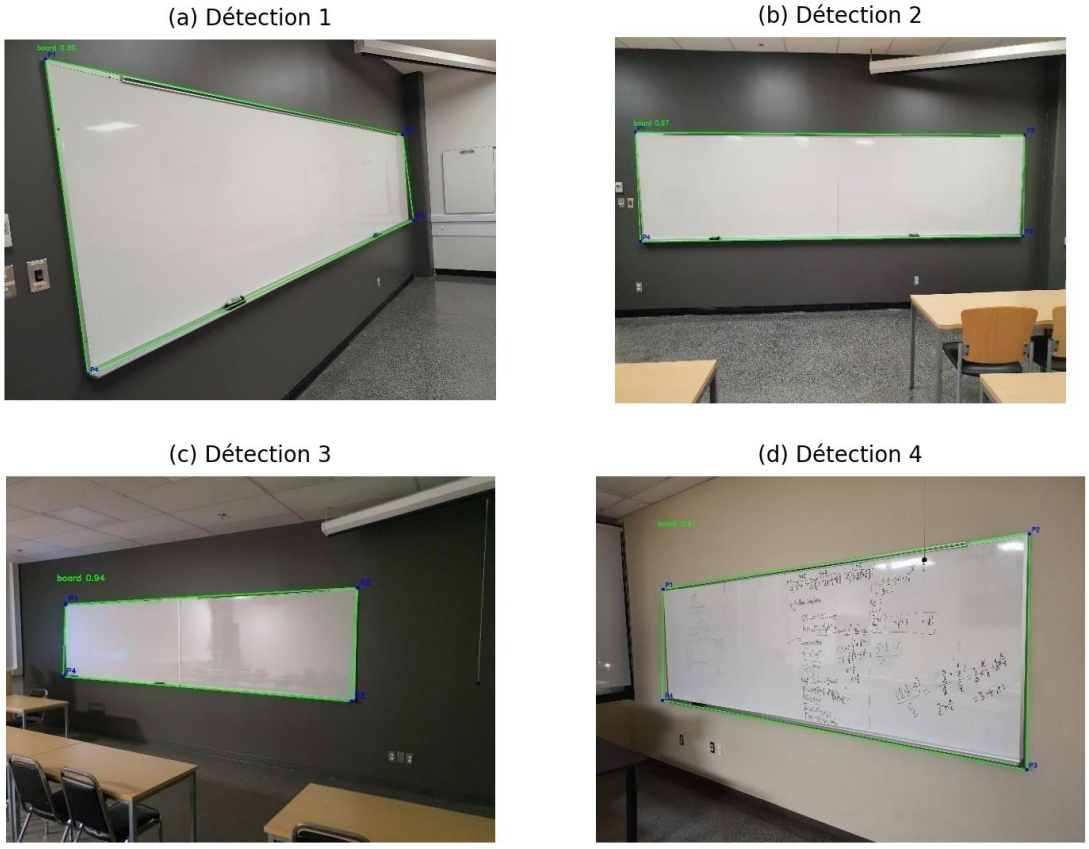

Figure 4.2 - Exemples de sorties d'inférence : (a) forte perspective; (b) vue frontale; (c) vue éloignée; (d) tableau comportant du contenu manuscrit

Après l'inférence, le masque est converti en contour. Un quadrilatère est estimé, les coins sont ordonnés, puis une homographie transforme la région en vue frontale. Cette séquence associe la robustesse de la segmentation à la précision de la géométrie projective.

4.5.1 Chaîne d'inférence reproductible

La procédure d'inférence commence par le chargement explicite du fichier de poids et par la définition de la taille d'image. Chaque fichier de validation est lu une seule fois, puis transmis au modèle sans mise à jour des poids. Les prédictions sont ramenées au repère d'origine et associées à l'identifiant de l'image. Cette association permet de comparer la sortie à l'annotation correspondante.

Pour la méthode YOLO seule, le masque produit par le réseau est comparé directement à la vérité terrain. Pour la méthode hybride, le même masque alimente le calcul des coins et la rectification géométrique. La comparaison utilise donc la même image de départ et la même partition. Ce contrôle est nécessaire pour que la différence observée soit attribuable au traitement ajouté plutôt qu'à un changement de données.

La méthode classique reçoit elle aussi les 80 images de validation. Elle applique les paramètres du tableau 2.5, recherche les contours et estime le quadrilatère sans utiliser le masque neuronal. Les trois méthodes sont ainsi évaluées sur le même corpus. Le mémoire ne présente pas d'optimisation indépendante des seuils sur un jeu réservé; cette limitation doit être prise en compte dans la comparaison.

4.5.2 Calcul des métriques et agrégation

Pour chaque image, la prédiction est associée à la région annotée. Une détection correcte contribue aux vrais positifs lorsque la classe et le recouvrement satisfont le seuil considéré. Une prédiction sans correspondance contribue aux faux positifs; une annotation non retrouvée contribue aux faux négatifs. Les valeurs agrégées produisent la précision et le rappel présentés au chapitre 5.

La mAP@0,50 calcule l'Average Precision pour un seuil d'IoU de 0,50. La mAP@0,50-0,95 moyenne ensuite la performance sur plusieurs seuils croissants et pénalise davantage les limites imprécises. Comme le corpus ne contient qu'une classe, la moyenne sur les classes correspond à la performance de board. L'IoU moyen compare directement la surface prédite et l'annotation.

Le temps d'inférence est rapporté par image sur CPU. Il doit être séparé du temps total d'un scénario complet, qui comprendrait la lecture, le post-traitement, la rectification et l'écriture du fichier. La valeur de 476,7 ms décrit donc la mesure conservée, non une

latence de bout en bout. Une future expérience devra chronométrer chaque composant et répéter la mesure après une phase d'échauffement.

Tableau 4.5 - Synthèse du protocole d'évaluation

Élément du protocole	Valeur ou règle	Portée
Corpus évalué	80 images de validation	Validation interne
Classe	board	Une seule classe
Taille d'entrée	640 × 640	Inférence YOLO
Métriques	Précision, rappel, mAP, IoU	Détection et segmentation
Temps	476,7 ms/image sur CPU	Inférence documentée
Test externe	Absent	Généralisation non démontrée

Le tableau 4.5 rassemble les informations nécessaires à l'interprétation des résultats. Il évite de présenter les métriques sans rappeler le sous-ensemble, la classe et la plateforme de mesure.

CHAPITRE 5
RÉSULTATS ET DISCUSSION

5.1 Protocole d'évaluation

Toutes les métriques quantitatives présentées dans ce chapitre sont calculées sur les 80 images de validation. La précision mesure la proportion de prédictions correctes parmi les détections; le rappel mesure la proportion de tableaux retrouvés; l'Intersection over Union (IoU) mesure le recouvrement entre la région prédite et la vérité terrain; l'Average Precision résume la courbe précision-rappel; la mAP moyenne cette valeur sur les classes et, pour mAP@0,50-0,95, sur plusieurs seuils d'IoU (Padilla et al., 2021; Ultralytics, 2023).

$$IoU = |A\_prédite \cap A\_vérité| / |A\_prédite \cup A\_vérité| \quad (5.1)$$

Les ensembles A_prédite et A_vérité représentent respectivement la région prédite et l'annotation de référence.

$$Précision = VP / (VP + FP) ; Rappel = VP / (VP + FN) \quad (5.2)$$

VP, FP et FN désignent les vrais positifs, les faux positifs et les faux négatifs.

5.1.1 Lecture conjointe des métriques

Aucune métrique ne résume seule la qualité du système. Un rappel élevé signifie que les tableaux annotés sont retrouvés, mais il peut être obtenu au prix de nombreuses détections erronées. La précision complète donc le rappel en mesurant la proportion de prédictions pertinentes. Dans un système de captation, un faux négatif entraîne la perte d'une capture, tandis qu'un faux positif peut provoquer l'archivage d'une région qui n'est pas le tableau.

L'IoU évalue la qualité spatiale. Une détection peut être considérée comme correcte à un seuil de 0,50 tout en étant insuffisante pour une rectification précise si elle décale fortement un bord. C'est pourquoi la mAP@0,50-0,95 est plus informative que la mAP@0,50 pour ce projet: elle répète l'évaluation avec des seuils de recouvrement plus exigeants et reflète mieux la stabilité des limites.

L'Average Precision correspond à l'aire sous une courbe précision-rappel interpolée. Dans le cas d'une seule classe, la mAP est égale à l'AP de board pour le seuil ou

l'ensemble de seuils considéré. Cette simplification ne doit pas masquer la taille modeste du corpus: une valeur élevée sur 80 images de validation n'a pas la même portée qu'une valeur obtenue sur plusieurs établissements.

$$AP = \int_0^1 p(r)\, dr \qquad (5.3)$$

La fonction p(r) représente la précision interpolée en fonction du rappel r; l'intégrale résume la courbe sur l'intervalle de rappel.

5.1.2 Unité d'analyse et seuils

L'unité d'analyse est l'image de validation. Chacune contient au moins un tableau annoté; le corpus ne comprend donc aucune scène négative entièrement dépourvue de tableau. Cette composition ciblée permet d'évaluer la localisation et la segmentation lorsque la cible est présente, mais elle ne mesure pas le comportement du modèle en son absence et peut favoriser une prédiction positive par défaut. Il serait incorrect de calculer une spécificité ou un nombre de vrais négatifs sans définir un ensemble négatif. Une évaluation complémentaire devra inclure des scènes de salle sans tableau et rapporter le nombre de fausses activations par image.

Le seuil de confiance influence le compromis entre précision et rappel. Un seuil faible conserve davantage de candidats, ce qui peut réduire les faux négatifs mais augmenter les faux positifs. Un seuil élevé produit l'effet inverse. Les courbes d'évaluation intègrent ce classement des prédictions; pour une utilisation réelle, un seuil opérationnel devrait être choisi sur un ensemble de développement distinct du test.

Les seuils d'IoU ne décrivent pas la même exigence. À 0,50, une région partiellement décalée peut encore être acceptée. À 0,75 ou 0,90, la frontière doit être beaucoup plus proche de l'annotation. Le maintien d'une mAP élevée sur l'intervalle 0,50-0,95 est donc cohérent avec l'objectif de rectification, mais il ne remplace pas l'inspection des images où une petite erreur de coin peut affecter le contenu.

