

# 이미지 저작권 판별을 위한 기계학습 적용과 분석

김수인<sup>1</sup>, 이상우<sup>2</sup>, 김학희<sup>3</sup>, 김원겸<sup>2</sup>, 황두성<sup>1</sup>

<sup>1</sup>단국대학교 소프트웨어학과

<sup>2</sup>(주)에이아이딥

<sup>3</sup>한국문학예술저작권협회

kimsooinj@gmail.com, lswross012@aideep.ai, hskim@kolaa.kr,

wgkim@aideep.ai, dshwang@dankook.ac.kr

## Application and Analysis of Machine Learning for Discriminating Image Copyright

Soojin Kim<sup>1</sup>, Sangwoo Lee<sup>2</sup>, Hakhee Kim<sup>3</sup>,

Wongyum Kim<sup>2</sup>, Doosung Hwang<sup>1</sup>

<sup>1</sup>Dept. of Software Science, Dankook University

<sup>2</sup>AiDEEP

<sup>3</sup>Korea Literature, Academic works and Art Copyright Association

### 요 약

본 논문은 이미지 저작권 유무 판별을 분류 문제로 정의하고 기계학습과 합성곱 신경망 모델을 적용하여 해결한다. 학습을 위해 입력 데이터를 고정된 크기로 변환하고 정규화 과정을 수행하여 학습 데이터셋을 준비한다. 저작권 유무 판별 실험에서 SVM,  $k$ -NN, 랜덤포레스트, VGG-Net 모델의 분류 성능을 비교 분석한다. VGG-Net C 모델의 결과가 다른 알고리즘과 비교 시 10.65% 높은 성능을 나타냈으며 배치 정규화 층을 이용하여 과적합 현상을 개선했다.

### 1. 서론

인터넷 사용자 증가에 따라 저작권 이미지에 대한 무분별한 사용이 사회 문제로 대두되고 있다. 이미지의 저작권 유무 판단을 위해 평균 해시 매칭(average hash matching), 특징화, CBIR(content based image retrieval) 응용[1], 딥러닝 기반 이미지 분류 모델 등이 사용되었다.

평균 해시 매칭은 이미지 특징을 해시값으로 표현해 유사도를 비교하는 이미지 검색 방법이다. 전체 픽셀(pixel)의 평균을 임계값으로 설정해 임계값보다 낮은 픽셀 값은 0, 큰 값은 1로 변환하여 해시값을 표현한다. 이미지 특징화는 컬러 히스토그램, HOG(Histogram of Oriented Gradients) 특징 등 영상처리 기법을 이용한다.

CBIR은 컴퓨터 비전 기술을 사용해 이미지를 검색하는 기술이다. 색상, 모양 텍스처(texture) 등 이미지의 콘텐츠 특징을 이용하여 이미지를 검색한다. CBIR은 딥러닝과 함께 사용되어 이미지 저작권 유무 판단 시스템으로 확장될 수 있다. 오토인코더(autoencoder) 기반 이미지 특징화와 합성곱 신경망(convolution neural network) 모델을 이용한 이미지 분류를 통해 저작권 유무를 판단할 수 있다.

이승재 외 4인은 ILSVRC(ImageNet Large Scale Visual Recognition Challenge) 이미지 분류 대회에서 사용된 기계학습과 합성곱 신경망 모델을 분석했다[2]. 2010년 트론토 대학은 선별 탐색 후보 영역 선별을 통한 SVM(Support Vector Machine) 기반 회귀 모델을 활용해 우수한 분류 성능을 보고했다. 2012년에는 토론토 대학이 제안한 AlexNet[3] 합성곱 신경망 모델이 16%의 가장 낮은 오차율을 기록했다. 이후 2개의 합성곱 필터(filter)와 하나의 풀링(pooling) 계층으로 구성된 블록을 사용한 VGG-Net[4]이 높은 분류 성능으로 주목받았다.

구글은  $1 \times 1$ ,  $3 \times 3$ ,  $5 \times 5$  크기의 합성곱 필터를 사용한 인셉션(inception) 모듈 기반 합성곱 신경망 구조를 제안했다. 2015년도 대회에서 마이크로소프트는 잔여 블록(residual block)을 이용한 ResNet[5]을 제안하여 기울기 소실 문제를 보완시켰다.

