Reanimating Images using Neural Representations of Dynamic Stimuli (BrainNRDS)

Jacob Yeung, Andrew F. Luo, Gabriel Sarch,
Margaret H. Henderson, Deva Ramanan, Michael J. Tarr

Carnegie Mellon University

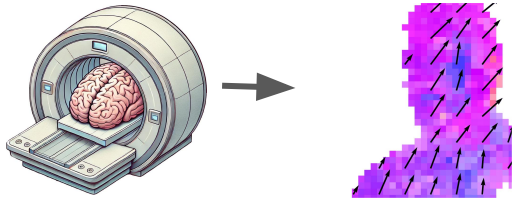
ML Machine Learning Department | Carnegie Mellon University School of Computer Science

Carnegie Mellon University Robotics Institute

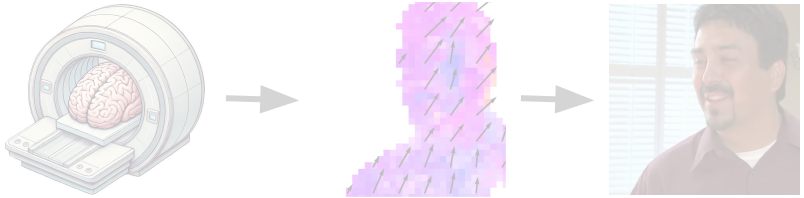
Neuroscience Institute

Roadmap

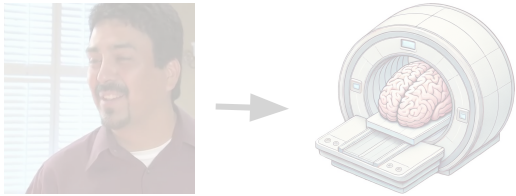
1. Decoding motion from the brain



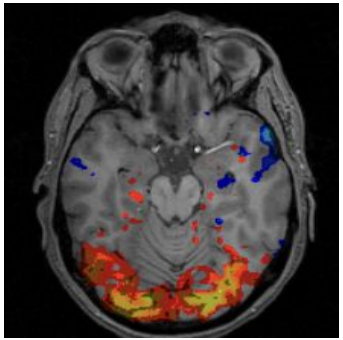
2. Decoding video from the brain



3. Predicting brain from video

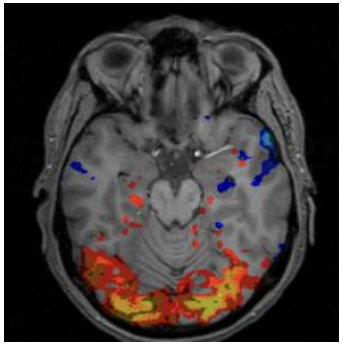


Background: functional Magnetic Resonance Imaging (fMRI)



- fMRI scanners measure changes in blood oxygenation at a 2-3 mm resolution every 2 seconds [100x100x100xT/2]
- Blood oxygenation responds slowly, making it challenging (but interesting) to measure fast neural changes

Background: functional Magnetic Resonance Imaging (fMRI)

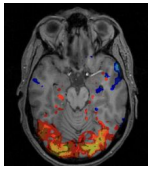


- fMRI scanners measure changes in blood oxygenation at a 2-3 mm resolution every 2 seconds [100x100x100xT/2]
- Blood oxygenation responds slowly, making it challenging (but interesting) to measure fast neural changes

Neural Encoding and Decoding with Deep Learning for Dynamic Natural Vision

Haiguang Wen[1,2], Junxing Shi[1,2], Yizhen Zhang[1,2], Kun-Han Lu[1,2], Jiayue Cao[2,3] and Zhongming Liu[1,2,3]

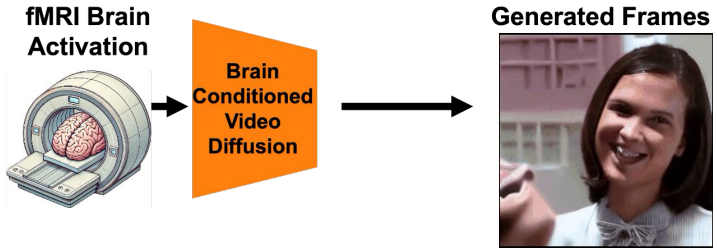
[1]School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47906, USA, [2]Purdue Institute for Integrative Neuroscience, Purdue University, West Lafayette, IN 47906, USA and [3]Weldon School of Biomedical Engineering, Purdue University, West Lafayette, IN 47906, USA

( , )

- Paired video-fMRI dataset of ~1000 2-second video clips

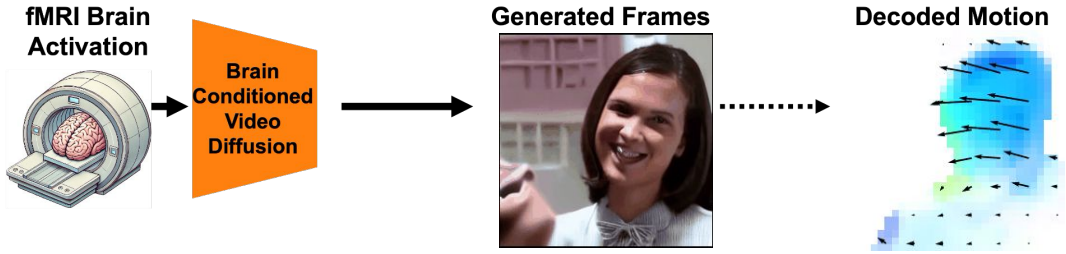
Decoding video motion

MindVideo
(NeurIPS ‘23)



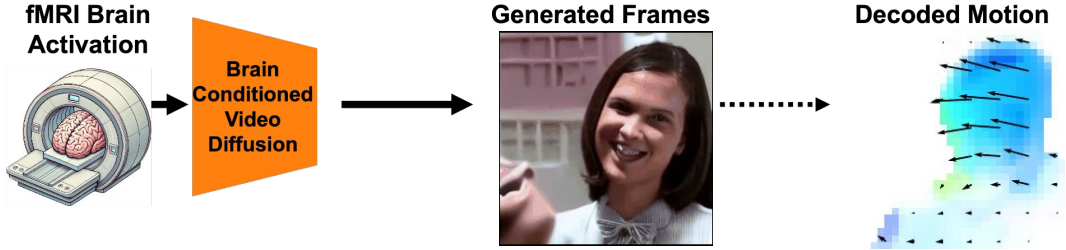
Decoding video motion

MindVideo
(NeurIPS ‘23)

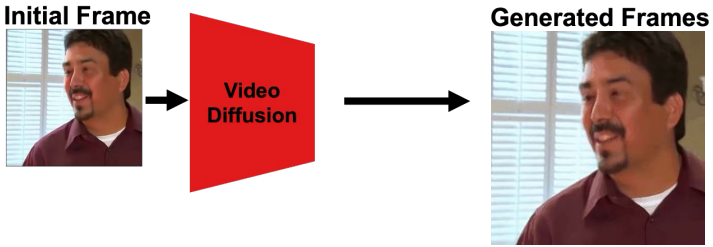


Decoding video motion

MindVideo
(NeurIPS ‘23)

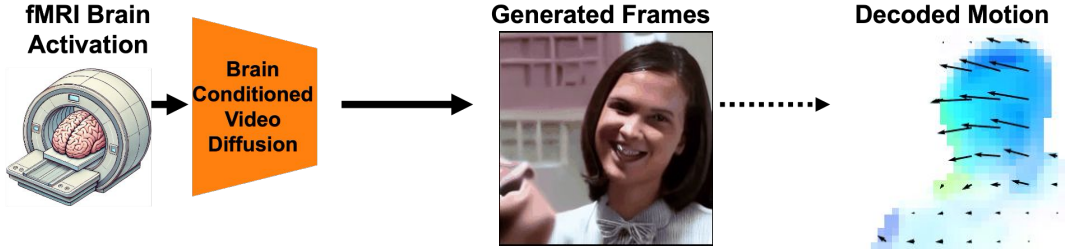


Stable Video Diffusion (SVD)

