International Journal of Engineering and Information Systems (IJEAIS)
ISSN: 2643-640X
Vol. 10 Issue 6 June - 2026, Pages: 137-144

Weakly Supervised and Noisy Label Learning: When Ground Truth Is Imperfect

Virendra Tank

Shri Mahaveer College, Jaipur, India

vtank87@gmail.com 0009-0003-9126-0982

Abstract: *The perfect, ideal clean labelled data assumption does not hold in actual machine learning problems. Here, we survey several techniques for learning from imperfect supervision such as missing annotation, noisy label and weak supervision cues. We review multi-instance learning frameworks, junk label de-noising and outlier removal methods, week supervision based object detection & segmentation algorithms as well as crowd-based label aggregation schemes. By dissecting the theory and phenomena learned with modern deep learning on theories in medical imaging and web-scale data, we establish that reliable knowledge can be extracted autonomously from unreliable supervision. This review presents recent progress in weakly supervised learning and discusses the main challenges and future perspectives for learning with imperfect ground truth.*

Keyword: Multiple Instance Learning, Weakly Supervised Object Detection, Learning Label Noise, Crowdsourced Label

1. Introduction

The key to the success of deep learning is well known to be the existence of massive-scale, annotated data [1]. However, good quality annotations are expensive to obtain, time consuming and often demand domain expertise [2]. Labels are often imperfect: in the real world, labels can be incomplete or noisy, or may correspond to a coarser granularity than desired [3]. Clinical imaging depends author assumption of medical radiologists for labeling [4], and labels in web-scale data are bound to be incorrect due to automated or crowd-sourced annotations [5].

Weakly supervised learning tackles situations when supervision is not complete or is imprecise, e.g. image-level labels for object localization and bag-level labels for instance classification [6]. Learning with noisy labels addresses the problem of learning robust models when a large proportion of training labels are wrong [7]. These paradigms are becoming particularly crucial with the explosion of machine learning to very large scale datasets where full annotation is not realistic [8].

Recent work on deep learning has led to complex methods of imperfect supervision. Self-supervised pre-training can be viewed as similar to multi view learning since it learns different representations without any labeled data [9], but with more robust initializations for better adaptation to label noise. By attending the inter-relationship between data instances, attention mechanisms and MIL frameworks can pool weak signals to make fine-grained predictions [10]. Meta-learning techniques variably weight training examples according to an estimated label quality [11]. This work synthesizes this variety of methods and is an in-depth review on how to learn when ground truths are imperfect.

2. Multiple Instance Learning

2.1 Foundations and Formulation

Multiple Instance Learning (MIL) for addressing problems where training data is organized as bags of instances, and labels are only available at the bag level instead of individual instances [12]. The MIL assumption is known as standard when one bag is defined to be positive if it contains at least one instances but all are negatives [13]. This framework innately represents a variety of practical problems, such as drug activity prediction, content-based image retrieval and medical image analysis [14].

In terms, if $B_i = {x_{i1}, x_{i2}, \dots, x_{in_i}}$ is a bag and we have its bag-level label Y_i, MIL learns a function that can predict the labels of each bag meeting at most one positive instances[15]. The major problem is credit assignment; that is, which examples inside the positive bags are considered as truly positives [16]. Standard MIL techniques, such as Diverse Density measures which discover areas in the feature space where positive bag instances are concentrated [17], and mi-SVM, with the instance labels being latent variables that are simultaneously optimized with the classifier [18].

2.2 Deep Multiple Instance Learning

MIL has been revolutionized by deep learning, where instance representation and aggregation function can be learned in an end-to-end manner [19]. The attention-based MIL models learn a weight for each instance according to its relevance to the bag label [10]. The attention mechanism calculates the importance scores over instances so that GNN can emphasis discriminative instances and ignore irrelevant ones [20].

Attention-based MIL The attention-based MIL framework consists of two essential modules: instance-lever feature extractor (usually a deep neural network), and an attention based pooling [10]. It aggregates instance features with attention weights

calculated in a gating manner to form bag-level representation for classification [21]. This method has been shown remarkably successful in computational pathology, treating gigapixel whole slide images as bags of tissue patches [22].

