

000
001
002
003
004
005
006
007
008
009
010
011
012
013
014
015

Abstract

In this supplementary document, we first provide more details about our proposed novel approach and related analysis, and then present more numerical and visual comparison results across multiple datasets. Specifically, We offer further analysis of annotation inconsistency in Sec. 1, pose distribution illustration of datasets in Sec. 2, visual show of projection-prediction distance in Sec. 3, more quantitative evaluations in Sec. 4, and comprehensive qualitative evaluations in Sec. 5 on all four datasets against all aforementioned approaches.

1. Annotation Inconsistency

As we explained in our main draft, the landmark annotation inconsistency issue is inevitable, though each annotation may seem reasonable to the given image. DAD-3DHeads [6] incorporates the FLAME fitting method to help annotation. However, it still suffers from this annotation inconsistency problem. To illustrate this, we run the procedure as shown in Fig. 1. Given an image with the annotated landmark, we first obtain a fitted mesh through the FLAME fitting, and then project the mesh to another view based on the corresponding camera parameters. Finally, we can extract landmarks from the projected mesh for the new view. As shown in the zoom-in inset of Fig. 1, the projected landmark does not fit the image. For instance, the points of the mouth area indicate the multiview inconsistency caused by annotation inconsistency. Motivated by this observation, we propose to train facial landmark detectors via multiview consistent synthetic data.

2. Dataset Distribution

In addition to the multiview consistency, another benefit of our synthetic dataset is the removal of pose distribution bias in training data. General datasets, such as DAD-3DHeads [6], are biased in the small range of head pose distribution. As shown in Fig. 2, the histograms of pitch and

054
055
056
057
058
059
060
061
062
063
064
065
066
067
068
069
070
071
072
073
074
075
076
077
078
079
080
081
082
083
084
085
086
087
088
089
090
091
092
093
094
095
096
097
098
099
100
101
102
103
104
105
106
107

Anonymous CVPR submission

Paper ID 4794

yaw angles of head pose in DAD-3DHeads [6] indicate approximate normal distributions with means at around zero degrees. In contrast, the pose distributions of our synthetic dataset are much more balanced across the full range, thus, help the model generate better estimations.

3. Projection-Prediction Distance

With the ready of the multiview consistent synthetic dataset of well-balanced pose distribution, we propose to incorporate multiview consistency into landmark detection by minimizing the distance of predicted landmark and projected landmark of the given image. As shown in Fig. 3, predictions and projections are denoted as green points and white points respectively, between which are the distances shown in red lines. DAD-3DNet+(Ours) generates much more consistent results indicated by shorter red lines in Fig. 3. Next, we will provide more numerical and visual results to demonstrate the superiority of our approach.

4. More Quantitative Evaluations

Table 1. Facial landmark detection result (NME) on DAD-3DHeads-Syn. Lower values mean better results.

Method	DAD-3DHeads-Syn
FAN [2]	2.826
Dlib [5]	3.023
3DDFA-V2 [4]	2.840
3DDFA [3]	3.174
3DDFA+	2.981
DAD-3DNet [6]	2.373
DAD-3DNet+	2.211

In this section, we will first provide the landmark detection comparison on DAD-3DHeads-Syn covering all algorithms shown previously in the main paper. We further present pose estimation comparisons on both DAD-3DHeads-Syn and MultiFace [7] across the same existing