5.2 Résultats quantitatifs

Tableau 5.1 - Performances du modèle YOLO sur l'ensemble de validation

Métrique	Valeur	Ensemble
Précision	0,987	Validation (80 images)

Métrique	Valeur	Ensemble
Rappel	1,000	Validation (80 images)
mAP@0,50	0,992	Validation (80 images)
mAP@0,50-0,95	0,967	Validation (80 images)
Temps d'inférence	476,7 ms/image	CPU

Le tableau 5.1 indique que les 80 tableaux présents ont été détectés, ce qui conduit à un rappel de 1,000. La précision de 0,987 est la métrique agrégée fournie par la validation Ultralytics sur la courbe précision-rappel. Elle ne doit pas être reconstruite directement à partir d'un seul seuil. Au seuil opérationnel résumé dans le tableau 5.2, 80 vrais positifs et un faux positif donnent une précision de 80/81 = 0,988 après arrondissement à trois décimales. La mAP@0,50-0,95, plus exigeante que la mAP@0,50, demeure élevée, ce qui traduit une localisation stable sur plusieurs seuils de recouvrement.

Tableau 5.2 - Synthèse des détections observées

Réalité / prédiction	Tableau détecté	Non détecté
Tableau annoté (80 images)	80 vrais positifs	0 faux négatif
Région sans correspondance	1 faux positif	Vrais négatifs non définis

Le tableau 5.2 distingue les 80 tableaux annotés des prédictions supplémentaires sans correspondance. Le faux positif observé apparaît dans une image positive; il ne provient pas d'une scène entièrement sans tableau. Les vrais négatifs restent non définis, puisque le corpus ne comprend pas d'ensemble négatif et que le protocole ne décompose pas chaque image en régions candidates exhaustives.

5.2.1 Interprétation de la précision et du rappel

Le rappel de 1,000 indique qu'aucune des 80 annotations n'est restée sans détection associée. Ce résultat répond directement au risque de perdre une capture de tableau. Il ne signifie cependant pas que chaque limite est parfaite; une prédiction peut compter comme vraie positive tout en présentant un décalage spatial mesuré par l'IoU. Le rappel doit donc être lu avec les métriques de segmentation.

La précision agrégée de 0,987 provient de la courbe de validation. Au seuil opérationnel retenu pour la matrice du tableau 5.2, le faux positif observé conduit à une

précision de 0,988. Dans un usage interactif, cette erreur peut être détectée par un aperçu avant l'archivage; dans une chaîne entièrement automatique, elle pourrait générer un fichier inutile. Le seuil devrait donc être ajusté selon le coût relatif d'une capture manquée et d'une capture supplémentaire.

Le nombre limité d'erreurs empêche une analyse statistique fine de leurs causes. Un seul faux positif ne permet pas d'estimer de manière stable une probabilité par type de salle ou d'objet. Il est néanmoins important de le conserver dans le rapport plutôt que d'arrondir la précision à 1,000, car il révèle que la détection n'est pas parfaite.

5.2.2 Interprétation de la mAP et de l'IoU

La différence entre la mAP@0,50 de 0,992 et la mAP@0,50-0,95 de 0,967 est de 0,025. Cette baisse est attendue lorsque les seuils deviennent plus exigeants. Son amplitude limitée suggère que la majorité des prédictions conserve un bon recouvrement, mais le résultat reste agrégé et ne permet pas de localiser les images responsables de la diminution.

L'IoU moyen de YOLO seul, égal à 0,910 dans la comparaison, indique que les masques recouvrent largement les annotations. L'approche hybride porte cette moyenne à 0,940. L'amélioration absolue de 0,030 peut paraître plus faible que le gain obtenu par rapport à la méthode classique, mais elle concerne une base déjà élevée et correspond à un ajustement des frontières utile à la rectification.

Une moyenne ne décrit pas les extrêmes. L'écart-type de 0,020 pour l'approche hybride indique une dispersion réduite dans le corpus, sans fournir à lui seul la valeur minimale ou maximale. Une future analyse devrait publier la distribution complète, les quartiles et les images associées aux plus faibles IoU afin de relier statistique et inspection visuelle.

5.2.3 Performance temporelle

$$Débit \approx 1000 / 476,7 = 2,10 \text{ images/s} \qquad (5.4)$$

Le numérateur convertit une seconde en millisecondes; le calcul donne un débit théorique fondé sur le seul temps d'inférence CPU.

Le débit de 2,10 images par seconde ne correspond pas à une vidéo de 25 ou 30 images par seconde. Il convient davantage à une capture déclenchée ou à un échantillonnage

espacé. La latence perçue pourrait être supérieure si l'on ajoute la lecture de la caméra, la rectification et l'écriture. À l'inverse, une exécution optimisée ou un accélérateur pourrait réduire le temps, mais cette hypothèse n'est pas démontrée par les mesures disponibles.

Une évaluation temporelle complète devrait séparer prétraitement, inférence, post-traitement géométrique et sauvegarde. Chaque étape devrait être chronométrée sur plusieurs répétitions, après échauffement, en rapportant moyenne, médiane et écart-type. La consommation de mémoire et l'utilisation du processeur compléteraient la mesure pour un déploiement sur du matériel de salle de classe.

5.3 Comparaison des approches

Tableau 5.3 - Comparaison quantitative des méthodes sur l'ensemble de validation

Méthode	mAP@0,50	mAP@0,50-0,95	IoU moyen	Écart-type IoU
Classique	0,820	0,610	0,680	0,110
YOLO seul	0,992	0,967	0,910	0,040
Hybride	0,995	0,972	0,940	0,020

Le tableau 5.3 et la figure 5.1 utilisent exactement les mêmes valeurs. La méthode classique est pénalisée par les variations d'éclairage et les contours incomplets. YOLO améliore fortement la détection, tandis que l'approche hybride accroît surtout l'IoU grâce à la rectification des limites du tableau.

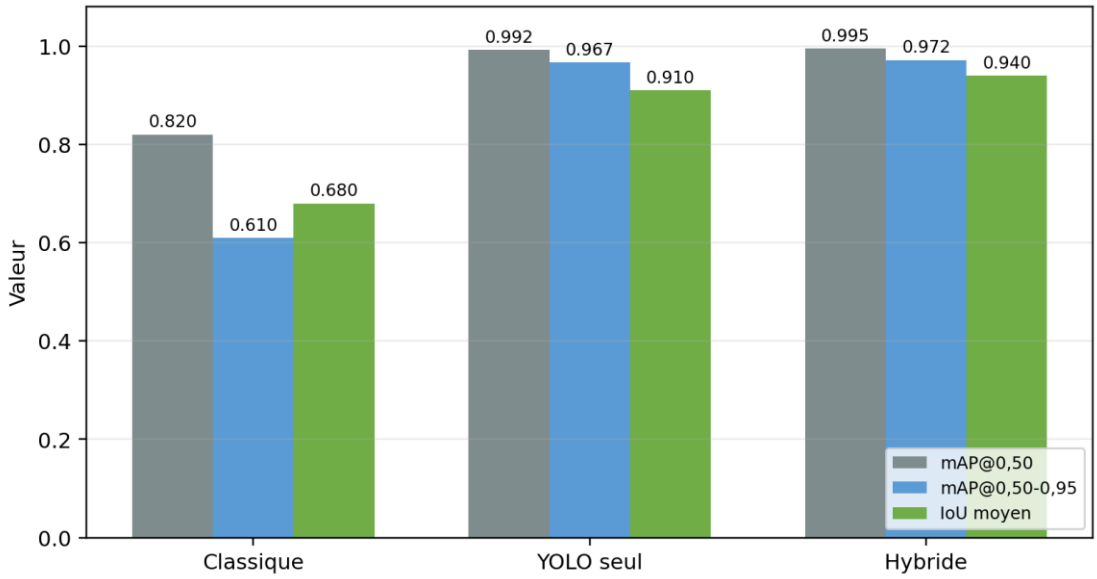

Figure 5.1 - Comparaison cohérente des performances des trois approches

5.3.1 Apport de la segmentation profonde

Le passage de la méthode classique à YOLO seul produit les gains les plus importants. La mAP@0,50 augmente de 0,172 et la mAP@0,50-0,95 de 0,357. L'IoU moyen progresse de 0,230, tandis que l'écart-type diminue de 0,070. Ces différences sont cohérentes avec la capacité du modèle à utiliser des indices visuels plus riches que les seuls contours et lignes.

La méthode classique dépend de contrastes locaux et de paramètres fixes. Lorsque le cadre est interrompu par un reflet ou que le mur présente d'autres structures rectangulaires, la sélection du quadrilatère devient ambiguë. Le réseau apprend au contraire une représentation de la surface, du cadre et du contexte. Il demeure sensible aux données d'entraînement, mais il réduit la dépendance à une frontière parfaitement visible.

Ce résultat ne rend pas les méthodes classiques inutiles. Canny, la morphologie et l'analyse des contours restent nécessaires pour interpréter et transformer le masque. Ils fournissent aussi une référence de comparaison compréhensible. L'expérience montre plutôt que leur utilisation seule est insuffisante dans le corpus étudié.

5.3.2 Apport de la composante géométrique

La différence entre YOLO seul et l'approche hybride est plus modeste pour la mAP, car la présence du tableau est déjà correctement détectée. La mAP@0,50 augmente de 0,003 et la mAP@0,50-0,95 de 0,005. Le gain d'IoU moyen atteint 0,030, ce qui

correspond mieux au rôle de la composante géométrique: affiner les limites et préparer une vue rectifiée.

La diminution de l'écart-type de 0,040 à 0,020 suggère une plus grande régularité des recouvrements. L'homographie et l'ordonnancement des coins imposent une structure rectangulaire cohérente aux prédictions. Cette contrainte peut corriger de petites irrégularités du masque, à condition que la région initiale soit suffisamment proche du tableau.

La géométrie ne peut toutefois pas récupérer une détection absente ou entièrement erronée. Si le réseau sélectionne un autre objet rectangulaire, l'homographie produira une vue frontale de cet objet. Le contrôle de classe, de confiance et de cohérence spatiale reste donc indispensable avant la rectification.

