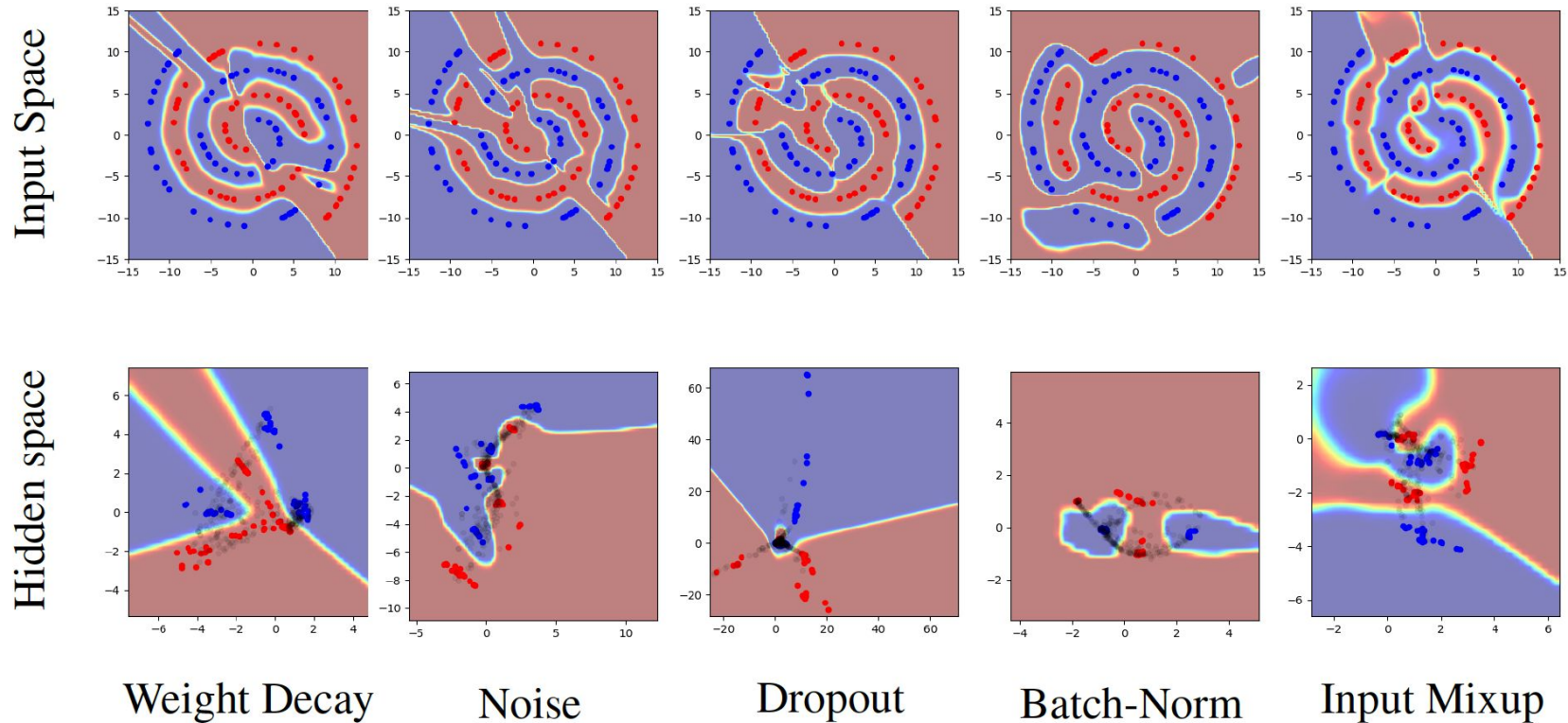


Manifold Mixup

—

Alex Lamb*, Vikas Verma*, Christopher Beckham, Amir Najafi, Ioannis Mitliagkas, David Lopez-Paz, Yoshua Bengio

