Key points

* Use max caption token of 32 (2sd above mean of 22), only 1% of captions have higher than this but the highest is 77!
* When processing data, just run the weights and bias upload. This is useful because it means it is saved straight away.
* Upload to wandb when home – it should now get it in the format [seq\_len, batch\_size, d\_model
* Struggled getting the right tensor shape and uploading to wandb
* Split data into:
* - Training: 129 examples
* - Testing: 15 examples
* Epoch 1/10, train Loss: 15.8046, Accuracy: 11.07%
* Epoch 1/10, test Loss: 8.7054, Accuracy: 57.50%
* Epoch 2/10, train Loss: 7.3679, Accuracy: 47.50%
* Epoch 2/10, test Loss: 4.8320, Accuracy: 57.29%
* Epoch 3/10, train Loss: 5.0734, Accuracy: 49.90%
* Epoch 3/10, test Loss: 3.5314, Accuracy: 60.00%
* Epoch 4/10, train Loss: 3.9320, Accuracy: 50.97%
* Epoch 4/10, test Loss: 2.9773, Accuracy: 59.17%
* Epoch 5/10, train Loss: 3.3844, Accuracy: 52.45%