AI 视觉功能的可行性构建方法——基于多尺度机器感知与三层映射模型的工程实现路线图

YunFei M. Wang
（王·Micheal·雲飛）
Independent Researcher

摘要

本文是“机器感知与认知构建”系列研究的第二篇，旨在系统论述 AI 视觉功能的工程化构建方案。视觉是人类获取外界信息的最主要感官通道，据估计人类约 80%的外界信息通过视觉获得，其信息带宽与空间精度远超其他感官。基于首篇确立的“物理信号—神经编码—主观知觉”三层可计算映射理论框架与“分层基准对照组”实验设计，本文进一步将其实践化为视觉模态的完整工程技术蓝图。我们系统阐述了视觉感知的生物学基础（从视网膜光感受器到视皮层腹侧与背侧通路）与心理物理学核心维度（亮度、颜色、空间频率、对比度、运动、深度等），提出了以“单变量纵深实验”为核心的多层级实验设计方案，并围绕“绝对色觉者—普通人群—视觉功能障碍者”的分层原则设计了全谱系数据采集方案。本文的核心贡献在于：为 AI 视觉功能的实质性研发提供了一份涵盖理论建模、实验室建设（含光学暗室、电磁屏蔽、视觉刺激呈现系统）、设备选型与参数标准（含高刷新率显示器、光谱辐射计、眼动仪、EEG/fMRI 同步记录系统）、人员配置、实验流程、伦理合规与社会协作的全要素、全链条实施方案，为构建可验证、可解释且与人类视觉认知对齐的 AI 视觉感知模型奠定了可操作的科学与工程基础。

关键词

机器视觉；三层映射模型；视觉心理物理学；视觉神经编码；分层基准对照组；单变量纵深实验；仿生机器人

1 引言：视觉在机器认知体系中的核心地位

视觉是人类认知世界的最主要感官通道。从进化的角度看，灵长类动物视觉系统的演化与大脑新皮层的扩展密切相关；从个体发展的角度看，婴幼儿通过视觉建立对物体恒常性、空间关系与因果联系的最初理解。视觉不仅提供了关于世界的最丰富信息流，更是高级认知功能（如物体识别、场景理解、社会认知、审美判断）的基础。

构建 AI 的视觉功能，具有独特的科学价值与技术挑战：

• **信息带宽优势：**视觉系统是人类五感中信息传输带宽最高的通道。人类视网膜拥有约 1.2 亿个感光细胞，经视神经传入大脑的信息速率估算可达约 10Mbps，远超听觉（约 1kbps 数量级）与触觉。这意味着视觉 AI 模型必须处理极高维度的数据流。

• **空间精度要求：**人类视觉系统在中央凹区域具有极高的空间分辨能力（约 1 弧分，相当于在 5 米距离分辨 1.5 毫米的细节），对 AI 的空间采样率与计算精度提出了严苛要求。

• **多层次处理架构：**视觉信息处理从视网膜开始，经外侧膝状体（LGN）中继至初级皮层（V1），再分流至腹侧通路（“是什么”通路，负责物体识别、颜色与精细形状分析）与背侧通路（“在哪里/如何”通路，负责空间定位、运动感知与动作引导）。这一多层次、双通路的架构为 AI 视觉建模提供了丰富的计算灵感，同时也意味着单一的统一模型难以覆盖全部视觉功能。

• **色彩感知的独特性：**色彩并非物理世界的固有属性，而是人类视觉系统对特定波段电磁辐射的主观建构。人类通过三种视锥细胞（L、M、S 型，分别对长、中、短波长敏感）的比值编码来感知颜色，个体间因视锥细胞光色素表达差异而存在巨大的色觉差异。这一“主观建构”特性使得视觉感知的建模必须从物理光学与神经编码两个层面同时入手。

本系列首篇已确立了“机器感知—神经编码—主观知觉”三层映射作为构建任何 AI 感官功能的通用理论框架。第三篇以听觉为范例，系统展示了从理论到工程的完整实施方案。本文将此框架扩展至视觉模态，系统解决从物理光学信号到预测的视觉神经编码、再到预测的主观视觉体验的全部关键工程问题。

本章核心论点：构建 AI 视觉功能的本质，是为 AI 建立一套“物理光学信号（含多尺度空间结构）→预测的人类视觉神经编码→预测的主观视觉体验”的可计算映射系统。该系统必须以人类视觉感知的全谱系数据为基准，以严谨的单变量纵深实验为方法，以分层招募的受试者群体为数据源，最终输出一个可公开验证、可持续迭代的 AI 视觉感知模型。

1.1 框架定位

在“键-值-信息-相”框架中，本文的物理信号层对应键操作值获得的原始反馈，神经编码层对应信息整合为相的中介环节，主观知觉层对应相的可报告投影。传感器是键，被测量的光学现象指向值，三层映射的输出是相。本文不展开该框架的哲学论证，详见文献[1]。

