Channel-Wise Contrastive Learning for Learning with Noisy Labels

Hui Kang[1]   **Sheng Liu**[2*]   **Huaxi Huang**[3*]   **Tongliang Liu**[1†]
[1]The University of Sydney   [2]New York University   [3]Data61, CSIRO

Abstract

In real-world datasets, noisy labels are pervasive. The challenge of learning with noisy labels (LNL) is to train a classifier that discerns the actual classes from given instances. For this, the model must identify features indicative of the authentic labels. While research indicates that genuine label information is embedded in the learned features of even inaccurately labeled data, it's often intertwined with noise, complicating its direct application. Addressing this, we introduce channel-wise contrastive learning (CWCL). This method distinguishes authentic label information from noise by undertaking contrastive learning across diverse channels. Unlike conventional instance-wise contrastive learning (IWCL), CWCL tends to yield more nuanced and resilient features aligned with the authentic labels. Our strategy is twofold: firstly, using CWCL to extract pertinent features to identify cleanly labeled samples, and secondly, progressively fine-tuning using these samples. Evaluations on several benchmark datasets validate our method's superiority over existing approaches.

1   Introduction

Deep neural networks (DNNs) have been at the forefront of numerous breakthroughs in diverse areas of study, showcasing exemplary performances across a myriad of tasks [20, 14, 31, 11]. A significant portion of their success can be attributed to the availability of vast volumes of high-quality annotated data. However, obtaining such expansive and impeccably labeled datasets can often pose significant financial constraints or, in some scenarios, may be entirely unfeasible.

Compounding this challenge, several datasets, as highlighted by sources like [34, 40], are amassed through avenues such as search engines or web crawlers. This modus operandi inevitably introduces a plethora of noisy labels. Training DNNs on these compromised datasets can lead to a counterproductive phenomenon: due to the intricate model capacity of DNNs, they tend to adapt excessively to these erroneous labels. This maladaptation subsequently results in compromised model generalization.

The paradigm of learning with noisy labels (LNL) [1] poses intricate challenges. The end goal is to sculpt a classifier that remains resilient against the pitfalls of misleading data emerging from inaccurate labels. Moreover, a successful model would need to deduce the genuine labels predicated upon the features gleaned from the input data. Achieving this necessitates that the derived features are predominantly infused with precise label information. This underscores the critical nature of directing the model to extract and prioritize clean label data in LNL scenarios. The research community has been relentless in this pursuit, exploring an array of methodologies [30, 47, 17, 18, 28] to distill pure label information from the mire of noisy training data.

Delving deeper into the anatomy of a DNN, it becomes apparent that the channels of its deep features often resonate with distinct visual patterns [33, 46, 48]. A plethora of studies [6, 5, 49, 13] reinforce

---

[*]Equal contribution.

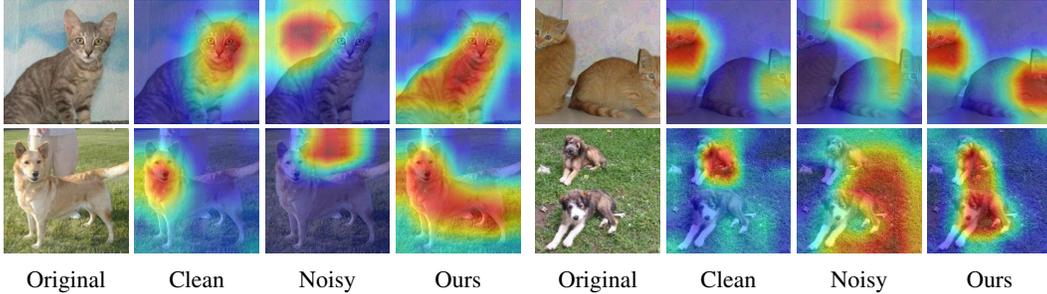[†]Contact person (tongliang.liu@sydney.edu.au).

Figure 1: The Grad-CAM [32] results of ResNet18 models trained on Dogs-vs-Cats [12]. 'Clean' represents the model trained with clean labels using CE loss. 'Noisy' denotes the model trained with 40% symmetric label noise using CE loss. 'Ours' showcases the model trained with 40% symmetric label noise using CWCL loss. Under label noise, CWCL can extract similar or even more clean label information than the model trained with clean labels.

the idea that intricate local features — like the unique characteristics of a bird's tail — are sequestered within specific channels. These features play pivotal roles in distinguishing between intricate subjects, such as different avian species. While these channels are treasure troves of clean label information, the full potential of this channel-wise data remains untapped in the context of LNL. The quest to devise strategies that can effectively mine these channels within noisy datasets stands as a tantalizing research challenge.

Recent innovations in the realm of self-supervised learning [8, 43, 19] have imbued optimism. These methodologies have demonstrated their prowess in amplifying the performance of DNNs, even in the absence of labels. By employing a contrastive loss mechanism on an unlabeled dataset, these self-supervised models have managed to rival, if not surpass, the efficacy of their supervised counterparts that train on labeled data. This observation elucidates the sheer potency of contrastive learning as a tool to unearth the pristine label information embedded within training data.

