

A Compact Embedding for Facial Expression Similarity - Supplementary

Anonymous CVPR submission

Paper ID 2958

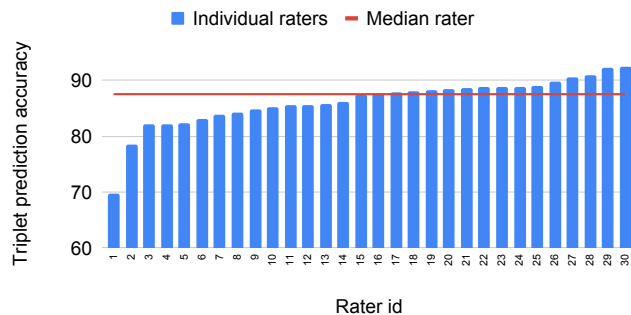


Figure 1: Triplet prediction accuracy of individual raters on the FEC test set (blue bars). The red line shows the median accuracy.

1. Triplet prediction accuracy of human raters

To estimate how good humans are at identifying the most similar pair in a triplet, we computed the accuracy values for individual raters based on how often they agree with the maximum-voted label. Figure 1 shows the accuracy values for all the 30 raters who contributed to the FEC test set annotations. The mean and median values are 86.2% and 87.5%.

2. Visualization of FECNet embeddings space

Figure 2 shows a 2D t-SNE [1] visualization of the learned FECNet-16d embedding space using the AffectNet [3] validation set. The color of the bounding box of a face represents the emotion label of that face. Please zoom in to see the face images clearly.

3. Photo album summarization

We created ten photo albums (100-200 images in each album) by downloading images of ten celebrities using Google image search, and generated two visual summaries for each album by clustering the FECNet-16d and AFFNet-CL-P embeddings, respectively. We showed these two summaries to ten human raters and asked them which one is better. For each decision, the winner and loser get a score of 1 and 0, respectively. If it is a tie, then both of them

Album	KM	DT	CB	GB	BO
FECNet-16d vs AFFNet-CL-P	10-0	5-1	9-1	9-0	5-2
Album	JL	LS	HC	LJ	JC
FECNet-16d vs AFFNet-CL-P	9-0	1-6	10-0	1-4	7-1

Table 1: Scores received by the FECNet-16d and AFFNet-CL-P embeddings in the photo album summarization task.

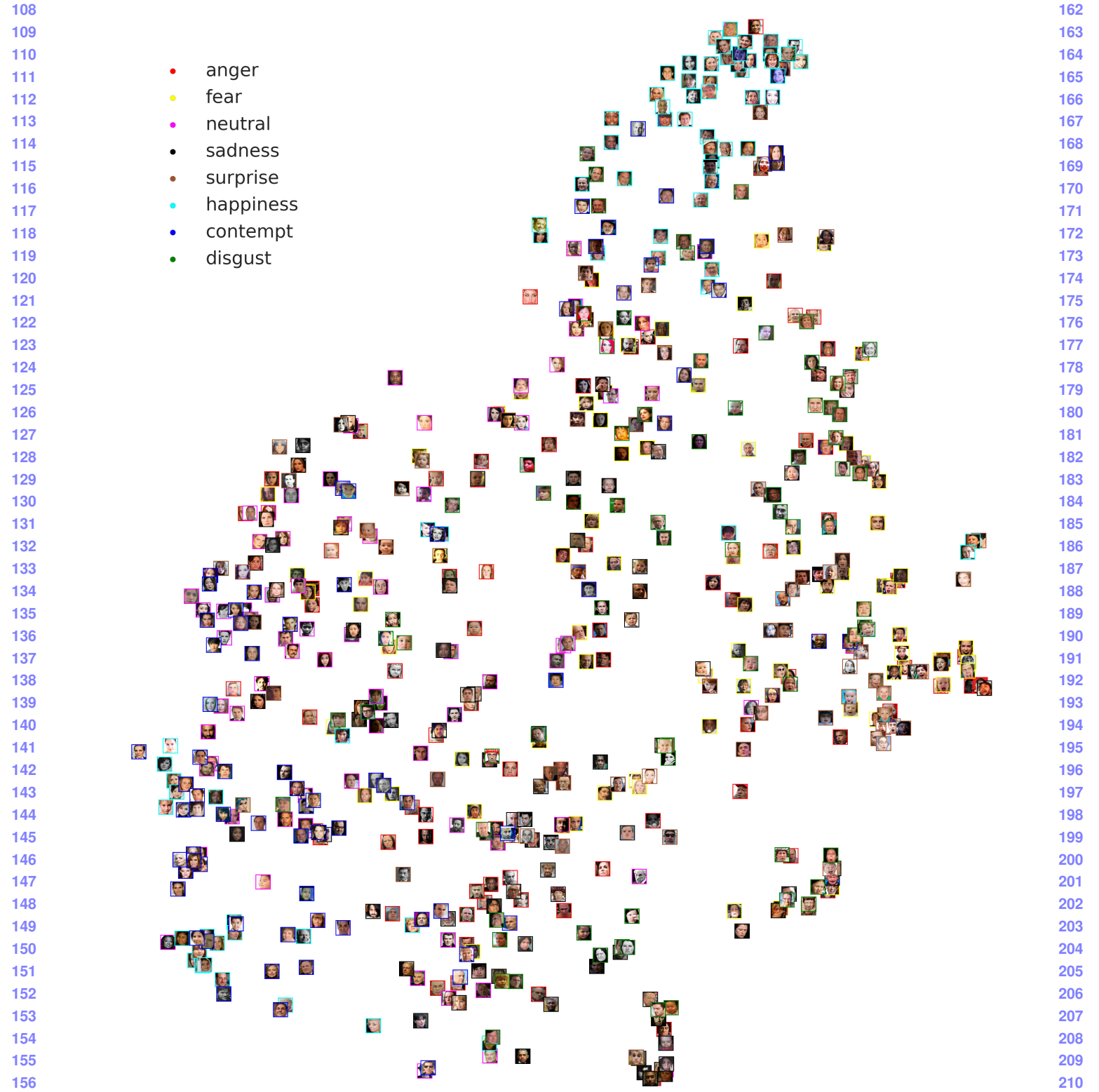
get a score of 0. Table 1 shows the overall scores for the FECNet-16d and AFFNet-CL-P embeddings for all ten albums. As we can see, human prefer the summaries generated by the proposed FECNet-16d embedding for eight out of ten albums. Figures 3 and 4 show the summary results for all ten albums. For most albums, the expression content is more diverse in the summaries generated by the proposed FECNet-16d embedding.

