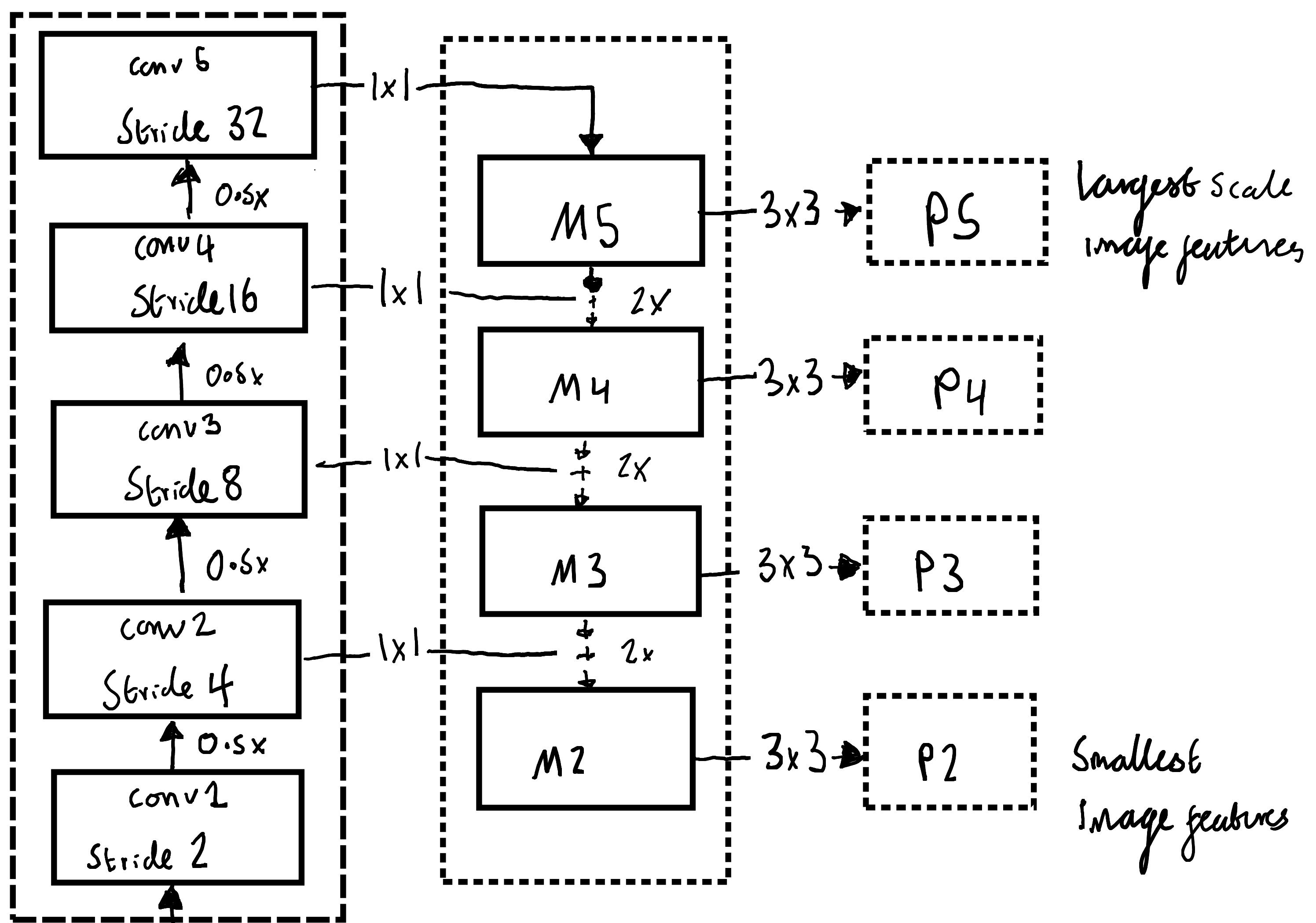
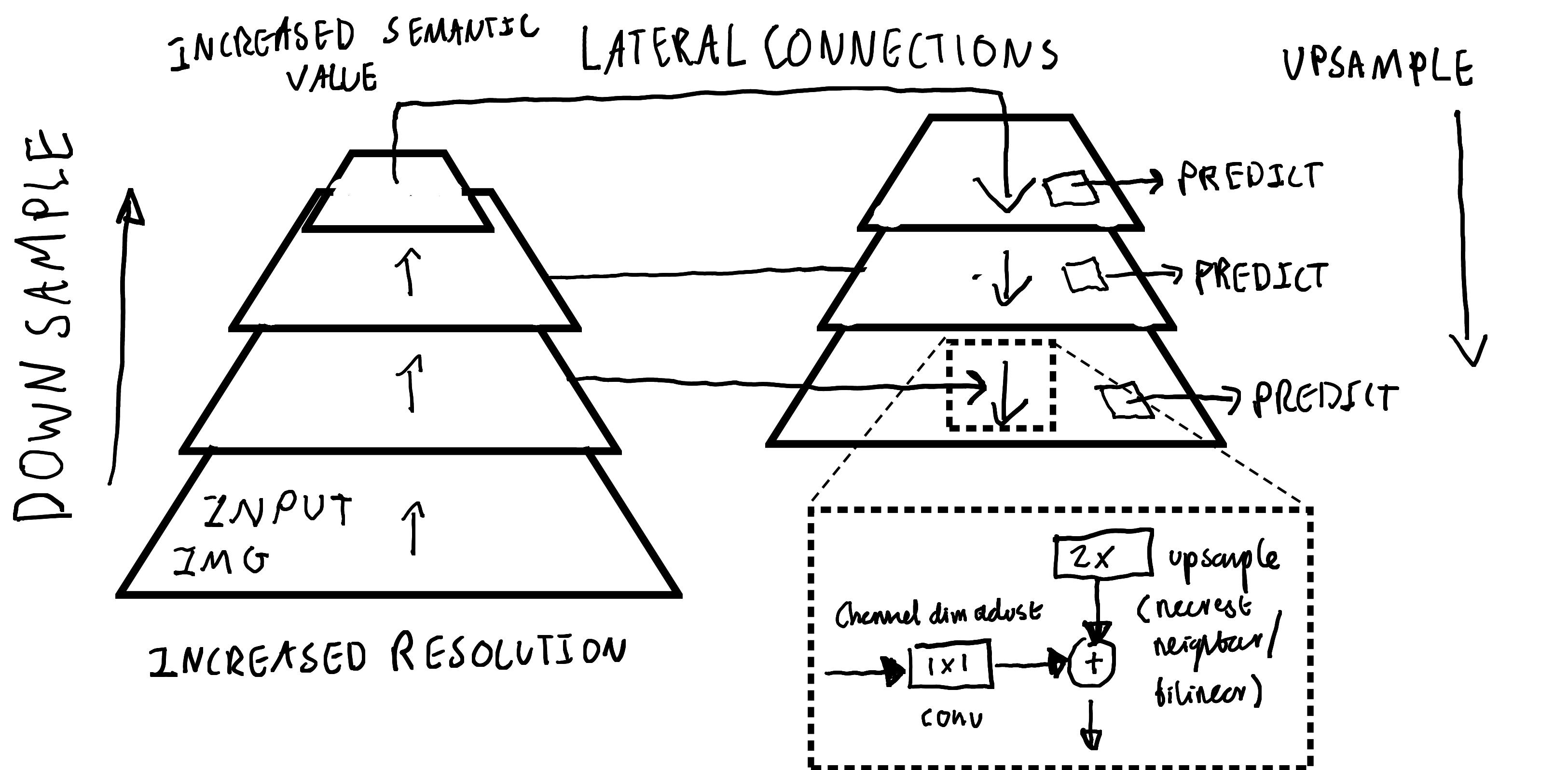
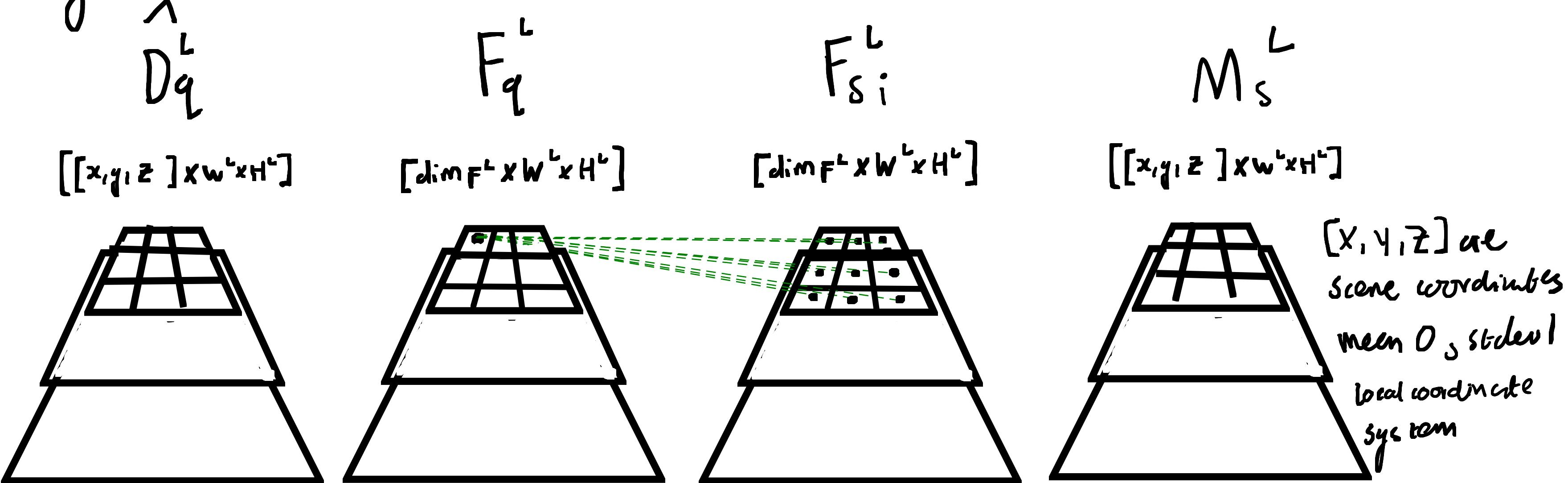


Dense scene matching



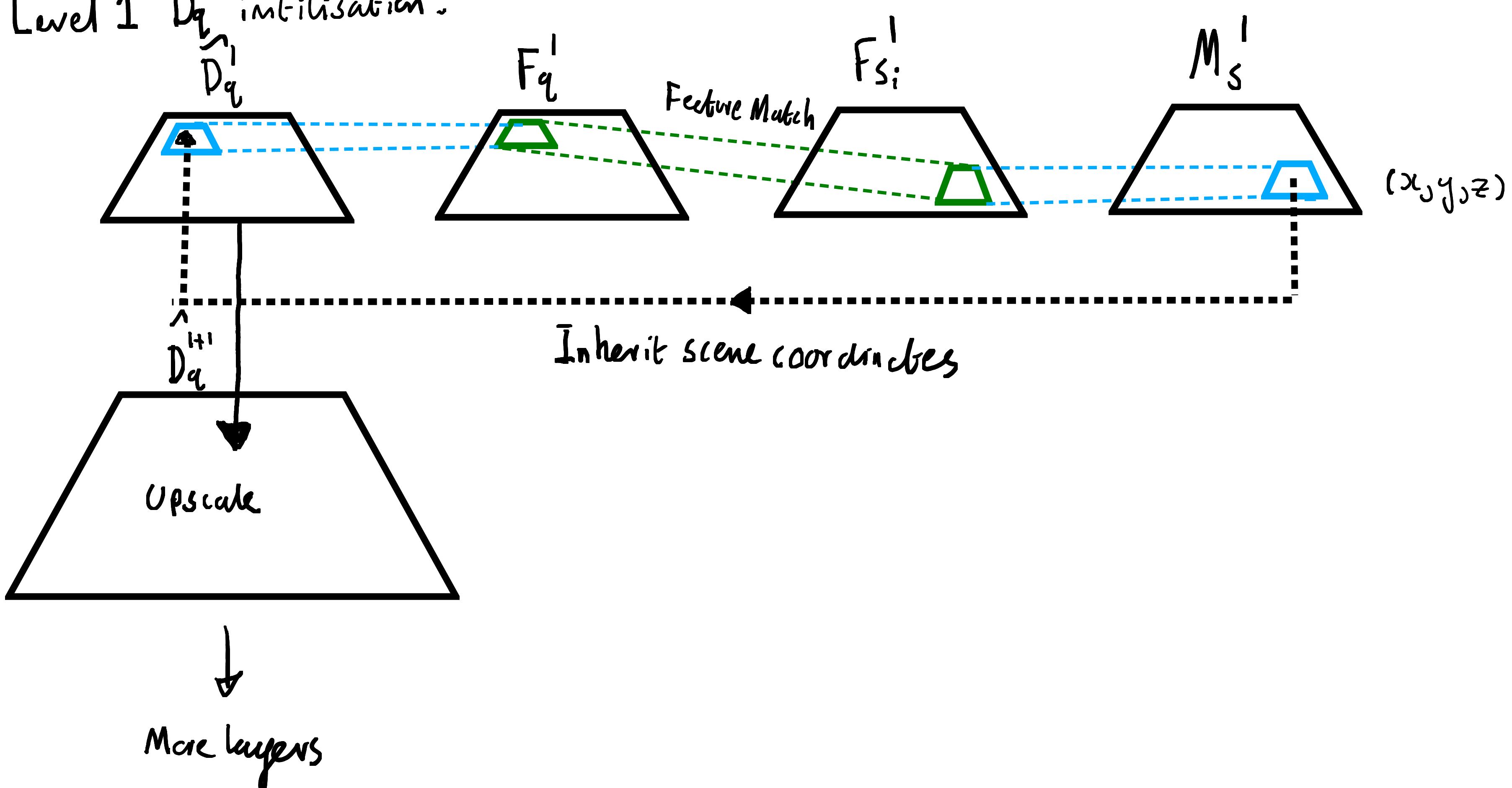
Stride wrt
Input image.

Feature Correlation is defined as the cosine similarity between pixel features



Say we have initial query coordinate map D_q^L , query feature pyramid F_q^L ,
 $\&$ a retrieved scene feature $F_{s_i}^L$ with scene coordinate maps M_s^L at level L .
 Level L has spatial resolution $[w^L \times H^L]$. From the coarsest level $L=1$ we can estimate
 D_q^1 from finding pixel correspondences between $F_q^1 \& F_{s_i}^1$, for every pixel, & extracting
 the coordinates for matching features from $M_s^1 = D_q^1 \cup F_q^1 \cup F_{s_i}^1$ & M_s^1 have the same dimensions for
 $w^L \& H^L$.

Level 1 D_q^1 initialisation:



Once we have our full upscaled \hat{D}^l , for each level l , project 3D points from query pixel q_t^l using \hat{D}_t^l to each scene pixel $p_{s,i}^l$. Then consider an $d \times d$ search window around point $p_{s,i}^l$. Correlation is set to zero if q_t^l is out of the view in scene feature map $F_{s,i}^l$. If there are N scene images then we have $N \times d \times d$ scene coordinate correlations per query pixel, per level.

