A Lightweight Deep Learning Model for EEG Classification Across Visual Stimuli

Yi Liu
University of Southern Queensland
Toowoomba, Australia
Yi.liu@Unisq.edu.au

Steven Goh
University of Southern Queensland
Toowoomba, Australia
Steven.goh@Unisq.edu.au

Tobias Low
University of Southern Queensland
Toowoomba, Australia
Tobias.low@Unisq.edu.au

Zach Quince
University of Southern Queensland
Toowoomba, Australia
Zach.quince@Unisq.edu.au

Shoryu Teragawa
Dalian University of Technology
Dalian, China
frozen@mail.dlut.edu.cn

***Abstract*—Visual stimuli have a multifaceted impact on brain activity, yet the nuanced differences in how various types of stimuli affect electroencephalogram (EEG) signals are still under investigation. This study endeavors to classify EEG signals in response to a range of visual stimuli by crafting a lightweight deep learning model. Utilizing the N170 EEG dataset from the ERP core, which encompasses recordings from 40 healthy participants exposed to roughly 10-minute sessions of randomly presented sets of normal and scrambled photographs. Each set consisted of images portraying either normal or scrambled representations of faces and cars, encapsulating four unique visual stimuli. By harnessing the EEG data from the 40 participants, our ResNet18-based model attained an impressive average classification accuracy of 98.13% for face images and 97.81% for car images, significantly surpassing the performance of traditional machine learning models. Notably, this marks the inaugural application of the ResNet18 model to the N170 dataset classification experiment within the ERP core. The findings of this study enrich our comprehension of the brain's distinct cognitive responses to these stimuli and the manifestation of these differences in EEG signals. The successful deployment of this model paves the way for furthering the exploration and development of brain-computer interface technologies.**

***Keywords*—EEG, Deep learning, Classification, Visual stimuli.**

I. Introduction

Electroencephalogram (EEG), with its origins dating back to the 19th century, was initially observed by Richard Caton in 1875 through the recording of subtle physiological electrical activity in exposed brains of animals, particularly dogs and monkeys [1]. In the realm of neuroscience, EEG stands as a fundamental tool for elucidating brain functions. Its waveform patterns, upon analysis, facilitate the establishment of correlations between EEG readings and the reception of visual information, as well as human behavioral patterns [2]. As humans undergo specific training and acquire new skills, their brains exhibit corresponding functional or structural modifications, a phenomenon termed brain plasticity [3]. Within the intricate realm of brain signaling, lies the potential to reconstruct stimulus contours.

Contemporary research concerning the impact of external stimuli on EEG changes primarily focuses on two key areas. First, it delves into studying alterations in EEG patterns associated with the onset and progression of neurodegenerative conditions like stroke, epilepsy, dementia, among others [4]. Second, it involves the application of fixed-pattern training to EEG data. This method aims to elicit specific EEG components by exposing individuals to particular stimuli, thereby enabling the exploration of Event-Related Potentials (ERPs) and leveraging these components in tasks involving human-machine cooperation under Brain-Computer Interface (BCI) systems [5].

In the realm of contemporary neuroscience research, understanding the intricate relationship between external stimuli and EEG responses remains a pivotal yet challenging endeavor. Human visual cognition undeniably excels in discerning and interpreting visual stimuli, exhibiting remarkable prowess in distinguishing between clear, structured images and cluttered, ambiguous ones. However, while human visual perception serves as an exemplar of efficient processing, the translation of this complex cognitive process into computational paradigms poses considerable challenges. Leveraging computational capabilities to effectively and accurately classify EEG signals across diverse stimulus types, therefore, emerges as a crucial frontier in contemporary neuroscience research. The ability to decode and interpret EEG signals corresponding to various visual stimuli represents a critical step towards uncovering the underlying mechanisms of human cognition, perception, and information processing.

II. Related Works

The study of electroencephalography (EEG) has a rich historical backdrop. In 1897, Adolf Beck conducted pioneering experiments involving diverse stimuli on monkeys, successfully capturing distinctive forms of electrical activity on the cortical surface of the brain corresponding to different stimuli [6][7]. Subsequently, in 1914, British scientists utilized EEG recordings from animals to document changes during epileptic seizures, uncovering the existence of action potentials [8]. Similarly, in 1951, neurological experts observed analogous responses in human EEG recordings, correlating to various rhythmic stimuli, thus unveiling Event-Related Potentials (ERPs) [9]. In the domain of visual stimuli, since the early 21st century, researchers have utilized EEG signals to delve into cognition-based automated methods for investigating visual classification tasks [10]. Studies have preliminary successfully decoded human visual cognitive information from EEG signals [11]. From a time series perspective, human brain activity induced by visual stimuli unfolds in distinct stages aligning with various cognitive processes [12]. Particularly noteworthy is the discovery that at different event-related sites, the human brain exhibits a more pronounced negative peak in response to faces compared to other object categories at the N170 stage [13]. These findings serve as foundational pillars for constructing classification systems grounded in EEG signals.

