Fine tuning Donut model

I followed the tutorial : <https://www.philschmid.de/fine-tuning-donut>

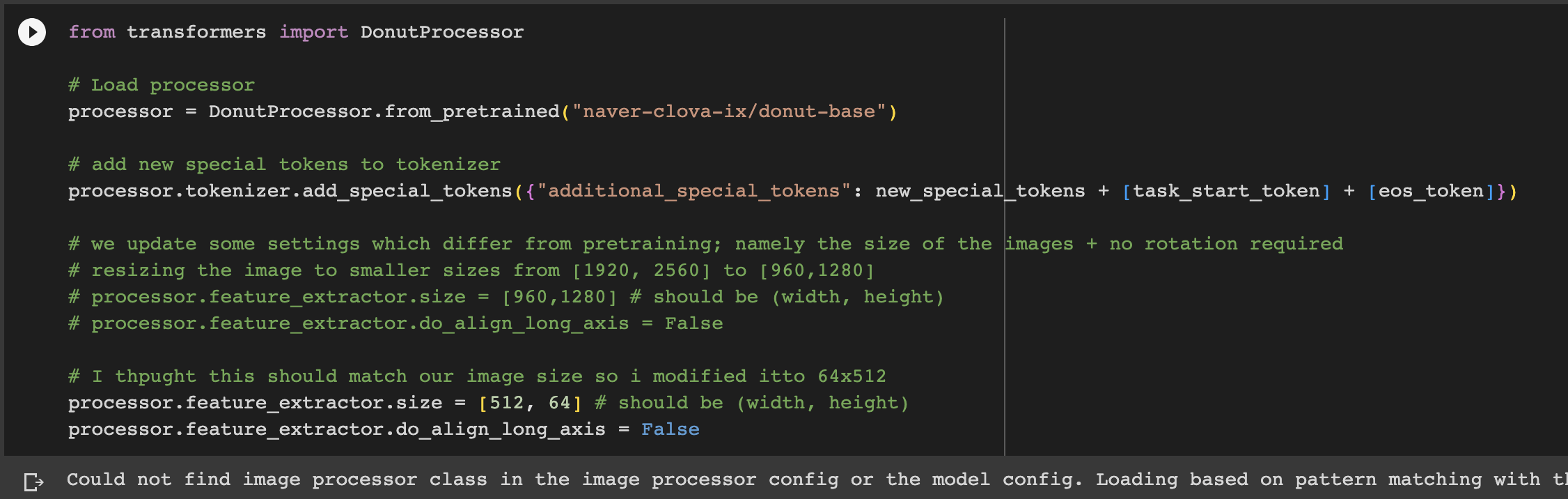
The hugging face model : <https://huggingface.co/naver-clova-ix/donut-base>

The code is on the github : <https://github.com/clovaai/donut>

Data set is built with a dataset object from dataset library with the following features :

* Our gray scale images : loaded as PIL images in RGB (model requires it)
* Labels : preprocessed and tokenized text appearing handwritten on the picture

The preprocessing and tokenisation is done and then a processor is built :



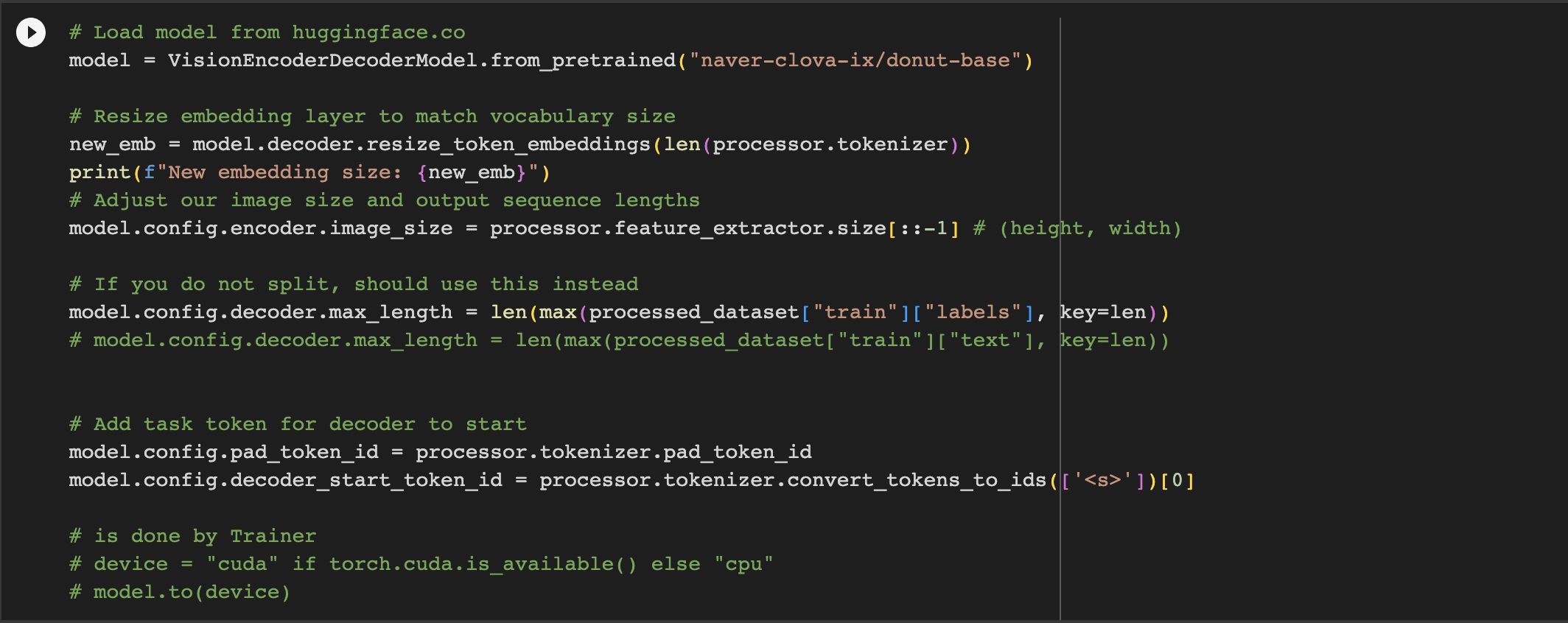
The size of the images are input into the configuration so that i matches the input layer

