

---

# 科学家发现深度神经网络对幻觉轮廓“视而不见”

作者：writer 来源：科学网

本文原地址：<https://www.iikx.com/news/progress/22279.html>

*本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！*

科学家发现深度神经网络对幻觉轮廓“视而不见”。

近日，中科院自动化所研究员曾毅团队研究发现，从经典的到最先进的深度神经网络都难以像人一样具有较好的幻觉轮廓识别能力。相关研究成果发表于细胞出版社旗下期刊《模式》。

神经网络和深度学习模型在过去十年中看似取得巨大成功，在许多给定的视觉任务中在指定方面超过了人类表现。然而，神经网络的性能仍然会随着各种图像扭曲和损坏而降低。一个非常极端的例子是对抗攻击，通过在图片上施加人眼难以察觉的微扰，能够使神经网络模型彻底失效。而人类的视觉系统在这些问题上具有高度鲁棒性，说明深度学习与生物视觉系统相比仍然存在根本性缺陷。

为此，曾毅团队提出了一种名为交错光栅扭曲的图像干扰方法，作为量化神经网络模型幻觉轮廓感知能力的工具。结果表明，大多数预训练模型的表现接近随机。另外可以观测到当交错光栅之间的距离较小时，存在一些模型的结果与其他模型的分布有较显著的差别。他们最终发现，使用深度增强技术训练的模型相比其他模型能够显著增强模型对交错光栅扭曲数据集的识别。

该研究还招募了24名人类受试者，以评估不同的参数设置下，人类的幻觉轮廓感知能力以及其对数字和图像识别的影响。研究发现，即使是当前最先进的深度学习算法在交错光栅效应的识别上也与人类水平相距甚远。

论文第一作者、中科院自动化所工程师范津宇认为，该研究结合了认知科学和人工智能，提出了将传统机器视觉数据集转换成认知科学中的交错光栅幻觉图像，并首次对大量的公开预训练神经网络模型的幻觉轮廓感知能力的量化测量，从神经元动力学角度和行为学角度两个检验深度学习和神经网络模型对幻觉轮廓的感知。

这项研究从认知科学的角度检验和部分重新审视了当前看似成功的人工神经网络模型，并且证明人工神经网络模型与生物视觉处理过程仍然存在着很大差距，大脑运作的机理和智能的本质将继续启发人工智能，特别是神经网络的研究。曾毅说。

在他看来，要想从本质上取得突破，人工智能需要借鉴和受自然演化、脑与心智的启发，建立智能的理论体系，这样的人工智能才会有长远的未来。(来源：中国科学报 田瑞颖)

相关论文信息：<https://doi.org/10.1016/j.patter.2023.100695>

---

作者：曾毅等 来源：《模式》

更多 科学进展 请访问 <https://www.iikx.com/news/progress/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](http://www.iikx.com)转发