전이학습을 활용한 Image CAPTCHA 해독 프로그램 구현

정민혁, 최진우
경희대학교 컴퓨터공학과
Jeongmh09@naver.com , jinwoochoi@khu.ac.kr

Image CAPTCHA decoder implementation using transfer learning

Minhyeok Jeong, Jinwoo Choi
Department of Computer Science and Engineering, KyungHee University

요 약

CAPTCHA(Completely Automated Public Turing test to tell Computers and Human Apart)는 사용자를 사람인지 컴퓨터 프로그램인지 구분하기 위해 사용되는 방법의 통칭으로, 다양한 방면으로 활용되지만 주로 로그인이나 회원가입 등에 하나의 절차로 사용된다. Image CAPTCHA는 그리드로 나뉜 하나의 사진 혹은 여러 장의 사진에서 특정 대상이 존재하는 사진을 모두 고르도록 요구하는 CAPTCHA로 Google이 제공하는 reCAPTCHA가 대표적 예시이다. 이에 본 논문에서는 전이학습을 통해 Image CAPCHA에 주로 등장하는 물체들을 학습시킨 모델을 설계하여 Image CAPCHAT를 해독하는 어플리케이션을 설계하고자 한다.

1. 서 론

CAPTCHA(Completely Automated Public Turing test to tell Computers and Human Apart)는 튜링 테스트의 한 종류로, 사용자와 컴퓨터 프로그램을 구분하기 위한 용도로 사용되는 방법의 통칭이다. 튜링 테스트는 1950년 앨런 튜링(Alan Turing)에 의해 제시된 개념으로, 인간이 컴퓨터와 인간을 구분하는데 있어서 기계가 인간이 할 수 있는 것을 할 수 있는지 그 능력을 검증하는 테스트이다. 튜링 테스트의 주체는 인간 평가자인 것에 반해, CAPTCHA는 기계가 컴퓨터와 인간 사용자를 구별한다는 점에서 CAPTCHA는 역 튜링 테스트(Reverse Turing Test)라고 부르기도 한다.

컴퓨터 프로그램을 구분하기 위한 CAPTCHA를 되려 컴퓨터 프로그램으로 해독하는 것은 역설적으로 들릴 수 있으나, AI를 이용하여 CAPTCHA를 풀려는 시도는 과거부터 존재해왔다. Ahn, Blum and Langford의 논문에서 저자들은 AI로 CAPTCHA를 풀지 못하면 컴퓨터와 사람을 구분할 방법이 존재한다는 것이며, 풀었을 경우에는 다른 복잡한 인식 문제에 적용 가능한 AI의 발전을 의미하는 윈-윈 상황을 가져온다고 주장한바 있다[1].

가장 많이 볼 수 있는 CAPTCHA의 형태로는 이미지에 적힌 텍스트를 읽고 그대로 입력하거나 해당 질문의 답을 기입하는 텍스트 CAPTCHA나, 물체를 포함한 사진을 활용한 Image CAPTCHA가 있다.

시간이 흐를수록 머신 러닝과 OCR 프로그램의 발전으로 기

계들도 높은 수준의 판독률을 보이게 되면서 낮은 복잡도의 CAPTCHA는 개선이나 다른 보안 방식으로 대체됨이 요구된다. 따라서 본 논문에서는 기계학습을 이용하여 가장 많이 사용되는 Image CAPTCHA인 reCAPTCHA V2의 해독 모델을 설계하여 사용자들에게 기존 Image CAPTCHA가 안전하지 않음을 보이고자 한다.

2. 관련연구

2.1 reCAPTCHA

CAPTCHA의 한 종류인 reCAPTCHA는 Luis von Ahn, David Abraham, Manuel Blum, Michael Crawford, Ben Maurer, Colin McMillen, and Edison Tan에 의해 개발되었고, 2009년부터 구글에 소유되어 무료로 사용가능한 서비스이다. V2는 Image CAPTCHA의 형태를 띄며, 그리드 상에서 문제에서 제시하는 물체를 포함한 이미지를 찾아 클릭하는 방식이다. 인증과정에서 생기는 사용자의 불편함을 없애기 위해 사용자의 행동에 점수를 매기고 이를 근거로 작동하는 reCAPTCHA V3까지 출시된 상태이나, V2가 여전히 많이 사용되고 있다.

