

Common Training Options Reference

You've now seen how common parameters affect model training.

This reading serves as a reference with descriptions of the model options discussed in the video **and introduces a few additional training options**. Finally, it summarizes practical steps to address overfitting.

Training Option	Syntax	Description
Solver	"solverName"	The algorithm that determines how the model optimizes weights. Sometimes also called "optimizer." We recommend using the "adam" solver unless you have a specific reason for changing it.
Validation Data	"ValidationData"	A subset of the training data set aside to evaluate the model's performance during training. It does not contribute to weight values but instead serves as an external check of loss, helping to prevent overfitting.
Shuffle	"Shuffle"	How often the data is randomly split into mini-batches. We recommend "every-epoch" to avoid bias towards the initial mini-batches.
Training Plots	"Plots"	This option allows you to monitor training progress. We recommend using "training-progress" plots to diagnose and address training issues.
Mini-Batch Size	"MiniBatchSize"	The number of images in each subset of the training images, or mini-batch. Increasing the mini-batch size may help improve irregular or jagged training plots. Decreasing the mini-batch size will use less memory and help reduce overfitting, but introduces more randomness.
Validation Frequency	"ValidationFrequency"	The number of iterations between each calculation of validation loss. We recommend you set the validation frequency to calculate validation loss once or twice per epoch.
Neural Network to Return	"OutputNetwork"	Determines which model weights are used in the final, trained model. We recommend "best-validation-loss" to ensure that the weights that resulted in the lowest validation loss are returned, which helps prevent overfitting.
Initial Learning Rate	"InitialLearnRate"	The step size your model uses when optimizing weights. Step sizes that are too large risk failing to converge on effective weights. Step sizes that are too small can take a long time to train and lose some of the stochasticity necessary to find effective weights. We recommend you start with the default value and adjust based on what you observe in the training plot.
Maximum Epochs	"MaxEpochs"	The length of training, as determined by the total number of epochs. One epoch is the number of iterations to cycle through all mini-batches (i.e. once through the full training set). Increase the maximum epochs if the validation loss is still decreasing at the end of training.
L2 Regularization	"L2Regularization"	A mathematical way to reduce overfitting. The details of this algorithm are beyond this course's scope, but increasing the regularization value will help combat overfitting. Typical regularization values range from 0 to 0.1.
Validation Patience	"ValidationPatience"	Training will stop early if the validation loss does not improve for the number of subsequent calculations specified by the validation patience. This saves training resources; we recommend using this option if the training will take hours or days and you want a way to stop early.