Watson visual recognition이 어떻게 작동을 하는 지 간략하게 이야기 하겠습니다

다른 이미지 분석 알고리즘과 비슷하게 왓슨 역시 CNN을 사용하였습니다.

하나의 이미지를 따와서 그 값을 convolution layer과 같은 여러가지 은닉층들 여느 신경망 알고리즘과 매우 유사한 구조인데 CNN은 convolution layer와 max-pooling layer로 구성됨을 확인할 수 있고 convolution layer를 통해 넘긴 값을 Relu창을 통해, Relu라는 것은 이상치를 없애주는 활성화함수를 의미하구요 Max pooling하는 결과를 거쳐서 하나의 값들로 수렴하게 되면 그 값들을 labeling 해줍니다. Fully Connected Neural Network라는 것은 그 과정을 의미하는데 가장 정확도가 높은 값을 산출해주는 과정을 의미합니다. 작동방식을 간단하게 보여드리면, filter가 이런식으로 searching을 하게 되면, 하나 하나 searching을 할 때마다 특정한 값을 산출합니다. 그 값들을 이용해서 전체 searching을 할 때마다, 사용자가 지정한 임의의 가중치를 부여하여 결과를 산출하면은 이러한 하나의 convolution layer의 값을 얻게 됩니다. 구체적인 작동 방식은 이러한 3x3 input 이미지가 주어지고 1로 구성된 필터값을 적용하면 실제로 결과 값은 6이고 해당 필터가 stride 1로 한칸씩 넘어가는 식으로 계산이 되면 6, 8, 9, 6으로 값을 구성합니다. 이 값을 또한 Max - pooling으로 넘기게 되면, 예를 들면 4x4 이미지가 주어졌다고 했을 때, 2x2 filter, stride=2로 분석을 하면 가장 최대값을 산출하여 결과적으로 6, 8, 3, 5를 가져오게 됩니다. 이러한 여러가지 과정을 거쳐서 값을 산출해주면 그 값이 어디에 labeling이 되는 지를 확인해주는 알고리즘입니다.

실제로 왓슨 같은 경우도 그러한 방식을 이용하여 구성을 했는데 이미지가 주어졌을 때 그 이미지를 쪼개고 쪼개서 하나의 값을 산출해줍니다. 여기까지는 사실 일반적인 cNN구조랑 되게 비슷한 구조인데 이 값에 가중치를 부여해서 왓슨만의 Semantics – aware, 의미를 매칭시켜서 가장 맞는 것을 산출해주는 결과입니다. 예를 들면 기본적으로 이미지가 주어졌을 때, 일반적인 cNN은 가장 의미가 있는 그 확률을 산출해주는 건데 의미를 고려한 가중치를 부여하여 가설 검정을 다시 합니다. 그래서 좀 더 확률을 개선시키는, 이 파란색 확률이 이제 개선된 확률을 의미하는데요 이런 식으로 좀 더 개선된 값을 내줍니다. 이 FT의 경우 이러한 가중치를 부여하지 않은 결과인데, 가중치를 부여한 HL의 결과는 군집화가 조금 더 깔끔하게 됨을 확인할 수 있습니다. 이러한 확률 값을 이용해서 이 사과의 이미지가 주어졌을 때 그 분석한 확률들을 보여줍니다. 가장 매칭이 잘 된 것을 보여줘서 예를 들면, fruit, apple같은 가장 확률이 높은 후보군을 산출해줍니다. 피자라는 것은 정확히 분석을 했는데 페파로니인지 소시지인지 이러한 분석을 해주는 방식으로 작동을 합니다.