

직물 이미지 결함 탐지를 위한 딥러닝 기술 연구: 트랜스포머 기반 이미지 세그멘테이션 모델 실험*

이현상** · 하성호*** · 오세환****

<목 차>

I. 서론	IV. 연구 결과
II. 선행 연구	4.1 Pixel F1 Score 검증 결과
2.1 섬유 이미지 결함 탐지	4.2 mIOU 검증 결과
2.2 이미지 세그멘테이션	4.3 결함 탐지 모델 추론 결과
III. 연구 방법론	V. 결론
3.1 데이터셋	참고문헌
3.2 연구 방법론	<Abstract>
3.3 실험 환경 및 절차	

I. 서론

섬유 산업에서 직물의 결함은 제품의 상업적 가치와 소비자 만족도에 직접적인 영향을 미치기 때문에, 품질 관리는 매우 중요한 과정이라고 할 수 있다(Shahrabadi et al., 2022). 직물의 결함은 스크래치, 얼룩, 탄 자국 등 다양한 형태로 나타날 수 있으며, 이는 원료의 품질, 제조 과정 중의 기계적 문제, 또는 작업자의 실수 등 여러 원인에 의해 발생할 수 있다. 전통적인 직물 결함 관리 방식에서는 작업자가 직물을 육

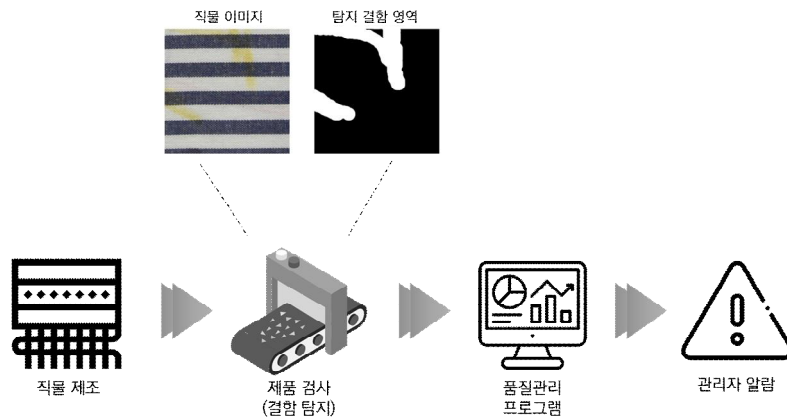
안으로 검사하여 결함을 찾아내고 분류한다. 이 방식은 작업자의 경험과 집중력에 크게 의존하며, 오랜 시간 동안 지속적인 주의를 요구하기 때문에 피로로 인한 오류가 발생할 수 있다. 또한, 결함 탐지를 위한 명확한 기준이 없기 때문에 새로운 작업자가 품질 관리를 하는 경우 큰 문제가 발생할 수 있다. 이러한 방식은 대량 생산 환경에서는 시간과 비용 면에서 비효율적일 수 있다. 이에 최근에는 컴퓨터 비전과 기계 학습 기술을 활용한 AOI(Automatic Optical Inspection) 시스템을 연구하고 있다(Voronin et

* 이 논문은 2023년도 산업통상자원부 산업혁신기반구축사업 재원으로 수행된 연구임(P0014711).

** 경북대학교 경영학부 박사과정, coolwin20@naver.com(주저자)

*** 경북대학교 경영학부 교수, hsh@knu.ac.kr(교신저자)

**** 경북대학교 경영학부 부교수, sehawno@knu.ac.kr



<그림 1> 직물 자동 결함 탐지 프로세스

al., 2021; Almeida et al., 2021; Wei et al., 2022; Liu et al., 2022; Huang et al., 2021). 이와 같이 섬유 산업에서는 다양한 빅데이터를 활용하여 연구를 수행 중이며, 국내에서도 섬유 빅데이터 플랫폼을 비롯하여 다양한 분야에서 머신러닝 시스템을 연구하고 있다(한국섬유경제신문, 2022; 김선아 등., 2017, 이동훈 등., 2020). 기존 섬유 분야의 머신러닝 기반 시스템들은 <그림 1>과 같이 고해상도 카메라와 이미지 처리 알고리즘을 사용하여 직물의 이미지를 캡처하고, 다양한 유형의 결함을 신속하고 정확하게 탐지할 수 있다.

기존 이미지 분류 방식 및 세그멘테이션 기반 직물 결함 탐지 접근법은 두 가지 측면에서 한계점이 존재한다. 첫 번째, 이미지 분류의 경우 직물의 결함에 대한 사실 여부를 알 수 있으나, 결함 객체의 위치를 특정할 수 없다. 이에 섬유 이미지 특성에 대한 정밀한 분석이 불가능하며, 시스템 활용의 측면에서도 한계점을 가지고 있다. 객체 탐지의 경우에도 결함의 위치를 파악할 수 있으나, 완전한 세부 영역을 포착하는 것은 불가능하다는 한계점을 가지고 있다.

두 번째, 기존 연구의 대부분이 트랜스포머 기반 딥러닝 이미지 분석 방법론을 활용하지 않았다. 최신 딥러닝 이미지 분석 연구에서는 ViT(Vision Transformer), DINOv2 등 트랜스포머 기법을 기반으로 고차원의 이미지 특성을 학습할 수 있는 모델을 개발했다(Dosovitskiy et al., 2020; Oquab et al., 2023). 이를 직물 이미지 결함 탐지와 같은 문제에 적용하면 기존 연구 대비 고성능의 모델을 만들 수 있을 것으로 기대된다. 본 연구는 위의 한계점을 보완하기 위해서 트랜스포머 기반의 딥러닝 이미지 세그멘테이션 모델을 실험하여, 고성능의 직물 결함 탐지 모델을 개발하는 것을 목표로 한다.

