Input

$$128 \times 10 \times 400 \times 352$$

Output channel #64, kernel(3,3,3), stride(2,1,1), padding(1,1,1)
Output channel #64, kernel(3,3,3), stride(1,1,1), padding(0,1,1)
Output channel #64, kernel(3,3,3), stride(2,1,1), padding(1,1,1)

Output

$$C \times D \times H \times W$$

Calculation:

- Since Output channel has 64 channels so C must be 64
- Consider 3 dimension separately, kernel 3 with stride 1 and padding 1 does not change dimension so H and W remains the same
- After first layer 10 becomes (10 + 2*1 - 3) / 2 + 1 = 5
- After second layer 5 - 2 = 3
- Finally (3 + 2 * 1 - 3) / 2 + 1 = 2

Thus final output has dimension

$$64 \times 2 \times 400 \times 352$$