1.2 结构安排

本文按以下序列展开：

- 第 2 章：视觉感知的生物学与心理物理学基础。
- 第 3 章：AI 视觉功能构建的理论框架——三层映射模型（含视频模态扩展）。
- 第 4 章：实验设计——分层基准与单变量纵深实验。
- 第 5 章：实验室工程设计与建设方案。
- 第 6 章：科研团队配置与技能要求。
- 第 7 章：伦理框架与受试者安全保障。
- 第 8 章：社会协作与资源整合方案。
- 第 9 章：项目阶段规划与目标成果。
- 第 10 章：结论。

2 视觉感知的生物学与心理物理学基础

2.1 人类视觉系统的信息处理层级

理解人类视觉的“信号翻译”过程，是构建 AI 视觉模型的第一性原理前提。人类视觉系统可划分为以下功能层级：

眼光学与视网膜层（光信号→神经信号）

角膜与晶状体将外界光线聚焦于视网膜，形成倒置的二维光学图像。视网膜感光细胞（视杆细胞与三种视锥细胞）将光子能量转化为化学信号，进而产生电信号。视网膜内神经网络（水平细胞、双极细胞、无长突细胞、神经节细胞）对视觉信号进行初步的空间、时间与色彩编码。神经节细胞的轴突汇聚成视神经，将编码后的神经脉冲信号传入脑中的外侧膝状体（LGN）。

外侧膝状体（LGN）与初级视皮层（V1）层（初级特征提取）

LGN 作为视觉信息的中继站，对来自双眼的视网膜信号进行初步整合。V1（初级视皮层）具有较小的时空整合感受野和很高的空间分辨能力，偏好编码空间方位和局部视觉信息。V1 神经元对边缘朝向、空间频率、运动方向等基本视觉特征具有选择性响应。

高级视皮层层（客体识别与空间感知）

腹侧通路（V1→V2→V4→下颞叶皮层 IT）：负责物体识别、颜色感知与精细形状分析。V4 区神经元具有较大的感受野，偏好编码颜色与形状特征；IT 区神经元对复杂物

体（如面孔、工具）具有高度选择性响应。

背侧通路（V1→V2→MT/MST→顶叶皮层）：负责空间定位、运动感知与动作引导。MT 区（中颞区）神经元对运动方向和速度高度敏感。

经验与注意调控层

视觉皮层通过自上而下的注意机制（如额叶—顶叶注意网络）对早期视觉加工进行调制。长期视觉训练（如画家、调色师）可改变 V4 等区域的颜色辨别精度与神经响应强度。

这一多层级、双通路的加工架构提示我们：AI 视觉模型若仅停留在“图像像素→语义标签”的直接映射，将永远无法触及视觉认知的本质。我们必须分层建模，从物理光学层逐步逼近知觉层。

2.2 视觉的心理物理学核心维度

心理物理学揭示，人类对视觉刺激的主观感知可分解为若干独立且可量化的维度。这些维度是我们设计“单变量纵深实验”的直接依据。本章心理物理学阈值数据引自视觉心理物理学标准文献[10]；典型值受测量范式、适应水平与受试者筛选标准影响，工程实施时应以原文条件为准。

感知维度	对应物理参数	心理物理学特性	关键阈值/差异阈值参考值
亮度感知	光强度（cd/m²）、照度（lux）	幂函数关系（史蒂文斯定律）；韦伯分数约 0.08—0.15（取决于适应水平）	暗适应绝对阈值约 10^{-6} cd/m²；明适应差异阈限约 2—5%
颜色感知（色相）	主波长（nm）	三维颜色空间（CIELAB、CIECAM02）；个体差异显著	色差 JND 在均匀背景、中央凹注视、中等亮度条件下 $\Delta u'v' \approx 0.001$[10]
颜色感知（饱和度）	色纯度/色度坐标偏移	非均匀感知空间；受亮度与色相调制	饱和度 JND 约 0.01—0.02 u′v′单位
空间频率与对比敏	空间频率（cycle/°）、对比度	对比敏感度函数（CSF）呈带通特性，峰值约 2—5	最小可分辨视角约 1 弧分（20/20 视力对应）

感度	（Michelson 对比度）	cycle/°；峰值位置与高度随适应亮度变化	
运动感知	速度（°/s）、方向（°）	速度感知遵循近似的比例关系；方向分辨力约 1—3°	最小可检测运动速度约 0.01—0.03°/s（中央凹）
深度感知	双眼视差	视差—深度映射近似线性	立体视锐度约 2—10 弧秒（中央凹，取决于测量范式与受试者群体，详见[10]）

这些维度各有独立而清晰的物理控制变量，可单独操作、单独测量，是“单变量纵深实验”的理想条件。每一维度的完整研究都将产出该维度的“物理—感知映射曲线”，构成 AI 视觉模型的“原子模块”。

2.3 视觉个体差异的谱系分布

与所有感官一样，人类视觉能力存在巨大的个体差异，这决定了 AI 视觉模型必须具备“全谱系适配”能力：

绝对色觉者（L1 基准组）

经长期专业训练的色彩工作者（画家、彩色漫画家、调色师、影视/游戏美工、美术指导、园艺师等），其颜色辨别阈限可达普通人的 2—3 倍（即阈值更低、辨别更精细）。部分个体可能具有超常的色觉能力（“色彩感知的超能力者”），源于视锥细胞光色素的遗传变异。此类个体提供了视觉感知“最高分辨率”的基准信号，是校准 AI 视觉模型上限的“黄金标准”。

