

La détection des polypes dans les vidéos issues de coloscopies

Comme on a pu le voir précédemment, le temps de calcul pour traiter une image fixe est de moins de 40 millisecondes, ce qui traduit une capacité à traiter toutes les images à la vitesse minimale de 25 images par seconde. On peut donc s'intéresser à la capacité de notre outil à détecter des polypes dans des vidéos issues de coloscopies, et surtout s'intéresser à la compatibilité avec une utilisation clinique de notre outil.

En faisant des essais sur des vidéos, il est alors possible de se rapprocher des cas réels de coloscopie et de prouver l'intérêt de la méthode de détection des polypes pour des examens en conditions hospitalières.

Mais il ne faut pas uniquement se concentrer sur le dernier point de la compatibilité avec une utilisation clinique, on garde toujours en mémoire toutes les contraintes de notre outil, que l'on rappelle ici :

- être capable de détecter les polypes en temps réel (c'est-à-dire que le temps pour traiter une image issue de coloscopie doit être inférieur à 40 millisecondes)
- être compatible avec du matériel dont les ressources de calculs sont standards et limitées
- obtenir des performances élevées en termes de sensibilité et de spécificité
- être compatible avec l'utilisation en routine clinique (et donc ne pas modifier la pratique clinique du médecin) et alerter le médecin sur la présence de polypes durant l'examen.

Dans cette partie, on va donc se consacrer aux performances dans les vidéos, et pour cela, on va introduire dans un premier temps de nouvelles métriques permettant de mieux traduire ces performances. On introduit également une nouvelle base de données contenant des vidéos de coloscopies. Par la suite, on présentera les résultats de cette méthode sur cette base de données puis les différentes optimisations mises en place et possibles afin d'améliorer les performances de l'outil.

3.1 La détection dans des vidéos

3.1.1 Introduction de nouvelles métriques pour la mesure de performances dans les vidéos

Tout comme avec les images fixes, il faut des métriques pour mesurer les performances de notre outil. Précédemment, afin de classer un résultat de détection sur une image, on s'est basé sur la méthode dite du centroïde, c'est-à-dire que pour classer un résultat, on regardait sur la vérité terrain si le centre de la région d'intérêt détectée se situait bien sur la zone contenant le polype. Mais cette méthode est stricte et surtout, ne rend pas compte de l'aspect visuel de la détection, qui ne traduit donc pas le fait d'attirer l'attention du médecin sur la zone de l'image contenant le polype à détecter dans une vidéo. On se propose alors, en plus d'utiliser une partie des métriques présentées précédemment (Précision, Rappel et F_1 Score), d'introduire de nouvelles métriques utilisables cliniquement [ABSM⁺17a], qui permettent de mieux traduire les performances de la méthode d'un point de vue clinique, et donc de mieux interpréter les résultats de la méthode utilisée (il s'agit ici d'être plus clair pour les médecins). Ces nouvelles métriques sont alors les suivantes :

- Le **taux de détection des polypes (Polyp Detection Rate, PDR)**. Il permet de montrer si la méthode a été capable de détecter le polype au moins une fois dans la séquence.
- Le **temps de calcul moyen par image (Mean Processing Time, MPT)**. Il correspond au temps moyen mis pour traiter une image. Il est associé à la contrainte temps réel pour l'utilisation clinique. Si on considère que les vidéos sont enregistrées à 25 images par secondes, alors la contrainte temps réel est de 40 millisecondes maximum.
- Le **nombre moyen de faux positifs par image (Mean Number of False Positive, MNFP)**. Il s'agit avec cette mesure de savoir si la méthode utilisée pour détecter les polypes ne perturbe pas l'attention du médecin en affichant de fausses détections.
- Le **temps de réaction (Reaction Time, RT)**. Cela représente le délai entre la première apparition d'un polype dans la vidéo et la première bonne détection de ce polype par la méthode utilisée, mesure effectuée à la fois en nombre d'images et en secondes (si on considère la vidéo défilant à 25 images par secondes). La figure 3.1 illustre comment est calculé le temps de réaction.

Ce sont donc ces métriques que l'on va utiliser par la suite pour mesurer les performances de notre outil et comparer les résultats des différentes optimisations par rapport à la méthode de base.

Afin de pouvoir calculer les différentes métriques (et en particulier le Rappel et la Précision), il faut pouvoir classer correctement les différentes détections renvoyées par le système, c'est-à-dire si une région d'intérêt détectée se classe comme vraie positive, fausse positive, fausse négative ou vraie négative. Pour cela, nous décidons de mettre au point un nouveau processus permettant de classer au mieux les zones détectées. Ce processus

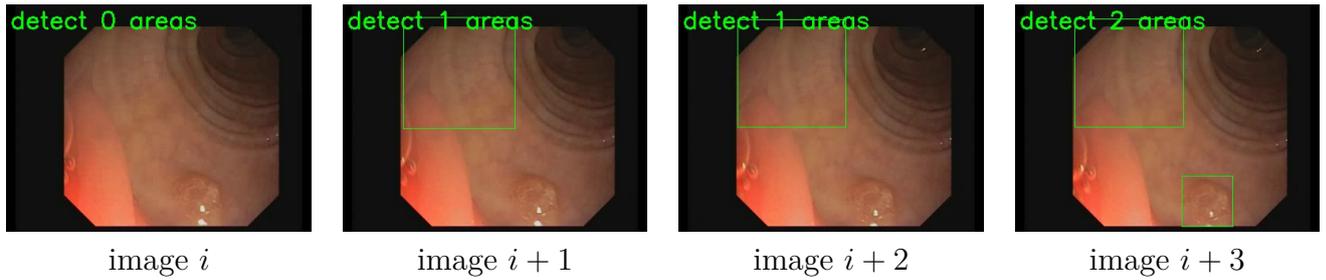


FIGURE 3.1 – Illustration du temps de réaction. Le polype apparaît pour la première fois à l'image i et est correctement détecté à l'image $i + 3$, il s'est donc écoulé un total de 4 images pour avoir une détection correcte après la première apparition du polype, donc un temps de réaction de 160 millisecondes ($4 \times 40ms = 160ms$)

est présenté dans la figure 3.2. En effet, précédemment, pour classer une détection, on se basait sur la méthode dite du centroïde : on comparait la position du centre de la région d'intérêt détectée par rapport à la vérité terrain. Mais cette méthode était trop stricte et ne traduisait pas l'aspect suivant : attirer l'oeil du médecin sur le polype. Avec le processus présenté ici, on tient compte de cela.

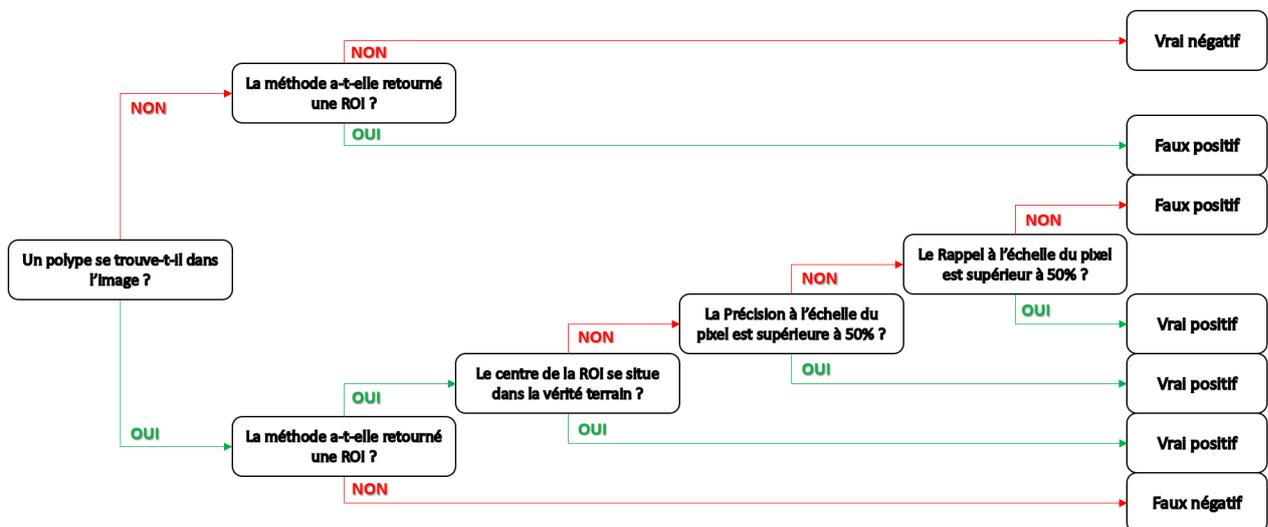


FIGURE 3.2 – Processus de classification d'une région d'intérêt (ROI)

Ce processus fonctionne à deux échelles : d'abord à l'échelle de l'image, puis à l'échelle du pixel. En s'appuyant sur la vérité terrain, on sait si un polype est présent ou non sur l'image traitée. On va ensuite enchaîner les différentes étapes de notre arbre de décisions afin de savoir si la région d'intérêt détectée par notre outil est à considérer comme vrai positif, vrai négatif, faux positif ou faux négatif. C'est donc en se basant sur les résultats obtenus, à travers ce processus, qu'on va pouvoir calculer les différentes métriques qui vont refléter les performances de notre méthode de détection des polypes.

De plus, l'intérêt de proposer un nouveau processus afin de mieux classer les régions d'intérêt détectées est d'accorder plus d'importance à l'aspect visuel de la détection et de traduire indirectement la capacité de la méthode à attirer l'oeil du médecin sur le polype sans nécessairement encadrer complètement le polype.

Base de données vidéo

Afin de pouvoir tester notre méthode, on se doit de disposer d'une base de données composée de vidéos issues de coloscopies et non plus seulement d'images fixes. C'est pour cela qu'on décide d'utiliser celle présentée dans [ABSM⁺17b]. Cette base de données dispose de 18 vidéos uniques extraites de coloscopies autour de polypes et de la vérité terrain associée pour chaque image des vidéos, il sera alors possible de se rendre compte des qualités de détection de notre méthode. Ces vidéos possèdent une résolution de 768 pixels par 576 pixels et ont été obtenues grâce à la combinaison d'un endoscope Olympus QF190 et d'un système de numérisation d'image vidéo Exera III. La figure 3.3 représente le profil de la vidéo numéro 18, c'est-à-dire la présence ou non d'un polype à l'écran (0 si le polype est absent, 1 s'il est présent). Plus loin en annexe, les figures A.1.a, A.2.a, A.3.a, A.4.a, A.5.a, A.6.a, A.7.a, A.8.a, A.9.a, A.10.a, A.11.a, A.12.a, A.13.a, A.14.a, A.15.a, A.16.a, A.17.a et A.18.a représentent les profils des 18 vidéos qui permettent de s'apercevoir qu'elles représentent des cas bien réels et variés. Au total, on dispose donc de 11 954 images afin de mener les tests.

