Convolutional neural networks on eye tracking trajectories in patients with spatial neglect

Benedetta Franceschiello1,2, Alexia Bourgeois3, Astrid Minier1,2, Micah M. Murray1,2,4,5, Pierre Pouget6, Paolo Bartolomeo6, and Fabio Anselmi7,8

1 The LINE (Laboratory for Investigative Neurophysiology), Radiology Department, University Hospital Center and University of Lausanne, Lausanne, Switzerland
2 Ophthalmology Department, Fondation Asile des aveugles and University of Lausanne, Lausanne, Switzerland
3 Laboratory of Cognitive Neurorehabilitation, Faculty of Medicine, University of Geneva, 1206 Geneva, Switzerland
4 Sensory, Perceptual, and Cognitive Neuroscience Section, Center for Biomedical Imaging (CIBM), Lausanne, Switzerland
5 Department of Hearing and Speech Sciences, Vanderbilt University, Nashville, TN, USA
6 Sorbonne Université, Inserm U 1127, CNRS UMR 7225, Paris Brain Institute, ICM, Hôpital de la Pitié-Salpêtrière, 75013 Paris, France
7 Center for Neuroscience and Artificial Intelligence, Department of Neuroscience, Baylor College of Medicine, Baylor Plaza, 77030, Houston, USA.
8 Center for Brains, Minds, and Machines | McGovern Institute for Brain Research at MIT, Cambridge, USA.

Abstract. Eye-movement trajectories are rich behavioral data, providing a window onto how the brain processes information. Analyses of these trajectories can be automated and benefit from machine learning algorithms. Among those, deep learning has recently proven very successful, setting new state-of-art results in many computer vision applications, including medical diagnosis systems. In this paper, we address the challenge of diagnosing and quantifying signs of visuospatial neglect from saccadic eye trajectories recorded in healthy controls and in brain-damaged patients with spatial neglect. We show how machine learning techniques, such as deep networks, can predict the patient’s status with unprecedented accuracy, benchmarking the algorithm prediction with structural Magnetic Resonance Images (MRI) of the patients’ brain lesions and their Diffusion Tensor Imaging (DTI) tracts. Preliminary evidence of correlation between MRI data and the algorithm scores suggest that a quantitative prediction of the patients’ impairment based only onto the behavioral data of eye trajectories seem possible, therefore opening to new horizons in the field of non-invasive diagnostics.

Keywords: neglect, bio-markers, eye-tracking, deep networks, structural lesion, diffusion tensor imaging

1 Introduction

Vision provides a window into the brain. The retina is part of the central nervous system, and the eye is non-invasively and readily accessible. The behavioral performance
of the eye measured through saccadic eye-movements has been the focus of decades of research linking functional oculomotor behavior to brain impairments (1). Saccadic eye-movements can be a precursor of brain pathology and may also constitute an important bio-marker for early diagnosis of brain impairments. They may also be particularly affected after a focal brain lesion, for instance in patients suffering from neglect (5; 21). Left spatial neglect is a frequent, but still poorly understood neurological syndrome that is characterized by a lack of awareness of contralesional stimuli following right hemispheric damage (3). A high degree of overlap between attentional orienting deficits in neglect and their oculomotor performance has been demonstrated. Thus, neglect patients exhibit saccadic impairments including direction-specific deficits of saccadic production (6; 27), saccadic amplitude or difficulty retaining locations across saccades (17). From a clinical perspective, the diagnosis of neglect is crucial, as the syndrome is associated with poor functional outcomes (18). Nonetheless, the neural mechanisms underlying neglect remain still debated. Neglect has been linked with structural damage of key parietal regions such as the temporoparietal junction (TPJ) or the inferior parietal lobule (13). Other studies indicated that damage of long range of white matter tracts connecting frontal and parietal areas may represent crucial antecedent of neglect (4). Neglect has been reported following damage to the superior longitudinal fasciculus (SLF), or the inferior fronto-occipital fasciculus (IFOF) disrupting connectivity within attentional network of the brain (26; 25). Interestingly, these studies highlighted that this syndrome could result from the disruption of quite diverse areas of the brain and their networks, rather than the dysfunction of a single cortical region (2).