Tableau 5.4 - Gains absolus entre les méthodes comparées

Comparaison	Δ mAP@0,50	Δ mAP@0,50-0,95	Δ IoU moyen	Δ écart-type IoU
YOLO - classique	+0,172	+0,357	+0,230	-0,070
Hybride - YOLO	+0,003	+0,005	+0,030	-0,020
Hybride - classique	+0,175	+0,362	+0,260	-0,090

Le tableau 5.4 présente des différences absolues calculées à partir du tableau 5.3. Aucune nouvelle expérience n'est introduite. Cette présentation rend visible la contribution dominante de la segmentation profonde et la contribution complémentaire de la géométrie.

5.3.3 Prudence statistique

Les valeurs comparatives sont observées sur un seul partage du corpus. Sans répétition avec plusieurs graines ou plusieurs partitions, il n'est pas possible de construire un intervalle de confiance sur le gain entre les méthodes. Une différence de 0,003 en mAP@0,50 peut être opérationnellement faible et doit être interprétée avec davantage de prudence que le gain de 0,230 en IoU par rapport à la méthode classique.

Un test statistique par image nécessiterait de conserver les IoU individuels pour chaque méthode. On pourrait alors analyser les différences appariées, leur médiane et leur distribution, plutôt que de comparer uniquement des moyennes. Ces données détaillées ne

figurent pas dans les journaux disponibles; le mémoire limite donc ses conclusions aux tendances descriptives.

La cohérence entre le tableau 5.3 et la figure 5.1 a été vérifiée. Les deux supports représentent exactement les mêmes valeurs et remplissent des fonctions différentes: le tableau permet la lecture précise, tandis que la figure facilite la comparaison visuelle. Cette cohérence corrige l'ambiguïté signalée dans l'évaluation initiale.

5.4 Analyse qualitative et robustesse

Les observations qualitatives confirment une bonne stabilité sous éclairage normal, à distance modérée et en perspective oblique. Les difficultés apparaissent lorsque le tableau est fortement occulté, que les reflets effacent une partie de sa frontière ou qu'un objet rectangulaire concurrence visuellement la région d'intérêt. La figure 5.2 présente trois catégories d'échec de l'approche hybride; les cadres verts correspondent à ses prédictions, conformément au code couleur de la figure.

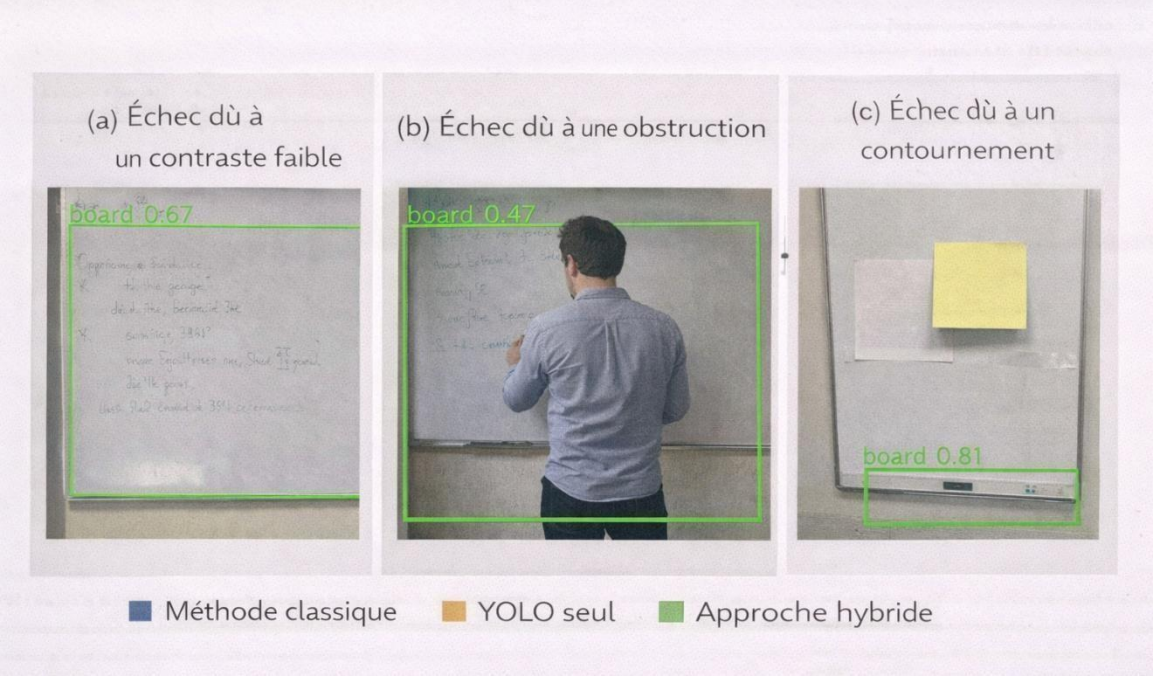

Figure 5.2 - Cas d'échec de l'approche hybride (cadres verts) : (a) faible contraste; (b) occlusion; (c) objet rectangulaire concurrent

Le temps moyen de 476,7 ms par image correspond à environ 2,10 images par seconde sur CPU. Cette cadence convient à une capture périodique ou déclenchée à la demande, mais elle ne justifie pas encore l'expression « temps réel » pour une vidéo fluide.

La dispersion de l'IoU complète cette lecture. L'écart-type passe de 0,110 pour la méthode classique à 0,040 pour YOLO seul, puis à 0,020 pour l'approche hybride. La rectification n'améliore donc pas seulement la valeur moyenne: elle réduit aussi la variabilité spatiale observée dans ce corpus.

5.4.1 Faible contraste et reflets

Le faible contraste réduit la différence entre la surface du tableau et son environnement. Pour la méthode classique, les gradients deviennent faibles et les segments de bord peuvent être interrompus. Le réseau profond peut utiliser la texture, le cadre et le contexte, mais une scène très sombre ou surexposée peut aussi s'éloigner des exemples d'entraînement. La figure 5.2a illustre cette difficulté sans fournir un taux d'échec par luminosité.

Les reflets produisent un problème différent: ils créent une région localement claire qui peut effacer le contenu et une partie de la frontière. Un masque global peut rester correct si les autres bords sont visibles. La rectification ne restaure toutefois pas les informations déjà saturées dans l'image. Une amélioration de contraste ou une fusion de plusieurs captures pourrait être étudiée après la localisation.

L'absence d'étiquettes de condition empêche de comparer numériquement les images à faible luminosité et les images normales. La conclusion doit donc rester qualitative. Une collecte future devrait définir des seuils reproductibles, par exemple à partir de la luminance moyenne et de la proportion de pixels saturés, puis publier les métriques pour chaque groupe.

5.4.2 Occlusion et objets concurrents

Une occlusion masque une partie de la surface ou du cadre. La figure 5.2b montre une personne devant le tableau. Le modèle doit alors compléter implicitement la région à partir des parties visibles. Si l'occlusion atteint plusieurs coins, le masque peut devenir trop irrégulier pour la géométrie. Le rectangle minimal offre une solution de repli, mais il peut inclure une zone extérieure au tableau.

Les objets rectangulaires concurrents, comme une affiche ou un écran, partagent une géométrie avec la cible. La méthode classique risque de les privilégier si leurs contours sont plus nets. Le réseau utilise le contexte, mais un objet peu représenté dans l'entraînement

peut encore recevoir une confiance élevée. La figure 5.2c illustre la nécessité de vérifier la classe, l'aire et la position.

Un scénario à plusieurs tableaux n'est pas évalué séparément. La règle de conservation de la région principale peut convenir lorsque l'un d'eux domine l'image, mais elle peut supprimer un second tableau utile. Un système interactif devrait alors afficher les candidats ou autoriser plusieurs captures. Cette fonctionnalité appartient à la conception future et ne doit pas être attribuée au prototype actuel.

5.4.3 Perspective et distance

La perspective oblique transforme le rectangle du tableau en quadrilatère. C'est précisément la situation visée par l'homographie. Tant que les quatre coins sont localisés, la géométrie peut rétablir une vue frontale. Lorsque l'angle devient extrême, la surface visible se réduit, les détails sont comprimés et l'ordre des coins devient plus sensible au bruit.

La distance influence le nombre de pixels représentant le tableau et le contenu. Un tableau éloigné peut être détecté, mais l'écriture risque de devenir illisible après agrandissement. Le protocole actuel mesure la localisation, non la lisibilité du texte. Une évaluation pédagogique devrait ajouter une mesure de résolution utile ou une tâche de lecture après rectification.

L'absence de mesure de distance et d'angle dans les métadonnées empêche d'identifier un seuil opérationnel. Une future acquisition pourrait enregistrer la pose de la caméra ou estimer l'angle à partir du quadrilatère. Les métriques pourraient ensuite être présentées par intervalles d'angle et de proportion de surface occupée.

Tableau 5.5 - Taxonomie des facteurs de difficulté

Facteur	Effet attendu	Module principalement touché	Mesure future
Faible contraste	Contours incomplets	Classique et segmentation	Luminance et IoU
Reflets	Frontière et contenu saturés	Segmentation et lisibilité	Pixels saturés
Occlusion	Masque irrégulier	Segmentation et géométrie	Taux d'occlusion
Objet concurrent	Faux positif	Sélection de la région	Précision par scène
Perspective forte	Coins instables	Géométrie	Angle et erreur de coin
Grande distance	Perte de détails	Acquisition	Pixels du tableau

Le tableau 5.5 transforme l'analyse qualitative en plan de mesure. Il ne présente pas de nouveaux effectifs; il précise quelles métadonnées devront être collectées pour quantifier chaque facteur sans inventer des informations absentes du corpus actuel.

5.5 Limites de validité

La première limite concerne l'absence de jeu de test indépendant. Les métriques décrivent une validation interne et peuvent surestimer la performance attendue dans un nouvel établissement. La deuxième limite est la composition exclusivement positive du corpus: chaque image contient un tableau et aucune scène entièrement sans tableau n'a été incluse. Ce choix empêche d'estimer directement la propension du modèle à produire une détection positive par défaut. La troisième limite est la taille modérée du jeu de données et l'absence de comptage systématique par condition d'éclairage, angle ou occlusion. Enfin, les comparaisons reposent sur un seul corpus et les résultats ne valident ni l'OCR ni une chaîne complète de diffusion pédagogique.

