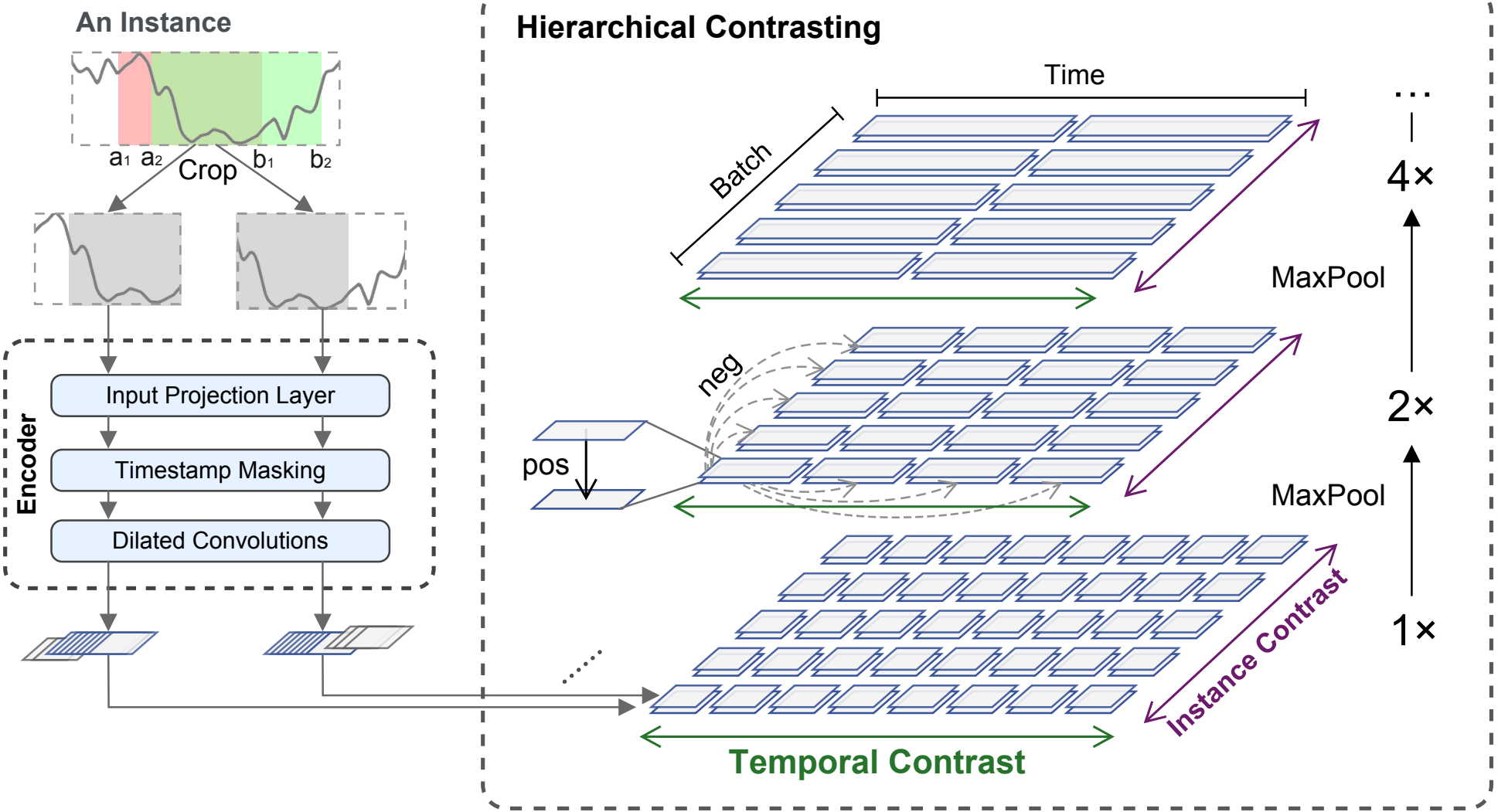
An Instance
a1 a2
Crop
b1 b2
Encoder
Input Projection Layer
Timestamp Masking
Dilated Convolutions
Hierarchical Contrasting
Time
Batch
MaxPool
MaxPool
4×
2×
1×
neg
pos
Instance Contrast
Temporal Contrast