Convolutional neural network layers

In this notebook, we will build the convolutional neural network layers. This will be followed by a spatial batchnorm, and then in the final notebook of this assignment, we will train a CNN to further improve the validation accuracy on CIFAR-10.

CS231n has built a solid API for building these modular frameworks and training them, and we will use their very well implemented framework as opposed to "reinventing the wheel." This includes using their Solver, various utility functions, their layer structure, and their implementation of fast CNN layers. This also includes nndl.fc_net, nndl.layers, and nndl.layer_utils. As in prior assignments, we thank Serena Yeung & Justin Johnson for permission to use code written for the CS 231n class (cs231n.stanford.edu).

In [1]:
```
## Import and setups

import time
import numpy as np
import matplotlib.pyplot as plt
from nndl.conv_layers import *
from cs231n.data_utils import get_CIFAR10_data
from cs231n.gradient_check import eval_numerical_gradient, eval_numerica
l_gradient_array
from cs231n.solver import Solver

%matplotlib inline
plt.rcParams['figure.figsize'] = (10.0, 8.0) # set default size of plots
plt.rcParams['image.interpolation'] = 'nearest'
plt.rcParams['image.cmap'] = 'gray'

# for auto-reloading external modules
# see http://stackoverflow.com/questions/1907993/autoreload-of-modules-i
n-ipython
%load_ext autoreload
%autoreload 2

def rel_error(x, y):
  """ returns relative error """
  return np.max(np.abs(x - y) / (np.maximum(1e-8, np.abs(x) + np.abs(y
))))
```

Implementing CNN layers

Just as we implemented modular layers for fully connected networks, batch normalization, and dropout, we'll want to implement modular layers for convolutional neural networks. These layers are in `nndl/conv_layers.py`.

Convolutional forward pass

Begin by implementing a naive version of the forward pass of the CNN that uses `for` loops. This function is `conv_forward_naive` in `nndl/conv_layers.py`. Don't worry about efficiency of implementation. Later on, we provide a fast implementation of these layers. This version ought to test your understanding of convolution. In our implementation, there is a triple `for` loop.

After you implement `conv_forward_naive`, test your implementation by running the cell below.

In [2]:
```
x_shape = (2, 3, 4, 4)
w_shape = (3, 3, 4, 4)
x = np.linspace(-0.1, 0.5, num=np.prod(x_shape)).reshape(x_shape)
w = np.linspace(-0.2, 0.3, num=np.prod(w_shape)).reshape(w_shape)
b = np.linspace(-0.1, 0.2, num=3)

conv_param = {'stride': 2, 'pad': 1}
out, _ = conv_forward_naive(x, w, b, conv_param)
correct_out = np.array([[[[-0.08759809, -0.10987781],
                           [-0.18387192, -0.2109216 ]],
                          [[ 0.21027089,  0.21661097],
                           [ 0.22847626,  0.23004637]],
                          [[ 0.50813986,  0.54309974],
                           [ 0.64082444,  0.67101435]]],
                         [[[-0.98053589, -1.03143541],
                           [-1.19128892, -1.24695841]],
                          [[ 0.69108355,  0.66880383],
                           [ 0.59480972,  0.56776003]],
                          [[ 2.36270298,  2.36904306],
                           [ 2.38090835,  2.38247847]]]])

# Compare your output to ours; difference should be around 1e-8
print('Testing conv_forward_naive')
print('difference: ', rel_error(out, correct_out))
```

```
Testing conv_forward_naive
difference:  2.2121476417505994e-08
```

Convolutional backward pass

Now, implement a naive version of the backward pass of the CNN. The function is `conv_backward_naive` in `nndl/conv_layers.py`. Don't worry about efficiency of implementation. Later on, we provide a fast implementation of these layers. This version ought to test your understanding of convolution. In our implementation, there is a quadruple `for` loop.

After you implement `conv_backward_naive`, test your implementation by running the cell below.

In [3]:
```
x = np.random.randn(4, 3, 5, 5)
w = np.random.randn(2, 3, 3, 3)
b = np.random.randn(2,)
dout = np.random.randn(4, 2, 5, 5)
conv_param = {'stride': 1, 'pad': 1}

out, cache = conv_forward_naive(x,w,b,conv_param)

dx_num = eval_numerical_gradient_array(lambda x: conv_forward_naive(x, w
, b, conv_param)[0], x, dout)
dw_num = eval_numerical_gradient_array(lambda w: conv_forward_naive(x, w
, b, conv_param)[0], w, dout)
db_num = eval_numerical_gradient_array(lambda b: conv_forward_naive(x, w
, b, conv_param)[0], b, dout)

out, cache = conv_forward_naive(x, w, b, conv_param)
dx, dw, db = conv_backward_naive(dout, cache)

# Your errors should be around 1e-9'
print('Testing conv_backward_naive function')
print('dx error: ', rel_error(dx, dx_num))
print('dw error: ', rel_error(dw, dw_num))
print('db error: ', rel_error(db, db_num))
```

```
Testing conv_backward_naive function
dx error:  3.549178100098112e-10
dw error:  1.0251075899212212e-09
db error:  2.6840994798677012e-11
```

Max pool forward pass

In this section, we will implement the forward pass of the max pool. The function is `max_pool_forward_naive` in `nndl/conv_layers.py`. Do not worry about the efficiency of implementation.

After you implement `max_pool_forward_naive`, test your implementation by running the cell below.

In [4]:
```
x_shape = (2, 3, 4, 4)
x = np.linspace(-0.3, 0.4, num=np.prod(x_shape)).reshape(x_shape)
pool_param = {'pool_width': 2, 'pool_height': 2, 'stride': 2}

out, _ = max_pool_forward_naive(x, pool_param)

correct_out = np.array([[[[-0.26315789, -0.24842105],
                          [-0.20421053, -0.18947368]],
                         [[-0.14526316, -0.13052632],
                          [-0.08631579, -0.07157895]],
                         [[-0.02736842, -0.01263158],
                          [ 0.03157895,  0.04631579]]],
                        [[[ 0.09052632,  0.10526316],
                          [ 0.14947368,  0.16421053]],
                         [[ 0.20842105,  0.22315789],
                          [ 0.26736842,  0.28210526]],
                         [[ 0.32631579,  0.34105263],
                          [ 0.38526316,  0.4       ]]]])

# Compare your output with ours. Difference should be around 1e-8.
print('Testing max_pool_forward_naive function:')
print('difference: ', rel_error(out, correct_out))
```

```
Testing max_pool_forward_naive function:
difference:  4.1666665157267834e-08
```

Max pool backward pass

In this section, you will implement the backward pass of the max pool. The function is `max_pool_backward_naive` in `nndl/conv_layers.py`. Do not worry about the efficiency of implementation.

After you implement `max_pool_backward_naive`, test your implementation by running the cell below.

In [8]:
```
x = np.random.randn(3, 2, 8, 8)
dout = np.random.randn(3, 2, 4, 4)
pool_param = {'pool_height': 2, 'pool_width': 2, 'stride': 2}

dx_num = eval_numerical_gradient_array(lambda x: max_pool_forward_naive(
x, pool_param)[0], x, dout)

out, cache = max_pool_forward_naive(x, pool_param)
dx = max_pool_backward_naive(dout, cache)

# Your error should be around 1e-12
print('Testing max_pool_backward_naive function:')
print('dx error: ', rel_error(dx, dx_num))
```

```
Testing max_pool_backward_naive function:
dx error:  3.275614875475113e-12
```

Fast implementation of the CNN layers

Implementing fast versions of the CNN layers can be difficult. We will provide you with the fast layers implemented by cs231n. They are provided in `cs231n/fast_layers.py`.

