

메타데이터 분석과 합성 판별 인공지능을 활용한 C2C 거래 사기 방지 시스템 설계 및 구현

권혁찬*¹, 남주영*², 정용태*³, 최태용*⁴, 정기현**

Design and Implementation of a Fraud Prevention System for C2C Transactions using Metadata Analysis and Synthetic Discriminative AI

Hyukchan Kwon*¹, Juyoung Nam*², Yongtae Jung*³, Taeyong Choi*⁴, and Kihyun Jung**

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(2021R111A3049788)

요 약

본 연구에서는 C2C 플랫폼에 대한 신뢰도를 향상하고 안전한 거래 환경을 조성하는 것을 목표로, CNN 기반의 AI 모델과 이미지 메타 데이터 분석을 결합한 이미지 무결성 탐지 시스템을 개발하고자 한다. 이를 위해 추출한 메타데이터를 분석하여 판단기준을 세우고 해당 이미지의 메타데이터를 비교, 검증하여 무결성을 검증하고 도용 여부를 확인한다. 이미지 합성 여부 판별을 위해 이미지 전처리 단계에서 HLS 채널을 활용한 이미지를 기반으로 Resnet-50, 자체 CNN 및 ViT 모델을 사용한다. 실험 결과 Resnet-50 모델과 ViT 모델이 높은 정확도를 보여주고, EXIF 메타데이터를 추출하여 사진의 출처와 도용 여부를 판별하는 기능을 구현하였다.

Abstract

This study aims to enhance the reliability of C2C platforms and create a secure trading environment by developing an image integrity detection system that combines a CNN-based AI model with image metadata analysis. To achieve this, extracted metadata is analyzed to establish evaluation criteria, and the metadata of a given image is compared and verified to assess its integrity and detect plagiarism. For image forgery detection, the system utilizes images processed with the HLS channel in the preprocessing stage and employs ResNet-50, a custom CNN, and ViT models. Experimental results show that the ResNet-50 and ViT models achieve high accuracy, successfully extracting EXIF metadata to determine image sources and detect plagiarism.

Keywords

image integrity detection system, CNN-based AI models, EXIF metadata, resnet-50, vit model, HLS channels, image sythesis detection, safe transaction environment

* 국립안동대학교 소프트웨어융합학과

- ORCID¹: <https://orcid.org/0009-0007-0060-7872>

- ORCID²: <https://orcid.org/0009-0003-6117-7598>

- ORCID³: <https://orcid.org/0009-0007-0744-9811>

- ORCID⁴: <https://orcid.org/0009-0001-2314-3851>

** 국립경국대학교(국립안동대학교) 소프트웨어융합학과 교수
(교신저자)

- ORCID: <https://orcid.org/0000-0002-0662-8355>

• Received: Jan. 09, 2025, Revised: Apr. 09, 2025, Accepted: Apr. 12, 2025

• Corresponding Author: Kihyun Jung

Dept. of Software Convergence, Gyeongbuk National University
(Andong National University), Korea

Tel.: +82-54-820-7968, Email: kingjung@anu.ac.kr

I. 서 론

우리는 1970년대부터 벼룩시장, 아나바다 운동 등 오프라인 형태로 C2C(Consumer-to-Consumer) 거래를 경험해오고 있다. 2000년대 초 인터넷과 모바일 기술의 상용화로 오프라인 C2C 거래는 줄어들고, 온라인 형태의 C2C가 자연스럽게 등장하며 빠르게 성장했다. 초기에는 블로그나 네이버 카페와 같은 웹 플랫폼에서 거래가 이루어졌으나, 스마트폰 확산과 모바일 앱 기반 플랫폼(예: 당근마켓, 중고나라)의 등장으로 공간적 제약이 해소되며 C2C 거래 시장은 더욱 확장되었다. 이를 대표하는 사례로 리셀 시장과 중고 거래 시장이 있다.

2020년 코로나바이러스 확산으로 경제와 사회 전반에 부정적 영향을 미치며 소비심리가 위축되었으나, ‘중고 시장은 불황을 먹고 자란다’는 말처럼 중고 거래는 오히려 급성장했다[1]. 2020년 9월 1,170만 명이던 중고 거래 플랫폼 이용자 수는 2022년 8월 1,928만 명으로 증가하며 2년간 약 65% 성장했다. 이는 중고 거래가 가격 대비 성능이 우수한 합리적 소비로 인식되고, 친환경적 소비 트렌드와 맞물리며 현대 소비자들의 일상적 소비 형태로 자리 잡았음을 보여준다. 그러나 시장 규모 확장과 함께 C2C 거래의 문제점도 두드러졌다. 가장 큰 문제는 누구나 쉽게 판매자와 구매자가 될 수 있다는 점이다. 신뢰도가 검증된 기업과 개인이 거래하는 B2C(Business-to-Consumer)와 달리, C2C는 개인 간 신뢰에 의존한다. 온라인 환경 특성상 제한된 정보만으로 거래가 이루어져 사기 위험이 크다. 특히, 거래 품목을 확인하는 주요 수단인 이미지는 AI와 이미지 합성 기술의 발전으로 육안으로 합성 여부를 판별하기 어렵고, 도용된 경우 이를 알아차리기 더욱 힘들다. 실제로 온라인 C2C 거래 사기 피해가 급증하고 있다[2]. 최근 9년간(2023년 기준) 중고 거래 사기는 81.4% 증가했으며, 2022년 하루 평균 228건, 총 8만 3,214건이 접수되었다. 피해 금액도 2014년 278억 원에서 2021년 3,606억 원으로 13배 상승하였다.

현재 C2C 사기 방지를 위해 경찰청 ‘사이버안전국’ 홈페이지, ‘더치트(The Cheat)’ 사이트, ‘사이버캡’ 앱 등에서 판매자의 전화번호나 계좌번호를 조

회해 최근 3개월간 사기 신고 여부를 확인할 수 있다. 그러나 이는 신고된 사례에 의존하며 실시간 탐지가 불가능하다. 또한, 전자상거래법은 B2C 보호에 초점을 맞추어 C2C 피해자는 법적 보호를 받기 어렵고[3], 사기 신고 후에도 배상 절차가 복잡하여 피해금 회수는 불확실한 상황이다.

