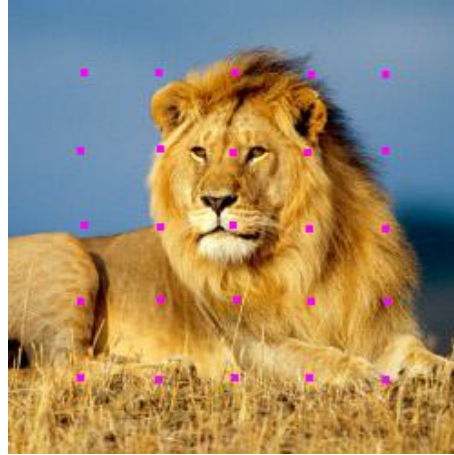# Solution 1

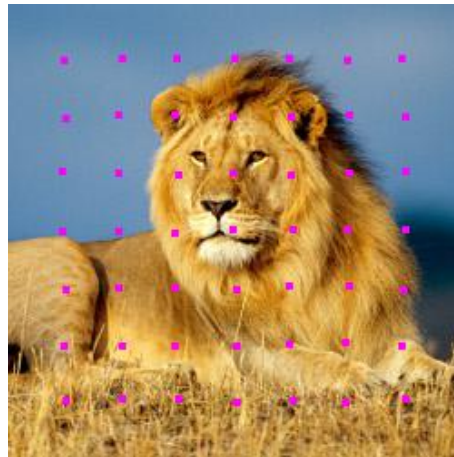**(a)** The results are shown as follow.

Figure 1: prob1a_25_centers

Figure 2: prob1a_49_centers

(b) The results are shown as follow. Value of **spatial_weight** is 0.3, which is chosen based on experiment result. Larger the value is, more blocks the results have. So, the value should be a small constant. On the contrary, if the value is too small, we can not find a integral cluster. Points in the same scatter in many places.
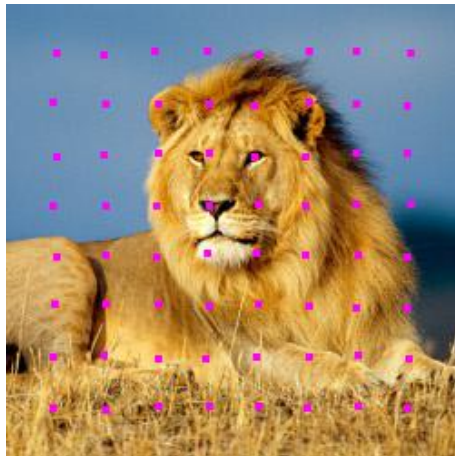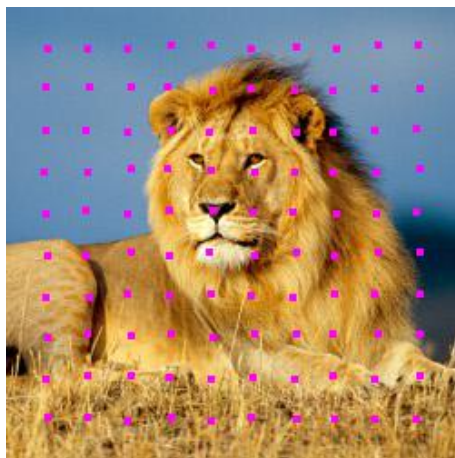
Figure 3: prob1a_64_centers

