

Object detection for defect detection

Kim Ji Won , Jung Hyunjin , Kim Young-soon

May 2023

Abstract

Transformer 구조는 현재 computer vision 분야에서 다양한 방식으로 사용되고 다양한 문제를 해결해오고 있다. attention이라는 개념을 필두로 CNN의 기본 가정을 무시하고 object detection에서 뛰어난 성능을 보이고 있다. 이에 따라서 구리 축관 예측 문제에서도 Transformer를 사용해 성능 향상을 하려한다.

1 Introduction

4차 산업혁명은 제조업뿐만 아니라 다양한 산업 분야에서도 큰 변화를 가져오고 있다. 인공지능, 빅데이터, 클라우드 컴퓨팅, 로봇 공학 및 사물인터넷(IoT) 등의 기술을 기본으로 하며, 기존의 산업 구조를 완전히 변화시키고 새로운 경제 성장 모델을 창출한다. 4차 산업혁명으로 인해 스마트 팩토리 같은 첨단화된 제조공정은 대부분 자동화되어 제품을 대량 생산하고 있다. 스마트 팩토리는 생산 공정을 자동화하여 생산성을 향상시키고, 스마트 팩토리의 세 가지 핵심 요소는 사물인터넷, 클라우드 컴퓨터, 빅데이터 분석기술이다. 삼성전자, Intel, Siemens 등 다양한 기업들이 생산 프로세스 개선과 생산성 증가를 위해 스마트 팩토리를 도입하고 있다. 생산 라인의 자동화, 생산 데이터의 수집과 분석, 로봇 기술 등을 활용하여 생산 공정을 최적화한다. 삼성전자는 스마트 팩토리 도입으로 생산성이 15% 향상되었고, Siemens의 경우 생산 라인의 운영 시간이 20% 증가했다. 이러한 예는 생산 프로세스 개선을 목적으로 하는 기업들이 스마트 팩토리 도입을 통해 얻을 수 있는 잠재적 이점을 보여준다. 컴퓨터 비전 기술을 이용한 산업용 시각 검사는 수십 년간 고려되어온 접근 방식이다. 최근 딥러닝의 발전으로 인해 단순한 환경에서부터 복잡한 환경에서까지 자동 시각 검사가 가능해졌다. 딥러닝 기반의 컴퓨터 비전에서 컨벌루션 신경망(CNN)은 최근 10년간 컴퓨터 비전 분야의 대표적 방법이었다.

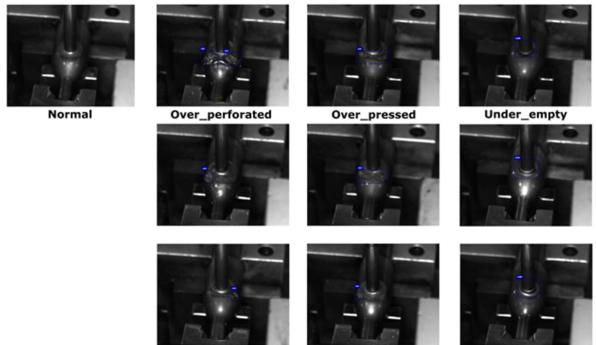


Figure 1: Class label

Abushahma 외(2019)는 객체탐지(Object Detection) 분야에서 R-CNN과 그 파생 모델들의 개념과 구조를 자세히 설명하고, Faster R-CNN, RPN 등을 활용하여 객체 탐지의 정확도와 속도를 향상시킨다. 뿐만 아니라 최근 등장한 어텐션 기반 비전 트랜스포머 아키텍처는 이미지 분류, 객체 탐지, 세분화와 같은 일반적인 컴퓨터 비전 작업에서 CNN을 능가하는 결과를 보여주었다. 그러나 뛰어난 성능에도 불구하고, 비전 트랜스포머를 실제 산업용 시각 검사에 적용하는 경우는 드물다. 이는 해당 방법들이 효과적으로 작동하기 위해서는 막대한 양의 데이터가 필요하다는 가정 때문인 것으로 보인다. 본 연구의 분석 목적은 실제 공정 이미지 자료를 통한 용접의 결함 여부를 분류하는 것이다. 제품의 표면 결함 이미지는 제품에서 용접을 기준으로 전공정과 후공정으로 나누어 20일 간의 총 31,193개 자료가 수집되었다. 그 중 1231개의 불량데이터가 있으며, 불량 데이터 Figure 1은 과적합과 미적합으로 분류된다.

