CIFAR-10 Classification on Symmetric and Asymmetric Noise

Achraf Anas El Houssaini - Moetaz Mohamed

December 19, 2023

Abstract

This study addresses the challenge of learning with label noise using the CIFAR-10 dataset, a common benchmark in deep learning research. Recognizing the critical impact of label noise on the performance of machine learning models, we explore various methodologies to mitigate its effects. Our approach involves introducing label noise at different levels (10%, 30%, 50%, 80%, and 90%) to the CIFAR-10 dataset, both symmetrically and asymmetrically. We employ several deep learning techniques, adjusting hyperparameters, different loss functions and implementing strategies to enhance noise robustness. Through extensive experimentation, we evaluate the performance of these models, focusing on their ability to adapt to and learn from noisy labels. Our results offer valuable insights into the trade-offs between model complexity, training time, and accuracy in a noisy-label environment. This report contributes to the growing body of knowledge on handling label noise in machine learning, presenting practical approaches for improving model resilience in real-world scenarios where perfect data labeling is often unattainable.

1 Introduction

Deep learning has revolutionized numerous fields with its ability to learn complex patterns from large datasets. Its applications span a wide range, from image and speech recognition to natural language processing and autonomous vehicles. The success of deep learning largely depends on the quality and quantity of the data used for training models. However, real-world datasets often contain inaccuracies, particularly in the form of label noise, which can significantly degrade the performance of these models.

Label noise refers to errors in the labels of training data. In machine learning, accurately labeled data is crucial for training robust and effective models. When labels are noisy or incorrect, the model might learn incorrect patterns, leading to reduced accuracy and reliability. This issue is particularly significant in datasets where manual labeling is involved, as human error and subjectivity can introduce inconsistencies. The CIFAR-10 dataset, central to this study, is a staple in deep learning research for benchmarking image classification models. Comprising 60,000 32x32 color images across 10 classes, it offers a balanced and diverse set of images for assessing model performance. The dataset is split into 50,000 training images and 10,000 test images, covering a range of everyday objects from vehicles to animals.

In this project, we delve into the challenge of learning with label noise using the CIFAR-10 dataset. We introduce both symmetric and asymmetric noise at various levels (10%, 30%, 50%, 80%, and 90%) to investigate their impacts on deep learning models. Symmetric noise involves random mislabeling of classes without bias, whereas asymmetric noise simulates more realistic scenarios where certain classes are more prone to being mislabeled as specific other classes.

The project's scope includes evaluating the effectiveness of different deep-learning strategies and techniques in mitigating the impact of label noise and improving model accuracy and reliability.

2 Literature Review

Recently, deep learning has seen a lot of new methods that make models better at dealing with wrong labels in data. This review looks at important studies that have worked on this problem, using the CIFAR-10 dataset as a key tool for testing how good these methods are.

Robust loss functions such as Symmetric Cross Entropy (SCE) [WMC+19], Reverse Cross Entropy (RCE), and Normalized Cross Entropy (NCE) [MHW+20] have been identified as key contributors in this area. SCE addresses over-confidence in traditional cross-entropy by combining it with reverse cross-entropy, enhancing model robustness against noisy labels. RCE reduces the severity of penalties for incorrect class predictions, offering a counterbalance to standard cross-entropy. NCE normalizes loss values, making models less sensitive to data distribution scale, beneficial in imbalanced datasets with label noise.

The distinction between active and passive losses offers a refined view on managing label noise. Active losses, like Cross Entropy (CE), Normalized Cross Entropy (NCE), Focal Loss (FL), and Normalized Focal Loss (NFL), focus on optimizing probabilities for the correct class label. This approach enhances the model's accuracy for the correct label without directly suppressing incorrect label probabilities. In contrast, passive losses such as Mean Absolute Error (MAE), Normalized MAE (NMAE), Reverse Cross Entropy (RCE), and Normalized RCE (NRCE) do more. They not only optimize for the correct class but also actively reduce probabilities for incorrect classes. This makes them particularly effective in scenarios where it's crucial to lessen the impact of incorrect labels.

The Active Passive Loss (APL) [MHW+20] framework combines a robust active loss with a robust passive loss. This combination aims to enhance the strengths of both, ensuring robustness and comprehensive learning in environments with noisy labels. APL effectively addresses the limitations inherent to active and passive losses separately, resulting in improved performance in handling noisy labels.

In the work by Forouzesh et al. [For23], Xavier initialization is shown to play a pivotal role in models trained on noisy labels. It helps to keep the gradients well-scaled, thereby preventing the common issues of gradient vanishing or exploding. This balanced approach to weight setting is crucial for deep learning models to differentiate between useful signals and label noise during training, promoting better generalization despite the presence of label errors.

In conclusion, recent developments in deep learning have significantly advanced the field's ability to handle noisy labels in datasets like CIFAR-10. Key breakthroughs include robust loss functions such as Symmetric Cross Entropy, Reverse Cross Entropy, and Normalized Cross Entropy, which enhance model robustness and accuracy in the face of label noise. Additionally, the distinction between active and passive losses offers a nuanced approach to managing label inaccuracies, with techniques like the Active Passive Loss framework integrating these concepts for improved performance.

3 Methodology

3.1 Data Description and Noise Incorporation

The CIFAR-10 dataset comprises 60,000 32x32 color images evenly distributed across 10 classes, each class representing different entities such as airplanes, automobiles, birds, cats, deer, dogs, frogs, horses, ships, and trucks. This dataset is particularly suited for developing machine learning models because it presents real-world complexities within a manageable scope. With 50,000 images designated for training and 10,000 for testing, CIFAR-10 provides a robust foundation for assessing the performance of algorithms under various conditions. In the context of our research, we will introduce label noise into the CIFAR-10 dataset to simulate the challenges models face when training data is imperfect. Label noise is categorized into two types:

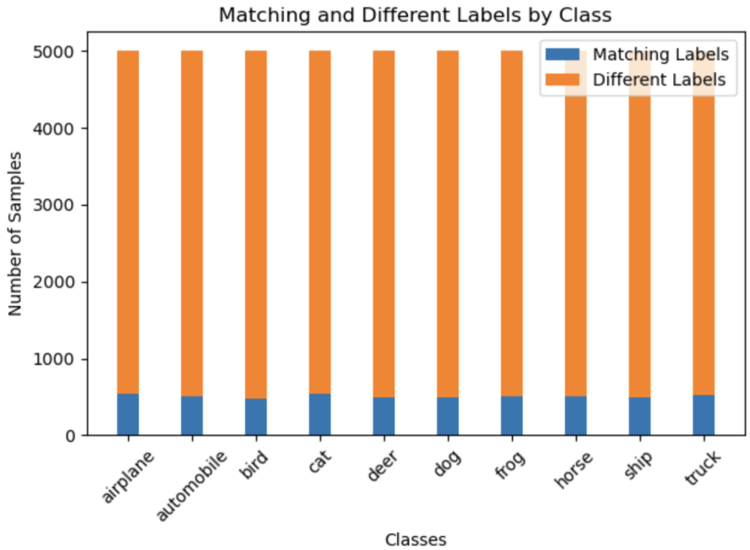

Figure 1: Symmetric noise at 90% on CIFAR-10 dataset.

3.2 Symmetric label noise

For this type, we will flip the correct label of an image to any other incorrect label with a probability corresponding to the noise level For example, if = 0.1, then for every 100 images of a bird, 90 will retain the correct label while 10 will be assigned a random incorrect label.

Figure 1 represents the distribution of matching and different labels in the CIFAR-10 dataset after the introduction of a high level (90%) of symmetric label noise. In this scenario, each class's original labels have been randomly flipped to any other label with a 90% probability. The blue sections of the bars show the small number of samples that have retained their correct label, whereas the overwhelming orange sections indicate the samples that have been assigned a different label due to the noise. With such a high level of noise, almost all the data points have been mislabeled, which is uniformly distributed across all classes. This visual representation highlights the extreme challenge that symmetric noise at this level presents to any classification algorithm, as the majority of the training data does not reflect the true attributes of each class. Such conditions require robust noise-handling mechanisms within the learning model to discern the underlying patterns despite the prevalent label inaccuracies.

3.3 Asymmetric label noise

This type reflects a structured mislabeling where certain classes are systematically mislabeled as specific other classes. Based on the noise level , we will flip the labels of the images as follows: trucks to automobiles, birds to airplanes, deer to horses, cats to dogs, and dogs to cats. For instance, if = 0.1 , then out of 100 bird images, 90 would still be labeled as birds, but 10 would be incorrectly labeled as airplanes.

Figure 2 illustrates the distribution of matching and different labels for each class in the CIFAR-10 dataset with a high level (90%) of asymmetric label noise introduced. Each bar represents one of the ten classes, with the blue portion indicating the number of samples that retained their original, correct label and the orange portion showing the number of samples that received a different, incorrect label due to the noise. At this extreme noise level, a significant portion of the original labels have been altered, as evidenced by the substantial orange segments in each bar. For instance, the 'bird' class would have had most of its instances relabeled as 'airplane', and similar transformations can be observed for other specified pairs, such as 'truck' to 'automobile' and 'cat' to 'dog'. The graph starkly demonstrates the challenge posed by such a high degree of label noise, where the vast majority of the data has been mislabeled, presenting a severe test for any noise-robust machine learning algorithm tasked with learning under these conditions. We will implement the noise in a controlled fashion at

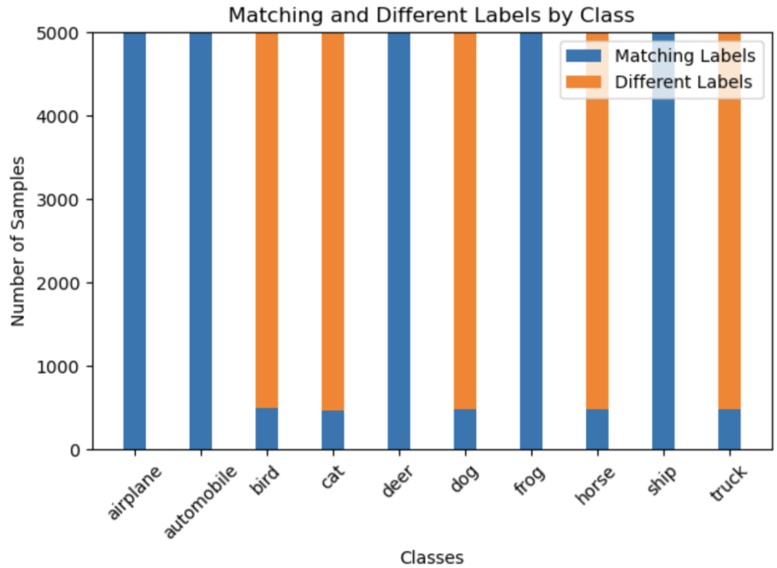

Figure 2: Asymmetric noise at 90% on CIFAR-10 dataset.

five distinct levels: 10%, 30%, 50%, 80%, and 90%. These levels are chosen to analyze the robustness of the algorithm across a spectrum from minimal to extreme noise conditions. The noise will be programmatically added to the training set before the learning process begins, ensuring that the validation and testing sets remain clean for accurate evaluation of the model's performance.

3.4 Data Processing and Manipulation

For data processing, the images were first converted into tensors and then normalized. This normalization involved adjusting the pixel values to have a mean and standard deviation of 0.5, ensuring that the input data for the neural network had a consistent scale and distribution. This step is crucial for facilitating effective learning and generalization.

In terms of data handling, the datasets were organized into batches for processing. The batch size was set at 64 for both training and test sets. This batch size strikes a balance between memory efficiency and the ability to accurately estimate the gradient during training. By processing data in batches, the model can efficiently learn from the dataset while managing computational resources effectively.

3.5 Base: CNN + Cross Entropy

In our research, we present our first model, a Convolutional Neural Network (CNN) designed for image classification tasks. It consists of three sequential sets of convolutional layers, each succeeded by a max-pooling layer to reduce spatial dimensions. The first set includes two convolutional layers, each with 32 filters, capturing basic image features. The second set escalates the complexity, featuring 64 filters in each of its two convolutional layers, allowing the network to detect more intricate patterns. The third set further deepens the feature extraction with 128 filters in each layer. Following these convolutional layers, the model transitions into dense layers. The feature maps are flattened and passed through two linear layers, reducing the dimensions to 128 nodes and finally to 10 output nodes corresponding to class probabilities. Throughout the network, ReLU activation functions are employed to introduce non-linearity, crucial for learning complex patterns. This model's architecture, devoid of batch normalization or dropout, emphasizes a straightforward yet effective approach for image classification, suitable for fundamental applications or less complex datasets.

