

学术硕士研究生

学位论文

**猕猴V4双光子成像数据的深度学习模型研究**

**作 者：林泽瀚**

**导 师：吕海东**

**培养单位：心理学部**

**学 号：202121061031**

**学科专业：心理学**

**完成日期：2024年5月**

**北京师范大学学位论文原创性声明**

本人郑重声明： 所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不含任何其他个人或集体已经发表或撰写过的作品成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。

学位论文作者签名：　　　　　　　　　　日期： 年 月 日

**学位论文使用授权书**

学位论文作者完全了解北京师范大学有关保留和使用学位论文的规定，即：学校有权保留并向国家有关部门或机构送交论文的复印件和电子版，允许学位论文被查阅和借阅；学校可以公布学位论文的全部或部分内容，可以允许采用影印、缩印或其它复制手段保存、汇编学位论文。保密的学位论文在解密后适用于本授权书。

本人签名： 日期：

导师签名： 日期：

猕猴V4双光子成像数据的深度学习模型研究

摘 要

视觉作为人类感知世界的主要方式，承担着向大脑传输大约80%的外部信息的任务，在信息处理中处于核心地位。大脑的视觉皮层可以划分为功能各异的背侧和腹侧两条路径，它们并行处理不同的视觉属性。本项研究着重于视觉皮层的第四区（V4），它位于腹侧通路中，承担着识别物体颜色和形状信息的任务。神经网络模型与视觉系统的紧密联系为现代人工神经网络的设计提供了灵感，因此深入探讨两者之间的关联对于科学发展具有重大意义。在先行研究中，猕猴V4区的电生理数据被用来建立神经元编码模型，这一模型能够生成使神经元放电率最大化的偏好图像。我们尝试将这一模型扩展到双光子钙成像数据上，以探索这种成像技术是否适用于神经网络模型的开发。此外，我们还利用双光子钙成像的优势进行了深入分析，以验证模型的可复现性，并寻找提高模型精度的新方法。研究结果表明，通过调整刺激条件以适应细胞的感受野特性，可以提升模型精度。模型所生成的偏好图像在不同初始条件下展现出了一定的可重复性。然而，同一组细胞在不同数据集下生成的偏好图像相似度有限，实验表明通过交叉训练可以在一定程度上增强图像间的一致性。此外，我们在比较偏好图像相似度的指标时发现，与结构相似性指数（SSIM）相比，CLIP度量的相似度更符合人的视觉感知。这项研究不仅提高了我们对视觉系统机制的理解，同时也证实了深度卷积神经网络（DCNN）模型在模拟V4神经元钙成像结果方面的适用性，能够为V4的进一步研究提供帮助。

关键词**：**腹侧通路, 猕猴视皮层V4区, 双光子钙成像, 深度学习, 人工神经网络

DEEP LEARNING MODEL RESEARCH ON MACAQUE V4 TWO-PHOTON IMAGING DATA

**ABSTRACT**(四号加粗，居中)

Visual perception serves as the primary modality through which humans sense the world, responsible for transmitting approximately 80% of external information to the brain and occupying a central position in information processing. The visual cortex of the brain is divided into the functionally distinct dorsal and ventral pathways, which process different visual attributes in parallel. This research focuses on the fourth area of the visual cortex (V4), which is part of the ventral stream and is tasked with recognizing the color and shape information of objects. The relationship between neural network models and visual system has provided inspiration for the design of modern artificial neural networks, making the exploration of interconnection significantly important for scientific progress. In previous studies, electrophysiological data from the V4 area of macaques was used to build a neuronal encoding model capable of producing images that maximize the firing rates of neurons. We have attempted to extend this model to two-photon calcium imaging data to investigate the suitability of this imaging technique for the development of neural network models. Additionally, we have taken advantage of the benefits of two-photon calcium imaging to conduct in-depth analyses to verify the reproducibility of the model and to seek new methods to improve its accuracy. The study shows that model precision can be enhanced by adjusting the stimulus conditions to cater to the specific receptive field properties of the cells. The preference images generated by the model demonstrate high reproducibility under various initial conditions. However, the similarity of preference images generated by the same set of cells across different datasets was limited. Moreover, when comparing indicators of preference image similarity, we discovered that similarity metrics from CLIP are more aligned with human visual perception than the Structural Similarity Index (SSIM). This research not only advances our understanding of the mechanisms of the visual system but also provides valuable insights for the development of neural network models in the field of artificial intelligence.

**KEY WORDS:** Ventral pathway, Area V4 in macaque, Two-photon calcium imaging, Deep learning, Artificial neural network

目 录（小二号，黑体，居中）

[第1章 研究背景 1](#_Toc155680736)

[1.1 视觉并行通路理论 1](#_Toc155680737)

[1.2 V4区的位置及主要功能区 2](#_Toc155680738)

[1.2.1 V4区的定位 2](#_Toc155680739)

[1.2.2 V4的功能和分区 2](#_Toc155680740)

[1.3 神经网络模型与视觉系统的联系 3](#_Toc155680741)

[1.3.1 神经网络模型的起源 3](#_Toc155680742)

[1.3.2 神经元DCNN模型 3](#_Toc155680743)

[1.3.3 组里的相关工作 4](#_Toc155680744)

[1.4 科学问题与研究思路 4](#_Toc155680745)

[第2章 实验操作步骤 6](#_Toc155680746)

[2.1 实验材料 6](#_Toc155680747)

[2.2 动物手术 6](#_Toc155680748)

[2.3 麻醉猴内源光学成像 6](#_Toc155680749)

[2.3.1 内源光学成像原理 6](#_Toc155680750)

[2.3.2 内源信号采集 7](#_Toc155680751)

[2.3.3 视觉刺激-内源信号光学成像 7](#_Toc155680752)

[2.4 麻醉猴双光子成像 8](#_Toc155680753)

[2.4.1 双光子成像原理 8](#_Toc155680754)

[2.4.2 双光子信号采集 8](#_Toc155680755)

[2.4.3 视觉刺激-双光子成像 9](#_Toc155680756)

[第3章 数据分析方法 10](#_Toc155680757)

[3.1 OI数据分析-SVM map 10](#_Toc155680758)

[3.2 双光子成像数据处理 10](#_Toc155680759)

[3.3 深度学习模型构建 11](#_Toc155680760)

[3.3.1 神经元DCNN编码模型 11](#_Toc155680761)

[3.3.2 AlexNet预训练卷积层的选择 12](#_Toc155680762)

[3.3.3 单位点与多位点模型 13](#_Toc155680763)

[3.3.4 模型评估与训练 13](#_Toc155680764)

[3.3.5 偏好图像生成 14](#_Toc155680765)

[3.3.6 卷积核消融实验 14](#_Toc155680766)

[3.3.7 图像相似度评估 15](#_Toc155680767)

[第4章 实验结果 17](#_Toc155680768)

[4.1 改变输入图片提升模型精度 17](#_Toc155680769)

[4.1.1 目标 17](#_Toc155680770)

[4.1.2 操作 17](#_Toc155680771)

[4.1.3 结论 18](#_Toc155680772)

[4.2 不同数据集的比较 18](#_Toc155680773)

[4.2.1 目标 18](#_Toc155680774)

[4.2.2 基本结果比较 19](#_Toc155680775)

[4.2.3 最优刺激比较 21](#_Toc155680776)

[4.2.4 结论 22](#_Toc155680777)

[第5章 讨论 23](#_Toc155680778)

[5.1 23](#_Toc155680779)

[参考文献 25](#_Toc155680780)

[附 录 26](#_Toc155680781)

[攻读学位期间取得的学术成果 27](#_Toc155680782)

[致 谢 28](#_Toc155680783)

图 目 录

[图1 神经元编码模型框架 12](#_Toc156399203)

[图2 CLIP结构图[10] 15](#_Toc156399204)

[图3 crop示例图，从左到右分别是crop90,crop150,crop200,crop250和初始图片 18](#_Toc156399205)

[图5 数据8,9刺激图片对比 21](#_Toc156399206)

[图7 数据8,9神经元反应统计分析 21](#_Toc156399207)

[图9 数据8,9 trial correlation 对比 22](#_Toc156399208)

[图10 单独训练下的偏好刺激对比，每一张图代表在数据8,9上训练的两个模型针对某个特定神经元生成的偏好图片对比 22](#_Toc156399209)

[图11 交叉训练下的偏好刺激对比，每一张图代表在交叉数据上训练的两个模型针对某个特定神经元生成的偏好图片对比 23](#_Toc156399210)

[图12 部分神经元随机初始值偏好图片对比 24](#_Toc156399211)

[图13 偏好图片进行SSIM计算 24](#_Toc156399212)

[图15 通过掩码后进行比较 25](#_Toc156399213)

表 目 录

[表1 Crop 训练效果图 18](#_Toc156573780)

[表1 SSIM评估结果表 23](#_Toc156573781)

[表2 CLIP评估结果表 25](#_Toc156573782)

# 研究背景

(黑体三号，居中，章序与标题之间空1个汉字符）

## 视觉并行通路理论

哺乳类动物的神经系统由无数精密相连的神经网络构成，其中大脑扮演着系统的指挥官角色，掌管着生命体的全部活动，并管理着与外部世界的信息交换。在漫长的演化旅程中，人类及其他非人类灵长动物的大脑架构演变成了一座错综复杂的神经迷宫。20世纪初，Brodmann基于大脑皮层的细胞构造，将其精心划分为43个区域，这一数字在后续研究中已经增至52个。这些划分不仅定义了大脑皮层的地图，而且与大脑的功能紧密相连，每个区域都承担着特定的角色。

对于人类以及其他高等灵长类动物来说，视觉系统是接受外界信息的主要通道。超过七成的感官信息是通过视觉得到的，这些信息涉及物体的亮度、形状、颜色、运动（方向与速度），以及立体视觉，即深度感知。相关的大脑区域主要集中于枕叶、颞叶和顶叶。过去数十年的科研揭示了视觉信息并行处理路径的理论，这一理论深刻总结了视觉信息在大脑各部位的传递流程，并对理解视觉处理的生物学机制提供了极为重要的理论支撑。

1982年，Mishkin和Ungerleider 根据行为学，电生理，形态学和脑损伤研究结果，定位了视觉皮层的并行通路：一条位于大脑腹侧，被称为腹侧视觉通路（简称腹侧通路，ventral visual pathway），主要包括纹状皮层（striate cortex，又称为初级视觉皮层V1），纹外皮层（extrastriate cortex 包括V2、V3、V4），延伸至下颞叶区域（TEO和TE），主要负责物体视觉（object vision），这些区域与腹侧边缘系统和额叶的链接可能是物体信息与情绪，动机等认知功能联合的结构基础；另一条通路位于大脑背侧，称为背侧视觉通路（dorsal visual pathway），从纹状皮层，纹前皮层至下顶叶区域，主要负责空间视觉（spatial vision），用于定位特定的物体，指导相关的行为。同年年末，根据过去十几年非人灵长类动物的电生理和形态学研究结果，Van Essen 和 Maunsell 对视觉皮层并行通路进行了更精细地描述：腹侧通路主要涉及脑区为V1，V4，VP和IT，背侧通路为V1，MT和MST，并区分了腹背侧通路主要加工的视觉信息，认为腹侧通路主要涉及加工颜色和形状信息，背侧主要涉及加工运动信息。

