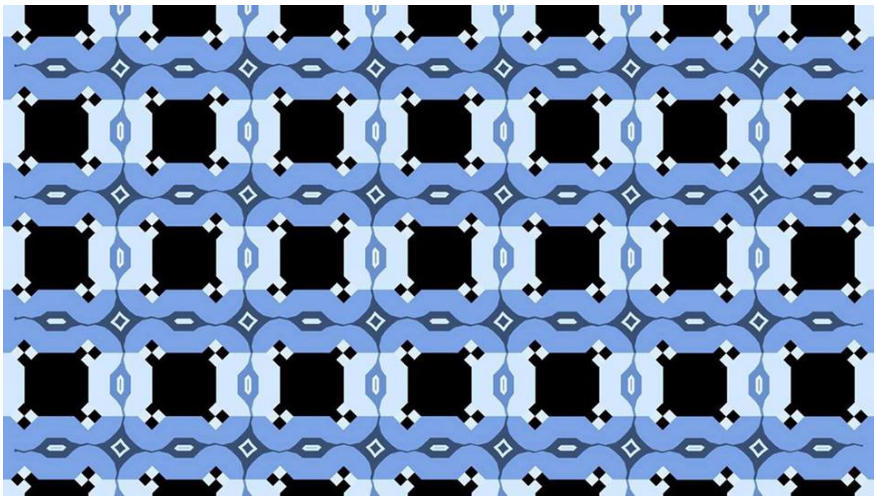
How Human and Deep Learning Perception are Very Different

Carlos E. Perez Follow

Apr 9, 2018 · 8 min read



https://twitter.com/victoria1skye

How do we perceive the world? To understand this, let us explore how we incorrectly perceive the world. The 'glitches in the matrix' shall reveal to us the nature of our perception.

Victoria Syke created the above optical illusion that works astonishingly well to confuse our perception. The illusion here is that **the dark blue lines run parallel to each other.** You can prove this to yourself by either scrolling the image so that it aligns with the top of the browser window or you can look at the image from one of the edges.

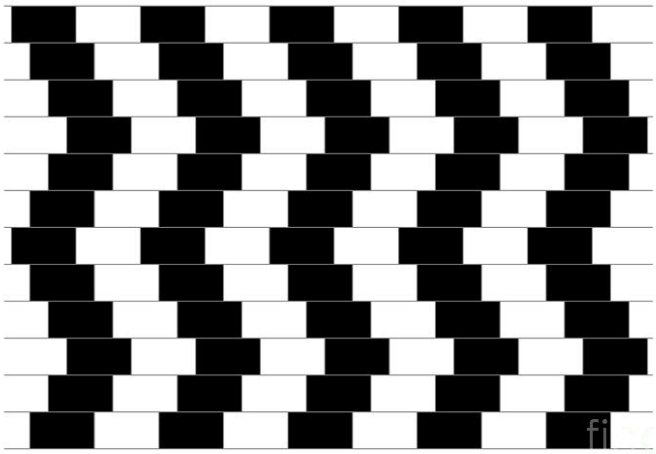
What I want to know is, why is this illusion so effective? What in our own cognitive processes is creating the confusion?

In the illusion above, you will notice that each element in the band with alternating light blue and black boxes appears smaller in size in a specific direction. In addition to this, you will also notice that the image in the dark blue band has lines that that are of a different height than the previous one. These two illusions combine with each other to give the illusion that a band in continuously trending upwards or downwards.

The light blue boxes do appear parallel even if you rotate the image by 90 degrees. That is because the dark blue boxes always appear the same size and the lines inside them are also at the same level.

Victoria Syke was inspired to create this image from two sources. Richard Gregory's observation of the Cafe Wall Illusion and Akiyoshi Kitaoka's Fringe Edge Illusion.

