Image Description Application for Dementia Prevention

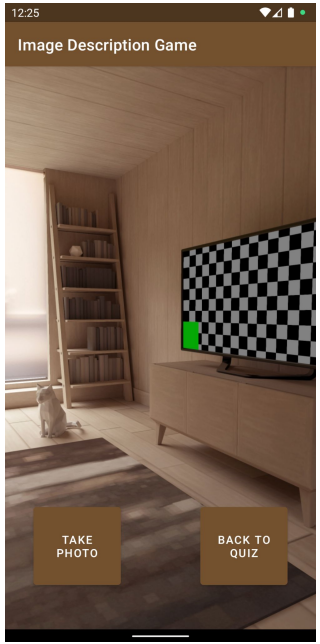
Seonghae Jo
Mobile Computing and its Application
Mini Project Final Presentation

Goal & Key Idea

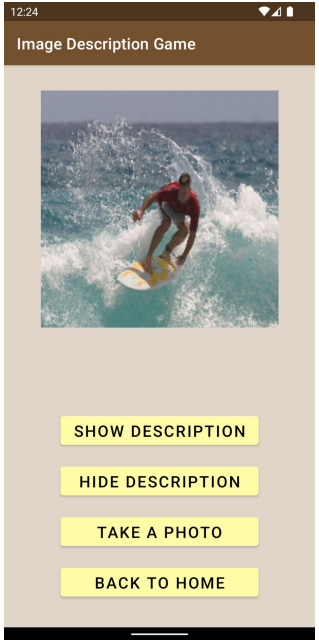
- Goal 1 : Help the users train their brain
  - Generate the image caption by using DNN model.
  - Application should help users to compare their own answer and the image caption.

- Goal 2 : More interaction with the user and the real world context
  - Application should provide the image capturing function.
  - Application should generate the caption for the image captured by user.