기존 연구는 이러한 사기 문제를 해결하고자 텍스트 분석과 이미지 분석을 시도했다. 예를 들어, 리뷰 텍스트를 LSTM과 Word2Vec으로 분석해 사기 패턴을 탐지하거나[4], CNN으로 합성 이미지를 판별하는 연구가 있었다[11]. 그러나 텍스트 분석은 문맥 이해 부족으로 오탐지가 빈번했고, 이미지 분석은 메타데이터를 활용하지 않아 도용 여부를 확인하지 못했다. EXIF 메타데이터를 활용한 연구는 촬영기기 식별과 같은 포렌식 목적에 국한되어 C2C 사기 방지에 적용되지 않고 있다[7][10].

이에 본 연구는 온라인 C2C 거래 사기의 심각성을 인식하고, 기존 연구의 한계를 극복하는 방향을 모색한다. 구체적으로, EXIF 메타데이터를 활용해 이미지의 무결성과 도용 여부를 검증하고, CNN 기반 AI 모델과 결합해 합성 여부를 실시간 탐지하는 시스템을 제안한다. 예를 들어, 1:1 채팅방에서 송수신된 물품 이미지를 분석해 동일 이미지 ID가 여러 거래에서 사용되었는지, 또는 합성 흔적이 있는지를 판별한다. 이는 텍스트 의존적 접근법의 주관성과 CNN 단독 사용의 한계를 넘어, 객관적이고 정확한 사기 방지 솔루션을 제공한다. 본 연구는 거짓 정보 유통을 차단하고, C2C 플랫폼의 신뢰도를 높여 안전한 거래 환경을 조성하는 것을 목표로 개발되었다.

II. 관련 연구

2.1 C2C 관련 사기 방지 메타데이터 기술

C2C 거래 사기 방지를 위해 메타데이터 기술이 활용되고 있으며, 대표적으로 리뷰 분석을 통해 판매자의 사기 가능성을 탐지하는 방법이 있다. 예를 들어, LSTM 모델과 Word2Vec을 결합하여 웹 크롤링된 리뷰에서 키워드와 유사 단어를 추출하고, 오타를 정제하여 사기성 패턴을 탐지한다[4].

이 과정에서 정제된 단어는 Word2Vec으로 사전 학습된 LSTM 모델을 통해 원래 리뷰의 단어와 대체되며, 거래 전 사기 가능성을 줄이는 데 기여한다. 그러나 이 접근법은 몇 가지 한계를 안고 있다. 첫째, 언어 모델의 문맥 이해 부족으로 인해 ‘급매’와 같은 단어가 정상 거래인지 사기인지 구분하기 어렵다. 둘째, 리뷰는 주관적 표현과 감정이 포함되어 객관적 판단에 한계가 있다.

이에 비해 본 연구는 텍스트 리뷰 대신 이미지의 EXIF 메타데이터를 활용하여 사기를 방지하는 새로운 접근법을 제안한다. 이미지 메타데이터는 주관적 해석의 여지가 적고, 카메라 제조사, 촬영 날짜, 이미지 ID, 해상도 등 객관적 정보를 제공한다[5]. 이는 리뷰 기반 방법의 문맥 의존성과 주관성 문제를 극복하며, 거래 전 사기를 보다 정확히 예방할 수 있는 기반을 마련한다. 기존 연구가 텍스트에 의존하며 발생하는 불확실성을 줄이고, 이미지 기반의 신뢰성 높은 분석을 통해 차별성을 확보한다.

2.2 EXIF 기반 이미지 분석

EXIF(Exchangeable Image File Format)는 일본 전자 산업진흥협회(JEIDA)에서 개발된 포맷으로, 초기에는 날짜, 시간, 섬네일과 같은 기본 메타데이터를 저장했으나, 현재는 카메라 설정, 저작권 정보 등 다양한 데이터를 포함한다[6]. EXIF를 활용한 기존 연구는 주로 이미지의 출처와 무결성 분석에 초점을 맞췄다. 예를 들어, Kee 등과 Gloe는 EXIF의 섬네일 정보, 메타데이터 개수, 파일 구조를 분석하여 이미지 촬영 기기와 위치를 식별하려 했다[7][8]. Castiglione 등은 온라인 소셜 네트워크에서 이미지 포맷과 JPEG 메타데이터를 분석해 출처를 추적했다[9]. Mullan 등은 JPEG 헤더의 DQT(Define Quantization Tables)와 EXIF 태그를 활용하여 Apple 스마트폰 기기를 분류했으며[10], Giudice 등은 머신러닝과 메타데이터를 결합해 이미지 업로드 위치(iOS, 웹, Android)를 판별했다[11].

그러나 이들 연구는 몇 가지 문제점을 드러낸다. 첫째, EXIF 기반 연구는 주로 촬영기기 식별 등과 같은 포렌식 목적에 국한되어 C2C 거래 사기 방

에 직접 적용되지 않았다. 둘째, 합성 이미지 탐지와 연계가 부족해 도용 여부를 넘어선 사기성 콘텐츠 판별에 한계가 있었다. 셋째, 메타데이터의 무결성 검증이 간과되어 조작된 EXIF 데이터에 취약할 수 있다.

본 연구는 이러한 한계를 해결하고자 EXIF 메타데이터를 활용해 이미지의 무결성과 도용 여부를 검증한다. 예를 들어, 동일 이미지 ID나 해상도가 여러 거래에서 반복될 경우 도용 가능성을 탐지하며, 이를 CNN 기반 합성 탐지와 결합해 사기 여부를 종합적으로 판단한다. 이는 기존 연구의 포렌식 중심 접근법을 넘어 C2C 거래 환경에 특화된 실용적 해결책을 제시한다.

2.3 CNN

CNN(Convolutional Neural Network)은 이미지 및 영상 처리에서 패턴 인식과 분류를 위해 주로 사용되는 인공지능 모델이다. CNN은 컨볼루션 계층, 풀링 계층, 완전 연결 계층으로 구성되며, 입력 데이터의 지역적 특징을 추출하고 크기를 축소하며 최종 분류를 수행한다[12]. 컨볼루션 계층은 필터를 통해 패턴을 인식하고, 풀링 계층은 데이터 크기를 줄여 학습 효율성을 높이며, 완전 연결 계층은 결과를 출력한다. CNN은 번역 불변성(이미지 위치 변화에도 동일한 특징 추출 가능)을 갖추어 이미지 처리에 강점을 보인다[13][14].

