



Correction to slide 33 (highlighted in RED)

| Method | Layer | Size ($C \times T \times H \times W$) |
|-----------------------|--|---|
| Early Fusion (2D CNN) | Input | $3 \times 20 \times 64 \times 64$ |
| | Conv2D (3×3 , $3 \times 20 \rightarrow 12$) | $12 \times 64 \times 64$ |
| | Pool2D (4×4) | $12 \times 16 \times 16$ |
| | Conv2D (3×3 , $12 \rightarrow 24$) | $24 \times 64 \times 64$, $24 \times 16 \times 16$ |
| | GlobalAvgPool | $24 \times 1 \times 1$ |
| Late Fusion (2D CNN) | Input | $3 \times 20 \times 64 \times 64$ |
| | Conv2D (3×3 , $3 \rightarrow 12$) | $12 \times 20 \times 64 \times 64$ |
| | Pool2D (4×4) | $3 \times 20 \times 16 \times 16$, $12 \times 20 \times 16 \times 16$ |
| | Conv2D (3×3 , $12 \rightarrow 24$) | $24 \times 20 \times 64 \times 64$, $24 \times 20 \times 16 \times 16$ |
| | GlobalAvgPool | $24 \times 1 \times 1 \times 1$ |
| 3D CNN | Input | $3 \times 20 \times 64 \times 64$ |
| | Conv3D ($3 \times 3 \times 3$, $3 \rightarrow 12$) | $12 \times 20 \times 64 \times 64$ |
| | Pool3D ($4 \times 4 \times 4$) | $12 \times 5 \times 16 \times 16$ |
| | Conv3D ($3 \times 3 \times 3$, $12 \rightarrow 24$) | $24 \times 5 \times 64 \times 64$, $24 \times 5 \times 16 \times 16$ |
| | GlobalAvgPool | $24 \times 1 \times 1 \times 1$ |