Malgré ces limites, les observations sur le sous-ensemble de validation indiquent que la segmentation profonde améliore la localisation par rapport à la méthode classique et que la correction projective augmente l'IoU moyen. Cette conclusion reste descriptive et limitée aux images positives du corpus étudié.

5.5.1 Validité interne

La validité interne concerne l'attribution des différences à la méthode plutôt qu'à un facteur extérieur. L'utilisation du même sous-ensemble de validation pour les trois approches renforce la comparaison. La cohérence des valeurs entre tableau et figure réduit également le risque d'erreur de transcription. En revanche, l'absence de journal détaillé pour certains paramètres et l'absence de répétitions limitent la capacité à isoler l'effet de chaque choix.

La qualité de la vérité terrain est une seconde menace. Les polygones ont été annotés pour représenter la surface du tableau, mais aucun accord inter-annotateurs n'est disponible. Une annotation légèrement différente peut modifier l'IoU, surtout à des seuils élevés. Le gain de l'approche hybride doit donc être lu en tenant compte de cette incertitude non quantifiée.

La sélection des poids au moyen de la validation peut introduire un optimisme. Le sous-ensemble n'ajuste pas directement les gradients, mais il intervient dans le choix du modèle retenu. La correction recommandée est de séparer développement et test, ou d'utiliser une validation croisée imbriquée lorsque le corpus est trop petit pour réserver un grand test indépendant.

5.5.2 Validité externe

La validité externe concerne la généralisation à d'autres salles, caméras, tableaux et pratiques d'enseignement. Les images proviennent d'un corpus local dont la distribution par condition n'est pas documentée. Les résultats ne permettent donc pas d'affirmer la même performance dans un établissement utilisant des tableaux d'une autre couleur, des caméras différentes ou un éclairage très variable.

Un test externe devrait être collecté après la finalisation du modèle, dans des salles et avec des appareils absents de l'entraînement. Les images devraient être regroupées par établissement pour éviter qu'une même scène apparaisse dans plusieurs partitions. La performance globale serait alors complétée par des résultats par condition et par site.

La généralisation temporelle mérite aussi une vérification. Une salle peut changer de configuration, les tableaux peuvent être remplacés et l'éclairage varier selon la saison. Un suivi longitudinal permettrait de déterminer si le modèle conserve sa performance ou nécessite une mise à jour. Cette question dépasse la validation interne présentée, mais elle est essentielle pour un système déployé.

5.5.3 Validité de construit et conclusion

La validité de construit vérifie que les métriques représentent bien l'objectif. La mAP et l'IoU mesurent la détection et le recouvrement, mais pas directement la lisibilité des notes, la satisfaction de l'enseignant ou l'utilité pour les étudiants. Une image correctement segmentée peut contenir une écriture floue, tandis qu'une image légèrement décalée peut rester utile pédagogiquement.

Une évaluation complète devrait donc combiner mesures de vision et critères d'usage. On pourrait mesurer la proportion de contenu visible après rectification, demander à des lecteurs de transcrire un échantillon ou comparer le temps nécessaire pour retrouver une

information. Ces mesures doivent être conçues comme une nouvelle étude; elles ne sont pas déduites des résultats actuels.

Tableau 5.6 - Réponse aux questions d'évaluation

Question	Élément de réponse	Niveau de preuve
Le tableau est-il détecté ?	Rappel 1,000 sur 80 images	Validation interne
Les faux positifs sont-ils maîtrisés ?	Précision agrégée 0,987; précision au seuil opérationnel 0,988 (un faux positif)	Validation interne
Les limites sont-elles précises ?	mAP@0,50-0,95 0,967; IoU hybride 0,940	Validation interne
L'hybride améliore-t-il la géométrie ?	Gain d'IoU de 0,030 sur YOLO	Comparaison descriptive
Le système fonctionne-t-il en vidéo fluide ?	Non démontré; 476,7 ms/image sur CPU	Mesure locale
L'OCR est-il validé ?	Non implémenté	Hors périmètre

Le tableau 5.6 relie les conclusions aux preuves disponibles et évite les généralisations excessives. La contribution démontrée concerne la localisation et la rectification dans le corpus de validation, non l'ensemble d'une plateforme pédagogique.

5.5.4 Implications pour le scénario pédagogique

Les résultats soutiennent un scénario de capture à la demande. Le rappel élevé réduit le risque de manquer un tableau, et la rectification améliore la présentation de la surface. Le temps CPU reste compatible avec une action ponctuelle. Dans ce cadre, une courte latence peut être acceptable si l'enseignant déclenche la capture avant l'effacement.

L'intégration à un cours doit néanmoins conserver un mécanisme de vérification. Un aperçu de l'image rectifiée permettrait de confirmer que le contenu est complet. Cette interaction serait particulièrement utile dans les cas de reflets, d'occlusion ou d'objet concurrent. Le prototype de vision fournit la base technique, mais l'interface et son évaluation auprès des utilisateurs restent à développer.

La conservation des images soulève enfin des questions de classement et d'accès. Des métadonnées comme le cours, la date, la séance et l'ordre de capture faciliteraient la consultation. Des règles de confidentialité seraient nécessaires si des informations

personnelles apparaissent au tableau. Ces fonctions appartiennent à la future couche applicative et ne modifient pas l'interprétation des métriques.

Le déclenchement de la capture devrait également être testé selon plusieurs modalités: action manuelle, intervalle régulier ou détection d'un changement important du contenu. La mesure porterait sur le nombre d'images utiles, les doublons produits et les captures manquées. Cette étude permettrait de choisir un mécanisme compatible avec le rythme du cours sans multiplier inutilement les fichiers.

Enfin, l'image rectifiée devrait être conservée avec l'image source ou avec les paramètres permettant de la recalculer. Cette traçabilité autorise une correction si les coins ont été mal estimés et évite de présenter une transformation irréversible comme une vérité. Elle s'inscrit dans le principe général de vérification humaine retenu pour le scénario pédagogique.

CHAPITRE 6
CONCLUSION ET PERSPECTIVES

6.1 Synthèse

Le travail a conduit à un système de captation de tableaux fondé sur la combinaison de YOLOv8s-seg et d'opérations OpenCV. Le modèle localise la surface du tableau; le module géométrique en extrait les coins et applique une homographie. L'évaluation interne sur 80 images montre une détection élevée et une amélioration de l'IoU pour l'approche hybride.

6.1.1 Réponse à la problématique

La problématique posée était de conserver automatiquement le contenu d'un tableau pédagogique à partir d'une caméra standard malgré les variations de scène. Les résultats montrent qu'une combinaison de segmentation profonde et de géométrie projective permet de localiser la surface, d'en estimer les limites et de produire une vue frontale. La réponse est positive dans le cadre de la validation interne, sans être généralisée à toutes les salles.

L'approche classique fournit une chaîne interprétable, mais elle dépend fortement de la continuité des contours. YOLOv8s-seg apporte la robustesse visuelle nécessaire pour localiser la région, tandis que l'homographie impose une structure géométrique utile à la captation. L'association répond ainsi à deux dimensions complémentaires du problème: reconnaître le tableau et normaliser sa perspective.

La solution obtenue ne doit pas être décrite comme une plateforme pédagogique complète. Elle ne reconnaît pas le texte, ne gère pas les utilisateurs et ne diffuse pas les fichiers. Elle constitue un noyau de vision qui fournit une image rectifiée à ces services futurs. Cette délimitation est essentielle pour que les contributions et les résultats correspondent aux fonctions effectivement évaluées.

6.1.2 Principaux résultats

Sur les 80 images de validation, le modèle obtient une précision de 0,987 et un rappel de 1,000. La mAP@0,50 atteint 0,992 et la mAP@0,50-0,95 atteint 0,967. Ces valeurs décrivent une détection et une segmentation élevées pour la classe board dans le corpus

étudié. Elles sont accompagnées de l'information essentielle qu'aucun jeu de test indépendant n'a été constitué.

La comparaison montre un IoU moyen de 0,680 pour la méthode classique, de 0,910 pour YOLO seul et de 0,940 pour l'approche hybride. Le gain principal provient de la segmentation profonde; la géométrie apporte ensuite un gain absolu de 0,030 et réduit l'écart-type à 0,020. Cette lecture évite d'attribuer tout le résultat à une seule composante.

Le temps d'inférence CPU de 476,7 ms par image correspond à environ 2,10 images par seconde. Il permet une capture ponctuelle, mais ne démontre pas un traitement vidéo fluide. L'usage du terme interactif renvoie donc au scénario de déclenchement et de consultation, non à une cadence de diffusion en temps réel.

6.2 Contributions

Les contributions principales sont la formalisation de la problématique en salle de classe, la constitution d'un jeu de 416 images annotées, l'implémentation d'un pipeline hybride modulaire et l'analyse comparative de trois stratégies. La contribution scientifique doit toutefois être comprise dans le cadre d'une validation interne.

6.2.1 Contribution méthodologique

La première contribution est la définition d'un pipeline où les responsabilités du réseau et de la géométrie sont séparées. Le masque neuronal traite la variabilité visuelle; l'analyse de contour et l'homographie produisent une sortie interprétable. Cette séparation permet de comparer les composantes, de diagnostiquer les erreurs et de remplacer un module sans redéfinir toute la chaîne.

La deuxième contribution méthodologique est l'explicitation du protocole. Le mémoire identifie les 336 images d'entraînement, les 80 images de validation et l'absence de test. Il définit les métriques et précise leur ensemble de calcul. Les informations non conservées, comme le détail du matériel d'entraînement ou les effectifs par condition, sont déclarées au lieu d'être reconstruites.

6.2.2 Contribution expérimentale

La contribution expérimentale comprend la constitution et l'annotation polygonale de 416 images de tableaux. Ce corpus permet d'entraîner une classe spécialisée et de comparer

trois approches sur les mêmes images. Sa taille et son origine locale limitent la généralisation, mais il fournit une base cohérente pour une première validation.