C2C 사기 방지에서 CNN은 합성 이미지를 판별하는 데 유용하다. 그러나 기존 연구에서는 주로 이미지 자체의 시각적 특징에만 의존했으며, 메타데이터와의 통합이 부족했다. 또한, CNN의 학습 데이터 의존성이 높아 품질 낮은 데이터나 부족한 데이터셋에서는 성능이 저하될 수 있다.

본 연구는 CNN을 활용해 합성 이미지를 탐지하면서, EXIF 메타데이터를 추가 분석하여 이미지의 출처와 무결성을 확인한다. 이는 CNN 단독 사용의 한계를 보완하고, 데이터 품질에 덜 의존적인 객관적 판단 기준을 제공한다. 기존 CNN 연구와 달리 메타데이터와의 융합으로 사기 탐지 정확도를 높이는 데 중점을 두고 있다.

2.4 HLS 채널

HLS 채널은 색조(Hue), 명암(Luminance), 포화도(Saturation)로 구성된 색상 공간 모델로, 디지털 영상 처리에서 활용된다. 색조는 0~360도 범위로 색상을 나타내며(빨강 0, 초록 120 등), 명암은 밝기를, 포화도는 색상의 강도를 표현한다. 본 연구에서는 HLS 채널을 이미지 전처리 단계에 적용하여 CNN 학습 데이터를 생성한다.

기존 연구에서는 HLS를 활용한 사례가 드물었으며, 주로 RGB 채널 기반 전처리에 의존했다. 그러나 RGB는 색상과 밝기의 상호작용을 분리하기 어려워 합성 이미지 판별에서 세부적인 특징 추출에 한계가 있었다. 반면, HLS는 색상과 밝기를 독립적으로 분석할 수 있어 합성 여부를 더 정밀히 탐지할 수 있다.

본 연구는 HLS 채널을 도입함으로써 CNN의 합성 이미지 판별 성능을 강화하고, EXIF 메타데이터와 결합하여 이미지의 신뢰성을 높인다. 이는 RGB 중심의 기존 접근법 대비 차별화된 전처리 방식으로, 사기 방지 시스템의 정확도를 개선한다.

III. 설계 및 구현

본 연구의 설계 및 구현 과정은 합성 판별 인공지능 파트와 메타데이터 기술을 활용한 도용 판별로 구성된다. 순서는 인공지능 설계 및 구현, 메타데이터 기술로 진행한다.

3.1 인공지능 아키텍처 설계

본 연구의 이미지 합성 여부 판별 인공지능 설계 단계에서 아키텍처를 설계하기 위해 파이프-필터 패턴을 사용한다. 그림 1은 파이프-필터 패턴을 사용해서 인공지능의 아키텍처를 다이어그램으로 작성한 것이다.

전형적인 소프트웨어 아키텍처의 설계 방법 중 하나인 파이프-필터 패턴은 데이터 스트림 작업을 개별적인 요소로 필터 컴포넌트로 캡슐화하여 데이터를 송수신하여 처리한다. 본 연구의 다이어그램에서는 총 7단계로, 데이터 수집 및 생성, 데이터 전처리, 학습 데이터 생성, 모델 학습, 모델 생성 및 검증, 모델 추론, 모델 API화 단계로 설계하였다. 다음은 설계한 다이어그램의 단계에 따라 세부적인 내용을 설명하고자 한다.

3.2 데이터 수집

합성 이미지를 판별하는 CNN 모델을 학습시키기 위한 데이터로 합성 이미지와 비합성 이미지를 수집한다. 합성이미지 6,117장, 비합성 이미지 6,117장으로 비합성 이미지는 AiHub 사이트의 상품 이미지 데이터를 통해 수집하고 합성이미지는 Lorem Picsum 사이트의 무작위 이미지를 불러와 기존 비합성 이미지 6,117장을 모든 무작위 이미지를 통하여 Python의 Pillow 이미지 모듈을 통해 각각의 이미지의 크기에 맞는 이미지를 무작위로 불러와 상품 이미지의 객체만 모두 무작위 이미지 배경과 합성을 하여 합성 이미지 데이터셋을 구축했다.

그림 2와 그림 3은 수집 및 생성된 이미지 데이터들을 보여주고 있다.

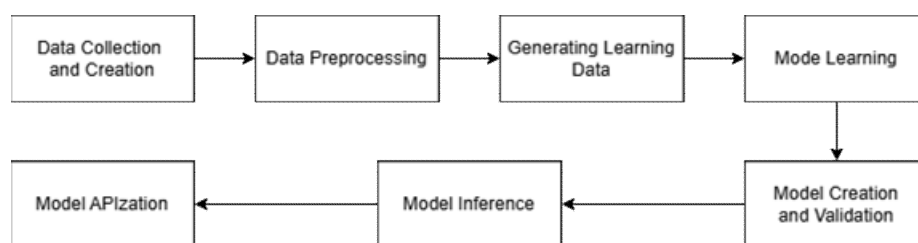


그림 1. 파이프-필터 패턴 다이어그램
Fig. 1. Pipe-Filter pattern diagram



그림 2. 합성 이미지 데이터
Fig. 2. Fake image data

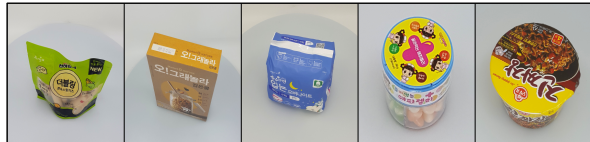


그림 3. 비합성 이미지 데이터
Fig. 3. Non-fake image data

각 합성 이미지와 비합성 이미지 데이터셋은 대부분 사물이 사진의 가운데에 위치하는 이미지들이며 파일 형식은 PNG 파일이다. 데이터셋의 대부분의 이미지는 밝은 환경에서 수집 및 생성되었기 때문에 HLS 채널을 활용하여 데이터 전처리 단계를 수행한다면 CNN 모델의 특징 추출 단계에서 효과가 있다고 예상된다. 이러한 방식으로 구성된 데이터셋을 사용하여 본 연구의 핵심적인 부분 중 하나인 합성 이미지 판별 모델을 구현한다.

