

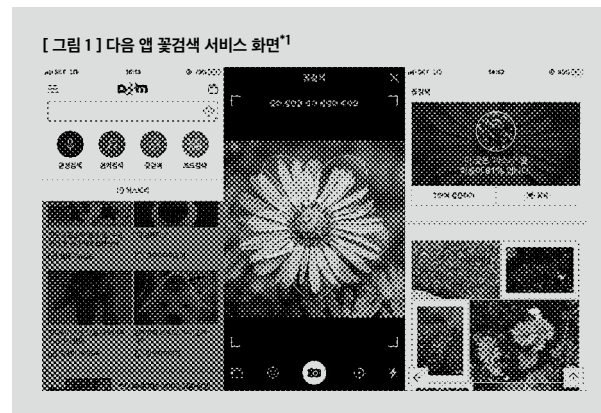
# 꽃이름이 궁금해? 다음 앱 시에 물어봐!

때늦은 꽃샘추위가 지나가고 완연한 봄이 도래한 듯 전국 곳곳에서 설레는 봄꽃 축제 소식이 들려왔다. 길을 걷다 주위를 둘러보면 민들레, 철쭉 등 익숙한 꽃도 있지만 이름 모를 예쁜 꽃이 한 가득이다. 하지만 이제 이들의 정체를 알기 위해 더 이상 어르신들에게 물어보거나 꽃 도감을 찾아볼 필요가 없게 됐다. 카카오에서 다음 앱을 통해 꽃 사진을 찍으면 실시간으로 꽃의 이름을 알려주는 꽃검색 서비스를 제공하고 있기 때문이다.

## 꽃검색

[그림 1]은 다음 앱 '꽃검색'의 실제 실행 화면을 보여준다.

꽃검색 서비스는 약 5년 전부터 사내에서 기획 및 연구되어 온 기술이다. 연구 초기에는 전통적인 영상 인식 기술로 개발되어 그 성능이 만족스럽지 않아 실제 서비스로 제공되지 못했다. 그러나 딥러닝(deep learning) 기술이 약 2년 전부터 비약적으로 발전하면서 이를 꽃검색 서비스에 적용하게 되었고, 이를 통해 성능이 대폭 향상되면서 정식 서비스로 출시하게 된 것이다. 이번 글에서는 초기 꽃검색의 알고리즘부터 딥러닝 기술을 적용한 현재의 꽃검색 알고리즘까지의 발전 과정을 살펴보고자 한다. 아울러 지속적으로 발전하는 최신 딥러닝 기술을 활용한 꽃검색 서비스의 미래에 대해서도 이야기해보도록 하겠다.



## 딥러닝 이전의 꽃검색 알고리즘

꽃검색 서비스는 쉽게 말하자면 꽃 품종을 분류하는 서비스다. 즉, 주어진 사진에 존재하는 꽃이 어떤 품종에 속하는지를 분류해주는 기술인 것이다. 최근 딥러닝 기술이 발전하기 이전에는 꽃검색 서비스 개발을 위해 전통적인 영상 인식 기술이 이용되었는데, 이는 크게 두 단계로 구성된다. 첫 번째 단계는 사진에서 꽃 영역을 찾아내는 분할(segmentation) 과정이다. 사용자가 꽃 사진을 찍으면 그 중 어느 부분이 꽃에 해당하는지 구분해주는 것이다. 두 번째 단계는 분할된 영역을 이용해 해당 꽃이 어떤 품종에 속하는지 알아내는 분류(classification) 과정이다.

첫 번째 단계인 꽃 분할을 위해선 우선 입력된 사진을 슈퍼 픽셀(superpixel) 단위로 나눈 후, 각 슈퍼 픽셀이 전경(foreground)에 속하는지 배경(background)에 속하는지를 구별한다. 이후 그래프컷(Graph Cut) 기술을 적용해 사진에서 꽃에 해당하는 영역을 분할한다.

