Signal-Level Fusion With Convolutional Neural Networks for Capacitively Coupled ECG in the Car

Christoph Hoog Antink, Erik Breuer, Durmus Umutcan Uguz, Steffen Leonhardt

Medical Information Technology,
Helmholtz-Institute, RWTH Aachen University, Aachen, Germany

Abstract

Unobtrusive measurement technologies for vital signs, such as capacitively coupled electrocardiography (cECG), allow for health monitoring outside the clinical domain, for example in automotive applications. While cECG has the potential to deliver accurate information about the driver's heart, the signal quality is very volatile compared to standard ECG and can be corrupted easily by motion artifacts. In this work, we present a signal-level fusion algorithm based on a convolutional neural network (CNN) to locate individual heartbeats in three-channel cECG signals. To design and optimize the network's structure, data from the PhysioNet / CinC challenge "Robust Detection of Heart Beats in Multimodal Data" was used as an independent source. To train and test the algorithm, we used cECG data from six subjects and three different driving scenarios (highway, city, and proving ground) that is freely available as part of our UnoViS-database. Data consisted of 31 recordings with a total duration of 13.4 hours. Leave-one-subject-out cross validation was performed to assess the algorithm's performance. We achieved a sensitivity of 88.0% and a positive predictive value of 95.2% compared to the reference ECG, with a root-mean-square R-Peak localization error of 20.81 ms. The developed algorithm is available for download via the UnoViS-Website.

1. Introduction

Since the first introduction in 1969 [1], capacitive electrodes have been tested as replacement for wet electrodes in Electrocardiography (ECG). Their advantage of not requiring direct skin contact makes capacitively coupled ECG (cECG) suitable for unobtrusive measurements in several application scenarios such as beds, chairs, and car seats [2]. Unlike wet electrodes, the performance of capacitive electrodes does not degrade over time due to drying out. However, their coupling is much more susceptible to motion artifacts and even minute physiological movements of the body. Therefore, not just the signal quality but also the availability of a reliable signal depends on the coupling with the subject and measurement conditions like humidity and temperature. Clinical studies have shown that waveform-dependent diagnostics cannot be performed at present, although cECG can be used for the analysis of interval lengths [3]. cECG is a trade-off between unobtrusiveness and signal quality, where the latter might be compensated with fusion algorithms. For one, cECG might be fused with other unobtrusive modalities to increase overall coverage [4]. For another, recent cECG systems often contain multiple electrodes and can thus derive multiple "leads" [3, 5]. However, unlike standard multi-lead ECG system, the main purpose is not to deriving additional diagnostic information, but to increase redundancy and accommodate for variations in subject's height or position when integrated into, for example, chairs or mattresses.

In this work, we develop and test an algorithm for the detection of heartbeats in multi-channel cECG data based on the concept of convolutional neural networks (CNNs) [6]. In particular, the algorithm is designed to perform signal-level fusion, as this allows access to the largest amount of raw information [7]. If fusion is performed on higher levels (feature or decision), part of this information is usually lost, while the level of abstraction is generally increased.

"Robust Detection of Heart Beats in Multimodal Data" is a general problem in medical signal analysis and was thus topic and title of the PhysioNet / CinC Challenge 2014 [8, 9]. In the course of the challenge, several algorithms were developed and tested on a hidden dataset containing various multimodal clinical signals. As a follow-up to the original challenge, an algorithm for beat detection was developed that uses individual fully connected neural networks (FCNs), which are then fused using gradient boosted trees [10]. Similarly, the work presented in [11] uses a CNN for ECG annotation, while the majority of recent works that use CNNs are aimed at direct classification, for example of cardiac arrhythmias.