본 논문에서는 이미지 저작권 유무 판별을 분류 문제로 정의하고 기계학습과 VGG-Net 모델을 적용하여 해결한다. VGG-Net 모델은 합성곱 신경망 구조 변화와 배치 정규화(batch normalization)에 따른 분류 성능을 분석한다. 2절에서 수집된 데이터셋의 특성과 전처리 방법을 제안한다. 3절에서는 저작권

유무 판별 문제 정의와 분류 모델 별 성능을 비교 분석하며 마지막 절에서는 향후 연구 방향과 결론을 서술한다.

## 2. 제안 방법

### 2.1 데이터셋 구성

저작권 이미지는 한국 문학 예술 저작권 협회[6]로부터 제공받았다. 해당 데이터셋은 블루 캔버스 연대, 한국 사진 작가 협회, 한국 문화 예술 협회에서 제공받은 이미지와 교과서 스캔본으로 구성된다. 비저작권 이미지는 ILSVRC2012 ImageNet을 이용했다. 표 1은 데이터셋 구성이며 그림 1은 수집된 이미지의 예시이다.

표 1. 데이터셋 구성

저작권	제공 기관	이미지 개수
유	블루 캔버스 연대	229,756
	한국 사진 작가 협회	24,925
	한국 문화 예술 협회	301,758
	교과서 스캔본	363
무	ILSVRC2012	1,500,000
	합계	2,056,802

### 2.2 데이터 전처리

그림 2는 수집된 이미지의 크기 분포이다. 대부분의 이미지가 0부터 10,000 픽셀 범위의 너비와 높이를 갖는다. 분류 모델 학습을 위해 다양한 크기의 이미지를 고정된 크기의 벡터로 표현해야 한다. 이미지를  $256 \times 256$  크기로 변환했으며, 일반화된 모델 성능을 위해 Z-score 표준화를 적용했다. 표준화

된 이미지의 픽셀 값은 평균 0, 표준 편차 1인 정규 분포 형태를 갖는다.

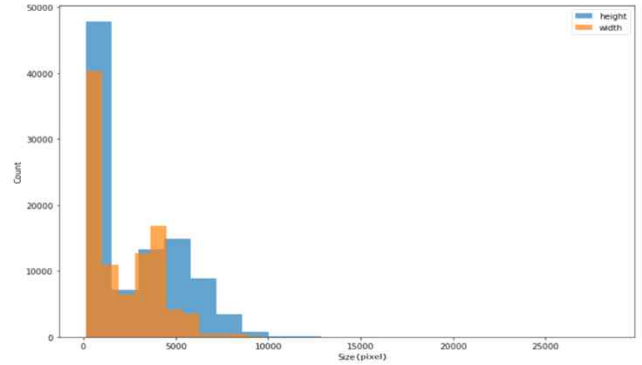


그림 2. 이미지 크기 분포

## 3. 실험 및 결과

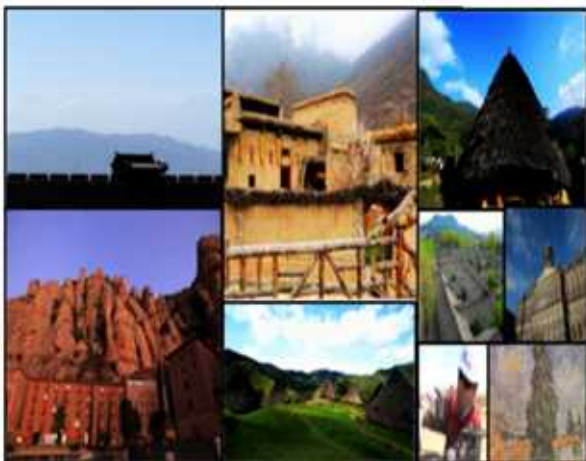
### 3.1 분류 문제 정의

저작권과 비저작권 이진 분류 문제 구성을 위해 클래스마다 무작위로 50,000개의 데이터를 선택해 학습 데이터를 구성했다. 모델 평가는 5-식 교차 검증(5-way cross validation)을 수행했으며 성능 지표는 정확도(accuracy), 정밀도(precision), 재현율(recall), F1-score를 이용했다.

학습 알고리즘은  $k$ -NN( $k$ -Nearest Neighbor), 랜덤포레스트(Random Forest), SVM, VGG-Net C 모델을 사용한다. VGG-Net 이외의 분류 모델은 이미지를 1차원으로 평탄화해 실험을 진행했다.

