코딩교육 열풍과 현주소 - 7

매의 눈, 인공지능의 이미지 처리

글. 김승욱 Rloha 대표, 데이터 분석 교육 및 컨설팅
'빅데이터 분석, R좀 R려줘', 'R 데이터 분석' 등 관련 칼럼 및 강의 진행

언제부턴가 스마트폰은 사람의 얼굴을 인식하기 시작했다. 예를 들면 [그림 1]과 같이 사람의 얼굴에 네모 상자를 그려주고, 사용자는 해당 네모에 터치를 하면 얼굴을 기준으로 초점 조절을 하여 사용자로 하여금 보다 만족스러운 사진을 찍을 수 있도록 도와준다는 것이다. 더 나아가서 요즘 사진 애플리케이션에는 얼굴을 인식하는 것 이상의 기능을 제공한다. 우선 얼굴 인식을 하고 나면 모자를 씌우거나 독특한 효과를 적용하는 기능을 말하며, 독자들도 한 번 즈음은 경험해보았을 것이다.

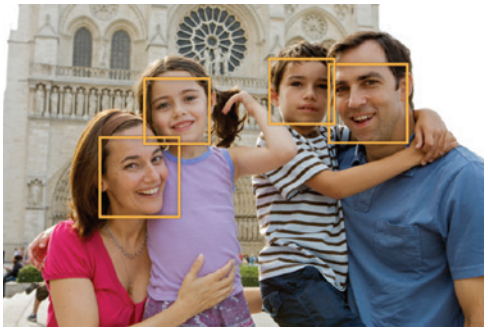

그림 1. 이미지 얼굴 인식 예시

그림 2. 사진 보정 애플리케이션 예시

그러면 인공지능은 어떻게 이미지를 인식하는 것일까? 우선 이미지는 어떻게 이루어져 있는지 알아야 하겠다. 기본적으로 이미지는 숫자로 이루어져 있다. 좀 더 구체적으로 말하자면 이미지는 픽셀(pixel, 화소)마다 색상 정보를 담고 있는데 그 정보가 숫자로 되어 있다는 것이다. 이를 조금 더 크게 본다면 이미지는 숫자의 조합이며, 우리는 이를 별도의 변환을 통해 알록달록한 색상을 볼 수 있고,

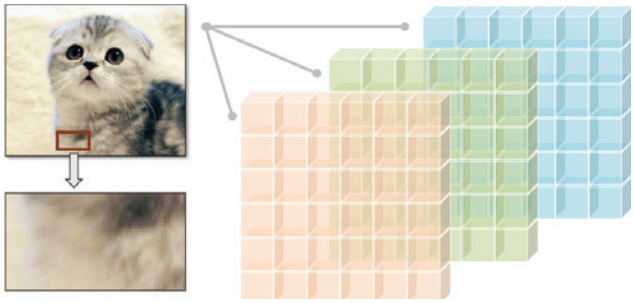

그림 3. 이미지 데이터 구조 예시

이미지에 사람이 있는지 고양이가 있는지 알아볼 수 있다. 이를 도식화한 것이 다음 [그림 3]의 예시가 되겠다.

그림 4. 이미지 엣지 추출 예시

각 픽셀에는 3개의 숫자(RGB, 빨강/초록/파랑)가 들어있는 것이 기본이며 자료구조의 관점에서 본다면, 같은 크기의 2차원 행렬 3개가 적층 되어있는 것과 같다고 할 수 있다. 즉, 카메라로 찍은 사진이 1,000만 화소라고 한다면, 해당 사진 파일에는 약 3,000만 개의 숫자가 들어있다고 할 수 있다. 기술의 발전에 따라 카메라에 사용되는 부품의 성능이 좋아져서 촬영하는 사진이나 영상의 해상도가 높은데 10년 전의 스마트폰과 지금의 스마트폰으로 찍은 사진의 용량을 비교해보면 확실히 체감이 되지 않을까 한다. 이렇게 이미지를 처리하기 위해서는 이미지의 숫자를 얼마나 잘 이해하고 처리하느냐가 관건이다. 고전적인 이미지 처리는 사람이 직접 이미지의 색상이 급격하게 바뀌는 엣지(edge)[1]를 찾아내고 그것을 기반으로 이미지 분류를 하는 것이 고작이었다.

그런데 딥러닝을 도입하게 되면서 엣지 기반 이미지 특징 추출을 인공지능에 맡겼더니 놀라운 성능 향상을 확인할 수 있었다. 여기서 핵심이 되는 기술이 바로 CNN(Convolutional Neural Network)이다.

