U-EEG: A Deep Learning Autoencoder for the Detection of Ocular Artifact in EEG Signal

J. Onners, M. Alam, B. Cichy, N. Wymbs and J. Lukos

Intelligent Sensing Branch, Naval Information Warfare Center Pacific, San Diego, California, USA
{jeffrey.c.onners, mohammad.alam, benjamin.cichy, nicholas.wymbs, jlukos}@niwc.navy.mil

Abstract— Current methods to remove blink artifacts from electroencephalography (EEG) brain signals, such as independent component analysis (ICA), have reduced functionality when processing EEG datasets that have a relatively small number of EEG electrodes. We evaluated an approach to address this limitation by modifying a U-Net convolutional neural network (CNN) to extract blink artifacts from EEG signals contaminated with electrooculography (EOG) signal. Crucially, U-Nets have high prediction accuracy even when trained on small datasets (i.e., low EEG electrode count). Our model (*U-EEG*) removed ocular artifacts from contaminated signals with a higher accuracy compared to the industry standard ICA algorithm for EEG analysis. With respect to an artifact-free baseline signal, we found that U-EEG had lower root mean square error (RMSE) for artifact detection compared to ICA (U-EEG: *3.134* ± 0.893; ICA: *5.793* ± 1.769). Our findings highlight the utility of data-driven models for EEG artifact correction, and the potential these methods have for the development of a real-time brain-computer interface (BCI) for real-world solutions.

I. Introduction

Deep Neural Net research has improved BCI performance in EEG signal analysis. Neural signals within an EEG waveform are naturally low in amplitude and highly susceptible to noise, impeding data analysis and potential BCI effectiveness. Recordings from EEG electrodes contain a mixture of both desired neural activity and undesired noise, such as cardiac or eye blinks. For instance, a single eye blink can produce a signal-to-noise ratio low enough to practically eliminate useful neural activity in the recorded signal. Without correction, the resulting artifacts can hinder a BCI's ability to account for specific neural cues. Most BCI testing is done in laboratory settings where noise can be minimized, but solutions are still needed so that BCIs can be reliably deployed in natural settings containing more noise. In neuroscience, ICA blind source separation has been a popular solution for EEG noise reduction to negate artifacts in BCI applications [1]. While the ICA-based technique is generally effective, correction accuracy is dependent on the number of electrodes used and costly manual intervention. Additional reference signals are also recommended for best performance. For ocular artifact, an EOG reference is derived from additional reference sensors placed around the eyes. While references provide a solution for a controlled laboratory setting, they may be inconvenient for a fieldable BCI with real-time filtering capability.

In this paper, we introduce a deep learning autoencoder U-Net, called U-EEG, to demonstrate blink artifact separation from contaminated EEG signals. The trained U-EEG model does not require EOG noise references during testing, which may be more suitable than ICA for real-time use. Furthermore, we show that our data-driven approach to filtering blink artifacts from EEG signal, even without EOG noise references, is more accurate compared to the standard ICA approach with additional EOG noise references. In reviewing the effectiveness of U-EEG artifact isolation and separation, we show the utility of deep learning methods in artifact extraction, which ultimately provides motivation for additional research on U-Nets for BCI development.

II. Methods

2.1. Dataset

We utilized an existing dataset from Klados et al. [2] to develop a U-Net model that is capable of correcting eye blink artifact from neural EEG signal. The dataset included 27 subjects wearing 19 electrodes on the scalp using standard 10-20 electrode placement [3]. Additional electrodes were placed above and below the left eye to record vertical EOG (VEOG) data, and on the outer canthi of both eyes to obtain horizonal EOG (HEOG) data. EEG recordings were sampled at 200 Hz and band-pass filtered at 0.5–40 Hz to remove external noise. EOG recordings were band-pass filtered at 0.5–5 Hz. The recordings were separated into two 30 s sessions: a closed-eye session to gather blink-free EEG signals (*Pure*), and an open-eye session to obtain data with real blinks. We later use *Pure* blink-free data as the ground truth for machine learning training. Klados and colleagues [2] used *Pure* data to generate artificially *Contaminated* EEG signals using the equation below:

$$Contaminated_{i,j} = Pure_{i,j} + a_j(VEOG) + b_j(HEOG) \quad (1)$$

where i and j subscripts refer to the subject number and electrode on the scalp respectively. *Contaminated* and *Pure* are 2D arrays with a 19 electrode by 6000 (200 Hz × 30 s) shape; *VEOG* and *HEOG* are two 1D vectors with 6000 samples; and contamination coefficients *a* and *b* are separate scalar values for each electrode. Ultimately, by using the simulated *Contaminated* data as the model input, and with *Pure* as the ground truth, we can objectively compare how well each of the artifact removal algorithms work [2]. Given that the Klados dataset was collected in a controlled experiment setting, it may not fully reflect noisy eyeblink signals collected in a

naturalistic environment. However, the following procedure that we describe reflects a needed first step to show feasibility for understanding CNN capability to estimate EEG signal noise.

2.2. EEGLAB - ICA Decomposition

Our baseline for machine learning comparison is EEGLAB's ICA decomposition tool, which calculates a mixing matrix that maximizes component independence by computing weights until the rate of change drops below a user-specified threshold [1][4]. This allows for the isolation and removal of noise signal, which in our case is the VEOG and HEOG noise reference signals. As described in the Introduction, it is possible to perform ICA-based blink artifact removal without an EOG reference by filtering with the electrode signal closest to the eyes (left and right frontal pole electrodes, Fp1 and Fp2, located on the forehead above the eyes) [4]. For this approach, however, further study is needed. In the present work, our reported results reflect using EOG noise references with ICA.

2.3. U-EEG

Our architecture was derived from the U-Net model, a specialized CNN pipeline designed for biomedical image segmentation [5]. We decided to use a U-Net implementation due to its ability to generate an accurate prediction using a relatively small dataset, and maintain an output shape equal to its input, allowing us to simply subtract the predicted artifact from the *Contaminated* input and obtain a blink-free EEG matrix.

Figure 1 displays a model diagram for U-EEG using the Klados dataset [2]. The model input uses a 3-dimensional shape of *electrode* X *sample* X *channel*. Here, *sample* refers to the product of the duration and the sampling frequency of the EEG signal, and *channel* refers to the depth of the feature-map. The U-Net configuration calculates weights and pools (i.e., down-samples) data from the electrode and sample dimensions by increasing depth to the *channel* dimension, which is consistent with a standard CNN. However, once the data reaches the mid-block *M* (Block 5 in Figure 1), the output is then up-sampled to match the corresponding down-sampling step *M*-1 (Block 4), and both outputs are merged. The outputs are now stacked across the channel dimension and can be fed into the next input block *M*+1 (Block 6). To summarize, this pattern of data flow repeats as follows:

$$B_{M+N} = Conv\left(Concatenate\left(B_{M-N}, B_{M+(N-1)}\right)\right) \quad (2)$$

where *B* is the block, and *N* ranges from 1 to *M-1*. This process essentially reincorporates past feature information to generate a more accurate output. Additionally, we applied zero-padding to ensure down-sampling did not distort data at the edges of the EEG epochs or result in a fractional dimension.

Table 1. U-EEG input and output dimensions. Electrodes (E); Input Samples (S); Channel Multiplier (C); Electrode Padding (Pe), Sample Padding (Ps)

