

从图像到知觉：通过重构图像产生知觉属性

Pablo Hernández-Cámara*, Jesus Malo, Valero Laparra

Image Processing Lab, University of Valencia

* Corresponding author: pablo.hernandez-camara@uv.es

摘要

一些科学家建议人类视觉感知可能源于图像统计学，在早期视觉中形成有效的神经表征。在这项工作中，我们对一种可以涵盖视网膜-初级视觉皮层（V1 皮层）中几个已知事实的仿生结构——感知网络进行了端到端优化，用于与图像重建相关的不同任务：自动编码、去噪、去模糊和稀疏正则化。我们的结果显示，编码器阶段（类似于 V1 的层）始终在图像失真方面表现出与人类感知判断的最高相关性，尽管在初始化或训练中未使用感知信息。这种一致性在中等噪声、模糊和稀疏性方面表现出最优。这些发现表明视觉系统可能被调整以去除这些特定级别的失真和稀疏性，且仿生模型可以在无需人为干预的情况下学习感知度量。

关键词：感知表示；视觉感知；仿生模型；自监督学习；自动编码器

介绍

一方面，根据经典的高效编码假设 (H. B. Barlow et al., 1961; H. Barlow, 2001)，许多研究人员表明，生物视觉系统的某些行为可以从自然图像的统计规律中推导出来。例子包括：基于一个非彩色和两个彩色宽带光谱敏感度的色彩编码 (Buchsbaum & Gottschalk, 1983) 及其相关的非线性 (von der Twer & MacLeod, 2001; Laparra et al., 2012)；非色和空间色频分析器的出现 (Olshausen & Field, 1996)，包括其带宽 (Atick, Li, & Redlich, 1992)、适应性 (Gutmann et al., 2014) 和非线性响应 (Schwartz & Simoncelli, 2001)。另一方面，自编码器也捕捉它们被训练来重建的图像的统计信息。最近的研究表明，去噪和去模糊自编码器可以重现人类的对比敏感度函数 (Li et al., 2022)，并表现出类似人类的色彩错觉 (Gomez-Villa et al., 2020)。同样，压缩自编码器开发出与人类对图像失真的判断相一致的非欧几里得度量 (Hepburn et al., 2021)，而在执行低级和中级视觉任务上训练的网络也会引发与感知一致的失真度量 (Kumar et al., 2022; Hernández-Cámara et al., 2025)。

这些发现引出了一个问题：像 PerceptNet (Hepburn et al., 2020) 这样的生物启发的早期视觉系统架构，能否在没有明确的感知监督的情况下学习感知距离？在这项工作中，我们在图像重建、去噪、去模糊和稀疏正则化等任务上训练 PerceptNet。然后，我们分析这些目标如何影响与人类感知判断的一致性。我们的结果显示，与人类评价的最强相关性出现在 PerceptNet 的 V1 阶段。值得注意的是，这种一致性在中等噪声、模糊和稀疏水平上表现出最佳值。

通过展示自编码器能够学习类似人类的感知特性，我们的研究提供了对视觉的计算和神经生物学机制的见解。此外，它表明生物启发的架构可能启用跨任务通用的感知度量，而无需感知的人类标注数据。

方法

我们的方法基于 PerceptNet，这是一种生物启发模型，旨在模拟直到初级视皮质的视觉通路 (Hepburn et al., 2020)。为了以自监督的方式训练模型，我们实现了一种自动编码器架构，使用 PerceptNet 作为编码器，并附加一个 PerceptNet 的逆版本作为解码器。这个逆模型反映了原始的 PerceptNet，但将池化操作替换为上采样，并将分裂归一化替换为逆分裂归一化，后者执行乘法缩放而不是除法。在训练过程中，编码器和解码器的参数是联合学习的。

我们使用来自 ImageNet 数据集的大约 200,000 张自然图像来训练模型。每个模型都经过调整超参数以优化性能，并训练到收敛。我们使用四种不同的目标来训练模型：

- 图像重建：模型训练以最小化输入图像和重建图像之间的均方误差 (MSE)。
- 去噪：模型经过训练用于重构被高斯噪声污染的图像的干净版本，噪声用其标准差 (σ) 来参数化。
- 去模糊：模型被训练为重建模糊图像的清晰版本（通过与高斯核卷积），由卷积核的标准差 (σ) 参数化。
- 稀疏性：模型在重建图像的同时，通过在编码器激活的平均绝对值上添加 L_1 惩罚来鼓励稀疏表示。稀疏性通过使用超参数 (λ) 来缩放 L_1 惩罚来进行参数化。

