ABSTRACT

The habilitation thesis authored by Associate Professor Dr. Bogdan – Cosmin MOCANU presents the author's significant accomplishments from both academic and scientific perspectives, together with his aspirations for career advancement. Notably, all contributions presented at the professional or research level have been attained after finishing his doctoral studies. Furthermore, the thesis provides insights into the trajectory of PhD. MOCANU's career, highlighting his commitment to ongoing growth and excellence in academia and research.

As a researcher, he possesses an **extensive record of publications** in prestigious journals and conferences across the field. His scientific activity can be summarized in **13 publications in Q1/Q2 journals** (*e.g.,* Pattern Recognition Letter, Multimedia Tools and Application, Image and Visual Computing, IEEE Access, Sensors, Journal of Ambient Intelligence and Smart Environments etc.) and participation in major **class A+** computer vision conferences such as: International Conference on Computer Vision (ICCV) and European Conference on Computer Vision (ECCV). As a senior researcher, his activity has resulted in a broad portfolio consisting of **66 scientific publications**, to which he has contributed as either the lead or co-author. Among these publications, **18 papers** have been featured in esteemed journals indexed by ISI. Furthermore, his engagement with academic discourse is underscored by the presentation of **42 papers** at esteemed conferences indexed by ISI. This active engagement in conference presentations not only demonstrates his dedication to advancing knowledge but also underscores his proficiency in disseminating findings and collaboration with peers on a global level. He has been involved in **16 research projects at national and international levels**, as follows: as a project manager, he coordinated three national projects; as a principal investigator, he led the technical team of the partner in two international research projects, and he have contributed as a scientific researcher to **11 international and/or national projects**.

The *interdisciplinary* of the thesis is given by the fact that it connects and integrates different subjects and techniques. The subjects being treated are situated at the intersection of two recent tendencies in science: artificial intelligence and online video streaming platforms. His findings showcased hold relevance for both academic and industrial sectors, with the potential for various industrial applications on the horizon.

The rest of the manuscript is organized into eight chapters as follows. *Chapter 1* realizes an introduction of all subjects covered by the thesis and details the problematic and content of each section.

Chapter 2 offers a concise overview of his educational journey and professional trajectory that he followed until now. With a meticulous focus on research projects and didactic activities, it offers an in-depth exploration of the multifaceted roles he has assumed, encompassing rigorous tasks such as peer reviewing, project evaluation, student coordination.

The next section (*Chapter 3*) introduces a subtitle synchronization and positioning system designed to increase the accessibility of deaf and hearing-impaired people to multimedia documents. The main contributions concern: a novel synchronization algorithm able to robustly align, without any human intervention, the closed caption with the audio transcript and a timestamp refinement technique that adjusts the subtitle segments duration with respect to the audiovisual recommendations.

Chapter 4 introduces the ***DEEP-HEAR*** framework, a multimodal dynamic subtitle positioning system designed to increase the accessibility of deaf and hearing-impaired people to multimedia documents. The proposed system exploits both computer vision algorithms and deep convolutional neural networks specifically designed and tuned to detect and recognize the identity of the active speaker.

Chapter 5 introduces the ***DEEP-AD*** framework, a multimodal advertisement insertion system dedicated to online video platforms. The framework is designed from the viewer's perspective, in terms of commercial contextual relevance and degree of intrusiveness. The proposed algorithm exploits various deep convolutional neural networks, involved at several stages. The video stream is first divided into shots based on a graph partition method. The video shots are then clustered into scenes/story units with the help of an agglomerative clustering methodology taking as input visual, audio, and semantic features. Furthermore, to facilitate the user's access to multimedia documents a novel thumbnail extraction method is proposed based on both semantic representativeness and visual quality information. Finally, the optimal advertisement insertion points are determined based on the ads' temporal distribution, commercial diversity, and degree of intrusiveness.

