https://doi.org/10.7735/ksmte.2025.34.4.234

J. Korean Soc. Manuf. Technol. Eng.
ISSN 2508-5107(Online)

실시간 품질 분류를 위한 다면 영상 기반 딥러닝 시스템 개발

김누림[a], 김정인[b,*]

Development of a Deep Learning System for Real-Time Quality Classification Using Multi-View Images

Nu-Lim Kim[a], Jungin Kim[b,*]

[a] School of Electronic Engineering, Hankyong National University
[b] AI Research, Hankyong National University

ARTICLE INFO

Article history:

Received	26	June	2025
Revised	16	July	2025
Accepted	21	July	2025

Keywords:
Quality classification
Composite multi-view image
Deep learning
Lightweight model
Real-time processing

ABSTRACT

Conventional deep learning-based grading systems for agricultural products typically rely on top-view images captured by conveyor belts, limiting surface inspection. Although random rolling has been introduced to expose different sides, it often results in redundant or inconsistent coverage, reducing classification reliability. This study proposes a multi-view image acquisition and classification system utilizing the structural features of the Calistar grader. Four industrial cameras—two positioned above and two below—captured images from four distinct angles, which were merged into a single composite multiview image. This composite image was processed using a ConvNeXt V2 model, converted to ONNX, and optimized with TensorRT for real-time deployment. The experimental results demonstrated a classification accuracy of approximately 99% and a processing rate of five objects per second. These findings validate the effectiveness of the system in accurately grading agricultural products in real-time and demonstrate its potential applicability to irregularly shaped 3D objects beyond onions.

1. 서 론

최근 인공지능 기술의 발전에 따라 농산물 품질 평가 분야에서도 다양한 자동화 사례가 등장하고 있다[1]. 특히 딥러닝 기반 영상 처리 기술의 발달로, 농산물의 외관을 기준으로 품질을 평가하고 등급을 자동으로 분류하려는 연구가 활발히 진행되고 있다[2,3]. 농산물의 선별 공정은 수확 직후부터 최종 유통 단계에 이르기까지 품질을 일관되게 유지하고 소비자의 신뢰를 확보하기 위한 핵심 단계로 작용한다[4].

최근에는 이러한 공정의 자동화를 통해 인력 의존도를 줄이고, 품질 평가의 일관성을 확보하려는 노력이 이어지고 있다. 자동 선별 시스템이 실제 환경에서 성능을 발휘하기 위해서는 단순한 분류 정확도뿐 아니라, 모델의 추론 속도, 연산 효율성, 다양한 외관 특성을 고려한 처리 방식 등이 종합적으로 고려되어야 한다[5,6]. 특히 산업 현장에서는 초당 다수의 농산물을 연속적으로 처리할 수 있는 시스템 성능이 요구되며, 이는 단순히 모델을 복잡하게 만드는 방식으로 해결하기 어렵다. 복잡한 모델 구조는 높은 연산량과 메모리 사용을 동반해 실제 임베디드 시스템 환경에서는 적용이

* Corresponding author. Tel.: +82-31-610-4867
E-mail address: jik@hknu.ac.kr (Jungin Kim).

제한된다. 또한 농산물은 3차원 형태의 비정형 물체이기 때문에 일정한 위치에 특징이 고정되어 있지 않다. 예를 들어 고등어처럼 특정 부위(눈 등)의 신선도만으로 전체적인 품질 평가가 가능한 것과는 대조적으로, 양파는 색상만으로 품질 등급의 구분이 어려우며, 등급 결정에 영향을 주는 결함의 위치가 이미지마다 달라질 수 있다. 이로 인해 단일 시점이나 단일 이미지 채널만으로는 품질 관련 정보를 충분히 포착하기 어렵다[7]. 따라서 모델의 복잡성을 증가시키는 것보다는, 다양한 시점에서 이미지를 확보하거나 물체의 표면 전체를 고려할 수 있는 촬영 시스템의 도입이 더욱 효과적이라고 할 수 있다.

이러한 문제를 해결하기 위해 최근에는 농산물의 외관을 다각도에서 촬영하여 더 많은 정보를 확보하고, 이를 기반으로 분류 성능을 향상시키려는 연구들이 활발히 이루어지고 있다. 예를 들어 회전 스테이지를 이용해 다양한 각도에서 영상을 획득하거나, 복수의 카메라를 배치해 여러 방향의 이미지를 동시에 수집하는 방식이 대표적이다. 하지만 기존 연구들에는 다음과 같은 한계가 존재한다. 첫째, 동일 객체에 대한 시점 간 정보 통합이 이루어지지 않거나, 단순히 각 이미지를 독립적으로 처리하여 영상 간 상호 보완 효과를 활용하지 못하는 경우가 많다. 둘째, 높은 분류 정확도를 달성하더라도, 실제 산업 현장에서 요구되는 실시간 처리 성능이나 전체 공정 속도와의 연계가 부족한 경우가 있다. 예를 들어, Lee와 Min[8]의 연구는 고해상도 이미지에서 추출된 세부 텍스처 및 형상 정보를 활용하였으나 조명 변화에 대한 일반화 성능에 한계가 있었고, Lee 등[9]은 ONNX로 변환한 경량 모델을 Jetson Nano에서 실행하여 빠른 추론 속도를 달성했지만, 정면 이미지만을 활용하여 측면이나 후면 결함을 반영하기 어려운 한계를 보였다. Gan 등[10]은 네 개의 카메라로 획득한 독립 뷰를 각각 처리한 후 결과를 단순 합산하는 방식으로 다각도 인식을 시도했지만, 객체 간 경계가 모호할 경우 등급 혼동이 발생하는 문제가 있었다. 이처럼 다양한 시도를 통해 다각도 촬영 기법이 연구되었으나, 여전히 시점 간 정보 통합과 실시간성 확보 측면에서는 기술적 한계가 존재한다.

