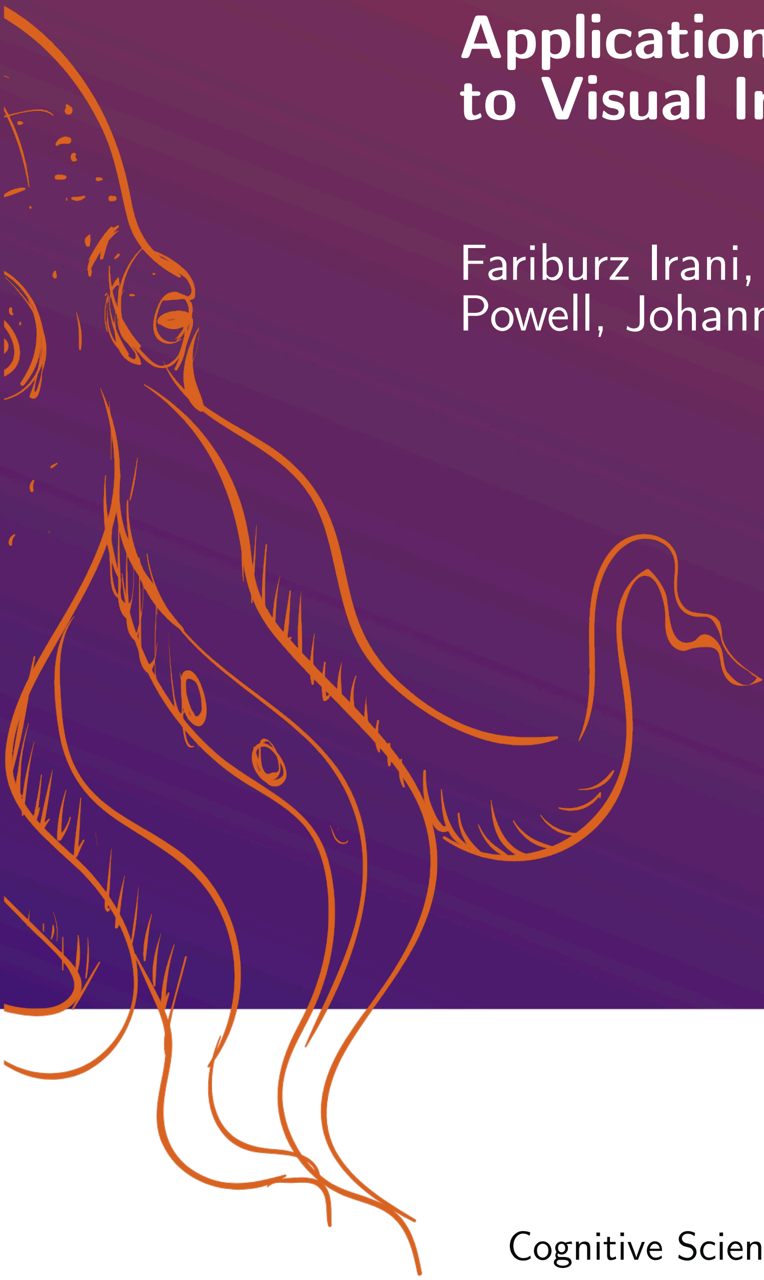
COGNITIVE SCIENCE STUDENT JOURNAL

Cognitive Science Student Journal
Osnabrück University

Do You See What I See? - Application of Neural Networks to Visual Imagery

Fariburz Irani, Febryeric Parantean, Piper Powell, Johanna Venzke, Junpei Zhong



Cognitive Science Student Journal 2023, Number 1

COGNITIVE SCIENCE STUDENT JOURNAL

Irani, F., Parantean, F. Powell, P., Venzke, J., & Zhong, J. (2023). Do you see what i see? - Application of neural networks to visual imagery. *Cognitive Science Student Journal 2023, 1*. 1-8.

Institut für Kognitionswissenschaft
Universität Osnabrück
49069 Osnabrück
Germany
https://www.ikw.uni-osnabrueck.de

Storage and cataloging done by Osnabrück University

UNIVERSITÄT OSNABRÜCK

The project is funded by *Stiftung Innovation in der Hochschullehre*

Stiftung
Innovation in der
Hochschullehre

Abstract

This paper discusses the application of four types of neural networks – standard artificial neural networks (ANNs), convolutional neural networks (CNNs), recurrent neural networks (RNNs), and spiking neural networks (SNNs) to visual imagery tasks. It aims to provide an overview of visual imagery and neural network approaches to provide a solid starting point for future work at the intersection of these fields. Visual imagery involves a user of a brain-computer interface (BCI) system controlling that BCI by imagining specific images that the system then deciphers from their neurological data (usually electroencephalographic data – EEG) as commands. The paper begins with a summary of each neural network type discussed and provides an overview on the background of visual imagery tasks in general. The application of each network type to visual imagery is then discussed, with attention paid to their relative advantages and disadvantages in general and specifically in that application. The paper concludes with a summary of findings and a general recommendation regarding which network type to use. We conclude that CNNs are ill-suited for visual imagery tasks, but that ANNs, RNNs, and SNNs are all valid options, with SNNs (once further developed) being the frontrunner.

