

Introduction to mixed sample data augmentation

2021.11.26

발표자 : 정기원

dia517@korea.ac.kr

Data Mining & Quality Analytics Lab.

1. Introduction

2. Mix-up Algorithm

3. Advanced Mix-up Algorithms

4. Conclusion

❖ 발표자 소개



- 정기원
- 고려대학교 산업경영공학 전공
- Data Mining & Quality Analytics Lab (김성범 교수님 연구실)
- 석사 과정 재학 중 (2020.03 ~)

✓ 관심 연구 분야

- deep learning for signal processing and time series analysis
- advanced data augmentation

❖ 딥러닝 모델 과적합(Overfitting) 문제

- 데이터 수에 비해 많은 파라미터 수를 갖는 딥러닝 모델은 과적합 문제에 빠지기 쉬움
- 과적합된 모델의 경우, 학습에서 보지 못한 새로운 데이터에 대한 예측 성능이 상당히 낮음
- 이러한 문제를 해결하기 위해 **규제화(regularization)** / **데이터 증강(data augmentation)** 기법을 적용

regularization method

- ✓ weight decay
- ✓ dropout
- ✓ early stopping
- ✓ etc..

data augmentation method

- ✓ mix-up
- ✓ autoaugment
- ✓ flipping & cropping
- ✓ etc..

Mix-up Algorithm

❖ Mix-up 알고리즘

- 2021.11.23 기준 설명 2702회 인용
- 다양한 연구 분야에서 활발하게 사용되고 있는 데이터 증강기법

mixup: BEYOND EMPIRICAL RISK MINIMIZATION

Hongyi Zhang
MIT

Moustapha Cisse, Yann N. Dauphin, David Lopez-Paz*
FAIR

ABSTRACT

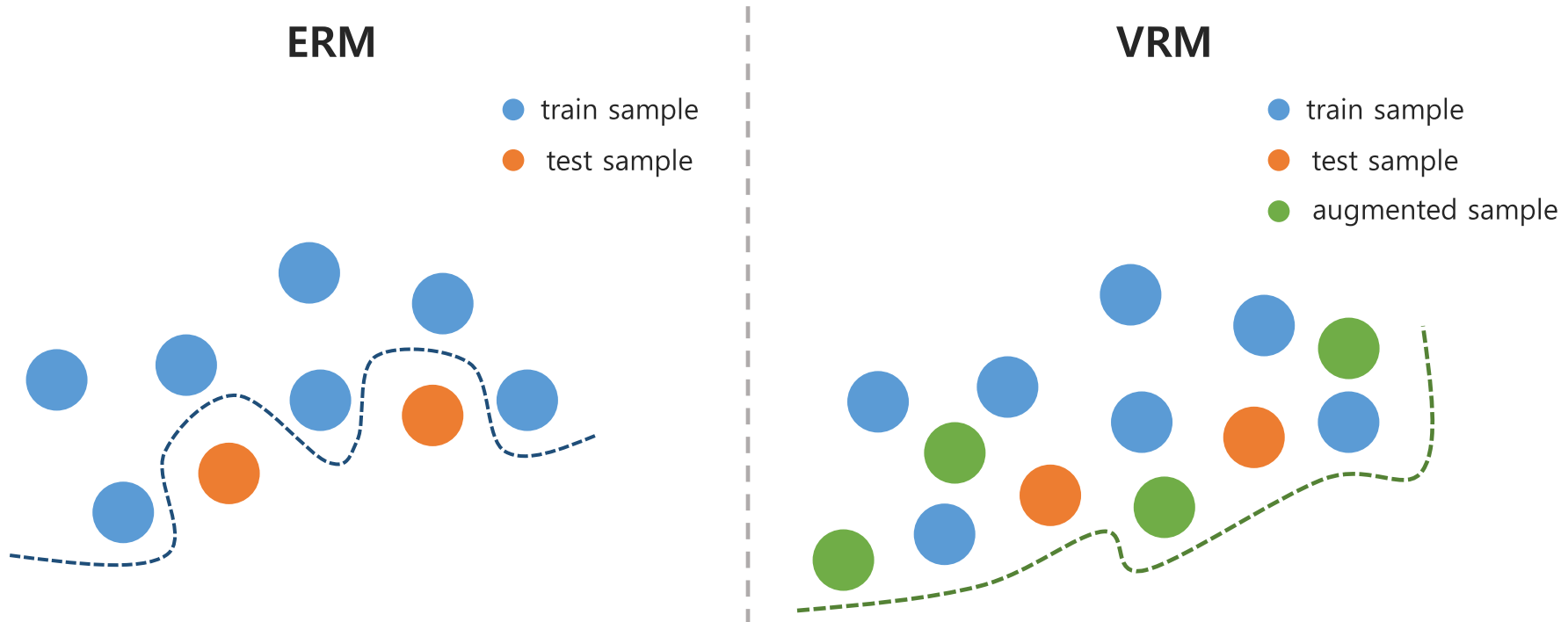
Large deep neural networks are powerful, but exhibit undesirable behaviors such as memorization and sensitivity to adversarial examples. In this work, we propose *mixup*, a simple learning principle to alleviate these issues. In essence, *mixup* trains a neural network on convex combinations of pairs of examples and their labels. By doing so, *mixup* regularizes the neural network to favor simple linear behavior in-between training examples. Our experiments on the ImageNet-2012, CIFAR-10, CIFAR-100, Google commands and UCI datasets show that *mixup* improves the generalization of state-of-the-art neural network architectures. We also find that *mixup* reduces the memorization of corrupt labels, increases the robustness to adversarial examples, and stabilizes the training of generative adversarial networks.

Zhang, H., Cisse, M., Dauphin, Y. N., & Lopez-Paz, D. (2017). mixup: Beyond empirical risk minimization. arXiv preprint arXiv:1710.09412.

Mix-up Algorithm

❖ ERM(Empirical Risk Minimization) 과 VRM(Vicinal Risk Minimization)

- 신경망 모델은 주어진 훈련 데이터셋에 대한 에러를 최소화하는 방향으로 학습을 진행
- ERM 기반 학습은 훈련 데이터에 과적합 되는 위험이 존재
- VRM 기반 학습은 주어진 훈련 데이터의 근방 분포를 활용하여 신경망 학습의 도움을 주고자 함



Mix-up Algorithm

❖ Mix-up 알고리즘 동작 원리

- 두 데이터 샘플로부터 선형 보간법을 통해 새로운 샘플을 생성하는 데이터 증강 기법
- 간단한 동작 원리에도 좋은 일반화 성능을 보장하여 다양한 딥러닝 연구 분야에서 사용되고 있는 기법

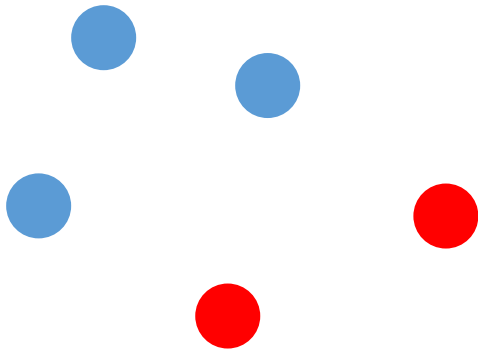
Mix-up formulation

$$\begin{aligned}\tilde{x} &= \lambda x_i + (1 - \lambda)x_j, \\ \tilde{y} &= \lambda y_i + (1 - \lambda)y_j,\end{aligned}$$

where x_i, x_j are raw input vectors

where y_i, y_j are one-hot label encodings

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



Mix-up Algorithm

❖ Mix-up 알고리즘 동작 원리

- 두 데이터 샘플로부터 선형 보간법을 통해 새로운 샘플을 생성하는 데이터 증강 기법
- 간단한 동작 원리에도 좋은 일반화 성능을 보장하여 다양한 딥러닝 연구 분야에서 사용되고 있는 기법

