Journal of Korea Multimedia Society Vol. 21, No. 8, August 2018(pp. 858-863)
https://doi.org/10.9717/kmms.2018.21.8.858

국방분야 비인가 이미지 파일 탐지를 위한 다중 레벨 컨볼루션 신경망 알고리즘의 구현 및 검증

김 영 수†

Implementation and Verification of Multi-level Convolutional Neural Network Algorithm for Identifying Unauthorized Image Files in the Military

Youngsoo Kim†

ABSTRACT

In this paper, we propose and implement a multi-level convolutional neural network (CNN) algorithm to identify the sexually explicit and lewdness of various image files, and verify its effectiveness by using unauthorized image files generated in the actual military. The proposed algorithm increases the accuracy by applying the convolutional artificial neural network step by step to minimize classification error between similar categories. Experimental data have categorized 20,005 images in the real field into 6 authorization categories and 11 non-authorization categories. Experimental results show that the overall detection rate is 99.51% for the image files. In particular, the excellence of the proposed algorithm is verified through reducing the identification error rate between similar categories by 64.87% compared with the general CNN algorithm.

Key words: Multi-level CNN, Convolutional Neural Network, Image Processing, Intelligent Information System, Military Application

1. 서 론

최근 몇 년 동안 우리나라는 경제 발전과 민주화로 인해 양성평등의 요구가 높아지면서 여성들의 군 입대 비율이 지속적으로 증가하고 있다. 이러한 여성 인력이 국방의 많은 분야에서 긍정적인 발전에 기여하는 반면, 성폭력과 같은 부작용의 발생도 꾸준히 증가하고 있다. 또한 정보통신기술 (ICT)의 발전과 인프라의 확장으로 다양한 경로를 통한 음란성 이미지의 유통과 공유가 이루어져 건전한 군대의 병영 문화에 위협이 되고 있다. 이러한 음란성 파일은 성폭력 사고를 유발할 뿐만 아니라 군 업무 네트워크에 커다란 트래픽을 발생시켜 정상적인 업무를 방해한다. 또한, 북한의 인공기나 독재자 김정은의 사진과 같은 불온성 이미지 자료들이 군대 내에서 확산·유통되는 것은 군인들의 대적관을 약화시킬 수 있으므로 지휘관은 이러한 자료들의 유입을 사전에 확인하고 예방하는 것이 점점 더 중요한 문제로 부각되고 있다.

우리 군의 군사보안업무훈령에 따르면, 군에서는 누구나 외부(인터넷)에서 군 내부 네트워크(인트라넷)로 파일을 가져 오기위해서는 보안 관리자의 승인을 받아야한다. 그러나 외부로부터 비인가 이미지 파일의 유입을 완벽하게 차단하는 것은 실질적으로

※ Corresponding Author : Youngsoo Kim, Address: (700-842) Danjae-ro 635, Sangdang-gu, Chongju, Korea, TEL: +82-43-290-6548, FAX : +82-43-297-0479, E-mail : yskim42@afa.ac.kr

Receipt date : Jul. 12, 2018, Approval date : Jul. 24, 2018
† Dept. of Computer & Information Engineering, Korea Air Force Academy

불가능하다. 왜냐하면, 컴퓨터 긴급 대응팀(CERT, Computer Emergency Response Team)의 검사관이 외부에서 오는 많은 양의 이미지 파일을 일일이 육안으로 확인하는 것은 현실적으로 불가능하기 때문이다. 따라서 이러한 인가되지 않은 음란성과 불온성 이미지와 같은 비인가 이미지 파일들에 대한 자동화된 검증체계가 요구된다.