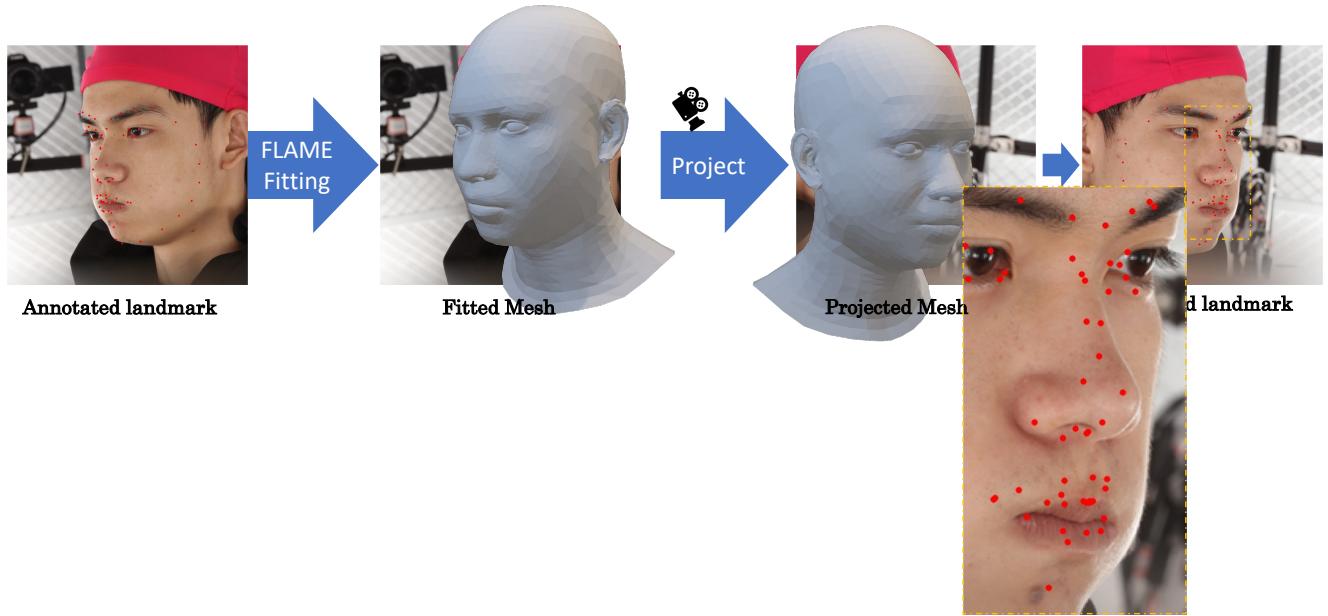
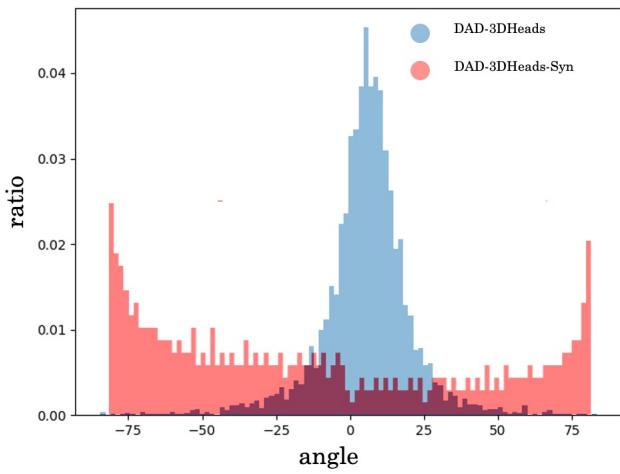


Figure 1. Multiview inconsistency caused by landmark annotation inconsistency.

Pitch Distribution



Yaw Distribution

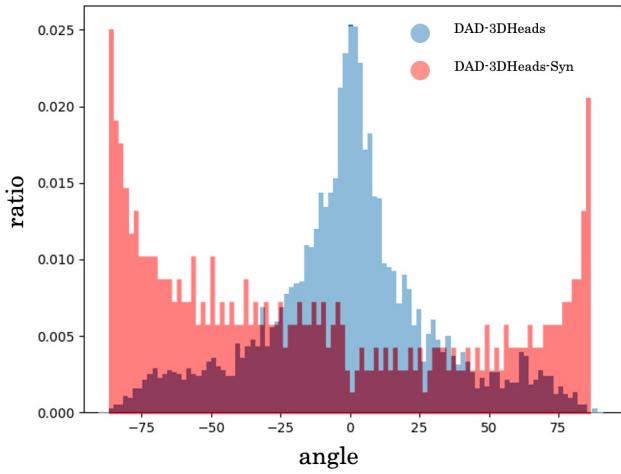


Figure 2. Pose distributions of datasets. Our DAD-3DHeads-Syn is much more balanced in pose distributions.

methods.

