Apprentissage profond: Raisons qui nous ont fait reconsidérer les réseaux de neurones.

Ludovic Trottier

April 8, 2016

Plan

Apprentissage profond (Deep Learning)

Sur Wikipédia, on peut lire:

1. Deep structured learning, hierarchical learning, deep machine learning
2. Attempt to model high-level abstractions in data by using multiple processing layers.
3. Deep learning < Machine Learning < Representation Learning (replacing handcrafted features with learned representations)
4. Buzzword, rebranding of neural network

Est-ce vraiment un rebranding?

10 BREAKTHROUGH TECHNOLOGIES 2013

MIT Technology Review

Introduction | The 10 Technologies | Past Years

Deep Learning

With massive amounts of computational power, machines can now recognize objects and translate speech in real time. Artificial intelligence is finally getting smart. →

Temporary Social Media

Messages that quickly self-destruct could enhance the privacy of online communications and make people freer to be spontaneous. →

Prenatal DNA Sequencing

Reading the DNA of fetuses will be the next frontier of the genomic revolution. But do you really want to know about the genetic problems or musical aptitude of your unborn child? →

Additive Manufacturing

Skeptical about 3-D printing? GE, the world's largest manufacturer, is on the verge of using the technology to make jet parts. →

Baxter: The Blue-Collar Robot

Rodney Brooks's newest creation is easy to interact with, but the complex innovations behind the robot show just how hard it is to get along with people. →

Memory Implants

A maverick neuroscientist believes he has deciphered the code by which the brain forms long-term memories. Next: testing a prosthetic implant for people suffering from long-term memory loss. →

Smart Watches

The designers of the Pebble watch realized that a mobile phone is more useful if you don't have to take it out of your pocket. →

Ultra-Efficient Solar Power

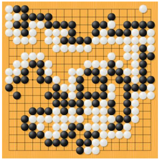
Doubling the efficiency of a solar cell would completely change the economics of renewable energy. Nanotechnology just might make it possible. →

Big Data from Cheap Phones

Collecting and analyzing information from simple cell phones can provide surprising insights into how people move about and behave – and even help us understand the spread of diseases. →

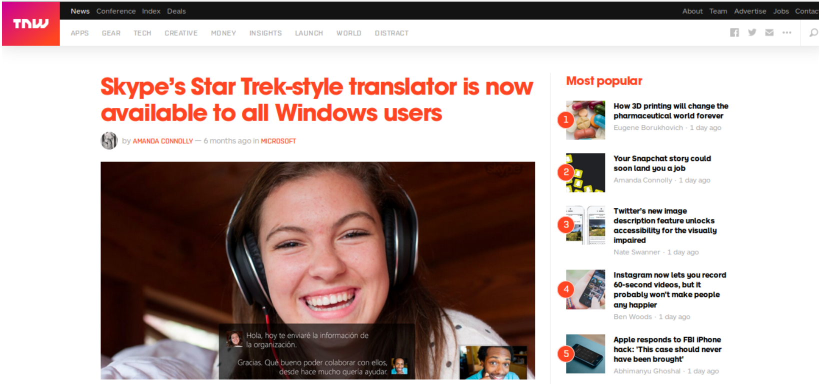
Supergrids

A new high-power circuit breaker could finally make highly efficient DC power grids practical. →



"In this paper we describe our Go program, AlphaGo. This program was based on general-purpose AI methods, using **deep neural networks** to mimic expert players, and further improving the program by learning from games played against itself."

Youtube: Google DeepMind Challenge Match: Lee Sedol vs AlphaGo



"Recent improvements in speech recognition, made possible by the introduction of **deep neural networks** combined with Microsofts proven statistical machine translation technology, allow for better translation outcomes, making meaningful one-on-one conversation possible."

http://blogs.skype.com/2014/12/15/skype-translator-how-it-works/

MIT Technology Review

Log in / Register Search

Subscribe

Topics+ The Daily Magazine Business Reports More+

Computing

Facebook Launches Advanced AI Effort to Find Meaning in Your Posts

A technique called deep learning could help Facebook understand its users and their data better.

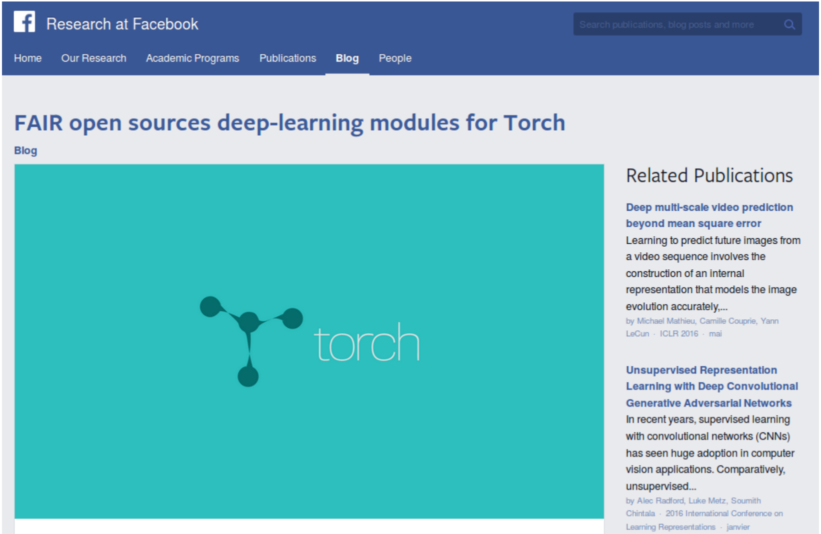
by Tom Simonite September 20, 2013

Facebook is set to get an even better understanding of the 700 million people who use the social network to share details of their personal lives each day.