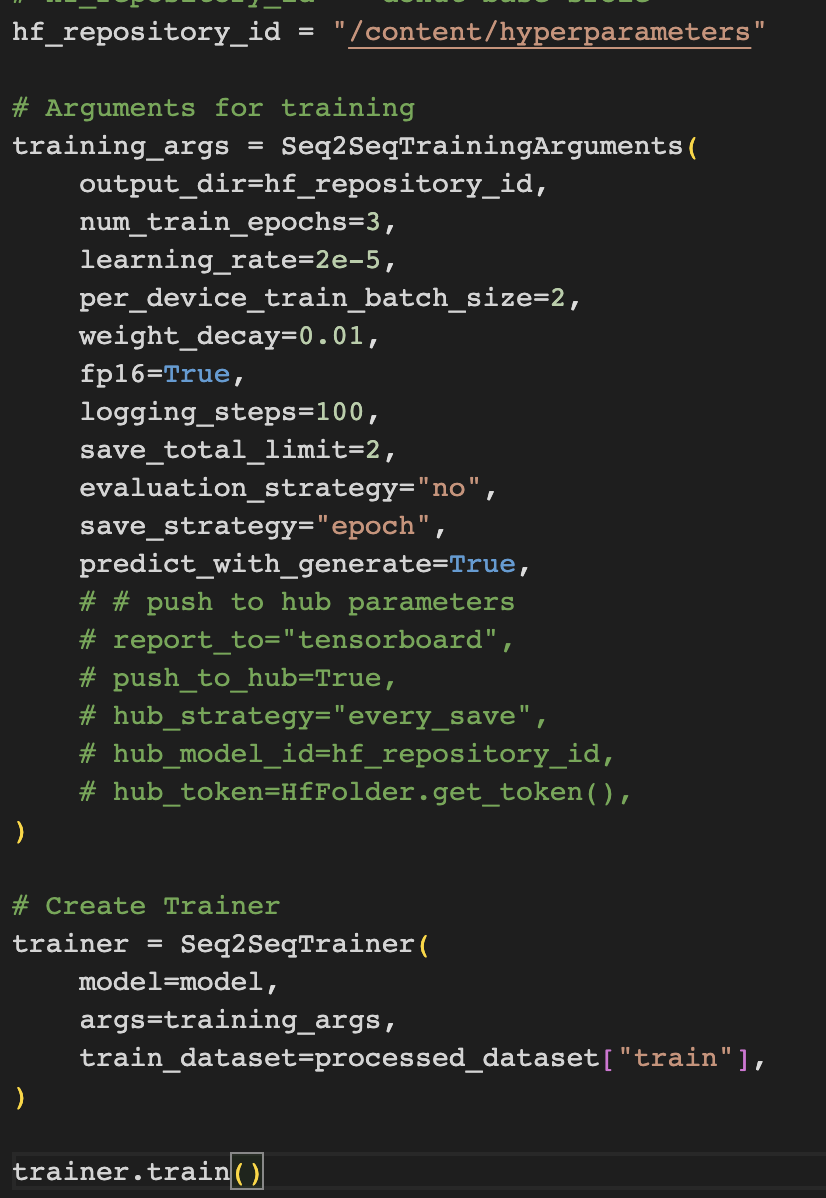


All images of the data set are preprocssed, i played with the dimension of pixel\_values using .permute and size but it did not change anything.

Model :

Model is downloaded from huggingface api and we can see that model’s encoder, decoder and configuration is built from processor.



The Runtime error is generated by the last line of code : trainer.train().

We can see there is a mismatch of dimensions between tensors.

This could come from the three arguments provided for trainer :

* training\_args
* model
* processed\_dataset