普通视觉者（L2 主体组）

占人群绝大多数，视觉阈值和分辨力在统计上呈正态分布。受年龄（晶状体黄化、瞳孔缩小、视细胞密度下降）、性别、文化背景（色彩命名差异）等因素影响。提供视觉感知的“平均模型”与“主流分布”数据。

视觉功能边界者（L3 边界组）

包括色盲/色弱（先天性色觉障碍，人群中男性约 8%、女性约 0.5%）、近视/远视/散光（屈光不正）、弱视、斜视、老年性黄斑变性、青光眼等患者。其数据用于定义感知空间的边界与模型的鲁棒性下限，对开发包容性视觉 AI（如无障碍设计、辅助技术）至关重要。特别地，色觉异常者提供了颜色感知空间“降维”或“扭曲”的自然实验数

据，对理解颜色编码的本质具有特殊价值。

2.4 视觉研究实验方法的已有基础

视觉科学已有数百年的实验传统，为我们的工作提供了丰富的方法参考。心理物理学的经典方法如恒定刺激法、调整法、极限法（阶梯法）在视觉阈值测量中广泛应用。近年来，自适应阶梯法（如 QUEST、MLDS）在颜色辨别、对比敏感度等精细测量中展现出更高效率。视觉诱发电位（VEP）与心理物理测量相结合的方法，已成功用于客观测定颜色辨别阈值。这些方法均可直接借鉴并整合至我们的三层映射数据采集框架中。

3 AI 视觉功能构建的理论框架：三层映射模型

本节将首篇确立的三层映射理论框架具体化至视觉模态。

3.1 物理信号层：多尺度视觉数据的系统采集

AI 视觉的物理信号层，目标是建立能够完整描述视觉场景的“视觉数字孪生”表征体系。视觉的“物理层”需包含从宏观到微观的多尺度数据维度:

宏观场景层

高分辨率二维图像（RGB/多光谱，空间分辨率需匹配人类中央凹 acuity，即至少 60 pixels/°）；三维场景几何（通过结构光/激光雷达/双目立体视觉获取的点云或深度图）；场景光照条件（环境光光谱功率分布、光源方向与强度、阴影与高光分布）；物体空间布局与相对位置。

介观特征层

边缘与轮廓（朝向、曲率、连续性）；纹理特征（空间频率、方向性、规则性）；颜色分布（色相直方图、饱和度分布、颜色对比度）；表面反射特性（漫反射/镜面反射比例、BRDF 参数）；运动场（光流、运动方向与速度分布）。

微观光学参数层

各像素/体系的光谱反射率曲线（可见光及扩展波段）；表面微观几何（粗糙度、凹凸纹理的统计参数）；材料介电常数与折射率（对透明/半透明物体）。

这些数据由高保真传感器阵列（多光谱相机、高分辨率 RGB 相机、深度相机、光谱辐射计、偏振相机）协同采集，经同步时钟精确标注后，形成标准化的多格式数据包。

需要限定“类人”原则在此层的适用边界。本文的“类人”指神经编码层的可通约性，而非对前端传感器的仿生约束。类脑视觉工程中已有大量工作同时追求前端仿生（如事件

相机模拟视网膜）与表征对齐；本文的分工是：前端允许使用超越人类生理限制的科学仪器（覆盖完整电磁谱、全频段振动等），后端通过神经编码映射使感知输出与人类可通约。这一分工与已有类脑实践不矛盾，但侧重不同——本文不要求传感器模拟生物结构，只要求其输出能够被映射为人类可解释的神经编码模式。

3.2 神经编码层：从光学到视觉神经活动的预测模型

神经编码层的目标，是构建一个能够将物理光学描述符映射为预测的人类视觉神经群体活动模式的计算模型。

在展开具体架构之前，需要说明本层与已有神经对齐研究的分工。2026 年前后，以 ReAlnets 为代表的工作已直接在非侵入式 EEG 上对齐视觉模型的多层表征，报告了表征相似度的显著提升；预测编码卷积网络的对照研究也表明，递归动力学能增强模型与人类 EEG 响应的对齐，且这一效应在困难刺激上更为明显【11】。这些工作解决的是“模型内部表征能否与大脑视觉皮层对齐”的问题。本文的神经编码层与它们在方法上同构——同样以 EEG/fMRI/MEG 作为训练目标——但目标不同：已有对齐工作通常止于“模型—大脑”的单向相似度，本文要求的是“物理信号→神经编码→主观知觉”的完整可计算链路，神经编码层只是中间环节，其输出必须进一步支撑主观知觉层的预测。这一区分决定了本层不是对已有对齐研究的重复，而是为后续的“体验可通约”目标提供中介表征。

