Brain-agnostic 3DCNNs learn naturalistic emotion from 7t fMRI

Joshua Lunger[1] Tohya Tanemura [1] Aditya Rajeev [1] Samuel Kostousov [1] Jurgen Germann [1, 2]

[1]University of Toronto [2]Krembil Brain Institute

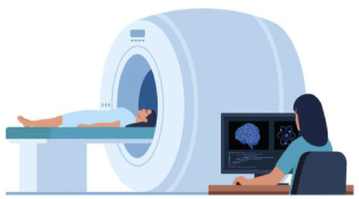
Introduction

We show brain-agnostic machine learning algorithms learn representations of naturalistic emotion. Our findings motivate clinical application in mental healthcare.

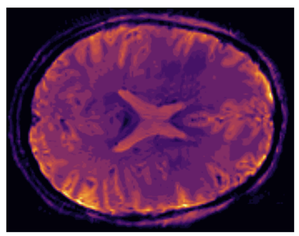
Problem Definition

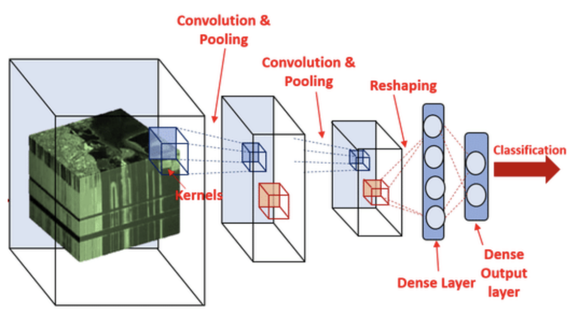
Given 3D fMRI images $\mathbf{X} \in \mathbb{R}^{x \times y \times z}$ and emotion labelings $\mathbf{y} \in \{1, \ldots, K\}$, we learn a mapping $g : \mathbf{X} \to \hat{\mathbf{y}}$ parameterized by a brain-agnostic 3D-convolutional neural network, where $\hat{\mathbf{y}} = \arg\max f(\mathbf{X}; \theta)$ represents the predicted emotion class, and $f(\mathbf{X}; \theta) \in \mathbb{R}^K$ is the network's output logits.

FMRI and Emotion Annotations



Model Parameterized by 3DCCNN

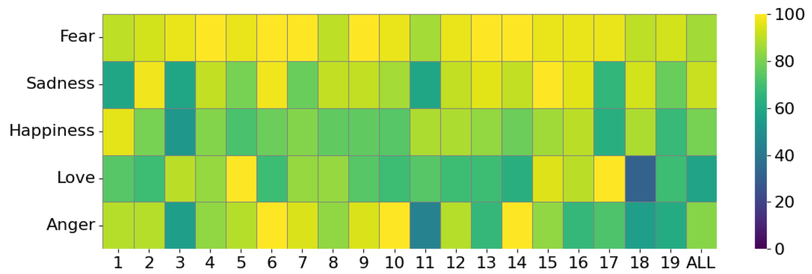




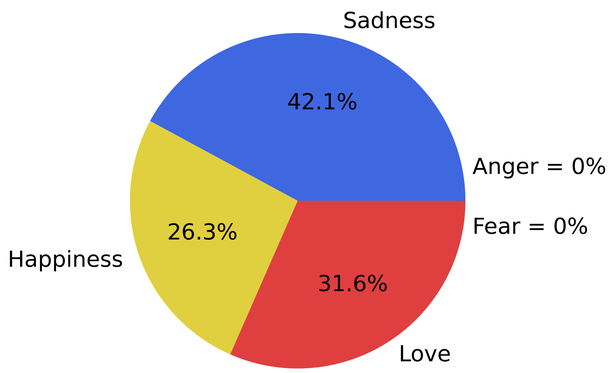
We parameterized our model with a brain-agnostic 3DCNN.

The model was trained using stochastic minibatch gradient descent with categorical cross-entropy loss and optimized with the Adam optimizer at a learning rate of 0.001. Training was conducted for 50 epochs on a NVIDIA Quadro M4000.

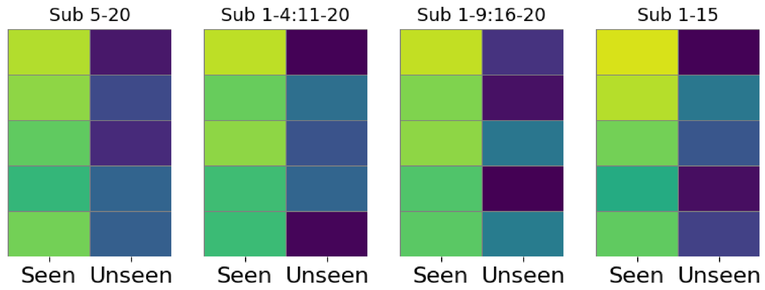
Inference is Powerful



Emotion-wise classification accuracy to held out data within-subject.



Subject 18 misclassification of love scenes.



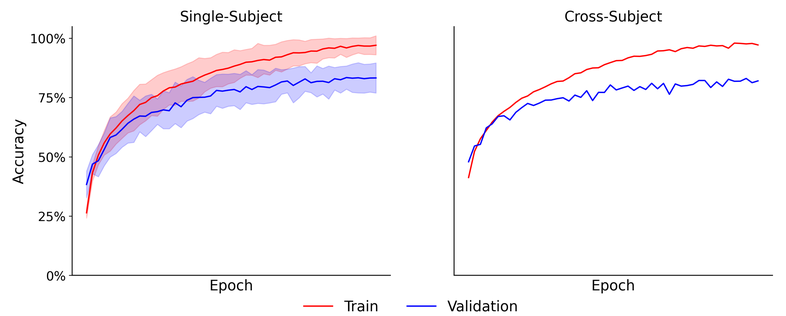
Emotion-wise classification accuracy to unseen subjects.

Inference is Consistent with Neurobiology

Our model's performance varies significantly across individuals. Indeed, naturalistic emotional responses exhibit high inter-individual variability[2, 3]. Notably, our model consistently performs well in detecting fear, which is known to be the most reliably observed emotional signal across individuals[4, 5]. Conversely, the cross-subject model struggles with predicting love scenes, reflecting neurobiological evidence that love is a learned emotion with high individual variability[6, 7].

Emotion Clustering is Difficult

Performance Generalizes Across Subjects



Train and validation accuracy over epochs. Single-Subject accuracy is averaged over 19 models trained on 1 subject each. Cross-Subject accuracy is 1 model trained on all 19 subjects.

Clinical Applications are Promising

Learning naturalistic emotion representations could aid care workers in developing new treatments for mental healthcare. Notably, the single-subject model trained exclusively on Subject 19 performs poorly in classifying love scenes compared to other single-subject models. By analyzing these misclassifications, we can gain insight into this subject's unique emotional responses, potentially identifying neural differences or deficits.

Moreover, models that generalize to unseen subjects could enable the development of consumer applications leveraging neural data for human preferences. However, we find that our model struggles to generalize across subjects, highlighting the need for further techniques to enhance cross-subject generalization.

References

[1] Hanke, M., Baumgartner, F., Ibe, P. et al. (2014). *A high-resolution 7-Tesla fMRI dataset from complex natural stimulation with an audio movie.* Sci Data, *1*, 140003. https://doi.org/10.1038/sdata.2014.3

[2] Tovote, P., Fadok, J., & Lüthi, A. (2015). Neuronal circuits for fear and anxiety. *Nature Reviews Neuroscience*, *16*, 317–331. https://doi.org/10.1038/nrn3945

[3] Steimer, T. (2002). The biology of fear- and anxiety-related behaviors. *Dialogues in Clinical Neuroscience*, *4*(3), 231-249. https://doi.org/10.31887/DCNS.2002.4.3/tsteimer

[4] Burunat, E. (2016). Love is not an emotion. *Psychology*, *7*, 1883-1910. https://doi.org/10.4236/psych.2016.714173

[5] Rinne, P., Lahnakoski, J. M., Saarimäki, H., Tavast, M., Sams, M., & Henriksson, L. (2024). Six types of love differentially recruit reward and social cognition brain areas. *Cerebral Cortex*, *34*(8), Article bhae331. https://doi.org/10.1093/cercor/bhae331

[6] Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., & Barrett, L. F. (2012). The brain basis of emotion: A meta-analytic review. *Behavioral and Brain Sciences*, *35*(3), 121-143. https://doi.org/10.1017/S0140525X11000446

[7] Saarimäki, H., Ejtehadian, L. F., Glerean, E., Jääskeläinen, I. P., Vuilleumier, P., Sams, M., & Nummenmaa, L. (2018). Distributed affective space represents multiple emotion categories across the human brain. *Social Cognitive and Affective Neuroscience*, *13*(5), 471-482.