

الجمهورية الجزائرية الديمقراطية الشعبية

وزارة التعليم العالي والبحث العلمي

UNIVERSITÉ BADJI MOKHTAR - ANNABA
BADJI MOKHTAR – ANNABA UNIVERSITY



جامعة باجي مختار – عنابة

Faculté : TECHNOLOGIE

Département : ELECTRONIQUE

Domaine : SCIENCES ET

TECHNIQUES

Filière : Télécommunication

Spécialité : Système et Réseaux de
lécommunication

Mémoire

Présenté en vue de l'obtention du Diplôme de Master

Thème :

**Systeme intelligent basé sur la vision pour la détection des
chutes des personnes âgées**

Présenté par : *Youla Chaïma*

Haiahem Ghozlane

Encadrant : *NASRI Seif Allah El Mesloul*

MCB

UBMA

Co-Encadrant : *BENRACHOU Djamel Eddine Maître de recherche*

QUT Australie

Jury de Soutenance :

FEZARI Mouhamed	Prof	UBMA	Président
NASRI Seif Allah El Mesloul	MCB	UBMA	Encadrant
BENRACHOU Djamel Eddine	Maître de recherche	QUT Australie	Co-encadrant
BOULMAIZ Amira	MCB	UBMA	Examineur

Année Universitaire : 2021/2022

Remercîment

Nous remercions tout d'abord ALLAH, LE TOUT PUISSANT qui nous a aidé et donné la persévérance et la patience durant ces longues années d'étude.

Nous tenons à remercier sincèrement nos Encadrants **Seif Allah Elmesloul Nasri** et **Djamel Eddine Benrachou** pour avoir accepté à nous encadrer dans cette étude et qui se sont toujours montrés à l'écoute et très disponible tout au long de la réalisation de ce mémoire, ainsi pour l'inspiration, l'aide, le soutien et l'encouragement. Et sans ces contributions ce mémoire n'aurait jamais vu le jour. Nous tenons à remercier également tous les membres du jury, leurs précieux conseils et remarques ont contribué à l'amélioration de la qualité scientifique de ce manuscrit.

Nous souhaitons adresser nos vifs remerciements les plus sincères aux personnes qui nous ont apporté leur aide et qui ont contribué à l'élaboration de ce mémoire ainsi qu'à la réussite de ce formidable semestre universitaire. Sans oublier les enseignants de la faculté technologie et le personnel administratif de l'université Badji Mokhtar Annaba, et plus précisément notre département d'électronique et nos enseignants de la spécialité -télécommunication-.

Merci à toutes et à tous.

Dédicace

Je dédie ce travail à mes chers parents :

Maman Feriel

Ma source de vie et d'amour qu'Allah la protège et la garde pour moi

Mon très chère Papa Mourad

Tu as toujours été à mes côtés pour me soutenir et m'encourager que ce travail traduit ma gratitude et mon affection, je t'aime papa

A mon frère Mohamed Ilyes et toute mes sœurs exceptionnellement Lydia, Julia, Maya qui m'ont chaleureusement encouragé tout au long de mon parcours. Ainsi que mes amies de ma vie Yousra et Yousra A tous les membres de ma famille et toute personnes qui porte le nom

YOULA

A tous ceux que j'aime et ceux qui ont partagé avec moi les moments d'émotion lors de la réalisation de ce projet

A Ghazlane mon binôme je te souhaiter pleine de succès et de réussite dans tes activités futures

Chaïma

Dédicace

Du profond de mon cœur, je dédie ce travail

À ma maman, qui m'a soutenu et encouragée durant ces années d'études. Ton affection me couvre, ta bienveillance me guide et ta présence à mes côtés a toujours été ma source de force pour affronter les différents obstacles.

À mon père, aucune dédicace ne saurait exprimer mon respect, Tu as toujours été à mes côtés pour me soutenir et m'encourager, je ne pouvais espérer avoir un père plus formidable que toi. J'ai eu beaucoup de chance que le bon Dieu t'a choisi pour que tu sois mon papa. Je vous remercie pour le soutien que vous me portez depuis mon enfance et j'espère que votre bénédiction m'accompagne toujours. Puisse Dieu, le Très Haut, vous accorder santé, bonheur et longue vie.

À mon frère ishak, ma sœur Amani qui m'ont encouragé sans cesse et cru en moi, À mes très chers amis, mes proches et à ceux qui me donnent de l'amour et de la vivacité.

Sans oublier mon binôme Chaïma, pour son soutien moral, sa patience et sa compréhension tout au long de ce projet.

À tous, je présente mes sincères remerciements et ma profonde gratitude.

Ghozlane

Abstract

As we move towards a post-Covid world we are pushed to think about more efficient means for creating more healthy lifestyles. Providing healthcare outside of the hospital settings is particularly crucial for one of the vulnerable segments of our society who are the elderly people living alone or in retirement homes. In this context, remote health monitoring systems provide a promising means for maintaining a close and continuous monitoring of long-term health and wellbeing conditions for individuals in a pandemic-like situation as well as in “normal” courses of everyday experiences, particularly for older adults. An old person living alone might be subject to a sudden falling in his living house. A fall for an old person may cause enormous effects especially if no emergency is provided urgently. In fact, fall detection systems stand as one of the important applications that can save thousands of people lives. This research project provides a brief review about fall detection systems. Based on that, a cost-effective fall detection system centred on vision and combining three benchmark deep learning model (Tiny YOLO, AlphaPose and ST-GCN) has been proposed and evaluated. Furthermore, a dataset with multiple scenarios has been created to evaluate the adopted system. Obtained results show the effectiveness of the joint model with certain limitations in some particular cases such as camera rotation and low light conditions of the captured environment.

Key words: Fall detection, Data set, Deep Learning, Tiny YOLO, AlphaPose, ST-GCN.

Résumé

Alors que nous nous dirigeons vers un monde post-covid, nous sommes poussés à réfléchir à des moyens plus efficaces de créer des modes de vie plus sains. Offrir des soins de santé en dehors du milieu hospitalier est particulièrement cruciale pour l'un des segments vulnérables de notre société, à savoir les personnes âgées vivant seules ou en maisons de retraite. Dans ce contexte, les systèmes de télésurveillance médicale constituent un moyen prometteur de maintenir un suivi continu des conditions de santé et de bien-être à long terme des individus dans une situation de type pandémie ainsi que dans les expériences de vie quotidiennes, en particulier pour les personnes âgées. Une personne âgée vivant seule peut être victime d'une chute soudaine dans sa maison. La chute d'une personne âgée peut entraîner des conséquences énormes, surtout si aucune mesure d'urgence n'est prise. En fait, les systèmes de détection des chutes constituent l'une des applications importantes qui peuvent sauver la vie de milliers de personnes. Ce projet de recherche fournit un bref aperçu sur les systèmes de détection des chutes. Sur cette base, un système de détection de chute moins coûteux et efficace centré sur la vision et combinant trois modèles d'apprentissage profond de référence (Tiny YOLO, AlphaPose et ST-GCN) a été proposé et évalué. En outre, une base de données avec de multiples scénarios a été créée pour évaluer le système adopté. Les résultats obtenus montrent l'efficacité du modèle conjoint avec certaines limitations dans certains cas particuliers tels que la rotation de la caméra et les conditions de faible luminosité de l'environnement capturé.

Mots clés : Détection de chute, base de données, apprentissage profond, Tiny YOLO, AlphaPose, ST-GCN.

ملخص

بينما نتحرك نحو العالم نحو مرحلة ما بعد كوفيد، فإننا مضطرون للتفكير في وسائل أكثر فعالية لخلق أنماط حياة أكثر صحة. يعد توفير الرعاية الصحية خارج المستشفيات أمرًا بالغ الأهمية بشكل خاص لأحد الشرائح الضعيفة في مجتمعنا، ألا وهي كبار السن الذين يعيشون بمفردهم أو في دار دور المسنين. في هذا السياق، توفر أنظمة مراقبة الصحة عن بُعد وسيلة واعدة للحفاظ على المراقبة الوثيقة والمستمرة للصحة على المدى الطويل للأفراد في وضع الحجر الصحي وكذلك في الأوضاع العادية للتجارب اليومية، خاصةً بالنسبة لكبار السن. قد يتعرض المسن الذي يعيش بمفرده لسقوط مفاجئ في منزل معيشتة. قد يتسبب سقوط شخص مسن في حدوث آثار خطيرة خاصة إذا لم يتم توفير النجدة بشكل مستعجل. في الواقع، تعتبر أنظمة الكشف عن السقوط أحد التطبيقات المهمة التي يمكن أن تنقذ حياة الآلاف من الناس. يقدم هذا المشروع البحثي مراجعة موجزة حول أنظمة الكشف عن السقوط. بناءً على ذلك، تم اقتراح وتقييم نظام للكشف عن السقوط فعال من حيث التكلفة يبني فعاليته على الصورة ويجمع بين ثلاثة نماذج قياسية للتعلم العميق (Tiny YOLO وAlphaPose وST-GCN). علاوة على ذلك، تم إنشاء قاعدة بيانات ذات سيناريوهات متعددة لتقييم النظام المعتمد. تظهر النتائج التي تم الحصول عليها فعالية النموذج المشترك مع بعض القيود في بعض الحالات المعينة مثل دوران الكاميرا وظروف الإضاءة المنخفضة للبيئة الملتقطة.

الكلمات المفتاحية: الكشف عن السقوط، قاعدة البيانات، التعلم العميق، Tiny YOLO، AlphaPose، ST-GCN.

Liste des tableaux

Tableau 1 Description de la base de données collectées	52
Tableau 2 Description matériel du PC utilisé	52

Liste des figures

Figure 1: Vue d'ensemble du système de détection et de prévention des chutes. [4].....	18
Figure 2 : L'architecture commune des systèmes de surveillance médicale [8].....	22
Figure 3: Les facteurs de chutes.....	25
Figure 4: (A) Une carte et (B) un histogramme des publications sur la détection des chutes par pays et régions de 1945 à 2020. [3].....	26
Figure 5 Classification des systèmes de détection de chutes.	27
Figure 6: Dispositifs médicaux et de soins de santé portables portés sur des parties du corps.[16]	28
Figure 7: Un exemple d'un système de détection basé sur des capteurs ambiants. [2].....	29
Figure 8 :Un exemple d'un système basé sur le traitement d'image. [2]	30
Figure 9: Détection humaine avec les boîtes de délimitations. [22]	33
Figure 10: Estimation de pose avec 17 point. [22].....	34
Figure 11: Modèle de détection de pose humaine base sur la méthode RMPE. [25].....	35
Figure 12 : Un exemple de calcul de l'intersection sur les unions pour diverses boîtes englobantes. [30]	37
Figure 13: Détection d'objets par le modèle YOLO. [27].....	38
Figure 14: Résultats TinyYOLO vs Résultats YOLO.[32].....	39
Figure 15: L'architecture de modèle TinyYOLO. [32]	39
Figure 16 : Estimation de pose multi-personnes avec AlphaPose.[35]	40
Figure 17: ST-GCN d'une séquence squelettique indiquer les articulations du corps. [38] ...	41
Figure 18 : Réseau de neurones artificiels simplifié du modèle utilisé.	44
Figure 19: Illustration du digramme globale des modèles utilisés.	45
Figure 20 Architecture YOLO v3. [39]	46

Figure 21 Fausse détection des points clés pour les poses de la personne par AlphaPose. [41]	
.....	47
Figure 22 : Modèle ST-GCN appliqué à des estimations de pose pour classifier les actions humaines. [38].....	48
Figure 23 : Application du modèle ST-GCN.....	48

Liste des abréviations

BAN: Body Area Network.

PAN: Personal Area Network.

IA: Intelligence artificielle.

ML: Machine learning.

DL: Deep learning.

YOLO: You Look Only Once.

Tiny YOLO : Une variante du détecteur d'objets « YOLO ».

CNN: Convolutional Neural Networks.

COCO: Common Objects in Context.

SPPE : Estimateur de la pose d'une seule personne.

SSTN : réseau de transformateurs spatiaux symétriques.

PGPG : générateur de propositions guidées par la pose.

RMPE : Estimation régionale des poses de plusieurs personnes.

ST-GCN : Réseaux convolutifs spatiaux graphiques temporels.

IoU : Intersection sur l'Union.

MOTA : Métrique de précision du suivi d'objets multiples.

(mAP) : Métrique de La précision moyenne.

Sommaire

Abstract	5
Résumé.....	6
ملخص	7
Liste des tableaux	8
Liste des figures.....	9
Liste des abréviations	11
Sommaire	12
Introduction générale :	15
Chapitre 1 :	17
État de l’art sur les systèmes de détection des chutes des personnes âgées	17
1.1 Introduction.....	17
1.2 Les problèmes, les tendances et les défis des systèmes de détection de chute humaine :	17
1.2.1 Les problèmes :	18
1.2.2 Les tendances :	19
1.2.3 Les défis :	20
1.3 L’architecture commune des systèmes de télésurveillance médicale :	21
1.3.1 Tier 1 : Les Capteurs :	22
1.3.2 Tier 2 : La passerelle :	22

1.3.3 Tier 3 : Le traitement et l'analyse des données :	23
1.4 Les personnes âgées et la chute :	23
1.5 Classification des systèmes de détection des chutes :	26
1.5.1 Système basé sur des dispositifs portables :	27
1.5.2 Système basé sur des capteurs ambiants :	28
1.5.3 Système basé sur le traitement d'image (système par vision) :	29
1.5.4 Fusion des systèmes de détection des chutes :	30
1.6 Conclusion :	31
Chapitre 2 : Contexte de recherche	32
2.1 Introduction :	32
2.2 Les étapes de la détection de chute humaine :	32
2.2.1 La détection de personne :	33
2.2.2 La détection de squelette (estimation de pose humaine) :	34
2.2.3 La prédiction de mouvement de personne :	36
2.2.4 Métriques de précision des modèles de détection d'objets :	36
2.3 Les techniques utilisées pour détecter et prédire la chute humaine :	38
2.3.1 Tiny-YOLO oneclass :	38
2.3.2 AlphaPose :	40
2.3.3 Le modèle ST-GCN :	40
2.5 Conclusion :	43
Chapitre 3 :	43
Méthodologie & Résultats	43

3.1 Introduction :	43
3.2 Diagramme de l'Algorithme :	44
3.3 Challenge et Scenarios de teste :	49
3.4 Les bases de données :	50
3.4.1 La base de données COCO	50
3.4.2 La base de données Le2i	51
3.5 Implémentation	52
3.6 Résultats et Discussion :	53
3.7 Conclusion :	59
Conclusion générale et perspectives	60
Références :	62

Introduction générale :

Dans le monde, le nombre total de personnes âgées augmente plus rapidement que les autres groupes d'âge [1]. Par conséquent, la part des personnes âgées augmente presque partout. En 2015, une personne sur huit dans le monde était âgée de 60 ans ou plus. D'ici 2030, une personne sur six appartiendra à ce groupe d'âge dans le monde. Selon l'OMS, 37 millions de chutes sont suffisamment graves pour nécessiter des soins médicaux chaque année¹. Ce sont les sujets de plus de 65 ans qui sont les plus exposés. On admet que 25 à 30 % des personnes de plus de 65 ans vivant chez elles chutent au moins une fois par an.

Cette problématique d'ordre international lance un appel à la communauté scientifique pour trouver une solution intelligente qui profite de l'avancement rapide des nouvelles technologies de communication et intelligence artificielle.

Les systèmes de détection des chutes humaines sont là dont l'objectif de sauver des vies. Des dizaines d'équipes de recherche à travers le monde s'intéressent à l'application de l'intelligence artificielle pour la détection automatique des chutes des personnes, qui permet le déclenchement rapide des processus de secours. C'est exactement dans ce contexte que se situe notre travail de recherche.