Block No.	Input Dimensions	Layers	Output Dimensions
0	E, S, None	ExpandDims ZeroPadding	(E+Pe), (S+Ps), 1
1	(E+Pe), (S+Ps), 1	Conv MaxPooling	(E+Pe), (S+Ps)/2, C
2	(E+Pe), (S+Ps)/2, C	Conv MaxPooling	(E+Pe), (S+Ps)/4, C*2
3	(E+Pe), (S+Ps)/4, C*2	SeparableConv Dropout AveragePooling	(E+Pe)/3, (S+Ps)/8, C*4
4	(E+Pe)/3, (S+Ps)/8, C*4	Conv Dropout AveragePooling	(E+Pe)/3, (S+Ps)/16, C*8
5	(E+Pe)/3, (S+Ps)/16, C*8	Conv SpatialDropout	(E+Pe)/3, (S+Ps)/16, C*16
6	(E+Pe)/3, (S+Ps)/16, C*16	Unsample Concatenate Conv	(E+Pe)/3, (S+Ps)/8, C*8
7	(E+Pe)/3, (S+Ps)/8, C*8	Upsample Concatenate SeparableConv	(E+Pe), (S+Ps)/4, C*4
8	(E+Pe), (S+Ps)/4, C*4	Upsample Concatenate Conv	(E+Pe), (S+Ps)/2, C*2
9	(E+Pe), (S+Ps)/2, C*2	Upsample Concatenate Conv	(E+Pe), (S+Ps), C
10	(E+Pe), (S+Ps)/2, C	Conv Cropping	E, S, None

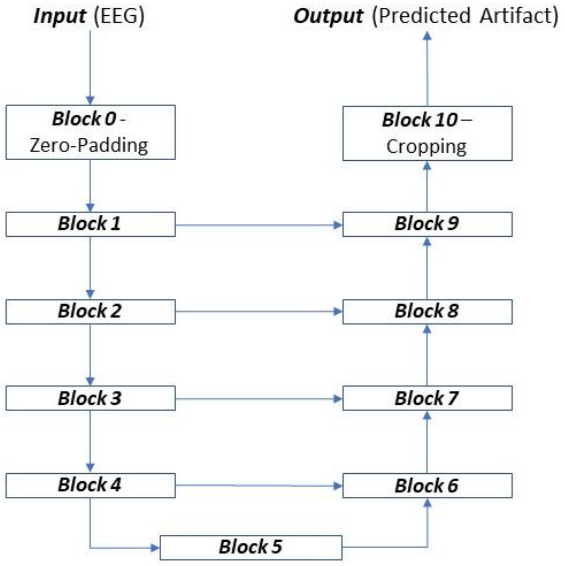

Figure 1. Diagram of the U-EEG architecture

Many machine learning classifier algorithms, such as EEGNet [6], use a SoftMax layer to compress an output to a fixed range. This procedure could be used to detect blinks, but is insufficient for our interest in estimating the full EOG waveform mask at each electrode. In contrast, the U-Net configuration is useful for our task as it ultimately allows us to easily up-sample back to the original EEG shape and subtract the EOG mask from the *Contaminated* EEG signal. This action makes the network more comparable to a filter than a classifier.

We derived our model from an existing open-source Keras [7] based U-Net [8] that was designed to reproduce the original biomedical image segmentation network [5]. Previous work by other investigators [9] similarly utilized a U-Net for artifact removal, but modified the input data to fit into with the original U-Net layers as if the inputs were images. We instead modified the layers to accept unmodified EEG. Table 1 shows the CNN layers used in each block, and the resulting data input and output dimensions.

We wanted to test architectures that include both blink-sensitive and general-use convolutional blocks, such that datasets used to classify other artifacts could utilize the general-use kernels. We used a specific kernel shape to cover the largest possible amplitude of a blink (2 s). We placed this filter at one of the model's lower blocks, since pooling would truncate the data and allow us to cover larger durations with a smaller kernel, saving computation time. We chose another block to cover fewer temporal samples, but more electrode information to increase spatial resolution. We expected that electrodes further from the eyes would show a smaller but non-zero blink component, which is potentially more difficult to detect. We reasoned that increasing the size of the kernel to cover more electrodes could help the network recognize the spatial propagation of noise to more distal electrode locations.

We tested both these customized 'blink-sensitive' and simpler N x N filters, and generally found that they varied in training time, but with little to no noticeable improvement in accuracy. Ultimately, we chose to only use the more generic kernel shapes in our final model. This is not necessarily a downside, as it leads us to believe that U-EEG can be implemented to detect other artifacts. More testing is needed to verify this.

