# Recent Advances in Deep Learning: An Overview

**Matiur Rahman Minar** MINAR09.BD@GMAIL.COM
**Jibon Naher** JIBON.NAHER09@GMAIL.COM
*Department of Computer Science and Engineering*
*Chittagong University of Engineering and Technology*
*Chittagong-4349, Bangladesh*

**Editor:**



## Abstract

Deep Learning is one of the newest trends in Machine Learning and Artificial Intelligence research. It is also one of the most popular scientific research trends now-a-days. Deep learning methods have brought revolutionary advances in computer vision and machine learning. Every now and then, new and new deep learning techniques are being born, outperforming state-of-the-art machine learning and even existing deep learning techniques. In recent years, the world has seen many major breakthroughs in this field. Since deep learning is evolving at a huge speed, its kind of hard to keep track of the regular advances especially for new researchers. In this paper, we are going to briefly discuss about recent advances in Deep Learning for past few years.

**Keywords:** Neural Networks, Machine Learning, Deep Learning, Recent Advances, Overview.

## 1. Introduction

The term "Deep Learning" (DL) was first introduced to Machine Learning (ML) in 1986, and later used for Artificial Neural Networks (ANN) in 2000 (Schmidhuber, 2015). Deep learning methods are composed of multiple layers to learn features of data with multiple levels of abstraction (LeCun et al., 2015). DL approaches allow computers to learn complicated concepts by building them out of simpler ones (Goodfellow et al., 2016). For Artificial Neural Networks (ANN), Deep Learning (DL) aka hierarchical learning (Deng and Yu, 2014) is about assigning credits in many computational stages accurately, to transform the aggregate activation of the network (Schmidhuber, 2014). To learn complicated functions, deep architectures are used with multiple levels of abstractions i.e. non-linear operations; e.g. ANNs with many hidden layers (Bengio, 2009). To sum it accurately, Deep Learning is a sub-field of Machine Learning, which uses many levels of non-linear information processing and abstraction, for supervised or unsupervised feature learning and representation, classification and pattern recognition (Deng and Yu, 2014).

Deep Learning i.e. Representation Learning is class or sub-field of Machine Learning. Recent deep learning methods are mostly said to be developed since 2006 (Deng, 2011). This paper is an overview of most recent techniques of deep learning, mainly recommended for upcoming researchers in this field. This article includes the basic idea of DL, major approaches and methods, recent breakthroughs and applications.

Overview papers are found to be very beneficial, especially for new researchers in a particular field. It is often hard to keep track with contemporary advances in a research area, provided that field has great value in near future and related applications. Now-a-days, scientific research is an attractive profession since knowledge and education are more shared and available than ever. For a technological research trend, its only normal to assume that there will be numerous advances and improvements in various ways. An overview of an particular field from couple years back, may turn out to be obsolete today.

Considering the popularity and expansion of Deep Learning in recent years, we present a brief overview of Deep Learning as well as Neural Networks (NN), and its major advances and critical breakthroughs from past few years. We hope that this paper will help many novice researchers in this field, getting an overall picture of recent Deep Learning researches and techniques, and guiding them to the right way to start with. Also we hope to pay some tributes by this work, to the top DL and ANN researchers of this era, Geoffrey Hinton (Hinton), Juergen Schmidhuber (Schmidhuber), Yann LeCun (LeCun), Yoshua Bengio (Bengio) and many others who worked meticulously to shape the modern Artificial Intelligence (AI). Its also important to follow their works to stay updated with state-of-the-art in DL and ML research.

In this paper, firstly we will provide short descriptions of the past overview papers on deep learning models and approaches. Then, we will start describing the recent advances of this field. We are going to discuss Deep Learning (DL) approaches, deep architectures i.e. Deep Neural Networks (DNN) and Deep Generative Models (DGM), followed by important regularization and optimization methods. Also, there are two brief sections for open-source DL frameworks and significant DL applications. Finally, we will discuss about current status and the future of Deep Learning in the last two sections i.e. Discussion and Conclusion.

## 2. Related works

There were many overview papers on Deep Learning (DL) in the past years. They described DL methods and approaches in great ways as well as their applications and directions for future research. Here, we are going to brief some outstanding overview papers on deep learning.

讨论了主要用于自然语言处理（NLP）的DL模型和体系结构。 他们展示了DL在各个NLP领域中的应用，比较了DL模型，并讨论了可能的未来趋势

Young et al. (2017) talked about DL models and architectures, mainly used in Natural Language Processing (NLP). They showed DL applications in various NLP fields, compared DL models, and discussed possible future trends.

Zhang et al. (2017) discussed state-of-the-art deep learning techniques for front-end and back-end speech recognition systems.

讨论了前端和后端语音识别系统的最新深度学习技术。

介绍了有关DL的最新技术概述。 他们还讨论了用于深度学习的开源DL框架和其他技术细节。

Zhu et al. (2017) presented overview on state-of-the-art of DL for remote sensing. They also discussed open-source DL frameworks and other technical details for deep learning.

以时序方式描述了深度学习模型的演变。 简要介绍了模型，并介绍了DL研究的突破。 如果想以进化的方式了解深度学习的起源，那么这篇论文将是不错的阅读。 他们还提到了神经网络的优化和未来的研究

Wang et al. (2017a) described the evolution of deep learning models in time-series manner. The briefed the models graphically along with the breakthroughs in DL research. This paper would be a good read to know the origin of the Deep Learning in evolutionary manner. They also mentioned optimization and future research of neural networks.

详细讨论了深度网络和生成模型。 他们从机器学习（ML）的基础知识，深度架构的优缺点出发，全面总结了最近的DL研究和应用。

Goodfellow et al. (2016) discussed deep networks and generative models in details. Starting from Machine Learning (ML) basics, pros and cons for deep architectures, they concluded recent DL researches and applications thoroughly.

发布了使用卷积神经网络（CNN）和递归神经网络（RNN）的深度学习（DL）模型的概述。 他们从表示学习的角度描述了DL，展示了DL技术是如何工作的以及如何在各种应用中成功使用，并基于无监督学习（UL）预测了未来的学习。 他们还指出了参考文献中DL的重大进展的文章

LeCun et al. (2015) published a overview of Deep Learning (DL) models with Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN). They described DL from the perspective of Representation Learning, showing how DL techniques work and getting used successfully in various applications, and predicting future learning based on Unsupervised Learning (UL). They also pointed out the articles of major advances in DL in the bibliography.

深入研究了深度学习以及CNN，RNN和深度强化学习（RL）。 他强调了序列处理RNN，同时指出了基本DL和NN的局限性以及改进它们的技巧

Schmidhuber (2015) did a generic and historical overview of Deep Learning along with CNN, RNN and Deep Reinforcement Learning (RL). He emphasized on sequence-processing RNNs, while pointing out the limitations of fundamental DL and NNs, and the tricks to improve them.

详细描述了神经网络以及代码和示例。 他还在某种程度上讨论了深度神经网络和深度学习。

Nielsen (2015) described the neural networks in details along with codes and examples. He also discussed deep neural networks and deep learning to some extent.

涵盖了基于时间进展的神经网络的历史和演变，机器学习方法以及在神经网络中深度学习的用途。

Schmidhuber (2014) covered history and evolution of neural networks based on time progression, categorized with machine learning approaches, and uses of deep learning in the neural networks.

Deng and Yu (2014) described deep learning classes and techniques, and applications of DL in several areas. 描述了深度学习课程和技术，以及DL在多个领域的应用。

从表示学习的角度对DL算法（即有监督和无监督的网络，优化和训练模型）进行了快速概述。 他专注于深度学习的许多挑战，例如 大型模型和数据的缩放算法，减少优化难度，设计有效的缩放方法等，以及乐观的DL研究

Bengio (2013) did quick overview on DL algorithms i.e. supervised and unsupervised networks, optimization and training models from the perspective of representation learning. He focused on many challenges of Deep Learning e.g. scaling algorithms for larger models and data, reducing optimization difficulties, designing efficient scaling methods etc. along with optimistic DL researches.

讨论了表示和特征学习又称为深度学习。 他们从应用，技术和挑战的角度探索了各种方法和模型。

Bengio et al. (2013) discussed on Representation and Feature Learning aka Deep Learning. They explored various methods and models from the perspectives of applications, techniques and challenges.

Deng (2011) gave an overview of deep structured learning and its architectures from the perspectives of information processing and related fields. 从信息处理和相关领域的角度概述了深度结构化学习及其架构

Arel et al. (2010) provided a short overview on recent DL techniques. 提供了有关最新DL技术的简短概述

Bengio (2009) discussed deep architectures i.e. neural networks and generative models for AI. 讨论了深度架构，即AI的神经网络和生成模型。

All recent overview papers on Deep Learning (DL) discussed important things from several perspectives. It is necessary to go through them for a DL researcher. However, DL is a highly flourishing field right now. Many new techniques and architectures are invented, even after the most recently published overview paper on DL. Also, previous papers focus from different perspectives. Our paper is mainly for the new learners and novice researchers who are new to this field. For that purpose, we will try to give a basic and clear idea of deep learning to the new researchers and anyone interested in this field.

## 3. Recent Advances

In this section, we will discuss the main recent Deep Learning (DL) approaches derived from Machine Learning and brief evolution of Artificial Neural Networks (ANN), which is the most common form used for deep learning.

### 3.1 Evolution of Deep Architectures

Artificial Neural Networks (ANN) have come a long way, as well as other deep models. First generation of ANNs was composed of simple neural layers for Perceptron. They were limited in simples computations. Second generation used Backpropagation to update weights of neurons according to error rates. Then Support Vector Machine (SVM) surfaced, and surpassed ANNs for a while. To overcome the limitations of backpropagation, Restricted 受限玻尔兹曼机 Boltzmann Machine was proposed, making the learning easier. Other techniques and neural networks came as well e.g. Feedforward Neural Networks (FNN), Convolutional Neural Netowrks (CNN), Recurrent Neural Networks (RNN) etc. along with Deep Belief Networks, Autoencoders and such (Hinton, *The next generation of neural networks*). From that point, ANNs got improved and designed in various ways and for various purposes.

Schmidhuber (2014), Bengio (2009), Deng and Yu (2014), Goodfellow et al. (2016), Wang et al. (2017a) etc. provided detailed overview on the evolution and history of Deep Neural Networks (DNN) as well as Deep Learning (DL). Deep architectures are multilayer non-linear repetition of simple architectures in most of the cases, which helps to obtain highly complex functions out of the inputs (LeCun et al., 2015).

## 4. Deep Learning Approaches

Deep Neural Networks (DNN) gained huge success in Supervised Learning (SL). Also, Deep Learning (DL) models are immensely successful in Unsupervised, Hybrid and Reinforcement Learning as well (LeCun et al., 2015).

### 4.1 Deep Supervised Learning

当标记数据并将分类器用于类或数字预测时，将应用监督学习。

Supervised learning are applied when data is labeled and the classifier is used for class or numeric prediction. LeCun et al. (2015) provided a brief yet very good explanation of supervised learning approach and how deep architectures are formed. Deng and Yu (2014) mentioned many deep networks for supervised and hybrid learning and explained them e.g. Deep Stacking Network (DSN) and its variants. Schmidhuber (2014) covered all neural networks starting from early neural networks to recently successful Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Long Short Term Memory (LSTM) and their improvements.