II. 선행 연구

2.1 섬유 이미지 결함 탐지

기존 연구에서는 다양한 딥러닝 기반 방법론을 활용하여 섬유의 결함을 탐지하려 했다. 직

물 결함 탐지를 자동화하는 데 있어 딥러닝 접근 방식을 활용한 연구는 식물의 복잡한 구조로 인해 자동 결함 탐지가 어려운 한계점을 극복하기 위해서 노력했다(Voronin et al., 2021). 첫 번째 단계에서는 블록 기반 알파-투팅을 사용한 새로운 결합된 지역 및 전역적 변환 도메인 기반 이미지 개선 알고리즘을 사용하여 전처리를 수행하고, 두 번째 단계에서는 식물 결함을 정확하게 탐지하기 위해 CNN(Convolutional Neural Network) 기반 아키텍처를 기반으로 한 신경망을 구축했다. 여기서 CNN이란 이미지의 지역 정보를 필터로 취합하여 특성을 학습하는 딥러닝 알고리즘을 의미한다(Lecun et al., 1998). 식물 결함 객체의 위치를 파악하기 위해서, 해당 연구는 서포트 벡터 머신(SVM)을 사용하여 결정 경계를 특정할 수 있는 새로운 방식을 개발했다. FNR(False Negative Reduction) 기법을 활용하여 기존 방법론의 정확도를 개선한 연구의 경우, CNN 아키텍처를 기반으로 식물의 결함 탐지를 시도했다(Almeida et al., 2021). FNR이라는 것은 작업자의 추가적인 검토를 가정하고 딥러닝 모델이 실제 식물의 결함을 잡지 못하는 오류를 극한으로 줄이는 방식을 의미한다. 이 경우 FP(False Positive) 오류, 즉 정상 이미지를 결함으로 분류하는 오류가 증가하지만, 이는 시스템에서 작업자의 검토로 보완이 가능하다는 주장이다. 연구 결과로써, FP 오류가 크게 증가하였음에도 50가지 유형의 결함을 95% 이상의 정확도로 탐지할 수 있었다. 식물 결함의 다중 분류 문제에서 MLMA-Net(Multi-Level Multi-Attentional Network) 기법을 활용한 연구는 결합 클래스의 데이터 수가 적거나, 동시에 다중

레이블이 존재하는 경우의 어려움을 지적했다(Wei et al., 2022). 이를 해결하기 위해, 작은 크기의 결함을 탐지하는 이미지 특성 표현(feature representations)을 개선하고, 여러 결함에 대한 고해상도 특성 맵을 활용하여 결함 상태에 대한 어텐션 레이어(attention layer)의 효과를 극대화하는 차별화된 표현을 생성하는 것을 목표로 MLMA-Net을 개발했다. 또한, 식물 결함 이미지에 대한 다중 라벨 객체 탐지 데이터셋(DHU-ML1000)을 구축하여 제안된 모델의 성능을 검증했다. 결론적으로 MLMA-Net가 더욱 뚜렷한 이미지 특성을 추출하고 실제 데이터셋에서 최신 접근법보다 더 나은 성능을 나타냈다.

딥러닝 기반 식물 결함 탐지 연구는 위의 연구처럼 이미지 분류의 접근 방법론을 활용하지만, 객체 탐지나 이미지 세그멘테이션(image segmentation) 기반 접근법을 활용하기도 한다. 딥러닝 기반 식물 결함 탐지 모델을 연구한 논문은 식물 산업에서 복잡한 형태와 다양한 종류의 식물 결함 탐지가 어려운 작업임을 제시하며, 이를 해결하기 위해 제안된 많은 방법들이 탐지 속도와 정확도가 매우 낮다고 지적했다(Liu et al., 2022). 이에 대한 해결책으로, YOLO(You Only Look Once)v4 알고리즘을 개선하여 더 높은 정확도를 가진 식물 결함 탐지 방법을 제안했다. YOLO 기법이란 이미지를 단일 신경망을 통해 처리하고, 그 결과로부터 바운딩 박스(bounding boxes)와 클래스 확률을 직접 추론하는 방식으로, 매우 빠른 속도로 객체를 탐지할 수 있기 때문에 실시간 시스템에 적합하다고 볼 수 있다. 해당 연구에서 개선된 YOLOv4 알고리즘은 MaxPool 대신 SoftPool

을 사용하는 새로운 SPP(Spatial Pyramid Pooling) 구조를 채택했으며, 기존 알고리즘 대비 mAP를 6% 개선했다. 딥러닝 이미지 세그멘테이션 모델 기반 직물 결함 탐지를 연구한 논문은 CNN 기반 다중 결함 레이블 분류 및 세그멘테이션 모델을 개발했다(Huang et al., 2021). 연구의 결과로써, 학습 수렴 속도 및 적은 수의 데이터 학습의 성능 측면에서 최신 방법론 대비 우수한 성능을 달성했다.

