INCM

Sathvika Miryala , Rishab Bhattacharya

November 2024

1 Aim

To evaluate the effectiveness of Latent Predictive Learning (LPL) in producing invariant representations without collapse in sensory networks and replicating neuronal selectivity changes observed in the primate inferotemporal cortex under altered temporal conditions. This work further extends LPL to auditory stimuli and spiking neural networks (SNNs) to assess its generalizability and limitations.

2 Introduction and Motivation

In biological systems, sensory networks can consistently recognize objects despite variations in presentation, such as changes in orientation or lighting. This ability has inspired models in visual domains. However, artificial neural networks, though effective, rely on backpropagation—a mechanism not proven feasible in biological systems. This limitation highlights the need for alternative learning mechanisms to model brain-like invariant object representations.

Using Hebbian and BCM rules alone for learning has not yielded good results, as they often lead to representational collapse. To address this, the model combines Hebbian learning with predictive terms, designing a local learning rule for sensory networks that shows promising results. Previous research has shown that Hebbian learning alone leads to representational collapse because it lacks a mechanism to counterbalance the pull effect of features. However, this approach adds a predictive term (derived from latent inputs), which generates an opposite push effect by capturing slowly changing features in the input, thereby preventing the collapse of representations.

While the original study focused on visual stimuli, real-world sensory processing often requires integrating information from multiple modalities. This project extends the model to analyze its performance with audio stimuli. We aim to examine the generalizability of the LPL approach and its relevance to auditory perception tasks, thereby advancing our understanding of how neural networks might handle multimodal sensory inputs.

3 Methods

To demonstrate that Latent Predictive Learning (LPL) produces disentangled representations in sensory networks without representational collapse, we implemented a series of experiments involving .

Architecture

The model architecture we designed and used for experiment is a **convolutional neural network** with 2 layers and relu as the activation function which is very similar to the brain operates on the signals it gets and then with 2 fully connected layers(embedding & output class predictions) **Combined Learning Rule** we used

$$\Delta W = \eta \cdot (\Delta W_{\text{hebbian}} + \Delta W_{\text{predictive}})$$

$$\Delta W_{\text{hebbian}} = \alpha \cdot x^T \cdot y_{\text{true}}$$

$$\Delta W_{\text{predictive}} = \beta \cdot x^T \cdot (y_{\text{pred}} - y_{\text{true}})$$

Here, x represents the flattened embedding from the CNN, y_{true} is the one-hot encoded target, and y_{pred} is the softmax output of the network. The coefficients α and β were set to 0.1 and 0.01, respectively, as optimized hyperparameters.

Data sets

We utilized the **MNIST dataset**, includes 60,000 training and 10,000 testing grayscale images of handwritten digits, resized to 64×64 for compatibility with the CNN architecture.

Neuronal selectivity in IT Cortex

For the understanding of the ability of LPL to replicate the results performed on the primates on their Inferotemporal cortex neuronal selectivity we used the following:

We used a synthetic dataset based on Moving MNIST to simulate swap and nonswap conditions. Each sequence in the dataset consists of two MNIST digits: a preferred digit (P) and a non-preferred digit (N), moving across a 64×64 canvas. The digits move with random initial positions and velocities to create dynamic input sequences. In the swap condition, at a designated frame (e.g., frame 10), the positions and identities of P and N digits are swapped, introducing an incorrect temporal association. In the nonswap condition, no swapping occurs, maintaining correct temporal associations throughout the sequence. Each sequence contains 20 frames, and we generated 500 sequences for both swap and nonswap conditions to ensure robust analysis. For each frame in a sequence, the network output was recorded. Responses of neurons corresponding to preferred (P) and non-preferred (N) digits were measured. Selectivity (S) was computed as:

$$S = P_{\text{response}} - N_{\text{response}}$$

Spiking Neural Networks Experiment

The spiking neural network (SNN) consists of 100 excitatory neurons and 25 inhibitory neurons. The excitatory neurons are modeled with a membrane potential that evolves based on input currents and resets upon firing, while the inhibitory neurons provide feedback to regulate the network's activity. The network receives input from five distinct input populations, each delivering time-varying signals via Poisson-distributed spike trains. The synaptic connections in the SNN include input synaptic weights (W_{input}), which are initialized with random values bounded between 0 and 5. Recurrent connections are divided into three types: excitatory-to-excitatory (W_{EE}), which are positive and sparse; excitatory-to-inhibitory (W_{EI}), which are also positive and sparse; and inhibitory-to-excitatory (W_{IE}), which are negative and sparse. Input signals are generated as sinusoidal waveforms, encoded into spike trains using Poisson statistics. These signals are simulated over 100,000 ms with a timestep (dt) of 1 ms. The resulting binary spike trains effectively mimic realistic neural activity patterns, enabling the network to process dynamic input signals.