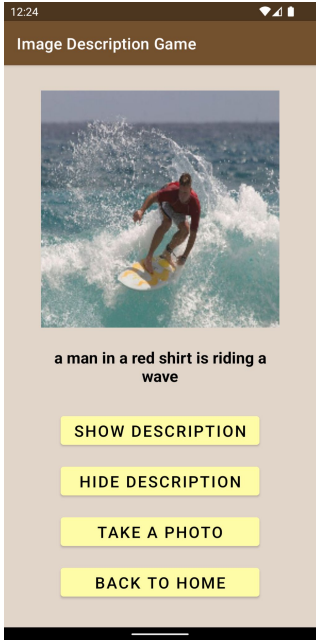
Usage Scenario



Take a photo

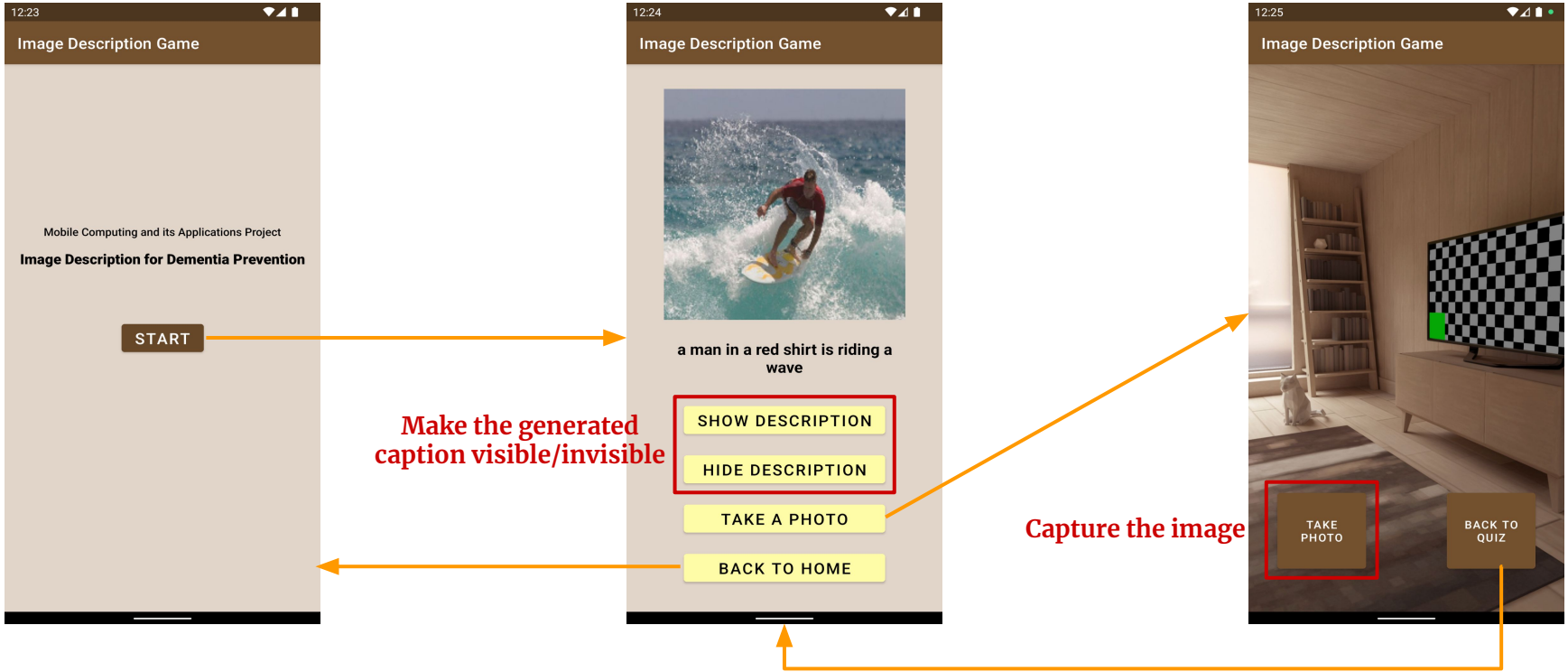


Try to describe the image

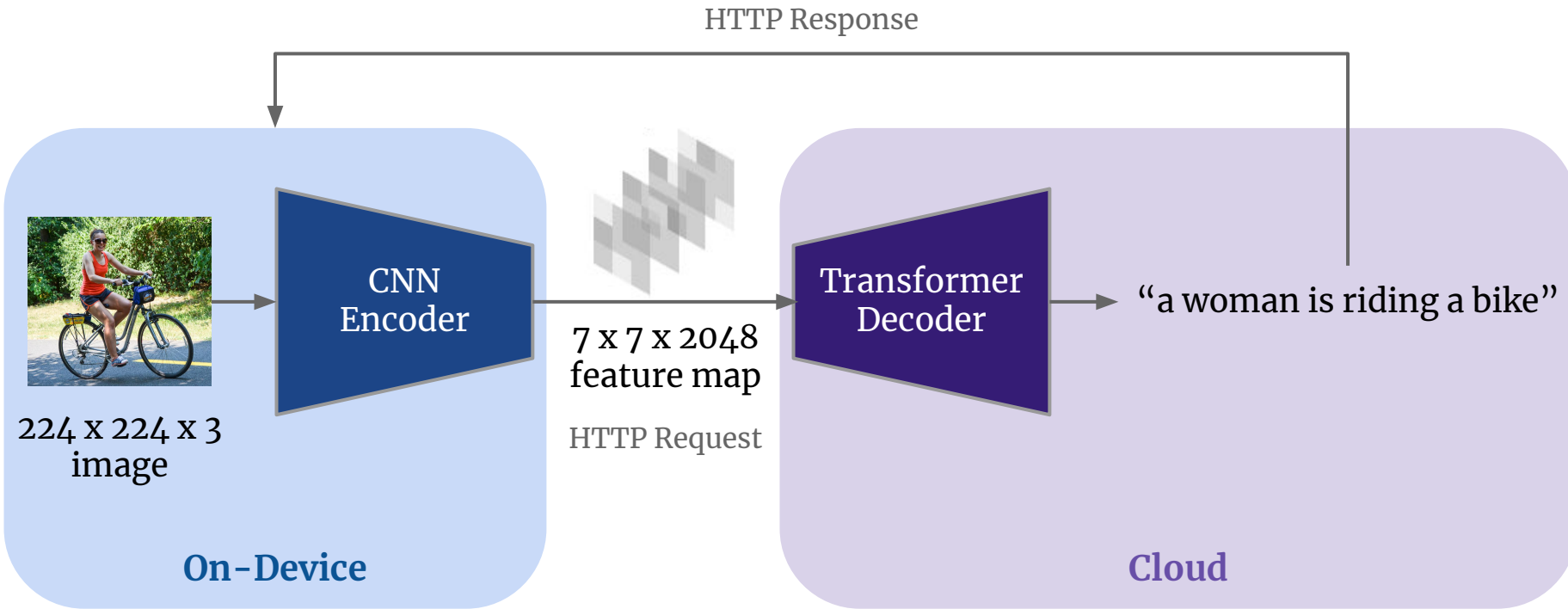


Compare with the generated caption

App Functionality



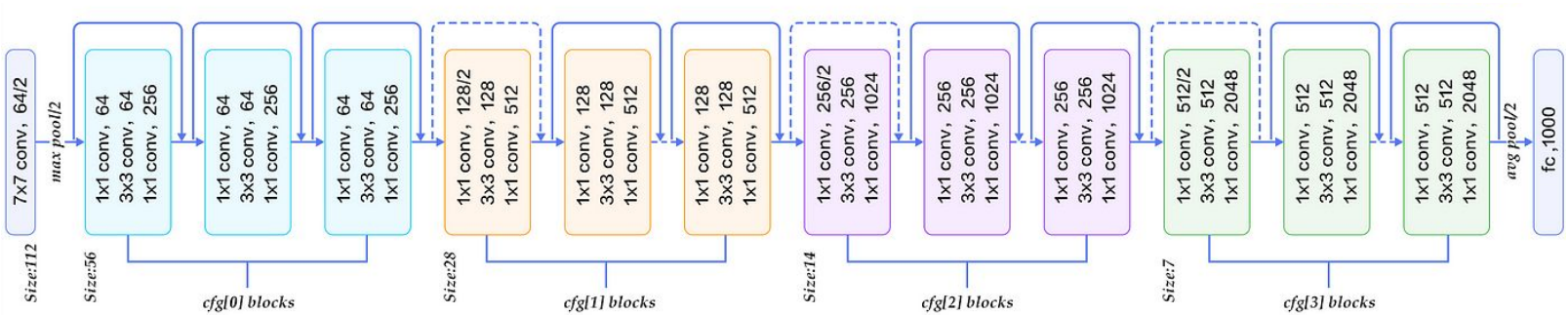
System Architecture



Encoder Model

ResNet50

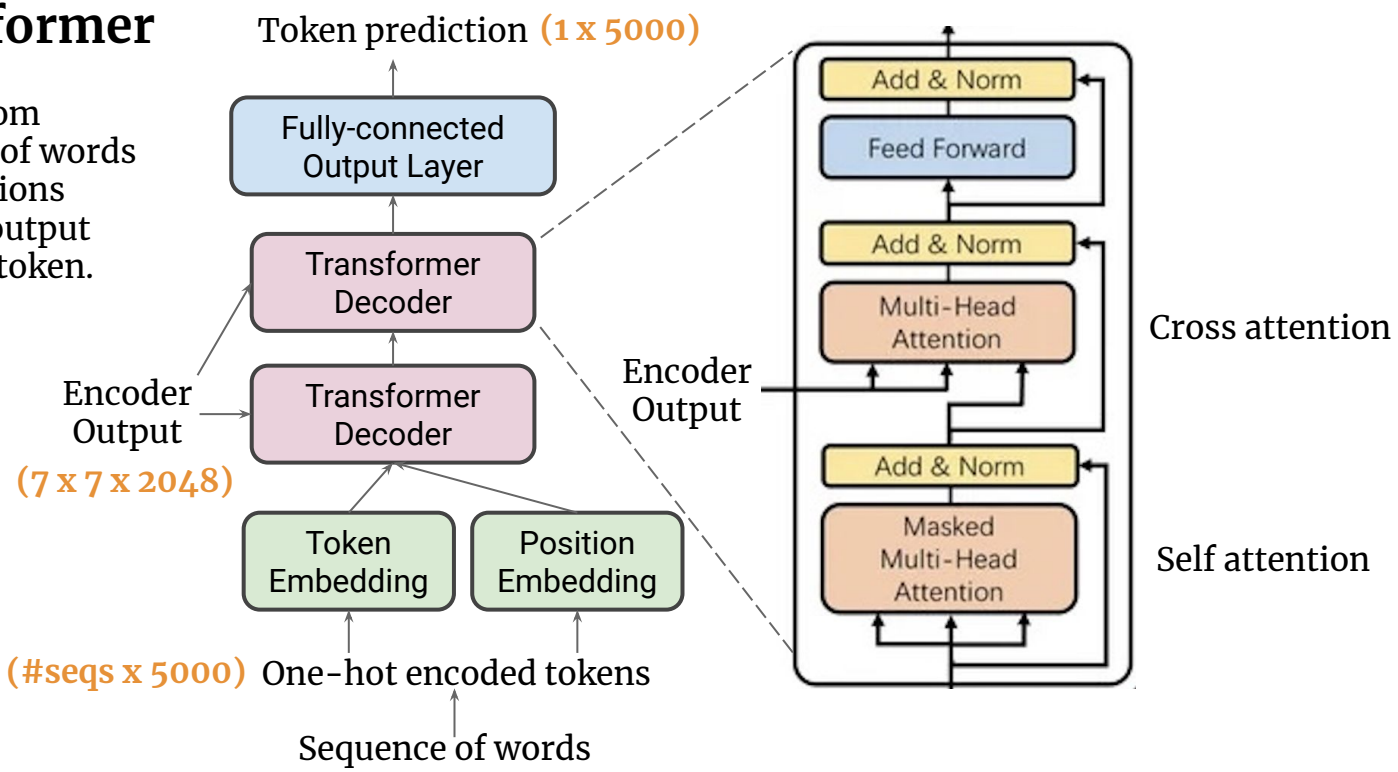
- Image feature extractor
- Pre-trained by ImageNet dataset
- Quantized to 8-bit integer
- Input shape : 224 x 224 x 3
- Output shape : 7 x 7 x 2048



Decoder Model

Two-Layer Transformer

