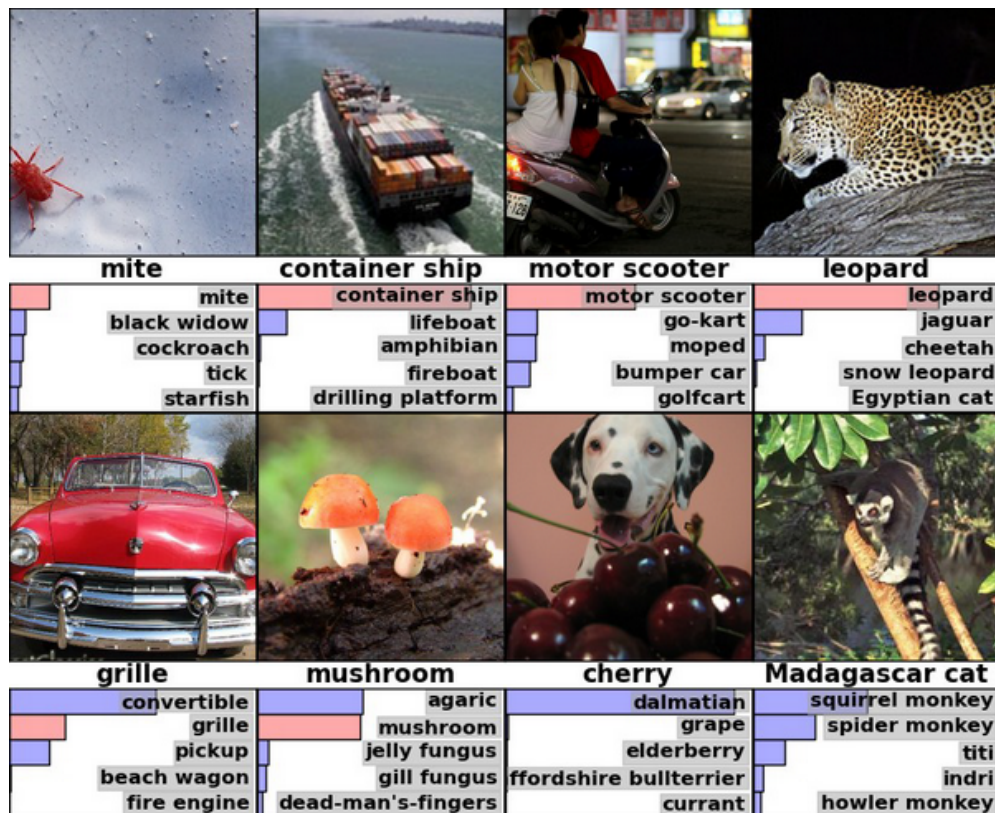


Suddenly, a leopard print sofa appears

If you have been around all the machine learning and artificial intelligence stuff, you surely have already seen this:



Or, if you haven't, there are some deep convolutional network result samples from ILSVRC2010, by Hinton and Krizhevsky

Let's look for a moment at the top-right picture. There's a leopard, recognized with substantial confidence, and then two much less probable choices are jaguar and cheetah.

And this is, if you think about it for a bit, kinda cool. Do *you* know how to tell apart those three big and spotty kitties? Because I totally don't. There must be differences, of course — maybe something subtle and specific, that only a skilled zoologist can perceive, like general body shape or jaw size, or tail length — or maybe is it context/background, because leopards inhabit forests and are more likely to be found laying on a tree, when cheetahs live in savanna? Either way, for a machine learning algorithm, this looks very impressive to me. After all, we're still facing lots of really annoying and foolish errors like [this one](#). So, is that the famous deep learning approach? Are we going to meet human-like machine intelligence soon?

Well... turns out, maybe not so fast.

Just a little zoological fact

Let's take a closer look at these three kinds of big cats again. Here's the jaguar, for example:



It's the biggest cat on both Americas, which also has a curious habit of killing its prey by puncturing their skull and brain (that's not really the little fact we're looking for). It's the most massive cat in comparison with leopard and cheetah, and its other distinguishing features are dark eyes and larger jaw. Well, that actually looks pretty fine-grained.