为了评估感知对齐度，我们计算模型每一层激活变化与人类平均意见评分 (MOS) 之间的相关性，评分来自 TID2013，这是一个标准的图像质量评估 (IQA) 数据库 (Ponomarenko et al., 2015)。这使我们能够分析不同目标如何影响在不同层出现的人类对齐的感知表示。

结果

我们首先研究在更简单的图像重建目标下，模型各层与人类 MOS 的相关性是如何变化的。图 2 显示了在编码器的最终层中相关性明显达到峰值，与 PerceptNet 的 V1 阶段相符。

然后，我们通过关注编码器输出分析不同训练目标的效果。图 1 左侧显示了相关性如何随着噪声级别的增加而增加，在大约 $\sigma = 0.1$ 处达到最大值，然后开始下降。这表明适度的噪声水平会改善感知对齐，但过多的噪声会降低这种对齐。

图 1 中心显示，使用 $\sigma \leq 1$ 的轻微模糊会增加相关性，但较强的模糊会降低相关性。这表明模型从学习逆转轻微退化中获益，但严重的模糊会削弱其与人类感知对齐的能力。

最后，图 1 右边显示，尽管稀疏性对比之前的目标影响较小，但适度的稀疏性可以改善相关性，而更高的稀疏度 ($\lambda > 0.1$) 则会降低相关性。这表明，在稀疏性能够提升表示的前提下，它存在一个平衡点，过度强调稀疏性反而会降低性能。

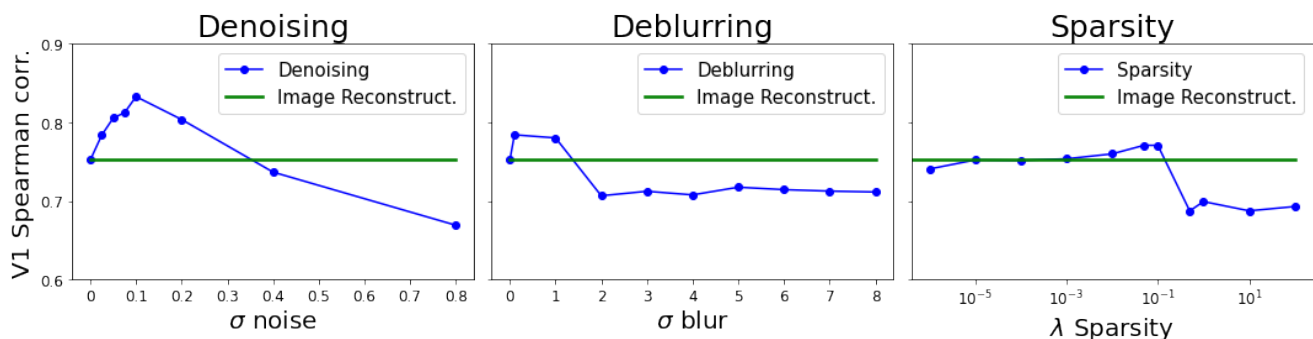


Figure 1: 在编码器末端（类似 V1 层）的 TID2013 斯皮尔曼相关性，作为去噪（左）、去模糊（中）和稀疏性（右）不同训练参数的函数。

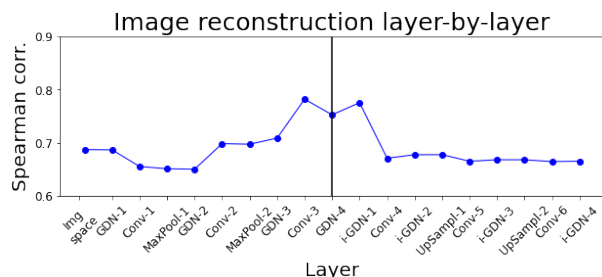


Figure 2: 在训练模型重建自然图像时，TID2013 的斯皮尔曼相关逐层分析。

这些发现表明，知觉来源于高效的编码策略，在这种策略中，大脑平衡了信息保留与噪声抑制。非单调效应表明，知觉通过中等程度的正则化而不是极端约束得以优化。

结论

我们的研究表明，仅通过自监督学习而无需感知监督，生物启发的模型可以发展出与人类视觉一致的感知表示。具体来说，我们展示了当 PerceptNet 被作为自编码器并配以适当的正则化进行训练时，它表现出的感知属性与人类判断有很强的相关性。

