

# 자동차 대시보드 분류를 위한 메타모픽 테스트 기반의 CNN 알고리즘 비교 분석

## A Comparative Analysis Based on Metamorphic Testing of CNN Algorithms for Classification of Vehicle Dashboards

정종민<sup>1</sup>, 김동헌<sup>2\*</sup>(Jong Min Jeong<sup>1</sup> and Dong Hun Kim<sup>2,\*</sup>)<sup>1</sup>Department of Mechatronics Engineering, Kyungnam University<sup>2</sup>Department of Electrical Engineering, Kyungnam University

**Abstract:** This paper proposes an effective classification system for vehicle dashboards using various pretrained convolutional neural network (CNN) models and presents a comparative analysis of their performance. Focusing on five types of dashboards that are visually indistinguishable, the CNNs are used to extract detailed features and develop an automated classification system. Nine pretrained CNN models are evaluated based on accuracy, loss rate, and F1-score, with training enhanced by emphasizing regions of interest. In particular, the study validates the robustness and consistency of the models through metamorphic testing, which demonstrates that the CNN model works reliably on deformed data in a real environment and can maintain high classification accuracy. The results highlight the significant improvements of the proposed system over visual inspection in reliability and efficiency. This research offers a pathway for automating material classification in manufacturing sites and illustrates the practical applicability of AI-based classification technologies.

**Keywords:** convolutional neural network (CNN), classification, dash boards, pre-trained CNN, metamorphic test

### I. 서론

최근 제조업은 기술의 발전과 함께 급격한 변화를 겪고 있으며, 보다 효율적이고 안정적인 시스템을 구축하기 위한 노력이 지속되고 있다. 4차 산업혁명의 도래와 함께 스마트 제조, 로봇 공학, 사물 인터넷(IoT), 빅데이터 등의 기술이 제조업에 접목되면서 생산성과 품질을 극대화하려는 시도가 이어지고 있다. 이러한 기술 혁신은 전통적인 제조업의 경계를 허물고 있으며, 다양한 산업 분야에서 새로운 비즈니스 모델과 생산 방식을 창출하고 있다.

스마트 제조는 디지털 기술을 활용하여 제조 과정의 모든 단계를 자동화하고 최적화하는 것을 목표로 한다. 이를 통해 생산 과정의 효율성을 높이고, 제품의 품질을 향상시키며, 생산 비용을 절감할 수 있다. 예를 들어, IoT 기술을 통해 제조 설비의 상태를 실시간으로 모니터링하고, 빅데이터 분석을 통해 최적의 생산 조건을 도출할 수 있다. 또한, 로봇 공학을 이용한 자동화 생산 라인은 인간의 개입을 최소화하고, 작업의 일관성을 유지할 수 있게 한다.

특히 자동차 산업은 이러한 변화의 중심에 서 있다. 자동차는 단순한 이동 수단을 넘어 스마트 디바이스로 진화하고 있으며, 자율주행차와 전기차의 등장은 이러한 변화를 더욱 가속화시키고

있다. 자율주행차는 고도의 센서 기술과 인공지능을 결합하여 운전자의 개입 없이도 안전하게 주행할 수 있는 차량을 의미하며, 전기차는 친환경적이고 지속 가능한 에너지를 사용하는 차량으로서 주목받고 있다. 이처럼 자동차의 발전은 내부 구조와 구성 요소에도 큰 변화를 요구하고 있다.

자동차의 내장재는 운전자의 안전과 편안함을 보장하는 중요한 요소이다. 내장재의 품질은 차량의 전체적인 완성도와 직결되며, 이를 적절히 관리하고 분류하는 것은 매우 중요한 일이다. 내장재는 대시보드, 시트, 도어 트림, 카펫 등 다양한 부품으로 구성되며, 각각의 부품은 고유의 특성과 기능을 가지고 있다. 예를 들어, 시트는 운전자의 편안함을 위하여 인체 공학적으로 설계되어야 하며, 대시보드는 안전성을 고려한 내구성과 충격 흡수 기능을 가져야 한다.

그러나 전통적인 내장재 분류 방식은 주로 수작업에 의존하고 있으며, 이는 시간과 비용 측면에서 비효율적일 뿐만 아니라, 인간의 주관적인 판단에 의존하기 때문에 일관성을 유지하기 어렵다는 문제가 있다. 수작업 분류 과정에서는 작업자의 경험과 능력에 따라 분류 결과가 달라질 수 있으며, 이는 제품의 일관성을 저해하는 요인이 된다. 또한, 수작업 분류는 작업자의 피로도와 집중력 저하로 인해 오류 발생 가능성이 높아진다.

\*Corresponding Author

Manuscript received November 21, 2024; revised January 18, 2025; accepted February 15, 2025

정종민: 경남대학교 메카트로닉스공학과 석사과정(sd610713@naver.com, ORCID<sup>®</sup> 0009-0008-7524-1630)

김동헌: 경남대학교 전기공학과 교수(dhkim@kyungnam.ac.kr, ORCID<sup>®</sup> 0000-0003-0615-9736)

※ 본 과제(결과물)는 교육부와 경상남도의 재원으로 지원을 받아 수행된 경상남도 지역혁신중심 대학지원체계(RISE)의 연구결과입니다.



(A) DB\_A. (b) DB\_B. (c) DB\_C. (d) DB\_D. (e) DB\_E.

