Weakly Supervised Image Classification using GANs

Jordan Axelrod
Dept. of Computer Science
Duke University
jordan.axelrod@duke.edu[*]

Rathi Kashi
Dept. of Computer Science
Duke University
rathi.kashi@duke.edu

Rajas Pandey
Dept. of Computer Science
Duke University
rajas.pandey@duke.edu

Abstract

Image classification tasks today require large labelled datasets to train which are expensive to obtain. In some domain specific applications like medical images, it might not be possible to have large datasets. Our paper addresses this problem through the lens of Weakly Supervised Learning, which requires only a small sample of clean labelled images for classification. The lack of data problem is addressed by having a GAN architecture, which both augments the training sample as well as regularizes the clean network. Our model achieves a top 1 accuracy of 89.86% using only 50% of the clean training sample of FashionMNIST, which is comparable to published models. We also present findings on image generation through a Weakly Supervised GAN architecture, which can be used to increase the training sample for domain specific applications.

1 Introduction

Computer vision tasks such as object recognition and image classification require annotated data which is difficult to obtain. Thus, many papers still rely on datasets like ImageNet which are relatively small by modern standards [8]. The challenge with larger datasets is the cost of annotation is very high. For this reason, models which can learn the underlying structures of a large dataset from a smaller sample of correctly labelled data are becoming popular in the recent times [3] [4].

Neural networks also require a large amount of data to learn and generalise to data outside the training data and give meaningful insights. This becomes a constraint in some applications such as medical imaging where the cost of generating large datasets is extremely high. It is also constrained by the standard way of training a neural network, which requires labels for all the data used in the training process. For images in fields like medical imaging, this labelling process involves categorising data or segmenting objects within the images which requires expertise [10].

Our paper discusses two approaches of addressing these challenges. First, we discuss weakly supervised image classification using only a small subset of correctly labelled images. Based on the work of Hu et al [4], we artificially introduce noisy labels which help regularise the network which learns the clean labels. This is done by having a two headed classification network consisting of a clean and residual net. The clean net learns the mapping to correct labels while the residual net learns the mapping between the residual of the correct and noisy labels. We discuss the performance of this model over the FashionMNIST dataset.

Then, we discuss a technique of artificially enhancing the size of our dataset using only a small number of real images. Using a technique based on Mao et al [10], we train a GAN to generate images from FashionMNIST dataset. We also train a classifier which works in conjunction with the discriminator to identify both the class of the image and whether it is real or fake.

Leveraging the complimentarity between the architectures, we propose a model which replaces the residual network in noise regularized classification with a GAN supervised model. The clean network

[*] The authors would like to thank Sayan Mandal for useful comments at various stages of our work.

then gets to learn not only the structure of the exiting training data but also some new generated data. With this regularization based on real clean data, real clean data with noisy labels, and fake generated data with noisy labels, our hypothesis is that the clean network will learn to classify the images better.

Our proposed architecture can leverage a small percentage of clean data to train a robust classifier, leveraging information from noisy data and new generated data. We report metrics like accuracy and average precision for different levels of clean labels in the dataset. The main contribution of this paper is the novel architecture which leverages a pre-trained and fine-tuned GAN architecture to regularize a classification network. In addition, we also provide insights on modelling and training a GAN network which can act as a classifier using only a small sample of real images.

1.1 Motivation and importance of the problem

Most methods of computer vision rely on fully supervised learning, which requires detailed and correct classification labelling, annotations using bounding boxes or pixelwise segmentation [2]. These methods are known to be expensive, time-consuming and error prone. In addition, the models under such constraints only learn the most discriminative parts of the images and ignore other features which might provide complementary information [3].

In addition, we have discussed the problem of lack of availability of large datasets for domain specific applications such as medical imaging. Thus, the main motivations of this problem are the challenges that 1) domain specific data may not be available in large quantities, and 2) annotating and labelling large datasets is an expensive process. To address these challenges, we utilise weakly supervised learning which uses small subset of labelled data and a large corpus of unlabelled data.

2 Related works

Our work is based on two papers whose results we replicate. The first paper is by Hu et al [4] which proposes an end-to-end trainable pipeline which makes use of information from the noisy labels. The model does not rely on putting assumptions on the nature and distribution of class labels, as has been the standard practice in weakly supervised classification [5], [12]. This work is similar to DivideMix [7] model where the objective was to avoid overfitting on the noisy data using information from the clean dataset. In this paper, we use a similar architechture but knowledge from noisy data to avoid overfitting the clean network.