두 번째 단계인 꽃 품종 분류 과정에서는 BOF(Bag of Features) 방법이 사용된다. 먼저 분할된 꽃 영역에서 다양한

특징(feature)을 뽑은 후 이 특징들을 클러스터링(clustering)해 얻은 대표 특징을 모아 놓은 코드북(codebook)을 생성한다. 이렇게 하면 입력된 각각의 꽃 사진은 코드북에 있는 코드단어(codeword)들의 조합으로 표현될 수 있다. 수많은 꽃 사진의 데이터베이스로부터 계산된 조합을 이용해 SVM(Support Vector Machine) 분류기를 학습시키면 이후 주어진 꽃의 사진이 어떤 품종에 속하는지 분류해낼 수 있는 것이다.

하지만, 위와 같이 전통적인 영상 인식 기술을 이용해 개발된 꽃검색 서비스는 사용자에게 제공하기에 성능이 충분치 않았다. 또한 적은 수의 공용 꽃 데이터베이스를 이용해 개발되었기 때문에 알고리즘이 분류할 수 있는 꽃의 품종이 102종 밖에 되지 않아 다양한 꽃의 품종을 검색하는 것 역시 쉽지 않았다.

## 딥러닝 기술의 등장

CNN(Convolutional Neural Networks)으로 대표되는 딥러닝 기술의 등장과 대규모 사진 데이터 수집이 가능하게 되면서 여러 영상 인식 및 검색 분야에서 큰 성능 향상이 이루어졌다. 특히, 2012년 ILSVRC 대회(ImageNet Large Scale Visual Recognition Challenge)에서 사진 분류 문제에 대해 CNN을 바탕으로 한 알렉스넷(AlexNet)<sup>2</sup>이 압도적인 격차로 우승을 하면서 딥러닝 기술이 본격적으로 주목받기 시작했다.

딥러닝 열풍의 선두주자인 알렉스넷은 이후 제안된 딥러닝 네트워크들의 초석이 된 모델로, 기울기 소실(vanishing gradient) 문제를 해결하기 위해 활성화 함수로 시그모이드(sigmoid)나 쌍곡 탄젠트(hyperbolic tangent) 대신 ReLU(Rectified Linear Unit)를 사용하였다. 특히 과적합(overfitting)을 피하기 위해 드롭아웃(dropout), 데이터 증대(data augmentation) 기법을 사용하는 등 현재는 당연히 여겨지는 기법들이 대부분 이때 도입되었다. 그 후 2014년에는 VGG넷(VGGNet)<sup>3</sup>이 등장해 더욱 뛰어난 성능 향상을 보였다. 알렉스넷은 컨볼루션 층(convolution layer)에서 큰 필터(11x11)를 사용한다. 비해 VGG넷은 모든 레이어에서 가장 작은 필터인 3x3을 사용하였다. 작은 크기의 컨볼루션 필터는 네트워크의 깊이가 증가함에 따라 비선형성(non-linearity)이 증가하여 더욱 복잡한 특징을 추출할 수 있게 해주고 상대적으로 파라미터의 개수가 줄어들어 네트워크 정규화(regularization)에 도움이 되는 장점을 지니고 있다.

하지만 2014년 ILSVRC 대회에서 우승을 차지한 모델은 구글넷(GoogleNet)<sup>4</sup>이다. 일반적으로 네트워크가 깊을수록(deeper) 그리고 레이어가 넓을수록(wider) 성능이 향상되지만, 실제로는 무작정 네트워크의 깊이나 넓이를 늘릴 수