One underlying culprit behind DNNs' propensity to overfit to label-noise is their reliance on label-dependent supervision losses [2, 25], such as the cross entropy (CE) loss. This mechanism introduces gradients biased by erroneous labels, marring the optimization process. Recent innovations [42, 25] have leveraged the tenets of self-supervised techniques, with contrastive learning being a noteworthy contributor, to counter this overfitting menace. The approach taken by [25] involves the pre-training of a base model using an unsupervised contrastive loss, followed by the selection of untarnished data pairs for subsequent finetuning. Complementarily, Ref. [42] integrated contrastive loss as a complementary regularizer for the CE loss, aiming to achieve robust LNL representation learning. Intriguingly, both these methods [42, 25] deploy instance-wise contrastive learning (IWCL) mechanisms to delineate between diverse data instances.

However, while IWCL offers respite from the overfitting challenge by bypassing label dependency, it does not inherently cater to the meticulous extraction of features pivotal to genuine labels. This realization paves the way for the potential inception of an avant-garde self-supervised learning paradigm. It should be meticulously crafted to extensively sift through the data and harvest authentic label information. With this vision, we put forth our innovative channel-wise contrastive learning (CWCL) approach. This technique is tailor-made to hone in on fine-grained and resilient features, setting a new benchmark for LNL scenarios.

Channel-wise contrastive learning is rooted in two fundamental principles: 1) Collaborative training employing CE loss and channel-wise contrastive loss. 2) Progressive confident-sample finetuning. In its operational paradigm, while concurrently trained with CE loss, CWCL delves into individual feature channels, spearheading contrastive learning across every layer or block of the DNN architecture. This unique approach mandates augmented positive channel pairs to gravitate towards each other, whilst ensuring that the negative counterparts maintain discernible distance. A consequential outcome of this mechanism, as illustrated in Figure 1, is the emergence of a more varied spectrum of feature channels compared to conventional LNL methodologies.

Enhancing the channel diversity within the corresponding feature maps across layers infuses the input samples or features with profound complexity. This escalating data complexity inversely impacts

the model's inherent complexity, ensuring its attenuation. Such a dynamic is pivotal in subduing the model's inclination to overfit when trained on data tainted with noisy labels [3]. This mechanism acts as a conduit for the LNL model, enabling it to sift and retain pristine label information from a sea of noise.

Furthermore, drawing from our earlier discussions, the profound depths of feature channels are reservoirs of nuanced and discriminatory data. CWCL capitalizes on this characteristic, meticulously mining these channels to harvest untainted label information. Progressive confident-sample finetuning embarks on a mission to further elevate the prowess of the LNL model. It operates by meticulously curating confident samples from the model's previous training epoch and subsequently leveraging them to nurture the succeeding epoch's model training. This stage is orchestrated through a synergistic training paradigm employing both CE and supervised contrastive loss [23]. This sequential fine-tuning refines the LNL model incrementally.

To our discernment, CWCL stands as a pioneering methodology that contemplates the challenge of learning features from noisy labels through the lens of a channel-centric perspective. Our pivotal contributions encapsulated within this research paradigm are:

- The inception of a channel-wise contrastive loss mechanism, meticulously tailored to extract and amplify clean label information within the realm of the LNL task.

- A novel approach towards progressive confident-sample finetuning, designed to incessantly refine the LNL model, thereby enhancing its resilience against label noise.

- Our empirical evaluations spanning multiple datasets attest to the formidable efficacy of our proposed strategy. Furthermore, our method consistently eclipses performances set by contemporary state-of-the-art methodologies.

2 Related Work

2.1 Learning with Noisy Labels

The landscape of learning with noisy labels has evolved considerably, and a plethora of methods have emerged. These methodologies can be primarily partitioned into two overarching paradigms: model-based and model-free strategies.

Model-based Strategies: Central to these techniques is the quest to decipher and estimate the intrinsic noise transition probabilities. This is accomplished by characterizing the intricate interplay between the clean and noisy labels [30, 36, 38, 41, 27]. Underlying this premise is the assumption that the noisy label manifests as a conditional probability distribution derivative of the pristine labels. Taking illustrative instances, Ref. [16] championed a noise adaptation layer, positioned atop the classification model, dedicated to the learning of transition probabilities. In a nuanced approach, T-revision [39] leveraged fine-tuned slack variables to intuitively gauge the noise transition matrix devoid of anchoring points. Another noteworthy mention is the model proposed by [27], where label noise is encapsulated within a sparse over-parameterized framework.

Model-free Strategies: Distancing themselves from explicit noise modeling, these techniques harness the deep models' inherent memorization capabilities [2] to counteract the deleterious effects of noisy labels [17, 24, 3, 36, 21]. A case in point is the Co-teaching method [17] where a pair of deep networks embark on a collaborative journey, educating one another through the selective exchange of low-loss instances within mini-batches. Enriching this foundation, DivideMix [24] marries the essence of Co-teaching with the sophistication of two Beta Mixture Models. Further, the PES initiative [3] delves into the intricacies of the progressive early halting of deep architectures, establishing diversified stopping points for distinct network segments. The proposed CWCL method finds its lineage in this model-free territory. It aspires to unearth channel-centric pristine label information via contrastive learning, thereby stifling the negative reverberations of noisy labels during the classification training phase.

---

[3]A quintessential learning task demands a harmonious alignment between model and data complexity. Overfitting is inevitable when a highly intricate model grapples with simplistic data, whereas an elementary model runs the risk of underfitting when confronted with convoluted data sets.