최근 인공지능 기술의 비약적인 발전으로 컨볼루션 신경망(Convolutional neural network, 이하 'CNN')은 이미지 탐지 및 분류에서 압도적인 성능을 보이고 있으며 다양한 영역에서 활용되고 있다.[1] 본 논문에서는 이러한 CNN을 기반으로 많은 양의 비인가 이미지 파일을 식별하고 차단하기 위해 기존 CNN 알고리즘을 응용한 다중 레벨 CNN 알고리즘을 제안하고 실제 국방 인트라넷 내에서 시스템을 구현하여 그 효율성을 입증한다. Google의 오픈 소스 기계학습 개발 프레임워크인 Tensorflow[2]를 사용하여 다중 레벨 CNN 모델을 설계하였으며 데이터 집합은 실제 현장에서 발생한 비인가 이미지 유입 사례를 토대로 11개의 비인가 범주와 6개의 인가 범주로 구성된 17가지 범주로 구성하였다. 학습 데이터의 부족한 문제점을 극복하고 시스템의 성능을 향상시키기 위해 Google의 Inception[3] 모델을 초기값으로 적용하여 학습을 통한 최적화 모델을 도출하였다.

이 논문의 나머지 부분은 다음과 같이 구성된다. 2장에서는 관련 연구에 대해 살펴보고, 3장에서는 비인가 이미지 파일을 탐지하기 위해 제안된 다중 레벨 CNN 알고리즘과 구현되는 시스템 아키텍처에 대해 설명한다. 4장에서는 실제 발생된 데이터 집합을 바탕으로 비인가 파일 탐지 시스템의 성능을 확인하고 제안된 알고리즘의 효용성을 분석한다. 마지막으로 5 장에서는 이 논문의 제안 결과를 요약하고 향후 연구 과제에 대해 논의한다.

2. 관련 연구

인터넷 웹 사이트에서의 성인물 탐지를 비롯한 여러 영역에서 음란성 이미지 파일 탐지에 대한 많은 연구를 통해 다양한 알고리즘과 전략이 제안되었다.

그동안 음란성 이미지를 식별하는 데에는 주로 픽셀 기반 처리 방법이 사용됐으며 대표적으로 이미지에서 피부색의 비율을 추출하여 차별화하는 방법이 사용되었다[4]. Rowley는 음란성 여부를 판단하기 위해 피부의 색 뿐만 아니라 모양과 같은 다른 제약조건을 추가하여 좀 더 정확성을 기하는 방법을 제안하였다[5]. 그러나 이러한 알고리즘들은 색상 변조에 취약하고 음란성을 판단하는 다른 많은 변수가 있기 때문에 좋은 성능을 기대하기가 어려웠다.

Avila는 음란 이미지를 탐지하기 위한 BossaNova라는 컨텐츠 기반의 개념을 제안하였다[6]. BossaNova는 색상과 모양을 모두 나타내는 HueSIFT Descriptor을 사용하므로써 모양이나 에지만을 표현하는 표준 BoW (Bag of Words) 모델의 성능을 능가하였다. 또한, Caetano는 BossaNova와 함께 Binary Descriptor를 적용하여 이미지 분류에 보다 적합하도록 확장시켰으나 근본적은 성능향상의 한계를 보였다[7].

최근에는 이미지 분류에서 CNN이 픽셀 기반 알고리즘의 한계를 뛰어 넘는 탁월한 성능으로 인해 다양한 영역에 적용되는 여러 가지 사례들이 발표되었다. Ciresan은 CNN을 사용하여 독일 교통 신호를 인식하고, Kang은 CNN을 통해 담배 회사의 소송 문서의 이미지를 우수한 성능으로 분류하기도 하였다[8,9]. CNN은 기본적으로 이미지 분류에 우수한 성능을 보이지만 적용 도메인의 특성을 고려하여 알고리즘을 최적화하고 성능 향상을 도모할 필요가 있다[10]. 본 논문에서는 국방분야의 비인가 이미지 파일 탐지를 위한 최적화 알고리즘으로서 다중 레벨 CNN 알고리즘을 제안하고 실 데이터를 활용하여 성능을 검증한다.

3. 다중 레벨 컨볼루션 신경망 알고리즘

모든 유형의 CNN은 기본적으로 컨볼루션(Convolution) 계층, 풀링(Pooling) 계층 및 완전 연결된(Fully Connected) 계층으로 구성된다. 컨볼루션 계층은 컨볼루션 연산을 통해 입력 영상으로부터 특징맵(Feature Map)을 추출하는 과정으로 각 화소에 동일한 필터를 적용한다. 컨볼루션이란, 영상 처리 분야의 입력 영상에 필터 (마스크)를 적용하여 영상을 필터링하거나 특징을 추출하는 일련의 처리를 의미하며, 수식 (1)과 같이 나타낼 수 있다.

$$S(t) = (x * w)(t) = \int x(a)w(t-a)da \qquad (1)$$

Fig. 1. An example of the multi-level CNN structure.

