Visual Image Reconstructed Without Semantics from Human Brain Activity Using Linear Image Decoders and Nonlinear Noise Suppression

Qiang Li[α,β]
[α]Image Processing Laboratory, University of Valencia, Valencia, Spain
[β]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 while considering semantic information. This advancement not only enables the recreation of visual content but also provides valuable insights into the intricate processes occurring within high-order functional brain regions, contributing to a deeper understanding of brain function. However, considering fusion semantics in reconstructing visual images from brain activity is semantic-to-image guide reconstruction and basically ignores underlying neural computational mechanisms, which is actually not real reconstruction from brain activity. In response to this limitation, our study introduces a novel approach that combines linear mapping with nonlinear reconstruction to reconstruct visual images perceived by 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 linear mapping 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 without semantic information. Moreover, the model has strong neural explanatory power because it shares structural and functional similarities with the visual brain.**

***Index Terms*—Visual Image Reconstruction, Brain Activity, Linear Decoding, Deep Autoencoder Denoised Neural Networks, Neural Explanatory, Functional Similarities**

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. The 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 elucidate the complex relationship between visual experiences and brain states, encoding and decoding models were employed to explain the statistical dependence based on features derived from both visual experience and neural signals [2].

In the nascent stages of this exploration, linear mapping decoders held sway as a prominent choice [3]. However, 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. The Blood-Oxygen-Level-Dependent (BOLD) signals, commonly employed in neuroimaging studies, are susceptible to both physiological noise stemming from bodily functions like heart rate variations [4], and physical noise introduced by the MRI machinery [5]. Secondly, linear decoders often falter when tasked with processing high-dimensional and nonlinear functional magnetic resonance imaging (fMRI) data [6, 7]. 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.

With the advent of deep learning models, a transformative shift occurred, giving rise to a diverse array of nonlinear decoding models. This paradigm shift yielded remarkable outcomes, culminating in superior performance in the realm of reconstructing visual content from neural activity [3, 8–17], and among others, have collectively propelled the understanding and application of nonlinear decoding models in the reconstruction of visual information from brain activity. Even though the nonlinear deep net models have nice performance in reconstructing visual images, they lack neural explanations. 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 [5], intrinsic neural noise [18, 19], and even physiological factors like heart rate fluctuations [4]. Overcoming these challenges remains a central focus, as researchers endeavor to refine the precision and fidelity of visual image reconstruction from brain

[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.

activity. Last but not least, 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 [20, 21]. Extensive research has underscored the hierarchical nature of representations within the visual cortex. 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.

In order to make models have better reconstruction while also having good neural explanations, we try to overcome the linear mapping noise problem, considering recent advancements in image denoising have illuminated a path forward. These denoising models can be broadly categorized into three distinct groups. First, we have *Fully-Supervised Denoising Models*, exemplified by deep convolutional neural networks like DnCNN [22], FFDNet [23], CBDNet [24], and SGNet [25]. 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*, 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 [26] and Self2Self [27]. While promising, self-supervised models also grapple with the challenge of restricted applicability. In navigating the landscape of denoising models, 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 to overcome the above-mentioned problems, 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 BOLD signal, a hallmark of 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 models 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 without semantic information. 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 in the domain of computer vision [3, 8–11, 13–17, 20, 28–30]. 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 fMRI. 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 [3]. However, it became evident that linear models 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 ushered in a new era of effective reconstruction models. 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, it usually leverages convolutional neural network (CNN) models to extract characteristic features from images and successfully reconstruct 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)[31], Recurrent Neural Networks (RNNs)[32], Spiking Neural Networks (SNNs) [33], and Diffusion models [30] have made substantial strides in enhancing the performance of visual image reconstruction from brain activity.

III. Methods and Results

A. BOLD5000

Cheng et al. have made the BOLD5000 dataset openly accessible to the public [34]. 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. These images 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. The availability of the dataset 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.

