|  |  |  |
| --- | --- | --- |
| Jumping NLP Curves: A Review of Natural Language Processing Research | Natural language processing (NLP) is a theory-motivated range of computational techniques for the automatic analysis and representation of human language. NLP research has evolved from the era of punch cards and batch processing (in which the analysis of a sentence could take up to 7 minutes) to the era of Google and the likes of it (in which millions of webpages can be processed in less than a second). This review paper draws on recent developments in NLP research to look at the past, present, and future of NLP technology in a new light. Borrowing the paradigm of `jumping curves' from the field of business management and marketing prediction, this survey article reinterprets the evolution of NLP research as the intersection of three overlapping curves-namely Syntactics, Semantics, and Pragmatics Curves which will eventually lead NLP research to evolve into natural language understanding. | @article{cambria2014jumping,  title={Jumping NLP curves: A review of natural language processing research},  author={Cambria, Erik and White, Bebo},  journal={IEEE Computational intelligence magazine},  volume={9},  number={2},  pages={48--57},  year={2014},  publisher={IEEE}  } |
| NLP | In recent years, advances in machine learning have led to significant and widespread improvements in how we interact with our world. One of the most portentous of these advances is the field of deep learning. Based on artificial neural networks that resemble those in the human brain, deep learning is a set of methods that permits computers to learn from data without human supervision and intervention. Furthermore, these methods can adapt to changing environments and provide continuous improvement to learned abilities. Today, deep learning is prevalent in our everyday life in the form of Google’s search, Apple’s Siri, and Amazon’s and Netflix’s recommendation engines to name but a few examples. When we interact with our email systems, online chatbots, and voice or image recognition systems deployed at businesses ranging from healthcare to financial services, we see robust applications of deep learning in action. | @book{kamath2019deep,  title={Deep learning for NLP and speech recognition},  author={Kamath, Uday and Liu, John and Whitaker, James},  volume={84},  year={2019},  publisher={Springer}  } |
| Deep residual networks for pre-classification based Indian language identification | This paper proposes a pre-classification based language identification (LID) system for Indian languages. In this system, firstly, languages are pre-classified into tonal and non-tonal categories and then individual languages are identified from the languages of the respective category. In this work, language discriminating ability of various acoustic features like, pitch Chroma, mel-frequency Cepstral coefficients (MFCCs) and their combination has been investigated. The system performance has been analyzed for features extracted using different analysis units, like, syllables and utterances. The effectiveness of deep residual networks (ResNets) model in identification of Indian languages has been studied. Also, the system performance has been compared with the performances of other deep neural network architectures like, Convolutional Neural network (CNN) model, cascade CNN-long short-term memory (LSTM) model and shallow architecture like, ANN. Experiments have been carried out on NIT Silchar language database (NITS-LD) and OGI-Multilingual database (OGI-MLTS). Experimental analysis suggests that proposed ResNets model, based on syllable-level features, outperforms the other models. The pre-classification module provides accuracies of 96.6%, 93.2% and 90.6% for NITS-LD, and 92.1%, 89.3% and 85.4% for OGI-MLTS database, with 30s, 10s and 3s test data respectively. The pre-classification module helps to improve the system performance by 3.8%, 4.1% and 4.3% for 30s, 10s and 3s test data respectively. For OGI-MLTS database, the respective improvements are 6.8%, 6.5% and 5.4%. | @article{bhanja2019deep,  title={Deep residual networks for pre-classification based Indian language identification},  author={Bhanja, Chuya China and Bisharad, Dipjyoti and Laskar, Rabul Hussain},  journal={Journal of Intelligent \& Fuzzy Systems},  volume={36},  number={3},  pages={2207--2218},  year={2019},  publisher={IOS Press}  } |
| End-to-end scene text recognition | This paper focuses on the problem of word detection and recognition in natural images. The problem is significantly more challenging than reading text in scanned documents, and has only recently gained attention from the computer vision community. Sub-components of the problem, such as text detection and cropped image word recognition, have been studied in isolation [7, 4, 20]. However, what is unclear is how these recent approaches contribute to solving the end-to-end problem of word recognition. We fill this gap by constructing and evaluating two systems. The first, representing the de facto state-of-the-art, is a two stage pipeline consisting of text detection followed by a leading OCR engine. The second is a system rooted in generic object recognition, an extension of our previous work in [20]. We show that the latter approach achieves superior performance. While scene text recognition has generally been treated with highly domain-specific methods, our results demonstrate the suitability of applying generic computer vision methods. Adopting this approach opens the door for real world scene text recognition to benefit from the rapid advances that have been taking place in object recognition | @inproceedings{wang2011end,  title={End-to-end scene text recognition},  author={Wang, Kai and Babenko, Boris and Belongie, Serge},  booktitle={2011 International Conference on Computer Vision},  pages={1457--1464},  year={2011},  organization={IEEE}  } |