2.3 Applications and Extensions

MIL has found tremendous success in the field of medical imaging where acquiring pixel-level labels for regions is costly [23]. Weakly supervised histopathology image analysis learns from slide-level diagnostic labels to train models capable of localizing cancerous regions on the patch level [24]. This has facilitated computer diagnosis systems that rival or surpass human pathologist performance [22].

Recent extensions are instance-level MIL, which predicts both bag and instance labels [25], and multi-instance multi-label learning, where bags have multiple simultaneous labels [26]. DeepMIL has also been extended to video understanding, when videos are considered as bags of frame or segment [27] and for document classification, where documents are represented as a bag of sentence or passage [28].

3. Learning with Label Noise and Outliers

3.1 Sources and Characteristics of Label Noise

Label noise is caused by various factors: human mistakes, automatic errors in labeling process, disturbances within data collection procedures and ambiguity or subjectivity inside labels's assignment tasks [29]. Two noise models are widely used to describe corruption of labels: symmetric noise (which consists in that any label can be flipped to any other with the same probability), and asymmetric noise (in which labels tend to confuse more with classes semantically similar [30]). Noise in the real-world tends to be class-biased due to a systematic annotated bias [7].

Effect of label noise on deep learning is very devastating: with enough capacity, neural networks can memorize any kind of label-noise mappings that results in poor generalization [31]. Nevertheless, deep networks follow a memorization rule that they tend to learn clean patterns before fitting distorted labels, which implies the potential for early stopping or sample selection [32]. Motivated by this learning behavior, many noise-robust training methods have been proposed[33].

3.2 Noise - Robust Loss Functions

Strong agnostic loss functions naturally offer some immunity to label noise because they automatically attenuate the impact of suspect examples [34]. Symmetric cross-entropy adds to the standard cross-entropy another reverse-oriented term of cross entropy that measures how much and in which direction this model is away from a learned model [35]. This regularizes prevents the model from overfitting to noisy labels by discouraging confident predictions on likely noisy labels [36].

The Generalized Cross Entropy (GCE) discriminating loss protagonist the tunable connections between mean absolute error and cross-entropy, which effectively balances noise robustness with discriminative capability [37]. The loss function smoothly changes from cross-entropy (for clean data) to MAE (for noisy data) according to confidence of prediction [37]. Focal Loss was also proposed for class imbalance, but shows robustness to noise by down-weighting high-confidence predictions that may occur due to fitting the noise [38].

3.3 Sample Selection and Reweighting

Sample selection methods determine and remove examples likely to have been mislabeled during training [39]. Small-loss selection: small-loss selection takes advantage of the fact that neural networks fit clean examples before noisy ones and select training examples with the smallest losses [32]. Co-teaching uses two networks that train on mini-batches chosen by the peer network, using disagreement to discard noise [39].

MentorNet: Inspired by meta-learning, a "mentor" network learns to predict those samples that should be more weighted in training the "student" network[40]. This mentor is trained on a small clean validation set to learn characteristics of examples that are accurately labeled (which is used for cleaning the poison [11]). Confident learning models the joint distribution of noisy and true labels, thereby supporting principled label error detection and correction [5].

By meta-training, the sample-wise reweighting methods aim to learn how to reweigh samples such that performance on a clean validation set can be maximized [11]. By casting reweighting as a meta-learning task, these methods dynamically pick example importance proportional to their contribution to performance on new data [41]. Indeed, this approach is shown to be effective even with small size validation set by which it gives the gradient-based feedback for weight optimization [42].

3.4 Noise Modeling and Correction

The estimation of the noise transition matrix explicitly models the label contamination process, i.e., a confusion matrix that represents the probability of observed labels given the original one is learned[30]. After obtaining an approximated transition matrix, they can be adjusted by marginalizing on possible true labels with the respect to the transition probability weighted.[43] This matrix

can be estimated by various approaches, for example, method using anchor points (samples confidently assigned to certain classes) or iterative estimation process [44].

Direct noisy label cleaning: Noisy labels are directly rectified during the training. Knowledge distillation based co-training employs teacher model's predictions to iteratively re-label the training data [33]. Self-training based methods sequentially purify pseudo-labels using model predictions until incrementally evolving the label quality [46]. Nonetheless, these strategies are prone to error propagation if the model has already been trained on noisy labels [47].

