

SUIVI SIMPLE EN LIGNE ET EN TEMPS RÉEL AVEC UNE MÉTRIQUE D'ASSOCIATION PROFONDE

Nicolai Wojke†, Alex Bewley, Dietrich Paulus†

Université de Koblenz-Landau †, Université de technologie du Queensland

ABSTRAIT

Le suivi simple en ligne et en temps réel (SORT) est une approche pragmatique du suivi d'objets multiples en mettant l'accent sur des algorithmes simples et efficaces. Dans cet article, nous intégrons des informations d'apparence pour améliorer les performances de SORT. Grâce à cette extension, nous sommes en mesure de suivre les objets pendant des périodes d'occlusion plus longues, réduisant ainsi le nombre de commutateurs d'identité. Dans l'esprit du cadre d'origine, nous plaçons une grande partie de la complexité de calcul dans une étape de pré-formation en fil où nous apprenons une métrique d'association profonde sur un ensemble de données de ré-identification de personne à grande échelle. Lors de l'application en ligne, nous établissons des associations mesure-piste en utilisant les requêtes du voisin le plus proche dans l'espace d'apparence visuelle. L'évaluation expérimentale montre que nos extensions réduisent le nombre de changements d'identité de 45%,

Termes de l'index - Vision par ordinateur, suivi d'objets multiples, association de données

1. INTRODUCTION

En raison des progrès récents de la détection d'objets, le suivi par détection est devenu le principal paradigme du suivi d'objets multiples. Dans ce paradigme, les trajectoires d'objets se trouvent généralement dans un problème d'optimisation global qui traite des lots vidéo entiers à la fois. Par exemple, les formules de réseaux de flux [1, 2, 3] et les modèles graphiques probabilistes [4, 5, 6, 7] sont devenus des cadres populaires de ce type. Cependant, en raison du traitement par lots, ces méthodes ne sont pas applicables dans les scénarios en ligne où une identité cible doit être disponible à chaque étape de temps. Les méthodes plus traditionnelles sont le suivi d'hypothèses multiples (MHT) [8] et le filtre d'association de données probabilistes conjoint (JPDAF) [9]. Ces méthodes effectuent une association de données image par image. Dans le JPDAF, une hypothèse d'état unique est générée en pondérant des mesures individuelles par leurs probabilités d'association. Dans MHT, toutes les hypothèses possibles sont suivies, mais des schémas d'élagage doivent être appliqués pour la traitabilité des calculs. Les deux méthodes ont été récemment revisitées dans un scénario de suivi par détection [10, 11] et ont donné des résultats prometteurs. Cependant, les performances de ces méthodes s'accompagnent d'une complexité de calcul et de mise en œuvre accrue.

Le suivi simple en ligne et en temps réel (SORT) [12] est un

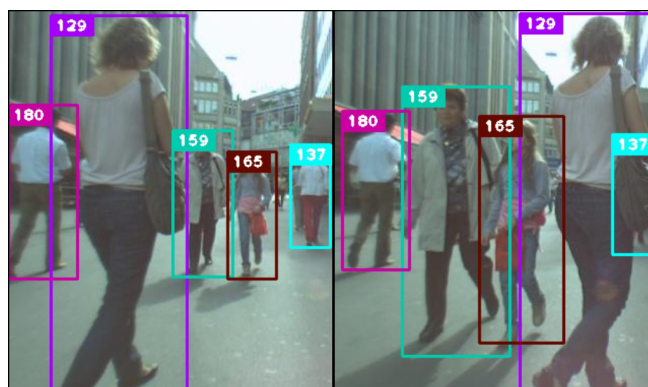


Fig. 1: Exemple de sortie de notre méthode sur le jeu de données de défi MOT [15] dans une situation de suivi commune avec occlusion fréquente.

cadre beaucoup plus simple qui effectue le filtrage de Kalman dans l'espace image et l'association de données image par image en utilisant la méthode hongroise avec une métrique d'association qui mesure le chevauchement de la boîte englobante. Cette approche simple permet d'obtenir des performances favorables à des fréquences d'images élevées. Sur l'ensemble de données de défi MOT [13], SORT avec un détecteur de personnes à la pointe de la technologie [14] se classe en moyenne plus haut que MHT sur les détections standard. Ceci souligne non seulement l'influence des performances du détecteur d'objet sur les résultats globaux du suivi, mais c'est aussi un aperçu important du point de vue des praticiens.

