```
ImageBindModel(
  (modality_preprocessors): ModuleDict(
    (vision): RGBDTPreprocessor(
      (cls_token): tensor((1, 1, 1280), requires_grad=True)

      (rgbt_stem): PatchEmbedGeneric(
        (proj): Sequential(
          (0): PadIm2Video()
          (1): Conv3d(3, 1280, kernel_size=(2, 14, 14), stride=(2, 14, 14), bias=False)
        )
      )
      (pos_embedding_helper): SpatioTemporalPosEmbeddingHelper(
        (pos_embed): tensor((1, 257, 1280), requires_grad=True)

      )
    )
    (text): TextPreprocessor(
      (pos_embed): tensor((1, 77, 1024), requires_grad=True)
      (mask): tensor((77, 77), requires_grad=False)

      (token_embedding): Embedding(49408, 1024)
    )
    (audio): AudioPreprocessor(
      (cls_token): tensor((1, 1, 768), requires_grad=True)

      (rgbt_stem): PatchEmbedGeneric(
        (proj): Conv2d(1, 768, kernel_size=(16, 16), stride=(10, 10), bias=False)
        (norm_layer): LayerNorm((768,), eps=1e-05, elementwise_affine=True)
      )
      (pos_embedding_helper): SpatioTemporalPosEmbeddingHelper(
        (pos_embed): tensor((1, 229, 768), requires_grad=True)

      )
    )
    (depth): RGBDTPreprocessor(
      (cls_token): tensor((1, 1, 384), requires_grad=True)

      (depth_stem): PatchEmbedGeneric(
        (proj): Conv2d(1, 384, kernel_size=(16, 16), stride=(16, 16), bias=False)
        (norm_layer): LayerNorm((384,), eps=1e-05, elementwise_affine=True)
      )
      (pos_embedding_helper): SpatioTemporalPosEmbeddingHelper(
        (pos_embed): tensor((1, 197, 384), requires_grad=True)

      )
    )
```

```
    (thermal): ThermalPreprocessor(
      (cls_token): tensor((1, 1, 768), requires_grad=True)

      (rgbt_stem): PatchEmbedGeneric(
        (proj): Conv2d(1, 768, kernel_size=(16, 16), stride=(16, 16), bias=False)
        (norm_layer): LayerNorm((768,), eps=1e-05, elementwise_affine=True)
      )
      (pos_embedding_helper): SpatioTemporalPosEmbeddingHelper(
        (pos_embed): tensor((1, 197, 768), requires_grad=True)

      )
    )
    (imu): IMUPreprocessor(
      (pos_embed): tensor((1, 251, 512), requires_grad=True)
      (cls_token): tensor((1, 1, 512), requires_grad=True)

      (imu_stem): PatchEmbedGeneric(
        (proj): Linear(in_features=48, out_features=512, bias=False)
        (norm_layer): LayerNorm((512,), eps=1e-05, elementwise_affine=True)
      )
    )
  )
  (modality_trunks): ModuleDict(
    (vision): LoRA_SimpleTransformer(
      (lora_model): SimpleTransformer(
        (pre_transformer_layer): Sequential(
          (0): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (1): EinOpsRearrange()
        )
        (blocks): Sequential(
          (0): BlockWithMasking(
            (attn): _LoRALayer(
              (w): MultiheadAttention(
                (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
              )
              (w_a): Linear(in_features=1280, out_features=4, bias=False)
              (w_b): Linear(in_features=4, out_features=1280, bias=False)
            )
            (drop_path): Identity()
            (norm_1): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
            (mlp): Mlp(
              (fc1): Linear(in_features=1280, out_features=5120, bias=True)
              (act): GELU(approximate='none')
              (fc2): Linear(in_features=5120, out_features=1280, bias=True)
              (drop): Dropout(p=0.0, inplace=False)
```

```
          )
          (norm_2): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (prev_attn): MultiheadAttention(
            (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
          )
        )
        (1): BlockWithMasking(
          (attn): _LoRALayer(
            (w): MultiheadAttention(
              (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
            )
            (w_a): Linear(in_features=1280, out_features=4, bias=False)
            (w_b): Linear(in_features=4, out_features=1280, bias=False)
          )
          (drop_path): Identity()
          (norm_1): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (mlp): Mlp(
            (fc1): Linear(in_features=1280, out_features=5120, bias=True)
            (act): GELU(approximate='none')
            (fc2): Linear(in_features=5120, out_features=1280, bias=True)
            (drop): Dropout(p=0.0, inplace=False)
          )
          (norm_2): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (prev_attn): MultiheadAttention(
            (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
          )
        )
        (2): BlockWithMasking(
          (attn): _LoRALayer(
            (w): MultiheadAttention(
              (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
            )
            (w_a): Linear(in_features=1280, out_features=4, bias=False)
            (w_b): Linear(in_features=4, out_features=1280, bias=False)
          )
          (drop_path): Identity()
          (norm_1): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (mlp): Mlp(
            (fc1): Linear(in_features=1280, out_features=5120, bias=True)
            (act): GELU(approximate='none')
            (fc2): Linear(in_features=5120, out_features=1280, bias=True)
            (drop): Dropout(p=0.0, inplace=False)
```

```
          )
          (norm_2): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (prev_attn): MultiheadAttention(
            (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
          )
        )
        (3): BlockWithMasking(
          (attn): _LoRALayer(
            (w): MultiheadAttention(
              (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
            )
            (w_a): Linear(in_features=1280, out_features=4, bias=False)
            (w_b): Linear(in_features=4, out_features=1280, bias=False)
          )
          (drop_path): Identity()
          (norm_1): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (mlp): Mlp(
            (fc1): Linear(in_features=1280, out_features=5120, bias=True)
            (act): GELU(approximate='none')
            (fc2): Linear(in_features=5120, out_features=1280, bias=True)
            (drop): Dropout(p=0.0, inplace=False)
          )
          (norm_2): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (prev_attn): MultiheadAttention(
            (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
          )
        )
        (4): BlockWithMasking(
          (attn): _LoRALayer(
            (w): MultiheadAttention(
              (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
            )
            (w_a): Linear(in_features=1280, out_features=4, bias=False)
            (w_b): Linear(in_features=4, out_features=1280, bias=False)
          )
          (drop_path): Identity()
          (norm_1): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (mlp): Mlp(
            (fc1): Linear(in_features=1280, out_features=5120, bias=True)
            (act): GELU(approximate='none')
            (fc2): Linear(in_features=5120, out_features=1280, bias=True)
            (drop): Dropout(p=0.0, inplace=False)
```

```
          )
          (norm_2): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (prev_attn): MultiheadAttention(
            (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
          )
        )
        (5): BlockWithMasking(
          (attn): _LoRALayer(
            (w): MultiheadAttention(
              (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
            )
            (w_a): Linear(in_features=1280, out_features=4, bias=False)
            (w_b): Linear(in_features=4, out_features=1280, bias=False)
          )
          (drop_path): Identity()
          (norm_1): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (mlp): Mlp(
            (fc1): Linear(in_features=1280, out_features=5120, bias=True)
            (act): GELU(approximate='none')
            (fc2): Linear(in_features=5120, out_features=1280, bias=True)
            (drop): Dropout(p=0.0, inplace=False)
          )
          (norm_2): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (prev_attn): MultiheadAttention(
            (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
          )
        )
        (6): BlockWithMasking(
          (attn): _LoRALayer(
            (w): MultiheadAttention(
              (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
            )
            (w_a): Linear(in_features=1280, out_features=4, bias=False)
            (w_b): Linear(in_features=4, out_features=1280, bias=False)
          )
          (drop_path): Identity()
          (norm_1): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (mlp): Mlp(
            (fc1): Linear(in_features=1280, out_features=5120, bias=True)
            (act): GELU(approximate='none')
            (fc2): Linear(in_features=5120, out_features=1280, bias=True)
            (drop): Dropout(p=0.0, inplace=False)
```

```
          )
          (norm_2): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (prev_attn): MultiheadAttention(
            (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
          )
        )
        (7): BlockWithMasking(
          (attn): _LoRALayer(
            (w): MultiheadAttention(
              (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
            )
            (w_a): Linear(in_features=1280, out_features=4, bias=False)
            (w_b): Linear(in_features=4, out_features=1280, bias=False)
          )
          (drop_path): Identity()
          (norm_1): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (mlp): Mlp(
            (fc1): Linear(in_features=1280, out_features=5120, bias=True)
            (act): GELU(approximate='none')
            (fc2): Linear(in_features=5120, out_features=1280, bias=True)
            (drop): Dropout(p=0.0, inplace=False)
          )
          (norm_2): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (prev_attn): MultiheadAttention(
            (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
          )
        )
        (8): BlockWithMasking(
          (attn): _LoRALayer(
            (w): MultiheadAttention(
              (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
