Knowledge Distillation for Anomaly Detection

DMQA Open Seminar

2023.11.03

Data Mining & Quality Analytics Lab.

백민지

발표자 소개

❖ 백민지(Minji Baek)

- 고려대학교 산업경영공학과 재학 중
- Data Mining & Quality Analytics Lab.(김성범 교수님)
- 석사과정(2023.03 ~)

❖ 연구 관심 분야

- Machine Learning & Deep Learning Algorithms
- Anomaly Detection

❖ E-mail

- hotbluechip@korea.ac.kr

Contents

Introduction

Introduction

Anomaly Detection

❖ 이상치 탐지(Anomaly detection)란?

학습 데이터 셋에서 비정상 샘플을 감지해 찾아내는 것

Introduction

Anomaly Detection

❖ **이상치 종류**

Novelty : 데이터의 본질적인 특성은 같지만, 유형이 다른 관측치
정상 분포내의 새로운 비정상 샘플

김성범교수님. [김성범[교수 / 산업경영공학부]]. (2022,May 29). *[핵심 머신러닝] Anomaly Detection (개요, 확률분포기반)*[Video]. YouTube. https://youtu.be/TqSwuCX7Lds?si=_OM86vBT-Wsfj3A8

Introduction

Anomaly Detection

❖ **이상치 종류**

Anomaly : 대부분의 데이터와 특성이 다른 관측치
정상 데이터 분포와 다른 분포를 가지는 것

김성범교수님. [김성범[교수 / 산업경영공학부]]. (2022,May 29). *[핵심 머신러닝] Anomaly Detection (개요, 확률분포기반)*[Video]. YouTube. https://youtu.be/TqSwuCX7Lds?si=_OM86vBT-Wsfj3A8

Introduction

Anomaly Detection

❖ **이상치 종류**

Outlier : 대부분의 데이터와 본질적인 특성이 다른 관측치
정상 데이터 분포에서 벗어나 오염을 일으키는 샘플

김성범교수님. [김성범[교수 / 산업경영공학부]]. (2022,May 29). *[핵심 머신러닝] Anomaly Detection (개요, 확률분포기반)*[Video]. YouTube. https://youtu.be/TqSwuCX7Lds?si=_OM86vBT-Wsfj3A8

Introduction

Anomaly Detection

❖ 이상치 종류

Out-of-distribution : 학습 데이터 셋과 다른 종류의 데이터 셋

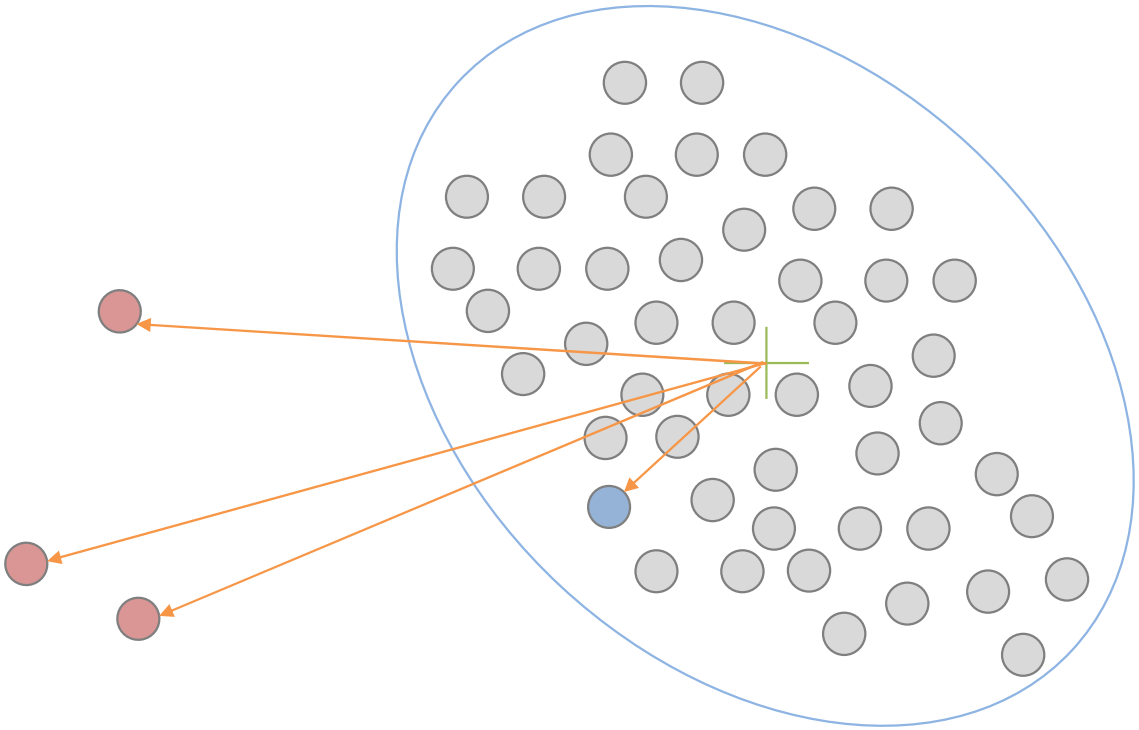
Introduction

Anomaly Detection

❖ **Density/Distance-based Methods**

데이터의 밀도 또는 거리 척도를 통해서 Majority 군집을 생성하여 이상치를 탐지

- Kernel Density Estimation
- Gaussian Mixture Estimation
- K-Nearest Neighbor
- LOF(Local Outlier Factor)

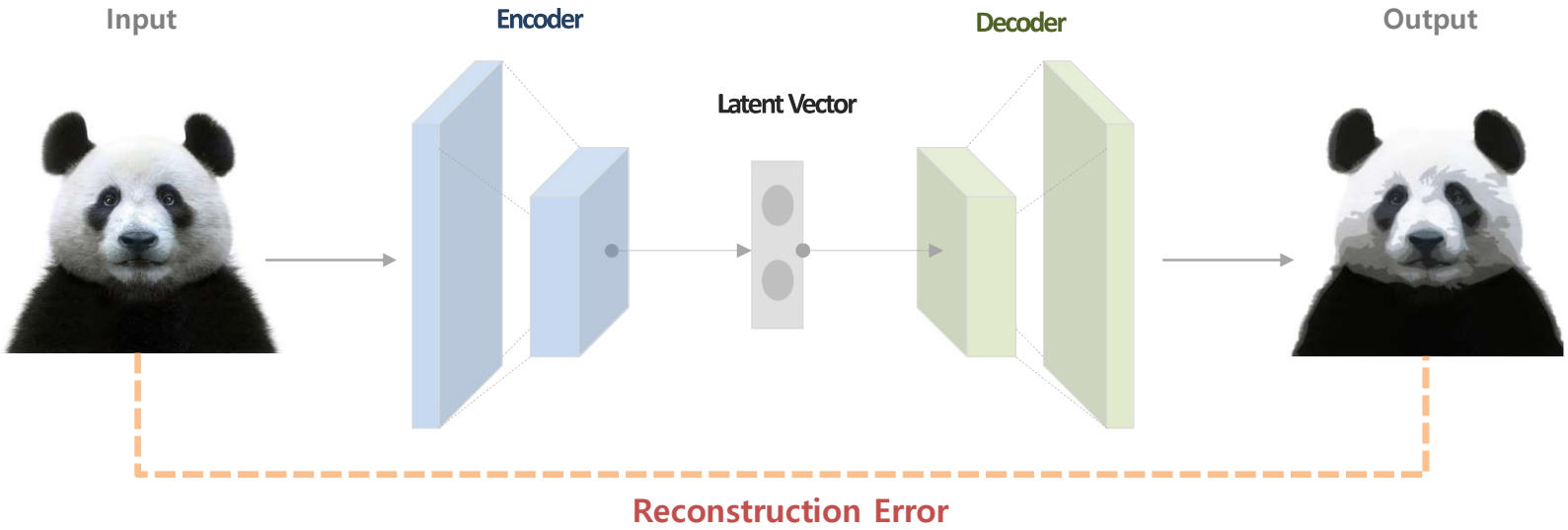


Introduction

Anomaly Detection

❖ **Reconstruction-based Methods**

- 고차원 데이터에서 주로 사용하는 방법론이며 Autoencoder 및 PCA 등의 모델을 사용
- 데이터를 압축/복원하고 Reconstruction error 기반으로 이상탐지

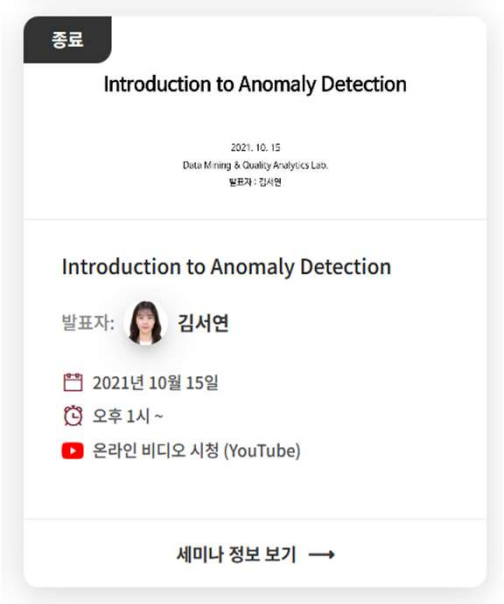


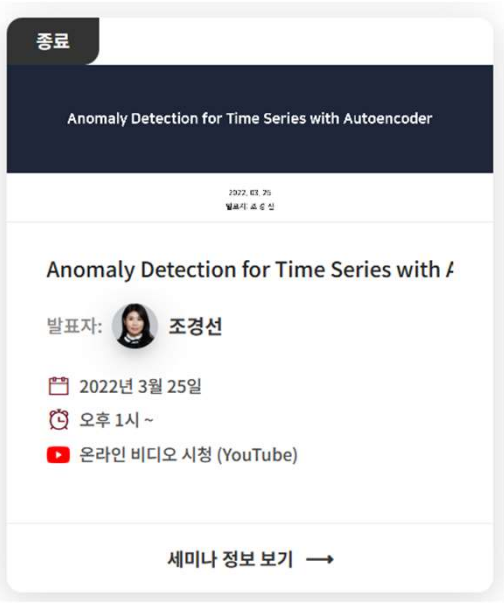
Reconstruction Error가 특정 임계값보다 큰 경우 이상치로 판단

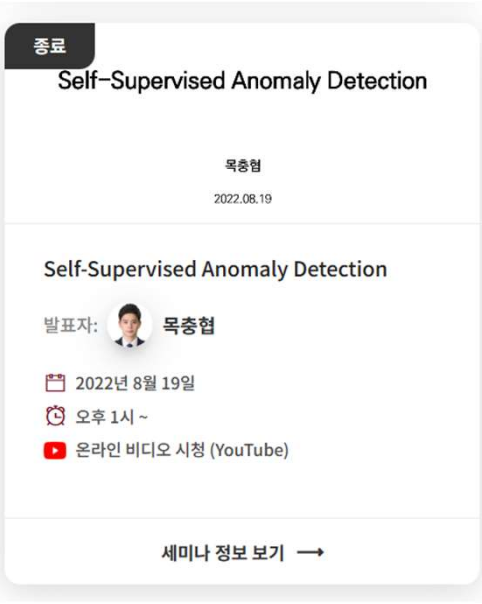
Introduction

Anomaly Detection

❖ 이상치 탐지(Anomaly Detection) 관련 세미나 참고





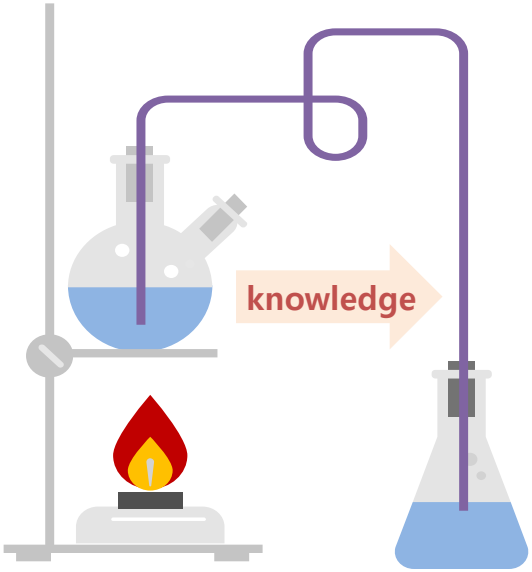


Introduction

What is Knowledge Distillation?

❖ Concept of knowledge distillation

지식(Knowledge) + 증류(Distillation)의 합성어

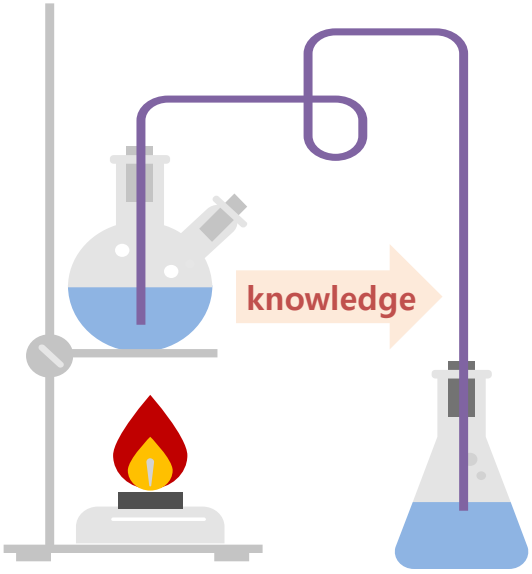


Introduction

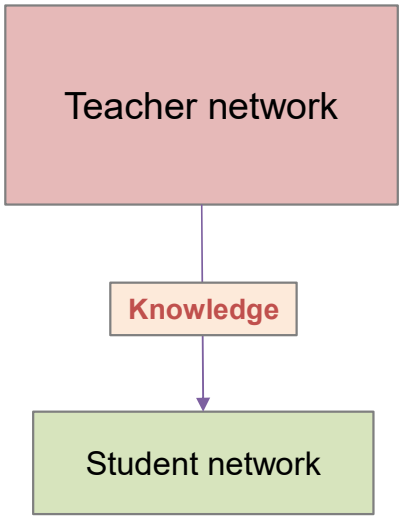
What is Knowledge Distillation?

❖ Concept of knowledge distillation

지식(Knowledge) + 증류(Distillation)의 합성어



"학습된 모델로부터 지식을 추출하는 것"



➢ **Teacher Network(T)**

- ✓ 복잡하고 큰 모델
- ✓ 성능이 좋음
- ✓ 컴퓨팅 리소스가 큼

➢ **Student Network(S)**

- ✓ 단순하고 작은 모델
- ✓ 추론이 빠름
- ✓ T보다 성능이 낮음

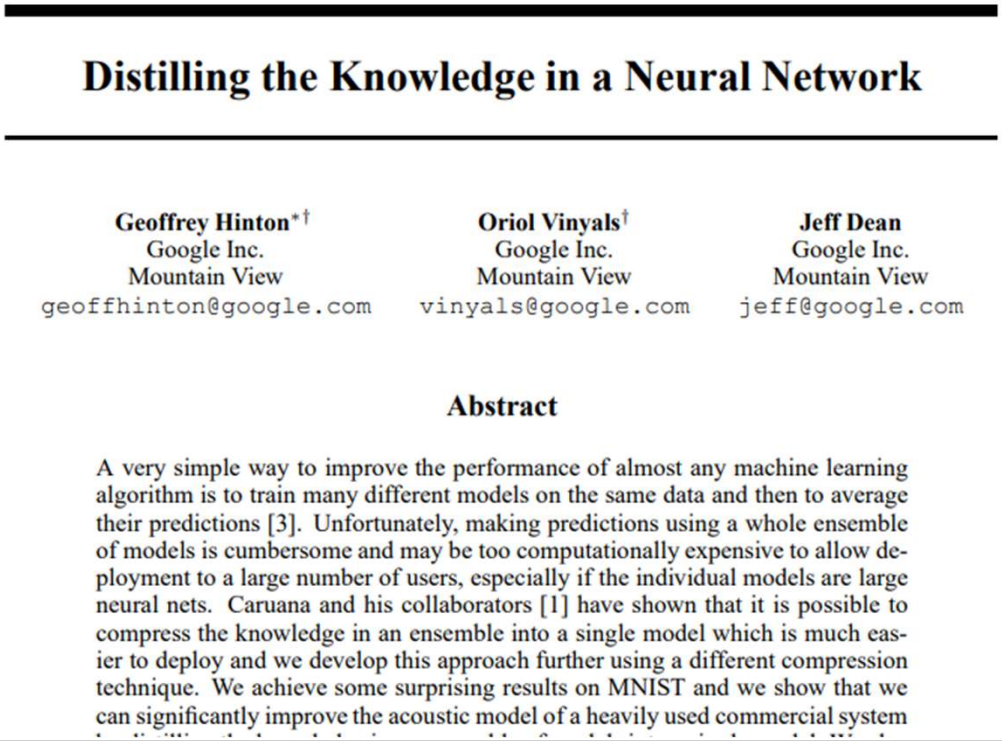
잘 학습된 큰 모델(Teacher Network)에서 증류한 지식을 작은 모델(Student)로 전달하는 것

Introduction

What is Knowledge Distillation?

❖ **Distilling the knowledge in a neural network** [1]

- 2014 Neural Information Processing Systems(NeurIPS) workshop에서 발표된 논문 (2023년 10월 30일 기준 16288회 인용)
- Knowledge distillation 개념을 딥러닝 모델에 처음 적용시킨 논문



Distilling the Knowledge in a Neural Network

Geoffrey Hinton[*†]
Google Inc.
Mountain View
geoffhinton@google.com

Oriol Vinyals[†]
Google Inc.
Mountain View
vinyals@google.com

Jeff Dean
Google Inc.
Mountain View
jeff@google.com

Abstract

A very simple way to improve the performance of almost any machine learning algorithm is to train many different models on the same data and then to average their predictions [3]. Unfortunately, making predictions using a whole ensemble of models is cumbersome and may be too computationally expensive to allow deployment to a large number of users, especially if the individual models are large neural nets. Caruana and his collaborators [1] have shown that it is possible to compress the knowledge in an ensemble into a single model which is much easier to deploy and we develop this approach further using a different compression technique. We achieve some surprising results on MNIST and we show that we can significantly improve the acoustic model of a heavily used commercial system

[1] Hinton, G., Vinyals, O., & Dean, J. (2015). Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531.

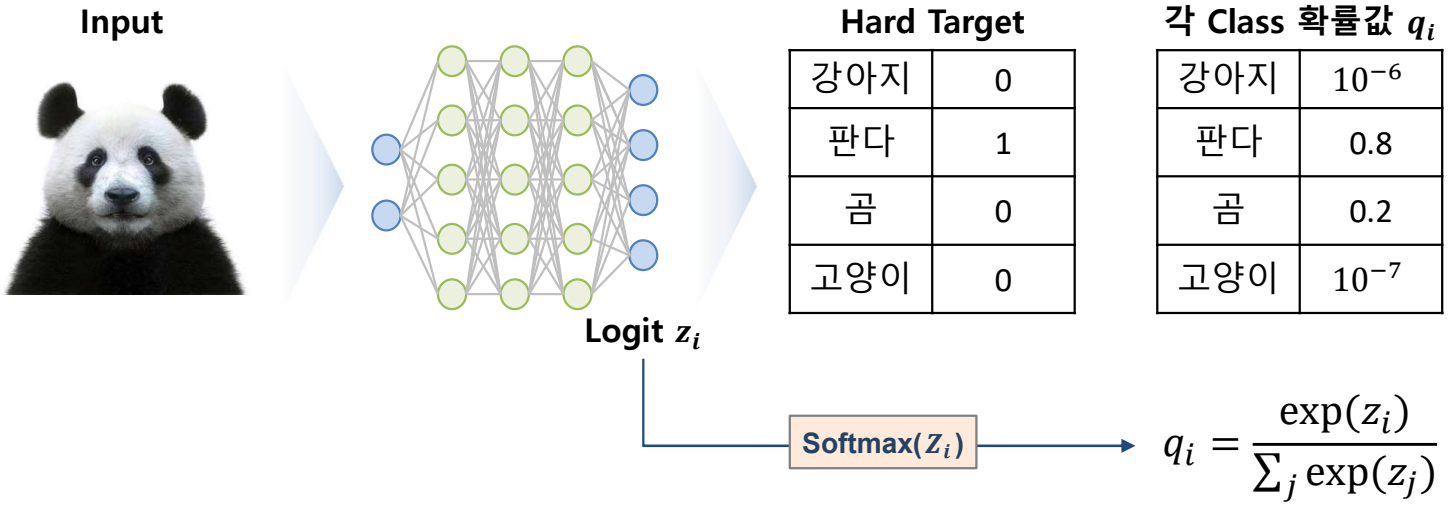
Introduction

What is Knowledge Distillation?

❖ Distilling the knowledge in a neural network [1]

➢ Soft Target 개념

Teacher 모델에서 생성된 클래스 확률을 Student 모델을 학습하기 위한 Soft Target으로 사용



Hard Target

강아지	0
판다	1
곰	0
고양이	0

각 Class 확률값 q_i

강아지	10^{-6}
판다	0.8
곰	0.2
고양이	10^{-7}

일반적으로 Softmax layer를 통해 구한 확률값은 너무 작아서 제대로 모델에 반영하기 어려움

[1] Hinton, G., Vinyals, O., & Dean, J. (2015). Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531.

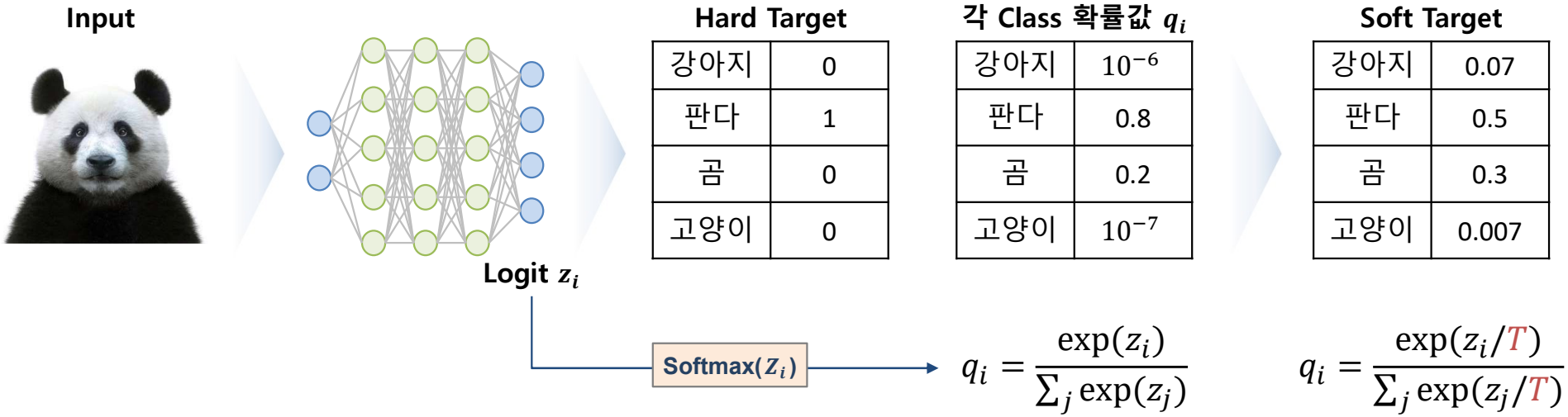
Introduction

What is Knowledge Distillation?

❖ Distilling the knowledge in a neural network [1]

➢ Soft Target 개념

Teacher 모델에서 생성된 클래스 확률을 Student 모델을 학습하기 위한 Soft Target으로 사용



Hard Target

강아지	0
판다	1
곰	0
고양이	0

각 Class 확률값 q_i

강아지	10^{-6}
판다	0.8
곰	0.2
고양이	10^{-7}

Soft Target

강아지	0.07
판다	0.5
곰	0.3
고양이	0.007

T (Temperature) : 값이 1에 가까울수록 Hard해지고, 작아질수록 Soft해지는 하이퍼파라미터

낮은 입력 값의 출력을 더 크게 만들어주고 큰 입력 값의 출력은 작게 만들어, Soft Target을 사용하는 이점을 최대화 함

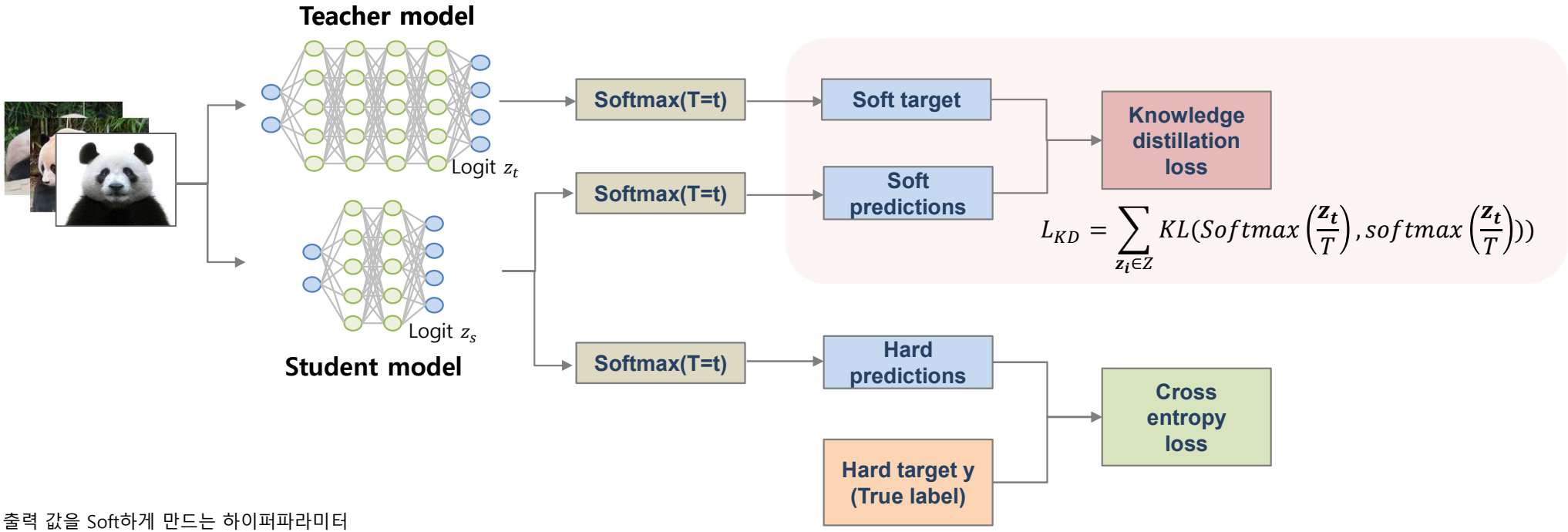
[1] Hinton, G., Vinyals, O., & Dean, J. (2015). Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531.

Introduction

What is Knowledge Distillation?

❖ Distilling the knowledge in a neural network [1]

➢ 모델 구조



T : 출력 값을 Soft하게 만드는 하이퍼파라미터
z_t : Teacher 모델의 Logit값
z_s : Student 모델의 Logit값
y : True label

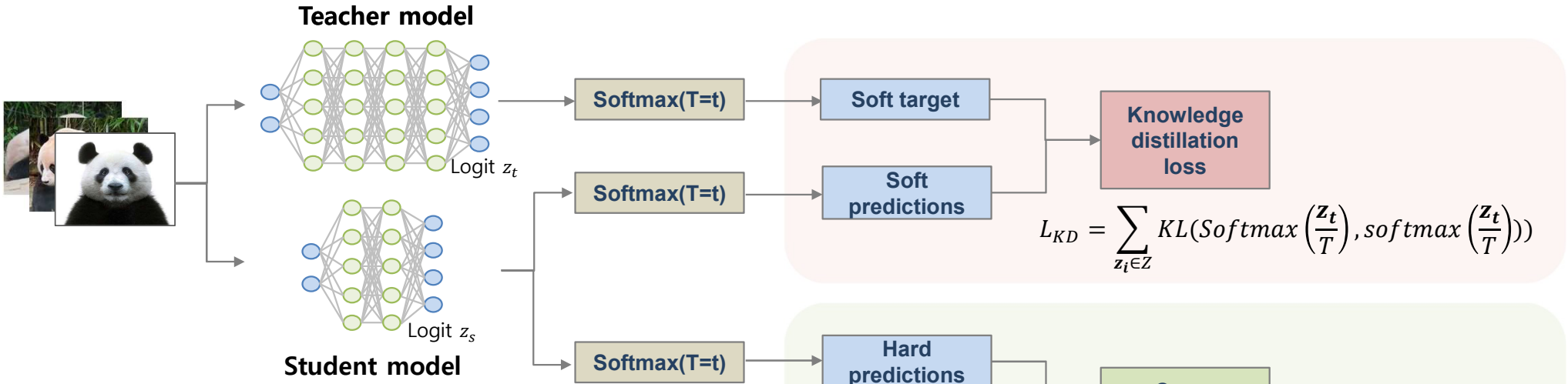
[1] Hinton, G., Vinyals, O., & Dean, J. (2015). Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531.

Introduction

What is Knowledge Distillation?

❖ Distilling the knowledge in a neural network [1]

➢ **모델 구조**



T : 출력 값을 Soft하게 만드는 하이퍼파라미터
z_t : Teacher 모델의 Logit값
z_s : Student 모델의 Logit값
y : True label

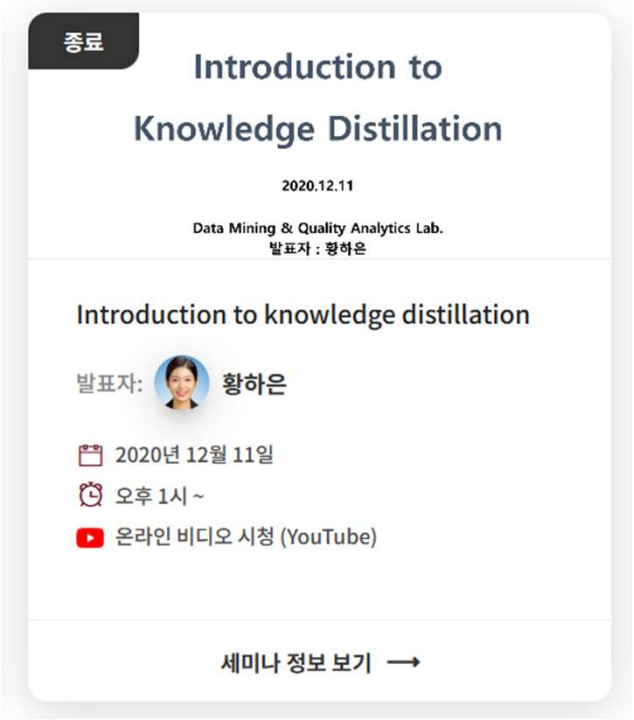
$$\textbf{Total loss} = L_{KD} + \lambda \cdot L_{CE}$$

[1] Hinton, G., Vinyals, O., & Dean, J. (2015). Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531.

Introduction

What is Knowledge Distillation?

❖ **Knowledge Distillation 관련 세미나 참고**



Knowledge Distillation for Anomaly Detection

Knowledge Distillation for Anomaly Detection

Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings

❖ **Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings** [2]

- 2020 CVPR에 게재된 논문 (2023년 10월 30일 기준 426회 인용)
- Knowledge distillation 기반 Anomaly Detection을 처음 제안한 논문

Uninformed Students: Student–Teacher Anomaly Detection with Discriminative Latent Embeddings

Paul Bergmann Michael Fauser David Sattlegger Carsten Steger

MVTec Software GmbH
www.mvtec.com
{paul.bergmann, fauser, sattlegger, steger}@mvtec.com

***Abstract*—We introduce a powerful student–teacher framework for the challenging problem of unsupervised anomaly detection and pixel-precise anomaly segmentation in high-resolution images. Student networks are trained to regress the output of a descriptive teacher network that was pre-trained on a large dataset of patches from natural images. This circumvents the need for prior data annotation. Anomalies are detected when the outputs of the student networks differ from that of the teacher network. This happens when they fail to generalize outside the manifold of anomaly-free training data. The intrinsic uncertainty in the student networks is used as an additional scoring function that indicates anomalies. We compare our method to a large number of existing deep learning based methods for unsupervised anomaly detection. Our experiments demonstrate improvements over state-of-the-art methods on a number of real-world datasets, including the recently introduced MVTec Anomaly Detection dataset that was specifically designed to benchmark anomaly segmentation algorithms.**

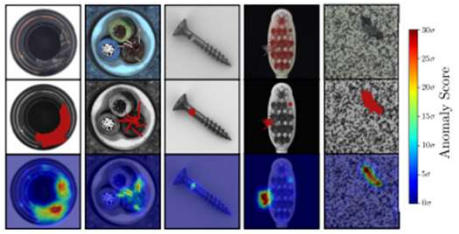


Figure 1: Qualitative results of our anomaly detection method on the MVTec Anomaly Detection dataset. **Top row:** Input images containing defects. **Center row:** Ground truth regions of defects in red. **Bottom row:** Anomaly scores for each image pixel predicted by our algorithm.

(GANs) [31, 32] or Variational Autoencoders (VAEs) [5, 36]. These detect anomalies using per-pixel reconstruc-

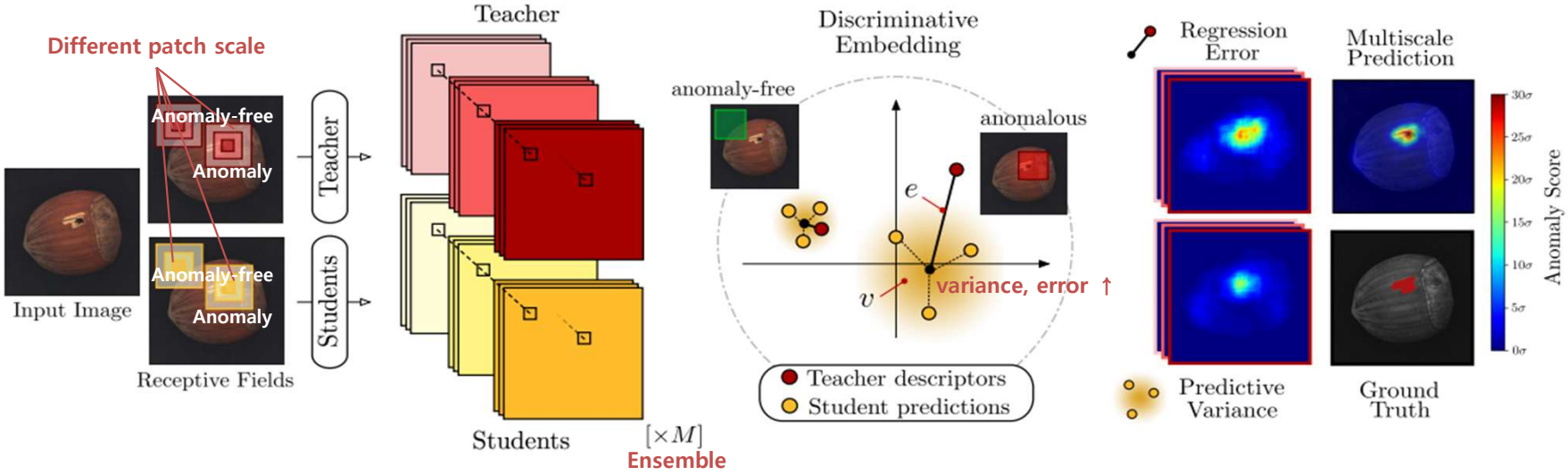
[2] Bergmann, P., Fauser, M., Sattlegger, D., & Steger, C. (2020). Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 4183-4192).

Knowledge Distillation for Anomaly Detection

Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings

❖ **Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings** [2]

➢ **Inference 단계 흐름**



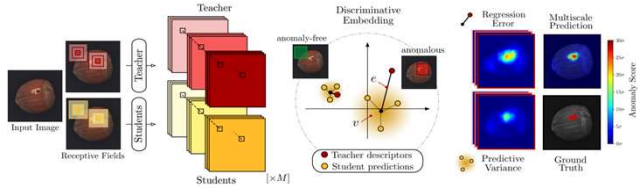
✓ 기존 Generative 모델에서 Downsampling시 해상도 문제를 해결하기 위해 다양한 Receptive field를 갖는 Multiple student ensemble 도입

✓ Anomaly detection를 Feature regression problem 관점으로 봄

[2] Bergmann, P., Fauser, M., Sattlegger, D., & Steger, C. (2020). Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 4183-4192).

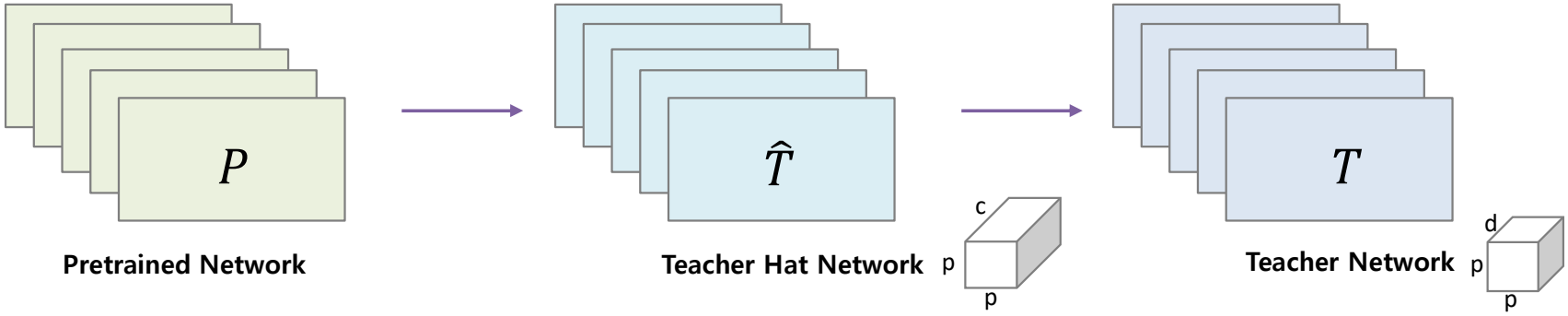
Knowledge Distillation for Anomaly Detection

Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings



❖ **Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings** [2]

➢ **Training teacher network**



✓ 사전 학습된 네트워크를 사용하여 Teacher network T를 효율적으로 구축하기 위해 $\hat{T}$ 도출

✓ T는 Convolution 과 Max pooling을 이용하여 p*p*c 를 차원 d의 metric으로 embedding 되도록 $\hat{T}$ 훈련

✓ $\hat{T}$에서 강력한 Descriptors를 도출하기 위해 Metric Learning과 Knowledge distillation을 도입하고, Loss term을 구함

[2] Bergmann, P., Fauser, M., Sattlegger, D., & Steger, C. (2020). Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 4183-4192).

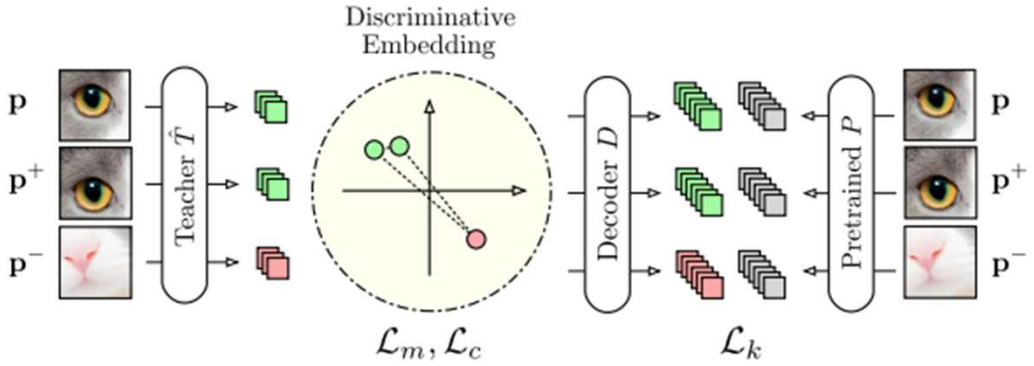
Knowledge Distillation for Anomaly Detection

Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings

❖ **Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings** [2]

➢ **Final training loss**

최종 손실 함수 $L(\hat{T})$ = Knowledge distillation loss + Metric learning loss + Compactness loss



$$L(\hat{T}) = \lambda_k L_k(\hat{T}) + \lambda_m L_m(\hat{T}) + \lambda_c L_c(\hat{T})$$

[2] Bergmann, P., Fauser, M., Sattlegger, D., & Steger, C. (2020). Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 4183-4192).

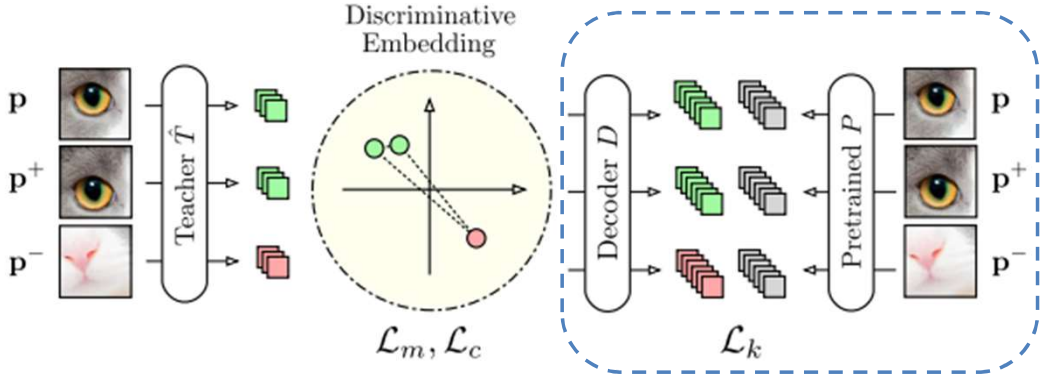
Knowledge Distillation for Anomaly Detection

Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings

❖ **Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings** [2]

➢ **Knowledge distillation**

$$L(\hat{T}) = \lambda_k L_k(\hat{T}) + \lambda_m L_m(\hat{T}) + \lambda_c L_c(\hat{T})$$



✓사전 학습된 P의 Knowledge를 추출하기 위해 P의 output과 $\hat{T}$로부터 얻은 Decoding된 Descriptor의 차이를 최소화 함

$$L_k(\hat{T}) = \left\| D\left(\hat{T}(p)\right) - P(p) \right\|^2$$

(D : $\hat{T}$ 의 d차원 출력을 Pretrained network의 Descriptor의 출력 차원으로 Decoding하는 Fully connected network)

[2] Bergmann, P., Fauser, M., Sattlegger, D., & Steger, C. (2020). Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 4183-4192).

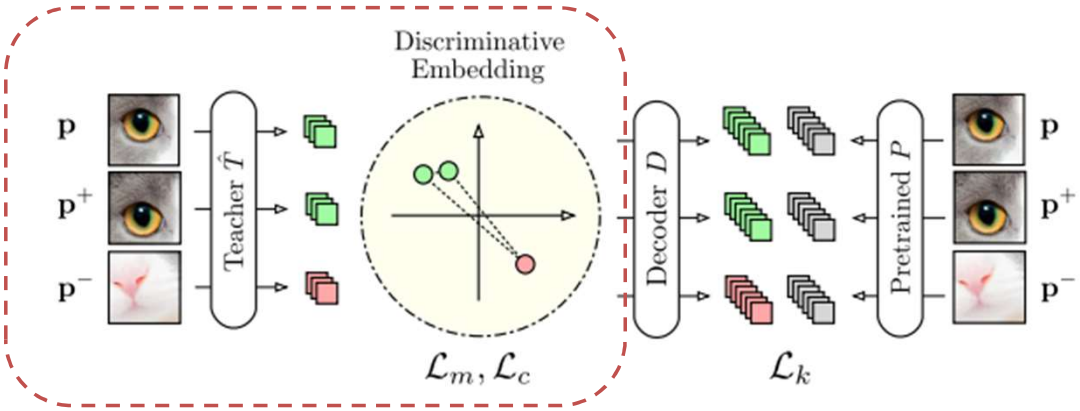
Knowledge Distillation for Anomaly Detection

Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings

❖ **Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings** [2]

➢ **Metric learning**

$$L(\hat{T}) = \lambda_k L_k(\hat{T}) + \lambda_m L_m(\hat{T}) + \lambda_c L_c(\hat{T})$$



✓ 사전 학습된 네트워크를 사용할 수 없는 경우 Self-supervised 방식으로 Local image descriptor를 학습하는 방법

✓ Triplet learning을 통해 얻은 Discriminative embedding을 사용

$$L_m(\hat{T}) = \max\{0, \delta + \delta^+ - \delta^-\}$$

δ : Margin parameter

$\delta^+ = \|\hat{T}(p) - \hat{T}(p^+)\|^2$

$\delta^- = \min\{\|\hat{T}(p) - \hat{T}(p^-)\|^2, \|\hat{T}(p^+) - \hat{T}(p^-)\|^2\}$

[2] Bergmann, P., Fauser, M., Sattlegger, D., & Steger, C. (2020). Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 4183-4192).

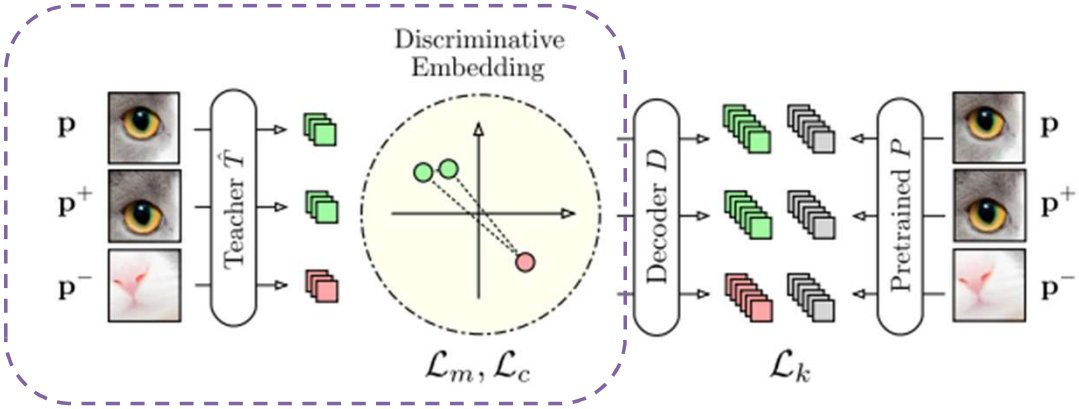
Knowledge Distillation for Anomaly Detection

Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings

❖ **Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings** [2]

➢ **Descripitor Compactness**

$$L(\hat{T}) = \lambda_k L_k(\hat{T}) + \lambda_m L_m(\hat{T}) + \lambda_c L_c(\hat{T})$$



✓Descriptor간의 상관관계를 최소화하여 간결성을 높이고 불필요한 중복성을 제거

$$L_c(\hat{T}) = \sum_{i \neq j} c_{ij}$$

(c_{ij} : 현재 미니 배치의 모든 Descriptor $\hat{T}(p)$에 대해 계산된 상관관계 행렬의 항목)

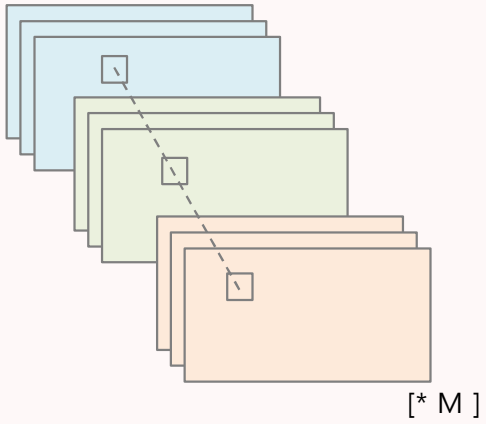
[2] Bergmann, P., Fauser, M., Sattlegger, D., & Steger, C. (2020). Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 4183-4192).

Knowledge Distillation for Anomaly Detection

Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings

❖ Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings [2]

➢ Training student network



Student network ensemble

✓ Teacher와 동일한 네트워크 구조를 갖는 무작위로 초기화된 M개의 Student networks $S_i(i = 1,2, \dots, M)$의 앙상블 구조

✓ 입력 이미지에 대한 로컬 이미지 영역에 대해 가능한 회귀 대상 공간에 대한 예측 분포를 출력함

✓ 크기 p의 제한된 Receptive field를 사용하여 patch를 자르지 않고도 단일 순방향 패스 만으로 밀도 높은 예측 가능

✓ Student network는 전부 Anomaly-free한 Training data를 학습함

✓ Student에서 예측된 값과 그에 상응하는 Teacher descriptor 차이를 최소화 하는 방향으로 학습됨

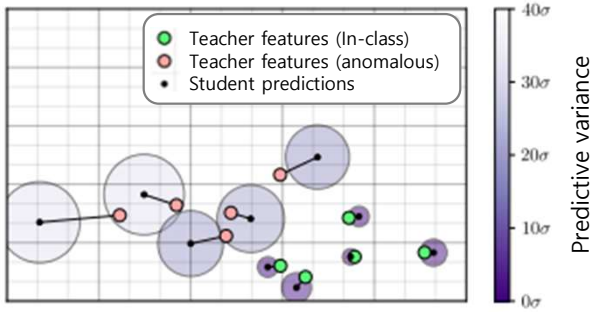
[2] Bergmann, P., Fauser, M., Sattlegger, D., & Steger, C. (2020). Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 4183-4192).

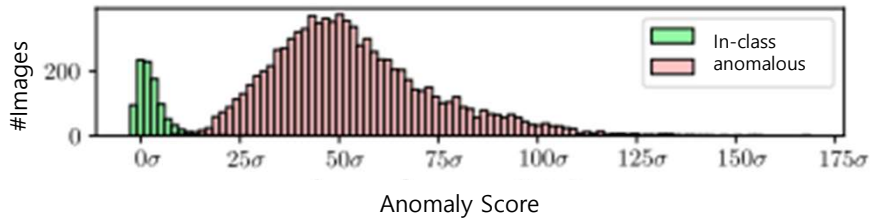
Knowledge Distillation for Anomaly Detection

Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings

❖ Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings [2]

➢ Scoring for Anomaly detection





✓ **Regression error**

$$e_{(r,c)} = \left\| \mu_{(r,c)} - \left(y^T_{(r,c)} - \mu\right) diag(\sigma)^{-1} \right\|_2^2$$

$$= \left\| \frac{1}{M} \sum_{i=1}^{M} \mu^{s_i}_{(r,c)} - \left(y^T_{(r,c)} - \mu\right) diag(\sigma)^{-1} \right\|_2^2$$

✓ **Predictive variance**

$$v_{(r,c)} = \frac{1}{M} \sum_{i=1}^{M} \left\| \mu^{s_i}_{(r,c)} \right\|_2^2 - \left\| \frac{1}{M} \sum_{i=1}^{M} \mu^{s_i}_{(r,c)} \right\|_2^2$$

✓ **Anomaly Score**

$$\tilde{e}_{(r,c)} + \tilde{v}_{(r,c)} = \frac{e_{(r,c)} - e_\mu}{e_\sigma} + \frac{v_{(r,c)} - v_\mu}{v_\sigma}$$

$\mu^{s_i}_{(r,c)}$: i번째 Student에서 만들어진 (r,c)픽셀 위치 예측값
$y^T_{(r,c)}$: (r,c)픽셀 위치 대응하는 Teacher descriptor
μ : 모든 Descriptor 평균값
σ : 모든 Descriptor 표준편차
e_μ, e_σ : 모든 $e_{(r,c)}$의 평균과 표준편차
v_μ, v_σ : 모든 $v_{(r,c)}$의 평균과 표준편차

[2] Bergmann, P., Fauser, M., Sattlegger, D., & Steger, C. (2020). Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 4183-4192).

Knowledge Distillation for Anomaly Detection

Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings

❖ **Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings** [2]

➢ **Experiment with MVTec Anomaly Detection Dataset**

	Category	Ours $p = 65$	1-NN	OC-SVM	K-Means	ℓ_2-AE	VAE	SSIM-AE	AnoGAN	CNN-Feature Dictionary
Textures	Carpet	**0.695**	0.512	0.355	0.253	0.456	0.501	0.647	0.204	0.469
	Grid	0.819	0.228	0.125	0.107	0.582	0.224	**0.849**	0.226	0.183
	Leather	**0.819**	0.446	0.306	0.308	**0.819**	0.635	0.561	0.378	0.641
	Tile	**0.912**	0.822	0.722	0.779	0.897	0.870	0.175	0.177	0.797
	Wood	0.725	0.502	0.336	0.411	**0.727**	0.628	0.605	0.386	0.621
Objects	Bottle	**0.918**	0.898	0.850	0.495	0.910	0.897	0.834	0.620	0.742
	Cable	**0.865**	0.806	0.431	0.513	0.825	0.654	0.478	0.383	0.558
	Capsule	**0.916**	0.631	0.554	0.387	0.862	0.526	0.860	0.306	0.306
	Hazelnut	**0.937**	0.861	0.616	0.698	0.917	0.878	0.916	0.698	0.844
	Metal nut	**0.895**	0.705	0.319	0.351	0.830	0.576	0.603	0.320	0.358
	Pill	**0.935**	0.725	0.544	0.514	0.893	0.769	0.830	0.776	0.460
	Screw	**0.928**	0.604	0.644	0.550	0.754	0.559	0.887	0.466	0.277
	Toothbrush	**0.863**	0.675	0.538	0.337	0.822	0.693	0.784	0.749	0.151
	Transistor	0.701	0.680	0.496	0.399	**0.728**	0.626	0.725	0.549	0.628
	Zipper	**0.933**	0.512	0.355	0.253	0.839	0.549	0.665	0.467	0.703
	Mean	**0.857**	0.640	0.479	0.423	0.790	0.639	0.694	0.443	0.515

성능 지표 : Area Under PRO Curve

[2] Bergmann, P., Fauser, M., Sattlegger, D., & Steger, C. (2020). Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 4183-4192).

Knowledge Distillation for Anomaly Detection

Multiresolution knowledge distillation for anomaly detection

❖ **Multiresolution knowledge distillation for anomaly detection** [3]

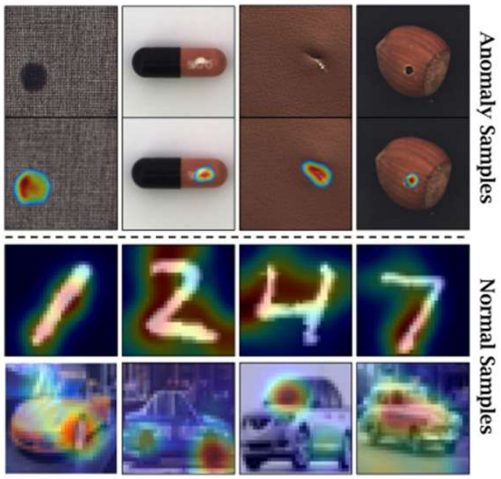
- 2021 CVPR에 게재된 논문 (2023년 10월 30일 기준 238회 인용)
- Knowledge distillation 기반 Anomaly Detection에서 마지막 Layer 뿐만 아니라 중간 Layer의 Feature도 이용

Multiresolution Knowledge Distillation for Anomaly Detection

Mohammadreza Salehi, Niousha Sadjadi*, Soroosh Baselizadeh*, Mohammad Hossein Rohban, Hamid R. Rabiee
Sharif University of Technology
(smrsalehi, nsadjadi, baselizadeh)@ce.sharif.edu, (rohban, rabiee)@sharif.edu

Abstract

Unsupervised representation learning has proved to be a critical component of anomaly detection/localization in images. The challenges to learn such a representation are two-fold. Firstly, the sample size is not often large enough to learn a rich generalizable representation through conventional techniques. Secondly, while only normal samples are available at training, the learned features should be discriminative of normal and anomalous samples. Here, we propose to use the "distillation" of features at various layers of an expert network, pre-trained on ImageNet, into a simpler cloner network to tackle both issues. We detect and localize anomalies using the discrepancy between the expert and cloner networks' intermediate activation values given the input data. We show that considering multiple intermediate hints in distillation leads to better exploiting the expert's knowledge and more distinctive discrepancy compared to solely utilizing the last layer acti-



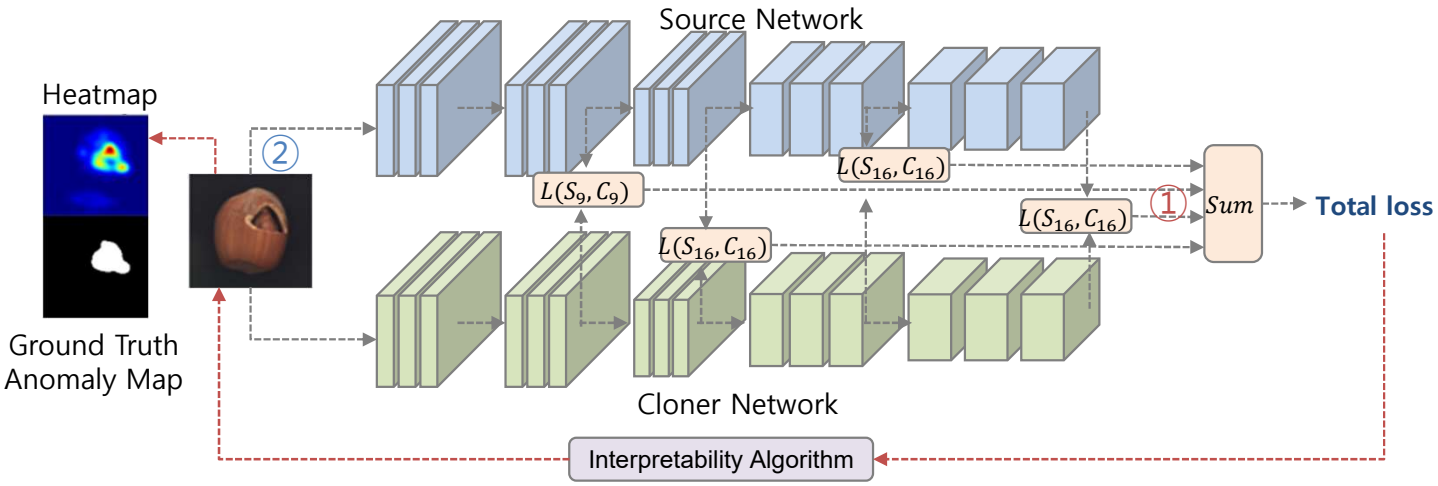
[3] Salehi, M., Sadjadi, N., Baselizadeh, S., Rohban, M. H., & Rabiee, H. R. (2021). Multiresolution knowledge distillation for anomaly detection. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 14902-14912).

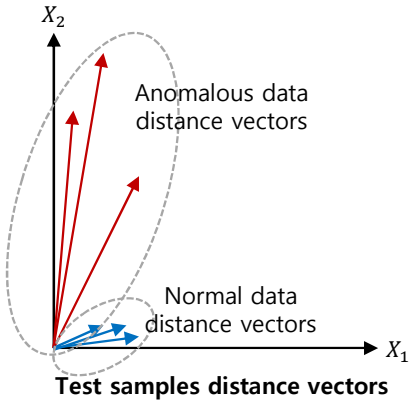
Knowledge Distillation for Anomaly Detection

Multiresolution knowledge distillation for anomaly detection

❖ **Multiresolution knowledge distillation for anomaly detection** [3]

➢ **아이디어 제안**





✓ **이전 Uninformed students 논문에서의 방법론에 문제 제기**

① 마지막 Layer 지식만 전달 → Local Minimum 발생 → Intermediate Layer 사용

② Patch 크기 제한 → Localization과 Detection 성능 저하 → 이미지 단위로 학습

✓ **Anomaly : Teacher-Student간 최대화**

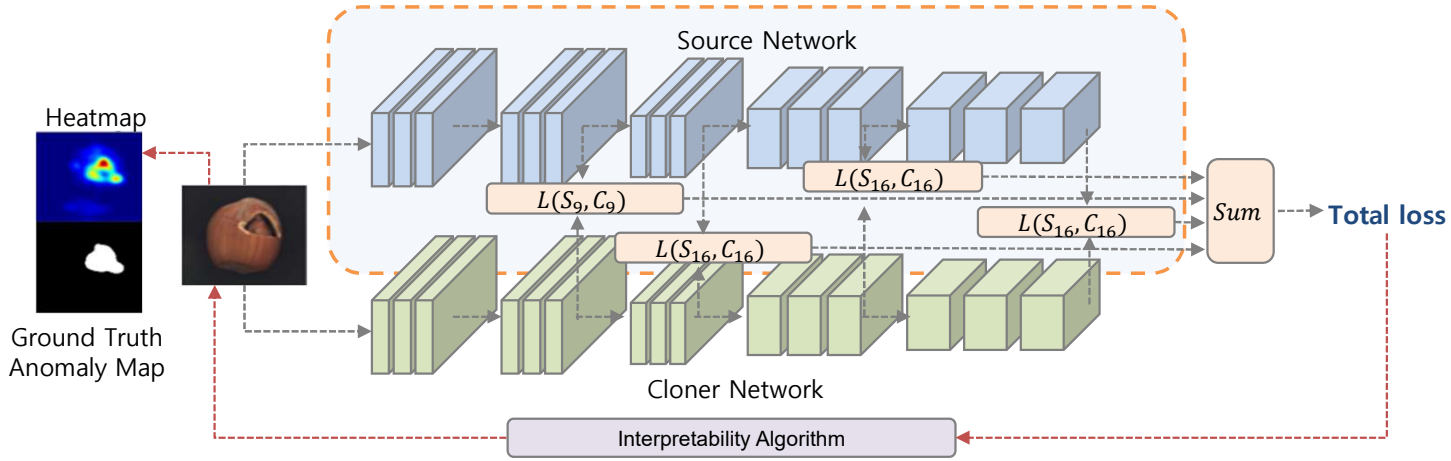
[3] Salehi, M., Sadjadi, N., Baselizadeh, S., Rohban, M. H., & Rabiee, H. R. (2021). Multiresolution knowledge distillation for anomaly detection. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 14902-14912).

