

000
001
002
003
004
005
006
007
008
009
010
011
012
013
014
015
016
017
018
019
020
021
022
023
024
025
026
027
028
029
030
031
032
033
034
035
036
037
038
039
040
041
042
043
044
045
046
047
048
049
050
051
052
053054
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

The Whole is greater than the sum of its Nonrigid Parts

Supplementary Material

Anonymous CVPR submission

Paper ID 5649

In this supplementary we provide:

1. Analysis of our network; We provide an ablation experiment introspecting the influence of the full shape Q on the network reconstructions. Additionally, we provide robustness analysis of our trained network in Section 1.
2. Additional visualizations of the network reconstructions in Section 2.
3. Visualizations of the dense correspondence results from the partial shape to the full shape in Section 3

We will publish code allowing for a reproduction of all results contained here and in the main manuscript upon acceptance.

1. Analysis

1.1. Comparison with a fixed template baseline

As described in the main manuscript, in order to predict the completion of a partial shape P , our method requires a full reference shape Q of the same subject in an arbitrary pose. We motivate this setting by a requirement for a completion that is faithful to the subject shape. This is different from previous completion methods which can only approximate or hallucinate missing details.

Here we would like to support this claim experimentally, by comparing with a baseline which uses a fixed template. Specifically, instead of providing a full shape Q of the same subject as the partial shape P , we provide a *fixed* full template T for all inputs. With this modification, the ablation network is trained with the triplets $\{(P_n, T, R_n)\}_{n=1}^N$, where N is the size of the training set. At inference time, we use the same template T to make a prediction for a given input part P . We chose the template to be the first subject from the FAUST Projections dataset, in its null pose. Both the original and the fixed-template networks were trained on the FAUST Projections training set, with identical parameters and for the same number of epochs, as described in

Section 3.6 in the paper. Table 1 summarizes the prediction errors of both methods, Figure 3 compares the partial correspondence results and Figure 2 shows visual comparison. The results clearly show the benefit of utilizing the shared geometry between the part and a full non-rigid observation of it. In particular, we receive a noticeable improvement in correspondence prediction as well as a lower reconstruction error across all metrics. Perhaps more importantly, Figure 2 demonstrates the main motivation of our framework: a completion that respects the fine details of the underlying shape. To further emphasize this effect, we magnify the face regions of each shape, showing the loss in detail achieved with the alternative training method.



Figure 1. Constant Template used in Ablation Fixed-Template experiment

Figure 2 implies how powerful our method is when it comes to the reconstruction of fine details, such as the facial structure and delicate body features. We verify that acquiring access to a full observation in inference time can significantly improve the reliability of the reconstruction for a network trained to utilize such information. In the absence of this full observation at inference time, the ablation network can only utilize the input part and the acquired statistics of the training examples, encoded in the network weights. While this later information can be used for coarse completion, we evidence it is not sufficient for accurate completion.