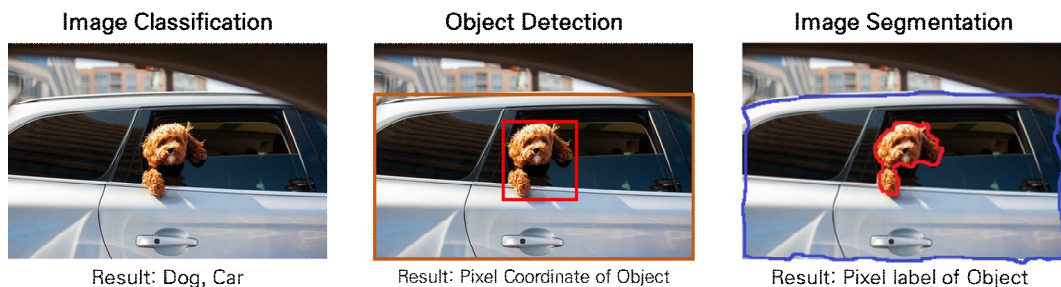
2.2 이미지 세그멘테이션

이미지 세그멘테이션이란 <그림 2>와 같이 이미지를 픽셀 수준에서 분석하여, 이미지 내의 각 객체의 정확한 형태를 파악하는 과정을 의미한다(Minaee et al., 2021). 이 작업은 객체의 "윤곽"을 정확하게 구분하며, 이를 통해 객체의 정확한 모양과 크기를 알 수 있다. 세그멘테이션은 크게 두 가지 유형이 있는데, 첫 번째는 시맨틱 세그멘테이션(Semantic Segmentation)으로 모든 픽셀을 분류하여 같은 범주에 속하는 픽셀들을 같은 라벨로 표시한다. 시맨틱 세그멘테이션은 개별 객체를 구분하지 않고, 같은 종류의 객체들을 하나의 집합으로 간주한다. 두 번째로 인스턴스 세그멘테이션(Instance

Segmentation)은 같은 범주에 속하는 개별 객체들을 서로 다른 인스턴스로 구분한다. 이미지 세그멘테이션의 경우 AI 모델이 픽셀 단위의 클래스를 예측하기 때문에 이미지 분류 및 객체 탐지 모델에 비교적 정밀한 객체의 영역을 탐지할 수 있다. 여기서 이미지 분류란 주어진 이미지가 어떤 범주에 속하는지를 결정하는 것을 의미하고, 객체 탐지란 이미지 내의 개별 객체를 식별하고, 그 위치를 박스 형태로 표시하는 것을 말한다(Lu et al., 2007; Zhao et al., 2019).

최신 이미지 세그멘테이션 모델 개발 연구에서는 주로 트랜스포머 기법을 기반으로 레이어를 구성한다(Chen et al., 2017; Xie et al., 2021; Cheng et al., 2021; Oquab et al., 2023). 트랜스포머 기법이란 NLP(Natural Language Processing) 분야에서 혁신을 가져온 모델 아키텍처로, 어텐션 메커니즘(attention mechanism)을 활용하여 입력 시퀀스의 다른 부분에 가중치를 다르게 주면서 중요한 특성에 집중할 수 있게 한다(Vaswani et al., 2017).

Deeplabv3 논문은 DCNN(Deep Convolutional Neural Network)에서 필터의 시야(field-of-view)를 명시적으로 조정하고 모델 출력 해상도를 제어하는 강력한 도구인 확장



<그림 2> 이미지 분류, 객체 탐지, 이미지 세그멘테이션

컨볼루션(atrous convolution)에 대해 연구했다(Chen et al., 2017). 이 연구는 객체를 다양한 스케일(scale)에서 세그멘테이션하는 문제를 다루기 위해 병렬 형태의 확장 컨볼루션을 사용하는 모듈을 설계했다. 또한, 다양한 확장 비율을 채택하여 멀티스케일 컨텍스트(multi-scale context)를 포착하는데 성공했다. 이 연구에서 제안하는 DeepLabv3 시스템은 이전 버전의 DeepLab을 크게 개선하고, PASCAL VOC 2012 시맨틱 이미지 세그멘테이션 벤치마크에서 다른 최첨단 모델들과 비교할 수 있는 성능을 달성했다. Segformer 연구는 트랜스포머와 다층 MLP(Multi-Layer Perceptron) 디코더(decoder)를 결합한 새로운 의미론적 분할 프레임워크를 제시했다(Xie et al., 2021). 해당 프레임워크는 두 가지 주요 특징을 가지고 있다. 첫째, 포지셔널 인코딩(positional encoding)이 필요 없는 새로운 계층적 구조의 트랜스포머 인코더(encoder)를 사용하여 다양한 이미지 스케일의 특성을 출력한다. 이는 테스트 이미지 해상도가 학습 이미지 해상도와 다를 때 포지셔널 코드의 보간을 보완하는 방식으로 성능 저하를 방지할 수 있다. 두 번째, 제안된 MLP 디코더는 모델 복잡성을 줄이고 다양한 레이어로부터 정보를 집계하여 로컬 어텐션(local attention)와 글로벌 어텐션(global attention)을 모두 결합하여 강력한 데이터 표현을 제공한다. SegFormer는 ADE20K, Cityscapes, COCO-Stuff와 같은 공개 데이터셋에서 모델 크기, 실행 시간, 정확도 측면에서 뛰어난 성능을 달성했다. Maskformer 연구는 기존의 픽셀 단위 분류 방식이 아닌 새로운 마스크 분류 방식을 제안했다(Cheng et al., 2021). 이 연구는 의미론

적 분할(semantic segmentation)과 인스턴스 분할(instance segmentation) 작업을 통합하여 단일 모델, 손실 함수, 학습 절차를 구현하는 것이 가능하다고 주장했다. 마스크 분류(mask classification)는 이미지를 n 개의 영역으로 나누고 각 영역에 대해 하나의 글로벌 클래스 레이블을 예측하는 방식이며, 트랜스포머 디코더를 사용하여 n 개의 세그먼트 임베딩(segment embedding)을 계산하고, 이를 통해 n 개의 클래스 확률 예측과 n 개의 마스크 임베딩(mask embedding)을 생성하는 레이어로 구성된다. MaskFormer는 ADE20K 데이터셋에서 55.6 mIoU(mean Intersection over Union), COCO 데이터셋에서 52.7 PQ(Panoptic Quality)의 최고 성능을 달성하며, 기존의 픽셀 단위 분류 모델보다 더 효율적임을 증명했다.