B. Linear and nonlinear decoder training

1) Linear/nonlinear encoding plus decoding: Here, we considered linear regression [35] 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:

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

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:

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

[3]http://sun.cs.princeton.edu/
[4]https://cocodataset.org/
[5]http://www.image-net.org/

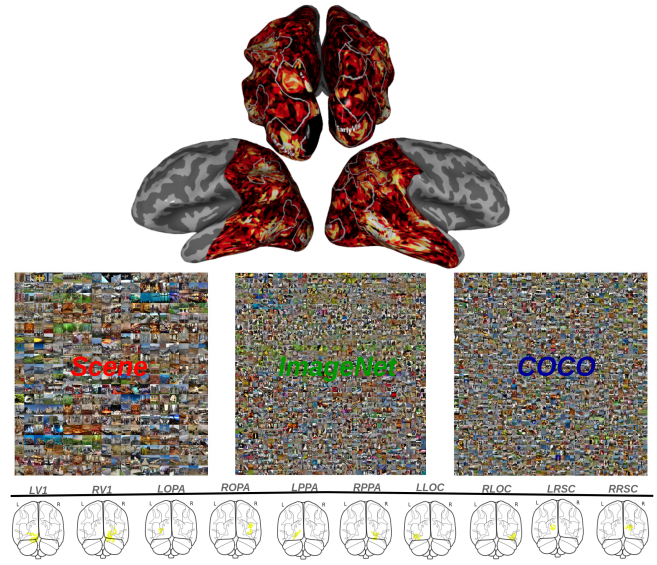

Fig. 1: **Nature scenes selective ROIs.** Regions of interest (ROIs) exhibiting the highest color intensity signify specialized regions within the human visual cortex that are selective for scenes. The brain's activity is depicted on a flattened surface, and corresponding brain images are presented, 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.

