A Survey of Un-, Weakly-, and Semi-Supervised Learning Methods for Noisy, Missing and Partial Labels in Industrial Vision Applications

Niclas Simmler*, Pascal Sager†, Philipp Andermatt‡, Ricardo Chavarriaga†, Frank-Peter Schilling†, Matthias Rosenthal†, and Thilo Stadelmann†§

*HSLU
Lucerne, Switzerland
niclas.simmler@stud.hslu.ch

†ZHAW Datalab
Winterthur, Switzerland
{sage, chav, scik, rosn, stdm}@zhaw.ch

‡KITRO SA
Zurich, Switzerland
philipp.andermatt@kitro.ch

§ECLT
Venice, Italy
Fellow

***Abstract*—When applying deep learning methods in an industrial vision application, they often fall short of the performance shown in a clean and controlled lab environment due to data quality issues. Few would consider the actual labels as a driving factor, yet inaccurate label data can impair model performance significantly. However, being able to mitigate inaccurate or incomplete labels might also be a cost-saver for real-world projects. Here, we survey state-of-the-art deep learning approaches to resolve such missing labels, noisy labels, and partially labeled data in the prospect of an industrial vision application. We systematically present un-, weakly, and semi-supervised approaches from 'A' like anomaly detection to 'Z' like zero-shot classification to resolve these challenges by embracing them.**

***Index Terms*—deep learning, computer vision, label quality**

I. Introduction

Utilizing machine learning in an industrial application poses additional challenges compared to research lab environments [1], [2], e.g., in the form of data quality and data quantity issues [3]. "Garbage in, Garbage out" is an often stressed dictum in machine learning – even more so in industrial applications, where data samples and labels collection is difficult and costly [4]. Supervised learning approaches for deep neural networks not only require large amounts of data but also reliably labeled ones. Usually, the data labeling process is conducted manually by individual experts and may involve complex decisions based on years of expert training. To scale-up data labeling in a cost-efficient manner, this process is increasingly outsourced to external contractors (the market for data labeling is expected to reach USD 1.2bn by 2023 [5]). These factors can compromise the labeling quality, often leading to incomplete or uncertain, i.e., *noisy labels*. Incomplete labels can constitute *partially labeled* samples or altogether *missing labels* in a dataset.

To illustrate the importance of clean labels, Fig. 1 shows how model performance is affected by noisy labels. Here, we randomly flipped the labels of a proportion of training data for an exemplary binary classification task in visual quality control. The results show robustness to low proportions of noisy labels (less than 20% noise). However, with increased noise the performance rapidly decreases, reaching a level of guessing if more than 30% of the labels are noisy.

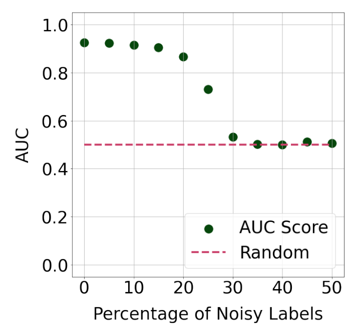

Fig. 1. Exemplary binary classification results showing the effect of increasing label noise on model performance (AUC [6]) in a real-world visual quality control application [7].

This paper presents a survey of methods aimed at coping with the aforementioned challenges related to data labeling. We focus on methods for computer vision in industrial applications, comprising both image classification and semantic segmentation. The surveyed methods include approaches based on unsupervised, weakly supervised, and semi-supervised learning that lend themselves to real-world applications.

II. Problem Description

The issue of data quality can be caused by factors such as (i) inaccurate or inappropriate sensors causing inconclusive readings; (ii) human mistakes while performing repetitive tasks during data collection; or (iii) financial restriction and too tightly scheduled time-frames for data collections. Frequently, data sample quality (i.e., sensor readings or image quality) is the predominantly optimized part of the data collection process. However, the often neglected aspect of label quality manifests in the three issues of *missing*, *noisy* or *partial labels*. They are the focus of this paper and described in the following paragraphs. Per issue, we present solutions from the following domains, where applicable (see Tab. I):

TABLE I
OVERVIEW OF IDENTIFIED ISSUES AND THEIR RESPECTIVE STATE-OF-THE-ART SOLUTIONS.