La plupart des solutions proposées pour la détection des chutes sont basées soit sur des capteurs portatifs tels que les accéléromètres et/ou gyroscopes pour détecter le mouvement de la personne, soit sur des capteurs ambiants (présence, pression) placés à l'environnement où sur la personne elle-même. Bien que ces systèmes puissent assurer une détection prometteuse des chutes, les capteurs portatifs présentent des limites d'utilisation. D'une part, ils représentent une solution peu pratique, étant donné que plusieurs personnes âgées refusent de porter ces dispositifs car ils ne sont généralement confortables. Pour les capteurs ambiants l'installation est assez difficile avec des résultats parfois insatisfaisants en raison de plusieurs facteurs externes affectant la détection précise. Une tendance vers la fusion des capteurs a montré son efficacité dans ce domaine, mais c'est une solution coûteuse qui exige des interventions et l'entretien continue souvent. Dans des cas elle ne respecte pas la confidentialité des personnes.

¹ World Health Organization (OMS). Les chutes. Aide-mémoire. 2012;344:1-4.

Notre système de détection des chutes adopté dans ce projet est basé sur un dispositif à faible coût comprenant un ordinateur pour le traitement de données et une seule caméra. Cet appareil peut être installé dans les murs ou les plafonds et surveiller une pièce sans intervention humaine.

En outre, les personnes surveillées à domicile ne sont pas tenues de porter des dispositifs. Ainsi, le système est capable d'effectuer une détection de chute continue même sous conditions difficiles comme il est le cas de faible éclairage.

Ce manuscrit est divisé en trois parties. Le premier chapitre présente un état de l'art sur les systèmes de détection des chutes des personnes âgées. Dans cette partie nous présentons les problèmes, les tendances et les défis des systèmes de détection de chute. Nous allons décrire aussi l'architecture commune de ces systèmes ainsi que leur classification à partir des capteurs utilisés.

Le deuxième chapitre explique les étapes de construction du système de détection de chute adopté, qui sont respectivement la détection de personne avec le modèle Tiny YOLO oneclass puis la détection de squelette avec AlphaPose et on termine avec le modèle ST-GCN pour la prédiction et la classifier l'action humaine dans le but de détecter la chute.

Pour le troisième chapitre nous allons présenter la méthodologie suivie afin d'arriver à notre objectif. De plus nous présentons les résultats d'évaluation du modèle composite à partir des images tirées des séquences vidéo résultantes.

C'est dans ce chapitre que nous dévoilons notre base de données créer pour évaluer la performance du système de détection des chutes présenté dans ce projet de fin d'étude.

En discutant les résultats avec une évaluation du système sur deux différentes bases de données, nous terminerons ce manuscrit par une conclusion générale.

Chapitre 1 :

État de l'art sur les systèmes de détection des chutes des personnes âgées

1.1 Introduction

Le développement de systèmes de détection et de prévention des chutes est devenu un sujet de recherche brûlant au cours des dernières années. Le premier système de détection des chutes a été mis au point au début des années 1970. Ce premier système basic se base sur une alarme qui se déclenche lorsqu'un bouton d'un transmetteur est enfoncé. Depuis, diverses approches sont utilisées pour développer les systèmes de détection des chutes humaines. Actuellement, ces systèmes sont classés en deux grandes catégories : les systèmes portables et les systèmes non portables. Dans ce chapitre nous présentons une explication détaillée sur les systèmes de détection de chute humaine en mettant l'accent sur les systèmes basés sur la détection des chutes par vision.

1.2 Les problèmes, les tendances et les défis des systèmes de détection de chute humaine :

L'intégration d'un système de prévention et de détection de chutes (Fall Prevention FP, Fall Detection FD) est nécessaire pour envoyer des alarmes en cas d'urgence, diminuer le temps d'intervention et augmenter les chances de survivre. Certains chercheurs ont classé les systèmes existants en fonction de deux variétés : types de chutes et types de capteur utilisé [3].

La figure (01) montre une vue d'ensemble de ces systèmes qui utilisent un ensemble de capteurs portables avec un analyse de données qui se base sur l'apprentissage machine pour la détection de l'action de chute.

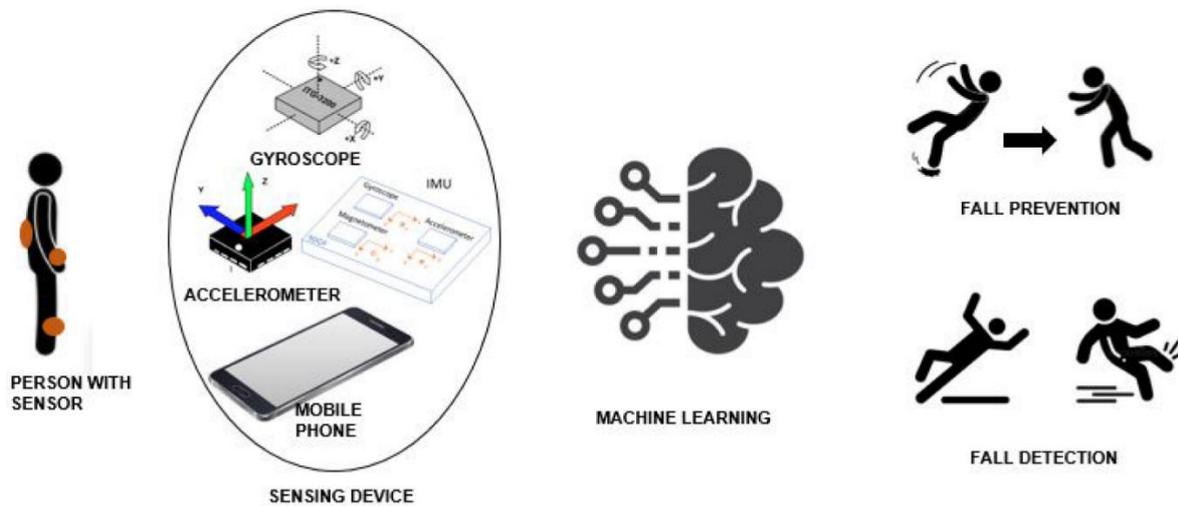


Figure 1: Vue d'ensemble du système de détection et de prévention des chutes. [4]

Malgré l'innovation continue dans les systèmes de détection de chute, cette évolution s'accompagne toujours avec certaines limites et défis qui doivent être discutées afin de pouvoir améliorer l'efficacité de détection et s'adapter avec les moyens existants et les différentes variétés de chutes.

Dans les sous-sections suivantes, nous présentons les problèmes, les tendances et les défis des systèmes de détection des chutes humaines [2] .

1.2.1 Les problèmes :

Les systèmes de détection des chutes actuelles présentent plusieurs problèmes qui empêchent le déploiement rapide de cette technologie dans la vie réelle des patients et des personnes âgées, nous pouvons citer les exemples suivants :

- Les remplacements fréquents des piles ou le besoin de chargement des batteries pour les capteurs portables.
- L'inconfort de l'utilisateur et le problème confidentialité et la vie privée de l'utilisateur, dont plusieurs utilisateurs refusent d'être sous contrôles des caméras connectés à l'internet.

- Les coûts élevés des ensembles des capteurs et des caméras dans le cas où le système de détection se base sur un ensemble de dispositifs fusionnés.
- La configuration compliquée qui nécessite l'intervention d'un expert pour l'installation et la maintenance du système.
- Le taux de calcul des données qui nécessite des machines performantes et coûteuses.

En fait, tous les systèmes non portables (ambiants) ne parviennent pas à détecter les chutes au-delà d'une distance limitée de dizaines de mètres.

Cependant, les systèmes de détection de chutes souffrent du phénomène des fausses alarmes à cause de taux de précision diminués de certains systèmes appliqués. Parfois, la fiabilité et la robustesse des appareils sont sacrifiées au profit de la portabilité.

Une augmentation de taux d'erreur est aussi remarquée à l'environnement extérieur ouvert et dynamique ou d'autres facteurs rentrent dans l'équation de détection des chutes. On doit aussi citer le manque de bases de données divers et publiques qui empêchent aussi le développement de systèmes de détection de chute notamment les systèmes de détection par vision. Ces bases de données doivent contenir des données réelles sur les flux et un cadre commun d'évaluation entre les différents systèmes.

1.2.2 Les tendances :

Afin de trouver des solutions à ces problèmes préalablement cités, plusieurs études ont été menées. La fusion de capteurs est l'une des tendances les plus populaires dans le développement de systèmes de détection de chute, Il s'agit d'une combinaison de nombreux capteurs, systèmes et algorithmes. Thinganos et al. [5] présente une comparaison entre trois schémas de fusion de données proposés et une étude dans laquelle un seul type de capteur et d'algorithme est utilisé, fournissant des informations utiles sur le problème de la détection des chutes.

Ainsi, pour diminuer le prix et augmenter le taux d'utilisation, des études intégrant la détection des chutes dans les smartphones ont été faites depuis 2009 [6]. Côté matériel, les smartphones sont naturellement faits pour l'acquisition de données et la transmission sans fil. De plus, de nos jours, les capteurs inertiels sont toujours intégrés aux smartphones.

Pour réduire le taux de fausses alarmes des techniques d'apprentissage automatique commencent à être appliquées dans les dispositifs de détection des chutes. Bien que les méthodes traditionnelles basées sur des seuils soient capables de détecter quand une chute se produit, le taux de fausses alarmes est toujours un problème. Chaque utilisateur est unique en taille et en poids et se comporte différemment dans une variété d'environnements de vie. Un seuil unique n'est ni suffisant ni précis. L'approche d'apprentissage automatique est plus sophistiquée et donc plus adaptative et conduit à de meilleures performances. Actuellement, plusieurs méthodes d'apprentissage automatique sont proposées, telles que les arbres de décision, la régression non linéaire, le réseau bayésien dynamique, et beaucoup plus. On ne trouve pas une seule méthode reconnue comme la plus efficace, des nouvelles approches sont encore introduites. Heureusement, un référentiel de chutes du monde réel est en cours de développement [2] et il y aura bientôt suffisamment de données de chutes réelles pour former et tester tous les algorithmes.

Il existe une nouvelle tendance vers un appareil EEG (Électroencéphalographie) portable pour faciliter la distinction entre les chutes réelles et les AVQ (Programme d'aide à la vie quotidienne) similaires. L'EEG (alias électroencéphalogramme) est une méthode non invasive largement utilisée pour mesurer la dynamique et les performances cérébrales [2].

Une tendance vers le suivi dans les voitures autonomes, est une caractéristique importante utilisée pour déduire la vitesse et la direction d'un objet en mouvement, généralement un piéton ou une voiture. Il aide à prévoir avec précision la trajectoire future d'un objet détecté. Cette technique peut être appliquée pour la détection des chutes des personnes pour une précision élevée [7].

1.2.3 Les défis :

Le sujet de la détection des chutes a été beaucoup étudié au cours des deux dernières décennies et de nombreuses expérimentations ont été proposées. Le développement rapide des nouvelles technologies maintient ce sujet très actif dans la communauté de la recherche scientifique. Bien que de nombreux progrès ont été réalisés, il reste encore divers défis ouverts, dont nous mentionnons : La rareté des données de chutes réelles, la détection en temps non réel, la sécurité et la confidentialité, des limitations de l'emplacement et l'évolutivité et la flexibilité dans les plateformes de fusion de capteurs.

Un autre défi très important dans les systèmes de détection de chute par vision c'est l'angle de vue et la direction de personne pendant le processus de suivi (la rotation de la personne suivie), les renseignements fournis ne tiennent pas compte du fait que la personne est face à la caméra ou non. Cela rend difficile pour l'algorithme de séparer les angles supplémentaires tels que 0° et 180°. La plupart des travaux antérieurs sur les algorithmes de suivi tentent d'être invariants aux nuisances qui perturbent ses performances telles que l'occlusion partielle, changement d'éclairage, déformation du corps, la capacité d'identifier et de dépasser de tels problèmes au cours du suivi conduira à une meilleure représentation, ce qui rendra l'algorithme plus puissant dans des situations différentes de la vie réelle [7].

1.3 L'architecture commune des systèmes de télésurveillance médicale :

Les systèmes de détection de chute font partie de la télésurveillance médicale (systèmes de télésanté). L'idée de détecter ou de signaler instantanément une dépend principalement de la surveillance à distance des personnes âgées ou les gens fragiles qui tombent souvent. Les systèmes de télésurveillance médicale dans leur ensemble dépendent d'une architecture commune de trois niveaux qui fonctionnent ensemble pour l'observation et la détection, suivi par la transmission des données sensorielles du patient jusqu'à la collection et l'analyse de ces dernières. Ces niveaux sont respectivement : capteurs, passerelle, et traitement et analyse des données. La figure (02) illustre de manière simple les composants de chaque niveau et l'interaction entre eux et aussi les technologies utilisées dans chaque niveau [8].

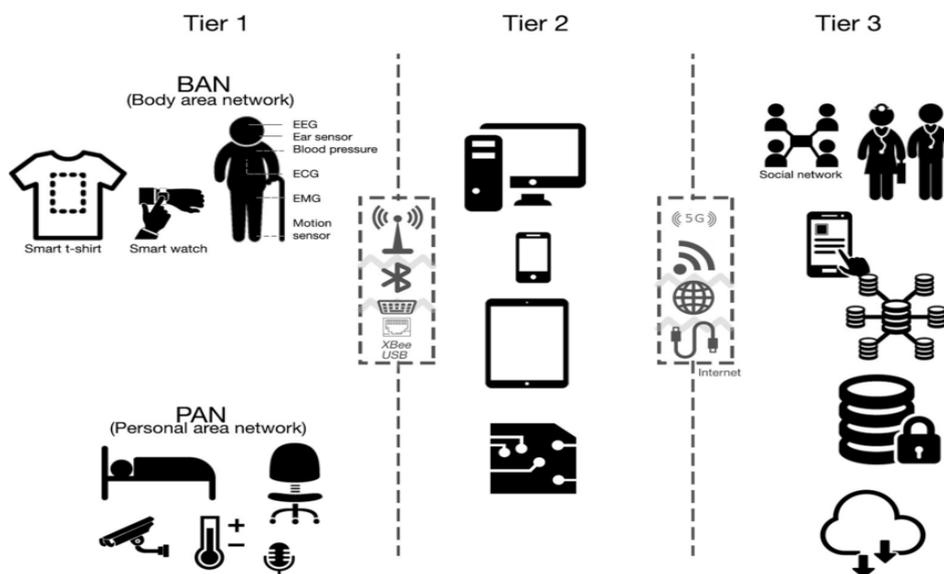


Figure 2 : L'architecture commune des systèmes de surveillance médicale [8]

1.3.1 Tier 1 : Les Capteurs :

Ce niveau est le système d'acquisition de données. L'évolution croissante de l'industrie microélectronique a ouvert la voie pour le développement de différents types de capteurs qui peuvent collecter des données clinique ou données environnementale par rapport aux patients. Ces capteurs sont devenus largement accessible à l'industrie et aux particuliers à domicile où ils sont chargés de collecter les signaux médicaux des patients. On peut décrire la procédure de collection des données par ces capteurs en deux types ; Le premier en les attachant les capteurs au corps du patient pour former ce qu'on appelle un réseau corporel ou BAN (Body Area Network) et le deuxième type des capteurs est déployer dans l'environnement intérieur du patient en tant que PAN (Personal Area Network) ou capteurs ambiants.

1.3.2 Tier 2 : La passerelle :

Une passerelle est généralement utilisée pour la collection des données à partir des capteurs, et aussi pour le transfert sur un serveur distant situé à l'hôpital ou vers l'unité de surveillance. Le transfert de données peut être effectué vers un espace cloud afin d'éviter le problème des capacités limitées des unités de traitement locaux. Il existe des systèmes qui effectuent l'analyse localement puis transfèrent les données dérivées à un serveur distant pour une analyse profonde tout en assurant un historique d'enregistrement comme il en existe d'autres

qui effectuent l'analyse localement pour une prise de décision plus rapide et une meilleure utilisation de la bande passante. Dans ce cas seulement des mises à jour partielles sont envoyées dans le cloud.