Tout en obtenant de bonnes performances globales en termes de précision et d'exactitude de suivi, SORT renvoie un nombre relativement élevé de commutateurs d'identité. En effet, la métrique d'association employée n'est précise que lorsque l'incertitude d'estimation de l'état est faible. Par conséquent, SORT a un défaut de suivi à travers les occlusions comme ils apparaissent généralement dans les scènes de caméra à vue frontale. Nous surmontons ce problème en remplaçant la métrique d'association par une métrique plus informée qui combine les informations de mouvement et d'apparence. En particulier, nous appliquons un réseau de neurones convolutifs (CNN) qui a été formé pour discriminer les piétons sur un ensemble de données de ré-identification de personne à grande échelle. Grâce à l'intégration de ce réseau, nous augmentons la robustesse contre les erreurs et les occlusions tout en maintenant le système facile à mettre en œuvre, efficace et applicable aux scénarios en ligne.

2. SORTWITH DEEP ASSOCIATION METRIC

Nous adoptons une méthodologie conventionnelle de suivi d'hypothèse unique avec filtrage récursif de Kalman et association de données image par image. Dans la section suivante, nous décrivons plus en détail les principaux composants de ce système.

2.1. Gestion des pistes et estimation de l'état

La gestion des pistes et le cadre de filtrage de Kalman sont pratiquement identiques à la formulation originale de [12]. Nous supposons un scénario de suivi très général où la caméra n'est pas calibrée et où nous n'avons aucune information sur le mouvement de l'ego disponible. Bien que ces circonstances posent un défi au cadre de filtrage, c'est la configuration la plus courante prise en compte dans les récents benchmarks de suivi d'objets multiples [15]. Par conséquent, notre scénario de suivi est défini sur l'espace d'états à huit dimensions $(u, v, \gamma, h, \dot{x}, \dot{y}, \dot{\gamma}, \dot{h})$ qui contient la position centrale du cadre de sélection (u, v) , rapport hauteur / largeur, hauteur

γ , et leurs vitesses respectives en coordonnées d'image. Nous utilisons un filtre de Kalman standard avec un mouvement à vitesse constante et un modèle d'observation linéaire, où nous prenons les coordonnées de délimitation (u, v, γ, h) comme observations directes de l'état de l'objet.

Pour chaque piste k on compte le nombre d'images depuis la dernière association de mesure réussie une_k . Ce compteur est incrémenté lors de la prédiction du filtre de Kalman et remis à 0 lorsque la trace a été associée à une mesure.

Pistes dépassant un âge maximum prédéfini UNE_{max} sont considérés comme ayant quitté la scène et sont supprimés de la piste ensemble. De nouvelles hypothèses de trace sont lancées pour chaque détection qui ne peut pas être associée à une trace existante. Ces nouvelles pistes sont classées comme provisoires lors de leurs trois premières images. Pendant ce temps, nous nous attendons à une association de mesure réussie à chaque pas de temps. Les pistes qui ne sont pas correctement associées à une mesure dans leurs trois premières images sont supprimées.

2.2. Problème d'affectation

Une manière conventionnelle de résoudre l'association entre les états de Kalman prédits et les mesures nouvellement arrivées est de construire un problème d'affectation qui peut être résolu en utilisant l'algorithme hongrois. Dans cette formulation de problème, nous intégrons des informations de mouvement et d'apparence par la combinaison de deux mesures appropriées.