The fast convolution implementation depends on a Cython extension; to compile it you need to run the following from the `cs231n` directory:

```
python setup.py build_ext --inplace
```

NOTE: The fast implementation for pooling will only perform optimally if the pooling regions are non-overlapping and tile the input. If these conditions are not met then the fast pooling implementation will not be much faster than the naive implementation.

You can compare the performance of the naive and fast versions of these layers by running the cell below.

You should see pretty drastic speedups in the implementation of these layers. On our machine, the forward pass speeds up by 17x and the backward pass speeds up by 840x. Of course, these numbers will vary from machine to machine, as well as on your precise implementation of the naive layers.

In [6]:
```
from cs231n.fast_layers import conv_forward_fast, conv_backward_fast
from time import time

x = np.random.randn(100, 3, 31, 31)
w = np.random.randn(25, 3, 3, 3)
b = np.random.randn(25,)
dout = np.random.randn(100, 25, 16, 16)
conv_param = {'stride': 2, 'pad': 1}

t0 = time()
out_naive, cache_naive = conv_forward_naive(x, w, b, conv_param)
t1 = time()
out_fast, cache_fast = conv_forward_fast(x, w, b, conv_param)
t2 = time()

print('Testing conv_forward_fast:')
print('Naive: %fs' % (t1 - t0))
print('Fast: %fs' % (t2 - t1))
print('Speedup: %fx' % ((t1 - t0) / (t2 - t1)))
print('Difference: ', rel_error(out_naive, out_fast))

t0 = time()
dx_naive, dw_naive, db_naive = conv_backward_naive(dout, cache_naive)
t1 = time()
dx_fast, dw_fast, db_fast = conv_backward_fast(dout, cache_fast)
t2 = time()

print('\nTesting conv_backward_fast:')
print('Naive: %fs' % (t1 - t0))
print('Fast: %fs' % (t2 - t1))
print('Speedup: %fx' % ((t1 - t0) / (t2 - t1)))
print('dx difference: ', rel_error(dx_naive, dx_fast))
print('dw difference: ', rel_error(dw_naive, dw_fast))
print('db difference: ', rel_error(db_naive, db_fast))
```

```
Testing conv_forward_fast:
Naive: 5.189369s
Fast: 0.017614s
Speedup: 294.618046x
Difference:  2.22569363396174e-11

Testing conv_backward_fast:
Naive: 8.141040s
Fast: 0.009410s
Speedup: 865.156507x
dx difference:  5.039667080792084e-12
dw difference:  9.669798837675978e-13
db difference:  9.278045008676634e-14
```

In [7]:
```
from cs231n.fast_layers import max_pool_forward_fast, max_pool_backward_
fast

x = np.random.randn(100, 3, 32, 32)
dout = np.random.randn(100, 3, 16, 16)
pool_param = {'pool_height': 2, 'pool_width': 2, 'stride': 2}

t0 = time()
out_naive, cache_naive = max_pool_forward_naive(x, pool_param)
t1 = time()
out_fast, cache_fast = max_pool_forward_fast(x, pool_param)
t2 = time()

print('Testing pool_forward_fast:')
print('Naive: %fs' % (t1 - t0))
print('fast: %fs' % (t2 - t1))
print('speedup: %fx' % ((t1 - t0) / (t2 - t1)))
print('difference: ', rel_error(out_naive, out_fast))

t0 = time()
dx_naive = max_pool_backward_naive(dout, cache_naive)
t1 = time()
dx_fast = max_pool_backward_fast(dout, cache_fast)
t2 = time()

print('\nTesting pool_backward_fast:')
print('Naive: %fs' % (t1 - t0))
print('speedup: %fx' % ((t1 - t0) / (t2 - t1)))
print('dx difference: ', rel_error(dx_naive, dx_fast))
```

```
Testing pool_forward_fast:
Naive: 0.352915s
fast: 0.003795s
speedup: 92.996922x
difference:  0.0

Testing pool_backward_fast:
Naive: 1.359755s
speedup: 115.850942x
dx difference:  0.0
```

Implementation of cascaded layers

We've provided the following functions in `nndl/conv_layer_utils.py`:

```
- conv_relu_forward
- conv_relu_backward
- conv_relu_pool_forward
- conv_relu_pool_backward
```

These use the fast implementations of the conv net layers. You can test them below:

In [11]:
```
from nndl.conv_layer_utils import conv_relu_pool_forward, conv_relu_pool
_backward

x = np.random.randn(2, 3, 16, 16)
w = np.random.randn(3, 3, 3, 3)
b = np.random.randn(3,)
dout = np.random.randn(2, 3, 8, 8)
conv_param = {'stride': 1, 'pad': 1}
pool_param = {'pool_height': 2, 'pool_width': 2, 'stride': 2}

out, cache = conv_relu_pool_forward(x, w, b, conv_param, pool_param)
dx, dw, db = conv_relu_pool_backward(dout, cache)

dx_num = eval_numerical_gradient_array(lambda x: conv_relu_pool_forward(
x, w, b, conv_param, pool_param)[0], x, dout)
dw_num = eval_numerical_gradient_array(lambda w: conv_relu_pool_forward(
x, w, b, conv_param, pool_param)[0], w, dout)
db_num = eval_numerical_gradient_array(lambda b: conv_relu_pool_forward(
x, w, b, conv_param, pool_param)[0], b, dout)

print('Testing conv_relu_pool')
print('dx error: ', rel_error(dx_num, dx))
print('dw error: ', rel_error(dw_num, dw))
print('db error: ', rel_error(db_num, db))
```

```
Testing conv_relu_pool
dx error:  3.359808302469702e-07
dw error:  1.365292727246627e-09
db error:  6.638788554262562e-08
```

In [12]:
```
from nndl.conv_layer_utils import conv_relu_forward, conv_relu_backward

x = np.random.randn(2, 3, 8, 8)
w = np.random.randn(3, 3, 3, 3)
b = np.random.randn(3,)
dout = np.random.randn(2, 3, 8, 8)
conv_param = {'stride': 1, 'pad': 1}

out, cache = conv_relu_forward(x, w, b, conv_param)
dx, dw, db = conv_relu_backward(dout, cache)

dx_num = eval_numerical_gradient_array(lambda x: conv_relu_forward(x, w, b, conv_param)[0], x, dout)
dw_num = eval_numerical_gradient_array(lambda w: conv_relu_forward(x, w, b, conv_param)[0], w, dout)
db_num = eval_numerical_gradient_array(lambda b: conv_relu_forward(x, w, b, conv_param)[0], b, dout)

print('Testing conv_relu:')
print('dx error: ', rel_error(dx_num, dx))
print('dw error: ', rel_error(dw_num, dw))
print('db error: ', rel_error(db_num, db))
```

```
Testing conv_relu:
dx error:  1.2226166840761237e-09
dw error:  1.214310971991104e-09
db error:  1.798610564672917e-11
```

What next?

We saw how helpful batch normalization was for training FC nets. In the next notebook, we'll implement a batch normalization for convolutional neural networks, and then finish off by implementing a CNN to improve our validation accuracy on CIFAR-10.

Spatial batch normalization

In fully connected networks, we performed batch normalization on the activations. To do something equivalent on CNNs, we modify batch normalization slightly.

Normally batch-normalization accepts inputs of shape `(N, D)` and produces outputs of shape `(N, D)`, where we normalize across the minibatch dimension `N`. For data coming from convolutional layers, batch normalization accepts inputs of shape `(N, C, H, W)` and produces outputs of shape `(N, C, H, W)` where the `N` dimension gives the minibatch size and the `(H, W)` dimensions give the spatial size of the feature map.