            )
            (w_a): Linear(in_features=1280, out_features=4, bias=False)
            (w_b): Linear(in_features=4, out_features=1280, bias=False)
          )
          (drop_path): Identity()
          (norm_1): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (mlp): Mlp(
            (fc1): Linear(in_features=1280, out_features=5120, bias=True)
            (act): GELU(approximate='none')
            (fc2): Linear(in_features=5120, out_features=1280, bias=True)
            (drop): Dropout(p=0.0, inplace=False)
```

```
          )
          (norm_2): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (prev_attn): MultiheadAttention(
            (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
          )
        )
        (9): BlockWithMasking(
          (attn): _LoRALayer(
            (w): MultiheadAttention(
              (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
            )
            (w_a): Linear(in_features=1280, out_features=4, bias=False)
            (w_b): Linear(in_features=4, out_features=1280, bias=False)
          )
          (drop_path): Identity()
          (norm_1): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (mlp): Mlp(
            (fc1): Linear(in_features=1280, out_features=5120, bias=True)
            (act): GELU(approximate='none')
            (fc2): Linear(in_features=5120, out_features=1280, bias=True)
            (drop): Dropout(p=0.0, inplace=False)
          )
          (norm_2): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (prev_attn): MultiheadAttention(
            (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
          )
        )
        (10): BlockWithMasking(
          (attn): _LoRALayer(
            (w): MultiheadAttention(
              (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
            )
            (w_a): Linear(in_features=1280, out_features=4, bias=False)
            (w_b): Linear(in_features=4, out_features=1280, bias=False)
          )
          (drop_path): Identity()
          (norm_1): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (mlp): Mlp(
            (fc1): Linear(in_features=1280, out_features=5120, bias=True)
            (act): GELU(approximate='none')
            (fc2): Linear(in_features=5120, out_features=1280, bias=True)
            (drop): Dropout(p=0.0, inplace=False)
```

```
          )
          (norm_2): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (prev_attn): MultiheadAttention(
            (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
          )
        )
        (11): BlockWithMasking(
          (attn): _LoRALayer(
            (w): MultiheadAttention(
              (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
            )
            (w_a): Linear(in_features=1280, out_features=4, bias=False)
            (w_b): Linear(in_features=4, out_features=1280, bias=False)
          )
          (drop_path): Identity()
          (norm_1): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (mlp): Mlp(
            (fc1): Linear(in_features=1280, out_features=5120, bias=True)
            (act): GELU(approximate='none')
            (fc2): Linear(in_features=5120, out_features=1280, bias=True)
            (drop): Dropout(p=0.0, inplace=False)
          )
          (norm_2): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (prev_attn): MultiheadAttention(
            (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
          )
        )
        (12): BlockWithMasking(
          (attn): _LoRALayer(
            (w): MultiheadAttention(
              (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
            )
            (w_a): Linear(in_features=1280, out_features=4, bias=False)
            (w_b): Linear(in_features=4, out_features=1280, bias=False)
          )
          (drop_path): Identity()
          (norm_1): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (mlp): Mlp(
            (fc1): Linear(in_features=1280, out_features=5120, bias=True)
            (act): GELU(approximate='none')
            (fc2): Linear(in_features=5120, out_features=1280, bias=True)
            (drop): Dropout(p=0.0, inplace=False)
```

```
          )
          (norm_2): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (prev_attn): MultiheadAttention(
            (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
          )
        )
        (13): BlockWithMasking(
          (attn): _LoRALayer(
            (w): MultiheadAttention(
              (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
            )
            (w_a): Linear(in_features=1280, out_features=4, bias=False)
            (w_b): Linear(in_features=4, out_features=1280, bias=False)
          )
          (drop_path): Identity()
          (norm_1): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (mlp): Mlp(
            (fc1): Linear(in_features=1280, out_features=5120, bias=True)
            (act): GELU(approximate='none')
            (fc2): Linear(in_features=5120, out_features=1280, bias=True)
            (drop): Dropout(p=0.0, inplace=False)
          )
          (norm_2): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (prev_attn): MultiheadAttention(
            (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
          )
        )
        (14): BlockWithMasking(
          (attn): _LoRALayer(
            (w): MultiheadAttention(
              (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
            )
            (w_a): Linear(in_features=1280, out_features=4, bias=False)
            (w_b): Linear(in_features=4, out_features=1280, bias=False)
          )
          (drop_path): Identity()
          (norm_1): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (mlp): Mlp(
            (fc1): Linear(in_features=1280, out_features=5120, bias=True)
            (act): GELU(approximate='none')
            (fc2): Linear(in_features=5120, out_features=1280, bias=True)
            (drop): Dropout(p=0.0, inplace=False)
```

```
          )
          (norm_2): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (prev_attn): MultiheadAttention(
            (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
          )
        )
        (15): BlockWithMasking(
          (attn): _LoRALayer(
            (w): MultiheadAttention(
              (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
            )
            (w_a): Linear(in_features=1280, out_features=4, bias=False)
            (w_b): Linear(in_features=4, out_features=1280, bias=False)
          )
          (drop_path): Identity()
          (norm_1): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (mlp): Mlp(
            (fc1): Linear(in_features=1280, out_features=5120, bias=True)
            (act): GELU(approximate='none')
            (fc2): Linear(in_features=5120, out_features=1280, bias=True)
            (drop): Dropout(p=0.0, inplace=False)
          )
          (norm_2): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (prev_attn): MultiheadAttention(
            (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
          )
        )
        (16): BlockWithMasking(
          (attn): _LoRALayer(
            (w): MultiheadAttention(
              (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
            )
            (w_a): Linear(in_features=1280, out_features=4, bias=False)
            (w_b): Linear(in_features=4, out_features=1280, bias=False)
          )
          (drop_path): Identity()
          (norm_1): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (mlp): Mlp(
            (fc1): Linear(in_features=1280, out_features=5120, bias=True)
            (act): GELU(approximate='none')
            (fc2): Linear(in_features=5120, out_features=1280, bias=True)
            (drop): Dropout(p=0.0, inplace=False)
```

```
          )
          (norm_2): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (prev_attn): MultiheadAttention(
            (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
          )
        )
        (17): BlockWithMasking(
          (attn): _LoRALayer(
            (w): MultiheadAttention(
              (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
            )
            (w_a): Linear(in_features=1280, out_features=4, bias=False)
            (w_b): Linear(in_features=4, out_features=1280, bias=False)
          )
          (drop_path): Identity()
          (norm_1): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (mlp): Mlp(
            (fc1): Linear(in_features=1280, out_features=5120, bias=True)
            (act): GELU(approximate='none')
            (fc2): Linear(in_features=5120, out_features=1280, bias=True)
            (drop): Dropout(p=0.0, inplace=False)
          )
          (norm_2): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (prev_attn): MultiheadAttention(
            (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
          )
        )
        (18): BlockWithMasking(
          (attn): _LoRALayer(
            (w): MultiheadAttention(
              (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
            )
            (w_a): Linear(in_features=1280, out_features=4, bias=False)
            (w_b): Linear(in_features=4, out_features=1280, bias=False)
          )
          (drop_path): Identity()
          (norm_1): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (mlp): Mlp(
            (fc1): Linear(in_features=1280, out_features=5120, bias=True)
            (act): GELU(approximate='none')
            (fc2): Linear(in_features=5120, out_features=1280, bias=True)
            (drop): Dropout(p=0.0, inplace=False)
```

```
          )
          (norm_2): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (prev_attn): MultiheadAttention(
            (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
          )
        )
        (19): BlockWithMasking(
          (attn): _LoRALayer(
            (w): MultiheadAttention(
              (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
            )
            (w_a): Linear(in_features=1280, out_features=4, bias=False)
