Yoshua Bengio

Tél. : +1 514 343-6804
Yoshua.Bengio@umontreal.ca
www.iro.umontreal.ca/~bengioy

Formation

1992 – 1993 **Bourse post-doctorale,** AT&T Bell Laboratories, NJ, É.-U.
Algorithmes d'apprentissage et de vision | Groupe de Larry Jackel et Yann LeCun

1991 – 1992 **Bourse post-doctorale**, MIT, MA, É.-U.
Bourse CRSNG | Apprentissage statistique / Données séquentielles
Department of Brain and Cognitive Sciences | Groupe de Michael I. Jordan

1988 – 1991 **Doctorat en informatique,** Université McGill, Montréal
Bourse CRSNG | Réseaux de neurones et modèles de Markov

1986 – 1988 **Maîtrise en informatique**, Université McGill, Montréal
Reconnaissance vocale par réseaux de neurones

1982 – 1986 **B. Ing. en génie informatique,** Université McGill, Montréal
Programme d'études spécialisées

Postes universitaires

Depuis 2002 **Professeur titulaire**, Université de Montréal

1997 – 2002 **Professeur agrégé,** Université de Montréal

1993 – 1997 **Professeur adjoint,** Université de Montréal

Titres et distinctions

2018 – 2023 **Titulaire d'une Chaire en IA** CIFAR-Canada (IACC) (1,25 M$)

Depuis 2016 **Directeur scientifique,** IVADO, Institut de valorisation des données

2016 – 2023 **Boursier,** Fonds d'excellence en recherche Apogée Canada (93,6 M$)
Demandeur principal, *Données au service des Canadiens : apprentissage profond et optimisation aux fins de la révolution du savoir* ; la plus importante bourse reçue par l'UdeM.

Depuis 2014 **Codirecteur,** programme Apprentissage automatique, apprentissage biologique, CIFAR
Anciennement appelé Calcul Neuronal et Perception Adaptative (CNPA) et dirigé au départ par Geoff Hinton, ce programme a financé les premières percées en apprentissage profond.

Depuis 2013 **Fondateur et président,** ICLR (International Conference on Learning Representations)

2012 – 2013 **Membre du comité des prix**, Association d'informatique canadienne
Également membre des comités des prix du meilleur article de NeurIPS 2012 et d'ICML 2016 ainsi que des comités de sélection des prochains présidents de programme (2013-2018)

Depuis 2010 **Membre du conseil d'administration,** Fondation Neural Information Processing Systems (NeurIPS, anciennement NIPS)

2009 **Président,** NeurIPS
Neural Information Processing Systems (NeurIPS) est une conférence scientifique de très haut niveau – la plus importante dans le domaine (plus de 1 000 propositions retenues) – dont les critères d'examen et d'acceptation se comparent à ceux des meilleures revues (taux d'acceptation de 20 à 25 %). Avec 56 articles publiés dans le cadre des congrès au fil des ans, Bengio se distingue comme l'un des contributeurs les plus prolifiques de la communauté NeurIPS.

2008 **Coprésident de programme,** NeurIPS 2008

2005 – 2015 **Chaire de recherche industrielle**, CRSNG

Depuis 2004 **Boursier principal**, CIFAR (Institut canadien de recherches avancées)

Depuis 2000 **Chaire de recherche du Canada en algorithmes d'apprentissage statistique**
Niveau 2, 2000 – 2005 ; niveau 1, depuis 2006.

1999 – 2009 **Membre du conseil d'administration,** Centre de recherches mathématiques (UdeM)

1993 **Fondateur et directeur scientifique, Mila - Institut québécois d'intelligence artificielle**
D'abord appelé LISA (fondé en 1993), Mila rassemble des chercheurs de l'Université de Montréal et de l'Université McGill au sein d'un organisme sans but lucratif. Comptant 300 chercheurs et 15 professeurs principaux, il s'agit du plus important centre de recherche universitaire sur l'apprentissage profond dans le monde. Mila a produit des articles avant-gardistes dans le domaine, notamment sur l'introduction à l'apprentissage profond (2006), l'apprentissage par curriculum (2009), la puissance des fonctions ReLU pour les réseaux profonds (2011) et les progrès réalisés grâce aux réseaux génératifs adverses (GAN) et à la traduction automatique neuronale (2014).

Autres expériences et titres professionnels

Rédacteur en chef intérimaire, *Journal of Machine Learning Research* (*JMLR*), *Neural Computation, Foundations and Trends in Machine Learning* et *Computational Intelligence*. Membre en 2012 du comité de nomination du rédacteur en chef du *JMLR*.

Rédacteur en chef adjoint, *Machine Learning, IEEE Trans. on Neural Networks*

Président de secteur ou membre du comité de programme pour différentes conférences scientifiques, dont NeurIPS 1995 et 2004; ICONIP 1996; IJCNN 2000; AISTATS 2001; ICPR 2002; ICML 2003, 2004, 2006, 2008, 2012, 2013,

2014 et 2015; CAp 2004, 2006, 2010, 2011. Organisation de la plupart des ateliers initiaux sur l'apprentissage profond à NeurIPS et ICML depuis 2007 et au premier symposium NeurIPS sur l'apprentissage profond en 2016.

Membre de comités de sélection de boursiers pour le FQRNT (1999-2000) au Québec et le CRSNG (2000-2003, 2006-2007) au Canada.

Cofondateur de plusieurs entreprises émergentes, dont Element AI (2016), qui a obtenu un investissement de série A record de 135 M$. Direction des efforts pour connecter Mila à l'écosystème entrepreneurial en IA et faire de Montréal un pôle mondial de l'intelligence artificielle en y attirant les laboratoires de recherche en IA de Microsoft, Google, Facebook, DeepMind, Samsung et Thales.

Prix et récompenses

2018	**Prix d'excellence pour l'ensemble des réalisations**, Association pour l'intelligence artificielle au Canada
2018	**Médaille du 50e anniversaire,** ministère des Relations internationales et de la Francophonie
2017	**Prix Marie-Victorin,** Prix du Québec La plus haute distinction de la province de Québec dans le domaine des sciences
2017	**Scientifique de l'année,** Radio-Canada
2017	**Membre de la Société royale du Canada**
2017	**Officier de l'Ordre du Canada**
2015	**Les 10 découvertes qui ont marqué l'année 2015,** *La Recherche* Pour ses travaux sur les minimums locaux des réseaux de neurones
2009	**Prix Urgel-Archambault,** ACFAS

Bourses et subventions

Actuelles

2017 – 2022	Réseaux stratégiques, CRSNG, 5,5 M$ pour 5 ans
2017 – 2020	Apprentissage profond, Groupe Samsung, 550 k$US/an pour 3 ans
2016 – 2021	Don non affecté, Microsoft, 1,2 M$/an pour 5 ans
2016 – 2023	Données au service des Canadiens, Fonds d'excellence en recherche Apogée Canada, 93,6 M$
2016 – 2019	Initiative sur la cyberinfrastructure, FCI, 5 M$
2016 – 2019	Google Focused Research Award, 250 k$US/an
2016 – 2022	Bourse de R-D collaborative en soins de santé, Imagia, 300 k$ pour 6 ans
2014 – 2019	Subvention à la découverte, CRSNG, 76 k$/an pour 5 ans
Depuis 2006	Chaire de recherche du Canada, 200 k$/an

Antérieures

2017, 2018	Don non affecté, Panasonic, 200 k$US en 2017 et 300 k$US en 2018
2017	Don non affecté pour l'équipement, Facebook, 1,5 M$
2016 – 2018	Subvention de R-D coopérative, CRSNG (en partenariat avec IBM), 200 k$/an
2015 – 2018	CRSNG + subvention de R-D coopérative IBM, 800 k$ pour 3 ans
2015 – 2018	Projet de processeur neuromorphique, Groupe Samsung, 100 k$/an pour 3 ans
2014 – 2018	Bourses de la Nuance Foundation (2), 200 k$/an pour 4 ans
2016	Commandite de recherche, Panasonic, 250 k$
2016	Subvention d'équipement, CRSNG, 135 k$
2014 – 2016	Apprentissage profond, Groupe Samsung, 500 k$US/an pour 2 ans
2014, 2015	Google Focused Research Award, 200 k$/an
2014	Don universitaire, Facebook, 50 k$
2013 – 2016	Subventions de projets stratégiques (2), CRSNG, 240 $k et 220 $k/an pour 3 ans
2012	Subvention De l'idée à l'innovation, CRSNG, 124 k$
2011 – 2016	Subvention de R-D coopérative, CRSNG-Ubisoft, 50 k$ et 80 k$/an
2011 – 2016	Chaire industrielle CRSNG-Ubisoft, 350 k$/an pour 5 ans
2010, 2011, 2013	Subvention d'engagement partenarial, CRSNG, 25 $k
2009 – 2012	Subvention de projet stratégique, CRSNG, 70 % de 120k$/an pour 3 ans
2009 – 2014	Subvention à la découverte, CRSNG, 70k$/an pour 5 ans
2008 – 2010	Subvention de projet stratégique, CRSNG, 50 % de 99 k$/an pour 2 ans
2008	Google Research Award, 50 k$
2007 – 2009	Subvention de R-D coopérative, CRSNG, 50 % de 73 k$/an pour 2 ans
2005 – 2010	Chaire industrielle CRSNG-CGI, 150 k$/an pour 5 ans
2004 – 2009	Subvention à la découverte, CRSNG, 56 k$/an pour 5 ans
2004 – 2006	Subvention de R-D coopérative, CRSNG, 56 k$/an pour 2 ans
2003 – 2005	Subvention de R-D coopérative, CRSNG, 45 k$/an pour 2 ans
2002 – 2008	Subvention IRSC NET, 5 % de 250 k$/an pour 6 ans
2000 – 2005	Chaire de recherche du Canada, 100 k$/an
1999 – 2011	Subvention MITACS RCE, 30 % de 130 k$/an pour 11 ans
1999 – 2008	Bell University Labs, 75 k$/an pour 10 ans
1993 – 2005	Subvention IRIS RCE, 30 % de 150 k$/an pour 11 ans

Contributions professionnelles et scientifiques

En décembre 2018, Google Scholar a répertorié plus de 164 700 citations de publications scientifiques de Yoshua Bengio, avec un indice h de 135 et près de 55 143 citations en 2018 seulement. Yoshua Bengio était, en 2018, le chercheur en informatique ayant le plus grand nombre de citations récentes par jour, parmi les chercheurs avec un indice h supérieur ou égal à 100. Une liste complète de ces chercheurs se trouve au lien suivant : http://www.iro.umontreal.ca/~bengioy/citation-rate-CS-13dec2018.html.

Sa carrière en recherche comprend les moments forts suivants, qui sont pour la plupart axés sur le développement de l'apprentissage profond. Ils se distinguent par des contributions majeures aux domaines des réseaux de neurones récurrents, du traitement du langage naturel et de l'apprentissage non supervisé.

1989 – 1998 Les réseaux convolutifs et récurrents combinés à l'alignement probabiliste (HMM) pour modéliser des séquences, sujet principal de sa thèse de doctorat (1991); NIPS 1988, NIPS 1989, Eurospeech 1991, PAMI 1991 et IEEE Trans. Neural Nets 1992. Ces architectures ont d'abord été appliquées à la **reconnaissance vocale** dans le cadre de son doctorat (et redécouvertes après 2010), puis avec l'équipe de Yann LeCun à la **reconnaissance de l'écriture manuscrite et à l'analyse de documents** (l'article le plus cité est « *Gradient-based learning applied to document recognition* », 1998, avec plus de 15 000 citations).

1991 – 1995 Articles sur l'art d'**apprendre à apprendre** en collaboration avec Samy Bengio, amorcés au IJCNN 1991 avec « *Learning a synaptic learning rule* ». L'idée d'apprendre à apprendre (notamment par la rétropropagation) est maintenant très populaire, mais, au début des années 1990, la puissance des ordinateurs n'était pas suffisante pour y arriver.