3.3 데이터 전처리

합성 이미지와 비합성 이미지는 모델에 따라 이미지의 크기를 리사이징하는 단계와 HLS 채널로 변환하는 단계, 이미지 라벨링 단계로 구성되어 데이터 전처리를 진행하여 학습데이터셋과 검증데이터셋을 생성한다. 학습데이터셋이 인공지능 모델에 가장 적합한 크기의 이미지를 입력받아야 하므로 이미지의 크기를 조정하고 각 입력 이미지 채널을 특정 평균과 표준편차로 정규화하여 모델의 학습을 안정화한다. ImageNet 데이터셋으로 Pre-trained된 모델인 Resnet에 적합한 이미지를 형식을 입력하기 위해 Resnet의 Pre-trained된 이미지 형식을 도입하여 이미지를 처리한다. 이미지의 크기는 224x224로 리사이징을 진행하고 Pytorch로 인공지능을 학습하기 위해 텐서로 변환한다. 정규화 작업은 ImageNet 데이터셋에 대한 모든 이미지의 각 채널별 픽셀 값의

평균과 표준편차로 평균은 [0.485, 0.456, 0.406], 표준편차는 [0.229, 0.224, 0.225]로 처리하여 진행한다. HLS 채널로의 변환은 OpenCV를 사용하여 이미지를 받아 각 채널로 분리하는 작업을 진행한다. HLS 채널로 변환하면 각 채널은 그림 4에 나타나는 것과 같이 이미지에 대해 변환이 적용된다.



그림 4. HLS 채널 시각화
Fig. 4. HLS channel visualization

본 연구에서의 HLS 조정 값으로는 명암, 포화도의 값을 조정을 진행한다. 명암의 값은 1.2로 20 퍼센트의 증가를, 포화도면에서는 10 퍼센트의 감소를 설정한다. HLS 조정값은 GPT-4를 활용하여 합성 이미지 판별에 가장 적합한 HLS 채널값을 설정하였다.

3.4 인공지능 학습

본 연구의 인공지능의 학습 단계에서는 데이터 전처리 단계를 진행한 합성 이미지와 비합성 이미지를 학습 데이터셋과 검증 데이터셋으로 분리하고 CNN 모델이 합성 여부를 판별할 수 있게 만드는 과정이다. 데이터셋의 빠르고 효과적인 처리를 위해 데이터셋의 인스턴스를 파일로 만들 수 있는 Pickle 라이브러리를 활용하였다. 훈련 데이터셋의 비율은 70퍼센트 나머지 검증 및 테스트 데이터셋의 비율은 30퍼센트, 배치 사이즈는 32를 사용한다. 합성 여부 판별을 위한 CNN 모델로는 Resnet-50, 자체 CNN 모델을 사용하여 구현하였다.

3.4.1 Resnet 모델 구현 및 학습

첫 번째 모델로 Pre-trained된 Resnet-50을 사용한다. 모델의 정답 라벨이 0,1인 합성, 비합성 두 개이므로 모델의 마지막 레이어를 이진 분류에 맞게 조정하고 클래스 수를 1로 설정하여 본 연구에 맞게 데이터 전처리를 완료한 학습데이터로 훈련을 진행한다.

Epoch는 100, 옵티마이저는 Adam, Loss는 이진 분류이므로 BCELoss로 설정한다. ImageNet으로 미리 학습된 모델로 파라미터의 가중치를 가져오기 때문에 Early-Stopping 함수를 사용하여 3번의 성능 향상인 검증 Loss값의 감소가 없는 경우, 훈련을 중지하게 설정하였다. 학습된 Resnet-50 모델로부터 Sigmoid 함수를 사용하여 추론 결과로 합성 여부를 출력한다.

3.4.2 자체 CNN 모델 구현 및 학습

두 번째 CNN 모델은 컨볼루션 레이어, 풀링 레이어, 완전 연결 레이어, 배치 정규화, 드롭 아웃으로 구성하였다. 각 컨볼루션 레이어는 Relu 활성화 함수와 배치 정규화를 사용하고 풀링 레이어로 연산량을 감소시킨다. 완전 연결 레이어로 마지막 출력을 얻기 위해 Sigmoid 함수를 사용하여 이진 분류를 수행한다. 드롭-아웃을 적용하여 무작위로 일부 뉴런을 비활성화하여 과적합을 방지한다. 훈련 설정으로는 Epoch는 100, 옵티마이저는 Adam, Loss는 이진 분류이므로 BCELoss로 설정하고 훈련 중지 설정으로 Early-Stopping 함수를 적용한다. 다음 그림 5는 모델에 대한 요약 정보를 보여준다.

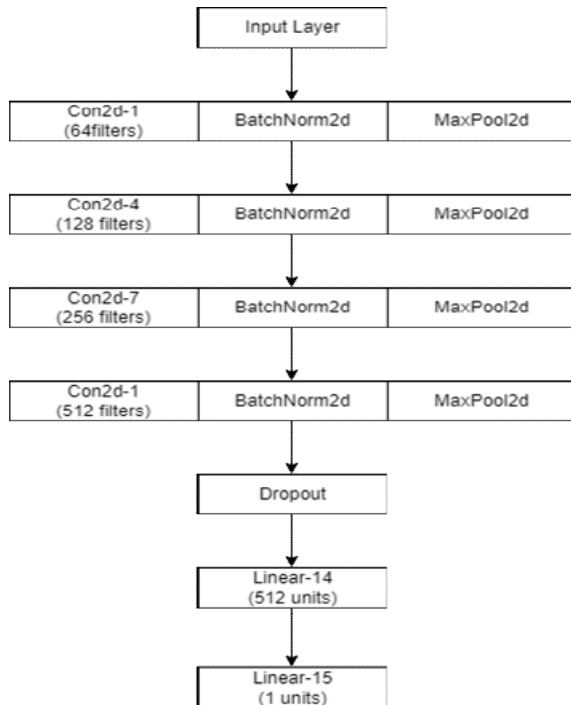


그림 5. CNN 모델 요약

Fig. 5. CNN model summary

3.4.3 ViT 모델 구현 및 학습

세 번째 모델은 Hugging Face에서 제공하는 오픈 소스 모델인 ViTForImageClassification을 활용하여 본 연구의 데이터에 대해 재학습을 진행한다. 훈련 설정은 Epoch는 100, 옵티마이저는 Adam, Loss는 CrossEntropyLoss를 사용하고 이전 모델 훈련과 같이 Early-Stopping 함수를 사용해서 효율적인 모델 훈련을 진행한다. 다음 그림 6는 ViT 모델을 호출하고 이진 분류에 맞게 출력 레이어를 수정한 코드이다.

