Based on Iron Scraps Truck Load Image A Study on the Load Weight Prediction Model

Yeong-Won Park†, Dae-Ho Kim, Tae-Gyu Kim, Eun-Ki Hong
Department of Industrial Engineering, Konkuk University

철스크랩 트럭 적재 이미지 기반 적재 중량 예측 모델 연구

박영원†, 김대호, 김태규, 홍은기
건국대학교 산업공학과

Accurate estimation of scrap metal truck load weight is essential for efficient logistics and pricing in the recycling industry. However, current methods rely on physical scales, causing delays and resource burdens. This study proposes a deep learning-based image classification model that predicts the weight category (light, medium, heavy) from top-view truck images. Using the CoaT-Lite Medium architecture, which combines convolution and attention mechanisms, the model achieved 64.38% accuracy. Grad-CAM was used to visualize key visual features influencing predictions. The results demonstrate the feasibility of non-contact, image-based weight estimation in industrial settings.

Keywords : Scrap Metal, Image Classification, Weight Estimation, CoaT-Lite Medium

1. 서 론

철스크랩은 제철산업을 비롯한 다양한 산업 분야에서 핵심적인 재활용 자원으로 활용되며, 그 물류 관리 효율성은 기업의 원가 경쟁력 및 자원 순환 시스템에 중요한 영향을 미친다. 특히, 철스크랩을 운반하는 트럭의 적재 중량은 거래 단가 및 운송 효율성과 직결되므로, 이를 정확히 파악하는 것은 산업 현장에서 매우 중요한 과제이다. 그러나 현재 다수의 현장에서는 적재 중량 측정을 위해 별도의 지게차 무게 측정기나 차량 저울에 의존하고 있으며, 이로 인해 시간 지연, 인력 소요, 장비 설치의 어려움 등 다양한 문제가 발생하고 있다.

이러한 상황에서, 적재된 철스크랩 트럭의 이미지만으로 중량을 예측할 수 있다면 물류 자동화 수준을 크게 향상시킬 수 있다. 본 연구는 이를 가능케 하기 위해, 트럭의 적재 이미지 데이터를 활용하여 철스크랩 중량을 예측하는 딥러닝 기반의 분류 모델을 개발하고자 한다. 특히, 이미지 기반 추정이라는 비접촉식 방식은 기존의 물리적 계측 방식보다 빠르고 유연하며, 추가적인 장비 설치 없이도 적용이 가능하다는 점에서 산업적 실효성이 높다.

최근 몇 년 간 인공지능(AI), 특히 컴퓨터 비전 기반 기술은 이미지로부터 정량적인 정보를 추출하고 예측하는 데 있어 괄목할 만한 성과를 보이고

있다. 특히, Convolutional Neural Network(CNN) 기반의 모델은 영상 내 객체의 형태, 크기, 밀집도 등을 분석하여 구조화된 정보를 얻는 데 유용하며, 이 기술을 물류 산업에 접목시킨 연구도 꾸준히 확대되고 있다. 하지만 철스크랩처럼 비정형적이고 복잡한 형태를 가진 소재에 대해 중량을 예측하는 연구는 아직 초기 단계에 머물러 있으며, 데이터 수집 및 처리, 모델 개선 등의 과제가 남아 있다.

본 연구는 이러한 문제의식에서 출발하여, 실제 산업 현장에서 수집된 트럭 적재 이미지를 기반으로 철스크랩 중량을 세 단계(저중량/ 중중량/ 고중량)로 구간 분류하는 모델을 설계하고 그 가능성을 검토하는 것을 목적으로 한다. 향후 본 연구가 성공적으로 수행된다면, 철스크랩뿐 아니라 다양한 산업 소재에 대한 이미지 기반 계량 예측 기술로 확장될 수 있을 것이다.

2. 연구 방법론

2.1 기존 연구

기존의 트럭 적재 중량 측정은 로드셀이나 압력 센서와 같은 센서 기반의 계측 방식을 중심으로 이루어져 왔다. 이러한 방식은 높은 정확도를 제공한다는 장점이 있으나, 장비 설치와 유지에 상당한 비용이 소요되며, 물류 환경이나 산업 현장의 제약에 따라 설치가 어려운 경우도 많다. 이에 따라, 비접촉식 계측 방식의 대안으로 이미지 기반 예측 기법이 점차 주목받고 있다.

최근에는 딥러닝 기술의 발전과 함께 이미지 데이터를 활용하여 물체의 형태, 면적, 색상, 밀도와 같은 시각적 정보를 학습하고, 이를 바탕으로 수치 예측이나 분류 작업을 수행하는 연구가 활발히 진행되고 있다. 예를 들어, Chen 등(2022)은 단일 카메라 영상을 활용하여 건설 폐기물이 적재된 트럭의 부피를 예측하는 모델을 제안하였으며, 실제 산업 현장에 적용 가능성을 확인하였다[1]. Sun 등(2021)은 광산 트럭에 실린 자재의 적재량을 예측하기 위해 VGG16 기반의 딥러닝 모델과 회귀 기법을 결합한 접근을 제시하였고, 실제 환경에서 안정적인 예측 성능을 검증하였다[2].

