A Survey on Improving Classification Accuracy using DNN

Uma Malusare[1]**, Prof. Dr. B. K. Sarkar**[2]
Department of Computer Engineering, Padmbhushan Vasantdada Patil Institute of Technology,Bavdhan Pune, Maharashtra[1 2]

Abstract— **Deep Neural Networks (DNNs) have demonstrated impressive performance in complex learning tasks like image classification or voice recognition. However, because of their multilayer nonlinear structure, they are not transparent, That is, it's hard to understand what makes them happen to a particular classification or recognition decision, given a new sample of data. Recently, several approaches have been proposed to understand and interpret the reasoning embodied in a DNN for a single test image. These methods quantify the "Importance" of individual pixels in relation to the classification decision and allow visualization in terms of heatmap in Pixel / input space. Although the utility of heatmaps can be judged subjectively by a human, a measure of objective quality is missing. n this article, we present a general methodology based on the region disruption for the evaluation of ordered collections of pixels such as heatmaps. We compare heatmaps calculated by three different the methods on SUN397, ILSVRC2012 and MIT place datasets. Our main result is that the relevance of the recently proposed layer-wise propagation algorithm provides qualitatively and quantitatively a better explanation of what made a DNN happen to a particular the classification decision that the approach focused on sensitivity or the method of devolution.**

Keywords: *Convolutional neural networks, explaining classification, image classification, interpretable machine learning, relevance models*

I INTRODUCTION

In recent years, Deep Neural Networks (DNNs) have emerged as the method of choice for perception tasks such as voice recognition and image classification. In essence, a DNN is a very complex nonlinear function, which makes it difficult to understand how a particular classification is sure. This lack of transparency is a major obstacle to the adoption of in-depth learning in industry, government and health care where the cost of errors is high. Deep Neural Network (DNN) used as a classifier of pixels. The the network calculates the probability that a pixel is a membrane, using as input the intensities of the image in a square window centered on the pixel itself. An image is then segmented by classifying its pixels. DNN is formed on a different cell with similar characteristics, in which the membranes have been annotated manually. Since DNN training methodologies (unsupervised pretraining, dropout, parallelization, GPUs, etc.) have been improved, DNNs are recently able to harvest extremely large amounts of training data and can thus achieve record performances in many research fields. At the same time, DNNs are generally conceived as black box methods, and users might consider this lack of transparency a drawback in practice. Namely, it is difficult to intuitively and quantitatively understand the result of DNN inference, i.e., for an individual novel input data point, what made the trained DNN model arrive at a particular response. Note that this aspect differs from feature selection, where the question is: which features are on average salient for the ensemble of training data?

In DNN a large body of work is dedicated to visualize particular neurons or neuron layers, they focus here on methods that visualize the impact of particular regions of a given and fixed single image for a prediction of this image. Zeiler and Fergus have proposed in their work a network propagation technique to identify patterns in a given input image that are linked to a particular DNN prediction. This method runs a backward algorithm that reuses the weights at each layer to propagate the prediction from the output down to the input layer, leading to the creation of meaningful patterns in input space.

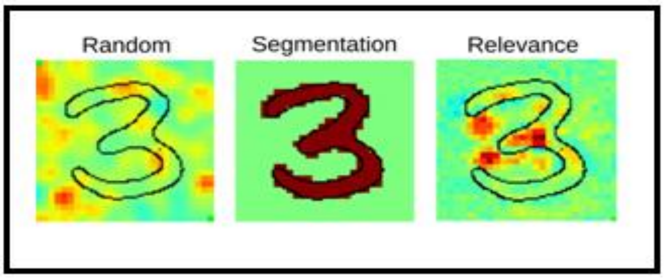

Figure 1: Comparison of three exemplary heatmaps for the image of a "3."

This approach was designed for a particular type of neural network, namely, convolution nets with max pooling and rectified linear units. A limitation of the deconvolution method is the absence of a particular theoretical criterion that would directly connect the predicted output to the produced pattern in a quantifiable way. Furthermore, the usage of image-specific information for generating the back projections in this method is limited to max-pooling layers alone. Previous work has focused on understanding nonlinear learning methods such as DNNs or kernel methods.

In figure 1, they will denote the visualizations produced by the above methods as heatmaps. While per se a heatmap is an interesting and intuitive tool that can already allow achieving transparency, it is difficult to quantitatively evaluate the quality of a heatmap. In other words we may ask: what exactly makes a “good” heatmap? A human may be able to intuitively assess the quality of a heatmap, e.g., by matching with a prior of what is regarded as being relevant. For practical applications, however, an automated objective and quantitative measure for assessing heatmap quality becomes necessary. Note that the validation of heatmap quality is important if we want to use it as input for further analysis. For example, we could run computationally more expensive algorithms only on relevant regions in the image, where relevance is detected by a heatmap.

II LITERATURE SURVEY

Wojciech Samek, Alexander Binder, Grégoire Montavon Sebastian Lapuschkin, and Klaus-Robert Müller [1], author presents a general methodology based on region perturbation for evaluating ordered collections of pixels such as heatmaps. They compare heatmaps computed by three different methods on the SUN397, ILSVRC2012, and MIT Places data sets.

Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton [2], they trained a large, deep convolutional neural network to classify the 1.2 million high-resolution images in the ImageNet LSVRC-2010 contest into the 1000 different classes. They achieved top-1 and top-5 error rates of 37.5% and 17.0% which is considerably better than the previous state-of-the-art. To make training faster, we used non-saturating neurons and a very efficient GPU implementation of the convolution operation.

