Visual Image Reconstruction from Human Brain Activity using Linear Image Decoders plus Nonlinear Noise Suppression

Qiang Li
Image Processing Laboratory, University of Valencia, Valencia, Spain
* *(Current Address) Tri-Institutional Center for Translational Research in Neuroimaging and Data Science (TReNDS)*
Georgia State University, Georgia Institute of Technology, Emory University, Atlanta, GA, USA

***Abstract*—In recent years, substantial strides have been made in the field of visual image reconstruction, particularly in its capacity to generate high-quality visual representations from human brain activity. This advancement not only enables the recreation of visual content but also provides valuable insights into the intricate processes occurring within functional brain regions, contributing to a deeper understanding of brain function. However, linear decoding methods, which were once prominent in this field, have faced challenges due to their susceptibility to noise. As a result, they have gradually fallen out of favor. In response to this limitation, our study introduces a novel approach that combines linear mapping with nonlinear reconstruction techniques to recreate visual images perceived by research subjects based on their brain activity patterns. The primary challenge associated with linear mapping lies in its susceptibility to noise interference. To address this issue, we leverage a flexible denoised deep convolutional neural network, which surpasses the performance of traditional linear mapping. Our investigation encompasses three distinct linear mapping approaches, as well as the training of shallow and deep autoencoder denoised neural networks, including a pre-trained state-of-the-art denoised neural network. The outcome of our study reveals that the amalgamation of linear image decoding with nonlinear noise reduction significantly enhances the quality of reconstructed images from human brain activity. This suggests that our methodology holds promise for decoding intricate perceptual experiences directly from brain activity patterns. Moreover, our analysis highlights an interesting observation: during the encoder phases of the denoised autoencoder, color information emerges as the most crucial factor, while both color and texture information assume equal importance during the decoder phases. This observation underscores the distinct manner in which Alexnet neural networks represent feature information, setting the stage for further investigation and innovation in this domain.**

***Index Terms*—visual image reconstruction, human brain activity, linear decoding, deep autoencoder denoised neural networks**

I. Introduction

The human brain, a marvelously intricate organ, serves as the foundation for cognitive information processing, orchestrating the multifaceted tapestry of our daily experiences. As elucidated by Mark and colleagues [1], visual perception emerges as a pivotal component of cognitive function, intricately intertwined with both Short-Term Memory[1], characterized by the transient storage of a limited volume of information, and Long-Term Memory[2], denoting the enduring retention of knowledge over extended periods [1]. To unravel the intricate relationship between visual experiences and our cerebral states, researchers have embarked on a comprehensive exploration, employing both linear and nonlinear encoding and decoding models, with a particular focus on the latter approach.

In the nascent stages of this exploration, linear mapping decoders held sway as a prominent choice [2]. However, a formidable limitation loomed large, marked by the pronounced presence of noise in the reconstructed visual imagery. With the advent of deep learning methodologies, a transformative shift occurred, giving rise to a diverse array of nonlinear decoding techniques. This paradigm shift yielded remarkable outcomes, culminating in superior performance in the realm of reconstructing visual content from neural activity. Noteworthy contributions from researchers such as Yusuke et al.[3], Yoichi et al.[4], Guohua et al.[5, 6], Rufin[7], Ziqi [8], Shinji [9], Yukiyasu [10], Bertrand [11], Thomas [12], Kendrick [13], and Garrett [14], among others, have collectively propelled the understanding and application of nonlinear decoding models in the reconstruction of visual information from brain activity.

The primary challenge associated with linear decoders lies in their vulnerability to noise, which can emanate from various sources. Firstly, noise can originate from intrinsic brain activity itself. BOLD signals, commonly employed in neuroimaging studies, are susceptible to both physiological noise stemming from bodily functions like heart rate variations [15], and physical noise introduced by the MRI machinery [16]. Secondly, linear decoders often falter when tasked with processing high-dimensional and nonlinear functional magnetic resonance imaging (fMRI) data [17]. The intricate nature of neural activity patterns necessitates more sophisticated decoding approaches. Furthermore, the issue of noise is exacerbated by the intrinsic redundancy prevalent in natural images. This redundancy, while advantageous for human recognition, poses a substantial challenge for linear decoders, engendering pronounced noise in the reconstructed images.

However, recent advancements in image denoising techniques have illuminated a path forward. These denoising models

[1]Short term memory, often known as primary or active memory, is the ability to store and store a small quantity of information within the mind.

[2]Long-term storage refers to information storage over a long period of time. This kind of memory is stable and can last for many years.

can be broadly categorized into three distinct groups. First, we have "Fully-Supervised Denoising Models," exemplified by deep convolutional neural networks like DnCNN [18], FFDNet [19], CBDNet [20], and SGNet [21]. These models exhibit exceptional denoising capabilities and have been trained using noisy-to-clean image pairs. While they excel in denoising tasks, they are constrained by the requirement for meticulously curated paired images during training, which can be challenging to collect in real-world scenarios. Second, there is the "Noise2Noise" category, where models are trained with noisy-noisy image pairs. However, this approach, although conceptually intriguing, possesses limited practical applications compared to the fully supervised paradigm. Finally, we encounter the "Self-Supervised Models," a burgeoning approach gaining popularity. These models are designed to work with a single noisy image, exemplified by DIP [22] and Self2Self [23]. While promising, self-supervised models also grapple with the challenge of restricted applicability. In navigating the landscape of denoising techniques, it becomes evident that each category presents its unique strengths and limitations. The choice of the most suitable denoising model hinges on the specific application context and the availability of paired or noisy-noisy data, ultimately shaping the trajectory of advancements in this domain.

In this research endeavor, we introduce a novel hybrid approach, combining both linear and nonlinear image decoding methodologies, with the overarching aim of reconstructing visual experiences from recorded brain activity. The pivotal metric underpinning our investigation is the Blood-Oxygen-Level-Dependent (BOLD) signal, a hallmark of functional magnetic resonance imaging (fMRI) that reflects variations in deoxyhemoglobin concentrations attributable to localized changes in blood flow and oxygenation levels.

Our initial focus gravitates toward linear decoders, primarily tasked with establishing the mapping between patterns of brain activity and corresponding image pixels. This preliminary mapping serves as the foundation upon which subsequent nonlinear decoding techniques are employed. These nonlinear decoders assume the vital roles of denoising and enhancing the reconstructed images, engendering heightened clarity and fidelity.

The principal objective that steers this research voyage is the reconstruction of visual images using the invaluable information gleaned from fMRI data. This aspiration propels us to unravel the intricate relationship between neural activity and visual perception. Moreover, we embark on a secondary mission, wherein we systematically investigate the influence of noise on the quality and precision of our reconstructed results. This dual-pronged pursuit promises to advance our understanding of the complexities inherent in the reconstruction process while shedding light on the impact of noise in this context.