### 4.2 Deep Unsupervised Learning

When input data is not labeled, unsupervised learning approach is applied to extract features from data and classify or label them. LeCun et al. (2015) predicted future of deep learning in unsupervised learning. Schmidhuber (2014) described neural networks for unsupervised learning as well. Deng and Yu (2014) briefed deep architectures for unsupervised learning and explained deep Autoencoders in detail. 简要介绍了用于无监督学习的深度架构，并详细解释了深度自动编码器。

### 4.3 Deep Reinforcement Learning 深度强化学习

Reinforcement learning uses reward and punishment system for the next move generated by the learning model. This is mostly used for games and robots, solves usually decision making

强化学习对学习模型所产生的下一步行动使用奖惩系统。 这主要用于游戏和机器人，通常解决决策问题（Li，2017年）。
Schmidhuber（2014）描述了强化学习（RL）中深度学习的进展以及RL的深度前馈神经网络（FNN）和递归神经网络（RNN）的使用。
Li（2017）讨论了深度强化学习（DRL）及其架构，例如 深度Q网络（DQN）及其在各个领域中的应用。

problems (Li, 2017). Schmidhuber (2014) described advances of deep learning in Reinforcement Learning (RL) and uses of Deep Feedforward Neural Netowrk (FNN) and Recurrent Neural Network (RNN) for RL. Li (2017) discussed Deep Reinforcement Learning(DRL), its architectures e.g. Deep Q-Network (DQN), and applications in various fields.

Mnih et al. (2016) proposed a DRL framework using asynchronous gradient descent for DNN optimization. （2016）提出了一种使用异步梯度下降进行DNN优化的DRL框架。

van Hasselt et al. (2015) proposed a DRL architecture using deep neural network (DNN).
2015年）提出了使用深度神经网络（DNN）的DRL体系结构

## 5. Deep Neural Networks

In this section, we will briefly discuss about the deep neural networks (DNN), and recent improvements and breakthroughs of them. Neural networks work with functionalities similar to human brain. These are composed on neurons and connections mainly. When we are saying deep neural network, we can assume there should be quite a number of hidden layers, which can be used to extract features from the inputs and to compute complex functions. Bengio (2009) explained neural networks for deep architectures e.g. Convolutional Neural Networks (CNN), Auto-Encoders (AE) etc. and their variants. Deng and Yu (2014) detailed some neural network architectures e.g. AE and its variants. Goodfellow et al. (2016) wrote and skillfully explained about Deep Feedforward Networks, Convolutional Networks, Recurrent and Recursive Networks and their improvements. Schmidhuber (2014) mentioned full history of neural networks from early neural networks to recent successful techniques.

解释了用于深度架构的神经网络

详细介绍了一些神经网络架构，例如 AE及其变体

详细介绍了这几种网络以及他们的进步

提到了神经网络的完整历史，从早期的神经网络到最近成功的技术

### 5.1 Deep Autoencoders 深度自动编码器

Autoencoders (AE) are neural networks (NN) where outputs are the inputs. AE takes the original input, encodes for compressed representation and then decodes to reconstruct the input (Wang). In a deep AE, lower hidden layers are used for encoding and higher ones for decoding, and error back-propagation is used for training (Deng and Yu, 2014). Goodfellow et al. (2016)

#### 5.1.1 Variational Autoencoders 变异自动编码器

Variational Auto-Encoders (VAE) can be counted as decoders (Wang). VAEs are built upon standard neural networks and can be trained with stochastic gradient descent (Doersch, 2016)

#### 5.1.2 Stacked Denoising Autoencoders 堆叠式降噪自动编码器

In early Auto-Encoders (AE), encoding layer had smaller dimensions than the input layer. In Stacked Denoising Auto-Encoders (SDAE), encoding layer is wider than the input layer (Deng and Yu, 2014).

#### 5.1.3 Transforming Autoencoders 转换自动编码器

Deep Auto-Encoders (DAE) can be transformation-variant, i.e., the extracted features from multilayers of non-linear processing could be changed due to learner. Transforming Auto-Encoders (TAE) work with both input vector and target output vector to apply transformation-invariant property and lead the codes towards a desired way (Deng and Yu, 2014).

### 5.2 Deep Convolutional Neural Networks

Four basic ideas make the Convolutional Neural Networks (CNN), i.e., local connections, shared weights, pooling, and using many layers. First parts of a CNN are made of convolutional and pooling layers and latter parts are mainly fully connected layers. Convolutional layers detect local conjunctions from features and pooling layers merge similar features into one (LeCun et al., 2015). CNNs use convolutions instead of matrix multiplication in the convolutional layers (Goodfellow et al., 2016).

卷积层从特征中检测出局部连词，而池化层将相似的特征合并为一个特征CNN在卷积层中使用卷积代替矩阵乘法

Krizhevsky et al. (2012) presented a Deep Convolutional Neural Network (CNN) architecture, also known as AlexNet, which was a major breakthrough in Deep Learning (DL). The network composed of five convolutional layers and three fully connected layers. The architecture used Graphics Processing Units (GPU) for convolution operation, Rectified Linear Units (ReLU) as activation function and Dropout (Srivastava et al., 2014) to reduce overfitting.

Iandola et al. (2016) proposed a small CNN architecture called SqueezeNet.

Szegedy et al. (2014) proposed a Deep CNN architecture named Inception. An improvement of Inception-ResNet is proposed by Dai et al. (2017).

Redmon et al. (2015) proposed a CNN architecture named YOLO (You Only Look Once) for unified and real-time object detection. 用于统一和实时的对象检测。

一种可视化CNN中活动的方法。

Zeiler and Fergus (2013) proposed a method for visualizing the activities within CNN.

Gehring et al. (2017) proposed a CNN architecture for sequence-to-sequence learning. 用于序列到序列学习的CNN架构

Bansal et al. (2017) proposed PixelNet, using pixels for representations.

Goodfellow et al. (2016) explained the basic CNN architecures and the ideas. Gu et al. (2015) presented a nice overview on recent advances of CNNs, multiple variants of CNN, its architectures, regularization methods and functionality, and applications in various fields.

#### 5.2.1 Deep Max-Pooling Convolutional Neural Networks

Max-Pooling Convolutional Neural Networks (MPCNN) operate on mainly convolutions and max-pooling, especially used in digital image processing. MPCNN generally consists of three types of layers other than the input layer. Convolutional layers take input images and generate maps, then apply non-linear activation function. Max-pooling layers down-sample images and keep the maximum value of a sub-region. And fully-connected layers does the linear multiplication (Masci et al., 2013a). In Deep MPCNN, convolutional and max-pooling layers are used periodically after the input layer, followed by fully-connected layers (Giusti et al., 2013).

最大池卷积神经网络（MPCNN）主要对卷积和最大池进行操作，特别是在数字图像处理中使用。 MPCNN通常由输入层以外的三种类型的层组成。 卷积层获取输入图像并生成地图，然后应用非线性激活函数。 最大池化层对图像进行下采样，并保持子区域的最大值。 全连接层进行线性乘法在Deep MPCNN中，卷积和最大池层在输入层之后定期使用，然后是全连接层

#### 5.2.2 Very Deep Convolutional Neural Networks 超深度卷积神经网络

Simonyan and Zisserman (2014b) proposed Very Deep Convolutional Neural Network (VD-CNN) architecture, also known as VGG Nets. VGG Nets use very small convolution filters and depth to 16-19 weight layers.

Conneau et al. (2016) proposed another VDCNN architecture for text classification which uses small convolutions and pooling. They claimed this architecture is the first

VDCNN to be used in text processing which works at the character level. This architecture is composed of 29 convolution layers.

### 5.3 Network In Network

Lin et al. (2013) proposed Network In Network (NIN). NIN replaces convolution layers of traditional Convolutional Neural Network (CNN) by micro neural networks with complex structures. It uses multi-layer perceptron (MLPConv) for micro neural networks and global average pooling layer instead of fully connected layers. Deep NIN architectures can be made from multi-stacking of this proposed NIN structure (Lin et al., 2013).

NIN通过具有复杂结构的微神经网络取代了传统卷积神经网络（CNN）的卷积层。它使用用于多层神经网络的多层感知器（MLPConv）和全局平均池层，而不是完全连接的层。可以通过对该提议的NIN结构进行多次堆叠来构建深度NIN架构

### 5.4 Region-based Convolutional Neural Networks 基于区域的卷积神经网络

Girshick et al. (2014) proposed Region-based Convolutional Neural Network (R-CNN) which uses regions for recognition. R-CNN uses regions to localize and segment objects. This architecture consists of three modules i.e. category independent region proposals which defines the set of candidate regions, large Convolutional Neural Network (CNN) for extracting features from the regions, and a set of class specific linear Support Vector Machines (SVM) (Girshick et al., 2014).

#### 5.4.1 Fast R-CNN

提出了基于快速区域的卷积网络。该方法利用R-CNN架构并产生快速结果。快速R-CNN由卷积和池化层，区域提案以及一系列完全连接的层组成

Girshick (2015) proposed Fast Region-based Convolutional Network (Fast R-CNN). This method exploits R-CNN (Girshick et al., 2014) architecture and produces fast results. Fast R-CNN consists of convolutional and pooling layers, proposals of regions, and a sequence of fully connected layers (Girshick, 2015).

#### 5.4.2 Faster R-CNN

基于区域的快速卷积神经网络，该网络使用区域提议网络（RPN）进行实时目标检测。 RPN是一个完全卷积的网络，可准确高效地生成区域建议

Ren et al. (2015) proposed Faster Region-based Convolutional Neural Networks (Faster R-CNN), which uses Region Proposal Network (RPN) for real-time object detection. RPN is a fully convolutional network which generates region proposals accurately and efficiently (Ren et al., 2015).

#### 5.4.3 Mask R-CNN

提出了基于遮罩区域的卷积网络（Mask R-CNN）实例对象分割。 Mask R-CNN扩展了Faster R-CNN体系结构，并为对象蒙版使用了额外的分支

He et al. (2017) proposed Mask Region-based Convolutional Network (Mask R-CNN) instance object segmentation. Mask R-CNN extends Faster R-CNN (Ren et al., 2015) architecture, and uses an extra branch for object mask (He et al., 2017).

#### 5.4.4 Multi-Expert R-CNN

提出了基于多专家区域的卷积神经网络，该网络利用了快速R-CNN架构。 ME R-CNN通过选择性和详尽搜索生成兴趣区域（RoI）。 此外，它使用每个RoI多专家网络而不是单个每个RoI网络。 每个专家都是Fast R-CNN的全连接层的相同架构

Lee et al. (2017) proposed Multi-Expert Region-based Convolutional Neural Networks (ME R-CNN), which exploits Fast R-CNN (Girshick, 2015) architecture. ME R-CNN generates Region of Interests (RoI) from selective and exhaustive search. Also it uses per-RoI multi-expert network instead of single per-RoI network. Each expert is the same architecture of fully connected layers from Fast R-CNN (Lee et al., 2017).

### 5.5 Deep Residual Networks 深度残留网络

He et al. (2015) proposed Residual Networks (ResNets) consists of 152 layers. ResNets have lower error and easily trained with Residual Learning. More deeper ResNets achieve more better performance (He). ResNets are considered an important advance in the field of Deep Learning.

#### 5.5.1 Resnet in Resnet

Targ et al. (2016) proposed Resnet in Resnet (RiR) which combines ResNets (He et al., 2015) and standard Convolutional Neural Networks (CNN) in a deep dual stream architecture (Targ et al., 2016).

#### 5.5.2 ResNeXt

Xie et al. (2016) proposed ResNeXt architecture. ResNext exploits ResNets (He et al., 2015) for repeating layers with split-transform-merge strategy (Xie et al., 2016).