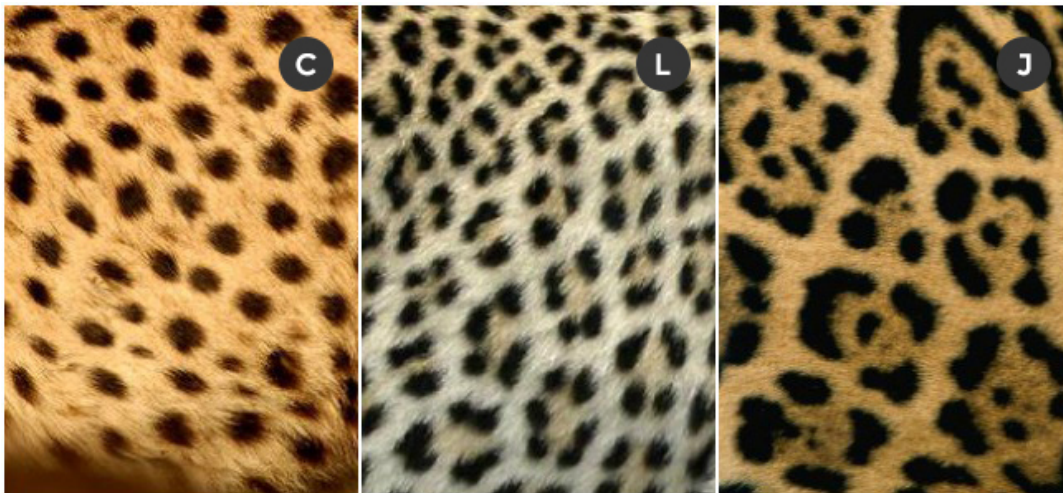


Then, the leopard. It's a bit smaller than jaguar and generally more elegant, considering, for example, its smaller paws and jaw. And also yellow eyes. Cute.



And the smallest of the pack, the cheetah, that actually looks quite different from the previous two. Has a generally smaller, long and slim body, and a distinctive face pattern that looks like two black tear trails running from the corners of its eyes.

And now for the part I've purposely left out: black spotty print pattern. It's not completely random, as you might think it is — rather, black spots are combined into small groups called “rosettes”. You can see that jaguar rosettes are large, distinctive and contain a small black spot inside, while leopard rosettes are significantly smaller. As for the cheetah, its print doesn't contain any, just a scatter of pure black spots.



See how those three prints actually differ (also, thanks [Imgur](#) for educating me and providing the pictures).

Suspicion grows

Now, I have a little bit of bad feeling about it. What if this is *the only thing* our algorithm does — just treating these three pictures like shapeless pieces of texture, knowing nothing about leopard's jaw or paws, its body structure at all? Let's test this hypothesis by running a pre-trained convolutional network on a very simple test image. We're not trying to apply any visual noise, artificial occlusion or any other tricks to mess with image recognition — that's just a simple image, which I'm sure everyone who reads this page will recognize instantly.

Here it is:



We're going to use [Caffe](#) and its pre-trained CaffeNet model, which is actually different from Hinton and Krizhevsky's AlexNet, but the principle is the same, so it will do just fine. Aaand here we go:

```
import numpy as np
import matplotlib.pyplot as plt

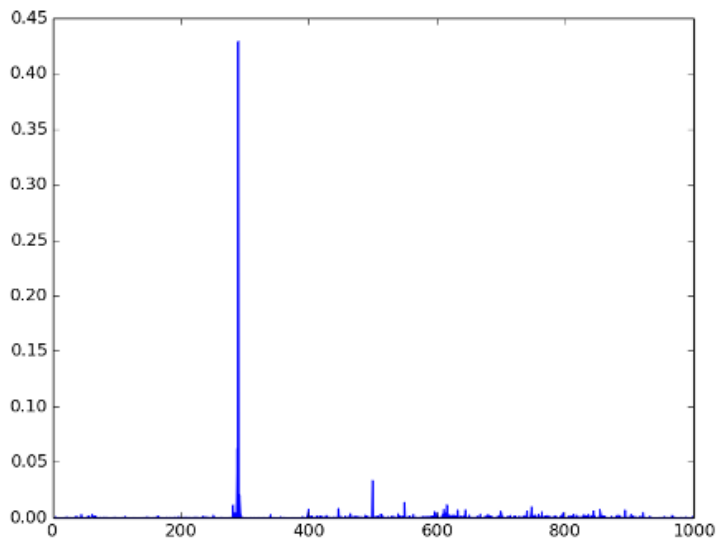
caffe_root = '../'
import sys
sys.path.insert(0, caffe_root + 'python')

import caffe

MODEL_FILE = '../models/bvlc_reference_caffenet/deploy.prototxt'
PRETRAINED = '../models/bvlc_reference_caffenet/bvlc_reference_caffenet.caffemodel'
IMAGE_FILE = '../sofa.jpg'

caffe.set_mode_cpu()
net = caffe.Classifier(MODEL_FILE, PRETRAINED,
                      mean=np.load(caffe_root + 'python/caffe/imagenet/ilsvrc_2012_mean.npy').mean(1).mean(1),
                      channel_swap=(2, 1, 0),
                      raw_scale=255,
                      image_dims=(500, 500))
input_image = caffe.io.load_image(IMAGE_FILE)
prediction = net.predict([input_image])
plt.plot(prediction[0])
print 'predicted class:', prediction[0].argmax()
plt.show()
```