模型架构设计（概念）

- 输入层：上述物理信号层的全部标准化描述符。

- 中间层：模拟视觉通路各关键节点的神经计算，包含：视网膜模型（感光细胞的光子→电信号转换、感受野中心—环绕拮抗结构、ON/OFF 通路分离）；LGN 模型（空间信息的初步中继与调制）；V1 模型（朝向选择性滤波器组、空间频率选择性、简单/复杂细胞响应）；V2/V4 模型（颜色编码、双对抗细胞、形状编码、曲率与轮廓整合）；MT 模型（运动方向与速度选择性）；IT 模型（复杂物体特征的稀疏编码）。

- 输出层：预测的各视觉脑区神经群体活动模式（可表示为特定电极通道的局部场电位功率谱、多单元放电率、fMRI BOLD 信号的空间分布等）。

模型训练策略

使用同步记录的人类受试者视觉神经生理数据（EEG、MEG、fMRI、颅内电生理）作为训练目标。可采用任务驱动的编码模型方法：让模型在图像识别等任务中学习，同时约束其内部表征与大脑视觉皮层的表征对齐。

3.3 主观知觉层：从神经活动到视觉体验的预测模型

主观知觉层模型的目标，是将神经编码层输出的活动模式“翻译”为人类可报告的主观视觉体验。

本层的必要性来自视觉工程中一个长期未被架构性回应的缺口：客观指标与人类感知之间的系统性偏离。PSNR、SSIM、LPIPS 等指标长期被用作图像处理模型的主要优化目标，但在生成模型时代，其与人类主观感知的对应关系已被反复表明不足；大规模用户研究甚至报告 SSIM 在某些数据集上与人类感知呈负相关，ICML 2026 的立场论文据此指出这些指标的“指导力已显著减弱”【12】。本文的主观知觉层不是要用“预测的主观报告”取代所有客观指标，而是为感知对齐提供一个比标量指标更细粒度的监督信号：其必要性恰恰在于，单一标量指标不足以追踪人类感知的结构，而一个从神经编码到主观报告的显式预测层可以。

输出维度

• 亮度（连续数值预测）；颜色（CIELAB/LMS 锥体响应空间坐标、语义颜色命名）；形状（几何描述、物体类别）；空间深度（距离、相对位置）；运动（速度、方向）；材质属性（光滑、粗糙、金属感、透明感）；情绪效价（愉悦度—唤醒度）。

模型训练策略

以受试者主观报告作为训练目标，采用多任务学习框架，同时预测多个感知维度。核心重点为个体差异适配模块，可基于个体预先标定参数（5 分钟颜色辨别测试获取的色觉敏感度曲线），完成模型预测的个性化校准。

3.4 三层模型整合：AI 视觉“感知模拟器”的完整运行逻辑

三层模型形成数据—模型—验证的闭环，完整运行逻辑如下：

物理光学数据（输入）→【物理信号层】光学描述符标准化→【神经编码层模型】预测视觉神经活动模式→【主观知觉层模型】预测主观视觉报告→输出：AI 对“这个视觉场景是什么体验”的完整预测。

验证闭环：将模型预测的神经活动模式与实际同步采集的人类视觉神经数据进行比对，将模型预测的主观报告与实际主观报告比对，误差返回优化各层模型参数。

3.5 视频模态扩展：整体结构定局部细节

视觉感知不限于静帧图像。视频理解是视觉感知在时间维度上的延伸，但在当前主流的 AI 视频理解范式中，普遍采用“逐帧分析+特征拼接”的自下而上路径：提取每一帧的局部特征，再从统计规律中“猜”帧与帧之间的时序结构。这种方法将视频降维为图像序列，导致其对时序因果、节奏、镜头语言、剪辑意图等信息维度的理解十分有限。

近年视频理解研究已从不同路径走向“事件级结构提取”：Event-VStream 将连续视频表示为离散的语义连贯事件序列，仅在事件边界触发语言模型生成；EventLens 研究从视觉证据中恢复事件边界、单事件进展与多事件关系的结构化时序理解；VideoScaffold 以预测引导的分割动态细化事件边界【13】。这些工作的共同判断与本节一致——视频不应被降维为图像序列。本节的三层映射扩展与它们的区别在于：将结构提取与“主观知觉层”的感知对齐目标绑定，而不仅是提升问答或检索性能。

本方案的三层映射框架为视频理解提供了另一种路径：自上而下，结构先行。不是让模型从像素序列中“猜”结构，而是把时序结构作为结构化信息直接提取，逐帧处理只用在真正需要像素级细节的环节。

