Table . Comparison of Methods for fMRI Decoding and Visual Image Reconstruction

Category	Representative Method	Dataset	Innovation	Advantages	Disadvantages	Results
Traditional Methods	Gabor-wavelet RF Model [9]	Vim-1	Early visual features are mapped.	Preset image sets can be recognized.	Pixel reconstruction cannot be performed.	Category recognition is achieved.
	Bayesian CCA Model [10]	Vim-1	Sparse mappings are extracted.	Simple pixel images can be reconstructed.	Reconstruction is limited by fine lines and high noise.	Rough image reconstruction is obtained.
Deep Learning (Non-generative)	Voxels → Pixels Bidirectional Model [15]	GOD	Unlabeled data are used for training.	Rough structures can be reconstructed.	Rough structures are the only ones reconstructed.	Shape similarity is achieved.
	NEI-GNN Multi-scale [16]	GOD	Structural details are optimized.	Good texture can be obtained.	Semantic consistency is lacking.	Accurate structure reconstruction is achieved.
Generative (Encoder-Decoder type)	Foreground-Attention & Loop-Enc-Dec [20]	GOD	A visual attention mechanism for guidance is introduced.	Attended regions can be effectively reconstructed.	Generalization is limited by dataset diversity.	Clear structure with limited scene coverage is obtained.
Generative (GAN type)	Dual-guided Brain Diffusion Model [35]	NSD	Multi-stage guidance control is applied.	Randomness and blur can be improved.	Performance is slightly worse in complex scenes.	High semantic consistency is achieved.
Generative (Diffusion Models)	MindDiffuser[34]	NSD	Dual guidance for semantics and structure is applied.	Randomness and blur can be improved.	Regularization adaptability is insufficient.	Good detail in simple scenes is achieved.
	STTM[41]	NSD, GOD	Cross-subject transfer and high/low-level joint learning are performed.	Cross-subject adaptation transfer can be realized.	Structural fidelity still needs improvement.	Good results in both low-level pixels and high-level semantics are obtained.
	LDM[44]	NSD	Simple linear mapping for reconstruction is applied.	High resolution and high semantic fidelity can be achieved.	Reconstructed visual images lack natural appearance.	High-resolution reconstruction is obtained.
	MindEye2[45]	NSD	A short training period for good reconstruction is required.	Cross-subject shared model can be applied.	Reconstruction is limited to natural scenes and susceptible to noise.	High precision with few samples is achieved.