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

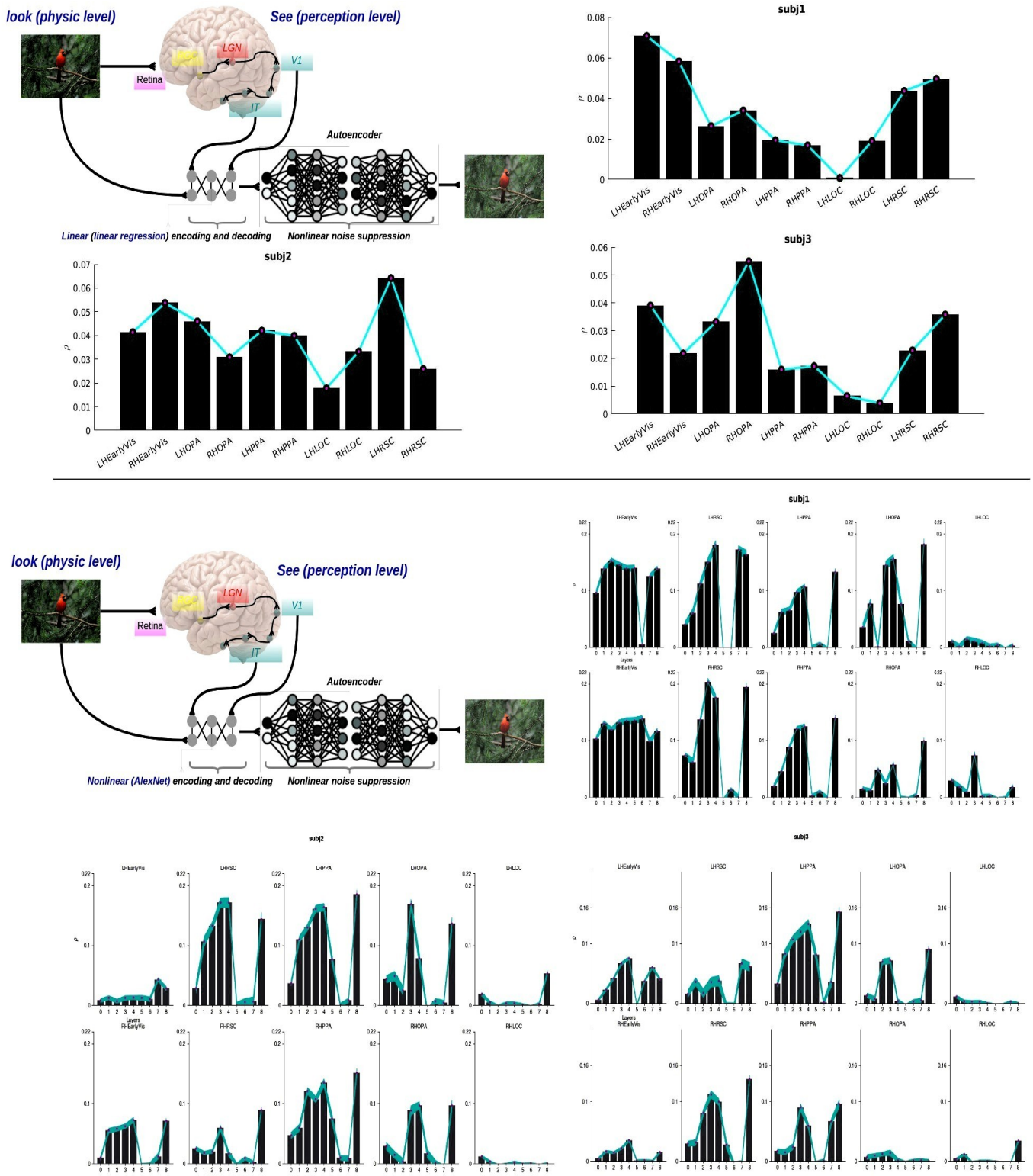

Fig. 2: **The architecture of linear/nonlinear image decoders plus nonlinear noise suppression.** The top figure initially gives an overview of the fundamental structure of linear image decoders (linear regression) combined with a nonlinear noise suppression model (autoencoder), and subsequent figures show the linear encoding accuracy observed in various subjects. Along the x-axis, we have the regions of interest (ROIs), while the y-axis represents the correlation (rho) 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. In the below figure, we replaced the linear image decoder with the nonlinear image decoder model AlexNet, and then the related encoding accuracy across various ROIs with different subjects was measured.

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. The nonlinear deep net model was used to extract the visual image features from each layer of AlexNet, more specifically, after the training process, a straightforward feedforward pass involving an input image was used to extract features and perform image denoising using the deep nets. Each unit within the deep nets 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. Then linear regression was applied between the extracted features and real neural signals. The related predicted neural responses from different visual regions across different subjects with each layer of deep net were present at the bottom of Fig. 2. In general, we see that RSC and OPA high-order visual regions contribute to high decoding accuracy, which totally makes sense considering RSC and OPA are mainly sensitivity-selected responses to complex natural scenes, and the predictive neural response gradually increases with layers in deep nets.

2) Color and texture in the visual brain and deep nets: To ascertain the internal features within deep neural networks, we employed the DeepDream methodology as a conventional means for feature visualization in neural networks [36, 37]. This entails forwarding an initialization image through the network and subsequently calculating the gradient of the image with respect to the activations of a designated layer. Subsequently, the image undergoes iterative modifications to enhance the strength of these activations in the context of AlexNet and Autoencoder, employing the Matlab implementation of the DeepDream method[6]. This process serves to depict images that optimally stimulate specific neurons or layers within these networks. These visual representations exhibit a degree of consistency with research asserting functional parallels between the visual cortex and deep neural networks, such as the emergence of progressively more abstract features [38–40].