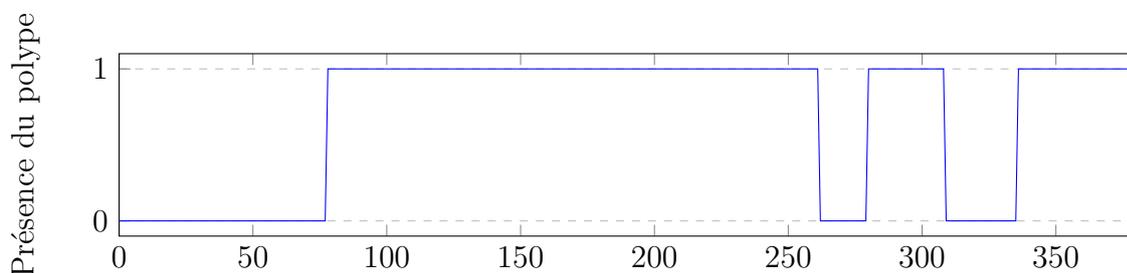


FIGURE 3.3 – Profil de la vidéo 18 (381 images)

Sur ces données, on a décidé d'aller plus loin en classant qualitativement ces 18 vidéos selon qu'il est plus ou moins difficile de détecter le polype. Les critères suivants ont été utilisés :

- La taille du polype, plus il est gros, plus il sera facile de le détecter (et vice versa)
- La netteté de l'image
- La présence de matière non désirée sur l'image (résidus, bulles)

Cela permet alors de classer ces vidéos en leur accordant une note allant de 0 à 3, 0 indiquant que le polype est facile à détecter, 3 indiquant qu'il est difficile à détecter dans

la vidéo. Cette classification est disponible dans le tableau 3.1.

On s’aperçoit donc que certains polypes sont plus difficiles que d’autres à détecter dans les vidéos disponibles dans cette base de données (environ 28%).

Numéro de vidéo	1	2	3	4	5	6	7	8	9
Niveau de difficulté	1	1	0	0	1	0	2	2	0
Numéro de vidéo	10	11	12	13	14	15	16	17	18
Niveau de difficulté	1	1	1	2	2	2	1	1	0

TABLE 3.1 – Niveaux de difficulté de détection des polypes dans les vidéos

Protocole de mesure

Afin de mesurer de manière adaptée les performances de notre méthode, on a mis en place un protocole pour avoir des mesures comparables entre elles. Tout d’abord, côté mesure des performances de détection, on utilise le processus de classification d’une région d’intérêt détectée (figure 3.2) ainsi que toutes les métriques classiques présentées jusque là en se basant sur les résultats du dit processus. En ce qui concerne la mesure des temps de calcul, on prend la moyenne des temps de calcul sur les 18 vidéos après 10 exécutions. Cela permet de rendre moins sensible les mesures aux variations de charge de travail que peut subir l’ordinateur sur lequel on effectue nos tests. Concernant la machine utilisée, il s’agit d’une station de calculs de marque Dell ayant les caractéristiques suivantes : un processeur Intel Xeon E5-1603 v3 (fonctionnant à une fréquence de 2.80 Giga-Hertz), une mémoire vive de 32 Giga-Octets et tournant sous Ubuntu 14.04 LTS (64 bits). L’outil a été développé en utilisant le langage Python, interprété dans sa version 2.7.6 et utilisant la version 3.1.0 d’OpenCV.

3.1.2 Résultats préliminaires de détection dans les vidéos

Avec ces métriques et cette base de données, on est donc capable de mener des tests préliminaires afin de regarder les performances de notre méthode. La méthode développée est décrite dans le tableau 3.2. Cette méthode permet donc de détecter les polypes dans les vidéos de coloscopies en utilisant les deux types de caractéristiques utilisées pour l’apprentissage : les LBP et les caractéristiques pseudo-Haar.

Quand on parle par la suite des mesures de performances de notre outil (en termes de détection et de temps de calcul), il s’agit de l’évaluation de la totalité des différentes étapes présentées dans le tableau 3.2 pour aboutir au résultat final, c’est-à-dire la détection (ou non) d’un polype.

Maintenant que l’on dispose d’une base de données, de métriques adaptées et de la méthode à tester, nous proposons des résultats préliminaires sur cette dernière.

-
1. Une étape de prétraitement de l'image pour préparer la détection (extraction de canal)
 2. Une étape de détection cascade à différentes échelles qui permet de trouver les régions d'intérêt sur l'image contenant de possibles polypes
 3. Une étape d'affichage permettant de retourner le résultats de la détection à l'utilisateur
-

TABLE 3.2 – Algorithme de détection des polypes dans les vidéos

Résultats

Comme vu juste avant, on va effectuer une évaluation préliminaire des performances de la méthode. Pour y parvenir, on va donc utiliser la base de données présentée ainsi que les différentes métriques disponibles. Côté méthode de détection, on va utiliser les deux types de classifications qui sont à notre disposition : la classification utilisant les motifs locaux binaires (LBP) et la classification utilisant les caractéristiques pseudo-Haar effectuée sur la base de données CVC-Clinic (comme dans la chapitre précédent). Pour ces deux types de classification et pour le test, on insiste sur le fait qu'aucune optimisation n'est utilisée. Autrement dit, on n'utilise pas l'apprentissage actif ici car on cherche juste à s'intéresser aux performances brutes de notre méthode. Les résultats sont alors disponibles dans le tableau 3.3.

Méthode	PDR	MPT	MNFP	Précision	Rappel	F ₁ Score	RT
LBP	100%	140ms	9.9	4.48%	55.33%	8.29%	4.8 (0.2s)
Haar	100%	26ms	5.8	9.20%	70.65%	16.27%	6.8 (0.3s)

TABLE 3.3 – Résultats préliminaires de la détection dans les vidéos sans optimisations

Suivant les résultats du tableau 3.3, on s'aperçoit que quelque soit le type de classification utilisé, on est capable d'attirer l'attention du médecin sur chacun des polypes présents dans les vidéos (ce qui correspond au PDR de 100%). Côté temps de calcul, on constate que la classification avec les caractéristiques pseudo-Haar est capable de détecter les polypes en temps réel (MPT de 26 millisecondes), mais aussi propose de meilleures performances de détection que la classification utilisant les motifs locaux binaires. Cependant, à l'inverse, on peut noter que la classification utilisant les motifs locaux binaires est plus rapide pour trouver le polype après sa première apparition dans la vidéo (cela se voit en comparant les temps de réaction entre les deux types de classification).

De plus, quand on compare les temps de calcul pour les LBP par rapport à ceux du chapitre précédent, on s'aperçoit qu'ils sont différents. Cela s'explique par le fait qu'avant, on effectuait la détection à un seul facteur d'échelle, ici, on l'effectue à 3 facteurs d'échelle différents. On a décidé de le faire à 3 échelles différents car on s'est aperçu qu'avec la classification utilisant les paramètres pseudo-Haar, les temps de calcul étaient très faibles,

et qu'en faisant la détection à 3 facteurs d'échelles différents, on rend notre méthode plus robuste pour les différentes tailles de polypes.

On voit donc ici qu'on obtient des résultats intéressants (F_1 Score de 8.29% pour la classification utilisant les LBP et de 16.27% pour celle utilisant les paramètres pseudo-Haar) mais qui peuvent être améliorés par plusieurs optimisations. On va donc s'intéresser par la suite à toutes ces améliorations possibles qui vont permettre d'obtenir de meilleures performances pour la détection des polypes colorectaux dans les vidéos de coloscopies.

3.2 Amélioration des performances dans les vidéos

Dans cette partie, on va s'intéresser à toutes les améliorations qu'il est possible de mettre en place afin d'améliorer les performances de la méthode. Parmi ces améliorations, il y a bien évidemment l'apprentissage actif, qui permet d'obtenir de meilleures performances comme on a pu le voir avec les images fixes. Mais on va aussi introduire d'autres améliorations qui s'appliquent bien à la vidéo (et non plus aux images fixes) comme la cohérence temporelle, la fusion spatiale des zones détectées et des étapes de prétraitements.

3.2.1 Fusion spatiales des zones détectées

On va ici s'intéresser à la première optimisation mise en place afin d'améliorer les résultats de la détection. A la vue du grand nombre de fausses détections de la méthode, cette première optimisation va consister à fusionner spatialement les zones les plus proches afin de réduire le nombre de régions d'intérêt renvoyé par la méthode. On a donc un nouvel algorithme mis en place, disponible dans le tableau 3.4. Ce critère de fusion se base sur deux éléments : le nombre de régions d'intérêt proche et le recouvrement de ces zones entre elles, dans notre cas, on a décidé de fixer ces paramètres à 2 (car on effectue 3 détections à différents facteurs d'échelles) et avec un taux de recouvrement de 70% entre ces zones (car si on prenait en dessous de cette valeur, on ne serait pas assez sélectif, et au dessus, on deviendrait trop sélectif). On teste donc les avantages de cette fusion sur l'ensemble des 18 vidéos. Les résultats sont disponibles dans les tableaux 3.5 et 3.6.

