

일반논문 (Regular Paper)

방송공학회논문지 제23권 제5호, 2018년 9월 (JBE Vol. 23, No. 5, September 2018)

<https://doi.org/10.5909/JBE.2018.23.5.669>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

딥러닝 기반 암세포 사진 분류 알고리즘

서 영 민^{a)}, 한 종 기^{a)†}

Deep Learning Algorithm to Identify Cancer Pictures

Young-Min Seo^{a)} and Jong-Ki Han^{a)†}

요 약

본 논문에서는 고해상도 자궁경부암 세포사진을 CNN(Convolution Neural Network)을 통해 효과적으로 인식 및 분류하는 방법을 소개한다. 이때 고려되는 세포의 종류는 Ascus, Inflammation, RCC, Normal로 네 가지가 있다. 본 논문에서는 먼저 기존의 고해상도 이미지를 분류하는 알고리즘을 소개하고, 이 방법을 이용하여 고해상도 세포사진을 분류하는 과정에서 어떤 정보의 손실이 발생하는지 분석한 후, 이를 해결하기 위한 방법을 제시한다. 이를 위해서 본 논문에서 제안하는 학습 모델에서는 dilated convolution을 이용하여 고해상도 사진의 정보의 손실을 최소한으로 줄임과 동시에 학습속도 빠르게 하는 알고리즘을 제시한다. 또한 이미지 전처리 과정으로 임계치를 사용함으로써 암세포를 판단하는데 혼란을 줄 수 있는 부분을 제거함으로써 인식률을 향상시킨다. 본 논문에서 제시되는 실험 결과를 통해, 제안한 알고리즘이 기존 기술보다 높은 인식률을 제공하는 것을 확인할 수 있었다.

Abstract

CNN (Convolution Neural Network) is one of the most important techniques to identify the kind of objects in the captured pictures. Whereas the conventional models have been used for low resolution images, the technique to recognize the high resolution images becomes crucial in the field of artificial intelligence. In this paper, we proposed an efficient CNN model based on dilated convolution and thresholding techniques to increase the recognition ratio and to decrease the computational complexity. The simulation results show that the proposed algorithm outperforms the conventional method and the thresholding technique enhances the performance of the proposed model.

Keyword : CNN, Dilated Convolution, Image Recognition

a) 세종대학교 전자정보통신공학과(Sejong University, Dept. of Electrical Engineering)

† Corresponding Author : 한종기(Jong-Ki Han)

E-mail: hjk@sejong.edu

Tel: +82-2-3408-3739

ORCID:<https://orcid.org/0000-0002-5036-7199>

※ This work was partly supported by the National Research Foundation of Korea (NRF) under Grant NRF-2018R1A2A2A05023117 and partly by Institute for Information & communications Technology Promotion (IITP) under Grant 2017-0-00486 funded by the Korea government (MSIT).

· Manuscript received July 11, 2018; Revised August 22, 2018; Accepted August 22, 2018.

I. 서 론

최근 CNN을 이용한 연구가 활발히 진행함에 따라 여러 가지 이미지 인식에 관한 연구 결과들이 생겨나고 있으며, 특히 암세포 이미지 인식 및 분류에 대한 연구들이 급속한 발전을 이루었다^{[1][2][3]}. 의학적인 관점에 있어 아직까지는 딥러닝 기술이 인간의 손을 거치지 않아도 될 정도의 인식률 또는 정확성의 신뢰를 주지 못한다. 그러나 암세포 분류에 있어 병리과 의사가 암세포를 구별하는데 딥러닝 기술을 통한 암세포 사진 분류를 1차적으로 마친 뒤, 정확도가 낮거나 애매한 사진들을 중점적으로 살펴 본다면 병리과 의사의 피로와 소모시간을 획기적으로 줄일 수 있다. 예를 들어, 현재 IBM의 패턴인식 기술로 사진을 높은 인식률로 판단하는 기술^[4]과 같이 인간의 수고를 덜 수 있으며, 이와 더불어 연구의 발전 속도가 급속도로 빨라지고 있는 이 시점에서 앞으로 CNN을 통한 의료 영상 인식의 잠재성은 아주 크다고 볼 수 있다^[5].

본 논문에서 고려하는 암세포 사진은 해상도가 2048×1536으로 상당히 높은 편이며, 이렇게 고해상도인 의료 영상을 기준의 알고리즘을 사용하여 분류할 경우 인식률이 기대에 미치지 못하는 현상이 발생한다. 본 논문에서 고려하는 기준 알고리즘들에는 대표적으로 CIFAR-10^[6]가 있는데, CIFAR-10 알고리즘은 입력 영상의 해상도가 32×32일 때 동작되도록 최적화 되어 있다. 따라서, 본 논문에서 고려하는 의료 영상을 CIFAR-10으로 분류하기 위해서, 2048×1536 해상도인 영상을 32×32로 다운 사이징 시킨 후, CIFAR-10의 테스트 영상으로 사용하였다. 한편, 다른 기존 연구로는 Aditya Golatkar^[7]가 사용한 Inception 모델^{[8][9][10]}이 있는데, 이 연구에서 사용된 방법은 고해상도의 의료 영상들을 Inception 모델의 입력 영상 해상도에 맞게 299×299 크기로 분할시키는데, 이때 분할된 이미지들끼리 영상 정보를 50% 오버랩되도록 분할한 후, 이 분할 영상들을 테스트 영상으로 사용하였다. 이 방법은 효과적으로 보일 수 있지만, 오버랩을 시킬 경우 암세포들이 겹쳐지기 때문에 만일 하나의 암세포가 정상세포와 겹쳐질 경우 사진 상으로 정확한 판단이 불가능하기 때문에, 일부 제한적인 암세포 사진에서만 사용될 수 있다.

본 논문에서 제안하는 알고리즘에서는 고해상도 의료 이

미지를 훈련 영상으로 이용하여 CNN 모델을 학습시켰으며, 이 경우 발생하는 정보의 손실을 최소화시키기 위해서, 불필요한 노이즈를 전처리과정으로 제거하는 과정을 사용하였다. 제안하는 알고리즘의 첫 번째 단계는 이미지 전처리 과정으로써 임계값을 주어 사진에 존재하는 물결 같은 노이즈나 핵의 모양을 알아볼 수 없게 방해하는 노이즈들을 보이지 않게 하였다. 이것은 Kim, Gang Beak이 저술한 논문의 자궁암세포 판단근거에 입각했다^[11]. 두 번째 단계에서는 고해상도 영상들을 대상으로 Convolution 계산을 효율적으로 수행하기 위해서 Dilated Convolution 기법을 이용하였다. Dilated Convolution은 수용필드영역은 늘리면서 Parameter의 개수는 유지하는, 입력과 출력의 이미지 사이즈가 유지되어야 할 때 사용되는 Convolution 방법이다.

본 논문의 구성은 다음과 같다. II장에서는 기존 이미지 분류 알고리즘들에 대해 설명하고 발생하는 문제점을 알아본다. III장에서는 고해상도 영상 분류를 효율적으로 수행하기 위한 dilated convolution 기법에 대해서 설명한다. 여기서 개념 및 간단한 실험 결과를 제시함으로써, 본 논문에서 제안하는 모델의 핵심 기술에 대한 이해를 돋고자 한다. IV장에서는 본 논문에서 제안하는 이미지 분류 알고리즘에 대해 설명하고 해당 알고리즘이 갖고 있는 기존 알고리즘 대비 장점에 대해 설명한다. 또한, 이 장에서는 이미지 전처리 과정으로 해당 세포 사진의 노이즈를 없애고 불필요 영역을 폐기하는 방법에 대해 설명한다. V장에서는 구체적인 실험 결과와 기존 알고리즘들과의 인식률 및 학습시간을 비교하며 차이에 대한 원인 분석을 실시한다. 마지막 VI장에서는 논문의 간단한 결론을 맺는다.