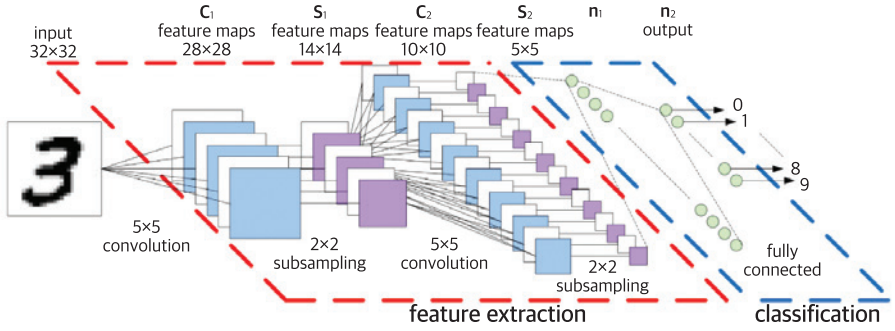

그림 5. 숫자 이미지 인식 CNN 구조 예시

앞에서 언급했던 신경망 기반 특징 추출이 합성곱(컨볼루션, Convolution) 연산을 통해서 이루어지는데 특징 추출이라고 해서 대단한 기술이 들어가는 것이 아니라 단순히 특정 원소(이미지 데이터로 따지면 픽셀) 주변의 숫자를 더하거나 최댓값을 뽑는 단순한 사칙연산을 하는 것이다.

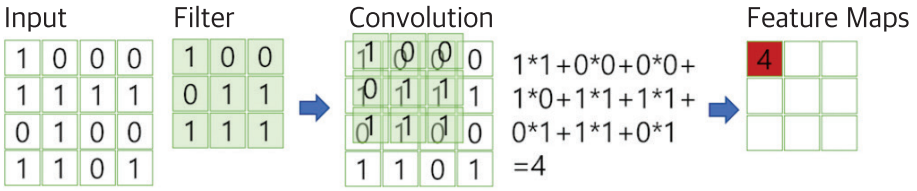

그림 6. 합성곱 연산 수행 예제

1) 정확하게는 엣지를 기반으로 한 필터 뱅크이다. 이는 필터 기반의 특징 공간(Feature Space)의 근간이 되며 특징 공간의 정보를 취합하여 최종적으로 이미지 분류를 하게 된다.

이렇게 수학 얘기만 하면 딱히 와닿지 않을 수 있다. 그렇다면 다음 QR 코드를 타고 가서 영상을 꼭 보는 것을 추천한다.

해당 영상에서는 강아지 사진을 수동 세절기로 가로 세로로 분할한다. 그런데 강아지 사진의 일부를 다시 이어 붙여도 얼추 강아지라는 것을 알아볼 수 있다.

그림 7. 사진으로 알아보는 합성곱 기반 특징 추출 예시

이렇게 이미지를 분류하는 기술은 CNN의 등장으로 성능이 비약적으로 발전하게 된다. 대표적인 이미지 분류 대회인 ILSVRC에서 처음으로 CNN이 사용된 2012년도 우승팀의 인공지능 모델은 11년도 모델의 오분류율 25.8%에서 16.4%까지 낮아졌다. 3년 후인 15년도 우승팀 모델의 오분류율은 3.57%를 기록했는데 이는 사람의 오분류율로 알려진 5%보다 낮은 수치이다.

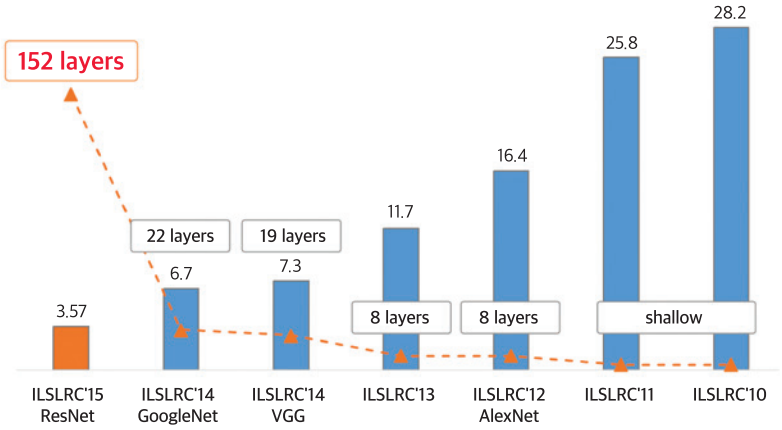

그림 8. ILSVRC 대회 역대 모델 성능

여기까지 보면 이미지 처리가 꽤 쉬워 보일 수 있지만 결코 그렇지 않다. 이미지 처리 또한 자연어 처리만큼 많은 난관에 부딪힌다. 관련하여 유명한 예시가 있는데 검색엔진 구글에서 머핀을 검색하면 치와와와 머핀이 섞여 나오기도 했다.