2.   Algorithm

The proposed algorithm architecture is shown in Figure 1. In essence, a neural network architecture is used in
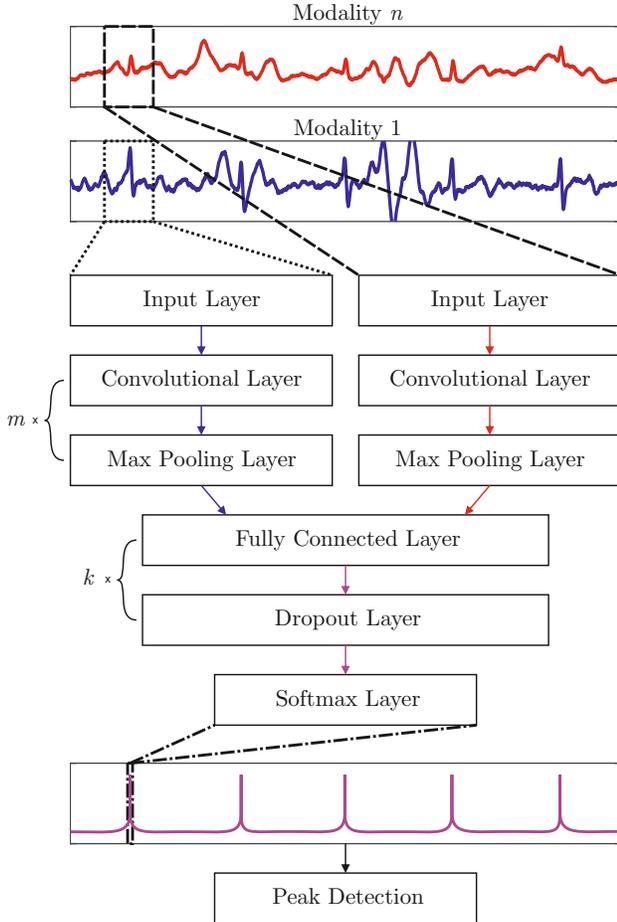

Figure 1. Schematic overview of the algorithm.

a moving window approach to preprocess a multichannel and potentially multimodal cardiac signal. For each of the n channels, we implemented a separate structure consisting of m convolutional layers and max pooling layers. The outputs of the last layers are then fused in a common structure consisting of k fully connected / dropout layers, which ultimately feed into a softmax layer. In the ideal case, the output of this layer is 1 if a heartbeat is located *exactly* in the center of the window and 0 otherwise. Since this will not be the case in a real application scenario, we applied a basic peak detection algorithm with adjustable threshold t and a minimum peak distance of 400 ms to the output signal obtained from the softmax layer.

To train the neural network, we created a balanced training set based on the input data and the reference annotation. For each annotated heartbeat, two windows of 101 samples each (404 ms) are taken from the signals. The "positive" window is centered on the annotation and assigned the desired output of 1. The "negative" window is shifted by a random offset of at least 20 samples (80 ms) and assigned the desired output of 0. In other words, negative output was defined for windows which have an uncentered peak or no peak at all.

3.   Data

To design the network architecture and tune the hyperparameters, the public training set from the PhysioNet / CinC Challenge 2014 was used as an independent source [8, 9]. This dataset consists of 200 recordings with a mean duration of 563.1 s. All recordings contain at least one ECG channel and one channel related to mechanical cardiac activity (arterial blood pressure, general blood pressure, etc.). In addition, recordings might contain other cardiac related signals (stroke volume), and signals that potentially contain cardiac components such as the electroencephalogram. For the purpose of this study, we only used the ECG channel and one blood pressure related signal. Recording sampling rate varies between 250 Hz and 360 Hz; in this study, all signals were downsampled to 200 Hz. For all recordings, a total of 151,032 manually verified heartbeat annotations are provided.

We evaluated the proposed architecture with data acquired using a cECG system integrated into the drivers seat of a car [5]. For reference, a conductive ECG using adhesive electrodes was used. In total six subjects participated in the study, which consisted of three different driving scenarios, namely driving on a highway, through city traffic, and on an automotive proving ground. In total, 13.4 hours worth of data is available and can be downloaded as part of our UnoViS database (dataset `UnoViS_auto2012`) [12]. The sampling rate for this dataset is 1 kHz and downsampling to 250 Hz was applied, as well as a second-order Butterworth bandpass filter with a passband of 8 to 20 Hz.