纵观这些研究，不同时间和方式的研究发现均支持以下理论即视皮层不同脑区接受信息的来源有所不同，传递信息的去向也不尽相同。正是这种神经信号处理通路的不同为处理不同的视觉信息提供了必备的神经基础，让视觉信息在视皮层中的分类和加工更加高校，同时也使视皮层中的信息交流更加准确。在本研究中，我们重点关注腹侧视觉通路中的V1和V4脑区，并尝试使用神经网络模型来拟合V1、V4神经元的反应。

## V4区的位置及主要功能区

### V4区的定位

在视觉科学领域，视皮层四区（V4区）最初被认定为颜色处理的关键脑区。早期的研究，如Semir Zeki等人的工作，通过损伤实验确定了位于猕猴脑内月状沟前和上颞叶沟之间的一个独立脑区，并将其命名为V4区。后续研究通过单细胞记录技术揭示，V4区的神经元对特定波长的视觉刺激表现出响应，表明这些神经元具有颜色选择性。更为有趣的是，Zeki发现这些具有颜色选择性的神经元按功能柱状结构分布。

随着研究的深入，科学家们在V4区发现了对朝向和形状有选择性反应的神经元。因此，位于视觉处理腹侧通路中的V4区，现在被普遍认为是处理物体颜色和形状信息的重要脑区。

关于V4区的确切定位和结构细节，科学界依然存在一些争议。然而，关于V4区的大致位置和分区已达成较广泛的共识。V4区主要位于脑的腹侧部位，处于月状沟（lunate）与颞上沟（STS）之间，后缘与V2区相连，前缘则延伸至内颞叶区域IT。V4区的一部分隐藏于月状沟和颞上沟的沟壑中，因此在大脑皮层表面并不容易观察到完整的V4区。

### V4的功能和分区

V4区位于视觉信息传递路径的中枢地带，扮演着将初级视觉皮层（V1区）的输入与内颞叶区域（IT区）的高级视觉处理连接起来的关键角色。在视觉信息从简单到复杂的转化过程中，V4不仅仅是一个信息的中转站，而是在这一连串复杂且细致的处理过程中，担任分工与整合的复杂任务。这种高效的神经机制为大脑的高级区域处理极其复杂的视觉刺激奠定了基础，使我们能在丰富多变的环境中形成接近实际情景的认知。

研究揭示了V4区内部神经元的多样性：不仅有对特定颜色和形状敏感的选择性神经元，还有对复杂图案和运动趋势作出反应的神经元。这种多样性让我们能迅速而准确地在自然环境中辨识物体。特别地，V4区的颜色选择性神经元机制支撑了我们在五彩斑斓的世界中维持颜色恒定性的能力。此外，V4区的深度信息处理能力，使我们能够感知物体的立体结构。

来自我们实验室的最新研究进一步拓宽了对V4区功能的认识，发现这一区域内还存在对颜色和运动方向敏感的选择性神经元。这表明，V4区不仅参与颜色、形状、深度、朝向的处理，还涉及运动方向的识别。V4区内不同种类神经元的独特属性共同构成了一个复杂且精细的视觉信息处理网络。

## 神经网络模型与视觉系统的联系

### 神经网络模型的起源

20世纪中叶，Hubel等人[1]在猫的初级视觉皮层（V1）中，发现了两种主要的细胞类型——简单细胞和复杂细胞。其中，简单细胞会对在某一个特殊空间位置的亮暗光栅有反应，每个细胞对于光栅的朝向有选择性，并且随着逐渐偏离偏好方向反应值降低（调谐曲线）。而对复杂细胞来说，没有那么多严格的反应特性，它们仍然拥有朝向选择性，但是与其他方向的反应差别不大。1980年，Fukushima 提出了Neocognitron模型[2]，是现代CNN的先驱。它包含两种主要的细胞类型。S细胞模拟的是简单细胞，重复出它的特性，二维的权重被分配到输入图片的每个位置。S细胞的平面具有视野的布局，所有细胞共享相同的偏好视觉特征，并且在一个层中存在多个平面。C细胞模拟的是复杂细胞，它得到将同一平面不同位置的简单细胞经过非线性运算的结果。一层的简单细胞和复杂细胞代表V1的基础运算，Neocognitron 仅仅是重复上面的过程。因此，第一层的复杂细胞作为第二层简单细胞的输入。到了19世纪90年代，许多类似的层级模型被提出，其中最突出的是HMAX，它对简单细胞集合的反应值进行使用简单的“max”操作确定复杂细胞的反应，在图像多变时表现优异。由于有图片对应的人类实验数据，模型的结果能够直接与人类快速视觉分类能力对比，类似的模型可以被用来拟合V4神经元对复杂形状刺激的反应[3]。

在2014年的一项研究中，研究人员记录了猕猴在观看复杂物体图片时的细胞活动，并对真实的V4和IT神经元的反应与人工神经网络的反应进行了回归分析。研究发现，那些在物体识别任务中表现更佳的神经网络，往往能够更准确地预测神经反应[4]，相似的结果也在视频分类任务中观察到[5] 。文章提到，作者构建的人工神经网络中的倒数第一层在预测IT神经元活动方面表现最优，而倒数第二层则更适合预测V4神经元的活动。这些发现在一定程度上表明，人工神经网络拥有与视觉系统腹侧通路类似的层次结构与处理特性。

### 神经元DCNN模型

在2019年，Dicarlo等研究人员[6] 通过向猕猴的V4区域植入一个由96个微型电极组成的阵列，成功采集了猴子神经元对图片刺激的放电反应数据。基于这些数据，他们设计了一个人工神经网络模型，这个模型不仅能够精准预测神经元的放电模式，还能够根据这些预测生成专门的控制图片，这些控制图片的目的是为了最大化特定神经元的放电率。

在合成这些控制图像时，研究团队采用了两种方法：一种是“拉伸”法（stretch control），该方法致力于最大程度地增加目标神经元的活动，尽管在此过程中其他神经元的活动也可能会有所提高。另一种是“独热群”法（one–hot–population control），这种方法在增强目标神经元响应的同时，还会刻意抑制周围其他神经元的活动。

虽然这些生成的刺激图像在外观上可能不够自然，也不直接关联到网络训练时使用的图像，但它们在实际应用中却能有效促使神经元活动超过常规水平。这一结果进一步验证了人工神经网络在捕捉视觉处理流程中基础构成方面的有效性，突显了其在模仿生物视觉系统方面的潜力。

前人的研究认为V4脑区是视觉物体识别中的一个关键环节，其神经元表现出对各种刺激特征（如朝向、颜色、弧度、空间频率（SF）、纹理以及亮度）的多维度选择性。然而，究竟哪些特征是V4神经元的主要处理对象，以及这些特征与V4脑区已知的功能结构（包括朝向、颜色、曲线和运动）之间的具体关联，目前仍然没有很明确的结论。

### 组里的相关工作

在前人研究的基础上[6]，我们采用了双光子钙信号成像技术，记录了522个分布于不同功能区的V4神经元对超过800张自然图像和人工刺激的响应。通过深度卷积神经网络（DCNN）对这些V4神经元的活动进行建模，我们对模型中的神经元感受野进行了深入分析。分析结果揭示出曲率、方向和颜色是构成神经元群体反应的三个主要维度，并且这些特征与神经元所处的功能区域高度一致。对神经元反应进行的降维分析进一步支持了这一发现。

该研究确认了颜色、方向和圆弧形状是V4神经元的关键反应特性，而这些特性恰好也是神经元所在位置的主要功能特征。这表明V4区域的功能架构反映了其主要处理的视觉信息特征。同时，我们的工作验证了DCNN模型在模拟V4神经元的钙信号成像结果方面的适用性，为未来深入探索V4区域的视觉特性奠定了坚实的基础。

## 科学问题与研究思路

在上一阶段的研究中，我们构建了一个深度学习模型，用于处理V4区双光子成像数据。该模型经过适量双光子数据训练后，能够准确拟合神经元活动，并生成针对特定神经元的优化刺激。本研究旨在基于现有模型进行深入探索，挖掘其潜在应用，并提高模型的拟合精度。

**问题1：DCNN神经元模型是否适用于双光记录的神经元反应？**

Dicarlo等研究人员[6] 在猕猴电生理学数据上已经取得了卓越的成果。但是，是否可以将DCNN神经元模型应用于双光子记录的神经元活动数据，并取得同样的效果呢？

双光子钙成像技术的一个显著优点是能够在不损伤细胞的情况下，进行长期和重复的实验，从而获取更加可靠的神经元活动数据。那么，双光子钙成像技术的先进性是否能够为我们提供更加丰富的结果？此外，我们亦需评估双光子钙成像中trial相关性对结果的具体影响程度。

**研究思路：**

在数据采集阶段，我们计划执行多次实验，从多个trials中收集数据，并将这些数据平均化以构建数据集。我们将采用随机抽样方法来评估增加trials数量对模型性能的影响。

**问题2：现有的模型有哪些问题，怎样提高？**

与Dicarlo[6]团队在猕猴电生理数据上的成果相比，我们观察到模型在双光子数据上的精度较低。这引发了关于模型当前局限性的思考，并促使我们审视实验方案的设计合理性。如何才能进一步提升模型的精度呢？

**研究思路：**

考虑到感受野的特性，我们将尝试改变刺激的呈现方式，比如裁剪、置灰等，以探索最适合模型训练的刺激形态。

对比同一组神经元在不同刺激集下的反应数据。通过分析两组数据的差异，我们希望能够更深入地理解模型的工作原理，引导后续实验的设计，从而探究提高模型精度的可能方案。

# 实验操作步骤

## 实验材料

在本次研究中，所有参与实验的动物均源于国内一流的猕猴繁育基地，并在北京师范大学的认知神经科学与学习国家重点实验室的专业动物设施内得到妥善饲养。在实验过程中，所有涉及动物的操作均严格遵守美国国立卫生研究院（NIH）制定的动物使用规范，并已获得北京师范大学实验动物管理和伦理委员会的正式审查与授权。本研究的数据收集涉及到3只动物，包括2只恒河猴和1只食蟹猴，他们的3个大脑半球均参与了研究，我们在所有的大脑半球上进行了内源信号光学成像以及双光子钙信号成像实验。

## 动物手术

动物手术主要包括术前准备、手术过程和术后维护。

动物先用氯胺酮（10mg/kg）在猴笼进行初始麻醉。之后转移到手术室，动物头部固定于立体定位仪上，在呼吸机和异氟烷（1.5%-2.5%）麻醉。

对动物进行开颅手术，开颅窗口直径为22mm，窗口中心根据研究脑区（V1或V4）有所调整，距中缝18-35mm，距后ridge 15-25mm。打开硬脑膜，根据lunate, STS 和IOS的脑沟位置确定视皮层V1, V2和V4区。我们先对暴露视皮层进行内源信号光学成像，之后依据光学成像所得的功能图，对目标功能区进行定点病毒注射。工具病毒为AAV9.Syn.GCaMP6S.WPRE.SV40 (CS1282, titer 3.34e13 (GC/ml), Addgene)，注射深度为500μm, 每点注射量为500nL。病毒注射后密封颅骨窗口，并在6周后再次手术植入一个密封光学窗口（直径13mm）。之后所有的内源信号光学成像和双光子钙信号成像都在这密封的光学窗口上进行，每次成像实验间隔7-10天。