Troubling Properties of Deep Networks



Issues with Current Methods

- Real data points occupy large volume in the space
- Decision boundary is close to the data
- Data points from off the manifold occupy region overlapping with real data points

Improving Representations with Manifold Mixup

- Simple Algorithm - just a few lines of code
- Great Results
- Surprising Properties - backed by rigorous theory

Manifold Mixup - Simple Algorithm

- On each update, select a random layer uniformly (including the input).
- Sample $\lambda \sim \text{Beta}(\alpha, \alpha)$
- Mix between two random examples from the minibatch at the selected layer with weights λ and $(1-\lambda)$.
- Mix the labels for those two examples in the same way to construct a soft target, yielding the *manifold mixup loss*, which compares the soft target with the output obtained with the mixed layer.

Manifold Mixup - Great Results

PreActResNet18	Test Error (%)	Test NLL
No Mixup	4.83 ± 0.066	0.190 ± 0.003
AdaMix‡	3.52	NA
Input Mixup†	4.20	NA
Input Mixup ($\alpha = 1$)	3.82 ± 0.048	0.186 ± 0.004
<i>Manifold Mixup</i> ($\alpha = 2$)	<u>2.95 ± 0.046</u>	<u>0.137 ± 0.003</u>

PreActResNet34	Test Error (%)	Test NLL
No Mixup	4.64 ± 0.072	0.200 ± 0.002
Input Mixup ($\alpha = 1$)	2.88 ± 0.043	0.176 ± 0.002
<i>Manifold Mixup</i> ($\alpha = 2$)	<u>2.54 ± 0.047</u>	<u>0.118 ± 0.002</u>

Wide-Resnet-28-10	Test Error (%)	Test NLL
No Mixup	3.99 ± 0.118	0.162 ± 0.004
Input Mixup ($\alpha = 1$)	2.92 ± 0.088	0.173 ± 0.001
<i>Manifold Mixup</i> ($\alpha = 2$)	<u>2.55 ± 0.024</u>	<u>0.111 ± 0.001</u>

(a) CIFAR-10

PreActResNet18	Test Error (%)	Test NLL
No Mixup	24.01 ± 0.376	1.189 ± 0.002
AdaMix‡	20.97	n/a
Input Mixup†	21.10	n/a
Input Mixup ($\alpha = 1$)	22.11 ± 0.424	1.055 ± 0.006
<i>Manifold Mixup</i> ($\alpha = 2$)	<u>20.34 ± 0.525</u>	<u>0.912 ± 0.002</u>

PreActResNet34	Test Error (%)	Test NLL
No Mixup	23.55 ± 0.399	1.189 ± 0.002
Input Mixup ($\alpha = 1$)	20.53 ± 0.330	1.039 ± 0.045
<i>Manifold Mixup</i> ($\alpha = 2$)	<u>18.35 ± 0.360</u>	<u>0.877 ± 0.053</u>

Wide-Resnet-28-10	Test Error (%)	Test NLL
No Mixup	21.72 ± 0.117	1.023 ± 0.004
Input Mixup ($\alpha = 1$)	18.89 ± 0.111	0.927 ± 0.031
<i>Manifold Mixup</i> ($\alpha = 2$)	<u>18.04 ± 0.171</u>	<u>0.809 ± 0.005</u>