Issue	Solutions	Supervision	Key literature
Missing labels	Contrastive learning, clustering	Unsupervised	[8]–[11]
Missing labels	Label propagation, label regularization, game theory, self-training	Semi-supervised	[12]–[19]
Noisy labels	Mixture modeling, collaborative unsupervised domain adaption	Unsupervised	[20]–[23]
Noisy labels	Confident learning, rank pruning	Weakly supervised	[24], [25]
Noisy labels	Outlier/anomaly detection, mixture modeling	Semi-supervised	[26]–[30]
Partially labeled data	Change detection, multi-task learning, multi-instance learning, refinement of object-based class-activation maps	Weakly supervised	[31]–[39]

unsupervised learning (learning from unlabeled data); *semi-supervised learning* (training on a small subset of labeled data first and subsequently utilizing similar unlabeled data); and *weak supervision*, which however is an umbrella term [40] for (i) *incomplete* supervision (e.g., when not all objects in an image are labeled), (ii) *inexact* supervision (e.g., coarse or loose markings of defects), and (iii) *inaccurate* supervision (e.g., labels containing mistakes).

Missing labels: Resource availability, including financial resources or access to expert labels, constrain real-world projects. This often leads to compromises, e.g., to label only a subset of the data and use machine learning methods to increase the labeled proportion of the data. Various approaches to overcome missing labels are described in Sec. III.

Noisy labels: To speed up the collection process, data labeling is often parallelized through the use of multiple experts (or observers). This approach has multiple benefits as (i) it can speed up the labeling process by splitting the work and save time; (ii) depending on the setup, multiple observers can cross-validate results to ensure quality; and (iii) if appropriately applied, multiple opinions on the same sample can mitigate misjudgment and increase trust in the labels. However, this kind of crowdsourcing has its drawbacks since observers may be unreliable and biased [41]. Manual labeling by multiple experts can also cause disagreement in how a sample is labeled (i.e., *inter-observer variability* [42]), especially when the subject of labeling involves considerable study of the sample. This issue will manifest itself in an uncertainty on the label itself. We cover approaches to resolve respective label noise in Sec. IV.

Partially labeled data for semantic segmentation: Fully supervised training of image segmentation models requires pixel-level annotations, i.e., assigning a semantic label (e.g., "carrot", or "person") to every single pixel in the image. While detailed pixel-level annotations would yield better models, this is a very time-consuming, hence expensive, process. Alternatively, weak annotations such as scribble annotations [43], point annotations [44], bounding boxes, or image-level labels can be used. Collecting bounding boxes is about 10–15 times faster/cheaper than pixel-level annotation [44], [45]. Image-level labeling, point annotation, and scribble annotation take even less time (around 1-2 seconds per image) [44]. Approaches for reduced labelings are surveyed in Sec. V.

III. MISSING LABELS

Typical approaches to resolve missing label issues are based on the idea of "finding similar samples", i.e., contrastive learning (unsupervised), or "label propagation" (semi-supervised).

In the *unsupervised* contrastive learning [11] approach, a model is trained to discriminate between similar and different images that are all derived from the same unlabeled data by mere augmentation (picking two random images yields a dissimilar pair, taking an image and its augmented version produces a matched pair). In order to classify images using the learned representations, fine-tuning the model using a tiny set of labeled data is required. In recent years, the performance of these methods has increased significantly. The SimCLR [8] framework is able to outperform supervised methods on ImageNet [46]. However, these results may not apply to industrial applications, as they can only be learned with very large batch sizes, i.e., a lot of data and long training times. SimCLRv2 [9] is even larger and more complex. Subsequent knowledge distillation shrinks the model and simplifies deployment. Another approach based on unlabeled data is to utilize a clustering algorithm to group similar features. One of the most recent works in this area is SwAV [10]. This method predicts the cluster assignment of a view from the representation of another view. Compared to SimCLR, SwAV achieves a slightly higher score on ImageNet. Its disadvantage is the higher complexity, as not only are two views compared, but all of them are clustered.

For a thorough evaluation of *semi-supervised* learning in the area of image classification, see Ref. [47]. The approaches used are typically split into two categories: (i) the addition of an unsupervised loss term, or (ii) the assignment of pseudo-labels to the unlabeled examples. Popular examples in the first category are the "consistency loss" between the outputs of a network on random perturbations of the same image [12], or the "mean teacher" method [48], which replaces output averaging by averaging of network parameters. The second category uses the regular classifiers to infer pseudo-labels of unlabeled examples by choosing the most confident class [13], [49]. These pseudo-labels are treated like standard labels in the cross-entropy loss. More recently, the "noisy student" training [14] showed improved ImageNet classification performance by training using an EfficientNet [50] model.