그림 1. 대시보드.

Fig. 1. Dash board (DB).

하나의 자동차에서 모델마다 각기 다른 내장재를 사용하는 경우가 존재한다. 그러한 이유로 현재 자동차 내장재를 생산하고 있는 회사에서 제품 생산과 조립 과정 중에 같은 자동차에서 사용하는 다양한 내장재 모델이 혼재해 있다. 또한, 같은 생산 라인에서 여러 자동차 모델의 내장재가 제조되고 있고, 조립은 다른 공정 라인에서 하고 있어 조립 전 구별이 어려워 많은 문제가 발생한다.

그림 1은 자동차 대시보드로 5개의 사진이 육안으로는 모두 같아 보이지만 실제로는 재질과 모양이 모두 다르다.

이러한 문제를 해결하기 위하여 인공지능과 딥러닝 알고리즘을 이용한 내장재 분류 시스템이 제안되고 있다. 딥러닝 알고리즘은 대량의 데이터를 학습하여 높은 정확도의 분류 작업을 수행할 수 있는 장점이 있다. 이는 내장재의 종류를 자동으로 분류하는 데 매우 유용하게 활용될 수 있다. 딥러닝은 다양한 층을 가진 인공신경망을 통해 데이터를 학습하고, 이를 바탕으로 패턴을 인식하여 높은 정확도의 예측 및 분류를 가능하게 한다.

본 연구에서는 AI 딥러닝 알고리즘을 적용한 자동차 내장재 분류 시스템을 개발하고, 이를 통해 기존 분류 방식의 한계를 극복하고자 한다. 이를 위하여 다양한 내장재 이미지 데이터를 수집하고, 해당 데이터를 기반으로 딥러닝 모델을 학습시켜 정확한 분류가 가능하도록 하였다. 특히, CNN (Convolutional Neural Network)과 같은 딥러닝 모델을 활용하여 내장재의 세부적인 특징을 추출하고, 이를 통해 높은 정확도의 분류를 목표로 한다.

본 논문의 구성은 다음과 같다. 2장에서는 연구 배경과 관련 연구를 다루고, 3장에서는 실험 전 데이터 준비과정을 거친다. 4장에서는 수행된 실험 결과를 분석하고, 마지막으로 5장에서는 결론 및 향후 연구 방향에 대해 논의한다.

## II. 연구 배경과 관련 연구

### 1. 메타모픽 테스트의 필요성

AI 딥러닝 알고리즘을 이용한 분류 시스템의 경우, 테스트 과정에서 오라클 문제(test oracle problem)가 발생할 수 있다. 오라클 문제란 시스템 출력의 올바른 참조 값을 정의하거나 확보하기 어려운 상황을 의미하며, 이는 다양한 환경 조건에서 입력 데이터가 변형되거나, 대규모 데이터셋에 대해 참조 값을 수작업으로 검증하는 데 막대한 시간과 비용이 소요되는 경우에

발생한다. 실제 제조 환경에서 내장재 분류 시스템은 조명, 각도, 배경과 같은 외부 조건의 변화에 따라 입력 데이터가 달라질 수 있다. 이러한 상황은 기존 테스트 방식으로는 신뢰성과 일관성을 확보하기 어렵다. 메타모픽 테스트는 다양한 입력 데이터 변형에도 시스템이 일관된 성능을 유지할 수 있는지 검증함으로써, 실제 제조 환경에서 안정적인 AI 기반 분류 시스템을 구현하는 데 중요한 역할을 할 수 있다. 본 연구는 이러한 문제 해결을 위하여 메타모픽 테스트를 적용하여 기존 방식의 한계를 극복하고, 신뢰성과 효율성을 동시에 확보할 수 있는 접근법이 필요하다.

### 2. 메타모픽 테스트를 활용한 관련 연구 동향

CNN을 사용한 자동차의 대시보드 분류 연구는 자동차 생산자와 관련된 구체적인 문제기 때문에 다양하게 이루어지지 않고 있다. 제안하는 문제와 대상은 다르지만, CNN 기반의 분류와 메타모픽 테스트를 사용한 관련된 연구는 다음과 같다.

[1]에서는 메타모픽 테스트가 테스트 오라클 문제를 해결하기 위한 효과적인 방법으로 제안되었으며, 기존의 성공적인 테스트 케이스를 기반으로 새로운 테스트 케이스를 생성함으로써 기존 테스트에서 발견되지 않은 오류를 탐지할 수 있음을 보여주었다. 이 기법은 이진 탐색, 그래프 최단 경로 탐색, 가우스 소거법 등 다양한 알고리즘에 적용 가능성을 입증하며, 테스트 자동화와 신뢰성 향상에 기여하였다.

[2]는 CNN 기반 이미지 분류 모델의 신뢰성을 평가하기 위해 복잡 변형 규칙을 활용한 메타모픽 테스트 방법을 제안하였다. CIFAR-10 데이터셋과 ResNet-101 모델을 사용하여 노이즈와 확대 변형 등 단일 변형 규칙과 복합 규칙을 적용한 결과, 메타모픽 관계를 유지하며 모델의 일관성을 확인하였다. 이 연구는 기존 단순 변형 규칙보다 복잡하고 현실적인 테스트 환경을 제공함으로써 CNN 모델의 강건성과 신뢰성을 평가하는 효과적인 접근법을 제시하였다.

