

1 We thank all reviewers for their positive and constructive feedback. Reviewers find strengths of the work to be a unified  
2 framework that includes not only flows and VAEs, but also surjections (R1,R2,R3,R4). Some reviewers note that this  
3 framework outlines a modular API that is both appealing and practical (R1,R3) and some note that they expect the  
4 community to build upon the work (R2,R4). We address individual concerns below.

#### 5 **Reviewer 1:**

- 6 • *Special case:* We apologize if our statement came across as dismissive of the previous work. That was certainly  
7 not our intent and we acknowledge that all these works have significant contributions in addressing the short-  
8 comings of flow models. We wanted to suggest that models like Augmented Normalizing Flows, Continuously  
9 Indexed Flows and others fit nicely in our SurVAE Flow framework and can be derived architecturally using  
10 SurVAE blocks. We will rephrase our exposition to better reflect this.
- 11 • *Diffusion perspective:* Thanks for bringing this to our attention. We will add a reference to (Tabak and  
12 Vanden-Eijnden 2010) in the updated version.
- 13 • *Niche problems:* Our intention was to demonstrate that common (surjective) operations such as abs, max, sort,  
14 etc. can be useful transformations in generative models. This showcases the fact that the framework can be  
15 used to model data that would be more difficult to model with regular flows and therefore suggests that further  
16 development of novel SurVAE layers might be worthwhile.
- 17 • *Permutation invariance:* The resulting model is permutation invariant  $p(Tx) = p(x)$ . Computing the true  
18 likelihood requires an intractable sum over all permutations. When the likelihood is approximated using  
19 Monte Carlo samples, you will get an unbiased approximation of the true likelihood. However, this unbiased  
20 approximation is also permutation invariant (when the random seed is not fixed).
- 21 • *Stochastic inverse parameterization:* We will (as much as the page limit allows) add a discussion on this in the  
22 main text. In our experiments we used only very simple stochastic inverse distributions (as discussed in App.  
23 J: Experimental Details).
- 24 • *Additional references:* We will add references to (Wu et al. 2020) and (Sohl-Dickstein et al. 2015). Thanks for  
25 pointing these out.

#### 26 **Reviewer 2:**

- 27 • *Image experiments:* In the image experiments, the max pooling surjection performs similarly to the slicing  
28 surjection (which corresponds to the regular multi-scale architecture). We do not claim that max pooling  
29 surjection gives improvements over baselines. Rather, we demonstrate that max pooling is a viable alternative  
30 for constructing multi-scale flows. Note that we use a max pooling layer with a simple stochastic inverse with  
31 no extra parameters (see App. J). Improvements can likely be made by using more sophisticated choices.
- 32 • *Computational complexity:* The reviewer is correct, the computational complexity is similar to other flow  
33 models (and always the same for models that we explicitly compare). We will add more details on this.
- 34 • *Delta functions:* We are unsure exactly which derivations are being referred to, but if it is the derivations of the  
35 surjective layers in App. E, F, G, H, more thorough derivations, leading to the same final results, can be made  
36 using the following steps: 1) Specify Gaussian distributions with standard deviation  $\sigma$  in place of the Dirac  
37 deltas. 2) Write out the likelihood contribution term and cancel terms in the numerator and denominator to get  
38 rid of  $\sigma$ . 3) Take the limit as  $\sigma \rightarrow 0$ , to obtain the likelihood contribution for the surjective transformation.
- 39 • *Neural Statistician:* The reviewer raises a nice point and it will be interesting to study this further. However,  
40 providing an informed discussion on this question requires a careful and controlled set of experiments with  
41 clearly defined tasks which we shall try to accommodate in our revision.
- 42 • *Experiment bolding:* In the table, we are comparing the baseline to the same model but with a max pooling  
43 surjection. The bold font thus only compares the performance of these two models. We apologize for this  
44 confusion and will clarify this in the revised version.

#### 45 **Reviewer 3:**

- 46 • *NICE:* Thanks for pointing out the derivation in NICE. We will include discussion of this.
- 47 • *Image experiments:* See Point 1 for Reviewer 2. Reviewer 3 makes a valid point about qualitatively comparing  
48 the latent spaces learnt using max pooling vs. slicing. This is worth exploring as a future direction.

49 **Reviewer 4:** It seems you forgot to include the Weaknesses section, which makes it hard for us to address your concerns.  
50 However, we hope that the other reviews, together with our response, will alleviate some of your concerns.

51 **All:** We have also noted all minor comments and will use these to improve the paper. Thanks again for your work.