Here's the result:



>> predicted class: 290

```

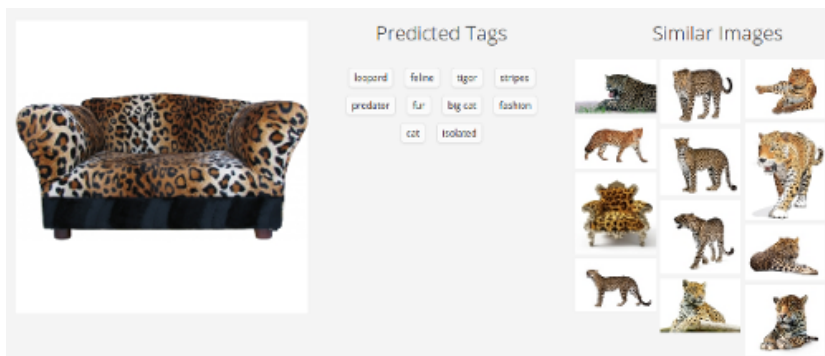
288 n02128385 leopard, Panthera pardus
289 n02128757 snow leopard, ounce, Panthera uncia
290 n02128925 jaguar, panther, Panthera onca, Felis onca
291 n02129165 lion, king of beasts, Panthera leo
292 n02129604 tiger, Panthera tigris

```

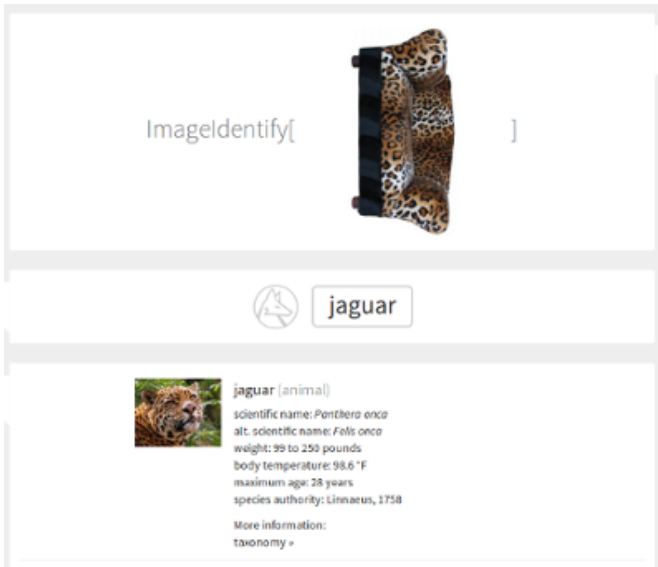
Whoops.

But wait, maybe that's just CaffeNet thing? Let's check something third-party:

[Clarifai](#) (those guys did great on the latest ImageNet challenge)



[Brand new Stephen Wolfram's ImageIdentify](#)



Okay, I cheated a bit: on the last picture the sofa is rotated by 90 degrees, but that's really simple transformation that should not change the recognition output so radically. I've also tried [Microsoft](#) and [Google](#) services and nothing have beaten rotated leopard print sofa. Interesting result, considering all the “*{Somebody}'s Deep Learning Project Outperforms Humans In Image Recognition*” headlines that's been around for a while now.

Why is this happening?