How do we calculate the spatial averages? First, notice that for the `C` feature maps we have (i.e., the layer has `C` filters) that each of these ought to have its own batch norm statistics, since each feature map may be picking out very different features in the images. However, within a feature map, we may assume that across all inputs and across all locations in the feature map, there ought to be relatively similar first and second order statistics. Hence, one way to think of spatial batch-normalization is to reshape the `(N, C, H, W)` array as an `(N*H*W, C)` array and perform batch normalization on this array.

Since spatial batch norm and batch normalization are similar, it'd be good to at this point also copy and paste our prior implemented layers from HW #4. Please copy and paste your prior implemented code from HW #4 to start this assignment. If you did not correctly implement the layers in HW #4, you may collaborate with a classmate to use their implementations from HW #4. You may also visit TA or Prof OH to correct your implementation.

You'll want to copy and paste from HW #4:

```
- layers.py for your FC network layers, as well as batchnorm and dropout.
- layer_utils.py for your combined FC network layers.
- optim.py for your optimizers.
```

Be sure to place these in the `nndl/` directory so they're imported correctly. Note, as announced in class, we will not be releasing our solutions.

If you use your prior implementations of the batchnorm, then your spatial batchnorm implementation may be very short. Our implementations of the forward and backward pass are each 6 lines of code.

CS231n has built a solid API for building these modular frameworks and training them, and we will use their very well implemented framework as opposed to "reinventing the wheel." This includes using their Solver, various utility functions, their layer structure, and their implementation of fast CNN layers. This also includes nndl.fc_net, nndl.layers, and nndl.layer_utils. As in prior assignments, we thank Serena Yeung & Justin Johnson for permission to use code written for the CS 231n class (cs231n.stanford.edu).

In [3]:
```
## Import and setups

import time
import numpy as np
import matplotlib.pyplot as plt
from nndl.conv_layers import *
from cs231n.data_utils import get_CIFAR10_data
from cs231n.gradient_check import eval_numerical_gradient, eval_numerica
l_gradient_array
from cs231n.solver import Solver

%matplotlib inline
plt.rcParams['figure.figsize'] = (10.0, 8.0) # set default size of plots
plt.rcParams['image.interpolation'] = 'nearest'
plt.rcParams['image.cmap'] = 'gray'

# for auto-reloading external modules
# see http://stackoverflow.com/questions/1907993/autoreload-of-modules-i
n-ipython
%load_ext autoreload
%autoreload 2

def rel_error(x, y):
  """ returns relative error """
  return np.max(np.abs(x - y) / (np.maximum(1e-8, np.abs(x) + np.abs(y
))))
```

```
The autoreload extension is already loaded. To reload it, use:
  %reload_ext autoreload
```

Spatial batch normalization forward pass

Implement the forward pass, `spatial_batchnorm_forward` in `nndl/conv_layers.py`. Test your implementation by running the cell below.

In [10]:
```
# Check the training-time forward pass by checking means and variances
# of features both before and after spatial batch normalization

N, C, H, W = 2, 3, 4, 5
x = 4 * np.random.randn(N, C, H, W) + 10

print('Before spatial batch normalization:')
print('  Shape: ', x.shape)
print('  Means: ', x.mean(axis=(0, 2, 3)))
print('  Stds: ', x.std(axis=(0, 2, 3)))

# Means should be close to zero and stds close to one
gamma, beta = np.ones(C), np.zeros(C)
bn_param = {'mode': 'train'}
out, _ = spatial_batchnorm_forward(x, gamma, beta, bn_param)
print('After spatial batch normalization:')
print('  Shape: ', out.shape)
print('  Means: ', out.mean(axis=(0, 2, 3)))
print('  Stds: ', out.std(axis=(0, 2, 3)))

# Means should be close to beta and stds close to gamma
gamma, beta = np.asarray([3, 4, 5]), np.asarray([6, 7, 8])
out, _ = spatial_batchnorm_forward(x, gamma, beta, bn_param)
print('After spatial batch normalization (nontrivial gamma, beta):')
print('  Shape: ', out.shape)
print('  Means: ', out.mean(axis=(0, 2, 3)))
print('  Stds: ', out.std(axis=(0, 2, 3)))
```

```
Before spatial batch normalization:
  Shape:  (2, 3, 4, 5)
  Means:  [ 9.79075644  9.28707227 10.38253694]
  Stds:  [3.41122223 4.21319692 4.87804522]
After spatial batch normalization:
  Shape:  (2, 3, 4, 5)
  Means:  [6.63358257e-16 1.09634524e-16 1.97758476e-16]
  Stds:  [0.99999957 0.99999972 0.99999979]
After spatial batch normalization (nontrivial gamma, beta):
  Shape:  (2, 3, 4, 5)
  Means:  [6. 7. 8.]
  Stds:  [2.99999871 3.99999887 4.99999895]
```

Spatial batch normalization backward pass

Implement the backward pass, `spatial_batchnorm_backward` in `nndl/conv_layers.py`. Test your implementation by running the cell below.

In [11]:
```
N, C, H, W = 2, 3, 4, 5
x = 5 * np.random.randn(N, C, H, W) + 12
gamma = np.random.randn(C)
beta = np.random.randn(C)
dout = np.random.randn(N, C, H, W)

bn_param = {'mode': 'train'}
fx = lambda x: spatial_batchnorm_forward(x, gamma, beta, bn_param)[0]
fg = lambda a: spatial_batchnorm_forward(x, gamma, beta, bn_param)[0]
fb = lambda b: spatial_batchnorm_forward(x, gamma, beta, bn_param)[0]

dx_num = eval_numerical_gradient_array(fx, x, dout)
da_num = eval_numerical_gradient_array(fg, gamma, dout)
db_num = eval_numerical_gradient_array(fb, beta, dout)

_, cache = spatial_batchnorm_forward(x, gamma, beta, bn_param)
dx, dgamma, dbeta = spatial_batchnorm_backward(dout, cache)
print('dx error: ', rel_error(dx_num, dx))
print('dgamma error: ', rel_error(da_num, dgamma))
print('dbeta error: ', rel_error(db_num, dbeta))
```

```
dx error:  1.4107596948671627e-08
dgamma error:  5.0961529429881404e-11
dbeta error:  9.645459815074127e-12
```

Convolutional neural networks

In this notebook, we'll put together our convolutional layers to implement a 3-layer CNN. Then, we'll ask you to implement a CNN that can achieve > 65% validation error on CIFAR-10.

CS231n has built a solid API for building these modular frameworks and training them, and we will use their very well implemented framework as opposed to "reinventing the wheel." This includes using their Solver, various utility functions, their layer structure, and their implementation of fast CNN layers. This also includes nndl.fc_net, nndl.layers, and nndl.layer_utils. As in prior assignments, we thank Serena Yeung & Justin Johnson for permission to use code written for the CS 231n class (cs231n.stanford.edu).

If you have not completed the Spatial BatchNorm Notebook, please see the following description from that notebook:

Please copy and paste your prior implemented code from HW #4 to start this assignment. If you did not correctly implement the layers in HW #4, you may collaborate with a classmate to use their layer implementations from HW #4. You may also visit TA or Prof OH to correct your implementation.

You'll want to copy and paste from HW #4:

```
- layers.py for your FC network layers, as well as batchnorm and dropout.
- layer_utils.py for your combined FC network layers.
- optim.py for your optimizers.
```

Be sure to place these in the `nndl/` directory so they're imported correctly. Note, as announced in class, we will not be releasing our solutions.