본 연구에서는 이러한 문제를 해결하기 위해 Calistar 선별기의 구조적 특성을 활용한 새로운 영상 획득 방식을 제안한다. Calistar 선별기는 본래 중량 측정을 위해 작물을 일시적으로 들어 올리는 기능을 수행하는 장치로, 이 들어올림 동작을 이용하면 하부 촬영이 가능하다는 점에 착안하였다. 구체적으로는, 양파가 컨베이어를 따라 이동하는 동안 상부에 위치한 두 대의 카메라로 상단 이미지를 먼저 촬영하고, 들어 올려진 상태에서 하부에 배치된 두 대의 카메라로 하단 이미지를 추가로 촬영함으로써 총 네 방향의 이미지를 수집할 수 있는 시스템을 설계하였다.

본 연구의 차별점은 다음과 같다. 첫째, 획득된 4장의 이미지를 한 장의 다면 통합 이미지(composite multi-view image)로 통합하여 CNN 기반 딥러닝 모델의 입력으로 사용함으로써, 시점 간의 정보 통합을 가능하게 하였다. 둘째, ONNX 형식으로 모델을 변환하고 TensorRT를 적용하여 추론 속도를 향상시키고 경량화를 구현함으로써, 1초당 5개 물체를 처리할 수 있는 실시간 품질 분류 성능을 달성하였다. 결과적으로, 단일 이미지 기반의 기존 품질 분류 모델과 달리, 다면 영상 정보의 통합과 효율적인 모델 구조 설계를 통해 분류 정확도와 실시간 추론 성능을 동시에 만족시키는 새로운 시스템을 제안한다.

2. 관련 연구

실제 산업 현장에서 농산물의 외관 품질을 선별하여 시장에 유통하기 위해서는, 컨베이어 벨트를 따라 이동하는 농산물을 실시간으로 촬영하고, 딥러닝 기반 영상 처리로 즉시 추론할 수 있는 경량화된 시스템이 필요하다. 최근 연구들은 모델을 최적화하고, 색상·형태·크기·표면 결함 등의 다양한 특징을 균형 있게 인식할 수 있도록 학습하는 방법에 집중하고 있다[11]. 이러한 시스템에서 높은 정확도와 실시간성을 동시에 확보하기 위해서는, 농산물의 모든 면을 반영할 수 있는 영상 데이터 확보와 함께 초당 복수의 객체를 처리할 수 있는 추론 속도가 중요하다.

2.1 농산물 품질 평가 및 선별 기술

Lee와 Min[8]은 사과 이미지를 축소하고 RGB 및 R/G/B/Gray의 다섯 채널로 분리하여 CNN 기반의 품질 등급 분류를 수행하였다. 실험 결과, Blue 채널 하나만으로도 RGB 전체와 유사한 정확도(98%)를 달성하였다. 그러나 입력 영상의 축소와 채널 분리에 따라 고해상도 텍스처 및 형상 정보가 손실되어, 미세한 표면 결함을 판별하는 데에는 한계가 있었다. 또한 실제 조명 조건과의 차이로 인해 일반화 성능 측면에서도 제약이 있었다.

Lee 등[9]은 U2-Net을 활용해 사과의 영역을 분할하고 이를 크롭하여 해상도를 조정한 이미지를 학습 데이터로 사용하였다. 이후 6종의 CNN 모델 중 가장 높은 성능을 보인 RegNet-Y를 ONNX 형식으로 변환하여 Jetson Nano에서 실행하였으며, 0.056초/장, 95%의 정확도를 보고하였다. 그러나 실제 산업 환경에서는 입력 획득, 추론, 제어 등 전체 공정의 응답 시간이 중요하므로, 단일 추론 시간만으로는 실시간 처리 능력을 평가하기 어렵다. 또한, 모든 데이터가 정면 이미지에 한정되어 있어 측면이나 후면의 결함은 탐지할 수 없는 한계가 있다.

Xu 등[12]은 산업용 CCD 카메라를 활용해 감귤을 상단에서 촬영하고, 양방향 필터링, 그레이스케일 변환, Canny 에지 검출, HSV 변환 등을 적용하여 4단계 품질 등급을 분류하는 경량 시스

템을 제안하였다. 이 연구에서는 딥러닝 기술이 적용되지 않았으며, 0.08초로 제시된 값은 초기 실행 시간으로, 실제 결과 전송과 후속 처리를 포함한 전체 지연 시간은 명확히 제시되지 않았다. 따라서 실제 산업 환경에서 요구되는 실시간 처리에는 제한이 있을 수 있음을 시사하였다.

2.2 다각도 영상 수집 및 처리 기술

Gan 등[10]은 딸기밭을 주행하는 플랫폼에 4대의 카메라를 V자 형태(지면 700 mm 높이, 60º 각도)로 배치하여, 네 방향에서 촬영된 이미지를 활용하였다. Faster R-CNN을 통해 꽃, 미숙과, 숙과를 탐지한 후, Optical Flow를 이용해 동일 객체를 추적하여 중복 계수를 제거하였다. 그러나 각 이미지를 독립적으로 처리하고 결과만 합산하는 구조이기 때문에, 과실의 숙성 단계가 변하는 중간 과정에서는 등급 분류에 혼동이 발생할 수 있었다.

Masoudi 등[13]은 회전 스테이지에 올려진 토마토를 고정된 카메라로 상·중·하 3개 위치에서 30도 간격으로 회전시켜 총 36장의 멀티뷰 이미지를 획득하였다. 수집된 영상은 SIFT 기반 3차원 구조 복원 기법과 표면 복원 알고리즘을 통해 정밀하게 재구성되었으며, 이를 통해 높은 정확도의 품질 분석이 가능하였다. 그러나 샘플당 약 75초가 소요되어, 대량 처리가 요구되는 산업 현장에서는 실시간 적용이 어렵다는 한계가 있었다.

