

addition to coloring the topological landscape profile, we also annotate the basins with the critical points, including saddle points (orange dots) and minima (red dots). Interestingly, the distribution (or density) of saddle points and minima reflects local characteristics of the loss landscape, such as locally sharp or locally flat.

4. Empirical Evaluation

4.1. Visualizing Different Physical Constraints

In our first evaluation, we look at a set of physics-informed neural network (PINN) models trained to solve simple convection problems (Krishnapriyan et al., 2021). Here we aim to investigate the PINN’s soft regularization and how it helps (or fails to help) the optimizer find an optimal solution to a seemingly simple convection problem. We show how the shape and complexity of our topological landscape profiles change as a physical “wave speed” parameter is increased and the PINN fails to solve this seemingly simple physical problem. Specifically, we consider the one-dimensional convection problem, a hyperbolic partial differential equation that is commonly used to model transport phenomena:

$$\frac{\partial u}{\partial t} + \beta \frac{\partial u}{\partial x} = 0, \quad x \in \Omega, \quad t \in [0, T] \quad (2)$$

$$u(x, 0) = h(x), \quad x \in \Omega \quad (3)$$

where β is the convection coefficient and $h(x)$ is the initial condition. The general loss function for this problem is

$$L(\theta) = \frac{1}{N_u} \sum_{i=1}^{N_u} (\hat{u} - u_0^i)^2 + \frac{1}{N_f} \sum_{i=1}^{N_f} \lambda_i \left(\frac{\partial \hat{u}}{\partial t} + \beta \frac{\partial \hat{u}}{\partial x} \right)^2 + L_B \quad (4)$$

where $\hat{u} = NN(\theta, x, t)$ is the output of the NN, and L_B is the boundary loss. While increasing the physical wave speed parameter, β , should not necessarily make this a harder problem to solve, it can make PINN models harder to train. Interestingly, Krishnapriyan et al. (2021) related these failure modes to changes in the corresponding loss landscape, showing that it becomes increasingly complicated, such that optimizing the model becomes increasingly difficult. Here we explore these failure modes in more detail using three-dimensional and four-dimensional Hessian-based loss landscapes, finding more variability in the shape of loss landscapes near the transition between high and low-performing models.

In Figure 3, we show a heat map corresponding to the average relative error across different values of the physical wave speed parameter and across different learning rates. Interestingly, we observe that the error increases with this physical parameter, but more slowly for higher learning rates. The smallest learning rate displays higher error rates even for smaller values of the physical parameter. When looking at the loss landscapes, we observe consistently more funnel-like loss landscapes for the smaller values of β , corresponding to lower error (Figure 3.1). In contrast, we observe a consistently more bowl-like loss landscape for the larger values of β , corresponding to higher error (Figure 3.3). The funnel-like landscapes correspond to when the PINN models find a physically reasonable solution, albeit constrained to a smaller space of solutions by the physical wave speed parameter.

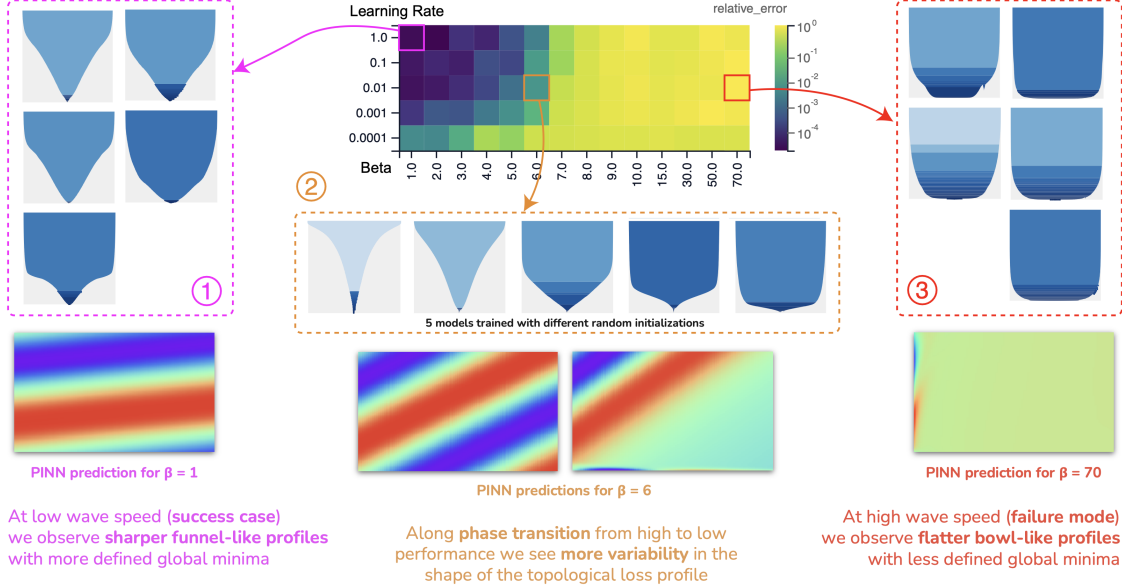


Figure 3: Analyzing the loss function of a physics-informed neural network (PINN) trained to solve simple physical convection problems. See Section 4.1 for details.