Dinov2 연구는 자연어의 사전 학습 방식이 이미지에도 효과적으로 적용될 수 있음을 주장하며, 이는 이미지를 다루는 모든 시스템에서 분포와 분야에 상관없이 적용할 수 있는 범용 이미지 특성 학습 모델인 Dinov2를 제시했다(Oquab et al., 2023). 특히, 다양한 출처에서 처리된 충분한 데이터에 대해 훈련될 때 자기 지도 학습 방법이 효과적일 수 있다고 주장했다. Dinov2 모델은 이미지와 텍스트 사전학습을 위해서 DINO(Distillation with NO labels)와 iBOT(image BERT Pretraining with Online Tokenizer) 손실 함수를 결합하고 SwAV(Swapping Assignments between Views)의 중심화를 추가하는 등의 방법으로 이미지 특성을 학습했다. 또한, 특성을 확산시키는 정규화와 소규모의 고해상도 학습 단계를 추가했다.

Ⅲ. 연구 방법론

3.1 데이터셋

본 연구에서 활용한 데이터셋은 ZJU-Leaper Dataset으로 다양한 직물 데이터 분류 및 세그멘테이션 정보를 제공한다(Zhang et al., 2020). 해당 연구의 경우 직물의 자동 결함 탐지를 위한 다중 단계 모델 및 효율적 학습의 관점에서 구축한 새로운 벤치마크 데이터셋을 제시했다. 총 데이터셋은 98,777건으로 <그림 3>과 같이 직물의 이미지와 레이블링 정보가 포함되어 있다. 직물 유형의 경우 5개 그룹으로 Stripe, Dot, Plaid, Floral 등의 다양한 범주를 포함하고 있어, 모델 학습에서의 일반화 수준이 높을 것으로 기대된다. 결함의 경우에도 얼룩, 자국, 스크래치 등 다양한 형태가 존재하며, 세그멘테이션의 경우에는 데이터셋의 레이블이 단일 클래스

로 구성되어 있어, 해당 기준으로 세그멘테이션 모델을 설계했다. 98,777건의 데이터 중 결함 레이블링이 있는 데이터는 총 26,077건으로 해당 데이터셋을 딥러닝 모델 학습용으로 활용했다. 데이터 분할의 경우 학습, 검증, 테스트 데이터셋에 22,520건, 1,186건, 2,371건으로 구성하여 딥러닝 모델 학습 및 테스트 과정을 수행했다. 데이터 전처리의 경우 같은 이름의 이미지 및 레이블 파일을 매칭하고, 레이블 데이터는 이진 클래스로 마스킹하여 딥러닝 모델에 입력했다. 또한, 이미지 해상도는 가로 448, 세로 448로 크기를 맞추고, 레이블 해상도는 가로 128, 세로 128 크기로 조정했다.

3.2 모델링

본 연구의 딥러닝 모델의 경우 비교 분석을 위해 Deeplabv3+MobileViT, SegformerB0,



<그림 3> 결함 데이터 예시

SegformerB1, Dinov2의 4가지 기법을 활용했다. Deeplabv3의 경우 CNN 기반 세그멘테이션 기법으로, 트랜스포머 기반 3가지 기법들과 비교를 위해서 함께 분석했다. 모든 분석은 파이썬을 활용하여 진행했으며, 허깅페이스(Huggingface)에서 제공하는 모델 백본(backbone)을 파이토치(Pytorch)를 통해 로드하여 분석을 수행했다. 모델 로드 후 각 모델 구조에 맞는 세그멘테이션 헤드를 설정하여 본 연구의 클래스, 입력 이미지 해상도, 출력값 형태 등을 조정했다. 하이퍼파라미터 설정의 경우 <표 1>처럼 같은 조건에서의 실험을 위해 동일해서 진행했으며, 각 모델의 학습 과정을 고려하여 세부 수치를 조정했다. Epoch는 10으로 설정하여 딥러닝 모델이 충분히 학습될 수 있도록 Earlystop 설정을 통해 Validation Loss가 최적화 되었는지 확인했고, Epoch 1회 학습에 소요되는 시간을 고려하여 Validation Step을 1000으로 설정해 Epoch 중간에 검증 및 모델이 저장되도록 설계했다. Early Stop 기능의 경우 Validation Step에서 3번 이상 Loss가 개선되지 않으면 학습을 조기 중단하도록 설정했다. 최종적으로 가중치 최적화 기법은 아담 옵티마이저(Adam Optimizer)를, 손실 함수는 이진 크로스 엔트로피(Binary CrossEntropy)로 설정했다. 학습 결과 <표 1>과 같이, Epoch 10회 범위 안에

서 Validation Loss가 최적화되어 Earlystop 기능에 따라 학습이 중단되었으며, SegformerB1 모델이 비교 모델 대비 Validation Loss 수렴 전까지 많은 수의 학습이 수행되었음을 알 수 있었다.

