Deep Multi-Task Learning for SSVEP Detection and Visual Response Mapping

Hong Jing Khok[†β] **Victor Teck Chang Koh**[α] **Cuntai Guan**[β]
[†]DAMO Academy, Alibaba Group
[α]Department of Ophthalmology, National University Hospital, Singapore
[β]School of Computer Science and Engineering, Nanyang Technological University, Singapore

***Abstract*—Glaucoma is an eye disease that occurs without the onset of symptoms at initial, and late diagnosis results in irreversible degeneration of retinal ganglion cells. Standard automated perimetry is the gold standard for assessing glaucoma; however, the examination is subjective, where responses can fluctuate each time the test is performed, significantly confounding the test's interpretation. In this study, we present our approach that aims to provide a rapid point-of-care diagnostics for glaucoma patients by eliminating the cognitive aspect in existing visual field assessment. Unlike existing methods that mostly report the foveal target detection's accuracy, we employed a multi-task learning architecture that efficiently captures signals simultaneously from the fovea and the neighboring targets in the peripheral vision, generating a visual response map. Furthermore, we designed a multi-task learning module that learns multiple tasks in parallel efficiently. We evaluated our model classification on a 40-classes dataset, with yields 92% and 95% in accuracy and F1 score respectively. Our model is able to perform on a calibration-free user-independent scenario, which is desirable for clinical diagnostics. Our proposed approach could be a stepping stone for an objective assessment of glaucoma patients' visual field.**

***Index Terms*—SSVEP, Multi-task Learning, Convolutional Neural Network**

I. Introduction

Glaucoma is a worldwide leading cause of irreversible vision loss, possibly affecting 111.8 million people worldwide in 2040 [1]. As glaucomatous visual field losses progress without noticeable initial symptoms, this results in late diagnosis, where the degeneration of retinal ganglion cells has already occurred with irreparable consequences [2]. Many glaucoma suspects are suffering silently from peripheral vision loss, with up to 50% of previously undiagnosed glaucoma patients already had significant visual field defect at diagnosis [3].

The key diagnostic sign for a glaucoma patient is peripheral vision loss, referring to the maximum angle field of vision from the center of fixation for each eye. Glaucoma suspects are advised to assess their visual functions early and regularly. This assessment is done by standard automated perimetry (SAP); it is the gold standard performed to assess and diagnose the disease's severity. The assessment of blind areas in the visual field is used to monitor visual function in glaucoma. The procedure duration is approximately 10-minutes per eye, where the patient is required to look at a large semi-circular bowl that covers their entire field of view. The system will present a series of stimuli (spots of light), one at a time while the patient has to maintain fixation at a central target during the entire procedure. The patient responds by clicking a button when a stimulus is seen, and this process yields a map of the seeing parts of the patient's field of view.

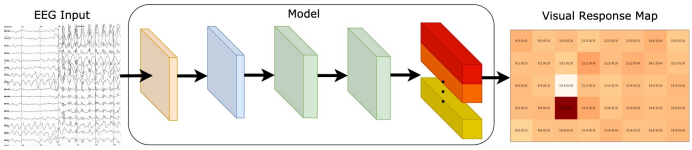

Fig. 1. We present a multi-task learning framework for generating a visual response map potentially suitable for providing glaucoma diagnostics.

SAP is a subjective examination where varying responses may be obtained each time the test is performed or even during the same test [4]. These fluctuations usually increase with the severity of the disease. It has been the biggest drawback of this visual field assessment, as it may significantly confound the test's interpretation. These conflicting results may hinder the physician's decision to make a diagnosis and order multiple subsequent tests. In order to reduce results variability, patients must first be educated to use the system; this would take multiple sessions, depending on the patient's ability to carry out the procedure. Approximately ten visual field tests are needed to achieve an accurate prediction point-wise and mean sensitivity for a typical glaucoma patient in the clinic [5], which in turn can lead to a delay in diagnosis. This highlights the current limitations of visual field testing. A study [6] investigates patients' experiences regarding their glaucoma follow-up indicates that patients find visual field examination by SAP challenging or uncomfortable. Some patients who underwent multiple tests describe their feelings of anxiety. As a patient's cognitive ability is involved in the assessment, it can produce inaccurate test results due to the patient being inexperienced with the system and examination procedure. A patient could lose fixation to the central target due to fatigue and distraction, which could also affect the test results [7]. These factors highlight a need for a technological solution that provides an objective assessment that is highly desirable to improve visual field test efficacy.