4. Weakly Supervised Object Detection and Segmentation

4.1 Weakly Supervised Object Detection

Weakly Supervised Object Detection (WSOD) tries to learn object detectors with only image-level labels for object presence, but no bounding box supervision [48]. This is a difficult task because the model has to learn both what object it's looking at and where the object resides [49]. Earlier WSOD techniques employed selective search to produce state-to-the-art object proposals and chose the most discriminative ones among them relative to image-level labels [50].

WSOD with deep learning methods exploit Class Activation Maps (CAMs) to indicate discriminative regions by projecting learned features back into the input space [51]. CAMs locate in approximate object regions, in many cases only emphasizing the most discriminative region and not whole objects [52]. Some multiple instance learning (MIL) frameworks treat images as bags of region proposals and learn to select and refine the proposals that are best explanations for image labels[48].

Many recent WSOD methods use online instance classifier refinement to refine pseudo ground truth bounding boxes during training iteratively [53]. Proposal cluster learning groups proposals that are similar across images to discover common object patterns [54]. SSL is able to remove noise and encourage an accurate localization [23].

4.2 Weakly Supervised Semantic Segmentation

Weakly Supervised Semantic Segmentation (WSSS) methods infer pixel and instance level segmentations from weak supervision, such as image-level tags, bounding boxes, scribbles or points [55]. Image-level supervision is the most difficult setting, as models must learn what object categories exist and where they are [56]. CAM-based approaches for which initial localization cues are provided that are iteratively refined during training and pseudo label generation [51].

AffinityNet, which trains a network to predict the semantic context between neighboring pixels and thereby propagate CAM activations for generating pixel-wise segmentation masks [57]. This research observes object bounds better than CAMs alone by modelling local pixel relationships [58]. Further post-processing using CRFs [60, 61] or dense CRFs also has demonstrated the capability in enhancing boundary accuracy [59].

Self-training with uncertainty estimation gradually enlarges the confident predictions to those uncertain areas [46]. The model first creates high confidence pseudo-labels for easy regions, and then utilizes them as supervision to segment more difficult parts [60]. Contrastive learning of two different augmentations promotes appearance invariant segmentations [9].

4.3 Point and Scribble Supervision

Point supervision offers object locations without extent, forcing models to generalize from sparse point annotations to full segmentation [55]. This kind of supervision turns out to be very efficient, since the annotator only needs one click to label an object instead of drawing boundaries [2]. Point-supervised methods use point annotations as initial points for the segmentation algorithm [55] or as constraints in an optimization framework.

Sketch reversion relies on loose, undetailed contours in place of boundary tracing attentive to details [55]. Scribbles require less annotation than full mask and offer more information than the points or boxes [2]. Such a loss function regularizes the task by promoting smooth segmentations following scribble constraints [59]. Scribble labels are propagated to unlabeled pixels on a graph by feature similarity [58].

5. Crowdsourced Label Aggregation

5.1 Challenges in Crowdsourcing

Crowdsourcing facilitates both fast and large scale data annotation, due to your ability to break down station labeling into numerous tasks handled by non-professional workers [1]. Nevertheless, crowdsourced labels tend to be high noisy because of different worker expertise, attention and motivation [7]. Workers can be biased differently, possess different levels of domain knowledge and have preferences for certain type of errors [29]. Malicious workers can return arbitrary, or adversarial labels which allows them to complete tasks fast [5].

Reliability of crowd-annotations is maintained by quality control measures. Redundancy, in which multiple workers annotate each example, makes it possible to identify outlier responses and aggregate via majority voting [5]. Questions with known answers in the

gold standard are used to measure worker quality, and identify unreliable workers [8]. Annotation quality and worker engagement are influenced by attention checks and task design [2].

5.2 Label Aggregation Methods

Majority voting by simple majority pools all labels, choosing the most frequent answer among workers [1]. Although simple, this does not distinguish between workers based on their expertise and reliability [8]. Weighted voting systems give weight to agents according to how closely the agent's output matches those of other agents or the correctness on validation examples [5].

