1,

Some issues:

(1) There is a minor mistake in the paper that the input image should be 227 X 227 X 3 rather than 224 X 224 X 3.

(2) The caption under Figure 2 is “The network’s input is 150,528-dimensional, and the number of neurons in the network’s remaining layers is given by 253,440–186,624–64,896–64,896–43,264– 4096–4096–1000”. The first number “253440” here is wrong.

(3) Pay attention to the 2-tower architecture where feature maps are splited, the stride of convolutional filters and pooling and whether convolution is with padding or not.

(4) Do not count bias.

Conv layer:

Units = f_h * f_w * N
Weights = h * w * C * N
Connections = h * w * C * N * X * Y

h: filter height, w: filter width, C: number of input channels, N: number of filters
H: input height, W: input width,
f_h: height of feature map,
f_w: width of feature map,
S_x: stride along width dimension,
S_y: stride along height dimension,
X: number of possible positions for convolution along width dimension,
Y: number of possible positions for convolution along height dimension,
X = [(W - 1) / S_x + 1] (with padding)
Y = [(H - 1) / S_y + 1] (with padding)
X = [(W - [w/2] - [w/2] - 1) / S_x + 1] (without padding)
Y = [(H - [h/2] - [h/2] - 1) / S_y + 1] (without padding)
where [x] means taking the floor interger, e.g., [2.6] = 2, [3.1] = 3.

FC layer:

Units = # Output
Weights = # Input * # Output
Connections = # Weights

	# Units	# Weights	# Connections
Convolution Layer 1	55 * 55 * 96 = 290400	11 * 11 * 3 * 96 = 34848	11 * 11 * 3 * 96 * 55 * 55 = 105415200 (Single tower here!)
Convolution Layer 2	27 * 27 * 256 = 186624	5 * 5 * 48 * 256 = 307200	5 * 5 * 48 * 128 * 2 * 27 * 27 = 223948800 (Two towers here!)
Convolution Layer 3	13 * 13 * 384 = 64896	3 * 3 * 256 * 384 = 884736	3 * 3 * 128 * 384 * 2 * 13 * 13 = 149520384 (Single tower here!)
Convolution Layer 4	13 * 13 * 384 = 64896	3 * 3 * 192 * 384 = 663552	3 * 3 * 192 * 192 * 2 * 13 * 13 = 112140288 (Two towers here!)
Convolution Layer 5	13 * 13 * 256 = 43264	3 * 3 * 192 * 256 = 442368	3 * 3 * 192 * 128 * 2 * 13 * 13 = 74760192 (Two towers here!)
Fully Connected Layer 1	4096	9216 * 4096 = 37748736	9216 * 4096 = 37748736
Fully Connected Layer 2	4096	4096 * 4096 = 16777216	4096 * 4096 = 16777216
Output Layer	1000	4096 * 1000 = 4096000	4096 * 1000 = 4096000

2,

(a) You can reduce the number of parameters by e.g., reducing the size of the fully connected layer. As you can tell from the above table, two FC layers occupy a lot of parameters, especially FC layer 1.

(b) You can reduce the connections by use fewer number of filters for the convolutional layers. As you can see from the above table, the convolutional layers have a lot of connections and each of them is approximately one add-multiplication operation.