

1 We thank the reviewers for their useful comments. As the most positive and confident reviewer, **R3**, raised no concerns
2 (we were delighted to read **R3** has found our work significant and impactful), below we focus on **R4** and **R5**'s reviews.

3 **R4.** The reference for the complexity of joint alignment of N time-series via DTW is [4]. “ \circ ” = function composition.
4 Re Eq. (3): indeed, we define θ_i only later (line 157); we’ll fix this, thanks. As suggested, we will drop the LHS.
5 **Eq. (4) is correct:** as written there, $T^\theta(x) = \phi^\theta(x, t = 1)$; note it is an **integral equation**; *i.e.*, its solution, $\phi^\theta(x, \cdot)$
6 appears both outside and inside the integral (our notation is standard for such equations). Lines 157/159: agreed. Space
7 limits prevent us from detailing how CPAB warp/gradient are evaluated; in short, the CPA structure lends itself to
8 highly-efficient and highly-accurate solvers of the associated integral equations. For details, cf. [2, 1]. According
9 to [2], the **CPAB complexity** is the sum of two **linear** terms, one w.r.t. # intervals in Ω , and one w.r.t. the # of points
10 in the signal, where the first term is negligible (unless the signal is very short); moreover, as points (and signals) are
11 parallelized over via GPU, the proportionality constant of the 2nd term is small, yielding excellent timings (line 253).
12 The complexity of each gradient component is similar and the components are parallelized over. True, DBA has no
13 **Hyper-parameters (HPs)** and SDTW has 1 HP, but DBA clearly under-performs when compared with SDTW/DTAN,
14 and, unlike DTAN, neither DBA nor SDTW generalizes. DTAN has 3 HPs: 1 for Ω , and 2 for the prior. The effect of
15 the last two is studied in our supmat. Other HPs (*e.g.*: the choice of f_{loc} ; # of layers/neurons) are common in DL and
16 are related to optimization and generalization capabilities of the model; **The choice of Ω** determines $\dim(\theta)$ (*i.e.*, the
17 CPAB’s expressiveness) and thus in practice we set it according to the training data availability: when data is scarce it is
18 prudent to use a coarse Ω (hence low $\dim(\theta)$) to avoid over-fitting (note $\dim(\theta) = \#$ of neurons in the last FC layer).

19 **R4. DBA/SDTW are applicable to test data only in the limited sense** that new optimization problems can be
20 formulated and solved from scratch (see lines 3/32/70, and, especially, 196) and if we ignore the fact they need test-data
21 labels; *i.e.*, we agree that for DBA/SDTW “it suffices to recompute an alignment” but note that (1) **on test data the**
22 **difference in speed is huge** due to DTAN’s fast forward pass (vs. DBA/SDTW’s expensive computation of either
23 DTW of each test signal to the train-set barycenter (BC) or a test-set BC) and (2) that, during test, DTAN does not
24 need class labels; *i.e.*, for multi-class signals, DBA/SDTW/DTAN all require labels during training (as the alignment is
25 within class). During test, however, only DBA/SDTW (but not DTAN) require test-data labels so they can recompute
26 DTW between the new signal and the **correct** train-set BC. Without these labels, they must first solve an error-prone
27 classification problem. We believe single-class/multi-class is the appropriate terminology (note that aligning together
28 signals from different classes is usually undesirable). DTAN handles **Multivariate Data** easily, in the same manner
29 Spatial Transformer Nets handles RGB images. **Variable-length (VL):** Our experiments focused on fixed-length
30 signals. For both f_{loc} and CPAB, VL is a non-issue: by dropping the boundary conditions (as done in the supmat), any
31 time interval can be warped towards any other, even if they are of different lengths. We agree, however, that for VL the
32 loss itself needs to be modified accordingly. In any case, please note SDTW had assumed a predefined fixed-length
33 mean signal and they (like the authors of DBA) experimented only on datasets of within-dataset fixed lengths.

34 **R5.** As requested, we here add another evaluation using
35 t-SNE visualization (Fig. 1). We do not assume data is
36 available at large scale (though we **can and do** handle
37 large data sets as well); *e.g.*, UCR contains 85 different
38 datasets, many of which include only few exemplars per
39 class (‘ECGFiveDays’, Fig 1., main paper, had only ~ 10
40 samples per class). We believe our experiments section
41 was extensive and thorough, and we refer the reviewer
42 to our supmat which includes more analyses and results.
43 The 1-NN experiment is the standard benchmark for time-
44 series alignment/averaging (and is not used as a benchmark
45 for best classification results). **We also included a CNN**
46 **vs. DTAN-CNN evaluation** (lines 289-297); DBA/SDTW
47 cannot be used for improving a CNN this way. We address
48 works akin to [3] in lines 100-105; particularly, as [3] pre-
49 dicted **pairwise** warps between 3D shapes using templates,
50 while DTAN learns joint-alignment of multiple 1D signals
51 without one, the two methods are quite different.

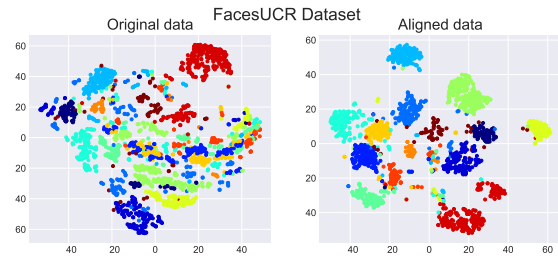


Figure 1: t-SNE visualization of the original and aligned **test data** of the challenging 11-class ‘FacesUCR’ dataset. No class labels were used during DTAN alignment of the test data (it is used here only for visualization). The t-SNE highlights how DTAN decreases within-class variance while increasing inter-class variance. For DBA/SDTW, handling such multi-class test data alignment requires solving new optimization problems as well as (known or estimated) class labels of the test signals.

52 References

- 53 [1] O. Freifeld. Deriving the CPAB derivative. Technical report, Ben-Gurion University, 2018.
54 [2] Freifeld et al. Transformations based on continuous piecewise-affine velocity fields. *IEEE TPAMI*, 2017.
55 [3] Groueix et al. 3d-coded: 3d correspondences by deep deformation. In *ECCV*, 2018.
56 [4] L. Wang and T. Jiang. On the complexity of multiple sequence alignment. *Journal of computational biology*, 1994.