Algorithm 1. Frame saving function

```

def initialize_model():
    model_id = "google/vit-base-patch16-224"
    model = ViTForImageClassification.from_pretrained(model_id)
    # Set the model classifier to a new classifier.
    model.classifier = torch.nn.Linear(model.config.hidden_size, 2)
    feature_extractor = ViTFeatureExtractor.from_pretrained(model_id)
    return model, feature_extractor

model, feature_extractor = initialize_model()
  
```

그림 6. ViT 모델 출력 레이어 수정 코드

Fig. 6. ViT model output layer modification code

3.5 검증데이터 정확도 확인 및 결과 분석

본 연구에서 Resnet-50, CNN 모델, ViT 모델 세 가지 모델의 훈련을 학습 데이터와 검증 데이터를 사용해서 진행하였다. Resnet-50 모델은 검증 정확도가 약 99퍼센트로 굉장히 높은 정확도를 보였으며 검증 손실값 또한 0.0022의 값으로 굉장히 낮은 값을 보였다. 본 연구에서 사용한 데이터셋으로 추가 학습을 진행한 Resnet-50 모델은 합성 이미지 판별에 좋은 결과를 보일 것으로 분석된다. 다음 그림 7은 Resnet-50의 훈련 결과를 나타낸다.

자체로 설계한 CNN 모델은 훈련을 진행한 다른 모델에 비해 정확도가 약 49퍼센트로 기대치보다 낮은 정확도를 보인다. 검증 손실값 또한 다른 모델에 비해 0.0137로 비교적 높은 손실값을 가진다. 다음 그림 8은 자체로 설계한 CNN 모델의 훈련 결과를 나타낸다.

```

Train Loss=0.0000, Train Acc=1.0000, Val Loss=0.0001, Val Acc=1.0000
loss decreased (0.000148 --> 0.000114). Saving model ...
poch 12: 100%|██████████| 268/268 [19:09<00:00, 4.29s/it]
: 100%|██████████| 58/58 [00:39<00:00, 1.47it/s]
Train Loss=0.0000, Train Acc=1.0000, Val Loss=0.0002, Val Acc=1.0000
ing counter: 1 out of 3
poch 13: 100%|██████████| 268/268 [17:25<00:00, 3.90s/it]
: 100%|██████████| 58/58 [00:38<00:00, 1.50it/s]
Train Loss=0.0025, Train Acc=0.9993, Val Loss=0.0131, Val Acc=0.9951
ing counter: 2 out of 3
poch 14: 100%|██████████| 268/268 [17:08<00:00, 3.84s/it]
: 100%|██████████| 58/58 [00:39<00:00, 1.46it/s]
Train Loss=0.0073, Train Acc=0.9978, Val Loss=0.0022, Val Acc=0.9995

```

그림 7. Resnet-50의 훈련 결과
Fig. 7. Resnet-50 result

```

: 100%|██████████| 268/268 [00:43<00:00, 6.16it/s]
|██████████| 58/58 [00:11<00:00, 5.21it/s]
oss=0.0112, Train Acc=0.5012, Val Loss=0.0307, Val Acc=0.4937
unter: 1 out of 3
: 100%|██████████| 268/268 [00:45<00:00, 5.87it/s]
|██████████| 58/58 [00:07<00:00, 8.14it/s]
oss=0.0121, Train Acc=0.5012, Val Loss=0.0177, Val Acc=0.4937
unter: 2 out of 3
: 100%|██████████| 268/268 [00:41<00:00, 6.49it/s]
|██████████| 58/58 [00:07<00:00, 7.73it/s]
oss=0.0127, Train Acc=0.5012, Val Loss=0.0137, Val Acc=0.4937

```

그림 8. CNN 모델의 훈련 결과
Fig. 8. CNN model result

마지막 모델인 ViT 모델이다. 검증 정확도가 약 99퍼센트로 Resnet-50의 모델과 비슷한 검증 정확도를 보이며, 검증 손실값은 0.0554의 값을 보였다. 하지만 훈련을 진행할수록 검증 손실값이 증가하며 검증 정확도가 1에 수렴하여 과적합의 문제가 있는 것으로 판단된다. 다음 그림 9는 ViT 모델의 훈련 결과를 나타낸다.

```

ch 1: 0%|██████████| 0/268 [00:00<?, ?it/s]
ch 1: 100%|██████████| 268/268 [37:12<00:00, 8.33s/it]
100%|██████████| 58/58 [02:00<00:00, 2.07s/it]
in Loss=0.0867, Train Acc=0.9996, Val Loss=0.0554, Val Acc=0.9967
oss decreased (inf --> 0.055354). Saving model ...
ch 2: 100%|██████████| 268/268 [47:15<00:00, 10.58s/it]
100%|██████████| 58/58 [03:27<00:00, 3.59s/it]
in Loss=0.0842, Train Acc=1.0001, Val Loss=0.0596, Val Acc=1.0000
g counter: 1 out of 2
ch 3: 100%|██████████| 268/268 [1:01:26<00:00, 13.75s/it]
100%|██████████| 58/58 [03:28<00:00, 3.59s/it]
in Loss=0.1528, Train Acc=0.9994, Val Loss=0.1012, Val Acc=1.0000
g counter: 2 out of 2
ng initiated. Training stopped.

```

그림 9. ViT 모델의 훈련 결과
Fig. 9. ViT model result

세 가지 모델의 훈련을 진행하고 정확도가 가장 높고 검증 손실값이 낮은 Resnet-50을 통해 테스트를 진행하였다. 이전 학습 데이터셋을 구성할 때 테스트 데이터로 분리한 데이터셋을 사용하였고 실제 테스트 결과 시 모든 합성 여부를 판별하는데 문제가 없었지만 실제 인터넷 이미지에 대한 테스트를

진행할 때 대체로 판별 여부를 판단하지 못 하는 문제가 있다.

