Impact of Noisy Labels in Learning Techniques: A Survey

Conference Paper · January 2020

DOI: 10.1007/978-981-15-0694-9_38

3 authors:

Nitika Nigam
Indian Institute of Technology (Banaras Hindu University) Varanasi

Tanima Dutta
Indian Institute of Technology (Banaras Hindu University) Varanasi

Hari prabhat Gupta
Indian Institute of Technology (Banaras Hindu University) Varanasi

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

Impact of Noisy Labels in Learning Techniques: A Survey

Nitika Nigam, Tanima Dutta and Hari Prabhat Gupta

Abstract Noisy data is the main issue in classification. The possible sources of noise label can be insufficient availability of information or encoding/communication problems, or data entry error by experts/non-experts, etc., which can deteriorate the model's performance and accuracy. However, in a real-world dataset, like Flickr, the likelihood of containing the noisy label is high. Initially, few methods such as identification, correcting and elimination of noisy data was used to enhance the performance. Various machine learning algorithms are used to diminish the noisy environment, but in the recent studies, deep learning models are resolving this issue. In this survey, a brief introduction about the solution for the noisy label is provided.

Keywords Noisy Labels, Deep-learning approach, Non-deep learning approach

1 Introduction

Deep learning is one of the latest emerging area of machine learning which enables computer to learn from the experience [Goodfellow et al.(2016)Goodfellow, Bengio, Courville, and Bengio]. In general, the deep learning model imitates the working of human brain to process the data for decision making. The term learning refers to be an iterative process which helps to enhance the knowledge gain. So, the capability of good decision making depends upon the datasets. The dataset is the collection of real-world data which is used in deep learning model. It is further classified into two parts: training dataset and test dataset [Goodfellow et al.(2016)Goodfellow, Bengio, Courville, and Bengio]. The training dataset is used in learning process of model for classification of new samples. Therefore, the noiseless dataset is essential requirement for training model. But, the analysis of real-world applications always produces a large set of values that will not inevitably have truth values, known as noisy data. The "Noisy data" referred as mislabeled data which was discovered in 1980s, a [Oja(1980), Angluin and Laird(1988)] and yet the research is also active in current years [Sun et al.(2018)Sun, Xu et al., Han et al.(2018)Han, Yao, Yu, Niu,

Department of Computer Science and Engineering, IIT (BHU) Varanasi, India, E-mail: (nitikanigam.rs.cse18, tanima.cse, hariprabhat.cse)@iitbhu.ac.in

Xu, Hu, Tsang, and Sugiyama]. The presence of noise in the dataset obscured the relationship between the features of an object and its class [Hickey(1996)]. This will increases the complexity problem in classification of dataset.
Noise is categorized into an attribute (feature) noise, and class noise[Zhu and Wu(2004)]. The modification in observed values are known as feature noise and mislabelling of a class in the observed labels, allocated to an instance are known as class noise [Zhu and Wu(2004)]. In [Quinlan(1986)] the non-systematic errors are refer as noise, in which the values of attributes, class attributes or both have been altered. Some authors [Sun et al.(2018)Sun, Xu et al.] categories the noise into two parts: symmetric and asymmetric noise. The noise due to uniform distribution are known as symmetric noise and asymmetric noise occur due to fixed-rule flipping [Sun et al.(2018)Sun, Xu et al.].
The class noise is said to be label noise if observed labels are polluted, i.e., incorrectly labeled [Frénay and Verleysen(2014)]. The root of label noise involves data entry error by a specialist or annotators, inadequate availability of resources or knowledge provided to tag a label for a particular object, inter-experts advice on the identical article or low budget non-experts are employed for labeling the data [Brodley and Friedl(1999)]. In [Frénay and Verleysen(2014)] the communication and encoding of data are reported to be mislabeled noisy data, in the real-world applications approximately 5% of encoding and communication fault refers to noise [Orr(1998)]. The existence of noise in the dataset leads to degradation of performance [Nettleton et al.(2010)Nettleton, Orriols-Puig, and Fornells, Yao et al.(2019)Yao, Wang, Tsang, Zhang, Sun, Zhang, and Zhang] in classification models. This paper focuses on label noise which is a subset of class noise. The essential demand is to learn from the accurate data as the noisy data are the mislabeled data which furnish the wrong information. Thus, the accurate prediction depends on the learning of a machine with the trustworthiness of labeled data. But, the probability of getting noiseless data is inadequate in case of real-world applications e.g. Flickr provides the image dataset but mostly consist of the noisy labeled dataset as the data has been indicated by humans.
The main contribution of this paper are summarized as follows:

