Disagreement Augmentation: A Socratic Approach to Knowledge Distillation

Anonymous ICCV submission

Paper ID 15298

Abstract

Disagreement plays a fundamental role in the learning process, driving deeper understanding and improved generalization. Inspired by the Socratic method, we introduce Disagreement Augmentation (DA), a novel approach to knowledge distillation that leverages disagreement between teacher and student models as a learning signal. Traditional distillation methods primarily focus on aligning the student with the teacher, minimizing divergence to transfer knowledge effectively. However, this approach may overlook critical underrepresented or ambiguous regions of the data distribution. Our method actively augments training samples to maximize disagreement between the student and teacher, encouraging the student to resolve conflicting predictions and develop a more robust approximation of the teacher. We evaluate DA in both an image classification setting and a reinforcement learning setting, demonstrating improved student model performance over typical baselines. These results highlight the potential of disagreement as a powerful augmentation strategy in knowledge distillation. Code and implementation details are available on GitHub.

1. Introduction

Disagreement is a catalyst for learning. This principle, rooted in the Socratic method, highlights the role of productive conflict in refining ideas and uncovering deeper truths. Socrates, through his method of dialectical questioning, often encouraged his students to confront contradictions in their beliefs, leading to a richer understanding of complex concepts. This pedagogical approach, centered on optimizing the interplay between opposing perspectives, inspires a new direction in knowledge distillation. We propose that, much like in Socratic dialogue, fostering disagreement between the teacher and student models can drive learning and enhance model performance.

Knowledge distillation traditionally aims to minimize the divergence between a large, well-trained teacher model and a smaller student model, transferring the teacher's expertise to create a compact, deployable version of the original system [9]. This approach focuses on alignment, where the student learns to emulate the teacher's soft predictions, thereby inheriting its generalization capabilities. However, this paradigm overlooks the potential benefits of disagreement—particularly as a mechanism to explore underrepresented or ambiguous aspects of the data distribution [21]. By intentionally optimizing for areas where the student and teacher disagree, we aim to emulate the Socratic process, leveraging conflict as a driver of more robust learning.

In this work, we introduce a novel method of data augmentation rooted in disagreement. Our approach, Disagreement Augmentation (DA), augments training samples to maximize divergence between the student and teacher models. These disagreement-optimized examples challenge the student to reconcile conflicting predictions, encouraging it to develop a more nuanced approximation of the teacher. This method of structured disagreement offers a complementary perspective to traditional distillation methods.

Beyond its conceptual motivation, DA is designed with practical advantages. In contrast to standard data augmentation, which typically applies predefined transformations (e.g., cropping, rotation, or noise injection) [2], DA generates task-specific augmentations tailored to expose the weaknesses of the student model. This targeted approach encourages the student to learn from its mistakes more effectively, leading to improved generalization and robustness. Moreover, DA aligns with recent efforts in self-supervised learning and contrastive learning, where the introduction of difficult training examples has been shown to enhance representation learning [1, 6].

We demonstrate the effectiveness of this approach across multiple domains, showing that DA improves both generalization and robustness. Our results suggest that disagreement-driven augmentation can serve as a valuable tool in knowledge distillation, offering a novel perspective on how models can learn more efficiently from one another. Through this work, we aim to bridge the gap between classical pedagogical insights and modern machine learning methodologies, reinforcing the notion that structured conflict—when properly harnessed—can be a powerful driver

ICCV #15298

of progress.

2. Related Work

Knowledge distillation, introduced by Hinton et al. [9], traditionally aims to transfer knowledge from a large, well-trained teacher model to a smaller student model by minimizing the divergence between their outputs. While this approach has been effective for model compression, recent work suggests that direct output matching may not always be optimal [10, 21]. The field has since evolved to recognize that valuable information exists not just in the teacher's predictions but also in the underlying learning process. Efficient knowledge transfer remains a key challenge, particularly under constraints of limited data or computational resources.