4.1. Landmark Detection Results

In the main paper, we have shown the quantitative results of landmark detection on DAD-3DHeads [6], FaceScape [8], MultiFace [7]. Here, we provide an extra test on our DAD-3DHeads-Syn dataset. As shown in Tab. 1, the algorithms incorporated with our plug-in module, 3DDFA+(Ours) and DAD-3DNet+(Ours), generate much better numerical results than their base models. Both of them make important improvements in the NME of

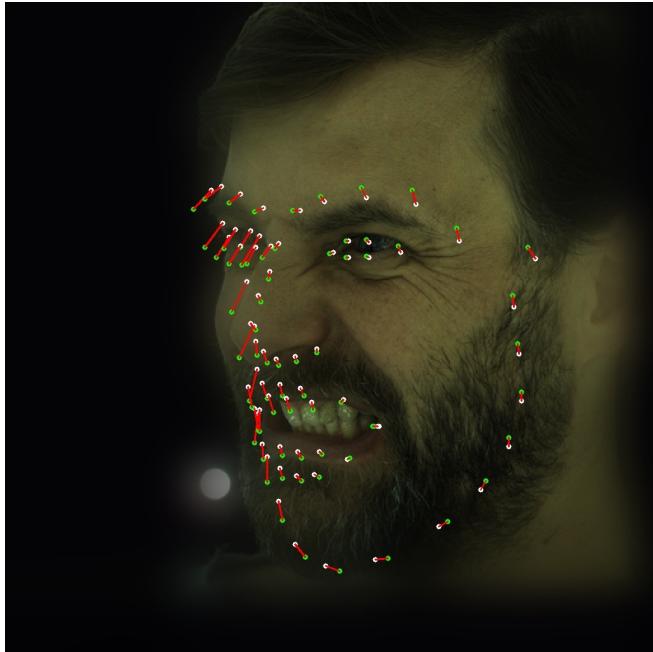
landmark accuracy.

4.2. Pose Estimation Results

Table 2 shows the pose estimation on DAD-3DHeads-Syn, and MultiFace [7], DAD-3DNet+(Ours) achieves 24.4%¹ and 19.0%² on DAD-3DHeads-Syn, and MultiFace [7] respectively. Also, the plug-in module improves significantly on almost all of metrics across pitch, roll, and yaw angles, except yaw estimation on DAD-3DHeads-Syn.

¹Head pose error drops from 7.412 to 5.958.

²Head pose error drops 14.962 to 12.578.

216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236

DAD-3DNet



DAD-3DNet+

237
238
239

Figure 3. Comparisons between DAD-3DNet and DAD-3DNet+(Ours) on projection-prediction distance. Predictions and projections are denoted as green points and white points respectively, between which are the distances shown in red lines.

242
243

Table 2. Head pose estimation results (head pose error) on DAD-3DHeads-Syn, and MultiFace [7]. Lower values mean better results.

