

We want to determine the weights W between the first and second layers in a CNN, where the first layer is stimulated by white noise, and the second layer consists of the retina's bipolar cells. How can we determine this? Experimentally, we know the white noise stimuli, and we can also measure the electrophysiological data of the second layer neurons. We believe that the W that maximizes $I(x, y)$ might be the one that biology actually uses.

The W that maximizes $I(x, y)$ leads us to deduce that in the Fourier space, y components are independent. This can be verified experimentally, and it has been found to be correct.

The biological significance of this is that organisms have adopted such W :

1. W maximizes mutual information, meaning it maximizes the amount of information transferred between the two layers.
2. W allows the second layer neurons to encode different frequencies.

The method of maximizing mutual information between two layers can also be used to train artificial neural networks, but clearly, it is not as fast as backpropagation.