II. 기존의 CIFAR-10을 이용한 세포사진 분류

본 장에서는 기존 CNN 모듈 중에 하나인 CIFAR-10^[12]를 구성하고 있는 학습 모델에 대해 설명한다. 그림 1은 CIFAR-10에서 사용되는 CNN 모델의 구조이다. 간단히 모델의 구조에 대해서 말하자면, 첫 번째 convolution layer에

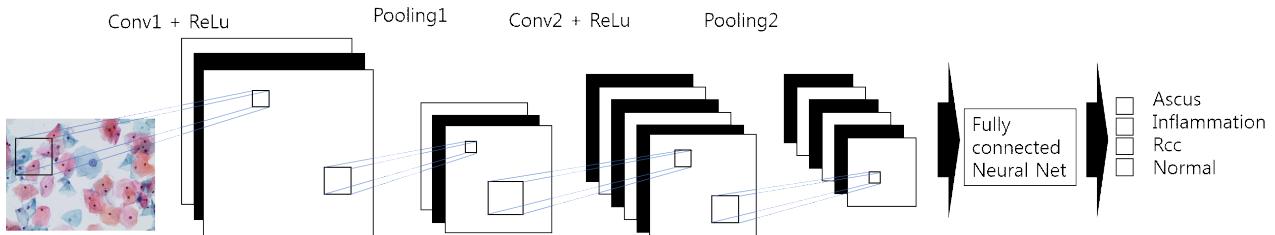


그림 1. CIFAR-10에서 사용하는 CNN 구조
 Fig. 1. CNN Structure used in CIFAR-10

서 하나의 이미지를 64개의 특징들로 맵핑하고, 그 결과 정 보에 대해 max pooling을 수행한다. 그 다음에 존재하는 convolution layer에서는 32개의 특징들을 64개의 특징들로 맵핑하고, 이 결과를 다시 max Pooling한다. 이런 식으로 다섯 단까지 반복하여 382개의 특징들로 맵핑하고, 마지막 단(layer)에서 Fully Connected Layer를 구성한다. 본 논문에서 구분하고자 하는 암세포의 유형이 네 가지이기 때문에 Full Connected Layer이후 4가지의 클래스로 최종 맵핑

한다. 또한 Convolution의 padding size는 2이며, Kernel size는 5이다. 그리고 Pooling의 Stride는 2이다.

이 CIFAR-10 구조를 학습시키는 과정에서 테스트 데이터 세트는 본 논문의 저자들이 사용하는 고해상도 자궁경부암 사진으로 구성하였다. 이때 CIFAR-10은 32×32 이미지 사이즈를 훈련 및 테스트 영상으로 사용하도록 설계되었기 때문에, 본 논문에서 구현할 때도, 입력 의료 영상의 해상도를 32×32로 축소시킨 후 사용하였다. 아래의 그림

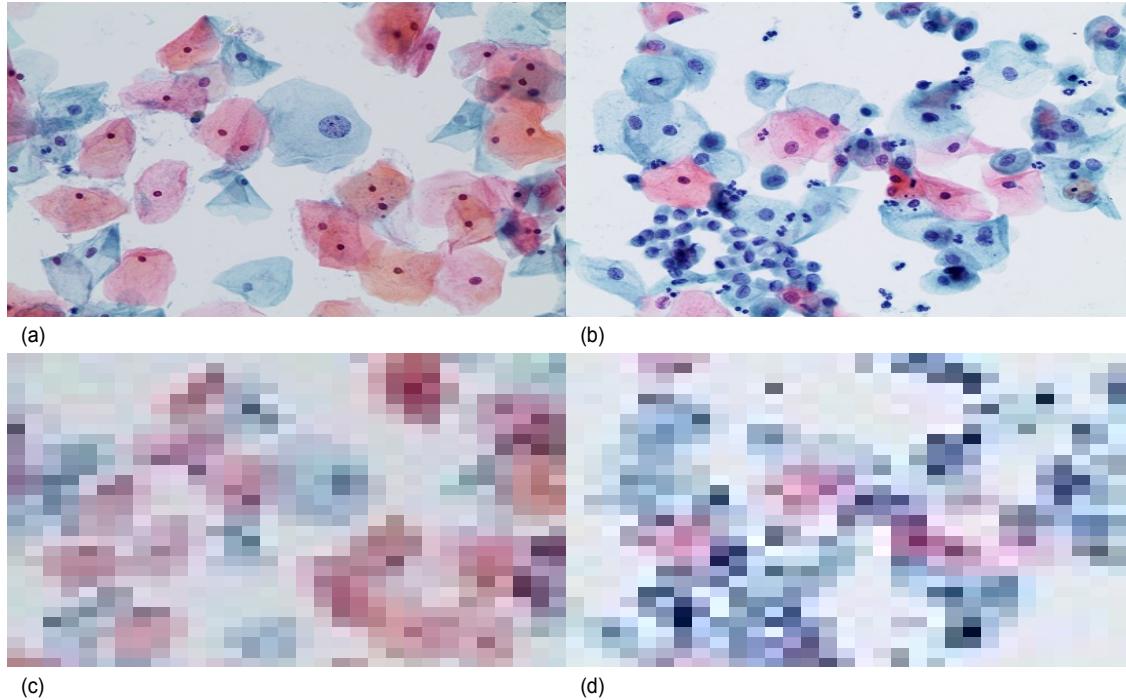


그림 2. CIFAR-10 모델에 사용하기 위해 축소된 세포사진들의 예시, (a) 고해상도 세포 사진 예시, (b) 고해상도 세포사진 예시, (c) 축소된 세포사진 예시, (d) 축소된 세포사진 예시

Fig. 2. Examples of original pictures and the reduced pictures containing cells, (a) high resolution picture containing the cells, (b) high resolution picture containing the cells, (c) low resolution picture containing the cells, (d) low resolution picture containing the cells

그림 2는 CIFAR-10을 이용하기 위해서, 원본사진과 축소된 사진들의 비교 예시들이다. 그림 2 (a)와 그림 2 (b)는 원본 세포 사진들이며, 그림 2 (c)와 그림 2 (d)는 32×32 해상도로 축소된 세포 사진을 의미한다. 이 그림에서 확인할 수 있듯이, 32×32 의 축소된 사진을 보면 핵의 모양이나 사진의 형태들이 blurred 되어 있는 것을 알 수 있다. 이로 인해 암세포를 판단하기 위한 근거인 세포들의 모양이나 구조가 저해상도로 변환되었기 때문에 CIFAR-10 모델에서 암세포의 유무 및 종류를 판단하는 것이 어려울 것으로 예상된다.

III. Dilated Convolution

1. Dilated Convolution의 개념

본 논문에서는 고해상도 영상을 구분(identify)하면서 계산 복잡도를 낮추기 위해 기존의 convolution을 사용하는 대신 dilated convolution을 사용한다. Dilated convolution

이란 확장된 컨볼루션이라고도 불리며 입력 이미지가 고해상도이며 출력이미지 또한 이를 유지해야 할 필요가 있을 때 효과적인 성능을 보인다^{[13][14]}. Dilated convolution을 이해하기 위해 우선 수용필드영역(Receptive Field)의 개념을 이해해야 하는데 이에 대한 설명은 다음 그림 3과 같다.

