2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

WaveFusion Squeeze-and-Excitation: Towards an Accurate and Explainable Deep Learning Framework in Neuroscience

Michael Briden[1] and Narges Norouzi[1]

Abstract— **We introduce WaveFusion Squeeze-and-Excite, a multi-modal deep fusion architecture, as a practical and effective framework for classifying and localizing neurological events. WaveFusion SE is composed of lightweight CNNs for per-lead time-frequency analysis and an attention network called squeeze and excitation network with a temperature factor for effectively integrating lightweight modalities for final prediction. Our proposed architecture demonstrates high accuracy in classifying subjects' anxiety levels with an overall accuracy of 97.53%, beating prior approaches by a considerable margin. As will also be demonstrated in the paper, our approach allows for real-time localization of neurological events during the inference without any additional post-processing. This is a great step towards an explainable DL framework for neuroscience applications.**

I. Introduction

Electroencephalogram (EEG) is used as a non-invasive technique for studying brain activity by measuring electric potential differences generated by neuronal activity [1]. EEG displays high temporal resolution making it ideal for capturing fast and dynamic cognitive events [2] and can also be transformed to analyze the important frequency and power features. Combined with MRI and computational models, EEG can be used to answer the "inverse problem" of localizing the sources of neurological activities within the brain and allowing researchers to localize neurological activities associated with many cognitive tasks, disorders, drug side-effects, and epilepsy [3].

EEG plays a vital role in diagnosing epilepsy, detecting, and localizing seizure foci. A standard method for detecting influential regions of brain activity is through Independent Component Analysis (ICA) [4], [5]. In addition to diagnosis and localization, EEG can be used to assess the effectiveness of removing epileptic foci [6]. Additionally, the localization of neurological activity plays a role in understanding different psychological disorders [7].

While source localization of EEG offers many benefits for neurological analysis, it is also an uphill task, and many Machine Learning (ML) techniques have been developed to automate the process. Researchers have investigated using Random Forest (RF) classifiers [8], Support Vector Machines (SVMs) [9] and Neighborhood Component Analysis (NCA) [10] to tackle the problem. In addition to classical ML approaches, many Deep Learning (DL) techniques have also been developed to localize EEG signals. Hussein et al. created a Long Short-Term Memory (LSTM) model to determine if EEG signals were sources for epileptic activity [11]. Cui et al. created a two-part Spatio-Temporal Neural Network (NN) model that takes in EEG signals and generates source locations through regression [12]. Daoud and Bayoumi experimented with two architectures for automatic epileptic focus localization [13].

(1) Michael Briden (email: mbriden@ucsc.edu) and Narges Norouzi (email: nanorouz@ucsc.edu) are the with UC Santa Cruz, Santa Cruz, CA 95064 USA

A. Our Contribution

Very little work has been done to use the DL framework to localize and explain neurological phenomena. As such, we propose WaveFusion Squeeze Excite (SE) architecture and outline the process for classifying neurological events and identifying influential regions of the brain using the Continuous Morlet Wavelet Transform (CMWT) and attention mechanisms [14]. First, we outline the process for building compact EEG data formats, using CMWT, that are compatible with Guided Back-Propagation (GBP) for Convolutional Neural Network (CNN) activations. We then explain WaveFusion SE architecture, WaveFusion SE's training routine, and how the model uses attention mechanisms and conventional visualization techniques to classify and localize neurological phenomena.

II. Materials and Methods

A. Dataset

We have demonstrated the effectiveness of our framework on a dataset from the Max Planck Institute Leipzig Mind-Brain-Body (LEMON). The study is conducted to learn the relationship between mental and somatic well-being [15]. The dataset contains the results of six psychological tests, MRI, and EEG readings for 227 patients. The subjects' anxiety levels are labeled as "Mild", "Moderate", and "Severe" based upon the State-trait anxiety inventory X2 score (a test taken by the subject before the experiment). The EEG recordings are 16 minutes long with 62 leads arranged in the 10-20 comprehensive localization system. Amplitude resolution is 0.1μV, and recordings are made with a bandpass filter between 0.015 Hz and 1 kHz and a 2500 Hz sampling rate. Readings are taken with eyes closed for 8 minutes and eyes open for another 8 minutes. The readings are then segmented into 16 blocks. Each block is 60-second long with Eyes Closed (EC) or Eyes Open (EO). The readings are then down-sampled to 250 Hz and bandpass filtered to 1-45 Hz with a Butterworth filter.

