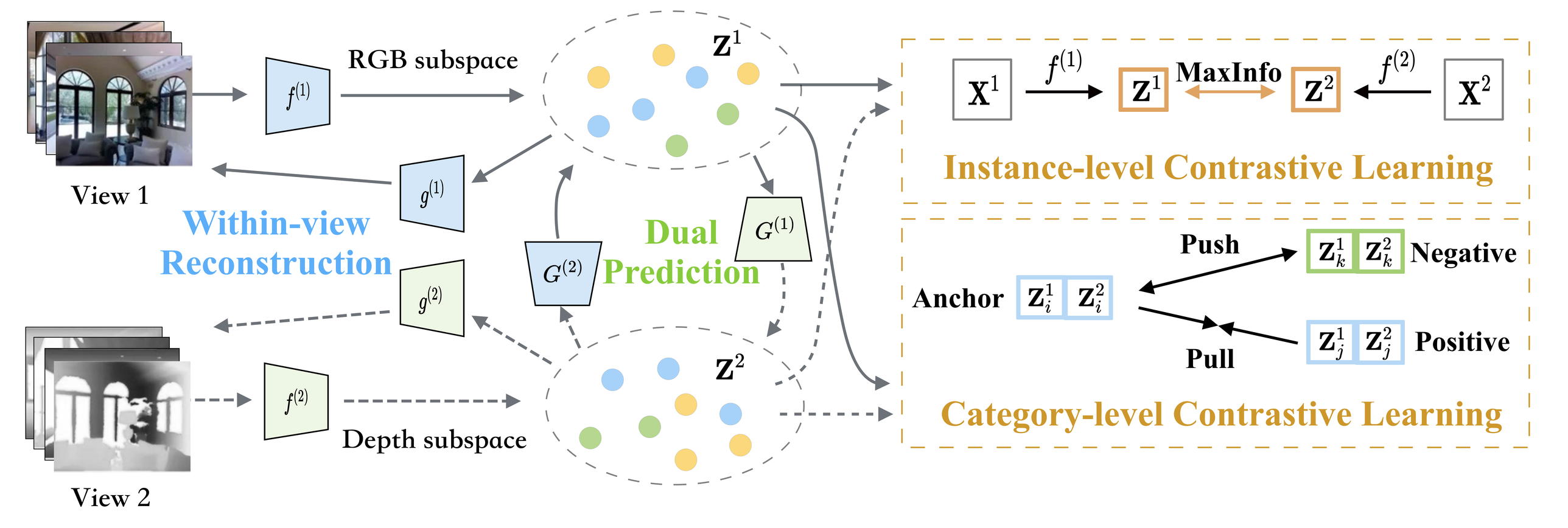
View 1
$f^{(1)}$
RGB subspace
$\mathbf{Z}^1$
$g^{(1)}$
Within-view
Reconstruction
$g^{(2)}$
$G^{(2)}$
Dual
Prediction
$G^{(1)}$
$\mathbf{Z}^2$
$f^{(2)}$
Depth subspace
View 2
$\mathbf{X}^1$
$f^{(1)}$
$\mathbf{Z}^1$
MaxInfo
$\mathbf{Z}^2$
$f^{(2)}$
$\mathbf{X}^2$
Instance-level Contrastive Learning
Push
$\mathbf{Z}_k^1$ $\mathbf{Z}_k^2$
Negative
Anchor
$\mathbf{Z}_i^1$ $\mathbf{Z}_i^2$
Pull
$\mathbf{Z}_j^1$ $\mathbf{Z}_j^2$
Positive
Category-level Contrastive Learning