In other words, since the solution is constrained by the physical parameter, perturbing the model results in a faster increase in the loss, given that the physical problem is no longer satisfied. In contrast, the more bowl-like landscapes correspond to the failure to find a reasonable solution, such that perturbing the model does not immediately change the already high loss. Note, the landscapes corresponding to these failure modes also include more saddle points and are otherwise more complex.

To verify that these representations are stable across different model initializations, we show five different landscapes for each hyperparameter configuration, corresponding to the same model trained using different random seeds. We see the landscapes look similar across different random seeds for the low and high values of the physical wave speed parameter. Moreover, we observe more variation in the loss landscapes near the transition from low to high error (Figure 3.2). This suggests that while the error starts to increase near the transition, only some of the models are failing whereas other may be finding physically reasonable solutions, as indicated by the funnel-like loss landscapes.

In Figure A.5 we compare the topological landscape profiles based on three- and four-dimensional loss landscapes. An important insight here is that, in higher dimensions, we observe many more critical points and that the basins in the much spikier landscape can be mapped back to the wider basins in the topological landscape profiles based on the three-dimensional loss landscapes. Overall the global shape of the topological landscape profile looks similar when comparing the same random seed. Moreover, this highlights an important feature of our new representations—the ability to visualize higher-dimensional loss landscapes, i.e., sampling along more than just one or two dimensions.

4.2. Visualizing Loss Landscapes Over Training

In our second evaluation, we explore how loss landscapes change throughout training and across different learning rates. To do this, we study UNet models with a learnable CRF-RNN layer (Avaylon et al., 2022) trained on the Oxford-IIIT Pet dataset (Parkhi et al., 2012). We trained the models using five different random seeds across seven different learning rates for 30 epochs. For each checkpoint, we computed two-dimensional loss landscapes based on the top two Hessian eigenvectors. The model was perturbed using a distance of 0.01 and layerwise normalization was adopted (Li et al., 2018).

In Figure 4 and Figure B.6, we show the same heat map corresponding to average test accuracy over training and across different learning rates. We observe that the test accuracy improves over training, with some variation across the different learning rates. In Figure 4, we consider how the loss landscape changes over training. When looking at the loss landscapes for three different random seeds, after zooming in, we observe an initially shallow loss landscape but with the global minimum at a much higher loss compared to the end of training. As training proceeds, we see that the global minimum becomes lower, but the basin itself becomes deeper with the edges remaining at much higher loss values. As the global minimum continues to drop, we also observe additional flattening of the basin, such that all points have a much lower loss compared to the beginning of training. Interestingly, the flat basin at a much higher loss corresponds to a phase of learning where perturbing the model in any one direction doesn't really increase the already high loss. After five epochs, the much deeper basin reflects a less stable model, where perturbing the model results in relatively higher loss. As training proceeds, we observe a flattening of the basin, which

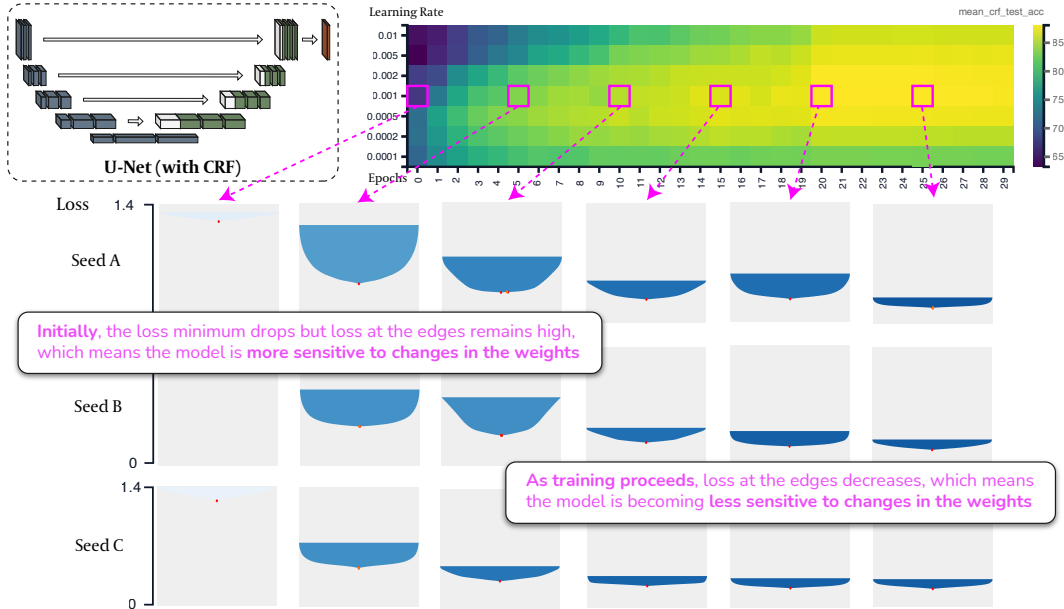


Figure 4: Loss landscapes over training for UNet models with a CRF layer trained on the Oxford-IIIT Pet dataset. See Section 4.2 for details.