术后会定期对小室 (chamber) 进行清理维护，确保皮层的状态保持正常。

## 麻醉猴内源光学成像

### 内源光学成像原理

内源性光学成像技术是通过监测大脑皮层血氧水平的动态变化来揭示群体神经元对视觉刺激的响应机制。这种响应体现为神经活动的增强，随之带来的是局部脑组织的耗氧量上升，从而促使含氧血红蛋白向去氧血红蛋白转换，这一过程影响了血液对光的吸收与反射特性。在实验设置中，我们选择了波长为630纳米的红光作为照射源，这是因为去氧血红蛋白对此波长的红光吸收率较高，而反射率较低。随后，利用高灵敏度的CCD成像系统来精确记录由此产生的反射光强度变化，从而实时追踪和分析脑内血氧变化，为理解视觉刺激与神经元活动间的相互作用提供了一种手段。

### 内源信号采集

在手术当天，为了进行内源性光学成像，我们采取了无成像窗的技术手段。显露出的皮层被玻璃片覆盖和平压，而玻璃片与皮层之间的空隙被琼脂填补，确保成像区域的稳定。随后，在成像窗植入之后，便可进行直接成像。在成像过程中，动物起初采用异氟烷吸入麻醉，随后转为静脉推注麻醉剂（第一例实验中是丙泊酚与舒芬太尼的混合，第二和第三例实验则使用了氯胺酮），同时维库溴铵的注射用以保持眼球位置的固定。动物的生命体征，包括心率、体温、血氧饱和度以及呼出气中二氧化碳水平被持续监测，确保麻醉水平的适宜，并据此调整麻醉剂的用量。

为了确保动物的瞳孔充分敞开，我们使用了0.5mg/ml的阿托品溶液，并通过精心选择的角膜接触镜来调整视线聚焦在57厘米远的屏幕上。利用632纳米波长的红光进行照射，皮层反射率的变化——即内源性血氧信号——被Optical Imaging公司的Imager 3001系统以4赫兹的频率捕获。图像的分辨率为540×654像素，这对应于一个15.5毫米×19毫米的成像区域。在每个实验周期中，视觉刺激持续3.5秒钟，血氧信号的记录从刺激前的0.5秒开始直到刺激结束，总共记录16帧图像，每次刺激之间的间隔时间大约为6秒。

### 视觉刺激-内源信号光学成像

在本次研究中，视觉刺激的产生由Cambridge Research Systems Ltd. 的ViSaGe系统控制，并通过位于动物正前方57厘米的21英寸CRT显示器（SONY CPD-G520型号）呈现给实验动物。这款显示器具备100Hz的刷新率，并且经过精确的gamma校正，以确保图像质量的准确性和稳定性。我们使用了动态的全屏正弦波光栅刺激来获得Orientation map和 Color map。这些光栅刺激设计有两个空间频率——0.25和1周期/度，并在8个不同的方向上移动（以45度为间隔，从0度至315度），所有刺激的时间频率均设为4周期/秒。无论是黑白光栅还是红绿光栅，它们的平均亮度都被设定为28.92坎德拉/平方米。

为了生成 Curvature map，我们采用了运动的全屏形状刺激，如圆形和三角形。这些形状刺激的直径均为2.5度视角，它们按照3度的步进间隔排列。刺激中的亮区亮度高达111.2坎德拉/平方米，其边缘宽度为0.2度，而背景亮度则保持在20.57坎德拉/平方米。在每一次试验中，这些形状刺激将在随机选定的一个方向上移动（从8个可能的方向中选择），运动速度统一为4度/秒。

通过这种方法，我们能够创造出一系列刺激，旨在精确测定和映射动物视觉皮层的反应，进而分析其对不同形状、颜色和运动方向的敏感度。

## 麻醉猴双光子成像

### 双光子成像原理

双光子激发技术是一种独特的成像方法，它依赖于两个低能量光子—通常是红外或近红外光—在极小体积内几乎同时与单个荧光分子相互作用。这一过程中，这两个光子共同提供足够的能量使荧光分子激发到较高能级，然后这个分子在返回基态时发射一个光子，产生荧光。这个精巧的物理现象使得科学家能够穿透更深的组织层，以实现对活体大脑内部结构更为精细的成像，进而在三维空间中探测细胞层面的活动。

双光子成像系统通过强度极高的飞秒激光脉冲、高数值孔径物镜以及敏感的光电倍增管（PMT）检测器相结合，实现了对生物组织的高分辨率成像。在此技术的帮助下，我们可以以精确的控制和优异的深度穿透能力，捕获立体图像，并洞察到不同深度的神经元活动。

在神经科学领域，钙离子的浓度变化被广泛地作为神经元活动的一个可靠指标。通过双光子钙成像，我们能够实时追踪神经元内的钙浓度变化。这项技术涉及先将细胞膜上的钙通道转染以发出绿色荧光，然后使用双光子显微镜发出的红外激发光线捕捉这些荧光信号。这样，我们可以通过观察荧光强度的变化，得到神经元活动强度的直观表示，从而在神经信号传递的复杂网络中，精准地描绘出活动图谱。

### 双光子信号采集

为了进行双光子显微成像实验，动物的准备和麻醉流程与内源信号光学成像的步骤保持一致。双光子显微镜使用的是Bruker Nano公司生产的Prairie Ultima IV型（In Vivo）双光子显微镜，而激光发射设备则是COHERENT品牌的Chameleon激光器。在这些成像会话中，激光的波长被设定在980纳米，功率范围介于30至45毫瓦。显微镜配备了Nikon品牌的16倍水浸物镜（数值孔径为0.8），使得对应的成像视野达到830微米乘以830微米的大小。

双光子扫描采取的是Galvo模式，以每秒1.3帧的速率进行，图像分辨率为512像素乘以512像素。为了同步视觉刺激与显微镜的成像过程，我们精确地将每次视觉刺激的开始与双光子扫描的下一帧启动时间对齐。

由于显微镜物镜的设计是垂直于样本的，因此在成像时，我们采用特制的旋转夹具将动物的头部旋转至水平位置，与显微镜镜头对齐，同时让动物的身体侧卧以维持其稳定。成像过程中，由于动物的呼吸和心跳，脑组织可能会发生轻微的位移。我们实时监控细胞的形态，并在必要时进行在线调整，以确保皮层在XY平面上的位移不超过50微米，在Z轴上的位移不超过150微米。

在本研究中，我们在3个半球的8个不同位置进行了详尽的双光子成像（如图Figure Sx所示）。成像深度范围在155至295微米之间，每个成像点只进行了一次特定深度的双光子成像。

### 视觉刺激-双光子成像

在本项研究中，视觉刺激是由ViSaGe系统产生，并通过位于试验动物正前方57厘米处的一台21英寸DELL E1913Sf LCD显示器进行展示。该显示器的刷新率设定为60Hz，经过精确的gamma校正，以保障视觉刺激的色彩和亮度准确传达，最大亮度调至80.82坎德拉/平方米。所有的视觉刺激都是在一个中性灰色背景上进行展示的，该背景亮度为40.41坎德拉/平方米。为了实验的需要，我们仅向动物的左眼展示视觉刺激，而右眼则保持闭合。

为了精确定位神经元的感受野，初始步骤中，我们通过手动移动一个圆形方波运动光栅（直径1到3度）并实时观察皮层的荧光响应来识别感受野的大致区域。确定感受野后，我们以此区域为中心，在一个5乘5的网格内，随机在其中一个格点上展示直径1到3度的视觉刺激。这些刺激包括三种形状——方波光栅、三角形和圆形，以及四个不同的运动方向（从45度至315度），且运动速度固定为每秒4度。每一刺激的呈现时间为1.5秒，在此期间会运动两次，其间隔为0.8秒。刺激展示完毕后，我们会立即对数据进行处理，以确定当前成像区域内细胞群的感受野位置和范围（采用高斯函数拟合，并以高斯函数高度的一半为阈值）。

在探究神经元对自然视觉刺激的响应方面，我们构建了一套包含800张彩色图片的自然客体刺激库，这些图片涵盖了450类不同物体，全部来源于www.freepngs.com网站。对于V4区域的5个位点，刺激库中还额外增加了50种人工形状刺激，所有刺激均未经亮度调整。在实验过程中，所有刺激的大小被统一调整到感受野直径的80%，并以随机顺序进行呈现。每个刺激的展示时间为2秒，每两次刺激之间间隔3.4秒。为了避免视觉适应现象，刺激的大小会在呈现前1秒内从感受野直径的80%逐渐增大至95%，然后在后1秒再缩小至80%。在一轮刺激展示完毕后，我们会重新测定感受野的位置，进而准备下一次的实验周期。通常情况下，一个实验会完成5个周期。在两次实验之后，我们会获得10个周期的数据，并将这些数据用于深度卷积神经网络（DCNN）模型的训练，以此来识别出各个神经元的偏好图像。在第三轮实验中，我们将一些神经元的偏好图像加入到刺激集中，以进一步验证模型的预测效果。

# 数据分析方法

## OI数据分析-SVM map

在我们的研究中，领域的性质与尺寸主要通过支持向量机（SVM）图来确定。SVM算法相较于传统的差异图像，更有效地抑制了血管引起的噪声和低频信号干扰，因此在提高信噪比和区分皮层不同功能区域方面表现出更加优越的性能（参考文献：Xiao et al., 2008）。所有的光学成像（OI）数据分析工作均在MATLAB 2017版本（The MathWorks, Natick, MA）中完成。

首先，我们计算每个试验周期中对每种刺激的内源性信号响应，即ΔR/R值。这里的ΔR/R定义为（第8到第16帧的平均反射率 - 第1到第4帧的平均反射率）除以第1到第4帧的平均反射率。接着，利用这些响应图，我们进一步生成SVM图。具体来说，通过SVM算法对两类刺激的响应图进行分类，每个像素点在分类中的权重反映了对应皮层区域在区分这两类刺激上的贡献程度。权重的大小指示了区分度的高低，而权重的正负值表示了该像素点对特定刺激的偏好。由此，我们可以将所有像素点的权重绘制成一幅图，揭示皮层的功能分布。用于此部分的Matlab SVM程序由Chih-Jen Lin提供（LIBLINEAR: A Library for Large Linear Classification, 2008; 可在https://www.csie.ntu.edu.tw/~cjlin/liblinear/获取）。

对于每一张生成的原始SVM图，我们首先应用高斯低通滤波（滤波器尺寸为5，标准差为1）进行平滑处理。然后，利用绿色血管图将成像区域划分为皮层区和背景区，其中背景区包括血管区和chamber外区域。随后，我们采用背景区灰度值分布的标准差对SVM图进行归一化处理（即计算z-score），并以±2倍标准差作为确定领域显著性的阈值（领域尺寸需大于50像素）。同时，我们还将各个双光子成像图像与血管图对齐，以此计算各细胞在各光学成像图中的z-score值。

## 双光子成像数据处理

我们的双光子成像数据分析同样在MATLAB 2017环境中完成。图像校准与细胞识别的方法继承自我们先前的研究。为了补偿成像过程中可能出现的皮层移动引起的图像模糊，我们将同一视场下的各个会话图像（包括跨日数据）都与一个统一的模板对齐，该模板是第一个会话前1000帧图像的平均值。这样的对齐通常能纠正小于20像素的位移。

