Categorization of seen images from brain activity using sequence models

Arash Jamalian, Rafi Ayub, Faraz Fadavi
{arashj, rafiayub, fadavi} @stanford.edu

Motivation

Background

- The human visual system is incredibly complex and is not yet fully understood how it can recognize objects.
- For biologically-inspired deep learning models for visual recognition to continue to be successful, we must study the brain further.

Problem

- Many studies have decoded object category from visual brain activity, but most do not incorporate temporal information of fMRI.

Approach

- We apply an LSTM classifier to predict seen object category from the brain activity of visual areas using fMRI.

Methods

Dataset

- BOLD5000: 4 subjects were shown images from ImageNet, COCO, and Scene UNderstanding (SUN) in the scanner.
- Images were labeled with three super categories: Animal, Artifact, Scene.
- fMRI timeseries was extracted from visual areas: parahippocampal place area, retrosplenial complex, occipital place area, lateral occipital complex, and early visual area in each hemisphere, resulting in ten total ROIs per subject
- 13136 samples x 5 timepoints/TRs x 6960 total voxels

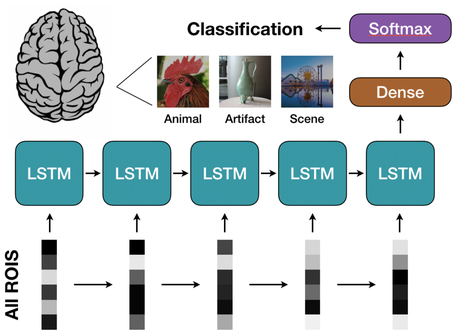
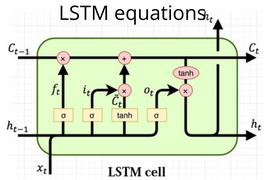


Figure 1: *LSTM classifier model. Dropout was including before the dense layer. The optimal model after hyperparameter tuning had an LSTM size of 12 units, dense 32, dropout of 49%, recurrent dropout of 55%, Adam optimizer with learning rate 0.0016, decay 0.001, and batch size of 208.*



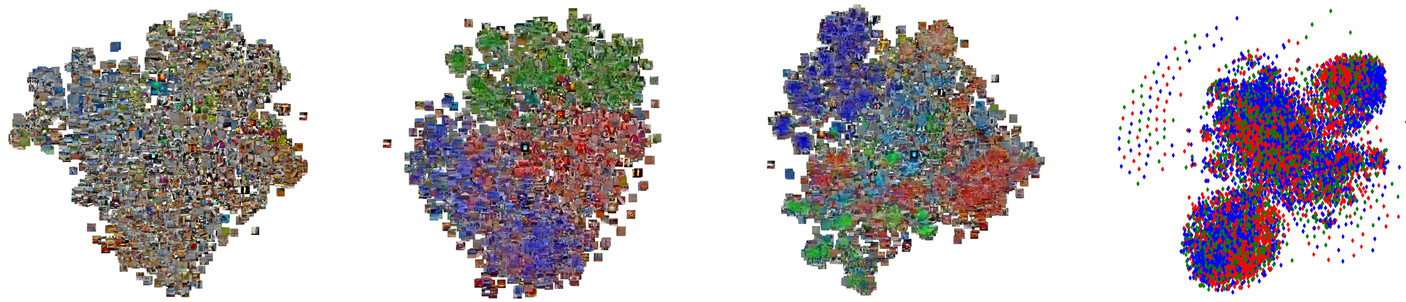
$$i_t = \sigma(x_t U^i + h_{t-1} W^i)$$
$$f_t = \sigma(x_t U^f + h_{t-1} W^f)$$
$$o_t = \sigma(x_t U^o + h_{t-1} W^o)$$
$$\tilde{C}_t = \tanh(x_t U^g + h_{t-1} W^g)$$
$$C_t = \sigma(f_t * C_{t-1} + i_t * \tilde{C}_t)$$
$$h_t = \tanh(C_t) * o_t$$

	Animal	Artifact	Scene	Subject Total
Subject 1	1227	1014	1420	3661
Subject 2	1227	1014	1420	3661
Subject 3	1227	1014	1420	3661
Subject 4	726	585	842	2153
Total	**4407**	**3627**	**5102**	**13136**

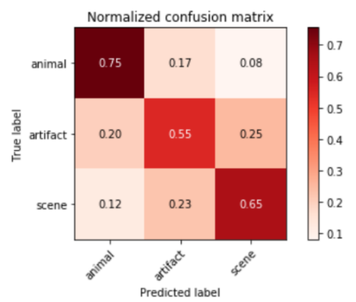
Table 1: Number of images from each class shown to each subjects.