Recent studies have explored alternative approaches to distillation by considering additional aspects of model behavior beyond simple output alignment [19]. For instance, Maroto et al. [14] demonstrated that knowledge distillation can improve adversarial robustness, while Goldblum et al. [4] showed that adversarially robust teachers yield more resilient student networks. Of particular relevance to our work is the use of decision boundary information to enhance distillation [8], which conceptually aligns with our disagreement-based augmentation method.

Distillation has also been investigated in the context of adversarial defense. Methods such as Adversarial Diffusion Distillation [18] and Adversarially Robust Distillation [4] demonstrate how transferring robustness properties from a teacher to a student can significantly improve model reliability. These approaches highlight the importance of decision boundaries in distillation, reinforcing the idea that structured exploration of disagreement can lead to better knowledge transfer.

In parallel, advancements in data-free model extraction have shown that student models can be trained without requiring access to the original training data. Truong et al. [20] introduced Data-Free Model Extraction (DFME), a technique that synthesizes queries to extract knowledge from black-box models. This method builds on data-free knowledge distillation by leveraging generative models to construct inputs that maximize disagreement between the teacher (victim) and student (stolen) models. Similarly, Fang et al. [3] proposed Data-Free Adversarial Distillation, which employs adversarial techniques to generate informative samples for distillation without original data. Both approaches align with our work by demonstrating how disagreement can guide knowledge transfer, reinforcing the role of structured model divergence in improving student learning.

Our work builds directly on the Committee Disagreement Sampling approach introduced by Goldfeder et al. [5]. Their method identifies regions of the input space where knowledge transfer is most needed by analyzing disagreement between multiple student models. While their work focused on exact parameter reconstruction, we adapt this technique for the more flexible problem of knowledge distillation. This is closely related to adversarial sample generation, where model disagreement often highlights decision boundary regions susceptible to adversarial attacks [13]. By combining insights from disagreement-based learning and adversarial robustness, our method introduces a novel framework for enhancing knowledge transfer through structured exploration of model differences.

3. Methodology

3.1. Classification Experimental Setup

We conducted our image classification experiments on the CIFAR-100 dataset [11], with three configurations of student/teacher pairs: Resnet8x4/Resnet32x4, VGG8/VGG13, and ShuffleNet-V2/Resnet32x4 [7, 19]. We used the original knowledge distillation method proposed by Hinton et al. [9], though our augmentation should be compatible with more modern techniques as well. The student model was trained to minimize the weighted sum of the knowledge distillation loss and cross-entropy loss. Typical image augmentations were performed in both the baseline and DA experiments, such as random cropping and horizontal flipping. Experiments were run on a NVIDIA RTX 4090. All training runs used an SGD optimizer, a batch size of 64, 240 training epochs, an initial learning rate of 0.05, and learning rate decay at epochs 150, 180 and 210. The learning rate here refers to the typical student learning rate, not the DA learning rate α. Both DA experiments and baselines without DA were run 5 times each to ensure statistical reliability, with results reported as the mean and standard deviation across these runs.

3.2. Disagreement Augmentation Algorithm

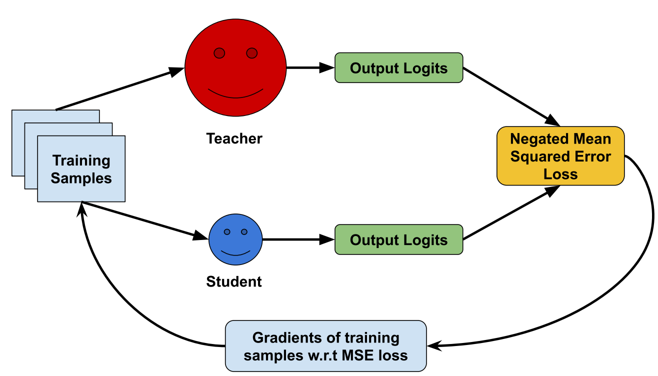


Figure 1. Schematic of the recursive DA algorithm. In practice only one epoch of augmentation occurs per batch.

The Disagreement Augmentation algorithm is designed to optimize input data by emphasizing areas of disagreement between a teacher model and a student model. The process begins by freezing the weights of both the teacher (T) and student (S) models to ensure that the augmentation process only modifies the input batch (I).