1993 – 1995 Reconnaissance de la **difficulté fondamentale de l'apprentissage dans les réseaux récurrents** et les autres modèles d'apprentissage automatique des dépendances temporelles, associée à la disparition et à l'explosion des gradients : ICNN 1993, NIPS 1993, NIPS 1994, *IEEE Transactions on Neural Nets* 1994 et NIPS 1995. Ces articles ont eu un impact majeur et ont motivé des articles ultérieurs sur les architectures pouvant favoriser l'apprentissage des dépendances à long terme et tenir compte de la disparition et de l'explosion des gradients. L'article paru dans *IEEE Transactions* en 1994 apporte une contribution subtile mais importante en montrant, au moyen de la théorie des systèmes dynamiques, que la condition requise pour stocker des bits d'information de manière fiable à long terme donne également lieu à la disparition de gradients. L'article de NIPS 1995 présente l'utilisation de différentes échelles de temps comme une solution à la disparition des gradients.

1999 – 2014 Analyse du rôle des **représentations distribuées** pour contourner la **malédiction de la dimension** en permettant d'améliorer exponentiellement la généralisation au-delà des régions d'où proviennent les exemples d'entraînement. Cette série d'articles explique également comment les méthodes basées sur la généralisation locale, comme les plus proches voisins et les SVM à noyau gaussien, n'ont pas cette capacité de généralisation. L'article accepté à NIPS 1999 a présenté, pour la première fois, des réseaux de neurones autorégressifs pour estimer la densité (l'ancêtre des modèles NADE et PixelRNNN/PixelCNN). Les articles des NIPS 2004, 2005 et 2011 à ce sujet montrent comment les réseaux de neurones peuvent apprendre une métrique locale, ce qui permet de conférer la puissance de la généralisation des représentations distribuées aux méthodes à noyau et aux diverses méthodes d'apprentissage. Un autre article de NIPS 2005 montre les limites fondamentales des méthodes à noyau en raison d'une généralisation de la malédiction de la dimension (la malédiction de fonctions très variables, qui ont de nombreux hauts et bas). Enfin, l'article d'ICRL 2014 démontre que, dans le cas des réseaux linéaires par partie (comme ceux des ReLU), les régions (parties linéaires) que distingue un réseau à une couche cachée sont exponentielles en nombre de neurones (alors que le nombre de paramètres est quadratique en nombre de neurones et qu'une méthode locale à noyau nécessite un nombre exponentiel d'exemples pour obtenir le même type de fonction).

2000 – 2008 **Intégration de mots aux réseaux de neurones et aux modèles de langue neuronaux.** L'article de NIPS 2000 présente pour la première fois l'apprentissage de mots intégrés dans un réseau de neurones qui modélise des données

linguistiques. La version 2003 dans la revue *JMLR* pousse ce sujet plus loin (ces deux articles ont généré environ 3 000 citations) et introduit également l'idée de **descente de gradient stochastique (SGD) asynchrone** pour l'entraînement distribué des réseaux neuronaux. L'intégration de mots est devenue l'une des méthodes les plus courantes de l'apprentissage profond en matière de traitement des langues, ce qui a engendré un nouveau sous-domaine de la linguistique informatique. Introduction de l'utilisation de l'échantillonnage préférentiel (AISTATS 2003, *IEEE Trans. on Neural Nets* 2008) et d'un modèle probabiliste hiérarchique (AISTATS 2005) pour accélérer les calculs et traiter des vocabulaires plus importants.

2006 – 2014 Démonstration de **l'avantage théorique de la profondeur** pour la généralisation. La présentation à NIPS 2006, qui démontrait expérimentalement l'avantage de la profondeur, est l'un des articles les plus cités dans le domaine (plus de 2 600 citations). L'article de NIPS 2011 montre comment des réseaux plus profonds somme-produit peuvent représenter des fonctions qui nécessiteraient un modèle considérablement plus grand si le réseau était peu profond. Enfin, l'article de NIPS 2014 sur le nombre de régions linéaires des réseaux de neurones profonds extrapole l'article d'ICLR 2014 mentionné plus haut en montrant que le nombre de parties linéaires produites par un réseau linéaire par partie augmente exponentiellement, à la fois en largeur des couches et en nombre de couches, c'est-à-dire en profondeur. Ainsi, les fonctions représentées par ces réseaux sont généralement impossibles à saisir efficacement avec les méthodes à noyau (à moins d'utiliser un réseau de neurones entraîné comme noyau).

2006 – 2014 **Apprentissage profond non supervisé** basé sur des autoencodeurs (avec le cas particulier des réseaux génératifs adverses comme modèles de décodeur uniquement, voir plus bas). L'article de NIPS 2006 traite du pré-entraînement glouton pour chaque couche, aussi bien dans le cas d'apprentissage supervisé que dans le cas d'apprentissage non supervisé avec des autoencodeurs. L'article du ICML 2008 présente les autoencodeurs débruitants et les articles NIPS 2013, ICML 2014 et JMLR 2014 exposent la théorie qui y est associée en les décrivant comme des modèles probabilistes appropriés, tout en proposant des solutions de rechange à l'entraînement par maximum de vraisemblance.

2014 Déconstruction du **mythe des minimums locaux** concernant l'optimisation des réseaux de neurones dans l'article de NIPS 2014 sur les points de selle et démonstration que le grand nombre de paramètres rend très improbable l'existence de mauvais minimums locaux.

2014 Invention des **réseaux génératifs adverses (GAN)**, avec un article à NIPS 2014, où sont présentées de nombreuses innovations relatives à l'entraînement de modèles génératifs profonds qui s'écartent du principe de maximum de vraisemblance et même du principe classique d'une seule fonction objective (et plutôt de multiples modèles entraînés selon la théorie des jeux, chacun avec son objectif). C'est actuellement l'un des domaines de recherche les plus populaires de l'apprentissage profond avec plus de 6 000 citations provenant principalement d'articles qui traitent des variantes des GAN, et ayant réussi à produire des images synthétiques d'un réalisme impressionnant que l'on n'aurait pas crues possibles il y a quelques années seulement.

2014 – 2016 Présentation de **l'attention douce basée sur le contenu** et de ses retombées pour la **traduction automatique neuronale**. L'équipe de Bengio a d'abord présenté l'architecture encodeur-décodeur (maintenant appelée séquence à séquence) (EMNLP 2014), puis a fait grimper le score BLEU grâce l'attention douce basée sur le contenu (ICLR 2015). Ces éléments constituent maintenant les fondements de la plupart des systèmes de traduction automatique commerciaux, un autre sous-domaine créé à l'aide de ces techniques.

Étudiants de deuxième cycle et postdoctoraux

Actuels

Postdoctorants : Min Lin, Devansh Arpit, Jason Jo, Joseph Paul Cohen, Mirco Ravanelli, Jonathan Binas

Doctorants : Guillaume Alain, Bart Merrienboer, Jessica Thompson, Taesup Kim, Julian Vlad Serban, Dmitrii Serdiuk, Saizheng Zhang, Benjamin Scellier, Dzmitry Bahdanau, Sarath Chandar Anbil Parthipan, Chinnadhurai Sankar, Sandeep Subramanian, Zhouhan Lin, Yaroslav Ganin, Tong Che, Tristan Sylvain, Sherjil Ozair, Akram Erraqabi, Valentin Thomas, William Fedus, Giancarlo Kerg, Salem Lahlou, Rim Assouel, Alex Lamb.

Étudiants en maîtrise: Stephanie Larocque, Philippe Lacaille, Anirudh Goyal, Francis Dutil, Samuel Lavoie-Marchildon, Rithesh Kumar, Barghav Kanuparthi.

Antérieurs (diplômés)

Postdoctorants : Devon Hjelm (2018), Simon Blackburn (2018), Adriana Romero Soriano (2017), Philemon Brakel (2017), Nicolas Ballas (2017), Sungjin Ahn (2016), Asja Fischer (2016), Jorg Bornschein (2015), Kyung-Hyun Cho (2015), Jyri Kivinen (2014), Heng Luo (2013), Aaron Courville (2011), Antoine Bordes (2011), Joseph Turian (2010), Michael Mendel (2010), Jerome Louradour (2008), Marina Sokolova (2007), Pierre-Jean L'Heureux (2006), Christopher Kermorvant (2005), Xiangdong Wang (2003), Gilles Caporossi (2002), Ichiro Takeuchi (2001), Takafumi Kanamori (2001), Claude Nadeau (2000), Stephen Langdell (2000), Holger Schwenk (1997), Samy Bengio (1996).

Doctorants : Vincent Dumoulin (2018), Laurent Dinh (2018), Junyoung Chung (2018), Caglar Gulcehre (2018), David Warde-Farley (2017), Li Yao (2017), Mehdi Mirza (2017), Yann Dauphin (2015), Xavier Glorot (2015), Razvan Pascanu (2014), Ian Goodfellow (2014), Guillaume Desjardins (2014), Nicolas Boulanger-Lewandoski (2013), Philippe Hamel (2012), Olivier Delalleau (2012), James Bergstra (2011), Dumitru Erhan (2011), François Rivest (2010), Nicolas Chapados (2009), Hugo Larochelle (2009), Nicolas Le Roux (2008), Julie Carreau (2008), Narjes Boufaden (2005), Pascal Vincent (2003), Charles Dugas (2003), Joumana Ghosn (2002), Steven Pigeon (2001), François Gingras (1999).

Étudiants en maîtrise : Olexa Bilaniuk (2018), Dong-Hyun Lee (2018), Kelvin Xu (2017), Soroush Mehri (2016), Samira Shabanian (2016), Jose Rodriguez Sotelo (2016), Kyle Kastner (2016), David Krueger (2016), Matthieu Courbariaux (2015), Pierre Luc Carrier (2014), Eric Thibodeau-Laufer (2014), Nicholas Leonard (2014), Valentin Bisson (2012), François Savard (2011), Olivier Breuleux (2010), Guillaume Desjardins (2009), Pierre-Antoine Manzagol (2007), Dumitru Erhan (2006), Marie Ouimet (2004), Christian Dorion (2004), Maryse Boisvert (2004), Fréderic Morin (2004), Francis Piérault (2003), Jean-François Paiement (2003), Jean-Sébastien Senecal (2003), Lynian Meng (2002), Nicolas Chapados (2000) Vincent-Philippe Lauzon (1999), Simon Latendresse (1999), Julien Desaulnier (1998).

Liste partielle de coauteurs

Yann LeCun, Geoff Hinton, Aaron Courville, Pascal Vincent, Vladimir Vapnik, Leon Bottou, Hugo Larochelle, Ronan Collobert, Ian Goodfellow, Antoine Bordes, Nicolas Le Roux, Samy Bengio, James Bergstra, Yves Grandvalet, Xavier Glorot, Jason Weston, Douglas Eck, Marco Gori, Juergen Schmidhuber, Dumitru Erhan, Olivier Chapelle, Lise Getoor, Thomas Breuel, Joseph Turian, Patrice Marcotte, Balazs Kegl, Tomas Mikolov, David Warde-Farley, Guido Montufar, Gal Chechik, Andrew Fitzgibbon, Patrick Haffner, Razvan Pascanu, Guillaume Desjardins, Patrice Simard, Salah Rifai, Pascal Lamblin, Kyunghyun Cho, Heng Luo, Yann Dauphin, Jean-Luc Gauvain, Renato De Mori, Paolo Frasconi, Caglar Gulcehre, Dzmitry Bahdanau, Jason Yosinski, Frederic Bastien, Jan Chorowski, Jorg Bornschein, Gregoire Mesnil, Nicolas Boulanger-Lewandowski, Junyoung Chung, Li Yao, Kelvin Xu, Alessandro Sordoni, Sherjil Ozair, Richard Zemel, Sepp Hochreiter, Saizheng Zhang, Dmitriy Serkyuk, Vincent Dumoulin, Chris Pal, Joelle Pineau, Jamie Kiros, Asja Fischer, Jeff Clune, Li Deng, Bing Xu, Laurent Dinh, Takeuchi Ichiro, Patrice Marcotte, Felix Hill, Heng Luo, Nicholas Leonard, Stephan Gouws

Contributions à la recherche

Publications dans des revues avec comité de lecture

[1] I. Hubara, M. Courbariaux, D. Soudry, R. El-Yaniv, and Y. Bengio, "Quantized neural networks: Training neural networks with low precision weights and activations," *Journal of Machine Learning Research*, vol. 18, no. 187, pp. 1–30, 2018.

[2] C. Gulcehre, S. Chandar, K. Cho, and Y. Bengio, "Dynamic neural turing machine with continuous and discrete addressing schemes," *Neural Computation*, vol. 30, no. 4, pp. 857–884, 2018.

[3] G. Derevyanko, S. Grudinin, Y. Bengio, and G. Lamoureux, "Deep convolutional networks for quality assessment of protein folds," *Bioinformatics*, 2018.