Fig. 1. Representative example of a metal-scrap truck image (for illustration purposes)

Fig. 1처럼 금속 스크랩이나 폐기물과 같은 대상에 대해서도 이미지 기반 예측 기법을 적용한 연구들이 일부 존재한다. Zang 등(2024)은 금속 스크랩 컨테이너의 이미지를 분석하여 적재 수준을 단계적으로 분류하는 모델을 개발하였으며, 다양한 환경에서도 높은 분류 성능을 보인 바 있다[3]. 또한, dos Santos 등(2024)은 딥러닝과 인공신경망을 결합한 모델을 통해 금속 스크랩의 분류 및 중량 예측을 동시에 수행할 수 있는 방법을 제시하였다[4].

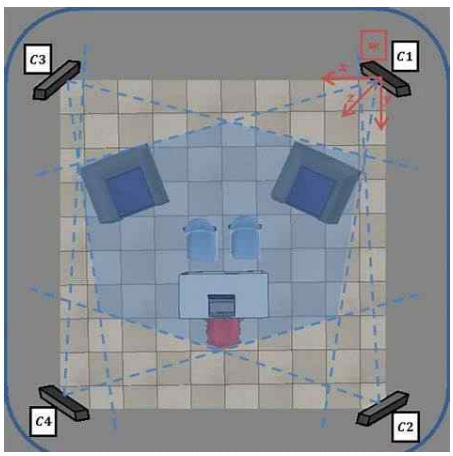

Fig. 2 .Diagram of a multi-view RGB D camera setup and 3D reconstruction process[5] (for illustration purposes)

한편, 트럭 단위의 적재량을 예측하기 위해 RGB 또는 깊이 기반 영상 데이터를 활용한 연구들도 이루어졌다. Kwak 등(2022)은 RGB-D 카메라(Fig. 2 참조)를 이용해 트럭의 적재 부피를 3차원으로 재구성하고 이를 기반으로 예측하는 모델을 제안하였다[6].

이처럼 기존 연구들은 정형화된 형태의 적재물이

나 제어된 환경에서 수집된 이미지를 대상으로 수행된 경우가 대부분이었다. 조명, 배경, 촬영 각도가 일정하게 유지되는 조건 하에서 모델을 훈련시키고 적용한 사례가 많으며, 산업 현장에서 흔히 접할 수 있는 비정형적이고 혼재된 형태의 적재물을 대상으로 한 연구는 상대적으로 부족한 실정이다. 따라서 본 연구는 이러한 현실적인 한계를 고려하여, 보다 복잡하고 불규칙한 철스크랩 적재 이미지를 기반으로 실용적인 예측 기법을 탐구하고자 한다.

2.2 이론적 배경

2.2.1 컴퓨터 비전 기술

컴퓨터 비전(Computer Vision)은 사람의 시각적 인식을 컴퓨터가 모방하도록 설계된 기술 분야로, 디지털 이미지나 동영상을 처리하여 그 안의 의미 있는 정보를 인식하고 해석하는 데 목적이 있다. 이 기술은 산업 자동화, 자율주행, 의료 진단, 스마트 팩토리 등 다양한 분야에 활용되고 있으며, 특히 딥러닝 기술의 발전과 함께 비약적인 성장을 이룩하였다.

전통적인 컴퓨터 비전 기술은 엣지 검출(edge detection), 색상 분할(color segmentation), 템플릿 매칭(template matching) 등과 같이 정해진 알고리즘에 기반한 방식이었으나, 최근에는 Convolutional Neural Network(CNN)과 같은 딥러닝 기반의 접근 방식이 주류가 되었다. CNN은 이미지의 국소적인 특징을 필터링하고 추상화하여 계층적으로 표현할 수 있어, 사람보다 뛰어난 인식 정확도를 달성할 수 있다. 이러한 모델은 이미지 분류, 객체 탐지(Object Detection), 세분화(Segmentation) 등 다양한 태스크에 활용될 수 있다.

산업 현장에서의 컴퓨터 비전은 품질 검사, 부품 인식, 이상 감지 등 자동화된 판단을 가능케 하며, 물류 분야에서는 물체의 위치 추적, 적재 상태 인식, 공간 최적화 등의 역할을 수행한다. 본 연구에서도 트럭 이미지 내 철스크랩의 형태적 특성과 적재 밀도를 분석하여, 이를 정량적 중량으로 환산하기 위한 기반 기술로 컴퓨터 비전을 활용하고자 한다.