We used both the uncontaminated EEG signal (*Pure*) and EOG signal (*Contaminated* – *Pure*, at every channel) as our ground truth. We generally found improved accuracy when using the EOG matrix as our ground truth. We believe this is due to the EOG being a more robust and distinct signal compared to the uncontaminated EEG. Thus, our final model used the EOG signal as its ground truth. We derived the U-EEG predicted pure signal ($U\text{-}EEG_{Pure}$) by subtracting its predicted EOG artifact ($U\text{-}EEG_{EOG}$) from the original *Contaminated* signal. Note that ICA estimates the *Pure* signal (ICA_{Pure}), but the ICA estimated EOG (ICA_{EOG}) can be similarly derived. This can be simplified as follows:

$$U\text{-}EEG_{pure} = Contaminated - U\text{-}EEG_{EOG} \quad (3)$$

$$ICA_{EOG} = Contaminated - ICA_{Pure} \quad (4)$$

We tested different segment lengths by splitting data from each subject into multiple batches. Our final implementation used 15 s (3000 samples) of data per batch. The ICA filter was given the full length of data for each subject (28 to 32 seconds), because accuracy increases with longer segments. The filtered signals were then sliced into smaller segments equal to the shape of the U-EEG output matrix to compare accuracy against each batch. We used exponential linear unit (ELU) activation functions instead of rectified linear unit (RELU) functions to conserve negative voltages and linear activations at the first and last blocks. Blocks 3 through 5 utilized a dropout rate of 0.7 to limit overfitting. We used the mean square error loss function to calculate the model's weights.

We divided the data 80/10/10 corresponding to training, validation, and testing, respectively. Note that this means $U\text{-}EEG_{EOG}$, $U\text{-}EEG_{Pure}$, ICA_{EOG} *and* ICA_{Pure} represent the estimated EOG and filtered EEG output from the testing data only, or last 10% of data from the Klados dataset [2].

We utilized early stopping, so our model would continue training until the validation loss stopped improving over a user selected number of epochs. We found 20 epochs to be effective, since the dropout layers could produce an increased loss between epochs. A summary of all relevant hyperparameters and variables can be seen in Table 2.

III. Results

Our U-EEG neural net was able to detect and remove EOG noise from *Contaminated* EEG signals with higher

Table 2. U-EEG hyperparameters and variable values

Parameter	Value
E (Number of Input Electrodes)	19
Pe (Electrode Dim. Padding)	2
S (Number of Input Samples)	3000
Ps (Sample Dim. Padding)	8
C (Channel Multiplier)	6
Number of Training Batches	71
Number of Validation Batches	8
Number of Testing Batches	8
Optimizer	Adam
Loss Function	Mean Squared Error
Dropout Rate	0.7
Layer Padding Type	Same
Trainable Params	487,229

accuracy than EEGLAB ICA. We compare the error of both filtering methods (RMSE) by contrasting the estimated clean signal (*U-EEG*$_{Pure}$ and *ICA*$_{Pure}$) from the original eyes-closed ground truth (*Pure*) testing data (Table 3).

Using ICA, the RMSE across all data and channels was 5.044 µV (SEM ± 2.844 µV), and **5.793 µV** (SEM ± 1.769 µV) when using only the testing data. For U-EEG, we found an aggregate RMSE of **3.134 µV** (SEM ± 0.893 µV) for test accuracy reflecting the variation in error due to variation in weights learned.

Figure 2 and Figure 3 illustrate plots comparing the filtered U-EEG and ICA outputs from randomly selected subjects. Figure 2 shows a comparison of *U-EEG*$_{Pure}$ and *ICA*$_{Pure}$ on test subject 0, depicting randomly selected electrodes Fp2 and P8 within a randomly selected time window. The top subplot shows *Contaminated*, *Pure*, and the filtered outputs. The remaining subplots ignore Contaminated to highlight the precision of each filtering method. Figure 3 shows the true and estimated artifacts (EOG, *U-EEG*$_{EOG,}$ *ICA*$_{EOG}$) on randomly selected electrodes (Fp1, F4, and C4) for test subject 4.

