

Diffeomorphic Temporal Alignment Nets

**Ron Shapira Weber¹, Matan Eyal¹, Nicki Skaftø Detlefsen²,
Oren Shriki¹ and Oren Freifeld¹**

1. Ben-Gurion University

2. Technical University of Denmark

October 2019

Author's summary (NeurIPS 2019)

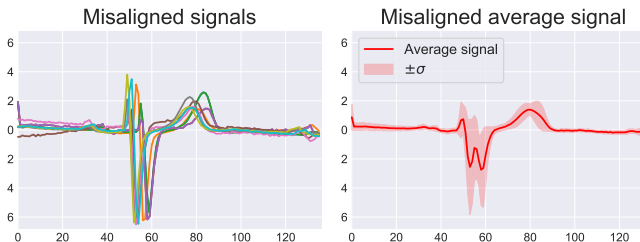


Outline

- 1 Nonlinear Misalignment
- 2 Preliminaries
- 3 Diffeomorphic Temporal Alignment Nets
- 4 Experiments and Results

Problem Formulation - Time-Series Joint Alignment

- Time-series data often presents a significant amount of nonlinear misalignment.
- This may include simple phase-shift between similar instances, artifacts due to sampling methods and/or stretching and squeezing of different regions of the signal.
- Even trivial task such as computing the arithmetic mean (i.e. average signal) is hard under nonlinear misalignment.



Problem Formulation - Time-Series Joint Alignment

- Let $(\mathbf{U}_i)_{i=1}^N$ be a set of N time-series observations. The nonlinear misalignment could be written as:

$$(\mathbf{U}_i)_{i=1}^N = (\mathbf{V}_i \circ \mathbf{W}_i)_{i=1}^N \quad (1)$$

- Where \mathbf{U}_i is the i^{th} misaligned signal, \mathbf{V}_i is the i^{th} latent aligned signal and \mathbf{W}_i is a latent warp of the domain of \mathbf{V}_i .
- For technical reasons, the misalignment is usually viewed in terms of $T_i \triangleq \mathbf{W}_i^{-1}$, the inverse warp of \mathbf{W}_i , implicitly suggesting \mathbf{W}_i is invertible.
- It is also typically assumed that $(T_i)_{i=1}^N$ belong to some nominal family of warps, parametrized by $\boldsymbol{\theta}$:

$$(\mathbf{V}_i)_{i=1}^N = (\mathbf{U}_i \circ T^{\boldsymbol{\theta}_i})_{i=1}^N, \quad T_i = T^{\boldsymbol{\theta}_i} \in \mathcal{T} \quad \forall i \in (1, \dots, N). \quad (2)$$

Problem Formulation - Time-Series Joint Alignment

- We argue that this problem should be seen as a learning one, mostly due to the need for generalization, and define:

Definition (the joint-alignment problem)

Given the observations $(U_i)_{i=1}^N$, infer the latent $(T^{\theta_i})_{i=1}^N \subset \mathcal{T}$.

- To solve the joint alignment problem one must define a:
 - Warp family - \mathcal{T}
 - Model capable of generalizing the learned alignment.
 - Metric measuring the misalignment.
- We will define each in following slides.

Preliminaries - Spatial/Temporal Transformer Nets

- Introduced by Jaderberg et al. (2015), the Spatial Transformer Nets (STN) is an end-to-end differential module that allows for certain invariances to spatial warps.
- Parameterized by any differential transformation family (*i.e.* Affine), the STN Learns input-dependent spatial transformation to allow for robust spatial invariance in deep neural networks.

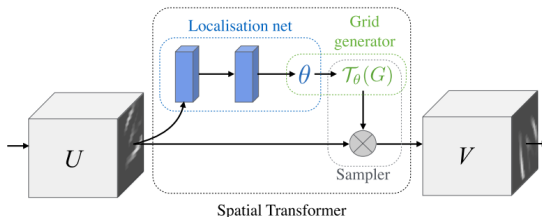


Figure taken from: Jaderberg et al. NIPS (2015)

Preliminaries - Spatial/Temporal Transformer Nets

- Temporal Transformer Nets (TTN), the time-series analog of STN.
- In more detail, let U denote the input of the TT layer. Its output consists of $\theta = f_{\text{loc}}(U)$ and $V = U \circ T^\theta$ (the latter, *i.e.*, the warped signal, is what is being passed downstream the TTN), where $T^\theta \in \mathcal{T}$ is a 1D warp parameterized by θ . The function $f_{\text{loc}} : U \mapsto \theta$ is itself a neural net called the localization net.