A new research group within the company is working on an emerging and powerful approach to artificial intelligence known as deep learning, which uses simulated networks of brain cells to process data. Applying this method to data shared on Facebook could allow for novel features and perhaps boost the company's ad targeting.

Deep learning has shown potential as the basis for software that could work out the emotions or events described in text even if they aren't explicitly referenced, recognize objects in photos, and make sophisticated predictions about people's likely future behavior.

The eight-person group, known internally as the AI team, only recently started work, and details of its experiments are still secret. But Facebook's chief technology officer, Mike Schroepfer, will say that one obvious way to


Research at Facebook
Search publications, blog posts and more
Home
Our Research
Academic Programs
Publications
Blog
People
FAIR open sources deep-learning modules for Torch
Blog
torch
Related Publications
Deep multi-scale video prediction beyond mean square error
Learning to predict future images from a video sequence involves the construction of an internal representation that models the image evolution accurately,...
by Michael Mathieu, Camille Couprie, Yann LeCun · ICLR 2016 · mai
Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks
In recent years, supervised learning with convolutional networks (CNNs) has seen huge adoption in computer vision applications. Comparatively, unsupervised...
by Alec Radford, Luke Metz, Soumith Chintala · 2016 International Conference on Learning Representations · janvier

Computing

IBM Pushes Deep Learning with a Watson Upgrade

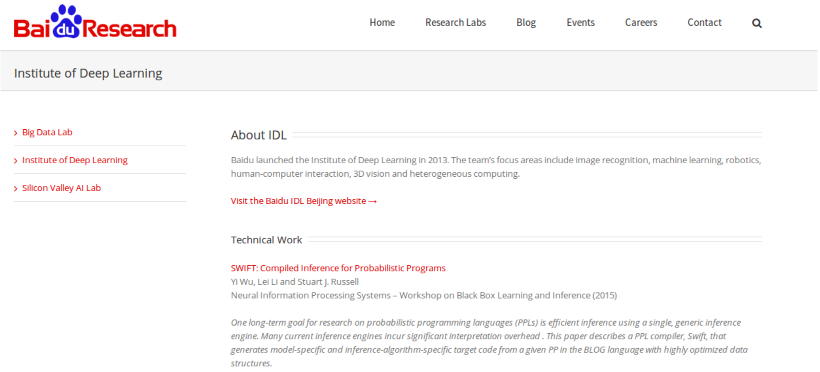
IBM is combining different AI techniques, including deep learning, in the commercial version of Watson.

by Will Knight   July 9, 2015

IBM's *Jeopardy!*-playing computer system, Watson, combined two separate areas of artificial intelligence research with winning results. Natural language understanding was merged with statistical analysis of vast, unstructured piles of text to find the likely answers to cryptic *Jeopardy!* clues.

Now IBM aims to add another powerful AI technique, known as deep learning, to the commercial version of Watson. The move could make the platform considerably smarter and more useful, and points to a promising future direction for AI research.

In its effort to commercialize Watson, IBM has made some of the features developed for the *Jeopardy!* challenge, as well as some new ones, available to developers via a cloud application programming interface (API). It has now added three deep-learning-based features to this Watson API: translation, speech-to-text, and text-to-speech. These could be used to build, for example, apps or websites that offer translation or transcription services. But developers could also connect them to other Watson services that parse questions and search for answers in large amounts of text. This could lead to an app that makes



Baidu Research

Home | Research Labs | Blog | Events | Careers | Contact

Institute of Deep Learning

- Big Data Lab
- Institute of Deep Learning
- Silicon Valley AI Lab

About IDL

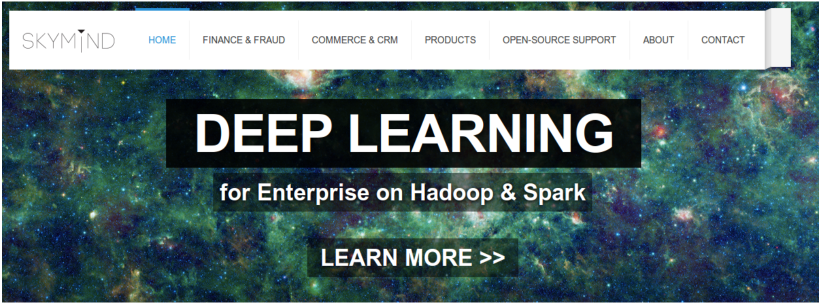
Baidu launched the Institute of Deep Learning in 2013. The team's focus areas include image recognition, machine learning, robotics, human-computer interaction, 3D vision and heterogeneous computing.

Visit the Baidu IDL Beijing website →

Technical Work

SWIFT: Compiled Inference for Probabilistic Programs
Yi Wu, Lei Li and Stuart J. Russell
Neural Information Processing Systems – Workshop on Black Box Learning and Inference (2015)

One long-term goal for research on probabilistic programming languages (PPLs) is efficient inference using a single, generic inference engine. Many current inference engines incur significant interpretation overhead . This paper describes a PPL compiler, Swift, that generates model-specific and inference-algorithm-specific target code from a given PP in the BLOG language with highly optimized data structures.


SKYMIND
HOME
FINANCE & FRAUD
COMMERCE & CRM
PRODUCTS
OPEN-SOURCE SUPPORT
ABOUT
CONTACT
DEEP LEARNING
for Enterprise on Hadoop & Spark
LEARN MORE >>

phoneArena.com

Site Search

HOME PHONES TABLETS NEWS REVIEWS VIDEOS

Home › News › Microsoft's Cortana to receive deep-learning and object recognition technologies

HOME
PHONES
CARRIERS
VERIZON
AT&T
SPRINT
T-MOBILE
MORE...
BRANDS
SAMSUNG
MOTOROLA
HTC
LG
BLACKBERRY
APPLE
MICROSOFT
MORE...
TABLETS

Microsoft's Cortana to receive deep-learning and object recognition technologies

Posted: 15 Jul 2014, 05:01, by Paul.K

Tags: Windows + Microsoft +

Even in Beta, Microsoft's Cortana has been showing potential to be just as adequate of an assistant as both Siri and Google Now. Recent revealings show that she is not done growing yet.