The second implements what the authors call WSGAN [9]. This Generative Adversarial Network uses a generator network to produce new images in the same distribution as the small labelled training set to create a classifier for unlabelled images. This model has uses in domains where it is difficult to get high quality large data sets. It improves on the classification accuracy by data augmentation through generation. The loss function in this paper is motvated from the cyclic consistency loss in the CycleGAN model [1].

Weakly supervised classification broadly falls under the categories of incomplete , inexact and inaccurate supervision [14]. The noise regularization model falls under the category of inexact supervision as we rely on imperfect labels to improve our classificaiton [13]. The GAN is trained as an incomplete supervision process, because we rely only on a small amount of labelled data to create that classifier.

The GAN model is also trained based on the idea of pseudo-labelling [6] where some labels are first assigned to the unlabelled data. Then the classifier learns to identify the correct classes using some labelled data and the pseudo-labels. This achieves a better performance compared to just the labelled data.

3 Details of the project

3.1 Noise regularization model

This model takes in a batch of noisy and clean data in a 9:1 ratio. This batch is fed into a backbone CNN (which in our model is ResNet 18). The CNN extracts features from the input and feeds the noisy points to the residual net and the clean points to the clean net. The noisy and clean net have the

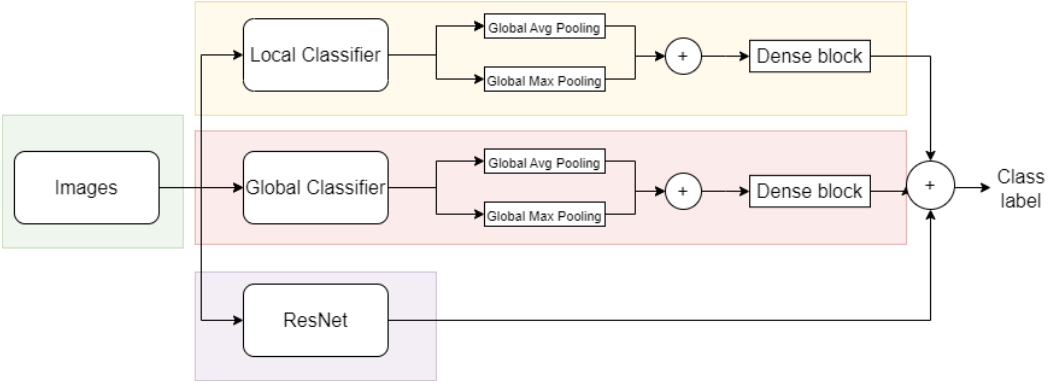

Figure 1: The architecture of the discriminator. This consists of three separate classifiers looking at the image in different ways.

same architecture consisting of a linear layer, an activation function and a linear layer, where the final feature size is reduced to 10. The noisy and clean points are added and passed through a sigmoid for classifier that is supervised by the noisy points. The clean points are passed through a sigmoid supervised by the clean data points. The trained clean net serves as the main classifier for test points while the residual net is the regularization for the network. The loss functions for the residual net and the clean net are as follows (cross entropy loss):

$$L_{noise} = \frac{-1}{N_n} \sum_{i \in D_n} (y_i \ln h_i + (1 - y_i) \ln(1 - h_i)) \tag{1}$$

$$L_{clean} = \frac{-1}{N_c} \sum_{j \in D_c} (v_j \ln g_j + (1 - v_j) \ln(1 - g_j)) \tag{2}$$

where D_n are the noisy data points, D_c are the clean data points, y_i are the noisy labels, v_j are the clean labels and these labels supervise the residual net h and the clean network g respectively.

3.2 Weakly supervised GAN model

The Discriminator implemented in [9] is able to easily identify fake images from real images. It uses three separate classifiers and uses all of their outputs to classify images. Two of the classifiers, the local and global classifiers, utilise attention mechanisms to classify images. The classifiers use pixel and spatial attention. Pixel attention pays attention to the different channels of each pixel, and spatial attention looks at the attention between each pixel. Convolutional blocks are used to down sample the images. The global classifier consists of two of these convolutional and attention layers, while the local has only one. Figure 2 shows the architecture of these classifiers.

