



Centre d'Etudes Doctorales : Sciences et Techniques de l'Ingénieur

**AVIS DE SOUTENANCE
THESE DE DOCTORAT**

Présentée par

Mr : SOUFIANE HOURRI

Discipline : Informatique

Spécialité : Informatique

Sujet de la thèse : Etude des Techniques de Deep Learning en Reconnaissance Automatique du Locuteur.

Formation Doctorale : Sciences de l'ingénieur Sciences Physiques, Mathématiques et Informatique.

**Thèse présentée et soutenue le lundi 22 février 2021 à 10h au Centre de conférences devant le jury
composé de :**

Nom Prénom	Titre	Etablissement	
Aicha MAJDA	PES	Faculté des Sciences et Techniques de Fès	Présidente
Ahmed BEN HAMIDA	PES	Ecole Nationale d'Ingénieurs de Sfax	Rapporteur
Mohamed BAHAJ	PES	Faculté des Sciences et Techniques de Settet	Rapporteur
Abderrahim BEN ABBOU	PES	Faculté des Sciences et Techniques de Fès	Rapporteur
Khalid ABBAD	PH	Faculté des Sciences et Techniques de Fès	Examineur
Jamal KHARROUBI	PES	Faculté des Sciences et Techniques de Fès	Directeur de thèse

Laboratoire d'accueil : Laboratoire Systèmes Intelligents et Applications.

Etablissement : Faculté des Sciences et Techniques de Fès



Centre d'Etudes Doctorales : Sciences et Techniques de l'Ingénieur

Titre de la thèse : Etude des Techniques de Deep Learning en Reconnaissance Automatique du Locuteur.

Nom du candidat : Soufiane HOURRI

Spécialité : Informatique

Résumé de la thèse

La reconnaissance automatique du locuteur (RAL), parfois appelée biométrie du locuteur, implique l'identification, la vérification (authentification), la classification et, par extension, la segmentation, le suivi et la détection des locuteurs. Il s'agit d'un terme générique utilisé pour toute procédure impliquant la reconnaissance de l'identité d'une personne sur la base de sa voix. Dans ce contexte, le deep learning a suscité beaucoup plus d'intérêt de la part des chercheurs en traitement de la parole, et il a été introduit récemment dans la RAL. Dans la plupart des cas, les modèles de deep learning sont adaptés des applications de reconnaissance automatique de la parole (RAP) et appliqués à la RAL, et ils ont montré leur capacité à concurrencer les approches de l'état de l'art. Néanmoins, l'utilisation de deep learning dans la RAL est toujours liée à la RAP. D'autre part, les modèles de deep learning sont maintenant considérés comme état de l'art dans nombreux domaines de la reconnaissance des formes. En RAL, plusieurs architectures ont été étudiées, telles que les réseaux de neurones profonds (DNN), les réseaux de croyances profonds (DBN), les machines de Boltzmann restreintes (RBM), etc. tandis que les réseaux de neurones convolutifs (CNN) sont les modèles les plus utilisés en traitement d'images.

L'objectif de cette thèse est d'étudier les modèles de deep learning pour le domaine de la RAL. Pour cette raison, nous avons proposé une nouvelle façon d'utiliser les DBN et les DNN dans la RAL, dans le but d'extraire les caractéristiques profondes du locuteur (DeepSF). Par la suite, nous avons proposé une nouvelle utilisation des CNN pour le problème de la RAL. Bien qu'ils soient particulièrement conçus pour les problèmes de traitement d'images, les CNN ont récemment été appliqués à la RAL en utilisant des spectrogrammes comme images d'entrée. Nous pensons que cette approche n'est pas optimale car elle peut entraîner deux erreurs cumulatives dans la résolution d'un problème de traitement d'images et de RAL. C'est pourquoi nous avons développé une nouvelle méthode qui permet d'utiliser les CNN sans utiliser d'images.

Les résultats de la thèse représentent une découverte importante pour comprendre comment les modèles de deep learning peuvent être adaptés au problème de la RAL.

Mots clés : *Reconnaissance Automatique du Locuteur, Vérification Automatique du Locuteur, Apprentissage Profond, Réseau de Neurones Profond, Réseau de Croyance profond, Réseau de Neurones Convolutif, Machine de Boltzmann Restreinte.*