Time Series Data Augmentation for Neural Networks by Time Warping with a Discriminative Teacher

Brian Kenji Iwana and Seiichi Uchida
Kyushu University

KYUSHU UNIVERSITY

One Problem with Time Series Recognition and Deep Learning

- Time series classification datasets are often small.
- This is not ideal for modern deep learning methods.
- Example:
  The UCR Time Series Archive contains 128 datasets.
  But, only **12 out of the 128** have more than 1,000 training patterns (and 1,000 is tiny).

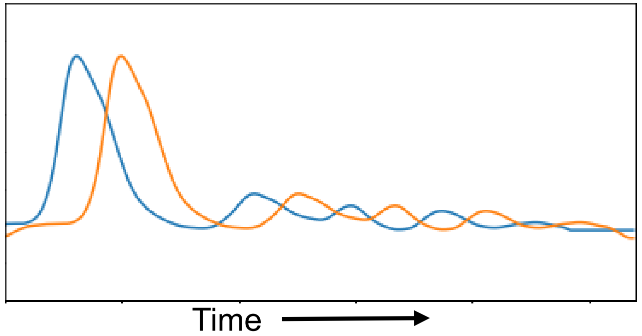
Time Series Data Augmentation

- In **image** recognition, data augmentation is normal and is used in most SotA methods.
  - Examples
    - **VGG** – random scaling, cropping, and color augmentation
    - **AlexNet** – cropping, mirroring, and color augmentation
    - **ResNet** – scaling, cropping, and color augmentation
    - **DenseNet** – translation and mirroring
    - **Inception** – cropping and mirroring
- By comparison, outside of NLP, data augmentation in **time series** recognition isn't as widely used and there are no standard methods.

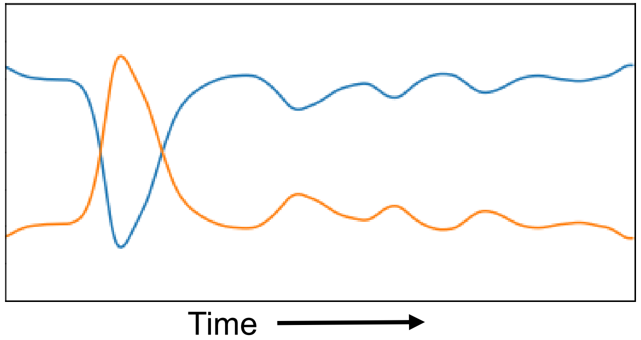
Random Transformations

Most of the commonly used **time series** data augmentations are adaptations borrowed from **image** recognition. Examples:

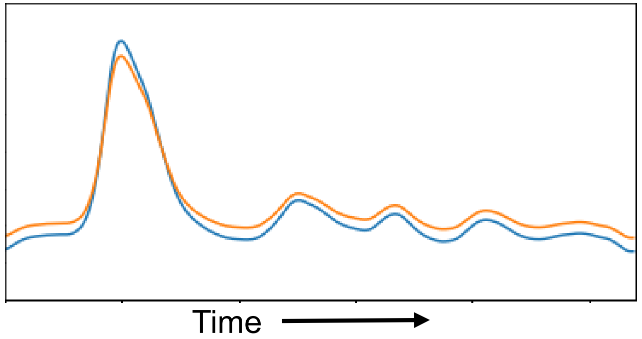
Window Slicing (Crop)



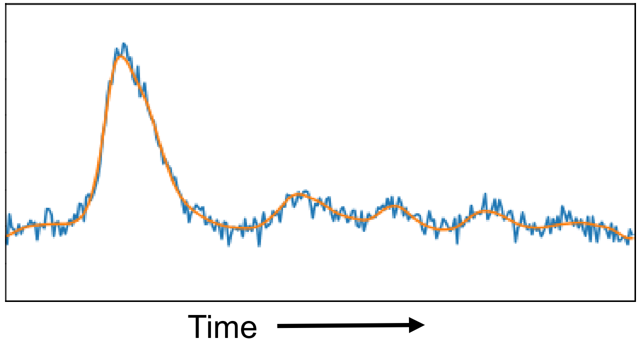
Rotation (Mirroring)



