Journal of Biomedical Engineering Research 42: 201-210 (2021)
http://dx.doi.org/10.9718/JBER.2021.42.5.201

학 술 논 문

콘볼루션 신경망(CNN)과 다양한 이미지 증강기법을 이용한 혀 영역 분할

안일구 · 배광호 · 이시우*

한국한의학연구원 한의약데이터부

Tongue Image Segmentation Using CNN and Various Image Augmentation Techniques

Ilkoo Ahn, Kwang-Ho Bae and Siwoo Lee*

Korean Medicine Data Division, Korea Institute of Oriental Medicine
(Manuscript received 28 June 2021 ; revised 17 September 2021 ; accepted 5 October 2021)

Abstract: In Korean medicine, tongue diagnosis is one of the important diagnostic methods for diagnosing abnormalities in the body. Representative features that are used in the tongue diagnosis include color, shape, texture, cracks, and tooth marks. When diagnosing a patient through these features, the diagnosis criteria may be different for each oriental medical doctor, and even the same person may have different diagnosis results depending on time and work environment. In order to overcome this problem, recent studies to automate and standardize tongue diagnosis using machine learning are continuing and the basic process of such a machine learning-based tongue diagnosis system is tongue segmentation. In this paper, image data is augmented based on the main tongue features, and backbones of various famous deep learning architecture models are used for automatic tongue segmentation. The experimental results show that the proposed augmentation technique improves the accuracy of tongue segmentation, and that automatic tongue segmentation can be performed with a high accuracy of 99.12%.

Key words: Tongue segmentation, Tongue diagnosis, Image augmentation, Transfer learning, Convolutional neural network

I. 서 론

한의학의 주요 진단법에는 망진(望診), 문진(聞診), 문진(問診), 그리고 절진(切診)이 있으며 그 중 망진은 눈으로 환자를 관찰하는 진단법이다. 망진 중에서 자주 쓰이는 진단법 중 하나는 설진(舌診)이다. 한의학에서 혀에는 내장의 기능과 기혈이 반영되어 있기 때문에 설진은 질병의 경중 정도를 판단하는 중요 진단법으로 수행돼 왔다[1-3].

전통적으로 설진은 관찰자에 의해 주관적인 관찰과 경험에 의해 수행되었기 때문에 관찰자 마다 진단결과가 다를 수 있었다. 이러한 문제를 극복하고자 객관적이고 신뢰할만한 컴퓨터 계산 기반의 시스템에 대한 연구들이 수행되어 왔다[3-6]. 이러한 혀 영상 분석의 필수 단계가 혀 영역 분할(tongue segmentation)이다.

혀 영역을 자동으로 추출하기 위해 다양한 알고리즘들이 시도되어 왔다. Kim[7]등은 혀 영상의 과분할(over-segmentation), 영역 융합(region merging), 곡선 피팅(curve fitting)의 조합으로 혀 영역을 추출하였다. Lee 등[8]은 혀 영상의 HSV 컬러모델의 Hue을 특징(feature)으로 혀 영역을 추출하였다. 하지만 단순히 Hue만을 특징으로 추출하는 것은 오류가 많은 방법이다. Han[9]은 능동 윤곽선 모델(Active Contour Model)을 사용하여 혀 영역을 추출하였다. 능동 윤곽선 모델은 주성분분석(Principal component analysis) 기반의 윤곽선 에지(edge)에 수직방향에 위치한 에지를 찾는 기법

*Corresponding Author : Siwoo Lee
Dasan building #202, Korea Institute of Oriental Medicine, 1672, Yuseong-daero, Yuseong-gu, Daejeon 34054, Republic of Korea
Tel: +82-42-868-9555
E-mail: bfree@kiom.re.kr
본 연구는 한국한의학연구원의 '빅데이터 기반 한의 예방치료 원천기술 개발(KSN2022120)'의 지원을 받아 수행되었음.

으로 이는 혀 윤곽선 주변의 노이즈에 민감한 방법이다.

최근 비약적인 발전으로 각광받고 있는 딥러닝 기술은 컴퓨터비전과 영상처리 분야에서 좋은 성능을 보여주고 있다. 딥러닝 기반 컴퓨터비전 기술은 기존 머신러닝 알고리즘을 대체하면서 갈수록 뛰어난 성능 향상을 보이고 있다. Alexnet[10], vgg[11], inception[12], 그리고 resnet[13]과 같이 이미지 분류에서 우수한 성능을 보인 컨볼루션 신경망(CNN) 모델들 이후 U-Net[14], LinkNet[15], PSPNet[16], Attention U-net[17]과 같은 영상 분할에서 우수한 성능을 보이는 딥러닝 네트워크가 등장했다.

딥러닝이 좋은 성능을 내기 위한 필수 전제는 많은 양의 데이터셋이 있어야 한다는 것이다. 새로 입력된 영상이 학습에 사용된 훈련데이터셋(training set)과 많은 차이를 보이면 새로 입력된 영상을 오분류 할 가능성이 높아진다. 이를 학습에 사용된 훈련데이터셋에 과적합(overfitting)되었다고 하며, 컴퓨터 비전 분야에서 이러한 문제를 해결하기 위한 가장 좋은 해결책은 데이터 증강(data augmentation) 기법이다. 영상 데이터 증강 기법은 훈련데이터셋의 다양한 변이를 발생시켜 훈련된 모델이 새로운 영상도 올바르게 예측하도록 돕는 기법이다. 다양한 변이에는 좌우/상하반전, 영상회전, 색변환, 블러링(blurring), 무작위로 자르기(random crop), 히스토그램 처리(histogram processing), 영상 와핑(image warping) 등 매우 다양하다. 딥러닝의 훈련데이터셋의 학습에서는 이러한 증강기법을 단독으로 사용하지 않고 여러 기법을 조합하여 적용한다. 이 때, 임의의 증강기법을 사용하면 데이터의 속성을 반영하지 못하여 최고성능을 낼 수 있는 가능성이 줄어든다. 따라서 목표하고자 하는 데이터의 특성에 적절한 증강기법의 조합을 잘 찾아내는 것이 중요하다.

