Check for updates

Journal of KIIT. Vol. 18, No. 3, pp. 127-137, Mar. 31, 2020. pISSN 1598-8619. eISSN 2093-7571
http://dx.doi.org/10.14801/jkiit.2020.18.3.127

뼈 나이 진단을 위한 대비 강화 및 슈퍼픽셀 기반 심층 특징 퓨전

김민정*, 손창환**, 추현승***

Deep Feature Fusion Based on Contrast Enhancement and Superpixels for Bone Age Diagnosis

Min-Jeong Kim*, Chang-Hwan Son**, and Hyunseung Choo***

이 논문은 2017년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임.
(No. 2017R1D1A3B03030853)

요 약

뼈 나이는 성장기 아동의 발육 상태를 예견하는 객관적인 지표로 널리 사용되고 있다. 하지만 기존의 뼈 나이 진단은 영상의학과 전문의가 수작업으로 진단하기 때문에 시간이 오래 걸리는 단점이 있다. 따라서 본 논문에서는 뼈 나이 진단 자동화를 위한 대비 강화 및 슈퍼픽셀 기반의 심층 특징 퓨전 기법을 제시하고자 한다. 핵심 아이디어는 X선 영상에서 뼈 나이에 대한 정보를 담고 있는 영역은 한정적이며, 대비가 낮은 X선 영상에서 추출된 특징은 최종 뼈 나이 진단 성능의 저하를 가져올 수 있다는 것이다. 이러한 아이디어를 딥러닝 프레임워크에 반영하기 위해, 제안한 심층 학습 모델은 두 종류의 심층 합성곱 신경망으로 구성된 두 개의 스트림을 형성한다. 실험 결과를 통해서 제안한 대비 강화 및 슈퍼픽셀 기반 심층 특징 퓨전 기법이 기존의 전이 학습 기반의 심층 학습 기법에 비해 약 9.7개월에서 약 8개월 수준까지 뼈 나이의 정확도를 높일 수 있었다.

Abstract

Bone age is widely used as an objective indicator for predicting the developmental status of growing children. However, the existing bone age diagnosis is a long time because it is manually diagnosed by radiologists. Therefore, this paper proposes a deep feature fusion based on contrast enhancement and superpixels for bone age diagnosis. The key idea is that the area containing information on bone ages in X-ray images is limited, and features extracted from low contrast X-ray images can lead to poor final bone age diagnostic performance. In order to reflect these ideas in the deep learning framework, the proposed deep learning model forms two streams composed of two types of deep convolutional neural networks. Based on the experimental results, the proposed deep feature fusion based on contrast enhancement and superpixel method is able to improve the accuracy of bone age from about 9.7 months to about 8 months, compared to the existing deep learning method based on transfer learning.

Key words

feature fusion, bone age, deep convolution neural networks, superpixels, transfer learning

* 군산대학교 소프트웨어융합공학과 학사과정
- ORCID https://orcid.org/0000-0002-5380-9773
** 군산대학교 소프트웨어융합공학과 교수(교신저자)
- ORCID: https://orcid.org/0000-0001-7077-3074
*** 성균관대학교 소프트웨어학과 교수
- ORCID: http://orcid.org/0000-0002-6485-3155

· Received: Nov. 21, 2019, Revised: Jan. 09, 2020, Accepted: Jan. 12, 2020
· Corresponding Author: Chang-Hwan Son
Dept. of Software Convergence Engineering, Republic of Korea, Korea,
Tel.: 063-469-8915, Email: cson@kunsan.ac.kr

Ⅰ. 서 론

뼈가 성장하는 속도는 생활환경이나 정신 연령에 의해서 개인별로 상이하게 나타난다. 이는 달력상의 나이가 동일할지라도 신체 나이에 해당하는 뼈 나이는 다를 수 있음을 의미한다. 뼈 나이는 사춘기 발현, 치아 발달, 키 성장과 밀접한 관련이 있다. 또한 뼈 나이를 추정하면 향후 어른이 되었을 때 저신장증 유무를 판별하는 데 도움이 될 수 있다[1]. 그뿐만 아니라 뼈 나이는 초경 등과 같은 생식계의 발달과 성장호르몬의 기능과도 깊은 연관성이 있다. 이 때문에 생식계 호르몬의 이상 상황 감지, 내분비계의 질환이나 만성질환에 의한 전신 발육 지연 상태를 예견하는 객관적인 지표로 사용되고 있다[2].

