

# SimCLR, 외관검사 트렌드

김동원

# 목차

SimCLR

외관검사, 컴퓨터 비전 트렌드

## A Simple Framework for Contrastive Learning of Visual Representations

이 논문에서는 이미지의 표현을 학습하는 간단한 contrastive learning 프레임워크를 제시한다.

이미지의 표현을 학습하기 위한 두가지 방법

Generative

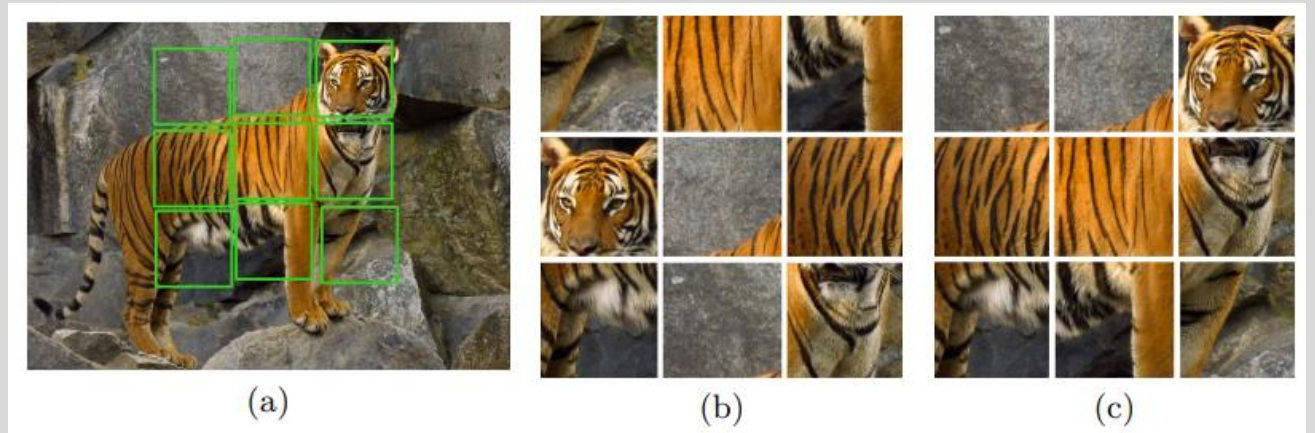
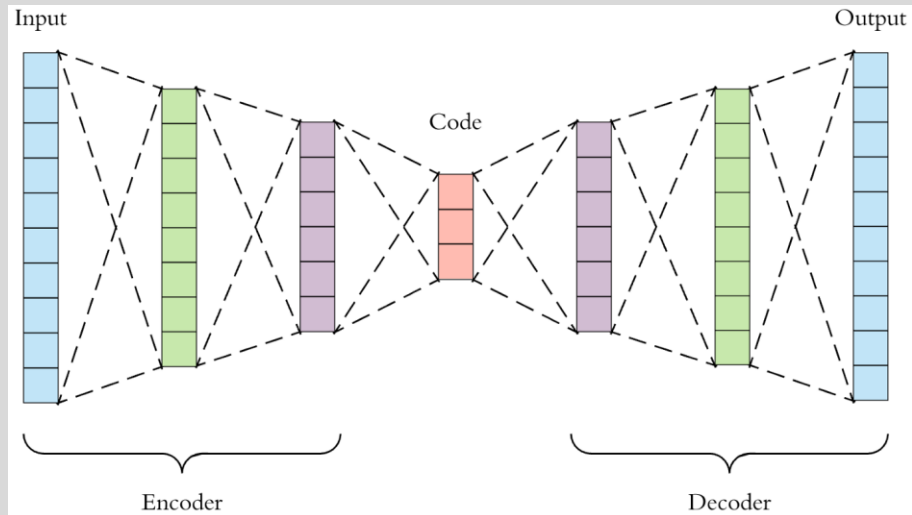
데이터 생성 기반의 방법으로 시간과 연산에 있어 비용이 크다

E.g) Variational auto-encoder

discriminative

지도학습과 유사하지만 라벨이 없는 데이터를 사용하여 pretext task를 진행하여 학습하는 방법

E.g) Self-supervised learning



# SimCLR

## 모델 구조

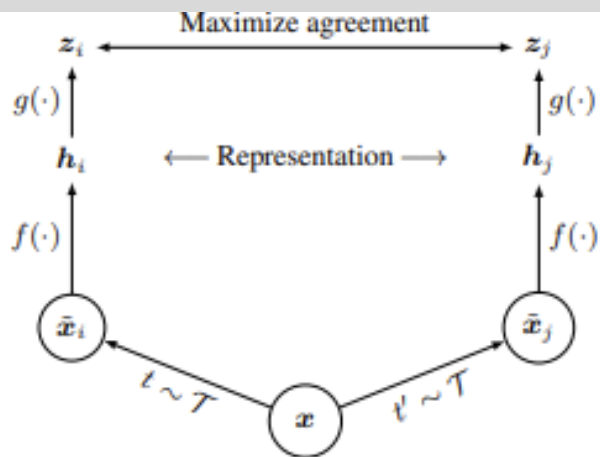


Figure 2. A simple framework for contrastive learning of visual representations. Two separate data augmentation operators are sampled from the same family of augmentations ( $t \sim \mathcal{T}$  and  $t' \sim \mathcal{T}$ ) and applied to each data example to obtain two correlated views. A base encoder network  $f(\cdot)$  and a projection head  $g(\cdot)$  are trained to maximize agreement using a contrastive loss. After training is completed, we throw away the projection head  $g(\cdot)$  and use encoder  $f(\cdot)$  and representation  $h$  for downstream tasks.

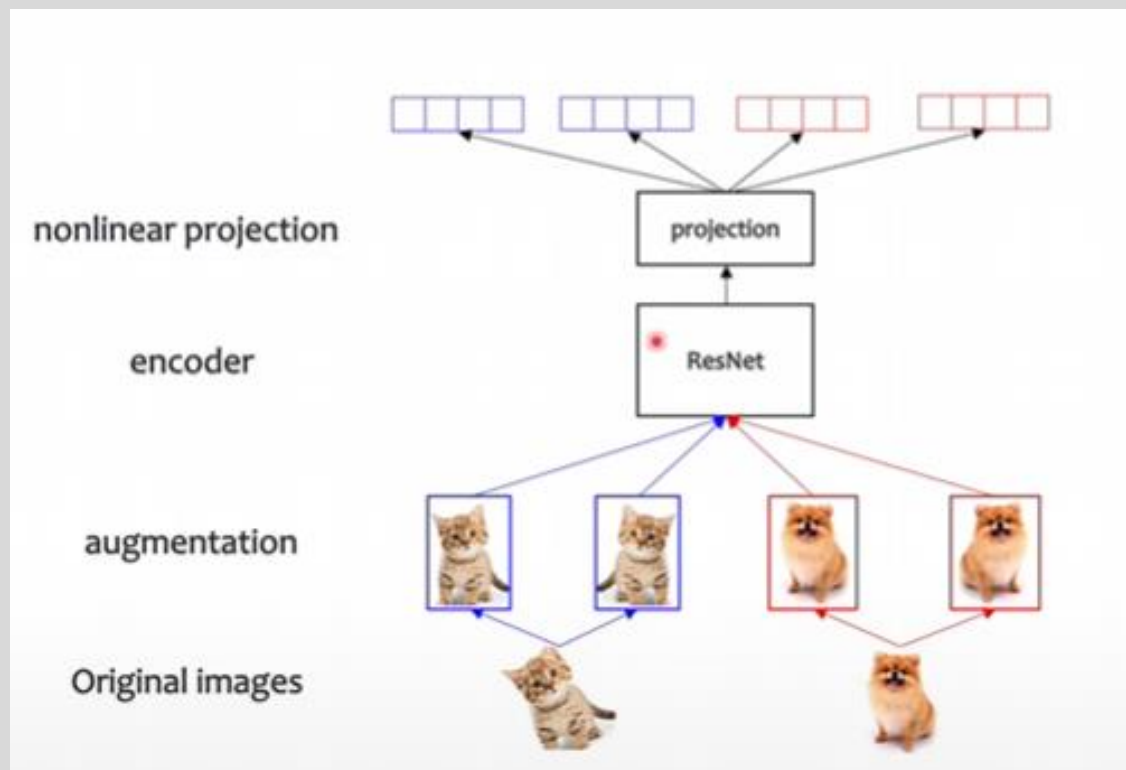
X라는 이미지를 확률적으로 다르게 변형시킨  $x_i$ 과  $x_j$ 에 대해  $f(\cdot)$ 라는 encoder와 같은 구조를 통해 표현을 학습시키고  $g(\cdot)$ 라는 MLP을 통해 비선형적 변환된 것을 비교한다.

같은  $x$ 에서 변형된  $x_i$ 와  $x_j$ 를 positive pair라고 하고 서로 다른 이미지에서 변형된 이미지를 negative pair라고 하는데 positive pair는 가깝게, negative pair는 멀게 학습시킨다.

N개의 이미지를 샘플링할 경우 이미지 변형을 통해 2N개의 데이터셋이 되고 2(N-1)개의 negative pair를 가지게 된다.

# SimCLR

모델 구조



# SimCLR

모델 구조

$$\ell_{i,j} = -\log \frac{\exp(\text{sim}(\mathbf{z}_i, \mathbf{z}_j)/\tau)}{\sum_{k=1}^{2N} \mathbb{1}_{[k \neq i]} \exp(\text{sim}(\mathbf{z}_i, \mathbf{z}_k)/\tau)}$$

$$\text{sim}(\mathbf{u}, \mathbf{v}) = \mathbf{u}^\top \mathbf{v} / \|\mathbf{u}\| \|\mathbf{v}\|$$

Positive pair과 negative pair의 유사도를 계산하여 loss를 구한다 NT Xent라고 부른다

Sim()은 코사인 유사도와 계산방법이 같고,  $\tau$ 는 temperature parameter로 정규화를 시켜주고 보통 0.1에서 1.0사이의 값이 사용된다

# SimCLR

Data augmentation

이미지 변형 방법으로 random crop, color distortion, gaussian blur를 사용하였다.