- We review about the noisy data and their effect in classification problems.
- This paper also focus on the various approaches to resolve the aforementioned problem.

The rest of the paper is organized as follows. The next section describe about the noise, its source and consequence of noisy labels. Section 3 is the related work, classified into a deep learning approach and non-deep learning approach to resolve the problem of classification with corrupted labels. We conclude the paper in section 4.

2 Noisy Labels: Definition, Source, and Consequences

Noise is an irregular patterns present in the dataset but is not part of real data. In [Hickey(1996)], noise is defined as the ambiguous relation between the features and its class. The ubiquity of noise in the data may alter the essential characteristic of an object. Due to this, the performance of classification may degrades [Hickey(1996)]. There are following challenges due to noise which may affect the intrinsic properties in classification problem are discussed as follows:

- It may create small clusters of data points of a particular class in the area of domain which belongs to the different class.
- It eliminates instances located in key areas within an appropriate class.
- It also disrupts the boundary of the classes and increases overlapping among them.

The noise is classified into attribute noise and class noise, which are decribed as follows [Zhu and Wu(2004), Teng(1999)]:

- The class noise occurs due to misclassification and inconsistent examples. The source of class noise also includes contradictory examples and mislabelling of an instance. In general, it occur on the boundaries of classes where the characteristics similarity between the datapoints is maximum.
- Attribute noise incorporates unavailable values, error in values and incomplete information. An alteration in the values of attributes of an instance are also referred as attribute noise.
  Attribute noise are more harmful than class noise because erroneous attribute values is uncertain and random.

A lot of work done had been done in the field of removing the class noise which results in improving the classification accuracy. The difficult task is to remove attribute noise and still is in research. The class labels are intentionally corrupted by an adversary, which is known as Noisy labels [Biggio et al.(2011)Biggio, Nelson, and Laskov].
The source of noise are also relevant which are discussed as follows [Frénay and Verleysen(2014)]:

- Distribution.
- Data entry error.
- Inefficient data description to tag a class of a instance.
- Subjective error.
- Data communication and encoding problem.
- Non-experts decision making.
- Instances overlapping near boundary.
- Biological artifacts.

The presence of noise is a key issue in the real-world dataset and has many drawbacks. The consequences of label noise are given by:

- Decrement in the performance of classification [Liu and Zhang(2012)].Some authors shown experimentally that noisy label leads to degradation in performance [Nettleton et al.(2010)Nettleton, Orriols-Puig, and Fornells,Yao et al.(2019)Yao, Wang, Tsang, Zhang, Sun, Zhang, and Zhang].
- The complexity of learned models is increased, for instance, the increment in the number of nodes of decision trees [Frénay and Verleysen(2014)].
- The examined instances of the possible classes can be changed [Frénay and Verleysen(2014)].

3 Approaches used to resolve noisy data

The problem of learning with noisy labels is a serious issue in any classification model. There are following sources such as Mechanical Turk or any crowdsourcing platform, i.e., Google and Microsoft for obtaining large dataset, but the possibility of noise in the dataset is very high. Numerous non-deep learning and deep learning approaches have been proposed to resolve this problem in various applications. Here is a concise explanation of the related work below and in Table 1 summary of some methods are shown.

