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

Author's summary - NeurIPSi 2019

Problem Formulation - Time-Series Joint Alignment

- Goal: statistical analysis of time-series data.
- Problem: temporal misalignment confounds statistical analysis.
- Solution: align the data.

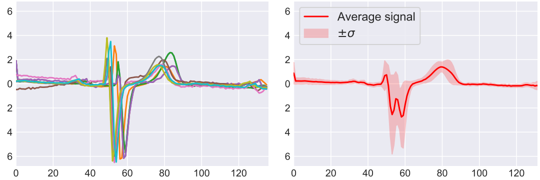
Problem Formulation - Time-Series Joint Alignment

- Goal: statistical analysis of time-series data.
- Problem: temporal misalignment confounds statistical analysis.

- Solution: align the data.

Problem Formulation - Time-Series Joint Alignment

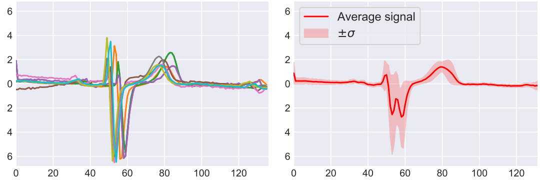
- Goal: statistical analysis of time-series data.
- Problem: temporal misalignment confounds statistical analysis.



- Solution: align the data.

Problem Formulation - Time-Series Joint Alignment

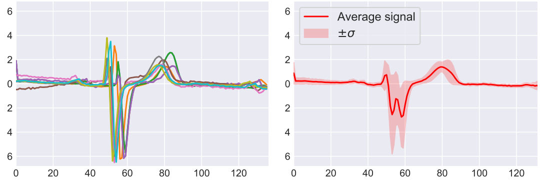
- Goal: statistical analysis of time-series data.
- Problem: temporal misalignment confounds statistical analysis.



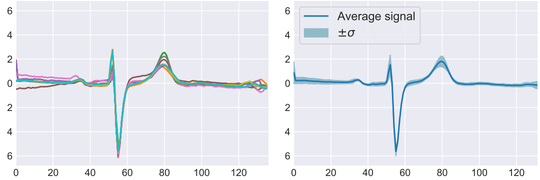
- Solution: align the data.

Problem Formulation - Time-Series Joint Alignment

- Goal: statistical analysis of time-series data.
- Problem: temporal misalignment confounds statistical analysis.



- Solution: align the data.



Problem Formulation - Time-Series Joint Alignment

- Goal: A system that can align large time-series ensembles.
- Problem: Existing alignment methods suffer from one or more of the following problems:
  - Computationally expensive
  - Don't scale well with N (# of signals) and/or T (signal length)
  - Can't handle multiple classes
  - No generalization
- Our proposed solution:
  Diffeomorphic Temporal Alignment Nets (DTAN).

Problem Formulation - Time-Series Joint Alignment

- Goal: A system that can align large time-series ensembles.
- Problem: Existing alignment methods suffer from one or more of the following problems:
  - Computationally expensive
  - Don't scale well with N (# of signals) and/or T (signal length)
  - Can't handle multiple classes
  - No generalization
- Our proposed solution:
  Diffeomorphic Temporal Alignment Nets (DTAN).

Problem Formulation - Time-Series Joint Alignment

- Goal: A system that can align large time-series ensembles.
- Problem: Existing alignment methods suffer from one or more of the following problems:
  - Computationally expensive
  - Don't scale well with N (# of signals) and/or T (signal length)
  - Can't handle multiple classes
  - No generalization
- Our proposed solution:
  Diffeomorphic Temporal Alignment Nets (DTAN).

Problem Formulation - Time-Series Joint Alignment

- Goal: A system that can align large time-series ensembles.
- Problem: Existing alignment methods suffer from one or more of the following problems:
  - Computationally expensive
  - Don't scale well with N ($\#$ of signals) and/or T (signal length)
  - Can't handle multiple classes
  - No generalization
- Our proposed solution: Diffeomorphic Temporal Alignment Nets (DTAN).

Problem Formulation - Time-Series Joint Alignment

- Goal: A system that can align large time-series ensembles.
- Problem: Existing alignment methods suffer from one or more of the following problems:
  - Computationally expensive
  - Don't scale well with N ($\#$ of signals) and/or T (signal length)
  - Can't handle multiple classes
  - No generalization
- Our proposed solution:
  Diffeomorphic Temporal Alignment Nets (DTAN).

