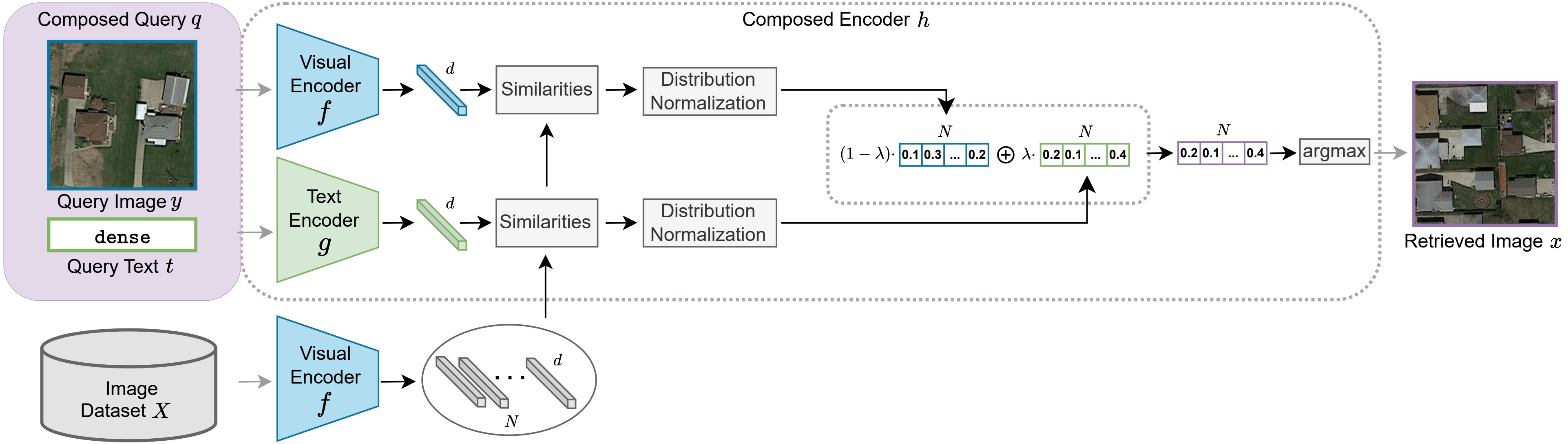
Composed Query q
Query Image y
dense
Query Text t
Composed Encoder h
Visual Encoder f
d
Similarities
Distribution Normalization
Text Encoder g
d
Similarities
Distribution Normalization
N
$(1-\lambda)\cdot$
0.1
0.3
...
0.2
$\oplus$
N
$\lambda\cdot$
0.2
0.1
...
0.4
N
0.2
0.1
...
0.4
argmax
Retrieved Image x
Image Dataset X
Visual Encoder f
d
N