Human Visual-Inspired Noise Further Improves Denoising Autoencoder

Author: Gengsheng Li
Email: ligengsheng2024@ia.ac.cn
Department: Laboratory of Brain Atlas and Brain-inspired Intelligence, Institute of Automation, Chinese Academy of Sciences, Beijing, China

***Abstract*—This report provides an in-depth analysis of Mahsa Mayeli's "Neurophysiology of Visual Perception" [1] and explores its implications for artificial intelligence visual system design. The article comprehensively introduces the neurophysiological foundations of the human visual system, including the anatomical structure of the visual system, visual pathways, object recognition mechanisms, the relationship between eye movements and perception, and the constructive nature of visual perception. Inspired by this, I designed a novel noise application method that simulates key features of the human visual system for training the denoising autoencoder. This human visual-inspired noise application method incorporates non-uniform noise, color attenuation, and eye movement blur, aiming to more accurately mimic the natural visual characteristics of the human eye. Preliminary results indicate that denoising autoencoders pre-trained using this noise application method demonstrate significant advantages in downstream tasks.**

I. Overview of the Literature

THE literature, "Neurophysiology of Visual Perception", provides a comprehensive introduction of the neurophysiological foundations of the human visual system. The article covers the following main aspects:

Neuroanatomy of the Visual System: The article details the entire visual pathway from the eyeball to the visual cortex in the brain. It particularly emphasizes the structure of the retina, noting that the foveal region has the highest density of photoreceptors and neurons, which determines the high-resolution center and low-resolution periphery of human vision.

Visual Pathways and Brain Regions: The author introduces two main visual processing pathways: the ventral pathway ("what" pathway) responsible for object recognition, and the dorsal pathway ("where" pathway) responsible for spatial and motion information processing. The article also discusses in detail the functions of the primary visual cortex (V1) and higher visual areas (such as V2, V4, V5).

Object Recognition and Scene Segmentation: The article explores the complex process of object recognition, including separating objects from the background (scene segmentation) and category-specific recognition. The author points out that this process involves bottom-up Gestalt rules and top-down attention mechanisms, and provides corresponding introductions.

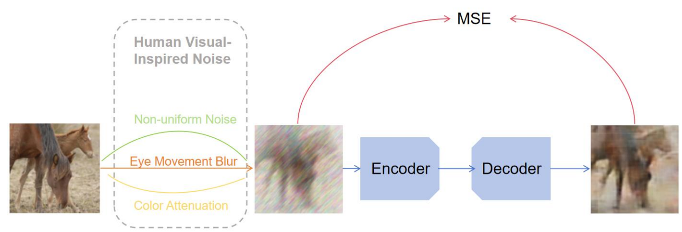

Fig. 1. The pipeline of training denoising autoencoder with human visual-inspired noise.

Eye Movements and Perception: The article emphasizes the crucial role of eye movements (such as microsaccades, drifts, and tremors) in maintaining visual perception. The author notes that these tiny eye movements are essential for preventing visual adaptation and maintaining object contrast.

Constructive Nature of Visual Perception: The article posits that visual perception is an active construction process, with the brain reconstructing visual scenes based on input, expectations, and prior experiences, emphasizing the predictive and adaptive nature of the visual system.

II. Implications for AI

The "Neurophysiology of Visual Perception" article provides rich inspiration for designing visual AI systems:

Non-uniform Processing: The center-periphery difference in the human retina suggests that visual AI systems can adopt similar non-uniform processing strategies. This can be achieved through designing attention mechanisms or using non-uniform noise.

Adaptive Mechanisms: The importance of eye movements in maintaining visual perception suggests that visual AI systems should include dynamic adaptive mechanisms to prevent adaptation to static stimuli.

Parallel Processing Pathways: The existence of ventral and dorsal pathways suggests that visual AI systems can adopt multi-task learning frameworks to simultaneously process object recognition and spatial-motion information. In this case, the system could learn something more meaningful with the help of various supervision tasks.

Predictive Coding: The constructive nature of visual perception inspires us to incorporate predictive coding mechanisms in visual AI systems, enabling the system to make predictions and decisions based on prior knowledge and contextual information.

These insights provide a theoretical foundation for designing AI models that more closely resemble the human visual system. By integrating non-uniform processing and adaptive mechanisms, we can develop denoising autoencoders that more closely approximate the human visual system. These denoising autoencoders have been tested and demonstrate performance far superior to traditional denoising autoencoders in the downstream task.

III. Denoising Autoencoder with Human Visual-Inspired Noise

Based on the aforementioned insights, I designed a denoising autoencoder that mimics features of the human visual system. While maintaining consistency with traditional denoising autoencoders in its main encoding-decoding structure, it incorporates characteristics of the human visual system in its noise application method through the following steps:

1) **Non-uniform Noise**
   Simulating the distribution differences of retinal cone and rod cells, less noise is applied to the central area of the image and more noise to the peripheral areas.
2) **Color Attenuation**
   Simulating the distribution differences of retinal cone and rod cells, the color sensitivity is gradually faded from the center to the edges of the image.
3) **Eye Movement Blur**
   Simulating the blur effect during rapid eye movements (saccades), a specific directional blur is added to the image.