Another method to infer the unknown labels is label propagation, where labels of labeled samples are propagated to unlabeled samples in close proximity (defined relationally or in terms of similarity). In Ref. [15], it is performed on a large image dataset with convolutional neural network (CNN) [51] descriptors for few-shot learning (FSL). Unseen images are classified via online label propagation, which requires storing the entire dataset while the network is trained in advance and descriptors are fixed. In Ref. [16], label propagation on the training set is performed offline while training the network, such that inference is possible without accessing the original training set. A transductive label propagation method is used, based on the manifold assumption (i.e., that similar examples should get the same prediction), to make predictions on the entire dataset and use them to generate pseudo-labels for the unlabeled data for training. The authors improve the performance on several datasets, especially in the few-labels regime. Ref. [17] proposes a Transductive Propagation Network (TPN) that performs end-to-end labeling of unlabeled images. The network performs the feature extraction using a standard CNN and the graph construction in one. A benchmark on *miniImageNet* [52] and *tieredImageNet* [53] shows superior performance compared to other state-of-the-art FSL algorithms especially using zero to five shots, which means it works especially well if only few labels are available. However, a typical issue with FSL is that the training and test samples are disjoint [18]. This causes the feature extractor of a TPN to produce embeddings that are seemingly uncorrelated for unseen classes. This manifests as a disadvantage in terms of robustness when the TPN tries to propagate the labels during graph construction. The Embedding Propagation Network (EPNet) [18] addresses this shortcoming of TPNs by applying the propagation at embedding creation time, thus locating an image's embedding close to images with similar features in embedding space, resulting in closer labels in their respective space. EPNet achieves superior performance over the TPN architecture in one- and five-shot benchmarking.

The Graph Transduction Game (GTG) [19] is a popular method in the category of label propagation and can be seen as a special case of relaxation labeling [54]–[56], which in turn addresses the problem of label disambiguation. GTG's general idea is to propagate contextual (i.e., relational) information of labeled instances to classify the unlabeled ones consistently. While, in general, label propagation methods are based on graph Laplacian regularization, GTG is based on non-cooperative game theory. It has been used for the determination of pseudo-labels [57], however in this case, the network is pre-trained, such that the graph remains fixed and there is no weighting mechanism. In general, recent years have seen a steep rise in the application of graph neural networks [58], [59], including graph convolutional neural networks [60] to solve problems which can be represented using graph-structured data. Knowledge graphs can be used as extra information to guide zero-shot classification [61], [62]. The similarity between images in the dataset is also useful in the case of few-shot learning [52]. Ref. [63] proposes to build a weighted fully-connected image network based on similarity and perform message passing in the graph for few-shot recognition. Ref. [64] selects some related entities to build a sub-graph based on object detection results and applies a gated graph neural network to the extracted graph for prediction. Finally, Ref. [65] proposes to build a knowledge graph where the entities are the different categories.

The abundance of solutions regarding missing labels promises that this issue can be resolved in real world tasks.

IV. Noisy Labels

Despite that neural networks exhibit certain robustness towards label noise (cp. Fig. 1 and Refs. [20], [66]), the problem of noisy labels is striking. Various methods in all three learning categories exist to identify and resolve this issue.

Ref. [20] introduces an *unsupervised* approach while suggesting that noisy labeled samples are harder to learn by a model than correctly labeled ones. This allows for noise identification by fitting a mixture model on the loss values and subsequently using the model's posterior probabilities to identify noisy labels. Unfortunately, the authors have not yet been able to replicate the modeling approach's performance on other datasets apart from CIFAR-10 & CIFAR-100 [67]. Collaborative unsupervised domain adaption [21] can mitigate label noise in an unsupervised manner when applied on a real-world dataset. The approach is based on unsupervised domain adaption, which aims to transfer knowledge from a labeled source domain to an unlabeled target domain [22]. The authors evaluate their method by benchmarking it on a medical image diagnosis dataset consisting of H&E stained colon histopathology slides [23] with a convincing performance.

An approach rooted in the area of *weak supervision* is based on Confident Learning (CL) [24], where the authors do not focus on a particular loss function or model architecture. It uses out-of-sample prediction probabilities that are obtained using cross-validation on noisy labeled data. CL performs (i) the estimation of the joint distribution of noisy labels and true labels, (ii) the identification and pruning of noisy samples, and (iii) the re-training and re-weighting of samples with a new estimated latent prior to identify label noise in a dataset. The main advantage of CL is the absence of hyperparameters and that it does not require guaranteed clean labels. As a result, the authors can test on many publicly available datasets for label noise using CL and, e.g., found an abundant amount of label errors in ImageNet, CIFAR, and even the MNIST dataset [68]. CL is available as an open-source Python package.

