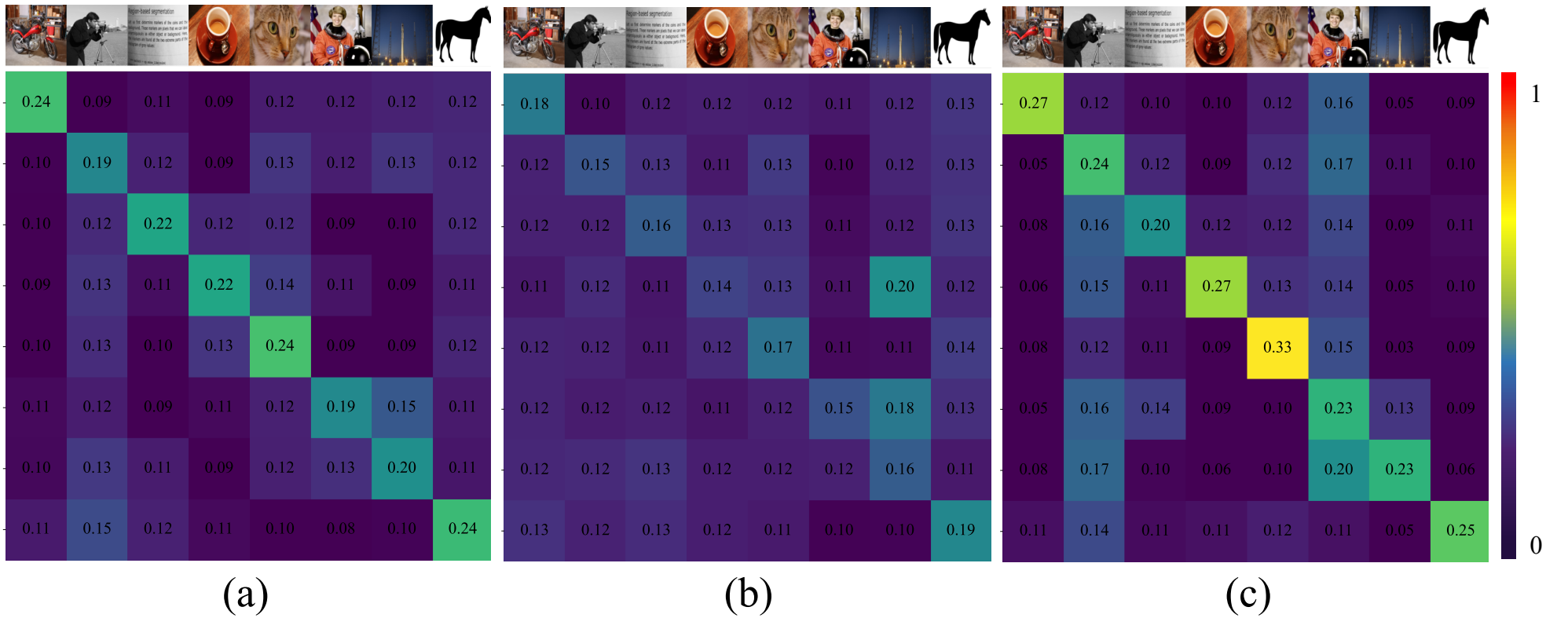

Full precision
PTQ quantization
P4Q quantization
0.24 0.09 0.11 0.09 0.12 0.12 0.12 0.12
0.10 0.19 0.12 0.09 0.13 0.12 0.13 0.12
0.10 0.12 0.22 0.12 0.12 0.09 0.10 0.12
0.09 0.13 0.11 0.22 0.14 0.11 0.09 0.11
0.10 0.13 0.10 0.13 0.24 0.09 0.09 0.12
0.11 0.12 0.09 0.11 0.12 0.19 0.15 0.11
0.10 0.13 0.11 0.09 0.12 0.13 0.20 0.11
0.11 0.15 0.12 0.11 0.10 0.08 0.10 0.24
0.18 0.10 0.12 0.12 0.12 0.11 0.12 0.13
0.12 0.15 0.13 0.11 0.13 0.10 0.12 0.13
0.12 0.12 0.16 0.13 0.13 0.11 0.12 0.13
0.11 0.12 0.11 0.14 0.13 0.11 0.20 0.12
0.12 0.12 0.11 0.12 0.17 0.11 0.11 0.14
0.12 0.12 0.12 0.11 0.12 0.15 0.18 0.13
0.12 0.12 0.13 0.12 0.12 0.12 0.16 0.11
0.13 0.12 0.13 0.12 0.11 0.10 0.10 0.19
0.27 0.12 0.10 0.10 0.12 0.16 0.05 0.09
0.05 0.24 0.12 0.09 0.12 0.17 0.11 0.10
0.08 0.16 0.20 0.12 0.12 0.14 0.09 0.11
0.06 0.15 0.11 0.27 0.13 0.14 0.05 0.10
0.08 0.12 0.11 0.09 0.33 0.15 0.03 0.09
0.05 0.16 0.14 0.09 0.10 0.23 0.13 0.09
0.08 0.17 0.10 0.06 0.10 0.20 0.23 0.06
0.11 0.14 0.11 0.11 0.12 0.11 0.05 0.25
1
0
(a)
(b)
(c)