ÉCOLE POLYTECHNIQUE DE L'UNIVERSITÉ DE NANTES
DÉPARTEMENT D'INFORMATIQUE

RAPPORT DE RECHERCHE ET DÉVELOPPEMENT

Analyse visuelle de la satisfaction client en magasin

Suivi de personnes dans une vidéo

Élodie BOUILLETEAU

2 décembre 2018

encadré par Nicolas NORMAND

coordinateur : Stéphane GAUDIN

POLYTECH NANTES

UNIVERSITÉ DE NANTES

Analyse visuelle de la satisfaction client en magasin

Suivi de personnes dans une vidéo

Élodie BOUILLETEAU

Résumé

Le sujet traité dans ce rapport est l'analyse de la satisfaction client à l'aide d'une caméra placé devant un stand en magasin. La problématique découlant de ce sujet est la détermination et le suivi de personnes dans une vidéo. La problématique a été volontairement simplifié à un domaine de recherche plus précis puisque le sujet englobe différents domaines comme la détection de visage ou encore la détection d'émotion.
Les objectifs qui ont été fixés sont les suivants :

- Déterminer les indicateurs pertinents d'analyse de la satisfaction
- Déterminer les critères d'évaluations des méthodes
- Trouver des méthodes de suivi
- Création de la vidéo qui soit conforme au contexte attendu
- Évaluer les méthodes de suivi sur la vidéo avec les critères d'évaluations définie plus tôt
- Récupérer les résultats de la méthode et calculé les indicateurs pertinents pour la satisfaction client (ex : nombre de temps moyen d'un client passé devant la caméra)
- Faire une démonstration.

Pour répondre à la problématique, les recherches se sont concentrées sur les méthodes / techniques de détection de personnes dans une image et de suivi de personnes dans une vidéo. Les méthodes abordées comptent parmi les plus performantes ou les plus récentes, mais ne possède pas de code source réutilisable. Nous recherchons avant tout une méthode dont le code source est réutilisable afin de faire nos propres évaluations via cette méthode.
Suite à la recherche de documentations sur le domaine du suivi de personnes dans une vidéo, nous présentons 3 méthodes dont le code source est réutilisable et regardons les avantages et les désavantages de chacune des méthodes.

Classification : Computing methodologies/Artificial intelligence/Computer vision/Computer vision tasks/Activity recognition and understanding Mots clés : vidéo, suivi de personne, comparaison de méthode de suivi, convolution...

Remerciements

Je tiens à remercier les personnes qui m'ont aidé tout au long de ce projet de recherche et de développement ainsi que les personnes qui m'ont conseillés pour la rédaction de ce rapport.

Tout d'abord, j'adresse mes remerciements à mon professeur et superviseur Nicolas NORMAND qui m'a conseillé et guidé pour la prise de décision et les recherches sur la problématique : suivi de personnes dans une vidéo.

Je tiens à remercier mon maître de contrat professionnel et coordinateur, Stéphane GAUDIN, pour ces nombreux conseils notamment de rédaction, et pour son soutien à mon égard.

Table des matières

Préambule

L'objet du rapport est de trouver une solution répondant à la problématique de suivi de personnes dans une vidéo, qui s'inscrit dans un sujet plus large de la satisfaction de la clientèle.

Cette solution pourra servir de base pour de future recherche comme sur la détection de visage ou d'émotion qui s'inscrive dans le sujet de la satisfaction client sur un stand en magasin.

Étant en contrat professionnel pour la dernière année d'étude à Polytech Nantes, je pouvais proposer un sujet venant de l'entreprise. Le sujet a été proposé par mon maître de contrat professionnel qui utilise l'opportunité du projet de recherche et développement comme un moyen d'analyser les possibilités de réponses déjà existante dans le domaine de suivi d'une personne dans une vidéo.

Ce sujet est une découverte pour moi, puisque je ne connaissais pas du tout les technologies de détection et de suivi personnes dans une vidéo. De plus, les réseaux de neurones de convolution qui ont révolutionné ce domaine m'étaient inconnus.

1

Introduction

L'introduction générale traite de la problématique du sujet qui va être développée dans ce rapport, les objectifs plus précis que l'on s'est fixés, le travail qui a été réalisé et les contributions associés. La dernière section de l'introduction détaillera l'organisation logique du rapport et présentera ses différents chapitres.

1.1 Présentation de la problématique

L'analyse de la satisfaction client en magasin s'inscrit dans un contexte plus large du groupement Système U. Pour un contexte économique tendu, le groupement Système U ambitionne de gagner des parts de marchés pour atteindre 12.5% en 2022.

Selon une étude Gartner, 64 % des consommateurs considèrent désormais l'expérience en magasin comme étant plus importante que le prix du choix de l'enseigne.

Étudié l'analyse de la satisfaction client en magasin correspond dans ce contexte à simuler un stand d'animation en magasin, qui serait filmé afin de mesurer des indicateurs sur la satisfaction client. La matérialisation de ce stand en magasin peut être réalisée avec l'installation d'une caméra sur une table. La caméra filme le passage de personnes devant la table.

La problématique étudiée dans ce rapport est l'analyse de la satisfaction client dans une vidéo où l'on aperçoit le passage de client devant un stand d'un magasin. Cette problématique soulève de nombreuses réflexions sur les méthodes de détection que nous pouvons utiliser.

La problématique donnée est trop vaste pour un sujet de projet de recherche et développement de 5 ème année en informatique. Avec mon superviseur et mon coordinateur, nous avons décidés de restreindre la problématique à la détection et au suivi de personnes dans une vidéo.

Les principaux indicateurs pertinents que nous pouvons obtenir avec la détection et le suivi de personnes dans une vidéo sont :

- Le nombre de personnes différentes vu dans la vidéo,
- L'arrêt de la personne dans la vidéo,
- Le temps moyen de passage d'une personne dans la vidéo,
- Le rapprochement de la personne par rapport à l'emplacement du stand.

Ces indicateurs seuls ne permettent pas de déterminer une valeur de satisfaction client pertinente. Cependant, si on associe ces indicateurs à un questionnaire d'analyse sur la satisfaction client, il est possible d'y ajouter des informations complémentaires pouvant se révéler utile. De plus, à l'avenir, on peut imaginer de nouveaux modules complémentaires qui viendront enrichir l'existent issu de cette recherche basé sur le suivi de personnes. Ces modules pourront être :

- La détection de visage,
- La détection d'émotion,
- La détection de caractéristique d'une personne (âge, genre...).

Nous pouvons découper la problématique de détection et suivi de personnes en plusieurs sous-problèmes. Le premier problème qui se pose est l'analyse de vidéo image par image. Comment détecter et reconnaître une personne ? Il s'agit du suivi de personne. Le deuxième problème soulevé est la mesure des indicateurs pertinents. Comment peut-on mesurer ces indicateurs avec les résultats en sortie d'un algorithme de suivi de personnes ?

1.2 Contexte

Cette partie me permet de présenter en détails le contexte dans lequel évolue le projet.

Le contexte du projet est d'analyser une vidéo d ?un stand prise dans un magasin. Deux questions se posent : Où le stand doit-il être placé dans le magasin pour que le suivi de personnes sur la vidéo soit optimal ? Comment la vidéo d'évaluation doit-être prise afin de correspondre au contexte ?

Pour répondre à ces deux questions, nous émettons des hypothèses sur le placement du stand. La première hypothèse est que le stand soit placer dans un rayon du magasin où il y a le plus de passants, ensuite la deuxième hypothèse consiste à se dire que la caméra ne doit observer seulement les passants qui passent devant le stand. De plus, il faut éviter d ?avoir un fond arrière où l'on verrait des personnes d'un autre stand ou d'une file d'attente, cela pourrait nuire à la détection de personnes.

La vidéo d'évaluation doit être prise avec une caméra classique fixe à l'intérieur d'un bâtiment et éclairer d'une lumière artificielle afin de correspondre au mieux au contexte du sujet.

1.3 Objectifs poursuivis

Cette partie permet de définir les objectifs liés à la problématique. Nous verrons à la fin s'ils ont été atteints.

L'objectif principal poursuivi est de déterminer les critères d'évaluations des propositions trouvées. Une fois, les critères déterminés, notre prochain objectif est de trouvé un algorithme/une méthode de suivi de personnes.

Une fois les méthodes récupérées et implémentées, notre objectif est de les évaluer sur une vidéo qui tient compte du contexte. La vidéo devra être créée au préalable et devra être filmée selon des critères bien précis définis dans la partie contexte.

Nous prendrons la méthode ayant les meilleurs résultats de performance.

A l'aide des résultats de cette méthode, on pourra calculer les indicateurs de la satisfaction client. Par exemple, on peut imaginer avoir en sortie un tableau avec en ligne les identifiants des clients qui sont passés devant la caméra et en colonne le nombre d ?images où l'on voit le client sur la vidéo. On peut calculé le temps moyen de passage devant la caméra par personne ainsi que le nombre de personnes différentes qui sont passées devant celle-ci avec le nombre de passages par personnes et la durée de la vidéo.

Le dernier objectif est de faire une démonstration des résultats de l'analyse de la vidéo.

1.4 Travail réalisé

Dans cette partie, je vais expliquer le travail que j'ai réalisé tout au long de ce projet.

Tout d'abord, mon premier travail a été la compréhension du sujet, ce qui a abouti à un réajustement du sujet afin de trouver une problématique adaptée.

Ensuite, après avoir déterminé de manière plus précise le sujet, j'ai recherché des articles sur la détection et le suivi de personnes dans une vidéo afin de pouvoir répondre à la problématique. Cette recherche a abouti sur la présentation des grandes méthodes de détection et de suivi utilisées et créées par les chercheurs.

Une fois les recherches finies, j'ai regardé l'existence de 3 propositions faites par des chercheurs ou étudiants sur le suivi de personnes dans une vidéo. J'ai analysé les propositions en les comparant entre elles avant de pouvoir les tester afin de savoir si ces méthodes sont performantes sur une vidéo tenant compte du contexte du sujet.

Le but que je poursuis pour la suite est d'évaluer les 3 propositions sur la vidéo et de calculer des indicateurs pertinents de la satisfaction client.

1.5 Contribution

Pour l'instant, l'évaluation n'ayant pas encore été effectuée, je ne peux montrer de résultats.

1.6 Plan de l'étude

Dans cette partie, je vais vous présenter le plan d'étude du rapport. Pour la première partie du projet de recherche et développement, le plan se limite à l'état de l'art et les propositions issus de l'état de l'art.

Le chapitre 2 étudie un ensemble de propositions de la littérature scientifique qui porte sur la problématique de suivi de personnes dans une vidéo. L'analyse conjointe de ces dernières permet de dresser un bilan de l'état de l'art et de proposer des pistes de recherches.

La conclusion permet de synthétiser les apports de ce travail et d'ouvrir des voies d'investigations supplémentaires.

2

État de l'art

L'objectif dans cette partie est d'étudier les connaissances liées à la problématique de détection et de suivi d'individus dans une vidéo. Le suivi de personnes est lié à la détection et la reconnaissance d'un individu d'une image à l ?autre de la vidéo.

A notre connaissance, les études menées sur la **détection de personnes** appartenant au domaine de la vision par ordinateur, ont commencées à la fin des années 1990. La détection de personnes est une méthode spécifique au domaine de la détection d'objets. Elle consiste à détecter la présence d'un humain et sa localisation dans une image numérique. En règle générale, la détection de personnes se fait en détectant des humains posant debout ou en train de marcher.

La variabilité d'apparence des êtres humains, l'articulation du corps humain, l'occultation par des objets ou par d'autres humains sont des éléments qui rendent la détection de personnes très difficile. De plus, les données d'entrées (images et vidéos) sont rarement de bonne qualité, ce qui rajoute encore une difficulté. Le problème que pose cette méthode de détection est de trouver une représentation des humains ni trop générique mais assez discriminante pour n'observer que des humains et ne pas confondre les humains de l'image avec d'autres objets.