La comparaison ne se limite pas à la mAP. Elle inclut l'IoU moyen et sa dispersion, ce qui met en évidence la contribution de la rectification. Les tableaux de gains absolus et la taxonomie des échecs relient les résultats numériques aux comportements observés. Cette articulation répond à la nécessité de bonifier la discussion plutôt que de présenter une simple liste de valeurs.

6.2.3 Contribution applicative

La contribution applicative est un prototype de captation fondé sur des outils accessibles et une caméra standard. L'image rectifiée peut être sauvegardée et servir de ressource de cours. Le système réduit la dépendance à un tableau numérique interactif spécialisé et ouvre la voie à une intégration dans des environnements pédagogiques existants.

Cette contribution reste volontairement circonscrite. L'OCR, le classement automatique, la gestion des droits et l'interface utilisateur ne sont pas évalués. Leur mention dans l'architecture sert à montrer les interfaces futures, non à gonfler le périmètre du prototype. La distinction entre réalisé et envisagé est maintenue dans l'ensemble du mémoire.

6.3 Perspectives

Les travaux futurs devraient d'abord constituer un jeu de test externe provenant d'autres salles et documenter précisément le nombre d'images par condition. Une validation croisée ou plusieurs répétitions avec des graines différentes permettraient d'estimer l'incertitude. L'optimisation du modèle et l'évaluation sur un GPU documenté pourraient ensuite mesurer le potentiel de traitement vidéo. Enfin, un module OCR, notamment pour l'écriture manuscrite et les expressions mathématiques, pourrait être développé et évalué séparément avant son intégration au pipeline.

6.3.1 Extension et documentation du corpus

La priorité est de collecter un test externe dans plusieurs établissements. Le nouvel ensemble devra combiner des images positives et des scènes négatives sans tableau, notamment des murs, des portes, des écrans et des affiches susceptibles de provoquer une

fausse détection. Chaque image devrait être accompagnée d'étiquettes de salle, caméra, session, présence du tableau, luminosité, reflet, occlusion, angle et distance. Ces métadonnées permettraient un découpage par groupe, une mesure des fausses activations et une évaluation par condition.

La procédure d'annotation devrait inclure une double lecture d'un échantillon et une mesure d'accord. Les règles concernant les cadres, les surfaces multiples et les occlusions devraient être illustrées dans un guide. Un contrôle automatique détecterait les polygones invalides, tandis qu'une inspection visuelle confirmerait les cas difficiles.

6.3.2 Renforcement expérimental

Plusieurs répétitions avec des graines différentes permettraient d'estimer la variance de l'entraînement. Une validation croisée par salle ou par session serait préférable à un découpage aléatoire image par image. Les IoU individuels devraient être conservés afin de publier médiane, quartiles, intervalles de confiance et comparaisons appariées.

Des études d'ablation isoleraient l'effet du transfert d'apprentissage, des augmentations, du post-traitement morphologique et du rectangle minimal. La comparaison pourrait aussi inclure une architecture de segmentation supplémentaire. Ces expériences devraient utiliser un protocole fixé avant l'ouverture du test pour éviter d'adapter les choix aux résultats finaux.

6.3.3 Optimisation et déploiement

L'optimisation pourrait explorer un modèle plus léger, la quantification et la réduction de la résolution lorsque la qualité reste acceptable. Les temps devraient être mesurés sur le CPU documenté, sur un GPU identifié et sur le matériel cible. Le rapport devrait séparer latence d'inférence, latence de rectification et latence de bout en bout.

Un prototype d'interface pourrait offrir un aperçu, une confirmation et un choix entre plusieurs régions. Une étude auprès d'enseignants mesurerait la charge d'interaction, le temps de capture et la valeur pédagogique des images. Les fonctions de stockage devraient intégrer authentification, durée de conservation et suppression.

6.3.4 OCR et exploitation du contenu

L'OCR devrait être développé comme une expérience distincte. Les données devraient comporter une transcription de référence et des catégories pour l'écriture manuscrite, le

texte imprimé et les expressions mathématiques. La performance serait mesurée par taux d'erreur de caractères et de mots, complétée par une mesure adaptée aux formules.

La rectification produite par le système pourrait servir d'entrée à l'OCR. Une étude d'ablation comparerait l'image brute, l'image simplement recadrée et l'image rectifiée afin de mesurer le bénéfice réel de la géométrie pour la reconnaissance. Jusqu'à la réalisation de cette étude, aucune performance OCR n'est revendiquée.

Tableau 6.1 - Feuille de route des travaux futurs

Priorité	Action	Résultat attendu
1	Constituer un test externe et documenter les conditions	Mesure de généralisation
2	Répéter les entraînements et conserver les sorties par image	Incertitude quantifiée
3	Mesurer la chaîne complète sur du matériel identifié	Latence de déploiement
4	Évaluer une interface avec des enseignants	Utilité et ergonomie
5	Développer et évaluer l'OCR séparément	Texte exploitable avec métriques dédiées

Le tableau 6.1 ordonne les perspectives selon leur dépendance. La généralisation et la reproductibilité précèdent l'optimisation, puis l'évaluation d'usage et l'OCR. Cette séquence évite d'ajouter des fonctions sur une base expérimentale insuffisamment documentée.

6.3.5 Conclusion générale

Le mémoire démontre la faisabilité d'une captation de tableaux fondée sur une approche hybride. La segmentation profonde fournit la robustesse nécessaire aux scènes réelles, et la géométrie projective améliore la précision spatiale et prépare une image frontale. Les résultats sont solides pour le corpus de validation, à condition de conserver leur portée interne.

La principale leçon méthodologique est que la performance doit être accompagnée d'une description précise des données, des métriques et des limites. En indiquant les informations manquantes, le document fournit une base plus rigoureuse pour la réplication. Le prolongement naturel consiste à tester le système dans de nouvelles salles avant d'étendre ses fonctions pédagogiques.

BIBLIOGRAPHIE

Ali, M. L., et Zhang, Z. (2024). The YOLO framework: A comprehensive review of evolution, applications, and benchmarks in object detection. Computers, 13(12), 336. https://doi.org/10.3390/computers13120336

Ballard, D. H. (1981). Generalizing the Hough transform to detect arbitrary shapes. Pattern Recognition, 13(2), 111-122. https://doi.org/10.1016/0031-3203(81)90009-1

Bay, H., Ess, A., Tuytelaars, T., et Van Gool, L. (2008). SURF: Speeded up robust features. Computer Vision and Image Understanding, 110(3), 346-359.

Bérard, F. (2003). The Magic Table: Computer-vision based augmentation of a whiteboard for creative meetings. Proceedings of the Projector-Camera Systems Workshop.

Bochkovskiy, A., Wang, C.-Y., et Liao, H.-Y. M. (2020). YOLOv4: Optimal speed and accuracy of object detection. arXiv:2004.10934.

Bradski, G. (2000). The OpenCV Library. Dr. Dobb's Journal of Software Tools.

Canny, J. (1986). A computational approach to edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 8(6), 679-698. https://doi.org/10.1109/TPAMI.1986.4767851

Cheng, B., Girshick, R., Dollár, P., Berg, A. C., et Kirillov, A. (2021). Boundary IoU: Improving object-centric image segmentation evaluation. Proceedings of CVPR, 15334-15342.

Dalal, N., et Triggs, B. (2005). Histograms of oriented gradients for human detection. Proceedings of CVPR, 886-893.

Diwan, T., Anirudh, G., et Tembhurne, J. V. (2023). Object detection using YOLO: Challenges, architectural successors, datasets and applications. Multimedia Tools and Applications, 82, 9243-9275. https://doi.org/10.1007/s11042-022-13644-y

Duda, R. O., et Hart, P. E. (1972). Use of the Hough transformation to detect lines and curves in pictures. Communications of the ACM, 15(1), 11-15.

Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., et Zisserman, A. (2010). The PASCAL Visual Object Classes challenge. International Journal of Computer Vision, 88, 303-338.

Girshick, R. (2015). Fast R-CNN. Proceedings of ICCV, 1440-1448.

Golovchinsky, G., Carter, S., et Biehl, J. (2009). Beyond the drawing board: Toward more effective use of whiteboard content. arXiv:0911.0039.

Gonzalez, R. C., et Woods, R. E. (2018). Digital Image Processing (4e éd.). Pearson.

Goodfellow, I., Bengio, Y., et Courville, A. (2016). Deep Learning. MIT Press.

Gor, N., Raval, J., Kalra, R., et Shah, N. (2012). Whiteboard capture and processing for e-learning. International Journal of Computer Applications, 49(17), 10-14.

Haralick, R. M., Sternberg, S. R., et Zhuang, X. (1987). Image analysis using mathematical morphology. IEEE Transactions on Pattern Analysis and Machine Intelligence, 9(4), 532-550.

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. Nature, 585, 357-362. https://doi.org/10.1038/s41586-020-2649-2

Hartley, R., et Zisserman, A. (2004). Multiple View Geometry in Computer Vision (2e éd.). Cambridge University Press.

He, K., Gkioxari, G., Dollár, P., et Girshick, R. (2017). Mask R-CNN. Proceedings of ICCV, 2961-2969.

He, L.-W., et Zhang, Z. (2004). Real-time whiteboard capture and processing using a video camera for teleconferencing. Microsoft Research Technical Report MSR-TR-2004-91.

Jocher, G., Chaurasia, A., et Qiu, J. (2023). Ultralytics YOLO [Logiciel]. Zenodo. https://doi.org/10.5281/zenodo.7347926

Kirillov, A., Wu, Y., He, K., et Girshick, R. (2020). PointRend: Image segmentation as rendering. Proceedings of CVPR, 9799-9808.

Krizhevsky, A., Sutskever, I., et Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems, 25.

Kuo, W., Angelova, A., Malik, J., et Lin, T.-Y. (2019). ShapeMask: Learning to segment novel objects by refining shape priors. Proceedings of ICCV, 9207-9216.

LeCun, Y., Bottou, L., Bengio, Y., et Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324.