2.3 본 연구의 차별성

본 연구에서는 기존 연구의 한계를 극복하기 위해, Calistar 선별기의 구조적 특성을 활용하여 양파의 상·하단 네 방향 이미지를 획득하고 이를 하나의 다면 통합 이미지(composite multi-view image)로 통합하여 CNN 기반 딥러닝 모델인 ConvNeXt V2의 입력으로 활용하였다. 기존 연구들은 다각도 이미지를 개별적으로 처리하거나, 정면 이미지에만 의존하는 방식이 대부분이었으나, 본 연구는 촬영된 다면 정보를 이미지 내에서 통합함으로써 위치가 일정하지 않은 비정형 3D 물체에 대해서도 패턴화된 입력 구성을 제공한다는 점에서 차별성을 가진다. 본 연구와 기존 연구의 주요 내용은 Table 1에 정리하였다.

3. 방법론

본 연구에서는 실제 산업 현장에서 적용 가능한 농산물 품질 선별 시스템을 구현하기 위해, 컨베이어 기반의 다면 영상 획득 및 딥러닝 추론 시스템을 설계하였다. 초기 설계한 컨베이어 벨트를 따라 이동하는 원물이 상부 카메라에 의해 고정된 자세로 촬영되는 구조의 기본 시스템은 다음과 같은 두 가지 문제점을 내포하고 있다.

첫째, 원물의 전체 표면 정보를 획득하지 못한다는 점이다. 고정된 자세로 이동하면서 상부만 촬영되기 때문에, 하부나 후면 등 비노출면의 결함은 탐지할 수 없다. 이를 보완하고자 원물을 컨베이어 상에서 무작위로 굴려 촬영하는 방식도 시도되었으나, 이 방식은 표면 전체가 고르게 촬영되지 않거나 특정 면만 반복적으로 촬영되는 문제가 있었다. 예를 들어 하나의 원물은 동일한 면이 중복 촬영될 수 있으며, 다른 원물은 네 면이 각각 한 번씩 촬영될 수 있어 영상 데이터 간 구성의 일관성이 확보되지 않았다.

둘째, 기존 시스템의 추론 속도는 실제 산업 현장의 컨베이어 벨트 속도를 따라가지 못하였다. 특히 Python 기반 모델은 고속

Table 1 Comparison of related works

Related work	Target object	Input type / View	Processing speed	Core technique	Limitations
Lee et al., (2022)	Apple	Downscaled front view; R/G/B/Gray channels	-	CNN with separated color channels	Loss of fine texture and shape; limited generalization in real lighting
Lee et al., (2024)	Apple	Cropped and resized front images	0.056 sec/image	RegNet-Y, ONNX Runtime	Only single inference time measured; no side/rear defect detection
Xu et al., (2025)	Citrus	Top-down single view (industrial CCD camera)	0.08 sec (init only)	Lightweight with filtering techniques	No deep learning; actual latency may exceed industrial requirements
Gan et al., (2020)	Strawberry	4-view (V-shape camera setup)	-	Faster R-CNN + optical flow	No maturity phase tracking across views
Masoudi et al., (2024)	Tomato	36 multi-view via rotation platform	~75 sec/sample	3D reconstruction with SIFT	Too slow for real-time; impractical for mass throughput
Proposed method	Onion	Composite image from 4-directional views	105–124 ms/ 5 samples	ConvNeXt V2; real-time optimized model	Multi-view integration with real-time inference; robust to 3D variability

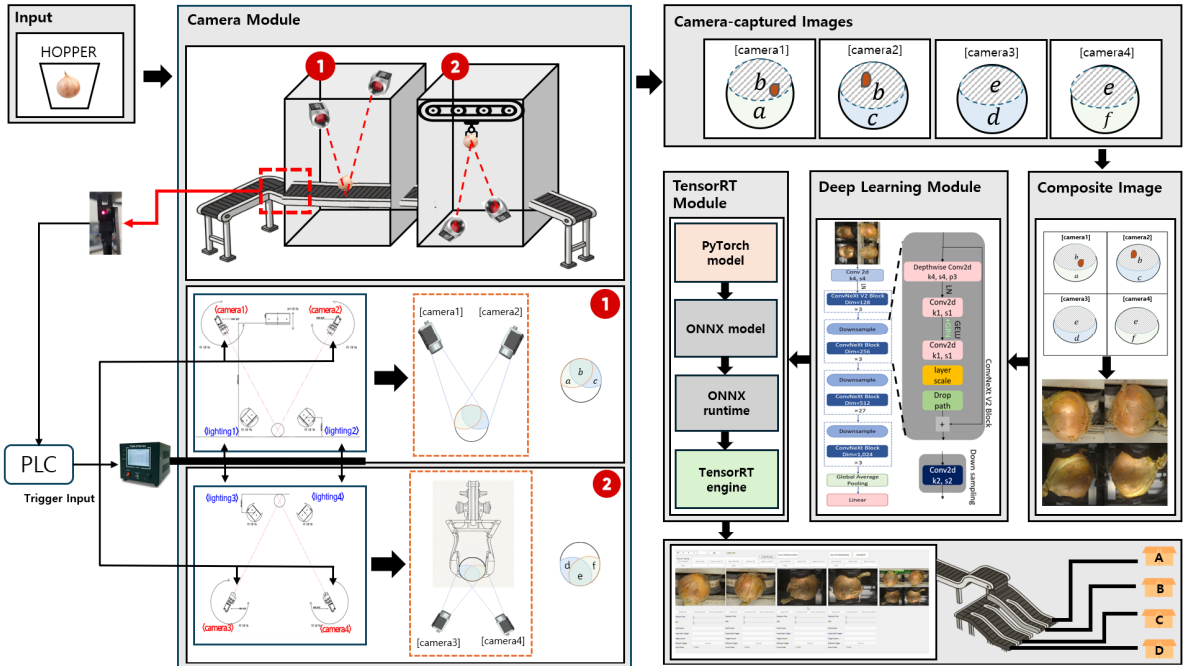


Fig. 1 System structure

처리가 요구되는 환경에서 실시간 응답 성능이 부족하였으며, 추론 시간의 편차도 컸다. 이를 개선하기 위해 ONNX Runtime과 TensorRT 기반의 고속 추론 구조를 도입하였다.

이러한 문제를 해결하고자 본 연구에서는 Fig. 1과 같이 전체 시스템 구조를 설계하고, 촬영 시스템, 다면 통합 이미지 구성, 딥러닝 모델 적용, 소프트웨어 구현, 실시간 최적화까지 일련의 과정을 단계적으로 수행하였다.