그림 3은 기존의 3×3 convolution과 3×3 dilated convolution을 사용할 때, 수용 필드의 예시들을 비교하는 그림이다. 그림 3 (a)에서 보여주는 기존의 convolution의 수용 필드들과 달리 그림 3 (b)에서 보여주는 dilated convolution의 수용필드는 같은 3×3 의 커널이지만 그 사이에 hole value로 0을 넣고 간격을 둔 채로 convolution 수행함으로써 수용필드가 달라지는 것을 확인할 수 있다. 따라서 dilated convolution을 사용함으로써 연산량은 낮추면서, 고해상도 영상내의 특성들을 효율적으로 학습하는 것이 가능하다. 이때 수용필드 내에 0을 삽입하는 개수에 따라 rate 조절이 가능하다. 즉, dilated convolution은 필터 내부에 제로 패딩을 줌으로써 receptive field를 강제적으로 늘리는 방법이다. 좀 더 자세한 비교를 위해 다음 세 가지 그림과 함께 설명을 실시한다.

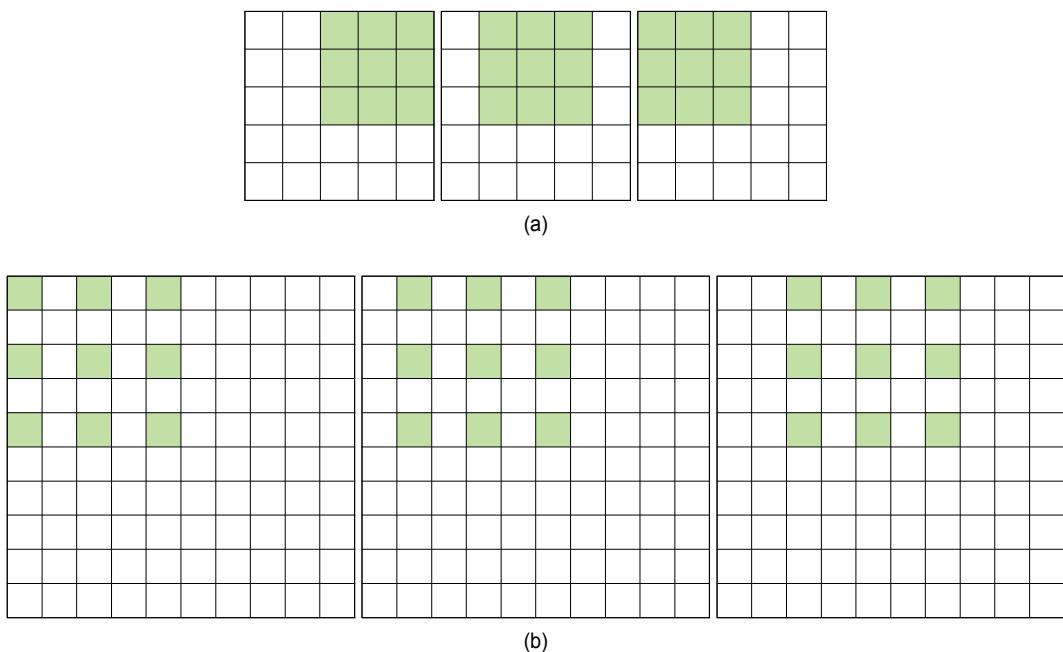


그림 3. 기존의 convolution과 dilated convolution의 수용필드 예시들, (a) 기존 convolution의 수용필드, (b) Dilated convolution의 수용필드

Fig. 3. Examples of receptive fields between conventional convolution and dilated convolution, (a) Receptive field of conventional convolution, (b) Receptive field of dilated convolution

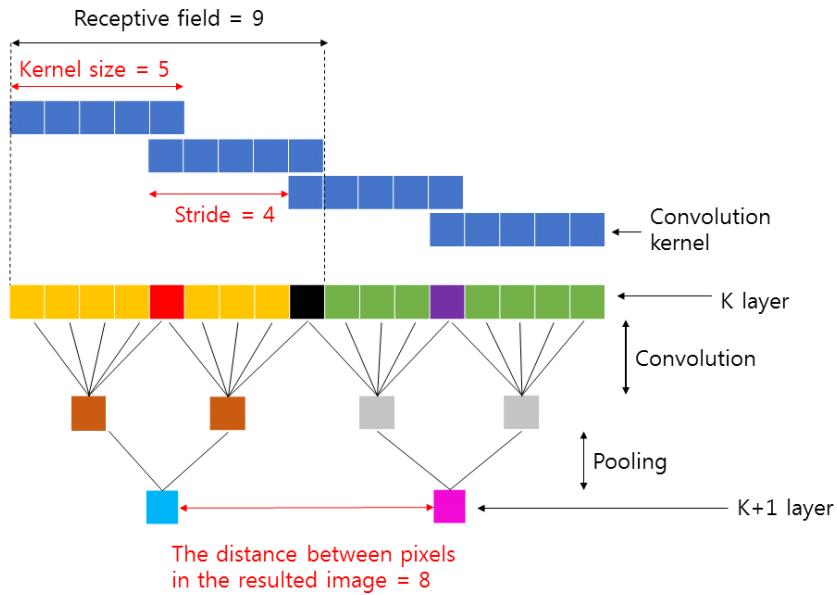


그림 4. 수용필드가 9인 일반적인 convolution의 예시
 Fig. 4. Example for conventional convolution with 9 receptive fields

2. Dilated Convolution의 예시

그림 4는 일반적인 convolution과 pooling을 수행한 결과이다. 그림 4에서 보이듯이 수용필드영역을 9로 맞추기 위해 stride를 4로 설정하고 kernel의 크기를 5로 설정했다. 그 이유는 커널사이즈가 5일 때 Stride가 4여야 수용필드가 9가 되기 때문이다. 이것을 점선으로 구분해 놓았고, 붉은 박스와 검은 박스, 보라색 박스는 커널이 Stride에 따라 겹치는 부분이다. 또한 노란영역의 박스들과 초록박스들의 영역은 convolution 이전의 K layer와 convolution이후의 갈색과 회색의 박스로 구분했다. pooling을 거친 뒤의 최종 블록은 하늘색과 자주색으로 K+1 layer에 정의했다. 이렇게 하면 마지막 출력단의 이웃 node들은 원 영상 좌표상에서 8픽셀 떨어져있는 것과 같게 된다. 결과적으로 출력단 영상 정보의 해상도가 매우 낮아지게 된다.

그림 5는 이전 그림 4의 단점을 보완하기 위해 kernel 사이즈를 9로 증가시키고, stride를 1로 설정한 후, convolution을 수행하는 과정의 개념도이다. Pooling을 제거하였으나 사이즈가 줄지 않아 연산량이 많아진다. 마찬가지로 초록색 블록은 Stride에 의해 겹치는 부분이고 노란색과 보라색 블록은 겹치지 않는 kernel이다. 그림 5에서 볼 수

있듯이, 마지막 출력단의 이웃 node들은 원 영상 좌표상에서 1픽셀 만큼 떨어져있는 것과 같게 된다. 이와 같은 수행 과정에서 출력단의 node 정보들 간의 영상 좌표 거리는 좁아졌지만, 학습시켜야 할 parameter들이 크게 증가하기 때

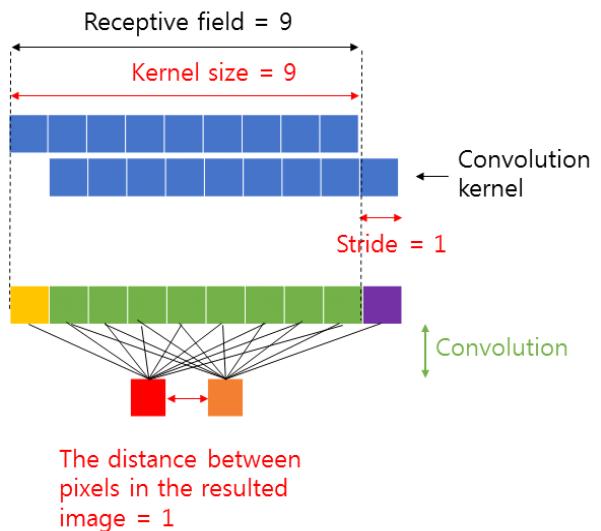


그림 5. Parameter가 조정된 일반적인 convolution의 예시
 Fig. 5. Example of conventional convolution having the modified parameters

문에 연산량 복잡도와 over-fitting의 위험성도 역시 증가하게 된다.