In [14]:
```
# As usual, a bit of setup

import numpy as np
import matplotlib.pyplot as plt
from nndl.cnn import *
from cs231n.data_utils import get_CIFAR10_data
from cs231n.gradient_check import eval_numerical_gradient_array, eval_nu
merical_gradient
from nndl.layers import *
from nndl.conv_layers import *
from cs231n.fast_layers import *
from cs231n.solver import Solver

%matplotlib inline
plt.rcParams['figure.figsize'] = (10.0, 8.0) # set default size of plots
plt.rcParams['image.interpolation'] = 'nearest'
plt.rcParams['image.cmap'] = 'gray'

# for auto-reloading external modules
# see http://stackoverflow.com/questions/1907993/autoreload-of-modules-i
n-ipython
%load_ext autoreload
%autoreload 2

def rel_error(x, y):
  """ returns relative error """
  return np.max(np.abs(x - y) / (np.maximum(1e-8, np.abs(x) + np.abs(y
))))
```

```
The autoreload extension is already loaded. To reload it, use:
  %reload_ext autoreload
```

In [15]:
```
# Load the (preprocessed) CIFAR10 data.

data = get_CIFAR10_data()
for k in data.keys():
  print('{}: {} '.format(k, data[k].shape))
```

```
X_train: (49000, 3, 32, 32)
y_train: (49000,)
X_val: (1000, 3, 32, 32)
y_val: (1000,)
X_test: (1000, 3, 32, 32)
y_test: (1000,)
```

Three layer CNN

In this notebook, you will implement a three layer CNN. The `ThreeLayerConvNet` class is in `nndl/cnn.py`. You'll need to modify that code for this section, including the initialization, as well as the calculation of the loss and gradients. You should be able to use the building blocks you have either earlier coded or that we have provided. Be sure to use the fast layers.

The architecture of this CNN will be:

conv - relu - 2x2 max pool - affine - relu - affine - softmax

We won't use batchnorm yet. You've also done enough of these to know how to debug; use the cells below.

Note: As we are implementing several layers CNN networks. The gradient error can be expected for the `eval_numerical_gradient()` function. If your `W1 max relative error` and `W2 max relative error` are around or below 0.01, they should be acceptable. Other errors should be less than 1e-5.

In [40]:
```
num_inputs = 2
input_dim = (3, 16, 16)
reg = 0.0
num_classes = 10
X = np.random.randn(num_inputs, *input_dim)
y = np.random.randint(num_classes, size=num_inputs)

model = ThreeLayerConvNet(num_filters=3, filter_size=3,
                          input_dim=input_dim, hidden_dim=7,
                          dtype=np.float64)
loss, grads = model.loss(X, y)
for param_name in sorted(grads):
    f = lambda _: model.loss(X, y)[0]
    param_grad_num = eval_numerical_gradient(f, model.params[param_name
], verbose=False, h=1e-6)
    e = rel_error(param_grad_num, grads[param_name])
    print('{} max relative error: {}'.format(param_name, rel_error(param
_grad_num, grads[param_name])))
```

```
W1 max relative error: 0.00030248109387705127
W2 max relative error: 0.0027583677513153887
W3 max relative error: 5.8482760346196364e-05
b1 max relative error: 1.5460761129401907e-05
b2 max relative error: 3.9361785518893287e-07
b3 max relative error: 1.0454244302204221e-09
```

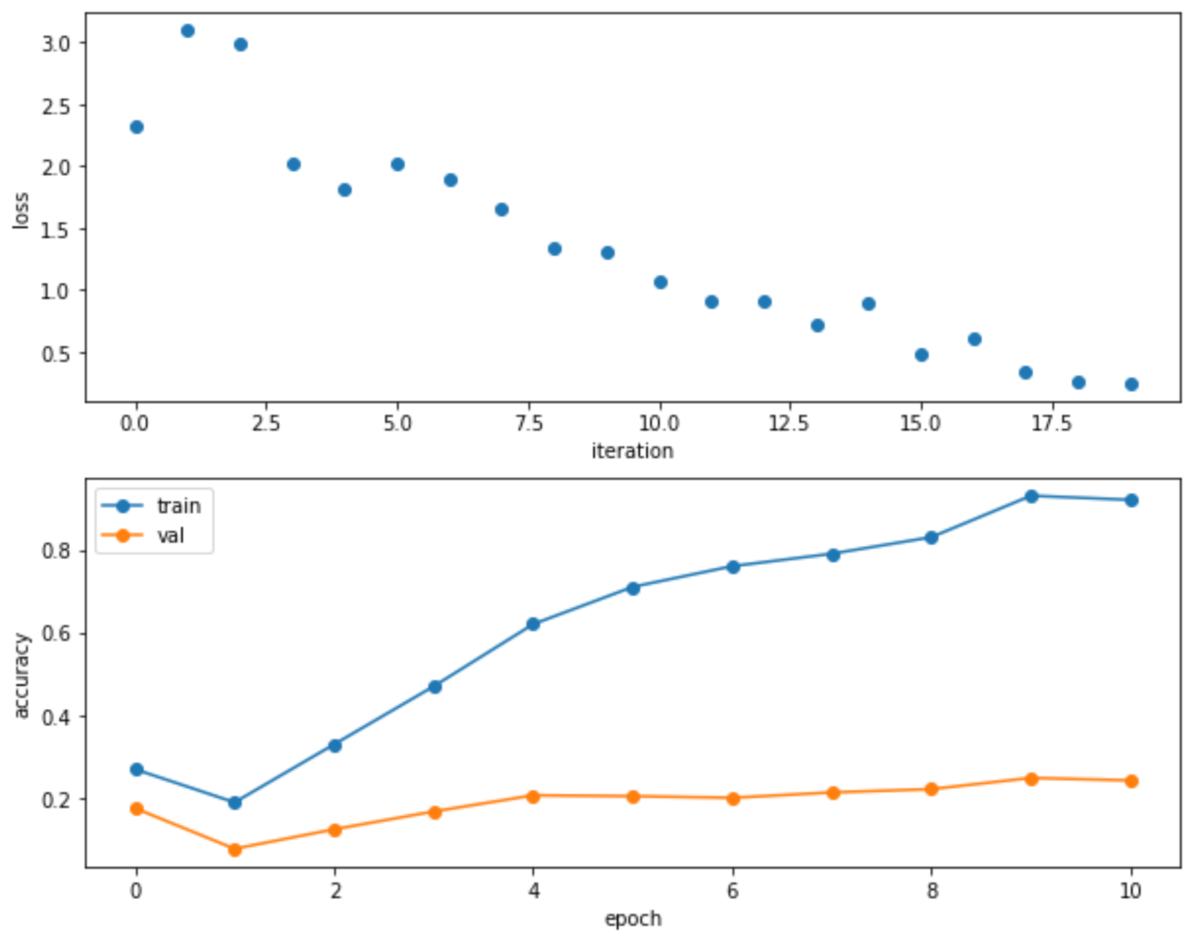
Overfit small dataset

To check your CNN implementation, let's overfit a small dataset.

