frontiers | Frontiers in Computational Neuroscience

TYPE Mini Review
PUBLISHED 06 July 2023
DOI 10.3389/fncom.2023.1153572

OPEN ACCESS

EDITED BY
Si Wu,
Peking University, China

REVIEWED BY
Benoit R. Cottereau,
UMR5549 Centre de Recherche Cerveau et Cognition (CerCo), France

*CORRESPONDENCE
Marco Tamietto
✉ marco.tamietto@unito.it;
✉ m.tamietto@tilburguniversity.edu
Alessia Celeghin
✉ alessia.celeghin@unito.it

†These authors have contributed equally to this work

RECEIVED 29 January 2023
ACCEPTED 19 June 2023
PUBLISHED 06 July 2023

CITATION
Celeghin A, Borriero A, Orsenigo D, Diano M, Méndez Guerrero CA, Perotti A, Petri G and Tamietto M (2023) Convolutional neural networks for vision neuroscience: significance, developments, and outstanding issues.
Front. Comput. Neurosci. 17:1153572.
doi: 10.3389/fncom.2023.1153572

Convolutional neural networks for vision neuroscience: significance, developments, and outstanding issues

Alessia Celeghin[1*†], Alessio Borriero[1†], Davide Orsenigo[1], Matteo Diano[1], Carlos Andrés Méndez Guerrero[2], Alan Perotti[3], Giovanni Petri[3] and Marco Tamietto[1,4*]

[1]Department of Psychology, University of Torino, Turin, Italy, [2]Institut des Sciences Cognitives Marc Jeannerod, CNRS, Université de Lyon, Lyon, France, [3]CENTAI Institute, Turin, Italy, [4]Department of Medical and Clinical Psychology, and CoRPS–Center of Research on Psychology in Somatic Diseases–Tilburg University, Tilburg, Netherlands

Convolutional Neural Networks (CNN) are a class of machine learning models predominately used in computer vision tasks and can achieve human-like performance through learning from experience. Their striking similarities to the structural and functional principles of the primate visual system allow for comparisons between these artificial networks and their biological counterparts, enabling exploration of how visual functions and neural representations may emerge in the real brain from a limited set of computational principles. After considering the basic features of CNNs, we discuss the opportunities and challenges of endorsing CNNs as *in silico* models of the primate visual system. Specifically, we highlight several emerging notions about the anatomical and physiological properties of the visual system that still need to be systematically integrated into current CNN models. These tenets include the implementation of parallel processing pathways from the early stages of retinal input and the reconsideration of several assumptions concerning the serial progression of information flow. We suggest design choices and architectural constraints that could facilitate a closer alignment with biology provide causal evidence of the predictive link between the artificial and biological visual systems. Adopting this principled perspective could potentially lead to new research questions and applications of CNNs beyond modeling object recognition.

KEYWORDS

Convolutional Neural Networks (CNN), visual system, ventral stream, blindsight, superior colliculus, pulvinar, V1-independent vision

The place of Convolutional Neural Networks between neuroscience and cognitive sciences

The brain processes multidimensional and context-dependent information about the world to generate appropriate behaviors. Models in cognitive sciences capture principles of brain information processing but typically overlook fine details about the spatiotemporal implementation of neuronal functions or biological components. On the other hand,

neurobiological models recapitulate dynamics of action potentials or signal propagation across neuronal populations. However, they have limited success in understanding the computations that support complex behaviors in real-life contexts (Kriegeskorte and Golan, 2019).

In parallel to research in neuroscience, CNNs have become a powerful tool in machine learning and AI that can attain human-like performance. In fact, CNNs can approximate functions in complex and real-world tasks, such as visual recognition (Krizhevsky et al., 2012), language processing (Wolf et al., 2020), or motor learning (Mnih et al., 2015; Dhawale et al., 2017), in ways resembling biological agents. Fueled by the current success of AI and computer vision, current studies suggest that CNNs can potentially bridge the gap between "disembodied" descriptions of cognitive functions and neurobiological models, thus offering a new framework for predicting brain information processing (Kriegeskorte and Golan, 2019). At its core, this framework explains sensory, cognitive, and motor functions in terms of local computations that emerge from experience in networks of units aggregated in multiple (i.e., deep) layers.

Nevertheless, the synergy between neuroscience and AI remains elusive unless the opportunities and challenges of reframing classic questions in neuroscience as deep learning problems are considered (Saxe et al., 2020). Why should we study biological brains through the lens of CNNs? Which new research question do CNNs allow to emerge in neuroscience? What do they offer more than, or differently from, traditional models in terms of predictions, interpretability, or explanatory power? In this short review, we summarize basic concepts and describe recent progress at the intersection between neuroscience and AI, limiting our discussion to the primate visual system and its functions. We then outline some principles in visual neuroscience that still need to be systematically integrated into current CNN models of the primate visual brain. These principles are cornerstones to improve the neurobiological realism of CNNs, and we propose examples of design choices and architectural constraints that may permit a closer match to biology. To this end, we would like to contribute to setting a roadmap for vision neuroscientists interested in drawing on CNNs toolkit.

Basic principles of CNNs for vision neuroscience

Convolutional Neural Networks are a particular class of artificial neural networks inspired by the architecture and basic functions of biological vision (LeCun et al., 1989; Gu et al., 2018). In contrast to non-convolutional fully connected networks, where each neuron in a layer is connected to all neurons in the previous layer, CNNs employ local connectivity and shared weights through convolutional layers. In general, a CNN consists of many processing units akin to neurons, arranged in interconnected layers typically interpreted as being analogous to brain areas, and with connections defined by weights that mimic the integration and activation properties of synapses. The output of one stage of operations is typically a non-linear combination of the input received and is then passed on to the next layer. This circuit motif is repeated several times and creates a hierarchical organization until the cascade culminates with a discriminative classification or regression generated by the last layers used for readout. A convolutional layer contains many filters with distinct receptive fields that are applied to the input image through a convolution operation, which allows the network to capture spatial and temporal patterns in the input data. The filter's weights are the learnable parameters of these layers, with the learning process typically managed by standard gradient descent algorithms (LeCun et al., 2015; Yamins and DiCarlo, 2016; Barrett et al., 2019; Kragel et al., 2019; Hasson et al., 2020).

Historically, the development of CNNs was informed by tuning properties of simple and complex neurons in the primary visual cortex (V1), the major cortical target of retinal information, modeled through handcrafted Gabor filters (Fukushima, 1980; Riesenhuber and Poggio, 2000). This bottom-up approach proved effective in modeling V1 properties, but had limited success when extended to higher-order cortical areas along the visual ventral stream (Gallant et al., 1996; Kriegeskorte, 2009). These limitations contributed to shifting the focus toward a goal-driven approach by maximizing, for example, the classification accuracy (Yamins and DiCarlo, 2016). Instead of characterizing the coding properties of individual neurons to enforce model parameters, the workflow of the goal-driven approach reverses the order: first, optimize the CNN to perform an ecologically relevant visual task, then compare artificial networks to real neural data.

Goal-driven CNNs learn to map input patterns (e.g., raw images) to output classifications (e.g., sorting natural images according to categories like faces, objects, and animals). They learn through training, in the form of supervised feedback or reward signals. Throughout this process, the network self-organizes, meaning that computations emerge spontaneously during training and weights change with repeated exposures to labeled or rewarded images (e.g., using the backpropagation algorithm) (LeCun et al., 2015). On the one hand, the classical pattern of interleaving convolutional with pooling layers produces filters with increasing receptive fields as the network layers are traversed forward. On the other hand, the goal-driven learning process does not explicitly enforce any kind of structure in the internal computations. The researcher examines the population-level description of how computations arise organically over the course of network training. The underlying assumption is that hidden layers of a good network model would functionally behave like real neurons in the corresponding neural structures. It has been observed that CNNs commonly learn hierarchies of abstraction, with the first layers acting as Gabor filters while deeper layers become detectors of more complex patterns (Yamins and DiCarlo, 2016). This emergent property as sparked speculations as to whether trained CNNs can functionally behave like biological neurons in the corresponding neural structures (Khaligh-Razavi and Kriegeskorte, 2014; Kuzovkin et al., 2018; Van Dyck et al., 2021).

We argue that the bottom-up and the goal-driven approaches are not mutually exclusive. It is not futile, indeed, to characterize the coding properties of individual neurons, especially at the early stages of visual analyses (i.e., in the retina, subcortical structures receiving direct retinal input, or V1). Design choices that permit a closer match to biology also improve subsequent model fit to neural data or contribute to explaining the diversity of structural and functional properties in the visual brain. For

example, when connections between the first layers are constrained with a bottleneck that reduces neurons at the virtual retinal output, consistent with the anatomy of the optic nerve, early layers of the CNN exhibit spontaneously concentric center-surround responses, as in the thalamus, whereas later layers are tuned to orientations, as in V1 (Lindsey et al., 2019). Likewise, the relationship between high selectivity for orientation in simple cells and low selectivity to spatial phase in complex cells of V1 has long been debated, as some mammals lack orientation maps. Different forms of pooling implemented in a Sparse Deep Predictive Coding (SDPC) model account for the emergence of complex cells in V1, both with and without orientation maps (Boutin et al., 2022). Pooling in the feature space is responsible for the formation of orientation maps, whereas pooling in the retinotopic space is related to the emergence of complex cells. Therefore, CNN approaches to this issue suggest that the presence or absence of orientation maps results from diverse strategies employed by different species to achieve invariance in complex natural stimuli.

Convolutional Neural Networks can thus differ considerably in their objective function (the goal of the system), learning rule (how parameters are updated to improve the goal), and circuit architecture (how units are arranged and connected), which are the three central components specified by design (Richards et al., 2019). From an engineering perspective, CNNs are built to solve a task better than prior models, with less computational effort or fewer training examples. Exceeding human performance is desirable, and biological plausibility is not a driving or required factor. Conversely, a closer correspondence with biological brains is paramount from a neuroscientific standpoint. In the latter case, CNNs can be useful if they incorporate elements that parallel the architecture and principles of functioning of the biological visual system. By doing so, CNNs can offer mechanistic hypotheses and enable empirical exploration of how a pattern of behaviors and neural representations may arise in the real brain from a limited set of computational principles.