Deeplabv3+MobileViT 모델은 애플에서 개발한 MobileViT의 모델 백본에 Deeplabv3 세그멘테이션 헤드를 연결한 것으로 허깅페이스에서 사전학습 모델 및 전체 백본 구조를 공개하고 있다(Apple, 2021). MobileViT란 ViT 구조에서 모바일 기기에서의 사용을 위해 효율적으로 경량화한 모델로, CNN의 지역적 특성 필터 처리를 전역적으로 처리하는 방식을 활용한다(Mehta et al., 2021). 모델 버전은 MobileViT-S로 <표 2>에서와 같이, 640만 개의 가중치를 가지고 있다. SegformerB0, B1은 Nvidia에서 개발한 모델을 B0에서 B5까지 버전 별로 제공하고 있는 허깅페이스 공개 모듈을 활용했다(Nvidia, 2021). B0의 경우 약 370만 개의 가중치를 가지고 있으며, B1의 경우 약 1,400만 개의 가중치를 가지고 있다. Dinov2는 메타에서 개발한 이미지 텍스트 사전학습 모델로, 허깅페이스에서 제공 중인 모델 백본을 활용했다(Meta, 2023). 모델 버전은 ViT-S/14 distilled로 2,100만 개의 가중치를 가지고 있다.

<표 1> 하이퍼 파라미터 명세 및 학습 수렴(Earlystop) 시점

하이퍼 파라미터	수치	DeepLabv3	SegformerB0	SegformerB1	Dinov2
Batch_size	8	-	-	-	-
Epoch	10	4	5	7	4
Validation Step	1000	13000	15000	20000	12000
Learning Rate	0.0001	-	-	-	-

<표 2> 실험 모델 가중치 수

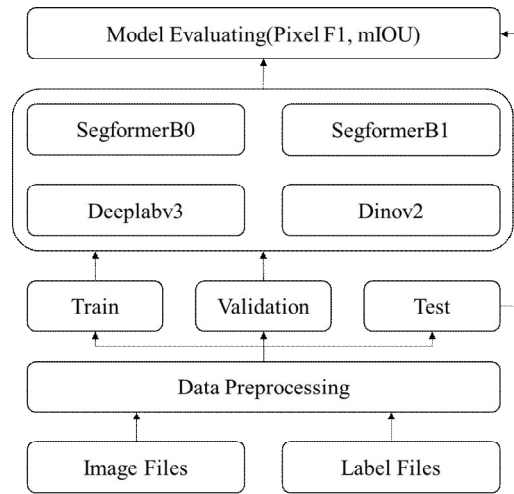
모델명	가중치 수(단위: 백만)
Deeplabv3+MobileViT	6.4
SegformerB0	3.7
SegformerB1	14
Dinov2	21

3.3 실험 설계

본 연구의 실험 절차는 <그림 4>와 같다. 이 이미지 및 레이블 파일을 가지고 해상도 조정, 이진 클래스 마스킹 등의 전처리 후 학습, 검증, 테스트 데이터셋을 구축했다. 학습 및 검증 데이터셋을 4가지 모델에 입력하여 모델 학습을 수행했고, 테스트 데이터를 가지고 픽셀 F1 Score와 mIOU를 계산하는 방식으로 모델의 성능을 평가했다. 여기서 픽셀의 F1 Score란 식 (1)과 같이 모든 레이블 픽셀에 대한 TP(True Positive), TN(True Negative), FP(False Positive), FN(False Negative) 경우의 수를 구하여 정밀도와 재현율을 구하고, 둘의 조화 평균을 계산하는 방식이다. IOU는 식 (2)와 같이 이미지 세그멘테이션에서 예측 영역과 실제 영역의 교집합으로 계산되며, 전체 경우의 수로 평균한 값을 mIOU라고 한다.

$$(1) F1\ Score = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

$$(2) IOU\ Score = \frac{pred \cap true}{pred \cup true}$$



<그림 4> 실험 절차

IV. 실험 결과

4.1 Pixel F1 Score 검증 결과

Pixel F1 Score 성능 평가 결과 <표 1>과 같이 나타났다. 결과적으로 SegformerB1 모델이 81.84%로 가장 우수한 성능을 달성했다. 픽셀 F1 Score는 식물 결함 영역을 예측하는 것 자체에 대한 중요도가 높기 때문에, SegformerB1의 결과는 식물 결함에 대해 민감하게 예측했다는 것을 알 수 있다. 또한, 픽셀 F1 Score는 결함 영역과 배경을 제대로 구분하고, 클래스 불균형

상황에서 조화 평균을 통해서 수치를 보정하기 때문에, 작물 결함 영역을 픽셀 단위에서 안정적으로 예측이 가능하다는 것으로 해석된다. SegformerB0 모델의 경우 370만 개의 가중치를 가지고 있는 모델로 성능 평가 시 효율적인 결과가 나타났으나, SegformerB1에 비해서는 다소 부족한 정확도인 77.71%로 측정되었다. Deeplabv3 모델은 일반적으로 세그멘테이션 문제에서 활용하는 CNN 기반 아키텍처 중 하나이며 MobileViT 백본을 활용했으나, 가중치가 640만 개 수준에서 370만 개인 SegformerB0 보다 다소 저조한 성능인 77.42%로 나타났다. Dinov2 모델의 경우에는 2,100만 개의 가장 큰 가중치를 가지고 있으나, 가장 낮은 성능인 62.63%로 확인되었다. 이는 모델의 과적합과 함께 작물 결함 탐지 문제의 특수성으로 인해, 이미지 텍스트 사전학습 방식이 효과적이지 못했음을 시사한다.