具体而言，视频理解可以被拆解为以下几个结构化维度的提取：

- 时序因果：对应第二层因果连贯性约束，提取事件 A 导致事件 B 的因果链，而非逐帧猜测“因在前果在后”。

- 节奏：对应投影参数组中的时间维度（剪辑频率、镜头时长分布、运动速度曲线）。

- 镜头语言：对应投影参数组（Z 区间、深度表达机制、投影类型），作为镜头级的结构化参数被提取。

- 剪辑意图：对应坍缩链，将一次剪辑记录为一次选择事件（选了切到 B 镜头，封存了不切的可能性）。

- 音画协同：对应通感机制，听觉编码与视觉编码在结构化层面进行跨模态对齐。

在这一扩展下，视频不再被视为“需要被逐帧消化的像素流”，而是“可以被结构化提取的信息包”。应用层不需要逐帧分析，训练层也不需要逐帧标注——训练的是“结构提取器”（输出场景图、因果链、节奏参数、镜头参数），标注的是场景级结构参数而非帧级文本描述。逐帧处理仅作为第三层局部工具，在需要“微表情是什么”“纹理是什么”时按需触发。这一机制与结构约束先行的视觉生成框架[9]在理解方向上形成对称：生成是“结构→投影→细节”，理解是“整体结构→关系约束→按需细节”。

本模态的感知数据接入长程智能体状态管理与分层生成架构的协作机制，已在文献[1]中系统阐述。

4 实验设计：分层基准与单变量纵深实验

4.1 核心实验原则

- **原则一：单变量纵深实验优先于多变量黑箱采集。**坚决避免在初期实验中同时操纵多个变量（如同时改变颜色、亮度、空间频率、运动速度）。实验设计必须从“最小感知

单元”出发，每次实验仅改变一个物理维度，系统性地探索该维度从阈值以下到阈上至感知极限的完整区间。

- **原则二：分层基准对照组保障模型的全谱系覆盖。**数据采集必须同时包含 L1 绝对色觉者（提供高精度基准）、L2 普通视觉者（提供主体统计分布）、L3 视觉功能边界者（定义模型鲁棒性边界）。

- **原则三：神经信号与主观报告的严格时间同步。**所有视觉刺激呈现、神经生理记录（EEG/MEG/fMRI）、眼动追踪、行为响应，必须由同一主时钟同步触发与标记。

4.2 受试者招募与分组方案

L1 基准组（绝对色觉者/视觉专家，每组 5—10 人）

来源：画家、彩色漫画家、各行业调色师、影视/游戏美工、美术指导、园艺师等与色彩应用高度相关的从业者。招募方式：通过行业协会（如中国美术家协会、中国电影美术学会）、艺术院校定向邀请。入选标准：通过标准化颜色辨别测试（Farnsworth-Munsell 100 Hue 测试，总错误分≤16 为常用筛选阈值，实际分组时应按年龄匹配的常模调整）；视觉 acuity≥20/20（矫正后）；色觉正常（经石原色觉检查板确认）。报酬标准：专家级咨询费，建议 1000—2000 元/小时。

L2 主体组（普通视觉者，每组 100—300 人）

来源：社会公开招募，涵盖不同年龄（18—70 岁）、性别、职业、地域文化背景。入选标准：矫正视力≥20/25；无已知色觉障碍；无严重眼病。报酬标准：100—200 元/小时。

L3 边界组（视觉功能边界者，每组 20—50 人）

分类子组：1）先天性色盲/色弱者（红绿色盲、蓝黄色盲等）；2）近视/远视/散光/弱视/斜视患者；3）老年性视觉衰退者（≥65 岁）；4）黄斑变性/青光眼等眼病患者。来源：与三甲医院眼科合作，由科室医师推荐（匿名化脱敏后接入实验）。报酬标准：考虑医疗陪同与交通，建议 200—300 元/小时。

4.3 单变量纵深实验流程标准化——以“颜色辨别”为例

实验前准备（约 20 分钟）

受试者签署知情同意书，填写视觉史问卷；进行基础视觉筛查（视力、色觉、视野）；在暗室中适应 5—10 分钟（视杆细胞暗适应）；佩戴高密度 EEG 电极帽（64—128 导），进入电磁屏蔽/暗室环境；完成眼动追踪系统校准（精度要求 0.25°—0.5°）。

实验正式阶段（约 60—90 分钟）

每一运行块固定一个颜色中心（如 CIE 标准 illuminant D65 下的特定色度坐标）；呈现一系列颜色刺激，在 CIELAB 或 u′v′色度图中沿特定方向（如红—绿轴、蓝—黄轴）精细步进；步长根据预实验确定，覆盖从 JND 以下到明显可辨的完整区间。每个刺激呈现持续 200—500ms，刺激间隔 1000—1500ms（含随机抖动）。

每次刺激后，受试者需完成三项操作：a）二选一判断："与上一颜色相同/不同？"（测量辨别阈限）；b）颜色命名（开放回答或从预设色名列表中选择）；c）在色度图上标出感知到的颜色位置。同时全程进行 EEG 连续记录，标记每个刺激起始点与受试者响应的精确时间戳。

实验后处理（约 10 分钟）

受试者完成主观疲劳问卷；完成数据备份与离线质检。

4.4 数据同步采集与质量保障

同步精度：视觉刺激呈现系统（显示器/投影仪）、EEG 采集卡、眼动追踪系统、行为响应盒，必须由同一主时钟源同步。所有通道的时间戳偏差需<1ms。

视觉呈现环境：实验需在符合 CIE 标准的光学暗室中进行。环境光照度需严格控制（如<1 lux）；显示器需经光谱辐射计校准，确保色温、亮度、gamma 曲线符合国际标准。