Scaling



Jittering (Noise)



Orange is original
Blue is generated

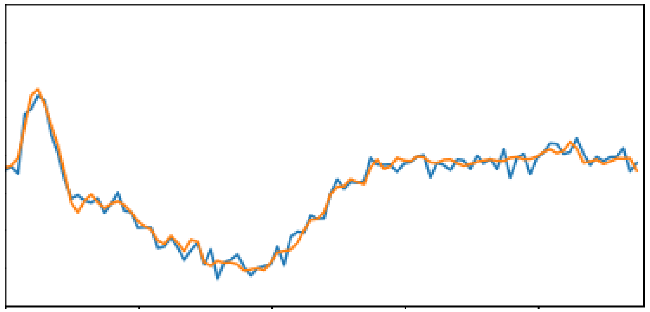
Problem with Random Transformations

The assumptions of each random transformation doesn’t always hold for every **time series** domain.

Ex: Jittering (adding nose) doesn’t always apply

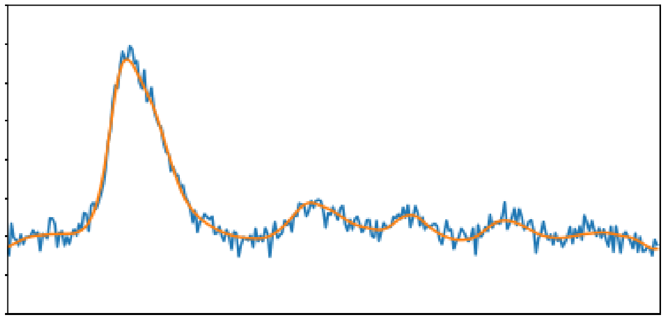
UCR ECG200

- Noise can be expected in ECG patterns



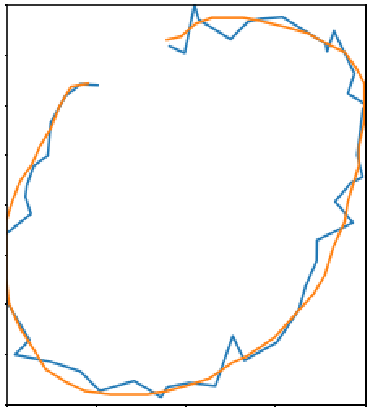
UCR 50words

- In this dataset, all the patterns are smooth



Unipen 1b, the letter “O”

- Jittering doesn’t make as much sense for pen-tip



Orange is original
Blue is generated

Other Time Series DA Methods

One solution is to generate patterns from the distribution of features in the training set

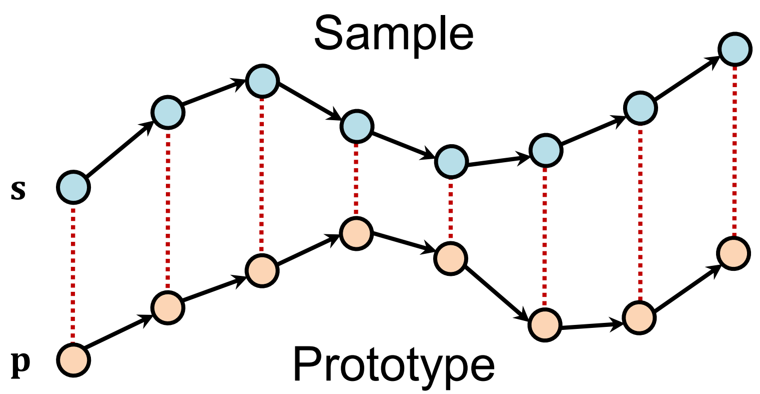
Examples:

- GANs
- Domain specific generative models
- Pattern mixing
- Signal decomposition
- etc.

