

# 비지도 학습 기반 Anomaly Detection을 위한 Autoencoder 모델의 손실 함수 적용에 대한 연구

배주원, 정운수, 남승우, 한병길\*

한국전자통신연구원\*

wjb0406@etri.re.kr yoonsu@etri.re.kr swnam@etri.re.kr \*kilyhan@etri.re.kr

## A Study on the Application of Loss Function in Autoencoder Model for unsupervised Learning-Based Anomaly Detection

Ju-Won Bae, Yun-Su Chung, Seung-Woo Nam, Byung-Gil Han\*

\*Electronics and Telecommunications Research Institute

### 요약

공장에서 생산된 제품의 정상 상태와 이상 상태를 구분하기 위해 Autoencoder를 사용한 비지도 학습에 대한 연구가 많이 이루어지고 있다. 비지도 학습은 데이터 수집에 필요한 시간과 비용을 크게 줄일 수 있다는 장점이 있으나, 일반적으로 비정상 데이터를 학습하지 않기 때문에 지도 학습 기반의 이상 탐지보다는 정확도가 낮은 편이다. 비지도 학습의 성능을 높이기 위해 본 논문에서는 Autoencoder를 사용하여 CE(Cross Entropy), MSE(Mean Squared Error), SSIM(Structural Similarity Index Map)과 같은 다양한 손실 함수(Loss Function)를 적용해서 이상 탐지 모델의 학습 성능을 비교한다. 비교 결과 CE 함수를 사용할 때 가장 정확한 출력 이미지를 만들어냈음을 확인했으며, SSIM의 경우 입력 이미지의 구조적인 특징에 집중하여 이미지를 생성하기 때문에 이상 검출을 위한 기준값 설정을 쉽게 수행할 수 있다는 점을 확인하였다.

### I. 서론

최근 비지도 학습 기반 이상 검출을 사용하는 방법에 대한 많은 연구가 이루어지고 있다[1, 2]. 이상 검출이란 같은 종류의 데이터 안에서 비정상 데이터를 찾아내는 기술로, 공장 생산 라인 비전 검사에서 비정상 제품 이미지를 자동으로 찾아낼 수 있어 비전 검사의 자동화에 사용될 수 있는 유용한 기술이다. 대표적으로 사용하는 모델은 Autoencoder로 입력 이미지를 그대로 출력하는 방향으로 학습하는데, 제품 이미지를 모델에 입력할 때 출력되는 이미지와 입력 이미지와의 차이를 계산해 이상 여부를 판단한다. 이 방법은 정상 이미지만을 학습하여 비정상 이미지를 수집하는데 필요한 인적, 물적 비용을 크게 줄일 수 있다는 장점이 있다. 하지만 지도 학습, 준지도 학습보다 낮은 생성 정확도를 가진다는 한계가 있어 정확도를 높이는 방법이 필요하다.

따라서 본 논문에서는 비지도 학습 기반 이상 검출을 위한 Autoencoder 모델의 생성 이미지 정확도를 높이기 위해 입력과 출력의 차이를 정의하는 손실 함수를 다양하게 적용하여 비교한다. 연구에는 동일한 모델에 MSE, CE, SSIM[3]을 각각 손실 함수로 사용하여 학습하며 입력한 이미지에 대해 출력된 이미지를 비교하고 분석한다.

### II. 연구 방법 및 내용

손실 함수는 모델에서 출력된 값과 실제 정답을 정의하는 함수이며, 모델은 손실 함수를 통해 계산되는 오차를 줄이는 방향으로 학습된다. 본 논문에서는 대표적인 손실 함수인 MSE, CE와 휘도와 대비를 통해 이미지의 구조 정보를 도출할 수 있는 SSIM을 사용한다.

#### 1. MSE

MSE는 실제 정답과 출력값의 오차에 제곱 후 평균을 취한 값으로, 통계

적인 추정치의 정확도에 대한 척도로 많이 사용되며 공식은 아래와 같다.

$$MSE = \frac{1}{2} \sum_n (y_n - t_n)^2$$

여기서  $y$ 는 모델의 출력이며,  $t$ 는 실제 정답 데이터이다.

#### 2. CE

CE는 분류 모델에서 가장 많이 사용되며 서로 다른 두 확률에 대해 같은 사건이 가지는 정보량을 계산한 것으로 공식은 다음과 같다.

$$CE = - \sum_n t_n \log y_n$$

여기서  $y$ 는 모델의 출력값이며,  $t$ 는 실제 정답 데이터이다. CE는 모델의 출력에 자연 로그를 취하여 정답과 곱한 형태로 계산한다.

