Deep Convolutional Neural Network Applied to Electroencephalography: Raw Data vs Spectral Features

D. Truong[1*], M. Milham[2], S. Makeig[1], A. Delorme[1]

[1] Swartz Center for Computational Neuroscience, Institute for Neural Computation, University of California San Diego
[2] Center for the Developing Brain, Child Mind Institute

Introduction

The success of deep learning in other domains has encouraged neuroscience researchers, specifically in electrophysiological neuroimaging, to start applying deep learning to make predictions on EEG data. Research remains open on the network architecture and the feature space that is most effective for EEG decoding.

This work compares deep learning using minimally processed EEG raw data versus deep learning using EEG spectral features, using two different deep convolutional neural networks (CNN). We apply them on a sex classification task (female vs. male) on 24-channel EEG from a large corpus of 1,574 participants.

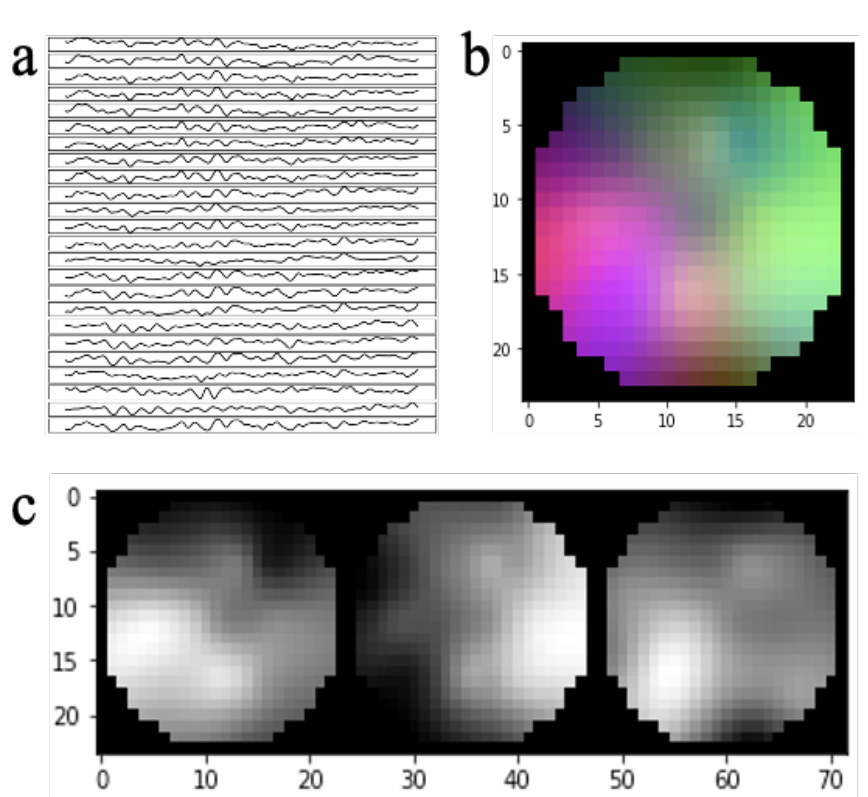
We improve on state-of-the-art classification performance for this type of classification problem, and show that in all cases, raw data classification leads to superior performance as compared to spectral EEG features.

Interestingly we show that a high-performing CNN model in computer vision tailored to process EEG spectral features has increased performance when applied to raw data classification. Our approach suggests that the same CNN used to process EEG spectral features yield superior performance when applied to EEG raw data.

Data

EEG recordings. The publicly available high-density EEG data were recorded at a sampling rate of 500 Hz with a bandpass of 0.1 to 100 Hz, using a 128-channel EEG geodesic hydrogel system by Electrical Geodesics Inc. (EGI) (Alexander et al., 2017). We only considered the resting state data during eye-closed condition.

Raw data preprocessing. We minimally preprocessed the raw EEG: baseline removal, down-sampling to 128 Hz, band-pass filtering between 0.25–25 Hz, re-referencing to the averaged mastoids, and automatically rejecting bad channels (removed channels were then interpolated). We segmented eye-closed data periods into non-overlapping 2-s windows: each preprocessed 2-s epoch was used as a sample for our final dataset.

