View Meta-Reviews

Paper ID
765

Paper Title
Adversarial Style Augmentation for Domain Generalization

Track Name
Main Track

META-REVIEWER #4

META-REVIEW QUESTIONS

2. Summary of the PC reviews and discussion among PC members. Please touch on the following aspects whenever relevant: (1) summarize the main claims/contributions of the paper, (2) assess the strengths and weaknesses of the paper with respect to review criteria, and (3) summarize the PC discussion if the discussion provides additional information beyond the reviews.
This work proposes an Adversarial Style Augmentation (ASA) technique for domain generalization. The proposed technique achieves SoTA performance on different DG tasks including classification, retrieval and segmentation. All four reviewers acknowledged the authors' rebuttal response and two of them raised their scores by 1. One of the main concerns raised about the PACS results has been adequately addressed by the authors. While the reviewers were mostly positive, none argued strongly for accepting the paper.

View Reviews

Paper ID
765

Paper Title
Adversarial Style Augmentation for Domain Generalization

Track Name
Main Track

Reviewer #1

Questions

1. {Summary} Please briefly summarize the main claims/contributions of the paper in your own words. (Please do not include your evaluation of the paper here).
This paper proposes a novel method for DG by Adversarial Style Augmentation. A large space is explored by adversarial learning and better performance is achieved than baseline methods.

2. {Strengths and Weaknesses} Please provide a thorough assessment of the strengths and weaknesses of the paper, touching on each of the following dimensions: novelty, quality, clarity, and significance.
Strength
1. This method proposes a novel method based on adversarial style augmentation.
2. The proposed method is simple yet effective.
3. The proposed method achieves better performance than baselines.

Weakness
1. This motivation of exploring a broader style space has been explored in previous method [1], while the author has not discussed with this work.
2. It is interesting to see how is the generated images with different styles like to better understand the proposed method.

[1] Learning to Diversify for Single Domain Generalization

3. {Questions for the Authors} Please carefully describe questions that you would like the authors to answer during the author feedback period. Think of the things where a response from the author may change your opinion, clarify a confusion or address a limitation. Please number your questions.
1. What is the difference between this work and previous work [1]?
2. How is the generated images with different styleslike?
3. Most of experiments are conducted under single source generalization setting, what is the performance of this method under leave-one-domain-out generalization setting on PACS?

After rebuttal
Part of my concerns are addressed, I raise my score to borderline accept.
4. The results of DSU in table is different from Table 3 of its original paper, what is the reason?

[1] Learning to Diversify for Single Domain Generalization

4. {Evaluation: Novelty} How novel are the concepts, problems addressed, or methods introduced in the paper?
Good: The paper makes non-trivial advances over the current state-of-the-art.

5. {Evaluation: Quality} Is the paper technically sound?
Good: The paper appears to be technically sound. The proofs, if applicable, appear to be correct, but I have not carefully checked the details. The experimental evaluation, if applicable, is adequate, and the results

convincingly support the main claims.

6. {Evaluation: Significance} How do you rate the likely impact of the paper on the AI research community?
Good: The paper is likely to have high impact within a subfield of AI OR modest impact across more than one subfield of AI.

7. {Evaluation: Clarity} Is the paper well-organized and clearly written?
Good: The paper is well organized but the presentation has minor details that could be improved.

8. (Evaluation: Reproducibility) Are the results (e.g., theorems, experimental results) in the paper easily reproducible? (It may help to consult the paper's reproducibility checklist.)checklist.)
Good: key resources (e.g., proofs, code, data) are available and sufficient details (e.g., proofs, experimental setup) are described such that an expert should be able to reproduce the main results.

9. {Evaluation: Resources} If applicable, how would you rate the new resources (code, data sets) the paper contributes? (It might help to consult the paper's reproducibility checklist)
Good: The shared resources are likely to be very useful to other AI researchers.

10. {Evaluation: Ethical considerations} Does the paper adequately address the applicable ethical considerations, e.g., responsible data collection and use (e.g., informed consent, privacy), possible societal harm (e.g., exacerbating injustice or discrimination due to algorithmic bias), etc.?
Good: The paper adequately addresses most, but not all, of the applicable ethical considerations.

11. (OVERALL EVALUATION) Please provide your overall evaluation of the paper, carefully weighing the reasons to accept and the reasons to reject the paper.
Borderline accept: Technically solid paper where reasons to accept, e.g., good novelty, outweigh reasons to reject, e.g., fair quality. Please use sparingly.

