

Adversarial Learning-Based Image Correction Methodology for Deep Learning Analysis of Heterogeneous Images

Junwoo Kim[†] · Namgyu Kim^{††}

ABSTRACT

The advent of the big data era has enabled the rapid development of deep learning that learns rules by itself from data. In particular, the performance of CNN algorithms has reached the level of self-adjusting the source data itself. However, the existing image processing method only deals with the image data itself, and does not sufficiently consider the heterogeneous environment in which the image is generated. Images generated in a heterogeneous environment may have the same information, but their features may be expressed differently depending on the photographing environment. This means that not only the different environmental information of each image but also the same information are represented by different features, which may degrade the performance of the image analysis model. Therefore, in this paper, we propose a method to improve the performance of the image color constancy model based on Adversarial Learning that uses image data generated in a heterogeneous environment simultaneously. Specifically, the proposed methodology operates with the interaction of the 'Domain Discriminator' that predicts the environment in which the image was taken and the 'Illumination Estimator' that predicts the lighting value. As a result of conducting an experiment on 7,022 images taken in heterogeneous environments to evaluate the performance of the proposed methodology, the proposed methodology showed superior performance in terms of Angular Error compared to the existing methods.

Keywords : Adversarial Learning, Color Constancy, Heterogeneous Images, Illumination Estimation, Image Correction

이질적 이미지의 딥러닝 분석을 위한 적대적 학습기반 이미지 보정 방법론

김 준 우[†] · 김 남 규^{††}

요 약

빅데이터 시대의 도래는 데이터에서 스스로 규칙을 배우는 딥러닝의 비약적인 발전을 가능하게 하였으며, 특히 CNN 알고리즘이 거둔 성과는 모델의 구조를 넘어 소스 데이터 자체를 조정하는 수준에 이르렀다. 하지만 기존의 이미지 처리 방법은 이미지 데이터 자체를 다룰 뿐, 해당 이미지가 생성된 이질적 환경을 충분히 고려하지 않았다. 이질적 환경에서 촬영된 이미지는 동일한 정보임에도 촬영 환경에 따라 각 이미지의 특징(Feature)이 상이하게 표현될 수 있다. 이는 각 이미지가 갖는 상이한 환경 정보뿐 아니라 이미지 고유의 정보조차 서로 상이한 특징으로 표현되며, 이로 인해 이들 이미지 정보는 서로 잡음(Noise)으로 작용해 모델의 분석 성능을 저해할 수 있음을 의미한다. 따라서 본 논문은 이질적 환경에서 생성된 이미지 데이터들을 동시에 사용하는 엔드-투-엔드(End-To-End) 구조의 적대적 학습(Adversarial Learning) 기반의 이미지 색 항상성 모델 성능 향상 방안을 제안한다. 구체적으로 제안 방법론은 이미지가 촬영된 환경인 도메인을 예측하는 '도메인 분류기'와 조명 값을 예측하는 '조명 예측기'의 상호 작용으로 동작하며, 도메인 분류의 성능을 떨어뜨리는 방향의 학습을 통해 도메인 특성을 제거한다. 제안 방법론의 성능을 평가하기 위해 이질적 환경에서 촬영된 이미지 데이터 셋 7,022장에 대한 색 항상성 실험을 수행한 결과, 제안 방법론이 기존 방법론에 비해 Angular Error 측면에서 우수한 성능을 나타냄을 확인하였다.

키워드 : 적대적 학습, 색 항상성, 이질적 이미지, 조명 추정, 이미지 보정

1. 서 론

사람의 신경망과 유사한 구조를 지닌 딥러닝(Deep Learning)

은 데이터의 집합에서 숨은 의미를 찾는 기계 학습 알고리즘이다. 따라서 모델의 성능 향상을 위해서는 데이터 자체의 양과 품질이 매우 중요한데, 최근 다양한 빅데이터 관련 응용 및 인프라를 통해 방대한 양의 데이터가 생성되면서 딥러닝의 발전이 가속화하고 있다. 특히 딥러닝은 정형 데이터는 물론 이미지와 텍스트로 대표되는 비정형 데이터까지 그 분석 범위를 넓혀 나가고 있으며, 이미지 분야에서 CNN (Convolutional Neural Network) 알고리즘의 발전은 이미지 분류, 객체 탐지, 미술치료[1], 영화 평점 예측[2] 등 학계와 업계를 가리지 않고 매우 다양한 응용 분야에서 괄목할 만한 성과를 거두고 있다.

* 이 논문은 2021년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임(NRF-2021S1A5A2A01061459).

** 이 논문은 과학기술정보통신부와 정보통신산업진흥원의 '고성능 컴퓨팅 지원 사업'의 지원을 받아 수행하였다.

† 비 회 원 : 국민대학교 비즈니스IT 석사과정

†† 정 회 원 : 국민대학교 비즈니스IT전문대학원 교수

Manuscript Received : September 28, 2021

Accepted : October 11, 2021

* Corresponding Author : Namgyu Kim(ngkim@kookmin.ac.kr)



Fig. 1. Similar Image Information Photographed in a Heterogeneous Environment

하지만 이미지 데이터는 언어라는 일관된 약속에 따라 생성되는 텍스트 데이터와 달리 촬영 환경에 따라 데이터의 형태와 구조가 매우 이질적으로 인식되며, 이로 인해 동일한 정보임에도 촬영 환경에 따라 각 이미지의 특징(Feature)이 상이하게 표현될 수 있다. 이는 각 이미지가 갖는 상이한 환경 정보뿐 아니라 이미지 고유의 정보조차 서로 상이한 특징으로 표현됨을 의미하며, 이로 인해 이들 이미지 정보는 서로 잡음(Noise)으로 작용해 결과적으로 모델의 분석 성능을 저해하는 요소로 작용할 수 있다.