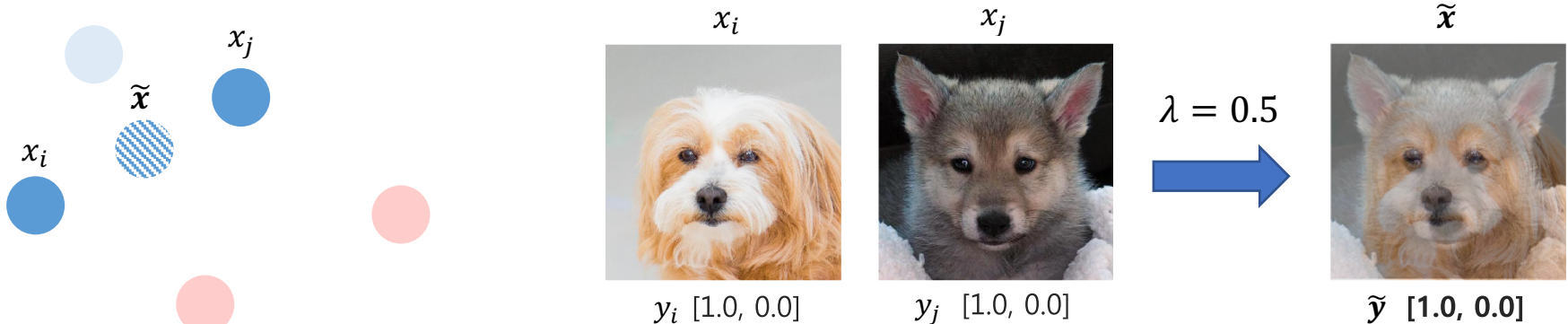
Mix-up formulation

$$\tilde{x} = \lambda x_i + (1 - \lambda)x_j,$$
$$\tilde{y} = \lambda y_i + (1 - \lambda)y_j,$$

where x_i, x_j are raw input vectors

where y_i, y_j are one-hot label encodings

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



Mix-up Algorithm

❖ Mix-up 알고리즘 동작 원리

- 두 데이터 샘플로부터 선형 보간법을 통해 새로운 샘플을 생성하는 데이터 증강 기법
- 간단한 동작 원리에도 좋은 일반화 성능을 보장하여 다양한 딥러닝 연구 분야에서 사용되고 있는 기법

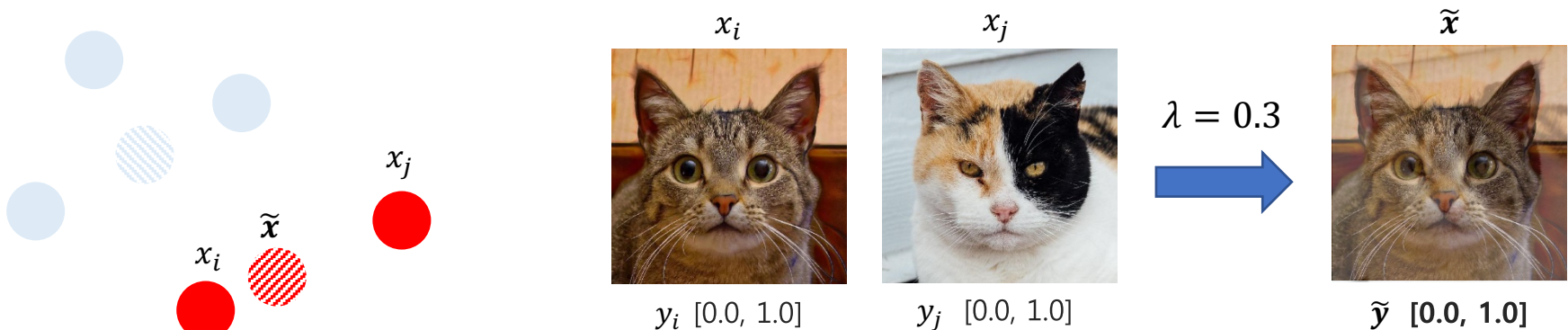
Mix-up formulation

$$\tilde{x} = \lambda x_i + (1 - \lambda)x_j,$$
$$\tilde{y} = \lambda y_i + (1 - \lambda)y_j,$$

where x_i, x_j are raw input vectors

where y_i, y_j are one-hot label encodings

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



Mix-up Algorithm

❖ Mix-up 알고리즘 동작 원리

- 두 데이터 샘플로부터 선형 보간법을 통해 새로운 샘플을 생성하는 데이터 증강 기법
- 간단한 동작 원리에도 좋은 일반화 성능을 보장하여 다양한 딥러닝 연구 분야에서 사용되고 있는 기법

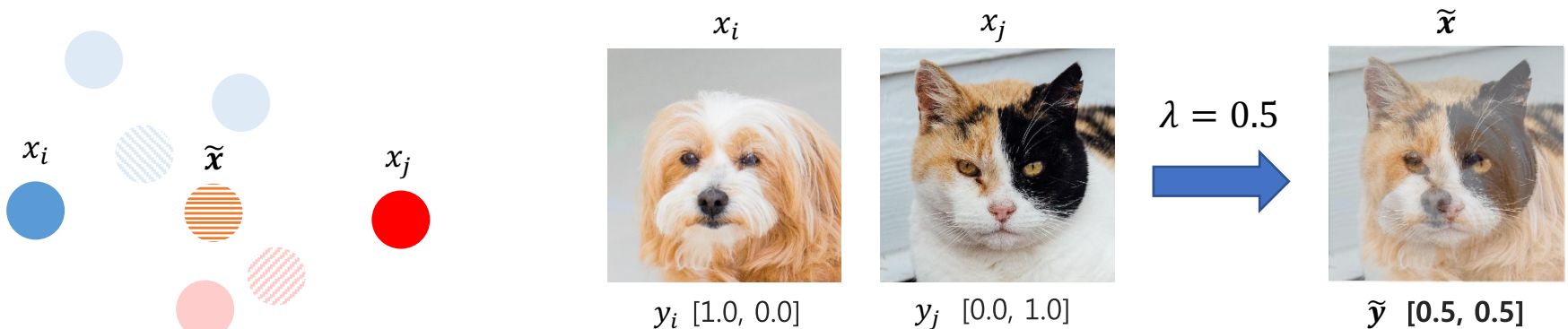
Mix-up formulation

$$\tilde{x} = \lambda x_i + (1 - \lambda)x_j,$$
$$\tilde{y} = \lambda y_i + (1 - \lambda)y_j,$$

where x_i, x_j are raw input vectors

where y_i, y_j are one-hot label encodings

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

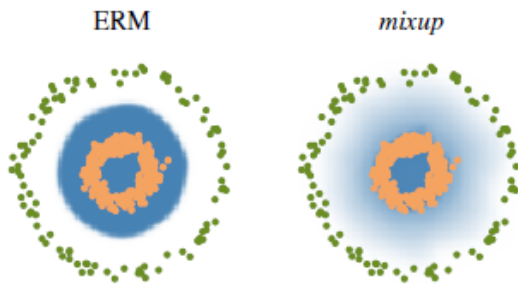


Mix-up Algorithm

❖ Mix-up 알고리즘 실험 결과

- ERM 학습 모델과 비교해 mixup을 적용한 모델의 경우, 부드러운 결정 경계선(Decision Boundary)를 생성
- CIFAR 이미지 데이터셋에 대해서 mixup을 적용했을 때, ERM 학습 모델보다 좋은 예측 성능을 보임

Decision Boundary Visualization

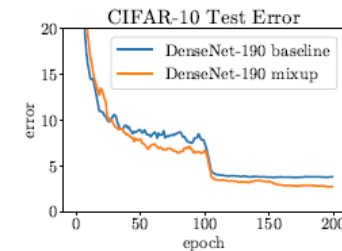


(b) Effect of *mixup* ($\alpha = 1$) on a toy problem. Green: Class 0. Orange: Class 1. Blue shading indicates $p(y=1|x)$.

Errors in CIFAR Datasets

Dataset	Model	ERM	<i>mixup</i>
CIFAR-10	PreAct ResNet-18	5.6	4.2
	WideResNet-28-10	3.8	2.7
	DenseNet-BC-190	3.7	2.7
CIFAR-100	PreAct ResNet-18	25.6	21.1
	WideResNet-28-10	19.4	17.5
	DenseNet-BC-190	19.0	16.8

(a) Test errors for the CIFAR experiments.