Now, here's a guess. Imagine a simple supervised classifier, without going into model specifics, that accepts a bunch of labeled images and tries to extract some inner structure (a set of features) from that dataset to use for recognition. During the learning process, a classifier adjust its parameters using prediction/recognition error, and here's when dataset size and structure matter. For example, if a dataset contains 99 leopards and only one sofa, the simplest rule that tells a classifier to always output “leopard” will result in 1% recognition error while staying not intelligent at all.

And that seems to be exactly the case, both for our own visual experience and for ImageNet dataset. Leopard sofas are rare things. There simply aren't enough of them to make difference for a classifier; and black spot texture makes a very distinctive pattern that is otherwise specific to a leopard category. Moreover, being faced with *different* classes of big spotted cats, a classifier can benefit from using these texture patterns, since they provide simple distinguishing features (compared with the others like the size of the jaw). So, our algorithm works just like it's supposed to. Different spots make different features, there's little confusion with other categories and sofa example is just an anomaly. Adding enough sofas to the dataset will surely help (and then the size of the jaw will matter more, I guess), so there's no problem at all, it's just how learning works.

Or is it?

What we humans do

Remember your first school year, when you learned digits in your math class.

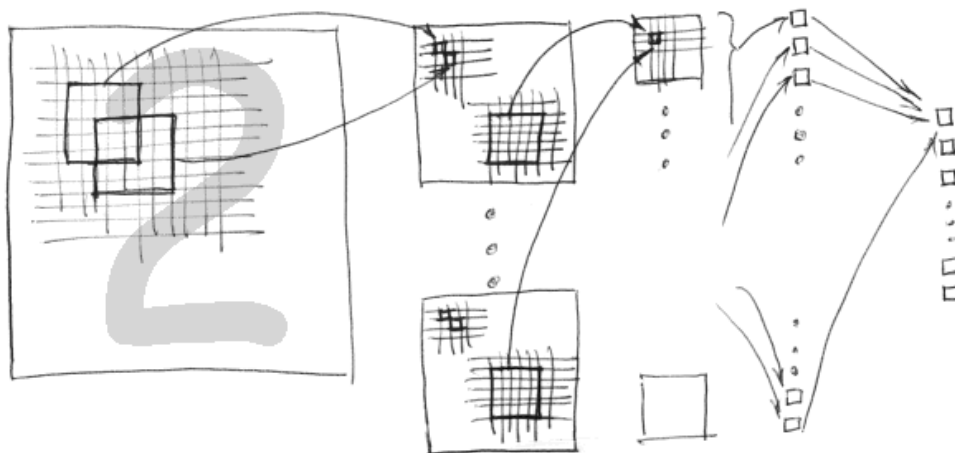
When each student was given a heavy book of MNIST database, hundreds of pages filled with endless hand-written digit series, 60000 total, written in different styles, bold or italic, distinctly or sketchy. The best students were also given an appendix, “Permutation MNIST”, that contained the same digits, but transformed in lots of different ways: rotated, scaled up and down, mirrored and skewed. And you had to scan through *all* of them to pass a math test, where you had to recognize just a small subset of length 10000. And just when you thought the nightmare was over, a language class began, featuring not ten recognition categories, but twenty-five instead.

So, are you going to say that was not the case?

It's an interesting thing: looks like we don't really need a huge dataset to learn something new. We perceive digits as abstract concepts, Plato's ideal forms, or actually rather a spatial combinations of ones, like “a straight line”, “a circle”, “an angle”. If an image contains two small circles placed one above the other, we recognize an eight; but when none of the digit-specific elements are present, we consider the image to be not a digit at all. This is something a supervised classifier never does — instead, it tries to put the image into the closest category, even if likeness is negligible.

Maybe MNIST digits is not a good example — after all, we all have seen a lot of them in school, maybe enough for a huge dataset. Let's get back to our leopard print sofa. Have you seen a lot of leopards in your life? Maybe, but I'm almost sure that you've seen "faces" or "computers" or "hands" a lot more often. Have you actually seen such a sofa before — even once? Can't be one hundred percent confident for myself, but I think I have not. And nevertheless, despite this total lack of visual experience, I don't consider the image above a spotty cat in a slightest bit.

Convolutional networks make it worse



Deep convolutional networks are long-time ImageNet champions. No wonder; they are designed to process images, after all. If you are not familiar with the concept of CNNs, here's a quick reminder: they are locally-connected networks that use a set of small filters as local feature detectors, convolving them across the entire image, which makes these features translation-invariant (which is often a desired property). This is also a lot cheaper than trying to put an entire image (represented by $1024 \times 768 \approx 800,000$ naive pixel features) into a fully-connected network. There are other operations involved in CNNs feed-forward propagation step, such as subsampling or pooling, but let's focus on convolution step for now.