Knowledge Distillation for Anomaly Detection

Multiresolution knowledge distillation for anomaly detection

❖ **Multiresolution knowledge distillation for anomaly detection** [3]

➢ **Teacher(Source) Network**



✓ Source Network의 중간 Layer의 knowledge도 Cloner Network로 전달함 (Intermediate feature 사용)

✓ Intermediate Feature들을 Cloner Network의 각 Layer에서 추출된 Feature와 일치시킴으로써 정상 샘플을 학습하게 함

✓ 서로 다른 Layer를 모방하여 다양한 level에서 Cloner Network를 학습 가능

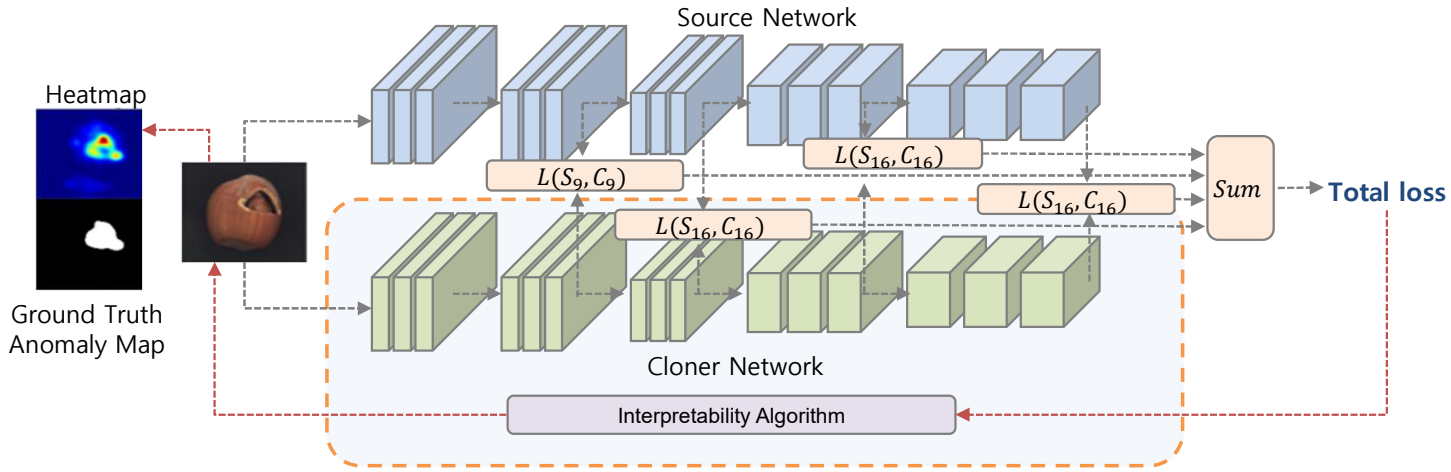
[3] Salehi, M., Sadjadi, N., Baselizadeh, S., Rohban, M. H., & Rabiee, H. R. (2021). Multiresolution knowledge distillation for anomaly detection. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 14902-14912).

Knowledge Distillation for Anomaly Detection

Multiresolution knowledge distillation for anomaly detection

❖ **Multiresolution knowledge distillation for anomaly detection** [3]

➢ **Student(Cloner) Network**



✓ 정상 데이터로만 학습하며, Source Network의 포괄적인 거동을 모방하도록 같은 구조로 설계

✓ Intermediate feature 사용

✓ Cloner network의 구조는 Knowledge distillation을 위해 Source network보다 단순하게 설계됨

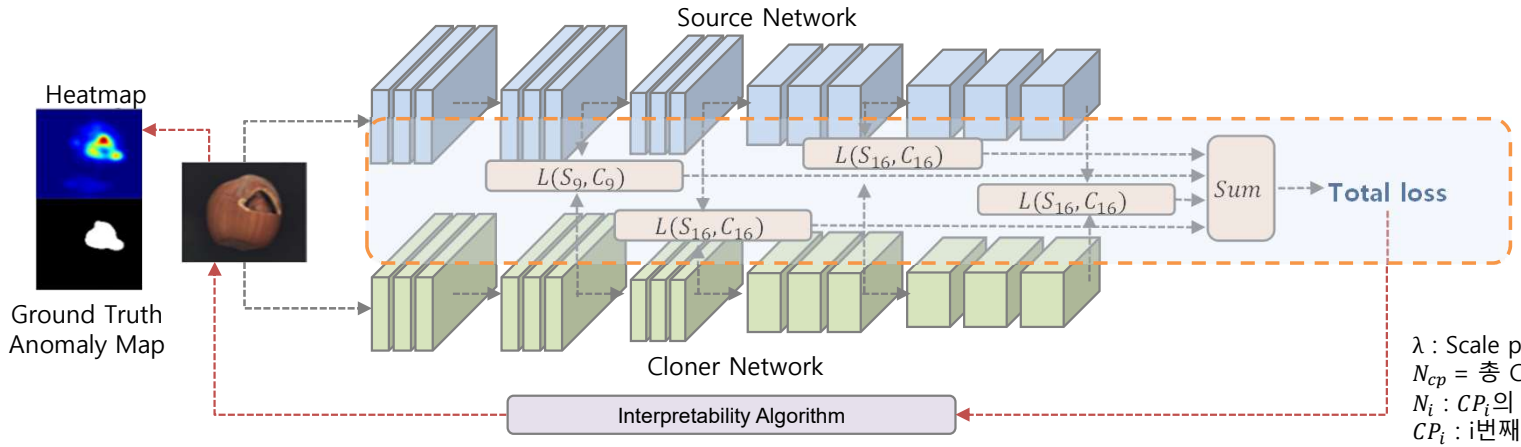
[3] Salehi, M., Sadjadi, N., Baselizadeh, S., Rohban, M. H., & Rabiee, H. R. (2021). Multiresolution knowledge distillation for anomaly detection. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 14902-14912).

Knowledge Distillation for Anomaly Detection

Multiresolution knowledge distillation for anomaly detection

❖ **Multiresolution knowledge distillation for anomaly detection** [3]

➢ **Total loss**



λ : Scale parameter
N_{cp} = 총 Critical layer 개수
N_i : CP_i의 뉴런수
CP_i : i번째 중요 계층
$a_s^{CP_i}(j)$: CP_i 의 j번째 Source 활성화값
$a_c^{CP_i}(j)$: CP_i 의 j번째 Cloner 활성화값

✓ 손실함수를 최소화하여 Cloner network가 Source network와 유사한 Intermediate feature를 생성하도록 함

$$L_{val} = \sum_{i=1}^{N_{cp}} \frac{1}{N_i} \sum_{j=1}^{N_i} (a_s^{CP_i}(j) - a_c^{CP_i}(j))^2$$

$$L_{total} = L_{val} + \lambda L_{dir}$$

$$L_{dir} = 1 - \sum_i \frac{vec(a_s^{CP_i})^T \cdot vec(a_c^{CP_i})}{\|vec(a_s^{CP_i})\| \|vec(a_c^{CP_i})\|}$$

[3] Salehi, M., Sadjadi, N., Baselizadeh, S., Rohban, M. H., & Rabiee, H. R. (2021). Multiresolution knowledge distillation for anomaly detection. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 14902-14912).

Knowledge Distillation for Anomaly Detection

Multiresolution knowledge distillation for anomaly detection

❖ Multiresolution knowledge distillation for anomaly detection [3]

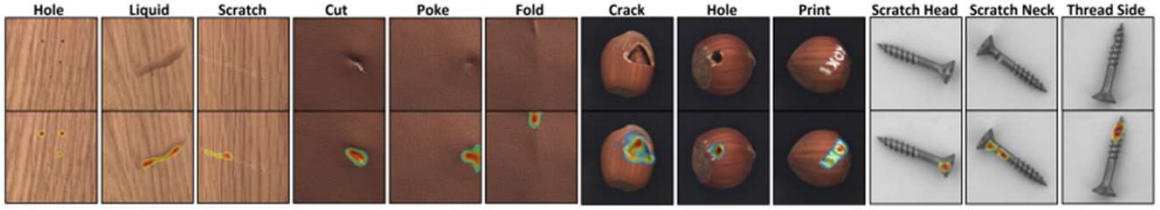
➢ Anomaly Detection

✓ 비정상 샘플을 탐지를 위해 테스트 샘플을 Source/Cloner network에 제공

✓ 정상샘플만 학습했기에 Cloner network는 불일치를 만들고, Total loss가 임계 값을 넘어가게 됨

임계 값 > Total loss : 비정상, 임계 값 < Total loss : 정상

➢ Anomaly Localization



✓ 입력 값에 대한 Total loss의 미분 값이 픽셀의 중요도를 제공 → 기울기를 증가시키는 비정상 영역을 찾음

✓ 먼저 Attribution map ∧을 얻음 → 노이즈 제거를 위해 Gaussian 필터 적용 후 Localization map L_{map} 을 얻음

$$\wedge = \frac{\partial L_{total}}{\partial x} \qquad M = g_\sigma \wedge \qquad L_{map} = (M \ominus B) \oplus B$$

g : 표준편차의 Gaussian 필터
B : 이진 map(구조화요소)

[3] Salehi, M., Sadjadi, N., Baselizadeh, S., Rohban, M. H., & Rabiee, H. R. (2021). Multiresolution knowledge distillation for anomaly detection. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 14902-14912).

