Avoiding Post-processing with Event-Based Detection in Biomedical Signals - Supplementary Material

Nick Seeuws, Maarten De Vos, and Alexander Bertrand

I. Simulated data generation

Background signals are 20 s segments taken from the Computing in Cardiology 2017 Challenge signals [1] (we only retain the signals that are labeled as "clean" or "atrial fibrillation"). For training and validation data, the Challenge training set is used and randomly split 80-20 into training and validation. The Challenge validation set is used for our test set. The segments are drawn from the original data set with a 5 s stride.

For every background segment, target events are added in randomly with 20% chance (to reflect a natural sparsity of target events common in biomedical signal processing tasks), taken from the Physionet MIT-BIH Noise Stress Test Database [2]. This process happens once for every background segment at the start of training, and the resulting signals and targets are used as finite data sets for experiments. In the cases where targets are added, either one or two segments (chosen randomly) are taken from the noise database of random duration, uniformly distributed between 1 s and 6.7 s. The target events are added to the background signal with a signal-to-noise ratio (treating the target artefact events as noise) uniformly random between -6 and 6. Target events for the training and test set are not taken from the same noise recording. To add more of a challenge to this simulation, additional artefact events are added to the background signal (in a similar fashion) with shorter duration, ranging between 0.5 s and 1 s.

II. Network and training details

A. Simulated events

Simulated events are detected based on a single-channel input. The backbone's base "building block" is constructed using a 1D (temporal) convolution, a batch-normalization layer, and an ELU activation function. Every down- and upsample step is taken with a factor 4. The full backbone architecture is shown in Table I. Two convolution layers with kernel size 7 are used in the event-based model on the output at stage 2' to produce the center and duration signals. A single convolution layer with kernel size 7 is used by the epoch-based models to produce their outputs. The networks are trained on input signal segments of 20 s (as explained above). The event-based loss function is described in the main text. The epoch-based loss function is binary cross-entropy. Note that the epoch-based approach also generates an output signal at 1/16th the original sampling frequency (and not at the full original frequency as would be common in traditional segmentation approaches). We also tested an epoch-based approach at full resolution, and observed no significant difference in performance between the two output resolutions. For the sake of simplicity (and to save computational resources), we make the two (backbone) architectures (epoch-based and event-based) identical.

TABLE I: Simulated events backbone. The different stages are connected in a U-Net-like way, stage 4 for example concatenates the features of stage 4 and upsampled features of stage 5. Every stage besides 0 contains a batch-normalization layer between the convolution and nonlinearity. Stages $()'$ have two convolution-normalization-nonlinearity blocks. The *stride factor* indicates the total "downsample factor" for the given stage. This backbone produces a features signal at 1/16th the original sampling frequency.

Stage	#Filters	Kernel size	Stride factor
0	32	20	1
1	64	20	4
2	64	15	16
3	64	15	64
4	64	10	256
5'	64	5	1024
4'	64	10	256
3'	64	15	64
2'	64	15	16

B. EEG artefacts

Artefact events are detected based on a single-channel input. The backbone's base "building block" is constructed using a separable 1D convolution, a batch-normalization layer, and an ELU activation function. Every down- or upsample step is taken with a factor 4. At the original input resolution (before the first downsampling step), the networks are using one such

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 802895) and from the Flemish Government (AI Research Program).

N. Seeuws , A. Bertrand and M. De Vos are with the Dept. of Electrical Engineering (ESAT), Stadius Center for Dynamical Systems, Signal Processing and Data Analytics (STADIUS), KU Leuven, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium

M. De Vos is also with the Dept. of Development and Regeneration, Faculty of Medicine, KU Leuven

A. Bertrand and N. Seeuws are affiliated to Leuven.AI - KU Leuven institute for AI, B-3000, Leuven, Belgium.

TABLE II: Artefact detection backbone. The different stages are connected in a U-Net-like way, stage 5 for example concatenates the features of stage 5 and upsampled features of stage 6. Every stage besides 0 contains two separable convolutions with the given hyper-parameters. Stages $()'$ have a dropout layer between the two convolutions. The *stride factor* indicates the total "downsample factor" for the given stage. This backbone produces a features signal at 1/16th the original sampling frequency.