[3]에서는 딥러닝 프레임워크의 검증을 위해 메타모픽 테스트를 활용하여 AlexNet과 SVM을 비교하며, 시스템 수준, 데이터셋 수준, 데이터 항목 수준의 세 가지 변형 관계를 정의하고 이를 적용하였다. 실험 결과, 딥러닝 분류기의 초기 정확도가 데이터 품질을 개선하고 훈련 데이터셋을 확장한 후 대폭 향상되었다. 이는 메타모픽 테스트가 대규모 데이터와 복잡한 딥러닝 모델의 검증에 효과적임을 보여주며, 다양한 딥러닝 애플리케이션에 활용될 가능성을 제시하였다.

### 3. 사전학습(Pre-trained) CNN

Pre-trained CNN은 일반적으로 대규모 이미지 분류 문제를 해결하기 위하여 대량의 데이터셋에서 미리 훈련되어 저장된 네트워크를 의미한다. 원본 데이터셋이 충분히 크고 다양한 경우, 사전 훈련된 네트워크가 학습한 특성의 계층 구조는 실제 데이터에도 일반화되어 효율적인 모델로 사용될 수 있다. 본 연구에서는 VGG16, VGG19, Inception, Xception, ResNet152, ResNet50V2 DensNet121, DensNet201 MobileNetV2, 총 9개의 사전 학습된 CNN을 사용한다. 9개 모델 모두 ImageNet을 사용하여 사전학습되며 ImageNet에서 학습된 가중치는 일반적인 이미지 분류 작업에 대한 강력한 기초를 제공한다[4].

제안된 CNN 모델들은 다양한 실험과 연구를 통해 충분히 검증되어, 성능 비교 및 분석을 위한 기준으로 활용될 수 있다.

이러한 모델들은 새로운 모델의 성능이 기존 모델과 얼마나 차이가 나는지 검증하는 중요한 기준점 역할을 한다. 또한, 비교적 단순한 구조를 가지기 때문에 실험 환경의 재현성이 높아 실험 결과를 반복적으로 확인하고 검증하기에 효율적이다.

3.1 VGG

VGG 모델은 네트워크의 깊이와 성능 간 관계를 탐구한 CNN 구조로, 그림 2와 같이 일관되게 작은 3x3 커널을 사용하여 깊은 네트워크를 구성한다[5]. 작은 커널을 반복 사용함으로써 복잡한 패턴 학습과 연산 효율성을 높인다. 본 연구에 사용된 VGG 모델은 VGG16과 VGG19로, 각각 16층과 19층의 깊이를 가지며 이미지 분류 등 다양한 컴퓨터 비전 작업에서 널리 활용된다.

3.2 Inception

Inception 모델은 다양한 크기의 합성곱 필터를 병렬로 연결하여, 여러 스케일의 특징을 동시에 학습할 수 있는 능력을 갖추고 있다. 이러한 구조는 모델이 더욱 유연하게 특징을 추출할 수 있도록 하며, 이를 통해 성능을 크게 향상시킬 수 있다. 그림 3을 보면 입력 특징 맵에 대해 1x1, 3x3, 5x5 합성곱을 병렬로 적용하는 모습을 볼 수 있다. Inception 모델은 1x1 합성곱을 추가로 사용하여 차원을 축소하고, 연산량을 줄이도록 설계한다[6].

3.3 Xception

Xception은 깊이별 분리 합성곱을 사용하여 Inception 모듈을 대체한 구조이다[7]. 그림 4에서 깊이별 합성곱은 각 입력 채널에 3x3 합성곱을 독립적으로 적용해 특징 맵을 생성하고, 점별 합성곱은 모든 채널에 1x1 합성곱을 수행하여 채널 수를 조절한다. 이를 통해 파라미터 수와 계산 복잡도를 줄이면서 성능을 유지하고, 간결한 모델 구조를 제공한다.

3.4 ResNet

깊은 모델은 복잡한 패턴을 학습할 수 있지만, 깊이가 늘어날수록 기울기 소실 문제에 직면한다. ResNet은 이를 해결하기 위하여 잔차 블록(residual block)을 사용한다[8,9]. 그림 5 (a)는 일반적인 방식으로 가중치 층과 ReLU 활성화 함수를

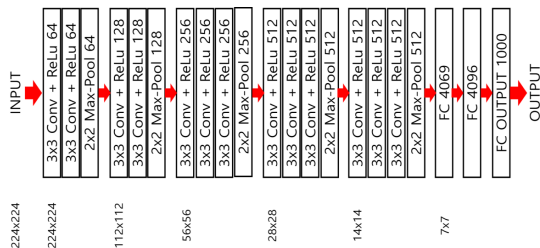


그림 2. VGG 구조. Fig. 2. VGG architectures.

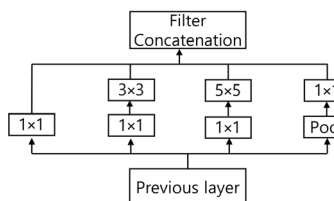


그림 3. Inception 구조. Fig. 3. Inception architectures.

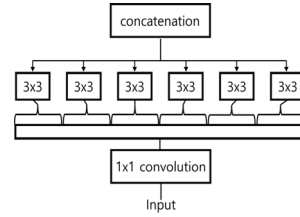


그림 4. Xception 구조. Fig. 4. Xception architectures.