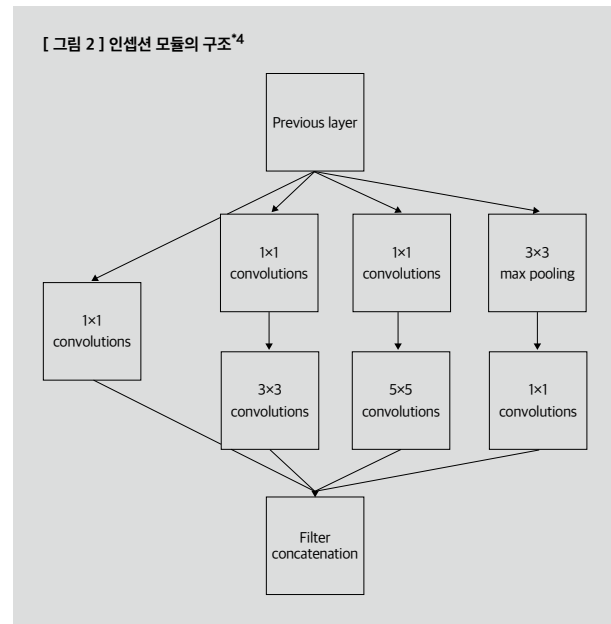
글 | 변민우 dylan.byeon@kakaocorp.com

꽃검색 서비스를 개발하면서 머신러닝에 관심을 갖기 시작하였으며, 현재는 비주얼 서치 관련 서비스를 개발하고 있습니다. 알고리즘과 머신러닝에 관심이 많으며 TopCoder, Kaggle과 같은 대회 플랫폼에서 다른 개발자들과 경쟁하며 배우는 것 또한 좋아합니다.

글 | 홍은빈 erin.hong@kakaocorp.com

딥러닝 기반 영상처리를 전공한 후 현재 카카오 멀티미디어처리파트에서 다양한 인공지능 기반 시스템을 연구 & 개발하고 있습니다. 급속도로 발전하는 딥러닝 기술을 매일 공부하고 따라가기도 벅차지만 앞으로도 신입과 같은 열정을 갖고 열심히 배움으로써 더욱 편리한 서비스를 만들어내고 싶습니다.

없다. 기술기 소실이나 과적합 문제가 발생하기 때문이다. 구글넷은 이를 해결하기 위해 인셉션 모듈(inception module)이라는 개념을 도입했다(그림 2) 참고). 인셉션 모듈은 같은 레이어에 서로 다른 크기를 갖는 여러 개의 컨볼루션 층들을 병렬적으로 이어놓은 구조를 말하는데 이때, 각 컨볼루션 필터(convolution filter) 앞에 1x1 컨볼루션 필터를 붙여 연산량을 크게 줄였다. 이를 통해 알렉스넷에 비해 12배나 적어진 파라미터 수를 가지고도 더 좋은 성능을 얻을 수 있다. 구글에서는 2014년 발표된 초기 구글넷을 인셉션-v1(Inception-v1)이라 명명하고 이후 꾸준히 개선된 인셉션 구조를 발표하였다. 2015년 12월에는 큰 사이즈의 컨볼루션 필터를 여러 개의 작은 컨볼루션 필터들로 대체하는 컨볼루션 팩토라이제이션(convolution factorization) 기법, 레이블 스무딩(label smoothing), 배치 정규화(batch normalization) 등의 개념이 모두 결합된 인셉션-v3(Inception-v3) 모델<sup>5</sup>이 발표되었다.



### 꽃검색 서비스의 시작

2015년 초부터 카카오는 딥러닝 기반의 꽃검색 서비스를 제공하기 위해 국내에 주로 서식하는 약 500여 가지의 꽃 품종에 대한 꽃 사진 수집을 통해 십여 만장의 꽃 품종 분류 데이터 셋(data set)을 구축하였다. 딥러닝 학습을 위한 데이터 셋 구축 과정에서 여러 이슈가 발생했는데 이 중 하나는 꽃 품종의 중복이다. 인터넷에서 수집한 대량의 꽃 사진 데이터 셋 중 실제로 같은 품종임에도 비슷한 이름의 다른 품종으로 분류되는 경우가 있었다. 예를 들어, ‘물봉선 꽃’과 ‘물봉선화’는 실제로 같은 품종의 꽃이지만 두 가지 이름이 혼용되고 있었고 거의 동일한 꽃 사진이 각각 다른 품종으로

분류되면서 딥러닝 학습을 방해하는 경우가 존재했다. 이를 해결하기 위해 500여 종에 달하는 꽃 품종을 사람의 눈으로 직접 비교해 위와 같이 혼동되는 품종의 사진 데이터를 병합하는 작업을 수행하였다.