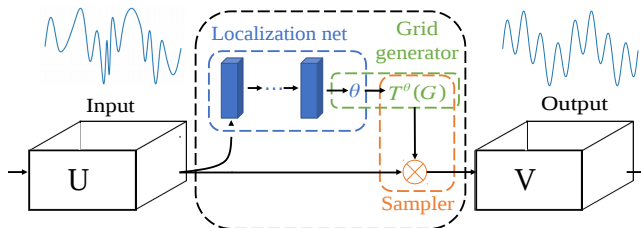


Figure taken from: Skaftve et al. CVPR (2018)

Preliminaries - Diffeomorphisms

- As previously mentioned, \mathcal{T} needs to be specified. In the context of time warping, diffeomorphisms are a natural choice (Mumford & Desolneux, 2010).

Definition (Diffeomorphisms)

A diffeomorphism is a differentiable invertible map with a differentiable inverse.

- Unfortunately, most representations of highly-expressive diffeomorphisms are complicated and computationally expensive.
- In our case, since the proposed method explicitly incorporates them in a DL architecture, it is even more important to drastically reduce the computational difficulties.

Preliminaries - CPAB

- We now define the warp family to be incorporated in our model.
- Following Skafté et al. (CVPR, 2018) incorporation of diffeomorphisms in STNs, the CPAB warps that had been proposed by Freifeld et al. (ICCV, 2015; PAMI, 2017) and are also used in this work.
- While any diffeomorphism could be integrated into DTAN, we chose CPAB warps since they:
 - 1 Are highly expressive and efficient.
 - 2 Offer an efficient and highly-accurate way to evaluate $x \mapsto \nabla_{\theta} T^{\theta}(x)$ (i.e. CPAB gradient).

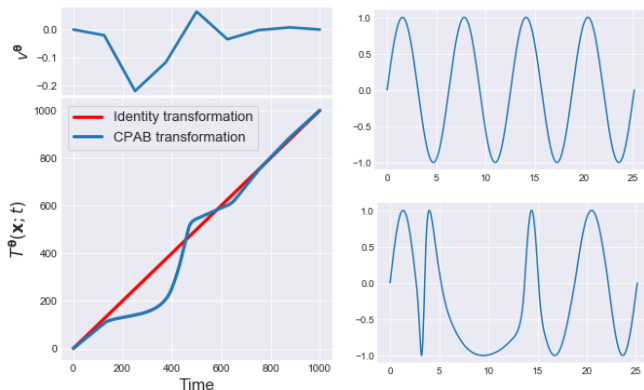
Preliminaries - CPAB

- The name CPAB, short for CPA-Based, is due to the fact that these warps are based on Continuous Piecewise-Affine (CPA) velocity fields. The term “piecewise” is w.r.t. a partition, denoted by Ω , of the signal's domain into subintervals.
- Let \mathcal{V} denote the linear space of CPA velocity fields w.r.t. such a fixed Ω , let $d = \dim(\mathcal{V})$, and let $v^\theta : \Omega \rightarrow \mathbb{R}$, a velocity field parametrized by $\theta \in \mathbb{R}^d$, denote the generic element of \mathcal{V} , where θ stands for the coefficient w.r.t. some basis of \mathcal{V} is

$$\mathcal{T} \triangleq \{T^\theta : x \mapsto \phi^\theta(x; 1) \text{ s.t. } \phi^\theta(x; t) = x + \int_0^t v^\theta(\phi^\theta(x; \tau)) d\tau\}; \quad (3)$$

Preliminaries - CPAB

- Left: An illustration of a CPAB warp (relative to the identity transformation) with its corresponding CPA velocity field (above). Right: a sine wave before (top) and after (bottom) being warped by the presented CPAB transformation.



Diffeomorphic Temporal Alignment Nets

- As mentioned before, to solve the joint alignment problem one must define a:
 - Warp family - \mathcal{T}
 - Model capable of generalizing the learned alignment.
 - Metric measuring the misalignment.
- Thus, we set:
 - Warp family - CPAB.
 - Model - Temporal Transformer Nets (TTN).
 - Metric measuring the misalignment - least squares (to be define below).