Similar to Ref. [20], but in a *semi-supervised* manner, is the approach followed by the authors of Ref. [27]. The proposed DivideMix architecture models the loss on a sample level using a mixture model to separate clean (labeled) samples from noisy (unlabeled) samples. To prevent confirmation bias, the authors propose to train two networks simultaneously, both generating the sample split for the other network for further training. Both networks then co-refine and co-guess on labeled and unlabeled samples to improve with each iteration. DivideMix achieves outstanding performance on CIFAR-10 and CIFAR-100 and thus outperforms Ref. [20].

Another prominent semi-supervised methodology is to model the problem of noise detection as anomaly (or: outlier) detection (AD) [28]. Ref. [26] proposes a respective conditional variational autoencoder (CVAE) system fit for real-world applications. Benchmarked on both MNIST and the newer MNIST fashion dataset [69], it compares favourably with other well-known AD algorithms like ν-SVM [29] and Isolation Forest (IF) [30]. The authors also present a real-world application using data collected from CMS experiment at the CERN Large Hadron Collider [70]. Despite the absence of hyperparameter optimization, their method shows great performance compared to ordinary methods on both the artificial datasets and real-world data from CERN.

A promising procedure that can be fitted to any method during training is called **S**harpness-**A**ware **M**inimization (SAM) [71]. SAM aims to find model parameters during gradient-descent that produce consistently flat minima [72] and thus can improve any training on noisy labels.

The availability of potential real-world solutions in all three surveyed supervision domains stresses this issue's feasibility to be resolved.

V. Partially Labeled Data

Here, we survey approaches for learning pixel-level classifications from image-level labels. Some of these approaches formulate the problem as a multi-instance learning (MIL) problem [73], where a multi-class MIL loss is designed for training the network [33], [34]. Generally, *weakly supervised* approaches try to infer predictions with high information content from labels with low information content. This is especially interesting for semantic segmentation, where generating pixel-level ground truth data is very time-consuming and labor-intensive. Likewise, image-based change detection has similar properties, with the additional challenge of comparing two images and identifying relevant differences.

Ref. [32] uses a directed acyclic graph (DAG) to perform weakly supervised change detection on image series. On top of the DAG, a conditional random field (CRF) model is defined [74] that helps to refine the change mask. Generally, all weakly supervised approaches face the challenge that predictions are coarse. Thus, using a CRF as post-processing is an easy and effective way to increase performance. Ref. [35] proposes a simple architecture that can leverage information from different annotations such as image-level labels, bounding-box labels, and pixel-level labels to improve a semantic segmentation network. The authors use an arbitrary, fully convolutional network to predict the segmentation masks. Afterward, a segmentation mask is fed in an annotation-specific loss module. Depending on the label's form, a different loss function is applied to improve the segmentation network. They can show that this method can effectively make use of training data with different levels of supervision.

Other approaches use a two-stage approach that generates object-based labels from class activation maps and then trains segmentation networks based on those maps. The class-activation map (CAM) [75] method uses global average pooling, typically applied as a structural regularizer for CNNs, to identify discriminative image regions. This allows using weakly supervised object localization based on networks trained with image-level labels. The coarse response maps are exploited to perform image segmentation using different approaches, including (i) the use of CAMs as the supervisory signal [36], (ii) progressive region refinement based on iterative mining of features [37], or (iii) learning pixel affinity to identify significant regions or propagate pixel-wise information [38], [39]. Further, graph neural networks can be employed to utilize non-local information in the images. In Refs. [76], [77] for example, a Graph-LSTM model is presented to incorporate long-term dependencies together with spatial connections by building graphs and apply the LSTM to propagate neighborhood information globally. Similar ideas have been applied in the case of 3D semantic segmentation and point cloud classification (see e.g., Refs. [78], [79]).

Ref. [31] proposes the weakly supervised change detection method W-CDNet that can be trained with image-level labels. It uses a siamese architecture [80] to compare features from two different images. The change segmentation and classification (CSC) module forces the model to highlight relevant changes. It consists of a custom remapping block that enforces a strong separation between background and foreground pixels, a CRF-RNN layer [81] that refines the change mask, and a classifier that predicts the image-level classification for the image pair. W-CDNet can be trained with both binary image-level labels (describe whether the image contains any relevant change at all) and semantic image-level labels (classify the relevant changes), as well as with full supervision. W-CDNet makes use of existing architectures like U-Net [82] and VGG16 [83], which allows the use of pre-trained weights to speed up the training process. One disadvantage of this method is that the predicted change mask is always a binary mask and not a semantic segmentation mask. Thus, in order to perform semantic change detection, one would have to employ an additional semantic segmentation network on top of the change detector.

