An EEG Image-search Dataset: A First-of-its-kind in IR/IIR

NAILS: Neurally Augmented Image Labelling Strategies

Graham Healy, Cathal Gurrin, Alan F. Smeaton
Insight Centre for Data Analytics, Dublin City University
Glasnevin, Dublin, Ireland
Graham.Healy,cathal.gurrin,alan.smeaton@dcu.ie

Zhengwei Wang, Tomas Ward
Department of Electronic Engineering, Maynooth University
Co Kildare, Ireland
wang.zhengwei.2014@mumail.ie,tomas.ward@nuim.ie

Abstract

In this work we emphasize the need for and we describe a first-of-its-kind RSVP (Rapid Serial Visual Presentation) - EEG (Electroencephalography) dataset to be released as part of the NTCIR-13 NAILS (Neurally Augmented Image Labelling Strategies) task at the NTCIR-13 participation conference. The dataset is used to support a collaborative evaluation task in which participating researchers benchmark machine-learning strategies against each other. The experimental protocol used to capture the dataset is designed to encompass a broad range of image search activities and coincident neural signals. Here, we outline the experimental protocol used to capture the dataset alongside discussing the motivation behind its construction.

CCS Concepts •**Computing methodologies** → *Machine learning;* •**Information systems** → *Image search;* •**Hardware** → *Emerging interfaces;*

Keywords Brain-computer Interface, Information Retrieval, Signal Processing, Machine Learning

ACM Reference format:
Graham Healy, Cathal Gurrin, Alan F. Smeaton and Zhengwei Wang, Tomas Ward. 2017. An EEG Image-search Dataset: A First-of-its-kind in IR/IIR. In *Proceedings of CHIIR Workshop on Challenges in Bringing Neuroscience to Research in Human-Information Interaction, Oslo, Norway, March 2017 (NeuroIIR 2017),* 2 pages.
DOI: n/a

1 Introduction

EEG (Electroencephalography) has recently become an accessible method for researchers and users to build and operate BCI (Brain-Computer Interface) applications, primarily because of the availability of a new generation of low-cost devices. While the initial use of such techniques began in clinical/rehabilitative settings for the purposes of augmenting communication and control, a recent trend has been to use such signals and methods in new domains, such as image annotation, which relies on the identification of target brain events to trigger labeling (Marathe et al. 2015), (Pohlmeyer et al. 2011), (Healy and Smeaton 2011). This trend is particularly relevant to the multimedia IR (Information Retrieval) and HII (Human-information Interaction) communities as in recent years, EEG has become a promising technology for several applications. These include annotating multimedia content, identifying when a user's attention is drawn to something in the real world, or even as a source of wearable sensor data to be indexed for later retrieval or analysis.

NTCIR (NII Testbeds and Community for Information access Research) is a conference (18-month schedule) that brings together researchers to develop evaluation methodologies and performance measures for IA (Information Access) technologies. This results in an active research community in which findings based on comparable experimental results are shared and exchanged in an open manner. One topical focus of this is mining knowledge from a large amount of human generated data. NAILS is an affiliated workshop to support the collaborative evaluation of best-practice strategies for RSVP-EEG image search applications, where researchers benchmark their machine-learning strategies. In this work, we outline the experimental protocol used to capture the dataset for this workshop alongside discussing the motivation behind its construction.

NeuroIIR 2017, Oslo, Norway
2017. n/a…$n/a
DOI: n/a

2 Motivation

Using EEG signals it is possible to detect attention-related events that are understood to be indicative of the piquing a user's interest – or more specifically the allocation of their attention to one particular stimulus as opposed to some other. One characteristic pattern of activity, commonly known as the P300 (Polich 2007), has been a focus of investigation as it can be used as an index of attentional resource allocation to a stimulus such as an attentionally captivating image (potentially due to its infrequency) presented on a screen. This finding has enabled BCI systems to leverage the ability of a user to be able to guide their attention in such a manner so as to be able to provide relevance judgements/ratings on visual stimuli. For example, a user can actively 'look out' for a particular type of image so that when relevant images appear in a high-speed visual presentation sequence known as RSVP (Rapid Serial Visual Presentation), they will subsequently elicit a P300 response that can be detected using signal processing and machine-learning methods. Ultimately this allows the image to be 'neurally' labelled by the participant.

While systems like these have been explored in a proof-of-concept manner in BCI research using a multitude of image-search tasks, the datasets used usually remain unshared between studies, making it difficult to meaningfully compare the machine-learning and feature-processing strategies used, to find those that offer an optimal generalisability both across tasks and subjects. EEG responses are rife with variability for numerous reasons, such as differences between experimental participants, between task parameters, or changes that can even occur over the course of an experiment. Such sources of variability impede systematic identification of best-practice methods and strategies in signal processing and machine learning for using neural responses from image signals to label them. This is what the NAILS dataset seeks to redress, that is to provide a common dataset to allow researchers to investigate best-practice strategies for RSVP-EEG image search applications utilising a range of image-search tasks (in a repeated-measures design).

Table 1. NAILS Tasks. *standard images were extracted in a balanced manner from the remaining visual categories in the dataset. For the Places365 dataset there are 364 categories and for the VEDAI dataset there are 8. BA (Balanced Accuracy) is shown for naive linear SVM models tested on withheld blocks within search tasks for one participant.