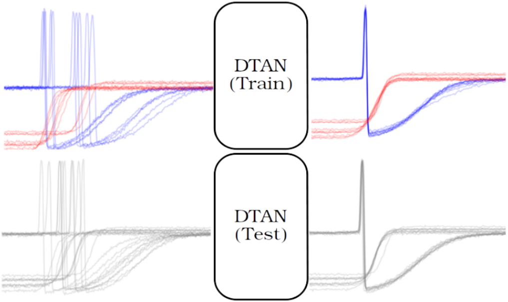
Problem Formulation - Time-Series Joint Alignment

- Goal: A system that can align large time-series ensembles.
- Problem: Existing alignment methods suffer from one or more of the following problems:
  - Computationally expensive
  - Don't scale well with N ($\#$ of signals) and/or T (signal length)
  - Can't handle multiple classes
  - No generalization
- Our proposed solution:
  Diffeomorphic Temporal Alignment Nets (DTAN).

Problem Formulation - Time-Series Joint Alignment

- Goal: A system that can align large time-series ensembles.
- Problem: Existing alignment methods suffer from one or more of the following problems:
  - Computationally expensive
  - Don't scale well with N ($\#$ of signals) and/or T (signal length)
  - Can't handle multiple classes
  - No generalization
- Our proposed solution:
  Diffeomorphic Temporal Alignment Nets (DTAN).

Diffeomorphic Temporal Alignment Nets - Overview



- A Deep Learing (DL)-based system that aligns time-series ensembles within-class in an input-dependent manner.
- Based on the Temporal Transformer Nets (TTN), the time-series analog of Spatial Transformer Nets [Jaderberg et al., NIPS 2015] .

Preliminaries - CPAB

- The diffeomorphism family used in this paper is CPAB [Freifeld et al., ICCV 2015; PAMI 2017] .
- CPAB - warps which are *based* on the integration of 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.

Preliminaries - CPAB

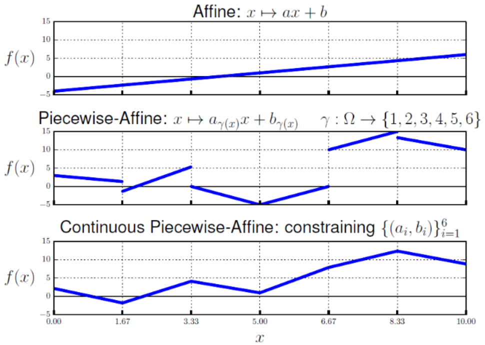
- The diffeomorphism family used in this paper is CPAB [Freifeld et al., ICCV 2015; PAMI 2017] .
- CPAB - warps which are *based* on the integration of Continuous Piecewise-Affine (CPA) velocity fields.
- The term "piecewise" is w.r.t. a partition, denoted by Ω, of the signal's domain into subintervals.

Preliminaries - CPAB

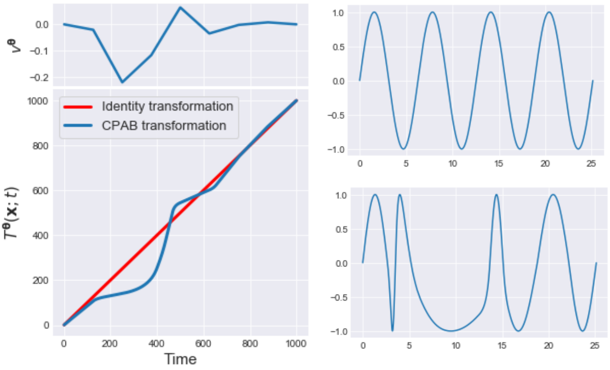
- The diffeomorphism family used in this paper is CPAB [Freifeld et al., ICCV 2015; PAMI 2017] .
- CPAB - warps which are *based* on the integration of Continuous Piecewise-Affine (CPA) velocity fields.
- The term “piecewise” is w.r.t. a partition, denoted by Ω, of the signal’s domain into subintervals.

Preliminaries - CPAB

- The diffeomorphism family used in this paper is CPAB [Freifeld et al., ICCV 2015; PAMI 2017] .
- CPAB - warps which are *based* on the integration of 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.



CPAB - 1D Example



Loss Function - Unsupervised / Semi-supervised Framework

- DTAN is set to minimize the within-class variance of the warped signals:

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

- As well as a regularization term on the warps:

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

- Where Σ_{CPA} is a CPA covariance matrix [Freifeld et al. ICCV 2015; PAMI 2017] associated with a zero-mean Gaussian smoothness prior over CPA fields.

Loss Function - Unsupervised / Semi-supervised Framework

- DTAN is set to minimize the within-class variance of the warped signals:

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

- As well as a regularization term on the warps:

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

- Where Σ_{CPA} is a CPA covariance matrix [Freifeld et al. ICCV 2015; PAMI 2017] associated with a zero-mean Gaussian smoothness prior over CPA fields.