Leopards (or jaguars) are complex 3-dimensional shapes with quite a lot of degrees of freedom (considering all the body parts that can move independently). These shapes can produce a lot of different 2d contours projected on the camera sensor: sometimes you can see a distinct silhouette featuring a face and full set of paws, and sometimes it's just a back and a curled tail. Such complex objects can be handled by a CNN very efficiently by using a simple rule: "take all these little spotty-pattern features and collect as many matches as possible from the entire image". CNNs local filters ignore the problem of having different 2d shapes by not trying to analyze leopard's spatial structure at all — they just look for black spots, and, thanks to nature, there are a lot of them at any leopard picture. The good thing here is that we don't have to care about object's pose and orientation, and the bad thing is that, well, we are now vulnerable to some specific kinds of sofas.

And this is really not good. CNN's usage of local features allows to achieve transformation invariance — but this comes with the price of not knowing neither object structure nor its orientation. CNN cannot distinguish between a cat sitting on the floor and a cat sitting on the ceiling upside down, which might be good for Google image search but for any other application involving interactions with actual cats it's kinda not.

If that doesn't look convincing, take a look at [Hinton's paper](#) from 2011 where he says that convolutional networks are doomed precisely because of the same reason. The rest of the paper is about an alternative approach, his [capsule theory](#), which is definitely worth reading too.

We're doing it wrong

Maybe not all wrong, and of course, convolutional networks are extremely useful things, but think about it: sometimes it almost looks like we're already there. We're using huge datasets like ImageNet, organize competitions and challenges, where we, for example, have decreased MNIST recognition error rate from 0.87 to 0.23 ([in three years](#)) — considering that no one really knows what error rate a human brain can achieve. There's a lot of talk about GPU implementations — like it's just a matter of computational power now, and the theory is all fine. It's not. And the problem won't be solved by collecting even larger datasets and using more GPUs, because leopard print sofas are inevitable. There always going to be an anomaly; lots of them, actually, considering all the things painted in different patterns. Something has to change. Good recognition algorithms have to understand the structure of the image and to be able to find its elements like paws or face or tail, despite the issues of projection and occlusion.

So I guess, there's still a lot of work to be done.



Join the discussion...

LOG IN WITH

OR SIGN UP WITH DISQUS 

**Terrence Andrew Davis** • 3 years ago

Yer a nigger. The nigger wants me to put Markov Artificial Intelligence onto my SETI. Just shoot me! These niggers are knuckle-dragging retards.

12 ^ | ▾ • Reply • Share ›

**Trolls** → Terrence Andrew Davis • 3 years ago

do you have schizophrenia, loser?

1 ^ | ▾ • Reply • Share ›

**Morley** → Trolls • 3 years ago

He actually does. Google for TempleOS.

18 ^ | ▾ • Reply • Share ›

**Trolls** → Morley • 3 years ago

haha. this schizo thinks random numbers are god talking to him, that's gold. you would think with so much time on his hands, designing his retard operating system and all, he would understand how they are generated. called this a loon mile away.

1 ^ | ▾ • Reply • Share ›

**Vidar Hokstad** → Trolls • 3 years ago

Apart from your distasteful approach to dealing with the mentally ill, I find it amusing that you find the idea of random numbers as a channel from a deity to be something worth singling out.

As an atheist, if I were to try to come up with a hypothesis for a potential deity that is communicating with its creation yet has somehow evaded the collection of more obvious evidence, then subtle manipulation of "random" (and you'll note if you look into it that while there may - quite likely - be issues with the amount of entropy in the random number generation Terry uses for this, his "words from God" are derived from use of timers rather than statically predictable pseudo-random sequences) numbers would be an approach that would be at least potentially viable (assuming said communications were kept sufficiently unclear and woolly to be possible to write off as nonsensical).

I have met plenty of people without mental illness with far more unreasonable religious beliefs.

2 ^ | ▾ • Reply • Share ›

**Trolls** → Vidar Hokstad • 3 years ago

then you're a fucking stupid atheist, learn what a seed is. and i give a fuck about your feelings.