II. Related Work

The realm of visual image reconstruction has recently emerged as a prominent and vibrant area of research, further catalyzed by the resounding success of deep learning techniques in the domain of computer vision [3–14, 24]. The endeavor to faithfully and precisely reconstruct visual images from neural activity represents a formidable confluence of neuroscience and machine learning, offering a challenging frontier.

This multidisciplinary undertaking encounters multifaceted challenges from the neuroscientific perspective. Notable among these challenges are the issues surrounding the sheer magnitude of voxels and regions of interest, the intricacies inherent in the encoding and decoding mechanisms of visual information within the brain, and the abundant facets of brain function that remain cloaked in the shroud of ambiguity, awaiting deeper understanding.

The coalescence of these challenges engenders a rich tapestry of complexities, beckoning both neuroscientists and machine learning practitioners to explore uncharted territories. As we navigate this intricate landscape, the pursuit of accurate visual image reconstruction stands as a testament to the symbiotic relationship between neuroscience and machine learning, with each field poised to enrich the other through collaborative exploration and discovery.

Previous research has shed light on the existence of a meaningful correspondence between visual cortical activities and the perception of visual stimuli. It has been demonstrated that the experienced visual content can be decoded from functional magnetic resonance imaging (fMRI) records. In the early stages of mapping estimation, linear models, such as linear regression, took precedence. These algorithms typically operate by extracting specific characteristics of the visual content, such as multi-scale local image features, and subsequently learning a linear mapping between these image characteristics and fMRI signals [4]. However, it became evident that linear techniques faced substantial challenges when tasked with reconstructing complex images, especially natural images, as they tended to yield outcomes laden with significant noise.

In contrast, the paradigm shift inspired by the discovery of homogeneity between the hierarchical structures of the brain and artificial deep neural networks (DNN) ushered in a new era of effective reconstruction techniques. Deep neural networks, hinging on this fundamental similarity, emerged as a potent tool for decoding visual information from brain activity. This transition marks a significant advancement in visual image reconstruction methodologies.

For example, Shen et al. leveraged convolutional neural network (CNN) models to extract characteristic features from images and successfully reconstructed natural visual content by mapping these CNN-based features to fMRI signals. The development of DNN architectures has facilitated the learning of nonlinear mappings from brain signals to visual content in an end-to-end fashion. Diverse DNN techniques, including Auto-encoders, Generative Adversarial Networks (GANs)[25], Recurrent Neural Networks (RNNs)[26], and Spiking Neural Networks (SNNs) [27], have made substantial strides in enhancing the performance of visual image reconstruction from brain activity.

Despite these significant advancements, challenges persist. Most notably, the reconstructed images are often derived

from brain activity patterns that do not perfectly align with the visual stimuli images. Furthermore, the presence of pervasive noise, emanating from various sources, poses a formidable obstacle. These noise sources encompass artifacts from scanning machines [16], intrinsic neural noise [28, 29], and even physiological factors like heart rate fluctuations [15]. Overcoming these challenges remains a central focus, as researchers endeavor to refine the precision and fidelity of visual image reconstruction from brain activity.

In the framework proposed by [6], a deep neural network (DNN)-based decoder is presented, demonstrating its prowess in effectively learning nonlinear mappings from fMRI signals to visual images. Meanwhile, [30] has ventured into the territory of bidirectional mapping, utilizing an unsupervised Generative Adversarial Network (GAN) to learn the intricate interplay between fMRI signals and stimulus images. These recent advancements in decoding methodologies have ushered in a new era characterized by improved image quality and the ability to reconstruct images in a more natural and faithful manner compared to their linear counterparts.

However, it is essential to acknowledge that the task of learning to map the corresponding stimuli from fMRI recordings presents a formidable challenge. The crux of this challenge predominantly lies in the complexity inherent to brain activity within the visual cortex, rather than the mere volume of information at hand. Extensive research has underscored the hierarchical nature of representations within the visual cortex. Furthermore, studies have posited that harnessing information from multiple visual fields holds the key to enhancing the performance of image reconstruction.

In light of these insights, it becomes evident that simplistic decoding models may fall short of achieving accurate reconstructions. It is imperative to consider and incorporate hierarchical information, a crucial dimension of data encoding, when striving to faithfully reconstruct visual images from fMRI data. This perspective underscores the rich tapestry of investigations into the hierarchy of data encoding within visual regions, a terrain ripe with potential for advancing the field of visual image reconstruction.

III. Methods and Results

A. BOLD5000

Cheng et al. have made the BOLD5000 dataset openly accessible to the public [31]. This dataset serves as a valuable resource for the task of reconstructing visual images from brain activity patterns. It comprises data collected from four subjects, each aged between 24 to 27 years, with either normal vision or vision corrected to normal. These individuals participated in 15 scanning sessions, during which they were exposed to a total of 5254 images.

To curate the stimulus images used in these sessions, the researchers drew from three well-established computer vision datasets. These datasets include the SUN dataset, consisting of 1000 scene images[3], the COCO dataset[4], and the ImageNet dataset, encompassing 1916 images[5]. The amalgamation of these diverse datasets served to enrich the variety of visual stimuli presented to the subjects during the scanning sessions.

This dataset's availability and the careful selection of stimuli from multiple sources contribute to its significance as a valuable resource for investigations in the field of visual image reconstruction, as illustrated in Fig. 1.

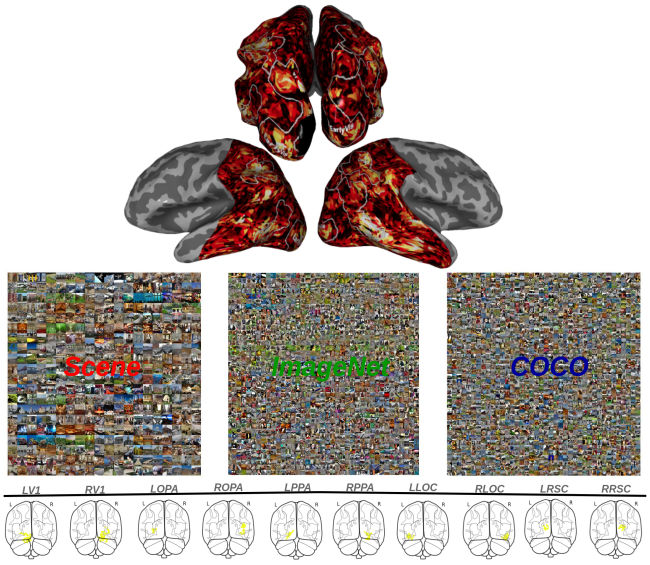

Fig. 1: **Nature scenes selective ROIs.** Regions of interest (ROIs) exhibiting the highest color intensity signify specialized regions within the human visual cortex that are selective for scenes. The brain's activity is depicted on a flattened surface, and corresponding brain images are presented, aligning with stimuli featuring natural scenes in the middle of the visual field. It's worth noting that "LH" designates the left hemisphere, whereas "RH" signifies the right hemisphere. Within these visual regions, "V1" denotes the primary visual cortex, responsible for early-stage visual processing. "OPA" stands for the occipital place area, a region specialized in processing information related to places and locations. "PPA" represents the parahippocampal place area, another region dedicated to scene and place recognition. "LOC" corresponds to the lateral occipital complex, known for its involvement in object recognition and scene perception. Finally, "RSC" signifies the retrosplenial cortex, which plays a role in spatial memory and navigation within the human brain.