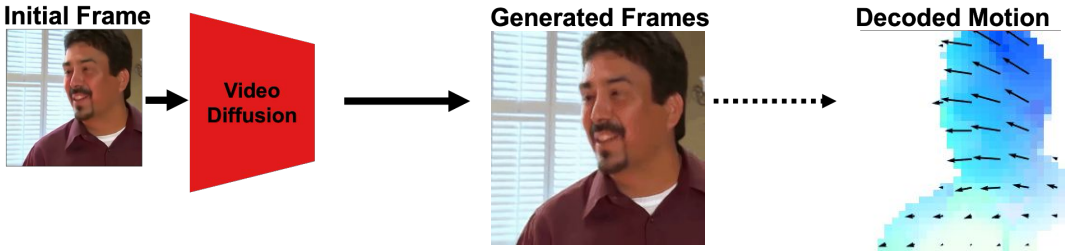


Decoding video motion

MindVideo
(NeurIPS ‘23)

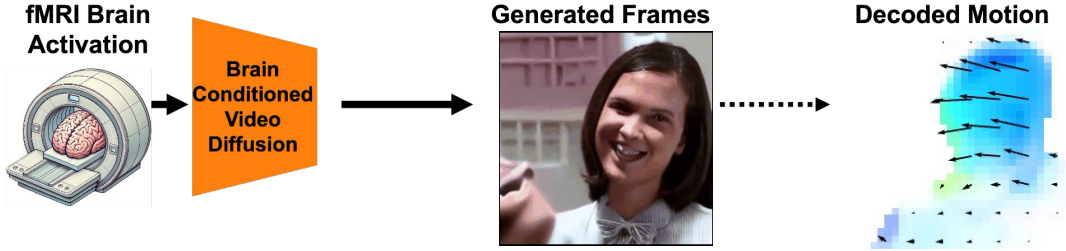


Stable Video Diffusion (SVD)

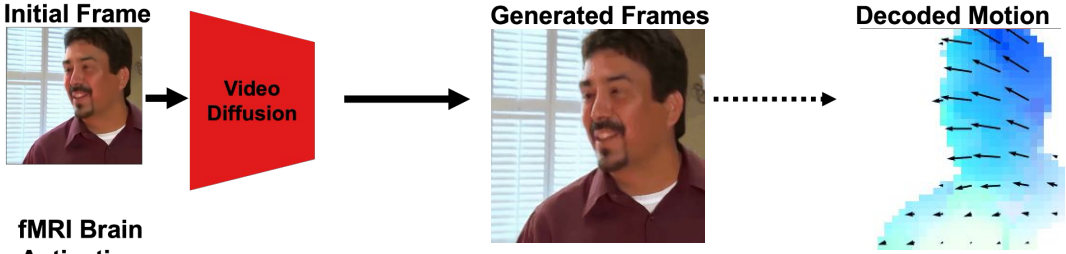


Decoding video motion

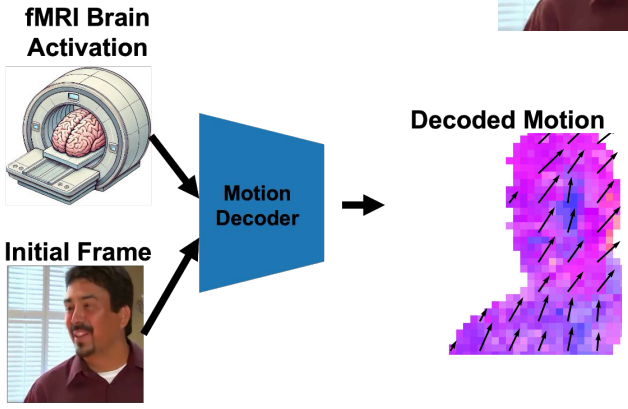
MindVideo
(NeurIPS '23)



Stable Video Diffusion (SVD)

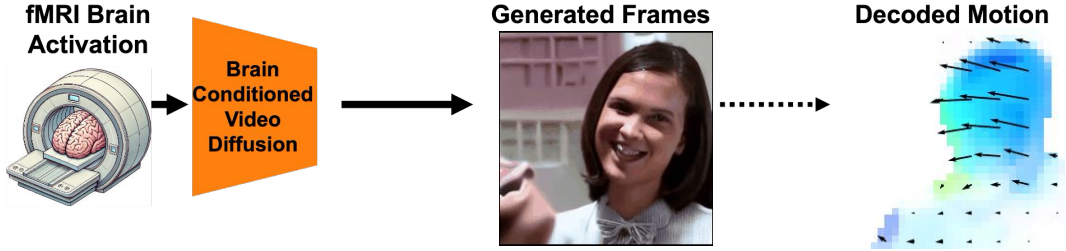


BrainNRDS (Ours)

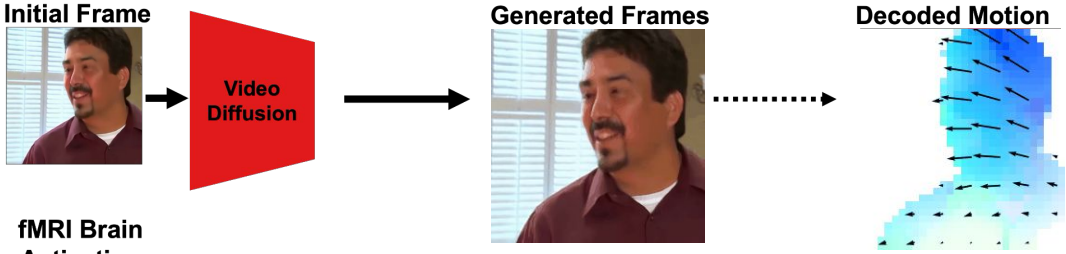


Decoding video motion

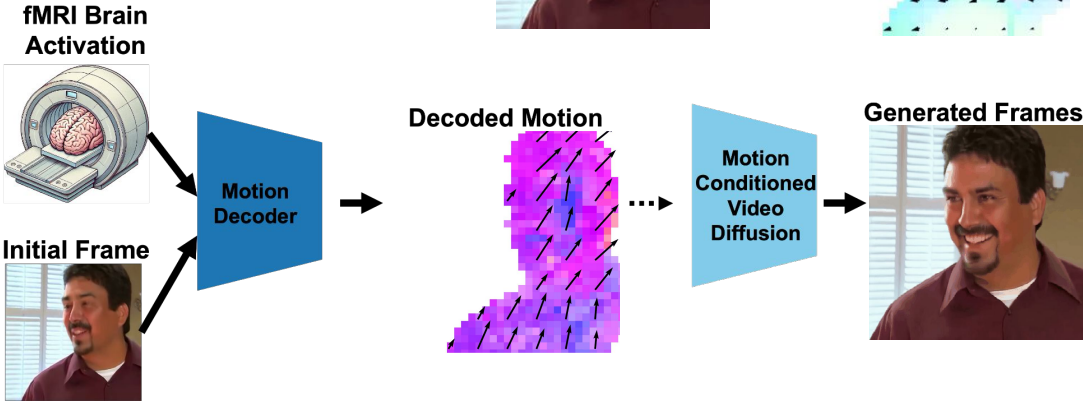
MindVideo
(NeurIPS '23)