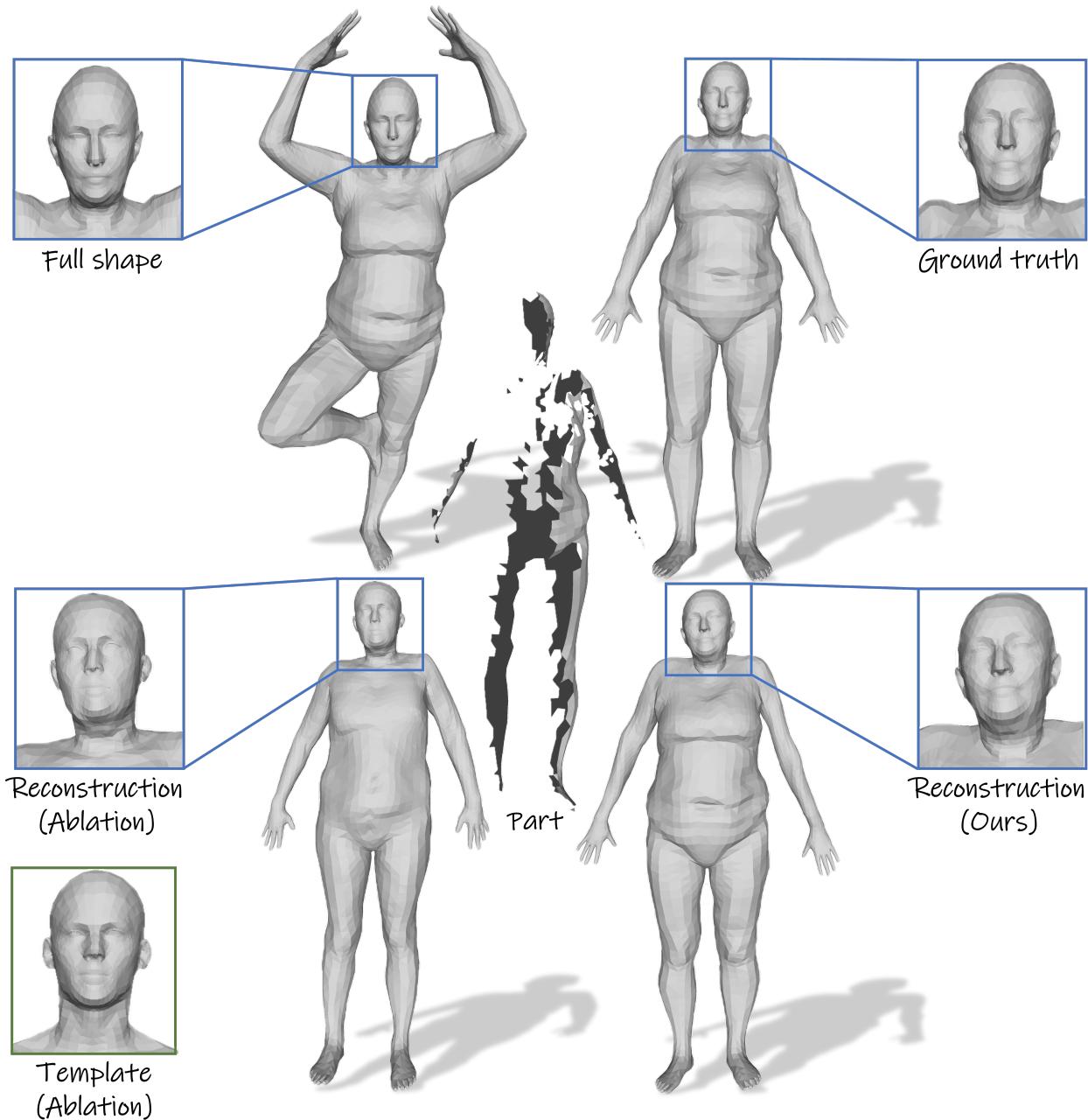


Figure 2. Comparison with Fixed-Template ablation experiment.

1.2. Robustness Analysis

We turn to analyze the robustness and stability of our proposed method, in hopes of shedding light of its possible applicability in real world conditions. Three specific aspects of the method were inspected empirically, each allowing for a realization of some non-optimal condition commonly found in real scans. The following experiments utilize a network trained over the FAUST train set. The realization is provided over a test-set of 200 single-view pro-

jected scans produced from 10 azimuthal viewpoints around 2 human subjects exhibiting 10 different poses. The relevant full shapes were taken from the FAUST dataset, and are completely disjoint from our train set, as they contain unique subjects and poses. Each scan P is matched with all possible poses Q of the same subject, achieving a total of 2000 inputs. We utilize a descriptive partial set of the evaluation metrics proposed in section 4.3 of the paper to evaluate each experiment.

216
217
218
219
220

Error	Euclidean distance GT and reconstruction [cm]	Volumetric err. mean \pm std [%]	Directional Chamfer distance GT to reconstruction [cm]	Directional Chamfer distance reconstruction to GT [cm]	Full Chamfer distance [cm]
Ablation	3.74	17.63 ± 7.41	3.00	2.32	5.32
Ours	2.94	7.05 ± 3.45	2.42	1.95	4.37

Table 1. **Comparison with Fixed-Template Ablation Experiment.** We evaluate our method against an ablation experiment, repeating exactly the same training except of one significant difference: instead of providing the full shape Q_n as described in the main paper, we provided a *constant* full template T in each of the training examples $\{(P_n, T, R_n)\}_{n=1}^N$. The template T is used in inference as well, to predict the completion of a given input part P . We report the prediction errors on FAUST test set, while both networks were trained on FAUST train set. The first and second rows summarize the ablation errors and our method errors, respectively.

