

Reconnaissance d'actions dans les vidéos de sport amateur

Mots clés : apprentissage profond, données vidéo, reconnaissance d'actions, apprentissage par transfert, adaptation de domaine.

1 A propos

Sport Amat¹ naît du constat de la mise en invisibilité du sport amateur dans les médias et sur les réseaux sociaux. Pour palier à ce déséquilibre, nous proposons la mise en place de solutions de captation vidéo sur des infrastructures sportives. Les vidéos sont ensuite accessibles en direct en et VOD à travers un site et une application. Les objectifs de Sport Amat sont :

- Promouvoir, démocratiser, rendre accessible et inclusif le sport amateur en apportant des outils et des moyens aux acteurs et actrices de cet écosystème (petits clubs, collectifs, clubs ruraux, handisport. .).
- Accompagner via le numérique un retour au physique. Un moyen de créer du lien, de fidéliser et de favoriser les dynamiques en faisant venir les sportifs et le public dans les gymnases.

2 Description du sujet

L'analyse automatique des données vidéo trouve de nombreuses applications dans le sport : identification des temps forts, extraction de statistiques de jeu, assistance à l'arbitrage, assistance à l'entraînement par l'analyse des gestes techniques ou des tactiques de jeu, etc. De nombreux travaux ont abordé ces aspects (voir par ex. [5, 10], allant du suivi de joueurs dans les vidéos à la caractérisation de gestes techniques, en passant par la reconnaissance d'actions de différents niveaux (simples ou complexes, individuelles ou de groupe). Ces travaux se sont appuyés sur la mise à disposition de diverses bases de vidéos annotées, concernant principalement une pratique professionnelle des sports les plus populaires. Par ailleurs, des implémentations de différentes méthodes de l'état de l'art sont librement disponibles, comme [PaddlePaddle](#) (qui s'intéresse entre autres au football et au basket), [MTV](#) [11] ou [MVD](#) [9].

Le développement récent de caméras rapides haute résolution et abordables a provoqué une explosion de la quantité de vidéos sportives disponibles, surtout concernant les sportifs amateurs. Cela reflète le grand intérêt du public pour ces nouveaux contenus. Ces vidéos, très peu annotées, diffèrent de façon souvent significative de leurs équivalents professionnels et cela rend difficile non seulement l'application directe mais aussi l'adaptation des méthodes existantes à ces contenus.

1. <https://sportamat.com/>

Dans ce stage nous cherchons à améliorer les performances des méthodes de reconnaissance d'actions dans des vidéos de sports, mises au point en général sur des contenus annotés professionnels, lorsqu'on les applique à des vidéos de sports amateurs. Les sports privilégiés sont le basket et le football en salle (ou **futsal**). Le basket et le football (*soccer*) sont deux des sports les plus suivis au monde, et il existe déjà un nombre important de travaux dédiés à la détection et au suivi des joueurs ou à la reconnaissance d'actions, s'appuyant sur des bases annotées de matches professionnels comme FineBasketball [4], NPUBasketball [8], ComprehensiveSoccer [12], SoccerNet [3], SSET [2], SoccerDB [7], SoccerNet-v2 [1]. Or, les vidéos de matches amateurs présentent des caractéristiques différentes : fonds variés, mouvements différents des joueurs non professionnels, conditions de prise de vue, etc. Par ailleurs, pour le futsal on observe des différences supplémentaires : le jeu se déroule à l'intérieur (contrairement au football), le terrain est autre, les mouvements sont différents sur sol dur, la caméra est fixe et grand angle, etc.

Le défi est d'identifier les aspects qui posent le plus de difficultés et de proposer des améliorations adaptées sans exploiter un nombre significatif d'annotations pour le sport amateur car celles-ci ne sont pas disponibles. Nous explorerons dans ce but des méthodes de transfert d'apprentissage [13, 6], s'appuyant éventuellement en partie sur des modèles génératifs.

3 Profil recherché

Étudiant(e) en 2ème année de Master, dernière année d'école d'ingénieur ou équivalent :

- Avec une expérience de stage(s) passé(s).
- Bonne maîtrise du domaine de l'apprentissage profond.
- Bonne connaissances de Python et de la librairie PyTorch.
- Une bonne maîtrise de l'anglais technique est indispensable.

4 Organisation

Le stage, financé par la société Sport Amat, se déroulera principalement au sein de l'équipe de recherche Vertigo² du laboratoire CEDRIC³ du CNAM Paris⁴, en co-encadrement avec Sport Amat. Plusieurs enseignants-chercheurs et doctorants de l'équipe Vertigo travaillent sur l'apprentissage profond pour la compréhension et la structuration des données multimédia (image/vidéo/son).