그림 9. 예전 구글 이미지(머핀) 검색 결과

성능이 좋지 않은 모델은 이와 같이 머핀에 박혀있는 초코와 치와와의 눈코입을 같은 것으로 인식하고 처리한다. 이렇게 사물과 동물이면 제대로 분류를 못 해도 큰 문제는 없지만, 흑인 사진을 고릴라로 분류한 사고사례는 구글 입장에서 꽤 난처한 문제였다. 이는 충분한 성능을 갖추지 못한 인공지능을 사용했을 때 발생할 수 있는 문제점이라고 볼 수 있겠다.

자연어에서의 오탈자 처리와 같이 이미지 처리에서도 전처리(preprocessing) 작업이 필요하다. 이미지를 일정 규격으로 자르거나, 노이즈가 있는 이미지를 처리하는 등 다양한 작업이 필요한데 이런 전처리 작업을 담당하는 인공지능도 연구되고 있다. 대표적으로 저품질 이미지를 고품질 이미지로 바꾸는 것인데, 잡티를 제거(denoising)하거나 고화질 변환(Super-Resolution 또는 upscaling)이 있다.

고화질 변환의 경우 waifu2x[2] 라는 사이트에서 무료로 제공하고 있으며 해상도 향상은 최대 2배까지 지원한다. 예를 들어 EBS 로고(200×100픽셀)를 가지고 실험한 결과를 다음 그림에서 볼 수 있는데 도형과 글자의 테두리를 보면 보다 선명해진 것을 알 수 있다.

그림 10. waifu2x 사이트를 통해 얻은 고품질 이미지(좌)와 원본(우)

다음으로 이미지 전처리 이후의 단계를 알아보자. 이미지에 고양이가 있는지 강아지가 있는지 단순히 확인하고 분류(classification)를 하는 것이 첫 번째 일이다. 다음은 이미지 어느 부분에 고양이가 있는지 대략 파악[3](localization)을 하는 것이 그다음 기술이 되겠고, 그 이후로는 고양이가 몇 마리나 있는지, 혹시나 다른 사물과 같이 있지는 않은지 알아보는 기술 등 점점 고차원의 기술을 사용하게 된다.

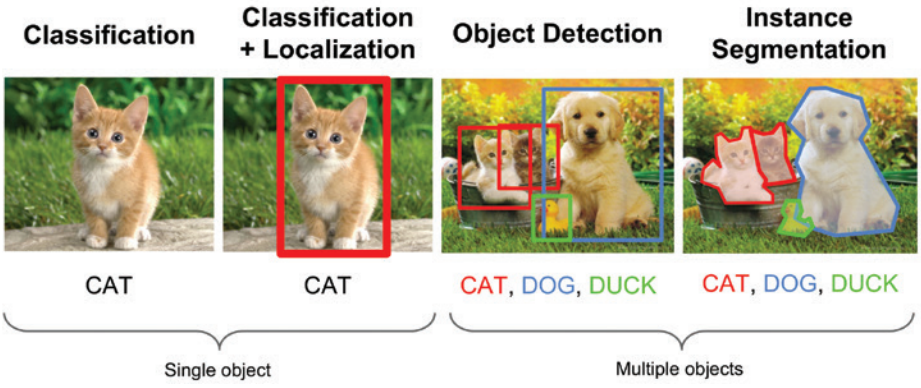

그림 11. 이미지 인식 종류와 개념 (https://light-tree.tistory.com/75)

성능이 좋은 인공지능을 만들기 위해서는 다양한 방법이 있다. 대표적으로 세 가지를 꼽자면, 앞에서 소개한 전처리. 보다 고도화된 신경망 구조 사용. 마지막으로 정해진 신경망 구조에서 각종 설정을 조금씩 바꿔보면서 최적의 결과를 도출하기 위해 다양한 시도를 하는 방법[4]도 있다. 그런데 이미지 처리용 인공지능의 경우 방법이 하나 더 있다. 바로 Data Augmentation이다.

Data Augmentation은 데이터 증폭이라고 하며, 한정된 이미지 데이터를 다양한 조건으로 변형시켜 마치 매우 많은 데이터가 있는 것처럼 데이터 개수를 부풀리는 방법이다. 일단 인공지능은 데이터가 많으면 학습이 대체로 잘 되는 편이니 특히나 데이터가 모자란 경우 데이터 증폭을 하는 것을 권장하며, 이렇게 생성된 데이터를 인공지능 학습에

2) http://waifu2x.udp.jp
3) 정확하게는 해당 사물을 감싸는 네모 상자를 그릴 수 있는 네 꼭지점의 좌표 산출
4) 정확하게는 하이퍼파라미터 튜닝(Hyperparameter Tuning) 또는 하이퍼파라미터 최적화(Hyperparameter Optimization)라고 한다.

그림 12. 영화 기생충 포스터 이미지의 데이터 증폭 예시

활용할 경우 기존 데이터로만 학습했을 때 보다 성능을 많이 올릴 수 있다는 연구 결과가 많으니 관련 논문[5]을 참고하도록 하자.

좀 더 쉽게 말해서, 인공지능에 정방향의 기생충 영화 포스터를 보여주고 학습을 시킨다면 일단 인공지능은 어떤 그림이 기생충 포스터인지 알게 된다. 그런데 사람은 기생충 포스터를 뒤집어 놓아도, 일부를 찢어버려도, 색이 바래도 기생충 포스터인지 아닌지 단박에 알아차린다. 인공지능도 이렇게 사람처럼 똑똑하게 하려면 단순히 정방향의 포스터 이미지로 학습을 하는 것이 아니라 이리저리 돌리고 뒤집고 색칠한 이미지도 넣어줘야 심지어 좀 찢어지고 구겨진 포스터라도 어느 정도 인공지능이 인식을 할 수 있게 된다. 조금 과장을 보태자면, 정방향 이미지로 학습한 인공지능은 이미지를 1°만 틀어버려도 다른 이미지로 인식하기도 한다.

이 정도 되면 어떤 사람들은 이런 기술이 좋은 것은 알겠는데 도대체 어디에 쓸 수 있는 것인지 의문을 가지기도 한다. 보통 인공지능은 우선으로 반복적이고 지루한 인간의 업무를 대체하는 경우가 많은데, 가끔은 장애인을 위한 기술로도 응용이 된다. 그럼 관련 응용 사례를 이미지 처리 분야와 영상처리 분야를 나눠서 알아보자.

이미지의 경우는 X-ray 촬영 사진을 해석하여 골절 부위나 기타 병변을 찾아낼 수 있으며, 안구 사진을 통해서 검안자의 당뇨병 여부를 진단하기도 한다. 앞에서 언급했던 카메라의 얼굴 인식은 물론이고 더 나아가서 Face ID같이 사용자의 얼굴을 보다 상세하게 기록하여 본인 인증의 용도로 사용하기도 한다. 사실 이미지 처리로 의외의 혜택을 받는 사람은 시각장애인이다. 시각장애인의 경우(전맹 기준)는 인터넷이나 Facebook 같은 SNS를 사용하면서 밖으로 드러나 있지 않은 이미지의 추가 설명문인 대체 텍스트[6]에 의존해서 세상을 읽는다. 여기서 보는 것이 아니라 읽는 것인데 보통 이 대체 텍스트에는 글자가 전혀 없거나 있더라도 단순히 '사진' 정도만 적혀 있는 경우가 많다. 그런데 페이스북의 경우 다음과 같이 표기해주고 있다. 아주 자세하게 표현되어 있지는 않지만, 조금이나마 정보를 담고 있다는 사실이 그나마 다행이랄까?

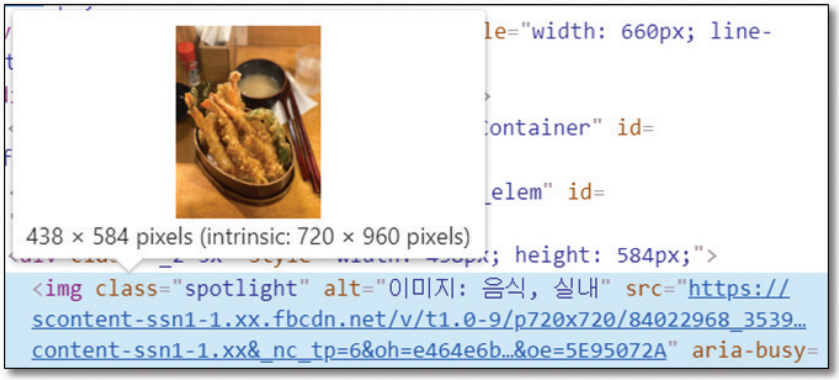

그림 13. Facebook의 인공지능 기반 이미지 설명 자동 태깅 예시

다음은 영상을 알아보자. 기왕 시각장애인을 위한 기술이 나온 김에 비슷한 사례를 소개하자면 수화를 들 수 있다. 수화는 보통 청각장애인을 위해서 사용하는 대화 수단이지만, 수화를 잘 모르는 사람이나 시각장애인의 경우 수화를 빠르게 이해할 방법이 없다. 그래서 이 수화를 영상으로 인식하여 각 수화에 맞는 단어나 문장으로 바꿔주는 기술도 있다. 물론 그 이전에 수화를 사용하는 특수 장갑을 끼면 해당 장갑에서 나오는 데이터를 인공지능 모델이 분석하여 스피커로 문장을 말하는 사례도 있지만 개인적으로 가장 마음에 드는 사례는 영상으로 수화를 인식하여 인공지능 스피커에 명령을 내리는 것이 아닐까 한다. 해당 영상은 다음 그림의 QR 코드로 확인할 수 있다.

5) https://arxiv.org/pdf/1708.04896.pdf, https://arxiv.org/pdf/1805.09501.pdf 이외에도 다양한 논문이 있다.
6) HTML img 태그의 alt 속성에 입력되는 값

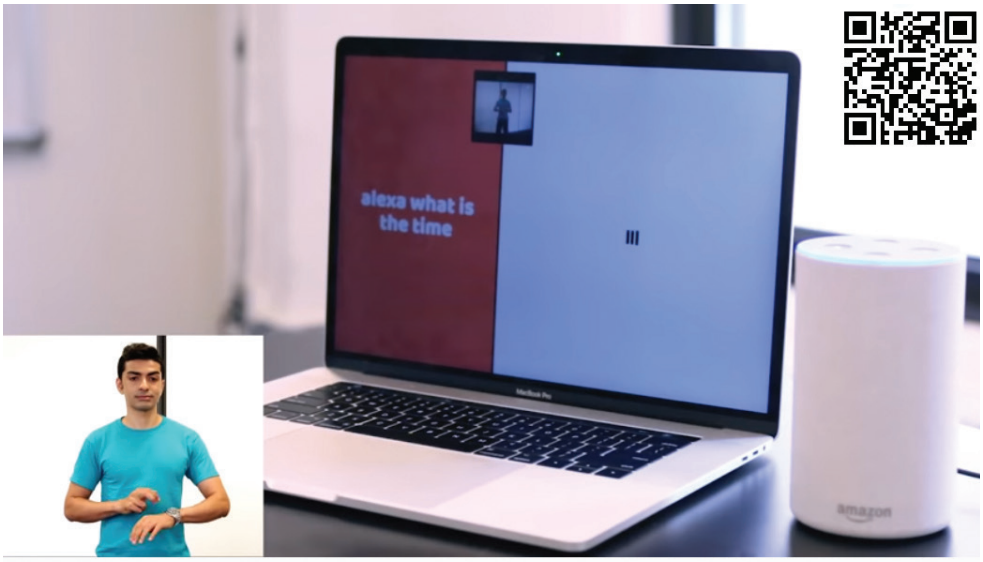

그림 14. 수화 인식 사례와 QR코드

영상이라고 하면 요즘 굉장히 화제가 되는 자율주행 자동차를 빼놓을 수 없다. 자동차의 자율주행에는 다양한 기술이 사용되긴 하지만 자율주행 자동차의 시야를 담당하는 영상처리를 빼놓을 수 없기 때문이다. 영상 속 이미지 인식으로 어떤 부분이 차선인지 차량인지 사람인지 제대로 분간을 할 수 있어야 운행을 할 수 있기 때문이다.

그림 15. 자율주행 영상 처리 예시

실제 자율주행 구현을 위해서는 도로의 표지판과 신호등까지 전부 해석하고 운행을 해야 하기에 단순히 차선과 차량을 구분하는 것만으로 해결되는 문제는 아니다. 관련 기술과 내용은 너무나 방대하기에 자율주행의 초기 단계부터 어떻게 기술이 발전되어 가는지 잘 보여주는 '기계는 어떻게 생각하는가?'라는 책을 추천하고자 한다. 내용이 조금 어려울 수 있겠지만 어떤 기술이 적용되고 그에 따른 성능이 어떻게 향상되었는지 현대 인공지능을 훑어볼 수 있는 재미있는 책이라고 할 수 있다.

아마 앞에서 소개한 책을 읽게 되면 독자분들은 자율주행의 근간이 되는 기술이 이미지/영상 처리가 아니라는 것을 깨달을 것이다. 자율주행의 핵심은 강화학습이라는 것인데 이 강화학습은 어떤 것이고, 어디에 활용되는지 보다 자세한 내용은 다음 호에서 알아보고자 한다.