

Images In The Test Gallery	Image Type
A	Bathroom
B	Bathroom
C	Kitchen
D	Outdoor

Test Case (TC)	Rank 1	Rank 2	Rank 3	Rank 4
1 Trying to retrieve image A	A	C	B	D
2 Trying to retrieve image B	A	B	C	D
3 Trying to retrieve image C	C	B	A	D
4 Trying to retrieve image D	C	B	D	A

			Image A and C were the Best recommendations in Test cases 1 and 3
Recall @ K = 1	Test Type = Fei Fei	2/4	
Recall @ K = 2	Test Type = Fei Fei	3/4	What has changed here? In test case 2, Image B was in the top 2. Hence, Considered a success.
Recall @ K = 1	Test Type = Category Based	3/4	Notice that TC2 returns A Instead of B. Still, because the Gold image has the same category as image B (Bathroom), this is counted as a success.
Recall @ K = 2	Test Type = Category Based	4/6	How do we arrive to this denominator? Notice that TC 3 and 4 only have 1 category for the gold recommendation. To penalize the retrieval of B in TC3 or C in TC4 would be unfair, as the system is forced to propose 2 items. So in TC3, we retrieved 1 out of 1 possible relevant item. In TC4, we retrieved 0 of 1 possible relevant item. In TC1, we retrieved 1 of 2. In TC2, 2 out of 2.