3.6 Method 1: Regularized CNN + Symmetric Cross Entropy Loss

The second model represents an evolution of the CNN base model, introducing advanced features to enhance the network's robustness and generalization capabilities. This model maintains the foundational architecture of its predecessor, comprising three sets of convolutional layers, each followed by a max-pooling layer. However, it incorporates several key enhancements that differentiate it from the initial model.

One of the primary additions is the integration of batch normalization after each convolutional layer. Batch normalization standardizes the inputs to each layer, effectively reducing internal covariate shift. This is crucial as it leads to faster convergence during training, stabilizes the learning process, and enhances overall network performance. By normalizing the layer inputs, the model can use higher learning rates and be less sensitive to the initial weights, contributing to an overall improvement in training efficiency.

Another modification in the second model is the inclusion of dropout layers. These dropout layers, implemented with a rate of 0.25, randomly deactivate a subset of neurons during the training phase. This approach is important in preventing the network from becoming overly dependent on specific neurons, promoting the development of more generalized features. Dropout serves as a form of regularization, reducing the likelihood of overfitting and enhancing the model's ability to perform well on unseen data.

The model incorporates a new loss function called Symmetric Cross Entropy (SCE) Loss, designed to improve the model's handling of both asymmetric and symmetric label noise. This loss function combines two components: Categorical Cross-Entropy (CCE) and Reverse Cross-Entropy (RCE). CCE is the standard component for measuring the difference between predicted probabilities and actual labels, which is effective in clean label scenarios. However, its performance can be compromised in the presence of label noise. To address this, RCE is introduced. RCE calculates the entropy between the predicted probability distribution and one-hot encoded labels, encouraging the model to produce less confident predictions when encountering noisy labels. This reduces the negative impact of incorrect labels on the learning process[KDWG20]. The SCE loss is a weighted sum of CCE and RCE, controlled by the hyperparameters 'alpha' and 'beta', allowing for a balance between learning from accurate labels and mitigating the effects of label noise.

Finally, The model is optimized using the Adam optimizer. For symmetric noise, the learning rate is set to 0.001, and for asymmetric noise, it is set to 0.0001. This decision was based on extensive testing of various rates: 0.1, 0.01, 0.001, and 0.0001. In the context of asymmetric noise, the most effective combination for Symmetric Cross Entropy (SCE) was found to be $\alpha = 6$, $\beta = 0.1$, with a learning rate of 0.0001 in Adam. For symmetric noise, the optimal combination in SCE was determined as $\alpha = 0.1$, $\beta = 1$, with a learning rate of 0.001 in Adam. We also experimented with a more complex model of 8 layers but it was overfitting and had worst results compared to the less complex cnn with 6 layers.

3.7 Method 2: Advanced Regularized CNN + Normalized Cross Entropy and Reverse Cross Entropy Loss + Kaiming and Xavier Initialization

The Third Model, Advanced Regularized CNN, represents a significant evolution from the Regularized CNN model, particularly in its architectural framework and regularization strategies. Unlike the Regularized CNN, the Advanced Regularized CNN is more complex and does not incorporate dropout. It is structured into three main sequential blocks, each featuring two convolutional branches followed by a max-pooling layer. This design, with its systematically increasing depth of feature maps, enables the model to progressively extract and refine crucial features from the input data, an essential aspect in handling noisy labels. The convolutional layers are further enhanced by fully connected layers towards the end, which process the extracted features for classification tasks.

Furthermore, the Advanced Regularized CNN incorporates batch normalization after every convolutional layer and in the fully connected layers. This consistent application throughout the model strategically manages internal covariate shifts, ensuring stable and efficient learning, especially with datasets contaminated by noisy labels. Additionally, this model introduces a novel weight initialization technique, employing Kaiming and Xavier initialization for convolutional and linear layers, respectively. In contrast, the Regularized CNN model does not specify a particular weight initialization strategy. This focused approach in weight initialization is vital for optimizing network performance from the onset of training.

To address noisy labels, the Advanced Regularized CNN combines two robust loss functions using the Active Passive Loss (APL) concept. The first, Reverse Cross-Entropy (RCE), enhances robustness against noisy labels by focusing on the probabilities of incorrect labels and penalizing high-confidence wrong predictions. This approach, as detailed in [MHW+20], mathematically proves the robustness of RCE against noisy labels. The second loss function, Normalized Cross Entropy (NCE), refines the traditional Cross Entropy loss by normalizing the loss values, thereby making it more reliable in scenarios with noisy data. However, as shown in [MHW+20], robustness alone is not sufficient for effective training on noisy labels. The study reveals that both NCE and RCE, despite their robustness, can lead to underperformance due to underfitting issues.

To mitigate this, the Advanced Regularized CNN employs the APL methodology, combining the strengths of active and passive losses. The active loss, in this case, NCE, maximizes the probability of the correct class, enhancing precision. The passive loss, RCE, reduces probabilities for incorrect classes, addressing the underfitting issue prevalent in models using solely robust active losses [MHW+20]. This balanced approach, leveraging the robustness of active loss and the error mitigation of passive loss, culminates in a model adept at handling noisy datasets while maintaining learning depth and accuracy. The new loss function derived from APL methodology is called NCEandRCE

Finally, the model is optimized using Stochastic Gradient Descent (SGD), with a momentum parameter set to 0.9 and weight decay finely tuned to 10^{-4}. After thorough testing, we chose a learning rate of 0.01 because it performs significantly better than both 0.001 and 0.1, making it the optimal choice. Moreover, a batch size of 64 was selected, which yielded better results than batch sizes of 128 and 256. This smaller batch size is instrumental in reducing overfitting, an important consideration when training with noisy labels. The model's adaptability to different types of noisy data is further enhanced by these configurations. Alpha was chosen to be 10 and beta was equal to 1. These configurations were chosen based on the advised learning rates in [MHW+20]. We experienced adding dropout for the model, it fixed overfitting, but the model was underfitting and took double the number of epochs to give the same result as without dropout.