D'après les résultats obtenus (et présentés dans les tableaux 3.5 et 3.6), la fusion spatiale des zones détectées permet d'améliorer nettement les résultats de notre outil sans en dégrader certains aspects. En effet, le temps de calcul moyen pour traiter une image ne change presque pas, donc cet aspect qui nous est essentiel n'est pas affecté. Côté détection, on est toujours capable d'alerter le médecin sur la présence des polypes (PDR de 100%) mais en n'envoyant beaucoup moins de fausses informations (au regard du MNFP), donc l'information renvoyée au médecin est beaucoup plus pertinente et claire. Au niveau performances de détection, le F_1 Score augmente nettement (on passe du simple au double), certes on détecte moins de polype mais on est plus précis sur la détection (au regard des valeurs du rappel et de la précision), ce qui est un aspect important pour le médecin. La seule variable qui se dégrade est le temps de réaction qui augmente (mais reste inférieur à 1 seconde), malgré l'avantage de réduire les fausses détections, la fusion spatiale va aussi

1. Une étape de prétraitement de l'image pour préparer la détection (extraction de canal)
2. Une étape de détection cascade à différentes échelles qui permet de trouver les régions d'intérêt sur l'image contenant de possibles polypes
3. Une étape de fusion spatiale des zones détectées afin de limiter le nombre de fausses détections et ainsi réduire le nombre de régions d'intérêt affichées à l'utilisateur
4. Une étape d'affichage permettant de retourner le résultats de la détection à l'utilisateur

TABLE 3.4 – Algorithme de détection des polypes dans les vidéos avec fusion spatiale

Fusion	PDR	MPT	MNFP	Précision	Rappel	F ₁ Score	RT
Sans	100%	140ms	9.9	4.48%	55.33%	8.29%	4.8 (0.2s)
Avec	100%	140ms	3.2	12.41%	54.58%	20.22%	7.2 (0.3s)

TABLE 3.5 – Comparaison des performances de la méthode avec fusion spatiale pour la classification utilisant les motifs locaux binaires (LBP)

Fusion	PDR	MPT	MNFP	Précision	Rappel	F ₁ Score	RT
Sans	100%	26ms	5.8	9.20%	70.65%	16.27%	6.8 (0.3s)
Avec	100%	26ms	1.4	23.29%	46.82%	31.10%	17.5 (0.7s)

TABLE 3.6 – Comparaison des performances de la méthode avec fusion spatiale pour la classification utilisant les paramètres pseudo-Haar

faire disparaître des bonnes détections, ce qui va alors augmenter le temps de réaction sur certaines vidéos.

Par la suite, on se basera donc sur cet algorithme (présenté tableau 3.4) qui sert de base dans notre méthode. De plus, c'est sur celui là qu'on appliquera les prochaines améliorations.

3.2.2 Apprentissage actif

Ici, on va s'intéresser à l'influence de l'apprentissage actif sur les performances de notre méthode pour les vidéos. Pour rappel, l'apprentissage actif est une méthode qui va permettre de réinjecter des exemples négatifs dans la phase d'apprentissage afin d'améliorer les performances en réduisant le nombre de fausses détections. Et comme on a pu le voir pour les images fixes, l'apprentissage actif va certainement améliorer nos performances.

On va donc comparer les différentes méthodes de classification utilisées avec et sans apprentissage actif sur la base de données de 18 vidéos avec les métriques introduites

précédemment. Les résultats de cette expérimentation sont disponibles dans les tableaux 3.7 et 3.8. Afin d'identifier si la classification utilisée a bénéficié de l'étape d'apprentissage actif ou non, il suffit de regarder le numéro (par exemple le 1 dans *LBP N 1*), ce numéro traduit le nombre de répétitions d'apprentissage actif appliqué à cette classification.

Méthode	PDR	MPT	MNFP	Précision	Rappel	F ₁ Score	RT
<i>LBPN0</i>	100%	140ms	3.2	12.41%	54.58%	20.22%	7.2 (0.3s)
<i>LBPN1</i>	100%	160ms	1.2	23.89%	46.03%	31.46%	13.5 (0.5s)
<i>LBPN2</i>	100%	162ms	0.9	23.42%	35.08%	28.08%	29.4 (1.2s)

TABLE 3.7 – Résultats de la détection dans les vidéos avec l'apprentissage actif pour la classification utilisant les motifs binaires locaux (LBP)

Méthode	PDR	MPT	MNFP	Précision	Rappel	F ₁ Score	RT
<i>HaarN0</i>	100%	26ms	1.4	23.29%	46.82%	31.10%	17.5 (0.7s)
<i>HaarN1</i>	100%	23ms	0.6	35.58%	42.29%	38.65%	25.8 (1.0s)

TABLE 3.8 – Résultats de la détection dans les vidéos avec l'apprentissage actif pour la classification utilisant les paramètres pseudo-Haar

Comme on peut le voir avec les tableaux 3.7 et 3.8, l'apprentissage actif n'impacte pas le PDR, on est donc capable d'alerter le médecin sur la présence du polype dans chacune des 18 vidéos. Côté performances :

- Avec la classification utilisant les motifs binaires locaux (tableau 3.7), quand on passe de la classification sans apprentissage actif à celle avec une étape de d'apprentissage actif, les résultats sont nettement améliorés : on a une augmentation du F₁ Score, qui se traduit par l'augmentation de la précision. Conséquence directe de cet effet, on réduit le nombre de MNFP, donc moins de fausses détections sont retournées au médecin. Cependant, avec une deuxième étape d'apprentissage actif, les performances commencent à se dégrader, même si le nombre de fausses détections diminue encore, il est donc intéressant de garder cette classification.
- Avec la classification utilisant les paramètres pseudo-Haar (tableau 3.8), on observe le même effet qu'avec les motifs binaires locaux pour une étape d'apprentissage actif, on a une augmentation du F₁ Score grâce à l'augmentation de la précision, et donc réduit le nombre de MNFP.

Dans les deux types de classification utilisés, on note une augmentation du temps de réaction. Cela est dû au fait que le classificateur devient de plus en plus sélectif, et donc va mettre un peu plus de temps pour trouver le polype après sa première apparition à l'image. Cependant, on reste de l'ordre de la seconde, ce qui est tout à fait acceptable.

3.2.3 Cohérence temporelle

On va maintenant s'intéresser ici à une des autres optimisations mise en place afin d'améliorer les performances de la méthode, et qui s'adapte uniquement aux vidéos : il s'agit de la cohérence temporelle.

Contrairement aux images fixes disponibles dans la base CVC-ColonDB [BSV12] qui ne se suivent pas forcément dans la prise de vue, la vidéo possède l'avantage d'avoir des images qui se suivent avec une variation minimale de l'information entre deux images successives. Il peut donc être intéressant ici d'utiliser cette faible variation entre la succession d'images afin d'améliorer nos performances.

On va donc mettre au point une cohérence temporelle basée sur la mise en mémoire des précédentes détections dans la vidéo afin de réduire le nombre de fausses détections. Cette cohérence temporelle vient donc se greffer à la méthode présentée précédemment qui, elle, ne change pas (qui utilise la fusion spatiale et si désiré, l'apprentissage actif). Le fonctionnement du nouvel algorithme est décrit dans le tableau 3.9.

-
1. Une étape de prétraitement de l'image pour préparer la détection
 2. Une étape de détection cascade qui permet de trouver les régions d'intérêt sur l'image contenant de possibles polypes (avec ou sans apprentissage actif)
 3. Une étape de fusion spatiale des zones détectées afin de limiter le nombre de fausses détections et ainsi réduire le nombre de régions d'intérêt affichées à l'utilisateur
 4. Une étape de fusion temporelle des zones détectées pour l'image en cours avec les zones détectées des deux précédentes images
 5. Une étape d'affichage permettant de retourner le résultats de la détection à l'utilisateur
-

TABLE 3.9 – Algorithme de détection des polypes dans les vidéos

Fonctionnement

On va alors s'intéresser à la mise au point de la cohérence temporelle. Elle se base sur l'utilisation des résultats de la détection sur 3 images successives. Pour obtenir le résultat de la détection dans l'image i , on va superposer sur l'image i les résultats des détections qui ont eu lieu dans les images i , $i - 1$ et $i - 2$. De cette superposition, on va fusionner ces différentes zones (de la même manière que la fusion spatiale) afin de réduire les fausses détections. Et c'est ce résultat qui est alors renvoyé à l'utilisateur. En voici un exemple sur la figure 3.4 qui illustre son fonctionnement : lors de la détection sur l'image i , on regarde si des détections ont eu lieu au même endroit ou a proximité sur les 2 précédentes images, si oui, dans ce cas la zone est considérée comme une bonne détection et est affichée à l'utilisateur. Sinon, la zone n'est pas affichée (mais est gardée en mémoire pour les

prochaines détections au cas où).

L'utilisation de 3 images se justifie par le fait d'utiliser un vote à la majorité pour fusionner les zones détectées, donc on a besoin d'un nombre impair d'images. On aurait pu utiliser plus d'images, mais cela risque d'augmenter le temps de réaction de la méthode, et comme on désire un temps de réaction des plus faibles, on décide alors de n'utiliser que 3 images, ce qui semble être un bon compromis.

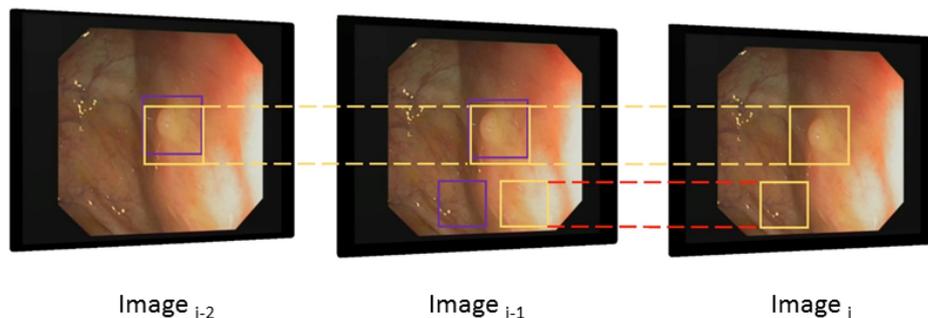


FIGURE 3.4 – Illustration de la cohérence temporelle

Effet de la cohérence temporelle sur la détection

La cohérence temporelle va avoir pour but de lisser les résultats de détection, afin d'apporter une meilleure stabilité dans le résultat affiché à l'écran. Cela peut se voir indirectement sur les profils résultants des détections dans les vidéos (figure A.1 à A.18). La figure 3.5 (extraite de la figure A.5) illustre le lissage observé. Sur la figure 3.5.b, on peut observer le profil de la détection sans cohérence temporelle (1 si le polype est bien trouvé, 0 sinon) et sur la figure 3.5.c, on peut observer le profil de la détection avec la cohérence temporelle. On voit donc que la cohérence temporelle permet de "lisser" les résultats bruts, elle supprime les oscillations que l'on peut observer dû au changement de la présence du polype dans la vidéo. Cela permet donc d'avoir un meilleur rendu visuel pour le médecin en n'ayant pas des boîtes qui enchainent apparitions et disparitions.