L'une des premières méthodes proposées est la détection d'objets en utilisant une transformée de Hough sur plusieurs images qui sont prises de différents points de vue. La transformée de Hough est une technique utilisée lors du pré-traitement d'images numérique pour reconnaître les formes. Cette méthode permet de détecter des piétons, mais n'est pas exclusive à celui-ci. En 1998, [HW98] Heisele et Wöhler conçoivent une méthode de détection et de classification à l'aide des mouvements des jambes des piétons par rapport au sol. Ces méthodes sont cependant spécifiques et ne sont pas génériques.

A partie des années 2000, les méthodes de détection de personnes ont évoluées avec l'arrivée de la méthode de Viola-Jones [VJ01], qui est étendu en 2005 à la détection de personne en utilisant le mouvement. Cette technique détecte les objets visuels rapidement et possède un taux

de précision élevé.

Lorsque les réseaux de neurones de convolution ont été découverts, cela à permis une avancée fulgurante dans le domaine de la reconnaissance d'images et de vidéos. Les réseaux de neurones de convolution sont performants et rapides, c'est pourquoi les dernières méthodes de détection utilisent ces réseaux. Depuis l'utilisation de ces réseaux de neurones dans le domaine de la détection d'objets en 2013, la moyenne de précision a augmentée. Elle est supérieure à 50 %.

Le réseau de neurones de convolution est un réseau inspiré du cerveau humain et notamment du lien entre les yeux et le cerveau. Tout humain bien constitué possède des yeux et un cerveau. Lorsqu'une personne regarde une image, elle sait reconnaître un objet dans cette image. Si la personne regarde une vidéo, elle peut suivre du regard un objet en particulier car elle sait le reconnaître image par image et associer les objets de chaque image entre elles. Il s'agit de quelque chose d'instinctif, de naturel chez l'homme.

Le réseau neuronal convolutif fonctionne sur le même principe. L'objectif est d'apprendre à reconnaître un objet dans une image. Pour cela, le réseau contient plusieurs couches de neurones.

LeNet [Haf98] est le premier petit réseau neuronal convolutif fonctionnel. Il est utilisé pour reconnaitre des chiffres écrit à la main. LeNet a été créé en 1990 par LeCun et al. La structure du réseau neuronal convolutif de LeNet est composée de 5 couches : 2 fois une couche de convolution suivie d'une couche de regroupement et d'une couche entièrement connectée.

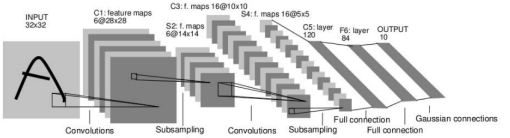

FIGURE 2.1 : Architecture de LeNet (Source : « *Graphics in Gradient-based learning applied to document recognition* », page 7)

Les couches entièrement connectées sont constituées de poids. A une paire de valeurs d'entrée et de sortie, on associe un poids. Il faut savoir que le nombre de paramètres augmente rapidement.

Les couches convolutives consistent en un ensemble de filtres qui capture une certaine structure dans l'image. Une fois qu'un filtre a appris à capturer une certaine structure, il est capable de la trouver n'importe où dans l'image, car le filtre sera appliqué à toutes les positions, pixels de l'image.

Les couches de regroupement permettent de réduire la taille des données. Le regroupement est effectué pour chaque valeur d'entrée. Le regroupement maximum, par exemple, récupère la valeur maximum pour une valeur d'entrée.

Pour le réseau neuronal convolutif LeNet, la première couche de convolution cherche une première structure dans l'image puis on regroupe les structures trouvées à l'aide de la deuxième couche de regroupement. La troisième couche de convolution trouve de nouvelles struc-

tures dans l'image regroupée et regroupe ces structures avec la cinquième couche de regroupement. La dernière couche entièrement connectées récupère les valeurs des structures regroupées et en conclu une valeur chiffré.

La résolution du problème de détection d'objets (boîte de détection et localisation) dans une image avec un simple réseau neuronal convolutif standard ne fonctionnera pas car on ne connaît pas à l'avance le nombre d'occurrences d'objets présent dans l'image. Et donc, on ne peut pas fixer la valeur de la couche de sortie du réseau.

Une première solution à ce problème serait de récupérer différentes régions d'intérêts et d'utiliser un réseau neuronal convolutif sur ces régions pour classifier la présence ou non d'un objet dans ces régions. Le problème de cette approche est qu'un objet peut avoir différentes localisations dans une image, et donc, il faudrait sélectionner un nombre de régions d'intérêts élevées, ce qui prendrait beaucoup trop de temps.

A notre connaissance, voici les 4 méthodes principales de détection d'occurrences d'objets utilisant un réseau neuronal convolutif qui soit relativement rapide et efficace :

1. **R-CNN** :

Ross Girshick et al. [Mal14] ont proposé une méthode de classification d'objets que l'on peut séparer en 3 partie.

Dans la première partie, ils utilisent la recherche sélective pour extraire 2000 régions d ?intérêts dans l'image. Au lieu de classifier un grand nombre de régions, on classifie seulement 2000 régions.

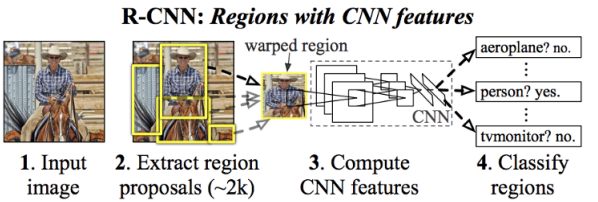

FIGURE 2.2 : Architecture simplifié du R-CNN (Source : « *Graphics in Rich feature hierarchies for accurate object detection and semantic segmentation* », page 1)

Dans la deuxième partie, ils transforment les 2000 régions en une région candidate. Cette région candidate passe dans un réseau neuronal convolutif qui extrait les caractéristiques de l'image et donne en sortie un vecteur de caractéristiques de dimension 4096. Ils ont utilisé l'implémentation Caffe du réseau neuronal convolutif décrit par Krizhevsky et al [SH12].

Dans la dernière partie, ils ingèrent les caractéristiques dans un SVM (un séparateur à vaste marge) afin de classifier la présence d'un objet dans la région proposée. Un séparateur à vaste marge est une technique d'apprentissage supervisé destinée à résoudre des problèmes de discrimination et de régression. L'algorithme prédit également 4 valeurs qui permettent d'augmenter la précision du cadre de sélection.

Par exemple, l'algorithme prédit la présence d'une personne, mais la tête de cette personne dans la région proposée peut être coupée en deux. L'intérêt est que les 4 valeurs de décalage prédit permettent d'ajuster le cadre

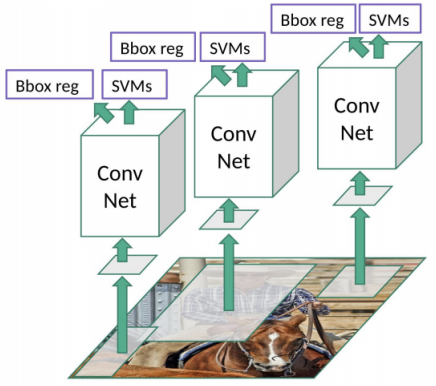

FIGURE 2.3 : R-CNN (Source : « *Graphics in Rich feature hierarchies for accurate object detection and semantic segmentation* »)

de sélection.

La partie apprentissage de la méthode R-CNN prend énormément de temps puisque l'on doit classifier 2000 régions par image. Elle ne peut pas être implémentée en temps réel car elle a une vitesse de test de 47 secondes par image.

2. **Fast R-CNN** :

La méthode Fast R-CNN [Gir15] est réalisée par les mêmes auteurs que la méthode R-CNN. Ils voulaient corriger le problème de lenteur d'apprentissage de l'ancienne méthode.

La méthode se déroule en 3 parties :

La première étape consiste à ingérer dans un réseau neuronal convolutif l'image complète et en sortir un tableau de caractéristiques de convolution (un tableau qui contient les caractéristiques nécessaire à la détection d'un objet). Par exemple, on veut détecter un vélo. Dans le réseau, on va d'abord détecter les courbes, puis les associées pour avoir des cercles et enfin, avec le tableau de caractéristiques remplis de cercles et de traits, on détecte le vélo.

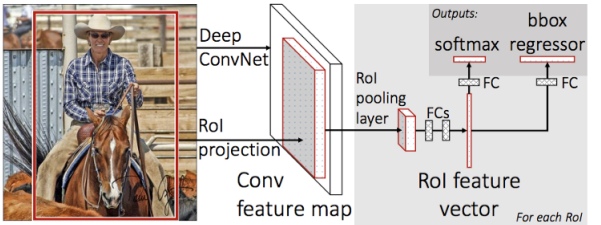

FIGURE 2.4 : Fast R-CNN (Source : « *Graphics in Fast R-CNN* »,page 2)

La deuxième étape consiste à identifier les régions de propositions à partir du tableau de caractéristiques fournit en sortie. Les régions d'intérêts sont dimensionnées en taille fixe carrée avec la couche de regroupement des régions d'intérêts (RoI pooling layer) afin qu'elles puissent être insérer dans la couche entièrement connectée.

La troisième étape consiste à récupérer le vecteur de caractéristiques de chaque région. A l'aide de ce vecteur, on prédit la classe de la région proposée avec la couche contenant la fonction softmax utilisée pour la classification multiple et on prédit les valeurs de décalage pour le cadre de sélection.

La méthode Fast R-CNN est plus rapide que R-CNN

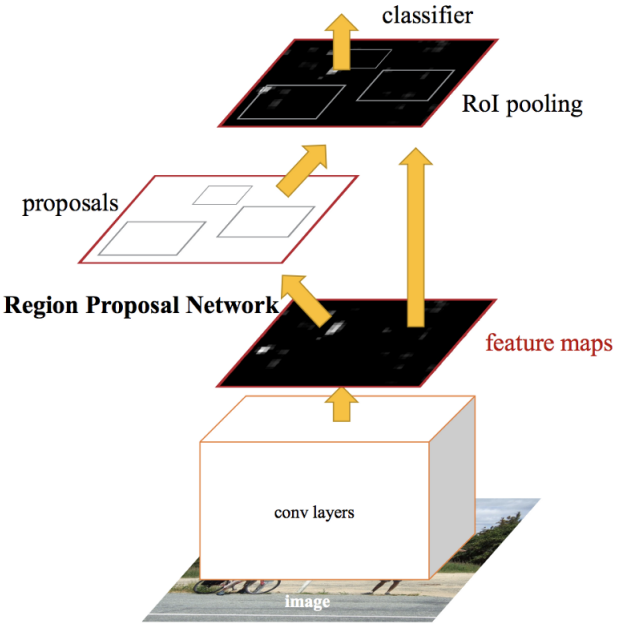

FIGURE 2.5 : Faster R-CNN (Source : « *Graphics in Faster R-CNN : Towards Real-Time Object* »,page 3)

car on n'insère pas les 2000 régions dans le réseau de neurone mais on insère toute l'image.

3. **Faster R-CNN** :

Faster R-CNN [Sun17] est une méthode créée par Shaoqing Ren et al. La différence entre cette méthode et les deux méthodes R-CNN et Fast R-CNN est que les deux méthodes utilisent la recherche sélective pour trouver les régions d'intérêts, alors que cette méthode utilise un réseau.

La méthode peut se séparer en 3 parties :

La première partie sert à insérer l'image dans un réseau neuronal convolutif et à en tirer un tableau de caractéristiques.

La deuxième partie consiste à utiliser un autre réseau neuronal convolutif pour déterminer les régions d'intérêts.

La troisième partie permet de remodeler les régions d'intérêts via une couche de regroupement RoI (region of interest). Les nouvelles régions sont utilisées pour prédire la classe des régions d'intérêts et prédire les valeurs de décalage pour le cadre de sélection.

Cette méthode est plus rapide que les précédentes et permet de faire de la détection en temps réel.

4. **YOLO** :

La méthode You Look Only Once [Far16] n'utilise pas de régions proposées dans son réseau neuronal convolutif contrairement aux méthodes précédentes. Dans cette méthode, seul un réseau neuronal convolutif prédit les régions de boîtes englobantes et les probabilités de prédiction d'appartenance à une classe pour chacune des boîtes.