### 3.2 학습 알고리즘 성능 비교

표 2는 학습 알고리즘별 분류 성능 비교이다. 평



(a) 저작권



(b) 비저작권

그림 1. 수집된 이미지 예시

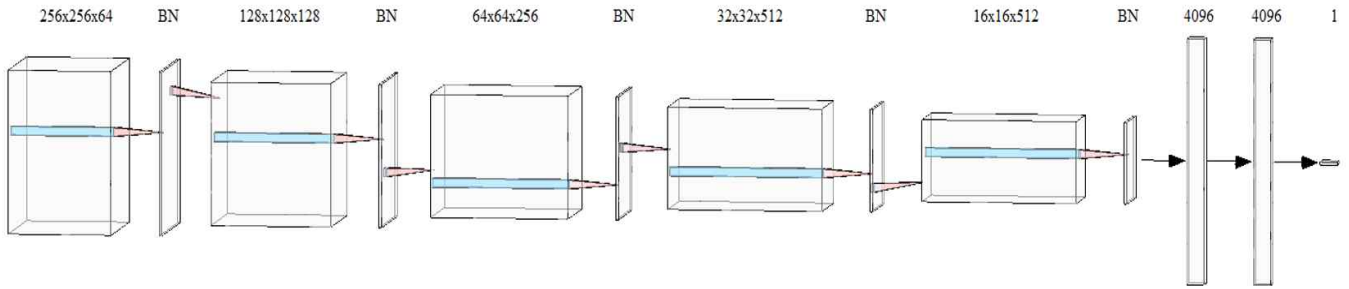


그림 3. 배치 정규화가 적용된 VGG-Net C 모델 구조

가 성능에서 VGG-Net C(VGG-C) 모델이 77.3%의 정확도와 78.0%의 F1-score로 가장 우수한 성능을 보였다. SVM과 랜덤포레스트(RF)는 각각 64.2%, 66.5%의 정확도를 얻었으며  $k$ -NN의 경우 정확도 59.9%로 가장 낮은 분류 성능을 보고했다. VGG-Net C, SVM,  $k$ -NN은 학습과 평가 성능을 비교 했을 때 평가 지표의 차이가 높아 과적합(overfitting)이 분석된다.

표 2. 학습 알고리즘별 평가 성능 비교

Model		SVM	$k$ -NN	RF	VGG-C
Metric					
Train	Accuracy	0.992	0.834	0.701	1.000
	Precision	0.999	1.000	0.702	1.000
	Recall	0.994	0.668	0.701	1.000
	F1-score	0.992	0.801	0.702	1.000
Test	Accuracy	0.642	0.599	0.665	0.773
	Precision	0.687	0.608	0.668	0.774
	Recall	0.673	0.601	0.666	0.786
	F1-score	0.680	0.594	0.664	0.780

### 3.3 VGG-Net 구조 분석

합성곱 신경망은 계층 수가 증가하면 높은 과적합 현상이 나타날 수 있다. VGG-Net 모델 구조와 배치 정규화 층에 따른 성능 변화 비교를 통해 과적합을 분석한다. 표 3은 합성곱 계층 변화에 따른 VGG-Net 모델 구조이다. C 모델을 제외한 모든 모델의 합성곱 필터의 크기는  $3 \times 3$ 이며 스트라이드(stride)는 1이다. 최대 풀링(max pooling) 계층은  $3 \times 3$  크기의 필터와 스트라이드 2로 구성된다. 모델 A부터 E로 갈수록 합성곱 계층 수가 증가하며 완전 연결 층의 구조는 동일하다.

표 4는 VGG-Net 모델별 성능 평가이다. C 모델이 가장 높은 분류 성능을 나타냈다. 합성곱 계층

표 3. VGG-Net 모델별 구조

VGG-Net Architecture				
A	B	C	D	E
conv3-64	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
Max Pooling				
conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
Max Pooling				
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 conv1-256	conv3-256 conv3-256	conv3-256 conv3-256 conv3-256 conv3-256
Max Pooling				
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
Max Pooling				
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
Max Pooling				
Fully Connected Layer-4096				
Fully Connected Layer-4096				
Fully Connected Layer-1000				
Sigmoid				

수가 적은 A 모델은 C와 비교하여 약 1.75%의 정확도 차이를 보고했다. 그러나 A 모델에서도 학습과 평가 성능의 차이가 높아 과적합 현상을 보였다. 표 5는 VGG-Net C 모델에 배치 정규화 층을 추가했을 때 나타나는 성능 변화이다. 모든 합성곱 계층 이후에 배치 정규화 층을 추가했으며 완전 연결 층 간 드롭 아웃(dropout) 비율은 0.5로 설정하였다(그림 3). 배치 정규화 전과 비교했을 때 정확도와 F1-score가 각각 2.9%, 5.3% 향상되었다.