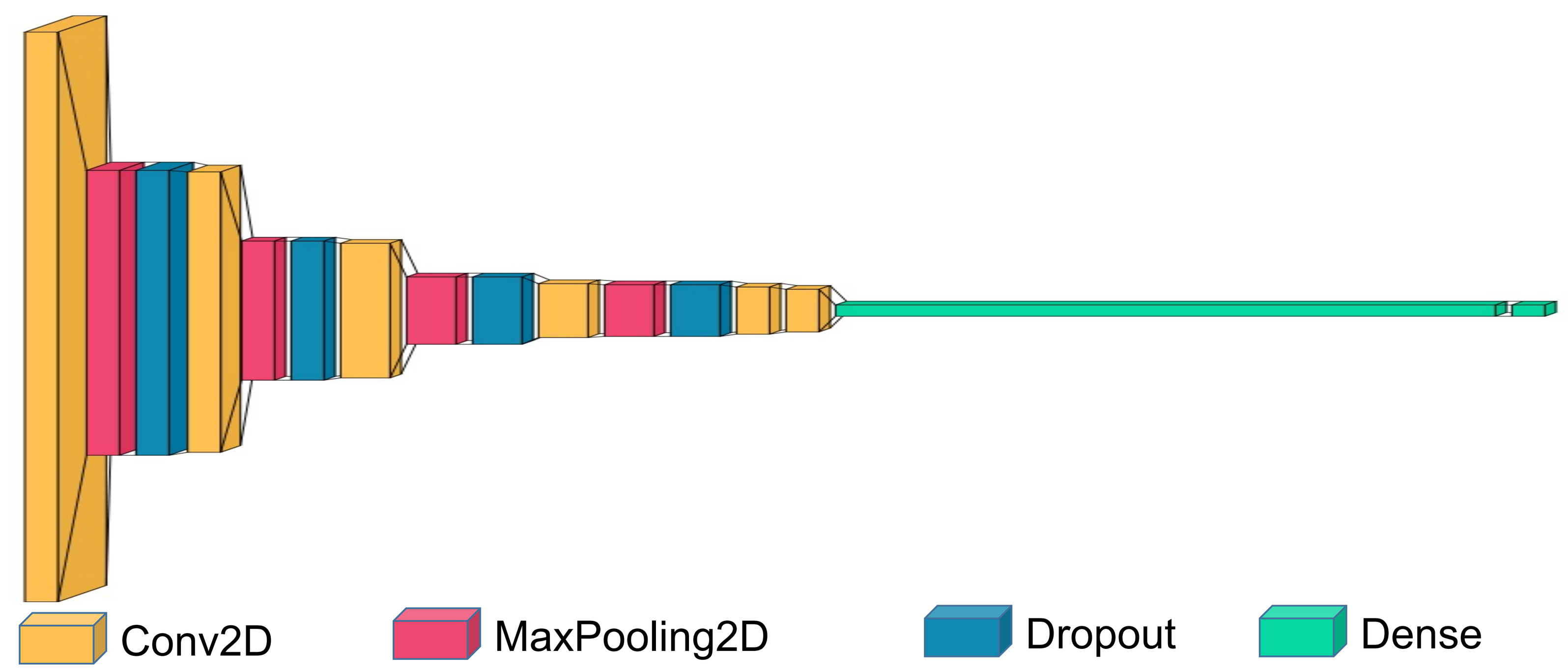


Raw and spectral data samples. (a) shows a 24x256 raw EEG sample. (b) shows a combined 24x24x3 scalp topography spectral data. (c) shows a 24x72 scalp topography of the three frequency bands represented side-by-side.