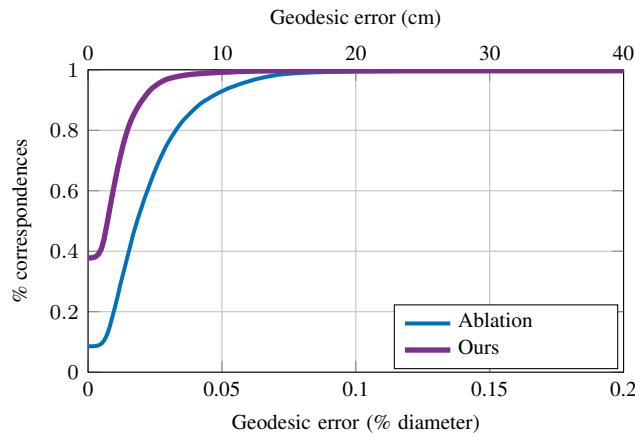


Figure 3. **Comparison with Fixed-Template Ablation Experiment.** Partial correspondence error evaluated on FAUST Projections dataset.

Residual Noise In this experiment, we attempt to emulate various artifacts commonly found in segmented depth scans. We corrupt the vertices of each partial input shape with various degrees of additive white Gaussian noise, with standard deviations in the range [0-4] cm. The corrupted partial shapes are fed to the network, together with the full shapes. Averaged reconstruction statistics are displayed graphically in Figure 4. As apparent from the figure, the method accuracy only slightly declines with the increase of the noise.

Downsampling We address the network’s ability to infer on partial shapes with decreasing degrees of resolution. For each partial shape in the mentioned test set, we decimate at random some percentage of the existing vertices, and infer on the resultant set. As can be seen in Figure 5, even under a majority decimation of the vertices, the proposed network is able to recover well the ground truth shape.

Projection Angle Finally, we examine the dependency of our network to the projection angle. We note that due to the different projection angles and poses, it is not unreasonable that some angles hold a higher degree of information

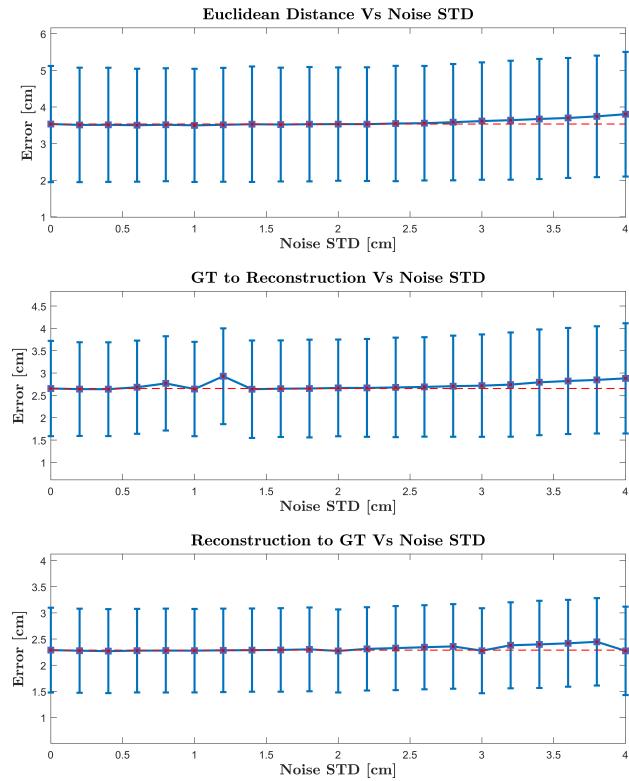


Figure 4. **Robustness to Noise.** Three reconstruction metrics evaluated on completions originating from corrupted partial shapes with varying levels of additive white Gaussian noise. A baseline with the evaluation realized with no noise is marked with a dashed red line.

relevant for reconstruction than others. Ideally, we would like to enable the network a reliable reconstruction at every angle, regardless if the information seen is the back, front or sides of a shape. We partition the 2000 completions received over the test set into their corresponding projection angles, and accumulate the errors over each partition. The result is displayed in Figure 6. The received error distribution is close to uniform, attributing to the method’s azimuthal invariancy.

