

Manifold Mixup: Encouraging Meaningful On-Manifold Interpolation as a Regularizer

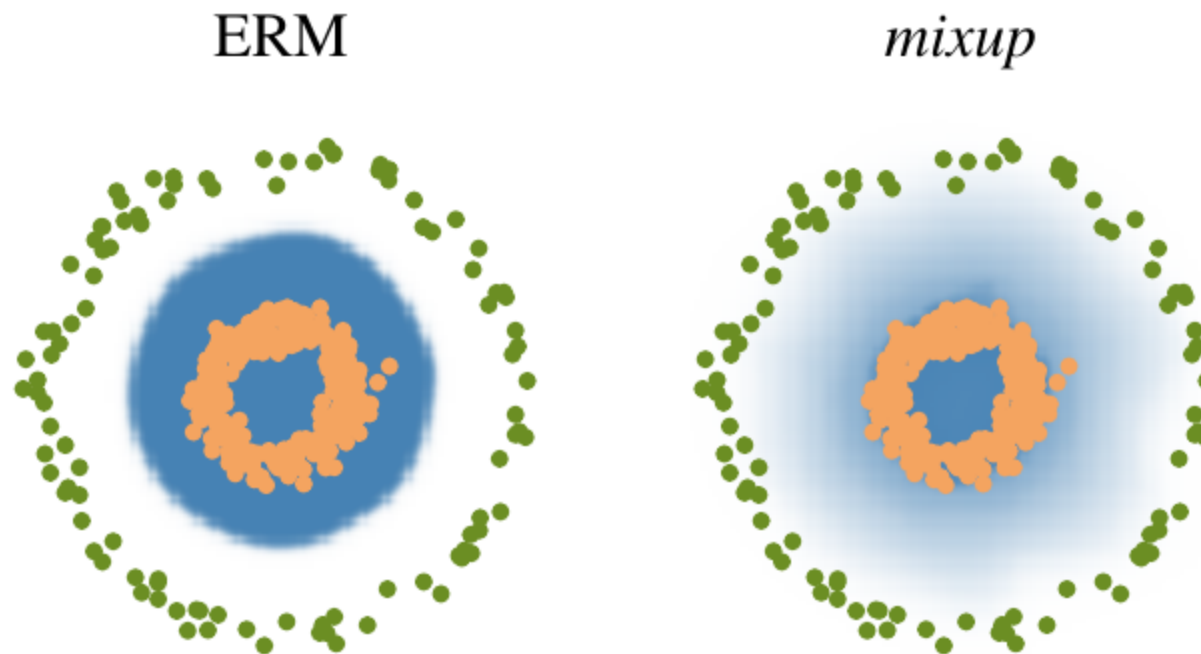
**Vikas Verma, Alex Lamb, Christopher Beckham,
Aaron Courville, Ioannis Mitliagkas, Yoshua
Bengio**

input mixup

$$\lambda \sim \text{Beta}(\alpha, \alpha)$$

$$x = \lambda x_i + (1 - \lambda)x_j$$

$$y = \lambda y_i + (1 - \lambda)y_j$$

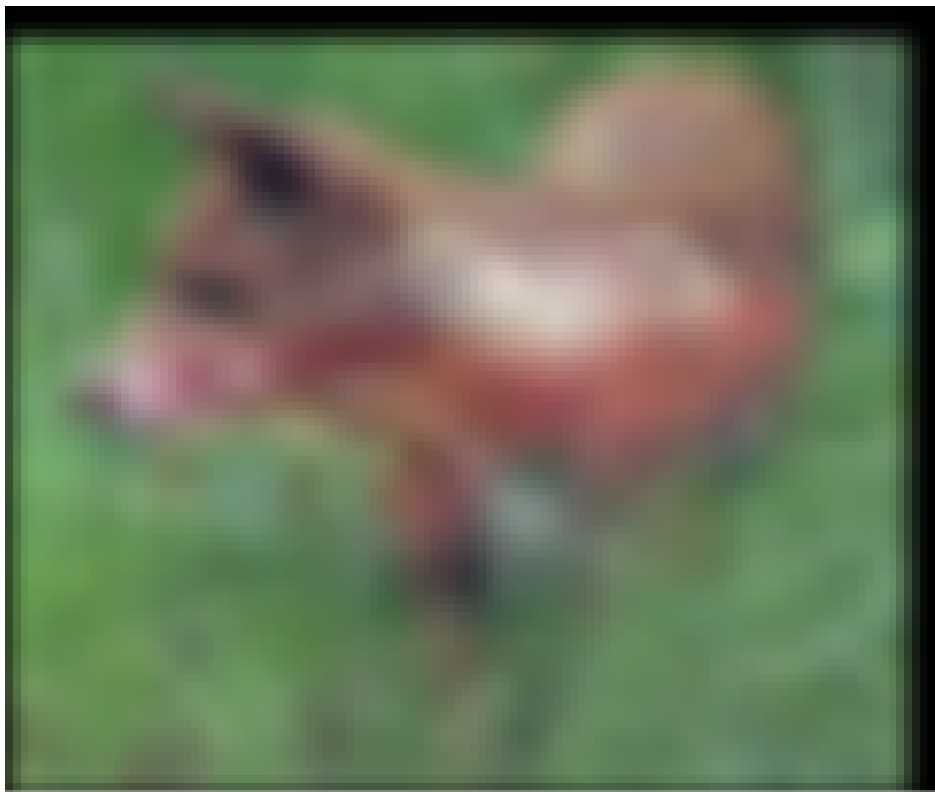


(b) Effect of *mixup* ($\alpha = 1$) on a toy problem. Green: Class 0. Orange: Class 1. Blue shading indicates $p(y = 1|x)$.