이러한 현상은 도로의 흰색 차선이라는 동일한 대상이 촬영 환경에 따라 각각 다르게 표현된 Fig. 1에서 확인할 수 있다. Fig. 1에서 왼쪽 이미지는 비교적 흐린 날에 촬영된 이미지로, 흰색 차선을 다소 어둡게 나타내고 있다. 반면 가운데 이미지는 흰색 차선을 매우 밝으면서 다소 붉은 빛을 띤 형태로 표현하고 있고, 오른쪽 이미지는 가로등 조명으로 인해 흰색 차선을 다소 노란 빛을 띤 형태로 표현한다. 본 예는 날씨, 조명, 촬영 도구 등 촬영 환경의 이질적 특성으로 인해 도로의 흰색 차선이 일관된 색상과 윤곽으로 표현되지 않는 현상을 보이며, 이러한 현상은 결과적으로 자율주행의 차선 인식 성능에 악영향을 미칠 수 있다.

이러한 문제를 해결하기 위해 학습에 사용되는 데이터 자체의 일관성을 향상시키기 위한 방법, 즉, 이질적 환경에서 촬영된 이미지를 동일 환경에서 촬영된 것처럼 변환하는 방법이 연구되어 왔다. 대표적으로 다양한 빛 또는 조명 하에서 서로 다르게 나타나는 물체의 색상을 일관되게 표현하는 방식으로 촬영 환경의 이질성을 제거하는 색 항상성(Color Constancy)은 컴퓨터 비전 애플리케이션의 필수 처리 과정으로, 사람의 시각이 인지하는 것과 동일한 이미지를 생성하기 위해 널리 사용된다. 색 항상성은 크게 통계기반과 학습기반으로 나뉘며, 통계기반 기법은 주로 'Intel'사의 'OpenCV' 라이브러리와 'MathWorks'사의 'MATLAB'을 활용하여 이미지 전처리 단계처럼 수행되고 있다. 통계기반 기법에 비해 상대적으로 최근에 연구되기 시작한 학습기반 기법은 CNN을 활용하는 연구들이 주를 이루고 있으며, 간단한 네트워크를 통해 기존 기법 대비 우수한 성능을 나타낸다는 장점이 있다[4].

그러나 학습기반 기법 또한 일반화 측면의 한계, 즉 특정 환경에서 촬영된 이미지에 대한 학습을 통해 생성한 모델을

다른 환경에서 촬영된 이미지의 분석에 사용하는 경우 성능이 낮게 나타난다는 한계를 갖고 있다. 이를 극복하기 위해서는 이미지의 촬영 환경이 분석 모델에 주는 영향을 최소화할 수 있는 강인한(Robust) 학습 방법론이 마련되어야 한다. 즉, 이미지 데이터 집단에 대해 각 집단이 소유한 도메인(Domain)의 특징은 배제하고, 이미지에 포함된 객체가 도메인과 무관하게 갖는 고유의(Intrinsic) 특징만을 추출하여 학습이 이루어져야 한다. 이와 관련하여 기술적으로 두 가지 도전과제가 있는데, 첫 번째는 딥러닝을 통해 얻은 고차원 특징에서 도메인 특징과 고유의 특징을 구분해 내기가 어렵다는 것이다. 두 번째는 학습의 시간성으로, 모델의 학습이 진행됨에 따라 특성이 계속해서 갱신되고 그 과정에서 달라지는 특성과 유지되는 특성을 일관되게 포착하기가 어렵다는 것이다.

이러한 도전과제를 해결하기 위해 본 연구에서는 이질적 환경에서 생성된 이미지 데이터들을 동시에 사용하는 엔드-투-엔드(End-To-End) 구조의 적대적 학습(Adversarial Learning)[5] 기반의 이미지 색 항상성 모델 성능 향상 방안을 제시한다. 구체적으로 제안 모델은 이미지가 촬영된 환경, 즉 도메인을 예측하는 '도메인 분류기'와 조명 값을 예측하는 '조명 예측기'의 상호 작용으로 동작하며, 도메인 분류의 성능을 떨어뜨리는 방향의 학습을 통해 도메인 특성을 제거한다.

본 논문의 이후 구성은 다음과 같다. 2장에서는 딥러닝 기반의 색 항상성 및 적대적 학습에 대해 기존 연구를 정리하고, 3장에서는 본 연구에서 제안하는 이질적 이미지의 딥러닝 분석을 위한 이미지 보정 방법론을 소개한다. 4장에서는 제안 방법론의 성능 평가를 위한 실험 과정 결과를 제시하고, 5장에서는 본 연구의 기여와 한계를 요약한다.

2. 관련 연구

2.1 색 항상성

색 항상성은 다양한 조명 아래에서도 객체의 색상을 일관되게 인식하도록 만들어 주는 인간의 시각 보상 인지 능력을 뜻한다. 한편 카메라는 렌즈를 통해 들어오는 물체의 반사광을 물체의 색상으로 나타내기 때문에, 촬영 당시의 조명에 따라 표현되는 이미지의 색상을 그대로 받아들인다. 따라서 인간의 시각과 달리, 카메라는 조명의 영향으로 인해 목표 객체의 본

질적 특성을 제대로 인식하지 못하는 어려움이 있다. 이러한 어려움을 해결하고자 하는 연구 분야를 색 항상성 또는 화이트 밸런스(White Balance)라고 하며, 크게 통계기반과 학습기반으로 나뉜다. 또한 이들 과정은 조명을 예측한 뒤, 예측한 조명을 통해 원본 이미지를 보정하는 2단계의 과정으로 구성된다.

전통적 통계기반의 색 항상성 기술은 이미지 색상 분포에 대한 특정 통계적 가정을 통해 조명을 추정한다. Gray-World[6] 알고리즘은 한 이미지의 평균 색상이 무채색이라는 가정을 통해 RGB 채널의 각 평균을 계산하여 조명을 추정한다. White-Point 또는 Max-RGB[7]라 불리는 알고리즘은 RGB 채널의 가장 밝은 지점들이 조명의 완전 반사에 의해 생성된다고 가정하고 반사에 의한 색상을 해석하여 조명을 추정한다. Gray-Edge[8] 알고리즘은 이미지 테두리의 평균이 무채색이라고 가정하고, RGB 채널의 각 테두리 색의 평균을 계산해 조명을 추정한다. 비교적 최근에 등장한 Grey-Pixel[9] 알고리즘은 이미지로부터 반사율이 무채색과 가장 유사한 픽셀들의 집합을 추출하고, 해당 픽셀들의 평균 반사율을 조명으로 추정한다. 통계기반 기법은 관측을 통한 비교적 단순한 가정으로 인해, 단 한 장의 데이터만으로도 적용할 수 있다는 장점이 있다. 하지만 통계기반 방법은 색 분포가 매우 다양하거나 무채색이 없는 이미지 데이터가 입력되는 등 해당 가정이 성립되지 않는 경우에는 적용이 어렵다는 한계를 갖는다. 따라서 최근에는 이러한 한계를 극복하고자 데이터에서 스스로 규칙을 찾는 학습기반 기법이 선호되고 있다.