In [33]:
```
num_train = 100
small_data = {
  'X_train': data['X_train'][:num_train],
  'y_train': data['y_train'][:num_train],
  'X_val': data['X_val'],
  'y_val': data['y_val'],
}

model = ThreeLayerConvNet(weight_scale=1e-2)

solver = Solver(model, small_data,
                num_epochs=10, batch_size=50,
                update_rule='adam',
                optim_config={
                  'learning_rate': 1e-3,
                },
                verbose=True, print_every=1)
solver.train()
```

```
(Iteration 1 / 20) loss: 2.315616
(Epoch 0 / 10) train acc: 0.270000; val_acc: 0.176000
(Iteration 2 / 20) loss: 3.098809
(Epoch 1 / 10) train acc: 0.190000; val_acc: 0.078000
(Iteration 3 / 20) loss: 2.981527
(Iteration 4 / 20) loss: 2.020423
(Epoch 2 / 10) train acc: 0.330000; val_acc: 0.125000
(Iteration 5 / 20) loss: 1.809810
(Iteration 6 / 20) loss: 2.019159
(Epoch 3 / 10) train acc: 0.470000; val_acc: 0.168000
(Iteration 7 / 20) loss: 1.898138
(Iteration 8 / 20) loss: 1.655982
(Epoch 4 / 10) train acc: 0.620000; val_acc: 0.207000
(Iteration 9 / 20) loss: 1.347849
(Iteration 10 / 20) loss: 1.307547
(Epoch 5 / 10) train acc: 0.710000; val_acc: 0.205000
(Iteration 11 / 20) loss: 1.071353
(Iteration 12 / 20) loss: 0.919177
(Epoch 6 / 10) train acc: 0.760000; val_acc: 0.201000
(Iteration 13 / 20) loss: 0.917780
(Iteration 14 / 20) loss: 0.716143
(Epoch 7 / 10) train acc: 0.790000; val_acc: 0.214000
(Iteration 15 / 20) loss: 0.891181
(Iteration 16 / 20) loss: 0.484528
(Epoch 8 / 10) train acc: 0.830000; val_acc: 0.222000
(Iteration 17 / 20) loss: 0.606501
(Iteration 18 / 20) loss: 0.347264
(Epoch 9 / 10) train acc: 0.930000; val_acc: 0.249000
(Iteration 19 / 20) loss: 0.258791
(Iteration 20 / 20) loss: 0.253796
(Epoch 10 / 10) train acc: 0.920000; val_acc: 0.243000
```

In [34]:
```
plt.subplot(2, 1, 1)
plt.plot(solver.loss_history, 'o')
plt.xlabel('iteration')
plt.ylabel('loss')

plt.subplot(2, 1, 2)
plt.plot(solver.train_acc_history, '-o')
plt.plot(solver.val_acc_history, '-o')
plt.legend(['train', 'val'], loc='upper left')
plt.xlabel('epoch')
plt.ylabel('accuracy')
plt.show()
```



Train the network

Now we train the 3 layer CNN on CIFAR-10 and assess its accuracy.

In [35]:
```
model = ThreeLayerConvNet(weight_scale=0.001, hidden_dim=500, reg=0.001)

solver = Solver(model, data,
                num_epochs=1, batch_size=50,
                update_rule='adam',
                optim_config={
                  'learning_rate': 1e-3,
                },
                verbose=True, print_every=20)
solver.train()
```

```
(Iteration 1 / 980) loss: 2.304795
(Epoch 0 / 1) train acc: 0.097000; val_acc: 0.119000
(Iteration 21 / 980) loss: 2.108474
(Iteration 41 / 980) loss: 1.699203
(Iteration 61 / 980) loss: 1.748868
(Iteration 81 / 980) loss: 1.617362
(Iteration 101 / 980) loss: 2.079706
(Iteration 121 / 980) loss: 1.497607
(Iteration 141 / 980) loss: 1.826750
(Iteration 161 / 980) loss: 2.141741
(Iteration 181 / 980) loss: 1.882131
(Iteration 201 / 980) loss: 1.991266
(Iteration 221 / 980) loss: 1.667279
(Iteration 241 / 980) loss: 1.549250
(Iteration 261 / 980) loss: 1.755482
(Iteration 281 / 980) loss: 1.586607
(Iteration 301 / 980) loss: 1.235908
(Iteration 321 / 980) loss: 1.873307
(Iteration 341 / 980) loss: 1.878106
(Iteration 361 / 980) loss: 1.570670
(Iteration 381 / 980) loss: 1.969980
(Iteration 401 / 980) loss: 1.601958
(Iteration 421 / 980) loss: 1.607514
(Iteration 441 / 980) loss: 1.916853
(Iteration 461 / 980) loss: 1.843414
(Iteration 481 / 980) loss: 1.688600
(Iteration 501 / 980) loss: 1.533797
(Iteration 521 / 980) loss: 1.565566
(Iteration 541 / 980) loss: 1.648820
(Iteration 561 / 980) loss: 1.347325
(Iteration 581 / 980) loss: 1.796001
(Iteration 601 / 980) loss: 1.601116
(Iteration 621 / 980) loss: 1.906355
(Iteration 641 / 980) loss: 1.398132
(Iteration 661 / 980) loss: 1.730645
(Iteration 681 / 980) loss: 1.624738
(Iteration 701 / 980) loss: 1.296591
(Iteration 721 / 980) loss: 1.644106
(Iteration 741 / 980) loss: 1.596834
(Iteration 761 / 980) loss: 1.588982
(Iteration 781 / 980) loss: 1.433749
(Iteration 801 / 980) loss: 1.925017
(Iteration 821 / 980) loss: 1.767805
(Iteration 841 / 980) loss: 1.556007
(Iteration 861 / 980) loss: 1.475954
(Iteration 881 / 980) loss: 1.558819
(Iteration 901 / 980) loss: 1.493340
(Iteration 921 / 980) loss: 1.783761
(Iteration 941 / 980) loss: 1.554866
(Iteration 961 / 980) loss: 1.604871
(Epoch 1 / 1) train acc: 0.433000; val_acc: 0.457000
```

Get > 65% validation accuracy on CIFAR-10.

In the last part of the assignment, we'll now ask you to train a CNN to get better than 65% validation accuracy on CIFAR-10.

Things you should try:

- Filter size: Above we used 7x7; but VGGNet and onwards showed stacks of 3x3 filters are good.
- Number of filters: Above we used 32 filters. Do more or fewer do better?
- Batch normalization: Try adding spatial batch normalization after convolution layers and vanilla batch normalization after affine layers. Do your networks train faster?
- Network architecture: Can a deeper CNN do better? Consider these architectures:
  - [conv-relu-pool]xN - conv - relu - [affine]xM - [softmax or SVM]
  - [conv-relu-pool]XN - [affine]XM - [softmax or SVM]
  - [conv-relu-conv-relu-pool]xN - [affine]xM - [softmax or SVM]

Tips for training

For each network architecture that you try, you should tune the learning rate and regularization strength. When doing this there are a couple of important things to keep in mind:

- If the parameters are working well, you should see improvement within a few hundred iterations
- Remember the coarse-to-fine approach for hyperparameter tuning: start by testing a large range of hyperparameters for just a few training iterations to find the combinations of parameters that are working at all.
- Once you have found some sets of parameters that seem to work, search more finely around these parameters. You may need to train for more epochs.