3.0.1 Non-Deep Learning Approach

Some non-deep learning strategies have been proposed to deal with noisy label dataset. These are statistical methods to reduce the noise and to enhance the performance.

In 1999, the identification of mislabeled noise was proposed for a small dataset. The author applied ensemble classifiers such as decision tree, linear classifier, and k-nearest neighbor (KNN) in addition with majority and consensus filtering scheme [Brodley and Friedl(1999)][Verbaeten and Van Assche(2003),Khoshgoftaar et al.(2005)Khoshgoftaar, Zhong, and Joshi].Majority and non-objection are the two threshold systems for classification of noise. After identification and correction of noisy labels, the uncomplicated approach is to eliminate the noise, here in [Zhu et al.(2003)Zhu, Wu, and Chen] mislabeled data is first recognized and then eliminated from the large dataset. The error count variables are used to compute the noisy label instances, and the higher probability occurrence tagged as the noisy data. Another method was proposed for correcting the mislabeled data [Teng(1999)] by comparing the polished (clean) dataset with the noisy dataset.
Bagging and boosting are the two strategies for the detection of noisy data in which weights are assigned to the training sample. The weights are updated on each turn which are labeled as noisy labels if the count value exceeded a threshold value.[Freund et al.(1996)Freund, Schapire et al.,Friedman et al.(2000)Friedman, Hastie, Tibshirani et al.,Zhong et al.(2005)Zhong, Tang, and Khoshgoftaar]. Adaboosting algorithm is used for identification and detection of noise by comparison with the clean dataset for only binary classes and in [Bootkrajang and Kabán(2013)] robust boosting algorithm is integrated with Adaboost for identification of mislabeled data. A ranking based method is also used for the detection of noise. An ensemble-based ranking method is recommended in [Sluban et al.(2014)Sluban, Gamberger, and Lavrač], noisy instances are rank on the occurance of inaccurate predictions done by a learner.
Some research also concentrates on the relabeling of noisy data, which helps in eliminating the noisy data from crowdsourcing or any website. In [Lin et al.(2014)Lin, Weld et al.], the author proposed a non-deep learning approach that attempts relabeling of noisy labels and on each round, the noisy labels, as well as predictors, are updated. There is some probabilistic modeling based method which was provided by [Raykar et al.(2010)Raykar, Yu, Zhao, Valadez, Florin, Bogoni, and Moy] for supervised learning to determine the problem generated by multiple experts data. In this, the noisy labels are removed by majority votes. The same concept is extended by [Yan et al.(2014)Yan, Rosales, Fung, Subramanian, and Dy] for the multi-annotators problem through supervised and semi-supervised learning. The additional work done was that a model was proposed to learn from unlabeled data in the semi-supervised environment as well as to handle missing annotators.

Some approaches are based on noise rate estimation which is shown in [Menon et al.(2015)Menon, Rooyen, Ong, and Williamson]. The author applied class-probability estimator to learn from altered binary labels by utilizing the order-statistics on a predicted array of numbers.

Table 1 Summary of methods for Noisy labels

Statistical Method		Deep learning method	
Surrogate loss	Majority and non-objective methods [Verbaeten and Van Assche(2003)]	Robust loss Function	1. Defense mechanism for various architecture [Vu and Tran(2018)] 2. Noisy level reweight [Li et al.(2017)Li, Yang, Song, Cao, Luo, and Li]
Bagging and Boosting	Data Cleaning Method [Sun et al.(2007)Sun, Zhao, Wang, and Chen, Karmaker and Kwek(2006)]		3.Cross-entropy loss [Joulin et al.(2016)Joulin, van der Maaten, Jabri, and Vasilache].
	Data tolerant Method [Bootkrajang and Kabán(2013), Cantador and Dorronsoro(2005), Oza(2004), Cantador and Dorronsoro(2005)]		4. Global ratio [Izadinia et al.(2015)Izadinia, Russell, Farhadi, Hoffman, and Hertzmann] 5. Bootstrap function [Reed et al.(2014)Reed, Lee, Anguelov, Szegedy, Erhan, and Rabinovich]
Probabilistic Method	Cluster based approach [Bouveyron and Girard(2009)]	Modelling the latent labels	1. Parallel classifier
Noise rate estimation method	Noise elimination approach [Hickey(1996)]		