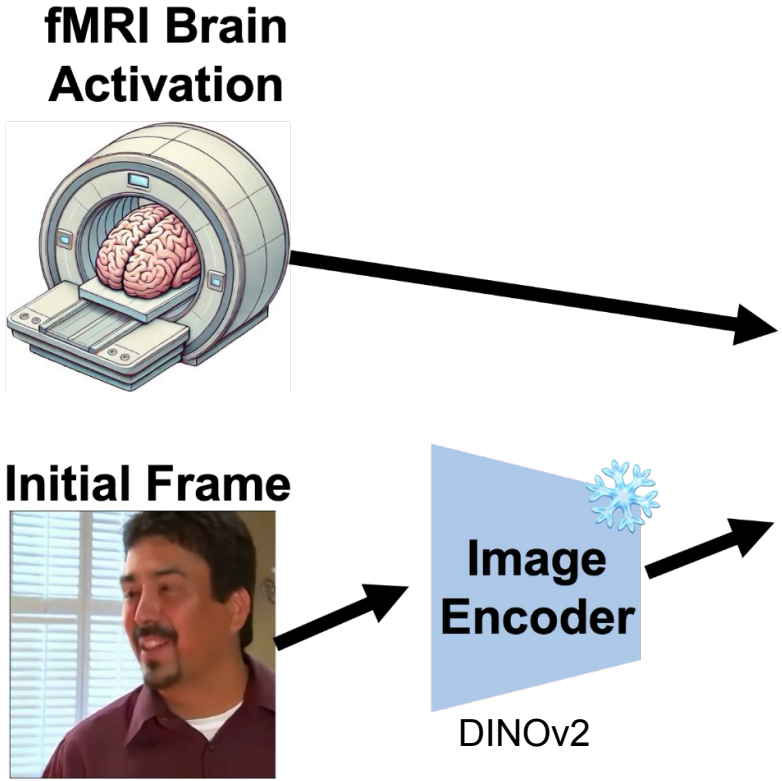
Stable Video Diffusion (SVD)



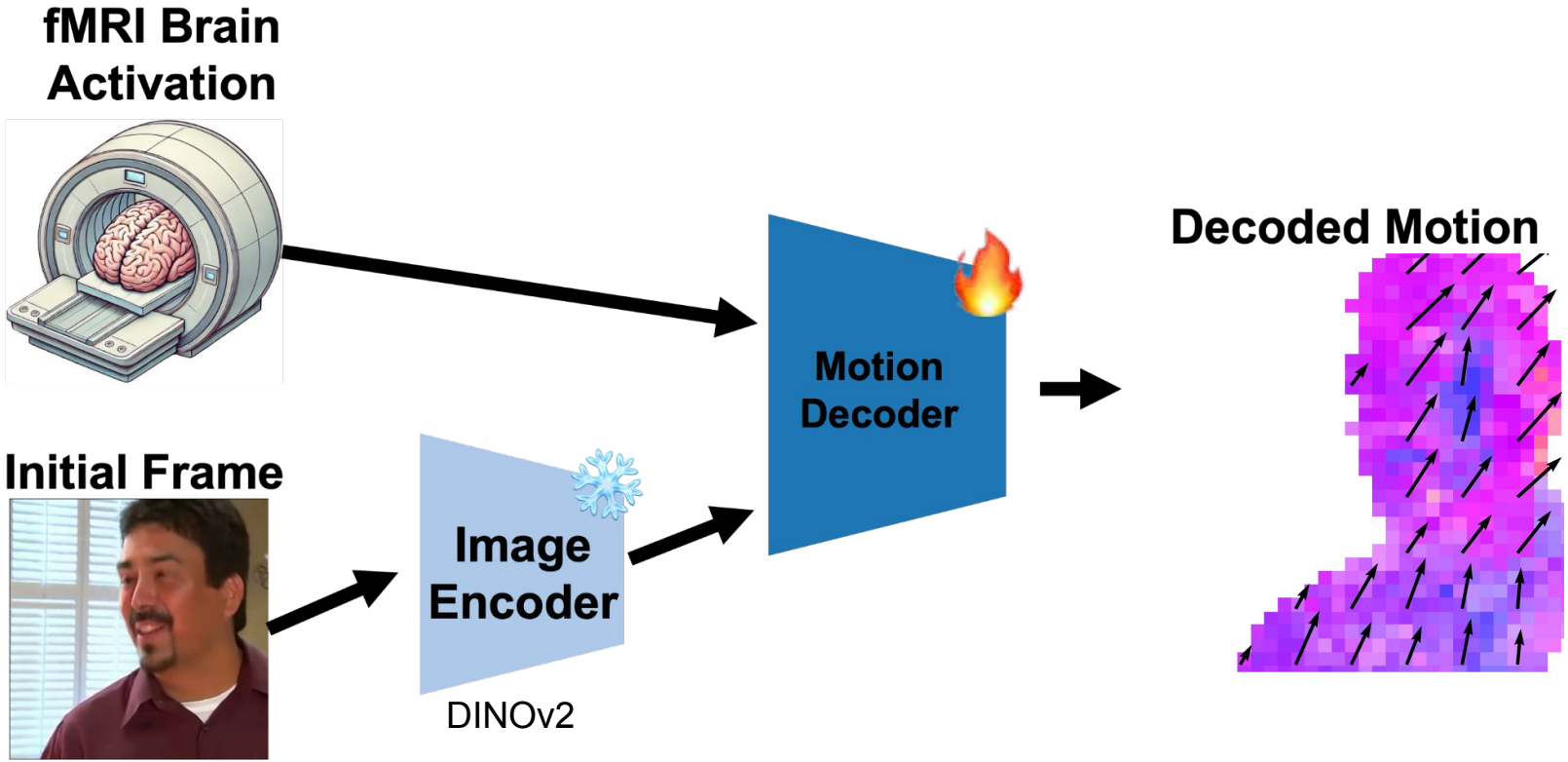
BrainNRDS
(Ours)



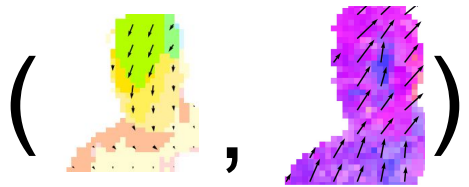
Our method for decoding motion from fMRI data



