



AVIS DE SOUTENANCE DE THESE

Le Doyen de la Faculté des Sciences Dhar El Mahraz –Fès – annonce que

Mme **CHERRATE Meryem**
Soutiendra : le **Mardi 21/04/2026 à 09H30**
Lieu : **FSDM – Centre Visioconférence**

Une thèse intitulée :

**Reconnaissance Automatique des Gestes de la langue des Signes :
Approches de Machine Learning et de Deep Learning Basées sur les
Images et les Séquences Vidéos**

En vue d'obtenir le **Doctorat**

FD : Sciences et Technologies de l'Information et de la Communication
Spécialité : Informatique

Devant le jury composé comme suit :

Nom et prénom	Etablissement	Grade	Qualité
SATORI KHALID	Faculté des Sciences Dhar EL Mahraz, Fès	PES	Président
BENKIRANE SAID	Ecole Supérieure de Technologie, Essaouira	PES	Rapporteur
GUEZZAZ AZIDINE	Ecole Supérieure de Technologie, Essaouira	MCH	Rapporteur
EL GAROUANI SAID	Faculté des Sciences Dhar EL Mahraz, Fès	PES	Rapporteur
DAROUICHI AZIZ	Faculté des Sciences et Techniques, Marrakech	MCH	Examineur
YAHYAOUY ALI	Faculté des Sciences Dhar EL Mahraz, Fès	PES	Examineur
AARAB ABDELLAH	Faculté des Sciences Dhar EL Mahraz, Fès	PES	Invité
SABRI MY ABDELOUAHED	Faculté des Sciences Dhar EL Mahraz, Fès	PES	Directeur de thèse



Résumé :

La reconnaissance automatique de la langue des signes est une tâche complexe impliquant l'analyse conjointe des caractéristiques spatiales et temporelles des gestes. Cette thèse propose une approche progressive visant d'abord la reconnaissance de signes à partir d'images statiques, avant de s'étendre à la reconnaissance dynamique sur séquences vidéo, en ciblant les langues des signes marocaines et américaines.

Dans un premier temps, des modèles de classification basés sur des réseaux de neurones convolutifs (CNN) ont été développés pour extraire des caractéristiques discriminantes des images de signes marocains et américains. Ces modèles ont permis d'établir une base solide pour la reconnaissance visuelle statique.

Ensuite, afin de capturer la dimension temporelle inhérente aux séquences vidéo de signes, une architecture hybride combinant CNN pour l'extraction spatiale et des réseaux de neurones récurrents à mémoire longue (LSTM) pour la modélisation temporelle a été conçue. Cette approche spatio-temporelle vise à apprendre efficacement les dépendances séquentielles entre les frames tout en exploitant la richesse des caractéristiques visuelles.

La robustesse du système a été évaluée dans des conditions réelles, notamment dans un environnement de station-service, caractérisé par des variations d'éclairage, d'arrière-plan et de positionnement. Les expérimentations montrent une amélioration significative des performances par rapport aux méthodes statiques, avec une précision élevée dans la reconnaissance des gestes.

Les principales contributions de cette thèse se traduisent par le développement d'un modèle hybride CNN-LSTM optimisé pour la reconnaissance de la langue des signes américaine (ASL), ainsi qu'une validation expérimentale approfondie réalisée sur des données collectées dans des conditions variées.

Mots clés :

Reconnaissance automatique ; Langue des signes ; Séquences vidéo ; Signes marocains ; Signes américains ; Réseaux de neurones convolutifs (CNN) ; Réseaux de mémoire à long terme (LSTM) ; Modèle hybride spatio-temporel ; Reconnaissance dynamique ; Apprentissage profond ; temps réel ; Station-service.



AUTOMATIC RECOGNITION OF SIGN LANGUAGE GESTURES: MACHINE LEARNING AND DEEP LEARNING APPROACHES BASED ON IMAGES AND VIDEO SEQUENCES

Abstract:

Automatic sign language recognition is a complex task that requires the joint analysis of spatial and temporal features of gestures. This thesis proposes a progressive approach, initially focusing on the recognition of isolated signs from static images and then extending to dynamic recognition on video sequences, targeting both Moroccan and American Sign Languages.

First, convolutional neural network (CNN)-based classifiers were developed to extract discriminative features from images of Moroccan and American signs. These models established a solid foundation for static visual recognition.

Subsequently, to capture the temporal dimension inherent in sign language video sequences, a hybrid architecture combining CNNs for spatial feature extraction and long short-term memory (LSTM) recurrent neural networks for temporal modeling was designed. This spatio-temporal approach aims to effectively learn sequential dependencies between frames while exploiting rich visual information.

The system's robustness was evaluated in real-world conditions, notably in a gas station environment characterized by variable lighting, backgrounds, and positioning. Experimental results demonstrate significant performance improvements compared to static methods, achieving high accuracy in gesture recognition.

The main contributions of this thesis lie in the development of a hybrid CNN-LSTM model optimized for recognizing American Sign Language (ASL), along with an in-depth experimental validation conducted on datasets collected under diverse conditions.

Key Words:

Automatic recognition; Sign language; Video sequences; Moroccan sign language; American sign language; Convolutional neural networks (CNN); Long short-term memory (LSTM) networks; Spatio-temporal hybrid model; Dynamic recognition; Deep learning; Real-world environment; Gas station.