# Loss Function 변화에 따른 VT-ADL 모델 성능 비교 분석

김남중[O], 박창준[**], 박준휘[***], 이재현[****], 곽정환(교신저자)[*]
[O]국립한국교통대학교 소프트웨어학과,
[*]국립한국교통대학교 소프트웨어학과,
[**]국립한국교통대학교 교통·에너지융합학과,
[***]국립한국교통대학교 AI·로봇공학과,
[****]국립한국교통대학교 컴퓨터공학과
e-mail: knj95@kakao.com[O], jgwak@ut.ac.kr[*]

# Comparative Analysis of VT-ADL Model Performance Based on Variations in the Loss Function

Namjung Kim[O], Changjoon Park[**], Junhwi Park[***], Jaehyun Lee[****], Jeonghwan Gwak(Corresponding Author)[*]
[O]Dept. of Software, Korea National University of Transportation,
[*]Dept. of Software, Korea National University of Transportation,
[**]Dept. of IT·Energy Convergence, Engineering, Korea National University of Transportation,
[***]Dept. of AI·Robotics Engineering, Korea National University of Transportation,
[****]Dept. of Computer Engineering, Korea National University of Transportation

● 요 약 ●

본 연구에서는 Vision Transformer 기반의 Anomaly Detection and Localization (VT-ADL) 모델에 초점을 맞추고, 손실 함수의 변경이 MVTec 데이터셋에 대한 이상 검출 및 지역화 성능에 미치는 영향을 비교 분석한다. 기존의 손실 함수를 KL Divergence와 Log-Likelihood Loss의 조합인 VAE Loss로 대체하여, 성능 변화를 심층적으로 조사했다. 실험을 통해 VAE Loss로의 전환은 VT-ADL 모델의 이상 검출 능력을 현저히 향상시키며, 특히 PRO-score에서 기존 대비 약 5%의 개선을 보였다는 점을 확인하였다. 이러한 결과는 손실 함수의 최적화가 VT-ADL 모델의 전반적인 성능에 중요한 영향을 미칠 수 있음을 시사한다. 또한, 이 연구는 Vision Transformer 기반 모델의 이상 검출과 지역화 작업에 있어서 손실 함수 선택의 중요성을 강조하며, 향후 관련 연구에 유용한 기준을 제공할 수 있을 것으로 기대된다.

**키워드:** Vision Transformer, 이상 탐지(Anomaly Detection), 지역화(Localization), 손실함수(Loss Function), MVTec

## I. Introduction

Computer Vision의 이상 탐지 Task는 정상 클래스에 대비되어, 사전에 정의되지 않은 목적으로 생성되거나 특징을 보유하고 있는 이미지 혹은 일정 영역을 식별하고 탐지하는 것으로 정의한다. 더 나아가, 탐지하는 것에 국한되지 않고 위치 정보를 소실하지 않고, 지역화(Localization)를 통해 이미지의 정확한 영역에 이상을 표시하는 것이 최근 가장 활발한 연구 주제 중 하나이다. 따라서, 본 논문에서는 생산 공정에서 발생할 수 있는 불량품이 포함된 MVTec[1] 데이터셋을 활용하여 Vision Transformer for Image Anomaly Detection and Localization(VT-ADL)의 정밀한 이상 탐지 성능과 지역화 성능을 고도화하기 위하여 Loss Function의 변화에 따른 성능을 분석한다.

본 논문에서 활용한 VT-ADL(A Vision Transformer Network for Image Anomaly Detection and Localization) 모델은 기존 자연어처리 분야에서 주로 사용되던 Transformer 아키텍처를 Computer Vision Task에 응용하고자 Vision Transformer(ViT)를 이상 탐지 및 지역화에 고도화 된 딥러닝 모델로 변형한 것이다[2, 3, 4].

딥러닝에서 사용되는 손실함수(Loss Function)는 모델 Parameter를 최적화하기 위하여 모델 훈련 중 사용한다. 이는 모델의 기대 출력값과 예측 출력값 사이의 차이를 측정하고, 그 차이를 최소화하는

것을 목표로 한다[5]. VT-ADL 논문에서 사용하는 손실함수로는 Mean-Squared Error(MSE), Structural Similarity Index Measure(SSIM), Log-likelihood Loss(LL)가 있으며 위 3가지의 손실함수에 가중치를 두어 그 수치를 합한 형태로 사용하고 있다.

본 논문에서는, 해당 손실함수들의 가중치를 조정하거나 손실함수를 교체함으로 Reconstruction의 고도화 및 성능 향상이 가능한지 알아보고자 한다.

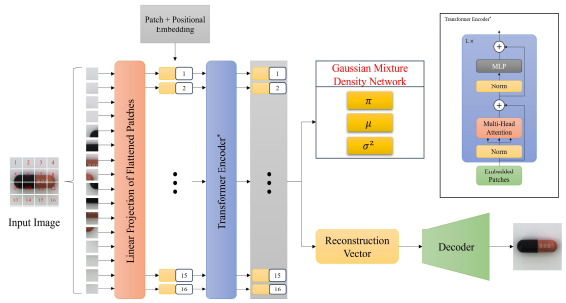


Fig. 1. VT-ADL Architecture

## II. The Proposed Method

### 1. Dataset

Table 1. MVTec AD 데이터셋 구성[1]

| Category | Train | Test (정상) | Test (비정상) |
|---|---|---|---|
| Bottle | 209 | 20 | 63 |
| Cable | 224 | 58 | 92 |
| Capsule | 219 | 23 | 109 |
| Carpet | 280 | 28 | 89 |
| Grid | 264 | 21 | 57 |
| Hazelnut | 391 | 40 | 70 |
| Leather | 245 | 32 | 92 |
| Metal Nut | 220 | 22 | 93 |
| Pill | 267 | 26 | 141 |
| Screw | 320 | 41 | 119 |
| Tile | 230 | 33 | 84 |
| Toothbrush | 60 | 12 | 30 |
| Transistor | 213 | 60 | 40 |
| Wood | 247 | 19 | 60 |
| Zipper | 240 | 32 | 119 |
| **Total** | **3629** | **467** | **1258** |