3.0.2 Deep learning Approach

Different techniques have been proposed to deal with noisy labels in various applications using machine learning models. In recent years, the deep learning model overshadow the previous techniques.Noisy data problem is resolve by using deep learning techniques. It is classified into two methods which is robust-loss function and modeling the latent labels. In [Azadi et al.(2015)Azadi, Feng, Jegelka, and Darrell] CNN deep learning model has been used for image classification, in which an auxiliary image regularization procedure is used to depreciate the noisy data. Bootstrapping and importance re-weighting [Reed et al.(2014)Reed, Lee, Anguelov, Szegedy, Erhan, and Rabinovich,Liu and Tao(2016)] are the techniques to resolve the dilemma of statistical outliers, i.e., random classification noise. The bootstrapping [Reed et al.(2014)Reed, Lee, Anguelov, Szegedy, Erhan, and Rabinovich] method was given by Yarowsky in 1995, is a self-learning method which helped in handling subjective as well as unlabeled images. Re-weighting is another

technique which is used to alleviate the problem to depreciate the noisy labels [Liu and Tao(2016)]. The information is acquired from a small clean dataset and with the help of knowledge graph, noisy labels have been updated.
The method to protect from acquiring the noisy labeled dataset by deep neural network, defense mechanism (adversarial attack and distillation) have been proposed. In [Vu and Tran(2018)], the author introduced multiple defense mechanisms to protect the deep architecture from learning the noisy data by using robust loss function for adversarial attacks besides of using cross-entropy softmax loss function. Above literature is based on robust loss function to diminish the effect of noise.
Some work has been done in the field of modeling the latent labels to guide the classifiers accurately. It also helps in establishing a change for adaption from latent labels to noisy labels. The first work was performed for random and symmetric noise in which [Mnih and Hinton(2012)] proposed a latent model for aerial images. This approach was extended for different noise, in [Sukhbaatar et al.(2014)Sukhbaatar, Bruna, Paluri, Bourdev, and Fergus] used the convolutional neural network to train on the small and large noisy dataset. Two models have been proposed to minimize the noisy data which significantly improves the performance. In [Chen and Gupta(2015)] two steps strategy is used to train the convolutional neural network. In the first step, the simple web images have been taken from google image search which is learned by CNN's to grab the knowledge. Further, in the following step, Flickr images are used to train, but it is challenging to train CNN's with noisy data. Thus, the refining method has been done before provided as input to CNN. The refining method is based on the similarity relationship graph; if there is an existence of error, then it is back-propagated through graph as to get the accurate classification.
Some methods are based on electing the samples to mark noisy labels for example in [Malach and Shalev-Shwartz(2017)] decoupling method is used which maintain two predictors. The updation of predictors are depended upon disagreement samples. But, this approach is depended upon selecting sample and bias selected sample may lead to an error. Co-teaching method is given by [Han et al.(2018)Han, Yao, Yu, Niu, Xu, Hu, Tsang, and Sugiyama], in which two distinct deep neural networks (DNN) have been trained concurrently, and in every mini batch dataset, they teach each other. The updation policy for a noisy label is the same as [Malach and Shalev-Shwartz(2017)] i.e., on disagreement noisy data is updated.There is a probabilistic approach to eliminate the noisy data using deep learning, in [Rodrigues and Pereira(2018)] an algorithm has been stated which clean the noisy labels on the postulate of majority voting method.
In [Sun et al.(2018)Sun, Xu et al.], an unsupervised method is to determine the find best stop time before learning the noisy data. This method is known as Limited Gradient Descent (LGD) which is applied to both symmetric and noisy asymmetric labels. A reverse sample is created from a given dataset which is different from the main pattern; the analysis significantly determines that only large scaled clean patterns are to be learned.