TaskID	*Dataset*	*Target*	*Standards*	*BA*
1	Places365	Wind Farm	Field Road	.86
2	Places365	Wind Farm	*All Categories	.79
3	ImageNet	Keyboard	Instruments	.80
4	ImageNet	Macaw	Birds	.86
5	VEDAI	Plane	Pickup	.86
6	VEDAI	Plane	*	.89

3 NAILS Data Set & Collection

The NAILS dataset contains EEG responses to 116,640 images, in total, from 12 experimental participants. Data collection has been carried out with approval from Dublin City University's Research Ethics Committee (DCUREC/2016/099). Each participant completed 6 different search tasks (for a particular type of target — see Table 1), where each search task was divided into 9 (approximately 35 second) blocks which were completed in a self-paced manner so as to alleviate strain on participants. In each search task, a participant searched for a known type of target (e.g. an airplane), and was instructed to covertly count occurrences of target images in the RSVP sequence so as to maintain their attention on the task. In each RSVP block, images were presented successively at a rate of 6 Hz with target (search-relevant) images randomly interspersed amongst standard (non-search relevant) images with a percentage of 5% across all blocks. In each block, 180 images (9 targets/171 standards) were presented in rapid succession on screen. Per participant, there were 486/9234 target/standard examples available. As contaminant eye-movement related activity on the EEG can often contain useful information, epochs (from -1000ms, 2000ms) containing such activity were excluded as they might encourage developed strategies to utilise these non-neural sources of discriminative information. Epochs were filtered to exclude those with a peak-to-peak amplitude greater than 70 μV on EOG and frontal EEG channels. ICA (Independent Component Analysis) was used alongside a wavelet based analysis to confirm that the remaining epochs did not contain non-neural sources of discriminative information. For the workshop's collaborative evaluation, this dataset was split into a training/testing set, where 15/285 target/standard trials from each search task were selected to act as a withheld test set in the evaluation. Competing organisations in the collaborative evaluation using the supplied training data (remaining epochs from blocks not used to extract test set data) needed to build machine-learning models that maximise a BA (balanced accuracy) score on the entire withheld testing set. That means for an evaluation run, an organisation needed to submit binary predictions for the 21,600 examples given in the test set (1080/20520 targets/standards respectively). There were more than 2500/50000 target/standard training examples available across all participants for model training. In Figure 1, we show a characteristic P3b response occurring for one experimental participant. Applying a naive classification pipeline using downsampled time-series data along with a linear SVM we find the signals generated in the tasks are suitable to be used with even very simple ML strategies (Table 1). These measures, in part verify that the chosen tasks are eliciting the expected characteristic oddball P300 response i.e. it is possible to do the search tasks. Images tasks were constructed using freely available datasets (Razakarivony and Jurie 2016; Russakovsky et al. 2015; Zhou et al. 2016). These were selected as a good choice given they are commonly used datasets with well-researched characteristics that are representative of the visual content typically encountered in multimedia-IR tasks whilst remaining similar to content used in previous RSVP-BCI studies.

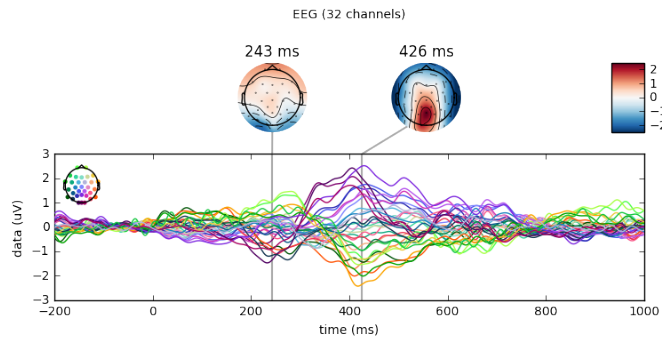

Figure 1. Butterfly plot (ERP average) of target epochs across all blocks minus average standard epochs across all blocks. Plots are generated using CAR (common average reference). Characteristic P3b activity can be seen at posterior scalp sites approximately between 300ms and 600ms following target detection (peaking at 426ms). The colors on time-series plots indicate electrode location on scalp (upper left).

4 Expectations and Conclusions

In this paper we have described the motivation behind the creation of the NAILS dataset and detailed key parameters in its construction. Future work will investigate (and similarly) construct datasets where target detection from neural signals may be difficult due to the presence of attentionally captivating non-target images.

Acknowledgments

This research was supported by Science Foundation Ireland under grant number SFI/12/RC/2289.

References

Graham Healy and Alan F Smeaton. 2011. Eye fixation related potentials in a target search task. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society.* 4203–4206.

Amar R. Marathe, Anthony J. Ries, Vernon J. Lawhern, Brent J. Lance, Jonathan Touryan, Kaleb McDowell, and Hubert Cecotti. 2015. The effect of target and non-target similarity on neural classification performance: a boost from confidence. *Frontiers in Neuroscience* 9 (2015), 270.

Eric A Pohlmeyer, Jun Wang, David C Jangraw, Bin Lou, Shih-Fu Chang, and Paul Sajda. 2011. Closing the loop in cortically-coupled computer vision: a brain-computer interface for searching image databases. *Journal of neural engineering* 8 3 (2011).

John Polich. 2007. Updating P300: an integrative theory of P3a and P3b. *Clin Neurophysiol* 118, 10 (Oct 2007), 2128–2148.

Sebastien Razakarivony and Frederic Jurie. 2016. Vehicle detection in aerial imagery : A small target detection benchmark. *Journal of Visual Communication and Image Representation* 34 (2016), 187 – 203.

Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. 2015. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)* 115, 3 (2015), 211–252.

Bolei Zhou, Aditya Khosla, Àgata Lapedriza, Antonio Torralba, and Aude Oliva. 2016. Places: An Image Database for Deep Scene Understanding. *CoRR* abs/1610.02055 (2016).