# Rotational Data Augmentation for Electroencephalographic Data

Mario Michael Krell[1] and Su Kyoung Kim[2]

***Abstract*— *Motivation:* For deep learning on image data, a common approach is to augment the training data by artificial new images, using techniques like moving windows, scaling, affine distortions, and elastic deformations. In contrast to image data, electroencephalographic (EEG) data suffers even more from the lack of sufficient training data. *Methods:* We suggest and evaluate rotational distortions similar to affine/rotational distortions of images to generate augmented data. *Results:* Our approach increases the performance of signal processing chains for EEG-based brain-computer interfaces when rotating only around y- and z-axis with an angle around $\pm 18$ degrees to generate new data. *Conclusion:* This shows that our processing efficient approach generates meaningful data and encourages to look for further new methods for EEG data augmentation.**

## I. INTRODUCTION

Brain-computer interfaces (BCIs) link a user and external systems by detecting a specific brain activity (e.g., electroencephalogram (EEG)), which is correlated with the users intent (e.g., attention, movement intention, etc.). In the last decade, a great progress has been achieved by using machine learning techniques in EEG-based BCI applications. In particular, a single-trial detection of event-related potentials (ERPs), which correlates cognitives processes, allows to deliver the users intent to external systems per event in real time. For example, embedded brain-reading, which uses single trials from EEG signals to infer human's intentions in real time, allows to adapt the human-machine (or computer) interaction [1].

Due to high sampling frequency (up to 5kHZ), EEG data is usually first decimated, followed by a frequency filtering or a frequency transform. For reducing the number of electrodes (up to 256), numerous spatial filters have been developed, e.g., [2]–[5]. Spatial filters linearly combine the data from several sensors to create a reduced set of pseudo sensors that condense the relevant information (as dimensionality reductions like principal component analysis also do but only on the spatial component of the data). However, these algorithms are no spatial transformations and do not consider the true position of the sensors.

EEG data processing chains are usually hand crafted and the optimization is usually very difficult for various reasons: a) EEG data is very noisy and non-stationary, b) there are large differences in data between different recording days with the same subject and different subjects, and c) real-world applications do not often allow to record a large number of labeled training samples for reasonable recording time. Especially in case of stroke rehabilitation, it is more difficult to record a sufficient amount of data, because the patient (subject) cannot perform a large number of events (movements) due to fatigue.

Several approaches have been applied to overcome these problems, e.g., ensemble learning, transfer learning, and online learning. In this paper, we propose a data augmentation approach for EEG data, which, to the best of our knowledge, has not yet been applied in EEG-based BCI applications. Our approach is to modify existing data to increase its amount and to support the learning algorithm in learning data invariances (data augmentation).

In contrast, for image data it is common to apply different distortions [6], scaling, or moving windows/pixel shifts to create additional data and make the data processing more robust/invariant to these transformations. This is especially important for deep learning, which requires a lot of data and to provide the network with examples of invariances which it is supposed to learn. Krizhevsky et al. generated new images at run time with translations and horizontal reflections without great processing effort and increased the number of images by a factor of 2048 [7]. Additionally, they randomly altered the intensities of the RGB channels in the images. For the first approach they report that "without this scheme, (their) network suffers from substantial overfitting, which would have forced (them) to use much smaller networks" and for the second approach they state that it "reduces the top-1 error rate by over 1%".[1] For EEG data processing, there are only a few publications on deep learning and all suffer from the low number of samples. Most of them focus only on the temporal but not the spatial aspect of the data, even though it is quite common to apply spatial filtering on EEG data [8]–[10]. Especially, no data augmentation has been used. In [11], a deep learning approach was applied which was motivated by common EEG data processing chains and which included temporal and spatial filtering but still suffered from an insufficient amount of data and not using augmentation techniques.

In this paper, we propose rotational data augmentation for EEG data, (Section II). We evaluate our approach by comparing different parametrizations and by analyzing the effect of data dimensionality on classification performance

This work was supported by the Federal Ministry of Education and Research (BMBF, grant no. 01IM14006A) and by a fellowship within the FITweltweit program of the German Academic Exchange Service (DAAD).

[1]Mario Michael Krell is with the International Computer Science Institute, University of California Berkeley, USA and the Robotics Group, University of Bremen, Bremen, Germany.