Loss Function - Unsupervised / Semi-supervised Framework

- DTAN is set to minimize the within-class variance of the warped signals:

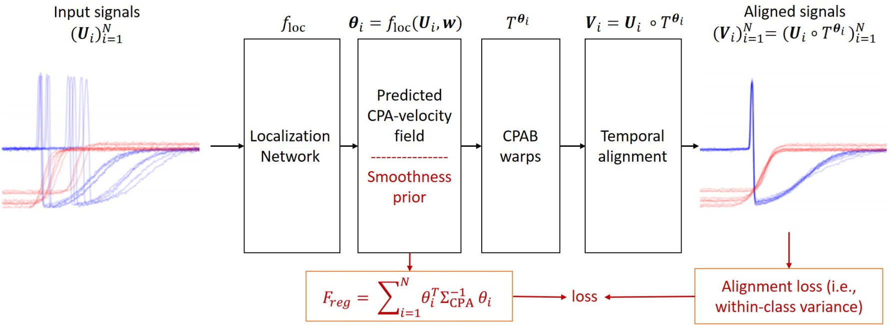
$$F_{\text{data}}\left(\boldsymbol{w}, (\boldsymbol{U}_i)_{i=1}^N\right) \triangleq \sum_{k=1}^{K} \tfrac{1}{N_k} \sum_{i:z_i=k} \left\| \boldsymbol{V}_i\left(\boldsymbol{U}_i;\boldsymbol{w}\right) - \tfrac{1}{N_k} \sum_{j:z_j=k} \boldsymbol{V}_j(\boldsymbol{U}_j;\boldsymbol{w}) \right\|_{\ell_2}^2$$

- As well as a regularization term on the warps:

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

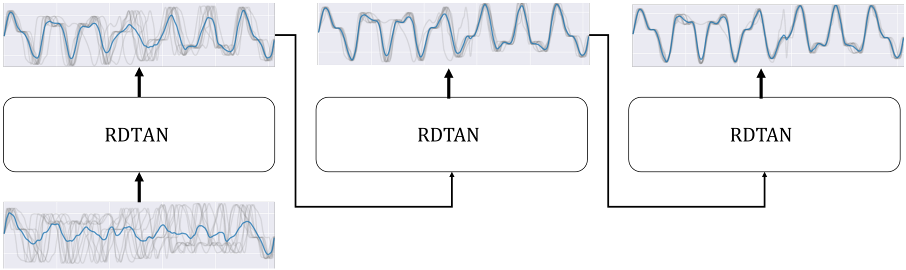
- Where $\boldsymbol{\Sigma}_{\text{CPA}}$ is a CPA covariance matrix [Freifeld et al. ICCV 2015; PAMI 2017] associated with a zero-mean Gaussian smoothness prior over CPA fields.

Diffeomorphic Temporal Alignment Nets



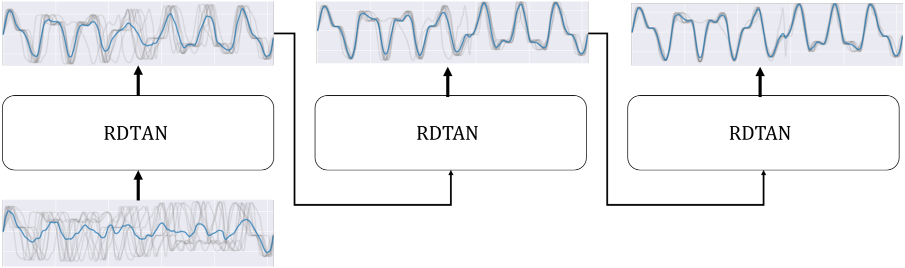
- f_{loc} - 1D-CNN consisting of 3 conv-layers (128-64-64 filters, respectively).
- Final layer: $d = \dim(\boldsymbol{\theta}) = 32$ with $tanh$ activation.

Recurrents DTANs



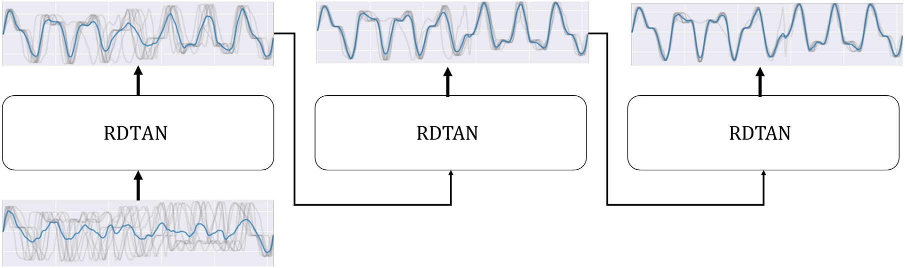
- RDTAN - recurrent warps of the input signals.
- Shared locnet - # of trainable parameters does not increase.
- Easier optimization framework.
- Implies a non-stationary velocity field.

