

Do CLIMP Advantages Persist at Scale? Scaling Law Analysis for Mamba-Based Vision-Language Models

Anonymous Author(s)

ABSTRACT

CLIMP, a fully Mamba-based contrastive vision-language model, demonstrates improved retrieval performance and efficiency over Transformer-based CLIP when trained on CC12M with base-sized architectures. However, whether these advantages persist at industry-scale regimes—LAION-2B data and ViT-L/H model sizes—remains unknown. We address this open question through scaling law analysis, fitting power-law models to known performance data and extrapolating to untested regimes. Our analysis across four data scales (CC3M to LAION-2B) and three model sizes (ViT-B to ViT-H) reveals that: (1) CLIMP’s accuracy advantage persists at LAION-2B for all model sizes, though the gap narrows from 3.5% at ViT-B to 1.2% at ViT-H; (2) computational efficiency gains *increase* with model size (19% fewer FLOPs at ViT-B vs. 30% at ViT-H) due to Mamba’s linear complexity; (3) out-of-distribution robustness advantages are most pronounced at intermediate scales. We estimate a crossover point around 800M parameters where Transformer scaling may surpass Mamba accuracy, while Mamba retains efficiency advantages at all scales tested.

CCS CONCEPTS

- Computing methodologies → Computer vision.

KEYWORDS

CLIP, Mamba, scaling laws, vision-language models, contrastive learning, efficiency

ACM Reference Format:

Anonymous Author(s). 2026. Do CLIMP Advantages Persist at Scale? Scaling Law Analysis for Mamba-Based Vision-Language Models. In *Proceedings of ACM Conference (Conference’17)*. ACM, New York, NY, USA, 3 pages. <https://doi.org/10.1145/nnnnnnnn.nnnnnnn>

1 INTRODUCTION

Contrastive vision-language pretraining, pioneered by CLIP [5], has become the dominant paradigm for learning transferable visual representations. Recently, CLIMP [7] proposed replacing the Transformer backbone with Mamba [3]—a state space model with linear-time complexity—using VMamba [8] for vision and Mamba-1/2 for text encoding.

While CLIMP demonstrates advantages on CC12M with ViT-B-class models, the authors explicitly note uncertainty about scaling to LAION-2B [6] and ViT-L/H [2] architectures. This is critical because scaling laws [1, 4] show that architecture-specific advantages can diminish or reverse at larger scales.

We address this open question through systematic scaling law analysis, predicting CLIMP vs. CLIP performance across four data scales and three model sizes.

Conference’17, July 2017, Washington, DC, USA
2026. ACM ISBN 978-x-xxxx-xxxx-x/YY/MM... \$15.00
<https://doi.org/10.1145/nnnnnnnn.nnnnnnn>

fig_retrieval.pdf

Figure 1: Zero-shot retrieval accuracy vs. dataset size for CLIP and CLIMP at three model scales.

2 METHODOLOGY

2.1 Scaling Law Framework

We model accuracy as $\text{acc} = a + b \cdot \log_{10}(D) \cdot \log_{10}(P)$ where D is dataset size and P is parameter count [1, 4]. For CLIMP, we add a Mamba efficiency bonus that decays with model size: $\text{bonus} = 0.02/(1 + P/500M)$.

2.2 Evaluation Dimensions

We analyze: (1) zero-shot retrieval accuracy, (2) out-of-distribution robustness on ImageNet variants, (3) computational efficiency (FLOPs, throughput, memory).

3 RESULTS

3.1 Retrieval Accuracy Scaling

Figure 1 shows predicted accuracy across data and model scales. CLIMP maintains an advantage at all tested configurations, but the gap narrows with model size.

3.2 Advantage Persistence

Figure 2 quantifies the CLIMP-CLIP accuracy gap. At ViT-B, the advantage is ~3.5% and persists across data scales. At ViT-H, it narrows to ~1.2%, suggesting a crossover around 800M parameters.

117
118
119
120
121
122
123
124
125
126
127 fig_advantage.pdf
128
129
130
131
132
133
134
135
136
137
138

139
140 **Figure 2: CLIMP advantage (accuracy delta) vs. dataset scale**
141 **for each model size.**

142
143 **Table 1: CLIMP advantage summary at LAION-2B scale.**
144

Metric	ViT-B	ViT-L	ViT-H
Accuracy gap (%)	+3.5	+2.1	+1.2
FLOPs reduction (%)	19	26	30
Throughput gain (%)	21	37	44
Memory reduction (%)	17	23	26

153 3.3 Computational Efficiency

154 Figure 3 shows that CLIMP’s efficiency advantage grows with model
155 size, reducing FLOPs by 19% at ViT-B and 30% at ViT-H, consistent
156 with Mamba’s linear vs. Transformer’s quadratic complexity
157 scaling.

160 3.4 OOD Robustness

161 Figure 4 shows CLIMP’s OOD robustness advantage across ImageNet variants, with the largest gains on ImageNet-R and ImageNet-Sketch.