B. Linear decoders training plus nonlinear decoders testing

1) Linear encoding plus decoding: Here, we considered linear regression [32] as a linear image decoder. In mathematical terms, we assumed that a linear mapping exists between the latent vectors of visual image stimuli S and the corresponding brain response vectors R, such that:

$$R_i = S_i \cdot W + \epsilon_i \tag{1}$$

[3]http://sun.cs.princeton.edu/

[4]https://cocodataset.org/

[5]http://www.image-net.org/

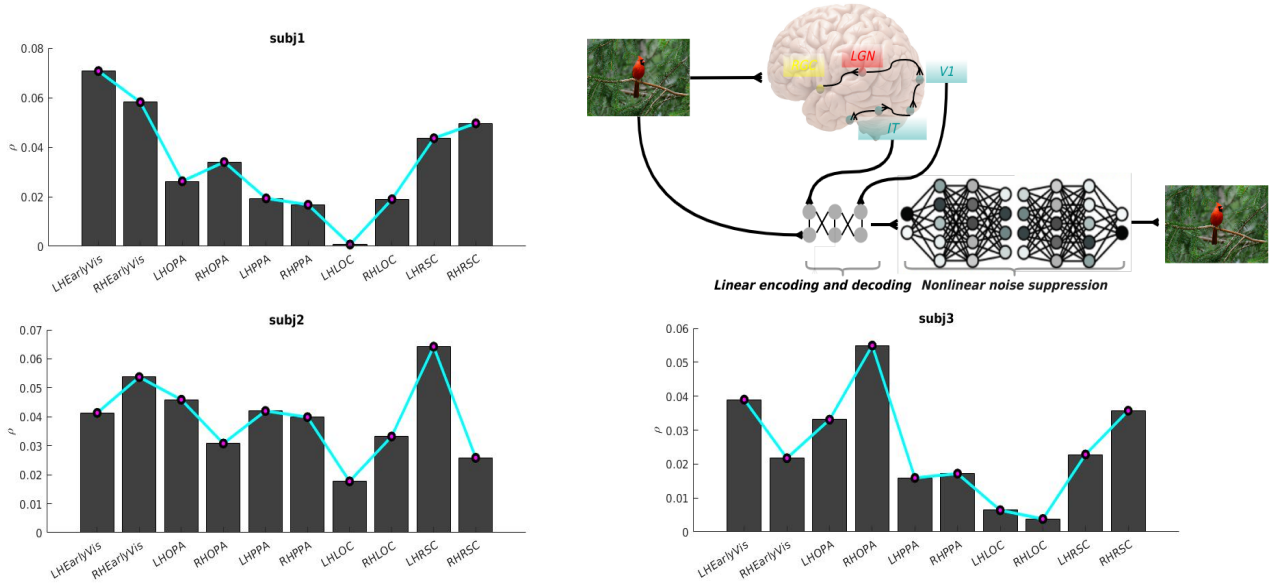

Fig. 2: **The architecture of linear image decoders plus nonlinear noise suppression.** In the second figure, the top row provides an overview of the fundamental structure of linear image decoders combined with nonlinear noise suppression techniques. Moving to the first and second rows in the top row of figures, they illustrate the linear encoding accuracy observed in different subjects. Along the x-axis, we find the regions of interest (ROIs), while the y-axis represents the correlation (ρ) between the predicted and actual BOLD responses. These visualizations offer insights into the performance of linear encoding within various ROIs across different subjects. The correlation values (ρ) serve as a metric for assessing the accuracy of predictions in capturing the real BOLD responses.

Where $R_i \sim \mathcal{N}(\beta^T R, \sigma^2)$, $\epsilon_i \sim \mathcal{N}(\beta^T R, \sigma^2)$, and $i \in \mathbb{R}^{+\infty}$ indicate number of stimuli. Training the brain decoder consists of finding the optimal mapping W by solving for W:

$$\begin{aligned} S^{\mathrm{T}} R &= S^{\mathrm{T}} S \cdot W \\ W &= \left(S^{\mathrm{T}} S\right)^{-1} \cdot S^{\mathrm{T}} R \end{aligned} \tag{2}$$

where $S^T S$ is the covariance matrix of the latent vectors used for training. Mathematically, testing the brain decoder means using the learned weights W to get the latent vector S for each new brain activation pattern R. Here the weight estimation is performed using $L2$-regularized linear least-squares regression. The optimal regularization parameter for each model is determined by 10-fold cross validation of training data and shared across all voxels. Starting again from Eq.1, we now solve for S:

$$\begin{aligned} RW^{\mathrm{T}} &= S.WW^{\mathrm{T}} \\ S &= RW^{\mathrm{T}}.\left(WW^{\mathrm{T}}\right)^{-1} \end{aligned} \tag{3}$$

In Fig. 2, each linear model predicted voxel responses to new image inputs. This is followed by the integration of the predicted voxel responses from individual models using a weighted linear average. Based on Pearson's correlation coefficient (Pearson's correlation between measured and predicted voxel responses) in model training, the weight is calculated for each voxel in the dataset. Overall, we found that linear decoding accuracy varies between individuals, with RSC and OPA having better decoding accuracy than other regions of the high-level visual cortex.

2) Nonlinear noise suppression: In this context, we employed two types of denoised neural networks for our analysis: one set of autoencoder-denoised neural networks that we trained in-house, and another set consisting of a state-of-the-art deep denoised neural network that had already undergone training.

Autoencoder-denoised neural networks: In this study, we harnessed the power of denoised Convolutional Neural Networks (CNNs), a cutting-edge advancement in artificial intelligence known as Autoencoder. These networks were utilized to address denoising challenges, particularly for images degraded by Gaussian white noise. The core concept behind Autoencoders lies in having the generator's target coincide with its input. Essentially, the generator's mission is to encode the input data into a compressed latent representation and then proficiently decode it back into the original images.

To facilitate the training process, we leveraged the Caltech-101 dataset, a comprehensive collection comprising 101 distinct classes, including a wide range of objects such as animals, airplanes, and flowers, among others, each exhibiting high shape variability [33]. Notably, the dataset encompasses varying numbers of images per category, ranging from 31 to 800, with most images characterized by medium resolution, approximately 300x300 pixels. In preparation for network training, we resized the input images to dimensions of 224 by 224 pixels.