13. (CONFIDENCE) How confident are you in your evaluation?
Quite confident. I tried to check the important points carefully. It is unlikely, though conceivable, that I missed some aspects that could otherwise have impacted my evaluation.

14. (EXPERTISE) How well does this paper align with your expertise?
Very Knowledgeable: This paper significantly overlaps with my current work and I am very knowledgeable about most of the topics covered by the paper.

16. I acknowledge that I have read the author's rebuttal (if applicable) and made changes to my review as needed.
Agreement accepted

Reviewer #6

Questions

1. {Summary} Please briefly summarize the main claims/contributions of the paper in your own words. (Please do not include your evaluation of the paper here).
The paper proposes AdvStyle to tackle domain generalization. Since feature statistics in neural networks have been shown to be relevant to domain information, AdvStyle maximizes the task loss with respect to feature statistics to search for the most sensitive direction that can generate novel "domains" that confuse the model. By using such adversarial feature statistics for training, AdvStyle achieves clear improvements over previous style-based augmentation methods on image classification and instance retrieval.

2. {Strengths and Weaknesses} Please provide a thorough assessment of the strengths and weaknesses of the paper, touching on each of the following dimensions: novelty, quality, clarity, and significance.
Strengths

The idea of using adversarial gradient to find the worst-domain feature statistics is interesting and clearly differs from existing feature statistics-based augmentation methods. The findings could be of interest to the community.

The comparison with some feature statistics-based augmentation methods is comprehensive. Overall, the results are good and convincing.

Weaknesses

The noise vectors (i.e., epsilon_mu and epsilon_sigma) used in eq. 4 & 5 are not well evaluated. Since adding perturbations to feature statistics helps improve DG, as shown in previous work, one might be wondering if AdvStyle works because of the adversarial style or the Gaussian noises.

The connection built between AdvStyle and the IRM loss is good but perhaps misplaced. It would be better to move eq. 6 and the discussion to a small section, after introducing the proposed method.

3. {Questions for the Authors} Please carefully describe questions that you would like the authors to answer during the author feedback period. Think of the things where a response from the author may change your opinion, clarify a confusion or address a limitation. Please number your questions.
- Could the authors justify that the improvements are brought by the adversarial style rather than the noises?
- Fig. 2 is a bit confusing. Is the SAM module applied to one or multiple layers?

*** Post-rebuttal updates ***
The reviewer has read the rebuttal and other reviewers' comments. There is no critical issue about the novelty and results. The reviewer's rating remains unchanged.

4. {Evaluation: Novelty} How novel are the concepts, problems addressed, or methods introduced in the paper?
Good: The paper makes non-trivial advances over the current state-of-the-art.

5. {Evaluation: Quality} Is the paper technically sound?
Excellent: I am confident that the paper is technically sound, and I have carefully checked the details of proofs, if applicable. The experimental evaluation, if applicable, is comprehensive and the results are compelling.

6. {Evaluation: Significance} How do you rate the likely impact of the paper on the AI research community?
Good: The paper is likely to have high impact within a subfield of AI OR modest impact across more than one subfield of AI.

7. {Evaluation: Clarity} Is the paper well-organized and clearly written?
Good: The paper is well organized but the presentation has minor details that could be improved.

8. (Evaluation: Reproducibility) Are the results (e.g., theorems, experimental results) in the paper easily reproducible? (It may help to consult the paper's reproducibility checklist.)checklist.)
Good: key resources (e.g., proofs, code, data) are available and sufficient details (e.g., proofs, experimental setup) are described such that an expert should be able to reproduce the main results.

9. {Evaluation: Resources} If applicable, how would you rate the new resources (code, data sets) the paper contributes? (It might help to consult the paper's reproducibility checklist)
Good: The shared resources are likely to be very useful to other AI researchers.

10. {Evaluation: Ethical considerations} Does the paper adequately address the applicable ethical considerations, e.g., responsible data collection and use (e.g., informed consent, privacy), possible societal harm (e.g., exacerbating injustice or discrimination due to algorithmic bias), etc.?
Good: The paper adequately addresses most, but not all, of the applicable ethical considerations.

11. (OVERALL EVALUATION) Please provide your overall evaluation of the paper, carefully weighing the reasons to accept and the reasons to reject the paper.
Weak Accept: Technically solid, modest-to-high impact paper, with no major concerns with respect to quality, reproducibility, and if applicable, resources, ethical considerations.

13. (CONFIDENCE) How confident are you in your evaluation?
Quite confident. I tried to check the important points carefully. It is unlikely, though conceivable, that I missed some aspects that could otherwise have impacted my evaluation.