Le rôle de passerelle peut être simplement un appareil dédié à un smartphone connecté à internet et situé à la place du patient [9]. On peut classer les modes de communication en deux : unidirectionnel, des capteurs vers le serveur via la passerelle ou bidirectionnel où les informations peuvent être envoyées aux capteurs nœud par la passerelle afin de modifier certains paramètres [10].

1.3.3 Tier 3 : Le traitement et l'analyse des données :

On peut considérer ce niveau comme le terminal distant qui peut être à l'hôpital. Les signaux médicaux ou environnementaux collectés par les capteurs sont généralement transmis à un système distant serveur ce système implique quatre fonctionnalités principales, la première est l'exécution des algorithmes pour l'interprétation des données collectés, la deuxième est de générer des alertes en cas d'urgence et puis l'enregistrement des données de chaque patient dans une base de données pour les prochaines analyses, et finalement la quatrième fonctionnalité qui est responsable de fournir une interface utilisateur graphique GUI pour permettre la surveillance en temps réel de l'état de patient.

1.4 Les personnes âgées et la chute :

L'Organisation Mondiale de la Santé estime que la proportion de personnes âgées de plus de 60 ans va doubler d'ici 2050 et représenter 22% de la population totale. Il y aura donc deux milliards de personnes âgées dans 30 ans [11]. Cela représente un défi pour les centres médicaux et de soins car les patients âgés sont connus d'être les plus sensibles aux pathologies. Les personnes âgées sont aussi les plus susceptibles aux chutes. La chute correspond à l'action de tomber au sol indépendamment de sa volonté [13]. La chute expose les personnes âgées à des risques et des graves complications qui peuvent être parfois mortelles. Certains âgés préfèrent de vivre seuls, ce qui augmente les taux de risque d'accidents domestiques, d'où vient la nécessité d'appliquer un système de détection de chute domestique.

Les facteurs et les risques de la chute :

Comme à tout âge, tomber peut entraîner des conséquences physiques parfois graves chez les personnes âgées. Il arrive qu'une chute ne se traduise que par une douleur persistante et modérée que la personne supporte plusieurs jours avant d'accepter de consulter et de découvrir l'existence d'une fissure ou d'une fracture osseuse. Avec l'âge la perception de la douleur se modifie et peut fausser la gravité des conséquences d'une chute. Parmi les conséquences de la chute, certaines surviennent lorsque la personne âgée est incapable de se relever et reste au sol longtemps avant que quelqu'un vienne la secourir. Cette immobilisation prolongée au sol peut augmenter le risque et provoquer des complications médicales.

Selon l'Organisation mondiale de la santé (OMS), environ 646 000 chutes mortelles surviennent chaque année dans le monde, dont la majorité sont subies par des adultes de plus de 65 ans (OMS, 2018)[1]. Cela en fait la deuxième cause de décès par blessure non intentionnelle, suivi des accidents de la route. Globalement, les chutes constituent un problème majeur de santé publique pour les personnes âgées [14].

Les chutes peuvent avoir aussi des conséquences sociales, financières et médicales néfastes. Après une première chute, l'indépendance de la personne diminue à cause de la peur de tomber à nouveau, ce qui l'empêche de rester active. Par conséquent, ces personnes âgées ont besoin de soins permanents, ce qui coûte cher pour la société ainsi que pour l'état.

La chute est causée par plusieurs facteurs intrinsèques et extrinsèques, la Figure (03) montre les principaux facteurs de risque qui perturbent l'autonomie des personnes âgées.

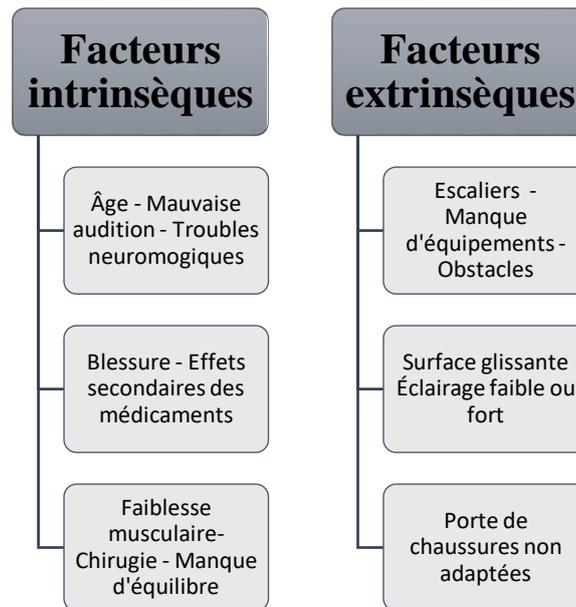


Figure 3: Les facteurs de chutes.

Une étude statistique montre qu'un adulte sur trois âgé de 65 ans ou plus tombe au moins une fois par an. Par ailleurs, le risque de chute est multiplié par 20 après une première chute et le risque de décès s'accroît d'un facteur de quatre dans l'année qui suit la chute. L'impact et les conséquences d'une chute peuvent être réduits lorsque l'événement est détecté en temps réel et que des soins médicaux sont fournis rapidement [15]. Pour améliorer la qualité de vie des personnes âgées indépendants, il faut d'abord identifier les conséquences des chutes, ensuite examiner les facteurs de risques et enfin proposer des solutions afin de réduire indirectement les dommages faits à la personne.

La personne tombe par terre et au moment de cette phase, des blessures graves peuvent se produire. La dernière phase (Post-chute) se déclenche juste après la chute. Pendant cette phase, l'accélération du corps diminue et la personne est en état de choc. De nombreux travaux de recherche ont été menés pour réduire les conséquences des chutes et améliorer la qualité de vie des personnes âgées. Les chercheurs ont proposé un dispositif communicant qui s'intègre dans la vie de la personne dans l'objectif d'alerter en cas de chute. À chaque événement anormal, une alarme, contenant l'état de la personne, est envoyée aux soignants. Cette alarme est générée à l'aide d'un système de détection de chute automatique. Dans le cas d'un évènement critique, une décision sera envoyée aux personnes concernées (infirmier, proches) sous forme d'une alarme.

La solution donc est un système de détection de chutes dans le but est la surveillance et la prévention. L'alerte d'une chute instantanément est considérablement une prévention des complications médicales et aussi un secours en temps réel pour augmenter les possibilités d'aide et d'intervention rapide.

1.5 Classification des systèmes de détection des chutes :

Au cours des dernières décennies, beaucoup d'efforts ont été déployés dans le domaine de détection de chute pour améliorer la précision des systèmes de détection et de prédiction des chutes ainsi que pour diminuer les fausses alarmes. La figure (4) montre les 25 premiers pays en termes de nombre de publications sur la détection des chutes de 1945 à 2020. La plupart des publications proviennent des États-Unis, suivis par l'Angleterre, la Chine et l'Allemagne. Les données présentées indiquent que les pays développés investissent davantage dans la recherche dans ce domaine par rapport aux autres pays. En raison d'un niveau de vie plus élevé et de meilleures ressources médicales, les habitants des pays développés sont plus susceptibles d'avoir une espérance de vie plus longue, ce qui entraîne un vieillissement de la population plus élevé dans ces pays [3].

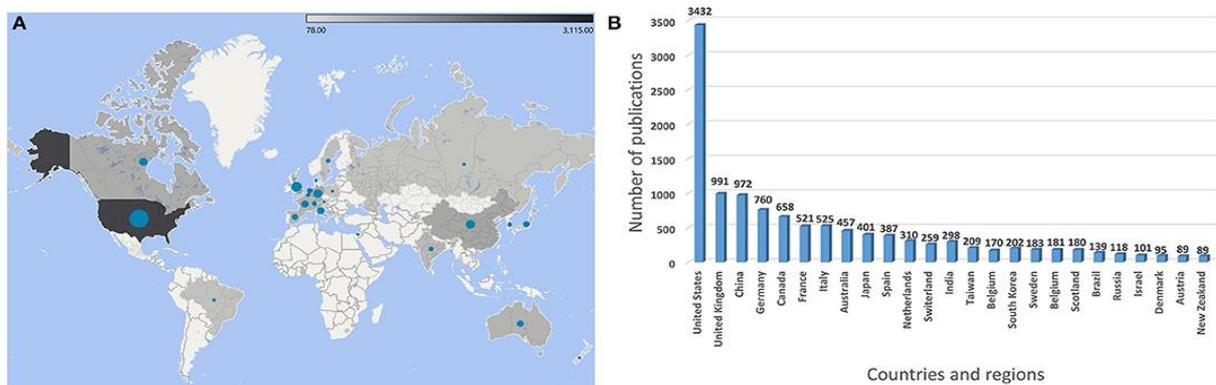


Figure 4: (A) Une carte et (B) un histogramme des publications sur la détection des chutes par pays et régions de 1945 à 2020. [3]

Une classification générale catégorise les systèmes avec capteurs permettant de détecter les chutes en trois catégories : systèmes à base de capteurs portables, systèmes à base de capteurs ambiants et les systèmes basés sur la vision. En effet, cette classification permet une meilleure compréhension des systèmes liés aux chutes. Nous l'illustrons dans la Figure (05).

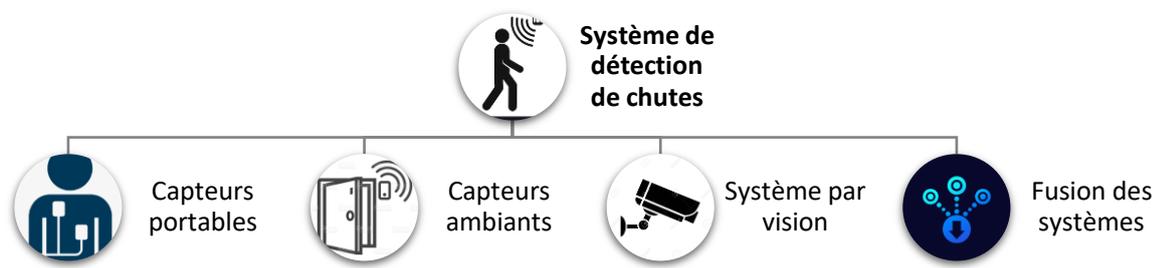


Figure 5 Classification des systèmes de détection de chutes.

1.5.1 Système basé sur des dispositifs portables :

Les appareils portables sont considérés comme l'un des principaux types de capteurs pour la détection des chutes et ils ont été largement étudiés en raison des avantages de mobilité, portabilité, du faible coût et de la disponibilité. Plusieurs études ont été menées pour étudier les dispositifs portables, qui sont considérés comme une voie prometteuse pour étudier la détection et la prédiction des chutes [3].

Un capteur porté par la personne est un dispositif électronique posé sur le corps de la personne ou sur ces vêtements pour réaliser une sorte de diagnostic et de surveillance non instructifs et non invasifs, en collectant des signaux corporels. Ces dispositifs fournissent des informations sur les mouvements de la personne ainsi que sa température et sa tension, ce qui est utile pour la détection de la chute.

L'idée du capteur portable est de comparer l'activité normale de la personne (s'allonger, s'asseoir et marcher) avec la chute. En général, les accéléromètres et les gyroscopes sont utilisés pour mesurer l'accélération et l'orientation de chaque partie du corps.

Un capteur portable est un dispositif autonome en énergie, il possède sa propre source d'énergie (pile ou batterie). Il contient une unité de traitement et une unité de communication. Il transforme les grandeurs physiques observées (température, mouvement, etc.) en données numériques compréhensibles par l'unité de traitement. Cette dernière est l'unité principale du système portable. Elle sert à gérer les périphériques, à programmer l'ordonnancement des

tâches, à traiter les données et à gérer l'énergie (mise en veille et réveil du capteur selon la situation). Elle traite les données soit en local, ce qui nécessite de l'espace en mémoire et la communication avec un serveur. L'unité de communication est responsable de la transmission de données au serveur via une technologie sans fil comme le Wi-Fi ou le Bluetooth.

La plupart des systèmes récents se composent d'une combinaison de deux capteurs ou plus, y compris un accéléromètre et un gyroscope. Dans de nombreux cas, tels systèmes sont également accompagnés d'un dispositif de protection par exemple un airbag embarqué prêt à se déployer en cas de chute. [2]. La figure (06) montre des exemples sur quelques dispositifs portables portés sur des parties du corps.

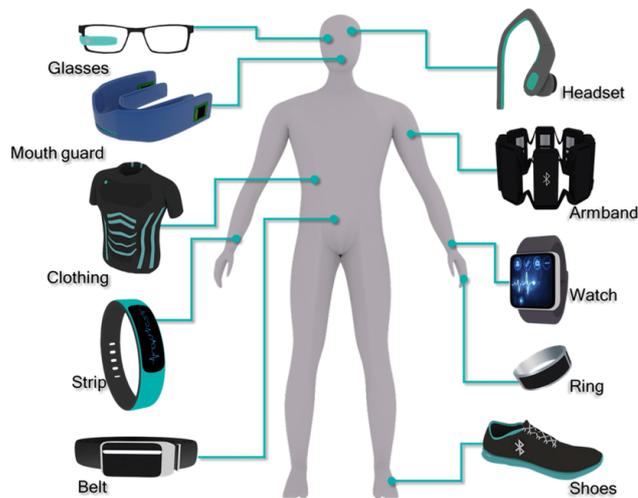


Figure 6: Dispositifs médicaux et de soins de santé portables portés sur des parties du corps.[16]

1.5.2 Système basé sur des capteurs ambiants :

Pour éviter les problèmes des capteurs portables, les chercheurs ont développé une autre solution de surveillance et de détection des chutes. Ce type de systèmes repose sur des capteurs ambiants, installés directement chez la personne âgée ou dans les maisons de retraite. Les données recueillies à partir des capteurs sont envoyées à un ordinateur pour les traiter et déclencher des alarmes dans les cas d'urgence. Ces capteurs peuvent être des capteurs de pression et de vibration au sol, des capteurs infrarouges muraux ou placés au plafond ou des capteurs acoustiques. La figure (07) montre un exemple de ses systèmes.

Les systèmes basés sur des capteurs ambiants sont sensibles à l'environnement extérieure, puisqu'ils tiennent compte des facteurs environnementaux dynamiques qui pourrait affecter la détection [2].

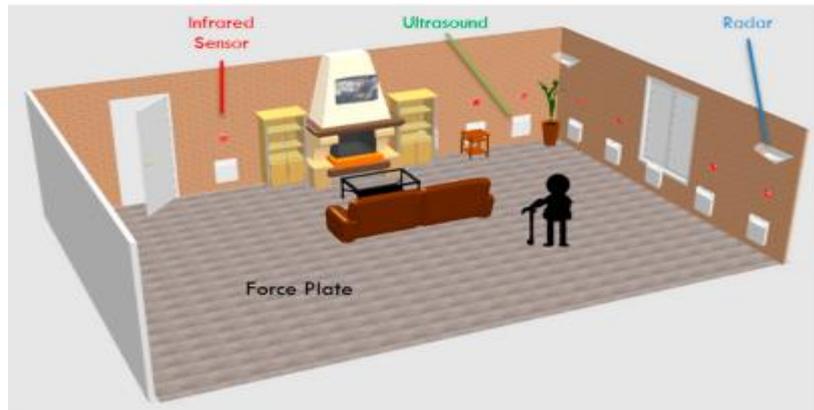


Figure 7: Un exemple d'un système de détection basé sur des capteurs ambiants. [2]

Contrairement aux autres techniques, les problèmes de confidentialité sont éliminés en utilisant ces capteurs ambiants. Mais, ce type de capteurs est sensible au bruit. Il est facilement perturbé par l'environnement et il est aussi coûteux. Le développement des technologies de détection des chutes suit une progression notable. Par ailleurs, plusieurs idées d'utilisation de capteurs ambiants restent restreintes à la recherche scientifique sans avoir recours à la commercialisation [3].