2 Related work

2.1 Defect detection

결합감지는 4차 산업 혁명 시대에 가장 중요한 task 중 하나이다. 머신러닝 부터 딥러닝 까지 다양한 method 를 활용해 정확도를 올려왔고 이에따라 우리도 사용하려고 한다.

2.2 CNN approaches

AlexNet이 소개된 이후로, CNN을 활용한 object detection은 점진적으로 defect detection에 확장되기 시작했다. VGG와, Resnet을 활용한 detection은 성능이 좋았고, 여기서는 VGG-16을 classification에 사용했다. Ale 외(2018)는 RetinaNet 모델을 사용하여 도로상의 손상을 자동으로 탐지하는 방법을 제안하였다 . 데이터 증강 및 클래스 가중치 조절 등의 기술을 추가로 활용하여 기존 방법들보다 높은 정확도와 빠른 속도를 동시에 달성하였다. 위 모델은 의학 분야에서도 활용되었다. Tiwari 외(2022)는 YOLO와 RetinaNet을 양상별하여 X-선 이미지에서 COVID-19 병변을 감지하는 방법을 제안하였다

2.3 transformer approaches

더 최근에는 Transformer 모형 구조가 점진적으로 computer vision task에 사용되고 있었다. 이러한 모형들은 최근에 ViT나 Swin transformer 등에 적용되며 state-of-the-art를 만들었고 우리의 task 역시 Object detection에 주안점을 둔다.

3 Method

3.1 Data sampling

우리의 데이터셋은 over와 under 두 가지 class에 대해서 Detection을 진행한다. 1월1일부터 1월 31일 까지 3,1000개의 구리축관에 대해 결합 감지를 진행한다. 하지만 imbalance 문제로 인해 우리는 복원 추출과 비복원 추출을 함께 mini-batch에다가 적용하여 데이터셋 불균형 문제를 해결하려 한다. dataset에 따라서

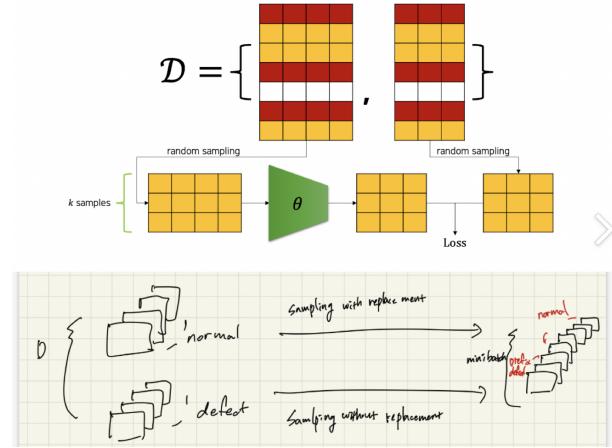


Figure 2: sampling

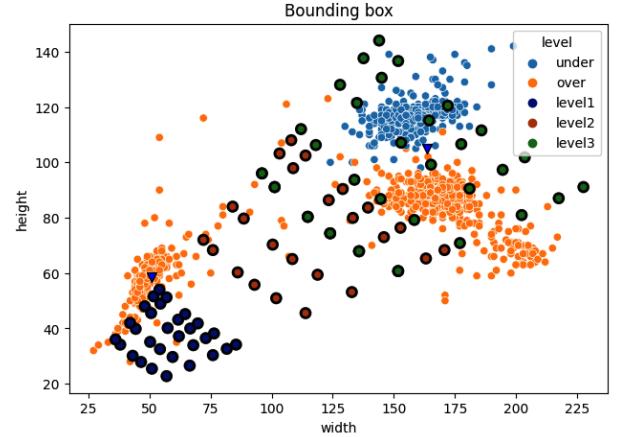


Figure 3: anchor

3.2 Anchor Generator

정확도를 향상시키기 위해 사전 bounding box를 만들었다. level 1, level 2, level 3를 나눠 anchor generator Figure 3를 통해 anchor를 생산하고 최대한 분포를 담을 수 있게 했다.

