

Du repérage sémantique robuste d'actions vers leur détection dans les vidéos

Soutenance de thèse Guillaume Vaudaux-Ruth

8/12/2021 - 10h

Palaiseau - salle Marcel Pierre

Devant le jury composé de :

Catherine Achard

(Directrice de thèse - Université Paris-Sorbonne)

Jean-Philippe Vandeborre

(Rapporteur - Université des Sciences et Technologies de Lille)

Quoc Cuong Pham

(Rapporteur – CEA)

Stéphane Doncieux

(Examineur – Université Paris-Sorbonne)

Alexandre Boulch

(Examineur – Valeo)

Adrien Chan-Hon-Tong

(Invité - ONERA)

Résumé :

La compréhension de vidéos nécessite une caractérisation à la fois spatiale et temporelle de leur contenu. Ainsi, face au succès des méthodes d'apprentissage statistique par réseaux de neurones pour l'analyse automatique d'images, ces méthodes ont rapidement été étendues au contexte spatio-temporel. La dimension temporelle introduit de nouvelles problématiques algorithmiques, que ce soit pour la caractérisation des vidéos ou pour l'extraction d'informations sémantiques. Dans cette thèse, nous nous intéressons à la détection temporelle d'actions qui vise, non seulement à reconnaître les actions présentes dans une vidéo, mais aussi à en connaître les limites temporelles. Plus précisément, nous étudions l'impact que peut avoir la recherche d'une localisation temporelle fine sur la robustesse de l'extraction sémantique.

Mots clés :

deep learning, video processing, action detection, action spotting, deep reinforcement learning