2.2 Feature Channels in DNNs

A retrospective glance at the annals of DNN research [44, 50, 32] unveils a captivating narrative: network features inherently encode hierarchical information. While the shallower layers resonate with rudimentary semantic nuances like edges and colors, the intermediate strata capture localized or fragmented data. In stark contrast, the pinnacle layers distill high-order semantic facets, painting a comprehensive portrayal of objects. The orchestration of DNNs is such that each layer usually houses multiple filters, each begetting a characteristic feature map. These maps, often envisaged as 2D matrices, amalgamate, forming a 3D tensor when emanating from multiple filters within a singular layer. This assembly culminates in each DNN layer exuding a 3D tensor representation, with individual channels mirroring distinct visual patterns [33, 46].

In this realm, several pioneering investigations [6, 5, 49, 13] have spotlighted the profound significance of deep feature channels, particularly in fine-grained image classification endeavors. Their revelations underscore that specific DNN feature channels harbor subtle yet quintessential discriminative part information - a linchpin for intricate image identification. Harnessing such channels promises marked performance leaps in fine-grained image discernment. Drawing inspiration from these insights, we hypothesize that these deep feature channels are reservoirs of untainted label information. Consequently, we champion a novel framework, tailored to extract and leverage these granular discriminative features from channels, catering explicitly to LNL tasks.

2.3 Contrastive Learning

Contrastive learning has etched itself as a formidable contender in the representation learning arena [19, 8, 4, 23]. The core tenet of these methodologies hinges on amplifying the congruence between positive instance pairs while concurrently diluting the affinity between negative counterparts. Incredibly, sans label supervision, contrastive learning has par excellence, rivaling supervised learning across a gamut of tasks.

In the realm of learning with noisy labels, pioneering contributions by [42] and [25] have harnessed the prowess of contrastive learning to counter the overfitting scourge induced by noise-laden labels. To elucidate, Ref. [25] employed contrastive learning as a precursor to base model pre-training, subsequently cherry-picking pristine pairs for the fine-tuning phase. Contrastingly, Ref. [42] wove the contrastive loss into the cross-entropy loss fabric, culminating in robust LNL representation learning. These ventures predominantly orbit around instance-wise contrastive learning, emphasizing differentiating disparate instances. Breaking from this mold, our proposition stands unparalleled as the maiden foray into explicit channel-centric feature learning for LNL. Given that specific feature channels encapsulate fine-grained discriminative nuances, our methodology promises a richer label information extraction than its instance-wise contrastive learning counterparts.

3 Methodology

In this section, we first provide the problem definition of a LNL task. Next, we present our proposed channel-wise contrastive learning (CWCL) approach. Finally, we introduce a progressive confident sample fine-tuning technique to further enhance the performance of the LNL classifier.

3.1 Problem Definition

In the context of learning with noisy labels, the true distribution of training data is typically represented by $\mathcal{D} = \{(x, y) \,|\, x \in \mathcal{X}, y \in 1, \ldots, K\}$. Here, $\mathcal{X}$ denotes the sample space, and $1, \ldots, K$ represents the label space consisting of K classes. However, due to label errors during data collection and dataset construction, the actual distribution of the label space is often unknown. Therefore, we have to rely on a noisy dataset $\tilde{\mathcal{D}} = \{(x, \tilde{y}) \,|\, x \in \mathcal{X}, \tilde{y} \in 1, \ldots, K\}$ with corrupted labels $\tilde{y}$ to train the model. Our goal is to develop an algorithm that can learn a robust deep classifier from these noisy data to accurately classify query samples.
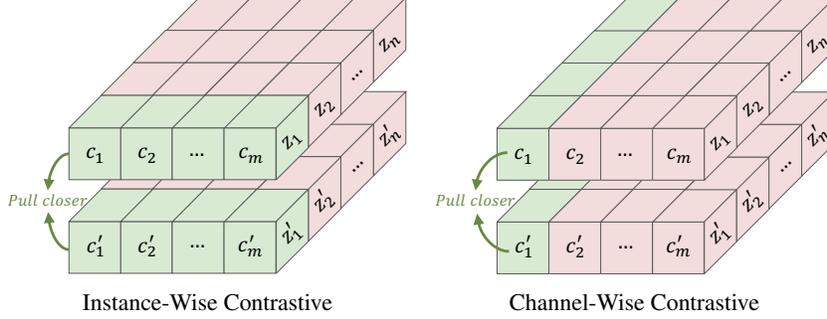
Figure 2: Comparison between instance-wise and channel-wise contrastive loss. z is the feature representation for different instances, and c refers to different feature channels. For instance-wise contrastive loss, two views of the same instance, z_i and z_i', constitute the positive pair. While for channel-wise contrastive loss, two views of the same channel, c_i and c_i', form the positive pair.