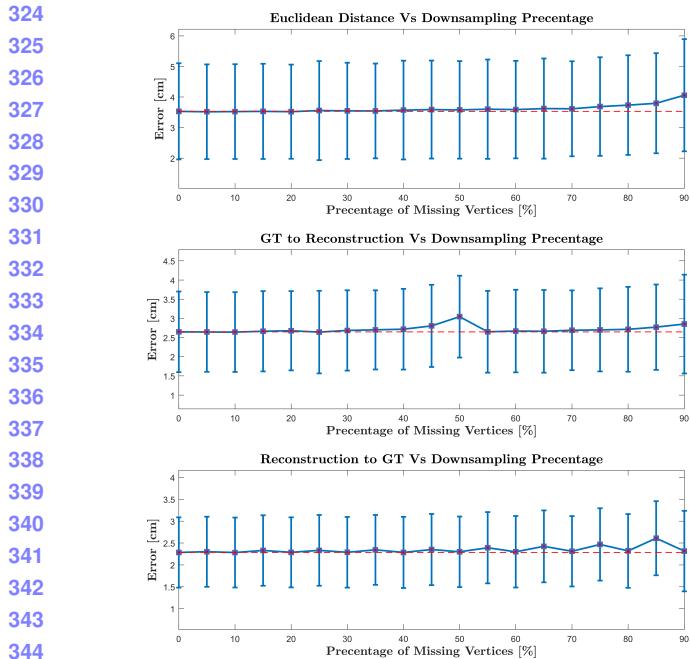


Figure 5. **Robustness to Downsampling.** Three reconstruction metrics evaluated on completions originating from decimated partial shapes with varying levels of vertex erasure. A baseline with the evaluation realized with no decimation is marked with a dashed red line.

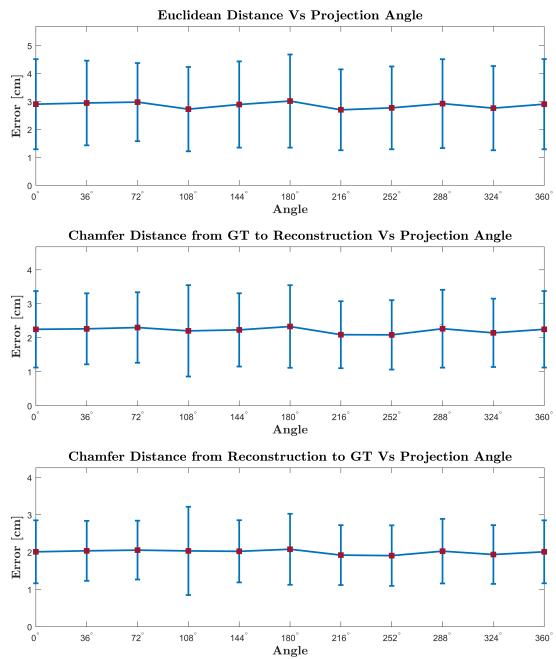


Figure 6. **Robustness to Projection Angle.** Three reconstruction metrics evaluated on different groups of the test-set, partitioned by the projection angle. We note a close to uniform distribution over the different angles, attributing to a azimuthal invariancy.

