Table of Contents

[Depth 3 1](#_Toc195964488)

[Depth 6 2](#_Toc195964489)

# Depth 3

==========================================================================

Layer (type:depth-idx) Param #

==========================================================================

EncoderDecoder --

├─Encoder: 1-1 --

│ └─ModuleList: 2-1 --

│ │ └─EncoderLayer: 3-1 3,152,384

│ │ └─EncoderLayer: 3-2 3,152,384

│ │ └─EncoderLayer: 3-3 3,152,384

│ │ └─EncoderLayer: 3-4 3,152,384

│ │ └─EncoderLayer: 3-5 3,152,384

│ │ └─EncoderLayer: 3-6 3,152,384

│ └─LayerNorm: 2-2 1,024

├─Decoder: 1-2 --

│ └─ModuleList: 2-3 --

│ │ └─DecoderLayer: 3-7 4,204,032

│ │ └─DecoderLayer: 3-8 4,204,032

│ │ └─DecoderLayer: 3-9 4,204,032

│ │ └─DecoderLayer: 3-10 4,204,032

│ │ └─DecoderLayer: 3-11 4,204,032

│ │ └─DecoderLayer: 3-12 4,204,032

│ └─LayerNorm: 2-4 1,024

├─Sequential: 1-3 --

│ └─Embeddings: 2-5 --

│ │ └─Embedding: 3-13 4,257,792

│ └─PositionalEncoding: 2-6 --

│ │ └─Dropout: 3-14 --

├─Sequential: 1-4 --

│ └─Embeddings: 2-7 --

│ │ └─Embedding: 3-15 3,268,608

│ └─PositionalEncoding: 2-8 --

│ │ └─Dropout: 3-16 --

├─Generator: 1-5 --

│ └─Linear: 2-9 3,274,992

==========================================================================

Total params: 54,941,936

Trainable params: 54,941,936

Non-trainable params: 0

# Depth 6

==========================================================================

Layer (type:depth-idx) Param #

==========================================================================

EncoderDecoder --

├─Encoder: 1-1 --

│ └─ModuleList: 2-1 --

│ │ └─EncoderLayer: 3-1 --

│ │ │ └─MultiHeadedAttention: 4-1 --

│ │ │ │ └─ModuleList: 5-1 --

│ │ │ │ │ └─Linear: 6-1 262,656

│ │ │ │ │ └─Linear: 6-2 262,656

│ │ │ │ │ └─Linear: 6-3 262,656

│ │ │ │ │ └─Linear: 6-4 262,656

│ │ │ │ └─Dropout: 5-2 --

│ │ │ └─PositionwiseFeedForward: 4-2 --

│ │ │ │ └─Linear: 5-3 1,050,624

│ │ │ │ └─Linear: 5-4 1,049,088

│ │ │ │ └─Dropout: 5-5 --

│ │ │ └─ModuleList: 4-3 --

│ │ │ │ └─SublayerConnection: 5-6 --

│ │ │ │ │ └─LayerNorm: 6-5 1,024

│ │ │ │ │ └─Dropout: 6-6 --

│ │ │ │ └─SublayerConnection: 5-7 --

│ │ │ │ │ └─LayerNorm: 6-7 1,024

│ │ │ │ │ └─Dropout: 6-8 --

│ │ └─EncoderLayer: 3-2 --

│ │ │ └─MultiHeadedAttention: 4-4 --

│ │ │ │ └─ModuleList: 5-8 --

│ │ │ │ │ └─Linear: 6-9 262,656

│ │ │ │ │ └─Linear: 6-10 262,656

