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]