4.   Evaluation Method

For evaluation, we applied the same metric as was used in the PhysioNet / CinC Challenge 2014: A heartbeat was considered correctly located (true positive, TP), if it was located within a tolerance window of 150 ms of the annotation. If no beat lies within that window it is considered a false negative (FN). If a beat is detected outside the window or if multiple beats lie within the window, the number of false positives (FP) is increased accordingly. Based on these values, the metrics sensitivity (SE) and positive pre-

dictive value (PPV) are calculated,

$$SE = \frac{TP}{TP + FN}, \tag{1}$$

$$PPV = \frac{TP}{TP + FP}, \tag{2}$$

and the "Score" is defined as the arrhythmic mean of SE and PPV. In addition, all beats located within tolerance are evaluated in terms of the temporal root-mean-square error (RMSE) e_{RMS} from the exact reference locations.

Leave-one-out cross validation was used to assess the algorithms performance. For the challenge dataset this meant that for each iteration, 199 recordings were used for training, while the remaining recording was used for evaluation. In the case of the UnoViS-dataset, if one recording was used for evaluation, only recordings of *other* subjects were used for training to prevent overfitting.

5.   Results and Discussion

In the first step, we explored a variety of neural network architectures, see Table 1. In particular, architectures con-

		ECG		BP		Fusion	
m	k	Score	e_{RMS}	Score	e_{RMS}	Score	e_{RMS}
0	1	89.96	55.16	83.90	48.66	86.56	40.05
0	2	90.05	49.91	83.36	49.15	89.24	37.61
0	4	89.82	48.61	84.05	48.67	90.69	36.75
0	8	80.84	50.91	83.63	47.62	90.16	38.37
1	2	90.76	**30.37**	83.34	46.70	94.78	43.96
2	2	90.16	33.12	83.87	45.33	94.96	43.59
4	2	89.95	35.90	83.41	45.73	**95.09**	42.86
8	2	89.32	31.02	82.78	45.25	93.51	42.14

Table 1. Results of heart beat detection on the challenge dataset for ECG, BP, and fusion of both when using different neural network architectures. Note that $m = 0$ signifies that only FCN and no CNN layers were used.

taining only FCNs (i.e. $m = 0$) were compared to architectures using CNNs (i.e. $m > 0$). It can be argued that CNNs are better suited for the task of heartbeat localization in monomodal cardiac signals. In particular in terms of e_{RMS}, we achieved a reduction of approximately 40 % for the ECG by adding CNN layers compared to the FCN-only architecture. For the BP signal, the difference is less pronounced and results are inferior compared to the ECG. This confirms the observation shared by many participants of the original challenge that the ECG is the most reliable single source to locate heartbeats in the given dataset [9]. In terms of sensor fusion, the CNN architecture shows an increase in Score from approximately 90 % to 95 %, outperforming all monomodal approaches. However, a slight deterioration in terms of e_{RMS} (from about 38 ms for FCN to 43 ms) can be observed as well.

Based on the observations made on the challenge dataset, an architecture with 4 CNN layers was chosen. Three different scenarios were evaluated, namely processing each lead individually, fusing two leads, and fusing all three leads; Figure 2 shows the used architecture for the latter case. Here, n_N is the number of neurons, n_F is the number of filters, n_K is the kernel width, and n_P is the pooling size. For all neuronal layers, the rectified linear unit is used as activation function and the "adam" optimizer was used to minimize the categorical crossentropy with eight epochs of training. This algorithm can be downloaded from the UnoViS-website www.medit.hia.rwth-aachen.de/unovis/.