Pour incorporer les informations de mouvement, nous utilisons la distance de Mahalanobis (au carré) entre les états de Kalman prédits et les mesures nouvellement arrivées:

$$r\acute{e}(1)(i, j) = (r\acute{e}_j - y_{je})^T S_{je}^{-1} (r\acute{e}_j - y_{je}), \quad (1)$$

où nous désignons la projection du j -distribution des pistes dans l'espace de mesure par (y_{je}, S_{je}) et le j -détection de la boîte englobante par $r\acute{e}_j$. La distance de Mahalanobis prend en compte l'incertitude d'estimation de l'état en mesurant combien

écarts types la détection est éloignée de l'emplacement moyen de la trace. En outre, en utilisant cette métrique, il est possible d'exclure des associations improbables en limitant la distance Mahalanobis à un 95% intervalle de confiance calculé à partir de l'inverse

χ^2 Distribution. Nous notons cette décision par un indicateur

$$b_{i,j}^{(1)} = 1[r\acute{e}(1)(i, j) \leq t^{(1)}] \quad (2)$$

qui évalue à 1 si l'association entre le j -th piste et i -La détection est admissible. Pour notre espace de mesure à quatre dimensions, le seuil de Mahalanobis correspondant est $t^{(1)} = 9.4877$.

Alors que la distance de Mahalanobis est une métrique d'association appropriée lorsque l'incertitude de mouvement est faible, dans notre formulation de problème d'espace image, la distribution d'état prédite obtenue à partir du cadre de filtrage de Kalman ne fournit qu'une estimation approximative de l'emplacement de l'objet. En particulier, un mouvement de caméra non pris en compte peut introduire des déplacements rapides dans le plan de l'image, faisant de la distance de Mahalanobis une métrique plutôt non renseignée pour le suivi des occlusions. Par conséquent, nous intégrons une deuxième métrique dans le problème d'affectation. Pour

chaque détection de boîte englobante $r\acute{e}_j$, nous calculons un descripteur d'apparence r avec $\|r\| = 1$. De plus, nous gardons une galerie $R =$

$\{r_k\}_{k=1}^L$ du dernier $L_k = 100$ descripteurs d'apparence associés pour chaque piste k . Ensuite, notre deuxième mesure métrique la plus petite distance cosinus entre le j -th piste et j -ème détection dans l'espace d'apparence:

$$r\acute{e}(2)(i, j) = \min \{1 - r^T r_k / (r^T r_k) \mid k \in R_{je}\}. \quad (3)$$

Encore une fois, nous introduisons une variable binaire pour indiquer si une association est admissible selon cette métrique

$$b_{i,j}^{(2)} = 1[r\acute{e}(2)(i, j) \leq t^{(2)}] \quad (4)$$

et nous trouvons un seuil approprié pour cet indicateur sur un ensemble de données d'apprentissage séparé. En pratique, nous appliquons un CNN pré-entraîné pour calculer les descripteurs d'apparence de boîte englobante. L'architecture de ce réseau est décrite dans la section 2.4.

En combinaison, les deux mesures se complètent en servant différents aspects du problème d'affectation. D'une part, la distance de Mahalanobis fournit des informations sur les emplacements possibles des objets en fonction du mouvement qui sont particulièrement utiles pour les prévisions à court terme. D'autre part, la distance cosinus prend en compte les informations d'apparence qui sont particulièrement utiles pour récupérer des identités après des occlusions à long terme, lorsque le mouvement est moins discriminant. Pour créer le problème d'association, nous combinons les deux métriques en utilisant une somme pondérée

$$c_{i,j} = \lambda d^{(1)}(i, j) + (1 - \lambda) d^{(2)}(i, j) \quad (5)$$

où nous appelons une association un \square irrecevable si elle est dans la région des deux métriques:

$$b_{je,j}^{(2)} \quad b_{i,j}^{(m)} \quad (6)$$

Contribution: Suivi des Indices $T = \{1, \dots, N\}$, Indices de détection $ré$
 $\{1, \dots, M\}$, Age maximum UNE_{max}