Dawid-Skene model directly estimates confusing matrices of workers and true labels together using expectation-maximization [8]. This generative process captures the reliability pattern for each worker, including systemic bias and varying level of expertise across classes [29]. Extensions include worker skill, task complexity and time dynamics of the workers performance [7].

GLAD (Generative model of Labels, Abilities and Difficulties) is another approach that takes account of both worker capability and task difficulty as latent variables [8]. This method acknowledges the fact that some examples are more challenging to label correctly, even for experienced workers [5]. Extensions to deep learning replace hand-engineered difficulty models with learned features, increasing scalability and flexibility [7].

5.3 Learning from Crowds

End-to-end learning from crowd labels additionally merges label aggregation with model training [8]. They jointly infer true labels and model parameter, instead of aggregating real labels before training [5]. The model is trained with all the crowdsourced labels at once, which might be able to infer some patterns that would not have been discovered by simple aggregation [7].

Crowd Layer a new layer for modeling worker effects in multiple-choice tasks was proposed by is a neural network, named as Crow-Layer, which captures the labeling process of each worker considering their confusion matrices [8]. This method avoids explicit aggregation as it allows direct training on disagreeing annotations [29]. The model is able to learn the task as well as individual worker labeling biases [7].

Recent approaches exploit self-supervised learning to provide strong initializations which are less affected by the label noise of crowds [9]. Contrastive pretraining acquires feature representations in the unlabeled domain that generalize well to the noisy crowd sourced setting [9]. Thus, such a self-supervision and noise-robust way to train model can effectively learn from imperfect crowd annotations [33].

6. Applications in Medical Imaging

6.1 Challenges in Medical Image Annotation

Annotating medical images requires domain experts and poses the challenge to create large-scale data, which is very costly [4]. There is limited time available from Radiologists and pathologists for annotating the datasets leading to bottleneck in dataset generation [23]. Inter-observer variability (in which experts fail to agree on the ground truth) results in intrinsic label noise even in curated datasets [4].

In many medical image analysis tasks there are subtle and hard to-localize findings which renders weak supervision inherent [22]. For example, chest X-ray diagnosis can be formed in the image level but to localize certain pathologies would need minute annotation [4]. Whole slide histopathology images are composed of millions of cells, which makes exhaustive annotation impractical [24].

6.2 Weakly Supervised Medical Image Analysis

The focus of computational pathology shifted towards the multiple instance learning paradigms [22]. Then WSIs are split into patches (considered as bags) with their slide-level diagnostic annotations [24]. MIL with attention learns to find diagnosis specific regions without specifying pathologists for their annotations [10].

Weakly supervised lesion detection in medical imaging leverages image-level disease labels to learn models that localize the pathological findings [4]. CAM-based methods arouse suspicious regions in the chest X-rays, CT and MRI [51]. These estimates are interpretable to the clinicians, who can validate the model rationale [52].

Bootstrapping is of critical importance to noisy label learning in medical imaging since expert annotations frequently conflict [23]. Annotator uncertainty modelling and learning from multiple noisy annotations makes the model more robust compared to assuming that single annotations are ground truth [29]. Generative, Bayesian approaches that model uncertainty in the model and annotations allow principled reasoning under imperfect supervision [34].

6.3 Label Efficient Medical AI

Semi-supervised learning is achieved by supplementing small sets of labeled medical images with large amounts of unlabeled data [46]. Consistency regularization incentivizes predictions that are invariant to augmentations of same image, making efficient use of

unlabeled data [60]. Pseudo-labeling is used to produce training signal from model predictions on unlabeled datasets and extend empirical support for the training set [33].

Active learning (AL) asks queries (selectively requests annotations) in "the quite informative" places to the maximize label efficiency[2]. Uncertainty-based selection selects images at which the model is least confident, and diversity-based methods guarantee that enough variety of samples are included in diverse regions of the input space [39]. They work by mitigating the annotation cost by directing expert attention towards valuable examples [11].

7. Applications in Web-Scale Datasets

7.1 Automated Web-Scale Annotation

Web-UVD builds on Web-vision datasets that utilize automated annotation from meta-data, alt-text and surrounding text [1]. If image collection is unlimited, it allows the creation of datasets containing billions of images, but introduces a great deal label noise due to irrelevant or incorrect metadata [5]. Social media hashtags are inherently noisy in terms of image content that they describe, because they often convey sentiment or context instead of actual visual content[7].

