University of Tehran

College of Engineering

School of Electrical & Computer Eng

Principles of Cognitive Science

Assignment 2 Phase Two

Name

Reza Chehreghani

Student ID

810101401

May 28, 2025

Contents

List of Figures

Replicating the Closed-Loop Design

To replicate the closed-loop design described in Ponce et al. (2019), I loaded the provided dataset and identified each main and auxiliary section using their respective start and end tags (226–227 for main sections; 228–229 for auxiliary sections). Within each section, I extracted the timestamps of all successful trials (marked by the event sequence 100 → 200 → 223) and computed the firing rate for each trial based on the spike count divided by its duration.

For each block, I then calculated the average firing rate for the main (synthetic) trials and for the auxiliary (natural) trials. The resulting firing rates across all sections are visualized in the following diagram:

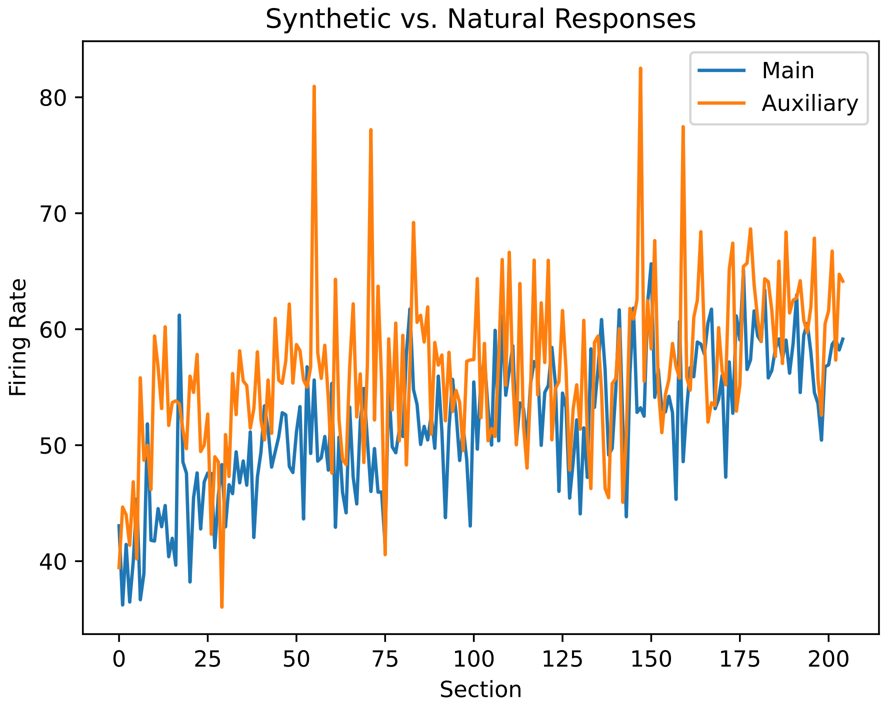

Figure 1: Average firing rates for main (synthetic) and auxiliary (natural) sections across all recording blocks. Each point reflects the mean response within a block.

The plot above shows that over successive blocks, the firing rate in the main sections increases, reflecting the genetic algorithm's ability to evolve progressively better image codes that more effectively drive neuronal responses. Notably, the auxiliary responses (to a fixed set of natural images) also show some increase, which was not initially expected, as one might anticipate adaptation effects reducing the response over time. However, this pattern is consistent with observations from the closed-loop paper, where optimizing synthetic stimuli can reveal the neuron's maximal activation potential.

These results demonstrate that the closed-loop approach successfully navigates the high-dimensional stimulus space, leveraging only the recorded neural responses as feedback, and converges on synthetic stimuli that elicit stronger neural responses. This highlights the power of evolutionary search combined with generative models to uncover the true tuning properties of individual neurons, as shown in the original experiment.

Alternative Fitness Criteria

One limitation of the XDream approach is that, although the evolved images elicit very high firing rates, they often lack semantic or perceptual coherence, making it difficult to infer the neuron's genuine tuning. In other words, driving a neuron maximally does not guarantee that the resulting pattern corresponds to any recognizable feature or object.

To address this, one could augment the standard activation-based fitness $F_{\text{spike}}(z)$ with a *realism* term drawn from the GAN's discriminator output. Recall that a GAN pairs a generator G with a discriminator D, where $D(x)$ estimates the likelihood that x belongs to the natural image manifold. We define

$$F(z) \;=\; F_{\text{spike}}(z) \;\times\; \big[D\big(G(z)\big)\big]^{\alpha},$$

where $\alpha > 0$ controlling the weight on realism. This joint criterion still rewards strong neural drive but biases evolution toward images that reside on the manifold of naturalistic stimuli, making the resulting patterns easier to interpret in terms of familiar shapes, textures, and object categories.

Using such a realism-augmented fitness implies that we're probing not only the neuron's maximal activation axes but also its sensitivity to natural image statistics — essentially targeting the property of "tuned selectivity" rather than mere excitability. Under this scheme, the final generation would consist of the most realistic stimuli that also elicit high firing rates, revealing the neuron's preferred natural exemplars. These images can then be inspected for consistent features to infer the neuron's functional role in object recognition and categorical coding.