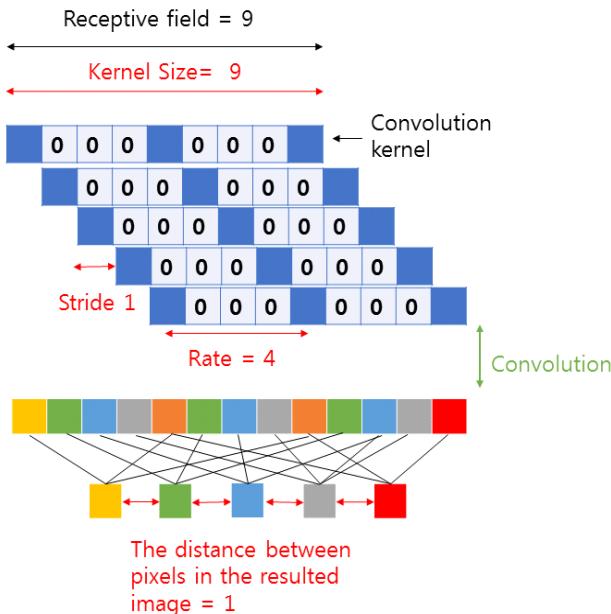


그림 6. 수용필드가 9인 dilated convolution의 예시
Fig. 6. Example for dilated convolution with 9 receptive field

그림 4와 그림 5의 예시의 단점을 극복하기 위해, dilated convolution을 사용한 예시를 그림 6에서 설명하고 있다. 그림 6에서 설명하는 dilated convolution에서는 수용필드의 크기를 9로 설정하고, kernel의 크기는 9로, stride는 1로, 0을 채워 넣는 rate는 4로 설정함으로써, 출력단의 이웃

node들은 원 영상 좌표상에서 1픽셀 만큼 떨어져있으면서도 계산 복잡도는 증가하지 않는 장점을 갖게 된다. 위 그림에서 첫 번째 dilated convolution의 결과를 노란색으로, 두 번째를 초록색으로, 세 번째를 하늘색으로, 네 번째를 회색으로, 마지막 다섯 번째를 빨간색으로 구분하였다. 중간의 13칸짜리 긴 막대는 모든 dilated convolution의 kernel을 겹쳐 놓은 것을 도식화 한 것이고 겹치는 부분에 대해서는 노란색과 빨간색이 겹치므로 주황색으로 표현하였다.

다음 수식 (1)은 입력 신호 $x[i]$ 에 대해 확장비율 r 을 갖는 dilated convolution의 결과 신호 $y[i]$ 를 얻는 과정을 표한한 식이다.

$$y[i] = \sum_{q=1}^S x[i + r \cdot q]w[q] \quad (1)$$

여기서 w 는 convolution 필터이며, S 는 필터길이, r 은 stride의 길이이다. 예를 들어 일반적인 convolution에서는 출력신호가 식 (2)와 같이 계산되지만,

$$y[i] = x[i+1]w[1] + x[i+2]w[2] + \dots + x[i+s]w[s] \quad (2)$$

$r=3$ 으로 세팅된 경우의 dilated convolution에서는 식 (3)과 같이 계산된다.

$$y[i] = x[i+3 \times 1]w[1] + x[i+3 \times 2]w[2] + \dots + x[i+3 \times s]w[s] \quad (3)$$

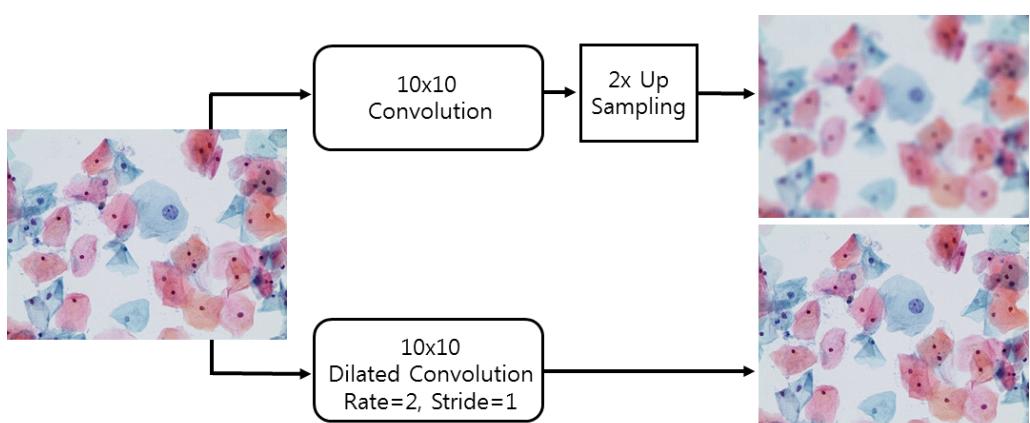


그림 7. Dilated convolution과 일반적인 convolution을 적용한 결과 영상들의 비교
Fig. 7. Comparison between images resulted from dilated convolution and conventional convolution

식 (3)에서는 원래 신호 좌표에서 3칸씩 건너뛰면서 계산한 결과를 출력값으로 얻게 된다.

그림 7은 10×10 크기의 kernel 크기를 갖는 일반적인 convolution과 dilated convolution을 사용하였을 때의 결과 이미지를 간의 차이를 비교하는 그림이다. 일반적인 convolution을 적용했을 경우, 결과 이미지의 해상도를 원 영상의 해상도로 유지하기 위해 2배의 up-sampling을 실시하였다. 이에 반해 dilated convolution을 적용한 경우에는 결과 영상의 해상도가 입력 영상의 해상도와 같이 때문에 별도의 up-sampling 과정이 불필요하다는 것을 알 수 있다.

3. Dilated Convolution의 변수들

Dilated convolution에서 사용하는 수용필드(receptive field) 및 관련된 변수값들을 계산하는 수식들은 다음과 같다^[15].

$$n_{output} = \left[\frac{n_{input} + 2p - k}{s} \right] + 1 \quad (4)$$

수식 (4)에서 n_{output} 은 output node의 수, n_{input} 은 input node의 수, k 는 kernel의 크기, p 는 padding의 크기, s 는 stride 크기이다. 위 방정식에서는 input node의 수와 convolution의 특성들을 기반으로 output node 수를 계산한다.

$$j_{output} = j_{input} \times s \quad (5)$$

수식 (5)에서 j 는 jump를 의미하는 약어로써 두 개의 연속적인 node들의 거리를 의미한다. j_{input} 은 input node들이 input image 내 공간좌표 상에서 떨어져 있는 정도, j_{output} 은 output node들이 image 내 공간좌표 상에서 떨어져 있는 정도이다. 위 방정식에서는 input에서 stride의 크기를 적용할 때 건너 뛴 node들의 개수를 곱함으로써 출력단의 jump를 계산한다.

$$r_{output} = r_{input} + (k-1) \times j_{input} \quad (6)$$

수식 (6)에서 r_{input} 은 input feature map의 수용필드 크기이고, r_{output} 은 output feature map의 수용필드 크기이다.