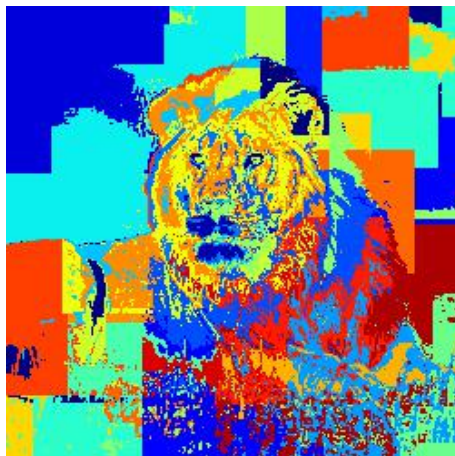Figure 4: prob1a_81_centers

Figure 5: prob1a_100_centers

Figure 6: prob1b_25

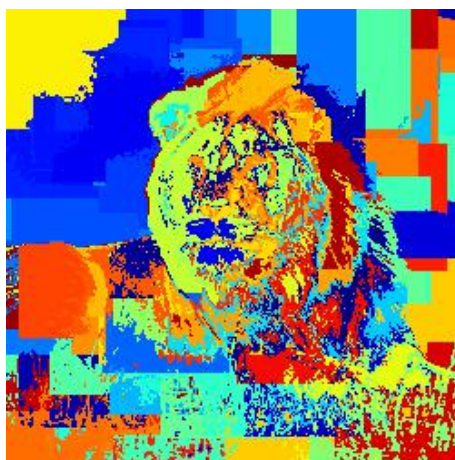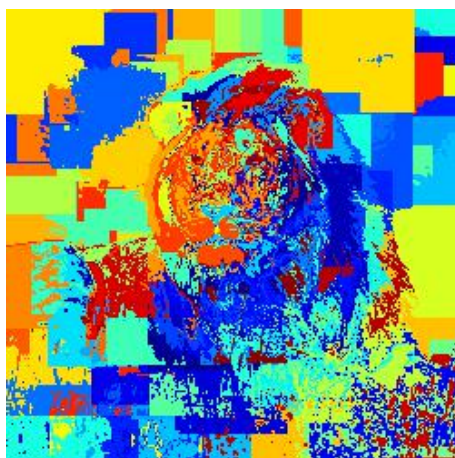

Figure 7: prob1b_49

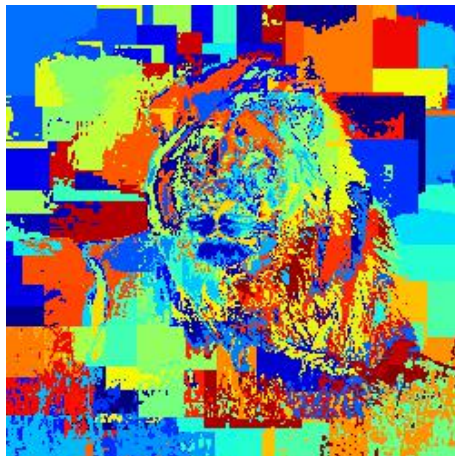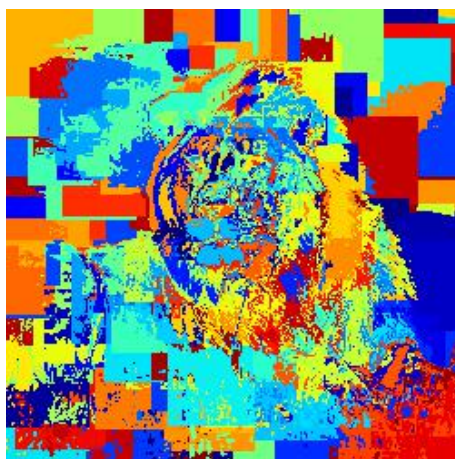

Figure 8: prob1b_64

Figure 9: prob1b_81

Figure 10: prob1b_100

Solution 2

(a)   I try different batch size, learning rate and number of hidden units and the results are shown as follow. The limits of the uniform distribution is for the stable of training process. For example, if the initial value is too large compared with the gradient, then training process will take too much time. And if the initial in different layer have huge difference, the changing of value of these parameters in training process are not in the same rate, which makes training process unstable.

(b)   I did some experiment with different learning rate and batch size under condition of using momentum or using SGD. The result is shown as follow. Basically, momentum help training process become quick and stable and have better performance in both loss value and accuracy when training or validating.
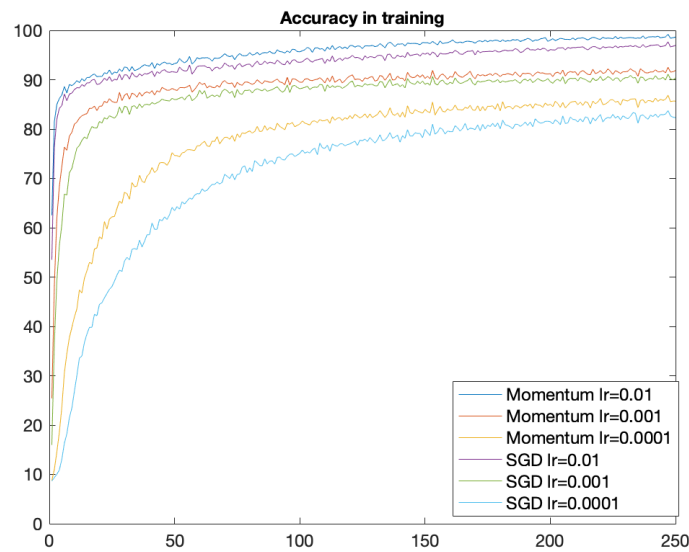