3.2 Channel-Wise Contrastive Loss

Following previous traditional instance-wise contrastive learning (IWCL) methods [8, 19, 10, 9], we often perform random data augmentation twice on each image in a mini-batch of N images $\{x_1, x_2, \ldots, x_N\}$, resulting in a larger batch of $2N$ images. For convenience, we consider the images x_i and x_{i+N} as two different augmented versions (views) of the same original image, which together form a positive pair. The representation vector from projection head is denoted as $z = Proj(Enc(x))$. The length of z is M (i.e. number of channels). The instance-wise contrastive loss, also known as InfoNCE loss [29] is expressed as

$$\mathcal{L}_{IWCL} = -\sum_{i=1}^{N} \log \frac{\exp\left(sim\left(z_i, z_{i+N}\right)/\tau\right)}{\sum_{k=1, k\neq i}^{2N} \exp\left(sim\left(z_i, z_k\right)/\tau\right)} \tag{1}$$

where τ is a temperature hyper-parameter, and sim represents the cosine similarity. Intuitively, $\mathcal{L}_{IWCL}$ encourages the encoder network to learn similar representation for different augmentations from the same image while increasing the difference between representations of the augmentations from different images.

As for proposed channel-wise contrastive loss, referring to Figure 2 and Equation 1, it can be summarized as

$$\mathcal{L}_{CWCL} = -\sum_{i=1}^{M} \log \frac{\exp\left(sim\left(c_i, c_{i+M}\right)/\tau\right)}{\sum_{k=1, k\neq i}^{2M} \exp\left(sim\left(c_i, c_k\right)/\tau\right)} \tag{2}$$

From Equation 2, we can know that, $\mathcal{L}_{CWCL}$ encourages the encoder network to learn similar representation for different augmentations from the same channel while increasing the difference between representations of the augmentations from different channels.

We extract representations from L intermediate layers of the network (e.g. $layer\{1, \ldots, 4\}$ of ResNet18) and apply each to $\mathcal{L}_{CWCL}$, and use CE loss ($\mathcal{L}_{CE}$) at the fully connected layer to do classification. So the combined total loss of this stage ($\mathcal{L}_{Stage1}$) is

$$\mathcal{L}_{Stage1} = (1-\lambda)\mathcal{L}_{CE} + \frac{\lambda}{L}\sum_{l=1}^{L}\mathcal{L}_{CWCL} \tag{3}$$

where λ is a loss balance hyper-parameter. Training the LNL model using Equation 3, we can obtain a classifier that learns to mine the clean label information from the noisy training data under the supervision of supervised CE loss and unsupervised CWCL loss.

3.3 Progressive Confident Sample Finetuning

Confident samples are characterized by high prediction probabilities concerning their associated labels. In a bid to augment robustness, we employ two distinct data augmentations for every input sample.

Table 1: Comparison of test accuracy using different methods on CIFAR-10 dataset with varying noise types and levels. The baseline results are taken from [42]. We use ResNet18 as the architecture, whereas all other methods in the comparison use PreAct ResNet18. The mean and standard deviation over 3 runs are reported. The best results are highlighted in bold.

| Method | CIFAR-10 | | | | | |
| | Symmetric | | | | | Asymmetric |
	0%	20%	40%	60%	80%	40%
CE	93.97±0.22	88.51±0.17	82.73±0.16	76.26±0.29	59.25±1.01	83.23±0.59
Forward	93.47±0.19	88.87±0.21	83.28±0.37	75.15±0.73	58.58±1.05	82.93±0.74
GCE	92.38±0.32	91.22±0.25	89.26±0.34	85.76±0.58	70.57±0.83	82.23±0.61
Co-teaching	93.37±0.12	92.05±0.15	87.73±0.17	85.10±0.49	44.16±0.71	77.78±0.59
LIMIT	93.47±0.56	89.63±0.42	85.39±0.63	78.05±0.85	58.71±0.83	83.56±0.70
SLN	93.21±0.21	88.77±0.23	87.03±0.70	80.57±0.50	63.99±0.79	81.02±0.25
SL	94.21±0.13	92.45±0.08	89.22±0.08	84.63±0.21	72.59±0.23	83.58±0.60
APL	93.97±0.25	92.51±0.39	89.34±0.33	85.01±0.17	70.52±2.36	84.06±0.20
CTRR	94.29±0.21	93.05±0.32	92.16±0.31	87.34±0.84	83.66±0.52	89.00±0.56
CWCL	**96.71±0.02**	**94.04±0.16**	**93.46±0.08**	**91.87±0.04**	**86.31±0.70**	**92.71±0.27**

The subsequent predicted label is ascertained by averaging predictions across these augmentations. This strategy not only stabilizes predictions but also elevates performance, as corroborated by empirical studies. On gathering these high-confidence samples, the classifier is then trained by viewing them as pristine, noise-free data.

A noteworthy challenge that surfaces is the potential variance in the number of confident samples across different classes. Training the model directly on this set might propagate severe class imbalance issues. To navigate this predicament, we incorporate a class balance sampler in the dataloader. This ensures an equitable representation from all classes during the model's training phase, thus bolstering its learning efficacy.

Since we consider confident samples as clean data, we use supervised contrastive loss ($\mathcal{L}_{SupCon}$) [23] instead of $\mathcal{L}_{CWCL}$. Same as $\mathcal{L}_{Stage1}$, we also extract representations from L intermediate layers of the network and apply each to $\mathcal{L}_{SupCon}$, and use CE loss ($\mathcal{L}_{CE}$) at the fully connected layer to do classification. So the combined total loss of this stage ($\mathcal{L}_{Stage2}$) is

$$\mathcal{L}_{Stage2} = (1 - \lambda)\,\mathcal{L}_{CE} + \frac{\lambda}{L}\sum_{l=1}^{L}\mathcal{L}_{SupCon} \qquad (4)$$

Based on the confident sample selector trained in stage one using Equation 3, we continuously train the model using Equation 4 each time and select the confident sample from the currently trained model as next-time input. Such a progressive confident-sample finetuning strategy can gradually improve the quality of the confident sample and further boost the performance of a LNL model.

