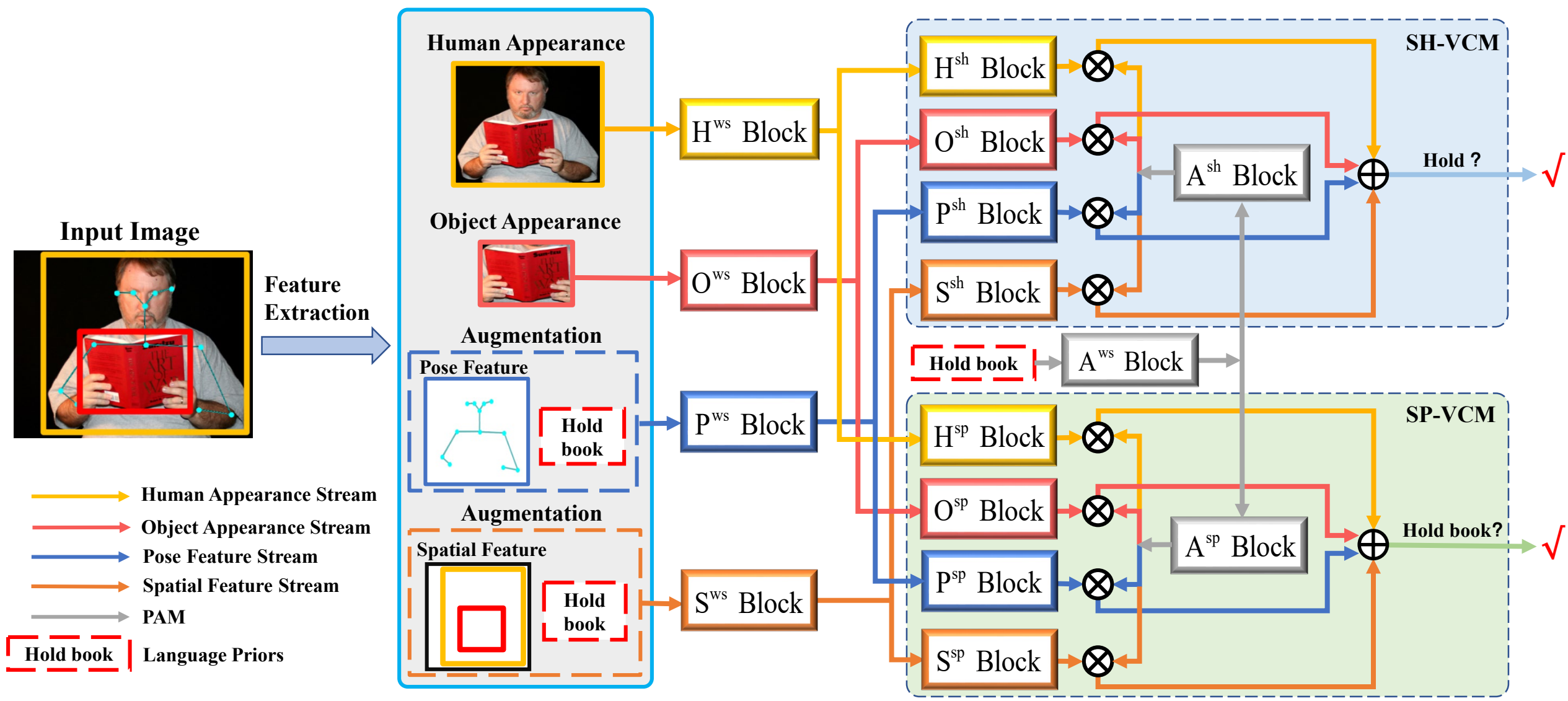
Input Image
Feature Extraction
Human Appearance
Object Appearance
Augmentation
Pose Feature
Hold book
Augmentation
Spatial Feature
Hold book
Human Appearance Stream
Object Appearance Stream
Pose Feature Stream
Spatial Feature Stream
PAM
Hold book
Language Priors
H^ws Block
O^ws Block
P^ws Block
S^ws Block
SH-VCM
H^sh Block
O^sh Block
P^sh Block
S^sh Block
A^sh Block
Hold ?
√
Hold book
A^ws Block
SP-VCM
H^sp Block
O^sp Block
P^sp Block
S^sp Block
A^sp Block
Hold book?
√