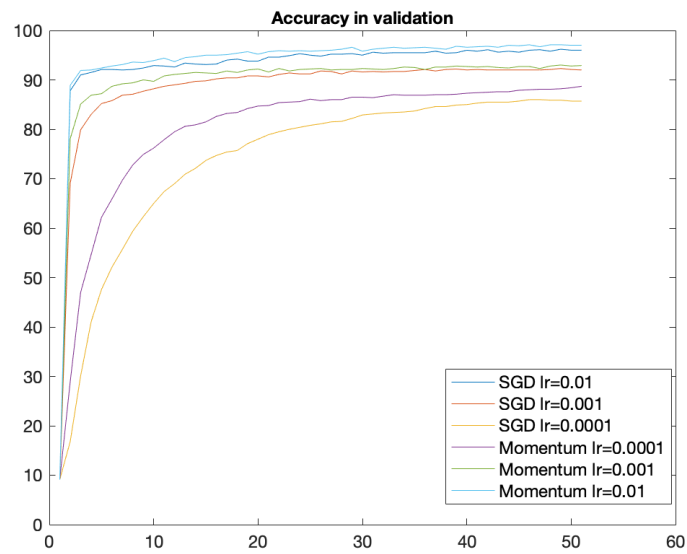Figure 11: Accuracy in Training with Different Learning Rate

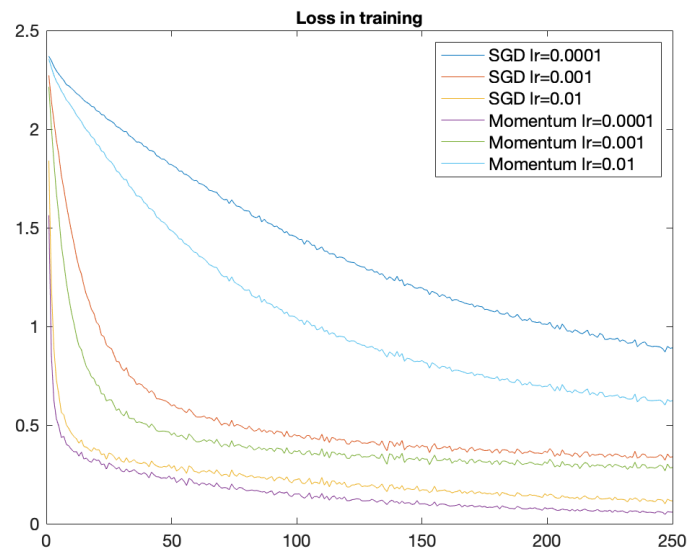Figure 12: Accuracy in Validation with Different Learning Rate

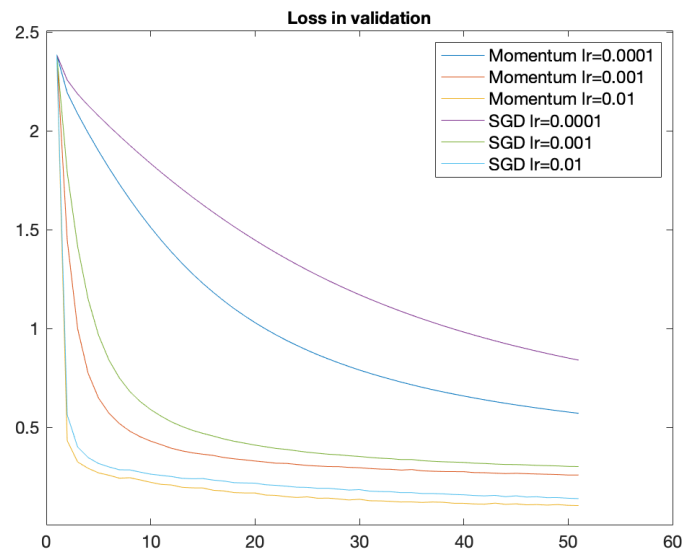Figure 13: Loss in Training with Different Learning Rate