학습기반 기법은 데이터에서 학습한 모델을 통해 조명을 추정하는 방식으로, 크게 Gamut Mapping[10]과 회귀 추론으로 나뉜다. Gamut Mapping은 Gamut이라 불리는 제한된 색 공간을 만든 뒤, 새로운 입력 이미지의 색 공간에 기존에 만들어진 공간에 투영하는 방식이다. 이러한 방식은 기존의 색 공간이 새로 들어온 이미지의 색 공간과 유사한 특성을 갖는 경우 효과적으로 작동하지만, 특성이 상이한 경우에는 성능이 낮아진다는 단점이 있다. 회귀 추론 방식은 이미지의 특징을 직접 학습하여 조명의 벡터(Vector) 값을 추정하는 방법으로, 서포트 벡터 머신(Support Vector Machine)[11], 의사결정나무(Decision Tree)[12] 등의 머신러닝(Machine Learning) 모델들을 주로 사용한다.

최근에는 학습 데이터의 전체적인 특성을 활용해 최대한의 효율을 내기 위한 시도로, 딥러닝을 사용한 회귀 추론 방식이 활발하게 시도되고 있다. 색 항상성에 CNN 알고리즘을 적용한 최초의 연구[13]는 이미지 전체에서 통계적 속성을 추출한 이전 연구들과 달리 패치(Patch) 단위 학습을 통해 이미지의 지역적 변화를 감지하는 방식으로 우수한 성능을 나타냈다. 이외에도 CNN 구조를 개선해 성능을 향상시키는 방법으로 다양한 학습기반 방법들[14-17]이 연구되고 있다.

2.2 적대적 학습

생성적 적대 신경망(Generative Adversarial Network)[18]으로 잘 알려진 적대적 네트워크는 의미없는 노이즈로부터

실제와 유사한 이미지를 생성하기 위해 사용된다. 적대적 네트워크는 구체적으로 생성자(Generator)와 판별자(Discriminator)의 모듈로 구성되며, 생성자는 가짜 이미지를 최대한 진짜와 유사한 모습으로 생성하고 판별자는 가짜 이미지와 진짜 이미지를 최대한 정확하게 판별해 내는 최소최대화(MiniMax) 학습을 수행한다. 이를 통해 생성자는 의미 없는 노이즈 이미지로부터 실제와 유사한 이미지를 만들어 낼 수 있다. 이러한 이중 네트워크 구조는 이미지 생성뿐만 아니라 가짜뉴스 판별[19], 시계열 합성[20], 자연어 처리[21], 도메인 적용(Domain Adaptation)[5, 22] 등 다양한 응용 분야에서 널리 활용되고 있다.

다양한 응용 분야 중 도메인 적용은 전이학습(Transfer Learning)의 하위분야로, 충분한 양의 데이터를 지닌 소스 도메인을 활용하여 데이터가 비교적 적거나 정답 레이블이 없는 타겟 도메인의 문제를 효율적으로 해결한다. 즉, 소스 도메인 데이터의 분류 성능을 유지하면서 타겟 도메인 데이터에서도 분류의 성능을 높이는 것을 목표로 한다. 이 과정에서 학습 전 소스 도메인의 특징(Feature)과 타겟 도메인의 특징은 그 분포가 상이하게 나타나는데, 이러한 현상, 즉 도메인 변화(Domain Shift)[23] 현상을 극복하고 상이한 도메인의 특성 분포를 동일하게 만드는 학습이 도메인 적용의 핵심이다. 도메인 시프트를 해결하는 방식은 소스 도메인에서 학습된 모델에 타겟 도메인 데이터를 입력으로 넣어 추출되는 특징을 소스 도메인의 특징 분포와 동일하게 만드는 방식, 그리고 상이한 두 도메인 데이터를 학습에 동시에 사용하여 두 도메인의 특징 분포 중 공통된 것만 추출하는 방식이 있다. 본 논문은 후자의 방식을 적용해, 두 도메인이 공유하는 공통된 특징 분포를 추출하고자 한다.

3. 제안 방법론

3.1 방법론 개요

본 연구에서 제안하는 방법론의 전체 개요는 Fig. 2와 같다. 방법론은 크게 Phase 1의 데이터 가공 단계와 Phase 2의 학습 단계로 구성되며, 본 장에서는 방법론의 각 단계별로 구체적인 프로세스를 설명한다.

본 연구에서는 상이한 상황에서 촬영된 이미지 집합을 구성하기 위해, 서로 다른 두 기종의 카메라를 통해 촬영한 이미지 셋을 각각 'Domain 1'과 'Domain 2'로 명명하여 학습에 사용한다. 우선 촬영된 이미지들을 패치 단위로 자르고 ①, 여러 패치 중 필요한 부분만을 선택하여 학습용 데이터 셋을 만든다 ②[24]. 이와 동시에 각 이미지의 RGB 레이블과 소스를 식별한 뒤 ③, 각 패치에 대해 RGB 레이블과 도메인 값을 태깅(Tagging)한다 ④. Phase 2의 첫 단계에서는 이렇게 태깅된 데이터 셋으로부터 CNN을 통해 특징을 추출하는 작업이 이루어진다 ⑤. 마지막으로 적대적 학습 단계에서는 이렇게 추출된 특징을 사용하여 조명 예측(A)과 도메인 분류(B)의 작업을 동시에 수행한다 ⑥. 이러한 전체 과정을 통해 일반화 성능이 높은 색 항상성 모델을 구축할 수 있다.