For each sample in the LEMON dataset, CMWT is applied with 32 scales, a Gaussian width of 0.4 Hz, and a center frequency of 1 Hz to the per-lead EEG readings. The absolute value of the CMWTs is computed to give us 61 32×250

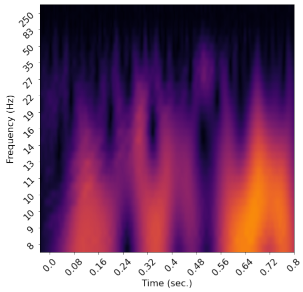

Fig. 1. An "inferno" sequential colormap of a scalogram generated by the CMWT.

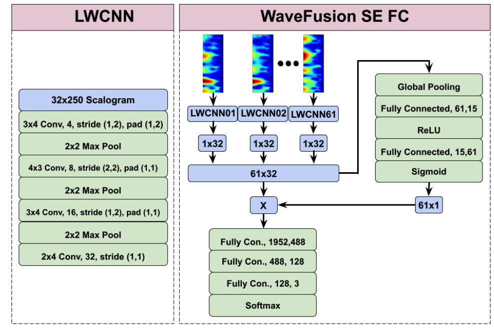

Fig. 2. WaveFusion SE FC Model with Squeeze and Excitation Network and three fully-connected layers for classification.

scalograms per sample. Fig. 1 shows an "inferno" sequential colormap example of a 32×250 scalogram with scales converted to frequency. The scalograms are concatenated to create $61 \times 32 \times 250$ tensors for inference.

III. Models and Algorithms

This section describes the WaveFusion SE framework (architecture, Squeeze-and-Excite, Networks (SENs), and training) and compares the WaveFusion SE architectures with 3D-CNN models. We also describe the effectiveness of each model in identifying influential regions of brain activity.

A. WaveFusion SE Framework

To combine the strengths of a CNN architecture and a robust fusion technique for combining information about neurological activities in different brain regions, we created the WaveFusion SE framework. WaveFusion SE uses a lightweight attention module to treat each EEG lead as a modality with an individually-trained CNN. In the following sections, we discuss the components of WaveFusion SE, such as LWCNNs, SEN, and WaveFusion SE architecture.

1) Lightweight CNN: WaveFusion SE comprises 61 Lightweight 2D-CNNs (LWCNNs) consisting of four convolution layers with ReLU activations and three max-pool layers. (Fig. 2). WaveFusion SE takes in $61 \times 32 \times 250$ scalogram tensors. The LWCNNs are indexed 0 to 60 with the i^{th} LWCNN, taking in the scalograms generated by the i^{th} EEG lead. Each LWCNN generates a 1×32 feature map which are then combined using weighted concatenation before being sent to a final classification layer during which the influences of some LWCNNs are suppressed while others are boosted. The weights for LWCNN feature maps are calculated using an SEN.

2) Squeeze and Excitation Network (SEN): A lightweight attention module is used to weigh the feature maps generated by the 61 LWCNNs. Attention modules are often used in image segmentation models and allow for CNN architectures to focus on essential details in an image by up and down weighting CNN activations [14]. Hu et al. propose a lightweight attention network called "squeeze and excitation network" which can be used in between DL model layers in order to up weight important CNN channels before they are sent to the next CNN layer [16].

We employ SEN as a lightweight attention module built atop the LWCNN models. Given an input of 61 scalograms denoted $\boldsymbol{X} = [\boldsymbol{x}_0, \boldsymbol{x}_1, ...\boldsymbol{x}_{60}] \in \mathbb{R}^{61\times32\times250}$, the tensor of feature maps $\boldsymbol{U} = [\boldsymbol{u}_0, \boldsymbol{u}_1, ...\boldsymbol{u}_{60}] \in \mathbb{R}^{61\times32}$ is computed when the i^{th} LWCNN model maps an input $\boldsymbol{x}_i \in \mathbb{R}^{32\times250}$ to a feature map $\boldsymbol{u}_i \in \mathbb{R}^{1\times32}$. Attention weight, π_i for each feature map, is computed by first passing $\boldsymbol{U}$ through a global pooling layer which provides a summary of the amount of information in each feature map. Then, the 61×1 tensor of averages is sent to an encoder-decoder model, which contains two fully-connected layers with ReLU activation in between them. The first layer acts as an encoder and further condenses the input size of 61×1 to 15×1. The second fully connected layer acts as a decoder and expands the final output back to 61×1 where sigmoid activation computes attention weights, π_i, for each of the 61 feature maps. Then, each of the 61 1×32 feature maps in $\boldsymbol{U}$ is multiplied by its corresponding attention weight. $\boldsymbol{U}$ is then flattened and sent to the final classification layers.

