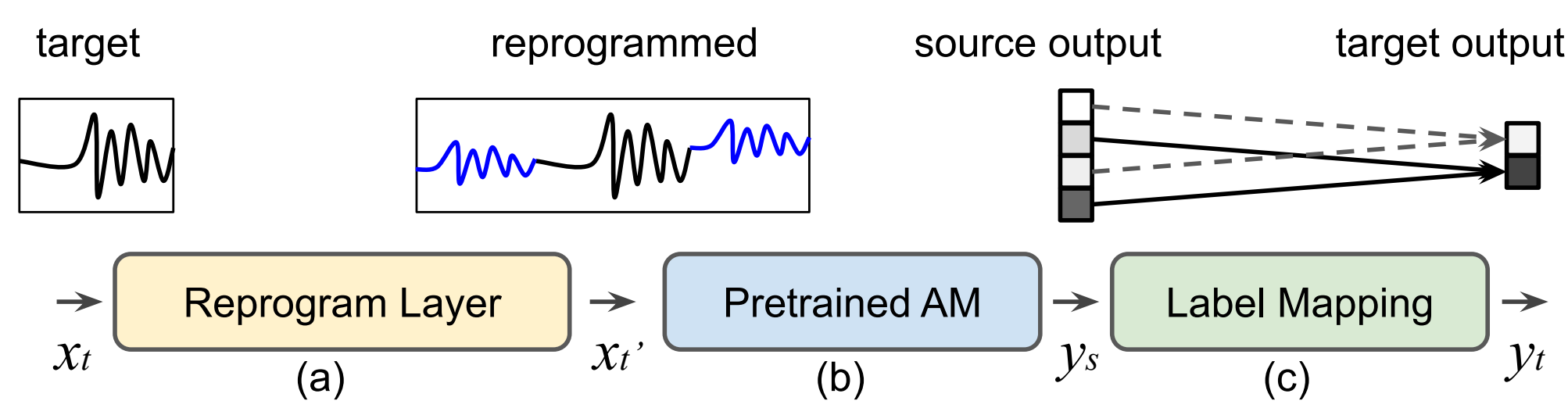


## OVERVIEW

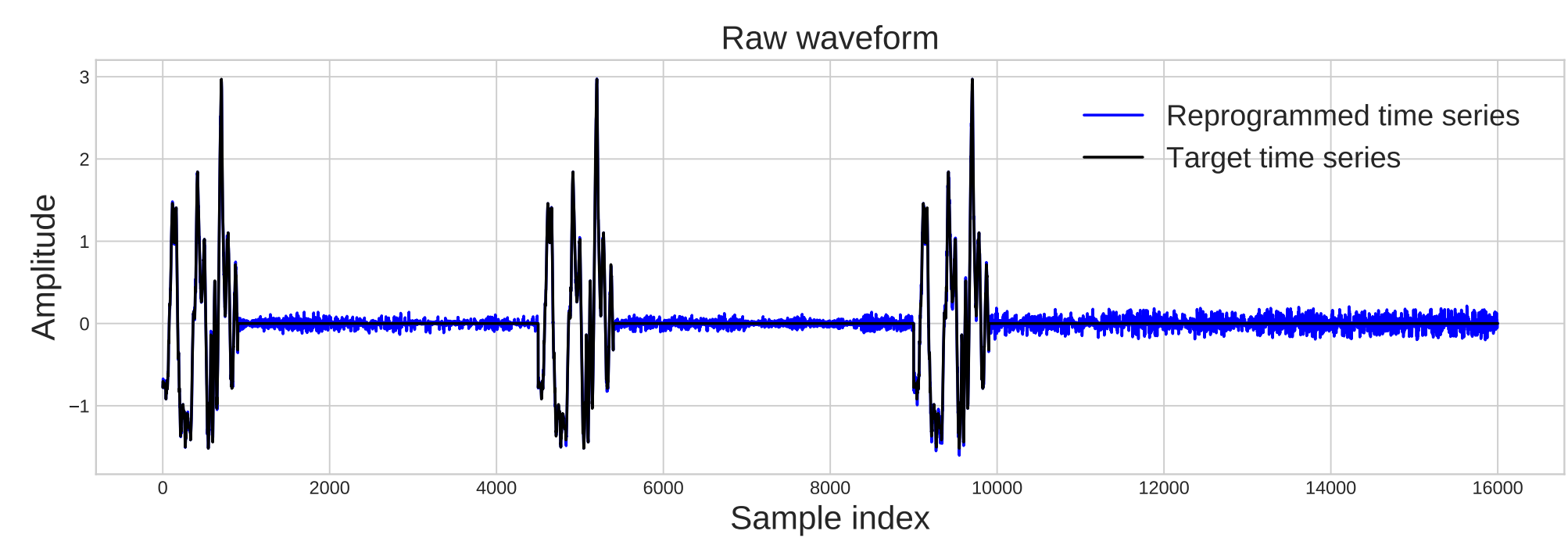
We propose V2S, an unified approach to reprogram large-scale pre-trained acoustic models for different time series classification tasks. To the best of our knowledge, V2S is the first framework that enables reprogramming for time series tasks.