Recurrents DTANs



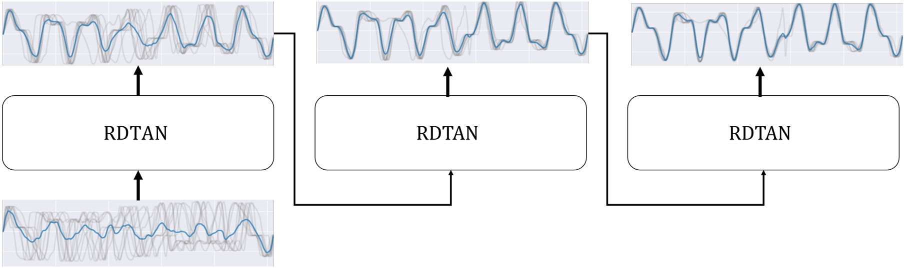
- RDTAN - recurrent warps of the input signals.
- Shared locnet - # of trainable parameters does not increase.
- Easier optimization framework.
- Implies a non-stationary velocity field.

Recurrents DTANs



- RDTAN - recurrent warps of the input signals.
- Shared locnet - # of trainable parameters does not increase.
- Easier optimization framework.
- Implies a non-stationary velocity field.

Recurrents DTANs



- RDTAN - recurrent warps of the input signals.
- Shared locnet - # of trainable parameters does not increase.
- Easier optimization framework.
- Implies a non-stationary velocity field.

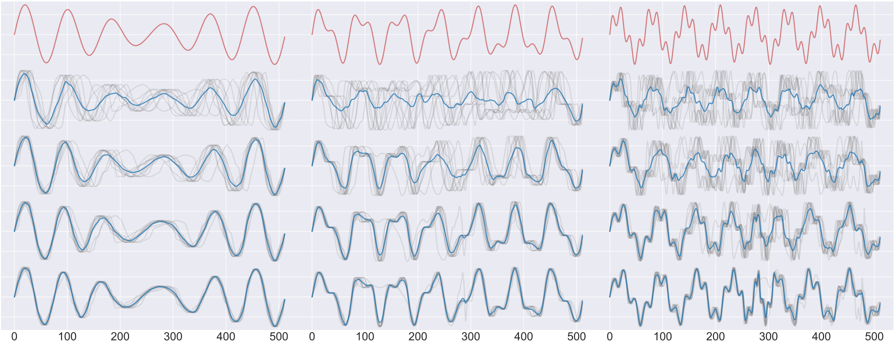
Experiments - Synthetic Data

- RDTAN Joint Alignment of Synthetic Data.

- Latent source signals are known (Red).
- Data - source signals warped with random transformations (Grey).
- Underlying source is unraveled by averaging the aligned signals (Blue).

Experiments - Synthetic Data

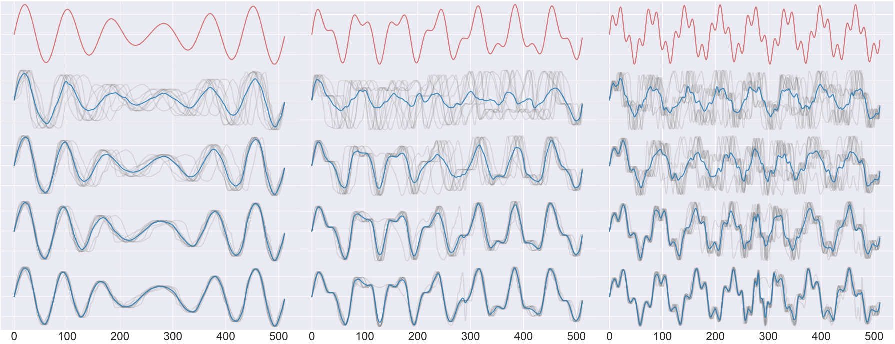
- RDTAN Joint Alignment of Synthetic Data.



- Latent source signals are known (Red).
- Data - source signals warped with random transformations (Grey).
- Underlying source is unraveled by averaging the aligned signals (Blue).