Our method for decoding motion from fMRI data



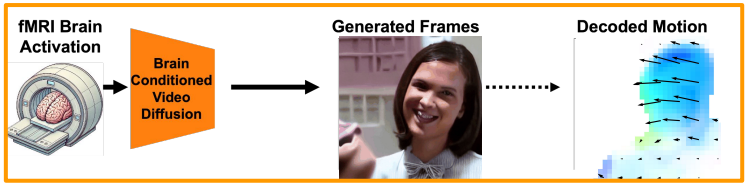
Quantitative motion decoding evaluation

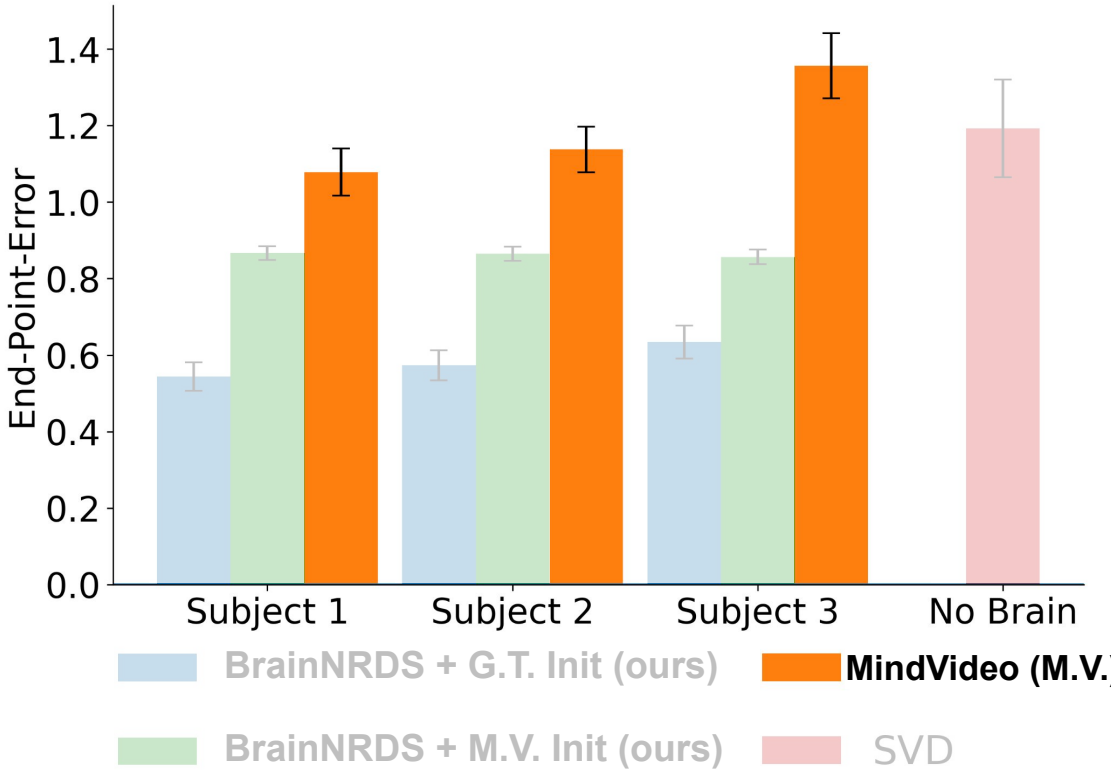


ground-truth prediction

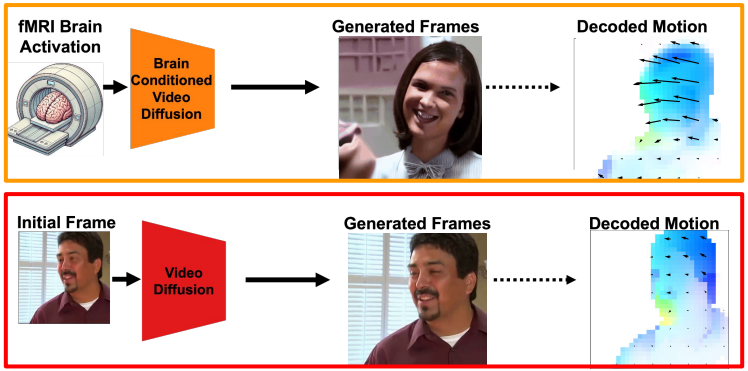
EPE or End-Point error

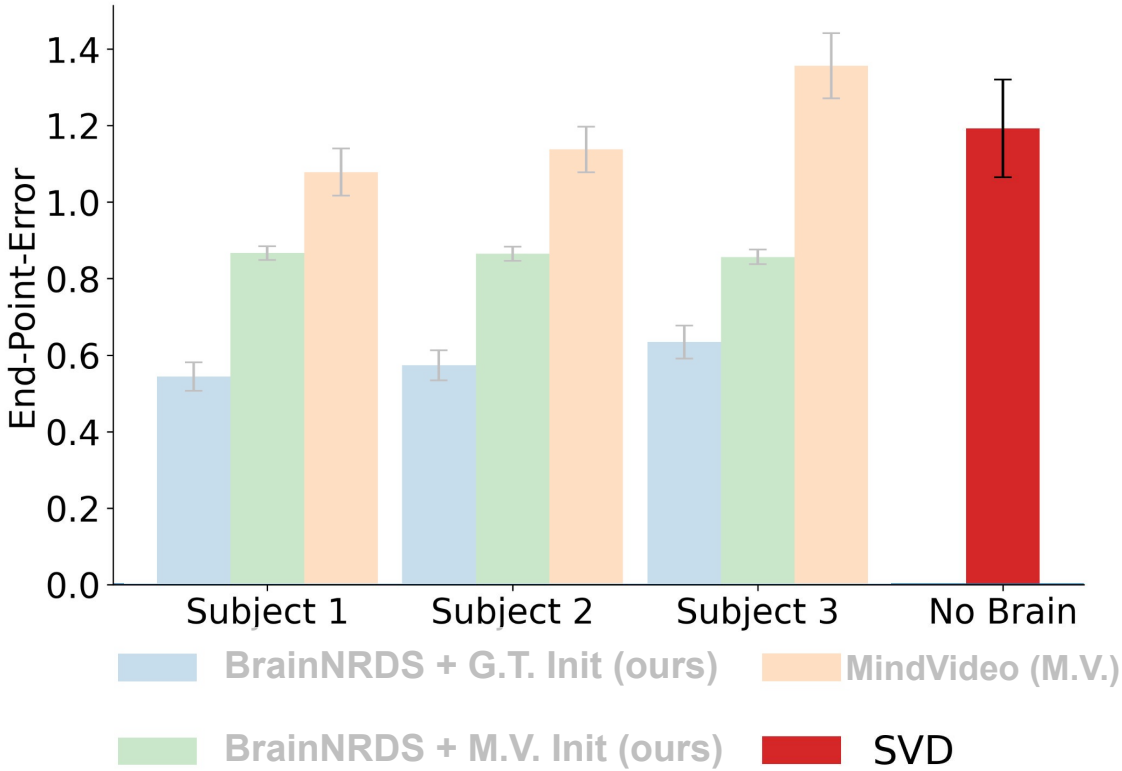
Quantitative motion decoding evaluation



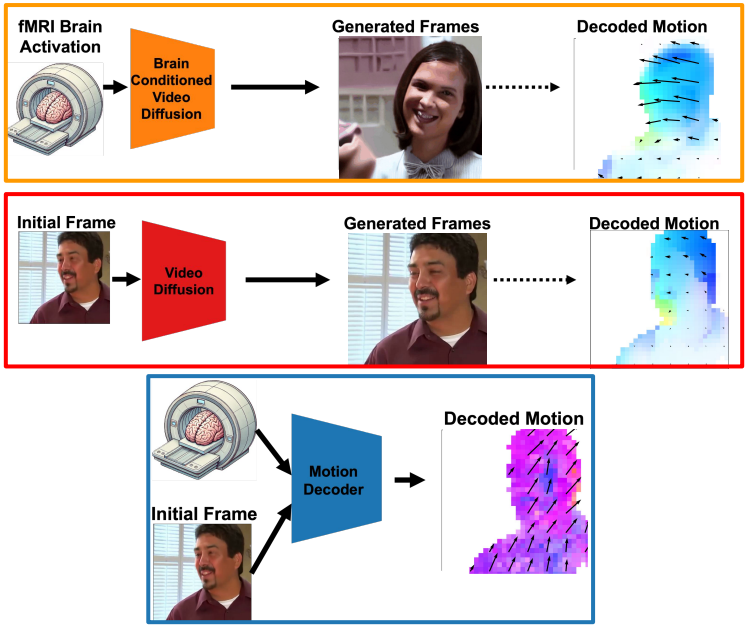


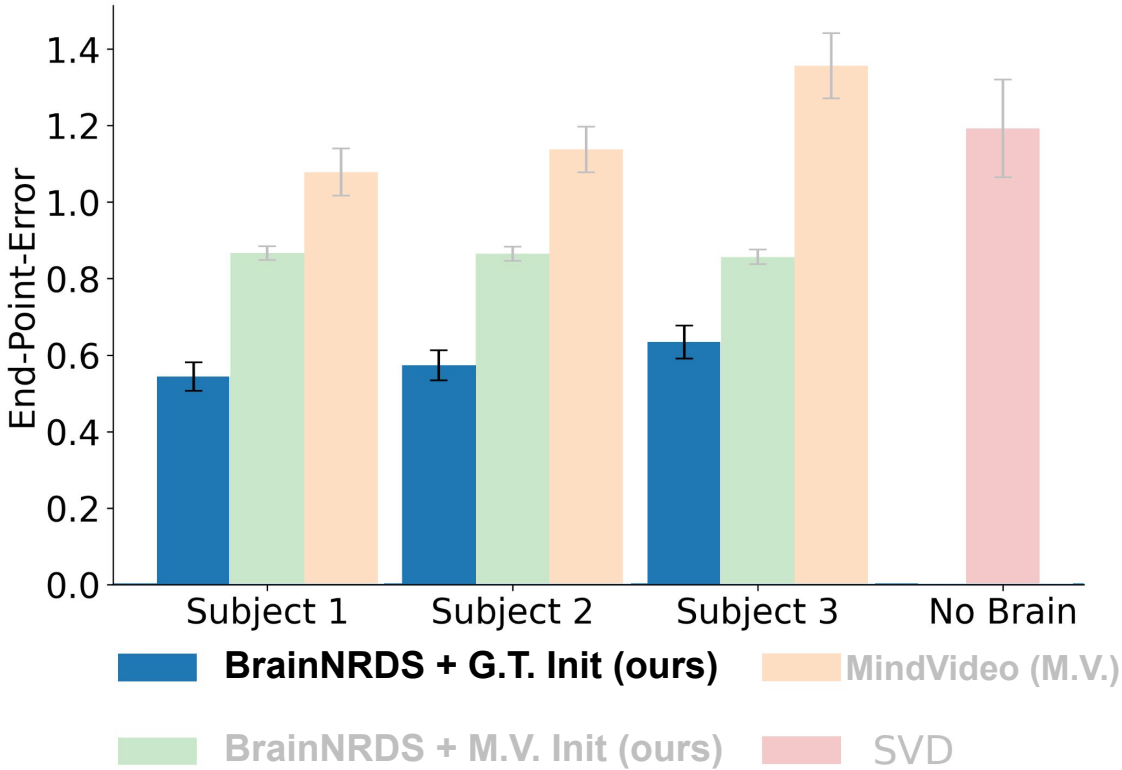
Quantitative motion decoding evaluation



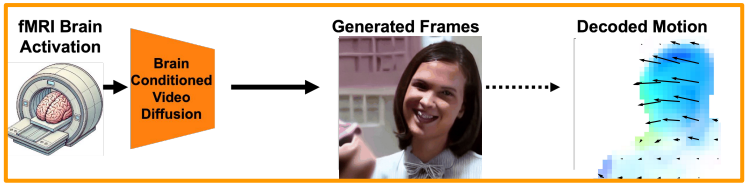


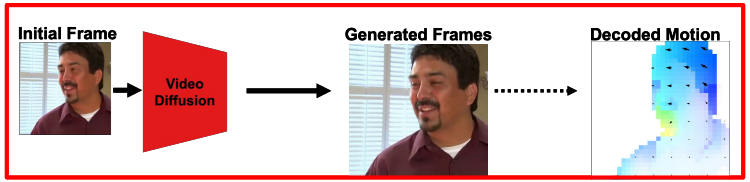
Quantitative motion decoding evaluation

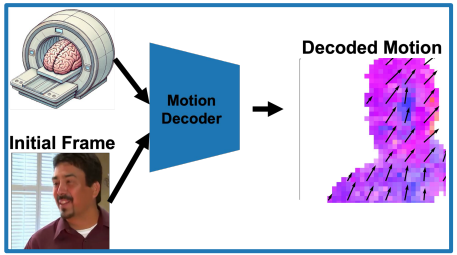