Li, M., Lv, T., Chen, J., Cui, L., Lu, Y., Florencio, D., Zhang, C., Li, Z., et Wei, F. (2023). TrOCR: Transformer-based optical character recognition with pre-trained models. Proceedings of AAAI, 37(11), 13094-13102. https://doi.org/10.1609/aaai.v37i11.26538

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., et Zitnick, C. L. (2014). Microsoft COCO: Common objects in context. Proceedings of ECCV, 740-755. https://doi.org/10.1007/978-3-319-10602-1_48

Long, J., Shelhamer, E., et Darrell, T. (2015). Fully convolutional networks for semantic segmentation. Proceedings of CVPR, 3431-3440.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 60, 91-110.

Otsu, N. (1979). A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man, and Cybernetics, 9(1), 62-66.

Padilla, R., Netto, S. L., et da Silva, E. A. B. (2021). A survey on performance metrics for object-detection algorithms. Electronics, 10(3), 279.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., … Chintala, S. (2019). PyTorch: An imperative style, high-performance deep learning library. Advances in Neural Information Processing Systems, 32.

Prewitt, J. M. S. (1970). Object enhancement and extraction. Dans B. S. Lipkin et A. Rosenfeld (dir.), Picture Processing and Psychopictorics (p. 75-149). Academic Press.

Redmon, J., Divvala, S., Girshick, R., et Farhadi, A. (2016). You only look once: Unified, real-time object detection. Proceedings of CVPR, 779-788.

Redmon, J., et Farhadi, A. (2018). YOLOv3: An incremental improvement. arXiv:1804.02767.

Ronneberger, O., Fischer, P., et Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. Proceedings of MICCAI, 234-241.

Rosten, E., et Drummond, T. (2006). Machine learning for high-speed corner detection. Proceedings of ECCV, 430-443.

Sapkota, R., Ahmed, D., et Karkee, M. (2024). Comparing YOLOv8 and Mask R-CNN for instance segmentation in complex orchard environments. Artificial Intelligence in Agriculture, 13, 84-99. https://doi.org/10.1016/j.aiia.2024.07.001

Serra, J. (1982). Image Analysis and Mathematical Morphology. Academic Press.

Smith, R. (2007). An overview of the Tesseract OCR engine. Proceedings of ICDAR, 629-633.

Sobel, I., et Feldman, G. (1968). A 3x3 isotropic gradient operator for image processing. Stanford Artificial Intelligence Project.

Sonka, M., Hlavac, V., et Boyle, R. (2015). Image Processing, Analysis, and Machine Vision (4e éd.). Cengage Learning.

Szeliski, R. (2022). Computer Vision: Algorithms and Applications (2e éd.). Springer.

Ultralytics. (2023). YOLOv8 documentation: Training, segmentation and validation. https://docs.ultralytics.com

Wang, C.-Y., Bochkovskiy, A., et Liao, H.-Y. M. (2023). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. Proceedings of CVPR, 7464-7475.

Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J. M., et Luo, P. (2021). SegFormer: Simple and efficient design for semantic segmentation with transformers. Advances in Neural Information Processing Systems, 34, 12077-12090.

Zhang, Z., et He, L.-W. (2003). Whiteboard scanning and image enhancement. Microsoft Research Technical Report MSR-TR-2003-39.

ANNEXES

Annexe A Principales formules

Cette annexe rassemble les expressions utilisées dans le mémoire et précise leur interprétation. Elle ne remplace pas les explications données dans les chapitres 2 et 5; elle fournit un aide-mémoire pour la réplication du traitement.

$$g(x,y) = 1 \text{ si } f(x,y) \geq T; 0 \text{ sinon} \qquad (A.1)$$

f(x,y) est l'intensité du pixel, T le seuil et g(x,y) la valeur binaire.

Le seuillage transforme une image en deux ensembles de pixels. Dans le pipeline classique, le choix de T détermine directement la quantité de contour conservée. Un seuil trop faible admet le bruit et les reflets; un seuil trop élevé peut supprimer une frontière faiblement contrastée.

$$\rho = x\cos(\theta) + y\sin(\theta) \qquad (A.2)$$

ρ désigne la distance à l'origine et θ l'orientation de la normale à la droite.

La transformée de Hough regroupe dans un espace paramétrique les pixels susceptibles d'appartenir à une même droite. Les maxima représentent des lignes candidates. Leur sélection doit être complétée par des contraintes de longueur, d'orientation et de cohérence géométrique.

$$x' \sim Hx \qquad (A.3)$$

x et x′ sont des coordonnées homogènes; H est la matrice projective 3 × 3.

L'égalité est définie à un facteur d'échelle près. Quatre correspondances de points non dégénérées permettent d'estimer l'homographie. Dans ce travail, les points sources sont les coins du tableau et les points d'arrivée ceux d'un rectangle frontal.

$$IoU = |A\_prédite \cap A\_vérité| \,/\, |A\_prédite \cup A\_vérité| \qquad (A.4)$$

A_prédite est la région produite et A_vérité la région annotée.

L'IoU varie entre 0 et 1. Une valeur nulle indique l'absence de recouvrement; une valeur égale à 1 indique une superposition parfaite. La mesure pénalise à la fois les pixels manquants et les pixels ajoutés autour de la vérité terrain.

$$Précision = VP \,/\, (VP + FP)\ ;\ Rappel = VP \,/\, (VP + FN) \qquad (A.5)$$

VP, FP et FN désignent les vrais positifs, faux positifs et faux négatifs.

La précision répond à la question « parmi les tableaux annoncés, quelle proportion est correcte? ». Le rappel répond à la question « parmi les tableaux présents, quelle proportion est retrouvée? ». Leur lecture conjointe évite de favoriser une méthode qui réduit artificiellement ses détections ou qui accepte trop de régions.

$$AP = \int_0^1 p(r)\, dr \qquad (A.6)$$

p(r) représente la précision interpolée en fonction du rappel r.

Dans la pratique, l'intégrale est approchée à partir des points de la courbe précision-rappel. La mAP agrège ensuite les AP. La mAP@0.5 utilise un seuil IoU de 0,5, tandis que la mAP@0.5:0.95 moyenne plusieurs seuils et évalue plus strictement la localisation.

Annexe B Organisation du jeu de données

Le répertoire contient images/train, images/val, labels/train, labels/val et data.yaml. Chaque fichier d'annotation associe l'identifiant de classe 0 aux coordonnées normalisées du polygone décrivant le tableau.

La structure parallèle des dossiers garantit une correspondance univoque entre une image et son annotation. Le nom de base doit être identique; seule l'extension change. Avant tout entraînement, un contrôle automatique doit repérer les images sans étiquette, les étiquettes sans image et les fichiers impossibles à lire.

Tableau B.1 - Structure logique du corpus expérimental

Répertoire ou fichier	Contenu	Contrôle recommandé
images/train	336 images destinées à l'apprentissage	Ouverture, dimensions et doublons
labels/train	Polygones associés aux images d'apprentissage	Format, classe et coordonnées
images/val	80 images destinées à la validation	Absence de fuite avec train
labels/val	Polygones de vérité terrain pour la validation	Correspondance univoque
data.yaml	Chemins et définition de la classe board	Chemins résolus et index de classe

Le tableau B.1 relie chaque emplacement du corpus au contrôle à effectuer avant l'entraînement.

B.1 Fiche descriptive du corpus

La fiche descriptive distingue les informations effectivement conservées de celles qui devront être ajoutées lors d'une extension. Le corpus comporte 416 images positives, chacune contenant au moins un tableau, et une seule classe. Aucune image négative sans tableau n'a été incluse. La répartition vérifiée est de 336 images pour l'entraînement et 80 pour la validation. Aucun ensemble de test indépendant n'a été conservé dans les éléments disponibles.

Tableau B.2 - Fiche descriptive minimale du jeu de données

Champ	Valeur ou état
Tâche	Segmentation d'instances d'une surface de tableau
Classe	board (identifiant 0)
Taille totale	416 images
Entraînement	336 images
Validation	80 images
Test indépendant	Non constitué
Unité d'annotation	Polygone de la surface visible du tableau
Conditions par image	Non étiquetées de façon systématique

Champ	Valeur ou état
Droits et consentement	À documenter avant diffusion externe du corpus

Le tableau B.2 distingue les valeurs vérifiées des informations encore manquantes. Les conditions d'acquisition sont visibles dans certains exemples, mais leurs effectifs ne peuvent pas être déduits de quatre illustrations. Une future version du corpus devra associer à chaque image des métadonnées explicites: salle, dispositif, distance, angle, éclairage, reflets, occlusion et nombre de tableaux.

B.2 Procédure de contrôle des annotations

Le premier contrôle est syntaxique. Chaque ligne doit commencer par l'identifiant de classe, suivi d'un nombre pair de coordonnées normalisées. Les valeurs doivent rester dans l'intervalle [0, 1]. Un polygone trop court ou de surface nulle doit être rejeté.

Le deuxième contrôle est visuel. Les polygones sont superposés aux images et examinés par lots. L'inspection doit rechercher les bords manquants, l'inclusion excessive du mur, les sommets mal ordonnés et les surfaces partiellement cachées. Les cas ambigus sont consignés avant correction.

Le troisième contrôle porte sur la cohérence du partage. Deux vues presque identiques d'une même scène ne doivent pas se trouver de part et d'autre de la séparation entraînement-validation. Si des séquences sont utilisées, la séparation doit être faite par séquence ou par salle plutôt que par image. Le tableau B.3 synthétise les décisions à prendre lorsqu'un contrôle révèle un écart.

Tableau B.3 - Liste de vérification de la qualité des annotations

Étape	Question de contrôle	Décision en cas d'écart
Correspondance	Chaque image possède-t-elle une annotation?	Retirer ou annoter le fichier
Validité	Les coordonnées sont-elles normalisées et dans l'image?	Corriger le polygone
Sémantique	Le polygone suit-il la surface utile?	Réannoter selon la consigne
Cohérence	Les cas semblables sont-ils traités pareil?	Arbitrage et mise à jour du guide
Séparation	Existe-t-il des doublons entre train et val?	Regrouper puis refaire le partage

Annexe C Commande d'entraînement

La commande ci-dessous reproduit les paramètres vérifiés du mémoire. Elle suppose que l'environnement Ultralytics est installé, que le fichier data.yaml est accessible et que les poids initiaux yolov8s-seg.pt ont été obtenus.

```
yolo segment train model=yolov8s-seg.pt data=dataset_seg/data.yaml imgsz=640
epochs=100 batch=8
```

C.1 Paramètres et portée

Tableau C.1 - Lecture de la commande d'entraînement

Paramètre	Valeur	Interprétation
model	yolov8s-seg.pt	Poids de départ du petit modèle de segmentation
data	dataset_seg/data.yaml	Description du corpus et de la classe
imgsz	640	Taille nominale des images d'entrée
epochs	100	Nombre maximal d'époques
batch	8	Nombre d'images traitées par lot

Le tableau C.1 explicite le rôle de chaque argument de la commande. Celle-ci ne suffit pas à elle seule à garantir une reproduction exacte. La graine, les versions complètes des bibliothèques, le matériel d'entraînement et certains réglages internes ne sont pas documentés dans les journaux disponibles. Une nouvelle expérience devra les exporter avec la configuration finale.

