



Image Encoder

Residual Transformer Layer 1



Interpreter

summer swing summer and summer - summer and
summer - summer red - summer - summer - summer
- summer - summer - summer - summer - summer

Residual Transformer Layer 18



Interpreter

a person standing in front of a tree with a house

Residual Transformer Layer 22



Interpreter

a person riding a bike on a dirt road

Residual Transformer Layer 23



Interpreter

a person riding a motorcycle on a dirt path

Residual Transformer Layer 24



Interpreter

a man riding a motorcycle on a dirt path .

.
.
.
.
.