Learning Invariant Features through Topographic Filter Maps

Koray Kavukcuoglu, Marc’Aurelio Ranzato, Rob Fergus, Yann LeCun

Several recently-proposed architectures for high-performance object recognition are composed of two main stages: a feature extraction stage that extracts locally-invariant feature vectors from regularly spaced image patches, and a somewhat generic supervised classifier. The first stage is often composed of three main modules: (1) a bank of filters (often oriented edge detectors); (2) a non-linear transform, such as a point-wise squashing functions, quantization, or normalization; (3) a spatial pooling operation which combines the outputs of similar filters over neighboring regions. We propose a method that automatically learns such feature extractors in an unsupervised fashion by simultaneously learning the filters and the pooling units that combine multiple filter outputs together. The method automatically generates topographic maps of similar filters that extract features of orientations, scales, and positions. These similar filters are pooled together, producing locally-invariant outputs. The learned feature descriptors give comparable results as SIFT on image recognition tasks for which SIFT is well suited, and better results than SIFT on tasks for which SIFT is less well suited.