YOLO prend une image en entrée et la sépare en une grille de SxS. Chaque case de la grille propose m boîtes englobantes. Pour chacune des boîtes englobantes, le réseau calcule une probabilité d'appartenance à une classe ainsi que les valeurs de décalage de la boîte englobante. Les boîtes englobantes ayant une valeur de probabilité supérieure à un certain seuil sont sélectionnées pour localiser l'objet dans l'image.

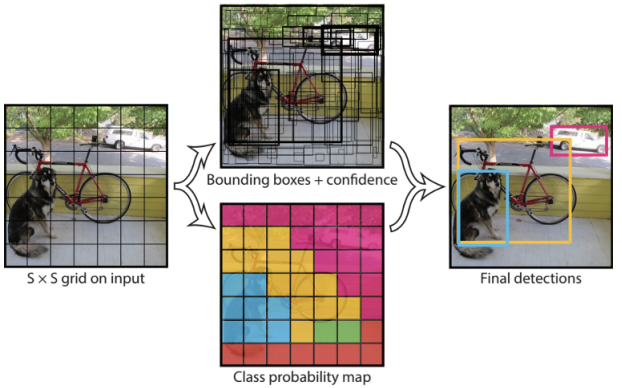

FIGURE 2.6 : YOLO (Source : « *Graphics in You Only Look Once : Unified, Real-Time Object Detection* »,page 2)

L'inconvénient de cette méthode est qu'elle ne peut pas détecter de petit objet tel que des oiseaux par exemple.

Les méthodes de détection d'objets sont efficaces mais ne répondent seulement qu'à la moitié de la problématique reformulée qui est la détection et le suivi de personnes dans une vidéo. Le suivi de personne se fait avec une ré-identification du client image par image. La différence entre le suivi et la simple détection de personne est que nous analysons une vidéo et non plus une seule image. Les algorithmes de détection d'objets dans une image numérique ne sont pas suffisants pour répondre à notre problématique.

Après avoir présenté une vision générale de quelques méthodes sur la détection d'objets dans une image numérique. On va s'orienter sur le deuxième problème de suivi de personne image par image. Suivre le client image par image est important pour pouvoir identifier un client et connaître son nombre de passages, son temps moyen de passages dans la vidéo. En terme général, avoir le suivi des personnes permet de récupérer un indicateur sur le nombre d'individus différents qui sont passés devant la caméra avec un taux d ?erreur plus ou moins varié.

Faire un état de l'art sur toutes les méthodes de suivi d'individu est difficile car depuis les années 2000, ce domaine est en pleine expansion et chaque chercheur réinvente une nouvelle méthode ou en modifie une déjà existante. Il est difficile de faire une liste complète des méthodes créée au fil du temps. De plus, il est compliqué de connaître la date exacte de modifications effectuées sur ces méthodes.

Il existe 2 types de méthodes de suivi : les méthodes en ligne qui reçoivent les données image par image et les méthodes hors ligne qui reçoivent toute la vidéo et peuvent donc utiliser des méthodes de prédiction pour prédire l'emplacement de l'objet à la prochaine image en fonction de toutes les images précédentes.

D'après l'article "Tracking the Trackers :An Analysis of the State of the Art in Multiple Object Tracking" [LLTAM17], qui compare les différentes méthodes de suivi de personnes dans une vidéo. Avoir une base de données normalisée de comparaison de méthodes de suivi est essentiel.

Dans cet article, les auteurs comparent les 10 premières méthodes de suivi issu des classements d'une base de données MOT (Multiple Objet Tracking). Ces méthodes possèdent les mesures de performances les plus élevés. MOT MOT15 et MOT16 sont des bases de données de vidéos de 2015 et 2016 sur lesquelles certaines méthodes de suivi se sont entraînées et ont été testé.

Il y a deux types d'approche pour le suivi de personnes. La première consiste à associer deux images entre elles. Il s'agit de l'association de données. Cette approche a été utilisée principalement avant 2015. L'objectif est de trouver une solution optimale pour résoudre le problème d'association de données.

Voici quelques méthodes présentées dans l'article utilisant l'association de données comme approche :

- DP-NMS [Fow11] : modèle graphique qui relie les détections dans un ensemble cohérent de trajectoires résolue via l'algorithme des k plus court chemins. La figure 2.7 représente l'architecture de cette méthode. On y trouve un nœud début (S), un nœud fin (T), 3 images successives. Les liens rouges représentent le lien entre l'image sans détection et l'image avec les détections de chaque objet. Les liens bleus représentent l'association entre les objets de deux images successives.

- LP2D [Ros11] : modèle linéaire résout avec l'algorithme du simplex. La figure 2.8 correspond à l'architecture de la méthode LP2D. Dans un graphe orienté, où 6 nœuds bleu correspondent à un objet de l'image et sont représentés par un nœud début et fin, où le nœud source (t) est relié à tous les objets "fin", où le nœud suivant (t) est relié à tous les objets "début", on a une association entre l'objet de l'image t-1 et les objets de l'image t.

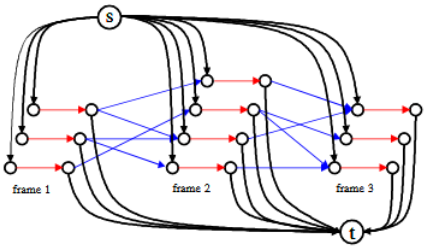

FIGURE 2.7 : Architecture DP-NMS (Source : « *Graphics in Globally-optimal greedy algorithms for tracking a variable number of objects* »,page 4)

- DCO-X [SR16] modèle utilisant un champ conditionnelle aléatoire.

- OVBT [APH13] : modèle bayésien varié.

- SMOT [SC11] : modèle de mouvement. Le mouvement permet de distinguer les objets ayant la même apparence.

L'association de données avec ces modèles sont basée sur des distances simples entre les données de détection des deux images ou des liens d'apparences faibles. Ces modèles ne sont pas très performants.

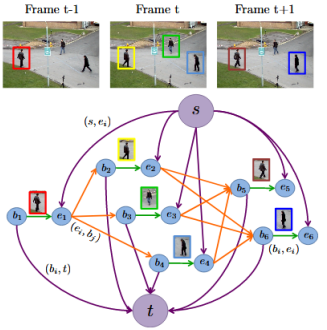

FIGURE 2.8 : Architecture LP2D (Source : « *Graphics in Everybody needs somebody : Modeling social and grouping behavior on a linear programming multiple people tracker* »,page 3)

La deuxième approche traitée par les auteurs de l'article de comparaison des méthodes de suivi de personnes, est la méthode d'affinité et d'apparence.

Récemment, on voit apparaître la construction de coût de similarité entre deux détections basé sur des indices d'apparence fort. Les traqueurs ont de meilleures performances et peuvent analyser des scénarios plus complexes comme la présence de plusieurs personnes avec occlusions.

Voici les 6 meilleures méthodes présentées dans l'article qui utilisent l'affinité et l'apparence comme approche :

- LINF1 [Ler17] : modèle d'apparence clairsemé. La figure 2.9 représente l'architecture de la solution LINF1. La première étape consiste à calculer les

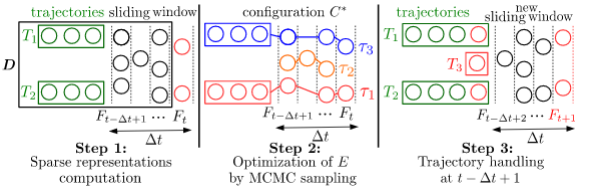

FIGURE 2.9 : Architecture LINF1 (Source : « *Graphics in Improving Multi-frame Data Association with Sparse Representations for Robust Near-online Multi-object Tracking* »,page 4)

"représentations" des détections à partir de la dernière image. (Cela correspond au rond rouge). La deuxième étape consiste à lier les "représentations" des détections de chaque image entre elles via une formule d'optimisation d'un paramètre E. La dernière étape consiste à définitivement estimé les trajectoires des détections dans la première image de la fenêtre glissante. (Les rectangles vert).

- MHT-DAM [Reh15] : modèle de mise à jour d'apparence en ligne. La figure 2.10 représente l'architecture de la méthode MHT-DAM. L'étape 1 consiste à suivre un ensemble d'hypothèses après le déclenchement du test au temps k. L'étape 2 consiste à récupérer les zones de contrôles des hypothèses de k+1 sur l'image k avec des seuils différents. L'étape 3 consiste à construire les arbres d'hypothèses correspondant. Chaque nœud est associé à une observation dans k. La dernière étape n'est pas présente dans la figure. Elle consiste à choisir le meilleur chemin

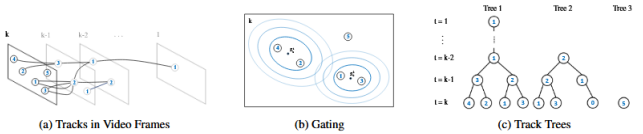

FIGURE 2.10 : Architecture MHT (Source : « *Graphics in Multiple Hypothesis Tracking Revisited* »,page 3)

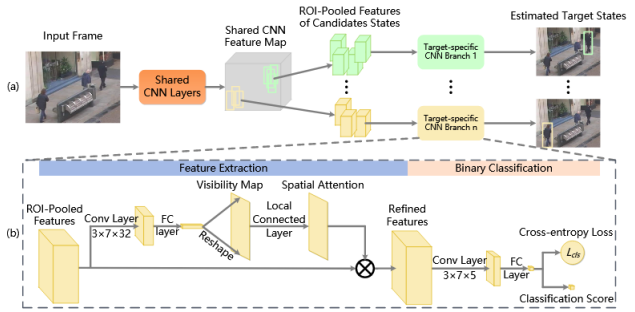

FIGURE 2.11 : Architecture oICF (Source : « *Graphics in Online multi-person tracking using Integral Channel Features* »,page 4)

dans l'arbre.

- oICF [Are16] : modèle d'apparence de fonction de canaux intégrés. La figure 2.11 représente l'architecture de la méthode oICF.

- NOMT [Cho15] : modèle de flux local agrégé de trajectoires de points d'intérêt à long terme. La figure 2.12 représente l'architecture de la méthode NOMT. Les modèles suivant utilisent les réseaux neuronaux.

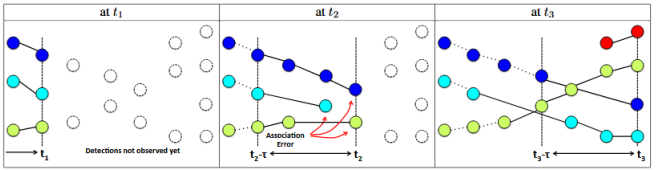

FIGURE 2.12 : Architecture NOMT (Source : « *Graphics in Near-Online Multi-target Tracking with Aggregated Local Flow Descriptor* »,page 2)

- MDPNN16 [Sav17] : modèle exploitant les réseaux de neurones récurrents pour coder l'apparence, le mouvement et les interactions. La figure 2.13 représente l'architecture de la méthode MDPNN16.

- JMC [Sch17] : modèle utilisant l'appariement profond pour améliorer la mesure d'affinité. La figure 2.14 représente l'architecture de la méthode JMC.

L'utilisation de réseau neuronal augmente grandement les performances. Les méthodes présentées dans cet état de l'art ne sont pas toutes pertinentes seules les dernières utilisant les réseaux de neurones le sont.

Je vais vous présenter un algorithme nommé SORT - Simple Online Real-time Tracking [Upc16] qui permet de traquer les personnes en temps réel dans une vidéo. Cet algorithme sert de base pour 2 des propositions suivantes.

La méthode consiste à détecter, propager l ?état des objets dans les images futurs, associer les détections actuelles aux objets existants et gérer la durée de vie des

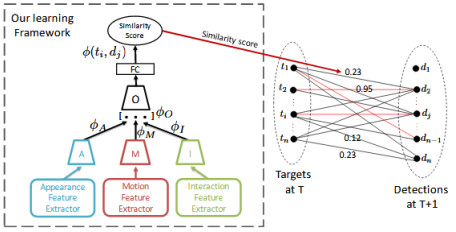

FIGURE 2.13 : Architecture MDPNN16 (Source : « *Graphics in Tracking the Untrackable : Learning to Track Multiple Cues with Long-Term Dependencies* »,page 3)

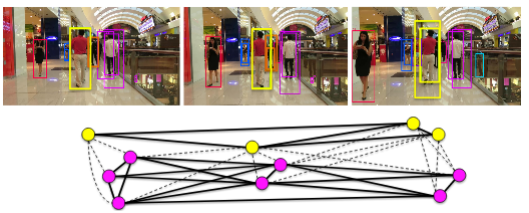

FIGURE 2.14 : Architecture JMC (Source : « *Graphics in Multiple People Tracking by Lifted Multicut and Person Re-identification* »,page 3)

objets suivis.