The figure 3 reveals that task-demand convolutional neural networks exhibit distinctive features in both low-level and high-level phases of neural networks. For instance, the alexnet and autoencoders primarily manifest white-black (WB), red-green (RG), and yellow-blue (YB) within the convolutional layer 1 (conv1). This pattern aligns with the chromatic-opponent mechanisms found in the primate visual system [41, 42] (refer to Figs. 3**A**, 3**B**, 3**C**). 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 (Fig. 3**A**). 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.

[6]https://www.mathworks.com/help/deeplearning/ref/deepdreamimage.html

C. 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.

1) 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. Furthermore, it shares both structural and functional similarities with the human visual system [29]. Furthermore, the previous study shows that visual brain performance enhances image quality processing when input images are of low quality [43], as shown in Fig. 4**A**.

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 [44]. 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.

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],

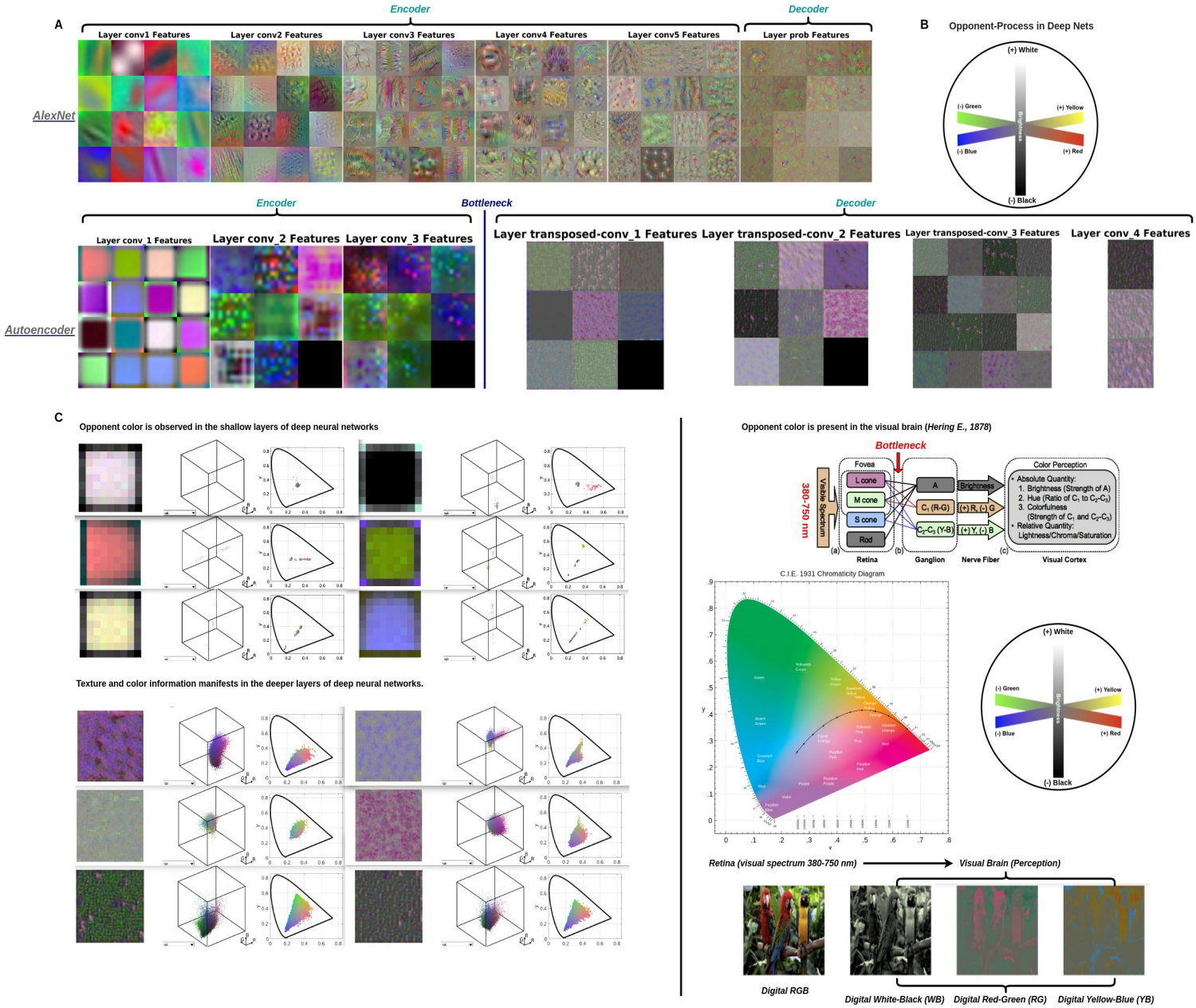

Fig. 3: **Color and texture information in the visual brain and deep nets.** As shown in **A**, the features in AlexNet and Autoencoder were present in both the encoding and decoding phases. We clearly see that color and texture information were present in both deep nets, and the color information was mainly kept in the shallow layers of deep nets, while texture information was mainly kept in the deep layers of AlexNet for better classification of image categories, but both texture and color information were kept in the Autoencoder for better reconstruction of original images. The opponent color process is illustrated in **B**, and it is primarily constructed through anti-inhibition pairwise color channels, including white-black (WB), red-green (RG), and yellow-blue (YB). The chosen opponent color distributions in the shallower layers of deep nets have been thoughtfully represented in our plots. Specifically, we have presented distributions for WB, RG, and YB in the left panel of **C**. These choices align with the chromatic-opponent mechanisms observed within the primate visual system, enhancing the relevance and significance of our visualizations in the right panel of **C**. We provide insight into the color distribution of the receptive fields in the deeper layer. Notably, in this representation, the chromatic-opponent characteristics appear to have dissipated, leading to a more uniform color and texture distribution. The 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. Moreover, we present opponent color processing from the digital RGB images shown on the right side of **C** to better show the opponent color process in the visual brain.