For each iteration of augmentation, the input batch is forward-propagated through the teacher and student models to compute their respective output logits, denoted as L_T and L_S. These logits are then normalized to ensure they are on a comparable scale. The algorithm computes a disagreement loss l as the negative Mean Squared Error (MSE) between the normalized logits of the teacher and student: $l = -\text{MSE}(L_S, L_T)$. This loss function incentivizes maximizing the discrepancy between the models' predictions.

The disagreement loss is backpropagated to compute gradients with respect to the input batch I. These gradients are then used to update I directly, employing a fixed learning rate α. This process is repeated for a predefined number of epochs e, iteratively refining the input batch to amplify disagreement between the models.

Once the iterations are complete, the optimized input batch I is returned as the final output of the algorithm, and used to train the student in typical knowledge distillation fashion. This approach ensures that the augmented data emphasizes areas where the teacher and student models diverge, challenging the student model to learn more robust and generalizable features.

Algorithm 1 Disagreement Augmentation Algorithm

Require: Student S, teacher T, input batch I, learning rate α, epochs e
  procedure DA(I, S, T, α, e)
    Freeze weights of S and T
    for each epoch i in 1 to e **do**
      Forward-propagate I through S and T
      Compute logits: $L_S = S(I)$, $L_T = T(I)$
      Normalize logits: $L_S \leftarrow \text{Normalize}(L_S)$, $L_T \leftarrow \text{Normalize}(L_T)$
      Compute disagreement loss: $l = -\text{MSE}(L_S, L_T)$
      Back-propagate l and compute gradient w.r.t. I
      Update I using α
    end for
    Return I
  end procedure

3.3. Policy Distillation Setup

To extend our method to reinforcement learning environments, we modified the original single-environment policy distillation methodology introduced by Rusu et al. [17]. Our setup consists of three main stages: online data collection, disagreement augmentation, and policy distillation. Both the student and teacher models are 4 layer deep Q-networks (DQNs), with 3 convolutional layers and one feed-forward layer [15]. The teacher DQN consists of 1.6 million paramaters, while the student has only 1% of that with roughly 16,000 (varies slightly across environments due to different action spaces). Students are trained for 500 epochs with a batch size of 32, a learning rate of 0.0001, and an SGD optimizer.

3.3.1. Online Data Collection

In this stage, a teacher pre-trained on an Atari environment is used to collect environment states. We used pre-trained teachers from RL Baselines3 Zoo [16]. The teacher interacts with the environment to generate trajectories, which are stored in a replay memory buffer. These stored states serve as the foundation for training the student. During each epoch of distillation, 54,000 environment states are generated, each consisting of 4 contiguous frames of the Atari game. The replay buffer has a capacity of 540,000 states, which it maintains by removing excess states when new ones are added in a first-in-first-out manner.

To ensure diversity in the collected training data, we introduced a 5% exploration rate during data collection. Specifically, for each action taken by the teacher model, there is a 5% probability of selecting a random action instead of the teacher's optimal policy decision. This controlled exploration helps capture a broader range of environment states, including suboptimal transitions that can improve the robustness of the student model.

3.3.2. Disagreement Augmentation

Once a batch of environment states is retrieved from the replay memory, we apply DA to emphasize areas in the state space where the student diverges from the teacher. The only difference between this instance of DA and what we used in the image classification setting is that we maximize the Kullback-Leibler divergence (KLD) between the student and teacher Q-values rather than the mean squared error. We did this to remain consistent with the policy distillation setting, as using KLD as the distillation loss is shown to improve performance over MSE [17]. For each batch of states, with probability $p = 0.3$ they undergo 1 epoch of DA with $\alpha = 0.001$ for Ms. Pacman and Space Invaders and $\alpha = 0.00001$ for Beam Rider.

3.3.3. Policy Distillation

Following the DA step, the policy distillation process takes place. The student policy is trained on the disagreement-augmented states, minimizing the KL divergence loss between its action distribution and that of the teacher. Unlike Rusu et al., who used RMSProp, we instead optimize the student's policy using SGD. This results in a training procedure that is more sensitive to the nuanced differences between the teacher and student policies, ensuring improved convergence dynamics.