1.5.3 Système basé sur le traitement d'image (système par vision) :

Pour améliorer la solution basée sur les capteurs ambiants et éviter les problèmes de dépendance des capteurs portables, des chercheurs ont développé des solutions basées sur la vision. Comme le montre la figure (08). La détection basée sur la vision est une autre méthode importante. De nombreux efforts dans ce sens ont été démontrés, et dont certains montrent des performances prometteuses [2].

Pour résumer, les systèmes de détection de chute basés sur le traitement d'images sont en quelque sorte similaires aux systèmes ambiants. Ils partagent des avantages dans l'analyse des facteurs environnementaux que les appareils portables n'ont pas. En outre, ils ont tous les deux le problème de la restriction intérieure, des angles morts et les dépenses énormes. Cependant, les systèmes de traitement d'image ont leurs propres avantages uniques. Grâce à

la reconnaissance des formes, les systèmes de traitement d'image peuvent identifier, suivre et surveiller l'utilisateur cible même lorsqu'il y a plusieurs personnes ou animaux de compagnie dans la zone surveillée. Toutefois, ces systèmes d'enregistrement vidéo détaillés sont toujours soumis à des préoccupations en matière de confidentialité ainsi qu'à la nécessité d'un calibrage entre plusieurs caméras et d'algorithmes complexes de traitement d'images en temps réel qui consomment une énorme quantité d'espace et de puissance de calcul.

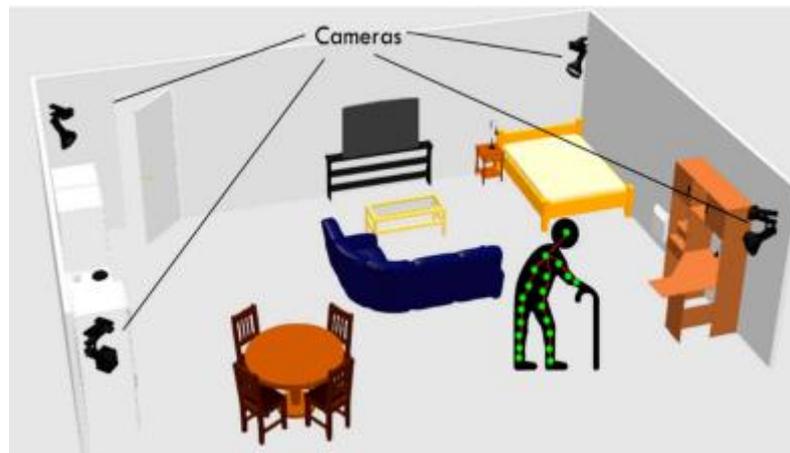


Figure 8 :Un exemple d'un système basé sur le traitement d'image. [2]

La méthode utilisée dans le système de surveillance basé sur la vision est plus détaillée et claire parce qu'elle suit les articulations des personnes. Elle peut même tracer la posture humaine pour déterminer sa position. Notre étude se base sur cette méthode d'où plus de détails sont présentés dans le prochain chapitre.

1.5.4 Fusion des systèmes de détection des chutes :

La fusion de capteurs fournit une approche plus robuste pour la détection des chutes chez les personnes âgées. L'utilisation de divers capteurs peut se compléter dans différentes situations. Ainsi, au lieu de compter sur un seul capteur, qui peut ne pas être fiable dans des cas particuliers, le système dépend de différents types de capteurs qui peuvent collecter des données diverses dans différentes conditions. Le résultat va être un système plus robuste qui réduit les fausses alarmes avec une grande précision [3].

Les systèmes fusionnés de détection des chutes forment un réseau de capteurs travaillant en corrélation pour détecter une chute. Dans [17], les données d'accélération sont observées à partir du nœud de capteur fixé sur le corps et transmises à la station de base, qui est un

ordinateur, pour détecter les chutes, tandis que la force du signal RF (Radio Fréquence) est utilisée pour localiser l'utilisateur. Les résultats obtenus indiquent qu'avec une telle combinaison, les activités normales ne produisent pas une fausse détection de chute. Dans [18], une méthode de détection des chutes à l'aide d'un système de localisation intérieure combinant l'ultra-large bande (UWB) et accéléromètre est présentée.

En effet, les systèmes combinés ont neutralisé certains points faibles de chaque système indépendant et apporté plus de précision dans la détection des chutes. La vulnérabilité est toujours évidente. À vrai dire, les systèmes combinés (fusionnés) sont incapables de faire la différence entre une chute accidentelle et une activité semblable auto-initiée. En outre, les systèmes combinés sont très coûteux et moins ergonomiques.

1.6 Conclusion :

Dans ce chapitre nous avons commencé par mentionner les défis et les différents problèmes existants dans les systèmes de détection de chute humaine et en suite nous avons présenté les tendances actuelles dans ce domaine pour bien comprendre les motivations de notre projet. Nous avons aussi discuté les chutes et leurs facteurs de risque par rapport aux personnes âgées.

Une classification des systèmes de détection des chutes a été présentée avec un aperçu général pour chaque catégorie.

Le prochain chapitre concentre sur les systèmes de détection des chutes basés sur la vision, qui la méthode adoptée dans notre projet de fin d'étude.

Chapitre 2 : Contexte de recherche

2.1 Introduction :

Dans ce chapitre, nous présentons les trois étapes de détection de chute par vision de l'approche adoptée, qui sont respectivement : la détection de personne, la détection de squelette (estimation de pose) et finalement la prédiction de mouvement humain. La chute est parmi les mouvements que le modèle adopté peut détecter.

Nous allons aussi expliquer les techniques de traitement d'image utilisées dans ce projet : En premier lieu l'utilisation de TinyYOLO oneclass qui est utilisé comme méthode efficace pour la détection des personnes, puis AlphaPose pour l'estimation de pose et ensuite le modèle ST-GCN qui peut prédire les mouvements des personnes détectées.

2.2 Les étapes de la détection de chute humaine :

Beaucoup de travaux sont proposés sur les systèmes de détection de chute en utilisant différents capteurs et différents algorithmes. La détection des chutes chez l'humain en général comporte plusieurs méthodes et systèmes, la méthode utilisée dans notre thèse se compose de trois phases comme suite : la détection de personne, détection de squelette, la prédiction de mouvement de personne. Cette méthode détecte la chute en temps réel d'une manière efficace et fiable avec un taux minimal de fausses alarmes. La détection et le suivi humains sont généralement considérés comme les deux premières actions d'une vidéosurveillance et peuvent alimenter des modules de raisonnement de haut niveau tels que la prédiction d'action et l'analyse dynamique [19].

Notre modèle présente plusieurs avantages par rapport aux systèmes à base de classificateurs simples. Il analyse l'image entière au moment de l'essai, de sorte que ses prédictions sont éclairées par le contexte global de l'image. Il fait également des prédictions avec une seule évaluation de réseau contrairement à des systèmes comme R-CNN[20] qui nécessitent des milliers d'évaluations des réseaux de neurones pour l'analyse d'une seule image [19]. Ce qui rend le model adopté extrêmement rapide, plus de 1000x que R-CNN et 100x plus rapide que Fast R-CNN[21].

2.2.1 La détection de personne :

Au cours des dernières années, la détection de personnes dans une scène vidéo a attiré plus d'attention en raison d'une large gamme d'application, y compris la détection d'événements anormaux, le dénombrement des personnes dans une foule, l'identification individuelle, la classification selon le sexe et la détection des chutes chez les personnes âgées. [22]

La détection et le suivi des humains font partie des systèmes de vision par ordinateur pour localiser et suivre les personnes en vidéo. La détection des personnes est la tâche de localiser toutes les instances d'êtres humains présents dans une image. Elle est faite en recherchant tous les zones de l'image, à toutes les échelles possibles, et en comparant une petite zone à chaque emplacement avec les modèles connus de personnes. Le suivi des personnes est le processus d'association temporelle des détections dans une séquence vidéo. Ce projet adopte un nouveau modèle Tiny-YOLO oneclass pour détecter l'être humain. Le modèle Tiny-YOLO oneclass est entraîné sur la base de données ensemble COCO [23] pour une détection de personnes plus robuste sous différents angles de vision.

La technique de détection des personnes possède deux approches différentes : la première est la détection des boîtes de délimitations. La boîte de délimitation est un rectangle sur les images, décrivant l'objet d'intérêt (dans ce cas, les personnes) dans chaque image en définissant ses coordonnées X et Y ainsi que la largeur et la hauteur comme le montre la figure (09). La deuxième approche est une approche de segmentation consiste à diviser une image en fonction des caractéristiques des pixels. Cette approche regroupe les pixels avec des attributs similaires et crée un masque de pixels pour l'apparence de chaque objet [22]. Dans notre cas le modèle TinyYOLO se base sur la première méthode avec plus de développement.

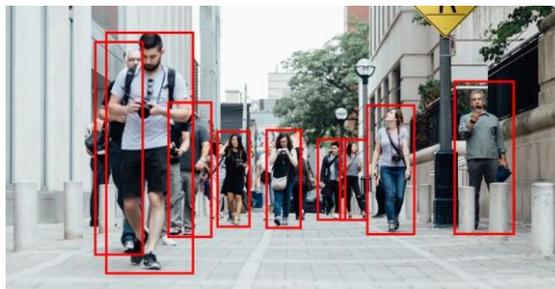


Figure 9: Détection humaine avec les boîtes de délimitations. [22]

Tiny-YOLO est une approche basée sur l'intelligence artificielle en appliquant un seul réseau neuronal à l'image complète. Ce réseau divise l'image en régions et prédit les boîtes de délimitation et les probabilités pour chaque région. Ces boîtes de délimitation sont pondérées par les probabilités prévues, afin de détecter la personne.

2.2.2 La détection de squelette (estimation de pose humaine) :

Il est choisi d'utiliser les données des poses humaines afin de détecter la chute d'une façon indépendante de l'environnement et ces conditions d'éclairage qui se présentent parmi les grands obstacles de détection de chute dans plusieurs travaux de recherche. La dynamique des squelettes du corps humain fournit des informations importantes pour la reconnaissance de l'action humaine, car elles peuvent s'adapter aux circonstances dynamiques et aux antécédents complexes.

L'estimation de pose consiste à localiser les points clés du corps à partir d'images fixes. Comme les points clés du corps sont interconnectés, il est souhaitable de modéliser les relations structurelles entre les points clés du corps pour une meilleure amélioration des performances de localisation. La figure (10) illustre une estimation de pose avec 17 points clés.



Figure 10: Estimation de pose avec 17 point. [22]

La pose de squelette représente l'orientation de l'être humain dans un format graphique. L'estimation de la posture humaine est un ensemble de coordonnées qui peuvent être connectées pour décrire la posture d'une personne. Chaque coordonnée du squelette est

appelée partie (joint ou clé). Une connexion valide entre deux parties est appelée une paire (des membres).

- **L'estimation des poses de plusieurs personnes dans une image :**

L'estimation des poses de plusieurs personnes dans les images a connu des progrès remarquables avec des grandes performances dans les dernières années. L'estimation de la pose humaine est un défi fondamental dans la vision par ordinateur. En pratique, reconnaître la pose de plusieurs personnes est beaucoup plus difficile que de reconnaître la pose d'une seule personne dans une image.

Dans le présent projet, un nouveau cadre régional d'estimation des poses de plusieurs personnes (RMPE) est utilisé, ce qui surperforme les méthodes de pointe pour plusieurs personnes, l'estimation de la posture humaine avec précision et plus d'efficacité.

Cette méthode peut capturer automatiquement les modèles intégrés dans la configuration spatiale des articulations ainsi que leur dynamique temporelle. C'est la force des réseaux neuronaux artificiels (DL). la figure (11) montre le principe de cette méthode.

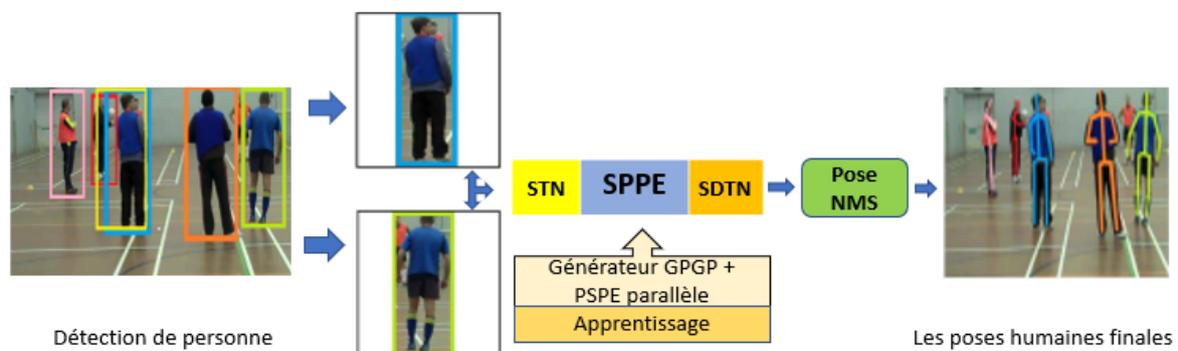


Figure 11: Modèle de détection de pose humaine base sur la méthode RMPE. [25]

Le SSTN (symmetric spatial transformer network) se compose de STN et SDTN (Spatial De-Transformer Network) est une procédure inverse de STN qui sont attachés avant et après la SPPE (single-person pose estimator).

STN reçoit les propositions humaines et le SDTN génère des propositions de pose. Le SPPE en parallèle agit comme un régulateur supplémentaire pendant la phase d'apprentissage.

En suite la pose paramétrique NMS (p-Pose Non-Maximum Suppression) est réalisée pour éliminer les estimations de pose redondantes. Contrairement à l'apprentissage traditionnelle, le module SSTN+SPPE est appreni avec des images générées par le générateur PGPG (pose-guided proposals generator) [25].

2.2.3 La prédiction de mouvement de personne :

La prédiction de l'action humaine est devenue une recherche active ces dernières années, car elle joue un rôle important dans la vidéo, l'action humaine peut être suivie avec plusieurs façons, notamment l'apparence, la profondeur, les flux optiques et les squelettes corporels Parmi ces modalités, la dynamique de squelette humaine qui génère des informations importantes pour l'estimation de l'action humaine.

Dans ce travail, un nouveau modèle de squelettes dynamiques est utilisé. Il est appelé les réseaux convolutifs spatiaux temporelle (ST-GCN). Il conduit à une plus grande puissance expressive avec une grande capacité de généralisation. Ce modèle est conçu pour extraire l'information des articulations. En outre La structure graphique est adaptable en fonction d'apprentissage ce qui fait une amélioration remarquable dans la prédiction de mouvement humain.

Le ST-GCN dépasse les limites des méthodes précédentes en apprenant automatiquement les modèles spatiaux et temporels à partir de données. Cette formulation conduit non seulement à une plus grande capacité d'expression, mais aussi à une plus grande capacité de généralisation. Ce modèle est entraîné sur la base de données, le2i [26]. Il résulte des améliorations substantielles par rapport aux méthodes classiques en termes de prédiction de mouvement [27].

2.4.4 Métriques de précision des modèles de détection d'objets :

De nombreux algorithmes de détection d'objets, tels que Faster R-CNN, MobileNet SSD [28] et YOLO [19] utilisent mAP pour évaluer leurs modèles. Le mAP est également utilisé dans plusieurs références tels que Pascal, VOC, COCO, et encore plus. [29]

La précision moyenne (mAP) ou parfois simplement appelée AP est une métrique utilisée pour mesurer la performance des modèles effectuant des tâches de de détection et de segmentation d'objets. En fait, la précision des prédictions est le pourcentage de validation de ces prédictions. La formule mAP est basée sur les sous-métriques suivantes :

Intersection sur l'Union (IoU) Pour chaque boîte de délimitation, ou le chevauchement est mesuré entre la boîte de délimitation prévue et la boîte de délimitation réel. Comme le montre la figure (12).