뼈 나이 진단은 일반적으로 영상의학과 전문의가 진단 대상자의 양손 또는 좌측 손의 손가락과 손목이 나오도록 X선 영상을 촬영해서 추정한다. 대표적인 진단 기법으로는 GP(Greulich-Pyle) 기법이 있다[3]. GP 기법은 뼈 나이의 기준이 되는 도감(Atlas)과 진단 대상자의 X선 영상을 대조하여 가장 근접한 도감의 뼈 나이를 대상자의 뼈 나이로 진단하는 기법을 말한다. 하지만 GP 기법은 영상의학과 전문의가 진단 대상자의 X선 영상을 도감에 일일이 비교해야 하는 번거로움이 따르며 무엇보다 뼈 나이를 추정하는 데 시간이 오래 걸리는 단점이 있다. 또한, 동일한 X선 영상에 대해 영상의학과 전문의마다 서로 다른 진단을 내릴 수 있다.

따라서 본 논문에서는 기존의 GP 기법과 같이 영상의학과 전문의의 도움 없이 X선 영상으로부터 컴퓨터 비전 기술을 사용해서 뼈 나이를 자동으로 추정할 수 있는 뼈 나이 추정 기법을 제안하고자 한다. X선 영상으로부터 뼈 나이를 진단하기 위해, 고전적인 컴퓨터 비전을 기술을 사용할 수 있다. 예를 들어, 스케일 불변 특징 변환(Scale invariant feature transform)[4], 국부 이진 패턴(Local binary pattern)[5], 스팔스 코드(Sparse code)[6] 등과 같은 수작업 기반의 특징 추출 기법을 사용해서 특징을 추출한 후, 서포트 벡터 머신(Support vector machine)[7]과 같은 분류기를 학습하여 뼈 나이를 추정할 수 있다. 하지만 최신 컴퓨터 비전 기술은 딥 러닝(Deep learning)을 사용하는 게 대세이다. 특히 심층 합성곱 신경망은 영상 분류와 영상 복원에 탁월한 성능을 보이고 있다[8]-[10]. 따라서 본 연구에서도 심층 합성곱 신경망 기법을 사용하고자 한다. 특히 대비 강화 및 슈퍼픽셀 기반 심층 특징 퓨전 기법을 이용한 뼈 나이 진단 기법을 제안하고자 한다. 이를 통해 기존의 수작업 기반의 뼈 나이 진단 기법을 보완함으로써 더욱 신속하고 정교한 의료 진단 서비스를 제공하고자 한다.

1.1 제안한 접근 방법

제안한 기법의 첫 번째 핵심은 뼈 나이를 진단하는 데 있어서 결정적인 단초를 줄 수 있는 관심 영역을 찾아서 뼈 나이를 추정하는 딥 러닝 모델을 제시하는 데 있다. X선 영상에서 뼈 나이를 진단하는 데 중요한 정보를 담고 있는 관심 영역은 손가락과 손목 영역이다. 손가락과 손목을 제외한 배경 영역은 뼈 나이와 관련된 정보가 없다. 따라서 X선 영상 전체에서 특징을 추출해서 뼈 나이를 추정하는 것보다 배경 영역을 제외한 관심 영역에서 뼈 나이를 추정하는 것이 훨씬 더 정확할 것으로 예측된다.

제안한 기법의 두 번째 핵심은 입력 X선 영상이 대비가 낮기 때문에 추출된 특징의 구별성이 떨어지는 문제점을 해결하는 데 있다. 비록 X선 영상에서 배경 영역과 관심 영역을 구분해서 특징을 추출할지라도 X선 영상 자체의 대비가 낮기 때문에 그 구별성이 떨어질 수 있다. 따라서 제안한 기법에서는 이런 문제점을 발견하고 X선 영상의 대비를 높여 추출된 특징의 구별력을 제고하고 최종 뼈 나이 추정 기술의 정확도를 높이고자 한다.

이러한 연구 동기와 아이디어를 딥 러닝 모델을 통해 실현하고자 두 종류의 스트림(Stream)으로 구성된 대비 강화 및 슈퍼픽셀 기반 심층 특징 퓨전 기법을 제안하고자 한다. 여기서 스트림이란 합성곱 계층(Convolution layer), 풀링 계층(Pooling layer) 등으로 구성된 심층 합성곱 신경망 모델을 일컫는다. 두 종류의 스트림에서 첫 번째 스트림은 입력 X선 영상에서 배경 영역과 관심 영역으로 분할된 영상

을 입력으로 받아 심층 특징(Deep features)을 추출하고 반면에 두 번째 스트림에서는 대비가 강화된 입력 X선 영상을 입력으로 받아 심층 특징을 추출한다. 배경 영역과 관심 영역으로 분할하기 위해 K-평균 군집 알고리즘(K-means clustering)[11]과 슈퍼픽셀(Super-pixels) 알고리즘[12]를 사용하고 대비 강화를 위해서는 히스토그램 평탄화(Histogram equalization) [13]을 사용한다. 그리고 두 종류의 스트림에서 추출된 특징을 퓨전하기 위해, 즉 두 종류의 스트림을 결합하기 위해 연결 계층(Concatenation layer)을 도입해서 최종 종단 간 학습(End-to-end learning) 모델을 완성한다.

