Literature survey on deep learning and graph data

April 2017

Issues to be considered:

- We will review most recent papers that combine deep learning and graph embedding. However, we may not focus on graph embedding, as this area has been studied a lot and is a quite crowded area now.

1 Deep learning and graph data

1.1 Anomaly detection

1. **Lstm-based encoder-decoder for multi-sensor anomaly detection**[80]

   Malhotra, Pankaj and Ramakrishnan, Anusha and Anand, Gaurangi and Vig, Lovekesh and Agarwal, Puneet and Shroff, Gautam

   arXiv 2016

   Mechanical devices such as engines, vehicles, aircrafts, etc., are typically instrumented with numerous sensors to capture the behavior and health of the machine. However, there are often external factors or variables which are not captured by sensors leading to time-series which are inherently unpredictable. For instance, manual controls and/or unmonitored environmental conditions or load may lead to inherently unpredictable time-series. Detecting anomalies in such scenarios becomes challenging using standard approaches based on mathematical models that rely on stationarity, or prediction models that utilize prediction errors to detect anomalies. We propose a Long Short Term Memory Networks based Encoder-Decoder scheme for Anomaly Detection (EncDec-AD) that learns to reconstruct 'normal' time-series behavior, and thereafter uses reconstruction error to detect anomalies. We experiment with three publicly available quasi predictable time-series datasets: power demand, space shuttle, and ECG, and two real-world engine datasets with both predictive and unpredictable behavior. We show that EncDec-AD is robust and can detect anomalies from predictable, unpredictable, periodic, aperiodic, and quasi-periodic time-series. Further, we show that EncDec-AD is able to detect anomalies from short time-series (length as small as 30) as well as long time-series (length as large as 500).

2. **MS-LSTM: A multi-scale LSTM model for BGP anomaly detection**[27]

   Cheng, Min and Xu, Qian and Lv, Jianming and Liu, Wenyin and Li, Qing and Wang, Jianping

   International Conference on Network Protocols (ICNP) 2016

   Detecting anomalous Border Gateway Protocol (BGP) traffic is significantly important in improving both security and robustness of the Internet. Existing solutions apply classic classifiers to make real-time decision based on the traffic features of present moment. However, due to the frequently happening burst and noise in dynamic Internet traffic, the decision based on short-term features is not reliable. To address this problem, we propose MS-LSTM, a multi-scale Long Short-Term Memory (LSTM) model to consider the Internet flow as a multi-dimensional time sequence and learn the traffic pattern from historical features in a sliding time window. In addition, we find that adopting different time scale to preprocess the traffic flow has great impact on the performance of all classifiers. In this paper, comprehensive experiments are conducted and the results show that a proper time scale can improve about 10

3. **Anomaly detection in ECG time signals via deep long short-term memory networks**[20]

   Chauhan, Sucheta and Vig, Lovekesh

   Data Science and Advanced Analytics (DSAA), 2015

   Electrocardiography (ECG) signals are widely used to gauge the health of the human heart, and the resulting time series signal is often analyzed manually by a medical professional to detect any arrhythmia that the patient may have suffered. Much work has been done to automate the process of analyzing ECG signals, but most of the research involves extensive preprocessing of the ECG data to derive vectorized features and subsequently designing a classifier to discriminate between healthy ECG signals and those indicative of an Arrhythmia. This approach requires knowledge and data of the different types of Arrhythmia for training. However, the heart is a complex organ and there are many different and new types of Arrhythmia that can occur which were not part of the original training set. Thus, it may be more prudent to adopt an anomaly detection approach towards analyzing ECG signals. In this paper, we utilize a deep recurrent neural network architecture with Long Short Term Memory (LSTM) units to develop a predictive model for healthy ECG signals. We further utilize the probability distribution of the prediction errors from these recurrent models to indicate normal or abnormal behavior. An added advantage of using LSTM networks is that the ECG signal can be directly fed into the network without any elaborate preprocessing as required by other techniques. Also, no prior information about abnormal signals is needed by the networks as they were trained only on normal data. We have used the MIT-BIH Arrhythmia Database to obtain ECG time series data for both normal periods and for periods during four different types of Arrhythmias, namely Premature Ventricular Contraction (PVC), Atrial Premature Contraction (APC), Paced Beats (PB) and Ventricular Couplet (VC). Results are promising and indicate that Deep LSTM models may be viable for detecting anomalies in ECG signals.

4. **Long short term memory networks for anomaly detection in time series**[81]

   Malhotra, Pankaj and Vig, Lovekesh and Shroff, Gautam and Agarwal, Puneet

   European Symposium on Artificial Neural Networks 2015

   Long Short Term Memory (LSTM) networks have been demonstrated to be particularly useful for learning sequences containing longer term patterns of unknown length, due to their ability to maintain long term memory. Stacking recurrent hidden layers in such networks also enables the learning of higher level temporal features, for faster learning with sparser representations. In this paper, we use stacked LSTM net- works for anomaly/fault detection in time series. A network is trained on non-anomalous data and used as a predictor over a number of time steps. The resulting prediction errors are modeled as a multivariate Gaussian distribution, which is used to assess the likelihood of anomalous behav- ior. The efficacy of this approach is demonstrated on four datasets: ECG, space shuttle, power demand, and multi-sensor engine dataset.

5. **Recurrent Neural Networks for anomaly detection in the Post-Mortem time series of LHC superconducting magnets** [135]

   Wielgosz, Maciej and Skoczeń, Andrzej and Mertik, Matej

   arXiv 2017

   This paper presents a model based on Deep Learning algorithms of LSTM and GRU for facilitating an anomaly detection in Large Hadron Collider superconducting magnets. We used high resolution data available in Post Mortem database to train a set of models and chose the best possible set of their hyper-parameters. Using Deep Learning approach allowed to examine a vast body of data and extract the fragments which require further experts examination and are regarded as anomalies. The presented method does not require tedious manual threshold setting and operator attention at the stage of the system setup. Instead, the automatic approach is proposed, which achieves according to our experiments accuracy of 99%. This is reached for the largest dataset of 302 MB and the following architecture of the network: single layer LSTM, 128 cells, 20 epochs of training, look$_b ack = 16, look_a head = 128, grid = 100 and optimizer Adam. All the experiments were erun on GPU Nvidia Tesla K80$

6. **Anomaly detection in aircraft data using Recurrent Neural Networks (RNN)** [88]

Nanduri, Anvardh and Sherry, Lance

Integrated Communications Navigation and Surveillance (ICNS), 2016

Anomaly Detection in multivariate, time-series data collected from aircraft's Flight Data Recorder (FDR) or Flight Operational Quality Assurance (FOQA) data provide a powerful means for identifying events and trends that reduce safety margins. The industry standard "Exceedance Detection" algorithm uses a list of specified parameters and their thresholds to identify known deviations. In contrast, Machine Learning algorithms detect unknown unusual patterns in the data either through semi-supervised or unsupervised learning. The Multiple Kernel Anomaly Detection (MKAD) algorithm based on One-class SVM identified 6 of 11 canonical anomalies in a large dataset but is limited by the need for dimensionality reduction, poor sensitivity to short term anomalies, and inability to detect anomalies in latent features. This paper describes the application of Recurrent Neural Networks (RNN) with Long Term Short Term Memory (LTSM) and Gated Recurrent Units (GRU) architectures which can overcome the limitations described above. The RNN algorithms detected 9 out the 11 anomalies in the test dataset with Precision = 1, Recall = 0.818 and F1 score = 0.89. RNN architectures, designed for time-series data, are suited for implementation on the flight deck to provide real-time anomaly detection. The implications of these results are discussed.

7. **Unsupervised Anomaly Detection in Sequences Using Long Short Term Memory Recurrent Neural Networks**[9] Blaisten-Barojas, Estela GMU PhD Thesis Paper

   Long Short Term Memory (LSTM) recurrent neural networks (RNNs) are evaluated for their potential to generically detect anomalies in sequences. First, anomaly detection techniques are surveyed at a high level so that their shortcomings are exposed. The shortcomings are mainly their inflexibility in the use of a context 'window' size and/or their suboptimal performance in handling sequences. Furthermore, high-performing techniques for sequences are usually associated with their respective knowledge domains. After discussing these shortcomings, RNNs are exposed mathematically as generic sequence modelers that can handle sequences of arbitrary length. From there, results from experiments using RNNs show their ability to detect anomalies in a set of test sequences. The test sequences had different types of anomalies and unique normal behavior. Given the characteristics of the test data, it was concluded that the RNNs were not only able to generically distinguish rare values in the data (out of context) but were also able to generically distinguish abnormal patterns (in context). In addition to the anomaly detection work, a solution for reproducing computational research is described. The solution addresses reproducing compute applications based on Docker container technology as well as automating the infrastructure that runs the applications. By design, the solution allows the researcher to seamlessly transition from local (test) application execution to remote (production) execution because little distinction is made between local and remote execution. Such flexibility and automation allows the researcher to be more confident of results and more productive, especially when dealing with multiple machines.

8. **Anomaly Detection in Automobile Control Network Data with Long Short-Term Memory Networks**[119]

   Taylor, Adrian and Leblanc, Sylvain and Japkowicz, Nathalie

   Data Science and Advanced Analytics (DSAA), 2016

   Modern automobiles have been proven vulnerable to hacking by security researchers. By exploiting vulnerabilities in the car's external interfaces, such as wifi, bluetooth, and physical connections, they can access a car's controller area network (CAN) bus. On the CAN bus, commands can be sent to control the car, for example cutting the brakes or stopping the engine. While securing the car's interfaces to the outside world is an important part of mitigating this threat, the last line of defence is detecting malicious behaviour on the CAN bus. We propose an anomaly detector based on a Long Short-Term Memory neural network to detect CAN bus attacks. The detector works by learning to predict the next data word originating from each sender on the bus. Highly surprising bits in the actual next word are flagged as anomalies. We evaluate the detector by synthesizing anomalies with modified CAN bus data. The synthesized anomalies are designed to mimic attacks reported in the literature. We show that the detector can detect anomalies we synthesized with low false alarm rates. Additionally, the granularity of the bit predictions can provide forensic investigators clues as to the nature of flagged anomal

9. **Multilevel Anomaly Detection for Mixed Data**[34]

   Do, Kien and Tran, Truyen and Venkatesh, Svetha

   arXiv 2016

   Anomalies are those deviating from the norm. Unsupervised anomaly detection often translates to identifying low density regions. Major problems arise when data is high-dimensional and mixed of discrete and continuous attributes. We propose MIXMAD, which stands for MIXed data Multilevel Anomaly Detection, an ensemble method that estimates the sparse regions across multiple levels of abstraction of mixed data. The hypothesis is for domains where multiple data abstractions exist, a data point may be anomalous with respect to the raw representation or more abstract representations. To this end, our method sequentially constructs an ensemble of Deep Belief Nets (DBNs) with varying depths. Each DBN is an energy-based detector at a predefined abstraction level. At the bottom level of each DBN, there is a Mixed-variate Restricted Boltzmann Machine that models the density of mixed data. Predictions across the ensemble are finally combined via rank aggregation. The proposed MIXMAD is evaluated on high-dimensional realworld datasets of different characteristics. The results demonstrate that for anomaly detection, (a) multilevel abstraction of high-dimensional and mixed data is a sensible strategy, and (b) empirically, MIXMAD is superior to popular unsupervised detection methods for both homogeneous and mixed data.

10. **Deep model based domain adaptation for fault diagnosis**[76]

    Lu, Weining and Liang, Bin and Cheng, Yu and Meng, Deshan and Yang, Jun and Zhang, Tao

    IEEE Transactions on Industrial Electronics 2017

    Abstract: In recent years, machine learning techniques have been widely used to solve many problems for fault diagnosis. However, in many real-world fault diagnosis applications, the distribution of the source domain data (on which the model is trained) is different from the distribution of the target domain data (where the learned model is actually deployed), which leads to performance degradation. In this paper, we introduce domain adaptation, which can find the solution to this problem by adapting the classifier or the regression model trained in a source domain for use in a different but related target domain. In particular, we proposed a novel deep neural network model with domain adaptation for fault diagnosis. Two main contributions are concluded by comparing to the previous works: first, the proposed model can utilize domain adaptation meanwhile strengthening the representative information of the original data, so that a high classification accuracy in the target domain can be achieved, and second, we proposed several strategies to explore the optimal hyperparameters of the model. Experimental results, on several real-world datasets, demonstrate the effectiveness and the reliability of both the proposed model and the exploring strategies for the parameters.

11. **Deep structured energy based models for anomaly detection**[145]

    Zhai, Shuangfei and Cheng, Yu and Lu, Weining and Zhang, Zhongfei

    ICML 2016

    In this paper, we attack the anomaly detection problem by directly modeling the data distri- bution with deep architectures. We propose deep structured energy based models (DSEBMs), where the energy function is the output of a deterministic deep neural network with structure. We develop novel model architectures to inte- grate EBMs with different types of data such as static data, sequential data, and spatial data, and apply appropriate model architectures to adapt to the data structure. Our training algorithm is built upon the recent development of score match- ing (Hyva rinen, 2005), which connects an EBM with a regularized autoencoder, eliminating the need for complicated sampling method. Statistically sound decision criterion can be derived for anomaly detection purpose from the perspective of the energy landscape of the data distribution. We investigate two decision criteria for performing anomaly detection: the energy score and the reconstruction error. Extensive empirical studies on benchmark tasks demonstrate that our proposed model consistently matches or outperforms all the competing methods.

12. **Transferred Deep Learning for Anomaly Detection in Hyperspectral Imagery**[69]

    Li, Wei and Wu, Guodong and Du, Qian

    IEEE Geoscience and Remote Sensing Letters 2017

In this letter, a novel anomaly detection framework with transferred deep convolutional neural network (CNN) is proposed. The framework is designed by considering the following facts: 1) a reference data with labeled samples are utilized, because no prior information is available about the image scene for anomaly detection and 2) pixel pairs are generated to enlarge the sample size, since the advantage of CNN can be realized only if the number of training samples is sufficient. A multilayer CNN is trained by using difference between pixel pairs generated from the reference image scene. Then, for each pixel in the image for anomaly detection, difference between pixel pairs, constructed by combining the center pixel and its surrounding pixels, is classified by the trained CNN with the result of similarity measurement. The detection output is simply generated by averaging these similarity scores. Experimental performance demonstrates that the proposed algorithm outperforms the classic Reed-Xiaoli and the state-of-the-art representation-based detectors, such as sparse representation-based detector (SRD) and collaborative representation-based detector.

13. **Research on Healthy Anomaly Detection Model Based on Deep Learning from Multiple Time-Series Physiological Signals**[133]

    Wang, Kai and Zhao, Youjin and Xiong, Qingyu and Fan, Min and Sun, Guotan and Ma, Longkun and Liu, Tong

    Scientific Programming 2016

    Health is vital to every human being. To further improve its already respectable medical technology, the medical community is transitioning towards a proactive approach which anticipates and mitigates risks before getting ill. This approach requires measuring the physiological signals of human and analyzes these data at regular intervals. In this paper, we present a novel approach to apply deep learning in physiological signals analysis that allows doctor to identify latent risks. However, extracting high level information from physiological time-series data is a hard problem faced by the machine learning communities. Therefore, in this approach, we apply model based on convolutional neural network that can automatically learn features from raw physiological signals in an unsupervised manner and then based on the learned features use multivariate Gauss distribution anomaly detection method to detect anomaly data. Our experiment is shown to have a significant performance in physiological signals anomaly detection. So it is a promising tool for doctor to identify early signs of illness even if the criteria are unknown a priori.

14. **Spatial–temporal convolutional neural networks for anomaly detection and localization in crowded scenes** [151]

    Shifu Zhou, Wei Shen, Dan Zeng, Mei Fang, Yuanwang Wei, Zhijiang Zhang

    Journal of Signal Processing: Image Communication 2016

    Abnormal behavior detection in crowded scenes is extremely challenging in the field of computer vision due to severe inter-object occlusions, varying crowd densities and the complex mechanics of a human crowd. We propose a method for detecting and locating anomalous activities in video sequences of crowded scenes. The key novelty of our method is the coupling of anomaly detection with a spatial– temporal Convolutional Neural Networks (CNN), which to the best of our knowledge has not been previously done. This architecture allows us to capture features from both spatial and temporal dimensions by performing spatial–temporal convolutions, thereby, both the appearance and motion information encoded in continuous frames are extracted. The spatial–temporal convolutions are only performed within spatial–temporal volumes of moving pixels to ensure robustness to local noise, and increase detection accuracy.

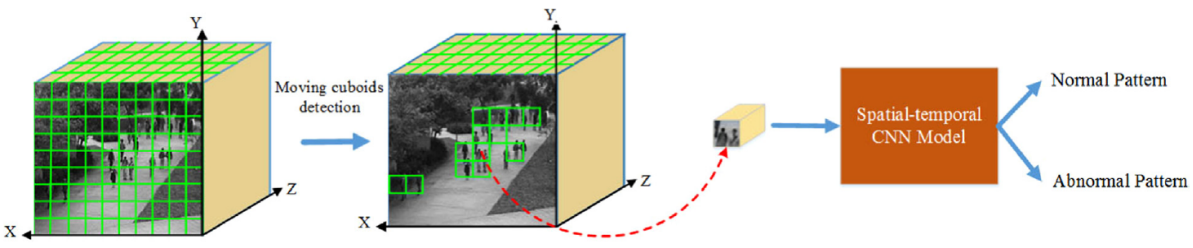

Figure 1: Overview of the proposed approach for anomaly detection.

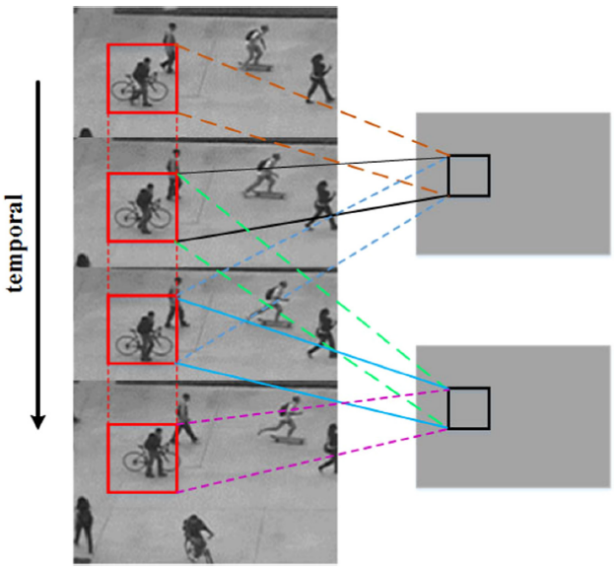

Figure 2: Illustration of the spatial–temporal convolution across both spatial and temporal domains. In this example, there are two different 3D kernels, whose temporal dimension is both3. That is to say, each feature map is obtained by performing spatial–temporal convolutions across 3 adjacent frames.

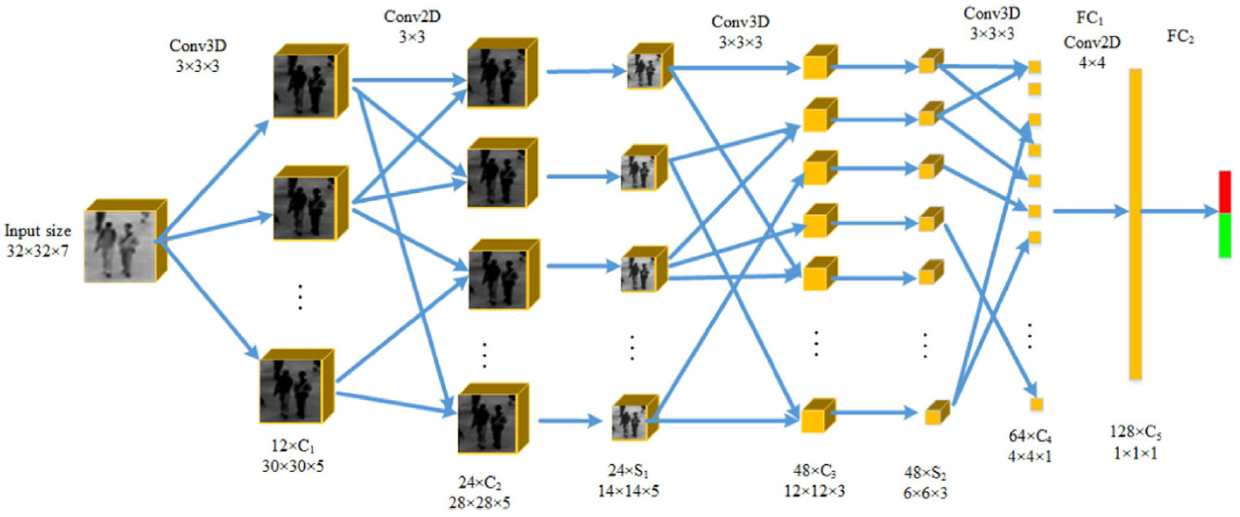

Figure 3: The spatial–temporal CNN structure used in this paper. It contains 4 convolutional layers, 2 subsampling layers and 2 fully connected layers.

15. **Deep-Anomaly: Fully Convolutional Neural Network for Fast Anomaly Detection in Crowded Scenes** [108]

Mohammad Sabokrou, Mohsen Fayyaz, Mahmood Fathy, Reinhard Klette

This paper presents an efficient method for detecting and localizing anomalies in videos showing crowded scenes. Research on fully convolutional neural networks (FCNs) has shown the potentials of this technology for object detection and localization, especially in images. We investigate how to involve temporal data, and how to transform a supervised FCN into an unsupervised one such that the resulting FCN ensures anomaly detection. Altogether, we propose an FCN-based architecture for anomaly detection and localization in crowded scenes videos. For reducing computations and, consequently, improving performance both with respect to speed and accuracy, we investigate the use of cascaded out-layer detection. Our architecture includes two main components, one for feature representation, and one for cascaded out-layer detection.

Experimental results on Subway and UCSD benchmarks confirm that the detection and localization accuracy of our method is comparable to stateof-the-art methods, but at a significantly increased speed of 370 fps.

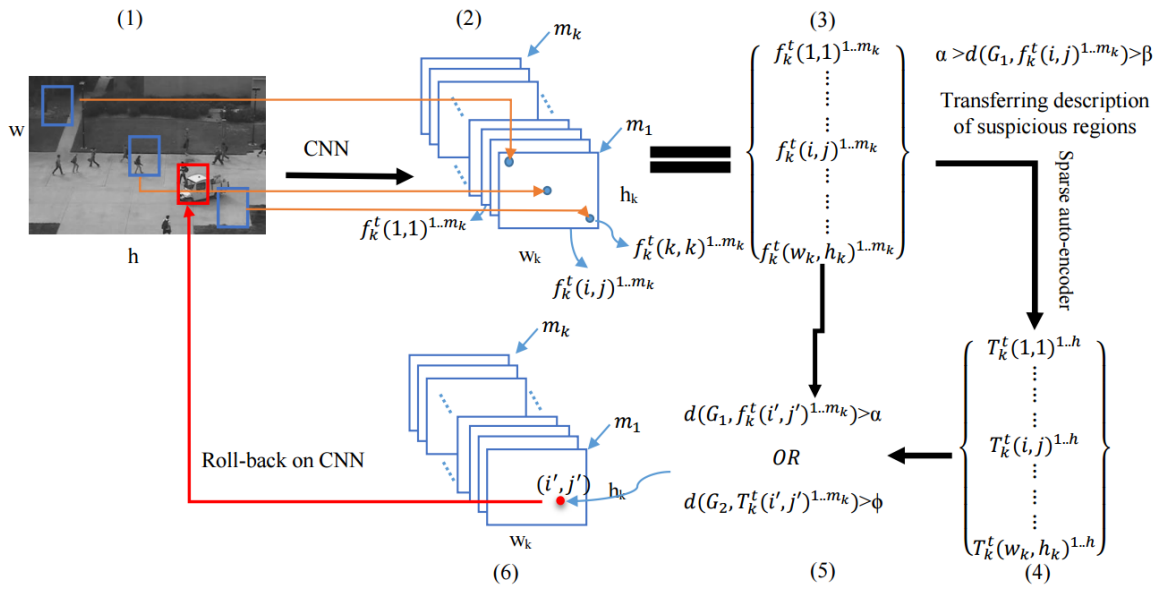

Figure 4: Scheme of proposed method. (1) Input video frame of size $w_0 \times h_0$, (2,3) Description of regions of size $h_k \times w_k$ generated by the k^{th} layer of the FCN. (4) Transformed feature domain using a sparse auto-encoder (for enhancing the features). (5) Joint anomaly detector. (6) Positions of those descriptions which identify anomalies.

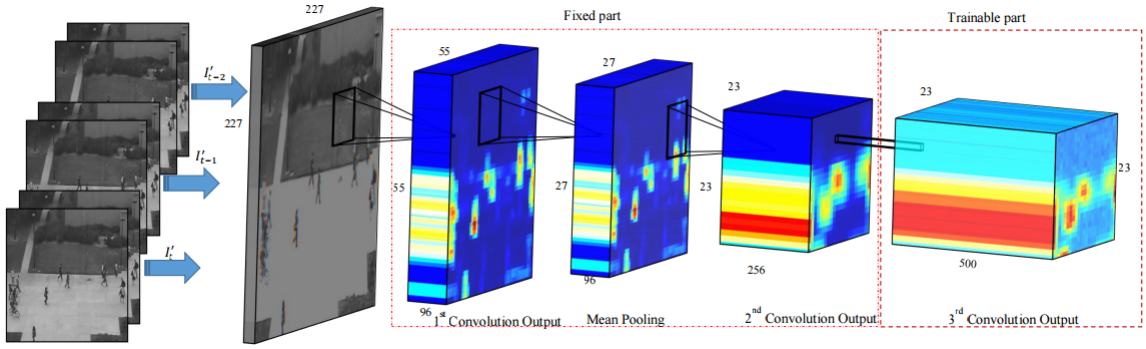

Figure 5: Proposed FCN structure for detecting anomalies. This FCNN is only used for regional feature extraction. Later on, two Gaussian classifiers are embedded for labeling anomaly regions, as discussed before.

