**Tutorials on Assignment 1**

Herein, we offer two tutorials on Assignment 1, covering both handcrafted and deep learning features. It is worth mentioning that the aim of the tutorials is to provide you with examples and help you better finish Assignment 1. However, it does not imply any restrictions. For example, the handcrafted feature used in Assignment 1 is not restricted to SIFT, and the deep learning network used is not restricted to VGG11 backbones. Moreover, we have no restrictions on programming language.

In the tutorial of handcrafted feature, we provide the SIFT based version. In particular, this is only a showcase, and would not help you achieve better performance if directly adopting the SIFT feature. The following directions which may help you better achieve the goal include, but are not limited to,

1) Use the instance location information of the query image, which is stored in txt files, to extract more specific information.

2) Use sliding windows on the gallery images to extract the features from the windows to achieve elaborate search.

3) Other feature extraction methods could also be applied. You could also combine various features for better performance.

4) The matching method in the tutorial is quite straightforward and you could explore other matching methods for better results.

In the tutorial of deep learning features, we extract features from the last layer of pretrained VGG11 backbone. In particular, this is only a showcase, and would not help you achieve better performance if directly adopting the features. The following directions which may help you better achieve the goal include, but are not limited to,

1) In the showcase, we adopt VGG11 as our backbone. You could use other backbones (*e.g.* ResNet, XceptionNet, AlexNet, ...) for feature extraction.

2) In the showcase, we extract features from the last layer of the backbone. You could extract features from more than one layer and combine them to perform feature similarity measurement.

3) In the showcase, we feed the whole gallery image into the neural network, which will lead to poor instance search performance. You could preprocess the gallery images by cropping them using windows with different sizes and shapes to get more fine-grained features.

4) The showcase employs the cosine similarity to measure the similarity scores between features. You could use other metrics to achieve better instance search results.

Please email our Teaching Assistant: Linfeng ZHENG(Email: lfzheng3-c@my.cityu.edu.hk) or Bolin CHEN (Email: bolinchen3-c@my.cityu.edu.hk) for any technical questions.