CNN as (partial) models of the visual brain: why and how

The visual system is typically represented as a constellation of different but interconnected maps harbored in anatomically distinguishable areas that analyses diverse input features, such as curvature, color, or motion. The division of labor across these areas is classically charted at the cortical level in two "pathways" or "streams," the dorsal and ventral, originally conceived to progress linearly and hierarchically from a common antecedent in V1 (Ungerleider and Mishkin, 1982; Goodale and Milner, 1992). Along the ventral stream, which courses from V1 to areas downstream up to the temporal pole, retinotopy decreases, receptive fields become progressively larger, and neural responses are increasingly complex and invariant to low-level changes in the input space. This cascade culminates at the apex of the ventral stream with "concept" cells that are tuned to specific (sub)categories, such as (famous) faces, bodies, or places (Quiroga et al., 2005).

The designed architecture of CNNs, as described herein, parallels that of the ventral stream along several dimensions: hierarchical sequence of organized stages, loose correspondence between different layers and visual areas such as V1, V2, V4 and IT, progressive increase of receptive field size and complexity. These features, combined with the evidence that artificial networks trained on an ecologically relevant task attain human-level performance and learn abstraction hierarchies, make CNNs credible candidates for modeling the ventral stream. However, these structural features and the objective function of CNNs are built by the experimenter. Arguably, artificial networks should exhibit additional properties and representations that are not explicitly engineered and that match those found in biological brains.

Assessing the behavioral correspondence

The equivalence between CNNs and biological brains can be profitably understood in the context of the behavioral outcomes they produce, beyond the overall accuracy in image classification for which the network has been explicitly optimized. CNNs trained for generic object recognition develop representations and categorical similarity that relate closely to human perceptual shape and semantic judgments (Kubilius et al., 2016). CNNs match human and non-human primate error patterns across object categories, viewpoint variations and similarity judgments (Rajalingham et al., 2018). However, a more fine-grained analysis of discrepancies at the level of individual images, typically achieved by comparing confusion matrices, reveals that artificial and biological agents make errors on different images. In comparison to humans, CNNs (i) rely more on texture to classify images, (ii) are more affected by perturbations that degrade image quality, like pixelate noise, spatial frequency filtering or occlusions, and (iii) exhibit robustness and generalizability still lower than biological vision (Ghodrati et al., 2014; Geirhos et al., 2017, 2018; Wichmann et al., 2017; Tang et al., 2018). While we acknowledge that gross similarities in object recognition between artificial and biological neural networks are encouraging, the extent to which existing CNNs reproduce the multiple ways biological agents classify natural images, especially at the level of single items, should not be overstated. These areas of mismatch are important endeavors to steer future research and to improve both neurobiological plausibility and predictive power of CNN models.

Examining neural correspondence

The overall similarities between CNNs and humans at the behavioral outcomes level motivate comparing their internal processing stages and representational transformations. How well the features learned by CNNs can predict brain responses? To what extent do top-down goals imposed at the output of the CNNs cause hidden layers to respond like real neurons at different stages along the ventral stream hierarchy?

One standard approach is to assess through a regression procedure the correspondence between multi-unit neuronal activity in different ventral stream areas of the primate brain, and the activity of artificial units in different layers of the CNN (Schrimpf et al., 2020). It turns out that neural activity at early

stages of the visual hierarchy, like V1, is well predicted by early layers of CNNs that develop Gabor-wavelet-like activation patterns (Cadena et al., 2019). Intermediate areas, like V4, that respond to complex curvature features are best reconstructed from activity in intermediate layers of CNNs, and top hidden layers of CNNs end up being predictive of infero-temporal (IT) neurons (Yamins et al., 2014; Anand et al., 2021). Similar CNN models trained on object categorization also predict responses at early and late stages of the human ventral stream at the aggregate population level of fMRI or MEG data (Cichy et al., 2016; Eickenberg et al., 2017). In this context, Brain-Score is a recent platform to systematically compare different artificial networks for object recognition on how well they approximate the brain's mechanisms of the ventral stream according to multiple neural and behavioral benchmarks (Schrimpf et al., 2018).

Another popular approach to assess the correspondence between artificial networks and the brain is representational similarity analysis (RSA) (Kriegeskorte, 2008). RSA builds up a distance matrix that represents how dissimilar are the responses for every pair of images presented to an "observer." Observers can be either biological brains (or specific brain areas) of different species or artificial networks (or their layers). As the dissimilarity is expressed in relative values, it abstracts from the specific methods (neural spikes, fMRI) or the nature of the observer wherein activity is recorded. For example, RSA was used to relate object representation in the IT cortex of humans and monkeys presented with the same images of real-world objects (Kriegeskorte, 2009). The same method has been used to compare different CNNs with representations in the human and monkey IT (Kriegeskorte and Kievit, 2013; Khaligh-Razavi and Kriegeskorte, 2014). These studies showed that better performing CNN models are also more similar to IT, as they develop greater clustering across categories and are also more sensitive to fine-grained dissimilarities within categories (e.g., faces and bodies form subclusters within animate items). In general, it seems that biological and artificial networks both impose upon the visual input certain categorical differentiations that are important for successful behavior.

Notions of the visual brain commonly overlooked in CNN models

In this section, we outline several notions informing the anatomical and physiological properties of the visual system that still need to be systematically transposed in CNN models through corresponding architectural and computational solutions, respectively. These principles are variably rooted in the process of phylogenetic evolution or acquired from learning in critical developmental periods. They offer insights into the complex interplay between the integration and segregation of functions within the visual system and the constraints that enable the brain to remodel itself through plasticity to compensate for the effects of lesions. Incorporating these notions in CNNs would advance our mechanistic explanation of how complex computations are possible using the machinery available to the biological brains and their driving forces across the life span.

Multiple routes bypass V1 and target higher-order visual cortices

There are multiple routes through which the visual input reaches the cortex from the retina (Pessoa and Adolphs, 2010; Baldwin and Bourne, 2020). The best-studied route targets V1 after an intermediate relay in the lateral geniculate nucleus of the thalamus (LGN). Standard CNNs loosely model this retino-geniculo-striate pathway with an initial front-end that approximates the retina and the two early layers thereafter. However, multiple pathways bypass V1 and target extra-striate visual areas (including ventral stream areas) through direct and indirect connections from LGN, the pulvinar and the superior colliculus (Bridge et al., 2016; Tamietto and Morrone, 2016; Bruni et al., 2018; McFadyen et al., 2019, 2020). Each of these subcortical structures receives direct projections from the retina. Such projections however, come from different classes of retinal ganglion cells (M, P, and K) specialized to respond to specific visual features (**Figure 1**).

The characterization of these V1-independent pathways in current CNN models of the visual system is important for several reasons. First, V1-independent pathways are not simply vestigial from a functional and anatomical perspective. After a lesion to V1, extra-striate areas remain responsive from 20% (Schmid et al., 2010) to 80% (Girard et al., 1992) of their pre-lesional activity. The retinal projection to the superior colliculus alone comprises about 100,000 fibers, which is more than the whole human auditory nerve. Second, these alternative pathways contribute to many important functions such as orientation, motion discrimination, object categorization and emotion processing, as these abilities can be retained in patients with V1 damage (Ajina et al., 2015a; de Gelder et al., 2015; Hervais-Adelman et al., 2015; Van den Stock et al., 2015; Ajina and Bridge, 2018; Celeghin et al., 2019). Third, it is becoming increasingly clear that the subcortical structures from which V1-independent pathways originate are not passive relay centers. Instead, they seem endowed with the necessary infrastructure and computational capabilities to instantiate complex analyses of the visual input (Bridge et al., 2016; Georgy et al., 2016; Basso et al., 2021; Carretié et al., 2021; Isa et al., 2021). Lastly, mounting evidence indicates that retino-recipient structures, like the superior colliculus or the pulvinar, provide the developmental foundation of what later in life become complex visual and attentional functions typically ascribed to higher-order cortical areas (Warner et al., 2012; Alves et al., 2022). For example, the superior colliculus has been proposed to establish new-born preferences for faces and facial expressions, and contribute to the maturation of "face patches" in areas of the ventral stream, such as the fusiform gyrus (Johnson, 2005; McFadyen et al., 2020). The pulvinar, through its direct connections to the area middle temporal (MT), drives the early maturation of the dorsal stream, which sustains global motion perception and serves visuo-motor integration (Warner et al., 2015; Kwan et al., 2021).

To the best of our knowledge, only one study has built a neurobiologically inspired CNN that simulates the physiological, anatomical, and connectional properties of the retino-collicular circuit and its contribution to facial expression categorization (Méndez et al., 2022; **Figure 2**). The model consists of a frontend that emulates retinal functions of M, P, and K pathways, along

