Noname manuscript No.
(will be inserted by the editor)

Semi-supervised deep neural network solution for regression and classification on stainless metal parts in manufacturing

Wenbin Zhang · Jochen Lang

Received: date / Accepted: date

Abstract Deep learning [1] been successful in many domain, it always gives incredible results while traditional machine learning algorithms relied on tons of knowledge base and reaching unacceptable results. However, deep learning does not perform well if there is lack of labeled data, especially in the industrial world, mannually labeling the data is time consuming and costly. The objective of my study is going to make use of the small amount of labeled dataset to generate unlimited number of unlabeled augmented data [2] can be applied to a semi-supervised learning setting. In another word, without putting enormous afford for labeling unlabeled data, the network should be able to achieve the same accuracy as the network which having huge labeled data. Hopefully, the semi-supervised network with unlabeled augmentations can outperform the state-of-art supervised solution. In addition to build the model base on small amount of labeled data, the model should also be capable to perform correctly on the image containing polished stainless steel metal surface. The proposed semi-supervised deep learning solution will open a new window for future research, showcasing well performance can be achieved without putting great amount of afford on labeling the data.

Keywords Deep learning · Semi-supervised learning · Image auto-augmentation

1 Introduction

Enclosures Direct Inc.(EDI) is a manufacturer of enclosures used to house and protect electrical and electronic devices. They have developed various industrial robotics based automation cells for its fabrication process. One of those cells utilizes a computer vision system to guide a robot in a welding

Wenbin Zhang
University of Ottawa
E-mail: wzhan133@uottawa.ca

application which comprised with a laser and an industrial camera. The laser is combined with an optical line generator to cast a line of structured light across a weld joint. An image of the structured light is captured using the camera and processed with specialized software tat was developed by the company. The software is designed to mathematically model the structured light in the image as a series of line equations; eventually it tends to find the coordinates of the intersection of the joint from the image which captured from the camera. However, the system has reliability issues when the material has reflective property such as polished stainless steel or the imaged has been disturbed by the ambient light. Even though, under all circumstance, the image of the structured light is clear to naked eyes and the location of the weld joint is obvious, the existing system failed due to the algorithm cannot analyze the image correctly. The company has archived data set of over 50,000 images that have been captured and stored during the production process. They would like to see a more robust machine learning solution can be designed to replace the current computer vision algorithm. In addition, a classification on different types of welding joints is also a desirable feature with the new model such as butt joint, corner joint, lap joint, tee-joint, edge joint, etc.

The main challenges involved to solve this problem can be listed as below:

- Construct a suitable deep learning model to best fit the main problem which including the regression problem as finding expected coordinates from the image and the classification problem as identifying the image into corresponding category according to the type of welding seams detected from the image.
- Auto generating unlimited number of augmented data from existing limited dataset.
- Finding the optimized solution to utilize both labeled data and unlimited number of unlabeled augmented data.
- Overcome the reflection affects by the stainless metal material.
- Propose a transfer learning and deep learning robust model that can efficiently solve the issue.

2 Literature Review

The main challenges in this project can be divided into two major problems which is constructing a network to solve image classification and logistic regression visual tasks and utilizing limited number of labeled data to work with unlabeled data efficiently. Base on those two different targets, my goal would be producing a framework that can satisfy both of desired requirements.

Image classification which can be defined as a task of categorizing images into corresponding numbers of predefined classes. Predicting the coordinates of the intersection of two metal sheets can be seen as an image regression problem. Recently, more and more work shows that feature extraction based deep learning structure is one of the best solution for image classification and logistic regression. According to Rawat et al.(2017) [3], deep learning models that exploit multiple layers of nonlinear information processing for feature extraction and classification have been shown to overcome this challenge. It extracts such high-level, abstract features from raw data. Many of these factors of variation can be identified only using nearly human-level sophisticated understanding of data. A quintessential model of deep learning is multi layer perceptron (MLP) [4] which is a feedforward artificial neural network. It is a mathematical function which mapping some set of data as input to expected format of output values. The function is formulated by composing numbers of sub functions. The idea of learning the correct representation of data provides a perspective on deep learning. Another point of view about deep learning is to enable computers to learn multi-step computer programs. After executing another set of instructions in parallel, each presentation layer can be considered the state of the computer's memory. Networks with greater depth can execute more instructions in sequence. Sequential instructions have a strong function because subsequent instructions can refer to the results of earlier instructions. According to this deep learning perspective, not all information in a layer of activation necessarily encodes a variation factor that interprets the input. The representation also stores state information that helps to execute programs that understand the input. This status information can be similar to a counter or pointer in a conventional computer program. It has nothing to do with the input, but it helps the model organize its processing. Over the past decade, there are plenty of deep learning models hugely influential to the development of the field. Among them, convolutional neural network(CNN) [5] [6] a special kind of deep neural network architecture that capable of achieving record-breaking performance using both supervised learning and unsupervised learning. The convolution kernel sharing parameters and the sparseness of the inter-layer connection in the hidden layer make the convolutional neural network smaller. There are large numbers of powerful CNN architectures which laid the foundation of today's computer vision achievements, such as LeNet-5 [7], AlexNet [8], VGGNet [9], GoogLeNet [10], ResNet [11], etc.