### 5.6 Capsule Networks

提出了Capsule Networks（CapsNet），一种具有两个卷积层和一个完全连接层的体系结构CapsNet通常包含多个卷积层，最后在胶囊层上。 CapsNet被认为是深度学习的最新突破之一，因为据说这是在卷及神经网络的局限性之上建立的。 它使用胶囊层而不是神经元层，其中胶囊是一组神经元。

Sabour et al. (2017) proposed Capsule Networks (CapsNet), an architecture with two convolutional layers and one fully connected layer. CapsNet usually contains several convolution layers and on capsule layer at the end (Xi et al., 2017). CapsNet is considered as one of the most recent breakthrough in Deep Learning (Xi et al., 2017), since this is said to be build upon the limitations of Convolutional Neural Networks (Hinton). It uses layers of capsules instead of layers of neurons, where a capsule is a set of neurons. Active lower level capsules make predictions and upon agreeing multiple predictions, a higher level capsule becomes active. A routing-by-agreement mechanism is used in these capsule layers. An improvement of CapsNet is proposed with EM routing (Anonymous, 2018b) using Expectation-Maximization (EM) algorithm.

活跃的较低级别的胶囊会做出预测，并且在同意多个预测后，更高级别的胶囊会变得活跃。在这些胶囊层中使用按协议路由机制。提出了一种使用期望最大化（EM）算法的EM路由改进CapsNet的方法

### 5.7 Recurrent Neural Networks

递归神经网络（RNN）更适合于顺序输入，例如语音和文本以及生成序列。

RNN过去由于梯度消失和爆炸问题而难

Recurrent Neural Networks (RNN) are better suited for sequential inputs like speech and text and generating sequence. A Recurrent hidden unit can be considered as very deep feedforward network with same weights when unfolded in time. RNNs used to be difficult to train because of gradient vanishing and exploding problem (LeCun et al., 2015). Many improvements were proposed later to solve this problem.

提供了递归和递归神经网络和体系结构及其变体以及相关门控和内存网络的详细信息

Goodfellow et al. (2016) provided details of Recurrent and Recursive Neural Networks and architectures, its variants along with related gated and memory networks.

Karpathy et al. (2015) used character-level language models for analyzing and visualizing predictions, representations training dynamics, and error types of RNN and its variants e.g. LSTMs 使用字符级语言模型来分析和可视化RNN及其变体的预测，表示训练动态和错误类型

Józefowicz et al. (2016) explored RNN models and limitations for language modelling.

### 5.7.1 RNN-EM

提出了带有外部存储器的递归神经网络（RNNEM），以提高RNN的存储容量。比其他RNN更好地实现了最新的语言理解能力。

Peng and Yao (2015) proposed Recurrent Neural Networks with External Memory (RNN-EM) to improve memory capacity of RNNs. They claimed to achieve state-of-the-art in language understanding, better than other RNNs.

### 5.7.2 GF-RNN

提出的门控反馈递归神经网络（GF-RNN），通过将多个递归层与全局门控单元堆叠在一起，扩展了标准RNN

Chung et al. (2015) proposed Gated Feedback Recurrent Neural Networks (GF-RNN), which extends the standard RNN by stacking multiple recurrent layers with global gating units.

### 5.7.3 CRF-RNN

提出了将条件随机场作为递归神经网络（CRF-RNN）的方法，该方法结合了卷积神经网络（CNN）和条件随机场（CRF）进行概率图形建模

Zheng et al. (2015) proposed Conditional Random Fields as Recurrent Neural Networks (CRF-RNN), which combines the Convolutional Neural Networks (CNNs) and Conditional Random Fields (CRFs) for probabilistic graphical modelling.

### 5.7.4 Quasi-RNN

提出的准递归神经网络（QRNN）用于神经序列建模，跨时间步并行应用。

Bradbury et al. (2016) proposed Quasi Recurrent Neural Networks (QRNN) for neural sequence modelling, appling parallel across timesteps.

## 5.8 Memory Networks

提出了用于提问（QA）的内存网络。 内存网络由内存，输入特征图，泛化，输出特征图和响应组成

Weston et al. (2014) proposed Memory Networks for question answering (QA). Memory Networks are composed of memory, input feature map, generalization, output feature map and response (Weston et al., 2014) .

### 5.8.1 Dynamic Memory Networks

提出了用于QA任务的动态内存网络（DMN）。 DMN具有四个模块，即输入，问题，情节记忆，输出

Kumar et al. (2015) proposed Dynamic Memory Networks (DMN) for QA tasks. DMN has four modules i.e. Input, Question, Episodic Memory, Output (Kumar et al., 2015).

## 5.9 Augmented Neural Networks

很好地介绍了注意力和增强递归神经网络，即神经图灵机（NTM），注意力接口，神经程序员和自适应计算时间。 增强型神经网络通常由逻辑功能和标准神经网络体系结构等额外属性组成

Olah and Carter (2016) gave nice presentation of Attentional and Augmented Recurrent Neural Networks i.e. Neural Turing Machines (NTM), Attentional Interfaces, Neural Programmer and Adaptive Computation Time. Augmented Neural Networks are usually made of using extra properties like logic functions along with standard Neural Network architecture (Olah and Carter, 2016).

### 5.9.1 Neural Turing Machines

提出的神经图灵机（NTM）架构，由神经网络控制器和存储库组成。 NTM通常将RNN与外部内存库结合在一起

Graves et al. (2014) proposed Neural Turing Machine (NTM) architecture, consisting of a neural network controller and a memory bank. NTMs usually combine RNNs with external memory bank (Olah and Carter, 2016).

### 5.9.2 Neural GPU

神经GPU，解决了NTM的并行问题

Kaiser and Sutskever (2015) proposed Neural GPU, which solves the parallel problem of NTM (Graves et al., 2014).

### 5.9.3 Neural Random-Access Machines

神经随机存取机，它使用外部可变大小的随机存取存储器

Kurach et al. (2015) proposed Neural Random Access Machine, which uses an external variable-size random-access memory.

### 5.9.4 Neural Programmer

神经编程器，一种具有算术和逻辑功能的增强型神经网络。

Neelakantan et al. (2015) proposed Neural Programmer, an augmented neural network with arithmetic and logic functions.

### 5.9.5 Neural Programmer-Interpreters

可以学习的神经程序解释器（NPI）。 NPI由循环内核，程序存储器和特定于域的编码器组成

Reed and de Freitas (2015) proposed Neural Programmer-Interpreters (NPI) which can learn. NPI consists of recurrent core, program memory and domain-specific encoders (Reed and de Freitas, 2015).

## 5.10 Long Short Term Memory Networks

提出了长期短期记忆（LSTM），它克服了递归神经网络（RNN）的错误回流问题。 LSTM基于递归网络以及基于梯度的学习算法。LSTM引入了自环来产生路径，从而梯度可以流动

Hochreiter and Schmidhuber (1997) proposed Long Short-Term Memory (LSTM) which overcomes the error back-flow problems of Recurrent Neural Networks (RNN). LSTM is based on recurrent network along with gradient-based learning algorithm (Hochreiter and Schmidhuber, 1997) LSTM introduced self-loops to produce paths so that gradient can flow (Goodfellow et al., 2016).

Greff et al. (2017) provided large-scale analysis of Vanilla LSTM and eight LSTM variants for three uses i.e. speech recognition, handwriting recognition, and polyphonic music modeling. They claimed that eight variants of LSTM failed to perform significant improvement, while only Vanilla LSTM performs well (Greff et al., 2015).

提供了Vanilla LSTM和八个LSTM变体的大规模分析，用于三种用途，即语音识别，手写识别和和弦音乐建模。 他们声称，LSTM的八个变体均未取得重大改进，而只有Vanilla LSTM表现良好

Shi et al. (2016b) proposed Deep Long Short-Term Memory (DLSTM), which is a stack of LSTM units for feature mapping to learn representations (Shi et al., 2016b).

### 5.10.1 Batch-Normalized LSTM

提出了批量归一化LSTM（BN-LSTM），它对循环神经网络的隐藏状态使用了批量归一化。

Cooijmans et al. (2016) proposed batch-normalized LSTM (BN-LSTM), which uses batch-normalizing on hidden states of recurrent neural networks.

### 5.10.2 Pixel RNN

像素递归神经网络（PixelRNN），由多达十二个二维LSTM层组成

van den Oord et al. (2016b) proposed Pixel Recurrent Neural Networks (PixelRNN), made of up to twelve two-dimensional LSTM layers.

### 5.10.3 Bidirectional LSTM

建议的双向LSTM（BLSTM）递归网络与动态贝叶斯网络（DBN）一起用于上下文相关的关键字检测。

Wöllmer et al. (2010) proposed Bidirection LSTM (BLSTM) Recurrent Networks to be used with Dynamic Bayesian Network (DBN) for context-sensitive keyword detection.

#### 5.10.4 Variational Bi-LSTM

提出的变体Bi-LSTM，它是双向LSTM体系结构的变体。 变分Bi-LSTM使用变分自动编码器（VAE）在LSTM之间创建信息交换的渠道，以学习更好的表示形式

Shabanian et al. (2017) proposed Variational Bi-LSTMs, which is a variant of Bidirectional LSTM architecture. Variational Bi-LSTM creates a channel of information exchange between LSTMs using Variational Auto-Encoders (VAE), for learning better representations (Shabanian et al., 2017).

### 5.11 Googles Neural Machine Translation

提出了用于自动翻译的Google神经机器翻译（GNMT）系统，该系统结合了编码器网络，解码器网络和关注网络，遵循通用的序列到序列学习框架

Wu et al. (2016) proposed Googles Neural Machine Translation (GNMT) System for automated translation, which incorporates an encoder network, a decoder network and an attention network following the common sequence-to-sequence learning framework.

### 5.12 Fader Networks

Fader Networks（推子网络），一种新型的编码器-解码器体系结构，可通过更改属性值来生成输入图像的逼真的变化。

Lample et al. (2017) proposed Fader Networks, a new type of encoder-decoder architecture to generate realistic variations of input images by changing attribute values.

### 5.13 Hyper Networks

HyperNetworks（超网络），可以为其他神经网络生成权重，例如静态超网络，卷积网络，递归网络的动态超网络

Ha et al. (2016) proposed HyperNetworks which generates weights for other neural networks, such as static hypernetworks convolutional networks, dynamic hypernetworks for recurrent networks.

Deutsch (2018) used Hyper Networks for generating neural networks.

### 5.14 Highway Networks

公路网，它使用选通单元来学习调节信息。 跨几层的信息流称为信息高速公路

Srivastava et al. (2015) proposed Highway Networks, which uses gating units to learn regulating information through. Information flow across several layers are called information highways (Srivastava et al., 2015).

#### 5.14.1 Recurrent Highway Networks

循环公路网（RHN），扩展了长期短期记忆（LSTM）体系结构。RHN suse循环过渡内的高速公路层

Zilly et al. (2017) proposed Recurrent Highway Networks (RHN), which extend Long Short-Term Memory (LSTM) architecture. RHNs use Highway layers inside the recurrent transition (Zilly et al., 2017).

### 5.15 Highway LSTM RNN

高速公路长短期存储器（HLSTM）RNN，它扩展了具有门控方向连接的深LSTM网络，即高速公路，位于相邻层的存储单元之间

Zhang et al. (2016c) proposed Highway Long Short-Term Memory (HLSTM) RNN, which extends deep LSTM networks with gated direction connections i.e. Highways, between memory cells in adjacent layers.

