

| Method | Compression Rate | Speed-up | Accuracy |
|------------------------|------------------|----------|--------------|
| Sparse Pruning | 4× | None | 91.57 |
| Low-Rank Factorization | 2.4× | 1.02× | 93.44 |
| Knowledge Distillation | 21.3 × | 1.15× | 90.57 |