Spatial dimension:

We can give every point in our three-dimensional space an *(x, y, z)* coordinate. Since these points are represented as three real numbers, we say that it belongs to the set ℝ. These coordinates are usually not very meaningful, but they encode useful notions of distance and magnitude.

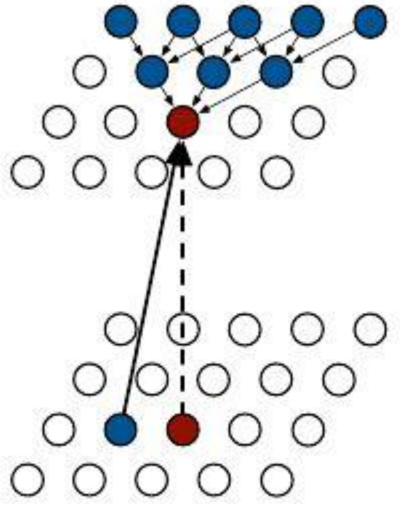
The important thing above is that when we had *three* real-valued numbers, we could represent it in our three-dimensional space. Let's now think about an image. An image is simply a big collection of pixels, with each pixel representing an intensity in some range. Here spatial Dimension is Row and Column and depth dimension is 3 (R, G, B).

LSTM:

LSTM recurrent unit tries to "remember" all the past knowledge that the network is seen so far and to "forget" irrelevant data.

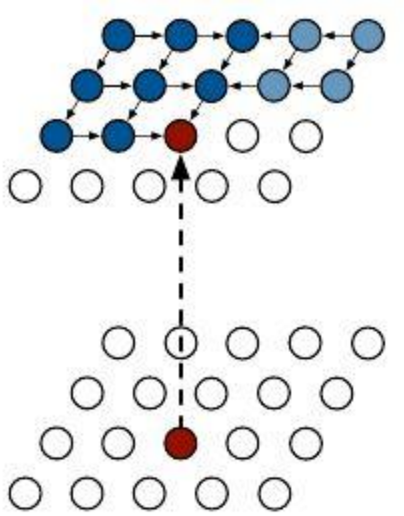
4 gates control how much information have to memorize.

Row LSTM: Hidden state (i, j) = Hidden state (i-1, j-1), Hidden state (i-1, j+1), Hidden state(i-1, j), p(i, j)



Row LSTM

Bi-Diagonal LSTM: In diagonal BLSTM, the hidden state of a pixel (i, j) depends on pixel (i, j-1) and on pixel (i-1, j). As bidirectional LSTM covers forward and backward dependencies, all the previously generated pixels are included in the 'context' / 'history' used in predicting value for a pixel.



The CIFAR-10 dataset

The CIFAR-10 dataset consists of 60000 32x32 color images in 10 classes, with 6000 images per class. There are 50000 training images and 10000 test images

Some image generation process:

1. VAE (Variational Auto Encoder): Generate image efficiently and quickly but blurry.

2. GAN (Generative Adversarial Networks): Generate more sharper image than VAE but difficult to optimize.

3. Auto Regressive Model (Ex. Pixel CNN, Pixel RNN): Simple and stable training process. Problem is it is inefficient during sampling.