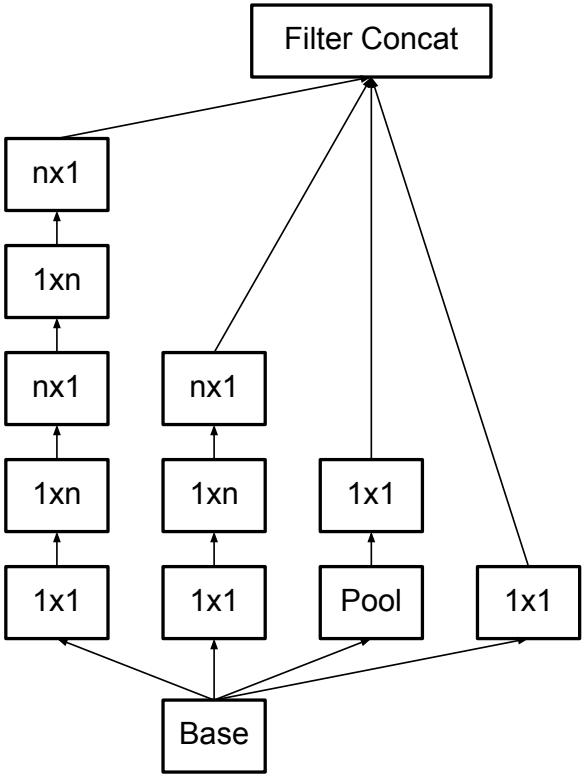


Figure 6. Inception modules after the factorization of the $n \times n$ convolutions. In our proposed architecture, we chose $n = 7$ for the $17 \times 17$ grid. (The f lter sizes are picked using principle 3)

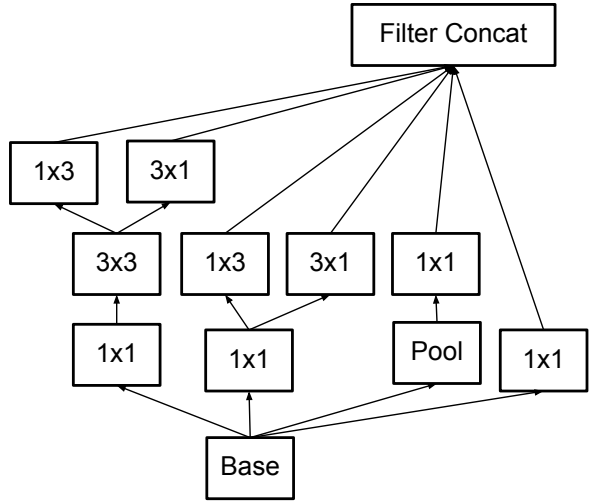


Figure 7. Inception modules with expanded the f lter bank outputs. This architecture is used on the coarsest $(8 \times 8)$ grids to promote high dimensional representations, as suggested by principle 2 of Section 2. We are using this solution only on the coarsest grid, since that is the place where producing high dimensional sparse representation is the most critical as the ratio of local processing (by $1 \times 1$ convolutions) is increased compared to the spatial aggregation.