Stage	#Filters	Kernel size	Stride factor
0	32	20	1
1	64	20	4
2	64	15	16
3	64	15	64
4	64	10	256
5	64	5	1024
6	64	5	4096
5'	64	5	1024
4'	64	10	256
3'	64	15	64
2'	64	15	16

block, at every other stage using two such blocks (two at the "downward" path, and two at the "upward" path). The full backbone architecture is shown in Table II. Two "size 1" convolutions are used in the event-based model on the output at stage 2' to produce the center and duration signals. A single "size 1" convolution is used by the epoch-based model on the output at stage 2' to produce its output. Both networks are trained on input signal segments of 200 s and tested using a full EEG channel recording. The event-based loss function is described in the main text. The epoch-based loss function is binary cross-entropy with label smoothing, similar to the U-Net of [3]. Similar to the experiment on simulated events the epoch-based approach produces its output signal at 1/16th the original sampling frequency. For the EEG artefacts, we also did not observe a significant difference with an output at this resolution and an epoch-based output at full resolution.

For epoch-based post-processing, a median filter with a size corresponding to 0.1 s is used. The shortest event in the artefact data set lasts 0.2 s, so it is considered a reasonable filter size. The filter is used *after* thresholding the model at a certain operating threshold.

The Artefact Corpus [4] is randomly split 64-16-20 for training-validation-test sets, ensuring that the proportion of events is similar among the sets. Recordings of the same individual are not split among the sets.

C. EEG Seizures

The seizure detection backbone is constructed following the network of [3]. This backbone is made up of a channel-independent encoder, and channel information is only merged at the deepest part of the backbone and the skip connections. Attention Gating [5] is used in the skip connections to merge channel information. For the event-based model, the original epoch-based network is only used until stage "4", at 1/256th of the original sampling frequency. After this stage, the event-based framework's center and duration heads are appended. This is done to limit computational and memory footprint, since there is negligible performance impact compared to applying the heads at the original sampling frequency (as is done in the U-Nets of [3]). To clarify, the event-based approach generates an output (the center and duration signals) at 1/256th of the original sampling frequency, while the epoch-based benchmark for the EEG seizures (the approach of [3]) generates an output at the original sampling frequency. The full backbone architecture is shown in Table III. The epoch-based seizure detector is the original model of [3], with an output at the same resolution as the original input and using an ensemble of three networks (three "views"). Similar to artefact detection, the seizure detection networks are trained on input segments 200 s, and tested on full recordings.

TABLE III: Seizure detection backbone. Stage 4 and 4' are connected in a U-Net-like way, concatenating the features of stage 4 and upsampled features of stage 5". Every stage contains convolution layers with the given hyper-parameters. The $()'$ stages have two convolution-normalization-nonlinearity blocks with the given parameters. Stage 4' combines the features of stage 4 using Attention Pooling [5] before processing. Stage 5' has a dropout layer after its two convolution blocks and performs max-pooling over the different EEG channels before processing. All encoder stages are channel-independent (using the same convolution filter on every channel), the decoder stages merge channel-level information. The *stride factor* indicates the total "downsample factor" for the given stage. This backbone produces a features signal at 1/256th the original sampling frequency.

Stage	#Filters	Kernel size	Stride factor
0	16	15	1
1	32	15	4
2	64	15	16
3	64	7	64
4	128	3	256
5	128	3	1024
5'	64	3	1024
4'	64	5	256

The seizure data set is split into training, validation, and test sets. The training and validation sets are taken from the original training set of [6]. This original set is randomly split 80-20 for training and validation purposes, ensuring that both training and validation sets contain a similar proportion of seizure-containing recordings. Our test set is the original test set of [6].