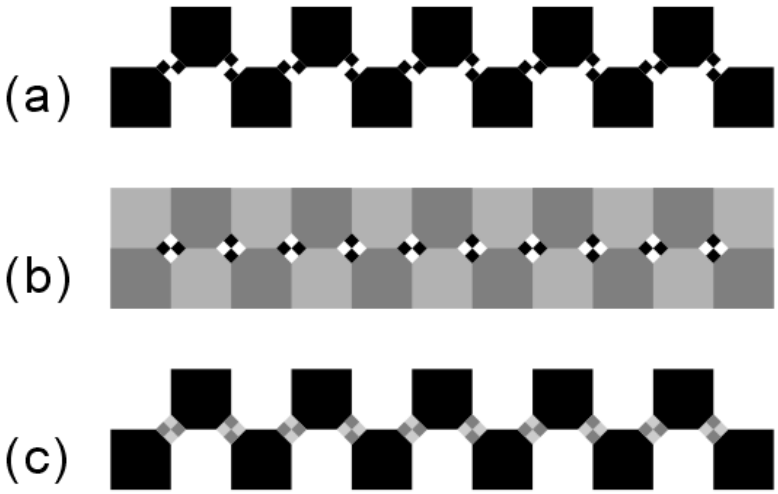
The Cafe Wall illusion's effect reveals itself when the luminance of the mortar between the bricks have a luminance value between black and white:



https://fineartamerica.com/featured/cafe-wall-illusion-spl-and-photo-researchers.html

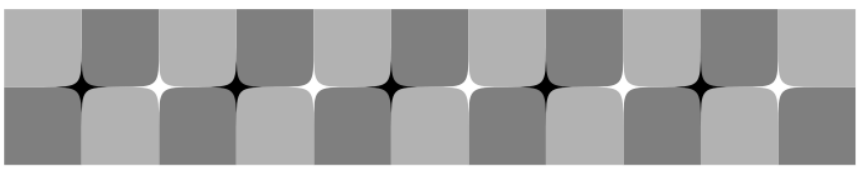
The effect of which is that each brick appears to be progressively bigger (or smaller) than the brick that is adjacent to it.

Syke also leveraged Akiyoshi Kitaoka's Fringe Edge illusion



http://www.psy.ritsumei.ac.jp/~akitaoka/fringede.html

and the Y-junction illusion:



http://www.psy.ritsumei.ac.jp/~akitaoka/Yjunctione.html

to enhance the effect even further. BTW, the effect also works in the vertical direction.

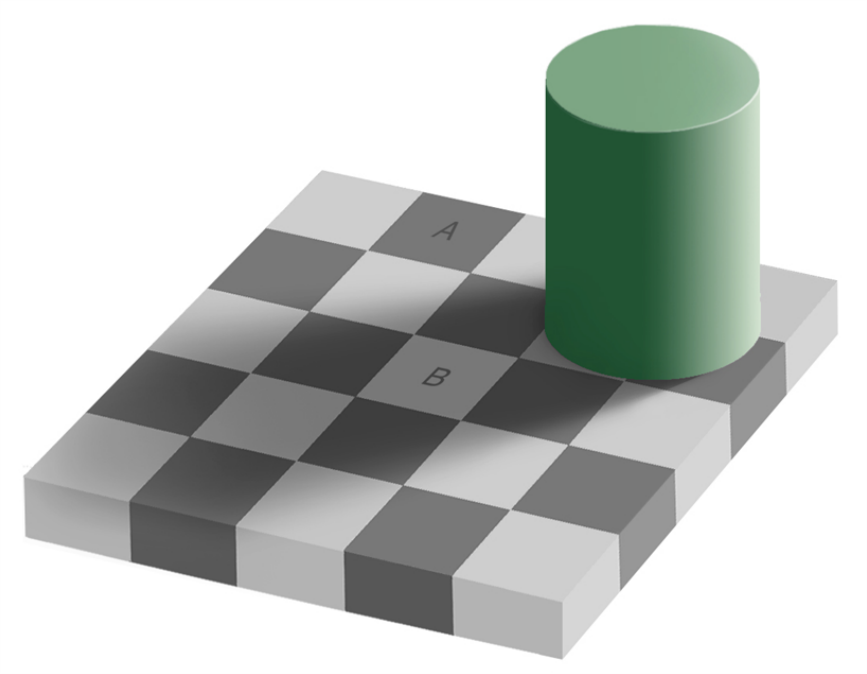
The mind apparently does not see an image as a whole. Rather it sees the image as a composition of images and recognizing the adjacent relationships of one to another. Why do adjacent relationships have such a strong effect on our visual perception? We have evolved to take advantage of affordances to allow our brain to reconstruct images more quickly. Said differently, our brains immediately recognizes patterns that facilitates our interpretation of a scene. Our visual perception performs a kind of semantic inference automatically such that higher level semantic patterns can't be ignored. That is why an illusion like this cannot be "unseen" no matter how much we convince ourselves that the lines are indeed parallel.