│ │ │ │ │ └─Linear: 6-11 262,656

│ │ │ │ │ └─Linear: 6-12 262,656

│ │ │ │ └─Dropout: 5-9 --

│ │ │ └─PositionwiseFeedForward: 4-5 --

│ │ │ │ └─Linear: 5-10 1,050,624

│ │ │ │ └─Linear: 5-11 1,049,088

│ │ │ │ └─Dropout: 5-12 --

│ │ │ └─ModuleList: 4-6 --

│ │ │ │ └─SublayerConnection: 5-13 --

│ │ │ │ │ └─LayerNorm: 6-13 1,024

│ │ │ │ │ └─Dropout: 6-14 --

│ │ │ │ └─SublayerConnection: 5-14 --

│ │ │ │ │ └─LayerNorm: 6-15 1,024

│ │ │ │ │ └─Dropout: 6-16 --

│ │ └─EncoderLayer: 3-3 --

│ │ │ └─MultiHeadedAttention: 4-7 --

│ │ │ │ └─ModuleList: 5-15 --

│ │ │ │ │ └─Linear: 6-17 262,656

│ │ │ │ │ └─Linear: 6-18 262,656

│ │ │ │ │ └─Linear: 6-19 262,656

│ │ │ │ │ └─Linear: 6-20 262,656

│ │ │ │ └─Dropout: 5-16 --

│ │ │ └─PositionwiseFeedForward: 4-8 --

│ │ │ │ └─Linear: 5-17 1,050,624

│ │ │ │ └─Linear: 5-18 1,049,088

│ │ │ │ └─Dropout: 5-19 --

│ │ │ └─ModuleList: 4-9 --

│ │ │ │ └─SublayerConnection: 5-20 --

│ │ │ │ │ └─LayerNorm: 6-21 1,024

│ │ │ │ │ └─Dropout: 6-22 --

│ │ │ │ └─SublayerConnection: 5-21 --

│ │ │ │ │ └─LayerNorm: 6-23 1,024

│ │ │ │ │ └─Dropout: 6-24 --

│ │ └─EncoderLayer: 3-4 --

│ │ │ └─MultiHeadedAttention: 4-10 --

│ │ │ │ └─ModuleList: 5-22 --

│ │ │ │ │ └─Linear: 6-25 262,656

│ │ │ │ │ └─Linear: 6-26 262,656

│ │ │ │ │ └─Linear: 6-27 262,656

│ │ │ │ │ └─Linear: 6-28 262,656

│ │ │ │ └─Dropout: 5-23 --

│ │ │ └─PositionwiseFeedForward: 4-11 --

│ │ │ │ └─Linear: 5-24 1,050,624

│ │ │ │ └─Linear: 5-25 1,049,088

│ │ │ │ └─Dropout: 5-26 --

│ │ │ └─ModuleList: 4-12 --

│ │ │ │ └─SublayerConnection: 5-27 --

│ │ │ │ │ └─LayerNorm: 6-29 1,024

│ │ │ │ │ └─Dropout: 6-30 --

│ │ │ │ └─SublayerConnection: 5-28 --

│ │ │ │ │ └─LayerNorm: 6-31 1,024

│ │ │ │ │ └─Dropout: 6-32 --

│ │ └─EncoderLayer: 3-5 --

│ │ │ └─MultiHeadedAttention: 4-13 --

│ │ │ │ └─ModuleList: 5-29 --

│ │ │ │ │ └─Linear: 6-33 262,656

│ │ │ │ │ └─Linear: 6-34 262,656

│ │ │ │ │ └─Linear: 6-35 262,656

│ │ │ │ │ └─Linear: 6-36 262,656

│ │ │ │ └─Dropout: 5-30 --

│ │ │ └─PositionwiseFeedForward: 4-14 --

│ │ │ │ └─Linear: 5-31 1,050,624

│ │ │ │ └─Linear: 5-32 1,049,088

│ │ │ │ └─Dropout: 5-33 --

│ │ │ └─ModuleList: 4-15 --

│ │ │ │ └─SublayerConnection: 5-34 --

│ │ │ │ │ └─LayerNorm: 6-37 1,024

│ │ │ │ │ └─Dropout: 6-38 --

│ │ │ │ └─SublayerConnection: 5-35 --

│ │ │ │ │ └─LayerNorm: 6-39 1,024

│ │ │ │ │ └─Dropout: 6-40 --