### 5.16 Long-Term Recurrent CNN

长期递归卷积网络（LRCN），该方法使用CNN作为输入，然后使用LSTM进行递归序列建模并生成预测

Donahue et al. (2014) proposed Long-term Recurrent Convolutional Networks (LRCN), which uses CNN for inputs, then LSTM for recurrent sequence modeling and generating predictions.

### 5.17 Deep Neural SVM

深度神经支持向量机（DNSVM），它使用支持向量机（SVM）作为顶层在深度神经网络中进行分类

Zhang et al. (2015a) proposed Deep Neural Support Vector Machines (DNSVM), which uses Support Vector Machine (SVM) as the top layer for classification in a Deep Neural Network (DNN)

### 5.18 Convolutional Residual Memory Networks

卷积残差存储网络，它将存储机制合并到卷积神经网络（CNN）中。 它通过长期的短期记忆机制增强了卷积残差网络

Moniz and Pal (2016) proposed Convolutional Residual Memory Networks, which incorporates memory mechanism into Convolutional Neural Networks (CNN). It augments convolutional residual networks with a long short term memory mechanism (Moniz and Pal, 2016).

### 5.19 Fractal Networks

分形网络，即FractalNet，作为残差网络的替代方案。 他们声称可以训练超深度神经网络，而无需残留学习。 分形是由简单扩展规则生成的重复体系结构

Larsson et al. (2016) proposed Fractal Networks i.e. FractalNet, as an alternative to residual nets. They claimed to train ultra deep neural networks without residual learning. Fractals are repeated architecture generated by simple expansion rule (Larsson et al., 2016).

### 5.20 WaveNet

WaveNet，深度神经网络，用于生成原始音频。 WaveNet由一叠卷积层和用于输出的softmax分布层组成

van den Oord et al. (2016a) proposed WaveNet, deep neural network for generating raw audio. WaveNet is composed of a stack of convolutional layers, and softmax distribution layer for outputs (van den Oord et al., 2016a).

Rethage et al. (2017) proposed a WaveNet model for speech denoising.

提出了用于语音降噪的WaveNet模型

### 5.21 Pointer Networks

指针网络（Ptr-Nets），通过使用称为“指针”的softmax概率分布解决了表示可变字典的问题

Vinyals et al. (2017) proposed Pointer Networks (Ptr-Nets), which solves the problem of representing variable dictionaries by using a softmax probability distribution called ”Pointer”.

## 6. Deep Generative Models

深度生成模型

In this section, we will briefly discuss other deep architecures which uses multiple levels of abstraction and representation similar to deep neural networks, also known as Deep Generative Models (DGM). Bengio (2009) explained deep architectures e.g. Boltzmann Machines (BM) and Restricted Boltzmann Machines (RBM) etc. and their variants.

详细解释了深入生成模型，例如 受限制和不受限制的Boltzmann机器及其变体，Deep Boltzmann机器，Deep Belief网络（DBN），定向生成网络和生成随机网络等。

Goodfellow et al. (2016) explained deep generative models in details e.g. Restricted and Unrestricted Boltzmann Machines and their variants, Deep Boltzmann Machines, Deep Belief Networks (DBN), Directed Generative Nets, and Generative Stochastic Networks etc.

提出的辅助深度生成模型，其中使用辅助变量扩展了深度生成模型。 辅助变量通过随机层进行变异分布并跳过连接

Maaløe et al. (2016) proposed Auxiliary Deep Generative Models where they extended Deep Generative Models with auxiliary variables. The auxiliary variables make variational distribution with stochastic layers and skip connections (Maaløe et al., 2016).

Rezende et al. (2016) developed a class for one-shot generalization of deep generative models.

开发了一类用于深度生成模型的一次性概括。

### 6.1 Boltzmann Machines

Boltzmann机器是用于学习任意概率分布的连接主义方法，该方法使用最大似然原理进行学习

Boltzmann Machines are connectionist approach for learning arbitrary probability distributions which use maximum likelihood principle for learning (Goodfellow et al., 2016).

### 6.2 Restricted Boltzmann Machines

受限玻尔兹曼机（RBM）是一种特殊类型的马尔可夫随机场，包含一层随机隐藏单元，即潜在变量和一层可观察变量

Restricted Boltzmann Machines (RBM) are special type of Markov random field containing one layer of stochastic hidden units i.e. latent variables and one layer of observable variables (Deng and Yu (2014), Goodfellow et al. (2016)).

Hinton and Salakhutdinov (2011) proposed a Deep Generative Model using Restricted Boltzmann Machines (RBM) for document processing.

提出了使用受限玻尔兹曼机（RBM）进行文件处理的深度生成模型

### 6.3 Deep Belief Networks

Deep Belief Networks (DBN) are generative models with several layers of latent binary or real variables (Goodfellow et al., 2016).

深度信念网络（DBN）是具有多层潜在二进制或实变量的生成模型

Ranzato et al. (2011) built a deep generative model using Deep Belief Network (DBN) for images recognition.

使用深度信念网络（DBN）建立了用于图像识别的深度生成模型。

### 6.4 Deep Lambertian Networks

深度朗伯网络（DLN），该网络是一个多层生成模型，其中潜在变量为反照率，表面法线和光源。 DLN是朗伯反射率与高斯受限玻尔兹曼机和深信度网络的组合

Tang et al. (2012) proposed Deep Lambertian Networks (DLN) which is a multilayer generative model where latent variables are albedo, surface normals, and the light source. DLN is a combination of lambertian reflectance with Gaussian Restricted Boltzmann Machines and Deep Belief Networks (Tang et al., 2012).

### 6.5 Generative Adversarial Networks

提出了生成对抗网络（GAN），用于估算具有对抗过程的生成模型。 GAN体系结构由对抗对手的生成模型（即用于学习模型或数据分布的判别模型）组成（Goodfellow等，2014）。 Mao等人为GAN提出的一些其他改进

Goodfellow et al. (2014) proposed Generative Adversarial Nets (GAN) for estimating generative models with an adversarial process. GAN architecture is composed of a generative model pitted against an adversary i.e. a discriminative model to learn model or data distribution (Goodfellow et al., 2014). Some more improvements proposed for GAN by Mao et al. (2016), Kim et al. (2017) etc.

Salimans et al. (2016) presented several methods for training GANs.

#### 6.5.1 Laplacian Generative Adversarial Networks

Laplacian生成对抗网络（LAPGAN）的深度生成模型使用了拉普拉斯金字塔框架内的卷积网络

Denton et al. (2015) proposed a Deep Generative Model (DGM) called Laplacian Generative Adversarial Networks (LAPGAN) using Generative Adversarial Networks (GAN) approach. The model also uses convolutional networks within a Laplacian pyramid framework (Denton et al., 2015).

### 6.6 Recurrent Support Vector Machines

递归支持向量机（RSVM），它使用递归神经网络（RNN）从输入序列中提取特征，并使用标准支持向量机（SVM）进行序列级目标识别。

Shi et al. (2016a) proposed Recurrent Support Vector Machines (RSVM), which uses Recurrent Neural Network (RNN) for extracting features from input sequence and standard Support Vector Machine (SVM) for sequence-level objective discrimination.

## 7. Training and Optimization Techniques

训练和优化技术

In this section, we will provide short overview on some major techniques for regularization and optimization of Deep Neural Networks (DNN).

### 7.1 Dropout

Dropout来防止神经网络过度拟合。 Dropout是一种神经网络模型平均正则化方法，通过向其隐藏单元中添加噪声来实现。 在训练过程中，它会随机从神经网络中删除单位以及连接。 辍学可以与任何类型的神经网络一起使用，甚至可以在RBM等图形模型中使用。 最近提出的一种改进的辍学改进是递归神经网络的Fraternal Dropout

Srivastava et al. (2014) proposed Dropout to prevent neural networks from overfitting. Dropout is a neural network model-averaging regularization method by adding noise to its hidden units. It drops units from the neural network along with connections randomly during training. Dropout can be used with any kind of neural networks, even in graphical models like RBM (Srivastava et al., 2014). A very recent proposed improvement of dropout is Fraternal Dropout (Anonymous, 2018a) for Recurrent Neural Networks (RNN).

### 7.2 Maxout

提出了Maxout，这是一种与Dropout一起使用的新激活功能（Srivastava等，2014）。 Maxout的输出是一组输入中的最大值，这对Dropout的模型平均有利

Goodfellow et al. (2013) proposed Maxout, a new activation function to be used with Dropout (Srivastava et al., 2014). Maxout's output is the maximum of a set of inputs, which is beneficial for Dropout's model averaging (Goodfellow et al., 2013).

### 7.3 Zoneout

提出了Zoneout，一种递归神经网络（RNN）的正则化方法。 在进行类似于Dropout的训练时，Zoneout随机使用噪声（Srivastava等，2014），但保留隐藏的单位而不是删除

Krueger et al. (2016) proposed Zoneout, a regularization method for Recurrent Neural Networks (RNN). Zoneout uses noise randomly while training similar to Dropout (Srivastava et al., 2014), but preserves hidden units instead of dropping (Krueger et al., 2016).

### 7.4 Deep Residual Learning

提出了用于深度神经网络（DNN）的深度残差学习框架，该框架称为ResNet，具有较低的训练误差

He et al. (2015) proposed Deep Residual Learning framework for Deep Neural Networks (DNN), which are called ResNets with lower training error (He).

### 7.5 Batch Normalization

提出了批量归一化，一种通过减少内部协变量偏移来加速深度神经网络训练的方法。 Ioffe（2017）提出了批量重归一化的扩展方法。

Ioffe and Szegedy (2015) proposed Batch Normalization, a method for accelerating deep neural network training by reducing internal covariate shift. Ioffe (2017) proposed Batch Renormalization extending the previous approach.

### 7.6 Distillation

蒸馏技术，它是将知识从高度正则化的模型（即神经网络）集成到压缩的较小模型中。

Hinton et al. (2015) proposed Distillation, from transferring knowledge from ensemble of highly regularized models i.e. neural networks into compressed and smaller model.

### 7.7 Layer Normalization

提出的层归一化，用于加快深度神经网络（特别是RNN）的训练，并解决了批量归一化的局限性

Ba et al. (2016) proposed Layer Normalization, for speeding-up training of deep neural networks especially for RNNs and solves the limitations of batch normalization (Ioffe and Szegedy, 2015).

395/5000例如Theano（Bergstra等人，2011），Tensorflow（Abadi等人，2016），PyTorch，PyBrain（Schaul等人，2010），Caffe（Jia等人，2014），Blocks and Fuel（van Merri ¨enboer等，2015），CuDNN（Chetlur等，2014），Honk（Tang和Lin，2017），ChainerCV（Niitani等，2017），PyLearn2，Chainer，火炬，霓虹灯等.Bahrampour等 。 （2015）对几种深度学习框架进行了比较研究

## 8. Deep Learning frameworks

有很多可用于深度学习的开源库和框架。 它们大多数是为python编程语言而构建的。

There are a good number of open-source libraries and frameworks available for deep learning. Most of them are built for python programming language. Such as Theano (Bergstra et al., 2011), Tensorflow (Abadi et al., 2016), PyTorch, PyBrain (Schaul et al., 2010), Caffe (Jia et al., 2014), Blocks and Fuel (van Merriënboer et al., 2015), CuDNN (Chetlur et al., 2014), Honk (Tang and Lin, 2017), ChainerCV (Niitani et al., 2017), PyLearn2, Chainer, torch, neon etc.

Bahrampour et al. (2015) did a comparative study of several deep learning frameworks.