- 1: Calculer la matrice des coûts $C = [c_{i,j}]$ en utilisant Eq. 5
- 2: calcul de la matrice de porte $B = [b_{i,j}]$ en utilisant Eq. 6 3: Initialiser l'ensemble de correspondances $M \leftarrow \emptyset$
- 4: Initialiser un ensemble de détections inégales $U \leftarrow D$
- 5: pour $n \in \{1, \dots, UNE_{max}\}$ faire
 - 6: Sélectionnez les morceaux par âge $T_n \leftarrow \{j \in T \mid une_i = n\}$
 - sept: $[X_{i,j}] \leftarrow \quad \quad \quad$
 - 8: $M \leftarrow M \cup \{ \text{col}(U, n), b_{i,j} \in \text{col}(C, T_n, U) \}$
 - 9: $U \leftarrow U \setminus \{j \mid \sum_{j \in b_{j \in j} \cdot X_{i,j} > 0\}$
- dix: fin pour
- 11: revenir M, U

2.3. Cascade assortie

Le listing 1 décrit notre algorithme d'appariement. En entrée, nous fournissons l'ensemble de la piste T et détection $r \in$ indices ainsi que l'âge maximum UNE_{\max} . Aux lignes 1 et 2, nous calculons la matrice de coût d'association et la matrice d'association. Nous itérons ensuite sur l'âge de la piste r pour résoudre un problème d'affectation linéaire pour des pistes d'âge croissant. En ligne 6, nous sélectionnons le sous-ensemble de pistes T_n qui n'ont pas été associés à une détection dans le dernier n cadres. À la ligne 7, nous résolvons le linéaire comme-signalisation entre les pistes dans T_n et des détections inégales U .

Nom	Taille du patch / foulée	Taille de sortie
Conv 1	$3 \times 3/1$	$32 \times 128 \times 64$
Conv 2	$3 \times 3/1$	$32 \times 128 \times 64$
Piscine Max 3	$3 \times 3/2$	$32 \times 64 \times 32$
Résiduel 4	$3 \times 3/1$	$32 \times 64 \times 32$
Résiduel 5	$3 \times 3/1$	$32 \times 64 \times 32$
Résiduel 6	$3 \times 3/2$	$64 \times 32 \times 16$
Résiduel 7	$3 \times 3/1$	$64 \times 32 \times 16$
Résiduel 8	$3 \times 3/2$	$128 \times 16 \times 8$
Résiduel 9	$3 \times 3/1$	$128 \times 16 \times 8$
Dense 10		128
Lot et γ_2 normalisation		128

Aux lignes 8 et 9, nous mettons à jour l'ensemble des correspondances et des détections inégales, que nous renvoyons après achèvement à la ligne 11. Notez que cette cascade de correspondance donne la priorité aux pistes de plus petit âge, c'est-à-dire aux pistes qui ont été vues plus récemment.

Cela permet de tenir compte des changements d'apparence soudains, par exemple dus à une occlusion partielle avec une géométrie de scène statique, et d'augmenter la robustesse contre une initialisation erronée.

En utilisant de simples requêtes de voisinage le plus proche sans apprentissage métrique supplémentaire, une application réussie de notre méthode nécessite une intégration de fonctionnalité bien discriminante pour être entraînée à plat, avant l'application de suivi en ligne réelle. À cette fin, nous utilisons un CNN qui a été formé sur un ensemble de données de ré-identification de personnes à grande échelle [21] qui contient plus de 1 100 000 images de 1 261 piétons, ce qui le rend bien adapté à l'apprentissage métrique profond dans un suivi des personnes. le contexte.