최근 제안한 접근 방법과 아주 유사한 논문이 발표되었지만[14][15], 몇 가지 점에서 차이가 있다. 첫째, 최근에 발표된 뼈 나이 기법들은 학습 기반으로 배경 영역을 찾기 위해 수작업 기반의 라벨링 작업을 요구한다. 하지만 수작업 기반의 라벨링 영상을 만드는 과정은 많은 시간과 인력을 요구한다. 반면에 제안한 방법은 수작업 기반이 아닌, 슈퍼픽셀과 같은 군집화 알고리즘을 사용해서 배경과 관심 영역을 자동으로 추출하는 방식을 취한다. 둘째, 최근 발표된 뼈 나이 기법들은 기하학적 위치 보정 등과 같은 전처리 작업에 초점을 두고 있을 뿐, 학습 네트워크는 기존의 모델을 그대로 사용한다. 반면에 본 논문에서는 다중 스트림 기반의 학습 네트워크를 새롭게 제시한다. 즉, 슈퍼픽셀과 대비강화 영상으로부터 추출된 특징 맵을 퓨전할 수 있는 학습 네트워크를 새롭게 제시함으로써, 다중 스트림 기반의 특징 퓨전이 최종 특징 구별력 제고와 뼈 나이 진단 성능 향상을 유도할 수 있음을 보이고자 한다.

1.2 기여도

- 본 연구에서는 X선 영상 전체에서 특징을 추출하는 것보다는 배경 영역과 관심 영역을 분리한 다음 특징을 추출하는 접근 전략이 뼈 나이 진단 성능을 더 개선할 수 있다는 새로운 결과를 제시한다. 또한, 대비가 낮은 X선 영상에서 특징을 바로 추출하는 것보다 대비 강화를 적용한 X선 영상에서 특징을 추출하는 게 최종 뼈 나이 진단 성능을 향상할 수 있음을 본 연구를 통해 확인할 수 있다.
- 본 연구에서는 뼈 나이 진단을 위한 대비 강화 및 슈퍼픽셀 기반의 심층 특징 퓨전 기술을 새롭게 제안하고 기존의 전이 학습(Transfer learning)[16] 기반의 딥 러닝 기법보다 뼈 나이 정확도 측면에서 9.7개월에서 약 8개월 수준까지 끌어올릴 수 있음을 보여 준다.
- 본 연구에서 학습된 뼈 나이 진단 전용 딥 러닝 모델을 배포함으로써, 기존의 수작업 기반의 뼈 나이 진단 기술을 자동화하고 진단 소요 시간을 단축하여 보다 신속하고 정교한 의료 서비스를 제공할 수 있을 것으로 기대된다.

II. 관련 연구

2.1 전이 학습 기반 뼈 나이 진단

전이 학습이란 이미지넷(ImageNet)과 같은 대규모 영상 데이터 셋을 대상으로 미리 학습된 파라미터의 값을 다른 응용 분야의 새로운 영상 데이터 셋을 학습할 때 초기치로 사용하는 기법을 말한다[8]. 일반적으로 딥 러닝의 파라미터의 수는 많기 때문에 난수로 생성된 초기치보다는 미리 학습된 파라미터의 값을 활용하면 학습 속도를 높이고 최적의 해를 찾는 데 더 효과적일 수 있다. 대표적인 전이 학습 모델에는 VGG[8], ResNet[17] 등이 있다.

VGG 모델의 구조는 그림 1과 같다. VGG 모델은 영상 입력 계층(Image input layer), 합성곱 계층, 풀링 계층, 정류 선형 유닛 계층(Rectified linear unit layer), 완전연결 계층(Fully-connected layer), 소프트맥스 계층(Softmax layer), 분류 계층(Classification layer)으로 이루어져 있다. 그리고 마지막 완전연결 계층의 출력 수는 1000개의 노드로 이루어져 있다. 즉, 1000개의 부류(Class)를 분류할 수 있다. 하지만 본 연구에서 뼈 나이 진단 기술은 이산적인 라벨 값을 추정하는 영상 분류가 목적이 아니라 연속적으로 변하는 값을 추정해야 되는 회귀(Regression) 문제에 속한다. 따라서 그림 1의 VGG 모델은 뼈 나이 진단을 위해 그대로 사용될 수 없고 일부 계층의 수정이 필요하다[16].