Evaluation on Audio

This experiment aimed to classify speakers using audio data from the LibriSpeech dataset, leveraging a combination of deep learning methods and biologically inspired learning rules. The LibriSpeech dataset, specifically the train-clean-100 and test-clean subsets, was utilized for this purpose. To facilitate the handling of the data, a custom PyTorch Dataset class was implemented to load and preprocess the audio samples effectively. Initially, we tried with a maximum of 5 speakers was selected for classification experiments, which was later extended to include 41 speakers to assess performance.

The preprocessing pipeline for the audio data involved transforming the waveforms into Mel-spectrograms, converting amplitude values to decibels, resizing the inputs to a consistent dimension of 224×224 pixels, and normalizing them to ensure uniformity across the dataset.

4 Results

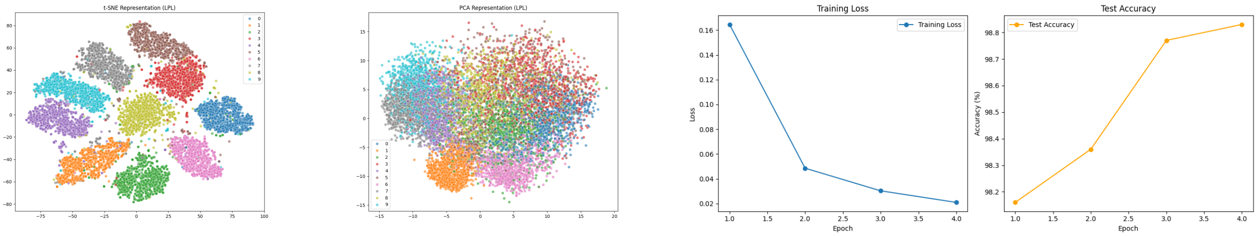
Experiment 1: Evaluating the Disentangled Representations formed using LPL.

After training, we analyzed the learned representations (x_{flat}) using **t-SNE** and **PCA** to evaluate their disentanglement properties. As shown in Figure 1, the LPL rule effectively created well-separated clusters, demonstrating its ability to disentangle input images. The results demonstrate that LPL effectively disentangles the input images, creating clear and distinct clusters in the embedding space.

Experiment 2: Neuronal selectivity by DNNs using LPL

Next, we simulated the experiment by Li & DiCarlo on primates and analyzed how well the constructed DNN immitated neuronal selectivity.

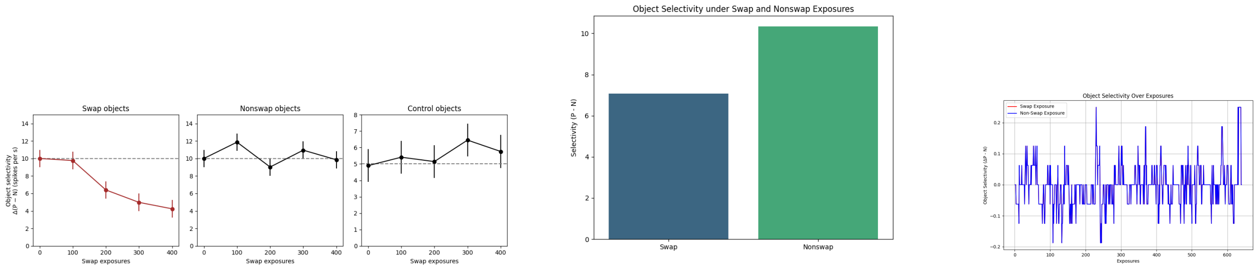
.As we can see in Fig 2, In the swap condition, average selectivity was significantly reduced, consistent with experimental findings. This reduction resulted from a combination of decreased responses to preferred stimuli (P) and increased responses to non-preferred stimuli (N) at the swap position. Conversely, in the nonswap condition, average selectivity remained stable, demonstrating the network's ability to preserve correct temporal associations when there were no disruptions in temporal continuity.Bar plot confirmed that selectivity was notably lower in the swap condition. Additionally, a detailed dynamics analysis revealed that the reduced selectivity in the swap condition stemmed directly from changes in the responses to P and N stimuli over time. This provided evidence that the network's altered behavior under the swap condition was due to specific



(a) t-SNE visualization (b) PCA visualization (c) Loss and accuracy while training the DNN