[2]Su Kyoung Kim is with the Robotics Innovation Center (RIC), German Research Center for Artificial Intelligence (DFKI) GmbH, Robert–Hooke–Str. 1, 28359, Bremen, Germany. su-kyoung.kim@dfki.de

[1] The training data consisted of 1.2 million images. The Alexnet has 60 million parameters, 650000 neurons, and 14 layers where 8 of them had weights.

(Section III). We provide a conclusion and outlook in Section IV.

## II. Rotational Data Augmentation

In this section, we give an overview of the structure of EEG data and introduce our approach. EEG data can be seen as two-, three-, or four-dimensional data.

The most important dimension is the temporal dimension because of its accuracy of sampling data with up to 5kHZ that is directly related to the current brain activity.

The other dimension corresponds to the spatial component of the data (different sensors/electrodes). This dimension is usually handled as a linear list with arbitrary sorting (1D). However, as regards content, it corresponds to the sensors on the head surface (2D) with positions in the 3D space. In most EEG-based BCI applications/processing cases, electrode positions and the underlying spatial relations are not considered. In fact, there are correlations in data between neighboring electrodes and hence it should be taken into consideration. There are, of course, approaches that consider the spatial relations between electrodes (e.g, source localization methods, connectivity methods, etc). However, such methods require a considerable amount of expert knowledge and computational power compared to the classical BCI applications.

For EEG data processing, spatial robustness is a major issue. When an EEG cap slightly shifts during experiments over time, it is not easy to find the original places of electrodes and to reset the current positions of electrodes to their original positions accordingly. Hence, there can be differences in electrode positions between a first and second recording session even during the same recording day (spatial shifts within session and between sessions). Furthermore, individually different head shapes of subjects can contribute to differences in electrode positions between subjects (spatial shifts between subjects).

Our approach to handle this issue is to generate artificial data associated with differently shifted electrode positions without much effort. This enriches the training data with examples of shifted caps and enables the resulting classifier to become robust against slight variations of electrode positions. For reasons of simplicity, the way to generate new data restricts to rotations around the three main axes of the head as displayed in Fig. 1.[2] Since electrode positions and head-shape are usually unknown, we use the standard positions, according to the extended 10-20 system. The new positions can be determined by standard rotation matrices for the respective axes. The data of the rotation is obtained by applying the interpolation based on radial basis functions (RBF) from SciPy (scipy.interpolate.Rbf) [13]. For this interpolation, it is crucial that the data is normalized beforehand such that electrodes show comparable ranges. For each time point, the current amplitudes are taken and a new interpolation is generated. This approach is evaluated with EEG data obtained from 5 subjects (details, Section III-A).

[2] This graphic was created with Brainstorm [12], which is documented and freely available for download online under the GNU general public license (`http://neuroimage.usc.edu/brainstorm`).

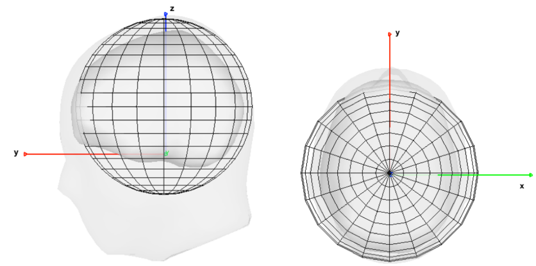

Fig. 1. Spherical Head Model – The x-axis points to the right, the y-axis to the front, and the z-axis runs through the vertex.

## III. Evaluation

In this section, the proposed approach is evaluated with the P300 data, which was generated by the oddball paradigm (Section III-A). We analyzed different parameters and properties of the rotational distortion. All data processing was performed with the open-source signal processing and classification environment pySPACE [14] on a high performance cluster.[3]

### A. Dataset, Preprocessing, Classification

We used data generated from an experimental scenario[4] as described in [1], which contains a P300 oddball paradigm. In the experimental scenario, the subject saw task-irrelevant event (standard) every second with a latency jitter of $\pm 100\,$ms. With a probability of $1/6$, task-relevant event (target) was displayed, which required a reaction from the subject (see Fig. 2). Based on this reaction to the targets, we can infer the true label for standards and targets. When the subject correctly responded to targets, we can ensure that targets were correctly perceived. The perceived task-relevant event leads to a specific pattern in the brain, called P300. In this scenario, the continuous EEG was recorded from 5 subjects. Two recording sessions were collected per subject on two different days. Each session consists of five runs and each run contains 720 standards and 120 targets.

