

폐 결절 검출을 위한 합성곱 신경망의 성능 개선

김한웅 · 김병남 · 이지은 · 장원석 · 유선국

연세대학교 의과대학 의학공학교실

Performance Improvement of Convolutional Neural Network for Pulmonary Nodule Detection

HanWoong Kim, Byeongnam Kim, JeeEun Lee, Won Seuk Jang and Sun K. Yoo

Department of Medical Engineering, Yonsei University College of Medicine

(Manuscript received 25 July 2017 ; revised 13 September 2017 ; accepted 11 October 2017)

Abstract: Early detection of the pulmonary nodule is important for diagnosis and treatment of lung cancer. Recently, CT has been used as a screening tool for lung nodule detection. And, it has been reported that computer aided detection(CAD) systems can improve the accuracy of the radiologist in detection nodules on CT scan. The previous study has been proposed a method using Convolutional Neural Network(CNN) in Lung CAD system. But the proposed model has a limitation in accuracy due to its sparse layer structure. Therefore, we propose a Deep Convolutional Neural Network to overcome this limitation. The model proposed in this work is consist of 14 layers including 8 convolutional layers and 4 fully connected layers. The CNN model is trained and tested with 61,404 regions-of-interest (ROIs) patches of lung image including 39,760 nodules and 21,644 non-nodules extracted from the Lung Image Database Consortium(LIDC) dataset. We could obtain the classification accuracy of 91.79% with the CNN model presented in this work. To prevent overfitting, we trained the model with Augmented Dataset and regularization term in the cost function. With L1, L2 regularization at Training process, we obtained 92.39%, 92.52% of accuracy respectively. And we obtained 93.52% with data augmentation. In conclusion, we could obtain the accuracy of 93.75% with L2 Regularization and Data Augmentation.

Key words: Pulmonary nodule detection, Convolutional Neural Network, Machine learning

237

1. 서 론

폐암은 전 세계적으로 가장 사망률이 높은 암의 하나이다. 국립암센터 통계에 따르면 폐암은 2014년 암 발생 순위에서 위암 다음으로 높은 발생율을 보였으며, 인구 10만명당 34.4명의 사망자로 가장 높은 암사망률을 보이고 있다. 또한 폐암의 경우 생존율도 낮아서, 2010년부터 2014년까지 전체 암환자의 5년 상대 생존율은 70.3%인 반면 폐암환자

의 5년 상대 생존율은 25.1% 밖에 되지 않는다.

이러한 폐암 환자의 생존율은 조기 진단의 여부에 크게 영향을 받는다. 그리고 폐암 조기 진단에 중요한 임상적 지표의 하나로 고립성 폐 결절이 있다. 최근 고립성 폐 결절의 관측은 대부분 CT 영상을 통해 이루어지며, Awai 등의 연구에서[1] CAD 시스템을 통해 판독자의 폐 결절 검출능력이 향상 될 수 있음이 보고되었다. 이에 따라 방사선 전문의를 보조 할 수 있는 CAD 시스템의 개발 또한 중요한 분야로 인식되고 있다. CAD 시스템은 의료영상에서 병변으로 의심되는 영역을 자동으로 검출하며, 일반적으로 폐 영상에 대하여는 결절 영역을 찾는 단계, 커널 설계 및 특징 추출단계, 결절분류단계로 구성된다.

한편, 인공 신경망의 하나인 합성곱 신경망은 최근 영상 인식 분야에서 높은 정확도를 보이며 기존의 기계학습 알고

Corresponding Author : Sun K. Yoo
Department of Medical Engineering, Yonsei University College of Medicine
TEL: +82-2-2228-1919 / E-mail: sunkyoo@yuhs.ac
본 연구는 산업통상자원부 및 한국산업기술평가관리원의 산업핵심 기술개발사업(MOTIE/KEIT 10048528)의 일환으로 수행하였음.

리즘으로 해결하지 못한 문제를 해결하는데 사용되고 있다. 예로, MNIST 데이터셋을 분류하는 합성곱 신경망의 경우 약 0.3%[2]의 오차율을 보이고 있으며, cifar10 데이터셋을 분류하는 합성곱 신경망의 경우에도 사람이 직접 분류 하였을 때의 오차율인 6%[3]보다 낮은 3.5% 이하의 오차율을 보이고 있다[4].

이에 따라 합성곱 신경망을 일반적인 영상인식 문제 외의 의료영상에 적용하려는 연구들이 진행되고 있다. 예를 들어, Wei Li 등은 기존의 CAD 시스템의 특징추출 단계에서 커널을 직접 설계해야 하는 어려움을 합성곱 신경망으로 대체 하였으며[5], Qing Li 등은 간질성 폐질환(interstitial lung disease)을 갖고있는 폐 영상 분류에서 커널 설계에 기반한 기존의 방법들보다 합성곱 신경망이 더 높은 정확도를 나타냄을 보였다[6]. 그러나 제안된 합성곱 신경망 모델들은 두 개 이하의 합성곱 레이어로 구성되어 있어 분류 성능에 한계를 갖고 있다. 이에 본 연구에서는 기존 CAD 시스템의 커널 설계 단계를 대신함과 동시에 향상된 정확도를 보이는 합성곱 신경망 모델을 제안한다. 본 연구에서 제안하는 합성곱 신경망 모델은 총 14개의 레이어로 구성되어 있으며 모델의 정확도를 향상시키기 위하여 과적합을 완화시키기 위한 방법들을 바탕으로 학습되었다.

들의 판독결과와 함께 본인의 판독결과를 검토하여 최종 판독을 내린다. 판독결과 3 mm 이상의 크기를 갖는 결절에 대하여는 각 방사선 전문의들이 그린 외곽선이 포함되어 있으며, 3 mm 이하의 결절과 결절이 아닌 병변에 대하여는 해당 결절의 위치를 하나의 좌표로 제공한다.

2. Preprocessing

LIDC-IDRI 데이터셋에 포함 되어있는 흉부 CT 영상들은 다양한 CT 영상장비로 수집되었다. 따라서 각 영상들을 합성곱 신경망의 학습과 평가에 사용하기 전에 그림 1과 같이 전 처리 과정을 수행하였다.

우선 그림 1(b)와 같이 대상 영상들의 픽셀 값을 하운스필드 단위로 정규화 한다. 하운스필드 단위는 CT 영상에서 각 픽셀의 x-ray 감쇠정도를 나타내는 양적 지표이다. 예를 들어, CT로 촬영한 영상에서 공기에 해당하는 픽셀은 -1000의 하운스필드 단위 값을 갖고 물에 해당하는 픽셀은 0의 하운스필드 단위를 갖게 된다. 본 연구에서는 LIDC-IDRI 데이터셋에 포함된 영상들의 각 픽셀 값에 DICOM tag로부터 얻을 수 있는 Rescale Slope 값을 곱한 뒤, Rescale Intercept 값을 더하여 하운스필드 단위 값을 갖는 영상으로 변환하였다.

하운스필드 단위계로 정규화 된 CT 영상은 grayscale 영상으로 표현된다. 이 때 특정 범위의 하운스필드 값을 grayscale로 변환하기 위하여 window level과 window width를 선택한다. window level은 grayscale로 나타낼 하운스필드 유닛 값의 중간 값을 말하며 window width는 특정 이미지에서 grayscale로 나타낼 하운스필드 유닛의 크기를 말한다. 본 연구에서는 window level과 window width를 일반적인 폐 CT 영상 디스플레이에 사용되는 -600과 1500으로 각각 설정하였다[8]. 따라서 그림 1(c)에 나타난 것과 같이 -1350 이하의 HU 값을 갖는 픽셀은 모두 검정색으로 나타나며 150 이상의 HU 값을 갖는 픽셀은 모두 흰색으로 표현된다. 추가적으로 각 영상들의 Pixel Spacing 값이 모두 다르므로 그림 1(d)와 같이 이를 보정하기 위하

II. 연구 방법

1. LIDC-IDRI Dataset

본 연구에서는 합성곱 신경망 학습을 위하여 LIDC-IDRI[7] 데이터셋을 활용한다. LIDC-IDRI 데이터셋은 1010 case의 흉부 CT 영상으로 구성 되어있으며, 각 CT 영상에 대하여 네 명의 흉부 방사선 전문의가 작성한 판독결과를 포함 하고있다. 판독과정은 두 단계로 진행된다. 첫 번째 단계에서는 각 방사선 전문의들이 독립적으로 CT 영상을 검사하고 병변을 “3 mm이상의 결절”, “3 mm미만의 결절”, “결절이 아닌 병변”의 세가지 경우로 분류한다. 그리고 두 번째 단계에서는 각 방사선 전문의들이 다른 방사선 전문의

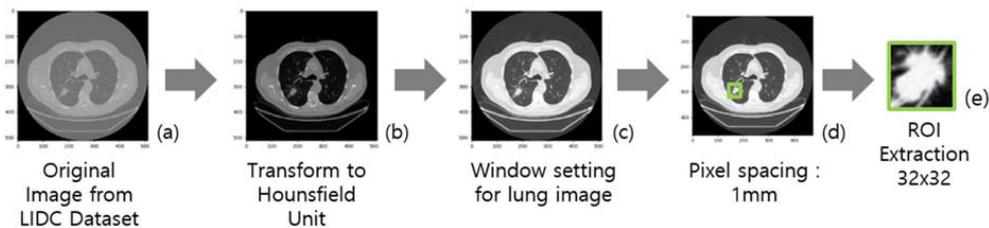


그림 1. LIDC 데이터셋의 전처리 과정 및 ROI 추출 과정. (a) 전처리의 이미지. (b) Hounsfield Unit으로 변경된 이미지. (c) window setting을 적용한 이미지. (d) pixel spacing 을 1 mm로 변경한 이미지. (e) 추출된 폐 결절 이미지.

Fig. 1. Preprocessing and ROI extraction in LIDC Dataset. (a) Original image. (b) Image converted to Hounsfield Unit. (c) Image with lung window setting. (d) Image with pixel spacing of 1 mm. (e) extracted ROI image of a pulmonary nodule.

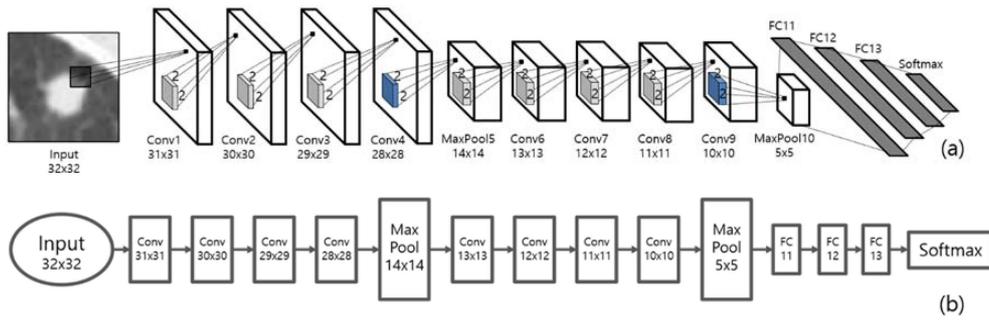


그림 2. 폐 결절 분류를 위한 합성곱 신경망의 구조를 나타내는 그림 (a)와 동일한 합성곱 신경망의 레이어 구조를 나타내는 플로우 차트 (b).
 Fig. 2. Architecture of the Convolutional Neural Network for nodule classification (a) and a flow chart of the layers constituting the Convolutional Neural Network (b).

여 모든 영상에서 한 픽셀이 1 mm의 물리적 크기를 갖도록 전처리 한다.

이렇게 전처리 된 영상에서 방사선 전문의의 판독 결과를 바탕으로 그림 1(e)와 같이 합성곱 신경망의 학습과 평가에 사용할 ROI 영상을 추출한다. ROI 영상은 이후 Data Augmentation 과정을 고려하여 36 × 36 크기로 추출하였다. 이렇게 추출된 ROI 영상은 전체 61404장이며 39760 장의 결절 ROI 영상과 21644장의 비 결절 ROI 영상으로 구성 되어있다.

3. 합성곱 신경망

최근 합성곱 신경망에 대한 연구가 진행됨에 따라 합성곱 신경망 모델의 분류 성능을 향상시키는 다양한 방법이 제안되고 있다. 이에 본 연구에서는 기존의 합성곱 신경망보다 더 많은 합성곱 커널을 사용하며, 더 깊은 층으로 구성된 합성곱 신경망을 사용하였다.

본 연구에서 제안하는 합성곱 신경망 모델은 그림 2와 같이 총 14개의 레이어로 구성된다. 이 중 8개 레이어는 합성곱을 수행하는 레이어이며, 마지막 네 개 레이어는 Fully Connected 레이어이다. 추가적으로 다섯 번째 레이어와 열 번째 레이어는 입력영상을 subsampling 하여 크기를 축소시키는 max pooling 레이어로 구성된다.

첫 번째 에서 네 번째 레이어는 2*2 크기를 갖는 160개의 합성곱 커널로 구성된다. 다섯 번째 레이어는 2*2 크기의 max pooling 레이어이며 입력 영상의 크기를 반으로 줄여 14*14 크기의 영상을 출력한다. 여섯 번째에서 아홉 번째 레이어는 2*2 크기의 합성곱 커널로 구성된다. 이 때 합성곱 커널의 개수는 320개이며, 아홉 번째 레이어 이후에 max pooling 레이어가 뒤따른다. 이들 합성곱 레이어와 max pooling 레이어는 입력된 폐 결절 영상에서 결절과 비 결절 분류에 사용할 특징을 추출하며, 학습과정을 통하여 더 좋은 특징을 선정한다.

열한 번째에서 열 네 번째 레이어는 이전 레이어의 모든

뉴런들과 연결 되어있는 Fully Connected 레이어로 구성된다. 이들 레이어는 각각 512개, 256개, 128개의 노드로 구성되어 있으며, 앞의 학습과정을 통하여 선택된 특징을 바탕으로 결절과 비 결절을 분류한다. 그리고 마지막 출력 레이어는 Softmax 레이어로 입력 영상이 결절 혹은 비 결절일 확률을 출력하는 2개의 노드로 구성 되어있다.

4. 정규화

학습 데이터로 학습과정을 거친 신경망은 일반적으로 학습 데이터에 과적합 된다. 그리고 과적합 된 신경망 모델은 평가 데이터에 대하여 낮은 성능을 보인다. 따라서 과적합을 완화시키기 위한 방법들이 제안되어 왔으며, 그 중 하나로 norm penalty를 활용한 비용함수의 정규화가 있다. 이러한 정규화 방법은 기존의 비용함수에 norm penalty 항을 추가하여 모델의 능력을 제한한다.

일반적으로 사용하는 정규화 방법에는 L1 norm 정규화 방법과 L2 norm 정규화 방법이 있다. L1 norm 정규화 방법은 식 (1)과 같이 가중치 벡터 \vec{w} 에 대한 L1 norm, $\|\vec{w}\|_1$ 에 상수항 λ 를 곱하여 norm penalty 항을 만든다.

$$L1 \text{ Norm penalty, } \lambda \|\vec{w}\|_1 = \lambda \sum_i |w_i| \quad \text{식 (1)}$$

반면, L2 norm 정규화 방법은 식 (2)와 같이 가중치 벡터 \vec{w} 에 대한 L2 norm, $\|\vec{w}\|_2$ 에 상수항 λ 를 곱하여 norm penalty 항을 만든다. L1 norm은 가중치 벡터의 모든 항의 절댓값의 합을 말하며, L2 norm은 가중치 벡터의 모든 항의 제곱의 합을 말한다. 상수항 λ 는 hyperparameter로 비용함수의 정규화 강도를 조절하는 데에 사용한다.

$$L2 \text{ Norm penalty, } \lambda \|\vec{w}\|_2^2 = \lambda \sum_i w_i^2 \quad \text{식 (2)}$$

norm penalty 항이 추가된 비용함수로 신경망 모델을 학습하면 비용함수와 함께 모델의 가중치가 감소하는 방향으

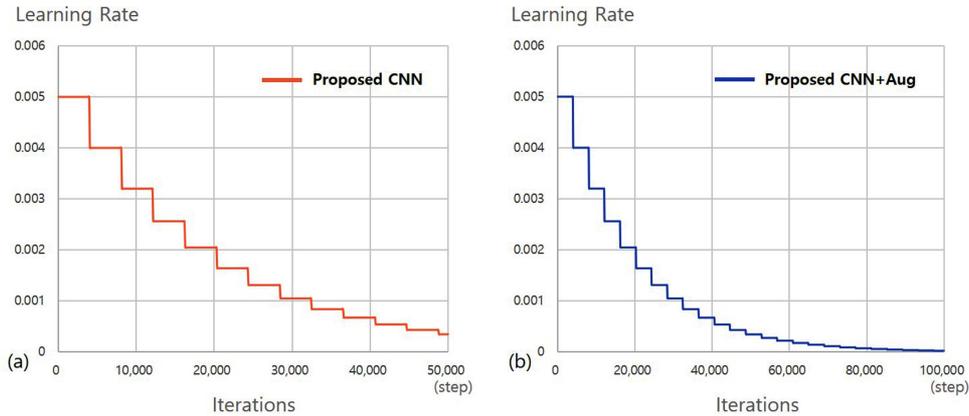


그림 3. Data Augmentation을 적용하지 않은 합성곱 신경망(a)과 Data Augmentation을 적용한 합성곱 신경망(b)의 학습에 따른 learning rate 변화.

Fig. 3. Changes of the learning rate in training process for Convolutional Neural Network without Data Augmentation (a) and with Data Augmentation (b).

로 학습된다. 그리고 이는 모델이 학습 데이터에 과적합 되는 것을 방지한다.

5. Data augmentation

신경망 모델의 과적합 문제는 더 많은 데이터로 모델을 학습하면 완화 시킬 수 있다. 그러나 일반적으로 학습에 사용가능한 데이터의 양은 제한 되어있다. 이를 해결하기 위하여 영상인식 분야에서는 Data Augmentation 기법을 사용한다. 영상인식 분야에서 Data Augmentation은 입력영상에 translation과 같은 연산을 수행하여 적용한다. 이와 같이 Data Augmentation이 적용된 학습 데이터로 신경망 모델을 학습하는 경우 과적합의 완화와 함께 모델의 평가데이터에 대한 분류 능력이 향상된다.

본 연구에서는 앞서 36×36 크기로 추출된 ROI 영상에서 무작위로 32×32 영상을 다시 추출하였다. 이후 수평방향으로 flip 연산을 하여 Data augmentation을 적용하였다. 이는 결과적으로 학습데이터를 32배 늘리는 효과를 가져온다.

III. 연구 결과 및 고찰

본 연구에서 제안하는 합성곱 신경망의 학습과 평가는 LIDC 데이터셋에서 추출한 데이터를 사용하였다. 데이터셋은 전체 61404 장의 ROI 영상으로 구성되어 있으며, 52623장의 학습데이터와 8781장의 평가 데이터로 나누어 실험을 진행하였다. ROI 영상은 LIDC 데이터셋에 포함된 방사선 전문의의 판독결과를 바탕으로 36×36 크기로 추출하였다. 추출된 영상은 Data Augmentation 기법을 적용하여 32×32 크기의 영상으로 변환한 뒤 합성곱 신경망 학

습에 사용되었다. 학습된 합성곱 신경망은 추출된 36×36 영상의 중심영역만 잘라내어 32×32 영상으로 만든 뒤 평가하였다. 실험은 4개의 GTX1080ti와 Ubuntu 16.04의 OS에서 진행하였으며, 알고리즘은 python3.5와 tensorflow1.01로 구현되었다.

본 연구에서는 다음의 6가지 방법에 대하여 실험을 진행하였다. 첫 번째 방법은 제안된 합성곱 신경망에 과적합 완화 방법을 적용하지 않고 학습하였으며, 두 번째 방법은 제안된 합성곱 신경망에 L1 정규화 방법을 적용하여 학습하였다. 세 번째 방법은 L2 정규화 방법을 적용하여 학습하였으며, 네 번째 방법은 Data Augmentation 방법을 적용하여 학습하였다. 다섯 번째 방법은 Data Augmentation과 L1 정규화 방법을 함께 적용하여 합성곱 신경망을 학습하였으며, 마지막으로 여섯 번째 방법은 Data Augmentation과 L2 정규화 방법을 함께 적용하여 학습하였다. 합성곱 신경망을 학습하기 위한 momentum 값은 0.95로 설정하였으며, 초기 learning rate은 0.005로 설정하였다. Learning rate는 매 5 epoch 마다 0.8배로 감소하도록 설정하였다. 학습 step 에 따른 learning rate의 변화는 그림 3에서 확인할 수 있다. 그림 3(a)는 Data Augmentation이 적용되지 않은 모델들의 learning rate 변화를 나타내며, 그림 3(b)는 Data Augmentation이 적용된 모델들의 learning rate 변화를 나타낸다. 제안된 합성곱 신경망은 64 batch size로 학습하였으며, Augmentation이 적용되지 않은 경우 50,000 step을 학습하였다. Data Augmentation이 적용된 경우 학습데이터의 증가에 따라 100,000 step으로 학습 횟수를 증가시켰다. 학습에 따른 모델의 성능 변화는 그림 4에 나타내었다. 그림 4(a)는 Data Augmentation이 적용되지 않은 모델의 학습에 따른 정확도 변화를 나타내며,

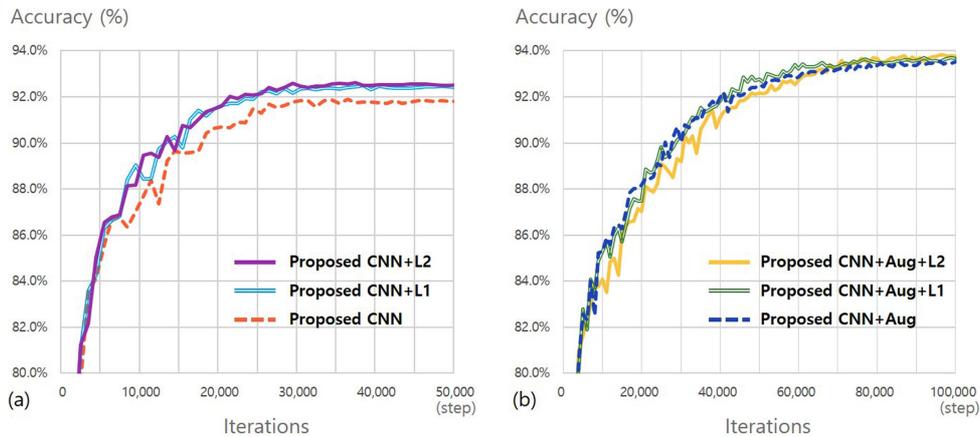


그림 4. Data Augmentation을 적용하지 않은 합성곱 신경망(a)과 Data Augmentation을 적용한 합성곱 신경망(b)의 학습에 따른 분류 정확도 변화.

Fig. 4. Changes of the classification accuracy in training process of Convolutional Neural Network without Data Augmentation(a) and with Data Augmentation (b).

그림 4(b)는 Data Augmentation이 적용된 모델의 학습에 따른 정확도 변화를 나타낸다. 제안된 합성곱 신경망 모델은 50,000 step의 학습 이후 91.79%의 정확도를 보였다. 이후 동일한 합성곱 신경망 모델에 대하여 L1 정규화 기법을 적용한 경우, 92.39%로 정확도 향상을 보였으며, L2 정규화 기법을 적용한 경우에도 92.52%로 정확도 향상을 확인할 수 있다. L1, L2 정규화 기법을 사용하지 않고 Data Augmentation 만을 적용하여 학습한 경우 합성곱 신경망의 정확도는 93.52% 였다. 최종적으로 Data Augmentation과 정규화 기법을 같이 적용한 합성곱 신경망의 정확도는 각각 93.66%와 93.75%였다.

IV. 결 론

본 논문에서는 폐 영상 CAD 시스템에서 결절과 비 결절로 구성된 ROI 영상을 분류하는 합성곱 신경망을 제안하였다. 제안된 합성곱 신경망 모델은 50,000 step의 학습 이후 91.79%의 정확도를 보였다. 과적합을 완화시키기 위하여 L1 정규화 기법과 L2 정규화 기법을 추가하였을 때의 정확도는 각각 92.39%와 92.52%로 향상되었으며 Data-Augmentation 기법을 적용하였을 때 93.52%의 정확도를 보였다. 최종적으로 Data Augmentation과 L1, L2 정규화 방법을 함께 적용하여 학습시킨 합성곱 신경망 모델은 각각 93.66%와 93.75%의 정확도를 보였다.

Reference

[1] Awai K., Murao K., Ozawa A., Komi M., Hayakawa H., Hori, S., & Nishimura, Y. "Pulmonary nodules at chest CT:

effect of computer-aided diagnosis on radiologists' detection performance." *Radiology*, vol. 230, no. 2, pp. 347-352, 2004.

[2] Liang M. and Hu X. "Recurrent convolutional neural network for object recognition." in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, USA, 2015, pp. 3367-3375.

[3] Karpathy, Andrej. "Lessons learned from manually classifying CIFAR-10." *Published online at <http://karpathy.github.io/2011/04/27/manually-classifying-cifar10>*, 2011.

[4] Graham, Benjamin. "Fractional max-pooling." *arXiv preprint arXiv:1412.6071*, 2014.

[5] Li, W., Cao, P., Zhao, D., and Wang, J. "Pulmonary nodule classification with deep convolutional neural networks on computed tomography images." *Computational and mathematical methods in medicine*, vol. 2016, 7, Article ID: 6215085, 2016.

[6] Li Q., Cai W., Wang X., Zhou Y., Feng D. D., and Chen M. "Medical image classification with convolutional neural network." In *Proc. Control Automation Robotics & Vision (ICARCV), 2014 13th International Conference on*. IEEE, Singapore. 2014, pp. 844-848.

[7] Armato S. G., McLennan G., Bidaut L., McNitt-Gray M. F., Meyer C. R., Reeves A. P. and Kazerooni. "The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans." *Medical physics*, vol. 38, no. 2, pp. 915-931, 2011.

[8] Romans Lois E, *Computed tomography for technologists: a comprehensive text*, Baltimore, Maryland: Wolters Kluwer Health, 2010, pp. 36.