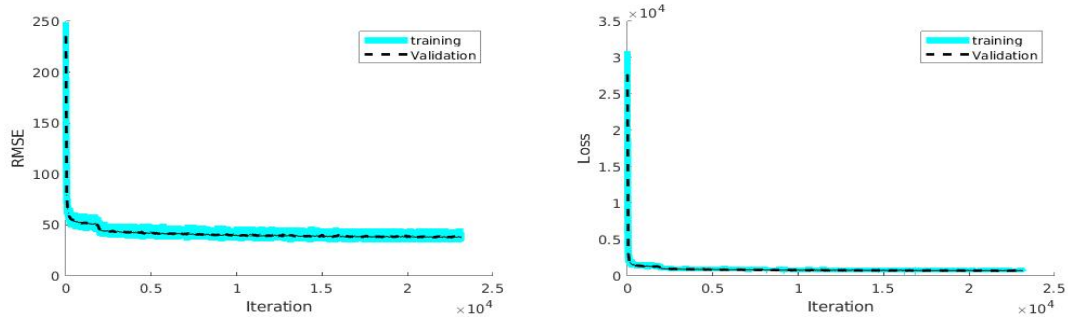

Fig. 3: **Training, validation, and loss curves of the autoencoder.** The left figure shows training and validation curves during training and validating the autoencoder. The y-axis refers to RMSE, and the x-axis refers to iteration. The right figure shows the loss function during training and validation. The y-axis refers to loss value, and the x-axis refers to iteration number.

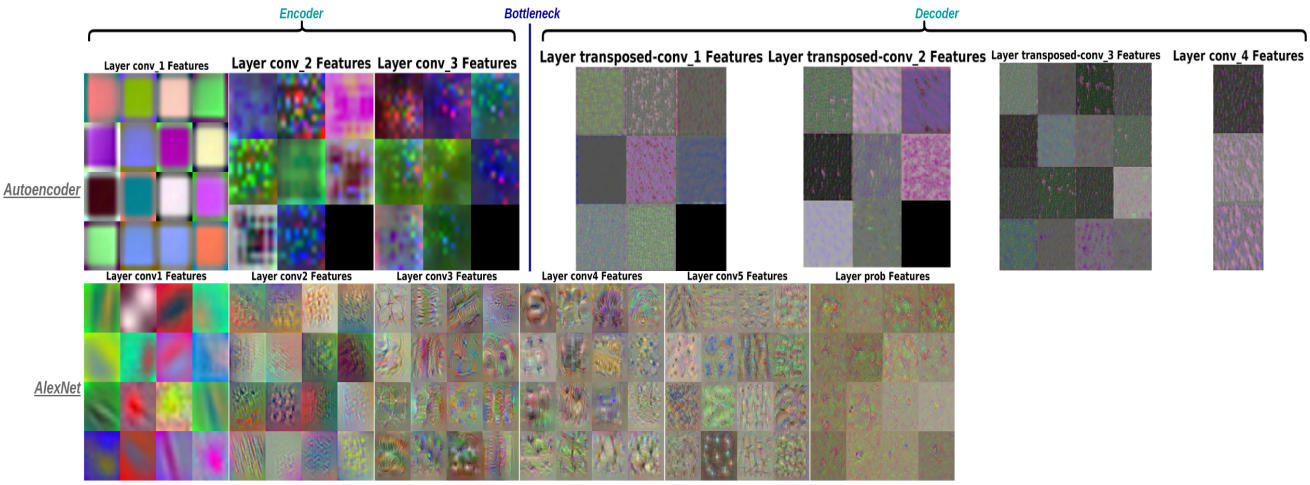

Fig. 4: **The features in the convolution layers are learned from images by Autoencoder.** In the top row, these characteristics offer valuable insights into the effectiveness of denoising when applied to paired clean-noised images. Notably, when the denoising process is initiated at the encoder phase and progresses to capture more abstract properties of the denoised image in higher layers, it leads to the emergence of more distinctive texture features in the reconstructed denoised images. Remarkably, when employing Alexnet, the resulting features exhibit striking disparities in both low-level and high-level neural networks.

To introduce controlled noise and further enhance the diversity of our training and validation samples, we intentionally added Gaussian white noise to create degraded images. Specifically, we applied noise with a mean value of 0.5 and a variance of 0.6. Additionally, we employed data augmentation techniques to augment our training and validation datasets, enriching the learning process and enhancing the network's ability to handle noise and variations effectively.

Thirdly, we trained networks on images using a regression output layer architecture with a six-layer encoder plus a six-layer decoder (convolution layer with 16 filters of size [3 3], max pooling layer with pool size [2 2] and stride [2 2], convolution layer with 8 filters of size [3 3], max pooling layer with pool size [2 2] and stride [2 2], convolution layer with 8 filters of size [3 3], max pooling layer with pool size [2 2] and stride [2 2]), and the decoder component of the architecture mirrors the encoder (upsample transposed convolution layer with 8 filters of size [2 2], upsample transposed convolution layer with 8 filters of size [2 2], upsample transposed convolution layer with 16 filters of size [2 2], convolution layer with 3 filters of size [3 3]). We trained neural networks with their loss function to optimize mean-squared-error (MSE). All layers are convolutional or up-convolutional. Because the autoencoder networks work with color images and accept input images of fixed sizes, if the input images were grayscale images, then we needed to convert them into RGB images by stacking three copies of the grayscale image G to create a corresponding 3-D array RGB = [G G G], and resizing them to fit into the input image size of the autoencoder network. The input image size for the autoencoder is 224-by-224. For the training options, the ADAM stochastic gradient descent was used. The momentum value was set to 0.9000, and the gradient threshold method was set to $L2$ norm. The minimum batch size was 800, and the maximum number of epochs was $1,000$. The initial learning rate was 0.001, and it stayed the same throughout training. The training data were shuffled before each training epoch, and the validation data were shuffled before each network validation.