Firstly, the company intends to integrate a whole lot of academic data into Bing, as part of the "Microsoft Academic Search" project. Cortana, being powered by Bing, will receive the full benefits of having quick access to that data. The project's future holds the development of a community portal for academic workers, where any of the researchers can control how much of their personal data is visible. This is sure to bring a new level of collaboration in the science community, and we would love to see it

MIT Technology Review

Log in / Register Search

Subscribe

Topics+ The Daily Magazine Business Reports More+

A View from **Emerging Technology from the arXiv**

Deep-Learning Robot Takes 10 Days to Teach Itself to Grasp

Leave a human baby with some toys and it'll quickly learn to pick them up. Now a robot with deep-learning capabilities has done the same thing.

October 5, 2015

One of the goals of general purpose robots is to interact in an intelligent way with everyday objects. But robotic grasping skills are embarrassingly poor. Ask a robot to pick up a TV remote or a bottle of water or a toy gun and it will endlessly fumble with it—unless specifically programmed to pick up that object in a specially controlled environment.

That's in stark contrast to human grasping capabilities. A human baby quickly learns to grasp such objects, even in the most cluttered and unstructured environments.

And therein lies an important clue. Could robots learnt to grasp like babies, by repeated trial and error?

Today, Lerrel Pinto and Abhinav Gupta at Carnegie Mellon University in Pittsburgh show how this is possible. These guys have equipped a robot called Baxter with powerful deep learning capabilities, placed a table full of ordinary objects in front of it and then left it to learn, like a baby playing in a high chair.

Baxter is a modern two-armed industrial robot designed to perform repeatable tasks in environments such as factory floors. Each arm has a standard two-fingered parallel gripper and a high resolution camera to allow the robot to see what it is grasping close up. It also has a Microsoft Kinect sensor to provide an overview of the table in front of it.

Research at Facebook

Home | Our Research | Academic Programs | **Publications** | Blog | People

DeepFace: Closing the Gap to Human-Level Performance in Face Verification

Publication

Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, Lior Wolf
Conference on Computer Vision and Pattern Recognition (CVPR) · 24 juin 2014

Applied Machine Learning, Facebook AI Research (FAIR)

Abstract

In modern face recognition, the conventional pipeline consists of four stages: detect => align => represent => classify. We revisit both the alignment step and the representation step by employing explicit 3D face modeling in order to apply a piecewise affine transformation, and derive a face representation from a nine-layer deep neural network. This deep network involves more than 120 million parameters using several locally connected layers without weight sharing, rather than the standard convolutional layers. Thus we trained it on the largest facial dataset to-date, an identity labeled dataset of four million facial images belonging to more than 4,000 identities.

The learned representations coupling the accurate model-based alignment with the large facial database generalize remarkably well to faces in unconstrained environments, even with a simple classifier. Our method reaches an accuracy of 97.35% on the Labeled Faces in the Wild (LFW) dataset, reducing the error of the current state of the art by more than 27%, closely approaching human-level performance.

J'aime | Partager | 62

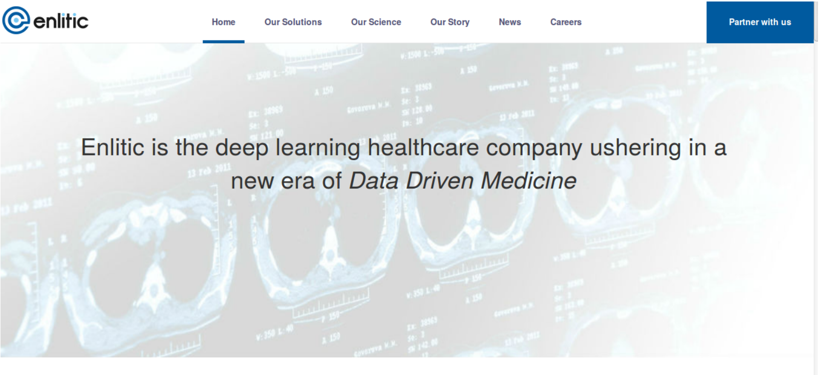
Resources

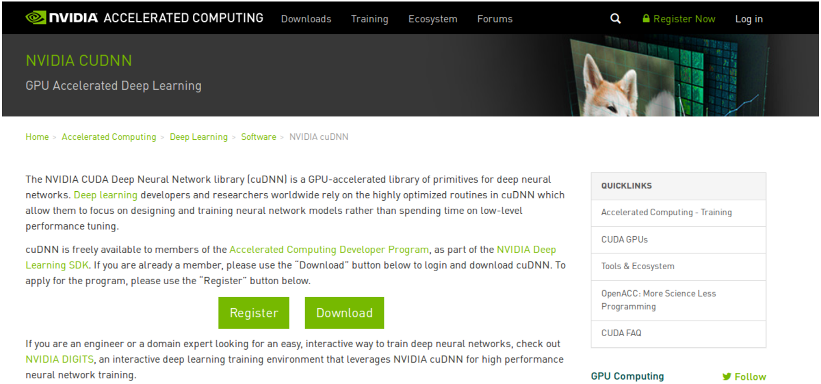
Download Paper

Related Publications

Deep multi-scale video prediction beyond mean square error
Learning to predict future images from a video sequence involves the construction of an internal representation that models the image evolution accurately,...
by Michael Mathieu, Camille Couprie, Yann LeCun · ICLR 2016 · mai

Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks
In recent years, supervised learning with convolutional networks (CNNs)


enlitic
Home
Our Solutions
Our Science
Our Story
News
Careers
Partner with us
Enlitic is the deep learning healthcare company ushering in a new era of *Data Driven Medicine*


NVIDIA ACCELERATED COMPUTING
Downloads
Training
Ecosystem
Forums
Register Now
Log in
NVIDIA CUDNN
GPU Accelerated Deep Learning
Home > Accelerated Computing > Deep Learning > Software > NVIDIA cuDNN
The NVIDIA CUDA Deep Neural Network library (cuDNN) is a GPU-accelerated library of primitives for deep neural networks. Deep learning developers and researchers worldwide rely on the highly optimized routines in cuDNN which allow them to focus on designing and training neural network models rather than spending time on low-level performance tuning.
cuDNN is freely available to members of the Accelerated Computing Developer Program, as part of the NVIDIA Deep Learning SDK. If you are already a member, please use the "Download" button below to login and download cuDNN. To apply for the program, please use the "Register" button below.
Register
Download
If you are an engineer or a domain expert looking for an easy, interactive way to train deep neural networks, check out NVIDIA DIGITS, an interactive deep learning training environment that leverages NVIDIA cuDNN for high performance neural network training.
QUICKLINKS
Accelerated Computing - Training
CUDA GPUs
Tools & Ecosystem
OpenACC: More Science Less Programming
CUDA FAQ
GPU Computing
Follow

Tutoriel:

`http://www.makeuseof.com/tag/create-neural-paintings-deepstyle-ubuntu/`

http://www.makeuseof.com/tag/create-neural-paintings-deepstyle-ubuntu/

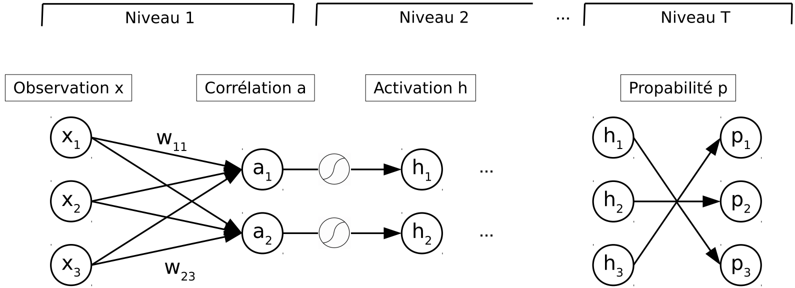
Ce qu'on peut en retirer...

Le Deep Learning c'est:

1. Une technologie applicable dans plusieurs domaines
   1.1 Jeu de Go
   1.2 reconnaissance vocale (Cortana, Skype, Google Voice)
   1.3 traitement automatique du langage naturel (Watson, Facebook, Twitter)
   1.4 vision numérique (DeepFace)
   1.5 robotique (grasping, Google Car)
   1.6 système d'aide à la décision (soins de santé)
   1.7 art (Neural Style)
   1.8 ...
2. Une approche d'apprentissage automatique nécessitant GPU + données massives.
3. Un lien étroit entre la recherche et l'industrie.

Réseau de neurones pré 2006

Réseau de neurones pré 2006



Corrélation

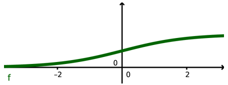
$a_i = x_1 w_{i1} + x_2 w_{i2} + x_3 w_{i3} = \langle x,w \rangle$

Activation

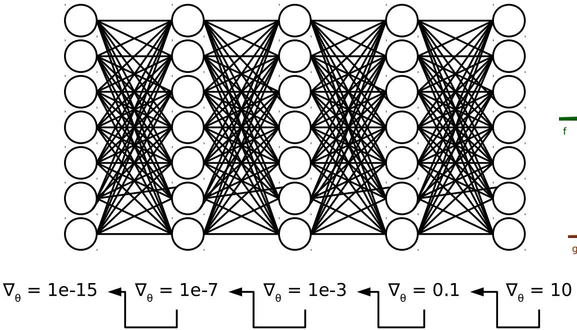
Logistique: $h_i = 1 / (1 + \exp(-a_i))$

$\langle x,w \rangle = 0$ ssi orthogonaux

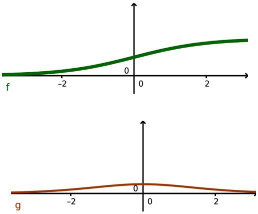
$\langle x,w \rangle = ||x||\ ||w||$ ssi colinéaire



Problème 1: Vanishing Gradient



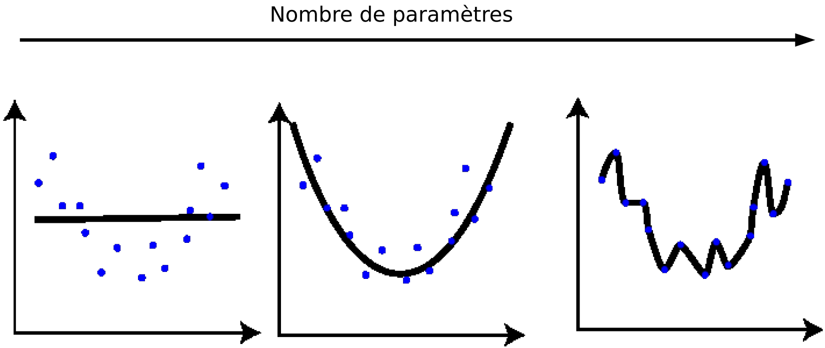
Fonction logistique



Sa dérivé

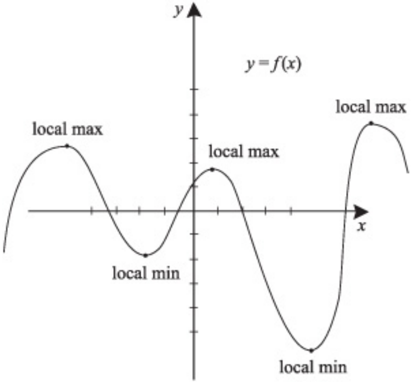
Conséquence: Les niveaux inférieurs prennent un temps fou à converger.