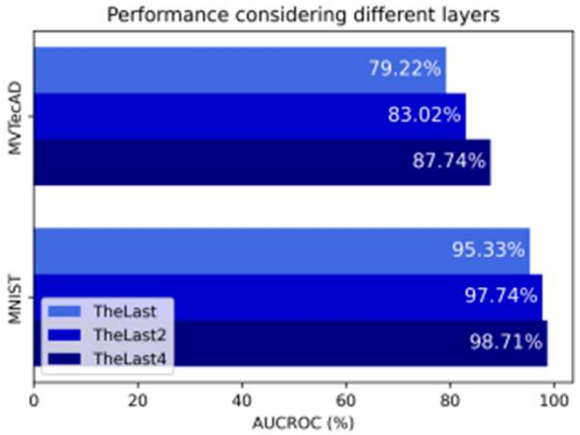
Knowledge Distillation for Anomaly Detection

Multiresolution knowledge distillation for anomaly detection

❖ **Multiresolution knowledge distillation for anomaly detection** [3]

➢ **Ablation Studies**

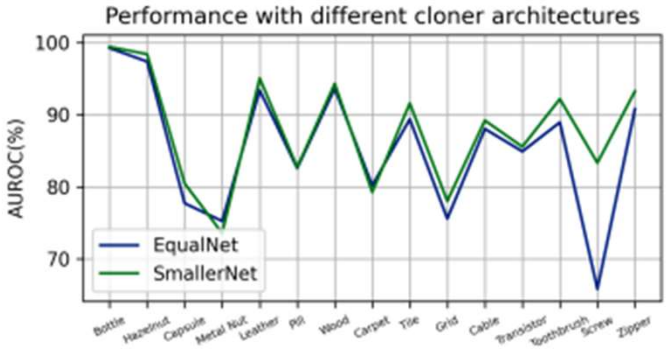
① Intermediate Knowledge



✓ Intermediate layer 개수에 따른 성능 변화

→ 개수가 많아질수록 성능이 더 높음

② Distillation Effect(Compact C)



✓ Cloner network 크기에 따른 Distillation 효과

→ 작은 크기의 Network에서 성능이 더 높음

[3] Salehi, M., Sadjadi, N., Baselizadeh, S., Rohban, M. H., & Rabiee, H. R. (2021). Multiresolution knowledge distillation for anomaly detection. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 14902-14912).

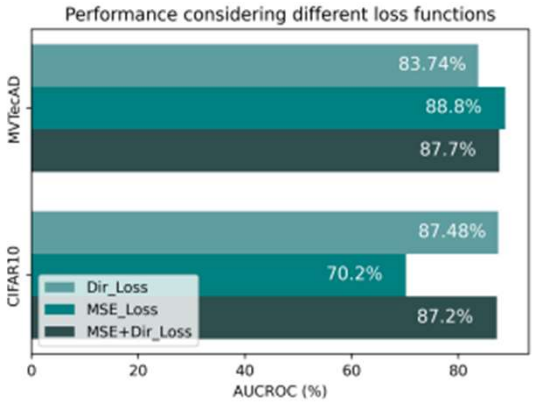
Knowledge Distillation for Anomaly Detection

Multiresolution knowledge distillation for anomaly detection

❖ **Multiresolution knowledge distillation for anomaly detection** [3]

➢ **Ablation Studies**

③ L_{dir}, L_{val}



✓ 손실 함수 구성 요소의 효과

→ 다양한 손실함수를 사용하는 것이 성능이 더 높음

(***MSE loss***(L_{val})만 이용하는 경우는 편차가 큼)

④ **Localization using Interpretability Methods**

Method	Gradients	SmoothGrad	GBP
Without Gaussian Filter	86.16%	86.97%	84.38%
With Gaussian Filter	90.51%	90.54%	90.08%

✓ Localization을 위한 해석 가능한 방법

→ 잘못 계산된 그라데이션 픽셀을 버리는 SmoothGrad와 간단한 Gradients 성능이 더 좋음

[3] Salehi, M., Sadjadi, N., Baselizadeh, S., Rohban, M. H., & Rabiee, H. R. (2021). Multiresolution knowledge distillation for anomaly detection. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 14902-14912).

Knowledge Distillation for Anomaly Detection

Multiresolution knowledge distillation for anomaly detection

❖ **Multiresolution knowledge distillation for anomaly detection** [3]

➢ **Experiments**

Dataset	Method	0	1	2	3	4	5	6	7	8	9	Mean
MNIST[24]	ARAE[38]	99.8	99.9	96.0	97.2	97.0	97.4	99.5	96.9	92.4	98.5	97.5
	OCSVM[14]	99.5	99.9	92.6	93.6	96.7	95.5	98.7	96.6	90.3	96.2	96.0
	AnoGAN[41]	96.6	99.2	85.0	88.7	89.4	88.3	94.7	93.5	84.9	92.4	91.3
	DSVDD[33]	98.0	99.7	91.7	91.9	94.9	88.5	98.3	94.6	93.9	96.5	94.8
	CapsNet$_{PP}$ [25]	99.8	99.0	98.4	97.6	93.5	97.0	94.2	98.7	99.3	99.0	97.7
	OCGAN[31]	99.8	99.9	94.2	96.3	97.5	98.0	99.1	98.1	93.9	98.1	97.5
	LSA[1]	99.3	99.9	95.9	96.6	95.6	96.4	99.4	98.0	95.3	98.1	97.5
	CAVGA-D$_u$[47]	99.4	99.7	98.9	98.3	97.7	96.8	98.8	98.6	98.8	99.1	98.6
	U-Std[9]	99.9	99.9	99	99.3	99.2	99.3	99.7	99.5	98.6	99.1	**99.35**
	OURS	99.82 ± 0.023	99.82 ± 0.017	97.79 ± 0.272	98.75 ± 0.098	98.43 ± 0.096	98.16 ± 0.182	99.43 ± 0.038	98.38 ± 0.178	98.41 ± 0.157	98.1 ± 0.152	98.71
Fashion-MNIST[49]	ARAE[38]	93.7	99.1	91.1	94.4	92.3	91.4	83.6	98.9	93.9	97.9	93.6
	OCSVM[14]	91.9	99.0	89.4	94.2	90.7	91.8	83.4	98.8	90.3	98.2	92.8
	DAGMM[59]	30.3	31.1	47.5	48.1	49.9	41.3	42.0	37.4	51.8	37.8	41.7
	DSEBM[51]	89.1	56.0	86.1	90.3	88.4	85.9	78.2	98.1	86.5	96.7	85.5
	DSVDD[33]	98.2	90.3	90.7	94.2	89.4	91.8	83.4	98.8	91.9	99.0	92.8
	LSA[1]	91.6	98.3	87.8	92.3	89.7	90.7	84.1	97.7	91.0	98.4	92.2
	OURS	92.5 ± 0.298	99.21 ± 0.064	92.48 ± 0.255	93.8 ± 0.095	92.95 ± 0.159	98.21 ± 0.157	84.87 ± 0.126	99.02 ± 0.331	94.33 ± 0.164	97.51 ± 0.055	**94.49**
CIFAR-10[23]	ARAE[38]	72.2	43.1	69.0	55.0	75.2	54.7	70.1	51.0	72.2	40.0	60.23
	OCSVM[14]	63.0	44.0	64.9	48.7	73.5	50.0	72.5	53.3	64.9	50.8	58.56
	AnoGAN[41]	67.1	54.7	52.9	54.5	65.1	60.3	58.5	62.5	75.8	66.5	61.79
	DSVDD[33]	61.7	65.9	50.8	59.1	60.9	65.7	67.7	67.3	75.9	73.1	64.81
	CapsNet$_{PP}$[25]	62.2	45.5	67.1	67.5	68.3	63.5	72.7	67.3	71.0	46.6	61.2
	OCGAN[31]	75.7	53.1	64.0	62.0	72.3	62.0	72.3	57.5	82.0	55.4	65.66
	LSA[1]	73.5	58.0	69.0	54.2	76.1	54.6	75.1	53.5	71.7	54.8	64.1
	DROCC[19]	81.66	76.74	66.66	67.13	73.62	74.43	74.43	71.39	80.02	76.21	74.23
	CAVGA-D$_u$[47]	65.3	78.4	76.1	74.7	77.5	55.2	81.3	74.5	80.1	74.1	73.7
	GT[18]	76.2	84.8	77.1	73.2	82.8	84.8	82	88.7	89.5	83.4	82.3
	U-Std[9]	78.9	84.9	73.4	74.8	85.1	79.3	89.2	83	86.2	84.8	81.96
	OURS	90.53 ± 0.158	90.35 ± 0.797	79.66 ± 0.415	77.02 ± 0.51	86.71 ± 0.346	91.4 ± 0.279	88.98 ± 0.2	86.78 ± 0.595	91.45 ± 0.148	88.91 ± 0.349	**87.18**

✓ 제시한 모델이 Fashion-MNIST, CIFAR-10에서는 가장 높은 성능을, MNIST에서도 우수한 성능을 보임

[3] Salehi, M., Sadjadi, N., Baselizadeh, S., Rohban, M. H., & Rabiee, H. R. (2021). Multiresolution knowledge distillation for anomaly detection. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 14902-14912).

Summary

Summary

Knowledge distillation for anomaly detection

❖ Knowledge distillation for anomaly detection

Anomaly Detection과 Knowledge distillation 개념을 설명하고,
Teacher-Student network를 적용하여 Anomaly Detection을 Regression 관점에서 바라본 연구들을 소개함.

✓ Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings

- Knowledge distillation 기반 Anomaly Detection을 처음 제안한 논문
- 다양한 Receptive field를 갖는 Multiple student ensemble 도입
- Triplet learning을 통해 얻은 Discriminative embedding을 사용

✓ Multiresolution knowledge distillation for anomaly detection

- Intermediate feature를 사용함으로써 더 많은 Knowledge를 distillation함
- Localization과 Detection 성능 향상을 위해 이미지 그대로 사용

고맙습니다.