In this paper the authors discuss the various challenges of visualizing the loss landscape. To solve these challenges, they introduced a new visualization scheme called ‘filter normalization.” Using their strategy, they also explore how the loss landscape changes for different neural net architectures. To start, the authors point to various theoretical studies that discuss our ability to optimize loss functions, and that these studies make restrictive assumptions. They then go on to discuss the challenges of visualizing the loss landscape in 1 dimension. Referring to work done by Ian Goodfellow, they point out that 1D plots 1) make it difficult to visualize the convex/non-convex structure of the landscape, and 2) do not consider batch normalization or invariance symmetries in a network. These cause 1D plots to be misleading. The next type of visualization the authors discuss are contour plots with random directions. These fall short because of 1) computational burden and 2) that they generally don’t create high quality images of small regions. This low quality makes it hard to capture the complexity of the loss surface.

Next, they discussed how the random directions approach leads to scale invariance in the weights of the network. To remove this scaling effect, the authors use filter-wise normalized directions to great effect. Using this scheme of visualization, the authors were able to determine that 1) the sharpness of minimizers correlate well with generalization error, 2) the depth of a network has a great effect on the loss surface when skip connections are not used (convex to chaotic), 3) skip connections in a network reduce the chaotic nature of the loss surface in deep networks dramatically, and 4) wider networks with more filters in each layer also decrease the chaotic behavior. In addition to these findings, they were able to quantify the convexity of an area using the ratio between the minimum and maximum eigenvalues of the Hessian. They also used PCA directions to effectively visualize the loss trajectory during training.