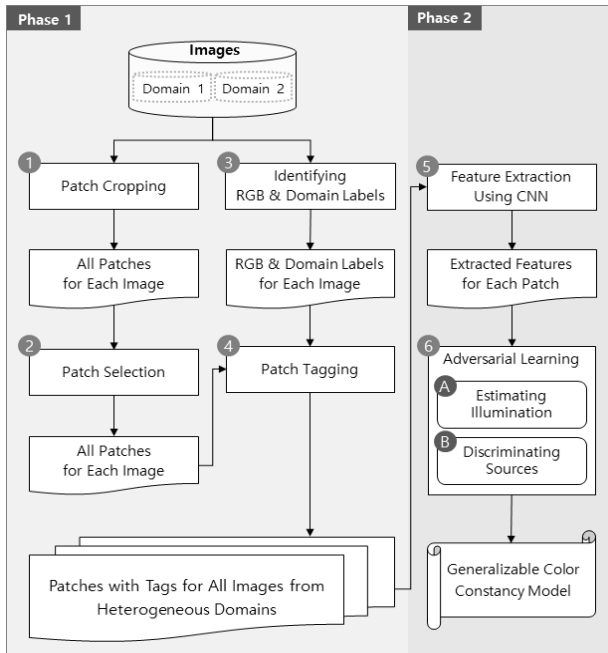


Fig. 2. Overview of the Proposed Methodology

3.2 패치 생성 및 레이블 태깅

본 절에서는 이미지를 여러 패치로 구분하고 각 패치에 대해 레이블을 태깅하여 학습 데이터를 준비하는 Phase 1의 과정을 소개한다. Phase 1은 색 항상성 기술에 적대적 학습을 적용하기 위한 데이터 가공 단계로, Fig. 3은 전체 과정을 하나의 예를 통해 보이고 있다. Fig. 3의 좌측 상단은 이미지 원본을 보이며, 태그인 'Domain 1'은 해당 이미지의 출처, 즉 해당 이미지를 촬영한 카메라가 무엇인지를 나타낸다. 또 다른 태그인 'RGB: [243, 228, 86]'은 해당 이미지가 어떤 상황에서 촬영되었는지, 즉 촬영에 사용된 조명의 R, G, B 값을 나타낸다.

'Image 1'은 (4*4)의 크기를 갖는 이미지로, 16개의 픽셀(Pixel)로 구성되어 있다. 각 픽셀은 빨강(R), 초록(G), 파랑(B) 각각에 대한 값을 가지므로, Image 1에 대한 정보는 Fig. 3의 좌측 하단과 같이 (4*4*3)의 배열로 나타낼 수 있다. 다음으로 R, G, B 각각에 대해 인접한 픽셀을 그룹화하여 패치를 생성하는데, Fig. 3의 예에서는 (2*2)의 크기, 즉 4개의 픽셀을 묶어서 하나의 패치를 생성하게 된다. 그 결과로 원본 이미지의 정보는 Fig. 3의 우측 하단과 같이 (2*2*3)의 배열로 나타나며, 이때 각 셀에 표시된 숫자는 해당 셀에 포함되는 픽셀의 값을 더한 값이다. 예를 들어 Fig. 3 우측 하단 그림의 첫 번째 셀의 18은 좌측 하단 그림의 1, 2, 13, 2를 합하여 산출된 값이다. Fig. 3 우측 하단 그림은 4개 패치의 R 값을 보여주고 있으며, 각 패치는 R 값 외에 G와 B의 값도 갖게 된다.

다음 과정은 각 패치의 밝기를 계산하는 과정으로, 그림에서는 생략되어 있으나 각 패치의 R, G, B 값을 더하여 해당 패치의 총 밝기를 산출한다. 예를 들어 Fig. 3 우측 하단에서

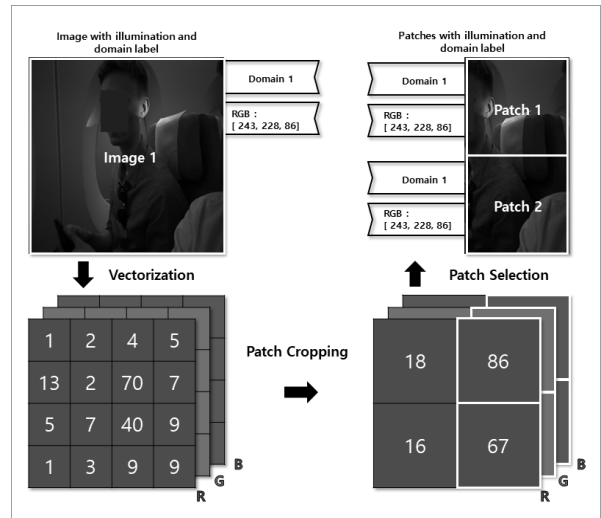


Fig. 3. Patch Creation and Label Tagging

첫 번째 패치의 값이 R=18, G=20, B=15라면, 해당 패치의 밝기는 RGB 채널의 픽셀 강도 값의 총합인 53으로 계산된다. 통계적 알고리즘을 사용한 기존 연구[25]에서 가장 밝은 픽셀 값을 갖는 이미지가 조명 추정에 유용함이 증명되었으므로, 본 연구에서도 조명 예측에 가장 영향력이 큰 데이터를 식별하기 위해 가장 밝은 값을 갖는 패치를 선택하여 학습 데이터로 사용한다. 이러한 기준으로 RGB 채널의 픽셀 강도 값의 총합이 가장 큰 두 개의 패치를 선택한 결과가 Fig. 3의 우측 상단에 표시되어 있다.

한편 이렇게 선택된 'Patch 1'과 'Patch 2'는 모두 'Image 1'로부터 생성된 것이므로, 이들 두 패치는 모두 'Image 1'의 레이블을 상속받아 'RGB: [243, 228, 86]'와 'Domain 1'의 레이블 값을 태그로 갖게 되며, 이는 두 개의 패치 모두 'Domain 1'의 카메라를 사용하여 촬영되었으며 촬영 당시 조명 값이 'RGB: [243, 228, 86]'라는 정보를 나타낸다. 이들 두 패치와 각각의 레이블 값은 이후 과정에서 적대적 학습을 위한 입력 데이터로 사용된다.