	DAD-3DHeads-Syn				MultiFace [7]			
	Pitch	Roll	Yaw	Overall	Pitch	Roll	Yaw	Overall
FAN [1]	21.938	13.093	17.002	17.344	16.840	5.913	21.074	14.609
Dlib [5]	14.525	11.472	8.272	11.430	23.506	4.303	11.093	12.966
3DDFA-V2 [4]	24.428	9.133	19.791	17.784	20.607	8.751	17.418	15.592
3DDFA [3]	24.418	9.364	19.750	17.834	29.059	12.077	17.382	19.506
3DDFA+	22.841	9.008	18.321	16.723	28.086	10.260	16.292	18.213
DAD-3DNet [6]	8.440	11.822	2.183	7.412	23.477	7.285	14.123	14.962
DAD-3DNet+	6.348	8.914	2.613	5.958	21.019	5.808	11.906	12.578

244
245
246
247
248
249
250
251
252
253

5. More Qualitative Evaluations

258
259
260
261
262
263
264
265
266
267
268
269

In this section, we provide even more extensive visual comparisons against aforementioned approaches, FAN [1], Dlib [5], 3DDFA [3], 3DDFA-V2 [4], and DAD-3DNet [6], covering all the four datasets, DAD-3DHeads-Syn, DAD-3DHeads [6], FaceScape [8], and MultiFace [7]. Specifically, we provide five pairs of comparisons, 3DDFA+(Ours) vs. 3DDFA [3], DAD-3DNet+(Ours) vs. Dlib [5], DAD-3DNet+(Ours) vs. FAN [1], DAD-3DNet+(Ours) vs. 3DDFA-V2 [4], and DAD-3DNet+(Ours) vs. DAD-3DNet [6]. As shown in Fig. 4, Fig. 5, Fig. 6, Fig. 7, Fig. 8, Fig. 9, Fig. 10, Fig. 11, Fig. 12, Fig. 13, Fig. 14, Fig. 15, Fig. 16, Fig. 17, Fig. 18, Fig. 19, Fig. 20, Fig. 21, Fig. 22,

and Fig. 23, methods with our plug-in module always show higher quality of landmark detection.

References

- [1] Adrian Bulat and Georgios Tzimiropoulos. Binarized convolutional landmark localizers for human pose estimation and face alignment with limited resources. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3706–3714, 2017.
- [2] Adrian Bulat and Georgios Tzimiropoulos. How far are we from solving the 2d & 3d face alignment problem?(and a dataset of 230,000 3d facial landmarks). In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1021–1030, 2017.

270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290291
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

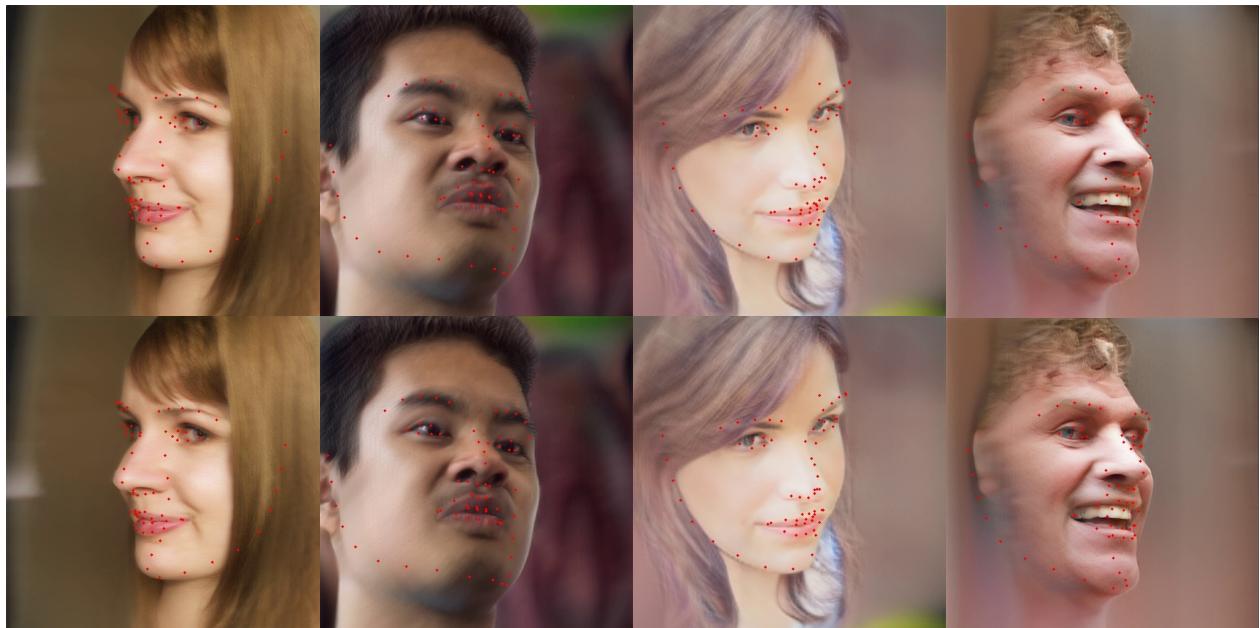
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
3433DDFA
3DDFA+