여기에서 x는 입력 이미지를, w는 필터를, S(t)는 입력에 대한 특징 맵을 나타낸다.

풀링 계층은 데이터를 감소시키기 위해 최대 풀링(Maximum Pooling) 또는 평균 풀링(Average Pooling)과 같은 하위 샘플링을 통해 하나의 픽셀에 복수의 픽셀을 매핑하는 추상화 과정을 수행한다. 이러한 컨볼루션 및 풀링 계층은 데이터를 크게 줄이고 의미 있는 특징에 집중하여 식별 성능을 향상시키기 위해 지속적으로 교차 적용된다.

마지막 단계로, 완전 연결 계층은 기존 신경망의 다중 계층 퍼셉트론과 동일하며 학습은 역전파 알고리즘을 통한 입력과 출력 간의 오차를 최소화하는 방향으로 반복된다. 입력이 수신되면 각 범주에 대한 확률은 수식 (2)로 계산되며 가장 높은 확률을 가진 범주로 자동 분류된다.

$$P(y=i|x,w_1,\cdots,w_M,b_1,\cdots,b_M)=\frac{e^{w_i x+b_i}}{\sum_{j=1}^{M} e^{w_j x+b_j}} \quad (2)$$

제안된 다중 레벨 CNN은 위에서 설명한 일반적인 CNN보다 탐지 성능을 향상시키기 위해 다단계 학습 모델을 가진다. 분류할 범주들 사이의 유사성으로 인해 혼란스러운 범주가 있는 경우, 첫 번째 단계의 CNN을 통해 그룹 범주가 만들어져 분류된다. 그런 다음 그룹 범주로 분류된 데이터 집합은 두 번째 단계의 CNN을 통해 하위 세부 범주로 분류되어 성능을 향상시킨다. 예를 들어, 북한의 독재자 김정은과 일반 대중의 얼굴이라는 두 개의 유사한 범주가 있다고 하자. Fig. 1에 표현된 것처럼 첫 번째 단계로는 FACE 그룹으로 범주화하여 첫 번째 단계의 CNN을 통해 다른 범주와 함께 분류된다. 그런 다음 FACE 범주로 분류된 데이터들을 대상으로 다시 두 번째 단계의 CNN을 통해 김정은과 일반 대중을 나누어 분류한다. 이러한 과정을 통하여 혼동을 유발할 수 있는 유사 범주 간의 식별오차를 줄여나갈 수 있으며, 필요에 따라 반복된다.

4. 비인가 이미지 파일 탐지 시스템

전체 비인가 파일 탐지 프로세스는 Fig. 2에서 보는 바와 같다. 인터넷이나 USB 또는 기타 이동식 미디어를 통해 외부에서 이미지 파일을 가져 와서 사용자 PC에 저장하면 PC에 미리 설치되어있는 에이전트가 사용자의 정보와 함께 파일을 탐지 서버로 전송한다. 다중 레벨 CNN 기반의 식별 프로세스는 서버에서 수행되고 결과는 CERT로 전송된다.

4.1 다중 레벨 CNN을 활용한 비인가 이미지 파일 탐지

비인가 이미지 파일 탐지 프레임워크는 Google의 오픈 소스 기반 기계 학습 라이브러리인 Tensorflow를 활용하여 구현되었다. 이미지의 픽셀 값은 파이썬(Python) 수치 분석을 위한 다차원 배열 표준 패키지(Numpy Array)의 벡터 형식으로 변환되고 컨볼루션되어 28×28 레이어를 생성한다. 각각의 레이어는 픽셀 임계값 이진화(Binarization)를 통하여 경계 값이 추출되고 각 신경망 노드와 함께 동시 다중 그래픽 연산을 통해 분석을 수행한다. 이러한 연산을 통