Elle utilise la méthode de détection Faster Region CNN [Sun17] qui consiste à extraire des caractéristiques des régions proposées puis de classer ces régions. Le filtre de Kalman est utilisé pour prédire les états des objets dans l'image suivante. L'association entre les cibles existantes et les détections est résolu en estimant pour chaque cible les cadres de sélection de chaque cible dans l'image actuelle. Ensuite, une matrice de coût des affectations est calculée en tant que distance d'intersection sur l'union (IOU) entre chaque détection et tous les cadres de sélection prévus des cibles existantes.

L'état de l'art permet d'avoir une vision plus ou moins générale des méthodes utilisées pour la détection et le suivi de personnes des plus anciennes au plus performantes.

Notre problématique est de trouver une méthode de suivi de personnes qui soit performante et récente puis de calculer des indicateurs plus ou moins pertinents vis-à-vis de la satisfaction client avec les données des résultats en sortie des différentes méthodes.

La plupart des solutions présentent dans l'état de l'art ne possèdent pas de code source, ce qui rend difficile la tâche d'intégration de la méthode à ce projet. Nous avons besoin d'une méthode de suivi de personnes qui soit récente, performante et dont le code est réutilisable.

2.1 Critères d'évaluation

Avant de détailler les 3 propositions, nous avons déterminé les critères d'évaluations qui permettront d'évaluer les propositions sur une vidéo de test enregistrée au préalable. Les critères d ?évaluation sont issus du MOTChallenge qui est un site contenant des données de tests de vidéos annotées pour tester les différents algorithmes de suivi de personnes.

Les critères d'évaluations des propositions retenus sont :

- FP : le taux de faux positifs est le nombre totale d'occurrence où un objet est détecté alors qu'il n'existe pas
- FP : le taux de faux négatives est le nombre totale d'occurrence où un objet existant n'est pas détecté
- ID Sw : le taux de changement d'identité est le nombre de fois où un objet est assigné à une nouvelle identité dans son suivi
- MOTA (Multiple object tracking accuracy) : taux de confiance de suivi d'objets multiple est une combinaison de 3 taux, MOTA = 1 - $\frac{\sum_t (fn_t + fp_t + idSw_t)}{\sum_t g_t}$ où pour chaque image t, g_t est le nombre d'objet présent, fn_t le nombre de faux négatifs, fp_t le nombre de faux positifs et $idSw_t$ le nombre d'identité changé
- Hz : la vitesse du traqueur qui est mesurer en Hz (image par seconde)

A ces critères, nous ajouterons le calcul des indicateurs pour savoir quelle proposition est la plus performante.

Par la suite, nous allons présenter les différentes propositions de méthodes de suivi de personnes et analyser ses méthodes pour savoir lesquelles sont celles qui répondent le mieux à notre problématique.

2.2 Proposition 1 : SORT avec une métrique d'association en profondeur

2.2.1 Présentation

Cette solution [Pau17] est basé sur la méthode Simple Online and Real-time Tracking (SORT) [Upc16] qui utilise des algorithmes de détections simples, en y ajoutant une information d'apparence et de mouvement pour détecter et suivre une personne dans une vidéo. SORT détermine la position des personnes dans une image via le filtre de Kalman, puis associe les données de détections image par image avec la méthode Hongroise qui utilise une métrique qui mesure le chevauchement du cadre de sélection des personnes.

La méthode d'apprentissage utilisé est d'apprendre hors ligne une métrique d'association approfondie sur un ensemble de données où les personnes sont ré-identifiées. L'application en ligne de la méthode s'effectue via des requêtes de recherche du k plus proches voisins sur une matrice d'association de mesures de suivi de détections dans un espace visuelle d'apparence.

Le système SORT avec une métrique d'association en profondeur peut être décrit en 4 composantes :

1. **Suivi de piste et estimation d'état** : Le scénario de suivi de piste est défini avec 8 dimensions d'état d'espaces qui sont la position centrale du cadre entourant une personne (u,v), un ratio d'aspect (γ), la hauteur du cadre (h) et leurs vitesses directionnelles respective associés ($\widetilde{x}$,$\widetilde{y}$,$\widetilde{\gamma}$,$\widetilde{h}$). Les coordonnées du cadre (u, v, γ, h) sont déterminées comme observation directe de l'objet, à l'aide d'un filtre de Kalman standard avec un mouvement à vitesse constante et un modèle d'observation linéaire. Le principe de ce composant est que pour chaque suivi ont compte le nombre d'images associées depuis la dernière bonne mesure d'association du filtre de Kalman. Lorsqu'une détection ne peut pas être associée à un suivi déjà existant, un nouveau suivi est créé et au bout de 3 images successives où le suivi est considéré comme une tentative, si le suivi n'a pas de mesure d'association valide alors on le supprime.

2. **Problème d'affectation** : L'algorithme Hongrois résout le problème d'association entre les mesures prédites du filtre de Kalman et les nouvelles mesures. Cette méthode intègre une combinaison des mesures de mouvement et d'apparence. La mesure de mouvement basé sur la distance Mahalanobis, donne l'emplacement des objets en fonction de leurs mouvements, ce qui est performant pour la reconnaissance à court terme. La mesure d'apparence basée sur la distance cosinus, donne des informations d'apparence, ce qui est performant pour la reconnaissance après de longues occlusions.

3. **Cascade d'association** : Au lieu d'associer les mesures au suivi en un seul bloc, on résout l'association en résolvant plusieurs sous-problèmes.

4. **Descripteur d'apparence profonde** : Un réseau neuronal convolutif est utilisé hors ligne pour apprendre les métriques profondes d'association dans un contexte de suivi de personne. En ligne, leur méthode utilise des requêtes du k plus proches voisins pour déterminer les détections similaires image par image.

2.2.2 Analyse

Intérêts de la proposition *1*

L'un des plus grands intérêts de la proposition est de pouvoir suivre les personnes malgré un long moment d'occlusion. La proposition est simple à mettre en place et peut être lancé en temps réel avec Nvidia GeForce GTX 1050 mobile GPU. De plus, le réseau pré-entrainé est disponible sur leur projet GitHub https://github.com/nwojke/deep_sort.

Le code est disponible en python 2.7 et 3. Une démonstration est utilisable sur les données MOT16 pour tester le modèle. Le code est open source.

Limites de la proposition *1*

La puissance nécessaire au calcul peut poser problème car nous n'avons pas de GPU mais un simple CPU qui ne fait pas de calcule parallèle contrairement au GPU.

2.3 Proposition 2 : Suivi de multiple objets en temps réel C++SORT

2.3.1 Présentation

La méthode [Mur17] est proposé par un étudiant en thèse dont le sujet est de combiner des techniques de détection, de prédiction et d'association pour créer une méthode de suivi de multiple objets.

Dans cette méthode, la tâche de détection et de suivi d'objet sont séparée. La méthode SORT - Simple Online and Real-time tracking est utilisée, cependant la méthode est étendue en utilisant d'autres mesures de similarité que l'IoU (intersection sur l'union).

Le principe de la méthode est de gardée un suivi de chaque objets en modélisant son mouvement avec une prédiction à l'aide d'un filtre de Kalman ou d'un filtre de particule. Pour chaque image, les objets sont détectés, de nouvelles localisations des suivis d'objets déjà suivi sont prédites et enfin, les détections et les suivis d'objet sont associés basé sur la similarité des cadres. Les prédictions sont mis-à-jour avec les nouvelles détections qui leurs sont associés. De nouvelles prédictions sont initialisées pour chaque détection non associée à un suivi. Les prédictions non utilisées sont supprimées.

2.3.2 Analyse

Intérêts de la proposition *2*

La proposition est codé en c++ ce qui permet une réutilisation du code plus simple. De plus, le code est disponible sur le projet GitHub `https://github.com/samuelmurray/tracking-by-detection` en open source.

Limites de la proposition *2*

La limite de cette proposition est qu'elle ne marche pas en temps réel. L'association d'une méthode de détection aux techniques de prédictions et d'associations prend trop de temps.

2.4 Proposition 3 : Ensemble de filtres de corrélation par noyau pour le suivi d'objets à grande vitesse

2.4.1 Présentation

La méthode [Seo18] proposée utilise un ensemble de filtre de corrélation par noyau qui gèrent les variations d'échelles et les mouvements des cibles (personnes).

3 filtres de corrélation par noyau sont appliquées successivement sur une image, au lieu de les appliquer tous sur la même image. Le premier filtre RSt apprend la zone cible et son fond, le deuxième filtre Rs se concentre sur l'apprentissage de l'échelle de la cible et le dernier filtre RLt se concentre sur l'apprentissage de la zone cible et sur un fond plus grand que celui du premier filtre.

La méthode fonctionne par étape est représenté à la figure : la première image est utiliser pour initialiser l'al-

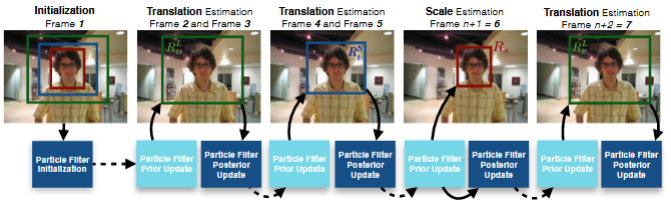

FIGURE 2.15 : EnKCF (Source : « *Graphics in Ensemble of Kernelized Correlation Filters for High-Speed Object Tracking* »,page 4)

gorithme de suivi et le filtre de particule. Pour les 6 prochaines images chacun des 3 filtres est déployer successivement les changements d'échelle d'une cible. Puis on répète l'ordre des 3 filtres.

«Le fait qu'une matrice circulante puisse être diagonalisée par transformée discrète de Fourier est la solution pour réduire la complexité de toute méthode de suivi basée sur un filtre de corrélation. Les éléments non diagonaux deviennent nuls alors que les éléments diagonaux représentent les valeurs propres de la matrice circulante. Les valeurs propres sont égales à la transformation DFT des éléments de l'échantillon de base (x). Le filtre de corrélation par noyau, en particulier, applique un noyau à x pour le transformer en un domaine plus discriminant. La matrice circulante est ensuite formée en appliquant des décalages cycliques sur le noyau x»,(Source : « *Citation in Ensemble of Kernelized Correlation Filters for High-Speed Object Tracking* »

2.4.2 Analyse

Intérêts de la proposition *3*

La proposition 3 possède un code source sur un GitHub `https://github.com/buzkent86/EnKCF_Tracker` réutilisable mais la méthode n'est pas facilement compréhensible. La méthode est implémentée en c++. De plus, il est possible de l'utilisée en temps réel. L'un des objectifs des chercheurs ayant implémenté cette méthode est de pouvoir utiliser une méthode de suivi de personnes sur des systèmes limités en calcul.

Limites de la proposition *3*

La méthode ne gère pas les occlusions.

2.5 Récapitulatif

À l'issue de ce travail bibliographique et critique, un résumé des éléments intéressants et des éléments manquants de chaque proposition est mis en évidence avec une présentation condensée des arguments ayant abouti à cette dichotomie.

Les éléments seront présentés de manière synthétique dans un tableau comparatif (cf. tableaux 2.1 et 2.2 à titres d'exemples simplifiés).