Overall, the Advanced Regularized CNN stands apart from the CNN Reguralized model with its modular architecture, heightened focus on dropout, specific weight initialization techniques, targeted use of batch normalization, and a more robust loss function. These features underscore a strategic approach tailored for complex datasets like CIFAR-10, emphasizing efficient feature extraction and robust regularization to effectively prevent overfitting.

4 Experimentation and Results

We trained the model with 10 different seeds and selected the best-performing iteration to ensure the robustness of our results. These training sessions were executed on Google Colab Pro, utilizing the high-performance NVIDIA Tesla V100 GPU.

4.1 Symmetric Noise

4.1.1 Base: CNN + Cross Entropy

The training loop was designed to evaluate the robustness of the model against symmetric label noise across varying noise levels, ranging from 0% to 90%. After some manual tuning we set a fixed number of epochs (n=15) to ensure consistency in the training duration across all noise levels. The model was optimized using the Adam optimizer with a learning rate of 0.001, which is a common choice due to its adaptive learning rate properties that help in converging faster and more efficiently.

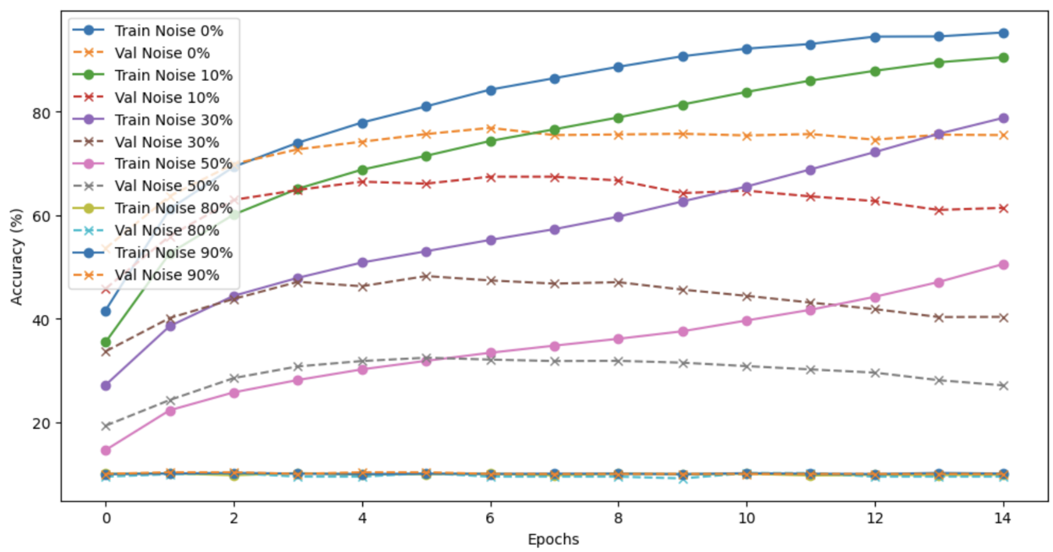

Figure 3: Training and Validation Accuracy of Symmetric noise for base model.

As depicted in the graph, the training and validation accuracies were tracked for each noise level at one epoch intervals. It is observed that with an increase in label noise, both training and validation accuracies deteriorate. However, the training accuracy consistently remains higher than the validation accuracy, indicating a degree of overfitting, particularly after more epochs. This overfitting is expected, as the model starts memorizing the noisy labels.

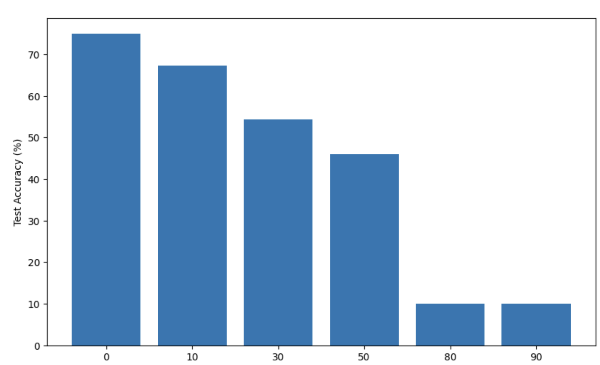

Figure 4: Testing Accuracy of Symmetric Noise for base model.

The experimental evaluation extended to testing the models on a dataset with non-noisy labels to establish a performance baseline. Notably, in the absence of label noise, the models demonstrated significantly higher test accuracies, emphasizing the detrimental impact of symmetric noise on model training. The stark contrast in performance between the non-noisy dataset and those with increasing levels of symmetric noise further accentuates this point. As the noise level escalates to 80% and 90%, the test accuracy sharply declines, approaching the lower bound of 10% accuracy, which is the expected outcome for random guesses in a 10-category classification task. This baseline comparison is critical as it clearly illustrates the extent to which symmetric noise can degrade the learning capability of a model, bringing its performance down to a mere random chance when the noise is exceedingly high, whereas the same model can achieve substantially higher accuracy when trained with clean, accurate labels.

4.1.2 Method 1: Regularized CNN + Symmetric Cross Entropy Loss

Training Method 1 with symmetric noise, we observe its learning capabilities are robust at lower noise levels. As noise is introduced, the validation accuracy starts to fall below training accuracy, indicating the beginning of overfitting to noisy labels. This trend is more pronounced when noise levels reach 50%, where the validation accuracy significantly lags, suggesting the model struggles with generalizing from highly noisy training data.