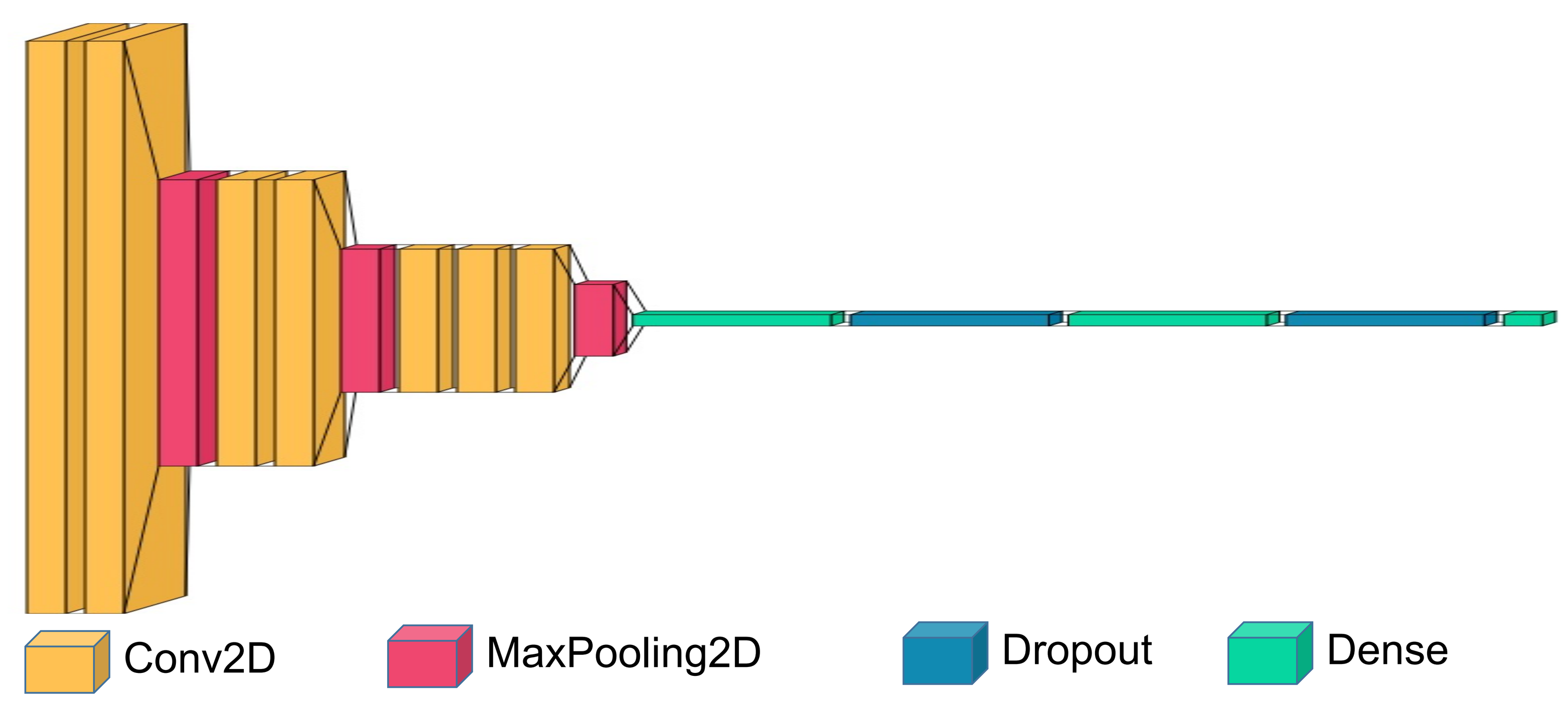
EEG-PSD images-based features. We extracted power spectral density (PSD) scalp maps in three EEG bands, theta (4-7 Hz), alpha (7-13 Hz), and beta (13-25 Hz) for each channel, then plotted the power spectrum heat maps for the three EEG bands. Image pixel values were rescaled to be between 0 and 255 and values outside the disk outlining the head limit were set to 0. The three scalp topographies may be combined into a chromatic image or placed side by side to form a 2-D image.

Convolutional neural networks

We based our models on two main network architectures. One of them was from Putten et al. (2018) tailored to process raw data on the same classification task; the other was derived from the VGG16 vision network (Simonyan and Zisserman, 2015) which has been used to process EEG spectral features (Siddharth et al., 2019). We call our application of the CNN model described in (Putten et al., 2018) to raw data R-SCNN (Raw/Sex CNN). R-SCNN repurposed for spectral data is called S-SCNN. Similarly, we refer to our modified VGG16 model applied to raw EEG data as R-VGG and use S-VGG for the version of the model trained on spectral data.



R-SCNN and S-SCNN schematic. Each model consists of 6 convolutional layers. Each of the first 4 convolutional layers is followed by a max pooling layer then a dropout layer using a 25% dropout rate. Strides and padding are different between the two models in some layers. The classification layer is a 2-unit fully-connected layer with a softmax activation.