III. Precision-recall curves for simulated events

Full precision-recall curves for the different approaches used for simulated events (evaluated with a 0.75 IoU threshold criterion) can be found in Figure 1. These curves are produced by sweeping over different confidence thresholds, i.e., sweeping the range [0, 1] on the models' output confidence. The event-based approach produces a traditional, intuitive precision-recall curve with high precision and low recall at one end of the curve, and low precision combined with high recall at the other end. The epoch-based approaches do not show this behavior and seem to have a single Pareto-optimal operating

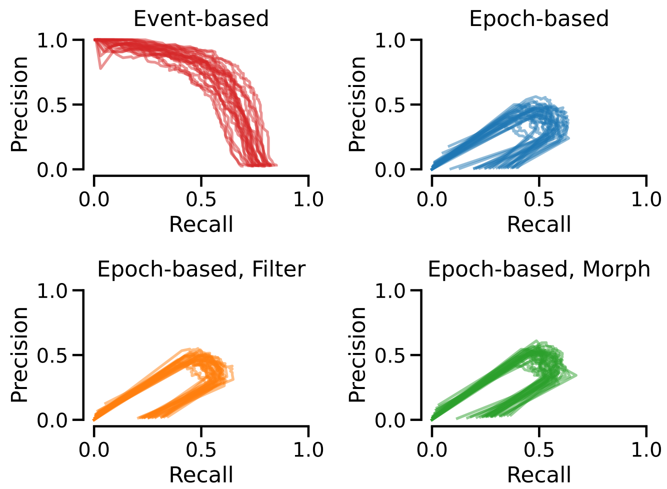

Fig. 1: Precision-recall curves for the detection of simulated events, evaluated with a 0.75 IoU threshold criterion. Every line is a separate run (data generation + training)

point (also observed for the real-world EEG artefacts in Figure 6 of the main text).

IV. Absolute errors for center and duration prediction

In the main text, we report *relative* errors for center and duration predictions because of the wide range in ground-truth durations. For completeness, we show the absolute center errors in Figure 2 and absolute duration errors in Figure 3.

V. Stable focal loss

The softplus function is a part of focal loss. Instabilities are expected when computing the gradient of the focal loss because of this softplus function. The softplus function can be found in the "cross-entropy factor" of the focal loss formulation (taking the logarithm of a sigmoid activation):

$$\log \sigma(l(x)) = \log e^l - \log(e^l + 1) = l - \text{softplus}(l)$$

and

$$\log(1 - \sigma(l(x))) = \log(1 - \frac{e^l}{e^l + 1})$$
$$= \log 1 - \log(e^l + 1) = -\text{softplus}(l)$$

The softplus function $\log(e^{(\cdot)} + 1)$ can be rewritten in a numerically more stable form that avoids a possibly exploding gradient as follows:

$$\text{softplus}_{stable}(x) = \log(e^{-|x|} + 1) + \max(x, 0)$$

References

[1] G. D. Clifford *et al.*, "Af classification from a short single lead ecg recording: The physionet/computing in cardiology challenge 2017," in *2017 Computing in Cardiology (CinC)*, IEEE, 2017, pp. 1–4.

[2] G. B. Moody *et al.*, "A noise stress test for arrhythmia detectors," *Computers in cardiology*, vol. 11, no. 3, pp. 381–384, 1984.

[3] C. Chatzichristos *et al.*, "Epileptic seizure detection in EEG via fusion of multi-view attention-gated U-Net deep neural networks," in *2020 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, 2020, pp. 1–7. DOI: 10.1109/SPMB50085.2020.9353630.

[4] A. Hamid *et al.*, "The Temple University artifact corpus: An annotated corpus of EEG artifacts," in *2020 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, IEEE, 2020, pp. 1–4.

[5] J. Schlemper *et al.*, "Attention gated networks: Learning to leverage salient regions in medical images," *Medical image analysis*, vol. 53, pp. 197–207, 2019.

[6] N. Neurotech and NeuroTechX. "Neureka^{TM} 2020 epilepsy challenge." (), [Online]. Available: https://neureka-challenge.com/ (visited on 06/12/2020).