Experiments - Synthetic Data

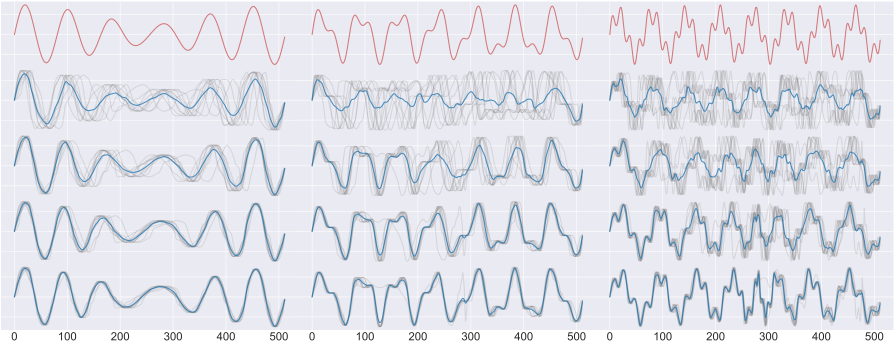
- RDTAN Joint Alignment of Synthetic Data.



- Latent source signals are known (Red).
- Data - source signals warped with random transformations (Grey).
- Underlying source is unraveled by averaging the aligned signals (Blue).

Experiments - Synthetic Data

- RDTAN Joint Alignment of Synthetic Data.



- Latent source signals are known (Red).
- Data - source signals warped with random transformations (Grey).
- Underlying source is unraveled by averaging the aligned signals (Blue).

Synthetic Data - Timing

Alignment timing per test set (in [sec])				
length \ # of signals	10	10^2	10^3	10^4
64	0.003	0.003	0.007	0.109
128	0.003	0.004	0.012	0.211
256	0.014	0.038	0.042	0.455
512	0.003	0.007	0.084	0.660

- DTAN joint alignment timing w.r.t. signal's length and size of the test-set (16 sets in total).

Joint Alignment of Real-World data

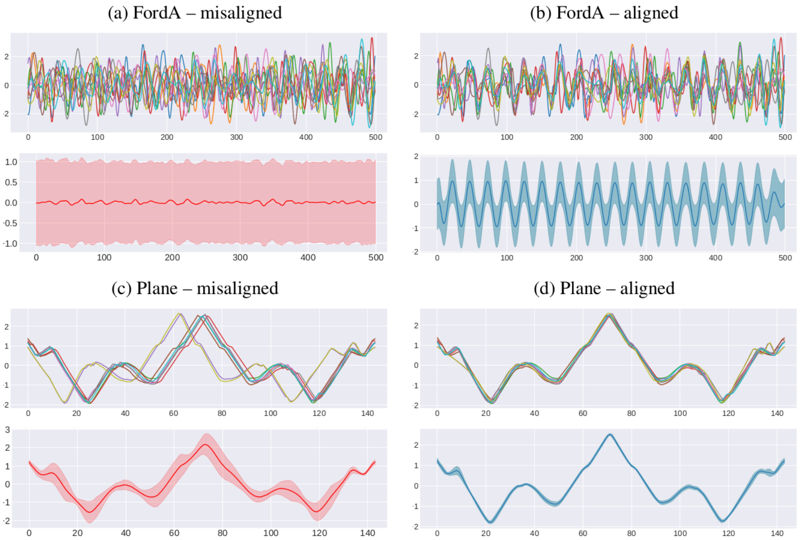
- The UCR time-series classification archive (Chen *et al*. 2015) contains 85 real-world datasets.
- ECGFiveDays dataset:

Joint Alignment of Real-World data

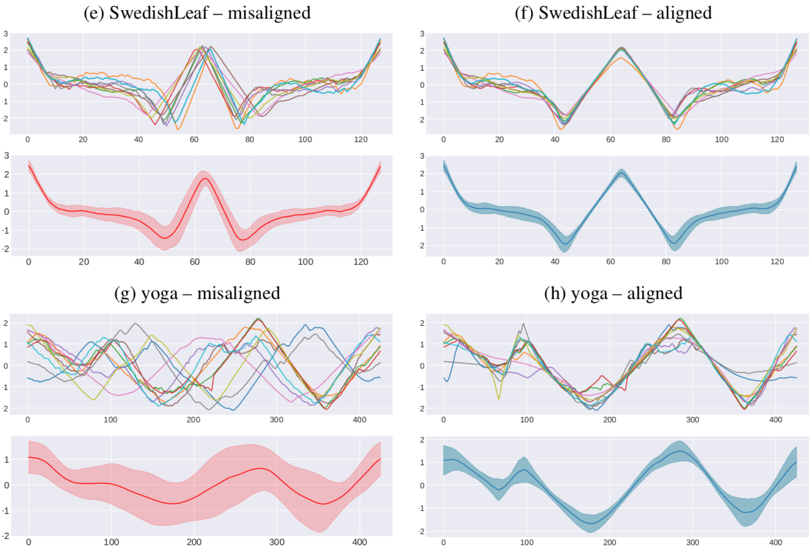
- The UCR time-series classification archive (Chen *et al.* 2015) contains 85 real-world datasets.
- ECGFiveDays dataset:

(a) Train (b) Test

Joint Alignment: More Results



Joint Alignment: More Results



Regularization effect - W/O

- Regularization term:

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

- Smoothness prior on the warps prevents unrealistic deformations.