显示器技术参数：推荐使用高刷新率（≥120Hz）、高色准（ΔE<1）、高亮度均匀性（≥95%）的科研级显示器。

EEG 数据质量：每运行块结束后实时检查数据，若某导联阻抗超过阈值或伪迹比例>20%，需在下一运行块前重新调整电极。

5 实验室工程设计与建设方案

5.1 实验室选址与环境要求

选址原则

优先选址于大学或科研院所园区内，便于跨学科合作与受试者招募；需远离强振动源（地铁、大型机械设备）与强电磁干扰源（高压线、无线基站）；预留未来扩容空间，保障长期实验开展。

环境影响评估

开展环境光照监测，确认夜间无杂散光侵入；完成环境振动监测，满足光学实验台隔振标准；开展电磁环境评估，超标区域增设电磁屏蔽设施。

5.2 实验室核心区域功能划分

主视觉实验舱（30—50m²）

完全遮光（光学暗室），墙面与设备采用低反射率哑光黑色涂装；地面铺设减振台或光学隔振平台；配备高精度视觉刺激呈现系统（显示器/投影仪/OLED 阵列）；支持受试者头部固定（下颌托/头托）与自由观看两种实验模式。

视觉神经记录与控制室（15—30m²）

紧邻主实验舱，通过电磁屏蔽观察窗连通；布置 EEG/MEG/fNIRS 数据采集主机、受试者监控系统、视觉刺激控制电脑、眼动追踪主机、数据存储服务器。

设备预备与校准室（10—20m²）

存放备用显示设备、光谱辐射计、色度计、校准色卡、EEG 电极等精密器材；配备恒温恒湿存储柜，保障光学设备稳定性。

受试者接待与筛查室（10—15m²）

用于受试者登记、问卷填写、知情同意签署；开展基础视力、色觉、眼部基础筛查，筛选合格实验对象。

数据处理与计算中心（15—30m²）

部署多 GPU 高性能计算服务器、PB 级大容量存储阵列；配备 UPS 不间断电源与恒温空调系统，保障算力设备稳定运行。

5.3 核心设备清单与关键技术参数

设备类别	设备名称	关键技术参数要求	预算估算（万元）
视觉刺激呈现	科研级高刷新率显示器	≥120Hz 刷新率，1920×1080 以上分辨率，ΔE<1，亮度均匀性≥95%，10-bit 色深	5—15
视觉刺激呈现	高色准投影系统（可选）	用于大视野刺激，色温 6500K±200K，均匀性≥90%	10—30

视觉刺激呈现	光谱辐射计	波长范围 380—780nm，带宽≤5nm，精度±2%	10—20
视觉刺激呈现	色度计/亮度计	测量色度坐标(x,y)、色温、亮度(cd/m²)	3—8
神经生理	高密度脑电图系统	≥128 导，DC 耦合，采样率≥16kHz，共模抑制比≥120dB	80—200
神经生理	功能性近红外光谱系统（可选）	多通道连续波，采样率≥50Hz	100—300
眼动追踪	高精度眼动仪	采样率≥500Hz（推荐 1000—2000Hz），精度 0.15—0.5°，与 EEG 同步	15—40
环境控制	光学暗室与电磁屏蔽	遮光率≥99.99%，电磁屏蔽效能≥80dB	30—80
环境控制	光学隔振平台	固有频率≤5Hz，隔振效率≥90%@10Hz	5—15
数据同步	主时钟与同步发生器	多路 TTL 输出，精度≤10μs	2—5
数据处理	高性能计算服务器	多 GPU（≥4×A100/H100），内存≥512GB	60—150
数据处理	海量存储阵列	容量≥500TB，读写带宽≥5GB/s	15—40
总预算	估测		约 350—900 万元

5.4 安全与应急设施

实验舱与控制室各安装 2 处紧急停止按钮，可快速终止实验运行；严格管控视觉刺激参数，显示器亮度、对比度、频闪频率均控制在安全阈值内，规避癫痫诱发风险；实验区域配齐消防器材、医疗急救用品，建立应急处置流程。

6 科研团队配置与技能要求

岗位	人数	核心职责	关键技能与资质要求
项目首席科学家	1	整体方案设计、项目统筹与技术领导	视觉神经科学、心理物理学、计算视觉领域教授/研究员，具备跨学科项目统筹经验
视觉心理物理学方向 PI	1	实验方案设计、心理物理学数据采集与分析	博士学位，5 年以上视觉心理物理学实验研究经验
计算神经科学方向 PI	1	三层映射神经编码模型构建与迭代优化	博士学位，熟练掌握 EEG/fMRI 神经数据处理与建模方法
色彩科学/光学方向 PI	1	实验室建设、设备校准、色彩光学体系搭建	博士学位，熟悉 CIE 色彩标准、光学测量与设备校准规范
数据科学家/机器学习专家	2—3	模型代码实现、数据训练、算法迭代优化	博士学位，精通 PyTorch/TensorFlow 框架，具备感知建模经验
实验技术员/EEG 技术员	2—4	日常实验执行、设备调试维护、数据初筛质检	本科及以上，经过专业培训，熟练掌握 EEG 设备操作与实验流程
科研助理/行政协调	1—2	受试者招募、伦理对接、项目行政与资源协调	本科及以上，具备项目管理、文书对接与沟通协调能力