2.6 Conclusion

Nous avons détaillés les avantages et les inconvénients des 3 propositions abordés qui possèdent un code source

Proposition	Avantages	Inconvénients
Proposition 1	Gestion des occlusions Facilement mise en place Code source disponible en python	Puissance de calcul sur GPU
Proposition 2	Code source disponible en c++ Gère les occlusions	Ne gère pas le temps réel
Proposition 3	Code source disponible en c++ Gère le temps réel Utilisable sur un système limités en calcul	Ne gère pas les occlusions

TABLE 2.1 : Tableau comparatif des propositions étudiées

Avantages	Proposition 1	Proposition 2	Proposition 3
Code source disponible	$\surd$	$\surd$	$\surd$
Gère les occlusions	$\surd$	$\surd$	
Calcule GPU	$\surd$	$\surd$	
Calcule CPU	$\approx$	$\approx$	$\surd$
Temps réels	$\surd$		$\surd$
Méthode facilement compréhensible	$\surd$	$\surd$	

TABLE 2.2 : Tableau comparatif des avantages des propositions étudiées

réutilisable. À l'issue de ce travail de recherche bibliographique, il apparaît que plusieurs propositions peuvent servir de base à la résolution de notre problème de suivi de personnes.

Par la suite, nous allons essayer d'implémenté une de ces propositions et observé s'il est possible de la faire fonctionner sur nos propres données de test.

3

Conclusion

En conclusion, le travail de recherche effectué sur la détection de personnes dans une image a permis de prendre connaissance des techniques de détections avec des régions proposées englobantes qui sont utilisées dans la plupart des méthodes de détection de personnes. Des régions sont proposées, puis des vecteurs de caractéristiques sont associés à ces régions, puis on calcule la probabilité d'une région d'appartenir à une classe spécifique. Au final, en sortie, on obtient la localisation de la région englobante d'un objet et sa classe.

La partie sur le suivi d'individus est plus complexe car de nombreuses techniques sont utilisées mais les principales techniques consistent à récupérer les détections image par image et à les associées. Pour cela, il existe deux méthodes : l'association de données classique entre 2 vecteurs de caractéristiques et l'association d'affinité où la méthode prédit les futures détections en s'aidant des anciennes puis associe les prédictions avec les détections pour avoir un suivi de la détection d ?un objet. L'association d'affinité peut aussi se faire en prédisant plusieurs prédictions pour un objet et en choisissant la bonne prédiction à chaque image.

Mon idée a été de trouver 3 propositions possédant un code source et qui pouvaient être intéressantes. Suite à la présentation des 3 propositions, je cherche maintenant à évaluer ces trois propositions sur les critères d'évaluations définis dans la partie critères d'évaluations avec une vidéo que je vais créer pour répondre au contexte spécifique du sujet.

Le résultat de l'évaluation nous montrera qu'elle méthode répond au mieux à la problématique.

3.1 Enseignements

Ce travail a permis de me renseigner sur les méthodes de détection et de suivi de personnes. Il s ?agit d ?un domaine que je ne connaissais absolument pas. De plus, j'ai compris qu'un projet de recherche et de développement se concentre sur la lecture d'articles scientifiques pour en trouver des éléments qui peuvent nous aider à proposer

notre propre solution. Pour ma part, il existait déjà des méthodes implémentées par des chercheurs qui semble performantes, j'ai préféré les réutilisées au lieu de réinventé toute la méthode.

3.2 Perspectives de recherche

Comme perspective de recherche, nous pouvons très bien imaginer l ?ajout de modules à ce projet pour améliorer les résultats produits afin de satisfaire l'analyse de la satisfaction client en magasin.

Nous pouvons ajouter le module de détection de visages qui permettra de savoir si la personne regarde en direction du stand ou non. Nous pouvons aussi ajouter un module de détection de caractéristiques pour affiner les statistiques résultant de l ?analyse de la satisfaction client en magasin. Par exemple, on pourrait connaître le nombre de femme et d ?homme passés devant le stand.

Bibliographie

[APH13] Y. Ban ; S. Ba ; X. Alameda-Pined and R. Horaud. The way they move : Tracking multiple targets with similar appearance. *ICCV*, 2013. 20

[Are16] Hilke Kieritz ; Stefan Becker ; Wolfgang Hübner ; Michael Arens. Online multi-person tracking using integral channel features. *2016 13th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 2016. 21

[Cho15] Wongun Choi. Near-online multi-target tracking with aggregated local flow descriptor. *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015. 22

[Far16] Joseph Redmon ; Santosh Divvala ; Ross Girshick ; Ali Farhadi. You only look once : Unified, real-time object detection. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. 18

[Fow11] Hamed Pirsiavash ; Deva Ramanan ; Charless C. Fowlkes. Globally-optimal greedy algorithms for tracking a variable number of objects. *CVPR 2011*, 2011. 20

[Gir15] Ross Girshick. Fast r-cnn. *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015. 17

[Haf98] Y. Lecun ; L. Bottou ; Y. Bengio ; P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 96 :2278 – 2324, 1998. 15

[HW98] B. Heisele and C. Wöhler. Motion-based recognition of pedestrians. *Proceedings. Fourteenth International Conference on Pattern Recognition*, 2, 1998. 14

[Ler17] Loïc Fagot-BouquetRomaric AudigierYoann DhomeYoann DhomeFrédéric Lerasle. Improving multi-frame data association with sparse representations for robust near-online multi-object tracking. *European Conference on Computer Vision*, 2017. 21

[LLTAM17] Konrad Schindler ; Daniel Cremers ; Ian Reid ; Stefan Roth Laura Leal-Taixe ; Anton Milan. Tracking the trackers : An analysis of the state of the art in multiple object tracking. 2017. 19

[Mal14] Ross Girshick ; Jeff Donahue ; Trevor Darrell ; Jitendra Malik. Rich feature hierar-

chies for accurate object detection and semantic segmentation. *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014. 16

[Mur17] Samuel Murray. Real-time multiple object tracking - a study on the importance of speed. 2017. 26

[Pau17] Nicolai Wojke ; Alex Bewley ; Dietrich Paulus. Simple online and realtime tracking with a deep association metric. *2017 IEEE International Conference on Image Processing (ICIP)*, 2017. 24

[Reh15] Chanho Kim ; Fuxin Li ; Arridhana Ciptadi ; James M. Rehg. Multiple hypothesis tracking revisited. *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015. 21

[Ros11] Laura Leal-Taixé ; Gerard Pons-Moll ; Bodo Rosenhahn. Everybody needs somebody : Modeling social and grouping behavior on a linear programming multiple people tracker. *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, 2011. 20

[Sav17] Amir Sadeghian ; Alexandre Alahi ; Silvio Savarese. Tracking the untrackable : Learning to track multiple cues with long-term dependencies. *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017. 22

[SC11] C. Dicle ; M. Sznaier and O. Camps. Everybody needs somebody : Modeling social and grouping behavior on a linear programming multiple people tracker. *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, 2011. 20

[Sch17] Siyu Tang ; Mykhaylo Andriluka ; Bjoern Andres ; Bernt Schiele. Multiple people tracking by lifted multicut and person re-identification. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 22

[Seo18] Burak UzkentYoung-Woo SeoYoung-Woo Seo. Enkcf : Ensemble of kernelized correlation filters for high-speed object tracking. 2018. 26

[SH12] A. Krizhevsky ; I. Sutskever and G. Hinton. Imagenet classification with deep convolutional neural networks. *In NIPS*, 2012. 16

[SR16] A. Milan ; K. Schindler ; and S. Roth. Multi-target tracking by discrete-continuous energy minimization. *TPAMI*, 2016. 20

[Sun17] Shaoqing Ren ; Kaiming He ; Ross Girshick ; Jian Sun. Faster r-cnn : Towards

real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39 :1137 – 1149, 2017. 18, 23

[Upc16] Alex Bewley; Zongyuan Ge; Lionel Ott; Fabio Ramos; Ben Upcroft. Simple online and realtime tracking. *2016 IEEE International Conference on Image Processing (ICIP)*, 2016. 22, 24

[VJ01] Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. *Conference on Computer Vision and Pattern Recognition*, 2001. 14

Table des figures

Liste des tableaux

Fiches de lecture

A.1 Detecting and Tracking of Multiple People in Video based on Hybrid Detection and Human Anatomy Body Proportion

A.1.1 Référence de l'article

Titre : Detecting and Tracking of Multiple People in Video based on Hybrid Detection and Human Anatomy Body Proportion.

Auteurs : El Maghraby Amr, Abdalla Mahmoud, Enany Othman et Y. EL Nahas Mohamed.

Université : Zagazig University et Elazhar University

Journal : International Journal of Computer Applications (0975 - 8887) - Volume 109 - No. 17

Date de parution : Janvier 2015

Termes généraux : Video Processing, Computer vision systems, Human detection and tracking, Clustering.

Mots Clés : Video Processing, Human detection and tracking, Viola-Jones upper body, Skin detection, Computer vision systems, Biometrics.

Lien : http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.695.6435&rep=rep1&type=pdf

A.1.2 Situation des auteurs

Trois des auteurs ont étudié dans le domaine du système de l'ingénierie informatique dont l'un a étudié à l'université de Elazhar. Abdalla Mahmoud a étudié l'ingénierie de communication. Cette thèse est un achèvement du travail de recherche qu'ils avaient commencé. Ils ont déjà publié deux articles scientifiques. L'un des articles traite

d'un système de détection hybride de visage utilisant la combinaison de la méthode Viola-Jones et une méthode de détection de la peau. Le second article traite de la détection et l'analyse d'informations sur les parties d'un visage en utilisant Viola-Jones et une approche géométrique.

A.1.3 Introduction

Cette thèse traite d'une méthode scientifique totalement automatisée mis en place par les auteurs pour répondre à une problématique récurrente dans le domaine de l'interaction Homme-Machine : la détection et le suivi de personne. Plus précisément, ils travaillent sur la détection et le suivi de plusieurs personnes en mouvement sur une vidéo.

A.1.4 Méthode de détection

Leur méthode se découpe en 3 phases :

1. La première phase consiste à analyser la vidéo image par image et d'y appliquer un algorithme. Cet algorithme va appliquer des détecteurs primaires sur les images basé sur l'algorithme de détecteurs d'objets en cascade de Viola-Jones afin de détecter le haut du corps humain qu ?il va placer dans une boîte.

2. Suite à la première phase, les auteurs ont récupéré des suites d'images contenant des boîtes entourant le haut des corps humains détectés par l'algorithme. A l'aide des proportions anatomique du corps humain, ils sont parvenus à localiser la position de la tête et du visage dans la boîte. Le résultat de cette étape retourne des détections d'humain positif et négatif.

3. La troisième étape sert à identifier au mieux la différence entre les bonnes et les mauvaises détections. Pour cela, ils vont utiliser la détection de la couleur de peau sur la partie du visage qui a été localisé à l ?étape 2.

A.1.5 Méthode de traçage

Le suivi se fait en répétant la méthode de détection sur chaque image de la vidéo. A chaque image, ils vont récupérer des informations sur les blocs valides de détections du haut du corps humain et en faire des moyennes de largeur et longueur du visage. Puis, ils qualifient les images en fonction de leurs moyennes via la méthode de classification (k-means). De cette manière, pour chaque image, ils déterminent l'appartenance des blocs à une personne. On peut donc suivre les traces de cette personne.

A.1.6 Conclusion

Cette thèse est relativement complète et explique très bien le processus de détection. Cette méthode correspond parfaitement à mon sujet d'analyse de la satisfaction client via une vidéo puisque je vais avoir besoin de trouver une méthode de détection de personnes en mouvement sur une vidéo.

A.2 Rapid object detection using a boosted cascade of simple features

A.2.1 Référence de l'article

Titre : Rapid object detection using a boosted cascade of simple features.

Auteurs : Paul Viola et Michael Jones.

Conférence : Conference on Computer Vision and Pattern Recognition

Date de parution : 2001

Lien : https://ieeexplore.ieee.org/document/990517

A.2.2 Situation des auteurs

Paul Viola, ancien professeur au MIT et vice-président des sciences pour Amazon Air est chercheur en vision par ordinateur. Michael Jones travaillais en 2001 pour le laboratoire Compaq CRL situé à Cambridge.

A.2.3 Introduction

Cette publication traite d ?un framework robuste de détection d ?objet visuel rapide et possédant un taux de précision élevé mis en place par les auteurs.

A.2.4 Définition

Image intégrale

L'image intégrale peut être calculée à partir d'une image en utilisant quelques opérations par pixel. Une fois calculée, chacune de ces caractéristiques peut être calculée à n'importe quelle échelle ou emplacement en temps constant.