Problème 2: Sur-apprentissage



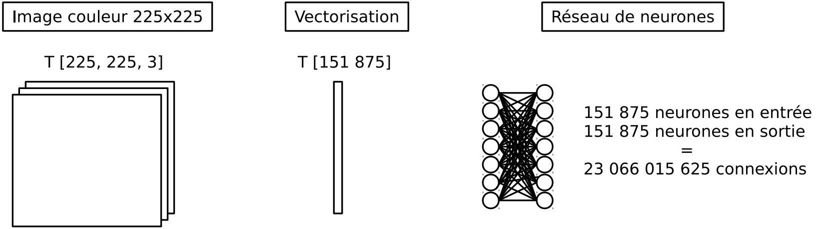
Conséquence: Pour la même quantité d'observations, plus il y a de niveaux/neurones, plus le réseau sur-apprend.

Problème 3: Fonction objective truffée de minimums locaux



Conséquence: Plus il y a de niveaux/neurones, plus la fonction est difficile.

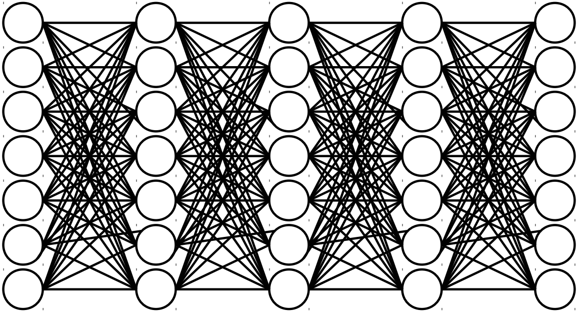
Problème 4: Malédiction de la dimensionalité



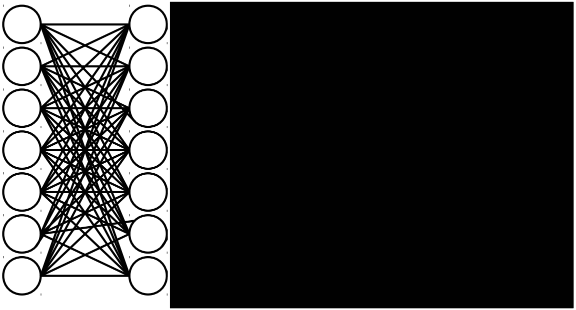
Conséquence: Plus la dimension de l'observation augmente, plus il y a de connexions à apprendre.

Qu'est-ce qui est arrivé en 2006 ?

Apprentissage vorace des niveaux de façon non-supervisée (Hinton et al., 2006)

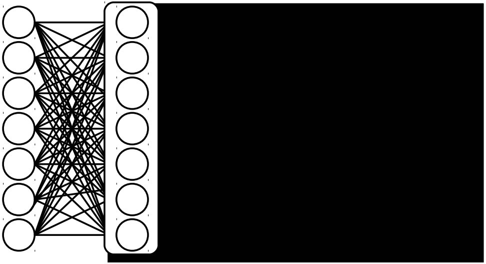


Qu’est-ce qui est arrivé en 2006 ?



Qu’est-ce qui est arrivé en 2006 ?

Machine de Boltzmann Restreinte

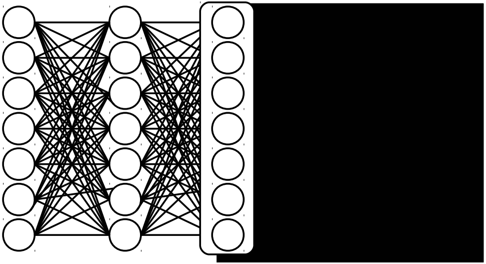


Au lieu d’avoir une distribution de probabilité conditionnelle Pr(y|x), on a maintenant une distribution de densité marginale p(x).

- Le but est maintenant de maximiser p(x) en fonction des connexions.
- Classification $\rightarrow$ Estimation de densité

Qu'est-ce qui est arrivé en 2006 ?

Machine de Boltzmann Restreinte

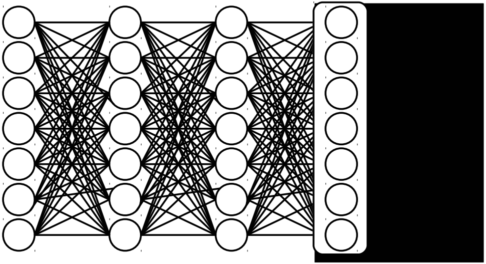


Au lieu d'avoir une distribution de probabilité conditionnelle Pr(y|x), on a maintenant une distribution de densité marginale p(x).

- Le but est maintenant de maximiser p(x) en fonction des connexions.
- Classification → Estimation de densité

Qu’est-ce qui est arrivé en 2006 ?

