

Avis de Soutenance

Madame Nahid NAZIFI

Sciences et Technologies Industrielles

Soutiendra publiquement ses travaux de thèse intitulés

Estimation Visuelle de l'Odométrie et de la Profondeur Basée sur les Transformateurs pour l'Endoscopie Capsulaire Sans Fil

dirigés par Monsieur Driss BOUTAT, Monsieur Helder ARAUJO et Monsieur Omar TAHRI

Ecole doctorale : Mathématiques, Informatique, Physique Théorique et Ingénierie des Systèmes - MIPTIS

Unité de recherche : PRISME - Laboratoire Pluridisciplinaire de Recherche en Ingénierie des Systèmes et Mécanique
Energétique

Cotutelle avec l'université "Université de Coimbra" (PORTUGAL)

Soutenance prévue le **lundi 27 janvier 2025** à 14h00

Lieu : INSA Centre Val De Loire, 88 boulevard Lahitolle, 18000 BOURGES

Salle : des Conseils

Composition du jury proposé

M. Driss BOUTAT	INSA Centre Val de Loire	Directeur de thèse
M. Helder ARAUJO	Electrical and Computer Engineering Department	Co-directeur de thèse
M. Omar TAHRI	Université de Bourgogne	Co-directeur de thèse
M. Fabrice MERIAUDEAU	Université de Bourgogne	Examineur
Mme Alice OTHMANI	Université Paris-Est Créteil	Examinatrice
M. Desiré SIDIBÉ	Université Evry Val Essonne	Rapporteur
M. Ibrahima FAYE	Université de Technologie de Petronas (Universiti Teknologi PETRONAS)	Rapporteur

Mots-clés : Deep-learning, Endoscopie par capsule, Odométrie visuelle, Transformers, Auto-supervisé, estimation de la pose et de la profondeur

Résumé :

L'estimation précise de la pose et de la profondeur pour l'endoscopie par capsule (Wireless Capsule Endoscopy, WCE) demeure un défi majeur en raison de la nature non structurée et pauvre en textures du tractus gastro-intestinal (GI). Cette thèse explore l'utilisation d'architectures basées sur les transformateurs pour l'estimation auto-supervisée de la profondeur et de la pose monoculaires en WCE. Contrairement aux méthodes traditionnelles d'odométrie visuelle, qui reposent sur des techniques basées sur des points d'intérêt, les approches proposées exploitent le Pyramid Vision Transformer (PVT) et le Swin Transformer pour extraire des caractéristiques hiérarchiques multi-échelles, renforçant ainsi la robustesse de la localisation visuelle dans des environnements complexes du GI. Le premier cadre, EndoTranSfm, utilise l'architecture PVTv2 comme base pour l'estimation de la pose et de la profondeur. Il intègre des pertes photométriques adaptées à la luminosité, une perte de cohérence géométrique et une régularisation de la douceur pour affiner les prédictions sans nécessiter de données annotées. EndoTranSfm démontre des performances compétitives par rapport à EndoSfMLearner, notamment dans des scénarios de faible éclairage et de textures répétitives. En s'appuyant sur ces bases, la deuxième approche explore l'intégration du Swin Transformer. Cette architecture introduit un mécanisme d'attention basé sur des fenêtres glissantes, améliorant l'efficacité computationnelle et la scalabilité. Les versions améliorées du modèle basé sur Swin intègrent des mécanismes d'attention spatiale pour se concentrer sur les régions anatomiques significatives, améliorant ainsi l'estimation de la profondeur et de la pose. Les évaluations comparatives sur les ensembles de données EndoSLAM, SimCol et ColonDepth mettent en évidence les atouts de ces méthodes, notamment leur capacité à généraliser sur des données synthétiques et réelles. Les résultats montrent que les méthodes basées sur les transformateurs sont comparables aux performances de pointe dans les tâches d'odométrie visuelle pour la WCE, avec des améliorations notables en précision de rotation et en prédiction de profondeur. Cette recherche souligne le potentiel de l'apprentissage auto-supervisé et des architectures basées sur les transformateurs pour faire progresser les technologies d'imagerie médicale.