Performances sans apprentissage actif

Maintenant que la cohérence temporelle est mise en place, on peut donc s'intéresser à son influence sur les performances sans utiliser l'apprentissage actif pour le moment, c'est-à-dire qu'on cherche à voir simplement quels peuvent être les bénéfices d'utiliser une telle optimisation pour améliorer les résultats de base de notre méthode avec déjà un minimum de fausses détections. On va donc comparer les résultats de la détection avec et sans cohérence temporelle sur les deux types de paramètres utilisés pour l'apprentissage (les motifs locaux binaires (LBP) et les paramètres pseudo-Haar). Comme précédemment, ces tests

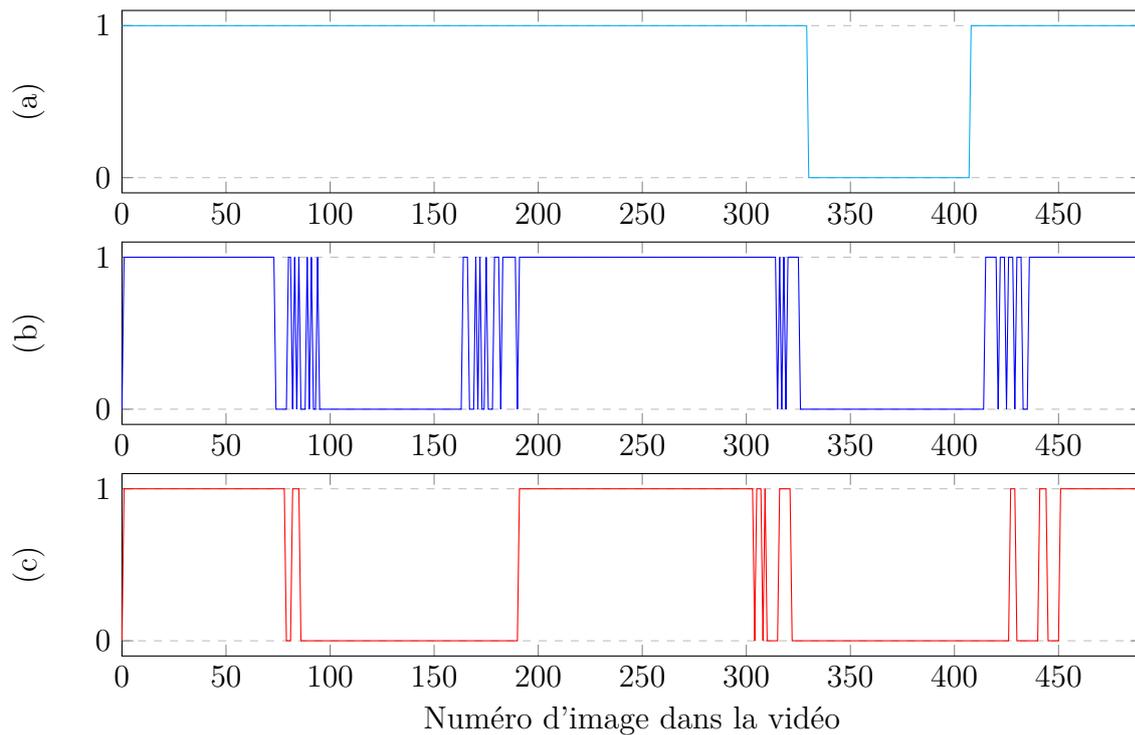


FIGURE 3.5 – Effet de la cohérence temporelle pour la vidéo 5 (490 images), (a) représente la vérité terrain, (b) représente les résultats de détection sans cohérence temporelle, (c) représente les résultats de détection avec cohérence temporelle

seront menés sur l'ensemble des 18 vidéos à notre disposition et évalués avec les métriques associées. Les résultats de ces différents tests sont consignés dans les tableaux 3.10 et 3.11.

Cohérence	PDR	MPT	MNFP	Précision	Rappel	F ₁ Score	RT
Sans	100%	140ms	3.5	12.42%	54.65%	20.24%	7.2 (0.3s)
Avec	100%	140ms	1.9	16.25%	41.25%	23.31%	35.0 (1.4s)

TABLE 3.10 – Influence de la cohérence temporelle sur les résultats utilisant la classification basée sur les LBP sans apprentissage actif

Cohérence	PDR	MPT	MNFP	Précision	Rappel	F ₁ Score	RT
Sans	100%	26ms	1.4	23.29%	48.82%	31.10%	17.5 (0.7s)
Avec	100%	26ms	0.9	27.02%	39.61%	32.12%	38.3 (1.5s)

TABLE 3.11 – Influence de la cohérence temporelle sur les résultats utilisant la classification basées les paramètres pseudo-Haar sans apprentissage actif

A partir des résultats disponibles dans les tableaux 3.10 et 3.11, on s'aperçoit toujours qu'on est capable d'attirer l'attention du médecin sur le polype (PDR de 100%), ce qui

veut dire que la cohérence temporelle n'impacte pas le but premier de l'outil : alerter sur la présence du polype et sa position. Côté performance de détection, la cohérence temporelle permet de réduire le nombre moyen de fausses détections par image, qui pour les LBP descend en dessous de 2, et pour les paramètres pseudo-Haar descend en dessous de 1. Cela va donc impacter le F_1 Score qui va augmenter, ce qui traduit une amélioration globale des performances de détection. Concernant le temps de calcul, on constate que dans dans les deux cas, le temps de calcul n'augmente pas de façon sensible pour être visible (les variations des temps de calcul sont inférieurs à la milliseconde). Par contre, la cohérence temporelle va augmenter le temps de réaction de notre méthode, ce qui est tout a fait logique car il faut plus d'images pour valider le résultat de la détection. La figure 3.6 permet de se rendre compte de l'influence de la cohérence temporelle sur les régions d'intérêts retournées au médecin.

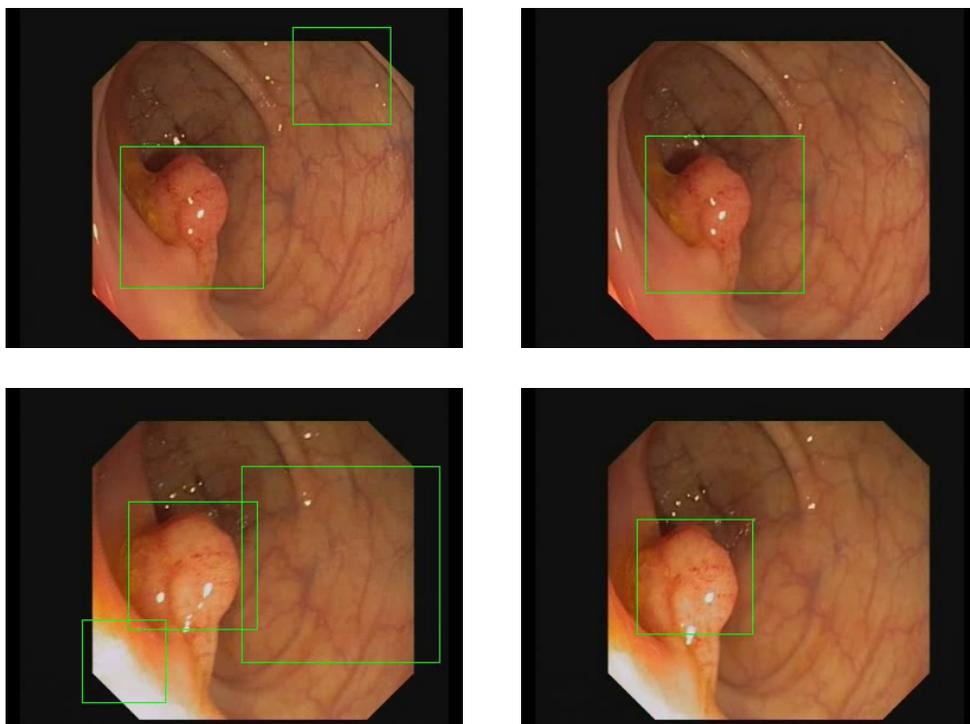


FIGURE 3.6 – Images illustrant la cohérence temporelle, à gauche des résultats de détection sans cohérence temporelle, à droite des résultats avec cohérence temporelle

Temps de réaction

Une chose que l'on a pu constater sur la quasi totalité des vidéos, c'est que l'ajout de la cohérence temporelle a eu un effet sur le temps de réaction de notre outil. En effet, comme on peut le constater sur la tableau 3.12, la moyenne des temps de réaction des 18 vidéos passe du simple au double avec la cohérence. Mais comme il s'agit d'une moyenne, et que l'augmentation de la valeur de cette métrique n'est pas négligeable, on peut alors se demander si ce n'est pas une ou deux vidéos qui mettent à mal le temps de réaction.

Dans le tableau 3.13, on peut voir les temps de réaction pour chacune des vidéos sans la cohérence temporelle et dans le tableau 3.14, les temps de réaction avec la cohérence temporelle.