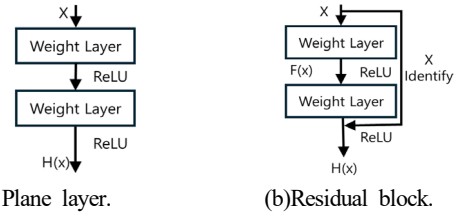


그림 5. 기존방법과 잔차 블록. Fig. 5. Plane layer and residual block.

연속적으로 적용하며, 그림 5 (b)는 잔차 블록 구조로, 스킵 연결을 통해 입력 X가 그대로 더해져 신호가 보존된다. 이를 통해 깊은 네트워크에서도 정보 손실이 줄어 성능이 향상된다.

3.5 DenseNet

ResNet은 스킵 연결을 통해 원천 정보를 뒤 레이어로 직접 전달하지만, 단순 덧셈으로 결합되어 정보 흐름에 한계가 있다. 반면, DenseNet은 그림 6과 같이 특성 맵 연결(concatenate) 방식을 사용하여 이전 레이어를 모든 이후 레이어에 직접 연결한다[10,11]. 이를 통해 정보와 기울기 소실을 방지하고, 적은 채널 수로 파라미터와 연산량을 줄여 높은 성능과 학습 효율성을 유지한다.

3.6 MobileNet

MobileNet은 Xception에서 사용된 깊이별 분리 합성곱의 연산효율이 뛰어난 장점에 집중하여 모바일 기기에서 동작 가능한 정도로 경량화된 네트워크를 설계하였다. MobileNet은 깊이별 분리 합성곱을 매우 단순하고 일관되게 전체 네트워크에서 사용한다. 하나의 깊이별 합성곱과 그다음 이어지는 1x1 포인트 합성곱으로 구성된 기본 구조가 모든 합성곱 계층에 적용된다. 이 구조는 매우 단순한 방식으로 계산 복잡도를 크게 줄이는데 초점을 맞춘다. 그림 7은 MobileNet의 네트워크 구조이다[12].

III. 데이터 준비

본 연구에서는 AI 딥러닝 알고리즘을 활용한 분류 시스템을 제한한다. 5종의 자동차 대시보드를 대상으로 동일한 환경에서 이미지를 캡처하여 데이터셋을 구축한다.

1. 데이터 수집 및 전처리

본 연구에서는 데이터셋의 양이 제한적인 상황에서 ROI (Region of Interest) 기법을 활용하여 대시보드의 중요한 패턴 부분만을 선택하여 데이터를 수집하였다. ROI는 이미지에서 핵심적인 특징을 강조하면서도 불필요한 정보를 배제하는 효과적인 접근 방식으로 이를 통해 대시보드의 세부 패턴을 명확하게 추출할 수 있었다. 추가적인 데이터를 수집할 경우

비슷한 패턴의 반복으로 인해 모델이 과적합(overfitting)에 취약해질 가능성이 높았다. 따라서 본 연구에서는 데이터 증강 기법을 적용하여 기존 데이터셋을 다양화하고 학습 데이터의 양을 효과적으로 늘리는 방식을 채택하였다. 이를 통해 데이터 제한 문제를 해결하면서도 모델의 일반화 성능을 유지할 수 있었다.

5개의 대시보드를 각각의 클래스로 구분하여 각 클래스당 100장의 이미지를 수집하였고, 총 500장의 데이터셋을 구축하였다. 구축한 데이터셋은 그림 8과 같이 ROI를 사용하여 추출한 후, 7:2:1의 비율로 훈련(train), 검증(validation), 테스트 데이터셋으로 나누었다. 각 클래스당 70장의 훈련 이미지, 20장의 검증 이미지, 10장의 테스트 이미지를 사용하였으며, 전체 데이터셋은 훈련용 350장, 검증용 100장, 테스트용 50장으로 구성되었다.

본 연구에서는 모델의 안정성과 메모리 효율성을 위하여 하이퍼파라미터를 Dense 128, BatchSize 32, Epochs는 50으로 설정하였다.

2. 데이터 증강

학습 과정에서 데이터셋의 부족이나 불균형으로 인한 과적합을 방지하기 위해 데이터 증강 기법을 사용한다. 데이터 증강은 수집한 이미지 데이터셋에 다양한 변형 기법을 적용하여 데이터를 확장함으로써, 모델이 더 많은 데이터를 학습하고 일반화 성능을 향상시키는 방법이다[13]. 본 연구에서는 ImageDataGenerator를 활용하여 원본 데이터를 미리 복사해 저장하지 않고, 훈련 중 매 배치(batch)마다 이미지를 불러올 때 랜덤하게 증강하여 모델에 입력하는 방식을 채택하였다. ImageDataGenerator는 딥러닝에서 이미지 데이터를 실시간으로 증강하거나 전처리하는 데 사용되는 유용한 도구로 모델 학습 성능을 높이고 과적합을 방지하는 역할을 한다. 이를 통해 회전, 확대/축소, 좌우 및 상하 이동, 전단 변환, 수평 반전, 빈 영역 채우기 등의 증강 기법을 훈련 및 검증 단계의 데이터셋에 적용하였다. 증강 과정에서는 각 기법을 랜덤하게 선택하여 데이터를 변형하였으며, 이러한 방식을 통해 생성된 데이터의 다양성을 극대화하였다. 증강 후 측정 결과, 원본 데이터셋의 약 50배에 달하는 데이터가 생성되었으며, 이를 활용하여 과적합을 효과적으로 방지할 수 있었다.