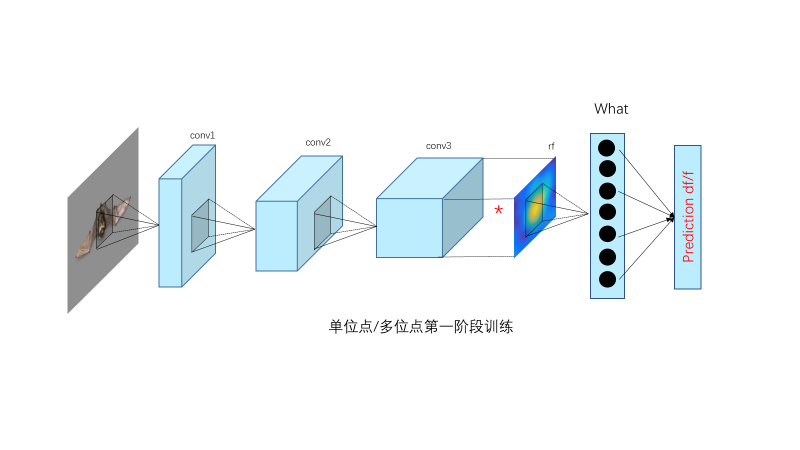
细胞的识别是基于其对视觉刺激的反应。具体操作中，我们首先将每个刺激的各次实验结果取平均，然后以刺激前两帧的平均图像作为基线，以刺激后第二、第三帧的平均图像作为反应图。通过从反应图中减去基线图，我们获得了用于细胞识别的差分图像。细胞是否被识别的标准是，在某个差分图像中至少有30个相连像素点的灰度值超过了整个图像灰度的平均值加上2.5倍的标准差。

一旦确定了细胞的位置，我们便计算该细胞对于各种视觉刺激的反应强度，即荧光变化率（ΔF/F0）。这里的ΔF表示的是刺激反应图中细胞灰度的变化值，F0则是基线图中细胞的灰度均值。如果一细胞对于所有实验刺激的平均反应强度均未超过0.3，那么这一细胞将不被纳入进一步分析中。细胞对各种视觉刺激的平均反应将被用来进行主成分分析（PCA）以及深度卷积神经网络（DCNN）模型的训练与验证。

## 深度学习模型构建

### 神经元DCNN编码模型

我们所采用的神经元编码模型是基于Bashivan等人在2019年提出的原型[6]。此模型结构包括AlexNet的部分预训练卷积层，一个代表感受野（RF）的“where”层，以及一个“what”层[7]，旨在构建从刺激图像到AlexNet特征空间的映射，最终映射到神经元的响应空间。与[6]中的原始模型相比，本研究中的刺激图像在送入模型之前并未进行鱼眼变换处理。模型的拟合过程分为两个阶段。在第一阶段，我们保持AlexNet的预训练卷积层参数不变，仅对“where”层和“what”层进行训练，直至模型达到收敛，这与文献[6]中的做法相似。在此基础上，我们引入了第二阶段的训练，允许AlexNet的卷积核参数参与调整，直到再次达到收敛。这一增加的微调阶段显著提高了编码模型的拟合度，EVE值提升了0.05至0.1。模型训练的损失函数由四个部分构成：第一部分是真实神经响应与预测响应间的均方误差​；第二部分是减少不必要边缘的空间滤波器拉普拉斯正则化项；第三部分是空间滤波器L2正则化项；第四部分是通道滤波器的L2正则化项。（公式待补）



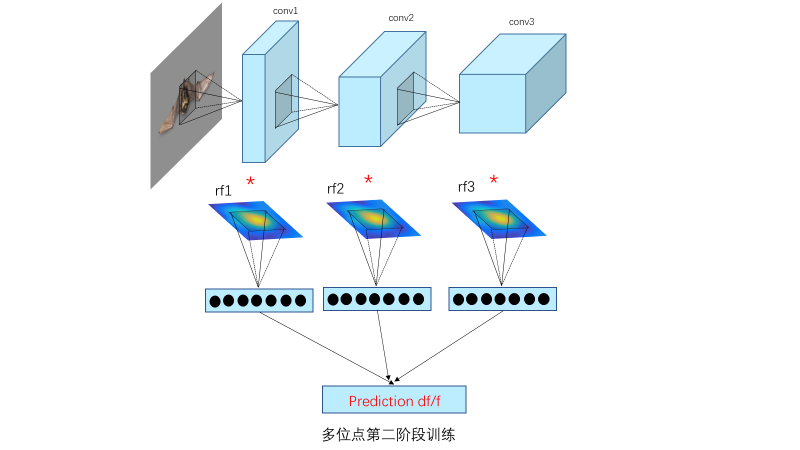


图1 神经元编码模型框架

### AlexNet预训练卷积层的选择

在本研究中，我们对V1和V4神经元的模型拟合进行了深入探讨，这涉及到为每个位点精选最适合的AlexNet卷积层。为了高效地挑选出对应于每个位点的最优卷积层，我们首先采用稀疏随机映射技术[8]来对AlexNet各卷积层的特征进行降维处理，确保降维后的维数符合Johnson-Lindenstrauss引理对样本量的要求。继而，我们利用降维后的特征在岭回归模型中预测神经元的响应，通过双重交叉验证方法为每个神经元选定最佳的正则化参数alpha。最终，我们根据各个位点所有神经元平均拟合精度的评估，挑选出表现最佳的卷积层，以此作为编码模型中AlexNet的特征输出层。在本研究中，V1区域前三个位点的最优AlexNet卷积层分别是第1、第2和第2层，值得注意的是，第三个位点位于V1和V2区域的交界。而V4区域的后五个位点，其最优卷积层均为第三层。这一筛选结果不仅映射了AlexNet与灵长类动物视觉腹侧通路之间的层级对应关系，而且与前人的研究结果[9]保持了一致性。

### 单位点与多位点模型

在本研究中，我们采用了单位点模型和多位点模型两种不同的方法来训练神经网络，以探究神经元在信息处理上的机制。

单位点模型专注于单个位点的神经元反应数据。在这种模型训练方法中，我们仅将同一位点的神经元反应数据输入模型进行训练。在此模型框架下，所有神经元共享相同的卷积层参数，而where和what层的参数则针对每个神经元独立优化。图4和图5展示的结果是基于这种单位点模型得出的。

为了更深入地比较V4区域内不同功能区神经元（通常分布于不同位点）在信息处理上的差异，我们设计并实施了多位点模型。在这个模型的训练过程中，所有神经元（包括V1和V4区的）被同时纳入训练。与单位点模型相同，不同位点对应的最优AlexNet卷积层也被保留。在多位点模型中，所有神经元共享卷积层参数，但where和what层的参数依然是针对每个神经元独立进行优化的。图6展示的结果是基于多位点模型得出的，我们发现多位点模型的总体预测精度略低于单位点模型（详见图Sx）。

总的来说，通过这两种模型的对比分析，我们可以更加深入地理解和揭示不同功能区神经元在信息处理上的独特机制和差异。

### 模型评估与训练

在本研究中，我们从800个自然视觉刺激及其对应的神经元响应中随机选取了720组作为训练集，而将余下的80个自然刺激以及额外的50个针对V4区域的人工刺激及其神经响应作为验证集。在模型的训练过程中，我们特别在验证集上对损失函数中的超参数进行了细致的网格搜索，其中的范围设定在[0.01, 0.1, 1]，和的范围同样设定在[0.01, 0.1, 1]。这一过程旨在寻找到能够实现最优拟合精度的超参数组合。

为了有效地进行模型优化，我们将学习率设定为1e-4，并选用了批处理大小为50的Adam优化器。为了防止模型过拟合，我们采取了早停策略：即如果在连续500次迭代中均方误差（MSE）的下降幅度持续小于5e-5，那么训练将被提前终止。

在量化模型拟合精度方面，我们采用了可解释方差的解释比例（EVE，Explainable Variance Explained）作为指标。EVE计算公式为：



其中，表示真实神经响应值与模型预测值之间的均方误差，是神经元在不同刺激下响应的方差，而则是神经元在同一刺激下响应的方差在所有刺激上的平均值。通过这种方式，我们能够更加精确地评估并解释模型的预测效果。

### 偏好图像生成

为了直观地展示每个神经元所编码的特征，我们受到文献[1]的启发，采用了梯度上升法在图像像素空间对训练好的编码模型进行优化，从而为每个神经元生成其偏好图像。这一过程的核心是对以下损失函数进行优化：



此处，L的第一项旨在最大化目标神经元的预测响应，而第二项则是总变异（total variation）正则项，其作用是为了避免生成图像中出现过多的高频噪声。在这里，我们设置了 = 500。图像像素空间初始状态为随机均匀分布的噪声，学习率设定为0.01。整个优化过程包括200次迭代。

通过这种方法，我们能够生成和提取每个神经元的偏好图像，这些图像反映了神经元对特定视觉特征的响应模式，有助于我们更深入地理解神经元的编码机制。

### 卷积核消融实验

在我们已经训练好的多位点模型中，我们采取了一种特殊的方法来评估第一卷积层各卷积核的重要性。具体来说，我们逐个对第一卷积层的每个卷积核进行“损伤”处理，即将单个卷积核的权重设置为0，然后重新评估编码模型在验证集上的表现。这一过程的关键在于观察当特定卷积核被移除后，模型的可解释方差的解释比例（EVE）会发生怎样的变化。

我们以“损伤”某个卷积核所导致的EVE损失与总EVE的比例作为衡量该卷积核在神经元编码模型中重要性的指标。这种方法允许我们量化每个卷积核对于模型整体性能的贡献，从而更深入地理解模型的工作机制以及各卷积核在编码过程中的作用。通过这种系统的分析，我们能够揭示多位点模型中不同卷积核对神经元响应编码的影响，进而为深入理解神经网络的工作原理提供新的见解。

### 图像相似度评估

为了全面评估图像间的相似度，我们采用了两种衡量方法：CLIP（Contrastive Language–Image Pre-training）模型和SSIM（Structural Similarity Index Measure）。CLIP作为一种基于深度学习的方法，能够从内容和语义的角度分析图像相似性；而SSIM则是一种传统的图像处理技术，用于从结构信息的角度评价图像质量。

#### CLIP

CLIP 模型结合了视觉和语言两个组件。视觉组件通常是一个卷积神经网络（CNN），用于从图像中提取特征。语言组件通常是一个变换器（Transformer）架构，用于处理描述图像的文本。这两个组件共同训练，通过对比学习来匹配图像和相关描述，如图2所示。

