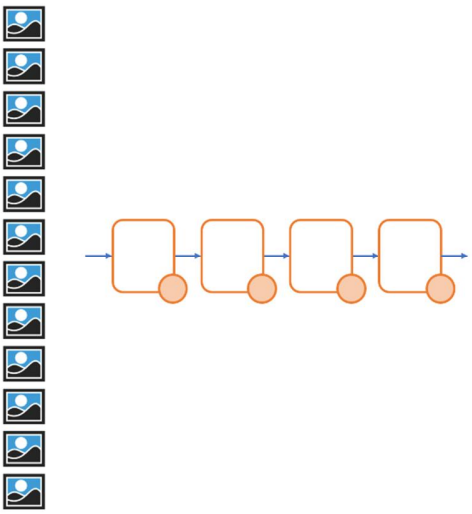
Training Options

A basic understanding of how training works can help you sift through the many training options available. The next few pages will break down some of the important ones:

- Mini-batches
- Learning rates
- Gradient clipping
- Algorithms

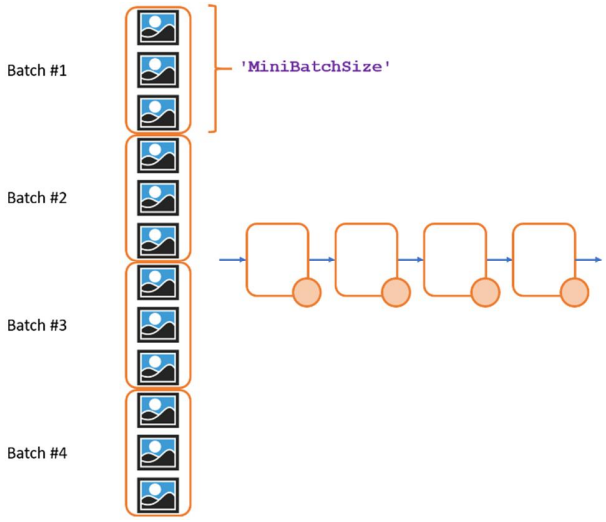
Mini-batches

1.



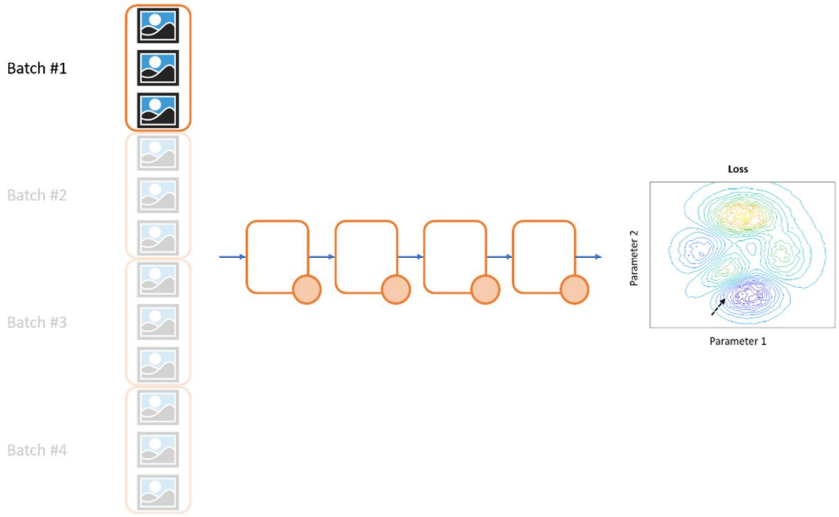
To get good results with deep networks, you generally need a lot of training data. Calculating the loss and the gradient by running all the training examples through the network would be expensive.

2.



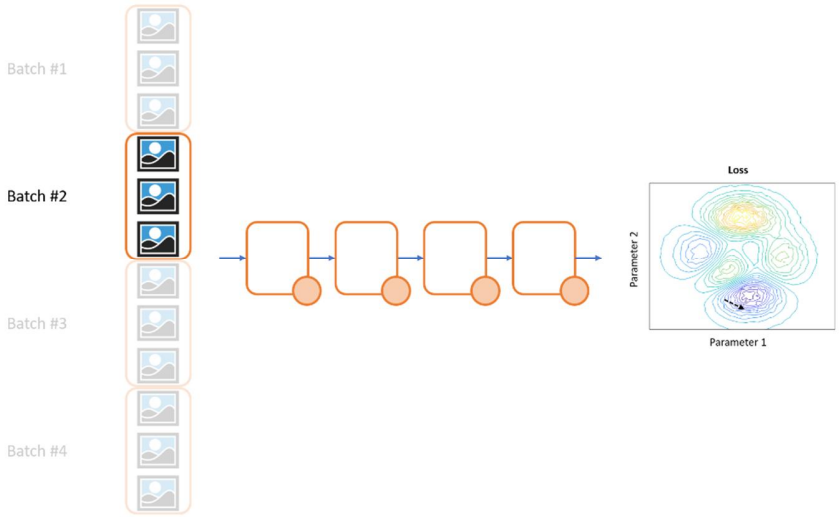
For this reason, the training set is split up into mini-batches.

3.



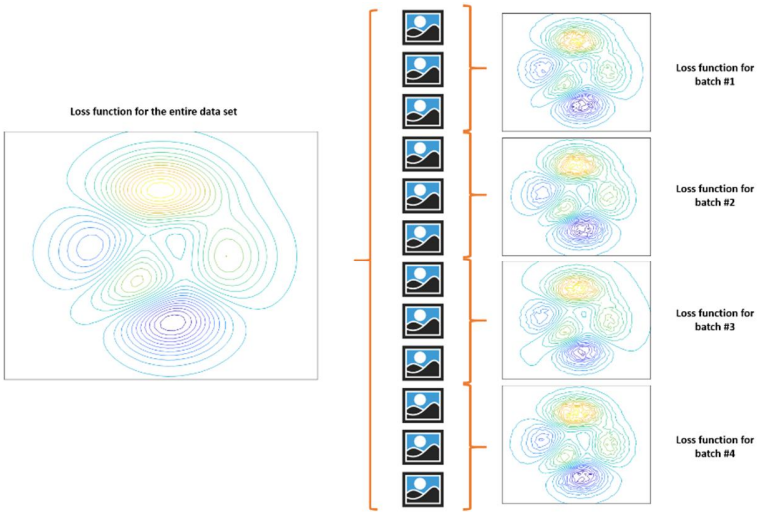
Each iteration of the training is performed with a different mini-batch.

4.



The loss and gradient calculated for each mini-batch is an approximation to the loss and gradient for the full training set.

5.



*This is known as **stochastic gradient descent**.*