한편, 메타모픽 테스트는 테스트 데이터셋에서만 변형을 적용하는 방식으로 수행되었다. 학습 전 데이터 증강은 학습 및 검증 단계의 데이터셋에만 적용되었으며, 이를 통해 만들어진 모델들은 메타모픽 테스트를 이용하여 변형된 테스트 데이터셋에서 성능을 평가하였다. 이 과정은 학습과 테스트 단계에서 데이터 증강이 모델의 성능과 일반화 능력 향상에 기여하였다.

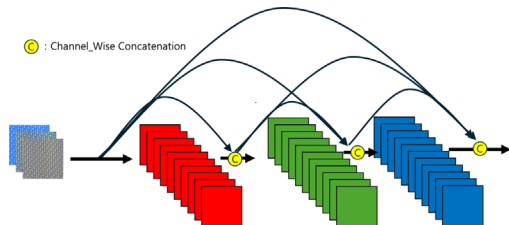


그림 6. DenseNet 구조.  
Fig. 6. DenseNet architectures.

Type / Stride	Filter Shape	Input Size
Conv / s2	3 × 3 × 3 × 32	224 × 224 × 3
Conv dw / s1	3 × 3 × 32 dw	112 × 112 × 32
Conv / s1	1 × 1 × 32 × 64	112 × 112 × 32
Conv dw / s2	3 × 3 × 64 dw	112 × 112 × 64
Conv / s1	1 × 1 × 64 × 128	56 × 56 × 64
Conv dw / s1	3 × 3 × 128 dw	56 × 56 × 128
Conv / s1	1 × 1 × 128 × 128	56 × 56 × 128
Conv dw / s2	3 × 3 × 128 dw	56 × 56 × 128
Conv / s1	1 × 1 × 128 × 256	28 × 28 × 128
Conv dw / s1	3 × 3 × 256 dw	28 × 28 × 256
Conv / s1	1 × 1 × 256 × 256	28 × 28 × 256
Conv dw / s2	3 × 3 × 256 dw	28 × 28 × 256
Conv / s1	1 × 1 × 256 × 512	14 × 14 × 256
5× Conv dw / s1	3 × 3 × 512 dw	14 × 14 × 512
Conv / s1	1 × 1 × 512 × 512	14 × 14 × 512
Conv dw / s2	3 × 3 × 512 dw	14 × 14 × 512
Conv / s1	1 × 1 × 512 × 1024	7 × 7 × 512
Conv dw / s2	3 × 3 × 1024 dw	7 × 7 × 1024
Conv / s1	1 × 1 × 1024 × 1024	7 × 7 × 1024
Avg Pool / s1	Pool 7 × 7	7 × 7 × 1024
FC / s1	1024 × 1000	1 × 1 × 1024
Softmax / s1	Classifier	1 × 1 × 1000

그림 7. MobileNet 구조.  
Fig. 7. MobileNet architectures

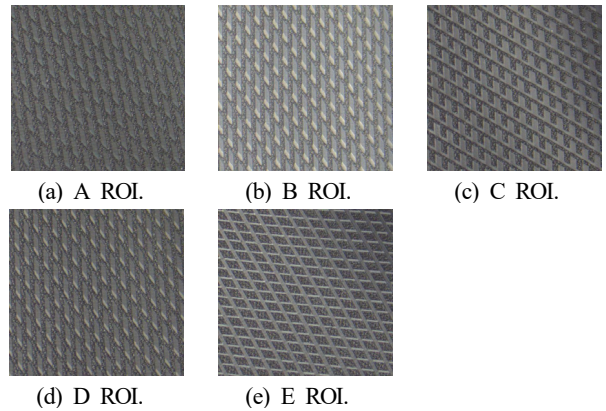


그림 8. 대시보드 ROI 사진.  
Fig. 8. Dashboard ROI.

IV. 실험 및 결과

본 연구에서는 사전 학습된 9개의 CNN 모델들을 통해 분류 성능을 비교하였다. 비교한 모델들은 정확도, 손실률, F1-Score 등의 성능 지표를 바탕으로 성능을 비교 분석하였다. F1-Score의 경우 메타모픽 테스트 케이스 생성 기법을 사용하여 실제 입력 환경에서 모델의 성능이 유지될 수 있는지 확인한다.

1. 정확도와 손실률

정확도와 손실률은 모델의 성능을 분석하는데 사용하는 지표 중 하나이다. 정확도는 식 (1)과 같이 분류 문제에서 모델이 전체 데이터 중에서 올바르게 예측한 데이터의 비율을 의미한다.

$$Accuracy = \frac{Number\ of\ Correct\ Predictions}{Total\ Number\ of\ Predictions} \times 100 \quad (1)$$

학습 정확도는 모델이 학습 데이터에서 얼마나 잘 학습했는지를, 검증 정확도는 모델이 학습한 내용을 새로운 데이터에 얼마나 잘 일반화할 수 있는지를 나타낸다.