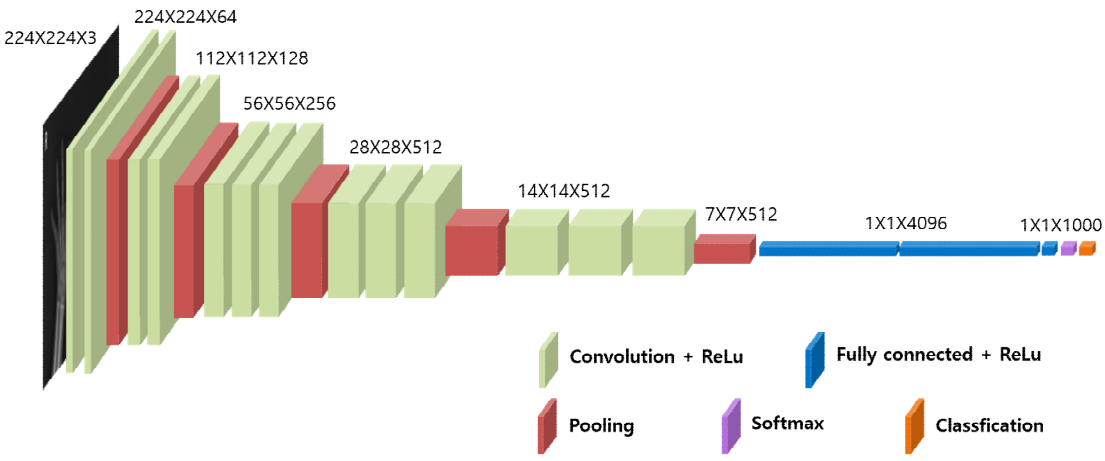

그림 1. VGG 네트워크의 구조 [8]
Fig. 1. Architecture of the VGG network [8]

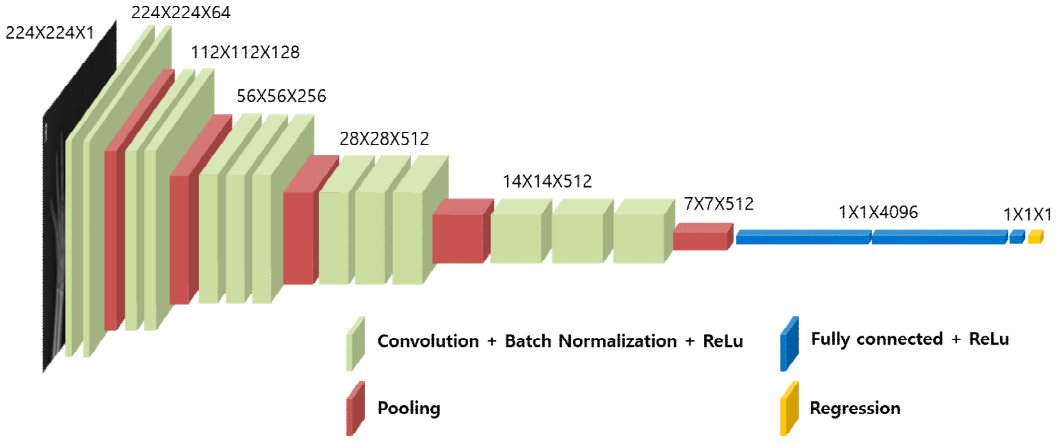

그림 2. 뼈 나이 진단을 위해 수정된 VGG-16 모델의 구조 [16]
Fig. 2. Architecture of the modified VGG-16 network for bone age diagnosis [16]

그림 2는 뼈 나이 진단에 맞게 회귀 모델로 변형된 VGG-16 모델을 보여주고 있다. 첫째, 그림 1과의 차이점은 마지막 세 개 계층인 완전연결 계층, 소프트맥스 계층, 로그 손실 계층을 제거하고 1개의 노드로 구성된 완전연결계층과 회귀 계층(Regression layer)이 추가된 것을 볼 수 있다. 회귀 계층에서는 영상의학과 전문의가 X선 영상을 보고 측정한 뼈 나이 값과 최종 완전연결계층의 추정된 값과의 오차를 계산한다. 본 연구에서 사용된 오차는 평균 제곱 오차(MSE, Mean Squared Error)이다.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - y_i^{'}\right)^2 \quad (1)$$

여기서 n은 총 데이터 개수이다. 그리고 y_i는 영상의학과 전문의가 측정한 뼈 나이이며 y'_i는 마지막 완전연결계층을 통과해 된 뼈 나이 값이다.

둘째, X선 영상은 그레이 영상이기 때문에 입력계층에 입력되는 채널 수도 3채널에서 단일 채널로 변경된 것을 볼 수 있다. 또한, 학습 수렴 속도를 높이기 위해 배치 정규화 계층(Batch normalization layer)을 합성곱 계층과 정류 선형 유닛 계층 사이에 추가한 것도 볼 수 있다. 참고로 참고문헌 [16]의 모델은 특징 맵을 기하학적으로 변형하는 공간 변형 계층(Spatial transformation layer)을 포함하고 있지만 본 연구에서는 X선 영상이 이미 정렬되어 있어서 배제하였다.