同时组建跨学科顾问团队，包含视觉神经生理学顾问、眼科医学顾问、色彩行业顾问、数据安全与伦理法律顾问、机器人学顾问，为项目提供全方位专业支撑。

7 伦理框架与受试者安全保障

7.1 伦理审查与合规框架

所有实验方案实施前，必须通过机构 IRB 或独立伦理委员会正式审批；严格遵循《中

华人民共和国个人信息保护法》及 GDPR 相关规范，全程保障实验合规性。

7.2 受试者知情同意与权益保护

知情同意书需完整载明研究目的、实验流程、刺激类型、实验时长、潜在不适感、随时退出权利、数据使用范围、报酬标准等核心信息，确保受试者完全知情、自愿参与。针对 L3 视觉障碍受试者，采用简洁通俗、图文结合的方式解读实验内容，配套专属辅助服务，保障其合法权益。

7.3 实验过程的安全监控

建立双重视觉安全管控机制，所有视觉刺激参数提前核验，杜绝频闪、强光等风险刺激；实验室通过监控设备实时观察受试者状态，受试者出现眼疲劳、头痛、眩晕等不适时，立即终止实验并做好记录与安抚。

7.4 数据隐私与匿名化处理

所有原始实验数据统一进行匿名化编码处理，剔除个人敏感信息；公开发布的数据集、研究成果均需经过伦理委员会二次审核，严格规范数据使用与传播范围。

7.5 安全阈值外推与被动校准（交叉引用）

视觉实验涉及强光、频闪等刺激，癫痫诱发风险与光毒性损伤属于超阈值区域。本模态适用文献[1]第 4.3 节“外推优先、被动校准”原则：安全阈值内的亮度、对比度、频闪参数严格按 5.4 节控制；超阈值视觉损伤数据只能通过安全区间外推或真实临床事件（如光敏性癫痫发作、职业性光损伤）被动接入。架构层面，频闪频率、亮度上限作为硬约束固化于实验控制系统（频闪参数按光敏性癫痫安全标准设定），操作员无权解除；制度层面，被动临床数据接入须经第三方伦理监督。触觉模态的完整表述详见文献[6]第 7.5 节。

8 社会协作与资源整合方案

8.1 核心协作机构

合作方类型	具体机构示例	合作内容
高等院校	美术学院、心理学院、信息工程学院、光学工程系	L1/L2 受试者招募、基础理论支撑、学术联合研究
医疗机构	三甲医院眼科、视光中心	L3 障碍组受试者引荐、临床眼科学指导、实验安全医疗支撑

行业协会	中国美术家协会、中国电影美术学会、印刷行业协会	L1 色彩专家招募、行业专业技术指导
企业合作	显示设备制造商、色彩管理企业、传感器厂商	精密设备定制、技术迭代、硬件适配优化
政府与社会组织	残联、科技部、社科联	政策扶持、经费申报、无障碍技术推广支撑

8.2 项目社会价值

激活社会人力资源，为艺术、设计、文化领域从业者提供灵活就业与收入渠道；推动包容性 AI 技术发展，为色盲友好界面、视障辅助设备等无障碍产品提供科学支撑；通过科普开放日、成果展览等形式普及视觉科学与 AI 认知知识，提升公众科学素养；培育视觉科学、人工智能、光学工程交叉复合型人才，填补行业人才缺口。

8.3 项目经济价值

模型成果可广泛落地于计算机视觉、AR/VR、智能显示、自动驾驶、数字文创、智能安防等万亿级产业市场，具备极强的产业转化潜力；项目可形成核心专利、行业标准、专属数据集等知识产权，构建技术壁垒；长期可赋能视觉康复、智能显示产业升级，提升全民视觉体验与生活品质，创造持续性社会经济收益。

9 项目阶段规划与目标成果

9.1 阶段划分与里程碑

阶段	**周期**	**核心任务**	**里程碑目标**
第一阶段：基建与筹备期	0—12 个月	实验室建设、设备采购安装、伦理审批、团队组建、L1 专家资源库搭建	实验室验收合格，设备调试完成，获取伦理批件，完成首批预实验数据采集
第二阶段：单维纵深实验期（核心攻坚）	12—48 个月	逐维度完成亮度、颜色、空间频率、对比度、运动、深度等单变量纵深实验，完成全分层受试者数据采集	建成 L1/L2/L3 全谱系视觉数据集，上线第一版视觉神经编码层模型

第三阶段：多维交互实验期	48—72 个月	开展颜色×亮度、空间频率×对比度等双变量、多变量交互实验，完善模型多维适配能力	完成全维度视觉感知模型迭代，通过多场景实验验证
第四阶段：整合与验证期	72—84 个月	整合全部实验数据，优化三层映射模型，完成成果梳理、论文发表与模型开源	发布 AI 视觉感知模拟器 1.0 版本，完成系列学术论文投稿，核心模型开源落地

