

Article

Speaker Recognition Using Constrained Convolutional Neural Networks in Emotional Speech

Nikola Simić ^{1,*} , Siniša Suzić ¹, Tijana Nosek ¹, Mia Vujović ¹, Zoran Perić ², Milan Savić ³ and Vlado Delić ¹ 

¹ Faculty of Technical Sciences, University of Novi Sad, Trg Dositeja Obradovica 6, 21000 Novi Sad, Serbia; sinisa.suzic@uns.ac.rs (S.S.); tijana.nosek@uns.ac.rs (T.N.); miavujovic@uns.ac.rs (M.V.); vlado.delic@uns.ac.rs (V.D.)

² Faculty of Electronic Engineering, University of Nis, Aleksandra Medvedeva 14, 18000 Nis, Serbia; zoran.peric@elfak.ni.ac.rs

³ Faculty of Sciences and Mathematics, University of Pristina in Kosovska Mitrovica, Ive Lole Ribara 29, 38220 Kosovska Mitrovica, Serbia; milan.savic1@pr.ac.rs

* Correspondence: nikolasimic@uns.ac.rs

Abstract: Speaker recognition is an important classification task, which can be solved using several approaches. Although building a speaker recognition model on a closed set of speakers under neutral speaking conditions is a well-researched task and there are solutions that provide excellent performance, the classification accuracy of developed models significantly decreases when applying them to emotional speech or in the presence of interference. Furthermore, deep models may require a large number of parameters, so constrained solutions are desirable in order to implement them on edge devices in the Internet of Things systems for real-time detection. The aim of this paper is to propose a simple and constrained convolutional neural network for speaker recognition tasks and to examine its robustness for recognition in emotional speech conditions. We examine three quantization methods for developing a constrained network: floating-point eight format, ternary scalar quantization, and binary scalar quantization. The results are demonstrated on the recently recorded SEAC dataset.

Keywords: speaker recognition; convolutional neural network; quantization; emotional speech



Citation: Simić, N.; Suzić, S.; Nosek, T.; Vujović, M.; Perić, Z.; Savić, M.; Delić, V. Speaker Recognition Using Constrained Convolutional Neural Networks in Emotional Speech.

Entropy **2022**, *24*, 414. <https://doi.org/10.3390/e24030414>

Academic Editor: Antonio M. Scarfone

Received: 19 January 2022

Accepted: 28 February 2022

Published: 16 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).