EE3 Deep Learning Interim Report

Alexander Montgomerie-Corcoran

CID: 01052454, am9215@ic.ac.uk

Abstract

This work investigates verification, matching and retrieval tasks for the N-HPatches dataset, a noisy version of the HPatches dataset [1]. A baseline approach is taken by first performing denoising on each patch, then deriving a descriptor for a patch. The baseline method achieves 0.086 mAP for matching, 0.64 mAP for retrieval and TODO for verification.

1. Introduction

1.1. Dataset

The HPatches dataset consists of sequences of patches extracted from various images. For every reference image, there are a set of patch sequences for the same subject with either viewpoint changes or illumination changes. These illumination variations can be characterised as varying brightness across the reference image. Viewpoint changes come from different perspectives of the subject, as well as artificial geometric noise applied to the patches. These non-reference sequences are ranked by their viewpoint and illumination variation, and are either *easy, hard* or *tough*. N-HPatches applies an additional random pixel noise to each patch, with the intensity of this noise based on the patch's rank.

1.2. Tasks

The N-HPatches dataset is used for three different problem classes.

The patch verification task consists of classifying the similarity of individual patches. A good descriptor is able to tell if two patches are the same, regardless of any noise.

Image matching tasks use a sequence of patches to classify the similarity of the image they are extracted from. This descriptor will take a sequence of both reference patches and test patches, and derive the similarity of these sequences using a given function. This is usually done by applying a metric to a sequence of descriptors.

The patch retrieval task attempts to find a patch within a group of patches which correspond to the reference patch.

2. Baseline Approach

The baseline method separates the problem into denoising and descriptor learning, and employs popular methods for each.

2.1. Patch Denoising

The patch denoising network makes use of U-Net [2], a popular image segmentation network with use in denoising tasks aswell. This network utilises convolutional layers to extract the major features of the image. Through upsampling and layer concatenation, these features are recombined to produce a reconstructed image with less noise.

This network is trained on noisy images evaluated against noiseless images. L2 loss is employed. Figure **??** shows that it is able to converge to reach both low validation and training loss, however looking a example outputs in figure **??**, there is obvious variation between the example and denoised image. As the validation loss decreases, this suggests that a more appropriate loss function is needed.

2.2. Patch Descriptor

A network is defined to extract discriminant features of each patch. This network is derived from L2-net ... The network consisted of convolutional layers which extract shift-invariant features of the patch. A fully connected layer is used to obtain the most significant features. This network is trained in using triplets , with a reference, positive and negative example given. The network learns features such that the L2 distance is maximised between reference and negative features, and minimised between positive and reference features.

There are improvements to be made to the loss function for this network also.

3. Improved Approach

To improve on the de-noising network, loss functions such as L1 loss and SISSM will be investigated, as they have proven more appropriate for image reconstruction tasks [**?**].

References

[1] V. Balntas, K. Lenc, A. Vedaldi, and K. Mikolajczyk. Hpatches: A benchmark and evaluation of handcrafted and learned local descriptors. In *CVPR*, 2017.

[2] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, 2015.