Steady-state visual-evoked potential (SSVEP) is an oscillatory stimulus-response evoked by certain repetitive stimuli with a constant frequency. These are recorded in the electroencephalogram (EEG) and can be detected from the primary

visual cortex. These produce a brain response that has the frequency that matches the rapid flickering stimulation in amplitude and phase. With this method, the need for cognitive processing can be eliminated, removing the requirement for the patient to click a button when a stimulus is presented. Instead, we extract a stimulus's presence by detecting the stimulus' frequency from the EEG, which we can yield a map of the patient's visible field of vision.

SSVEP has many neuroscience applications, such as visual spelling [8], [9], and decoding user intends to operate assistive devices [10], [11]. SSVEP is popular due to its ease of recording and high signal-to-noise ratio [12]. Despite a few studies that use SSVEP on glaucoma applications, using SSVEP to assist in glaucoma diagnosis could yield significant potential. A study [13] shows that brain-computer interface devices were able to discriminate glaucomatous eyes from healthy eyes. Another study [14] shows that it is possible to use SSVEP responses to detect visual signals of varying view angles in the peripheral field. These studies suggest that SSVEP could be promising for objectively assessing visual function loss and detecting glaucomatous damage.

Most studies have been devoted to identifying a single flickering target where the focus is on delivering reliable SSVEP responses detected on fovea vision; hence stimuli at the peripheral vision are considered noise signals. In this study, we focus on detecting the peripheral field to diagnose glaucoma patients by capturing signals from the fovea and the neighboring targets in the peripheral vision. A fundamentally different way to detect SSVEP was proposed in this study, employing the use of Multi-task Learning (MTL) [15], which is to train a neural network with multiple related objectives (or tasks) while sharing a common network structure. MTL can determine how tasks are related without being given a specific knowledge for task relatedness. Training a network that learns to predict multiple tasks in one network performs better than training multiple separate networks to predict multiple tasks [15]. In SSVEP detection, we utilize MTL to improve SSVEP classification performance. The shared layers enabled the model to decode SSVEP from raw EEG, and task-specific layers will learn to separate different target frequencies. The number of outputs corresponds to the number of target frequencies in the dataset, where each output is the probability of the SSVEP frequency present in the EEG. MTL will allow us to identify a patient's visual field by detecting multiple SSVEP targets at once, thus cutting down the patient's assessment time.

Convolutional neural networks (CNN) have pushed computer vision tasks' performance because of its remarkable ability to learn directly from images end-to-end without handcrafting features [16]. However, the applications of CNN on SSVEP are still at the beginning stage. Traditionally, the go-to SSVEP detection techniques have been statistical and correlation-based such as canonical-correlation analysis, where its objective is to find the maximum correlation between the signal and the target frequency [17]. However, studies [10], [18] have shown that CNN can provide significant improvement in SSVEP classification performance as compared to traditional SSVEP detection techniques. Interestingly, the top-performing methods in various fields have adopted dilated convolution for efficient dense feature extraction; it has been beneficial in image semantic segmentation [19], speech recognition [20] and signals processing [21]. That is because dilated convolutions are effective feature extractors due to its capability to expand the receptive field without significantly increasing computational cost [22]. As such, our MTL model is a deep CNN model with dilated convolutions.

The main idea of our study is illustrated in Fig. 1 with the aim to implement a system that produces visual field tests results that are more reliable as we eliminate the cognitive aspect in the existing visual field assessment. By doing so, patients do not have to learn to use the system, and test results are not affected by patients being distracted or feeling uncomfortable. Since our MTL model is capable of detecting multiple stimuli, we can reduce visual field assessment time and produce reliable test results. This could be potentially suitable for providing a rapid point-of-care diagnostics for glaucoma patients.

