We sincerely thank all reviewers for their time and helpful comments. Below please see our response to the comments.

Q1 (**R1**). “what kind of guarantees," & “relations between domains and assumptions on test data." The assumptions was given in lines 47-52: domain-specific distributions are random draws from a “mother" distribution. We have asymptotic guarantees (as the number of domains and sample size in each domain increases). With a graphical representation to describe essential properties of the mother distribution (assumed in lines 105-106), domain adaptation (DA) reduces to inference on the graphical model, where all relevant information for the label in the target domain is exploited.

Q2 (**R1**) "successful if the target data distribution is a convex combination of source distributions?" Yes, this is a special case satisfying our assumption. However, in complex situations, one may need to use supernodes for sets of the variables, because of the independent change assumption in the graphical representation (lines 109 & 120-124).

Q3 (**R1**)what if changes in new data are broader Nice point. It violates our assumption in lines 47-52, and our method may fail. This situation is like extrapolation, and additional assumptions may be needed. Will discuss it.

Q4 (**R1**)“causality," “interventions," and “negative transfer." Interesting point. The connection between interventions and distribution changes has been discussed in the causality community; see, e.g., “Interventions and causal inference” (by Eberhardt and Scheines, 2007). Even if the graphical model is a causal representation, our technical assumptions are still needed to ensure successful transfer (please also refer to **Q3**).

Q5 (**R1**&**R2**&**R4**)“synthetic tailored to the assumptions" & “image example with two variables." Synthetic experiments aim to verify the validity of the proposal. Inevitably, we need to consider image classification, an important problem in DA, in which it is natural to consider image pixels together as a supernode (L122-124). We use $Y \to X$ in light of previous work, which is consistent with the results of our procedure, since $P(Y)$ does not change in the image datasets.

Q6 (**R1**) single-source case. For instance, if we know $P(Y)$ does not change (as an additional constraint), then we know the structure and can do adaptation. Without such constraints, multiple domains would be needed.

Q7 (**R1**&**R4**) Influence of errors in graph learning. Thanks for raising this practical issue. We have conducted quantitative analysis following your suggestions and found that the adaptation performance indeed becomes worse when the result of the first step is wrong. The results will be included in the paper. For instance, when the direction between X_2 and X_3 in the graph is wrong, the prediction accuracy drops to $92(6.3)$, $82.36(10.33)$, and $83.1(9.95)$, with 2, 4, and 9 source domains, respectively.

Q8 (**R1**)results in Table 1. As confirmed by additional experiments, it is because of randomness in the generated datasets. we keep the target domains to be those in the 9 source experiment, and randomly sample 2/4/9 source domains. The results of our method on 2, 4, and 9 source domains are $81.04(12.71)$, $82.88(10.31)$, and $83.90(9.02)$. However, the comparisons of the methods are still fair because they deal with the same datasets.

Q9 (**R2**&**R1**)computational complexity. Our procedure is essentially local graph learning (focusing on only Y and variables in its Markov blanket) and inference on graphical models. We will include the complexity for both modules. The complexity is not very sensitive to the original dimensions of the data, but to how large the Markov blanket is. Empirically, it took around 6 hours on the Wifi data with 69 variables. The LVGAN training is efficient on GPUs. It takes about 20 mins on the WiFi data and 60 mins on the digits data.

Q10 (**R2**))learning a graphical model over images. Please refer to **Q5**.

Q11 (**R2**)Different causal structure graph learning on simulated data. We appreciate your insight and valuable suggestions. We have done additional experiments regarding this point, please refer to **Q7**. Following your suggestions, we will include the results and discussions in the paper. We applied the whole procedure to learn the graph and do inference. We will also include the performance of graph learning in the paper.

Q12 (**R2**)visualize relevant features in real datasets. We have shown the graph learned on the Wifi data in the Appendix A6. However, we consider image pixels together as a supernode.

Q13 (**R2**)office and sentiment. Thanks for your suggestion. We are playing with the datasets you recommended.

Q14 (**R3**) relevant works [1][2]. Many thanks for recommending the recent works. We will cite them and discuss how they are connected to our work. [1] uses the source domains to find invariant features and use them to predict in the target domain, and is closely related, although we make use of both invariant and changing features. The two-sample test method [2] is actually a special case of our conditional independence test-based method, because the dependence between a continuous variable (e.g. X) and discrete variable (domain index C) is related to the discrepancy between $P(X|C = 1)$ and $P(X|C = 0)$. At the same time, we hope that the reviewer kindly understands why the two recent works were not discussed: the first was available on arxiv this year and the other was published at last year’s NeurIPS.

Q15 (R3) motivation of latent-variable GAN. We develop the latent-variable GAN model because it is an implicit model that is flexible in modeling distributions, especially for images. We agree that other methods, including GP-based ones, may work well in various scenarios. Will make it explicit in the paper.

Q16(R3) Multi-environment causal discovery. Yes. In fact, our augmented graph learning method is extended from a recent multi-environment causal discovery method CD-NOD. The details are given in Sec 3.1 and Appendix A3.

Q17(R4) DA performance on DAG. Thanks for your encouraging comments. Please refer to **Q7**.

Q18(R4) Enhance when DAG is not identifiable in images. Great idea! Learning a graph in hidden representation of images can further enhance our method.