In [42]:
```
# ================================================================ #
# YOUR CODE HERE:
#   Implement a CNN to achieve greater than 65% validation accuracy
#   on CIFAR-10.
# ================================================================ #

optimizer = 'adam'
best_model = None

weight_scale = 0.01
learning_rate = 1e-3
lr_decay = 0.9

model = ThreeLayerConvNet(weight_scale=0.001, hidden_dim=500, reg=0.001)

solver = Solver(model, data,
                num_epochs=10, batch_size=100,
                update_rule=optimizer,
                optim_config={
                  'learning_rate': learning_rate,
                },
                lr_decay=lr_decay,
                verbose=True, print_every=50)

solver.train()

# ================================================================ #
# END YOUR CODE HERE
# ================================================================ #
```

```
(Iteration 1 / 4900) loss: 2.304510
(Epoch 0 / 10) train acc: 0.116000; val_acc: 0.122000
(Iteration 51 / 4900) loss: 1.770843
(Iteration 101 / 4900) loss: 1.831444
(Iteration 151 / 4900) loss: 1.723793
(Iteration 201 / 4900) loss: 1.553610
(Iteration 251 / 4900) loss: 1.393221
(Iteration 301 / 4900) loss: 1.565930
(Iteration 351 / 4900) loss: 1.698151
(Iteration 401 / 4900) loss: 1.492429
(Iteration 451 / 4900) loss: 1.508485
(Epoch 1 / 10) train acc: 0.489000; val_acc: 0.492000
(Iteration 501 / 4900) loss: 1.556234
(Iteration 551 / 4900) loss: 1.306318
(Iteration 601 / 4900) loss: 1.353494
(Iteration 651 / 4900) loss: 1.351386
(Iteration 701 / 4900) loss: 1.308536
(Iteration 751 / 4900) loss: 1.467527
(Iteration 801 / 4900) loss: 1.188758
(Iteration 851 / 4900) loss: 1.412666
(Iteration 901 / 4900) loss: 1.445009
(Iteration 951 / 4900) loss: 1.344242
(Epoch 2 / 10) train acc: 0.563000; val_acc: 0.550000
(Iteration 1001 / 4900) loss: 1.275696
(Iteration 1051 / 4900) loss: 1.150632
(Iteration 1101 / 4900) loss: 1.368649
(Iteration 1151 / 4900) loss: 1.295056
(Iteration 1201 / 4900) loss: 1.222005
(Iteration 1251 / 4900) loss: 1.155882
(Iteration 1301 / 4900) loss: 1.030409
(Iteration 1351 / 4900) loss: 1.143373
(Iteration 1401 / 4900) loss: 1.340394
(Iteration 1451 / 4900) loss: 1.295085
(Epoch 3 / 10) train acc: 0.571000; val_acc: 0.569000
(Iteration 1501 / 4900) loss: 1.128084
(Iteration 1551 / 4900) loss: 1.125984
(Iteration 1601 / 4900) loss: 1.079257
(Iteration 1651 / 4900) loss: 1.204651
(Iteration 1701 / 4900) loss: 1.215688
(Iteration 1751 / 4900) loss: 1.169037
(Iteration 1801 / 4900) loss: 0.942539
(Iteration 1851 / 4900) loss: 0.866118
(Iteration 1901 / 4900) loss: 1.085536
(Iteration 1951 / 4900) loss: 0.968670
(Epoch 4 / 10) train acc: 0.635000; val_acc: 0.610000
(Iteration 2001 / 4900) loss: 1.087900
(Iteration 2051 / 4900) loss: 1.114967
(Iteration 2101 / 4900) loss: 1.006622
(Iteration 2151 / 4900) loss: 1.175996
(Iteration 2201 / 4900) loss: 0.824987
(Iteration 2251 / 4900) loss: 0.857084
(Iteration 2301 / 4900) loss: 1.419721
(Iteration 2351 / 4900) loss: 1.150189
(Iteration 2401 / 4900) loss: 1.239157
(Epoch 5 / 10) train acc: 0.648000; val_acc: 0.631000
(Iteration 2451 / 4900) loss: 1.304620
(Iteration 2501 / 4900) loss: 1.046581
```

```
(Iteration 2551 / 4900) loss: 0.855849
(Iteration 2601 / 4900) loss: 1.007044
(Iteration 2651 / 4900) loss: 1.064739
(Iteration 2701 / 4900) loss: 1.069459
(Iteration 2751 / 4900) loss: 0.930649
(Iteration 2801 / 4900) loss: 1.077328
(Iteration 2851 / 4900) loss: 0.816714
(Iteration 2901 / 4900) loss: 1.113363
(Epoch 6 / 10) train acc: 0.680000; val_acc: 0.630000
(Iteration 2951 / 4900) loss: 0.965668
(Iteration 3001 / 4900) loss: 1.093905
(Iteration 3051 / 4900) loss: 1.146155
(Iteration 3101 / 4900) loss: 0.916298
(Iteration 3151 / 4900) loss: 1.013372
(Iteration 3201 / 4900) loss: 0.860397
(Iteration 3251 / 4900) loss: 1.040853
(Iteration 3301 / 4900) loss: 0.906793
(Iteration 3351 / 4900) loss: 1.052177
(Iteration 3401 / 4900) loss: 0.909095
(Epoch 7 / 10) train acc: 0.721000; val_acc: 0.633000
(Iteration 3451 / 4900) loss: 1.053566
(Iteration 3501 / 4900) loss: 0.856925
(Iteration 3551 / 4900) loss: 0.934538
(Iteration 3601 / 4900) loss: 0.896635
(Iteration 3651 / 4900) loss: 1.040976
(Iteration 3701 / 4900) loss: 0.856940
(Iteration 3751 / 4900) loss: 0.988786
(Iteration 3801 / 4900) loss: 1.018906
(Iteration 3851 / 4900) loss: 1.081226
(Iteration 3901 / 4900) loss: 0.875528
(Epoch 8 / 10) train acc: 0.717000; val_acc: 0.657000
(Iteration 3951 / 4900) loss: 0.853751
(Iteration 4001 / 4900) loss: 0.763060
(Iteration 4051 / 4900) loss: 0.895589
(Iteration 4101 / 4900) loss: 1.066831
(Iteration 4151 / 4900) loss: 0.882167
(Iteration 4201 / 4900) loss: 0.678572
(Iteration 4251 / 4900) loss: 0.934127
(Iteration 4301 / 4900) loss: 0.907197
```

```
--------------------------------------------------------------------------
----
KeyboardInterrupt                         Traceback (most recent call l
ast)
<ipython-input-42-8da3df2a5f4d> in <module>()
     23                 verbose=True, print_every=50)
     24 
---> 25 solver.train()
     26 
     27 # ==============================================================
=== #

~/GoogleDrive/UCLA/Academics/Junior/Q2/EC ENGR C247/HW5/cs231n/solver.p
y in train(self)
    262 
    263         for t in range(num_iterations):
--> 264             self._step()
    265 
    266             # Maybe print training loss

~/GoogleDrive/UCLA/Academics/Junior/Q2/EC ENGR C247/HW5/cs231n/solver.p
y in _step(self)
    185             dw = grads[p]
    186             config = self.optim_configs[p]
--> 187             next_w, next_config = self.update_rule(w, dw, confi
g)
    188             self.model.params[p] = next_w
    189             self.optim_configs[p] = next_config

~/GoogleDrive/UCLA/Academics/Junior/Q2/EC ENGR C247/HW5/nndl/optim.py i
n adam(w, dw, config)
    197     config['v'] = config['beta1'] * config['v'] + (1 - config[
'beta1']) * dw
    198     config['a'] = config['beta2'] * config['a'] + (1 - config[
'beta2']) * dw * dw
--> 199     v_corrected = config['v'] / (1 - config['beta1']**config[
't'])
    200     a_corrected = config['a'] / (1 - config['beta2']**config[
't'])
    201     next_w = w - v_corrected * config['learning_rate'] / (np.sq
rt(a_corrected) + config['epsilon'])

KeyboardInterrupt: 
```

```
import numpy as np
from nndl.layers import *
import pdb

"""
This code was originally written for CS 231n at Stanford University
(cs231n.stanford.edu).  It has been modified in various areas for use
in the
ECE 239AS class at UCLA.  This includes the descriptions of what code
to
implement as well as some slight potential changes in variable names
to be
consistent with class nomenclature.  We thank Justin Johnson & Serena
Yeung for
permission to use this code.  To see the original version, please
visit
cs231n.stanford.edu.
"""

def conv_forward_naive(x, w, b, conv_param):
  """
  A naive implementation of the forward pass for a convolutional
layer.

  The input consists of N data points, each with C channels, height H
and width
  W. We convolve each input with F different filters, where each
filter spans
  all C channels and has height HH and width HH.

  Input:
  - x: Input data of shape (N, C, H, W)
  - w: Filter weights of shape (F, C, HH, WW)
  - b: Biases, of shape (F,)
  - conv_param: A dictionary with the following keys:
    - 'stride': The number of pixels between adjacent receptive fields
in the
      horizontal and vertical directions.
    - 'pad': The number of pixels that will be used to zero-pad the
input.

  Returns a tuple of:
  - out: Output data, of shape (N, F, H', W') where H' and W' are
given by
    H' = 1 + (H + 2 * pad - HH) / stride
    W' = 1 + (W + 2 * pad - WW) / stride
  - cache: (x, w, b, conv_param)
  """
  out = None
  pad = conv_param['pad']
```

```
  stride = conv_param['stride']

  # ================================================================ #
  # YOUR CODE HERE:
  #   Implement the forward pass of a convolutional neural network.
  #   Store the output as 'out'.
  #   Hint: to pad the array, you can use the function np.pad.
  # ================================================================ #

  n, c, xh, xw = x.shape
  wf, wc, wh, ww = w.shape
  hout, wout = 1 + (xh + 2*pad - wh)//stride, 1 + (xw + 2*pad - ww)//
stride

  out = np.zeros((len(x), wf, hout, wout))

  # pad the input
  x_padded = np.pad(x, ((0, 0), (0, 0), (pad, pad), (pad, pad)),
'constant', constant_values=0)

  # loop over all inputs
  for i in range(n):
      # loop over filters
      for j in range(wf):
          # loop over each cell in output
          for hi in range(hout):
              for wi in range(wout):
                  out[i, j, hi, wi] = np.sum(x_padded[i, :, hi*stride:
hi*stride + wh, wi*stride: wi*stride + ww] * w[j]) + b[j]

  # ================================================================ #
  # END YOUR CODE HERE
  # ================================================================ #

  cache = (x, w, b, conv_param)
  return out, cache


def conv_backward_naive(dout, cache):
  """
  A naive implementation of the backward pass for a convolutional
layer.

  Inputs:
  - dout: Upstream derivatives.
  - cache: A tuple of (x, w, b, conv_param) as in conv_forward_naive

  Returns a tuple of:
  - dx: Gradient with respect to x
  - dw: Gradient with respect to w
```

```
  - db: Gradient with respect to b
  """
  dx, dw, db = None, None, None

  N, F, out_height, out_width = dout.shape
  x, w, b, conv_param = cache

  stride, pad = [conv_param['stride'], conv_param['pad']]
  xpad = np.pad(x, ((0,0), (0,0), (pad,pad), (pad,pad)),
mode='constant')
  num_filts, _, f_height, f_width = w.shape

  # ================================================================ #
  # YOUR CODE HERE:
  #   Implement the backward pass of a convolutional neural network.
  #   Calculate the gradients: dx, dw, and db.
  # ================================================================ #

  xn, xc, xh, xw = xpad.shape
  d_n, d_f, d_h, d_w = dout.shape
  wf, wc, wh, ww = w.shape

  dw = np.zeros(w.shape)
  dx = np.zeros(xpad.shape)
  db = np.zeros(b.shape)

  for i in range(d_n):
      for j in range(d_f):
          for k in range(d_h):
              for l in range(d_w):
                  dw[j] += xpad[i, :, k*stride: k*stride + wh,
l*stride: l*stride + ww] * dout[i, j, k, l]
                  db[j] += dout[i, j, k, l]
                  dx[i, :, k*stride: k*stride + wh, l*stride: l*stride
+ ww] += dout[i, j, k, l] * w[j]

  # the pad is used to help calculate the gradient, so we now remove
the pad from the gradient
  xpad_h, xpad_w = xpad.shape[2], xpad.shape[3]
  dx = dx[:, :, pad:xpad_h - pad, pad:xpad_w - pad]

  # ================================================================ #
  # END YOUR CODE HERE
  # ================================================================ #

  return dx, dw, db


def max_pool_forward_naive(x, pool_param):
  """
```

```
    A naive implementation of the forward pass for a max pooling layer.

    Inputs:
    - x: Input data, of shape (N, C, H, W)
    - pool_param: dictionary with the following keys:
      - 'pool_height': The height of each pooling region
      - 'pool_width': The width of each pooling region
      - 'stride': The distance between adjacent pooling regions

    Returns a tuple of:
    - out: Output data
    - cache: (x, pool_param)
    """
    out = None

    # ================================================================= #
    # YOUR CODE HERE:
    #   Implement the max pooling forward pass.
    # ================================================================= #

    p_h, p_w, stride = pool_param['pool_height'],
  pool_param['pool_width'], pool_param['stride']
    n, c, h, w = x.shape
    h_new, w_new = (h - p_h)//stride + 1, (w - p_w)//stride + 1

    out = np.zeros((n, c, h_new, w_new))

    for i in range(n):
        for j in range(c):
            for k in range(h_new):
                for l in range(w_new):
                    out[i, j, k, l] = np.max(x[i, j, k*stride: k*stride
  + p_h, l*stride: l*stride + p_w])

    # ================================================================= #
    # END YOUR CODE HERE
    # ================================================================= #
    cache = (x, pool_param)
    return out, cache

  def max_pool_backward_naive(dout, cache):
    """
    A naive implementation of the backward pass for a max pooling layer.

    Inputs:
    - dout: Upstream derivatives
    - cache: A tuple of (x, pool_param) as in the forward pass.

    Returns:
    - dx: Gradient with respect to x
```

```
  """
  dx = None
  x, pool_param = cache
  pool_height, pool_width, stride = pool_param['pool_height'],
pool_param['pool_width'], pool_param['stride']

  # ================================================================ #
  # YOUR CODE HERE:
  #   Implement the max pooling backward pass.
  # ================================================================ #

  dout_n, dout_c, dout_w, dout_h = dout.shape

  dx = np.zeros(x.shape)

  for i in range(dout_n):
      for j in range(dout_c):
          for k in range(dout_w):
              for l in range(dout_h):
                  slice = x[i, j, k*stride: k*stride + pool_height,
l*stride: l*stride + pool_width]
                  mask = slice == np.max(slice)
                  dx[i, j, k*stride: k*stride + pool_height, l*stride:
l*stride + pool_width] += mask * dout[i, j, k, l]

  # ================================================================ #
  # END YOUR CODE HERE
  # ================================================================ #

  return dx

def spatial_batchnorm_forward(x, gamma, beta, bn_param):
  """
  Computes the forward pass for spatial batch normalization.

  Inputs:
  - x: Input data of shape (N, C, H, W)
  - gamma: Scale parameter, of shape (C,)
  - beta: Shift parameter, of shape (C,)
  - bn_param: Dictionary with the following keys:
    - mode: 'train' or 'test'; required
    - eps: Constant for numeric stability
    - momentum: Constant for running mean / variance. momentum=0 means
that
      old information is discarded completely at every time step,
while
      momentum=1 means that new information is never incorporated. The
      default of momentum=0.9 should work well in most situations.
    - running_mean: Array of shape (D,) giving running mean of
features
```

```
    - running_var Array of shape (D,) giving running variance of
features

  Returns a tuple of:
  - out: Output data, of shape (N, C, H, W)
  - cache: Values needed for the backward pass
  """
  out, cache = None, None

  # ================================================================== #
  # YOUR CODE HERE:
  #   Implement the spatial batchnorm forward pass.
  #
  #   You may find it useful to use the batchnorm forward pass you
  #   implemented in HW #4.
  # ================================================================== #

  n, c, h, w = x.shape
  x = x.transpose(0, 2, 3, 1).reshape((n*h*w, c))
  out, cache = batchnorm_forward(x, gamma, beta, bn_param)
  out = out.reshape(n, h, w, c).transpose(0, 3, 1, 2)