Since the global classifier is deeper it is looking at more interactions between the entire image, while the local classifier looks at local interactions between pixels and channels. The last classifier is a Resnet18 that was pretrained using contrastive learning, leading it to be able to tell the difference between the classes of the images. This high powered classifier is easily able to classifiy real and fake images early on in training, this leads to gradient saturation when training the Generator in this architecture. It is very difficult to train a good generator in these conditions. To alleviate this issue we pretrain a generator with a simple discriminator consisting of 4 convolutional layers and a linear output layer. As the authors point out, it is difficult to train a GAN on both class and real/fake classification at the same time. Training in this way generally produces poor images. The authors prescribe a MSE loss between the real and fake flags. We found this to be insufficient to train the Generator well. Instead, we use a Wasserstien loss function with gradient penalization, WGAN-GP, to train the Generator in the first step of training. Figure 3 shows the images generated after 100 epochs.

Using this pretrained generator we train the Classifier to identify real and fake images as well as the class of the real images. We use two loss functions a supervised and unsupervised function.

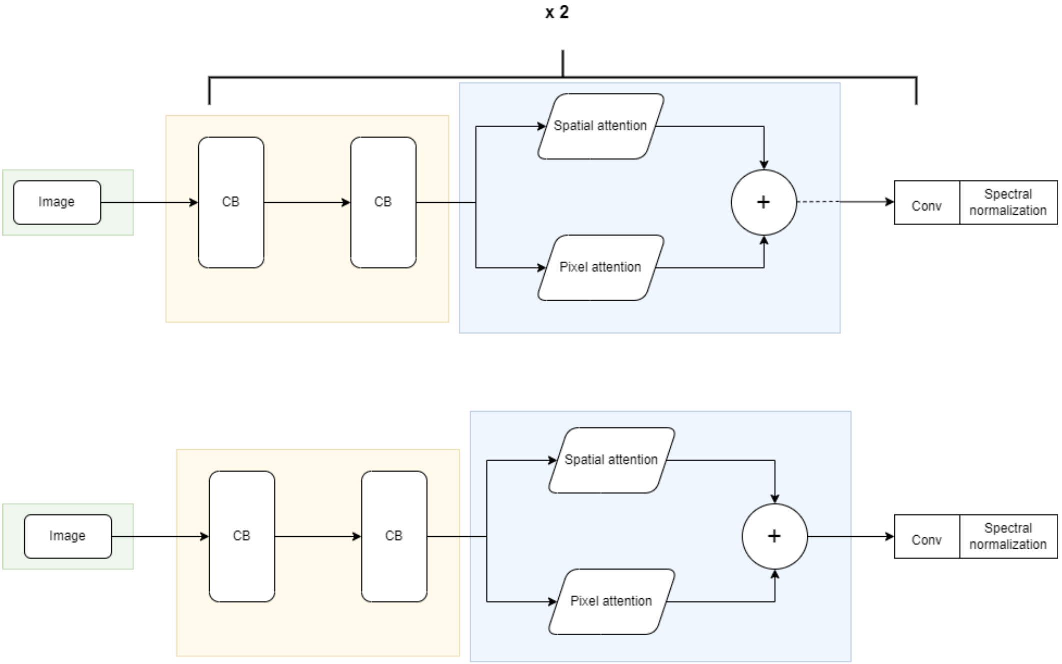

Figure 2: The architecture of the global and local classifiers. These consist of down sampling convolutional layers feeding into two separate attention mechanisms. The global classifier contains two of these blocks.

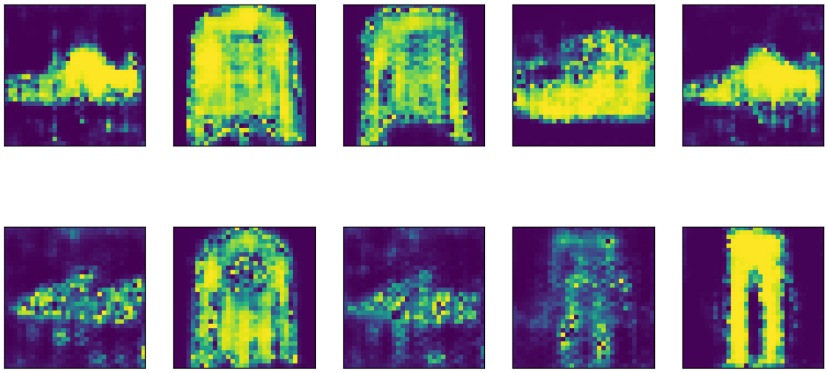

Figure 3: Generate images on the fashionMNIST Dataset after 100 epochs of training

$$L_{supervised} = -\,\mathbb{E}_{x,y\sim p_{data}(x,y)} \log D_1(x)$$

This function is the Cross Entropy loss of the output labels of the real images. The second loss function classifies real and fake images.