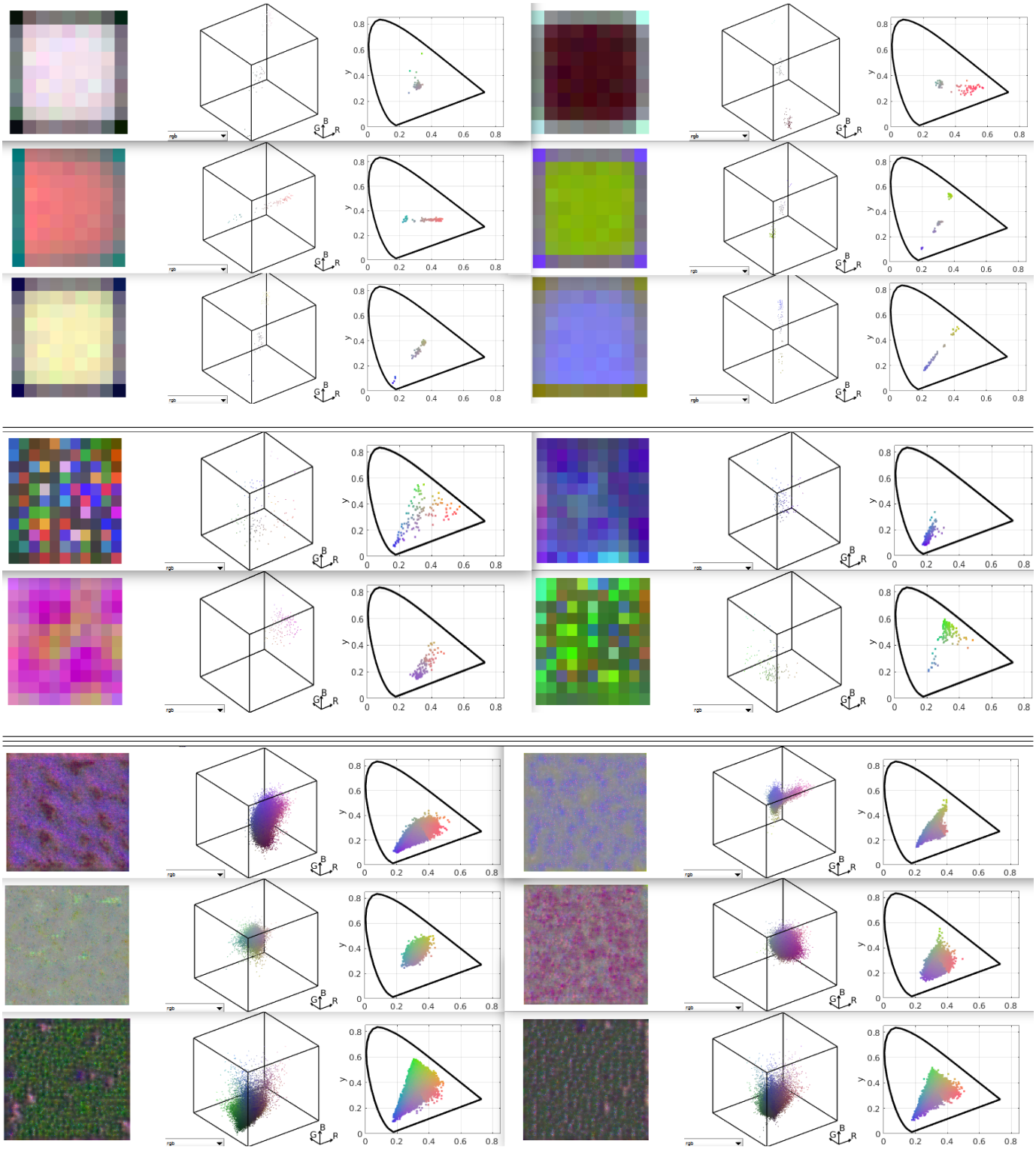

Fig. 5: **The feature color distribution of receptive fields at the different convolutional layers is learned from images by the Autoencoder.** The chosen color distributions have been thoughtfully represented in our plots. Specifically, we have presented distributions for white-black (WB) in the first row, red-green (RG) in the second row, and yellow-blue (YB) in the third row. These choices align with the chromatic-opponent mechanisms observed within the primate visual system, enhancing the relevance and significance of our visualizations. Below the first pair of horizontal lines, we provide insight into the color distribution of the receptive fields in the conv2 layer. Notably, in this representation, the chromatic-opponent characteristics appear to have dissipated, leading to a more uniform color distribution. Beneath the second set of three horizontal lines, we delve further into the receptive fields of the transposed conv2 layer. In this visualization, we offer a glimpse into both the color distribution and texture information. What's intriguing here is the gradual recovery of both color and texture information, suggesting a multi-faceted process of transformation and reconstruction. Each figure within these visual representations encompasses a receptive field, showcasing the intricate interplay of colors within a three-dimensional RGB color gamut. Additionally, we provide a chromaticity diagram to provide a more comprehensive understanding of color dynamics, arranged from left to right.

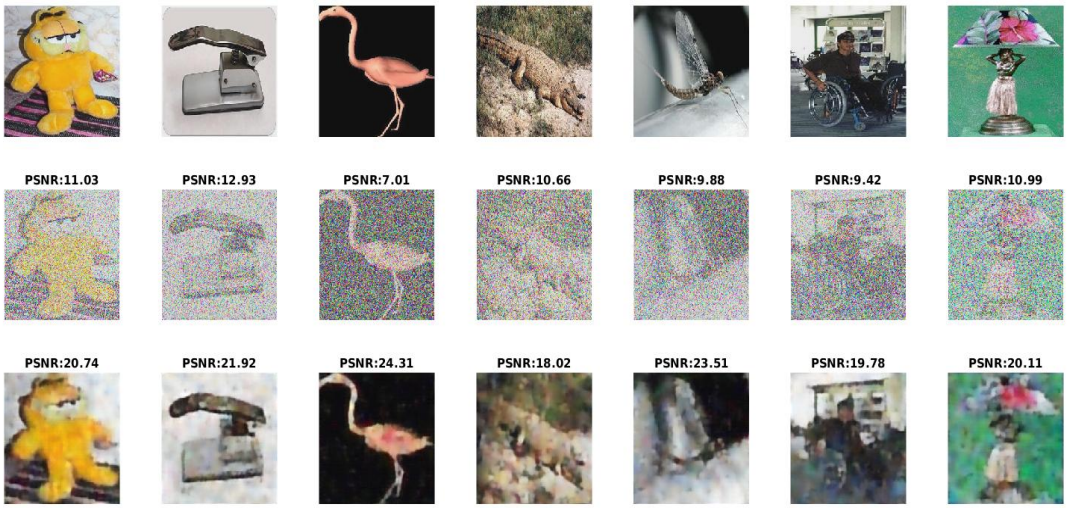

Fig. 6: **The denoised images are from a trained autoencoder neural network.** The clean image (top row) was degraded with Gaussian white noise models (middle row), then denoised with a well-trained Autoencoder neural network (bottom row). The evaluation metric, peak signal-to-noise ratio (PSNR), is reported in each image title.