In 2006, research from Columbia University's Laboratory of Intelligent Imaging and Neural Computing unveiled significant differences in EEG signals between facial and automobile image stimuli at 170ms and 330ms post-stimulus. This discovery led to the proposition of a non-invasive neuroscientific method based on EEG signals for perceptual decision-making, successfully distinguishing faces and automobiles [14]. Subsequent studies from the Biotechnology Collaboration Research Institute at the University of California, Santa Barbara, reaffirmed EEG's classification prowess for object recognition tasks (faces/cars). These studies identified distinct N100 and N170 components evoked by the face/car paradigm, achieving robust binary classification performance [15]. Meanwhile, researchers from the University of Trento in Italy utilized spectral features of EEG signals to predict three categories—animals, plants, and tools—with an average accuracy of 80% [16]. At the 2008 International Conference on Computer Vision and Pattern Recognition (CVPR), a collaborative study between the University of Washington and Microsoft Research showcased a groundbreaking integration of brain EEG signal analysis with a visual category recognition system based on the Pyramid Match Kernel (PMK). Impressively, it achieved over 91% accuracy in a three-category classification task encompassing animals, non-biological objects, and human faces [17]. Moreover, during a conference on human-computer interaction, researchers revealed a study combining spatial projection algorithms [18] and Regularized Linear Discriminant Analysis (RLDA) for EEG signal analysis. By leveraging averaged Event-Related Potential (ERP) responses to improve the EEG signal-to-noise ratio, this method attained over 90% accuracy in classifying four categories (faces, vehicles, animals, and buildings), marking a significant leap in visual classification using EEG signals [19].

In 2015, Stanford University's Center for Language and Information shared an EEG dataset featuring signals from 10 participants. This dataset encompassed 72 images across six categories as stimulus sources. Statistical analysis methods achieved an 81.06% accuracy in binary face and object classification and a 40.68% accuracy in six-category recognition [20]. Additionally, researchers from the University of Lebanon conducted preliminary studies utilizing high-density EEG signals from 256 leads, successfully distinguishing animal and non-animal image categories from EEG signals with an 82.7% classification accuracy [21]. In 2017, the Perception, Research, and Cognition Lab at the University of Central Florida curated an EEG dataset reflecting human brain responses to object recognition information. This dataset involved a remarkable 40-class classification task, marking a significant stride in the quantity of classification. They devised an automated visual classification approach employing recurrent neural networks to learn brain activity manifold for specific visual object categories. Leveraging the learned visual category representations for classifying diverse category-specific EEG data, they achieved an average accuracy of 83% [22]. In 2018, Zhong et al. [23] leveraged an enhanced LSTM (eLSTM) methodology on the ImageNet-EEG dataset, aiming to diminish the reliance on high-density EEG arrays, thereby achieving a remarkable classification accuracy of 96.2% across forty categories. To exploit the temporal variability in visual features, Jiao et al. [24] compared the performance of Linear Discriminant Analysis (LDA), LSTM, and CNN on the same dataset, with CNNs achieving the highest classification accuracy of 83.10% in 2019. They further refined the classification models and incorporated a CNN-based visual guidance mechanism, elevating the accuracy to 85.50%. Mukherjee et al. [25] introduced a sophisticated deep recurrent architecture comprising Bi-LSTM units on the ImageNet-EEG dataset. They employed knowledge distillation techniques for training the deep cognitive model CogniNet, achieving an apex accuracy of 89.6%. In 2020, Cudlenco et al. [26] meticulously curated a dataset from the publicly accessible ImageNet repository, augmented by their personally captured photographs, to compile six distinct categories of visual stimuli: flora, aircraft, automobiles, parks, beaches, and urban environments. The researchers meticulously recorded the electroencephalogram (EEG) signals elicited from volunteers in response to these varied visual stimuli. Subsequently, they embarked on a comprehensive analysis and comparison of the efficacy of several rudimentary classifiers traditionally employed in such tasks, including Ridge Regression (RR), Gabor filters, Convolutional Neural Networks (CNN), and Long Short-Term Memory (LSTM) networks. Despite the conventional wisdom that CNNs are predominantly suited for image data, they demonstrated commendable performance in interpreting EEG signals [27]. Zheng et al. [28] proposed an end-to-end attention-based Bi-LSTM-AttGW method, achieving an unprecedented accuracy of 99.50% on the ImageNet-EEG dataset. Despite the plethora of research utilizing the ImageNet-EEG dataset for classification purposes, with accuracies approaching theoretical limits, the dataset has recently been critiqued for its experimental design flaws. Stimuli presented in different blocks belonged to varying categories, leading to artificially inflated classification accuracies [22]. Despite considerable advancements in the classification of EEG signals elicited by visual stimuli, notable challenges remain. First and foremost, the accuracy of existing models leaves ample scope for enhancement. Moreover, the complexity of the models developed thus far is significant, posing challenges to the efficient utilization of computational resources. Additionally, inherent issues within the design of the ImageNet-EEG dataset raise concerns regarding the reliability of research built upon this foundation.