Chapter 6 tackles the issue of multi-modal emotion recognition. A novel end-to-end multimodal emotion recognition methodology is introduced, based on audio and visual fusion designed to leverage the mutually complementary nature of features while maintaining the modality-specific information. The proposed method integrates spatial, channel and temporal attention mechanisms into a visual 3D convolutional neural network and temporal attention into an audio 2D convolutional neural network to capture the intra-modal features characteristics. Further, the inter-modal information is captured with the help of an audio-video cross-attention fusion technique that effectively identifies salient relationships across the two modalities.

Chapter 7 serves to briefly present his academic and research journey, with the focus put on professional experience, the obtained results, and the prominence of the proposed

contributions. Moreover, it accentuates the visibility of his activity within the specialized area of artificial intelligence.

Finally, *Chapter 8* concludes the manuscript and highlights the main contributions proposed in this work. Additionally, it provides insights into future research directions, focusing on advancements in methodology dedicated to online video streaming platforms.

The current thesis, entitled "*Multimodal Deep Learning Technologies Dedicated to Online Video Streaming Platforms*," presents the author's professional accomplishments following the attainment of his PhD title. It attests to the originality and significance of the author's academic, scientific, and professional endeavors. This document serves as validation of the author's autonomous progression in both future research and his university career within the domain of *Electronics, Telecommunications, and Information Technologies*.

REZUMAT

Teza de abilitare elaborată de Conf. Dr. Ing. Bogdan-Cosmin MOCANU prezintă realizările semnificative ale autorului din punct de vedere academic și științific, alături de aspirațiile sale pentru carieră viitoarea. Este demn de menționat că toate contribuțiile prezentate, atât la nivel profesional, cât și în domeniul cercetării, au fost obținute după finalizarea studiilor doctorale. Mai mult, teza oferă o perspectivă asupra parcursului profesional al domnului MOCANU, evidențiind angajamentul acestuia față de dezvoltarea continuă și excelența în mediul academic și de cercetare.

Teza include o selecție din rezultatele cercetării autorului. Acesta deține un **portofoliu extins de publicații** în reviste și conferințe prestigioase din domeniu. Activitatea sa științifică poate fi rezumată în **13 publicații în reviste indexate Q1/Q2** (de exemplu, Pattern Recognition Letter, Multimedia Tools and Application, Image and Visual Computing, IEEE Access, Sensors, Journal of Ambient Intelligence and Smart Environments etc.) și participări la conferințe de prestigiu indexate **clasă A+** precum: International Conference on Computer Vision (ICCV) and European Conference on Computer Vision (ECCV). Contribuția sa academică poate fi rezumată în **66 de publicații științifice**, la care a contribuit ca autor principal sau co-autor. Dintre acestea, **18 articole** au fost publicate în reviste de prestigiu indexate de ISI, iar **42 de lucrări** au fost publicate la conferințe de prestigiu indexate de ISI. El a fost implicat în **16 proiecte de cercetare** la nivel național și internațional, astfel: în calitate de manager de proiect, a coordonat trei proiecte naționale; în calitate de investigator principal, a condus echipa tehnică a partenerului în două proiecte internaționale și a contribuit ca cercetător științific la 11 proiecte internaționale și/sau naționale.

Interdisciplinaritatea tezei este dată de faptul că aceasta conectează și integrează diferite domenii și tehnici. Subiectele tratate se află la intersecția a două tendințe recente în știință: inteligența artificială și platformele online de streaming video. Rezultatele prezentate sunt relevante atât pentru sectorul academic, cât și pentru cel industrial, anticipându-se câteva potențiale aplicații industriale.

Manuscrisul este organizat în opt capitole după cum urmează. *Capitolul 1* realizează o introducere a temelor abordate, cu detalierea problematicii și conținutului fiecărei secțiuni.

Capitolul 2 oferă un rezumat al parcursului său educațional și a traiectoriei profesionale pe care a urmat-o. Accentul este pus pe proiectele de cercetare la care acesta a participat și pe activitatea didactică.

