Dataset containing Parquet Files - figure out how to use

decode bytes into images

↓

change images into arrays

images labelled 0, 1, 2, 3
- examine class distribution as part of preprocessing
- if class imbalance can change to binary classification problem

data is already split into training and testing - just have to set up X_train y_train X_test y_test and make training and testing datasets

need to batch and shuffle data, use batch size 32 first

Model architecture - insp. from previous project

Sequential CNN

input layer - (256, 256, 3)

↓

Conv2D (32 (3,3) relu)

↓

MaxPooling2D

↓

Conv2D (64 (3,3) relu)

↓

MaxPooling2D

↓

Conv2D (128 (3,3) relu)

↓

MaxPooling2D

↓

Flatten

↓

Dense (256, relu)

↓

Dropout (?) * maybe try 0.5 first

↓

Dense (1, sigmoid) - change probabilities into decision

model.compile(adam, binarycrossentropy, accuracy)

model.fit (... epochs = 20)
- use earlystopping to prevent overfitting - patience of 3-5

make confusion matrix and classification report

Also plot loss and accuracy during training - make sure to save model history for this!!

make saliency maps (or try to)