En rouge, la prédiction - En vert, la vérité terrain.



Figure 12 : Un exemple de calcul de l'intersection sur les unions pour diverses boîtes englobantes. [30]

Rappel (Recall) détermine la qualité de l'identification de tous les éléments positifs. C'est le pourcentage de vrai positif détecté parmi toutes les vérités de terrain pertinentes (mesure la capacité du modèle à prédire l'ensemble des résultats attendus).

Précision (Precision) Il s'agit du pourcentage de prédictions positives correctes (L'exactitude des prédictions). Un modèle peut avoir une très bonne précision mais avoir un mauvais rappel [22].

Une autre métrique est aussi utilisé pour evaluer les modeles de détecction appelé MOTA (Multiple Object Tracking Accuracy). Il mesure la précision globale du tracker et de la détection. Il traite à la fois la sortie du tracker et la sortie de détection.

2.3 Les techniques utilisées pour détecter et prédire la chute humaine :

Nous avons adopté un modèle de détection et de suivi des chutes humaines qui utilise TinyYOLO oneclass pour détecter chaque personne dans l'image et AlphaPose pour obtenir les postures squelettiques, puis le modèle ST-GCN pour prédire les mouvements de chaque personne détectée.

2.3.1 Tiny-YOLO oneclass :

Le modèle de détection YOLO comporte 24 couches convolutionnelles suivies de 2 couches entièrement connectées. Il utilise des couches convolutionnelles alternées 1×1 pour réduire l'espace des caractéristiques entre les couches. Les couches convolutives sont pré-entraînées sur la base de données ImageNet [31] à la moitié de la résolution (image d'entrée 224×224), puis la résolution est doublée pour l'entraînement à la détection.

Le système applique le modèle sur une image en balayant plusieurs emplacements et échelles. Il utilise une approche totalement différente, il applique un réseau neuronal unique à l'image complète. Ce réseau divise l'image en régions et prédit des cases limitatives et des probabilités pour chaque région. Ces cases limitatives sont pondérées par les probabilités prédites. Les régions de score élevé de l'image sont considérées comme des détections. La figure (13) suivante montre une détection d'objet par le modèle YOLO.

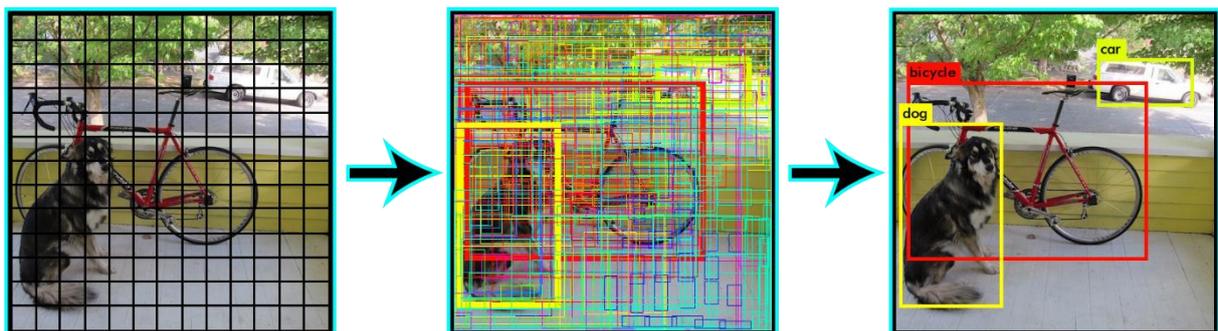


Figure 13: Détection d'objets par le modèle YOLO. [27]

Dans notre projet un nouveau modèle **TinyYOLO** oneclass appelé aussi Tiny-darknet a été utiliser pour détecter chaque personne dans l'image. Le modèle est entraîné sur la base de

données COCO [23] pour une détection de personne plus robuste dans des postures d'angles différents [32].

TinyYOLO est l'un des détecteurs d'objets basés sur l'apprentissage profond les plus rapides en surpassant la précision des autres modèles légers [32]. C'est la version légère de YOLO avec plus de rapidité et moins de calcul. Les résultats de TinyYOLO sont également très similaires à ceux de YOLO. La figure (14) suivante compare les deux résultats.

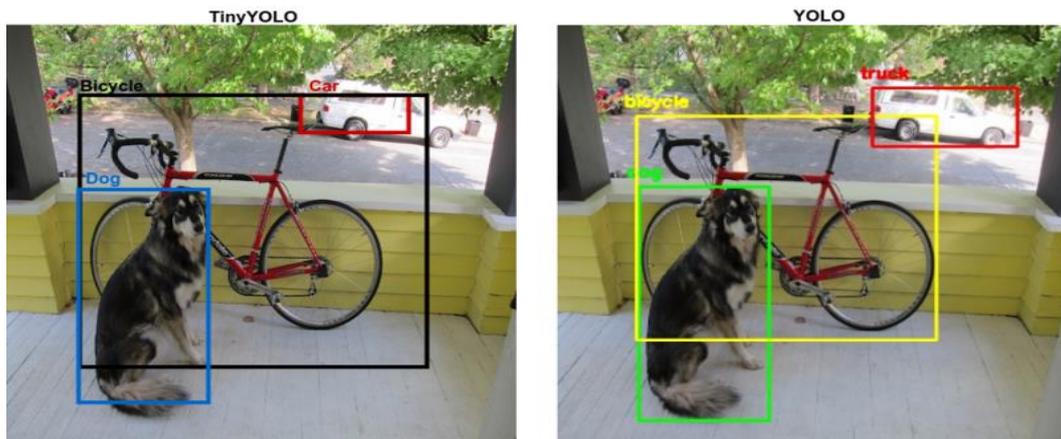


Figure 14: Résultats TinyYOLO vs Résultats YOLO.[32]

Tiny-YOLO fonctionne sur les mêmes principes que YOLO mais avec un nombre réduit de paramètres. Il n'a que 9 couches convolutives, par rapport aux 24 couches de YOLO. La figure (15) illustre l'architecture de TinyYOLO

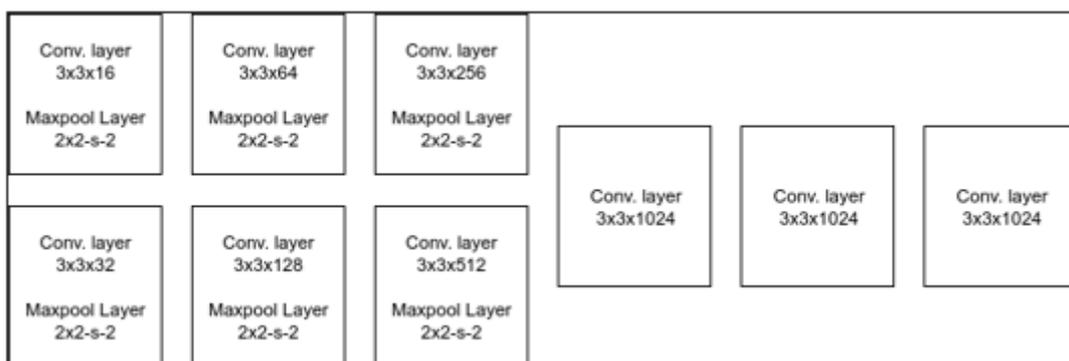


Figure 15: L'architecture de modèle TinyYOLO. [32]

2.3.2 AlphaPose :

AlphaPose est un estimateur de pose multi-personnes précis, qui est le premier système open-source qui atteint plus de 70 mAP (72,3 mAP) sur la base de données COCO et plus de 80 mAP (82,1 mAP) sur la base de données MPII [33]. Pour faire correspondre les poses qui conviennent à la même personne à travers les cadres pour prendre la posture squelettique.

AlphaPose fournit aussi un tracker de pose en ligne efficace appelé Pose Flow [34]. Il est le premier tracker de pose en ligne open-source qui atteint à la fois plus de 60 mAP (66,5 mAP) et plus de 50 MOTA (58,3 MOTA) sur la base de données Pose Track [35].

AlphaPose est un système d'estimation des poses 2D multi-personnes. Le module Symétrique SSTN + SPPE reçoit en entrée les boîtes de délimitation humaine obtenues par le détecteur des personnes. Pour chaque boîte limitative détectée, la pose humaine correspondante sera prédite par SPPE. Les poses humaines redondantes sont éliminées par le NMS de pose paramétrique pour obtenir les poses humaines finales.

Fonctionnalité d'AlphaPose :

La fonctionnalité d'AlphaPose est la détection précise de points clés multiples en temps réel avec une entrée (Image, vidéo, liste d'images) et une sortie (Image de base + affichage/enregistrement de points clés (PNG, JPG, AVI, ...), enregistrement de points clés (JSON), prend en charge plusieurs formats) [36].

La figure (16) présente une estimation de pose avec AlphaPose.



Figure 16 : Estimation de pose multi-personnes avec AlphaPose.[35]

2.3.3 Le modèle ST-GCN :

ST-GCN (Réseaux convolutifs spatiaux graphiques temporels) est un modèle d'apprentissage automatique qui détecte les actions humaines à partir d'informations squelettiques. ST-GCN

est désormais transféré à MMSkeleton [37]. ST-GCN continue à se développer comme une boîte à outils open source flexible pour la compréhension des actions humaines basée sur la détection des squelettes. La prédiction d'action basée sur le squelette et les articulations trajectoires du corps humain sont plus robustes au changement d'éclairage et les variations de scène

Ce modèle est utilisé dans notre projet pour prédire l'action à partir de toutes les 30 images de chaque piste de personne surveillée. Le ST-GCN apprend automatiquement les caractéristiques temporelles et spatiales des données squelettiques et obtiennent des performances remarquables pour la reconnaissance des actions humains. Le modèle prend désormais en charge 7 actions : debout, marcher, s'asseoir, s'allonger, se lever, assis, tomber. C'est une formulation graphique générique pour la modélisation des squelettes dynamiques, qui est le premier qui applique des réseaux de neurones basés sur des graphes pour cette tâche. Ce modèle réalise des améliorations significatives par rapport aux méthodes traditionnelles Par ce qu'il dépasse leurs limites en apprenant automatiquement les modèles spatiaux et temporels des données. Cette formulation conduit non seulement à un plus grand pouvoir d'expression mais aussi à une plus grande capacité de généralisation [38].

La figure (17) montre l'utilisation de ST-GCN pour estimer les actions à partir des graphes temporels spatiales d'une séquence squelettique.

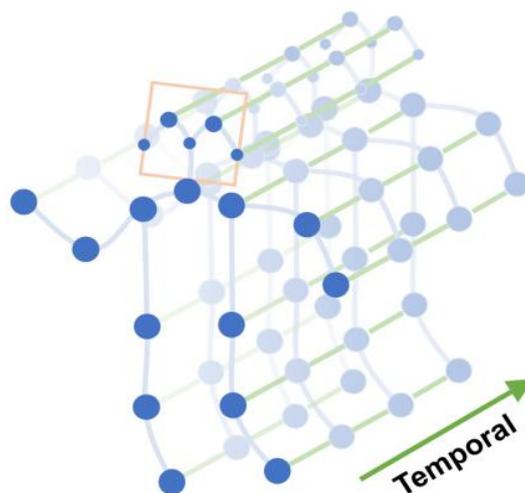


Figure 17: ST-GCN d'une séquence squelettique indiquer les articulations du corps. [38]

ST-GCN permet une vitesse d'entraînement beaucoup plus rapide avec moins de paramètres. Les expériences montrent que le modèle STGCN capte les corrélations spatio-temporelles complètes d'une manière efficace.[38]

L'architecture du réseau de neurone ST-GCN

Le modèle ST-GCN est au noyau de l'algorithme de prédiction, ce modèle apprend les caractéristiques saillantes en utilisant les bords (edges) et les nœuds (nodes).

Dans ce qui suit, nous expliquerons rapidement l'architecture ST-GCN.

Le ST-GCN partage des poids dans différents nœuds afin de maintenir une cohérence dans l'échelle des données en entrée des différentes articulations des corps détectés au préalable. Au cours des expériences menées, les caractéristiques des squelettes introduites dans la couche d'entrée ont été normalisées par lots pour accélérer les tâches de formation/prédiction.

Le modèle ST-GCN est composé de 9 couches dans le ST-GCN (ou unités ST-GCN). Les trois premières couches ont 64 canaux de sortie. Les trois couches suivantes ont 128 canaux de sortie. Et les trois dernières couches ont 256 canaux de sortie ; Ces couches ont 9 tailles de noyau temporel.

Le mécanisme resnet est appliqué à chaque unité ST-GCN. Dans les couches 4ème et 7ème couche de convolution temporelle sont définies sur 2 en tant que couche de regroupement. Après cela, un regroupement global a été effectué sur le tenseur résultant pour obtenir un vecteur de caractéristiques à 256 dimensions pour chaque séquence.

Enfin, ST-GCN a une couche Soft Max pour une prédiction probabiliste, et le modèle est formé avec le gradient descendant stochastique avec un taux d'apprentissage de 0,01 et un apprentissage de 0,1 entraîne sur 10 époques.

Au cours de la phase d'entraînement/validation, plusieurs transformations ont été appliquées sur les images, telles que le changement d'angle, la translation et l'échelle, etc.

2.5 Conclusion :

Dans ce chapitre nous avons présenté un modèle de détection de chute qui utilise : Tiny-YOLO oneclass pour détecter chaque personne et AlphaPose pour estimer la posture squelettique, puis le modèle ST-GCN pour prédire l'action des personnes détectées. Nous avons aussi expliqué le mode de fonctionnement des trois modèles utilisés en cascade pour détecter efficacement les chutes des personnes.

Chapitre 3 :

Méthodologie & Résultats

3.1 Introduction :

Ce chapitre présente la méthodologie construite pour la réalisation de l'algorithme dédié aux trois tâches principales, à savoir : (a)-la détection des corps humains ; (b)-la construction et la détection du squelette humain ; et (c)-la prédiction de chute des corps détectés dans les phases (a) et (b). L'algorithme présenté précédemment dans les chapitres 1 et 2, est un algorithme composite (ou un réseau de neurones composite) qui associe trois modèles neuronaux, qui sont : (1)-Tiny YOLO-Mono Classe (c'est-à-dire que le modèle est testé sur une seule classe d'objet – la classe d'intérêt, à savoir le corps humain);(2)-AlphaPose; et (3)-ST-GCN.

Trois phases ont été réalisées pour la prédiction des chutes de personnes, à savoir les phases de formation/apprentissage, de validation et de test. Il est important de noter que les modèles Tiny YOLO-Mono Class et AlphaPose n'ont pas été formés de bout en bout dans ce travail car cette phase prend du temps et nécessite plusieurs outils d'accélération basés sur GPU pour atteindre une précision supérieure à 90%. Cependant, nous avons téléchargé les poids des modèles préformés à partir du référentiel open source. La mixture des trois modèles susmentionnés est entraînée et évaluée sur des images statiques.

Dans le cadre de ce travail, nous avons testé cette mixture sur des séquences vidéo de la base de données Le2i et aussi sur notre base de données que nous avons collectées avec de camera d'un téléphone mobile, avec une basse résolution et sous différentes conditions environnementales.

3.2 Diagramme de l'Algorithme :

Pour la détection de chute nous avons utilisé trois modèles Tiny YOLO Oneclass, AlphaPose, et ST-GCN. Ces modèles se basent sur une architecture neuronale. La figure (18) illustre un réseau de neurone appliqué sur une séquence vidéo de 25 images par second.