실험에 사용된 데이터셋의 경우 15가지 종류의 카테고리로 구성되어있으며, 훈련 데이터의 경우 모두 정상 클래스로만 이루어져 있다. 이를 통해 정상 분포를 학습하게 되므로 정상 상태를 벗어나는 비정상 상태를 감지하는 데 유용하다. 이미지 크기의 경우 딥러닝 모델에 입력시 512×512로 조정된다.

### 2. Experiment

본 실험에서는 Log-Likelihood Loss를 Variational AutoEncoder(VAE) Loss로 교체하여 비교하였다. 학습된 VAE는 데이터를 생성할 때 잠재 변수를 샘플링하여 가우시안 분포로부터 새로운 데이터를 생성하므로 이를 원본 데이터와 비교하여 Loss를 계산한다. 이를 통해 현저한 성능의 차이를 보일 것으로 판단하였다.

실험 환경으로는 Ubuntu 22.04 LTS 운영체제를 사용하였으며 GPU는 RTX 3090Ti 2대를 사용하였다. 하이퍼파라미터의 경우 Epoch 400, Patch size 64, Batch size 8, Learning rate 0.0001로 설정하였다.

## III. Experiment Results

Table 2. PRO Score 결과

| Category | Original Loss (MSE + SSIM + LL) | Proposed Loss (MSE + VAE Loss) |
|---|---|---|
| Bottle | **0.949** | 0.918 |
| Cable | 0.776 | **0.865** |
| Capsule | 0.672 | **0.916** |
| Carpet | **0.773** | 0.695 |
| Grid | **0.871** | 0.819 |
| Hazelnut | 0.897 | **0.937** |
| Leather | 0.728 | **0.819** |
| Metal Nut | 0.726 | **0.895** |
| Pill | 0.705 | **0.935** |
| Screw | **0.928** | **0.928** |
| Tile | 0.796 | **0.912** |
| Toothbrush | **0.901** | 0.863 |
| Transistor | **0.796** | 0.701 |
| Wood | **0.781** | 0.725 |
| Zipper | 0.808 | **0.933** |
| Means | 0.807 | **0.857** |

본 실험의 결과는 Table. 2와 같다. 각 카테고리 별로 성능은 다 달랐으나 VAE 손실함수를 사용했을 때 평균적으로 성능이 더 준수하였다. 실험을 위해 제안된 Loss의 경우 MSE와 VAE를 통해 학습된 분포를 KL Divergence Loss와 LL의 조합인 VAE Loss로 구성되어있는데, 여기서 KL Divergence가 잠재 변수의 분포를 정규화함으로써 모델이 더 안정적으로 학습하고 다양한 데이터를 생성할 수 있도록 도운 것으로 보인다. 이를 통해 비정상 데이터를 더 잘 검출할 수 있었던 것으로 해석된다.

## IV. Conclusions and Future Work

본 논문에서는 VT-ADL의 손실함수 변화와 가중치 수정에 따른 모델의 성능을 확인하였으며, 이상 탐지 및 지역화의 고도화 가능성을 확인하였다. 본 논문에서 제안한 VAE Loss로 기존 LL Loss를 대체하였을 때 PRO score가 약 5% 더 향상되는 것을 확인하였다.. 추후 연구에서는 VT-ADL 외에도 Vision Transformer 기반 이상

탐지 모델들에 대한 연구 및 성능 향상을 위한 모듈 변경 등을 진행할 예정이다.

## ACKNOWLEDGEMENT

This work was supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korean government (MSIT) (No. 2014-3-00077).

## REFERENCES

[1] P. Bergmann, M. Fauser, D. Sattlegger and C. Steger, "MVTec AD — A Comprehensive Real-World Dataset for Unsupervised Anomaly Detection," *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9584-9592, 2021, doi: 10.1109/CVPR.2019.00982.

[2] P. Mishra, R. Verk, D. Fornasier, C. Piciarelli and G. L. Foresti, "VT-ADL: A Vision Transformer Network for Image Anomaly Detection and Localization," *2021 IEEE 30th International Symposium on Industrial Electronics (ISIE),* pp. 01-06, 2021, doi: 10.1109/ISIE45552.2021.9576231.

[3] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. Gomez, Ł. Kaizer and I. Polosukhin, "Attention is All You Need," *Advances in Neural Information Processing Systems 30*, Vol. 1, pp. 5999-6009, Dec. 2017.

[4] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszhoreit and N. Houlsby," An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," *In International Conference on Learning Representations*, 2021.

[5] J. Terven, D. M. Cordova-Esparza, A. Ramizez-Pedraza and E. A. Chavez-Urbiola, "Loss functions and metrics in deep learning. A review." *arXiv preprint arXiv:2307.02694*, 2023.