
Regularization With Stochastic Transformations and Perturbations for Deep Semi-Supervised Learning

School of Industrial and Management Engineering, Korea University

Jihyun Kim

Contents

- ❖ Research Purpose
- ❖ Proposed Unsupervised Loss Function
- ❖ Experiments and Results
- ❖ Conclusion

Research Purpose

❖ Regularization With Stochastic Transformation and Perturbations for Deep Semi-Supervised Learning (NeurIPS, 2016)

- 2022년 08월 04일 기준 658회 인용됨

Regularization With Stochastic Transformations and Perturbations for Deep Semi-Supervised Learning

Mehdi Sajjadi

Mehran Javanmardi

Tolga Tasdizen

Department of Electrical and Computer Engineering
University of Utah
{mehdi, mehran, tolga}@sci.utah.edu

Abstract

Effective convolutional neural networks are trained on large sets of labeled data. However, creating large labeled datasets is a very costly and time-consuming task. Semi-supervised learning uses unlabeled data to train a model with higher accuracy when there is a limited set of labeled data available. In this paper, we consider the problem of semi-supervised learning with convolutional neural networks. Techniques such as randomized data augmentation, dropout and random max-pooling provide better generalization and stability for classifiers that are trained using gradient descent. Multiple passes of an individual sample through the network might lead to different predictions due to the non-deterministic behavior of these techniques. We propose an unsupervised loss function that takes advantage of the stochastic nature of these methods and minimizes the difference between the predictions of multiple passes of a training sample through the network. We evaluate the proposed method on several benchmark datasets.

Research Purpose

Introduction

❖ Regularization With Stochastic Transformation and Perturbations for Deep Semi-Supervised Learning (NeurIPS, 2016)

- Image Classification Task에서 Semi-Supervised Learning을 위한 방법론 제시
 - Convolutional Neural Network를 이용하여 모델을 구축하는 상황 가정
 - ✓ 대량의 Labeled 데이터를 필요로 하는 지도 학습의 한계 극복
 - ✓ Unlabeled 데이터를 이용해 데이터의 본질적 특성을 모델링하고, 소량의 Labeled 데이터만을 통해 모델의 일반화 성능 향상
- 예측모델의 Stochastic Nature에 의한 Non-Deterministic Behavior를 최소화 함으로써 모델의 일반화 성능 및 안정성을 높임

Stochastic: 랜덤한 결과, 불확실성을 포함한 다양한 프로세스를 의미
본 논문에서는 같은 Input이더라도 다른 Output이 나올 수 있는 예측모델의 성질에 대해 “Stochastic Nature”라 명명

Non-Deterministic: 같은 Input이더라도 다른 Output이 나올 수 있는 결과의 가능성을 의미
본 논문에서는 예측 모델의 Stochastic한 성격에 의한 Non-Deterministic Behavior를 최소화 하는 방안 제시

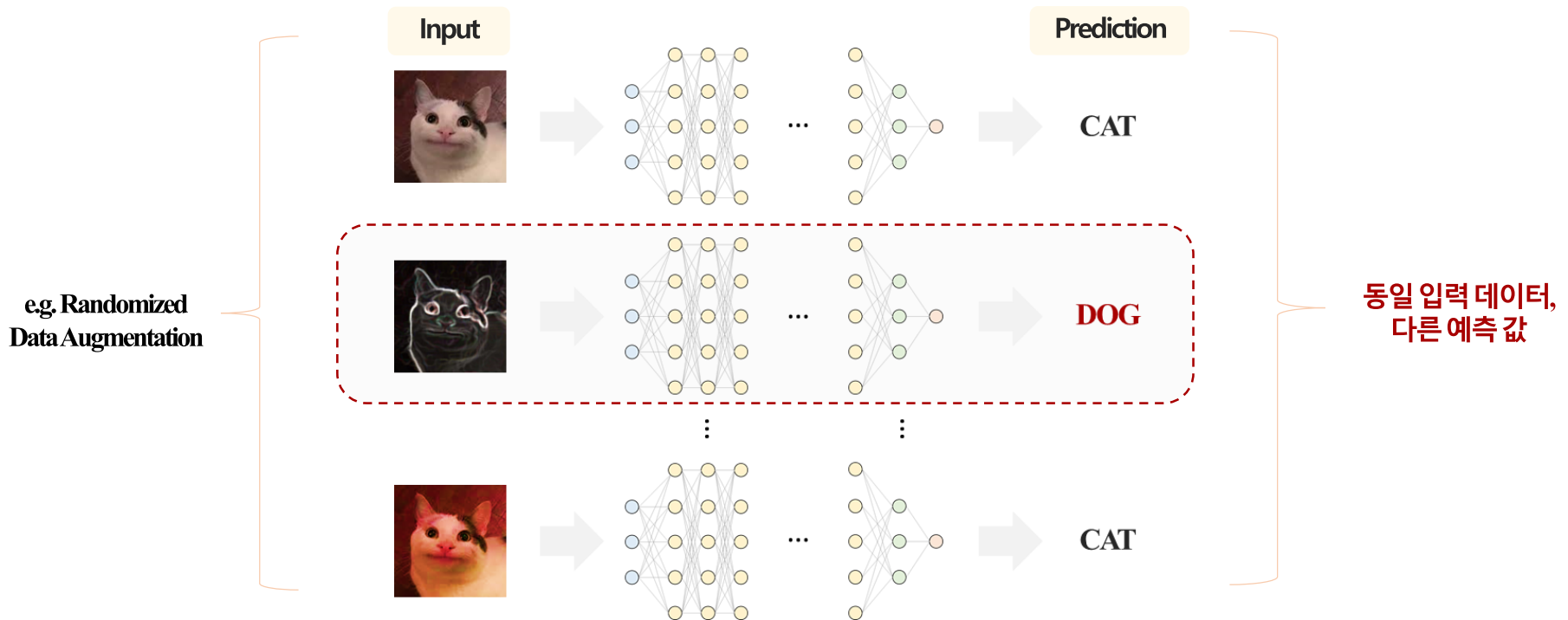
Research Purpose

Problem Setup

Transformations

- ❖ Problem: 데이터를 변형하거나 모델에 작은 변화가 생기면 예측 값이 다를 수 있음
 - **Randomization**은 모델의 일반화 성능을 높이기 위해 필요한 요소이지만, 동일한 입력 데이터에 대해서 다른 예측 값을 도출하는 문제 야기 (1/2)

1. Transformations: Any Linear/Non-linear Data Augmentation

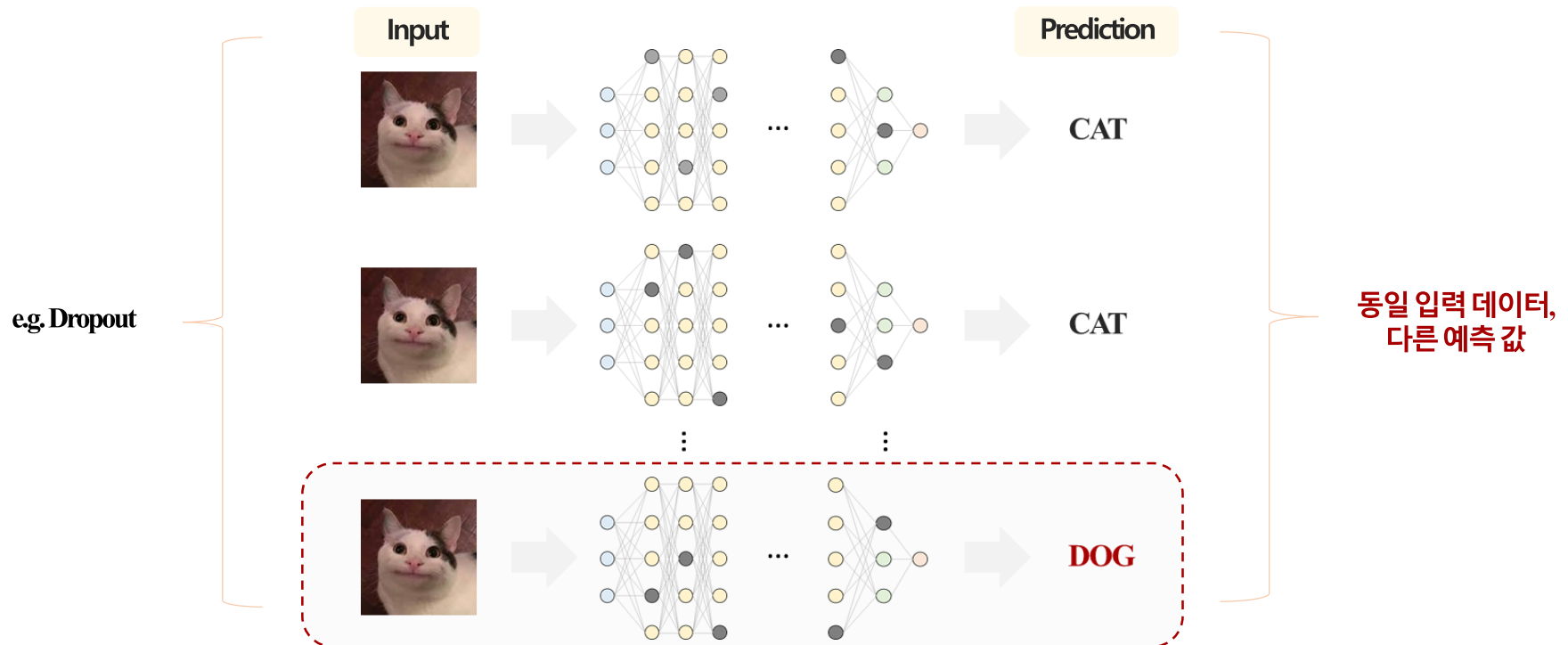


Research Purpose

Problem Setup

- ❖ **Problem:** 데이터를 변형하거나 **모델에 작은 변화가 생기면 예측 값이 다를 수 있음**
- **Randomization**은 모델의 일반화 성능을 높이기 위해 필요한 요소이지만, 동일한 입력 데이터에 대해서 다른 예측 값을 도출하는 문제 야기 (2/2)

2. Perturbations: Dropout, Randomized Max Pooling, and etc.



Proposed Unsupervised Loss Function

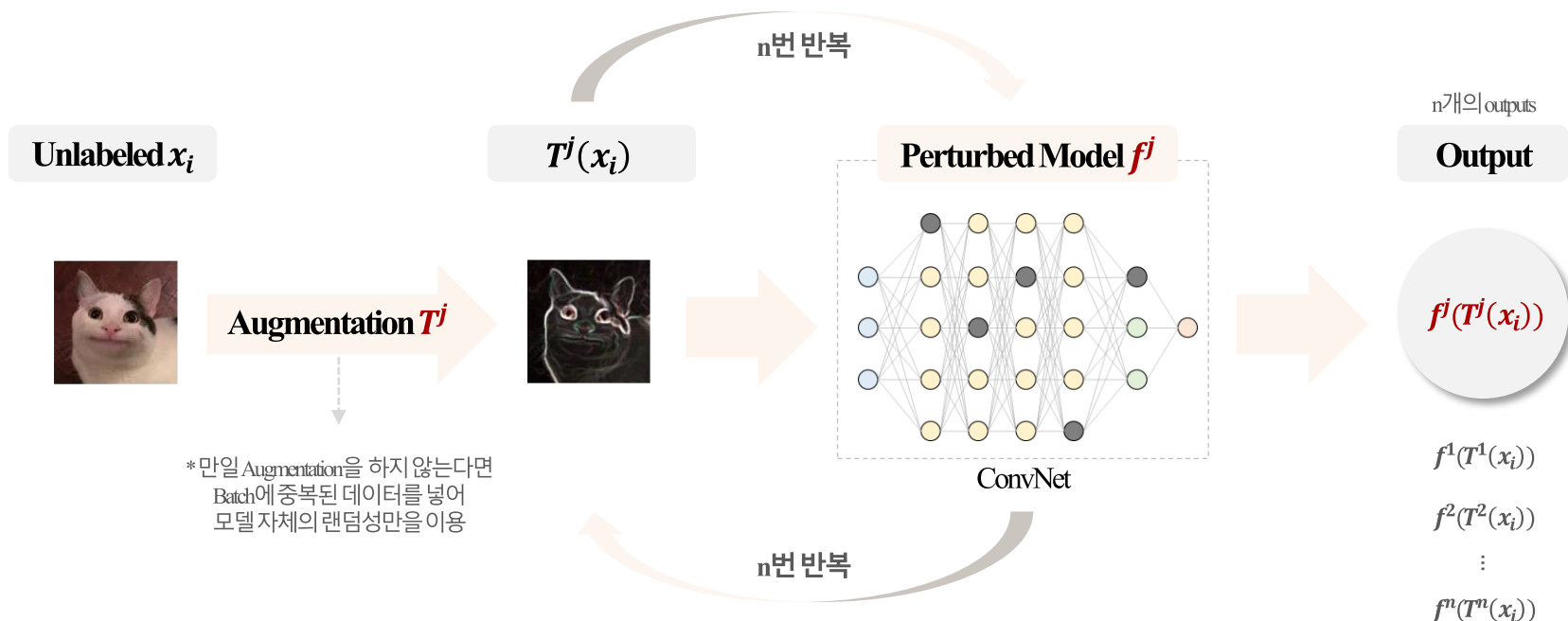
Method

❖ Solution: 같은 입력 데이터에 대해 동일한 결과를 내기 위한 Unsupervised Loss 제안

- 1. TS Loss Function (1/2)

TS: Transformation/Stability

- ✓ 학습 시 동일 데이터를 n 번 반복 입력함으로써 도출되는 n 개 예측 값들의 MSE 값을 최소화
- ✓ 예측 값에 차이가 없도록 규제함으로써 모델 및 데이터의 랜덤성을 제어하여 예측 일관성 확보



Proposed Unsupervised Loss Function

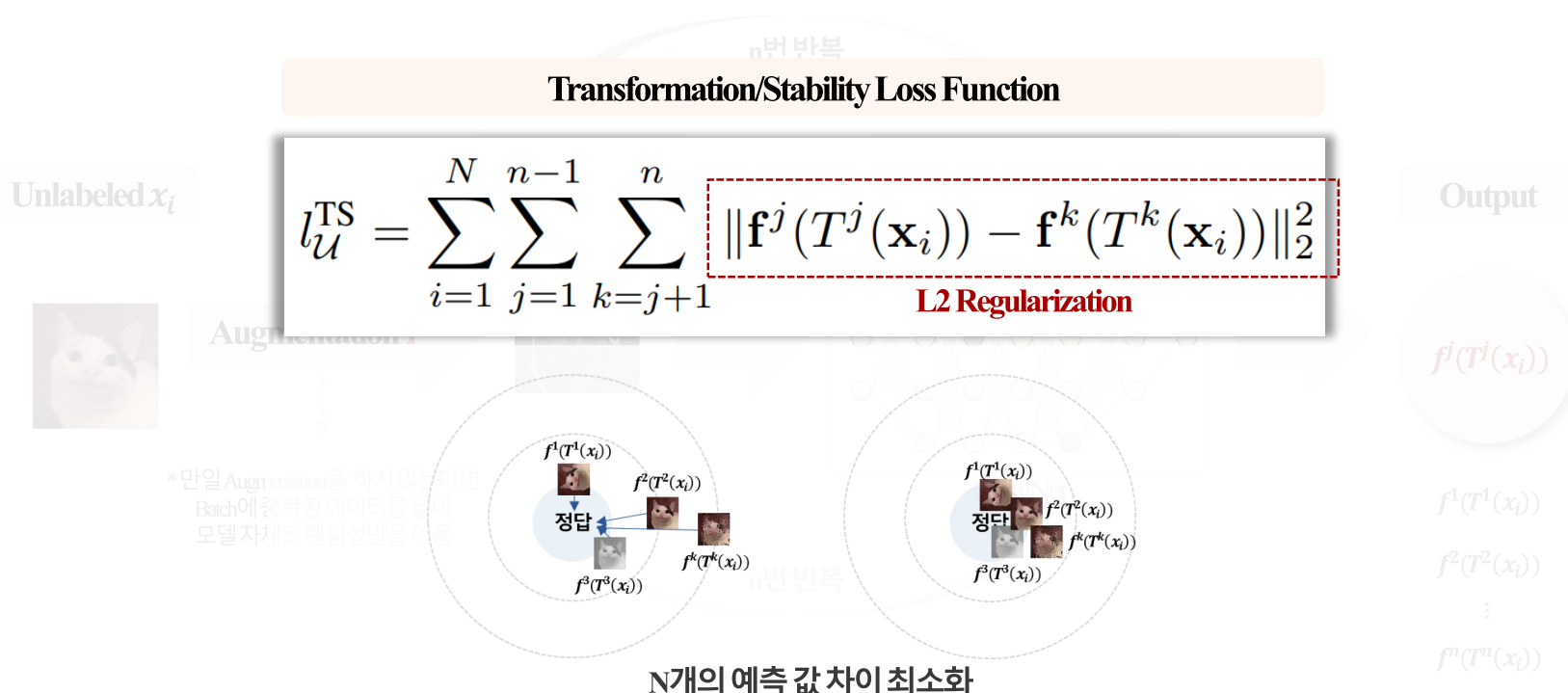
Method

❖ **Solution: 같은 입력 데이터에 대해 동일한 결과를 내기 위한 Unsupervised Loss 제안**

• **1. TS Loss Function (1/2)**

TS: Transformation/Stability

- ✓ 학습 시 동일 데이터를 n번 반복 입력함으로써 도출되는 n개 예측 값들의 MSE 값을 최소화
- ✓ 예측 값에 차이가 없도록 규제함으로써 모델 및 데이터의 랜덤성을 제어하여 예측 일관성 확보



Proposed Unsupervised Loss Function

Method

❖ Solution: 같은 입력 데이터에 대해 동일한 결과를 내기 위한 Unsupervised Loss 제안

• 2. ME Loss Function (1/2)

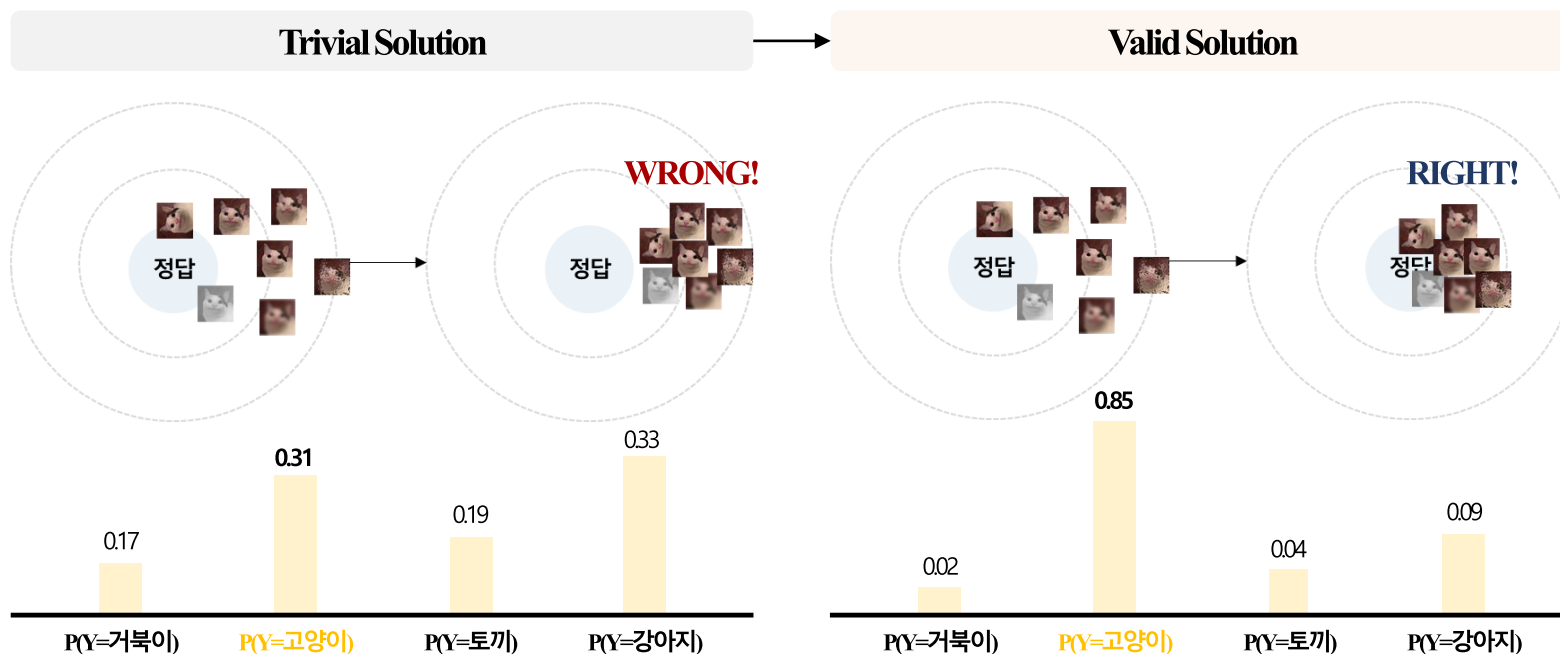
ME: Mutual Exclusivity

- ✓ TS 손실함수와 더불어 Mutual Exclusivity 손실함수를 더하여 학습하는 아이디어 제안

→ 제안된 TS 손실함수의 Trivial Solution을 방지하기 위함

- ✓ 따라서 정답과 관련 있으면서도 일관성 있는 예측 값으로 학습 유도

예측 벡터 내 개별 원소에 대해서는 규제를 부과하지 않아, 일관된 예측 성능을 보이기는 하지만 실제 결과값과는 무관한 값으로 학습될 위험



Proposed Unsupervised Loss Function

Method

❖ Solution: 같은 입력 데이터에 대해 동일한 결과를 내기 위한 Unsupervised Loss 제안

- 2. ME Loss Function (1/2)

ME: Mutual Exclusivity

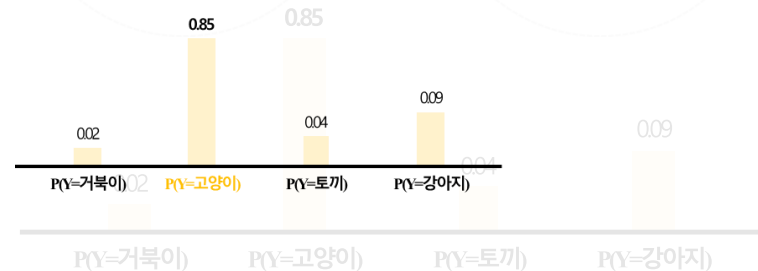
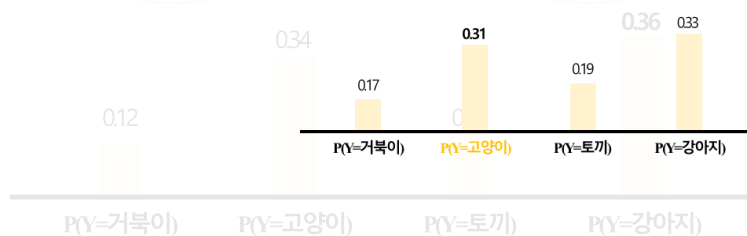
- ✓ TS 손실함수와 더불어 Mutual Exclusivity 손실함수를 더하여 학습하는 아이디어 제안
→ 제안된 TS 손실함수의 Trivial Solution을 방지하기 위함
- ✓ 따라서 정답과 관련 있으면서도 일관성 있는 예측 값으로 학습 유도

Mutual Exclusivity Loss Function

$$l_{\mathcal{U}}^{\text{ME}} = \sum_{i=1}^N \sum_{j=1}^n \left(- \sum_{k=1}^C f_k^j(\mathbf{x}_i) \prod_{l=1, l \neq k}^C (1 - f_l^j(\mathbf{x}_i)) \right)$$

Trivial Solution: -0.42

Valid Solution: -0.72



Proposed Unsupervised Loss Function

Method

❖ Solution: 같은 입력 데이터에 대해 동일한 결과를 내기 위한 Unsupervised Loss 제안

- **Total Unsupervised Loss: TS Loss Function + ME Loss Function**

- ✓ 논문에서는 $\lambda_1 = 0.1, \lambda_2 = 1$ 로 고정한 뒤 실험 진행

- ✓ 더불어 Semi-Supervised Learning을 위한 Total Loss는 아래와 같음

→ 본 논문에서는 Unsupervised Loss가 어떻게 Supervised Loss와 함께 사용되는지는 명쾌하게 설명하지 않았으나
저자의 추가 발표자료에서 아래와 같이 언급됨

Unsupervised Loss Function

$$l_u = \lambda_1 l_u^{ME} + \lambda_2 l_u^{TS}$$

Total Loss Function

$$l = l_L + \lambda l_U$$

Experiments and Results

Experiments

❖ 제안한 두 개의 Loss Functions: (1) TS Loss, (2) ME Loss를 이용하여 정확도 향상

- 실험에서 사용된 ConvNet은 (1) Cuda-ConvNet, (2) Sparse Convolutional Networks 두 가지

AlexNet이 대표적인 예이며, Batch Size 128 기준,
TS Loss에서 중복 데이터(n)을 4로 설정

Fractional Max-pooling을 이용하며, TS Loss에서
중복 데이터(n)을 5로 설정

- 분석 데이터셋: MNIST, CIFAR10, CIFAR100, SVHN, NORB and ILSVRC 2012 Challenge

MNIST

	labeled data only	transform /stability loss	mut-excl loss [30]	both losses	ladder net. [28]	ladder net baseline [28]
100 :	5.44 ± 1.48	0.76 ± 0.61	3.92 ± 1.12	0.55 ± 0.16	0.89 ± 0.50	6.43 ± 0.84
all:	0.32 ± 0.02	0.29 ± 0.02	0.30 ± 0.03	0.27 ± 0.02	-	0.36

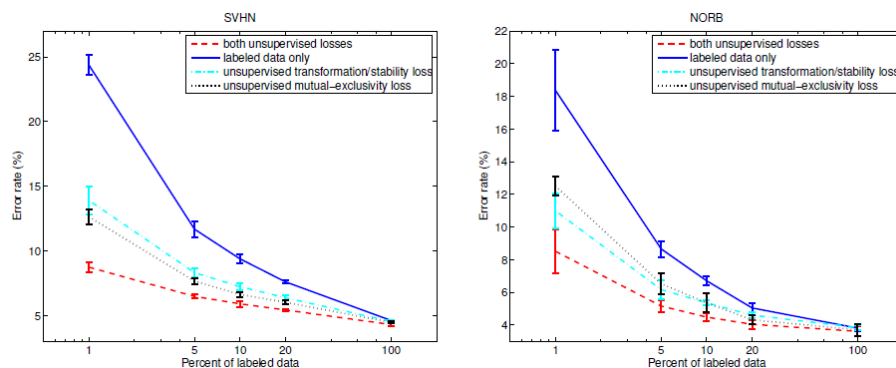
- ✓ Sparse Convolutional Network 이용
- ✓ 데이터 증강(Transformations)은 사용하지 않음
→ 따라서 $T^j(x_i)$ 는 Identity Function으로서 기능하고, 오로지 Dropout(Perturbations)만 이용
- ✓ 2개의 손실함수를 함께 사용하였을 때 Error Rate 0.55% 까지 줄이며 SOTA 성능 달성
→ Labeled 데이터를 100개 사용 (10개 Class에서 10개의 Labeled 데이터 무작위 추출)

Experiments and Results

Experiments

- ❖ 제안한 두 개의 Loss Functions: (1) TS Loss, (2) ME Loss를 이용하여 정확도 향상
 - 실험에서 사용된 ConvNet은 (1) Cuda-ConvNet, (2) Sparse Convolutional Networks 두 가지
 - 분석 데이터셋: MNIST, CIFAR10, CIFAR100, **SVHN**, **NORB** and ILSVRC 2012 Challenge

SVHN and NORB



- ✓ **Cuda-ConvNet 이용**: 2 Conv Layers with 64 maps and kernel size=5 and 2 Locally Connected Layers with 32 maps and kernel size=3
- ✓ 데이터 증강 기법으로는 Random Crop과 Random Rotation을 이용
Perturbations으로는 Dropout과 Random Max-Pooling 이용
- ✓ Labeled 데이터 사용 비율을 1%, 5%, 10%, 20%, 100%로 달리 해가며 실험 진행
→ 2개의 Loss Term을 모두 이용하면, **Labeled 데이터가 적어도 상대적으로 좋은 Error Rate 산출**
(NORB는 1%만을 이용해도 지도학습과 성능 유사)

Experiments and Results

Experiments

❖ 제안한 두 개의 Loss Functions: (1) TS Loss, (2) ME Loss를 이용하여 정확도 향상

- 실험에서 사용된 ConvNet은 (1) Cuda-ConvNet, (2) Sparse Convolutional Networks 두 가지
- 분석 데이터셋: MNIST, **CIFAR10**, CIFAR100, SVHN, NORB and ILSVRC 2012 Challenge

CIFAR10

	transformation/stability+mutual-exclusivity	ladder networks [28]
labeled data only:	13.60 \pm 0.24	23.33 \pm 0.61
semi-supervised:	11.29 \pm 0.24	20.40 \pm 0.47

Error Rates on test data for CIFAR10 with 4000 labeled samples (mean % \pm std)

- ✓ **Sparse Convolutional Network 이용**
- ✓ 데이터 증강 기법으로는 Translations, Rotations, Flipping, Stretching, Shearing Operations 이용
Perturbations으로는 Dropout 이용
- ✓ Semi-Supervised 방법론 사용에 있어, **Unlabeled Data에 대해서 2가지 Loss Term을 모두 적용하였을 때 성능 향상이 이루어짐**
→ SOTA(Error Rate: 3.47%)와 비교하였을 때 3.18%를 달성하였으며, 모델이 커지면 3.00%까지 성능 향상

Experiments and Results

Experiments

❖ 제안한 두 개의 Loss Functions: (1) TS Loss, (2) ME Loss를 이용하여 정확도 향상

- 실험에서 사용된 ConvNet은 (1) Cuda-ConvNet, (2) Sparse Convolutional Networks 두 가지
- 분석 데이터셋: MNIST, CIFAR10, CIFAR100, SVHN, NORB and ILSVRC 2012 Challenge

ImageNet								
	rep 1	rep 2	rep 3	rep 4	rep 5	mean ± std	mutual xcl [30]	[21] ~1.5% of data
labeled only:	45.73	46.15	46.06	45.57	46.08	45.91 ± 0.25	45.63	85.9
semi-sup:	39.50	39.99	39.94	39.70	40.08	39.84 ± 0.23	42.90	84.2

- ✓ AlexNet 이용
- ✓ 데이터 증강 기법으로는 Random Translations, Flipping, Color noise 이용
Perturbations으로는 Dropout 이용
- ✓ 학습 데이터 중 Labeled 데이터를 10% 이용하였을 때 20 epochs 만에 수렴
- ✓ Class 개수가 많음에도 불구하고, 제안 방법론의 성능 향상 확인

Conclusion

Discussion and Conclusion

- ❖ **Randomization 단점 보완을 위해 2개의 Unsupervised Loss Function 제안**
 - **Unsupervised Loss: TS Loss Function + ME Loss Function**
 - ✓ 제안 방법론을 여러 이미지 벤치마크 데이터셋과 ConvNet 프레임워크를 통해 제안
 - ✓ TS Loss Function, ME Loss Function 2가지를 모두 사용하였을 때 정확도가 향상됨을 실험적으로 확인
 - ✓ Labeled 데이터가 적은 상황에서 Semi-Supervised Learning 방법론을 사용 가능함을 증명
 - Stochastic Transformations 혹은 Perturbations에 의해 **모델의 일반화 성능을 높이는 것은 물론, 예측 값이 달라질 수 있는 단점을 보완한 방법론**

Thank You