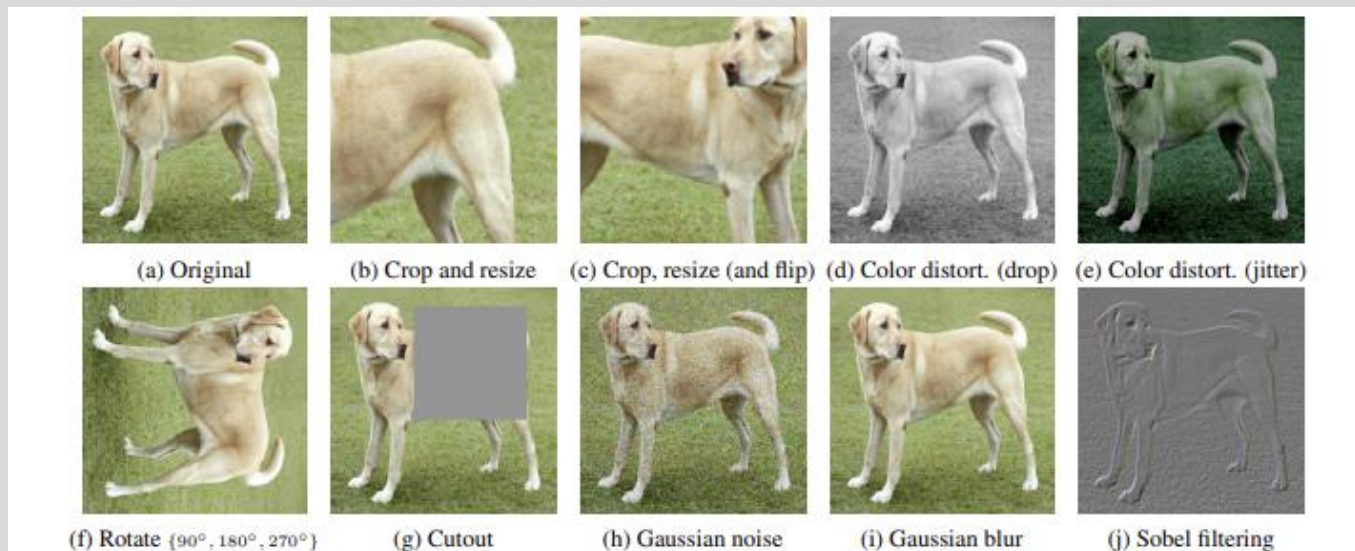


Figure 4. Illustrations of the studied data augmentation operators. Each augmentation can transform data stochastically with some internal parameters (e.g. rotation degree, noise level). Note that we *only* test these operators in ablation, the *augmentation policy* used to train our models only includes *random crop* (with *flip* and *resize*), *color distortion*, and *Gaussian blur*. (Original image cc-by: Von.grzanka)

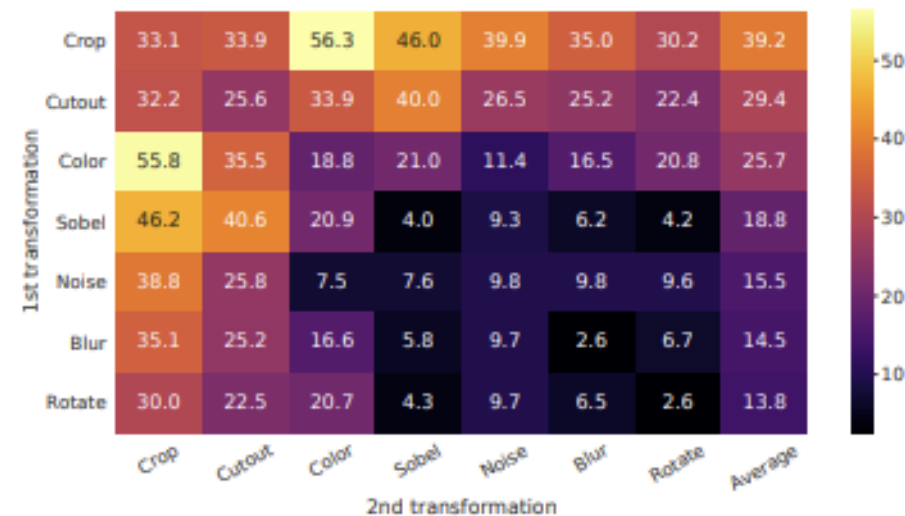
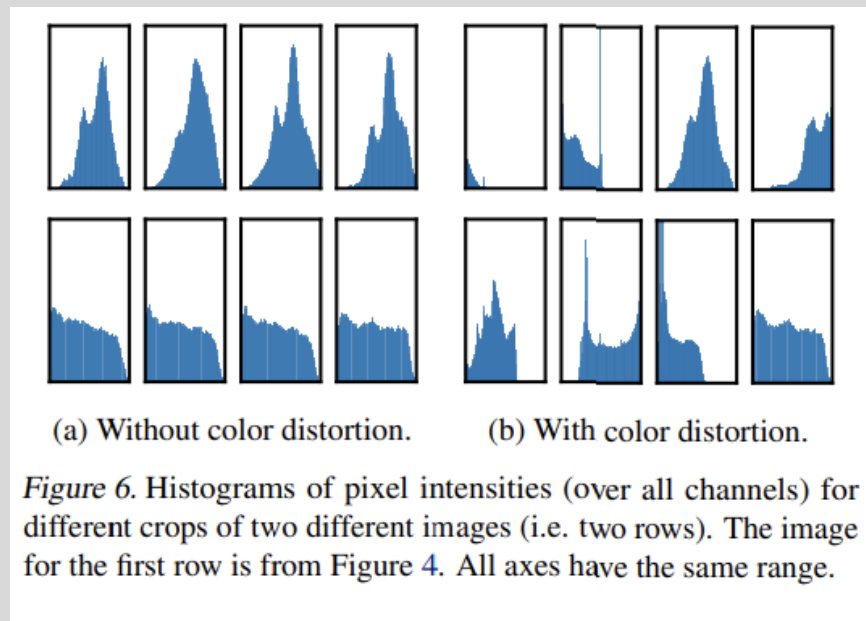
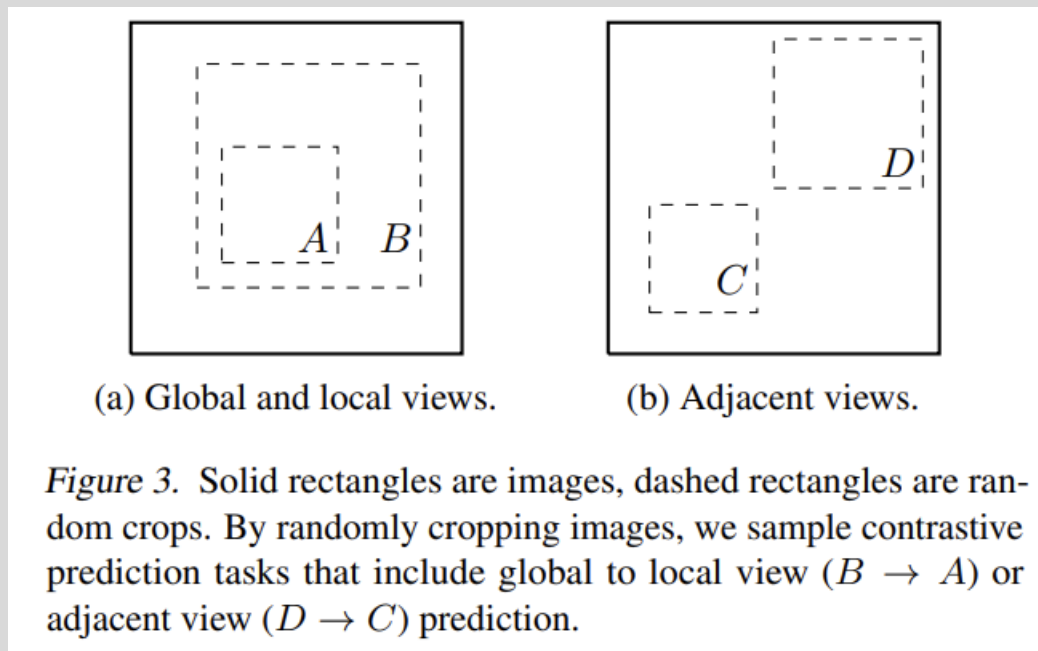


Figure 5. Linear evaluation (ImageNet top-1 accuracy) under individual or composition of data augmentations, applied only to one branch. For all columns but the last, diagonal entries correspond to single transformation, and off-diagonals correspond to composition of two transformations (applied sequentially). The last column reflects the average over the row.

# SimCLR

Data augmentation



이미지를 crop한 경우 색상에 의존하여 추론하게 되는 경우가 있기 때문에 색상을 조정하는 변형을 같이 사용할 경우 성능이 좋아지게 된다.



# SimCLR

Data augmentation

지도 학습과는 다르게 color distortion을 심하게 할 수록 좋은 성능을 보였다

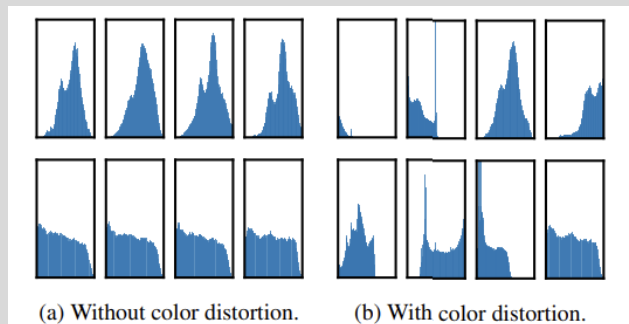


Figure 6. Histograms of pixel intensities (over all channels) for different crops of two different images (i.e. two rows). The image for the first row is from Figure 4. All axes have the same range.

Methods	Color distortion strength					AutoAug
	1/8	1/4	1/2	1	1 (+Blur)	
SimCLR	59.6	61.0	62.6	63.2	64.5	61.1
Supervised	77.0	76.7	76.5	75.7	75.4	77.1

Table 1. Top-1 accuracy of unsupervised ResNet-50 using linear evaluation and supervised ResNet-50<sup>5</sup>, under varied color distortion strength (see Appendix A) and other data transformations. Strength 1 (+Blur) is our default data augmentation policy.

Name	Negative loss function	Gradient w.r.t. $\mathbf{u}$
NT-Xent	$\mathbf{u}^T \mathbf{v}^+ / \tau - \log \sum_{\mathbf{v} \in \{\mathbf{v}^+, \mathbf{v}^-\}} \exp(\mathbf{u}^T \mathbf{v} / \tau)$	$(1 - \frac{\exp(\mathbf{u}^T \mathbf{v}^+ / \tau)}{Z(\mathbf{u})}) / \tau \mathbf{v}^+ - \sum_{\mathbf{v}^-} \frac{\exp(\mathbf{u}^T \mathbf{v}^- / \tau)}{Z(\mathbf{u})} / \tau \mathbf{v}^-$
NT-Logistic	$\log \sigma(\mathbf{u}^T \mathbf{v}^+ / \tau) + \log \sigma(-\mathbf{u}^T \mathbf{v}^- / \tau)$	$(\sigma(-\mathbf{u}^T \mathbf{v}^+ / \tau)) / \tau \mathbf{v}^+ - \sigma(\mathbf{u}^T \mathbf{v}^- / \tau) / \tau \mathbf{v}^-$
Margin Triplet	$-\max(\mathbf{u}^T \mathbf{v}^- - \mathbf{u}^T \mathbf{v}^+ + m, 0)$	$\mathbf{v}^+ - \mathbf{v}^-$ if $\mathbf{u}^T \mathbf{v}^+ - \mathbf{u}^T \mathbf{v}^- < m$ else $\mathbf{0}$

Table 2. Negative loss functions and their gradients. All input vectors, i.e.  $\mathbf{u}$ ,  $\mathbf{v}^+$ ,  $\mathbf{v}^-$ , are  $\ell_2$  normalized. NT-Xent is an abbreviation for “Normalized Temperature-scaled Cross Entropy”. Different loss functions impose different weightings of positive and negative examples.

다른 loss값을 취했을 때의 성능을 보면 NT Xent가 가장 성능이 좋다

Margin	NT-Logi.	Margin (sh)	NT-Logi.(sh)	NT-Xent
50.9	51.6	57.5	57.9	63.9

Table 4. Linear evaluation (top-1) for models trained with different loss functions. “sh” means using semi-hard negative mining.

