IJCSE International Journal of Computer Sciences and Engineering Open Access
Review Paper | Vol.-7, Issue-6, June 2019 | E-ISSN: 2347-2693

Convolution Neural Network Based Enhanced Learning Classification Using Privileged Information

Hardikkumar M. Dhaduk[1*], Mallikarjuna Shastry P.M[2]

[1,2] Department of Computing and Information Technology, REVA University, Bangalore, India

Corresponding Author:dhadukhardik111@gmail.com

DOI: https://doi.org/10.26438/ijcse/v7i6.353357 | Available online at: www.ijcseonline.org

Accepted: 13/Jun/2019, Published: 30/Jun/2019

***Abstract*—** The accuracy of data-driven teaching methods is often unsatisfactory when training data are insufficient either in amount or quality. Usually incorporate privileged information (PI), tags, properties or attributes manually labeled to improve the learning of classification. The manual labeling process, however, takes time and works intensively. In addition, manually labeled privileged information may not be rich Sufficient due to personal knowledge limitations. In this approach, classifier learning is enhanced by exploring untagged corporate privileged information (PI), which can effectively eliminate reliance on manually labeled data and enhance privileged information. We treat each selected privileged information as a subcategory in detail and for each subcategory we learn one classifier independently. Classifiers are integrated for all sub-categories to form a more powerful category classifier. In this CNN classifier approach, in particular, to learn the optimum output based on the pictures chosen. The superiority of this proposed approach is demonstrated by extensive experiments on two benchmark data sets.

***Keywords*—** Untagged corpora, Transfer learning, privileged Information, Neural network

I. INTRODUCTION

A big number of computer vision problems can be transformed into classification problems [1],[5],[6]. Data-driven approaches to classification learning have been predominant in recent decades [7],[8],[9],[12]. A set of training samples (x1, y1) … (xn, yn) teaches the classifier. Data-driven methods, despite the achievement attained, become very brittle and susceptible to overfitting when either the amount or quality of training data is insufficient. Unfortunately, in many real-world applications, this is often the case. To relieve this restriction, it is a natural answer to incorporate extra privileged information [13][14]. The learner, for example, can also leverage object attributes (e.g. "cheese" and "veg") as well as Image functions and labels in the training phase (e.g. "hamburger"). In recognition of human action, in addition to the RGB features and human action labels, human common positions can be incorporated into the classifier training. In practice, the tags, characteristics, attributes, positions, or pictures of internet or data set context may be the privileged information.

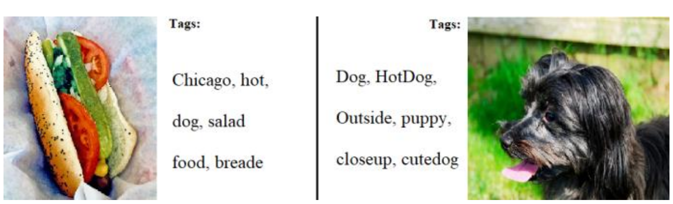

Fig. 1: Textual tags (privileged information) are included for "Flickr" Website for image sharing. The tags (metadata) are both useful and noisy.

Learning classification with privileged information, however, is a challenging issue. There are three aspects to the difficulty. First, it is very expensive to manually label privileged information. Secondly, it is available only during training and during testing it is not seen. In order to predict the category label, we cannot combine the privileged information with input features. Third, the quality of the PI collected is excessively dependent on PI learning classifiers. Flickr images on the website tend to have multiple text metadata, as shown in Figure 1. In practice, these textual metadata can be associated with noise. If we failed to remove noise and could get even worse in extreme times, the accuracy and robustness of the learned classifier would be significantly reduced.

Extracting and leveraging useful privileged information in this approach to improve classifier learning with privileged information for learning better object recognition systems: attributes, annotator rationale, bounding boxes for objects, and text descriptions. This approach extracts privileged data from untagged corpora that find privileged data from manually marked descriptions as opposed to prior works. We also concentrate on encoding privileged data in the classifier framework during training, which is distinct from prior works that generally encode in the classifier parameters privileged data during training. Consider a critical issue primarily in this work. The problems are generally noisy privileged data from untagged corporations, how to get cleared of noise, select helpful privileged data, purify noise, and select helpful photos to learn robust classifiers.