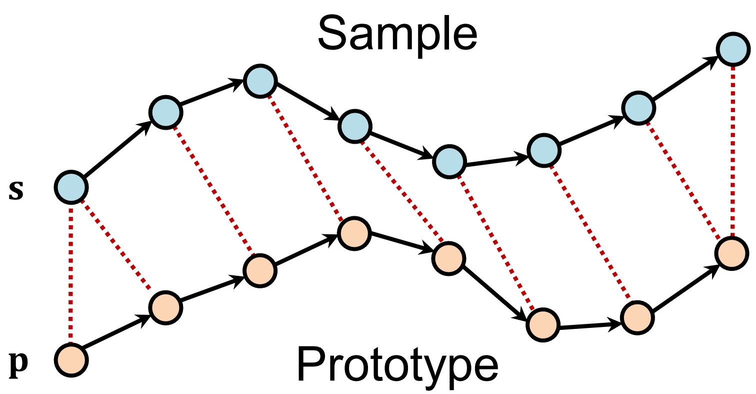
Guided Warping For Time Series Data Augmentation

Dynamic Time Warping (DTW)

Dynamic time warping (DTW) uses **dynamic programming** to match elements of time series. It is typically used as a global distance measure, but we use it for the **dynamic alignment** property.



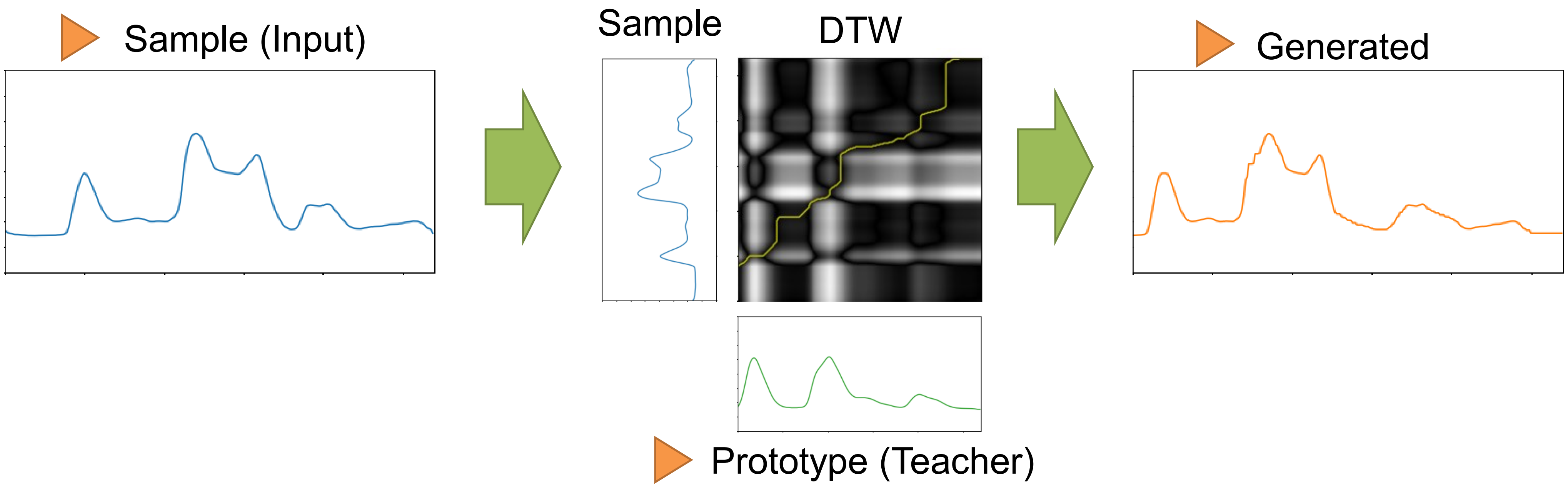
Linear Matching



Dynamic Matching

Guided Warping (Proposed)

In this work, we exploit the alignment property of DTW to align the elements of two time series.