[4] M. Ravanelli, P. Brakel, M. Omologo, and Y. Bengio, "Light gated recurrent units for speech recognition," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 2, no. 2, pp. 92–102, 2018.

[5] H. Choi, K. Cho, and Y. Bengio, "Fine-grained attention mechanism for neural machine translation," *Neurocomputing*, vol. 284, pp. 171–176, 2018.

[6] X.-Y. Zhang, F. Yin, Y.-M. Zhang, C.-L. Liu, and Y. Bengio, "Drawing and recognizing chinese characters with recurrent neural network," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 849–862, 2018.

[7] M. Drozdzal, G. Chartrand, E. Vorontsov, M. Shakeri, L. D. Jorio, A. Tang, A. Romero, Y. Bengio, C. Pal, and S. Kadoury, "Learning normalized inputs for iterative estimationin medical image segmentation," *Medical Image Analysis*, vol. 44, pp. 1–13, 2018.

[8] P. D. Luna, J.Wei, Y. Bengio, A. Aspuru-Guzik, and E. Sargent, "Use machine learning to find energy materials," *Nature*, vol. 552, pp. 23–27, 2017.

[9] F. Hill, K. Cho, S. Jean, and Y. Bengio, "The representational geometry of word meanings acquired by neural machine translation models," *Machine Translation*, vol. 31, pp. 1–16, 2017.

[10] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.-M. Jodoin, and H. Larochelle, "Brain tumor segmentation with deep neural networks," *Medical Image Analysis*, vol. 35, pp. 18 – 31, 2017.

[11] Y. Bengio, T. Mesnard, A. Fischer, S. Zhang, and Y.Wu, "STDP-compatible approximation of back-propagation in an energy-based model," *Neural Computation*, vol. 29, no. 3, pp. 555–577, 2017.

[12] Ç. Gül.ehre, O. Firat, K. Xu, K. Cho, and Y. Bengio, "On integrating a language model into neural machine translation," *Computer Speech Language*, vol. 45, p. 137–148, 2017.

[13] X.-Y. Zhang, Y. Bengio, and C.-L. Liu, "Online and offline handwritten Chinese character recognition: A comprehensive study and new benchmark," *Pattern Recognition*, vol. 61, pp. 348–360, 2017.

[14] X.-Y. Zhang, G.-S. Xie, C.-L. Liu, and Y. Bengio, "End-to-end online writer identification with recurrent neural network," *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 2, pp. 285–292, 2017.

[15] H. Choi, K. Cho, and Y. Bengio, "Context-dependent word representation for neural machine translation," *Computer Speech & Language*, vol. 45, pp. 149–160, 2017.

[16] O. Firat, K. Cho, B. Sankaran, F. T. Y. Vural, and Y. Bengio, "Multi-way, multilingual neural machine translation," *Computer Speech & Language*, 2016.

[17] Y. Bengio, "Springtime for AI: The rise of deep learning," *Scientific American*, June 2016.

[18] G. Alain, Y. Bengio, L. Yao, J. Yosinski, E. Thibodeau-Laufer, S. Zhang, and P. Vincent, "GSNs: generative stochastic networks," *Information and Inference*, 2016.

[19] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.-M. Jodoin, and H. Larochelle, "Brain tumor segmentation with deep neural networks," *Medical Image Analysis*, 2016.

[20] X.-Y. Zhang, G.-S. Xie, C.-L. Liu, and Y. Bengio, "End-to-end online writer identification with recurrent neural networks," *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 2, pp. 285–292, 2016.

[21] F. Hill, K. Cho, A. Korhonen, and Y. Bengio, "Learning to understand phrases by embedding the dictionary," *Transactions of the Association for Computational Linguistics*, vol. 4, pp. 17–30, 2016.

[22] Ç. Gül.ehre and Y. Bengio, "Knowledge matters: Importance of prior information for

[23] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[24] K. Cho, A. Courville, and Y. Bengio, "Describing multimedia content using attentionbased encoder-decoder networks," *IEEE Transactions on Multimedia*, vol. 17, no. 11, pp. 1875–1886, 2015.

[25] I. J. Goodfellow, D. Erhan, P.-L. Carrier, A. Courville, M. Mirza, B. Hamner,W. Cukierski, Y. Tang, D. Thaler, D.-H. Lee, Y. Zhou, C. Ramaiah, F. Feng, R. Li, X.Wang, D. Athanasakis, J. Shawe-Taylor, M. Milakov, J. Park, R. Ionescu, M. Popescu, C. Grozea, J. Bergstra, J. Xie, L. Romaszko, B. Xu, Z. Chuang, and Y. Bengio, "Challenges

in representation learning : A report on three machine learning contests," *Neural Networks*, vol. 64, pp. 59–63, 2015.

[26] S. Ebrahimi Kahou, X. Bouthillier, P. Lamblin, Ç. Gül.ehre, V. Michalski, K. R. Konda, S. Jean, P. Froumenty, Y. Dauphin, N. Boulanger-Lewandowski, R. Chandias Ferrari, M. Mirza, D.Warde-Farley, A. Courville, P. Vincent, R. Memisevic, C. Pal, and Y. Bengio, "Emonets : Multimodal deep learning approaches for emotion recognition in video," *Journal on Multimodal User Interfaces*, pp. 1–13, 2015.

[27] F. Rivest, J. F. Kalaska, and Y. Bengio, "Conditioning and time representation in long short-term memory networks," *Biological Cybernetics*, vol. 108, no. 1, pp. 23–48, 2014.

[28] G. Mesnil, Y. Dauphin, K. Yao, Y. Bengio, L. Deng, D. Hakkani-Tur, X. He, L. Heck, G. Tur, D. Yu, and G. Zweig, "Using recurrent neural networks for slot filling in spoken language understanding," *IEEE Tr. ASSP*, 2015.

[29] G. Alain and Y. Bengio, "What regularized auto-encoders learn from the data-generating distribution," in *Journal of Machine Learning Research* [98], pp. 3563–3593.

[30] A. Courville, G. Desjardins, J. Bergstra, and Y. Bengio, "The spike-and-slab RBM and extensions to discrete and sparse data distributions," *IEEE Tr. PAMI*, vol. 36, no. 9, pp. 1874–1887, 2014.

[31] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI)*, vol. 35, no. 8, pp. 1798–1828, 2013.

[32] A. Bordes, X. Glorot, J. Weston, and Y. Bengio, "A semantic matching energy function for learning with multi-relational data," *Machine Learning: Special Issue on Learning Semantics*, 2013.

[33] G. Mesnil, A. Bordes, J. Weston, G. Chechik, and Y. Bengio, "Learning semantic representations of objects and their parts," *Machine Learning: Special Issue on Learning Semantics*, 2013.

[34] O. Delalleau, E. Contal, E. Thibodeau-Laufer, R. Chandias Ferrari, Y. Bengio, and F. Zhang, "Beyond skill rating: Advanced matchmaking in ghost recon online," *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 4, pp. 167–177, Sept. 2012.

[35] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *Journal of Machine Learning Research*, vol. 13, pp. 281–305, Feb. 2012.

[36] H. Larochelle, M. Mandel, R. Pascanu, and Y. Bengio, "Learning algorithms for the classification restricted boltzmann machine," *JMLR*, vol. 13, pp. 643–669, Mar. 2012.

[37] Y. Bengio, N. Chapados, O. Delalleau, H. Larochelle, and X. Saint-Mleux, "Detonation classification from acoustic signature with the restricted Boltzmann machine," *Computational Intelligence*, vol. 28, no. 2, 2012.

[38] O. Breuleux, Y. Bengio, and P. Vincent, "Quickly generating representative samples from an RBM-derived process," *Neural Computation*, vol. 23, pp. 2053–2073, Aug. 2011.

[39] J. Bergstra, Y. Bengio, and J. Louradour, "Suitability of V1 energy models for object classification," *Neural Computation*, vol. 23, p. 774–790, Mar. 2011.

[40] M. Mandel, R. Pascanu, D. Eck, Y. Bengio, L.M. Aeillo, R. Schifanella, and F.Menczer, "Contextual tag inference," *ACM T. Multimedia Comp., Comm. & Appl.*, vol. 7S, p. 1–32, Oct. 2011.

[41] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," in *Journal of Machine Learning Research* [98], pp. 3371–3408.

[42] H. Larochelle, Y. Bengio, and J. Turian, "Tractable multivariate binary density estimation and the restricted boltzmann forest," *Neural Computation*, vol. 22, pp. 2285–2307, Sept. 2010.

[43] N. Le Roux and Y. Bengio, "Deep belief networks are compact universal approximators," *Neural Computation*, vol. 22, pp. 2192–2207, Aug. 2010.

[44] Y. Bengio, O. Delalleau, and C. Simard, “Decision trees do not generalize to new variations,” *Computational Intelligence*, vol. 26, pp. 449–467, Nov. 2010.
[45] D. Erhan, Y. Bengio, A. Courville, P.-A. Manzagol, P. Vincent, and S. Bengio, “Why does unsupervised pre-training help deep learning?” in *Journal of Machine Learning Research* [98], pp. 625–660.
[46] F. Rivest, J. Kalaska, and Y. Bengio, “Alternative time representations in dopamine models,” *Journal of Computational Neuroscience*, vol. 28, no. 1, pp. 107–130, 2009.
[47] Y. Bengio, “Learning deep architectures for AI,” *Foundations and Trends in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009. [Also published as a book by Now Publishers, 2009.]
[48] C. Dugas, Y. Bengio, F. Belisle, C. Nadeau, and R. Garcia, “Incorporating functional knowledge in neural networks,” *The Journal of Machine Learning Research*, vol. 10, pp. 1239–1262, June 2009.
[49] J. Carreau and Y. Bengio, “A hybrid Pareto mixture for conditional asymmetric fat-tailed distribution*,” IEEE Transactions on Neural Networks*, vol. 20, no. 7, pp. 1087–1101, 2009.
[50] J. Carreau and Y. Bengio, “A hybrid pareto model for asymmetric fat-tailed data: the univariate case,” *Extremes*, vol. 12, no. 1, pp. 53–76, 2009.
[51] H. Larochelle, Y. Bengio, J. Louradour, and P. Lamblin, “Exploring strategies for training deep neural networks,” in *Journal of Machine Learning Research* [98], pp. 1–40.
[52] Y. Bengio and O. Delalleau, “Justifying and generalizing contrastive divergence,” *Neural Computation*, vol. 21, pp. 1601–1621, June 2009.
[53] N. Le Roux and Y. Bengio, “Representational power of restricted Boltzmann machines and deep belief networks,” *Neural Computation*, vol. 20, pp. 1631–1649, June 2008.
[54] Y. Bengio and J.-S. Sénécal, “Adaptive importance sampling to accelerate training of a neural probabilistic language model,” *IEEE Trans. Neural Networks*, vol. 19, no. 4, pp. 713–722, 2008.
[55] Y. Bengio, “Neural net language models,” *Scholarpedia*, vol. 3, no. 1, p. 3881, 2008.
[56] Y. Bengio, “On the challenge of learning complex functions,” in *Computational Neuroscience: Theoretical Insights into Brain Function* (P. Cisek, J. Kalaska, and T. Drew, eds.), Progress in Brain Research, Elsevier, 2007.
[57] N. Chapados andY. Bengio, “Noisy k best-paths for approximate dynamic programming with application to portfolio optimization,” *Journal of Computers*, vol. 2, no. 1, pp. 12–19, 2007.
[58] Y. Bengio, M. Monperrus, and H. Larochelle, “Nonlocal estimation of manifold structure,” *Neural Computation*, vol. 18, no. 10, pp. 2509–2528, 2006.
[59] D. Erhan, P.-J. L’Heureux, S. Y. Yue, and Y. Bengio, “Collaborative filtering on a family of biological targets.,” *Journal of Chemical Information and Modeling*, vol. 46, no. 2, pp. 626–635, 2006.
[60] M. C. Zaccaro, H. Boon, M. Pattarawarapan, Z. Xia, A. Caron, P.-J. L’Heureux, Y. Bengio, K. Burgess, and H. U. Saragori, “Selective small molecule peptidomimetic ligands of trkc and trka receptors afford discrete or complete neurotrophic activities,” *Chemistry & Biology*, vol. 12, no. 9, pp. 1015–1028, 2005.
[61] P.-J. L’Heureux, J. Carreau, Y. Bengio, O. Delalleau, and S. Y. Yue, “Locally linear embedding for dimensionality reduction in QSAR,” *Journal of Computer-Aided Molecular Design*, vol. 18, pp. 475–482, 2004.
[62] Y. Bengio, O. Delalleau, N. Le Roux, J.-F. Paiement, P. Vincent, and M. Ouimet, “Learning eigenfunctions links spectral embedding and kernel PCA,” *Neural Computation*, vol. 16, no. 10, pp. 2197–2219, 2004.
[63] Y. Bengio and Y. Grandvalet, “No unbiased estimator of the variance of K-fold cross-validation,” *Journal of Machine Learning Research*, vol. 5, pp. 1089–1105, 2004.