S-VGG and R-VGG schematic. Each model consists of 7 convolutional layers, separated by max pooling operations between every 2-3 layers. Convolutional filter configurations are different from that of R-SCNN and S-SCNN. Three fully-connected layers are interleaved by dropout layers with dropout rate 50%.

Experiment

From 127,174 samples generated from 1574 participants (50% female), we split the balanced data into training, validation, and test sets in size ratio 60:30:10. Each sample received a binary label, indicating a male (0) or a female (1).

All models were trained on a single NVIDIA V100 SMX2 GPU (32 GB) with Python 3.7.10 and PyTorch 1.3.1. During training, the validation data were used to assess its performance and to inform a stopping rule.

Overfitting is a common issue in deep learning. One common practice to avoid overfitting is early stopping, in which training is stopped (and model performance evaluated) when validation accuracy starts to plateau or decrease as training accuracy continues to grow. We saved intermediary models during training and since each of the models overfit the data at different rates, we chose to evaluate the models at different training epochs.

Results

Evaluation metric. Per-sample prediction accuracy was reported for all models. We also obtained a final performance estimate for the test dataset by taking the mean gender probability p_{ave} of the first 40 2-s samples for each subject; if $p_{ave} > 0.5$, the subject was classified as 1 (female) and as 0 (male) otherwise. We refer to this as per-subject performance.

Model	Per-sample	Per-subject
R-SCNN	80.6 (79.7 to 81.5)	85.1 (84.3 to 85.9)
R-VGG	**83.1** (82.7 to 83.4)	**87.0** (86.6 to 87.4)
S-SCNN	79.0 (78.7 to 79.3)	83.2 (82.1 to 84.3)
S-VGG	77.1 (76.8 to 77.4)	81.3 (80.0 to 82.6)

Models' classification accuracy in percentage. 95% confidence interval is indicated in parenthesis. Bolded values indicate best performance.

Performance on raw EEG. Only per-subject classification results (using votes on the first 40 data segments) were reported in (Putten et al., 2018), achieving 81% prediction accuracy. Here we report both per-sample and per-subject classification performance and show improvement in prediction accuracy for both models trained on the raw EEG data. R-SCNN achieved 85.1% accuracy, while R-VGG achieved 87% per-subject performance. Because there is no overlap between the 95% confidence intervals (equivalent to an unpaired parametric t-test), the difference is statistically significant at the p=0.05 threshold.

Performance on spectral EEG. S-VGG performed significantly worse than S-SCNN on per-sample performance. R-SCNN significantly (about 1%) outperforms S-SCNN on per-sample performance. In all cases, R-VGG gave significantly better performance.

Conclusions

We have shown that a neural network tailored to process simple EEG spectral features gave improved (sex) classification performance when applied instead to the raw data. This model also outperformed a previous sex classification deep learning approach. Our approach suggests that the same convolutional networks used to process image-based inputs, specifically scalp images of EEG spectral features, can give superior performance when applied instead to EEG raw data. Despite the popularity of reusing vision inspired convolutional neural network architecture to process spectral scalp topographies, such preprocessing is not warranted.

Acknowledgements

Expanse supercomputer time was provided via XSEDE allocations and NSG (the Neuroscience Gateway). We thank Amitava Majumdar, Subhashini Sivagnanam, and Kenneth Yoshimoto for providing computational resources.

We also thank professor Xiaolong Wang of UC San Diego for his teaching and guidance.

References

Alexander, L., Escalera, J., Ai, L. et al., An open resource for transdiagnostic research in pediatric mental health and learning disorders. Scientific Data, 2017. 4, 170181.

Siddharth, Jung, T., Sejnowski, T., J., Utilizing Deep Learning Towards Multi-modal Bio-sensing and Vision-based. IEEE Transactions on Affective Computing, 2019.

Simonyan, K., & Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. CoRR, 2015. abs/1409.1556.

Van Putten, M. J. A. M., Olbrich, S. and Arns, M., Predicting sex from brain rhythms with deep learning. Scientific Reports, 2018. 8.

* Dung (Young) Truong [dutruong@sccn.ucsd.edu]

EEG: Analytical approaches and applications, Virtual symposium, Sapien Labs, September 1-2, 2021