This effect also occurs in the audio domain. There is an auditory illusion known as the Shepard Tone. The illusion is created by having three tones that are ascending. A higher tone that becomes quieter, a middle tone that has constant loudness and a lower tone that becomes louder. The brain is tricked by hearing two tones that are always ascending. This is best illustrated in this video (Start at 0:40):

The illusions in the image and auditory regimes reveals to us insights on how the mind perceives its world. Our minds sees images and sounds relative to each other and makes an imagined prediction of a progression even when that progression does not exist. The mind cannot override the affordances it sees and therefore proceeds with an incorrect reconstruction. You can look at the image above but you cannot unsee the lines that are tilting. If you look at the image at a distance or look at it at an angle, you see the image without the affordances and see thus reconstruct reality correctly.

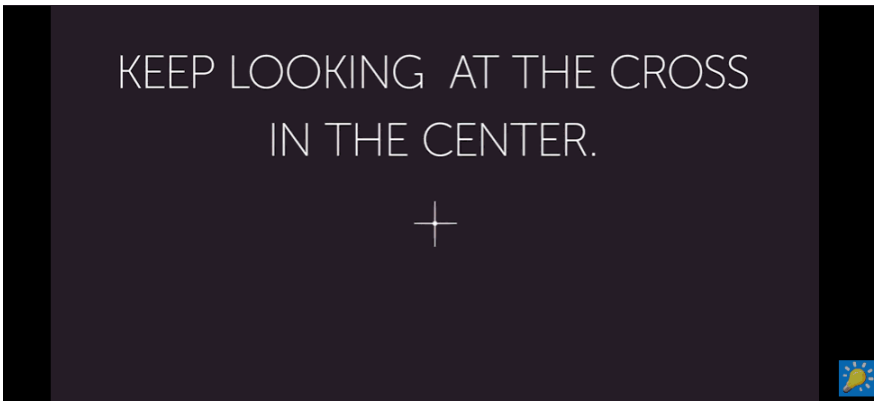
But why is relative size important to our biology? We can learn from art as to what elements lead to a perception of depth: overlapping objects, diminishing scale, atmospheric perspective, vertical placement and linear perspective. The brain makes use of these affordances to reconstruct a 3D representation of the world. We are embodied in a 3D world and our senses are designed to comprehend and interact with

that world. Affordances that are clues to the 3D structure of objects are the sources of optical illusions. The checker board shadow illusion is one of the more well known examples of this:



A and B are of the same shade.

Here is another illusion that illustrates how the brain must be given sufficient time to correctly re-construct its perception:

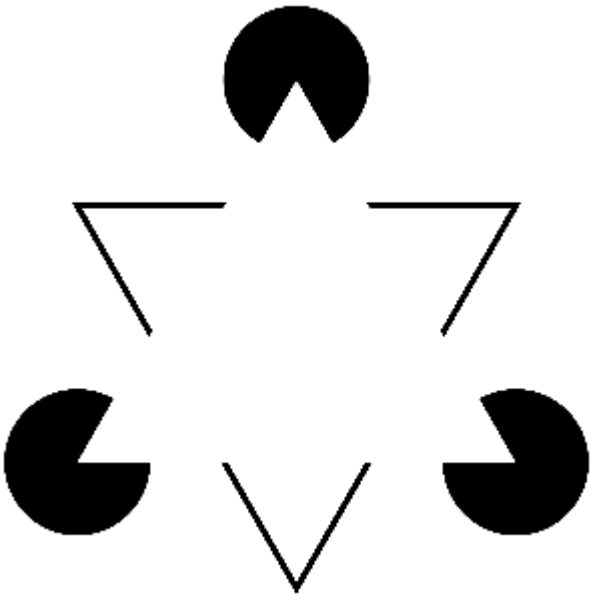


Click to make enlarge. Source: https://www.youtube.com/watch?v=LcpliVYfEqk

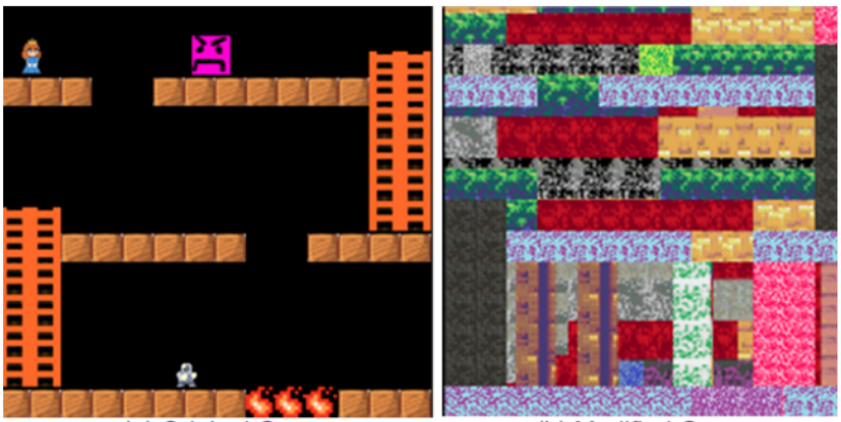
In the above experiment, as you focus on the center, you will notice that the faces in your periphery become distorted. The images are flashed fast enough that our brain see the cross talk between the two images and is not fast enough reconstruct it correctly.

Unlike Deep Learning networks that actually do capture images in its entirety, biological brains will use affordances (i.e. shortcuts and heuristics) to construct patterns that it will use for perception. Deep Learning networks are trained specifically using networks that ignore certain invariances (i.e. translation for ConvNets). Biological brains appear to work differently, rather than ignore invariances we are hardwired to make use of patterns that convey semantics. DL networks are not trained to identify affordances that lead to pattern identification

that leads to semantic interpretation. To achieve the kind of visual perception we find in humans, we must train networks to learn some basic human image recognition skills such as occlusion, perspective, and shadows:



To illustrate how very different a Deep Learning system's visual cognition is from that of humans, a recent paper "Investigating Human Priors for Playing Video Games" investigates removing human affordances for playing a game:



(a) Original Game (b) Modified Game

Arcade games were modified to re-render the game's textures. In the modified game, humans performed extremely poorly. In contrast, a Deep Learning system performed equivalently for both games. Deep Learning systems do not need to use human priors. On the flip side, a human can learn a game with less trials because we can exploit the use of existing human priors (or affordances). What this should tell you is that humans learn quickly by using our existing priors.

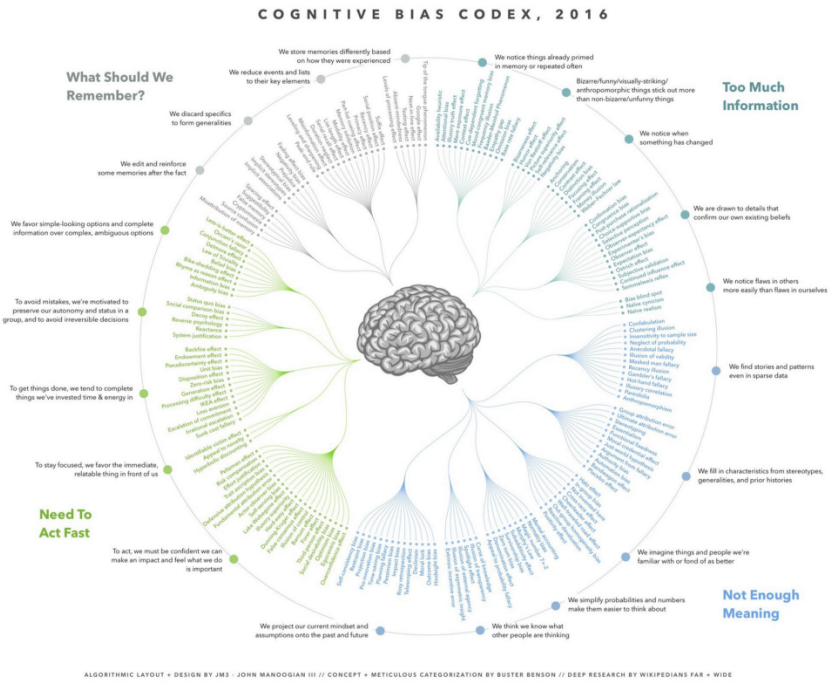
DeepMind's Pyschlab is a setup to explore the difference between Deep Learning and Human visual recognition. Psychlab contains many experiments that a human and a machine can perform. By examining

the difference in performance, we can learn the cognitive differences between the two systems. In general, its observed that humans employ a mix of parallel and sequential processing. This can be discerned from a slowdown in the performance of tasks as compared to a machine that employs only parallel processing:

> *In humans, this data has suggested a difference between **parallel and serial attention**. Agents appear only to have parallel mechanisms. Identifying this difference between humans and our current artificial agents shows a path toward improving future agent designs.*

Another paper from DeepMind published in BioArxiv "Prefrontal cortex as a meta-reinforcement learning system" proposes that the brain uses two different reinforcement learning systems. Reinforcement learning in biological brains are postulated to be driven by dopamine releases. This is the standard model of reward driven learning. DeepMind's proposal is that there are two RL system, one RL system is based on the standard dopamine model and a second RL system is found in the prefrontal cortex. The prefrontal cortex learning is influenced by the first system. Effectively, the standard dopamine model has learned human priors (or affordances) and employs this to guide the more dynamic learning of the prefrontal cortex.

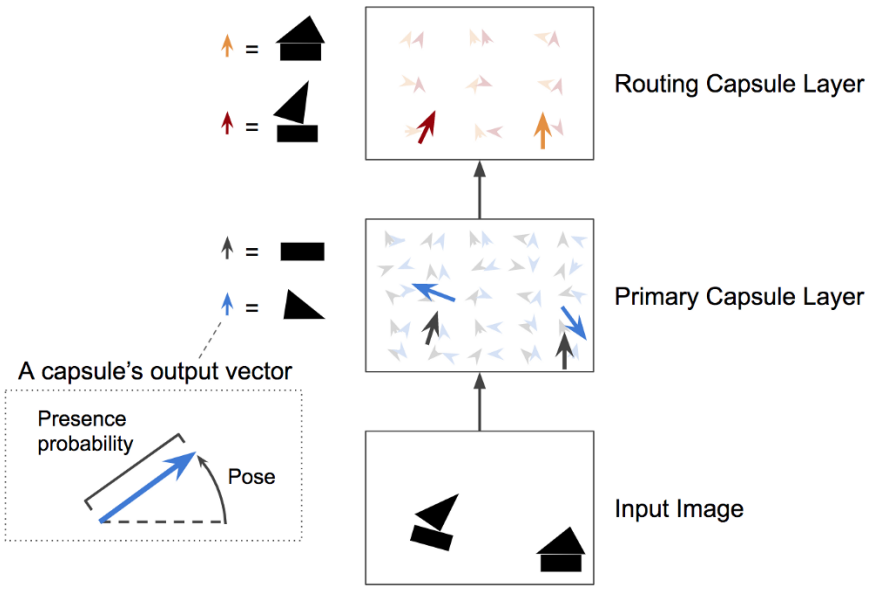
So whenever we see something, we can only see it with human priors engaged. However as you can see in the example of the faces, there is a cognitive process at work that attempts to reconstruct what it sees. Stop that reconstruction process too quickly and you see how it can incorrectly make errors. Our brain employs heuristics all the time and we find that these heuristics can fail in many ways.



https://betterhumans.coach.me/cognitive-bias-cheat-sheet-55a472476b18 Buster Benson

Geoffrey Hinton may be on the right track with his Capsule Network. In Capsule networks, there are two important stages. A first stage that is

able to recognize parts of objects using a ConvNet and then a second stage that votes on which composition of recognized objects is the most likely one that is perceived. This two stage process, one of object recognition and then followed by inference seems to be gaining traction in the research community.



Watch this excellent video explaining Capsule Nets: https://www.youtube.com/watch?v=pPN8d0E3900

In the 1980's a new field emerged out of the advances of supercomputers, this was known as computational science and it differed to the existing approaches to science (i.e. Theoretical and Experimental). Computational science explored physical systems through simulation by computer. In the same way, research in Deep Learning is now encroaching into the fields of neuroscience and psychology. That is, we are beginning to understand our own nature as we compare our simulations with ourselves.

In summary, the emerging research trend in Deep Learning is to begin to dig deeper into the precise nature of human perception and to identify how it differs from Deep Learning perception. From the perspective of a Deep Learning researcher, it is not enough to understand the mathematics and the technology, but one must have some familiarity with the characteristics of basic human perception. It is well established that adversarial features are problematic for Deep Learning. To solve problems like this, we need to understand why it isn't a problem for humans. This is indeed exactly what Geoffrey Hinton argued about in his lecture about "What's wrong with Convolutional Networks."