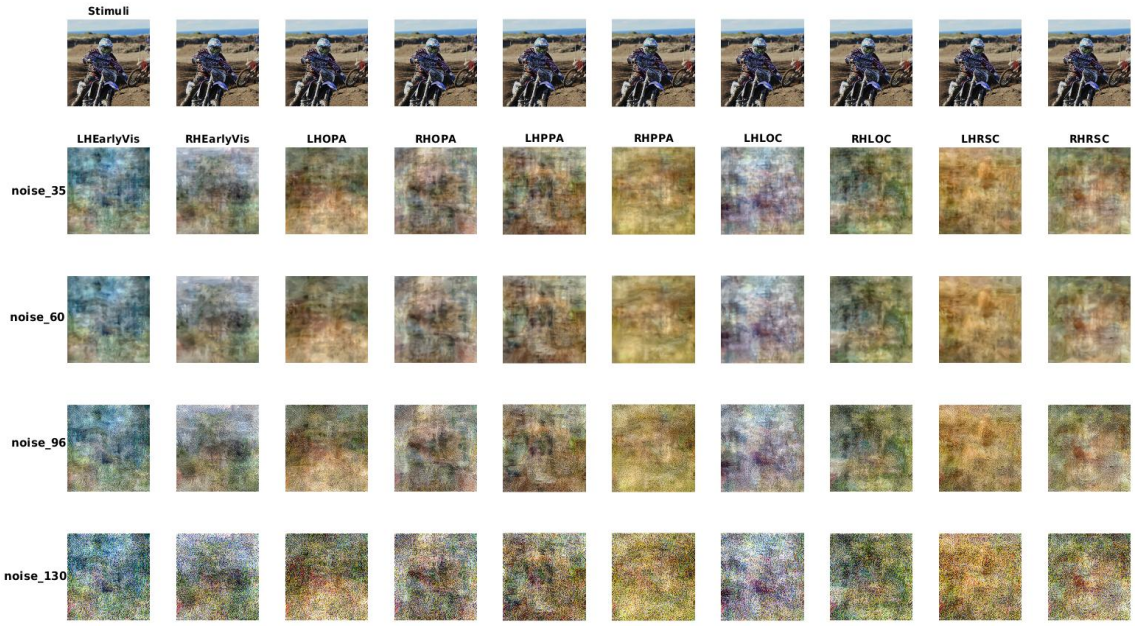

Fig. 7: **Visual image reconstruction from brain activity**. The top row shows stimuli for subjects. The remaining rows show reconstructed images from varying brain regions, such as, from left to right, they are LHV1, RHV1, LHOPA, RHOPA, LHPPA, RHPPA, LHLOC, RHLOC, LHRSC, and RHRSC, respectively. The left side of the first column labels showed varying suppression noise levels (e.g., 35, 60, 96, 130) with a pre-trained state-of-the-art deep convolution neural network, and the 60 noise level is the best choice for suppressing noise.

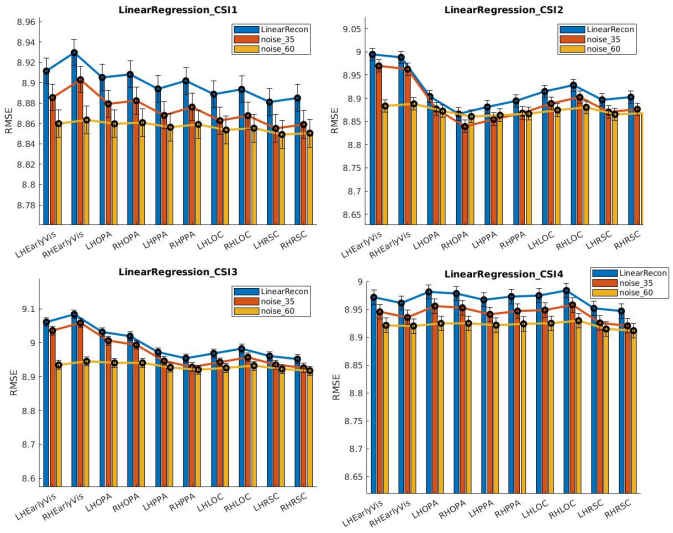

Fig. 8: **Visual image reconstruction quality was evaluated for four subjects via linear regression plus nonlinear denoised processing across different brain regions.** We reported the Root Mean Square Error (RMSE) for various brain regions that were reconstructed across four subjects. In each mini-group, the first bar represents linear reconstruction without any denoising processing. The second bar in each mini-group corresponds to denoising using dCNN with a noise level of 35 applied after linear reconstruction. Finally, the third bar in each mini-group reflects the same processing, but with a noise level of 60. This data provides valuable insights into the impact of denoising on image quality across different brain regions.

Pre-trained state-of-the-art deep convolutional denosied neural networks: In our study, we harnessed the power of a pre-trained deep convolutional neural network known as DUBD to effectively mitigate noise [34]. This choice was deliberate, as the DUBD model has consistently outperformed other state-of-the-art denoised neural networks [34]. For a comprehensive understanding of the network's architecture and loss function, we refer readers to the detailed description provided in [34].

Our experimentation involved subjecting images to varying noise levels, specifically 35, 60, 96, and 130, in order to assess the denoising performance of the DUBD model. We quantified the model's effectiveness in reducing noise and enhancing image quality using the Root Mean Square Error (RMSE) metric. Please refer to Fig.7 and Fig.8 for visual representations of these results. For additional reconstruction images, we direct readers to the Appendix section of this study.

IV. Conclusions and Discussion

After the training process, a straightforward feedforward pass involving an input image was used to extract features and perform image denoising using the CNN. Each unit within the CNN became active when a natural image was input into the system. The fluctuating representation of a specific feature within the image was captured by an activation weight generated as the image passed through each unit. These units, sharing the same kernel within a single layer, collectively produced an output feature map for each image. The output of each layer was referred to as the result of the rectified linear function before undergoing max-pooling.

Figure 4 reveals that task-demand convolutional neural networks exhibit distinctive features in both low-level and high-level phases of neural networks. For instance, autoencoders primarily manifest white-black (WB), red-green (RG), yellow-blue (YB), and purple colors within the convolutional layer 1. This pattern aligns with the chromatic-opponent mechanisms found in the primate visual system [35] (refer to Fig. 5). In layers conv2 and conv3, they feature blob-like texture information while retaining color information, although the chromatic-opponent information gradually diminishes, leading to a more uniform color distribution.

In contrast, within the decoder block, all layers predominantly exhibit both color and texture information to effectively reconstruct denoised images. This observation suggests that both color and texture features play pivotal roles in composing images. In essence, color introduces significant changes in the encoder layers, while both color and texture contribute significantly in the decoder blocks to restore clean images. Comparatively, Alexnet networks primarily share color information but diverge significantly in terms of texture information, reflecting their distinct objectives in neural network architecture.

In Figure 7, we presented the results of visual image reconstruction from various brain regions using linear regression in combination with different levels of denoising processes. Leveraging contextual information from the stimuli displayed in the top row, we were able to predict and reconstruct images from brain activity. To assess the quality of reconstruction, we computed the Root Mean Square Error (RMSE) between the stimuli and the reconstructed images, as illustrated in Figure 8.

Our findings revealed that the high-level visual cortex exhibited fewer errors in image reconstruction compared to the low-level visual cortex. From a biological perspective, this observation aligns with our understanding that natural images, such as objects, scenes, or more complex visual content, are typically represented in high-level brain regions. For example, the Occipital Place Area (OPA) is primarily activated by scene stimuli, while the Parahippocampal Place Area (PPA) is primarily activated by place-related stimuli. Furthermore, our results demonstrated that noise could be effectively reduced, leading to an enhancement in image quality, when employing linear regression in conjunction with a denoising process.