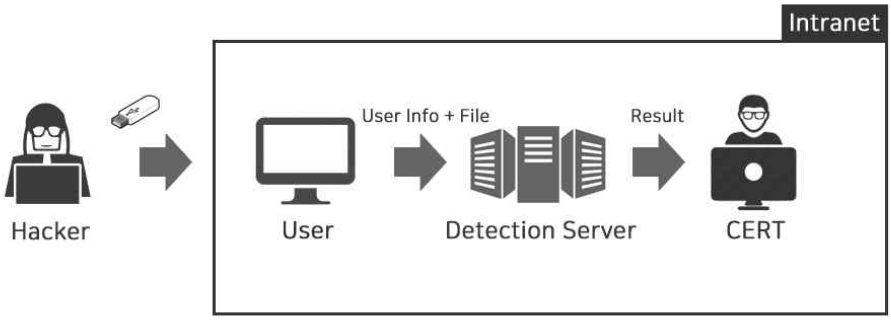

Fig. 2. Overall Process of the unauthorized image files detection system.

하여 최종적으로 학습된 데이터 집합에서 유사성을 발견하고 결과 값을 만들어 낸다.

전체적인 학습 데이터의 범주는 Table 1과 같이 17개의 범주로 나뉘며, 그 중 6개 범주는 '인가'이고 11개 범주는 '비인가'이다. 인가 범주는 군인들 간에 공개적으로 공유되는 건전한 이미지들을 기반으로 구성되고 비인가 범주는 음란성과 불온성과 같은 인가되지 않은 이미지들을 기반으로 한다. 비교적 비인가 이미지를 세분화한 것은 정확도를 높이기 위한 것이다.

시스템이 우수한 탐지 성능을 갖기 위해서는 일반적으로 많은 데이터와 이를 학습하기 위한 고성능의 컴퓨팅 자원 및 시간이 필요하다. 우리는 이러한 학습 데이터, 컴퓨팅 자원 및 시간을 절약하고 시스템의 탐지성능을 향상시키기 위해 Google의 이미지에 대한 우수한 학습 결과 모델인 Inception (버전 3)을 본 시스템의 학습 모델의 초기 값으로 활용하였다. Inception 모델은 Google의 CNN 기반 학습모델로서 총 1,000개의 범주로 구성되고 ImageNet Challenge의 데이터 집합에 대하여 93.9%의 높은 정확도를 제공한다[3]. 그런 다음 17개 범주에 대한 모델을 학습하면서 비인가 범주 당 약 5,000개의 데이터를 추가하여 학습시켰다. 이를 통해 처음부터 학습을 하는 대신 새로운 데이터로 재학습하는 단계를 적용한 전이 학습(Transfer Learning)이 수행되었다.

'인가' 및 '비인가'의 판별 공식은 수식 (3)과 같다. 어떤 i 번째 이미지 파일이 입력으로 들어올 때, 각 확률은 FACE(PUBLIC)와 FACE(JONG-EUN KIM)를 하나의 FACE 범주로 결합한 16개의 범주로 0에서 J까지 각 범주별 확률값을 계산한다. 계산된 확률은 0과 1 사이의 값으로 스케일되어 가중치 w로 치환된다. 해당 범주가 '인가' 범주에 속하면 C는 1이고, 그렇지 않으면 -1이 된다. 최종적으로 합계 결과가 0 보다 크면 '인가'로 결정되고, 0 보다 작 으면 '비인가' 으로 결정된다. 이는 같은 그림에서 각각의 범주가 서로 다른 비율로 나타날 수 있기 때문이다. 예를 들어 어떤 여성 사진이 있을 때, 얼굴은 작고 섹시한 몸매가 크거나 또는 그 반대 인 경우에 분류 결과는 각 범주별 비율의 크기에 따라 달라진다. 그리고 가장 높은 확률의 범주가 FACE 인 경우에는 다음 단계로 FACE(PUBLIC)과 FACE(JONG-EUN KIM)의 확률을 CNN을 통해 다시 비교하여 가장 높은 확률을 갖는 범주로 최종 결정된다.