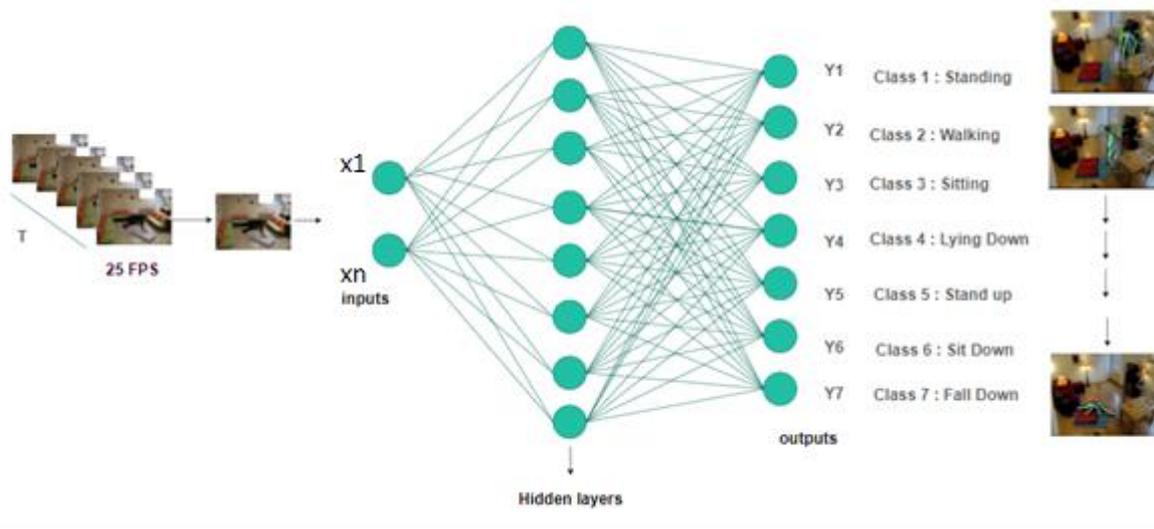


Figure 18 : Réseau de neurones artificiels simplifié du modèle utilisé.

Dans cette illustration le réseau de neurones entièrement connecté comporte trois couches (une couche d'entrée, une couche cachée et une couche de sortie) - la dimension des entrées est équivalente aux caractéristiques extraites (n - caractéristiques) par le réseau de couches convolution. Cependant, nous présentons ici la partie de classification de manière simplifiée, afin de démontrer le nombre de classes en sortie du réseau composite.

Avec ce réseau de neurones, sept classes sont obtenues : debout, marche, assis, couché, se lever, s'asseoir, tomber.

La figure (19) illustre un diagramme des étapes principales qui forment ensemble le model globale adopté pour la détection des chutes humaines.

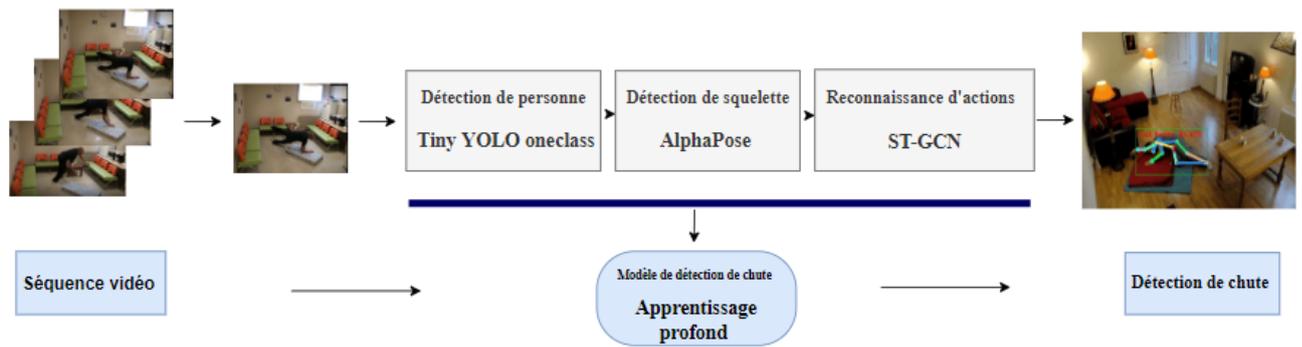


Figure 19: Illustration du digramme globale des modèles utilisés.

Étape 1 : La détection des personnes (TinyYOLO v3) :

La détection de personnes utilise des algorithmes avancés de vision par ordinateur pour identifier les personnes sur une vidéo. Il s'agit de la première phase de notre projet.

L'algorithme d'apprentissage profond utilisé appartient à la famille d'algorithmes YOLO. Cette famille d'algorithmes tire parti de l'utilisation de réseaux neuronaux convolutifs pour la détection d'objets. Ils sont l'un des algorithmes de détection d'objets les plus rapides disponibles et constituent un bon choix pour la détection en temps réel, sans perte de précision [39].

La détection d'objets basée sur le réseau neuronal YOLOv3 (une variante des approches YOLO) qui traite toute l'image instantanément ce qui permet une détection en temps réel sur des vidéos. L'architecture YOLO-v3 est illustrée à la figure (20).

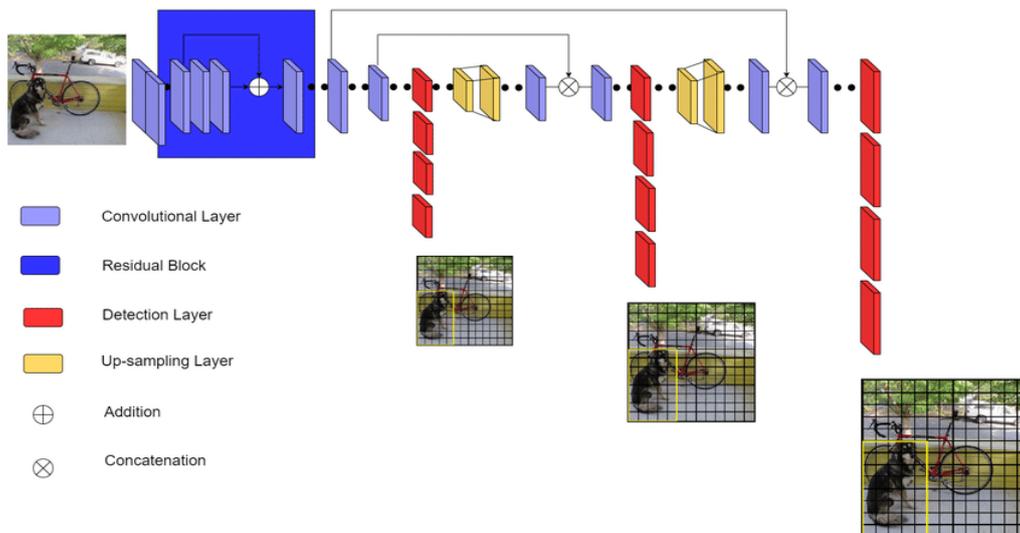


Figure 20 Architecture YOLO v3. [39]

La caractéristique unique de YOLOv3 est qu'il effectue la détection à trois échelles différentes, qui sont précisément données par le sous-échantillonnage des dimensions de l'image d'entrée par 32, 16 et 8 respectivement [40].

L'apprentissage des caractéristiques est effectué par les couches de convolution. Aucune couche entièrement connectée n'est utilisée, ce modèle est donc indépendant de la taille des images.

YOLOv3 est une famille d'architectures et de modèles de détection d'objets pré-entraînés sur le l'ensemble de données COCO. Dans notre projet nous allons utiliser la version Tiny qui se compose de 9 couches seulement et demande moins de ressource de calcul.

Etape 2 : la détection du squelette (AlphaPose) :

Après la détection de personne nous avons la deuxième phase qui est la détection de pose squelettique. AlphaPose est un système d'estimation de pose basé sur l'apprentissage profond. Cette étape est propos de l'implémentation d'AlphaPose sur Pytorch avec des poids pré-entraînés dans l'ensemble de données COCO.

AlphaPose peut facilement estimer la pose de plusieurs personnes en temps réel sur des vidéos de déférentes natures. AlphaPose permet une détection de pose avec un bon fps (frame per second) [15 fps jusqu'à 30 fps] , même avec une gamme moyenne de GPU [41].

L'expérimentation visent à résoudre le problème de l'estimation de la pose de plusieurs personnes dans la nature. Notre méthode adopte l'approche basée sur les graphes pour associer les articulations aux propositions humaines d'une manière globalement idéale. Le problème principal est que la détection légèrement erronée des humains entraîne également une détection erronée des poses.

Le model AlphaPose se compose de trois éléments : Le réseau de transformateurs spatiaux symétriques (SSTN), Pose Paramétrique-Suppression-Non-Maximal (NMS) et le générateur de propositions guidées par la pose (PGPG).

Le système d'apprentissage profond AlphaPose est assez précis. Il donne facilement toutes les estimations de pose correctes dans presque toutes les trames de la vidéo d'entrée. Parfois ce système n'est pas capable d'estimer la pose dans des cas certain de complexité élevée comme montre la figure (21).



Figure 21 Fausse détection des points clés pour les poses de la personne par AlphaPose. [41]

Bien que AlphaPose estime correctement les poses de tous les spectateurs, les estimations de la pose de l'acrobate sont très erronées lors de l'exécution. Cela signifie que le modèle a encore quelques défis qui doivent être surmontés afin d'élargir dans le domaine d'estimation de pose basée sur l'apprentissage profond. Dans notre cas d'étude, les mouvements des personnes sont assez simples pour cela AlphaPose ne présente pas des limitations par rapport à la détection des chutes.

Etape 3 : La reconnaissance des actions (ST-GCN) :

On applique dans cette étape un modèle d'apprentissage automatique ST-GCN pour la reconnaissance des actions squelettiques en utilisant la librairie PyTorch. Ce modèle détecte les actions des personnes à partir des informations squelettiques obtenues à l'étape

précédente. La figure (22) décrit le fonctionnement du modèle pour classifier les actions à base de l'information de la pose détectée.

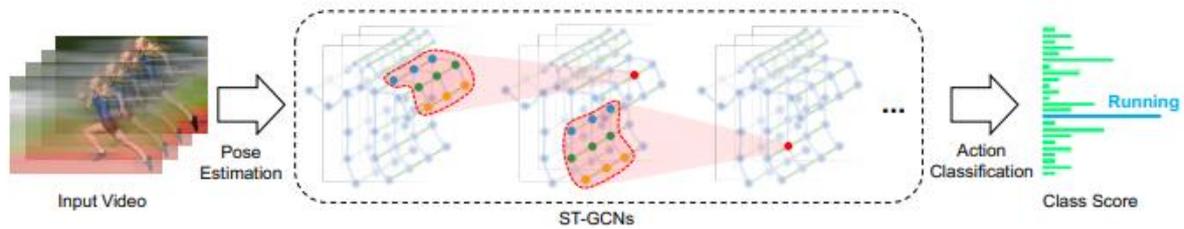


Figure 22 : Modèle ST-GCN appliqué à des estimations de pose pour classifier les actions humaines. [38]

Le ST-GCN est entraîné sur la base de données Le2i. Le modèle est testé en utilisant la même base de données (Le2i) et aussi sur la base de données que nous créer dans le cadre de notre projet de fin d'étude. ST-GCN va être appliquer sur différents scénarios capturés par nous même dans le but est de tester l'efficacité et le rendement du modèle avec des données de différents types. La figure (23) montre l'application de modèle ST-GCN sur une séquence vidéo.

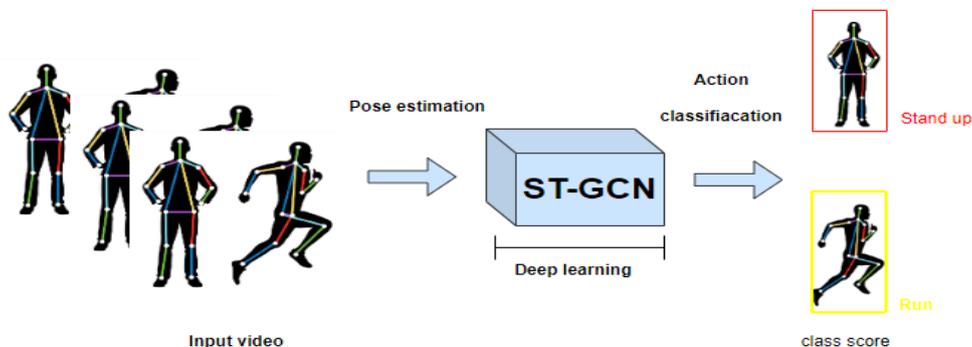


Figure 23 : Application du modèle ST-GCN.

Etape 4 : L'étape de détection de la chute des personnes détectées :

L'étape de détection de chute est la somme des sorties des étapes 1, 2, et 3. Cette phase est considérée comme notre objectif principal que nous voulons atteindre. Après avoir détecté la personne, détecter la pose du squelette, et reconnaître les actions nous arrivons à détecter la l'action instantanée de la personne dont la chute fait partie.

3.3 Challenge et Scenarios de teste :

Afin de valider l'efficacité de notre modèle composite, nous avons testé le modèle sur plusieurs scénarios dont chaque scénario présente des défis différents par rapport aux autres scénarios.

Scenarios de test :

Scenario 1 : Ce scenario traite le problème de rotation de la caméra.

Scenario 2 : Ce scenario traite le problème d'occlusion, où la personne est partiellement cachée par des objets de différentes tailles.



Scenario 1: Camera Rotation



Scenario 2: Occlusion

Scenario 3 : Ce scenario traite le problème de lumière ambiante moindre. Les vidéos de cette catégorie sont capturés dans une chambre à faible lumière.

Scenario 4 : Ce scenario comprend un homme âgé qui simulent une situation de chute. Le visage de cette personne est entièrement caché tout au long de la séquence vidéo.



**Scenario 3:
Lighting Conditions**



**Scenario 4:
Eldery + Occlusion**

3.4 Les bases de données :

3.4.1 La base de données COCO :

L'algorithme comprend trois modèles principaux d'apprentissage profond premièrement, une version simplifiée de Tiny-YOLO v3 one-classe [40] pour la détection des personnes comme une première étape très importante dans le cadre de la détection de chute avec des ressources de calcul modestes et aussi avec une complexité réduite du réseau YOLO. Dans les étapes d'apprentissage et de validation, la base de données utilisé est COCO person keypoints [42] avec une augmentation de données (incluant la rotation des textures, etc.), afin d'apporter beaucoup plus de robustesse au modèle par rapport aux différents changements ambiants, tel que les variations de pose et d'angle.

COCO est un ensemble de données de détection, de segmentation et de sous-titrage d'objets à grande échelle. COCO a plusieurs caractéristiques:

- Segmentation d'objet
- Reconnaissance en contexte

- Segmentation des superpixels
- 330K Images (>200K étiquetées)
- 1,5 million d'instances d'objets
- 80 catégories d'objets
- 91 catégories de trucs
- 5 légendes par image
- **250 000 personnes avec des points clés** (utilisé pour l'entraînement de Tiny-YOLO)

3.4.2 La base de données Le2i

La deuxième base de données utilisé pour l'entraînement du modèle est Le2i [43]. Cette base de données est capturée par une seule caméra avec une résolution de 320×240 pixels. Le2i contient 191 vidéos enregistrés avec une réalisation et un paramétrage de vidéosurveillance réelle. Les vidéos de cette base de données ont été enregistré à partir de différents lieux, permettant de définir plusieurs protocoles d'évaluation (« Maison », « Salle de café », « Bureau » et « Salle de cours »).

Base de données collectée par les étudiantes

Cette base de données fut collectée dans le cadre de cette étude dans le but de mesurer la capacité de généralisation des modèles d'apprentissage profond à détecter, localiser, définir les poses, et reconnaître les actions des personnes.