Machine de Boltzmann Restreinte

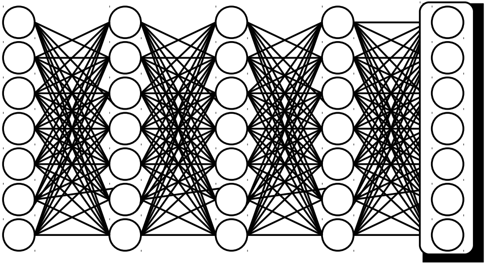


Au lieu d’avoir une distribution de probabilité conditionnelle Pr(y|x), on a maintenant une distribution de densité marginale p(x).

- Le but est maintenant de maximiser p(x) en fonction des connexions.
- Classification $\rightarrow$ Estimation de densité

Qu'est-ce qui est arrivé en 2006 ?

Machine de Boltzmann Restreinte

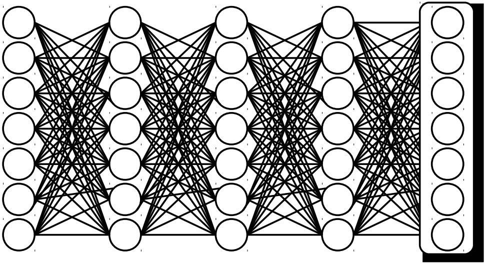


Au lieu d'avoir une distribution de probabilité conditionnelle Pr(y|x), on a maintenant une distribution de densité marginale p(x).

- Le but est maintenant de maximiser p(x) en fonction des connexions.
- Classification $\rightarrow$ Estimation de densité

Qu'est-ce qui est arrivé en 2006 ?

Machine de Boltzmann Restreinte



Au lieu d'avoir une distribution de probabilité conditionnelle Pr(y|x), on a maintenant une distribution de densité marginale p(x).

Conséquence:

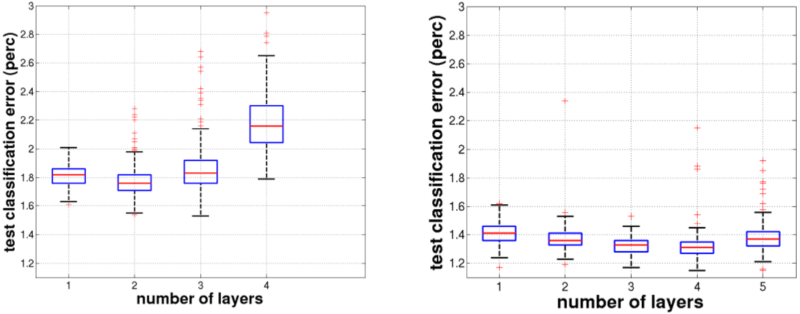
1. Possibilité d'apprendre sur des données non-étiquetées.
2. Excellente initialisation pour l'apprentissage supervisé.
3. Première fois que les réseaux profonds fonctionnent.

Impact majeur

1. Greedy layer-wise training of deep networks (Bengio et al., 2007).
2. Unsupervised learning of invariant feature hierarchies with applications to object recognition (Ranzato et al., 2007).
3. Unsupervised feature learning for audio classification using convolutional deep belief networks (Lee et al., 2009)
4. Sparse feature learning for deep belief networks (Boureau et al., 2008)
5. Training restricted Boltzmann machines using approximations to the likelihood gradient (Tieleman, 2008)
6. Large-scale deep unsupervised learning using graphics processors (Raina et al., 2009)
7. Representational power of restricted Boltzmann machines and deep belief networks (Le Roux and Bengio, 2008)

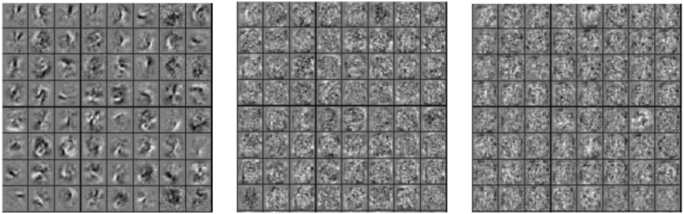
Pourquoi ça fonctionne?

Après 4 ans de recherche, on comprend que l'entraînement de Hinton est une forme de régularisation (Erhan et al., 2010).



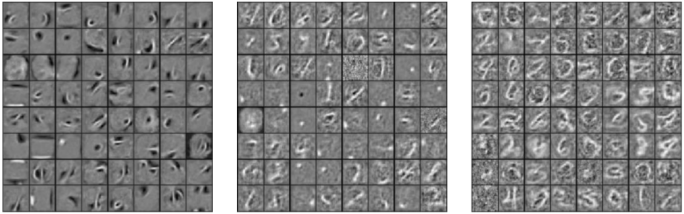
Exemple de vanishing gradient (Erhan et al., 2010)

Réseau 1



Niveau 1 Niveau 2 Niveau 3

Réseau 2



Niveau 1 Niveau 2 Niveau 3

Une nouvelle percée

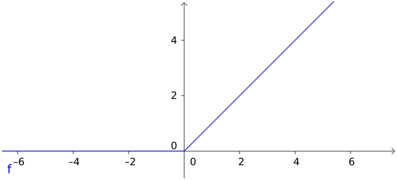
- Les nombreux travaux de recherche entre 2006 et 2010 nous ont montré l'importance de contrôler le vanishing gradient.
- Le pré-entraînement est long et fastidieux.
- En 2010, une nouvelle fonction d'activation ayant des propriétés intéressantes est proposée: la Rectified Linear Unit (ReLU) (Nair and Hinton, 2010):

$$f(x) = \begin{cases} x & \text{si } x \geq 0 \\ 0 & \text{sinon} \end{cases}$$

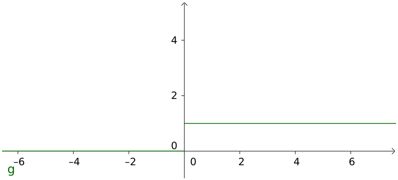
Propriétés intéressantes de la ReLU

La ReLU a plusieurs propriétés intéressantes:

1. Biologiquement plausible: one-sided comparé à l'anti-symétrique tanh.
2. Activations creuses: 50% des neurones ont une activation nulle après l'initialisation aléatoire des poids.
3. Pas de vanishing gradient: La dérivé vaut 1 partout dans la portion positive.
4. Complexité faible: comparaison.