훈련, 테스트, 결과 분석을 통해 사전 학습된 모델은 높은 정확도와 낮은 검증 손실값을 보이지만 검증되지 않거나 모델 구조의 복잡도와 깊이에 따른 충분한 데이터셋의 수집, 생성, 정제가 되지 않는다면 과적합 문제와 실제 데이터를 제대로 판별하지 못하는 한계가 있다. 결과적으로, CNN 기반 모델만으로 합성 여부를 판별하는 작업은 정확하고 검증된 합성, 비합성 이미지 데이터셋을 수집, 생성, 정제하는 과정과 데이터셋에 맞는 적합한 모델을 선정하는 것이 중요하다는 점을 나타낸다.

3.6 메타데이터 분석

본 논문에서는 EXIF 태그 중 카메라 제조사, 날짜와 시간, 이미지 ID, 사진 크기, 너비, 높이 등의 방면으로 연구를 진행하여 EXIF 데이터를 기반으로 메타데이터가 조작된 여부를 판단하여 도용의 여부를 판단하고자 한다.

표 1. 메타데이터가 포함된 예시 테이블
Table 1. Example table with metadata

Image	
ImageWidth	4000
ImageLength	3000
Make	samsung
Model	SM-A336N
Orientation	Horizontal(normal)
XResolution	72/1
YResolution	72/1
ResolutionUnit	inches
Software	A336NKSU5DWL1
Date Time	2024-01-23 3:29:57 PM
YCbCrPositioning	Centered
Photo	
ExposureTime	1/60 sec
FNumber	f/1.8
ExposureProgram	Auto
ISOSpeedRatings	80
DateTimeOriginal	2024-01-23 3:29:57 PM
DateTimeDigitized	2024-01-23 3:29:57 PM
ShutterSpeed	1/1 s
Apreture	f/1.69
MeteringMode	Center weighted average
FocalLength	4.65mm
SceneCaptureType	Standard

스마트폰이나 카메라를 통하여 사진을 촬영하는 경우에 표 1의 메타데이터를 확인해보면 날짜와 시간 정보, 카메라 설정, 저작권, 카메라 제조사, 카메라 모델, 이미지 ID 등 여러 메타데이터가 포함된다.

본 논문에서는 각 스마트폰이나 카메라 기기마다 카메라의 해상도나 너비와 높이가 달라지는 부분을 집중적으로 연구를 진행하였다. 아이폰을 예로 들면 아이폰 2G/3G 사진 해상도는 2MP(Mega Pixel)이며 1660x1200 픽셀의 해상도를 가지고 아이폰 3GS 사진 해상도는 3.2MP이며 2048x1536 픽셀의 해상도를 가지고 있다. 현재 나오는 기종들의 해상도는 12MP로 나오고 있으며 4032x3024의 픽셀의 해상도를 가지고 출시하고 있다. 이를 통해, 아이폰과 갤럭시 기종은 각 고유의 해상도를 가지고 있다는 것을 알 수 있다. 그림 10의 이미지를 확인하면 현재 나오는 12MP를 가진 아이폰으로 사진을 촬영 후 웹사이트나 소셜 네트워크 등에 업로드 후 재차 다른 기기로 내려받으면 SNS는 1080x1440으로 변환되었고 네이버는 773x103, 800x1067의 크기로 변환된 것을 확인할 수 있다. 이에 대해 다른 기기들과 사진들로 확인한 결과 모두 같은 결과로 나왔다.

	사진 크기	수평, 수직 해상도 픽셀	이미지 ID	카메라제원 유무	용량
원본	3024 x 4032	72 72	2024-03-09 x	o	2.74MB
카카오프로필 PC 다운	3024 x 4032	72 72	2024-03-09 x	o	2.74MB
카카오프로필 PC 다운	1080 x 1440	96 96	x	x	576KB
카카오프로필 모바일 다운	3024 x 4032	확인 불가	2024-03-09 확인 불가	x	2.9MB
카카오프로필 모바일 다운	1080 x 1440	확인 불가	다운한 날짜로 지 x	x	600KB
구글(원본파일) PC 다운					
구글(다른 이름 저장) PC 다운					
구글(원본파일) 모바일 다운					
구글(다른 이름 저장) 모바일 다운					
네이버(원본파일) PC 다운	3024 x 4032	72 72	2024-03-09 x		2.74MB
네이버(다른 이름 저장) PC 다운	773 x 1031	96 96	x	x	
네이버(원본파일) 모바일 다운	3024 x 4032	x			
네이버(다른 이름 저장) 모바일 다운	800 x 1067	x	x	x	
네이버 M to M (원본)	3024 x 4032				
네이버 M to M (다른 이름 저장)	800 x 1067	x	다운한 날짜로 지 x	x	227KB
네이버 M to PC (원본)	3024 x 4032	72 72	2024-03-09 x	x	
네이버 M to PC (다른 이름 저장)	773 x 1031	96 96	x	x	177KB
정종대 (안드로이드)	사진 크기	수평, 수직 해상도 픽셀	이미지 ID	카메라제원 유무	용량
원본	4000 x 3000	150 150	2024-01-23 J48ESOAR1PM o		2.95MB
카카오프로필 PC 다운	4000 x 3000	150 150	2024-01-23 J48ESOAR1PM o		2.95MB
카카오프로필 PC 다운	1400 x 1050	1400 1050	x	x	280KB
카카오프로필 모바일 다운	4000 x 3000	확인 불가	2024-01-23 J48ESOAR1PM o		2.95MB
카카오프로필 모바일 다운	1400 x 1050	확인 불가	x	x	280KB
구글(원본파일) PC 다운	1400x1050				
구글(다른 이름 저장) PC 다운					
구글(원본파일) 모바일 다운					
구글(다른 이름 저장) 모바일 다운					
네이버(원본파일) PC 다운	4000 x 3000	150 150	2024-01-23 J48ESOAR1PM o		2.95MB
네이버(다른 이름 저장) PC 다운	773 x 580	96 96	x	x	61.1KB
네이버(원본파일) 모바일 다운	4000 x 3000	확인 불가	2024-01-23 J48ESOAR1PM o		2.95MB
네이버(다른 이름 저장) 모바일 다운	800 x 600	x	x	x	
네이버 M to M (원본)	4000 x 3000	확인 불가	2024-01-23 J48ESOAR1PM o		2.95MB
네이버 M to M (다른 이름 저장)	800 x 600	x	x	x	
네이버 M to PC (원본)	4000 x 3000	150 150	2024-01-23 J48ESOAR1PM o		2.95MB
네이버 M to PC (다른 이름 저장)	773 x 580	96 96	x	x	61.1KB

그림 10. 이미지 메타데이터 분석
Fig. 10. Image metadata analysis

이미지가 웹사이트나 소셜 네트워크 등에 업로드가 되면 이미지 처리 및 저장 공간, 응답 속도를 더

욱 효율적으로 사용하기 위해서 이미지의 크기를 줄이는 이미지 압축을 진행하여 해상도 및 크기, 용량을 감소시켜 웹사이트에 업로드 된다[15]. 이로 인해 표 2에서 볼 수 있듯이 웹사이트에 업로드된 이미지는 EXIF 태그가 거의 모두 사라진다는 부분을 본 연구에서는 집중하여 탐색했고 이미지의 고유 ID와 사진 크기, 너비, 높이에 좀 더 높은 비중을 두어 이미지의 도용 여부 판단 기준을 잡았다.