$$L_{unsupervised} = \mathbb{E}_{x\sim p_x}(D_2(x)-b)^2 + \mathbb{E}_{z\sim p_z}(D_2(G_{best}(z))-a)^2$$

Where a and b are the labels we want the classifier to produce. While training the Discriminator, it can again easily classify real and fake images after few iterations of training. Once this happens we update the generator with this fixed new classifier.

3.3 Proposed model

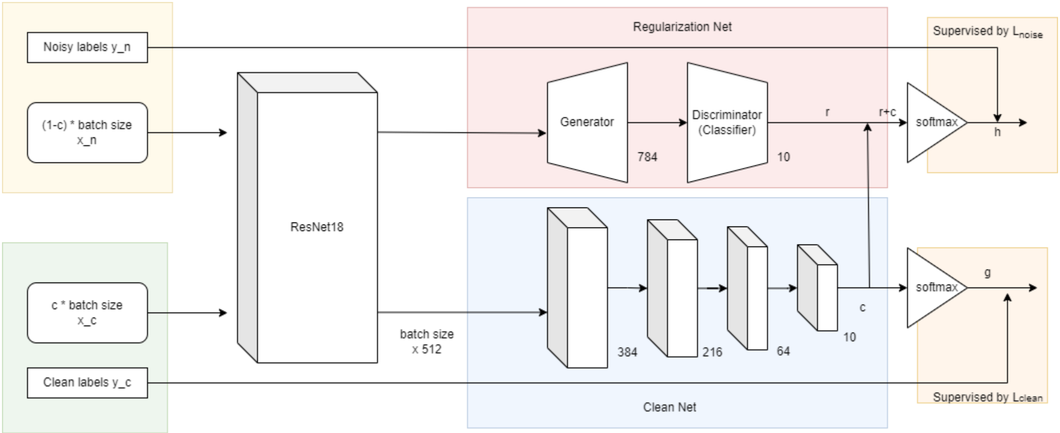

Figure 4: The architecture of our proposed model. The residual net of the Noise Regularization model is replaced with a Generator - Discriminator pair which acts as a regularizer. The numbers beside the layers represent the size of outputs from these layers.

Our proposed model architecture (Figure 4) is derived from the architecture of the noise regularization model. We modify the clean net structure to include more layers to learn the features better. We also replace the residual net of the original paper with our pre-trained Weakly Supervised GAN classifier.

The loss functions are the same as Equations 1 and 2 and are used to fine tune the GAN as well on the noisy labels. The overall loss function is

$$Loss = \alpha L_{clean} + L_{noise} \tag{3}$$

where α is a hyperparameter. We experimentally determined α to be 0.2, which means that loss is clean data is penalised a fifth of the loss in the noisy set.

The model is fed the feature embeddings generated from the fine tuned ResNet18 outputs. In this architecture, the model is getting regularizaed through both the noisy labels as well as generated fake images. Thus the clean network is more robust to noise in the data and trains to learn the important features from the clean data.

4 Experimental results

4.1 Dataset

We used the FashionMNIST dataset from the PyTorch library [11]. For the regularization process, we split the training data into certain percentages of clean and noisy labels. Noisy labels are created using a random permutation of the image targets.

For the weakly supervised GAN, we augment the training data using cropped images, adding jitter, flipping horizontally and vertically, resizing and normalizing the images. These augmented samples are used to train the generator and improve the generation output. The images are sent in batches where one clean image and it's transformed images pass through the network at the same time, so that the netowrk learns more robust features about the images.

4.2 Training procedure

We used a pre-trained ResNet18 model to generate features from images. We trained GAN models separately and used the best performing GAN model in our final regularization task.