16. **Deep-Cascade: Cascading 3D Deep Neural Networks for Fast Anomaly Detection and Localization in Crowded Scenes** [109]

Mohammad Sabokrou, Mohsen Fayyaz, Mahmood Fathy, and Reinhard Klette

IEEE Transactions on Image Processing 2017

This paper proposes a fast and reliable method for anomaly detection and localization in video data showing crowded scenes. Time-efficient anomaly localization is an ongoing challenge and subject of this paper. We propose a cubic-patchbased method, characterised by a cascade of classifiers, which makes use of an advanced feature-learning approach. Our cascade of classifiers has two main stages. First, a light but deep 3D auto-encoder is used for early identification of "many" normal cubic patches. This deep network operates on small cubic patches as being the first stage, before carefully resizing the remaining candidates of interest, and evaluating those at the second stage using a more complex and deeper 3D convolutional neural network (CNN). We divide the deep autoencoder and the CNN into multiple sub-stages, which operate as cascaded classifiers. Shallow layers of the cascaded deep networks (designed as Gaussian classifiers, acting as weak

single-class classifiers) detect "simple" normal patches, such as background patches and more complex normal patches, are detected at deeper layers. It is shown that the proposed novel technique (a cascade of two cascaded classifiers) performs comparable to current top-performing detection and localization methods on standard benchmarks, but outperforms those in general with respect to required computation time.

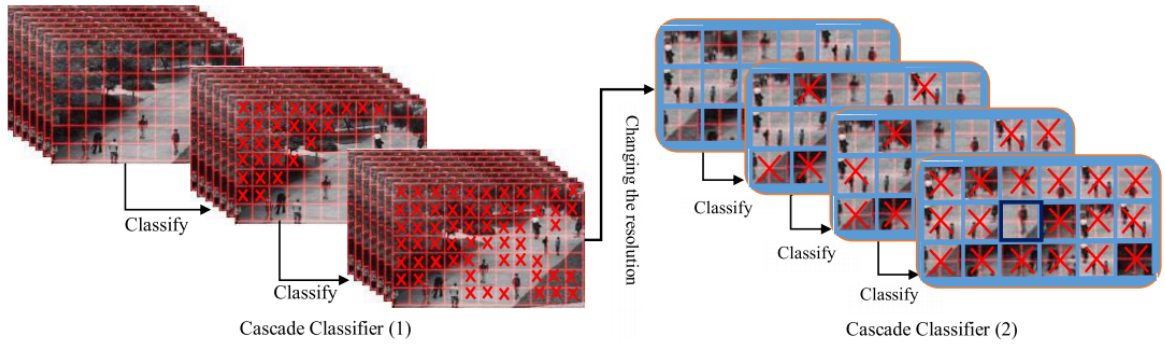

Figure 6: Our cascade for anomaly detection. Top:Processing of small patches for early rejection of normal patches. Bottom:Resized remaining patches and further rejection of normal patches.

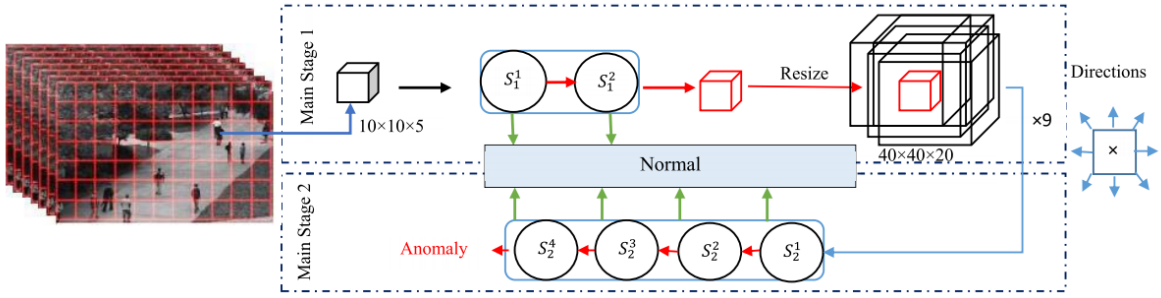

Figure 7: Work-flow of the proposed system.Top: Stage 1 has a cascaded stack-auto-encoder with two (S_1^1 and S_1^2)layers as sub-stages.Bottom: Stage 2 has a cascaded CNN with four sub-stages (S_2^1, S_2^2, S_2^3 and S_2^4)

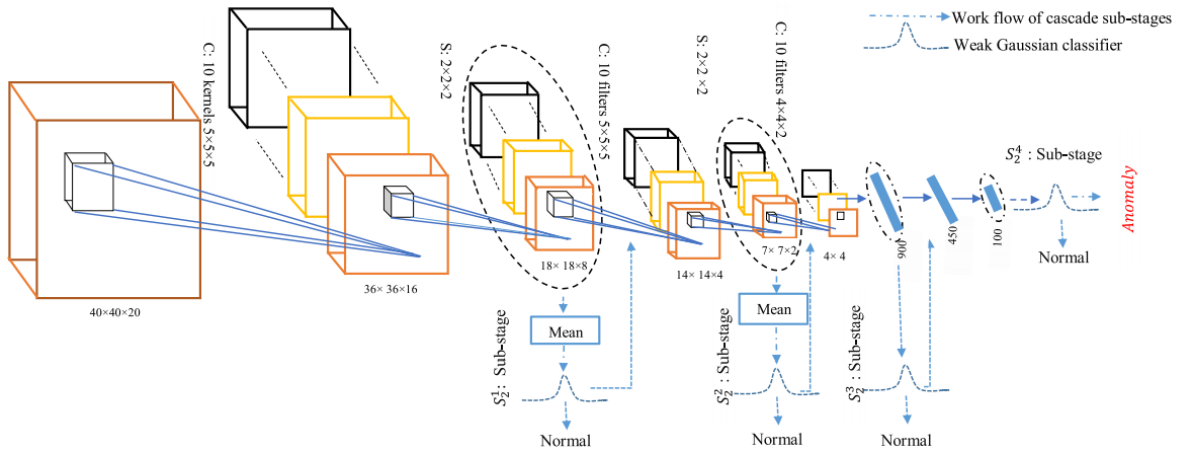

Figure 8: This sketch of our CNN architecture illustrates subsequent processing of one cubic kernel in the first convolution layer; C and S stands for "convolution" and "sub-sampling" (non-padded convolution is used). The mean operation is defined by channel-wise averaging.

17. **Deep Representation for Abnormal Event Detection in Crowded Scenes** [38]

MM '16, Proceedings of the 2016 ACM on Multimedia Conference

Abnormal event detection is extremely important, especially for video surveillance. Nowadays, many detectors have been proposed based on hand-crafted features. However, it remains challenging to effectively

distinguish abnormal events from normal ones. This paper proposes a deep representation based algorithm which extracts features in an unsupervised fashion. Specially, appearance, texture, and short-term motion features are automatically learned and fused with stacked denoising autoencoders. Subsequently, long-term temporal clues are modeled with a long shortterm memory (LSTM) recurrent network, in order to discover meaningful regularities of video events. The abnormal events are identified as samples which disobey these regularities. Moreover, this paper proposes a spatial anomaly detection strategy via manifold ranking, aiming at excluding false alarms. Experiments and comparisons on real world datasets show that the proposed algorithm outperforms state of the arts for the abnormal event detection problem in crowded scenes.

1.2 Graph embedding techniques

This subsection reviews deep learning techniques that are specifically designed for spatial or spatio-temporal graphps

1. **Graph-structured representations for visual question answering** [120]

   Damien Teney, Lingqiao Liu, Anton van den Hengel

   CVPR 2016

   This paper proposes to improve visual question answering (VQA) with structured representations of both scene contents and questions. A key challenge in VQA is to require joint reasoning over the visual and text domains. The predominant CNN/LSTM-based approach to VQA is limited by monolithic vector representations that largely ignore structure in the scene and in the form of the question. CNN feature vectors cannot effectively capture situations as simple as multiple object instances, and LSTMs process questions as series of words, which does not reflect the true complexity of language structure. We instead propose to build graphs over the scene objects and over the question words, and we describe a deep neural network that exploits the structure in these representations. This shows significant benefit over the sequential processing of LSTMs. The overall efficacy of our approach is demonstrated by significant improvements over the state-of-the-art, from 71.2% to 74.4% in accuracy on the "abstract scenes" multiple-choice benchmark, and from 34.7% to 39.1% in accuracy over pairs of "balanced" scenes, i.e. images with fine-grained differences and opposite yes/no answers to a same question.

2. **Structural-RNN: Deep Learning on Spatio-Temporal Graphs** [54]

   Ashesh Jain, Amir R. Zamir, Silvio Savarese, and Ashutosh Saxena

   CVPR 2016

   Deep Recurrent Neural Network architectures, though remarkably capable at modeling sequences, lack an intuitive high-level spatio-temporal structure. That is while many problems in computer vision inherently have an underlying high-level structure and can benefit from it. Spatio-temporal graphs are a popular tool for imposing such high-level intuitions in the formulation of real world problems. In this paper, we propose an approach for combining the power of high-level spatio-temporal graphs and sequence learning success of Recurrent Neural Networks (RNNs). We develop a scalable method for casting an arbitrary spatio- temporal graph as a rich RNN mixture that is feedforward, fully differentiable, and jointly trainable. The proposed method is generic and principled as it can be used for transforming any spatio-temporal graph through employing a certain set of well defined steps.

   In high-level steps, given an arbitrary st-graph, we first roll it out in time and decompose it into a set of contributing factor components. The factors identify the independent components that collectively determine one decision and are derived from both edges and nodes of the st-graph. We then semantically group the factor components and represent each group using one RNN, which results in the desired RNN mixture. The main challenges of this transformation problem are: 1) making the RNN mixture as rich as possible to enable learning complex functions, yet 2) keeping the RNN mixture scalable with respect to size of the input st-graph. In order to make the resulting RNN mixture rich, we liberally represent each spatio-temporal factor (including node factors, temporal edge factors, and spatiotemporal edge factors) using one RNN. On the other hand, to keep the overall mixture scalable but not lose the essential learning capacity, we utilize "factor sharing" (aka clique templates [54, 42, 53]) and allow the factors with similar semantic functions to share an RNN. This results in a rich and

scalable feedforward mixture of RNNs that is equivalent to the provided st-graph in terms of input, output, and spatiotemporal relationships. The mixture is also fully differentiable, and therefore, can be trained jointly as one entity.

3. **Large-Scale Embedding Learning in Heterogeneous Event Data** [45]

   Huan Gui, Jialu Liu, Fangbo Tao, Meng Jiang, Brandon Norick and Jiawei Han

   ICDM 2016

   Heterogeneous events, which are defined as events connecting strongly-typed objects, are ubiquitous in the real world. We propose a HyperEdge-Based Embedding (HEBE) framework for heterogeneous event data, where a hyperedge represents the interaction among a set of involving objects in an event. The HEBE framework models the proximity among objects in an event by predicting a target object given the other participating objects in the event (hyperedge). Since each hyperedge encapsulates more information on a given event, HEBE is robust to data sparseness. In addition, HEBE is scalable when the data size spirals. Extensive experiments on large-scale real-world datasets demonstrate the efficacy and robustness of HEBE.

4. **Heterogeneous Network Embedding via Deep Architectures** [19]

   Shiyu Chang, Wei Han, Jiliang Tang, Guo-Jun Qi, Charu C. Aggarwal, Thomas S. Huang

   KDD2015

   Data embedding is used in many machine learning applications to create low-dimensional feature representations, which preserves the structure of data points in their original space. In this paper, we examine the scenario of a heterogeneous network with nodes and content of various types. Such networks are notoriously difficult to mine because of the bewildering combination of heterogeneous contents and structures. The creation of a multidimensional embedding of such data opens the door to the use of a wide variety of off-the-shelf mining techniques for multidimensional data. Despite the importance of this problem, limited efforts have been made on embedding a network of scalable, dynamic and heterogeneous data. In such cases, both the content and linkage structure provide important cues for creating a unified feature representation of the underlying network. In this paper, we design a deep embedding algorithm for networked data. A highly nonlinear multilayered embedding function is used to capture the complex interactions between the heterogeneous data in a network. Our goal is to create a multi-resolution deep embedding function, that reflects both the local and global network structures, and makes the resulting embedding useful for a variety of data mining tasks. In particular, we demonstrate that the rich content and linkage information in a heterogeneous network can be captured by such an approach, so that similarities among cross-modal data can be measured directly in a common embedding space. Once this goal has been achieved, a wide variety of data mining problems can be solved by applying off-the-shelf algorithms designed for handling vector representations. Our experiments on real-world network datasets show the effectiveness and scalability of the proposed algorithm as compared to the state-of-the-art embedding methods.

5. **DeepWalk: Online Learning of Social Representations** [98]

   Bryan Perozzi, Rami Al-Rfou, Steven Skiena

   KDD 2014

   The authors present DeepWalk, a novel approach for learning latent representations of vertices in a network. These latent representations encode social relations in a continuous vector space, which is easily exploited by statistical models. DeepWalk generalizes recent advancements in language modeling and unsupervised feature learning (or deep learning) from sequences of words to graphs. DeepWalk uses local information obtained from truncated random walks to learn latent representations by treating walks as the equivalent of sentences.

6. **Line: Large-scale information network embedding** [118]

   Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, Qiaozhu Mei

   WWW 2015

This paper studies the problem of embedding very large information networks into low-dimensional vector spaces, which is useful in many tasks such as visualization, node classification, and link prediction. Most existing graph embedding methods do not scale for real world information networks which usually contain millions of nodes. In this paper, we propose a novel network embedding method called the "LINE," which is suitable for arbitrary types of information networks: undirected, directed, and/or weighted. The method optimizes a carefully designed objective function that preserves both the local and global network structures. An edge-sampling algorithm is proposed that addresses the limitation of the classical stochastic gradient descent and improves both the effectiveness and the efficiency of the inference. Empirical experiments prove the effectiveness of the LINE on a variety of real-world information networks, including language networks, social networks, and citation networks. The algorithm is very efficient, which is able to learn the embedding of a network with millions of vertices and billions of edges in a few hours on a typical single machine. The source code of the LINE is available online.

7. **Traversing Knowledge Graphs in Vector Space** [46]

   Kelvin Guu, John Miller, Percy Liang

   2015 Conference on Empirical Methods on Natural Language Processing (EMNLP)

   Path queries on a knowledge graph can be used to answer compositional questions such as "What languages are spoken by people living in Lisbon?". However, knowledge graphs often have missing facts (edges) which disrupts path queries. Recent models for knowledge base completion impute missing facts by embedding knowledge graphs in vector spaces. We show that these models can be recursively applied to answer path queries, but that they suffer from cascading errors. This motivates a new "compositional" training objective, which dramatically improves all models' ability to answer path queries, in some cases more than doubling accuracy. On a standard knowledge base completion task, we also demonstrate that compositional training acts as a novel form of structural regularization, reliably improving performance across all base models (reducing errors by up to 43%) and achieving new state-of-the-art results.

8. **node2vec: Scalable Feature Learning for Networks** [44]

   Aditya Grover, Jure Leskovec

   KDD 2016

   Prediction tasks over nodes and edges in networks require careful effort in engineering features used by learning algorithms. Recent research in the broader field of representation learning has led to significant progress in automating prediction by learning the features themselves. However, present feature learning approaches are not expressive enough to capture the diversity of connectivity patterns observed in networks. Here we propose node2vec, an algorithmic framework for learning continuous feature representations for nodes in networks. In node2vec, we learn a mapping of nodes to a low-dimensional space of features that maximizes the likelihood of preserving network neighborhoods of nodes. We define a flexible notion of a node's network neighborhood and design a biased random walk procedure, which efficiently explores diverse neighborhoods. Our algorithm generalizes prior work which is based on rigid notions of network neighborhoods, and we argue that the added flexibility in exploring neighborhoods is the key to learning richer representations. We demonstrate the efficacy of node2vec over existing state-ofthe-art techniques on multi-label classification and link prediction in several real-world networks from diverse domains. Taken together, our work represents a new way for efficiently learning stateof-the-art task-independent representations in complex networks.

9. **subgraph2vec: Learning Distributed Representations of Rooted Sub-graphs from Large Graphs** [89]

   Annamalai Narayanan, Mahinthan Chandramohan, Lihui Chen, Yang Liu, Santhoshkumar Saminathan

   arXiv 2016

   This paper presents subgraph2vec, a novel approach for learning latent representations of rooted subgraphs from large graphs inspired by recent advancements in Deep Learning and Graph Kernels. These latent representations encode semantic substructure dependencies in a continuous vector space, which

is easily exploited by statistical models for tasks such as graph classification, clustering, link prediction and community detection. subgraph2vec leverages on local information obtained from neighborhoods of nodes to learn their latent representations in an unsupervised fashion. We demonstrate that subgraph vectors learnt by our approach could be used in conjunction with classifiers such as CNNs, SVMs and relational data clustering algorithms to achieve significantly superior accuracies. Also, we show that the subgraph vectors could be used for building a deep learning variant of Weisfeiler-Lehman graph kernel. Our experiments on several benchmark and large-scale real-world datasets reveal that subgraph2vec achieves significant improvements in accuracies over existing graph kernels on both supervised and unsupervised learning tasks. Specifically, on two real world program analysis tasks, namely, code clone and malware detection, subgraph2vec outperforms state-of-the-art kernels by more than 17% and 4%, respectively.

10. **Distributed Representation of Subgraphs** [1]

    Bijaya Adhikari, Yao Zhang, Naren Ramakrishnan, B. Aditya Prakash

    arXiv 2017

    Network embeddings have become very popular in learning effective feature representations of networks. Motivated by the recent successes of embeddings in natural language processing, researchers have tried to find network embeddings in order to exploit machine learning algorithms for mining tasks like node classification and edge prediction. However, most of the work focuses on finding distributed representations of nodes, which are inherently ill-suited to tasks such as community detection which are intuitively dependent on subgraphs. Here, we propose sub2vec, an unsupervised scalable algorithm to learn feature representations of arbitrary subgraphs. We provide means to characterize similarties between subgraphs and provide theoretical analysis of sub2vec and demonstrate that it preserves the so-called local proximity. We also highlight the usability of sub2vec by leveraging it for network mining tasks, like community detection. We show that sub2vec gets significant gains over state-of-the-art methods and node-embedding methods. In particular, sub2vec offers an approach to generate a richer vocabulary of features of subgraphs to support representation and reasoning.

11. **Learning of multimodal representations with random walks on the click graph** [136]

    Wu, Fei and Lu, Xinyan and Song, Jun and Yan, Shuicheng and Zhang, Zhongfei Mark and Rui, Yong and Zhuang, Yueting

    IEEE Transactions on Image Processing 2016

    In multimedia information retrieval, most classic approaches tend to represent different modalities of media in the same feature space. With the click data collected from the users' searching behavior, existing approaches take either one-to-one paired data (text–image pairs) or ranking examples (text–query–image and/or image–query–text ranking lists) as training examples, which do not make full use of the click data, particularly the implicit connections among the data objects. In this paper, we treat the click data as a large click graph, in which vertices are images/text queries and edges indicate the clicks between an image and a query. We consider learning a multimodal representation from the perspective of encoding the explicit/implicit relevance relationship between the vertices in the click graph. By minimizing both the truncated random walk loss as well as the distance between the learned representation of vertices and their corresponding deep neural network output, the proposed model which is named multimodal random walk neural network (MRW-NN) can be applied to not only learn robust representation of the existing multimodal data in the click graph, but also deal with the unseen queries and images to support crossmodal retrieval. We evaluate the latent representation learned by MRW-NN on a public large-scale click log data set Clickture and further show that MRW-NN achieves much better crossmodal retrieval performance on the unseen queries/images than the other state-of-the-art methods.

12. **Context-Dependent Knowledge Graph Embedding** [77]

    Yuanfei Luo, Quan Wang, Bin Wang, Li Guo

    EMNLP 2015

Authors consider the problem of embedding knowledge graphs (KGs) into continuous vector spaces. Existing methods can only deal with explicit relationships within each triple, i.e., local connectivity patterns, but cannot handle implicit relationships across different triples, i.e., contextual connectivity patterns. This paper proposes context-dependent KG embedding, a twostage scheme that takes into account both types of connectivity patterns and obtains more accurate embeddings. We evaluate our approach on the tasks of link prediction and triple classification, and achieve significant and consistent improvements over state-of-the-art methods.

13. **Structural Deep Network Embedding** [132]

    Daixin Wang, Peng Cui, Wenwu Zhu

    KDD 2016

    Network embedding is an important method to learn low-dimensional representations of vertexes in networks, aiming to capture and preserve the network structure. Almost all the existing network embedding methods adopt shallow models. However, since the underlying network structure is complex, shallow models cannot capture the highly non-linear network structure, resulting in sub-optimal network representations. Therefore, **how to find a method that is able to effectively capture the highly non-linear network structure and preserve the global and local structure is an open yet important problem**. To solve this problem, in this paper we propose a Structural Deep Network Embedding method, namely SDNE. More specifically, we first propose a semi-supervised deep model, which has multiple layers of non-linear functions, thereby being able to capture the highly non-linear network structure. Then we propose to exploit the first-order and second-order proximity jointly to preserve the network structure. The second-order proximity is used by the unsupervised component to capture the global network structure. While the first-order proximity is used as the supervised information in the supervised component to preserve the local network structure. By jointly optimizing them in the semi-supervised deep model, our method can preserve both the local and global network structure and is robust to sparse networks.

14. **Deep Coevolutionary Network: Embedding User and Item Features for Recommendation** [31]

    Hanjun Dai, Yichen Wang, Rakshit Trivedi, Le Song

    KDD 2017

    Recommender systems often use latent features to explain the behaviors of users and capture the properties of items. As users interact with different items over time, user and item features can influence each other, evolve and co-evolve over time. The compatibility of user and item's feature further influence the future interaction between users and items.

    Recently, point process based models have been proposed in the literature aiming to capture the temporally evolving nature of these latent features. However, these models often make strong parametric assumptions about the evolution process of the user and item latent features, which may not reflect the reality, and has limited power in expressing the complex and nonlinear dynamics underlying these processes.

    To address these limitations, we propose a novel deep coevolutionary network model (DeepCoevolve), for learning user and item features based on their interaction graph. DeepCoevolve use recurrent neural network (RNN) over evolving networks to define the intensity function in point processes, which allows the model to capture complex mutual influence between users and items, and the feature evolution over time. We also develop an efficient procedure for training the model parameters, and show that the learned models lead to significant improvements in recommendation and activity prediction compared to previous state-of-the-arts parametric models.

15. **SEANO: Semi-supervised Embedding in Attributed Networks with Outliers** [72]

    iongqian Liang, Peter Jacobs, Srinivasan Parthasarathy

    Network embedding has attracted an increasing amount of attention in recent years due to its wide-ranging applications in graph mining tasks such as vertex classification, community detection, and

network visualization. While embedding homogeneous networks has been widely studied, few methods have examined the embedding of partially labeled attributed networks (PLAN) that arise in a semi-supervised setting. In this paper, we propose a novel framework, called Semi-supervised Embedding in Attributed Networks with Outliers (SEANO), to learn a robust low-dimensional vector representation that captures the topological proximity, attribute affinity and label similarity of vertices in a PLAN while accounting for outliers. We design a tree-shaped deep neural network with both a supervised and an unsupervised component. These components share the first several layers of the network. We alternate training between the two components to iteratively push information regarding network structure, attributes, and labels into the embedding. Experimental results on various datasets demonstrate the advantages of SEANO over state-of-the-art methods in semisupervised classification under both transductive and inductive settings. This paper also shows as a byproduct that SEANO can significantly outperform other methods when applied to the task of outlier detection. Finally, we present the use of SEANO in a challenging real-world setting - flood mapping of satellite images. Qualitatively, this paper finds that SEANO is able to outperform state-of-the-art remote sensing algorithms on this task.

16. **Scalable out-of-sample extension of graph embeddings using deep neural networks** [56]

    Aren Jansen, Greg Sell, Vince Lyzinski

    Several popular graph embedding techniques for representation learning and dimensionality reduction rely on performing computationally expensive eigendecompositions to derive a nonlinear transformation of the input data space. The resulting eigenvectors encode the embedding coordinates for the training samples only, and so the embedding of novel data samples requires further costly computation. In this paper, we present a method for the out-of-sample extension of graph embeddings using deep neural networks (DNN) to parametrically approximate these nonlinear maps. Compared with traditional nonparametric out-of-sample extension methods, we demonstrate that the DNNs can generalize with equal or better fidelity and require orders of magnitude less computation at test time. Moreover, we find that unsupervised pretraining of the DNNs improves optimization for larger network sizes, thus removing sensitivity to model selection.

17. **Revisiting semi-supervised learning with graph embeddings** [142]

    Zhilin Yang, William W. Cohen, Ruslan Salakhutdinov

    ICML 2016

    This paper presents a semi-supervised learning framework based on graph embeddings. Given a graph between instances, we train an embedding for each instance to jointly predict the class label and the neighborhood context in the graph. We develop both transductive and inductive variants of our method. In the transductive variant of our method, the class labels are determined by both the learned embeddings and input feature vectors, while in the inductive variant, the embeddings are defined as a parametric function of the feature vectors, so predictions can be made on instances not seen during training. On a large and diverse set of benchmark tasks, including text classification, distantly supervised entity extraction, and entity classification, we show improved performance over many of the existing models.