II. RELATED WORK

This method is associated with latest works of image sub-categorization [15], which assume that in each category there are several sub-categories. However, only the visual features were used by the subcategories discovered by [15][16], the subcategories discovered have the category information. Information about the subcategory of refinement remains unavailable. Ristin et al. [17] The Random Forest Framework has been adopted and a regular objective function has been proposed to improve classification accuracy, taking into account the relationship between categories and subcategories. Method may classify images into subcategories as opposed to previous works, but it must rely on WordNet expert knowledge manually labeled [18] to obtain information in the subcategory of semantic refinement. We eliminate dependence on manually marked data in our job and suggest that untagged corpora to extract textual information.

This approach has to do with semi-supervised methods [19], poorly supervised methods [20], and supervised methods [21]. A various regularized multitask learning algorithm was suggested by Luo et al. [22]. By exploiting the common framework of these tasks, it learns a discriminatory subspace shared by various classification functions. In [23], Doulamis et al proposed a semi-supervised deep learning paradigm for object classification and monitoring. This method addresses the main deep learning issues by allowing non-supervised data to initially configure the network. Carneiro et al. suggested a probabilistic approach for the annotation and retrieval of semi-conductor pictures in annotation and retrieval as classification problems where each class is described as a set of pictures with a common semi-conductor label. Gao et al. [20] Proposed leveraging of picture attributes to weakly monitor the dictionary learning process without actually needing labels. In Joulin et al., huge, weakly labeled picture collections were leveraged to know excellent visual features. To discard outliers of images, the writers suggested a fresh content-based picture filtering algorithm. The strategy provided features and mixed two unexpected methods of clustering DBSCAN and spectral clustering. DBSCAN algorithm is used to extract outliers from the collected information set and spectral clustering distinguishes noise-free picture information set in separate classifications each representing distinctive geometric views of cultural heritage artifacts.

This strategy is nearer to teaching techniques that use privileged data to change the learning of classifiers [13],[14]. Li et al. [14] suggested a technique for categorizing pictures by integrating textual features (obtained from the text descriptions that surround them) and dealing with noise in loose picture tags at the same time. Similarly, method [13] and embrace various kinds of privileged data to improve the learning of classifiers. During practice, all techniques in [13] encode privileged data in the parameters of the classification. The disadvantage is that the quality of the privileged information gathered is overly dependent on these methods. Due to the complexity of the Internet, it is difficult to select helpful privileged information from the text descriptions that surround it, with a lot of noise involved. The precision of the learned classifier will be significantly decreased if we have failed to during exercise, filter out the noisy privileged information.

III. METHODOLOGY

Our proposed strategy primarily comprises of three main phases

A. *Discovering Privileged Information*

Recent research inspired [Divvala et al., 2014], this approach uses Google Open-Image Corpora [Lin et al., 2012] Discover a comprehensive Explanatory vocabulary all the category-specific Variations in shape. Parts-of-speech (POS) gram data are treated specifically as privileged information. For instance, given a category (e.g., "hamburger") and its respective portion of the speech tag (e.g., ' Cheese, VERB '), all its events are noted. in the attributes with part of the speech tag, rational annotator, bounding object boxes and text descriptions. Of all the occurrences that were retrieved for that category.

B. *Purifying PI (Privileged Information)*

Not all privileged candidates for data are helpful, but some noise may also be included. It will harm both precision and robustness by using the noisy PI to improve classifier teaching. To this end, prior to studying classifiers, we need to distinguish helpful PI from noise from the appropriate view, our fundamental concept is to filter the noisy PI.

C. *Learning Integrated Classifier*

Each selected privileged information will be treated as a sub-category for the target category in this part. Suppose we get M subcategories; we collect the top few candidate pictures from open images for each subcategory. Although the image engine rated the transferred images, due to the accessible image error index, some noisy images may still be included. We need to prune noise before learning an integrated classifier.

Learning using privileged information

Given a set of vector-represented N training examples, we assume a supervised situation of binary classification X= {x1... xN} rollover Rd and the mark Y= {y1...yN} The task is to learn the prediction function F: X= R from space F of possible functions, e.g. all linear classifiers. For instance, bag-of-visual-words histograms [6], we will consider the examples as images and their representation as calculated from the content of the image.

Adopting the privileged information setting, we receive additional information with regard to the training set, we also assume that this is a Characteristic of vectors, $X^* = \{x^*_1 \dots x^*_N\} \subset X^* \subset R^{d^*}$, where any x^*_i unaffected Encodes additional x_i information we have. Note that we do not make any additional assumptions about this privileged data. In particular, x^*_i Perhaps from the original image it is not computable, but it reflects a very different type of information. In general, X* will also be different from X, for example, X cannot be applied to X * or vice versa for defined functions.