[64] R. Collobert, Y. Bengio, and S. Bengio., "Scaling large learning problems with hard parallel mixtures," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 17, no. 3, pp. 349–365, 2003.
[65] C. Nadeau and Y. Bengio, "Inference for the generalization error," *Machine Learning*, vol. 52, no. 3, pp. 239–281, 2003.
[66] J. Ghosn and Y. Bengio, "Bias learning, knowledge sharing," *IEEE Transactions on Neural Networks*, vol. 14, pp. 748–765, July 2003.
[67] Y. Bengio, R. Ducharme, P. Vincent, and C. Jauvin, "A neural probabilistic language model," *Journal of Machine Learning Research*, vol. 3, pp. 1137–1155, 2003.
[68] Y. Bengio and N. Chapados, "Extensions to metric-based model selection," *Journal of Machine Learning Research*, vol. 3, pp. 1209–1227, Mar. 2003. *Special Issue on Feature Selection.*
[69] I. Takeuchi, Y. Bengio, and T. Kanamori, "Robust regression with asymmetric heavy-tail noise distributions," *Neural Computation*, vol. 14, no. 10, pp. 2469–2496, 2002.
[70] R. Collobert, S. Bengio, and Y. Bengio, "Parallel mixture of SVMs for very large scale problems," *Neural Computation*, vol. 14, no. 5, pp. 1105–1114, 2002.
[71] O. Chapelle, V. Vapnik, and Y. Bengio, "Model selection for small-sample regression," *Machine Learning Journal*, vol. 48, no. 1, pp. 9–23, 2002.
[72] P. Vincent and Y. Bengio, "Kernel matching pursuit," *Machine Learning*, vol. 48, pp. 165–187, 2002.
[73] N. Chapados and Y. Bengio, "Cost functions and model combination for var–based asset allocation using neural networks," *IEEE Transactions on Neural Networks*, vol. 12, no. 4, pp. 890–906, 2001.
[74] Y. Bengio, V.-P. Lauzon, and R. Ducharme, "Experiments on the application of IOHMMs to model financial returns series," *IEEE Transactions on Neural Networks*, vol. 12, no. 1, pp. 113–123, 2001.
[75] H. Schwenk and Y. Bengio, "Boosting neural networks," *Neural Computation*, vol. 12, no. 8, pp. 1869–1887, 2000.
[76] Y. Bengio, "Gradient-based optimization of hyperparameters," *Neural Computation*, vol. 12, no. 8, pp. 1889–1900, 2000.
[77] S. Bengio and Y. Bengio, "Taking on the curse of dimensionality in joint distributions using neural networks," *IEEE Transactions on Neural Networks, special issue on Data Mining and Knowledge Discovery*, vol. 11, no. 3, pp. 550–557, 2000.
[78] Y. Bengio, "Markovian models for sequential data," *Neural Computing Surveys*, vol. 2, pp. 129–162, 1999.
[79] S. Bengio, Y. Bengio, J. Robert, and G. Bélanger, "Stochastic learning of strategic equilibria for auctions," *Neural Computation*, vol. 11, no. 5, pp. 1199–1209, 1999.
[80] L. Bottou, P. Haffner, P. Howard, P. Simard, and Y. Bengio, "High quality document image compression with DjVu," *Journal of Electronic Imaging*, vol. 7, no. 3, pp. 410–425, 1998.
[81] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, pp. 2278–2324, Nov. 1998.
[82] Y. Bengio, F. Gingras, B. Goulard, and J.-M. Lina, "Gaussian mixture densities for classification of nuclear power plant data," *Computers and Artificial Intelligence, special issue on Intelligent Technologies for Electric and Nuclear Power Plants*, vol. 17, no. 2–3, pp. 189–209, 1998.
[83] F. Gingras and Y. Bengio, "Handling asynchronous or missing financial data with recurrent networks," *International Journal of Computational Intelligence and Organizations*, vol. 1, no. 3, pp. 154–163, 1998.
[84] Y. Bengio, "Using a financial training criterion rather than a prediction criterion," *International Journal of Neural Systems*, vol. 8, no. 4, pp. 433–443, 1997. *Special issue on noisy time-series.*

[85] Y. Bengio and P. Frasconi, “Input/Output HMMs for sequence processing,” *IEEE Transactions on Neural Networks*, vol. 7, no. 5, pp. 1231–1249, 1996.
[86] Y. Bengio and P. Frasconi, “Diffusion of context and credit information in Markovian models,” *Journal of Artificial Intelligence Research*, vol. 3, pp. 249–270, 1995.
[87] Y. Bengio, Y. LeCun, C. Nohl, and C. Burges, “Lerec: A NN/HMM hybrid for on-line handwriting recognition,” *Neural Computation*, vol. 7, no. 6, pp. 1289–1303, 1995.
[88] S. Bengio, Y. Bengio, and J. Cloutier, “On the search for new learning rules for ANNs,” *Neural Processing Letters*, vol. 2, no. 4, pp. 26–30, 1995.
[89] Y. Bengio, P. Simard, and P. Frasconi, “Learning long-term dependencies with gradient descent is difficult,” *IEEE Transactions on Neural Networks*, vol. 5, no. 2, pp. 157–166, 1994. *Special Issue on Recurrent Neural Networks*, March 94.
[90] Y. Bengio, “A connectionist approach to speech recognition,” *International Journal on Pattern Recognition and Artificial Intelligence*, vol. 7, no. 4, pp. 647–668, 1993. Special issue entitled Advances in Pattern Recognition Systems using Neural Networks.
[91] Y. Bengio, M. Gori, and R. De Mori, “Learning the dynamic nature of speech with back-propagation for sequences,” *Pattern Recognition Letters*, vol. 13, no. 5, pp. 375–385, 1992. *Special issue on Artificial Neural Networks*.
[92] Y. Bengio, R. De Mori, G. Flammia, and R. Kompe, “Phonetically motivated acoustic parameters for continuous speech recognition using artificial neural networks,” *Speech Communication*, vol. 11, no. 2–3, pp. 261–271, 1992. *Special issue on neurospeech*.
[93] Y. Bengio, R. De Mori, G. Flammia, and R. Kompe, “Global optimization of a neural network-hidden Markov model hybrid,” *IEEE Transactions on Neural Networks*, vol. 3,
[94] Y. Bengio and Y. Pouliot, “Efficient recognition of immunoglobulin domains from amino-acid sequences using a neural network,” *Computer Applications in the Biosciences*, vol. 6, no. 2, pp. 319–324, 1990.
[95] P. Cosi, Y. Bengio, and R. De Mori, “Phonetically-based multi-layered networks for acoustic property extraction and automatic speech recognition,” *Speech Communication*, vol. 9, no. 1, pp. 15–30, 1990.
[96] Y. Bengio and R. D. Mori, “Use of multilayer networks for the recognition of phonetic features and phonemes,” *Computational Intelligence*, vol. 5, pp. 134–141, 1989.
[97] Y. Bengio, R. Cardin, R. De Mori, and E. Merlo, “Programmable execution of multi-layered networks for automatic speech recognition,” *Communications of the Association for Computing Machinery*, vol. 32, no. 2, pp. 195–199, 1989.

Articles dans des comptes rendus de conférence avec comité de lecture

[98] A. Gonzalez-Garcia, J. van de Weijer, and Y. Bengio, “Image-to-image translation for cross-domain disentanglement,” in *NeurIPS’2018*.
[99] R. Zhang, T. Che, Z. Ghahramani, Y. Bengio, and Y. Song, “MetaGAN: An adversarial approach to few-shot learning,” in *Advances in Neural Information Processing Systems*, vol. 31, 2018.
[100] T. Kim, J. Yoon, O. Dia, S. Kim, Y. Bengio, and S. Ahn, “Bayesian Model-Agnostic Meta-Learning,” in *NeurIPS’2018*.
[101] N. R. Ke, A. Goyal, O. Bilaniuk, J. Binas, M. C. Mozer, C. Pal, and Y. Bengio, “Sparse Attentive Backtracking: Temporal Credit Assignment Through Reminding,” in *NeurIPS’2018.*
[102] J. Sacramento, R. Ponte Costa, Y. Bengio, and W. Senn, “Dendritic error backpropagation in deep cortical microcircuits,” in *NeurIPS’2018*.
[103] M. Belghazi, A. Baratin, S. Rajeswar, S. Ozair, Y. Bengio, D. Hjelm, and A. Courville, “Mutual information neural estimation,” in *ICML’2018*.
[104] N.Ke,K. Zolna,A. Sordoni, Z. Lin,A. Trischler,Y. Bengio, J. Pineau, L. Charlin, and C. Pal, “Focused hierarchical rnns for conditional sequence processing,” in *ICML’2018*.

[105] F. Bordes, T. Berthier, L. D. Jorio, P. Vincent, and Y. Bengio, "Iteratively unveiling new regions of interest in deep learning models," in *Medical Imaging with Deep Learning, MIDL'2018*.
[106] M. Ravanelli, D. Serdyuk, and Y. Bengio, "Twin regularization for online speech recognition," in *Interspeech, 2018*.
[107] T. Parcollet, Y. Zhang, C. Trabelsi, M.Morchid, R. deMori, G. Linares, and Y. Bengio, "Quaternion convolutional neural networks for end-to-end automatic speech recognition," in *Interspeech, 2018*.
[108] R. D. Hjelm, A. P. Jacob, A. Trischler, T. Che, K. Cho, and Y. Bengio, "Boundary seeking GANs," in *ICLR'2018* (conference track).
[109] S. Subramanian, A. Trischler, Y. Bengio, and C. Pal, "Learning general purpose distributed sentence representations via large scale multi-task learning," in *ICLR'2018* (conference track).
[110] D. Serdyuk, N. R. Ke, A. Sordoni, A. Trischler, C. Pal, and Y. Bengio, "Twin networks: Matching the future for sequence generation," in *ICLR'2018* (conference track), 2018.
[111] K. Zolna, D. Arpit, D. Suhubdy, and Y. Bengio, "Fraternal dropout," in *ICLR'2018* (conference track).
[112] C. Trabelsi, O. Bilaniuk, Y. Zhang, D. Serdyuk, S. Subramanian, J. F. Santos, S.Mehri, N. Rostamzadeh, Y. Bengio, and C. Pal, "Deep complex networks," in *ICLR'2018* (conference track).
[113] P. Veli ̌ckovi ́c, G. C. Preixens, A. C. Paga, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," in *ICLR'2018* (conference track).
[114] S. Jastrzebski, D. Arpit, N. Ballas, V. Verma, T. Che, and Y. Bengio, "Residual connections encourage iterative inference," in *ICLR'2018* (conference track).
[115] S. Mimilakis, K. Drossos, J. Santos, G. Schuller, T. Virtanen, and Y. Bengio, "Monaural singing voice separation with skip-filtering connections and recurrent inference of time frequency mask," in Proc. *ICASSP'2018*.
[116] V. Thomas, E. Bengio, W. Fedus, J. Pondard, P. Beaudoin, H. Larochelle, J. Pineau, D. Precup, and Y. Bengio, "Disentangling the independently controllable factors of variation by interacting with the world," in *NIPS'2017*, p. arXiv :1802.09484, Feb. 2018.
[117] Ç. Gül.ehre, F. Dutil, A. Trischler, and Y. Bengio, "Plan, attend, generate: Planning for sequence-to-sequence models," in *NIPS'2017,* pp. 5480–5489, 2017. arxiv: 1706.05087.
[118] A. Lamb, D. R. Hjelm, Y. Ganin, J. P. Cohen, A. Courville, and Y. Bengio, "GibbsNet: Iterative adversarial inference for deep graphical models," in *NIPS'2017*, pp. 5095–5104, 2017.
[119] A. Goyal, A. Sordoni, M.-A. Côté, N. R. Ke, and Y. Bengio, "Z-forcing: Training stochastic recurrent networks," in *NIPS'2017*, pp. 6716–6726, 2017. arXiv: 1711.05411.
[120] I. V. Serban, A. Sordoni, R. Lowe, L. Charlin, J. Pineau, A. Courville, and Y. Bengio, "A hierarchical latent variable encoder-decoder model for generating dialogues," in *Thirty-First AAAI Conference on Artificial Intelligence*, p. 1583, 2017.
[121] L. Dinh, R. Pascanu, S. Bengio, and Y. Bengio, "Sharp minima can generalize for deep nets," in *Proceedings of the 34th International Conference on Machine Learning (ICML'17)* [315], pp. 1019–1028. arXiv :1703.04933.
[122] D. Krueger, N. Ballas, S. Jastrzebski, D. Arpit, M. S. Kanwal, T. Maharaj, E. Bengio, A. Fischer, A. Courville, S. Lacoste-Julien, and Y. Bengio, "A closer look at memorization in deep networks," in P*roceedings of the 34th International Conference on Machine Learning (ICML'17)* [315]. arxiv :1706.05394.
[123] M. Ravanelli, P. Brakel, M. Omologo, and Y. Bengio, "A network of deep neural networks for distant speech recognition," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, no. 1358, 2017.
[124] S. Jegou, M. Drozdzal, D. Vazquez, A. Romero, and Y. Bengio, "The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation," in *Workshop on Computer Vision in Vehicle Technology at CVPR 2017*.

