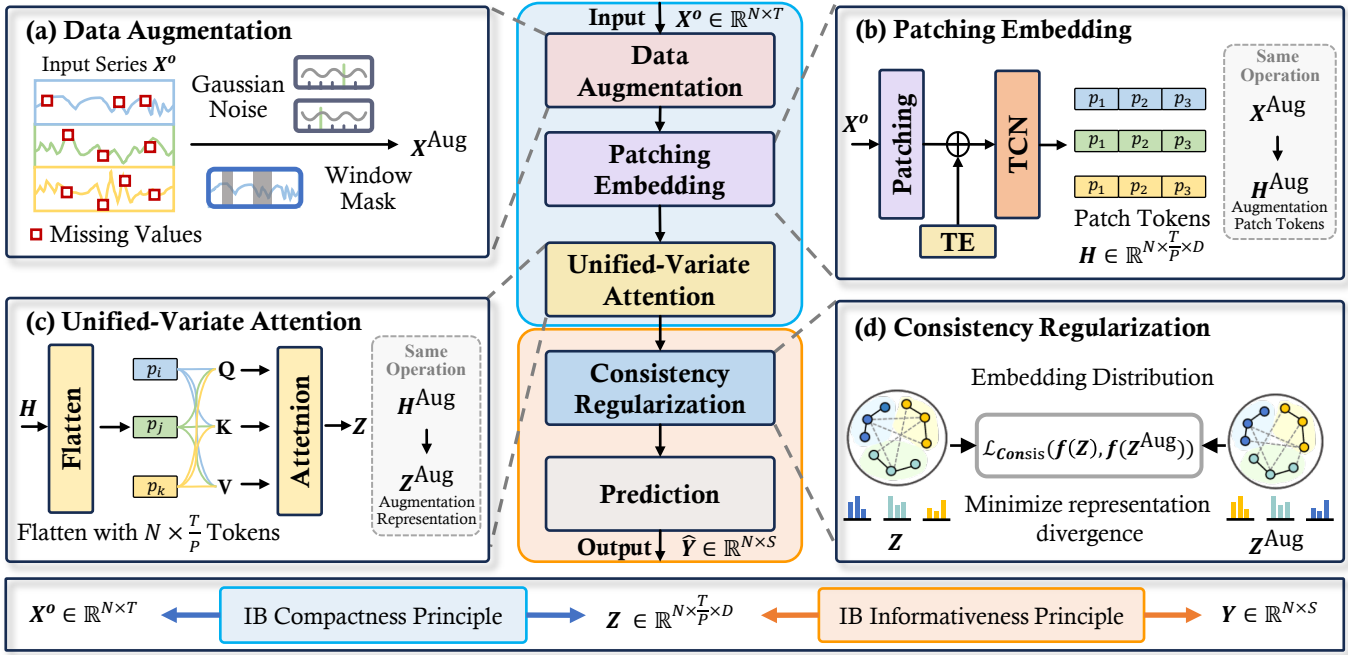

(a) Data Augmentation
Input Series X^o
Gaussian Noise
Window Mask
X^{Aug}
Missing Values
Input $X^o \in \mathbb{R}^{N\times T}$
Data Augmentation
Patching Embedding
Unified-Variate Attention
Consistency Regularization
Prediction
Output $\widehat{Y} \in \mathbb{R}^{N\times S}$
(b) Patching Embedding
X^o
Patching
TCN
TE
p_1 p_2 p_3
Patch Tokens
$H \in \mathbb{R}^{N\times \frac{T}{P}\times D}$
Same Operation
X^{Aug}
H^{Aug}
Augmentation Patch Tokens
(c) Unified-Variate Attention
H
Flatten
p_i
p_j
p_k
Q
K
V
Attetnion
Z
Flatten with $N \times \frac{T}{P}$ Tokens
Same Operation
H^{Aug}
Z^{Aug}
Augmentation Representation
(d) Consistency Regularization
Embedding Distribution
$\mathcal{L}_{Consis}(f(Z), f(Z^{\text{Aug}}))$
Minimize representation divergence
Z
Z^{Aug}
$X^o \in \mathbb{R}^{N\times T}$
IB Compactness Principle
$Z \in \mathbb{R}^{N\times \frac{T}{P}\times D}$
IB Informativeness Principle
$Y \in \mathbb{R}^{N\times S}$