손실률은 분류 문제에서 사용되는 손실 함수 중 하나인 크로스 엔트로피를 사용한다. 식 (2)는 손실률의 수식으로 N은 샘플의 수를, C는 클래스의 수를 나타내며,  $y_{i,c}$ 는 실제 라벨의 의미하며 샘플이 클래스 c에 속하면 1, 그렇지 않으면 0이 된다.  $\hat{y}_{i,c}$ 는 샘플 i에 대해 모델이 예측한 클래스 c의 확률이다.

$$Loss = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C y_{i,c} \log(\hat{y}_{i,c}) \quad (2)$$

손실률은 모델이 예측한 값과 실제 값 사이의 차이를 수치적으로 표현하며, 이 차이가 클수록 손실값이 커지고 차이가 작을수록 손실값이 작아진다. 즉 손실률이 낮을수록 모델의 예측이 더 정확하다는 의미이다.

표 1은 다양한 딥러닝 모델에 대한 학습과 검증을 이용하여 정확도와 손실률을 비교한 것이다. ResNet152V2, Xception과 InceptionV3이 검증 정확도 및 손실에서 가장 높은 성능을 보여주며, 특히 ResNet152V2는 손실 값에서 매우 우수한 결과를 보였다 이 두 모델은 복잡한 이미지 분류 문제에서 가장 강력한 성능을 제공할 수 있다. DenseNet 계열 모델들 또한 안정적이고 좋은 성능을 보여준다. 이들은 높은 검증 정확도와 낮은 손실 값을 기록했으며, 최신 모델들에 비해 효율적인 학습 성능을 보인다. VGG16과 VGG19는 과거에는 널리 사용되었지만, 최신 모델들에 비해 성능이 비교적 떨어진다. 특히 손실 값이 크게 높고, 검증 정확도 또한 다른 최신 모델들에 비해 낮다. 모든 모델들을 종합적으로 비교했을 때, ResNet152V2와 Xception이 최고 성능을 보였으며, MobileNetV2는 경량 모델 중에서 성능과 효율성 측면에서 인상적이다. 하지만 정확도와 손실률은 데이터가 불균형한 경우에는 적합한 평가 지표가 아닐 수 있다. 이러한 문제를 방지하고자 F1-Score 또한 성능지표로 사용한다.

2. F1-Score

F1-Score는 정밀도(precision)와 재현율(recall) 간의 조화평균을 계산하여 모델의 성능을 종합적으로 평가하는 지표이다. 정밀도는 모델이 예측한 긍정 클래스 중에서 실제로 올바르게 예측한 비율을 나타내며, 재현율은 실제로 긍정 클래스인 샘플 중에서 모델이 올바르게 예측한 비율을 의미한다. 정밀도와 재현율은 다음과 같은 식 (3), 식 (4)로 정의된다[14].

$$Precision = \frac{TruePositive}{TruePositives + FalsePositive} \quad (3)$$

$$Recall = \frac{TruePositive}{TruePositives + FalseNegatives} \quad (4)$$

F1-Score는 정밀도와 재현율 간의 조화평균을 구하여, 두 값이 균형을 이루는지를 평가하는 역할을 한다. 특히, 정밀도만 높고 재현율이 낮거나, 반대로 재현율만 높고 정밀도가 낮을 때, F1-Score는 이를 보정하여 한쪽으로 치우친 평가를 방지한다. F1-Score는 식 (5)와 같다.

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (5)$$

F1-Score는 0에서 1 사이의 값으로 표현되며, 값이 1에 가까울수록 모델의 성능이 더 우수하다는 것을 의미한다. 그림 10은 메타모픽을 적용하지 않은 테스트 셋의 경우 Xception과 DenseNet201 가장좋은 점수를 보여주고 그 다음으로 Inception과 DenseNet121이 좋은 성능을 보여주고 있다. MobileNetV2의 경우 정확도와 손실률에서 좋은 평가를 보였으나 F1-Score에서 낮은 점수를 보여주며 정밀도와 재현율 간의 불균형이 존재할 가능성을 보여주었다,

표 1. 모델 정확도와 손실률.

Table 1. Model accuracy and loss.

Model	Accuracy		Loss	
	Train	Val.	Train	Val.
DenseNet121	0.977	0.977	0.017	0.045
DenseNet201	0.979	0.970	0.052	0.067
ResNet50V2	0.981	0.992	0.016	0.063
ResNet152V2	0.979	0.985	0.022	0.050
MobileNetV2	0.990	0.979	0.026	0.045
VGG16	0.902	0.875	0.057	0.062
VGG19	0.972	0.948	0.094	0.160
InceptionV3	0.983	0.963	0.057	0.095
Xception	0.989	0.985	0.051	0.065

3. 메타모픽 테스트(Metamorphic Test)

메타모픽 테스트는 모델이 변형된 입력 데이터에서도 일관된 성능을 유지하는지 평가하는 기법으로, 주로 딥러닝 기반 이미지 분류 모델의 강건성을 확인하는 데 활용된다. 메타모픽 테스트의 핵심은 변형된 입력 데이터와 원본 데이터에 대해 모델이 얼마나 일관된 성능을 발휘하는지를 측정하는 것이다. 메타모픽 테스트는 입력 변형에 따른 성능 저하를 정량적으로 분석함으로써, 모델이 다양한 환경에서 얼마나 잘 작동하는지 파악할 수 있다[15].