## 9. Applications of Deep Learning

在本节中，我们将简要讨论深度学习架构的一些近期杰出应用。自从深度学习（DL）开始以来，DL方法就以监督学习，无监督学习，半监督学习或强化学习的形式用于各个领域。 从分类和检测任务开始，DL应用程序在各个领域迅速传播

In this section, we will briefly discuss some recent outstanding applications of Deep Learning architectures. Since the beginning of Deep Learning (DL), DL methods are being used in various fields in forms of supervised, unsupervised, semi-supervised or reinforcement learning. Starting from classification and detection tasks, DL applications are spreading rapidly in every fields.

Such as -

- 图像分类与识别 image classification and recognition (Simonyan and Zisserman (2014b), Krizhevsky et al. (2012), He et al. (2015))
- 视频分类 video classification (Karpathy et al., 2014)
- 序列产生 sequence generation (Graves, 2013)
- 缺陷分类 defect classification (Masci et al., 2013b)
- text, speech, image and video processing (LeCun et al., 2015) 文字，语音，图像和视频处理
- 文字分类 text classification (Conneau et al., 2016)
- 语音处理 speech processing (Arel et al., 2009)
- 语音识别和口语理解 speech recognition and spoken language understanding (Hinton et al. (2012), Zhang et al. (2015b), Zhang et al. (2016c), Zhang et al. (2016c), Zhang et al. (2015a), Shi et al. (2016a), Mesnil et al. (2015), Peng and Yao (2015), Amodei et al. (2015))
- 文字转语音生成 text-to-speech generation (Wang et al. (2017b), Arik et al. (2017))
- 查询分类 query classification (Shi et al., 2016b)
- 句子分类 sentence classification (Kim, 2014)
- 句子建模 sentence modelling (Kalchbrenner et al., 2014)
- 字处理 word processing (Mikolov et al., 2013a)
- 前提选择 premise selection (Alemi et al., 2016)
- 文件和句子处理 document and sentence processing (Le and Mikolov (2014), Mikolov et al. (2013b))
- 生成图像标题 generating image captions (Vinyals et al. (2014), Xu et al. (2015))

- 摄影风格转移 photographic style transfer (Luan et al., 2017)
- 自然图像流形 natural image manifold (Zhu et al., 2016)
- 图像着色 image colorization (Zhang et al., 2016b)
- 图像问题解答 image question answering (Yang et al., 2015)
- 生成纹理和风格化的图像 generating textures and stylized images (Ulyanov et al., 2016)
- 视觉和文字问题解答 visual and textual question answering (Xiong et al. (2016), ?DBLP:journals/corr/AntolALMBZP15))
- 视觉识别和描述 visual recognition and description (Donahue et al. (2014), Razavian et al. (2014), Oquab et al. (2014))
- 目标检测 object detection (Lee et al. (2017), Ranzato et al. (2011), Redmon et al. (2015), Liu et al. (2015))
- 文件处理 document processing (Hinton and Salakhutdinov, 2011)
- 角色动作合成与编辑 character motion synthesis and editing (Holden et al., 2016)
- 歌唱综合 singing synthesis (Blaauw and Bonada, 2017)
- 人身识别 person identification (Li et al., 2014)
- 人脸识别和验证 face recognition and verification (Taigman et al., 2014)
- 视频中的动作识别 action recognition in videos (Simonyan and Zisserman, 2014a)
- 人体动作识别 human action recognition (Ji et al., 2013)
- 动作识别 action recognition (Sharma et al., 2015)
- 对运动捕捉序列进行分类和可视化 classifying and visualizing motion capture sequences (Cho and Chen, 2013)
- 笔迹生成和预测 handwriting generation and prediction (Carter et al., 2016)
- 自动化和机器翻译 automated and machine translation (Wu et al. (2016), Cho et al. (2014), Bahdanau et al. (2014), Hermann et al. (2015), Luong et al. (2015))
- 命名实体识别 named entity recognition (Lample et al., 2016)
- 移动视觉 mobile vision (Howard et al., 2017)
- 对话代理 conversational agents (Ghazvininejad et al., 2017)
- 称遗传变异 calling genetic variants (Poplin et al., 2016)
- 癌症检测 cancer detection (Cruz-Roa et al., 2013)
- X射线CT重建 X-ray CT reconstruction (Kang et al., 2016)
- 癫痫发作的预测 Epileptic Seizure Prediction (Mirowski et al., 2008)

硬件加速

- hardware acceleration (Han et al., 2016)

机器人技术

- robotics (Lenz et al., 2013)

to name a few.

提供了各种类别的DL应用程序的详细列表，例如 语音和音频处理，信息检索，对象识别和计算机视觉，多模式和多任务学习

Deng and Yu (2014) provided detailed lists of DL applications in various categories e.g. speech and audio processing, information retrieval, object recognition and computer vision, multimodal and multi-task learning etc.

使用深度强化学习（DRL）来掌握游戏已经成为当今的热门话题。 时不时地，由DNN和DRL创建的AI机器人仅经过数小时的培训，就在战略游戏和其他游戏中击败了人类世界的冠军和大师。

Using Deep Reinforcement Learning (DRL) for mastering games has become a hot topic now-a-days. Every now and then, AI bots created with DNN and DRL, are beating human world champions and grandmasters in strategical and other games, from only hours of training. For example, AlphaGo and AlphaGo Zero for game of GO (Silver et al. (2017b), Silver et al. (2016), Dong et al. (2017)), Dota2 (Batsford (2014)), Atari (Mnih et al. (2013),Mnih et al. (2015), van Hasselt et al. (2015)), Chess and Shougi (Silver et al., 2017a).

## 10. Discussion

尽管深度学习在许多领域都取得了巨大的成功，但还有很长的路要走。 还有很多房间需要改进。 至于限制，列表也很长。 例如，Nguyen等。 （2014年）表明，在识别图像时，很容易愚弄深度神经网络（DNN）。 还有其他问题，例如所学功能的可移植性（Yosinski等，2014）。 黄等。（2017）提出了一种用于神经网络的网络攻击的体系结构，他们认为防御这些攻击需要未来的工作。张等。 （2016a）提

Though Deep Learning has achieved tremendous success in many areas, it still has long way to go. There are many rooms left for improvement. As for limitations, the list is quite long as well. For example, Nguyen et al. (2014) showed that Deep Neural Networks (DNN) can be easily fooled while recognizing images. There are other issues like transferability of features learned (Yosinski et al., 2014). Huang et al. (2017) proposed an architecture for adersarial attacks on neural networks, where they think future works are needed for defenses against those attacks. Zhang et al. (2016a) presented an experimental framework for understanding deep learning models. They think understanding deep learning requires rethinking generalization.

Marcus（2018）对深度学习（DL），其功能，局限性和性质进行了重要回顾。 他强烈指出了DL方法的局限性，即需要更多数据，容量有限，无法处理层次结构，在开放式推理中苦苦挣扎，不够透明，无法与先验知识很好地集成以及无法 从因果关系中区分因果关系（Marcus，2018）。 他还提到DL假设世界稳定，工作近似，难以设计，并且存在过度宣传的潜在风险。 Marcus（2018）认为DL需要重新概念化，并在无监督学习，符号操作和混合模型中寻找可能性，并从认知科学和心理学领域获得见识并应对更大胆的挑战。

Marcus (2018) gave an important review on Deep Learning (DL), what it does, its limits and its nature. He strongly pointed out the limitations of DL methods, i.e., requiring more data, having limited capacity, inability to deal with hierarchical structure, struggling with open-ended inference, not being sufficiently transparent, not being well integrated with prior knowledge, and inability to distinguish causation from correlation (Marcus, 2018). He also mentioned that DL assumes stable world, works as approximation, is difficult to engineer and has potential risks as being an excessive hype. Marcus (2018) thinks DL needs to be reconceptualized and to look for possibilities in unsupervised learning, symbol manipulation and hybrid models, having insights from cognitive science and psychology and taking bolder challenges.

## 11. Conclusion

Although Deep Learning (DL) has advanced the world faster than ever, there are still ways to go. We are still away from fully understanding of how deep learning works, how we can get machines more smarter, close to or smarter than humans, or learning exactly like human. DL has been solving many problems while taking technologies to another dimension. However, there are many difficult problems for humanity to deal with. For example, people are still dying from hunger and food crisis, cancer and other lethal diseases etc. We hope deep learning and AI will be much more devoted to the betterment of humanity, to carry out the hardest scientific researches, and last but not the least, to make the world a more better place for every single human.


## Acknowledgments

We would like to thank Dr. Mohammed Moshiul Hoque, Professor, Department of CSE, CUET, for introducing us to the amazing world of Deep Learning.


## References


Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Gregory S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian J. Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Józefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Rajat Monga, Sherry Moore, Derek Gordon Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul A. Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda B. Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *CoRR*, abs/1603.04467, 2016.

Alexander A. Alemi, François Chollet, Geoffrey Irving, Christian Szegedy, and Josef Urban. Deepmath - deep sequence models for premise selection. *CoRR*, abs/1606.04442, 2016.

Dario Amodei, Rishita Anubhai, Eric Battenberg, Carl Case, Jared Casper, Bryan Catanzaro, Jingdong Chen, Mike Chrzanowski, Adam Coates, Greg Diamos, Erich Elsen, Jesse Engel, Linxi Fan, Christopher Fougner, Tony Han, Awni Y. Hannun, Billy Jun, Patrick LeGresley, Libby Lin, Sharan Narang, Andrew Y. Ng, Sherjil Ozair, Ryan Prenger, Jonathan Raiman, Sanjeev Satheesh, David Seetapun, Shubho Sengupta, Yi Wang, Zhiqian Wang, Chong Wang, Bo Xiao, Dani Yogatama, Jun Zhan, and Zhenyao Zhu. Deep speech 2: End-to-end speech recognition in english and mandarin. *CoRR*, abs/1512.02595, 2015.

Anonymous. Fraternal dropout. *International Conference on Learning Representations*, 2018a. URL `https://openreview.net/forum?id=SJyVzQ-C-`.

Anonymous. Matrix capsules with em routing. *International Conference on Learning Representations*, 2018b. URL `https://openreview.net/forum?id=HJWLfGWRb`.

Itamar Arel, Derek Rose, and Tom Karnowski. A deep learning architecture comprising homogeneous cortical circuits for scalable spatiotemporal pattern inference. In *in Proc. NIPS Workshop Deep Learn. Speech*, pages 1–8, 2009.

Itamar Arel, Derek C. Rose, and Thomas P. Karnowski. Research frontier: Deep machine learning–a new frontier in artificial intelligence research. *Comp. Intell. Mag.*, 5 (4):13–18, November 2010. ISSN 1556-603X. doi: 10.1109/MCI.2010.938364. URL `http://dx.doi.org/10.1109/MCI.2010.938364`.

Sercan Ömer Arik, Mike Chrzanowski, Adam Coates, Greg Diamos, Andrew Gibiansky, Yongguo Kang, Xian Li, John Miller, Jonathan Raiman, Shubho Sengupta, and Mohammad Shoeybi. Deep voice: Real-time neural text-to-speech. *CoRR*, abs/1702.07825, 2017.

Lei Jimmy Ba, Ryan Kiros, and Geoffrey E. Hinton. Layer normalization. *CoRR*, abs/1607.06450, 2016.

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *CoRR*, abs/1409.0473, 2014.

Soheil Bahrampour, Naveen Ramakrishnan, Lukas Schott, and Mohak Shah. Comparative study of caffe, neon, theano, and torch for deep learning. *CoRR*, abs/1511.06435, 2015.

Aayush Bansal, Xinlei Chen, Bryan C. Russell, Abhinav Gupta, and Deva Ramanan. Pixelnet: Representation of the pixels, by the pixels, and for the pixels. *CoRR*, abs/1702.06506, 2017.