Regularization effect - W/O

- Regularization term:

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

- Smoothness prior on the warps prevents unrealistic deformations.

Regularization effect - W/O

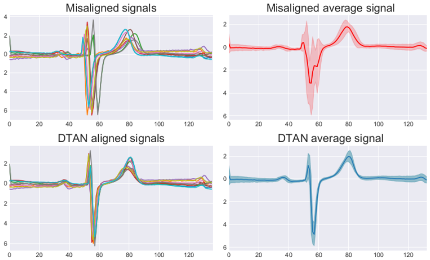
- Regularization term:

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

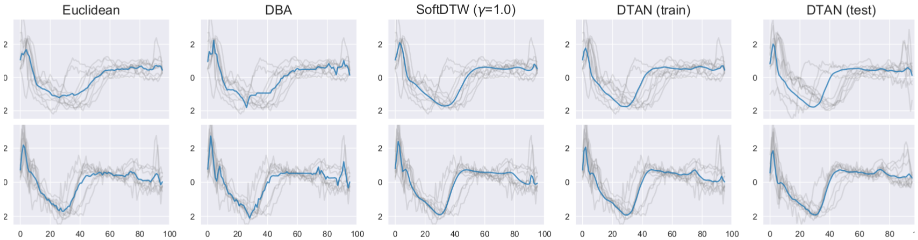
- Smoothness prior on the warps prevents unrealistic deformations.

(c) ECGFiveDays - Without regularization

(d) With regularization

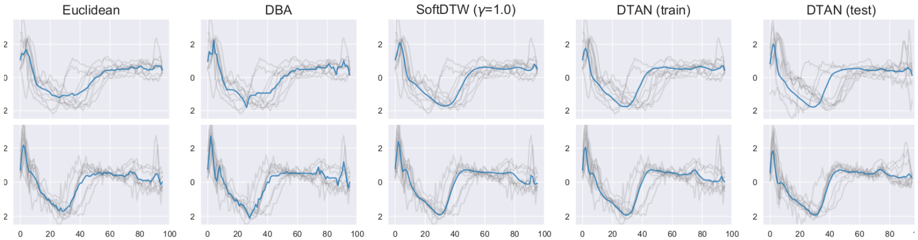


Time-Series Averaging Comparison



- Euclidean mean - serves as baseline ('ECG200' dataset).
- Dynamic Time Warping Barycenter Averaging - DBA [Petitjean, *Pattern Recognition* 2011] and SoftDTW [Cuturi, ICML 2017] ; No generalization, single class time-series averaging methods.

Time-Series Averaging Comparison



- Euclidean mean - serves as baseline ('ECG200' dataset).
- Dynamic Time Warping Barycenter Averaging - DBA [Petitjean, *Pattern Recognition* 2011] and SoftDTW [Cuturi, ICML 2017] ; No generalization, single class time-series averaging methods.

Nearest Centroid Classification (NCC) experiment

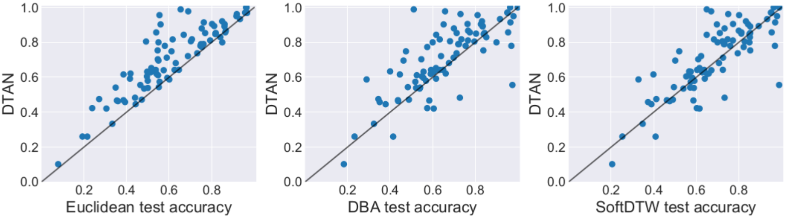
- NCC - 1-NN, using each class mean signal as the train set.
- Compared DTAN NCC while using euclidean distance to DBA and SoftDTW while measuring DTW distance.
- DTANs test accuracy compared with: Euclidean (93% of the datasets), DBA (77%) and SoftDTW (62%).

Nearest Centroid Classification (NCC) experiment

- NCC - 1-NN, using each class mean signal as the train set.
- Compared DTAN NCC while using euclidean distance to DBA and SoftDTW while measuring DTW distance.
- DTANs test accuracy compared with: Euclidean (93% of the datasets), DBA (77%) and SoftDTW (62%).