4 Conclusion

The presence of noise in data is a common problem that produces several negative consequences in classification problems. This survey summarized that the noisy data is a complex problem and harder to provide an accurate solution. In general, the data of real-world application is the key source of noisy data. There are two approaches to handle noisy labels. In the deep learning approach, different architectures are implemented for the elimination of noisy labels. The method for elimination of noisy labels in deep learning approach are further classified into a robust-loss function and modeling latent variable. The statistical-based methods have been discussed in the non-deep learning approach in which mostly algorithms were based on majority voting mechanism, bagging and boosting method, noise rate estimation and the probabilistic method.

All these approaches improve the performance but can't eliminate the noisy labels completely. In recent work Limited Gradient Descent method has been used for getting the best stop time before learning the noisy labels without eliminating the noisy labels. The work on symmetric noise and asymmetric noise has been ignored. In future, the scope of work could be done in the above field.

Acknowledgements This work is supported by Science and Engineering Research Board (SERB) file number ECR/2017/002419, project entitled as A Robust Medical Image Forensics System for Smart Healthcare, and scheme Early Career Research Award.

References

[Angluin and Laird(1988)] Angluin D, Laird P (1988) Learning from noisy examples. Machine Learning 2(4):343–370

[Azadi et al.(2015)Azadi, Feng, Jegelka, and Darrell] Azadi S, Feng J, Jegelka S, Darrell T (2015) Auxiliary image regularization for deep cnns with noisy labels. arXiv preprint arXiv:151107069

[Biggio et al.(2011)Biggio, Nelson, and Laskov] Biggio B, Nelson B, Laskov P (2011) Support vector machines under adversarial label noise. In: Asian Conference on Machine Learning, pp 97–112

[Bootkrajang and Kabán(2013)] Bootkrajang J, Kabán A (2013) Boosting in the presence of label noise. arXiv preprint arXiv:13096818

[Bouveyron and Girard(2009)] Bouveyron C, Girard S (2009) Robust supervised classification with mixture models: Learning from data with uncertain labels. Pattern Recognition 42(11):2649–2658

[Brodley and Friedl(1999)] Brodley CE, Friedl MA (1999) Identifying mislabeled training data. Journal of artificial intelligence research 11:131–167

[Cantador and Dorronsoro(2005)] Cantador I, Dorronsoro JR (2005) Boosting parallel perceptrons for label noise reduction in classification problems. In: International Work-Conference on the Interplay Between Natural and Artificial Computation, Springer, pp 586–593

[Chen and Gupta(2015)] Chen X, Gupta A (2015) Webly supervised learning of convolutional networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp 1431–1439

[Frénay and Verleysen(2014)] Frénay B, Verleysen M (2014) Classification in the presence of label noise: a survey. IEEE transactions on neural networks and learning systems 25(5):845–869

[Freund et al.(1996)Freund, Schapire et al.] Freund Y, Schapire RE, et al. (1996) Experiments with a new boosting algorithm. In: Icml, Citeseer, vol 96, pp 148–156

[Friedman et al.(2000)Friedman, Hastie, Tibshirani et al.] Friedman J, Hastie T, Tibshirani R, et al. (2000) Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). The annals of statistics 28(2):337–407

[Goodfellow et al.(2016)Goodfellow, Bengio, Courville, and Bengio] Goodfellow I, Bengio Y, Courville A, Bengio Y (2016) Deep learning, vol 1. MIT press Cambridge