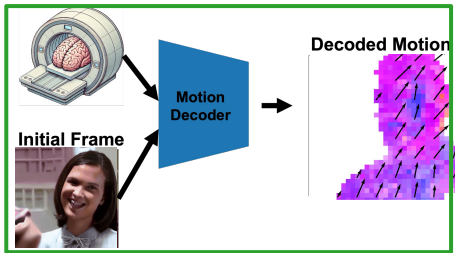


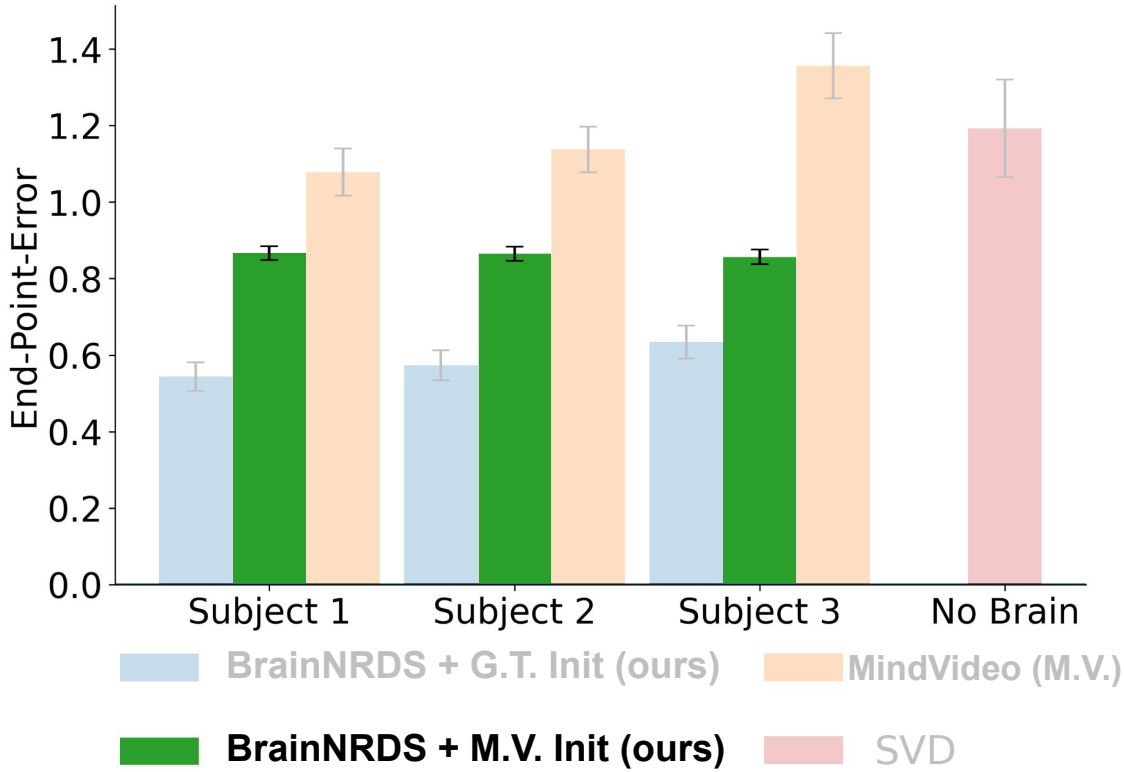
Quantitative motion decoding evaluation



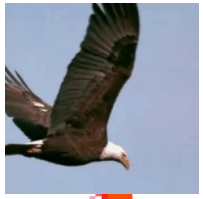
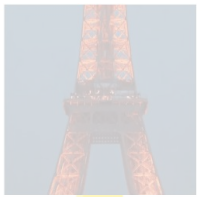
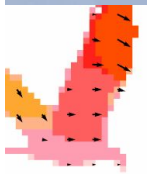
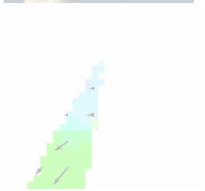
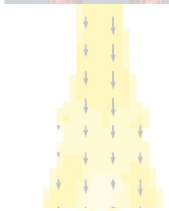
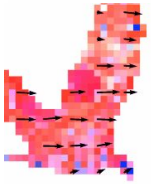
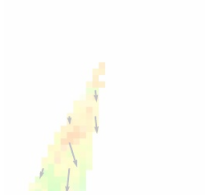
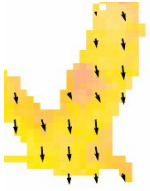
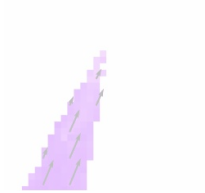




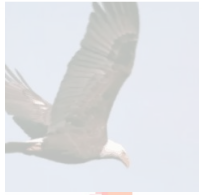
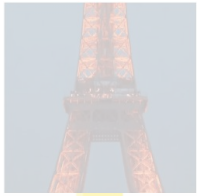
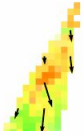