| Fast-RCNN | This paper proposes a Fast Region-based Convolutional Network method (Fast R-CNN) for object detection. Fast R-CNN builds on previous work to efficiently classify object proposals using deep convolutional networks. Compared to previous work, Fast R-CNN employs several innovations to improve training and testing speed while also increasing detection accuracy. Fast R-CNN trains the very deep VGG16 network 9x faster than R-CNN, is 213x faster at test-time, and achieves a higher mAP on PASCAL VOC 2012. Compared to SPPnet, Fast R-CNN trains VGG16 3x faster, tests 10x faster, and is more accurate. Fast R-CNN is implemented in Python and C++ (using Caffe) and is available under the open-source MIT License at https://github.com/rbgirshick/fast-rcnn. | @inproceedings{girshick2015fast,  title={Fast r-cnn},  author={Girshick, Ross},  booktitle={Proceedings of the IEEE international conference on computer vision},  pages={1440--1448},  year={2015}  } |
| Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning | Remarkable progress has been made in image recognition, primarily due to the availability of large-scale annotated datasets and deep convolutional neural networks (CNNs). CNNs enable learning data-driven, highly representative, hierarchical image features from sufficient training data. However, obtaining datasets as comprehensively annotated as ImageNet in the medical imaging domain remains a challenge. There are currently three major techniques that successfully employ CNNs to medical image classification: training the CNN from scratch, using off-the-shelf pre-trained CNN features, and conducting unsupervised CNN pre-training with supervised fine-tuning. Another effective method is transfer learning, i.e., fine-tuning CNN models pre-trained from natural image dataset to medical image tasks. In this paper, we exploit three important, but previously understudied factors of employing deep convolutional neural networks to computer-aided detection problems. We first explore and evaluate different CNN architectures. The studied models contain 5 thousand to 160 million parameters, and vary in numbers of layers. We then evaluate the influence of dataset scale and spatial image context on performance. Finally, we examine when and why transfer learning from pre-trained ImageNet (via fine-tuning) can be useful. We study two specific computer-aided detection (CADe) problems, namely thoraco-abdominal lymph node (LN) detection and interstitial lung disease (ILD) classification. We achieve the state-of-the-art performance on the mediastinal LN detection, and report the first five-fold cross-validation classification results on predicting axial CT slices with ILD categories. Our extensive empirical evaluation, CNN model analysis and valuable insights can be extended to the design of high performance CAD systems for other medical imaging tasks. | @article{shin2016deep,  title={Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning},  author={Shin, Hoo-Chang and Roth, Holger R and Gao, Mingchen and Lu, Le and Xu, Ziyue and Nogues, Isabella and Yao, Jianhua and Mollura, Daniel and Summers, Ronald M},  journal={IEEE transactions on medical imaging},  volume={35},  number={5},  pages={1285--1298},  year={2016},  publisher={IEEE}  } |
| EESEN: End-to-end speech recognition using deep RNN models and WFST-based decoding | The performance of automatic speech recognition (ASR) has improved tremendously due to the application of deep neural networks (DNNs). Despite this progress, building a new ASR system remains a challenging task, requiring various resources, multiple training stages and significant expertise. This paper presents our Eesen framework which drastically simplifies the existing pipeline to build state-of-the-art ASR systems. Acoustic modeling in Eesen involves learning a single recurrent neural network (RNN) predicting context-independent targets (phonemes or characters). To remove the need for pre-generated frame labels, we adopt the connectionist temporal classification (CTC) objective function to infer the alignments between speech and label sequences. A distinctive feature of Eesen is a generalized decoding approach based on weighted finite-state transducers (WFSTs), which enables the efficient incorporation of lexicons and language models into CTC decoding. Experiments show that compared with the standard hybrid DNN systems, Eesen achieves comparable word error rates (WERs), while at the same time speeding up decoding significantly | @inproceedings{miao2015eesen,  title={EESEN: End-to-end speech recognition using deep RNN models and WFST-based decoding},  author={Miao, Yajie and Gowayyed, Mohammad and Metze, Florian},  booktitle={2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)},  pages={167--174},  year={2015},  organization={IEEE}  } |
| LSTM | Neural networks have become increasingly popular for the task of language modeling. Whereas feed-forward networks only exploit a fixed context length to predict the next word of a sequence, conceptually, standard recurrent neural networks can take into account all of the predecessor words. On the other hand, it is well known that recurrent networks are difficult to train and therefore are unlikely to show the full potential of recurrent models. These problems are addressed by a the Long Short-Term Memory neural network architecture. In this work, we analyze this type of network on an English and a large French language modeling task. Experiments show improvements of about 8 % relative in perplexity over standard recurrent neural network LMs. In addition, we gain considerable improvements in WER on top of a state-of-the-art speech recognition system. | @inproceedings{sundermeyer2012lstm,  title={LSTM neural networks for language modeling},  author={Sundermeyer, Martin and Schl{\"u}ter, Ralf and Ney, Hermann},  booktitle={Thirteenth annual conference of the international speech communication association},  year={2012}  } |