[Han et al.(2018)Han, Yao, Yu, Niu, Xu, Hu, Tsang, and Sugiyama] Han B, Yao Q, Yu X, Niu G, Xu M, Hu W, Tsang I, Sugiyama M (2018) Co-sampling: Training robust networks for extremely noisy supervision. arXiv preprint arXiv:180406872

[Hickey(1996)] Hickey RJ (1996) Noise modelling and evaluating learning from examples. Artificial Intelligence 82(1-2):157–179

[Izadinia et al.(2015)Izadinia, Russell, Farhadi, Hoffman, and Hertzmann] Izadinia H, Russell BC, Farhadi A, Hoffman MD, Hertzmann A (2015) Deep classifiers from image tags in the wild. In: Proceedings of the 2015 Workshop on Community-Organized Multimodal Mining: Opportunities for Novel Solutions, ACM, pp 13–18

[Joulin et al.(2016)Joulin, van der Maaten, Jabri, and Vasilache] Joulin A, van der Maaten L, Jabri A, Vasilache N (2016) Learning visual features from large weakly supervised data. In: European Conference on Computer Vision, Springer, pp 67–84

[Karmaker and Kwek(2006)] Karmaker A, Kwek S (2006) A boosting approach to remove class label noise 1. International Journal of Hybrid Intelligent Systems 3(3):169–177

[Khoshgoftaar et al.(2005)Khoshgoftaar, Zhong, and Joshi] Khoshgoftaar TM, Zhong S, Joshi V (2005) Enhancing software quality estimation using ensemble-classifier based noise filtering. Intelligent Data Analysis 9(1):3–27

[Li et al.(2017)Li, Yang, Song, Cao, Luo, and Li] Li Y, Yang J, Song Y, Cao L, Luo J, Li LJ (2017) Learning from noisy labels with distillation. In: Proceedings of the IEEE International Conference on Computer Vision, pp 1910–1918

[Lin et al.(2014)Lin, Weld et al.] Lin CH, Weld DS, et al. (2014) To re (label), or not to re (label). In: Second AAAI conference on human computation and crowdsourcing

[Liu and Zhang(2012)] Liu H, Zhang S (2012) Noisy data elimination using mutual k-nearest neighbor for classification mining. Journal of Systems and Software 85(5):1067–1074

[Liu and Tao(2016)] Liu T, Tao D (2016) Classification with noisy labels by importance reweighting. IEEE Transactions on pattern analysis and machine intelligence 38(3):447–461

[Malach and Shalev-Shwartz(2017)] Malach E, Shalev-Shwartz S (2017) Decoupling" when to update" from" how to update". In: Advances in Neural Information Processing Systems, pp 960–970

[Menon et al.(2015)Menon, Rooyen, Ong, and Williamson] Menon A, Rooyen BV, Ong CS, Williamson B (2015) Learning from corrupted binary labels via class-probability estimation. In: Bach F, Blei D (eds) Proceedings of the 32nd International Conference on Machine Learning, PMLR, Lille, France, Proceedings of Machine Learning Research, vol 37, pp 125–134, URL http://proceedings.mlr.press/v37/menon15.html

[Mnih and Hinton(2012)] Mnih V, Hinton GE (2012) Learning to label aerial images from noisy data. In: Proceedings of the 29th International conference on machine learning (ICML-12), pp 567–574

[Nettleton et al.(2010)Nettleton, Orriols-Puig, and Fornells] Nettleton DF, Orriols-Puig A, Fornells A (2010) A study of the effect of different types of noise on the precision of supervised learning techniques. Artificial intelligence review 33(4):275–306

[Oja(1980)] Oja E (1980) On the convergence of an associative learning algorithm in the presence of noise. International Journal of Systems Science 11(5):629–640

[Orr(1998)] Orr K (1998) Data quality and systems theory. Communications of the ACM 41(2):66–71

[Oza(2004)] Oza NC (2004) Aveboost2: Boosting for noisy data. In: International Workshop on Multiple Classifier Systems, Springer, pp 31–40