Tom Batsford. Calculating optimal jungling routes in dota2 using neural networks and genetic algorithms. *Game Behaviour*, 1(1), 2014. URL `https://computing.derby.ac.uk/ojs/index.php/gb/article/view/14`.

Yoshua Bengio. URL `http://www.iro.umontreal.ca/~bengioy/yoshua_en/index.html`. MILA, University of Montreal, Quebec, Canada.

Yoshua Bengio. Learning deep architectures for ai. *Found. Trends Mach. Learn.*, 2(1):1–127, January 2009. ISSN 1935-8237. doi: 10.1561/2200000006. URL `http://dx.doi.org/10.1561/2200000006`.

Yoshua Bengio. Deep learning of representations: Looking forward. *CoRR*, abs/1305.0445, 2013.

Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning: A review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(8): 1798–1828, August 2013. ISSN 0162-8828. doi: 10.1109/TPAMI.2013.50. URL `http://dx.doi.org/10.1109/TPAMI.2013.50`.

James Bergstra, Olivier Breuleux, Frédéric Bastien, Pascal Lamblin, Razvan Pascanu, Guillaume Desjardins, Joseph P. Turian, David Warde-Farley, and Yoshua Bengio. Theano: A cpu and gpu math compiler in python. 2011.

Merlijn Blaauw and Jordi Bonada. A neural parametric singing synthesizer. *CoRR*, abs/1704.03809, 2017.

James Bradbury, Stephen Merity, Caiming Xiong, and Richard Socher. Quasi-recurrent neural networks. *CoRR*, abs/1611.01576, 2016.

Shan Carter, David Ha, Ian Johnson, and Chris Olah. Experiments in handwriting with a neural network. *Distill*, 2016. doi: 10.23915/distill.00004. URL `http://distill.pub/2016/handwriting`.

Sharan Chetlur, Cliff Woolley, Philippe Vandermersch, Jonathan Cohen, John Tran, Bryan Catanzaro, and Evan Shelhamer. cudnn: Efficient primitives for deep learning. *CoRR*, abs/1410.0759, 2014.

Kyunghyun Cho and Xi Chen. Classifying and visualizing motion capture sequences using deep neural networks. *CoRR*, abs/1306.3874, 2013.

Kyunghyun Cho, Bart van Merrienboer, Çaglar Gülçehre, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *CoRR*, abs/1406.1078, 2014.

Junyoung Chung, Caglar Gulcehre, Kyunghyun Cho, and Yoshua Bengio. Gated feedback recurrent neural networks. In *Proceedings of the 32Nd International Conference on International Conference on Machine Learning - Volume 37*, ICML'15, pages 2067–2075. JMLR.org, 2015. URL `http://dl.acm.org/citation.cfm?id=3045118.3045338`.

Alexis Conneau, Holger Schwenk, Loïc Barrault, and Yann LeCun. Very deep convolutional networks for text classification. *CoRR*, abs/1606.01781, 2016.

Tim Cooijmans, Nicolas Ballas, César Laurent, and Aaron C. Courville. Recurrent batch normalization. *CoRR*, abs/1603.09025, 2016.

Angel Alfonso Cruz-Roa, John Edison Arevalo Ovalle, Anant Madabhushi, and Fabio Augusto González Osorio. *A Deep Learning Architecture for Image Representation, Visual Interpretability and Automated Basal-Cell Carcinoma Cancer Detection*, pages 403–410. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013. ISBN 978-3-642-40763-5. doi: 10.1007/978-3-642-40763-5_50. URL `https://doi.org/10.1007/978-3-642-40763-5_50`.

Jifeng Dai, Haozhi Qi, Yuwen Xiong, Yi Li, Guodong Zhang, Han Hu, and Yichen Wei. Deformable convolutional networks. *CoRR*, abs/1703.06211, 2017.

Li Deng. An overview of deep-structured learning for information processing. 01 2011.

Li Deng and Dong Yu. Deep learning: Methods and applications. *Found. Trends Signal Process.*, 7(3–4):197–387, June 2014. ISSN 1932-8346. doi: 10.1561/2000000039. URL `http://dx.doi.org/10.1561/2000000039`.

Emily L. Denton, Soumith Chintala, Arthur Szlam, and Robert Fergus. Deep generative image models using a laplacian pyramid of adversarial networks. *CoRR*, abs/1506.05751, 2015.

Lior Deutsch. Generating neural networks with neural networks. *CoRR*, abs/1801.01952, 2018. URL `https://arxiv.org/abs/1801.01952`.

Carl Doersch. Tutorial on variational autoencoders. *CoRR*, 1606.05908, 2016. URL `https://arxiv.org/abs/1606.05908`.

Jeff Donahue, Lisa Anne Hendricks, Sergio Guadarrama, Marcus Rohrbach, Subhashini Venugopalan, Kate Saenko, and Trevor Darrell. Long-term recurrent convolutional networks for visual recognition and description. *CoRR*, abs/1411.4389, 2014.

Xiao Dong, Jiasong Wu, and Ling Zhou. Demystifying alphago zero as alphago GAN. *CoRR*, abs/1711.09091, 2017.

Jonas Gehring, Michael Auli, David Grangier, Denis Yarats, and Yann N. Dauphin. Convolutional sequence to sequence learning. *CoRR*, abs/1705.03122, 2017.

Marjan Ghazvininejad, Chris Brockett, Ming-Wei Chang, Bill Dolan, Jianfeng Gao, Wen-tau Yih, and Michel Galley. A knowledge-grounded neural conversation model. *CoRR*, abs/1702.01932, 2017.

Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.

Ross B. Girshick. Fast R-CNN. *CoRR*, abs/1504.08083, 2015.

Alessandro Giusti, Dan C. Ciresan, Jonathan Masci, Luca Maria Gambardella, and Jürgen Schmidhuber. Fast image scanning with deep max-pooling convolutional neural networks. In *IEEE International Conference on Image Processing, ICIP 2013, Melbourne, Australia, September 15-18, 2013*, pages 4034–4038, 2013. doi: 10.1109/ICIP.2013.6738831. URL `https://doi.org/10.1109/ICIP.2013.6738831`.

Ian Goodfellow, David Warde-Farley, Mehdi Mirza, Aaron Courville, and Yoshua Bengio. Maxout networks. In Sanjoy Dasgupta and David McAllester, editors, *Proceedings of the 30th International Conference on Machine Learning*, volume 28 of *Proceedings of Machine Learning Research*, pages 1319–1327, Atlanta, Georgia, USA, 17–19 Jun 2013. PMLR. URL `http://proceedings.mlr.press/v28/goodfellow13.html`.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 2672–2680. Curran Associates, Inc., 2014. URL `http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf`.

Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

Alex Graves. Generating sequences with recurrent neural networks. *CoRR*, abs/1308.0850, 2013.

Alex Graves, Greg Wayne, and Ivo Danihelka. Neural turing machines. *CoRR*, abs/1410.5401, 2014.

Klaus Greff, Rupesh Kumar Srivastava, Jan Koutník, Bas R. Steunebrink, and Jürgen Schmidhuber. LSTM: A search space odyssey. *CoRR*, abs/1503.04069, 2015.

Klaus Greff, Rupesh Kumar Srivastava, Jan Koutník, Bas R. Steunebrink, and Jürgen Schmidhuber. LSTM: A search space odyssey. *IEEE Trans. Neural Netw. Learning Syst.*, 28(10):2222–2232, 2017. doi: 10.1109/TNNLS.2016.2582924. URL `https://doi.org/10.1109/TNNLS.2016.2582924`.

Jiuxiang Gu, Zhenhua Wang, Jason Kuen, Lianyang Ma, Amir Shahroudy, Bing Shuai, Ting Liu, Xingxing Wang, and Gang Wang. Recent advances in convolutional neural networks. *CoRR*, abs/1512.07108, 2015.

David Ha, Andrew M. Dai, and Quoc V. Le. Hypernetworks. *CoRR*, abs/1609.09106, 2016.

Song Han, Xingyu Liu, Huizi Mao, Jing Pu, Ardavan Pedram, Mark A. Horowitz, and William J. Dally. EIE: efficient inference engine on compressed deep neural network. *CoRR*, abs/1602.01528, 2016.

Kaiming He. URL `http://kaiminghe.com`. Facebook AI Research (FAIR).

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015.

Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross B. Girshick. Mask R-CNN. *CoRR*, abs/1703.06870, 2017.

Karl Moritz Hermann, Tomás Kociský, Edward Grefenstette, Lasse Espeholt, Will Kay, Mustafa Suleyman, and Phil Blunsom. Teaching machines to read and comprehend. *CoRR*, abs/1506.03340, 2015.

Geoffrey Hinton. URL `http://www.cs.toronto.edu/~hinton/`. University of Toronto (U of T), Ontario, Canada. Google Brain Team.

Geoffrey Hinton and Ruslan Salakhutdinov. Discovering binary codes for documents by learning deep generative models. *Topics in Cognitive Science*, 3(1): 74–91, 2011. ISSN 1756-8765. doi: 10.1111/j.1756-8765.2010.01109.x. URL `http://dx.doi.org/10.1111/j.1756-8765.2010.01109.x`.

Geoffrey Hinton, Li Deng, Dong Yu, George Dahl, Abdel rahman Mohamed, Navdeep Jaitly, Andrew Senior, Vincent Vanhoucke, Patrick Nguyen, Tara Sainath, and Brian Kingsbury. Deep neural networks for acoustic modeling in speech recognition. *Signal Processing Magazine*, 2012.

Geoffrey Hinton, Oriol Vinyals, and Jeffrey Dean. Distilling the knowledge in a neural network. In *NIPS Deep Learning and Representation Learning Workshop*, 2015. URL `http://arxiv.org/abs/1503.02531`.

Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Comput.*, 9 (8):1735–1780, November 1997. ISSN 0899-7667. doi: 10.1162/neco.1997.9.8.1735. URL `http://dx.doi.org/10.1162/neco.1997.9.8.1735`.

Daniel Holden, Jun Saito, and Taku Komura. A deep learning framework for character motion synthesis and editing. *ACM Trans. Graph.*, 35(4):138:1–138:11, July 2016. ISSN 0730-0301. doi: 10.1145/2897824.2925975. URL `http://doi.acm.org/10.1145/2897824.2925975`.

Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *CoRR*, abs/1704.04861, 2017.

Sandy H. Huang, Nicolas Papernot, Ian J. Goodfellow, Yan Duan, and Pieter Abbeel. Adversarial attacks on neural network policies. *CoRR*, abs/1702.02284, 2017.

Forrest N. Iandola, Matthew W. Moskewicz, Khalid Ashraf, Song Han, William J. Dally, and Kurt Keutzer. Squeezenet: Alexnet-level accuracy with 50x fewer parameters and <1mb model size. *CoRR*, abs/1602.07360, 2016.

Sergey Ioffe. Batch renormalization: Towards reducing minibatch dependence in batch-normalized models. *CoRR*, abs/1702.03275, 2017.

Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Francis Bach and David Blei, editors, *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 448–456, Lille, France, 07–09 Jul 2015. PMLR. URL `http://proceedings.mlr.press/v37/ioffe15.html`.

Shuiwang Ji, Wei Xu, Ming Yang, and Kai Yu. 3d convolutional neural networks for human action recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(1): 221–231, January 2013. ISSN 0162-8828. doi: 10.1109/TPAMI.2012.59. URL `http://dx.doi.org/10.1109/TPAMI.2012.59`.

Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross B. Girshick, Sergio Guadarrama, and Trevor Darrell. Caffe: Convolutional architecture for fast feature embedding. *CoRR*, abs/1408.5093, 2014.

Rafal Józefowicz, Oriol Vinyals, Mike Schuster, Noam Shazeer, and Yonghui Wu. Exploring the limits of language modeling. *CoRR*, abs/1602.02410, 2016.

Lukasz Kaiser and Ilya Sutskever. Neural gpus learn algorithms. *CoRR*, abs/1511.08228, 2015.

Nal Kalchbrenner, Edward Grefenstette, and Phil Blunsom. A convolutional neural network for modelling sentences. *CoRR*, abs/1404.2188, 2014.

Eunhee Kang, Junhong Min, and Jong Chul Ye. Wavenet: a deep convolutional neural network using directional wavelets for low-dose x-ray CT reconstruction. *CoRR*, abs/1610.09736, 2016.

Andrej Karpathy, George Toderici, Sanketh Shetty, Thomas Leung, Rahul Sukthankar, and Li Fei-Fei. Large-scale video classification with convolutional neural networks. In *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, CVPR '14, pages 1725–1732, Washington, DC, USA, 2014. IEEE Computer Society. ISBN 978-1-4799-5118-5. doi: 10.1109/CVPR.2014.223. URL `http://dx.doi.org/10.1109/CVPR.2014.223`.

Andrej Karpathy, Justin Johnson, and Fei-Fei Li. Visualizing and understanding recurrent networks. *CoRR*, abs/1506.02078, 2015.

Taeksoo Kim, Moonsu Cha, Hyunsoo Kim, Jung Kwon Lee, and Jiwon Kim. Learning to discover cross-domain relations with generative adversarial networks. *CoRR*, abs/1703.05192, 2017.

Yoon Kim. Convolutional neural networks for sentence classification. *CoRR*, abs/1408.5882, 2014.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, NIPS'12, pages 1097–1105, USA, 2012. Curran Associates Inc. URL `http://dl.acm.org/citation.cfm?id=2999134.2999257`.

David Krueger, Tegan Maharaj, János Kramár, Mohammad Pezeshki, Nicolas Ballas, Nan Rosemary Ke, Anirudh Goyal, Yoshua Bengio, Hugo Larochelle, Aaron C. Courville, and Chris Pal. Zoneout: Regularizing rnns by randomly preserving hidden activations. *CoRR*, abs/1606.01305, 2016.

Ankit Kumar, Ozan Irsoy, Jonathan Su, James Bradbury, Robert English, Brian Pierce, Peter Ondruska, Ishaan Gulrajani, and Richard Socher. Ask me anything: Dynamic memory networks for natural language processing. *CoRR*, abs/1506.07285, 2015.

Karol Kurach, Marcin Andrychowicz, and Ilya Sutskever. Neural random-access machines. *CoRR*, abs/1511.06392, 2015.

Guillaume Lample, Miguel Ballesteros, Sandeep Subramanian, Kazuya Kawakami, and Chris Dyer. Neural architectures for named entity recognition. *CoRR*, abs/1603.01360, 2016.

Guillaume Lample, Neil Zeghidour, Nicolas Usunier, Antoine Bordes, Ludovic Denoyer, and Marc'Aurelio Ranzato. Fader networks: Manipulating images by sliding attributes. *CoRR*, abs/1706.00409, 2017. URL `http://arxiv.org/abs/1706.00409`.

Gustav Larsson, Michael Maire, and Gregory Shakhnarovich. Fractalnet: Ultra-deep neural networks without residuals. *CoRR*, abs/1605.07648, 2016.

Quoc V. Le and Tomas Mikolov. Distributed representations of sentences and documents. *CoRR*, abs/1405.4053, 2014.

Yann LeCun. URL `http://yann.lecun.com`. New York University (NYU), NY, USA. Facebook AI Research (FAIR).

Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521:436 EP –, 05 2015. URL `http://dx.doi.org/10.1038/nature14539`.

Hyungtae Lee, Sungmin Eum, and Heesung Kwon. ME R-CNN: multi-expert region-based CNN for object detection. *CoRR*, abs/1704.01069, 2017.

Ian Lenz, Honglak Lee, and Ashutosh Saxena. Deep learning for detecting robotic grasps. *CoRR*, abs/1301.3592, 2013.

Wei Li, Rui Zhao, Tong Xiao, and Xiaogang Wang. Deepreid: Deep filter pairing neural network for person re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Columbus, USA, June 2014.

Yuxi Li. Deep reinforcement learning: An overview. *CoRR*, abs/1701.07274, 2017.

Min Lin, Qiang Chen, and Shuicheng Yan. Network in network. *CoRR*, abs/1312.4400, 2013.

Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott E. Reed, Cheng-Yang Fu, and Alexander C. Berg. SSD: single shot multibox detector. *CoRR*, abs/1512.02325, 2015.

Fujun Luan, Sylvain Paris, Eli Shechtman, and Kavita Bala. Deep photo style transfer. *CoRR*, abs/1703.07511, 2017.

Minh-Thang Luong, Hieu Pham, and Christopher D. Manning. Effective approaches to attention-based neural machine translation. *CoRR*, abs/1508.04025, 2015.

Lars Maaløe, Casper Kaae Sønderby, Søren Kaae Sønderby, and Ole Winther. Auxiliary deep generative models. In *Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48*, ICML'16, pages 1445–1454. JMLR.org, 2016. URL `http://dl.acm.org/citation.cfm?id=3045390.3045543`.

Xudong Mao, Qing Li, Haoran Xie, Raymond Y. K. Lau, and Zhen Wang. Multi-class generative adversarial networks with the L2 loss function. *CoRR*, abs/1611.04076, 2016.

Gary Marcus. Deep learning: A critical appraisal. *CoRR*, abs/1801.00631, 2018. URL `https://arxiv.org/abs/1801.00631`.

Jonathan Masci, Alessandro Giusti, Dan C. Ciresan, Gabriel Fricout, and Jürgen Schmidhuber. A fast learning algorithm for image segmentation with max-pooling convolutional networks. In *IEEE International Conference on Image Processing, ICIP 2013, Melbourne, Australia, September 15-18, 2013*, pages 2713–2717, 2013a. doi: 10.1109/ICIP.2013.6738559. URL `https://doi.org/10.1109/ICIP.2013.6738559`.

Jonathan Masci, Ueli Meier, Gabriel Fricout, and Jürgen Schmidhuber. Multi-scale pyramidal pooling network for generic steel defect classification. In *The 2013 International Joint Conference on Neural Networks, IJCNN 2013, Dallas, TX, USA, August 4-9, 2013*, pages 1–8, 2013b. doi: 10.1109/IJCNN.2013.6706920. URL `https://doi.org/10.1109/IJCNN.2013.6706920`.

Grégoire Mesnil, Yann Dauphin, Kaisheng Yao, Yoshua Bengio, Li Deng, Dilek Z. Hakkani-Tür, Xiaodong He, Larry P. Heck, Gökhan Tür, Dong Yu, and Geoffrey Zweig. Using recurrent neural networks for slot filling in spoken language understanding. *IEEE/ACM Trans. Audio, Speech & Language Processing*, 23(3):530–539, 2015.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *CoRR*, abs/1301.3781, 2013a.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. Distributed representations of words and phrases and their compositionality. *CoRR*, abs/1310.4546, 2013b.

Piotr Mirowski, Yann LeCun, Deepak Madhavan, and Ruben Kuzniecky. Comparing svm and convolutional networks for epileptic seizure prediction from intracranial eeg. In *Proc. Machine Learning and Signal Processing (MLSP'08)*. IEEE, 2008.

Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin A. Riedmiller. Playing atari with deep reinforcement learning. *CoRR*, abs/1312.5602, 2013.

Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin Riedmiller, Andreas K. Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharshan Kumaran, Daan Wierstra, Shane Legg, and Demis Hassabis. Human-level control through deep reinforcement learning. *Nature*, 518:529 EP –, 02 2015. URL `http://dx.doi.org/10.1038/nature14236`.

Volodymyr Mnih, Adrià Puigdomènech Badia, Mehdi Mirza, Alex Graves, Timothy P. Lillicrap, Tim Harley, David Silver, and Koray Kavukcuoglu. Asynchronous methods for deep reinforcement learning. *CoRR*, abs/1602.01783, 2016.

Joel Moniz and Christopher J. Pal. Convolutional residual memory networks. *CoRR*, abs/1606.05262, 2016.

Arvind Neelakantan, Quoc V. Le, and Ilya Sutskever. Neural programmer: Inducing latent programs with gradient descent. *CoRR*, abs/1511.04834, 2015.

Anh Mai Nguyen, Jason Yosinski, and Jeff Clune. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. *CoRR*, abs/1412.1897, 2014.

Michael A. Nielsen. *Neural Networks and Deep Learning*. Determination Press, 2015. `http://neuralnetworksanddeeplearning.com`.

Yusuke Niitani, Toru Ogawa, Shunta Saito, and Masaki Saito. Chainercv: a library for deep learning in computer vision. *CoRR*, abs/1708.08169, 2017.

Chris Olah and Shan Carter. Attention and augmented recurrent neural networks. *Distill*, 2016. doi: 10.23915/distill.00001. URL `http://distill.pub/2016/augmented-rnns`.

Maxime Oquab, Leon Bottou, Ivan Laptev, and Josef Sivic. Learning and transferring mid-level image representations using convolutional neural networks. In *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, CVPR '14, pages 1717–1724, Washington, DC, USA, 2014. IEEE Computer Society. ISBN 978-1-4799-5118-5. doi: 10.1109/CVPR.2014.222. URL `http://dx.doi.org/10.1109/CVPR.2014.222`.

Baolin Peng and Kaisheng Yao. Recurrent neural networks with external memory for language understanding. volume abs/1506.00195, 2015.

Ryan Poplin, Dan Newburger, Jojo Dijamco, Nam Nguyen, Dion Loy, Sam S. Gross, Cory Y. McLean, and Mark A. DePristo. Creating a universal snp and small indel variant caller with deep neural networks. *BioRxiv*, 2016. URL `https://doi.org/10.1101/092890`.

M. Ranzato, J. Susskind, V. Mnih, and G. Hinton. On deep generative models with applications to recognition. In *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition*, CVPR '11, pages 2857–2864, Washington, DC, USA, 2011. IEEE Computer Society. ISBN 978-1-4577-0394-2. doi: 10.1109/CVPR.2011.5995710. URL `http://dx.doi.org/10.1109/CVPR.2011.5995710`.

Ali Sharif Razavian, Hossein Azizpour, Josephine Sullivan, and Stefan Carlsson. CNN features off-the-shelf: an astounding baseline for recognition. *CoRR*, abs/1403.6382, 2014.

Joseph Redmon, Santosh Kumar Divvala, Ross B. Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. *CoRR*, abs/1506.02640, 2015.

Scott E. Reed and Nando de Freitas. Neural programmer-interpreters. *CoRR*, abs/1511.06279, 2015.

Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. In *Neural Information Processing Systems (NIPS)*, 2015.

Dario Rethage, Jordi Pons, and Xavier Serra. A wavenet for speech denoising. *CoRR*, abs/1706.07162, 2017.

Danilo Rezende, Shakir, Ivo Danihelka, Karol Gregor, and Daan Wierstra. One-shot generalization in deep generative models. In Maria Florina Balcan and Kilian Q. Weinberger, editors, *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 1521–1529, New York, New York, USA, 20–22 Jun 2016. PMLR. URL `http://proceedings.mlr.press/v48/rezende16.html`.