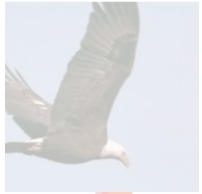
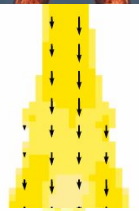
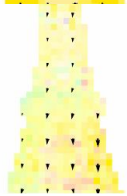
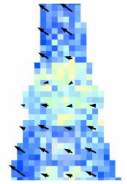
Comparison with and without neural data

	Multiple plausible actions	Camera motion overriding plausible motion	Ambiguity of motion of stationary objects
Ground Truth Video			
Ground Truth Flow			
Flow with Brain (Ours)			
Flow without Brain (SVD)			

Comparison with and without neural data

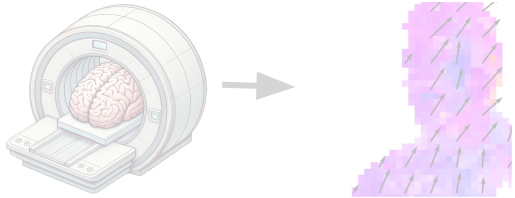
	Multiple plausible actions	Camera motion overriding plausible motion	Ambiguity of motion of stationary objects
Ground Truth Video			
Ground Truth Flow			
Flow with Brain (Ours)			
Flow without Brain (SVD)			

Comparison with and without neural data

	Multiple plausible actions	Camera motion overriding plausible motion	Ambiguity of motion of stationary objects
Ground Truth Video			
Ground Truth Flow			
Flow with Brain (Ours)			
Flow without Brain (SVD)			