또 다른 이슈 하나는 데이터 셋 구축을 위해 인터넷으로부터 꽃 사진을 수집하는 과정에서 발생한 품종별 사진 데이터 수의 불균형 문제(data imbalance problem)다. ‘벚꽃’이나 ‘장미’와 같이 사람들에게 널리 알려지고 주위에서 흔히 볼 수 있는 꽃의 경우 인터넷에 업로드된 다량의 사진 데이터가 존재한다. 하지만 ‘금강제비 꽃’이나 ‘제주소황금’과 같이 한국에서 쉽게 발견할 수 없는 희귀한 품종은 사진 데이터의 수집이 어렵기 때문에 품종별 수집된 사진의 개수에 있어 크게 차이가 나게 된다. 이러한 불균형 문제를 해결하고 품종별 사진 수를 유사하게 조절하기 위해 사진의 수가 적은 품종을 기준으로 다른 품종의 사진을 무작위로 제거하여 수를 줄여주거나(undersampling), 사진의 수가 많은 품종을 기준으로 다른 품종의 사진을 변형해 수를 늘려주는 방법(oversampling) 등이 필요했다.

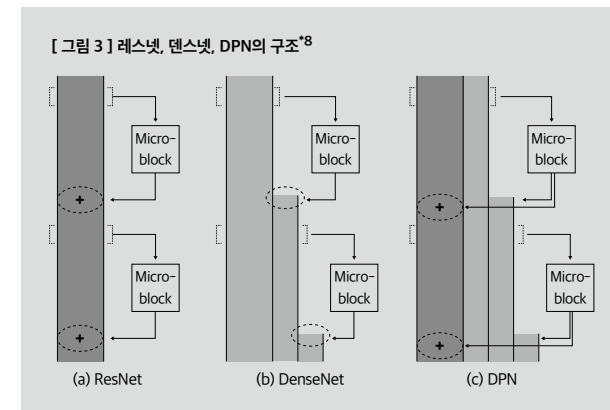
위와 같은 과정을 거쳐 최종적으로 구축된 10만 장 이상의 대규모 꽃 사진 데이터 셋은 2015년 당시 가장 높은 성능을 보인 CNN 모델을 바탕으로 한 꽃 품종 분류 네트워크를 학습하기 위해 사용되었다. 또한 꽃검색 서비스에 꽃이 아닌 사진이 제공되는 경우 임의의 꽃 품종을 출력하는 경우를 막기 위해 사진의 꽃 여부를 판단하는 이진 분류기를 추가로 학습 및 적용하였다. 아울러 분류 성능을 높이기 위한 다양한 테크닉을 추가로 개발 및 적용하였다. 예를 들어 다양한 데이터 증대 기법을 이용해 꽃 사진 데이터 셋을 변형함으로써 각도, 날씨, 조명 등 보다 다양한 꽃 사진 촬영 환경에 대해 강인한 분류 성능을 얻을 수 있었던 것이다. 그 결과, 기존 전통적인 컴퓨터 비전(computer vision) 기술을 이용한 꽃 검색 알고리즘은 분류해야 할 꽃의 품종이 102종 밖에 되지 않음에도 서비스로 제공할 정도의 충분한 정확도를 얻지 못했다. 하지만 딥러닝을 이용한 꽃 품종 분류 알고리즘의 경우 500종이 넘는 품종을 분류해야 하는 훨씬 어려운 문제임에도 불구하고 기존 대비 분류 정확도를 30% 이상 개선할 수 있었고, 이를 바탕으로 다음 앱 꽃검색 서비스를 개시하게 됐다.

### 계속되는 딥러닝 기술의 발전

2014년의 구글넷 이후부터 현재까지 사진 분류를 위한 딥러닝 네트워크 구조는 끝없이 발전하고 있다. 특히, 2015년에는 레스넷(residual network, ResNet)<sup>6</sup>이라는 혁신적인 네트워크 구조가 등장했다. 이전의 딥러닝 네트워크에서는 레이어를 많이 쌓을수록 기술기 소실 문제나 성능 저하 문제가 발생하여 네트워크

깊이에 있어 한계가 존재했다. 인셉션 모듈을 이용해 이러한 문제를 해결하고자 한 구글넷과 달리 레스넷은 스킵 커넥션(skip-connection)이라는 새로운 구조를 이용해 이를 극복했다(그림 3(a)). 스킵 커넥션은 한 컨볼루션 레이어의 입력 값을 레이어의 출력에 직접 더해주는 구조로, 이를 통해 각 레이어는 입력과 출력의 차이 값(residual)만을 예측하도록 한다. 네트워크 후반부의 기울기(gradient)를 전반부에 손실 없이 전달함으로써 훨씬 효과적인 딥러닝 학습이 가능하게 한 것이다. 이러한 스킵 커넥션을 바탕으로 직렬로 연결한 152개의 레이어로 구성된 레스넷은 2015년 ILSVRC 대회에서 압도적인 성능으로 우승하였다.

