```python
import numpy as np
from numpy.core.defchararray import add
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, H, W = x.shape
    F, C, HH, WW = w.shape
    Hout = 1 + (H + 2 * pad - HH) // stride
    Wout = 1 + (W + 2 * pad - WW) // stride
    x_padded = np.pad(x, ((0, 0), (0, 0), (pad, pad), (pad, pad)))
    out = np.zeros((N, F, Hout, Wout))

```

```python
    for n in range(N):
        for f in range(F):
            for h in range(Hout):
                for j in range(Wout):
                    cur_x = x_padded[n, :, h * stride:h *
                                     stride + HH, j * stride:j * stride + WW]
                    out[n, f, h, j] = np.sum(cur_x * w[f]) + b[f]

    # ================================================================ #
    # 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.
    # ================================================================ #
    N, C, H, W = x.shape
    H_out = 1 + (H + 2 * pad - f_height) // stride
    W_out = 1 + (W + 2 * pad - f_width) // stride
    dxpad = np.zeros_like(xpad)
    dx = np.zeros(x.shape)
    dw = np.zeros(w.shape)
    db = np.zeros(b.shape)

    for n in range(N):
        for f in range(num_filts):
            db[f] += np.sum(dout[n, f])
            for h in range(H_out):
                hs = h*stride
                for j in range(W_out):
                    ws = j * stride
                    dw[f] += xpad[n, :, hs:hs + f_height,
```

```python
                                  ws:ws + f_width] * dout[n, f, h, j]
                    dxpad[n, :, hs:hs + f_height, ws:ws +
                          f_width] += w[f] * dout[n, f, h, j]
    dx = dxpad[:, :, pad:pad+H, pad:pad+W]
    # ================================================================ #
    # 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.
    # ================================================================ #
    N, C, H, W = x.shape
    pool_height = pool_param['pool_height']
    pool_width = pool_param['pool_width']
    stride = pool_param['stride']
    Hout = 1 + (H - pool_height) // stride
    Wout = 1 + (W - pool_width) // stride
    out = np.zeros((N, C, Hout, Wout))

    for n in range(N):
        for c in range(C):
            for j in range(Wout):
                for m in range(Hout):
                    mstride = m * stride
                    ws = j * stride
                    window = x[n, c, mstride:mstride +
                               pool_height, ws:ws+pool_width]
                    out[n, c, m, j] = np.max(window)
    # ================================================================ #
    # 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.

```

```python
    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.
    # ================================================================ #
    N, C, H, W = x.shape
    H_out = 1 + (H - pool_height) // stride
    W_out = 1 + (W - pool_width) // stride
    dx = np.zeros(x.shape)

    for n in range(N):
        for c in range(C):
            for h in range(H_out):
                hstride = h * stride
                for j in range(W_out):
                    wstride = j * stride
                    window = x[n, c, hstride:hstride +
                               pool_height, wstride:wstride+pool_width]
                    m = np.max(window)
                    dx[n, c, hstride:hstride+pool_height, wstride:wstride +
                        pool_width] += (window == m) * dout[n, c, h, j]
    # ================================================================ #
    # 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_new = x.transpose(0, 3, 2, 1).reshape((N*H*W, C))

    out, cache = batchnorm_forward(x_new, gamma, beta, bn_param)
    out = out.reshape(N, W, H, C).transpose(0, 3, 2, 1)

    # ================================================================ #
    # 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 forward pass you
    #   implemented in HW #4.
    # ================================================================ #
    N, C, H, W = dout.shape
    dout_new = dout.transpose(0,3,2,1).reshape((N*H*W, C))
    dx, dgamma, dbeta = batchnorm_backward(dout_new, cache)
    dx = dx.reshape(N, W, H, C).transpose(0, 3, 2, 1)
    # ================================================================ #
    # END YOUR CODE HERE
    # ================================================================ #

    return dx, dgamma, dbeta

```