#### 3. SSIM

SSIM은 단순 이미지 값 차이를 비교하는 것이 아닌 휘도와 대비를 사용하여 인간의 시각 시스템에 따라 이미지의 구조 정보를 도출하여 비교하는 방법이다. 그래서 이미지의 휘도, 대비, 구조를 비교하며 그 공식은 다음과 같다.

$$SSIM(y, t) = \frac{(2\mu_y\mu_t + C_1)(2\sigma_y\sigma_t + C_2)}{(\mu_y^2 + \mu_t^2 + C_1)(\sigma_y^2 + \sigma_t^2 + C_2)}$$

$$C_1 = (0.01 \times L)^2$$

$$C_2 = (0.03 \times L)^2$$

여기서  $y$ 는 모델의 출력 데이터이며,  $t$ 는 실제 정답 데이터이며  $\mu$ 는 이미지의 휘도,  $\sigma$ 는 이미지의 대비이다.  $C_1$ ,  $C_2$ 는 상수로 위에서 서술한 공식처럼 계산되는데  $L$ 은 동적 변수로 여기서는 1로 설정한다.

### III. 본론

#### 1. 학습 데이터 셋 및 학습 모델 설정

본 논문에서 사용하는 데이터 셋은 15개의 물체에 대한 원본 이미지, 손상된 이미지를 가진 MVTec AD-A[4]이다. 그 중 단일 제품에 대한 생성 이미지 비교를 위해 데이터 셋 내 유리병에 대한 학습 이미지 209장, 검증에 20장을 사용하며 테스트에는 조금 깨진 유리병( $I_S$ ), 많이 깨진 유리병( $I_L$ ), 오염물이 있는 유리병( $I_C$ ), 그리고 정상 이미지( $I_G$ )를 각각 20장을 사용한다. 학습 모델은 CNN(Convolutional Neural Network)를 사용하는 Autoencoder로 모델의 형태는 그림 1과 같다.

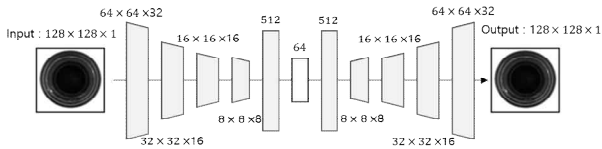


그림 1. Autoencoder 모델의 구조도

그림 1에서 나오는 각 층의 손실 함수는 LeakyReLU를 사용했으며, Optimizer로는 Adam을 사용했다. 학습 모델의 Batch Size는 64, 학습 Epochs는 200으로 같게 설정하여 이를 비교하였다.

#### 2. 실험 수행 및 결과 비교

본 논문에서는 동일한 모델에 다양한 손실 함수를 적용하여 학습하고 정상 제품 이미지와 손상된 제품 이미지들을 입력하여 출력된 이미지들을 비교한다. 각 모델에서 출력된 이미지들은 그림 2와 같다.

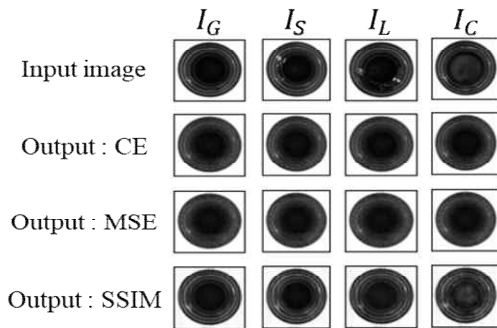


그림 2. 각 모델에 대한 입력 이미지와 출력 이미지 비교

그림 2의 각 모델에 대한 생성 이미지들은 학습한 병 이미지와 유사하게 출력하려 하기 때문에 비슷한 형태를 보였다. SSIM은 입력 이미지의 구조적인 특징들인 병 입구의 홈이나 열룩을 잘 생성했다. 하지만 입력한 이미지에는 없는 잡음들을 출력해 이미지 재생성에서 낮은 성능을 보이는 것을 확인했다. 그리고 MSE를 사용한 경우 다른 모델들보다 흐릿한 이미지를 생성하는 것으로 보였다. 반면, CE 모델은 다른 모델들에 비해 전체적으로 입력한 이미지와 비슷한 출력 이미지를 생성했다.