4   Experiment

4.1   Datasets and Implementation Details

Datasets: We evaluate our method on two synthetic datasets with different noise types and levels, CIFAR-10 and CIFAR-100 [15], as well as two real-world datasets, Animal-10N [34] and Clothing-1M [40]. CIFAR-10 and CIFAR100 both contain 50k training images and 10k testing images, each with a size of 32×32 pixels. CIFAR-10 has 10 classes, while CIFAR-100 contains 100 classes. The original labels of these two datasets are clean. Animal-10N has 10 animal classes with 50k training images and 5k test images, each with a size of 64×64 pixels. Its estimated noise rate is around 8%. Clothing-1M has 1 million training images and 10k test images with 14 classes crawled from online shopping web sites. Its estimated noise level is around 40%.

Synthetic Noise: Following previous works [17, 26, 37, 30], we explore two different types of synthetic noise with different noise levels for both CIFAR-10 and CIFAR-100 datasets. For symmetric label noise in both datasets, each label has the same probability of being flipped to any class, and we randomly select a certain percentage of training data to have their labels flipped, with the range being {20%, 40%, 50%, 60%, 80%}. For asymmetric label noise in CIFAR-10, we follow the labeling rule

Table 2: Comparison of test accuracy using different methods on CIFAR-100 dataset with varying noise types and levels. The baseline results are taken from [42]. We use ResNet18 as the architecture, whereas all other methods in the comparison use PreAct ResNet18. The mean and standard deviation over 3 runs are reported. The best results are highlighted in bold.

Method	CIFAR-100					
	Symmetric					Asymmetric
	0%	20%	40%	60%	80%	40%
CE	73.21±0.14	60.57±0.53	52.48±0.34	43.20±0.21	22.96±0.84	44.45±0.37
Forward	73.01±0.33	58.72±0.54	50.10±0.84	39.35±0.82	17.15±1.81	-
GCE	72.27±0.27	68.31±0.34	62.25±0.48	53.86±0.95	19.31±1.14	46.50±0.71
Co-teaching	73.39±0.27	65.71±0.20	57.64±0.71	31.59±0.88	15.28±1.94	-
LIMIT	65.53±0.91	58.02±1.93	49.71±1.81	37.05±1.39	20.01±0.11	-
SLN	63.13±0.21	55.35±1.26	51.39±0.48	35.53±0.58	11.96±2.03	-
SL	72.44±0.44	66.46±0.26	61.44±0.23	54.17±1.32	34.22±1.06	46.12±0.47
APL	73.88±0.99	68.09±0.15	63.46±0.17	53.63±0.45	20.00±2.02	52.80±0.52
CTRR	74.36±0.41	70.09±0.45	65.32±0.20	54.20±0.34	43.69±0.28	54.47±0.37
CWCL	**81.04±0.10**	**75.12±0.12**	**73.90±0.27**	**70.83±0.23**	**60.49±0.34**	**73.97±0.04**

proposed in [30], where we flip labels between TRUCK → AUTOMOBILE, BIRD → AIRPLANE, DEER → HORSE, and CAT ↔ DOG. We randomly choose 40% of the training data and flip their labels according to the asymmetric labeling rule. For asymmetric label noise in CIFAR-100, we also randomly select 40% of the training data and flip their labels to the next class in the label space.

Baseline Methods: To evaluate the effectiveness of our proposed method, we compare it against several baseline methods that address label noise. These methods include: 1) CE loss, which is the standard loss function used in supervised learning. 2) Forward correction [30], which corrects loss values based on an estimated noise transition matrix. 3) GCE [47], which combines the Mean Absolute Error (MAE) loss and CE loss to create a robust loss function. 4) Co-teaching [17], which trains two networks and selects small-loss examples to update. 5) LIMIT [18], which introduces noise to gradients to avoid memorization. 6) SLN [7], which adds Gaussian noise to labels to combat label noise. 7) SL [35], which uses CE loss and a Reverse Cross Entropy (RCE) loss as a robust loss function. 8) APL [28], which combines two mutually boosted robust loss functions NCE and RCE for training. 9) CTRR [42], which proposes a novel contrastive regularization function to address the memorization issue of LNL, and achieved state-of-the-art performance. The results of the baseline methods are taken from [42].

Implementation Details: In our study, to guarantee unbiased evaluations across different experiments, we consistently utilized the ResNet18 architecture [20] for all datasets. Throughout the training process, we adhered to a batch size of 128 and set the loss balance factor λ to 0.6. Optimization was achieved using the SGD optimizer with a momentum of 0.9, weight decay of 5e-4, and an initial learning rate of 0.1. Our training regimen spanned a total of 300 epochs: initially, 100 epochs leveraging the channel-wise contrastive loss were used to form a preliminary model adept at generating high-quality confident samples. These samples subsequently steered the next 200 epochs of training, employing the contrastive deep supervision method as proposed in [45]. To fortify the stability of our training dynamics, we adopted the exponential moving average (EMA) strategy [22]. Ensuring transparency, all our findings are presented as the mean and standard deviation derived from three independent runs.