(b) CIFAR-100

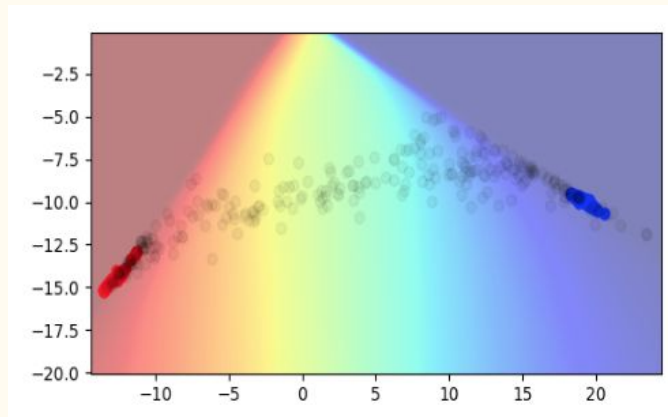
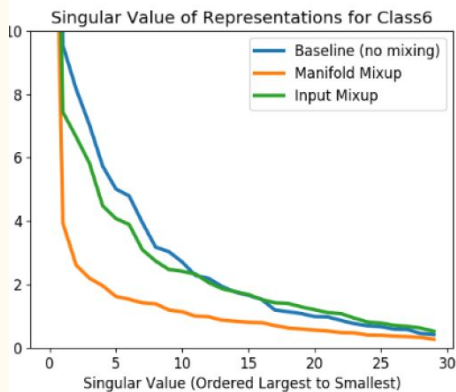
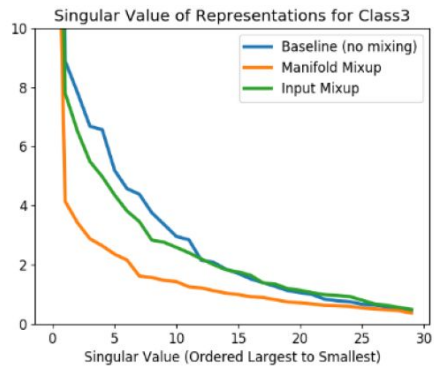
Massive gains
on likelihood

Also works on
SVHN,
Tiny-Imagenet,
Imagenet

Manifold Mixup - Great Results (external)

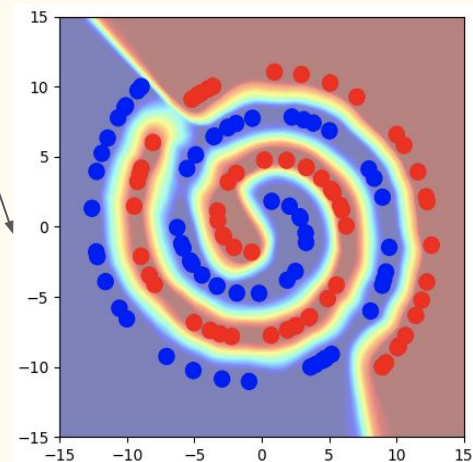
- Other labs have gotten great results with Manifold Mixup
- Handwriting Recognition (Moysset and Massina, ICDAR 2019)
- Convnets without Batch Normalization (Defazio & Bottou 2018)
- Prostate Cancer Segmentation with U-Net (Jung 2019)

Manifold Mixup - Surprising Properties



Hidden Space

Data Space



Manifold Mixup - Theory Justifying Properties

- When the manifold mixup loss is perfectly satisfied on a layer, the rest of the network becomes an implicit linear model, which we can call A.
- This can only be satisfied when $\dim(H) \geq d - 1$.
- The representations H have $\dim(H) - d + 1$ degrees of freedom.
- Implications: fitting the manifold mixup objective exactly is feasible in later layers, and concentrates the representations such that they have zero measure.

What can Manifold Mixup do for you (applied)?

- Massively improved likelihood, so any classification task where you use the probabilities will probably be helped.
- Tasks with small amounts of labeled data
- May also help with outliers / out-of-distribution, but needs to be studied more

What can you do for Manifold Mixup (theory)?

- Our theory makes very precise assumptions, can these be relaxed?
- Is there a way to generalize mixing to multiple layers or to RNNs (and understand it)?
- Lots of broader work on spectral properties of learned representations:
 - “An analytic theory of generalization dynamics and transfer learning in deep linear networks” (Lampinen and Ganguli 2019)
- Would be great to explicitly connect to Manifold Mixup!

Questions?

- Also if you have any questions, are curious about applying Manifold Mixup, or want to collaborate, reach out to:
 - vikasverma.iitm@gmail.com
 - lambalex@iro.umontreal.ca