대비 강화 및 슈퍼픽셀 기반 심층 특징 퓨전 기법

서론에서 언급했듯이 입력 X선 영상은 대체적으로 낮은 대비를 갖고 있다. 그림 3의 첫 번째 행은 촬영된 X선 영상이다. 그림에서 볼 수 있듯이 X선 영상의 대비가 낮은 것을 관찰할 수 있다. 이런 대비가 낮은 X선 영상에서 특징을 추출한다면 특징의 구별성이 떨어지는 문제가 발생할 수 있고, 이는 최종 뼈 나이 진단의 성능의 저하를 가져올 수 있다. 따라서 제안한 기법에서는 이런 문제점을 해결하고자 입력 X선 영상의 대비를 높여 추출된 특징의 구별력을 제고하고 최종 뼈 나이 추정 기술의 정확도를 높이고자 한다.

입력 영상의 대비를 강화하는 기법으로 히스토그램 평탄화 기법이 널리 사용되고 있다. 물론 레티넥스(Retinex)[18]과 같은 대비 강화 기법도 고려될 수 있다. 히스토그램 평탄화 기법은 입력 영상의 확률 분포의 모양이 균일 확률 분포(Uniform probability distribution)로 만드는 과정을 말한다. 다시 말하면, 영상의 모든 픽셀 값이 균등한 개수를 갖는 히스토그램을 갖도록 만드는 과정이라 볼 수 있다.