A.2.5 Méthode de détection

Leur méthode se découpe en 3 clés de contributions :

1. La première phase est l ?introduction d ?une nouvelle représentation d ?image appelée «Image intégrale», qui permet de calculer très rapidement les caractéristiques utilisées par le détecteur.

2. La deuxième phase est un algorithme d ?apprentissage, basé sur Ada-Boost, qui sélectionne un petit nombre de caractéristiques visuelles critiques de Haar et produit des classificateurs extrêmement efficaces.

3. La troisième étape est une méthode pour combiner des classificateurs dans une "cascade" qui permet d ?éliminer rapidement les régions d ?arrière-plan de l ?image tout en se concentrant sur les régions prometteuses.

Deux caractéristiques sont relevées pour détecter les visages : La première caractéristique mesure la différence d ?intensité entre la région des yeux et la région sur les

joues supérieures. La fonctionnalité tire profit de l?observation que la région des yeux est souvent plus sombre que les joues. La deuxième caractéristique compare les intensités dans les régions des yeux aux intensités sur la région du nez.

A.2.6 Performance

Fonctionnant sur des images de 384 x 288 pixels, les visages sont détectés à 15 images par seconde sur un Intel Pentium III 700 MHz classique.

A.2.7 Conclusion

Cet article explique une méthode rapide de détection de visage en sélectionnant des caractéristiques grâce à une variance de l?algorithme AdaBoost et en apprenant ses caractéristiques à un classificateur.

A.3 You Only Look Once : Unified, Real-Time Object Detection

A.3.1 Référence de l'article

Titre : You Only Look Once : Unified, Real-Time Object Detection

Auteurs : Joseph Redmon, Santosh Divvala, Ross Girshick et Ali Farhadi.

Université : University of Washington

Date de parution : 8 juin 2015

Date de dernière révision : 9 mai 2016

Lien : https://arxiv.org/abs/1506.02640

A.3.2 Situation des auteurs

Joseph Redmon est un informaticien passionné d?apprentissage automatique, d?analyse de données, ainsi que de conception et de mise oeuvre de programmes de bas niveau. Il travaille sur la vision par ordinateur. Il possède son propre site web : https://pjreddie.com/.

Santosh Divvala est chercheur scientifique chez AI2. Son intérêt principal est la vision par ordinateur, en particulier le problème de compréhension des images.

Ross Girshick est chercheur à Facebook AI Research (FAIR) et travaille sur la vision par ordinateur et l'apprentissage automatique.

Ali Farhadi est professeur associé au département d'informatique et d'ingénierie de l'Université de Washington. Il s'intéresse principalement à la vision par ordinateur, à l'apprentissage automatique, à l'intersection du langage naturel et de la vision, à l'analyse du rôle de la sémantique dans la compréhension visuelle et au raisonnement visuel

A.3.3 Introduction

Cette publication traite d'une nouvelle approche de détection d ?objets visuels en temps réel. Ils décrivent la détection d'objets comme un problème de régression dans des boîtes englobantes séparées dans l'espace associées à des probabilités de classe. Un seul réseau de neurones prédit les limites et les probabilités de classe directement à partir d'images complètes.

A.3.4 Méthode de détection

Leur méthode de détection se base sur un seul réseau de neurones. Ils unissent les composants de la détection d'objets dans un seul réseau. Le réseau utilise des caractéristiques de l'ensemble de l'image pour prédire des cadres de sélections. Il prédit tous les cadres de sélection de toutes les classes possibles sur une seule image en même temps.

L'algorithme va séparer l'image d'entrée en une grille de dimension SxS. Chaque case de la grille va prédire B boîtes et les scores de confiance pour ses boîtes. Ces scores de confiance reflètent le degré de confiance du modèle sur le fait que la boîte contient un objet et la précision avec laquelle il prédit la boîte. Je ne détaille pas le calcul du score de confidence.

Chaque boîte contient 5 prédictions : x, y, w, h et le score de confiance. Les coordonnées (x, y) représentent le centre de la boîte par rapport aux limites de la cellule de la grille. La largeur et la hauteur sont prédites par rapport à l'image entière. Chaque cellule de grille prédit également C probabilité conditionnelle, p(Classe i | Objet). Ces probabilités sont conditionnées par la cellule de la grille contenant un objet.

A.3.5 Désigne du réseau

Ce modèle de détection d'objets est implémenté comme un réseau de neurone de convolution et est évalué via le jeu de données de détection de PASCAL VOC. Les couches convolutives initiales du réseau extraient les caractéristiques de l'image, tandis que les couches entièrement connectées prédisent les probabilités et les coordonnées de sortie. Le réseau comporte 24 couches convolutives suivies de 2 couches entièrement connectées. Il utilise des couches de réduction 1x1 suivi de couches convolutives de 3x3.

A.3.6 Performance

Le modèle YOLO traite les images en temps réel à 45 images par seconde sans traitement par lots sur un Titan X GPU. Une version réduite du réseau, Fast YOLO, traite 155 images par seconde. Par rapport aux autres systèmes de détection, YOLO fait plus d'erreurs de localisation mais est moins susceptible de prédire de faux positifs. YOLO atteint plus de deux fois la précision moyenne des autres systèmes de détection en temps réel.

A.3.7 Conclusion

Cet article explique une méthode en temps réel de détection d'objet en utilisant un réseau de neurone de convolution sur l'image complète.

A.4 REAL-TIME MULTIPLE PEOPLE TRACKING WITH DEEPLY LEARNED CANDIDATE SELECTION AND PERSON RE-IDENTIFICATION

A.4.1 Référence de l'article

Titre : REAL-TIME MULTIPLE PEOPLE TRACKING WITH DEEPLY LEARNED CANDIDATE SELECTION AND PERSON RE-IDENTIFICATION

Auteurs : Long Chen, Haizhou Ai, Zijie Zhuang et Chong Shang.

Université : Tsinghua University

Date de parution : 12 septembre 2018

Sujets : Computer Vision and Pattern Recognition.

Lien vers le document : https://arxiv.org/abs/1809.04427v1

A.4.2 Situation des auteurs

Long Chen, Haizhou Ai, Zijie Zhuang et Chong Shang sont 4 chercheurs du département informatique de l'université de Tsinghua.

A.4.3 Introduction

Cette publication traite d'une nouvelle approche de détection et de suivi d'une personne en temps réel. Ils proposent de détecter les détections de personne non fiable en collectant les candidats à partir des résultats de détection et de suivi d'une personne. Dans certaine situation, les systèmes de détection et de suivi peuvent se compléter. D'une part, les détections fiables du traqueur peuvent être associées à court terme aux détections en cas de détection manquante ou non précise. D'autre part, les résultats fiables des détections permettent d'éviter les écarts de détection des traqueurs. Afin de sélectionner en temps réel le candidat optimal, il utilise une nouvelle fonction de scoring basé sur un réseau de neurone convolutif.

A.4.4 Méthode de détection

Leur méthode de détection se base sur la sélection de candidat entre les méthodes de détection et de traçage.

Tout d'abord, ils mesurent tous les candidats à l'aide d'un score de notation unifié. Pour former cette fonction de score, ils fusionnent un classificateur d'objets entraîné de manière discriminatoire avec le degré de confiance du traqueur. Puis, à l'aide de la suppression non maximal

effectué sur les scores estimés, ils obtiennent les candidats sans redondance.

Pendant la procédure d'entrainement de leur réseau, ils échantillonnaient aléatoirement des régions d'intérêts (candidat à classer avec les paramètres x0,y0,w,h) autour des vrais boîtes de détection et les considéraient comme des exemples positifs. Ils prenaient le même nombre de régions d'intérêts du fond de l'image (background) comme des exemples négatifs. De cette manière, le réseau apprend à reconnaître les localisations des objets.

A.4.5 Désigne du réseau

Leur classificateur se base sur le réseau de neurone entièrement convolutif basé sur une région de l'image (R-FCN). La carte des scores de l'image sont prédit en utilisant un réseau de neurone convolutif avec une architecture codeur-décodeur. La partie codeur est la couche légère centrale du réseau pour une performance en temps réel. Ils ajoutent la partie décodeur avec un sur-échantillonnage pour augmenter les cartes de scores des résolutions spatiales pour une classification future.

A.4.6 Méthode de traçage

La méthode de suivi consiste à prédire la localisation de chaque trace existante en utilisant le filtre de kalman. Ces prédictions sont adoptées pour éviter les détections fausses causées par des variations visuelles des propriétés des objets ou par les occlusions dans les scènes de passage de personnes. Cependant, ces prédictions ne sont pas utilisables à long termes pour le suivi. Afin de mesurer le degré de confiance du filtre de kalman dont la précision peut décroitre s'il n'est pas mis à jour par détection au bout d'un certain temps, il utilise un indice de confiance par sous trace en utilisant des informations temporel.

Une trace peut être séparée en plusieurs sous trace.

A.4.7 Méthode d'association

Nous obtenons un candidat obtenu via la méthode de détection et un candidat obtenu via la méthode de suivi. Ils vont utiliser la suppression non maximale pour sélectionner le candidat idéal.

A.4.8 Méthode de comparaison

Pour une meilleure précision, ils utilisent un réseau de neurone de comparaison des traits de caractères entre deux images afin de savoir s'il s'agit de la même personne.

A.4.9 Association hiérarchique des étapes de détection et de traçage

Premièrement, ils appliquent l'association de données sur les candidats issus de la détection en utilisant l'apparence avec un seuil Td pour la distance maximal.

Ensuite, ils associent les candidats restant avec les traces non associé basé sur la jointure entre les candidats de détection et les candidats de traçages. Ils ne mettent à jour les représentations d'apparence que lorsque les candidats de traçage sont associés à une détection. La mise à jour est effectuée en sauvegardant les caractéristiques de ré identification (ReID) de la détection associée.

Ensuite, les nouvelles traces sont initialisées avec les résultats des détections restantes.

Avec l'association de données hiérarchique, ils ont seulement besoin d'extraire les fonctionnalités ReID pour les candidats de détection une fois par image. En combinant cela avec l'ancienne fonction de scoring efficace et les degrés de confiance des sous trace, leur infrastructure peut fonctionner en temps réel.

A.4.10 Performance

Ils utilisent SqueezeNet [16], la couche centrale de R-FCN pour la performance en temps réel. Leur réseau de neurone convolutif, composé de SqueezeNet et du décodeur, ne coûte que 8 ms pour estimer les cartes de scores pour une image d'entrée de la taille de 1152x640 sur un GPU GTX1080Ti.

Ils fixent k=7 pour les cartes de scores sensible à la position, et entraîne le réseau en utilisant l'optimiseur RMSprop avec un taux d'apprentissage de 1e-4 et une taille de lot de 32 pour 20 000 itérations.

Les données de formation pour la classification des personnes sont collectées à partir de MSCOCO.

A.4.11 Conclusion

Cet article explique une méthode en temps réel de détection de personne en utilisant une infrastructure composé de plusieurs parties : détection de personnes ressemblantes, vérification de détection des personnes via l'association de candidats de détection et de traçage, utilisation de caractéristiques de ré identification.

Ils traitent les détections non fiable en sélectionnant les candidats parmi les sorties de détection et de suivi.

La fonction de notation pour la sélection des candidats est formulée par un R-FCN efficace, qui partage les calculs sur toute l'image.

Ils améliorent la capacité d'identification des occlusions en introduisant les fonctionnalités ReID pour l'association des données.

Pour conclure, le traqueur proposé permet d'obtenir des résultats en temps réel et des performances de pointes sur la référence MOT16.

A.5 Bounding Box Embedding for Single Shot Person Instance Segmentation

A.5.1 Référence de l'article

Titre : Bounding Box Embedding for Single Shot Person Instance Segmentation

Auteurs : Jacob Richeimer et Jonathan Mitchell.

Date de parution : 20 juillet 2018

Sujets : Computer Vision and Pattern Recognition.

Lien vers le document : https://arxiv.org/abs/1807.07674

A.5.2 Situation des auteurs

Jacob Richeimer est le directeur de l'entreprise nommé OCTI, INC. L'entreprise OCTI INC développe la technologie de vision par ordinateur et d'apprentissage automatique de l'application de messagerie vidéo en réalité augmentée.