Image-text matching in vision-language pretraining inherently addresses noisy correspondence via contrastive learning [9]. CLIP is trained with 400M image-text pairs from the internet, showing that scaling up with noisy supervision can be competitive with task curated datasets [9]. The contrastive loss is naturally tolerant to certain noises since the positive pairs must be similar to randomly sampled negatives. [9]

7.2 Handling Web-Scale Label Noise

Curriculum learning methods learn from easy examples to hard (or noisy) ones in the sense of introducing them step by step[40]. This method is based on the key insight that clean samples are often simpler to classify (due to which they prioritize high-quality training data in noise-infused examples) [32]. Self-paced learning generalizes this to instead let the model itself choose, in light of current ability, which examples to learn from [40].

Bootstrapping [9, 33] work iteratively enhances web-scale datasets (starting from zero or a small seed) by employing trained models to "relabel" or filter the1 training data. The model trained with the noisy inputs may still learn useful representations that support better noise detection in subsequent passes through the data [46]. This iterative purification gradually enhances the quality of the dataset without human intervention. [45].

Data augmentation acts as an implicit regularizer against noise on labels by the enlarged effective training set size and invariance towards samples [60]. For strong augmentation policies (e.g. mixup and cutmix) a model is encouraged to learn more robust features that are not tied as closely to potentially incorrect labels [38]. These techniques were crucial for learning from noisy web data [33].

8. Theoretical Perspectives and Guarantees

8.1 Learnability under Label Noise

Theoretical studies about learning with label noise examine when it is possible to learn good classifiers despite incorrect training labels [29]. Based on the NTM model, when noise-rates are less than a threshold for various classes, the optimal classifier is preserved [30]. Such a fact indicates that label noise is admissible to a certain degree, above which it changes the nature of the learning problem [7].

Statistical Consistency [34] A related question is whether the learning algorithm will converge to the correct classifier as sample size increases despite label noise? Many classical algorithms (such as logistic regression) continue to be statistically consistent in presence of symmetric noise [29]. However, the guarantees of consistency often require assumptions on the noise distribution which can be unrealistic in practice [30].

8.2 Sample Complexity

Label noise raises sample complexity i.e., the least amount of data that is needed to be observed in order to learn a True model [31]. The relationship between the noise rate and sample complexity is a function of both the learning algorithm, data distribution, and the properties of noise [7]. Theoretical lower bounds direct that sample complexity increases exponentially with noise rate in the worst case [29].

Active learning and selective sampling can help mitigate the increased sample complexity by directing labeling efforts for informative data [2]. The second case where clean and noisy labels are separable (albeit sub optimally) is known as selective learning and such framework has shown better sample complexity 2 than passive labels [39]. The findings of this paper will inspire practical prefilters to reduce label noise in learning [11].

9. Challenges and Future Directions

9.1 Open Problems

There are several issues which need to be addressed in weakly supervised and noisy learning. However, executives of weak supervision still need to be manually defined for tasks [3]. It is unclear how the tradeoff between reducing annotation costs and model quality changes across domains [2]. P. Warm Up Joint Learning It would be interesting to investigate if Weakly Supervised Object Localization with the output as a by-product could help joint learning [55] of multiple types of weak supervision (e.g., image labels, points, scribbles) in principled ways.

The extreme label noise, i.e. in presence of corruption rates approaching the 50% are a critical issue for current approaches [7]. Systematic biases are hard to discern from random noise but become very important when handling the data properly [29]. Adversarial label noise, for instance the one where corruptions are designed to impair model performance, calls for more robust defenses beyond standard noise handling approaches [31].

9.2 Emerging Directions

State-of-the-art pretrained models on large, noisy web data show remarkable ability to tolerate imperfect supervision [9]. How scale and pretraining help with label noise may have implications for more efficient training strategies [31]. Few-shot or zero-shot learning may also help to alleviate reliance on large-scale annotation when stripping down from foundation models [9].