L'architecture CNN de notre réseau est présentée dans le tableau 1. En résumé, nous employons un large réseau résiduel [22] avec deux couches convolutives suivies de six blocs résiduels. La carte globale des caractéristiques de la dimensionnalité 128 est calculé en couche dense 10. Un dernier lot et ℓ_2 la normalisation projette des fonctionnalités sur l'hypersphère de l'unité pour être compatible avec notre métrique d'apparence cosinus. Au total, le réseau a 2800864 paramètres et un passage avant de 32 boîtes englobantes prend environ 30 ms sur un GPU mobile Nvidia GeForce GTX 1050. Ainsi, ce réseau est bien adapté pour le suivi en ligne, à condition qu'un GPU moderne soit disponible. Bien que les détails de notre procédure de formation sortent du cadre de cet article, nous fournissons un modèle pré-formé dans notre dépôt GitHub.

		MOTA \uparrow	MOTP \uparrow	MT \uparrow	ML \downarrow	ID \downarrow	FM \downarrow	FP \downarrow	FN \downarrow	Durée \uparrow
KDNT [16]?	LOT	68,2	79,4	41,0%	19,0%	933	1093	11479	45605	0,7 Hz
LMP p.[17]?	LOT	71,0	80,2	46,9%	21,9%	434	587	7880	44564	0,5 Hz
MCMOT HDM [18]	LOT	62,4	78,3	31,5%	24,2%	1394	1318	9855	57257	35 Hz
NOMTwSDP16 [19]	LOT	62,2	79,6	32,5%	31,1%	406	642	5119	63352	3 Hz
EAMTT [20]	EN LIGNE	52,5	78,8	19,0%	34,9%	910	1321	4407	81223	12 Hz
POI [16]?	EN LIGNE	66,1	79,5	34,0%	20,8%	805	3093	5061	55914	10 Hz
TRIER [12]?	EN LIGNE	59,8	79,6	25,4%	22,7%	1423	1835	8698	63245	60 Hz
Deep SORT (le nôtre)?	EN LIGNE	61,4	79,1	32,8%	18,2%	781	2008	12852	56668	40 Hz

Tableau 2: Suivi des résultats sur le défi MOT16 [15]. Nous comparons à d'autres méthodes publiées avec des détections non standard. Le tableau complet des résultats se trouve sur le site Web du défi. Méthodes marquées d'un? utiliser les détections fournies par [16].

tory [†] avec un script qui peut être utilisé pour générer des fonctionnalités.

3. EXPÉRIENCES

Nous évaluons les performances de notre tracker sur le benchmark MOT16 [15]. Cette référence évalue les performances de suivi sur sept séquences de test difficiles, y compris des scènes de vue frontale avec une caméra en mouvement ainsi que des configurations de surveillance descendante. Comme entrée de notre outil de suivi, nous nous appuyons sur les détections fournies par Yu et al. [16]. Ils ont formé un RCNN Faster sur une collection d'ensembles de données publics et privés pour fournir d'excellentes performances. Pour une comparaison équitable, nous avons relancé SORT sur les mêmes détections.

L'évaluation des séquences de test a été réalisée en utilisant $\lambda = 0$ et $UNE_{\max} = 30$ cadres. Comme dans [16], les détections ont été seuillées à un score de confiance de 0,3. Le pa-
Les paramètres de notre méthode ont été trouvés sur des séquences de formation distinctes fournies par le benchmark. L'évaluation est effectuée selon les métriques suivantes:

- Précision de suivi multi-objets (MOTA): Résumé de la précision globale du suivi en termes de faux positifs, de faux négatifs et de commutateurs d'identité [23].
- Précision de suivi multi-objets (MOTP): Résumé de la précision de suivi globale en termes de chevauchement de la boîte englobante entre la vérité terrain et l'emplacement signalé [23].
- Surtout suivis (MT): pourcentage de pistes de vérité terrain ayant le même libellé pendant au moins 80% de leur durée de vie.
- Principalement perdu (ML): pourcentage de pistes de vérité terrain suivies pendant au plus 20% de leur durée de vie.
- Commutateurs d'identité (ID): nombre de fois où l'identité rapportée d'une piste de vérité terrain change.
- Fragmentation (FM): nombre de fois qu'une piste est interrompue par une détection manquante.