이번에는 입력 이미지와 출력 이미지의 정확한 값 비교를 위해 이미지 간 L1 거리( $D_{L1}$ )를 계산하여 입력한 이미지를 그대로 구현하는지 비교하였다. 여기서,  $I_G$ 의 L1 거리  $D_{L1}(I_G)$ 은 작을수록 모델이 이미지 재생성을 잘하는 것으로 볼 수 있다. 그리고 손상이 있는 이미지의 L1 거리  $D_{L1}(I_S)$ ,  $D_{L1}(I_L)$ ,  $D_{L1}(I_C)$ 는  $D_{L1}(I_G)$ 와의 차가 클수록 비정상 이미지를 입력했을 때 복원하는 정도가 크다는 것을 의미하는데, 이를 통해 이상 검출에서 비정상임을 판단할 수 있다. 이 수치를 통해 어떤 모델이 이상 검출에 좋은 성능을 보이는지 비교할 것이다. 계산한 L1 거리 비교 결과는 표 1과 같다.

	CE	MSE	SSIM
$D_{L1}(I_G)$	<b>512</b>	551	666
$D_{L1}(I_S)$	541	581	739
$D_{L1}(I_L)$	590	620	775
$D_{L1}(I_C)$	631	658	792
$D_{L1}(I_G) - D_{L1}(I_S)$	29	30	<b>73</b>
$D_{L1}(I_G) - D_{L1}(I_L)$	78	69	<b>109</b>
$D_{L1}(I_G) - D_{L1}(I_C)$	119	107	<b>126</b>

표 1. 각 모델의 출력 이미지 L1 거리( $D_{L1}$ ) 비교

표 1의 결과에서, CE를 사용한 모델은 모든 경우에서 가장 낮은 값을 보였다. 이는 CE 모델이 입력한 이미지를 가장 잘 복원한다는 것을 의미한다. 반면 SSIM은 모든  $D_{L1}$ 에서 가장 크게 계산되었으며, 그림 2에서 분석한 대로 이미지를 그대로 재생성하는 것에 낮은 성능을 보이는 것을 확인했다. 하지만  $D_{L1}(I_G)$ 과  $D_{L1}(I_S)$ ,  $D_{L1}(I_L)$ ,  $D_{L1}(I_C)$ 의 차는 가장 컸는데, SSIM이 손상된 이미지의 구조적인 특징을 잘 구현했기 때문에  $D_{L1}(I_G)$ 와 비교했을 때 큰 차이가 나타나는 것이다.

생성된 이미지들의 다양한 비교를 통해 CE 모델이 가장 정확하게 입력 이미지를 생성한다는 것을 알 수 있었다. 하지만  $D_{L1}(I_G)$ 과 손상된 이미지의  $D_{L1}$ 의 차는 SSIM이 가장 컸으며, 비정상 이미지를 입력했을 때 이상을 잘 찾아낼 수 있는 손실 함수는 SSIM으로 판단된다.

### IV. 결론

본 논문에서는 비지도 학습 기반의 이상 탐지의 성능을 높이기 위한 다양한 손실 함수 적용에 관한 연구를 진행하였다. 출력 이미지와 L1 점수 비교에서 CE 모델이 가장 정확한 이미지 재생성 성능을 보임을 확인했으며 또한 SSIM을 사용하면 이상 검출을 위한 큰 변화를 확인할 수 있다는 결과를 얻었다. 향후 실제 공장 비전 검사에 사용하기 위한 모델 최적화와 Hyperparameter 조절 등을 통해 빠르고 정확하게 비정상 제품을 검출하는 모델을 생성하는 연구를 진행할 예정이다.

### ACKNOWLEDGMENT

본 연구는 한국전자통신연구원 연구운영지원사업의 일환으로 수행되었음. [21ZD1130, 지능제어기반 스마트기계 및 로봇 기술 개발]

### 참고 문헌

- [1] Zhou, C., & Paffenroth, R. C., "Anomaly detection with robust deep autoencoders", In Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining, pp. 665-674, August, 2017.
- [2] Zong, B., Song, Q., Min, M. R., Cheng, W., Lumezanu, C., Cho, D., & Chen, H., "Deep autoencoding gaussian mixture model for unsupervised anomaly detection", In International Conference on Learning Representations, February, 2018.
- [3] Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P., "Image quality assessment: from error visibility to structural similarity", IEEE transactions on image processing, vol. 13, no. 4, pp. 600-612, 2004.
- [4] Bergmann, P., Fauser, M., Sattlegger, D., & Steger, C., "MVTec AD--A Comprehensive Real-World Dataset for Unsupervised Anomaly Detection", In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9592-9600, 2019.