We found that U-EEG is consistently more accurate than ICA, especially at electrodes where the blink component is highest in amplitude (Table 3). These results show that U-EEG provides a more accurate method for EEG noise removal relative to ICA-derived procedures. Further, the U-EEG procedure avoids the costly manual selection steps required for ICA, which is important for concerns related to reliability and reproducibility of neuroscience results.

IV. Summary

Despite these promising results, we imagine a few areas of improvement. Firstly, U-EEG was at a disadvantage due to its relatively small dataset. The fact that our U-Net performed better than ICA while only having access to limited data illustrates its value and the need to explore performance with larger datasets. We believe that it is either the case that: (1) U-Nets perform exceptionally well compared to other deep learning models specifically when data is limited (either low-resolution or has a small number of batches), but has a theoretical 'accuracy-cap' and will not improve by much with more data, or (2) U-Nets will continue to improve when utilizing expanded datasets.

Ocular artifacts are one example of the panoply of noise sources. The ability to separate other artifacts that are common to EEG (e.g., muscle-related noise from walking), and in particular, mobile EEG related to real-life situations, will bring us closer to real-time BCI capability. In the near future, we plan to test U-EEG in mobile environments, and to use U-EEG in series with an

Table 3. Artifact removal precision (RMSE, µV) on testing data

Electrode	*ICA*$_{Pure}$	*U-EEG*$_{Pure}$
Fp1	13.817	7.103
Fp2	13.845	6.990
F7	3.823	3.277
F3	3.895	3.167
Fz	1.860	1.561
F4	1.971	1.563
F8	1.256	1.148
T7	1.593	1.071
C3	0.977	0.749
Cz	1.072	0.906
C4	8.274	4.319
T8	8.413	4.233
P7	3.255	1.683
P3	3.299	1.663
Pz	1.178	1.063
P4	1.396	1.004
P8	6.118	3.207
O1	2.831	1.481
O2	1.478	1.028
Mean	***5.793*** ± 1.769	***3.134*** ± 0.893

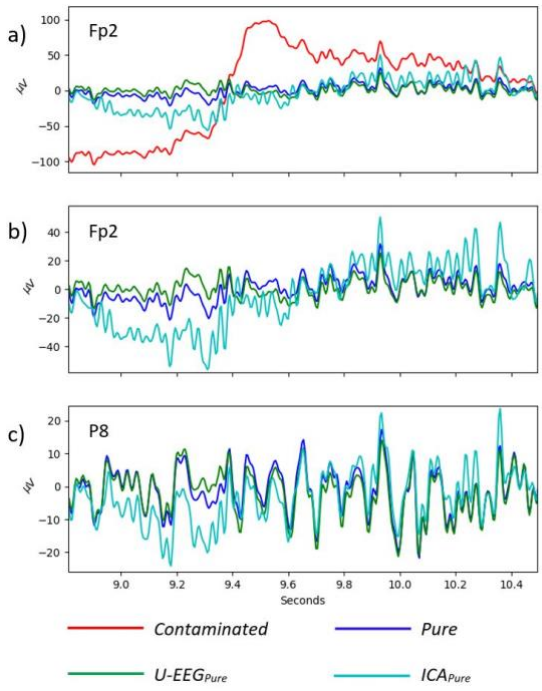

Figure 2. Comparison of U-EEG and ICA filtered EEG signal on test subject 0, depicting randomly selected electrodes Fp2 and P8 within a randomly selected time window (9-10.5 s). Subplot (a) shows *Contaminated*, *Pure*, and the filtered outputs at Fp2. Traces from Fp2 (b) and P8 (c) show the noise-free signal (*Pure*) and estimates *U-EEG*$_{Pure}$ and *ICA*$_{Pure}$. For display purposes, the y-axis scale showing EEG amplitude (µV) is not constant across subplots.