WaveFusion SE tends to overfit when the SEN over emphasizes a small number of channels while down-weighting others. To address this issue, the weights are flattened by using temperature τ within the sigmoid activation function:

$$\pi_i = \frac{e^{z_i/\tau}}{e^{z_i/\tau} + 1}$$

where z_i is the summed and weighted input to the last fully connected layer [17]. This flattening drives probability scores towards 0.5 and allows for even optimization across the LWCNN models.

3) WaveFusion SE Architectures: We implemented two forms of feature integration using SENs and the feature maps generated by LWCNNs. *WaveFusion SE FC* (Fig. 2) contains 61 LWCNNs, a SEN, and a three-layer fully-connected classifier. Once $\boldsymbol{U}$ is weighted, the tensor is flattened and sent to a three-layer fully-connected classifier with ReLU activation and a final softmax activation. We also experimented with different classifier architectures to identify the best classifier for the task, including a single fully-connected layer that we call *WaveFusion SE*. In the WaveFusion SE architecture, the tensor of feature maps $\boldsymbol{U}$

is weighted, flattened, and sent to a final fully-connected classification layer with a softmax activation.

B. Analysis of Scalograms with 3D-CNN

To compare the performance of WaveFusion SE in prediction and localization, we created Wavelet 3D-CNN (W3DCNN) model. The W3DCNN has four 3D convolutional layers with ReLU activations followed by 3D batch normalization. There are two 3D max-pooling layers to add 3D translational invariance. The $61 \times 32 \times 250$ tensors of data are fed directly to W3DCNN.

IV. Experiments and Results

We tested the classification and localization effects of WaveFusion SE FC and WaveFusion SE using the LEMON dataset. WaveFusion SE FC and Wavefusion SE are both trained with Adam optimizer, a learning rate of 0.001, weight decay of 7.5×10^{-4}, and a batch size of 500. A temperature of 35 was chosen experimentally. The dimensionality reduction of 61 to 15 within the SE modules was also chosen experimentally. The baseline comparison model, W3DCNN, is trained with Adam optimizer, a learning rate of 0.001, weight decay of 7.5×10^{-4}, and a batch size of 500.

A. Classification Results

Table I summarizes the F1 score and the percentage of correct classifications for Mild, Moderate, and Severe classes. The WaveFusion SE FC performs best during the overall classification task. The model also outperforms other architectures (WaveFusion SE and W3DCNN) for the classification of Mild and Moderate cases with 97.4% and 96.6% accuracy, respectively. The model also have a similar accuracy in classifying Severe cases to that of the W3DCNN architecture, mainly due to more distinguished features in this class.

B. Localizing Neurological Activities

After the WaveFusion SE models are trained, we exploit properties of the attention network to identify influential regions of brain activity. We then compare outputs to Saliency maps (SMs) generated by a trained W3DCNN model using GBP. Three samples from Mild, Moderate, and Severe classes are used for illustration. As will be described below, the WaveFusion SE models show a higher capability for localizing brain activity, both in terms of accuracy and computational cost. Our proposed approach can localize brain activity at the inference time without any additional post-processing.

1) Localization using WaveFusion SE: To localize neurological activities using the WaveFusion SE models, we use the 61 attention weights, π_i, generated during the inference phase and interpolate them onto brain and scalp models. The attention weights display desirable properties for comparing activity across regions. Since the weights are learned during model training, the influence of each LWCNN is adjusted automatically. Moreover, the SEN attention weights are proportional to the amount of activation in each channel but

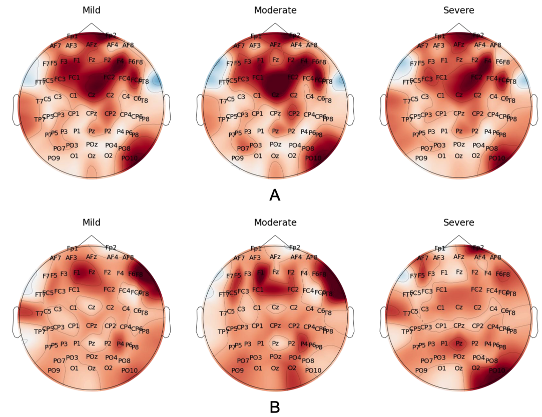

Fig. 3. Interpolation of most influential modalities onto a 2D topology map. A illustrates the interpolation of attention weights generated from the WaveFusion SE FC architecture. B shows the interpolation for the WaveFusion SE architecture.