C.2 Séquence de réplication recommandée

1. Créer un environnement Python isolé et enregistrer la version de l'interpréteur. 2. Installer des versions figées d'Ultralytics, PyTorch, OpenCV et NumPy. 3. Vérifier l'intégrité du corpus et conserver une empreinte des fichiers. 4. Exécuter l'entraînement avec une graine consignée. 5. Archiver la configuration, les courbes, les poids et les journaux. 6. Recalculer les métriques sur la validation sans modifier les seuils.

Pour comparer plusieurs méthodes, les prédictions par image doivent être conservées. Cette conservation permet de recalculer les métriques, d'étudier les cas communs et d'effectuer une comparaison appariée. Un simple tableau de moyennes ne permet pas ces analyses.

Tableau C.2 - Artefacts nécessaires à une réplication complète

Artefact	Moment de conservation	Utilité
Environnement figé	Avant l'entraînement	Recréer les dépendances
Empreinte du corpus	Avant l'entraînement	Vérifier l'identité des données
Configuration complète	Au lancement	Connaître tous les hyperparamètres
Courbes et journaux	Pendant l'entraînement	Diagnostiquer convergence et surapprentissage
Poids best et last	À la fin	Comparer sélection et dernière époque
Prédictions par image	Pendant l'évaluation	Analyser erreurs et incertitude

Le tableau C.2 précise à quel moment chaque artefact doit être conservé et à quelle analyse il pourra servir.

Annexe D Pseudocode de la chaîne de traitement

Le pseudocode suivant formalise la chaîne décrite dans les chapitres 3 et 4. Il s'agit d'une représentation de référence destinée à rendre les décisions explicites; il ne prétend pas reproduire ligne pour ligne un dépôt logiciel qui n'a pas été archivé avec le mémoire.

```
ENTRÉE : image I, poids du modèle M
SORTIE : image rectifiée R ou état d'échec

vérifier que I est lisible
redimensionner et normaliser I selon l'interface du modèle
P ← inférence_segmentation(M, I)
C ← filtrer(P, classe = board, seuil_confiance)
si C est vide : retourner ECHEC_AUCUN_TABLEAU
S ← sélectionner_région_principale(C)
B ← binariser_et_redimensionner_masque(S, taille_originale)
B ← fermeture_morphologique(B)
K ← plus_grand_contour_valide(B)
si aire(K) < seuil_aire : retourner ECHEC_REGION_TROP_PETITE
Q ← approximation_polygonale(K)
si nombre_sommets(Q) ≠ 4 : Q ← rectangle_orienté_minimal(K)
Q ← ordonner_coins(Q)
si quadrilatère_dégénéré(Q) : retourner ECHEC_GEOMETRIE
D ← dimensions_de_sortie(Q)
H ← calculer_homographie(Q, rectangle(D))
R ← appliquer_transformation(I, H, D)
enregistrer métadonnées, masque, coins et état
retourner R
```

D.1 Validation des entrées et de la prédiction

La validation de l'entrée distingue une image absente, un format non pris en charge et une image de dimensions nulles. Cette distinction évite d'attribuer au modèle un échec de lecture. Les dimensions originales sont conservées, car le masque retourné doit être remis à l'échelle avant la recherche du contour.

La sélection de la classe board doit être explicite. Si plusieurs régions sont présentes, la plus grande surface constitue une règle déterministe adaptée au scénario principal, mais elle ne couvre pas toutes les salles. L'interface future pourrait présenter les régions à l'utilisateur ou conserver toutes les instances.

D.2 Traitement du masque

Le redimensionnement du masque utilise les dimensions de l'image source. Après seuillage, une fermeture morphologique peut combler de petites discontinuités. La taille du noyau doit être fixée relativement à la résolution afin de ne pas fusionner des régions indépendantes.

Le contour principal est filtré par son aire et, si nécessaire, par sa position. Un seuil d'aire absolu dépend de la résolution; un rapport entre l'aire du contour et l'aire de l'image est plus transportable. La valeur utilisée doit être enregistrée avec la prédiction.

D.3 Construction du quadrilatère

L'approximation polygonale dépend d'un coefficient appliqué au périmètre. Une faible valeur conserve trop de sommets; une valeur élevée peut supprimer un coin. Le rectangle orienté minimal offre une solution de repli stable, mais il peut élargir la région quand le masque est incomplet.

L’ordre des coins est vérifié avant l’homographie. Une procédure classique utilise la somme $x + y$ pour repérer les coins supérieur gauche et inférieur droit, et la différence $y - x$ pour les deux autres. Une vérification d’orientation et d’aire signée limite les permutations lorsque la perspective est forte.

D.4 Traçabilité d’une exécution

Le tableau D.1 définit le noyau minimal du journal d’inférence. Il distingue les informations nécessaires à la reproductibilité de celles qui doivent être limitées pour protéger le contenu pédagogique.

Tableau D.1 - Champs minimaux du journal d'inférence

Champ de journal	Exemple de type	Justification
image_id	Texte	Relier la sortie à l'entrée
timestamp	Date-heure	Ordonner les traitements
model_version	Texte ou empreinte	Identifier les poids
confidence	Nombre réel	Analyser la sélection
mask_area_ratio	Nombre réel	Détecter une région incohérente
corner_coordinates	Liste de quatre points	Rejouer la rectification
status	Code contrôlé	Distinguer succès et causes d'échec
latency_ms	Nombre réel	Mesurer la performance

Annexe E Protocole détaillé d'évaluation

E.1 Préparation de l'évaluation

L'évaluation commence par le gel des données, des poids et de la configuration. Chaque méthode reçoit exactement les mêmes images de validation. Les sorties sont conservées sous une forme permettant de relier la prédiction à la vérité terrain.

Pour la segmentation, les coordonnées du masque sont remises à la résolution originale. La vérité terrain doit être convertie selon la même convention. Les mesures ne doivent pas mélanger des masques à 640 × 640 et des annotations aux dimensions natives.

E.2 Enregistrement par image

Tableau E.1 - Schéma proposé pour les résultats par image

Variable	Type	Rôle
image_id	Identifiant	Assurer la comparaison appariée
has_ground_truth	Booléen	Vérifier l'annotation attendue
prediction_count	Entier	Repérer absences et détections multiples
confidence	Réel	Étudier le seuil de décision
iou_mask	Réel	Mesurer le recouvrement
is_true_positive	Booléen	Calculer précision et rappel
latency_inference_ms	Réel	Mesurer le réseau
latency_geometry_ms	Réel	Mesurer le post-traitement
failure_category	Catégorie	Analyser les limites

Le tableau E.1 définit les variables nécessaires à une comparaison appariée. La conservation des valeurs par image permet de construire des intervalles de confiance par rééchantillonnage et d'appliquer des tests appariés. Elle facilite également une analyse qualitative reproductible: une image difficile peut être retrouvée sans dépendre d'une capture d'écran isolée.

E.3 Agrégation et comparaison

Les métriques sont d'abord calculées séparément pour chaque méthode. La précision, le rappel et les mAP sont accompagnés du nombre d'images et du seuil employé. Les gains absolus sont obtenus par différence de points, tandis qu'un gain relatif doit être annoncé comme tel et préciser son dénominateur.

Une comparaison robuste devrait répéter l'entraînement avec plusieurs graines. La moyenne décrit la tendance; l'écart-type décrit la variabilité. En l'absence de répétitions, comme dans les résultats disponibles, les écarts observés restent descriptifs et ne doivent pas être présentés comme une preuve de significativité.

E.4 Mesure de la latence

La latence est mesurée après une phase de chauffe afin de réduire l'effet du premier chargement. Le chronométrage distingue lecture, prétraitement, inférence, géométrie et

écriture. La moyenne, la médiane et un percentile élevé donnent une vue plus utile qu'une seule moyenne.

Tableau E.2 - Périmètres de mesure du temps

Mesure	Début	Fin
Prétraitement	Image chargée	Entrée prête pour le modèle
Inférence	Appel du modèle	Masques disponibles
Géométrie	Masque sélectionné	Image rectifiée
Bout en bout	Demande de capture	Fichier enregistré

Le tableau E.2 empêche de confondre le temps du réseau avec la latence complète ressentie pendant une capture.

E.5 Rapport final

Le rapport final doit présenter les métriques globales, les distributions par image et un échantillon raisonné de réussites et d'échecs. Les images choisies doivent illustrer des catégories définies à l'avance plutôt que seulement les meilleurs résultats.

La conclusion de l'évaluation rappelle enfin son domaine de validité: corpus local, validation interne et configuration documentée. Cette formulation n'empêche pas de montrer l'intérêt du prototype; elle sépare clairement la preuve obtenue des hypothèses qui restent à tester.

Annexe F Grille d'analyse des échecs et métadonnées futures

F.1 Catégories d'échec

La grille transforme l'observation qualitative en données analysables. Une image peut recevoir plusieurs étiquettes, par exemple faible contraste et reflet. La cause du défaut doit être distinguée de son effet sur la sortie.

Tableau F.1 - Taxonomie proposée des conditions difficiles

Code	Condition observée	Effet possible
LUM	Éclairage insuffisant ou non uniforme	Masque incomplet
REF	Reflet spéculaire	Rupture de frontière
OCC	Occlusion par une personne ou un objet	Contour tronqué
PER	Perspective prononcée	Coins instables
DIS	Tableau éloigné ou de petite taille	Absence de détection
MUL	Plusieurs tableaux visibles	Sélection de la mauvaise instance
BGD	Arrière-plan visuellement similaire	Faux positif

Le tableau F.1 attribue un code stable à chaque condition afin de rendre les analyses comparables entre plusieurs exécutions.

