**Making a personalised photo album using cluster learning**

I have almost 100,000 images saved in my phone. They include my daily shots; scans of my reading and notes; screenshots from when I look at social networking sites; images from my collection and so on. And many of the pictures taken on a daily basis, many duplicates of which were taken at the time, have not been deleted or managed since. The sheer volume of images makes them very difficult to find and takes up so much space that I can never go through them all. I often feel that the sorting function that comes with ios albums is ineffective for me.

Currently ios comes with several main types of image sorting.

1. geolocation-based sorting. The advantage is that it can be used as a kind of geolocation memo; the disadvantage is that it often includes a lot of duplicate travel photos.

2. Face recognition based classification. The advantage is that individual memories can be created for each person; the disadvantage is that it is not sufficiently human and many memories do not want to be seen repeatedly.

3. Classification based on timeline. It is a more traditional way of organising albums.

But I have a hard time organizing and stuffing my images with a specific classification. So clustering learning is a good direction to try.

Methods 1 and 3 above rely on metadata obtained at the same time as the photos, and 2 makes use of a deep learning model. A better way to organise photos is to extract semantic information from the images themselves.

Let's break this idea down into parts. Suppose we have a diversity of similar (as described above) photos. What features can our algorithmic model capture?

1. distinguish whether it is a photo or a screenshot

2. photos with text (notes, reading), which often act as memos

3. whether the image matches my aesthetic and is the one I would choose

With these evaluation criteria, I put the photos that meet the criteria into different folders for clustering. Clustering learning can be done using unsupervised classification based on the K-means algorithm. This is a classification model trained on my own preferences, so this database should be my own photo albums. To test the accuracy of the clustering, we can take new photos at random and see if the folders the model classifies for it meet my criteria.

By searching the internet, I found a relatively complete process for doing image clustering analysis using K-means. It includes the steps: 1) taking unlabelled images as input images and randomly selecting blocks of images to form a set of unlabelled images of the same size.

2) extracting the best clustering centres once using the K-means algorithm.

3) constructing a feature mapping function to extract the image features of the unlabelled image set.

4) performing pooling operation with normalization.

5) extracting the secondary best clustering centres using the K-means algorithm and employing convolution operations to extract the final image features and normalize the final image features.