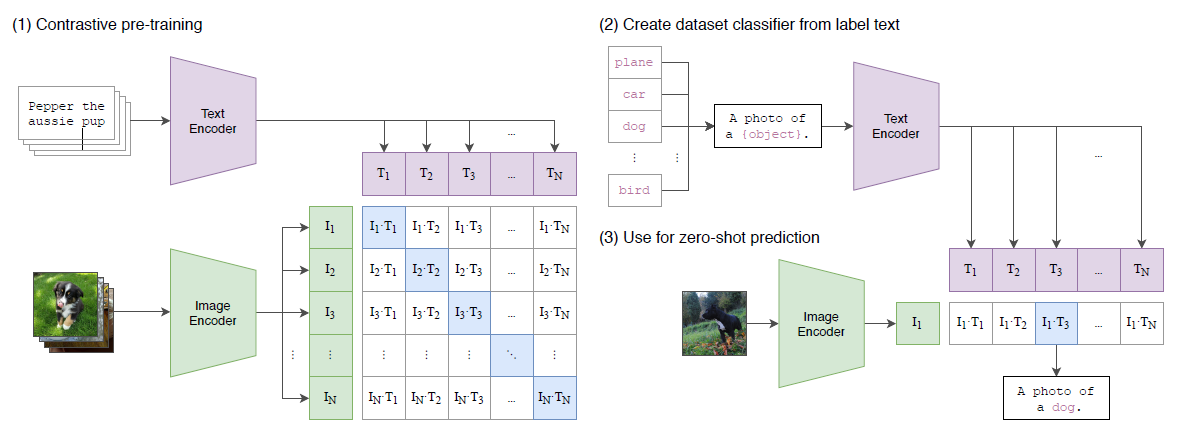


图2 CLIP结构图[10]

在训练过程中，模型学会将图像和文本映射到共同的特征空间中，从而能够通过计算特征向量之间的距离来评估它们的相似度。在这个共同的特征空间中，相似的图像和文本对应的特征向量彼此接近，而不相似的则相隔较远。在我们的研究中，我们使用 CLIP 模型来比较两张图片的相似度。首先，我们将每张图片分别输入到 CLIP 的视觉组件以提取特征向量。然后，我们使用余弦相似度来衡量这些特征向量之间的相似度。余弦相似度是衡量两个非零向量的夹角余弦值，用于度量它们在方向上的相似性。在使用CLIP模型提取的特征向量上应用余弦相似度的计算公式如下：



其中，表示向量和向量的点积，和分别代表向量和向量的欧几里得范数。具体来说，两个向量的点积定义为，其中和是向量和向量的第个分量。此外，向量的欧几里得范数被定义为, , 余弦相似度的值范围从 -1 到 1，其中 1 表示两个向量方向完全相同，0 表示两个向量正交，而 -1 表示两个向量方向完全相反。

#### SSIM

结构相似性指数（SSIM）是一种衡量两幅图像视觉效果相似度的指标，与传统的基于误差敏感度的指标（如均方误差MSE或峰值信噪比PSNR）不同，SSIM考虑了图像的结构信息，更加符合人类视觉系统的特性。SSIM假设人眼会从图像中提取结构信息，而这种信息包含了场景中的对象结构、亮度和对比度等视觉元素。

SSIM的基础是三个比较测量：亮度、对比度和结构。亮度比较函数考虑了图像的平均亮度，对比度比较函数评估了标准差（即图像的对比度），结构比较函数测量了样本的协方差（即图像的结构相似性）。SSIM指数将这三者结合成一个测量值，来评价图像质量。SSIM指数的计算公式如下：



其中，和分别表示图像和的平均亮度，和分别表示图像的方差，表示图像的协方差。常数和的值分别设置为 0.01 和 0.03 的平方，以稳定分母。

# 实验结果

## 改变输入图片可以提升模型精度

### 实验目标

本节的目标是通过改进刺激图像的展示方式来增强模型的预测准确性。在模型训练的过程中，我们观察到每张图像首先通过一个预训练的AlexNet进行特征抽取。模型随后尝试挖掘这些特征与相应神经元反应之间的内在联系，这一步骤涉及到了对神经元感受野的定位，以及对各个通道权重的估计，这个过程我们称之为“What and Where”。

在训练过程中，模型致力于从720张训练图像中解码出神经元的感受野，如图3所示。但神经科学的研究表明，每个神经元的感受野是独特的。我们在实验中已经将图像定位于神经元感受野的中心，因此，仅有位于中心的像素会投射到神经元的感受野上。因此，对于一个300px的图像，我们考虑采取一种称为“crop”的策略，即裁剪出中心的圆形区域，并将圆外的像素设置为统一的灰色背景。

通过这种刺激图像的精简方法，我们能够显著减少解码的复杂性，进而有效地提高模型的预测准确度。为了进一步提升模型性能，我们采用了这种裁剪刺激图像的手段，并期望能够提升模型的精度。

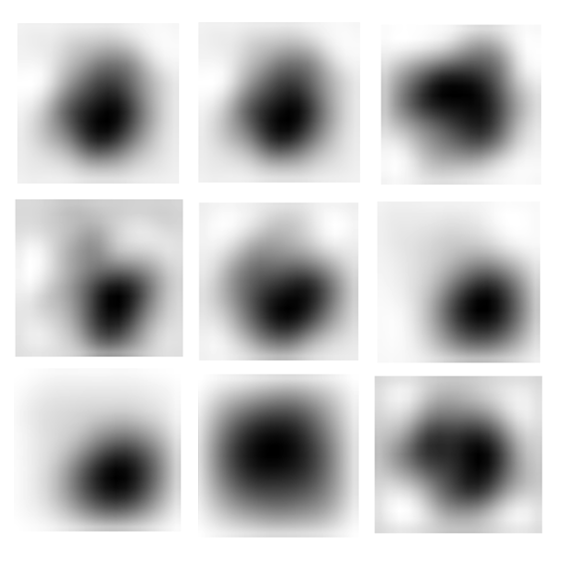


图3 感受野示意图，黑色部分代表模型拟合反应值取决的像素

### 具体操作

为了检验我们的猜想，我们在V1、V2、V4各取一个位点数据进行操作。我们将原有的刺激图像进行crop，再输入到模型中进行训练。由于我们不知道crop多少像素比较合适，因此我们尝试性地选取90，150，200，250px作为尝试大小，并比较不同尺寸的crop与初始图像训练后的PCC(Pearson correlation coefficient)，为了排除模型随机初始值的影响，我们分别训练了多次并取平均值，图片在经过不同程度的crop后的示例图如图3。

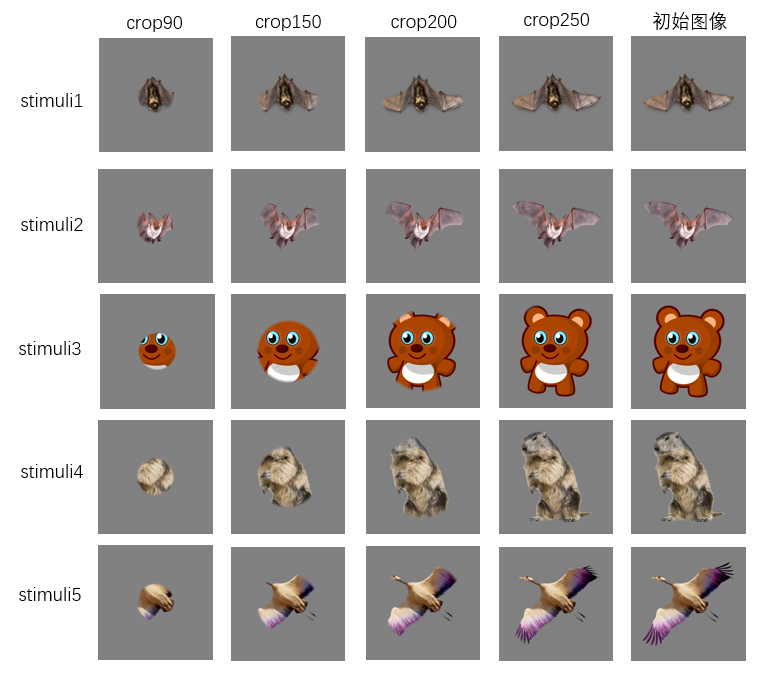


图4 crop示例图，从左到右分别是crop90,crop150,crop200,crop250和初始图片

在我们的实验中，不同尺寸的图像裁剪结果如表1所示。通过对比，我们注意到在特定的视觉皮层区域，如V1和V4，使用200像素的裁剪尺寸能够获得最高的皮尔逊相关系数（PCC），而在V2区域，150像素的裁剪尺寸则显现出最优的相关性。

需要特别指出的是，我们尝试了直接将300x300像素的图像裁剪至其他规模例如90x90像素。然而，考虑到特征提取时使用的AlexNet卷积层要求输入图像格式为299x299x3，直接裁剪的图像必须在输入卷积层之前进行放缩，这会导致图片经过重塑，从而使得神经元接收到的视觉刺激与原图差异加大，不利于我们的研究目标。实际测试也证实，这种直接裁剪的方法会显著降低模型的准确度。

基于上述发现，我们在实验设计中采取了一种更为精细的裁剪策略，以确保图像的有效信息最大化地保留，同时符合神经元实际观察到的刺激，从而提升了我们模型的性能和结果的可靠性。

表1 Crop 训练效果图

|  |  |
| --- | --- |
| 训练类型 | V1位点PCC V2位点PCC V4位点PCC |
| Crop90  Crop150  Crop200  Crop250  初始图像 | 0.8065 0.7470 0.4281  0.8375 **0.7510** 0.4972  **0.8405** 0.7330 **0.5068**  0.8403 0.7332 0.5063  0.8388 0.7335 0.4961 |

### 结论

通过深入比较模型在crop和初始的数据集上的训练结果，我们发现无论是在V1、V2还是V4区域的神经数据上，对刺激图片进行适当的crop均能显著提升模型的拟合精度。这一结果体现在皮尔逊相关系数（PCC）、解释方差分数（EVE）以及决定系数（R^2）的增加上。这些指标的提升不仅验证了我们在实验设计上的正确性，也印证了模型在处理经过优化的感受野时的有效性。

我们进一步观察到，不同脑区位点的最优裁剪尺寸存在差异，这可能与各区域神经元的固有特性相关。例如，V2区域的最佳裁剪尺寸与V1和V4不同，这暗示着不同视觉处理阶段的神经元对视觉信息的编码方式具有区别。在三个不同的位点上，PCC指标普遍提升，这进一步证实了初始图像中确实存在对神经网络模型解码有干扰作用的冗余像素信息。这一发现不仅验证了我们之前的假设，也为未来的研究提供了明确的指导：在设计神经网络模型时，应当细致考虑输入数据的质量和相关性，去除不必要的信息，以提高模型的性能和解码能力。

综上所述，这些实验结果不仅加深了我们对视觉编码机制的理解，也为构建更为精确的神经网络模型提供了实验依据。未来的研究可以在这一基础上探索更多的视觉刺激参数，进一步优化模型表现，以便更好地逼近真实的神经系统功能。

## 模型可重复性验证

### 实验目标

在我们的数据集里，数据8,9是同一批神经元细胞在两组不同的图片刺激集下采集到的结果。如果我们的模型泛化能力足够好以及实验设计合理，那么即使是不同的图片刺激集，训练得到的模型都能很好的得到每个细胞的特性。因此，针对这两组数据我们希望通过比较这两组数据的拟合精度情况，从而指导模型的改进和提升。

### 数据8，9训练结果比较

在对数据集8和9进行比较时，首先需要指出的是，两组数据所使用的刺激图片在分辨率上有所不同。具体来说，数据组8使用的刺激图片分辨率为300×300像素，而数据组9的图片分辨率为450×450像素。除此之外，两者在刺激内容上也存在显著差异：数据组9的刺激包含了背景元素，而数据组8的图片则以灰色背景为特征，正如图4所展示的。