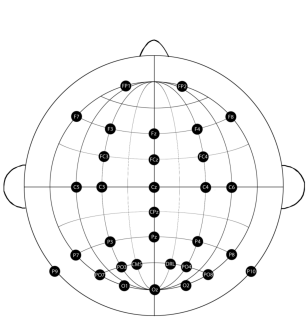
In response, this paper introduces a streamlined deep learning model tailored to a reputable dataset. This model is designed for the swift and precise classification of EEG signals across a spectrum of stimuli, promising to significantly influence future investigative endeavors.

III. Methods

A. Dataset

The data utilized in this study originate from the public dataset ERP Core [29], specifically aimed at evoking N170 in facial perception tasks. Access to this dataset is freely available at the URL: https://doi.org/10.18115/D5JW4R. This experiment involved 40 participants from the community of the University of California, Davis (25 females, 15 males; mean age = 21.5 years, SD = 2.87, range 18-30; 38 right-handed individuals). Each participant had native English proficiency, normal color perception, normal or corrected-to-normal vision, and no history of neurological injury or illnesses (as self-reported). Approval for this study was obtained from the Institutional Review Board at the University of California, Davis, and all participants provided informed consent.

The testing occurred in a dimly lit, sound-attenuated, and electrically shielded laboratory. Visual stimuli were presented to participants on a Hewlett-Packard ZR2440w liquid crystal display monitor with a resolution of 1280 × 1024, a refresh rate of 60 Hz, and a viewing distance of 100 centimeters. The stimuli were displayed on a medium-gray background (x = 0.35, y = 0.36, 25.9 cd/m2). The continuous EEG was captured utilizing a Biosemi ActiveTwo recording system equipped with active electrodes (Biosemi B.V., Amsterdam, the Netherlands). Recordings were obtained from 30 scalp electrodes, arranged within an elastic cap conforming to the International 10/20 System (FP1, F3, F7, FC3, C3, C5, P3, P7, P9, PO7, PO3, O1, Oz, Pz, CPz, FP2, Fz, F4, F8, FC4, FCz, Cz, C4, C6, P4, P8, P10, PO8, PO4, O2, as Figure 1), at a sampling rate of 1024Hz.



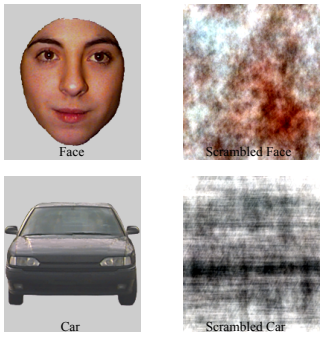


Fig. 1 (left). The position of the electrodes on the EEG recording instrument worn by the participants.

Fig. 2 (right). Sample pictures of each of the four categories.

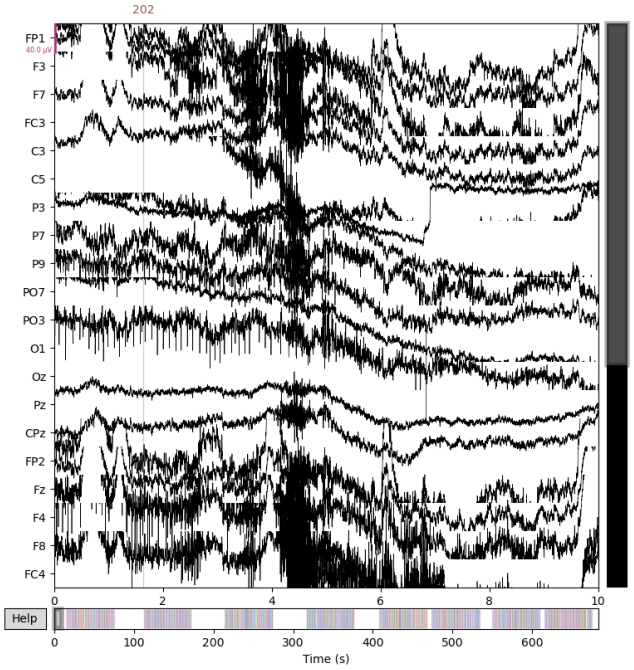


Fig. 3. Sub-001 raw EEG signal.