3.3 특징 추출 및 적대적 학습

본 절에서는 Phase 1을 통해 구축된 패치 단위 이미지에 적대적 학습을 적용하여 일반화 가능한 색 항상성 모델을 구축하는 Phase 2의 과정을 소개한다. Phase 2는 CNN을 통한 특징 추출 및 적대적 학습이 진행되는 제안 방법론의 학습 단계로, Fig. 4는 해당 과정을 도식화해 표현한 예이다. Fig. 4의 좌측에 나타난 4장의 패치는 3.2절의 결과물로, 'Domain 1'과 'Domain 2'로부터 각각 2장씩 생성되었다.

Fig. 4의 중앙에 위치한 CNN은 이러한 패치들 각각을 입력으로 받아 하나의 합성곱층(Convolutional Layer)과 맥스 풀링층(Max Pooling Layer)을 거쳐 합성곱 연산을 진행한 뒤, 다시 플랫튼 층(Flatten Layer)을 통해 3,080차원의 고차원 벡터를 추출한다. 이렇게 추출된 특징 벡터는 우측 상단의 도메인 분류기(Domain Discriminator)와 우측 하단의

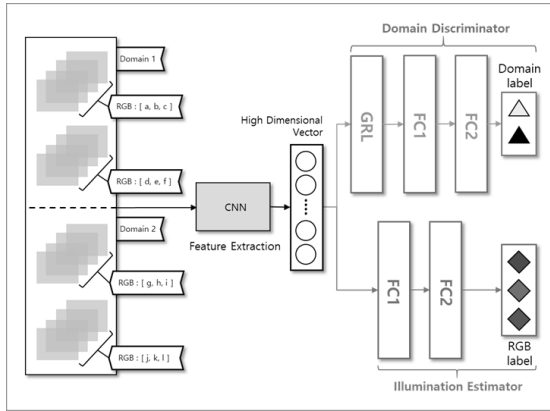


Fig. 4. Feature Extraction and Adversarial Learning

조명 예측기(Illumination Estimator)의 입력으로 사용된다.

조명 예측기는 완전연결계층(Fully Connected Layer) 2개를 쌓아서 구성되며, RGB 레이블에 해당하는 값을 예측하는 학습을 진행한다. 본 예에서는 각 패치들의 촬영 당시 조명 값으로 태깅된 'RGB: [a, b, c]', 'RGB: [d, e, f]', 'RGB: [g, h, i]', 그리고 'RGB: [j, k, l]'가 예측의 목표 값이다. 실제 학습에서는 개별 이미지를 32*32 크기의 패치로 구분하여 밝기 값이 가장 높은 100개의 패치만을 학습에 사용하며 [24], 이러한 반복 학습을 통해 CNN에서 추출된 고차원 벡터는 RGB 조명 값을 예측하도록 갱신된다.

도메인 분류기는 'GRL(Gradient Reversal Layer)[5]' 1개와 완전연결계층 2개로 구성되어 있으며, 패치의 도메인 레이블인 'Domain 1'과 'Domain 2'를 각각 예측하여 개별 패치들이 어떤 소스에서 생성된 것인지를 구분한다. 이때 도메인 분류기는 일반적인 딥러닝 모델과 달리 정답과 예측 사이의 손실을 늘리는 방향으로 학습을 진행한다. 즉, 학습을 통해 정답인 'Domain Label'을 잘 구분하지 못하도록 많은 파라미터들이 갱신되며, 결과적으로 이질적 환경에서 생성된 이미지 데이터들로부터 도메인 응집성이 높은 특성을 제거할 수 있다. 이러한 작업은 적대적 학습의 핵심 요소인 GRL을 통해 구현되며, 구체적으로 역전파(Back Propagation)시 이전 층에서 전달받은 기울기에 -1~0 사이의 음수 값을 곱하여 전달하는 방식으로 동작한다.

조명 예측기와 도메인 분류기가 동시에 상호 작용하며 적대적 학습이 이루어지기 때문에, 이를 통해 최종적으로 연계되는 모델은 입력 데이터의 도메인과 무관하게 조명의 RGB 값을 잘 예측할 수 있다. Fig. 5는 적대적 학습을 통한 제안 모델의 특징 갱신 과정 및 결과의 예를 나타낸다.

Fig. 5에서 학습 전 특징(Before feature)은 도메인별 특성이 초록색과 청색으로 명확히 구분되어 나타나는 반면, 적대적 학습이 진행될수록 이러한 특성은 열어져서 학습 후 특징(After feature)에서는 도메인별 특성이 제거된 것을 알 수 있다. 즉 학습 후 특징은 도메인의 특성을 배제한 이미지 고유의 특성만을 표현하며, 이를 통해 일반화 가능한 색 항상성 모델을 도출하게 된다.

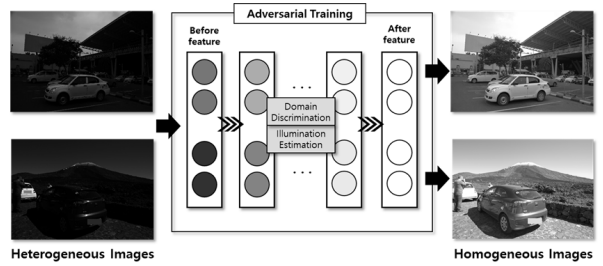


Fig. 5. Example of Feature Update in Adversarial Learning

4. 실험

4.1 실험 개요

본 장에서는 이질적 환경에서 촬영된 이미지에 대한 적대적 학습을 통해 모델의 일반화 성능을 향상시킨 제안 방법론의 성능을 평가한다.

Table 1은 실험에 사용된 벤치마크 데이터 셋[26]의 특징을 요약한다. 전체 데이터 수는 7,022장으로, 총 3대의 카메라를 통해 촬영되었다. 이 가운데 'C'사와 'N'사의 카메라를 통해 촬영된 이미지 4,902장을 모델 학습에 사용하였으며, 추론에는 학습에 사용되지 않은 'S'사 카메라의 촬영 이미지 2,120장을 사용하였다. 즉 'C'사와 'N'사 카메라 이미지에 대한 학습을 통해 도출한 모델이 'S'사 카메라 이미지의 분석에서도 높은 성능을 보인다면, 해당 모델은 일반화 가능성이 높은 모델로 인정받을 수 있다.