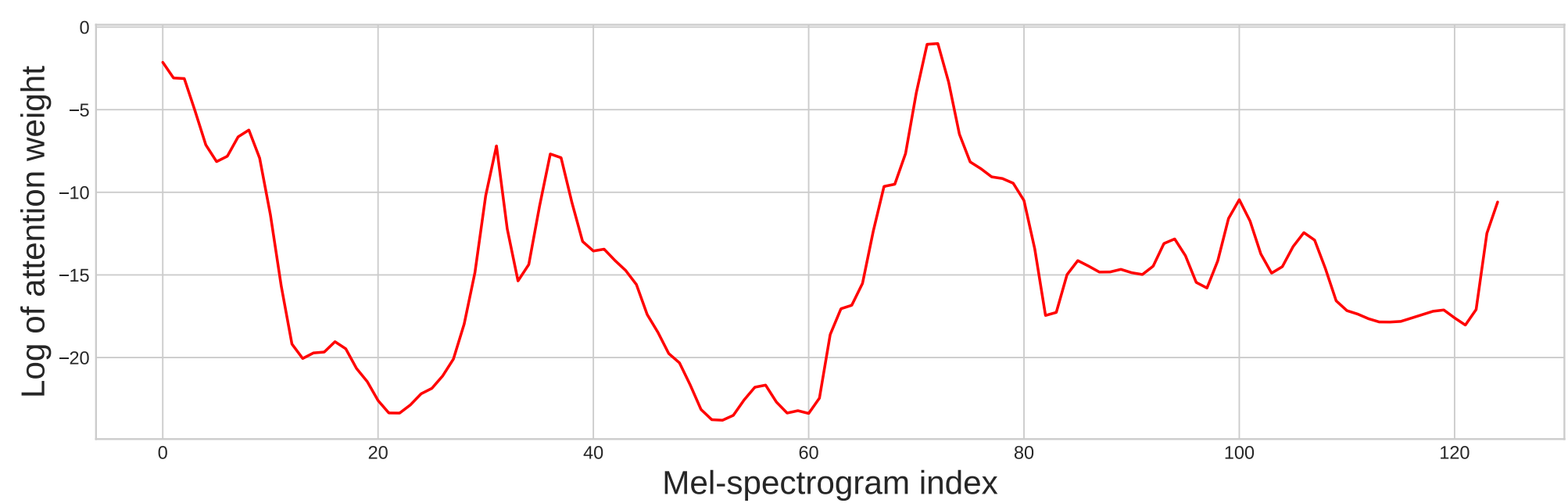
**Figure 1:** Schematic illustration of the proposed Voice2Series (V2S) framework: (a) trainable reprogram layer; (b) pre-trained acoustic model (AM); (c) source-target label mapping function.

