We thank all the reviewers for the unanimous positive comments! Below we address questions raised by each reviewer.

Reviewer 1

Q:"The $\sim$ operator on edges is never defined"

A: "$\sim$" means stitching two adjacent paths (i.e., they share one endpoint) into a longer path. We will add a definition in the paper.

Q:"We always assume f_{ij} is an isomorphism between D_i and D_j and $f_{ji} = f_{ij}^{-1}$" Is this really an okay assumption in practice? Is anything given up by this assumption? Most parametric maps will not be injective."

A: This is to define the cycle-consistency basis. Note that when using the cycle-consistency basis, we use a soft-constraint to allow approximation and deviations from injectivity.

Q:" It would be nice to show this form of cycle-consistency optimization enables novel capabilities, rather than just improved quantitative results over existing methods"

A: Improved testing accuracy indicates better-learned representations. We will add one paragraph in the conclusions to discuss this. A thorough study is left for future research.

Q:" It could have been better with some figures containing more detailed qualitative results. This could aid in making the experimental setups easier to understand and would improve result interpretability."

A: We will add visual comparisons between our approach and baseline approaches on dense flow prediction in the supplemental material.

Reviewer 2

Q: The setting of allowing network parameters to vary across different edges – this seems create a lot of individual networks, which is less optimal in the real-world use case. Also, I wonder if the networks prone to overfitting? and how to prevent it?

A: Our approach is a relaxation of enforcing identical weights by using a soft constraint to enforce the similarity between network weights. Overfitting is not an issue in our experiments. There is a tradeoff between the number of individual networks (e.g., sharing network weights among a subset of networks) and the prediction accuracy. Since the network weights are similar, one way to address the storage issue is to use weight quantization techniques on weights differences to compress the network weights. Note that our approach only uses one network for predicting network flows during testing time.

Reviewer 3

Q: The algorithm needs to be summarized more explicitly. I would prefer adding some pseudocode for better explanation.

A: We will add pseudo-code to reflect the procedure of (1) Initial cycles, (2) cycles via optimization, and (3) cycles via sampling.

Q: "More details should be included in experiments. For example, the hyperparameters in line 217 are unexplained, and necessary comparison in terms of running time is missing."

A: The hyperparameter in line 217 is set as $L = 10|\mathcal{E}|$ in all of the experiments. The insect table provides the running time for all the experiments in this paper.

Q:" There are some ambiguous expressions, such as "s" in Theorem 4.1."

A: s is defined precisely in the appendix depending on the geometric properties of mapping network $\{f_{ij}\}$. Under some mild assumptions given in the appendix, s is a constant. We provide some justifications for such assumptions but leave out from the main text due to the space limitation.

Q:" It is recommended to change the title so that highlights of the work could be well reflected."

A: We will add condition number into the title, e.g., "A Condition Number for Joint Optimization of Cycle-Consistent Networks".