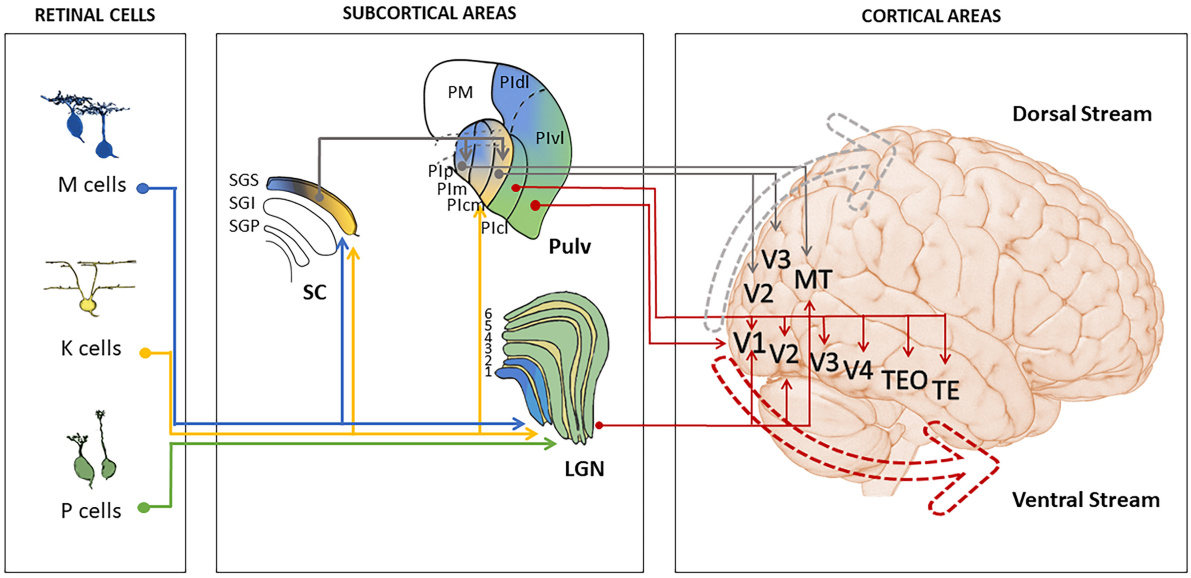

FIGURE 1
Connections from the retinal ganglion cells to the visual cortex intermediate relays in LGN, SC and Pulv. The blue arrow indicates projections from the M cells in the retina to the superficial layers of SC and magnocellular layers of the LGN. The Yellow arrow indicates the projection from K cells in the retina to the superficial layers of the SC and the intermediate layers of the LGN. The Green arrow indicates projection from the P cells in the retina to the magnocellular layers of the LGN. Gray arrows indicate projections originating from the superior colliculus and reaching the dorsal stream cortical areas via the pulvinar. The red arrows indicate projections from pulvinar subnuclei and LGN to areas along the cortical ventral stream. In LGN and superior colliculus, yellow layers indicate Koniocellular, blue Magnocellular, and green Parvocellular channels. In the pulvinar and SC these pathways are not clearly segregated and shaded blue-yellow; green-blue colors indicate the conjoint presence of the respective channels in given subdivisions. White denotes areas of the superior colliculus and pulvinar not interesting for the present purposes. PIcl, pulvinar inferior centro-lateral; PIcm, pulvinar inferior centro-medial; PIm, pulvinar inferior medial; PIp, pulvinar inferior posterior; PLdm, pulvinar lateral dorso-medial; PLvl, pulvinar lateral ventro-lateral; PM, pulvinar medial; TEO, temporal inferior posterior; TE, temporal inferior anterior.

with three layers analogous to the superficial strata of the primate superior colliculus that receive direct retinal information. This CNN matched error patterns and classification accuracy of patients with V1 damage, developed spontaneous tuning to low spatial frequencies in accordance with fMRI data, and generated saliency maps that directed attention to different facial features depending on expressions (Sahraie et al., 2010; Celeghin et al., 2015; Burra et al., 2019). These findings contribute to superseding a cortico-centric perspective on visual functions and to explore with CNNs the encoding of emotional information.

Dorsal and ventral stream: how many subsystems?

The division of extra-striate visual areas into dorsal and ventral streams is a crucial framework that has been heuristically seminal in visual neuroscience for the past four decades (Ungerleider and Mishkin, 1982; Goodale and Milner, 1992). However, some of its tenets have come under renewed scrutiny (de Haan and Cowey, 2011; Rossetti et al., 2017). For instance, the dorsal pathway is now conceived as a multiplicity of at least three segregate pathways based on different downstream projection targets that serve spatial working memory, visually guided action, and navigation (Kravitz et al., 2011). Similarly, the ventral stream has been proposed to encompass up to six distinct cortico-subcortical systems, each with specialized behavioral, cognitive, or affective functions (Kravitz et al., 2013). More radically, recent evidence suggests the existence of a third visual stream, terminating in the superior temporal sulcus (STS) (Pitcher and Ungerleider, 2021). This third stream appears specialized for the dynamic aspects of social perception and does not fit within the traditional dichotomy altogether.

As described previously, CNN applications have been essentially grounded on models of the ventral stream. However, there are interesting attempts to predict neural responses along the dorsal stream. Using an encoding model, a CNN has been trained to recognize actions in videos and map stimuli to their constituent features (Güçlü and van Gerven, 2017). These features were then regressed to fMRI activity in subjects watching natural movies. Through this method, it was possible to predict responses in the dorsal stream, with deeper layers corresponding to activity in downstream areas such as V3b and MT. Besides a few remarkable exceptions, most studies have generally failed to appreciate that the brain can exploit visual information to achieve different behavioral goals beyond foveal object recognition. These environmental constraints and adaptive pressures shape the functional segregation of different input properties at early encoding stages (Milner and Goodale, 1995). In this context, CNN models can be profitably applied to probe the development of specialized sub-pathways by investigating computational trade-offs and the underlying reasons for the emergence of specialized and segregated sub-systems. For example, a relatively generic architecture can be trained from the same starting point to perform different tasks. Then, the CNN is inspected to understand how many layers can be shared before performance declines and the network needs to split into specialized sub-streams to perform well on all tasks

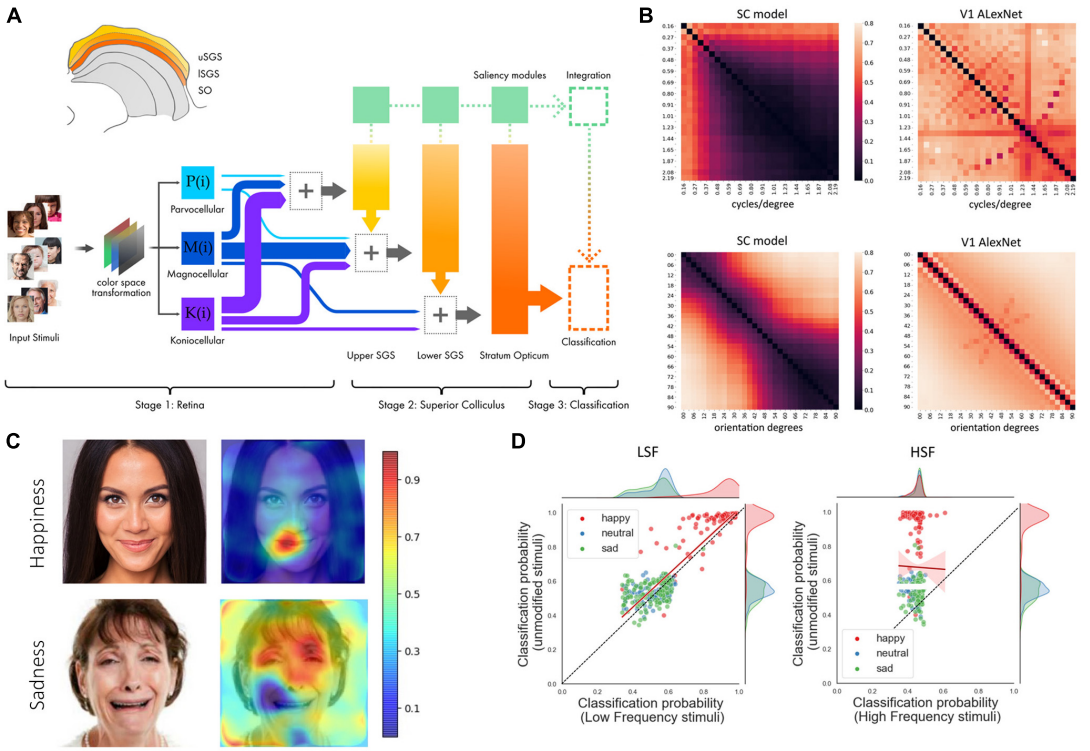

FIGURE 2
General overview of the model adopted in Méndez et al. (2022) and selected results. **(A)** In the upper left corner, an anatomical diagram of the SC, viewed from a coronal section, highlights the relevant superficial layers that have been modeled with corresponding layers in the CNN. Stage 1: input stimuli are color-transformed and processed. Each one of the three main P, M, and K channels is modeled by a function and the visual information is projected in different proportions to the appropriate SC layers. The varying width of the arrows represent the proportional contribution of each pathway to the corresponding superficial SC layer. Stage 2: the three layers composing the superficial SC are modeled with fractal convolutional blocks and saliency modules that guide attention to select image regions for further analysis. Stage 3: after the final SC layer, the network uses a global averaging stage before a classification layer, while the saliency masks are integrated and contribute to generate the final output of the network. **(B)** Spontaneous emergence of spatial frequency tuning and orientation sensitivity. Discrimination performance across gratings of different spatial frequencies (upper) and orientation (lower) in the SC model and in the AlexNet model that approximate V1 responses. **(C)** Example of bubbles analysis on three subjects from the test dataset. **(D)** Effect of face manipulation toward M and P channels in the SC model and in the fMRI activity of the human SC. Scatterplots and marginal distributions of classification probability for LSF and HSF filtered faces compared to original images. The *y*-axis corresponds to the probability of belonging to the correct category assigned by the model to each unfiltered image, while the *x*-axis represents the probability assigned when the instances are spatial frequency filtered. The red line and shaded area denote the best linear regression fit to the data and its 95% confidence interval. Figure modified from Méndez et al. (2022).

(Kell et al., 2018). Variants of this approach have been recently applied to study why and how face and object processing segregate in the visual system (Rawat and Wang, 2017; Dobs et al., 2019), or to provide computational foundations for dorsal and ventral streams to arise based on different goals and learning principles (Scholte et al., 2018).

Evaluating hierarchy and linearity of information integration

Uncertainty about how to aggregate the fractioned architecture of the visual brain into pathways also calls into question its hierarchical organization, which also assumes a serial progression of information and linear integration from lower-level to higher-order visual areas. For example, information exchange is reciprocal between adjacent structures, and, in most cases, backward projections outnumber forward projections (Angelucci et al., 2002). Moreover, "shortcut" connections link relatively distant areas: V1 projects directly to V3, V4, and MT; V2 to TEO; and V4 to TE. Developmentally, the traditional view of a hierarchical maturation, where V1 develops first followed by higher-order areas, is contradicted by recent evidence that MT matures in parallel due to driving pulvinar input in early postnatal phases (Bourne and Morrone, 2017). Finally, visual areas like prostriate exhibit response latency, receptive field characteristics, and projection patterns that, to some extent, contradict each other for classifying its hierarchical position and assignment to either the dorsal or ventral stream (Mikellidou et al., 2017; Tamietto and Leopold, 2018).