(b) Test error evolution for the best ERM and *mixup* models.

Mix-up Algorithm

❖ Mix-up 알고리즘 실험 결과

- 이미지 데이터가 아닌 Speech / Tabular 데이터에 mixup을 적용해도 예측 성능 향상을 보여줌
- mixup을 통해 불안정한 학습을 하는 GAN(Generative Adversarial Networks)이 안정적으로 학습하도록 도와줌

Errors in Speech/Tabular Dataset

Model	Method	Validation set	Test set
LeNet	ERM	9.8	10.3
	<i>mixup</i> ($\alpha = 0.1$)	10.1	10.8
	<i>mixup</i> ($\alpha = 0.2$)	10.2	11.3
VGG-11	ERM	5.0	4.6
	<i>mixup</i> ($\alpha = 0.1$)	4.0	3.8
	<i>mixup</i> ($\alpha = 0.2$)	3.9	3.4

Figure 4: Classification errors of ERM and *mixup* on the Google commands dataset.

Dataset	ERM	<i>mixup</i>
Abalone	74.0	73.6
Arcene	57.6	48.0
Arrhythmia	56.6	46.3

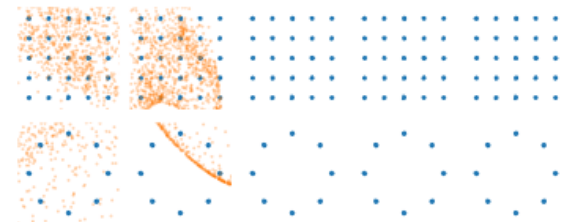
Dataset	ERM	<i>mixup</i>
Htru2	2.0	2.0
Iris	21.3	17.3
Phishing	16.3	15.2

Table 4: ERM and *mixup* classification errors on the UCI datasets.

Effect of mixup on GAN training

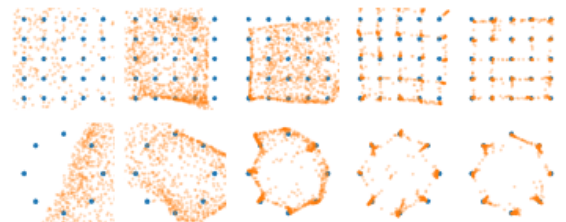
$$\max_g \min_d \mathbb{E}_{x,z} \ell(d(x), 1) + \ell(d(g(z)), 0)$$

ERM GAN



$$\max_g \min_d \mathbb{E}_{x,z,\lambda} \ell(d(\lambda x + (1 - \lambda)g(z)), \lambda)$$

mixup GAN ($\alpha = 0.2$)



Mix-up Algorithm

❖ 적대적 예제(Adversarial Examples)

- 다양한 연구에서 신경망 모델이 적대적 예제에 대해서 취약하다는 문제를 제기하고 있음
- 적대적 예제는 원본 이미지에 **아주 작은 변화(perturbation)**만 더해도 모델 성능이 크게 달라질 수 있는 이미지
 - 대표적인 적대적 예제 생성 방법이 **FGSM(Fast Gradient Sign Attack)** 알고리즘
 - FGSM은 원본 이미지에 대한 그레디언트를 계산하여 손실을 최대화하는 방향으로 이미지를 생성

Original Example



57.7% confidence

"Panda"

+ .007 ×



=

Adversarial Example



99.3% confidence

"gibbon"

Goodfellow, I. J., Shlens, J., & Szegedy, C. (2014). Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572*.

Mix-up Algorithm

❖ 적대적 예제(Adversarial Examples)

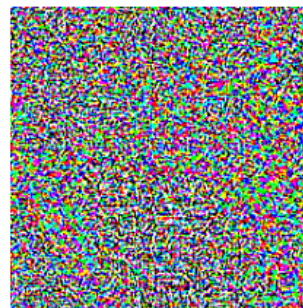
- 다양한 연구에서 신경망 모델이 적대적 예제에 대해서 취약하다는 문제를 제기하고 있음
- 적대적 예제는 원본 이미지에 **아주 작은 변화(perturbation)**만 더해도 모델 성능이 크게 달라질 수 있는 이미지
 - 대표적인 적대적 예제 생성 방법이 **FGSM(Fast Gradient Sign Attack)** 알고리즘
 - FGSM은 **원본 이미지**에 대한 **그레디언트를 계산하여 손실을 최대화하는 방향으로 이미지를 생성**

Original Example



x

ϵ
 $+ .007 \times$



$\text{sign}(\nabla_x J(\theta, x, y))$

=

Adversarial Example



$x + \epsilon \text{sign}(\nabla_x J(\theta, x, y))$

Goodfellow, I. J., Shlens, J., & Szegedy, C. (2014). Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572*.

Mix-up Algorithm

❖ 적대적 공격(Adversarial attack)에 강건한 Mix-up 알고리즘

- Adversarial example을 만들어내는 것을 적대적 공격(Adversarial attack)이라고 함
- 적대적 공격에는 white box 방식과 black box 방식이 있음
 - white box : 모델의 파라미터 및 그레디언트 정보를 **알고 있는 상황**에서 공격
 - black box : 모델의 파라미터 및 그레디언트 정보를 **알지 못하는 상황**에서 공격
- 믹스업을 통해 학습한 모델이 ERM 학습 모델에 비해 적대적 공격에 강건한 성능을 보여줌

Metric	Method	FGSM	I-FGSM
Top-1	ERM	90.7	99.9
	<i>mixup</i>	75.2	99.6
Top-5	ERM	63.1	93.4
	<i>mixup</i>	49.1	95.8

(a) White box attacks.

Metric	Method	FGSM	I-FGSM
Top-1	ERM	57.0	57.3
	<i>mixup</i>	46.0	40.9
Top-5	ERM	24.8	18.1
	<i>mixup</i>	17.4	11.8

(b) Black box attacks.

Table 3: Classification errors of ERM and *mixup* models when tested on adversarial examples.

Mix-up Algorithm

❖ Mix-up 알고리즘에 대한 추가적인 실험

- 2021.11.23 기준 설명 157회 인용
- mixup 방법론이 calibration과 over-confident에 대해 강건한 성능을 보장함을 실험을 통해 증명

On Mixup Training: Improved Calibration and Predictive Uncertainty for Deep Neural Networks

Sunil Thulasidasan^{*,1,2}, Gopinath Chennupati¹, Jeff Bilmes²,
Tanmoy Bhattacharya¹, Sarah Michalak¹

¹Los Alamos National Laboratory

²Department of Electrical and Computer Engineering, University of Washington

Abstract

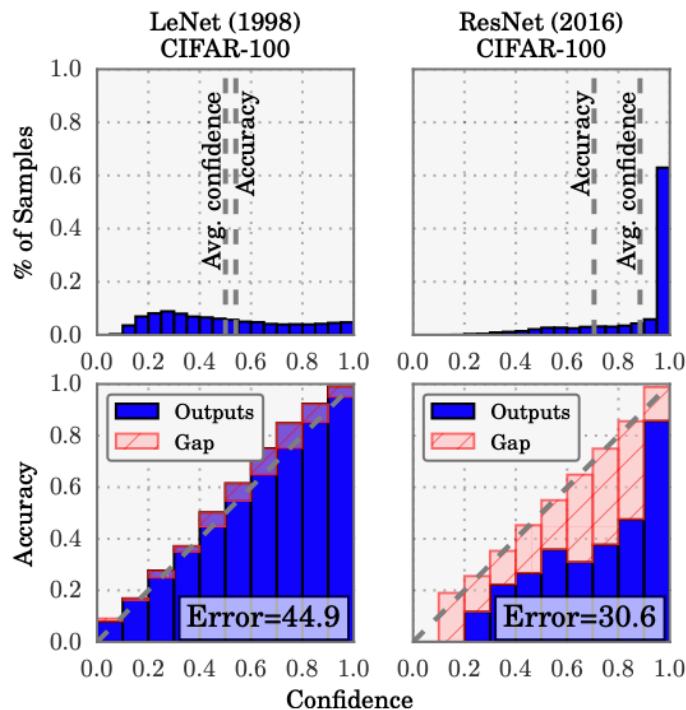
Mixup [40] is a recently proposed method for training deep neural networks where additional samples are generated during training by convexly combining random pairs of images and their associated labels. While simple to implement, it has been shown to be a surprisingly effective method of data augmentation for image classification: DNNs trained with mixup show noticeable gains in classification performance on a number of image classification benchmarks. In this work, we discuss a hitherto untouched aspect of mixup training – the calibration and predictive uncertainty of models trained with mixup. We find that DNNs trained with mixup are significantly better calibrated – i.e., the predicted softmax scores are much better indicators of the actual likelihood of a correct prediction – than DNNs trained in the regular fashion. We conduct experiments on a number of image classification architectures and datasets – including large-scale datasets like ImageNet – and find this to be the case. Additionally, we find that merely mixing features does not result in the same calibration benefit and that the label smoothing in mixup training plays a significant role in improving calibration. Finally, we also observe that mixup-trained DNNs are less prone to over-confident predictions on out-of-distribution and random-noise data. We conclude that the typical overconfidence seen in neural networks, even on in-distribution data is likely a consequence of training with hard labels, suggesting that mixup be employed for classification tasks where predictive uncertainty is a significant concern.