Input, n_N=101	Input, n_N=101	Input, n_N=101
CNN, n_F=10, n_K=20	CNN, n_F=10, n_K=20	CNN, n_F=10, n_K=20
Max Pooling, n_P=2	Max Pooling, n_P=2	Max Pooling, n_P=2
CNN, n_F=20, n_K=15	CNN, n_F=20, n_K=15	CNN, n_F=20, n_K=15
CNN, n_F=25, n_K=10	CNN, n_F=25, n_K=10	CNN, n_F=25, n_K=10
CNN, n_F=30, n_K=10	CNN, n_F=30, n_K=10	CNN, n_F=30, n_K=10

FCN, n_N=100, Dropout, 20%

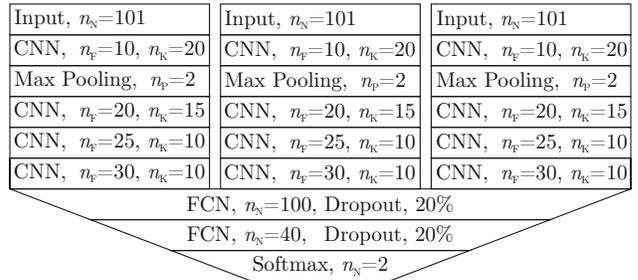FCN, n_N=40, Dropout, 20%

Softmax, n_N=2

Figure 2. Architecture of the neural network for fusing three channels of cECG.

Table 2 presents the results for all individual channels as well as for the fusion of two and all three channels. As

channel(s)	SE	PPV	Score	e_{RMS}
cECG 1	81.91	86.87	84.39	30.64
cECG 2	87.99	94.59	91.29	20.91
cECG 3	83.51	89.52	86.52	26.69
cECG 1-2	86.61	94.87	90.74	**20.74**
cECG 2-3	85.41	94.86	90.14	22.87
cECG 3-1	84.33	92.12	88.23	24.87
cECG 1-2-3	**88.02**	**95.17**	**91.60**	20.81

Table 2. Results of heart beat detection on individual cECG channel as well as fusion of two and three channels, $t = 0.7$.

observed in [5], notable differences between leads exist. While channel 1 shows the worst performance in terms of Score and e_{RMS}, channel 2 shows the best performance. If channels 1 and 2 are fused, the localization error is slightly improved to 20.74 ms. If all three channels are fused, a minimal gain in Score is achieved, resulting in the best value of 91.60. While a slight deterioration in e_{RMS} is observed as well, it still outperforms the individual channels and the combinations 2-3 and 3-1. Figure 3 shows the influence of the peak detection threshold t for the three-channel fusion scenario. As expected, an increase in t re-

duces both SE and e_{RMS}, while PPV is increased, resulting in an almost constant Score for $0.5 < t < 0.8$.
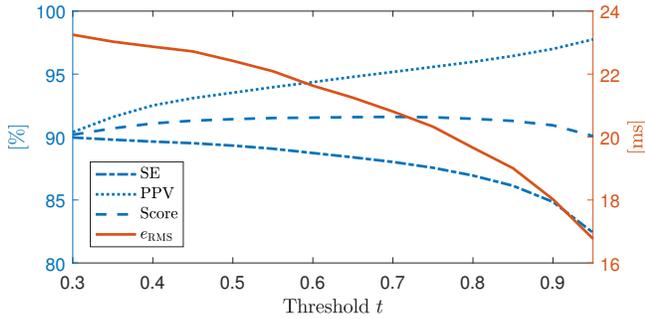

Figure 3. Influence of the threshold parameter t when fusing of all three cECG channels.

Both, the generation of the training set as well as the training of the network involve random components. We observed that, depending on the seed, results varied by about one percentage point and 1 ms respectively. Thus, ranking between the top scorers cECG 1, cECG 1-2, and cECG 1-2-3 is probably of little value. However, it is worth noting that fusion of all channels provides results that are at least as good as the best channel, thus avoiding so-called "catastrophic fusion". Moreover, since we trained permuting the input data, the algorithm is invariant to the channel ordering. This is important since the signal quality of multi-electrode cECG is known to be quite variable and the best channel might differ from subject to subject.