Concerning hierarchical organization and feed-forward vs. recursive interactions, standard CNNs approximate the initial stages of visual processing (~150 ms after stimulus onset) when the dominant direction of signal flow within occipito-temporal networks is feedforward (Tang et al., 2018; Semedo et al., 2022). However, rapid feedback interactions coexist with the initial feedforward sweep and can influence basic levels of visual processing, while at longer latencies information exchange gradually reverses to feedback (Bullier, 2001;

Semedo et al., 2022). Adding recurrent connections between layers improves equivalence with later stages of neural processing in both the dorsal and ventral stream (Shi et al., 2018; Kar et al., 2019; Kietzmann et al., 2019). One important extension of classical CNNs could thus be to systematically incorporate feedback connections and residual links that skip layers, as they seem to increase receptive field size of corresponding cell units (Jarvers and Neumann, 2019; Rawat and Wang, 2017). In fact, the comparison of different feedback and feedforward architectures suggests that including feedback connections and recurrence at either local and global network level (i.e., within and between layers, respectively) can improve network performance and robustness (Hasani et al., 2019).

Improving neurobiological plausibility of objective functions and learning rules

As previously mentioned, CNNs are grounded in three essential components: the objective function, the learning rules, and the network architecture (Richards et al., 2019). The principles of the primate visual brain discussed in the previous section can be mainly transposed in the next generation of CNNs through architectural solutions. We now complement the discussion with a focus on approaches aimed at improving the biological plausibility of artificial neural network through objective functions and learning rules.

Objective functions as ethological task constraints

Animals clearly possess objective functions crucial for survival, which rely variably on both evolution and learning processes. Examples of these functions include escaping predators, recognizing conspecifics, and seeking food. Task constraints are driving forces that shape brain architecture and functions to enhance fitness. Thus, they should be an integral component in developing neurobiologically plausible CNNs. For example, some authors have proposed that the expansion of the visual system, the rise of orbital convergence, and the development of foveal vision evolved to cope with evolutionary pressures favoring the emergence of visually guided reaching and grasping due to the arboreal lifestyle of early primates (Sussman, 1991). These developments have subsequently led to significant improvements in oculomotor behaviors, enabling more efficient visual search and precise localization of targets with minimal head or body movements (Schütz et al., 2011). Other accounts have suggested that detecting snakes before they strike was the primary selective pressure that drove the development of the anthropoids' visual system, with foveal vision linked to the development of trichromatic color perception (Isbell, 2006).

Recent examples provide insights into the potential benefits of including task constraints as neurobiologically meaningful objective functions in CNN design. Mnih et al. (2014) developed a recurrent CNN that learned to track a simple object without explicit training, reproducing foveation and saccading. Cheung et al. (2017) trained a CNN to perform a visual search task using a retinal front-end with a receptor lattice that could be moved across input images to mimic eye movement, foveating specific features and image parts. The optimization procedure resulted in a virtual retina displaying characteristics of the biological one, featuring high resolution and densely sampled fovea with small virtual receptive fields, and more coarsely sampled periphery with lower resolution and larger receptive fields. Notably, the CNN did not develop these biologically realistic properties when the system was endowed with additional actions, like zooming, which are absent in the biological visual system. These findings suggest that the receptor properties and arrangements in the primate retina can be profitably studied with CNNs, and that several aspects of primate vision may arise from evolutionary pressure to optimize visual world sampling through the integration of eye movements and fixations.

Introducing "visuo-motor" goals as objective functions seems especially ground-breaking when modeling dorsal stream responses that exploit visual information to guide subsequent actions. For example, motion parameters in retinal image must be integrated with oculomotor and vestibular signal to avoid collision, grasp objects, or stabilize gaze on items of interest while we move through the environment (Burr and Morrone, 2022). Accordingly, dorsal stream neuron response properties can be better predicted by a 3D Resnet model trained to orient itself during locomotion (i.e., estimating self-motion parameters from image sequences) compared to networks simply trained on action recognition (Mineault et al., 2021). Therefore, incorporating ethologically relevant tasks as objective functions in CNNs, such as navigation, object manipulation and visual search, can lead to a more comprehensive understanding of the visual brain, and it seems necessary to fully capture the diversity of biological vision, its organizing principles, and relation to other brain functions.

An intriguing research direction in computer vision is neuro-symbolic integration (Kroshchanka et al., 2022), which aims to combine connectionist models with structured knowledge representation. This approach aims to enforce a more structured representation learning by enriching the loss function used to train CNNs with additional domain-distilled information, such as taxonomical relations between objects (Chen et al., 2019; Bertinetto et al., 2020). Visual hierarchies (input images) are matched with ontological hierarchies (enriched loss functions), fostering robustness against adversarial attacks. Examining these models' behavioral and neural correspondence with the primate visual system presents a promising avenue for research.

Learning rules and synaptic weights

Learning rules guide the optimization of model parameters, expressed as synaptic weights, to achieve a specific objective function. CNNs typically employ backpropagation, a highly supervised learning process that provides explicit performance feedback (Lindsay, 2021). In contrast, unsupervised learning builds meaningful representations by utilizing the inherent structure of the data (i.e., without instructions). Vision neuroscience has traditionally emphasized unsupervised principles that modulate synaptic changes and local plasticity rules, such as Hebbian learning

and spike timing-dependent plasticity (Caporale and Dan, 2008; Chauhan et al., 2021). These principles are embodied in Spiking Neural Networks (SNNs), where network units process time-varying spikes at the input and output stages, mimicking time-dependent processing in natural vision (Maass, 1997; Tavanaei et al., 2019).

Backpropagation calculates gradients and adjusts weights between nodes during learning, relying on biologically unrealistic assumptions, such as symmetrical feedback weights and separate forward and backward information flows (McClelland et al., 1986; Bengio et al., 2017). Despite this, biological brains can approximate backpropagation learning when assumptions about inhibitory microcircuits, short-term plasticity, or feedback connections are considered (Körding and König, 2001; Roelfsema and van Ooyen, 2005; Lillicrap et al., 2016; Guerguiev et al., 2017; Scellier and Bengio, 2017; Whittington and Bogacz, 2017; Pozzi et al., 2018; Sacramento et al., 2018). Dynamic weight sharing has been recently proposed as a learning rule capable of accounting for local weight updates. It facilitates local weight adjustments via lateral connections, enabling local neuron subgroups to equalize weights through shared activity and anti-Hebbian learning (Pogodin et al., 2021). Artificial networks with dynamic weight sharing exhibit a better fit to ventral stream data, as measured by the Brain-Score, performing almost as well as traditional CNNs.

Recent studies on synaptic plasticity and learning have focused on top-down attentional mechanisms and predictive coding, both involving feedback connectivity and predicting varied activity distributions (Kwag and Paulsen, 2009; Yagishita et al., 2014; Bittner et al., 2017; Lacefield et al., 2019; Williams and Holtmaat, 2019). For instance, the influence of fronto-parietal attentional network over the visual system has been traditionally modeled in CNNs using saliency modules that guide visual selection for further processing of the most informative image parts (Itti et al., 1998). Additionally, attentional learning modules have been adapted to encode a topographic saliency map of the visual scene generated by the superior colliculus (Méndez et al., 2022).

Non-invasive imaging techniques, such as fMRI and wide-field calcium imaging, enable measuring the dynamics of representational changes and comparing learning trajectories during training. Estimating synaptic changes *in vivo* and relating them to behavioral performance can facilitate comparisons between artificial and biological brains based on learning procedures, rather than solely on the final representations they generate.

Toward causal evidence: closed-loop experiments and task constraints

Closed-loop experiments through image synthesis

Successful application of AI in neuroscience should permit moving the research agenda beyond correlations toward new approaches that gather causal evidence of the predictive link between artificial and biological brains. "Closed-loop" experiments harness CNNs activations to systematically manipulate brain activity in pre-defined visual regions of the brain, such as V1 and V4, according to the following logic (Bashivan et al., 2019; Ponce et al., 2019). First, a CNN presented with natural images is trained to predict neural activity recorded in the real brain, wherein the same images are shown to the animal. Then, the CNN is used to synthesize optimal images that maximally excite specific artificial units (or layers) by selecting their preferred features. Finally, when these synthetic images are shown to the real neurons, their responses are measured and found to match the predicted firing rate. This demonstrates that the CNN can capture the correspondence from pixels to neural responses (Olah et al., 2017; Walker et al., 2019). By enabling non-invasive control over brain activations, this closed-loop approach promises new causal insights into the interplay of multiple brain areas during visual processing. For example, the method permits stringent control of activity in one brain region while establishing the impact on the functioning of another related area.

Lesion analysis at the single neuron and population level

Lesion-symptom mapping is probably the most straightforward tool in neuroscience to establish the causal contribution of a neural structure to a given function and to investigate the plastic changes that intervene thereafter. However, this approach has inevitable limitations when applied to biological brains. Performing single-neuron ablations has traditionally proved challenging due to technical limitations (Wurtz, 2015). The advent of optogenetics, viral vectors, and two-photon stimulation techniques promise to overcome these challenges (Kinoshita and Isa, 2015; Kinoshita et al., 2019; Vanduffel and Li, 2020; Klink et al., 2021). However, these methods are still in infancy and their application to animal models phylogenetically proximal to humans has just begun. In primates, surgical lesions are still the most used approach at the areal or network level of analysis. Nevertheless, their precision and specificity vary depending on multiple factors, whereas, in humans, naturally occurring lesions obviously do not adhere to cytoarchitectonic or functional boundaries between areas.

