Mathematical Symbols in PEAR

Table S1: Mathematical Symbols in PEAR: Phrase-Based Hand-Object Interaction Anticipation (Part 1)

Symbol	Description
Input and Output Interaction Elements	
$\mathcal{I}$	Pre-interaction image (interaction scenario)
$\mathcal{E}$	Phrase (interaction prompt)
$\mathcal{E}_V$	Verb component of the phrase
$\mathcal{E}_N$	Noun component of the phrase
$\mathcal{T}$	Hand motion trends (part of interaction intention)
$\mathcal{O}$	Interaction hotspots (part of interaction intention)
$\mathcal{M}$	Manipulation trajectories (part of interaction manipulation)
$\mathcal{P}$	Hand poses (part of interaction manipulation)
$\mathcal{C}$	Hand Contact (part of interaction manipulation)
Feature Representations	
$\mathbf{H}$	Hand motion feature
$\mathbf{O}$	Contact area feature
$\mathbf{F}_t$	Refined hand motion feature
$\mathbf{F}_h$	Interaction hotspots feature
$\hat{\mathbf{F}}_t$	Updated hand motion feature after cross-attention
$\hat{\mathbf{F}}_h$	Updated interaction hotspots feature after cross-attention
$\mathbf{F}_i$	Manipulation feature
$\mathbf{F}_m$	Manipulation trajectory feature
$\mathbf{F}_p$	Hand pose feature
$\mathbf{F}_c$	Hand contact feature
$\overline{\mathbf{F}}_t$	Refined hand motion feature after residual connection
$\overline{\mathbf{F}}_h$	Refined interaction hotspots feature after residual connection
Attention and Neural Network Notations	
$\mathbf{y}_i$	Intermediate feature in self-attention block
$\mathbf{x}_i$	Feature at layer i in self-attention block
$\mathbf{x}_0$	Original hand motion pattern feature
$\mathbf{x}_n$	Corrected feature after self-attention blocks
MSA	Multi-head Self-Attention function
LN	Layer normalization function
FFN	Feed Forward Layer function
MCA	Multi-head Cross-Attention function
$\mathbf{W}^Q$	Linear transformation for query in MSA
$\mathbf{W}^K$	Linear transformation for key in MSA
$\mathbf{W}^V$	Linear transformation for value in MSA

Table S2: Mathematical Symbols in PEAR: Phrase-Based Hand-Object Interaction Anticipation (Part 2)

Symbol	Description
DEQ (Deep Equilibrium) Extraction Model Parameters	
$\mathbf{z}^k$	Hidden feature at layer k in DEQ model
$\mathbf{z}^*$	Equilibrium state in DEQ model
f_θ	Residual block function in DEQ model
g_θ	Function representing the residual of equilibrium equation
$\mathbf{z}_h^0$	Initial hotspots feature in DEQ model
$\mathbf{z}_h^{k+1}$	Hotspots feature at layer k+1 in DEQ model
$\mathbf{z}_h^*$	Equilibrium hotspots feature in DEQ model
$\mathbf{z}_t^0$	Initial hand motion feature in DEQ model
$\mathbf{z}_t^{k+1}$	Hand motion feature at layer k+1 in DEQ model
$\mathbf{z}_t^*$	Equilibrium hand motion feature in DEQ model
$\mathbf{F}_i^0$	Initial manipulation feature in DEQ model
$\mathbf{F}_i^{k+1}$	Manipulation feature at layer k+1 in DEQ model
$\mathbf{F}_i^*$	Equilibrium manipulation feature in DEQ model
f_ψ	Fusion unit in DEQ model
f_ϕ	Nonlinear functions in DEQ model
C-VAE (Conditional Variational Autoencoder) Parameters	
$\mathbf{\Gamma}$	Input to C-VAE
$\mathbf{\Gamma}_e$	Input to C-VAE for specific element type $e \in \{t, h, p, c, m\}$
$\mathbf{\Lambda}$	Condition for C-VAE
$\mathbf{\Theta}$	Features in latent space in C-VAE
μ	Mean parameter in C-VAE
σ	Co-variance parameter in C-VAE
$\hat{\mathbf{\Gamma}}$	Reconstructed result from C-VAE
$\hat{\mathbf{\Gamma}}_e$	Reconstructed result from C-VAE for specific element type $e \in \{t, h, p, c, m\}$
f_e	Encoding function in C-VAE
f_d	Decoding function in C-VAE
$\mathcal{N}$	Normal distribution
Hand Position and Trajectory Parameters	
$\mathbf{\Lambda}_t^i$	Condition for hand motion trend at time step i
$\mathbf{S}_t^i$	Hand location at predicted time step i
$\mathbf{S}_t^0$	Initial hand position
n_c	Contact time step
$\mathbf{S}_m^i$	Hand position in manipulation trajectory at time step i
n_m	Final predicted step of manipulation trajectory
MANO Hand Model Parameters	
θ	Pose parameters of MANO model
β	Shape parameters of MANO model

Table S3: Mathematical Symbols in PEAR: Phrase-Based Hand-Object Interaction Anticipation (Part 3)

Symbol	Description
Loss Functions	
$\mathcal{L}$	Total training loss
$\mathcal{L}_t$	Loss for hand motion trends
$\mathcal{L}_h$	Loss for interaction hotspots
$\mathcal{L}_p$	Loss for hand pose
$\mathcal{L}_c$	Loss for hand contact
$\mathcal{L}_m$	Loss for manipulation trajectory
ω_{1-5}	Hyper-parameters to balance the losses
$\mathcal{L}_{recon}$	Generic reconstruction loss in C-VAE
$\mathcal{L}_{recon,e}$	Reconstruction loss for specific element type $e \in \{t, h, p, c, m\}$
$\mathcal{L}_{kl}$	KL divergence loss in C-VAE
$\mathcal{L}_{vae}$	Generic VAE-based loss
$\mathcal{L}_{vae,e}$	VAE-based loss for specific element type $e \in \{t, h, p, c, m\}$
λ	Weight for KL divergence
$\mathcal{L}_{joint}$	Joint loss for hand pose coordinates
ζ	Weight for joint loss
$\mathbf{U}_{pred}$	Predicted 3D hand joint coordinates
$\mathbf{U}_{gt}$	Ground truth 3D hand locations
n_o	Number of vertices in MANO mesh
ρ_i	Ground truth contact probability
$\hat{\rho}_i$	Predicted probability of vertex in contact state
Evaluation Metrics	
ADE	Average Displacement Error
FDE	Final Displacement Error
n	Total number of time steps
$\mathbf{p}_t$	Ground truth position
$\hat{\mathbf{p}}_t$	Predicted position
$\mathbf{p}_n$	Ground truth position at time step n
$\hat{\mathbf{p}}_n$	Predicted position at time step n
SIM	Similarity metric
AUC-J	Area Under Curve-Judd metric
NSS	Normalized Scanpath Saliency metric
PA-MPJPE	Procrustes Aligned Mean Per Joint Position Error
n_p	Number of pose joints
$\mathbf{R}$	Rotation matrix
$\mathbf{v}$	Translation vector
$\mathbf{p}_i$	Ground truth joint position
$\hat{\mathbf{p}}_i$	Predicted joint position