4.2 Classification Performance Analysis

Results on Synthetic Datasets: Tables 1 and 2 elucidate the comparative performance of different methods on CIFAR-10 and CIFAR-100 across a diverse spectrum of label noise settings. A discernible trend from the results is the consistent superior performance of CWCL, underscoring its robustness against other contemporary methodologies. While CWCL's dominance is evident across all noise levels, its prowess is particularly pronounced in the CIFAR-100 dataset with symmetric 80% and asymmetric 40% noise settings. Here, CWCL surpasses the runner-up method by an impressive margin, nearing 20 percentage points in terms of accuracy. Such outstanding results underscore CWCL's adeptness in managing label noise, thereby ensuring enhanced classification outcomes even under intricate noise-dominant scenarios.

Table 3: Comparison of test accuracy using different methods on real-world datasets Animal-10N and Clothing-1M. The baseline results are taken from [42]. All methods use ResNet18 as the base model. The mean and standard deviation over 3 runs are reported. The best results are highlighted in bold.

Method	Animal-10N	Clothing-1M
CE	83.18±0.15	70.88±0.45
Forward	83.67±0.31	71.23±0.39
GCE	84.42±0.39	71.34±0.12
Co-teaching	85.73±0.27	71.68±0.21
SLN	83.17±0.08	71.17±0.12
SL	83.92±0.28	72.03±0.13
APL	84.25±0.11	72.18±0.21
CTRR	86.71±0.15	72.71±0.19
CWCL	**88.95±0.23**	**73.87±0.16**

Results on Real-world Datasets: Table 3 shines a light on the comparative analysis for real-world datasets, namely Animal-10N and Clothing-1M. To ensure an equitable comparison with previously established benchmarks, specific architectures were chosen: a randomly initialized ResNet18 for Animal-10N and an ImageNet pre-trained ResNet18 for Clothing-1M. Even under these settings, CWCL emerges as a top contender, surpassing other methods on both the aforementioned datasets. Its efficacy is especially noteworthy on the Clothing-1M dataset, an expansive collection of 1 million images. On this dataset, CWCL not only establishes its dominance but does so with an appreciable margin, achieving an accuracy that's over a percentage point higher than its closest competitor. This superior performance on a real-world, large-scale dataset further testifies to CWCL's robustness and adaptability.

4.3 Ablation Studies

Table 4: Ablation studies of the proposed CWCL under the supervised setting, experiments on CIFAR-100 are based on a ResNet18 backbone. CWCL(Y/N) indicates the progressive confident-sample finetuning stage is enabled/disabled.

Method	CIFAR-100			
	Symmetric			Asymmetric
	40%	60%	80%	40%
CE	52.48±0.34	43.20±0.21	22.96±0.84	44.45±0.37
CTRR	65.32±0.20	54.20±0.34	43.69±0.28	54.47±0.37
CWCL(N)	70.70±0.07	63.81±0.26	49.48±0.40	67.91±0.48
CWCL(Y)	73.90±0.27	70.83±0.23	60.49±0.34	73.97±0.04

We present an ablation study to delve deeper into the inner workings and contributions of various components of our proposed CWCL method. The results of this analysis are illustrated in Table 4.

Effectiveness of Channel-Wise Contrastive Learning: By examining the results of CWCL(N), it becomes evident that even without the integration of progressive confident sample finetuning, CWCL outperforms both CE and CTRR. This significant lead highlights the inherent strength and efficacy of the channel-wise contrastive learning approach. By leveraging this scheme, our model efficiently harnesses more discriminative label information, even in the presence of noisy training data, further attesting to its resilience and robustness.

Contribution of Progressive Confident Sample Finetuning: The results demonstrate that incorporating the progressive confident sample finetuning stage further refines the performance, as seen with CWCL(Y). This observation aligns with our expectations. After the initial training phase via channel-wise contrastive learning, the model possesses the capability to produce high-quality confident samples. These samples then serve as an invaluable resource during subsequent training stages, ensuring a more refined learning process.

Performance under High Noise Levels: An intriguing observation from our analysis is the enhanced performance of our model, especially when subjected to high noise levels. The progressive confident-

sample finetuning technique appears to be particularly effective in these scenarios. This suggests that our method is not only robust but is also adaptive, enhancing its capabilities further when faced with challenging noisy conditions.

In conclusion, our ablation study underscores the synergistic effect of combining channel-wise contrastive learning with progressive confident sample finetuning. Each component plays a pivotal role, contributing to the method's overall effectiveness and resilience against noise.

