Diffeomorphic Temporal Alignment Nets

Ron Shapira Weber[1], Matan Eyal[1], Nicki Skafte Detlefsen[2], Oren Shriki[1] and Oren Freifeld[1]

1. Ben-Gurion University
2. Technical University of Denmark

October 2019

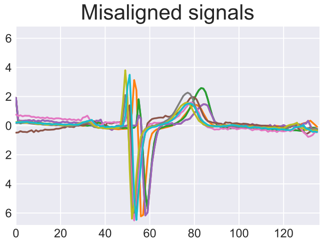
Author's summary (NeurIPS 2019)

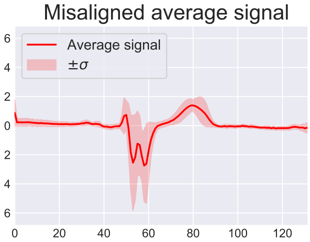
Outline

1. Nonlinear Misalignment
2. Preliminaries
3. Diffeomorphic Temporal Alignment Nets
4. Experiments and Results
5. Conclusion

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

$$(\boldsymbol{U}_i)_{i=1}^N = (\boldsymbol{V}_i \circ W_i)_{i=1}^N \tag{1}$$

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

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

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 $(\boldsymbol{U}_i)_{i=1}^{N}$, infer the latent $(T^{\boldsymbol{\theta}_i})_{i=1}^{N} \subset \mathcal{T}$.

- 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.
- Paramerized by any differential transformation family (*i.e.* Affine), the STN Learns input-depended spatial transformation to allow for robuts spatial invariance in deep neural networks.

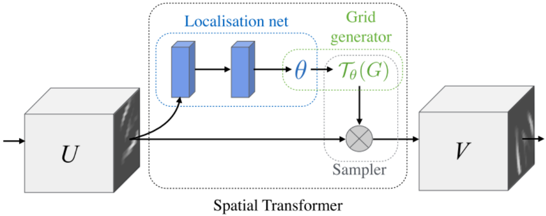


Figure taken from: Jaderberg et al. NIPS (2015)

Preliminaries - Spatial/Temporal Transformer Nets

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

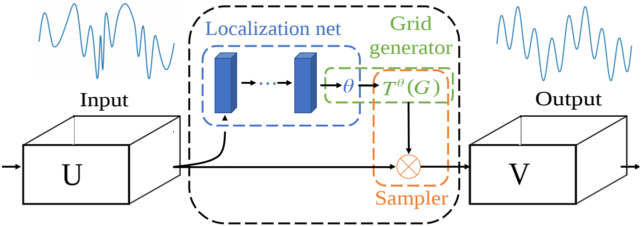


Figure taken from: Skafte et al. CVPR (2018)

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 Skafte 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 $x \mapsto \nabla_{\boldsymbol{\theta}} T^{\boldsymbol{\theta}}(x)$ (*i.e.* CPAB gradient).

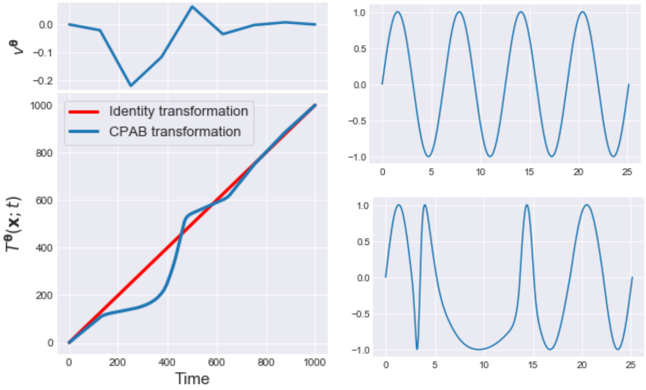
Preliminaries - CPAB

- 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 $v^{\boldsymbol{\theta}} : \Omega \to \mathbb{R}$, a velocity field parametrized by $\boldsymbol{\theta} \in \mathbb{R}^d$, denote the generic element of $\mathcal{V}$, where $\boldsymbol{\theta}$ stands for the coefficient w.r.t. some basis of $\mathcal{V}$ is