            (w_b): Linear(in_features=4, out_features=1280, bias=False)
          )
          (drop_path): Identity()
          (norm_1): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (mlp): Mlp(
            (fc1): Linear(in_features=1280, out_features=5120, bias=True)
            (act): GELU(approximate='none')
            (fc2): Linear(in_features=5120, out_features=1280, bias=True)
            (drop): Dropout(p=0.0, inplace=False)
          )
          (norm_2): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (prev_attn): MultiheadAttention(
            (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
          )
        )
        (20): BlockWithMasking(
          (attn): _LoRALayer(
            (w): MultiheadAttention(
              (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
            )
            (w_a): Linear(in_features=1280, out_features=4, bias=False)
            (w_b): Linear(in_features=4, out_features=1280, bias=False)
          )
          (drop_path): Identity()
          (norm_1): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (mlp): Mlp(
            (fc1): Linear(in_features=1280, out_features=5120, bias=True)
            (act): GELU(approximate='none')
            (fc2): Linear(in_features=5120, out_features=1280, bias=True)
            (drop): Dropout(p=0.0, inplace=False)
```

```
          )
          (norm_2): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (prev_attn): MultiheadAttention(
            (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
          )
        )
        (21): BlockWithMasking(
          (attn): _LoRALayer(
            (w): MultiheadAttention(
              (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
            )
            (w_a): Linear(in_features=1280, out_features=4, bias=False)
            (w_b): Linear(in_features=4, out_features=1280, bias=False)
          )
          (drop_path): Identity()
          (norm_1): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (mlp): Mlp(
            (fc1): Linear(in_features=1280, out_features=5120, bias=True)
            (act): GELU(approximate='none')
            (fc2): Linear(in_features=5120, out_features=1280, bias=True)
            (drop): Dropout(p=0.0, inplace=False)
          )
          (norm_2): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (prev_attn): MultiheadAttention(
            (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
          )
        )
        (22): BlockWithMasking(
          (attn): _LoRALayer(
            (w): MultiheadAttention(
              (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
            )
            (w_a): Linear(in_features=1280, out_features=4, bias=False)
            (w_b): Linear(in_features=4, out_features=1280, bias=False)
          )
          (drop_path): Identity()
          (norm_1): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (mlp): Mlp(
            (fc1): Linear(in_features=1280, out_features=5120, bias=True)
            (act): GELU(approximate='none')
            (fc2): Linear(in_features=5120, out_features=1280, bias=True)
            (drop): Dropout(p=0.0, inplace=False)
```

```
          )
          (norm_2): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (prev_attn): MultiheadAttention(
            (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
          )
        )
        (23): BlockWithMasking(
          (attn): _LoRALayer(
            (w): MultiheadAttention(
              (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
            )
            (w_a): Linear(in_features=1280, out_features=4, bias=False)
            (w_b): Linear(in_features=4, out_features=1280, bias=False)
          )
          (drop_path): Identity()
          (norm_1): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (mlp): Mlp(
            (fc1): Linear(in_features=1280, out_features=5120, bias=True)
            (act): GELU(approximate='none')
            (fc2): Linear(in_features=5120, out_features=1280, bias=True)
            (drop): Dropout(p=0.0, inplace=False)
          )
          (norm_2): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (prev_attn): MultiheadAttention(
            (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
          )
        )
        (24): BlockWithMasking(
          (attn): _LoRALayer(
            (w): MultiheadAttention(
              (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
            )
            (w_a): Linear(in_features=1280, out_features=4, bias=False)
            (w_b): Linear(in_features=4, out_features=1280, bias=False)
          )
          (drop_path): Identity()
          (norm_1): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (mlp): Mlp(
            (fc1): Linear(in_features=1280, out_features=5120, bias=True)
            (act): GELU(approximate='none')
            (fc2): Linear(in_features=5120, out_features=1280, bias=True)
            (drop): Dropout(p=0.0, inplace=False)
```

```
          )
          (norm_2): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (prev_attn): MultiheadAttention(
            (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
          )
        )
        (25): BlockWithMasking(
          (attn): _LoRALayer(
            (w): MultiheadAttention(
              (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
            )
            (w_a): Linear(in_features=1280, out_features=4, bias=False)
            (w_b): Linear(in_features=4, out_features=1280, bias=False)
          )
          (drop_path): Identity()
          (norm_1): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (mlp): Mlp(
            (fc1): Linear(in_features=1280, out_features=5120, bias=True)
            (act): GELU(approximate='none')
            (fc2): Linear(in_features=5120, out_features=1280, bias=True)
            (drop): Dropout(p=0.0, inplace=False)
          )
          (norm_2): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (prev_attn): MultiheadAttention(
            (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
          )
        )
        (26): BlockWithMasking(
          (attn): _LoRALayer(
            (w): MultiheadAttention(
              (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
            )
            (w_a): Linear(in_features=1280, out_features=4, bias=False)
            (w_b): Linear(in_features=4, out_features=1280, bias=False)
          )
          (drop_path): Identity()
          (norm_1): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (mlp): Mlp(
            (fc1): Linear(in_features=1280, out_features=5120, bias=True)
            (act): GELU(approximate='none')
            (fc2): Linear(in_features=5120, out_features=1280, bias=True)
            (drop): Dropout(p=0.0, inplace=False)
```

```
          )
          (norm_2): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (prev_attn): MultiheadAttention(
            (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
          )
        )
        (27): BlockWithMasking(
          (attn): _LoRALayer(
            (w): MultiheadAttention(
              (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
            )
            (w_a): Linear(in_features=1280, out_features=4, bias=False)
            (w_b): Linear(in_features=4, out_features=1280, bias=False)
          )
          (drop_path): Identity()
          (norm_1): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (mlp): Mlp(
            (fc1): Linear(in_features=1280, out_features=5120, bias=True)
            (act): GELU(approximate='none')
            (fc2): Linear(in_features=5120, out_features=1280, bias=True)
            (drop): Dropout(p=0.0, inplace=False)
          )
          (norm_2): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (prev_attn): MultiheadAttention(
            (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
          )
        )
        (28): BlockWithMasking(
          (attn): _LoRALayer(
            (w): MultiheadAttention(
              (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
            )
            (w_a): Linear(in_features=1280, out_features=4, bias=False)
            (w_b): Linear(in_features=4, out_features=1280, bias=False)
          )
          (drop_path): Identity()
          (norm_1): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (mlp): Mlp(
            (fc1): Linear(in_features=1280, out_features=5120, bias=True)
            (act): GELU(approximate='none')
            (fc2): Linear(in_features=5120, out_features=1280, bias=True)
            (drop): Dropout(p=0.0, inplace=False)
```

```
          )
          (norm_2): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (prev_attn): MultiheadAttention(
            (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
          )
        )
        (29): BlockWithMasking(
          (attn): _LoRALayer(
            (w): MultiheadAttention(
              (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
            )
            (w_a): Linear(in_features=1280, out_features=4, bias=False)
            (w_b): Linear(in_features=4, out_features=1280, bias=False)
          )
          (drop_path): Identity()
          (norm_1): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (mlp): Mlp(
            (fc1): Linear(in_features=1280, out_features=5120, bias=True)
            (act): GELU(approximate='none')
            (fc2): Linear(in_features=5120, out_features=1280, bias=True)
            (drop): Dropout(p=0.0, inplace=False)
          )
          (norm_2): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (prev_attn): MultiheadAttention(
            (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
          )
        )
        (30): BlockWithMasking(
          (attn): _LoRALayer(
            (w): MultiheadAttention(
              (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
            )
            (w_a): Linear(in_features=1280, out_features=4, bias=False)
            (w_b): Linear(in_features=4, out_features=1280, bias=False)
          )
          (drop_path): Identity()