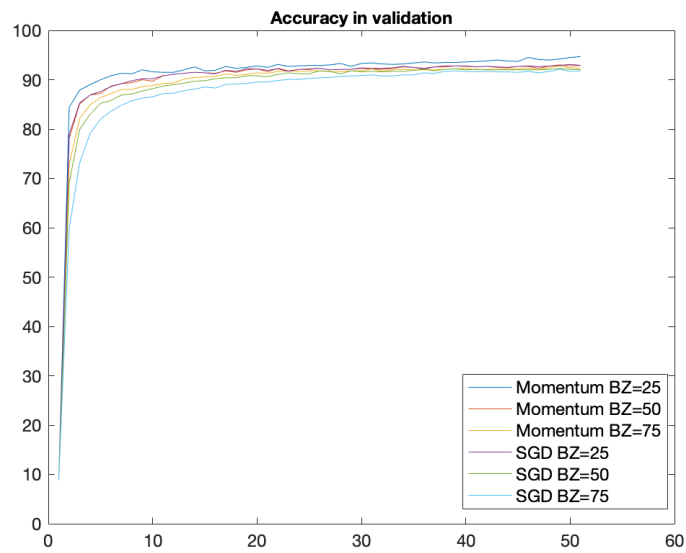Figure 14: Loss in Validation with Different Learning Rate

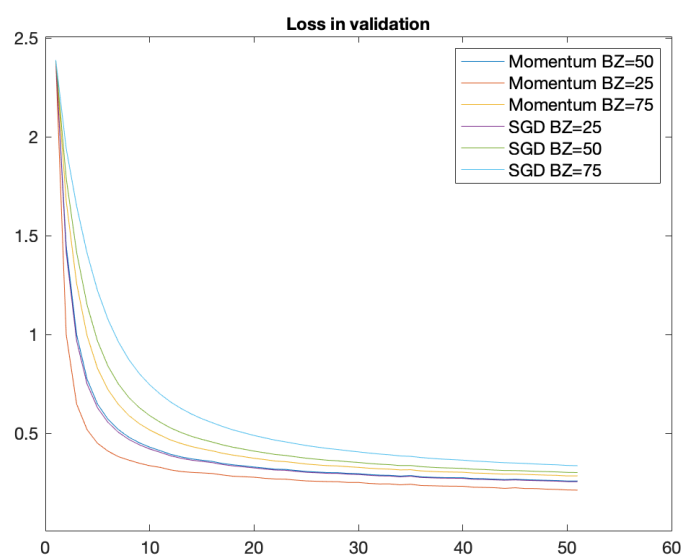Figure 15: Accuracy in Training with Different Batch Size

