

A. Vanilla Transformer

Output Vector

Nx

Add & Norm

Point-Wise Feed-Forward Network

Add & Norm

Multi-Head Full-Attention Mechanism

Canonical Point-Wise Dot-Product

Q K V

Positional Encoding

Input Sequence

B-Informer

Output Vector

Nx

Add & Norm

Point-Wise Feed-Forward Network

Add & Norm

Multi-Head ProbSparse-Attention Mechanism

Input Sequence

Positional Encoding

Scaled Point-Wise Dot-Product

Q

K

V

C. Autoformer

Output Vector

Nx

Add & Norm

Point-Wise Feed-Forward Network

Add & Norm

Multi-Head Auto-Correlation Mechanism

Positional Encoding

Input Sequence

Q

K

V

FFT

FFT⁻¹

TDA

D. Block-Recurrent Transformer

Horizontal Direction

The **Recurrent Layer** (orange box) processes the **Input Sequence** (blue circle with a slash) and **Positional Encoding** (blue circle with a plus) to produce the **Output Vector** (blue circle with a plus). The layer consists of $R \times$ (Recurrent) blocks. Each block contains a **Gate** (orange box) and a **Multi-Head Self-Attention** (blue box) block. The **Multi-Head Self-Attention** block is composed of **Cross-Attention** and **Full-Attention** sub-layers. The **Gate** block is followed by an **Add & Norm** (yellow box) block. The **Multi-Head Self-Attention** block is followed by a **Point-Wise Feed-Forward Network** (green box) and an **Add** (yellow box) block. The **Input Sequence** and **Positional Encoding** are combined and passed through a **Norm** (yellow box) block before entering the **Multi-Head Self-Attention** block.

Vertical Direction

The **Self-Attention Layer** (orange box) processes the **Current State Vectors** (blue circle with a plus) and **Positional Encoding** (blue circle with a slash) to produce the **Output Vector** (blue circle with a plus). The layer consists of $N \times$ (Self-Attention) blocks. Each block contains a **Gate** (orange box) and a **Multi-Head Self-Attention** (blue box) block. The **Multi-Head Self-Attention** block is composed of **Cross-Attention** and **Full-Attention** sub-layers. The **Gate** block is followed by an **Add & Norm** (yellow box) block. The **Multi-Head Self-Attention** block is followed by a **Point-Wise Feed-Forward Network** (green box) and an **Add** (yellow box) block. The **Current State Vectors** and **Positional Encoding** are combined and passed through a **Norm** (yellow box) block before entering the **Multi-Head Self-Attention** block.

E. THAT

The diagram illustrates the E. THAT architecture, which processes two different input representations: the Input Sequence (Nx) and the Input Sequence Transpose (Hx).

Temporal Module (Left): Processes the Input Sequence (Nx). It consists of a Multi-Head Full-Attention Mechanism (blue box) followed by two Add & Norm blocks (yellow boxes) and a HAR CNN (green box). The output of the HAR CNN is added to the input of the first Add & Norm block. The output of the second Add & Norm block is added to the input of the HAR CNN. The final output of the Temporal Module is a CNN (white box).

Channel Module (Right): Processes the Input Sequence Transpose (Hx). It consists of a Multi-Head Full-Attention Mechanism (blue box) followed by two Add & Norm blocks (yellow boxes) and a HAR CNN (green box). The output of the HAR CNN is added to the input of the first Add & Norm block. The output of the second Add & Norm block is added to the input of the HAR CNN. The final output of the Channel Module is a CNN (white box).

The outputs of both CNNs are concatenated and passed through a Softmax layer to produce the Output Vector.

The diagram illustrates the architecture of the Proposed Transformer, which processes an Input Sequence to produce an Output Vector. The architecture is divided into two main modules: the Temporal Module and the Channel Module.

Temporal Module: This module processes the Input Sequence through three parallel paths (Nx, Rx, and Mx) stacked vertically. Each path consists of a Multi-Head Auto-Correlation layer (blue), followed by a GBR CNN layer (green), and an Add & Norm layer (yellow). The output of the Nx path is added to the Input Sequence (indicated by a residual connection). The output of the Rx path is added to the output of the Nx path. The output of the Mx path is added to the output of the Rx path. The final output of the Temporal Module is then processed by a CNN layer.

Channel Module: This module processes the output of the Temporal Module through a single path (Hx) consisting of a Multi-Head Auto-Correlation Mechanism (blue), followed by a GBR CNN layer (green), and an Add & Norm layer (yellow). The output of the Hx path is added to the input of the Channel Module (indicated by a residual connection). The final output of the Channel Module is then processed by a CNN layer.

Input and Output: The Input Sequence is processed by a Gaussian Range Encoding layer (represented by a circle with a sine wave) and an Input Sequence Transpose layer (represented by a circle with a sine wave). The outputs of these two layers are added (indicated by a circle with a plus sign) to produce the input for the Temporal Module. The outputs of the CNN layers from both the Temporal and Channel Modules are concatenated and then passed through a Softmax layer to produce the final Output Vector.