# SimCLR

## SimCLR특징

큰 batch와 긴 학습을 시키는 것이 성능에 좋은 영향을 끼친다

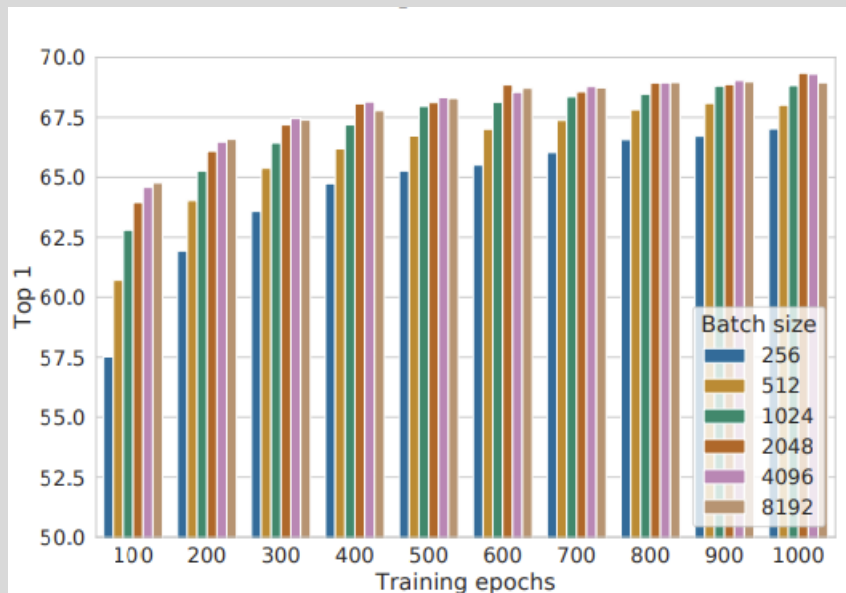


Figure 9. Linear evaluation models (ResNet-50) trained with different batch size and epochs. Each bar is a single run from scratch.<sup>10</sup>

모델의  $g()$ 에서 비 선형적인 MLP를 사용하는 것이 선형적 변환방법 보다 성능이 좋다

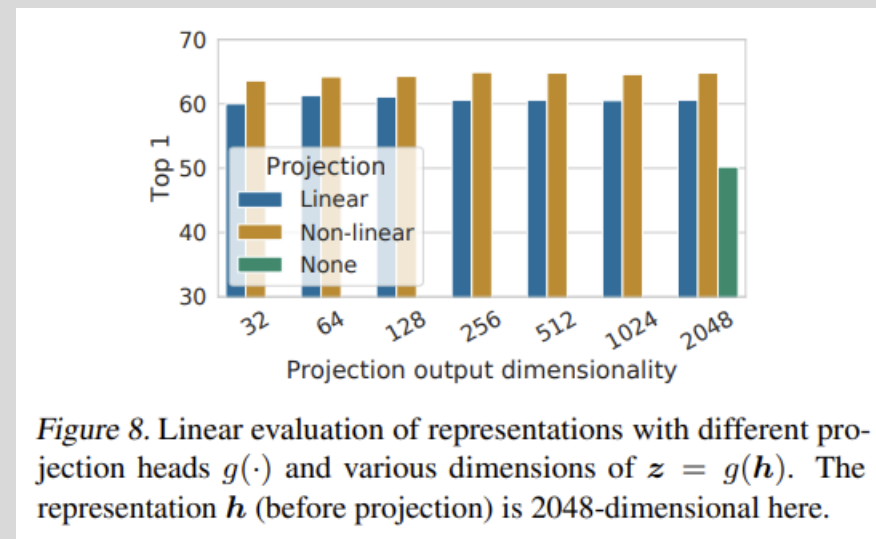


Figure 8. Linear evaluation of representations with different projection heads  $g(\cdot)$  and various dimensions of  $z = g(h)$ . The representation  $h$  (before projection) is 2048-dimensional here.

# SimCLR

학습환경, 파라미터 설정방법

학습은 32~128코어의 TPU환경에서 진행하였고, LARS라는 optimizer, weight decay, learning rate decay, warm up을 사용하고 4096사이즈의 배치로 100epoche 시행

TPU환경에서 데이터를 병렬적으로 학습시키는 경우 각 장치별로 배치 정규화를 진행하면, positive pair는 같은 장치에서 연산 되므로 모델의 학습과 상관없이 성능이 좋아지므로 모든 장치에 대해 배치 정규화를 실시한다

NT Xent와 많이 사용되는 LARS optimizer를 사용하는데 이는 큰 배치에서 안정적인 학습이 가능하고 레이어 마다 다른 학습률을 적용하여 모델의 각 레이어가 효과적으로 최적화 되기 때문에 사용된다고 한다.

# SimCLR

## 요약

1. Random crop와 color distortion은 상호보완적으로 사용하여 성능을 증가시켰고,
2. 이미지의 특징 벡터를 비선형적 벡터로 임베딩한 후 학습을 진행하였을 경우 성능이 증가하였다.
3. 지도학습보다 큰 배치사이즈와 긴 학습시에도 지속적인 성능향상을 보여줬다.

Contrastive learning의 의미와 self-supervised learning 원리를 알 수 있었다.

# 외관검사 컴퓨터 비전 트렌드

여러 분야에서 시행되고 단순 반복작업, 숙련된 노동자가 필요한 작업이고, 결과가 일정하지 못하다

95%이상이 사람이 눈으로 검사, 공장 자동화(FA)의 최후의 보루

제조업 종사자의 약 1/5

단순작업, 중노동, 숙련작업, 결과가 일정하지 않음



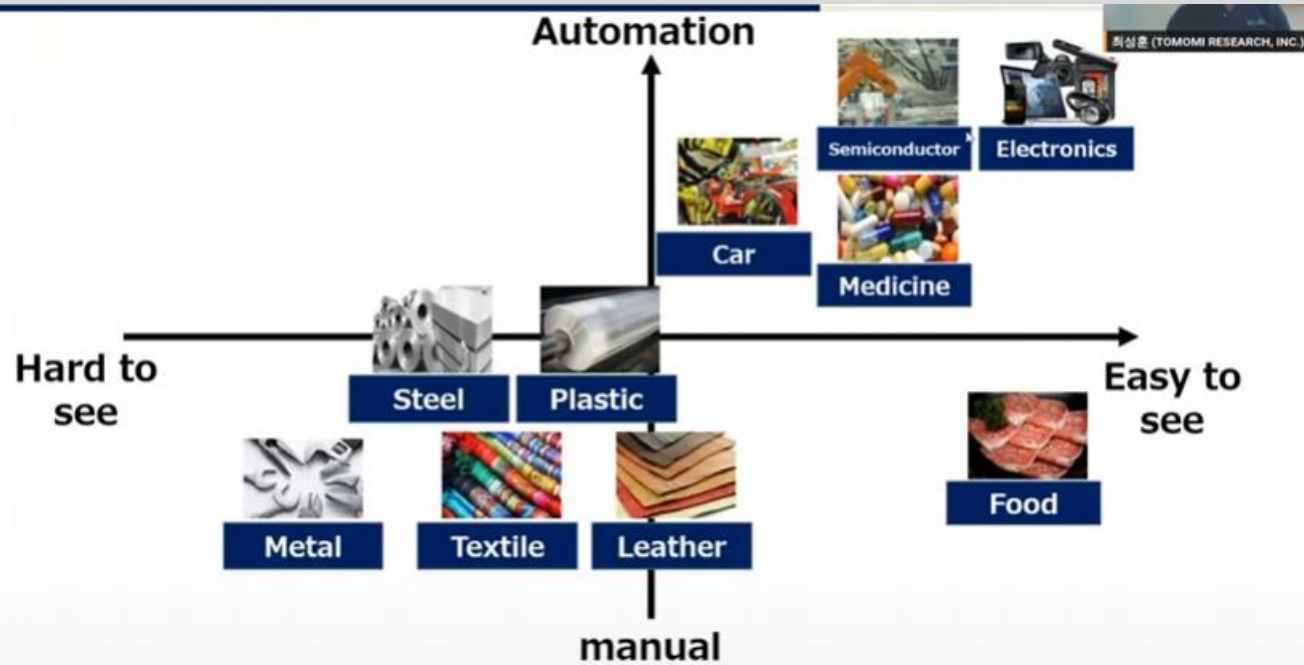
종업원의 1/5 : 약 140만

전세계로는 670만

검사원의  
고령화 및  
지원자 부족

단순한 작업  
숙련자가 필요  
실패가 허용되지  
않음

보이지않는 코스트 : 불량 클레임  
처리



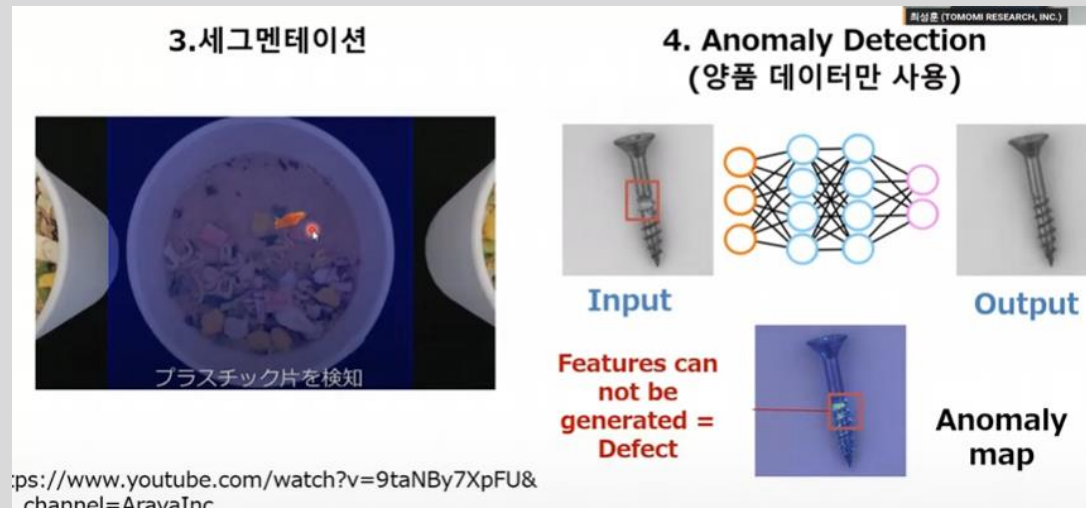
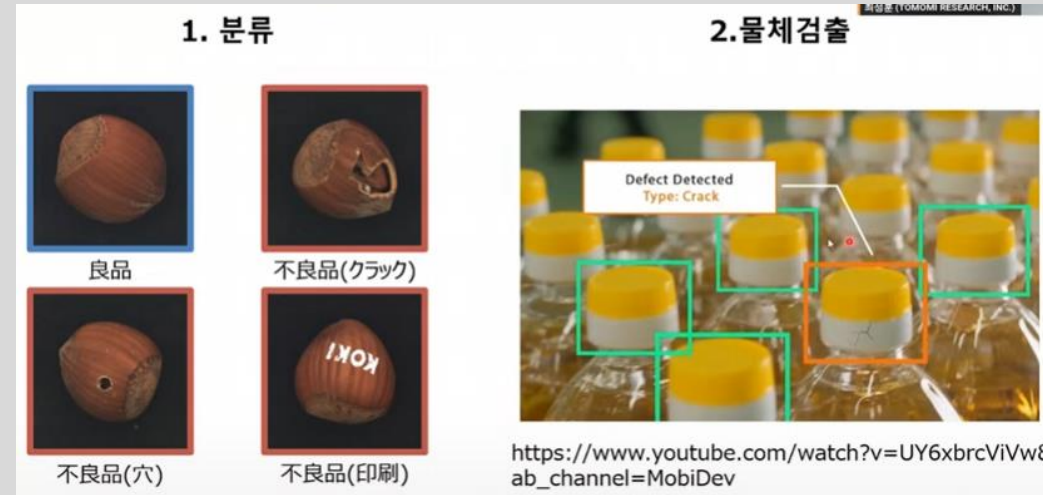
# 외관검사 컴퓨터 비전 트렌드

