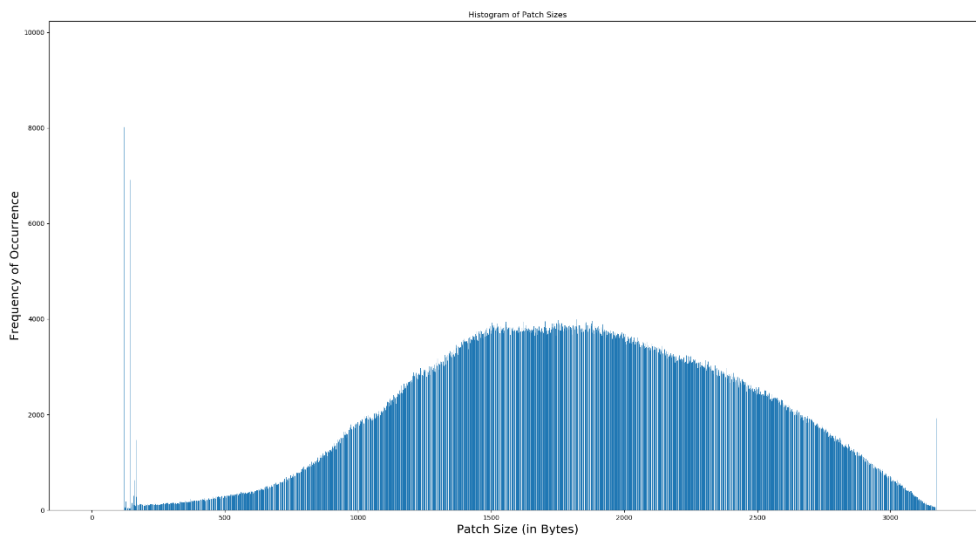


Dataset for the Image Codec

This dataset was prepared using images from the following datasets:

1. CLIC Dataset. (<http://www.compression.cc/challenge/>) (Challenge on Learned Image Compression) - High Quality images.
 - a. Professional Valid: 41 images
 - b. Professional Train: 585 images
 - c. Mobile Valid: 61 images
 - d. Mobile Train: 1048 images
2. DIV2K Dataset. (<https://data.vision.ee.ethz.ch/cvl/DIV2K/>) (DIVERse 2K resolution high quality images) - High Resolution images
 - a. Train: 800 images
 - b. Valid: 100 images
3. Ultra-Eye Dataset (<https://mmspg.epfl.ch/downloads/ultra-eye/>) (Ultra-Eye: UHD and HD images eye tracking dataset) – High Quality, high resolution.
 - a. HD: 38 images
 - b. UHD: 40 images

First, all images were separated in 32x32 patches, resulting in 6,231,440 patches. Each image was encoded losslessly in PNG format, and the size of each file is used as a proxy for the patch entropy (i.e., patches with a small file size are considered “low entropy”). The histogram of the whole dataset is shown here:



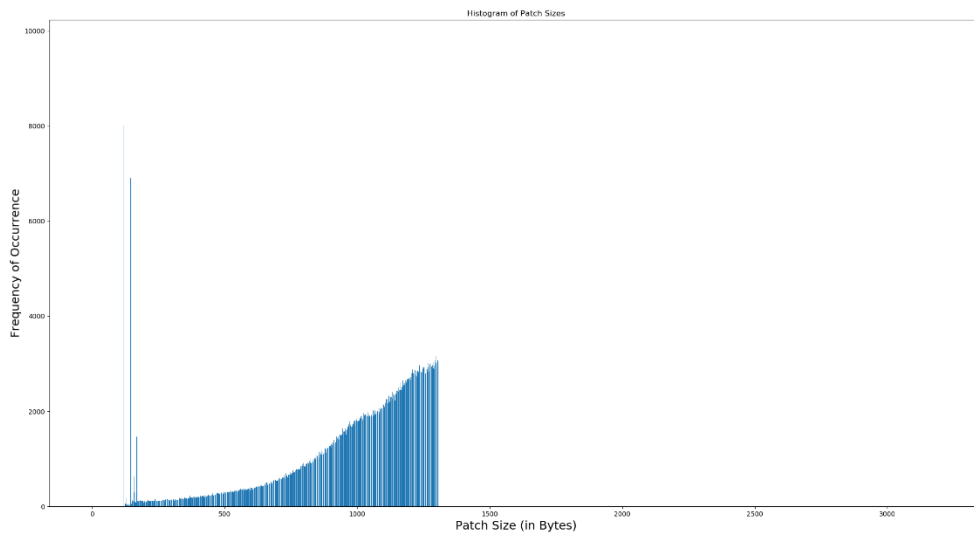
Histogram of entire Database

Then, each database was selected with around 1.25 million patches and the following rules:

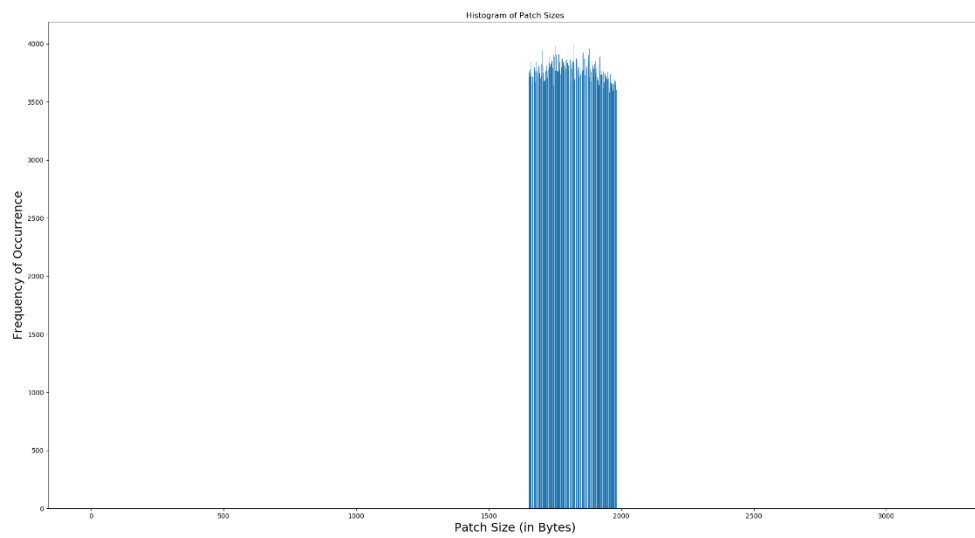
- Database0: 1,248,978 patches. All patches in the lower 20% entropy.
- Database1: 1,251,421 patches. All patches within 40% and 60% entropy.
- Database2: 1,248,725 patches. All patches in the top 20% entropy.
- Database3: 1,247,033 patches. 20% of the patches randomly drawn from the entire database.
- Database4: 1,246,698 patches. 20% of the patches randomly drawn from the top 50% entropy.

By construction there is no overlap among databases 0, 1 and 2, but there is an overlap of these databases with databases 3 and 4.

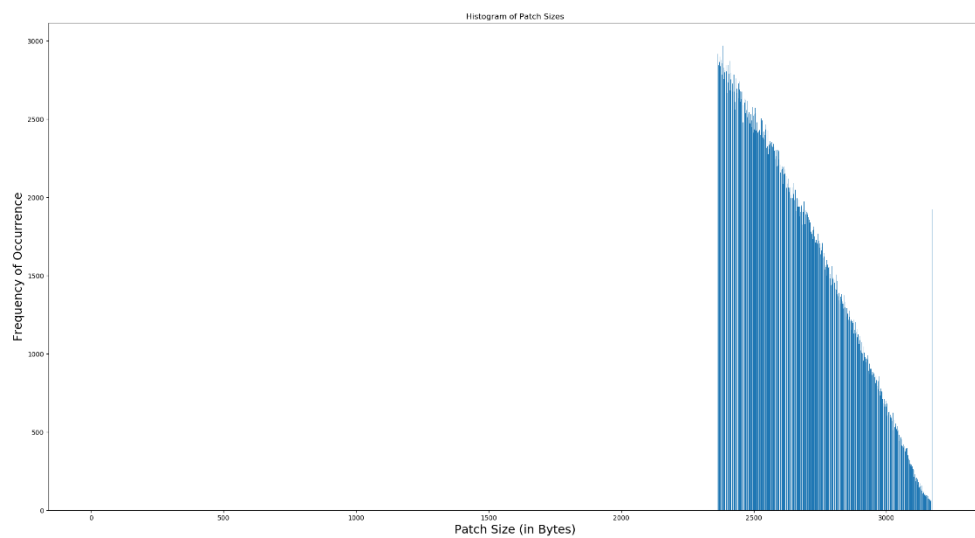
A histogram of each database is given:



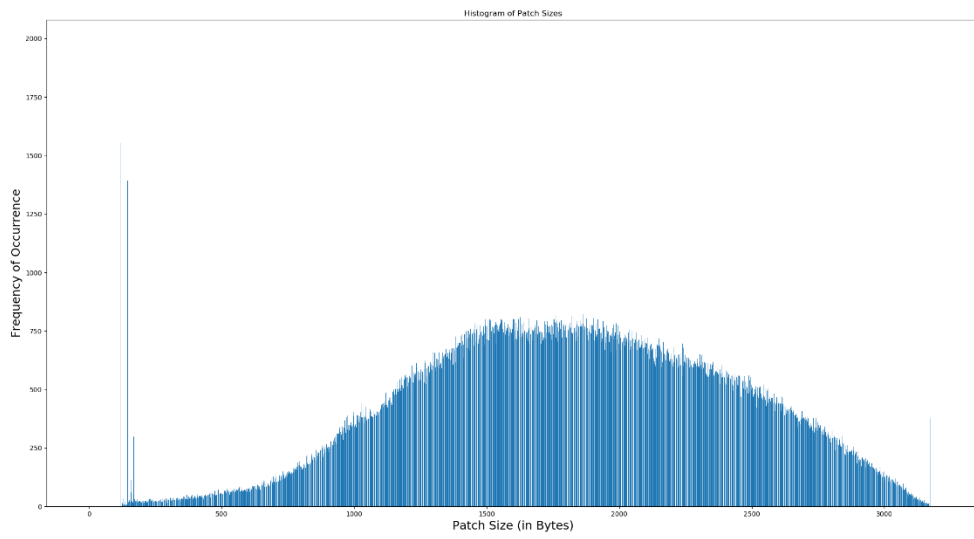
Histogram of Database0



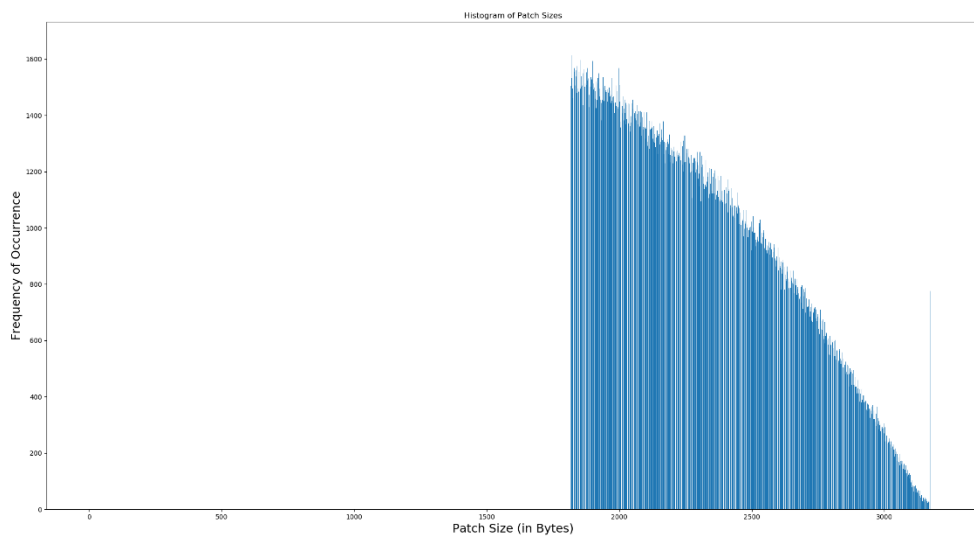
Histogram of Database1



Histogram of Database2



Histogram of Database3



Histogram of Database4