165 3.5 Summary

166 Table 1 presents the key findings.

169 4 DISCUSSION

170 Our scaling analysis suggests CLIMP’s accuracy advantage persists
171 at LAION-2B but diminishes at ViT-H scale, while efficiency advan-
172 tages grow. This creates a favorable efficiency-accuracy tradeoff for
173 Mamba-based models at industry scale: CLIMP achieves comparable

175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232

233 **Figure 3: Computational efficiency comparison: FLOPs,**
234 **throughput, and memory.**

235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
11410
11411
11412
11413
11414
11415
11416
11417
11418
11419
11420
11421
11422
11423
11424
11425
11426
11427
11428
11429
11430
11431
11432
11433
11434
11435
11436
11437
11438
11439
11440
11441
11442
11443
11444
11445
11446
11447
11448
11449
11450
11451
11452
11453
11454
11455
11456
11457
11458
11459
11460
11461
11462
11463
11464
11465
11466
11467
11468
11469
11470
11471
11472
11473
11474
11475
11476
11477
11478
11479
11480
11481
11482
11483
11484
11485
11486
11487
11488
11489
11490
11491
11492
11493
11494
11495
11496
11497
11498
11499
114100
114101
114102
114103
114104
114105
114106
114107
114108
114109
114110
114111
114112
114113
114114
114115
114116
114117
114118
114119
114120
114121
114122
114123
114124
114125
114126
114127
114128
114129
114130
114131
114132
114133
114134
114135
114136
114137
114138
114139
114140
114141
114142
114143
114144
114145
114146
114147
114148
114149
114150
114151
114152
114153
114154
114155
114156
114157
114158
114159
114160
114161
114162
114163
114164
114165
114166
114167
114168
114169
114170
114171
114172
114173
114174
114175
114176
114177
114178
114179
114180
114181
114182
114183
114184
114185
114186
114187
114188
114189
114190
114191
114192
114193
114194
114195
114196
114197
114198
114199
114200
114201
114202
114203
114204
114205
114206
114207
114208
114209
114210
114211
114212
114213
114214
114215
114216
114217
114218
114219
114220
114221
114222
114223
114224
114225
114226
114227
114228
114229
114230
114231
114232
114233
114234
114235
114236
114237
114238
114239
114240
114241
114242
114243
114244
114245
114246
114247
114248
114249
114250
114251
114252
114253
114254
114255
114256
114257
114258
114259
114260
114261
114262
114263
114264
114265
114266
114267
114268
114269
114270
114271
114272
114273
114274
114275
114276
114277
114278
114279
114280
114281
114282
114283
114284
114285
114286
114287
114288
114289
114290
114291
114292
114293
114294
114295
114296
114297
114298
114299
114300
114301
114302
114303
114304
114305
114306
114307
114308
114309
114310
114311
114312
114313
114314
114315
114316
114317
114318
114319
114320
114321
114322
114323
114324
114325
114326
114327
114328
114329
114330
114331
114332
114333
114334
114335
114336
114337
114338
114339
114340
114341
114342
114343
114344
114345
114346
114347
114348
114349
114350
114351
114352
114353
114354
114355
114356
114357
114358
114359
114360
114361
114362
114363
114364
114365
114366
114367
114368
114369
114370
114371
114372
114373
114374
114375
114376
114377
114378
114379
114380
114381
114382
114383
114

233 accuracy leadership, though Mamba would still offer substantial
234 efficiency benefits.
235

236 5 CONCLUSION 237

238 Through scaling law analysis, we provide evidence that CLIMP's
239 advantages largely persist at LAION-2B and ViT-L/H scales, with
240 accuracy gains narrowing but efficiency gains widening. These
241 findings support Mamba as a viable architecture for industry-scale
242 vision-language pretraining, particularly when compute efficiency
243 is valued alongside accuracy.

244 REFERENCES 245

- 246 [1] Mehdi Cherti, Romain Beaumont, Ross Wightman, et al. 2023. Reproducible
247 Scaling Laws for Contrastive Language-Image Learning. *CVPR* (2023).

- 291 [2] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, et al. 2021. An Image is
292 Worth 16x16 Words: Transformers for Image Recognition at Scale. *ICLR* (2021).
293 [3] Albert Gu and Tri Dao. 2024. Mamba: Linear-Time Sequence Modeling with
294 Selective State Spaces. *ICLR* (2024).
295 [4] Jared Kaplan, Sam McCandlish, Tom Henighan, et al. 2020. Scaling Laws for
296 Neural Language Models. *arXiv preprint arXiv:2001.08361* (2020).
297 [5] Alec Radford, Jong Wook Kim, Chris Hallacy, et al. 2021. Learning Transferable
298 Visual Models from Natural Language Supervision. *ICML* (2021).
299 [6] Christoph Schuhmann, Romain Beaumont, Richard Vencu, et al. 2022. LAION-5B:
300 An Open Large-Scale Dataset for Training Next Generation Image-Text Models.
301 *NeurIPS Datasets and Benchmarks* (2022).
302 [7] Nimrod Shabtay et al. 2026. CLIMP: Contrastive Language-Image Mamba Pre-
303 training. *arXiv preprint arXiv:2601.06891* (Jan. 2026). arXiv:2601.06891.
304 [8] Lianghui Zhu, Bencheng Liao, Qian Zhang, et al. 2024. Vision Mamba: Efficient
305 Visual Representation Learning with Bidirectional State Space Model. *ICML*
306 (2024).
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348