3.1 시스템 구성 개요

- Input: 원물은 호퍼(hopper)를 통해 시스템에 투입되며, 컨베이어 벨트를 따라 자동으로 이송된다. 대상은 정렬되지 않은 상태로 연속적으로 유입되며, 카메라 촬영 위치로 이동한다.
- Camera Module: 원물이 컨베이어 상에서 센서를 통과하면 트리거 신호가 PLC로 전달되고, PLC는 컨베이어 속도를 기준으로 촬영 위치에 도달하는 시점을 계산하여 카메라 및 조명 제어기에 동기화된 신호를 전달한다. 총 4대의 카메라(상부 2대, 하부 2대)가 사용되며, 각기 다른 시점에서 원물의 상단 및 하단 정보를 획득한다.
- Composite Multi-View Image & Deep Learning Module: 획득된 4장의 이미지는 다면 통합 이미지(composite multi-view image)로 구성되며, 다양한 시점 정보를 포함한 하나의 입력 영상으로 통합된다. 해당 이미지는 ConvNeXt V2 기반 딥러닝 분류 모델의 입력으로 사용된다.
- TensorRT Module: 학습은 PyTorch 환경에서 수행되고, 추론은 ONNX 형식으로 변환된 모델을 기반으로 ONNX Runtime과 TensorRT를 활용하여 고속화된다.
- Classifier: 최종 품질 등급(A, B, C, D)은 추론 결과를 기반으로 자동 분류되며, 등급별 컨베이어 라인을 통해 물리적으로 선별된다.

3.2 다면 촬영 시스템 설계

원물의 모든 면을 안정적으로 인식하고 평가하기 위해 Calistar 선별기의 구조를 응용하였다. 기존 방식은 정면 이미지만 활용하거나, 무작위 회전을 통해 다양한 시점을 촬영하고자 하였으나 회전 방향의 불확실성, 결함 부위 누락, 물리적 손상 등의 문제를 야기했다.

특히 단일 이미지 기반 평가 방식은 다음과 같은 한계를 지닌다. 예를 들어 동일한 결함이 두 시점에서 반복 촬영되면, 중복 결함으로 인식되어 낮은 등급이 부여될 수 있으며, 반대로 한쪽 면의 심각한 결함이 다른 시점 이미지와 평균화되어 품질 등급이 왜곡되는 경우도 발생할 수 있다.

이에 따라 본 연구에서는 Fig. 2에 제시된 바와 같이 Calistar 선별기의 무게 측정 동작을 활용하여, 원물을 들어 올린 상태에서 상·하단을 모두 노출시켜 촬영하는 방식으로 구조를 설계하였다.

3.3 카메라 및 조명 시스템 구성

Fig. 3은 상·하단 카메라 및 조명 시스템의 설계를 나타낸다. 고해상도 산업용 카메라 4대를 사용하였으며, 조명 균일성을 확보

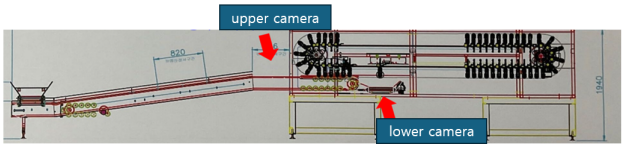


Fig. 2 Calistar sorter

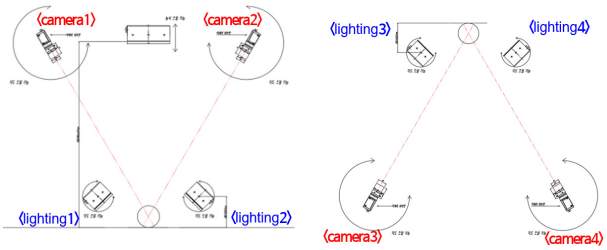


Fig. 3 Design and implementation of upper/lower camera and lighting system

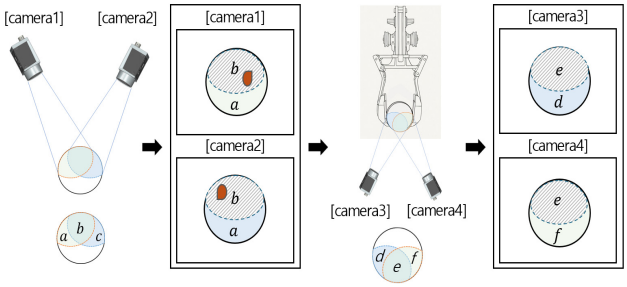


Fig. 4 Complete surface imaging method using upper and lower cameras

하기 위해 조명 제어 시스템을 별도로 설계·구현하였다. 이 구성은 카메라와 조명의 시점 정합을 통해 고품질 영상 데이터를 안정적으로 획득할 수 있도록 한다.

3.4 다면 통합 이미지 생성

4대의 카메라로부터 획득된 다각도 이미지는 Fig. 4와 같이 각기 다른 시점을 포함하고 있으며, Fig. 5에서는 이를 하나의 다면 통합 이미지로 구성하는 과정을 시각화하였다. 상부 카메라(camera 1, 2)는 상단 좌우 영역과 중복 영역을, 하부 카메라(camera 3, 4)는 하단 좌우 및 중심 영역을 각각 촬영한다. 이를 하나의 수평 방향 통합 영상으로 구성함으로써, 사람의 시점처럼 원물의 모든 면을 종합적으로 인식할 수 있는 구조를 구현하였다. 이러한 구성은 3차원 비정형 물체에 대해 일정한 시각적 패턴을 형성하며, ConvNeXt V2와 같은 CNN 기반 모델이 시각적 특징을 효과적으로 학습할 수 있는 입력을 제공한다.