$$d_i = \begin{cases} \text{인가}, & \text{if} \sum_{j=0}^{J} c_{ij} w_{ij} > 0 \\ \text{비인가}, & otherwise \end{cases} \quad (3)$$

5. 실험 및 결과 분석

제안된 비인가 파일 탐지 시스템은 Intel (R) Core (TM) i5-4560 CPU @ 3.30 GHz, Intel (R) HD 그래픽 4600 VGA 및 16 Gbytes RAM이 장착된 Windows 7(64 비트) PC에서 구현 및 실험되었다. 음란성과 불온성 관련된 45,062개의 이미지 파일이 군(軍) 내부망과 인터넷에서 수집되어 CNN 모델을 학습하기 위한 데이터 집합으로 사용되었다. 3장에서 언급 한 바와 같이, 학습된 CNN 모델 기반의 시스템은 3개월 동안 실제 국방 영역에 적용되었다. 35,786개의 이미지 파일이 탐지되었으며 이중에서 중복을 제외한 20,005개의 파일이 음란성 및 불온성 여부에 대한 판단이 이루어졌다.

우선, 일반적인 CNN 모델을 가지고 테스트를 수행하였다. Table 2에서 보이는 바와 같이 4가지 범주에서 279건의 비인가 이미지가 탐지되었다. 가장 빈번한 거짓 경보의 경우, CLOTHES(ORDINARY)의 122개 이미지가 CLOTHES(EROTIC) 범주로 분류된다. 이것은 기본적으로 CLOTHES(ORDINARY)와 CLOTHES(EROTIC)의 모호성에서 비롯되었다. 또한, FACE (JONG-EUN KIM) 범주와 FACE

Table 1. Categories by authorization vs. non-authorization

	Category
Authorization (6)	MILITARY ACTIVITY, SPORTS, EQUIPMENT, FACE(PUBLIC), CLOTHES (ORDINARY), ETC
Non-authorization (11)	SEXUAL ACTIVITY, QUISI-SEXUAL ACTIVITY, GENITAL EXPOSURE, ADULT CARTOON, FEMALE UPPER BODY, FEMALE LOWER BODY, CLOTHES(EROTIC), MALE WHOLE BODY, HOMO-SEXUALITY, FACE (JONG-EUN KIM), NORTH KOREAN FLAG

Table 2. Confusion matrix (general CNN → multi-level CNN)

Authorization / Non-authorization	CLOTHES (EROTIC)	FACE (JONG-EUN KIM)	OTHERS(B)
CLOTHES(ORDINARY)	122 → 15	-	-
FACE(PUBLIC)	-	81 → 7	-
OTHERS(A)	-	-	76

Table 3. Classification result of images

		Authorization	Non-authorization	Accuracy(%)
Authorization	general CNN	19,573	279	98.59
	multi-level CNN	19,754	98	99.51
Non-authorization		0	153	100

(PUBLIC) 범주의 혼란으로 인해 81개의 부정오류(False Negative)가 발생하였다. 이러한 부정오류 문제를 극복하기 위해, 3장에서 언급된 대로 다중 레벨 CNN이 적용되었다. 결과적으로 에로틱한 옷과 일반 옷의 혼동이 122에서 15로, 김정은과 일반 대중 사이의 혼동이 81에서 7로 급격히 감소하였다. 이를 통해서, 다중 레벨 CNN 알고리즘은 적용이 간편하면서도 유사도를 갖는 범주 간에 식별 오차를 줄이는데 매우 효과적임을 확인할 수 있다. 마지막의 76개의 오류는 유명인사 사진 또는 미술 사진과 같은 음란성의 관점이 사람마다 판단 기준이 다르기 때문에 발생한 것들이다. 그래서 이러한 경우에는 제안된 알고리즘을 적용하지 않았다.