The purpose of privileged information is to use privileged data, X* to learn a better classifier than to learn without it. It is clear that f: X → R itself cannot rely on the X* domain because at the time of the test this is not available. Therefore, this is influenced by privileged information, it must be our choice f ∈ F

Fig. 2: Three special kinds of privileged data that can assist you learn to understand objects better - attributes, bounding boxes for annotator objects, and text descriptions.

Figure 2 illustrates the three modalities. Additional sources of information are available. This privileged data helps us to differentiate in the training set between simple and hard examples. This knowledge enables us to identify and focus the learning step towards suitable components of training data in order to find a higher predictive quality function.

IV. CONVOLUTION NEURAL NETWORK COMPONENTS

Convolution types of neural network layers mainly include three different types, first layer of Convolution, second layer of pooling, and third layer of fully connected layer. Figure 3 Shows Yann LeCun's introduced architecture of LeNet.

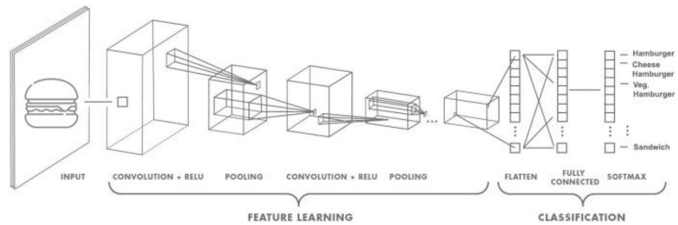

Fig. 3: LeNet network architecture

A. Convolutional Layer

Convolution layer is at the heart of the Convolution Neural Network, sharing characteristic weights and local connections. Each neuron of the same function map is used to get local features from different locations in the former layer, but their extraction in the former feature map is local features of the same positions for single neurons. The maps of the input function are first converted to a learned kernel for a new feature and the results are then transferred to a nonlinear activation function. By applying different kernels, we will obtain different feature maps. Basically, therefore, provide the core of the Convolution Neural Network with open images.

B. Pooling Layer

The sampling method is similar to fuzzy filtering. The impact of pooling layer is to obtain the secondary function, decrease the feature maps size, and boost the extraction feature's robustness. It is generally positioned between the two layers of Convolution. The move phase of the kernel determines the size of the pooling layer feature maps.

C. Fully-connected Layer

Usually one or more fully linked layers are the Convolution neural network classifier. In the previous layer, they take all the neurons and connect them to each current layer of neurons. There are no conserved spatial data in fully linked layers. A layer of production follows the last completely linked layer. SoftMax regression is commonly used for classification assignments as it generates a well-performed likelihood output distribution. Another frequently used method is SVM, which can be combined with CNNs to solve different classification tasks.

V. EXPERIMENTAL SETUP

We are learning three types of privileged information in this experimental setting approach, all of which can be handled in a unified structure where beforehand hand-crafted techniques were used. This approach considers annotation attributes, bounding box annotation, text overview and rationale as Sources of privileged information (PI) during training, but not during testing. As we will see, some ways of transferring the rank are more appropriate than others. In the subsequent subsections, we will discuss this. Note that where privileged information is useful, we also include results. The reason for this is to show, in addition to scientific honesty, that there is no negative transfer.

Methods: Using privileged information, we're looking at two learning methods: our proposed transfer learning technique and the CNN method based on PI. Compare the results with the CNN ranking when you learn directly on the original space X (CNN Pooling). Also provide the performance of CNN as a reference in the privileged X space, as if during training we had access to privileged information.

Evaluation accuracy: Our model is unconfident when classifying hamburgers (class 0), but confident when classifying hamburgers (class 1). In fact, 90% of the images classified as hamburgers are hamburgers. But 92 percent of all actual hamburgers are properly classified. These results are in line with model accuracy of 75 % and model loss of 52 % as we can see from Figure 4.

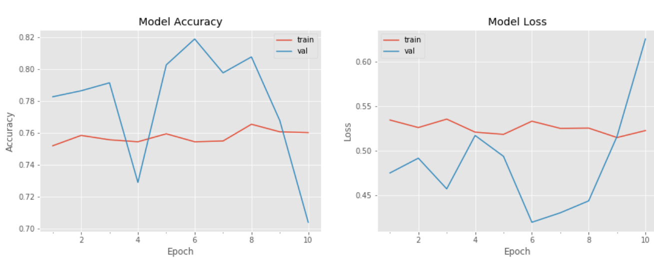

Fig. 4: Model accuracy and loss

A. *Attributes as PI (privileged information)*

Attribute annotation includes a description of the semantic properties of various objects such as shape, color etc. We use the data set of Open-Images v4. We depend on the default 2 test classes in combination with the dataset for which the annotation attribute is provided. With a total of 2,000 training images, the 2 classes are sandwich and hamburger. For model, the values of the expected characteristics are acquired and match the estimates of binary attribute probability in the pictures.