4. Experimental Results

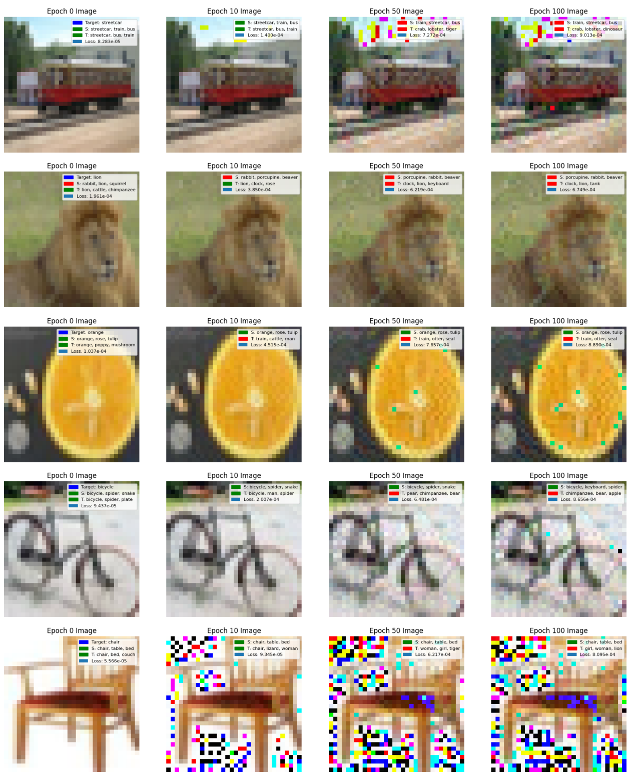


Figure 3. Examples of CIFAR-100 images undergoing various epochs of DA. The legend shows the ground truth target label, the Resnet8x4 student model's top 3 predictions, the Resnet32x4 teacher model's top 3 predictions, and the MSE loss between the student and teacher logits.

4.1. Hyperparamater Search for Classification

To optimize DA for our classification, we conducted a Bayesian hyperparameter search with Hyperband early stopping [12] over three hyperparameters: the number of epochs of augmentation per batch e, the learning rate of augmentation α, and the probability of augmentation per batch p. The search was conducted using a Resnet32x4 teacher and a Resnet8x4 student, with the goal of maximizing student validation accuracy. It found the ideal parameters to be $e = 1$, $\alpha = 0.01778$, and $p = 0.7374$. These are the parameters used in all classification experiments.

4.2. Classification Results

The results in Table 1 demonstrate that DA consistently improves student model performance across different architectures. For the Resnet32x4 to Resnet8x4 transfer, DA increases validation accuracy from 73.66% $\pm$ 0.26 to 74.59% $\pm$ 0.24, showing a clear performance gain. Similarly, for the VGG13 to VGG8 transfer, DA yields a modest improvement from 73.33% $\pm$ 0.25 to 73.76% $\pm$ 0.29. The most significant relative improvement occurs in the Resnet32x4 to ShuffleNet-V2 distillation, where DA raises accuracy from 71.67% $\pm$ 0.34 to 73.70% $\pm$ 0.19, suggesting that DA is particularly beneficial when distilling into more compact, efficiency-oriented architectures. These results highlight that disagreement-driven augmentation provides a complementary boost to standard knowledge distillation by encouraging more informative training dynamics. The improvements observed across all tested student models suggest that DA is a robust and effective augmentation strategy for classification tasks.

Table 1. Validation accuracy of baseline student models and student models trained with DA.