Méthode	Temps de réaction (images)
Sans cohérence	17.5 (0.7s)
Avec cohérence	38.3 (1.5s)

TABLE 3.12 – Influence de la cohérence temporelle sur le temps de réaction avec la classification utilisant les paramètres pseudo-Haar sans apprentissage actif

Numéro de vidéo	1	2	3	4	5	6	7	8	9
Temps de réaction (secondes)	0.6	3.44	0.08	0.16	0.08	0.08	0.08	3.44	0.32
Numéro de vidéo	10	11	12	13	14	15	16	17	18
Temps de réaction (secondes)	0.24	0.20	0.08	0.08	0.08	3.4	0.12	0.12	0

TABLE 3.13 – Temps de réaction pour chaque vidéo sans cohérence temporelle

Numéro de vidéo	1	2	3	4	5	6	7	8	9
Temps de réaction (secondes)	1.28	3.4	0.08	1.2	0.08	0.08	6.28	8.52	1.68
Numéro de vidéo	10	11	12	13	14	15	16	17	18
Temps de réaction (secondes)	1.12	0.28	0.08	0.16	0.12	3.48	0.56	0.56	0

TABLE 3.14 – Temps de réaction pour chaque vidéo avec cohérence temporelle

On constate alors que le temps de réaction augmente de manière importante sur deux vidéos en particulier. Si on retire ces deux vidéos de l'équation, on obtient alors un temps de réaction de 19.94 images, soit environ 0.8 secondes, plutôt proche de la valeur obtenue sans cohérence temporelle (qui est de 14.19 images, soit environ 0.6 secondes). On observe seulement une différence de 0.2 secondes, contrairement aux 0.8 secondes observées au début, c'est donc bien ces deux vidéos qui détériorent les performances du temps de réaction. On pourrait alors conclure que la cohérence temporelle ne détériore que faiblement la détection des polypes si on fait abstraction des vidéos 7 et 8. Mais avant de conclure cela, il serait intéressant de regarder en détail les deux vidéos en question afin d'identifier la (ou les) cause(s) qui augmente le temps de réaction dans ces deux vidéos. On va donc traiter ces deux cas séparément par la suite.

Cas de la vidéo 7

On s'intéresse au cas de la vidéo 7. On a pu constater que le temps de réaction sans cohérence temporelle est de 2 images, mais passe subitement à 157 avec la cohérence

temporelle. Quand on regarde le profil de cette vidéo (figure 3.8, on a 1 si le polype est présent dans l'image, 0 sinon), on s'aperçoit que le polype fait une série d'apparition et de disparition, et c'est ce qui semble être la cause de l'augmentation du temps de réaction. En effet, la cohérence temporelle a pour but de moyenniser les détections, alors si peu de détections sont faites, alors le résultat de la cohérence temporelle donnera l'impression que l'algorithme n'a rien détecté. Ce qui explique l'augmentation du temps de réaction dans cette vidéo puisque ce type de structure peut effectivement apparaître.

Mais ce n'est pas le profil de la vidéo qui est à remettre en cause, mais la qualité de l'image, si on se rappelle le tableau 3.1 qui notait la difficulté à détecter le polype dans les vidéos, la vidéo 7 avait obtenu un score élevé, synonyme d'une forte difficulté à trouver le polype. Cette note élevée est due tout simplement à la qualité de la vidéo. Quand on regarde la vidéo, on peut s'apercevoir de la présence de liquide qui passe devant l'endoscope (figure 3.7), et qui introduit un certain flou, ce qui peut alors mettre en échec la méthode, et cela à plusieurs reprises. Comme la cohérence temporelle requiert 3 images pour fonctionner, si on n'est plus capable d'effectuer des détections sur certaines images, alors on va clairement augmenter notre temps de réaction. Et c'est donc la raison pour laquelle le temps de réaction augmente dans cette vidéo.

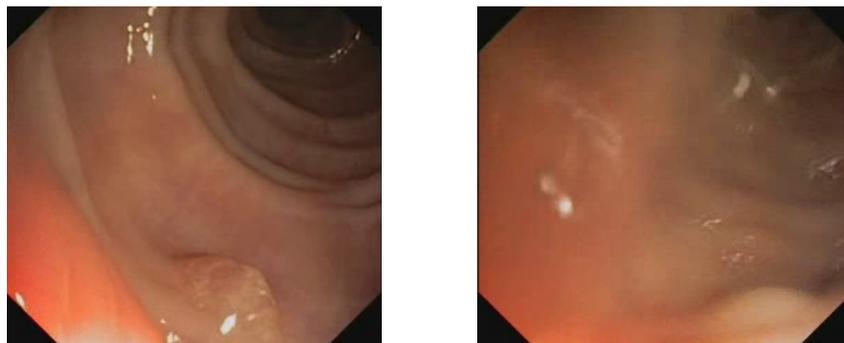


FIGURE 3.7 – Images extraites de la vidéo 7 (image nette à gauche et image floutée à cause d'un liquide à droite)

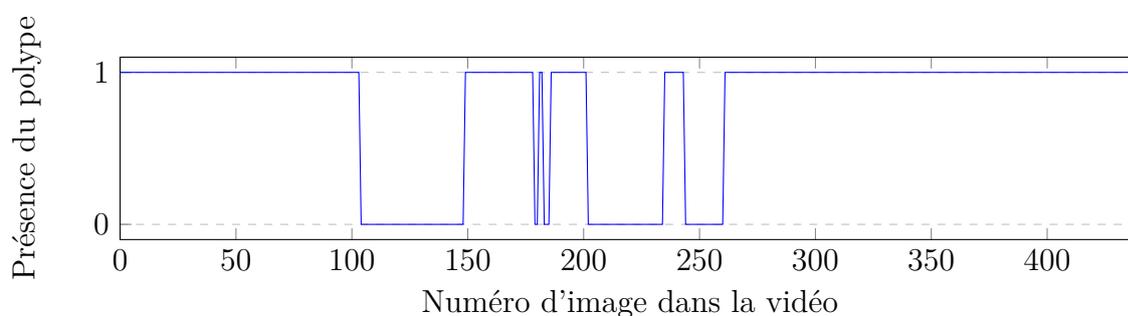


FIGURE 3.8 – Présence du polype dans la vidéo 7 (440 images)

Cas de la vidéo 8

Dans le cas de la vidéo 8, le temps de réaction sans cohérence temporelle est de 86 images, mais passe subitement à 213. On constate ici que de base, l'algorithme met du temps à trouver le polype (car il est peut être difficile à trouver). Comme précédemment, on peut regarder le profil de la vidéo (figure 3.10) pour obtenir quelques indices, mais il n'y a pas de successions d'apparitions et de disparitions du polype dans la vidéo. Cependant, cela nous indique donc qu'il faut s'intéresser à l'image en elle même. Et comme on peut le constater en regardant la figure 3.9, le polype est difficilement différentiable de la paroi, d'où le score qu'avait obtenu cette vidéo dans le tableau 3.1. Cela explique donc d'une part pourquoi le temps de réaction était déjà élevé au départ, mais qu'en rajoutant la cohérence temporelle, on va donc indéniablement augmenter le temps de réaction.



FIGURE 3.9 – Images extraites de la vidéo 8 (image éloignée du polype à gauche et image proche du polype à droite)

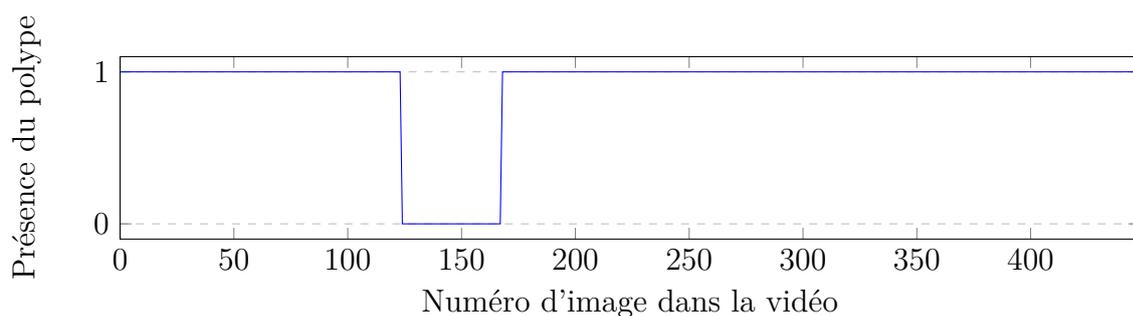


FIGURE 3.10 – Présence du polype dans la vidéo 8 (448 images)

Conclusion sur la cohérence temporelle

La cohérence temporelle permet clairement de réduire le nombre moyen de fausses détections par image et permet donc de réduire les perturbations renvoyées au médecin.

Cependant, la cohérence temporelle rend l'outil moins réactif (au sens de sa rapidité à détecter un polype après sa première apparition dans à l'image, autrement dit le temps de réaction). De plus, le rappel diminue lorsque l'on applique la cohérence temporelle. Mais on va pu voir précédemment que l'apprentissage actif permettait justement d'augmenter le rappel de la méthode. On va donc maintenant s'intéresser à l'association de la cohérence temporelle avec l'apprentissage actif.

De plus, la cohérence va augmenter de façon significative le temps de réaction dans certains cas : lorsque le polype va enchaîner disparitions et apparitions, lorsque la qualité de la vidéo n'est pas bonne (qui peut être lié à la qualité de la préparation du côlon) ou lorsque le polype est difficilement différentiable de la paroi du côlon. Cependant, ces cas restent assez rares, et malgré cela, la cohérence temporelle permet quand même d'améliorer les performances globales de la méthode (qui s'aperçoit par l'augmentation du F_1 Score).

3.2.4 Combinaison de l'apprentissage actif et de la cohérence temporelle

Tout comme pour le test précédent sur l'influence de la cohérence temporelle sur les performances de détection, on va maintenant effectuer un test pour analyser l'influence de l'apprentissage actif sur la méthode de détection dans les vidéos équipées de la cohérence temporelle dans le but d'améliorer les performances de la méthode. Ce test va donc aussi s'effectuer sur la base de données vidéo [ABSM⁺17b] avec les mêmes métriques. Les résultats de ce test sont disponibles dans le tableau 3.15

Méthode	PDR	MPT	MNFP	Précision	Rappel	F_1 Score	RT
LBP $N0$	100%	140ms	1.9	16.25%	41.25%	23.31%	35.0 (1.4s)
LBP $N1$	100%	160ms	1.1	27.11%	46.02%	34.12%	43.7 (1.7s)
LBP $N2$	100%	162ms	0.7	29.88%	34.96%	32.22%	45.9 (1.8s)
Haar $N0$	100%	26ms	0.9	27.02%	39.61%	32.12%	38.3 (1.5s)
Haar $N1$	100%	23ms	0.6	39.14%	42.56%	40.78%	27.3 (1.1s)

TABLE 3.15 – Résultats de la détection dans les vidéos avec la cohérence temporelle et l'apprentissage actif

D'après le tableau 3.15, on constate que l'apprentissage actif associé à la cohérence temporelle permet d'obtenir les meilleures performances. Concernant la classification avec les LBP, on s'aperçoit que l'association de l'apprentissage actif avec la cohérence temporelle influe sur les performances de la classification (la précision augmente mais le rappel diminue) et le temps de calcul moyen par image augmente. Cela s'explique par le fait que l'apprentissage actif augmente la précision (comme vu précédemment) mais fait chuter le rappel, et il en est de même pour la cohérence temporelle. Donc à partir du moment où les deux sont associés, ces deux paramètres vont évoluer comme prévu, mais le F_1 va

augmenter. Mais contrairement à la classification LBP, la classification utilisant les paramètres pseudo-Haar avec une étape d'apprentissage actif réunit les meilleurs résultats : à elle seule, elle possède un temps de calcul très faible par image, tout en ayant un F_1 Score assez élevé. C'est donc cette classification là qu'il faudrait favoriser pour effectuer la détection des polypes lors des examens de coloscopie.