<표 3> Pixel F1 Score 성능 평가

모델명	Pixel F1 Score
Deeplabv3+MobileViT	0.7742
SegformerB0	0.7771
SegformerB1	0.8184
Dinov2	0.6263

4.2 mIOU 검증 결과

mIOU 성능 평가 결과는 <표 4>와 같다. Pixel F1 Score의 결과와 동일하게 SegformerB1의 mIOU 수치가 83.61%로 가장 안정적인 성능이 나타났다. mIOU의 수치는 실제 결함 영역과 예측 영역의 교집합 정도를 의미하기 때문에, 클래스 불균형을 고려하는

Pixel F1 Score 대비 전반적인 결함 탐지 정확도가 나타난다고 볼 수 있다. 전반적으로 모든 모델이 Pixel F1 Score보다 높은 수치로 안정적인 결과가 나타나지만, SegformerB1의 Pixel F1 Score가 다른 모델 대비 높은 수준과 비교하면 mIOU에서는 SegformerB1의 결과가 다소 낮다는 것이 확인된다. 이는 SegformerB1의 결과가 작물 결함 영역에 대해 다른 실험 모델들 보다 민감하게 예측했다는 것을 알 수 있다. Dinov2의 경우 Pixel F1 Score에서는 매우 저조한 성능이 나타났으나, mIOU 성능에서는 70.83%로 전반적인 결함 영역 예측은 데이터의 특성을 잘 학습한 것으로 판단된다.

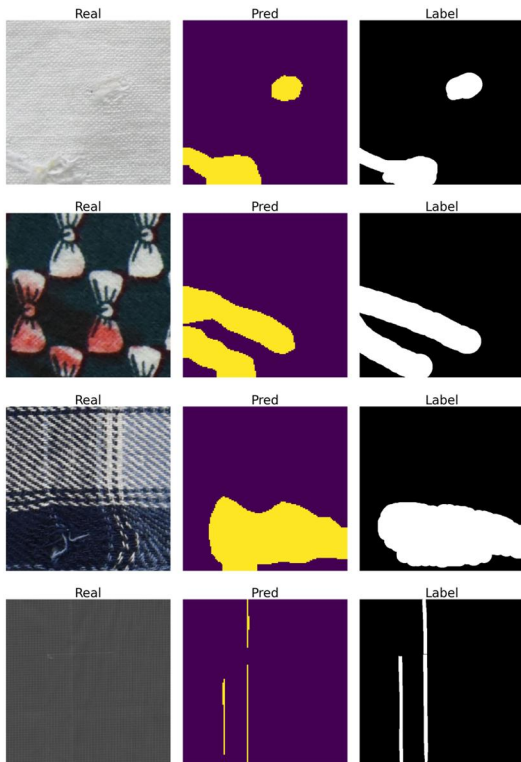
<표 4> mIOU 성능 평가

모델명	mIOU
Deeplabv3+MobileViT	0.8035
SegformerB0	0.8058
SegformerB1	0.8361
Dinov2	0.7083

4.3 결함 탐지 모델 추론 결과

<그림 5>는 본 연구의 SegformerB1 모델의 다양한 결함 탐지 예측 결과를 나타낸 것이다. 좌측의 Real 이미지가 실제 작물 이미지이고, Pred 이미지가 딥러닝 모델이 예측한 세그멘테이션 영역에 해당한다. 그리고 Label 이미지는 실제 데이터의 레이블링 세그멘테이션 영역을 나타낸다. 첫 번째 행의 이미지를 보면 흰색 평직에 놓린 자국들이 있는데, SegformerB1 모델은 해당 영역을 정확하게 예측한 것을 확인할 수 있다. 두 번째 행의 경우, 리본 무늬 패턴의 작물 이미지에 붉은색 얼룩이 있는 영역을 예

측한 결과로, 다소 직물의 패턴이 복잡하고 색상 차이가 있음에도 SegformerB1 모델이 이를 잘 예측한 것으로 나타났다. 세 번째 행은 식물 이미지 하단에 왜곡된 결함 부분이 보이는데, 레이블 영역대로 모델이 잘 예측한 것을 확인할 수 있다. 네 번째 행은 식물 결함 이미지에서 스크래치 영역을 확인할 수 있는데, 미세한 식물 결함의 영역도 모델이 성공적으로 예측한 것을 알 수 있다.



<그림 5> SegformerB1 결함 영역 예측 결과

V. 결론

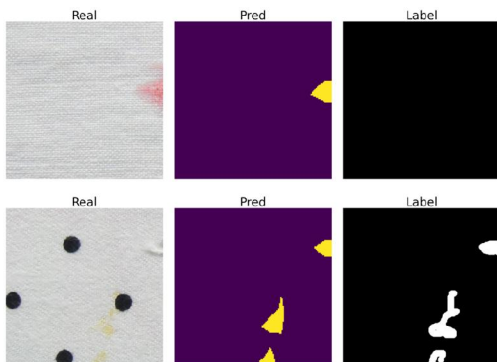
본 연구는 ZJU-Leaper 데이터셋을 활용하여

딥러닝 트랜스포머 기반 이미지 세그멘테이션 모델을 개발했다. 모델 개발을 위해서 Deeplabv3, SegformerB0, SegformerB1, Dinov2에 해당하는 4가지 딥러닝 기법을 실험했다. 그 결과 SegformerB1 모델이 Pixel F1 Score 기준 81.84%, mIOU 기준 83.61%로 가장 우수한 성능을 달성했다. SegformerB1 모델은 mIOU 보다 Pixel F1 Score가 다른 모델 대비 높은 수준으로 확인되며, 이는 본 연구에서 개발한 SegformerB1 모델이 식물 결함 영역을 민감하게 예측할 수 있다는 것으로 해석된다. SegformerB1 모델의 이미지 추론 분석 결과, 눌린 자국, 복잡한 패턴에서의 얼룩, 왜곡, 미세한 스크래치 등 다양한 패턴에서 예측할 수 있음을 확인했다.