3.5 ConvNeXt V2 기반 품질 분류 모델

ConvNeXt V2[14]는 전통적인 CNN 구조를 기반으로, ViT의

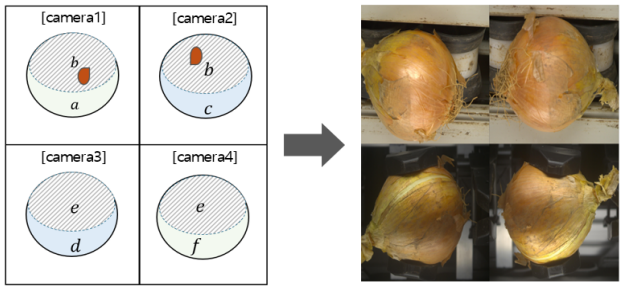


Fig. 5 Composite multi-view image generation

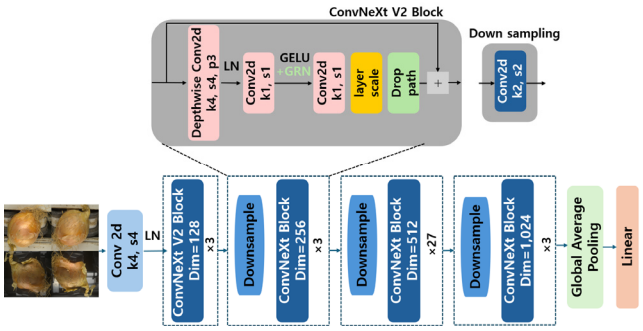


Fig. 6 ConvNeXt V2 architecture

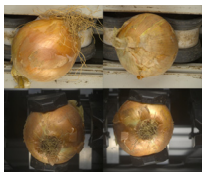
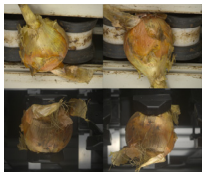
MAE(Masked AutoEncoder) 개념을 확장한 FCMAE(Fully Convolutional MAE) 방식으로 사전 학습된다. 입력 이미지의 일부를 마스킹하고 나머지 정보로 복원하는 학습 과정을 통해 표현 능력을 강화하였으며, GRN(Global Response Normalization) 레이어를 도입해 feature collapse를 방지하고 특징 표현의 다양성을 유지한다.

본 모델은 ImageNet-1K 벤치마크 기준 Top-1 정확도 88.9%를 기록하며, 당시 분류 분야에서 SOTA(state-of-the-art) 성능을 달성한 모델이다. 이는 ResNet-152×4, EfficientV2-XL, Swin Transformer-L 등 주요 CNN 및 트랜스포머 계열 모델들보다 우수한 성능으로, FCMAE(Fully Convolutional Masked AutoEncoder) 기반의 표현 학습 구조와 GRN(Global Response Normalization)을 통해 강력한 특징 추출 능력을 갖춘 것이 특징이다. 본 연구는 품질 등급 분류(classification) 문제를 다루므로, 높은 정확도와 효율성을 고려하여 ConvNeXt V2를 백본으로 채택하였으며, Fig. 6에 나타난 모델 구조를 기반으로 다면 통합 이미지를 입력으로 하는 학습을 수행하였다.

3.6 학습 데이터셋 구성

본 연구에서는 4장의 시점 이미지를 통합한 다면 통합 이미지 기반 RGB 데이터셋을 구축하였다. 총 6,329장의 등급별 양파 데이터를 사용하였으며, 학습/검증/테스트 세트는 각각 4,000장 / 800장 / 1,529장으로 구성되었다.

Table 2 Training dataset

Class (grade)	Train	Valid	Test	Sample image
A	1,000	200	453	
B	1,000	200	337	
C	1,000	200	337	
D	1,000	200	402	
Sum	4,000	800	1,529	6,329

등급은 시장에 실제로 유통되는 양파 등급 규격을 반영하여, AI 기반 자동 선별기에 적용 가능한 표준 규격의 개선 방안을 제시하였다. 상품 출하가 가능한 A등급, 결점이 존재하지만 식용 가능하며 시기에 따라 상품과로 전환될 수 있는 B등급, 심한 결점으로 상품성은 없지만 식용 가능한 C등급, 그리고 식용으로도 사용할 수 없는 폐기용 D등급으로 분류하였다[15]. (Table 2)

1) Grade A: 품종 고유의 완전한 대칭 형태와 선명한 색·우수한 광택을 지니며, 중결점 0%, 경결점 ≤ 5%인 양파
2) Grade B: 품종 고유의 거의 대칭 형태와 선명한 색·양호한 광택을 지니며, 중결점 0%, 경결점 ≤ 10%인 양파
3) Grade C: 품종 형태가 다소 비대칭이고 색·광택이 Grade B 기준에 미달하며, 중결점 ≤ 5%, 경결점 ≤ 20%인 양파
4) Grade D: 품종 형태가 심하게 비대칭이거나 색·광택이 부족하고, 흙이나 이물질이 일부 잔존할 수 있으며, 중결점 > 5% 또는 경결점 > 20%인 양파

3.7 품질 인식 소프트웨어 시스템

실제 산업 환경에 적용하기 위해, 카메라 제어, 영상 수집, 딥러닝 추론, 결과 표시를 하나의 파이프라인으로 통합한 소프트웨어를 개발하였다. C++ 기반 사용자 인터페이스와 Python 기반 딥러닝 추론 엔진을 연동하였으며, Fig. 7은 품질 선별기와 연결된 실행 화면을 나타낸다. 촬영된 원물의 이미지는 TCP/IP 통신을 통해 선별기 제어 프로그램에 전송되며, 각 카메라 이미지와 추론 결과가 실시간으로 표시된다.

