A.9 Performance with varying dataset size

Method	Low-Level				High-Level				Retrieval	
	PixCorr ↑	SSIM ↑	Alex(2) ↑	Alex(5) ↑	Incep ↑	CLIP ↑	Eff ↓	SwAV ↓	Image ↑	Brain ↑
All Data (High-Level)	.209	.318	92.8%	98.0%	94.5%	94.8%	.635	.361	97.2%	94.7%
Half Data (High-Level)	.149	.276	87.7%	94.3%	87.1%	90.1%	.738	.424	77.5%	60.8%
2-Sessions (High-Level)	.119	.281	81.0%	88.2%	79.2%	84.4%	.824	.472	17.9%	12.0%

Table 9: Quantitative comparison of MindEye performance with varying dataset sizes on Subject 1 with the high-level pipeline. Half Data corresponds to MindEye trained with half of the training samples randomly removed. 2-Sessions corresponds to MindEye trained with a random selection of 500 training image samples (or 1,500 training fMRI samples given 3 repetitions per image), equivalent to the number of samples collected across two scan sessions. Notably, image and brain retrieval metrics maintained state-of-the-art performance even when training the model with half of the training samples removed, and reconstruction performance remained competitive with previous models even with reduced training data. This suggests that our MindEye approach is flexible to being trained with smaller datasets.

A.10 Model size comparison with other methods

Method		Parameter Count
Lin et al.		2×1.17M deep models + StyleGAN
Takagi et al.	Low Level	37M linear regression model
	High Level	450M linear regression model
Ozcelik et al.	Low Level	1.45B linear regression model
	High Level	257 separate 12M linear regression models
MindEye	Low Level	206M residual MLP + CNN decoder model
	High Level	996M residual MLP + diffusion prior model

Table 10: Comparison of MindEye parameter count with other competing methods. Other methods primarily rely on linear regression or relatively small deep models.