14. (EXPERTISE) How well does this paper align with your expertise?
Expert: This paper is within my current core research focus and I am deeply knowledgeable about all of the topics covered by the paper.

16. I acknowledge that I have read the author's rebuttal (if applicable) and made changes to my review as needed.
Agreement accepted

Reviewer #7

Questions

1. {Summary} Please briefly summarize the main claims/contributions of the paper in your own words. (Please do not include your evaluation of the paper here).
This paper extends the idea of augmentation through style for domain generalization into the ideas the parameters can be inferred through a worst-case training process (adversarial training). The idea is fairly straightforward, and expected to be very effective (at it turns out). The writing is also fairly smooth. The experiments seems somewhat limited though.

======
Post-rebuttal

Thanks for the clarification, especially the empirical differences in other settings. Some more clarifications on the figure illustration could be more helpful. I improved the score.

2. {Strengths and Weaknesses} Please provide a thorough assessment of the strengths and weaknesses of the paper, touching on each of the following dimensions: novelty, quality, clarity, and significance.
- strengths:
- the idea of the paper is very intuitive and straightforward, it should be intuitively expected that the idea will work well.
- the writing of the paper is fairly straightforward, easy to follow.

- weakness
- the experiments are quite limited in comparison to standard DG experiments with tons of methods missing.
- For example, why the typical PACS results ResNet50 are above the 80 accuracy line, while the numbers reported in this paper are still mostly at 67.1, e.g., see table 1 [1]

[1] The Two Dimensions of Worst-case Training and the Integrated Effect for Out-of-domain Generalization

3. {Questions for the Authors} Please carefully describe questions that you would like the authors to answer during the author feedback period. Think of the things where a response from the author may change your opinion, clarify a confusion or address a limitation. Please number your questions.
why the typical PACS results ResNet50 are above the 80 accuracy line, while the numbers reported in this paper are still mostly at 67.1?

How does the thrid row of Figure 1 (the position of the blue and green circles) calculated exactly (in this figure, not the methods listed in the paper for generic)? Or if these are only for illustration purposes, please clarify it.

4. {Evaluation: Novelty} How novel are the concepts, problems addressed, or methods introduced in the paper?
Good: The paper makes non-trivial advances over the current state-of-the-art.

5. {Evaluation: Quality} Is the paper technically sound?
Good: The paper appears to be technically sound. The proofs, if applicable, appear to be correct, but I have not carefully checked the details. The experimental evaluation, if applicable, is adequate, and the results convincingly support the main claims.

6. {Evaluation: Significance} How do you rate the likely impact of the paper on the AI research community?
Good: The paper is likely to have high impact within a subfield of AI OR modest impact across more than one subfield of AI.

7. {Evaluation: Clarity} Is the paper well-organized and clearly written?
Good: The paper is well organized but the presentation has minor details that could be improved.

8. (Evaluation: Reproducibility) Are the results (e.g., theorems, experimental results) in the paper easily reproducible? (It may help to consult the paper's reproducibility checklist.)checklist.)
Good: key resources (e.g., proofs, code, data) are available and sufficient details (e.g., proofs, experimental setup) are described such that an expert should be able to reproduce the main results.

9. {Evaluation: Resources} If applicable, how would you rate the new resources (code, data sets) the paper contributes? (It might help to consult the paper's reproducibility checklist)
Good: The shared resources are likely to be very useful to other AI researchers.

10. {Evaluation: Ethical considerations} Does the paper adequately address the applicable ethical considerations, e.g., responsible data collection and use (e.g., informed consent, privacy), possible societal harm (e.g., exacerbating injustice or discrimination due to algorithmic bias), etc.?
Not Applicable: The paper does not have any ethical considerations to address.

11. (OVERALL EVALUATION) Please provide your overall evaluation of the paper, carefully weighing the reasons to accept and the reasons to reject the paper.
Borderline reject: Technically solid paper where reasons to reject, e.g., poor novelty, outweigh reasons to accept, e.g. good quality. Please use sparingly.

13. (CONFIDENCE) How confident are you in your evaluation?
Very confident. I have checked all points of the paper carefully. I am certain I did not miss any aspects that could otherwise have impacted my evaluation.

14. (EXPERTISE) How well does this paper align with your expertise?
Expert: This paper is within my current core research focus and I am deeply knowledgeable about all of the topics covered by the paper.