외관검사는 틀린 그림 찾기 문제와 같은 과정을 가진다

## 외관검사 수행방법의 변화



분류, 물체검출, 세그멘테이션은 지도 학습, anomaly detection은 비지도 학습이다  
불량률이 적으므로 지도 학습이 어렵다



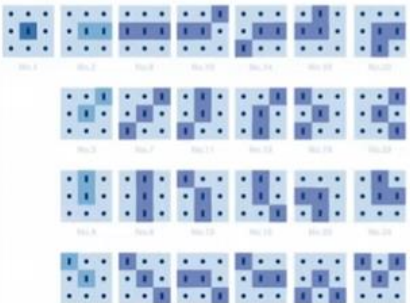


# 외관검사 컴퓨터 비전 트렌드

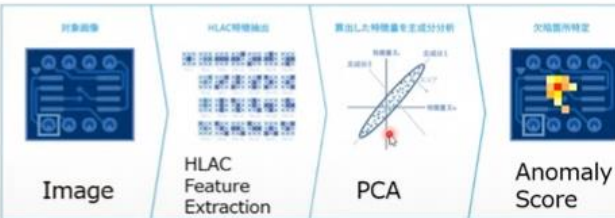
양품의 데이터만으로 학습하여 불량품을 찾는 방법

■ 2011年、HLAC(高次局所自己相関)

OK이미지→특징량 추출→ 통계처리→ Anomaly Detection



大津 展之 教授  
独立行政法人産業技術総合研究所 (AIST)

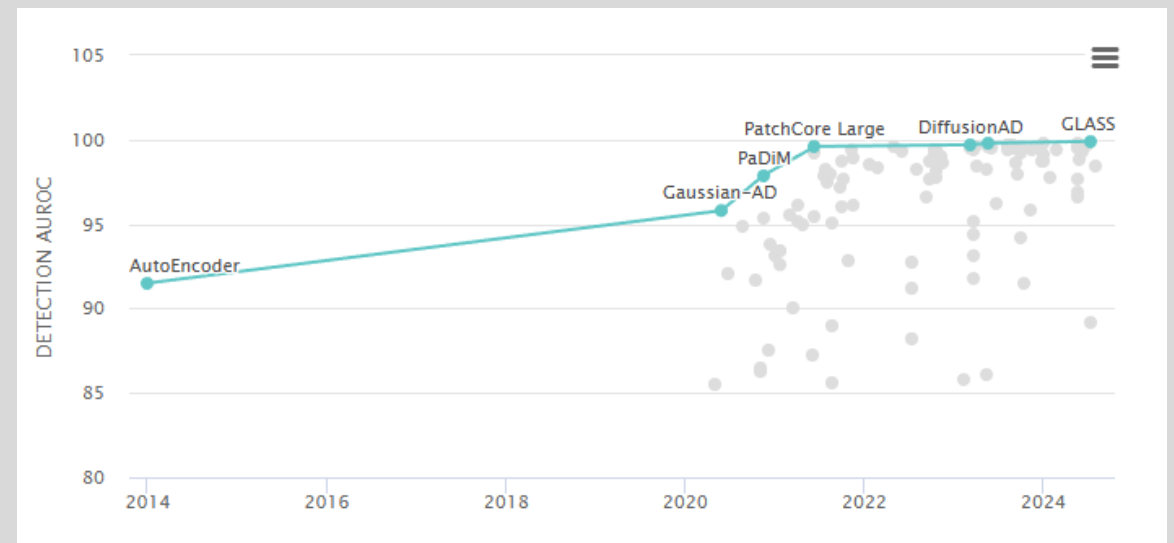
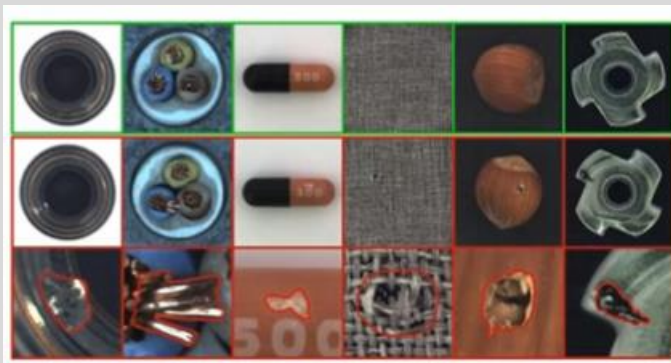


[https://www.aist.go.jp/pdf/aist\\_j/synthesiology/vol04\\_02/vol04\\_02\\_p70\\_p79.pdf](https://www.aist.go.jp/pdf/aist_j/synthesiology/vol04_02/vol04_02_p70_p79.pdf)

<https://zenn.dev/shinue/articles/40560352dcf70b>

Class	Image Size	Normal Num	Defective Num	Defective class	Total
bottle	900 x 900	229	63	3	292
cable	1024 x 1024	282	92	8	374
capsule	1000 x 1000	242	109	5	351
carpet	1024 x 1024	308	89	5	397
grid	1024 x 1024	285	57	5	342
hazelnut	1024 x 1024	431	70	4	501
leather	1024 x 1024	277	92	5	369
metal_nut	700 x 700	242	93	4	335
pill	800 x 800	293	141	7	434
screw	1024 x 1024	361	119	5	480
tile	840 x 840	263	84	5	347
toothbrush	1024 x 1024	72	30	1	102
transistor	1024 x 1024	273	40	4	313
wood	1024 x 1024	266	60	5	326
zipper	1024 x 1024	272	119	7	391

MVTec-AD 데이터셋을 통해 anomaly detection방법을 벤치마킹한다



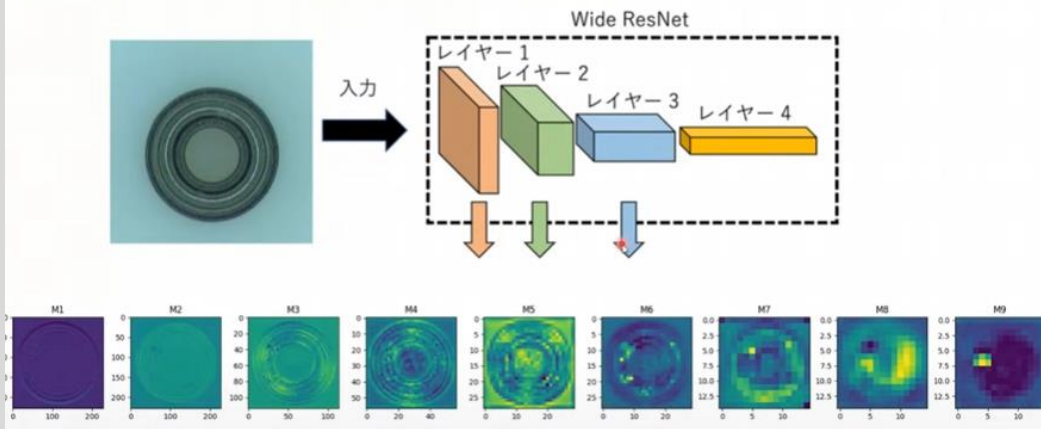
# 외관검사 컴퓨터 비전 트렌드

Auto-encoder는 MSE에서 SSIM의 loss를 변화시켰을 때 성능향상  
학습량과, 학습시간이 길어 다품종 소량생산 기업에 적용이 어려움

따라서 사전 학습된 네트워크를 통해 특징을 추출하는 방법 사용

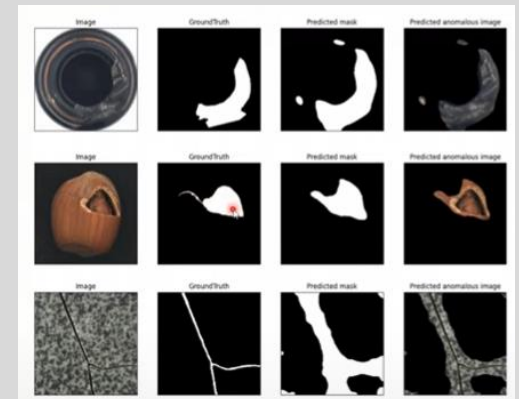
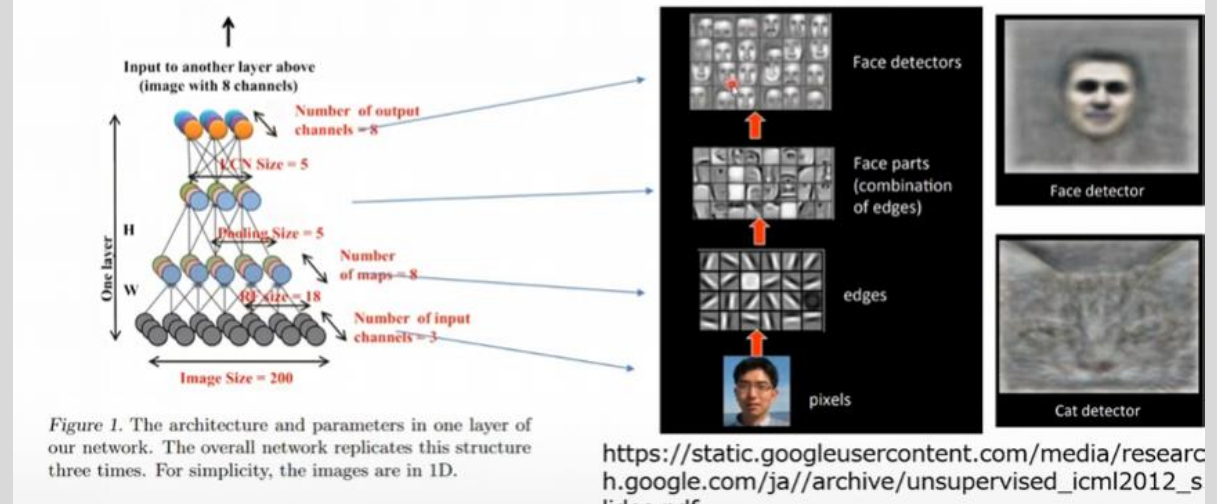
## 2. SPADE : Sub-Image Anomaly Detection with Deep Pyramid Correspondences [2020/05]

### Pretrained network as a Feature Extractor



아이디어

### Pretrained network as Feature Extractor





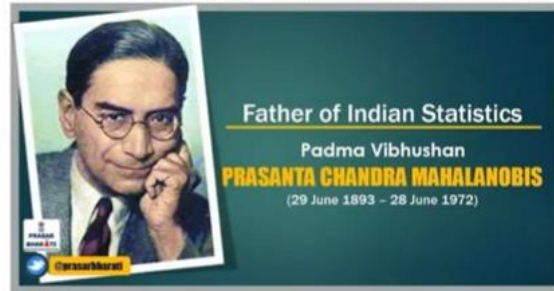
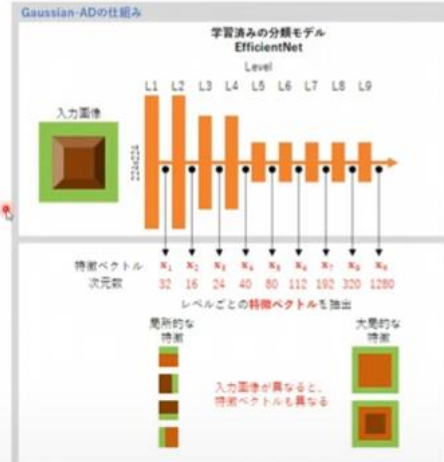
# 외관검사 컴퓨터 비전 트렌드

Gaussian AD는 사전 학습된 네트워크로 추출된 특징들에 마할라노비스 거리를 결합한 방법

## Pretrained Feature Extractor + Mahalanobis Distance

최성훈 (TOMOMI RESEARCH, INC.)

