



## OPEN ACCESS

EDITED AND REVIEWED BY  
Benjamin Thompson,  
University of Waterloo, Canada

\*CORRESPONDENCE  
Teng Li  
✉ liteng@ahu.edu.cn

RECEIVED 28 November 2023  
ACCEPTED 04 December 2023  
PUBLISHED 19 December 2023

CITATION  
Li T, Zhang Q, Nian F and Shan C (2023)  
Editorial: Neuroscience-driven visual  
representation. *Front. Neurosci.* 17:1345688.  
doi: 10.3389/fnins.2023.1345688

COPYRIGHT  
© 2023 Li, Zhang, Nian and Shan. This is an  
open-access article distributed under the terms  
of the [Creative Commons Attribution License  
\(CC BY\)](#). The use, distribution or reproduction  
in other forums is permitted, provided the  
original author(s) and the copyright owner(s)  
are credited and that the original publication in  
this journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted which  
does not comply with these terms.

# Editorial: Neuroscience-driven visual representation

Teng Li<sup>1,2\*</sup>, Qieshi Zhang<sup>3</sup>, Fudong Nian<sup>4</sup> and Caifeng Shan<sup>5,6</sup>

<sup>1</sup>Information Materials and Intelligent Sensing Laboratory of Anhui Province, Anhui University, Hefei, China, <sup>2</sup>Hefei Comprehensive National Science Center, Institute of Artificial Intelligence, Hefei, China, <sup>3</sup>Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences (CAS), Shenzhen, Guangdong, China, <sup>4</sup>School of Advanced Manufacturing Engineering, Hefei University, Hefei, Anhui, China, <sup>5</sup>College of Electrical Engineering and Automation, Shandong University of Science and Technology, Qingdao, China, <sup>6</sup>School of Intelligence Science and Technology, Nanjing University, Nanjing, China

## KEYWORDS

visual representation, computer vision, neuroscience, learning, deep neural networks

## Editorial on the Research Topic [Neuroscience-driven visual representation](#)

Visual representation learning seeks to mimic the human visual system using deep neural networks, enabling machines to interpret digital images and video for diverse applications from manufacturing to energy. However, major gaps remain compared to biological vision, and most representation learning methods do not sufficiently incorporate neuroscientific and psychological principles. Key open questions persist around designing optimized architectures to extract meaningful representations from complex 2D or 3D scenes containing numerous heterogeneous, unlabeled examples.

While deep learning has achieved state-of-the-art results across various vision tasks like classification, detection and segmentation, core challenges in representation learning need to be tackled to reach human-level visual understanding. For instance, handling unlabeled, unstructured data and generalizing learned patterns to novel datasets continue to pose difficulties. Furthermore, lack of model interpretability is an issue that integration of biological approaches could help address.

This research area aims to advance visual representation learning through synergistic fusion of deep neural networks with psychological and neuroscientific concepts. By providing a platform to exchange cutting-edge techniques spanning both data-driven and theory-driven disciplines, impactful progress can be made toward biomimetic visual systems. Realizing more efficient, generalizable, and explainable visual learning has the potential to profoundly transform capabilities in scientific imaging, manufacturing, transport and healthcare.

Enhanced analysis of facial imagery for health assessment. Building on advanced computer vision techniques, [Li et al.](#) present a facial analysis methodology using convolutional neural networks (CNNs) to detect depression. They introduce innovations including multi-head attention modules and region-specific tuning to improve CNN sensitivity in analyzing different facial areas tied to depression. With further research, such AI-based systems could assist in mental health evaluation and screening.

Multi-constraint modeling for 3D shape reconstruction. Reconstructing 3D structure from 2D image sequences is an important but challenging computer vision task. [Chen X. et al.](#) put forth a multi-constraint estimation algorithm that first extracts shape bases via sparse coding, then estimates 3D geometry through a penalized least-squares model incorporating orthogonal and similarity constraints. Experiments demonstrated higher accuracy compared to existing methods, showing the value of fusing multiple constraints.

