Paper Title: **Invertible Convolutional Flow**

Thanks to all the reviewers for their time and helpful comments!

Reviewer #1

Thank you for your valuable suggestions. We agree that including a table contrasting flow architectures, as you suggest, will greatly improve the presentation of past work. We are working to develop a clear diagram illustrating the CONF layer, and will include a diagram of this type in the camera ready. We will additionally reference and/or better motivate the design choices (eg, the use of ActNorm instead of batch normalization) for the camera ready.

Reviewer #2

Thank you for your detailed comments.

1. Details on 2D convolutions.
The results presented for the 1D case are based on the convolution-multiplication property and the discrete trigonometric transforms (DFT or DCT) of the signal. Since the multi-dimensional transforms can be expressed separably in terms of 1D transforms, the theoretical results presented in this work can therefore be extended to 2D or 3D convolutions and their corresponding block circulant or block Toeplitz matrices. In practice, we used 2D invertible convolutions and 2D DFT/DCT for image datasets, which are implemented using the convolution-multiplication property and the efficient 1D FFT algorithm, thanks to their separable property. The explanation of 2D convolutions will be greatly expanded to clarify these points in the final version.

2. Point-wise nonlinearities
The nonlinear gates can induce special properties on the intermediate activations by introducing extra terms in the loss functions that, as you mentioned, can be interpreted as regularizers on the *latent representation*. Indeed, the main novelty of this part is proposing an analytic approach to designing customized pointwise nonlinearities according to desired latent structures in the deep normalizing flow. This also helps better understand the role of nonlinear gates through the lens of their contribution to latent variables' distributions. As you suggested, we will revise Proposition 2 to better clarify these ideas.

3. Invertibilty of the convolutions
The log determinant Jacobian of the convolutions acts as a log-barrier in the objective function that in turn prevents the convolution kernel in the frequency domain, $w_f(n)$, from becoming zero, and hence guarantees the invertibility of the convolution transform. (Note that the guarantee holds for continuous time gradient descent. It is technically possible, though not observed in practice, that SGD could produce a non-invertible kernel.) This remark was moved to the appendix due to lack of space but will be incorporated back into the main body. Additionally, the space of non-invertible kernels is measure zero in the space of kernels (it's rare for an eigenvalue to be *exactly* zero), and so non-invertible kernels are unlikely to occur by chance.

Reviewer #3

Thank you for your comments and feedback. The expressivity/flexibility of the CONF is of a great deal of interest to us as well! In addition, we are very interested in better understanding the implicit bias over trained probabilistic models induced by this choice of architecture. We hope in future work to further explore these questions.