Teacher	Student	KD (%)	DA (%)
Resnet32x4	Resnet8x4	73.66 $\pm$ 0.26	**74.59** $\pm$ **0.24**
VGG13	VGG8	73.33 $\pm$ 0.25	**73.76** $\pm$ **0.29**
Resnet32x4	ShuffleNet-V2	71.67 $\pm$ 0.34	**73.70** $\pm$ **0.19**

4.3. Robustness to Disagreement Augmented Samples

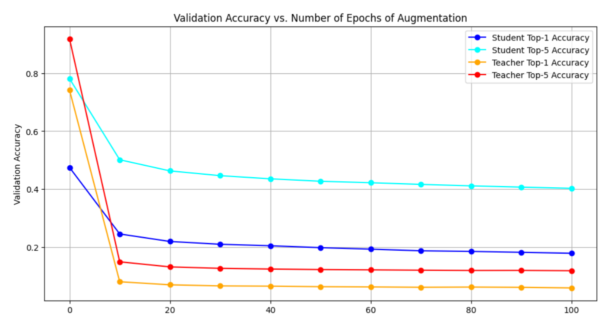


Figure 4. Validation accuracy an DA augmented validation set vs. number of epochs of augmentation.

We hypothesized that training a student model with disagreement augmented samples would result in a more robust model. To investigate, we evaluated the validation accuracy of a pre-trained Resnet32x8 teacher and a DA-trained Resnet8x4 student under varying levels of augmentation intensity, measured by the number of augmentation epochs. Here, augmentation occurs on the validation set to ensure that the evaluation reflects whether training with disagreement-augmented samples leads to improved robustness against such perturbations. Additionally, we compared the performance of the student model with its teacher to assess whether the knowledge distillation process, combined with disagreement-based augmentation, enables the student

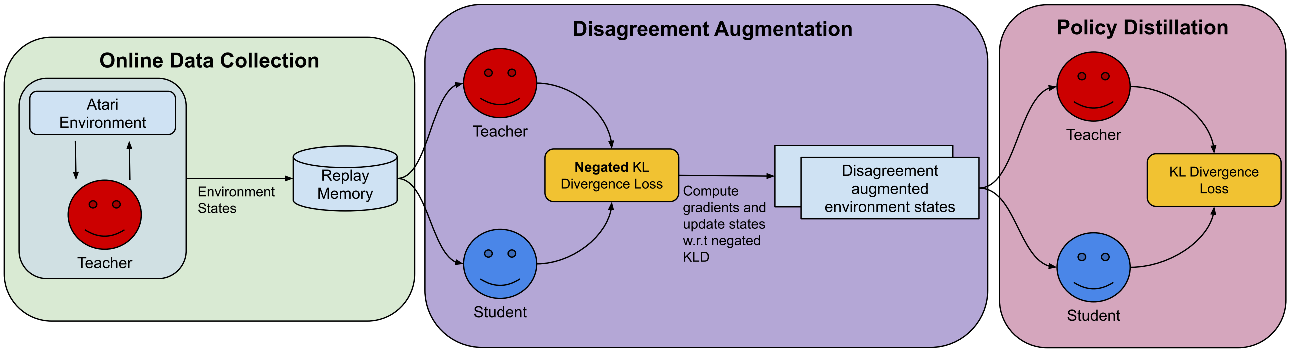


Figure 2. Policy distillation with Disagreement Augmentation.

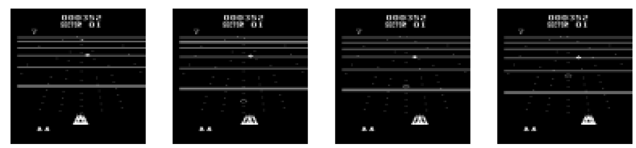
to achieve similar or superior resilience in handling these adversarial-like inputs. This approach allowed us to validate the hypothesis that disagreement-driven training fosters a more adaptable and robust student model.

4.4. Policy Distillation Results

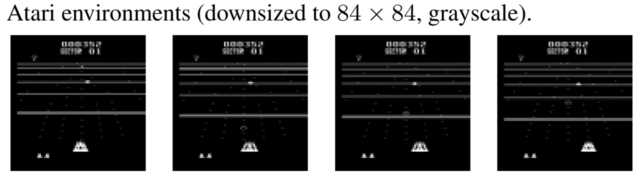
We evaluated DA in the policy distillation setting across three Atari environments: Beam Rider, Ms. Pacman, and Space Invaders. The goal of this experiment was to assess whether DA could improve the performance of a distilled student policy compared to standard behavior cloning using policy distillation.