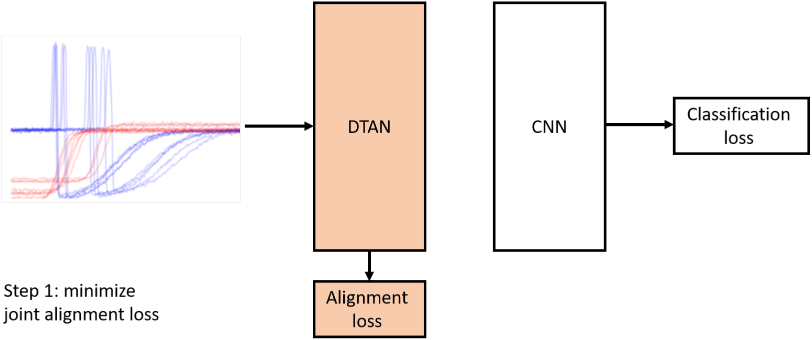
Nearest Centroid Classification (NCC) experiment

- NCC - 1-NN, using each class mean signal as the train set.
- Compared DTAN NCC while using euclidean distance to DBA and SoftDTW while measuring DTW distance.



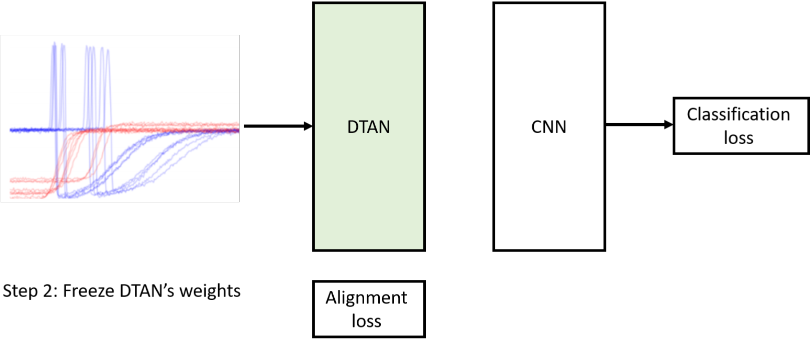
- DTANs test accuracy compared with: Euclidean (93% of the datasets), DBA (77%) and SoftDTW (62%).

CNN Classification Experiment



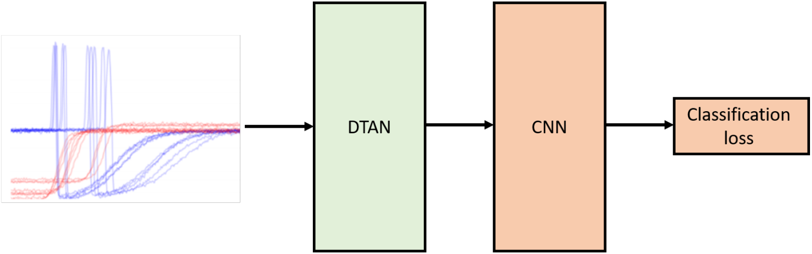
Step 1: minimize
joint alignment loss

CNN Classification Experiment



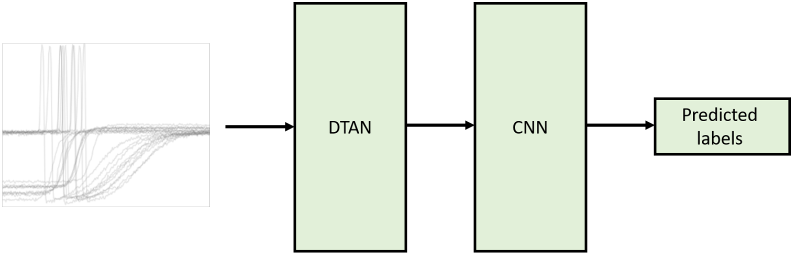
Step 2: Freeze DTAN's weights

CNN Classification Experiment

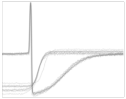


Step 3: Connect DTAN to classification network.
Minimize classification loss (i.e., cross-entropy)

CNN Classification Experiment



Step 4: use entire model – DTAN-CNN
For classification.



CNN Classification Results

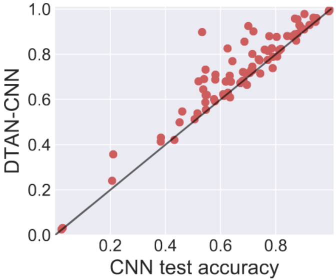
- DTAN-CNN compared with the same CNN without DTAN.
- DTAN-CNN achieved higher, or equal to, correct classification rates on 87% of the datasets (average test accuracy of 5 runs per set).

CNN Classification Results

- DTAN-CNN compared with the same CNN without DTAN.
- DTAN-CNN achieved higher, or equal to, correct classification rates on 87% of the datasets (average test accuracy of 5 runs per set).

CNN Classification Results

- DTAN-CNN compared with the same CNN without DTAN.
- DTAN-CNN achieved higher, or equal to, correct classification rates on 87% of the datasets (average test accuracy of 5 runs per set).

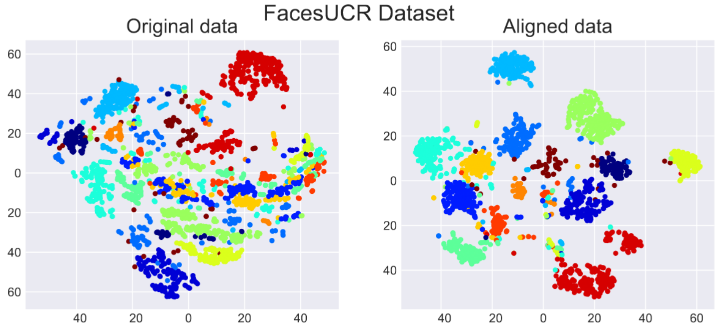


CNN Classification Results

- t-SNE visualization of the original and aligned data, illustrates how DTAN decreases intra-class variance while increasing inter-class one, thus improving the performance of classification net.

CNN Classification Results

- t-SNE visualization of the original and aligned data, illustrates how DTAN decreases intra-class variance while increasing inter-class one, thus improving the performance of classification net.



Conclusion

- Goal: statistical analysis of time-series data.
- Problem: temporal misalignment confounds statistical analysis.
- Our proposed solution: Diffeomorphic Temporal Alignment Nets (DTAN).
- Unsupervised/Semi-supervised DL framework for time-series joint alignment and averaging.
- Additional contributions: Smoothness prior, RDTAN and DTAN-CNN.
- Our code is publicly available at: https://github.com/BGU-CS-VIL/dtan

Thank You

Conclusion

- Goal: statistical analysis of time-series data.
- Problem: temporal misalignment confounds statistical analysis.
- Our proposed solution: Diffeomorphic Temporal Alignment Nets (DTAN).
- Unsupervised/Semi-supervised DL framework for time-series joint alignment and averaging.
- Additional contributions: Smoothness prior, RDTAN and DTAN-CNN.
- Our code is publicly available at: https://github.com/BGU-CS-VIL/dtan

Thank You

Conclusion

- Goal: statistical analysis of time-series data.
- Problem: temporal misalignment confounds statistical analysis.
- Our proposed solution: Diffeomorphic Temporal Alignment Nets (DTAN).
- Unsupervised/Semi-supervised DL framework for time-series joint alignment and averaging.
- Additional contributions: Smoothness prior, RDTAN and DTAN-CNN.
- Our code is publicly available at: https://github.com/BGU-CS-VIL/dtan

Thank You

Conclusion

- Goal: statistical analysis of time-series data.
- Problem: temporal misalignment confounds statistical analysis.
- Our proposed solution: Diffeomorphic Temporal Alignment Nets (DTAN).
- Unsupervised/Semi-supervised DL framework for time-series joint alignment and averaging.
- Additional contributions: Smoothness prior, RDTAN and DTAN-CNN.
- Our code is publicly available at: https://github.com/BGU-CS-VIL/dtan

Thank You

Conclusion

- Goal: statistical analysis of time-series data.
- Problem: temporal misalignment confounds statistical analysis.
- Our proposed solution: Diffeomorphic Temporal Alignment Nets (DTAN).
- Unsupervised/Semi-supervised DL framework for time-series joint alignment and averaging.
- Additional contributions: Smoothness prior, RDTAN and DTAN-CNN.
- Our code is publicly available at: https://github.com/BGU-CS-VIL/dtan

Thank You

Conclusion

- Goal: statistical analysis of time-series data.
- Problem: temporal misalignment confounds statistical analysis.
- Our proposed solution: Diffeomorphic Temporal Alignment Nets (DTAN).
- Unsupervised/Semi-supervised DL framework for time-series joint alignment and averaging.
- Additional contributions: Smoothness prior, RDTAN and DTAN-CNN.
- Our code is publicly available at: https://github.com/BGU-CS-VIL/dtan

Thank You

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/.

M. Cuturi and M. Blondel. Soft-dtw: a differentiable loss function for time-series. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 894–903. JMLR. org, 2017.

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.

F. Petitjean, A. Ketterlin, and P. Gançarski. A global averaging method for dynamic time warping, with applications to clustering. *Pattern Recognition*, 44(3):678–693, 2011.