## VISUALIZE V2S ATTENTION

We visualize the sequence-level attention weights of reprogrammed AMs (non-trainable in V2S).



**Figure 4:** Targeted (blue) and reprogrammed (black).



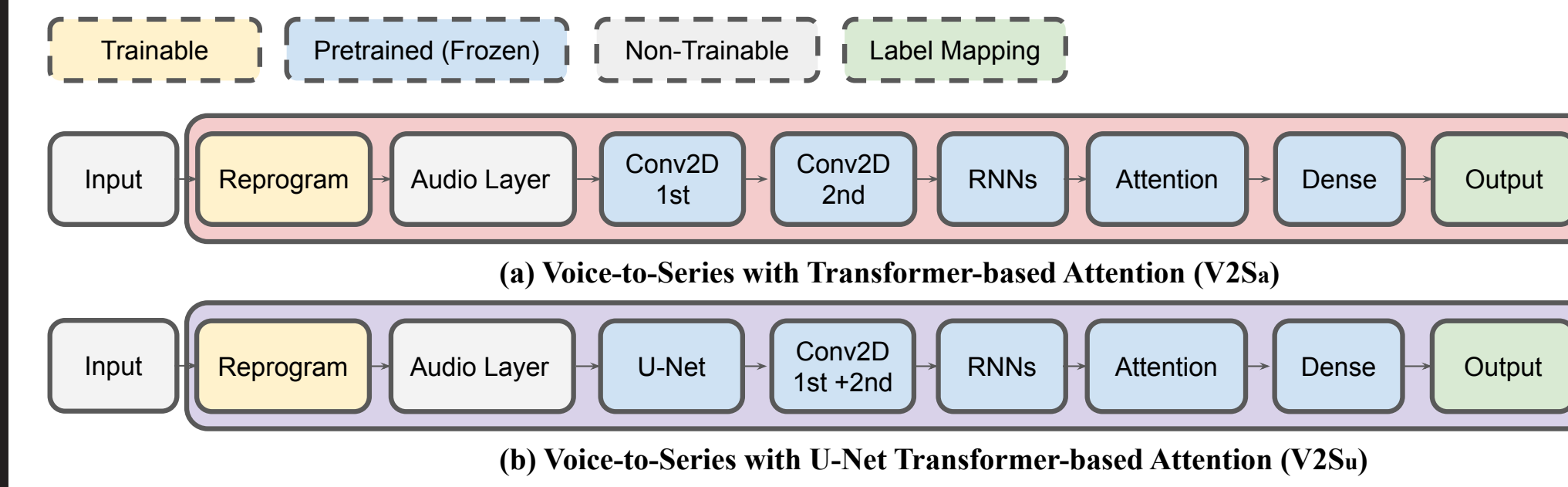
**Figure 5:** Attention weight of reprogrammed input.

## ICLR WORKSHOP 2021

[1] Yang et al. "voice2series: Reprogramming acoustic models for time series classification". *ICLR S2D-OLAD Workshop, Spotlight, 2021.*

## V2S ARCHITECTURE

For training the source model, we use a self-attention architecture for V2S reprogramming.