18. **Learning Deep Representations for Graph Clustering** [121]

    Fei Tian, Bin Gao, Qing Cui, Enhong Chen, Tie-Yan Liu

    AAAI 2014

    Recently deep learning has been successfully adopted in many applications such as speech recognition and image classification. In this work, we explore the possibility of employing deep learning in graph clustering. We propose a simple method, which first learns a nonlinear embedding of the original graph by stacked autoencoder, and then runs k-means algorithm on the embedding to obtain clustering result. We show that this simple method has solid theoretical foundation, due to the similarity between autoencoder and spectral clustering in terms of what they actually optimize. Then, we demonstrate that the proposed method is more efficient and flexible than spectral clustering. First, the computational complexity of autoencoder is much lower than spectral clustering: the former can be linear to the number of nodes in a sparse graph while the latter is super quadratic due to eigenvalue decomposition.

Second, when additional sparsity constraint is imposed, we can simply employ the sparse autoencoder developed in the literature of deep learning; however, it is nonstraightforward to implement a sparse spectral method. The experimental results on various graph datasets show that the proposed method significantly outperforms conventional spectral clustering, which clearly indicates the effectiveness of deep learning in graph clustering.

19. **Asymmetric Transitivity Preserving Graph Embedding** [92]

    Mingdong Ou, Peng Cui, Jian Pei, Wenwu Zhu

    KDD 2016

    Graph embedding algorithms embed a graph into a vector space where the structure and the inherent properties of the graph are preserved. The existing graph embedding methods cannot preserve the asymmetric transitivity well, which is a critical property of directed graphs. Asymmetric transitivity depicts the correlation among directed edges, that is, if there is a directed path from u to v, then there is likely a directed edge from u to v. Asymmetric transitivity can help in capturing structures of graphs and recovering from partially observed graphs. To tackle this challenge, we propose the idea of preserving asymmetric transitivity by approximating high-order proximity which are based on asymmetric transitivity. In particular, we develop a novel graph embedding algorithm, High-Order Proximity preserved Embedding (HOPE for short), which is scalable to preserve high-order proximities of large scale graphs and capable of capturing the asymmetric transitivity. More specifically, we first derive a general formulation that cover multiple popular high-order proximity measurements, then propose a scalable embedding algorithm to approximate the high-order proximity measurements based on their general formulation. Moreover, we provide a theoretical upper bound on the RMSE (Root Mean Squared Error) of the approximation. Our empirical experiments on a synthetic dataset and three real-world datasets demonstrate that HOPE can approximate the high-order proximities significantly better than the state-of-art algorithms and outperform the state-of-art algorithms in tasks of reconstruction, link prediction and vertex recommendation.

20. **GraRep: Learning Graph Representations with Global Structural Information** [140]

    Cheng Yang, Maosong Sun, Wayne Xin Zhao, Zhiyuan Liu, Edward Y.Chang

    arXiv 2016

    The accelerated growth of mobile trajectories in location-based services brings valuable data resources to understand users' moving behaviors. Apart from recording the trajectory data, another major characteristic of these location-based services is that they also allow the users to connect whomever they like. A combination of social networking and location-based services is called as location-based social networks (LBSN). As shown in previous works, locations that are frequently visited by socially-related persons tend to be correlated, which indicates the close association between social connections and trajectory behaviors of users in LBSNs. In order to better analyze and mine LBSN data, we present a novel neural network model which can joint model both social networks and mobile trajectories. In specific, our model consists of two components: the construction of social networks and the generation of mobile trajectories. We first adopt a network embedding method for the construction of social networks: a networking representation can be derived for a user. The key of our model lies in the component of generating mobile trajectories. We have considered four factors that influence the generation process of mobile trajectories, namely user visit preference, influence of friends, short-term sequential contexts and long-term sequential contexts. To characterize the last two contexts, we employ the RNN and GRU models to capture the sequential relatedness in mobile trajectories at different levels, i.e., short term or long term. Finally, the two components are tied by sharing the user network representations. Experimental results on two important applications demonstrate the effectiveness of our model. Especially, the improvement over baselines is more significant when either network structure or trajectory data is sparse.

21. **Walklets: Multiscale Graph Embeddings for Interpretable Network Classification** [99]

    Bryan Perozzi, Vivek Kulkarni, Steven Skiena

    arXiv 2016

This paper presents Walklets, a novel approach for learning multiscale representations of vertices in a network. These representations clearly encode multiscale vertex relationships in a continuous vector space suitable for multi-label classification problems. Unlike previous work, the latent features generated using Walklets are analytically derivable, and human interpretable. Walklets uses the offsets between vertices observed in a random walk to learn a series of latent representations, each which captures successively larger relationships. This variety of dependency information allows the same representation strategy to model phenomenon which occur at different scales. We demonstrate Walklets' latent representations on several multi-label network classification tasks for social networks such as BlogCatalog, Flickr, and YouTube. Our results show that Walklets outperforms new methods based on neural matrix factorization, and can scale to graphs with millions of vertices and edges.

22. **GraRep: Learning Graph Representations with Global Structural Information** [16]

    Shaosheng Ca, Wei Lu, Qiongkai Xu

    CIKM 2015

    This paper presents GraRep, a novel model for learning vertex representations of weighted graphs. This model learns low dimensional vectors to represent vertices appearing in a graph and, unlike existing work, integrates global structural information of the graph into the learning process. Authors also formally analyze the connections between their work and several previous research efforts, including the DeepWalk model of Perozzi et al. [20] as well as the skip-gram model with negative sampling of Mikolov et al. [18] Authors conduct experiments on a language network, a social network as well as a citation network and show that our learned global representations can be effectively used as features in tasks such as clustering, classification and visualization. Empirical results demonstrate that their representation significantly outperforms other state-of-the-art methods in such tasks.

1.3 Modeling directly on graphs

1.3.1 Spatial or spatio-temporal graphs

1.3.2 Other graphs

1. **Deep Graph Kernels** [139]

   Pinar Yanardag, S.V.N. Vishwanathan

   KDD 2015

   This paper presents Deep Graph Kernels, a unified framework to learn latent representations of sub-structures for graphs, inspired by latest advancements in language modeling and deep learning. Our framework leverages the dependency information between sub-structures by learning their latent representations. We demonstrate instances of our framework on three popular graph kernels, namely Graphlet kernels, Weisfeiler-Lehman subtree kernels, and Shortest-Path graph kernels. Our experiments on several benchmark datasets show that Deep Graph Kernels achieve significant improvements in classification accuracy over state-of-the-art graph kernels.

2. **Matching Node Embeddings for Graph Similarity** [91]

   Giannis Nikolentzos, Polykarpos Meladianos, Michalis Vazirgiannis

   AAAI 2017

   Graph kernels have emerged as a powerful tool for graph comparison. Most existing graph kernels focus on local properties of graphs and ignore global structure. In this paper, we compare graphs based on their global properties as these are captured by the eigenvectors of their adjacency matrices. We present two algorithms for both labeled and unlabeled graph comparison. These algorithms represent each graph as a set of vectors corresponding to the embeddings of its vertices. The similarity between two graphs is then determined using the Earth Mover's Distance metric. These similarities do not yield a positive semidefinite matrix. To address for this, we employ an algorithm for SVM classification using indefinite kernels. We also present a graph kernel based on the Pyramid Match kernel that finds an approximate correspondence between the sets of vectors of the two graphs. We further improve the

proposed kernel using the Weisfeiler-Lehman framework. We evaluate the proposed methods on several benchmark datasets for graph classification and compare their performance to state-of-the-art graph kernels. In most cases, the proposed algorithms outperform the competing methods, while their time complexity remains very attractive.

3. **Matching Node Embeddings for Graph Similarity** [105]

   Petar Ristos, Heiko Paulheim

   International Semantic Web Conference 2016

   Linked Open Data has been recognized as a valuable source for background information in data mining. However, most data mining tools require features in propositional form, i.e., a vector of nominal or numerical features associated with an instance, while Linked Open Data sources are graphs by nature. In this paper, we present RDF2Vec, an approach that uses language modeling approaches for unsupervised feature extraction from sequences of words, and adapts them to RDF graphs. We generate sequences by leveraging local information from graph sub-structures, harvested by Weisfeiler-Lehman Subtree RDF Graph Kernels and graph walks, and learn latent numerical representations of entities in RDF graphs. Our evaluation shows that such vector representations outperform existing techniques for the propositionalization of RDF graphs on a variety of different predictive machine learning tasks, and that feature vector representations of general knowledge graphs such as DBpedia and Wikidata can be easily reused for different tasks.

1.4 Other graphs

1. **Learning Convolutional Neural Networks for Graphs** [90]

   Mathias Niepert, Mohamed Ahmed, Konstantin Kutzkov

   ICML 2016

   Numerous important problems can be framed as learning from graph data. We propose a framework for learning convolutional neural networks for arbitrary graphs. These graphs may be undirected, directed, and with both discrete and continuous node and edge attributes. Analogous to image-based convolutional networks that operate on locally connected regions of the input, we present a general approach to extracting locally connected regions from graphs.

2. **Transfer Learning for Deep Learning on Graph-Structured Data** [66]

   Jaekoo Lee, Hyunjae Kim, Jongsun Lee, Sungroh Yoon

   Technical report 2016

   Graphs provide a powerful means for representing complex interactions between entities. Recently, new deep learning approaches have emerged for representing and modeling graphstructured data while the conventional deep learning methods, such as convolutional neural networks and recurrent neural networks, have mainly focused on the grid-structured inputs of image and audio. Leveraged by representation learning capabilities, deep learning-based techniques can detect structural characteristics of graphs, giving promising results for graph applications. In this paper, we attempt to advance deep learning for graph-structured data by incorporating another component: transfer learning. By transferring the intrinsic geometric information learned in the source domain, our approach can construct a model for a new but related task in the target domain without collecting new data and without training a new model from scratch.

3. **Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering** [32]

   Michaël Defferrard, Xavier Bresson, Vandergheynst Pierre

   NIPS 2016

   This work is interested in generalizing convolutional neural networks (CNNs) from low-dimensional regular grids, where image, video and speech are represented, to high-dimensional irregular domains, such as social networks, brain connectomes or words' embedding, represented by graphs. We present a formulation of CNNs in the context of spectral graph theory, which provides the necessary mathematical

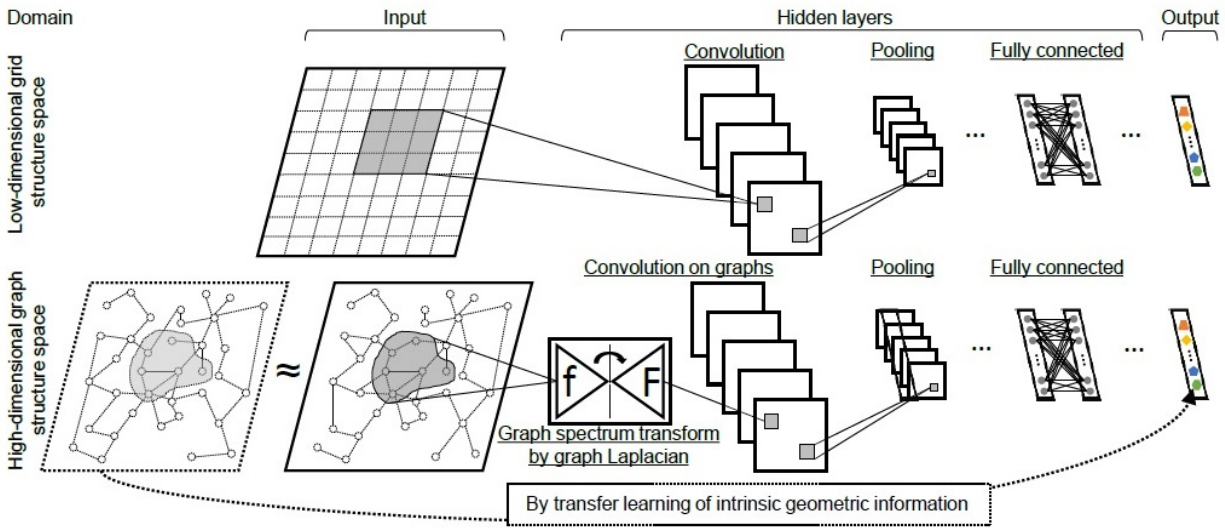

Figure 9: Conventional CNN works on a regular grid domain (top); proposed transfer learning framework for CNN, which can transfer intrinsic geometric information obtained from a source graph domain to a target graph domain (bottom) [66].

background and efficient numerical schemes to design fast localized convolutional filters on graphs. Importantly, the proposed technique offers the same linear computational complexity and constant learning complexity as classical CNNs, while being universal to any graph structure. Experiments on MNIST and 20NEWS demonstrate the ability of this novel deep learning system to learn local, stationary, and compositional features on graphs.

4. **Convolutional Networks on Graphs for Learning Molecular Fingerprints** [35]

   David Duvenaudy, Dougal Maclauriny, Jorge Aguilera-Iparraguirre, Rafael Gomez-Bombarelli, Timothy Hirzel, Alá an Aspuru-Guzik, Ryan P. Adams

   NIPS 2015

   This paper introduces a convolutional neural network that operates directly on graphs. These networks allow end-to-end learning of prediction pipelines whose inputs are graphs of arbitrary size and shape. The architecture we present generalizes standard molecular feature extraction methods based on circular fingerprints. They show that these data-driven features are more interpretable, and have better predictive performance on a variety of tasks

5. **Molecular graph convolutions: moving beyond fingerprints** [62]

   Steven Kearnes, authorKevin McCloskey, Marc Berndl, Vijay Pande, Patrick Riley

   Molecular "fingerprints" encoding structural information are the workhorse of cheminformatics and machine learning in drug discovery applications. However, fingerprint representations necessarily emphasize particular aspects of the molecular structure while ignoring others, rather than allowing the model to make data-driven decisions. We describe molecular graph convolutions, a machine learning architecture for learning from undirected graphs, specifically small molecules. Graph convolutions use a simple encoding of the molecular graph—atoms, bonds, distances, etc.—which allows the model to take greater advantage of information in the graph structure. Although graph convolutions do not outperform all fingerprint-based methods, they (along with other graph-based methods) represent a new paradigm in ligand-based virtual screening with exciting opportunities for future improvement.

6. **An Effective Neural Network Model for Graph-based Dependency Parsing** [97]

   Wenzhe Pei, Tao Ge, Baobao Chang

ACL 2015

Most existing graph-based parsing models rely on millions of hand-crafted features, which limits their generalization ability and slows down the parsing speed. In this paper, we propose a general and effective Neural Network model for graph-based dependency parsing. Our model can automatically learn high-order feature combinations using only atomic features by exploiting a novel activation function tanhcube. Moreover, we propose a simple yet effective way to utilize phrase-level information that is expensive to use in conventional graph-based parsers. Experiments on the English Penn Treebank show that parsers based on our model perform better than conventional graph-based parsers.

7. **Deep neural networks for learning graph representations** [17]

   Shaosheng Cao, Wei Lu, Qiongkai Xu

   AAAI 2016

   This paper proposes a novel model for learning graph representations, which generates a low-dimensional vector representation for each vertex by capturing the graph structural information. Different from other previous research efforts, this paper adopts a random surfing model to capture graph structural information directly, instead of using the samplingbased method for generating linear sequences proposed by Perozzi et al. (2014). The advantages of proposed approach will be illustrated from both theorical and empirical perspectives. This paper also gives a new perspective for the matrix factorization method proposed by Levy and Goldberg (2014), in which the pointwise mutual information (PMI) matrix is considered as an analytical solution to the objective function of the skipgram model with negative sampling proposed by Mikolov et al. (2013). Unlike their approach which involves the use of the SVD for finding the low-dimensitonal projections from the PMI matrix, however, the stacked denoising autoencoder is introduced in our model to extract complex features and model non-linearities. To demonstrate the effectiveness of the proposed model, this paper conducts experiments on clustering and visualization tasks, employing the learned vertex representations as features. Empirical results on datasets of varying sizes show that the proposed model outperforms other stat-of-the-art models in such tasks.

8. **Dynamic Edge-Conditioned Filters in Convolutional Neural Networks on Graphs** [113]

   Martin Simonovsky, Nikos Komodakis

   Accepted to CVPR 2017; extended version

   A number of problems can be formulated as prediction on graph-structured data. In this work, we generalize the convolution operator from regular grids to arbitrary graphs while avoiding the spectral domain, which allows us to handle graphs of varying size and connectivity. To move beyond a simple diffusion, filter weights are conditioned on the specific edge labels in the neighborhood of a vertex. Together with the proper choice of graph coarsening, we explore constructing deep neural networks for graph classification. In particular, we demonstrate the generality of our formulation in point cloud classification, where we set the new state of the art, and on a graph classification dataset, where we outperform other deep learning approaches.

9. **Leveraging Deep Neural Networks and Knowledge Graphs for Entity Disambiguation** [51]

   Hongzhao Huang, Larry Heck , Heng Ji

   Entity Disambiguation aims to link mentions of ambiguous entities to a knowledge base (e.g., Wikipedia). Modeling topical coherence is crucial for this task based on the assumption that information from the same semantic context tends to belong to the same topic. This paper presents a novel deep semantic relatedness model (DSRM) based on deep neural networks (DNN) and semantic knowledge graphs (KGs) to measure entity semantic relatedness for topical coherence modeling. The DSRM is directly trained on large-scale KGs and it maps heterogeneous types of knowledge of an entity from KGs to numerical feature vectors in a latent space such that the distance between two semantically-related entities is minimized. Compared with the state-ofthe-art relatedness approach proposed by (Milne and Witten, 2008a), the DSRM obtains 19.4% and 24.5% reductions in entity disambiguation errors on two publicly available datasets respectively.

10. **Geometric deep learning on graphs and manifolds using mixture model CNNs** [86]

    Federico Monti, Davide Boscaini, Jonathan Masci, Emanuele Rodola, Jan Svoboda, Michael M. Bronstein

    Deep learning has achieved a remarkable performance breakthrough in several fields, most notably in speech recognition, natural language processing, and computer vision. In particular, convolutional neural network (CNN) architectures currently produce state-of-the-art performance on a variety of image analysis tasks such as object detection and recognition. Most of deep learning research has so far focused on dealing with 1D, 2D, or 3D Euclidean structured data such as acoustic signals, images, or videos. Recently, there has been an increasing interest in geometric deep learning, attempting to generalize deep learning methods to non-Euclidean structured data such as graphs and manifolds, with a variety of applications from the domains of network analysis, computational social science, or computer graphics. In this paper, we propose a unified framework allowing to generalize CNN architectures to non-Euclidean domains (graphs and manifolds) and learn local, stationary, and compositional task-specific features. We show that various non-Euclidean CNN methods previously proposed in the literature can be considered as particular instances of our framework. We test the proposed method on standard tasks from the realms of image, graph and 3D shape analysis and show that it consistently outperforms previous approaches.

11. **A Hybrid Spectral Clustering and Deep Neural Network Ensemble Algorithm for Intrusion Detection in Sensor Networks** [78]

    Tao Ma, Fen Wang, Jianjun Cheng, Yang Yu and Xiaoyun Chen

    Sensors 2016

    The development of intrusion detection systems (IDS) that are adapted to allow routers and network defense systems to detect malicious network traffic disguised as network protocols or normal access is a critical challenge. This paper proposes a novel approach called SCDNN, which combines spectral clustering (SC) and deep neural network (DNN) algorithms. First, the dataset is divided into k subsets based on sample similarity using cluster centers, as in SC. Next, the distance between data points in a testing set and the training set is measured based on similarity features and is fed into the deep neural network algorithm for intrusion detection. Six KDD-Cup99 and NSL-KDD datasets and a sensor network dataset were employed to test the performance of the model. These experimental results indicate that the SCDNN classifier not only performs better than back propagation neural network (BPNN), support vector machine (SVM), random forest (RF) and Bayes tree models in detection accuracy and the types of abnormal attacks found. It also provides an effective tool of study and analysis of intrusion detection in large networks.

12. **Deep Convolutional Networks on Graph-Structured Data** [49]

    Mikael Henaff, Joan Bruna

    Deep Learning's recent successes have mostly relied on Convolutional Networks, which exploit fundamental statistical properties of images, sounds and video data: the local stationarity and multi-scale compositional structure, that allows expressing long range interactions in terms of shorter, localized interactions. However, there exist other important examples, such as text documents or bioinformatic data, that may lack some or all of these strong statistical regularities. In this paper we consider the general question of how to construct deep architectures with small learning complexity on general non-Euclidean domains, which are typically unknown and need to be estimated from the data. In particular, we develop an extension of Spectral Networks which incorporates a Graph Estimation procedure, that we test on large-scale classification problems, matching or improving over Dropout Networks with far less parameters to estimate.

13. **Adverse Drug Event Detection in Tweets with Semi-Supervised Convolutional Neural Networks** [67]

    Kathy Lee, Ashequl Qadir, Sadid A. Hasan, Vivek Datla

    WWW 2017

Current Adverse Drug Events (ADE) surveillance systems are often associated with a sizable time lag before such events are published. Online social media such as Twitter could describe adverse drug events in real-time, prior to official reporting. Deep learning has significantly improved text classification performance in recent years and can potentially enhance ADE classification in tweets. However, these models typically require large corpora with human expert-derived labels, and such resources are very expensive to generate and are hardly available. Semi-supervised deep learning models, which offer a plausible alternative to fully supervised models, involve the use of a small set of labeled data and a relatively larger collection of unlabeled data for training. Traditionally, these models are trained on labeled and unlabeled data from similar topics or domains. In reality, millions of tweets generated daily often focus on disparate topics, and this could present a challenge for building deep learning models for ADE classification with random Twitter stream as unlabeled training data. In this work, we build several semi-supervised convolutional neural network (CNN) models for ADE classification in tweets, specifically leveraging different types of unlabeled data in developing the models to address the problem. We demonstrate that, with the selective use of a variety of unlabeled data, our semi-supervised CNN models outperform a strong state-of the-art supervised classification model by +9.9% F1-score. We evaluated our models on the Twitter data set used in the PSB 2016 Social Media Shared Task. Our results present the new state-of-the-art for this data set.

14. **Tri-Party Deep Network Representation** [94]

    Shirui Pan, Jia Wu, Xingquan Zhu, Chengqi Zhang, Yang Wang

    IJCAI 2016

    Information network mining often requires examination of linkage relationships between nodes for analysis. Recently, network representation has emerged to represent each node in a vector format, embedding network structure, so off-the-shelf machine learning methods can be directly applied for analysis. To date, existing methods only focus on one aspect of node information and cannot leverage node labels. In this paper, we propose TriDNR, a tri-party deep network representation model, using information from three parties: node structure, node content, and node labels (if available) to jointly learn optimal node representation. TriDNR is based on our new coupled deep natural language module, whose learning is enforced at three levels: (1) at the network structure level, TriDNR exploits inter-node relationship by maximizing the probability of observing surrounding nodes given a node in random walks; (2) at the node content level, TriDNR captures node-word correlation by maximizing the co-occurrence of word sequence given a node; and (3) at the node label level, TriDNR models label-word correspondence by maximizing the probability of word sequence given a class label. The tri-party information is jointly fed into the neural network model to mutually enhance each other to learn optimal representation, and results in up to 79% classification accuracy gain, compared to state-of-the-art methods.

15. **A General Framework for Content-enhanced Network Representation Learning** [117]

    Xiaofei Sun, Jiang Guo, Xiao Ding, Ting Liu

    arXiv 2016

    This paper investigates the problem of network embedding, which aims at learning low-dimensional vector representation of nodes in networks. Most existing network embedding methods rely solely on the network structure, i.e., the linkage relationships between nodes, but ignore the rich content information associated with it, which is common in real world networks and beneficial to describing the characteristics of a node. In this paper, we propose content-enhanced network embedding (CENE), which is capable of jointly leveraging the network structure and the content information. Our approach integrates text modeling and structure modeling in a general framework by treating the content information as a special kind of node. Experiments on several real world net- works with application to node classification show that our models outperform all existing network embedding methods, demonstrating the merits of content information and joint learning.

16. **Unsupervised Deep Haar Scattering on Graphs** [26]

    Xu Chen, Xiuyuan Cheng, and Stephane Mallat

    NIPS 2014

The classification of high-dimensional data defined on graphs is particularly diffi- cult when the graph geometry is unknown. We introduce a Haar scattering transform on graphs, which computes invariant signal descriptors. It is implemented with a deep cascade of additions, subtractions and absolute values, which iteratively compute orthogonal Haar wavelet transforms. Multiscale neighborhoods of unknown graphs are estimated by minimizing an average total variation, with a pair matching algorithm of polynomial complexity. Supervised classification with dimension reduction is tested on data bases of scrambled images, and for signals sampled on unknown irregular grids on a sphere.

17. **Spectral Networks and Deep Locally Connected Networks on Graphs** [11]

    Joan Bruna, Wojciech Zaremba, Arthur Szlam, Yann LeCun

    Convolutional Neural Networks are extremely efficient architectures in image and audio recognition tasks, thanks to their ability to exploit the local translational invariance of signal classes over their domain. In this paper authors consider possible generalizations of CNNs to signals defined on more general domains without the action of a translation group. In particular, we propose two constructions, one based upon a hierarchical clustering of the domain, and another based on the spectrum of the graph Laplacian. This paper shows through experiments that for low dimensional graphs it is possible to learn convolutional layers with a number of parameters independent of the input size, resulting in efficient deep architectures.

18. **Semi-supervised Learning for Convolutional Neural Networks via Online Graph Construction** [5]

    Semi-supervised Learning for Convolutional Neural Networks via Online Graph Construction

    Sheng-Yi Bai, Sebastian Agethen, Ting-Hsuan Chao and Winston Hsu

    The recent promising achievements of deep learning rely on the large amount of labeled data. Considering the abundance of data on the web, most of them do not have labels at all. Therefore, it is important to improve generalization performance using unlabeled data on supervised tasks with few labeled instances. In this work, authors revisit graph-based semi-supervised learning algorithms and propose an online graph construction technique which suits deep convolutional neural network better. This paper considers an EM-like algorithm for semi-supervised learning on deep neural networks: In forward pass, the graph is constructed based on the network output, and the graph is then used for loss calculation to help update the network by back propagation in the backward pass. This paper demonstrates the strength of our online approach compared to the conventional ones whose graph is constructed on static but not robust enough feature representations beforehand.

