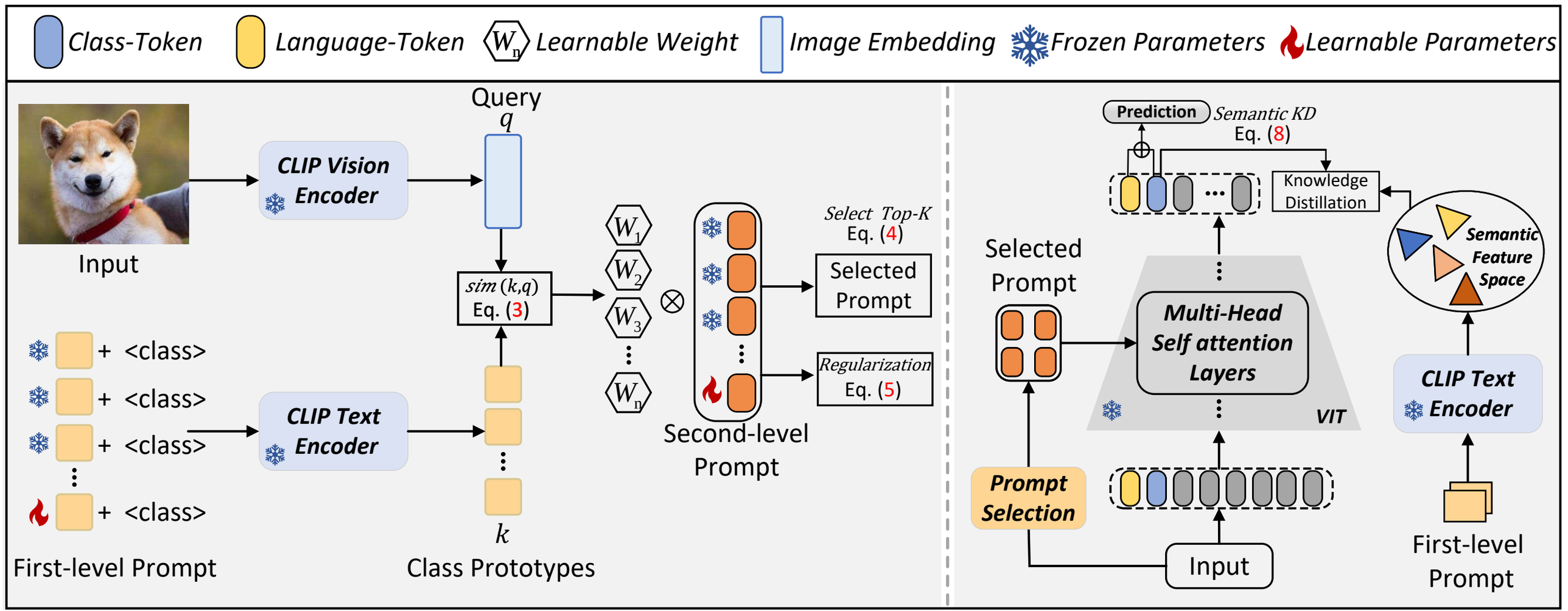
Class-Token
Language-Token
Learnable Weight
Image Embedding
Frozen Parameters
Learnable Parameters
Input
CLIP Vision Encoder
Query
q
sim (k,q)
Eq. (3)
+ <class>
+ <class>
+ <class>
+ <class>
First-level Prompt
CLIP Text Encoder
k
Class Prototypes
Second-level Prompt
Select Top-K
Eq. (4)
Selected Prompt
Regularization
Eq. (5)
Prediction
Semantic KD
Eq. (8)
Knowledge Distillation
Semantic Feature Space
Selected Prompt
Multi-Head Self attention Layers
VIT
CLIP Text Encoder
Prompt Selection
Input
First-level Prompt