│ │ └─EncoderLayer: 3-6 --

│ │ │ └─MultiHeadedAttention: 4-16 --

│ │ │ │ └─ModuleList: 5-36 --

│ │ │ │ │ └─Linear: 6-41 262,656

│ │ │ │ │ └─Linear: 6-42 262,656

│ │ │ │ │ └─Linear: 6-43 262,656

│ │ │ │ │ └─Linear: 6-44 262,656

│ │ │ │ └─Dropout: 5-37 --

│ │ │ └─PositionwiseFeedForward: 4-17 --

│ │ │ │ └─Linear: 5-38 1,050,624

│ │ │ │ └─Linear: 5-39 1,049,088

│ │ │ │ └─Dropout: 5-40 --

│ │ │ └─ModuleList: 4-18 --

│ │ │ │ └─SublayerConnection: 5-41 --

│ │ │ │ │ └─LayerNorm: 6-45 1,024

│ │ │ │ │ └─Dropout: 6-46 --

│ │ │ │ └─SublayerConnection: 5-42 --

│ │ │ │ │ └─LayerNorm: 6-47 1,024

│ │ │ │ │ └─Dropout: 6-48 --

│ └─LayerNorm: 2-2 1,024

├─Decoder: 1-2 --

│ └─ModuleList: 2-3 --

│ │ └─DecoderLayer: 3-7 --

│ │ │ └─MultiHeadedAttention: 4-19 --

│ │ │ │ └─ModuleList: 5-43 --

│ │ │ │ │ └─Linear: 6-49 262,656

│ │ │ │ │ └─Linear: 6-50 262,656

│ │ │ │ │ └─Linear: 6-51 262,656

│ │ │ │ │ └─Linear: 6-52 262,656

│ │ │ │ └─Dropout: 5-44 --

│ │ │ └─MultiHeadedAttention: 4-20 --

│ │ │ │ └─ModuleList: 5-45 --

│ │ │ │ │ └─Linear: 6-53 262,656

│ │ │ │ │ └─Linear: 6-54 262,656

│ │ │ │ │ └─Linear: 6-55 262,656

│ │ │ │ │ └─Linear: 6-56 262,656

│ │ │ │ └─Dropout: 5-46 --

│ │ │ └─PositionwiseFeedForward: 4-21 --

│ │ │ │ └─Linear: 5-47 1,050,624

│ │ │ │ └─Linear: 5-48 1,049,088

│ │ │ │ └─Dropout: 5-49 --

│ │ │ └─ModuleList: 4-22 --

│ │ │ │ └─SublayerConnection: 5-50 --

│ │ │ │ │ └─LayerNorm: 6-57 1,024

│ │ │ │ │ └─Dropout: 6-58 --

│ │ │ │ └─SublayerConnection: 5-51 --

│ │ │ │ │ └─LayerNorm: 6-59 1,024

│ │ │ │ │ └─Dropout: 6-60 --

│ │ │ │ └─SublayerConnection: 5-52 --

│ │ │ │ │ └─LayerNorm: 6-61 1,024

│ │ │ │ │ └─Dropout: 6-62 --

│ │ └─DecoderLayer: 3-8 --

│ │ │ └─MultiHeadedAttention: 4-23 --

│ │ │ │ └─ModuleList: 5-53 --

│ │ │ │ │ └─Linear: 6-63 262,656

│ │ │ │ │ └─Linear: 6-64 262,656

│ │ │ │ │ └─Linear: 6-65 262,656

│ │ │ │ │ └─Linear: 6-66 262,656

│ │ │ │ └─Dropout: 5-54 --

│ │ │ └─MultiHeadedAttention: 4-24 --

│ │ │ │ └─ModuleList: 5-55 --

│ │ │ │ │ └─Linear: 6-67 262,656

│ │ │ │ │ └─Linear: 6-68 262,656

│ │ │ │ │ └─Linear: 6-69 262,656

│ │ │ │ │ └─Linear: 6-70 262,656

│ │ │ │ └─Dropout: 5-56 --

│ │ │ └─PositionwiseFeedForward: 4-25 --

│ │ │ │ └─Linear: 5-57 1,050,624

│ │ │ │ └─Linear: 5-58 1,049,088

│ │ │ │ └─Dropout: 5-59 --

│ │ │ └─ModuleList: 4-26 --

