A MAHANALOBIS DISTANCE AND WHITENING

The Mahalanobis distance (Chandra et al., 1936; McLachlan, 1999) between two samples x and x' drawn from a random vector X with covariance Σ is defined as

$$D_M(x, x') = \sqrt{(x - x')^T \Sigma^{-1} (x - x')}$$

If the random vector X has identity covariance, it is simply the usual euclidian distance :

$$D_M(x, x') = \|x - x'\| \, .$$

Using the diagonalization of the coraviance matrix, $\Sigma = P \Lambda P^T$, the affine whitening operators of the random vector $\mathbf{X}$ are the operators

$$w : \mathbf{X} \mapsto O \Lambda^{-1/2} P^T (\mathbf{X} - \mu), \quad \forall O \in O_n(\mathbb{R}) \, . \tag{5}$$

For example, the PCA whitening operator is

$$w_{\mathrm{PCA}} : \mathbf{X} \mapsto \Lambda^{-1/2} P^T (\mathbf{X} - \mu)$$

and the ZCA whitening operator is

$$w_{\mathrm{ZCA}} : \mathbf{X} \mapsto P \Lambda^{-1/2} P^T (\mathbf{X} - \mu) \, .$$

For all whitening operator w we have

$$\|w(x) - w(x')\| = D_M(x, x')$$

since

$$\begin{aligned} \|w(x) - w(x')\| &= \|O \Lambda^{-1/2} P^T (x - x')\| \\ &= \sqrt{(x - x')^T P \Lambda^{-1/2} O^T O \Lambda^{-1/2} P^T (x - x')} \\ &= \sqrt{(x - x')^T P \Lambda^{-1} P^T (x - x')} \\ &= D_M(x, x') \, . \end{aligned}$$

B IMPLEMENTATION OF THE PATCHES K-NEAREST-NEIGHBORS ENCODING

In this section, we explicitly write the whitened patches with the whitening operator W. Recall that we consider the following set of euclidean pairwise distances:

$$\mathcal{C}_{i,x} = \{\|W p_{i,x} - W d\| \, d \in \mathcal{D}\} \, .$$

For each image patch we encode the K nearest neighbors of $W p_{i,x}$ in the set $W d, d \in \mathcal{D}$, for some $K \in 1 \dots |\mathcal{D}|$. We can use the square distance instead of the distance since it doesn't change the K nearest neighbors. We have

$$\|W p_{i,x} - W d\|^2 = \|W p_{i,x}\|^2 - 2 \langle p_{i,x}, W^T W d \rangle + \|W d\|^2$$

The term $\|W p_{i,x}\|^2$ doesn't affect the K nearest neighbors, so the K nearest neighbors are the K smallest values of

$$\left\{ \frac{\|W d\|^2}{2} + \langle p_{i,x}, -W^T W d \rangle, \; d \in \mathcal{D} \right\}$$

This can be implemented in a convolution of the image using $-W^T W d$ as filters and $\|W d\|^2 / 2$ as bias term, followed by a "vectorwise" non-linearity that binary encodes the K smallest values in the channel dimension. Once this is computed, we can then easily compute

$$\left\{ \frac{\|W d\|^2}{2} + \langle p_{i,x}, W^T W d \rangle, \; d \in \mathcal{D} \right\}$$

which is the quantity needed to compute the K nearest neighbors in the set of negative patches $\overline{\mathcal{D}}$. This is a computationally efficient way of doubling the number of patches while making the representation invariant to negative transform.

C ABLATION STUDY ON CIFAR-10

For this ablation study on CIFAR-10, the reference experiment uses $|\mathcal{D}| = 2048$ patches, a patch size $Q = 6$ a number of neighbors $K = 0.4 \times 2048 = 820$ and a whitening regularizer $\lambda = 1e-3$, and yields 82.5% accuracy. Figure 4 shows the results in high resolution.

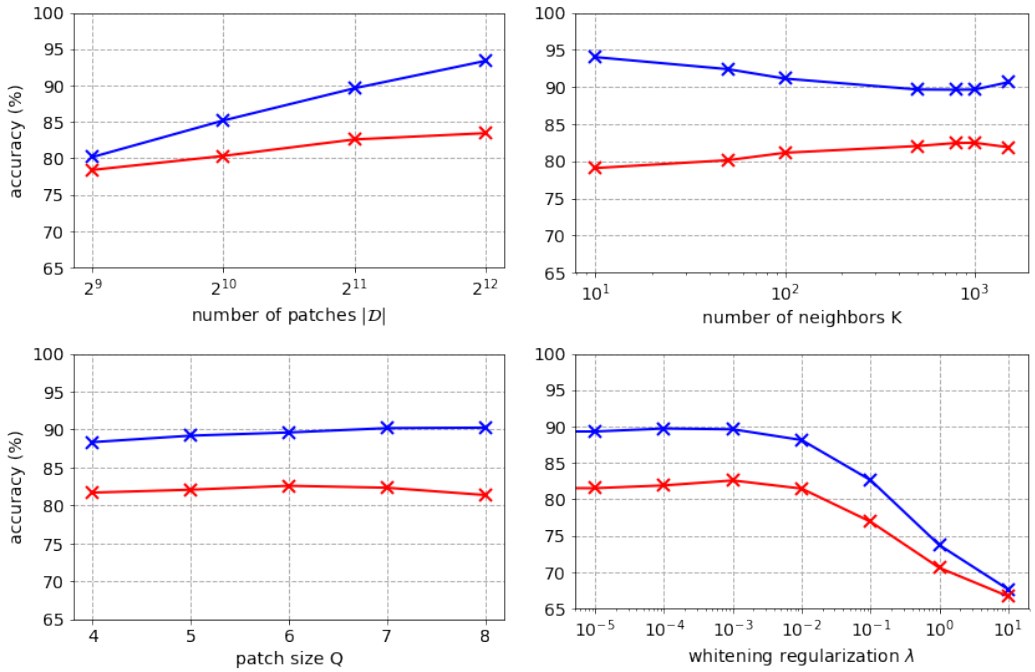

Figure 4: CIFAR-10 ablation experiments, train accuracies in blue, test accuracies in red. Number of patches $|\mathcal{D}|$ varies in $\{512, 1024, 2048, 4096\}$, number of neighbors K varies in $\{10, 50, 100, 500, 800, 1000, 1500\}$, patch size Q varies in $\{4, 5, 6, 7, 8\}$, whitening regularization λ varies in $\{0, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 10\}$.

D INTRINSIC DIMENSION ESTIMATE

The following estimate of the intrinsic dimension d_{int} is introduced in Levina and Bickel (2004) as follows

$$d_{\text{int}}(p) = \left(\frac{1}{K-1} \sum_{k=1}^{K-1} \log \frac{\tau_K(p)}{\tau_k(p)} \right)^{-1}, \tag{6}$$

where $\tau_k(p)$ is the euclidean distance between the patch p and it's k-th nearest neighbor int the training set.