We segmented the continuous EEG based on each event with a segment length of one second and normalized them to zero mean and a standard deviation of one. The sampling rate of the data was reduced from 1000Hz to 25Hz. Then the data was low-pass filtered with a cut-off frequency of 4Hz, which was chosen based on our previous evaluations.

After this preprocessing step, the data augmentation approaches were applied. Afterwards, the xDAWN spatial filter [3] was trained on the complete training data and then applied with 8 resulting channels. Afterwards, local straight lines were fitted for each channel and the respective slopes (but not the offsets) were concatenated as features (segment width of 400ms and step size of 120ms, see also [1]).

[3] The code is provided as part of this framework.

[4] Study was approved by the ethic committee of the Bremen University

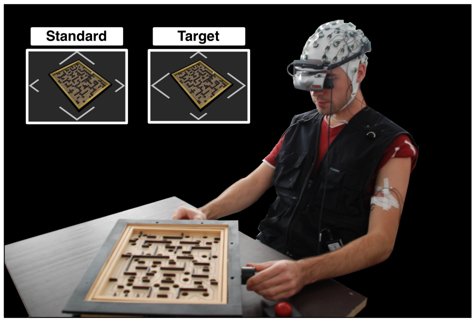


Fig. 2. Setup – EEG data is recorded while the subject plays a virtual Labyrinth game and reacts to the target stimuli by pressing the buzzer.

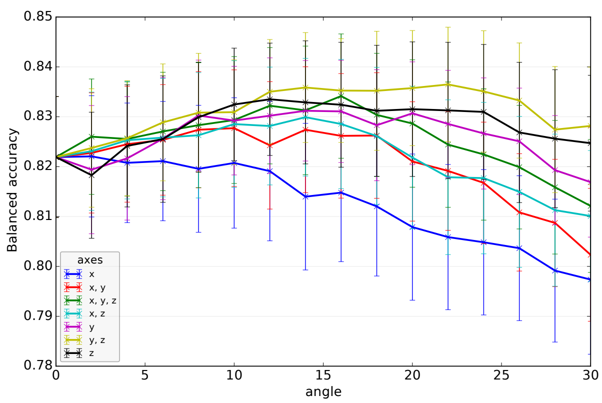


Fig. 3. Data augmentation with a rotation around the different axes: x, y, z and a rotation angle of *+angle* **and** *-angle* (combined). See also Fig. 1 for the meaning of the axes. The augmented data of the different rotation axes is combined for training. An angle of zero corresponds to the baseline with no data augmentation.

Features were normalized to zero-mean and unit-variance on the training data. The used classifier was a standard (affine) SVM implementation with a linear kernel [15] and limited number of iterations (100 times the number of samples). The regularization hyperparameter $C$ of the SVM was optimized using 5 fold cross validation with two repetitions and with the values $[10^0, 10^{-0.5}, \ldots, 10^{-4}]$.

For our evaluation, we train with the data of one recording session and then test on the other remaining recording session of the same subject. This results in 10 samples (5 subjects * 2 sessions). EEG recordings on two different days (2 sessions) lead to slight changes of the electrode positions.

To account for the unbalanced class ratio, we use the balanced accuracy as performance metric, which is the arithmetic mean of true positive and true negative rate [16].

### B. Rotation Axes

In this section, we analyzed the effect of the three rotation axes in the data augmentation. Here, we evaluated both single axis data and possible axes combinations. For the data augmentation, we took the original data and added an artificial sample for the chosen angle in positive and negative direction.

The results are depicted in Fig. 3. Note that using an angle of zero means that the data has not been augmented but kept as is. Using the x-axis reduced performance on average, whereas the augmentation around the y- and z-axis increased performance, in which the z-axis slightly outperformed the y-axis. The combination of y- and z-axis slightly increased classification performance, whereas adding data, augmented with a rotation around the x-axis, decreased classification performance in every combination. The best performance was achieved with an angle between 12° and 24° over all subjects.

We could also observe a variability between subjects in the performance. Such subject-specific performance should be considered, since we aim to configure the data augmentation as independent from data properties as possible. Here, a performance increase could be observed in 5 out of 10 cases. In the other cases, there is no change or a slight decrease. This absence of a substantial decrease is very important for the applicability of our approach.