Figure 4. Comparisons between 3DDFA [3] and 3DDFA+(Ours) on DAD-3DHeads-Syn.

344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366Dlib
DAD-3DNet+

Figure 5. Comparisons between Dlib [5] and DAD-3DNet+(Ours) on DAD-3DHeads-Syn.

367
368
369[3] Jianzhu Guo, Xiangyu Zhu, and Zhen Lei. 3ddfa. <https://github.com/cleardusk/3DDFA>, 2018.370
371
372
373
374
375[4] Jianzhu Guo, Xiangyu Zhu, Yang Yang, Fan Yang, Zhen Lei, and Stan Z Li. Towards fast, accurate and stable 3d dense face alignment. In *European Conference on Computer Vision*, pages 152–168. Springer, 2020.376
377[5] Davis E. King. Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research*, 10:1755–1758, 2009.[6] Tetiana Martyniuk, Orest Kupyn, Yana Kurlyak, Igor Krasheniy, Jiří Matas, and Viktoriia Sharmanśka. Dad-3heads: A large-scale dense, accurate and diverse dataset for 3d head alignment from a single image. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 20942–20952, 2022.

[7] Cheng-hsin Wuu, Ningyuan Zheng, Scott Ardisson, Rohan Bali, Danielle Belko, Eric Brockmeyer, Lucas Evans, Timo-

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
420421
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

Figure 6. Comparisons between FAN [1] and DAD-3DNet+(Ours) on DAD-3DHeads-Syn.

458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477

Figure 7. Comparisons between 3DDFA-V2 [4] and DAD-3DNet+(Ours) on DAD-3DHeads-Syn.

thy Godisart, Hyowon Ha, Alexander Hypes, Taylor Koska, Steven Krenn, Stephen Lombardi, Xiaomin Luo, Kevyn McPhail, Laura Millerschoen, Michal Perdoch, Mark Pitts, Alexander Richard, Jason Saragih, Junko Saragih, Takaaki Shiratori, Tomas Simon, Matt Stewart, Autumn Trimble, Xinshuo Weng, David Whitewolf, Chenglei Wu, Shou Yu, and Yaser Sheikh. Multiface: A dataset for neural face rendering. In *arXiv*, 2022.

- [8] Haotian Yang, Hao Zhu, Yanru Wang, Mingkai Huang, Qiu Shen, Ruigang Yang, and Xun Cao. Facescape: a large-scale high quality 3d face dataset and detailed riggable 3d face prediction. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 601–610, 2020.

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

DAD-3DNet



Figure 8. Comparisons between DAD-3DNet [6] and DAD-3DNet+(Ours) on DAD-3DHeads-Syn.

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

3DDFA+



Figure 9. Comparisons between 3DDFA [3] and 3DDFA+(Ours) on DAD-3DHeads [6].