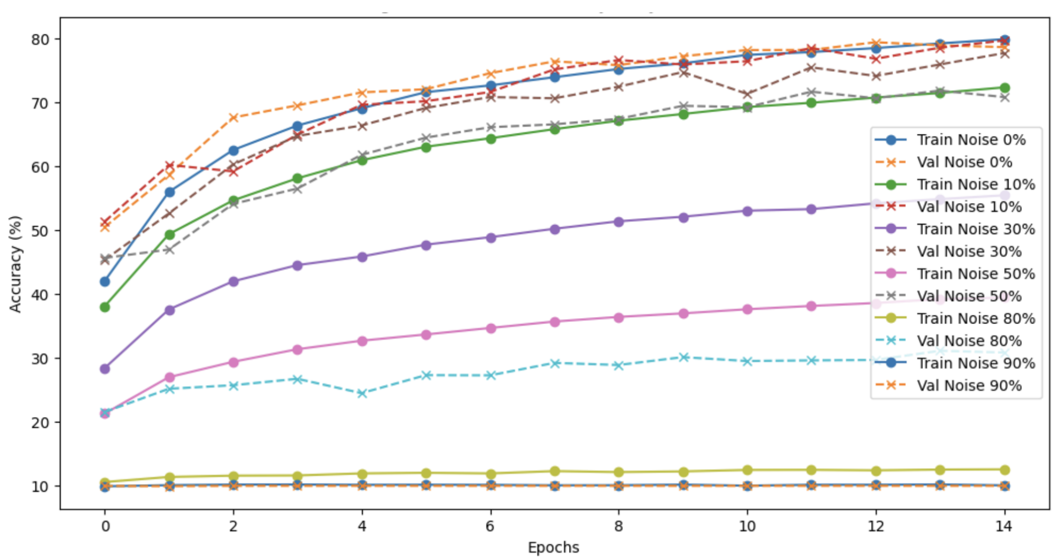

Figure 5: Training and Validation Accuracy of Symmetric noise for Method 1.

When testing Method 1 on clean data, the results demonstrate its resilience to a point. In an ideal setting with no noise, the model achieves its highest accuracy. It maintains a respectable performance up to a noise level of 80%. Beyond this, at 90% noise, there's a drastic decline in test accuracy, plummeting to 10%, which points to the model's inability to extract useful patterns amidst the noise.

Comparing the performance of Method 1 to the base model under symmetric noise conditions, it's evident that Method 1 significantly outperforms the base model across various noise levels. At 0% noise, Method 1 achieves a test accuracy of 84.65%, which is a marked improvement over the base model's 75.66%. This trend of superior performance continues at 10% noise, with Method 1 maintaining a high accuracy of 82.58%, compared to the base model's 67.348%. Even at 30% noise, Method 1 has a noticeable edge, with 76.25% accuracy over the base model's 56.83%. The gap narrows slightly at 50% noise, yet Method 1 still leads with 68.41% accuracy against the base model's 42.14%. At 80%, the model outperforms base model with 20% increase in performance as it reaches 32% accuracy. However, at an extreme noise level 90%, models' performances converge to an accuracy of 10.08%, indicating a threshold where noise overwhelms the learning capability of both models. Overall, Method 1's robustness to noise is significantly better than that of the base model.

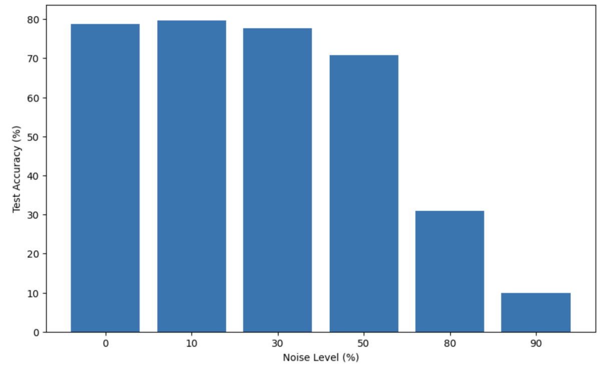

Figure 6: Testing Accuracy of Symmetric noise for Method 1.

4.1.3 Method 2: Advanced Regularized CNN + Normalized Cross Entropy and Reverse Cross Entropy Loss + Kaiming and Xavier Initialization

Method 2 exhibits a notable performance edge over the base model in handling symmetric label noise, highlighted by its superior test accuracy across all noise levels. With no noise present, Method 2 achieves a test accuracy of 82.378%, indicating a strong starting point compared to the base model's 75.66%. This advantage is maintained as noise is introduced, with Method 2 showing a minimal reduction to 81.19% at 10% noise level, contrasting the base model's more significant drop to 67.348%. The trend persists at higher noise levels, with Method 2 recording 76.58% and 70.99% at 30% and 50% noise levels, respectively, while the base model's performance deteriorates more drastically to 56.83% and 42.14%.

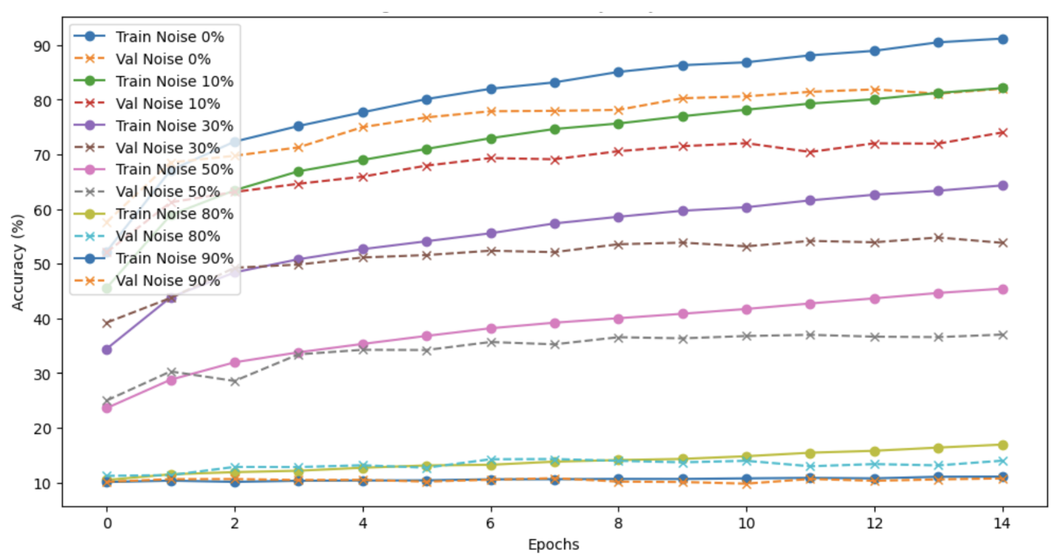

Figure 7: Training and Validation Accuracy of Symmetric noise for Method 2.