9.2 最终目标成果

- **理论成果：**构建完整的人类视觉感知“物理—神经—知觉”三层映射全谱系数据库，覆盖 L1/L2/L3 三类人群感知数据。

- **模型成果：**发布开源可用的 AI 视觉感知模拟器，包含完整推理代码、预训练模型参数与迭代优化方案。

- **标准规范成果：**形成 AI 视觉心理物理学实验设计、数据采集、模型评估的行业标准草案及全套标准化 SOP 文档。

- **社会应用成果：**搭建国际视觉 AI 交叉研究节点，产出可落地于视觉康复、智能显示、计算机视觉领域的产业化技术成果。

10 结论

本文系统论述了 AI 视觉功能构建的完整工程实施方案，从理论基础、三层映射模型架构、分层受试者实验设计、单变量纵深实验标准化流程、实验室基建方案、设备参数选型、跨学科团队配置、伦理合规体系、社会资源整合到项目阶段化规划，形成了全链条、可落地、可验证的 AI 视觉工程实施方案。

本文核心工程创新在于，以绝对色觉者标定人类视觉感知精度上限、以普通人群构建主流感知统计模型、以视觉功能障碍人群界定模型鲁棒性边界，实现了 AI 视觉模型对人类视觉认知全谱系的精准覆盖，突破了传统计算机视觉“像素—标签”黑箱建模的局限性。

AI 视觉功能的工程化构建，并非单一算法或数据的突破，而是融合精密光学工程、视觉神经科学、心理物理学、计算科学、伦理合规与社会协作的系统性工程。本文搭建的技术路线与实施体系，打通了基础感知理论到工程落地的壁垒，为构建可解释、可对齐、可迭代的类人视觉感知 AI 奠定了核心基础。

视觉作为人类核心认知通道的建模突破，将为听觉、嗅觉、味觉、触觉等全感官 AI 感知建模提供标准化、可复制的工程范式，最终推动具备完整类人感知能力的 AI 世界认知体系落地成型。

参考文献

[1] YunFei M. Wang 论经典 AI 训练方式认知局限的产生原因与破局思路——迈向基于多尺度机器感知与认知对象数据库的下一代人工智能[PP/OL]. AiraXiv, 2026. eprint: 2611.0009

[2] YunFei M. Wang 数学有效性的哲学解释：基于计算模拟视角的“键-值-信息-相”四层认知模式[预印本]. PhilPapers, 2026. 链接: https://philpapers.org/rec/WANPOT-8

[3] YunFei M. Wang AI 听觉功能的可行性构建方法——基于多尺度机器感知与三层映射模型的工程实现路线图[PP/OL]. AiraXiv, 2026. eprint: 2611.0011

[4] YunFei M. Wang AI 嗅觉功能的可行性构建方法——基于多尺度机器感知与三层映射模型的工程实现路线图[PP/OL]. AiraXiv, 2026. eprint: 2611.0012

[5] YunFei M. Wang AI 味觉功能的可行性构建方法——基于多尺度机器感知与三层映射模型的工程实现路线图[PP/OL]. AiraXiv, 2026. eprint: 2611.0013

[6] YunFei M. Wang AI 触觉功能的可行性构建方法——基于多尺度机器感知与三层映射模型的工程实现路线图[PP/OL]. AiraXiv, 2026. eprint: 2611.0014

[7] YunFei M. Wang 论 AI 通感（机器联觉）功能的可行性构建方法——基于五感感知体系与认知对象数据库的跨模态联想与验证机制[PP/OL]. AiraXiv, 2026. eprint: 2611.0015

[8] YunFei M. Wang 论 AI 认知对象数据库的设计与构建——基于五感感知映射的多模态认知对象数据体系与基础设施[PP/OL]. AiraXiv, 2026. eprint: 2611.0016

[9] YunFei M. Wang 结构约束先行的视觉生成：AI 绘画与视频生成的三层混合逻辑[PP/OL]. AiraXiv, 2026. eprint: 2611.0019

[10] Palmer S E. Vision Science: Photons to Phenomenology[M]. Cambridge: MIT Press, 1999.

[11] ReAlnets: Aligning Visual Neural Networks with Human EEG Representations[J]. Communications Biology, 2026.

[12] Position Paper on Perceptual Metrics in the Era of Generative Models[C]// Proceedings of the 43rd International Conference on Machine Learning (ICML 2026), 2026.

[13] Event-VStream: Event-Driven Video Stream Understanding[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2026 Findings), 2026.

AIGC 使用声明

本文作者提出了论文的核心观点、理论框架、方案设计思路、全部实验与工程逻辑。

在文本撰写过程中，作者使用了 DeepSeek AI 模型辅助进行文本结构的组织、部分章节的起草与修辞润色。作者对 AI 生成的所有内容进行了全面的审阅、批判性修改与最终定稿，并对论文的全部学术内容、观点、数据引用及结论负最终责任。