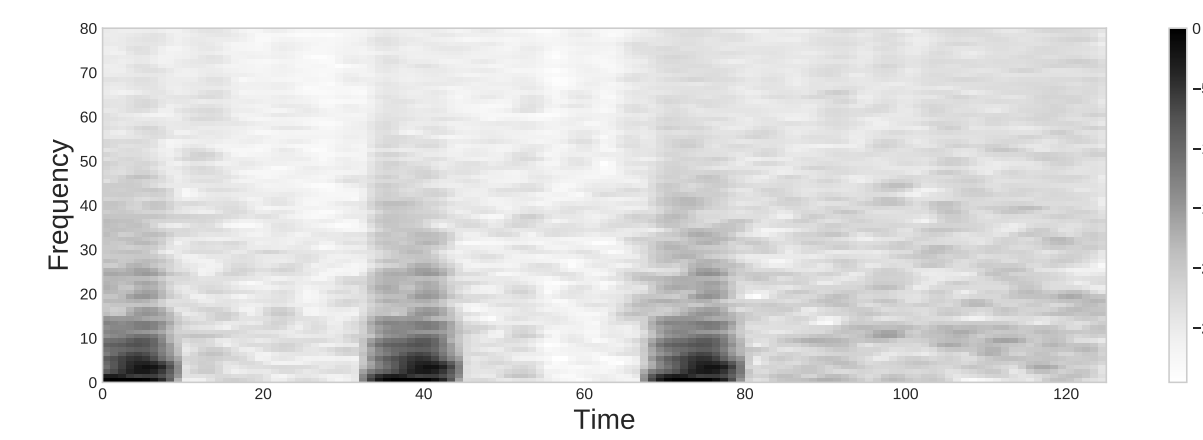


- (a) V2S with self-attention Bi-LSTM.
- (b) V2S with U-Net self-attention Bi-LSTM.

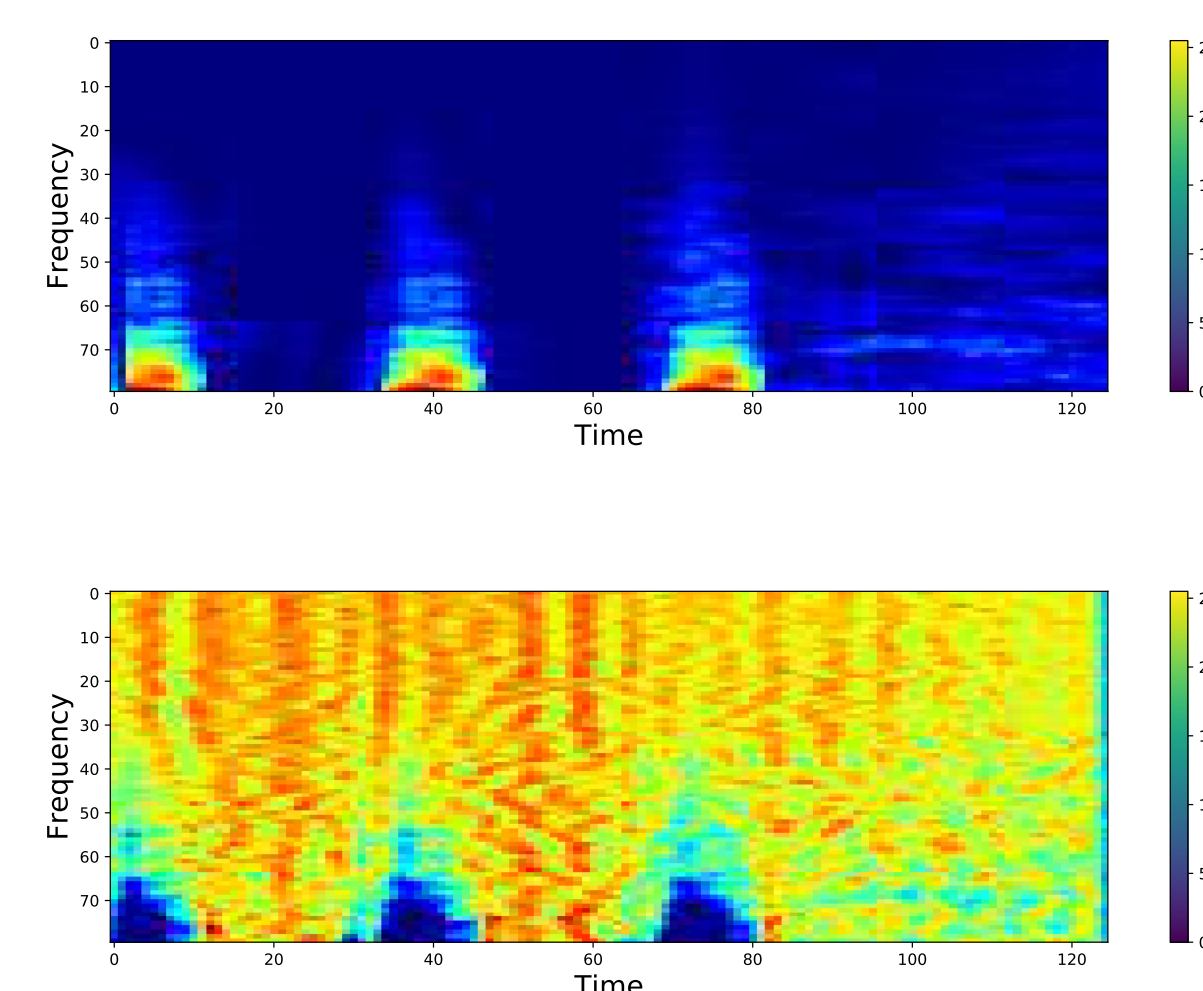
We evaluate different AMs and dataset, validate that V2S is a general framework, and further provide a theoretical justifications on reprogramming. V2S implementation and pretrained models will open source to the community.