648
649
650702
703
704651
652
653
654
655
656
657705
706
707
708
709
710
711658
659
660
661
662
663
664712
713
714
715
716
717
718665
666
667
668
669
670719
720
721
722
723
724

Dlib

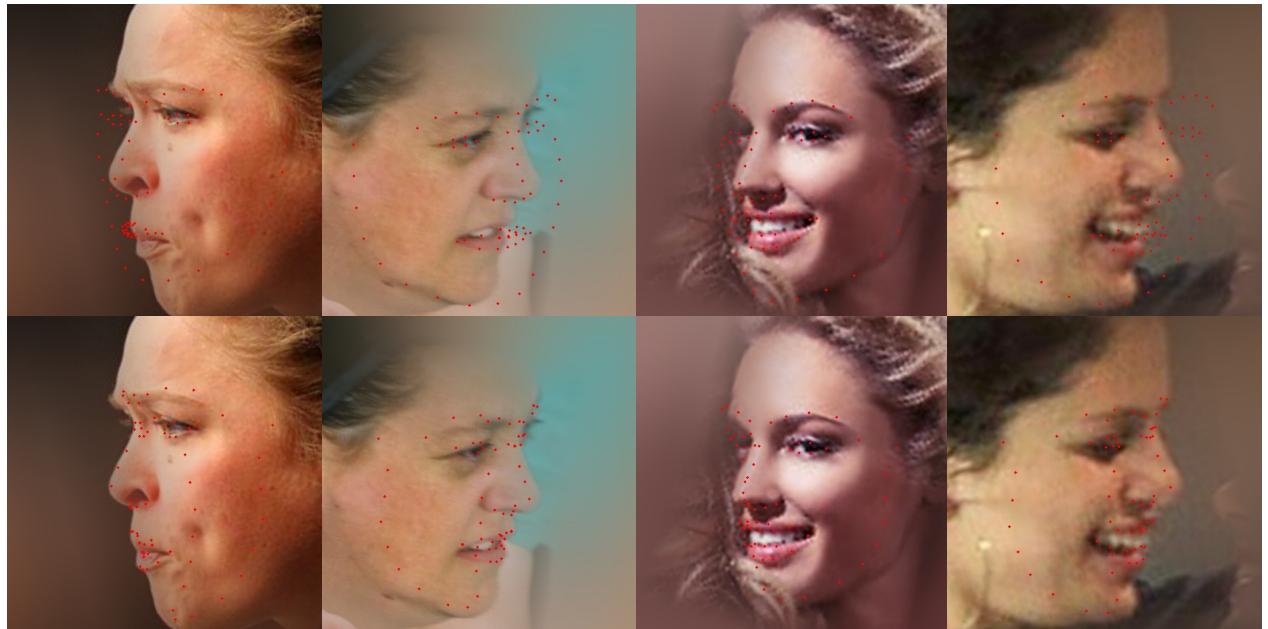


Figure 10. Comparisons between Dlib [5] and DAD-3DNet+(Ours) on DAD-3DHeads [6].

671

725
726672
673
674
675
676
677727
728
729
730
731678
679
680
681
682
683
684
685
686
687732
733
734
735
736
737
738
739
740

FAN



Figure 11. Comparisons between FAN [1] and DAD-3DNet+(Ours) on DAD-3DHeads [6].

