*MBS 96-22*

Unsupervised Learning with Statistical Certification

Charles Chubb, Zhong-Lin Lu, George Sperling

---

We have developed an explicit independence rejection algorithm that can extract the characteristic structures in an arbitrary population of images. The algorithm is based on the face that applying an orthonormal linear transformation B to an image v a vector Bv whose pixels also are ISNs. The apply this result, we choose an arbitrary, orthonormal basis B spanning the space of all 16 X 16 pixel image patches. Then we iterate the following 5-step procedure many times: (1) Sample. Grab a random 16 X 16 pixel image patch v from the target image population. Attempt to reject the (false) null hypothesis H o that v is devoid of structure (i.e., that v's pixels are jointly independent, identically distributed, random variables as follows: (2) Gaussian replacement. Transform v into gv by replacing the ith greatest of v's pixel values by the ith greatest value in a fresh sample of 256 ISN's. If Ho is true, gv's pixels are not ISNs. (3) Rotate. Apply the basis B to gv. Under H o, the resulting vector Bgv must still consist of ISNs. (4): Test H o. Statistically determine whether the pixels of Bgv are ISNs. (5a) Sufficient structure detected. If the p-value of the test is sufficiently small, reject H o conclude that B is sensitive to the structure in v, and return to step (1); otherwise, (5b) Insufficient structure detected. Update B by rotating it to decrease the current p-value and return to step (1). Independence rejection generates a basis B that efficiently embodies "structure" in the target images and provides a powerful statistical test to evaluate the presence/absence of image structures. The algorithm's capabilities are demonstrated in three test populations: natural images, faulty random number generators, and artificial images composed of mixtures of basis functions. (1) The IR Algorithm succeeds in rejecting the null hypothesis that the UNIX random number generator and () is truly random. (2) When images are composed by adding arbitrary pairs of component images, the IR Algorithm extracts the components. (3) For a set of 153,280 natural image patches, we use an update rule designed to produce a sparse representation, rotating Bgv to increase its kurtosis. The resulting basis B detects structure with p-value $< .005$ in 92% of the patches. The basis elements resemble the receptive fields of Vi simple cells. (4) We show how the IR algorithm can be embodied in a biologically neural network. Together , (3) and (4) suggest that the receptive fields of sensory neurons may have evolved to achieve an efficient sparse representation of their environment by cooperatively maximizing their collective statistical power to reject the null hypothesis that their input is devoid of structure.