[125] T. Kim, I. Song, and Y. Bengio, "Dynamic layer normalization for adaptive neural acoustic modeling in speech recognition," in *Interspeech 2017*, Aug. 2017.
[126] R. Lowe, M. Noseworthy, I. V. Serban, N. Angelard-Gontier, Y. Bengio, and J. Pineau, "Towards an automatic Turing test: Learning to evaluate dialogue responses," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, 2017*. Outstanding Paper award at ACL.
[127] A. Nguyen, J. Yosinski, Y. Bengio, A. Dosovitskiy, and J. Clune, "Plug & play generative networks: Conditional iterative generation of images in latent space," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
[128] Ç. Gül.ehre, M. Moczulski, F. Visin, and Y. Bengio, "Mollifying networks," in *International Conference on Learning Representations 2017* (Conference Track) [316].
[129] D. Warde-Farley and Y. Bengio, "Improving generative adversarial networks with denoising feature matching," in *International Conference on Learning Representations 2017* (Conference Track) [316].
[130] I. V. Serban, T. Klinger, G. Tesauro, K. Talamadupula, B. Zhou, Y. Bengio, and A. Courville, "Multiresolution recurrent neural networks: An application to dialogue response generation," *Thirty-First AAAI Conference on Artificial Intelligence*, p. 1641, 2017.
[131] D. Bahdanau, P. Brakel, K. Xu, A. Goyal, R. Lowe, J. Pineau, A. Courville, and Y. Bengio, "An actor-critic algorithm for sequence prediction," in *International Conference on Learning Representations 2017* (Conference Track) [316].
[132] J. Chung, S. Ahn, and Y. Bengio, "Hierarchical multiscale recurrent neural networks," in *International Conference on Learning Representations 2017* (Conference Track) [316].
[133] D. Krueger, T. Maharaj, J. Kramár, M. Pezeshki, N. Ballas, N. R. Ke, A. Goyal, Y. Bengio, A. Courville, and C. Pal, "Zoneout: Regularizing rnns by randomly preserving hidden activations," in *International Conference on Learning Representations 2017* (Conference Track) [316].
[134] Z. Lin, M. Feng, C. N. dos Santos, M. Yu, B. Xiang, B. Zhou, and Y. Bengio, "A structured self-attentive sentence embedding," in *International Conference on Learning Representations 2017* (Conference Track) [316].
[135] S. Mehri, K. Kumar, I. Gulrajani, R. Kumar, S. Jain, J. Sotelo, A. Courville, and Y. Bengio, "Samplernn: An unconditional end-to-end neural audio generation model," in *International Conference on Learning Representations 2017* (Conference Track) [316].
[136] A. Romero, P.-L. Carrier, A. Erraqabi, T. Sylvain, A. Auvolat, E. Dejoie,M.-A. Legault, M.-P. Dubé, J. G. Hussin, and Y. Bengio, "Diet networks : Thin parameters for fat genomic," in *International Conference on Learning Representations 2017* (Conference Track) [316].
[137] A. Lamb, A. Goyal, Y. Zhang, S. Zhang, A. Courville, and Y. Bengio, "Professor forcing: A new algorithm for training recurrent networks," in *NIPS'2016*.
[138] Y.Wu, S. Zhang, Y. Zhang, Y. Bengio, and R. Salakhutdinov, "On multiplicative integration with recurrent neural networks," in *NIPS'2016*.
[139] M. Arjovsky, A. Shah, and Y. Bengio. 2016. Unitary evolution recurrent neural networks. In Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48 (ICML'16).
[140] S. Zhang, Y. Wu, T. Che, Z. Lin, R. Memisevic, R. Salakhutdinov, and Y. Bengio, "Architectural complexity measures of recurrent neural networks," in *NIPS'2016*.
[141] I. Hubara, M. Courbariaux, D. Soudry, R. El-Yaniv, and Y. Bengio, "Binarized neural networks," in *NIPS'2016*.
[142] J. Chung, K. Cho, and Y. Bengio, "Nyu-mila neural machine translation systems for wmt'16," in *First Conference on Machine Translation*, 2016.
[143] Y. Zhang, M. Pezeshki, P. Brakel, S. Zhang, C. Laurent, Y. Bengio, and A. Courville, "Towards end-to-end speech recognition with deep convolutional neural networks," in *Interspeech* 2016, pp. 410–414, 2016.

[144] M. Ravanelli, P. Brakel, M. Omologo, and Y. Bengio, "Batch-normalized joint training for dnn-based distant speech recognition," in 2016 *IEEE Spoken Language Technology Workshop* (SLT), pp. 28–34, Dec. 2016
[145] M. Havaei, N. Guizard, N. Chapados, and Y. Bengio, "HeMIS: Hetero-modal image segmentation," in *Int. Conf. Medical Image Computing and Computer-Assisted Intervention*, MICCAI-2016, pp. 469–477, 2016.
[146] J. Serban, A. Sordoni, Y. Bengio, A. Courville, and J. Pineau, "Building end-to-end dialogue systems using generative hierarchical neural network models," in *AAAI'2016*.
[147] Y. Zhang, M. Pezeshki, P. Brakel, S. Zhang, C. Laurent, Y. Bengio, and A. Courville, "Towards end-to-end speech recognition with deep convolutional neural networks," in *Interspeech'2016*, pp. 410–414, 2016.
[148] D. B. J. Chorowski, D. Serdyuk, P. Brakel, and Y. Bengio, "End-to-end attention-based large vocabulary speech recognition," in *ICASSP'2016*, pp. 4945–4949, 2016.
[149] C. Laurent, G. Pereyra, P. Brakel, Y. Zhang, and Y. Bengio, "Batch normalized recurrent neural networks," in *ICASSP'2016*, pp. 2657–2661, 2016.
[150] J. Bornschein, S. Shabanian, A. Fischer, and Y. Bengio, "Training bidirectional Helmholtz machines," in *Proceedings of the 33nd International Conference on Machine Learning* (*ICML'2016*) [317]
[151] Ç. Gül.ehre, M. Moczulski, M. Denil, and Y. Bengio, "Noisy activation functions," in *Proceedings of the 33nd International Conference on Machine Learning* (*ICML'16*) [317].
[152] M. Pezeshki, L. Fan, P. Brakel, A. Courville, and Y. Bengio, "Deconstructing the ladder network architecture," in *Proceedings of the 33nd International Conference on Machine Learning (ICML'16)* [317].
[153] Z. Lin, M. Courbariaux, R. Memisevic, and Y. Bengio, "Neural networks with few multiplications," in *ICLR'2016*.
[154] J. Chung, K. Kastner, L. Dinh, K. Goel, A. Courville, and Y. Bengio, "A recurrent latent variable model for sequential data," in *NIPS'2015*.
[155] J. Chorowski, D. Bahdanau, D. Serdyuk, K. Cho, and Y. Bengio, "Attention-based models for speech recognition," in *NIPS'2015*.
[156] C. Gulcehre, K. Cho, R. Pascanu, and Y. Bengio, "Learned-norm pooling for deep feedforward and recurrent neural networks," in *Machine Learning and Knowledge Discovery in Databases* (*ECML/PKDD*), 2015.
[157] M. Courbariaux, Y. Bengio, and J.-P. David, "BinaryConnect: Training deep neural networks with binary weights during propagations," in *NIPS'2015*.
[158] Y. Dauphin, H. de Vries, and Y. Bengio, "Equilibrated adaptive learning rates for non-convex optimization," in *NIPS'2015*.
[159] A. Sordoni, Y. Bengio, H. Vahabi, C. Lioma, J. Grue Simonsen, and J.-Y. Nie, "A hierarchical recurrent encoder-decoder for generative context-aware query suggestion," in *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, pp. 553–562, 2015.
[160] S. Jean, O. Firat, K. Cho, R. Memisevic, and Y. Bengio, "Montreal neural machine translation systems for wmt15," in *Proceedings of the Tenth Workshop on Statistical Machine Translation*, pp. 134–140, 2015.
[161] S. Jean, K. Cho, R. Memisevic, and Y. Bengio, "On using very large target vocabulary for neural machine translation," in *ACL-IJCNLP'2015*. arXiv :1412.2007.
[162] J. Chung, Ç. Gül.ehre, K. Cho, and Y. Bengio, "Gated feedback recurrent neural networks," in *Proceedings of the 32nd International Conference on Machine Learning* (*ICML'1*5), pp. 2067–2075, 2015.
[163] D.-H. Lee, S. Zhang, A. Fischer, and Y. Bengio, "Difference target propagation," in *Machine Learning and Knowledge Discovery in Databases* (*ECML/PKDD*), 2015.

[164] L. Yao, S. Ozair, K. Cho, and Y. Bengio, "On the equivalence between deep nade and generative stochastic networks," in *Machine Learning and Knowledge Discovery in Databases*, 2014.
[165] J. Bornschein and Y. Bengio, "Reweighted wake-sleep," in *ICLR'2015*, arXiv: 1406.2751.
[166] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," in *ICLR'2015*, arXiv :1409.0473.
[167] A. Romero, N. Ballas, S. Ebrahimi Kahou, A. Chassang, C. Gatta, and Y. Bengio, "Fitnets: Hints for thin deep nets," in *ICLR'2015*, arXiv :1412.6550.
[168] G. Mesnil, S. Rifai, A. Bordes, X. Glorot, Y. Bengio, and P. Vincent, "Unsupervised learning of semantics of object detections for scene categorization," in *Pattern Recognition Applications and Methods, Advances in Intelligent Systems and Computing* (A. Fred and M. De Marsico, eds.), pp. 209–224, Springer International Publishing Switzerland, 2015.
[169] T. Raiko, L. Yao, K. Cho, and Y. Bengio, "Iterative neural autoregressive distribution estimator (NADE-k)," in *NIPS'2014*.
[170] [170] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" in *NIPS'2014*.
[171] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," in *NIPS'2014*.
[172] G. F. Montufar, R. Pascanu, K. Cho, and Y. Bengio, "On the number of linear regions of deep neural networks," in *NIPS'2014*.
[173] Y. Dauphin, R. Pascanu, C. Gulcehre, K. Cho, S. Ganguli, and Y. Bengio, "Identifying and attacking the saddle point problem in high-dimensional non-convex optimization," in *NIPS'2014*.
[174] A. Sordoni, Y. Bengio, and J.-Y. Nie, "Learning concept embeddings for query expansion by quantum entropy minimization," in *AAAI*, pp. 1586–1592, 2014.
[175] Y. Bengio, E. Thibodeau-Laufer, and J. Yosinski, "Deep generative stochastic networks trainable by backprop," in *ICML'2014*.
[176] M. Chen, K.Weinberger, F. Sha, and Y. Bengio, "Marginalized denoising auto-encoders for nonlinear representations," in *ICML'2014*.
[177] D. Warde-Farley, I. J. Goodfellow, A. Courville, and Y. Bengio, "An empirical analysis of dropout in piecewise linear networks," in *International Conference on Learning Representations* 2014 (Conference Track) [318].
[178] I. J. Goodfellow, M. Mirza, D. Xiao, A. Courville, and Y. Bengio, "An empirical investigation of catastrophic forgetting in gradient-based neural networks," in *International Conference on Learning Representations* 2014 (Conference Track) [318].
[179] Y. Bengio, L. Yao, and K. Cho, "Bounding the test log-likelihood of generative models," in *International Conference on Learning Representations* 2014 (Conference Track) [318].
[180] R. Pascanu, Ç. Gül.ehre, K. Cho, and Y. Bengio, "How to construct deep recurrent neural networks," in *International Conference on Learning Representations* 2014 (Conference Track) [318].
[181] R. Pascanu, G. Montufar, and Y. Bengio, "On the number of inference regions of deep feed forward networks with piece-wise linear activations," in *International Conference on Learning Representations* 2014 (Conference Track) [318].
[182] R. Pascanu and Y. Bengio, "Revisiting natural gradient for deep networks," in *International Conference on Learning Representations* 2014 (Conference Track) [318].
[183] I. J. Goodfellow, M. Mirza, A. Courville, and Y. Bengio, "Multi-prediction deep Boltzmann machines," in Advances in *Neural Information Processing Systems 26 (NIPS 2013)* [319].
[184] Y. Dauphin and Y. Bengio, "Stochastic ratiomatching of RBMs for sparse high-dimensional inputs," in *Advances in Neural Information Processing Systems 26 (NIPS 2013)* [319].