│ │ │ │ └─SublayerConnection: 5-60 --

│ │ │ │ │ └─LayerNorm: 6-71 1,024

│ │ │ │ │ └─Dropout: 6-72 --

│ │ │ │ └─SublayerConnection: 5-61 --

│ │ │ │ │ └─LayerNorm: 6-73 1,024

│ │ │ │ │ └─Dropout: 6-74 --

│ │ │ │ └─SublayerConnection: 5-62 --

│ │ │ │ │ └─LayerNorm: 6-75 1,024

│ │ │ │ │ └─Dropout: 6-76 --

│ │ └─DecoderLayer: 3-9 --

│ │ │ └─MultiHeadedAttention: 4-27 --

│ │ │ │ └─ModuleList: 5-63 --

│ │ │ │ │ └─Linear: 6-77 262,656

│ │ │ │ │ └─Linear: 6-78 262,656

│ │ │ │ │ └─Linear: 6-79 262,656

│ │ │ │ │ └─Linear: 6-80 262,656

│ │ │ │ └─Dropout: 5-64 --

│ │ │ └─MultiHeadedAttention: 4-28 --

│ │ │ │ └─ModuleList: 5-65 --

│ │ │ │ │ └─Linear: 6-81 262,656

│ │ │ │ │ └─Linear: 6-82 262,656

│ │ │ │ │ └─Linear: 6-83 262,656

│ │ │ │ │ └─Linear: 6-84 262,656

│ │ │ │ └─Dropout: 5-66 --

│ │ │ └─PositionwiseFeedForward: 4-29 --

│ │ │ │ └─Linear: 5-67 1,050,624

│ │ │ │ └─Linear: 5-68 1,049,088

│ │ │ │ └─Dropout: 5-69 --

│ │ │ └─ModuleList: 4-30 --

│ │ │ │ └─SublayerConnection: 5-70 --

│ │ │ │ │ └─LayerNorm: 6-85 1,024

│ │ │ │ │ └─Dropout: 6-86 --

│ │ │ │ └─SublayerConnection: 5-71 --

│ │ │ │ │ └─LayerNorm: 6-87 1,024

│ │ │ │ │ └─Dropout: 6-88 --

│ │ │ │ └─SublayerConnection: 5-72 --

│ │ │ │ │ └─LayerNorm: 6-89 1,024

│ │ │ │ │ └─Dropout: 6-90 --

│ │ └─DecoderLayer: 3-10 --

│ │ │ └─MultiHeadedAttention: 4-31 --

│ │ │ │ └─ModuleList: 5-73 --

│ │ │ │ │ └─Linear: 6-91 262,656

│ │ │ │ │ └─Linear: 6-92 262,656

│ │ │ │ │ └─Linear: 6-93 262,656

│ │ │ │ │ └─Linear: 6-94 262,656

│ │ │ │ └─Dropout: 5-74 --

│ │ │ └─MultiHeadedAttention: 4-32 --

│ │ │ │ └─ModuleList: 5-75 --

│ │ │ │ │ └─Linear: 6-95 262,656

│ │ │ │ │ └─Linear: 6-96 262,656

│ │ │ │ │ └─Linear: 6-97 262,656

│ │ │ │ │ └─Linear: 6-98 262,656

│ │ │ │ └─Dropout: 5-76 --

│ │ │ └─PositionwiseFeedForward: 4-33 --

│ │ │ │ └─Linear: 5-77 1,050,624

│ │ │ │ └─Linear: 5-78 1,049,088

│ │ │ │ └─Dropout: 5-79 --

│ │ │ └─ModuleList: 4-34 --

│ │ │ │ └─SublayerConnection: 5-80 --

│ │ │ │ │ └─LayerNorm: 6-99 1,024

│ │ │ │ │ └─Dropout: 6-100 --

│ │ │ │ └─SublayerConnection: 5-81 --

│ │ │ │ │ └─LayerNorm: 6-101 1,024

│ │ │ │ │ └─Dropout: 6-102 --

│ │ │ │ └─SublayerConnection: 5-82 --

│ │ │ │ │ └─LayerNorm: 6-103 1,024

│ │ │ │ │ └─Dropout: 6-104 --

│ │ └─DecoderLayer: 3-11 --

│ │ │ └─MultiHeadedAttention: 4-35 --

│ │ │ │ └─ModuleList: 5-83 --