Artificial networks can fulfill "*in silico* neurophysiology" at the single cell level exceedingly well, as we can characterize every unit's activity in response to predefined ablation and measure the impact on neural computation and performance (Barrett et al., 2019). This approach revealed that CNN accuracy drops as increasing numbers of neurons are deleted (Morcos et al., 2018). Moreover, networks that learn generalizable solutions are more robust to ablations than those that simply memories the training data (Zhou et al., 2018). Notably, neurons with clearly defined tuning properties are not more important for classification performance than those with complex or ambiguous tuning properties, as the latter often contains substantial amounts of task-relevant information. These findings contribute to reconsidering some basic assumptions in neurophysiology, where single-cell selectivity to stimulus features or categories has been traditionally regarded as the principal proxy to infer functions.

Drop-out, a randomized temporal ablation technique, is widely used in training artificial neural networks to ensure robustness. Although predominantly employed for regularization, it can also

profitably simulate "virtual lesions" and the resulting plasticity. To model neuroplasticity using CNNs, researchers can simulate the reorganization of neural connections and the emergence of new response properties following lesions by adjusting the network's architecture, connectivity, or learning rules. Neuroplasticity can be measured by evaluating the network's adaptability to external perturbations, such as introducing noise or limiting weight updates in a specific layer. This approach may reveal how mid-level features and new response properties emerge as the network compensates for the loss of specific neural elements. For example, how do response properties in intact visual structures change following brain damage? Neural tuning in extra-striate visual areas gradually recovers after V1 damage. However, this recovery does not lead extra-striate neurons to emulate the response properties of the damaged cortex but to resume their own original response properties (Maffei and Fiorentini, 1973; Guido et al., 1992). Furthermore, in humans, if V1 damage occurs in adulthood, the response in MT neurons for motion and contrast reshapes to resemble the response pattern of V1 in the intact brain (Ajina et al., 2015b).

Lesion analysis can be applied to deep learning architectures incorporating features like lateral connection or shortcut and learning through an online procedure, where weights are updated one data sample at a time. Investigating how shortcut connections serve as alternative information pathways could also yield valuable insights. CNNs can be exploited to address these issues with new tools offering insights into the mechanisms underlying neuroplasticity and potentially guiding the development of interventions to promote recovery after brain damage.

Conclusion

The resurgence of interest in neural networks has sparked both enthusiasm and skepticism regarding the relevance of CNNs in understanding biological brains (Michel et al., 2019). Experimenting with CNNs offers valuable insights for neuroscience, especially if biological credibility is recognized as a crucial factor in modeling network properties, and results are deployed in assays on biological brains. In turn, laboratory investigations should drive the design of future CNN models. This iterative process offers a principled perspective to specifying mechanistic hypotheses on how real brains may carry out visual and cognitive functions.

Longstanding questions include whether perceptual representations, like sensitivity to biological motion or face recognition, are innate or learned from experience (Behrmann and Avidan, 2022). While traditional supervised models used for explaining primate object recognition demand vast labeled data, primates develop sophisticated object understanding with limited training and less examples (Saxe et al., 2020). However, (quasi)innate behaviors can partly be conceived as learned on an evolutionary timescale, and the relationship between evolutionary and developmental variations can be reframed in CNNs. Evolutionary diversity can be addressed by changing architectural parameters that restructure the computational primitives of the network, while development can be modeled by modifying filter parameters and their learning algorithms, imitating synaptic weights.

To align CNNs with biology and to steer future directions, it seems useful to consider neural responses and their functions as an emerging consequence of the interplay between objective functions, learning rules and architecture. The environment and its constraints seem to provide guidance in identifying which objective functions are useful for biological brains to optimize. Accordingly, introducing ethologically relevant tasks as objective functions in CNNs, such as navigation, object manipulation, and visual search, can lead to a more comprehensive understanding of the visual brain.

By exploring unsupervised learning principles and spiking neural networks, researchers can better understand the role of local plasticity rules and time-varying signals in natural vision, thereby improving the neural correspondence of CNNs. This may involve exploring alternative learning algorithms, such as reinforcement learning, that incorporate elements of reward-based learning and decision-making. Analyzing the processing of affective signals also offers a testing ground for the proposal that representation formation is driven by the need to predict the motivational value of experience and its interface with attention (Mnih et al., 2015).

Architecturally, it is important to incorporate the role of subcortical structures and V1-independent pathways in visual processing. This would quantify the respective contributions of redundancy and synergy in the multiplicity of parallel routes that help decode visual stimuli (Nigam et al., 2019; Luppi et al., 2022). Systematically integrating feedback connections, recurrent connectivity motifs, and residual links can enhance performance, robustness, and equivalence with the brain's hierarchical organization and the balance between linear and recursive interactions. CNN models can be utilized to probe the development of functional segregation and the emergence of specialized subsystems through computational trade-offs.

The use of closed-loop experiments and lesion analysis can provide new causal insights into the predictive link between artificial and biological brains, the mechanisms underlying neuroplasticity, and the development of interventions to promote recovery after brain damage. For example, what is the impact of silencing a single structure on the computations performed in other parts of the network at both the aggregate level of layers and of single units? In CNN, this would imply assessing network robustness to "virtual ablations" of individual components and can help evaluate how the biological brain recruits plasticity.

Convolutional Neural Networks face the unavoidable trade-off between complexity, interpretability, and energy consumption (Petri et al., 2021). On this front, the sparsity of spike computing is central to information processing where computationally demanding tasks can be realized by a restricted subset of neurons and disentangled from millions of examples through "direct fit" (Olshausen and Field, 2004; Dalgleish et al., 2020; Hasson et al., 2020). Transformer models originally applied to natural language tasks are finding their way in the vision science community (Khan et al., 2022). Unlike CNNs, transformers support parallel processing, require minimal inductive biases for their design, and allow simultaneous processing of multiple modalities.

The development of CNNs is progressing rapidly and spreading in different directions and domains within neuroscience. The theoretical discussion and sober consideration of the promises and pitfalls accompanying these developments are needed to ensure that neuroscientists make informed use of CNNs as falsifiable

models of biological brains. By addressing these critical areas, researchers can harness the full potential of CNNs to bridge the gap between neuroscience and cognitive sciences, ultimately leading to a deeper understanding of the primate visual system and advancing artificial intelligence.

Author contributions

AC, CAMG, and MT contributed to the conception of the study. MT, AC, CAMG, and AB wrote the first draft of the manuscript. DO, MD, AP, and GP wrote sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

Funding

This study was supported by the European Research Council (ERC) Consolidator Grant 2017 "LIGHTUP" (772953) and PRIN 2017 grant from the Ministero dell'Università e della Ricerca (MIUR, Italy) (2017TBA4KS) to MT.

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

References

Ajina, S., and Bridge, H. (2018). Blindsight relies on a functional connection between hMT+ and the lateral geniculate nucleus, not the pulvinar. *PLoS Biol.* 16:e2005769. doi: 10.1371/journal.pbio.2005769

Ajina, S., Kennard, C., Rees, G., and Bridge, H. (2015b). Motion area V5/MT+ response to global motion in the absence of V1 resembles early visual cortex. *Brain* 138, 164–178. doi: 10.1093/brain/awu328

Ajina, S., Pestilli, F., Rokem, A., Kennard, C., and Bridge, H. (2015a). Human blindsight is mediated by an intact geniculo-extrastriate pathway. *Elife* 4:e08935. doi: 10.7554/eLife.08935

Alves, P. N., Forkel, S. J., Corbetta, M., and Thiebaut de Schotten, M. (2022). The subcortical and neurochemical organization of the ventral and dorsal attention networks. *Commun. Biol.* 5:1343. doi: 10.1038/s42003-022-04281-0

Anand, A., Sen, S., and Roy, K. (2021). Quantifying the brain predictivity of artificial neural networks with nonlinear response mapping. *Front. Comput. Neurosci.* 15:609721. doi: 10.3389/fncom.2021.609721

Angelucci, A., Levitt, J. B., Walton, E. J. S., Hupé, J.-M., Bullier, J., and Lund, J. S. (2002). Circuits for local and global signal integration in primary visual cortex. *J. Neurosci.* 22, 8633–8646. doi: 10.1523/jneurosci.22-19-08633.2002

Baldwin, M. K. L., and Bourne, J. A. (2020). "The evolution of subcortical pathways to the extrastriate cortex," in *Evolutionary neuroscience*, ed. J. H. Kaas (Cambridge, MA: Academic Press), 565–587. doi: 10.1016/B978-0-12-820584-6.00024-6

Barrett, D. G., Morcos, A. S., and Macke, J. H. (2019). Analyzing biological and artificial neural networks: Challenges with opportunities for synergy? *Curr. Opin. Neurobiol.* 55, 55–64. doi: 10.1016/j.conb.2019.01.007

Bashivan, P., Kar, K., and DiCarlo, J. J. (2019). Neural population control via deep image synthesis. *Science* 364:eaav9436. doi: 10.1126/science.aav9436

Basso, M. A., Bickford, M. E., and Cang, J. (2021). Unraveling circuits of visual perception and cognition through the superior colliculus. *Neuron* 109, 918–937. doi: 10.1016/j.neuron.2021.01.013

Behrmann, M., and Avidan, G. (2022). Face perception: Computational insights from phylogeny. *Trends Cogn. Sci.* 26, 350–363. doi: 10.1016/j.tics.2022.01.006

Bengio, Y., Goodfellow, I., and Courville, A. (2017). *Deep learning*, Vol. 1. Cambridge, MA: MIT Press.

Bertinetto, L., Mueller, R., Tertikas, K., Samangooei, S., and Lord, N. A. (2020). "Making better mistakes: Leveraging class hierarchies with deep networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, (Piscataway, NJ), 12506–12515.

Bittner, K. C., Milstein, A. D., Grienberger, C., Romani, S., and Magee, J. C. (2017). Behavioral time scale synaptic plasticity underlies CA1 place fields. *Science* 357, 1033–1036.