Thulasidasan, S., Chennupati, G., Bilmes, J., Bhattacharya, T., & Michalak, S. (2019). On mixup training: Improved calibration and predictive uncertainty for deep neural networks. arXiv preprint arXiv:1905.11001.

Mix-up Algorithm

❖ 모델 교정(Model Calibration) 필요성

- Calibration은 모델이 예측한 확률(Confidence score)과 실제 예측 정확도(accuracy)가 일치하도록 조정하는 것
- 최근 딥러닝 모델이 예측 성능은 향상되고 있으나, 자신이 예측한 결과를 과잉 확신하는 경향이 있음
 - 실제로 정답을 맞출 확률에 비해 모델이 출력한 예측 확률이 너무 높다는 문제가 있음



Guo C, Peiss G, Sun Y, & Weinberger K Q (2017 July). On calibration of modern neural networks. In International Conference on Machine Learning (pp. 1321-1330). PMLR.

Mix-up Algorithm

❖ Calibration 평가지표

- 논문에서는 3가지 평가지표를 사용함
 - **Reliability Diagrams** : 모델의 accuracy와 confidence 관계를 시각적으로 확인할 수 있는 그래프
 - **ECE(Expected Calibration Error)** : accuracy와 confidence 차이를 가중 평균한 값
 - **OE(Overconfidence Error)** : overconfidence한 경우에 대해서 accuracy와 confidence의 차이를 수치화한 값

Reliability Diagrams

$$acc(B_m) = \frac{1}{|B_m|} \sum_{i \in B_m} 1(\hat{y}_i = y_i)$$

$$conf(B_m) = \frac{1}{|B_m|} \sum_{i \in B_m} \hat{p}_i$$

ECE

$$\sum_{m=1}^M \frac{|B_m|}{n} |acc(B_m) - conf(B_m)|$$

OE

$$\sum_{m=1}^M \frac{|B_m|}{n} [conf(B_m) * \max(conf(B_m) - acc(B_m), 0)]$$

B_m : 모델의 output probability를 m개의 구간으로 나누었을 때, 해당 구간의 샘플 수

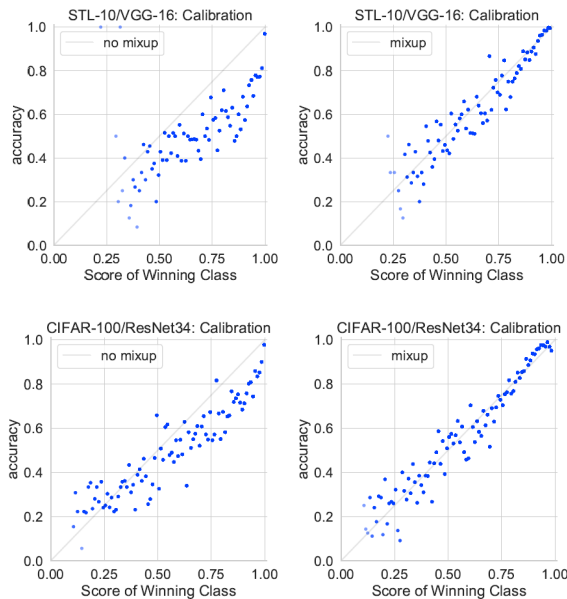
p_i : 샘플 i 의 confidence(probability)

Mix-up Algorithm

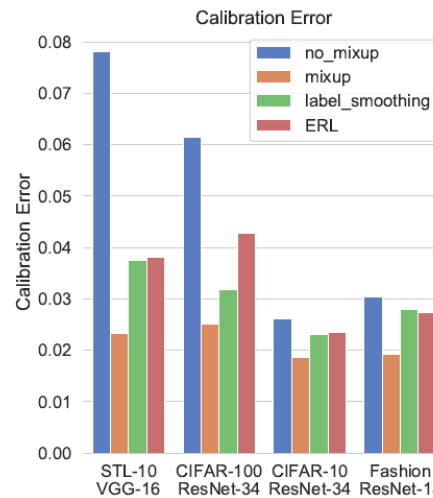
❖ mixup 알고리즘의 Calibration 효과

- label smoothing, ERL(entropy-regularized loss) 방법론과 각 평가지표에 대해 성능을 비교
- natural language data(STL-10), image data(CIFAR-100)에 대해서 mixup 방법론의 우수한 성능을 입증

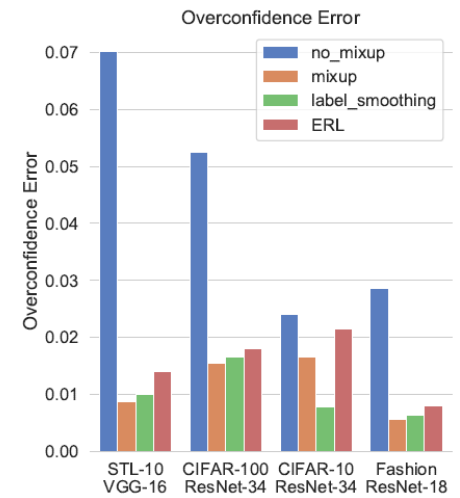
Reliability Diagrams



ECE



OE



Mix-up Algorithm

❖ mixup 알고리즘의 Calibration 효과

- label smoothing, ERL(entropy-regularized loss) 방법론과 각 평가지표에 대해 성능을 비교

왜 Mix-up 알고리즘이 모델의 일반화 성능에 도움이 될까?

➡ 학습하는 과정에서 **매번 새로운 샘플을 학습하기 때문에,**
모델은 거의 **무한한 수의 데이터를 학습하는 것과 같다.**

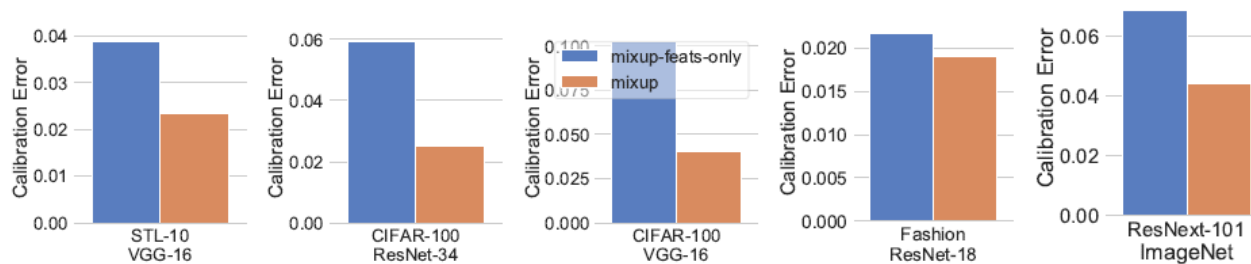
➡ 그렇다면 레이블(Y) 정보는 섞지 않고, 입력 데이터(X)만
섞어도 동일한 효과를 얻을 수 있지 않을까?