During each trial, participant completed viewing of 320 images presented in a random order across the four categories. Each category comprised 40 samples, each of which was presented twice. Example pictures of each category can be seen in Figure 2.

B. *Signal preprocessing*

The raw data consists of approximately 600-second EEG signals, interspersed with 320 nodes representing stimulus events occurring as images. The initial step involves segmenting the signals based on these events, aimed at obtaining segmented signals corresponding to the events. Figure 3 depicts the raw signal prior to segmentation, and Figure 4 displays the segmented signal corresponding to the first event (triggered by stimuli-79). The pronounced discrepancy between the two images is attributable to the presence of substantial noise in the signal.

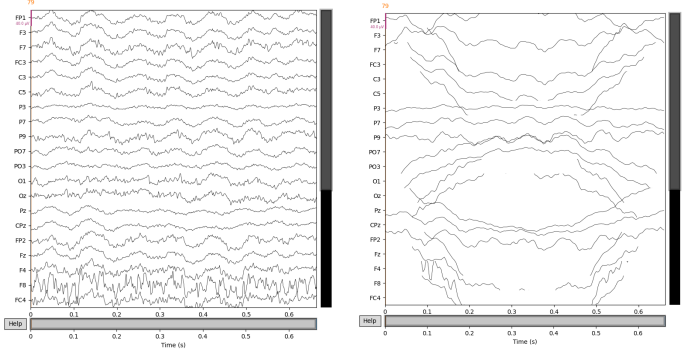


Fig. 4 (left). The EEG signal triggered by visual stimuli during the viewing of image number 79 by participant sub-001.

Fig. 5 (right). The signal in Figure 4 becomes purified after filtering and ICA analysis.

The original EEG data is often affected by various sources of noise, including muscle movements, electrode artifacts, and environmental interferences, which can disrupt the accurate analysis of brain signals. Filters play a crucial role in EEG preprocessing. The 0.5 to 70 Hz filter range corresponds to the common frequency range for most brain signal activities. Low-frequency signals (below 0.5 Hz) typically contain information related to basic physiological rhythms and subconscious processes, whereas high-frequency signals (above 70 Hz) are usually associated with muscle activity or electrode noise. Consequently, limiting the signal within the 0.5 to 70 Hz range aids in removing extraneous noise, preserving information relevant to the brain activity of interest. In addition to filters, Independent Component Analysis(ICA) [30] is a commonly employed technique in EEG preprocessing. ICA can decompose mixed signals into multiple independent components statistically. In EEG signal processing, ICA effectively distinguishes differences between brain signals and noise components. By identifying and removing these noise components, ICA assists in extracting pure brain signals, thereby enhancing the accuracy and reliability of subsequent analyses. Figure 5 shows the signals after filtering and ICA analysis. This approach efficiently eliminates noise, providing a reliable foundation for subsequent studies in neuroscience.

C. *Deep learning model*

The fundamental model utilized in this study is the ResNet-18 architecture, a relatively shallow network within the Residual Neural Network (ResNet) series [31], comprising 18 layers primarily constituted by fundamental residual blocks. This model utilized a pre-trained ResNet neural network to process segmented and denoised data into class categories

within a customized dataset class called 'CustomDataset'. The training dataset was organized into a DataLoader with a batch size of 32, indicating the model simultaneously processed 32 images per training iteration. Similarly, the validation dataset was configured as a DataLoader with a batch size of 32, facilitating model evaluation after each training epoch. For the testing dataset, a DataLoader with a batch size of 1 was employed, allowing the model to make predictions on individual images sequentially, enabling the computation of overall test accuracy and specific class predictions for each image. The PyTorch DataLoader class was employed in this code segment to handle data batch processing during training, validation, and testing phases. The significance of these varied batch sizes during training, validation, and testing procedures cannot be overstated. Larger training batches (32) generally contribute to faster training speeds as the model can concurrently process more images. Conversely, smaller testing batches (1) permit evaluations on individual samples, providing more accurate predictions for the entire test dataset. This batch size selection balances computational efficiency with the accuracy of model performance evaluation throughout training, validation, and testing phases.

The training process incorporated the use of a cross-entropy loss function and the Adam optimizer [32]. During training, the model iteratively optimized parameters by traversing the training dataset to minimize the loss function. Subsequently, in the validation phase, model performance was assessed on the validation set, computing accuracy metrics. Finally, in the testing phase, model performance was evaluated on the testing set, yielding the overall test accuracy and predicted classes for each image. This ResNet18-based model was designed for the classification of facial images into two categories, based on image clarity. Following training and validation, the model demonstrated its ability to categorize facial images with notable accuracy on the testing set. Figure 6 shows the structure of the model.