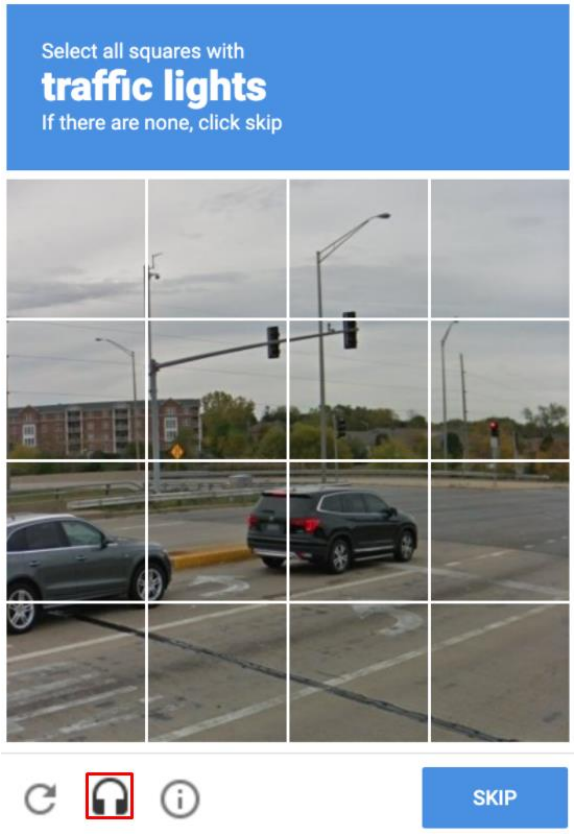

그림 1 reCAPTCHA V2의 예

2.2. 기계학습

기계학습(머신 러닝)은 컴퓨터과학의 한 분야로 경험을 통해 학습 또는 발전하는 알고리즘을 연구한다. AI연구의 한 분야로 간주되며, 모델을 구축하여 샘플 데이터를 통해 결과를 얻고, 이 결과를 이용하여 모델을 다시 수정하는 과정을 반복하면서 입력 값에 대해 원하는 결과를 출력하도록 수정하는 과정을 거친다. Tom M. Mitchell은 "어떤 경험이 작업의 평가를 더 좋게 만들면 그 경험을 통해 프로그램이 학습했다고 한다" 라고 컴퓨터의 학습을 정의하였다[2]. ("A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E.")

기계학습은 가진 데이터를 이용하여 통계학적 모델을 학습시켜 인공지능을 구현하는 것으로, 이에 사용되는 모델은 사람의 신경망을 모방하는 형태로 구축된다. 모델을 이루는 네트워크 레이어의 개수가 많은 경우 이 방법론을 딥 러닝이라고 부른다. 영상을 활용하는 기계학습에 자주 사용되는 네트워크 구조로 CNN이 있고, 본 논문에서도 CNN을 활용하여 모델을 설계한다.

2.2.1 CNN[3]

CNN은 딥 러닝에 주로 사용되는 네트워크 구조의 한 종류로, 커널의 공유되는 weight를 통한 convolution 연산을 거쳐, 모델에 있는 모든 노드들이 완전 연결되어 있는 것이 아닌 부분연결 구조를 가지는 모델의 복잡도가 낮은 모델이다. 기본적인 구조로 입력 계층과 컨볼루션 은닉 계층, 완전 연결 은닉 계층, 출력 계층의 형태로 네트워크가 구성되어 있다.

Convolution연산은 본래 영상처리나 신호처리 분야에서 많이 사용되었고, 따라서 CNN역시 영상과 관련된 작업에 강점을 보이며, Image Classification, Image Segmentation, Video Recognition등과 같은 연구 분야에 많이 사용된다.

2.2.2 전이학습

전이학습(Transfer learning)은 어떤 문제를 해결하기 위해 학

습하면서 배운 지식을 다른 문제에도 적용시킬 수 있을 것이라는 생각에서 비롯된 학습 방법의 한 갈래이다. 처음부터 하나하나 모든 무작위로 초기화한 상태에서 시작하여 학습하는 것이 아니라 이전에 잘 학습된 모델을 초기 설정으로 새로 적용하고자 하는 대상에 대해 학습을 진행하는 것으로 이 과정을 미세조정(fine-tuning)이라고 한다. 이는 데이터가 충분히 많은 양이 갖추어지지 않은 경우에도 효과적으로 학습을 할 수 있게 해주기 때문에 본 설계에서도 ImageNet 데이터셋으로 미리 학습된 신경망을 사용하여 미세조정을 거쳐서 모델을 구축한다.

3. 프로젝트 설계

3.1 Data Set

3.1.1 Training / Validation / Test Set

Image CAPTCHA에서 나오는 물체 찾기에는 다양한 유형이 있으나 주로 많이 출제되는 유형으로 도로/교통 관련 이미지를 사용하여 관련 물체를 찾게 하는 것이 많이 사용된다. 따라서 본 프로젝트에서는 학습할 클래스 범위를 Ambulance, Bus, Fire hydrant, Parking meter, Stop Sign, Traffic light, Vehicle registration plate, Street light, Bicycle, Limousine, House, Boat, Traffic sign, Motorcycle, Truck, Unicycle, Taxi, Vehicle, Wheel, Helmet, Helicopter, Car, Airplane, Aircraft로 우선 제한하여 학습을 진행한다.

학습에 필요한 이미지들은 구글에서 제공하는 Open Images V6에서 각 클래스 별로 200장씩 구하고, Training/Validation/Test 각각 7:1:2 비율로 나눈다.

3.1.2 CAPTCHA Images

reCAPTCHA 데이터셋은 대중에 공개된 것이 없기 때문에 직접 수집하거나 임의로 reCAPTCHA에서 사용되는 형식으로 제작할 필요가 있다. 그림 1에 나온 것처럼, 각 이미지는 4x4의 그리드 형식으로 영역이 나누어져 있고, 전체 이미지가 16구역으로 나뉘거나, 각 구역이 다른 이미지로 구성되어 있다. 따라서 Street View 이미지를 웹 크롤링을 통해 수집하고, 이 이미지를 바탕으로 사진을 재구성하는 프로그램을 작성한다. 준비한 이미지는 {target}_{ground truth}_{number}의 이름 형식으로 저장한다.

3.2 모델

Image Classification은 이미 연구가 많이 진행되어 있고, 높은 정확도를 가진 모델이 많이 나와있다. ResNet은 ILSVRC(ImageNet Large Scale Visual Recognition Challenge)에서 2015년 우승을 거머쥔 우수한 성능의 모델이다. 여기에 기존 ResNet이 학습하지 않은 클래스에 대해 미세조정(fine tuning)을 하는 전이학습을 진행하여 모델을 구축한다.

3.3 해독 알고리즘

메인 프로그램은 준비한 reCAPTCHA 이미지를 입력으로 넣으면, 4x4 각 영역에 속한 이미지를 앞서 학습한 모델을 통해 image classification 을 진행한다. 다음 표에 나온 순서대로 영역을 구분하고, 각 영역에서 나온 결과를 종합하여 어떤 영역에서 문제에서 요구한 물체가 나왔는지 출력한다.

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

표1. 그리드 영역 구분 번호

5. 참고문헌

1. von Ahn, Luis; Blum, Manuel; Hopper, Nicholas J.; Langford, John (May 2003). "CAPTCHA: Using Hard AI Problems for Security". Advances in Cryptology — EUROCRYPT 2003. EUROCRYPT 2003: International Conference on the Theory and Applications of Cryptographic Techniques. Lecture Notes in Computer Science. Vol. 2656. pp. 294–311.

2. Mitchell, T. M., & Mitchell, T. M. (1997). *Machine learning* (Vol. 1, No. 9). New York: McGraw-hill.

3. 오일석, (2017). 기계학습 pp. 23-25, 194-200, 225-229.

4. HOSSEN, Md Imran, et al. An Object Detection based Solver for {Google's} Image {reCAPTCHA} v2. In: *23rd international symposium on research in attacks, intrusions and defenses (RAID 2020)*. 2020. p. 269-284.