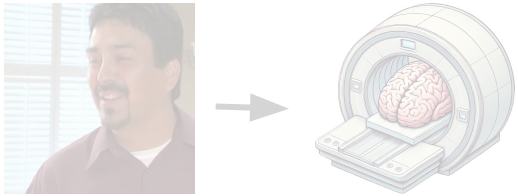
Roadmap

1. Decoding motion from the brain

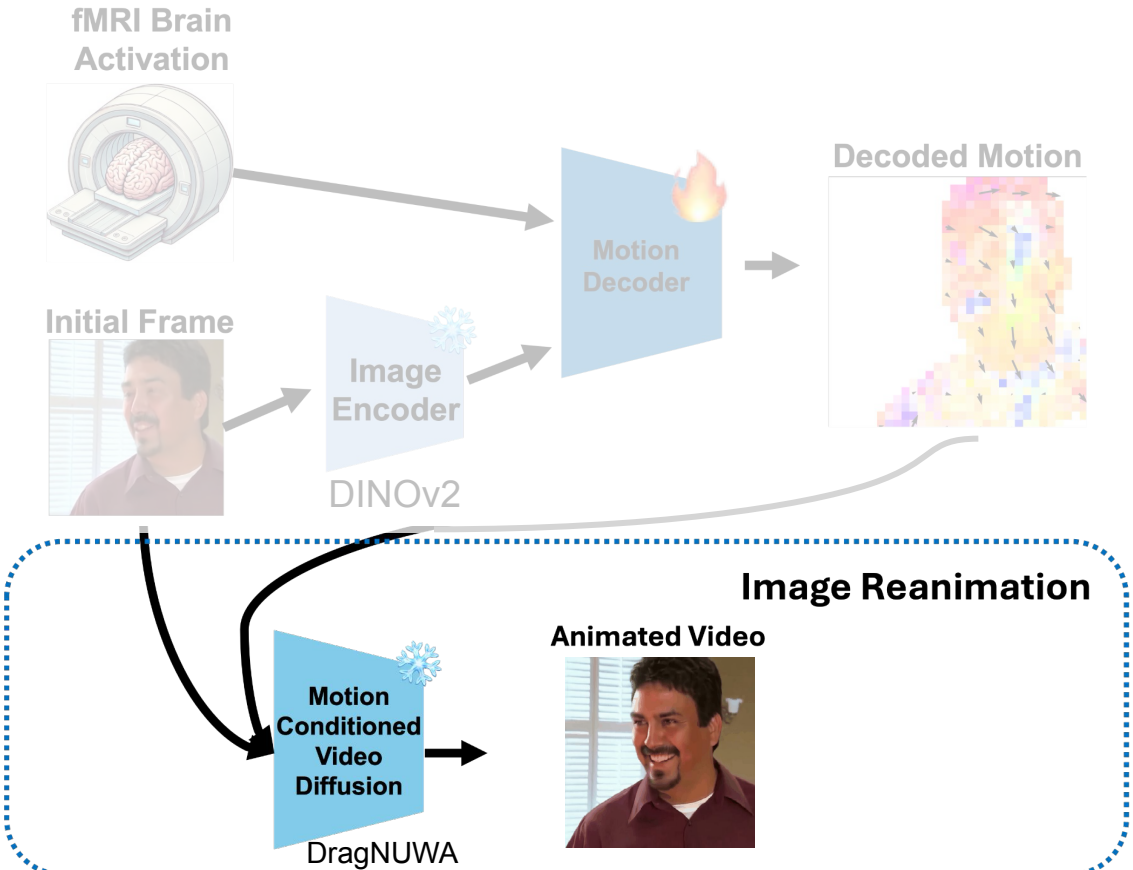


2. Decoding video from the brain

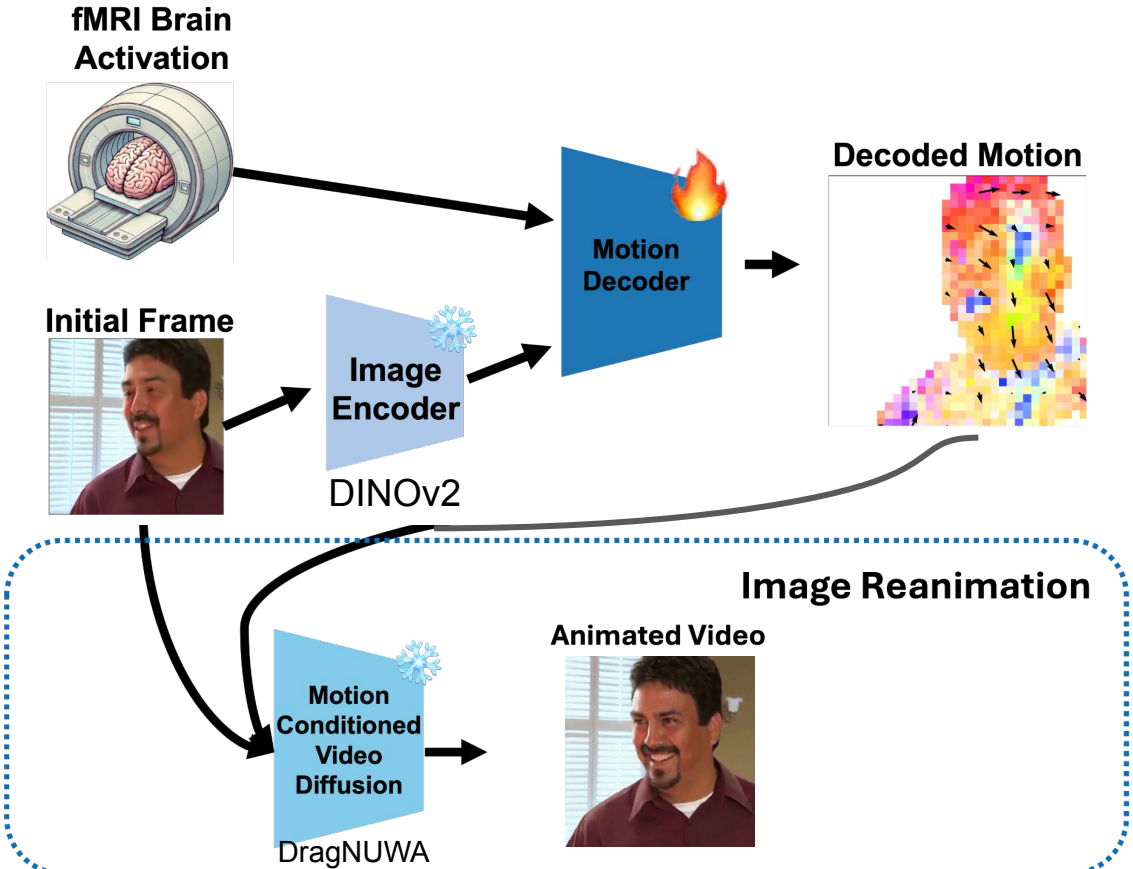
3. Predicting brain from video



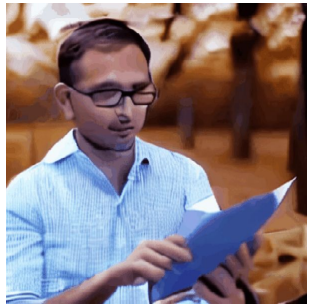
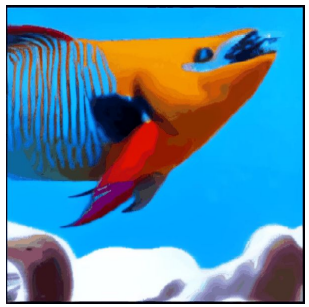
Reanimating frames using motion from fMRI data



Reanimating frames using motion from fMRI data

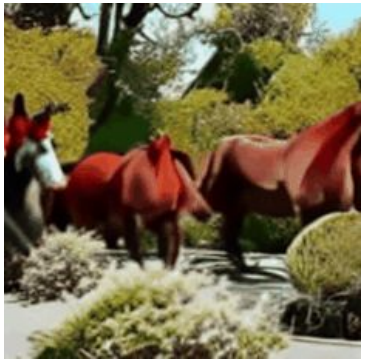


Generated video examples

G.T.	BrainNRDS + G.T. Initial Frame	BrainNRDS + MindVideo Initial Frame
		
		

Our method compared to standard approach

MindVideo

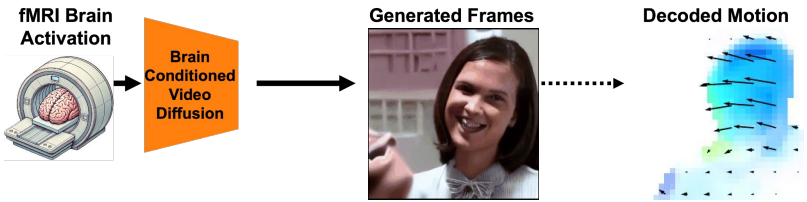


BrainNRDS

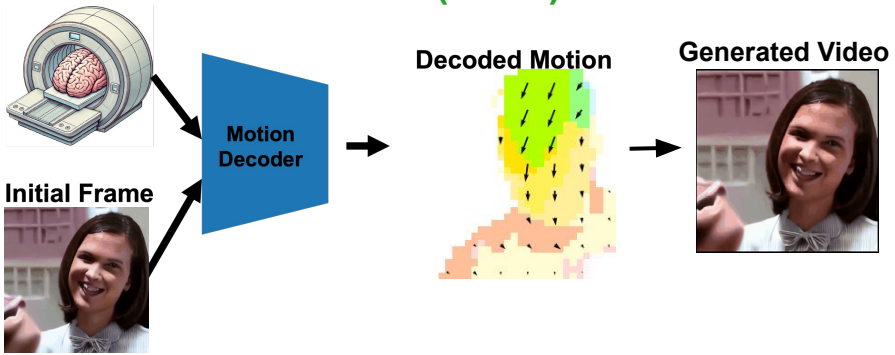
Our method performs best on video reconstruction metrics

Method	Video-based	Frame-based	
	VideoMAE CosSim↑	CLIP CosSim↑	Pixel SSIM↑
MindVideo	0.742±0.006	0.879±0.004	0.171±0.02
BrainNRDS (Ours)	**0.769±0.006**	**0.896±0.003**	**0.214±0.01**