19. **Gated graph sequence neural networks** [71]

    Li, Yujia and Tarlow, Daniel and Brockschmidt, Marc and Zemel, Richard

    ICLR 2016

    Graph-structured data appears frequently in domains including chemistry, natural language semantics, social networks, and knowledge bases. In this work, we study feature learning techniques for graph-structured inputs. Our starting point is previous work on Graph Neural Networks (Scarselli et al., 2009), which we modify to use gated recurrent units and modern optimization techniques and then extend to output sequences. The result is a flexible and broadly useful class of neural network models that has favorable inductive biases relative to purely sequence-based models (e.g., LSTMs) when the problem is graph-structured. We demonstrate the capabilities on some simple AI (bAbI) and graph algorithm learning tasks. We then show it achieves state-of-the-art performance on a problem from program verification, in which subgraphs need to be matched to abstract data structures.

20. **Learning Graphical State Transitions** [57]

    Johnson, Daniel D

    ICLR 2017

    Graph-structured data is important in modeling relationships between multiple entities, and can be used to represent states of the world as well as many data structures. Li et al. (2016) describe a

model known as a Gated Graph Sequence Neural Network (GGS-NN) that produces sequences from graph-structured input. In this work I introduce the Gated Graph Transformer Neural Network (GGT-NN), an extension of GGS-NNs that uses graph-structured data as an intermediate representation. The model can learn to construct and modify graphs in sophisti- cated ways based on textual input, and also to use the graphs to produce a variety of outputs. For example, the model successfully learns to solve almost all of the bAbI tasks (Weston et al., 2016), and also discovers the rules governing graphical formulations of a simple cellular automaton and a family of Turing machines.

21. **Semi-supervised classification with graph convolutional networks** [63]

    Kipf, Thomas N and Welling, Max

    ICLR 2017

    We present a scalable approach for semi-supervised learning on graph-structured data that is based on an efficient variant of convolutional neural networks which operate directly on graphs. We motivate the choice of our convolutional architecture via a localized first-order approximation of spectral graph convolutions. Our model scales linearly in the number of graph edges and learns hidden layer representations that encode both local graph structure and features of nodes. In a number of experiments on citation networks and on a knowledge graph dataset we demonstrate that our approach outperforms related methods by a significant margin.

22. **Protein contact prediction from amino acid co-evolution using convolutional networks for graph-valued images** [42]

    Golkov, Vladimir and Skwark, Marcin J and Golkov, Antonij and Dosovitskiy, Alexey and Brox, Thomas and Meiler, Jens and Cremers, Daniel

    NIPS 2016

    Proteins are responsible for most of the functions in life, and thus are the central focus of many areas of biomedicine. Protein structure is strongly related to protein function, but is difficult to elucidate experimentally, therefore computational structure prediction is a crucial task on the way to solve many biological questions. A contact map is a compact representation of the three-dimensional structure of a protein via the pairwise contacts between the amino acids constituting the protein. We use a convolutional network to calculate protein contact maps from detailed evolutionary coupling statistics between positions in the protein sequence. The input to the network has an image-like structure amenable to convolutions, but every "pixel" instead of color channels contains a bipartite undirected edge-weighted graph. We propose several methods for treating such "graph-valued images" in a convolutional network. The proposed method outperforms state-of-the-art methods by a considerable margin.

23. **Generative models of rich clubs in Hebbian neuronal networks and large-scale human brain networks** [127]

    Petra E. Vértes, Aaron Alexander-Bloch, and Edward T. Bullmore

    Phil. Trans. R. Soc. B

    Rich clubs arise when nodes that are 'rich' in connections also form an elite, densely connected 'club'. In brain networks, rich clubs incur high physical connection costs but also appear to be especially valuable to brain function. However, little is known about the selection pressures that drive their formation. Here, we take two complementary approaches to this question: firstly we show, using generative modelling, that the emergence of rich clubs in large-scale human brain networks can be driven by an economic trade-off between connection costs and a second, competing topological term. Secondly we show, using simulated neural networks, that Hebbian learning rules also drive the emergence of rich clubs at the microscopic level, and that the prominence of these features increases with learning time. These results suggest that Hebbian learning may provide a neuronal mechanism for the selection of complex features such as rich clubs. The neural networks that we investigate are explicitly Hebbian, and we argue that the topological term in our model of large-scale brain connectivity may represent an analogous connection rule. This putative link between learning and rich clubs is also consistent with predictions that integrative aspects of brain network organization are especially important for adaptive behaviour.

24. **DeepGraph: Graph Structure Predicts Network Growth** [68]

Cheng Li, Xiaoxiao Guo, Qiaozhu Mei

arXiv

The topological (or graph) structures of real-world networks are known to be predictive of multiple dynamic properties of the networks. Conventionally, a graph structure is represented using an adjacency matrix or a set of hand-crafted structural features. These representations either fail to highlight local and global properties of the graph or suffer from a severe loss of structural information. There lacks an effective graph representation, which hinges the realization of the predictive power of network structures. This paper proposes to learn the represention of a graph, or the topological structure of a network, through a deep learning model. This end-to-end prediction model, named DeepGraph, takes the input of the raw adjacency matrix of a real-world network and outputs a prediction of the growth of the network. The adjacency matrix is first represented using a graph descriptor based on the heat kernel signature, which is then passed through a multi-column, multi-resolution convolutional neural network. Extensive experiments on five large collections of real-world networks demonstrate that the proposed prediction model significantly improves the effectiveness of existing methods, including linear or nonlinear regressors that use hand-crafted features, graph kernels, and competing deep learning methods.

25. **Spectral Networks and Locally Connected Networks on Graphs** [12]

Joan Bruna, Wojciech Zaremba, Arthur Szlam, Yann LeCun

Convolutional Neural Networks are extremely efficient architectures in image and audio recognition tasks, thanks to their ability to exploit the local translational invariance of signal classes over their domain. In this paper we consider possible generalizations of CNNs to signals defined on more general domains without the action of a translation group. In particular, we propose two constructions, one based upon a hierarchical clustering of the domain, and another based on the spectrum of the graph Laplacian. We show through experiments that for low-dimensional graphs it is possible to learn convolutional layers with a number of parameters independent of the input size, resulting in efficient deep architectures.

26. **Graph Based Convolutional Neural Network** [37]

Michael Edwards, Xianghua Xie

BMVC 2016

The benefit of localized features within the regular domain has given rise to the use of Convolutional Neural Networks (CNNs) in machine learning, with great proficiency in the image classification. The use of CNNs becomes problematic within the irregular spatial domain due to design and convolution of a kernel filter being non-trivial. One solution to this problem is to utilize graph signal processing techniques and the convolution theorem to perform convolutions on the graph of the irregular domain to obtain feature map responses to learnt filters. We propose graph convolution and pooling operators analogous to those in the regular domain. We also provide gradient calculations on the input data and spectral filters, which allow for the deep learning of an irregular spatial domain problem. Signal filters take the form of spectral multipliers, applying convolution in the graph spectral domain. Applying smooth multipliers results in localized convolutions in the spatial domain, with smoother multipliers providing sharper feature maps. Algebraic Multigrid is presented as a graph pooling method, reducing the resolution of the graph through agglomeration of nodes between layers of the network. Evaluation of performance on the MNIST digit classification problem in both the regular and irregular domain is presented, with comparison drawn to standard CNN. The proposed graph CNN provides a deep learning method for the irregular domains present in the machine learning community, obtaining 94.23% on the regular grid, and 94.96% on a spatially irregular subsampled MNIST.

27. **A Review on Deep Learning Techniques Applied to Semantic Segmentation** [39]

A. Garcia-Garcia, S. Orts-Escolano, S.O. Oprea, V. Villena-Martinez, and J. Garcia-Rodriguez

arXiv 2017

Image semantic segmentation is more and more being of interest for computer vision and machine learning researchers. Many applications on the rise need accurate and efficient segmentation mechanisms: autonomous driving, indoor navigation, and even virtual or augmented reality systems to name a few. This demand coincides with the rise of deep learning approaches in almost every field or application target related to computer vision, including semantic segmentation or scene understanding. This paper provides a review on deep learning methods for semantic segmentation applied to various application areas. Firstly, we describe the terminology of this field as well as mandatory background concepts. Next, the main datasets and challenges are exposed to help researchers decide which are the ones that best suit their needs and their targets. Then, existing methods are reviewed, highlighting their contributions and their significance in the field. Finally, quantitative results are given for the described methods and the datasets in which they were evaluated, following up with a discussion of the results. At last, we point out a set of promising future works and draw our own conclusions about the state of the art of semantic segmentation using deep learning techniques.

28. **SyncSpecCNN: Synchronized Spectral CNN for 3D Shape Segmentation** [144]

    Li Yi, Hao Su, Xingwen Guo, Leonidas Guibas

    arXiv

    This paper studies the problem of semantic annotation on 3D models that are represented as shape graphs. A functional view is taken to represent localized information on graphs, so that annotations such as part segment or keypoint are nothing but 0-1 indicator vertex functions. Compared with images that are 2D grids, shape graphs are irregular and non-isomorphic data structures. To enable the prediction of vertex functions on them by convolutional neural networks, we resort to spectral CNN method that enables weight sharing by parameterizing kernels in the spectral domain spanned by graph laplacian eigenbases. Under this setting, our network, named SyncSpecCNN, strive to overcome two key challenges: how to share coefficients and conduct multi-scale analysis in different parts of the graph for a single shape, and how to share information across related but different shapes that may be represented by very different graphs. Towards these goals, we introduce a spectral parameterization of dilated convolutional kernels and a spectral transformer network. Experimentally we tested our SyncSpecCNN on various tasks, including 3D shape part segmentation and 3D keypoint prediction. State-of-the-art performance has been achieved on all benchmark datasets.

29. **SyncSpecCNN: Synchronized Spectral CNN for 3D Shape Segmentation** [110]

    Michael Schlichtkrull, Thomas N. Kipf, Peter Bloem, Rianne van den Berg, Ivan Titov, Max Welling

    arXiv2017

    Knowledge bases play a crucial role in many applications, for example question answering and information retrieval. Despite the great effort invested in creating and maintaining them, even the largest representatives (e.g., Yago, DBPedia or Wikidata) are highly incomplete. We introduce relational graph convolutional networks (R-GCNs) and apply them to two standard knowledge base completion tasks: link prediction (recovery of missing facts, i.e. subject-predicate-object triples) and entity classification (recovery of missing attributes of entities). R-GCNs are a generalization of graph convolutional networks, a recent class of neural networks operating on graphs, and are developed specifically to deal with highly multi-relational data, characteristic of realistic knowledge bases. Our methods achieve competitive results on standard benchmarks for both tasks.

30. **Multi-view Convolutional Neural Networks for 3D Shape Recognition** [115]

    Hang Su, Subhransu Maji, Evangelos Kalogerakis, Erik Learned-Miller

    ICCV 2015

    A longstanding question in computer vision concerns the representation of 3D shapes for recognition: should 3D shapes be represented with descriptors operating on their native 3D formats, such as voxel grid or polygon mesh, or can they be effectively represented with view-based descriptors? We address this question in the context of learning to recognize 3D shapes from a collection of their rendered views on 2D images. We first present a standard CNN architecture trained to recognize the shapes' rendered

views independently of each other, and show that a 3D shape can be recognized even from a single view at an accuracy far higher than using state-of-the-art 3D shape descriptors. Recognition rates further increase when multiple views of the shapes are provided. In addition, we present a novel CNN architecture that combines information from multiple views of a 3D shape into a single and compact shape descriptor offering even better recognition performance. The same architecture can be applied to accurately recognize human hand-drawn sketches of shapes. We conclude that a collection of 2D views can be highly informative for 3D shape recognition and is amenable to emerging CNN architectures and their derivatives.

31. **Image Style Transfer Using Convolutional Neural Networks** [40]

    Leon A. Gatys, Alexander S. Ecker, Matthias Bethge

    CVPR 2016

    Rendering the semantic content of an image in different styles is a difficult image processing task. Arguably, a major limiting factor for previous approaches has been the lack of image representations that explicitly represent semantic information and, thus, allow to separate image content from style. Here we use image representations derived from Convolutional Neural Networks optimised for object recognition, which make high level image information explicit. We introduce A Neural Algorithm of Artistic Style that can separate and recombine the image content and style of natural images. The algorithm allows us to produce new images of high perceptual quality that combine the content of an arbitrary photograph with the appearance of numerous wellknown artworks. Our results provide new insights into the deep image representations learned by Convolutional Neural Networks and demonstrate their potential for high level image synthesis and manipulation.

32. **Unifying local and non-local signal processing with graph CNNs** [102]

    Gilles Puy, Srdan Kitic, Patrick Pérez

    arXiv2017

    This paper deals with the unification of local and non-local signal processing on graphs within a single convolutional neural network (CNN) framework. Building upon recent works on graph CNNs, we propose to use convolutional layers that take as inputs two variables, a signal and a graph, allowing the network to adapt to changes in the graph structure. This also allows us to learn through training the optimal mixing of locality and non-locality, in cases where the graph is built on the input signal itself. We demonstrate the versatility and the effectiveness of our framework on several types of signals (greyscale and color images, color palettes and speech signals) and on several applications (style transfer, color transfer, and denoising).

33. **Spectral Graph Convolutions for Population-based Disease Prediction** [95]

    Sarah Parisot, Sofia Ira Ktena, Enzo Ferrante, Matthew Lee, Ricardo Guerrerro Moreno, Ben Glocker, Daniel Rueckert

    arXiv2017

    Exploiting the wealth of imaging and non-imaging information for disease prediction tasks requires models capable of representing, at the same time, individual features as well as data associations between subjects from potentially large populations. Graphs provide a natural framework for such tasks, yet previous graph-based approaches focus on pairwise similarities without modelling the subjects' individual characteristics and features. On the other hand, relying solely on subject-specific imaging feature vectors fails to model the interaction and similarity between subjects, which can reduce performance. In this paper, we introduce the novel concept of Graph Convolutional Networks (GCN) for brain analysis in populations, combining imaging and non-imaging data. We represent populations as a sparse graph where its vertices are associated with image-based feature vectors and the edges encode phenotypic information. This structure was used to train a GCN model on partially labelled graphs, aiming to infer the classes of unlabelled nodes from the node features and pairwise associations between subjects. We demonstrate the potential of the method on the challenging ADNI and ABIDE databases, as a proof of concept of the benefit from integrating contextual information in classification tasks. This has

a clear impact on the quality of the predictions, leading to 69.5% accuracy for ABIDE (outperforming the current state of the art of 66.8%) and 77% for ADNI for prediction of MCI conversion, significantly outperforming standard linear classifiers where only individual features are considered.

34. **Distance Metric Learning using Graph Convolutional Networks: Application to Functional Brain Networks** [65]

    Sofia Ira Ktena, Sarah Parisot, Enzo Ferrante, Martin Rajchl, Matthew Lee, Ben Glocker, Daniel Rueckert

    arXiv2017

    Evaluating similarity between graphs is of major importance in several computer vision and pattern recognition problems, where graph representations are often used to model objects or interactions between elements. The choice of a distance or similarity metric is, however, not trivial and can be highly dependent on the application at hand. In this work, we propose a novel metric learning method to evaluate distance between graphs that leverages the power of convolutional neural networks, while exploiting concepts from spectral graph theory to allow these operations on irregular graphs. We demonstrate the potential of our method in the field of connectomics, where neuronal pathways or functional connections between brain regions are commonly modelled as graphs. In this problem, the definition of an appropriate graph similarity function is critical to unveil patterns of disruptions associated with certain brain disorders. Experimental results on the ABIDE dataset show that our method can learn a graph similarity metric tailored for a clinical application, improving the performance of a simple k-nn classifier by 11.9% compared to a traditional distance metric.

35. **Robust Spatial Filtering with Graph Convolutional Neural Networks** [116]

    Felipe Petroski Such, Shagan Sah, Miguel Dominguez, Suhas Pillai, Chao Zhang, Andrew Michael, Nathan Cahill, Raymond Ptucha

    arXiv2017

    Convolutional Neural Networks (CNNs) have recently led to incredible breakthroughs on a variety of pattern recognition problems. Banks of finite impulse response filters are learned on a hierarchy of layers, each contributing more abstract information than the previous layer. The simplicity and elegance of the convolutional filtering process makes them perfect for structured problems such as image, video, or voice, where vertices are homogeneous in the sense of number, location, and strength of neighbors. The vast majority of classification problems, for example in the pharmaceutical, homeland security, and financial domains are unstructured. As these problems are formulated into unstructured graphs, the heterogeneity of these problems, such as number of vertices, number of connections per vertex, and edge strength, cannot be tackled with standard convolutional techniques. We propose a novel neural learning framework that is capable of handling both homogeneous and heterogeneous data, while retaining the benefits of traditional CNN successes. Recently, researchers have proposed variations of CNNs that can handle graph data. In an effort to create learnable filter banks of graphs, these methods either induce constraints on the data or require considerable preprocessing. As opposed to defining filters as spectral multipliers applied to the eigenvectors of the graph Laplacian, our framework, which we term Graph-CNNs, defines filters as polynomials of functions of the graph adjacency matrix. Graph-CNNs can handle both heterogeneous and homogeneous graph data, including graphs having entirely different vertex or edge sets. We compare Graph-CNN to traditional CNNs using the CIFAR-10 and Imagenet image classification datasets.

36. **3D Shape Segmentation via Shape Fully Convolutional Networks** [134]

    Pengyu Wang, Yuan Gan, Yan Zhang, Panpan Shui

    arVix2017

    This paper proposes a novel fully convolutional networks architecture for shapes, denoted as Shape Fully Convolutional Networks (SFCN). Similar to convolution and pooling operation on image, the 3D shape is represented as a graph structure in the SFCN architecture, based on which we first propose and implement shape convolution and pooling operation. Meanwhile, to build our SFCN architecture

in the original image segmentation FCN architecture, we also design and implement the generating operation with bridging function. This ensures that the convolution and pooling operation we designed can be successfully applied in the original FCN architecture. In this paper,we also present a new shape segmentation based on SFCN. In contrast to existing state-of-the-art shape segmentation methods that require the same types of shapes as input, we allow the more general and challenging input such as mixed datasets of different types of shapes. In our approach, SFCNs are first trained end-to-end, triangles-to-triangles by three low-level geometric features. Then, based on the trained SFCNs, we can complete the shape segmentation task with high quality. Finally, The feature voting-based multilabel graph cuts is adopted to optimize the segmentation results obtained by SFCN prediction. The experiment results show that our method can effectively learn and predict mixed shape datasets of either similar or different characters, and achieve excellent segmentation results.

37. **An efficient neural network based method for medical image segmentation** [124]

    Nima Torbati, Ahmad Ayatollahi, Ali Kermani

    Computers in Biology and Medicine 2014

    The aim of this research is to propose a new neural network based method for medical image segmentation. Firstly, a modified self-organizing map (SOM) network, named moving average SOM (MA-SOM), is utilized to segment medical images. After the initial segmentation stage, a merging process is designed to connect the objects of a joint cluster together. A two-dimensional (2D) discrete wavelet transform (DWT) is used to build the input feature space of the network. The experimental results show that MA-SOM is robust to noise and it determines the input image pattern properly. The segmentation results of breast ultrasound images (BUS) demonstrate that there is a significant correlation between the tumor region selected by a physician and the tumor region segmented by our proposed method. In addition, the proposed method segments X-ray computerized tomography (CT) and magnetic resonance (MR) head images much better than the incremental supervised neural network (ISNN) and SOM-based methods.

38. **ReSeg: A Recurrent Neural Network-based Model for Semantic Segmentation** [129]

    Francesco Visin, Marco Ciccone, Adriana Romero, Kyle Kastner, Kyunghyun Cho, Yoshua Bengio, Matteo Matteucci, Aaron Courville

    CVPR, Workshops, 2016

    This paper proposes a structured prediction architecture, which exploits the local generic features extracted by Convolutional Neural Networks and the capacity of Recurrent Neural Networks (RNN) to retrieve distant dependencies. The proposed architecture, called ReSeg, is based on the recently introduced ReNet model for image classification. We modify and extend it to perform the more challenging task of semantic segmentation. Each ReNet layer is composed of four RNN that sweep the image horizontally and vertically in both directions, encoding patches or activations, and providing relevant global information. Moreover, ReNet layers are stacked on top of pre-trained convolutional layers, benefiting from generic local features. Upsampling layers follow ReNet layers to recover the original image resolution in the final predictions. The proposed ReSeg architecture is efficient, flexible and suitable for a variety of semantic segmentation tasks. We evaluate ReSeg on several widely-used semantic segmentation datasets: Weizmann Horse, Oxford Flower, and CamVid; achieving stateof-the-art performance. Results show that ReSeg can act as a suitable architecture for semantic segmentation tasks, and may have further applications in other structured prediction problems. The source code and model hyperparameters are available on https://github.com/fvisin/reseg.

39. **Mapping Stacked Decision Forests to Deep and Sparse Convolutional Neural Networks for Semantic Segmentation** [104]

    David L Richmond, Dagmar Kainmueller, Michael Yang, Eugene W. Myers, and Carsten Rother

    arXiv 2015

    Authors consider the task of pixel-wise semantic segmentation given a small set of labeled training images. Among two of the most popular techniques to address this task are Random Forests (RF) and

Neural Networks (NN). The main contribution of this work is to explore the relationship between two special forms of these techniques: stacked RFs and deep Convolutional Neural Networks (CNN). We show that there exists a mapping from stacked RF to deep CNN, and an approximate mapping back. This insight gives two major practical benefits: Firstly, deep CNNs can be intelligently constructed and initialized, which is crucial when dealing with a limited amount of training data. Secondly, it can be utilized to create a new stacked RF with improved performance. Furthermore, this mapping yields a new CNN architecture, that is well suited for pixel-wise semantic labeling. We experimentally verify these practical benefits for two different application scenarios in computer vision and biology, where the layout of parts is important: Kinect-based body part labeling from depth images, and somite segmentation in microscopy images of developing zebrafish.

40. **Combining Fully Convolutional and Recurrent Neural Networks for 3D Biomedical Image Segmentation** [22]

    Jianxu Chen, Lin Yang, Yizhe Zhang, Mark Alber, Danny Z. Chen

    NIPS 2016

    Segmentation of 3D images is a fundamental problem in biomedical image analysis. Deep learning (DL) approaches have achieved state-of-the-art segmentation performance. To exploit the 3D contexts using neural networks, known DL segmentation methods, including 3D convolution, 2D convolution on planes orthogonal to 2D image slices, and LSTM in multiple directions, all suffer incompatibility with the highly anisotropic dimensions in common 3D biomedical images. In this paper, we propose a new DL framework for 3D image segmentation, based on a combination of a fully convolutional network (FCN) and a recurrent neural network (RNN), which are responsible for exploiting the intra-slice and inter-slice contexts, respectively. To our best knowledge, this is the first DL framework for 3D image segmentation that explicitly leverages 3D image anisotropism. Evaluating using a dataset from the ISBI Neuronal Structure Segmentation Challenge and in-house image stacks for 3D fungus segmentation, our approach achieves promising results comparing to the known DL-based 3D segmentation approaches.

41. **Recurrent Fully Convolutional Neural Networks for Multi-slice MRI Cardiac Segmentation** [101]

    Rudra P K Poudel, Pablo Lamata, Giovanni Montana

    MICCAI Workshop RAMBO 2016

    In cardiac magnetic resonance imaging, fully-automatic segmentation of the heart enables precise structural and functional measurements to be taken, e.g. from short-axis MR images of the left-ventricle. In this work we propose a recurrent fully-convolutional network (RFCN) that learns image representations from the full stack of 2D slices and has the ability to leverage inter-slice spatial dependences through internal memory units. RFCN combines anatomical detection and segmentation into a single architecture that is trained end-to-end thus significantly reducing computational time, simplifying the segmentation pipeline, and potentially enabling real-time applications. We report on an investigation of RFCN using two datasets, including the publicly available MICCAI 2009 Challenge dataset. Comparisons have been carried out between fully convolutional networks and deep restricted Boltzmann machines, including a recurrent version that leverages inter-slice spatial correlation. Our studies suggest that RFCN produces state-of-the-art results and can substantially improve the delineation of contours near the apex of the heart.

42. **Automatic Liver and Tumor Segmentation of CT and MRI Volumes Using Cascaded Fully Convolutional Neural Networks** [30]

    Patrick Ferdinand Christ, Florian Ettlingera, Felix Grün, Mohamed Ezzeldin A, Elshaer et. al.

    arXiv 2017

    Automatic segmentation of the liver and hepatic lesions is an important step towards deriving quantitative biomarkers for accurate clinical diagnosis and computer-aided decision support systems. This paper presents a method to automatically segment liver and lesions in CT and MRI abdomen images using cascaded fully convolutional neural networks (CFCNs) enabling the segmentation of large-scale

medical trials and quantitative image analyses. We train and cascade two FCNs for the combined segmentation of the liver and its lesions. As a first step, we train an FCN to segment the liver as ROI input for a second FCN. The second FCN solely segments lesions within the predicted liver ROIs of step 1. CFCN models were trained on an abdominal CT dataset comprising 100 hepatic tumor volumes. Validation results on further datasets show that CFCN-based semantic liver and lesion segmentation achieves Dice scores over 94% for the liver with computation times below 100s per volume. We further experimentally demonstrate the robustness of the proposed method on 38 MRI liver tumor volumes and the public 3DIRCAD dataset.