## NEURAL SALIENCY ANALYSIS

Class activation mapping (CAM) evaluates neural AM on Mel-features with the proposed V2S.



**Figure 6:** Mel-spectrogram of reprogrammed input.



**Figure 7:** CAM result from different CNN layer of AM.

## CONCLUSION AND FUTURE RESEARCH

The proposed V2S shows competitive results for time series classification; we also provide the first theoretical justification on reprogramming.

## POPULATION RISK VIA REPROGRAMMING

- We consider the RMSE denoted by  $\|f(x) - y\|_2$  for a given neural network classifier  $f$ .

1. The source risk is  $\epsilon_S: \mathbb{E}_{\mathcal{D}_S}[\ell(x_s, y_s)] = \epsilon_S$ .
2. The source-target label space has a specified surjective one-to-one label mapping function  $h_t$  for every target label  $t$ , such that  $\forall y_t \in \mathcal{Y}_T, y_t = h_t(\mathcal{Y}_S) \triangleq y_s \in \mathcal{Y}_S$ .
3. Based on reprogramming, the target loss function  $\ell_T$  with an additive input transformation function  $\delta$  can be represented as  $\ell_T(x_t + \delta, y_t) \stackrel{(a)}{=} \ell_T(x_t + \delta, y_s) \stackrel{(b)}{=} \ell_S(x_t + \delta, y_s)$ , where (a) is induced by label mapping (Assumption 2) and (b) is induced by reprogramming the source loss with target data.
4.  $\delta^* \triangleq \arg \min_{\delta} \mathbb{E}_{\mathcal{D}_T}[\ell_S(x_t + \delta, y_s)]$  is the minimizer of the target population risk with the