Le stage pourra être poursuivi par une thèse de doctorat en CIFRE, co-encadrée par le CEDRIC et Sport Amat.

2. <http://cedric.cnam.fr/lab/teams/vertigo-en/>

3. <http://cedric.cnam.fr/>

4. <http://www.cnam.fr/>

5 Encadrement

Marin Ferecatu (marin.ferecatu@cnam.fr)

Maître de conférences, HDR, CNAM Paris, 2 rue Conté, 75003 Paris
<http://cedric.cnam.fr/lab/teams/vertigo-en/>

Michel Crucianu (michel.crucianu@cnam.fr)

Professeur d'université, CNAM Paris, 2 rue Conté, 75003 Paris
<http://cedric.cnam.fr/lab/teams/vertigo-en/>

François-Xavier Bazantay (fxbazantay@sportamat.com)

Sport Amat, 25 rue veuve Aublet, 93230 ROMAINVILLE
<https://sportamat.com/>

Bernard Schoukroun (fbernardschoukroun@sportamat.com)

Sport Amat, 25 rue veuve Aublet, 93230 ROMAINVILLE
<https://sportamat.com/>

6 Comment faire acte de candidature

Transmettre aux membres de l'équipe d'encadrement CV, lettre de motivation, relevés de notes, éventuellement lettre(s) de recommandation. Privilégier pour cela l'envoi d'un lien vers une archive en ligne contenant ces documents.

7 Informations complémentaires

Niveau d'études : Bac+5 / Master

Type de contrat : Stage (6 mois)

Date de début : février/mars 2024 (flexible)

Lieu : Paris centre (CEDRIC/CNAM)

Télétravail partiel possible.

Références

- [1] A. Deliege, A. Cioppa, S. Giancola, M. J. Seikavandi, J. V. Dueholm, K. Nasrollahi, B. Ghanem, T. B. Moeslund, and M. Van Droogenbroeck. Soccernet-v2 : A dataset and benchmarks for holistic understanding of broadcast soccer videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4508–4519, 2021.
- [2] N. Feng, Z. Song, J. Yu, Y.-P. P. Chen, Y. Zhao, Y. He, and T. Guan. Sset : a dataset for shot segmentation, event detection, player tracking in soccer videos. *Multimedia Tools and Applications*, 79(39) :28971–28992, 2020.

- [3] S. Giancola, M. Amine, T. Dghaily, and B. Ghanem. Soccernet : A scalable dataset for action spotting in soccer videos. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 1711–1721, 2018.
- [4] X. Gu, X. Xue, and F. Wang. Fine-grained action recognition on a novel basketball dataset. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2563–2567. IEEE, 2020.
- [5] K. Host and M. Ivašić-Kos. An overview of human action recognition in sports based on computer vision. *Heliyon*, 8(6) :e09633, 2022.
- [6] M. Iman, H. R. Arabnia, and K. Rasheed. A review of deep transfer learning and recent advancements. *Technologies*, 11(2), 2023.
- [7] Y. Jiang, K. Cui, L. Chen, C. Wang, and C. Xu. Soccerdb : A large-scale database for comprehensive video understanding. In *Proceedings of the 3rd International Workshop on Multimedia Content Analysis in Sports*, pages 1–8, 2020.
- [8] C. Ma, J. Fan, J. Yao, and T. Zhang. Npu rgb-d dataset and a feature-enhanced lstm-dgcn method for action recognition of basketball players. *Applied Sciences*, 11(10) :4426, 2021.
- [9] R. Wang, D. Chen, Z. Wu, Y. Chen, X. Dai, M. Liu, L. Yuan, and Y.-G. Jiang. Masked video distillation : Rethinking masked feature modeling for self-supervised video representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6312–6322, June 2023.
- [10] F. Wu, Q. Wang, J. Bian, H. Xiong, N. Ding, F. Lu, J. Cheng, and D. Dou. A survey on video action recognition in sports : Datasets, methods and applications, 2022.
- [11] S. Yan, X. Xiong, A. Arnab, Z. Lu, M. Zhang, C. Sun, and C. Schmid. Multiview transformers for video recognition. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3323–3333, 2022.
- [12] J. Yu, A. Lei, Z. Song, T. Wang, H. Cai, and N. Feng. Comprehensive dataset of broadcast soccer videos. In *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, pages 418–423. IEEE, 2018.
- [13] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He. A comprehensive survey on transfer learning, 2020.