Loss Function

- Definition 1 requires the specification of \mathcal{T} and a loss function for estimating $(T^{\theta_i})_{i=1}^N$.
- Let U_i denote an input signal, let $\theta_i = f_{\text{loc}}(U_i, \mathbf{w})$ denote the corresponding output of the localization net $f_{\text{loc}}(\cdot, \mathbf{w})$ of weights \mathbf{w} , and let V_i denote the result of warping U_i by $T^{\theta_i} \in \mathcal{T}$; i.e., $V_i = U_i \circ T^{\theta_i}$.

Loss Function

- As the variance of the observed $(U_i)_{i=1}^N$ is (at least partially) explained by the latent warps, $(T^{\theta_i})_{i=1}^N$, we seek to minimize the empirical variance of the warped signals, $(V_i)_{i=1}^N$. In other words, our data term in this setting is

$$F_{\text{data}}(\mathbf{w}, (U_i)_{i=1}^N) \triangleq \frac{1}{N} \sum_{i=1}^N \left\| \mathbf{V}_i(U_i; \mathbf{w}) - \frac{1}{N} \sum_{j=1}^N \mathbf{V}_j(U_j; \mathbf{w}) \right\|_{\ell_2}^2 \quad (4)$$

- For multi-class problems, our data term is the sum of the within-class variances:

$$F_{\text{data}}(\mathbf{w}, (U_i)_{i=1}^N) \triangleq \sum_{k=1}^K \frac{1}{N_k} \sum_{i: z_i=k} \left\| \mathbf{V}_i(U_i; \mathbf{w}) - \frac{1}{N_k} \sum_{j: z_j=k} \mathbf{V}_j(U_j; \mathbf{w}) \right\|_{\ell_2}^2$$

- where K is the number of classes, z_i takes values in $\{1, \dots, K\}$.

Loss Function - Regularization

- In both the single- and multi-class cases, we also use a regularization term on the warp

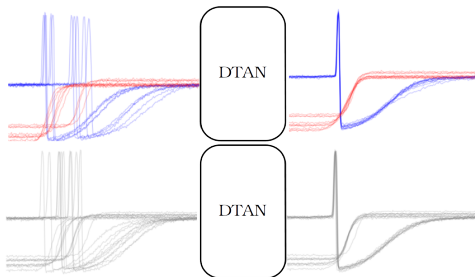
$$F_{\text{reg}}(\mathbf{w}, (U_i)_{i=1}^N) = \sum_{i=1}^N (\boldsymbol{\theta}_i(\mathbf{w}, U_i))^T \boldsymbol{\Sigma}_{\text{CPA}}^{-1} \boldsymbol{\theta}_i(\mathbf{w}, U_i) \quad (5)$$

- Where $\boldsymbol{\Sigma}_{\text{CPA}}$ is a CPA covariance matrix (proposed by Freifeld et al. (ICCV, 2015; PAMI, 2017) associated with a zero-mean Gaussian smoothness prior over CPA fields.
- $\boldsymbol{\Sigma}_{\text{CPA}}$ has two parameters: λ_{var} , which controls the overall variance, and λ_{smooth} , which controls the smoothness of the field.
- A small λ_{var} favors small warps (*i.e.*, close to the identity); similarly, the larger λ_{smooth} is, the more it favors CPA velocity fields that are almost purely affine.
- Our loss function, to be minimized over \mathbf{w} , is

$$F(\mathbf{w}, (U_i)_{i=1}^N) = F_{\text{data}}(\mathbf{w}, (U_i)_{i=1}^N) + F_{\text{reg}}(\mathbf{w}, (U_i)_{i=1}^N). \quad (6)$$

Diffeomorphic Temporal Alignment Nets

- Diffeomorphic Temporal Alignment Nets (DTAN) Allows for time-series nonlinear joint alignment in an input-dependent manner.
- In a single-class case, the method is unsupervised: the ground-truth alignments are unknown.
- In the multi-class case, it is semi-supervised in the sense that class labels (but not the ground-truth alignments) are used during learning.
- During test, however, the class labels are unknown.



Recurrents DTANs

- We also propose a recurrent variation of DTAN, R-DTAN.
- inspired by, how Lin et al. (CVPR, 2017) used a recurrent net with affine 2D warps, we propose to iteratively warp the input signal by DTAN, using the same localization net.
- By enforcing a zero-boundary condition ($U[0] = V[0], U[n] = V[n]$, where n is the signal's length), we can propagate the signal itself, thus avoiding the need for inverse composition.
- This allows for increasing the expressiveness of the CPAB diffeomorphisms without increasing the number of trainable weights or the partition Ω .

