

Fusion of Classical Digital Signal Processing and Deep Learning Methods (FTCAPPS)

Gomez, Angel; Sánchez, Victoria E.; Peinado, Antonio; Martín-Doñas, Juan M.; Gómez-Alanís, Alejandro; Villegas-Morcillo, Amelia; Rosello, Eros; Chica, Manuel; Garcia, Celia; Espejo, Ivan Lopez

Published in:
Proc. IberSPEECH 2022

DOI (link to publication from Publisher):
[10.21437/IberSPEECH.2022-48](https://doi.org/10.21437/IberSPEECH.2022-48)

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

[Link to publication from Aalborg University](#)

Citation for published version (APA):
Gomez, A., Sánchez, V. E., Peinado, A., Martín-Doñas, J. M., Gómez-Alanís, A., Villegas-Morcillo, A., Rosello, E., Chica, M., Garcia, C., & Espejo, I. L. (2022). Fusion of Classical Digital Signal Processing and Deep Learning Methods (FTCAPPS). In *Proc. IberSPEECH 2022* (pp. 237-240) <https://doi.org/10.21437/IberSPEECH.2022-48>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy

If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.



Fusion of Classical Digital Signal Processing and Deep Learning methods (FTCAPPS)

Angel M. Gomez¹, Victoria E. Sánchez¹, Antonio M. Peinado¹, Juan M. Martín-Doñas²,
Alejandro Gómez-Alanis³, Amelia Villegas-Morcillo¹, Eros Rosello¹, Manuel Chica¹,
Celia García¹, Iván López-Espejo⁴

¹ Universidad de Granada (Spain),

² Fundación Vicomtech, Basque Research and Technology Alliance (BRTA),
Mikeletegi 57, 20009 Donostia-San Sebastián (Spain),

³ Amazon Alexa (Germany),

⁴ Aalborg University (Denmark)

{amgg,victoria,amp,ameliamv, erosrosello, manuelc, cgrl4}@ugr.es,
jmmartin@vicomtech.org, agomezalanis@amazon.de, ivl@es.aau.dk

Abstract

The use of deep learning approaches in Signal Processing is finally showing a trend towards a rational use. After an effervescent period where research activity seemed to focus on seeking old problems to apply solutions entirely based on neural networks, we have reached a more mature stage where integrative approaches are on the rise. These approaches gather the best from each paradigm: on the one hand, the knowledge and elegance of classical signal processing and, on the other, the great ability to model and learn from data which is inherent to deep learning methods. In this project we aim towards a new signal processing paradigm where classical and deep learning techniques not only collaborate, but fuse themselves. In particular, we focus on two objectives: 1) the development of deep learning architectures based on or inspired by signal processing schemes, and 2) the improvement of current deep learning training methods by means of classical techniques and algorithms, particularly, by exploiting the knowledge legacy they treasure. These innovations will be applied to two socially and scientifically relevant topics in which our research group has been working for years. The first one is the enhancement of speech signal acquired under acoustic adverse conditions (e.g., noise, reverberation, other speakers, ...). The second one is the development of anti-fraud measures for biometric voice authentication, in which banking corporations and other large companies are strongly interested.

Index Terms: Machine Learning, Deep Neural Networks, Speech Enhancement, Multichannel speech processing, Voice Anti-spoofing

1. Introduction

In the last decade we have witnessed a radical change in the classical framework where, for the second half of the last century, signal processing founded itself as a new discipline. The need to overcome the limitations associated with the classical assumptions of linearity and stationarity, as well as the extensive use of second-order statistics, has paved the way for the irruption of machine learning techniques and, in particular, Deep Neural Networks (DNN) [1]. Although they were previously known and applied, neural networks have found fertile ground in the last decade thanks to the advances, both algorithmic and from hardware, needed to handle the huge amounts of data they

require.

Although, at first, there was a proliferation of solutions in which DNNs were conceived as black box units that performed a particular task from end-to-end, it was soon realized that these approaches wasted the accumulated in-domain knowledge we already had [2]. The efforts of the signal processing research community to take advantage of this knowledge within deep learning approaches have been remarkable in recent years [3, 4, 5, 6]. Our research team stakes on this integrative approach aiming to take a step further towards a new paradigm in which deep learning techniques are *seamlessly* incorporated into the existing signal processing theoretical body of knowledge.

To this end, we propose two lines of work: i) the design of new network architectures inspired by classical signal processing techniques, and ii) the development of new training methods that take advantage of available in-domain knowledge.

In the first case we look for new architectures, or the modification of existing ones, at the structural level, taking advantage of the available knowledge about the signals involved. Furthermore, we aim at the integration of advanced algorithms (e.g. adaptive filtering, Kalman filtering, etc.) at the layer (or cell) level, considering these as signal processing operators (in the same way that a convolutional layer can be understood as a filter-bank or a pooling as a decimation) [7]. Embedding these algorithms in a layer, or cell, would avoid the need for a heuristic adjustment of the algorithm parameters. These can directly be learned from data, while allowing us to non-linearly process signals, as the very DNNs naturally do.