II. Method

A. Data

In this study, we used an open-access dataset by Tsinghua University, HS-SSVEP [9], a 40-classes dataset for visual spelling tasks. The EEG was recorded using 64 channels at the 250-Hz sampling rate, and 35 healthy subjects participated in the experiment. The 40 stimuli flickered at frequencies between 8-15.8 Hz with an interval of 0.2-Hz; each target frequency was presented six times, totaling to 240 trials per subject. Each trial lasted 6-seconds, in which the first and the last 0.5-seconds were used for visual cue and rest.

Out of the 64 electrodes, we selected 11 from the occipital and parietal areas, namely P-z, PO-3/4/5/6/7/8/z, and O-1/2/z. EEG data were preprocessed by band-pass filtering using a 6th order Butterworth between 1 and 40 Hz. Then, the filtered data were segmented and removed the first and the last 0.5-seconds, and we selected the first 1-second for training and validation. Therefore, the input data dimension was 1000 time samples by 11 channels (N_{ch}).

B. Multi-task Learning Architecture

Traditionally, studies on SSVEP were focused on detecting the target stimulating frequency in the foveal, while the other stimuli are interference. In our study, we assess the peripheral field's visual responses by detecting the responses from the off-foveal targets. We proposed a multi-task learning architecture to handle the multi-label learning problem. This approach allowed us to identify the presence of multiple SSVEP frequencies in a single trial. During training, the model learns each SSVEP target separately in parallel and back-propagated jointly to improve the generalization of the EEG input at the shared hidden layer. Our implementation is built

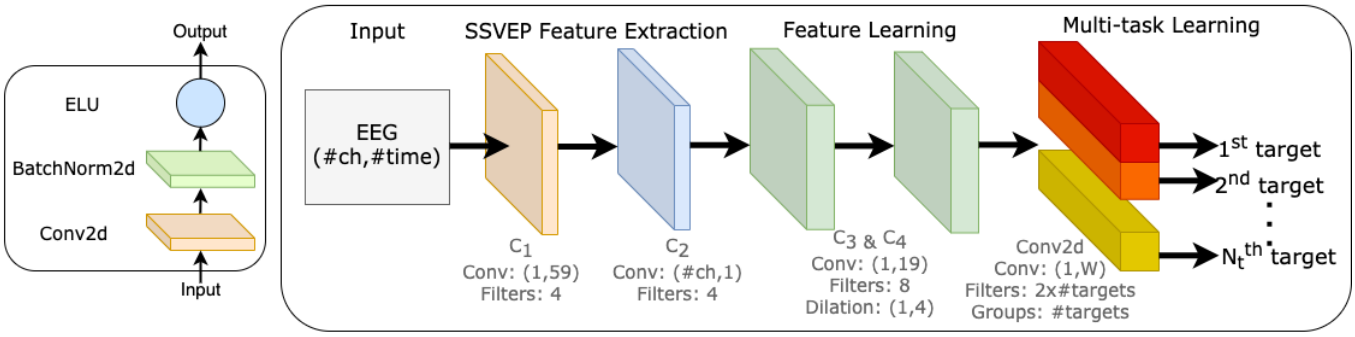

Fig. 2. Left: A convolution block. Right: Proposed multi-label SSVEP classifier with modified multi-tasking learning architecture.

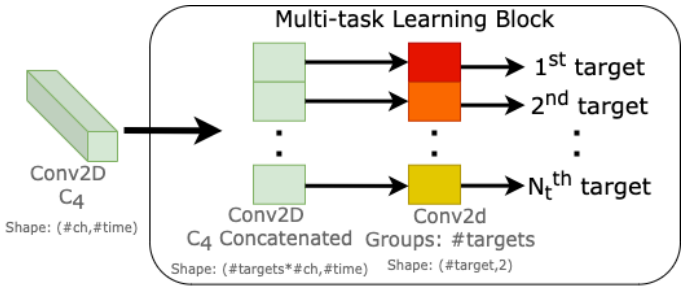

Fig. 3. Multi-task learning block that performs convolutions by groups, performing separate convolutions within a single convolution layer.

on the PyTorch framework and is made publicly available at a companion website[1].