Convolutional neural networks (CNNs) have proven extremely successful algorithms, particularly in image classification, even outperforming human observers and experts. If given sufficient data for training, CNNs are very versatile function approximators capable of detecting subtle patterns that are crucial for the final task (e.g. classification). Thanks to the enormous amount of data generated within the healthcare sector, recently CNNs have been very successfully employed in a variety of image-driven medical diagnosis domains (e.g. radiology (12; 28), ophthalmology (22)). Current research on machine learning techniques applied to eye-tracking data have hitherto focused on different fields, such as classification of eye movements (fixation, saccades, etc.). (29) or computer-assisted diagnosis tools (19). Other machine learning techniques provide early diagnosis methods for classification and detection of neurodevelopmental disorders (24), methods aiming at detecting the presence of strabismus (10), or deep learning methods for the detection of Alzheimer’s disease (8; 7), or Mild Cognitive Impairment (20) based on eye-movement behavior identification.

The contributions of the present work are threefold. First, the pre-processing of eye-tracker trajectories is often not detailed in its steps, despite data needing to undergo cleaning and re-organisation prior to analysis. Eye-tracking data typically contain errors and noise that must be accounted for (16). In this paper an overview of the pre-processing required is standardised, to provide a pipeline adaptable to other task-based eye-tracker measurements. Secondly, this paper demonstrates how modern machine learning algorithms, in particular deep convolutional networks, allow learning of representations of data features that are particularly effective in classifying pathological versus non-pathological conditions, in the case of neglect, from patterns of eye-
movements. Finally, the crucial interplay between classification based on eye-tracking trajectories and the anatomical markers as detected through structural magnetic resonance imaging (MRI) and diffusion tensor imaging (DTI) is investigated, establishing a structure-function link between the behavioral measurements such as eye-tracking and anatomical MRI-based ones. To the best of our knowledge, this is the first time that convolutional neural networks are used to determine and quantify the presence of neglect through eye-movement analysis during a visual search task and that a link between the algorithm outputs and anatomical markers is established.

2 Material and Methods

2.1 Behavioral and neuroimaging data collection

Participants. A subset of 6 right brain-damaged patients with neglect and 6 healthy individuals recruited for a previous study (9) were included in this analysis. Demographic and clinical characteristics of the patients are presented in Figure 2.1. Patients were considered as showing signs of neglect if they were impaired on at least two neuropsychological tests. Healthy individuals were age-matched with the patients (mean age 58 years, range 45-69, t < 1) and had no neurological or psychiatric history.

Apparatus, stimuli and procedure. Participants were asked to perform the visual search task described in detail in (9). Eye-movements were recorded at a sampling rate of 300 Hz with an eye-tracker Tobii TX300. Participants were required to maintain their gaze to the cue, presented centrally, and to freely move their eyes as soon as the cue disappeared to respond to the target. The cue lasted for 3000ms corresponded to a white circle, surrounded with eight peripheral circles. The circle representing the cue was imaginary subdivided in four quadrants. For each trial, one of these four quadrants was highlighted in white, serving as an voluntary, endogenous attentional cue, and indicating the most likely location of target appearance. The cue correctly indicated the target location on 73% of the trials (valid location). The target appeared in one of the three uncued quadrants (invalid location) on 18% of the trials. The target was not present in the remaining 9% of the trials (catch trials), which were included in the design to avoid guesses and anticipations. The target was then presented after a random interval ranging from 1000 to 2000 ms. The target was created by eliminating either the upper or the lower part (0.4° of visual angle) of one of the eight peripheral circles. The colour of each target and distractors changed randomly in each trial, thus inducing a serial, demanding search. Participants were asked to move a joystick up when the upper part of the circle was missing, or down when the lower part was missing, as fast and as accurately as possible with their right hand. The target disappeared when a response occurred or after 3000 ms if no response was made. This experiment was composed of a total of 176 trials.