Notamment : estimation de la pose humaine 3D mobile en temps réel, segmentation d'instances sémantiques et reconnaissance des actions squelettiques. Langues : Python. Outils : Keras, Tensorflow.

Jonathan Mitchell est ingénieur dans cette entreprise et se concentre sur la vision par ordinateur, l'apprentissage en profondeur (deep learning), et en particulier la détection de pose humaine et la segmentation d'instances.

A.5.3 Introduction

Cette publication présente une nouvelle approche pour la tâche de segmentation d'instances d'une personne en utilisant un modèle en single-shot (en un coup). Le modèle proposé emploi un réseau de neurone de convolution qui est entraîné pour prédire aussi bien les masques de segmentation par classe (ici les personnes) que les boîtes englobantes des instances d'objets (de personnes) auxquelles chaque pixel appartient. Les auteurs de cet article cherchent à associer un pixel à l'instance de l'objet (de la personne) auquel il appartient.

A.5.4 information

Deux approches de détection

1. L'approche "Top-down" consiste à d'abord localiser les instances de l'objet puis à obtenir le masque de pixel pour chaque instance détectée.

2. l'approche "Bottom-up" consiste à d'abord déterminer la classe d'objet de chaque pixel puis à les grouper en une seule instance d'objet.

Dans cet article, les auteurs adoptent l'approchent "Bottom-up" et propose une méthode simple qui ne

requiert qu'un minimum de calculs en plus des actuelle approches de l'état de l'art sur la segmentation sémantique catégorique.

L'approche "Bottom-up" demande, après la segmentation sémantique, d'ajouter des étapes supplémentaires de regroupement de pixels en instances.

A.5.5 Méthode de détection

Ils ont développés une approche "single-shot" de segmentation d'instance de personne. Pour une image donnée, cela consiste d'abord à classifier chaque pixel comme appartenant à une personne ou au fond de l'image, puis à regrouper les pixels qualifiés comme personne dans une instance de personne.

Segmentation sémantique de personne

Ils utilisent un réseau de neurone de convolution standard pour faire la segmentation. Ils prédissent pour chaque emplacement de pixel, la probabilité qu'il appartient à une instance de personne.

Proposition de boîte de détection

Dans le but de prédire l'instance de personne à laquelle appartient chaque pixel, chaque emplacement de pixel est associé à une "proposition" ou à une "ancre". Une "ancre" est une boîte englobante qui est centré sur ce pixel et à une largeur w et une hauteur h.

Pour chaque pixel, le réseau prédit les décalages (dx, dy, dw, dh) entre sa boîte d'ancrage et la boîte englobante de l'instance à laquelle il appartient.

Regroupement de pixels en instance

Dans cet article, les auteurs ont choisi de faire correspondre les coordonnées des pixels pour qu'elles soient comprises dans les coordonnées du cadre de sélection de l'instance à laquelle appartient chaque pixel.

Cette méthode n'est pas parfaite, en effet, il est possible que plusieurs instances se chevauchent et que les boîtes englobantes soient presque identiques.

Le point positif est que cette méthode est facile à implémenter à la suite des architectures de segmentation sémantique déjà existante.

La méthode regroupement des pixels à 2 étapes qui se suivent :

1. La sélection de boîte globale est la première étape. Les auteurs traitent la valeur de la probabilité attribuée à chaque emplacement de pixel comme la valeur de confiance associée au cadre de sélection prévu pour cet emplacement. Ils recueillent toutes les boîtes englobantes prédites qui correspondent aux pics locaux de la carte de segmentation sémantique et ont un indice de confiance supérieur à un seuil (t = 0.6). La suppression non-maximale est ensuite appliquée aux boîtes englobantes collectées

pour obtenir les détections globales de la boîte englobante Bg pour l'image donnée.

2. L'assignation de pixel à une instance est la seconde étape. Sp est l'ensemble des pixels trouvés comme personne. Chacun de ses pixels a besoin d'être assignés à une des instances globales des boîtes englobantes Bg de l'étape précédente. Pour chaque location de pixel xi dans Sp, ils prennent la boîte englobante correspondante bi, et effectue l'intersection sur l'union entre bi et chacune des boîtes englobantes globale Bg. La localisation du pixel est ensuite assignée à la boîte englobante Bg. Si toutes les boîtes de Bg se chevauchent avec bi avec un score IoU inférieur à un seuil tiou, alors la localisation du pixel xi est supprimée de Sp. Ce pixel est supposé être un résultat faux positif de la segmentation sémantique et n'est attribué à aucune des instances.

A.5.6 Jeu de données

Ils ont utilisé le jeu de données COCO pour l'apprentissage et l'évaluation. Ils ont réalisé l'apprentissage seulement avec les images d'apprentissage contenant des annotations de personnes, soit 64 115 images.

A.5.7 Architecture du modèle

Les auteurs utilisent le réseau de base ResNet-50, qu'ils ont choisi pour son équilibre riche en fonctionnalités et sa consommation de mémoire, auquel ils y attachent le module Atrous Spatial Pyramid Pooling et les couches de décodeurs DeepLabv3 +.

La seule divergence par rapport à DeepLabv3 + réside dans le fait qu'en plus de la couche de convolution finale 1x1 avec un filtre par classe (dans leur cas, il n'existe qu'une seule classe) au-dessus des cartes de caractéristiques de sortie du décodeur, ils disposent d'une couche de convolution supplémentaire 1x1 avec quatre filtres pour prédire les décalages de la boîte englobante dense.

A.5.8 Conclusion

Cette article présente une méthode unique pour la segmentation d'instances d'objets et montré son efficacité lors de la segmentation d'instances de personnes.

B

Planification

La figure B.1 présente le planning élaboré *a priori*...

La figure B.2 présente pour sa part le planning relevé au fur et à mesure de l'avancement du travail.

Discuter les différences entre les deux plannings et les leçons apprises sur la gestion d'un projet de recherche ou de R&D.

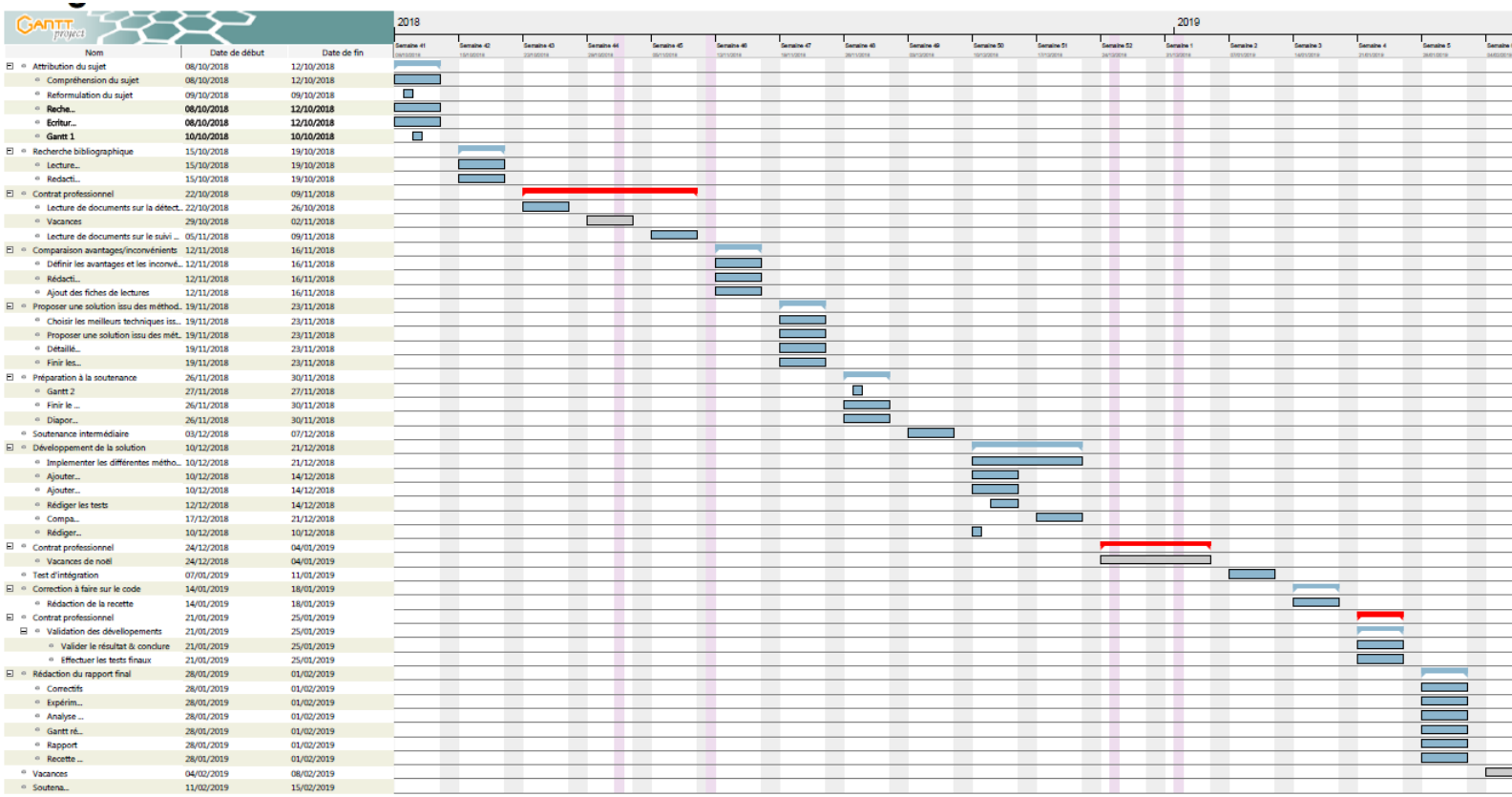

FIGURE B.1 : Planification prévisionnelle

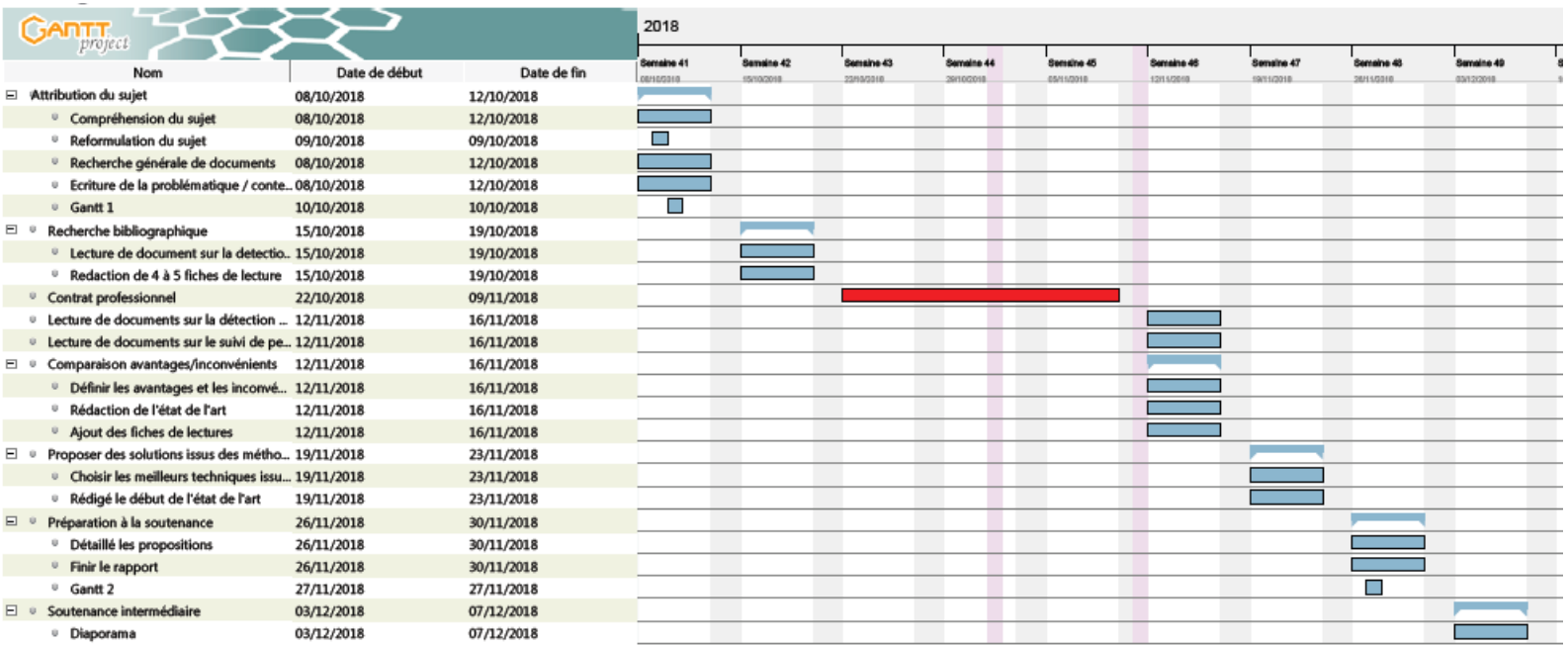

FIGURE B.2 : Planning effectif

Fiches de suivi

Fiche de suivi de la semaine 1
du 8 octobre 2018 au 12 octobre 2018

Temps de travail de Élodie BOUILLETEAU: 12 h 30 m
Travail effectué.