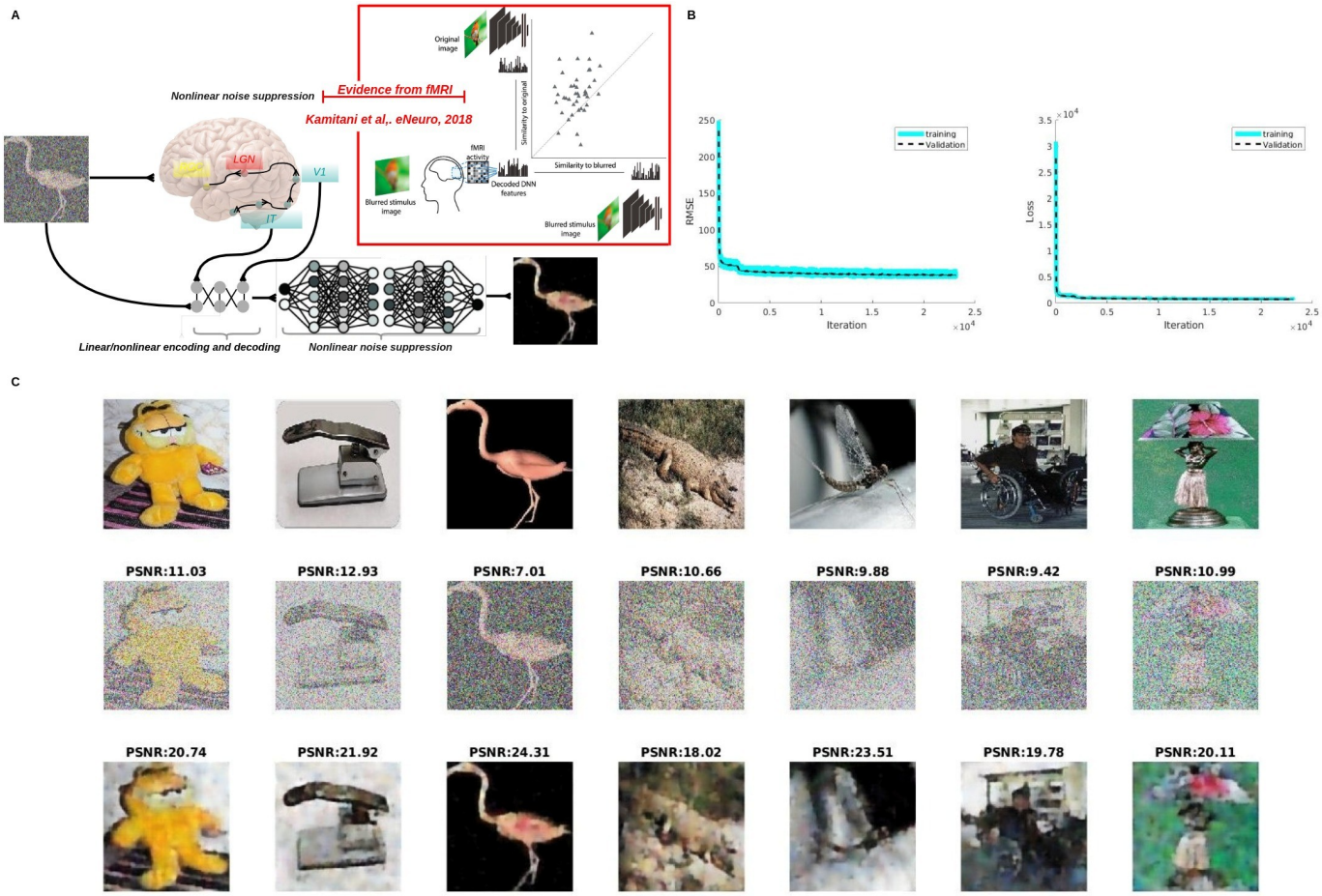

Fig. 4: **Nonlinear noise suppression in the visual brain and deep nets.** It has been previously discovered through related fMRI experiments that the visual brain performs denoise processes [43] that the visual brain may enhance noise-infected images. We trained a deep autoencoder to suppress noise in this instance, taking into account the structural and functional similarities between the visual cortex and the deep autoencoder. The training and validation curves were displayed in **B**, while the denoise performance was displayed in **C**.

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. The final denoise performance of deep nets is shown in Figs. 4**B,C**.

2) 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 [45]. This choice was deliberate, as the DUBD model has consistently outperformed other state-of-the-art denoised neural networks [45]. For a comprehensive understanding of the network's architecture and loss function, we refer readers to the detailed description provided in [45].