^ | ▾ • Reply • Share ›

**Matt Hebert** → Trolls • 3 years ago

There are far worse ideas than God communicating to someone with random numbers. (Not that I agree with the idea in any literal sense.)

Even if the random numbers produced by a seed are pseudorandom, is the seed itself pseudorandom or random?

Suddenly we're talking about free-will and determinism. What about true random numbers? Are there true random numbers? If they do exist, wouldn't true random numbers affect the selection of the seed, since they would effect the computer and the behavior of the person using it? What causes random behavior? What about the algorithm used to generate the psuedorandom numbers? Was it destined? If not, than the whole system, while patterned, is still a product of randomness, amplified.

I'm also fully aware you're trolling, but I felt like replying anyway.

^ | v • Reply • Share ›



Trolls → Matt Hebert • 3 years ago

all of your answers are in an intro level cs course, mr hebert. perhaps you should spend time there instead of Disqus

^ | v • Reply • Share ›



Stephen Stucky → Trolls • 3 years ago

Its amusing that you don't know that pre-mental illness this guy was at the literal top if his field. But its okay, Its not every day you get shown up by someone suffering a mental illness.

^ | v • Reply • Share ›



Ellie Kesselman → Trolls • 3 years ago

He does. This explains his particular situation in more detail: <http://motherboard.vice.com...>

Thank you, Morley.

2 ^ | v • Reply • Share ›



kek → Terrence Andrew Davis • 3 years ago

nice

1 ^ | v • Reply • Share ›



Dave Eckblad → Terrence Andrew Davis • 3 years ago

You're the best example of a horrible fake-as-fuck "Christian." Get lost ya loser. P.S. your OS is as useless as the Bible.

3 ^ | v • Reply • Share ›



Arthur Bugorski → Dave Eckblad • 3 years ago

You are picking on a mentally ill man. The classy thing to do is understand its the disease speaking and not him and hope he gets better. Be the bigger person, make the world a better place.

20 ^ | v • Reply • Share ›



Stephen Stucky → Dave Eckblad • 3 years ago

A) The God he is referring to is his Computer. He is convinced the random messages his computer makes are messages from a higher power.

B) He is obviously mentally ill, fuck off.

3 ^ | v • Reply • Share ›



k9s → Dave Eckblad • 3 years ago

LOL, I get a kick about of the idea that you sit around and complain about the religious being hate mongers. Meanwhile you are the one picking on the disabled kid. The irony is not lost on me.

1 ^ | v • Reply • Share ›



idonthaterryijustgrowweary • 3 years ago

Dammit, even when I leave HN Terry is here.

11 ^ | v • Reply • Share ›



Terrence Andrew Davis • 3 years ago



What kind of nigger would shape a signal from SETI with a Markov engine to improve the signal?

3 ^ | v • Reply • Share ›



phaed → Terrence Andrew Davis • 3 years ago

An atheist one.

2 ^ | v • Reply • Share ›



Alex Beschler • 3 years ago

Wow, this thread escalated quickly...



2 ^ | v • Reply • Share ›



Steven Fox • 3 years ago

I don't think the idea of CNN's need to be thrown out, our visual system is quite similar, the difference is we look for large patterns first rather than kernels & weight them differently.

We look for large patterns first and only use smaller details to distinguish closer classifications.

Humans also develop language, visual recognition and their conceptual ontologies simultaneously which is another way of closing this gap.

One thing that's become clear to me in the last number of months is that humans do learn via large training sets. My 16 month old son has seen enough dogs and sheep to differentiate them, but sometimes confused a lamb with a dog (due to their less differentiated shape)

He can distinguish between tigers and zebras, even if the tiger is depicted as black&white.

What I'm not sure of yet is his recognition of the similarities between lions, tigers and leopards & that these are big cats, but he does seem to recognise female lions, so the mane isn't his only classification feature.