다. 위 방정식에서는 input 수용필드 사이즈에 j_{input} 과 $(k-1)$ 을 곱한 결과를 더해주는 식이다.

$$start_{output} = start_{input} + \left(\frac{k-1}{2} - p \right) \times j_{input} \quad (7)$$

수식 (7)에서 $start$ 는 첫 번째 feature의 중앙점을 의미한다. $start_{input}$ 은 input feature map의 첫 번째 feature가 input image 상에서 갖는 공간적인 위치이며, $start_{output}$ 은 output feature map의 첫 번째 feature가 input image 상에서 갖는 공간적인 위치이다. 위 식 (7)은 $start_{input}$ 에서 커널의 반쪽만큼 convolution의 중심점이 대각선 아래로 이동하는 식이다.

IV. 제안하는 CNN 구조

1. 제안하는 CNN 모델 구조

그림 8은 본 논문에서 제안하는 CNN 모델의 구조를 개괄적으로 표현한 것이다. 본 논문에서는 고해상도 영상에 대해 빠르고 효과적으로 convolution을 수행하기 위해서 dilated convolution을 사용한다. Dilated convolution의 수행 과정 및 rate 값에 따른 결과값들의 특성은 앞 장에서 이미 설명하였다. 제안하는 학습 모델에서는 그림 8에서 표현했듯이 다양한 rate를 사용해서 dilated convolution을 수행한다. 구체적으로 3×3 크기의 커널에서 rate를 6, 12, 18, 24 비율로 dilated convolution을 수행한다. 여기서 rate란 앞의 3장에서 설명했듯이, kernel 크기가 확장(dilated)되는 비율이다. 예를 들면, rate가 6이라면, kernel을 구성하는 3×3 크기의 계수 사이마다 5개의 0을 삽입하여, 최종적으로 13×13 크기의 kernel을 구성하는 것을 의미한다. 또한 이 부분을 구현하기 위해서 본 연구자들은 Liang-Chieh Chen의 DeepLab에서 개발한 dilated convolution을 multi-scale에 적용하여 만든 ASPP(Atrous Spatial Pyramid Pooling)^{[16][17][18]} 방법을 사용하였다. 이 방법은 다양한 rate의 dilated convolution을 별별로 진행하여 다양한 receptive field를 분석할 수 있게 한다.

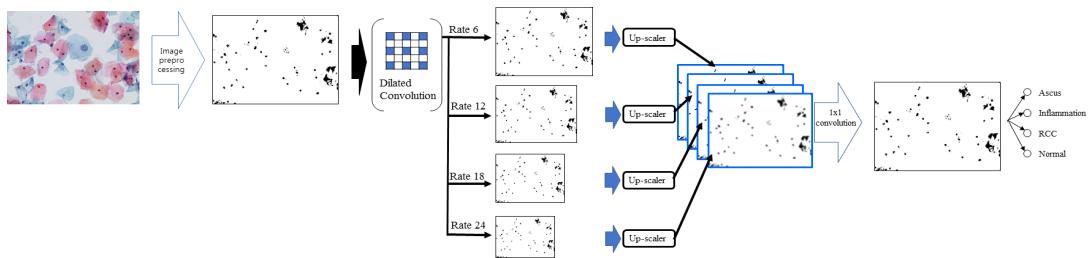


그림 8. 세포의 종류들을 구분하기 위해 제안된 학습 모델
Fig. 8. The proposed model to identify the kinds of the cell pictures

그림 8의 모델에서 rate 6, 12, 18, 24로 설정되어 dilated convolution을 수행하면 각각 다른 해상도의 영상 4장을 얻게 된다. 이 각 영상들을 up-sampling하여 해상도를 원래 영상의 해상도와 같게 변형시킨다. 이렇게 변형된 영상들을 결합시켜 3차원 구조의 데이터 형태로 만든 후, 이 데이터에 1×1 convolution을 적용한다. 1×1 convolution을 통해 얻어진 데이터를 사용하여 최종적인 세포 타입을 결정한다.

2. 이미지 전처리 과정

그림 8에서 설명한 CNN 학습 모델을 사용하여 4가지 종류의 암세포 종류를 학습할 때, 세포 사진 내의 핵의 이미지가 다른 노이즈에 영향 받지 않도록, 임계값을 사용하여 세포 핵의 모양만 검출하여 학습시키는 전처리 과정을 사용하였다. 이는 세포사진 내의 모두 화소값들을 사용할 필요 없이, 세포핵을 포함하는 영역만을 추출해서 제안하는

CNN 구조로 학습하는 것이 계산량 복잡도는 낮추고 학습률을 향상시키기 때문이다.

그림 9는 암세포를 포함하는 세포 사진의 예를 보여주고

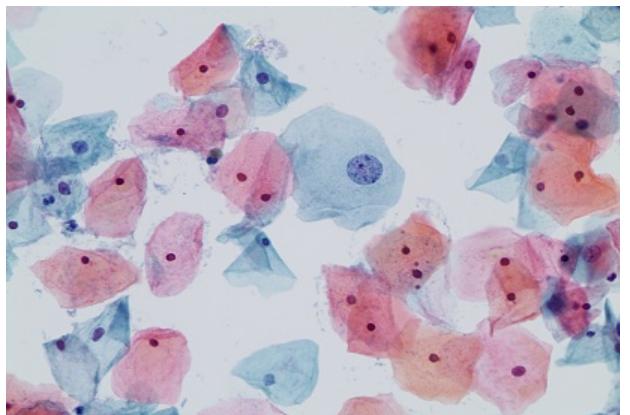


그림 9. 원본 암세포 사진
Fig. 9. Original picture representing cancer cells

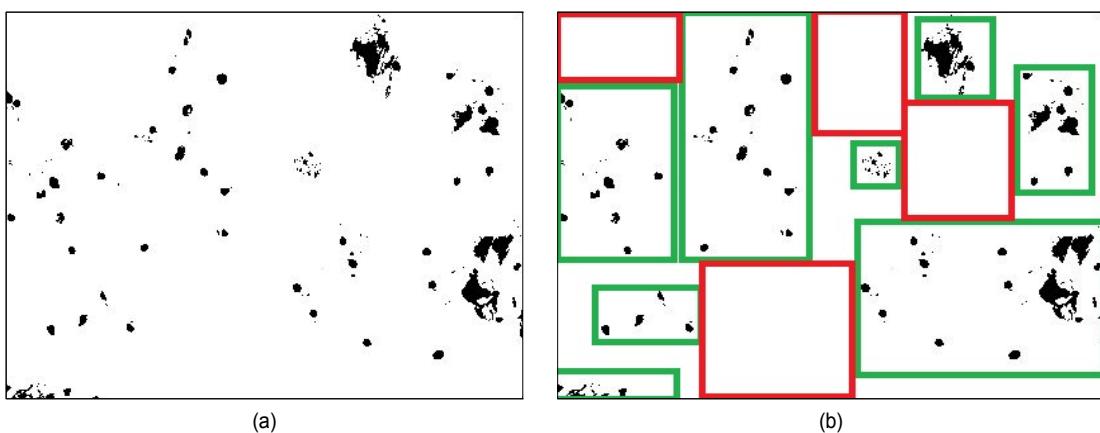


그림 10. (a) 임계치를 사용하여 전처리한 사진과 (b) 사진 내부의 활용 구간들
Fig. 10. (a) The preprocessed image and (b) the regions classified according to existence of cells

있다. 이 사진에서 확인할 수 있듯이, 다양한 모양과 크기의 세포핵이 존재하는 영역과 아무런 세포가 위치하지 않은 배경 영역이 존재한다. 본 연구에서는 미리 정해진 문턱치 값을 이용해서 세포핵의 모양을 검출하여 그림 10 (a)와 같은 사진을 만든 후, 이 사진을 기반으로 학습에 사용할 영역과 사용하지 않을 영역을 구분하는 그림 10 (b)를 만든다. 그림 10 (b)과 같은 훈련 데이터들 중에서 학습에 사용할 영역 데이터만을 그림 8의 학습 모델에 입력으로 사용하여 학습을 진행한다.

