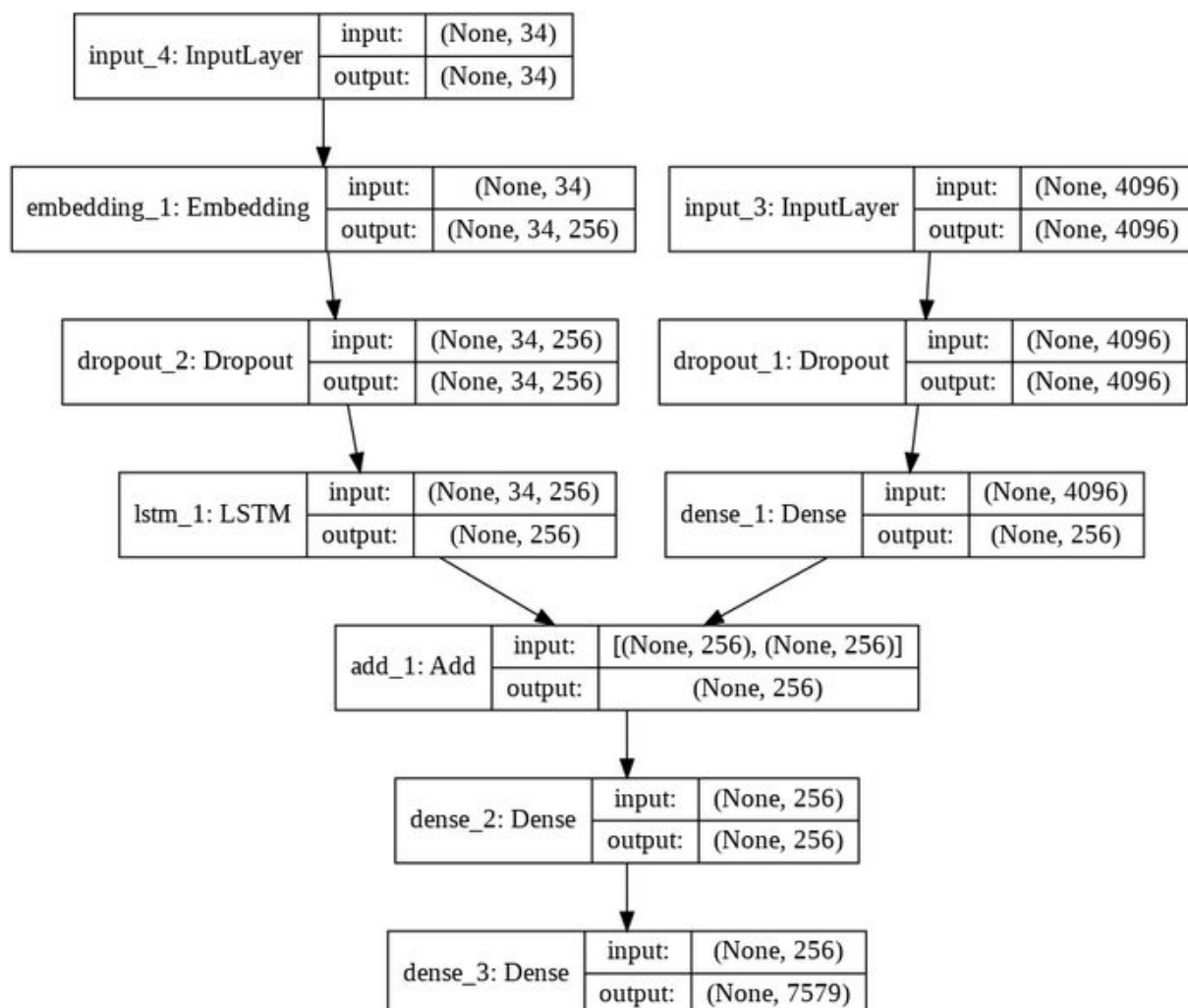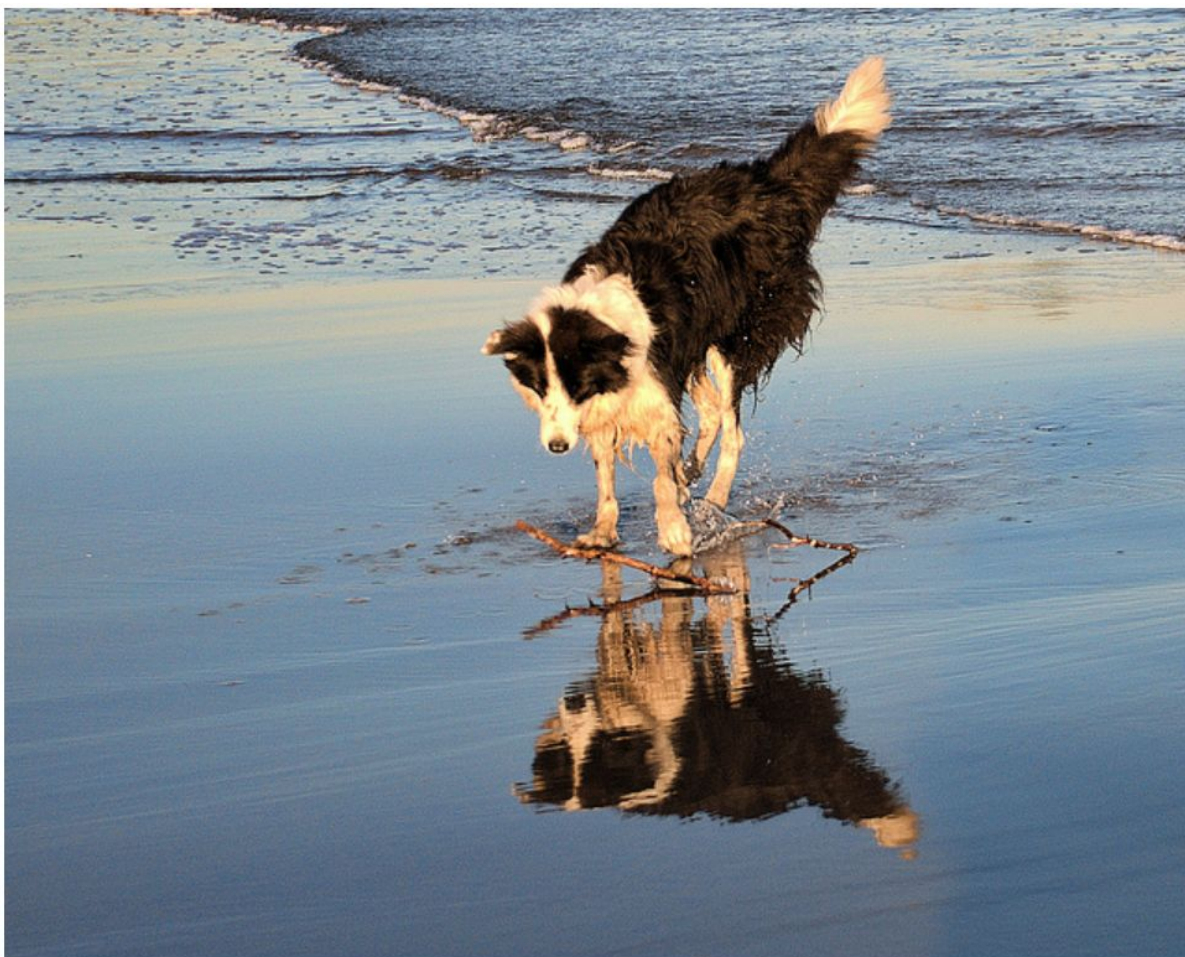# Image captioning using VGG and LSTM :

- **Photo Feature Extractor**. This is a 16-layer VGG model pre-trained on the ImageNet dataset. We have pre-processed the photos with the VGG model (without the output layer) and will use the extracted features predicted by this model as input.
- **Sequence Processor**. This is a word embedding layer for handling the text input, followed by a Long Short-Term Memory (LSTM) recurrent neural network layer.
- **Decoder (for lack of a better name)**. Both the feature extractor and sequence processor output a fixed-length vector. These are merged together and processed by a Dense layer to make a final prediction.

Results :

A black and White Dog near Sea

BLEU Measurements Results :

```
BLEU-1: 0.526051
BLEU-2: 0.279514
BLEU-3: 0.187424
BLEU-4: 0.080902
```

Maryam Meghdadi - Mahya Jamshidian