43. **Fully Convolutional Networks for Semantic Segmentation** [112]

    Evan Shelhamer, Jonathan Long, Trevor Darrell

    IEEE transactions on pattern analysis and machine intelligence

    Convolutional networks are powerful visual models that yield hierarchies of features. Authors show that convolutional networks by themselves, trained end-to-end, pixels-to-pixels, improve on the previous best result in semantic segmentation. Their key insight is to build "fully convolutional" networks that take input of arbitrary size and produce correspondingly-sized output with efficient inference and learning. We define and detail the space of fully convolutional networks, explain their application to spatially dense prediction tasks, and draw connections to prior models. We adapt contemporary classification networks (AlexNet, the VGG net, and GoogLeNet) into fully convolutional networks and transfer their learned representations by fine-tuning to the segmentation task. We then define a skip architecture that combines semantic information from a deep, coarse layer with appearance information from a shallow, fine layer to produce accurate and detailed segmentations. Our fully convolutional networks achieve improved segmentation of PASCAL VOC (30% relative improvement to 67.2% mean IU on 2012), NYUDv2, SIFT Flow, and PASCAL-Context, while inference takes one tenth of a second for a typical image.

44. **Automatic 3D liver location and segmentation via convolutional neural network and graph cut** [75]

    Fang Lu, Fa Wu, Peijun Hu, Zhiyi Peng, Dexing Kong

    Int J CARS 2017

    Purpose Segmentation of the liver from abdominal computed tomography (CT) images is an essential step in some computer-assisted clinical interventions, such as surgery planning for living donor liver transplant, radiotherapy and volume measurement. In this work, we develop a deep learning algorithm with graph cut refinement to automatically segment the liver in CT scans. Methods The proposed method consists of two main steps: (i) simultaneously liver detection and probabilistic segmentation using 3D convolutional neural network; (ii) accuracy refinement of the initial segmentation with graph cut and the previously learned probability map. Results The proposed approach was validated on forty CT volumes taken from two public databases MICCAI-Sliver07 and 3Dircadb1. For the MICCAI-Sliver07 test dataset, the calculated mean ratios of volumetric overlap error (VOE), relative volume difference (RVD), average symmetric surface distance (ASD), root-mean-square symmetric surface distance (RMSD) and maximum symmetric surface distance (MSD) are 5.9, 2.7%, 0.91, 1.88 and 18.94 mm, respectively. For the 3Dircadb1 dataset, the calculated mean ratios of VOE, RVD, ASD, RMSD and MSD are 9.36, 0.97%, 1.89, 4.15 and 33.14 mm, respectively. Conclusions The proposed method is fully automatic without any user interaction. Quantitative results reveal that the proposed approach is efficient and accurate for hepatic volume estimation in a clinical setup. The high correlation between the automatic and manual references shows that the proposed method can be good enough to replace the time-consuming and nonreproducible manual segmentation method.

45. **Constrained Convolutional Neural Networks for Weakly Supervised Segmentation** [96]

    ICCV 2015

    This paper presents an approach to learn a dense pixel-wise labeling from image-level tags. Each image-level tag imposes constraints on the output labeling of a Convolutional Neural Network (CNN) classifier. We propose Constrained CNN (CCNN), a method which uses a novel loss function to optimize

for any set of linear constraints on the output space (i.e. predicted label distribution) of a CNN. Our loss formulation is easy to optimize and can be incorporated directly into standard stochastic gradient descent optimization. The key idea is to phrase the training objective as a biconvex optimization for linear models, which we then relax to nonlinear deep networks. Extensive experiments demonstrate the generality of our new learning framework. The constrained loss yields state-of-the-art results on weakly supervised semantic image segmentation. We further demonstrate that adding slightly more supervision can greatly improve the performance of the learning algorithm.

46. **From Image-level to Pixel-level Labeling with Convolutional Networks** [100]

    Pedro O. Pinheiro, Ronan Collobert

    CVPR 2015

    This paper is interested in inferring object segmentation by leveraging only object class information, and by considering only minimal priors on the object segmentation task. This problem could be viewed as a kind of weakly supervised segmentation task, and naturally fits the Multiple Instance Learning (MIL) framework: every training image is known to have (or not) at least one pixel corresponding to the image class label, and the segmentation task can be rewritten as inferring the pixels belonging to the class of the object (given one image, and its object class). We propose a Convolutional Neural Network-based model, which is constrained during training to put more weight on pixels which are important for classifying the image. We show that at test time, the model has learned to discriminate the right pixels well enough, such that it performs very well on an existing segmentation benchmark, by adding only few smoothing priors. Our system is trained using a subset of the Imagenet dataset and the segmentation experiments are performed on the challenging Pascal VOC dataset (with no fine-tuning of the model on Pascal VOC). Our model beats the state of the art results in weakly supervised object segmentation task by a large margin. We also compare the performance of our model with state of the art fully-supervised segmentation approaches

47. **Convolutional Recurrent Neural Networks: Learning** [152]

    Zhen Zuo, Bing Shuai1, Gang Wang, Xiao Liu, Xingxing Wang, Bing Wang, Yushi Chen

    CVPR 2015

    In existing convolutional neural networks (CNNs), both convolution and pooling are locally performed for image regions separately, no contextual dependencies between different image regions have been taken into consideration. Such dependencies represent useful spatial structure information in images. Whereas recurrent neural networks (RNNs) are designed for learning contextual dependencies among sequential data by using the recurrent (feedback) connections. In this work, we propose the convolutional recurrent neural network (C-RNN), which learns the spatial dependencies between image regions to enhance the discriminative power of image representation. The C-RNN is trained in an end-to-end manner from raw pixel images. CNN layers are firstly processed to generate middle level features. RNN layer is then learned to encode spatial dependencies. The C-RNN can learn better image representation, especially for images with obvious spatial contextual dependencies. Our method achieves competitive performance on ILSVRC 2012, SUN 397, and MIT indoor

48. **End-to-End Learning of Deformable Mixture of Parts and Deep Convolutional Neural Networks for Human Pose Estimation** [141]

    Wei Yang, Wanli Ouyang, Hongsheng Li, Xiaogang Wang

    CVPR 2016

    Recently, Deep Convolutional Neural Networks (DCNNs) have been applied to the task of human pose estimation, and have shown its potential of learning better feature representations and capturing contextual relationships. However, it is difficult to incorporate domain prior knowledge such as geometric relationships among body parts into DCNNs. In addition, training DCNN-based body part detectors without consideration of global body joint consistency introduces ambiguities, which increases the complexity of training. In this paper, we propose a novel end-to-end framework for human pose estimation that combines DCNNs with the expressive deformable mixture of parts. We explicitly incorporate domain prior knowledge into the framework, which greatly regularizes the learning process and enables

the flexibility of our framework for loopy models or tree-structured models. The effectiveness of jointly learning a DCNN with a deformable mixture of parts model is evaluated through intensive experiments on several widely used benchmarks. The proposed approach significantly improves the performance compared with state-of-the-art approaches, especially on benchmarks with challenging articulations.

49. **Reseg: A recurrent neural network for object segmentation** [130]

    Francesco Visin, Kyle Kastner, Aaron Courville, Yoshua Bengio, Matteo Matteucci? Kyunghyun Cho

    CoRR 2015

    This paper proposes a structured prediction architecture for images centered around deep recurrent neural networks. The proposed network, called ReSeg, is based on the recently introduced ReNet model for object classification. We modify and extend it to perform object segmentation, noting that the avoidance of pooling can greatly simplify pixel-wise tasks for images. Each layer of the network is composed of four recurrent neural networks that sweep the image horizontally and vertically in both directions encoding patches, along with a final layer that resizes the prediction to be of the same size of the original image. This architecture is quite flexible and makes it suitable for a variety of structured prediction tasks. We evaluate ReSeg on the specific task of object segmentation with three widely-used image segmentation datasets, namely Weizmann Horse, Fashionista and Oxford Flower. The results suggest that ReSeg can perform successfully on the object segmentation task, and may have further applications in structured prediction at large.

50. **SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation** [4]

    Vijay Badrinarayanan, Alex Kendall, Roberto Cipolla

    arXiv 2015

    This paper presents a novel and practical deep fully convolutional neural network architecture for semantic pixel-wise segmentation termed SegNet. This core trainable segmentation engine consists of an encoder network, a corresponding decoder network followed by a pixel-wise classification layer. The architecture of the encoder network is topologically identical to the 13 convolutional layers in the VGG16 network [1]. The role of the decoder network is to map the low resolution encoder feature maps to full input resolution feature maps for pixel-wise classification. The novelty of SegNet lies is in the manner in which the decoder upsamples its lower resolution input feature map(s). Specifically, the decoder uses pooling indices computed in the max-pooling step of the corresponding encoder to perform non-linear upsampling. This eliminates the need for learning to upsample. The upsampled maps are sparse and are then convolved with trainable filters to produce dense feature maps. We compare our proposed architecture with the widely adopted FCN [2] and also with the well known DeepLab-LargeFOV [3], DeconvNet [4] architectures. This comparison reveals the memory versus accuracy trade-off involved in achieving good segmentation performance. SegNet was primarily motivated by scene understanding applications. Hence, it is designed to be efficient both in terms of memory and computational time during inference. It is also significantly smaller in the number of trainable parameters than other competing architectures and can be trained end-to-end using stochastic gradient descent. We also performed a controlled benchmark of SegNet and other architectures on both road scenes and SUN RGB-D indoor scene segmentation tasks. These quantitative assessments show that SegNet provides good performance with competitive inference time and most efficient inference memory-wise as compared to other architectures. We also provide a Caffe implementation of SegNet and a web demo at http://mi.eng.cam.ac.uk/projects/segnet/.

51. **Brain Tumor Segmentation Using Large Receptive Field Deep Convolutional Neural Networks** [53]

    Fabian Isensee, Philipp Kickingereder, David Bonekamp, Martin Bendszus, Wolfgang Wick, Heinz-Peter Schlemmer, Klaus Maier-Hein

    Bildverarbeitung für die Medizin 2017

Glioblastoma segmentation is an important challenge in medical image processing. State of the art methods make use of convolutional neural networks, but generally employ only few layers and small receptive fields, which limits the amount and quality of contextual information available for segmentation. In this publication we use the well known UNet architecture to alleviate these shortcomings. We furthermore show that a sophisticated training scheme that uses dynamic sampling of training data, data augmentation and a class sensitive loss allows training such a complex architecture on relatively few data. A qualitative comparison with the state of the art shows favorable performance of our approach.

52. **Improving automated multiple sclerosis lesion segmentation with a cascaded 3D convolutional neural network approach** [125]

    Sergi Valverde, Mariano Cabezas, Eloy Roura, Sandra González-Villà, Deborah Pareto, Joan C. Vilanova, LLuís Ramió-Torrentà, Àlex Rovira, Arnau Oliver, Xavier Lladó

    NeuroImage 2017

    This paper presents a novel automated method for White Matter (WM) lesion segmentation of Multiple Sclerosis (MS) patient images. The proposed approach is based on a cascade of two 3D patch-wise convolutional neural networks (CNN). The first network is trained to be more sensitive revealing possible candidate lesion voxels while the second network is trained to reduce the number of misclassified voxels coming from the first network. This cascaded CNN architecture tends to learn well from a small ($n \leq 35$) set of labeled data of the same MRI contrast, which can be very interesting in practice, given the difficulty to obtain manual label annotations and the large amount of available unlabeled Magnetic Resonance Imaging (MRI) data. We evaluate the accuracy of the proposed method on the public MS lesion segmentation challenge MICCAI2008 dataset, comparing it with respect to other state-of-the-art MS lesion segmentation tools. Furthermore, the proposed method is also evaluated on two private MS clinical datasets, where the performance of our method is also compared with different recent public available state-of-the-art MS lesion segmentation methods. At the time of writing this paper, our method is the best ranked approach on the MICCAI2008 challenge, outperforming the rest of 60 participant methods when using all the available input modalities (T1-w, T2-w and FLAIR), while still in the top-rank (3rd position) when using only T1-w and FLAIR modalities. On clinical MS data, our approach exhibits a significant increase in the accuracy segmenting of WM lesions when compared with the rest of evaluated methods, highly correlating ($r \geq 0.97$) also with the expected lesion volume.

53. **FusionNet: A deep fully residual convolutional neural network for image segmentation in connectomics** [103]

    Tran Minh Quan, David G. C. Hildebrand, Won-Ki Jeong

    arXiv2016

    Electron microscopic connectomics is an ambitious research direction with the goal of studying comprehensive brain connectivity maps by using high-throughput, nanoscale microscopy. One of the main challenges in connectomics research is developing scalable image analysis algorithms that require minimal user intervention. Recently, deep learning has drawn much attention in computer vision because of its exceptional performance in image classification tasks. For this reason, its application to connectomic analyses holds great promise, as well. In this paper, we introduce a novel deep neural network architecture, FusionNet, for the automatic segmentation of neuronal structures in connectomics data. FusionNet leverages the latest advances in machine learning, such as semantic segmentation and residual neural networks, with the novel introduction of summation-based skip connections to allow a much deeper network architecture for a more accurate segmentation. We demonstrate the performance of the pro- posed method by comparing it with state-of-the-art electron microscopy (EM) segmentation methods from the ISBI EM segmentation challenge. We also show the segmentation results on two different tasks including cell membrane and cell body segmentation and a statistical analysis of cell morphology.

54. **U-Net: Convolutional Networks for Biomedical Image Segmentation** [106]

    Olaf Ronneberger, Philipp Fischer, Thomas Brox

    MICCAI 2015

There is large consent that successful training of deep networks requires many thousand annotated training samples. In this paper, we present a network and training strategy that relies on the strong use of data augmentation to use the available annotated samples more efficiently. The architecture consists of a contracting path to capture context and a symmetric expanding path that enables precise localization. We show that such a network can be trained end-to-end from very few images and outperforms the prior best method (a sliding-window convolutional network) on the ISBI challenge for segmentation of neuronal structures in electron microscopic stacks. Using the same network trained on transmitted light microscopy images (phase contrast and DIC) we won the ISBI cell tracking challenge 2015 in these categories by a large margin. Moreover, the network is fast. Segmentation of a 512x512 image takes less than a second on a recent GPU. The full implementation (based on Caffe) and the trained networks are available at http://lmb.informatik.uni-freiburg.de/people/ronneber/u-net.

55. **VoxResNet: Deep Voxelwise Residual Networks for Volumetric Brain Segmentation** [21]

    Hao Chen, Qi Dou, Lequan Yu and Pheng-Ann Heng

    arXiv 2016

    Recently deep residual learning with residual units for training very deep neural networks advanced the state-of-the-art performance on 2D image recognition tasks, e.g., object detection and segmentation. However, how to fully leverage contextual representations for recognition tasks from volumetric data has not been well studied, especially in the field of medical image computing, where a majority of image modalities are in volumetric format. In this paper we explore the deep residual learning on the task of volumetric brain segmentation. There are at least two main contributions in our work. First, we propose a deep voxelwise residual network, referred as VoxResNet, which borrows the spirit of deep residual learning in 2D image recognition tasks, and is extended into a 3D variant for handling volumetric data. Second, an auto-context version of VoxResNet is proposed by seamlessly integrating the low-level image appearance features, implicit shape information and high-level context together for further improving the volumetric segmentation performance. Extensive experiments on the challenging benchmark of brain segmentation from magnetic resonance (MR) images corroborated the ef- ficacy of our proposed method in dealing with volumetric data. We believe this work unravels the potential of 3D deep learning to advance the recognition performance on volumetric image segmentation.

56. **V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation** [85]

    Fausto Milletari, Nassir Navab

    2016 IEEE Fourth International Conference on 3D Vision

    Convolutional Neural Networks (CNNs) have been recently employed to solve problems from both the computer vision and medical image analysis fields. Despite their popularity, most approaches are only able to process 2D images while most medical data used in clinical practice consists of 3D volumes. In this work we propose an approach to 3D image segmentation based on a volumetric, fully convolutional, neural network. The proposed CNN is trained end-to-end on MRI volumes depicting prostate, and learns to predict segmentation for the whole volume at once. We introduce a novel objective function, that authors optimize during training, based on Dice coefficient. In this way we can deal with situations where there is a strong imbalance between the number of foreground and background voxels. To cope with the limited number of annotated volumes available for training, we augment the data applying random non-linear transformations and histogram matching. We show in our experimental evaluation that our approach achieves good performances on challenging test data while requiring only a fraction of the processing time needed by other previous methods.

57. **3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation** [85]

    Çiçek, Özgün, Ahmed Abdulkadir, Soeren S. Lienkamp, Thomas Brox, Olaf Ronneberger

    arXiv 2016

    This paper introduces a network for volumetric segmentation that learns from sparsely annotated volumetric images. We outline two attractive use cases of this method: (1) In a semi-automated setup,

the user annotates some slices in the volume to be segmented. The network learns from these sparse annotations and provides a dense 3D segmentation. (2) In a fully-automated setup, we assume that a representative, sparsely annotated training set exists. Trained on this data set, the network densely segments new volumetric images. The proposed network extends the previous u-net architecture from Ronneberger et al. by replacing all 2D operations with their 3D counterparts. The implementation performs on-the-fly elastic deformations for efficient data augmentation during training. It is trained end-to-end from scratch, i.e., no pre-trained network is required. We test the performance of the proposed method on a complex, highly variable 3D structure, the Xenopus kidney, and achieve good results for both use cases.

58. **Deep Learning with Dynamic Computation Graphs** [74]

    Moshe Looks, Marcello Herreshoff, DeLesley Hutchins, Peter Norvig

    ICLR 2017

    Neural networks that compute over graph structures are a natural fit for problems in a variety of domains, including natural language (parse trees) and cheminformatics (molecular graphs). However, since the computation graph has a different shape and size for every input, such networks do not directly support batched training or inference. They are also difficult to implement in popular deep learning libraries, which are based on static data-flow graphs. We introduce a technique called dynamic batching, which not only batches together operations between different input graphs of dissimilar shape, but also between different nodes within a single input graph. The technique allows us to create static graphs, using popular libraries, that emulate dynamic computation graphs of arbitrary shape and size. We further present a high-level library of compositional blocks that simplifies the creation of dynamic graph models. Using the library, we demonstrate concise and batch-wise parallel implementations for a variety of models from the literature.

59. **Hybrid neural network for classification of graph structured data** [59]

    R. B. Gnana Jothi, S. M. Meena Rani

    International Journal of Machine Learning and Cybernetics 2014

    Multilayer perceptron (MLP) with recurrent architecture is proposed for stabilizing the state vector, which represents the characteristics of the nodes in a graph, to classify the graph structured data. M number of input and output networks are constructed for the M node undirected graphs for classifying graph structured data. Output of every input network represents the characteristics of the node as a state vector. The output of each input MLP is also taken as input for the same network along with output of neighboring node's MLP. Both the input and output networks are trained by backpropagation. The proposed approach is implemented on the standard benchmark classification problems namely mutagenesis problem, subgraph matching problem and clique problem. Simulation results show that best accuracy in classification is obtained with minimum computational complexity.

60. **FANNG: Fast Approximate Nearest Neighbour Graphs** [47]

    Harwood, Ben and Drummond, Tom CVPR 2016

    We present a new method for approximate nearest neighbour search on large datasets of high dimensional feature vectors, such as SIFT or GIST descriptors. Our approach constructs a directed graph that can be efficiently explored for nearest neighbour queries. Each vertex in this graph represents a feature vector from the dataset being searched. The directed edges are computed by exploiting the fact that, for these datasets, the intrinsic dimensionality of the local manifold-like structure formed by the elements of the dataset is significantly lower than the embedding space. We also provide an efficient search algorithm that uses this graph to rapidly find the nearest neighbour to a query with high probability. We show how the method can be adapted to give a strong guarantee of 100% recall where the query is within a threshold distance of its nearest neighbour. We demonstrate that our method is significantly more efficient than existing state of the art methods. In particular, our GPU implementation can deliver 90% recall for queries on a data set of 1 million SIFT descriptors at a rate of over 1.2 million queries per second on a Titan X. Finally we also demonstrate how our method scales to datasets of 5M and 20M entries.

61. **Graph Convolutional Encoders for Syntax-aware Neural Machine Translation** [6]

Joost Bastings, Ivan Titov, Wilker Aziz, Diego Marcheggiani, Khalil Sima'an

arXiv 2017

This paper presents a simple and effective approach to incorporating syntactic structure into neural attention-based encoderdecoder models for machine translation. We rely on graph-convolutional networks (GCNs), a recent class of neural networks developed for modeling graph-structured data. Our GCNs use predicted syntactic dependency trees of source sentences to produce representations of words (i.e. hidden states of the encoder) that are sensitive to their syntactic neighborhoods. GCNs take word representations as input and produce word representations as output, so they can easily be incorporated as layers into standard encoders (e.g., on top of bidirectional RNNs or convolutional neural networks). We evaluate their effectiveness with English-German and English-Czech translation experiments for different types of encoders and observe substantial improvements over their syntax-agnostic versions in all the considered setups.

62. **Encoding Sentences with Graph Convolutional Networks for Semantic Role Labeling** [83]

Diego Marcheggiani, Ivan Titov

arXiv 2017

Semantic role labeling (SRL) is the task of identifying the predicate-argument structure of a sentence. It is typically regarded as an important step in the standard natural language processing pipeline, providing information to downstream tasks such as information extraction and question answering. As the semantic representations are closely related to syntactic ones, we exploit syntactic information in our model. We propose a version of graph convolutional networks (GCNs), a recent class of multilayer neural networks operating on graphs, suited to modeling syntactic dependency graphs. GCNs over syntactic dependency trees are used as sentence encoders, producing latent feature representations of words in a sentence and capturing information relevant to predicting the semantic representations. We observe that GCN layers are complementary to LSTM ones: when we stack both GCN and LSTM layers, we obtain a substantial improvement over an already state-of-the-art LSTM SRL model, resulting in the best reported scores on the standard benchmark (CoNLL-2009) both for Chinese and English.

63. **Neural Graph Machines: Learning Neural Networks Using Graphs** [13]

Thang D. Bui, Sujith Ravi, Vivek Ramavajjala

arXiv 2017

Label propagation is a powerful and flexible semi-supervised learning technique on graphs. Neural networks, on the other hand, have proven track records in many supervised learning tasks. In this work, we propose a training framework with a graph-regularised objective, namely "Neural Graph Machines", that can combine the power of neural networks and label propagation. This work generalises previous literature on graph-augmented training of neural networks, enabling it to be applied to multiple neural architectures (Feed-forward NNs, CNNs and LSTM RNNs) and a wide range of graphs. The new objective allows the neural networks to harness both labeled and unlabeled data by: (a) allowing the network to train using labeled data as in the supervised setting, (b) biasing the network to learn similar hidden representations for neighboring nodes on a graph, in the same vein as label propagation. Such architectures with the proposed objective can be trained efficiently using stochastic gradient descent and scaled to large graphs, with a runtime that is linear in the number of edges. The proposed joint training approach convincingly outperforms many existing methods on a wide range of tasks (multi-label classification on social graphs, news categorization, document classification and semantic intent classification), with multiple forms of graph inputs (including graphs with and without node-level features) and using different types of neural networks.

64. **Bootstrapping Graph Convolutional Neural Networks for Autism Spectrum Disorder Classification** [2]

Rushil Anirudh, Jayaraman J. Thiagarajan

arXiv 2017

Using predictive models to identify patterns that can act as biomarkers for different neuropathoglogical conditions is becoming highly prevalent. In this paper, we consider the problem of Autism Spectrum Disorder (ASD) classification. While non-invasive imaging measurements, such as the rest state fMRI, are typically used in this problem, it can be beneficial to incorporate a wide variety of non-imaging features, including personal and socio-cultural traits, into predictive modeling. We propose to employ a graph-based approach for combining both types of feature, where a contextual graph encodes the traits of a larger population while the brain activity patterns are defined as a multivariate function at the nodes of the graph. Since the underlying graph dictates the performance of the resulting predictive models, we explore the use of different graph construction strategies. Furthermore, we develop a bootstrapped version of graph convolutional neural networks (G-CNNs) that utilizes an ensemble of weakly trained G-CNNs to avoid overfitting and also reduce the sensitivity of the models on the choice of graph construction. We demonstrate its effectiveness on the Autism Brain Imaging Data Exchange (ABIDE) dataset and show that the proposed approach outperforms state-of-the-art approaches for this problem.

65. **A Generalization of Convolutional Neural Networks to Graph-Structured Data** [48]

    Yotam Hechtlinger, Purvasha Chakravarti, Jining Qin

    arXiv 2017

    This paper introduces a generalization of Convolutional Neural Networks (CNNs) from low-dimensional grid data, such as images, to graph-structured data. We propose a novel spatial convolution utilizing a random walk to uncover the relations within the input, analogous to the way the standard convolution uses the spatial neighborhood of a pixel on the grid. The convolution has an intuitive interpretation, is efficient and scalable and can also be used on data with varying graph structure. Furthermore, this generalization can be applied to many standard regression or classification problems, by learning the the underlying graph. We empirically demonstrate the performance of the proposed CNN on MNIST, and challenge the state-of-the-art on Merck molecular activity data set.

66. **Graph based manifold regularized deep neural networks for automatic speech recognition** [122]

    Vikrant Singh Tomar, Richard C. Rose

    arXiv 2016

    Deep neural networks (DNNs) have been successfully applied to a wide variety of acoustic modeling tasks in recent years. These include the applications of DNNs either in a discriminative feature extraction or in a hybrid acoustic modeling scenario. Despite the rapid progress in this area, a number of challenges remain in training DNNs. This paper presents an effective way of training DNNs using a manifold learning based regularization framework. In this framework, the parameters of the network are optimized to preserve underlying manifold based relationships between speech feature vectors while minimizing a measure of loss between network outputs and targets. This is achieved by incorporating manifold based locality constraints in the objective criterion of DNNs. Empirical evidence is provided to demonstrate that training a network with manifold constraints preserves structural compactness in the hidden layers of the network. Manifold regularization is applied to train bottleneck DNNs for feature extraction in hidden Markov model (HMM) based speech recognition. The experiments in this work are conducted on the Aurora- 2 spoken digits and the Aurora-4 read news large vocabulary continuous speech recognition tasks. The performance is measured in terms of word error rate (WER) on these tasks. It is shown that the manifold regularized DNNs result in up to 37% reduction in WER relative to standard DNNs.