Experiments and Results

- We first explore DTAN time-series joint alignment of synthetic data. Unlike real-world data, here the latent warp signal is available for us.

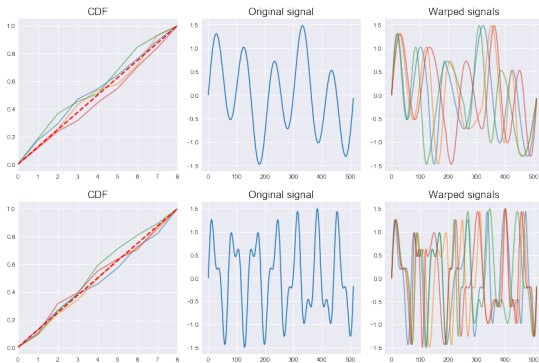
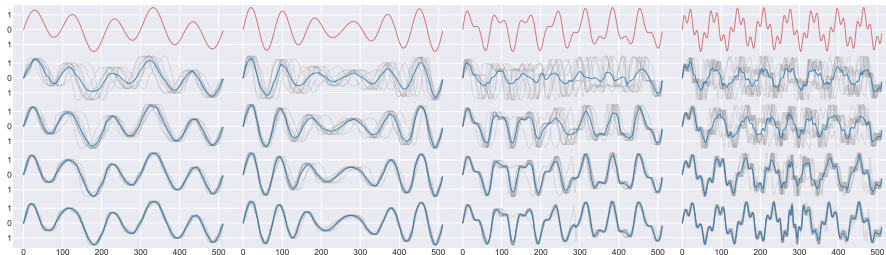


Figure: Synthetic data generation process. We create 4 latent signal (2 are presented here). We then sample CDFs from a Dirichlet distribution and use them as random warp.

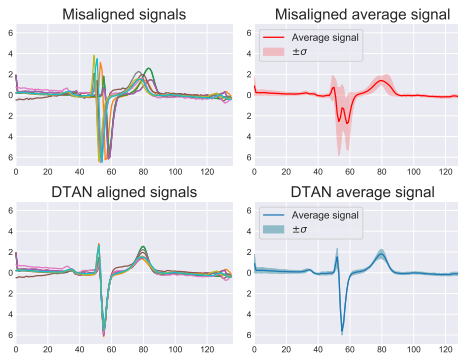
R-DTAN Joint Alignment of Synthetic Data

- Here we see the output of R-DTAN at each recurrence of the synthetic **test data**.
- The original signals are in red. Each column represents a different class. Note that the same model was trained on all classes.



Joint Alignment of Real-World data

- The UCR time-series classification archive (Chen et al, 2015) contains 85 real-world datasets (we used 84).
- Below: Learning to generalize time-series joint alignment from train to test signals on the ECGFiveDays dataset. Each panel depicts 10 random sample from the **test set** and their respective mean (shaded areas correspond to standard deviations).



- Y. Chen, E. Keogh, B. Hu, N. Begum, A. Bagnall, A. Mueen, and G. Batista. The ucr time series classification archive, July 2015. www.cs.ucr.edu/~eamonn/time_series_data/.
- O. Freifeld, S. Hauberg, K. Batmanghelich, and J. W. Fisher III. Highly-expressive spaces of well-behaved transformations: Keeping it simple. In *ICCV*, 2015.
- O. Freifeld, S. Hauberg, K. Batmanghelich, and J. W. Fisher III. Transformations based on continuous piecewise-affine velocity fields. *IEEE TPAMI*, 2017.
- M. Jaderberg, K. Simonyan, A. Zisserman, et al. Spatial transformer networks. In *Advances in neural information processing systems*, pages 2017–2025, 2015.
- C.-H. Lin and S. Lucey. Inverse compositional spatial transformer networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2568–2576, 2017.
- D. Mumford and A. Desolneux. *Pattern theory: the stochastic analysis of real-world signals*. AK Peters/CRC Press, 2010.

N. Scafte Detlefsen, O. Freifeld, and S. Hauberg. Deep diffeomorphic transformer networks. In *CVPR*, 2018.