This survey suggests that one can achieve similar performance using weaker labels than full (pixel-level) annotations for semantic segmentation tasks.

VI. Discussion and Conclusions

Our survey on the issues of missing labels, noisy labels, and partially labeled data in real-world applications of computer vision shows many potential solutions and plenty of active research. However, our survey also shows that there is no silver bullet for either issue: it all depends on the application setting. We conclude by formulating four hypotheses for the further adoption of deep learning in industrial practice:

Shift towards real-world benchmarks: Although various approaches have been tested on standard and artificial datasets, many have not yet seen a noteworthy real-world application. Nevertheless, we could find a few recent candidates in most categories. Overall, this trend [84] in the research community towards more *in*accurate and real-world-oriented datasets and benchmarks is promising and further underlines the importance of the issues covered in this paper. Traditionally, deep learning models rely on high-quality and large datasets. However, some of the presented methods allow for the potential application of these very models on scarce and unreliable data. This paradigm shift allows the introduction of deep learning methods in many more fields, which have been untouched by modern machine learning so far.

Missing labels sufficiently addressed: We covered solutions for missing labels in great detail and identified a large amount of promising research work in both the unsupervised and semi-supervised domains. The existence of this large number of potential solutions demonstrates that this specific issue has been considered a significant pain point in the community. Obtaining a large amount of data is hard enough, but labeling a sufficient amount of it is even harder, especially in a corporate setting compared to a global effort of volunteers. The surveyed solutions suggest that this problem can be mitigated.

Allowing label noise might pay off: We showed that there are unsupervised, weakly-supervised and semi-supervised methods to counteract noisy labels. Even though noisy labels are rightfully feared when applying deep learning models to real-world data, it is assuring that there are real-world proven methods to overcome the issue. If the intentional admittance of noisy labels in a dataset introduces a quicker turnaround on the data collection process, it can reduce costs and save valuable project time.

Pixel-level information potentially not necessary: When dealing with partially labeled data, we presented promising weakly supervised methods. For a real-world application, this implies that it might be easier and faster to label samples using weaker annotations (e.g., scribbles, image-level labels) rather than enforcing exhaustive labeling. Thus, relaxing the labeling requirements can be a considerable cost- and time-saver.

Given the number of potential solutions to crucial but sometimes overlooked (or: underrated) problems in recent years, we expect a greater adoption in deep learning applications in years to come, as the feasibility might incline more and more industries to introduce modern computer vision in their everyday processes.

Acknowledgments

This work is financially supported through Innosuisse grants 36777.1 IP-ICT "FWA" and 26025.1 PFES-ES "QualitAI".

References

[1] M. Braschler, T. Stadelmann, and K. Stockinger, Eds., *Applied Data Science - Lessons Learned for the Data-Driven Business*. Springer, 2019. [Online]. Available: https://doi.org/10.1007/978-3-030-11821-1

[2] T. Stadelmann, V. Tolkachev, B. Sick, J. Stampfli, and O. Dürr, "Beyond imagenet: deep learning in industrial practice," in *Applied Data Science*. Springer, 2019, pp. 205–232.

[3] T. Stadelmann, M. Amirian, I. Arabaci, M. Arnold, G. F. Duivesteijn, I. Elezi, M. Geiger, S. Lörwald, B. B. Meier, K. Rombach *et al.*, "Deep learning in the wild," in *Proc. ANNPR*. Springer, 2018, pp. 17–38.

[4] L. Hollenstein, L. Lichtensteiger, T. Stadelmann, M. Amirian, L. Budde, J. Meierhofer, R. M. Füchslin, and T. Friedli, "Unsupervised learning and simulation for complexity management in business operations," in *Applied Data Science*. Springer, 2019, pp. 313–331.

[5] C. Metz, "A.I. is learning from humans. many humans," *The New York Times*, Aug 2019. [Online]. Available: https://www.nytimes.com/2019/08/16/technology/ai-humans.html

[6] T. Fawcett, "An introduction to roc analysis," *Pattern recognition letters*, vol. 27, no. 8, pp. 861–874, 2006.

[7] Severstal: Steel Defect Detection Dataset. Kaggle.com. December 27, 2020. [Online]. Available: https://www.kaggle.com/c/severstal-steel-defect-detection/data

[8] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *Proc. ICML*, 2020, pp. 1597–1607.