Variations

Précédemment, on a pu voir le fonctionnement de la cohérence temporelle, où il s'agissait d'utiliser les résultats des deux précédentes détections pour valider ou non les détections dans l'image en cours. En se basant sur la figure 3.4, on utilisait les résultats de détection des images i , $i - 1$ et $i - 2$ pour valider les résultats de l'image i . Mais on peut aussi envisager quelques variations de cette méthode de cohérence temporelle : utiliser le résultats de ces 3 détections pour valider la détection de l'image $i - 1$ ou de l'image $i - 2$. Des tests ont été alors réalisés afin de regarder l'influence de ces variations. Ces tests ont été menés avec la classification utilisant les paramètres pseudo-Haar avec une étape d'apprentissage actif. Les résultats sont disponibles sur la figure 3.11.

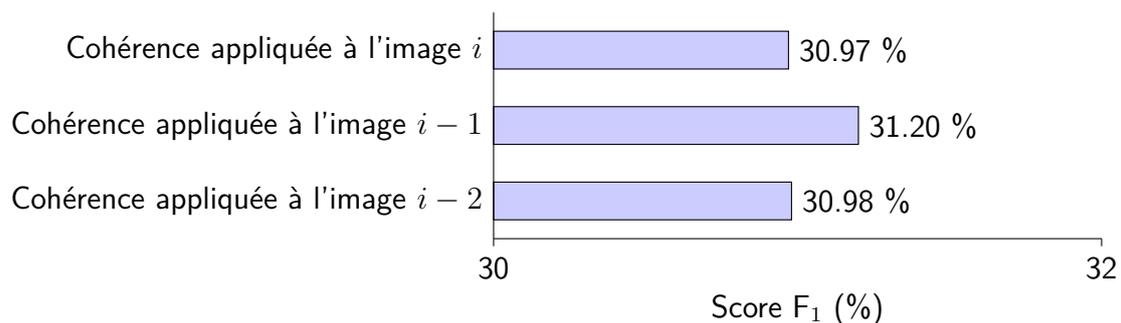


FIGURE 3.11 – Score F_1 pour différentes combinaisons de cohérence temporelle

On constate donc que l'utilisation des résultats de la détection dans les images autour de l'image qu'on traite actuellement (c'est-à-dire utiliser les résultats des images i et $i - 2$ pour valider les résultats de l'image $i - 1$) permet d'améliorer sensiblement les performances de détection de la méthode. Cependant, l'utilisation de ces deux variations va rajouter un certain retard dans la chaîne de traitement : dans un cas, il faut attendre l'image d'après et les résultats de la détection, donc cela va rajouter un retard d'une image (et dans l'autre, cela rajoute un retard de deux images). Ainsi, au rythme de 25 images par secondes (rythme auquel arrive les images dans le flot vidéo), on va rajouter un retard entre 40 et 80 millisecondes. Cela va donc jouer évidemment sur le temps de réaction et l'augmenter, mais cette augmentation reste assez négligeable devant le temps de réaction de l'outil.

3.2.5 Combinaison des différents types de descripteurs d'apprentissage

Côté apprentissage, on a vu que les deux types de paramètres utilisés (motifs binaires locaux et paramètres pseudo-Haar) permettaient d'obtenir des performances différentes mais plutôt acceptables (si on fait abstraction du temps de calcul moyen, qui est dépendant de la puissance de la machine utilisée). On peut alors se poser la question de savoir si l'association de ces deux types d'apprentissage ne permettrait pas d'améliorer les performances globales de la méthode. On s'est donc intéressé à cette combinaison en associant la classification *LBP $N2$* (classificateur utilisant les motifs binaires locaux avec une étape d'apprentissage actif) et *Haar $N1$* (classificateur utilisant les paramètres pseudo-Haar avec une étape d'apprentissage actif). Le choix de ces classificateurs n'est pas dû au hasard, pour chaque type de paramètres utilisés lors de l'apprentissage, ce sont eux qui permettent d'obtenir les F_1 Score les plus élevés. On mène donc les tests sur l'ensemble des 18 vidéos afin d'en calculer les performances. Les résultats de ce test sont disponibles dans le tableau 3.16.

Méthode	PDR	MPT	MNFP	Précision	Rappel	F_1 Score	RT
<i>LBP$N2$</i>	100%	162ms	0.7	29.88%	34.96%	32.22%	45.9 (1.8s)
<i>Haar$N1$</i>	100%	21ms	0.6	39.14%	42.56%	40.78%	27.3 (1.1s)
Combinaison	100%	185ms	1.1	30.39%	52.40%	38.47%	15.0 (0.6s)

TABLE 3.16 – Comparaison entre classifications classiques et combinaison de classifications (*LBP $N2$* & *Haar $N1$*)

On voit donc sur le tableau 3.16 que la combinaison des deux types de classification permet toujours d'alerter le médecin sur la présence du polype dans l'ensemble des 18 vidéos, mais qu'en plus, le Rappel augmente, ce qui signifie qu'on est capable de détecter plus de polype qu'auparavant mais de manière moins précise. On peut aussi voir que le temps de réaction diminue fortement, ce qui induit une meilleure rapidité dans la détection après la première apparition du polype à l'image. En somme, la combinaison des deux classifications est une piste à explorer car elle permet d'obtenir de meilleurs résultats. Cependant, le temps de calcul reste trop élevé pour faire du temps réel (on hérite du temps de calcul de la classification utilisant les motifs binaires locaux), et donc cela ne nous convient pas pour le moment.

3.2.6 Prétraitement de l'image

On a pu voir plusieurs pistes qui ont permis d'améliorer les performances de la méthode, ces méthodes se sont principalement concentrées sur l'amélioration de l'apprentissage actif et la réduction des différentes régions d'intérêt détectées en appliquant une fusion spatiale et une cohérence temporelle. Cependant, une autre piste peut être explorée pour améliorer les performances. Il s'agit d'appliquer un prétraitement à l'image avant d'effectuer la moindre détection. On va donc s'intéresser dans cette partie aux différentes

possibilités de prétraitement de l'image et leurs influences sur les performances de la méthode.

Influence du retrait d'artefacts

Une première piste de prétraitement vient lorsqu'on regarde les images issues de vidéo-coloscopies. On s'aperçoit qu'une bonne partie de l'image ne contient pas d'information permettant la détection des polypes. Il s'agit de tout le contour de l'image en noir (voir figure 3.12). Il peut être alors intéressant de comparer les performances de détection entre des images normales et des images détournées (voir figure 3.12). La comparaison est alors effectuée pour un seul type de classification pour voir la différence (la classification utilisant les paramètres pseudo-Haar est utilisée avec une étape d'apprentissage actif). Les résultats sont disponibles dans le tableau 3.17.

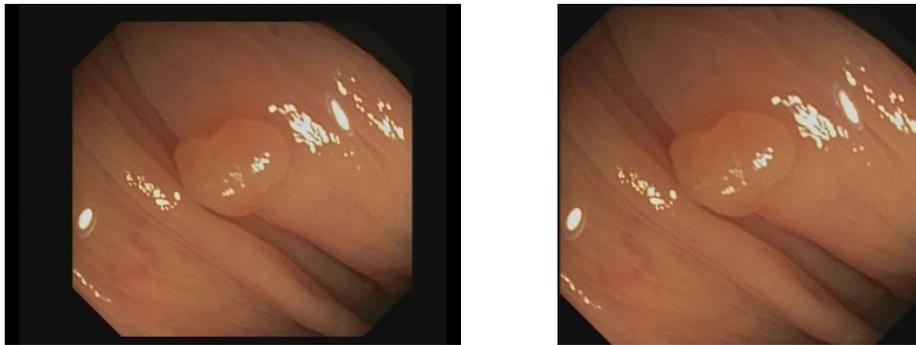


FIGURE 3.12 – Image avec les bordures (à gauche) et image sans (à droite)

Images	TP	FP	TN	FN	Précision	Rappel	F ₁ Score	PDR
Originales	4267	6634	702	5758	39.14%	42.56%	40.78%	100%
Sans contours	4496	5338	880	5529	45.72%	44.85%	45.28%	100%

TABLE 3.17 – Comparaisons des performances de détection entre images normales et images détournées

On constate alors qu'en retirant les bordures noires, on améliore globalement les performances de la détection. Certes, le nombre de bonnes détections diminue, mais le nombre de fausses détections diminue drastiquement, ce qui est un avantage car cela retourne moins d'informations erronées au médecin, ce qui peut améliorer les performances de l'examen.

Cela peut s'expliquer de la manière suivante : le retrait des bordures de l'image retire de l'information, mais pas une information si importante que ça. En retirant donc cette information, certes on perd en performance, mais on permet à l'algorithme de mieux se concentrer sur les zones à détecter et permet alors de diminuer le nombre de fausses détections.

Images	Temps de calcul
Originales	23 ms
Sans contours	16 ms

TABLE 3.18 – Comparaisons des temps de traitement entre images normales et images détournées

Mais de plus, en retirant ces bordures, on accélère le temps de traitement (figure 3.18) car l'image à traiter diminue en résolution, sans perdre d'information utile. Donc détourner les images de leurs contours noirs présente de nombreux avantages, et peut être une des solutions simples afin d'améliorer les performances de notre outil.