67. **Wavelets on Graphs via Deep Learning** [107]

    Raif M. Rustamov, Leonidas Guibas

    NIPS 2013

    An increasing number of applications require processing of signals defined on weighted graphs. While wavelets provide a flexible tool for signal processing in the classical setting of regular domains, the

existing graph wavelet constructions are less flexible – they are guided solely by the structure of the underlying graph and do not take directly into consideration the particular class of signals to be processed. This paper introduces a machine learning framework for constructing graph wavelets that can sparsely represent a given class of signals. Our construction uses the lifting scheme, and is based on the observation that the recurrent nature of the lifting scheme gives rise to a structure resembling a deep auto-encoder network. Particular properties that the resulting wavelets must satisfy determine the training objective and the structure of the involved neural networks. The training is unsupervised, and is conducted similarly to the greedy pre-training of a stack of auto-encoders. After training is completed, we obtain a linear wavelet transform that can be applied to any graph signal in time and memory linear in the size of the graph. Improved sparsity of our wavelet transform for the test signals is confirmed via experiments both on synthetic and real data.

68. **HD-CNN: Hierarchical Deep Convolutional Neural Network for Image Classification** [138]

    Zhicheng Yan, Vignesh Jagadeesh, Dennis Decoste, Wei Di, Robinson Piramuthu

    ICCV 2015

    Existing deep convolutional neural network (CNN) architectures are trained as N-way classifiers to distinguish between N output classes. This work builds on the intuition that not all classes are equally difficult to distinguish from a true class label. Towards this end, we introduce hierarchical branching CNNs, named as Hierarchical Deep CNN (HD-CNN), wherein classes that can be easily distinguished are classified in the higher layer coarse category CNN, while the most difficult classifications are done on lower layer fine category CNN. We propose utilizing a multinomial logistic loss and a novel temporal sparsity penalty for HD-CNN training. Together they ensure each branching component deals with a subset of categories confusing to each other. This new network architecture adopts coarseto-fine classification strategy and module design principle. The proposed model achieves superior performance over standard models. We demonstrate state-of-the-art results on CIFAR100 benchmark.

69. **The More You Know: Using Knowledge Graphs for Image Classification** [84]

    Kenneth Marino, Ruslan Salakhutdinov, Abhinav Gupta

    CVPR 2017

    One characteristic that sets humans apart from modern learning-based computer vision algorithms is the ability to acquire knowledge about the world and use that knowledge to reason about the visual world. Humans can learn about the characteristics of objects and the relationships that occur between them to learn a large variety of visual concepts, often with few examples. This paper investigates the use of structured prior knowledge in the form of knowledge graphs and shows that using this knowledge improves performance on image classification. We build on recent work on end-to-end learning on graphs, introducing the Graph Search Neural Network as a way of efficiently incorporating large knowledge graphs into a vision classification pipeline. We show in a number of experiments that our method outperforms standard neural network baselines for multi-label classification.

70. **Pancreas Segmentation in MRI Using Graph-Based Decision Fusion on Convolutional Neural Networks** [15]

    Jinzheng Cai, Le Lu, Zizhao Zhang, Fuyong Xing, Lin Yang, and Qian Yin

    2016 International Conference on Medical Image Computing and Computer-Assisted Intervention

    Automated pancreas segmentation in medical images is a prerequisite for many clinical applications, such as diabetes inspection, pancreatic cancer diagnosis, and surgical planing. In this paper, we formulate pancreas segmentation in magnetic resonance imaging (MRI) scans as a graph based decision fusion process combined with deep convolutional neural networks (CNN). Our approach conducts pancreatic detection and boundary segmentation with two types of CNN models respectively: (1) the tissue detection step to differentiate pancreas and non-pancreas tissue with spatial intensity context; (2) the boundary detection step to allocate the semantic boundaries of pancreas. Both detection results of the two networks are fused together as the initialization of a conditional random field (CRF) framework to obtain the final segmentation output. Our approach achieves the mean dice similarity coefficient

(DSC) 76.1 % with the standard deviation of 8.7 % in a dataset containing 78 abdominal MRI scans. The proposed algorithm achieves the best results compared with other state of the arts.

71. **Dynamic Graph Convolutional Networks** [82]

    Franco Manessi, Alessandro Rozza, and Mario Manzo

    arXiv 2017

    Many different classification tasks need to manage structured data, which are usually modeled as graphs. Moreover, these graphs can be dynamic, meaning that the vertices/edges of each graph may change during time. Our goal is to jointly exploit structured data and temporal information through the use of a neural network model. To the best of our knowledge, this task has not been addressed using these kind of architectures. For this reason, we propose two novel approaches, which combine Long Short-Term Memory networks and Graph Convolutional Networks to learn long short-term dependencies together with graph structure. The quality of our methods is confirmed by the promising results achieved.

72. **Training Deep Neural Networks via Optimization Over Graphs** [146]

    Guoqiang Zhang, W. Bastiaan Kleijn

    arXiv 2017

    In this work, authors propose to train a deep neural network by distributed optimization over a graph. Two nonlinear functions are considered: the rectified linear unit (ReLU) and a linear unit with both lower and upper cutoffs (DCutLU). The problem reformulation over a graph is realized by explicitly representing ReLU or DCutLU using a set of slack variables. We then apply the alternating direction method of multipliers (ADMM) to update the weights of the network layerwise by solving subproblems of the reformulated problem. Empirical results suggest that by proper parameter selection, the ADMM-based method converges considerably faster than gradient descent method.

73. **Graph Neural Networks and Boolean Satisfiability** [14]

    Benedikt Bünz, Matthew Lamm

    arXiv 2017

    In this paper, authors explore whether or not deep neural architectures can learn to classify Boolean satisfiability (SAT). We devote considerable time to discussing the theoretical properties of SAT. Then, we define a graph representation for Boolean formulas in conjunctive normal form, and train neural classifiers over general graph structures called Graph Neural Networks, or GNNs, to recognize features of satisfiability. To the best of our knowledge this has never been tried before. Our preliminary findings are potentially profound. In a weakly-supervised setting, that is, without problem specific feature engineering, Graph Neural Networks can learn features of satisfiability.

74. **Geometric Matrix Completion with Recurrent Multi-Graph Neural Networks** [87]

    Federico Monti, Michael M. Bronstein, Xavier Bresson

    arXiv 2017

    Matrix completion models are among the most common formulations of recommender systems. Recent works have showed a boost of performance of these techniques when introducing the pairwise relationships between users/items in the form of graphs, and imposing smoothness priors on these graphs. However, such techniques do not fully exploit the local stationarity structures of user/item graphs, and the number of parameters to learn is linear w.r.t. the number of users and items. We propose a novel approach to overcome these limitations by using geometric deep learning on graphs. Our matrix completion architecture combines graph convolutional neural networks and recurrent neural networks to learn meaningful statistical graph-structured patterns and the non-linear diffusion process that generates the known ratings. This neural network system requires a constant number of parameters independent of the matrix size. We apply our method on both synthetic and real datasets, showing that it outperforms state-of-the-art techniques.

75. **Knowledge Representation in Graphs using Convolutional Neural Networks** [128]

    Armando Vieira

    arXiv 2016

    Knowledge Graphs (KG) constitute a flexible representation of complex relationships between entities particularly useful for biomedical data. These KG, however, are very sparse with many missing edges (facts) and the visualisation of the mesh of interactions nontrivial. Here we apply a compositional model to embed nodes and relationships into a vectorised semantic space to perform graph completion. A visualisation tool based on Convolutional Neural Networks and Self-Organised Maps (SOM) is proposed to extract high-level insights from the KG. We apply this technique to a subset of CTD, containing interactions of compounds with human genes / proteins and show that the performance is comparable to the one obtained by structural models.

76. **Regular graphs maximize the variability of random neural networks** [131]

    Gilles Wainrib, Mathieu Galtier

    Physical Review E 2015

    In this work authors study the dynamics of systems composed of numerous interacting elements interconnected through a random weighted directed graph, such as models of random neural networks. We develop an original theoretical approach based on a combination of a classical mean-field theory originally developed in the context of dynamical spin-glass models, and the heterogeneous mean-field theory developed to study epidemic propagation on graphs. Our main result is that, surprisingly, increasing the variance of the in-degree distribution does not result in a more variable dynamical behavior, but on the contrary that the most variable behaviors are obtained in the regular graph setting. We further study how the dynamical complexity of the attractors is influenced by the statistical properties of the in-degree distribution.

77. **GRAM: Graph-based Attention Model for Healthcare Representation Learning** [28]

    Edward Choi, Mohammad Taha Bahadori, Le Song, Walter F. Stewart, Jimeng Sun

    arXiv 2016

    Deep learning methods exhibit promising performance for predictive modeling in healthcare, but two important challenges remain: -Data insufficiency:Often in healthcare predictive modeling, the sample size is insufficient for deep learning methods to achieve satisfactory results. -Interpretation:The representations learned by deep learning methods should align with medical knowledge. To address these challenges, we propose a GRaph-based Attention Model, GRAM that supplements electronic health records (EHR) with hierarchical information inherent to medical ontologies. Based on the data volume and the ontology structure, GRAM represents a medical concept as a combination of its ancestors in the ontology via an attention mechanism. We compared predictive performance (i.e. accuracy, data needs, interpretability) of GRAM to various methods including the recurrent neural network (RNN) in two sequential diagnoses prediction tasks and one heart failure prediction task. Compared to the basic RNN, GRAM achieved 10% higher accuracy for predicting diseases rarely observed in the training data and 3% improved area under the ROC curve for predicting heart failure using an order of magnitude less training data. Additionally, unlike other methods, the medical concept representations learned by GRAM are well aligned with the medical ontology. Finally, GRAM exhibits intuitive attention behaviors by adaptively generalizing to higher level concepts when facing data insufficiency at the lower level concepts.

2 Deep learning and conditional random fields

1. **Recurrent conditional random field for language understanding**[143]

    Yao, Kaisheng and Peng, Baolin and Zweig, Geoffrey and Yu, Dong and Li, Xiaolong and Gao, Feng Acoustics, Speech and Signal Processing (ICASSP) 2014

Recurrent neural networks (RNNs) have recently produced record setting performance in language modeling and word-labeling tasks. In the word-labeling task, the RNN is used analogously to the more traditional conditional random field (CRF) to assign a label to each word in an input sequence, and has been shown to significantly outperform CRFs. In contrast to CRFs, RNNs operate in an online fashion to assign labels as soon as a word is seen, rather than after seeing the whole word sequence. In this paper, we show that the performance of an RNN tagger can be significantly improved by incorporating elements of the CRF model; specifically, the explicit modeling of output-label dependencies with transition features, its global sequence-level objective function, and offline decoding. We term the resulting model a "recurrent conditional random field" and demonstrate its effectiveness on the ATIS travel domain dataset and a variety of web-search language understanding datasets.

2. **End-to-end sequence labeling via bi-directional lstm-cnns-crf**[79]

   Ma, Xuezhe and Hovy, Eduard arXiv 2016

   State-of-the-art sequence labeling systems traditionally require large amounts of task-specific knowledge in the form of hand-crafted features and data pre-processing. In this paper, we introduce a novel neutral network architecture that benefits from both word- and character-level representations automatically, by using combination of bidirectional LSTM, CNN and CRF. Our system is truly end-to-end, requiring no feature engineering or data pre-processing, thus making it applicable to a wide range of sequence labeling tasks. We evaluate our system on two data sets for two sequence labeling tasks — Penn Treebank WSJ corpus for part-of-speech (POS) tagging and CoNLL 2003 corpus for named entity recognition (NER). We obtain state-of-the-art performance on both the two data — 97.55% accuracy for POS tagging and 91.21% F1 for NER.

3. **Bidirectional LSTM-CRF models for sequence tagging**[52]

   Huang, Zhiheng and Xu, Wei and Yu, Kai arXiv 2015

   In this paper, we propose a variety of Long Short-Term Memory (LSTM) based models for sequence tagging. These models include LSTM networks, bidirectional LSTM (BI-LSTM) networks, LSTM with a Conditional Random Field (CRF) layer (LSTM-CRF) and bidirectional LSTM with a CRF layer (BI-LSTM-CRF). Our work is the first to apply a bidirectional LSTM CRF (denoted as BI-LSTM-CRF) model to NLP benchmark sequence tagging data sets. We show that the BI-LSTM-CRF model can efficiently use both past and future input features thanks to a bidirectional LSTM component. It can also use sentence level tag information thanks to a CRF layer. The BI-LSTM-CRF model can produce state of the art (or close to) accuracy on POS, chunking and NER data sets. In addition, it is robust and has less dependence on word embedding as compared to previous observations.

4. **Convolutional neural network based triangular CRF for joint intent detection and slot filling**[137]

   Xu, Puyang and Sarikaya, Ruhi Automatic Speech Recognition and Understanding (ASRU), 2013

   Abstract: We describe a joint model for intent detection and slot filling based on convolutional neural networks (CNN). The proposed architecture can be perceived as a neural network (NN) version of the triangular CRF model (TriCRF), in which the intent label and the slot sequence are modeled jointly and their dependencies are exploited. Our slot filling component is a globally normalized CRF style model, as opposed to left-to-right models in recent NN based slot taggers. Its features are automatically extracted through CNN layers and shared by the intent model. We show that our slot model component generates state-of-the-art results, outperforming CRF significantly. Our joint model outperforms the standard TriCRF by 1% absolute for both intent and slot. On a number of other domains, our joint model achieves 0.7-1%, and 0.9-2.1% absolute gains over the independent modeling approach for intent and slot respectively.

5. **Semantic image segmentation with deep convolutional nets and fully connected crfs**[23]

   Chen, Liang-Chieh and Papandreou, George and Kokkinos, Iasonas and Murphy, Kevin and Yuille, Alan L TPAMI 2014

   Deep Convolutional Neural Networks (DCNNs) have recently shown state of the art performance in high level vision tasks, such as image classification and object detection. This work brings together methods

from DCNNs and probabilistic graphical models for addressing the task of pixel-level classification (also called "semantic image segmentation"). We show that responses at the final layer of DCNNs are not sufficiently localized for accurate object segmentation. This is due to the very invariance properties that make DCNNs good for high level tasks. We overcome this poor localization property of deep networks by combining the responses at the final DCNN layer with a fully connected Conditional Random Field (CRF). Qualitatively, our "DeepLab" system is able to localize segment boundaries at a level of accuracy which is beyond previous methods. Quantitatively, our method sets the new state-of-art at the PASCAL VOC-2012 semantic image segmentation task, reaching 71.6% IOU accuracy in the test set. We show how these results can be obtained efficiently: Careful network re-purposing and a novel application of the 'hole' algorithm from the wavelet community allow dense computation of neural net responses at 8 frames per second on a modern GPU.

6. **Learning Sparse High Dimensional Filters: Image Filtering, Dense CRFs and Bilateral Neural Networks**[55]

   Jampani, Varun and Kiefel, Martin and Gehler, Peter V. CVPR 2016

   Bilateral filters have wide spread use due to their edge-preserving properties. The common use case is to manually choose a parametric filter type, usually a Gaussian filter. In this paper, we will generalize the parametrization and in particular derive a gradient descent algorithm so the filter parameters can be learned from data. This derivation allows to learn high dimensional linear filters that operate in sparsely populated feature spaces. We build on the permutohedral lattice construction for efficient filtering. The ability to learn more general forms of high-dimensional filters can be used in several diverse applications. First, we demonstrate the use in applications where single filter applications are desired for runtime reasons. Further, we show how this algorithm can be used to learn the pairwise potentials in densely connected conditional random fields and apply these to different image segmentation tasks. Finally, we introduce layers of bilateral filters in CNN and propose bilateral neural networks for the use of high-dimensional sparse data. This view provides new ways to encode model structure into network architectures. A diverse set of experiments empirically validates the usage of general forms of filters.

7. **A deep learning model integrating FCNNs and CRFs for brain tumor segmentation** [148]

   Xiaomei Zhao, Yihong Wu, Guidong Song, Zhenye Li, Yazhuo Zhang, Yong Fan

   arXiv 2017

   Accurate and reliable brain tumor segmentation is a critical component in cancer diagnosis, treatment planning, and treatment outcome evaluation. Build upon successful deep learning techniques, we propose a novel brain tumor segmentation method by integrating fully convolutional neural networks (FCNNs) and Conditional Random Fields (CRFs) in a unified framework to obtain segmentation results with appearance and spatial consistency. We train the deep learning based segmentation model using image patches and image slices in following steps: 1) training FCNNs using image patches; 2) training CRF-RNN using image slices of axial view with parameters of FCNNs fixed; and 3) fine-tuning the whole network using image slices. Our method could segment brain images slice-by-slice, much faster than those image patch based tumor segmentation methods. We have evaluated our method based on imaging data provided by the Multimodal Brain Tumor Image Segmentation Challenge (BRATS) 2013 and the BRATS 2016. The experimental results have demonstrated that our method could build a segmentation model with Flair, T1c, and T2 scans and achieve competitive performance as those built with Flair, T1, T1c, and T2 scans.

8. **End-to-End Training of Hybrid CNN-CRF Models for Stereo** [64]

   Patrick Knöbelreiter, Christian Reinbacher, Alexander Shekhovtsov, Thomas Pock

   arXiv 2016

   Authors propose a novel method for stereo estimation, combining advantages of convolutional neural networks (CNNs) and optimization-based approaches. The optimization, posed as a conditional random field (CRF), takes local matching costs and consistency-enforcing (smoothness) costs as inputs, both estimated by CNN blocks. To perform the inference in the CRF we use an approach based on linear programming relaxation with a fixed number of iterations. We address the challenging problem of

training this hybrid model end-to-end. We show that in the discriminative formulation (structured support vector machine) the training is practically feasible. The trained hybrid model with shallow CNNs is comparable to state-of-the-art deep models in both time and performance. The optimization part efficiently replaces sophisticated and not jointly trainable (but commonly applied) post-processing steps by a trainable, well-understood model.

9. **Conditional Random Fields as Recurrent Neural Networks** [149]

   Shuai Zheng, Sadeep Jayasumana, Bernardino Romera-Paredes, Vibhav Vineet, Zhizhong Su, Dalong Du, Chang Huang, Philip H. S. Torr

   ICCV 2015

   Pixel-level labelling tasks, such as semantic segmentation, play a central role in image understanding. Recent approaches have attempted to harness the capabilities of deep learning techniques for image recognition to tackle pixellevel labelling tasks. One central issue in this methodology is the limited capacity of deep learning techniques to delineate visual objects. To solve this problem, we introduce a new form of convolutional neural network that combines the strengths of Convolutional Neural Networks (CNNs) and Conditional Random Fields (CRFs)-based probabilistic graphical modelling. To this end, we formulate Conditional Random Fields with Gaussian pairwise potentials and mean-field approximate inference as Recurrent Neural Networks. This network, called CRF-RNN, is then plugged in as a part of a CNN to obtain a deep network that has desirable properties of both CNNs and CRFs. Importantly, our system fully integrates CRF modelling with CNNs, making it possible to train the whole deep network end-to-end with the usual back-propagation algorithm, avoiding offline post-processing methods for object delineation. We apply the proposed method to the problem of semantic image segmentation, obtaining top results on the challenging Pascal VOC 2012 segmentation benchmark.

10. **Higher Order Conditional Random Fields in Deep Neural Networks** [3]

    Anurag Arnab, Sadeep Jayasumana, Shuai Zheng, and Philip H.S. Torr

    European Conference on Computer Vision

    We address the problem of semantic segmentation using deep learning. Most segmentation systems include a Conditional Random Field (CRF) to produce a structured output that is consistent with the image's visual features. Recent deep learning approaches have incorporated CRFs into Convolutional Neural Networks (CNNs), with some even training the CRF end-to-end with the rest of the network. However, these approaches have not employed higher order potentials, which have previously been shown to significantly improve segmentation performance. In this paper, we demonstrate that two types of higher order potential, based on object detections and superpixels, can be included in a CRF embedded within a deep network. We design these higher order potentials to allow inference with the differentiable mean field algorithm. As a result, all the parameters of our richer CRF model can be learned end-to-end with our pixelwise CNN classifier. We achieve state-of-the-art segmentation performance on the PASCAL VOC benchmark with these trainable higher order potentials.

11. **Fast, Exact and Multi-scale Inference for Semantic Image Segmentation with Deep Gaussian CRFs** [18]

    Siddhartha Chandra, authorIasonas Kokkinos ECCV 2016

    In this work authors propose a structured prediction technique that combines the virtues of Gaussian Conditional Random Fields (G-CRF) with Deep Learning: (a) our structured prediction task has a unique global optimum that is obtained exactly from the solution of a linear system (b) the gradients of our model parameters are analytically computed using closed form expressions, in contrast to the memory-demanding contemporary deep structured prediction approaches [1, 2] that rely on back-propagation-through-time, (c) our pairwise terms do not have to be simple hand-crafted expressions, as in the line of works building on the DenseCRF [1, 3], but can rather be 'discovered' from data through deep architectures, and (d) out system can trained in an end-to-end manner. Building on standard tools from numerical analysis we develop very efficient algorithms for inference and learning, as well as a customized technique adapted to the semantic segmentation task. This efficiency allows us to

explore more sophisticated architectures for structured prediction in deep learning: we introduce multi-resolution architectures to couple information across scales in a joint optimization framework, yielding systematic improvements. We demonstrate the utility of our approach on the challenging VOC PASCAL 2012 image segmentation benchmark, showing substantial improvements over strong baselines. We make all of our code and experiments available at https://github.com/siddharthachandra/gcrf.

12. **Geometric deep learning: going beyond Euclidean data** [10]

    Michael M. Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, Pierre Vandergheynst

    arXiv 2016

    Many signal processing problems involve data whose underlying structure is non-Euclidean, but may be modeled as a manifold or (combinatorial) graph. For instance, in social networks, the characteristics of users can be modeled as signals on the vertices of the social graph. Sensor networks are graph models of distributed interconnected sensors, whose readings are modelled as time-dependent signals on the vertices. In genetics, gene expression data are modeled as signals defined on the regulatory network. In neuroscience, graph models are used to represent anatomical and functional structures of the brain. Modeling data given as points in a high-dimensional Euclidean space using nearest neighbor graphs is an increasingly popular trend in data science, allowing practitioners access to the intrinsic structure of the data. In computer graphics and vision, 3D objects are modeled as Riemannian manifolds (surfaces) endowed with properties such as color texture. Even more complex examples include networks of operators, e.g., functional correspondences or difference operators in a collection of 3D shapes, or orientations of overlapping cameras in multi-view vision ("structure from motion") problems. The complexity of geometric data and the availability of very large datasets (in the case of social networks, on the scale of billions) suggest the use of machine learning techniques. In particular, deep learning has recently proven to be a powerful tool for problems with large datasets with underlying Euclidean structure. The purpose of this paper is to overview the problems arising in relation to geometric deep learning and present solutions existing today for this class of problems, as well as key difficulties and future research directions.

13. **DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs** [24]

    Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, Alan L. Yuille arXiv 2016

    In this work authors address the task of semantic image segmentation with Deep Learning and make three main contributions that are experimentally shown to have substantial practical merit. First, we highlight convolution with upsampled filters, or 'atrous convolution', as a powerful tool in dense prediction tasks. Atrous convolution allows us to explicitly control the resolution at which feature responses are computed within Deep Convolutional Neural Networks. It also allows us to effectively enlarge the field of view of filters to incorporate larger context without increasing the number of parameters or the amount of computation. Second, we propose atrous spatial pyramid pooling (ASPP) to robustly segment objects at multiple scales. ASPP probes an incoming convolutional feature layer with filters at multiple sampling rates and effective fields-of-views, thus capturing objects as well as image context at multiple scales. Third, we improve the localization of object boundaries by combining methods from DCNNs and probabilistic graphical models. The commonly deployed combination of max-pooling and downsampling in DCNNs achieves invariance but has a toll on localization accuracy. We overcome this by combining the responses at the final DCNN layer with a fully connected Conditional Random Field (CRF), which is shown both qualitatively and quantitatively to improve localization performance. Our proposed "DeepLab" system sets the new state-of-art at the PASCAL VOC-2012 semantic image segmentation task, reaching 79.7% mIOU in the test set, and advances the results on three other datasets: PASCAL-Context, PASCAL-Person-Part, and Cityscapes. All of our code is made publicly available online.

14. **Deep Gaussian Conditional Random Field Network: A Model-based Deep Network for Discriminative Denoising** [126]

    Raviteja Vemulapalli, Oncel Tuzel, Ming-Yu Liu

CVPR 2016

This paper proposes a novel end-to-end trainable deep network architecture for image denoising based on a Gaussian Conditional Random Field (GCRF) model. In contrast to the existing discriminative denoising methods that train a separate model for each individual noise level, the proposed deep network explicitly models the input noise variance and hence is capable of handling a range of noise levels. Our deep network, which we refer to as deep GCRF network, consists of two sub-networks: (i) a parameter generation network that generates the pairwise potential parameters based on the noisy input image, and (ii) an inference network whose layers perform the computations involved in an iterative GCRF inference procedure. We train two deep GCRF networks (each network operates over a range of noise levels: one for low input noise levels and one for high input noise levels) discriminatively by maximizing the peak signal-to-noise ratio measure. Experiments on Berkeley segmentation and PASCALVOC datasets show that the proposed approach produces results on par with the stateof-the-art without training a separate network for each individual noise level.