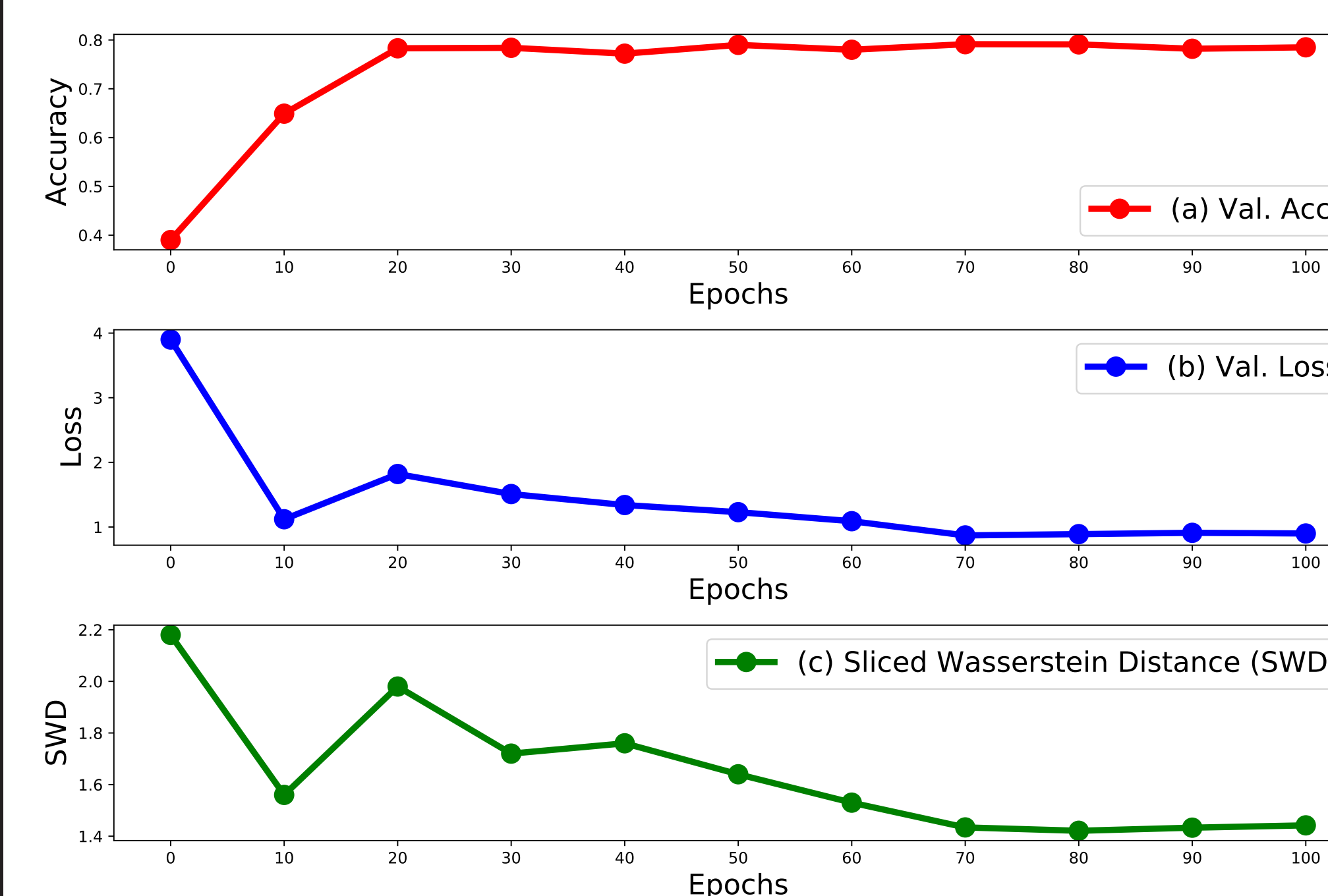
reprogramming loss objective.

5. Domain-independent drawing of source and target data: Let  $\Phi_S(\cdot)$  and  $\Phi_T(\cdot)$  denote the probability density function of source data and target data distributions over  $\mathcal{X}_S$  and  $\mathcal{X}_T$ . The joint probability density function is the product of their marginals, i.e.,  $\Phi_{S,T}(x_s, x_t) = \Phi_S(x_s) \cdot \Phi_T(x_t)$ .