698

752
753

699

754

700

755

701

756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778

3DDFA-V2

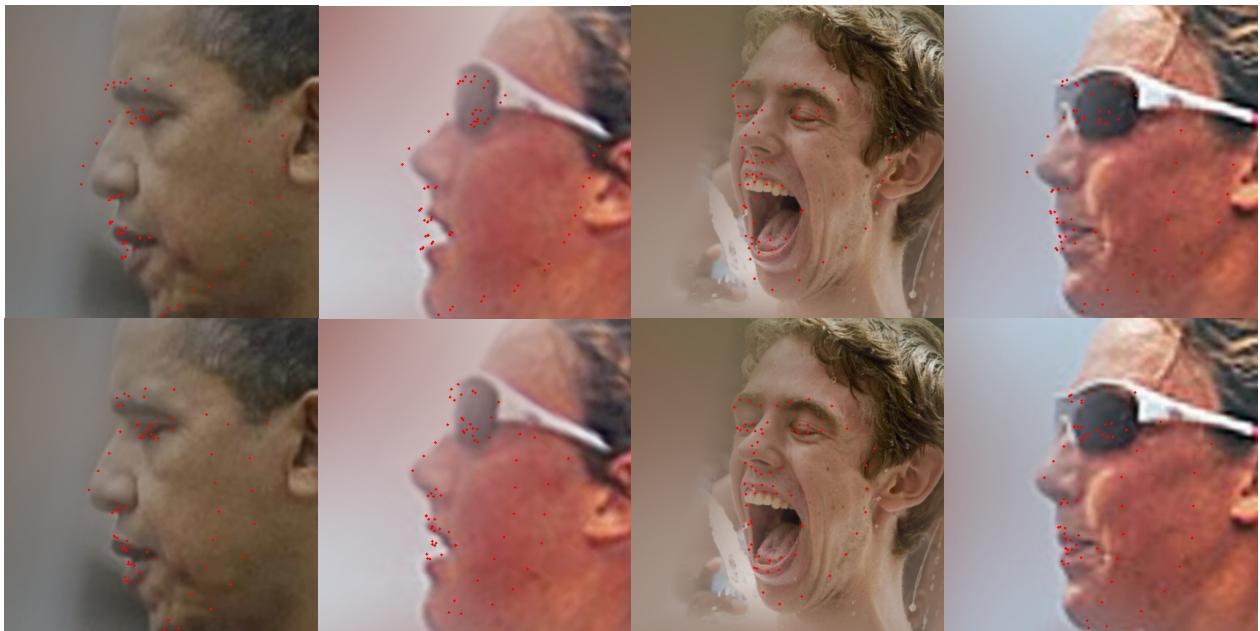


Figure 12. Comparisons between 3DDFA-V2 [4] and DAD-3DNet+(Ours) on DAD-3DHeads [6].

779
780

DAD-3DNet



Figure 13. Comparisons between DAD-3DNet [6] and DAD-3DNet+(Ours) on DAD-3DHeads [6].

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

DAD-3DNet+

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 14. Comparisons between 3DDFA [3] and 3DDFA+(Ours) on FaceScape [8].



Figure 15. Comparisons between Dlib [5] and DAD-3DNet+(Ours) on FaceScape [8].

972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994FAN
DAD-3DNet+

Figure 16. Comparisons between FAN [1] and DAD-3DNet+(Ours) on FaceScape [8].

995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
10213DDFA-V2
DAD-3DNet+

Figure 17. Comparisons between 3DDFA-V2 [4] and DAD-3DNet+(Ours) on FaceScape [8].

1022
1023
1024
10251026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
10481049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075

1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102

DAD-3DNet



Figure 18. Comparisons between DAD-3DNet [6] and DAD-3DNet+(Ours) on FaceScape [8].

1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129

3DDFA+



Figure 19. Comparisons between 3DDFA [3] and 3DDFA+(Ours) on MultiFace [7].

1130
1131
1132
1133

1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156

1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187

1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210

Dlib



Figure 20. Comparisons between Dlib [5] and DAD-3DNet+(Ours) on MultiFace [7].

1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237

FAN



Figure 21. Comparisons between FAN [1] and DAD-3DNet+(Ours) on MultiFace [7].

1238

DAD-3DNet+



1292

1239

1293

1240

1294

1241

1295

1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318

3DDFA-V2



Figure 22. Comparisons between 3DDFA-V2 [4] and DAD-3DNet+(Ours) on MultiFace [7].

1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345

DAD-3DNet



Figure 23. Comparisons between DAD-3DNet [6] and DAD-3DNet+(Ours) on MultiFace [7].

1346
1347
1348
1349

1400
1401
1402
1403