

Learning of Separable Filters by Stacked Fisher Convolutional Autoencoders

Arash Shahriari
arash.shahriari@anu.edu.au;csiro.au

Australian National University (ANU)
Commonwealth Scientific & Industrial
Research Organisation (CSIRO)
Canberra, Australia

Inspired by the overcomplete autoencoders, we introduce Fisher convolutional autoencoders to learn separable filters in a distributed network. These stacked autoencoders employ the linear discriminant analysis to impose the maximum distinction among texture classes whilst holds the minimum separation within each individual one. A network of stacked Fisher autoencoders learns banks of separable filters in parallel and makes an ensemble of deep features with higher separability for better classification. This adjusts the depth of stacks automatically with respect to the capability of each separable filter to extract high order convolutional features for the textures of dataset under study.

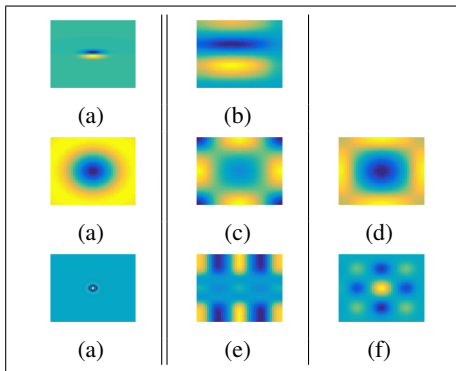


Figure 1: (a) initial filters; (b) UIUC; (c) KTH-TIPS2-a; (d) KTH-TIPS2-b; (e) FMD; (f) DTD.

The figure shows examples of initials and their corresponding learned separable filters from different texture banks. It can be seen that for some filters the changes across various datasets is smaller than the others because they are responsible to extract common features in the texture patterns. In contrast, the filters with high deformations usually connect to deeper stacks of autoencoders capturing high-order convolutional representations.

Dataset	DSIFT [1]	Proposed
UIUC	97.2 ± 0.8	90.1 ± 0.8
KTH-a	82.5 ± 5.3	85.5 ± 4.9
KTH-b	69.3 ± 0.9	70.1 ± 0.7
FMD	58.1 ± 1.7	71.8 ± 2.2
DTD	58.6 ± 2.0	59.1 ± 1.3

Table 1: Mean accuracy of texture recognition for dense SIFT and our descriptors.

We conduct our experiments on publicly available datasets varying in number of classes and quality of textures on a standard platform. The results prove supremacy of our descriptors over popular dense SIFT features for the purpose of texture understanding.

Dataset	SOA [1][2]	Ours
UIUC	99.3 ± 0.4	96.3 ± 0.1
KTH-a	84.7 ± 1.5	86.0 ± 2.3
KTH-b	81.1 ± 2.4	82.3 ± 0.9
FMD	82.4 ± 1.5	85.7 ± 3.0
DTD	74.7 ± 1.0	85.9 ± 1.4

Table 2: Mean accuracy of texture recognition for others (SOA) and our framework (Ours).

Our experiments also confirm that the Fisher convolutional autoencoders are successful on imposing distinction among highly-correlated texture patterns when joined the pre-trained deep local descriptors like DeCAF and VGG.

- [1] Ken Chatfield, Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Return of the devil in the details: Delving deep into convolutional nets. *arXiv preprint arXiv:1405.3531*, 2014.
- [2] Mircea Cimpoi, Subhransu Maji, Iasonas Kokkinos, and Andrea Vedaldi. Deep filter banks for texture recognition, description, and segmentation. *International Journal of Computer Vision*, pages 1–30, 2015.