Neuroimaging data collection. Lesion extent was determined for each patient by manually drawing the lesion borders onto the original 3DTI MRI (for the complete preprocessing pipeline, please see (9)). In order to reduce lesion-induced registration errors, spatial normalization was performed using a mask that excluded the damaged
Fig. 1. (Top). Demographic and clinical characteristics of neglect patients, with their performance on visuo-spatial tests. Characters with asterisks denote pathological performance compared to normative data. For line bisection, positive values indicate rightward deviations, negative values indicate leftward deviations. Scores for the landscape drawing (15) indicate the number of omitted left-sided details. I, ischemic; H, hemorrhagic; NA, not available. (Bottom-left) The preprocessing steps required to analyse the eye-tracker data is presented together with the corresponding parts of the visual search behavioral paradigm task they refer to. (Bottom-right) An example of clean trial, as well as the mean trajectories over all trials and subjects for each target.
areas of the brains, thereby preventing these areas from biasing the transformation. After normalization, the brain lesion was segmented, and its borders were redefined in the normalized brain. Diffusion Tensor Imaging (DTI) tractography was used to study long-range of sub-cortical white matter pathways. Again, for the complete preprocessing pipeline, see (9). The mean fractional anisotropy values of the three branches of the superior longitudinal fasciculus (SLF), the cingulum, and the inferior fronto-occipital fasciculus (IFOF) was extracted in the right hemisphere. These tracts were chosen on the basis of their implication in attention networks and in visual neglect (4).

2.2 Data processing and analysis

Preprocessing. Targets presented within the left visual field were considered for the analysis, as neglect signs concerned first and foremost attentional orienting towards the left, controlesional hemispace. Recorded eye-tracking trajectories underwent a preprocessing procedure to standardize the dataset before analyzing it. The code was developed with Matlab R2019b. First, the duration of the trajectories was standardised across trials, i.e. all trials were uniformed in order to be 9s long (9000 ms). Time series corresponding to trials ending before 9s were filled with NaN (not a number, missed recording). The analysis focused on those trials both valid and not missed by the subject, according to the definition given in section 2.1. Only time-stamps recorded after the first 3000 ms and before 9000 ms were first taken into account. A loop over the total number of targets seen by each subject was iterated to analyse all trajectories recorded per each subject. For each target \( i \) the set of trajectories coordinates \((Eye_x, Eye_y)\) was extracted and coordinates corresponding to outliers (such as points outside the screen or above 9000 ms) were respectively filled with “NaN” if acquired between 3000 and 9000 ms, otherwise removed. All missed recordings were then re-filled according to the following criteria: if a NaN was present at the beginning of the trial, the NaN value was replaced with the center of the screen coordinates, equal to (384, 512) pixels. If a NaN was in the middle of the time series, this value was linearly interpolated using the available neighbouring sampled data, corresponding to actual recordings. If the target was reached at a certain timestamp, all remaining points (to reach the 9000 ms upper limit) were filled with the target coordinates. The mean of the first 3000 ms of \( Eye_x \) and \( Eye_y \) was used as first timestamp of a shortened time series 6000 ms long, now cleaned-up and interpolated (see Figure 2.1). The post-processed time-series corresponds to the 6000 ms of the visual search task. Every trial for every subject and every target \( i \), \((Eye_x, Eye_y)\), was z-scored by subtracting its mean and dividing by its standard deviation. All pre-processing steps are schematized in Figure 2.1.

Previous use of machine learning techniques on eye-movements. Few studies have analyzed the correlation between brain dysfunctions and eye movements, using machine learning techniques. Among those, in (8), the authors use deep neural networks to identify patients with Alzheimer disease (AD): they first train a set of denoising sparse-auto-encoders and feed the output to a neural network with softmax classifier. The study highlights how in principle CNNs can be used early diagnosis of AD. CNNs were also employed in (23) and (10) to detect, from eye movements, developmental disorders and strabismus, respectively. These studies show how Deep Neural Networks (DNNs) are
able to capture subtle geometric features of eyes trajectories, to which common diagnostics could be blind to, to inform about the patient’s status.