The training trajectory of Method 2, as depicted in Figure 7, showcases the model's stability and resilience. The training and validation accuracies demonstrate a consistent decline rather than abrupt drops as noise levels increase, which points to the model's effectiveness in generalizing from noisy training data.

This is further reflected in the model's performance at extreme noise levels; while the base model's accuracy falls to near-random guessing at 10.08% for both 80% and 90% noise levels, Method 2 manages to sustain significantly higher accuracies of 39.23% and 11.59%, respectively. This sustained performance under high noise conditions reinforces Method 2s suitability for deployment in environments with uncertain data quality.

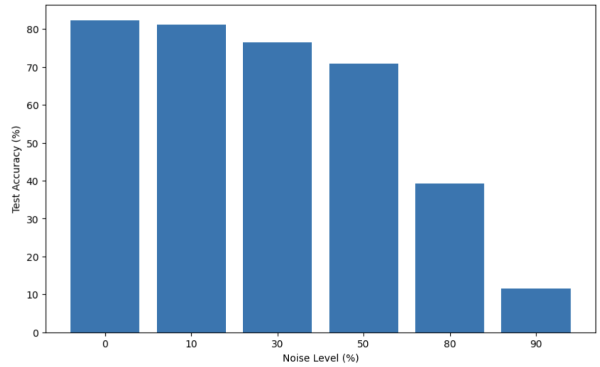

Figure 8: Testing Accuracy of Symmetric Noise for Method 2.

4.2 Asymmetric Noise

For the Asymmetric noise we applied the same training and testing configuration as the Symmetric noise experiment.

4.2.1 Base: CNN + Cross Entropy

The models subjected to asymmetric label noise over 15 epochs displayed a markedly better performance than those trained with symmetric noise. The structured nature of asymmetric noise allowed for a less severe degradation in accuracy as noise levels increased. Even at high noise levels, the models maintained accuracies significantly above the baseline of 10%, which represents random guessing in a 10-class classification task. This enhanced performance under asymmetric noise conditions is attributed to the model's ability to learn from the remaining structure in the data, despite the presence of label noise. The decision to extend training to 20 epochs was validated by the observed optimal results, demonstrating the effectiveness of longer training periods in combating asymmetric label noise.

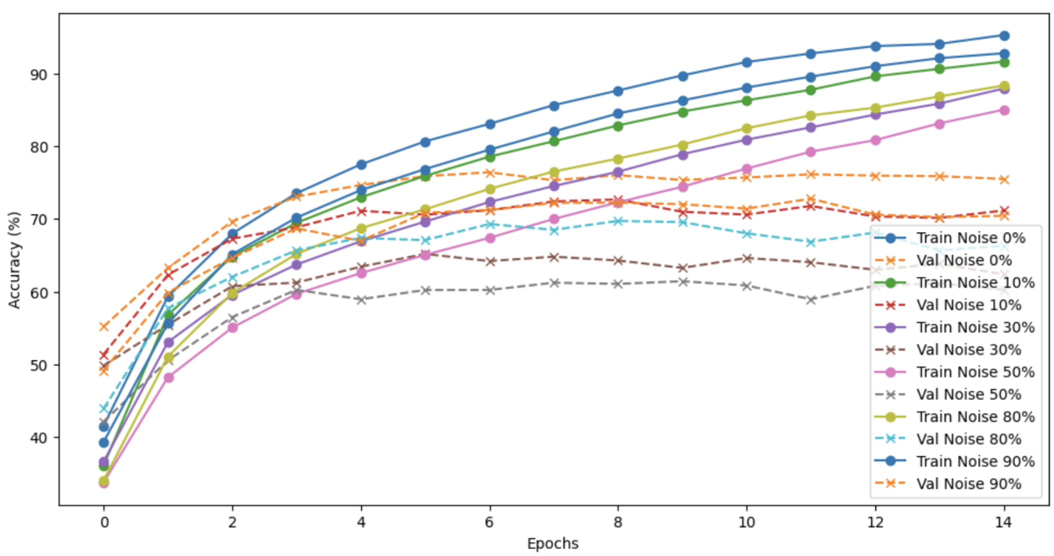

Figure 9: Training and Validation Accuracy of Asymmetric Noise for base model.

Figure 9 demonstrates the decline in model accuracy as label noise goes towards 50% noise level. At this point, the model is fed an equal mix of correct and incorrect labels, causing significant challenges in discerning the correct patterns. This leads to a notable decrease in validation accuracy, implying the model's inability to generalize well from such heavily corrupted data. The observed trend of higher training accuracy despite increased noise indicates a tendency towards overfitting, where the model learns the noise rather than the signal. This condition accentuates the need for more sophisticated methods to handle high levels of label noise, ensuring the model can learn effectively in spite of the corrupt data.

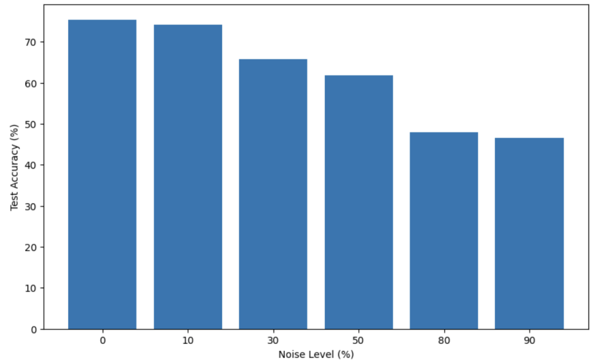

Figure 10: Training and Validation Accuracy of Asymmetric Noise for base model.

The test results for the models trained with asymmetric label noise show a markedly better resilience than those with symmetric noise. Despite the increase in noise levels, the accuracies significantly outperform the 10% random-guessing baseline, highlighting the models' robustness against asymmetric noise. This performance is a stark contrast to the much lower accuracies seen with symmetric noise, confirming that models handle asymmetric noise more effectively during testing.

4.2.2 Method 1: Regularized CNN + Symmetric Cross Entropy Loss

For 'Method 1', the resilience is evident up to a noise level of 30%, where the model maintains a high accuracy of 76.35%, only marginally less than the accuracy at 10% noise, which is 76.98%. This suggests that 'Method 1' can effectively learn from noisy data while mitigating the detrimental effects of asymmetric noise on model performance. Even at a 50% noise level, where the base model's performance significantly drops, 'Method 1' manages a respectable 63.378% accuracy, indicating its robustness up to this point of noise infusion.

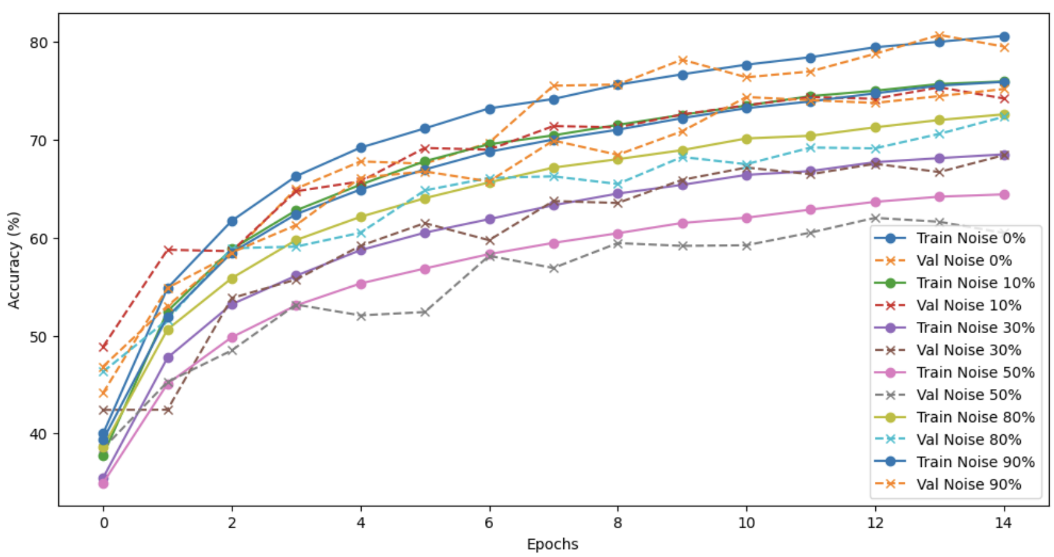

Figure 11: Training and Validation Accuracy of Asymmetric Noise for Method 1.

As the noise level reaches the extreme (80% and 90%), 'Method 1' shows a similar pattern to the base model, with accuracies dipping to just under 48%. This is expected since, at such high noise levels, the model is being trained on more incorrect than correct labels, making it exceedingly difficult for any model to discern the underlying pattern in the data.

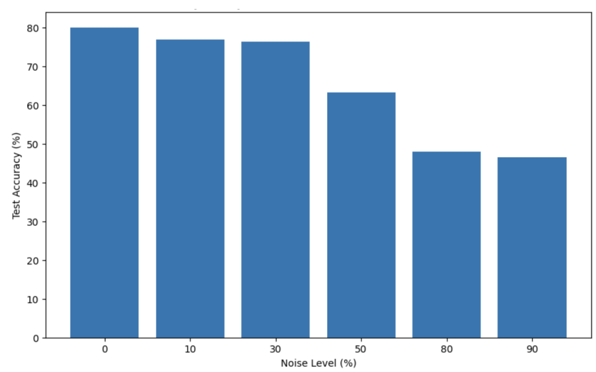

Figure 12: Testing Accuracy of Asymmetric Noise for Method 1.

The training graph for 'Method 1' indicates that the validation accuracy closely follows the training accuracy, unlike in the base model. This congruence between training and validation accuracy implies that 'Method 1' is effectively learning generalizable patterns from the noisy data, rather than merely memorizing the training set, which is a common problem when models are exposed to high noise levels. The closer alignment between the training and validation lines suggests that 'Method 1' can maintain its generalization capabilities even when trained on data with substantial label noise. This could be the result of an improved training algorithm, noise-robust loss function that helps the model ignore or correct noisy labels during training.

4.2.3 Method 2: Advanced Regularized CNN + Normalized Cross Entropy and Reverse Cross Entropy Loss + Kaiming and Xavier Initialization

Method 2 shows similar performance when compared to Method 1, especially in environments with moderate levels of noise. For instance, at 0% and 10% noise levels, Method 2 demonstrates superior accuracy, with 82.59% and 80.53% respectively, compared to Method 1's 80.09% and 76.98%.

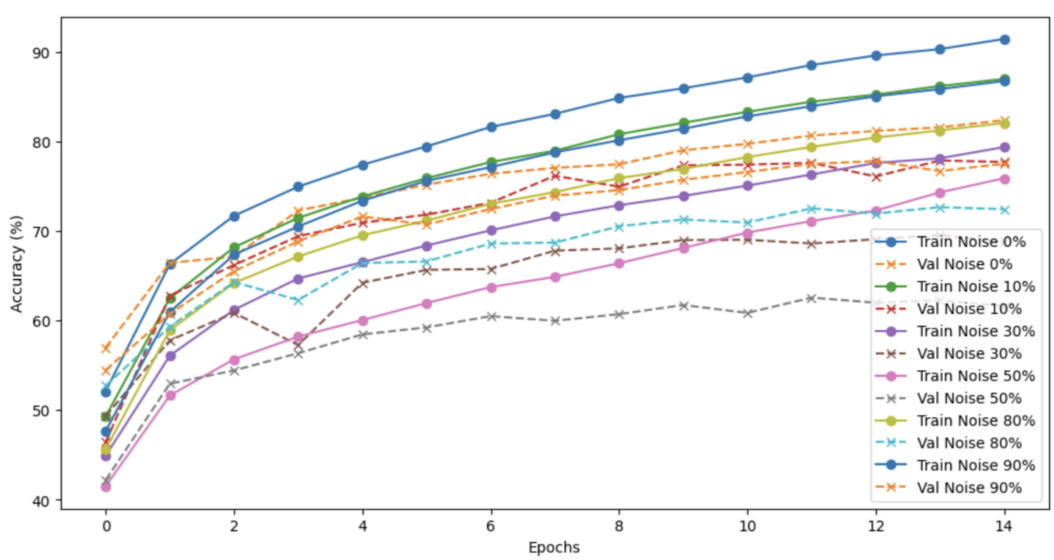

Figure 13: Training and Validation Accuracy of Asymmetric noise for method 2.

At a 30% noise level, 'Method 2' maintains a commendable accuracy of 76.45%, which is nearly on par with 'Method 1's 76.35%. This suggests that both methods are quite effective at dealing with moderate noise, possibly through sophisticated noise-filtering algorithms or robust training strategies that prevent the model from overfitting to the noisy data.

However, 'Method 2' faces a steeper decline in accuracy at the 50% noise level, recording 60.33% compared to 'Method 1's 63.378%. This could indicate that 'Method 1' might have a slightly more effective mechanism for handling higher noise levels, perhaps through a training regime that includes noise adaptation strategies or advanced regularization techniques.

Despite this, 'Method 2' stays competitive at extreme noise levels of 80% and 90%, with accuracies close to those of Method 1, suggesting that both methods have their merits and might employ different approaches to mitigate the impact of heavy label noise. Overall, Method 2 exhibits a strong ability to compete with Method 1, especially in scenarios with less noise, and maintains relatively close performance where the noise is substantial, making it a valuable approach in the arsenal of noise-robust machine learning methodologies.

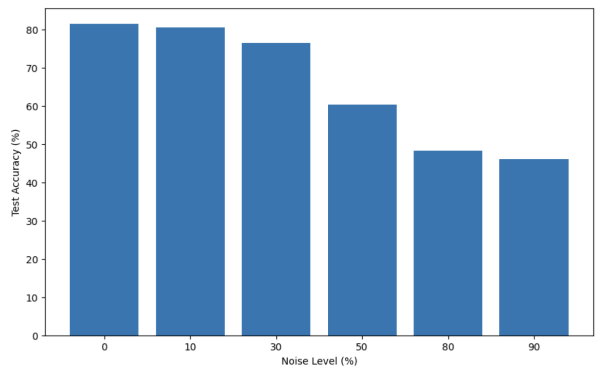

Figure 14: Testing Accuracy of Asymmetric Noise for method 2.

5 Discussion

In machine learning where tasks involve noisy data, the concept of entropy plays a crucial role. The symmetric cross entropy (SCE) is an amalgamation of reverse cross entropy (RCE) and traditional cross entropy. This blend is designed to enhance model robustness against label noise. However, a more advanced combination, which merges normalized cross entropy (NCE) with RCE, demonstrates even greater efficacy. According to recent research [MHW+20], normalizing any loss function theoretically boosts its robustness, a principle clearly evident when comparing the performance of SCE and the NCE+RCE combination.

Empirical results substantiate this theory, particularly in scenarios with high levels of noise. For instance, as the noise level escalates, a remarkable improvement in accuracy is observed when transitioning from SCE to the NCE+RCE combination. In experiments, with symmetric noise at an 80% level, Method 2 (employing NCE+RCE) achieved a 39.23% accuracy rate, outperforming SCE which capped at 30.08%. This 9.15% increase in accuracy underscores the effectiveness of the NCE+RCE approach under extreme noise conditions.

Interestingly, in asymmetric noise conditions, the results also support the superiority of NCE+RCE. For example, at a high noise level of 80%, Method 2 (NCE+RCE) attained an accuracy of 48.368%, compared to Method 1 (SCE) which achieved 47.92%. While the margin is smaller in asymmetric noise scenarios, it still indicates the enhanced capability of NCE+RCE in handling different types of noise.

However, the complexity of the model architecture and the use of dropout also significantly influence performance. In Method 1, which utilized a less complex architecture and incorporated dropout, SCE yielded better results in 50% compared to Method 3. After trials, we found that overfitting is more serious in asymmetric noise than symmetric. This explains why method 2 that had less complex architecture and dropout did it better on 50%.

Method	0	0.1	0.3	0.5	0.8	0.9
Base	75.66	67.34	56.83	42.14	10.08	10.08
Method 1	84.65	82.58	76.25	68.41	30.43	10.08
Method 2	82.37	81.19	76.58	70.99	39.23	11.59

Table 1: Models test accuracy on different Symmetric noise levels.

Method	0	0.1	0.3	0.5	0.8	0.9
Base	75.48	74.188	65.76	61.79	47.86	46.54
Method 1	80.09	76.98	76.35	63.37	47.92	46.62
Method 2	81.59	80.53	76.45	60.33	48.36	46.09

Table 2: Models test accuracy on different Asymmetric noise levels.

In conclusion, while the choice between SCE and NCE+RCE depends on the specific noise level and type, the latter generally offers enhanced robustness in higher noise environments. Nevertheless, factors such as model complexity, dropout usage, and susceptibility to overfitting must also be considered to optimize performance in various noisy conditions.

6 Conclusion

In conclusion, the Advanced Regularized CNN, distinguished by its innovative combination of NCE+RCE loss functions, superior weight initialization methods, and strategic application of batch normalization, has demonstrated remarkable robustness and efficiency in handling noisy datasets. Its modular architecture, tailored for complex environments like CIFAR-10, effectively prevents overfitting while ensuring deep and accurate learning. Empirical results show that this model outperforms the Regularized CNN, particularly in high-noise scenarios, thanks to its balanced approach in loss functions and focus on optimizing network performance right from the start of training. The careful configuration of learning parameters, such as the choice of SGD with fine-tuned momentum and weight decay, along with a smaller batch size, further enhances its adaptability to various types of noisy data. Method 2's strong performance across both symmetric and asymmetric noise conditions, as evidenced by the experimental results, highlights its potential as a robust solution in machine learning tasks involving noisy data. The research underscores the importance of considering model complexity, dropout usage, and the nature of noise in designing algorithms that can effectively manage challenging data environments.

References

[For23] Mahsa Forouzesh. Deep learning generalization with limited and noisy labels. Technical report, EPFL, 2023.

[KDWG20] Davood Karimi, Haoran Dou, Simon K Warfield, and Ali Gholipour. Deep learning with noisy labels: Exploring techniques and remedies in medical image analysis. *Medical image analysis*, 65:101759, 2020.

[MHW+20] 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.

[WMC+19] 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.