

## 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  $(N \cdot H \cdot W, C)$  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.

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

import time
import numpy as np
import matplotlib.pyplot as plt
from nndl.conv_layers import *
from utils.data_utils import get_CIFAR10_data
```

```

from utils.gradient_check import eval_numerical_gradient, eval_numerical_gradient_array
from utils.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-in-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))))

```

## 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 [ ]: # 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: [9.33882566 9.41475073 8.73902467]  
 Stds: [3.78556883 3.50655032 4.03980908]

After spatial batch normalization:

Shape: (2, 3, 4, 5)  
 Means: [-4.21884749e-16 4.44089210e-16 -3.77475828e-16]  
 Stds: [0.99999965 0.99999959 0.99999969]

After spatial batch normalization (nontrivial gamma, beta):

Shape: (2, 3, 4, 5)  
 Means: [6. 7. 8.]  
 Stds: [2.99999895 3.99999837 4.99999847]

## 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 [ ]:

```
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: 1.536124647280122e-08
dgamma error: 1.0642010844431384e-11
dbeta error: 5.238952335848806e-12
```