In the second line of work we try to imbue DNNs with in-domain knowledge by means of the training procedure. The easiest way to do so is through the loss function used as training criterion. Thus, we propose both to adapt known classical quality or performance metrics that distill prior knowledge available on the signal, and to generate others adapted to the problem under consideration. Also, we propose the development of new data generation techniques (data augmentation) that make training possible or that, conveniently used during training, reduce the risk of overfitting. These generation techniques are based on simulation and, therefore, require knowledge of the problem and the associated classical signal processing techniques.

The resulting developments will be applied on two topics of current interest in which our research team has been working for years: multichannel speech enhancement and detection of

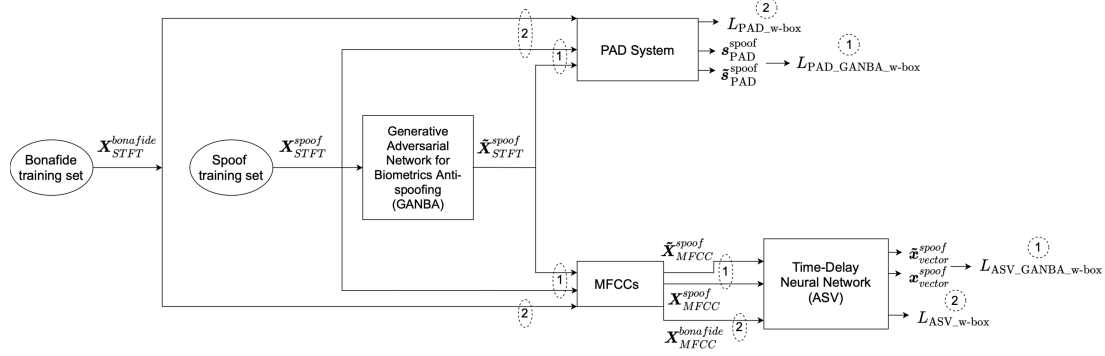


Figure 3: *Generative adversarial network for biometric anti-spoofing (GANBA) framework for white-box scenarios. Step 1: generator-only training (ASV and PAD parameters frozen). Step 2: discriminator (ASV + PAD) training. Encircled outputs corresponding to classical cross-entropy loss function.*

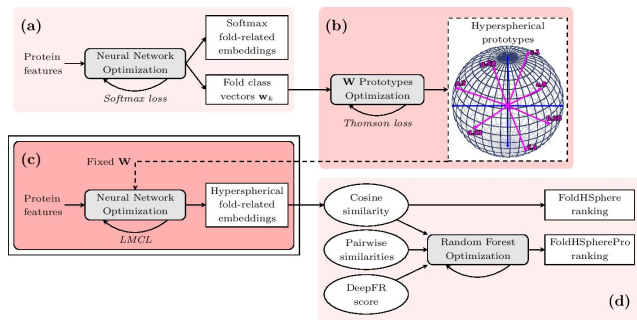


Figure 4: *Overview of the FoldHSphere approach for protein fold recognition. In the first stage a we train a neural network model using the softmax cross-entropy as loss function. We then optimize the position of the classes by our Thomson-based loss, so that they are maximally separated in the angular space. The resulting hyperspherical prototypes are used as a fixed non-trainable classification matrix.*

ing, improve the state-of-the-art systems [17] (Figure 3).

We have also investigated the use of the Large Margin Cosine Loss (LMCL) function for extraction of prototype class vectors in deep neural network training, as well as its improvement by means of the adaptation of quasi-optimal solutions to the Thomson problem in order to achieve more representative embeddings (Figure 4). We have also applied this approach to other topics as proteomic signal processing, where it was very successful for protein folding type classification [18].

Finally, techniques developed by our team have been adapted for participation in the DiCOVA 2021 Challenge, whose objective was the detection of COVID19 from audio signal (cough recordings in *track 1*). The proposed system integrated classical techniques and neural networks for signal pre-processing and cough segment detection, as well as a score fusion system between multiple classifiers [19].

4. Conclusions and future work

In this paper we have presented the project FTCAPPS, which will be executed in the period from July 2020 till June 2023. The project involves researchers from the University of Granada in collaboration with expert researchers from other countries and from the industry.

Currently the project is providing novel approaches which exploit this fused approach of paradigms, as shown in the previous section. We aim at going further in this line and achieve groundbreaking developments which seamlessly integrate DNN as part of speech signal processing. Updated information about this project can be found at <http://sigmat.ugr.es/proyectos/ftcapps>.