  # ================================================================== #
  # END YOUR CODE HERE
  # ================================================================== #

  return out, cache


def spatial_batchnorm_backward(dout, cache):
  """
  Computes the backward pass for spatial batch normalization.

  Inputs:
  - dout: Upstream derivatives, of shape (N, C, H, W)
  - cache: Values from the forward pass

  Returns a tuple of:
  - dx: Gradient with respect to inputs, of shape (N, C, H, W)
  - dgamma: Gradient with respect to scale parameter, of shape (C,)
  - dbeta: Gradient with respect to shift parameter, of shape (C,)
  """
  dx, dgamma, dbeta = None, None, None

  # ================================================================== #
  # YOUR CODE HERE:
  #   Implement the spatial batchnorm backward pass.
  #
  #   You may find it useful to use the batchnorm backward pass you
  #   implemented in HW #4.
```

```
# ================================================================ #

n, c, h, w = dout.shape
dout = dout.transpose(0, 2, 3, 1).reshape((n*h*w, c))
dx, dgamma, dbeta = batchnorm_backward(dout, cache)
dx = dx.reshape(n, h, w, c).transpose(0, 3, 1, 2)

# ================================================================ #
# END YOUR CODE HERE
# ================================================================ #

return dx, dgamma, dbeta
```

```
import numpy as np

from nndl.layers import *
from nndl.conv_layers import *
from cs231n.fast_layers import *
from nndl.layer_utils import *
from nndl.conv_layer_utils import *

import pdb

"""
This code was originally written for CS 231n at Stanford University
(cs231n.stanford.edu).  It has been modified in various areas for use
in the
ECE 239AS class at UCLA.  This includes the descriptions of what code
to
implement as well as some slight potential changes in variable names
to be
consistent with class nomenclature.  We thank Justin Johnson & Serena
Yeung for
permission to use this code.  To see the original version, please
visit
cs231n.stanford.edu.
"""

class ThreeLayerConvNet(object):
  """
  A three-layer convolutional network with the following architecture:

  conv - relu - 2x2 max pool - affine - relu - affine - softmax

  The network operates on minibatches of data that have shape (N, C,
H, W)
  consisting of N images, each with height H and width W and with C
input
  channels.
  """

  def __init__(self, input_dim=(3, 32, 32), num_filters=32,
filter_size=7,
               hidden_dim=100, num_classes=10, weight_scale=1e-3,
reg=0.0,
               dtype=np.float32, use_batchnorm=False):
    """
    Initialize a new network.

    Inputs:
    - input_dim: Tuple (C, H, W) giving size of input data
    - num_filters: Number of filters to use in the convolutional layer
    - filter_size: Size of filters to use in the convolutional layer
```

```
    - hidden_dim: Number of units to use in the fully-connected hidden
layer
    - num_classes: Number of scores to produce from the final affine
layer.
    - weight_scale: Scalar giving standard deviation for random
initialization
      of weights.
    - reg: Scalar giving L2 regularization strength
    - dtype: numpy datatype to use for computation.
    """
    self.use_batchnorm = use_batchnorm
    self.params = {}
    self.reg = reg
    self.dtype = dtype


    # ================================================================
#
    # YOUR CODE HERE:
    #   Initialize the weights and biases of a three layer CNN. To
initialize:
    #     - the biases should be initialized to zeros.
    #     - the weights should be initialized to a matrix with entries
    #         drawn from a Gaussian distribution with zero mean and
    #         standard deviation given by weight_scale.
    # ================================================================
#

    c, h, w = input_dim
    self.params["W1"] = np.random.normal(scale=weight_scale,
size=((num_filters, c, filter_size, filter_size)))
    self.params["b1"] = np.zeros(num_filters)

    # dimensions after convolution
    stride = 1
    pad = (filter_size - 1) // 2
    h_conv, w_conv = (h - filter_size + 2*pad) // stride + 1, (w -
filter_size + 2*pad) // stride + 1

    # dimensions after 2x2 max pooling
    stride = 2
    h_pool, w_pool = (h - 2) // stride + 1, (w - 2) // stride + 1

    # flatten feature axes
    input_dim = h_pool * w_pool * num_filters

    self.params["W2"] = np.random.normal(scale=weight_scale,
size=(input_dim, hidden_dim))
    self.params["b2"] = np.zeros(hidden_dim)
```

```
    self.params["W3"] = np.random.normal(scale=weight_scale,
size=(hidden_dim, num_classes))
    self.params["b3"] = np.zeros(num_classes)

    # ================================================================
#
    # END YOUR CODE HERE
    # ================================================================
#

    for k, v in self.params.items():
      self.params[k] = v.astype(dtype)


  def loss(self, X, y=None):
    """
    Evaluate loss and gradient for the three-layer convolutional
network.

    Input / output: Same API as TwoLayerNet in fc_net.py.
    """
    W1, b1 = self.params['W1'], self.params['b1']
    W2, b2 = self.params['W2'], self.params['b2']
    W3, b3 = self.params['W3'], self.params['b3']

    # pass conv_param to the forward pass for the convolutional layer
    filter_size = W1.shape[2]
    conv_param = {'stride': 1, 'pad': (filter_size - 1) / 2}

    # pass pool_param to the forward pass for the max-pooling layer
    pool_param = {'pool_height': 2, 'pool_width': 2, 'stride': 2}

    scores = None

    # ================================================================
#
    # YOUR CODE HERE:
    #   Implement the forward pass of the three layer CNN.  Store the
output
    #   scores as the variable "scores".
    # ================================================================
#

    conv_out, conv_cache = conv_relu_pool_forward(X, W1, b1,
conv_param, pool_param)
    # flatten all the feature maps into one axis
    flat_out = conv_out.reshape((conv_out.shape[0], -1))
    relu_out, relu_cache = affine_relu_forward(flat_out, W2, b2)
    scores, aff_cache = affine_forward(relu_out, W3, b3)
```

```
    # ================================================================
#
    # END YOUR CODE HERE
    # ================================================================
#

    if y is None:
      return scores

    loss, grads = 0, {}
    # ================================================================
#
    # YOUR CODE HERE:
    #   Implement the backward pass of the three layer CNN.  Store the
grads
    #   in the grads dictionary, exactly as before (i.e., the gradient
of
    #   self.params[k] will be grads[k]).  Store the loss as "loss",
and
    #   don't forget to add regularization on ALL weight matrices.
    # ================================================================
#

    loss, dout = softmax_loss(scores, y)
    loss += 0.5 * self.reg * (sum([np.sum(self.params["W" +
str(i)]**2) for i in range(1, 4)]))

    dx_aff, grads["W3"], grads["b3"] = affine_backward(dout,
aff_cache)
    dx_relu, grads["W2"], grads["b2"] = affine_relu_backward(dx_aff,
relu_cache)
    # resuscitate the feature map axes from the flattened
    dx_resuscitated = dx_relu.reshape(conv_out.shape)
    dx_conv, grads["W1"], grads["b1"] =
conv_relu_pool_backward(dx_resuscitated, conv_cache)

    for i in range(1, 4):
        grads["W" + str(i)] += self.reg * self.params["W" + str(i)]

    # ================================================================
#
    # END YOUR CODE HERE
    # ================================================================
#

    return loss, grads


pass
```