În următoarea secțiune (*Capitolul 3*), se introduce un sistem de sincronizare și poziționare a subtitrărilor conceput pentru a îmbunătăți accesibilitatea persoanelor surde și cu deficiențe de auz la documente multimedia. Principalele contribuții presupun: un nou algoritm de sincronizare a subtitrărilor capabil să alinieze, în mod robust, fără intervenția umană, subtitrările cu transcrierea audio și o nouă tehnică de rafinare a marcajelor temporale care ajustează durata segmentelor de subtitrare în conformitate cu recomandările audiovizuale.

Capitolul 4 introduce arhitectura ***DEEP-HEAR***, un sistem multimodal de poziționare dinamică a subtitrărilor, conceput pentru a îmbunătăți accesibilitatea persoanelor cu deficiențe de auz la documente multimedia. Sistemul propus exploatează atât algoritmi de viziune artificială, cât și rețele neurale convoluționale profunde special concepute pentru a detecta și recunoaște identitatea vorbitorului activ.

Capitolul 5 introduce sistemul ***DEEP-AD***, un sistem multimodal de inserție adaptivă a reclamelor dedicat platformelor online de redare a fluxurilor video. Arhitectura este concepută din perspectiva utilizatorului, în ceea ce privește relevanța contextuală comercială și gradul de intruziune. Algoritmul propus exploatează un set divers de rețele neurale convoluționale profunde. Fluxul video este mai întâi împărțit în planuri video pe baza unei metode de partiționare de tip graf. Ulterior, planurile video sunt grupate în scene/unități semantice cu ajutorul unei metodologii de clustering aglomerativ, care utilizează caracteristici vizuale și auditive. În plus, pentru a facilita accesul utilizatorului la documentele multimedia, este propusă o metodă nouă de extragere a imaginilor cheie, bazată atât pe reprezentativitatea semantică, cât și pe informațiile privind calitatea vizuală. În cele din urmă, punctele optime de inserare a reclamelor sunt determinate pe baza distribuției temporale, diversității comerciale și gradului de intruziune.

În *Capitolul 6* se abordează problema recunoașterii multimodale a emoțiilor. În acest capitol, se introduce o nouă metodologie de recunoaștere a emoțiilor end-to-end, bazată pe fuziunea audio și vizuală, concepută pentru a valorifica natura complementară a caracteristicilor, menținând în același timp informațiile specifice modalității. Metoda propusă integrează mecanisme de atenție spațială, de canal și temporală într-o rețea neuronală convoluțională 3D ce acceptă la intrare fluxuri video și atenție temporală într-o rețea neuronală convoluțională 2D pentru semnalul audio proiectate pentru a captura caracteristicile intra-modale. În plus, informația inter-modală este extrasă cu ajutorul unei tehnici de fuziune a atenției audio-video care identifică eficient relațiile semnificative între cele două modalități.

Capitolul 7 face o trecere în revistă a realizărilor autorului în plan academic, științific și profesional, făcând-se referire la activitatea de predare, la granturile în care a participat ca membru sau coordonator, la activitatea de îndrumare a studenților și la activitatea de publicare.

În final, *Capitolul 8* încheie manuscrisul și evidențiază principalele contribuții propuse în această lucrare. În plus, se oferă perspective asupra direcțiilor viitoare de cercetare, concentrându-se pe sistemele dedicate platformelor de streaming video.

Lucrarea cu titlul "*Multimodal Deep Learning Technologies Dedicated to Online Video Streaming Platform (Tehnologii multimodale de învățare profundă dedicate platformelor online de redare a fluxurilor video)*" prezintă în mod documentat realizările profesionale obținute de autor ulterior conferirii titlului de doctor în știință, certificând originalitatea și relevanța contribuțiilor sale academice, științifice și profesionale. Teza de abilitare anticipează o dezvoltare a viitoarei cariere academice a autorului în domeniul *Inginerie Electronică, Telecomunicații și Tehnologii Informaționale*.