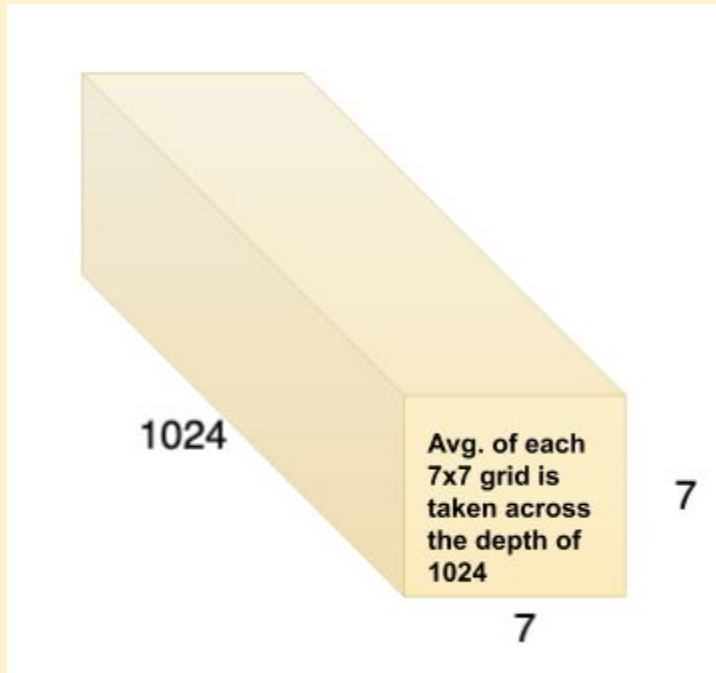


Average Pooling

How does average pooling reduce the output size?

1. At the final Inception Module, we have an output dimension of $1024 \times 7 \times 7$. If this was to directly interface with a Fully-Connected layer with a 1000 Neurons, we would get ~50 Million parameters
2. To reduce this number, Google added another layer which performs Average-pooling



3. By taking the average value of each 7×7 slice to get a 1×1 value across the depth, we are left with a $1024 \times 1 \times 1$ vector.
4. This vector of 1024 values interfaces with the Fully connected layer with a 1000 Neurons.
5. This gives us ~1 Million parameters, as opposed to the earlier seen 50 Million parameters.