Automated defect recognition for semiconductor quality control. As discussed by [Chen Y. et al.](#), precise identification of surface defects in semiconductor wafers is critical for controlling manufacturing quality. They develop a multi-scale visual perception network architecture for automated wafer defect pattern recognition. By effectively integrating fine-grained texture cues across resolutions, their approach achieved state-of-the-art accuracy on a real-world industry dataset, demonstrating feasibility for quality inspection.

Self-supervised representation learning from multimodal data. For human action recognition, [Yang et al.](#) present a novel framework applying contrastive self-supervised learning on paired unlabeled data (skeleton sequences and inertial sensor signals). Without requiring negative samples, they show superior cross-dataset retrieval and zero-shot transfer performance compared to previous multimodal methods. This highlights the promise of self-supervised techniques to improve model generalization.

Elucidating audiovisual processing in the brain. Understanding the complex neural mechanisms underlying sensory integration remains a key challenge in neuroscience. [Jiang et al.](#) combine functional MRI and EEG to construct brain networks involved in audiovisual processing. Through their novel dynamic analysis approach, they revealed early visual-auditory integration occurring prior to attentional effects. These insights shed light on the nature of inter-sensory interactions within the brain.

AI for detecting overloaded trucks to improve road safety. Excessively overloaded trucks pose critical challenges regarding road damage and traffic safety. [Sun et al.](#) develop an AI system to detect truck overloading by recognizing truck models from images and matching against weight data. Achieving 85–100% accuracy on small real-world datasets shows feasibility for automated enforcement on highways to improve infrastructure maintenance and prevent hazardous accidents.

More human-like image captioning via reinforced decoding. Generating textual descriptions for images, known as image captioning, requires modeling both visual concepts and language semantics. [Bai et al.](#) introduce techniques including guided decoding connections, DenseNets, and reinforcement learning to enhance contextual modeling and feature extraction. Superior results across standard captioning metrics represent tangible progress toward human-level visual understanding.

Targeted smoke reduction to maintain surgical visualization. As discussed by [Wang et al.](#), smoke generated during endoscopic procedures can severely obscure surgical sight. They create an enhanced classifier to detect smoke-filled frames prior to selective image enhancement, maximizing efficiency. Achieving high accuracy and speed shows promise for integrated, real-time de-smoking systems to improve situational awareness.

Intelligent product recognition to enable smart vending. As [Xu et al.](#) explored, computer vision powered by deep learning can enable emerging autonomous retail models like smart vending machines to accurately recognize products for automatic checkout and inventory status tracking, reducing overhead costs. Their results demonstrate the feasibility of AI to deliver advanced functionality without constant human intervention.

Multi-Scale adaptive learning for robust driving scene parsing. [Liu et al.](#) address core challenges in semantic segmentation for autonomous vehicle perception including variations in scale, occlusions and diverse appearances. Their multi-scale adaptive network dynamically selects the most relevant features across levels to accurately parse complex driving environments. State-of-the-art performance on automotive datasets confirms robustness, advancing safety for self-driving systems.

Group-based sparse modeling for image restoration. Recovering high-quality images from incomplete or corrupted inputs remains an active computer vision research area. [Ning et al.](#) propose a multi-scale group sparse residual constraint model exploiting patch correlations to effectively eliminate noise and fill in missing regions. Experiments show marked improvements in restoration fidelity compared to existing methods, enabled by joint image priors.

In conclusion, recent advances in visual representation learning could unlock transformative capabilities in transportation, manufacturing, healthcare, and scientific imaging. While progress has been made in tackling real-world vision tasks, continued research into dynamic models, multimodal fusion, and incorporating domain-specific constraints will be instrumental in achieving human-like scene understanding.

## Author contributions

TL: Writing—original draft, Writing—review & editing. QZ: Writing—original draft. FN: Data curation, Writing—original draft. CS: Writing—review & editing.

## Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work is supported by the University Synergy Innovation Program of Anhui Province (No. GXXT-2022-037).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.