본 논문에서는 일반영상에 대해 우수한 분할성능을 보인 U-Net[14], LinkNet[15], PSPNet[16], Attention U-net[17]을 이용하여 혀 영역 분할도 높은 성능으로 수행할 수 있음을 보인다(그림 1 참조). 또한 혀 영상의 특성에 적절한 이미지 증강기법들을 조합하면 보다 높은 성능으로 추출할 수 있음을 보인다. II 연구 방법에서 혀 영상 분할에 적절한 데이터 증강기법을 적용하고 증강된 데이터를 학습할 U-Net[14], LinkNet[15], PSPNet[16], Attention U-net[17] 모델에 대해 알아본다. III 연구결과에서는 데이터 증강기법과 전이 학습의 조합으로 혀 분할의 정확도를 알아본다.

II. 연구 방법

1. 혀 영상분할(tongue segmentation)을 위한 데이터 증강기법

혀 영상은 얼굴영상이나 일반 장면영상과 달리 일상적으로 노출되는 일이 적어 데이터가 부족한 실정이다. 따라서 딥러닝 모델을 훈련시키기 위해서는 기존 데이터의 특성을 유지하면서 새로운 다양성을 만들어내는 데이터 증강 기법을 적용하여 모델의 일반화 성능을 향상시키며 과적합을 예방할 필요가 있다. 다양한 데이터 증강 기법들 중 어떤 증강기법들을 조합하느냐에 따라 모델의 일반화 성능이 좌우된다. 본 연구에서는 혀 영상의 특성에 기반하여 혀 영역 분할 성능에 큰 영향을 미치는 증강기법들을 중심으로 설명한다.

(1) 어두운 영역 처리를 위한 감마보정(gamma correction)

혀 영상 촬영 시 입의 내부영역은 조명이 비치지 않아 어두울 수 있다(그림 2(a)의 노란색 사각형 영역 참조). 어두운 영역으로 인하여 혀 분할 성능이 감소할 수 있어 보정할 필요가 있다. 어두운 영역의 보정에 감마보정(gamma correction)이 적합하다. 감마보정은 영상의 밝기(intensity)를 비선형적으로 변형하여 영상을 보다 쉽게 인지하도록 돕는 기법이다.

$$o = c \cdot i^{\gamma} \tag{1}$$

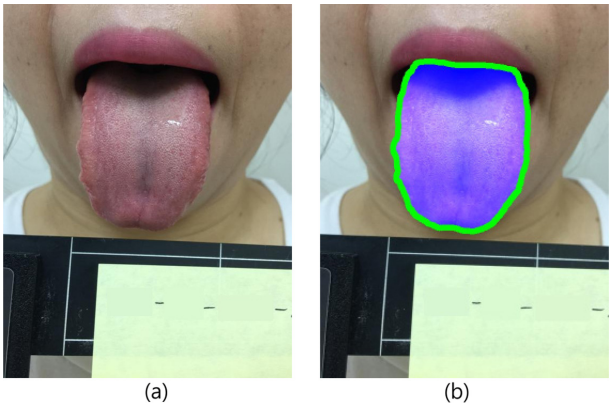

그림 1. 혀 영역 분할 예시 (a) 촬영된 혀 영상 (b) 분할된 혀 영역
Fig. 1. An example of tongue region segmentation (a) tongue image (b) segmented tongue region

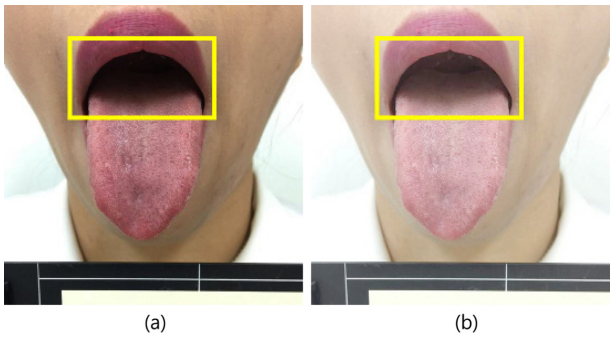

그림 2. 혀 영상의 감마보정 예시. (a) 원본 영상 (b) 감마보정된 영상
Fig. 2. An example of gamma correction of tongue image. (a) original image (b) gamma-corrected image

위 식에서 i와 o는 각각 입력 및 출력 픽셀값이며, c와 γ는 양의 상수이다. 다양한 보정 환경을 제공하기 위해 γ는 일정 범위 내의 무작위(random) 값을 사용할 수 있다.

(2) 치흔설(scalloped tongue)을 위한 탄성변환(elastic transformation)

치흔설은 혀의 가장자리에 울퉁불퉁한 모양의 치아의 흔적이 있는 것을 말한다(그림 3 참조). 한의학에서는 기허, 비허, 습성의 증상으로 보며[18,19], 양방에서는 영양불량, 특히 단백질 결핍이나 혀의 부종을 원인으로 보기도 한다[20].

일반적인 매끈한 혀 영상으로만 훈련데이터셋이 주어질 경우 치흔영역에서는 분할 정확도가 낮아질 수 있다. 치흔설을 훈련데이터셋에 반영하기 위해 탄성변환(elastic transformation)을 수행한다. 탄성변환은 각 픽셀의 수평과 수직 방향에 대해 무작위 강도(stress)의 변위(displacement) $\triangle x$, $\triangle y$를 생성하여 수행된다. $m \times {}^{a}n$ 크기의 영상의 각 픽셀에 대해 수평(x축)과 수직(y축) 방향으로 [-1,1]범위의 무작위 변수를 균등하게(uniformly) 생성한다. 이는 [0,1]범위의 무작위 변수를 균등하게 생성하는 $rand$ 함수를 이용해($rand(m,n) \times {}^{a}2\text{-}1$)으로 표현될 수 있다. 인접한 픽셀이 비슷한 변위를 갖도록 하기 위해 σ의 가우시안 필터를 이용하여 평활화(smoothing)시켜준다. 그 후 파라미터 α를 곱하여 변위의 크기를 조절한다. 수평과 수직방향의 변위를 각각 독립적으로 생성하기 위해 다른 무작위 변수가 생성되며 파라미터 σ와 α은 수평과 수직방향에서 같은 값으로 계산된다.

$$\Delta x = \{G(\sigma) * (rand(m,n) \times 2 - 1)\} \times \alpha \quad (2)$$

$$\Delta y = \{G(\sigma) * (rand(m,n) \times 2 - 1)\} \times \alpha \quad (3)$$

그림 3. 치흔설 예시. 혀 가장자리에 치흔(치아 자국)이 울퉁불퉁한 형태로 나타나 있다(노란색 색 U자 영역)
Fig. 3. An example of scalloped tongue. Teeth marks appear on the edge of the tongue in an uneven shape (the yellow U-shaped region)

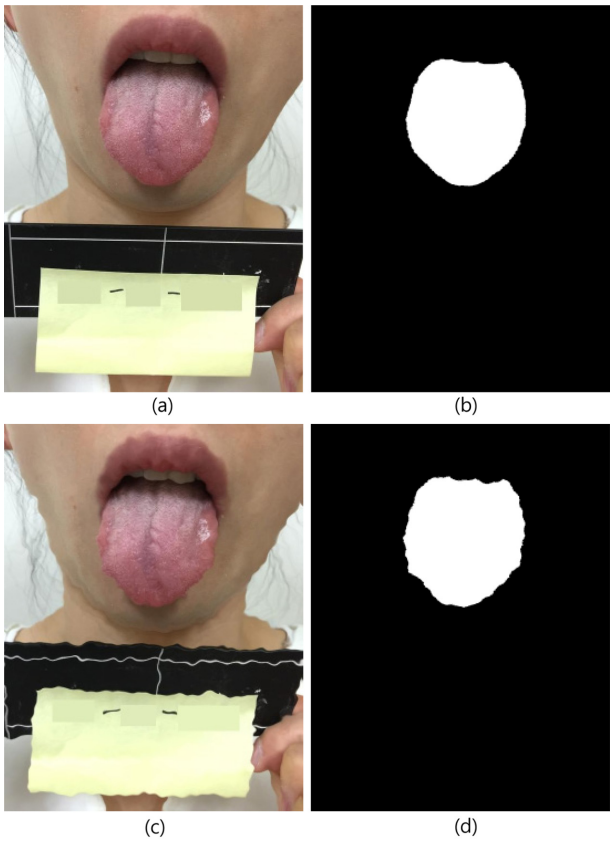

그림 4. 혀 영상의 탄성 변환 적용. (a)(b) 원본영상 (c)(d) 각각의 원본영상에 탄성변환이 적용된 영상
Fig. 4. Applying elastic transformation of tongue image. (a)(b) original images (c)(d) images with elastic transformation applied to each original image

$$I_{\alpha}(i + \Delta x(i,j), j + \Delta y(i,j)) = I(i,j) \quad (4)$$

그림 4는 일반적인 매끈한 혀 영상에 탄성변환을 수행한 것으로, 치흔설을 반영함을 알 수 있다.

이러한 증강기법 외에도 좌우반전(horizontal flip), 회전(rotation), 그리드 왜곡(grid distortion), 무작위 대비(random contrast)가 적용되었다. 훈련셋에 대해 증강기법이 적용된 순서는 좌우반전(horizontal flip), 회전(rotation), 그리드 왜곡(grid distortion), 탄성변환(elastic transformation), 무작위 대비(random contrast), 감마보정(gamma correction) 순이다.

2. 전이학습(transfer learning)과 백본 네트워크(backbone network)를 이용한 혀 영상분할

(1) 전이학습(transfer learning)

깊은 신경망(deep neural network) 모델이 좋은 성능을 내기 위해서는 어느정도 이상의 대용량 데이터가 필수적인 것으로 알려져 있다. 하지만 모든 분야에서는 데이터가 충분하지 않을 수 있으며, 충분하더라도 수집에 많은 시간과 비용이

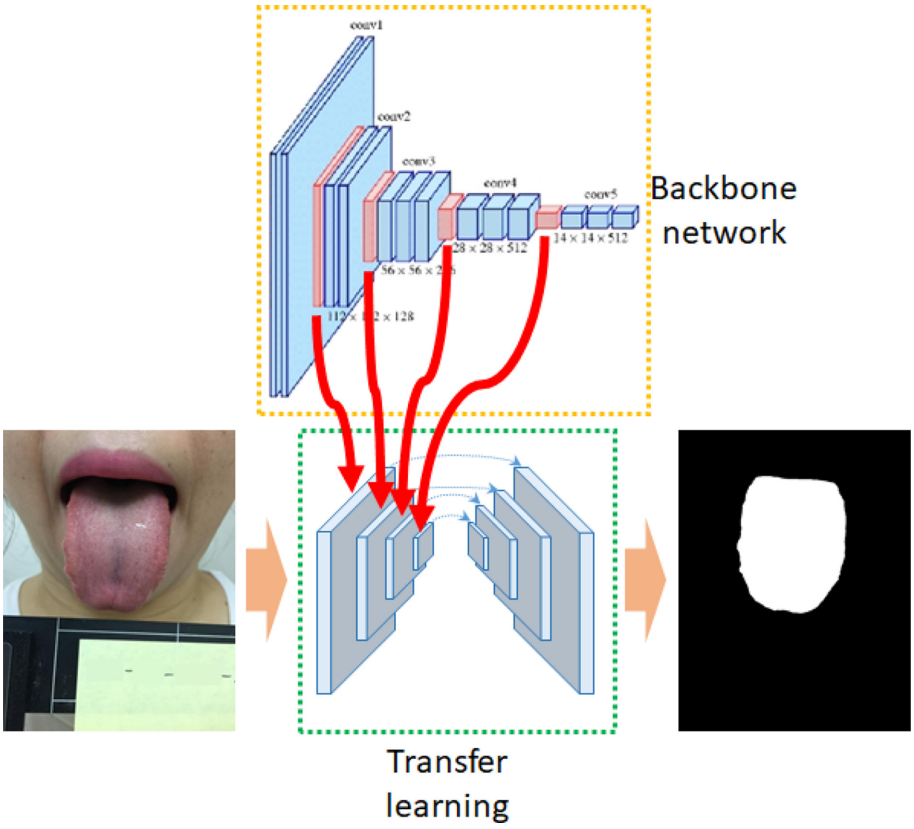

그림 5. 백본네트워크와 전이학습을 이용한 혀 영역 분할
Fig. 5. Tongue segmentation using Transfer Learning with backbone network

들 수 있다. 이를 해결하는 대안으로 전이학습(transfer learning)이 이용되고 있다(그림 5 참조). 전이학습은 사전에 학습된 기존 신경망 모델의 가중치를 현재 모델링하려는 신경망의 초기 가중치로 설정하여 학습하는 기법이다. 기존 신경망 모델의 가중치가 같은 종류(영상)의 데이터로부터 학습된 가중치이기 때문에 새로운 모델을 비교적 빠르게 학습할 수 있으며 대용량의 데이터가 없이도 성능이 좋은 것이 장점이다. 이 때 현재 학습하려는 모델의 데이터가 기존 신경망 모델 학습에 이용된 데이터와 비슷할수록 학습이 빠르고 성능이 좋은 것으로 알려져 있다. 본 연구에서는 영상분할에서 좋은 성능을 보인 U-net[14], Linknet[15], PSPNet[16], 그리고 Attention U-net[17]을 이용한다.

(1-1) U-net

U-net[14]은 Ronneberger 등이 제안한 U자 모양의 Fully-Convolutional Network 모델이다. 본래 의료영상을 위해 개발되었으나 좋은 성능으로 일반영상 분할에도 널리 쓰이고 있다. U-Net은 특징맵(feature map)의 가로와 세로 크기를 점차 줄여나가는 인코더와 특징맵의 가로와 세로의 크기를 늘려나가는 디코더로 구성되며 인코더와 디코더는 네트워크의 중앙을 기준으로 대칭이다(그림 6 참조). 인코더는 3×3 convolution, ReLU 활성화, max-pooling으로 구성된 블록이 4개 존재하며, 디코더는 3×3 convolution, up-convolution을 수행하는 블록 4개로 구성되며, up-convolution을 이용해 수축된 특징맵의 크기를 늘린다. 이 때 up-convolution 후

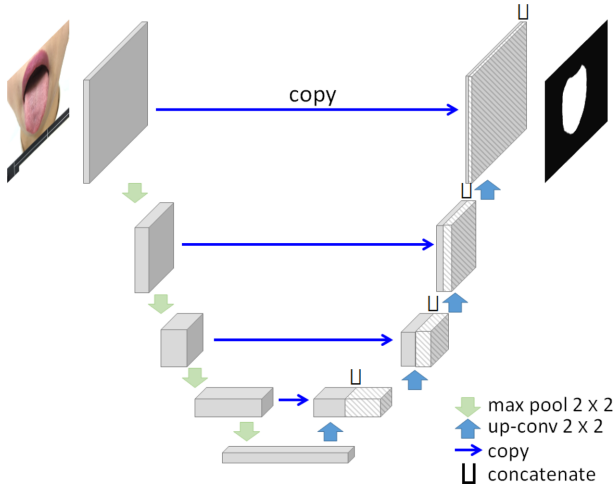

그림 6. U-net 네트워크
Fig. 6. U-net network

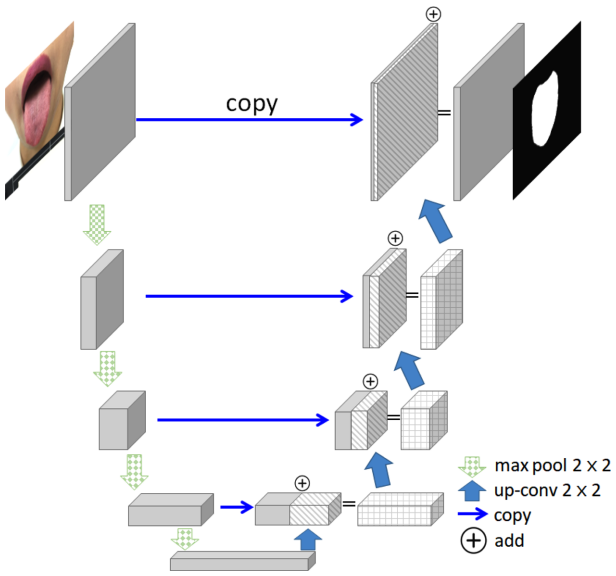

그림 7. Linknet 네트워크
Fig. 7. Linknet network

인코더에서 해당크기의 블록으로부터 특징맵을 스킵 커넥션(skip connection)을 통해 전달받는다. 스킵 커넥션이 없을 경우 객체의 국부적인 형태 정보가 소실되므로 스킵 커넥션을 통해 보다 정교한 분할이 이뤄질 수 있도록 한 것이 U-net의 특징이다.

(1-2) Linknet

Linknet[15]은 U-net과 비슷하지만 일부를 개량한 모델로 크게 2가지가 다르다(그림 7 참조). 첫째, U-net의 인코더의 콘볼루션 구조를 ResNet[13]의 스킵 커넥션 구조로 대체한 것이다. 즉, 연속된 콘볼루션 레이어 뿐 아니라 스킵 커넥션을 통해서도 특징맵이 전달된다. 둘째, U-net에서는 인코더의 특징맵과 디코더의 특징맵이 결합(concatenate)되었으나, Linknet에서는 두 특징맵이 합(add)으로 계산된다. 합으로 계산되므로 차원이 줄어들어 연산량이 줄어드는 효과가 있다. 또한 인코더를 ResNet[13]으로 설계했기 때문에 ResNet 모델에서 계산된 가중치(weights)를 사용하여 빠르게 학습할 수 있다.

(1-3) PSPNet

PSPNet (Pyramid Scene Parsing Network) [16]은 장면 분석(scene parsing)을 목적으로 제안된 모델로 국부적

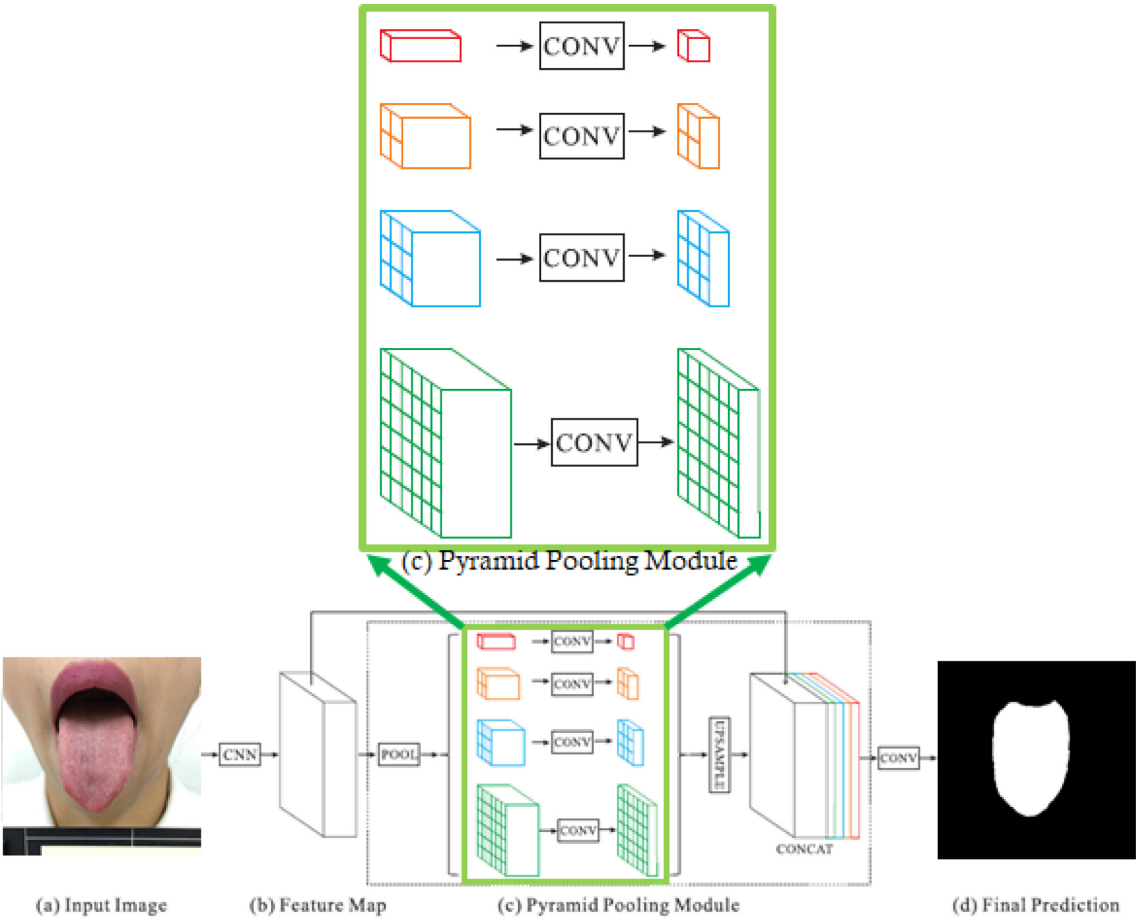

그림 8. PSPNet 네트워크
Fig. 8. PSPNet network

문맥 정보(Local context information)만 갖는 U-net과 달리 전역적 문맥 정보(Global Context Information)를 고려하도록 설계된 것이 가장 큰 특징이다. 예를 들어, 호수 위에 떠 있는 보트를 차로 인식하는 오류가 발생하는 이유는 물 위에 떠 있다는 정보가 없기 때문이다. 또한 고층빌딩의 유리에 반사된 하늘을 빌딩이 아닌 하늘로 인식하는 문제 역시 주변 정보가 있다면 해결될 수 있는 문제이다. 국부적 문맥정보만을 이용하는 일반 CNN 모델과 달리 전역적 문맥정보를 추가해주기 위해서 특징맵을 평균 풀링(average pooling)한 특징맵을 원래 특징맵에 추가한다는 것이 PSPNet의 특징이다. 구체적으로는 입력영상에 대해 그림 8(b)와 같은 특징맵을 ResNet[13]에 Dilated convolution을 사용하여 생성한다. 그리고 그림 8(c) 처럼 풀링연산으로 특징맵을 1×1×N, 2×2×N, 3×3×N, 6×6×N의 크기를 가진 특징맵을 생성한다. 이 특징맵들을 1×1 convolution을 적용하여 channel을 1로 만든다. 즉 특징맵의 정보를 압축한다. 그리고 Bilinear interpolation을 이용하여 각 특징맵의 크기를 처음 입력된 특징맵과 같은 크기로 만든 후 원래의 특징맵에 이어 붙인다.

(1-4) Attention U-net

Attention U-net[17]은 U-Net에 Attention Gate(AG)이라는 구조를 추가하여 성능을 높인 것이 특징인 네트워크로 그림 9(a)에 나타나 있으며, 그림 9(b)는 AG의 내부구조이다. AG의 개념은 영상의 모든 영역을 같은 가중치로 주목(attention)하기보다 목적에 맞는 일부 영역에 보다 주목하겠다는 것이며, AG가 주목할 영역을 학습하게 된다. U-net이 인코더의 특징맵과 디코더의 특징맵만을 사용하는 것과 달리, Attention U-net은 디코더의 특징맵 뿐 아니라 인코더의 특징맵과 디코더의 특징맵에 AG연산을 적용한 결과가 추가된다(그림 9(a) 참조). AG는 인코더 특징맵과 디코더 특징맵에 컨볼루션과 배치 정규화의 결과값을 합한 뒤 ReLU, 콘볼루션, Sigmoid 연산의 결과에 디코더 특징맵과 결합하는 구조이다(그림 9(b) 참조).

(2) 백본 네트워크 (backbone network)

전이학습에 더하여 모델 학습을 빠르게 하는 기법 중 하나는 백본 네트워크(backbone network)이다. 학습하려는 모델의 입력에 데이터셋을 입력시키기 전 모델의 가중치(weights)를 기존 딥러닝 모델의 가중치(weights)로 설정한 후 학습(finetuning)시키는 것이다. 본 연구에서는 네터워크에 따라 백본에 따라 초기화 되는 레이어가 가중치로 초기화되기도 하고 일부분만 초기화되기도 한다. 가능한 많은 레이어가 학습되도록 설정하였다. 본 연구에서는 imagenet에서 학습된 VGG[11], Inception[12], ResNet[13], Densenet[21], Inception-Resnet[22]을 이용하였다.

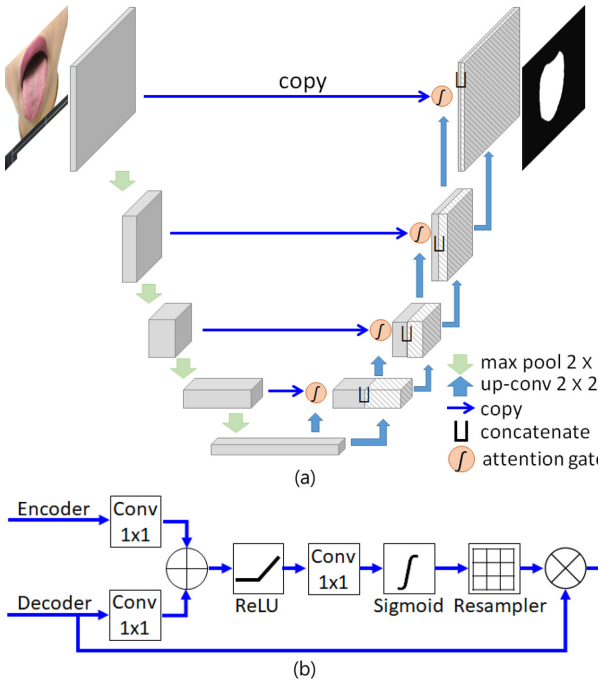

그림 9. Attention U-net 네트워크
Fig. 9. Attention U-net network

III. 연구 결과

1. 연구대상 및 데이터셋

본 연구는 ○○연구원의 ○○정보은행의 혀 영상을 분양받아 진행하였다. 이 혀 영상들은 2개 병원에서 확보한 것으로, 특별한 질환이 없는 일반 피험자와 직장 코호트에 참여한 피험자가 포함되었다. 데이터 수집을 위해 각각의 기관으로부터 기관생명윤리위원회의 승인을 받았다(KOMCIRB-150818-HR-035, GBIRB2015-70).

일반 피험자들의 혀 영상은 2015년 12월에서 2016년 7월까지 X병원을 방문한 52명의 피험자들로부터 확보되었으며, 직장 코호트 피험자들의 혀 영상은 2016년 8월에서 10월까지 Y병원을 방문한 100명의 코호트 피험자들로부터 취득되어 두 기관에서 총 152명의 혀영상이 취득되었다. X병원의 경우, 첫 방문(visit) 촬영 후 6개월 후 방문하여 촬영하여 총 2번 촬영되었으며, Y병원의 경우, 3번의 방문촬영이 있었고 첫 방문 후 2주, 4주 후 방문하였다. 각 방문마다 같은 대상자를 3번 촬영하였으며, 그 중 1번은 플래쉬 카메라 조명을 켜고 촬영하였다. 촬영기기는 삼성 갤럭시 S6를 이용하였으며 혀영상 촬영 전용 어플을 개발하여 촬영하였다. 대상자로 하여금 혀를 최대한 내밀도록 하고 어플 촬영 시 그려진 혀 라인에 맞도록 사마트폰을 혀 위로 위치시킨다. 혀가 화면

안에 위치하는지, 너무 어둡지 않은지 확인한 후, 혀영역을 손가락으로 터치하여 반드시 포커스(focus)가 혀에 맞춰지도록 한 후 촬영하였다. 혀영상 해상도는 가로 세로 1080 ×[a]1920이며 jpg파일로 저장되었다. 촬영 시에는 머리카락이 혀를 가리는 것을 방지하기 위해 헤어밴드를 착용하고 촬영하였다. 대상자 중 일부 대상자가 방문하지 못한 경우가 있어 152명으로부터 총 1528장의 영상을 취득하였다. 모든 혀영상의 혀 영역(ground truth)은 사람이 수동 추출하였다.

2. 평가방법

평가방법으로는 영상분할에서 자주 쓰이는 mean Intersection over Union (mIoU)와 F_1-score를 이용하였다.

$$mIoU=\frac{1}{2}\left(\frac{|F_g\cap F_p|}{|F_g\cup F_p|}+\frac{|B_g\cap B_p|}{|B_g\cup B_p|}\right) \quad (5)$$

$$F_1 \text{ score}=\frac{2|F_g\cap F_p|}{2|F_g\cap F_p|+|B_g\cap B_p|+|B_g\cap B_p|} \quad (6)$$

여기서 F_g와 B_g는 각각 실제 혀영역(ground-truth foreground)과 혀가 아닌 영역(ground-truth background)을 말하고, F_p와 B_p는 각각 모델에 의해 추정된 혀영역과 혀가 아닌 영역을 말한다. |•|는 집합의 원소개수(cardinality), 즉 픽셀수를 나타낸다.

3. 실험결과

훈련셋(training set), 검증셋(validation set), 테스트셋(test set)을 구성함에 있어 같은 대상자의 혀영상이 같은 셋에 포함되지 않도록 대상자를 기준으로 나누었다. 대상자 152명을 무작위로 훈련셋(training set), 검증셋(validation set), 테스트셋(test set)에 각각 91명, 30명, 31명으로 나누었다. 훈련셋, 검증셋, 테스트셋은 각각 433, 153, 178장이며, 훈련셋 433장은 데이터 증강기법을 통해 8660장이 최종 사용되었다. 훈련셋에 적용된 데이터 증강기법과 구체적인 설정은 표 1과 같다.

모델은 Keras 프레임워크를 이용하여 구현되었으며 3.6 GHz Intel Core i7 CPU, 64GB RAM의 윈도우 10 운영체제에서 NVIDIA GTX TITAN X 그래픽카드로 학습되었다. 0.0001의 학습률(learning rate)로 Adam 알고리즘으로 최적화되었으며 Batch 크기는 4~32 범위에서 백본 네트워크마다 다르게 설정하였다. 전이학습과 백본 네트워크에 따른 실험결과를 표 2와 표 3에 나타내었다.

표 1. 실험에 사용된 데이터 증강기법
Table 1. Data augmentation techniques used in the experiment

데이터 증강 기법(data augmentation technique)	설명 (Discription)
좌우반전(Horizontal Flip)	모든 영상에 대해 좌우반전되었다.
회전(Rotation)	모든 영상에 대해 0, 10, 20도로 회전되었다.
그리드 왜곡(Grid Distortion)	모든 영상에 대해 그리드 왜곡이 (-0.3,0.3)범위에서 수행되었다.
탄성 변환(Elastic Transform)	30%의 영상에 대해 식 (2)와 식 (3)의 σ와 α는 각각 6.0과 80으로 설정되었다.
명암 대비(Random Contrast)	40%의 영상에 대해 (-0.4, 0.4)의 범위의 무작위 값으로 명암 대비되었다.
감마 보정(Random Gamma Correction)	50%의 영상에 대해 식 (1)의 c는 1로 설정하였으며 γ는 [0.4, 0.6) 범위에서 무작위로 설정되었다.

표 2. 전이학습과 백본 네트워크에 따른 mIoU 성능. 굵은 글씨는 가장 좋은 결과를 나타낸다
Table 2. mIoU segmentation performance with various transfer learnings and backbone networks. Best results are marked in bold

	U-net		Linknet		PSPNet		Attention U-net	
Backbone	no Aug.	Aug.	no Aug.	Aug.	no Aug.	Aug.	no Aug.	Aug.
vgg16	97.26%	98.27%	97.79%	98.18%	97.50%	98.16%	96.39%	97.62%
vgg19	96.86%	98.23%	97.71%	98.03%	97.61%	98.14%	69.67%	97.64%
inceptionv3	97.78%	98.20%	97.40%	98.09%	97.35%	97.97%	96.51%	97.54%
resnet50	97.88%	98.08%	97.10%	98.02%	97.65%	98.07%	96.49%	97.64%
resnet101	97.86%	98.06%	97.28%	97.95%	97.54%	98.07%	96.72%	97.58%
inceptionresnetv2	97.85%	98.14%	96.54%	98.09%	96.13%	97.90%	91.38%	97.61%
densenet201	98.03%	98.02%	97.91%	98.19%	97.73%	98.14%	62.93%	97.61%
average	97.64%	98.14%	97.39%	98.08%	97.36%	98.07%	87.16%	97.61%

표 3. 전이학습과 백본 네트워크에 따른 F_1 score 성능. 굵은 글씨는 가장 좋은 결과를 나타낸다
Table 3. F_1 score segmentation performance with various transfer learnings and backbone networks. Best results are marked in bold

	U-net		Linknet		PSPNet		Attention U-net	
Backbone	no Aug.	Aug.	no Aug.	Aug.	no Aug.	Aug.	no Aug.	Aug.
vgg16	98.60%	99.12%	98.88%	99.07%	98.72%	99.07%	98.45%	98.82%
vgg19	98.39%	99.10%	98.83%	99.00%	98.77%	99.05%	82.69%	98.82%
inceptionv3	98.87%	99.08%	98.67%	99.03%	98.65%	98.97%	98.48%	98.78%
resnet50	98.92%	99.02%	98.52%	98.99%	98.81%	99.02%	98.49%	98.83%
resnet101	98.91%	99.01%	98.62%	98.96%	98.75%	99.02%	98.58%	98.79%
inceptionresnetv2	98.91%	99.05%	98.09%	99.03%	97.99%	98.93%	95.95%	98.81%
densenet201	99.00%	99.00%	98.94%	99.08%	98.85%	99.06%	77.60%	98.81%
average	98.80%	99.06%	98.65%	99.02%	98.65%	99.02%	92.89%	98.81%

그림 10. 테스트셋에 대해 F_1 score가 가장 낮은 3개와 가장 높은 3개의 분할 결과. 초록색 선은 예측을 나타내고 빨간색 선은 실제 혀 영역을 나타낸다
Fig. 10. The lowest 3 and highest 3 F_1 score segmentation results on test set. Green line indicates the prediction and red line means the ground truth

mIoU의 결과로 U-net의 vgg16 백본네트워크 설정 시 가장 좋은 98.27%의 성능을 보였으며, 그 뒤로 Linknet의 densenet201 백본이 98.19%의 성능을 보였다. F_1 score 결과에서도 U-net의 vgg16가 99.12%로 가장 좋은 성능을 보였다. 데이터 증강 (augmentation)기법을 적용하면 0.26% ~5.92%의 성능향상이 있었다.

정성적인 결과 측정을 위해 테스트셋에 대해 F_1 score 결과상 성능이 가장 나쁜 3장과 성능이 가장 좋은 3장을 그림 10에 보였다. 성능이 가장 좋지 않은 결과에서 알 수 있듯이 입 안의 어두운 영역에서의 분할 결과가 좋지 않음을 알 수 있다. 반면 대부분의 밝은 영역에서는 잘 검출되며 치흔이 있는 영역에서도 분할이 잘 되고 있음을 알 수 있다.

IV. 고찰 및 결론

본 연구는 혀 영상분석의 필수적 단계인 혀영역 분할을 수행한 연구이다. 영역분할에 사용한 CNN 알고리즘은 입력된 데이터(혀 영상)로부터 출력(실제 혀영역)으로의 비선형(non-linear)적이며 최적인 매칭함수 F(x)를 찾는 기법이다. CNN 알고리즘 중 U-net은 의료영상분할에 최적화된 기법으로 up-convolution으로 인해 발생하는 해상도 소실을 스킵 커넥션(skip connection)을 통해 방지하도록 한 것이 특징이다. Linknet은 U-net의 변형으로 인코더의 콘볼루션 구조를 스킵 커넥션으로 대체한 것이 특징이다. PSPNet은 일반 장면 분석(scene parsing)을 위해 제안된 CNN 알고리즘으로 전역적 문맥 정보(Global Context Information)를 포함하도록 설계된 것이 특징이다.

두 기관에서 대상자 152명으로부터 취득한 1528장의 영상을 분할한 결과, mIoU에서는 U-net의 vgg16백본네트워크 설정 시 가장 좋은 98.27%의 성능을 보였으며, F_1 score에서도 U-net의 vgg16 백본네트워크가 99.12%로 가장 좋은 성능을 보였다. mIoU상에서 증강된 데이터에 대해 vgg16이 2개의 네트워크에서 가장 좋은 백본네트워크였고, vgg19와 densenet201가 각각 1개의 네트워크에서 좋은 백본이었다. F_1 score상에서는 vgg16이 2개, resnet50와 densenet201이 각각 1개의 네트워크에서 좋은 백본이었으며, 이를 통해 vgg16은 이미지 분류 및 분할을 수행 시 기본 네트워크로 채택하기에 좋을 뿐 아니라 백본 네트워크로 채택하기 좋은, 잘 설계된 네트워크라고 판단된다. vgg19는 vgg16보다 미세하게 약한 성능을 보이는데, 이는 vgg19가 보다 많은 객체를 분류하기에 적합한 백본이라서 한 개의 객체를 분할하는 본 연구에는 vgg16이 더 적합한 것으로 판단된다. 같은 이유로 resnet50이 resnet101보다 성능이 좋은 것으로 판단되며, densenet 백본의 경우 본 실험에서는 densenet201만 사용했으나 깊이가 덜 깊은 백본으로 수행 시 densenet201보다 좋은 성능을 보일 것으로 판단된다.

데이터 증강을 통한 성능향상은 mIoU와 F_1 score에서 각각 12.34%, 6.92%가 있었다. mIoU상으로 최고성능과 최저성능의 차이가 약 0.53%이고 F_1 score에서 0.25%임을 감안할 때 데이터 증강의 효과가 네트워크 선택보다 더 크다는 것을 알 수 있었다.

U-net과 Linknet보다 최신의 구조인 PSPNet은 mIoU와 F_1 score에서 근소하게 낮은 성능을 보였다. 이는 PSPNet이 일반 장면에서 다양한 객체를 분할하는 경우와 같이 주변 객체와의 정보인 전역적 문맥정보에 강인한 것이 특징인데, 혀 영상분할에서는 분할 대상이 혀만 있기 때문에 PSPNet의 강점을 제대로 발휘하지 못하는 것으로 판단된다.

그림 10의 정성적 결과에서 성능이 낮은 영상들에서 알 수 있는 바와 같이, 분할 오류가 나는 곳은 입 안쪽에 위치한 영역이다. 구강 안쪽 영역에서는 조도가 너무 낮아 육안으로도 분별이 어렵다. 이는 테스트영상에 대해서도 감마보정을 전처리로 수행하면 보다 좋은 성능을 보일 것으로 생각된다. 테스트영상에 전처리를 수행하지 않고 해결하는 대안으로는 네트워크에 감마보정효과를 주는 네트워크를 추가하여 분할까지 수행하는 네트워크 개발을 들 수 있을 것이다.

본 연구의 제한점은 대상자수(152명)가 적다는 것을 들 수 있다. 같은 대상자로부터 여러차례 촬영하여 이미지수가 많지만, 대상자수가 늘어 보다 다양한 혀 영상을 취득한다면 일반화성능이 더 뛰어난 모델링이 가능할 것이다.

본 연구에서는 최근 영상분할에서 많이 이용되고 있는 콘볼루션 신경망(CNN) 기반의 다양한 아키텍쳐를 갖는 모델들을 이용하여 높은 성능으로 혀 영역 분할을 수행할 수 있음을 보였다. CNN기반의 영상분할기술이 잘 쓰이지 않았던 한의학 영상 분야에 적용가능함을 보였다는 점에 있어서 본 연구의 의의가 있다고 생각된다. 특히, 눈에 쉽게 보이는 혀 바깥쪽이 아닌 입 안쪽 어두운 영역에서 분할오류가 빈번하게 발생한다는 임상관찰결과를 바탕으로 이에 적절한 데이터 증강기술(감마보정)을 적용하였다는 점이 다른 혀 영역분할 연구들과 다른 점이라고 할 수 있다. 또한 치흔이라는 혀영상만의 특징에 맞는 데이터 증강기술(탄성변환)도 특징이라 할 수 있다. 후속 연구는 보다 다양한 환경에서 많은 데이터를 추가 수집하여 높은 성능의 전자동 분할 시스템 개발이 될 것이다. 또한 분할 성능개선뿐만 아니라 휴대용 단말기에 사용가능하도록 학습 파라미터의 용량을 제한하면서도 성능이 유지되는 모델 개발이 중요할 것으로 판단된다. 이러한 연구의 수행과 함께 혀 영역 추출 결과를 이용하여 혀의 색(color)과 형태(shape), 질감(texture)을 이용한 CNN기반의 분류[23]가 될 것이다. 이 분류의 목표변수는 설태, 체질, 한열이 될 수 있다.

References

[1] 손지희, 김진성, 박재우, 류봉하. 설진의 표준화를 위한 제언 : 설태 후박의 진단기준을 중심으로. 대한한방내과학회지. 2012;33(1):1-3.

[2] 신윤진, 김윤범, 남혜정, 김규석, 차재훈. 설진(舌診)의 진단적 의의에 대한 문헌고찰. 한방안이비인후피부과학회지. 2007;20(3):118-26.

[3] Kim GH, Park GM. 설진의 과거와 미래 전망. The Magazine of the IEIE. 2010;37(7):62-71.

[4] 최민, 양동민, 이규원. 미각 영역별 설색 분석을 이용한 디지털 설진 시스템 개발. 한국정보통신학회논문지. 2015 Feb; 19(2):428-34.

[5] 은성종, 김재승, 김근호, 황보택근. 설진 유효 분석을 위한 혀의 기하정보 추출 방법. 한국콘텐츠학회논문지. 2011 Dec; 11(12):522-32.

[6] 박진웅, 강선경, 김영운, 정성태. ASM과 SVM 을 이용한 설진 시스템 개발. 한국컴퓨터정보학회논문지. 2013 Apr;18(4):45-55.

[7] 김근호, 도준형, 유현희, 김종열. 설진 유효 영역 추출의 시스템적 접근 방법. 전자공학회논문지-SC. 2008 Nov;45(6):123-31.

[8] 이민택, 이규원. 영역 특징 학습을 이용한 혀의 자동 영역 분리 및 한의학적 설진 시스템. 한국정보통신학회논문지. 2016 Apr;20(4):826-32.

[9] 한영환. 능동 윤곽선 모델을 이용한 혀 영역의 검출. 재활복지공학회논문지. 2016 May;10(2):141-6.

[10] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems. 2012;25:1097-105.

[11] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. 2014 Sep 4.

[12] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. InProceedings of the IEEE conference on computer vision and pattern recognition 2015 (pp. 1-9).

[13] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. InProceedings of the IEEE conference on computer vision and pattern recognition 2016 (pp. 770-778).

[14] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. InInternational Conference on Medical image computing and computer-assisted intervention 2015 Oct 5 (pp. 234-241). Springer, Cham.

[15] Chaurasia A, Culurciello E. Linknet: Exploiting encoder representations for efficient semantic segmentation. In2017 IEEE Visual Communications and Image Processing (VCIP) 2017 Dec 10 (pp. 1-4). IEEE.

[16] Zhao H, Shi J, Qi X, Wang X, Jia J. Pyramid scene parsing network. InProceedings of the IEEE conference on computer vision and pattern recognition 2017 (pp. 2881-2890).

[17] Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, Mori K, McDonagh S, Hammerla NY, Kainz B, Glocker B. Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999. 2018 Apr 11.

[18] 이수정, 백상인, 이병권, 이아람, 김광록, 윤현민, 김원일. 男子齒痕舌 변증에 관한 임상적 고찰. Journal of Pharmacopuncture. 2010 Dec;13(4):91-108.

[19] 박세욱, 강경원, 강병갑, 김정철, 김보영, 고미미, 최동준, 조현경, 이인, 설인찬, 조기호. 중풍환자의 변증분형을 위한 설진에 관한 연구. 동의생리병리학회지. 2008 Feb;22(1): 262-6.

[20] 彭淸華. 望診. 서울, 청홍(지상사). pp 247, 2007.

[21] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. InProceedings of the IEEE conference on computer vision and pattern recognition 2017 (pp. 4700-4708).

[22] Szegedy C, Ioffe S, Vanhoucke V, Alemi AA. Inception-v4, inception-resnet and the impact of residual connections on learning. InThirty-first AAAI conference on artificial intelligence 2017 Feb 12.

[23] 박예랑, 김영재, 정준원, 김광기. 내시경의 위암과 위궤양 영상을 이용한 합성곱 신경망 기반의 자동 분류 모델. Journal of Biomedical Engineering Research. 2020;41(2):101-6.