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

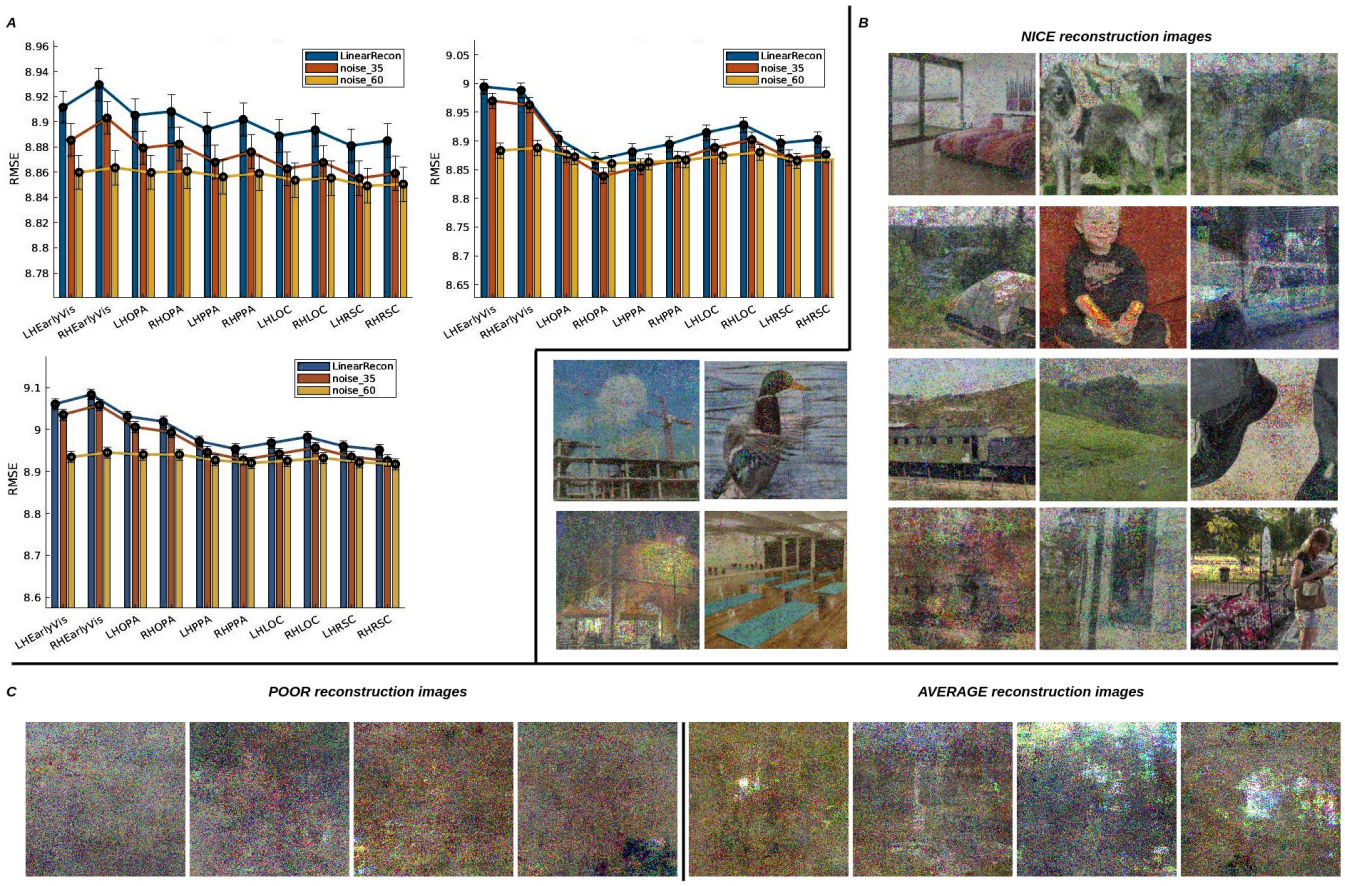

Fig. 5: **Visual image reconstruction from brain activity**. Visual image reconstruction quality was evaluated for four subjects via linear regression plus nonlinear denoised processing across different brain regions in **A**. We reported the Root Mean Square Error (RMSE) for various brain regions that 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, and the 60 noise level is the best choice for suppressing noise. This data provides valuable insights into the impact of denoising on image quality across different brain regions. The super-nice reconstruction images from brain activity were selected across different subjects, as shown in **B**. The very poor and average reconstruction images from brain activity were selected across different subjects, as shown in **C**.

quantified the model's effectiveness in reducing noise and enhancing image quality using the Root Mean Square Error (RMSE) metric, as shown in Fig.5**A**. Meanwhile, we presented the results of visual image reconstruction from various brain regions using linear regression in combination with different levels of nonlinear denoising processes. The findings indicate our capability to predict and reconstruct visual images based on brain activity, devoid of the incorporation of semantic information. Specifically, we have curated a selection of visually impressive reconstructions, average reconstruction images, and reconstructions exhibiting notably diminished quality, as shown in Figs.5**B**, 5**C**.