2.2.2 이미지 기반 분류

이미지 기반 분류는 디지털 이미지를 입력으로 받아, 해당 이미지가 속하는 클래스(범주)를 예측하는 대표적인 지도학습 문제다. 일반적으로 입력 이미지는 합성곱 신경망(CNN) 또는 Transformer 기반의 딥러닝 모델을 통해 특징 벡터로 변환되며, 최종적으로는 미리 정의된 레이블 중 하나로 분류된다. 이 방식은 고양이/개 이미지 분류, 질병 진단, 품질 불량 탐지, 교통 표지 인식 등 다양한 시각 인식 문제에 널리 활용되고 있으며, 비교적 적은 양의 데이터로도 높은 정확도를 낼 수 있는 것이 장점이다.

최근에는 이미지 기반 분류가 단순 시각적 구분을 넘어, 재고 상태 분류, 산업 현장 내 적재 이상 탐지, 생산 공정 중 분류 자동화 등 정량적 또는 구조적 판단이 요구되는 실용 분야로도 확장되고 있다. 이러한 맥락에서, 본 연구는 철스크랩이 적재된 트럭의 상단 이미지를 활용하여, 적재된 철스크랩의 무게를 저중량 / 중중량 / 고중량의 세 구간으로 분류하는 문제를 정의하였다.

이는 연속적인 중량값을 예측하는 회귀 문제보다, 상대적으로 분류에 강한 딥러닝 모델의 구조적 특성을 활용함으로써 모델의 수렴 가능성을 높이고, 실용적인 분류 정확도를 확보할 수 있는 방향이다. 또한 데이터 수가 제한적이고, 각 이미지 내의 조명, 각도, 적재 방식이 일정하지 않은 현실적 조건에서도, 분류 문제는 비교적 견고한 성능을 확보할 수 있는 장점이 있다. 따라서 이미지 기반 분류는 본 연구의 조건과 목적에 부합하는 적절한 문제 정의 방식이라 할 수 있다.

2.2.3 CoaT-Lite Medium

본 연구에서는 트럭 적재 이미지로부터 철스크랩의 중량 구간을 예측하기 위한 분류 모델로 CoaT-Lite Medium(Convolutional attention-based Transformer) 구조를 적용하였다. CoaT는 Convolution과 Transformer의 장점을 결합한 하이브리드 비전 모델로, 국소 특징 추출에 강한 CNN 구조와 전역적인 의존성 파악에 강한 Transformer 구조를 통합하여 다양한 시각 인식 과제에서 우수한 성능을 보여준다.

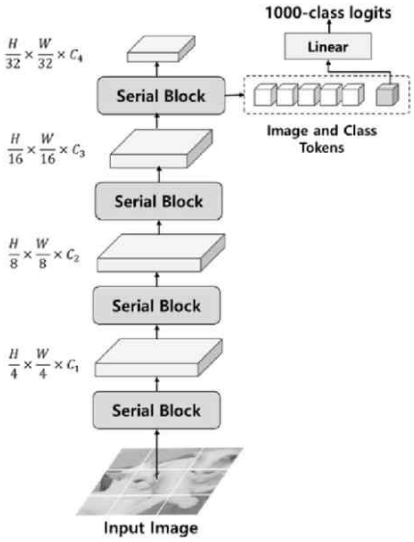

Fig. 3. CoaT-Lite Medium Architecture Overview[6]

특히, CoaT-Lite Medium의 전체 구조는 Fig. 3에 제시된 바와 같이 계층적으로 구성된 serial block과 attention 메커니즘을 통해 이미지로부터 다층적인 특징을 추출하도록 설계되어 있으며, 비교적 경량화된 구조임에도 이미지의 정형성과 배경 다양성이 큰 환경에서도 높은 정확도를 유지할 수 있다는 점에서 본 연구에 적합한 모델로 판단되었다.

철스크랩 이미지의 경우, 동일한 중량이라도 시각적으로는 형태가 다르거나 조명과 각도의 차이가 크기 때문에, 단순 CNN이나 Transformer 단독 구조보다 더 정교한 특징 추출 능력이 요구된다. CoaT는 이를 해결하기 위해 convolution block으로 로컬 정보를 추출한 뒤, attention 계층을 통해 전역적인 구조 정보를 보완적으로 처리한다.

또한 CoaT-Lite는 입력 이미지 해상도에 대한 유연성을 가지며, 연산 효율성과 정확도 사이의 균형을 잘 갖춘 구조이기 때문에, 학습 데이터가 제한된 본 연구의 조건에서도 수렴성과 일반화 성능 모두를 기대할 수 있다. 실제 실험 과정에서도 다른 경량 모델들보다 우수한 분류 정확도를 보여, 이후 모델 개선 및 현장 적용 가능성 측면에서도 긍정적인 결과를 제공하였다.