MindVideo

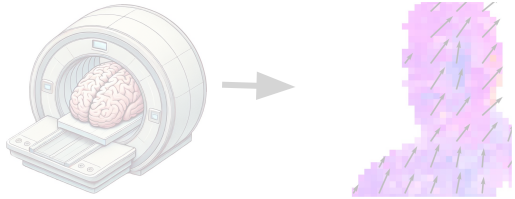


BrainNRDS (Ours) + MindVideo

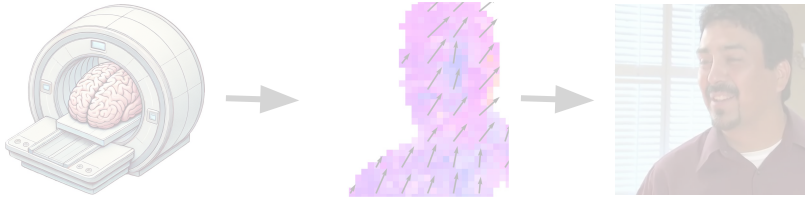


Roadmap

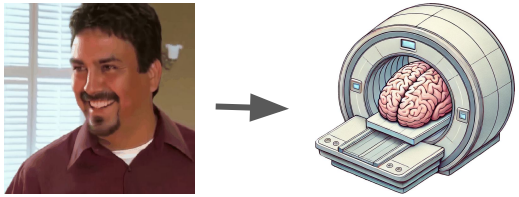
1. Decoding motion from the brain



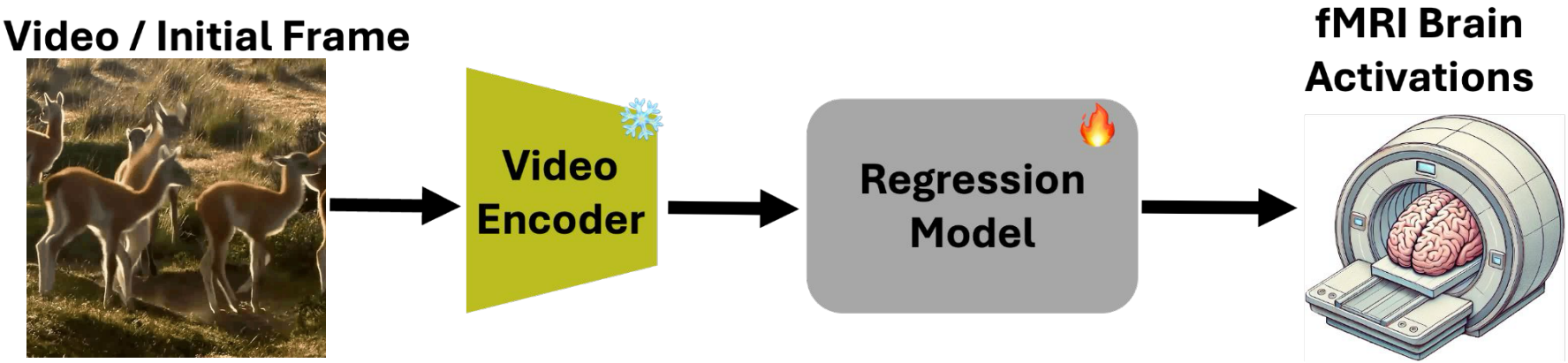
2. Decoding video from the brain



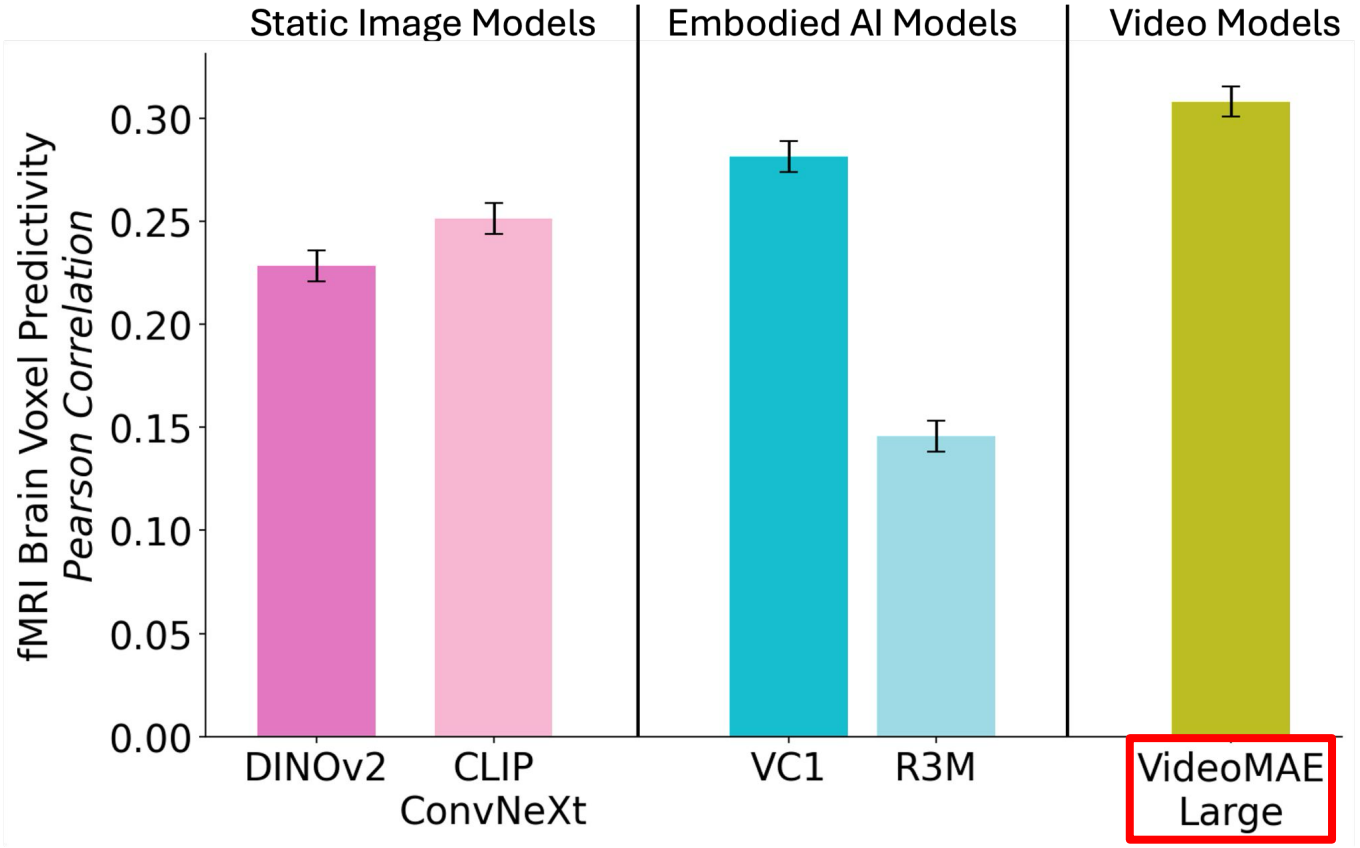
3. Predicting brain from video



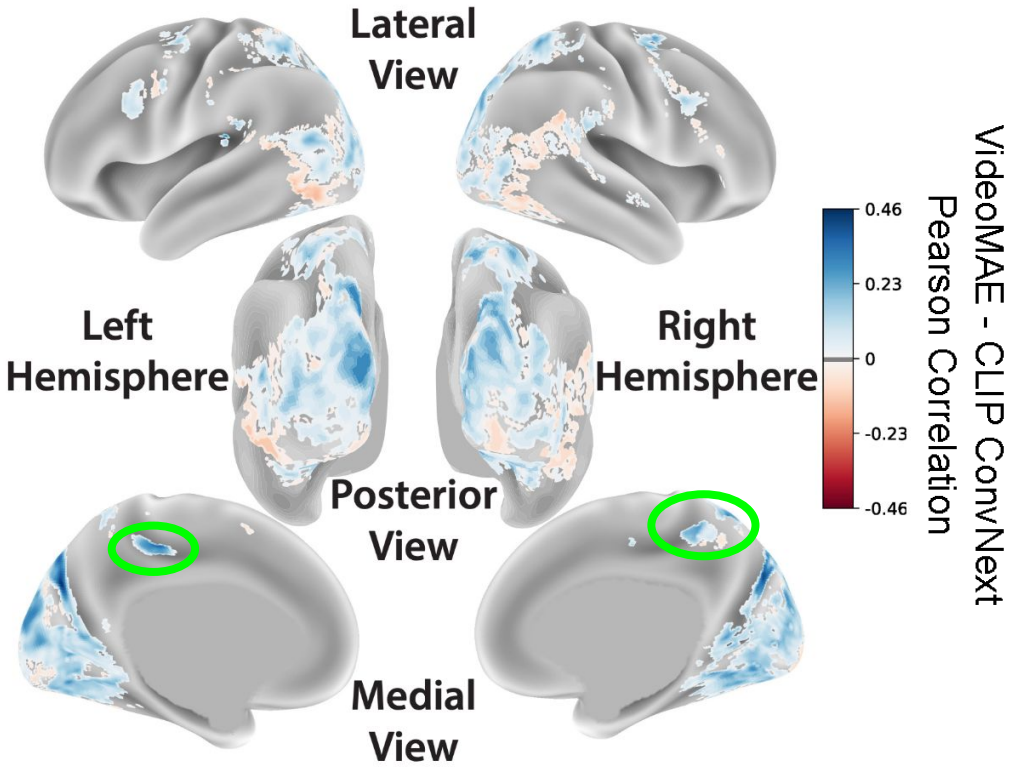
Identifying brain regions tuned to dynamic features



Brain Encoding Prediction Performance



Prediction difference between video and image model (VideoMAE vs CLIP ConvNeXt)



Video > Image

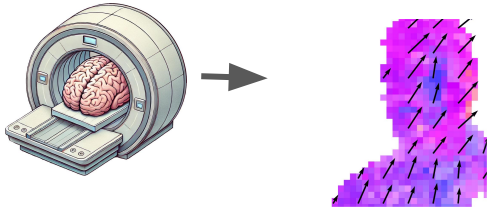
Video < Image

Video model better predicts:

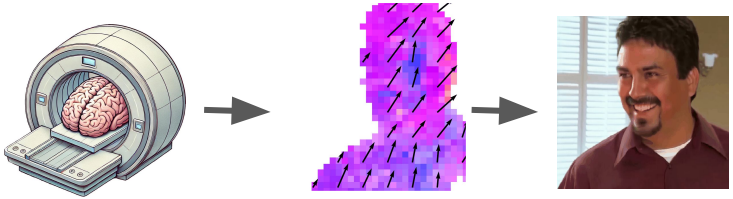
1. most of visual cortex
2. visuomotor regions.

A look back

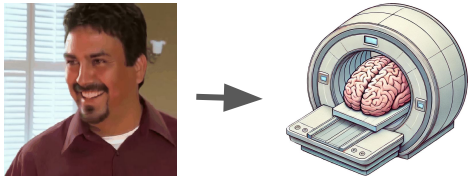
1. Decoding motion from the brain



2. Decoding video from the brain



3. Predicting brain from video



- Decoding with motion prior improves fMRI video generation
- Video encoders better predict brain responses

Reanimating Images using Neural Representations of Dynamic Stimuli

Poster Session: Today, 4:00-6:00pm, Poster #220 ExHall D

jacobyeung@cmu.edu

Project Page: