Optimization-based Whitening

Chenguang Zhu (SUNET ID: cgzhu)

1. Introduction

In natural image understanding, the whitening step plays an important role, especially within many unsupervised feature learning algorithms. Examples of these algorithms include ICA, TICA, Auto-encoder, and so forth. Current whitening techniques include Principal Component Analysis (PCA) and Zero-phase Component Analysis (ZCA). However, the drawbacks of these approaches are obvious: the time complexity is proportional to the cubic of the number of image pixels. When dealing with large images like photos in resolution 640*480, these methods are simply infeasible.

In this paper, we probe into ZCA-based whitening filters and observe strong signs of locality, based on which we propose incorporating local receptive fields into whitening to speed up. Secondly, we introduce convolution and symmetry into whitening to further simplify the procedure. Experimental results are reported for both the visual object recognition tasks CIFAR-10 and natural image dataset NIS.

Te rest of this paper is organized as following. We introduce ZCA-based whitening in Section 2. Then, we present optimization-based whitening in Section 3. We report experimental results in Section 4 and conclude the paper in Section 5.

2. Zero-phase Component Analysis

Zero-phase Component Analysis (ZCA) is widely used in current natural image understanding tasks. The goal of ZCA is to whiten images, i.e. applying linear transformation to decorrelate image features. To put it mathematically, suppose an image is represented by a random vector $X \in R^d$, the goal is to find a whitening filter $W \in R^{a*d}$ transforming X into dimension of a ($a = d$ in most cases), such that $\mathrm{Cov(WX) = I}$, i.e. the resulting data WX has an identity covariance matrix. Typically, X is pre-processed to ensure $EX = 0$, a.k.a removing DC-component. Therefore, it is sufficient to make $\mathrm{E(WXX^TW^T) = I}$ in order to make $\mathrm{Cov(WX) = I}$.

For real world data, there can only be limited number of images for training. Here, we overload the symbol X to represent the training data: $X \in R^{d*n}$, where there are n images. Then, the objective is to find $W \in R^{a*d}$ such that $\frac{WXX^TW^T}{n} = I$. If we define $\frac{XX^T}{n} = A$, the goal becomes:

$$WAW^T = I \tag{1}$$

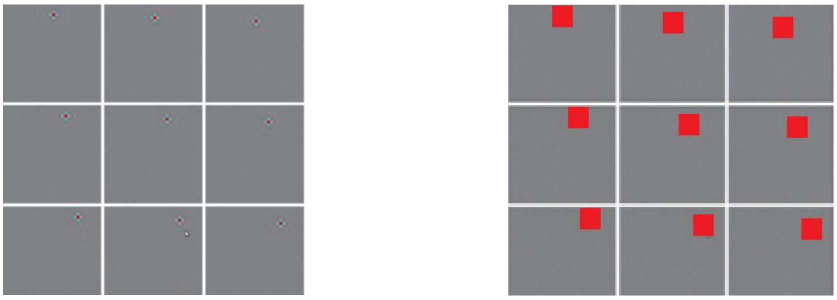
ZCA filters take advantage of the fact that A is a symmetric positive semi-definite matrix. It follows that we can diagonalize A: $A = EDE^T$, where E is an orthogonal matrix and D is a diagonal matrix. Then, the solution W_z given by ZCA is:

$$W_z = ED^{-\frac{1}{2}}E^T \tag{2}$$

It follows that $W_z A W_z^T = ED^{-\frac{1}{2}}E^T EDE^T ED^{-\frac{1}{2}}E^T = ED^{-\frac{1}{2}}DD^{-\frac{1}{2}}E^T = EE^T = I$.[1] In Figure 1(a),

[1] Here, we can deal with zeros on the diagonals of matrix D by adding a small identity matrix to D: $W_z = E(D + \epsilon I)^{-\frac{1}{2}}E^T$,

we display some columns of the resulted W_z (also called filters) calculated from CIFAR-10 dataset.



(a) (b)

Figure 1. (a) A selection of ZCA-based whitening filters for CIFAR-10 dataset. Each filter is displayed as a 32*32 patch. Brighter pixels represent values with larger absolute values. (b) For each filter in OWF scheme, we only parametrize values within a small neighbourhood (denoted by red), setting all other values to 0 (denoted by gray).

As clearly seen, the ZCA-based whitening filters present a strong sign of locality, which means that for each filter, only a tiny window contains values that are far away from zero. Because ZCA takes time proportional to d^3 (in diagonalization), it will be beneficial if we can take advantage of the locality property by directly optimizing over these local filter values.

3. Optimization-based Whitening Filter

In this section, we present the optimization-based whitening filter (OWF), to exploit the locality property of ZCA-based whitening filters. In particular, for each whitening filters, we only parametrize values within a tiny window of size $s * s$. For example, if an image has d pixels, there can be d filters, of which the windows are centered at d different locations in the image (Figure 1(b)). This technique is also called local receptive field (LRF).

Unfortunately, after applying LRF, we can no longer find a closed form solution of W like in ZCA. Instead, we can measure the quality of different W by looking at the distance from covariance matrix to identity matrix:

$$f(\theta) = ||W(\theta)AW(\theta)^T - I||_F^2 \quad (3)$$

Where we denote the set of parameters by θ and the induced sparse whitening filter matrix by $W(\theta)$. In each step of optimization, we calculate $\nabla_W f$, and then $\nabla_\theta f$, via collecting values in the corresponding windows in $\nabla_W f$.

$$\nabla_W f = \nabla_W \left||W(\theta)AW(\theta)^T - I|\right|_F^2 = \nabla_W tr(WAW^T - I)(WAW^T - I)^T$$

$$= \nabla_W (trWAW^T WAW^T - 2trWAW^T) = 4\mathrm{WA}(\mathrm{W}^\mathrm{T}\mathrm{WA} - \mathrm{I})$$

3.1 Variants of Optimization-based Whitening

For optimization-based whitening, there can be several variants to fine-tune from different aspects, many of which are to make the algorithm generate ZCA-like filters.

1. *Padding*

As each image has its size, the boundary effect is inevitable, i.e. how to deal with the case where the window of a filter exceeds the boundary of an image. Here, we propose three kinds of

where ϵ is a very small positive quantity (we set 10^{-7} in this paper). The resulting covariance matrix will be approximately I.

padding to overcome this issue (Figure 2):

i) Zero-padding: pad all pixels outside the image with zeros, so that only original image pixels fall into window. But the drawback is that windows on the boundary do not generate numbers in the same magnitude as those generated by other windows.

ii) Mirror-padding: flip the image along four boundary axes, and then flip them again along their axes. Thus, windows centered near the periphery reuse pixels near the boundary.

iii) Natural-padding: when pixels outside the images are available (like sampling from larger images), it is most straightforward to employ these pixels for padding. But this is not applicable when only small image patches are accessible.

(a) (b) (c)

Figure 2. Illustration of (a) zero-padding, (b) mirror-padding and (c) natural padding. Yellow box is the local receptive field window. Red box marks the image being processed. White area in (a) represents 0.

2. *Convolution*

From observing ZCA-based whitening filters, we found out that the windows from different filters contain similar values, i.e. windows are approximately translation invariant. This inspires us to greatly speed up the whitening process: just parametrize values in one window and copy them to all other windows. This process is also known as convolution.

3. *Symmetry of whitening filters*

One of the advantages of ZCA-based whitening filter is that the resulting W_z is symmetric: $W_z = W_z^T = ED^{-\frac{1}{2}}E^T$, in contrast to the non-symmetric PCA-based filters: $W_p = D^{-\frac{1}{2}}E^T$. In application, ZCA outperforms PCA by generating whitening filters that lead to better image classification results. In this sense, it is more desirable to symmetrize our whitening filter W. Specifically, there are two ways of achieving this goal:

i) Hard-symmetry constraint: After obtaining $\nabla_W f$ for loss function (3), we project it into the symmetric-matrix space: $(*_W)f \triangleq (\nabla_W f + (\nabla_W f)^T)/2$.

ii) Soft-symmetry constraint: Add a regularizer to punish solutions that are "too non-symmetric":

$$f_{sym}(\theta) = \left\|W(\theta)AW(\theta)^T - I\right\|_F^2 + C\left\|W(\theta) - W(\theta)^T\right\|_F^2 \quad (4)$$

(in experiment, we set coefficient $C = 0.01$)

4. Experiments

4.1 Datasets

In the experiments, we employ two datasets: CIFAR-10 [2] and Natural Image Statistics [1] dataset. CIFAR dataset is a collection of tiny color images of size 32*32, obtained by searching various online image search engines for all the non-abstract English nouns and noun phrases in the

WordNet lexical database. CIFAR-10 is a ten-class labeled subset of CIFAR, containing 50,000 images for training and 10,000 images for testing. Natural Image Statistics dataset also contains 60,000 images of size 32*32, (50,000 for training and 10,000 for testing), sampled from 10 large images downloadable from the website for book Natural Image Statistics [1].

4.2 Experimental Results

4.2.1 CIFAR-10 Dataset

On CIFAR dataset, we conducted experiments for ZCA-based whitening as well as different versions of our optimization-based whitening. We report the value for loss function (3) on both training data and test data in Table 1. As seen, on training set, ZCA performs almost perfectly (see footnote 1 for details). On the test set, all variants of OWF perform better. Within variants of OWF, mirror-padding can greatly improve the performance in almost all cases; convolution hurts the performance a bit; both hard and soft symmetrizing the whole whitening matrix lead to worse result than non-symmetrizing.

Additionally, we concatenated our model with the Topographical Independent Component Analysis (TICA) [3] based on local receptive fields to generate features for SVM classification. The results are reported in Table 1. As seen, among all the whitening methods, ZCA-based filters generate the features leading to best classification results. However, it's important to note that Optimization-based Whitening Filters use much fewer parameters than ZCA filters. For example, with $7^2/32^2 \approx 4.8\%$ of number of parameters (or degrees of freedom) as ZCA, the basic OWF leads to accuracy 3.75% less than ZCA; with only $7^2/(32^2 * 32^2) * \approx 0.0047\%$ of number of parameters as ZCA, mirror-padding convolutional OWF leads to accuracy 4.18% less than ZCA, which substantiates our assumption of locality and convolution in whitening filters, and manifests a good tradeoff when there is insufficient time or memory for executing ZCA. Within different variants of OWF, symmetry constraint drags down the accuracy a bit, while convolution is most helpful when mirror-padding scheme is carried out.

Padding	Convolution	Method-Name	Train	Test	Classification Accuracy on Test Set
/	/	ZCA	1	3296	56.93%
Zero-Pad	Non-Conv	Zero-OWF	352	1767	53.18%
		Zero-Hard-Sym	526	1660	51.49%
		Zerp-Soft-Sym	564	1654	51.72%
	Conv	Zero-Conv	1234	1825	48.92%
Mirror-Pad	Non-Conv	Mir-OWF	355	1766	51.88%
		Mir-Hard-Sym	369	1760	51.29%
		Mir-Soft-Sym	609	1649	49.51%
	Conv	Mir-Conv	831	1672	52.75%

Table 1 The values of loss function (3) for both training and testing sets on CIFAR-10 dataset, together with the classification results generated by whitening-TICA-SVM process. Window size for all variants of Optimization-based Window Filter is 7.

4.2.1 Natural Image Statistics Dataset

In Natural Image Statistics dataset, we can apply natural padding by sampling patches with margins. Specifically, we sample 60,000 patches of size 48*48 and deem the center 32*32 pixels as within boundary (the margin is thus 8 on each side). As seen from Table 2, natural padding can to a huge extent improve the results, even outperforming ZCA-based whitening filters on test data. Furthermore, compared to the experiments on CIFAR, our methods can almost perfectly decorrelate pixels for natural images dataset. We attribute this to the fact that CIFAR is a hand-engineering dataset and the properties of locality and convolution are more obvious in whitening natural images.

Padding	Convolution	Method-Name	Train	Test
/	/	ZCA	1	234
Zero-Pad	Non-Conv	Zero-OWF	282	370
		Zero-Hard-Sym	101	227
		Zero-Soft-Sym	52	190
	Conv	Zero-Conv	337	405
Natural-Pad	Non-Conv	Nat-OWF	27	176
	Conv	Nat-Conv	95	213

Table 2 The values of loss function (3) for both training and testing sets on NIS dataset. Window size for all variants of Optimization-based Window Filter is 7. Note that symmetry cannot be applied in natural padding because the whitening matrix is not a square matrix (input image dimension is larger than output dimension).

5. Conclusion

In this paper, we present the optimization-based whitening (OWF), which can accomplish the task of whitening natural images with much fewer parameters than traditional methods such as Zero-phase Component Analysis (ZCA). OWF applies local receptor fields, which lend it lots of variants such as convolution and symmetry. These variants are to simulate the various properties of ZCA filters, while circumventing the scalability issue faced by traditional whitening methods.

Experiments show that with optimization-based whitening, the difference of the resulting empirical covariance matrix (on training data) and identity matrix can be acceptablly small, while on test data, OWF can sometimes achieve better whitening effects, casting doubt on the generalization ability of ZCA. However, OWF is still dwarfed by ZCA with respect to classification accuracy on benchmark dataset. We attribute this to the large reduction in parameter set and will test the model on more datasets to check the validity of our assumption.

References

[1] A. Hyvärinen, J. Hurri, and P. O. Hoyer. Natural Image Statistics — A probabilistic approach to early computational vision. Springer-Verlag, 2009.

[2] A. Krizhevsky. Learning multiple layers of features from tiny images. Technical report, U. Toronto, 2009.

[3] Q. V. Le, J. Ngiam, Z. Chen, D. Chia, P. W. Koh, A. Y. Ng. Tiled convolutional neural networks. In *NIPS*, 2010.