2.2.4 Grad-CAM

Grad-CAM은 CNN 기반 모델의 결정 근거를 시각적으로 해석할 수 있도록 도와주는 시각화 기법으로, Selvaraju et al.(2017)에 의해 제안되었다. 이 방법은 특정 클래스에 대한 예측 점수의 gradient를 사용하여, 마지막 convolution layer의 feature map에서 중요도가 높은 영역을 강조한다. 이를 통해 모델이 어떤 부분을 중점적으로 바라보고 해당 클래스를 예측했는지를 직관적으로 파악할 수 있다.

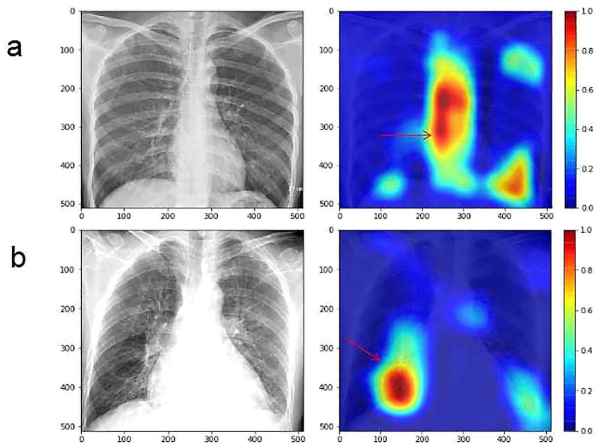

Fig. 4. Example of Grad-CAM visualization applied to CNN-based classification[7]

Grad-CAM은 특히 이미지 분류 문제에서 모델의 판단 과정을 해석하거나 오류 사례를 분석하는 데 효과적이다. Fig. 4는 Gao(2020)의 연구에서 실제 의료 영상을 대상으로 Grad-CAM을 적용한 예시로, CNN 모델이 질병을 예측할 때 주목하는 시각적 영역을 효과적으로 보여준다. 본 연구에서도 이와 유사한 방식으로, 학습된 분류 모델이 철스크랩 적재 이미지를 인식하고 분류하는 과정에서 어떤 영역에 주목하는지를 시각적으로 분석하고자 하였다. 이를 통해 모델 성능이 높은 이미지와 낮은 이미지 간의 주목 패턴 차이를 비교하고, 분류 오류의 원인을 탐색하는 데 활용하였다.

3. 연구 절차

Fig. 5는 데이터 수집 및 전처리, 모델 선택 및 학습(교차검증 및 하이퍼파라미터 튜닝 포함), 성능 평가 및 Grad-CAM 기반 해석으로 구성된 전체 연구 절차를 나타낸다.

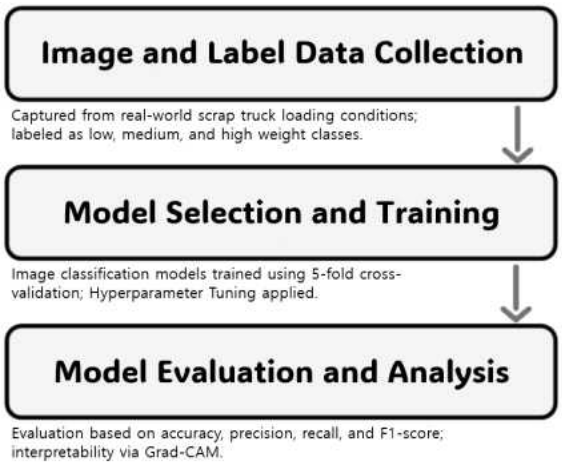

Fig. 5. Overall Research Process

3.1 데이터 정제

본 연구의 첫 번째 단계는 모델 학습 및 평가를 위한 기초 데이터를 확보하고 정제하는 과정과, 유사 연구 및 기술 사례를 분석하여 연구의 기술적 타당성과 적용 가능성을 확인하는 과정으로 구성된다.

먼저, 데이터 정제 측면에서는 별도의 데이터 수집 없이, 연구 협력처로부터 제공받은 철스크랩 트럭 적재 이미지와 해당 트럭의 중량 측정 이력 데이터를 전수 확보하여 활용하였다. 해당 데이터는 트럭 적재 상태가 다양한 시점에서 촬영된 이미지로 구성되어 있으며, 각각의 이미지에는 실측 중량 값이 대응되어 있다. 이 데이터를 기반으로 학습용 입력-정답 쌍(input-label pair)을 구성하였고, 이미지의 형식 일관성 확보 및 전처리 과정을 통해 모델 학습에 적합한 형태로 가공하였다.

Fig. 6. Example of original truck load image

전처리 작업은 다음과 같은 단계를 포함하였다. 첫째, 원본 이미지 파일(Fig. 6 참조) 내 불필요한 여백, 배경, 번호판 등 예측과 무관한 요소를 제거하거나 크롭(Crop)하여 트럭 적재함이 중심이 되도록 정렬하였다.(Fig. 7. 참조)

Fig. 7. Example of image after cropping

둘째, 중복 이미지, 각도 이탈, 정보 부족(트럭 식별 불가) 등의 문제를 가진 데이터를 식별하여 제외하거나 보완하였다.

셋째, 모델 입력에 적합하도록 모든 이미지를 224×224 해상도로 통일하여 크기 조정하였다.(Fig. 8 참조)

Fig. 8. Example of image after resizing(224x224)

3.2 모델 설계

본 연구에서는 다양한 이미지 분류용 딥러닝 모델을 실험적으로 비교하여 최적의 모델 구조를 도출하였다. 그 결과, CoaT-Lite Medium 모델이 가장 우수한 성능을 보였다. 이 모델은 제한된 데이터 환경에서도 다른 비교 모델들에 비해 안정적인 학습 성능을 나타냈으며, 특히 다양한 배경, 조명, 촬영 거리 등으로 인해 단일 특징만으로 분류하기 어려운 본 연구의 데이터 환경에서 뛰어난 일반화 성능을 기록하였다. 실제 실험 과정에서도 hierarchical attention을 활용하여 이미지 내 여러 위치 정보를 효과적으로 강조함으로써, 다른 모델들이 공통적으로 오답을 기록한 이미지 일부를 정확하게 예측하는 차별적 성능이 확인되었다. 따라서 본 연구에서는 CoaT-Lite Medium을 최종 분류 모델로 채택하였다.

3.3 모델 평가

본 실험은 Google Colab 환경에서 수행되었으며, PyTorch 2.0 기반의 학습 코드를 사용하였다. 주요 실험은 NVIDIA T4 GPU가 탑재된 환경에서 실행되었으며, Colab에서 제공하는 CUDA 11.8 및 Python 3.10 기반 환경을 바탕으로 하였다.

모델 학습 과정에서는 K-Fold Cross Validation (k=5)을 적용하여 학습 데이터의 일반화 성능을 검증하고, 데이터 수의 제한에도 불구하고 최대한 안정적인 평가가 가능하도록 구성하였다. 학습에는 CrossEntropyLoss를 손실 함수로 사용하고, 각 fold에 대해 10 epoch 동안 학습을 반복하였다. 실험 환경은 CUDA 지원 GPU 기반 환경에서 진행되었다.

데이터의 질적 분포 분석 결과, 일부 이미지들은 정답률이 0%로 모든 모델에서 오답으로 분류되는 것으로 나타났다. 이는 해당 이미지들이 본질적으로 모델이 학습하기 어려운 특성을 가지고 있음을 시사하며, 학습 성능 향상을 위해서는 향후 데이터 필터링 또는 난이도 기반 학습(curriculum learning)의 도입이 필요하다는 점을 확인할 수 있었다.

하이퍼파라미터 튜닝은 Optuna 프레임워크를 활용하여 자동화된 방식으로 진행하였으며, 주요 탐색 대상은 learning rate, dropout, weight decay 세 가지로 설정하였다. 옵티마이저는 AdamW로 고정하였다. 해당 튜닝은 각 fold의 검증 정확도 평균을 기준으로 최적 조합을 도출하는 방식으로 수행되었으며, 각 파라미터에 대한 최적 조합은 다음과 같다. 학습률은 7.58e-05로 설정되었으며, 드롭아웃 비율은 0.3374, 가중치 감쇠는 1.29e-04로 결정되었다. 해당 최적값을 바탕으로 학습된 모델은 평균 분류 정확도 64.38%를 기록하였다. 보다 구체적인 실험 결과는 4장에서 기술한다.

3.4 추가 연구

본 연구에서는 철스크랩이 적재된 트럭의 단일 이미지를 입력으로 하여, 해당 트럭의 중량을 저중량, 중중량, 고중량의 세 단계로 분류하는 딥러닝 기반 이미지 분류 모델을 개발하였다. 다만, 중량 분류의 정밀도 향상과 물리적 근거 보강을 위한 시도로서, 본 연구 초기에는 3D 복원 기반의 부피 추정과 밀도 기반 질량 계산 접근을 병행하여 실험적으로 시도하였다.

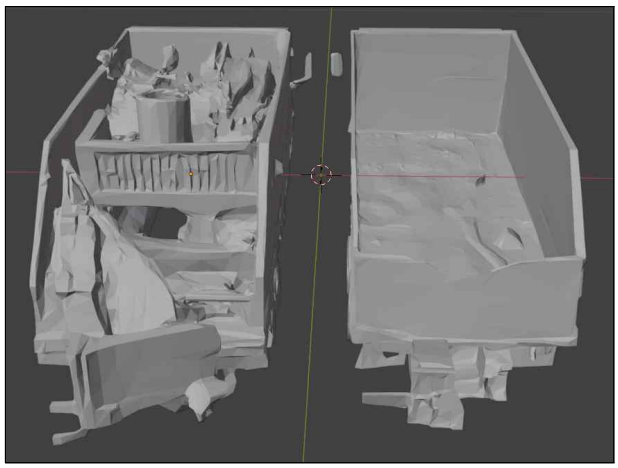

Fig. 9. Truck 3D mesh with open geometry causing inaccurate volume estimation

이 보조 실험에서는 철스크랩 적재 전과 적재 후 트럭의 단일 이미지를 입력으로 사용하여, 각각의 3D 메쉬 모델(Fig. 9 참조)을 생성하고 두 상태의 부피 차이를 계산함으로써 철스크랩의 부피를 추정하는 것을 목표로 하였다. 이후, 철스크랩의 종류에 따른 평균 밀도를 활용하거나 이미지 기반 클러스터링을 통해 밀도를 추정한 후, 물리식에 기

반한 중량 예측을 수행하고자 하였다.

그러나 실험 과정에서 다음과 같은 세 가지 기술적 한계가 발견되었으며, 이로 인해 해당 방식은 본 연구의 최종 중량 분류 모델에는 포함되지 않았다. 각 한계점은 아래와 같다.

3.4.1. 3D 메쉬 구조의 밀폐 불완전성

단일 이미지로부터 외부 서비스의 Image-to-3D 기능을 활용해 생성한 .obj 메쉬는 외형적으로는 트럭의 기본 형태를 재현하였으나, 대부분의 경우 메쉬 구조가 완전히 밀폐되어 있지 않았다. 특히 트럭 내부가 비어 있고, 철스크랩의 불규칙한 형상 또한 완전한 표면을 구성하지 못한 채 복원되었다. 이로 인해 부피를 계산할 경우, 내부 공간이 인식되지 않아 비현실적으로 낮은 부피값이 산출되었다. 실제로 철스크랩이 가득 적재된 모델의 경우에도 0.025m^3와 같은 수치가 출력되는 등, 부피 계산의 신뢰도가 현저히 낮았다.

3.4.2. 변환 과정에서의 구조 손실과 표현 제한

Pix2Vox 모델과 같은 3D 딥러닝 네트워크는 학습 입력으로 voxel 형태의 3D 데이터를 사용하므로, .obj 파일을 .binvox 포맷으로 변환하는 과정이 필요하다. 그러나 다음과 같은 문제점이 확인되었다.

우선, 복잡한 구조이거나 밀폐되지 않은 메쉬를 변환할 경우 변환 오류가 발생하거나, 변환 결과로 생성된 파일이 완전히 비어 있는 경우가 다수 있었다. 둘째, voxel 해상도가 제한적이다 보니, 철스크랩의 복잡하고 불규칙한 외형이 변환 과정에서 과도하게 단순화되어 있다. 이로 인해 실제 형상과는 다른 저해상도 구조로 바뀌면서 학습 효과를 저해하는 요인이 되었다.

이러한 구조 손실과 표현 제약은 voxel 기반 학습의 품질을 떨어뜨리는 원인이 되었다. 이와 같이, 본 연구에서 시도한 3D 복원 기반 중량 추정 방식은 부피 기반 질량 계산이라는 물리적 해석 가능성과 정량성 확보라는 측면에서 의미 있는 시도였으나, 현 시점에서는 데이터 품질, 구조 신뢰성, 학습 안정성 측면에서 실용화에는 미치지 못하였다. 그럼에도 본 실험은 향후 트럭 중량 예측 연구의 정밀화 및 실제 물류 시스템 내 응용 가능성을 확장할 수 있는 기반으로 작용할 수 있을 것이며, 관련 기술의 발전에 따라 본 연구 또한 다양한 방향으로 확장될 수 있을 것으로 판단된다.

4. 결과 및 분석

Table 1. Performance Comparison of Models

Model	Accuracy	F1-Score
CoaT-Lite Medium	**63.33%**	**0.6235**
RegNetY-002	57.52%	0.5657
Swin-Tiny	57.43%	0.5530
Dropout Rate	52.52%	0.4854
EfficientNetV2	50.48%	0.4544

CoaT-Lite Medium, RegNetY-002, Swin-Tiny, ConvNeXt-Tiny, EfficientNetV2 등 5개의 이미지 분류 모델에 대해 동일한 실험 조건(5-Fold 교차검증, Epoch=10) 하에서의 성능을 비교하였다. Table 1에 제시된 바와 같이, CoaT-Lite Medium 모델이 Accuracy 63.33%, F1-Score 0.6235로 가장 우수한 성능을 보였으며, 이는 다른 모델에 비해 철스크랩 적재 이미지의 중량 분류 문제에 보다 효과적으로 대응함을 시사한다. F1-Score는 클래스 간 불균형을 고려한 정밀도와 재현율의 조화평균으로, 모델의 전반적인 균형 잡힌 성능을 나타내는 지표로 활용되었다.