F.2 Distinction entre cause et effet

Une catégorie de scène comme REF décrit une cause potentielle. L'effet doit être enregistré séparément: détection absente, région excédentaire, région manquante, coins mal ordonnés ou rectification déformée. Cette séparation permet de constater qu'un même reflet n'entraîne pas toujours le même défaut.

Tableau F.2 - Taxonomie proposée des effets sur la sortie

Code d'effet	Description	Indicateur quantitatif associé
DET0	Aucune instance retournée	Faux négatif
DETX	Instance incorrecte retournée	Faux positif
MSK-	Partie du tableau absente du masque	Rappel de pixels ou IoU
MSK+	Région extérieure ajoutée au masque	Précision de pixels ou IoU
CRN	Coins incorrects ou mal ordonnés	Erreur moyenne de coin
REC	Rectification inutilisable	Contrôle géométrique ou jugement annoté

Le tableau F.2 complète la cause observée par un effet mesurable sur la sortie.

F.3 Schéma de métadonnées

Les métadonnées d'acquisition doivent être définies avant une nouvelle collecte. Elles servent à équilibrer le corpus, à produire des résultats par sous-groupe et à décrire la portée de l'expérience. Elles ne doivent pas être reconstruites de mémoire après l'entraînement.

Tableau F.3 - Champs proposés pour la prochaine collecte

Champ	Valeurs proposées	Mode d'obtention
room_id	Identifiant pseudonymisé	Saisi lors de la collecte
device_id	Modèle ou code du dispositif	Métadonnée de capture
distance_band	proche / moyenne / éloignée	Mesure ou estimation contrôlée
view_angle	faible / moyen / fort	Calcul ou annotation
lighting	uniforme / faible / mixte	Annotation
reflection	absent / léger / important	Annotation
occlusion	0 / <25 % / ≥25 %	Annotation
board_count	Entier	Annotation

Le tableau F.3 fournit le schéma de métadonnées à appliquer dès la prochaine collecte.

F.4 Plan d'analyse par sous-groupe

Chaque sous-groupe doit contenir assez d'images pour que sa métrique soit interprétable. Le rapport indiquera l'effectif en plus de la valeur. Lorsque deux facteurs se recouvrent, par exemple reflet et perspective, l'analyse doit signaler ce chevauchement au lieu d'attribuer toute la baisse à un seul facteur.

L'échantillonnage futur devrait réserver des salles ou des dispositifs entiers au test externe. Cette stratégie mesure mieux la généralisation qu'un partage aléatoire d'images très proches. Les métriques globales seraient alors complétées par les métriques des conditions difficiles et par leur intervalle d'incertitude.

Annexe G Matrice de traçabilité méthodologique

Cette annexe relie les objectifs du mémoire, les opérations réalisées, les preuves présentées et les limites déclarées. Elle facilite la vérification du raisonnement sans ajouter de nouveaux résultats expérimentaux.

G.1 Des objectifs aux preuves

Tableau G.1 - Correspondance entre objectifs, preuves et limites

Objectif	Opération	Preuve présentée	Limite associée
Localiser le tableau	Segmentation YOLOv8s-seg	Précision, rappel et mAP	Validation interne
Décrire la frontière	Prédiction d'un masque	IoU et exemples visuels	80 images de validation
Extraire quatre coins	Contours et approximation	Pipeline et cas illustrés	Erreur de coin non mesurée
Rectifier la vue	Homographie projective	Images avant-après	Qualité perceptuelle non notée
Comparer les approches	Classique, YOLO et hybride	Tableau comparatif	Une exécution documentée
Évaluer l'usage	Analyse de latence CPU	476,7 ms par image	Pas d'étude utilisateur

Le tableau G.1 montre que chaque objectif technique possède au moins une preuve dans le corps du mémoire. Elle montre également que la portée des preuves varie. Les métriques de segmentation sont quantitatives, tandis que la rectification est principalement illustrée. Une prochaine étude devrait ajouter une mesure d'erreur des coins et une évaluation de la lisibilité après transformation.

L'objectif d'usage pédagogique reste prospectif. Le système produit une image destinée à l'archivage, mais l'utilité pour les enseignants et les étudiants n'a pas été mesurée. Cette distinction évite de confondre faisabilité technique et bénéfice pédagogique démontré.

G.2 Discipline des affirmations

Une affirmation est considérée comme étayée lorsqu'elle renvoie à une mesure, à une figure, à un tableau, à une observation explicitement décrite ou à une source bibliographique. Les formulations absolues sont évitées lorsque la preuve ne couvre qu'un corpus local.

Tableau G.2 - Règles de formulation des conclusions

Type d'affirmation	Formulation retenue	Formulation à éviter
Performance	Le modèle atteint la valeur X sur la validation	Le modèle est toujours précis
Robustesse	Les exemples couvrent plusieurs variations observées	Le système résiste à toute condition
Généralisation	Elle reste à tester sur un corpus externe	Les résultats valent pour toutes les salles
Temps	La latence mesurée sur le CPU documenté est X	Le système fonctionne en temps réel

Type d'affirmation	Formulation retenue	Formulation à éviter
OCR	L'OCR constitue une perspective séparée	Le système comprend automatiquement le contenu

Les règles du tableau G.2 sont appliquées à la discussion et à la conclusion. Elles n'affaiblissent pas la contribution: elles indiquent avec précision ce qui a été observé et ce qui demeure une hypothèse de travail.

G.3 Traçabilité des données numériques

Les valeurs principales doivent pouvoir être retrouvées dans un tableau ou dans un journal. Les effectifs 416, 336 et 80 décrivent respectivement le corpus complet, l'entraînement et la validation. Les gains du système hybride sont calculés à partir des métriques comparatives, sans inventer de mesures intermédiaires.

Tableau G.3 - Registre des principales valeurs expérimentales

Valeur	Signification	Emplacement de vérification
416	Nombre total d'images	Chapitre 4 et annexe B
336	Images d'entraînement	Tableau 4.3 et annexe B
80	Images de validation	Tableau 4.3 et annexe B
100	Nombre maximal d'époques	Tableau 4.4 et annexe C
640	Taille nominale d'entrée	Tableau 4.4 et annexe C
476,7 ms	Latence d'inférence CPU	Chapitre 5

Le tableau G.3 centralise les principales valeurs numériques. Les effectifs par condition d'acquisition ne figurent pas dans ce registre, car ils n'ont pas été consignés. La version corrigée signale explicitement cette absence et propose un schéma de métadonnées pour la prochaine collecte. Cette transparence est préférable à une estimation fondée sur quelques images illustratives.

G.4 Vérification avant diffusion

Avant une nouvelle diffusion du mémoire, le document et les artefacts expérimentaux peuvent être contrôlés selon la liste suivante. Le contrôle documentaire porte sur les titres, la numérotation, les renvois et la bibliographie. Le contrôle scientifique porte sur la provenance des données, la cohérence des métriques et les limites.

Tableau G.4 - Liste de vérification documentaire et scientifique

Contrôle	Critère d'acceptation	Trace à conserver
Structure	Six chapitres cohérents et titres hiérarchisés	Table des matières
Équations	Numéro unique et explication des symboles	Liste de vérification
Figures et tableaux	Légende, numéro et renvoi dans le texte	Index et inspection visuelle
Citations	Chaque source mobilisée apparaît en bibliographie	Audit des références
Données	Effectifs cohérents et absence de valeur inventée	Fiche du corpus

Contrôle	Critère d'acceptation	Trace à conserver
Résultats	Même périmètre et mêmes seuils pour la comparaison	Fichiers par image
Limites	Portée interne et informations manquantes déclarées	Discussion
Fichiers finaux	Word et PDF lisibles, pagination identique	Copie archivée

Le tableau G.4 définit le contrôle final. Une version est considérée comme prête lorsque les deux formats présentent le même contenu, que les pages sont lisibles et que les index renvoient aux pages correctes. Les fichiers de données et de code, lorsqu'ils seront archivés, devront recevoir un identifiant de version distinct de celui du mémoire.

G.5 Synthèse de la traçabilité

Le fil directeur du travail peut ainsi être suivi depuis la problématique jusqu'aux preuves: détection et segmentation du tableau, consolidation géométrique, rectification, comparaison et analyse des limites. Les résultats justifient l'intérêt de l'approche hybride sur la validation disponible.

La traçabilité fait également apparaître les travaux indispensables pour transformer le prototype en solution généralisable: test externe, répétitions d'entraînement, métadonnées d'acquisition, mesure séparée de la géométrie et étude d'usage. Ces points sont intégrés à la feuille de route plutôt que présentés comme déjà résolus.

Annexe H Fiche de réplication à compléter

Cette fiche peut être remplie lors d'une nouvelle exécution afin de conserver les informations absentes des journaux actuels. Le tableau H.1 identifie l'environnement; les cases laissées en blanc signalent immédiatement les éléments qui limiteraient la comparaison.

Tableau H.1 - Identification de l'environnement de réplication

Rubrique	Information à consigner
Date et responsable	_____
Version Python	_____
Version Ultralytics	_____
Version PyTorch	_____
Version OpenCV	_____
Système et matériel	_____
Graine aléatoire	_____
Empreinte du corpus	_____
Empreinte des poids	_____

Tableau H.2 - Contrôles à effectuer pendant la réplication

Vérification	État ou valeur
Nombre d'images train / val / test	_____
Contrôle des doublons effectué	Oui / Non
Annotations contrôlées visuellement	Oui / Non — échantillon : _____
Configuration complète archivée	Oui / Non — chemin : _____
Poids best et last archivés	Oui / Non
Prédictions par image conservées	Oui / Non
Latence mesurée après chauffe	Oui / Non — nombre de répétitions : _____
Métriques recalculées sans modification	Oui / Non
Écarts par rapport au protocole	_____

Le tableau H.2 sert de liste de contrôle pendant l'exécution. La fiche complétée doit être archivée avec les poids, les courbes et les résultats par image. Elle permet de comparer deux exécutions sans dépendre de la mémoire de l'expérimentateur et renforce la reproductibilité du protocole.