15. **Brain Tumor Segmentation Using a Fully Convolutional Neural Network with Conditional Random Fields** [147]

    Xiaomei Zhao, Yihong Wu1, Guidong Song, Zhenye Li, Yong Fan, and Yazhuo Zhang International Workshop on Brainlesion 2016

    Deep learning techniques have been widely adopted for learning task-adaptive features in image segmentation applications, such as brain tumor segmentation. However, most of existing brain tumor segmentation methods based on deep learning are not able to ensure appearance and spatial consistency of segmentation results. In this study we propose a novel brain tumor segmentation method by integrating a Fully Convolutional Neural Network (FCNN) and Conditional Random Fields (CRF), rather than adopting CRF as a post-processing step of the FCNN. We trained our network in three stages based on image patches and slices respectively. We evaluated our method on BRATS 2013 dataset, obtaining the second position on its Challenge dataset and first position on its Leaderboard dataset. Compared with other top ranking methods, our method could achieve competitive performance with only three imaging modalities (Flair, T1c, T2), rather than four (Flair, T1, T1c, T2), which could reduce the cost of data acquisition and storage. Besides, our method could segment brain images slice-by-slice, much faster than the methods patch-by-patch. We also took part in BRATS 2016 and got satisfactory results. As the testing cases in BRATS 2016 are more challenging, we added a manual intervention post-processing system during our participation.

16. **Automatic Liver and Lesion Segmentation in CT Using Cascaded Fully Convolutional Neural Networks and 3D Conditional Random Fields** [29]

    Patrick Ferdinand Christ, Mohamed Ezzeldin A. Elshaer, Florian Ettlinger, Sunil Tatavarty, Marc Bickel, Patrick Bilic et. al.

    2016 International Conference on Medical Image Computing and Computer-Assisted Intervention

    Automatic segmentation of the liver and its lesion is an important step towards deriving quantitative biomarkers for accurate clinical diagnosis and computer-aided decision support systems. This paper presents a method to automatically segment liver and lesions in CT abdomen images using cascaded fully convolutional neural networks (CFCNs) and dense 3D conditional random fields (CRFs). We train and cascade two FCNs for a combined segmentation of the liver and its lesions. In the first step, we train a FCN to segment the liver as ROI input for a second FCN. The second FCN solely segments lesions from the predicted liver ROIs of step 1. We refine the segmentations of the CFCN using a dense 3D CRF that accounts for both spatial coherence and appearance. CFCN models were trained in a 2-fold cross-validation on the abdominal CT dataset 3DIRCAD comprising 15 hepatic tumor volumes. The results show that CFCN-based semantic liver and lesion segmentation achieves Dice scores over 94% for liver with computation times below 100 s per volume. Authors experimentally demonstrate the robustness of the proposed method as a decision support system with a high accuracy and speed for usage in daily clinical routine.

17. **Effective Semantic Pixel labelling with Convolutional Networks and Conditional Random Fields** [93]

Sakrapee Paisitkriangkrai, Jamie Sherrah, Pranam Janney and Anton Van-Den Hengel

CVPR 2015

Large amounts of available training data and increasing computing power have led to the recent success of deep convolutional neural networks (CNN) on a large number of applications. In this paper, we propose an effective semantic pixel labelling using CNN features, hand-crafted features and Conditional Random Fields (CRFs). Both CNN and hand-crafted features are applied to dense image patches to produce per-pixel class probabilities. The CRF infers a labelling that smooths regions while respecting the edges present in the imagery. The method is applied to the ISPRS 2D semantic labelling challenge dataset with competitive classification accuracy.

18. **Efficient Multi-Scale 3D CNN with Fully Connected CRF for Accurate Brain Lesion Segmentation** [60]

    Konstantinos Kamnitsas, Christian Ledig, Virginia F.J. Newcombe, Joanna P. Simpson, Andrew D. Kane, David K. Menon, Daniel Rueckert, Ben Glocker

    Medical Image Analysis 2017

    Authors propose a dual pathway, 11-layers deep, three-dimensional Convolutional Neural Network for the challenging task of brain lesion segmentation. The devised architecture is the result of an in-depth analysis of the limitations of current networks proposed for similar applications. To overcome the computational burden of processing 3D medical scans, we have devised an efficient and effective dense training scheme which joins the processing of adjacent image patches into one pass through the network while automatically adapting to the inherent class imbalance present in the data. Further, we analyze the development of deeper, thus more discriminative 3D CNNs. In order to incorporate both local and larger contextual information, we employ a dual pathway architecture that processes the input images at multiple scales simultaneously. For post-processing of the network's soft segmentation, we use a 3D fully connected Conditional Random Field which effectively removes false positives. Our pipeline is extensively evaluated on three challenging tasks of lesion segmentation in multi-channel MRI patient data with traumatic brain injuries, brain tumors, and ischemic stroke. We improve on the state-of-the-art for all three applications, with top ranking performance on the public benchmarks BRATS 2015 and ISLES 2015. Our method is computationally efficient, which allows its adoption in a variety of research and clinical settings. The source code of our implementation is made publicly available.

3 Deep learning and Markov random fields

1. **Combining markov random fields and convolutional neural networks for image synthesis**

    Li, Chuan and Wand, Michael

    CVPR 2016

    This paper studies a combination of generative Markov random field (MRF) models and discriminatively trained deep convolutional neural networks (dCNNs) for synthesizing 2D images. The generative MRF acts on higher-levels of a dCNN feature pyramid, controling the image layout at an abstract level. We apply the method to both photographic and non-photo-realistic (artwork) synthesis tasks. The MRF regularizer prevents over-excitation artifacts and reduces implausible feature mixtures common to previous dCNN inversion approaches, permitting synthezing photographic content with increased visual plausibility. Unlike standard MRF-based texture synthesis, the combined system can both match and adapt local features with considerable variability, yielding results far out of reach of classic generative MRF methods.

2. **Instance-level segmentation for autonomous driving with deep densely connected mrfs**

    Zhang, Ziyu and Fidler, Sanja and Urtasun, Raquel

    CVPR 2016

    Our aim is to provide a pixel-wise instance-level labeling of a monocular image in the context of autonomous driving. We build on recent work [Zhang et al., ICCV15] that trained a convolutional

neural net to predict instance labeling in local image patches, extracted exhaustively in a stride from an image. A simple Markov random field model using several heuristics was then proposed in [Zhang et al., ICCV15] to derive a globally consistent instance labeling of the image. In this paper, we formulate the global labeling problem with a novel densely connected Markov random field and show how to encode various intuitive potentials in a way that is amenable to efficient mean field inference [Krähenbühl et al., NIPS11]. Our potentials encode the compatibility between the global labeling and the patch-level predictions, contrast-sensitive smoothness as well as the fact that separate regions form

3. **Joint Training of a Convolutional Network and a Graphical Model for Human Pose Estimation** [123]

   Jonathan Tompson, Arjun Jain, Yann LeCun, Christoph Bregler

   NIPS 2014

   This paper proposes a new hybrid architecture that consists of a deep Convolutional Network and a Markov Random Field. We show how this architecture is successfully applied to the challenging problem of articulated human pose estimation in monocular images. The architecture can exploit structural domain constraints such as geometric relationships between body joint locations. We show that joint training of these two model paradigms improves performance and allows us to significantly outperform existing state-of-the-art techniques.

4. **Integration of Gibbs Markov Random Field and Hopfield-Type Neural Networks for Unsupervised Change Detection in Remotely Sensed Multitemporal Images** [41]

   Ashish Ghosh, Badri Narayan Subudhi, Lorenzo Bruzzone

   IEEE Transactions on Image Processing 2013

   The process of detecting changes in multitemporal, multi spectral remote sensing images is fundamental in many different applications, including environmental monitoring [1], land cover dynamics [2], forest monitoring [3], geographical survey [4], urban studies [5], etc. Generally, for change detection, a pair of multitemporal images acquired on the same geographical area at different times is analyzed. The most widely used unsupervised change detection scheme includes three fundamental steps: preprocessing, comparison and analysis [6]. The preprocessing stage includes normalization of the available multitemporal remote sensing images (i.e., co-registration, radiometric correction, geometric correction, atmospheric correction, etc. [7]), so that they can be used in the subsequent stages. In the comparison step, the multispectral remotely sensed images taken over the same geographical area at different times are compared using suitable mathematical operators like, single band differencing, vector differencing, ratioing, etc. Change vector analysis (CVA) [7] based difference image generation scheme is a popular technique. In the CVA-based scheme the difference image is generated by analyzing the magnitude of the spectral change vectors obtained from vector difference of each pair of corresponding pixels.

4 Deep learning and structured prediction

1. **End-to-End Learning for Structured Prediction Energy Networks**[8]

   Belanger, David and Yang, Bishan and McCallum, Andrew

   arXiv 2017

   Structured Prediction Energy Networks (Belanger and McCallum, 2016) (SPENs) are a simple, yet expressive family of structured prediction models. An energy function over candidate structured outputs is given by a deep network, and predictions are formed by gradient-based optimization. Unfortunately, we have struggled to apply the structured SVM (SSVM) learning method of Belanger and McCallum, 2016 to applications with more complex structure than multi-label classification. In general, SSVMs are unreliable whenever exact energy minimization is intractable. In response, we present end-to-end learning for SPENs, where the energy function is discriminatively trained by back-propagating through gradient-based prediction. This paper presents a collection of methods necessary to apply the technique to problems with complex structure. For example, we avoid vanishing gradients when learning SPENs

for convex relaxations of discrete prediction problems and explicitly train models such that energy minimization converges quickly in practice. Using end-to-end learning, we demonstrate the power of SPENs on 7-Scenes depth image denoising and CoNLL-2005 semantic role labeling tasks. In both, we outperform competitive baselines that employ more simplistic energy functions, but perform exact energy minimization. In particular, for denoising we achieve 40 PSNR, outperforming the previous state-of-the-art of 36.

2. **Learning deep structured models** [25]

   Liang-Chieh Chen, Alexander G. Schwing, Alan L. Yuille, Raquel Urtasun

   ICML 2015

   The goal of this paper [25] is to combine MRFs with deep learning to estimate complex representations while taking into account the dependencies between the output random variables. Towards this goal, Chen, Schwing, et al. [25] propose a training algorithm that is able to learn structured models jointly with deep features that form the MRF potentials.

3. **Structured Prediction Energy Networks** [7]

   David Belanger, Andrew McCallum

   ICML 2016

   This paper introduces structured prediction energy networks (SPENs), a flexible framework for structured prediction. A deep architecture is used to define an energy function of candidate labels, and then predictions are produced by using backpropagation to iteratively optimize the energy with respect to the labels. This deep architecture captures dependencies between labels that would lead to intractable graphical models, and performs structure learning by automatically learning discriminative features of the structured output. One natural application of our technique is multi-label classification, which traditionally has required strict prior assumptions about the interactions between labels to ensure tractable learning and prediction. We are able to apply SPENs to multi-label problems with substantially larger label sets than previous applications of structured prediction, while modeling high-order interactions using minimal structural assumptions. Overall, deep learning provides remarkable tools for learning features of the inputs to a prediction problem, and this work extends these techniques to learning features of structured outputs. Our experiments provide impressive performance on a variety of benchmark multi-label classification tasks, demonstrate that our technique can be used to provide interpretable structure learning, and illuminate fundamental trade-offs between feedforward and iterative structured prediction.

4. **Deep Learning and Structured Prediction for the Segmentation of Mass in Mammograms** [33]

   Neeraj Dhungel, Gustavo Carneiro, and Andrew P. Bradley

   MICCAI 2015

   In this paper, we explore the use of deep convolution and deep belief networks as potential functions in structured prediction models for the segmentation of breast masses from mammograms. In particular, the structured prediction models are estimated with loss minimization parameter learning algorithms, representing: a) conditional random field (CRF), and b) structured support vector machine (SSVM). For the CRF model, we use the inference algorithm based on tree re-weighted belief propagation with truncated fitting training, and for the SSVM model the inference is based on graph cuts with maximum margin training. We show empirically the importance of deep learning methods in producing state-of-the-art results for both structured prediction models. In addition, we show that our methods produce results that can be considered the best results to date on DDSM-BCRP and INbreast databases. Finally, we show that the CRF model is significantly faster than SSVM, both in terms of inference and training time, which suggests an advantage of CRF models when combined with deep learning potential functions.

5. **Learning Structured Inference Neural Networks with Label Relations** [50]

   Hexiang Hu, Guang-Tong Zhou1, Zhiwei Deng, Zicheng Liao, and Greg Mori

CVPR 2016

Images of scenes have various objects as well as abundant attributes, and diverse levels of visual categorization are possible. A natural image could be assigned with finegrained labels that describe major components, coarsegrained labels that depict high level abstraction, or a set of labels that reveal attributes. Such categorization at different concept layers can be modeled with label graphs encoding label information. In this paper, we exploit this rich information with a state-of-art deep learning framework, and propose a generic structured model that leverages diverse label relations to improve image classification performance. Our approach employs a novel stacked label prediction neural network, capturing both inter-level and intra-level label semantics. We evaluate our method on benchmark image datasets, and empirical results illustrate the efficacy of our model.

6. **BrainNetCNN: Convolutional neural networks for brain networks; towards predicting neurodevelopment** [61]

   Jeremy Kawaharaa, Colin J. Browna, Steven P. Millerb, Brian G. Bootha, Vann Chaub, Ruth E. Grunauc, Jill G. Zwickerc, Ghassan Hamarneh

   NeuroImage 2017

   This paper proposes BrainNetCNN, a convolutional neural network (CNN) framework to predict clinical neurodevelopmental outcomes from brain networks. In contrast to the spatially local convolutions done in traditional image-based CNNs, our BrainNetCNN is composed of novel edge-to-edge, edge-to-node and node-to-graph convolutional filters that leverage the topological locality of structural brain networks. We apply the BrainNetCNN framework to predict cognitive and motor developmental outcome scores from structural brain networks of infants born preterm. Diffusion tensor images (DTI) of preterm infants, acquired between 27 and 46 weeks gestational age, were used to construct a dataset of structural brain connectivity networks. We first demonstrate the predictive capabilities of BrainNetCNN on synthetic phantom networks with simulated injury patterns and added noise. BrainNetCNN outperforms a fully connected neural-network with the same number of model parameters on both phantoms with focal and diffuse injury patterns. We then apply our method to the task of joint prediction of Bayley-III cognitive and motor scores, assessed at 18 months of age, adjusted for prematurity. We show that our BrainNetCNN framework outperforms a variety of other methods on the same data. Furthermore, BrainNetCNN is able to identify an infant's postmenstrual age to within about 2 weeks. Finally, we explore the high-level features learned by BrainNetCNN by visualizing the importance of each connection in the brain with respect to predicting the outcome scores. These findings are then discussed in the context of the anatomy and function of the developing preterm infant brain.

7. **Composing graphical models with neural networks for structured representations and fast inference** [58]

   Matthew James Johnson, David Duvenaud

   NIPS 2016

   This paper proposes a general modeling and inference framework that combines the complementary strengths of probabilistic graphical models and deep learning methods. Our model family composes latent graphical models with neural network observation likelihoods. For inference, we use recognition networks to produce local evidence potentials, then combine them with the model distribution using efficient message-passing algorithms. All components are trained simultaneously with a single stochastic variational inference objective. We illustrate this framework by automatically segmenting and categorizing mouse behavior from raw depth video, and demonstrate several other example models.

8. **Deep Belief Network-Based Approaches for Link Prediction in Signed Social Networks** [73]

   Feng Liu, Bingquan Liu, Chengjie Sun, Ming Liu , Xiaolong Wang

   Entropy 2015

   In some online social network services (SNSs), the members are allowed to label their relationships with others, and such relationships can be represented as the links with signed values (positive or negative). The networks containing such relations are named signed social networks (SSNs), and some real-world

complex systems can be also modeled with SSNs. Given the information of the observed structure of an SSN, the link prediction aims to estimate the values of the unobserved links. Noticing that most of the previous approaches for link prediction are based on the members' similarity and the supervised learning method, however, research work on the investigation of the hidden principles that drive the behaviors of social members are rarely conducted. In this paper, the deep belief network (DBN)-based approaches for link prediction are proposed. Including an unsupervised link prediction model, a feature representation method and a DBN-based link prediction method are introduced. The experiments are done on the datasets from three SNSs (social networking services) in different domains, and the results show that our methods can predict the values of the links with high performance and have a good generalization ability across these datasets.

9. **A Deep Learning Approach to Link Prediction in Dynamic Networks** [70]

   Xiaoyi Li, Nan Du, Hui Li, Kang Li, Jing Gao and Aidong Zhang

   SIAM 2014

   Time varying problems usually have complex underlying structures represented as dynamic networks where entities and relationships appear and disappear over time. The problem of efficiently performing dynamic link inference is extremely challenging due to the dynamic nature in massive evolving networks especially when there exist sparse connectivities and nonlinear transitional patterns. In this paper, we propose a novel deep learning framework, i.e., Conditional Temporal Restricted Boltzmann Machine (ctRBM), which predicts links based on individual transition variance as well as influence introduced by local neighbors. The proposed model is robust to noise and have the exponential capability to capture nonlinear variance. We tackle the computational challenges by developing an efficient algorithm for learning and inference of the proposed model. To improve the effi- ciency of the approach, we give a faster approximated implementation based on a proposed Neighbor Influence Clustering algorithm. Extensive experiments on simulated as well as real-world dynamic networks show that the proposed method outperforms existing algorithms in link inference on dynamic networks.

10. **Structured Prediction with Convolutional Neural Networks for Multimodal Brain Tumor Segmentation** [36]

    Pavel Dvorak, Bjoern Menze

    MICCAI-BRATS 2015

    Most medical images feature a high similarity in the intensities of nearby pixels and a strong correlation of intensity profiles across different image modalities. One way of dealing with – and even exploiting – this correlation is the use of local image patches. In the same way, there is a high correlation between nearby labels in image annotation, a feature that has been used in the "local structure prediction" of local label patches. In the present study we test this local structure prediction approach for 3D segmentation tasks, systematically evaluating different parameters that are relevant for the dense annotation of anatomical structures. We choose convolutional neural network as learning algorithm, as it is known to be suited for dealing with correlation between features. We evaluate our approach on the public BRATS2014 data set with three multimodal segmentation tasks, being able to obtain state-of-the-art results for this brain tumor segmentation data set consisting of 254 multimodal volumes with computing time of only 13 seconds per volume.

11. **Deep Supervised and Convolutional Generative Stochastic Network for Protein Secondary Structure Prediction** [150]

    Jian Zhou, Olga G. Troyanskaya

    ICML 2014

    Predicting protein secondary structure is a fundamental problem in protein structure prediction. Here we present a new supervised generative stochastic network (GSN) based method to predict local secondary structure with deep hierarchical representations. GSN is a recently proposed deep learning technique (Bengio & Thibodeau-Laufer, 2013) to globally train deep generative model. We present the supervised extension of GSN, which learns a Markov chain to sample from a conditional distribution,

and applied it to protein structure prediction. To scale the model to full-sized, high-dimensional data, like protein sequences with hundreds of aminoacids, we introduce a convolutional architecture, which allows efficient learning across multiple layers of hierarchical representations. Our architecture uniquely focuses on predicting structured low-level labels informed with both low and high-level representations learned by the model. In our application this corresponds to labeling the secondary structure state of each amino-acid residue. We trained and tested the model on separate sets of non-homologous proteins sharing less than 30% sequence identity. Our model achieves 66.4% Q8 accuracy on the CB513 dataset, better than the previously reported best performance 64.9% (Wang et al., 2011) for this challenging secondary structure prediction problem.

12. **Structured Sequence Modeling with Graph Convolutional Recurrent Networks** [111]

    Youngjoo Seo, Michaël Defferrard, Pierre Vandergheynst, Xavier Bresson

    arXiv

    This paper introduces Graph Convolutional Recurrent Network (GCRN), a deep learning model able to predict structured sequences of data. Precisely, GCRN is a generalization of classical recurrent neural networks (RNN) to data structured by an arbitrary graph. Such structured sequences can represent series of frames in videos, spatio-temporal measurements on a network of sensors, or random walks on a vocabulary graph for natural language modeling. The proposed model combines convolutional neural networks (CNN) on graphs to identify spatial structures and RNN to find dynamic patterns. We study two possible architectures of GCRN, and apply the models to two practical problems: predicting moving MNIST data, and modeling natural language with the Penn Treebank dataset. Experiments show that exploiting simultaneously graph spatial and dynamic information about data can improve both precision and learning speed.

13. **Protein contact prediction from amino acid co-evolution using convolutional networks for graph-valued images** [43]

    Vladimir Golkov, Marcin J. Skwark, Antonij Golkov, Alexey Dosovitskiy, Thomas Brox, Jens Meiler, and Daniel Cremers

    NIPS 2016

    Proteins are responsible for most of the functions in life, and thus are the central focus of many areas of biomedicine. Protein structure is strongly related to protein function, but is difficult to elucidate experimentally, therefore computational structure prediction is a crucial task on the way to solve many biological questions. A contact map is a compact representation of the three-dimensional structure of a protein via the pairwise contacts between the amino acids constituting the protein. We use a convolutional network to calculate protein contact maps from detailed evolutionary coupling statistics between positions in the protein sequence. The input to the network has an image-like structure amenable to convolutions, but every "pixel" instead of color channels contains a bipartite undirected edge-weighted graph. Authors propose several methods for treating such "graph-valued images" in a convolutional network. The proposed method outperforms state-of-the-art methods by a considerable margin.

14. **A Deep Learning Network Approach to ab initio Protein Secondary Structure Prediction** [114]

    Matt Spencer, Jesse Eickholt, and Jianlin Cheng

    Transactions on Computational Biology and Bioinformatics 2015

    Ab initio protein secondary structure (SS) predictions are utilized to generate tertiary structure predictions, which are increasingly demanded due to the rapid discovery of proteins. Although recent developments have slightly exceeded previous methods of SS prediction, accuracy has stagnated around 80 percent and many wonder if prediction cannot be advanced beyond this ceiling. Disciplines that have traditionally employed neural networks are experimenting with novel deep learning techniques in attempts to stimulate progress. Since neural networks have historically played an important role in SS prediction, we wanted to determine whether deep learning could contribute to the advancement of

this field as well. We developed an SS predictor that makes use of the position-specific scoring matrix generated by PSI-BLAST and deep learning network architectures, which we call DNSS. Graphical processing units and CUDA software optimize the deep network architecture and efficiently train the deep networks. Optimal parameters for the training process were determined, and a workflow comprising three separately trained deep networks was constructed in order to make refined predictions. This deep learning network approach was used to predict SS for a fully independent test dataset of 198 proteins, achieving a Q3 accuracy of 80.7 percent and a Sov accuracy of 74.2 percent.

References

[1] B. Adhikari, Y. Zhang, N. Ramakrishnan, and B. A. Prakash. Distributed representation of subgraphs. *arXiv preprint arXiv:1702.06921*, 2017.

[2] R. Anirudh and J. J. Thiagarajan. Bootstrapping graph convolutional neural networks for autism spectrum disorder classification. *arXiv preprint arXiv:1704.07487v1*, 2017.

[3] A. Arnab, S. Jayasumana, S. Zheng, and P. H. Torr. Higher order conditional random fields in deep neural networks. In *European Conference on Computer Vision*, pages 524–540. Springer, 2016.

[4] V. Badrinarayanan, A. Kendall, and R. Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *arXiv preprint arXiv:1511.00561*, 2015.

[5] S.-Y. Bai, S. Agethen, T.-H. Chao, and W. Hsu. Semi-supervised learning for convolutional neural networks via online graph construction. *arXiv preprint arXiv:1511.06104*, 2015.

[6] J. Bastings, I. Titov, W. Aziz, D. Marcheggiani, and K. Sima'an. Graph convolutional encoders for syntax-aware neural machine translation. *arXiv preprint arXiv:1704.04675*, 2017.

[7] D. Belanger and A. McCallum. Structured prediction energy networks. In *Proceedings of the International Conference on Machine Learning*, 2016.

[8] D. Belanger, B. Yang, and A. McCallum. End-to-end learning for structured prediction energy networks. *arXiv preprint arXiv:1703.05667*, 2017.

[9] E. Blaisten-Barojas. *Unsupervised Anomaly Detection in Sequences Using Long Short Term Memory Recurrent Neural Networks*. PhD thesis, George Mason University, 2016.

[10] M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst. Geometric deep learning: going beyond euclidean data. *arXiv preprint arXiv:1611.08097*, 2016.

[11] J. Bruna, W. Zaremba, A. Szlam, and Y. LeCun. Spectral networks and deep locally connected networks on graphs. *arXiv preprint arXiv:1312.6203*, 2013.

[12] J. Bruna, W. Zaremba, A. Szlam, and Y. LeCun. Spectral networks and locally connected networks on graphs. *arXiv preprint arXiv:1312.6203*, 2013.

[13] T. D. Bui, S. Ravi, and V. Ramavajjala. Neural graph machines: Learning neural networks using graphs. *arXiv preprint arXiv:1703.04818*, 2017.

[14] B. Bünz and M. Lamm. Graph neural networks and boolean satisfiability. *arXiv preprint arXiv:1702.03592*, 2017.

[15] J. Cai, L. Lu, Z. Zhang, F. Xing, L. Yang, and Q. Yin. Pancreas segmentation in mri using graph-based decision fusion on convolutional neural networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 442–450. Springer, 2016.

[16] S. Cao, W. Lu, and Q. Xu. Grarep: Learning graph representations with global structural information. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, pages 891–900. ACM, 2015.

[17] S. Cao, W. Lu, and Q. Xu. Deep neural networks for learning graph representations. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, pages 1145–1152, 2016.

[18] S. Chandra and I. Kokkinos. Fast, exact and multi-scale inference for semantic image segmentation with deep gaussian crfs. In *European Conference on Computer Vision*, pages 402–418. Springer, 2016.

[19] S. Chang, W. Han, J. Tang, G.-J. Qi, C. C. Aggarwal, and T. S. Huang. Heterogeneous network embedding via deep architectures. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 119–128, 2015.

[20] S. Chauhan and L. Vig. Anomaly detection in ecg time signals via deep long short-term memory networks. In *Data Science and Advanced Analytics (DSAA), 2015. 36678 2015. IEEE International Conference on*, pages 1–7. IEEE, 2015.