Bourne, J. A., and Morrone, M. C. (2017). Plasticity of visual pathways and function in the developing brain: Is the pulvinar a crucial player? *Front. Syst. Neurosci.* 11:3. doi: 10.3389/fnsys.2017.00003

Boutin, V., Franciosini, A., Chavane, F., and Perrinet, L. U. (2022). Pooling strategies in V1 can account for the functional and structural diversity across species. *PLoS Comput. Biol.* 18:e1010270. doi: 10.1371/journal.pcbi.1010270

Bridge, H., Leopold, D. A., and Bourne, J. A. (2016). Adaptive pulvinar circuitry supports visual cognition. *Trends Cogn. Sci.* 20, 146–157. doi: 10.1016/j.tics.2015.10.003

Bruni, S., Gerbella, M., Bonini, L., Borra, E., Coudé, G., Ferrari, P. F., et al. (2018). Cortical and subcortical connections of parietal and premotor nodes of the monkey hand mirror neuron network. *Brain Struct. Funct.* 223, 1713–1729. doi: 10.1007/s00429-017-1582-0

Bullier, J. (2001). Integrated model of visual processing. *Brain Res. Rev.* 36, 96–107. doi: 10.1016/S0165-0173(01)00085-6

Burr, D., and Morrone, M. C. (2022). Vision: Neuronal mechanisms enabling stable perception. *Curr. Biol.* 32, R1338–R1340.

Burra, N., Hervais-Adelman, A., Celeghin, A., De Gelder, B., and Pegna, A. J. (2019). Affective blindsight relies on low spatial frequencies. *Neuropsychologia* 128, 44–49. doi: 10.1016/j.neuropsychologia.2017.10.009

Cadena, S. A., Denfield, G. H., Walker, E. Y., Gatys, L. A., Tolias, A. S., Bethge, M., et al. (2019). Deep convolutional models improve predictions of macaque V1 responses to natural images. *PLoS Comput. Biol.* 15:e1006897. doi: 10.1371/journal.pcbi.1006897

Caporale, N., and Dan, Y. (2008). Spike timing-dependent plasticity: A hebbian learning rule. *Annu. Rev. Neurosci.* 31, 25–46. doi: 10.1146/annurev.neuro.31.060407.12563

Carretié, L., Yadav, R. K., and Méndez-Bértolo, C. (2021). The missing link in early emotional processing. *Emot. Rev.* 13, 225–244.

Celeghin, A., Bagnis, A., Diano, M., Méndez, C. A., Costa, T., and Tamietto, M. (2019). Functional neuroanatomy of blindsight revealed by activation likelihood estimation meta-analysis. *Neuropsychologia* 128, 109–118. doi: 10.1016/j.neuropsychologia.2018.06.007

Celeghin, A., de Gelder, B., and Tamietto, M. (2015). From affective blindsight to emotional consciousness. *Conscious. Cogn.* 36, 414–425. doi: 10.1016/j.concog.2015.05.007

Chauhan, T., Masquelier, T., and Cottereau, B. R. (2021). Sub-optimality of the early visual system explained through biologically plausible plasticity. *Front. Neurosci.* 15:727448. doi: 10.3389/fnins.2021.727448

Chen, H. Y., Tsai, L. H., Chang, S. C., Pan, J. Y., Chen, Y. T., Wei, W., et al. (2019). Learning with hierarchical complement objective. *arXiv* [Preprint]. arXiv:1911.07257.

Cheung, B., Weiss, E., and Olshausen, B. (2017). Emergence of foveal image sampling from learning to attend in visual scenes. *arXiv* [Preprint]. arXiv:1611.09430v2.

Cichy, R. M., Khosla, A., Pantazis, D., Torralba, A., and Oliva, A. (2016). Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. *Sci. Rep.* 6:27755. doi: 10.1038/srep27755

Dalgleish, H. W., Russell, L. E., Packer, A. M., Roth, A., Gauld, O. M., Greenstreet, F., et al. (2020). How many neurons are sufficient for perception of cortical activity? *Elife* 9:e58889. doi: 10.7554/eLife.58889

de Gelder, B., Tamietto, M., Pegna, A. J., and Van den Stock, J. (2015). Visual imagery influences brain responses to visual stimulation in bilateral cortical blindness. *Cortex* 72, 15–26. doi: 10.1016/j.cortex.2014.11.009

de Haan, E. H. F., and Cowey, A. (2011). On the usefulness of 'what' and 'where' pathways in vision. *Trends Cogn. Sci.* 15, 460–466. doi: 10.1016/j.tics.2011.08.005

Dhawale, A. K., Smith, M. A., and Ölveczky, B. P. (2017). The role of variability in motor learning. *Annu. Rev. Neurosci.* 40, 479–498. doi: 10.1146/annurev-neuro-072116-031548

Dobs, K., Isik, L., Pantazis, D., and Kanwisher, N. (2019). How face perception unfolds over time. *Nat. Commun.* 10:1258. doi: 10.1038/s41467-019-09239-1

Eickenberg, M., Gramfort, A., Varoquaux, G., and Thirion, B. (2017). Seeing it all: Convolutional network layers map the function of the human visual system. *Neuroimage* 152, 184–194. doi: 10.1016/j.neuroimage.2016.10.001

Fukushima, K. (1980). A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol. Cybern.* 36, 193–202.

Gallant, J. L., Connor, C. E., Rakshit, S., Lewis, J. W., and Van Essen, D. C. (1996). Neural responses to polar, hyperbolic, and Cartesian gratings in area V4 of the macaque monkey. *J. Neurophysiol.* 76, 2718–2739. doi: 10.1152/jn.1996.76.4.2718

Geirhos, R., Janssen, D. H., Schütt, H. H., Rauber, J., Bethge, M., and Wichmann, F. A. (2017). Comparing deep neural networks against humans: Object recognition when the signal gets weaker. *arXiv* [Preprint]. arXiv:1706.06969.

Geirhos, R., Temme, C. R., Rauber, J., Schütt, H. H., Bethge, M., and Wichmann, F. A. (2018). "Generalisation in humans and deep neural networks," in *Proceedings of the 32nd international conference on neural information processing systems*, (Red Hook, NY), 7538–7550.

Georgy, L., Celeghin, A., Marzi, C. A., Tamietto, M., and Ptito, A. (2016). The superior colliculus is sensitive to gestalt-like stimulus configuration in hemispherectomy patients. *Cortex* 81, 151–161. doi: 10.1016/j.cortex.2016.04.018

Ghodrati, M., Farzmahdi, A., Rajaei, K., Ebrahimpour, R., and Khaligh-Razavi, S.-M. (2014). Feedforward object-vision models only tolerate small image variations compared to human. *Front. Comput. Neurosci.* 8:74. doi: 10.3389/fncom.2014.00074

Girard, P., Salin, P. A., and Bullier, J. (1992). Response selectivity of neurons in area MT of the macaque monkey during reversible inactivation of area V1. *J. Neurophysiol.* 67, 1437–1446. doi: 10.1152/jn.1992.67.6.1437

Goodale, M. A., and Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends Neurosci.* 15, 20–25. doi: 10.1016/0166-2236(92)90344-8

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., et al. (2018). Recent advances in convolutional neural networks. *Pattern Recogn.* 77, 354–377. doi: 10.1016/j.patcog.2017.10.013

Güçlü, U., and van Gerven, M. A. J. (2017). Increasingly complex representations of natural movies across the dorsal stream are shared between subjects. *Neuroimage* 145, 329–336. doi: 10.1016/j.neuroimage.2015.12.036

Guerguiev, J., Lillicrap, T. P., and Richards, B. A. (2017). Towards deep learning with segregated dendrites. *Elife* 6:e22901.

Guido, W., Spear, P. D., and Tong, L. (1992). How complete is physiological compensation in extrastriate cortex after visual cortex damage in kittens? *Exp. Brain Res.* 91, 455–466. doi: 10.1007/BF00227841

Hasani, H., Soleymani, M., and Aghajan, H. (2019). Surround Modulation: A bio-inspired connectivity structure for convolutional neural networks. *Neural Inform. Proc. Syst.* 32, 15877–15888.

Hasson, U., Nastase, S. A., and Goldstein, A. (2020). Direct fit to nature: An evolutionary perspective on biological and artificial neural networks. *Neuron* 105, 416–434. doi: 10.1016/j.neuron.2019.12.002

Hervais-Adelman, A., Legrand, L. B., Zhan, M., Tamietto, M., De Gelder, B., and Pegna, A. J. (2015). Looming sensitive cortical regions without V1 input: Evidence from a patient with bilateral cortical blindness. *Front. Integr. Neurosci.* 9:51. doi: 10.3389/fnint.2015.00051

Isa, T., Marquez-Legorreta, E., Grillner, S., and Scott, E. K. (2021). The tectum/superior colliculus as the vertebrate solution for spatial sensory integration and action. *Curr. Biol.* 31, R741–R762. doi: 10.1016/j.cub.2021.04.001

Isbell, L. A. (2006). Snakes as agents of evolutionary change in primate brains. *J. Hum. Evol.* 51, 1–35. doi: 10.1016/j.jhevol.2005.12.012

Itti, L., Koch, C., and Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* 20, 1254–1259.

