Re: Coupling-based Invertible Neural Networks Are Universal Diffeomorphism Approximators (ID=1064).

We thank the reviewers for reviewing our work. We will update the paper based on the suggestions.

<u>To Reviewer #1</u>

Q1-1. On what occasion would the diffeomorphic universality results be useful other than distribution approximation?
A1-1. As explained in lines 29–33 of the paper, we believe the result is relevant when these INNs are used to learn an invertible transformation, e.g., in feature extraction or independent component analysis.

Q1-2. Missing references [1, 2, 3].
A1-2. Thank you for pointing out the missing references. We will include them in an updated version: as for [1] and [3], we will explain the relation in Supplementary H. As for [2], we will introduce it in Section 5.2.

Q1-3. Can you comment on the argument of [4], and whether it contradicts the results presented in this work?
A1-3. We first note that it is hard to technically verify whether the argument of [4] is complete as it is not rigorously stated. That said, to our best understanding, the argument of [4] does not properly take into account the approximation perspective hence fails to prove the non-universality. More concretely, the counterexample proposed in [4] critically relies on the independence of the target distribution. However, as far as approximation is concerned, non-independent distributions can approximate independent distributions. Specifically, our Lemma 12 (Supplementary F) seems to circumvent the first case of the contradiction argument of [4] because Lemma 12 shows that we can approximate a nonlinear component-wise transformation by using affine coupling flows.

Q1-4. Does the difference between the permutation layers and the invertible linear layers essentially contribute to the diffeomorphism universality? Does this result imply that Glow has superior representation power over RealNVP?
A1-4. In terms of representation power, the difference between using the permutation group and using the general linear group is relatively small: as small as component-wise sign swapping (i.e., a layer to multiply some dimensions by -1). In fact, one can express the elementary operation matrices (hence the regular matrices) by combining affine coupling flows, permutations, and component-wise sign swapping. Therefore, we believe the difference between Glow and RealNVP is mainly in the efficiency of approximation rather than the capability of approximation.

Q1-5. Is it accurate enough to directly say Glow is universal (line 191) since the family of maps it uses is limited?
A1-5. We agree with your concern for confusion and would like to add to line 191 that the result may not immediately apply to the typical Glow models for image data that use the 1x1 invertible convolution layers and convolutional neural networks for the coupling layers. Our explanation presumed a situation where Glow (or other coupling flows) is applied to non-image data (e.g., [11, 13] in the paper). In this case, the 1x1 invertible convolution layers correspond to the general linear group, and our results apply. Extending our results to the case of image data is future work.

<u>To Reviewer #2</u>

Q2-1. Ideas for improving the presentation of lines 212–235.
A2-1. Thank you for your suggestions for improving the manuscript. We will add such explanations accordingly.

Q2-2. Bigoni et al, "Greedy inference with layers of lazy maps" may be related.
A2-2. Thank you for pointing out the connection. We will introduce the paper in Section 5.1 as existing work that proposed a distributionally universal class of CF-INNs (equipped with a KL-divergence approximation error bound).

<u>To Reviewer #3</u>

Q3-1. The results here are not quantitative. For another type of INNs (residual-flow based ones), some lower bounds on the number of layers required for distribution approximation are known (Kong and Chaudhuri, 2020).
A3-1. We believe that establishing the universality of a model class remains important as the first step toward understanding the representation power because the question of efficiency only makes sense when the model has universality. Nonetheless, we agree that quantitative evaluation is important, and we will investigate the question further in future work. Our results can provide a simple route to confirming the universality not only for the existing coupling-based flow layers, but also for those to be designed in the future for improved efficiency.

<u>To Reviewer #4</u>

Q4-1. Only limited architecture is analyzed.
A4-1. As other reviewers pointed out (e.g., Reviewer #2), the architecture of invertible neural networks based on affine coupling flows (ACFs) is widely adopted in practice, hence we believe the analysis is highly relevant to the community (please also see lines 19–20 of the paper and the references therein). Furthermore, the ACFs analyzed in our paper are often special cases of more sophisticated flow layer designs, thus the result readily extends to such other architectures as explained in lines 186–192 of the paper.

Q4-2. No experimental results are presented to demonstrate how the theory works on existing algorithms.
A4-2. We believe the present theoretical results have high importance by themselves as they tackle the long-standing question of the universality of coupling-based invertible neural networks. The literature on the empirical evaluation of the coupling-flow based INNs is relatively rich whereas these rigorous theoretical results have been missing ingredients, as Reviewer #2 also pointed out.