[185] Y. Bengio, L. Yao, G. Alain, and P. Vincent, "Generalized denoising auto-encoders as generative models," in *NIPS'2013*.
[186] N. Boulanger-Lewandowski, Y. Bengio, and P. Vincent, "High-dimensional sequence transduction," in Proc. *ICASSP 3*, 2013.
[187] I. J. Goodfellow, D. Warde-Farley, M. Mirza, A. Courville, and Y. Bengio, "Maxout networks," in *ICML'2013*, 2013.
[188] R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training recurrent neural networks," in *ICML'2013*.
[189] Y. Bengio, G. Mesnil, Y. Dauphin, and S. Rifai, "Better mixing via deep representations," in *Proceedings of the 30th International Conference on Machine Learning (ICML'13),* ACM, 2013.
[190] H. Luo, P. L. Carrier, A. Courville, and Y. Bengio, "Texture modeling with convolutional spike-and-slab RBMs and deep extensions," in *AISTATS'2013*.
[191] G. Mesnil, S. Rifai, A. Bordes, X. Glorot, Y. Bengio, and P. Vincent, "Unsupervised and transfer learning under uncertainty: from object detections to scene categorization," in *ICPRAM, 2013*.
[192] S. Rifai, Y. Bengio, Y. Dauphin, and P. Vincent, "A generative process for sampling contractive auto-encoders," in *Proceedings of the Twenty-nine International Conference on Machine Learning (ICML'12)* [320].
[193] N. Boulanger-Lewandowski, Y. Bengio, and P. Vincent, "Discriminative non-negative matrix factorization for multiple pitch estimation," in *ISMIR*, 2012.
[194] S. Rifai, Y. Bengio, A. Courville, P. Vincent, and M. Mirza, "Disentangling factors of variation for facial expression recognition," in *European Conference on Computer Vision*, 2012.
[195] A. Bordes, X. Glorot, J.Weston, and Y. Bengio, "Joint learning of words and meaning representations for open-text semantic parsing," *AISTATS'2012*.
[196] I. J. Goodfellow, A. Courville, and Y. Bengio, "Large-scale feature learning with spike-and-slab sparse coding," in *Proceedings of the Twenty-nine International Conference on Machine Learning (ICML'12)* [320].
[197] N. Boulanger-Lewandowski, Y. Bengio, and P. Vincent, "Modeling temporal dependencies in high-dimensional sequences: Application to polyphonic music generation and transcription," in *ICML'2012*.
[198] S. Rifai, Y. Dauphin, P. Vincent, Y. Bengio, and X. Muller, "The manifold tangent classifier," in *NIPS'2011*. Student paper award.
[199] J. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl, "Algorithms for hyper-parameter optimization," in *NIPS'2011*.
[200] G. Desjardins, A. Courville, and Y. Bengio, "On tracking the partition function," in *NIPS'2011*.
[201] O. Delalleau and Y. Bengio, "Shallow vs. deep sum-product networks," in *NIPS'2011*.
[202] Y. Bengio and O. Delalleau, "On the expressive power of deep architectures," in *ALT'2011*.
[203] Y. Bengio, "Deep learning of representations for unsupervised and transfer learning," in *JMLR W&CP: Proc. Unsupervised and Transfer Learning*, 2011.
[204] G. Mesnil, Y. Dauphin, X. Glorot, S. Rifai, Y. Bengio, I. Goodfellow, E. Lavoie, X. Muller, G. Desjardins, D.Warde-Farley, P. Vincent, A. Courville, and J. Bergstra, "Unsupervised and transfer learning challenge : a deep learning approach," in *JMLR W&CP: Proc. Unsupervised and Transfer Learning*, vol. 7, 2011.
[205] S. Rifai, G. Mesnil, P. Vincent, X. Muller, Y. Bengio, Y. Dauphin, and X. Glorot, "Higher order contractive auto-encoder," in *ECML PKDD*, 2011.
[206] A. Bordes, J. Weston, R. Collobert, and Y. Bengio, "Learning structured embeddings of knowledge bases," in *AAAI 2011*.
[207] A. Courville, J. Bergstra, and Y. Bengio, "Unsupervised models of images by spike-and-slab RBMs," in *ICML'2011*.

[208] Y. Dauphin, X. Glorot, and Y. Bengio, “Large-scale learning of embeddings with reconstruction sampling,” in *Proceedings of the Twenty-eighth International Conference on Machine Learning (ICML’11),* June 2011.
[209] S. Rifai, P. Vincent, X. Muller, X. Glorot, and Y. Bengio, “Contractive auto-encoders: Explicit invariance during feature extraction,” in *ICML’2011*.
[210] P. Hamel, S. Lemieux, Y. Bengio, and D. Eck, “Temporal pooling and multiscale learning for automatic annotation and ranking of music audio,” in *ISMIR’2011*.
[211] X. Glorot, A. Bordes, and Y. Bengio, “Domain adaptation for large-scale sentiment classification: A deep learning approach,” in *ICML’2011*.
[212] X. Glorot, A. Bordes, and Y. Bengio, “Deep sparse rectifier neural networks,” in *AISTATS’2011*.
[213] A. Courville, J. Bergstra, and Y. Bengio, “A spike and slab restricted Boltzmann machine,” in *JMLRW&CP: Proc. AISTATS’2011*, vol. 15, 2011. Recipient of People’s Choice Award.
[214] Y. Bengio, F. Bastien, A. Bergeron, N. Boulanger-Lewandowski, T. Breuel, Y. Chherawala, M. Cisse, M. Côté, D. Erhan, J. Eustache, X. Glorot, X. Muller, S. Pannetier Lebeuf, R. Pascanu, S. Rifai, F. Savard, and G. Sicard, “Deep learners benefit more from out-of-distribution examples,” in *JMLR W&CP: Proc. AISTATS’2011*.
[215] J. Bergstra, O. Breuleux, F. Bastien, P. Lamblin, R. Pascanu, G. Desjardins, J. Turian, D. Warde-Farley, and Y. Bengio, “Theano: a CPU and GPU math expression compiler,” in *Proc. SciPy*, 2010.
[216] M. Mandel, D. Eck, and Y. Bengio, “Learning tags that vary within a song,” in In *Proceedings of the 11th International Conference on Music Information Retrieval* (ISMIR), pp. 399–404, Aug. 2010.
[217] J. Turian, L. Ratinov, and Y. Bengio, “Word representations: A simple and general method for semi-supervised learning,” in *Proc. ACL’2010*, pp. 384–394, 2010.
[218] D. Erhan, A. Courville, Y. Bengio, and P. Vincent, “Why does unsupervised pre-training help deep learning?” in *JMLR W&CP: Proc. AISTATS’2010*, vol. 9, pp. 201–208, 2010.
[219] X. Glorot and Y. Bengio, “Understanding the difficulty of training deep feedforward neural networks,” in *JMLR W&CP: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics (AISTATS 2010*), vol. 9, pp. 249–256, May 2010.
[220] G. Desjardins, A. Courville, Y. Bengio, P. Vincent, and O. Delalleau, “Tempered Markov chain Monte Carlo for training of restricted Boltzmann machine,” in *AISTATS*, pp. 145–152, 2010.
[221] J. Bergstra and Y. Bengio, “Slow, decorrelated features for pretraining complex cell-like networks,” in *NIPS’2009*.
[222] A. Courville, D. Eck, and Y. Bengio, “An infinite factor model hierarchy via a noisy-or mechanism,” in *Neural Information Processing Systems Conference (NIPS)* 22, pp. 405–413, 2009.
[223] Y. Bengio, J. Louradour, R. Collobert, and J.Weston, “Curriculum learning,” in *ICML’2009*.
[224] D. Erhan, P.-A. Manzagol, Y. Bengio, S. Bengio, and P. Vincent, “The difficulty of training deep architectures and the effect of unsupervised pre-training,” in *Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics* (*AISTATS 2009*), pp. 153–160, Apr. 2009.
[225] J. Turian, J. Bergstra, and Y. Bengio, “Quadratic features and deep architectures for chunking,” in *Proc. NAACL-HLT’2009*, pp. 245–248, 2009.
[226] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, “Extracting and composing robust features with denoising autoencoders,” in *Cohen et al.* [321], pp. 1096–1103.
[227] H. Larochelle and Y. Bengio, “Classification using discriminative restricted Boltzmann machines,” in *Cohen et al.* [321], pp. 536–543.

[228] N. Le Roux, P.-A. Manzagol, and Y. Bengio, "Topmoumoute online natural gradient algorithm," in Advances in *Neural Information Processing Systems 20 (NIPS'2007*) (J. Platt, D. Koller, Y. Singer, and S. Roweis, eds.), pp. 849–856, Cambridge, MA: MIT Press, 2008.

[229] N. Le Roux, Y. Bengio, P. Lamblin, M. Joliveau, and B. Kégl, "Learning the 2-d topology of images," in *Platt et al.* [322], pp. 841–848.

[230] N. Chapados and Y. Bengio, "Augmented functional time series representation and forecasting with gaussian processes," in *Platt et al.* [322], pp. 265–272.

[231] H. Larochelle, D. Erhan, and Y. Bengio, "Zero-data learning of new tasks," in *AAAI Conference on Artificial Intelligence*, 2008.

[232] H. Larochelle, D. Erhan, A. Courville, J. Bergstra, and Y. Bengio, "An empirical evaluation of deep architectures on problems with many factors of variation," in *Proceedings of the 24th International Conference on Machine Learning (ICML'07*) (Z. Ghahramani, ed.), pp. 473–480, ACM, 2007.

[233] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, "Greedy layer-wise training of deep networks," in *Advances in Neural Information Processing Systems 19 (NIPS'06)* (B. Schölkopf, J. Platt, and T. Hoffman, eds.), pp. 153–160, MIT Press, 2007.

[234] N. Le Roux and Y. Bengio, "Continuous neural networks," in *Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics* (AISTATS'07), (San Juan, Porto Rico), Omnipress, Mar. 2007.

[235] N. Chapados and Y. Bengio, "Forecasting commodity contract spreads with Gaussian process," in *13th International Conference on Computing in Economics and Finance*, June 2007.

[236] J. Carreau andY. Bengio, "Ahybrid paretomodel for conditional density estimation of asymmetric fat-tail data," in *Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics (AISTATS'07),* (San Juan, Puerto Rico), Omnipress, Mar. 2007.

[237] Y. Bengio, N. Le Roux, P. Vincent, O. Delalleau, and P. Marcotte, "Convex neural networks," in *Advances in Neural Information Processing Systems 18 (NIPS'05)* (Y.Weiss,B. Schölkopf, and J. Platt, eds.), pp. 123–130, Cambridge, MA: MIT Press, 2006.

[238] Y. Bengio, O. Delalleau, and N. Le Roux, "The curse of highly variable functions for local kernel machines," in *Advances in Neural Information Processing Systems 18 (NIPS'2005)* (Y.Weiss, B. Schölkopf, and J. Platt, eds.), pp. 107–114, Cambridge, MA: MIT Press, 2006.