- Input : Feature map from encoder and Sequence of words
- Output : Token predictions
- Generate the decoder output until output is <END> token.



Decoder Model

Two-Layer Transformer

- Word sequence generator
  - Generate one token prediction for one inference
- Trained by Flickr8k dataset (in Colab)
- Inference on cloud
  - Converting into tflite model is successful but..
  - Android Tensorflow Lite does not support dynamic input size.
  - Hosted on the Docker container Tensorflow image
  - Flask web server to process HTTP request

Model Optimization

- Number of parameters (Encoder vs Decoder)

	Encoder	Decoder	Ratio
#params	23.58M	8.88M	2.65x

- Quantize encoder!
  - Encoder has 2.65 times more parameters than decoder.
  - Decoder is cloud offloaded, but encoder is on device.

Model Optimization

- Post-training Quantization
  - 32-bit float to 8-bit integer (8-bit fixed point)
  - tf.lite.Optimize.DEFAULT
- Comparison of Before/After Quantization of encoder

	Accuracy	Loss	Model Size	Latency
Original	0.383	3.059	94 MB	895ms
Quantized	0.387	3.056	24 MB	529ms
Ratio	1.01x	0.99x	0.26x	0.59x

Accuracy: Percentage of labels==preds
Loss: CE loss

Demo Video