<https://arxiv.org/pdf/2005.14140.pdf>



$$d = \sqrt{(\vec{x} - \vec{\mu})^T \Sigma^{-1} (\vec{x} - \vec{\mu})}$$

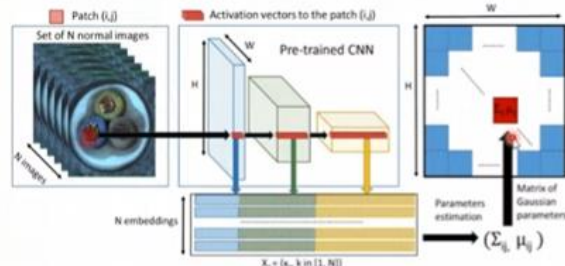
<https://qiita.com/makotoito/items/39bc64d30ce49a9edad8>

평균으로부터 얼마의 표준편차로부터 떨어져 있는지 측정

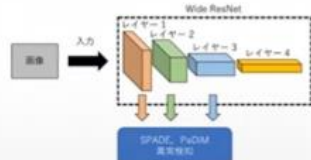
제품의 불량은 알 수 있지만 어디가 불량인지 확인 불가능

## Pretrained Feature Extractor : Mahalanobis Distance (PaDiM)

최성훈 (TOMOMI RESEARCH, INC.)



<https://arxiv.org/pdf/2011.08785.pdf>



- SPADE : slow inference due to kNN
- kNN -> Mahalanobis Distance
- Use the pretrained ResNet with ImageNet data without re-training
- Store the extracted features with normal image data  $y$  and its each pixel  $p$  to the mean  $\mu$  and covariant matrix  $\Sigma$ .
- At inference phase, measure the Mahalanobis distance between input data  $y$  and stored feature data  $(\mu, \Sigma)$ .

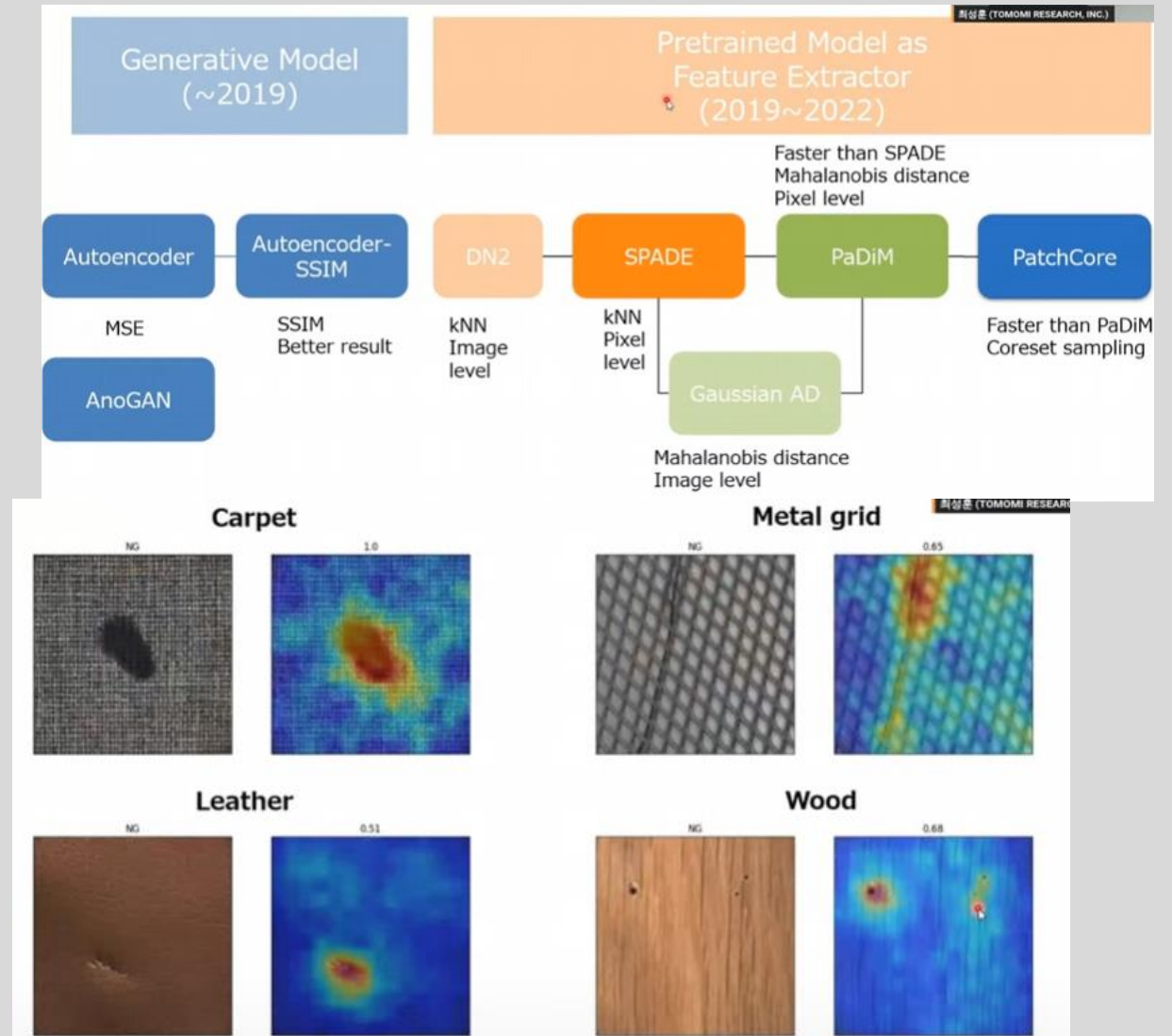
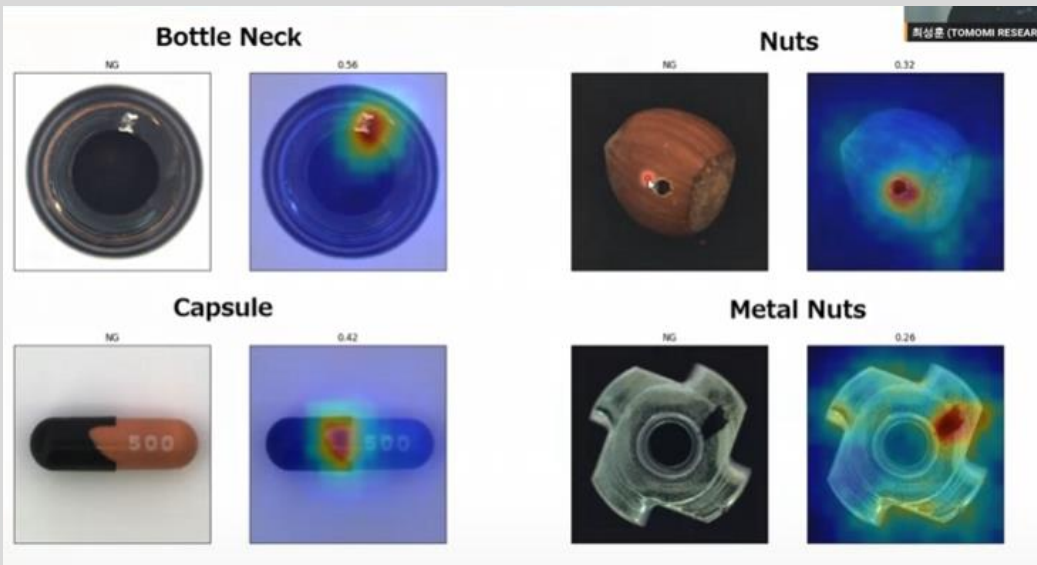
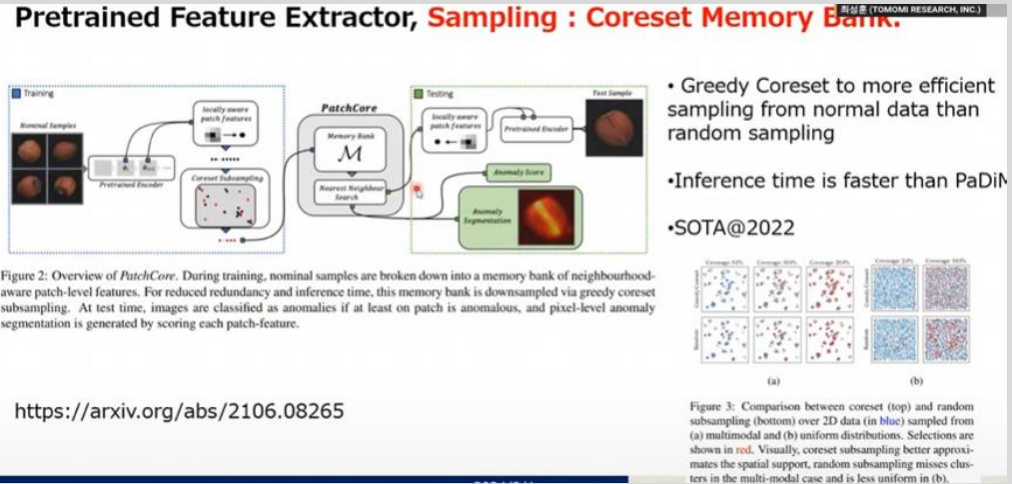
$$M(x_{ij}) = \sqrt{(x_{ij} - \mu_{ij})^T \Sigma_{ij}^{-1} (x_{ij} - \mu_{ij})}$$

PaDiM 모델은 사전 학습된 네트워크로 특징을 추출하고 그것을 작은 patch로 나누어 마할라노비스거리를 측정하는 방법으로 Gaussian AD의 단점을 극복

물체의 정렬에 따라 patch의 이상치를 탐지가 달라진다

# 외관검사 컴퓨터 비전 트렌드







Patchcore = PaDiM에서 patch를 겹쳐서 생성하는데 이때 메모리를 줄이기 위해 greedy coreset 알고리즘을 사용하는 방법

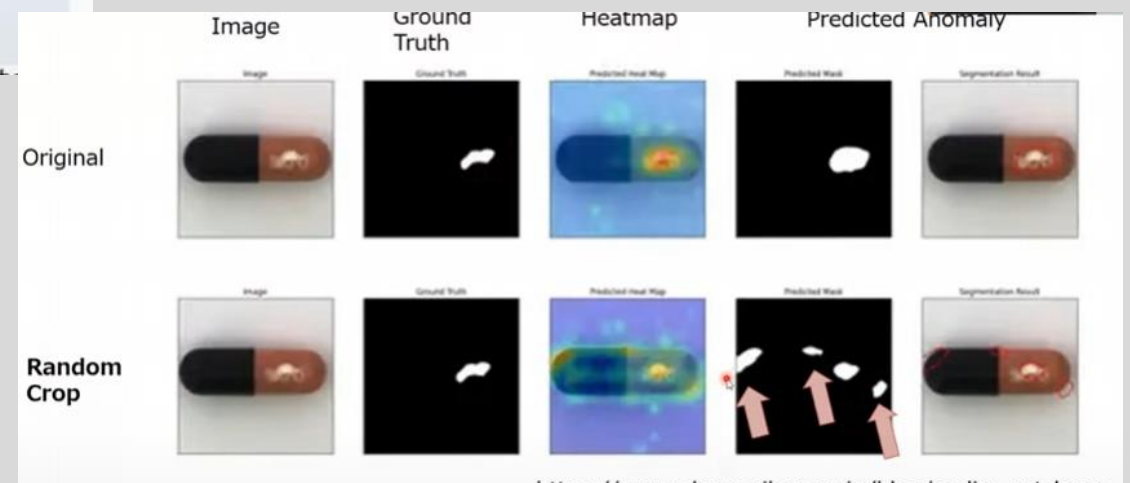


# 외관검사 컴퓨터 비전 트렌드

PaDiM은 물체의 위치 변화에 따라 성능이 크게 변한다

최성훈 (TOMOMI RESEARCH)