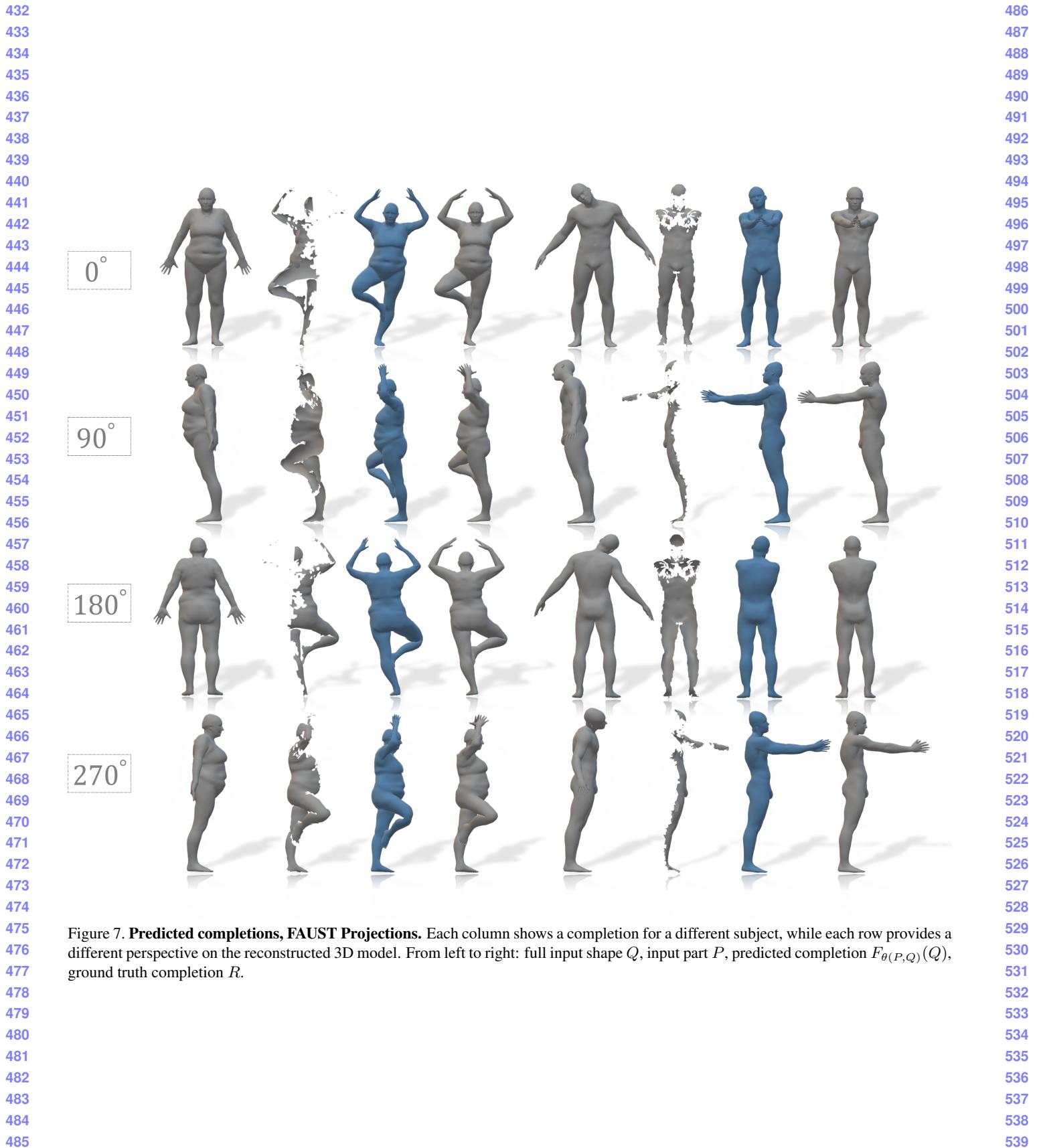
2. Additional Visualizations

Here we provide additional reconstructions that were not included in the main paper in order to save space. Figure 7 and Figure 8 visualize our network predictions for examples from **FAUST Projections** and **AMASS Projections**, respectively.

3. Non-Rigid partial correspondence

Figure 9 visualizes the dense correspondence between the input partial and full shape. As explained in the paper, we achieve this by using the network reconstruction as a proxy; For every point in the partial shape we calculate the nearest neighbor point in the reconstruction allowing us a recovery of a mapping between the partial shape to the reconstructed shape, which is by construction also the mapping between the part and the full input shape. In **Section 4.5** of the paper we evaluated the predicted correspondence numerically for **FAUST Projections** and **AMASS Projections** datasets, providing geodesic error graphs for both, in **Figure 4** and **Figure 5**, respectively. For completion, we show the results also qualitatively here.

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



475 **Figure 7. Predicted completions, FAUST Projections.** Each column shows a completion for a different subject, while each row provides a
476 different perspective on the reconstructed 3D model. From left to right: full input shape Q , input part P , predicted completion $F_{\theta(P,Q)}(Q)$,
477 ground truth completion R .