표 2. 메타데이터가 없는 예시 이미지

Table 2. Example image without metadata

Title	Content
FilePath	C:\User\admin\Desktop\...
FileName	channels4_profile.jpg
Name	channels4_profile
FileLocation	C:\User\admin\Desktop\...
FileSize	39409
DataCreation	2024-03-23 3:54:23 PM
DataLastAccess	2024-03-23 3:54:23 PM
DataLastWrite	2024-03-23 3:54:23 PM
PixelWidth	600
PixelHeight	600

위에서 언급했듯이 각 스마트폰 기종별 사진 해상도를 기준으로 도용의 여부를 판단하였다. 본 시스템에 이미지를 넣으면 메타데이터를 추출하여 이미지 ID의 유무를 판단하고 이미지 해상도를 기준으로 각 스마트폰 기종별 카메라 제조사와 해상도를 비교하여 이 기기가 찍은 사진이 그 기기의 우리가 세운 기준의 이미지 해상도가 동일 여부를 확인하여 도용의 여부를 판단한다.

3.7 EXIF 데이터 추출

그림 11에서는 Node.js 환경에서 소켓 통신을 사용하여 사용자가 이미지 파일을 선택하면 ‘sendImage’함수가 호출되는 프로세스를 나타내고 있다.

이 함수는 소켓을 통해 서버에 이미지 데이터를 비동기적으로 전송한다. 이미지 전송에 성공한 경우 그림 12에 따라, 서버로부터의 응답을 확인한 후 ‘extractImageMetadata’ 함수를 통해 이미지 파일로부터 EXIF 데이터의 추출을 진행한다.

Algorithm 2. Image transmission code

```

sendMessage() {
  if (this.newMessage.trim() !== "") {
    const message = {
      text: this.newMessage,
      timestamp: Date.now(),
      sender: this.userName,
      type: "text",
      file: this.selectedFile ?
      URL.createObjectURL(this.selectedFile) : null,
    };
    socket.emit("sendmessage", message, (res) => {
      console.log("Message sent!", message);
    });

    this.channelMessages[this.channel.id].push(message);
    this.scrollToBottom();
    this.newMessage = "";
    this.selectedFile = null;
  } else {
    console.error("Text send failed:", res.error);
  }
}

```

그림 11. 이미지 전송 코드
Fig. 11. Image transfer code

Algorithm 3. Image metadata extraction code

```

extractImageMetadata(file) {
  return new Promise((resolve, reject) => {
    // Extract EXIF data from the image file
    EXIF.getData(file, function() {
      // If extracted data is available, call resolve
      if (this.exifdata) {
        resolve(this.exifdata);
      } else {
        reject(new Error('Failed to extract image metadata.'));
      }
    });
  });
}

```

그림 12. 메타데이터 추출 코드
Fig. 12. Metadata extraction code

EXIF 데이터 추출 기능 구현을 위해 JavaScript에서 사용되는 'EXIF.js' 라이브러리를 활용하였다. EXIF.js 라이브러리는 웹 환경에서 이미지 파일의 메타데이터를 읽을 수 있도록 지원하며, 특히 'EXIF.getData' 함수를 사용하여 이미지 파일로부터 EXIF 데이터를 비동기적으로 추출한다. 이 함수는

이미지 파일 객체를 입력으로 받고, 해당 이미지 파일에 포함된 EXIF 정보를 'this.exifdata'를 통해 접근할 수 있다. 데이터 추출 로직은 콜백 함수 내에서 처리되며, 데이터가 성공적으로 추출되면 'Promise' 객체의 'resolve' 메소드를 호출하여 추출된 데이터를 반환한다. 만약 EXIF 데이터가 없거나 추출에 실패한 경우, 'reject' 메소드를 호출하여 오류 메시지와 함께 실패를 반환하도록 설계하였다.

3.8 전체 시스템 구현 결과

메타데이터 분석 기술과 합성 판별 인공지능을 활용하여 C2C 기반 거래에서 이미지를 이용한 사기를 방지하는 시스템을 구현하였다. 시스템의 주요 기능은 이미지 파일 업로드, 인공지능 활용 합성 판별, EXIF 데이터 추출, 이미지 파일 메타데이터 분석이다. 사용자가 이미지 파일을 업로드하면, Fastapi를 통해 연결된 인공지능이 합성 판별 작업을 수행하고 EXIF.js 라이브러리를 통해 데이터를 추출하고 비동기적으로 처리하여 결과를 출력한다. 다음 그림 13은 시스템의 실행결과이다.

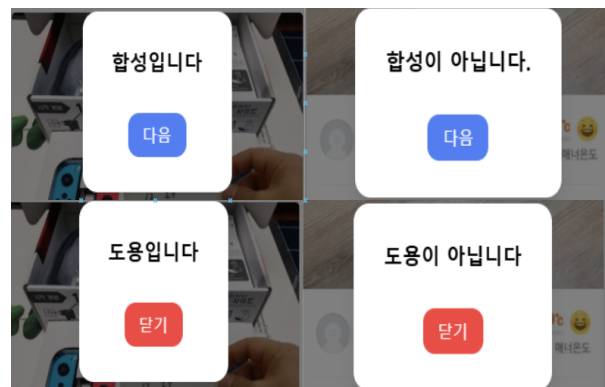


그림 13. 실행 결과
Fig. 13. Run result

IV. 결론 및 향후 과제

본 논문에서는 C2C 온라인 거래에서 발생할 수 있는 사기를 방지하기 위해 이미지 합성 판별과 메타데이터 분석을 통합한 시스템을 설계하고 구현하였다.