모델의 성능 평가에는 색 항상성 기술의 평가 지표로 널리 사용되는 Angular Error[14-16]를 사용한다. 해당 지표는 실제 조명(Ground Truth)의 RGB 값(\hat{y})과 예측 RGB 값(y) 사이의 각도를 계산하며[Equation (1)], 모델의 성능이 우수할수록 그 값이 작게 나타난다.

$$AE = \arccos\left(\frac{y \cdot \hat{y}}{\|y\| \|\hat{y}\|}\right) \quad (1)$$

본 실험에서는 비교 모델(A), (B), (C)와 제안 모델 (D)의 성능을 비교하였으며, Fig. 6은 성능 평가를 위한 실험의 전체 개요를 나타낸다. 비교 모델 (A)와 (B)는 각각 'C'사와 'N'사의 이미지만 사용하여 학습을 수행한 모델로, 학습은 제안 방법론의 '특징 추출기'인 패치 기반(Patch-based) CNN 구조를 이용한다. 비교 모델 (C)는 'C'사와 'N'사의 이미지를 모두 학습에 사용하지만, 도메인 정보를 활용하지 않고 단순히 두 이미지 집합을 통합하여 사용한다. 한편 (D)는 본 연구에서 제안하는 모델로, 적대적 학습을 통해 일반화 성능을 향

Table 1. Benchmark Dataset (INTEL-TAU)

Camera	Sample Size	Total
'C'사	2,109	7,022
'N'사	2,793	
'S'사	2,120	

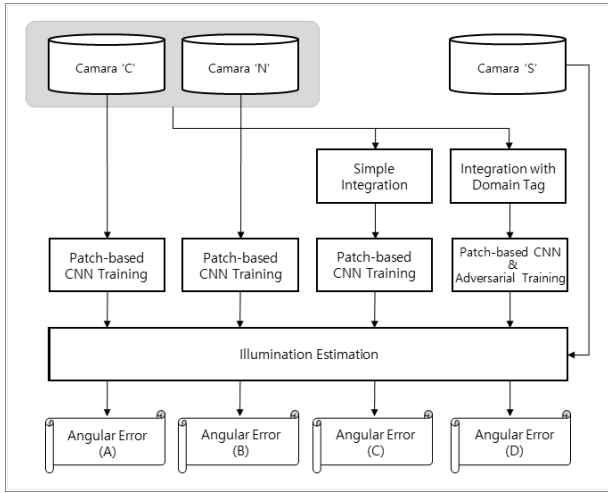


Fig. 6. Overall Process of Performance Evaluation

Table 2. Experimental Environments

HW	GPU	Tesla V100
	CPU	Xenon (R) 5120 2.20GHz
	Memory	180 GB
SW	Python	3.6.10
	Tensorflow-gpu	2.2.0
	Keras	2.3.0

상시킨 모델이다. 이들 네 가지 모델의 성능 평가를 위해 'S'사의 이미지에 대한 추론을 진행하며, 각 모델의 추론 결과에 대한 Angular Error를 비교한다.

Fig. 6의 'Patch-based CNN Training'의 세부 구조는 기본적으로 선행 연구[14]의 구조를 따르되, 일반화 성능 향상을 위해 합성곱층과 맥스풀링층 사이에 배치정규화층 (Batch Normalization Layer)을 삽입하였다. 또한 하이퍼 파라미터(Hyperparameter)로 배치 크기(Batch Size)는 64, 에폭(Epoch)은 10으로 지정하였다. 실험을 수행한 H/W 및 S/W 환경은 Table 2와 같다.

4.2 성능 평가 및 해석

본 실험에서 제시하는 비교 모델과 제안 모델은 모두 패치 단위로 학습이 이루어지며, 추론 시에도 패치 단위의 개별 입력은 자신의 원본 이미지의 RGB 값을 각각 예측한다. 이때 성능 평가에는 각 패치 단위로 정답과 예측 값의 Angular Error를 계산하는 방법(Each Patch), 또는 패치를 통합하여 각 이미지 단위로 정답과 예측 값의 Angular Error를 계산하는 방법이 사용될 수 있다. 패치의 통합은 각 이미지에 대해 밝기의 강도가 가장 높은 100개 패치의 RGB 값을 집계하는 방식으로 이루어지며, 집계에는 일반적으로 평균 풀링 (Average Pooling), 또는 중위수 풀링(Median Pooling) 방식이 널리 사용된다. Table 3은 위의 세 가지 방식에 따라 산출한 네 가지 모델 각각의 평균 Angular Error를 보이고 있으며, 이를 도식화한 결과는 Fig. 7과 같다.

Table 3. Experimental Results

Unit of Analysis	Model	Average Angular Error
Per Patch	(A) Camera 'C'	7.7160
	(B) Camera 'N'	6.2980
	(C) Simple Integration	7.5475
	(D) Proposed	5.8676
Per Image (Average Pooling)	(A) Camera 'C'	7.6938
	(B) Camera 'N'	6.2160
	(C) Simple Integration	7.5513
	(D) Proposed	5.7450
Per Image (Median Pooling)	(A) Camera 'C'	7.7077
	(B) Camera 'N'	6.2949
	(C) Simple Integration	7.4876
	(D) Proposed	5.8268