In our future research endeavors, we may explore the application of Principal Component Analysis (PCA) to reduce image redundancy, potentially leading to improved reconstruction performance. Additionally, it's worth noting that the suboptimal reconstruction results observed in this study could be attributed to the relatively low Signal-to-Noise Ratio (SNR) in the BOLD5000 dataset. Therefore, we should consider utilizing datasets with higher SNR levels, such as the Natural Scene Dataset [36], to potentially enhance our outcomes. Furthermore, it is imperative to investigate the interactions within high-order

visual networks during task states, as this exploration has the potential to enhance decoding accuracy. High-order brain networks are known to provide richer information compared to low-order brain networks [37–41]. Simultaneously, we must take into account bidirectional information interactions within the visual brain during visual decoding. This consideration may not only lead to improved decoding accuracy but also facilitate a more profound understanding of the information flow within the visual system during visual tasks [42, 43]. Lastly, we should also investigate the potential of deep approximation neural networks, as they have demonstrated remarkable achievements in various domains, and could be a promising avenue for future research in our field.

V. Acknowledgement

The author hereby declares that they have no conflicts of interest to disclose.

References

[1] M. Schurgin, "Visual memory, the long and the short of it: A review of visual working memory and long-term memory," *Attention, Perception, Psychophysics*, vol. 80, Apr. 2018.

[2] Q. Li, "Saliency prediction based on multi-channel models of visual processing," *Machine Vision and Applications*, vol. 34, May 2023.

[3] Y. Fujiwara, Y. Miyawaki, and Y. Kamitani, "Estimating image bases for visual image reconstruction from human brain activity," in *Advances in Neural Information Processing Systems*, Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams, and A. Culotta, Eds., vol. 22, Curran Associates, Inc., 2009.

[4] Y. Miyawaki *et al.*, "Visual image reconstruction from human brain activity using a combination of multiscale local image decoders," *Neuron*, vol. 60, pp. 915–29, Jan. 2009.

[5] G. Shen, T. Horikawa, K. Majima, and Y. Kamitani, "Deep image reconstruction from human brain activity," *PLOS Computational Biology*, vol. 15, e1006633, Jan. 2019.

[6] G. Shen, K. Dwivedi, K. Majima, T. Horikawa, and Y. Kamitani, "End-to-end deep image reconstruction from human brain activity," *Frontiers in Computational Neuroscience*, vol. 13, p. 21, 2019.

[7] R. VanRullen and L. Reddy, "Reconstructing faces from fmri patterns using deep generative neural networks," *Communications Biology*, vol. 2, May 2019.

[8] Z. Ren *et al.*, "Reconstructing seen image from brain activity by visually-guided cognitive representation and adversarial learning," *NeuroImage*, vol. 228, p. 117 602, Mar. 2021.

[9] S. Nishimoto, A. Vu, T. Naselaris, Y. Benjamini, B. Yu, and J. Gallant, "Reconstructing visual experiences from brain activity evoked by natural movies," *Current biology : CB*, vol. 21, pp. 1641–6, Sep. 2011.

[10] Y. Kamitani and F. Tong, "Decoding the visual and subjective contents of the human brain," *Nature neuroscience*, vol. 8, pp. 679–85, Jun. 2005.

[11] B. Thirion *et al.*, "Inverse retinotopy: Inferring the visual content of images from brain activation patterns," *NeuroImage*, vol. 33, pp. 1104–16, Jan. 2007.

[12] T. Naselaris, R. Prenger, K. Kay, M. Oliver, and J. Gallant, "Bayesian reconstruction of natural images from human brain activity," *Neuron*, vol. 63, pp. 902–15, Sep. 2009.

[13] K. Kay, T. Naselaris, R. Prenger, and J. Gallant, "Identifying natural images from human brain activity," *Nature*, vol. 452, pp. 352–5, Apr. 2008.

[14] G. Stanley, F. Li, and Y. Dan, "Reconstruction of natural scenes from ensemble responses in the lateral geniculate nucleus," *The Journal of neuroscience : the official journal of the Society for Neuroscience*, vol. 19, pp. 8036–42, Oct. 1999.

[15] C. Chang, J. Cunningham, and G. Glover, "Influence of heart rate on the bold signal: The cardiac response function," *NeuroImage*, vol. 44, pp. 857–69, Nov. 2008.

[16] T. Liu, "Noise contributions to the fmri signal: An overview," *NeuroImage*, vol. 143, Sep. 2016.

[17] Q. Li, V. D. Calhoun, T. D. Pham, and A. Iraji, "Exploring nonlinear dynamics in brain functionality through phase portraits and fuzzy recurrence plots," *bioRxiv*, 2023. DOI: 10.1101/2023.07.06.547922.

[18] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE Transactions on Image Processing*, vol. PP, Aug. 2016.

[19] K. Zhang, W. Zuo, and L. Zhang, "Ffdnet: Toward a fast and flexible solution for cnn based image denoising," *IEEE Transactions on Image Processing*, vol. PP, Oct. 2017.

[20] S. Guo, Z. Yan, K. Zhang, W. Zuo, and L. Zhang, "Toward convolutional blind denoising of real photographs," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2019, pp. 1712–1722.

[21] L.-Z. Chen, Z. Lin, Z. Wang, Y.-L. Yang, and M.-M. Cheng, "Spatial information guided convolution for real-time rgbd semantic segmentation," *IEEE Transactions on Image Processing*, vol. 30, pp. 2313–2324, 2021.

[22] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," *International Journal of Computer Vision*, vol. 128, Jul. 2020.

[23] Y. Quan, M. Chen, T. Pang, and H. Ji, "Self2self with dropout: Learning self-supervised denoising from single image," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2020, pp. 1887–1895.

[24] Q. Li, A. Gomez-Villa, M. Bertalmío, and J. Malo, "Contrast sensitivity functions in autoencoders," *Journal of Vision*, vol. 22, no. 6, pp. 8–8, May 2022.

[25] I. Goodfellow *et al.*, "Generative adversarial nets," in *Advances in Neural Information Processing Systems*, Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Q. Weinberger, Eds., vol. 27, Curran Associates, Inc., 2014.

[26] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[27] W. Maass, "Networks of spiking neurons: The third generation of neural network models," *Neural Networks*, vol. 10, no. 9, pp. 1659–1671, 1997, ISSN: 0893-6080.

[28] G. Raghavan and M. Thomson, "Neural networks grown and self-organized by noise," in *NeurIPS*, 2019.

[29] C. Hartmann, A. Lazar, B. Nessler, and J. Triesch, "Where's the noise? key features of spontaneous activity and neural variability arise through learning in a deterministic network," *PLoS computational biology*, vol. 11, e1004640, Dec. 2015.

