Despite excellent performance on stationary test sets, deep neural networks (DNNs) can fail to generalize to out-of-distribution (OoD) inputs, including natural