1 Introduction

Deciphering neurological data related to specific imagined "commands" by a user is a key method of communication with Brain-Computer Interfaces (BCIs). Many BCIs operate by deciphering data related to movements the user imagines enacting, a paradigm called Motor Imagery (MI) (Virgilio G. et al., 2020). Lately, however, a new method of imagination-based control has garnered interest – visual imagery. This newer method involves a user imagining one of a small set of images and the BCI analyzing their neurological data (most often electroencephalographic - EEG - data) to decide which they are currently thinking of in order to "hear" the appropriate command (Kosmyna et al., 2018). Like MI, this presents a classic classification task. Traditional machine learning methods such as support vector machines (SVMs) and K-Nearest-Neighbors (KNN) classifiers are often used in tasks of this sort, but the application of artificial neural networks has also been gaining ground in recent years (Luo et al., 2020; Virgilio G. et al., 2020). With the visual imagery field still relatively new and the field of artificial neural networks broad and in many places new itself, it is increasingly important to examine the relative advantages and disadvantages of the various network types to the kind of classification task presented by visual imagery in order to better inform future studies at the intersection of these fields. This will be the aim of this paper, which will first provide an overview of the network types and then discuss their application to visual imagery.

2 Network Types

Before discussing their application to visual imagery tasks, we first provide a brief overview of the four neural network types we will be discussing – classic artificial neural networks (ANNs), convolutional neural networks (CNNs), recurrent neural networks (RNNs), and the latest addition to the neural network family, spiking neural networks (SNNs). Note that in older literature, and in some modern work, classic ANNs are also referred to as multi-layer perceptrons (MLPs) or simply as neural networks (NNs).

2.1 Artificial Neural Networks (ANNs)

At its most basic level, an ANN is a computational model which mirrors the structure and connections of the human neuron system. Neurons in a real brain are represented as nodes in an ANN. Each node aggregates data from previous nodes (or from the original input itself) and performs simple operations on that data before passing it on to the next linked node, provided that the combined input to the node is high enough to cross a firing threshold (analogous to thresholds in real neurons). Nodes are arranged in layers, with the first being the input layer and the final being the output layer. Between these two layers are processing layers, called hidden layers, of which there can be as few as one or as many as deemed necessary for the present task. How the nodes are connected and the weights of those connections can be modified to govern how the neural network handles the data it is fed, optimizing it for the task at hand (Chai et al., 2012; Hramov et al., 2017).

2.2 Convolutional Neural Networks (CNNs)

Convolutional neural networks are a type of ANN in which early layers identify less complex patterns which are used in later layers to process and construct increasingly complex ones. Three different kinds of layers are typically seen in a CNN – convolutional, pooling, and fully connected. A convolutional layer, which the first layer typically is, computes the dot product of a filter kernel (a two-dimensional array of weights) and a data array (in the first layer, usually representing an image) to extract specific target features. Subsequent pooling layers reduce the model's complexity, often through subsampling and similar processes that output a reduced, more summarized representation of the data up to that layer. Two types of pooling layers are typically used – the Max Pooling layer, which returns the maximum value of the data covered by a kernel, and the Average Pooling layer, which returns the average of all the values of the data covered by the kernel. A fully connected layer(s) is typically placed at the end of the layer chain and often handles the majority of the classification work in a CNN, calculating the likelihood that the output (now in the form of extracted features) of earlier layers belongs to a certain class. While the first layer is usually a convolutional layer and the last layer a fully-connected layer, the intermediate layers can be in any arrangement and order of the layer types needed. Note that some layers of a CNN, such as the final fully connected layer, can often take the form of a basic ANN (Bagchi & Bathula, 2022; Kwon et al., 2020).

2.3 Recurrent Neural Networks (RNNs)

RNNs are also a kind of ANN whose hallmark is the addition of recurrence, which gives these networks a form of memory and also enables them to work on sequences of data (such as data with a temporal dimension). Whereas in standard ANNs activity proceeds from the input to the output in one direction, called "feeding forward" outputs in an RNN can be fed back into the model to inform earlier steps. This may occur between layers or on the level of the full model, in which the output of the model on the previous run is fed into the model as part of its input for the next run. This allows RNNs to have a form of memory in that they retain all previous information until the next step or time series. The incorporation of previous outputs into new inputs allows an RNN's "decisions" to consider not only the current input from the current step in the input sequence but also what it has learned from previous inputs as well. Various types of RNNs exist depending on the number of inputs and outputs involved, with the most basic form being the one-to-one model with only one input and only one output (Ma et al., 2018; Qiao et al., 2019).

2.4 Spiking Neural Networks (SNNs)

SNNs are the latest generation of ANNs, and the generation which most closely reflects the working of a real brain. Unlike earlier network types, these networks accept input data and produce output data only in the form of spike trains, which are similar to the action potentials used by real neurons. SNNs also lack the fixed layer format of earlier network types and they do not transmit data on a cyclical basis, but rather in response to discrete events occurring at specific time points (Paugam-Moisy & Bohte, 2012). The neurons that make up the model may be arranged in a cortex-like structure, such as in the Neu- Cube architecture (Kasabov, 2014), with the modeled neurons most commonly threshold-based (e.g. Leaky Integrate and Fire) models (Paugam-Moisy & Bohte, 2012). Because input data for SNNs must be in the form of spike trains, various encoding schemes can be applied to the raw data in preparation for the application of this kind of network. These principally include translations to rate codes (in which the presence, absence, and frequency of spikes is taken into account) or to temporal codes (in which the relative timing of spikes is the principal focus), with the latter being especially suited to SNNs (Tu et al., 2014). As with nearly all neural network types, optimization of the network for a specific task involves updating the connections and connection weights between neurons, a task often achieved in a SNN via a spike-timing-dependent plasticity (STDP) paradigm, which strengthens or weakens connections between neurons depending on when the pre- and postsynaptic neurons fired relative to each other (Paugam-Moisy & Bohte, 2012).

3 Application of Networks to Visual Imagery Data

With a basic overview of the network types in mind, we will now turn to their respective advantages and disadvantages, especially as they apply to the processing of the EEG data commonly collected and processed in visual imagery tasks.

3.1 Applying ANNs

Basic ANNs benefit from being simpler and easier to set up than the more advanced neural network forms. They are able to handle larger inputs and can usually achieve high accuracy and efficiency with less data. However, these networks do not handle uncorrelated variables well, limiting what they can glean from more complex input data, and as with most neural networks, they suffer from the infamous "black box" problem, in which the relationships between variables in the final model are not always clear. This can limit the conclusions that can be drawn about the relationships between inputs and outputs and what factors were more or less important to the task (Chai et al., 2012; Hramov et al., 2017). Specifically with regard to visual imagery EEG data, ANNs also require that some form of feature extraction process is run on the data before input, resulting in the loss of potentially relevant data contained in the raw EEG data stream. However, their relative simplicity and the fact that they do not require any specific hardware to run, the existence of well-defined training and programming paradigms for this network type, and the existence of predeveloped and easy-to-use tools for programming this kind of network (such as PyTorch and TensorFlow) maintain the ANN as a "safe" option for use in visual imagery and other classification tasks.

3.2 Applying CNNs

While CNNs are typically associated with processing image data, they have also been successfully applied to neurological data (Kalafatovich et al., 2020). CNNs excel at local feature extraction and, with clean EEG data, they can achieve an impressive level of accuracy in classification tasks (Waytowich et al., 2018). Their feature-based method may also allow for better insights into the most critical aspects of the data for the output in a given task. However, the presence of significant noise in data is a serious challenge for this kind of network, as the filter kernels in the convolutional layers are more easily hindered by significant noise in the data stream. This particularly affects the processing of EEG data, which has a low signal-to-noise ratio (Bagchi & Bathula, 2022). Large amounts of EEG training data are required to train the model well for an EEG application, and the performance of the ultimate decoding is directly related to the dataset size (Zhang et al., 2020). CNNs are also not optimized for temporal data and can only detect structural patterns, a key drawback when working with visual system data, which is thought to rely heavily on the temporal components of neural activity (Bang et al., 2021). CNNs are therefore a suboptimal choice for processing visual imagery EEG data since their high data requirements and lack of temporal component processing would yield considerable drawbacks when working with this kind of task.

3.3 Applying RNNs

RNNs excel when long-term dependencies in data are present and essential for classification, and when data is sequential, as both require a form of memory. Parameter tuning is less of a concern in this kind of network as well, since the memory component affords the network a unique form of adaptability. The memory component of an RNN can also allow it to overcome noise, crucial when working with EEG data, provided that noise was present in the training data and not only in the data the network later works with in application (Yeo, 2019). However, domain knowledge is usually still key for saving processing time and hand-crafted steps are often needed when applying RNNs to EEG data, limiting their ability to shed light on elements of visual information processing that are not already known. RNNs can also struggle with lengthy input data, and training an RNN still requires substantial time and data (Ma et al., 2018; Qiao et al., 2019). Visual imagery EEG data specifically may not be lengthy enough for the RNNs struggles with lengthy input data to be a concern, but it is also not clear if longterm dependencies are important in visual imagery classification tasks, meaning that one of RNNs hallmark advantages may not be useful in this context at all.

3.4 Applying SNNs

SNNs are able to work with both the spatial and the temporal components of input data, the former particularly in models such as NeuCube, in which artificial neurons are arranged in a cortex-like layout and data is fed into the neurons closest to the location of the electrodes which originally recorded that data (Kasabov, 2014). With respect to EEG data, which already presents as a form of spike train, this network type can also accept raw or nearly-raw input and has been shown to be successful when doing so (Taylor et al., 2014). Accepting raw input allows the network to take full advantage of input data and avoids the inevitable data loss associated with feature extraction. The ability to handle the temporal components of input data, which as previously discussed are considered key for visual information processing (Bang et al., 2021) is another key advantage of this kind of network. Their lack of a rigid layer system can also make these networks more adaptable, since the entire structure of the network can shift dynamically to best match the task at hand. These networks have also been shown to be highly successful even when fed lower amounts of training data (Taylor et al., 2014). However, SNNs do have the significant drawback of requiring specialized hardware to run optimally, and parameter optimization and conversion of data to spike train inputs (for data not already in a spikelike form, such as EEG data) is a challenge (Kasabov, 2014). SNNs therefore present numerous advantages when approaching visual imagery tasks, though their hardware requirements, the challenge of parameter tuning in these models, and the fact that few established tools are currently available for programming them may limit how broadly they can be applied at this time.

4 Conclusion

Each neural network type discussed has advantages and disadvantages both in general and specifically with respect to visual imagery tasks. CNNs and standard ANNs are suboptimal for visual imagery as they are unable to process the temporal components considered key to working with visual information processing. RNNs can handle the temporal component, but suffer from the same high training data and time demands of CNNs. However, these three network types are well-established with substantial knowledge and tools available for successfully programming them. CNNs specifically would not be recommendable for a visual imagery task, but the use of an RNN for the temporal component or an ANN for simplicity and relatively lower data and training time requirements would both be suitable options. SNNs have a few current limitations that may prevent their widespread use for the time being, principally the requirement of specialized hardware and the current lack of tools and resources for programming them. However, they present a potentially game-changing option due to their ability to handle both spatial and temporal data, their high level of adaptability, and their ability to work with less training data. Once SNNs themselves are further developed and better tools and methods become available for training and applying them, these networks will likely present the best option for processing data in tasks like visual imagery.

References

Bagchi, S., & Bathula, D. R. (2022). EEG-ConvTransformer for single-trial EEG-based visual stimulus classification. *Pattern Recognition*, *129*, 108757. https://doi.org/10.1016/j.patcog.2022.108757

Bang, J.-S., Jeong, J.-H., & Won, D.-O. (2021). Classification of visual perception and imagery based eeg signals using convolutional neural networks. *2021 9th International Winter Conference on Brain-Computer Interface (BCI)*, 1–6. https://doi.org/10.1109/BCI51272.2021.9385367

Chai, R., Ling, S. H., Hunter, G. P., & Nguyen, H. T. (2012). Mental non-motor imagery tasks classifications of brain computer interface for wheelchair commands using genetic algorithm-based neural network. *The 2012 International Joint Conference on Neural Networks (IJCNN)*, 1–7. https://doi.org/10.1109/IJCNN.2012.6252499

Hramov, A. E., Maksimenko, V. A., Pchelintseva, S. V., Runnova, A. E., Grubov, V. V., Musatov, V. Y., Zhuravlev, M. O., Koronovskii, A. A., & Pisarchik, A. N. (2017). Classifying the perceptual interpretations of a bistable image using eeg and artificial neural networks. *Frontiers in Neuroscience*, *11*, 674. https://doi.org/10.3389/fnins.2017.00674

Kalafatovich, J., Lee, M., & Lee, S.-W. (2020). Decoding visual recognition of objects from eeg signals based on attention-driven convolutional neural network. *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2985–2990. https://doi.org/10.1109/SMC42975.2020.9283434

Kasabov, N. K. (2014). NeuCube: A spiking neural network architecture for mapping, learning and understanding of spatio-temporal brain data. *Neural Networks*, *52*, 62–76. https://doi.org/10.1016/j.neunet.2014.01.006

Kosmyna, N., Lindgren, J. T., & Lécuyer, A. (2018). Attending to visual stimuli versus performing visual imagery as a control strategy for eeg-based brain-computer interfaces. *Scientific reports*, *8*(1), 1–14. https://doi.org/10.1038/s41598-018-31472-9

Kwon, B.-H., Jeong, J.-H., Cho, J.-H., & Lee, S.-W. (2020). Decoding of intuitive visual motion imagery using convolutional neural network under 3D-BCI training environment. *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2966–2971. http://arxiv.org/abs/2005.08879

Luo, Y., Fu, Q., Xie, J., Qin, Y., Wu, G., Liu, J., Jiang, F., Cao, Y., & Ding, X. (2020). Eeg-based emotion classification using spiking neural networks. *IEEE Access*, *8*, 46007–46016. https://doi.org/10.1109/ACCESS.2020.2978163

Ma, X., Qiu, S., Du, C., Xing, J., & He, H. (2018). Improving eeg-based motor imagery classification via spatial and temporal recurrent neural networks. *2018 40th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, 1903–1906. https://doi.org/10.1109/EMBC.2018.8512590

Paugam-Moisy, H., & Bohte, S. M. (2012). Computing with spiking neuron networks. *Handbook of Natural Computing*, *1*, pp. 335–376. https://doi.org/10.1007/978-3-540-92910-910

Qiao, K., Chen, J., Wang, L., Zhang, C., Zeng, L., Tong, L., & Yan, B. (2019). Category decoding of visual stimuli from human brain activity using a bidirectional recurrent neural network to simulate bidirectional information flows in human visual cortices. *Frontiers in neuroscience*, *13*, 692. https://doi.org/10.3389/fnins.2019.00692

Taylor, D., Scott, N., Kasabov, N., Capecci, E., Tu, E., Saywell, N., Chen, Y., Hu, J., & Hou, Z.-G. (2014). Feasibility of neucube snn architecture for detecting motor execution and motor intention for use in bci applications. *2014 International Joint Conference on Neural Networks (IJCNN)*, 3221–322. https://doi.org/10.1109/IJCNN.2014.6889936

Tu, E., Kasabov, N., Othman, M., Li, Y., Worner, S., Yang, J., & Jia, Z. (2014). Neucube (st) for spatio-temporal data predictive modelling with a case study on ecological data. *2014 international joint conference on neural networks (IJCNN)*, 638–645. https://doi.org/10.1109/IJCNN.2014.6889717

Virgilio G., C. D., Sossa A., J. H., Antelis, J. M., & Falcón, L. E. (2020). Spiking neural networks applied to the classification of motor tasks in EEG signals. *Neural Networks*, *122*, 130–143. https://doi.org/10.1016/j.neunet.2019.09.037

Waytowich, N., Lawhern, V. J., Garcia, J. O., Cummings, J., Faller, J., Sajda, P., & Vettel, J. M. (2018). Compact convolutional neural networks for classification of asynchronous steady-state visual evoked potentials. *Journal of Neural Engineering*, *15*(6), 066031. https://doi.org/10.1088/1741-2552/aae5d8

Cognitive Science Student Journal 2023, Number 1

Yeo, K. (2019). Short note on the behavior of recurrent neural network for noisy dynamical system. *arXiv preprint arXiv:1904.05158*. http://arxiv.org/abs/1904.05158

Zhang, K., Xu, G., Chen, L., Tian, P., Han, C., Zhang, S., & Duan, N. (2020). Instance transfer subject-dependent strategy for motor imagery signal classification using deep convolutional neural networks. *Computational and Mathematical Methods in Medicine*, *2020*, 1–10. https://doi.org/10.1155/2020/1683013

About the Journal

The 'Cognitive Science Student Journal' aims at giving its readers an insight into current research and cutting-edge topics at our institute from a student perspective as well as students a platform to publish their work. Its editorial board consists of seminar participants and instructors of the Institute of Cognitive Science.

Cognitive Science is taught as an interdisciplinary research field at University Osnabrück, investigating cognition and the mind as a joint research effort of Artificial Intelligence, Neuroscience, Computational Linguistics, Psychology, Neuroinformatics, and Philosophy of Mind.

The journal can be accessed via:
http://cogsci-journal.uni-osnabrueck.de

Find us on social media:
https://www.instagram.com/cogscistudentjournal/
https://www.linkedin.com/company/cognitive-science-student-journal/