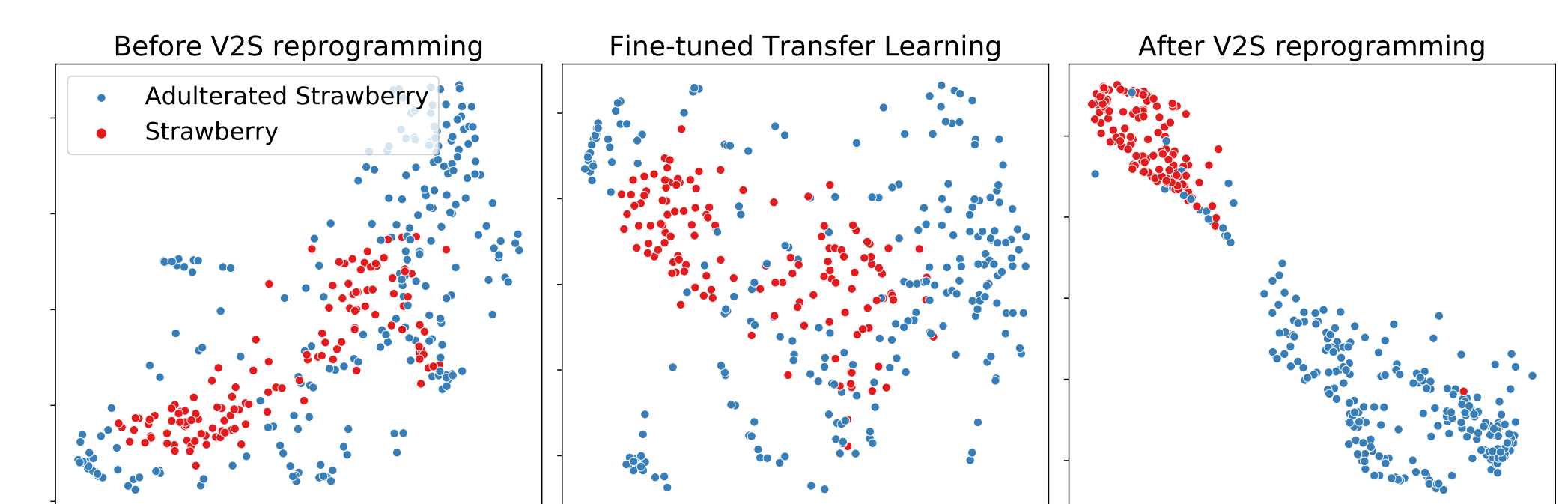
**Theorem 1:** The population risk for the target task via reprogramming a  $K$ -way source neural network classifier  $f_S(\cdot) = \eta(z_S(\cdot))$ , denoted by  $\mathbb{E}_{\mathcal{D}_T}[\ell_T(x_t + \delta^*, y_t)]$ , is upper bounded by

$$\mathbb{E}_{\mathcal{D}_T}[\ell_T(x_t + \delta^*, y_t)] \leq \underbrace{\epsilon_S}_{\text{source risk}} + 2\sqrt{K} \cdot \underbrace{\mathcal{W}_1(\mu(z_S(x_t + \delta^*)), \mu(z_S(x_s)))}_{\text{representation alignment loss via reprogramming}}$$

## PERFORMANCE DISCUSSION



**Figure 2:** Training-time reprogramming analysis using proposed V2S and DistalPhalanxTW dataset. All values are averaged over the training set. The rows are (a) validation (test) accuracy, (b) validation loss, and (c) sliced Wasserstein distance (SWD).



**Figure 3:** tSNE plots of the logit representations using the Strawberry training set and V2S, for the cases of before and after V2S reprogramming, and fine-tuned transfer learning from the same AM.

**Performance:** In UCR time series archive, V2S outperforms or is tied with the previous reported results on 22 datasets out of with 31 different tasks and improves their average accuracy by 1.72%.

**Model Selection:** Based on Theorem 1, one can leverage our derived risk bound for V2S model selection and AM design for reprogramming.

## MORE INFORMATION

**Email** [huckkiyang@gatech.edu](mailto:huckkiyang@gatech.edu)

**Note** This is a non-archival and preliminary venue. The full version is under review.

Our future work includes incorporating different time series and acoustic processing tasks. We will open source our code to the community.