CNN 기반의 이미지 합성 판별 기술과 EXIF 메타데이터를 활용한 기술을 사용하여 사용자가 신뢰할 수 있는 거래 시스템을 제공하는 것을 목표로 진행하였다. Resnet-50, 자체 설계한 CNN 모델, ViT 모델을 활용하여 합성 이미지와 비합성 이미지를 판별하였고 Resnet-50과 ViT 모델은 높은 성능을 나타내었다. 또한, 메타데이터를 활용하여 이미지의 도용 여부를 판단할 수 있는 기능을 성공적으로 구현하였다. 하지만 자체 설계한 CNN 모델은 기대했던 바와 다르게 낮은 정확도를 보이고 모든 모델이 실제 이미지에 대해서는 일부 판별이 실패하는 케이스가 있었다. 이는 과적합 문제와 확실하게 검증되지 않은 데이터셋의 문제로 인한 것으로 추정된다. 이를 해결하기 위해, 향후 다양하고 복잡한 환경에서 검증된 데이터셋을 확보하여 모델의 일반화능력을 향상하고 모델의 구조를 개선하는 작업을 진행하고자 한다. 결과적으로, 본 연구는 C2C 거래 환경에서 신뢰성을 높이기 위한 유의미한 시도로 기술의 개선과 데이터의 확대를 통해 효과적인 C2C 거래 환경에서의 사기 방지 시스템을 구축할 수 있을 것으로 기대된다.

References

- [1] E.-A. Park and E.-C. Kim, "Why do they buy used products? : Comparison between adolescents and adults", *Social Science Research Review*, Vol. 37, No. 1, pp. 5-31, Feb. 2021.
- [2] <https://platum.kr/192905> [accessed: Apr. 17, 2024]
- [3] J. H. Hyeon, D. Y. Im, C. Y. Lee, J. U. Lee, H. H. Lee, and W. Y. Choi, "Online direct transaction environment for fraud damage prevention suggestion: focusing on fraud red flags", *The Journal of Police Science*, Vol. 21, No. 1, pp. 249-272, Mar. 2021.
- [4] J. H. Lee, "Method for providing used goods trade service using fraud detection and appraisal based on blockchain with safe transaction", Patent KR102134792B1, Jul. 2020.
- [5] A. Piva, "An Overview on Image Forensics", *International Scholarly Research Notice*, Vol. 2013, pp. 1-22, Jan. 2013. <https://doi.org/10.1155/2013/496701>.
- [6] JEITA, "EXIF 2.2 Specification", Japan Electronics and Information Technology Industries Association, 2002.
- [7] E. Kee, M. K. Johnson, and H. Farid, "Digital image authentication from JPEG headers", *IEEE Transactions on Information Forensics and Security*, Vol. 6, No. 3, pp. 1066-1075, Sep. 2011. <https://doi.org/10.1109/TIFS.2011.2114354>.
- [8] T. Gloe, "Forensic analysis of ordered data structures on the example of JPEG files", 2012 IEEE International Workshop on Information Forensics and Security, Costa Adeje, Spain, pp. 139-144, Dec. 2012. <https://doi.org/10.1109/WIFS.2012.6412639>.
- [9] A. Castiglione, G. Cattaneo, and A. D. Santis, "A forensic analysis of images on Online Social Networks", 2011 Third International Conference on Intelligent Networking and Collaboration Systems, Fukuoka, Japan, pp. 679-684, Nov.-Dec. 2011. <https://doi.org/10.1109/INCoS.2011.17>.
- [10] P. Mullan, C. Riess, and F. Freiling, "Forensic source identification using JPEG image headers : The case of smartphones", *Digital Investigation*, Vol. 28, pp. S68-S76, Apr. 2019. <https://doi.org/10.1016/j.diin.2019.01.016>.
- [11] O. Giudice, A. Paratore, M. Moltisanti, and S. Battiato, "A Classification Engine for Image Ballistics of Social Data", *Image Analysis and Processing - ICIAP 2017*, Catania, Italy, Vol. 10485, pp. 625-636, Sep. 2017. https://doi.org/10.1007/978-3-319-68548-9_57.
- [12] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition", *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, pp. 770-778, Jun. 2016. <https://doi.org/10.1109/CVPR.2016.90>.

- [13] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition", arXiv preprint arXiv:1409.1556, Sep. 2014. <https://doi.org/10.48550/arXiv.1409.1556>.
- [14] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks", Advances in neural information processing systems 25, Lake Tahoe, Nevada, USA, pp. 1097-1105, Dec. 2012.
- [15] K. H. Sim, H. A. Ko, J. K. Kim, M. H. Zhao, and Y. H. Lim, "Adaptive Image Converting for More Efficient Mobile Web Service", Journal of Internet Computing and Services, Vol. 12, No. 2, pp. 37-46, Apr. 2011.

저자소개

권혁찬 (Hyukchan Kwon)



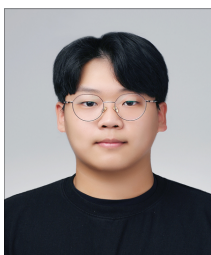
2025년 2월 : 국립안동대학교
멀티미디어공학과(공학사)
관심분야 : 기계학습, 딥러닝,
클라우드 컴퓨팅

남주영 (Juyoung Nam)



2025년 2월 : 국립안동대학교
멀티미디어공학과(공학사)
관심분야 : 웹 디자인, 웹
프론트엔드

정용태 (Yongtae Jung)



2025년 2월 : 국립안동대학교
멀티미디어공학과(재학중)
관심분야 : 웹 백엔드, 클라우드
컴퓨팅

최태용 (Taeyong Choi)



2025년 2월 : 국립안동대학교
멀티미디어공학과(공학사)
관심분야 : 웹 백엔드, 웹
프론트엔드

정기현 (Kihyun Jung)



2007년 8월 : 경북대학교
컴퓨터공학과(공학박사)
2022년 9월 ~ 현재 :
국립경국대학교(국립안동대학교)
소프트웨어융합학과 교수
관심분야 : 데이터보호,
스태가노그래피, 블록체인,
디지털위변조탐지기술, 보안프로그래밍