16. I acknowledge that I have read the author's rebuttal (if applicable) and made changes to my review as needed.
Agreement accepted

Reviewer #8

Questions

1. {Summary} Please briefly summarize the main claims/contributions of the paper in your own words. (Please do not include your evaluation of the paper here).
This paper proposes an Adversarial Style Augmentation (ASA) method for Domain Generalization (DG). The ASA generates style-related statistics perturbation via adversarial training to improve the model's robustness to style variations, thus improving the model's generalization ability. Extensive experiments show that the proposed method achieves SoTA on different DG tasks including classification, instance retrieval, and segmentation.

2. {Strengths and Weaknesses} Please provide a thorough assessment of the strengths and weaknesses of the paper, touching on each of the following dimensions: novelty, quality, clarity, and significance.
+ Overall, this paper is well-written and organized, making it easy to follow.

+ The motivation for generating styles for DG is clearly clarified and the proposed ASA is well presented.

+ This paper further proposes a simple yet effective module, namely AdvStyle, to enable end-to-end training of ASA.

+ The proposed method achieves SoTA on different DG tasks including classification, instance retrieval, and segmentation.

+ Ablation studies and visualization are comprehensive, which verifies the effectiveness of the proposed methods.

- Some more analysis of the experimental results can make the conclusion more persuasive.

3. {Questions for the Authors} Please carefully describe questions that you would like the authors to answer during the author feedback period. Think of the things where a response from the author may change your opinion, clarify a confusion or address a limitation. Please number your questions.
Q1: In Tab.4, please analyze why the AdvStyle inserted after Res-3&4 can also improve performance. Note that Res-3&4 usually do not encode style information but semantic information.
Q2: In the leave-one-domain-out setting, why is the proposed method less effective, e.g. only achieving similar results to DSU (Li et al. 2022) in Tab. A3?
Q3: In Fig. A2, the ResNet50+DSU seems to be more discriminative than ResNet50+AdvStyle, please analyze the reason.

---------- Comments After Author Response ----------
Q2 and Q3 are well addressed. For Q1, if the ASA can also work in semantic space, this paper doesn't have to be limited by the "style" augmentation. Probably the motivation and narrative can be further improved by taking account of the semantic augmentation. After reading all the reviews, I still believe that this paper is well-motivated and novel, with comprehensive and convincing experimental results. So I will keep my initial rating.

4. {Evaluation: Novelty} How novel are the concepts, problems addressed, or methods introduced in the paper?
Good: The paper makes non-trivial advances over the current state-of-the-art.

5. {Evaluation: Quality} Is the paper technically sound?
Good: The paper appears to be technically sound. The proofs, if applicable, appear to be correct, but I have not carefully checked the details. The experimental evaluation, if applicable, is adequate, and the results convincingly support the main claims.

6. {Evaluation: Significance} How do you rate the likely impact of the paper on the AI research community?
Good: The paper is likely to have high impact within a subfield of AI OR modest impact across more than one subfield of AI.

7. {Evaluation: Clarity} Is the paper well-organized and clearly written?
Excellent: The paper is well-organized and clearly written.

8. (Evaluation: Reproducibility) Are the results (e.g., theorems, experimental results) in the paper easily reproducible? (It may help to consult the paper's reproducibility checklist.)checklist.)
Good: key resources (e.g., proofs, code, data) are available and sufficient details (e.g., proofs, experimental setup) are described such that an expert should be able to reproduce the main results.

9. {Evaluation: Resources} If applicable, how would you rate the new resources (code, data sets) the paper contributes? (It might help to consult the paper's reproducibility checklist)
Good: The shared resources are likely to be very useful to other AI researchers.

10. {Evaluation: Ethical considerations} Does the paper adequately address the applicable ethical considerations, e.g., responsible data collection and use (e.g., informed consent, privacy), possible societal harm (e.g., exacerbating injustice or discrimination due to algorithmic bias), etc.?
Not Applicable: The paper does not have any ethical considerations to address.

11. (OVERALL EVALUATION) Please provide your overall evaluation of the paper, carefully weighing the reasons to accept and the reasons to reject the paper.

Weak Accept: Technically solid, modest-to-high impact paper, with no major concerns with respect to quality, reproducibility, and if applicable, resources, ethical considerations.

13. (CONFIDENCE) How confident are you in your evaluation?

Quite confident. I tried to check the important points carefully. It is unlikely, though conceivable, that I missed some aspects that could otherwise have impacted my evaluation.

14. (EXPERTISE) How well does this paper align with your expertise?

Knowledgeable: This paper has some overlap with my current work. My recent work was focused on closely related topics and I am knowledgeable about most of the topics covered by the paper.

16. I acknowledge that I have read the author's rebuttal (if applicable) and made changes to my review as needed.

Agreement accepted