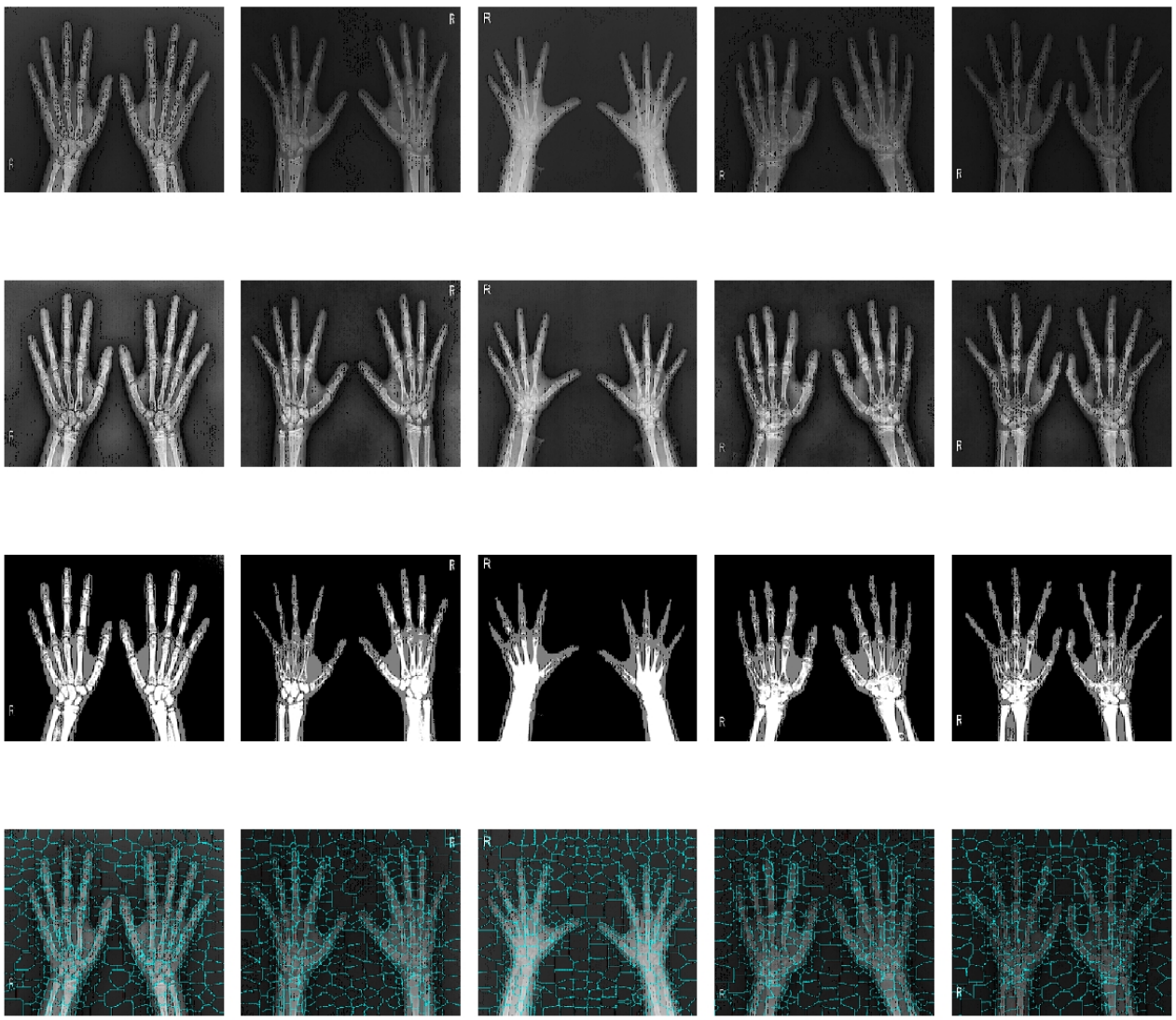

그림 3의 대비가 낮은 영상은 픽셀 값의 분포가 한쪽으로 쏠려 있기 때문에 히스토그램 평탄화를 적용하면 다양한 픽셀 값이 고르게 분포할 수 있으므로 대비가 증가될 수 있다.

본 연구에서 사용된 대비 강화 기법은 대비 제한 적응 히스토그램 평탄화(CLAHE, Contrast-Limited Adaptive Histogram Equalization) 기법이다[13]. 히스토그램 평탄화 기법은 영상의 대비를 강화할 수 있지만 굉장히 밝고 어두운 영역의 대비 증가에 한계가 있다. 따라서 이러한 단점을 보완하기 위해 대비 제한 적응 히스토그램 평탄화 기법이 제안되었다. 이 기법은 기존의 히스토그램 평탄화 기법과는 달리 누적 분포 함수(Cumulative distribution function)를 구할 때 영상 전체를 대상으로 하지 않고 국부영역만으로 국한한다. 그러나 국부 영역에서만 누적 분포 함수를 변환 함수로 사용한다면 잡음의 증폭을 초래할 수 있으므로 그 증폭을 제안할 필요가 있다. 이렇게 국부 영역을 대상으로 변환 함수를 구하고 잡음의 증폭되는 정도를 조절하는 기법이 바 로 대비 제한 적응 히스토그램 평탄화 기법이다.

그림 3의 두 번째 행은 대비 제한 적응 히스토그램 평탄화 기법을 적용한 결과이다. 첫 번째 행의 입력 영상과 비교했을 때, 영상의 국부 영역마다 대비가 증가한 것을 볼 수 있고 잡음의 증폭도 억제된 것을 볼 수 있다. 본 연구에서 히스토그램 평탄화를 사용하는 목적은 단순히 영상 강화의 목적이 아니다. 향후 실험 결과에서 확인하겠지만, 대비가 강화된 영상에서 특징을 추출하는 것이 특징의 구별력을 높이고 최종 뼈 나이 진단 성능까지 제고할 수 있음을 보이는 것에 있다. 즉, 대비 강화가 뼈 나이 추정에 얼마만큼의 성능 향상을 가져올지 그 유효성에 대해 파악하는 것에 있다.

3.2 관심 영역 추출

X선 영상에서 뼈 나이를 진단하는 데 중요한 정보를 담고 있는 영역은 손과 손목 영역이다. 배경 영역은 뼈 나이와 관련된 정보가 없다. 따라서 X선 영상 전체에서 특징을 추출해서 뼈 나이를 추정하는 것보다 배경 영역과 관심 영역에서 추출된 특징의 중요도를 반영할 수 있으면 뼈 나이 진단의 정확도를 높일 수 있을 것으로 기대된다. 따라서 본 연구에서는 입력 X선 영상에서 배경 영역과 관심 영역을 구분하고자 한다. 이를 위해, 대표적인 군집 알고리즘인 K-평균 군집 알고리즘[11]과 슈퍼픽셀 알고리즘[12]를 사용하고자 한다. 물론 관심 영역을 좀 더 섬세하게 분할하기 위해서는 라벨링 작업이 필요하다[14][15][19]-[21]. 예를 들어, 영상 분할이나 바운딩 박스 정보를 영상마다 추출해서 네트워크를 학습할 수 있지만, 상당한 시간이 소요된다. 실제로 좀 더 정확한 분할 영상을 만들기 위해서는 뼈마디 사이의 공간을 분할해야 되기 때문에 라벨링 작업이 상당히 번거롭고 많은 인력 동원이 요구된다. 따라서 본 연구에서는 근사적으로 배경과 관심 영역을 분리할 수 있는 K-평균 군집 알고리즘과 슈퍼픽셀 알고리즘을 사용하고자 한다.