### C. Data Reduction and Change of Data Dimensionality

In general, the dimensionality correlates with the cap configuration (i.e., cap with different numbers of electrodes). It is well known that machine learning algorithms behave differently depending on the ratio between dimension of the data after the preprocessing and number of provided samples for each class.

For the xDAWN in the processing chain, the use of a larger number of filters increases data dimension for the SVM classifier. For the case of using small rotation angles, the performance was reduced due to increased feature dimension (see, Fig. 4). For 16 filters, the dimension is doubled, and with 32 it is quadrupled. The same effect can be observed when reducing the data size (see Fig. 5). Especially when using one fifth of the data, the performance drops drastically. In contrast to the dimensionality increase, performance decreases with less data which is common and expected. Interestingly the large performance drop only occurs for small rotation angles and it is not relevant for larger angles over 10°.

This result is very positive for our augmentation strategy because it is still applicable when there is a lack of data. The reason for the performance drop for small angles is probably that the data augmentation is modeled by the classifier as noise and degrades the classifier model whereas larger angles are modeled like new data.

Further, we evaluated whether the data augmentation/interpolation is possible for smaller electrode constellations (32 and 19 electrodes according to the 10-20 system). This evaluation shows that the data augmentation did not reduce the performance for smaller electrode constellation, but there was also no relevant improvement between angles (see Fig. 6). This indicates that the interpolation is not good enough if the electrodes are positioned too sparse.

## IV. Conclusion

In this paper, we proposed the rotational data augmentation techniques for EEG data to generate new data without

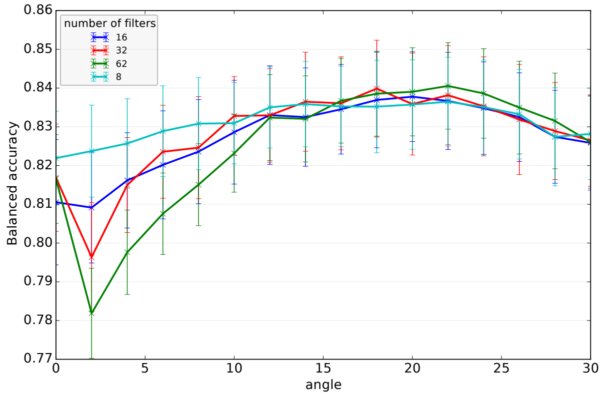


Fig. 4. Comparison of different numbers of spatial filters.

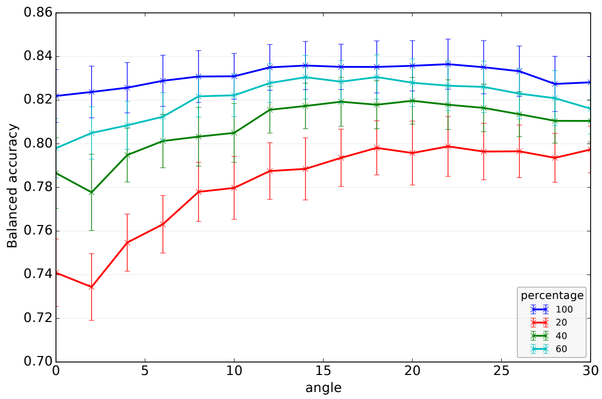


Fig. 5. Comparison of different percentages of used training data.

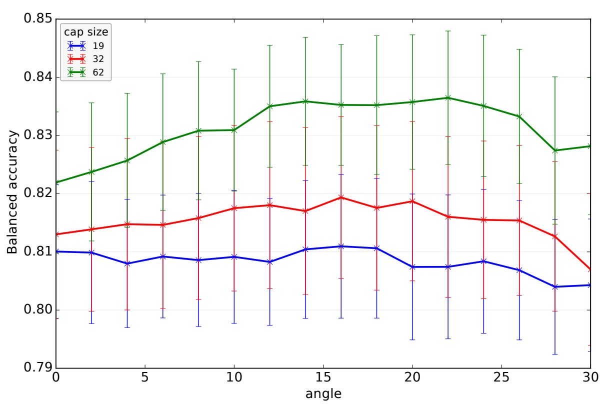


Fig. 6. Comparison of different cap configurations containing a different number of electrodes.

reducing the performance. We analyzed and compared the behavior of the used algorithm on real EEG data. Our analyses show that our proposed novel approach results in an increase of classification performance with a general setting of rotating only around y- and z-axis with an angle around 18 degrees. For the future, we want to analyze statistical significance, further augmentation strategies, and use our findings for deep learning. Furthermore, it would be interesting to investigate whether complex head models or aggregation of augmented results to one decision can further improve the performance.