[9] T. Chen, S. Kornblith, K. Swersky, M. Norouzi, and G. Hinton, "Big self-supervised models are strong semi-supervised learners," *arXiv:2006.10029*, 2020.

[10] M. Caron, I. Misra, J. Mairal, P. Goyal, P. Bojanowski, and A. Joulin, "Unsupervised learning of visual features by contrasting cluster assignments," *arXiv:2006.09882*, 2021.

[11] A. van den Oord, Y. Li, and O. Vinyals, "Representation learning with contrastive predictive coding," 2018.

[12] M. Sajjadi, M. Javanmardi, and T. Tasdizen, "Mutual exclusivity loss for semi-supervised deep learning," in *Proc. IEEE ICIP*, 2016, pp. 1908–1912.

[13] D.-H. Lee, "Pseudo-label : The simple and efficient semi-supervised learning method for deep neural networks," *Proc. ICML Workshop: Challenges in Representation Learning*, 2013.

[14] Q. Xie, M. Luong, E. H. Hovy, and Q. V. Le, "Self-training with noisy student improves imagenet classification," in *Proc. CVPR*, 2020, pp. 10 684–10 695.

[15] M. Douze, A. Szlam, B. Hariharan, and H. Jégou, "Low-shot learning with large-scale diffusion," in *Proc. CVPR*, 2018, pp. 3349–3358.

[16] A. Iscen, G. Tolias, Y. Avrithis, and O. Chum, "Label propagation for deep semi-supervised learning," in *Proc. CVPR*, 2019, pp. 5070–5079.

[17] Y. Liu, J. Lee, M. Park, S. Kim, E. Yang, S. J. Hwang, and Y. Yang, "Learning to propagate labels: Transductive propagation network for few-shot learning," *arXiv:1805.10002*, 2018.

[18] P. Rodríguez, I. Laradji, A. Drouin, and A. Lacoste, "Embedding propagation: Smoother manifold for few-shot classification," in *Proc. ECCV*, 2020, pp. 121–138.

[19] A. Erdem and M. Pelillo, "Graph transduction as a noncooperative game," *Neural Comput.*, vol. 24, no. 3, pp. 700–723, 2012.

[20] E. Arazo, D. Ortego, P. Albert, N. E. O'Connor, and K. McGuinness, "Unsupervised Label Noise Modeling and Loss Correction," *arXiv:1904.11238*, 2019.

[21] Y. Zhang, Y. Wei, P. Zhao, S. Niu, Q. Wu, M. Tan, and J. Huang, "Collaborative Unsupervised Domain Adaptation for Medical Image Diagnosis," *arXiv:1911.07293*, 2019.

[22] W. M. Kouw and M. Loog, "A review of domain adaptation without target labels," *arXiv:1901.05335*, 2019.

[23] Y. Zhang, H. Chen, Y. Wei, P. Zhao, J. Cao, X. Fan, X. Lou, H. Liu, J. Hou, X. Han, J. Yao, Q. Wu, M. Tan, and J. Huang, "From Whole Slide Imaging to Microscopy: Deep Microscopy Adaptation Network for Histopathology Cancer Image Classification," in *Proc. MICCAI*, 2019, pp. 360–368.

[24] C. G. Northcutt, L. Jiang, and I. L. Chuang, "Confident Learning: Estimating Uncertainty in Dataset Labels," *arXiv:1911.00068*, 2020.

[25] C. G. Northcutt, T. Wu, and I. L. Chuang, "Learning with confident examples: Rank pruning for robust classification with noisy labels," in *Proc. UAI*, 2017.

[26] A. A. Pol, V. Berger, C. Germain, G. Cerminara, and M. Pierini, "Anomaly detection with conditional variational autoencoders," in *Proc. IEEE ICMLA*, 2019, pp. 1651–1657.

[27] J. Li, R. Socher, and S. C. H. Hoi, "DivideMix: Learning with Noisy Labels as Semi-supervised Learning," *arXiv:2002.07394*, 2020.

[28] B. Frenay and M. Verleysen, "Classification in the Presence of Label Noise: A Survey," *IEEE TNNLS*, vol. 25, no. 5, pp. 845–869, May 2014.

[29] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural computation*, vol. 13, no. 7, pp. 1443–1471, 2001.

[30] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation-based anomaly detection," *ACM TKDD*, vol. 6, no. 1, pp. 1–39, 2012.