K-평균 군집 알고리즘은 미리 정해진 거리척도, 예를 들어 칼라 유사도(Color similarity)를 사용해서 입력 픽셀을 유사한 그룹으로 묶는다. 반면에 슈퍼픽셀 알고리즘은 K-평균 군집 알고리즘과 유사하지만 거리 척도를 적용하는 범위를 전체 영상이 아닌 국부 영역으로 제한한다. 이를 통해서 픽셀보다 좀 더 큰 슈퍼픽셀이라 불리는 군집을 형성하게 된다. 슈퍼픽셀 알고리즘에 사용된 거리 척도는 픽셀 단위의 공간적인 거리와 칼라 유사도를 결합한 방식을 채택하고 있다.

그림 3의 세 번째 행과 네 번째 행은 각각 K-평균 군집 알고리즘과 슈퍼픽셀 알고리즘을 적용한 결과 영상이다. K-평균 군집 알고리즘에서 군집의 개수를 3개로 설정하였고, 슈퍼픽셀 알고리즘에서 슈퍼픽셀의 개수는 500개로 설정하였다. 그림 3의 세 번째 행에서 볼 수 있듯이 입력 X선 영상이 배경 영역과 관심 영역, 즉 손과 손목을 포함한 영역과 뼈 부분을 구분한 것을 볼 수 있다. 그리고 네 번째 행에서는 입력 X선 영상이 슈퍼픽셀로 분할된 것을 볼 수 있다. 이때 각각의 슈퍼픽셀은 고유한 인덱스 값을 가지고 있기 때문에 배경 영역과 관심 영역을 나눌 수 있게 된다. 여기서 한 가지 강조할 점은 본 연구의 목적이 군집화가 아니라 X선 영상에서 추출된 특징이 관심 영역과 배경 영역에 따라

구별성이 높여지고 최종 뼈 나이 진단에 긍정적인 영향을 줄 수 있음을 관찰하기 위함이다.

3.3 학습 네트워크 구성

앞서 언급한 대비가 강화된 X선 영상과 군집화 기법이 적용된 분할 영상을 사용해서 더 정확한 뼈 나이 추정을 위해 학습 네트워크를 구성하고자 한다. 제안한 학습 네트워크는 그림 4와 같이 심층 합성곱 신경망으로 구성된 두 종류의 스트림을 갖는다. 그림에서 첫 번째 스트림은 대비가 강화된 X선 영상에서 유용한 특징을 뽑아내는 역할을 하고, 반면에 두 번째 스트림은 관심 영역과 배경 영역으로 구분된 군집 영상을 입력으로 받아 영역별 중요도를 판별한다. 그리고 퓨전 계층(Fusion layer)에서는 첫 번째 스트림에서 추출된 특징 맵과 두 번째 스트림에서 추출된 특징 맵을 결합함으로써, 공간 영역에 따른 특징 맵의 중요도를 부가할 수 있다. 그리고 완전연결계층과 회귀 계층을 통해서 최종 뼈 나이를 추정하게 된다.

3.4 퓨전 기법

그림 4에서 제안한 두 종류의 스트림을 결합하기 위한 퓨전 방식으로는 연결 계층을 고려할 수 있다.

그림 5는 연결 계층 기반의 퓨전 방식을 예시로 보여 주고 있다. 연결 계층이란 두 종류의 스트림에서 추출된 특징 맵을 채널 축으로 가지런히 쌓는 것을 말한다. 예를 들어, 그림 4의 각각의 스트림의 출력은 7×7 크기의 512개의 특징 맵을 가지고 있으므로 1024개의 특징 맵이 생겨나게 된다. 이를 통해 X선 영상의 특징 맵을 공간 영역별로 중요도에 대한 정보를 함께 결합할 수 있게 된다.