3.3 Swin Transformer

Swin transformer[1]를 활용해 우리는 모형을 적용했다. swin transformer 모형은 CNN에 비해 예측력이 높고 inductive bias 문제만 해결한다면 다양한 문제를

해결할 수 있다.

3.4 Losses

bounding box losses[2] 우리는 상대적인 좌표로 예측할 예정이기 때문에 아래와 같은 수식을 사용했다.

$$\begin{aligned}\hat{G}_x &= P_w d_x(P) + P_x \\ \hat{G}_y &= P_h d_y(P) + P_y \\ \hat{G}_w &= P_w \exp(d_w(P)) \\ \hat{G}_h &= P_h \exp(d_h(P))\end{aligned}\quad (1)$$

$$\begin{aligned}t_x &= (G_x - P_x) / P_w \\ t_y &= (G_y - P_y) / P_h \\ t_w &= \log(G_w / P_w) \\ t_h &= \log(G_h / P_h)\end{aligned}\quad (2)$$

bbox losses where $t_\star = (t_x, t_y, t_w, t_h)$

$$L1 \text{ loss} = \sum_{\star=1}^n |t_\star| \quad (3)$$

Class losses[3]

$$p_t = \begin{cases} p, & \text{if } y = 1 \\ 1 - p & \text{otherwise,} \end{cases} \quad (4)$$

$$FL(p_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t)$$

Total Losses

$$\begin{aligned}\lambda_{confidence} \sum_{i=0}^{S^2} \sum_{j=0}^B -\alpha_t(1 - p_t)^\gamma \log(p_t) \\ + \lambda_{localization} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{obj} |t_\star|\end{aligned}\quad (5)$$

3.5 overall Architecture

전체적인 architecture는 swin-L-fpn-retina[4]이다. feature pyramid network와 retina-net을 사용해서 만들었다.

4 Experiments

본 연구에서는 custom dataset으로 24,000장의 학습 데이터, 6,000의 검증 데이터 셋으로 구성되었고 결합이 있는 데이터는 1.2%이다. optimizer로 AdamW[5]를 사용하였고 초기값은 10^{-5} , weight decay 10^{-2} 를 기본값으로 사용하였다. scheduler는 cosine annealing을 사용했고 warm-up은 5,000 iterations 을 사용하였고 input size는 512×512 로 통일 하였다. backbone 모형은 ImageNet-1K를 활용하여 전이학습을 진행하였다. Augmentation으로는 HorizontalFlip과, $[0, 90]$ Range의 Rotation을 진행했다. Epoch은 50번

mini batch size	AP_{50}	Precision	Recall
defect sampling 5	93.39	80.30	95.60
defect sampling 3	96.39	83.30	98.60
defect sampling 2	92.30	81.30	93.60

Table 1: Results of Object detection validation set by different sampling

4.1 Defect sample size

mini batch 중 어느 정도의 비중으로 mini batch를 구성하는지 좋을지에 대한 연구에서는 Table 1에서 보는 것과 같이 mini batch에서 3개를 넣었을 때 가장 좋은 결과가 나왔다.

4.2 Results

Table 2 transformer based 모형을 사용하는 것이 가장 좋은 결과가 나왔다. 우리는 defect detection task 관련해서 swin transformer가 효과적인지 확인하기 위해 Resenet 관련 모형과 비교하였다. 그 결과 5가지 평가지표에 대해 모두 좋았다.