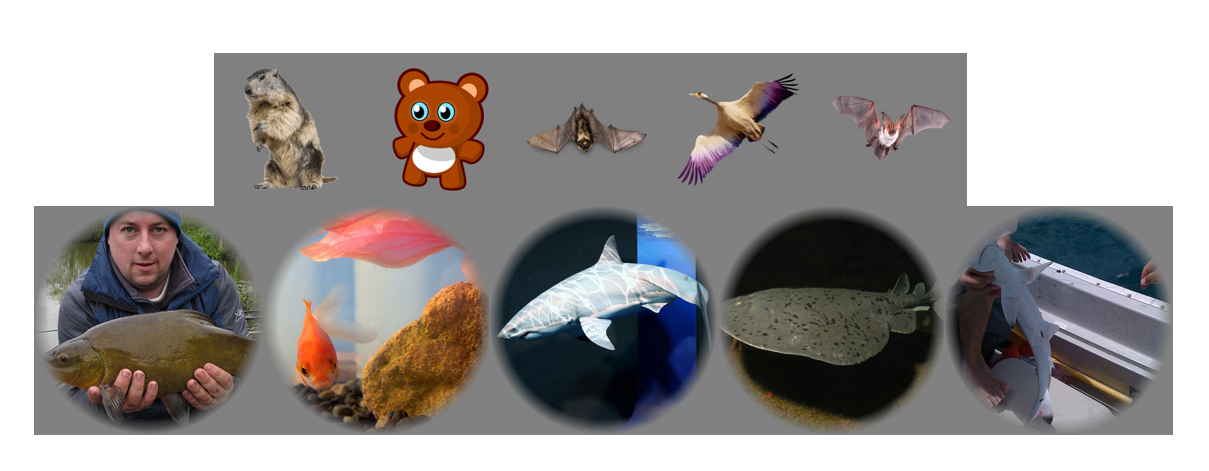


图5 数据8,9刺激图片对比

从统计学的视角进一步分析两类刺激集对113个神经元反应的影响，我们着手研究了它们在不同刺激集下的平均反应值和变异度。如图6所示，结果表明，当刺激图片含有背景时，神经元的反应强度通常较弱，且反应的变异度也较小，这两者的差异在统计上达到了显著水平(p < 0.005)。箱线图进一步证实了当刺激中包含背景元素时，神经元激发的反应强度普遍降低。

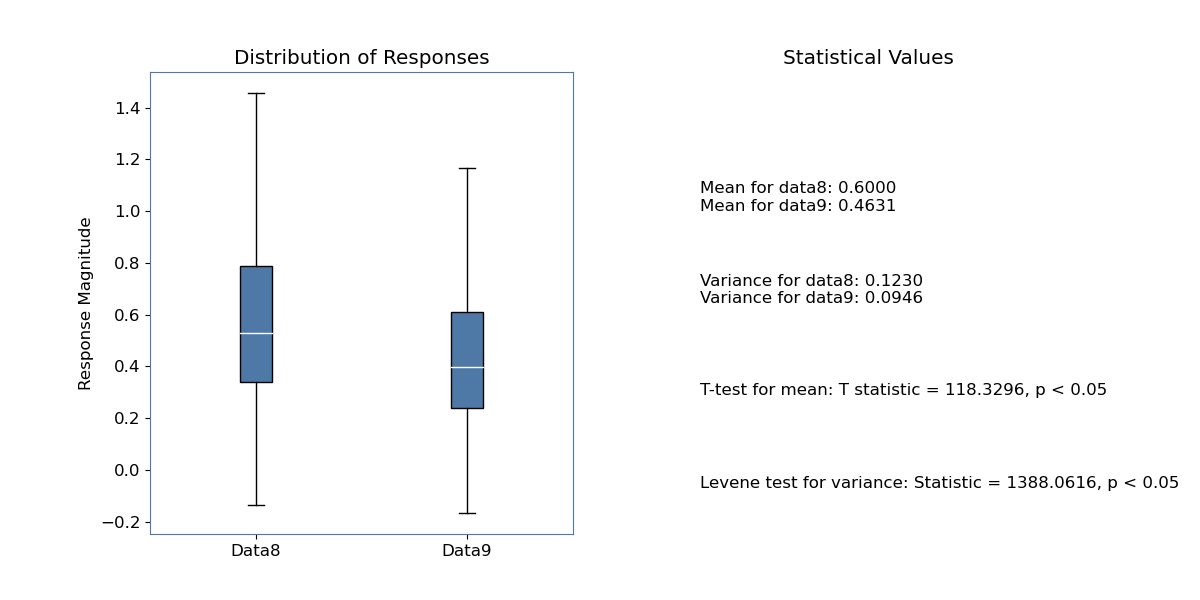
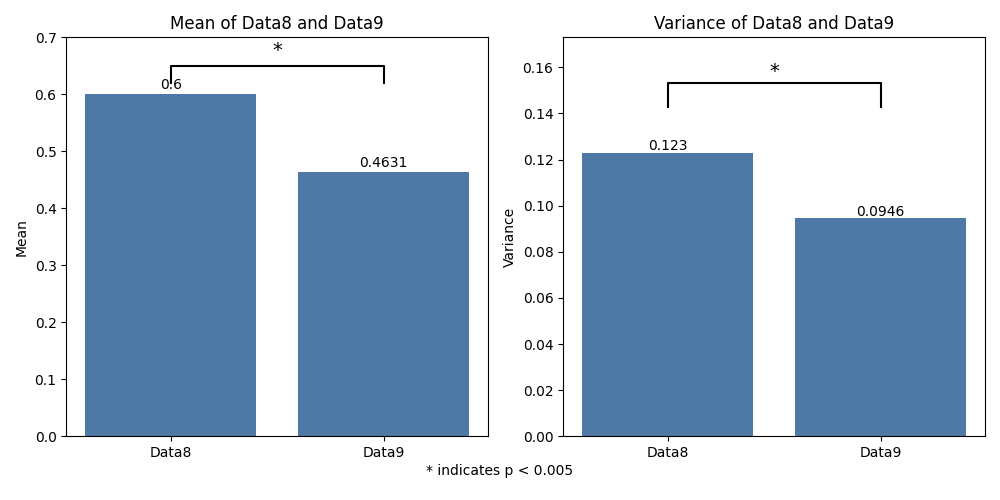


图6 数据8,9神经元反应统计分析

除了上述分析，我们还深入研究了数据组8和9中的试次相关性（trial correlation）。对每个神经元而言，我们计算了其在数据组8的刺激集下所有图片的试次相关性，并取其平均值；在数据组9中，我们采取了相同的方法。结果显示，与数据组8相比，数据组9中神经元的平均试次相关性显著降低。这表明，在多次实验中，神经元对于带有背景的图片刺激表现出较低的一致性反应。特别地，试次相关性低于0.2的神经元在数据组9中占据了不小的比例。这一趋势暗示，神经元对于包含背景元素的图片刺激反应的规律性较弱。为了直观展现这一差异，我们在图7中呈现了数据组8和9的神经元平均试次相关性和PCC的散点图。

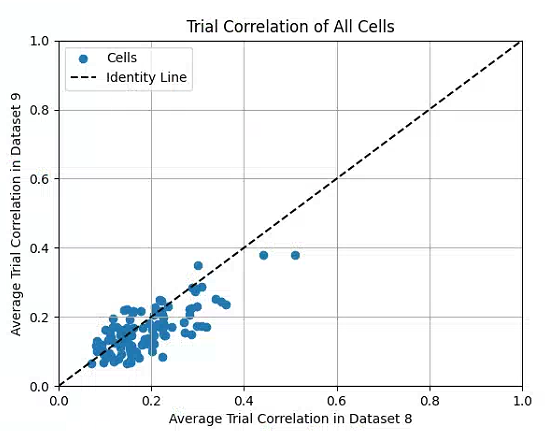
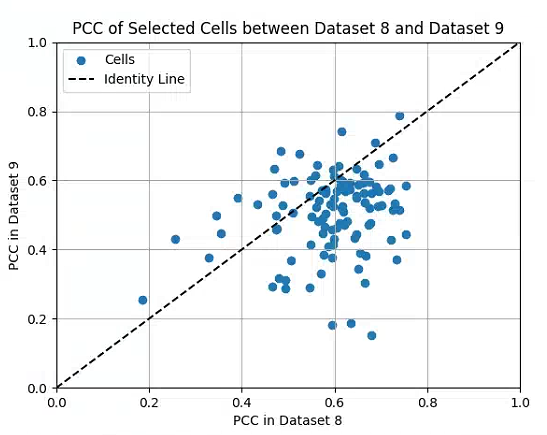
 

图7 数据8,9 trial correlation, PCC对比

### 最优刺激比较

#### 单独训练

在我们模型里面，最大的优势就是能够通过生成偏好图片来可视化结果，从而判断模型的准确性。数据8,9是同一批神经元位点针对不同刺激集的神经响应数据，我们首先尝试分析使用数据8,9各自训练的模型得到的最优图片，并可视化最优图片进行对比，观察模型的泛化能力，结果如下图所示。

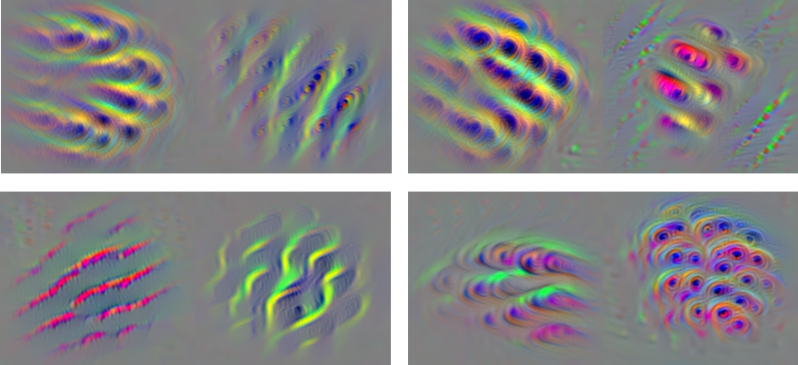


图8 单独训练下的偏好刺激对比，每一张图代表在数据8,9上训练的两个模型针对某个特定神经元生成的偏好图片对比

#### 交叉训练

在理想情况下，一个具有强大泛化能力的模型应当能够在不同的刺激集下保持对特定神经元偏好图像的一致识别。然而，通过直观观察两组刺激集中的偏好图像，我们注意到某些神经元在两组训练集下的偏好刺激存在显著差异，这与我们预期的一致性不符。在对比两组刺激集时，我们发现最显著的差异在于刺激集中是否包含背景信息。据此，我们推测这些背景元素的引入可能是导致模型在两个不同刺激集下产生不同偏好图像的原因。

为了深入探究这一现象，我们设计了一个交叉训练实验。鉴于两组数据集中都包含了800张自然景观刺激图像，我们从数据集8中随机选取400张图像及其对应的神经元反应，同时从数据集9中随机选取400张刺激图像及其神经元反应，组成一个新的、共800张图像的刺激集。剩余的刺激图像同样构成一个规模为800的数据集。我们将在这两个新的数据集上分别训练模型，并生成相应的偏好图像，最后比较两者的相似度，以期揭示背景因素对神经元偏好刺激识别差异的影响。