The convolutional neural network (CNN) architecture presented in Figure 6 encompasses various layers, including Conv2d, BatchNorm2d, ReLU activation, MaxPool2d, BasicBlock, and Linear layers. The model initiates with a Conv2d layer (kernel size of 7×7 and stride of 2), yielding an output shape of [-1, 64, 400, 400] with 9,408 parameters. Following the convolutional layer, BatchNorm2d and ReLU activation are applied.

Subsequently, a MaxPool2d layer downsamples the feature maps to an output shape of [-1, 64, 200, 200]. The network comprises multiple blocks, each consisting of two Conv2d layers (kernel size of 3×3), followed by BatchNorm2d and ReLU activation. These blocks, termed BasicBlock, contribute significantly to enhancing the model's depth and feature extraction capabilities. As the architecture progresses, it incorporates Conv2d layers with diverse output channels (64, 128, 256, 512) along with their respective BatchNorm2d and ReLU activation functions. The final layers include AdaptiveAvgPool2d, reshaping the tensor to [-1, 512, 1, 1], and a Linear layer with 2 output units, facilitating the ultimate classification into 2 categories. The Linear layer comprises 1,026 parameters. In summary, this model follows the architecture of a deep residual network (ResNet), utilizing skip connections via the BasicBlock modules, enabling effective learning and extraction of hierarchical features from input images. The network culminates in a linear classification layer utilized for final predictions. This architecture demonstrates robust feature extraction mechanisms tailored for image classification tasks. Its complex arrangement of convolutional and residual layers enhances the learning and representation of visual information.

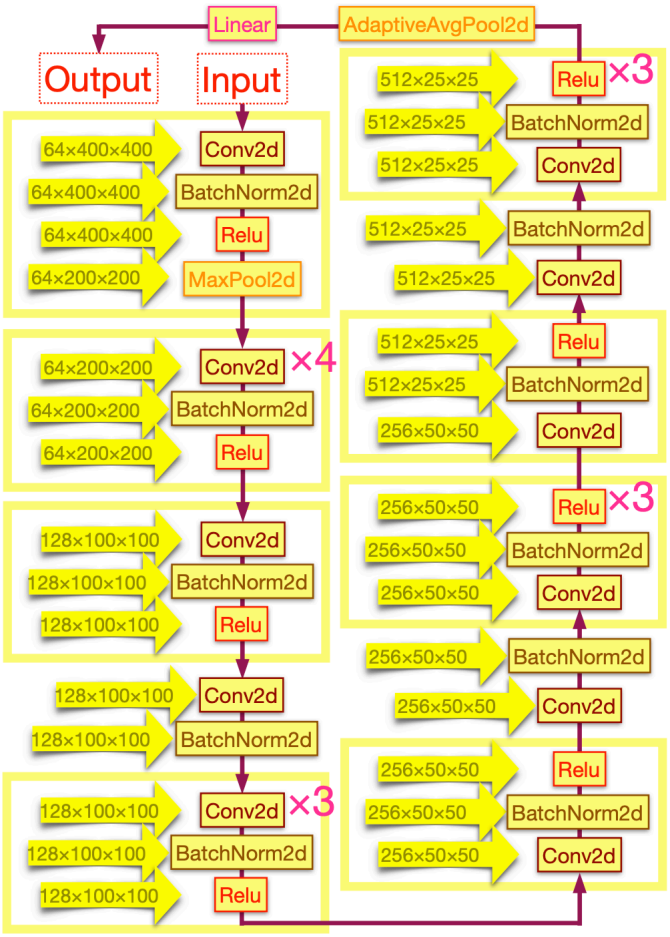


Fig.6. The structure of the model.

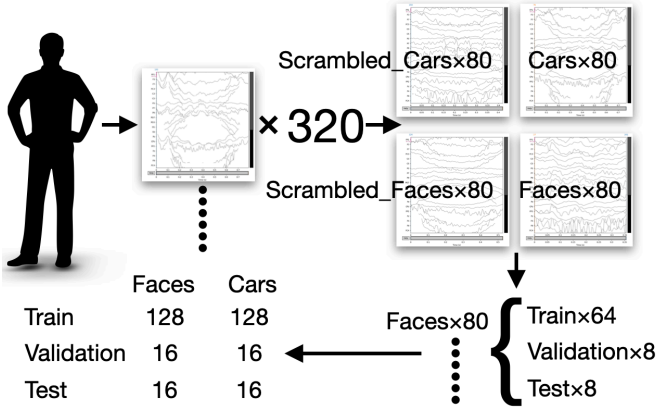


Fig.7. The schematic representation of the partitioning of the dataset into training, validation, and testing sets. Maintaining a balance between positive and negative examples in a dataset is crucial for ensuring robust model performance, enhancing generalization to new data, and upholding fairness in decision-making processes. Balanced datasets prevent model bias towards the majority class, enabling more accurate and reliable predictions across all categories. This balance also simplifies the evaluation of model performance using common metrics and facilitates the use of a wider range of machine learning algorithms.

