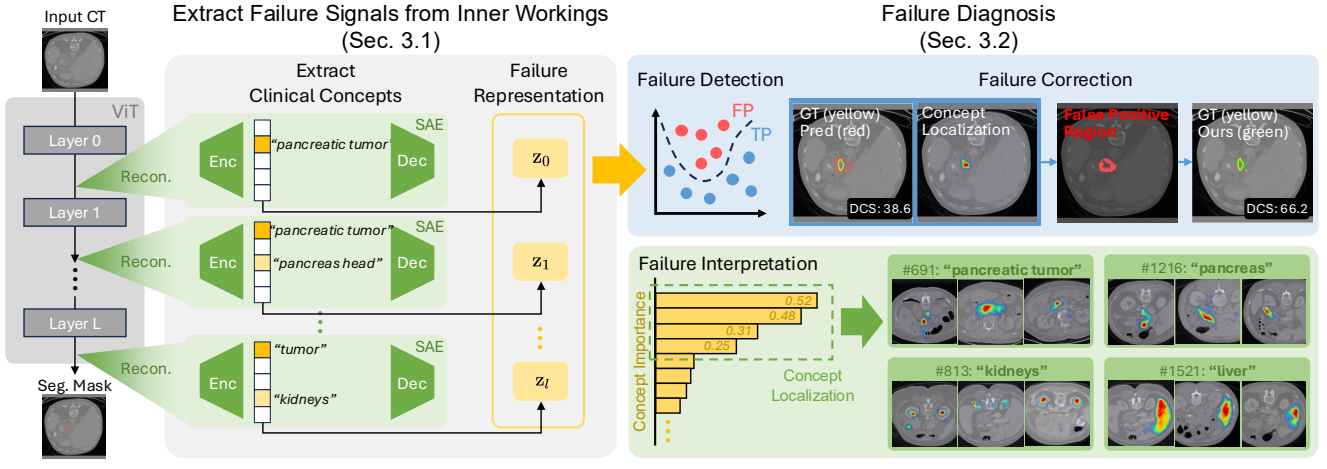
Input CT
ViT
Layer 0
Layer 1
Layer L
Recon.
Recon.
Recon.
Seg. Mask
Extract Failure Signals from Inner Workings
(Sec. 3.1)
Extract
Clinical Concepts
Failure
Representation
SAE
Enc
Dec
"pancreatic tumor"
"pancreatic tumor"
"pancreas head"
"tumor"
"kidneys"
$z_0$
$z_1$
$z_l$
Failure Diagnosis
(Sec. 3.2)
Failure Detection
FP
TP
Failure Correction
GT (yellow)
Pred (red)
DCS: 38.6
Concept
Localization
False Positive
Region
GT (yellow)
Ours (green)
DCS: 66.2
Failure Interpretation
Concept Importance
0.52
0.48
0.31
0.25
Concept
Localization
#691: "pancreatic tumor"
#1216: "pancreas"
#813: "kidneys"
#1521: "liver"