5 Conclusion

우리는 이러한 결과를 활용하여 구리 축관 결합 텁지에서도 swin transformer가 잘 작동한다는 것을 알 수 있었다. 본 감지는 3개의 레일에서 11초당 하나의 구리 축관을 생산하기 때문에 목표 FPS는 0.3이었다. 따라서

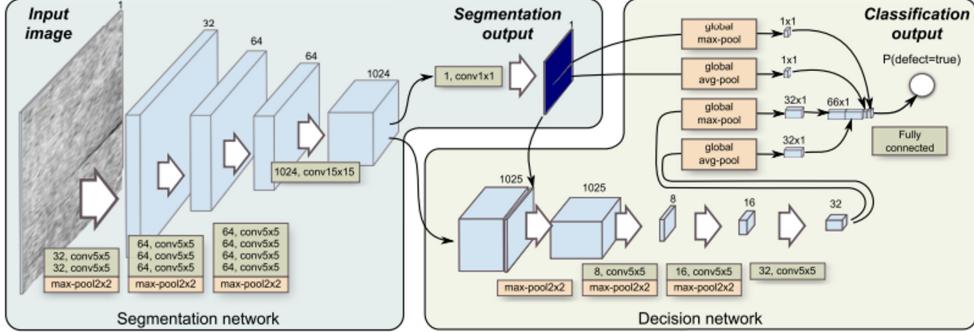


Figure 4: The Architecture. overall architecture

Model	AP50	AP_over	AP_under	Precision	Recall	params	FLOPS
ResNet152-FPN-Retina	92.59	92.59	92.59	83.03	89.23	17.5M	20G
ResNet152-PAFPN-Retina	92.06	92.06	92.06	86.68	86.84	17.5M	20G
SwinL-FPN-Retina	96.39	96.39	96.39	83.30	98.60	17.5M	20G
SwinL-PAFPN-Retina	94.46	94.46	94.46	93.35	94.52	17.5M	20G
EffNet-FPN-Retina	96.39	96.39	96.39	83.30	98.60	17.5M	20G
EffNet-FPN-Retina	96.39	96.39	96.39	83.30	98.60	17.5M	20G
ResNet152-FPN-Dense	92.59	92.59	92.59	83.03	89.23	17.5M	20G
ResNet152-PAFPN-Dense	92.06	92.06	92.06	86.68	86.84	17.5M	20G
SwinL-FPN-Dense	96.39	96.39	96.39	83.30	98.60	17.5M	20G
SwinL-PAFPN-Dense	94.46	94.46	94.46	93.35	94.52	17.5M	20G
EffNet-FPN-Dense	96.39	96.39	96.39	83.30	98.60	17.5M	20G
EffNet-FPN-Dense	96.39	96.39	96.39	83.30	98.60	17.5M	20G

Table 2: Results of Object detection validation set by different sampling

초당 FPS가 1초가 채 되지 않는 Swintransformer 모형의 적용은 충분하다.

References

- [1] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baineng Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10012–10022, 2021.
- [2] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587, 2014.
- [3] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017.
- [4] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *Proceedings of the IEEE confer-*

ence on computer vision and pattern recognition, pages 2117–2125, 2017.

- [5] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.
- [6] Jonathan Huang, Vivek Rathod, Chen Sun, Menglong Zhu, Anoop Korattikara, Alireza Fathi, Ian Fischer, Zbigniew Wojna, Yang Song, Sergio Guadarrama, et al. Speed/accuracy trade-offs for modern convolutional object detectors. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7310–7311, 2017.
- [7] Laha Ale, Ning Zhang, and Longzhuang Li. Road damage detection using retinanet. In *2018 IEEE International Conference on Big Data (Big Data)*, pages 5197–5200. IEEE, 2018.
- [8] Vinayak Tiwari, Amit Singhal, and Nischay Dhankhar. Detecting covid-19 opacity in x-ray images using yolo and retinanet ensemble. In *2022 IEEE Delhi Section Conference (DELCON)*, pages 1–5. IEEE, 2022.
- [9] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.
- [10] Jiangyun Li, Zhenfeng Su, Jiahui Geng, and Yixin Yin. Real-time detection of steel strip surface defects based on improved yolo detection network. *IFAC-PapersOnLine*, 51(21):76–81, 2018.