D. Training set, validation set and test set

The dataset employed comprises 320 pre-categorized electroencephalogram (EEG) images, categorized based on clarity into two groups: 'normal' and 'scrambled', as well as based on objects into two groups: 'faces' and 'cars'. Each participant engages in two category classification tests: one involving distinguishing between faces and scrambled faces, and the other involving distinguishing between cars and scrambled cars. From the 80 images per category, a random selection of 8 images is allocated for validation, another 8 for testing purposes, while the remaining 64 images constitute the training set for each classification group. Despite the relatively modest volume of the dataset employed in each individual trial, the study significantly enhances its reliability by replicating the experiment 40 times, utilizing a unique dataset for each iteration. Figure 7 illustrates the schematic representation of the partitioning of the dataset into training, validation, and testing sets.

V. Result and Discussion

Based on the EEG of the first participant (sub-001), figure 8 shows the iterative graphs of accuracy and loss values of faces classification and figure 9 shows the iterative graphs of accuracy and loss values of cars classification.

Overall, the model proposed in this study exhibited commendable performance across the 40 subjects evaluated. In the task of recognizing facial photographs, the average accuracy attained was 98.13%, with an average precision of 99.72%, average recall of 96.56% and an average F1-score of 97.50%. In comparison to facial recognition, the model achieved an average accuracy of 97.81%, an average precision of 99.44%, an average recall of 96.25%, and an average F1-score of 97.18% in the task of identifying cars. An accuracy rate approaching 100% indicates that the model exhibits near-perfect performance in the classification task on this dataset.

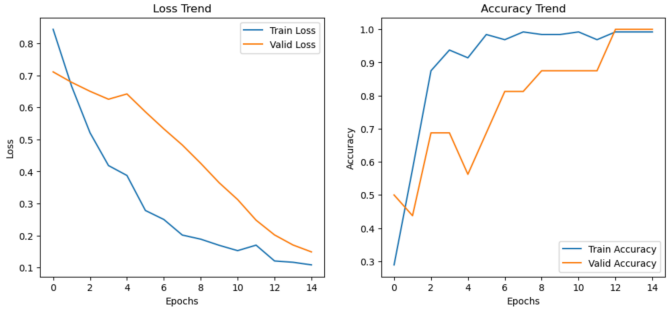


Fig.8. Accuracy and loss curves of faces classification in sub-001.

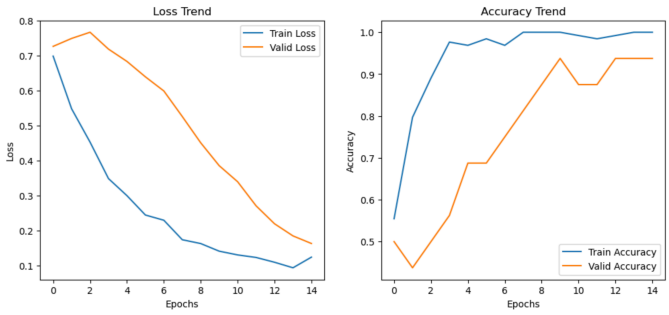


Fig.9. Accuracy and loss curves of cars classification in sub-001.

TABLE I
THE PERFORMANCE OF EEG IN SUB-001

Objective	*Accuracy*	*Precision*	*Recall*	*F1-score*
Faces/ Scrambled faces	100.00%	100.00%	100.00%	100.00%
Cars/Scrambled cars	100.00%	100.00%	100.00%	100.00%

Based on the EEG of the first participant (sub-001), table I shows the results of the experiment. All predictions are correct.

To highlight the advantages of the proposed model, this study conducted a comparative analysis employing two conventional machine learning methodologies. Initially, the Support Vector Machine (SVM) approach yielded an average accuracy of 81.25%, an average precision of 82.97%, an average F1-score of 81.02%, and an average recall of 85.63%. While the performance of this model surpasses that achievable through direct human observation, it is indispensable for advancing our understanding of the variations in human eye electroencephalogram responses under diverse stimuli. Given its accuracy approaching 100%, this model is exceptionally well-suited for in-depth exploration in this domain.