이미지 파일의 전체 분류 결과는 Table 3과 같다. 구현된 비인가 파일 탐지 시스템이 비인가 이미지를 완벽하게 탐지하는 동안 인가된 일부 이미지는 비인가 이미지로 인식하였다. 일반적인 CNN이 적용되었을 때 비인가 등급 분류 비율을 포함한 전체 정확도는 98.61%이다. 다중 레벨 CNN의 경우 정확도는 99.51%로서 0.90% 향상되었다. 군대의 특성으로 인해 음란성이거나 불온성인 파일이 많지 않아서 정확도가 크게 향상되지는 않았지만 다중 레벨 CNN은 부정오류의 비율을 크게 줄였음을 알 수 있다. 즉, 부정 오류는 279 건에서 98건으로 64.87% 감소하였다. CLOTHES 및 FACE에서는 Table 2와 Table 3에서 설명된 대로 203 건에서 22 건으로 89.16% 감소시켰다.

6. 결론 및 향후과제

본 논문에서는 국방 영역에서의 비인가 이미지 파일을 탐지하기 위해 다중 레벨 CNN 알고리즘을 제안하고 실제 국방 인트라넷에서 시스템을 구현하여 발생한 데이터 집합으로 그 유효성을 검증하였다. 제안된 알고리즘은 비인가 파일들을 99.51%의 매우 높은 정확도로 탐지할 뿐만 아니라 인가 이미지를 비인가 이미지로 인식하는 탐지 오류는 일반 CNN 대비 89.16% 감소시켜서 그 유효성을 확인할 수 있었다. 뿐만 아니라 다중 레벨 CNN 알고리즘은 적용이 간편하여 응용에 따라 쉽게 확장하여 적용할 수 있는 메커니즘이라 할 수 있다.

향후에는 좀 더 일반화되고 정확한 성능향상을 위하여 더 많은 데이터들을 수집하여 학습 모델을 보다 정확하게 모델링되어야 한다. 특히, 유사도가 높은 경우들을 더욱 많이 식별하여 다중 레벨 CNN을 통한 부정오류를 줄여 나아가야 할 필요가 있다.

REFERENCE

[1] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” *Proceedings of Computing Research Repository*, arxiv.org/abs/1512.03385, 2015.

[2] Google Tensorflow, https://www.tensorflow.org/, (accessed Jun., 21, 2018).

[3] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Auguelov, et al., “Going Deeper with

Convolutions," *Proceedings of Conference on Computer Vision and Pattern Recognition*, pp. 1-9, 2005.

[4] A. Abadpour and S. Kasaei, "Pixel-based Skin Detection for Pornography Filtering," *Journal of Electrical and Electronic Engineering*, Vol. 1, Issue 3, pp. 21-41, 2005.

[5] H. Rowley, Y. Jing, and S. Baluja, "Large Scale Image-based Adult-content Filtering," *Proceeding of International Conference of Vision theory and Applications*, pp. 290-296, 2006.

[6] S. Avila, N. Thome, M. Cord, E. Valle, and A. Araujo, "Pooling in Image Representation: The Visual Codeword Point of View," *Computer Vision and Image Understanding*, Vol. 117, Issue 5, pp. 453-465, 2013.

[7] C. Caetano, S. Avila, S. Guimaraes, and A. Araujo, "Pornography Detection Using Bossanova Video Descriptor," *Proceedings of the 22nd European Signal Processing Conference*, pp. 1681-1685, 2014.

[8] D. Ciresan, U. Meier, J. Masci, and J. Schmidhuber, "A Committee of Neural Networks for Traffic Sign Classification," *Proceeding of The International Joint Conference on Neural Networks*, pp. 1918-1921, 2011.

[9] L. Kang, J. Kumar, P. Ye, Y. Li, and D. Doermann, "Convolutional Neural Networks for Document Image Classification," *Proceeding of 22nd International Conference on Pattern Recognition*, pp. 3168-3172, 2014.

[10] J. Choi, L. Lee, Y. Chung, and D. Park, "Individual Pig Detection Using Fast Region-based Convolution Neural Network," *Journal of Korea Multimedia Society*, Vol. 20, No. 2, pp. 216-224, 2017.

김 영 수

1994년 공군사관학교 이학사
2001년 서강대학교 공학석사
2009년 한국과학기술원 공학박사
현재 공군사관학교 전산정보학과 교수
관심분야 : 인공지능, 영상처리, 컴퓨터 비전 등이다.