Les résultats de notre évaluation sont présentés dans le tableau 2. Nos adaptations réduisent avec succès le nombre de commutateurs d'identité. Par rapport à SORT, les commutateurs ID passent de 1423 à 781. Il s'agit d'une diminution d'environ 45%. En même temps,

la fragmentation des traces augmente légèrement en raison du maintien des identités des objets par des occlusions et des ratés. Nous constatons également une augmentation significative du nombre d'objets pour la plupart suivis et une diminution des objets pour la plupart perdus. Dans l'ensemble, grâce à l'intégration des informations d'apparence, nous réussissons à maintenir les identités grâce à des occlusions plus longues. Cela peut également être vu par une analyse qualitative de la sortie de suivi que nous fournissons dans le matériel supplémentaire. Un exemple de sortie de notre tracker est illustré à la figure 1.

Notre méthode est également un concurrent sérieux par rapport aux autres cadres de suivi en ligne. En particulier, notre approche renvoie le plus petit nombre de changements d'identité de toutes les méthodes en ligne tout en conservant des scores MOTA compétitifs, en suivant les fragmen- tations et les faux négatifs. La précision de suivi signalée est principalement affectée par un plus grand nombre de faux positifs. Compte tenu de leur impact global sur le score MOTA, l'application d'un seuil de confiance plus grand aux détections peut potentiellement augmenter les performances rapportées de notre algorithme dans une large mesure. Cependant, une inspection visuelle de la sortie de suivi montre que ces faux positifs sont principalement générés à partir de réponses sporadiques du détecteur à une géométrie de scène statique. En raison de notre âge de piste maximal autorisé relativement grand, ceux-ci sont plus communément joints aux trajectoires d'objets. En même temps, nous n'avons pas observé de pistes sautant fréquemment entre les fausses alarmes. Au lieu de cela, le traqueur générerait généralement des pistes stationnaires relativement stables à l'emplacement de l'objet signalé.

Notre implémentation fonctionne à environ 20 Hz avec environ la moitié du temps consacré à la génération de fonctionnalités. Par conséquent, étant donné un GPU moderne, le système reste efficace en termes de calcul et fonctionne en temps réel.

4. CONCLUSION

Nous avons présenté une extension de SORT qui incorpore des informations d'apparence grâce à une métrique d'association pré-entraînée. Grâce à cette extension, nous sommes en mesure de suivre des périodes d'occlusion plus longues, faisant de SORT un concurrent sérieux pour les algorithmes de suivi en ligne de pointe. Pourtant, l'algorithme reste simple à implémenter et s'exécute en temps réel.