B. *Textual description as PI (privileged information)*

A text description offers an additional perspective for an object's visual depiction. This can be used as privileged data in object classification tasks. The hamburger group has more than 7 classes: it has a total of 2643+ images, there are many images attached to each image with a text description. There is a comparatively tiny number of samples per class, around 250 samples. Text representation of data in bag-of-words as privileged information. For training, we use 250 photos per group. We're repeating the 10/Epoch split train.

C. *Bounding box as PI (privileged information)*

The bounding box for annotation is designed to capture the e xact location of an object in the image. Knowing the exact pl ace of the object in the training data is privileged information when performing object recognition of picture level. We use a subset of train-annotations-bbox classifications available f or annotation bounding box. Using nine million open images, we define 600 classes. The group of 6oo classes has 9 millio n open images: Soccer, toy, cat, vase, hair dryer, kangaroo, k nife, pencil case, tennis ball, high heels, sushi, tree, truck, vio lin, wine, wheel, pizza cutter, bread, lemon, dog, elephant, fl ower, furniture, airplane, spoon, swan, peanut, camera, flute, helmet, crown, etc. With too little bounding region, we ignor e few images and use 2643 images for further analysis. There are more than 7 classes in the hamburger group: it has a total of 2643 + images. Ignore uninformative annotation box ima ges and instead use other images. Use flicker data to remove boundary box areas and use the former to represent privilege d information data directly. We use 186 pictures from the req uired class and 186 randomly drawn samples to train from th e remaining classes. The remaining pictures are used for testi ng from the other categories in the required class and the sam e quantity. To obtain better performance statistics, we repeat the 10 Epoch train / test split.

As shown in Figure 5, it is useful to use Annotation of the boundary box as privileged classification information. Training loss is also very low and the accuracy of the data set is very high, so classification is more powerful. The PI methods in both spaces can take advantage of simple and hard samples. Credit this because in this experiment both spaces are of the same modality, i.e. privileged information is obtained from a subset of the same image properties used to represent the original data. Consequently, our fundamental hypothesis is fulfilled that in both modalities the same instances are easy and difficult.

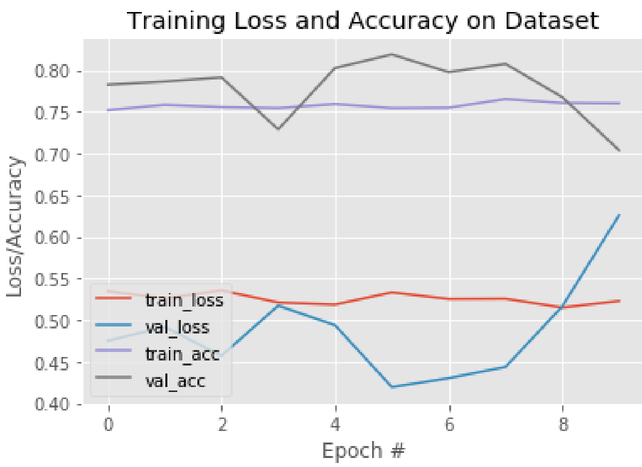

Fig. 5: Training loss and accuracy on dataset

VI. CONCLUSION

The studied extracting privileged information to improve visual object classification tasks classification learning. This approach has shown how it can be applied to a number of situations where separate methods were previously hand-crafted. This approach experiments show that prediction performance is often enhanced by using the proposed transfer learning technique and privileged information based enhance classification learning using convolution neural network (CNN).

REFERENCES

[1] S. Andrews, I. Tsochantaridis, and T. Hofmann. Support vector machines for multiple-instance learning. Advances in Neural Information Processing Systems, 561–568, 2002.

[2] A. Bergamo and L. Torresani. Exploiting weakly-labeled web images to improve object classification: a domain adaptation approach. Advances in Neural Information Processing Systems, 181–189, 2010.

[3] L. Niu, W. Li, D. Xu, and J. Cai, “Visual recognition by learning from web data via weakly supervised domain generalization,” IEEE Transactions on neural networks and learning systems, 22(9), 1985–1999, 2017.