3. Multi-Rate 과정

세포들을 분류할 때 우수한 성능을 얻으면서 수용영역을 늘리려면 convolution과 pooling의 반복이 필요하다. 즉, 계속적인 pooling이 반복되어야 하는데, 이 경우 반복되는 pooling 때문에 detail이 사라지게 된다. 이로 인해, 세포의 세부적인 정보보다는 사진 전체적인 윤곽 정보로 분류를 진행하게 된다. 이러한 문제점을 극복하기 위해서 본 연구에서는 multi-rate로 구성된 학습 모델을 이용한다. 본 연구에서는 DeepLab^o Image segmentation에서 사용하고 있는 ASPP(Atrous Spatial Pyramid Pooling)^{[16][17][18]} 방법을 이용하였다.

Multi-Rate를 활용하면 다양한 해상도에서 학습을 시키고 그것을 비교하게 됨으로써 세포 사진의 윤곽 정보 뿐만 아니라 세밀한 정보를 이용해서 세포 종류를 분류할 수 있게 된다. 또한 앞서 3장에서 설명한 Dilated convolution을 사용함으로써 수용영역은 효과적으로 늘리고 Pooling layer를 없앨 수 있다.

4. Up-scaling

Up-scaling이 필요한 이유는 픽셀단위 예측을 하기 위해 필요하다. Dilated convolution을 하고 나면 원본 이미지의 크기가 줄어들게 되는데 이 상태로 픽셀단위 예측을 하게 되면 정확도가 떨어지기 때문이다. Multi-rate 과정을 거친 feature는 다양한 rate별로 축소되었기 때문에 이 다양한 feature들을 일정한 크기로 up-scaling하여 겹치게 되면 밀도에 따라서 픽셀단위 예측을 할 수 있다. 이렇게 되면 보다

정확한 정보에 근거하여 예측할 수 있다.

5. 1x1 convolution

일반적으로 CNN 모델의 마지막 단에서는 fully connected layer 역할을 하는 layer가 필요하다. 그런데 fully connected layer는 다양한 크기의 입력이 아닌 고정된 크기의 입력밖에 받아들이지 못하고, 더욱이 위치정보가 사라지기 때문에, 본 연구에서는 segmentation에 우수한 성능을 제공하는 1x1 convolution을 사용한다. 이 1x1 convolution을 fully connected layer 대신 사용하게 되면 정해진 해상도의 이미지 및 다양한 해상도의 이미지들을 처리할 수 있으며, patch 단위로 영상을 처리하지 않고, 전체 영상을 한 번에 처리하기 때문에 연산량 관점에 있어서도 효율적이다. 다음 그림 11은 convolution된 각각의 feature가 1x1 convolution을 거쳐 하나의 feature로 나오는 과정을 도식화 한 것이다.

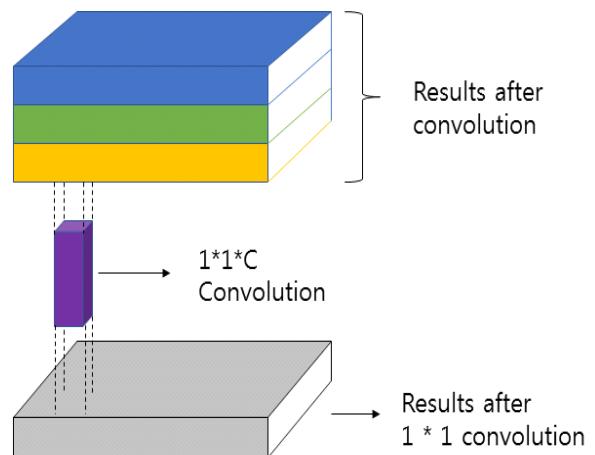


그림 11. 1x1 Convolution
Fig. 11. 1x1 Convolution

V. 실험 결과

본 장에서는 고해상도 세포 사진의 종류를 구분하는 CNN 모델들의 성능을 분석하기 위해서 기존 모델인 CIFAR-10 알고리즘과 제안하는 모델을 구현하여 여러 가

지 측면의 데이터를 분석하였다. 구체적으로는 (1) CIFAR-10 모델의 성능, (2) 그림 8의 dilated convolution을 사용하는 제안 모델의 성능, 그리고 (3) 임계치를 사용하여 필요한 영역만을 구분하여 그림 8의 모델을 사용한 알고리즘의 성능을 비교하였다.

사용된 세포 영상들은 해상도가 2048×1536 이며, 자궁경부에서 추출된 사진을 이용하였다. 사용된 사진들의 class 별 개수는 Ascus가 884장, Inflammation가 1062장, RCC가 438장, Normal가 266장이었다. Normal은 암세포가 없는 건강한 세포 사진을 의미한다. 각 class의 사진들 중 70%의 사진을 훈련 데이터로 사용하고 30%를 인식률 평가용 Test 데이터 사진으로 사용하였다. 이때 Learning rate는 0.1이고, batch size는 10으로 정했으며, epoch는 80회이다.

위와 같은 조건으로 실험을 실행한 결과를 표 1에 정리했다. 이 표에서 CIFAR-10의 인식률은 평균 21.0%이었고, 학습시간은 19분 22초 소모 되었다. CIFAR-10 알고리즘으로 실행했을 때는 입력 사진들의 해상도를 32×32 로 감소시킨 후 학습시켰기 때문에 인식률이 매우 낮음을 확인할 수 있었다. 그렇지만, 작은 영상을 훈련 데이터로 사용하였기 때문에, 메모리 소모 비율이 매우 낮고, 학습시간이 매우 짧은 장점이 있었다. 이 표의 실험 결과에서, dilated convolution을 사용하는 제안 모델을 사용하는 경우, 평균적으로 68.4%의 인식률을 얻었고, 학습에 소요되는 시간은 177분 34초임을 알 수 있었다. 이 모델에서는 원본사진의 고해상도를 유지한 채로 실험을 수행했고, 빠른 convolution을 수행하기 위해, dilated convolution을 이용하였다. 마지막 실험 결과인 임계값을 이용하여 중요한 세포 부분을 추출한 후 dilated convolution을 적용하는 제안 모델을 사용한 경우에는 평균적인 인식률이 70.9%로 향상되었으며, 이때 학습 시간이 141분 2초로 측정되었다. 이 실험에서 임계값은 125로 선택했으며, 다양한 사전 실험을 통해, 이 임계값이 세포 사진 내의 정보를 손실하지 않으면서 학습 시간을 단축시키는 것으로 확인할 수 있었다. 이 실험에서 인식률이 소폭 증가한 이유는 세포 사진의 세포벽이나 푸른색 또는 붉은색의 물결 같은 노이즈가 제거됨으로써 세포핵의 모양을 좀 더 확실하게 판단할 수 있었기 때문이다. 만약, Dilated convolution을 사용하지 않고 일반적인 Convolu-

tion으로 변환하여 실험할 경우 학습속도는 453분 19초로 증가하지만 아래 표와 같이 인식률 부분에 있어서는 72.3%로 Dilated convolution과 큰 차이가 없는 것을 확인 할 수 있었다. 그리고 제안하는 학습 모델구조는 그대로 유지한 뒤, Dilated convolution의 single rate 6으로 실험한 결과 인식률이 저조한 것을 알 수 있었다.