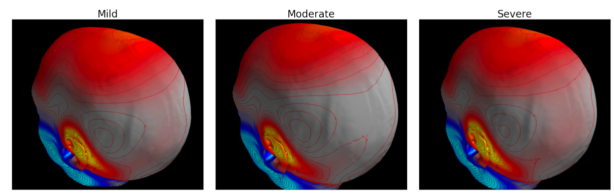

Fig. 4. Interpolation of most influential modalities onto a 3D scalp model constructed with MRI and EEG mappings.

are not prone to over-optimizing channel-specific details. Likewise, the SEN attention weights are limited to between 0 and 1, allowing for a fair comparison across modalities. Fig. 3.A illustrates the interpolation of π_i values generated from WaveFusion SE FC using the Mild, Moderate, and Severe examples onto 2D scalp topology maps using MNE software. In the Mild example, we can see a considerable amount of influence from the frontal and right occipital modalities, which is also shown to be an area of high activation in the W3DCNN Mild activation map. In the Moderate example, there is considerable influence from the FC6 region in addition to the frontal and occipital regions, which is shown to be an influential region in the average Moderate SM. Lastly, PO10 also illustrates high activation in both the Severe 2D topology map and the average SM. Fig. 3.B shows the interpolation for WaveFusion SE. Despite noticeable differences, both models focus primarily on the frontal and occipital regions. WaveFusion SE's attention can also be projected onto MRI images for more detailed analysis of influential brain regions (Fig. 4).

2) 3D-CNN Activation Visualization: We attempt to identify the leads and frequency ranges that contribute most to the model's predictions. We compute the average Mild, Moderate, and Severe activations of the W3DCNN model for samples in the test set. Next, we average over the time dimension to give us the average of the 32×61 frequency-by-lead activations for the Mild, Moderate, and Severe samples. Fig. 5 shows that activations range across leads for the Mild and Moderate samples while the highest level activation

TABLE I

F1 SCORE AND ACCURACY PER CLASS FOR WAVEFUSION AND W3DCNN MODELS ARE DISPLAYED IN COLUMNS 1 TO 6. OVERALL F1 SCORE AND OVERALL ACCURACY IS DISPLAYED IN COLUMN 8.

	Mild		Moderate		Severe		Overall
Model	**F1 Score**	**Accuracy**	**F1 Score**	**Accuracy**	**F1 Score**	**Accuracy**	**Accuracy**
WaveFusion SE FC	**0.979**	**97.4%**	**0.976**	**96.6%**	**0.971**	98.6%	**97.53%**
WaveFusion SE	0.914	91.0%	0.908	95.0%	0.889	85.2%	90.40%
W3DCNN	0.957	96.0%	0.963	94.0%	0.968	**98.8%**	96.27%

is centered between C3 and C1 for the Severe samples. Additionally, the region AF3 to FC3 receives the highest level of activation for Mild and Moderate samples.

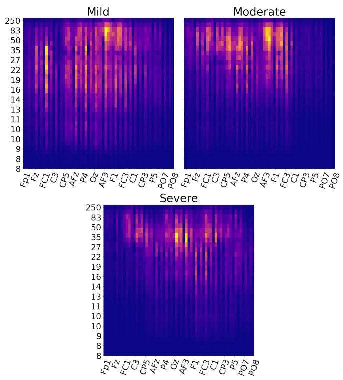

Fig. 5. Average W3DCNN activation for the Mild, Moderate and Severe classes.

V. FUTURE WORK

WaveFusion SE implements late-fusion techniques with 61 learners. As such, training is a lengthy process as the parameters of all LWCNNs are updated with each epoch. We aim to employ an early fusion approach based on multi-task learning to alleviate the number of the lightweight models needed at the inference time. Given WaveFusion SE's performance in classification and the ability to localize neurological activity, we seek to apply our fusion technique to the task of localizing seizure onset with both tomography-like techniques and intracranial EEG.