레스넷의 발명 이후 현재까지도 레스넷의 스킵 커넥션 개념을 활용한 다양한 네트워크가 제안되고 있다. 2017년에 발표된 덴스넷(densely connected network, DenseNet)<sup>7</sup>은 모든 레이어들을 다른 레이어와 직접적으로 연결해 네트워크 전반부에서 추출한 특징 맵(feature map)들을 계속 이어 붙여(concatenate) 후반부로 전달하여 새로운 특징을 계속 추출하는 네트워크다(그림 3(b)). 이러한 구조는 네트워크 후반부의 기울기 값을 보다 직접적으로 전달할 수 있어 레스넷에 비해 적은 파라미터 수로도 좋은 성능을 낸다는 장점이 있다. 최근 개발된 DPN(Dual Path Network)<sup>8</sup>은 이러한 레스넷과 덴스넷 구조를 결합함으로써 네트워크가 추출한 일반적인 특징은 skip-connection path를 통해 계속 재사용하면서 densely connected path를 통해 후반부에서 계속 새로운 특징을 추출하도록 유도(dual path)하였다(그림 3(c)). 이를 통해 각 레스넷과 덴스넷의 장점들만을 취하여 더욱 높은 사진 분류 성능을 얻게 되었다.



### 꽃검색은 계속 진화 중

지금까지 살펴본 바와 같이 영상 인식 및 사진 분류를 위한 딥러닝 기술은 매우 빠르게 진화하고 있으며, 이러한 기술의 발전 속도에 맞춰 카카오의 꽃검색 서비스 역시 발전하고 있다. 현재 다음 검색을

통해 서비스되고 있는 꽃검색은 약 500여 종의 꽃 품종에 대한 분류를 지원하고 있지만 실제 우리나라와 전 세계에 서식하는 꽃의 품종은 훨씬 다양하다. 카카오는 사용자가 보다 다양한 종류의 꽃 품종을 검색할 수 있도록 꽃 품종 데이터 셋의 규모를 현재의 500여 종에서 600여 종이 더 추가된 약 1,100여 종으로 확대하기 위해 더욱 많은 사진 데이터를 수집 중에 있다. 또한 현재 꽃검색 서비스에서 사용되고 있는 딥러닝 네트워크보다 높은 성능을 가진 최신 네트워크 구조를 바탕으로 다양한 딥러닝 기술을 활용한 새로운 꽃검색 알고리즘을 지속적으로 개발하고 있다. 이와 같이 카카오는 딥러닝 기술의 최신 트렌드를 적용하여 이용자들에게 더욱 똑똑하고 편리한 경험을 제공하고자 한다. 그렇기 때문에 꽃검색은 지금도 계속 진화 중이다.

<sup>1</sup> 참고 | [http://magazine.channel.daum.net/daumapp\\_notice/search\\_flower](http://magazine.channel.daum.net/daumapp_notice/search_flower) <sup>2</sup> 참고 | A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks", In NIPS, 2012. <sup>3</sup> 참고 | K. Simonyan, and A. Zisserman, "Very deep convolutional networks for large-scale image recognition", In ICLR, 2015. <sup>4</sup> 참고 | C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions", In CVPR, 2015. <sup>5</sup> 참고 | C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision", In CVPR, 2016. <sup>6</sup> 참고 | K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition", In CVPR, 2016. <sup>7</sup> 참고 | G. Huang, Z. Liu, K. Q. Weinberger, and L. van der Maaten, "Densely connected convolutional networks", In CVPR, 2017. <sup>8</sup> 참고 | Y. Chen, J. Li, H. Xiao, X. Jin, S. Yan, and J. Feng, "Dual path networks", In NIPS, 2017.