Mix-up Algorithm

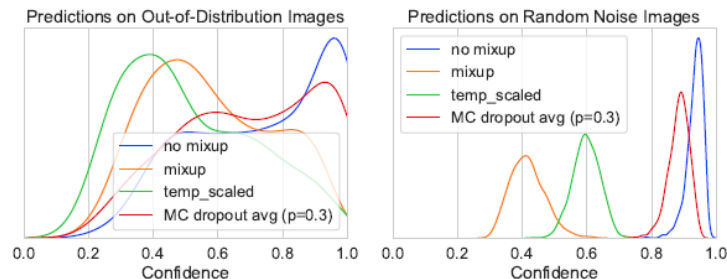
❖ 그 외 다양한 실험 결과

- 입력 데이터(X)만 섞는 것보다 레이블 정보(Y)도 함께 섞는 것이 calibration 효과가 우수함
- 믹스업을 통해 학습한 모델이 OOD(out-of-distribution), Random noise data에 강건함을 실험을 통해 확인

calibration performance- only features mixed vs full mixed



distribution of confidence on OOD, noise samples



out-of-category detection results

Method	AUROC(In/Out)	
	STL-10/ ImageNet	STL-10/ Gaussian
Baseline	80.57	73.28
Mixup ($\alpha=0.4$)	83.28	95.93
Temp. Scaling	56.2	54.2
Dropout($p=0.3$)	78.93	70.57

Mix-up Algorithm

❖ 그 외 Mixup 효과를 입증하는 논문

- 2021.11.19 기준 설명 33회 인용
- mixup 알고리즘의 효과를 수식적으로 증명한 논문
 - 적대적 손실(adversarial loss)의 upper bound를 낮출 수 있음을 증명

HOW DOES MIXUP HELP WITH ROBUSTNESS AND GENERALIZATION?

Linjun Zhang*
Rutgers University
linjun.zhang@rutgers.edu

Zhun Deng*
Harvard University
zhundeng@g.harvard.edu

Kenji Kawaguchi*
Harvard University
kkawaguchi@fas.harvard.edu

Amirata Ghorbani
Stanford University
amiratag@stanford.edu

James Zou
Stanford University
jamesz@stanford.edu

ABSTRACT

Mixup is a popular data augmentation technique based on taking convex combinations of pairs of examples and their labels. This simple technique has been shown to substantially improve both the robustness and the generalization of the trained model. However, it is not well-understood why such improvement occurs. In this paper, we provide theoretical analysis to demonstrate how using Mixup in training helps model robustness and generalization. For robustness, we show that minimizing the Mixup loss corresponds to approximately minimizing an upper bound of the adversarial loss. This explains why models obtained by Mixup training exhibits robustness to several kinds of adversarial attacks such as Fast Gradient Sign Method (FGSM). For generalization, we prove that Mixup augmentation corresponds to a specific type of data-adaptive regularization which reduces overfitting. Our analysis provides new insights and a framework to understand Mixup.

Zhang, L, Deng, Z, Kawaguchi, K, Ghorbani, A, & Zou, J. (2020). How Does Mixup Help With Robustness and Generalization?. arXiv preprint arXiv:2010.04819.

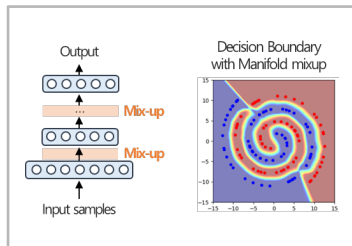
Advanced Mix-up Algorithm

❖ 다양한 Mix-up 알고리즘 고도화 방법론

- Mixup 알고리즘 수식을 개선한 다양한 방법론들이 제안되었음
- 다양한 데이터셋과 연구 분야에 mixup 알고리즘을 적용하여 각 방법론들의 우수성을 입증

모델링 고도화

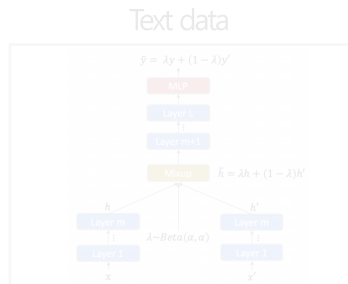
manifold mixup



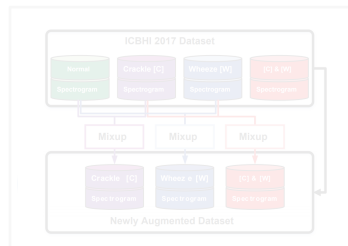
cutmix



적용 데이터셋 확장

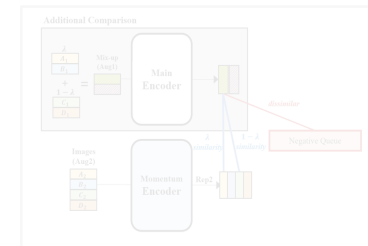


sound data

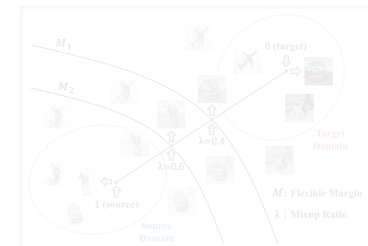


다른 연구 분야에 적용

contrastive learning



domain adaptation



Advanced Mix-up Algorithm

❖ Manifold Mixup 알고리즘

- 원본 데이터 공간이 아닌 매니폴드 공간 상에서 두 샘플을 섞는 것을 제안함
- 신경망 모델 중 무작위로 특정 레이어를 선택 후 해당 레이어로부터 출력된 히든 벡터를 섞는 방식으로 동작

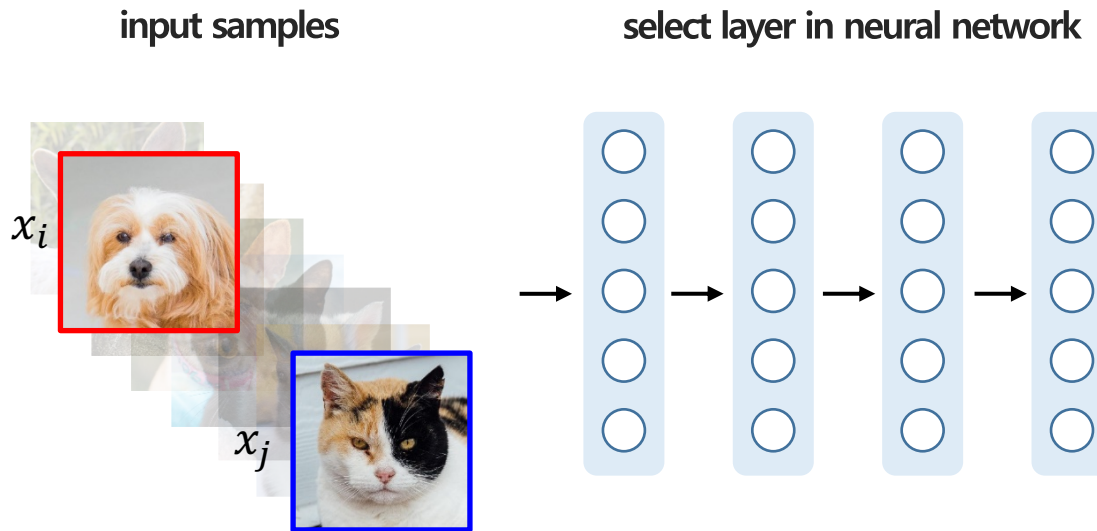
input samples



Advanced Mix-up Algorithm

❖ Manifold Mixup 알고리즘

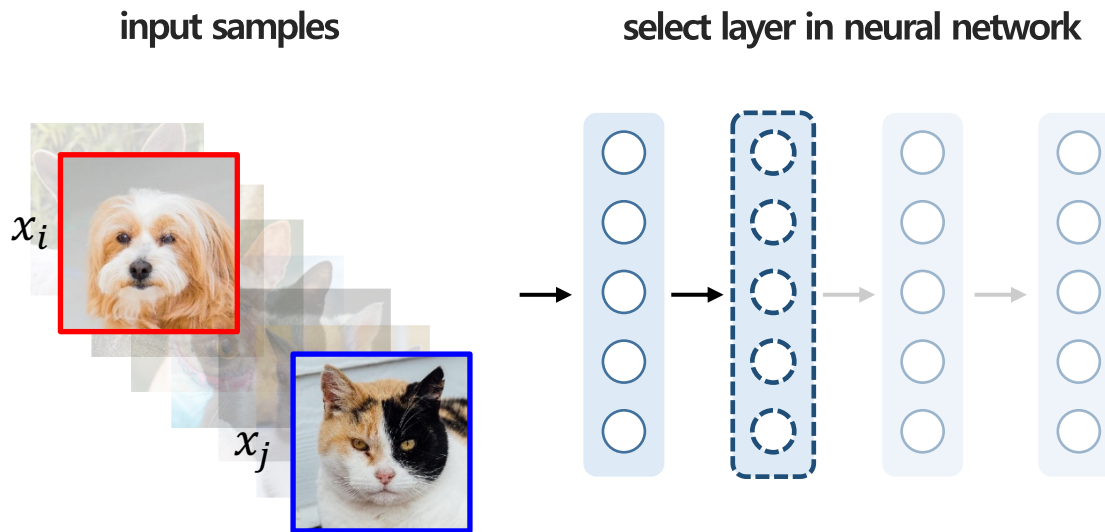
- 원본 데이터 공간이 아닌 매니폴드 공간 상에서 두 샘플을 섞는 것을 제안함
- 신경망 모델 중 무작위로 특정 레이어를 선택 후 해당 레이어로부터 출력된 히든 벡터를 섞는 방식으로 동작



Advanced Mix-up Algorithm

❖ Manifold Mixup 알고리즘

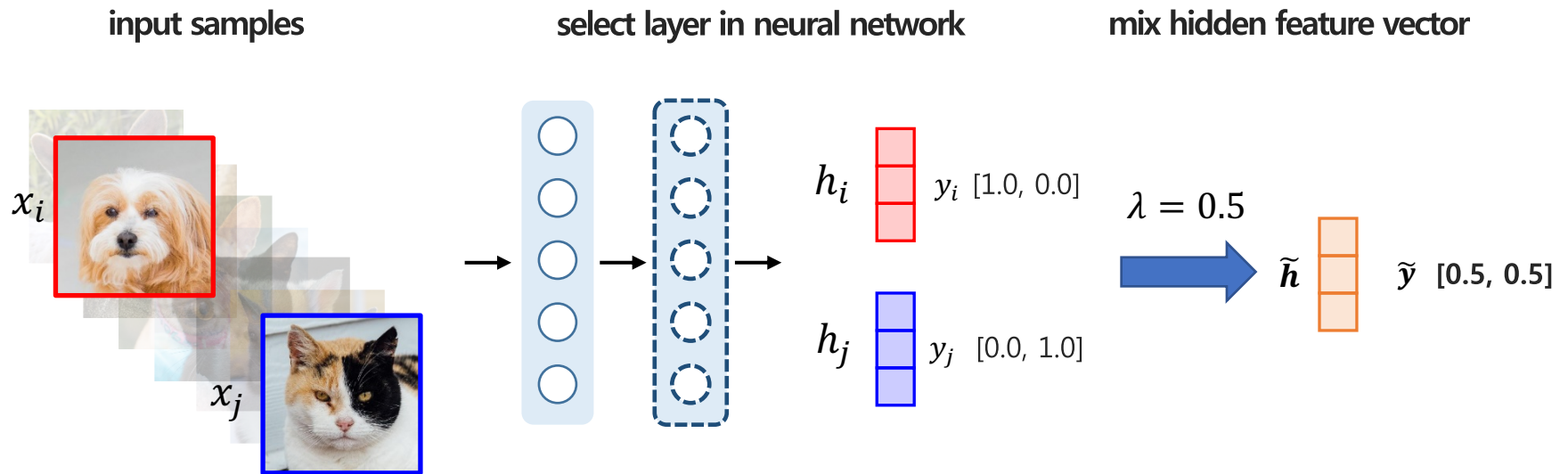
- 원본 데이터 공간이 아닌 매니폴드 공간 상에서 두 샘플을 섞는 것을 제안함
- 신경망 모델 중 무작위로 특정 레이어를 선택 후 해당 레이어로부터 출력된 히든 벡터를 섞는 방식으로 동작



Advanced Mix-up Algorithm

❖ Manifold Mixup 알고리즘

- 원본 데이터 공간이 아닌 매니폴드 공간 상에서 두 샘플을 섞는 것을 제안함
- 신경망 모델 중 무작위로 특정 레이어를 선택 후 해당 레이어로부터 출력된 히든 벡터를 섞는 방식으로 동작



Advanced Mix-up Algorithm

❖ Manifold Mixup 알고리즘 실험

- 매니폴드 믹스업을 통해 신경망 모델의 hidden representation과 decision boundaries에 도움을 준다고 주장
- 다양한 실험을 통해 제안 방법론의 기여점을 입증

Manifold Mixup improves the hidden representations and decision boundaries of neural networks at multiple layers.

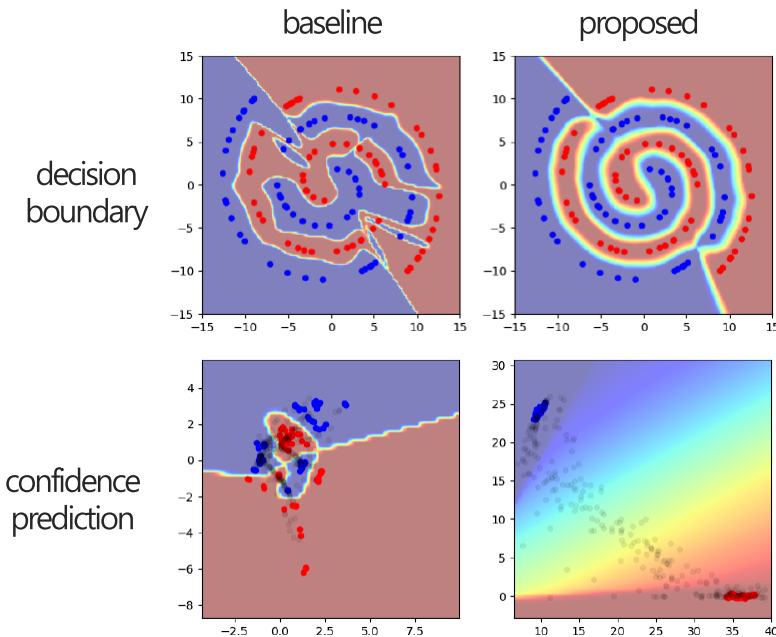
More specifically, *Manifold Mixup* improves generalization in deep neural networks because it:

- Leads to smoother decision boundaries that are further away from the training data, at multiple levels of representation. Smoothness and margin are well-established factors of generalization (Bartlett & Shawe-taylor, 1998; Lee et al., 1995).
- Leverages interpolations in deeper hidden layers, which capture higher level information (Zeiler & Fergus, 2013) to provide additional training signal.
- Flattens the class-representations, reducing their number of directions with significant variance (Section 3). This can be seen as a form of compression, which is linked to generalization by a well-established theory (Tishby & Zaslavsky, 2015; Shwartz-Ziv & Tishby, 2017) and extensive experimentation (Alemi et al., 2017; Belghazi et al., 2018; Goyal et al., 2018; Achille & Soatto, 2018).

Advanced Mix-up Algorithm

❖ Manifold Mixup 알고리즘 실험

- 매니폴드 믹스업을 통해 신경망 decision boundaries가 smooth하게 생성됨을 시각적으로 확인
- 신경망 모델이 Over-confident하는 문제를 완화할 수 있음을 시각적으로 확인
- 기존 Mixup 보다 좋은 예측 성능과 적대적 샘플에 강건함을 보임



classification error

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
Manifold Mixup ($\alpha = 2$)	2.95 ± 0.046	0.137 ± 0.003

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
Manifold Mixup ($\alpha = 2$)	2.54 ± 0.047	0.118 ± 0.002

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
Manifold Mixup ($\alpha = 2$)	2.55 ± 0.024	0.111 ± 0.001

(a) CIFAR-10

robustness to adversarial examples

CIFAR-10	FGSM
No Mixup	36.32
Input Mixup ($\alpha = 1$)	71.51
Manifold Mixup ($\alpha = 2$)	77.50
PGD training (7-steps)†	56.10

CIFAR-100	FGSM
Input Mixup ($\alpha = 1$)	40.7
Manifold Mixup ($\alpha = 2$)	44.96

SVHN	FGSM
No Mixup	21.49
Input Mixup ($\alpha = 1$)	56.98
Manifold Mixup ($\alpha = 2$)	65.91
PGD training (7-steps)†	72.80

Advanced Mix-up Algorithm

❖ CutMix 알고리즘

- 특정 샘플의 일부 영역을 다른 샘플의 일부 영역으로 대체하여 새로운 샘플을 생성
 - 전체 영역 중 선택된 일부 영역의 비율이 λ 값으로 사용
- CutMix는 Classification, Localization, Detection task에서 좋은 성능을 보여준다고 제안

x_i



y_i [1.0, 0.0]

x_j



y_j [0.0, 1.0]

Advanced Mix-up Algorithm

❖ CutMix 알고리즘

- 특정 샘플의 일부 영역을 다른 샘플의 일부 영역으로 대체하여 새로운 샘플을 생성
 - 전체 영역 중 선택된 일부 영역의 비율이 λ 값으로 사용
- CutMix는 Classification, Localization, Detection task에서 좋은 성능을 보여준다고 제안

x_i

1	1	1	0
1	1	1	0
1	1	1	0
0	0	0	0

M

y_i [1.0, 0.0]



y_j [0.0, 1.0]

Advanced Mix-up Algorithm

❖ CutMix 알고리즘

- 특정 샘플의 일부 영역을 다른 샘플의 일부 영역으로 대체하여 새로운 샘플을 생성
 - 전체 영역 중 선택된 일부 영역의 비율이 λ 값으로 사용
- CutMix는 Classification, Localization, Detection task에서 좋은 성능을 보여준다고 제안

x_i				x_j			
1	1	1	0	0	0	0	1
1	1	1	0	0	0	0	1
1	1	1	0	0	0	0	1
0	0	0	0	1	1	1	1
M				$(1 - M)$			
$y_i [1.0, 0.0]$				$y_j [0.0, 1.0]$			

Advanced Mix-up Algorithm

❖ CutMix 알고리즘

- 특정 샘플의 일부 영역을 다른 샘플의 일부 영역으로 대체하여 새로운 샘플을 생성
 - 전체 영역 중 선택된 일부 영역의 비율이 λ 값으로 사용
- CutMix는 Classification, Localization, Detection task에서 좋은 성능을 보여준다고 제안

$$\tilde{x} = M \odot x_i + (1 - M) \odot x_j$$

$$\tilde{y} = \lambda y_i + (1 - \lambda) y_j, \quad \lambda = \frac{\text{선택된 일부 영역}}{\text{전체 영역}} = \frac{9}{16}$$

x_i

1	1	1	0
1	1	1	0
1	1	1	0
0	0	0	0

M
 y_i [1.0, 0.0]

x_j

0	0	0	1
0	0	0	1
0	0	0	1
1	1	1	1

$(1 - M)$
 y_j [0.0, 1.0]



$\tilde{y}$ [0.56, 0.44]

Advanced Mix-up Algorithm

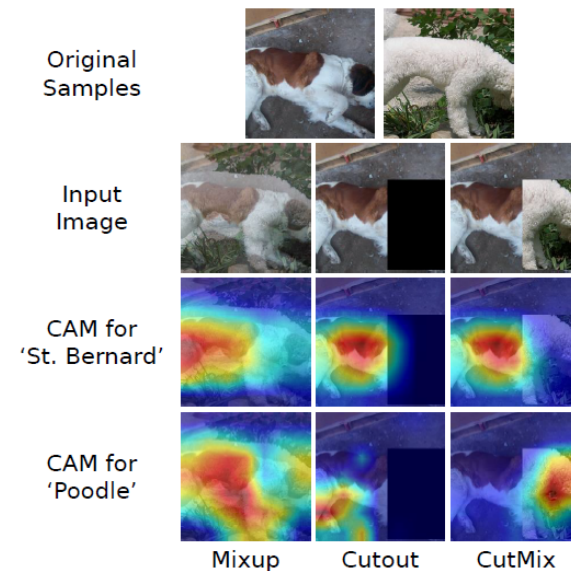
❖ CutMix 알고리즘 실험

- 다른 방법론들과 비교해 Classification, Localization, Detection task에서 좋은 성능을 보여줌
- CutMix를 통해 생성된 샘플로부터 각 클래스의 CAM이 적절하게 출력됨을 시각적으로 확인

results on each task

Image	ResNet-50	Mixup [47]	Cutout [3]	CutMix
				
Label	Dog 1.0	Dog 0.5 Cat 0.5	Dog 1.0	Dog 0.6 Cat 0.4
ImageNet Cls (%)	76.3 (+0.0)	77.4 (+1.1)	77.1 (+0.8)	78.6 (+2.3)
ImageNet Loc (%)	46.3 (+0.0)	45.8 (-0.5)	46.7 (+0.4)	47.3 (+1.0)
Pascal VOC Det (mAP)	75.6 (+0.0)	73.9 (-1.7)	75.1 (-0.5)	76.7 (+1.1)

CAM visualization



Advanced Mix-up Algorithm

❖ CutMix 알고리즘 한계점

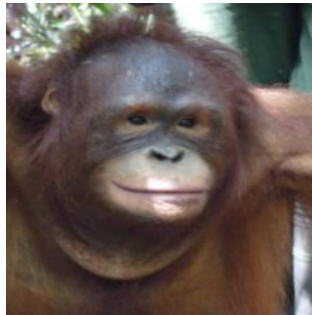
- 다른 샘플의 일부 영역으로 대체하는 과정에서 불필요한 정보를 포함하는 영역을 사용할 수 있음
- 이러한 한계점을 개선하기 위한 연구가 최근 활발하게 진행되고 있음

x_i



y_i [1.0, 0.0]

x_j



y_j [0.0, 1.0]

Advanced Mix-up Algorithm

❖ CutMix 알고리즘 한계점

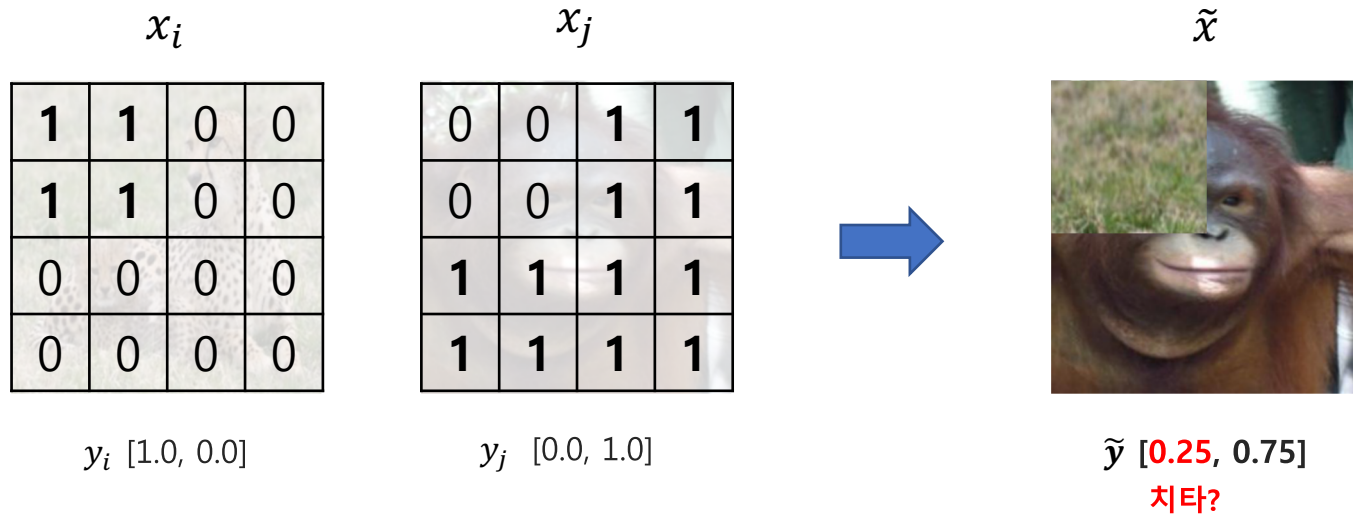
- 다른 샘플의 일부 영역으로 대체하는 과정에서 불필요한 정보를 포함하는 영역을 사용할 수 있음
- 이러한 한계점을 개선하기 위한 연구가 최근 활발하게 진행되고 있음

x_i				x_j			
1	1	0	0	0	0	1	1
1	1	0	0	0	0	1	1
0	0	0	0	1	1	1	1
0	0	0	0	1	1	1	1
y_i [1.0, 0.0]				y_j [0.0, 1.0]			