Influence du retrait de la lumière spéculaire

La seconde piste s'obtient aussi en regardant les images issues de coloscopies. On peut apercevoir sur plusieurs images la réflexion de la lumière de l'endoscope sur la paroi du côlon : il s'agit de la lumière spéculaire (voir la figure 3.13). On a pu voir que sur plusieurs images renvoyées par la méthode (utilisée pour la détection des polypes) que des régions d'intérêt se trouvent être des zones contenant de la lumière spéculaire et pas des polypes. On peut donc essayer de supprimer cette lumière dans l'optique d'améliorer les résultats de détection. Pour ce faire, on utilise une méthode mise au point par J. Bernal [BSV13] qui va permettre de détecter les zones contenant la lumière spéculaire et repeindre ces zones avec des pixels voisins afin de réduire leurs impacts sur la détection (voir la figure 3.13). La comparaison va alors être effectuée avec un seul type de classification pour voir la différence (la classification utilisant les paramètres pseudo-Haar est utilisée avec une étape d'apprentissage actif). Les résultats sont disponibles dans le tableau 3.19.

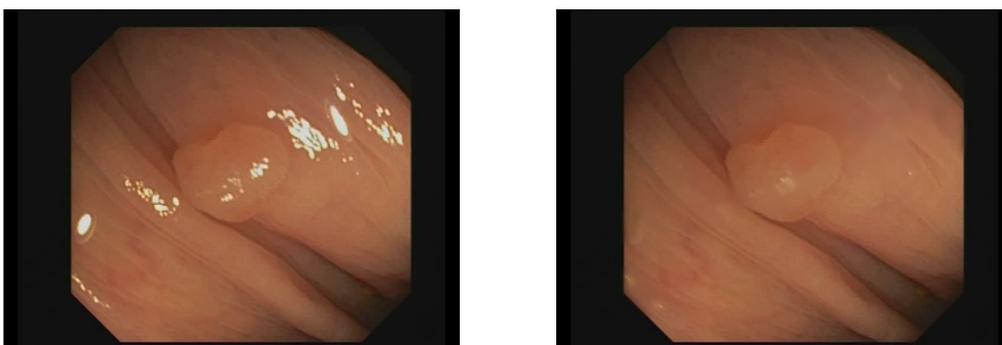


FIGURE 3.13 – Image avec lumière spéculaire (à gauche) et image inpaintée (à droite)

Selon le tableau 3.19, on constate alors une très légère détérioration des performances, en effet la Précision augmente mais le Rappel diminue, ce qui impacte évidemment le F_1 Score, dont la variation reste minime. Dans la réalité, on se rend compte qu'on diminue le nombre de bonnes détections mais on diminue surtout le nombre de fausses détections,

Images	TP	FP	TN	FN	Précision	Rappel	F ₁ Score	PDR
Originales	4267	6634	702	5758	39.14%	42.56%	40.78%	100%
Sans lumière spéculaire	4342	6934	669	5683	38.51%	43.31%	40.77%	100%

TABLE 3.19 – Comparaisons des performances de détection entre images normales et images sans lumière spéculaire

Images	Temps de calcul
Originales	23 ms
Sans contours	20 ms

TABLE 3.20 – Comparaisons des temps de traitement entre images normales et images détournées

cela traduit donc toujours la capacité à détecter les polypes tout en fournissant un faible nombre de fausses alertes, ce qui perturbe moins le médecin. On est alors capable de mieux attirer l'attention du médecin sur la présence du polype, sachant qu'on conserve toujours un PDR de 100%, indiquant qu'on alerte le médecin sur la présence de tous les polypes. L'obtention de tels résultats s'explique par le fait que de base, la méthode détectait certaines zones de lumière spéculaire comme étant des polypes, donc en retirant cette lumière spéculaire, on retire les fausses détections qui y sont liées, et on perturbe moins la détection. Il s'en suit alors une légère modification des performances, comme observée. De plus, d'après le tableau 3.20, on constate que le temps de calcul diminue, ce qui montre un second intérêt d'utiliser une telle optimisation.

Influence de la combinaison des deux phases de prétraitement

Précédemment, on a constaté que retirer les bordures permettait de diminuer de façon conséquente les fausses détections mais diminuait aussi les bonnes détections. En retirant la lumière spéculaire, on s'apercevait qu'on augmentait le nombre de bonnes détections mais aussi, on a eu quelques fausses détections. Naturellement, on peut se poser la question de combiner les deux. Pour cela, on utilise donc des images recadrées et retouchées (figure 3.14). On pourrait donc s'attendre à une chute de fausses détections, mais aussi une baisse des bonnes détections, mais moins conséquente qu'avec uniquement le retrait des bordures. Les essais ont été menés avec la classification utilisant les paramètres pseudo-Haar (avec une étape d'apprentissage actif). Les résultats sont disponibles dans le tableau 3.21.

On constate alors qu'on améliore bien les performances globales de la méthode (F₁ Score). On peut donc envisager d'intégrer une étape de prétraitement dans la chaîne de détection afin de retirer les bordures et la lumière spéculaire afin de limiter grandement le nombre de fausses détections, et donc de moins perturber le médecin durant son examen.

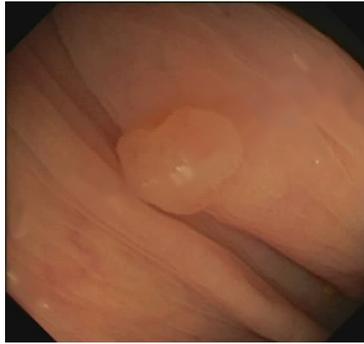


FIGURE 3.14 – Image sans bordure et sans lumière spéculaire

Images	TP	FP	TN	FN	Précision	Rappel	F ₁ Score	PDR
Originales	4267	6634	702	5758	39.14%	42.56%	40.78%	100%
Sans contours	4496	5338	880	5529	45.72%	44.85%	45.28%	100%
Sans lumière spéculaire	4342	6934	669	5683	38.51%	43.31%	40.77%	100%
Sans contours et sans lumière spéculaire	4630	5438	915	5395	45.98%	46.18%	46.09%	100%

TABLE 3.21 – Comparaisons des performances de détection entre images normales et images détournées

3.3 Conclusion sur la détection dans les vidéos

Dans cette partie, on a pu s'intéresser à la détection et la localisation des polypes dans les vidéos de coloscopies, vidéos qui se rapprochent de la réalité d'un examen, et qui vont permettre de prouver l'utilité d'un tel outil d'aide à la détection. On a donc pu voir dans cette partie quelle était la méthode retenue pour effectuer la détection des polypes dans les vidéos issues de coloscopies, méthode basée sur l'apprentissage machine pur utilisant soit les paramètres à motif binaire local ou les paramètres pseudo-Haar, et ainsi que ses limitations en terme de performances sur les vidéos grâce à l'introduction de nouvelles métriques dédiées et cliniquement utilisables (c'est-à-dire facilement interprétables par le médecin). Pour palier à ces limitations, plusieurs améliorations ont été développées pour améliorer les performances de notre outil.

La première amélioration est tout simplement la mise en place d'une fusion spatiale des régions d'intérêt détectées afin de réduire le nombre de régions d'intérêt inutiles et de réduire le nombre de fausses détection. On a ainsi vu que, quelque soit les paramètres d'apprentissage utilisés, les performances étaient améliorées. De plus, cette amélioration permettant un gain non négligeable en performances, il a été intégré et utilisé avec toutes les autres améliorations.

La seconde amélioration, qui avait été présentée dans la partie sur la détection dans les images fixes, est l'utilisation de l'apprentissage actif pour la détection dans les vidéos.

Encore une fois, l'apprentissage actif a montré ses avantages et a amélioré nettement les performances de notre méthode.

La troisième amélioration mise au point s'applique très bien au cas de la détection dans la vidéo. Il s'agit de la cohérence temporelle, cette amélioration permet d'obtenir de meilleurs résultats sur une image de la vidéo en se basant sur les détections obtenues dans les images précédentes, il s'agit donc d'une sorte de mémoire afin de valider un futur résultat. On a alors pu constater que la cohérence temporelle permettait d'améliorer les performances de notre outil.

Par la suite, on a combiné la seconde amélioration (l'apprentissage actif) avec la troisième (la cohérence temporelle). On a ainsi pu voir que cette combinaison permettait bel et bien d'améliorer les performances de notre outil et donc que ces deux améliorations ne sont pas néfastes l'une pour l'autre.

Cette association des trois améliorations montre de très bonnes performances (en particulier pour la classification utilisant les paramètres pseudo-Haar) et promet de bonnes choses pour un essai clinique. Mieux, l'ajout ces différentes améliorations n'impacte pas le temps de calcul. Il est tout à fait négligeable car les modifications des temps de calcul engendrées sont inférieures à la milliseconde.

On peut alors remarquer que toutes ces améliorations s'intéressent soit à l'étape de détection (l'apprentissage actif) ou à l'étape de post-traitement de la détection (fusion spatiale et cohérence temporelle). On s'est alors posé la question de savoir si il était possible de mettre au point des améliorations qui allaient se placer avant la détection, c'est-à-dire effectuer un pré-traitement sur l'image. On a donc relevé deux nouvelles améliorations possibles qui pourraient permettre d'avoir de meilleurs résultats.

La quatrième amélioration est le retrait de l'information inutile de l'image pour concentrer les recherches dans ce qui nous intéresse. On a alors retiré les bordures de chaque image car elles ne contiennent logiquement pas de polype et donc ne sont pas très utiles. On observe donc une amélioration des performances concernant la détection ainsi qu'une réduction du temps de calcul pour traiter une image. Cela veut dire qu'on peut consacrer un peu plus de temps à d'autres étapes de pré-traitement de l'image.

La cinquième amélioration est le retrait de la lumière spéculaire. Au cours des différents tests, la lumière spéculaire générait de nombreuses fausses détections, la question de la retirer fut judicieuse. En effet, le remplacement de cette lumière par les zones voisines permet d'obtenir des performances intéressante. Cependant, cette étape va rajouter nécessairement du temps de calcul, qui est au maximum de 19 millisecondes (rajouté aux 23 millisecondes, la contrainte de temps réel n'est plus atteinte).

Cependant, en combinant ces deux améliorations avec toutes les autres présentées, on améliore nettement les performances de notre outil, ce qui permet d'obtenir une méthode beaucoup plus robuste pour la détection des polypes les vidéos. De plus, avec toutes ces améliorations, l'outil est toujours capable de fonctionner presque en temps réel (en

utilisant uniquement les paramètres pseudo-Haar pour la détection). Les vidéos se rapprochant de cas réels, et avec notre méthode éprouvée, on peut donc envisager d'effectuer des essais cliniques en conditions réelles afin de se comparer au médecin et encore mieux prouver l'intérêt de notre outil de détection des polypes en temps réel.