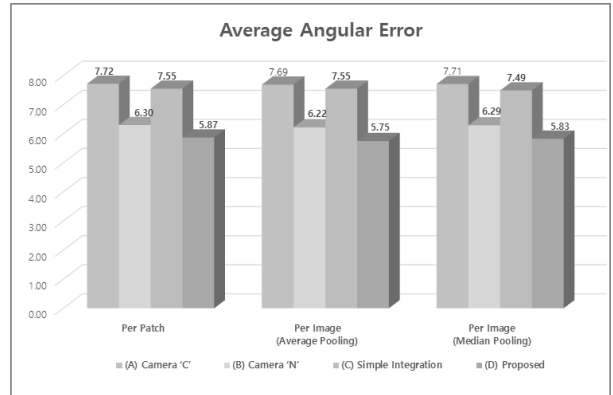


Fig. 7. Performance Comparison

Table 3과 Fig. 7에서 Angular Error의 산출 단위, 즉 패치 단위의 산출과 이미지 단위의 산출은 큰 차이를 나타내지 않았으며, 세 가지 모든 방식에서 제안 방법론의 Angular Error가 가장 낮은 것으로 나타났다. 이는 제안 방법론이 적대적 학습을 통해 도메인 응집성이 높은 특성을 제거함으로써 일반화 성능이 우수한 모델을 생성하였음을 나타낸다. 한편 각각 'C'사와 'N'사의 이미지만 사용하여 학습을 수행한 비교 모델 (A)와 (B)의 성능 차이가 크게 나타난 것은 이미지 소스의 특성에 기인한 것으로 볼 수 있다. 즉 학습에 사용된 'N'사의 이미지는 다소 청색을 띠고 추론에 사용된 'S'사의 이미지는 다소 청녹색을 띠었는데, 두 데이터 셋의 우연한 도메인 유사성으로 인해 (B) 모델의 성능이 과평가된(Overestimated) 것으로 해석된다.

색 항상성 프로세스의 효과는 실제 이미지의 비교를 통해 보다 직관적으로 확인할 수 있다. Fig. 8은 'S'사의 이미지에 대한 추론 실험에서 획득한 이미지로, 가장 왼쪽의 Original Image는 모델의 입력으로 사용된 이미지를, 그리고 바로 우측의 Ground Truth는 실제 조명의 RGB 값이 알려진 상태에서 색 항상성 기술 적용을 통해 보정한 정답 이미지를 나타

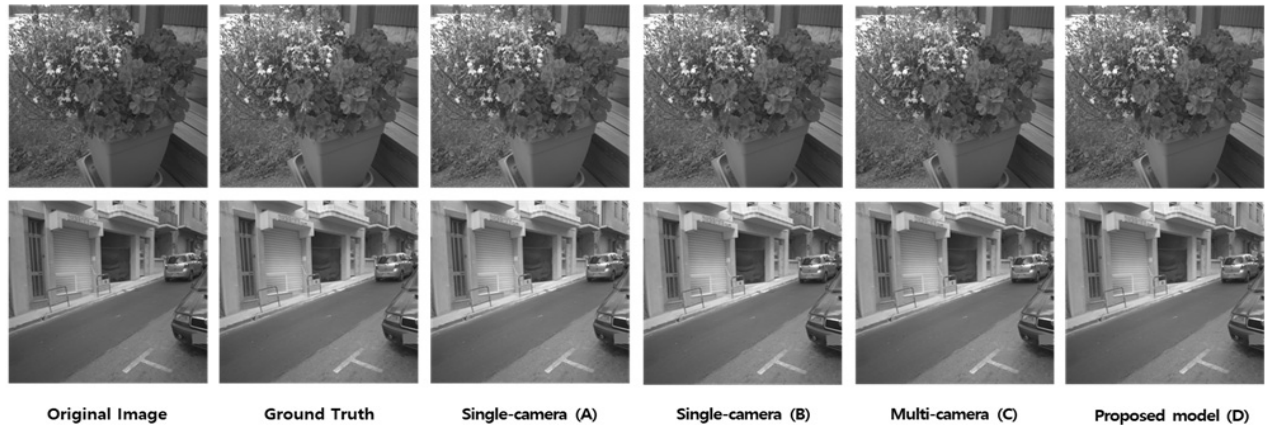


Fig. 8. Results of Applying Color Constancy

낸다. 우측의 네 열은 실제 조명의 RGB 값을 사용하지 않고 각각 비교 모델 (A), (B), (C)와 제안 모델 (D)을 통해 이미지를 보정한 결과를 나타낸다. 그림을 통해 네 가지 모델 모두 Original Image에서 나타난 초록빛을 효과적으로 제거하고 있으며, 특히 맨 우측의 제안 모델이 Ground Truth와 가장 유사한 이미지를 생성한 것을 육안으로도 확인할 수 있다.

본 장에서는 평균 Angular Error를 통한 정량적인 비교, 그리고 실제 이미지의 육안 확인을 통한 직관적인 비교를 통해, 제안 모델이 여러 비교 모델에 비해 우수한 일반화 성능을 보임을 확인하였다.

5. 결 론

본 연구는 방대한 이미지 데이터를 컴퓨터 비전 분야의 딥러닝 기술에 효과적으로 활용하기 위한 방안을 제안하고 있다. 이질적 환경에서 촬영된 이미지는 동일한 정보임에도 촬영 환경에 따라 각 이미지의 특징이 상이하게 표현될 수 있다. 이는 각 이미지가 갖는 상이한 환경 정보뿐 아니라 이미지 고유의 정보조차 서로 상이한 특징으로 표현되며, 이로 인해 이들 이미지 정보는 서로 잡음으로 작용해 모델의 분석 성능을 저해할 수 있음을 의미한다. 이에 본 논문에서는 적대적 학습을 통해 이질적 환경에서 생성된 이미지들의 색 항상성을 향상시키는 방안을 제안하였으며, 이질적 환경에서 촬영된 이미지 데이터 셋 7,022장에 대한 실험을 통해 제안 방법론의 우수성을 확인하였다.

본 연구에서는 이미지가 생성된 서로 다른 환경을 각각의 도메인으로 인식하여 적대적 학습을 적용하는 방안을 새롭게 제시하였으며, 이는 본 연구의 학술적 기여로 인정받을 수 있다. 또한 제안 방법론의 일반화 성능이 기존 방법론에 비해 'Angular Error' 측면에서 우수한 것으로 확인되었으며, 이는 본 연구의 성과가 실무적으로도 충분히 활용될 수 있음을 의미한다. 특히 제안 방법론의 모듈은 비교적 간단한 구조로 이루어져 있으므로, 일반화 성능을 향상시키고자 하는 다양한 딥러닝 응용 분야에 널리 활용될 수 있을 것으로 기대한다.