2 ^ | v • Reply • Share ›



Alfonso de la Osa → Steven Fox • 3 years ago

This comment is off topic. We were talking about Terry here. :D

^ | v • Reply • Share ›



Steven Fox → Alfonso de la Osa • 3 years ago

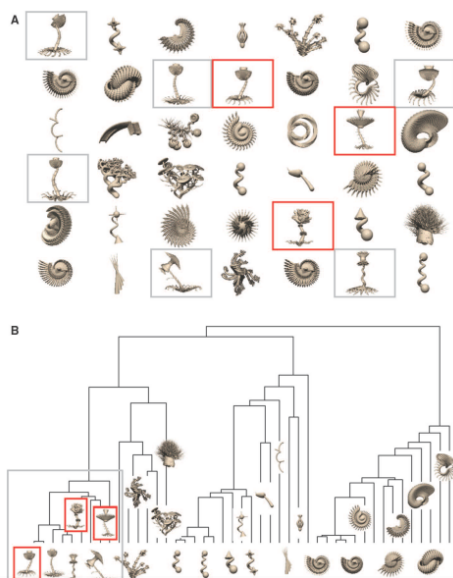
:-) I did wonder!!

^ | v • Reply • Share ›



anglrphish Mod → Steven Fox • 3 years ago

Please take a look at this brilliant illustration that was posted on HN:



[see more](#)

2 ^ | v • Reply • Share ›



Lexi • 3 years ago

Does the same thing happen with more common everyday items such as fashionable garments?



^ | v • Reply • Share ›



anglrphish Mod → Lexi • 3 years ago

This picture is recognized correctly both by Imageidentify and Clarifai. Maybe because of the human face part?

^ | v • Reply • Share ›



James Homer → Lexi • 3 years ago

This is definitely a problem area. I've analysed this as not fashionable.

1 ^ | v • Reply • Share ›

**Sigi Kiermayer** • 3 years ago

I had the same thoughts as well. A few weeks/months ago there was an article which talked about a picture which looks like color noise which was calculated in a way that it would change the output of a neural network. The image looked after the added noise image the same (for a human) but not for the neuronal network.

There was also a presentation of different projects which analyse different types of data: Image, Audio, Text. And i was wondering when the combination of all of them will happen. I think thats the next step. Give the image classifier a logic classifier etc.

^ | v • Reply • Share ›

**Sam Harrington** → Sigi Kiermayer • 2 years ago

Re: noise that looks like a sofa :P, adding some better priors seems to help sidestep that problem and generate good exemplar images for a given output - see <http://yosinski.com/deepvis> for more.

^ | v • Reply • Share ›

**Guest** • 3 years ago

"Something have to change." Grammatical error recognition rate got worse for the editor. Great article.

^ | v • Reply • Share ›

**anglrphish** Mod → Guest • 3 years ago

Whooops, thank you. Writing in English is still hard, but that was an especially dumb mistake. :-)

^ | v • Reply • Share ›

**Jakub Narebski** • 3 years ago

This reminds me of (being perhaps an urban legend) one student of History of Art recognizing churches on images not by their interior and art (which was intended to), but by the layout and number of pews...

^ | v • Reply • Share ›

**koko** • 3 years ago

What wasn't mentioned is that humans map perceived 2D images to their mental 3D representations of objects. Can a CNN learn that the reality is 3D from 2D images? To make things worse some objects like leopard can change shape of their body (move their body parts), this combined with different 3D angles of camera shots means that it is probably impossible for current CNNs to actually understand how a leopard really looks. A leopard with 7 legs and two tails would look perfectly fine to a CNN.

^ | v • Reply • Share ›

**Rainer Kordmaa** • a year ago

Kinda reminds me a story of how a neural network was trained by military to detect camouflaged tanks on terrain, except pictures with tanks were taken on a nice sunny day and pictures without on a cloudy day and instead of a tank detector they ended up with a sunny day detector

^ | v • Reply • Share ›

ALSO ON ROCKNROLLNERD.GITHUB.IO

Memory is a lazy mistress

6 comments • 3 years ago



docotor — Thanks for your efforts and samples. I think I'll need to read it few more times, as well as watching videos

...

Wrong whales

6 comments • 2 years ago



anglrphish — Hi! Sorry it took me that much time to answer. :-)Unfortunately, I've dropped out of the competition and ...

Oversimplified introduction to Boltzmann Machines

7 comments • 3 years ago



rukhreeves — Thank you for these posts! I came across RBMs recently and was completely thrown off by them even




...

Suddenly, a leopard print sofa appears

9 comments • 3 years ago



anglrphish — My thoughts exactly! Hope we're going to see the progress in this area yet.

 [Subscribe](#)  [Add Disqus to your site](#)[Add Disqus](#)[Add](#)  [Privacy](#)

Written by

Artem Khurshudov

Published 27 May 2015