5   Conclusion

In this research, we have introduced the Channel-Wise Contrastive Learning loss (CWCL) - a novel contrastive loss that offers a more refined approach to harnessing true label information, even amid a noisy environment. Unlike traditional methods that might struggle in the presence of noise, our empirical investigations have spotlighted the ability of CWCL to generate noise-resilient, as well as intricately detailed features. These features not only serve as robust indicators of true label information but also pave the way for the selection of pristine, high-caliber samples. The supremacy of CWCL over traditional instance-wise contrastive learning loss is not merely theoretical. Our empirical results are a testament to its prowess. A salient takeaway from our findings is how CWCL fosters feature diversity, a crucial attribute that bolsters the model's resistance against overfitting by augmenting feature complexity. However, our exploration did not stop at theoretical datasets. Recognizing the multifaceted nature of real-world data, we subjected CWCL to a series of rigorous tests, spanning a myriad of datasets varying in noise levels, noise types, and even those that mirror the unpredictable nature of real-world label noises. Our method showcased remarkable consistency and effectiveness across all evaluations, further solidifying its practical utility. Closing this chapter of research, we posit that the scope of CWCL is not limited to operating in isolation. Its adaptable nature makes it a prime candidate for integration with other existing methodologies. In scenarios plagued by label noise, which is a prevalent challenge in many machine learning endeavors, CWCL promises to be a potent tool, either standalone or in conjunction with other techniques, providing an edge in crafting more resilient and accurate models. We are optimistic about its potential and foresee its adoption in numerous forthcoming applications to tackle label noise challenges.

References

[1] Dana Angluin and Philip Laird. Learning from noisy examples. *Machine Learning*, 2(4):343–370, 1988.

[2] Devansh Arpit, Stanislaw Jastrzebski, Nicolas Ballas, David Krueger, Emmanuel Bengio, Maxinder S Kanwal, Tegan Maharaj, Asja Fischer, Aaron Courville, Yoshua Bengio, et al. A closer look at memorization in deep networks. In *ICML*, pages 233–242. PMLR, 2017.

[3] Yingbin Bai, Erkun Yang, Bo Han, Yanhua Yang, Jiatong Li, Yinian Mao, Gang Niu, and Tongliang Liu. Understanding and improving early stopping for learning with noisy labels. *NeurIPS*, 34:24392–24403, 2021.

[4] Mathilde Caron, Ishan Misra, Julien Mairal, Priya Goyal, Piotr Bojanowski, and Armand Joulin. Unsupervised learning of visual features by contrasting cluster assignments. *Advances in neural information processing systems*, 33:9912–9924, 2020.

[5] Dongliang Chang, Yifeng Ding, Jiyang Xie, Ayan Kumar Bhunia, Xiaoxu Li, Zhanyu Ma, Ming Wu, Jun Guo, and Yi-Zhe Song. The devil is in the channels: Mutual-channel loss for fine-grained image classification. *IEEE Transactions on Image Processing*, 29:4683–4695, 2020.

[6] Long Chen, Hanwang Zhang, Jun Xiao, Liqiang Nie, Jian Shao, Wei Liu, and Tat-Seng Chua. Sca-cnn: Spatial and channel-wise attention in convolutional networks for image captioning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5659–5667, 2017.

[7] Pengfei Chen, Guangyong Chen, Junjie Ye, Pheng-Ann Heng, et al. Noise against noise: stochastic label noise helps combat inherent label noise. In *International Conference on Learning Representations*, 2021.

[8] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PMLR, 2020.

[9] Ting Chen, Simon Kornblith, Kevin Swersky, Mohammad Norouzi, and Geoffrey E Hinton. Big self-supervised models are strong semi-supervised learners. *Advances in neural information processing systems*, 33:22243–22255, 2020.

[10] Xinlei Chen, Haoqi Fan, Ross Girshick, and Kaiming He. Improved baselines with momentum contrastive learning. *arXiv preprint arXiv:2003.04297*, 2020.

[11] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[12] Jeremy Elson, John (JD) Douceur, Jon Howell, and Jared Saul. Asirra: A captcha that exploits interest-aligned manual image categorization. In *Proceedings of 14th ACM Conference on Computer and Communications Security (CCS)*. Association for Computing Machinery, Inc., October 2007.

[13] Yu Gao, Xintong Han, Xun Wang, Weilin Huang, and Matthew Scott. Channel interaction networks for fine-grained image categorization. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 10818–10825, 2020.

[14] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587, 2014.

[15] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 249–256. JMLR Workshop and Conference Proceedings, 2010.

[16] Jacob Goldberger and Ehud Ben-Reuven. Training deep neural-networks using a noise adaptation layer. *ICLR*, 2017.

[17] Bo Han, Quanming Yao, Xingrui Yu, Gang Niu, Miao Xu, Weihua Hu, Ivor Tsang, and Masashi Sugiyama. Co-teaching: Robust training of deep neural networks with extremely noisy labels. *Advances in neural information processing systems*, 31, 2018.

[18] Hrayr Harutyunyan, Kyle Reing, Greg Ver Steeg, and Aram Galstyan. Improving generalization by controlling label-noise information in neural network weights. In *International Conference on Machine Learning*, pages 4071–4081. PMLR, 2020.

[19] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9729–9738, 2020.

[20] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[21] Huaxi Huang, Hui Kang, Sheng Liu, Olivier Salvado, Thierry Rakotoarivelo, Dadong Wang, and Tongliang Liu. Paddles: Phase-amplitude spectrum disentangled early stopping for learning with noisy labels. *arXiv preprint arXiv:2212.03462*, 2022.

[22] Pavel Izmailov, Dmitrii Podoprikhin, Timur Garipov, Dmitry Vetrov, and Andrew Gordon Wilson. Averaging weights leads to wider optima and better generalization. *arXiv preprint arXiv:1803.05407*, 2018.