본 연구에서는 다양한 CNN 모델에 대해 메타모픽 테스트를 이용하여 테스트셋을 증강하고 이를 통해 모델의 일반화 성능을 평가하였다. 메타모픽 테스트에서 사용하는 변형 기법을 간단한 변형(simple meta)과 극단적 변형(extreme meta)으로 나누었다. 간단한 변형은 이미지의 90° 회전, 270° 회전, 상하 반전, 좌우 반전, 크기 변환 등을 적용하여 테스트 셋을 6배로 늘렸다. 이러한 변형은 현실적인 데이터 변형을 모방하며, 모델이 이러한 간단한 변형에도 일관된 성능을 유지하는지 평가한다. 극단적 변형은 이미지의 상하좌우 각각 3픽셀 이동, 여백에 노이즈 추가, RGB 채널을 하나씩 삭제, 오브젝트에 노이즈 추가 등의 변형을 적용하여 테스트 셋을 10배로 증가시켰다. 이러한 변형은 데이터의 중요한 정보를 왜곡하거나 손상시켜, 모델의 강건성을 극한으로 평가하는 데 사용된다.

메타모픽 테스트에서 성능을 비교하기 위하여 F1-score를 기준으로 한 성능 유지 비율을 사용하였다. 식 (6)은 메타모픽 테스트의 F1-Score 와 일반 F1-Score의 비율로 x'와 원본 데이터 x의 예측결과를 비교하여, 변형 후 성능이 원본 대비 어느 정도 유지되는지를 평가한다.

$$MR_{F1} : \frac{F1(P(x), P(x'))}{F1(P(x), P(x))} \geq \theta \quad (6)$$

MR<sub>F1</sub>(Metamorphic Relation F1-Score)은 각 모델에 대해 계산된 변형 후 성능 유지 비율을 뜻한다, P(x)는 원본 입력 x에 대한 모델의 예측값, P(x')는 변형된 입력 x'에 대한 모델의 예측값, F1은 예측 성능을 나타내는 F1-Score이다. θ는 성능 유지 기준값으로, θ=0.9로 설정하여 변형 후 성능이 원본 성능의 90% 이상 유지되는지 확인하였다.

그림 9에서 CNN 모델에 대해 메타모픽 테스트를 적용한 결과, Xception, InceptionV3, DenseNet201 등은 대부분의 변형된 데이터에서도 높은 F1-Score를 보여주었다. 반면, DenseNet121,

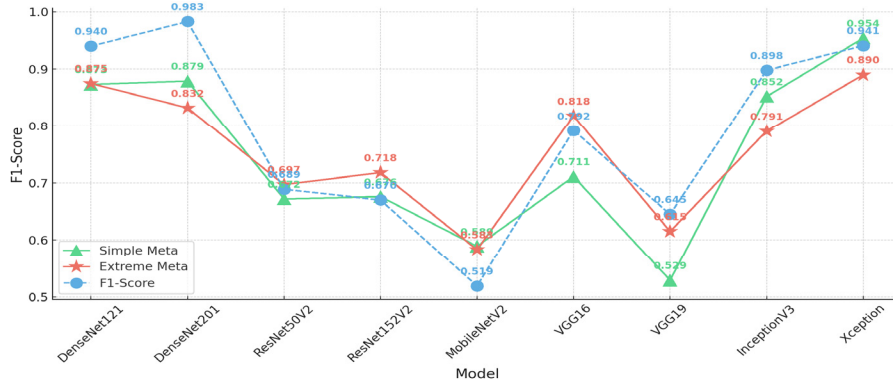


그림 9. 기본, 간단한 변형 그리고 극단적 변형의 F1-Score.

Fig. 9. Basic, simple, and extreme metamorphic F1-score.

표 2. 메타모픽 관계 F1-Score 유지율.

Table 2. Metamorphic relation F1-score retention.

Model	$MR_{F1}$	
	Simple Meta	Extreme Meta
DenseNet 121	0.93	0.93
DenseNet201	0.98	0.93
ResNet50V2	0.98	1.01
ResNet152V2	1.01	1.01
MobileNetV2	1.13	1.12
VGG16	0.89	0.81
VGG19	0.82	0.82
InceptionV3	0.95	0.88
Xception	1.01	0.94

ResNet50과 같은 얇은 모델은 깊은 모델에 비해 연산량이 적고, 데이터의 불균형이나 변형된 데이터에 대해 과적합되지 않는 특성이 있을 수 있다. 또한, 변형된 데이터셋이 모델의 일반화 능력을 더 효과적으로 평가했을 가능성도 제기된다. 이러한 결과는 특정 상황에서 얇은 모델이 더 효율적일 수 있음을 시사하며, 모델 구조의 복잡도가 항상 성능 향상을 보장하지 않는다는 점을 보여준다. 또한 표 2는  $MR_{F1}$ 가 계산된 것으로 Xception, InceptionV3, DenseNet201은  $MR_{F1} \geq 0.9$ 를 만족하여, 일관된 성능을 유지하는 것으로 나타났다. 결과는 변형된 데이터에서도 높은 성능을 유지할 수 있는 모델을 선택하는데 중요한 기준을 제공하며, 실질적인 응용에서의 일반화 성능을 평가하는 데 있어 메타모픽 테스트의 유용성을 확인할 수 있다.