Table 2. Hyperparameter Tuning Results

Parameters	Value
Accuracy	**64.38%**
Learning Rate	7.58e-05
Weight Decay	1.29e-04
Dropout Rate	0.3374

본 연구에서는 CoaT-Lite Medium 모델의 성능을 최적화하기 위해 Optuna를 활용한 하이퍼파라미터 튜닝을 수행하였다. 학습률(Learning Rate), 가중치 감쇠 계수(Weight Decay), 드롭아웃 비율(Dropout Rate)을 대상으로 10개의 Trial을 실험한 결과, Trial 6에서 정확도(Accuracy) 64.38%로 가장 우수한 성능을 기록하였다. Table 2에 제시된 바와 같이, 해당 Trial에서는 비교적 작은 학습률

(7.58e-05)과 적절한 드롭아웃 비율(0.3374), 안정적인 가중치 감쇠 계수(1.29e-04)를 사용하였으며, 이는 모델의 일반화 성능 향상에 약간의 긍정적인 영향을 미친 것으로 해석된다.

Grad-CAM 시각화 결과는 CoaT-Lite Medium 모델이 철스크랩 적재 이미지의 중량을 판단함에 있어 어떤 시각적 정보를 주요하게 활용하는지를 잘 보여준다.

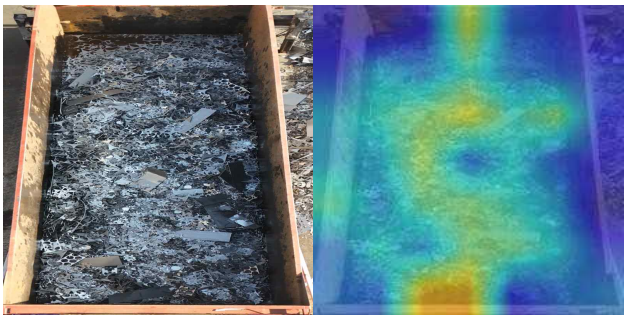

Fig. 10. Grad-CAM Visualizations for Correct Predictions (weight : High)
[left: original image / right: Grad-CAM]

Fig. 10의 고중량 사례에서는 적재함 전체에 걸쳐 넓고 균일하게 분포한 적재물 중에서도 중심부와 하단부에 대한 주목도가 특히 높게 나타났다. 이는 실제 중량이 클수록 적재물의 밀도와 부피가 하중 중심에 모이기 쉽다는 특성과 일치하며, 모델이 중량 판단에 있어 합리적인 시각 정보를 근거로 활용했음을 보여준다.

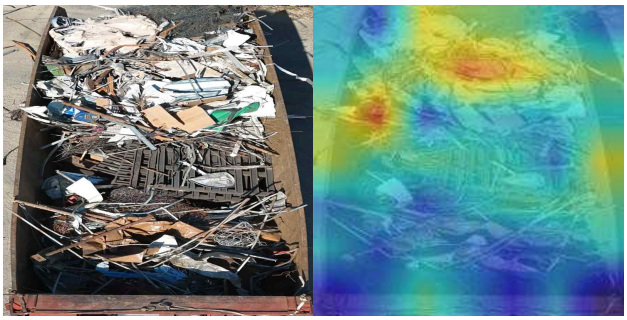

Fig. 11. Grad-CAM Visualizations for Correct Predictions (weight : Mid)
[left: original image / right: Grad-CAM]

Fig. 11의 중중량 사례에서는 철근 구조처럼 복잡하게 얽힌 선형 물체들이 분포한 적재함 중앙에 집중된 활성화가 관찰되었다. 이러한 구조물은 중량 측면에서 불균형한 질량 분포를 형성할 수 있는데, 모델은 해당 특징을 효과적으로 포착하고 중량 등급을 적절히 분류한 것으로 해석된다.

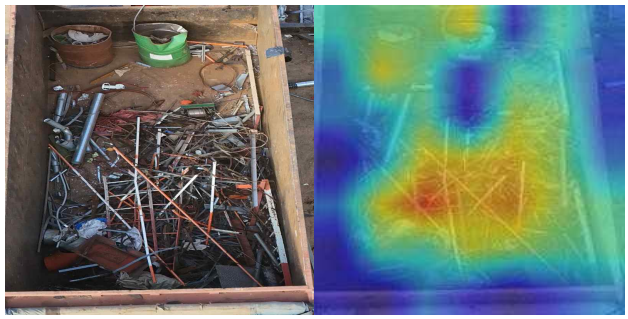

Fig. 12. Grad-CAM Visualizations for Correct Predictions (weight : Low)
[left: original image / right: Grad-CAM]

마지막으로 Fig. 12의 저중량 사례에서는 소규모 적재물 부분에 대해서 제한적으로 주목하는 한편, 그 외 대부분의 영역은 비활성화되어 있었다. 이는 모델이 가벼운 무게의 경우 적재물의 전체적인 부피나 존재 유무를 중점적으로 고려하여 분류를 수행했음을 보여준다.
이와 같은 Grad-CAM 결과는 모델이 시각적으로 일관성 있는 기준에 따라 적재물의 공간적 밀도와 구조를 고려하여 중량을 추론하고 있음을 의미하며, 정량 지표뿐만 아니라 정성적 해석 관점에서도 모델의 예측 근거에 대한 신뢰성을 높여준다.