## References


[1] E. A. Kirchner, S. K. Kim, S. Straube, A. Seeland, H. Wöhrle, M. M. Krell, M. Tabie, and M. Fahle, "On the applicability of brain reading for predictive human-machine interfaces in robotics," *PloS ONE*, vol. 8, no. 12, p. e81732, jan 2013.

[2] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Müller, "Optimizing Spatial filters for Robust EEG Single-Trial Analysis," *IEEE Signal Processing Magazine*, vol. 25, no. 1, pp. 41–56, 2008.

[3] B. Rivet, A. Souloumiac, V. Attina, and G. Gibert, "xDAWN Algorithm to Enhance Evoked Potentials: Application to Brain-Computer Interface," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 8, pp. 2035–2043, aug 2009.

[4] H. Woehrle, M. M. Krell, S. Straube, S. K. Kim, E. A. Kirchner, and F. Kirchner, "An Adaptive Spatial Filter for User-Independent Single Trial Detection of Event-Related Potentials," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 7, pp. 1696–1705, jul 2015.

[5] M. M. Krell, H. Wöhrle, and A. Seeland, "raxDAWN: Circumventing Overfitting of the Adaptive xDAWN," in *Proceedings of the 3rd International Congress on Neurotechnology, Electronics and Informatics*. SciTePress, 2015, pp. 68–75.

[6] P. Simard, D. Steinkraus, and J. Platt, "Best practices for convolutional neural networks applied to visual document analysis," in *Seventh International Conference on Document Analysis and Recognition, 2003*, vol. 1. IEEE Comput. Soc, aug 2003, pp. 958–963.

[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105.

[8] D. F. Wulsin, J. R. Gupta, R. Mani, J. A. Blanco, and B. Litt, "Modeling electroencephalography waveforms with semi-supervised deep belief nets: fast classification and anomaly measurement." *Journal of neural engineering*, vol. 8, no. 3, p. 036015, jun 2011.

[9] K. Li, X. Li, Y. Zhang, and A. Zhang, "Affective state recognition from EEG with deep belief networks," in *2013 IEEE International Conference on Bioinformatics and Biomedicine*. IEEE, dec 2013, pp. 305–310.

[10] X. An, D. Kuang, X. Guo, Y. Zhao, and L. He, "A Deep Learning Method for Classification of EEG Data Based on Motor Imagery," in *Intelligent Computing in Bioinformatics: 10th International Conference, ICIC 2014, Taiyuan, China, August 3-6*, D.-S. Huang, K. Han, and M. Gromiha, Eds. Cham: Springer International Publishing, 2014, pp. 203–210.

[11] H. Cecotti and A. Graser, "Convolutional Neural Networks for P300 Detection with Application to Brain-Computer Interfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 3, pp. 433–445, mar 2011.

[12] F. Tadel, S. Baillet, J. C. Mosher, D. Pantazis, and R. M. Leahy, "Brainstorm: A User-Friendly Application for MEG/EEG Analysis," *Computational Intelligence and Neuroscience*, vol. 2011, pp. 1–13, 2011.

[13] E. Jones, T. Oliphant, P. Peterson, and Others, "SciPy: Open source scientific tools for Python," 2001.

[14] M. M. Krell, S. Straube, A. Seeland, H. Wöhrle, J. Teiwes, J. H. Metzen, E. A. Kirchner, and F. Kirchner, "pySPACE a signal processing and classification environment in Python," *Frontiers in Neuroinformatics*, vol. 7, no. 40, pp. 1–11, dec 2013.

[15] C.-J. Hsieh, K.-W. Chang, C.-J. Lin, S. S. Keerthi, and S. Sundararajan, "A dual coordinate descent method for large-scale linear SVM," in *Proceedings of the 25th International Conference on Machine learning (ICML 2008)*. ACM Press, jul 2008, pp. 408–415.

[16] S. Straube and M. M. Krell, "How to evaluate an agent's behaviour to infrequent events? – Reliable performance estimation insensitive to class distribution," *Frontiers in Computational Neuroscience*, vol. 8, no. 43, pp. 1–6, jan 2014.