CNN Layer Sizing & Parameter-Count Report

This report summarizes the key formulas and procedures you need to:

1. Compute **output feature-map dimensions**
2. Compute **learnable parameter counts**

for **convolutional**, **pooling**, and **fully-connected** layers in any standard CNN.

---

1. Convolutional Layers

1.1. Definitions

- **Input feature-map**:
  width = W_{in},
  height = H_{in},
  depth (channels) = D_{in}.
- **Kernel (filter)**: size $F \times F$.
- **Number of filters**: K.
- **Stride**: S.
- **Padding**: P
  - “same” padding $\rightarrow P = \lfloor F/2 \rfloor$
  - “valid” padding $\rightarrow P = 0$

1.2. Output Dimensions

The output width W_{out} and height H_{out} are

$$W_{\text{out}} = \left\lfloor \frac{W_{\text{in}} - F + 2P}{S} \right\rfloor + 1, H_{\text{out}} = \left\lfloor \frac{H_{\text{in}} - F + 2P}{S} \right\rfloor + 1.$$

The output **depth** is simply the number of filters:

$$D_{\text{out}} = K.$$

1.3. Parameter Count (Revised)

Each convolutional layer has **one bias per filter**, so:

- **Weights per filter**: $F \times F \times D_{\text{in}}$
- **Biases per filter**: 1
- **Number of filters**: K

Total parameters for the layer

- **Factored form** (emphasizes “per-filter” cost):

$$\text{Params}_{\text{conv}} = K\big(F^2 \cdot D_{\text{in}} + 1\big).$$

- **Expanded form** (separates weights vs. biases):

$$\text{Params}_{\text{conv}} = K \cdot F^2 \cdot D_{\text{in}} + K$$

---

2. Pooling Layers (Max- or Avg-pool)

2.1. Definitions

- **Input feature-map**:
  width = W_{in},
  height = H_{in},
  depth = D_{in}.

- **Pool window**: size $F_p \times F_p$.
- **Stride**: S_p.
- **Padding**: usually $P_p = 0$ (no padding).

2.2. Output Dimensions

$$W_{\text{out}} = \left\lfloor \frac{W_{\text{in}} - F_p + 2P_p}{S_p} \right\rfloor + 1$$

$$H_{\text{out}} = \left\lfloor \frac{H_{\text{in}} - F_p + 2P_p}{S_p} \right\rfloor + 1$$

$$D_{\text{out}} = D_{\text{in}}.$$

2.3. Parameter Count

Pooling layers have **no learnable parameters**:

$$\text{Params}_{\text{pool}} = 0.$$

---

3. Fully-Connected (Dense) Layers

3.1. Definitions

- **Flattened input length**: N_{in}.
- **Number of neurons**: N_{out}.

3.2. Parameter Count

$$\text{Params}_{\text{FC}} = N_{\text{in}} \times N_{\text{out}} + N_{\text{out}}.$$

---

4. Putting It All Together

1. **Start** with your input size (W_0, H_0, D_0).
2. **For each convolutional layer**:
   - Compute (W_i, H_i, D_i) via the conv formula.
   - Compute its parameter count via $\text{Params}_{\text{conv}}$.
3. **For each pooling layer**:
   - Compute (W_i, H_i) via the pooling formula.
   - (No parameters.)
4. **Repeat** over all conv/pool blocks.
5. **Flatten** final (W_f, H_f, D_f) into:

$$N_{\text{in}} = W_f \times H_f \times D_f.$$

6. **For each FC layer**:
   - Apply the FC parameter count formula.
7. **Sum** all layer-wise parameter counts → total model parameters.

Example Workflow

> **Given** an input $224 \times 224 \times 3$, conv($K = 64$, $F = 7$, $S = 2$, same),
> pool($F_p = 3$, $S_p = 2$), conv($K = 128$, $F = 3$, $S = 1$, same), …

1. **Conv1**
   - $F = 7$, $S = 2$, $P = \lfloor 7/2 \rfloor = 3$
   - $$W_{\text{out}} = \lfloor (224 - 7 + 2 \cdot 3)/2 \rfloor + 1 = 112$$
   - $H_{\text{out}} = 112$, $D_{\text{out}} = 64$
   - Params = $64 \times (7^2 \times 3 + 1) = 9{,}472$
2. **Pool1**

- $F_p = 3$, $S_p = 2$, $P_p = 0$
- $$W_{\text{out}} = \lfloor (112 - 3)/2 \rfloor + 1 = 55$$
- $H_{\text{out}} = 55$, $D_{\text{out}} = 64$
- Params = 0

3. **Conv2**
   - $F = 3$, $S = 1$, $P = \lfloor 3/2 \rfloor = 1$
   - $W_{\text{out}} = 55$, $H_{\text{out}} = 55$, $D_{\text{out}} = 128$
   - Params = $128 \times (3^2 \times 64 + 1) = 73{,}856$

…and so on.