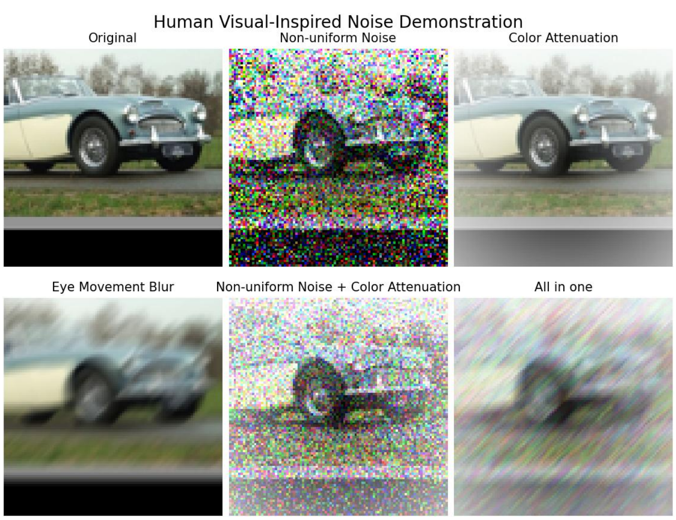

Fig. 2. The effects of different human visual-inspired noise.

Figure 1 illustrates the self-supervised training process of the proposed denoising autoencoder, while Figure 2 demonstrates the effects of different human visual-inspired noise.

To quickly validate the feasibility of the method, I used a relatively simple convolutional network as the backbone of the denoising autoencoder (the encoder consists of three convolutional layers plus pooling layers, and the decoder consists of three deconvolutional layers). The widely used STL-10 dataset was chosen as the experimental data for self-supervised training tasks.

A. Pre-training Stage Results

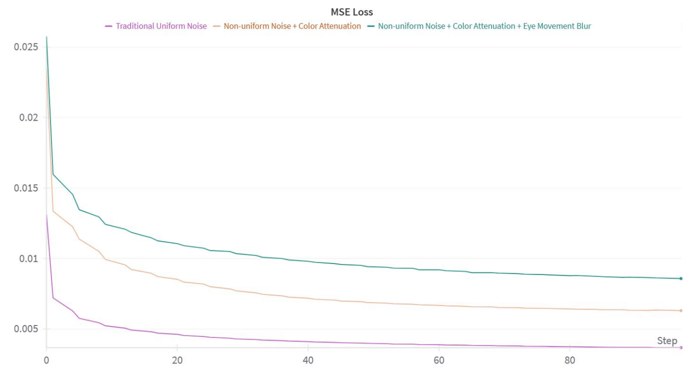

Fig. 3. The loss curves during the pre-training stage of different denoising autoencoder.

We used 100,000 unlabeled images from the STL-10 dataset to pre-train denoising autoencoders with three noise application methods, including a denoising autoencoder with traditional uniform noise, one with non-uniform noise and color attenuation, and one with non-uniform noise, color attenuation, and eye movement blur. Figure 3 shows the training process of denoising autoencoders with different noise application methods.

B. Results in Downstream Tasks

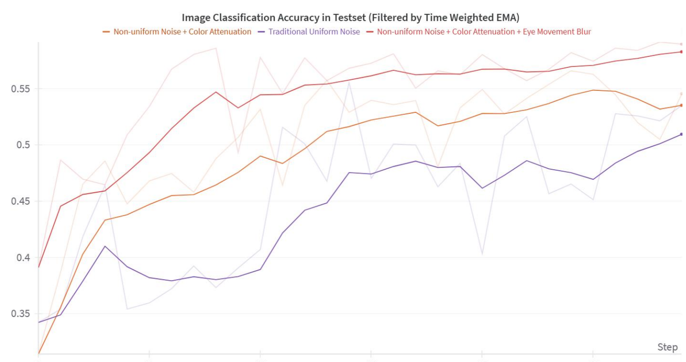

Fig. 4. The testset performance of denoising autoencoders with different noise application methods in the downstream image classification task.

To evaluate the pre-training effects of denoising autoencoders with different application methods, I adopted a linear-probe approach to use the pre-trained denoising autoencoders for downstream tasks. Due to time and computational resource constraints, I currently only selected the most representative task, i.e. image classification, as the downstream task for evaluation. Specifically, the encoder part of the pre-trained denoising autoencoder is frozen (its parameters no longer update), and a trainable linear layer is added to the output end of the encoder to map the image features extracted by the encoder to the target classification categories. Here, the labeled training set of STL-10 (5,000 images) was used for linear-probe training, and the labeled test set (8,000 images) was used to test the performance of the resulting classifier.

Figure 4 shows the testset performance of denoising autoencoders with different noise application methods in downstream tasks. It is evident that the denoising autoencoder using non-uniform noise, color attenuation, and eye movement blur demonstrates a clear advantage in the downstream image classification task, followed closely by the one using non-uniform noise and color attenuation, while the one using traditional uniform noise performs the worst. This result indicates that the human visual-inspired noise application method I proposed enables the denoising autoencoder to extract more useful and generalized features, thus exhibiting superior performance in downstream tasks.

IV. Conclusion

By applying neurophysiological knowledge to AI system design, I developed a denoising autoencoder that more closely resembles the human visual system. Preliminary results indicate that this biologically inspired method significantly outperforms traditional methods. Future research directions include experimenting with more complex network structures and testing on a wider range of downstream tasks.

References

[1] Mayeli, Mahsa. "Neurophysiology of visual perception." *Biophysics and neurophysiology of the sixth sense* (2019): 13-26.