

Self-Attention 딥러닝 모델 기반 산업 제품의 이상 영역 분할 성능 분석

박창준[○], 김남중*, 박준휘**, 이재현***, 곽정환(교신저자)*

[○]한국교통대학교 교통·에너지융합학과,

*한국교통대학교 소프트웨어학과,

**한국교통대학교 AI·로봇공학과,

***한국교통대학교 컴퓨터공학과

e-mail: cjp128@a.ut.ac.kr[○], jgwak@ut.ac.kr*

Performance Analysis of Anomaly Area Segmentation in Industrial Products Based on Self-Attention Deep Learning Model

Changjoon Park[○], Namjung Kim*, Junhwi Park**, Jaehyun Lee***, Jeonghwan Gwak(Corresponding Author)*

[○]Dept. of IT·Energy Convergence, Korea National University of Transportation,

*Dept. of Software, Korea National University of Transportation,

**Dept. of AI·Robotics Engineering, Korea National University of Transportation,

***Dept. of Computer Engineering, Korea National University of Transportation

● 요약 ●

본 논문에서는 Self-Attention 기반 딥러닝 기법인 Dense Prediction Transformer(DPT) 모델을 MVTEC Anomaly Detection(MVTEC AD) 데이터셋에 적용하여 실제 산업 제품 이미지 내 이상 부분을 분할하는 연구를 진행하였다. DPT 모델의 적용을 통해 기존 Convolutional Neural Network(CNN) 기반 이상 탐지 기법의 한계점인 지역적 Feature 추출 및 고정된 수용영역으로 인한 문제를 개선하였으며, 실제 산업 제품 데이터에서의 이상 분할 시 기존 주력 기법인 U-Net의 구조를 적용한 최고 성능의 모델보다 1.14%만큼의 성능 향상을 보임에 따라 Self-Attention 기반 딥러닝 기법의 적용이 산업 제품 이상 분할에 효과적임을 입증하였다.

키워드: 자기주의(Self-Attention), 밀집 예측 트랜스포머(Dense Prediction Transformer), 이상 분할(Anomaly Segmentation), 딥러닝(Deep Learning), 성능 분석(Performance Analysis)

I. Introduction

산업 제품의 생산을 위한 공정에서 발생하는 불량품을 초기에 색출하는 것은 산업 제품의 안정성과 신뢰성 증진을 위한 핵심적인 요인으로, 생산 공정 내 이상 제품 선별을 위한 모니터링 기법은 제조품의 품질 관리를 위해 필수적이다. 기존에는 불량품을 선별하기 위해 노동자의 육안 검사 또는 X-Ray와 같은 장비를 통한 검사를 진행하였으며, 이러한 작업을 계속해서 반복하는 경우 작업자의 피로도를 누적시키고 선별 정확도를 크게 저해한다[1]. 이에 따라 최근에는 해당 문제를 해결하기 위해 U-Net[2]과 같은 객체 분할(Segmentation)모델과 같은 딥러닝 기반 기법을 실제 생산 공정에 적용함으로써 불량품 분류의 자동화된 모니터링을 진행한다. U-Net은 Autoencoder 구조를 개선한 객체 분할모델로, Skip-Connection을 통해 기존 CNN 기반 Encoder-Decoder 구조의 한계점인 지역적인 정보 소실을 완화하지만, CNN의 본질적인 한계로 인해 필터의 크기가 고정되어 수용영역(Receptive Field)의 최적화가 어려우며, 전역적인

문맥(Context) 추출이 제한적이다. 따라서 본 논문에서는 더욱 향상된 불량품의 이상 영역 분할이 가능하도록 Self-Attention 기반 딥러닝 모델인 Dense Prediction Transformer(DPT)[3]를 이상 분할 분야에 적용한다. DPT의 적용을 통해 Vision Transformer(ViT)[4]의 Encoder 특징인 전역적인 Feature 추출 및 동적인 수용 영역 구현 능력을 활용함으로써 기존 CNN 기반 Encoder-Decoder 구조의 한계를 극복하고, Self-Attention 기반 딥러닝 모델의 적용이 산업 제품 이상 분할 성능 향상에 효과가 있음을 확인한다.

II. Experiment

실험을 위한 데이터셋으로 실제 산업 제품 이미지 데이터로 구성된 MVTEC AD[5]를 사용하였다. 데이터셋은 실제 제품과 텍스처에 대한 병, 케이블, 기축 표면 등 총 15개의 산업 제품 클래스로 이루어진 4,096개의 정상 이미지, 1,258개의 비정상 이미지를 포함하며, 해당

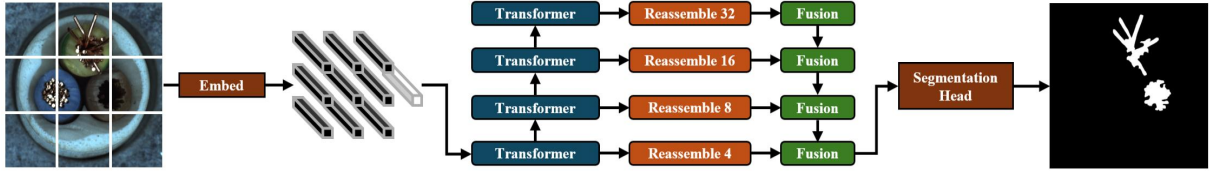


Fig. 1. DPT 모델의 전체 구조

데이터를 256×256의 크기로 Resize 한 후 각 클래스별 전체 데이터의 Train, Validation, Test 비율을 6:2:2로 설정하여 DPT 및 비교군인 U-Net 계열 모델의 학습을 진행하였다.