[239] Y. Bengio, H. Larochelle, and P. Vincent, "Non-local manifold Parzen windows," in *Advances in Neural Information Processing Systems 18 (NIPS'2005)* (Y.Weiss, B. Schölkopf and J. Platt, eds.), pp. 115–122,MIT Press, 2006.

[240] N. Chapados and Y. Bengio, "The k best-paths approach to approximate dynamic programming with application to portfolio optimization," in *AI06*, pp. 491–502, 2006

[241] Y. Grandvalet and Y. Bengio, "Semi-supervised Learning by Entropy Minimization," in *Advances in Neural Information Processing Systems 17 (NIPS'2004)* (L. Saul, Y. Weiss, and L. Bottou, eds.), (Cambridge, MA), MIT Press, Dec. 2005.

[242] F. Rivest, Y. Bengio, and J. Kalaska, "Brain inspired reinforcement learning," in *Advances in Neural Information Processing Systems 17 (NIPS'2004)* (L. Saul, Y.Weiss, and L. Bottou, eds.), (Cambridge, MA), pp. 1129–1136,MIT Press, Cambridge, 2005.

[243] Y. Bengio and M. Monperrus, "Non-local manifold tangent learning," in *Advances in Neural Information Processing Systems 17 (NIPS'2004)* (L. Saul, Y.Weiss, and L. Bottou, eds.), pp. 129–136, MIT Press, 2005.

[244] O. Delalleau, Y. Bengio, and N. Le Roux, "Efficient non-parametric function induction in semi-supervised learning," in *Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics (AISTATS'2005)* (R. G. Cowell and Z. Ghahramani, eds.), pp. 96–103, *Society for Artificial Intelligence and Statistics*, Jan. 2005.

[245] M. Ouimet and Y. Bengio, “Greedy spectral embedding,” in *Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics (AISTATS’2005)* (R. G. Cowell and Z. Ghahramani, eds.), pp. 253–260, 2005.

[246] F. Morin and Y. Bengio, “Hierarchical probabilistic neural network language model,” in *Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics (AISTATS’2005*) (R. G. Cowell and Z. Ghahramani, eds.), pp. 246–252, 2005.

[247] I. Bhattacharya, L. Getoor, and Y. Bengio, “Unsupervised sense disambiguation using bilingual probabilistic models,” in *Conference of the Association for Computational Linguistics (ACL’2004),* 2004.

[248] Y. Bengio, J.-F. Paiement, P. Vincent, O. Delalleau, N. Le Roux, and M. Ouimet, “Out-of-sample extensions for LLE, Isomap, MDS, Eigenmaps, and Spectral Clustering,” in *Advances in Neural Information Processing Systems 16 (NIPS’2003)* (S. Thrun, L. Saul and B. Schölkopf, eds.), MIT Press, 2004.

[249] Y. Bengio and Y. Grandvalet, “No unbiased estimator of the variance of k-fold cross-validation,” in *Advances in Neural Information Processing Systems* 16 (NIPS’03) (S. Thrun, L. Saul, and B. Schölkopf, eds.), (Cambridge, MA), MIT Press, Cambridge, 2004.

[250] N. Boufaden, Y. Bengio, and G. Lapalme, “Approche statistique pour le repérage de mots informatifs dans les textes oraux,” in *TALN’2004*, *Traitement Automatique du Langage Naturel*, 2004.

[251] Y. Bengio and J.-S. Sénécal, “Quick training of probabilistic neural nets by importance sampling,” in *Proceedings of the conference on Artificial Intelligence and Statistics (AISTATS)*, 2003.

[252] P. Vincent and Y. Bengio, “Manifold parzen windows,” in *Advances in Neural Information Processing Systems 15 (NIPS’2002)* (S. Becker and S. Thrun, eds.), (Cambridge, MA), pp. 825–832, MIT Press, 2003.

[253] R. Collobert, Y. Bengio, and S. Bengio, “Scaling large learning problems with hard parallel mixtures,” in *Pattern Recognition with Support Vector Machines* (S. Lee and A. Verri, eds.), vol. 2388 of *Lecture Notes in Computer Science*, pp. 8–23, Springer-Verlag, 2002.

[254] Y. Bengio, I. Takeuchi, and K. Kanamori, “The challenge of non-linear regression on large datasets with asymmetric heavy tails,” in *Proceedings of 2002 Joint Statistical Meetings*, pp. 193–205, American Statistical Association publ., 2002.

[255] Y. Bengio and N. Chapados, “Metric-based model selection for time-series forecasting,” in *Proc. of 2002 IEEE International Workshop on Neural Networks for Signal Processing*, pp. 13–24, IEEE Press, September 2002.

[256] P. Vincent and Y. Bengio, “K-local hyperplane and convex distance nearest neighbor algorithms,” in *Advances in Neural Information Processing Systems 14 (NIPS’2001)* (T. Dietterich, S. Becker, and Z. Ghahramani, eds.), (Cambridge, MA), pp. 985–992, MIT Press, 2002.

[257] R. Collobert, S. Bengio, and Y. Bengio, “A parallel mixture of SVMs for very large scale problems,” in *Advances in Neural Information Processing Systems 14 (NIPS’2001)* (T. Dietterich, S. Becker, and Z. Ghahramani, eds.), pp. 633–640, 2002.

[258] N. Chapados, Y. Bengio, P. Vincent, J. Ghosn, C. Dugas, I. Takeuchi, and L. Meng, “Estimating car insurance premia: a case study in high-dimensional data inference,” in *Advances in Neural Information Processing Systems 14 (NIPS’2001)* (T. Dietterich, S. Becker, and Z. Ghahramani, eds.), (Cambridge, MA), pp. 1369–1376, MIT Press, 2002.

[259] N. Boufaden, L. G., and Y. Bengio, “Topic segmentation: First stage of dialogue-based information extraction process,” in *Proceedings of the Natural Language Pacific Rim Symposium*, NLPRS-01, 2001.

[260] Y. Bengio, R. Ducharme, and P. Vincent, “A neural probabilistic language model,” in *Advances in Neural Information Processing Systems 13 (NIPS’2000)* (T. Leen, T. Dietterich, and V. Tresp, eds.), pp. 932–938, MIT Press, 2001.

[261] C. Dugas, Y. Bengio, F. Bélisle, and C. Nadeau, "Incorporating second-order functional knowledge for better option pricing," in *Advances in Neural Information Processing Systems 13 (NIPS'2000)* (T. Leen, T. Dietterich, and V. Tresp, eds.), pp. 472–478,MIT Press, 2001.
[262] Y. Bengio, "Probabilistic neural network models for sequential data," *in International Joint Conference on Neural Networks (IJCNN*), vol. 5, pp. 79–84, 2000.
[263] Y. Bengio, "Continuous optimization of hyper-parameters," in *International Joint Conference on Neural Networks* (*IJCNN*), vol. V, pp. 305–310, 2000.
[264] J. Ghosn and Y. Bengio, "Bias learning, knowledge sharing," in *International Joint Conference on Neural Networks* (*IJCNN*), vol. I, pp. 9–14, 2000.
[265] P. Vincent and Y. Bengio, "A neural support vector network architecture with adaptive kernels," in *International Joint Conference on Neural Networks* (*IJCNN*), vol. 5, pp. 5187–5192, 2000.
[266] N. Chapados and Y. Bengio, "VaR-based asset allocation using neural networks," in *Computational Finance*, 2000.
[267] F. Gingras, Y. Bengio, and C. Nadeau, "On out-of-sample statistics for time-series," in *Computational Finance*, 2000.
[268] Y. Bengio and S. Bengio, "Modeling high-dimensional discrete data with multi-layer neural networks," in *Advances in Neural Information Processing Systems* 12 (NIPS'99) (S. Solla, T. Leen, and K.-R. Müller, eds.), pp. 400–406, MIT Press, 2000.
[269] C. Nadeau and Y. Bengio, "Inference for the generalization error," in Advances in Neural Information Processing Systems 12 (NIPS'99) (S. Solla, T. Leen, and K.-R. Müller, eds.), pp. 307–313, MIT Press, 2000.
[270] S. Pigeon and Y. Bengio, "Binary pseudowavelets and application to bilevel image processing," in *Proceedings of the Data Compression Conference*, DCC'1999.
[271] Y. Bengio, S. Bengio, J. F. Isabelle, and Y. Singer, "Shared context probabilistic transducers," in *Advances in Neural Information Processing Systems* 10 (NIPS'1997) (M. Jordan, M. Kearns, and S. Solla, eds.), pp. 409–415, MIT Press, 1998.
[272] M. Bonneville, J. Meunier,Y. Bengio, and J. Soucy, "Support vector machines for improving the classification of brain pet images," *in SPIE Medical Imaging* 1998, (San Diego), 1998.
[273] H. Schwenk and Y. Bengio, "Training methods for adaptive boosting of neural networks," in *Advances in Neural Information Processing Systems 10 (NIPS'1997*) (M. Jordan, M. Kearns, and S. Solla, eds.), pp. 647–653, MIT Press, 1998.
[274] P. Haffner, L. Bottou, P.G.Howard, P. Simard,Y. Bengio, andY. L. Cun, "Browsing through high quality document images with DjVu," in *Proceedings of the Advances in Digital Libraries Conference (ADL'1998*), (Washington, DC, USA), p. 309, IEEE Computer Society, 1998.
[275] L. Bottou, P. G. Howard, and Y. Bengio, "The Z-coder adaptive binary coder," in *Proceedings of the Conference on Data Compression* (*DCC'98*), (Washington, DC, USA), p. 13, IEEE Computer Society, 1998.
[276] S. Pigeon and Y. Bengio, "A memory-efficient adaptive Huffman coding algorithm for very large sets of symbols," in *Proceedings of the Conference on Data Compression* (DCC'1998), p. 568, 1998.
[277] Y. LeCun, L. Bottou, and Y. Bengio, "Reading checks with multilayer graph transformer networks," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'1997),* vol. 1, pp. 151–154, Apr 1997.
[278] M. Rahim, Y. Bengio, and Y. LeCun, "Discriminative feature and model design for automatic speech recognition," in *In Proc. of Eurospeech*, pp. 75–78, 1997.
[279] L. Bottou, Y. Bengio, and Y. LeCun, "Global training of document processing systems using graph transformer networks," in *Proceedings of the Computer Vision and Pattern Recognition Conference* (CVPR'1997), (Puerto Rico), pp. 490–494, IEEE, 1997.