- tâche 1 : Compréhension du sujet ; simplicité ; achevée.
- tâche 2 : Reformulation du sujet ; difficulté : moyenne ; achevée.
- tâche 3 : Recherche générale du sujet sur Wikipédia ; simplicité ; achevée.
- tâche 4 : Écriture du contexte, problématique ; simplicité ; achevée.
- tâche 5 : Lecture de quelques documents (thèses, articles...) sur le sujet ; difficulté : moyenne ; réalisée à 30 %

Travail non effectué.
Échanges avec le commanditaire.

- Est ce que l'on garde l'analyse en temps réel ? ; Est ce que la restitution visuelle des résultats est importante ?
- Le temps réel n'est pas une priorité. ; La restitution est utile car elle permet de présenter le résultat au sein du pôle innovation de U GIE IRIS ;
- Les indicateurs sont trop nombreux. Il faut se spécialisé dans un niveau ;

Planification pour la semaine prochaine.

- recherches à effectuer ;
- articles à lire, comprendre et analyser ;
- proposé des indicateurs pertinents ;

Fiche de suivi de la semaine 2
du 15 octobre 2018 au 19 octobre 2018

Temps de travail de Élodie BOUILLETEAU: 10 h 30 m
Travail effectué.

- tâche 1 : Fiche de lecture sur deux documents scientifiques traitant de la méthode Viola-Jones ; difficulté : moyenne ; achevée.
- tâche 2 : Précision des problèmes à résoudre ; difficulté : moyenne ; achevée.
- tâche 3 : Recherche d'une banque de données de vidéos libre de droits ; réalisée à 75 % ; achevée.

Échanges avec le commanditaire.

- Précision des problèmes du sujet avec Nicolas NORMAND ;
  1. Méthode de détection de visage.
  2. Méthode d'identification de 2 personnes identiques d'une image à l'autre.
  3. Méthode de traçage de personne relié à la méthode d'identification.
  4. Sous-problème : la gestion des occlusions.
- Recherche générale sur toutes les méthodes possibles et performantes lié à notre sujet ;

Planification pour la semaine prochaine.

- Recherche sur les méthodes de détection, de traçage et d'identification récente et performante à effectuer ;
- articles à lire, comprendre et analyser ;

Fiche de suivi de la semaine 3
du 12 novembre 2018 au 18 novembre 2018

Temps de travail de Élodie BOUILLETEAU: 11 h 00 m
Travail effectué.

- tâche 1 : Fiche de lecture sur la méthode de détection YOLO ; difficulté : moyenne ; achevée.
- tâche 2 : Fiche de lecture 2 méthode de détection de personnes en temps réel ; difficulté : moyenne ; achevée.
- tâche 3 : Recherche d'une banque de données de vidéos libres de droits ; achevée.
- tâche 4 : Recherche d'article sur le tracking d'objet avec code source ; simplicité ; achevée.

Travail non effectué.

- tâche 1 : Rédaction de l'état de l'art, recherche bibliographique non terminée ;

Échanges avec le commanditaire.

- Concentration sur des méthodes avec le code source open source et réutilisable ;

Planification pour la semaine prochaine.

- Lire les articles de tracking de personne et en faire des fiches de lecture (3 max) ;
- Faire une présentation power-point d'un article ;
- Rédaction de l'état de l'art ;
- Faire le diagramme de Gantt ;
- Faire un Excel synthétique de toutes les solutions possibles avec leurs critères de sélection ;

Fiche de suivi de la semaine 4
du 19 novembre 2018 au 25 novembre 2018

Temps de travail de Élodie BOUILLETEAU: 12 h 15 m
Travail effectué.

- tâche 1 : Rédaction de l'état de l'art ; difficulté : moyenne ; achevée.
- tâche 2 : Recherche d'article sur le tracking d'objet avec code source ; simplicité ; achevée.
- tâche 3 : Présentation des 3 propositions ; simplicité ; achevée.

Travail non effectué.

- tâche 1 : Rédaction de la fiche de suivi de la semaine 47, oublie ;

Échanges avec le commanditaire.

- Explication des différentes approches sur le suivi de personnes ;

Planification pour la semaine prochaine.

- Finir la rédaction de l'état de l'art ;
- Rédiger la partie sur les propositions (3 max) ;
- Faire le diagramme de Gantt ;

Fiche de suivi de la semaine 5
du 26 novembre 2018 au 2 décembre 2018

Temps de travail de Élodie BOUILLETEAU: 12 h 00 m
Travail effectué.

- tâche 1 : Rédaction de l'état de l'art ; difficulté : moyenne ; achevée.
- tâche 2 : Rédiger la partie sur les propositions (3 max) ; difficile ; réalisée à 0 %.
- tâche 3 : Faire le diagramme de Gantt ; moyenne ; réalisée à 0 %.

Travail non effectué.

- tâche 1 : Présentation power-point , reporter au lundi de la semaine prochaine ;

Planification pour la semaine prochaine.

- Présentation power-point pour la soutenance ;

Le tableau C.1 récapitule le taux d'avancement du projet. Rappelons que le temps de travail théorique *minimal* correspond au temps indiqué sur la maquette pédagogique auquel on ajoute un strict minimum de 20 % correspondant au travail personnel hors emploi du temps. La partie « haute » de la fourchette correspond à 50 % de temps supplémentaire au titre du travail personnel.

	Temps prévu		Élodie BOUILLETEAU					
	bas	haut	hebdo.	Σ	%	hebdo.	Σ	%
Semaine	h : m	h : m	h : m	h : m		h : m	h : m	
1	10 : 00	12 : 30	12 : 30	12 : 30	125 (100)	:	:	
2	20 : 00	25 : 00	10 : 30	23 : 00	115 (92)	:	:	
3	30 : 00	37 : 30	11 : 00	34 : 00	113 (90)	:	:	
4	40 : 00	50 : 00	12 : 15	46 : 15	115 (92)	:	:	
5	50 : 00	62 : 30	12 : 00	58 : 15	116 (93)	:	:	

TABLE C.1 : Avancement du projet par rapport au temps de travail théorique minimal (respectivement haut)

D

Auto-contrôle et auto-évaluation

La figure D.1 permet d'énumérer un certain nombre de points importants dans les trois composantes du travail :

1. rapport ;
2. présentation orale ;
3. travail de fond ;

ainsi que d'évaluer notre niveau de satisfaction à l'issue de la phase I, composée de trois étapes :

1. étude préalable ;
2. étude bibliographique ;
3. conception générale.

Les points de satisfaction ou d'insatisfaction peuvent être approfondis.

PRD	Prénom	Nom
Binome 1		
Binome 2		

Notation	
A	Maîtrise dans l'application du savoir-faire requis
B	Application du savoir-faire requis
C	Insuffisances, lacunes à corriger dans l'application du savoir-faire requis *a minima*
D	Insuffisances flagrantes, inacceptables, voire travail absent

PROPOSITION DE NOTE (I)	**#DIV/0!**
PROPOSITION DE NOTE (II)	**#DIV/0!**
NOTE PROJET	**#NOM ?**

A	B	C	D	Remarque / Note / Commentaire

Phase I : Etude préalable, étude bibliographique et conception générale

Rapport	Organisation	Plan	Equilibre	?
			Cohérence	?
		Fluidité	Introductions (partielles)	?
			Transitions	?
			Conclusions (partielles)	?
		Tableaux, figures	Numérotés	?
			Légendés	?
			Référencés (non "en ligne")	?
	Rédaction	Orthographe	Coquilles	?
			Fautes évitables	?
			Franglais, jargon	?
		Rédaction	Aérée	?
			Absence de plagiat !	?
	Bibliographie	Références	Suffisantes (nombre, intérêt)	?
			Pérennes	?
			Complètes (auteurs, pages...)	?
			Conséquentes (volume)	?
		Références dans le texte		?

Proposition de note haute	0,00
Proposition de note basse	0,00
Proposition de note du jury	

Projection	Organisation	Plan	?
		Liaisons	?
		Numérotation	?
	Contenu	Informatif	?
		Concis	?
		Clair	?
		Orthographe	?
		Illustrations	?
Oral	Présentation	Aisance	?
		Tenue	?
		Articulation, compréhension	?
	Durée	Respect	?
		Temps de parole équilibré	?
	Réponses	Pertinence	?
		Argumentation	?

Proposition de note haute	0,00
Proposition de note basse	0,00
Proposition de note du jury	

Travail	Etude	Bibliographie	Adéquate	?
			Suffisante	?
		Cahier des charges	Clair	?
			Formalisé	?
		Hypothèses envisagées	Nombre	?
			Pertinence	?
			Analyse *a priori*	?
		Validation	Tableau comparatif	?
			Choix argumenté(s)	?
			Faisabilité	?
	Complexité	Temps consacré		?
		Résultats obtenus		?
		Difficulté	Intrinsèque	?
			Vis-à-vis du binôme	?
	Annexes	Fiches d'avancement	Régularité	?
			Détaillées	?
		Gantt	Prévisionnel et justifications	?
			Effectifs, erreurs, corrections	?

Proposition note haute	0,00
Proposition note basse	0,00
Proposition de note du jury	

Phase II : Conception détaillée, réalisation et recette

Rapport	Organisation	Plan	Equilibre	?
			Cohérence	?
		Fluidité	Introductions (partielles)	?
			Transitions	?
			Conclusions (partielles)	?
		Tableaux, figures	Numérotés	?
			Légendés	?
			Référencés (non "en ligne")	?
	Rédaction	Orthographe	Coquilles	?
			Fautes évitables	?
			Franglais, jargon	?
		Rédaction	Aérée	?
			Absence de plagiat !	?

Proposition de note haute	0,00
Proposition de note basse	0,00
Proposition de note du jury	

Projection	Organisation	Plan	?
		Liaisons	?
		Numérotation	?
	Contenu	Informatif	?
		Concis	?
		Clair	?
		Orthographe	?
		Illustrations	?
Oral	Présentation	Aisance	?
		Tenue	?
		Articulation, compréhension	?
	Durée	Respect	?
		Temps de parole équilibré	?
	Réponses	Pertinence	?
		Argumentation	?

Proposition de note haute	0,00
Proposition de note basse	0,00
Proposition de note du jury	

Travail	Conception	Générale	Clarté	?
			Niveau de détail adéquat	?
			Argumentation	?
			Validation	?
		Détaillée	Formalisation adéquate (algorithme...)	?
			Validation	?
	Réalisation	Développement	"Fonctionnalités"	?
			Volume, environnement...	?
		Tests (unitaires, d'intégration, fonctionnel, de performance...)		?
	Recette	Livraison du paquetage		?
		Redéploiement du paquetage		?
	Complexité	Temps consacré		?
		Résultats obtenus		?
		Difficulté	Intrinsèque	?
			Vis-à-vis du binôme	?
	Annexes	Fiches d'avancement	Régularité	?
			Détaillées	?
		Gantt	Prévisionnel et justifications	?
			Effectif, erreurs, corrections	?

Proposition de note haute	0,00
Proposition de note basse	0,00
Proposition de note du jury	

FIGURE D.1 : Points à contrôler à l'issue de la phase I