표 1. 알고리즘별 인식률 및 학습시간 비교

Table 1. Comparison between recognition ratios and consumed learning times for various algorithms

	Recognition Ratio	Consumed Learning time(min)
CIFAR-10	21.0%	19.22
Dilated Convolution with single rate	36.2%	217.11
Dilated Convolution with multi-rates	68.4%	177.34
Dilated Convolution with threshold	70.9%	141.02
Conventional Convolution	72.3%	453.19

본 연구에서 사용하는 세포 사진들의 암세포 타입은 Ascus, Inflammation, RCC, Normal 등 총 4가지인데, 각 타입별로 임계치를 이용하여 세포핵의 모양만을 추출한 영상들의 예를 그림 12에 보이고 있다. 이 사진들은 세 가지의 암세포 사진과 하나의 정상세포 사진을 비교하고 각각 이미지 전처리 과정으로 임계값을 125씩 준 사진들이다. 사진을 보는 바와 같이 임계값을 줌으로써 붉은색 또는 푸른색의 세포벽들이나 핵과 세포벽이 겹쳐지면서 발생하는 노이즈를 제거할 수 있음을 알 수 있다.

표 2에서는 제안하는 알고리즘의 세포별 인식률을 설명하고 있다. 이 실험 결과에서 Ascus 타입의 인식률은 72.1%이고, Inflammation 타입의 인식률은 78.3%이며, RCC 타입의 인식률은 67.7%이고, 정상세포인 Normal 타입의 인식률은 65.5%였다.

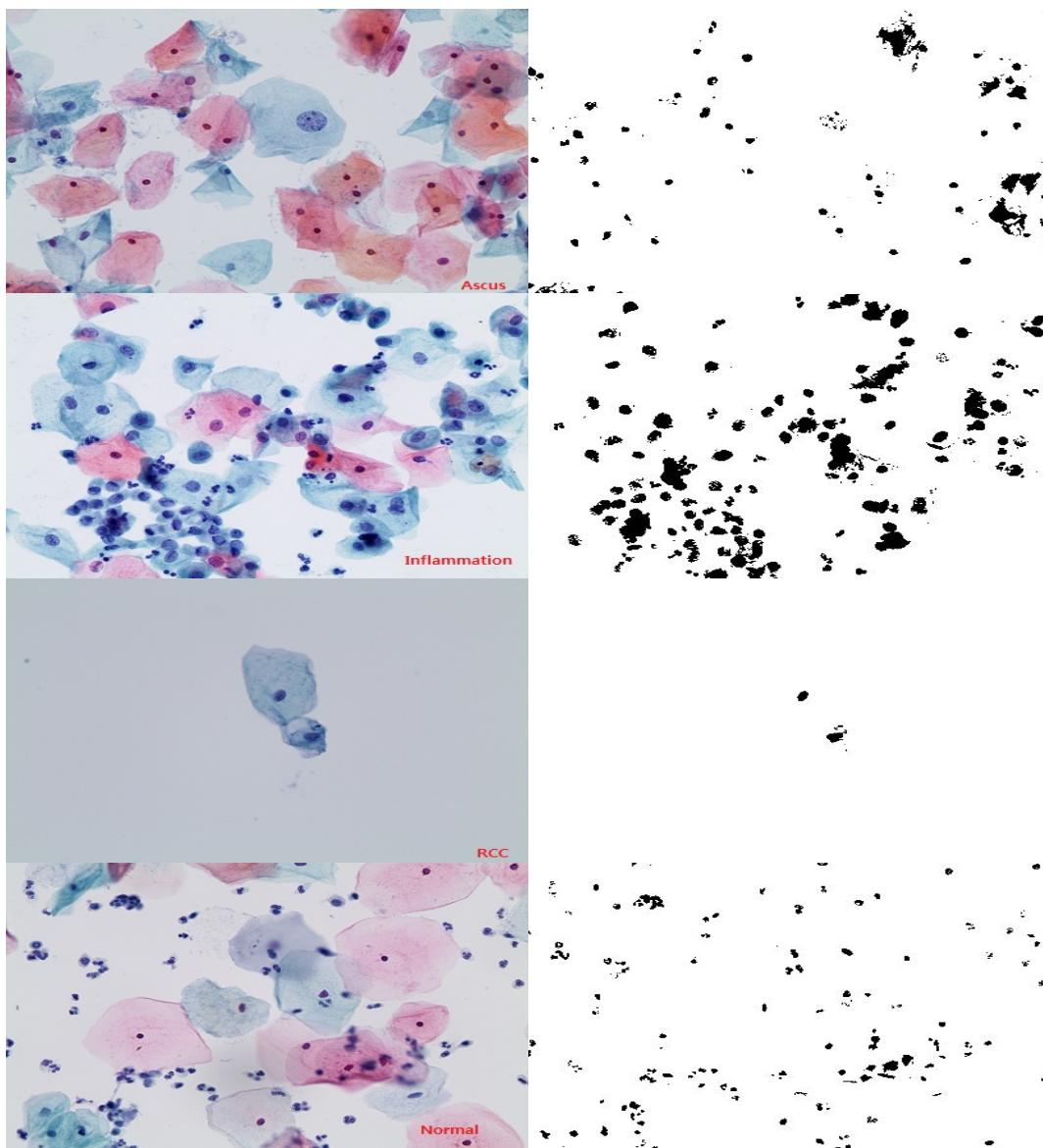


그림 12. 종류별 원본 암세포 사진과 임계값을 준 암세포 사진의 비교

Fig. 12. Examples of cancer cell pictures and the preprocessed pictures for four cancer types

표 2. 문턱치와 dilated convolution을 이용한 학습 모델의 세포 종류별 인식률

Table 2. Recognition ratios for cell types in CNN model based on dilated convolution and thresholding process

		Type recognized by using the proposed system			
		Ascus	Inflammation	RCC	Normal
Original Type of Cell	Ascus	72.1%	18.7%	3.1%	6.1%
	Inflammation	19.1%	78.3%	0.7%	1.9%
	RCC	10.3%	16.5%	67.7%	5.5%
	Normal	20.1%	10.6%	3.8%	65.5%

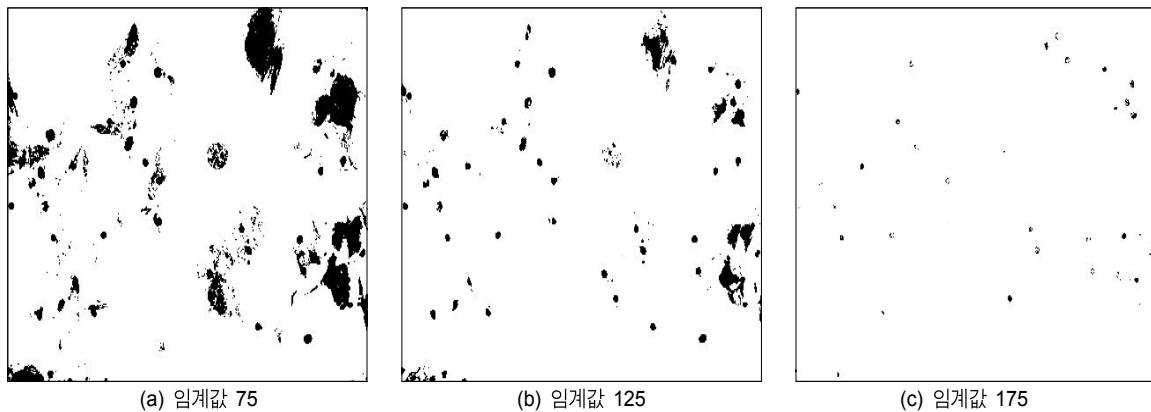


그림 13. 임계값 변화에 따른 세포 사진 비교

Fig. 13. Pictures resulted from the preprocessing module with various threshold values

다음 실험에서는 제안하는 학습 모델에서 임계값을 변화하면서, 이 값의 변화가 인식률에 어떤 영향을 끼치는가를 분석해 보았다. 위의 그림 13에서는 임계값을 다양하게 변화시키면서 전처리된 영상이 어떻게 변화되는지 확인하였다. 그림 13 (a), (b), (c)에서와 같이 임계값을 변화시키면, 전처리된 영상의 특성도 크게 달라지게 된다. 이렇게 다양한 임계값을 사용했을 경우, 본 논문에서 제안된 모델의 인식률 성능을 정리하면 표 3과 같다. 표 3에서 알 수 있듯이, 임계값을 75으로 주었을 때는 22.2%의 인식률이 측정되었고, 임계값을 175으로 설정하였을 경우는 인식률이 15.4%로 매우 낮았다. 이 결과는 적절하지 못한 임계값을 주었을 때는 이미지 내부의 정보가 상당히 손상되기 때문에 인식률이 대폭 감소하는 것을 의미한다.

표 3. 임계값 변화에 따른 인식률의 변화

Table 3 Recognition ratios according to threshold values

Threshold	75	125	175
Dilated convolution with threshold	22.2%	70.9%	15.4%

VI. 결 론

본 논문에서는 고해상도 영상의 분류를 위해서 dilated convolution이 사용되는 CNN 구조를 제안하였다. 이때, 계

산량을 추가로 감소시키고 성능을 향상시키기 위해서 문턱치를 사용한 전처리 기법을 사용하였다. 본 논문의 제안된 알고리즘을 사용함으로써 기존의 기술들보다 인식률이 크게 향상된 학습 모델을 얻을 수 있었다.

참 고 문 헌 (References)

- [1] Andre Esteva1, Brett Kuprel, Roberto A. Novoa, Justin Ko, Susan M. Swetter, Helen M. Blau & Sebastian Thrun, “Dermatologist-level classification of skin cancer with deep neural networks”, *Nature*, 542, pp. 115-118, Feb., 2017. [DOI: 10.1038/nature21056]
- [2] Fabio Alexandre Spanhol, Luiz S. Oliveira, Caroline Petitjean, “Breast cancer histopathological image classification using Convolutional Neural Networks”, 2016 International Joint Conference on Neural Networks (IJCNN), pp. 2560-2567, July, 2016. [DOI: 10.1109/IJCNN.2016.7727519]
- [3] Korsuk Sirinukunwattana, Shan E Ahmed Raza, Yee-Wah Tsang, “Locality Sensitive Deep Learning for Detection and Classification of Nuclei in Routine Colon Cancer Histology Images”, *IEEE transactions on Medical Imaging*, Vol. 35, Issue 5, pp. 1196-1206, May 2016. [DOI: 10.1109/TMI.2016.2525803]
- [4] Henschen, D, “IBM's Watson could be healthcare game changer.” November, 2013. Retrieved from <http://www.informationweek.com/software/information-management/ibms-watson-could-be-healthcare-game-changer/d/d-id/1108608>
- [5] Dorrier, J, “Exponential medicine: Deep learning AI better than your doctor at finding cancer.”, November, 2015. Retrieved from <https://singularityhub.com/2015/11/11/exponential-medicine-deep-learning-ai-better-than-your-doctor-at-finding-cancer/#sm.0001ivo53nz3nd33wx01kpv5qtzbs>
- [6] Alex Krizhevsky, “Convolutional Deep Belief Networks on CIFAR-10”, pp. 1-9, 2010, retrieved from <https://www.cs.toronto.edu/~kriz/>

conv-cifar10-aug2010.pdf

- [7] Aditya Golatkar, Deepak Anand, Amit Sethi, "Classification of Breast Cancer Histology using Deep Learning", International Conference Image Analysis and Recognition, Springer, pp. 837-844, June, 2018. [DOI https://doi.org/10.1007/978-3-319-93000-8_95]
- [8] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich, "Going Deeper with Convolutions", Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, pp. 1-9, June 2015. [DOI: 10.1109/CVPR.2015.7298594]
- [9] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, Zbigniew Wojna, "Rethinking the Inception Architecture for Computer Vision", Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, pp. 2818-2826, 2016. [DOI: 10.1109/CVPR.2016.308]
- [10] C Szegedy, S Ioffe, V Vanhoucke, AA Alemi, "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning", Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, pp. 4278 - 4284, 2017.
- [11] Kim, Kwang-Bae, "Detection and Recognition of Uterine Cervical Carcinoma Cells in Pap Smear Using Kapur Method and Morphological Features", The Journal of the Korea Information and Communications Society, 11th Issue, No. 10, pp. 1992-1998, October, 2007.
- [12] Krizhevsky, Alex, and Geoffrey Hinton, "Learning Multiple Layers of Features from Tiny Images", Technical report, University of Toronto, Vol. 1, No.4, pp.7, April, 2009.
- [13] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, Alan L. Yuille, "Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs", Proceedings of ICLR, pp. 1-14, May, 2015
- [14] G Fu, C Liu, R Zhou, T Sun, Q Zhang, "Classification for High Resolution Remote Sensing Imagery Using a Fully Convolutional Network", Remote Sensing, vol. 9, Issue 5, pp.1-21, May, 2017. [doi:10.3390/rs9050498]
- [15] Vincent Dumoulin and Francesco Visin, "A guide to convolution arithmetic for deep learning", arXiv 2016, arXiv:1603.07285, pp. 1-31, March 2016.
- [16] Wuzhen Shi, Feng Jiang, Debin Zhao, "Single image super-resolution with dilated convolution based multi-scale information learning inception module", 2017 IEEE International Conference on Image Processing (ICIP), pp. 977-981, Sept., 2017. [DOI: 10.1109/ICIP.2017.8296427]
- [17] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, Jiaya Jia, "Pyramid Scene Parsing Network", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6230 - 6239, July 2017. [DOI: 10.1109/CVPR.2017.660]
- [18] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, Alan L. Yuille, "DeepLab: Semantic Image Segmentation with Deep Convolution Nets, Atrous Convolution, and Fully Connected CRFs", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 40, Issue 4, pp. 834-848, April, 2018.

저자 소개

서영민



- 2014년 3월 ~ 현재 : 세종대학교 전자정보통신공학과 학사과정
- ORCID : <https://orcid.org/0000-0003-3898-8221>
- 주관심분야 : 딥러닝, 인공지능 알고리즘, 영상처리

한종기



- 1992년 : KAIST 전기및전자공학과 공학사
- 1994년 : KAIST 전기및전자공학과 공학석사
- 1999년 : KAIST 전기및전자공학과 공학박사
- 1999년 3월 ~ 2001년 8월 : 삼성전자 DM연구소 책임연구원
- 2001년 9월 ~ 현재 : 세종대학교 전자정보통신공학과 교수
- 2008년 9월 ~ 2009년 8월 : University California San Diego (UCSD) Visiting Scholar
- ORCID : <https://orcid.org/0000-0002-5036-7199>
- 주관심분야 : 비디오 코덱, 영상 신호처리, 정보 압축, 방송 시스템