본 연구의 시사점은 다음과 같다. 첫 번째, ZJU-Leaper 데이터셋을 활용하여 딥러닝 기반 이미지 세그멘테이션 분석의 가능성을 실증했다. ZJU-Leaper 데이터셋은 다양한 이미지 패턴에서의 눌린 자국, 얼룩, 왜곡, 스크래치 등의 결함 유형을 포함하고 있으며, 본 연구는 실험을 통해서 해당 결함 영역을 예측하는 딥러닝 모델의 성능을 검증했다. 두 번째, 트랜스포머 기반의 최신 딥러닝 이미지 세그멘테이션 모델들의 효과성을 검증했다. 결론적으로 CNN 기반의 Deeplabv3 대비 SegformerB0의 성능이 가중치가 270만 개 적음에도 불구하고 성능이 우수했으며, SegformerB1의 성능이 가장 높은 수준임을 알 수 있었다. 세 번째, 본 연구는 SegformerB1 모델을 개발하여 직물의 복잡한 패턴에서의 얼룩, 패턴 왜곡, 미세한 스크래치 등 육안으로도 제대로 영역을 파악하기 어려운 직물의 결함에 대해서도 영역을 성공적으로 예

측하는 것에 성공했다. ZJU-Leaper 데이터셋 및 본 연구의 딥러닝 접근법을 활용하면 실제 직물 공정 프로세스에서도 이를 활용할 수 있을 것으로 기대한다.

본 연구의 한계점은 다음과 같다. 첫 번째, <그림 6>를 보면 실제 데이터의 레이블 영역이 직물의 결함을 충분히 나타내지 못하며, 오히려 딥러닝 모델의 예측 영역이 직물 결함을 잘 예측한 것으로 보인다. 이는 ZJU-Leaper 데이터셋의 레이블링 품질이 다소 문제가 있으며, <그림 6>의 예시 외에도 결함 영역에 대한 세밀하지 못하거나 잘못된 레이블링 데이터들이 다수 확인되었다. 이와 같은 문제가 개선된다면 모델 예측 성능 지표의 개선과 함께 더욱 정밀하게 직물의 결함 영역을 예측할 수 있는 딥러닝 모델을 개발할 수 있을 것이다. 두 번째, 개인 연구자 단위의 컴퓨터 GPU 자원을 활용했기 때문에, base 이상 수준의 딥러닝 모델을 실험하는 것이 불가능했다. 또한, Epoch 및 Learning Rate 등 모델의 다양한 파라미터를 실험하는 것에서도 제약 사항이 존재했다. 충분한 GPU 자원을 활용한다면 Dinov2 등의 다른 최신 트랜스포머 모델의 우수한 성능을 달성할 수 있음이 기대된다.



<그림 6> 레이블링 데이터 확인

참고문헌

- 김선아, 김정원, 원동연, 최예림, “무슬림 관광객 증대를 위한 머신러닝 기반의 할랄푸드 분류 프레임워크,” 정보시스템연구, 제26권, 제3호, 2017, pp. 273-293.
- 이동훈, 김태형, “머신러닝 기법을 활용한 대졸구직자 취업 예측모델에 관한 연구,” 정보시스템연구, 제29권, 제2호, 2020, pp. 287-306.
- 한국섬유경제, “다이텍研, ‘섬유소재 빅데이터 통합 지원센터’ 섬유소재기업-바이어 ‘ON’”, 김진일 기고, 2022.06.27.
- Almeida, T., Moutinho, F., and Matos Carvalho, J. P., “Deep Learning for Fabric Defect Detection with False Negative Reduction,” *IEEE*, Vol. 9, 2021, pp. 81936-81945.
- Apple, “MobileViT + DeepLabV3 (small-sized model)”, <https://huggingface.co/apple/deeplabv3-mobilevit-small>, 2021.
- Chen, L., Papandreou, G., Schroff, F., and Adam, H., “Rethinking Atrous Convolution for Semantic Image Segmentation,” *arXiv preprint arXiv:1706.05587*, 2017.
- Cheng, B., Schwing, A., Kirillov, and A., “Per-Pixel Classification Is Not All You Need for Semantic Segmentation,” *Advances in Neural Information Processing Systems*, Vol. 34, 2021, pp. 17864-17875.
- Dosovitskiy, A., Beyer, L., Kolesnikov, A.,

