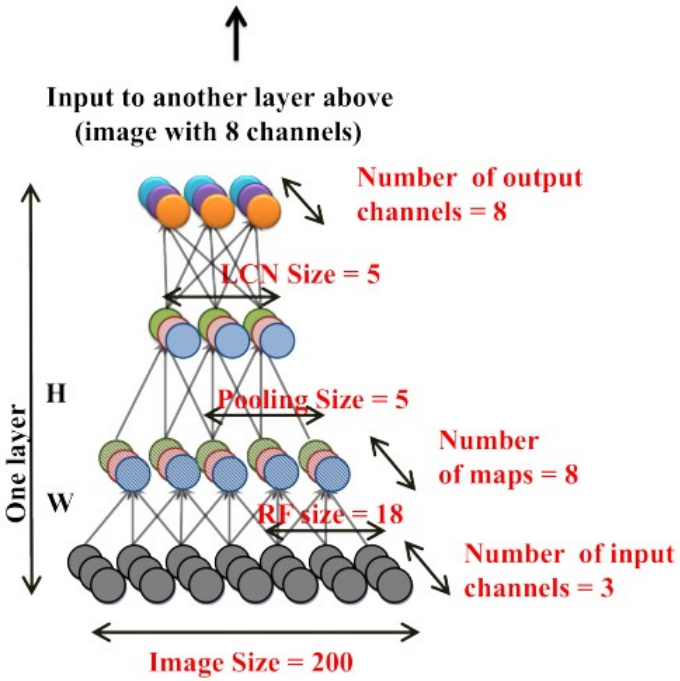
Building high-level features using large-scale unsupervised learning

1. GOAL

- Unlabeled images
- 10 million
- Wild dataset & ImageNet
- 37,000 example
- Accuracy 17.9%

2. Architecture



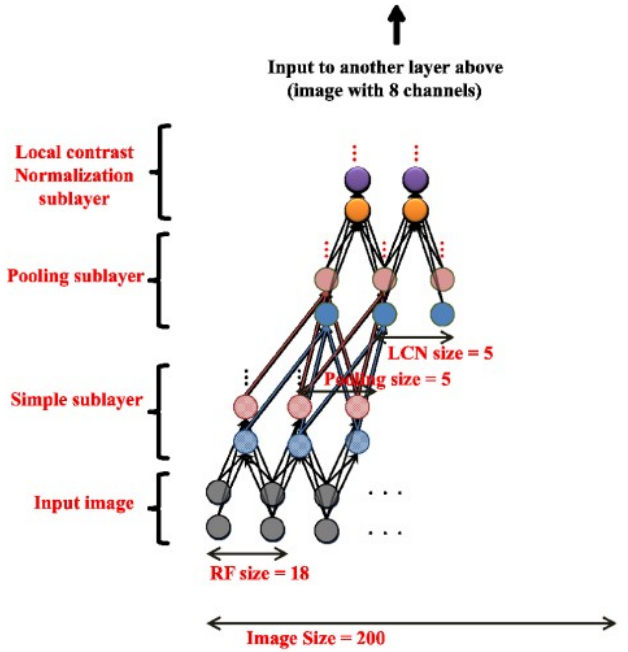


Figure 9. Diagram of the network we used with more detailed connectivity patterns. Color arrows mean that weights only connect to only one map. Dark arrows mean that weights connect to all maps. Pooling neurons only connect to one map whereas simple neurons and LCN neurons connect to all maps.

3. Filtering And Pooling

After two layer , part of the output is :

$$p_i(x^{(t)}; W, V) = \sqrt{\sum_{k=1}^{m} V_{ik} (\sum_{j=1}^{n} W_{kj} x_j^{(t)})^2}, \qquad (1)$$

Adding reconstitution cost :

$$\underset{W,\alpha}{\text{minimize}} \sum_i ||W^\top (\alpha W x^{(i)}) - x^{(i)}||_2^2 + \qquad (2)$$

$$\lambda \sum_j \sqrt{V_j (\alpha W x^{(i)})^2}$$

$$\text{subject to } ||W^{(k)}||_2 = 1, \forall k.$$

4. Local Contrast Normalization

Subtractive normalization:

$$g_{i,j,k} = h_{i,j,k} - \sum_{iuv} G_{uv} h_{i,j+u,i+v}$$

Divisive normalization（ c = 0.01 ）:

$$y_{i,j,k} = g_{i,j,k} / \max\{c, (\sum_{iuv} G_{uv} g^2_{i,j+u,i+v})^{0.5}\}$$

5. Optimization

Optimization problem:

$$\underset{W,\alpha}{\text{minimize}} \sum_i ||W^\top(\alpha W x^{(i)}) - x^{(i)}||_2^2 + \qquad (1)$$

$$\lambda \sum_j \sqrt{V_j (\alpha W x^{(i)})^2}$$

$$\text{subject to } ||W^{(k)}||_2 = 1, \forall k.$$

Deriving gradients:

$$\begin{aligned}\nabla_W F &= \nabla_W F + (\nabla_{W^T} F)^T \\ &= (W^T)(2(W^T W x - x))x^T + 2(Wx)(W^T W x - x)^T\end{aligned}$$