**Classifying healthy vs neglect patients from eyes tracking images.** In this section we discuss the methodology we applied to estimate healthy versus neglect patients’ status from a subset of eye-tracking trajectories, i.e. those corresponding to targets belonging to the left hemifield. Left hemifield targets are more challenging for patients affected by this syndrome. To this purpose, we formulated the estimation problem as a classification task, i.e. learning the mapping between an appropriate representation of the eyes’ trajectories and the patient’s label. In particular, as explained in the data pre-processing section, we restrict our analysis to the $x$ coordinates of the trajectories ($x$-projection). The resulting vectors have length $d = 2001$. Relevant trajectories’ features were extracted by passing these vectors through a network composed of a sequence of 3 blocks of convolutional layers, each one separated by a pooling layer halving the output vector dimension (details in Figure 2). The convolutional part of the network was followed by a classifier composed by a fully connected layer, a dropout layer and another fully connected layer with softmax nonlinearity, to predict the patients’ labels. All the computational models were implemented using Keras and Tensor Flow (11). The training has been performed on the 80% of the data samples (570 trajectories of dimension $d = 2001$), while the remaining, 141 trajectories, 20% of the total, has been employed for test evaluation only. At each of the 8 epochs, after a random permutation of the order of the trajectories, the training set was further split into 80% and 20% to validate the model (114 trajectories). To avoid overfitting, we used a 50% dropout regularization within the fully connected layers. An Adam optimizer was chosen with a standard batch size of 32 and a learning rate of 0.001 to minimize the Cross entropy loss.

### 3 Results

**Behavioral results.** In (9), oculomotor behaviour was studied together with manual responses on a visual search task in neglect to determine the relation between saccadic parameters and sub-clinical disorders of spatial attention. The study showed the occurrence of inappropriate rightward saccades during target selection: when left-sided targets were presented, saccades were equally likely to be performed towards the left side or towards the right side. Right-sided distractors may erroneously capture patients’ attention, leading to an over-exploration of the right hemispace, consistent with the magnetic attraction of gaze typically observed in neglect (14). Thus, pathological production of eye movements should be considered as a subtle manifestation of visuo-spatial disorders.

**Classification results.** The average classification accuracy obtained on the test set for 10 random data splitting with our final architecture (see Fig 2) was $74 \pm 5\%$. Overall the network performance in classification is very promising and support our methodology. The results we obtained are in line with performance reported in the literature, see 2.2 (although the visual search task is different): patients’ healthy vs unhealthy status can indeed be evaluated extracting salient features from the geometry of eye-trajectories in
visual tasks, using a deep convolutional analyzer. To determine the final network architecture we explored different outputs. In particular we analyzed separately \( x \) and \( y \) projections of the eyes trajectories and the instant moving direction angle \( \theta = \Delta x / \Delta y \). Results showed best performance for \( x \)-projections data and a drop of \(-5\) to \(-10\)% for \( y \)-projections or angle analysis, see [4]. Furthermore, the analyses with 2d trajectories, both with a pre-trained or fully trained VGG16 network, showed poorer performances slightly above chance 55/60%. This suggested that a 1d network trained on \( x \)-projections was the best architecture to employ. Finally, for each patient we produced a score evaluating the trained model on a set of test trajectories relative to that patient. The score is a measure of the probability \( 0 < c < 1 \) computed by the model to classify the patient for each test trajectory. We averaged the score among all test trajectories and repeated the procedure for seven trained models. The final score for each patient is the average score of the seven runs.

**Fig. 2.** (Left). The architecture we used in our experiments. Three convolutional blocks (light blue) with pooling are followed by two dense layers with a dropout regularization layer. Dimensions of the output at each layer are reported in the image. (Right) The network performance over 17 Runs for each epoch (1-8) is reported in the boxplot in green, while in blue we report the loss.