[31] P. Andermatt and R. Timofte, "A weakly supervised convolutional network for change segmentation and classification," in *Proc. ACCV*, 2020.

[32] S. H. Khan, X. He, F. Porikli, M. Bennamoun, F. Sohel, and R. Togneri, "Learning deep structured network for weakly supervised change detection," *arXiv:1606.02009*, 2016.
[33] P. H. O. Pinheiro and R. Collobert, "From image-level to pixel-level labeling with convolutional networks," *Proc. CVPR*, pp. 1713–1721, 2015.
[34] D. Pathak, P. Krähenbühl, and T. Darrell, "Constrained convolutional neural networks for weakly supervised segmentation," *Proc. ICCV*, pp. 1796–1804, 2015.
[35] L. Ye, Z. Liu, and Y. Wang, "Learning semantic segmentation with diverse supervision," 2018.
[36] A. Kolesnikov and C. H. Lampert, "Seed, expand and constrain: Three principles for weakly-supervised image segmentation," in *European conference on computer vision*. Springer, 2016, pp. 695–711.
[37] Y. Wei, X. Liang, Y. Chen, X. Shen, M. Cheng, J. Feng, Y. Zhao, and S. Yan, "Stc: A simple to complex framework for weakly-supervised semantic segmentation," *IEEE TPAMI*, vol. 39, no. 11, pp. 2314–2320, 2017.
[38] X. Wang, S. Liu, H. Ma, and M.-H. Yang, "Weakly-supervised semantic segmentation by iterative affinity learning," *Int J Comput Vis*, Jun. 2020. [Online]. Available: https://doi.org/10.1007/s11263-020-01293-3
[39] J. Ahn and S. Kwak, "Learning pixel-level semantic affinity with image-level supervision for weakly supervised semantic segmentation," *Proc. CVPR*, pp. 4981–4990, 2018.
[40] Z.-H. Zhou, "A brief introduction to weakly supervised learning," *National science review*, vol. 5, no. 1, pp. 44–53, 2018.
[41] P. Welinder, S. Branson, P. Perona, and S. Belongie, "The multidimensional wisdom of crowds," in *Proc. NIPS*, 2011.
[42] "Observer Variation - MeSH - NCBI." [Online]. Available: https://www.ncbi.nlm.nih.gov/mesh/68015588
[43] D. Lin, J. Dai, J. Jia, K. He, and J. Sun, "Scribblesup: Scribble-supervised convolutional networks for semantic segmentation," in *Proc. CVPR*, 2016, pp. 3159–3167.
[44] A. Bearman, O. Russakovsky, V. Ferrari, and L. Fei-Fei, "What's the point: Semantic segmentation with point supervision," in *Proc. ECCV*, 2016, pp. 549–565.
[45] T.-Y. Lin, M. Maire, S. J. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *Proc. ECCV*, 2014.
[46] J. Deng, W. Dong, R. Socher, L. Li, Kai Li, and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. CVPR*, Jun. 2009, pp. 248–255.
[47] A. Oliver, A. Odena, C. Raffel, E. D. Cubuk, and I. J. Goodfellow, "Realistic evaluation of deep semi-supervised learning algorithms," in *Proc. NeurIPS*, 2018, p. 3239–3250.
[48] A. Tarvainen and H. Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," in *Proc. NIPS*, 2017, p. 1195–1204.
[49] W. Shi, Y. Gong, C. Ding, Z. Ma, X. Tao, and N. Zheng, "Transductive semi-supervised deep learning using min-max features," in *Proc. ECCV*, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds., 2018, pp. 311–327.
[50] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. ICML*, 2019, pp. 6105–6114.
[51] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
[52] O. Vinyals, C. Blundell, T. Lillicrap, K. Kavukcuoglu, and D. Wierstra, "Matching networks for one shot learning," *arXiv:1606.04080*, 2016.
[53] M. Ren, E. Triantafillou, S. Ravi, J. Snell, K. Swersky, J. B. Tenenbaum, H. Larochelle, and R. S. Zemel, "Meta-learning for semi-supervised few-shot classification," *arXiv:1803.00676*, 2018.
[54] A. Rosenfeld, R. A. Hummel, and S. W. Zucker, "Scene labeling by relaxation operations," *IEEE Trans. Syst. Man Cybern.*, vol. 6, no. 6, pp. 420–433, 1976.
[55] R. A. Hummel and S. W. Zucker, "On the foundations of relaxation labeling processes," *IEEE TPAMI*, vol. 5, no. 3, pp. 267–287, 1983.