Category		Original PaDiM	Original PatchCore	Random Crop PaDiM	Random Crop PatchCore
cable		0.844	0.983	0.740	0.971
capsule		0.850	0.986	0.407	0.909
carpet		0.980	0.984	0.984	0.984
screw		0.710	0.975	0.480	0.742
wood		0.980	0.996	0.987	0.986
zipper		0.686	0.982	0.628	0.980

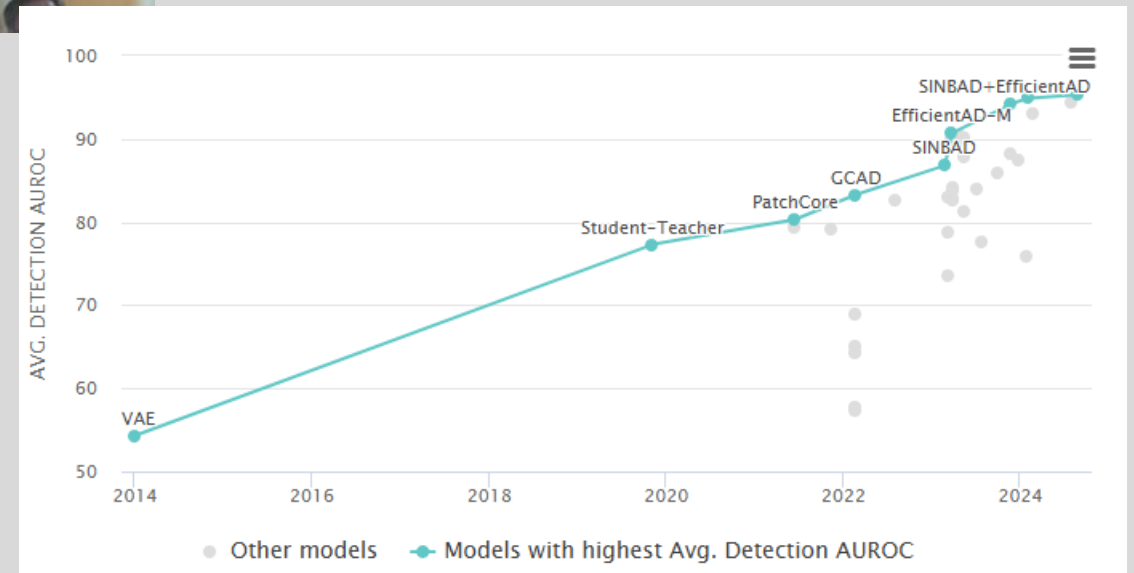
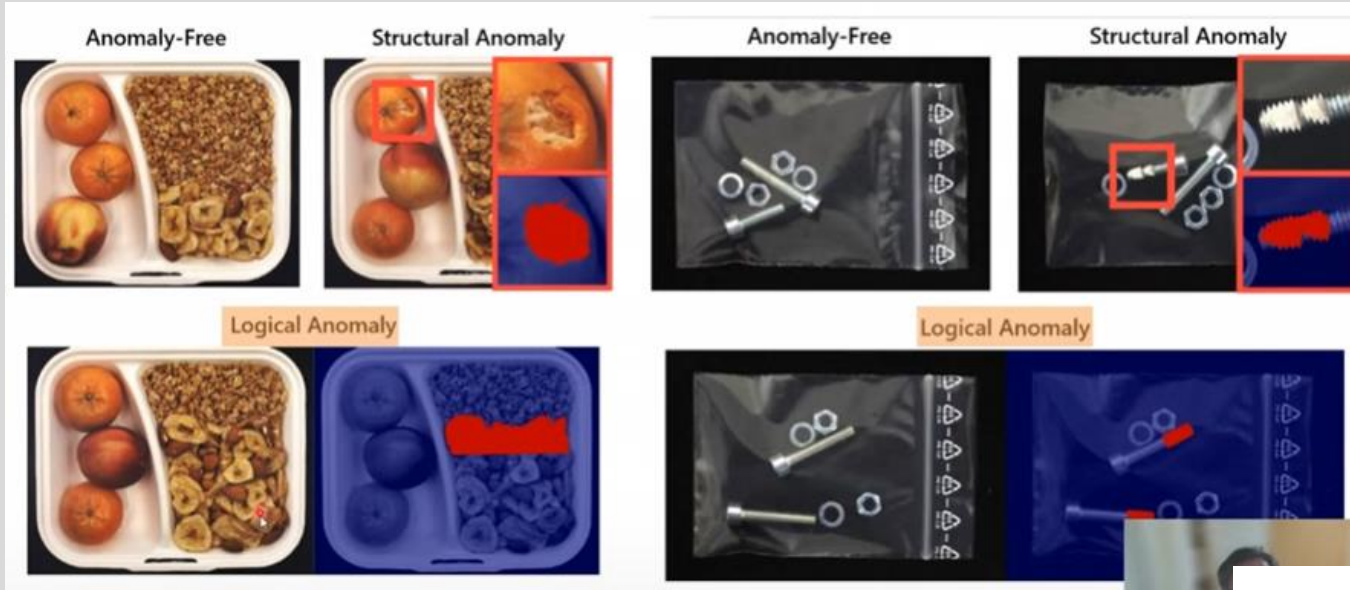


[https://www.choushiken.co.jp/blog/reading\\_patchcore/](https://www.choushiken.co.jp/blog/reading_patchcore/)



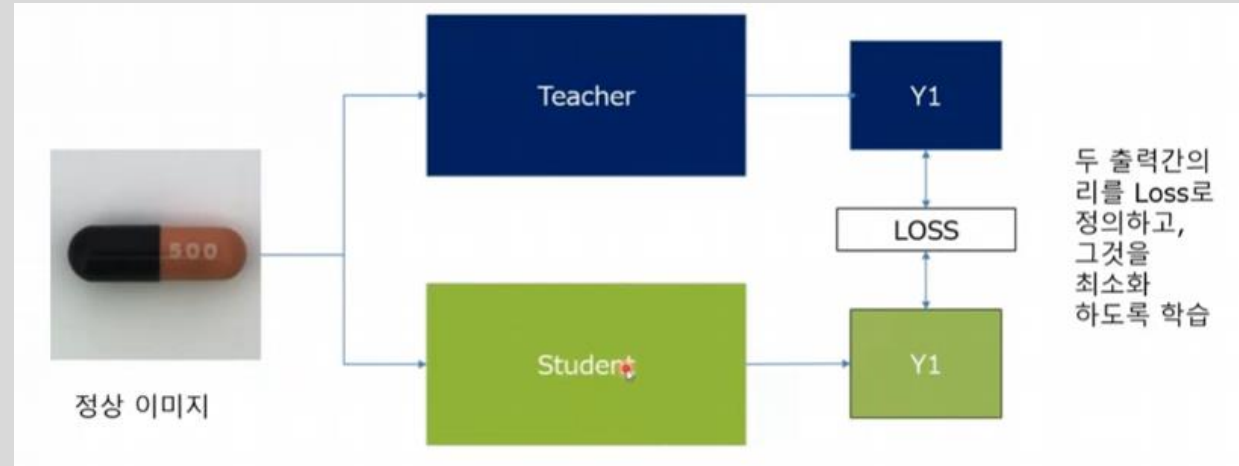
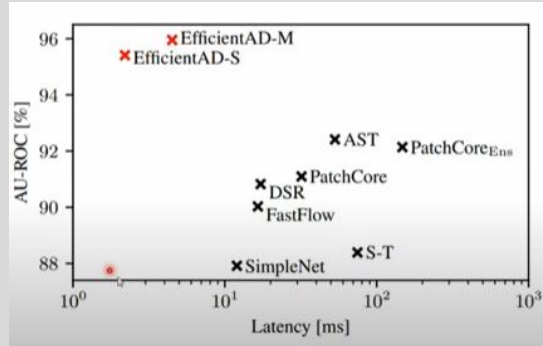
# 외관검사 컴퓨터 비전 트렌드

MVTec LOCO AD는 구조적 결함(긁힘, 찌그러짐)과 논리적 결함(잘못된 위치, 물체의 누락)의 결함을 포함 데이터 셋이다

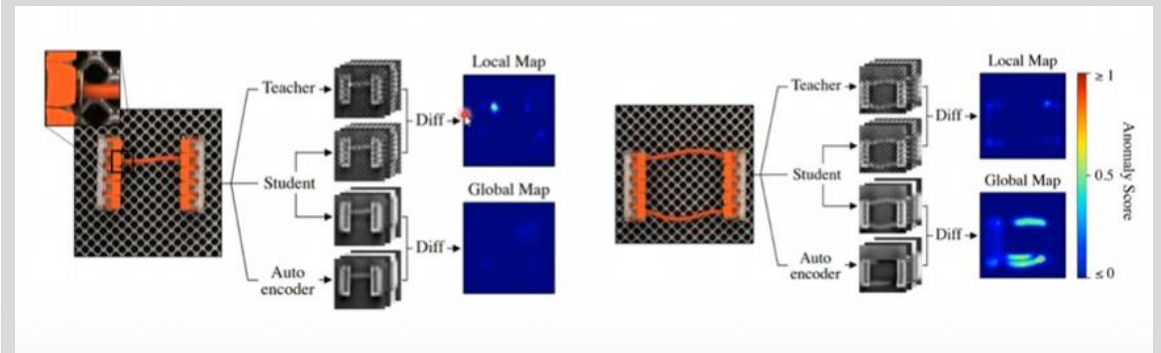
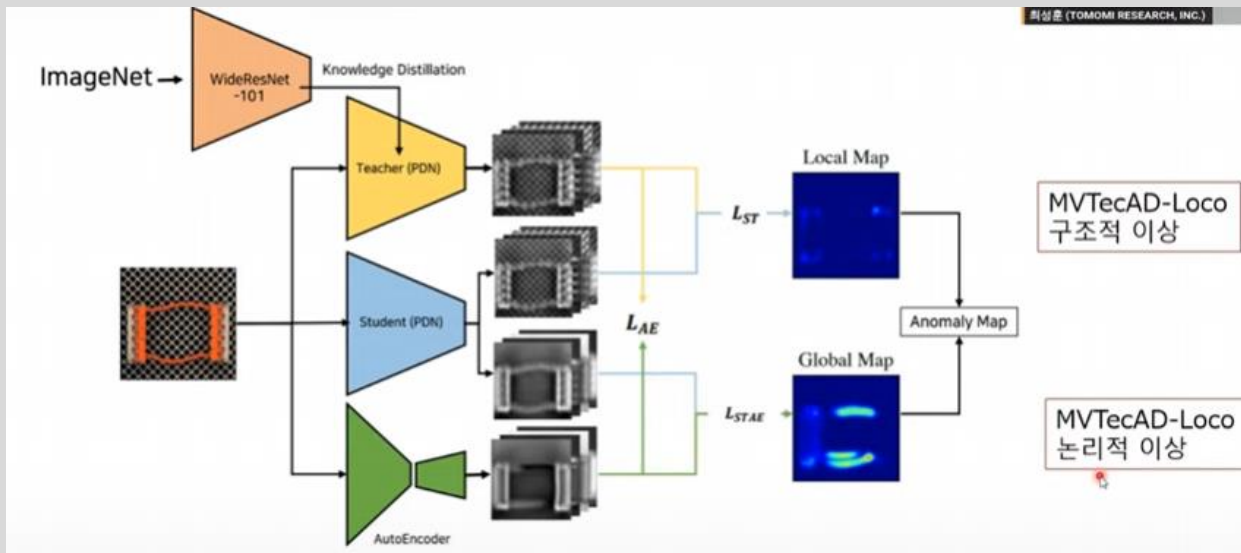