5. 결 론

본 연구는 철스크랩이 적재된 트럭의 상단 이미지를 입력으로 활용하여, 비접촉 방식으로 중량을 예측할 수 있는 이미지 기반 분류 모델의 가능성을 검토하고자 하였다. 이는 기존의 센서 기반 계측 방식이 갖는 장비 설치의 어려움, 비용 부담, 자동화 제약 등의 문제를 보완하고, 산업 현장 내 물류 자동화 수준을 제고할 수 있는 대안으로서의 실용성을 확인하고자 한 시도였다.

이를 위해 CoaT-Lite Medium 구조를 포함한 여러 딥러닝 분류 모델을 실험적으로 비교하였으며, 모델의 일반화 성능을 확보하기 위해 교차검증과 하이퍼파라미터 최적화를 병행하였다. 실험 결과

CoaT-Lite Medium은 제한된 학습 데이터와 다양한 이미지 변형 환경에서도 상대적으로 높은 예측 성능을 기록하였으며, Grad-CAM을 통한 시각적 분석을 통해 모델이 철스크랩 적재물의 구조적 특성을 효과적으로 학습하고 있음을 확인할 수 있었다.

한편, 본 연구 진행 중에, 3D 복원 기반 부피 추정과 물리적 밀도 추정을 병행하는 방식도 시도하였으나, 단일 이미지 기반의 3D 재구성이 갖는 기술적 한계—예: 메쉬 밀폐 불완전성, 구조 단순화, 변환 오류 등—로 인해 본 모델에는 포함되지 않았다. 이러한 시도는 향후 중량 예측 모델의 정밀도 향상과 해석 가능성을 확장하기 위한 기반 실험으로서의 의미를 가진다.

본 연구의 주요 의의는, 철스크랩처럼 비정형적이고 시각적 변동성이 큰 적재물에 대해 정량적 분류가 가능함을 보였다는 점에 있다. 이미지 기반 예측이라는 접근은 물류 시스템의 자동화 가능성을 확대하고, 현장 작업자의 판단 의존도를 줄이는 데 기여할 수 있다. 다만, 데이터의 양적 한계와 특정 이미지에서의 분류 실패 사례는 여전히 존재하였으며, 이는 향후 정제된 데이터 수집과 더불어, 객체 검출 기반의 분할(pre-segmentation), 다중 시점 이미지 활용 등의 기법을 통해 보완될 필요가 있다.

향후 연구에서는 다양한 산업 소재를 대상으로 본 연구에서 제안한 모델을 확장 적용하고, 실제 물류 시스템 내에서의 활용 가능성을 평가함으로써, 산업 전반의 스마트 계측 기술 발전에 기여할 수 있을 것으로 기대된다.

감사의 글

본 연구에 데이터를 제공해주신 (주)심팩에 감사드립니다.

References

[1] Chen, H., et al. (2022). Construction waste truck volume estimation using monocular images. Automation in Construction, 138, 104252.

[2] Sun, Z., et al. (2021). Truckload estimation in open-pit mines using VGG16 and regression. IEEE Access, 9, 113782-113791.

[3] Zang, Y., et al. (2024). Automatic fill-level estimation of scrap metal containers using CNNs. Resources, Conservation and Recycling, 200, 106805.

[4] dos Santos, R., et al. (2024). Multimodal weight estimation and classification of metal scrap using DenseNet and BPNN. Journal of Cleaner Production, 435, 140291.

[5] RGB-D Multi-View System with Full Scene 3D Reconstruction. Bing Images, https://tse4.mm.bing.net/th?id=OIP.OCFlI4CedYQ439wp6eAH2QHaHk&pid=Api (Accessed: June 20, 2025).

[6] Kwak, M., et al. (2022). Excavator truck volume estimation using RGB-D cameras and 3D reconstruction. Journal of Field Robotics, 39(8), 1201-1215.

[7] Gao, Y. (2020). Visualization of CNN models using Grad-CAM to classify PH from normal based on radiograph images [Figure]. Figshare. https://doi.org/10.6084/m9.figshare.12710725.v1

[8] Xu, W., Zhang, Z., Zhu, H., Zhang, C., & Li, Y. (2021). CoaT: Co-Scale Conv-Attentional Image Transformers. arXiv preprint arXiv:2104.06399.