- Weissenborn, D., Zhai, X., and Unterthiner, T., "Transformers for Image Recognition at Scale," *arXiv preprint arXiv: 2010.11929*, 2020.
- Huang, Y., Jing, J., and Wang, Z., "Fabric Defect Segmentation Method Based on Deep Learning," *IEEE Transactions on Instrumentation and Measurement*, Vol. 70, 2021, Article 1000315.
- LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P., "Gradient-Based Learning Applied to Document Recognition," *Proceedings of the IEEE*, Vol. 86, No. 11, 1998, pp. 2278-2324.
- Liu, Q., Wang, C., Li, Y., Gao, M., and Li, J., "A Fabric Defect Detection Method Based on Deep Learning," *IEEE*, Vol. 10, 2022, pp. 4284-4296.
- Lu, D., Weng, Q., and "A Survey of Image Classification Methods and Techniques for Improving Classification Performance," *International Journal of Remote Sensing*, Vol. 28, No. 5, 2007, pp. 823-870.
- Mehta, S., and Rastegari, M., "MobileViT: Light-Weight, General-Purpose, and Mobile-Friendly Vision Transformer," *arXiv preprint arXiv: 2110.02178*, 2021.
- Meta, "Vision Transformer (small-sized model trained using DINOv2)", <https://huggingface.co/facebook/dinov2-small>, 2023.
- Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N., and Terzopoulos, D., "Image Segmentation Using Deep Learning: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 44, No. 7, 2021, pp. 3523-3542.
- Nvidia, "SegFormer (b0-sized) model fine-tuned on ADE20k", <https://huggingface.co/nvidia/segformer-b0-finetuned-ade-512-512>, 2021.
- Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., and Fernandez, P et al., "DinoV2: Learning Robust Visual Features without Supervision," *arXiv preprint arXiv:2304.07193*, 2023.
- Shahrabadi, S., Castilla, Y., Guevara, M., Magalhães, L. G., Gonzalez, D., and Adão, T., "Defect Detection in the Textile Industry Using Image-Based Machine Learning Methods: A Brief Review," *the Journal of Physics: Conference Series*, 2022.
- Voronin, V., Sizyakin, R., Zhdanova, M., Semenishchev, E., Bezuglov, D., and Zelemskii, A., "Automated Visual Inspection of Fabric Image Using Deep Learning Approach for Defect Detection," *the Automated Visual Inspection and Machine Vision IV*, 2021.
- Wei, B., Xu, B., Hao, K., and Gao, L., "Textile

Defect Detection Using Multilevel and Attentional Deep Learning Network (MLMA-Net),” *Textile Research Journal*, Vol. 92, No. 19-20, 2022, pp. 3462-3477.

- Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J. M., and Luo, P., “SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers,” *Advances in Neural Information Processing Systems*, Vol. 34, 2021, pp. 12077-12090.
- Zhang, C., Feng, S., Wang, X., and Wang, Y., “ZJU-Leaper: A Benchmark Dataset for Fabric Defect Detection and a Comparative Study,” *IEEE Transactions on Artificial Intelligence*, Vol. 1, No. 3, 2020, pp. 219-232.
- Zhao, Z.-Q., Zheng, P., Xu, S., and Wu, X., “Object Detection with Deep Learning: A Review,” *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 30, No. 11, 2019, pp. 3212-3232.

이 현 상 (Lee, Hyun Sang)



경북대학교 경영학사와 석사학위를 취득하였다. 현재 경북대학교 박사과정을 수료했으며, 주요 관심분야는 경영정보, 딥러닝, 자연어 처리, 시계열 예측 등이다.

오 세 환 (Oh, Se Hwan)



현재 경북대학교 경영학부에서 부교수로 재직 중이다. 서울대학교 경제학부(학사)를 졸업했으며 카네기멜론대에 서 e-비즈니스 석사, 서울대학교에서 경영학(경영정보) 박사학위를 받았다. 주요 연구 관심분야는 전자상거래, 국제 무역에서의 IT 활용 등이다.

하 성 호 (Ha, Sung Ho)



연세대학교 경영학사와 한국과학기술원 석사와 박사학위를 취득하였다. 현재 경북대학교 경영학부 교수로 재직하고 있으며, 주요 관심 분야는 경영 분야에 적용되는 데이터 사이언스, 기계학습, 딥러닝 등이다.

<Abstract>

Deep Learning Models for Fabric Image Defect Detection: Experiments with Transformer-based Image Segmentation Models

Lee, Hyun Sang · Oh, Se Hwan · Ha, Sung Ho

Purpose

In the textile industry, fabric defects significantly impact product quality and consumer satisfaction. This research seeks to enhance defect detection by developing a transformer-based deep learning image segmentation model for learning high-dimensional image features, overcoming the limitations of traditional image classification methods.

Design/methodology/approach

This study utilizes the ZJU-Leaper dataset to develop a model for detecting defects in fabrics. The ZJU-Leaper dataset includes defects such as presses, stains, warps, and scratches across various fabric patterns. The dataset was built using the defect labeling and image files from ZJU-Leaper, and experiments were conducted with deep learning image segmentation models including Deeplabv3, SegformerB0, SegformerB1, and Dinov2.

Findings

The experimental results of this study indicate that the SegformerB1 model achieved the highest performance with an mIOU of 83.61% and a Pixel F1 Score of 81.84%. The SegformerB1 model excelled in sensitivity for detecting fabric defect areas compared to other models. Detailed analysis of its inferences showed accurate predictions of diverse defects, such as stains and fine scratches, within intricated fabric designs.

Keyword: Fabric Defect Detection, Deep Learning, Image Segmentation, ZJU-Leaper, Transformer, Segformer

* 이 논문은 2023년 11월 7일 접수, 2023년 11월 30일 1차 심사, 2023년 12월 17일 게재 확정되었습니다.