          (norm_1): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (mlp): Mlp(
            (fc1): Linear(in_features=1280, out_features=5120, bias=True)
            (act): GELU(approximate='none')
            (fc2): Linear(in_features=5120, out_features=1280, bias=True)
            (drop): Dropout(p=0.0, inplace=False)
```

```
          )
          (norm_2): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (prev_attn): MultiheadAttention(
            (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
          )
        )
        (31): BlockWithMasking(
          (attn): _LoRALayer(
            (w): MultiheadAttention(
              (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
            )
            (w_a): Linear(in_features=1280, out_features=4, bias=False)
            (w_b): Linear(in_features=4, out_features=1280, bias=False)
          )
          (drop_path): Identity()
          (norm_1): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (mlp): Mlp(
            (fc1): Linear(in_features=1280, out_features=5120, bias=True)
            (act): GELU(approximate='none')
            (fc2): Linear(in_features=5120, out_features=1280, bias=True)
            (drop): Dropout(p=0.0, inplace=False)
          )
          (norm_2): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
          (prev_attn): MultiheadAttention(
            (out_proj): NonDynamicallyQuantizableLinear(in_features=1280, out_features=1280,
bias=True)
          )
        )
      )
      (post_transformer_layer): EinOpsRearrange()
    )
  )
  (text): SimpleTransformer(
    (pre_transformer_layer): Sequential(
      (0): Identity()
      (1): EinOpsRearrange()
    )
    (blocks): Sequential(
      (0): BlockWithMasking(
        (attn): MultiheadAttention(
          (out_proj): NonDynamicallyQuantizableLinear(in_features=1024, out_features=1024,
bias=True)
        )
        (drop_path): Identity()
```

```
        (norm_1): LayerNorm((1024,), eps=1e-06, elementwise_affine=True)
        (mlp): Mlp(
          (fc1): Linear(in_features=1024, out_features=4096, bias=True)
          (act): GELU(approximate='none')
          (fc2): Linear(in_features=4096, out_features=1024, bias=True)
          (drop): Dropout(p=0.0, inplace=False)
        )
        (norm_2): LayerNorm((1024,), eps=1e-06, elementwise_affine=True)
      )
      (1): BlockWithMasking(
        (attn): MultiheadAttention(
          (out_proj): NonDynamicallyQuantizableLinear(in_features=1024, out_features=1024,
bias=True)
        )
        (drop_path): Identity()
        (norm_1): LayerNorm((1024,), eps=1e-06, elementwise_affine=True)
        (mlp): Mlp(
          (fc1): Linear(in_features=1024, out_features=4096, bias=True)
          (act): GELU(approximate='none')
          (fc2): Linear(in_features=4096, out_features=1024, bias=True)
          (drop): Dropout(p=0.0, inplace=False)
        )
        (norm_2): LayerNorm((1024,), eps=1e-06, elementwise_affine=True)
      )
      (2): BlockWithMasking(
        (attn): MultiheadAttention(
          (out_proj): NonDynamicallyQuantizableLinear(in_features=1024, out_features=1024,
bias=True)
        )
        (drop_path): Identity()
        (norm_1): LayerNorm((1024,), eps=1e-06, elementwise_affine=True)
        (mlp): Mlp(
          (fc1): Linear(in_features=1024, out_features=4096, bias=True)
          (act): GELU(approximate='none')
          (fc2): Linear(in_features=4096, out_features=1024, bias=True)
          (drop): Dropout(p=0.0, inplace=False)
        )
        (norm_2): LayerNorm((1024,), eps=1e-06, elementwise_affine=True)
      )
      (3): BlockWithMasking(
        (attn): MultiheadAttention(
          (out_proj): NonDynamicallyQuantizableLinear(in_features=1024, out_features=1024,
bias=True)
        )
        (drop_path): Identity()
        (norm_1): LayerNorm((1024,), eps=1e-06, elementwise_affine=True)
```

```
        (mlp): Mlp(
          (fc1): Linear(in_features=1024, out_features=4096, bias=True)
          (act): GELU(approximate='none')
          (fc2): Linear(in_features=4096, out_features=1024, bias=True)
          (drop): Dropout(p=0.0, inplace=False)
        )
        (norm_2): LayerNorm((1024,), eps=1e-06, elementwise_affine=True)
      )
      (4): BlockWithMasking(
        (attn): MultiheadAttention(
          (out_proj): NonDynamicallyQuantizableLinear(in_features=1024, out_features=1024,
bias=True)
        )
        (drop_path): Identity()
        (norm_1): LayerNorm((1024,), eps=1e-06, elementwise_affine=True)
        (mlp): Mlp(
          (fc1): Linear(in_features=1024, out_features=4096, bias=True)
          (act): GELU(approximate='none')
          (fc2): Linear(in_features=4096, out_features=1024, bias=True)
          (drop): Dropout(p=0.0, inplace=False)
        )
        (norm_2): LayerNorm((1024,), eps=1e-06, elementwise_affine=True)
      )
      (5): BlockWithMasking(
        (attn): MultiheadAttention(
          (out_proj): NonDynamicallyQuantizableLinear(in_features=1024, out_features=1024,
bias=True)
        )
        (drop_path): Identity()
        (norm_1): LayerNorm((1024,), eps=1e-06, elementwise_affine=True)
        (mlp): Mlp(
          (fc1): Linear(in_features=1024, out_features=4096, bias=True)
          (act): GELU(approximate='none')
          (fc2): Linear(in_features=4096, out_features=1024, bias=True)
          (drop): Dropout(p=0.0, inplace=False)
        )
        (norm_2): LayerNorm((1024,), eps=1e-06, elementwise_affine=True)
      )
      (6): BlockWithMasking(
        (attn): MultiheadAttention(
          (out_proj): NonDynamicallyQuantizableLinear(in_features=1024, out_features=1024,
bias=True)
        )
        (drop_path): Identity()
        (norm_1): LayerNorm((1024,), eps=1e-06, elementwise_affine=True)
        (mlp): Mlp(
```

```
          (fc1): Linear(in_features=1024, out_features=4096, bias=True)
          (act): GELU(approximate='none')
          (fc2): Linear(in_features=4096, out_features=1024, bias=True)
          (drop): Dropout(p=0.0, inplace=False)
        )
        (norm_2): LayerNorm((1024,), eps=1e-06, elementwise_affine=True)
      )
      (7): BlockWithMasking(
        (attn): MultiheadAttention(
          (out_proj): NonDynamicallyQuantizableLinear(in_features=1024, out_features=1024,
bias=True)
        )
        (drop_path): Identity()
        (norm_1): LayerNorm((1024,), eps=1e-06, elementwise_affine=True)
        (mlp): Mlp(
          (fc1): Linear(in_features=1024, out_features=4096, bias=True)
          (act): GELU(approximate='none')
          (fc2): Linear(in_features=4096, out_features=1024, bias=True)
          (drop): Dropout(p=0.0, inplace=False)
        )
        (norm_2): LayerNorm((1024,), eps=1e-06, elementwise_affine=True)
      )
      (8): BlockWithMasking(
        (attn): MultiheadAttention(
          (out_proj): NonDynamicallyQuantizableLinear(in_features=1024, out_features=1024,
bias=True)
        )
        (drop_path): Identity()
        (norm_1): LayerNorm((1024,), eps=1e-06, elementwise_affine=True)
        (mlp): Mlp(
          (fc1): Linear(in_features=1024, out_features=4096, bias=True)
          (act): GELU(approximate='none')
          (fc2): Linear(in_features=4096, out_features=1024, bias=True)
          (drop): Dropout(p=0.0, inplace=False)
        )
        (norm_2): LayerNorm((1024,), eps=1e-06, elementwise_affine=True)
      )
      (9): BlockWithMasking(
        (attn): MultiheadAttention(
          (out_proj): NonDynamicallyQuantizableLinear(in_features=1024, out_features=1024,
bias=True)
        )
        (drop_path): Identity()
        (norm_1): LayerNorm((1024,), eps=1e-06, elementwise_affine=True)
        (mlp): Mlp(
          (fc1): Linear(in_features=1024, out_features=4096, bias=True)
```

```
          (act): GELU(approximate='none')
          (fc2): Linear(in_features=4096, out_features=1024, bias=True)
          (drop): Dropout(p=0.0, inplace=False)
        )
        (norm_2): LayerNorm((1024,), eps=1e-06, elementwise_affine=True)
      )
      (10): BlockWithMasking(
        (attn): MultiheadAttention(
          (out_proj): NonDynamicallyQuantizableLinear(in_features=1024, out_features=1024,
bias=True)
        )
        (drop_path): Identity()
        (norm_1): LayerNorm((1024,), eps=1e-06, elementwise_affine=True)
        (mlp): Mlp(
          (fc1): Linear(in_features=1024, out_features=4096, bias=True)
          (act): GELU(approximate='none')
          (fc2): Linear(in_features=4096, out_features=1024, bias=True)
          (drop): Dropout(p=0.0, inplace=False)
        )
        (norm_2): LayerNorm((1024,), eps=1e-06, elementwise_affine=True)
      )
      (11): BlockWithMasking(
        (attn): MultiheadAttention(
          (out_proj): NonDynamicallyQuantizableLinear(in_features=1024, out_features=1024,
bias=True)
        )
        (drop_path): Identity()
        (norm_1): LayerNorm((1024,), eps=1e-06, elementwise_affine=True)
        (mlp): Mlp(
          (fc1): Linear(in_features=1024, out_features=4096, bias=True)
          (act): GELU(approximate='none')
          (fc2): Linear(in_features=4096, out_features=1024, bias=True)
          (drop): Dropout(p=0.0, inplace=False)
        )
        (norm_2): LayerNorm((1024,), eps=1e-06, elementwise_affine=True)
      )
      (12): BlockWithMasking(
        (attn): MultiheadAttention(
          (out_proj): NonDynamicallyQuantizableLinear(in_features=1024, out_features=1024,
bias=True)
        )
        (drop_path): Identity()
        (norm_1): LayerNorm((1024,), eps=1e-06, elementwise_affine=True)
        (mlp): Mlp(
          (fc1): Linear(in_features=1024, out_features=4096, bias=True)
          (act): GELU(approximate='none')
```

```
          (fc2): Linear(in_features=4096, out_features=1024, bias=True)
          (drop): Dropout(p=0.0, inplace=False)
        )
        (norm_2): LayerNorm((1024,), eps=1e-06, elementwise_affine=True)
      )
      (13): BlockWithMasking(
        (attn): MultiheadAttention(
          (out_proj): NonDynamicallyQuantizableLinear(in_features=1024, out_features=1024,
bias=True)
        )
        (drop_path): Identity()
        (norm_1): LayerNorm((1024,), eps=1e-06, elementwise_affine=True)
        (mlp): Mlp(
          (fc1): Linear(in_features=1024, out_features=4096, bias=True)
          (act): GELU(approximate='none')
          (fc2): Linear(in_features=4096, out_features=1024, bias=True)
          (drop): Dropout(p=0.0, inplace=False)
        )
        (norm_2): LayerNorm((1024,), eps=1e-06, elementwise_affine=True)
      )
      (14): BlockWithMasking(
        (attn): MultiheadAttention(
          (out_proj): NonDynamicallyQuantizableLinear(in_features=1024, out_features=1024,
bias=True)
        )
        (drop_path): Identity()
        (norm_1): LayerNorm((1024,), eps=1e-06, elementwise_affine=True)
        (mlp): Mlp(
          (fc1): Linear(in_features=1024, out_features=4096, bias=True)
          (act): GELU(approximate='none')
          (fc2): Linear(in_features=4096, out_features=1024, bias=True)
          (drop): Dropout(p=0.0, inplace=False)
        )
        (norm_2): LayerNorm((1024,), eps=1e-06, elementwise_affine=True)
      )
      (15): BlockWithMasking(
        (attn): MultiheadAttention(
          (out_proj): NonDynamicallyQuantizableLinear(in_features=1024, out_features=1024,
bias=True)
        )
        (drop_path): Identity()
        (norm_1): LayerNorm((1024,), eps=1e-06, elementwise_affine=True)
        (mlp): Mlp(
          (fc1): Linear(in_features=1024, out_features=4096, bias=True)
          (act): GELU(approximate='none')
          (fc2): Linear(in_features=4096, out_features=1024, bias=True)
```

```
          (drop): Dropout(p=0.0, inplace=False)
        )
        (norm_2): LayerNorm((1024,), eps=1e-06, elementwise_affine=True)
      )
      (16): BlockWithMasking(
        (attn): MultiheadAttention(
          (out_proj): NonDynamicallyQuantizableLinear(in_features=1024, out_features=1024,
bias=True)
        )
        (drop_path): Identity()
        (norm_1): LayerNorm((1024,), eps=1e-06, elementwise_affine=True)
        (mlp): Mlp(
          (fc1): Linear(in_features=1024, out_features=4096, bias=True)
          (act): GELU(approximate='none')
          (fc2): Linear(in_features=4096, out_features=1024, bias=True)
          (drop): Dropout(p=0.0, inplace=False)
        )
        (norm_2): LayerNorm((1024,), eps=1e-06, elementwise_affine=True)
      )
      (17): BlockWithMasking(
        (attn): MultiheadAttention(
          (out_proj): NonDynamicallyQuantizableLinear(in_features=1024, out_features=1024,
bias=True)
        )
        (drop_path): Identity()
        (norm_1): LayerNorm((1024,), eps=1e-06, elementwise_affine=True)
        (mlp): Mlp(
          (fc1): Linear(in_features=1024, out_features=4096, bias=True)
          (act): GELU(approximate='none')
          (fc2): Linear(in_features=4096, out_features=1024, bias=True)
          (drop): Dropout(p=0.0, inplace=False)
        )
        (norm_2): LayerNorm((1024,), eps=1e-06, elementwise_affine=True)
      )
      (18): BlockWithMasking(
        (attn): MultiheadAttention(
          (out_proj): NonDynamicallyQuantizableLinear(in_features=1024, out_features=1024,
bias=True)
        )
        (drop_path): Identity()
        (norm_1): LayerNorm((1024,), eps=1e-06, elementwise_affine=True)
        (mlp): Mlp(
          (fc1): Linear(in_features=1024, out_features=4096, bias=True)
          (act): GELU(approximate='none')
          (fc2): Linear(in_features=4096, out_features=1024, bias=True)
          (drop): Dropout(p=0.0, inplace=False)
```

```
        )
        (norm_2): LayerNorm((1024,), eps=1e-06, elementwise_affine=True)
      )
      (19): BlockWithMasking(
        (attn): MultiheadAttention(
          (out_proj): NonDynamicallyQuantizableLinear(in_features=1024, out_features=1024,
bias=True)
        )
        (drop_path): Identity()
        (norm_1): LayerNorm((1024,), eps=1e-06, elementwise_affine=True)
        (mlp): Mlp(
          (fc1): Linear(in_features=1024, out_features=4096, bias=True)
          (act): GELU(approximate='none')
          (fc2): Linear(in_features=4096, out_features=1024, bias=True)
          (drop): Dropout(p=0.0, inplace=False)
        )
        (norm_2): LayerNorm((1024,), eps=1e-06, elementwise_affine=True)
      )
      (20): BlockWithMasking(
        (attn): MultiheadAttention(
          (out_proj): NonDynamicallyQuantizableLinear(in_features=1024, out_features=1024,
bias=True)
        )
        (drop_path): Identity()
        (norm_1): LayerNorm((1024,), eps=1e-06, elementwise_affine=True)
        (mlp): Mlp(
          (fc1): Linear(in_features=1024, out_features=4096, bias=True)
          (act): GELU(approximate='none')
          (fc2): Linear(in_features=4096, out_features=1024, bias=True)
          (drop): Dropout(p=0.0, inplace=False)
        )
        (norm_2): LayerNorm((1024,), eps=1e-06, elementwise_affine=True)
      )
      (21): BlockWithMasking(
        (attn): MultiheadAttention(
          (out_proj): NonDynamicallyQuantizableLinear(in_features=1024, out_features=1024,
bias=True)
        )
        (drop_path): Identity()
        (norm_1): LayerNorm((1024,), eps=1e-06, elementwise_affine=True)
        (mlp): Mlp(
          (fc1): Linear(in_features=1024, out_features=4096, bias=True)
          (act): GELU(approximate='none')
          (fc2): Linear(in_features=4096, out_features=1024, bias=True)
          (drop): Dropout(p=0.0, inplace=False)
        )
```

```
        (norm_2): LayerNorm((1024,), eps=1e-06, elementwise_affine=True)
      )
      (22): BlockWithMasking(
        (attn): MultiheadAttention(
          (out_proj): NonDynamicallyQuantizableLinear(in_features=1024, out_features=1024,
bias=True)
        )
        (drop_path): Identity()
        (norm_1): LayerNorm((1024,), eps=1e-06, elementwise_affine=True)
        (mlp): Mlp(
          (fc1): Linear(in_features=1024, out_features=4096, bias=True)
          (act): GELU(approximate='none')
          (fc2): Linear(in_features=4096, out_features=1024, bias=True)
          (drop): Dropout(p=0.0, inplace=False)
        )
        (norm_2): LayerNorm((1024,), eps=1e-06, elementwise_affine=True)
      )
      (23): BlockWithMasking(
        (attn): MultiheadAttention(
          (out_proj): NonDynamicallyQuantizableLinear(in_features=1024, out_features=1024,
bias=True)
        )
        (drop_path): Identity()
        (norm_1): LayerNorm((1024,), eps=1e-06, elementwise_affine=True)
        (mlp): Mlp(
          (fc1): Linear(in_features=1024, out_features=4096, bias=True)
          (act): GELU(approximate='none')
          (fc2): Linear(in_features=4096, out_features=1024, bias=True)
          (drop): Dropout(p=0.0, inplace=False)
        )
        (norm_2): LayerNorm((1024,), eps=1e-06, elementwise_affine=True)
      )
    )
    (post_transformer_layer): EinOpsRearrange()
  )
  (audio): LoRA_SimpleTransformer(
    (lora_model): SimpleTransformer(
      (pre_transformer_layer): Sequential(
        (0): Identity()
        (1): EinOpsRearrange()
      )
      (blocks): Sequential(
        (0): BlockWithMasking(
          (attn): _LoRALayer(
            (w): MultiheadAttention(
```

```
            (out_proj): NonDynamicallyQuantizableLinear(in_features=768, out_features=768,
bias=True)
            )
            (w_a): Linear(in_features=768, out_features=4, bias=False)
            (w_b): Linear(in_features=4, out_features=768, bias=False)
          )
          (drop_path): Identity()
          (norm_1): LayerNorm((768,), eps=1e-06, elementwise_affine=True)
          (mlp): Mlp(
            (fc1): Linear(in_features=768, out_features=3072, bias=True)
            (act): GELU(approximate='none')
            (fc2): Linear(in_features=3072, out_features=768, bias=True)
            (drop): Dropout(p=0.0, inplace=False)
          )
          (norm_2): LayerNorm((768,), eps=1e-06, elementwise_affine=True)
          (prev_attn): MultiheadAttention(
            (out_proj): NonDynamicallyQuantizableLinear(in_features=768, out_features=768,
bias=True)
          )
        )
        (1): BlockWithMasking(
          (attn): _LoRALayer(
            (w): MultiheadAttention(
              (out_proj): NonDynamicallyQuantizableLinear(in_features=768, out_features=768,
bias=True)
            )
            (w_a): Linear(in_features=768, out_features=4, bias=False)
            (w_b): Linear(in_features=4, out_features=768, bias=False)
          )
          (drop_path): DropPath(drop_prob=0.009)
          (norm_1): LayerNorm((768,), eps=1e-06, elementwise_affine=True)
          (mlp): Mlp(
            (fc1): Linear(in_features=768, out_features=3072, bias=True)
            (act): GELU(approximate='none')
            (fc2): Linear(in_features=3072, out_features=768, bias=True)
            (drop): Dropout(p=0.0, inplace=False)
          )
          (norm_2): LayerNorm((768,), eps=1e-06, elementwise_affine=True)
          (prev_attn): MultiheadAttention(
            (out_proj): NonDynamicallyQuantizableLinear(in_features=768, out_features=768,
bias=True)
          )
        )
        (2): BlockWithMasking(
          (attn): _LoRALayer(
            (w): MultiheadAttention(
```

```
            (out_proj): NonDynamicallyQuantizableLinear(in_features=768, out_features=768,
bias=True)
          )
          (w_a): Linear(in_features=768, out_features=4, bias=False)
          (w_b): Linear(in_features=4, out_features=768, bias=False)
        )
        (drop_path): DropPath(drop_prob=0.018)
        (norm_1): LayerNorm((768,), eps=1e-06, elementwise_affine=True)
        (mlp): Mlp(
          (fc1): Linear(in_features=768, out_features=3072, bias=True)
          (act): GELU(approximate='none')
          (fc2): Linear(in_features=3072, out_features=768, bias=True)
          (drop): Dropout(p=0.0, inplace=False)
        )
        (norm_2): LayerNorm((768,), eps=1e-06, elementwise_affine=True)
        (prev_attn): MultiheadAttention(
          (out_proj): NonDynamicallyQuantizableLinear(in_features=768, out_features=768,
bias=True)
        )
      )
      (3): BlockWithMasking(
        (attn): _LoRALayer(
          (w): MultiheadAttention(
            (out_proj): NonDynamicallyQuantizableLinear(in_features=768, out_features=768,
bias=True)
          )
          (w_a): Linear(in_features=768, out_features=4, bias=False)
          (w_b): Linear(in_features=4, out_features=768, bias=False)
        )
        (drop_path): DropPath(drop_prob=0.027)
        (norm_1): LayerNorm((768,), eps=1e-06, elementwise_affine=True)
        (mlp): Mlp(
          (fc1): Linear(in_features=768, out_features=3072, bias=True)
          (act): GELU(approximate='none')
          (fc2): Linear(in_features=3072, out_features=768, bias=True)
          (drop): Dropout(p=0.0, inplace=False)
        )
        (norm_2): LayerNorm((768,), eps=1e-06, elementwise_affine=True)
        (prev_attn): MultiheadAttention(
          (out_proj): NonDynamicallyQuantizableLinear(in_features=768, out_features=768,
bias=True)
        )
      )
      (4): BlockWithMasking(
        (attn): _LoRALayer(
          (w): MultiheadAttention(
```

```
            (out_proj): NonDynamicallyQuantizableLinear(in_features=768, out_features=768,
bias=True)
          )
          (w_a): Linear(in_features=768, out_features=4, bias=False)
          (w_b): Linear(in_features=4, out_features=768, bias=False)
        )
        (drop_path): DropPath(drop_prob=0.036)
        (norm_1): LayerNorm((768,), eps=1e-06, elementwise_affine=True)
        (mlp): Mlp(
          (fc1): Linear(in_features=768, out_features=3072, bias=True)
          (act): GELU(approximate='none')
          (fc2): Linear(in_features=3072, out_features=768, bias=True)
          (drop): Dropout(p=0.0, inplace=False)
        )
        (norm_2): LayerNorm((768,), eps=1e-06, elementwise_affine=True)
        (prev_attn): MultiheadAttention(
          (out_proj): NonDynamicallyQuantizableLinear(in_features=768, out_features=768,
bias=True)
        )
      )
      (5): BlockWithMasking(
        (attn): _LoRALayer(
          (w): MultiheadAttention(
            (out_proj): NonDynamicallyQuantizableLinear(in_features=768, out_features=768,
bias=True)
          )
          (w_a): Linear(in_features=768, out_features=4, bias=False)
          (w_b): Linear(in_features=4, out_features=768, bias=False)
        )
        (drop_path): DropPath(drop_prob=0.045)
        (norm_1): LayerNorm((768,), eps=1e-06, elementwise_affine=True)
        (mlp): Mlp(
          (fc1): Linear(in_features=768, out_features=3072, bias=True)
          (act): GELU(approximate='none')
          (fc2): Linear(in_features=3072, out_features=768, bias=True)
          (drop): Dropout(p=0.0, inplace=False)
        )
        (norm_2): LayerNorm((768,), eps=1e-06, elementwise_affine=True)
        (prev_attn): MultiheadAttention(
          (out_proj): NonDynamicallyQuantizableLinear(in_features=768, out_features=768,
bias=True)
        )
      )
      (6): BlockWithMasking(
        (attn): _LoRALayer(
          (w): MultiheadAttention(
```

```
            (out_proj): NonDynamicallyQuantizableLinear(in_features=768, out_features=768,
bias=True)
          )
          (w_a): Linear(in_features=768, out_features=4, bias=False)
          (w_b): Linear(in_features=4, out_features=768, bias=False)
        )
        (drop_path): DropPath(drop_prob=0.055)
        (norm_1): LayerNorm((768,), eps=1e-06, elementwise_affine=True)
        (mlp): Mlp(
          (fc1): Linear(in_features=768, out_features=3072, bias=True)
          (act): GELU(approximate='none')
          (fc2): Linear(in_features=3072, out_features=768, bias=True)
          (drop): Dropout(p=0.0, inplace=False)
        )
        (norm_2): LayerNorm((768,), eps=1e-06, elementwise_affine=True)
        (prev_attn): MultiheadAttention(
          (out_proj): NonDynamicallyQuantizableLinear(in_features=768, out_features=768,
bias=True)
        )
      )
      (7): BlockWithMasking(
        (attn): _LoRALayer(
          (w): MultiheadAttention(
            (out_proj): NonDynamicallyQuantizableLinear(in_features=768, out_features=768,
bias=True)
          )
          (w_a): Linear(in_features=768, out_features=4, bias=False)
          (w_b): Linear(in_features=4, out_features=768, bias=False)
        )
        (drop_path): DropPath(drop_prob=0.064)
        (norm_1): LayerNorm((768,), eps=1e-06, elementwise_affine=True)
        (mlp): Mlp(
          (fc1): Linear(in_features=768, out_features=3072, bias=True)
          (act): GELU(approximate='none')
          (fc2): Linear(in_features=3072, out_features=768, bias=True)
          (drop): Dropout(p=0.0, inplace=False)
        )
        (norm_2): LayerNorm((768,), eps=1e-06, elementwise_affine=True)
        (prev_attn): MultiheadAttention(
          (out_proj): NonDynamicallyQuantizableLinear(in_features=768, out_features=768,
bias=True)
        )
      )
      (8): BlockWithMasking(
        (attn): _LoRALayer(
          (w): MultiheadAttention(
```

```
            (out_proj): NonDynamicallyQuantizableLinear(in_features=768, out_features=768,
bias=True)
            )
            (w_a): Linear(in_features=768, out_features=4, bias=False)
            (w_b): Linear(in_features=4, out_features=768, bias=False)
          )
          (drop_path): DropPath(drop_prob=0.073)
          (norm_1): LayerNorm((768,), eps=1e-06, elementwise_affine=True)
          (mlp): Mlp(
            (fc1): Linear(in_features=768, out_features=3072, bias=True)
            (act): GELU(approximate='none')
            (fc2): Linear(in_features=3072, out_features=768, bias=True)
            (drop): Dropout(p=0.0, inplace=False)
          )
          (norm_2): LayerNorm((768,), eps=1e-06, elementwise_affine=True)
          (prev_attn): MultiheadAttention(
            (out_proj): NonDynamicallyQuantizableLinear(in_features=768, out_features=768,
bias=True)
          )
        )
        (9): BlockWithMasking(
          (attn): _LoRALayer(
            (w): MultiheadAttention(
              (out_proj): NonDynamicallyQuantizableLinear(in_features=768, out_features=768,
bias=True)
            )
            (w_a): Linear(in_features=768, out_features=4, bias=False)
            (w_b): Linear(in_features=4, out_features=768, bias=False)
          )
          (drop_path): DropPath(drop_prob=0.082)
          (norm_1): LayerNorm((768,), eps=1e-06, elementwise_affine=True)
          (mlp): Mlp(
            (fc1): Linear(in_features=768, out_features=3072, bias=True)
            (act): GELU(approximate='none')
            (fc2): Linear(in_features=3072, out_features=768, bias=True)
            (drop): Dropout(p=0.0, inplace=False)
          )
          (norm_2): LayerNorm((768,), eps=1e-06, elementwise_affine=True)
          (prev_attn): MultiheadAttention(
            (out_proj): NonDynamicallyQuantizableLinear(in_features=768, out_features=768,
bias=True)
          )
        )
        (10): BlockWithMasking(
          (attn): _LoRALayer(
            (w): MultiheadAttention(
```

```
            (out_proj): NonDynamicallyQuantizableLinear(in_features=768, out_features=768,
bias=True)
            )
            (w_a): Linear(in_features=768, out_features=4, bias=False)
            (w_b): Linear(in_features=4, out_features=768, bias=False)
          )
          (drop_path): DropPath(drop_prob=0.091)
          (norm_1): LayerNorm((768,), eps=1e-06, elementwise_affine=True)
          (mlp): Mlp(
            (fc1): Linear(in_features=768, out_features=3072, bias=True)
            (act): GELU(approximate='none')
            (fc2): Linear(in_features=3072, out_features=768, bias=True)
            (drop): Dropout(p=0.0, inplace=False)
          )
          (norm_2): LayerNorm((768,), eps=1e-06, elementwise_affine=True)
          (prev_attn): MultiheadAttention(
            (out_proj): NonDynamicallyQuantizableLinear(in_features=768, out_features=768,
bias=True)
          )
        )
        (11): BlockWithMasking(
          (attn): _LoRALayer(
            (w): MultiheadAttention(
              (out_proj): NonDynamicallyQuantizableLinear(in_features=768, out_features=768,
bias=True)
            )
            (w_a): Linear(in_features=768, out_features=4, bias=False)
            (w_b): Linear(in_features=4, out_features=768, bias=False)
          )
          (drop_path): DropPath(drop_prob=0.100)
          (norm_1): LayerNorm((768,), eps=1e-06, elementwise_affine=True)
          (mlp): Mlp(
            (fc1): Linear(in_features=768, out_features=3072, bias=True)
            (act): GELU(approximate='none')
            (fc2): Linear(in_features=3072, out_features=768, bias=True)
            (drop): Dropout(p=0.0, inplace=False)
          )
          (norm_2): LayerNorm((768,), eps=1e-06, elementwise_affine=True)
          (prev_attn): MultiheadAttention(
            (out_proj): NonDynamicallyQuantizableLinear(in_features=768, out_features=768,
bias=True)
          )
        )
      )
      (post_transformer_layer): EinOpsRearrange()
    )
```

```
    )
    (depth): SimpleTransformer(
      (pre_transformer_layer): Sequential(
        (0): Identity()
        (1): EinOpsRearrange()
      )
      (blocks): Sequential(
        (0): BlockWithMasking(
          (attn): MultiheadAttention(
            (out_proj): NonDynamicallyQuantizableLinear(in_features=384, out_features=384,
bias=True)
          )
          (drop_path): Identity()
          (norm_1): LayerNorm((384,), eps=1e-06, elementwise_affine=True)
          (mlp): Mlp(
            (fc1): Linear(in_features=384, out_features=1536, bias=True)
            (act): GELU(approximate='none')
            (fc2): Linear(in_features=1536, out_features=384, bias=True)
            (drop): Dropout(p=0.0, inplace=False)
          )
          (norm_2): LayerNorm((384,), eps=1e-06, elementwise_affine=True)
        )
        (1): BlockWithMasking(
          (attn): MultiheadAttention(
            (out_proj): NonDynamicallyQuantizableLinear(in_features=384, out_features=384,
bias=True)
          )
          (drop_path): Identity()
          (norm_1): LayerNorm((384,), eps=1e-06, elementwise_affine=True)
          (mlp): Mlp(
            (fc1): Linear(in_features=384, out_features=1536, bias=True)
            (act): GELU(approximate='none')
            (fc2): Linear(in_features=1536, out_features=384, bias=True)
            (drop): Dropout(p=0.0, inplace=False)
          )
          (norm_2): LayerNorm((384,), eps=1e-06, elementwise_affine=True)
        )
        (2): BlockWithMasking(
          (attn): MultiheadAttention(
            (out_proj): NonDynamicallyQuantizableLinear(in_features=384, out_features=384,
bias=True)
          )
          (drop_path): Identity()
          (norm_1): LayerNorm((384,), eps=1e-06, elementwise_affine=True)
          (mlp): Mlp(
            (fc1): Linear(in_features=384, out_features=1536, bias=True)
```

```
          (act): GELU(approximate='none')
          (fc2): Linear(in_features=1536, out_features=384, bias=True)
          (drop): Dropout(p=0.0, inplace=False)
        )
        (norm_2): LayerNorm((384,), eps=1e-06, elementwise_affine=True)
      )
      (3): BlockWithMasking(
        (attn): MultiheadAttention(
          (out_proj): NonDynamicallyQuantizableLinear(in_features=384, out_features=384,
bias=True)
        )
        (drop_path): Identity()
        (norm_1): LayerNorm((384,), eps=1e-06, elementwise_affine=True)
        (mlp): Mlp(
          (fc1): Linear(in_features=384, out_features=1536, bias=True)
          (act): GELU(approximate='none')
          (fc2): Linear(in_features=1536, out_features=384, bias=True)
          (drop): Dropout(p=0.0, inplace=False)
        )
        (norm_2): LayerNorm((384,), eps=1e-06, elementwise_affine=True)
      )
      (4): BlockWithMasking(
        (attn): MultiheadAttention(
          (out_proj): NonDynamicallyQuantizableLinear(in_features=384, out_features=384,
bias=True)
        )
        (drop_path): Identity()
        (norm_1): LayerNorm((384,), eps=1e-06, elementwise_affine=True)
        (mlp): Mlp(
          (fc1): Linear(in_features=384, out_features=1536, bias=True)
          (act): GELU(approximate='none')
          (fc2): Linear(in_features=1536, out_features=384, bias=True)
          (drop): Dropout(p=0.0, inplace=False)
        )
        (norm_2): LayerNorm((384,), eps=1e-06, elementwise_affine=True)
      )
      (5): BlockWithMasking(
        (attn): MultiheadAttention(
          (out_proj): NonDynamicallyQuantizableLinear(in_features=384, out_features=384,
bias=True)
        )
        (drop_path): Identity()
        (norm_1): LayerNorm((384,), eps=1e-06, elementwise_affine=True)
        (mlp): Mlp(
          (fc1): Linear(in_features=384, out_features=1536, bias=True)
          (act): GELU(approximate='none')
```

```
          (fc2): Linear(in_features=1536, out_features=384, bias=True)
          (drop): Dropout(p=0.0, inplace=False)
        )
        (norm_2): LayerNorm((384,), eps=1e-06, elementwise_affine=True)
      )
      (6): BlockWithMasking(
        (attn): MultiheadAttention(
          (out_proj): NonDynamicallyQuantizableLinear(in_features=384, out_features=384,
bias=True)
        )
        (drop_path): Identity()
        (norm_1): LayerNorm((384,), eps=1e-06, elementwise_affine=True)
        (mlp): Mlp(
          (fc1): Linear(in_features=384, out_features=1536, bias=True)
          (act): GELU(approximate='none')
          (fc2): Linear(in_features=1536, out_features=384, bias=True)
          (drop): Dropout(p=0.0, inplace=False)
        )
        (norm_2): LayerNorm((384,), eps=1e-06, elementwise_affine=True)
      )
      (7): BlockWithMasking(
        (attn): MultiheadAttention(
          (out_proj): NonDynamicallyQuantizableLinear(in_features=384, out_features=384,
bias=True)
        )
        (drop_path): Identity()
        (norm_1): LayerNorm((384,), eps=1e-06, elementwise_affine=True)
        (mlp): Mlp(
          (fc1): Linear(in_features=384, out_features=1536, bias=True)
          (act): GELU(approximate='none')
          (fc2): Linear(in_features=1536, out_features=384, bias=True)
          (drop): Dropout(p=0.0, inplace=False)
        )
        (norm_2): LayerNorm((384,), eps=1e-06, elementwise_affine=True)
      )
      (8): BlockWithMasking(
        (attn): MultiheadAttention(
          (out_proj): NonDynamicallyQuantizableLinear(in_features=384, out_features=384,
bias=True)
        )
        (drop_path): Identity()
        (norm_1): LayerNorm((384,), eps=1e-06, elementwise_affine=True)
        (mlp): Mlp(
          (fc1): Linear(in_features=384, out_features=1536, bias=True)
          (act): GELU(approximate='none')
          (fc2): Linear(in_features=1536, out_features=384, bias=True)
```

```
          (drop): Dropout(p=0.0, inplace=False)
        )
        (norm_2): LayerNorm((384,), eps=1e-06, elementwise_affine=True)
      )
      (9): BlockWithMasking(
        (attn): MultiheadAttention(
          (out_proj): NonDynamicallyQuantizableLinear(in_features=384, out_features=384,
bias=True)
        )
        (drop_path): Identity()
        (norm_1): LayerNorm((384,), eps=1e-06, elementwise_affine=True)
        (mlp): Mlp(
          (fc1): Linear(in_features=384, out_features=1536, bias=True)
          (act): GELU(approximate='none')
          (fc2): Linear(in_features=1536, out_features=384, bias=True)
          (drop): Dropout(p=0.0, inplace=False)
        )
        (norm_2): LayerNorm((384,), eps=1e-06, elementwise_affine=True)
      )
      (10): BlockWithMasking(
        (attn): MultiheadAttention(
          (out_proj): NonDynamicallyQuantizableLinear(in_features=384, out_features=384,
bias=True)
        )
        (drop_path): Identity()
        (norm_1): LayerNorm((384,), eps=1e-06, elementwise_affine=True)
        (mlp): Mlp(
          (fc1): Linear(in_features=384, out_features=1536, bias=True)
          (act): GELU(approximate='none')
          (fc2): Linear(in_features=1536, out_features=384, bias=True)
          (drop): Dropout(p=0.0, inplace=False)
        )
        (norm_2): LayerNorm((384,), eps=1e-06, elementwise_affine=True)
      )
      (11): BlockWithMasking(
        (attn): MultiheadAttention(
          (out_proj): NonDynamicallyQuantizableLinear(in_features=384, out_features=384,
bias=True)
        )
        (drop_path): Identity()
        (norm_1): LayerNorm((384,), eps=1e-06, elementwise_affine=True)
        (mlp): Mlp(
          (fc1): Linear(in_features=384, out_features=1536, bias=True)
          (act): GELU(approximate='none')
          (fc2): Linear(in_features=1536, out_features=384, bias=True)
          (drop): Dropout(p=0.0, inplace=False)
```

```
        )
        (norm_2): LayerNorm((384,), eps=1e-06, elementwise_affine=True)
      )
    )
    (post_transformer_layer): EinOpsRearrange()
  )
  (thermal): SimpleTransformer(
    (pre_transformer_layer): Sequential(
      (0): Identity()
      (1): EinOpsRearrange()
    )
    (blocks): Sequential(
      (0): BlockWithMasking(
        (attn): MultiheadAttention(
          (out_proj): NonDynamicallyQuantizableLinear(in_features=768, out_features=768,
bias=True)
        )
        (drop_path): Identity()
        (norm_1): LayerNorm((768,), eps=1e-06, elementwise_affine=True)
        (mlp): Mlp(
          (fc1): Linear(in_features=768, out_features=3072, bias=True)
          (act): GELU(approximate='none')
          (fc2): Linear(in_features=3072, out_features=768, bias=True)
          (drop): Dropout(p=0.0, inplace=False)
        )
        (norm_2): LayerNorm((768,), eps=1e-06, elementwise_affine=True)
      )
      (1): BlockWithMasking(
        (attn): MultiheadAttention(
          (out_proj): NonDynamicallyQuantizableLinear(in_features=768, out_features=768,
bias=True)
        )
        (drop_path): Identity()
        (norm_1): LayerNorm((768,), eps=1e-06, elementwise_affine=True)
        (mlp): Mlp(
          (fc1): Linear(in_features=768, out_features=3072, bias=True)
          (act): GELU(approximate='none')
          (fc2): Linear(in_features=3072, out_features=768, bias=True)
          (drop): Dropout(p=0.0, inplace=False)
        )
        (norm_2): LayerNorm((768,), eps=1e-06, elementwise_affine=True)
      )
      (2): BlockWithMasking(
        (attn): MultiheadAttention(
          (out_proj): NonDynamicallyQuantizableLinear(in_features=768, out_features=768,
bias=True)
```

```
        )
        (drop_path): Identity()
        (norm_1): LayerNorm((768,), eps=1e-06, elementwise_affine=True)
        (mlp): Mlp(
          (fc1): Linear(in_features=768, out_features=3072, bias=True)
          (act): GELU(approximate='none')
          (fc2): Linear(in_features=3072, out_features=768, bias=True)
          (drop): Dropout(p=0.0, inplace=False)
        )
        (norm_2): LayerNorm((768,), eps=1e-06, elementwise_affine=True)
      )
      (3): BlockWithMasking(
        (attn): MultiheadAttention(
          (out_proj): NonDynamicallyQuantizableLinear(in_features=768, out_features=768,
bias=True)
        )
        (drop_path): Identity()
        (norm_1): LayerNorm((768,), eps=1e-06, elementwise_affine=True)
        (mlp): Mlp(
          (fc1): Linear(in_features=768, out_features=3072, bias=True)
          (act): GELU(approximate='none')
          (fc2): Linear(in_features=3072, out_features=768, bias=True)
          (drop): Dropout(p=0.0, inplace=False)
        )
        (norm_2): LayerNorm((768,), eps=1e-06, elementwise_affine=True)
      )
      (4): BlockWithMasking(
        (attn): MultiheadAttention(
          (out_proj): NonDynamicallyQuantizableLinear(in_features=768, out_features=768,
bias=True)
        )
        (drop_path): Identity()
        (norm_1): LayerNorm((768,), eps=1e-06, elementwise_affine=True)
        (mlp): Mlp(
          (fc1): Linear(in_features=768, out_features=3072, bias=True)
          (act): GELU(approximate='none')
          (fc2): Linear(in_features=3072, out_features=768, bias=True)
          (drop): Dropout(p=0.0, inplace=False)
        )
        (norm_2): LayerNorm((768,), eps=1e-06, elementwise_affine=True)
      )
      (5): BlockWithMasking(
        (attn): MultiheadAttention(
          (out_proj): NonDynamicallyQuantizableLinear(in_features=768, out_features=768,
bias=True)
        )
```

```
        (drop_path): Identity()
        (norm_1): LayerNorm((768,), eps=1e-06, elementwise_affine=True)
        (mlp): Mlp(
          (fc1): Linear(in_features=768, out_features=3072, bias=True)
          (act): GELU(approximate='none')
          (fc2): Linear(in_features=3072, out_features=768, bias=True)
          (drop): Dropout(p=0.0, inplace=False)
        )
        (norm_2): LayerNorm((768,), eps=1e-06, elementwise_affine=True)
      )
      (6): BlockWithMasking(
        (attn): MultiheadAttention(
          (out_proj): NonDynamicallyQuantizableLinear(in_features=768, out_features=768,
bias=True)
        )
        (drop_path): Identity()
        (norm_1): LayerNorm((768,), eps=1e-06, elementwise_affine=True)
        (mlp): Mlp(
          (fc1): Linear(in_features=768, out_features=3072, bias=True)
          (act): GELU(approximate='none')
          (fc2): Linear(in_features=3072, out_features=768, bias=True)
          (drop): Dropout(p=0.0, inplace=False)
        )
        (norm_2): LayerNorm((768,), eps=1e-06, elementwise_affine=True)
      )
      (7): BlockWithMasking(
        (attn): MultiheadAttention(
          (out_proj): NonDynamicallyQuantizableLinear(in_features=768, out_features=768,
bias=True)
        )
        (drop_path): Identity()
        (norm_1): LayerNorm((768,), eps=1e-06, elementwise_affine=True)
        (mlp): Mlp(
          (fc1): Linear(in_features=768, out_features=3072, bias=True)
          (act): GELU(approximate='none')
          (fc2): Linear(in_features=3072, out_features=768, bias=True)
          (drop): Dropout(p=0.0, inplace=False)
        )
        (norm_2): LayerNorm((768,), eps=1e-06, elementwise_affine=True)
      )
      (8): BlockWithMasking(
        (attn): MultiheadAttention(
          (out_proj): NonDynamicallyQuantizableLinear(in_features=768, out_features=768,
bias=True)
        )
        (drop_path): Identity()
```

```
      (norm_1): LayerNorm((768,), eps=1e-06, elementwise_affine=True)
      (mlp): Mlp(
        (fc1): Linear(in_features=768, out_features=3072, bias=True)
        (act): GELU(approximate='none')
        (fc2): Linear(in_features=3072, out_features=768, bias=True)
        (drop): Dropout(p=0.0, inplace=False)
      )
      (norm_2): LayerNorm((768,), eps=1e-06, elementwise_affine=True)
    )
    (9): BlockWithMasking(
      (attn): MultiheadAttention(
        (out_proj): NonDynamicallyQuantizableLinear(in_features=768, out_features=768,
bias=True)
      )
      (drop_path): Identity()
      (norm_1): LayerNorm((768,), eps=1e-06, elementwise_affine=True)
      (mlp): Mlp(
        (fc1): Linear(in_features=768, out_features=3072, bias=True)
        (act): GELU(approximate='none')
        (fc2): Linear(in_features=3072, out_features=768, bias=True)
        (drop): Dropout(p=0.0, inplace=False)
      )
      (norm_2): LayerNorm((768,), eps=1e-06, elementwise_affine=True)
    )
    (10): BlockWithMasking(
      (attn): MultiheadAttention(
        (out_proj): NonDynamicallyQuantizableLinear(in_features=768, out_features=768,
bias=True)
      )
      (drop_path): Identity()
      (norm_1): LayerNorm((768,), eps=1e-06, elementwise_affine=True)
      (mlp): Mlp(
        (fc1): Linear(in_features=768, out_features=3072, bias=True)
        (act): GELU(approximate='none')
        (fc2): Linear(in_features=3072, out_features=768, bias=True)
        (drop): Dropout(p=0.0, inplace=False)
      )
      (norm_2): LayerNorm((768,), eps=1e-06, elementwise_affine=True)
    )
    (11): BlockWithMasking(
      (attn): MultiheadAttention(
        (out_proj): NonDynamicallyQuantizableLinear(in_features=768, out_features=768,
bias=True)
      )
      (drop_path): Identity()
      (norm_1): LayerNorm((768,), eps=1e-06, elementwise_affine=True)
```

```
        (mlp): Mlp(
          (fc1): Linear(in_features=768, out_features=3072, bias=True)
          (act): GELU(approximate='none')
          (fc2): Linear(in_features=3072, out_features=768, bias=True)
          (drop): Dropout(p=0.0, inplace=False)
        )
        (norm_2): LayerNorm((768,), eps=1e-06, elementwise_affine=True)
      )
    )
    (post_transformer_layer): EinOpsRearrange()
  )
  (imu): SimpleTransformer(
    (pre_transformer_layer): Sequential(
      (0): Identity()
      (1): EinOpsRearrange()
    )
    (blocks): Sequential(
      (0): BlockWithMasking(
        (attn): MultiheadAttention(
          (out_proj): NonDynamicallyQuantizableLinear(in_features=512, out_features=512,
bias=True)
        )
        (drop_path): Identity()
        (norm_1): LayerNorm((512,), eps=1e-06, elementwise_affine=True)
        (mlp): Mlp(
          (fc1): Linear(in_features=512, out_features=2048, bias=True)
          (act): GELU(approximate='none')
          (fc2): Linear(in_features=2048, out_features=512, bias=True)
          (drop): Dropout(p=0.0, inplace=False)
        )
        (norm_2): LayerNorm((512,), eps=1e-06, elementwise_affine=True)
      )
      (1): BlockWithMasking(
        (attn): MultiheadAttention(
          (out_proj): NonDynamicallyQuantizableLinear(in_features=512, out_features=512,
bias=True)
        )
        (drop_path): DropPath(drop_prob=0.140)
        (norm_1): LayerNorm((512,), eps=1e-06, elementwise_affine=True)
        (mlp): Mlp(
          (fc1): Linear(in_features=512, out_features=2048, bias=True)
          (act): GELU(approximate='none')
          (fc2): Linear(in_features=2048, out_features=512, bias=True)
          (drop): Dropout(p=0.0, inplace=False)
        )
        (norm_2): LayerNorm((512,), eps=1e-06, elementwise_affine=True)
```

```
      )
      (2): BlockWithMasking(
        (attn): MultiheadAttention(
          (out_proj): NonDynamicallyQuantizableLinear(in_features=512, out_features=512,
bias=True)
        )
        (drop_path): DropPath(drop_prob=0.280)
        (norm_1): LayerNorm((512,), eps=1e-06, elementwise_affine=True)
        (mlp): Mlp(
          (fc1): Linear(in_features=512, out_features=2048, bias=True)
          (act): GELU(approximate='none')
          (fc2): Linear(in_features=2048, out_features=512, bias=True)
          (drop): Dropout(p=0.0, inplace=False)
        )
        (norm_2): LayerNorm((512,), eps=1e-06, elementwise_affine=True)
      )
      (3): BlockWithMasking(
        (attn): MultiheadAttention(
          (out_proj): NonDynamicallyQuantizableLinear(in_features=512, out_features=512,
bias=True)
        )
        (drop_path): DropPath(drop_prob=0.420)
        (norm_1): LayerNorm((512,), eps=1e-06, elementwise_affine=True)
        (mlp): Mlp(
          (fc1): Linear(in_features=512, out_features=2048, bias=True)
          (act): GELU(approximate='none')
          (fc2): Linear(in_features=2048, out_features=512, bias=True)
          (drop): Dropout(p=0.0, inplace=False)
        )
        (norm_2): LayerNorm((512,), eps=1e-06, elementwise_affine=True)
      )
      (4): BlockWithMasking(
        (attn): MultiheadAttention(
          (out_proj): NonDynamicallyQuantizableLinear(in_features=512, out_features=512,
bias=True)
        )
        (drop_path): DropPath(drop_prob=0.560)
        (norm_1): LayerNorm((512,), eps=1e-06, elementwise_affine=True)
        (mlp): Mlp(
          (fc1): Linear(in_features=512, out_features=2048, bias=True)
          (act): GELU(approximate='none')
          (fc2): Linear(in_features=2048, out_features=512, bias=True)
          (drop): Dropout(p=0.0, inplace=False)
        )
        (norm_2): LayerNorm((512,), eps=1e-06, elementwise_affine=True)
      )
```

```
        (5): BlockWithMasking(
          (attn): MultiheadAttention(
            (out_proj): NonDynamicallyQuantizableLinear(in_features=512, out_features=512,
bias=True)
          )
          (drop_path): DropPath(drop_prob=0.700)
          (norm_1): LayerNorm((512,), eps=1e-06, elementwise_affine=True)
          (mlp): Mlp(
            (fc1): Linear(in_features=512, out_features=2048, bias=True)
            (act): GELU(approximate='none')
            (fc2): Linear(in_features=2048, out_features=512, bias=True)
            (drop): Dropout(p=0.0, inplace=False)
          )
          (norm_2): LayerNorm((512,), eps=1e-06, elementwise_affine=True)
        )
      )
      (post_transformer_layer): EinOpsRearrange()
    )
  )
  (modality_heads): ModuleDict(
    (vision): Sequential(
      (0): LayerNorm((1280,), eps=1e-06, elementwise_affine=True)
      (1): SelectElement()
      (2): Linear(in_features=1280, out_features=1024, bias=False)
    )
    (text): SelectEOSAndProject(
      (proj): Sequential(
        (0): LayerNorm((1024,), eps=1e-06, elementwise_affine=True)
        (1): Linear(in_features=1024, out_features=1024, bias=False)
      )
    )
    (audio): Sequential(
      (0): LayerNorm((768,), eps=1e-06, elementwise_affine=True)
      (1): SelectElement()
      (2): Linear(in_features=768, out_features=1024, bias=False)
    )
    (depth): Sequential(
      (0): LayerNorm((384,), eps=1e-06, elementwise_affine=True)
      (1): SelectElement()
      (2): Linear(in_features=384, out_features=1024, bias=False)
    )
    (thermal): Sequential(
      (0): LayerNorm((768,), eps=1e-06, elementwise_affine=True)
      (1): SelectElement()
      (2): Linear(in_features=768, out_features=1024, bias=False)
    )
```

```
    (imu): Sequential(
      (0): LayerNorm((512,), eps=1e-06, elementwise_affine=True)
      (1): SelectElement()
      (2): Dropout(p=0.5, inplace=False)
      (3): Linear(in_features=512, out_features=1024, bias=False)
    )
  )
  (modality_postprocessors): ModuleDict(
    (vision): Normalize()
    (text): Sequential(
      (0): Normalize()
      (1): LearnableLogitScaling(logit_scale_init=14.285714285714285,learnable=True,
max_logit_scale=100)
    )
    (audio): Sequential(
      (0): Normalize()
      (1): LearnableLogitScaling(logit_scale_init=20.0,learnable=False, max_logit_scale=100)
    )
    (depth): Sequential(
      (0): Normalize()
      (1): LearnableLogitScaling(logit_scale_init=5.0,learnable=False, max_logit_scale=100)
    )
    (thermal): Sequential(
      (0): Normalize()
      (1): LearnableLogitScaling(logit_scale_init=10.0,learnable=False, max_logit_scale=100)
    )
    (imu): Sequential(
      (0): Normalize()
      (1): LearnableLogitScaling(logit_scale_init=5.0,learnable=False, max_logit_scale=100)
    )
  )
)
```