[Quinlan(1986)] Quinlan JR (1986) Induction of decision trees. Machine learning 1(1):81–106

[Raykar et al.(2010)Raykar, Yu, Zhao, Valadez, Florin, Bogoni, and Moy] Raykar VC, Yu S, Zhao LH, Valadez GH, Florin C, Bogoni L, Moy L (2010) Learning from crowds. Journal of Machine Learning Research 11(Apr):1297–1322

[Reed et al.(2014)Reed, Lee, Anguelov, Szegedy, Erhan, and Rabinovich] Reed S, Lee H, Anguelov D, Szegedy C, Erhan D, Rabinovich A (2014) Training deep neural networks on noisy labels with bootstrapping. arXiv preprint arXiv:14126596

[Rodrigues and Pereira(2018)] Rodrigues F, Pereira FC (2018) Deep learning from crowds. In: Thirty-Second AAAI Conference on Artificial Intelligence

[Sluban et al.(2014)Sluban, Gamberger, and Lavrač] Sluban B, Gamberger D, Lavrač N (2014) Ensemble-based noise detection: noise ranking and visual performance evaluation. Data mining and knowledge discovery 28(2):265–303

[Sukhbaatar et al.(2014)Sukhbaatar, Bruna, Paluri, Bourdev, and Fergus] Sukhbaatar S, Bruna J, Paluri M, Bourdev L, Fergus R (2014) Training convolutional networks with noisy labels. arXiv preprint arXiv:14062080

[Sun et al.(2007)Sun, Zhao, Wang, and Chen] Sun Jw, Zhao Fy, Wang Cj, Chen Sf (2007) Identifying and correcting mislabeled training instances. In: Future generation communication and networking (FGCN 2007), IEEE, vol 1, pp 244–250

[Sun et al.(2018)Sun, Xu et al.] Sun Y, Xu Y, et al. (2018) Limited Gradient Descent: Learning With Noisy Labels. arXiv preprint arXiv:181108117

[Teng(1999)] Teng CM (1999) Correcting noisy data. In: ICML, Citeseer, pp 239–248

[Verbaeten and Van Assche(2003)] Verbaeten S, Van Assche A (2003) Ensemble methods for noise elimination in classification problems. In: International Workshop on Multiple Classifier Systems, Springer, pp 317–325

[Vu and Tran(2018)] Vu TK, Tran QL (2018) Robust Loss Functions: Defense Mechanisms for Deep Architectures. In: 2018 10th International Conference on Knowledge and Systems Engineering (KSE), IEEE, pp 163–168

[Yan et al.(2014)Yan, Rosales, Fung, Subramanian, and Dy] Yan Y, Rosales R, Fung G, Subramanian R, Dy J (2014) Learning from multiple annotators with varying expertise. Machine Learning 95(3):291–327, DOI 10.1007/s10994-013-5412-1, URL https://doi.org/10.1007/s10994-013-5412-1

[Yao et al.(2019)Yao, Wang, Tsang, Zhang, Sun, Zhang, and Zhang] Yao J, Wang J, Tsang IW, Zhang Y, Sun J, Zhang C, Zhang R (2019) Deep learning from noisy image labels with quality embedding. IEEE Transactions on Image Processing 28(4):1909–1922

[Zhong et al.(2005)Zhong, Tang, and Khoshgoftaar] Zhong S, Tang W, Khoshgoftaar TM (2005) Boosted noise filters for identifying mislabeled data. Department of Computer Science and engineering, Florida Atlantic University

[Zhu and Wu(2004)] Zhu X, Wu X (2004) Class noise vs. attribute noise: A quantitative study. Artificial intelligence review 22(3):177–210

[Zhu et al.(2003)Zhu, Wu, and Chen] Zhu X, Wu X, Chen Q (2003) Eliminating class noise in large datasets. In: Proceedings of the 20th International Conference on Machine Learning (ICML-03), pp 920–927