Cette base de données collectée comprend 22 séquences vidéo, où chacune des séquences est enregistrée avec une résolution de 720×480 pixels et un fps de 25 images par seconde.

Notre base de données se compose de 7 classes qui sont :

- Children
- Elderly
- Multiperson
- Rotation
- Light conditions
- Occlusion
- Fusion of different classes

La collecte de ces données est faite sous différentes conditions (lumière, pose, rotation/angle de capture, occlusion, etc.) afin de défier les performances le modèle composite sous des conditions variables.

Le tableau suivant décrit quelques caractéristiques sur notre base de données.

Tableau 1 Description de la base de données collectées

Résolution	FPS	Référence de caméra	Nombre de personnes	Age	Classes	Nombre de séquence vidéo
720 × 480	30	OPPO A16	7 personnes	5 ans – 73 ans	7 classes	22

3.5 Implémentation

Le langage utilisé dans notre projet fin d'étude est Python 3 [44] avec Library PyTorch [45] et OpenCV [46] dédiés au traitement d'images.

Python est un langage de programmation puissant et facile à apprendre. Il dispose de structures de données de haut niveau et permet une approche simple mais efficace de la programmation orientée objet. PyTorch est une bibliothèque open-source utilisée dans le domaine d'apprentissage automatique et développée à l'aide de la bibliothèque Torch pour le programme python.

OpenCV (Open Source Computer Vision Library) est une bibliothèque logicielle open source de vision par ordinateur et d'apprentissage automatique. OpenCV a été construit pour fournir une infrastructure commune pour les applications de vision par ordinateur et pour accélérer l'utilisation de la perception artificielle.

L'implémentation du modèle a été réalisé sur une machine locale avec les performances décrites dans le tableau suivant :

Tableau 2 Description matériel du PC utilisé

Processeur	Intel(R) Core(TM) i7-10875H CPU @
-------------------	-----------------------------------

	2.30GHz 2.30 GHz
RAM	16.0 GB DDR4
Carte Graphique (GPU)	NVIDIA RTX 2070 super

Les tests sont effectués en principalement sur la machine mentionnée avec une accélération graphique (GPU, Graphics Processing Unit). Puis nous avons testé le modèle sur un CPU (Central Processing Unit) pour pouvoir évaluer l'efficacité de l'accélération graphique sur le temps d'exécution de notre algorithme.

Nous avons utilisé Sublime Text comme un éditeur de texte, qui est un logiciel d'édition de code qui nous permet d'écrire les codes d'une manière simple et efficace. Il est connu pour sa vitesse et sa fluidité. Sublime Text a été construit sur la base de Python et C++.

Mini anaconda est un logiciel open source utilisé dans notre projet comme environnement de programmation et d'exécution. Mini anaconda est une petite version d'Anaconda.

3.6 Résultats et Discussion :

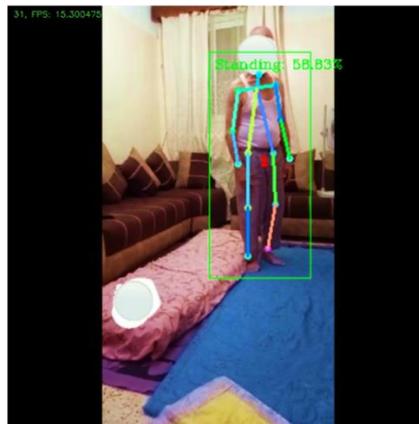
L'un des problèmes rencontrés par George Adaimi et A. Alahi [7] est que les informations squelettiques fournies ne prennent pas en compte le fait que la personne soit face à la caméra ou non. Ce problème a été surmonté par notre modèle adopté, car il détecte la chute indépendamment du positionnement de la personne devant la caméra. Grâce aux capacités des réseaux à base de Graphs (ST-GCN) pour l'extraction et l'apprentissage des caractéristiques spatio-temporelles de la silhouette humaine.

Les illustrations suivantes montrent les résultats de détection sur quelques images tirées à partir des vidéos obtenues après l'application de notre modèle.

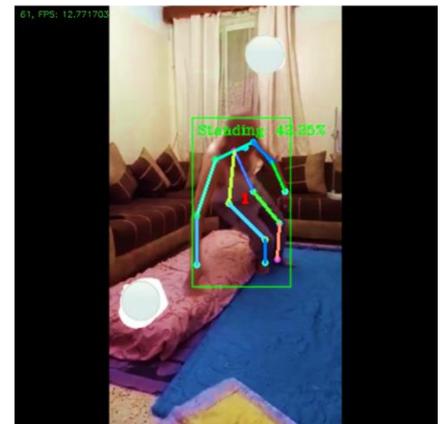
- Vidéo : Old person 1 | Cette vidéo se compose de 170 images



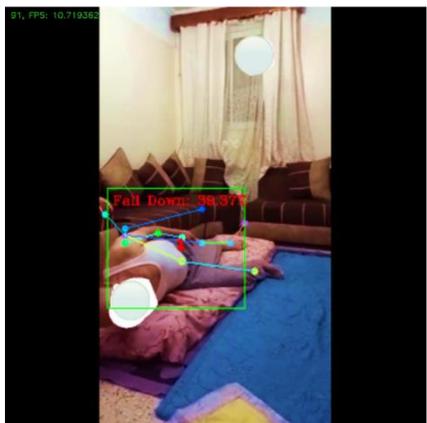
Frame 0



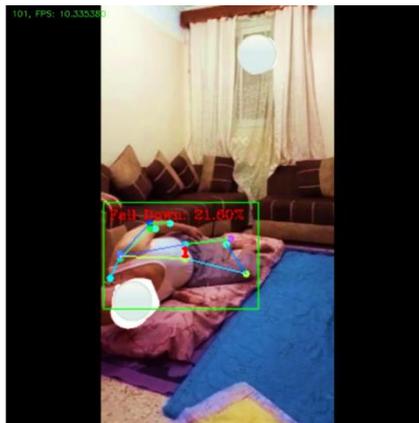
Frame 31



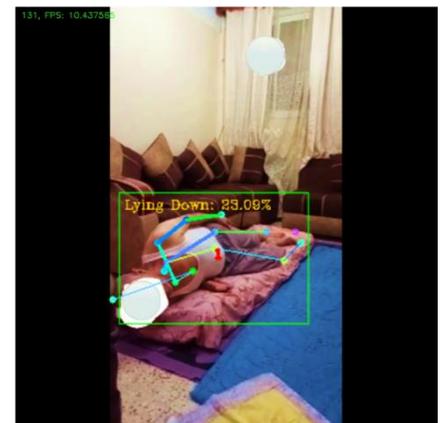
Frame 61



Frame 91



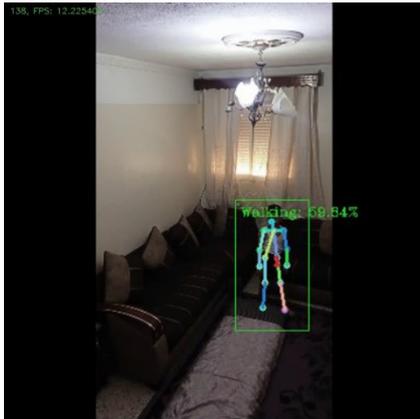
Frame 101



Frame 131

- Remarque : Le modèle a détecté la chute de la personne âgée comme montre l'exemple de la trame 91 et 101, avec un taux de 39.37% et 21.60%, respectivement. Nous remarquons aussi que le modèle ne détecte pas le mouvement de la personne dans la trame 0. Sur la trame 131, on remarque un changement d'état vers 'Lying down'.

- **Vidéo : Light Conditions 2** | Cette vidéo se compose de 295 images.



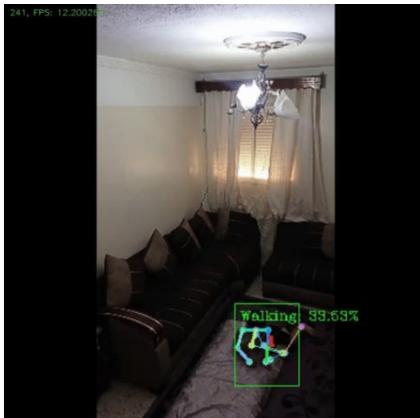
Frame 138



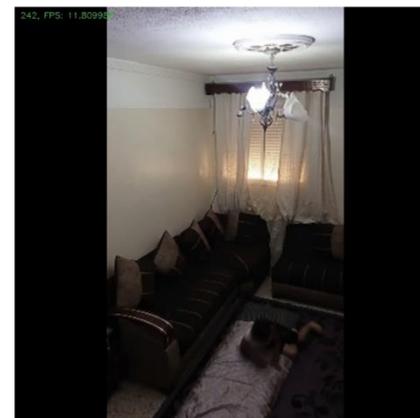
Frame 168



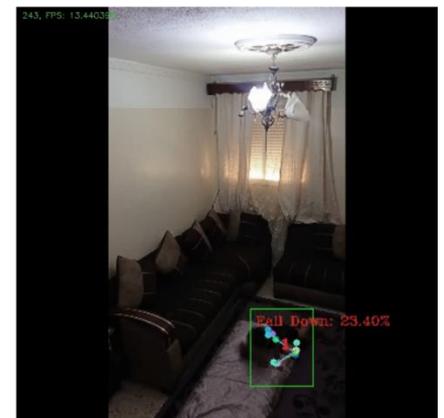
Frame 228



Frame 241



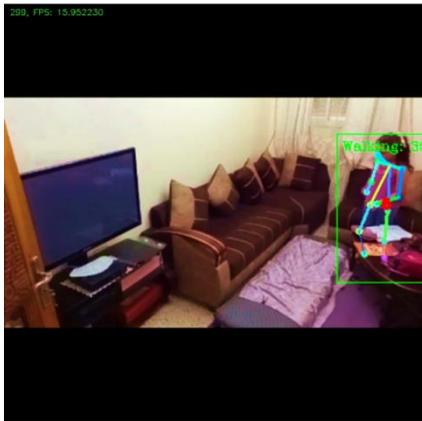
Frame 242



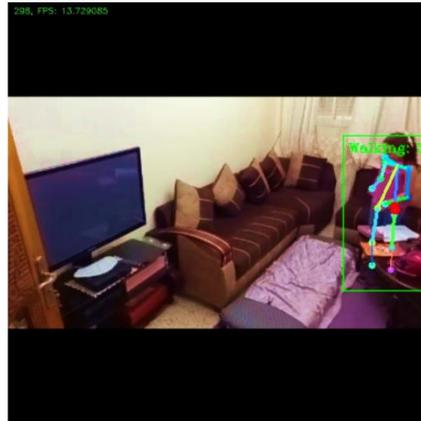
Frame 243

- Remarque : Dans cette séquence vidéo caractérisée par sa faible luminosité, le modèle appliqué détecte généralement le mouvement de la personne (garçon, 5 ans). L'action détecté sur les trames (138, 168, 228) est 'Walking'. On remarque une détection sur la trame 241 où la personne est en état initial de chute mais l'action détectée est toujours 'Walking'. Une absence de détection est notée sur la trame 242. La détection de chute se commence à partir de la trame 243.

- **Vidéo : Occlusion 2** | Cette vidéo se compose de 355 images.



Frame 297



Frame 298



Frame 299



Frame 353



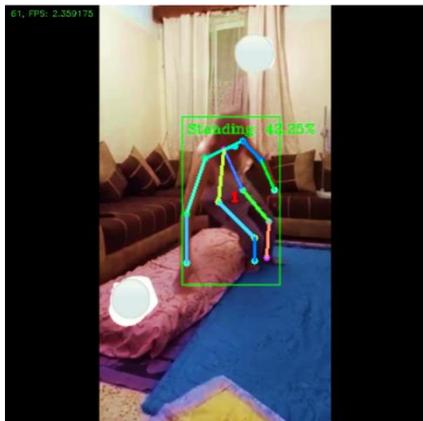
Frame 354



Frame 355

- Remarque : On remarque que le modèle détecte efficacement l'action de la personne (fille, 10 ans) dans chaque trame de la séquence vidéo malgré l'occlusion apparente (table) qui couvre les pieds de de la personne dans les premiers mouvements. Les trames (297 : 299) montrent bien cet exemple.

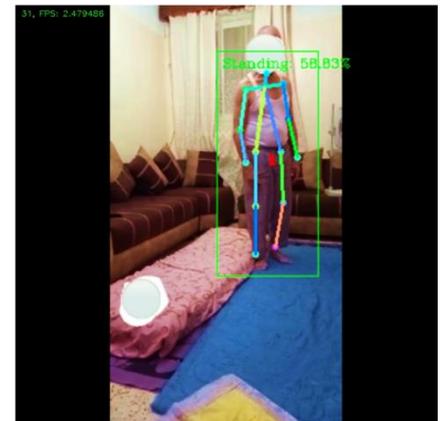
- Vidéo : Old person 1 | CPU



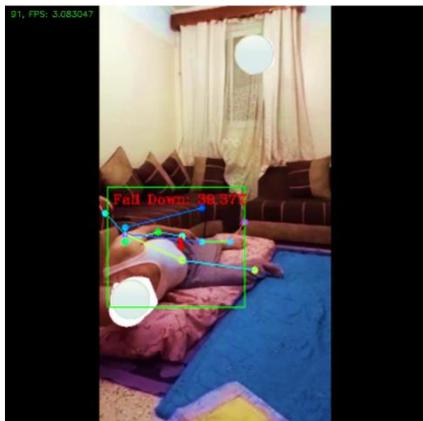
Frame 0



Frame 31



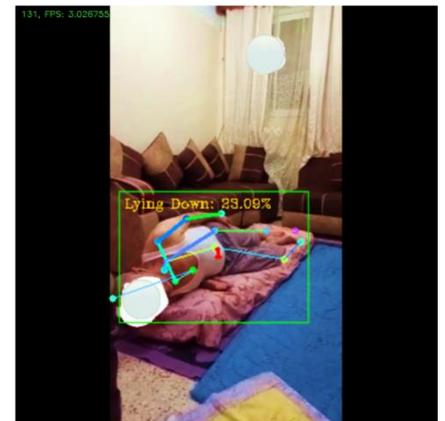
Frame 61



Frame 91



Frame 101



Frame 131

- Remarque : On peut noter clairement que le l'algorithme détecte les mêmes mouvements sur les mêmes trames en utilisant une exécution à base de CPU.

La seule différence est remarquée sur le FPS, où le CPU donne des résultats avec une accélération atténué par rapport à ceux de GPU [47].