Figure 1: Visualization for the representations with LPL rule

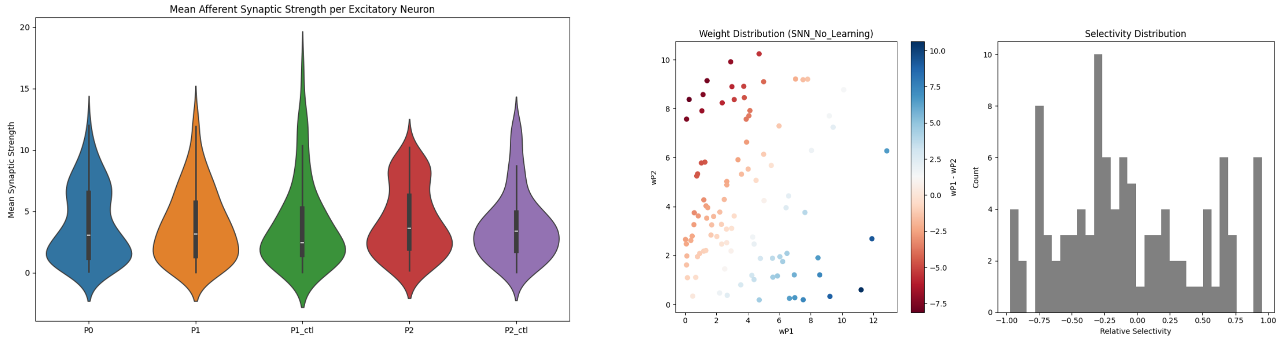
disruptions in the temporal association between input stimuli and their corresponding neuronal responses.
Experiment 3: LPL in SNNs



(a) Selectivity comparison (b) Selectivity comparison barplot (c) Selectivity analysis

Figure 2: Selectivity analysis for the IT cortex experiment in primates

To assess the effectiveness of the learning rule, we conducted simulations using Spiking Neural Networks (SNNs). The observed results, as depicted in Figure 3, did not meet the expected benchmarks outlined in the reference paper. Specifically, the weight selectivity demonstrated by the network was suboptimal.
Experiment 4: Evaluating on the audio data set We observed the following results, which were not satisfactory, as the accuracy was approximately 20% on the audio dataset. However, the model performed significantly better when evaluated on visual data. Below, we present the training loss and accuracy metrics. Additionally, we include t-SNE and PCA plots to evaluate the disentanglement of features in a low-dimensional space.



(a) weight distribution (b) weight selectivity SNN

Figure 3: Results when LPL was used with spiking neural networks

5 Discussion and Conclusions

The experiments conducted using the Latent Predictive Learning (LPL) framework demonstrate its ability to model neuronal selectivity changes and produce disentangled representations in neural networks. In the first experiment, LPL replicated changes in neuronal selectivity observed in primate IT cortex using swap and nonswap conditions in a synthetic Moving MNIST dataset. The swap condition led to reduced selectivity due to disrupted temporal associations, while the nonswap condition maintained stable selectivity, highlighting LPL's ability to preserve correct associations.

In the second experiment, LPL exhibited clear, well-separated clusters in t-SNE and PCA visualizations, validating its disentanglement capabilities. Comparisons with baseline models (backpropagation, Hebbian, or predictive learning alone) revealed that combining Hebbian and predictive components in LPL produces superior disentanglement and clustering, avoiding representational collapse.

However, applying LPL to spiking neural networks (SNNs) yielded less effective results, with weak relative selectivity and neuronal activity compared to the original study. Despite this, LPL successfully resulted in the disentangled representations and modeled selectivity changes observed in the experiment.Along with that it yielded bad results when applied to the audio data set.These findings highlight LPL's promise in modeling adaptive neural processes and unsupervised learning mechanisms. Future work should explore its application to more complex datasets with large data, real-world scenarios. Code and other artifacts are available on : **GitHub Repository Link**

References

[1] Li, N. & DiCarlo, J. J. Unsupervised natural experience rapidly alters invariant object representation in visual cortex. *Science* **321**, 1502–1507 (2008).

[2] Halvagal, M. S. & Zenke, F. The combination of Hebbian and predictive plasticity learns invariant object representations in deep sensory networks. *Nature Neuroscience* **26**, https://doi.org/10.1038/s41593-023-01460-y (2023).

[3] Illing, B., Bellec, G., Ventura, J. & Gerstner, W. Local plasticity rules can learn deep representations using self-supervised contrastive predictions, Conference on Neural Information Processing Systems .(NeurIPS 2021).