IV. Conclusions and Discussion

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. At last, it is possible to rebuild visual images from brain activity without the need for complex diffusion models or text-to-image methods. We argue that applying the text-to-image method to reconstruct visual images from brain activity may overlook the underlying neurological mechanism.

In our future research endeavors, we may explore the application of principal component analysis or salinecy mapping to complex nature images 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 [46], 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 [47–52]. 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 [53, 54]. 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 NeurIPS 2023 conference served as the primary inspiration for the main themes of the paper, and after returning from NeurIPS 2023, I made some improvements to the manuscript.

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] T. Naselaris, K. Kay, S. Nishimoto, and J. Gallant, "Encoding and decoding in fmri," *NeuroImage*, vol. 56, pp. 400–10, May 2011.

[3] 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.

[4] 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.

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

[6] 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.

[7] Q. Li, M. Seraji, V. D. Calhoun, and A. Iraji, "Complexity measures of psychotic brain activity in the fmri signal," *bioRxiv*, 2023. DOI: 10.1101/2023.11.10.566647.

[8] 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.

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

[10] 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.

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

[12] 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.

[13] 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.

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

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

[16] 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.

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

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

[19] 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.

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

[21] Q. Li, V. Calhoun, and A. Iraji, "Revealing complex functional topology brain network correspondences between humans and marmosets," *Neuroscience Letters*, vol. 822, p. 137 624, Jan. 2024.

[22] 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.

[23] 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.

[24] 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.

[25] 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.

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

[27] 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.

[28] 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.

[29] 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.

[30] F. Ozcelik and R. VanRullen, "Natural scene reconstruction from fmri signals using generative latent diffusion," *Scientific Reports*, vol. 13, 2023.

[31] 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.

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

[33] 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.

[34] 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.

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

[36] A. Mordvintsev, C. Olah, and M. Tyka, *Inceptionism: Going deeper into neural networks*, 2015. [Online]. Available: https://research.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html.

[37] C. Szegedy *et al.*, "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1–9.

[38] D. Yamins and J. DiCarlo, "Using goal-driven deep learning models to understand sensory cortex," *Nat. Neurosci.*, vol. 19, pp. 356–365, 2016.

[39] D. L. Yamins, H. Hong, C. Cadieu, and J. J. DiCarlo, "Hierarchical modular optimization of convolutional networks achieves representations similar to macaque it and human ventral stream," in *Advances in Neural Information Processing Systems*, C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Weinberger, Eds., vol. 26, Curran Associates, Inc., 2013.

[40] C. F. Cadieu *et al.*, "Deep neural networks rival the representation of primate it cortex for core visual object recognition," *PLOS Computational Biology*, vol. 10, no. 12, pp. 1–18, Dec. 2014.

[41] L. M. Hurvich and D. Jameson, "An opponent-process theory of color vision.," *Psychological review*, vol. 64, Part 1 6, pp. 384–404, 1957.

[42] 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.

[43] M. Abdelhack and Y. Kamitani, "Sharpening of hierarchical visual feature representations of blurred images," *eneuro*, vol. 5, ENEURO.0443–17.2018, May 2018.

[44] 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.

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

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

[47] 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.

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

[49] 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.

[50] Q. Li, G. Ver Steeg, and J. Malo, "Functional connectivity via total correlation: Analytical results in visual areas," *Neurocomputing*, vol. 571, p. 127 143, Dec. 2023.

[51] 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.

[52] Q. LI, V. Calhoun, A. R. Ballem, S. Yu, J. Malo, and A. Iraji, "Aberrant high-order dependencies in schizophrenia

resting-state functional MRI networks," in *NeurIPS 2023 workshop: Information-Theoretic Principles in Cognitive Systems*, 2023. [Online]. Available: https://openreview.net/forum?id=ZgMRaX02ck.

[53] 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.

[54] 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.