Sara Sabour, Nicholas Frosst, and Geoffrey E. Hinton. Dynamic routing between capsules. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA*, pages 3859–3869, 2017. URL `http://papers.nips.cc/paper/6975-dynamic-routing-between-capsules`.

Tim Salimans, Ian J. Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training gans. *CoRR*, abs/1606.03498, 2016.

Tom Schaul, Justin Bayer, Daan Wierstra, Yi Sun, Martin Felder, Frank Sehnke, Thomas Rückstieß, and Jürgen Schmidhuber. Pybrain. *J. Mach. Learn. Res.*, 11:743–746, March 2010. ISSN 1532-4435. URL `http://dl.acm.org/citation.cfm?id=1756006.1756030`.

J. Schmidhuber. Deep Learning. *Scholarpedia*, 10(11):32832, 2015. doi: 10.4249/scholarpedia.32832. URL `http://www.scholarpedia.org/article/Deep_Learning`. revision #152272.

Juergen Schmidhuber. URL `http://www.idsia.ch/~juergen`. IDSIA, USI. Dalle Molle Institute for Artificial Intelligence, Manno-Lugano, Switzerland.

Jürgen Schmidhuber. *Deep Learning in Neural Networks: An Overview*, volume abs/1404.7828. 2014. URL `http://arxiv.org/abs/1404.7828`.

Samira Shabanian, Devansh Arpit, Adam Trischler, and Yoshua Bengio. Variational bi-lstms. *CoRR*, 1711.05717, 2017. URL `https://arxiv.org/abs/1711.05717`.

Shikhar Sharma, Ryan Kiros, and Ruslan Salakhutdinov. Action recognition using visual attention. *CoRR*, abs/1511.04119, 2015.

Yangyang Shi, Kaisheng Yao, Hu Chen, Dong Yu, Yi-Cheng Pan, and Mei-Yuh Hwang. Recurrent support vector machines for slot tagging in spoken language understanding. In *HLT-NAACL*, pages 393–399. The Association for Computational Linguistics, 2016a.

Yangyang Shi, Kaisheng Yao, Le Tian, and Daxin Jiang. Deep LSTM based feature mapping for query classification. In *HLT-NAACL*, pages 1501–1511. The Association for Computational Linguistics, 2016b.

David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel, and Demis Hassabis. Mastering the game of go with deep neural networks and tree search. *Nature*, 529:484 EP –, 01 2016. URL `http://dx.doi.org/10.1038/nature16961`.

David Silver, Thomas Hubert, Julian Schrittwieser, Ioannis Antonoglou, Matthew Lai, Arthur Guez, Marc Lanctot, Laurent Sifre, Dharshan Kumaran, Thore Graepel, Timothy Lillicrap, Karen Simonyan, and Demis Hassabis. Mastering chess and shogi by self-play with a general reinforcement learning algorithm. *CoRR*, abs/1712.01815, 2017a. URL `https://arxiv.org/abs/1712.01815`.

David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, Yutian Chen, Timothy Lillicrap, Fan Hui, Laurent Sifre, George van den Driessche, Thore Graepel, and Demis Hassabis. Mastering the game of go without human knowledge. *Nature*, 550:354 EP –, 10 2017b. URL `http://dx.doi.org/10.1038/nature24270`.

Karen Simonyan and Andrew Zisserman. Two-stream convolutional networks for action recognition in videos. *CoRR*, abs/1406.2199, 2014a.

Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2014b.

Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15:1929–1958, 2014. URL `http://jmlr.org/papers/v15/srivastava14a.html`.

Rupesh Kumar Srivastava, Klaus Greff, and Jürgen Schmidhuber. Highway networks. *CoRR*, abs/1505.00387, 2015. URL `http://arxiv.org/abs/1505.00387`.

Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott E. Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. *CoRR*, abs/1409.4842, 2014.

Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, and Lior Wolf. Deepface: Closing the gap to human-level performance in face verification. In *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, CVPR '14, pages 1701–1708, Washington, DC, USA, 2014. IEEE Computer Society. ISBN 978-1-4799-5118-5. doi: 10.1109/CVPR.2014.220. URL `http://dx.doi.org/10.1109/CVPR.2014.220`.

Raphael Tang and Jimmy Lin. Honk: A pytorch reimplementation of convolutional neural networks for keyword spotting. *CoRR*, abs/1710.06554, 2017.

Yichuan Tang, Ruslan Salakhutdinov, and Geoffrey Hinton. Deep lambertian networks. 2, 06 2012.

Sasha Targ, Diogo Almeida, and Kevin Lyman. Resnet in resnet: Generalizing residual architectures. *CoRR*, abs/1603.08029, 2016.

Dmitry Ulyanov, Vadim Lebedev, Andrea Vedaldi, and Victor S. Lempitsky. Texture networks: Feed-forward synthesis of textures and stylized images. *CoRR*, abs/1603.03417, 2016.

Aäron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew W. Senior, and Koray Kavukcuoglu. Wavenet: A generative model for raw audio. *CoRR*, abs/1609.03499, 2016a.

Aäron van den Oord, Nal Kalchbrenner, and Koray Kavukcuoglu. Pixel recurrent neural networks. *CoRR*, abs/1601.06759, 2016b.

Hado van Hasselt, Arthur Guez, and David Silver. Deep reinforcement learning with double q-learning. *CoRR*, abs/1509.06461, 2015.

Bart van Merriënboer, Dzmitry Bahdanau, Vincent Dumoulin, Dmitriy Serdyuk, David Warde-Farley, Jan Chorowski, and Yoshua Bengio. Blocks and fuel: Frameworks for deep learning. *CoRR*, abs/1506.00619, 2015.

Oriol Vinyals, Alexander Toshev, Samy Bengio, and Dumitru Erhan. Show and tell: A neural image caption generator. *CoRR*, abs/1411.4555, 2014.

Oriol Vinyals, Meire Fortunato, and Navdeep Jaitly. Pointer networks. *CoRR*, 1506.03134, 2017. URL `https://arxiv.org/abs/1506.03134`.

Haohan Wang, Bhiksha Raj, and Eric P. Xing. On the origin of deep learning. *CoRR*, abs/1702.07800, 2017a.

Shenlong Wang. URL `http://www.cs.toronto.edu/~slwang/`. University of Toronto (U of T), Ontario, Canada.

Yuxuan Wang, R. J. Skerry-Ryan, Daisy Stanton, Yonghui Wu, Ron J. Weiss, Navdeep Jaitly, Zongheng Yang, Ying Xiao, Zhifeng Chen, Samy Bengio, Quoc V. Le, Yannis Agiomyrgiannakis, Rob Clark, and Rif A. Saurous. Tacotron: A fully end-to-end text-to-speech synthesis model. *CoRR*, abs/1703.10135, 2017b.

Jason Weston, Sumit Chopra, and Antoine Bordes. Memory networks. *CoRR*, abs/1410.3916, 2014.

Martin Wöllmer, Florian Eyben, Alex Graves, Björn Schuller, and Gerhard Rigoll. Bidirectional lstm networks for context-sensitive keyword detection in a cognitive virtual agent framework. *Cognitive Computation*, 2(3):180–190, Sep 2010. ISSN 1866-9964. doi: 10.1007/s12559-010-9041-8. URL `https://doi.org/10.1007/s12559-010-9041-8`.

Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V. Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, Jeff Klingner, Apurva Shah, Melvin Johnson, Xiaobing Liu, Lukasz Kaiser, Stephan Gouws, Yoshikiyo Kato, Taku Kudo, Hideto Kazawa, Keith Stevens, George Kurian, Nishant Patil, Wei Wang, Cliff Young, Jason Smith, Jason Riesa, Alex Rudnick, Oriol Vinyals, Greg Corrado, Macduff Hughes, and Jeffrey Dean. Google's neural machine translation system: Bridging the gap between human and machine translation. *CoRR*, abs/1609.08144, 2016.

Edgar Xi, Selina Bing, and Yang Jin. Capsule network performance on complex data. *CoRR*, abs/1712.03480v1, 2017. URL `https://arxiv.org/abs/1712.03480v1`.

Saining Xie, Ross B. Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. Aggregated residual transformations for deep neural networks. *CoRR*, abs/1611.05431, 2016.

Caiming Xiong, Stephen Merity, and Richard Socher. Dynamic memory networks for visual and textual question answering. *CoRR*, abs/1603.01417, 2016.

Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron C. Courville, Ruslan Salakhutdinov, Richard S. Zemel, and Yoshua Bengio. Show, attend and tell: Neural image caption generation with visual attention. *CoRR*, abs/1502.03044, 2015.

Zichao Yang, Xiaodong He, Jianfeng Gao, Li Deng, and Alexander J. Smola. Stacked attention networks for image question answering. *CoRR*, abs/1511.02274, 2015.

Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. How transferable are features in deep neural networks? In *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'14, pages 3320–3328, Cambridge, MA, USA, 2014. MIT Press. URL `http://dl.acm.org/citation.cfm?id=2969033.2969197`.

Tom Young, Devamanyu Hazarika, Soujanya Poria, and Erik Cambria. Recent trends in deep learning based natural language processing. *CoRR*, abs/1708.02709, 2017.

Matthew D. Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. *CoRR*, abs/1311.2901, 2013.

Chiyuan Zhang, Samy Bengio, Moritz Hardt, Benjamin Recht, and Oriol Vinyals. Understanding deep learning requires rethinking generalization. *CoRR*, abs/1611.03530, 2016a.

Richard Zhang, Phillip Isola, and Alexei A. Efros. Colorful image colorization. *CoRR*, abs/1603.08511, 2016b.

Shi-Xiong Zhang, Chaojun Liu, Kaisheng Yao, and Yifan Gong. Deep neural support vector machines for speech recognition. In *ICASSP*, pages 4275–4279. IEEE, 2015a.

Yu Zhang, Guoguo Chen, Dong Yu, Kaisheng Yao, Sanjeev Khudanpur, and James R. Glass. Highway long short-term memory rnns for distant speech recognition. *CoRR*, abs/1510.08983, 2015b.

Yu Zhang, Guoguo Chen, Dong Yu, Kaisheng Yao, Sanjeev Khudanpur, and James R. Glass. Highway long short-term memory RNNS for distant speech recognition. In *ICASSP*, pages 5755–5759. IEEE, 2016c.

Zixing Zhang, Jürgen T. Geiger, Jouni Pohjalainen, Amr El-Desoky Mousa, and Björn W. Schuller. Deep learning for environmentally robust speech recognition: An overview of recent developments. *CoRR*, abs/1705.10874, 2017.

Shuai Zheng, Sadeep Jayasumana, Bernardino Romera-Paredes, Vibhav Vineet, Zhizhong Su, Dalong Du, Chang Huang, and Philip H. S. Torr. Conditional random fields as recurrent neural networks. *CoRR*, abs/1502.03240, 2015.

Jun-Yan Zhu, Philipp Krähenbühl, Eli Shechtman, and Alexei A. Efros. Generative visual manipulation on the natural image manifold. *CoRR*, abs/1609.03552, 2016.

Xiao Xiang Zhu, Devis Tuia, Lichao Mou, Gui-Song Xia, Liangpei Zhang, Feng Xu, and Friedrich Fraundorfer. Deep learning in remote sensing: a review. *CoRR*, abs/1710.03959, 2017.

Julian Georg Zilly, Rupesh Kumar Srivastava, Jan Koutník, and Jürgen Schmidhuber. Recurrent highway networks. In *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017*, pages 4189–4198, 2017. URL `http://proceedings.mlr.press/v70/zilly17a.html`.