[280] H. Schwenk and Y. Bengio, "AdaBoosting neural networks: Application to on-line character recognition," in *International Conference on Artificial Neural Networks*, pp. 967–972, Springer Verlag, 1997.
[281] J. Ghosn and Y. Bengio, "Multi-task learning for stock selection," in *Advances in Neural Information Processing Systems 9 (NIPS'1996)* (M. Mozer, M. Jordan, and T. Petsche, eds.), pp. 946–952, MIT Press, Cambridge, MA, 1997.
[282] Y. Bengio, "Training a neural network with a financial criterion rather then a prediction criterion," in *Proceedings of the Fourth International Conference on Neural Networks in the Capital Markets (NNCM-96)* (A. Weigend, Y. Abu-Mostafa, and A.-P. Regenes, eds.), pp. 433–443, World Scientific, 1997.
[283] S. Bengio and Y. Bengio, "An EMalgorithm for asynchronous input/output hidden Markov models," in *International Conference on Neural Information Processing* (L. Xu, ed.), (Hong-Kong), pp. 328–334, 1996.
[284] Y. Bengio and F. Gingras, "Recurrent neural networks for missing or asynchronous data," in *Touretzky et al.* [323], pp. 395–401.
[285] S. El Hihi and Y. Bengio, "Hierarchical recurrent neural networks for long-term dependencies," in *Touretzky et al.* [323].
[286] Y. Bengio and P. Frasconi, "An input/output HMM architecture," in *Tesauro et al.* [324], pp. 427–434.
[287] Y. Bengio and P. Frasconi, "Diffusion of credit in Markovian models," in *Tesauro et al.* [324], pp. 553–560.
[288] L. Bottou and Y. Bengio, "Convergence properties of the K-means algorithm," in *Tesauro et al.* [324], pp. 585–592.
[289] S. Bengio, Y. Bengio, and J. Cloutier, "Use of genetic programming for the search of a new learning rule for neural networks," in *Proceedings of the First IEEE Conference on Evolutionary Computation*, pp. 324–327 vol.1, Jun 1994.
[290] P. Frasconi and Y. Bengio, "An EM approach to grammatical inference: Input/output HMMs," in *International Conference on Pattern Recognition (ICPR'1994*), (Jerusalem 1994), pp. 289–294, 1994.
[291] Y. LeCun and Y. Bengio, "Word-level training of a handritten word recognizer based on convolutional neural networks," in *International Conference on Pattern Recognition* (ICPR'94) (IEEE, ed.), (Jerusalem) 1994.
[292] Y. Bengio and Y. LeCun, "Word normalization for on-line handwritten word recognition," in *International Conference on Pattern Recognition (ICPR'94*), pp. 409-413, 1994.
[293] Y. Bengio, Y. LeCun, and D. Henderson, "Globally trained handwritten word recognizer using spatial representation, space displacement neural networks and hidden Markov models," in *Cowan et al.* [325], pp. 937–944.
[294] Y. Bengio and P. Frasconi, "Credit assignment through time: Alternatives to backpropagation," in *Cowan et al.* [325], pp. 75–82.
[295] Y. LeCun, Y. Bengio, D. Henderson, and A.Weisbuch, "On-line handwriting recognition with neural networks: spatial representation versus temporal representation," in *Proceedings of the International Conference on Handwriting and Drawing*, 1993.
[296] Y. Bengio, P. Frasconi, M. Gori, and G. Soda, "Recurrent neural networks for adaptive temporal processing," in *Proc. of the 6th Italian Workshop on Neural Networks*, WIRN-93 (E. Caianello, ed.), (Vietri, Italy), pp. 1183–1195, World Scientific Publ., 1993.
[297] S. Bengio, Y. Bengio, J. Cloutier, and J. Gecsei, "Generalization of a parametric learning rule," in *Proceedings of the International Conference on Artificial Neural Networks* 1993 (S. Gielen and B. Kappen, eds.), (Amsterdam, The Netherlands), pp. 502–502, Springer-Verlag, 1993.
[298] Y. Bengio, P. Frasconi, and P. Simard, "The problem of learning long-term dependencies in recurrent networks," in *IEEE International Conference on Neural Networks*, (San Francisco), pp. 1183–1195, IEEE Press, 1993. (invited paper).

[299] Y. Bengio, R. De Mori, G. Flammia, and R. Kompe, "Neural network - gaussian mixture hybrid for speech recognition or density estimation," in *Advances in Neural Information Processing Systems 4 (NIPS'91*) (J. E. Moody, S. J. Hanson, and R. P. Lipmann, eds.), (Denver, CO), pp. 175–182, Morgan Kaufmann, 1992.
[300] S. Bengio, Y. Bengio, J. Cloutier, and J. Gecsei, "Aspects théoriques de l'optimisation d'une règle d'apprentissage," in *Actes de la conférence Neuro-Nîmes 1992*, (Nîmes, France), 1992.
[301] Y. Bengio, S. Bengio, J. Cloutier, and J. Gecsei, "On the optimization of a synaptic learning rule," in in *Conference on Optimality in Biological and Artificial Networks*, 1992.
[302] Y. Bengio, R. De Mori, G. Flammia, and R. Kompe, "Global optimization of a neural network - hidden markov model hybrid," in *International Joint Conference on Neural Networks* (*IJCNN*), vol. 2, pp. 789–794, 1991.
[303] Y. Bengio, R. De Mori, G. Flammia, and R. Kompe, "A comparative study of hybrid acoustic phonetic decoders based on artificial neural networks," in *Proceedings of EuroSpeech*, 1991.
[304] Y. Bengio, S. Bengio, J. Cloutier, and J. Gecsei, "Learning a synaptic learning rule," in *International Joint Conference on Neural Networks* (*IJCNN*), pp. II–A969, 1991.
[305] Y. Bengio, R.D.Mori,G. Flammia, and R.Kompe, "Acomparative study on hybrid acoustic phonetic decoders based on artificial neural networks," in *Proceeding of EuroSpeech*, 1991.
[306] Y. Bengio, R. De Mori, and M. Gori, "Experiments on automatic speech recognition using bps," in *Parallel Architectures and Neural Networks* (E. Caianello, ed.), pp. 223–232, World Scientific Publ., 1990.
[307] Y. Bengio, R. Cardin, R. De Mori, and Y. Normandin, "A hybrid coder for hidden Markov models using a recurrent neural network," in *International Conference on Acoustics, Speech and Signal Processing*, (Albuquerque, NM), pp. 537–540, 1990.
[308] Y. Bengio, Y. Pouliot, S. Bengio, and P. Agin, "A neural network to detect homologies in proteins," in *Touretzky* [326], pp. 423–430.
[309] Y. Bengio, R. Cardin, and R. De Mori, "Speaker independent speech recognition with neural networks and speech knowledge," in *Touretzky* [326], pp. 218–225.
[310] Y. Bengio, P. Cosi, R. Cardin, and R. D. Mori, "Use of multi-layered networks for coding speech with phonetic features," in *Advances in Neural Information Processing Systems 1 (NIPS'88)* (D. Touretzky, ed.), (Denver, CO), pp. 224–231, Morgan Kaufmann, San Mateo, 1989.
[311] R. De Mori, Y. Bengio, and P. Cosi, "On the generalization capability of multilayered networks in the extraction of speech properties," in *Proceedings of the International Joint Conference on Artificial Intelligence*, (Detroit), pp. 1531–1536, IEEE, 1989.
[312] M. Gori, Y. Bengio, and R. DeMori, "BPS: A learning algorithm for capturing the dynamical nature of speech," in *International Joint Conference on Neural Networks* (*IJCNN*), (Washington D.C.), pp. 643–644, IEEE, New York, 1989.
[313] Y. Bengio, R. Cardin, P. Cosi, and R. De Mori, "Speech coding with multi-layer networks," in *International Conference on Acoustics, Speech and Signal Processing*, (Glasgow, Scotland), pp. 164–167, 1989.
[314] Y. Bengio and R. De Mori, "Use of neural networks for the recognition of place of articulation," in *International Conference on Acoustics, Speech and Signal Processing*, (New-York, NY), pp. 103–106, 1988.

Livres et chapitres de livres avec comité de lecture

[315] I. J. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[316] Y. Bengio, “Evolving culture vs local minima,” in *Growing Adaptive Machines: Integrating Development and Learning in Artificial Neural Networks*, no. also as ArXiv 1203.2990v1, pp. T. Kowaliw, N. Bredeche & R. Doursat, eds., Springer-Verlag, Mar. 2013.
[317] Y. Bengio and A. Courville, “Deep learning of Representations,” in *Handbook on Neural Information Processing*, vol. 49, Springer: Berlin Heidelberg, 2013.
[318] Y. Bengio, *Learning Deep Architectures for AI*. Now Publishers, 2009.
[319] Y. Bengio and Y. LeCun, “Scaling learning algorithms towards AI,” in *Large Scale Kernel Machines* (L. Bottou, O. Chapelle, D. DeCoste, and J.Weston, eds.),MIT Press, 2007.
[320] Y. Grandvalet and Y. Bengio, “Entropy Regularization,” in *Semi-Supervised Learning* (O. Chapelle, B. Schölkopf, and A. Zien, eds.), pp. 151–168, MIT Press, 2006.
[321] Y. Bengio, O. Delalleau, and N. Le Roux, “Label propagation and quadratic criterion,” in *Semi-Supervised Learning* (O. Chapelle, B. Schölkopf, and A. Zien, eds.), pp. 193–216, MIT Press, 2006.
[322] O. Delalleau, Y. Bengio, and N. Le Roux, “Large-scale algorithms,” in *Semi-Supervised Learning* (O. Chapelle, B. Schölkopf, and A. Zien, eds.), pp. 333–341, MIT Press, 2006.
[323] Y. Bengio, O. Delalleau, N. Le Roux, J.-F. Paiement, P. Vincent, and M. Ouimet, “Spectral dimensionality reduction,” in *Feature Extraction, Foundations and Applications* (I. Guyon, S. Gunn, M. Nikravesh, and L. Zadeh, eds.), Springer, 2006.
[324] Y. Bengio and Y. Grandvalet, “Bias in estimating the variance of k-fold cross-validation,” in *Statistical Modeling and Analysis for Complex Data Problem* (P. Duchesne and B. Remillard, eds.), pp. 75–95, Kluwer: Lawrence Erlbaum, 2004.
[325] C. Dugas, Y. Bengio, N. Chapados, P. Vincent, G. Denoncourt, and C. Fournier, “Statistical learning algorithms applied to automobile insurance ratemaking,” in *Intelligent and Other Computational Techniques in Insurance: Theory and Applications* (L. Jain and A. Shapiro, eds.), World Scientific Publishing Company, 2004.
[326] E. Trentin, F. Brugnara, Y. Bengio, C. Furlanello, and R. D. Mori, “Statistical and neural network models for speech recognition,” in *Connectionist Approaches to Clinical Problems in Speech and Language* (R. Daniloff, ed.), pp. 213–264, Lawrence Erlbaum, 2002.
[327] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” in *Intelligent Signal Processing*, pp. 306–351, IEEE Press, 2001. chap. 9.
[328] J. Schmidhuber, S. Hochreiter, and Y. Bengio, “Evaluating benchmark problems by random guessing,” in *Field Guide to Dynamical Recurrent Networks* (J. Kolen and S. Kremer, eds.), IEEE Press, 2001.
[329] S. Hochreiter, F. F. Informatik, Y. Bengio, P. Frasconi, and J. Schmidhuber, “Gradient flow in recurrent nets: the difficulty of learning long-term dependencies,” in *Field Guide to Dynamical Recurrent Networks* (J. Kolen and S. Kremer, eds.), IEEE Press, 2000.
[330] Y. LeCun, P. Haffner, L. Bottou, and Y. Bengio, “Object recognition with gradient-based learning,” in *Shape, Contour and Grouping in Computer Vision*, pp. 319–345, Springer, 1999.
[331] Y. Bengio, *Neural Networks for Speech and Sequence Recognition*. London, UK: International Thompson Computer Press, 1996.
[332] Y. LeCun and Y. Bengio, “Convolutional networks for images, speech, and time-series,” in *The Handbook of Brain Theory and Neural Networks* (M. A. Arbib, ed.), pp. 255–257, MIT Press, 1995.
[333] Y. LeCun and Y. Bengio, “Pattern recognition and neural networks,” in *The Handbook of Brain Theory and Neural Networks* (M. A. Arbib, ed.), pp. 711–714, MIT Press, 1995.
[334] Y. Bengio, “Radial basis functions for speech recognition,” in *Speech Recognition and Understanding: Recent Advances, Trends and Applications*, pp. 293–298, NATO Advanced Study Institute Series F: Computer and Systems Sciences, 1990.

[335] Y. Bengio and R. DeMori, “Speech coding with multilayer networks,” in *Neurocomputing: Algorithms, Architectures and Applications* (F. Fogelman Soulie and J. Herault, eds.), pp. 207–216, NATO Advanced Study Institute Series F: Computer and Systems Sciences, 1990.
[336] R. De Mori, Y. Bengio, and P. Cosi, “On the use of an ear model and multi-layer networks for automatic speech recognition,” in *Structural Pattern Analysis* (R. Mohr, T. Pavlidis, and A. Sanfelin, eds.), World Scientific, 1990.
[337] Y. Bengio and R. De Mori, “Connectionist models and their application to automatic speech recognition,” in *Artificial Neural Networks and Statistical Pattern Recognition: Old and New Connections* (I. K. Sethi and A. K. Jain, eds.), pp. 175–192, Elsevier, Machine Intelligence and Pattern Recognition Series, 1990.

Brevets

[338] Y. Bengio, L. Bottou, and P. G. Howard, “Z-coder: a fast-adaptive binary arithmetic coder.” U.S. Patent 6,188,334, February 13, 2001, along with patents 6,225,925, 6,281,817, and 6,476,740, 2001.
[339] Y. Bengio, L. Bottou, and Y. LeCun, “Module for constructing trainable modular network in which each module outputs and inputs data structured as a graph.” U.S. Patent 6,128,606, October 3, 2000.
[340] Y. Bengio, Y. LeCun, C. Nohl, and C. Burges, “Visitor registration system using automatic handwriting recognition.” Patent submitted in the U.S.A. in October 1994, submission number 1-16-18-1, 1994.