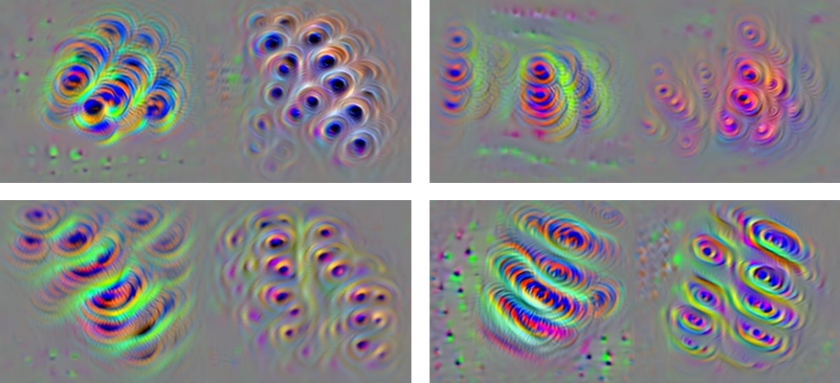


图9 交叉训练下的偏好刺激对比，每一张图代表在交叉数据上训练的两个模型针对某个特定神经元生成的偏好图片对比

#### 不同初始值对比

我们的模型通过迭代的方式生成偏好图片。模型在生成偏好图片的会随机初始一张图片。因此如果模型是有效且可重复的，那么对于同一个神经元的编码模型，不同的随机初始图片在经过迭代后都能得到相似度极高的图片。事实证明模型确实能够在不同初始图片下生成相似度极高的结果。图x是部分神经元在多次随机初始值下得到的例子。

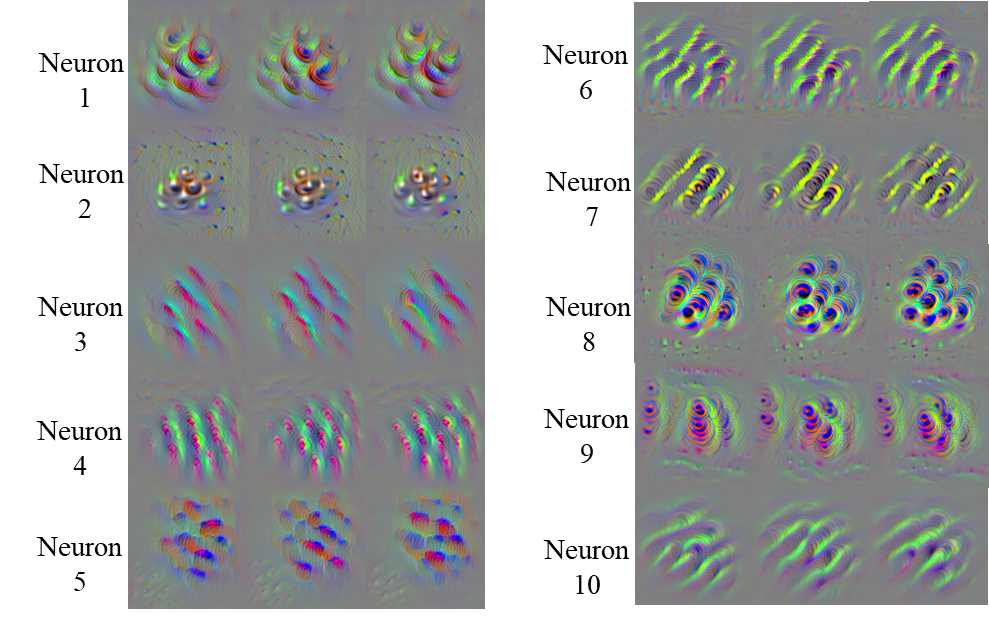


图10 部分神经元随机初始值偏好图片对比

#### 随机偏好图片对比

除了随机初始值对比之外，我们还希望引入随机偏好图片对比，由于随机初始值在肉眼观感上十分相似，能够作为相似度对比的上界，那么随机偏好图片对比应该能够作为相似度对比的下界。这是因为在理想状态下，两个神经元的偏好图片一般而言相似性不会很高，假如我们随机挑选两个不同的神经元偏好图片进行对比，可以作为一个参考下界。换句话说，我们的模型必须严格大于随机偏好图片对比才是有意义的。

因此我们遍历113个神经元，对于每个神经元都随机地在除了本身之外神经元集合中随机挑选一个进行偏好图片对比，对比示例如图11所示。

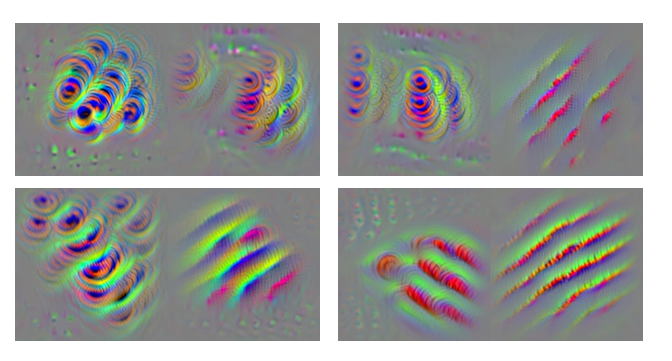


图11 随机对比示意图

#### SSIM对偏好图片相似性的评估

对于评估两张偏好图片的相似性，初步的方法通常是通过肉眼观察来进行。这种方法简单直接，依赖于个人的视觉感知和判断。然而，人对图像的感知受到多种因素的影响，如个人经验、情感、预期和视觉接受能力等，这些主观因素导致评估结果在不同观察者之间出现较大的变异。因此，尽管肉眼评估对于初步筛选有其便利性，但它的主观性和不一致性限制了其在需要精确和一致评估的情况下的应用。

表2 SSIM评估结果表

|  |  |
| --- | --- |
| 训练类型 | Maximum Minimum Average |
| 单独训练  交叉训练  不同初始值  随机对比 | 0.8631 0.3677 0.6271  **0.8822** **0.4088** **0.6447**  0.8007 0.3913 0.5953  0.8534 0.3182 0.5983 |

为了克服这些限制，我们想找到一种定量指标来衡量偏好图像之间的相似度，为我们提供一个客观的、数值化的相似度评分，能够帮助我们进一步评估模型的准确性和可重复性。

首先我们尝试了结构相似度(SSIM)作为指标，部分示例结果如图10所示，针对113个神经元，我们从SSIM最大值（Maximum SSIM）、SSIM最小值（Minimum SSIM）、平均SSIM（Average SSIM）分析，结果如表1所示。



图12 用SSIM计算偏好图片相似性

然而，当我们对随机图片进行SSIM对比的时候，我们发现SSIM对图像相似性的评估并不能很好地符合我们的肉眼观感，尤其是在随机初始值比较的时候分析结果我们发现，对于某一个神经元的偏好刺激，即使给定另一个随机神经元的偏好刺激，SSIM值也不会特别低。如图13所示，这与我们的肉眼观感违背。我们初步猜测是由于偏好图片的背景都是灰色导致两者的SSIM会偏高。

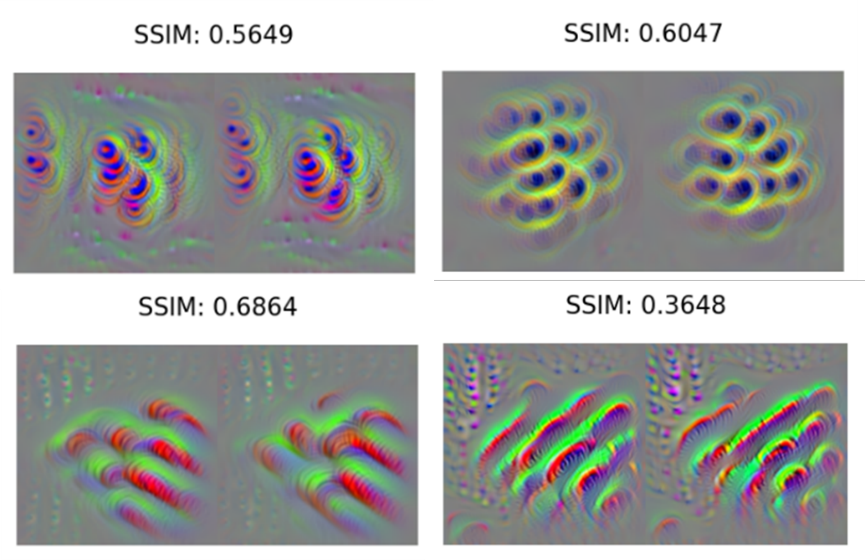


图13 随机初始值的SSIM对比,SSIM的值不符合肉眼观感

为了排除背景因素的干扰，考虑到我们的偏好图片像素值为300\*300px，我们对偏好图片进行掩码操作，只对图片中心半径为125px的圆形区域进行SSIM计算，掩码效果如图14所示。

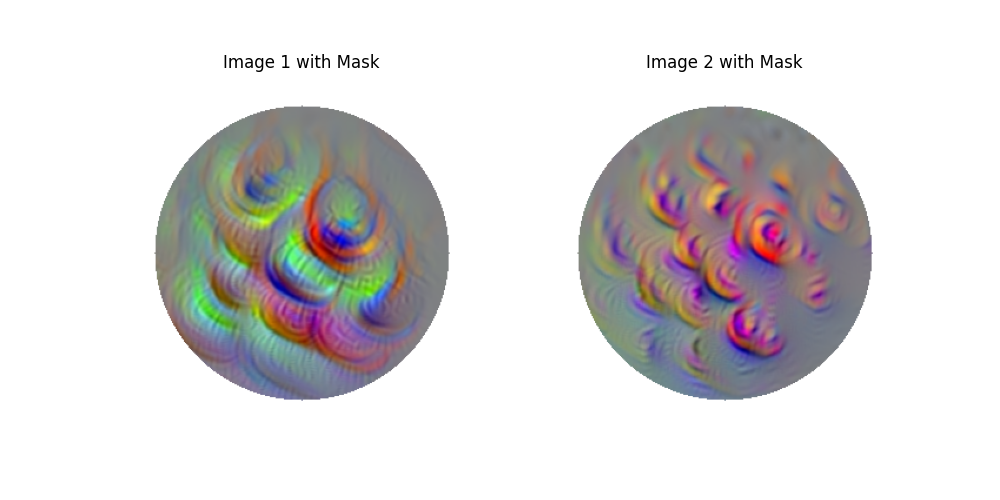


图14 通过掩码后进行比较

然而，经过测试掩码后，虽然观察到随机对比的值是最低的，这与我们的预期相符，然而不同初始值的对比仍然不高，并且结果不具有很好的区分度。带掩码的SSIM结果如表3所示。

回到SSIM的定义，它是基于亮度、对比度、结构进行比较。然而我们发现，生成出来的偏好图片有的在亮度和对比度上极为相似，而结构有所不同。而随机初始化的图片虽然结构相似，但是由于噪声的原因，在对比度和亮度可能差异很大，最终导致SSIM的得分很低，因此我们有必要寻找另一种评估指标针对偏好图片对比的场景。

表3 带掩码的SSIM评估结果表

|  |  |
| --- | --- |
| 训练类型 | Maximum Minimum Average |
| 单独训练  交叉训练  不同初始值  随机对比 | 0.8756 0.5595 0.7106  **0.8930** 0.5798 0.7170  0.8695 0.5619 0.6842  0.8019 **0.5859** 0.6837 |