Finally, it is interesting to note that results of heart beat detection from a cECG obtained in a driving vehicle are at least in the same range as results from noisy yet conductive and partly clinical ECG as provided by the PhysioNet / CinC Challenge 2014.

6. Conclusion and Outlook

In this work, we presented a signal-level fusion algorithm based on a CNN to locate individual heartbeats in 13.4 hours worth of capacitively coupled ECG data obtained from the seat of a driving car. We achieved a sensitivity of 88.0% and a positive predictive value of 95.2% compared to a conductive reference ECG, with a root-mean-square R-Peak localization error of 20.81 ms. While the fused results are only marginally better than the results of the optimal lead, the algorithm does not require additional lead switching or preselection.

To increase reproducibility and to encourage further development, both the source code as well as the dataset are freely available as part of the UnoViS database `www.medit.hia.rwth-aachen.de/unovis/`.

References

[1] Lopez A, Richardson PC. Capacitive Electrocardiographic and Bioelectric Electrodes. IEEE Transactions on Biomedical Engineering 1969;BME-16(1):99–99.

[2] Sun Y, Yu XB. Capacitive Biopotential Measurement for Electrophysiological Signal Acquisition: A Review. IEEE Sensors Journal 2016;16(9):2832–2853.

[3] Leicht L, Skobel E, Knackstedt C, Mathissen M, Sitter A, Wartzek T, Mohler W, Reith S, Leonhardt S, Teichmann D. Capacitive ECG Monitoring in Cardiac Patients During Simulated Driving. IEEE Transactions on Biomedical Engineering 2018;9294(APRIL):1–1.

[4] Hoog Antink C, Schulz F, Leonhardt S, Walter M. Motion Artifact Quantification and Sensor Fusion for Unobtrusive Health Monitoring. Sensors 2017;18(1):38.

[5] Wartzek T, Eilebrecht B, Lem J, Lindner HJ, Leonhardt S, Walter M. ECG on the Road: Robust and Unobtrusive Estimation of Heart Rate. IEEE Transactions on Biomedical Engineering 2011;58(11):3112–20.

[6] Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification With Deep Convolutional Neural Networks. In Pereira F, Burges CJC, Bottou L, Weinberger KQ (eds.), Advances in Neural Information Processing Systems 25. Curran Associates, Inc., 2012; 1097–1105.

[7] Hall DL, Llinas J. An Introduction to Multisensor Data Fusion. Proceedings of the IEEE 1997;85(1):6–23.

[8] Moody GB, Moody B, Silva I. Robust Detection of Heart Beats in Multimodal Data: The PhysioNet/Computing in Cardiology Challenge 2014. Computing in Cardiology 2014;41:549–552.

[9] Silva I, Moody B, Behar J, Johnson A, Oster J, Clifford GD, Moody GB. Robust Detection of Heart Beats in Multimodal Data. Physiological Measurement 2015;36(8):1629–1644.

[10] Vernekar S, Vijayasenan D, Ranjan R. A Novel Approach for Robust Detection of Heart Beats in Multimodal Data Using Neural Networks and Boosted Trees. Computing in Cardiology 2016;43:2–5.

[11] Vollmer M, Sodmann P, Caanitz L, Nath N, Kaderali L. Can Supervised Learning be Used to Classify Cardiac Rhythms? Computing in Cardiology 2017;44:2–5.

[12] Wartzek T, Czaplik M, Hoog Antink C, Eilebrecht B, Walocha R, Leonhardt S. UnoViS: The MedIT Public Unobtrusive Vital Signs Database. Health Information Science and Systems 2015;3(2):1–9.

Address for correspondence:

Christoph Hoog Antink
Medical Information Technology
Helmholtz-Institute, RWTH Aachen University
Pauwelsstr. 20 / D-52074 Aachen / Germany
hoog.antink@hia.rwth-aachen.de