Advanced Mix-up Algorithm

❖ CutMix 알고리즘 한계점

- 다른 샘플의 일부 영역으로 대체하는 과정에서 불필요한 정보를 포함하는 영역을 사용할 수 있음
- 이러한 한계점을 개선하기 위한 연구가 최근 활발하게 진행되고 있음

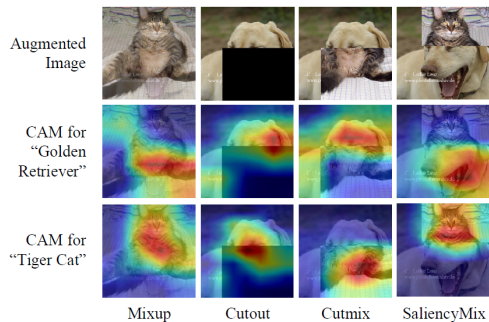
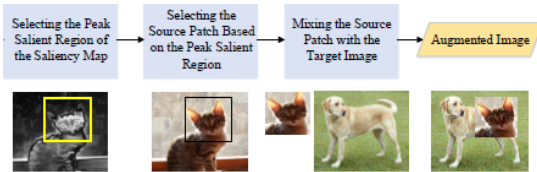


Advanced Mix-up Algorithm

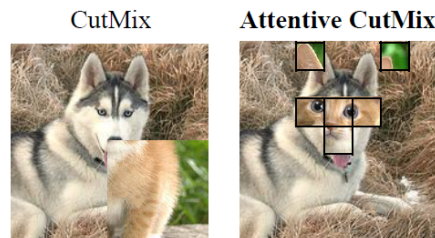
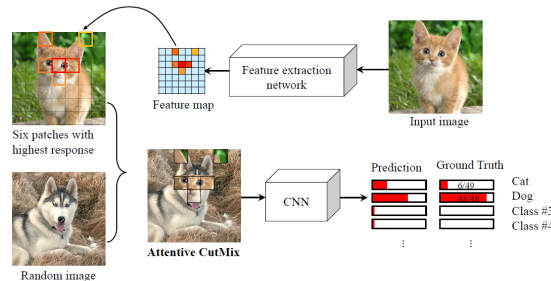
❖ 중요 영역을 고려한 개선된 Mix-up 알고리즘

- 최근 다양한 방법론에서 샘플 내 중요 영역 정보를 이용하여 두 샘플을 섞는 방안을 제안
- 각 방법론마다 중요 영역을 찾는 방법과 섞는 방식이 다름

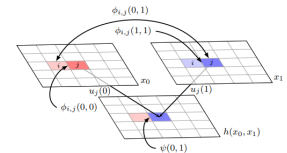
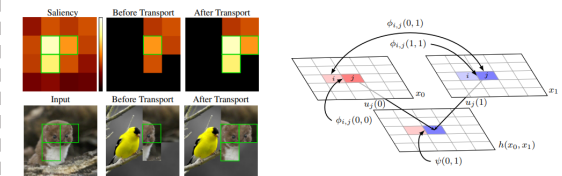
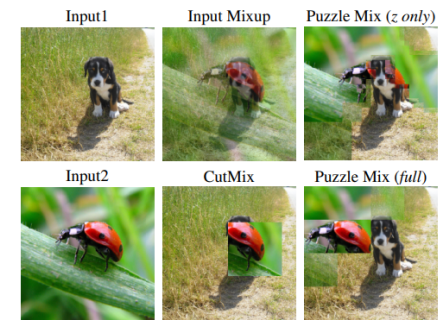
Saliency Mix



Attentive Mix



Puzzle Mix



Reference

- Zhang, H., Cisse, M., Dauphin, Y. N., & Lopez-Paz, D. (2017). mixup: Beyond empirical risk minimization. arXiv preprint arXiv:1710.09412.
- Goodfellow, I. J., Shlens, J., & Szegedy, C. (2014). Explaining and harnessing adversarial examples. arXiv preprint arXiv:1412.6572.
- Thulasidasan, S., Chennupati, G., Bilmes, J., Bhattacharya, T., & Michalak, S. (2019). On mixup training: Improved calibration and predictive uncertainty for deep neural networks. arXiv preprint arXiv:1905.11001.
- Guo, C., Peiss, G., Sun, Y., & Weinberger, K. Q. (2017, July). On calibration of modern neural networks. In International Conference on Machine Learning (pp. 1321-1330). PMLR.
- Zhang, L., Deng, Z., Kawaguchi, K., Ghorbani, A., & Zou, J. (2020). How Does Mixup Help With Robustness and Generalization?. arXiv preprint arXiv:2010.04819.
- Kim, S., Lee, G., Bae, S., & Yun, S. Y. (2020). MixCo: Mix-up Contrastive Learning for Visual Representation. arXiv preprint arXiv:2010.06300.
- Xu, M., Zhang, J., Ni, B., Li, T., Wang, C., Tian, Q., & Zhang, W. (2020, April). Adversarial domain adaptation with domain mixup. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 34, No. 04, pp. 6502-6509).
- Manifold mixup: Better representations by interpolating hidden states. In International Conference on Machine Learning (pp. 6438-6447). PMLR.
- Cutmix: Regularization strategy to train strong classifiers with localizable features. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 6023-6032).
- Chen, J., Yang, Z., & Yang, D. (2020). Mixtext: Linguistically-informed interpolation of hidden space for semi-supervised text classification. arXiv preprint arXiv:2004.12239.
- An improved adventitious lung sound classification using non-local block resnet neural network with mixup data augmentation. In Proc. Interspeech (Vol. 2020, pp. 2902-2906).
- Kim, J. H., Choo, W., & Song, H. O. (2020, November). Puzzle mix: Exploiting saliency and local statistics for optimal mixup. In International Conference on Machine Learning (pp. 5275-5285). PMLR.
- Uddin, A. F. M., Monira, M., Shin, W., Chung, T., & Bae, S. H. (2020). Saliencymix: A saliency guided data augmentation strategy for better regularization. arXiv preprint arXiv:2006.01791.
- Walawalkar, D., Shen, Z., Liu, Z., & Savvides, M. (2020). Attentive cutmix: An enhanced data augmentation approach for deep learning based image classification. arXiv preprint arXiv:2003.13048.

❖ 결론

- 모델의 일반화 성능 향상을 위한 데이터 증강기법 중 Mix-up 알고리즘에 대해 설명
- 다양한 실험 조건에서 좋은 성능을 보임을 확인
- 이후 Mix-up 알고리즘을 적용 및 개선한 여러 연구들이 진행되었음
 - 모델링 고도화 / 적용 데이터셋 확장 / 다른 연구 분야에 적용
 - Manifold mixup / CutMix 등 여러 개선된 Mix-up 알고리즘이 연구됨
- 최근에는 중요 영역 정보를 이용한 개선된 Mix-up 연구가 진행되고 있음
- 본 세미나를 통해 Mix-up에 대한 전반적인 이해를 돕고, 다양한 연구 분야에 활용되기를 기대