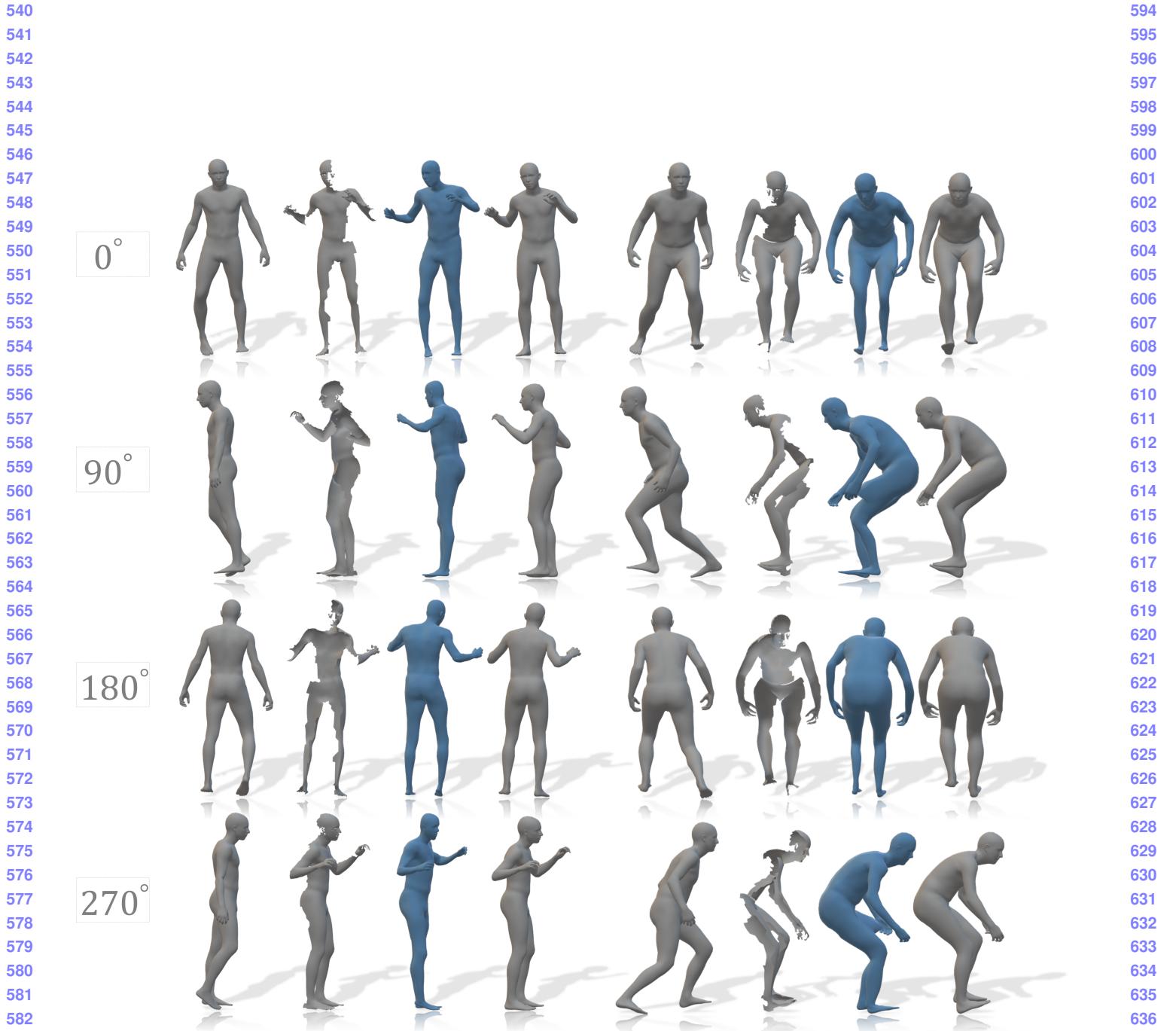


Figure 8. **Predicted completions, AMASS Projections.** Each column shows a completion for a different subject, while each row provides a different perspective on the reconstructed 3D model. From left to right: full input shape Q , input part P , predicted completion $F_{\theta(P,Q)}(Q)$, ground truth completion R .