The network is composed of four convolution blocks and one convolution layer. Each convolution block consists of a convolution layer, a batch normalization, and an exponential linear unit, as shown in Fig. 2. The fifth and final convolution layer is a multi-task learning block, where it learns to differentiate multiple SSVEP target frequencies. An illustration of the proposed network is shown in Fig. 2.

Convolution blocks C_1 to C_4 in the proposed MTL architecture are shared hidden layers responsible for learning EEG representations across all target frequencies. A study [23] has shown that the network can learn the EEG's features by convoluting across time, followed by convoluting across channels. C_1 block was designed to extract the spectral representation of the EEG input, as it performs convolution across the time dimension, capturing features from each EEG channel independently from the others. We have chosen the kernel size of 1×59 to allow the network to observe a few cycles of the lowest target frequency, as well as selecting based on experimentation and is therefore described in the Results section. C_2 block was designed for performing spatial filtering, as it performs convolutions across the channel dimension. The objective of this layer is to learn the weights of all channels at each time sample. The convolution kernel size is $N_{ch}\times1$ where N_{ch} is the number of channels. The purpose of C_3 and C_4 blocks are capturing the temporal patterns in each extracted feature maps. We have chosen kernel size 1×19 based on experimentation, further described in the Results section. We also explored different dilation configurations on C_3 and C_4 block. As the kernel size needed for signals is much larger compared to for images, dilation convolutions allow us to expand the receptive field, perform feature learning with a smaller kernel size, which produces a smaller model, and potentially increase performance. We will evaluate the effectiveness of dilation convolutions on this EEG dataset in the Results section.

In regular MTL architectures, each task is separated into task-specific layers, where each layer is responsible for learning to identify each task. In this study, we designed an MTL block that performs convolutions by groups, as shown in Fig. 3. By defining t groups, we are essentially performing t separate convolutions within a single convolution layer. This allows us to use the same model architecture and scale to any number of tasks effectively. To match the input size of the MTL layer's input, we expanded the output from convolution block C_4 by N_t folds by concatenation, where N_t is the number of targets. Next, we employ a group-wise convolution to split the input into N_t different groups of weights, each responsible for learning each target frequency separately. The result of N_t convolutions are concatenated to produce N_t binary output targets. This MTL block allowed us to train multiple tasks in parallel efficiently on a single GPU. We can dynamically tweak our MTL model for any number of tasks, which is potentially suitable for other MTL applications.

Dropout has been used in deep neural network training as a regularisation technique to reduce the network tendency to overfit during the training process. Therefore, we used it after convolution blocks C_2 and C_4, with the dropout set to 0.5.

C. Training Protocol

The model was trained for 100 iterations with a minibatch size of 64. We monitored the accuracy and F1 score on the validation set and used an early-stopping mechanism when the model stops improving for ten consecutive epochs. We used Adam optimizer with an initial learning rate value of 0.01. L2 penalty was added to reduce overfitting by fixing 0.05 to weight decay.

[1]Source code is available at jinglescode.github.io/ssvep-multi-task-learning.

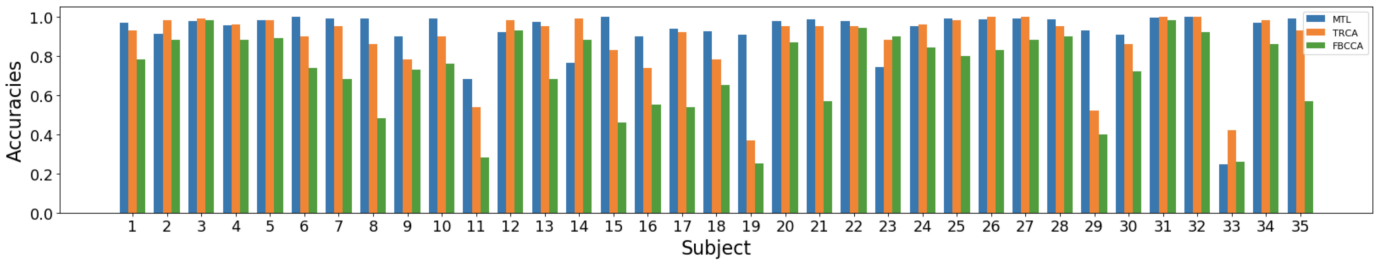

Fig. 4. Accuracies by MTL model, TRCA and FBCCA across subjects.

D. Evaluation

Many studies have been done to explore user-dependent SSVEP classification. In this study, we focused on user-independent SSVEP classification. Our aim is to build a calibration-free SSVEP classification system that is practical for clinical use, capable of diagnosing new patients. This eliminates the need for data collection and training for novel users. We evaluated our model's performance with the leave-one-subject-out method to determine our model's performance in a calibration-free user-independent scenario.

We applied the stratified K-fold technique to experiment and verify multiple model configurations such as kernel size, filter size, and dilation size. The performance of each model is determined by averaging across all subjects and 5-fold cross-validation. We used the accuracy metric to determine the model's ability to detect target frequencies and evaluate true positive performance. F1 score metric was also adopted to determine the model's ability to do well on both positive and negative classes in multiple output MTL scenarios.

III. Results

A. Performance

The classification performance on the 40-classes dataset is as shown in Table I. We compare our model's classification accuracy against Canonical Correlation Analysis (CCA), Filter Bank Canonical Correlation Analysis (FBCCA) [24], and Task-Related Component Analysis (TRCA) [25]. At 1-second data length, CCA, FBCCA, and TRCA yielded approximately 59%, 72%, and 88%, respectively. Using our proposed method, we can identify the foveal's target frequency effectively, with the best performing model configuration (*Net-4/4*) achieved an accuracy of 92.2%. Evidently, this result shows that our approach can be an alternative to identifying a single flickering target where the focus is on delivering reliable SSVEP responses detected on fovea vision. Our model's average cross-validation accuracy for each subject is compared against other methods, as shown in Fig. 4. By evaluating using the leave-one-subject-out method, our model was able to generalize to unseen test data, potentially little to no training and calibration are required for new users, suitable for other SSVEP classification tasks. Our model achieved 95.2% in the F1 score. This result shows the model's ability to do well in identifying both positive and negative classes in a multi-label classification scenario.

TABLE I
PERFORMANCE ON 40-CLASSES HS-SSVEP DATASET

Network	Filters	Dilation	# Params	Accuracy	F1
CCA	-	-	-	59%	-
FBCCA	-	-	-	72%	-
TRCA	-	-	-	88%	-
No Dilation	2	1	290,654	84.8%	90.1%
Net-2/2	2	2	280,414	88.4%	92.2%
Net-2/4	2	4	259,934	90.2%	93.4%
No Dilation	4	1	582,228	86.2%	90.9%
Net-4/2	4	2	561,748	90.0%	93.8%
Net-4/4	**4**	**4**	**520,788**	**92.2%**	**95.2%**
No Dilation	8	1	1,168,376	84.1%	90.4%
Net-8/2	8	2	1,127,416	89.0%	92.4%
Net-8/4	8	4	1,045,496	90.1%	93.6%

B. Effects of Kernel Size and Dilated Convolutions

We experimented with different kernel sizes and evaluated the effects of dilated convolutions. From our experiments, the kernel size of 1×59 for convolution blocks C_1 produced the most consistent performance with a lower standard deviation. Based on our observation, kernel size lower than 1×39 can cause a drop in performance, as single convolution might not be able to detect enough sufficient data points of the lowest frequency. As for convolution block C_3 and C_4, we expanded the receptive field by employing dilated convolutions. Thus, our model can effectively perform feature learning with a smaller kernel size of 1×19, improving classification performance while reducing the model size. We also observed that increasing the number of feature maps increases the model size, but it does not improve its performance.

C. Multi-label Classifcation and Visual Response Map

From a multi-label classifier perspective, we used the MTL approach to learn all 40-tasks, thus enabling a unified system to predict all target frequencies simultaneously. As such, this enables us to visualize what the user has seen with a visual response map. We selected 6-targets that are located around the center of the screen (Fig. 5), as this is the region of interest for visual field assessment in our future work. We have evaluated our models with the leave-one-subject-out method to exhibit the generality of our approach, and our approach requires little or no calibration data novel users. Additionally, we explored further on the models' ability to diagnose users who are not experienced in SSVEP-based BCIs. Specifically, we trained our model with the first eight participants from the dataset and

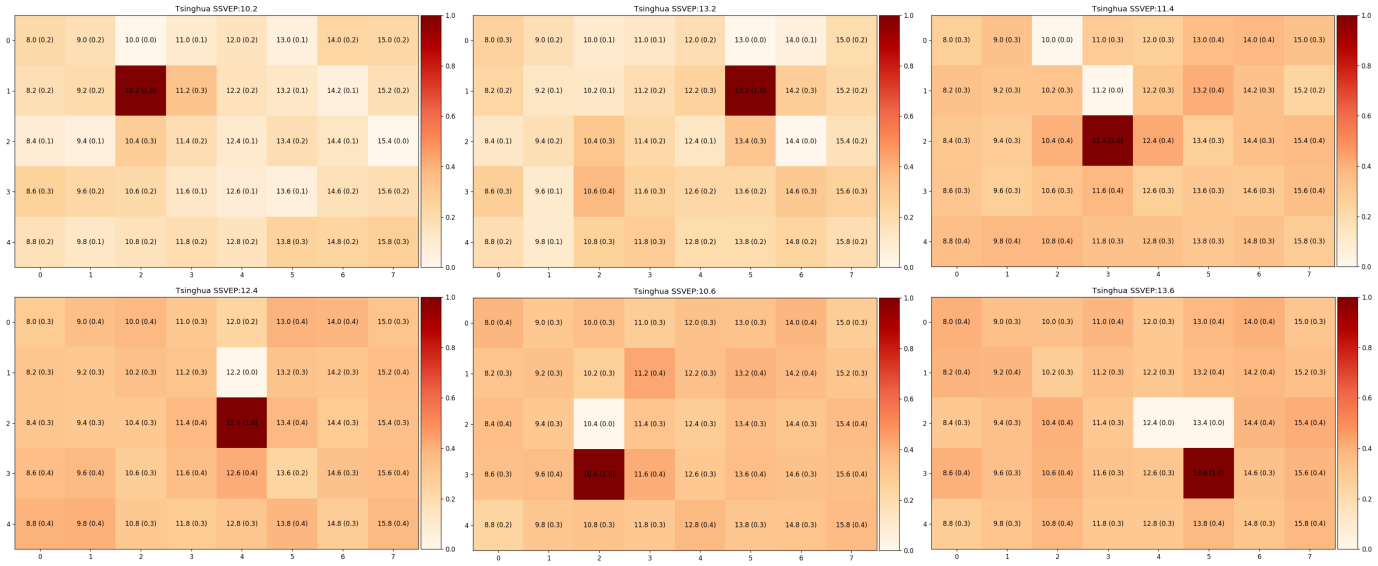

Fig. 5. Visual response maps of subjects unseen by model, predicted by proposed multi-task learning model.

validated the remaining 27 subjects. Each map is the average response of the 27 subjects for a target frequency: a darker shade denotes a stronger signal being detected. As the subjects recruited in the HS-SSVEP dataset have healthy eyesight and considering that all the stimuli are flickering simultaneously in a visual spelling task, expected results will be the intense signal concentration at the target frequency.

From our results, it is evident that the model has successfully identified the target frequencies with a strong signal, and some neighboring off-foveal targets around the target frequency were detected more than those there were further from the target frequency. We noticed that the frequency above the target is often zero; this was due to the loss function minimizing errors during the training. We also observed that some frequencies in the peripheral vision were detected more than others. However, we cannot verify why some frequencies were detected more than others on the visual response map due to the lack of eye tracker information. Thus we planned to verify this in our future work by creating our experiments and data collection. Nevertheless, our MTL model can generalize well on unseen subjects, performing reasonably well even though these participants are not experienced in SSVEP-based BCIs. This is desirable as these are the type participants that we will encounter in the clinic.

IV. Future Work

This study presented a deep learning method that potentially enables us to detect multiple SSVEP stimuli simultaneously, thus mapping a visual map of glaucoma patients, reducing visual field assessment time and produce reliable test results. The results in this study encourage and motivate us to apply MTL to a more challenging dataset to test our hypothesis. As such, we will design our experiments to collect data for this purpose, and utilizes what we have learned from this study in subsequent studies and verify the results presented in this paper with clinical results.

V. Conclusion

Traditionally, studies on steady-state visual-evoked potential (SSVEP) were focused on detecting the target stimulating frequency in the foveal while the other flicking stimuli are regarded as interference. Our study presented an end-to-end multi-task learning approach to detect the responses from the peripheral vision. Furthermore, we designed multi-task learning (MTL) block that learns multiple tasks in parallel efficiently, and we can dynamically tweak our MTL model to the number of tasks.

Our results show that our MTL model can perform in single target SSVEP classification, which could be potentially useful in other SSVEP applications and studies. We observed that the MTL approach is able to learn each target frequency separately, which allowed us to yield a map of the patient's visible field of vision.

In view of recent events during disease outbreak and pandemics where non-essential hospital appointments are recommended to be kept to a minimum, this assessment method can reduce the number of tests needed, thus minimizing any unnecessary or additional tests. In essence, this study enables our future work to potentially assess glaucoma patients' visual field to detect peripheral vision loss. To improve the reliability of the assessment results, utilizing SSVEP could eliminate a patient's ability to carry out the procedure and variability of the patient's mental state. Assessment time could be cut down by detecting multiple SSVEP targets at once and generating a visual response map. Our approach could be potentially suitable for providing a rapid point-of-care diagnostics for glaucoma patients.

VI. Acknowledgment

This research was supported by Alibaba Group Holding Limited, DAMO Academy, Health-AI Division under Alibaba-NTU Talent Program. The program is the collaboration between Alibaba Group and Nanyang Technological University, Singapore.

References

[1] Yih-Chung Tham, Xiang Li, Tien Y Wong, Harry A Quigley, Tin Aung, and Ching-Yu Cheng. Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. *Ophthalmology*, 121(11):2081–2090, 2014.

[2] Robert N Weinreb, Tin Aung, and Felipe A Medeiros. The pathophysiology and treatment of glaucoma: a review. *Jama*, 311(18):1901–1911, 2014.

[3] Jacqueline Chua, Mani Baskaran, Peng Guan Ong, Yingfeng Zheng, Tien Yin Wong, Tin Aung, and Ching-Yu Cheng. Prevalence, risk factors, and visual features of undiagnosed glaucoma: the singapore epidemiology of eye diseases study. *JAMA ophthalmology*, 133(8):938–946, 2015.

[4] Stuart K Gardiner, William H Swanson, Deborah Goren, Steven L Mansberger, and Shaban Demirel. Assessment of the reliability of standard automated perimetry in regions of glaucomatous damage. *Ophthalmology*, 121(7):1359–1369, 2014.

[5] Yukako Taketani, Hiroshi Murata, Yuri Fujino, Chihiro Mayama, and Ryo Asaoka. How many visual fields are required to precisely predict future test results in glaucoma patients when using different trend analyses? *Investigative ophthalmology & visual science*, 56(6):4076–4082, 2015.

[6] Fiona C Glen, Helen Baker, and David P Crabb. A qualitative investigation into patients' views on visual field testing for glaucoma monitoring. *BMJ open*, 4(1):e003996, 2014.

[7] Luciana M Alencar and Felipe A Medeiros. The role of standard automated perimetry and newer functional methods for glaucoma diagnosis and follow-up. *Indian journal of ophthalmology*, 59(Suppl1):S53, 2011.

[8] Han-Jeong Hwang, Jeong-Hwan Lim, Young-Jin Jung, Han Choi, Sang Woo Lee, and Chang-Hwan Im. Development of an ssvep-based bci spelling system adopting a qwerty-style led keyboard. *Journal of neuroscience methods*, 208(1):59–65, 2012.

[9] Yijun Wang, Xiaogang Chen, Xiaorong Gao, and Shangkai Gao. A benchmark dataset for ssvep-based brain–computer interfaces. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(10):1746–1752, 2016.

[10] No-Sang Kwak, Klaus-Robert Müller, and Seong-Whan Lee. A convolutional neural network for steady state visual evoked potential classification under ambulatory environment. *PloS one*, 12(2):e0172578, 2017.

[11] Krupal Sureshbai Mistry, Pablo Pelayo, Divya Geethakumari Anil, and Kiran George. An ssvep based brain computer interface system to control electric wheelchairs. In *2018 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, pages 1–6. IEEE, 2018.

[12] Sergio Parini, Luca Maggi, Anna C Turconi, and Giuseppe Andreoni. A robust and self-paced bci system based on a four class ssvep paradigm: algorithms and protocols for a high-transfer-rate direct brain communication. *Computational Intelligence and Neuroscience*, 2009, 2009.

[13] Masaki Nakanishi, Yu-Te Wang, Tzyy-Ping Jung, John K Zao, Yu-Yi Chien, Alberto Diniz-Filho, Fabio B Daga, Yuan-Pin Lin, Yijun Wang, and Felipe A Medeiros. Detecting glaucoma with a portable brain-computer interface for objective assessment of visual function loss. *JAMA ophthalmology*, 135(6):550–557, 2017.

[14] Noémie Hébert-Lalonde, Lionel Carmant, Dima Safi, Marie-Sylvie Roy, Maryse Lassonde, and Dave Saint-Amour. A frequency-tagging electrophysiological method to identify central and peripheral visual field deficits. *Documenta Ophthalmologica*, 129(1):17–26, 2014.

[15] Rich Caruana. Multitask learning. *Machine learning*, 28(1):41–75, 1997.

[16] Yann LeCun, Yoshua Bengio, et al. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, 3361(10):1995, 1995.

[17] Masaki Nakanishi, Yijun Wang, Yu-Te Wang, and Tzyy-Ping Jung. A comparison study of canonical correlation analysis based methods for detecting steady-state visual evoked potentials. *PloS one*, 10(10), 2015.

[18] Nicholas Waytowich, Vernon J Lawhern, Javier O Garcia, Jennifer Cummings, Josef Faller, Paul Sajda, and Jean M Vettel. Compact convolutional neural networks for classification of asynchronous steady-state visual evoked potentials. *Journal of neural engineering*, 15(6):066031, 2018.

[19] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2017.

[20] Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. Wavenet: A generative model for raw audio. *arXiv preprint arXiv:1609.03499*, 2016.

[21] Matthias Holschneider, Richard Kronland-Martinet, Jean Morlet, and Ph Tchamitchian. A real-time algorithm for signal analysis with the help of the wavelet transform. In *Wavelets*, pages 286–297. Springer, 1990.

[22] Fisher Yu and Vladlen Koltun. Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*, 2015.

[23] Siavash Sakhavi, Cuntai Guan, and Shuicheng Yan. Learning temporal information for brain-computer interface using convolutional neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 29(11):5619–5629, 2018.

[24] Xiaogang Chen, Yijun Wang, Shangkai Gao, Tzyy-Ping Jung, and Xiaorong Gao. Filter bank canonical correlation analysis for implementing a high-speed ssvep-based brain–computer interface. *Journal of neural engineering*, 12(4):046008, 2015.

[25] Hirokazu Tanaka, Takusige Katura, and Hiroki Sato. Task-related component analysis for functional neuroimaging and application to near-infrared spectroscopy data. *NeuroImage*, 64:308–327, 2013.