다. 추후 연구에서는 적대적 학습을 구성하는 하이퍼 파라미터 튜닝, 학습에 사용되는 이미지 구성의 다양화, 학습 및 추론 시간의 평가 등을 통해 제안 모델의 성능 및 견고성을 향상시키기 위한 방안이 모색되어야 한다.

References

- [1] T. Kim and N. Kim, "Deep learning-based professional image interpretation using expertise transplant," *Journal of Intelligence and Information Systems*, Vol.26, No.2, pp.79-104, 2020.
- [2] H. Kim, H. S. Oh, and D. Kim, "CNN architecture predicting movie rating from audience's reviews written in Korean," *KIPS Transactions on Computer and Communication Systems*, Vol.9, No.1, pp.17-24, 2020.
- [3] A. Gijsenij, T. Gevers, and J. V. Weijer, "Computational color constancy: Survey and experiments," *IEEE Transactions on Image Processing*, Vol.20, No.9, pp.2475-2489, 2011.
- [4] J. Xiao, S. Gu, and L. Zhang, "Multi-domain learning for accurate and few-shot color constancy," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp.3258-3267, 2020.
- [5] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, and V. Lempitsky, "Domain-adversarial training of neural networks," *Journal of Machine Learning Research*, Vol.17, No.1, pp.2096-2030, 2016.
- [6] G. Buchsbaum, "A spacial processor model for object color perception," *Journal of the Franklin institute*, Vol.310, No.1, pp.1-26, 1980.
- [7] E. Land, "The retinex theory of color vision," *Scientific American*, Vol.237, No.6, pp.108-128, 1977.
- [8] J. V. Weijer, T. Gevers, and A. Gijsenij, "Edge-based color constancy," *IEEE Transactions on Image Processing*, Vol.16, No.9, pp.2207-2214, 2007.

- [9] K. F. Yang, S. B. Gao, and Y. J. Li, "Efficient illuminant estimation for color constancy using grey pixels," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.2254-2263, 2015.
- [10] K. Barnard and G. Finlayson, "Shadow identification using colour ratios," in *Proceedings of the IS&T's Color and Imaging Conference*, Vol.2000, No.1, pp.97-101, 2000.
- [11] B. Funt and W. Xiong, "Estimating illumination chromaticity via support vector regression," in *Proceedings of the IS&T's Color and Imaging Conference*, Vol.2004, No.1, pp.47-52, 2004.
- [12] D. Cheng, B. Price, S. Cohen, and M. S. Brown, "Effective learning-based illuminant estimation using simple features," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.1000-1008, 2015.
- [13] S. Bianco, C. Cusano, and R. Schettini, "Color constancy using cnns," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp.81-89, 2015.
- [14] J. T. Barron, "Convolutional color constancy," in *Proceedings of the IEEE International Conference on Computer Vision*, pp.379-387, 2015.
- [15] J. T. Barron and Y. T. Tsai, "Fast fourier color constancy," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp.886-894, 2017.
- [16] Y. Hu, B. Wang, and S. Lin, "Fc4: Fully convolutional color constancy with confidence weighted pooling," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.4085-4094, 2017.
- [17] M. Rizzo, C. Conati, D. Jang, and H. Hu, "Cascading convolutional temporal colour constancy," *arXiv preprint arXiv:2106.07955*, 2021.
- [18] I. Goodfellow, J. P. Abadie, M. Mirza, B. Xu, D. W. Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *Advances in Neural Information Processing Systems*, pp.2672-2680, 2014.
- [19] Y. Wang, F. Ma, Z. jin, Y. Yuan, G. Xun, K. Jha, L. Su, and J. Gao, "Eann: Event adversarial neural networks for multimodal fake news detection," in *Proceedings of the 24th ACM SIGKDD International Conference On Knowledge Discovery & Data Mining*, pp.849-857, 2018.
- [20] D. Li, D. Chen, L. Shi, B. Jin, J. Goh, and S. K. Ng, "MAD-GAN: Multivariate anomaly detection for time series data with generative adversarial networks," in *International Conference on Artificial Neural Networks*, Springer, pp.703-716, 2019.
- [21] Z. Yang, J. Hu, R. Salakhutdinov, and W. Cohen, "Semi-supervised qa with generative domain-adaptive nets," *arXiv preprint arXiv:1702.02206*, 2017.
- [22] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by backpropagation," in *International Conference on Machine Learning*, pp.1180-1189, 2015.
- [23] B. Sun, J. Feng, and K. Saenko, "Return of frustratingly easy domain adaptation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol.30, No.1, 2016.
- [24] H. Yuzuguzel, "Learning colour constancy using convolutional neural networks," MS thesis, Tampere University of Technology, Finland, 2015.
- [25] H. R. V. Joze, M. S. Drew, G. D. Finlayson, and P. A. T. Rey, "The role of bright pixels in illumination estimation," in *Proceedings of the IS&T's Color and Imaging Conference*, Vol.2012, No.1, pp.41-46, 2012.
- [26] F. Laakom, J. Raitoharju, J. Nikkanen, A. Iosifidis, and M. Gabbouj, "Intel-tau: A color constancy dataset," *IEEE Access*, Vol.9, pp.39560-39567, 2021.



김 준 우

<https://orcid.org/0000-0001-5362-5766>
 e-mail : britz17@kookmin.ac.kr
 2021년 한국외국어대학교 언론정보전공 (학사)
 2010년 ~ 현 재 국민대학교 비즈니스IT 전문대학원 석사과정

관심분야 : Deep Learning, Data Mining



김 남 규

<https://orcid.org/0000-0002-7228-4562>
 e-mail : ngkim@kookmin.ac.kr
 1998년 서울대학교 컴퓨터공학과(학사)
 2000년 한국과학기술원 경영공학과(석사)
 2007년 한국과학기술원 경영공학과(박사)
 2007년 ~ 현 재 국민대학교 비즈니스IT 전문대학원 및 경영정보학부 교수

관심분야 : Deep Learning, Text Mining, Data Modeling