[4] R. C. Bunescu and R. J. Mooney. Multiple instance learning for sparse positive bags. International Conference on Machine Learning, 105–112, 2007.

[5] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(9): 1627– 1645, 2010.

[6] X. Hua and J. Li. “Prajna: Towards recognizing whatever you want from images without image labeling,” AAAI International Conference on Artificial Intelligence, 137–144, 2015.

[7] M. Jager, C. Knoll, and F. Hamprecht, “Weakly supervised learning of a classifier for unusual event detection,” IEEE Transactions on Image Processing, 17(9): 1700–1708, 2008.

[8] J. Fan, Y. Shen, C. Yang, and N. Zhou, “Structured max-margin learning for inter-related classifier training and multilabel image annotation,” IEEE Transactions on Image Processing, 20(3): 837–854, 2011.

[9] Y. Li, S. Wang, Q. Tian, and X. Ding, “Learning cascaded shared-boost classifiers for part-based object detection,” IEEE Transactions on Image Processing, 23(4): 1858–1871, 2014.

[10] A. Coates, H. Lee, and A. Y. Ng. An analysis of single-layer networks in unsupervised feature learning. Ann Arbor, 1001(481): 2, 2010.

[11] S. Andrews, I. Tsochantaridis, and T. Hofmann. Support vector machines for multiple-instance learning. Advances in Neural Information Processing Systems, 561–568, 2002.

[12] F. R. Bach and Z. Harchaoui. Diffrac: a discriminative and flexible framework for clustering. Advances in Neural Information Processing Systems, 49–56, 2008.

[13] Z. Wang and Q. Ji, “Classifier learning with hidden information,” IEEE International Conference on Computer Vision and Pattern Recognition, 4969–4977, 2015

[14] W. Li, L. Niu, and D. Xu, “Exploiting privileged information from web data for image categorization,” European Conference on Computer Vision, 437–452, 2014.

[15] M. Hoai and A. Zisserman. Discriminative sub-categorization. IEEE International Conference on Computer Vision and Pattern Recognition, 1666–1673, 2013.

[16] X. Wang, B. Wang, X. Bai, W. Liu, and Z. Tu. Max-margin multipleinstance dictionary learning. International Conference on Machine Learning, 846–854, 2013.

[17] M. Ristin, J. Gall, M. Guillaumin, and L. Van Gool. From categories to subcategories: large-scale image classification with partial class label refinement. IEEE International Conference on Computer Vision and Pattern Recognition, 231–239, 2015.

[18] G. A. Miller. Wordnet: a lexical database for english. Communications of the ACM, 38(11): 39–41, 1995.

[19] Y. Luo, D. Tao, B. Geng, C. Xu, and S. Maybank, “Manifold regularized multitask learning for semi-supervised multilabel image classification,” IEEE Transactions on Image Processing, 22(2): 523–536, 2013.

[20] Y. Gao, R. Ji, W. Liu, Q. Dai, and G. Hua, “Weakly supervised visual dictionary learning by harnessing image attributes,” IEEE Transactions on Image Processing, 23(12): 5400–5411, 2014.

[21] N. Doulamis and A. Doulamis, “Semi-supervised deep learning for object tracking and classification,” IEEE International Conference on Image Processing, 848–852, 2014.

[22] G. Carneiro, A. Chan, P. Moreno, and N. Vasconcelos, “Supervised learning of semantic classes for image annotation and retrieval,” IEEE Transactions on Pattern Analysis and Machine Intelligence, 29(3):394–410, 2007.

[23] Y. Gao, R. Ji, W. Liu, Q. Dai, and G. Hua, “Weakly supervised visual dictionary learning by harnessing image attributes,” IEEE Transactions on Image Processing, 23(12): 5400–5411, 2014.

Authors Profile

Hardikkumar M. Dhaduk Currently pursuing M. Tech in Data Engineering and cloud computing final year at REVA University, Bangalore, B. E. in Computer Engineering from Om Engineering College, Junagadh, Gujarat under Gujarat technological university and Diploma in Computer Engineering under Gujarat technological university.

Dr. Mallikarjuna Shastry P.M holds Ph. D. on Fault Tolerance in Distributed Systems from VTU, M. Tech. in CSE from MSRIT and B. E. in CSE from Karnataka University Dharwar. He has totally 23 years of vast experience in teaching in CSE/ISE. He has published and presented several papers in International Journals and Conferences respectively. His areas of Interest are, Distributed Computing, Cloud Computing, Computer Architecture, Information and Networks Security. Image processing. Currently he is guiding two candidates for Ph. D. under VTU.