Figure 16: Accuracy in Validation with Different Batch Size

Solution 3

I did some experiment with momentum and default parameters (lr=0.001, batch size=50) under condition of using convolutional layer or not. The result is shown as follow.
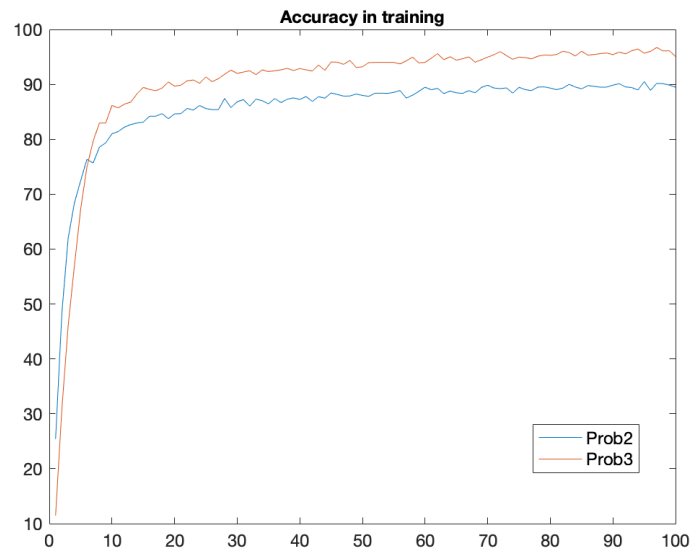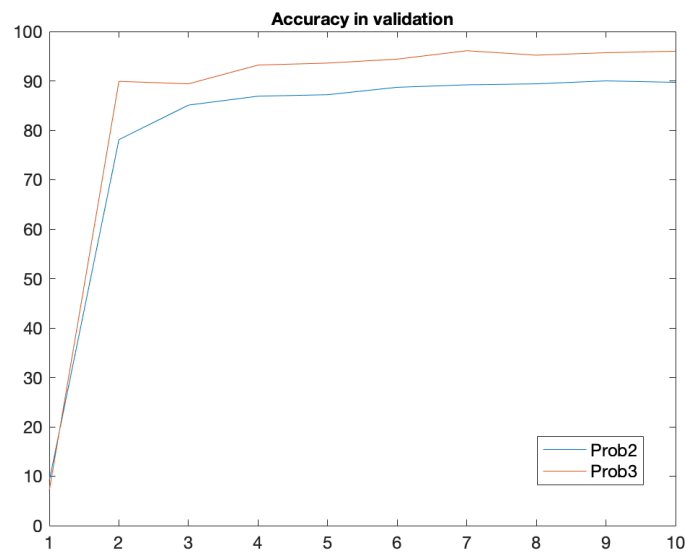

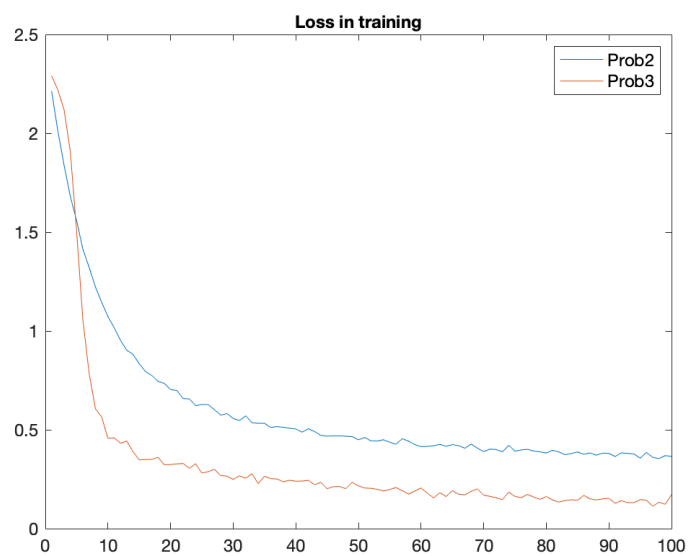Figure 17: Accuracy in Training

Figure 18: Accuracy in Validation
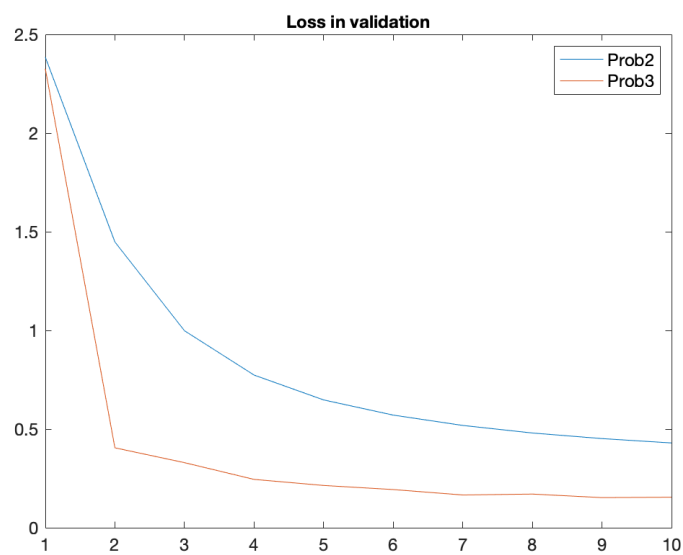
Figure 19: Loss in Training

Figure 20: Loss in Validation

Information

This problem set took approximately 18 hours of effort.