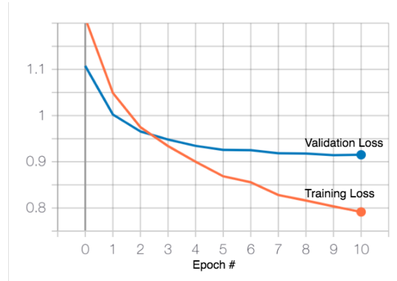
Results and Discussion

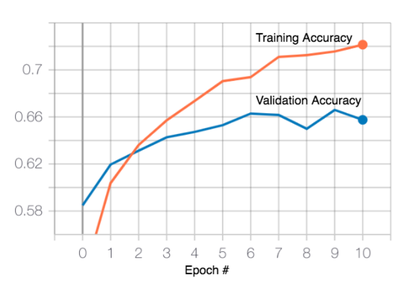
t-SNE



All images | *Animal, Artifact, Scene* | *Multiple subcategories* | *fMRI data*







Results:

- High noise levels in fMRI with small sample size require high dropout rate to prevent overfitting
- Achieve 68% accuracy with 65% average F1 score.
- Subjects have different distributions. Best performance on subject 1.
- Best performance: Animal category, Worst performance: Artifact category.

	Weighted Loss			Normal Loss		
	Precision	Recall	F_1 Score	Precision	Recall	F_1 Score
Animal	0.70	**0.75**	0.73	0.70	**0.79**	0.74
Artifact	0.52	**0.55**	0.53	0.55	**0.51**	0.53
Scene	0.73	**0.65**	0.69	0.73	**0.69**	0.71

Iteration	learning rate	Input dropout	Recurrent dropout	Batch size	LSTM hidden units	Val loss	Val accuracy
16	0.2317	0.54	0.27	234	4	1.14	0.581
28	0.0031	0.6	0.38	94	25	0.918	0.673
37	**0.0016**	**0.49**	**0.55**	**208**	**12**	**0.906**	**0.675**
39	0.0618	0.63	0.04	238	43	1.037	0.605
60	0.1296	0.27	0.39	62	155	1.29	0.325
81	0.0034	0.59	0.01	236	202	0.918	0.677
83	0.0980	0.18	0.33	57	248	1.297	0.396

	Subject 1	Subject 2	Subject 3	Subject 4	All Subjects
Training Size	2928	2928	2928	1722	**10508**
Validation Size	733	733	733	431	**2628**
Training Loss	0.720	0.749	0.747	0.689	**0.793**
Validation Loss	0.777	0.983	0.928	0.943	**0.912**
Training Accuracy	0.756	0.747	0.749	0.773	**0.719**
Validation Accuracy	0.746	0.648	0.664	0.654	**0.675**

Future Directions

- Acquiring more fMRI data from similar studies can help significantly reduce our overfitting
- Combining an LSTM with a generator network or autoencoder can allow use to generate images from fMRI data directly

References

1. Nadine Chang et al. “BOLD5000, a public fMRI dataset while viewing 5000 visual images”. In: Scientific data 6.1 (2019), p. 49.
2. Tomoyasu Horikawa and Yukiyasu Kamitani. “Generic decoding of seen and imagined objects using hierarchical visual features”. In: Nature Communications 8.May (2017), pp. 1–15. ISSN: 20411723. DOI: 10.1038/ncomms15037. arXiv: 1510.06479. URL: http://dx.doi.org/10.1038/ncomms15037.
3. Kai Qiao et al. “Category Decoding of Visual Stimuli From Human Brain Activity Using a Bidirectional Recurrent Neural Network to Simulate Bidirectional Information Flows in Human Visual Cortices”. In: Frontiers in Neuroscience 13.July (2019), pp. 1–15. DOI: 10.3389/fnins.2019.00692.
4. Guohua Shen et al. “Deep image reconstruction from human brain activity”. In: PLoS Computational Biology 15.1 (2019), pp. 1–23. ISSN: 15537358. DOI: 10.1371/journal.pcbi.1006633.
5. Kuan Han et al. “Variational autoencoder: An unsupervised model for encoding and decoding fMRI activity in visual cortex”. In: NeuroImage 198.January 2018 (2019), pp. 125–136. ISSN: 10959572. DOI: 10.1016/j.neuroimage.2019.05.039.