Richard Socher, Alex Perelygin, Jean Y. Wu, Jason Chuang, Christopher D. Manning, Andrew Y. Ng and Christopher Potts [3], the author introduces a Sentiment Treebank. It includes fine grained sentiment labels for 215,154 phrases in the parse trees of 11,855 sentences and presents new challenges for sentiment compositionality. To address them, they introduce the Recursive Neural Tensor Network.

S. Ji, W. Xu, M. Yang, and K. Yu [4], the author develops a novel 3D CNN model for action recognition. This model extracts features from both the spatial and the temporal dimensions by performing 3D convolutions, thereby capturing the motion information encoded in multiple adjacent frames. The developed model generates multiple channels of information from the input frames, and the final feature representation combines information from all channels.

Q. V. Le, W. Y. Zou, S. Y. Yeung, and A. Y. Ng, [5], author present an extension of the Independent Subspace Analysis algorithm to learn invariant spatio-temporal features from unlabeled video data. they discovered that, despite its simplicity, this method performs surprisingly well when combined with deep learning techniques such as stacking and convolution to learn hierarchical representations.

G. Montavon, M. L. Braun, T. Krueger, and K.-R. Müller [6], they report on a set of recent methods that can be universally used to make kernel methods more transparent. The author discuss relevant dimension estimation (RDE) that allows to assess the underlying complexity and noise structure of a learning problem and thus to distinguish high/low noise scenarios of high/low complexity respectively

Table 1: Literature Survey

Sr. No	Paper Title	Authors	Method Proposed	Disadvantages
1	Evaluating the Visualization of What a Deep Neural Network Has Learned	Wojciech Samek, Alexander Binder, Grégoire Montavon Sebastian Lapuschkin, and Klaus-Robert Müller	Presents a general methodology based on region perturbation for evaluating ordered collections of pixels such as heatmaps.	Multilayer nonlinear structure, do not transparent.
2	Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank	Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton	They trained a large, deep convolutional neural network to classify the 1.2 million high-resolution images in the ImageNet LSVRC-2010 contest into the 1000 different classes.	Less Accurate.
3	3D convolutional neural networks for human action recognition	Richard Socher, Alex Perelygin, Jean Y. Wu, Manning, Andrew Y. Ng and Christopher Potts	The author introduces a Sentiment Treebank.	Performance of this method is not good.

4	Learning hierarchical invariant spatio-temporal features for action recognition with independent subspace analysis	S. Ji, W. Xu, M. Yang, and K. Yu	The author develops a novel 3D CNN model for action recognition.	It only utilizes the most reliable direct competitive information.
5	Analyzing local structure in kernel-based learning: Explanation, complexity, and reliability assessment	Q. V. Le, W. Y. Zou, S. Y. Yeung, and A. Y. Ng	Present an extension of the Independent Subspace Analysis algorithm to learn invariant spatio-temporal features from unlabeled video data.	Because of space limitation only present 28 percent result.
6	Analyzing local structure in kernel-based learning: Explanation, complexity, and reliability assessment	G. Montavon, M. L. Braun, T. Krueger, and K.-R. Müller	They report on a set of recent methods that can be universally used to make kernel methods more transparent.	Time Consuming.

III CONCLUSION

In this paper, we have studied an orthogonal research direction in our manuscript, namely, we have contributed to furthering the understanding and transparency of the decision making implemented by a trained DNN: for this we have focused on the heatmap concept that, e.g., in a computer vision application, is able to attribute the contribution of individual pixels to the DNN inference result for a novel data sample. We tackled the so far open problem of quantifying the quality of a heatmap. We proposed a region perturbation strategy that is based on the idea that flipping the most salient pixels first should lead to high performance decay. A large AOPC value as a function of the perturbation steps was shown to provide a good measure for a very informative heatmap. We also showed quantitatively and qualitatively that sensitivity maps and heatmaps computed with the deconvolution algorithm are much noisier than heatmaps computed with the LRP method, and thus are less suitable for identifying the most important regions with respect to the classification task.

REFERENCES

[1] Wojciech Samek, Alexander Binder, Grégoire Montavon Sebastian Lapuschkin, and Klaus-Robert Müller, “Evaluating the Visualization of What a Deep Neural Network Has Learned”, 2016.

[2] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, “Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank”., 2013.

[3] S. Ji, W. Xu, M. Yang, and K. Yu, “3D convolutional neural networks for human action recognition,.” 2010.

[4] Q. V. Le, W. Y. Zou, S. Y. Yeung, and A. Y. Ng, “Learning hierarchical invariant spatio-temporal features for action recognition with independent subspace analysis,” 2011.

[5]G. Montavon, M. L. Braun, T. Krueger, and K.-R. Müller, “Analyzing local structure in kernel-based learning: Explanation, complexity, and reliability assessment,” 2013.

[6] D. Erhan, A. Courville, and Y. Bengio, “Understanding representations learned in deep architectures,” Dépt. d’Informatique Recherche Opérationnelle, 2010.

[7]G. Montavon, M. L. Braun, and K.-R. Müller, “Kernel analysis of deep networks.,”2011.

[8] D. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, “Deep neural networks segment neuronal membranes in electron microscopy images,” 2012.