#### CLIP 对偏好图片的评估

接下来，我们考虑采用CLIP进行图片相似的对比。与SSIM不同，CLIP 评估图片相似性的优势主要体现在其对图片的高层次语义内容的理解能力上。它特别适合于那些需要理解图片内容和上下文的应用，而 SSIM 更适合于需要评估和确保图像结构保真度的任务，CLIP的评估结果如表2所示。

深入分析CLIP的结果，我们发现它与我们的直观预期相吻合：由于不同初始值的图片尽管受到一些误差和噪声的干扰，但它们本质上是由相同模型参数生成的，因此它们之间的相似度应当是最高的。在交叉训练的情况下，每个训练集都含有400张带背景的和400张不带背景的图片，由于内容的相似性，排除了背景因素的影响，两个模型参数应当也产生较高的相似度。而单独训练的情况下，由于两组数据在背景上的差异，导致各自训练出的模型参数更适应于其特定的刺激形式，这会使得偏好图像的相似性降低。最后，随机对比的情况应当展现出最低的相似度，因为不同神经元对刺激的偏好和反应模式有显著差异，导致随机选取的偏好图像间的相似性最弱。CLIP的评估结果与这一系列预期保持一致。

表2 CLIP评估结果表

|  |  |
| --- | --- |
| 训练类型 | Maximum Minimum Average |
| 单独训练  交叉训练  不同初始值  随机对比 | 0.9354 0.5895 0.8549  0.9573 0.7505 0.8791  0.8953 0.6809 0.8333  **0.9908** **0.9247** **0.9718** |

#### 心理物理实验验证

### 结论

# 讨论

## 对后续实验的指导

在我们的实验中，我们注意到当双光子钙信号数据在带有背景的视觉刺激下采集时，神经反应相对于无背景刺激的情况较为减弱，并且试验间的相关性不高。据此，我们得出结论，在未来的研究中，收集自然刺激下的双光子钙信号数据时，应优先考虑使用无背景的刺激以获得更加清晰和一致的神经响应，这对于指导后续的实验采集具有重要的意义。

在我们进行双光子钙信号实验之前，我们会采用内源信号光学成像技术来确定每一组数据的功能性区域。实验发现，我们所使用的深度学习模型生成的偏好图像与这些功能区域的特征高度一致。例如，Color功能区域的偏好图像通常展现出鲜明的颜色特征，而Orientation功能区域则偏好含有明显朝向特征的图像。这意味着深度学习模型不仅为我们的实验策略提供了有效的验证手段，也指导了我们更加精确地理解神经元偏好。

此外，我们通过对图像进行裁剪的实验，能够有效地提高模型的精度，这确认了将图片放置于感受野中心的方法是有效且可行的。这种方法的成功应用进一步强化了我们对双光子钙信号数据采集策略的信心，为未来在类似研究领域内的工作提供了宝贵的实证基础。

## 研究的局限性

本研究所涉及的双光子成像区域相对于V4区域整体而言较小，因此在研究范围和细胞数量上存在一定局限。同时，成像深度也有限，目前只能覆盖V4区域的第二和第三层，无法对功能柱内不同层次的神经元进行详细记录和比较。

由于多个trial跨越不同的实验天，尽管我们对数据进行了校准，动物状态的日间变异和实验带来的某些噪声仍然无法完全消除。加之麻醉条件的特殊性，神经元对自然刺激的反应往往较传统光栅刺激来说更为微弱和不规律，导致trial correlation相对较低。因此，我们收集的神经元数据质量存在不足，这在一定程度上影响了模型训练的预期效果。

在生成神经元偏好图片方面，我们观察到生成的图片在对比度和亮度上具有较高的相似度，这些图片显示出一定的统一模式。虽然我们尝试通过心理物理实验和CLIP指标进行定量分析，但我们注意到，无论是直接训练、交叉训练、不同初始值对比，还是随机对比，CLIP的评估区分度都不高，导致最终评分差异不显著。这可能与偏好图片在背景、亮度和对比度方面的固有相似性有关。因此，一方面我们需要进一步优化评估指标，例如通过排除灰色背景的干扰，提高结果的区分度。另一方面，我们的偏好图片生成方法也有待改进，目前生成的图片内容相对抽象，与现实世界中的目标物体存在差异。鉴于我们使用的是自然图片作为训练刺激，生成更加逼真的图片将更具说服力。在这方面，我们可以借鉴一些研究人员使用的扩散模型，这类模型已被证明能生成更符合自然状态的图像，为我们提供了宝贵的启示。

## 探索新的研究方向

在实验的过程中，尽管我们尝试了多种方法，但并非所有尝试都获得了预期的结果。例如，我们已经能够训练模型生成那些能够最大化特定神经元放电速率的最优刺激。这引出了一个有趣的问题：是否存在某种抑制性刺激，能够使得神经元的放电速率降至最低？尽管我们可以根据神经元所在的功能区域，对其偏好的刺激进行预测，但是这种所谓的“最差刺激”的特征似乎更加难以捉摸和预测。如果我们能成功生成这类刺激，不仅能提高对神经元功能的精确控制，同时在医疗领域的应用也可能具有重要的影响。

此外，我们注意到，尽管我们使用了AlexNet的卷积层提取特征，这种方法与V1和V4神经元实际处理刺激的方式仍有不少差异。通过可视化对比我们在第二阶段训练后的卷积核，如图13中所示，我们展示了数据8卷积核的可视化示例。由于受限于AlexNet神经网络架构，卷积核的数量被固定为96个。在观察这些卷积核的模式时，我们发现存在一定的局限性。比如，有些卷积核并没有表现出颜色和方向的叠加效应，这可能导致部分刺激特征无法被有效捕捉。此外，鉴于这些卷积核是从网络上预先训练好的模型参数中提取的，训练刺激集与我们的实验任务之间存在差异，这些参数并不完全适用于我们的研究，最终可能引入一定的误差。因此，我们考虑是否可以通过调整卷积核使其更贴近V1和V4神经元的特征提取机制。虽然卷积核的调控和生成存在着一定的困难，我们经过一些实验并不能很好的实现这个目标。但是如果我们能够最终实现对这些神经元的完备模拟，那么许多实验可能会被模型所替代，这将在减少人力物力消耗方面带来巨大的优势。

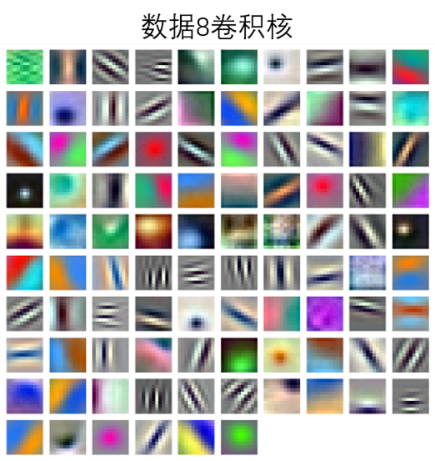


图15 两阶段训练的卷积核可视化结果（以数据8为例）

最后，我们还探索了图像重构的可能性。我们希望通过记录一个神经元群体的放电活动，我们能在训练有素的神经元编码模型中重构出引发该群体反应的图像。这种方法已经在处理fMRI数据的研究中得到应用，我们希望能将其扩展到双光子钙信号数据。这不仅能推进我们对视觉认知的理解，例如确定需要多少个细胞的反应才能重构出与原始刺激足够相似的图像，而且这项技术未来可能用于解码梦境，甚至于进行数据压缩——将数据编码为神经元反应的形式，再通过神经元编码模型解码得到原始数据。

# 参考文献

1. 1. Hubel, D.H. and T.N.J.T.J.o.p. Wiesel, *Receptive fields, binocular interaction and functional architecture in the cat's visual cortex.* 1962. **160**(1): p. 106.
2. 2. Fukushima, K.J.B.c., *Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position.* 1980. **36**(4): p. 193-202.
3. 3. Cadieu, C., et al., *A model of V4 shape selectivity and invariance.* 2007. **98**(3): p. 1733-1750.
4. 4. Yamins, D.L., et al., *Performance-optimized hierarchical models predict neural responses in higher visual cortex.* 2014. **111**(23): p. 8619-8624.
5. 5. Tacchetti, A., L. Isik, and T.J.P.c.b. Poggio, *Invariant recognition drives neural representations of action sequences.* 2017. **13**(12): p. e1005859.
6. 6. Bashivan, P., K. Kar, and J.J. DiCarlo, *Neural population control via deep image synthesis.* Science, 2019. **364**(6439).
7. 7. Klindt, D., et al., *Neural system identification for large populations separating “what” and “where”.* Advances in neural information processing systems, 2017. **30**.
8. 8. Li, P., T.J. Hastie, and K.W. Church. *Very sparse random projections*. in *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*. 2006.
9. 9. Güçlü, U. and M.A.J.J.o.N. van Gerven, *Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream.* Journal of Neuroscience, 2015. **35**(27): p. 10005-10014.
10. 10. Radford, A., et al. *Learning transferable visual models from natural language supervision*. in *International conference on machine learning*. 2021. PMLR.
11. 11. Schrimpf, M., et al., *Brain-Score: Which Artificial Neural Network for Object Recognition is most Brain-Like?* BioRxiv, 2018.
12. [11]

# 附 录

“附录”二字与题名间空一个汉字符位。宋体小四号（英文用Times New Roman体小四号），两端对齐书写，段落首行左缩进2个汉字符。行距20磅，段前段后0磅。

# 攻读学位期间取得的学术成果

**已发表（或正式接受）的学术论文：**（书写格式同参考文献）

**申请或已获得的专利：**（书写格式同参考文献，无专利时此项不必列出）

**参加的研究项目及获奖情况：**

# 致 谢

本篇论文的完成标志着我在北师大三年硕士研究生学习的顺利结束。这段学术之旅的每一步，都凝聚着诸多恩师与良朋的支持与鼓励，对此我心怀感激。

我首先要向我的恩师，吕海东教授，表达最深的敬意和感谢。吕老师不仅是我的学术导师，更是我心灵的引路人。在我刚踏入神经科学领域时的迷茫与徘徊中，是吕老师的悉心指导与不懈鼓励，点亮了我前行的灯塔。吕老师推荐的《视觉信息处理的脑机制》一书，极大地补充了我在视觉科学领域的知识空白，同时吕老师对科研的严谨态度与孜孜不倦的探索精神，更是我学术研究中的不竭动力。

对于实验室中的同仁，我同样怀有无尽的感激之情。每一位师兄师姐的悉心指导、每一次的共同探讨与协作，都让我在科研的道路上不断成长，开阔了视野。在实验的点滴中，我不仅提升了个人的实验技能，更对视觉科学有了更为深刻和全面的理解。

我还要特别感谢我的家人和朋友们。家人们无私的爱与支持，是我在追求学术梦想过程中最坚强的后盾。在每一次的迷惘与困惑中，家是我最温暖的避风港。父母亲的理解与支援，女朋友陈莹的鼓励与陪伴，都是我能够坚持下来的重要力量。同时，我也要感谢我的班主任郁曦老师，您的辛勤工作和教诲，让我们这个集体更加团结和谐，营造了一个有益于学术追求的优良环境。每一位朋友的关心与帮助，都为我的生活和研究增添了无限的色彩。

在此，我对所有给予我帮助与支持的人表示最诚挚的感谢和最崇高的敬意。这篇论文的每一个字节，都凝聚着大家的智慧与情谊。

2024年5月