|                  |              | MSCOCO (5K test set)               |      |      |                                |      |      |
|------------------|--------------|------------------------------------|------|------|--------------------------------|------|------|
| Model            | Pre-training | $\mathrm{Image} \to \mathrm{Text}$ |      |      | $\text{Text} \to \text{Image}$ |      |      |
|                  |              | R@1                                | R@5  | R@10 | R@1                            | R@5  | R@10 |
| ViT-B $/32$      | CLIP WIT     | 50.0                               | 75.0 | 83.3 | 30.4                           | 54.8 | 66.1 |
|                  | LAION-400M   | 53.5                               | 77.2 | 85.4 | 34.9                           | 60.3 | 71.1 |
|                  | LAION-2B-en  | 56.4                               | 79.6 | 87.4 | 38.7                           | 64.1 | 74.4 |
| ViT-B $/16$      | CLIP WIT     | 51.7                               | 76.8 | 84.3 | 32.7                           | 57.8 | 68.2 |
|                  | LAION-400M   | 56.5                               | 80.4 | 87.3 | 37.9                           | 63.2 | 73.3 |
| ViT-B $/16+$     | LAION-400M   | 58.6                               | 81.6 | 88.4 | 40.0                           | 65.5 | 75.1 |
| m ViT- $ m L/14$ | CLIP WIT     | 56.0                               | 79.5 | 86.9 | 35.3                           | 60.0 | 70.2 |
|                  | LAION-400M   | 59.3                               | 81.9 | 89.0 | 42.0                           | 67.2 | 76.6 |