The teacher policy in each case was a DQN model trained on the respective environment, while the student was a significantly smaller DQN model. The results are summarized in Table 2.

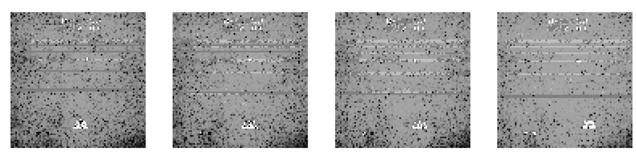
The results demonstrate that incorporating DA consistently improves student policy performance across all tested Atari environments. In Beam Rider, DA increases the student's average score from 2105.46 to 2663.53, raising its relative performance from 51.62% to 65.30% of the teacher's score. Similarly, in Ms. Pacman, DA enhances the student's performance from 3094.92 to 3341.92, yielding a relative improvement from 117.44% to 126.81%. The most notable increase occurs in Space Invaders, where DA raises the student's performance from 585.24 to 637.02, boosting the relative score from 112.24% to 122.17%. These results suggest that disagreement-driven augmentation enhances policy distillation by exposing the student to more informative training samples, leading to improved generalization and robustness.



4 frame stack from Beam Rider after typical preprocessing for Atari environments (downsized to 84×84, grayscale).



The same 4 frame stack after one iteration of DA, with $\alpha = 0.00001$.



Absolute difference between original and augmented frames, normalized to the range [0, 255] for visualization. This is equivalent to the (normalized) absolute value of the gradient of the input frames with respect to the negated KLD loss.

Figure 5. Example of how DA affects environment states. This was generated with a pre-trained teacher and a student trained with DA.

5. Discussion

5.1. Interpretation of Results

The results of our experiments demonstrate that incorporating DA into the knowledge distillation process significantly improves the generalization and robustness of student models. Across all tested configurations, models trained with DA consistently outperformed their baseline counterparts in both classification and reinforcement learning tasks. This suggests that the structured introduction of disagreement

Table 2. Policy distillation results across different Atari environments. Scores are calculated as average scores over 1000 episodes. The relative score represents student performance as a percentage of the teacher's performance.

Env	Method	Teacher Score	Student Score	Relative Score
Beam Rider	PD	4078.726	2105.460	51.62%
Beam Rider	PD + DA	4078.726	**2663.526**	**65.30%**
Ms. Pacman	PD	2635.370	3094.920	117.44%
Ms. Pacman	PD + DA	2635.370	**3341.920**	**126.81%**
Space Invaders	PD	521.405	585.235	112.24%
Space Invaders	PD + DA	521.405	**637.020**	**122.17%**

during training helps the student model better learn nuanced representations of the teacher's decision boundaries.

In the classification experiments, DA led to improved validation accuracy across all student architectures, with particularly strong gains in compact models such as ShuffleNet-V2. These improvements indicate that DA is especially beneficial for lightweight models, where standard distillation may struggle to fully capture the teacher's knowledge.

The reinforcement learning experiments further highlight the effectiveness of DA in policy distillation. In Atari environments, DA consistently improved student policy performance across all tested games, increasing relative student scores compared to standard policy distillation. Notably, the largest improvements were observed in Beam Rider and Space Invaders, where DA enhanced the student's ability to generalize across diverse game states. These results suggest that DA is not only beneficial in supervised learning but also in reinforcement learning settings, where effectively transferring policy knowledge remains a major challenge.

Furthermore, the robustness evaluation confirmed our hypothesis that disagreement-driven training fosters resilience to adversarial-like inputs. By augmenting the validation set to contain disagreement-optimized samples, we observed that DA-trained students were better equipped to reconcile these challenging inputs, achieving performance levels comparable to or surpassing their teachers. The improvements observed in both classification and reinforcement learning settings demonstrate that DA provides a generalizable mechanism for improving knowledge transfer.

5.2. Comparison with Previous Studies

Our findings align with and expand upon prior work that has explored the role of adversarial robustness in knowledge distillation. Although earlier studies, such as Goldblum et al. [4], demonstrated the benefits of robust teachers for improving student resilience, our method extends this concept by actively incorporating disagreement between models as a training signal. Compared to approaches like adversarially robust distillation, DA introduces a more generalizable framework that does not rely on predefined attack methods but instead leverages natural divergences between teacher and student predictions. This positions DA as a complementary and scalable strategy for enhancing robustness in distillation tasks.

In reinforcement learning, previous work on policy distillation has primarily focused on directly matching teacher policies [17], often struggling to capture uncertainty or out-of-distribution states effectively. Our results suggest that introducing structured disagreement in policy distillation improves knowledge transfer, potentially helping student policies generalize beyond trajectories demonstrated by the teacher. This complements recent studies on uncertainty-aware policy distillation, reinforcing the idea that controlled divergence can be a useful signal in both supervised and reinforcement learning settings.

5.3. Challenges and Limitations

Despite its promising results, DA is not without challenges. One limitation is the additional computational cost incurred during the augmentation process, as optimizing input batches over multiple epochs introduces overhead. While this cost was manageable in our experiments with CIFAR-100, scaling to larger datasets or models may require further optimization of the augmentation procedure. Similarly, in reinforcement learning environments, generating disagreement-optimized samples requires additional exploration, which can slow down training if not carefully managed.

Another limitation is the reliance on hyperparameter tuning to achieve optimal performance. As shown in our hyperparameter search, the number of augmentation epochs (e), learning rate (α), and probability of augmentation (p) are critical to the success of DA. While our method performed well across multiple settings, the need for manual tuning may limit accessibility. Automating or simplifying this tuning process could improve the scalability of the method, particularly for reinforcement learning applications where hyperparameter sensitivity is often high.

5.4. Future Directions

Future work could address the computational challenges of DA by exploring methods to reduce augmentation overhead, such as adaptive augmentation strategies that selectively apply disagreement-based modifications based on confidence thresholds. Additionally, while our reinforcement learning experiments demonstrated the benefits of DA in Atari environments, further research is needed to assess its effectiveness in more complex RL tasks, such as continuous control or multi-agent settings.

Another promising direction is extending DA to other domains, such as natural language processing (NLP) or self-supervised learning, where structured disagreement could help improve representation learning. For example, disagreement-based augmentation could be applied to NLP models by modifying token embeddings to create diverse training sequences, potentially leading to better generalization in text classification and translation tasks.

Finally, investigating the theoretical underpinnings of disagreement as a learning signal, particularly in the context of decision boundary exploration, could further refine and justify the approach. A deeper understanding of why and when DA is most effective could help develop more principled augmentation strategies that generalize across a broader range of learning tasks.

6. Conclusion

6.1. Summary of Findings

This work introduced Disagreement Augmentation (DA), a novel method for improving knowledge distillation by intentionally optimizing the input to maximize disagreement between teacher and student models. Inspired by the Socratic method, DA leverages structured conflict to challenge the student model, encouraging it to develop more robust and generalizable representations.

Experimental results on CIFAR-100 demonstrated that DA-trained students consistently outperformed baseline models in validation accuracy and robustness to disagreement-augmented samples. Furthermore, extending DA to reinforcement learning environments showed that disagreement-driven augmentation significantly enhances policy distillation. In Atari games, DA improved student policies across all tested environments, increasing their ability to generalize beyond the teacher's demonstrated trajectories. These results suggest that DA is a versatile augmentation strategy applicable to both supervised and reinforcement learning tasks.

6.2. Contributions

Our primary contributions are as follows:

- The introduction of Disagreement Augmentation as a generalizable data augmentation strategy for knowledge distillation across both classification and reinforcement learning.
- Empirical validation of DA's effectiveness, demonstrating improved generalization and robustness across multiple teacher-student configurations in classification tasks and enhanced policy transfer in reinforcement learning.
- A conceptual shift in knowledge distillation, emphasizing the role of structured disagreement as a catalyst for learning.

Acknowledgment

This work was built on this codebase [22] [23].

References

[1] 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. 1

[2] Ekin D. Cubuk, Barret Zoph, Dandelion Mane, Vijay Vasudevan, and Quoc V. Le. Autoaugment: Learning augmentation strategies from data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 1

[3] Gongfan Fang, Jie Song, Chengchao Shen, Xinchao Wang, Da Chen, and Mingli Song. Data-free adversarial distillation. *arXiv preprint arXiv:1912.11006*, 2019. 2

[4] Micah Goldblum, Liam Fowl, Soheil Feizi, and Tom Goldstein. Adversarially robust distillation. In *Proceedings of the AAAI conference on artificial intelligence*, pages 3996–4003, 2020. 2, 6

[5] Judah Goldfeder, Quinten Roets, Gabe Guo, John Wright, and Hod Lipson. Sequencing the neurome: Towards scalable exact parameter reconstruction of black-box neural networks. *arXiv preprint arXiv:2409.19138*, 2024. 2

[6] Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Guo, Mohammad Gheshlaghi Azar, et al. Bootstrap your own latent-a new approach to self-supervised learning. *Advances in neural information processing systems*, 33:21271–21284, 2020. 1

[7] 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. 2

[8] Byeongho Heo, Minsik Lee, Sangdoo Yun, and Jin Young Choi. Knowledge distillation with adversarial samples supporting decision boundary. In *Proceedings of the AAAI conference on artificial intelligence*, pages 3771–3778, 2019. 2

[9] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015. 1, 2

[10] Cheng-Yu Hsieh, Chun-Liang Li, Chih-Kuan Yeh, Hootan Nakhost, Yasuhisa Fujii, Alexander Ratner, Ranjay Krishna, Chen-Yu Lee, and Tomas Pfister. Distilling step-by-step!

outperforming larger language models with less training data and smaller model sizes. *arXiv preprint arXiv:2305.02301*, 2023. 2

[11] Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009. 2

[12] Lisha Li, Kevin Jamieson, Giulia DeSalvo, Afshin Rostamizadeh, and Ameet Talwalkar. Hyperband: A novel bandit-based approach to hyperparameter optimization. *Journal of Machine Learning Research*, 18(185):1–52, 2018. 4

[13] Aleksander Madry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, and Adrian Vladu. Towards deep learning models resistant to adversarial attacks. *arXiv preprint arXiv:1706.06083*, 2017. 2

[14] Javier Maroto, Guillermo Ortiz-Jiménez, and Pascal Frossard. On the benefits of knowledge distillation for adversarial robustness. *arXiv preprint arXiv:2203.07159*, 2022. 2

[15] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. Human-level control through deep reinforcement learning. *nature*, 518(7540):529–533, 2015. 3

[16] Antonin Raffin. Rl baselines3 zoo. `https://github.com/DLR-RM/rl-baselines3-zoo`, 2020. 3

[17] Andrei A Rusu, Sergio Gomez Colmenarejo, Caglar Gulcehre, Guillaume Desjardins, James Kirkpatrick, Razvan Pascanu, Volodymyr Mnih, Koray Kavukcuoglu, and Raia Hadsell. Policy distillation. *arXiv preprint arXiv:1511.06295*, 2015. 3, 6

[18] Axel Sauer, Dominik Lorenz, Andreas Blattmann, and Robin Rombach. Adversarial diffusion distillation. In *European Conference on Computer Vision*, pages 87–103. Springer, 2024. 2

[19] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 2

[20] Jean-Baptiste Truong, Pratyush Maini, Robert J Walls, and Nicolas Papernot. Data-free model extraction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4771–4780, 2021. 2

[21] Xiaohan Xu, Ming Li, Chongyang Tao, Tao Shen, Reynold Cheng, Jinyang Li, Can Xu, Dacheng Tao, and Tianyi Zhou. A survey on knowledge distillation of large language models. *arXiv preprint arXiv:2402.13116*, 2024. 1, 2

[22] Borui Zhao, Quan Cui, Renjie Song, Yiyu Qiu, and Jiajun Liang. Decoupled knowledge distillation. In *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*, pages 11953–11962, 2022. 7

[23] Borui Zhao, Quan Cui, Renjie Song, and Jiajun Liang. Dot: A distillation-oriented trainer. *arXiv preprint arXiv:2307.08436*, 2023. 7