For noise-robust model, neural architecture search may find a more label-corruption-free orphan [34] with better intrinsic robustness. Automated learning of augmentation policies may also detect transformations that yield maximum regularization against specific types of noise [60]. One potential approach is through human-AI collaboration frameworks, where model predictions are judiciously integrated with expert annotations to achieve the ideal balance of accuracy and cost [2].

The weakly supervised causal methodologies may separate between correlation and causation in the label noise, which could lead to more robustness [29]. Fairness-sensitive v weakly supervised learning is required to make sure that label noise does not adversely affect the model's performance towards a minority [7]. For weakly supervised learning, techniques that enable privacy (keeping) collaboration but not sharing potentially sensitive detailed annotations have also been proposed [8].

9.3 Evaluation and Benchmarks

Controlled noise benchmarks that are standardized across multiple methods are necessary to ensure that the comparison of noise-robustness between different methods is rigorous [1]. Artificial noise injection allows for a reproducible evaluation, but it might not resemble real world noise patterns [30]. The real world naturally noisy datasets are realistic evaluation, while the clean ground truth for improvement does not always exist [5].

Beyond accuracy, we require evaluation metrics for robustness, calibration and interpretability under weak supervision [34]. Methods should not be judged purely based on final performance but also on annotation efficiency, training stability and computational cost [2]. They would all benefit from wide baselines that cover multiple domains and different types of imperfect supervision [3].

10. Conclusion

Learning from imperfect supervision has been recognized as a key ability for scaling machine learning to real-world applications where clean and full labeled data is lacking or too costly. MIL offers a principled way for learning from bag-level labels, and attention-aware deep MIL has achieved great success in the context of medical imaging and other applications. Methods for dealing with label noise, robust loss functions and sample selection and noise models allow for training on large noisy datasets while retaining strong generalization.

Weakly supervised object detection and segmentation can alleviate the annotation requirement by learning from coarse labels such image tags or scribbles instead of exact boundaries. Crowdsourced label aggregation techniques can obtain accurate supervision from multiple noisy annotators thanks to probabilistic modeling and end-to-end training. Medical imaging applications utilize weak supervision to lower expert annotation overhead, whereas web-scale datasets show that large, noisy training data can even compete with carefully-generated supervision.

A theoretical analysis yields learnability and sample complexity under imperfect supervision; the results are used to design algorithms. Nonetheless, there are still major obstacles in extending to strong noise, multiple supervision types' fusion, fairness and privacy. In the future, we will also explore better use of foundation models for improving robustness as well as developing a framework of human-AI collaboration and constructing more comprehensive benchmarks to evaluate them rigorously.

Such models would be necessary as machine learning systems operate more and more in data-rich but annotation-poor regimes where it will become imperative to learn from imperfect supervision. The techniques we discuss is section 4 make progress toward this ideal, but challenge us to continue developing such methods in order to unleash the full potential of learning when ground truth labels are not perfect.