III. Proposed Method and Results

제안된 DPT 모델은 먼저 백본 모델인 ViT를 통해 Feature 추출을 진행한다. 이때 ViT의 Feature Map은 입력된 이미지의 해상도와 같은 수용영역을 가진다. 이후 {3,6,9,12}번째 Layer의 Feature Map을 Reassemble 블록과 연결해 낮은 수준부터 높은 수준의 해상도를 가지도록 Feature를 재조합한다. 이때 Reassemble 블록은 수식(1)을 따르며, Read 단계를 통해 Readout 토큰을 각 패치에 투영한 후 흩어진 패치들을 연결하여 입력된 Feature Map의 해상도를 설정한 s 와 \hat{D} 에 맞게 조절하는 Resample 과정을 진행한다.

$$Reassemble_x(t) = (Resample_s \cdot Concatenate \cdot Read)(t) \quad (1)$$

$$Reassemble : \mathbb{R}^{N_p+1 \times D} \rightarrow \mathbb{R}^{N_p \times D} \rightarrow \mathbb{R}^{\frac{H}{p} \times \frac{W}{p} \times D} \rightarrow \mathbb{R}^{\frac{H}{s} \times \frac{W}{s} \times \hat{D}} \quad (2)$$

수식 (2)는 Reassemble 진행 시 전체 Feature Map의 해상도 변화 과정을 표현한다. Reassemble 과정을 통해 추출된 각 해상도의 Feature Map은 Fusion 블록에 입력되어 Dense Prediction 구조를 구현하며, 최종적으로 Segmentation Head와 연결되어 이상 영역 분할을 진행한다.

Table 1.은 DPT 및 각 비교군 모델의 성능을 정리한 테이블이다. DPT 모델의 이상 영역 분할 성능이 U-Net 구조 모델의 최고 성능을 가지는 ResU-Net보다 평균 1.14%성능 향상을 보였으며, 이는 DPT 모델이 기존 U-Net과 같은 CNN 기반 모델의 한계를 극복하고, 산업 제품 이상 영역 분할에 더욱 효율적인 기법임을 의미한다.

Table 1. 클래스별 IoU Score 테이블

	U-Net	ResU-Net	DPT
Carpet	0.8973	0.9124	0.8963
Grid	0.9154	0.9017	0.9124
Leather	0.8998	0.9088	0.9017
Tile	0.9132	0.9103	0.9245
Wood	0.9011	0.9135	0.9387
Bottle	0.9187	0.9199	0.9472
Cable	0.8962	0.9022	0.9058
Capsule	0.9056	0.9156	0.9191
Hazelnut	0.9123	0.9041	0.9296
Metal Nut	0.9084	0.9067	0.9034
Pill	0.9145	0.9185	0.9159
Screw	0.9029	0.9111	0.9411
Toothbrush	0.9198	0.9072	0.9088
Transistor	0.9047	0.9168	0.9322
Zipper	0.9116	0.9033	0.9456
Average	0.9081	0.9101	0.9215

IV. Conclusions

본 논문에서는 산업 제품 생산 공정에서의 향상된 이상 제품 선별을 위해 DPT 모델을 이상 분할 기법에 적용하는 방법을 제안하였다. Transformer Encoder의 특징을 활용하여 기존 CNN의 한계를 극복 하였으며, 성능 분석 결과 실제 산업 제품 이상 영역 분할에 있어 성능 향상이 있음을 확인하였다. 추후 DPT 모델 구조를 응용하여 경량화 기법의 적용을 통해 더욱 강인하고 경량화된 이상 영역 분할 모델을 설계할 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] J. Liu, G. Xie, J. Wang, S. Li, C. Wang, F. Zheng and Y. Jjin, "Deep Industrial Image Anomaly Detection: A Survey," arXiv e-prints, arXiv: 2301.11514, January 2023.
- [2] O. Ronneberger, P. Fischer and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015, Vol. 9351, pp. 234-241, November 2015.
- [3] R. Ranftl, A. Bochikovsky and V. Koltun, "Vision Transformers for Dense Prediction," Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 12179-12188, October 2021.
- [4] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit and N. Houlsby, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," International Conference on Learning Representations ICLR 2021, January 2021.
- [5] P. Bergmann, K. Batzner, M. Fauser, D. Sattlegger and C. Steger, "The MVTEC Anomaly Detection Dataset: A Comprehensive Real-World Dataset for Unsupervised Anomaly Detection," International Journal of Computer Vision (IJCV), Vol. 129, pp. 1038-1059, January 2021.