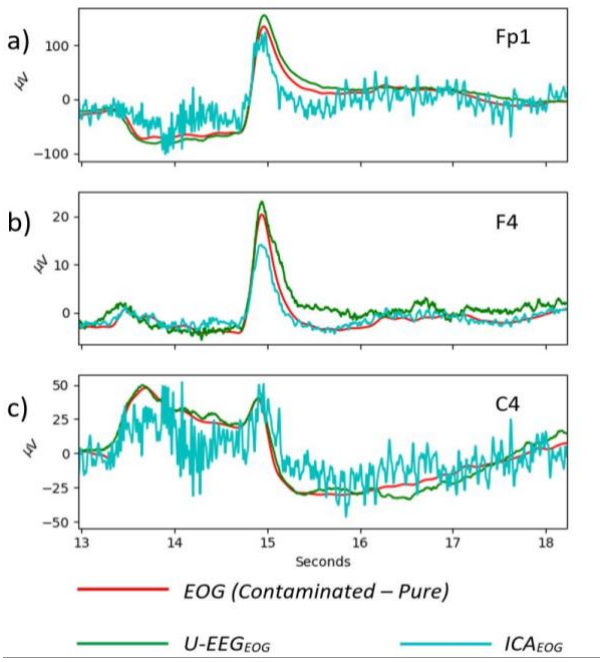

Figure 3. Comparison of U-EEG and ICA EEG artifact detection on test subject 4, depicting randomly selected electrodes Fp1, F4, and C4 within a randomly selected time window (13-18 s). Traces from Fp1, F4, and C4 (a-c) show the EOG artifact (*Contaminated – Pure*) and the EOG estimates *U-EEG$_{EOG}$* and *ICA$_{EOG}$*. For display purposes, the y-axis scale showing EEG amplitude (μV) is not constant across subplots.

EEG categorizer to confirm if artifact removal improves classification performance.

REFERENCES

[1] S. Makeig, A. J. Bell, T. Jung, and T. J. Seijnowski, "Independent component analysis of electroencephalographic data," *In Proceedings of the 8th International Conference on Neural Information Processing Systems (NIPS'95). MIT Press, Cambridge, MA, USA, 1995.* pp. 145-151.

[2] M. A. Klados and P. D. Bamidis, "A semi-simulated EEG/EOG dataset for the comparison of EOG artifact rejection techniques," *Data in brief*, June 29, 2016. vol. 8, pp. 1004–1006. *https://doi.org/10.1016/j.dib.2016.06.032*.

[3] G. H. Klem, H. O. Lüders, H. H. Jasper, and C. Elger, "The ten-twenty electrode system of the International Federation. The International Federation of Clinical Neurophysiology," *Electroencephalography and clinical neurophysiology. Supplement.* vol. 52, pp. 3-6. 1999. [Online], Available: *https://pubmed.ncbi.nlm.nih.gov/10590970/*.

[4] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *J Neurosci Methods.* vol. 134(1), pp. 9-21, 2004. *https://doi.org/10.1016/j.jneumeth.2003.10.009*.

[5] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Separation," *In Proceedings of the 18th International Conference on Medical Image Computing and Computer Assisted Intervention, MICCAI 2015, Munich, Germany, October 2015. https://arxiv.org/abs/1505.04597*.

[6] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: A Compact Convolutional Network for EEG-based Brain-Computer Interfaces," *J of Neural Eng*, 2018. *https://arxiv.org/abs/1611.08024*.

[7] F. Chollet *et al*., Keras, GitHub, 2015. [Online], Available: *https://github.com/fchollet/keras*.

[8] Zhixuhao, *Implementation of deep learning framework – Unet, using Keras*, GitHub. [Online], Available: *https://github.com/zhixuhao/unet* [Accessed: December 2019].

[9] N. Mashhadi, A. Z. Khuzani, M. Heidari, and D. Khaledyan, "Deep learning denoising for EOG artifacts removal from EEG signals," 2020. *https://arxiv.org/abs/2009.08809*.

[10] J. C. Bradford, J. R. Lukos, A. Passaro, A. Ries, and D. P. Ferris, "Effect of locomotor demands on cognitive processing," *Sci Rep* 9, 9234. 2019. *https://doi.org/10.1038/s41598-019-45396-5*.