Other than designing a proper deep learning neural network which can output multiple task results that including classification and regression, realizing the idea of how to learn from labeled and unlabeled data is the biggest challenge. Traditional classifiers usually need enough number of labeled data to train. However, in the real world, obtaining the labeled instances which composed by feature and label pairs are often expensive or time-consuming. On the other hand, unlabeled data is relatively cheaper to collect, but having lack of usage for analyzing features. Presently consistency regularization [12] [13] one of the most successful approaches which the model is trained to be robust to perturbations of its inputs and parameters.

3 Hypothesis

I will be responsible for designing an auto-augmentation process to enlarge the existing dataset without labeling the extra image. Trying to utilize the consistency smoothness enforcing [14] [15] [16] [17] [18] method to realize semi-supervised learning in this case. For the sub output of data regression for the coordinates, although, the manual labeling stage is not required, the pixel updated (the value for delta of x and y) should be able to traced back while randomly translating or cropping the image in order to update the ground truth of the coordinates. The semi-supervised learning approach will enforce the smoothness of the model by utilizing unlabeled data. Minimizing the divergence metric between the prediction of original input and the input with perturbation. Creating a deep learning model that can use an image captured from the camera as the input to generate the desired coordinates. Applying deep learning method to achieve the detection of a single point coordinates from the image which is a gray scaled high definition (1280 * 1024) image containing a laser stripe. The coordinates generated from the network should be as close as the coordinates calculated based on the existing algorithm which based on computer vision algorithm. In addition, the network should be capable for stainless materials which caused the current algorithm failed. The desired output should including both x and y coordinates and a softmax result to indicate the category of the image.

4 Conclusion and Justification

The outcome of this research project will aim to use a semi-supervised machine learning approach which will satisfy all the basic requirements to replace the existing traditional computer vision method. Some new features will be added to the new model, such as welding type classification, etc. The project will provide an opportunity for the research on consistency-based semi-supervised learning for regression and classification. If the final goal can be achieved, the new model can save a lot of time and cost on manually labeling the data. It can become a better solution for propotion of the industry.

References

1. Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. The MIT Press, 2016.
2. Ekin Dogus Cubuk, Barret Zoph, Dandelion Mané, Vijay Vasudevan, and Quoc V. Le. Autoaugment: Learning augmentation policies from data. *CoRR*, abs/1805.09501, 2018.
3. Waseem Rawat and Zenghui Wang. Deep convolutional neural networks for image classification: A comprehensive review. *Neural Computation*, 29:1–98, 06 2017.
4. M.W Gardner and S.R Dorling. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric Environment*, 32(14-15):2627–2636, aug 1998.
5. Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, Nov 1998.

6. Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012.
7. Jiuxiang Gu, Zhenhua Wang, Jason Kuen, Lianyang Ma, Amir Shahroudy, Bing Shuai, Ting Liu, Xingxing Wang, and Gang Wang. Recent advances in convolutional neural networks. *CoRR*, abs/1512.07108, 2015.
8. Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. *Commun. ACM*, 60(6):84–90, May 2017.
9. Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv 1409.1556*, 09 2014.
10. Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott E. Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. *CoRR*, abs/1409.4842, 2014.
11. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015.
12. Ben Athiwaratkun, Marc Finzi, Pavel Izmailov, and Andrew Gordon Wilson. Improving consistency-based semi-supervised learning with weight averaging. *CoRR*, abs/1806.05594, 2018.
13. Carlo Ciliberto, Alessandro Rudi, and Lorenzo Rosasco. A consistent regularization approach for structured prediction. *CoRR*, abs/1605.07588, 2016.
14. Philip Bachman, Ouais Alsharif, and Doina Precup. Learning with pseudo-ensembles. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 3365–3373. Curran Associates, Inc., 2014.
15. Samuli Laine and Timo Aila. Temporal ensembling for semi-supervised learning. *CoRR*, abs/1610.02242, 2016.
16. T. Miyato, S. Maeda, M. Koyama, and S. Ishii. Virtual adversarial training: A regularization method for supervised and semi-supervised learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(8):1979–1993, Aug 2019.
17. Kevin Clark, Minh-Thang Luong, Christopher D. Manning, and Quoc V. Le. Semi-supervised sequence modeling with cross-view training. *CoRR*, abs/1809.08370, 2018.
18. Antti Tarvainen and Harri Valpola. Weight-averaged consistency targets improve semi-supervised deep learning results. *CoRR*, abs/1703.01780, 2017.