- Résultats sur la base de données Le2i

Nous avons téléchargé les résultats de détection de chute, en utilisant la base de données Le2i, sur les liens YouTube suivants :

Vidéo 1 sur Le2i :

<https://youtu.be/K-KXFMEetzM>

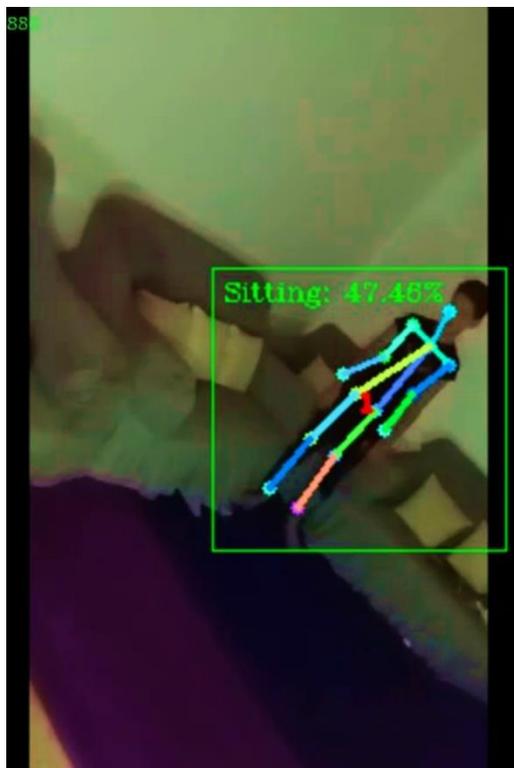
Vidéo 1 sur Le2i :

<https://youtu.be/hAlFlsLrHEw>

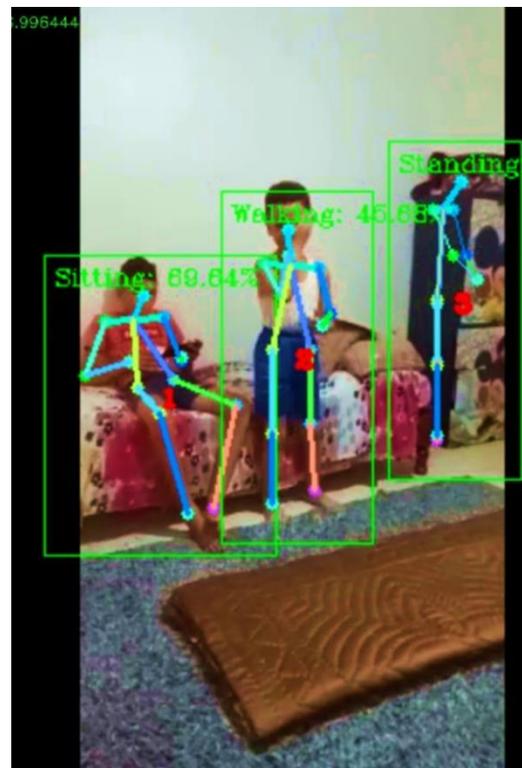
- Exemples des limitations de modèle :



(a)



(b)



(c)

Figure (24) : Exemples de détection erronés

La figure (24) regroupe trois exemples où le modèle n'a pas pu fonctionner correctement dans divers scénarios. Dans la trame (a), le modèle fait une détection erronée. Le personnage sur la télévision a été sujet de détection.

Pour la trame (b), le modèle détecte le mouvement 'Sitting' au lieu de détecter le mouvement 'Standing'. Dans la trame (c), on remarque que le modèle réagit par un faux positif, il détecte la petite armoire comme une personne.

- **Discussion générale :**

D'après les tests réalisés, il est bien remarqué que notre modèle donne des résultats de détection de chute efficace sur presque toutes les séquences vidéo que nous avons utilisé de notre base de données ou bien de la base de données Le2i. Des limitations de performance sont aussi remarquées surtout dans les scénarios qui incluent des défis élevés tels que : Rotation, Multi-personne, TV + personne et environnement à faible luminosité .

Le modèle s'exécute avec une vitesse d'accélération rapide sur le GPU, où la moyenne de FPS est égale à 20. Cependant, la moyenne de FPS sur le CPU tant vers 5 sur tous les vidéos testés.

3.7 Conclusion :

Dans ce dernier chapitre, nous avons présenté et testé un modèle de détection qui se base sur la vision pour la détection des chutes humaines. Nous avons adopté conjointement trois modèles d'apprentissage profond afin d'arriver à la phase de détection de chute.

La base de données que nous avons créée, a été aussi présentée dans ce chapitre. Cette base de données a été utilisé pour tester l'efficacité du modèle sur plusieurs scénarios.

Le modèle a prouvé sa performance dans l'ordre général mai avec quelques limitations notées sur quelques scénarios. Il est recommandé d'enrichir encore plus la base de données et élargir la plage des scénarios.

Conclusion générale et perspectives

Au cours des dernières années, les personnes âgées et vulnérables ont utilisé différents systèmes afin de trouver un mécanisme d'alerte en cas de chute dans leurs domiciles. Parfois la chute rend la personne incapable de bouger ou inconscient, ce problème qui limite plusieurs systèmes portables et les rend loin d'atteindre l'objectif de détection. D'autres systèmes ambiants ont dépassé cet obstacle mais ils sont pratiquement coûteux et sensibles aux changements. Pour surmonter ces limites susmentionnées, nous avons adopté un système intelligent basé sur la vision par ordinateur permettant d'analyser des vidéos captées par une seule caméra qui exploite l'efficacité de l'apprentissage machine dans le but d'améliorer la vie quotidienne des personnes âgées.

Dans la partie théorique, nous avons présenté l'état de l'art sur les systèmes de détection de chutes et leurs méthodes de fonctionnement. Puis dans le chapitre 2, nous avons présenté le fond technique de la méthode choisie pour notre projet de fin d'étude. Il s'agit des 3 étapes principales qui forment l'ensemble de notre modèle. 1) la détection des personnes, 2) la détection de la pose squelettique et 3) la prédiction de mouvement de la/les personnes détectée/s.

Pour la détection de personne, nous avons utilisé le modèle Tiny YOLO one-class préalablement entraîné sur la base de données COCO. Tiny YOLO a démontré une haute performance pour la détection des personnes dans les séquences vidéo utilisées. Pour la deuxième étape nous avons utilisé le modèle AlphaPose pour l'estimation de pose des personnes détectées dans la première étape. Le modèle ST-GCN est utilisé dans la troisième étape pour la prédiction des mouvements à partir des points clés des poses détectés. Le résultat de combinaison de ces trois modèles permet à la fin de détecter la personne, tracer sa pose squelettique, et prédire son mouvement instantané parmi 7 classes prédéfinies dont la chute fait partie.

Une base de données a été créée dans le cadre de ce projet de fin d'étude. Cette base de données est composée de 22 séquences vidéo divisées sur 7 classes différentes. Notre modèle composite a prouvé une efficacité pertinente sur l'ensemble des scénarios réalisés. Nous avons aussi testé le modèle sur la base de données Le2i qui a été déjà utilisée pour l'entraînement de

la troisième phase qui se base sur le modèle ST-GCN. Les tests ont été réalisés sur un PC personnel en utilisant deux unités de traitement de données (GPU & CPU). Le modèle de détection de chute a montré plus d'efficacité en utilisant le GPU par rapport au temps d'exécution et l'accélération de traitement de vidéo en temps réel.

Cette solution adoptée est classée moins coûteuse tout en assurant des résultats optimaux par rapport aux solutions existantes dans la littérature. C'est le modèle composite à base d'apprentissage profond qui est derrière la puissance de cette solution.

Il faut aussi noter que notre modèle présente des limitations dans quelques scénarios testés.

Afin d'augmenter l'efficacité de notre système, nous proposons de l'entraîner de plus en plus sur d'autres bases de données qui incluent divers scénarios.

Comme un futur travail, nous visons d'enrichir et organiser de davantage la base de données collectée avec d'autres séquences vidéo et la mettre plus tard à la disposition des chercheurs qui visent à optimiser les systèmes de détection des chutes humaines. Une implémentation de sur des systèmes embarqués à faible coût énergétique fait partie des perspectives visées.

Finalement, nous recommandons de mettre notre système dans un environnement IoT qui assure une communication instantanée et fiable afin de lancer des alertes instantanées au service des urgences médicales ainsi que les relatives de la personne qui a subi la chute.

Références :

- [1] United Nations. World Population Ageing; United Nations: New York, NY, USA, 2015.
- [2] WANG, Zhuo, RAMAMOORTHY, Vignesh, GAL, Udi, *et al.* Possible life saver: A review on human fall detection technology. *Robotics*, 2020, vol. 9, no 3, p. 55.
- [3] WANG, Xueyi, ELLUL, Joshua, *et* AZZOPARDI, George. Elderly fall detection systems: A literature survey. *Frontiers in Robotics and AI*, 2020, vol. 7, p. 71.
- [4] USMANI, Sara, SABOOR, Abdul, HARIS, Muhammad, *et al.* Latest research trends in fall detection and prevention using machine learning: A systematic review. *Sensors*, 2021, vol. 21, no 15, p. 5134.
- [5] TSINGANOS, Panagiotis *et* SKODRAS, Athanassios. On the comparison of wearable sensor data fusion to a single sensor machine learning technique in fall detection. *Sensors*, 2018, vol. 18, no 2, p. 592.
- [6] SPOSARO, Frank *et* TYSON, Gary. iFall: an Android application for fall monitoring and response. In : *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2009. p. 6119-6122.
- [7] ADAIMI, George *et* ALAHI, Alexandre. Learning nuisances to track pedestrians in autonomous vehicles. In : *18th Swiss Transport Research Conference*. 2018.
- [8] MARDINI, Mamoun T., IRAQI, Youssef, *et* AGOULMINE, Nazim. A survey of healthcare monitoring systems for chronically ill patients and elderly. *Journal of Medical Systems*, 2019, vol. 43, no 3, p. 1-21.
- [9] ALEMDAR, Hande *et* ERSOY, Cem. Wireless sensor networks for healthcare: A survey. *Computer networks*, 2010, vol. 54, no 15, p. 2688-2710.
- [10] SELIM, Bassant, IRAQI, Youssef, *et* CHOI, Ho-Jin. A multi-sensor surveillance system for elderly care. In : *2013 IEEE 15th International Conference on e-Health Networking, Applications and Services (Healthcom 2013)*. IEEE, 2013. p. 502-506.
- [11] OMS Vieillesse et santé. En ligne : <http://www.who.int/mediacentre/factsheets/fs404/fr/> Consulté Janvier 2022.
- [12] OMS personnes âgées. En ligne : https://www.who.int/health-topics/ageing#tab=tab_1. Consulté Janvier 2022.
- [13] DARGENT-MOLINA, P. *et* BREART, G. Epidemiology of falls and fall-related injuries in the aged. *Revue D'épidémiologie et de Santé Publique*, 1995, vol. 43, no 1, p. 72-83.
- [14] BLOOM, David E., BOERSCH-SUPAN, Axel, MCGEE, Patrick, *et al.* Population aging: facts, challenges, and responses. *Benefits and compensation International*, 2011, vol. 41, no 1, p. 22.

- [15] BEAUCHET, Olivier, ANNWEILER, Cédric, HUREAUX-HUYNH, Raphaëlle, *et al.* Médicaments et chute de la personne âgée. *Annales de Gérontologie*, 2008, vol. 1, no 1, p. 47-52.
- [16] GUK, K., HAN, G., LIM, J., *et al.* Evolution of wearable devices with real-time disease monitoring for personalized health care. *Nanomaterials (Basel)* 2019 May 29 ; 9 (6) doi: 10.3390/nano9060813.
- [17] ABBATE, Stefano, AVVENUTI, Marco, BONATESTA, Francesco, *et al.* A smartphone-based fall detection system. *Pervasive and Mobile Computing*, 2012, vol. 8, no 6, p. 883-899.
- [18] PATEL, Shyamal, PARK, Hyung, BONATO, Paolo, *et al.* A review of wearable sensors and systems with application in rehabilitation. *Journal of neuroengineering and rehabilitation*, 2012, vol. 9, no 1, p. 1-17.
- [19] YOLO: Real-Time Object Detection. En ligne ; <https://pjreddie.com/darknet/yolo/> Consulté Mai 2022.
- [20] GIRSHICK, Ross, DONAHUE, Jeff, DARRELL, Trevor, *et al.* Rich feature hierarchies for accurate object detection and semantic segmentation. In : *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014. p. 580-587.
- [21] GIRSHICK, Ross. Fast r-cnn. In : *Proceedings of the IEEE international conference on computer vision*. 2015. p. 1440-1448.
- [22] Technical Metrics Required to Assess of Human/People Detection. En ligne: <https://visailabs.com/technical-metrics-required-to-asses-of-people-detection-solution/>. Consulté Juin 2022.
- [23] LIN, Tsung-Yi, MAIRE, Michael, BELONGIE, Serge, *et al.* Microsoft coco : Common objects in context. In : *European conference on computer vision*. Springer, Cham, 2014. p. 740-755.
- [24] GARAU, Nicola, MARTINELLI, Giulia, BRÓDKA, Piotr, *et al.* PanopTOP: a framework for generating viewpoint-invariant human pose estimation datasets. In : *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021. p. 234-242
- [25] FANG, Hao-Shu, XIE, Shuqin, TAI, Yu-Wing, *et al.* Rmpe: Regional multi-person pose estimation. In : *Proceedings of the IEEE international conference on computer vision*. 2017. p. 2334-2343.
- [26] The Fall Detection Dataset. En ligne : <http://le2i.cnrs.fr/Fall-detection-Dataset?lang=fr>. Consulté Mai 2022.
- [27] YOLO Real-Time Object Detection disponible en ligne : <https://pjreddie.com/darknet/yolov2/> Consulté le 03 06 2020.
- [28] COHEN, Julia, CRISPIM-JUNIOR, Carlos, CHIAPPA, Jean-Marc, *et al.* MobileNet SSD : étude d'un détecteur d'objets embarquable entraîné sans images réelles. In : *ORASIS 2021*. 2021.
- [29] Evaluating multiple object tracking accuracy and performance metrics in a real-time setting. En ligne : <https://visailabs.com/evaluating-multiple-object-tracking-accuracy-and-performance-metrics-in-a-real-time-setting/> Consulté Juin 2022.
- [30] Extraction d'objets pour la cartographie par deep-learning : évaluation du modèle. En ligne : <https://makina-corpus.com/data-science/extraction-objets-cartographie-deep-learning> Consulté Juin 2022
- [31] ImageNet. En ligne : <https://www.image-net.org/> Consulté Juin 2022.

- [32] TinyYOLOv2 Architecture. En ligne : https://developer.ridgerun.com/wiki/index.php?title=GstInference/Supported_architectures/TinyYoloV2. Consulté juin 2022.
- [33] MPII Human Pose Dataset. En ligne : <http://human-pose.mpi-inf.mpg.de/> Consulté Juin 2022.
- [34] XIU, Yuliang, LI, Jiefeng, WANG, Haoyu, *et al.* Pose Flow: Efficient online pose tracking. *arXiv preprint arXiv:1802.00977*, 2018.
- [35] AlphaPose. En ligne : <https://github.com/Amanbhandula/AlphaPose> . Consulté mai 2022.
- [36] CHEN, Qilai, ZHANG, Ying, LIU, Shuzhi, *et al.* Switchable perovskite photovoltaic sensors for bioinspired adaptive machine vision. *Advanced Intelligent Systems*, 2020, vol. 2, no 9, p. 2000122.
- [37] MMSkeleton. En ligne : <https://github.com/open-mmlab/mmskeleton>. Consulté juin 2022.
- [38] YAN, Sijie, XIONG, Yuanjun, et LIN, Dahua. Spatial temporal graph convolutional networks for skeleton-based action recognition. In : *Thirty-second AAAI conference on artificial intelligence*. 2018.
- [39] VALDEZ, Paolo. Apple defect detection using deep learning based object detection for better post harvest handling. *arXiv preprint arXiv:2005.06089*, 2020.
- [40] YOLOv3 object detection. En ligne : <https://towardsdatascience.com/yolo-v3-object-detection-53fb7d3bfe6b>. Consulté Juin 2022.
- [41] Real-Time Pose Estimation using AlphaPose, PyTorch, and Deep Learning. En ligne : <https://debuggercafe.com/real-time-pose-estimation-using-alphapose-pytorch-and-deep-learning/>. Consulté Juin 2022.
- [42] COCO Dataset. <https://cocodataset.org/#home>. Consulté Avril 2022.
- [43] Le2i Dataset. <http://le2i.cnrs.fr/Fall-detection-Dataset>. Consulté Avril 2022.
- [44] Python. En ligne : <https://www.python.org/>. Consulté Mai 2022.
- [45] PyTorch. En ligne : <https://pytorch.org/>. Consulté Mai 2022.
- [46] Open CV. En ligne : <https://opencv.org/>. Consulté Mai 2022.
- [47] Comparaison entre le CPU et le GPU. En ligne : <https://www.icphs2015.info/principales-differences-entre-les-cpu-et-les-gpu/>. Consulté Juin 2022.