References

[1] Deng, J., et al. (2009). ImageNet: Large scale hierarchical image database. In IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255.
[2] Papadopoulos, D.P., et al. (2017). Click and Aspect: towards constructing an object-detector from image-level annotations. In: Proceedings of the IEEE/IAPR Int't IConf on Computer Vision, 4930-4939.
[3] Zhou, Z. H. (2018). A short gunman on weakly supervised learning. National Science Review, 5(1), 44-53.
[4] Rajpurkar, P., et al. (2017). CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning. arXiv preprint arXiv:1711.05225.
[5] Northcutt, C., et al. (2021). Confident learning: Estimating uncertainty for accurate dataset labels. Journal of Artifical Intelligence Research, 70, 1373-1411.
[6] Dietterich, T. G., et al. (1997). Multiple-instance problem solved using axis-parallel rectangles. Artificial Intelligence, 89(1-2), 31-71.
[7] Song, H., et al. (2022). Learning with noisy labels: A survey. IEEE Transactions on Neural Networks and Learning Systems 34(11), 8135-8153.
[8] Ratner, A., et al. (2017). Snorkel Snorkel: rapid training data creation with weak supervision. Proceedings of the VLDB Endowment, 11(3), 269-282.
[9] Chen, T., et al. (2020). Contrastive learning of visual representations in less than an hour. International Conference on Machine Learning, 1597–1607.
[10] Ilse, M., et al. (2018). Attention-based deep multiple instance learning. In International Conference on Machine Learning, 2127--2136.
[11] Ren, M., et al. (2018). Provable robustness of relu networks via maximization of linear regions. In Proceedings of the International Conference on Machine Learning, 4334-4343.
[12] Dietterich, T. G., et al. (1997). Multiple-instance learning with axis-parallel rectangles. Artificial Intelligence, 89(1-2), 31-71.
[13] Maron, O., & Lozano-Pérez, T. (1998). A framework for multiple-instance learning. In Advances in Neural Information Processing Systems, 10.
[14] Amores, J. (2013). Multiple instance learning: A survey and comparative study. Artificial Intelligence, 201, 81-105.
[15] Foulds, J., & Frank, E. (2010). A survey of multi-instance learning assumptions. Knowledge Engineering Review, 25(1), 1-25.
[16] Carbonneau, M. A., et al. (2018). Multiple instance learning: Foundations and algorithms Multiple instance learning: A survey of problem characteristics and applications. Pattern Recognition, 77, 329-353.
[17] Maron, O., & Ratan, A. L. (1998). Muilti-bag learning: a bag of local features model for natural scene classification. In International Conference on Machine Learning, 341-349.
[18] Andrews, S., et al. (2003). Multiple-instance Learning with Support Vector Machines. In: Advances in Neural Information Processing Systems, 15.
[19] Wu, J., et al. (2015). Deep multiple instance learning for image classification and auto-annotation. IEEE Conference on Computer Vision and Pattern Recognition, 3460-3469.
[20] Vaswani, A., et al. (2017). Attention is all you need. Advances In Neural Information Processing Systems, 30.
[21] Wang, X., et al. (2018). Revisiting multiple instance neural networks. Pattern Recognition, 74, 15-24.
[22] Campanella, G., et al. (2019). Clinical-grade computational pathology with weakly supervised deep learning on whole slide images. Nature Medicine, 25(8), 1301-1309.
[23] Krause, J., et al. (2016). Grader Variability and the Importance of Reference Standards for Evaluating Machine Learning Models for Diabetic Retinopathy. Ophthalmology, 125(8), 1264-1272.
[24] Courtiol, P., et al. (2019). Deep learning of the ARCHITECT mesothelioma tissue polypeptide antigen assay enhances the ability to predict survival in mesothelioma patients. Nature Medicine, 25(10), 1519-1525.
[25] Wei, X. S., et al. (2018). Selective Convolutional Descriptor Aggregation for Fine-Grained Image Retrieval. IEEE transactions on image processing : a publication of the IEEE Signal Processing Society, 26(6), 2868-2881.
[26] Zhou, Z. H., et al. (2012). Multi-instance multi-label learning. Artificial Intelligence, 176(1), 2291-2320.
[27] Sultani, W., et al. (2018). Anomaly detection in surveillance video signals from real-world scenario. Computer Vision and Pattern Recognition, 6479--6488.
[28] Angelidis, S., & Lapata, M. (2018). Multiple instance learning networks for fine-grained sentiment analysis. Transactions of the Association for Computational Linguistics, 6, 17–31.
[29] Frénay, B., & Verleysen, M. (2014). Learning with label noise: A survey. IEEE Transactions on Neural Networks and Learning Systems, 25(5), 845–869.
[30] Patrini, G., et al. (2017). Deep neural networks resistant to label noise via iterative weak learning. Computer Vision and Pattern Recognition, 1944–1952.
[31] Zhang, C., et al. (2017). Making sense of deep learning requires rethinking generalization. International Conference on Learning Representations.
[32] Arpit, D., et al. (2017). On the state of the art of evaluation in neural network compression. ICML'05: Proceedings of the 22nd international conference on Machine learning, 233-242.
[33] Li, J., et al. (2020). DivideMix: Learning uses model inductive biases for semi-supervised learning. International Conference on Learning Representations.
[34] Ghosh, A., et al. (2017). Stronger generalization for a class of K-tied transformers: model and representational benefits. AAAI Conference on Artificial Intelligence, 31 (1).
[35] Wang, Y., et al. (2019). Symmetric cross entropy Loss for robust learning with noisy labels. Computer Vision and Pattern Recognition, 1-8.
[36] Ma, X., et al. (2020). Learning with noisy labels Learning with Bad Training Data We begin by giving some results on the generalization of deep networks to investigate learning accurate models from a noisy labeling of data. In International Conference on Machine Learning, 6543–6553.