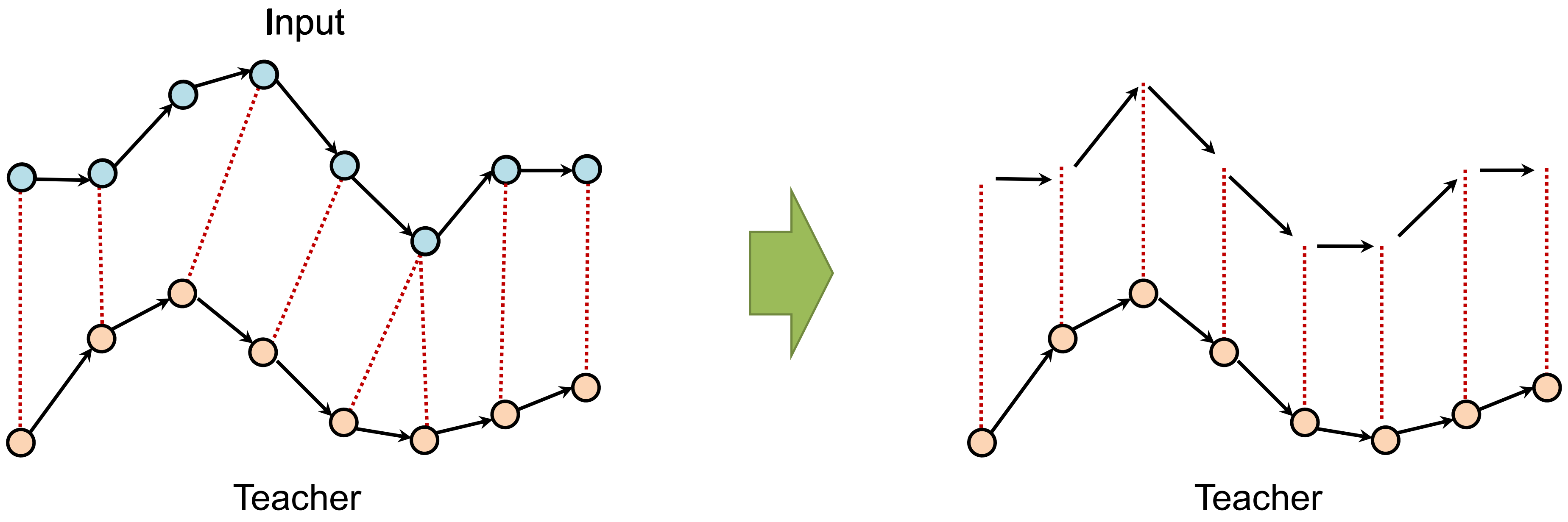


Note: the patterns are only warped in the time domain.

Guided Warping (Proposed)

Specifically, the elements of the input sample are warped to the time steps of the teacher sample using DTW.

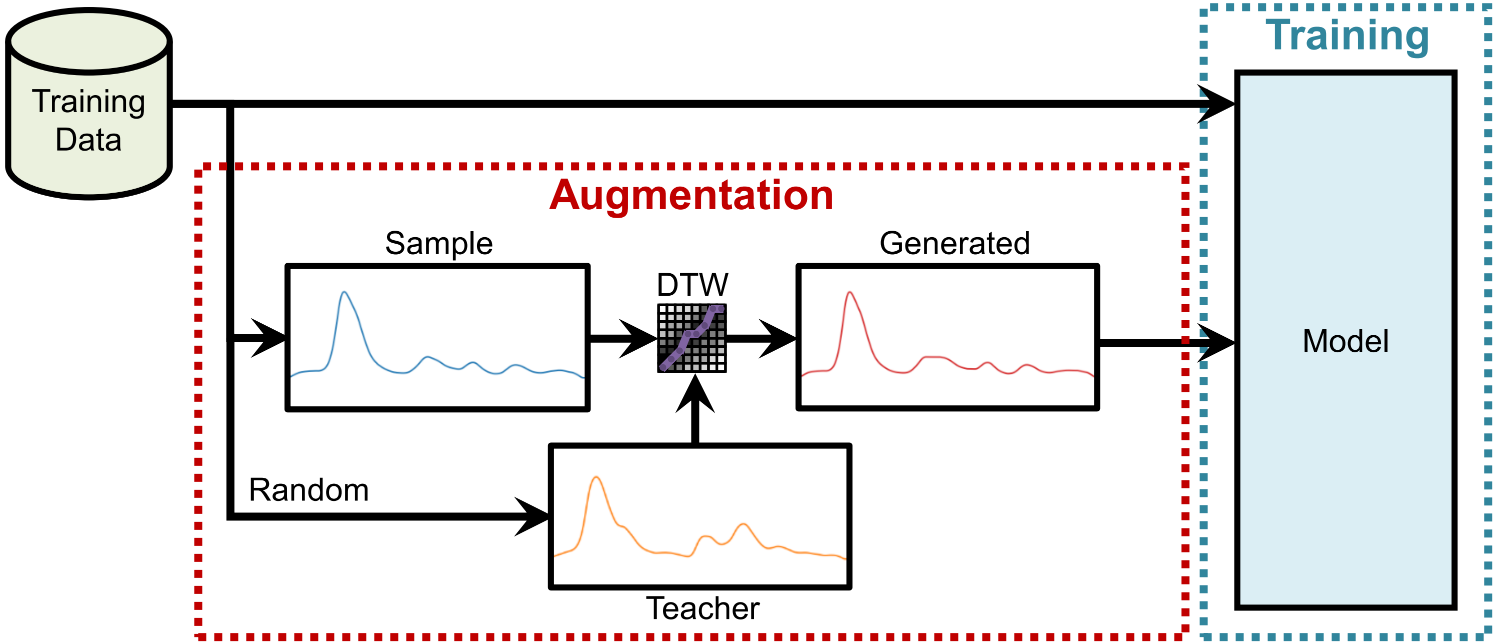
In this way, the generated pattern has the features of the input sample at the time steps of the teacher sample.



Guided Warping (Proposed)

In order to generate new samples, random samples from the same class are paired and then warped.

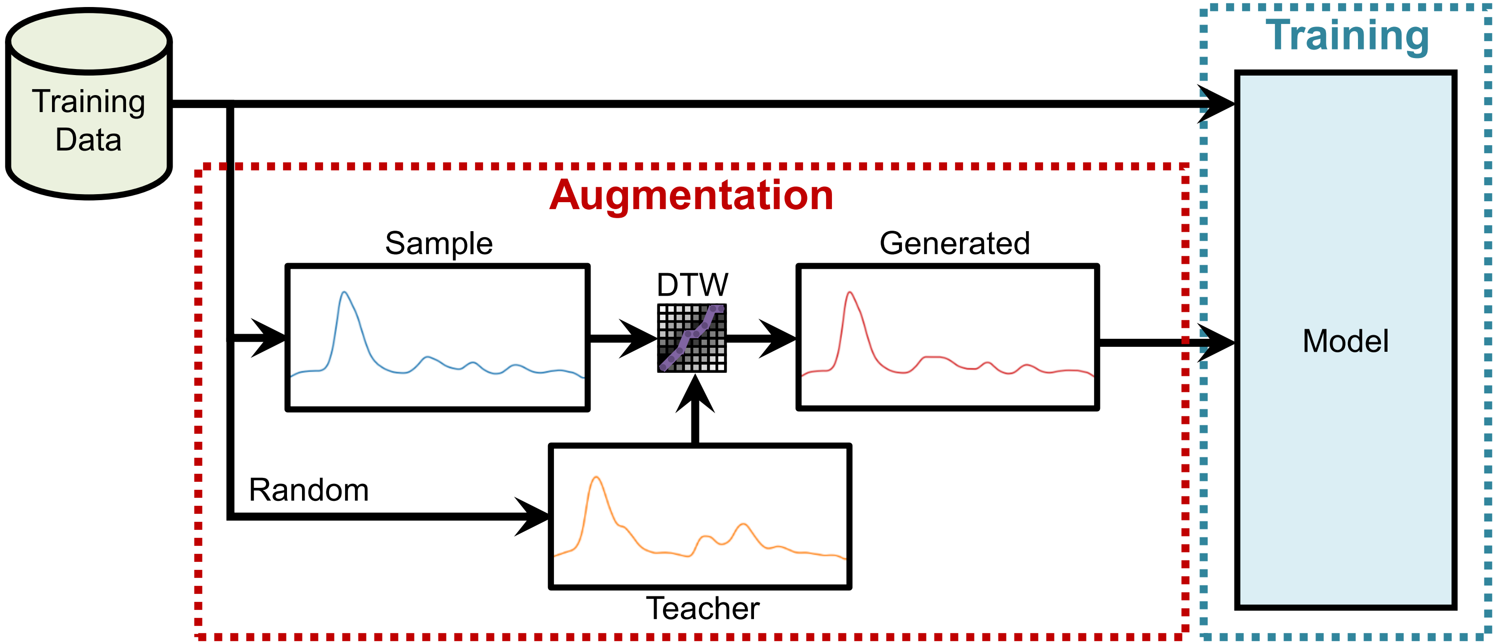
We can then use these new samples as augmented data.



Problem: Guided Warping

The problem is, just like many **pattern mixing** methods, the two intra-class samples are randomly selected.

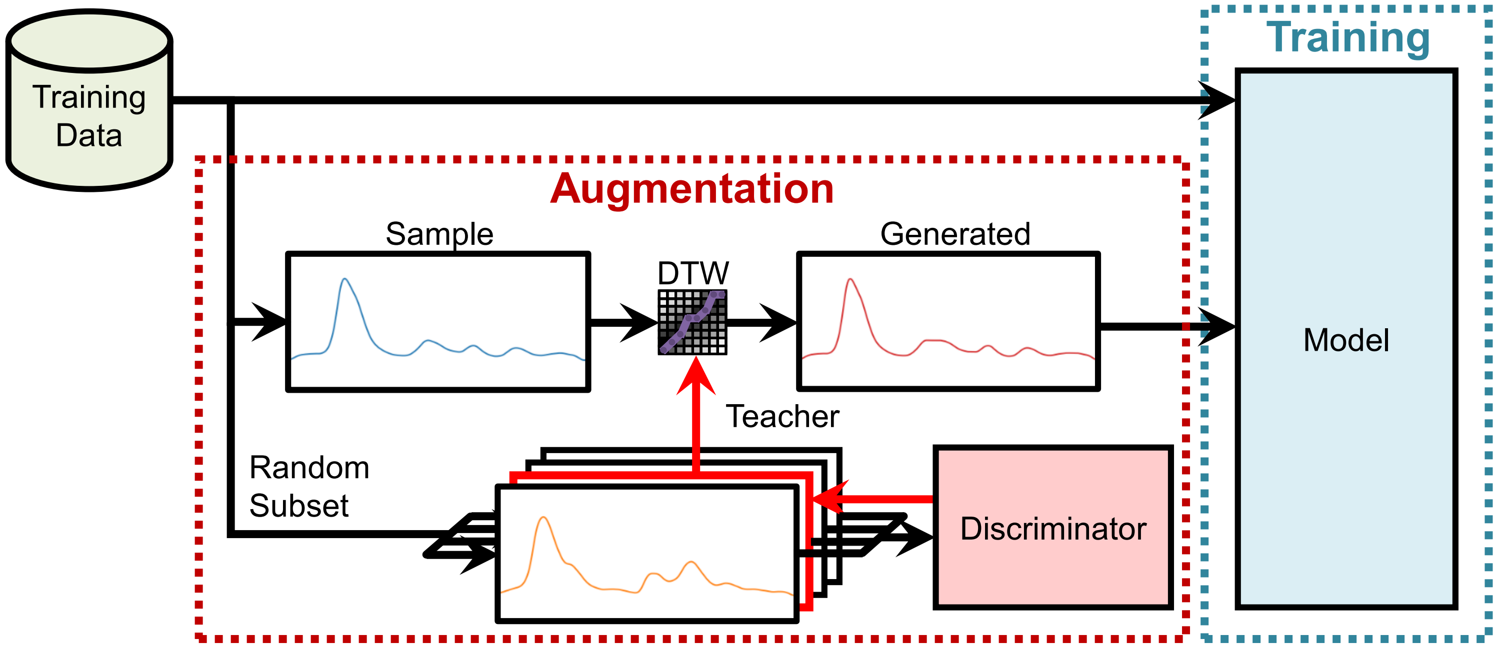
There is no consideration to which patterns are mixed.



Discriminative Guided Warping (Proposed 2)

Thus, we propose taking a small subset of the training data and use a **nearest centroid classifier** based on DTW distance to determine the **most discriminative pattern**.

This pattern is used as the teacher pattern.



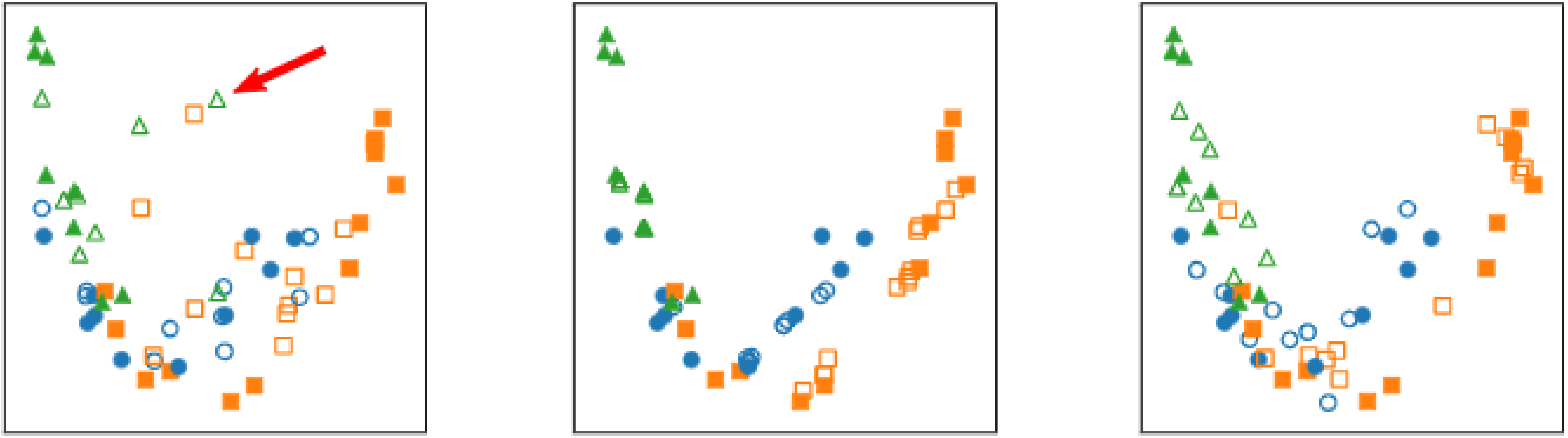
Results

Average of all 88 of the 2015 UCR TS Archive datasets

Method	CNN	LSTM
No Augmentation	76.44	57.24
Jittering	77.32	58.35
Rotation	74.84	54.78
Scaling	77.06	57.98
Magnitude Warping	78.30	58.04
Time Warping	78.10	52.80
Slicing	79.15	54.49
Window Warping	79.58	57.49
SPAWNER	78.84	**58.98**
wDBA	77.42	56.01
RGW-D (Proposed)	79.39	57.42
RGW-sD (Proposed)	79.27	56.43
DGW-D (Proposed)	80.12	56.01
DGW-sD (Proposed)	**80.17**	56.99

Comparison to Other Pattern Mixing Methods

SPAWNER | wDBA ASD | DGW-sD (Proposed)



- Strange patterns
- Redundant patterns
- More useful patterns for DA

Real | Generated

UCR CBF data

Summary

- We proposed a new data augmentation method for time series called guided warping. Guided warping warps the elements of one time series to the time steps of another using DTW.
- We demonstrated that the use of a discriminator can improve the performance of guided warping by selecting the most discriminate teacher.
- We evaluated the proposed method on 88 UCR time series datasets using a CNN and an LSTM.