We trained the WSGAN model in steps to get a final best GAN generator and classifier. We added it to the model and fine-tuned it further on the clean and noisy datasets to regularize the network.

Hyperparameters we used a batch size of 300, number of epochs was 10, learning rate and weight decay were both 0.0001. We set $\alpha = 0.2$ which is the penalty on the loss from the clean network. We used the Adam optimiser.

Finally, the data was tested on the whole 10,000 test samples of the FashionMNIST data and the metrics were reported.

4.3 Noise regularization model

We first implemented the standard noise regularization model with $\alpha = 0.2$ as mentioned in the[4] paper. The results on the top-1 accuracy for different levels of clean data are shown in Table 1. As we can see from the table, accuracy increases with increasing the percent of clean labels in the data, which is expected.

	Clean_pct	Accuracy
0	0.05	0.8373
1	0.10	0.8612
2	0.20	0.8714
3	0.50	0.88839996
4	0.70	0.88989997

Table 1: Top 1 accuracy as we change the percentage of clean labels in the Noise Regularization model with WSGAN

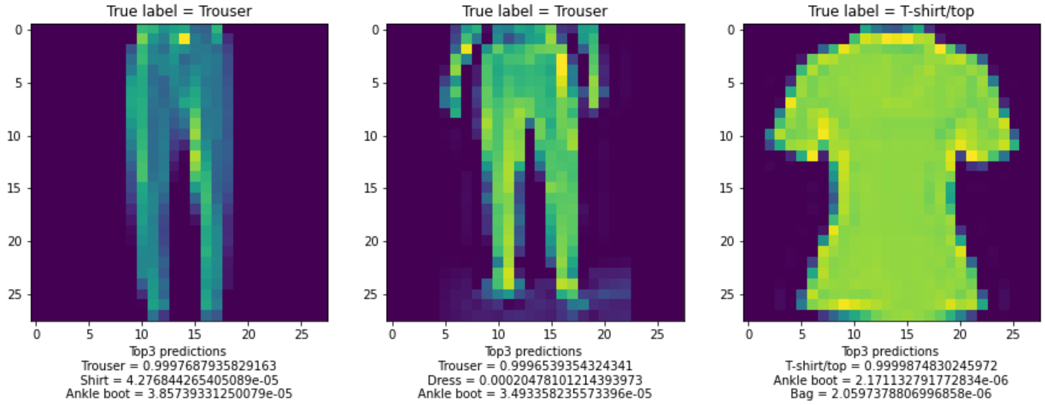

Figure 5: The top 3 predictions from Noise Regularization models along with the true labels.

To show the performance of our model, we present the top 3 predictions of the model on some samples from the test data in Figure 5.

4.4 Noise Regulartization with WSGAN

The accuracy metrics for our final model are presented in Table 2. The accuracy improves with increasing the size of clean data and our model outperforms the noise regularization model with over 50% of clean data. The generated images and top3 predictions on some samples are shown in Figure 6.

Our hypothesis is that as we increase the size of clean labels, our network learns to identify more features from the clean data. The GAN based regularizer is stricter than basic noise regularizer, and penalises performance both on noisy data from the training samples as well as generated samples. This helps the clean network to learn the features better.

	Clean_pct	Accuracy
0	0.05	0.778
1	0.10	0.8107
2	0.20	0.8293
3	0.50	0.8986
4	0.70	0.9068

Table 2: Top 1 accuracy as we change the percentage of clean labels

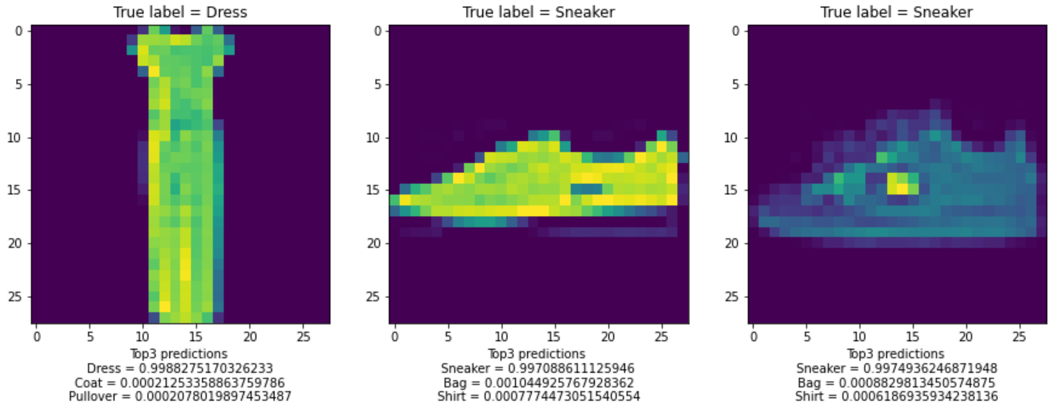

Figure 6: The top 3 predictions from Noise Regularization models along with the true labels.