[37] Zhang, Z., & Sabuncu, M. R. (2018). The generalised cross entropy loss for training deep neural networks with noisy labels. Proceedings of Neural Information Processing Systems, 31.
[38] Lin, T. Y., et al. (2017). Focal loss for dense object detection. International Conference on Computer Vision, 2980-2988.
[39] Han, B., et al. (2018). Co-teaching: Robust training of deep neural networks with extreme noisy labels. Advances in Neural Information Processing Systems, 31.
[40] Jiang, L., et al. (2018). MentorNet: Learning Data-Driven Curriculum for Very Deep Neural Networks on Corrupted Labels. In International Conference on Machine Learning, 2304--2313.
[41] Shu, J., et al. (2019). Meta-weight-net: Learning an explicit mapping for sample weighting. In: NeurIPS 32, pp.7475–7484.
[42] Liu, S., et al. (2020). Early-learning regularization inhibits noise-to-label memorization. (2020) Advances in Neural Information Processing Systems 33:20331-20342.
[43] Xia, X., et al. (2019). Are anchor points truly necessary in label-noise learning? In: NeurIPS 32:Proceedings of the 32nd Conference on Neural Information Processing Systems.
[44] Yao, Y., et al. (2020). Dual T: Training noisy labels with certified confidence in transition matrix. Parenas, I., Espinel, A., Chollet, C.: Problem solving with weak supervision.
[45] Tanaka, D., et al. (2018). Co-regularized learning with noisy labels. Computer Vision and Pattern Recognition, 5552–5560.
[46] Arazo, E., et al. (2019). Unsupervised label noise modeling and loss correction for text classification. In International Conference on Machine Learning, 312-321.
[47] Wei, H., et al. (2020). Noisy Label and Recovery by Consensus: [[https://arxiv.org/pdf/1803.05277.pdf]] [Code not provided in paper, sent author an email] A Unified Framework for Generalized Multi-class Object Recognition. Computer Vision and Pattern Recognition, 13726-13735.
[48] Bilen, H., & Vedaldi, A. (2016). Weakly supervised deep detection networks. C omputer V ision and Pattern Recognition, 2846?2854.
[49] Tang, P., et al. (2017). MIDN with online instance classifier refinement. In Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, 2843-2851.
[50] Uijlings, J. R., et al. (2013). Selective search for object recognition. International Journal of Computer Vision, 104(2), 154-171.
[51] Zhou, B., et al. (2016). Discriminative localization in CNNs for weakly-supervised segmentation of pulmonary nodules. Computer Vision and Pattern Recognition, 2921--2929.
[52] Selvaraju, R. R., et al. (2017). Grad-CAM: Why did you say that? Visual Explanations from Deep Networks via Gradient-based Localization. International Conference on Computer Vision, 618-626.
[53] Tang, P., et al. (2017). Online instance classifier refinement in a multiple instance detection network. Computer Vision and Pattern Recognition, 2843-2851.
[54] Tang, P., et al. (2018). PCL: Proposal cluster learning for weakly supervisedobject detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 42(1), 176-191.
[55] Bearman, A., et al. (2016). What to learn: Semantic segmentation with point supervision. In European Conference on Computer Vision, pages 549–565.
[56] Kolesnikov, A., & Lampert, C. H. (2016). Sow, re-grow and trim: Three principles for weakly-supervised image segmentation. European Conference on Computer Vision, 695--711.
[57] Ahn, J., and Kwak, S. (2018). Pixel-level semantic affinity learning via image-level supervision for weakly supervised semantic segmentation. Computer Vision and Pattern Recognition, 4981–4990.
[58] Huang, Z., et al. (2018). Deep seed region growing for weakly supervised semantic segmentation. Computer Vision and Pattern Recognition, 7014–7023.
[59] Krähenbühl, P., & Koltun, V. (2011). Efficient inference in fully connected CRFs with Gaussian edge potentials. In Neural Information Processing Systems 24.
[60] Lee, D. H. (2013). Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks