

Modèles d'Apprentissage Machine pour la Compréhension de Scènes

Soutenance d'Habilitation à Diriger des Recherches – Bertrand LE SAUX

18 décembre 2019, 14h

Salle Marcel Pierre

ONERA, Palaiseau

Devant le jury composé de :

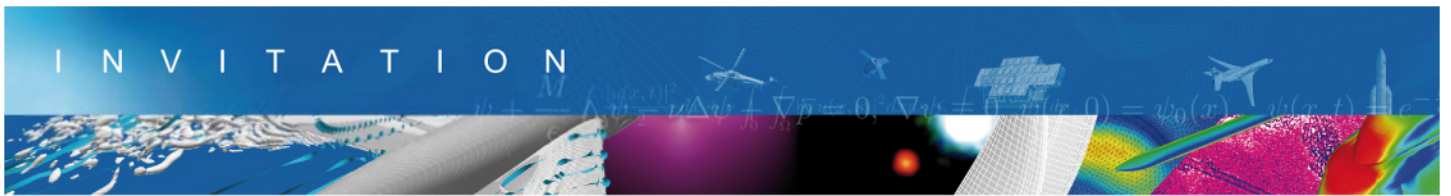
- M. Jocelyn CHANUSSOT (rapporteur),
- M. François FLEURET (examineur),
- Mme Sylvie LE HÉGARAT MASCLE (examinatrice),
- M. Renaud MARLET (examineur),
- Mme Florence TUPIN (rapporteuse),
- Mme Josiane ZERUBIA (rapporteuse).

Résumé :

La compréhension de scènes vise à répondre à la question : comment construire un modèle d'une région du monde réel afin d'y agir et d'y interagir ? Il s'agit donc d'extraire la sémantique et la géométrie des données disponibles : images, nuages de points 3D, etc. Dans ce but, plusieurs approches d'apprentissage machine sont présentées : elles diffèrent par la proportion d'a priori de conception et d'apprentissage introduits tout au long des algorithmes. Trois aspects du problème sont envisagés. Les premiers travaux visent à la compréhension du contenu sémantique des images, c'est à dire la classification, la détection d'objets et la segmentation sémantique. Puis, plusieurs approches d'apprentissage sont proposées pour l'observation de la Terre et la télédétection, notamment pour l'apprentissage interactif, la classification sémantique multimodale et la détection de changements sémantiques. Enfin, l'accent est mis sur la vision 3D, avec l'estimation de la profondeur à partir d'une seule image et la classification de nuages de points 3D par des réseaux de neurones. Ces approches variées reposent sur des mécanismes sous-jacents communs qui prennent une importance croissante. Elles réalisent une analyse multimodale pour bénéficier des données complémentaires disponibles, issues de capteurs différents mais aussi de sources et méta-données hétérogènes. Symétriquement, l'optimisation jointe d'objectifs multiples permet de régulariser l'apprentissage de modèles performants. Surtout, elles ont de plus en plus recours à une multiplicité des points de vue sur la scène pour relier, tant en apprentissage qu'en inférence, des invariances spatiales qui servent une analyse locale et une reconstruction sémantique globale. Cela est rendu possible par une intégration croissante de l'apparence et de la structure 3D, et conduit à une meilleure compréhension sémantique de la scène.

Abstract:

Scene understanding aims to answer the question : how to build a model of a real-world region in order to act and interact with it? It is therefore necessary to extract the semantics and geometry of the available data : images, 3D point-clouds, etc. For this purpose, several machine learning approaches are presented: they differ in the proportion of prior assumptions and learning introduced throughout the algorithms. Three aspects of the problem are envisaged. The first works aim at understanding the semantic content of images, through classification, object detection and semantic segmentation. Then, several learning approaches are proposed for Earth observation and remote sensing, notably for interactive learning, multi-modal semantic classification and semantic change detection. Finally, the focus is on 3D vision, with depth estimation from



a single image and classification of 3D point-clouds by neural networks. These various approaches are based on common underlying mechanisms that are becoming increasingly important. They perform a multimodal analysis in order to benefit from the available, complementary data, obtained from different sensors but also from heterogeneous sources and meta-data. Symmetrically, joint optimization of multiple objectives helps to regularize the learning of efficient models. Moreover, they increasingly rely on a multiplicity of points of view on the scene to relate, in both learning and inference, spatial invariances that serve a local analysis and a global semantic reconstruction. This is made possible by a growing integration of the appearance and 3D structure, and leads to a better semantic understanding of the scene.

Mots clés :

Apprentissage machine; Compréhension de scènes;
Vision par ordinateur; Observation de la Terre; Vision 3D;
Réseaux de neurones; Classification d'images; Détection d'objets.

Accès :

Salle Marcel Pierre
ONERA
6 chemin de la Vauve aux Granges, Palaiseau

<https://www.onera.fr/centres/palaiseau>