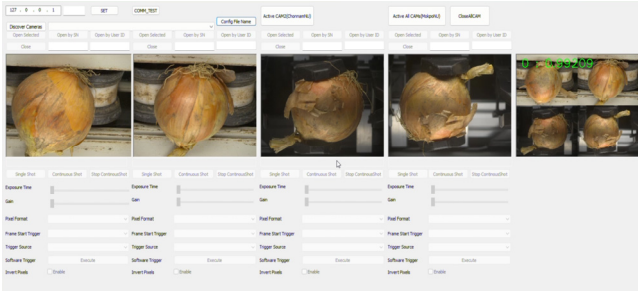


Fig. 7 User interface of the quality grading software

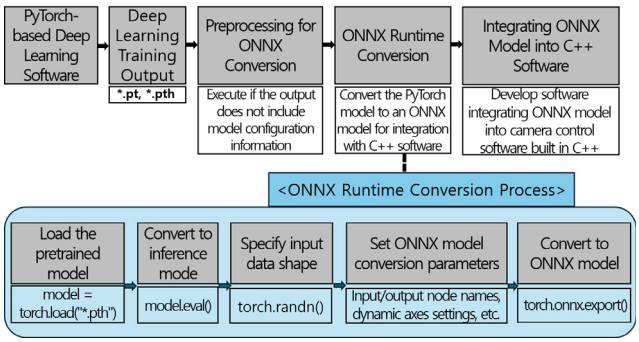


Fig. 8 ONNX model conversion and integration pipeline

고속 처리를 위해 학습된 PyTorch 모델을 Fig. 8과 같이 ONNX 형식으로 변환하고, ONNX Runtime을 통해 추론이 가능하도록 구성하였다. ONNX Runtime 기반 시스템은 추론 시간 편차(최대 392 ms)로 인해 컨베이어 속도를 안정적으로 추적하지 못하는 문제가 있어, TensorRT 엔진을 적용하여 ONNX 변환 모델을 최적화하였고, 추론 응답 속도의 안정성과 평균 처리 속도를 개선하였다[16].

4. 실험 및 성능 평가

본 장에서는 제안된 시스템의 성능을 검증하기 위해 두 가지 실험을 수행하였다. 첫 번째는 모델의 분류 정확도와 표현 특성에 대한 정량적 평가이며, 두 번째는 통합 시스템의 실시간 처리 성능을 확인하기 위한 실험이다. 실험은 ConvNeXt V2 기반의 딥러닝 분류 모델에 다면 통합 이미지(composite multi-view image)를 입력으로 사용한 학습 결과와, 산업 현장에서 요구되는 실시간 품질 선별 조건을 기준으로 진행되었다.

4.1 정량적 분류 성능 평가

제안된 모델의 품질 분류 성능을 정량적으로 평가하기 위해

Table 3 Performance metrics by quality grade

Metrics	A	B	C	D	Total
Recall	1.00	0.98	0.98	1.00	0.99
Precision	1.00	0.99	0.97	1.00	0.99
Accuracy	0.99				

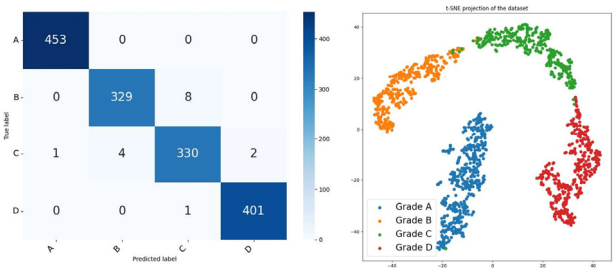


Fig. 9 Classification results: confusion matrix and t-SNE plot

Confusion Matrix 및 t-SNE 기반 시각화 기법을 활용하였다. Table 3은 각 등급별 정확도, 정밀도, 재현율 등을 요약한 결과이며, 전체 분류 정확도는 0.99로 나타났다. 모든 클래스(A, B, C, D)에 대해 정밀도 및 재현율이 0.97 이상으로 확인되었으며, 안정적인 분류 성능을 보였다.

Fig. 9는 t-SNE를 이용해 고차원 특징 표현을 2차원 공간에 시각화한 결과로, 각 품질 등급이 서로 명확하게 분리된 군집을 형성하는 것을 확인할 수 있었다. 이는 ConvNeXt V2 모델이 각 클래스의 특성을 효과적으로 학습하였음을 의미한다. 한편, Fig. 9의 t-SNE 시각화 결과에서는 군집 내에 일부 다른 등급이 혼재되어 나타나는 현상이 관찰되었다. 이는 첫째, 등급 분류 학습 과정에서 사용된 cross-entropy 손실 함수가 클래스 간 유사성이나 거리 정보를 명시적으로 반영하지 않기 때문에 오분류가 발생했을 가능성이 있으며, 둘째, 실제 수작업 분류 과정에서는 육안으로도 등급 간 구분이 모호한 샘플이 존재할 수 있기 때문이다.

4.2 실시간 추론 성능 실험

본 실험에서는 제안한 시스템의 실시간 품질 선별 가능 여부를 검증하였다. 실험은 ONNX Runtime과 TensorRT 엔진을 적용한 최종 통합 시스템 상에서 수행되었으며, 카메라 제어부터 등급 분류까지 전 과정을 포함한 응답 시간과 추론 처리 속도를 측정하였다. 본 시스템에서 사용된 카메라는 200 ms 간격으로 작동하며, 1초에 최대 5장을 촬영할 수 있다. 취득된 원본 이미지의 해상도는 979 × 819 pixels이며, RGB 3채널(24-bit color depth) 형식으로 저장된다. 해당 해상도는 실시간 추론 성능과 처리 안정성을 고려하여 설정된 값으로, 필요 시 해상도 조정이 가능하나 모델의 분류 정확도에는 유의미한 영향을 미치지 않음이 실험을 통해 확인되었다. Server 환경은 Table 4에 정리하였으며, 실험 조건은 아래와 같다.

Table 4 Server hardware specifications

Spec.	CPU	GPU
Server	AMD Ryzen 7 9700X 8-Core	RTX 4080 SUPER

Table 5 Inference time and latency comparison by platform

Metric	Total inference time range	Latency variance
Python-based	154 ms ~ 423 ms	269 ms
ONNX Runtime	32 ms ~ 392 ms	360 ms
TensorRT	25 ms ~ 42 ms	17 ms

- 양파 70개를 사용하여 총 5회 반복 측정
- 컨베이어 속도는 1초에 5개 처리 기준으로 설정

선별 대상물의 촬영은 IR 센서를 통해 자동으로 트리거되므로, 컨베이어 이송 속도가 증가하더라도 촬영 누락이 발생할 가능성은 매우 낮다. 단, 선별기의 물리적 작동 한계로 인해 1초당 5개를 실시간 처리 가능한 기준으로 설정하였다.