VI. CONCLUSION

We present WaveFusuion SE as a multi-modal fusion architecture to classify neurological events and identify regions of the brain that contribute most to the model's prediction. The WaveFusion SE uses a combination of LWCNN sub-models, trained independently for extracting localized time-frequency features, and SEN to allow for classifying and localizing neural activity. Using the proposed WaveFusion SE architecture, the classification and localization can be done in real-time, giving essential insights into the neural process. Our proposed framework demonstrates high accuracy in classifying subjects' anxiety levels with an overall accuracy of 97.53%, beating prior approaches by a large margin. The authors plan to pursue applying this technique in predicting and localizing the onset of a seizure.

REFERENCES

[1] S. Asadzadeh, T. Y. Rezaii, S. Beheshti, A. Delpak, and S. Meshgini, "A systematic review of EEG source localization techniques and their applications on diagnosis of brain abnormalities," *Journal of Neuroscience Methods*, p. 108740, 2020.

[2] M. X. Cohen, *Analyzing neural time series data: theory and practice*. MIT press, 2014.

[3] C. M. Michel and B. He, "EEG source localization," in *Handbook of clinical neurology*. Elsevier, 2019, vol. 160, pp. 85–101.

[4] D. Chen, S. Wan, and F. S. Bao, "Epileptic focus localization using EEG based on discrete wavelet transform through full-level decomposition," in *2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP)*. IEEE, 2015, pp. 1–6.

[5] Lisha Sun, Ying Liu, and P. J. Beadle, "Independent component analysis of EEG signals," in *Proceedings of 2005 IEEE International Workshop on VLSI Design and Video Technology, 2005.*, 2005, pp. 219–222.

[6] P. E. Coutin-Churchman, J. Y. Wu, L. L. Chen, K. Shattuck, S. Dewar, and M. R. Nuwer, "Quantification and localization of EEG interictal spike activity in patients with surgically removed epileptogenic foci," *Clinical neurophysiology*, vol. 123, no. 3, pp. 471–485, 2012.

[7] M. A. Klados, E. Paraskevopoulos, N. Pandria, and P. D. Bamidis, "The impact of math anxiety on working memory: A cortical activations and cortical functional connectivity EEG study," *IEEE Access*, vol. 7, pp. 15 027–15 039, 2019.

[8] A. Subasi, S. Jukic, and J. Kevric, "Comparison of EMD, DWT and WPD for the localization of epileptogenic foci using random forest classifier," *Measurement*, vol. 146, pp. 846–855, 2019.

[9] J. Cimbalnik, P. Klimes, V. Sladky, P. Nejedly, P. Jurak, M. Pail, R. Roman, P. Daniel, H. Guragain, B. Brinkmann *et al.*, "Multi-feature localization of epileptic foci from interictal, intracranial EEG," *Clinical Neurophysiology*, vol. 130, no. 10, pp. 1945–1953, 2019.

[10] S. Raghu and N. Sriraam, "Classification of focal and non-focal EEG signals using neighborhood component analysis and machine learning algorithms," *Expert Systems with Applications*, vol. 113, pp. 18–32, 2018.

[11] R. Hussein, H. Palangi, R. K. Ward, and Z. J. Wang, "Optimized deep neural network architecture for robust detection of epileptic seizures using EEG signals," *Clinical Neurophysiology*, vol. 130, no. 1, pp. 25–37, 2019.

[12] S. Cui, L. Duan, B. Gong, Y. Qiao, F. Xu, J. Chen, and C. Wang, "EEG source localization using spatio-temporal neural network," *China Communications*, vol. 16, no. 7, pp. 131–143, 2019.

[13] H. Daoud and M. Bayoumi, "Deep learning approach for epileptic focus localization," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 14, no. 2, pp. 209–220, 2019.

[14] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," *arXiv preprint arXiv:1706.03762*, 2017.

[15] A. Babayan, M. Erbey, D. Kumral, J. D. Reinelt, A. M. Reiter, J. Röbbig, H. L. Schaare, M. Uhlig, A. Anwander, P.-L. Bazin *et al.*, "A mind-brain-body dataset of MRI, EEG, cognition, emotion, and peripheral physiology in young and old adults," *Scientific data*, vol. 6, p. 180308, 2019.

[16] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.

[17] Y. Chen, X. Dai, M. Liu, D. Chen, L. Yuan, and Z. Liu, "Dynamic convolution: Attention over convolution kernels," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 11 030–11 039.