5. Acknowledgements

This work has been supported by the project PID2019-104206GB-I00 funded by MCIN/ AEI /10.13039/501100011033.

6. References

- [1] S. Haykin, "Neural Networks Expand SP's Horizons", *Signal Processing Magazine*, pp. 24 - 49, 1996.
- [2] E. Vincent, "Is audio signal processing still useful in the era of machine learning?," *2015 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 7-7, 2015.
- [3] Y. Xu, J. Du, L.R. Dai, C.H. Lee, "An experimental study on speech enhancement based on deep neural networks", *IEEE Signal Process. Letters* vol. 21, no. 1, pp. 65-68, Jan. 2014.
- [4] K. Tan, J. Chen, D. Wang, "Gated residual networks with dilated convolutions for monaural speech enhancement", *IEEE Trans. on Audio, Speech, and Language Processing* vol. 27, no. 1, pp. 189-198, Jan. 2019.
- [5] D. Wang, J. Chen, "Supervised speech separation based on deep learning: An overview", *IEEE Trans. on Audio, Speech, and Language Processing*, 2018.
- [6] J.M. Martín-Doñas, A.M. Gomez, J.A. Gonzalez, A.M. Peinado, "A Deep Learning Loss Function based on the Perceptual Evaluation of the Speech Quality", *IEEE Signal Processing Letters*, vol. 25, no. 11, pp. 1680-1684, Nov. 2018.
- [7] A.K. Maier, C. Syben, et al. "Learning with known operators reduces maximum error bounds", *Nature machine intelligence*, 1(8), 2019.
- [8] A.M. Gomez, V. Sanchez, et al. "Nuevas aproximaciones conexiones integradas para procesamiento de señal", Ministerio de Economía y Competitividad: Programa estatal de fomento de la investigación científica y técnica, 2016-2019, (TEC2016-80141-P).
- [9] D. Wang, G.J. Brown, "Computational Auditory Scene Analysis: Principles, Algorithms, and Applications", Hoboken, NJ, USA: Wiley, 2006.
- [10] Y. Qian, N. Chen, K. Yu, "Deep Features for automatic spoofing detection", *Speech Communication*, 85, 2016.

- [11] F. Ma, L. Chai, J. Du, D. Liu, Z. Ye, C.-H. Lee, "Acoustic Model Ensembling Using Effective Data Augmentation for CHiME-5 Challenge", in Proc. *InterSpeech* 2019.
- [12] S. Novoselov, V. Shchemelinin, A. Shulipa, A. Kozlov, I. Kremnev, "Triplet Loss Based Cosine Similarity Metric Learning for Text-independent Speaker Recognition", in Proc. *InterSpeech* 2018.
- [13] R. Li, M. Zhao, Z. Li, L. Li, Q. Hong, "Anti-Spoofing Speaker Verification System with Multi-Feature Integration and Multi-Task Learning", in Proc. *InterSpeech* 2019.
- [14] J.M. Martín-Doñas, A.M. Peinado, I. López-Espejo, A.M. Gómez, "Dual-channel eKF-RTF framework for speech enhancement with DNN-based speech presence estimation", in proceedings of *IBERSPEECH 2021*, pp. 31-35, 24-25 March 2021.
- [15] J.M. Martín-Doñas, J. Jensen, Z.-H. Tan, A. M. Gómez, A. M. Peinado, "Online multichannel speech enhancement based on recursive EM and DNN-based speech presence estimation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 3080-3094, 2020.
- [16] A. Gomez-Alanis, J. A. Gonzalez-Lopez, S. Pavankumar Dubagunta, A. M. Peinado, M. Magimai-Doss, "On Joint Optimization of Automatic Speaker Verification and Anti-spoofing in the Embedding Space", *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 1579-1593, 2021.
- [17] A. Gomez-Alanis, J. A. Gonzalez-Lopez, A. M. Peinado, "GANBA: Generative Adversarial Network for Biometric Anti-spoofing", *MDPI Applied Sciences*, vol. 12, no. 3, pp. 1454, 2022.
- [18] A. Villegas-Morcillo, V. Sanchez, A. M. Gomez, "FoldHSphere: Deep Hyperspherical Embeddings for Protein Fold Recognition", *BMC Bioinformatics*, 22(490), pp. 1-21. 2021.
- [19] D. Madhu R. Kamble, J. A. Gonzalez-Lopez, T. Grau, J. M. Espin, L. Cascioli, Y. Huang, A. Gomez-Alanis, J. Patino, R. Font, A. M. Peinado, A. M. Gomez, N. Evans, M. A. Zuluaga, M. Todisco, "PANACEA cough sound-based diagnosis of COVID-19 for the DiCOVA 2021 Challenge", in proceedings of *INTER-SPEECH 2021*, pp. 906-910, 2021.