표 4. VGG-Net 모델 평가 성능 비교

Model		VGG-Net				
		A	B	C	D	E
Train	Accuracy	0.988	0.899	1.000	0.996	0.914
	Precision	0.995	0.917	1.000	0.994	0.971
	Recall	0.995	0.920	1.000	0.974	0.890
	F1-score	0.995	0.919	1.000	0.928	0.928
Test	Accuracy	0.755	0.738	0.773	0.725	0.725
	Precision	0.770	0.737	0.774	0.776	0.755
	Recall	0.770	0.766	0.786	0.704	0.669
	F1-score	0.770	0.751	0.780	0.710	0.710

표 5. 배치 정규화 적용에 따른 모델 성능 변화

Model		VGG-Net C	VGG-Net C (+ Batch Norm)
Train	Accuracy	1.000	0.980
	Precision	1.000	0.993
	Recall	1.000	0.991
	F1-score	1.000	0.995
Test	Accuracy	0.773	0.801
	Precision	0.774	0.804
	Recall	0.786	0.790
	F1-score	0.780	0.819

그림 4와 5는 VGG-Net C 모델의 훈련 반복 수 변화에 따른 손실 그래프 변화이다. 그림 5는 배치 정규화 적용 전, 그림 6은 적용 후의 손실 그래프이다. 그림 5와 비교했을 때 그림 6에서 학습과 평가 데이터간 손실 격차가 감소되는 경향이 분석된다. VGG-Net C 모델의 과적합 현상이 배치 정규화 층을 이용하여 개선되었다.

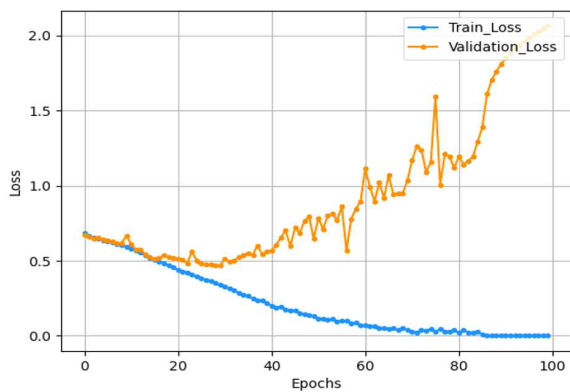


그림 4. 배치 정규화 전 최소 자승 손실 변화

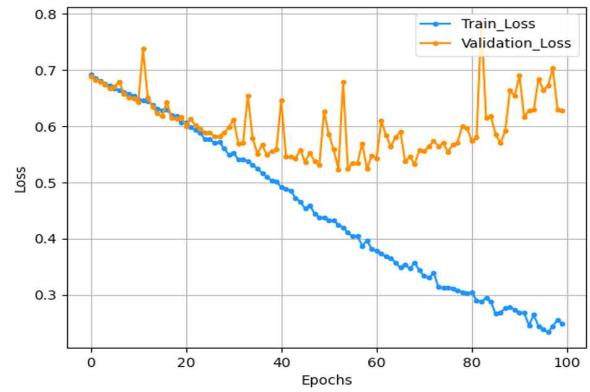


그림 5. 배치 정규화 후 최소 자승 손실 변화

### Acknowledgement

이 논문은 2021년도 정부(문화체육관광부)의 재원으로 한국저작권위원회의 지원을 받아 수행된 연구임(No.2021-ec-9500S2, 교육 콘텐츠에 대한 인공지능 기반 저작권 침해 의심요소 검출 및 대체 재료 콘텐츠 추천 기술 개발)

### 참고문헌

- [1] Saritha, R. Rani, Varghese Paul, and P. Ganesh Kumar. "Content based image retrieval using deep learning process." Cluster Computing 22.2 (2019): 4187-4200.
- [2] Lee, S. J., et al. "Technology Trends and Analysis of Deep Learning Based Object Classification and Detection." Electronics and Telecommunications Trends, vol. 33, no. 4, 한국전자통신연구원, Aug. 2018, pp. 33 - 42.
- [3] Krizhevsky, Alex, Ilya Sutskever, and G. Hinton. "Imagenet classification with deep convolutional networks." Proceedings of the Conference Neural Information Processing Systems (NIPS).
- [4] Simonyan, K. and Zisserman, A. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).
- [5] He, K. et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
- [6] KOLAA [Internet], [https://www.kolaa.kr/jsp/main/MainCtrl.jsp?ACT\\_CD=MVW](https://www.kolaa.kr/jsp/main/MainCtrl.jsp?ACT_CD=MVW)