# 외관검사 컴퓨터 비전 트렌드

빠른 추론속도를 필요하는 경우가 발생한다



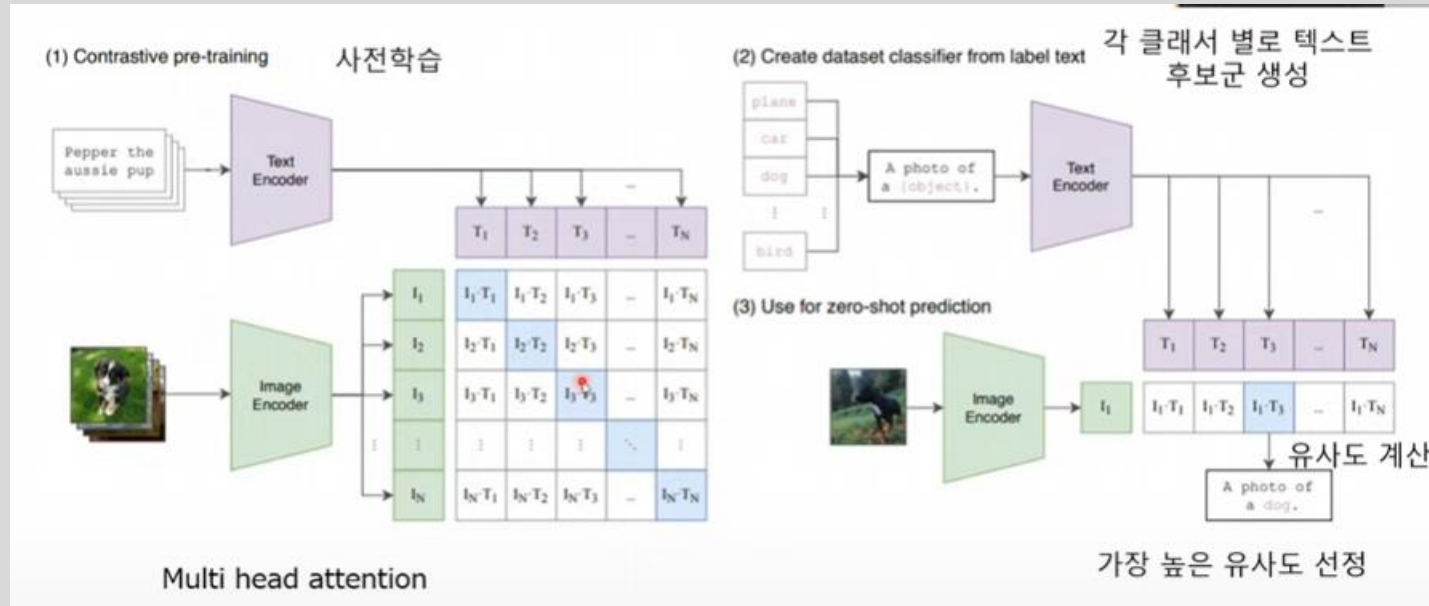
Efficient AD는 복잡하지만 성능이 좋은 teacher모델과 단순한 student 모델을 같이 학습시켜 student모델을 추론에 사용하는 방법



# 외관검사 컴퓨터 비전 트렌드

Efficient ad는 각 제품별로 각각의 모델을 구성해야 한다는 문제가 있다. 따라서 WinCLIP모델이 등장

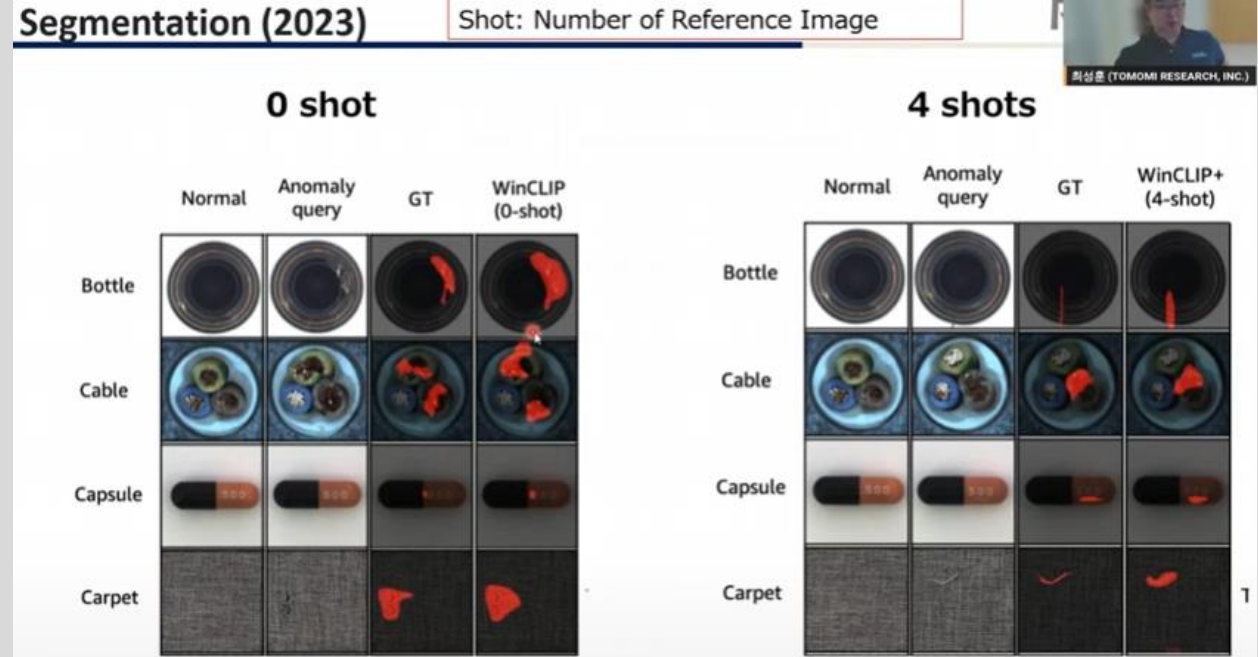
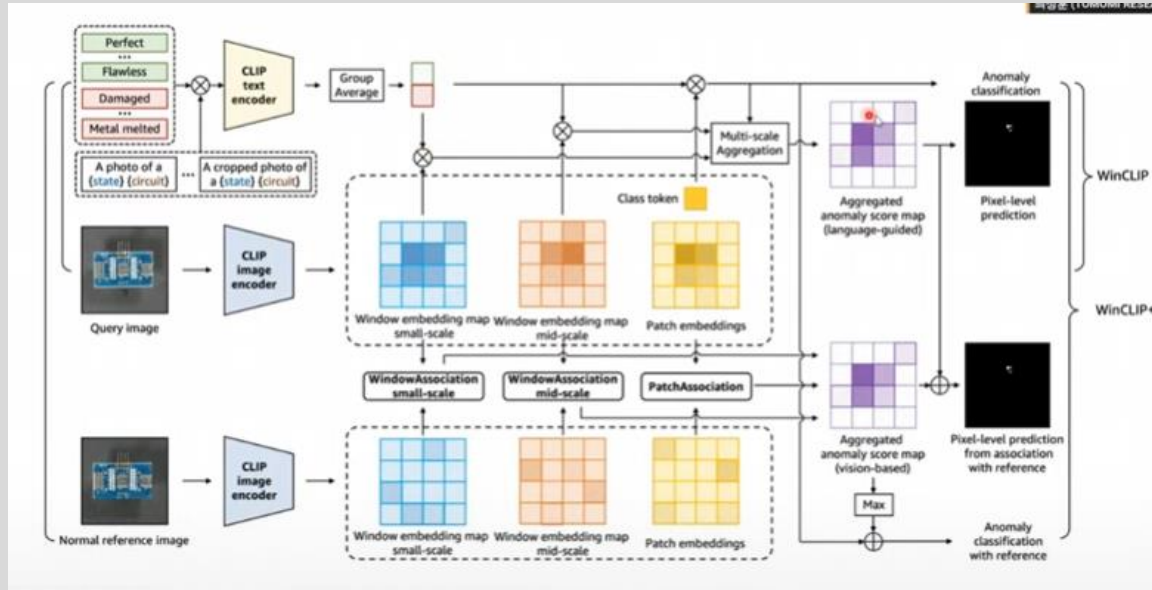
말로 표현한 것과 이미지의 상관관계와 이미지의 라벨을 학습하고 추론시에는 입력된 이미지와 가장 잘 어울리는 라벨을 반환한다.