一个关键发现是，与人类感知的一致性最高的情况始终出现在编码器阶段，这对应于 V1 的处理。这表明，当大脑的早期视觉表征被优化以用于重建时，它们可能自然地反映了环境的统计特性。有趣的是，适度的噪声、模糊和稀疏性增强了这种一致性，而过度的正则化则降低了它。这些结果支持了视觉系统可能被调整到去除特定水平的失真和稀疏性，以及生物启发模型可以在没有人为监督的情况下学习感知度量。

我们的研究结果补充并扩展了最近的工作，这些工作表明，自监督模型——例如复制人类对比敏感函数的去噪去模糊自编码器 (Li et al., 2022)，或者学习与人类一致的非欧几里得度量的压缩自编码器 (Hepburn et al., 2021)——可以在没有显式监督的情况下捕获感知属性。此外，我们对任务依赖的对齐模式的分析与研究相一致，这些研究表明，为低级和中级视觉任务训练的网络可以产生类似人类的失真度量 (Hernández-Cámara et al., 2025)。

通过展示像 PerceptNet 这样的生物学基础架构能够实现相似的对齐，我们的工作提供了关于可能支撑生物视觉的计算原则的进一步洞察。这也表明，仿生的感知度量可以跨任务和数据集进行泛化，提供强健、可解释

的感知模型，而无需依赖人工注释。

致谢

本工作部分得到了 MCIN/AEI/FEDER/UE 支持，资助项目编号 PID2020-118071GB-I00 和 PID2023-152133NB-I00，部分得到了西班牙 MIU 资助，资助项目编号 FPU21/02256，并部分得到了瓦伦西亚自治区项目 GV/2021/074, CIPROM/2021/056 的资助，以及 BBVA 科学基金项目：数学、统计学、计算机科学和人工智能 (VIS4NN) 资助。一些计算资源由阿尔特米萨提供，该项目通过物理学研究所 Corpuscular, IFIC (CSIC-UV) 由欧盟欧洲区域发展基金资助。

References

- Atick, J. J., Li, Z., & Redlich, A. N. (1992). Understanding retinal color coding from first principles. *Neural computation*, 4(4), 559–572.
- Barlow, H. (2001). Redundancy reduction revisited. *Network: computation in neural systems*, 12(3), 241.
- Barlow, H. B., et al. (1961). Possible principles underlying the transformation of sensory messages. *Sensory communication*, 1(01), 217–233.
- Buchsbaum, G., & Gottschalk, A. (1983). Trichromacy, opponent colours coding and optimum colour information transmission in the retina. *Proceedings of the Royal society of London. Series B. Biological sciences*, 220(1218), 89–113.
- Gomez-Villa, A., et al. (2020). Color illusions also deceive cnns for low-level vision tasks: Analysis and implications. *Vision Research*, 176, 156–174.
- Gutmann, M. U., et al. (2014). Spatio-chromatic adaptation via higher-order canonical correlation analysis of natural images. *PloS one*, 9(2), e86481.
- Hepburn, A., et al. (2020). Perceptnet: A human visual system inspired neural network for estimating perceptual distance. In *2020 IEEE International Conference on Image Processing (ICIP)* (pp. 121–125).
- Hepburn, A., et al. (2021). On the relation between statistical learning and perceptual distances. *ICLR 2022*.

- Hernández-Cámara, P., et al. (2025). Dissecting the effectiveness of deep features as metric of perceptual image quality. *Neural Networks*, 185, 107189.
- Kumar, M., et al. (2022). Do better imagenet classifiers assess perceptual similarity better? *arXiv preprint arXiv:2203.04946*.
- Laparra, V., et al. (2012). Nonlinearities and adaptation of color vision from sequential principal curves analysis. *Neural Computation*, 24(10), 2751–2788.
- Li, Q., et al. (2022). Contrast sensitivity functions in autoencoders. *Journal of Vision*, 22(6), 8–8.
- Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381(6583), 607–609.
- Ponomarenko, N., et al. (2015). Image database tid2013: Peculiarities, results and perspectives. *Signal processing: Image communication*, 30, 57–77.
- Schwartz, O., & Simoncelli, E. P. (2001). Natural signal statistics and sensory gain control. *Nature neuroscience*, 4(8), 819–825.
- von der Twer, T., & MacLeod, D. I. (2001). Optimal nonlinear codes for the perception of natural colours. *Network: Computation in Neural Systems*, 12(3), 395.