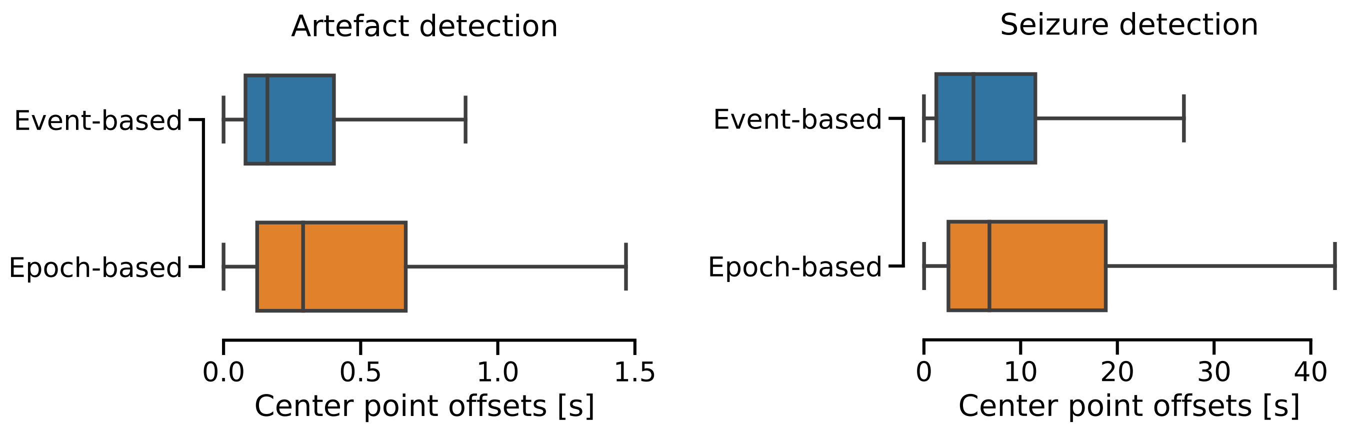

Fig. 2: Absolute values of the offsets between ground-truth and predicted centers. We consider all matched ground-truth and predicted events that show any overlap (green + orange class in Fig. 6 of the main text)

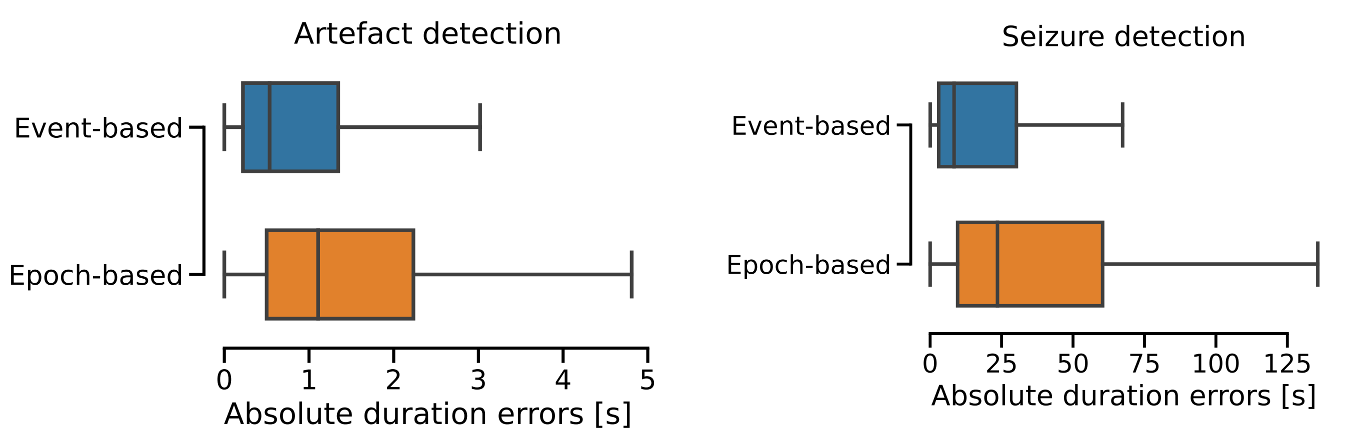

Fig. 3: Absolute values of the errors between ground-truth and predicted durations. We consider all matched ground-truth and predicted events that show any overlap (green + orange class in Fig. 6 of the main text)