

Chers collègues,

Je vous écris pour vous inviter tous à la soutenance de ma thèse doctorale, qui est prévue pour le 6 octobre dans l'après-midi. Le sujet a été préparé au sein du laboratoire IBISC, Université Paris-Saclay (Univ Evry), intitulé "***Deep Learning pour la fusion multimodale d'images: application à l'analyse de scènes extérieures dans des conditions difficiles.***"

INFO:

Date: **Vendredi, Octobre 6th 2023 at 14:00 pm**

Lieu: UFR Sciences et Technologies – **site Pelvoux, 40 rue du Pelvoux** – CE1455, 91020 Evry Cedex Courcouronnes

Salle: Hall - B, amphi **BX30**

Membres du jury:

- Mme. Anissa MOKRAOUI, Professor, Université Sorbonne Paris Nord (L2TI), Présidente
- M. Antoine MANZANERA, Professor, ENSTA Paris (U2IS), Examineur
- Mme. Sylvie CHAMBON, Maître de Conférences - HDR, Toulouse INP (IRIT), Rapporteur
- M. Yassine RUICHEK, Professor, Université de Technologie Belfort-Montbéliard (CIAD), Rapporteur

- M. Dro Désiré SIDIBÉ, Professor, University of Paris-Saclay (Univ Evry), Directeur de thèse
- Mme. Samia Bouchafa-Bruneau, Professor, University of Paris-Saclay (Univ Evry), Co-directrice
- M. Fabien Bonardi, Associate Professor, University of Paris-Saclay (Univ Evry), Co-encadrant

Résumé:

Les données visuelles multimodales peuvent fournir des informations différentes sur la même scène, améliorant ainsi la précision et la robustesse de l'analyse de scènes. Cette thèse se concentre principalement sur la façon d'utiliser efficacement les données visuelles multimodales telles que les images en couleur, les images infrarouges et les images de profondeur, et sur la façon de fusionner ces données visuelles pour une compréhension plus complète de l'environnement. Nous avons choisi la segmentation sémantique et la détection d'objets, deux tâches représentatives de la vision par ordinateur, pour évaluer et valider différentes méthodes de fusion de données visuelles multimodales. Ensuite, nous proposons un schéma de fusion RGB-D basé sur l'attention additive, considérant la carte de profondeur comme une modalité auxiliaire pour fournir des indices géométriques supplémentaires, et résolvant le coût élevé associé à l'auto-attention. Compte tenu de la complexité de la perception de scènes en conditions de faible luminosité, nous avons conçu un module de fusion croisée qui utilise l'attention de canal et spatiale pour explorer les informations complémentaires des paires d'images visible-infrarouge, améliorant ainsi la perception de l'environnement par le système. En fin, nous avons également abordé l'application des données visuelles multimodales dans l'adaptation de domaine non supervisée. Nous proposons d'utiliser des indices de profondeur pour guider le modèle à apprendre la représentation de caractéristiques

invariables au domaine. Les nombreux résultats expérimentaux indiquent que les méthodes proposées surpassent les autres méthodes sur plusieurs bases de données multimodales disponibles publiquement et peuvent être étendues à différents types de modèles, démontrant ainsi davantage la robustesse et les capacités de généralisation de nos méthodes dans les tâches de perception de scènes en extérieur.

Bien cordialement,

Dear colleagues,

I am writing to extend an invitation to you all for the defense of my doctoral thesis, which is scheduled for October 6th in the afternoon. The topic was prepared within the IBISC laboratory, University of Paris-Saclay (Univ Évry), titled “***Multi-modal Visual Data Fusion For Outdoor Scene Understanding in Challenging Weather Conditions.***”

INFO:

Data: **Friday, October 6th 2023 at 14:00 pm**

Loaction: Location: UFR Sciences and Technologies – **Pelvoux site, 40 rue du Pelvoux** – CE1455, 91020 Evry Cedex Courcouronnes

Room: Hall - B, amphitheater **BX30**

Jury members :

- Mrs. Anissa MOKRAOUI, Professor, Université Sorbonne Paris Nord (L2TI), President
- Mr. Antoine MANZANERA, Professor, ENSTA Paris (U2IS), Examiner
- Mrs. Sylvie CHAMBON, Maître de Conférences - HDR, Toulouse INP (IRIT), Reviewer
- Mr. Yassine RUICHEK, Professor, Université de Technologie Belfort-Montbéliard (CIAD), Reviewer

- Mr. Dro Désiré SIDIBÉ, Professor, University of Paris-Saclay (Univ Evry), Thesis Director
- Mrs. Samia Bouchafa-Bruneau, Professor, University of Paris-Saclay (Univ Evry), Co-director
- Mr. Fabien Bonardi, Associate Professor, University of Paris-Saclay (Univ Evry), Co-supervisor

Abstract : Multi-modal visual data can provide different information about the same scene, thus enhancing the accuracy and robustness of scene analysis. This thesis mainly focuses on how to effectively utilize multi-modal visual data such as color images, infrared images, and depth images, and how to fuse these visual data for a more comprehensive understanding of the environment. Semantic segmentation and object detection, two representative computer vision tasks, were selected for investigating and verifying different multi-modal visual data fusion methods. Then, we propose an additive-attention-based RGB-D fusion scheme, considering the depth map as an auxiliary modality to provide additional geometric clues, and solving the high cost associated with self-attention. Considering the complexity of scene perception under low-light conditions, we designed a cross-fusion module that uses

channel and spatial attention to explore the complementary information of visible-infrared image pairs, enhancing the system's perception of the environment. Additionally, we also researched the application of multi-modal visual data in unsupervised domain adaptation. We proposed to leverage depth cues to guide the model to learn domain-invariant feature representation. Extensive research results indicate that the proposed methods outperform others on multiple publicly available multi-modal datasets and can be extended to different types of models, which further demonstrating the robustness and generalization capabilities of our methods in outdoor scene perception tasks.

With best regards,

Sijie Hu