**CNN results and Neuroimaging.** Interestingly, preliminary results show a positive correlation (\( \rho = 0.78 \)) between the lesion volumes and algorithm score, despite the small lesion overlap. A partial correlation with the lesion volume as regressor showed a preliminary link between the algorithm score and RIFOF (\( \rho = -0.91 \)) and the Cingulum (\( \rho = -0.95 \)). However, we observe that the sample size of \( N = 6 \) is too small to establish a clear statistical relationship: this will be further investigated by enlarging the sample size to the full population in [6] and insure statistical robustness.
Fig. 3. Lesion overlap presented along different slices (top), and in the coronal - sagittal and axial view (center left). Each color indicates one subject, enhancing the different sizes and areas of the lesions across patients. For \( N = 6 \) patients (Center right) the correlation between the algorithm scores and the lesion volumes is presented (\( \rho = 0.78 \)), as well as (bottom) between the algorithm scores and RIFOF (\( \rho = -0.9140 \)) and the Cingulum (\( \rho = -0.9497 \)) respectively.
4 Discussion and conclusions

This paper demonstrates how deep convolutional networks allow learning of representations of data features which are particularly effective in classifying pathological versus non-pathological conditions. Although several studies highlighted pathological eye-movements behaviour as a consequence of impaired attentional shift of attention in this syndrome, the measurement of eye-movements behaviour as a tool to diagnose neurological pathology has only begun to be studied. To first address the research question, the pre-processing of eye-tracking recordings is clearly outlined in its dependency on the visual search task: the steps are versatile and easy adaptable to other visual studies. To the best of our knowledge, this is the first time that CNNs are used to determine the presence of neglect through eye-movements analysis during a visual search task. This results are confirmed by the correlation between the algorithm score output and the anatomical markers of the patients’ (MRI and DTI).

The neglect syndrome is characterized by unawareness of stimuli presented in the left, contralosional visual field, accompanied by impaired oculomotor performances. Spatial orienting responses are similarly biased to the right, i.e. the magnetic gaze attraction phenomenon (14). In agreement with this observation, the performance of the CNNs employed was much worse (average $-15\%$) on 2D trajectories then on the $x$-projections: patients saccades are compromised along the right-left direction with no bias on up and down directions. We counter-checked this observation by training a CNN only using the $y$-projections of the trajectories: results show a drop of $10\%$ in classification performance. The lower performance of the 2D images is due by the fact that $y$-coordinates act as distractor, since only the left-right direction carries useful information. The observation is also confirmed by the network performance using the instant angle direction of the saccades’ movements as input $\theta = \Delta y/\Delta x$. Performance drops of $(-5\%,-10\%)$, confirming that $y$ component acts as a distractor. These results thus suggest that typical features we are looking for classification are local, sometime subtle, right-left anomalous shifts. This observation was further confirmed by the fact that global features extractors (induced by a many layers network) were also not suitable with respect to the classification task: indeed architectures with bigger convolutional kernels size performed worse ($-20\%)$. We also tested on the shelf VGG16 (2D) models performing 1) full-training; 2) fine tuning on the features extracted from the pre-trained convolutional layers on imagenet. Results confirmed that 2D networks architectures perform poorly on our dataset.

Neglect is a multi-component syndrome, characterized by dissociation across patients’ performance and also between tests of spatial performance within patients. Structural and functional anatomical studies highlighted that numerous lesions at the cortical and subcortical levels could induce neglect. Our population shows substantial variability in lesion location. In our study, the classification score of eye-movements trajectories appears to be linked to different anatomical markers of the patients’ brain: the lesion volumes and the RIFOF and Cingulum tracts, once the lesion volume is regressed out in the correlation. Further investigations should consolidate this link, allowing to differentiate and predict patterns in agreement with the anatomical markers with unprecedented precision, especially as the neural substrates of neglect are still debated. Saccades are generally evaluated qualitatively by a qualified examiner and quantitatively by the eye-
tracker specialised in saccadometry. Only the differential diagnosis of ocular movements is subject to the interpretation of the medical doctor: our method introduces a quantitative evaluations of eye-movements trajectories. This cutting-edge classification represents therefore a powerful instrument to understand and predict the trajectory of patients’, opening to the possibility of a new computer-aid diagnosis tool in neglect.
Bibliography