Le tableau 3.22 résume les gains en performance de la méthode basée sur la classification utilisant les paramètres pseudo-Haar, classification, qui à l'heure actuelle, est la seule qui est capable de fonctionner en temps réel dans nos conditions expérimentales.

La tableau 3.22 montre clairement que les améliorations apportées à l'outil permettent d'obtenir de meilleurs résultats qu'à l'origine (on a un gain allant jusqu'à 29.82 points)

Amélioration	Gain en performances (points)
Aucune	+0
Fusion spatiale (1)	+14.83
Apprentissage actif (2) + 1	+22.38
Cohérence temporelle (3) + 1	+15.85
1 + 2 + 3	+24.51
Retrait d'artefacts (4) + 1 + 2 + 3	+29.01
Inpainting (5) + 1 + 2 + 3	+24.50
1 + 2 + 3 + 4 + 5	+29.82

TABLE 3.22 – Résumé des gains en performances suivant les optimisations (Gains relatifs basés sur le F_1 Score et sur la classification utilisant les paramètres pseudo-Haar)

On a donc mis au point une méthode permettant de localiser les polypes dans les vidéos de coloscopies, vidéos qui se rapprochent des cas réels. La méthode obtenue possède de très bonnes performances de détections (F_1 Score de 40.78% avec pour la classification Haar $N1$ avec cohérence temporelle) et alerte le médecin sur la présence de polypes dans chacune des vidéos (PDR de 100%). La méthode fonctionne aussi en temps réel (23 millisecondes pour traiter une image). On a donc une méthode qui possède de bonnes performances, fonctionne en temps réel, embarquable et qui est cliniquement utilisable. On a donc mis au point une méthode répondant à nos contraintes. Par rapport à l'état de l'art, il semble alors qu'on soit les seuls à proposer ce genre d'outil, qui sera bientôt testé en routine clinique.

3.4 Évaluation de la robustesse

On a vu précédemment que notre méthode de détection offrait de bonnes performances. Mais on a décidé de tester notre méthode sur une nouvelle base de données contenant à nouveau 18 vidéos [BHM⁺18] afin de voir si les performances restent constantes ou changent grandement. Contrairement à la première base de données [ABSM⁺17b] qui contenait 11 954 images pour 18 vidéos, la nouvelle base de données [BHM⁺18] contient

elle 29 657 images pour 18 vidéos. Parmi ces 29 657 images, 21 813 contiennent un polype, soit 73.55% des images. On va ainsi mesurer les performances de notre méthode sur ces nouvelles méthodes en utilisant une partie des métriques précédentes, qui sont la Précision, le Rappel, le F_1 Score, le temps de réaction (RT, en nombre d'images) et le taux de polypes détectés (PDR). Les résultats de ce test sont disponibles dans le tableau 3.23. Les tests menés se sont concentrés sur les classifications bénéficiant de la fusion spatiale, de la cohérence temporelle et de l'apprentissage actif (et ne bénéficiant d'aucune étape de prétraitement).

Méthode	Précision	Rappel	F_1 Score	RT	PDR
LBP $N0$	13.52%	39.11%	20.09%	42.4 (1.7s)	100%
LBP $N1$	32.22%	36.98%	34.44%	49.7 (2.0s)	100%
LBP $N2$	34.32%	37.14%	35.67%	40.7 (1.6s)	100%
Haar $N0$	24.49%	35.93%	29.13%	35.9 (1.4s)	94.4%
Haar $N1$	33.28%	37.26%	35.16%	25.8 (1.0s)	100%

TABLE 3.23 – Résultats de la détection sur une nouvelle base de données vidéos

D'après le tableau 3.23, on constate qu'à part pour la classification Haar $N0$, on est capable d'alerter l'oeil du médecin sur la position du polype dans l'image (PDR de 100%). Concernant le cas de la classification Haar $N0$, ceci est dû à la difficulté à trouver le polype dans cette vidéo (le PDR de 94.4% est dû au rapport 17 sur 18). Concernant les performances, on s'aperçoit, comme précédemment que c'est la classification Haar $N1$ qui permet d'obtenir les meilleurs résultats en termes de performances de détection et localisation, mais aussi en temps de réaction, on se rapproche de la seconde, ce qui est tout à fait acceptable. Cela confirme aussi que c'est la classification la plus judicieuse à utiliser pour détecter les polypes.

Mais plus encore, quand on compare les résultats du test sur la nouvelle base de données par rapport aux anciens résultats 3.24, on s'aperçoit que les performances gardent une certaine stabilité. En effet, on constate (en mettant à part la classification Haar $N0$), que les résultats sont proches entre les deux bases de données. Cela montre donc la robustesse de notre outil à traiter de nouveaux cas, et sa capacité à détecter et localiser les polypes en toutes circonstances si on utilise la classification Haar $N1$ qui permet de conjuguer bonnes performances et temps de calcul faible (qui est de 20 millisecondes sur la base [BHM⁺18] et de 23 millisecondes sur la base [ABSM⁺17b]) afin de fonctionner en temps réel.

3.5 Détection de la présence des polypes dans les vidéos

Jusque là, la méthode est utilisée pour localiser les polypes dans les vidéos afin d'attirer l'oeil du médecin sur le polype en lui-même. Sur l'aspect localisation, la méthode

Méthode	Précision	Rappel	F ₁ Score	RT	PDR
LBP <i>N0</i> [BHM ⁺ 18]	13.52%	39.11%	20.09%	42.4 (1.7s)	100%
LBP <i>N0</i> [ABSM ⁺ 17b]	16.25%	41.25%	23.31%	35.0 (1.4s)	100%
LBP <i>N1</i> [BHM ⁺ 18]	32.22%	36.98%	34.44%	49.7 (2.0s)	100%
LBP <i>N1</i> [ABSM ⁺ 17b]	27.11%	46.02%	34.12%	43.7 (1.7s)	100%
LBP <i>N2</i> [BHM ⁺ 18]	34.32%	37.14%	35.67%	40.7 (1.6s)	100%
LBP <i>N2</i> [ABSM ⁺ 17b]	29.88%	34.96%	32.22%	45.9 (1.8s)	100%
Haar <i>N0</i> [BHM ⁺ 18]	24.49%	35.93%	29.13%	35.9 (1.4s)	94.4%
Haar <i>N0</i> [ABSM ⁺ 17b]	27.02%	39.61%	32.12%	38.3 (1.5s)	100%
Haar <i>N1</i> [BHM ⁺ 18]	33.28%	37.26%	35.16%	25.8 (1.0s)	100%
Haar <i>N1</i> [ABSM ⁺ 17b]	39.14%	42.56%	40.78%	27.3 (1.1s)	100%

TABLE 3.24 – Comparaison des résultats de la détection sur les deux bases de données vidéos ([ABSM⁺17b] et [BHM⁺18])

montrait de bonnes performances. Mais on peut s'intéresser aussi aux performances de notre outil, non pas pour la localisation des polypes, mais pour la détection de la présence des polypes dans les vidéos, c'est-à-dire signaler au médecin le fait qu'un polype est visible à l'écran, mais sans le localiser.

Dans le cadre de notre collaboration avec l'institut Computer Vision Center de Barcelone, une comparaison entre notre méthode (nommée ALCAPOD pour l'occasion, Active Learning for Computer Aided POLyp Detection), la méthode développée par J. Bernal (nommée WM-DOVA [BSFE⁺15]) et les réseaux de neurones convolutifs (CNN) a été effectuée (présentée dans [BHM⁺18]) concernant la détection de la présence des polypes dans les vidéos. En effet, l'utilisation des réseaux de neurones est de plus en plus présente pour la détection d'objets ou de personnes et devient une des méthodes de référence. Pour mesurer les performances, on utilise toujours les mêmes métriques présentées jusqu'ici : la Précision, le Rappel, le F₁ Score, le temps de réaction (RT) et le temps de calcul moyen pour traiter une image (MPT). Les méthodes ont été testées sur la base de données présentées dans [BHM⁺18] qui contient 18 vidéos pour un total de 29 657 images. Les résultats sont disponibles dans le tableau 3.25.

D'après le tableau 3.25, on peut voir que notre outil, en termes de performance pour la détection de la présence des polypes dans les vidéos, se place bien par rapport aux méthodes utilisant les réseaux de neurones. On peut aussi voir que par rapport aux réseaux de neurones, on obtient une précision plus élevée au détriment du rappel et du temps de réaction. Cependant, on est capable de traiter les images beaucoup plus rapidement, si on compare notre méthode par rapport à celle utilisant les "CNN (FT) + C", on a un gain de 14 millisecondes sur le temps de calcul. On est alors parfaitement capable de détecter la présence de polypes dans les vidéos en temps réel avec de très bonnes performances. De plus, contrairement aux réseaux de neurones convolutifs, notre outil est embarquable, ce qui est un argument non négligeable pour l'utilisation clinique. Notre outil serait donc ici le seul à être retenu pour cette application particulière qu'est la détection de la présence de polypes colorectaux dans les vidéos.

Méthode	Précision	Rappel	F ₁ Score	RT	MPT
WM-DOVA	80.2%	64.7%	71.6%	0.53s	10.50s
ALCAPOD	81.4%	79.3%	80.3%	0.12s	0.020s
CNN (FT)	79.6%	86.9%	83.1%	0.06s	0.038s
CNN (FT) + C	80.0%	91.4%	85.3%	0.05s	0.034s
CNN (FT,DA)	80.4%	85.6%	82.9%	0.06s	0.038s
CNN (FT,DA) + C	80.5%	88.9%	84.5%	0.05s	0.028s
CNN (S)	75.6%	74.7%	75.1%	0.23s	0.038s
CNN (S) + C	75.5%	79.9%	77.6%	0.21s	0.033s
CNN (S,DA)	74.1%	86.3%	79.7%	0.07s	0.038s
CNN (S,DA) + C	73.4%	89.2%	80.5%	0.06s	0.033s

TABLE 3.25 – Comparaison des résultats de la détection de la présence des polypes dans les vidéos de la base [BHM⁺18]. FT : réglage fin, S : à partir de rien, DA : augmentation des données, C : compression