ReLU



Sa dérivée

Impact majeur de (Nair and Hinton, 2010)

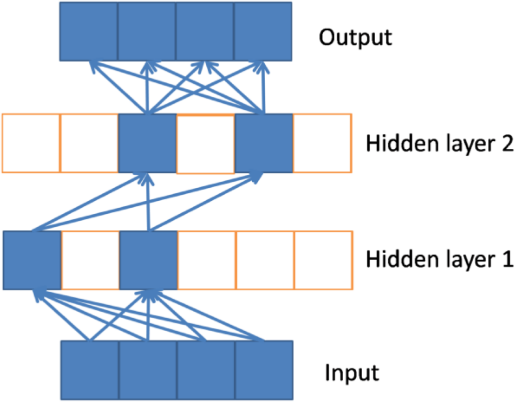
En 2011, pour la première fois on apprend un réseau de neurones profond sans entraînement non-supervisé vorace (Glorot et al., 2011).

On peut lire dans le résumé de l'article:

Even though they [rectifying neurons] can take advantage of semi-supervised setups with extra-unlabeled data, deep rectifier net can reach their best performance **without requiring any unsupervised pre-training** on purely supervised tasks with large labeled datasets. Hence, these results can be seen as **a new milestone** in the attempts at understanding the difficulty in training deep but purely supervised neural networks.

Observation empirique d'une autre propriété intéressante

Propriété: propagation de la sparsité à travers les niveaux (Glorot et al., 2011).

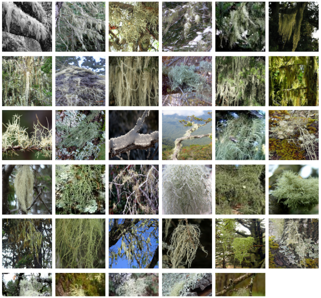


Exemple de tâche d'apprentissage

Exemple de tâche d'apprentissage

ImageNet Classification

1.4 millions d'images couleur entraînement (138 Go), 50 000 validation (6.3 Go), 100 000 test (13 Go). 1000 classes. Dimensions variables $\approx 256 \times 256$.

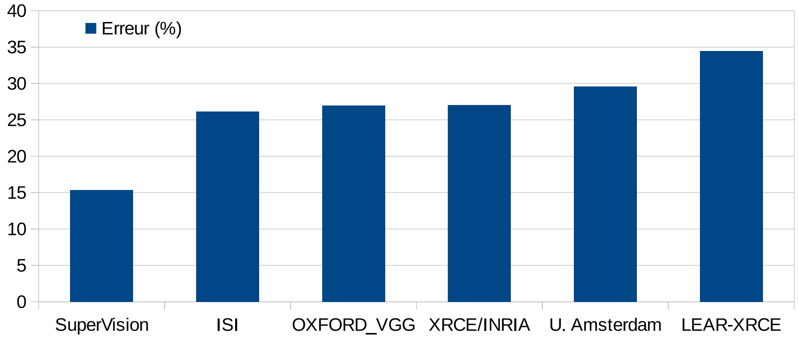


Lichen barbu

Lichen des caribous

Une redécouverte majeure

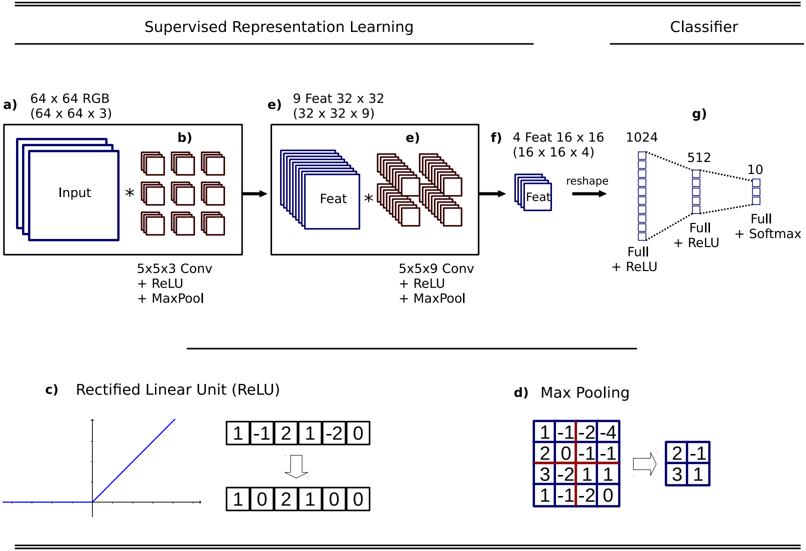
Krizhevsky et al. (2012) entraînement un réseau profond à convolution à 8 niveaux de 60 M paramètres et 650 000 neurones.