[21] H. Chen, Q. Dou, L. Yu, and P.-A. Heng. Voxresnet: Deep voxelwise residual networks for volumetric brain segmentation. *arXiv preprint arXiv:1608.05895*, 2016.

[22] J. Chen, L. Yang, Y. Zhang, M. Alber, and D. Z. Chen. Combining fully convolutional and recurrent neural networks for 3d biomedical image segmentation. In *Advances in Neural Information Processing Systems*, pages 3036–3044, 2016.

[23] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. Semantic image segmentation with deep convolutional nets and fully connected crfs. *arXiv preprint arXiv:1412.7062*, 2014.

[24] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *arXiv preprint arXiv:1606.00915*, 2016.

[25] L.-C. Chen, A. G. Schwing, A. L. Yuille, and R. Urtasun. Learning deep structured models. In *ICML*, pages 1785–1794, 2015.

[26] X. Chen, X. Cheng, and S. Mallat. Unsupervised deep haar scattering on graphs. In *Advances in Neural Information Processing Systems*, pages 1709–1717, 2014.

[27] M. Cheng, Q. Xu, J. Lv, W. Liu, Q. Li, and J. Wang. Ms-lstm: A multi-scale lstm model for bgp anomaly detection. In *Network Protocols (ICNP), 2016 IEEE 24th International Conference on*, pages 1–6. IEEE, 2016.

[28] E. Choi, M. T. Bahadori, L. Song, W. F. Stewart, and J. Sun. Gram: Graph-based attention model for healthcare representation learning. *arXiv preprint arXiv:1611.07012*, 2016.

[29] P. F. Christ, M. E. A. Elshaer, F. Ettlinger, S. Tatavarty, M. Bickel, P. Bilic, M. Rempfler, M. Armbruster, F. Hofmann, M. D'Anastasi, et al. Automatic liver and lesion segmentation in ct using cascaded fully convolutional neural networks and 3d conditional random fields. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 415–423. Springer, 2016.

[30] P. F. Christ, F. Ettlinger, F. Grün, M. E. A. Elshaera, J. Lipkova, S. Schlecht, F. Ahmaddy, S. Tatavarty, M. Bickel, P. Bilic, et al. Automatic liver and tumor segmentation of ct and mri volumes using cascaded fully convolutional neural networks. *arXiv preprint arXiv:1702.05970*, 2017.

[31] H. Dai, Y. Wang, R. Trivedi, and L. Song. Deep coevolutionary network: Embedding user and item features for recommendation. 2017.

[32] M. Defferrard, X. Bresson, and P. Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. In *30th Conference on Neural Information Processing Systems*, pages 3837–3845, 2016.

[33] N. Dhungel, G. Carneiro, and A. P. Bradley. Deep learning and structured prediction for the segmentation of mass in mammograms. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 605–612. Springer, 2015.

[34] K. Do, T. Tran, and S. Venkatesh. Multilevel anomaly detection for mixed data. *arXiv preprint arXiv:1610.06249*, 2016.

[35] D. K. Duvenaud, D. Maclaurin, J. Iparraguirre, R. Bombarell, T. Hirzel, A. Aspuru-Guzik, and R. P. Adams. Convolutional networks on graphs for learning molecular fingerprints. In *29th Conference on Neural Information Processing Systems*, pages 2224–2232, 2015.

[36] P. Dvorak and B. Menze. Structured prediction with convolutional neural networks for multimodal brain tumor segmentation. *Proceeding of the Multimodal Brain Tumor Image Segmentation Challenge*, pages 13–24, 2015.

[37] M. Edwards and X. Xie. Graph based convolutional neural network. *arXiv preprint arXiv:1609.08965*, 2016.

[38] Y. Feng, Y. Yuan, and X. Lu. Deep representation for abnormal event detection in crowded scenes. In *Proceedings of the 2016 ACM on Multimedia Conference*, pages 591–595. ACM, 2016.

[39] A. Garcia-Garciac, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, and J. Garcia-Rodriguez. A review on deep learning techniques applied to semantic segmentation. *arXiv:1704.06857v1*, 2017.

[40] L. A. Gatys, A. S. Ecker, and M. Bethge. Image style transfer using convolutional neural networks. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[41] A. Ghosh, B. N. Subudhi, and L. Bruzzone. Integration of gibbs markov random field and hopfield-type neural networks for unsupervised change detection in remotely sensed multitemporal images. *IEEE Transactions on Image Processing*, 22(8):3087–3096, 2013.

[42] V. Golkov, M. J. Skwark, A. Golkov, A. Dosovitskiy, T. Brox, J. Meiler, and D. Cremers. Protein contact prediction from amino acid co-evolution using convolutional networks for graph-valued images. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems 29*, pages 4222–4230. Curran Associates, Inc., 2016.

[43] V. Golkov, M. J. Skwark, A. Golkov, A. Dosovitskiy, T. Brox, J. Meiler, and D. Cremers. Protein contact prediction from amino acid co-evolution using convolutional networks for graph-valued images. In *Advances in Neural Information Processing Systems*, pages 4215–4223, 2016.

[44] A. Grover and J. Leskovec. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 855–864. ACM, 2016.

[45] H. Gui, J. Liu, F. Tao, M. Jiang, B. Norick, and J. Han. Large-scale embedding learning in heterogeneous event data. In *Data Mining (ICDM), 2016 IEEE 16th International Conference on*, pages 907–912. IEEE, 2016.

[46] K. Guu, J. Miller, and P. Liang. Traversing knowledge graphs in vector space. *EMNLP*, 2015.

[47] B. Harwood and T. Drummond. Fanng: Fast approximate nearest neighbour graphs. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5713–5722, 2016.

[48] Y. Hechtlinger, P. Chakravarti, and J. Qin. A generalization of convolutional neural networks to graph-structured data. *arXiv preprint arXiv:1704.08165v1*, 2017.

[49] M. Henaff, J. Bruna, and Y. LeCun. Deep convolutional networks on graph-structured data. *arXiv preprint arXiv:1506.05163*, 2015.

[50] H. Hu, G.-T. Zhou, Z. Deng, Z. Liao, and G. Mori. Learning structured inference neural networks with label relations. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2960–2968, 2016.

[51] H. Huang, L. Heck, and H. Ji. Leveraging deep neural networks and knowledge graphs for entity disambiguation. *arXiv preprint arXiv:1504.07678*, 2015.

[52] Z. Huang, W. Xu, and K. Yu. Bidirectional lstm-crf models for sequence tagging. *arXiv preprint arXiv:1508.01991*, 2015.

[53] F. Isensee, P. Kickingereder, D. Bonekamp, M. Bendszus, W. Wick, H.-P. Schlemmer, and K. Maier-Hein. Brain tumor segmentation using large receptive field deep convolutional neural networks. In *Bildverarbeitung für die Medizin 2017*, pages 86–91. Springer, 2017.

[54] A. Jain, A. R. Zamir, S. Savarese, and A. Saxena. Structural-rnn: Deep learning on spatio-temporal graphs. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5308–5317, 2016.

[55] V. Jampani, M. Kiefel, and P. V. Gehler. Learning sparse high dimensional filters: Image filtering, dense crfs and bilateral neural networks. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[56] A. Jansen, G. Sell, and V. Lyzinski. Scalable out-of-sample extension of graph embeddings using deep neural networks. *arXiv preprint arXiv:1508.04422*, 2015.

[57] D. D. Johnson. Learning graphical state transitions. *ICLR*, 2017.

[58] M. Johnson, D. K. Duvenaud, A. Wiltschko, R. P. Adams, and S. R. Datta. Composing graphical models with neural networks for structured representations and fast inference. In *Advances in Neural Information Processing Systems*, pages 2946–2954, 2016.

[59] R. G. Jothi and S. M. Rani. Hybrid neural network for classification of graph structured data. *International Journal of Machine Learning and Cybernetics*, 6(3):465–474, 2015.

[60] K. Kamnitsas, C. Ledig, V. F. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker. Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation. *Medical Image Analysis*, 36:61–78, 2017.

[61] J. Kawahara, C. J. Brown, S. P. Miller, B. G. Booth, V. Chau, R. E. Grunau, J. G. Zwicker, and G. Hamarneh. Brainnetcnn: Convolutional neural networks for brain networks; towards predicting neurodevelopment. *NeuroImage*, 146:1038–1049, 2017.

[62] S. Kearnes, K. McCloskey, M. Berndl, V. Pande, and P. Riley. Molecular graph convolutions: moving beyond fingerprints. *Journal of computer-aided molecular design*, 30(8):595–608, 2016.

[63] T. N. Kipf and M. Welling. Semi-supervised classification with graph convolutional networks. *ICLR*, 2017.

[64] P. Knöbelreiter, C. Reinbacher, A. Shekhovtsov, and T. Pock. End-to-end training of hybrid cnn-crf models for stereo. *arXiv preprint arXiv:1611.10229*, 2016.

[65] S. I. Ktena, S. Parisot, E. Ferrante, M. Rajchl, M. Lee, B. Glocker, and D. Rueckert. Distance metric learning using graph convolutional networks: Application to functional brain networks. *arXiv preprint arXiv:1703.02161*, 2017.

[66] J. Lee, H. Kim, J. Lee, and S. Yoon. Transfer learning for deep learning on graph-structured data. 2017.

[67] K. Lee, A. Qadir, S. A. Hasan, V. Datla, A. Prakash, J. Liu, and O. Farri. Adverse drug event detection in tweets with semi-supervised convolutional neural networks. In *Proceedings of the 26th International Conference on World Wide Web*, pages 705–714. International World Wide Web Conferences Steering Committee, 2017.

[68] C. Li, X. Guo, and Q. Mei. Deepgraph: Graph structure predicts network growth. *arXiv preprint arXiv:1610.06251*, 2016.

[69] W. Li, G. Wu, and Q. Du. Transferred deep learning for anomaly detection in hyperspectral imagery. *IEEE Geoscience and Remote Sensing Letters*, 2017.

[70] X. Li, N. Du, H. Li, K. Li, J. Gao, and A. Zhang. A deep learning approach to link prediction in dynamic networks. In *Proceedings of the 2014 SIAM International Conference on Data Mining*, pages 289–297. SIAM, 2014.

[71] Y. Li, D. Tarlow, M. Brockschmidt, and R. Zemel. Gated graph sequence neural networks. *ICLR*, 2016.

[72] J. Liang, P. Jacobs, and S. Parthasarathy. Seano: Semi-supervised embedding in attributed networks with outliers. *arXiv preprint arXiv:1703.08100*, 2017.

[73] F. Liu, B. Liu, C. Sun, M. Liu, and X. Wang. Deep belief network-based approaches for link prediction in signed social networks. *Entropy*, 17(4):2140–2169, 2015.

[74] M. Looks, M. Herreshoff, D. Hutchins, and P. Norvig. Deep learning with dynamic computation graphs. *arXiv preprint arXiv:1702.02181*, 2017.

[75] F. Lu, F. Wu, P. Hu, Z. Peng, and D. Kong. Automatic 3d liver location and segmentation via convolutional neural network and graph cut. *International Journal of Computer Assisted Radiology and Surgery*, pages 1–12, 2016.

[76] W. Lu, B. Liang, Y. Cheng, D. Meng, J. Yang, and T. Zhang. Deep model based domain adaptation for fault diagnosis. *IEEE Transactions on Industrial Electronics*, 64(3):2296–2305, 2017.

[77] Y. Luo, Q. Wang, B. Wang, and L. Guo. Context-dependent knowledge graph embedding. In *EMNLP*, pages 1656–1661, 2015.

[78] T. Ma, F. Wang, J. Cheng, Y. Yu, and X. Chen. A hybrid spectral clustering and deep neural network ensemble algorithm for intrusion detection in sensor networks. *Sensors*, 16(10):1701, 2016.

[79] X. Ma and E. Hovy. End-to-end sequence labeling via bi-directional lstm-cnns-crf. *arXiv preprint arXiv:1603.01354*, 2016.

[80] P. Malhotra, A. Ramakrishnan, G. Anand, L. Vig, P. Agarwal, and G. Shroff. Lstm-based encoder-decoder for multi-sensor anomaly detection. *arXiv preprint arXiv:1607.00148*, 2016.

[81] P. Malhotra, L. Vig, G. Shroff, and P. Agarwal. Long short term memory networks for anomaly detection in time series. In *Proceedings*, page 89. Presses universitaires de Louvain, 2015.

[82] F. Manessi, A. Rozza, and M. Manzo. Dynamic graph convolutional networks. *arXiv preprint arXiv:1704.06199*, 2017.

[83] D. Marcheggiani and I. Titov. Encoding sentences with graph convolutional networks for semantic role labeling. *arXiv preprint arXiv:1703.04826*, 2017.

[84] K. Marino, R. Salakhutdinov, and A. Gupta. The more you know: Using knowledge graphs for image classification. *arXiv preprint arXiv:1612.04844*, 2016.

[85] F. Milletari, N. Navab, and S.-A. Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *3D Vision (3DV), 2016 Fourth International Conference on*, pages 565–571. IEEE, 2016.

[86] F. Monti, D. Boscaini, J. Masci, E. Rodolà, J. Svoboda, and M. M. Bronstein. Geometric deep learning on graphs and manifolds using mixture model cnns. *arXiv preprint arXiv:1611.08402*, 2016.

[87] F. Monti, M. M. Bronstein, and X. Bresson. Geometric matrix completion with recurrent multi-graph neural networks. *arXiv preprint arXiv:1704.06803v1*, 2017.

[88] A. Nanduri and L. Sherry. Anomaly detection in aircraft data using recurrent neural networks (rnn). In *Integrated Communications Navigation and Surveillance (ICNS), 2016*, pages 5C2–1. IEEE, 2016.

[89] A. Narayanan, M. Chandramohan, L. Chen, Y. Liu, and S. Saminathan. subgraph2vec: Learning distributed representations of rooted sub-graphs from large graphs. *arXiv preprint arXiv:1606.08928*, 2016.

[90] M. Niepert, M. Ahmed, and K. Kutzkov. Learning convolutional neural networks for graphs. In *Proceedings of The 33rd International Conference on Machine Learning*, pages 2014–2023, 2016.

[91] G. Nikolentzos, P. Meladianos, and M. Vazirgiannis. Matching node embeddings for graph similarity. In *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

[92] M. Ou, P. Cui, J. Pei, Z. Zhang, and W. Zhu. Asymmetric transitivity preserving graph embedding. In *Proc. of ACM SIGKDD*, pages 1105–1114, 2016.

[93] S. Paisitkriangkrai, J. Sherrah, P. Janney, V.-D. Hengel, et al. Effective semantic pixel labelling with convolutional networks and conditional random fields. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 36–43, 2015.

[94] S. Pan, J. Wu, X. Zhu, C. Zhang, and Y. Wang. Tri-party deep network representation. *Network*, 11(9):12, 2016.

[95] S. Parisot, S. I. Ktena, E. Ferrante, M. Lee, R. G. Moreno, B. Glocker, and D. Rueckert. Spectral graph convolutions on population graphs for disease prediction. *arXiv preprint arXiv:1703.03020*, 2017.

[96] D. Pathak, P. Krahenbuhl, and T. Darrell. Constrained convolutional neural networks for weakly supervised segmentation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1796–1804, 2015.

[97] W. Pei, T. Ge, and B. Chang. An effective neural network model for graph-based dependency parsing. In *ACL*, pages 313–322, 2015.

[98] B. Perozzi, R. Al-Rfou, and S. Skiena. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 701–710. ACM, 2014.

[99] B. Perozzi, V. Kulkarni, and S. Skiena. Walklets: Multiscale graph embeddings for interpretable network classification. *arXiv preprint arXiv:1605.02115*, 2016.

[100] P. O. Pinheiro and R. Collobert. From image-level to pixel-level labeling with convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1713–1721, 2015.

[101] R. P. Poudel, P. Lamata, and G. Montana. Recurrent fully convolutional neural networks for multi-slice mri cardiac segmentation. *arXiv preprint arXiv:1608.03974*, 2016.

[102] G. Puy, S. Kitic, and P. Pérez. Unifying local and non-local signal processing with graph cnns. *arXiv preprint arXiv:1702.07759*, 2017.

[103] T. M. Quan, D. G. Hilderbrand, and W.-K. Jeong. Fusionnet: A deep fully residual convolutional neural network for image segmentation in connectomics. *arXiv preprint arXiv:1612.05360*, 2016.

[104] D. L. Richmond, D. Kainmueller, M. Y. Yang, E. W. Myers, and C. Rother. Relating cascaded random forests to deep convolutional neural networks for semantic segmentation. *arXiv preprint arXiv:1507.07583*, 2015.

[105] P. Ristoski and H. Paulheim. Rdf2vec: Rdf graph embeddings for data mining. In *International Semantic Web Conference*, pages 498–514. Springer, 2016.

[106] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241. Springer, 2015.

[107] R. Rustamov and L. J. Guibas. Wavelets on graphs via deep learning. In *Advances in neural information processing systems*, pages 998–1006, 2013.

[108] M. Sabokrou, M. Fayyaz, M. Fathy, et al. Fully convolutional neural network for fast anomaly detection in crowded scenes. *arXiv preprint arXiv:1609.00866*, 2016.

[109] M. Sabokrou, M. Fayyaz, M. Fathy, and R. Klette. Deep-cascade: Cascading 3d deep neural networks for fast anomaly detection and localization in crowded scenes. *IEEE Transactions on Image Processing*, 2017.

[110] M. Schlichtkrull, T. N. Kipf, P. Bloem, R. v. d. Berg, I. Titov, and M. Welling. Modeling relational data with graph convolutional networks. *arXiv preprint arXiv:1703.06103*, 2017.

[111] Y. Seo, M. Defferrard, P. Vandergheynst, and X. Bresson. Structured sequence modeling with graph convolutional recurrent networks. *arXiv preprint arXiv:1612.07659*, 2016.

[112] E. Shelhamer, J. Long, and T. Darrell. Fully convolutional networks for semantic segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(4):640–651, 2017.

[113] M. Simonovsky and N. Komodakis. Dynamic edge-conditioned filters in convolutional neural networks on graphs. *arXiv preprint arXiv:1704.02901*, 2017.

[114] M. Spencer, J. Eickholt, and J. Cheng. A deep learning network approach to ab initio protein secondary structure prediction. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 12(1):103–112, 2015.

[115] H. Su, S. Maji, E. Kalogerakis, and E. Learned-Miller. Multi-view convolutional neural networks for 3d shape recognition. In *Proceedings of the IEEE international conference on computer vision*, pages 945–953, 2015.

[116] F. P. Such, S. Sah, M. Dominguez, S. Pillai, C. Zhang, A. Michael, N. Cahill, and R. Ptucha. Robust spatial filtering with graph convolutional neural networks. *arXiv preprint arXiv:1703.00792*, 2017.

[117] X. Sun, J. Guo, X. Ding, and T. Liu. A general framework for content-enhanced network representation learning. *arXiv preprint arXiv:1610.02906*, 2016.

[118] J. Tang, M. Qu, M. Wang, M. Zhang, J. Yan, and Q. Mei. Line: Large-scale information network embedding. In *Proceedings of the 24th International Conference on World Wide Web*, pages 1067–1077. ACM, 2015.

[119] A. Taylor, S. Leblanc, and N. Japkowicz. Anomaly detection in automobile control network data with long short-term memory networks. In *Data Science and Advanced Analytics (DSAA), 2016 IEEE International Conference on*, pages 130–139. IEEE, 2016.

[120] D. Teney, L. Liu, and A. v. d. Hengel. Graph-structured representations for visual question answering. *CVPR*, 2016.

[121] F. Tian, B. Gao, Q. Cui, E. Chen, and T.-Y. Liu. Learning deep representations for graph clustering. In *AAAI*, pages 1293–1299, 2014.

[122] V. S. Tomar and R. C. Rose. Graph based manifold regularized deep neural networks for automatic speech recognition. *arXiv preprint arXiv:1606.05925*, 2016.

[123] J. J. Tompson, A. Jain, Y. LeCun, and C. Bregler. Joint training of a convolutional network and a graphical model for human pose estimation. In *Advances in neural information processing systems*, pages 1799–1807, 2014.

[124] N. Torbati, A. Ayatollahi, and A. Kermani. An efficient neural network based method for medical image segmentation. *Computers in biology and medicine*, 44:76–87, 2014.

[125] S. Valverde, M. Cabezas, E. Roura, S. González-Villà, D. Pareto, J. C. Vilanova, L. Ramió-Torrentà, À. Rovira, A. Oliver, and X. Lladó. Improving automated multiple sclerosis lesion segmentation with a cascaded 3d convolutional neural network approach. *NeuroImage*, 2017.

[126] R. Vemulapalli, O. Tuzel, and M.-Y. Liu. Deep gaussian conditional random field network: A model-based deep network for discriminative denoising. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4801–4809, 2016.

[127] P. E. Vértes, A. Alexander-Bloch, and E. T. Bullmore. Generative models of rich clubs in hebbian neuronal networks and large-scale human brain networks. *Phil. Trans. R. Soc. B*, 369(1653):20130531, 2014.

[128] A. Vieira. Knowledge representation in graphs using convolutional neural networks. *arXiv preprint arXiv:1612.02255*, 2016.

[129] F. Visin, M. Ciccone, A. Romero, K. Kastner, K. Cho, Y. Bengio, M. Matteucci, and A. Courville. Reseg: A recurrent neural network-based model for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 41–48, 2016.

[130] F. Visin, K. Kastner, A. Courville, Y. Bengio, M. Matteucci, and K. Cho. Reseg: A recurrent neural network for object segmentation. *CoRR, abs/1511.07053*, 2, 2015.

[131] G. Wainrib and M. Galtier. Regular graphs maximize the variability of random neural networks. *Physical Review E*, 92(3):032802, 2015.

[132] D. Wang, P. Cui, and W. Zhu. Structural deep network embedding. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1225–1234. ACM, 2016.

[133] K. Wang, Y. Zhao, Q. Xiong, M. Fan, G. Sun, L. Ma, and T. Liu. Research on healthy anomaly detection model based on deep learning from multiple time-series physiological signals. *Scientific Programming*, 2016, 2016.

[134] P. Wang, Y. Gan, Y. Zhang, and P. Shui. 3d shape segmentation via shape fully convolutional networks. *arXiv preprint arXiv:1702.08675*, 2017.

[135] M. Wielgosz, A. Skoczeń, and M. Mertik. Recurrent neural networks for anomaly detection in the post-mortem time series of lhc superconducting magnets. *arXiv preprint arXiv:1702.00833*, 2017.

[136] F. Wu, X. Lu, J. Song, S. Yan, Z. M. Zhang, Y. Rui, and Y. Zhuang. Learning of multimodal representations with random walks on the click graph. *IEEE Transactions on Image Processing*, 25(2):630–642, 2016.

[137] P. Xu and R. Sarikaya. Convolutional neural network based triangular crf for joint intent detection and slot filling. In *Automatic Speech Recognition and Understanding (ASRU), 2013 IEEE Workshop on*, pages 78–83. IEEE, 2013.

[138] Z. Yan, V. Jagadeesh, D. DeCoste, W. Di, and R. Piramuthu. Hd-cnn: Hierarchical deep convolutional neural network for image classification. In *International Conference on Computer Vision (ICCV)*, volume 2. Citeseer, 2015.

[139] P. Yanardag and S. Vishwanathan. Deep graph kernels. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1365–1374. ACM, 2015.

[140] C. Yang, M. Sun, W. X. Zhao, Z. Liu, and E. Y. Chang. A neural network approach to joint modeling social networks and mobile trajectories. *arXiv preprint arXiv:1606.08154*, 2016.

[141] W. Yang, W. Ouyang, H. Li, and X. Wang. End-to-end learning of deformable mixture of parts and deep convolutional neural networks for human pose estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3073–3082, 2016.

[142] Z. Yang, W. W. Cohen, and R. Salakhutdinov. Revisiting semi-supervised learning with graph embeddings. *arXiv preprint arXiv:1603.08861*, 2016.

[143] K. Yao, B. Peng, G. Zweig, D. Yu, X. Li, and F. Gao. Recurrent conditional random field for language understanding. In *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, pages 4077–4081. IEEE, 2014.

[144] L. Yi, H. Su, X. Guo, and L. Guibas. Syncspeccnn: Synchronized spectral cnn for 3d shape segmentation. *arXiv preprint arXiv:1612.00606*, 2016.

[145] S. Zhai, Y. Cheng, W. Lu, and Z. Zhang. Deep structured energy based models for anomaly detection. In *Proceedings of the 33nd International Conference on Machine Learning, ICML*, pages 19–24, 2016.

[146] G. Zhang and W. B. Kleijn. Training deep neural networks via optimization over graphs. *arXiv preprint arXiv:1702.03380*, 2017.

[147] X. Zhao, Y. Wu, G. Song, Z. Li, Y. Fan, and Y. Zhang. Brain tumor segmentation using a fully convolutional neural network with conditional random fields. In *International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pages 75–87. Springer, 2016.

[148] X. Zhao, Y. Wu, G. Song, Z. Li, Y. Zhang, and Y. Fan. A deep learning model integrating fcnns and crfs for brain tumor segmentation. *arXiv preprint arXiv:1702.04528*, 2017.

[149] S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, and P. H. Torr. Conditional random fields as recurrent neural networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1529–1537, 2015.

[150] J. Zhou and O. G. Troyanskaya. Deep supervised and convolutional generative stochastic network for protein secondary structure prediction. In *ICML*, pages 745–753, 2014.

[151] S. Zhou, W. Shen, D. Zeng, M. Fang, Y. Wei, and Z. Zhang. Spatial–temporal convolutional neural networks for anomaly detection and localization in crowded scenes. *Signal Processing: Image Communication*, 47:358–368, 2016.

[152] Z. Zuo, B. Shuai, G. Wang, X. Liu, X. Wang, B. Wang, and Y. Chen. Convolutional recurrent neural networks: Learning spatial dependencies for image representation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 18–26, 2015.