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

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 $(T^{\boldsymbol{\theta}_i})_{i=1}^{N}$.
- Let $\boldsymbol{U}_i$ denote an input signal, let $\boldsymbol{\theta}_i = f_{\text{loc}}(\boldsymbol{U}_i, \boldsymbol{w})$ denote the corresponding output of the localization net $f_{\text{loc}}(\cdot, \boldsymbol{w})$ of weights $\boldsymbol{w}$, and let $\boldsymbol{V}_i$ denote the result of warping $\boldsymbol{U}_i$ by $T^{\boldsymbol{\theta}_i} \in \mathcal{T}$; *i.e.*, $\boldsymbol{V}_i = \boldsymbol{U}_i \circ T^{\boldsymbol{\theta}_i}$.

Loss Function

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

$$F_{\text{data}}\left(\boldsymbol{w}, (\boldsymbol{U}_i)_{i=1}^N\right) \triangleq \tfrac{1}{N}\sum\nolimits_{i=1}^{N}\left\|\boldsymbol{V}_i(\boldsymbol{U}_i;\boldsymbol{w}) - \tfrac{1}{N}\sum\nolimits_{j=1}^{N}\boldsymbol{V}_j(\boldsymbol{U}_j;\boldsymbol{w})\right\|_{\ell_2}^2 \tag{4}$$

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

$$F_{\text{data}}\left(\boldsymbol{w}, (\boldsymbol{U}_i)_{i=1}^N\right) \triangleq \sum\nolimits_{k=1}^{K}\tfrac{1}{N_k}\sum\nolimits_{i:z_i=k}\left\|\boldsymbol{V}_i\left(\boldsymbol{U}_i;\boldsymbol{w}\right) - \tfrac{1}{N_k}\sum\nolimits_{j:z_j=k}\boldsymbol{V}_j(\boldsymbol{U}_j;\boldsymbol{w})\right\|_{\ell_2}^2$$

- 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

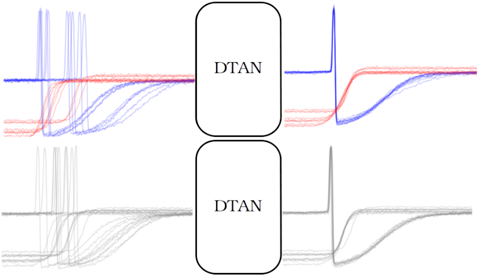
$$F_{\text{reg}}(\boldsymbol{w}, (\boldsymbol{U}_i)_{i=1}^N) = \sum_{i=1}^N (\boldsymbol{\theta}_i(\boldsymbol{w}, \boldsymbol{U}_i))^T \boldsymbol{\Sigma}_{\text{CPA}}^{-1} \boldsymbol{\theta}_i(\boldsymbol{w}, \boldsymbol{U}_i) \tag{5}$$

- 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 $\boldsymbol{w}$, is

$$F(\boldsymbol{w}, (\boldsymbol{U}_i)_{i=1}^N) = F_{\text{data}}(\boldsymbol{w}, (\boldsymbol{U}_i)_{i=1}^N) + F_{\text{reg}}(\boldsymbol{w}, (\boldsymbol{U}_i)_{i=1}^N) \,. \tag{6}$$

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 varition of DTAN, R-DTAN.
- inspired by, how Lin et al. (CVPR, 2017) used a recurrent net with affine 2D warps, we propose the iteratively warp the input signal by DTAN, using the same localization net.
- By enforcing a zero-boundary condition ($\boldsymbol{U}[0] = \boldsymbol{V}[0], \boldsymbol{U}[n] = \boldsymbol{V}[n]$, 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.

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