Anomaly detection에서는 양품과 불량품의 특징에 대한 텍스트와 이미지를 입력하여 학습시킨다

# 외관검사 컴퓨터 비전 트렌드

Anomaly detection에서는 양품과 불량품의 특징에 대한 텍스트와 이미지를 입력하여 학습시킨다

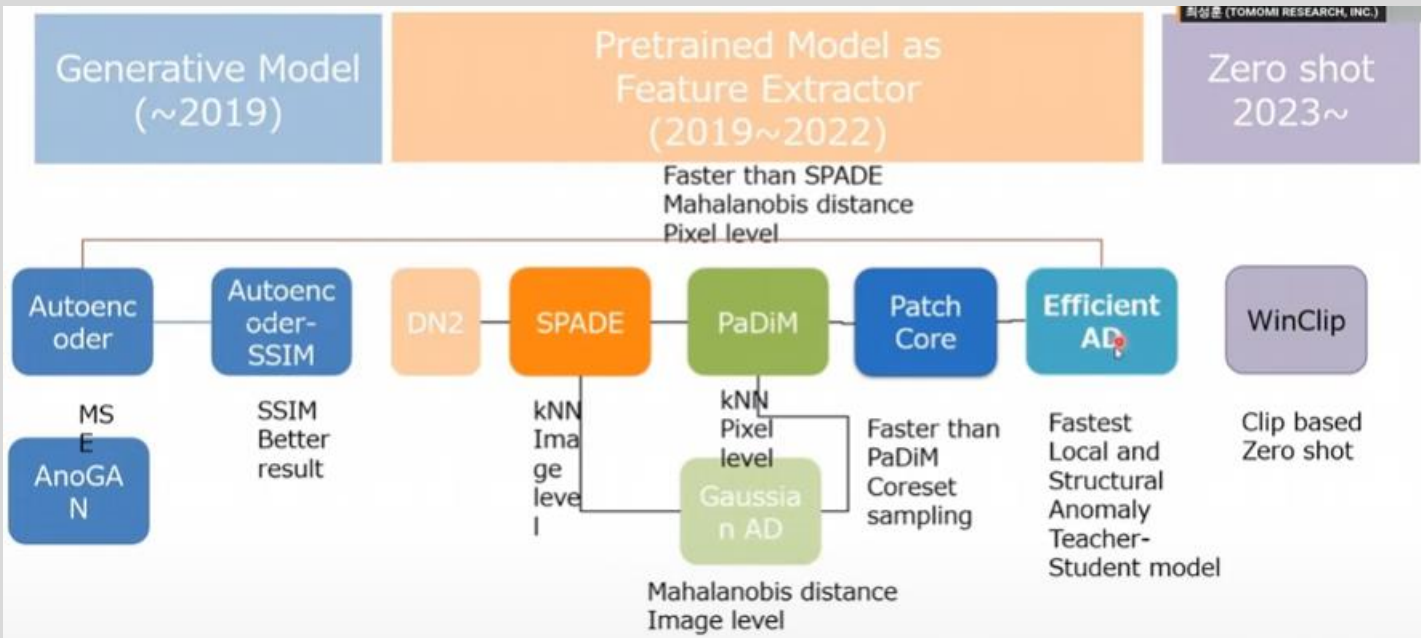
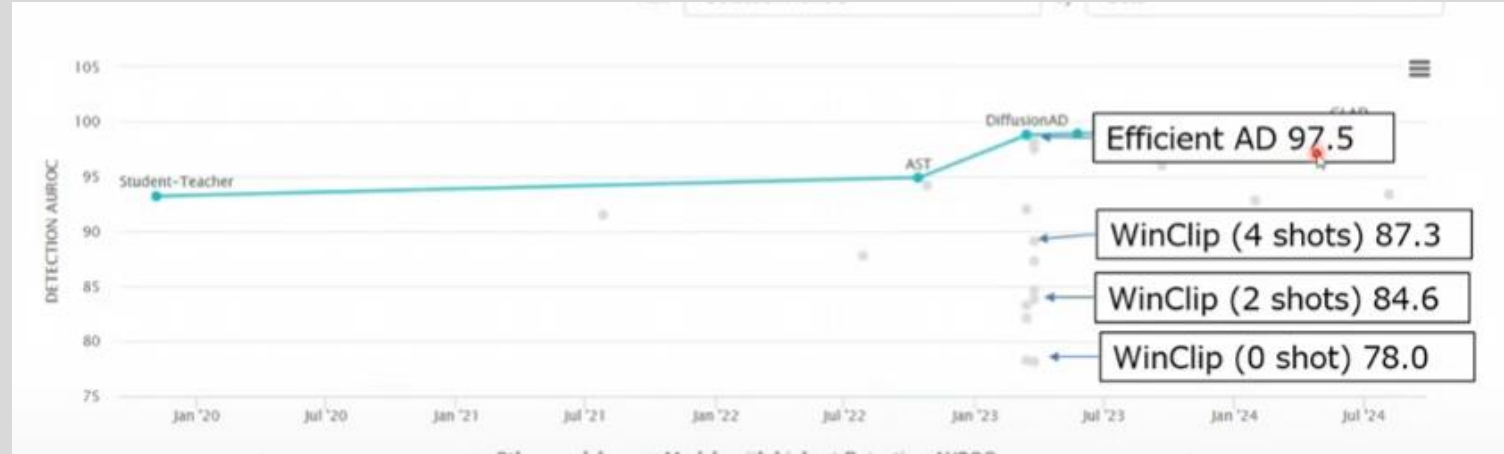
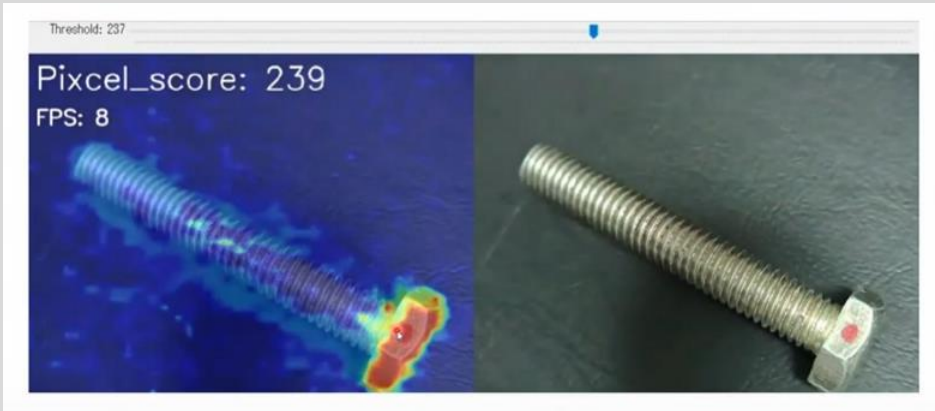


조명과 배경이 바뀌는 것이 가능하고, 감지하고 싶은 이상을 선택하는 것이 가능하다.



# 외관검사 컴퓨터 비전 트렌드

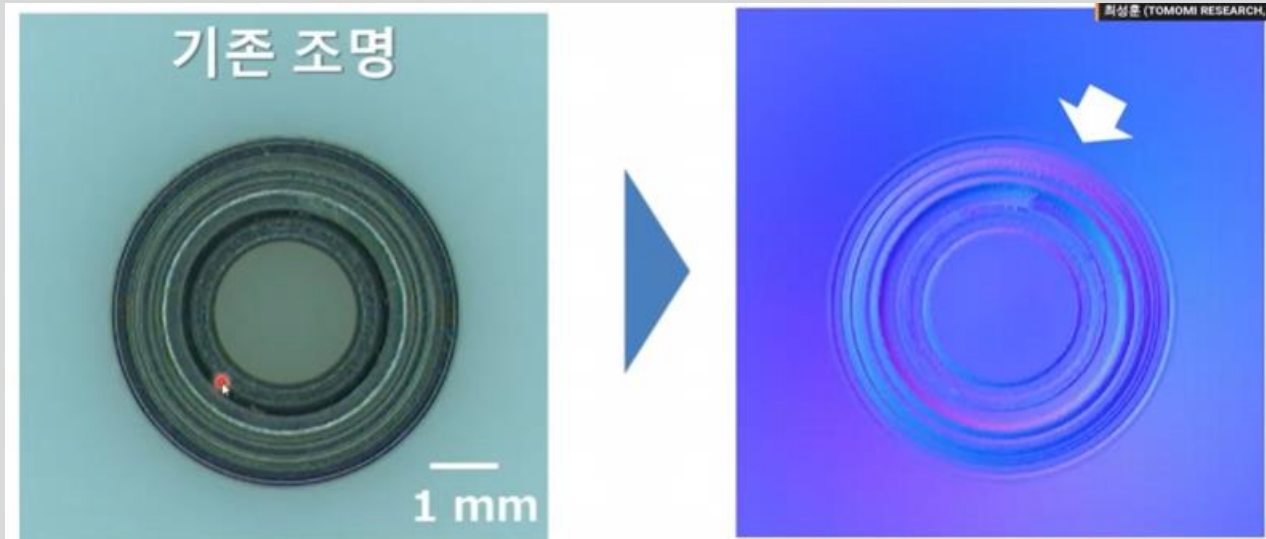
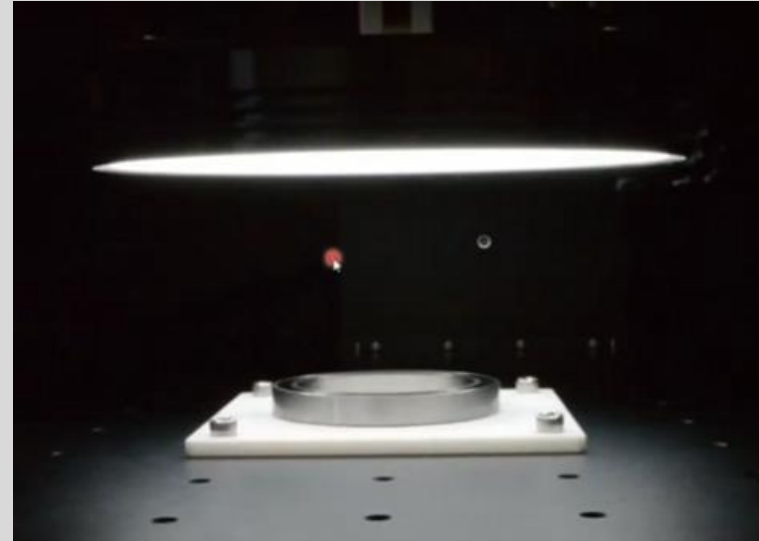
April GAN은 zero shot anomaly detection 모델로 히트맵을 통해 결함을 보여주는 모델





# 외관검사 컴퓨터 비전 트렌드

광택표면 물체는 잘 안보여 자동화가 안됨 -> 광원의 위치를 바꿔 여러 사진을 통해 검사



# 외관검사 컴퓨터 비전 트렌드

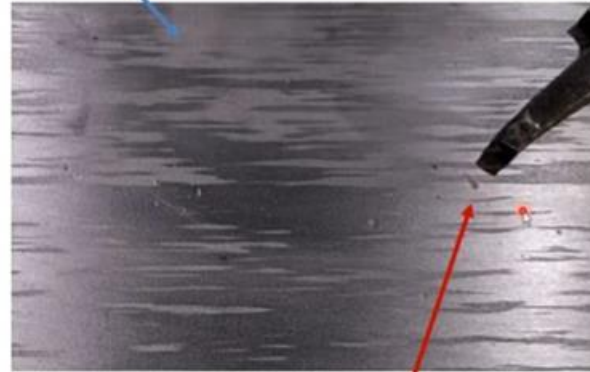
철강 업계 적용예시

■ 위험한 작업 -> 자동화의 요구가 높음.

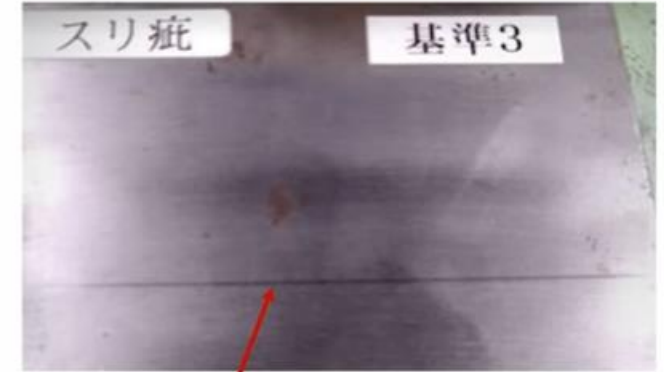


최성훈 (TOMOMI RESEARCH, INC.)

무해한 결함 : 실리콘 스케일

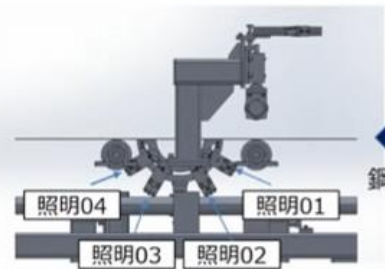


검출하고 싶은 결함

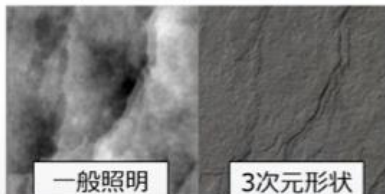


라인 형태의 굽힌 결함

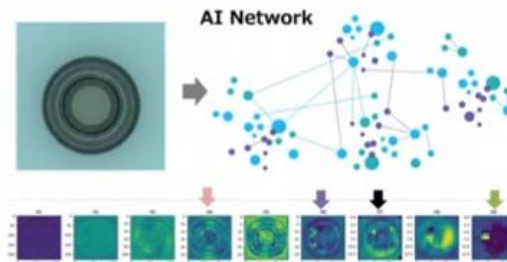
단일 카메라 및 복수 조명 및 화상처리를  
이용한 3차원 이미지 생성



銅板의進行方向

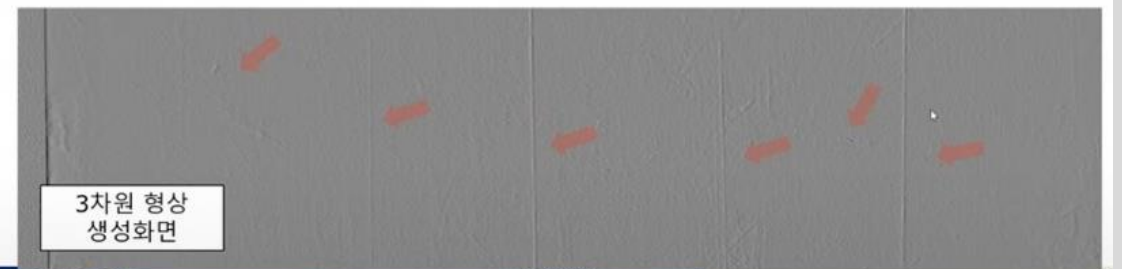


AI 이상검지를 이용한 결함 검출의 자동화



銅板의3次元形状

결함을 AI로 검출



최성훈 (TOMOMI RESEARCH, INC.)

2024/04