Jarvers, C., and Neumann, H. (2019). "Incorporating feedback in convolutional neural networks," in *Proceeding of the 2019 conference on cognitive computational neuroscience*, (Berlin), doi: 10.32470/ccn.2019.1191-0

Johnson, M. H. (2005). Subcortical face processing. *Nat. Rev. Neurosci.* 6, 766–774. doi: 10.1038/nrn1766

Kar, K., Kubilius, J., Schmidt, K., Issa, E. B., and DiCarlo, J. J. (2019). Evidence that recurrent circuits are critical to the ventral stream's execution of core object recognition behavior. *Nat. Neurosci.* 22, 974–983. doi: 10.1038/s41593-019-0392-5

Kell, A. J. E., Yamins, D. L. K., Shook, E. N., Norman-Haignere, S. V., and McDermott, J. H. (2018). A task-optimized neural network replicates human auditory behavior, predicts brain responses, and reveals a cortical processing hierarchy. *Neuron* 98, 630–644.e16. doi: 10.1016/j.neuron.2018.03.044

Khaligh-Razavi, S. M., and Kriegeskorte, N. (2014). Deep Supervised, but not unsupervised, models may explain IT cortical representation. *PLoS Comput. Biol.* 10:e1003915. doi: 10.1371/journal.pcbi.1003915

Khan, S., Naseer, M., Hayat, M., Zamir, S. W., Khan, F. S., and Shah, M. (2022). Transformers in vision: A survey. *ACM Comput. Surveys* 54, 1–41. doi: 10.1145/3505244

Kietzmann, T. C., Spoerer, C. J., Sörensen, L. K., Cichy, R. M., Hauk, O., and Kriegeskorte, N. (2019). Recurrence is required to capture the representational dynamics of the human visual system. *Proc. Natl. Acad. Sci. U.S.A.* 116, 21854–21863. doi: 10.1073/pnas.1905544116

Kinoshita, M., and Isa, T. (2015). "Potential of optogenetics for the behavior manipulation of non-human primates," in *Optogenetics*, eds H. Yawo, H. Kandori, and A. Koizumi (Tokyo: Springer), 279–290. doi: 10.1007/978-4-431-55516-2_19

Kinoshita, M., Kato, R., Isa, K., Kobayashi, K., Kobayashi, K., Onoe, H., et al. (2019). Dissecting the circuit for blindsight to reveal the critical role of pulvinar and superior colliculus. *Nat. Commun.* 10:135. doi: 10.1038/s41467-018-08058-0

Klink, P. C., Aubry, J., Ferrera, V. P., Fox, A. S., Froudist-Walsh, S., Jarraya, B., et al. (2021). Combining brain perturbation and neuroimaging in non-human primates. *Neuroimage* 235:118017. doi: 10.1016/j.neuroimage.2021.118017

Körding, K. P., and König, P. (2001). Supervised and unsupervised learning with two sites of synaptic integration. *J. Comput. Neurosci.* 11, 207–215.

Kragel, P. A., Reddan, M. C., LaBar, K. S., and Wager, T. D. (2019). Emotion schemas are embedded in the human visual system. *Sci. Adv.* 5:eaaw4358. doi: 10.1126/sciadv.aaw4358

Kravitz, D. J., Saleem, K. S., Baker, C. I., Ungerleider, L. G., and Mishkin, M. (2013). The ventral visual pathway: An expanded neural framework for the processing of object quality. *Trends Cogn. Sci.* 17, 26–49. doi: 10.1016/j.tics.2012.10.011

Kravitz, D., Saleem, K., Baker, C., and Mishkin, M. (2011). A new neural framework for visuospatial processing. *J. Vis.* 11, 319–319. doi: 10.1167/11.11.923.t

Kriegeskorte, N. (2008). Representational similarity analysis–connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* 2:4. doi: 10.3389/neuro.06.004.2008

Kriegeskorte, N. (2009). Relating population-code representations between man, monkey, and computational models. *Front. Neurosci.* 3:363–373. doi: 10.3389/neuro.01.035.2009

Kriegeskorte, N., and Golan, T. (2019). Neural network models and deep learning. *Curr. Biol.* 29, R231–R236. doi: 10.1016/j.cub.2019.02.034

Kriegeskorte, N., and Kievit, R. A. (2013). Representational geometry: Integrating cognition, computation, and the brain. *Trends Cogn. Sci.* 17, 401–412. doi: 10.1016/j.tics.2013.06.007

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Commun. ACM* 60, 84–90. doi: 10.1145/3065386

Kroshchanka, A., Golovko, V., Mikhno, E., Kovalev, M., Zahariev, V., and Zagorskij, A. (2022). "A neural-symbolic approach to computer vision," in *Open semantic technologies for intelligent systems OSTIS 2021. Communications in computer and information science*, Vol. 1625, eds V. Golenkov, V. Krasnoproshin, V. Golovko, and D. Shunkevich (Cham: Springer), doi: 10.1007/978-3-031-15882-7_15

Kubilius, J., Bracci, S., and Op de Beeck, H. P. (2016). Deep neural networks as a computational model for human shape sensitivity. *PLoS Comput. Biol.* 12:e1004896. doi: 10.1371/journal.pcbi.1004896

Kuzovkin, I., Vicente, R., Petton, M., Lachaux, J.-P., Baciu, M., Kahane, P., et al. (2018). Activations of deep convolutional neural networks are aligned with gamma band activity of human visual cortex. *Commun. Biol.* 1:107. doi: 10.1038/s42003-018-0110-y

Kwag, J., and Paulsen, O. (2009). The timing of external input controls the sign of plasticity at local synapses. *Nat. Neurosci.* 12, 1219–1221. doi: 10.1038/nn.2388

Kwan, W. C., Chang, C. K., Yu, H. H., Mundinano, I. C., Fox, D. M., Homman-Ludiye, J., et al. (2021). Visual cortical area MT is required for development of the dorsal stream and associated visuomotor behaviors. *J. Neurosci.* 41, 8197–8209. doi: 10.1523/JNEUROSCI.0824-21.2021

Lacefield, C. O., Pnevmatikakis, E. A., Paninski, L., and Bruno, R. M. (2019). Reinforcement learning recruits somata and apical dendrites across layers of primary sensory cortex. *Cell Rep.* 26, 2000–2008. doi: 10.1016/j.celrep.2019.01.093

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi: 10.1038/nature14539

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., et al. (1989). Backpropagation applied to handwritten zip code recognition. *Neural Comput.* 1, 541–551. doi: 10.1162/neco.1989.1.4.541

Lillicrap, T. P., Cownden, D., Tweed, D. B., and Akerman, C. J. (2016). Random synaptic feedback weights support error backpropagation for deep learning. *Nat. Commun.* 7:13276. doi: 10.1038/ncomms13276

Lindsay, G. W. (2021). Convolutional neural networks as a model of the visual system: Past, present, and future. *J. Cogn. Neurosci.* 33, 2017–2031. doi: 10.1162/jocn_a_01544

Lindsey, J., Ocko, S. A., Ganguli, S., and Deny, S. (2019). A unified theory of early visual representations from retina to cortex through anatomically constrained deep CNNs. *bioRxiv* [Preprint] doi: 10.1101/511535

Luppi, A. I, Mediano, P. A., Rosas, F. E., Holland, N., Fryer, T. D., O'Brien, J. T., et al. (2022). A synergistic core for human brain evolution and cognition. *Nat. Neurosci.* 25, 771–782. doi: 10.1038/s41593-022-01070-0

Maass, W. (1997). Networks of spiking neurons: The third generation of neural network models. *Neural Netw.* 10, 1659–1671. doi: 10.1016/s0893-6080(97)00011-7

Maffei, L., and Fiorentini, A. (1973). The visual cortex as a spatial frequency analyser. *Vis. Res.* 13, 1255–1267. doi: 10.1016/0042-6989(73)90201-0

McClelland, J. L., Rumelhart, D. E., and PDP Research Group (1986). *Parallel distributed processing*, Vol. 2. Cambridge, MA: MIT Press, 20–21.

McFadyen, J., Dolan, R. J., and Garrido, M. I. (2020). The influence of subcortical shortcuts on disordered sensory and cognitive processing. *Nat. Rev. Neurosci.* 21, 264–276. doi: 10.1038/s41583-020-0287-1

McFadyen, J., Mattingley, J. B., and Garrido, M. I. (2019). An afferent white matter pathway from the pulvinar to the amygdala facilitates fear recognition. *Elife* 8:e40766. doi: 10.7554/eLife.40766

Méndez, C. A., Celeghin, A., Diano, M., Orsenigo, D., Ocak, B., and Tamietto, M. (2022). A deep neural network model of the primate superior colliculus for emotion recognition. *Philos. Trans. R. Soc. B Biol. Sci.* 377:20210512. doi: 10.1098/rstb.2021.0512

Michel, M., Beck, D., Block, N., Blumenfeld, H., Brown, R., Carmel, D., et al. (2019). Opportunities and challenges for a maturing science of consciousness. *Nat. Hum. Behav.* 3, 104–107. doi: 10.1038/s41562-019-0531-8

Mikellidou, K., Kurzawski, J. W., Frijia, F., Montanaro, D., Greco, V., Burr, D. C., et al. (2017). Area prostriata in the human brain. *Curr. Biol.* 27, 3056–3060. doi: 10.1016/j.cub.2017.08.065

Milner, A. D., and Goodale, M. A. (1995). *The visual brain in action*. Oxford: Oxford Psychological Press.

Mineault, P. J., Bakhtiari, S., Richards, B. A., and Pack, C. C. (2021). Your head is there to move you around: Goal-driven models of the primate dorsal pathway. *bioRxiv* [Preprint] 34. doi: 10.1101/2021.07.09.451701

Mnih, V., Heess, N., and Graves, A. (2014). "Recurrent models of visual attention," in *Proceedings of the 27th international conference on neural information processing systems*, (Cambridge, MA), 27.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., et al. (2015). Human-level control through deep reinforcement learning. *Nature* 518, 529–533. doi: 10.1038/nature14236

Morcos, A. S., Barrett, D. G. T., Rabinowitz, N. C., and Botvinick, M. (2018). On the importance of single directions for generalization. *arXiv* [Preprint]. arXiv:1803.06959.

Nigam, S., Pojoga, S., and Dragoi, V. (2019). Synergistic coding of visual information in columnar networks. *Neuron* 104, 402–411. doi: 10.1016/j.neuron.2019.07.006

Olah, C., Mordvintsev, A., and Schubert, L. (2017). Feature visualization. *Distill* 2:e7. doi: 10.23915/distill.00007

Olshausen, B., and Field, D. (2004). Sparse coding of sensory inputs. *Curr. Opin. Neurobiol.* 14, 481–487. doi: 10.1016/j.conb.2004.07.007

Pessoa, L., and Adolphs, R. (2010). Emotion processing and the amygdala: From a 'low road' to 'many roads' of evaluating biological significance. *Nat. Rev. Neurosci.* 11, 773–782. doi: 10.1038/nrn2920

Petri, G., Musslick, S., Dey, B., Özcimder, K., Turner, D., Ahmed, N. K., et al. (2021). Topological limits to the parallel processing capability of network architectures. *Nat. Phys.* 17, 646–651. doi: 10.1038/s41567-021-01170-x

Pitcher, D., and Ungerleider, L. G. (2021). Evidence for a third visual pathway specialized for social perception. *Trends Cogn. Sci.* 25, 100–110. doi: 10.1016/j.tics.2020.11.006

Pogodin, R., Mehta, Y., Lillicrap, T., and Latham, P. E. (2021). Towards biologically plausible convolutional networks. *Adv. Neural Inform. Proc. Syst.* 34, 13924–13936.