실험 결과 Xception, InceptionV3, DenseNet201 알고리즘 모두 정확도와 F1-Score 높고 메타모픽 테스트를 사용한 F1-Score와  $MR_{F1}$  또한 높은 점수를 보여주며 알고리즘에 대한 높은 일관성과 강건성을 보여주었다. 이러한 실험 결과는 9개의 알고리즘 중 3개의 알고리즘이 본 연구에서 효과적으로 내장재를 분류하는 데에 적합한 모델임을 보여주었다.

V. 결론

본 연구에서는 AI 딥러닝 기반의 CNN 알고리즘을 활용하여 자동차 내장재, 특히 대시보드의 분류 문제를 해결하고자 하였다. 전통적인 수작업 분류 방식이 가지고 있던 비효율성과 일관성

문제를 개선하기 위하여 ResNet과 Xception 모델 등 총 9개의 사전 학습된 CNN 모델을 정확도, 손실률, F1-Score 그리고 메타모픽 테스트를 이용하여 다양한 데이터 변형에도 모델이 일관된 성능을 유지하는지 비교 분석하였다. 실험 결과, Xception, DenseNet201 모델은 정확도와 강건성 면에서 우수했으며, 메타모픽 테스트를 통한 성능 검증이 분류 시스템에서의 신뢰도를 높이는 데 중요한 역할을 한다는 점이 강조되었다. 이러한 결과는 CNN 모델이 자동차 내장재 분류에서 높은 성능을 보이며, 메타모픽 테스트를 통해 실제 환경에 적합한 모델을 선택할 수 있는 기준을 제공함으로써 기존의 수작업 방식을 대체할 수 있는 효과적인 방법임을 확인할 수 있었다.

향후 연구에서는 모델별 하이퍼파라미터와 레이어등을 수정하여 오차의 편차를 줄이고 모델들의 성능을 더 높여 정확도 및 F1-Score를 높이고자 한다.

REFERENCES

- [1] T. Y. Chen, S. C. Cheung, and S. M. Yiu “Metamorphic testing: A new approach for generating next test cases,” Technical Report HKUST-CS98-01, Dept. of Computer Science, The Hong Kong University of Science and Technology, 1998.
- [2] Y. W. Kim and D. S. Kang, “Generating composite variable rules for metamorphic testing of CNN image classification model,” *The Trans. of the Korean Institute of Electrical Engineers, Transactions on Computing Practices*, vol. 30, no. 2, pp. 77-83, 2024.
- [3] J. H. Ding, X. J. Kang, and X. H. Hu, “Validating a deep learning framework by metamorphic testing,” *IEEE/ACM 2nd Int. Workshop on Metamorphic Testing (MET)*, pp. 28-34, 2017.
- [4] S. Kornblith, J. Shlens, and Q. V. Le, “Do better imagenet models transfer better?,” *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2661-2671, 2019.
- [5] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” arXiv preprint arXiv:1409.1556, 2014.

- [6] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," *Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 2818-2826, 2016.
- [7] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1251-1258, 2017.
- [8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 770-778, 2016.
- [9] H. S. Kim, B. W. Cho, and B. J. Kang, "Deep learning-based landmark identification for the upper gastrointestinal track in endoscopic images," *Journal of Institute of Control, Robotics and Systems (in Korean)*, vol. 29, no. 11, pp. 908-9013, 2023.
- [10] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "densely connected convolutional networks," *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2261-2269, 2017.
- [11] C. H. Park and H. D. Choi, "Enhanced parallel sparse-MLP for monocular depth estimation of autonomous UAV," *Journal of Institute of Control, Robotics and Systems (in Korean)*, vol. 29, no. 11, pp. 928-935, 2023.
- [12] A. G. Howard, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," arXiv preprint arXiv:1704.04861, 2017.
- [13] Y. W. Choi, Y. W. Lee, and H. S. Chae, "Development of an image data augmentation apparatus to evaluate CNN mode," *Journal of Software Engineering Society*, vol. 29, no. 1, pp. 13-21, 2020.
- [14] R. Yacouby and D. Axman, "Probabilistic extension of precision, recall, and F1 score for more thorough evaluation of classification models," *Proc. of the First Workshop on Evaluation and Comparison of NLP Systems*, pp. 1-8, 2020.
- [15] B. W. Na and D. S. Kang, "A method for generating metamorphic test cases for CNN image classification models," *KIISE Trans. on Computing Practice*, vol. 28, no. 1, pp. 33-41, 2022.



### 정종민

2022년 경남대학교 전기공학과(공학사). 2023년~현재 경남대학교 대학원 메카트로닉스공학과(공학석사). 관심분야는 딥러닝 기반의 불량품 검출 및 분류, 지능로봇 제어.



### 김동현

2001년 한양대학교 전기공학과(공학박사). 2001년~2003년 미국 듀크 대학교 연구원. 2003년~2004년 미국 보스턴 대학교 연구원. 2004년~2005년 일본 동경대학교 박사 후 과정. 20012년~2013년 영국 에식스대학교 방문교수. 2005년~현재 경남대학교 전기공학과 교수. 관심분야는 이동로봇의 경로계획과 자율주행, 드론 및 무인수상선 제어, 지능로봇 제어, 딥러닝 기반의 불량품 검출 및 분류.