│ │ │ │ │ └─Linear: 6-105 262,656

│ │ │ │ │ └─Linear: 6-106 262,656

│ │ │ │ │ └─Linear: 6-107 262,656

│ │ │ │ │ └─Linear: 6-108 262,656

│ │ │ │ └─Dropout: 5-84 --

│ │ │ └─MultiHeadedAttention: 4-36 --

│ │ │ │ └─ModuleList: 5-85 --

│ │ │ │ │ └─Linear: 6-109 262,656

│ │ │ │ │ └─Linear: 6-110 262,656

│ │ │ │ │ └─Linear: 6-111 262,656

│ │ │ │ │ └─Linear: 6-112 262,656

│ │ │ │ └─Dropout: 5-86 --

│ │ │ └─PositionwiseFeedForward: 4-37 --

│ │ │ │ └─Linear: 5-87 1,050,624

│ │ │ │ └─Linear: 5-88 1,049,088

│ │ │ │ └─Dropout: 5-89 --

│ │ │ └─ModuleList: 4-38 --

│ │ │ │ └─SublayerConnection: 5-90 --

│ │ │ │ │ └─LayerNorm: 6-113 1,024

│ │ │ │ │ └─Dropout: 6-114 --

│ │ │ │ └─SublayerConnection: 5-91 --

│ │ │ │ │ └─LayerNorm: 6-115 1,024

│ │ │ │ │ └─Dropout: 6-116 --

│ │ │ │ └─SublayerConnection: 5-92 --

│ │ │ │ │ └─LayerNorm: 6-117 1,024

│ │ │ │ │ └─Dropout: 6-118 --

│ │ └─DecoderLayer: 3-12 --

│ │ │ └─MultiHeadedAttention: 4-39 --

│ │ │ │ └─ModuleList: 5-93 --

│ │ │ │ │ └─Linear: 6-119 262,656

│ │ │ │ │ └─Linear: 6-120 262,656

│ │ │ │ │ └─Linear: 6-121 262,656

│ │ │ │ │ └─Linear: 6-122 262,656

│ │ │ │ └─Dropout: 5-94 --

│ │ │ └─MultiHeadedAttention: 4-40 --

│ │ │ │ └─ModuleList: 5-95 --

│ │ │ │ │ └─Linear: 6-123 262,656

│ │ │ │ │ └─Linear: 6-124 262,656

│ │ │ │ │ └─Linear: 6-125 262,656

│ │ │ │ │ └─Linear: 6-126 262,656

│ │ │ │ └─Dropout: 5-96 --

│ │ │ └─PositionwiseFeedForward: 4-41 --

│ │ │ │ └─Linear: 5-97 1,050,624

│ │ │ │ └─Linear: 5-98 1,049,088

│ │ │ │ └─Dropout: 5-99 --

│ │ │ └─ModuleList: 4-42 --

│ │ │ │ └─SublayerConnection: 5-100 --

│ │ │ │ │ └─LayerNorm: 6-127 1,024

│ │ │ │ │ └─Dropout: 6-128 --

│ │ │ │ └─SublayerConnection: 5-101 --

│ │ │ │ │ └─LayerNorm: 6-129 1,024

│ │ │ │ │ └─Dropout: 6-130 --

│ │ │ │ └─SublayerConnection: 5-102 --

│ │ │ │ │ └─LayerNorm: 6-131 1,024

│ │ │ │ │ └─Dropout: 6-132 --

│ └─LayerNorm: 2-4 1,024

├─Sequential: 1-3 --

│ └─Embeddings: 2-5 --

│ │ └─Embedding: 3-13 4,257,792

│ └─PositionalEncoding: 2-6 --

│ │ └─Dropout: 3-14 --

├─Sequential: 1-4 --

│ └─Embeddings: 2-7 --

│ │ └─Embedding: 3-15 3,268,608

│ └─PositionalEncoding: 2-8 --

│ │ └─Dropout: 3-16 --

├─Generator: 1-5 --

│ └─Linear: 2-9 3,274,992

==========================================================================

Total params: 54,941,936

Trainable params: 54,941,936

Non-trainable params: 0