Ponce, C. R., Xiao, W., Schade, P. F., Hartmann, T. S., Kreiman, G., and Livingstone, M. S. (2019). Evolving images for visual neurons using a deep generative network reveals coding principles and neuronal preferences. *Cell* 177, 999–1009.e10. doi: 10.1016/j.cell.2019.04.005

Pozzi, I., Bohté, S., and Roelfsema, P. (2018). A biologically plausible learning rule for deep learning in the brain. *arXiv* [Preprint]. arXiv:1811.01768.

Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C., and Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature* 435, 1102–1107. doi: 10.1038/nature03687

Rajalingham, R., Issa, E. B., Bashivan, P., Kar, K., Schmidt, K., and DiCarlo, J. J. (2018). Large-scale, high-resolution comparison of the core visual object recognition behavior of humans, monkeys, and state-of-the-art deep artificial neural networks. *J. Neurosci.* 38, 7255–7269. doi: 10.1523/jneurosci.0388-18.2018

Rawat, W., and Wang, Z. (2017). Deep convolutional neural networks for image classification: A comprehensive review. *Neural Comput.* 29, 2352–2449. doi: 10.1162/neco_a_00990

Richards, B. A., Lillicrap, T. P., Beaudoin, P., Bengio, Y., Bogacz, R., Christensen, A., et al. (2019). A deep learning framework for neuroscience. *Nat. Neurosci.* 22, 1761–1770. doi: 10.1038/s41593-019-0520-2

Riesenhuber, M., and Poggio, T. (2000). *Computational models of object recognition in cortex: A review (CBCL Paper 190/AI Memo 1695)*. Cambridge, MA: MIT Press, doi: 10.21236/ADA458109

Roelfsema, P. R., and van Ooyen, A. (2005). Attention-gated reinforcement learning of internal representations for classification. *Neural Comput.* 17, 2176–2214. doi: 10.1162/0899766054615699

Rossetti, Y., Pisella, L., and McIntosh, R. D. (2017). Rise and fall of the two visual systems theory. *Ann. Phys. Rehabil. Med.* 60, 130–140. doi: 10.1016/j.rehab.2017.02.002

Sacramento, J., Ponte Costa, R., Bengio, Y., and Senn, W. (2018). "Dendritic cortical microcircuits approximate the backpropagation algorithm," in *Advances in neural information processing systems*, eds B. Samy, W. Hanna, L. Hugo, G. Kristen, C. Nicolò, and G. Roman (New York, NY: Curran), 31.

Sahraie, A., Hibbard, P. B., Trevethan, C. T., Ritchie, K. L., and Weiskrantz, L. (2010). Consciousness of the first order in blindsight. *Proc. Natl. Acad. Sci. U.S.A.* 107, 21217–21222.

Saxe, A., Nelli, S., and Summerfield, C. (2020). If deep learning is the answer, what is the question? *Nat. Rev. Neurosci.* 22, 55–67. doi: 10.1038/s41583-020-00395-8

Scellier, B., and Bengio, Y. (2017). Equilibrium propagation: Bridging the gap between energy-based models and backpropagation. *Front. Comput. Neurosci.* 11:24. doi: 10.3389/fncom.2017.00024

Schmid, M. C., Mrowka, S. W., Turchi, J., Saunders, R. C., Wilke, M., Peters, A. J., et al. (2010). Blindsight depends on the lateral geniculate nucleus. *Nature* 466, 373–377.

Scholte, H. S., Losch, M. M., Ramakrishnan, K., de Haan, E. H., and Bohte, S. M. (2018). Visual pathways from the perspective of cost functions and multi-task deep neural networks. *Cortex* 98, 249–261. doi: 10.1016/j.cortex.2017.09.019

Schrimpf, M., Blank, I., Tuckute, G., Kauf, C., Hosseini, E. A., Kanwisher, N., et al. (2020). Artificial neural networks accurately predict language processing in the brain. *bioRxiv* [Preprint] doi: 10.1101/2020.06.26.174482

Schrimpf, M., Kubilius, J., Hong, H., Majaj, N. J., Rajalingham, R., Issa, E. B., et al. (2018). Brain-score: Which artificial neural network for object recognition is most brain-like? *bioRxiv* [Preprint]. bioRxiv 407007.

Schütz, A. C., Braun, D. I., and Gegenfurtner, K. R. (2011). Eye movements and perception: A selective review. *J. Vis.* 11:9.

Semedo, J. D., Jasper, A. I., Zandvakili, A., Krishna, A., Aschner, A., Machens, C. K., et al. (2022). Feedforward and feedback interactions between visual cortical areas use different population activity patterns. *Nat. Commun.* 13:1099. doi: 10.1038/s41467-022-28552-w

Shi, J., Wen, H., Zhang, Y., Han, K., and Liu, Z. (2018). Deep recurrent neural network reveals a hierarchy of process memory during dynamic natural vision. *Hum. Brain Mapp.* 39, 2269–2282.

Sussman, R. W. (1991). Primate origins and the evolution of angiosperms. *Am. J. Primatol.* 23, 209–223. doi: 10.1002/ajp.1350230402

Tamietto, M., and Leopold, D. A. (2018). Visual cortex: The eccentric area prostriata in the human brain. *Curr. Biol.* 28, R17–R19. doi: 10.1016/j.cub.2017.11.006

Tamietto, M., and Morrone, M. C. (2016). Visual plasticity: Blindsight bridges anatomy and function in the visual system. *Curr. Biol.* 26, R70–R73. doi: 10.1016/j.cub.2015.11.026

Tang, H., Schrimpf, M., Lotter, W., Moerman, C., Paredes, A., Ortega Caro, J., et al. (2018). Recurrent computations for visual pattern completion. *Proc. Natl. Acad. Sci. U.S.A.* 115, 8835–8840. doi: 10.1073/pnas.1719397115

Tavanaei, A., Ghodrati, M., Kheradpisheh, S. R., Masquelier, T., and Maida, A. (2019). Deep learning in spiking neural networks. *Neural Netw.* 111, 47–63. doi: 10.1016/j.neunet.2018.12.002

Ungerleider, L. G., and Mishkin, M. (1982). "Two cortical visual systems," in *Analysis of visual behavior*, eds D. J. Ingle, M. A. Goodale, and R. J. W. Mansfield (Cambridge, MA: MIT Press), 549–586.

Van den Stock, J., Tamietto, M., Hervais-Adelman, A., Pegna, A. J., and de Gelder, B. (2015). Body recognition in a patient with bilateral primary visual cortex lesions. *Biol. Psychiatry* 77, e31–e33. doi: 10.1016/j.biopsych.2013.06.023

Van Dyck, L. E., Kwitt, R., Denzler, S. J., and Gruber, W. R. (2021). Comparing object recognition in humans and deep convolutional neural networks—an eye tracking study. *Front. Neurosci.* 15:750639. doi: 10.3389/fnins.2021.750639

Vanduffel, W., and Li, X. (2020). Optogenetics: Exciting inhibition in primates. *eLife* 9:e59381. doi: 10.7554/eLife.59381

Walker, E. Y., Sinz, F. H., Cobos, E., Muhammad, T., Froudarakis, E., Fahey, P. G., et al. (2019). Inception loops discover what excites neurons most using deep predictive models. *Nat. Neurosci.* 22, 2060–2065. doi: 10.1038/s41593-019-0517-x

Warner, C. E., Kwan, W. C., and Bourne, J. A. (2012). The early maturation of visual cortical area MT is dependent on input from the retinorecipient medial portion of the inferior pulvinar. *J. Neurosci.* 32, 17073–17085. doi: 10.1523/JNEUROSCI.3269-12.2012

Warner, C. E., Kwan, W. C., Wright, D., Johnston, L. A., Egan, G. F., and Bourne, J. A. (2015). Preservation of vision by the pulvinar following early-life primary visual cortex lesions. *Curr. Biol.* 25, 424–434. doi: 10.1016/j.cub.2014.12.028

Whittington, J. C., and Bogacz, R. (2017). An approximation of the error backpropagation algorithm in a predictive coding network with local hebbian synaptic plasticity. *Neural Comput.* 29, 1229–1262. doi: 10.1162/NECO_a_00949

Wichmann, F. A., Janssen, D. H. J., Geirhos, R., Aguilar, G., Schütt, H. H., Maertens, M., et al. (2017). Methods and measurements to compare men against machines. *Electron. Imaging* 29, 36–45. doi: 10.2352/issn.2470-1173.2017.14.hvei-113

Williams, L. E., and Holtmaat, A. (2019). Higher-order thalamocortical inputs gate synaptic long-term potentiation via disinhibition. *Neuron* 101, 91–102. doi: 10.1016/j.neuron.2018.10.049

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., et al. (2020). "Transformers: State-of-the-art natural language processing," in *Proceedings of the 2020 conference on empirical methods in natural language processing: System demonstrations*, (Miami, FL), doi: 10.18653/v1/2020.emnlp-demos.6

Wurtz, R. H. (2015). Using perturbations to identify the brain circuits underlying active vision. *Philos. Trans. R. Soc. B Biol. Sci.* 370:20140205. doi: 10.1098/rstb.2014.0205

Yagishita, S., Hayashi-Takagi, A., Ellis-Davies, G. C., Urakubo, H., Ishii, S., and Kasai, H. (2014). A critical time window for dopamine actions on the structural plasticity of dendritic spines. *Science* 345, 1616–1620.

Yamins, D. L. K., and DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nat. Neurosci.* 19, 356–365. doi: 10.1038/nn.4244

Yamins, D. L. K., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., and DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proc. Natl. Acad. Sci. U.S.A.* 111, 8619–8624. doi: 10.1073/pnas.1403112111

Zhou, B., Bau, D., Oliva, A., and Torralba, A. (2018). Interpreting visual representations of neural networks via network dissection. *J. Vis.* 18:1244. doi: 10.1167/18.10.1244