[23] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning. *Advances in neural information processing systems*, 33:18661–18673, 2020.

[24] Junnan Li, Richard Socher, and Steven CH Hoi. Dividemix: Learning with noisy labels as semi-supervised learning. In *ICLR*, 2020.

[25] Shikun Li, Xiaobo Xia, Shiming Ge, and Tongliang Liu. Selective-supervised contrastive learning with noisy labels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 316–325, 2022.

[26] Sheng Liu, Jonathan Niles-Weed, Narges Razavian, and Carlos Fernandez-Granda. Early-learning regularization prevents memorization of noisy labels. *Advances in neural information processing systems*, 33:20331–20342, 2020.

[27] Sheng Liu, Zhihui Zhu, Qing Qu, and Chong You. Robust training under label noise by over-parameterization. In *ICML*, 2022.

[28] Xingjun Ma, Hanxun Huang, Yisen Wang, Simone Romano, Sarah Erfani, and James Bailey. Normalized loss functions for deep learning with noisy labels. In *International conference on machine learning*, pages 6543–6553. PMLR, 2020.

[29] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018.

[30] Giorgio Patrini, Alessandro Rozza, Aditya Krishna Menon, Richard Nock, and Lizhen Qu. Making deep neural networks robust to label noise: A loss correction approach. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1944–1952, 2017.

[31] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015.

[32] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, pages 618–626, 2017.

[33] Marcel Simon and Erik Rodner. Neural activation constellations: Unsupervised part model discovery with convolutional networks. In *ICCV*, pages 1143–1151. IEEE Computer Society, 2015.

[34] Hwanjun Song, Minseok Kim, and Jae-Gil Lee. Selfie: Refurbishing unclean samples for robust deep learning. In *International Conference on Machine Learning*, pages 5907–5915. PMLR, 2019.

[35] Yisen Wang, Xingjun Ma, Zaiyi Chen, Yuan Luo, Jinfeng Yi, and James Bailey. Symmetric cross entropy for robust learning with noisy labels. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 322–330, 2019.

[36] Xiaobo Xia, Tongliang Liu, Bo Han, Chen Gong, Nannan Wang, Zongyuan Ge, and Yi Chang. Robust early-learning: Hindering the memorization of noisy labels. In *ICLR*, 2020.

[37] Xiaobo Xia, Tongliang Liu, Bo Han, Chen Gong, Nannan Wang, Zongyuan Ge, and Yi Chang. Robust early-learning: Hindering the memorization of noisy labels. In *International conference on learning representations*, 2021.

[38] Xiaobo Xia, Tongliang Liu, Bo Han, Nannan Wang, Mingming Gong, Haifeng Liu, Gang Niu, Dacheng Tao, and Masashi Sugiyama. Part-dependent label noise: Towards instance-dependent label noise. *NeurIPS*, 33:7597–7610, 2020.

[39] Xiaobo Xia, Tongliang Liu, Nannan Wang, Bo Han, Chen Gong, Gang Niu, and Masashi Sugiyama. Are anchor points really indispensable in label-noise learning? *Advances in neural information processing systems*, 32, 2019.

[40] Tong Xiao, Tian Xia, Yi Yang, Chang Huang, and Xiaogang Wang. Learning from massive noisy labeled data for image classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2691–2699, 2015.

[41] Yu Yao, Tongliang Liu, Bo Han, Mingming Gong, Jiankang Deng, Gang Niu, and Masashi Sugiyama. Dual t: Reducing estimation error for transition matrix in label-noise learning. *NeurIPS*, 33:7260–7271, 2020.

[42] Li Yi, Sheng Liu, Qi She, A Ian McLeod, and Boyu Wang. On learning contrastive representations for learning with noisy labels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16682–16691, 2022.

[43] Jure Zbontar, Li Jing, Ishan Misra, Yann LeCun, and Stéphane Deny. Barlow twins: Self-supervised learning via redundancy reduction. In *International Conference on Machine Learning*, pages 12310–12320. PMLR, 2021.

[44] Matthew D Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In *ECCV*, pages 818–833. Springer, 2014.

[45] Linfeng Zhang, Xin Chen, Junbo Zhang, Runpei Dong, and Kaisheng Ma. Contrastive deep supervision. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXVI*, pages 1–19. Springer, 2022.

[46] Xiaopeng Zhang, Hongkai Xiong, Wengang Zhou, Weiyao Lin, and Qi Tian. Picking deep filter responses for fine-grained image recognition. In *CVPR*, pages 1134–1142. IEEE Computer Society, 2016.

[47] Zhilu Zhang and Mert Sabuncu. Generalized cross entropy loss for training deep neural networks with noisy labels. *Advances in neural information processing systems*, 31, 2018.

[48] Heliang Zheng, Jianlong Fu, Tao Mei, and Jiebo Luo. Learning multi-attention convolutional neural network for fine-grained image recognition. In *ICCV*, pages 5219–5227. IEEE Computer Society, 2017.

[49] Heliang Zheng, Jianlong Fu, Zheng-Jun Zha, and Jiebo Luo. Learning deep bilinear transformation for fine-grained image representation. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

[50] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba. Learning deep features for discriminative localization. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2921–2929, 2016.