4. Expression retrieval

Figures 5 to 9 show the expression retrieval results (top-10 images from CelebA [2] dataset) of the FECNet-16d and AFFNet-CL-P embeddings for all the 25 expressions in the query set. In each figure, the images on the left are the query expressions. For each query, the top and bottom rows correspond to the FECNet-16d and AFFNet-CL-P embeddings, respectively. We can see that the proposed FECNet-16d embedding produces better matches when compared to the AFFNet-CL-P embedding.

References

- [1] Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9:2579–2605, 2008. 1, 2
- [2] Z. Liu, P. Luo, X. Wang, and X. Tang. Deep learning face attributes in the wild. In *ICCV*, 2015. 1
- [3] A. Mollahosseini, B. Hassani, and M. H. Mahoor. AffectNet: A database for facial expression, valence, and arousal computing in the wild. *CoRR*, abs/1708.03985, 2017. 1



108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161

Figure 2: 2D visualization of the learned FECNet-16d embedding space using t-SNE [1]. The color of the bounding box of a face represents the emotion label of that face. Please zoom in to see the face images clearly.

162
163
164
165
166
167
168
169
170
171
172
173
174
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



Kate McKinnon (KM)



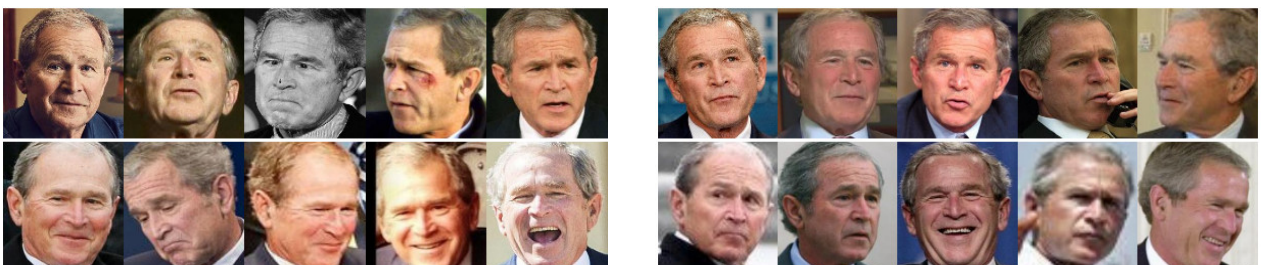
Jim Carrey (JC)



Jennifer Lawrence (JL)



Cardi B (CB)



George Bush (GB)

Figure 3: Summaries generated by the FECNet-16d (left) and AFFNet-CL-P (right) embeddings.

216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
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



Hillary Clinton (HC)



Barack Obama (BO)



Lilly Singh (LS)



Lebron James (LJ)



Donald Trump (DT)

Figure 4: Summaries generated by the FECNet-16d (left) and AFFNet-CL-P (right) embeddings.

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



Figure 5: Expression retrieval results - Images on the left are the query expressions. For each query, the top row corresponds to the FECNet-16d embedding and the bottom row corresponds to the AFFNet-CL-P embedding.

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



Figure 6: Expression retrieval results - Images on the left are the query expressions. For each query, the top row corresponds to the FECNet-16d embedding and the bottom row corresponds to the AFFNet-CL-P embedding.



Figure 7: Expression retrieval results - Images on the left are the query expressions. For each query, the top row corresponds to the FECNet-16d embedding and the bottom row corresponds to the AFFNet-CL-P embedding.



756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809

810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863

Figure 8: Expression retrieval results - Images on the left are the query expressions. For each query, the top row corresponds to the FECNet-16d embedding and the bottom row corresponds to the AFFNet-CL-P embedding.

864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917

918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971



Figure 9: Expression retrieval results - Images on the left are the query expressions. For each query, the top row corresponds to the FECNet-16d embedding and the bottom row corresponds to the AFFNet-CL-P embedding.