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

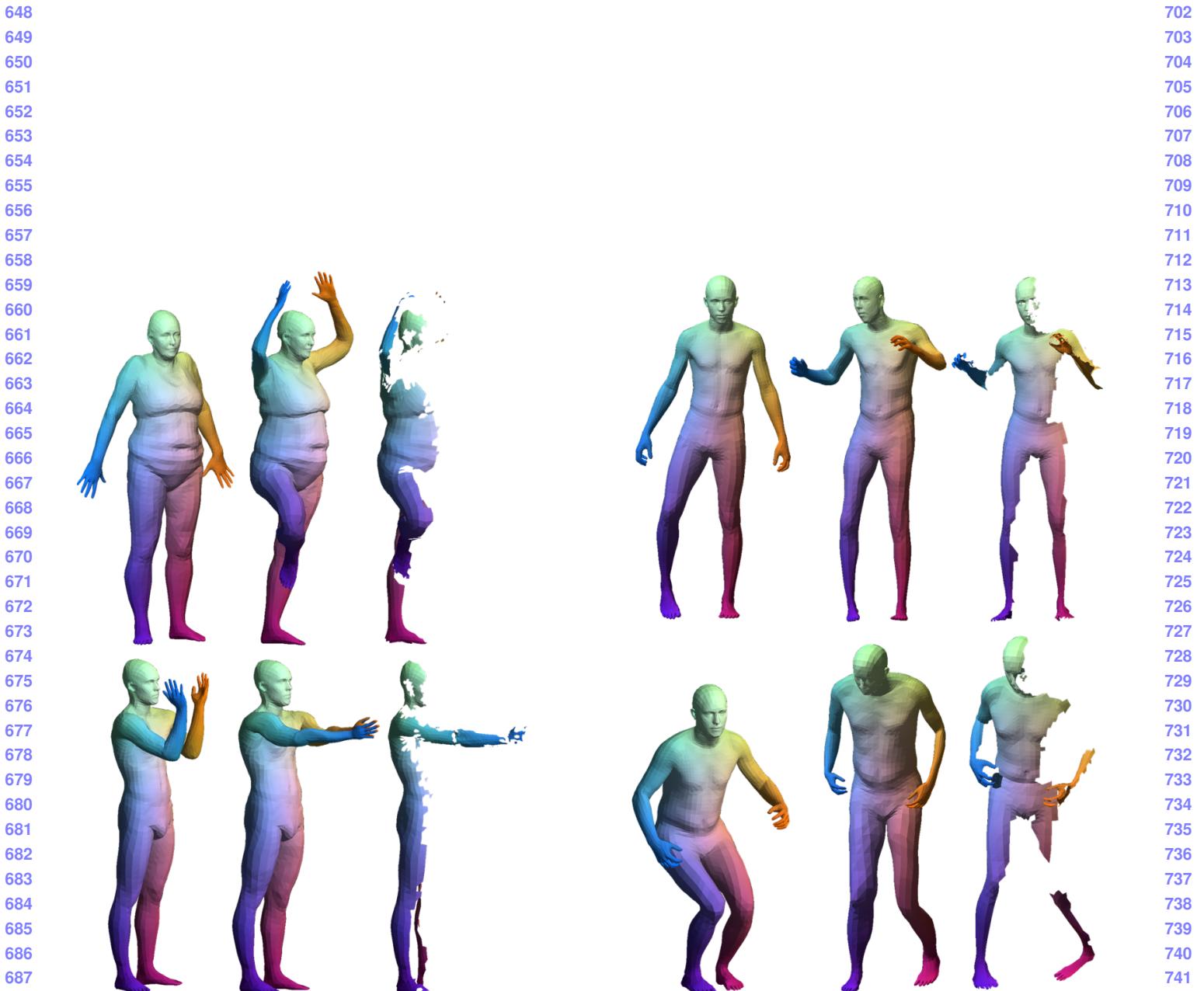


Figure 9. **Non-Rigid partial correspondence.** Left and right columns show the dense correspondence for FAUST Projections and AMASS Projections, respectively. From left to right: full input shape Q , our network completion $F_{\theta(P,Q)}(Q)$ and partial input shape P . Corresponding points are indicated by the same color.

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