The model demonstrated exceptionally high average accuracy rates in two classification tasks, facial recognition and vehicle identification, achieving 98.13% and 97.81% respectively, along with notably high average precision, recall, and F1 scores. This indicates the model's high reliability and accuracy for classification tasks within this dataset. Compared to traditional machine learning models such as Support Vector Machines (SVM), the superior accuracy in EEG signal classification suggests potential practical applications of this model in fields such as neuroscience research and Brain-Computer Interface (BCI) technologies, particularly in understanding how the brain processes different types of visual information.

IV. Conclusion

The advent of deep neural network technologies has significantly alleviated the necessity for manual feature extraction, demonstrating notable advantages in analyzing unstructured data. Remarkable performance achievements have been observed, particularly in domains such as computer vision and speech recognition [33]. Specifically, the escalating utilization of deep convolutional networks has effectively addressed numerous challenging image classification tasks [34, 35, 36], surpassing traditional machine learning methods that relied on manually curated feature selection for classification and identification [37]. The deep learning approach proposed in this study significantly contributes to the exploration of neural activity associated with visual stimuli by achieving high accuracy in classifying EEG data with the powerful ResNet18 architecture. Its capability to differentiate between EEG patterns triggered by clear and scrambled images within a concise training duration opens avenues for comprehensive investigations into how the brain responds to different visual stimuli. Consequently, this lightweight model serves as a valuable tool for further studies aiming to unravel the complex interplay between visual perception and neural processes.

References

[1] R. Caton, 'The Electric Currents of the Brain', American Journal of EEG Technology, vol. 10, no. 1, pp. 12–14, Mar. 1970.

[2] W. Klimesch, 'EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis', Brain Research Reviews, vol. 29, no. 2, pp. 169–195, Apr. 1999.

[3] J. D. Schaechter, 'Motor rehabilitation and brain plasticity after hemiparetic stroke', *Prog Neurobiol*, vol. 73, no. 1, pp. 61–72, May 2004.

[4] A. Horvath, A. Szucs, G. Csukly, A. Sakovics, G. Stefanics, and A. Kamondi, 'EEG and ERP biomarkers of Alzheimer's disease: a critical review', *Front Biosci (Landmark Ed)*, vol. 23, no. 2, pp. 183–220, Jan. 2018.

[5] A. Halpern, J. Martin, and T. Davis, 'An ERP Study of Major-Minor Classification in Melodies', *Music Perception*, vol. 25, pp. 181–191, Feb. 2008.

[6] A. Coenen, E. Fine, and O. Zayachkivska, 'Adolf Beck: a forgotten pioneer in electroencephalography', *J Hist Neurosci*, vol. 23, no. 3, pp. 276–286, 2014.

[7] W. Feichtinger, S. Szalay, A. Beck, P. Kemeter, and H. Janisch, 'Results of Laparoscopic Recovery of Preovulatory Human Oocytes From Nonstimulated Ovaries in an Ongoing in Vitro Fertilization Program', *Fertility and Sterility*, vol. 36, no. 6, pp. 707–711, Dec. 1981.

[8] E. Magiorkinis, A. Diamantis, K. Sidiropoulou, and C. Panteliadis, 'Highights in the history of epilepsy: the last 200 years', *Epilepsy Res Treat*, vol. 2014, p. 582039, 2014.

[9] J. M. Clark, 'Modularity, abstractness and the interactive brain', *Behav Brain Sci*, vol. 17, no. 1, pp. 67–68, Mar. 1994.

[10] M. G. Philiastides, R. Ratcliff, and P. Sajda, 'Neural Representation of Task Difficulty and Decision Making during Perceptual Categorization: A Timing Diagram', *J. Neurosci.*, vol. 26, no. 35, pp. 8965–8975, Aug. 2006.

[11] R. VanRullen and S. J. Thorpe, 'The time course of visual processing: from early perception to decision-making', *J Cogn Neurosci*, vol. 13, no. 4, pp. 454–461, May 2001.

[12] J. Liu, A. Harris, and N. Kanwisher, 'Stages of processing in face perception: an MEG study', *Nat Neurosci*, vol. 5, no. 9, Art. no. 9, Sep. 2002.

[13] G. Thierry, C. D. Martin, P. Downing, and A. J. Pegna, 'Controlling for interstimulus perceptual variance abolishes N170 face selectivity', *Nat Neurosci*, vol. 10, no. 4, pp. 505–511, Apr. 2007.

[14] M. G. Philiastides and P. Sajda, 'Temporal characterization of the neural correlates of perceptual decision making in the human brain', *Cereb Cortex*, vol. 16, no. 4, pp. 509–518, Apr. 2006.

[15] K. Das, B. Giesbrecht, and M. P. Eckstein, 'Predicting variations of perceptual performance across individuals from neural activity using pattern classifiers', *NeuroImage*, vol. 51, no. 4, pp. 1425–1437, Jul. 2010.

[16] B. Murphy, 'Distinguishing Concept Categories from Single-Trial Electrophysiological Activity', Jan. 2008.

[17] A. Kapoor, P. Shenoy, and D. Tan, 'Combining brain computer interfaces with vision for object categorization', in 2008 IEEE Conference on Computer Vision and Pattern Recognition, Jun. 2008, pp. 1–8.

[18] U. Hoffmann, J.-M. Vesin, and T. Ebrahimi, 'Spatial filters for the classification of event-related potentials', presented at the European Symposium on Artificial Neural Networks (ESANN 2006), Jan. 2006, pp. 47–52.

[19] P. Shenoy and D. S. Tan, 'Human-aided computing: utilizing implicit human processing to classify images', in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, in CHI '08. New York, NY, USA: Association for Computing Machinery, Apr. 2008, pp. 845–854.

[20] B. Kaneshiro, M. Perreau Guimaraes, H.-S. Kim, A. M. Norcia, and P. Suppes, 'A Representational Similarity Analysis of the Dynamics of Object Processing Using Single-Trial EEG Classification', *PLoS One*, vol. 10, no. 8, p. e0135697, 2015

[21] R. El-Lone, M. Hassan, A. Kabbara, and R. Hleiss, 'Visual objects categorization using dense EEG: A preliminary study', in 2015 International Conference on Advances in Biomedical Engineering (ICABME), Sep. 2015, pp. 115–118.

[22] C. Spampinato, S. Palazzo, I. Kavasidis, D. Giordano, M. Shah, and N. Souly, 'Deep Learning Human Mind for Automated Visual Classification'. arXiv, Oct. 22, 2019.

[23]S. Zhong, Y. Liu, Z. Zhou, and D. Hu, 'ELSTM-Based Visual Decoding from Singal-Trial EEG Recording', in 2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS), Nov. 2018, pp. 1139–1142.

[24] Z. Jiao, H. You, F. Yang, X. Li, and H. Zhang, Decoding EEG by Visual-guided Deep Neural Networks. 2019, p. 1393.

[25] P. Mukherjee, A. Das, A. K. Bhunia, and P. P. Roy, 'Cogni-Net: Cognitive Feature Learning Through Deep Visual Perception', in 2019 IEEE International Conference on Image Processing (ICIP), Sep. 2019, pp. 4539–4543.

[26] N. Cudlenco, N. Popescu, and M. Leordeanu, 'Reading into the mind's eye: Boosting automatic visual recognition with EEG signals', Neurocomputing, vol. 386, pp. 281–292, Apr. 2020.

[27] M. Nour, Ş. Öztürk, and K. Polat, 'A novel classification framework using multiple bandwidth method with optimized CNN for brain–computer interfaces with EEG-fNIRS signals', Neural Comput & Applic, vol. 33, no. 22, pp. 15815–15829, Nov. 2021.

[28] X. Zheng and W. Chen, 'An Attention-based Bi-LSTM Method for Visual Object Classification via EEG', Biomedical Signal Processing and Control, vol. 63, p. 102174, Jan. 2021.

[29] E. S. Kappenman, J. L. Farrens, W. Zhang, A. X. Stewart, and S. J. Luck, 'ERP CORE: An open resource for human event-related potential research', NeuroImage, vol. 225, p. 117465, Jan. 2021.

[30] A. Hyvärinen and E. Oja, 'A Fast Fixed-Point Algorithm for Independent Component Analysis', Neural Computation, vol. 9, no. 7, pp. 1483–1492, Jul. 1997.

[31] K. He, X. Zhang, S. Ren, and J. Sun, 'Deep Residual Learning for Image Recognition'. arXiv, Dec. 10, 2015.

[32] D. P. Kingma and J. Ba, 'Adam: A Method for Stochastic Optimization'. arXiv, Jan. 29, 2017.

[33] G. Hinton et al., 'Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups', IEEE Signal Processing Magazine, vol. 29, no. 6, pp. 82–97, Nov. 2012.

[34] A. Krizhevsky, I. Sutskever, and G. E. Hinton, 'ImageNet classification with deep convolutional neural networks', Commun. ACM, vol. 60, no. 6, pp. 84–90, May 2017.

[35] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, 'Densely Connected Convolutional Networks'. arXiv, Jan. 28, 2018.

[36] K. Simonyan and A. Zisserman, 'Very Deep Convolutional Networks for Large-Scale Image Recognition'. arXiv, Apr. 10, 2015.

[37] J. Schmidhuber, 'Deep learning in neural networks: An overview', Neural Networks, vol. 61, pp. 85–117, Jan. 2015.