mixup: Beyond Empirical Risk Minimization

Huang Kaiuan <https://arxiv.org/abs/1710.09412>

60% dog + 40% car



Manifold mixup

$$\lambda \sim \text{Beta}(\alpha, \alpha)$$

$$f(x) = g_k(h_k(x))$$

$$loss = l(g_k(\lambda h_k(x_i) + (1 - \lambda)h_k(x_j)), \lambda y_i + (1 - \lambda)y_j)$$

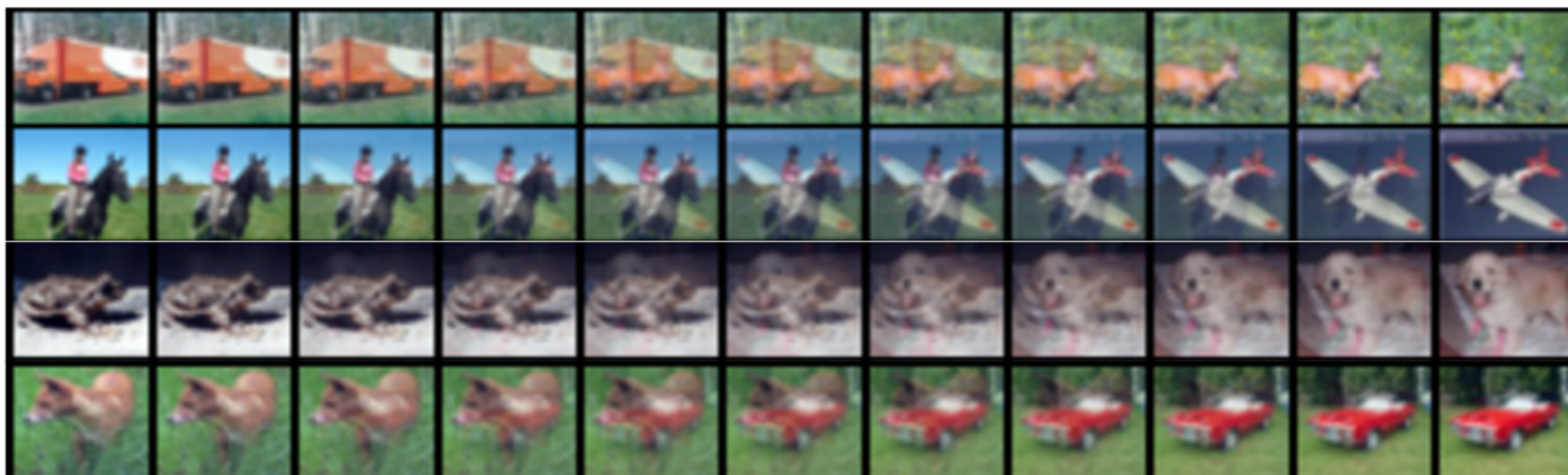


Figure 3: **Interpolations in the input space** with a mixing rate varied from 0.0 to 1.0.



Figure 4: **Interpolations in the hidden states** (using a small convolutional network trained to predict the input from the output of the second resblock). The interpolations in the hidden states show

Table 1: Supervised Classification Results on CIFAR-10. We note significant improvement with *Manifold Mixup* especially in terms of likelihood. Please refer to Appendix [A](#) for details on the implementation of *Manifold Mixup* and *Manifold Mixup* All layers.

Model	Test Acc	Test NLL
PreActResNet18		
No Mixup	94.88	0.2646
Input Mixup ($\alpha = 1.0$) (Zhang et al., 2017)	96.10	n/a
Input Mixup ($\alpha = 1.0$) (ours)	96.498	0.1945
<i>Manifold Mixup</i> ($\alpha = 2.0$)	97.104	0.1407
PreActResNet152		
No Mixup	95.797	0.1994
Input Mixup ($\alpha = 1.0$)	96.844	0.2312
<i>Manifold Mixup</i> ($\alpha = 2.0$)	97.238	0.1419
<i>Manifold Mixup</i> all layers ($\alpha = 6.0$)	97.622	0.0957

Table 3: Results on models trained on the normal CIFAR-100 and evaluated on a test set with novel deformations. The full version of this table and a more detailed description are in appendix table 7. *Manifold Mixup* (ours) consistently allows the model to be more robust to random shearing, rescaling, and rotation even though these deformations were not observed during training.

Test Set Deformation	No Mixup Baseline	Input Mixup $\alpha=1.0$	Input Mixup $\alpha=2.0$	<i>Manifold Mixup</i> $\alpha=2.0$
Rotation $U(-20^\circ, 20^\circ)$	52.96	55.55	56.48	60.08
Rotation $U(-60^\circ, 60^\circ)$	26.77	28.47	27.53	33.78
Shearing $U(-28.6^\circ, 28.6^\circ)$	55.92	58.16	60.01	62.85
Shearing $U(-57.3^\circ, 57.3^\circ)$	35.66	39.34	39.7	44.27
Zoom In (80% rescale)	47.95	52.18	50.47	52.7
Zoom Out (140% rescale)	19.34	41.81	42.02	45.29
Zoom Out (160% rescale)	11.12	25.48	25.85	27.02