[30] R. Beliy, G. Gaziv, A. Hoogi, F. Strappini, T. Golan, and M. Irani, "From voxels to pixels and back: Self-supervision in natural-image reconstruction from fmri," in *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds., vol. 32, Curran Associates, Inc., 2019.

[31] N. Chang, J. Pyles, A. Marcus, H. Mulam, M. Tarr, and E. Aminoff, "Bold5000, a public fmri dataset while viewing 5000 visual images," *Scientific Data*, vol. 6, Dec. 2019.

[32] G. Bontempi, *"Statistical foundations of machine learning" (2nd edition) handbook*. Feb. 2021.

[33] F. F. Li, R. Fergus, P. Perona, and D. Zekrifa, "Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories," *Computer Vision and Image Understanding*, vol. 106, pp. 59–70, Apr. 2013.

[34] J. Soh and N. Cho, "Deep universal blind image denoising," 25th International Conference on Pattern Recognition, Underline Science I, Jan. 2021.

[35] G. Buchsbaum and A. Gottschalk, "Trichromacy, opponent colours coding and optimum colour information transmission in the retina," *Proceedings of the Royal Society of London. Series B. Biological Sciences*, vol. 220, pp. 113–89, 1983.

[36] E. J. Allen *et al.*, "A massive 7t fmri dataset to bridge cognitive and computational neuroscience," *bioRxiv*, 2021.

[37] Q. Li, "Measuring functional connectivity of human intra-cortex regions with total correlation," in *Entropy 2021: The Scientific Tool of the 21st Century*, May 2021, p. 9797.

[38] Q. Li, "Functional connectivity inference from fmri data using multivariate information measures," *Neural Networks*, vol. 146, pp. 85–97, 2022.

[39] Q. Li, G. V. Steeg, S. Yu, and J. Malo, "Functional connectome of the human brain with total correlation," *Entropy*, vol. 24, no. 12, p. 1725, 2022.

[40] Q. Li, G. V. Steeg, and J. Malo, "Functional connectivity in visual areas from total correlation," *arXiv preprint arXiv:2208.05770*, 2022.

[41] Q. Li, S. Yu, K. H. Madsen, V. D. Calhoun, and A. Iraji, "Higher-order organization in the human brain from matrix-based rényi's entropy," in *2023 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW)*, 2023, pp. 1–5.

[42] Q. Li, "Bidirected information flow in the high-level visual cortex," in *Brain Informatics*, M. Mahmud, M. S. Kaiser, S. Vassanelli, Q. Dai, and N. Zhong, Eds., Cham: Springer International Publishing, 2021, pp. 57–66.

[43] Q. Li, "Investigate bidirectional functional brain networks using directed information," in *2022 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 2022, pp. 109–116.

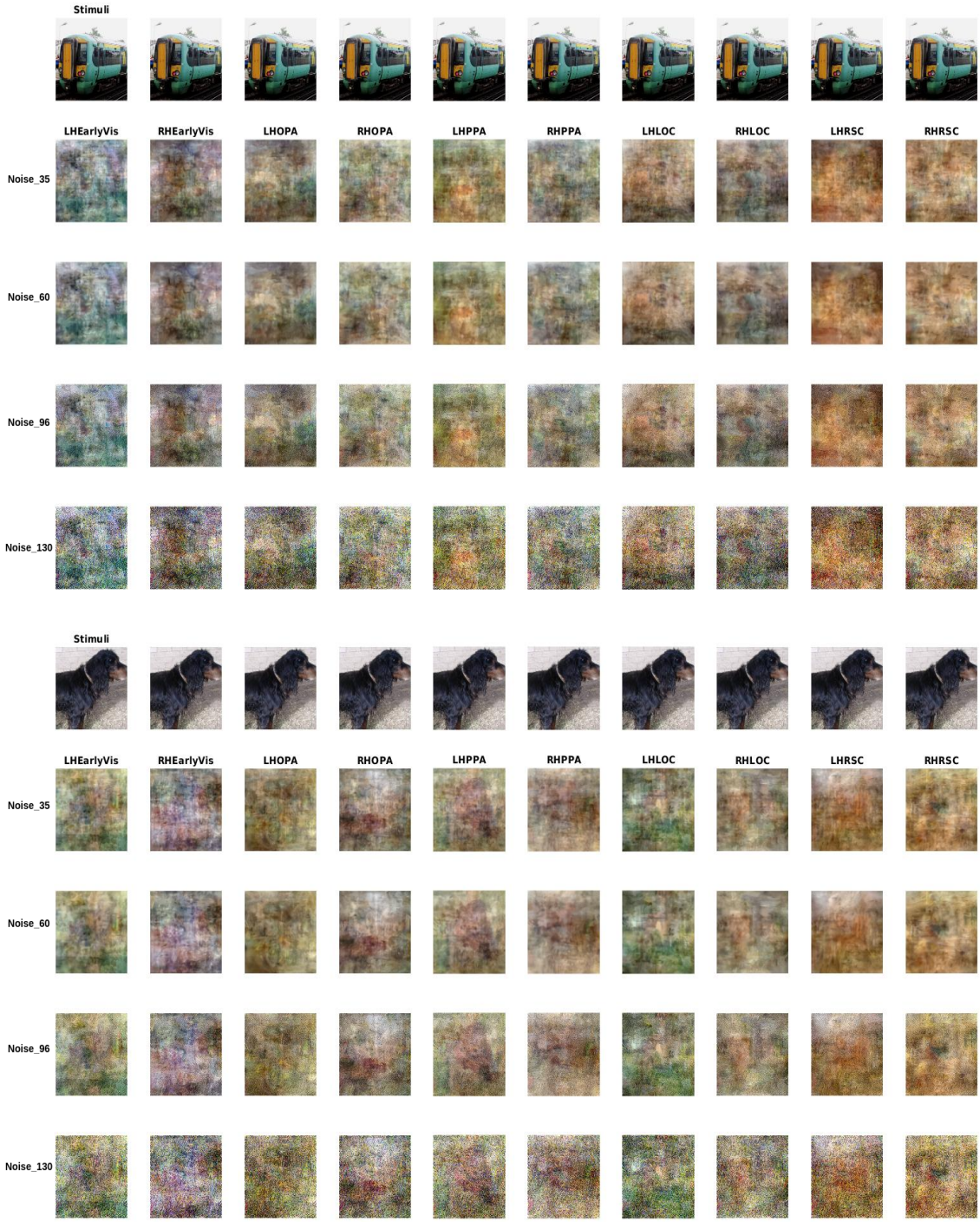

Fig. 9: **Reconstruction of Objects and Animals from Brain Activity Using Linear Regression and State-of-the-Art Denoised Neural Network (DUBD)**.

Fig. 10: **Reconstruction of Animals from Brain Activity Using Linear Regression and State-of-the-Art Denoised Neural Network (DUBD)**.