사전 테스트 결과, ONNX Runtime만을 사용할 경우 추론 시간의 편차가 크게 나타났으며, 일부 샘플은 200 ms를 초과하는 처리 시간이 측정되었다. 반면, TensorRT를 적용한 경우 모든 추론이 50 ms 이내에서 안정적으로 수행되었다.

Table 5는 Python 기반 모델, ONNX Runtime, TensorRT 세 가지 추론 플랫폼에 대해 전체 추론 시간 범위와 지연 시간 편차(latency variance)를 비교한 결과를 나타낸다. 실험 결과, Python 기반 모델은 평균 추론 속도는 느리고 시간 편차도 큰 반면, ONNX Runtime은 처리 속도는 향상되었으나 여전히 지연 편차가 큰 것으로 나타났다. 반면, TensorRT는 25~42 ms 사이의 안정적인 추론 속도를 유지하며 가장 낮은 시간 편차(17 ms)를 기록하였다. 이는 본 연구에서 적용한 TensorRT 기반 최적화 기법이 실시간 처리 성능과 예측 일관성 측면에서 산업 환경에 가장 적합함을 보여준다.

4.3 중복 촬영 여부 및 처리 속도 평가

첫 번째 실험에서는 70개의 샘플에 대해 5회 반복 측정을 통해 다면 통합 이미지 생성 과정에서 동일 면이 중복 촬영되는지 여부를 검증하였고, 중복 없이 4개의 시점 영상이 안정적으로 수집됨을 확인하였다.

두 번째 실험에서는 하나의 양파에 대해 등급 분류에 소요되는 전체 시간을 측정하였다. 추론만을 기준으로 할 때 소요 시간은 25~42 ms로 측정되었으며(Table 6), 전체 처리 파이프라인을 포함한 시간은 105~124 ms로 나타났다. 이는 산업 현장에서 요구하

Table 6 Elapsed time summary for each processing stage

System	CAM1	CAM2	CAM3	CAM4	Merged image write	OnPaint
Elapsed time (ms)	18~19			54~72	11	22
Total time (ms)	105~124					

Table 7 Server time profiling results

	Time-1 (ms)	Time-2 (ms)	Time-3 (ms)
Server	25~42	54~72	105~124

는 200 ms 이하 처리 조건을 안정적으로 충족함을 의미한다.

4.4 시스템 구간별 응답 시간 분석

Table 7은 전체 처리 과정을 세 가지 시간 구간으로 나누어 요약한 것이다. 실험 결과, 모든 구간에서 200 ms 이내로 처리가 완료되어, 1초당 5개 원물 처리 요건을 충족함을 확인하였다.

- Time-1: 추론에만 소요된 시간
- Time-2: 마지막 카메라(CAM4) 동작 후 선별기로 전송되기까지의 시간
- Time-3: 하나의 원물이 투입되어 품질이 선별되기까지의 전체 처리 시간

4.5 기존 연구와의 비교 분석

양파는 형태와 표면 상태의 편차가 크고, 정형화된 등급 기준이 부족하여 수작업 기준에서도 선별 일관성이 떨어지는 특성이 있다. 따라서, 동일한 작물을 대상으로 한 실용화 사례와의 비교는 제안 시스템의 산업 적용 가능성을 평가하는 데 중요한 기준이 된다. Song 등[17]은 컨베이어 기반 환경에서 동작하는 양파 품질 선별 시스템을 제안하였으며, 단일 시점 이미지를 활용해 정확도 86.37%, 처리 시간 55~500 ms를 보고하였다. 반면, 본 연구에서는 다면 통합 이미지(composite multi-view image)를 기반으로 CNN 분류 모델을 구성하고, TensorRT 최적화를 통해 정확도 99%, 처리 시간 105~124 ms의 일관된 성능을 달성하였다. 이는 동일한 작물과 유사한 환경에서 수행된 연구와 비교하여 정확도와 처리 속도 모두에서 향상된 결과로, 실제 산업 환경에서의 적용 가능성과 기술적 우수성을 동시에 확인할 수 있다.

5. 결 론

본 연구에서는 작물의 전체 표면 정보를 반영할 수 있는 영상 기반 품질 선별 시스템을 제안하였다. 기존의 산업용 자동 선별 시스템은 제한된 시야에서 획득한 이미지 또는 정면 시점의 영상만을 활용함으로써, 품질 판단의 정밀도 및 일관성 측면에서 한계를 지녀왔다. 이를 해결하기 위해 본 연구에서는 Calistar 선별기의 구조적 특성을 활용하여 상·하부 네 방향에서 영상을 획득하고, 이를 하나의 다면 통합 이미지(composite multi-view image)로 구성한 후 ConvNeXt V2 기반 딥러닝 모델의 입력으로 활용하였다.

실험 결과, 제안된 시스템은 약 99%의 품질 분류 정확도를 달성하였으며, 초당 5개 작물을 처리할 수 있는 추론 속도를 통해 실제 산업 환경에서 요구되는 실시간 대응 성능을 충족함을 확인하였다. 이는 다각도 정보 통합과 경량화 추론 구조가 결합된 시스템이 현장 적용 가능성을 갖추었음을 의미한다.

다만, 본 시스템은 고성능 서버 환경에서 운용되었기 때문에, 향후 저전력 임베디드 엣지 컴퓨팅 환경에서는 연산 자원의 제약으로 실시간 처리에 한계가 발생할 수 있다. 이에 따라 후속 연구에서는 딥러닝 모델의 경량화, 연산 구조의 최적화, 통신 지연 최소화 등을 통해 다양한 하드웨어 조건에서도 안정적으로 운용 가능한 시스템으로 확장할 계획이다.

아울러, 본 연구에서 제안한다면 통합 이미지 기반의 품질 인식 구조는 양파와 같은 구형 작물에만 국한되지 않으며, 표면이 불규칙하거나 형상이 비정형적인 3차원 물체에 대한 이상 탐지 및 품질 평가에도 활용 가능하다. 이러한 확장성을 바탕으로, 제안된 기술은 농업을 넘어 제조, 물류, 식품 가공 등 다양한 산업 분야에서도 적용 가능한 범용 품질 인식 시스템으로 발전할 수 있을 것으로 기대된다.

후 기

본 결과물은 농림축산식품부의 재원으로 농림식품기술기획평가원의 고부가가치식품기술개발사업의 지원을 받아 연구되었음(RS-2022-IP322054).

References

[1] Kamilaris, A., Prenafeta-Boldú, F. X., 2018, Deep Learning in Agriculture: A Survey, Comput. Electron. Agric., 147 70-90, https://doi.org/10.1016/j.compag.2018.02.016.

[2] Albahr, M., 2023, A Survey on Deep Learning and Its Impact on Agriculture: Challenges and Opportunities, Agriculture, 13:3 540, https://doi.org/10.3390/agriculture13030540.

[3] Hayat, A., Morgado-Dias, F., Choudhury, T., Singh, T. P., Kotecha, K., 2024, FruitVision: A Deep Learning Based

Automatic Fruit Grading System, Open Agric., 9:1 20220276, https://doi.org/10.1515/opag-2022-0276.

[4] Reachana, N., Mardy, S., 2024, The Critical Role of Post-Harvest Handling in Vegetables, Jurnal Semesta Ilmu Manajemen dan Ekonomi (J-SIME), 1:2 59-69, https://doi.org/10.71417/j-sime.v1i2.79.

[5] Qin, Y.-M., Tu, Y.-H., Li, T., Ni, Y., Wang, R.-F., Wang, H., 2025, Deep Learning for Sustainable Agriculture: A Systematic Review on Applications in Lettuce Cultivation, Sustainability, 17:7 3190, https://doi.org/10.3390/su17073190.

[6] Xiao, F., Wang, H., Xu, Y., Zhang, R., 2023, Fruit Detection and Recognition Based on Deep Learning for Automatic Harvesting: An Overview and Review, Agronomy, 13:6 1625, https://doi.org/10.3390/agronomy13061625.

[7] Rojas Santelices, I., Cano, S., Moreira, F., Peña Fritz, Á., 2025, Artificial Vision Systems for Fruit Inspection and Classification: Systematic Literature Review, Sensors, 25:5 1524, https://doi.org/10.3390/s25051524.

[8] Lee, E. J., Min, J. I., 2022, Explore Influential Color Channels for Quality Classification of Apples Based on CNN, Journal of Digital Contents Society, 23:8 1477-1484, https://doi.org/10.9728/dcs.2022.23.8.1477.

[9] Lee, J.-H., Lee, S.-C., Nguyen, B. N. H., Lee, J., Kwon, G.-J., Kim, J.-Y., 2024, Research on Exploring Optimal Deep Learning Models for Apple Quality Prediction Based on Embedded Systems, Journal of Korean Institute of Information Technology (JKIIT), 22:5 61-70, https://doi.org/10.14801/jkiit.2024.22.5.61.

[10] Gan, H., Lee, W. S., Peres, N., Fraisse, C., 2020, Development of a Multi-Angle Imaging System for Automatic Strawberry Flower Counting, Proc. 16th International Symposium on Artificial Intelligence and Mathematics(ISAIM), 1-8.

[11] Chuquimarca, L. E., Vintimilla, B. X., Velastin, S. A., 2024, A Review of External Quality Inspection for Fruit Grading Using CNN Models, Artif. Intell. Agric., 14 1-20, https://doi.org/10.1016/j.aiia.2024.10.002.

[12] Xu, M., Zhang, X., Zhan, C., Ge, J., Yang, H., 2025, Research on Citrus Grading System Based on Machine Vision, Syst. Sci. Control Eng., 13:1 2460443, https://doi.org/10.1080/21642583.2025.2460443.

[13] Masoudi, M., Golzarian, M. R., Lawson, S. S., Rahimi, M., Islam, S. M. S., Khodabakhshian, R., 2024, Improving 3D Reconstruction for Accurate Measurement of Appearance Characteristics in Shiny Fruits Using Post-Harvest Particle Film: A Case Study on Tomatoes, Comput. Electron. Agric., 224 109141, https://doi.org/10.1016/j.compag.2024.109141.

[14] Woo, S., Debnath, S., Hu, R., Chen, X., Liu, Z., Kweon, I. S., Xie, S., 2023, viewed 7 July 2025, ConvNeXt V2: Co-designing and Scaling Convolutional Networks with Masked Autoencoders, <https://arxiv.org/abs/2301.00808>.

[15] Kim, E. J., Ju, S. H., Go, Y., Kwon, Y. S., Na, H. Y., 2024, Improvement Strategies for Onion Quality Grading Standards Based on Domestic Distribution and Consumption Trends and AI-Based Automatic Sorting Machine Applications, Journal of the Korean Society of International Agriculture, 36:4 326-332, https://doi.org/10.12719/KSIA.2024.36.4.326.

[16] Nassef, L., Tarabishi, R. A., Abo Alnasor, S. A., 2024, Benchmarking NLP and Computer Vision Models on Domain-Specific Architectures: Standard vs. TensorRT-Optimized Performance, J. Electr. Syst., 20:11s 736-752, https://doi.org/10.52783/jes.7272.

[17] Song, Z., Buayai, P., Makino, K., Mao, X., 2025, Feature-adaptive Anomaly Detection Model for Onion Inspection System, Smart Agric. Technol., 11 100983, https://doi.org/10.1016/j.atech.2025.100983.

Nu-Lim Kim
Undergraduate in the School of Electronic Engineering, Hankyong National University. His research interests are Computer Vision, Robotics, and Reinforcement Learning.
E-mail: ssasink@hknu.ac.kr

Jungin Kim
Research Professor in AI Research, Hankyong National University. His research interests are Computer Vision, Robotics, and Vision Language Models.
E-mail: jik@hknu.ac.kr