5 Concluding remarks

Weak supervision is required because of lack of availablity of large, labelled datasets in specific domains. We address this challenge by proposing a netowrk which requires only a small percentage of clean labels to learn the classifier. This is achieved through a model which leverages noisy labels to regularize the training of the clean network. We enhanced this model through a generator to further use generated images for regularization of our model.

Overall, our implementations gave comparable results to state of art published models on the FashionMNIST dataset. At 50% clean labels, our model outperformed the original noise regularization model on the top 1 accuracy metric.

Further, the authors would like to study and quantify the effects of each of the components of the model through abalation experiments. The authors would also like to train the model further on other image datasets and compare performance against state-of-the art benchmarks.

Contributions

- Jordan Axelrod - Weakly Supervised GAN implementation, architecture design and writing
- Rathi Kashi - Noise Regularization model implementation, architecture design and writing
- Rajas Pandey - Combining NR model with WSGAN implementation, architecture design and writing

All authors had equal contributions in Conceptualization, Methodology and Investigation process.

References

[1] Amjad Almahairi, Sai Rajeshwar, Alessandro Sordoni, Philip Bachman, and Aaron Courville. Augmented cyclegan: Learning many-to-many mappings from unpaired data. In *International Conference on Machine Learning*, pages 195–204. PMLR, 2018.

[2] Ricardo Cabral, Fernando De la Torre, João Paulo Costeira, and Alexandre Bernardino. Matrix completion for weakly-supervised multi-label image classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(1):121–135, 2015.

[3] Weifeng Ge, Xiangru Lin, and Yizhou Yu. Weakly supervised complementary parts models for fine-grained image classification from the bottom up. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3034–3043, 2019.

[4] Mengying Hu, Hu Han, Shiguang Shan, and Xilin Chen. Weakly supervised image classification through noise regularization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11517–11525, 2019.

[5] Junjun Jiang, Jiayi Ma, Zheng Wang, Chen Chen, and Xianming Liu. Hyperspectral image classification in the presence of noisy labels. *IEEE Transactions on Geoscience and Remote Sensing*, 57(2):851–865, 2018.

[6] Dong-Hyun Lee et al. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks.

[7] Junnan Li, Richard Socher, and Steven C. H. Hoi. Dividemix: Learning with noisy labels as semi-supervised learning. 2020.

[8] Dhruv Mahajan, Ross Girshick, Vignesh Ramanathan, Kaiming He, Manohar Paluri, Yixuan Li, Ashwin Bharambe, and Laurens van der Maaten. Exploring the limits of weakly supervised pretraining. In *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018.

[9] Jiawei Mao, Xuesong Yin, Yuanqi Chang, and Qi Huang. Weakly-supervised generative adversarial networks for medical image classification. *CoRR*, abs/2111.14605, 2021.

[10] Jiawei Mao, Xuesong Yin, Yuanqi Chang, Qi Huang, Daoqiang Zhang, Jieyue Yu, and Yigang Wang. Weakly-supervised generative adversarial networks for medical image classification. *arXiv preprint arXiv:2111.14605*, 2021.

[11] Han Xiao, Kashif Rasul, and Roland Vollgraf. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms, 2017.

[12] Weining Zhang, Dong Wang, and Xiaoyang Tan. Robust class-specific autoencoder for data cleaning and classification in the presence of label noise. *Neural Processing Letters*, 50(2):1845–1860, 2019.

[13] Zhen-Yu Zhang, Peng Zhao, Yuan Jiang, and Zhi-Hua Zhou. Learning from incomplete and inaccurate supervision. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1017–1025, 2019.

[14] Zhi-Hua Zhou. A brief introduction to weakly supervised learning. *National science review*, 5(1):44–53, 2018.