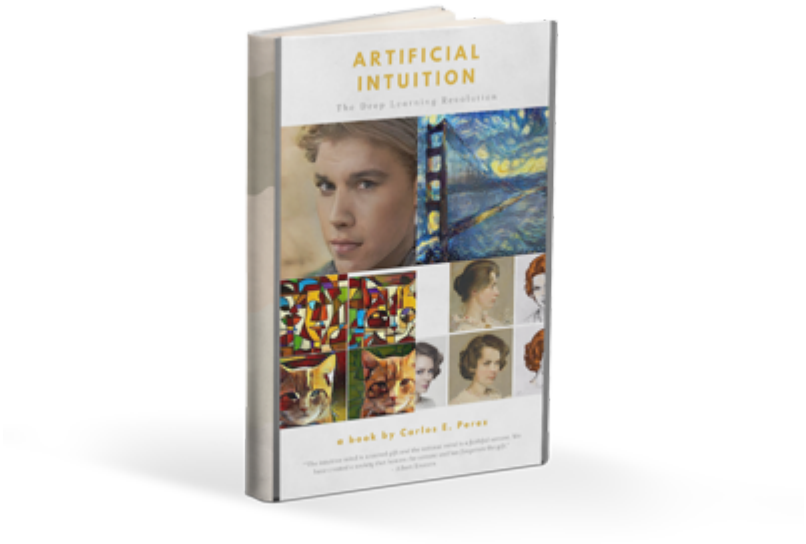
Further Reading

[1808.08750v1] Generalisation in humans and deep neural networks

Abstract: We compare the robustness of humans

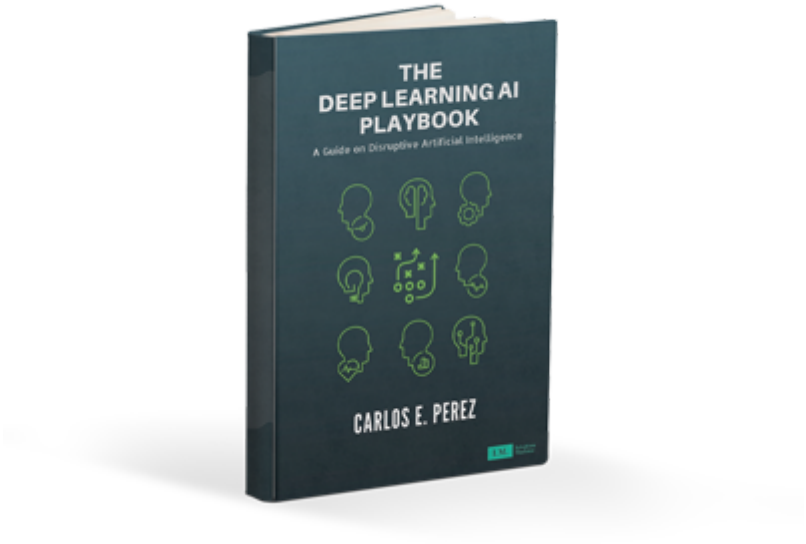
and current convolutional deep neural networks...

arxiv.org



Explore Deep Learning: Artificial Intuition: The Improbable Deep Learning Revolution

.



Exploit Deep Learning: The Deep Learning AI Playbook

Landmark Discovery for Image Modeling - Yuting Zhang's Homepage

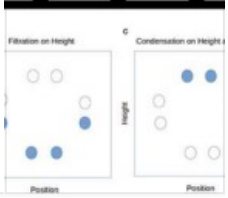
Deep neural networks can model images with rich latent representations, but they cannot naturally...

www.ytzhang.net

Attentional Bias in Human Category Learning: The Case of Deep Learning

Category learning performance is influenced by both the nature of the category's structure and t...

www.frontiersin.org



[1805.10734v1] A neural network trained to predict future video frames mimics critical...

Abstract: While deep neural networks take loose inspiration from neuroscience, it is an open...

arxiv.org

[1805.12177] Why do deep convolutional networks generalize so poorly to small image...

Abstract: Deep convolutional network architectures are often assumed to guarantee...

arxiv.org

Flash-Lag Effect

Demonstration of 'Flash-Lag'

www.michaelbach.de

Illusory Motion Reproduced by Deep Neural Networks Trained for Prediction

The cerebral cortex predicts visual motion to adapt human behaviour to surrounding objects moving...

www.frontiersin.org