[†] https://github.com/nwojke/deep_sort

5. RÉFÉRENCES

- [1] L. Zhang, Y. Li et R. Nevatia, «Global data association pour le suivi multi-objets à l'aide de flux réseau », dans *CVPR*, 2008, pp. 1–8.
- [2] H. Pirsiavash, D. Ramanan et CC Fowlkes, «Algorithmes gloutons optimaux à l'échelle mondiale pour suivre un nombre variable d'objets», dans *CVPR*, 2011, p. 1201–1208.
- [3] J. Berclaz, F. Fleuret, E. Türetken et P. Fua, «Multi-suivi des objets grâce à l'optimisation des k chemins les plus courts, » *IEEE Trans. Modèle Anal. Mach. Intell.*, vol. 33, non. 9, pp. 1806–1819, 2011.
- [4] B. Yang et R. Nevatia, «Un modèle CRF appris en ligne pour le suivi multi-cibles, "dans *CVPR*, 2012, pp. 2034–2041.
- [5] B. Yang et R. Nevatia, «Multi-target tracking by on-apprentissage en ligne de modèles de mouvement non linéaires et de modèles d'apparence robustes, » dans *CVPR*, 2012, pp. 1918-1925.
- [6] A. Andriyenko, K. Schindler et S. Roth, «Discrete-optimisation continue pour le suivi multi-cibles », dans *CVPR*, 2012, pp. 1926–1933.
- [7] A. Milan, K. Schindler et S. Roth, «Detection- and exclusion au niveau de la trajectoire dans le suivi d'objets multiples, "dans *CVPR*, 2013, pp. 3682–3689.
- [8] DB Reid, "Un algorithme pour suivre plusieurs cibles", *IEEE Trans. Autom. Contrôle*, vol. 24, non. 6, pp. 843–854, 1979.
- [9] TE Fortmann, Y. Bar-Shalom et M. Scheffe, «Sonar suivi de plusieurs cibles à l'aide d'une association de données probabilistes conjointe, " *IEEE J. Ocean. Eng.*, vol. 8, non. 3, pp. 173-184, 1983.
- [10] C. Kim, F. Li, A. Ciptadi et JM Rehg, «Multiple suivi des hypothèses revisitée, "dans *ICCV*, 2015, pp. 4696–4704.
- [11] SH Rezatofighi, A. Milan, Z. Zhang, Qi. Shi, An. Queue, et I. Reid, «Joint probabilistic data association revisited», dans *ICCV*, 2015, pp. 3047-3055.
- [12] A. Bewley, G. Zongyuan, F. Ramos et B. Upcroft, "Suivi simple en ligne et en temps réel", dans *ICIP*, 2016, pp. 3464–3468.
- [13] L. Leal-Taixé, A. Milan, I. Reid, S. Roth et K. Schindler, «MOTChallenge 2015: Vers une référence pour le suivi multi-cibles», *arXiv: 1504.01942 [cs]*, 2015.
- [14] S. Ren, K. He, R. Girshick et J. Sun, «Faster R-CNN: Vers une détection d'objets en temps réel avec des réseaux de proposition de région », dans *NIPS*, 2015.
- [15] A. Milan, L. Leal-Taixé, I. Reid, S. Roth et K. Schindler, «Mot16: Une référence pour le suivi multi-objets», *préimpression arXiv arXiv: 1603.00831*, 2016.
- [16] F. Yu, W. Li, Q. Li, Y. Liu, X. Shi et J. Yan, «Poi: Suivi d'objets multiples avec fonction de détection et d'apparence hautes performances », dans *ECCV*. Springer, 2016, p. 36–42.
- [17] M. Keuper, S. Tang, Y. Zhongjie, B. Andres, T. Brox, et B. Schiele, «Une formulation multi-coups pour la segmentation conjointe et le suivi de plusieurs objets», *préimpression arXiv arXiv: 1607.06317*, 2016.
- [18] B. Lee, E. Erdene, S. Jin, MY Nam, YG Jung et PK Rhee, «Suivi multi-objets multi-classes utilisant la détection de point de changement», dans *ECCV*. Springer, 2016, pp. 68–83.
- [19] W. Choi, «Suivi multi-cibles quasi en ligne avec descripteur de flux local fermé, "dans *ICCV*, 2015, pp. 3029–3037.
- [20] R. Sanchez-Matilla, F. Poiesi et A. Cavallaro, «En ligne suivi multi-cibles avec détections fortes et faibles », dans *Conférence européenne sur la vision par ordinateur*. Springer, 2016, pp. 84–99.
- [21] L. Zheng, Z. Bie, Y. Sun, J. Wang, C. Su, S. Wang et Q. Tian, «MARS: A video benchmark for large-scale person re-identification», dans *ECCV*, 2016.
- [22] S. Zagoruyko et N. Komodakis, «Wide residual net-travaille dans *BMVC*, 2016, p. 1–12.
- [23] K. Bernardin et R. Stiefelhausen, «Evaluating multiple performances de suivi d'objets multiples: les mesures CLEAR MOT, " *Processus vidéo d'image EURASIP J.*, vol. 2008, 2008.