[56] M. Pelillo, "The dynamics of nonlinear relaxation labeling processes," *J. Math. Imaging Vis.*, vol. 7, no. 4, pp. 309–323, 1997.
[57] I. Elezi, A. Torcinovich, S. Vascon, and M. Pelillo, "Transductive label augmentation for improved deep network learning," in *Proc. ICPR*, 2018, pp. 1432–1437.
[58] J. Zhou, G. Cui, Z. Zhang, C. Yang, Z. Liu, and M. Sun, "Graph neural networks: A review of methods and applications," *arXiv:1812.08434*, 2018.
[59] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, "A comprehensive survey on graph neural networks," *arXiv:1901.00596*, 2019.
[60] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proc. ICLR*, 2017.
[61] X. Wang, Y. Ye, and A. Gupta, "Zero-shot recognition via semantic embeddings and knowledge graphs," *arXiv:1803.08035*, 2018.
[62] M. Kampffmeyer, Y. Chen, X. Liang, H. Wang, Y. Zhang, and E. P. Xing, "Rethinking knowledge graph propagation for zero-shot learning," *arXiv:1805.11724*, 2018.
[63] V. Garcia and J. Bruna, "Few-shot learning with graph neural networks," *arXiv:1711.04043*, 2018.
[64] K. Marino, R. Salakhutdinov, and A. Gupta, "The more you know: Using knowledge graphs for image classification," in *Proc. CVPR*, 2017, pp. 20–28.
[65] C. Lee, W. Fang, C. Yeh, and Y. F. Wang, "Multi-label zero-shot learning with structured knowledge graphs," in *Proc. CVPR*, 2018, pp. 1576–1585.
[66] M. Li, M. Soltanolkotabi, and S. Oymak, "Gradient Descent with Early Stopping is Provably Robust to Label Noise for Overparameterized Neural Networks," *arXiv:1903.11680*, 2019.
[67] A. Krizhevsky, G. Hinton *et al.*, "Learning multiple layers of features from tiny images," University of Toronto, Tech. Rep., 2009.
[68] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
[69] H. Xiao, K. Rasul, and R. Vollgraf, "Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms," *arXiv:1708.07747*, 2017.
[70] R. Adolphi *et al.*, "The CMS experiment at the CERN LHC," *JINST*, vol. 3, p. S08004, 2008.
[71] P. Foret, A. Kleiner, H. Mobahi, and B. Neyshabur, "Sharpness-aware minimization for efficiently improving generalization," *arXiv preprint arXiv:2010.01412*, 2020.
[72] S. Hochreiter and J. Schmidhuber, "Flat minima," *Neural computation*, vol. 9, no. 1, pp. 1–42, 1997.
[73] T. G. Dietterich, R. H. Lathrop, and T. Lozano-Pérez, "Solving the multiple instance problem with axis-parallel rectangles," *Artificial intelligence*, vol. 89, no. 1-2, pp. 31–71, 1997.
[74] J. Lafferty, A. McCallum, and F. C. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *Proc. ICML*, 2001.
[75] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proc. CVPR*, 2016, pp. 2921–2929.
[76] X. Liang, X. Shen, J. Feng, L. Lin, and S. Yan, "Semantic object parsing with graph LSTM," *arXiv:1603.07063*, 2016.
[77] X. Liang, L. Lin, X. Shen, J. Feng, S. Yan, and E. P. Xing, "Interpretable structure-evolving LSTM," *arXiv:1703.03055*, 2017.
[78] X. Qi, R. Liao, J. Jia, S. Fidler, and R. Urtasun, "3d graph neural networks for rgbd semantic segmentation," in *Proc. ICCV*, 2017, pp. 5209–5218.
[79] L. Landrieu and M. Simonovsky, "Large-scale point cloud semantic segmentation with superpoint graphs," in *Proc. CVPR*, 2018, pp. 4558–4567.
[80] J. Bromley, I. Guyon, Y. LeCun, E. Säckinger, and R. Shah, "Signature verification using a" siamese" time delay neural network," *Proc. NIPS*, pp. 737–737, 1994.
[81] S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, and P. H. Torr, "Conditional random fields as recurrent neural networks," in *Proc. ICCV*, 2015, pp. 1529–1537.
[82] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. MICCAI*. Springer, 2015, pp. 234–241.
[83] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv:1409.1556*, 2014.
[84] L. Tuggener, J. Schmidhuber, and T. Stadelmann, "Is it enough to optimize cnn architectures on imagenet?" *arXiv preprint arXiv:2103.09108*, 2021.