Trois approches expliquent leur succès:

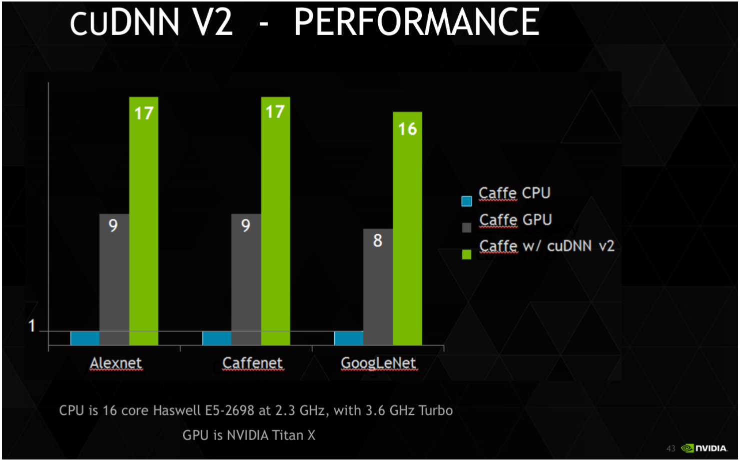
1. Convolution (ReLU, Pooling)
2. GPU
3. Dropout

Approche #1: Réseau de neurones à convolution

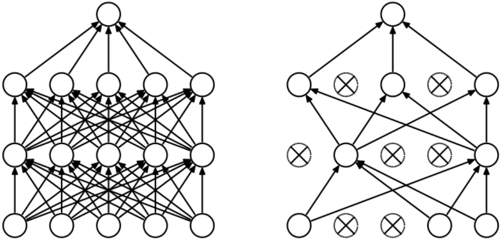


Approche #2: GPU

Entraîner le réseau de Krizhevsky et al. (2012) demande 6 jours sur deux GPUs.



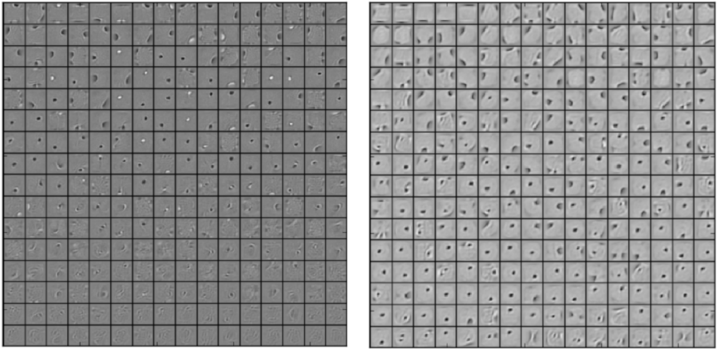
Approche #3: Dropout (Srivastava et al., 2014)



(a) Standard Neural Net (b) After applying dropout.

Figure 1: Dropout Neural Net Model. **Left**: A standard neural net with 2 hidden layers. **Right**: An example of a thinned net produced by applying dropout to the network on the left. Crossed units have been dropped.

Dropout (Srivastava et al., 2014)



(a) Without dropout (b) Dropout with $p = 0.5$.

Figure 12: Features learned on MNIST by 256 hidden unit RBMs. The features are ordered by L2 norm.

Pour aller plus loin...

IGGG Computer Vision + Applied Machine Learning Reading Group

Vendredi 15 Avril, PLT-3904, 3:30 PM:
Recent Advances in Convolutional Neural Networks.
`http://www2.ift.ulaval.ca/~pgiguere/rgroup/readingGroup2015.html`

Je parlerai des avancées entre 2012-2015 telle que:

1. Leaky ReLU, Noisy ReLU, Parametric ReLU, exponential LU
2. Drop-connect
3. Stochastic pooling, spatial pyramidal pooling
4. Batch Normalization
5. L'architecture Inception (oui, comme le film)

Conclusion

Pas tant un rebranding...

Oui, il y a des réseaux de neurones, mais pré 2006, il n'y avait pas:

1. Apprentissage non-supervisé vorace.
2. Fonction d'activation non-saturée.
3. Régularisation stochastique.
4. Support matériel pour les convolutions.
5. Réduction de dimensionnalité.
6. Et bien d'autres (à suivre le 15 Avril).

Découvrez par vous-même: `http://deeplearning.net/`

Questions?

References I

Bengio, Y., Lamblin, P., Popovici, D., Larochelle, H., et al. (2007). Greedy layer-wise training of deep networks. *Advances in neural information processing systems*, 19:153.

Boureau, Y.-l., Cun, Y. L., et al. (2008). Sparse feature learning for deep belief networks. In *Advances in neural information processing systems*, pages 1185–1192.

Erhan, D., Bengio, Y., Courville, A., Manzagol, P.-A., Vincent, P., and Bengio, S. (2010). Why does unsupervised pre-training help deep learning? *The Journal of Machine Learning Research*, 11:625–660.

Glorot, X., Bordes, A., and Bengio, Y. (2011). Deep sparse rectifier neural networks. In *International Conference on Artificial Intelligence and Statistics*, pages 315–323.

Hinton, G. E., Osindero, S., and Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. *Neural computation*, 18(7):1527–1554.

References II

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105.

Le Roux, N. and Bengio, Y. (2008). Representational power of restricted boltzmann machines and deep belief networks. *Neural computation*, 20(6):1631–1649.

Lee, H., Pham, P., Largman, Y., and Ng, A. Y. (2009). Unsupervised feature learning for audio classification using convolutional deep belief networks. In *Advances in neural information processing systems*, pages 1096–1104.

Nair, V. and Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pages 807–814.

References III

Raina, R., Madhavan, A., and Ng, A. Y. (2009). Large-scale deep unsupervised learning using graphics processors. In *Proceedings of the 26th annual international conference on machine learning*, pages 873–880. ACM.

Ranzato, M. A., Huang, F. J., Boureau, Y.-L., and LeCun, Y. (2007). Unsupervised learning of invariant feature hierarchies with applications to object recognition. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–8. IEEE.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1):1929–1958.

Tieleman, T. (2008). Training restricted boltzmann machines using approximations to the likelihood gradient. In *Proceedings of the 25th international conference on Machine learning*, pages 1064–1071. ACM.