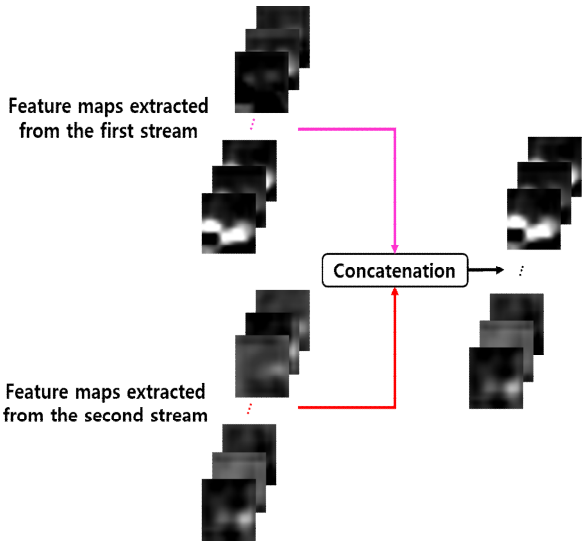

그림 5. 연결 계층을 통한 특징 퓨전
Fig. 5. Feature fusion via concatenation layer

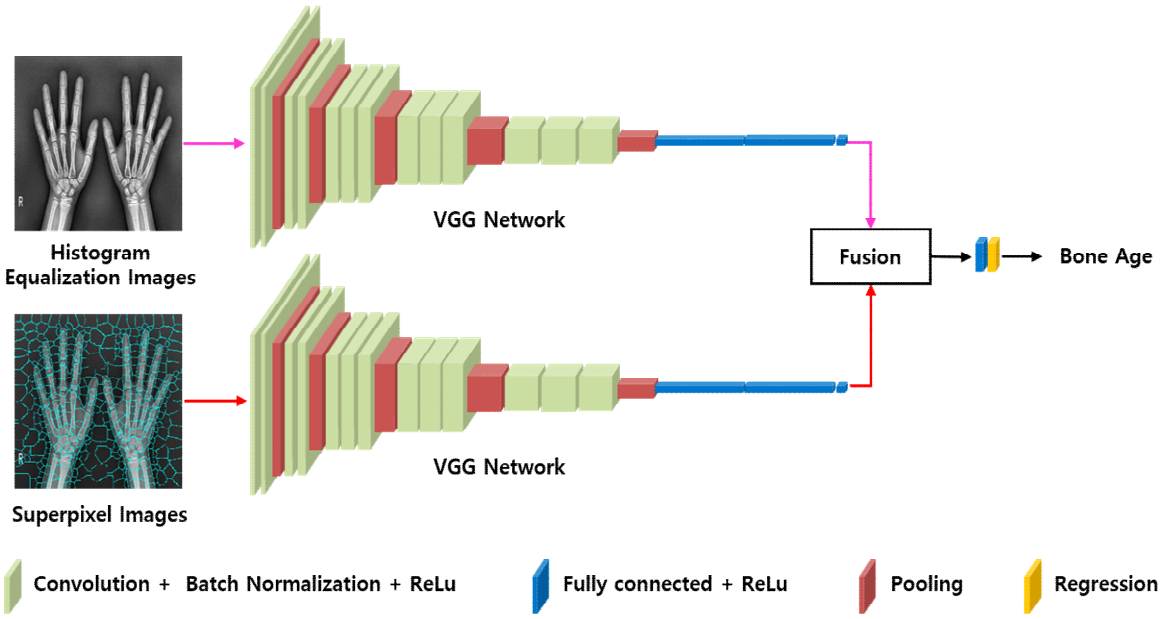

그림 4. 뼈 나이 진단을 위해 제안한 대비 강화 및 슈퍼픽셀 기반 심층 특징 퓨전
Fig. 4. Proposed deep feature fusion based on contrast enhancement and superpixels for bone age diagnosis.

이후 완전연결계층을 통해서 어떤 영역에서 특징을 추출할 때, 뼈 나이 진단에 더 효과적인지 학습을 통해 유도된다. 그리고 두 종류의 스트림을 퓨전을 할 때, 어느 계층에서 특징 맵을 추출해서 결합할지를 결정해야 된다. 일반적으로 학습된 심층 합성곱 신경망에서 특징 맵을 시각화해보면, 입력단 근처의 합성곱 계층에서는 에지와 같은 저수준의 특징을 뽑아내고 출력단 근처에 속한 완전연결계층에는 객체의 얼굴과 같은 고수준의 특징을 뽑아내는 경향이 있다[22]. 따라서 본 연구에서는 5번째로 위치한 풀링 계층과 마지막 완전연결계층에서 특징 맵을 각각 퓨전해 보고자 한다. 이를 통해 어떤 계층에서 퓨전을 하는 것이 뼈 나이 진단 성능이 더 효과적인지를 파악하고자 한다.

IV. 실험 및 결과

4.1 실험 환경

실험에 사용된 X선 영상은 영상 판독 전문업체로부터 제공 받았으며, 사용한 뼈 나이는 영상의학과 전문의에 의해 진단되었다. 영상의 개수는 총 2,328장이고 7:3의 비율로 훈련과 테스트 셋으로 랜덤하게 나누었다. 입력 X선 영상은 단일 채널을 갖는 그레이 영상이고 픽셀 당 12bit를 갖는다. 시뮬레이션 환경은 윈도우 운영체제이고 4개의 Titan-XP로 구성된 워크스테이션에서 학습되었다. 에폭의 수는 30으로 설정되었고 미니 배치 크기는 10으로 설정되었다. 심층 합성곱 신경망 툴로는 매트랩(Matlab2019b)를 사용하였고 미니 배치 확률적 경사하강법(Mini-batch stochastic gradient optimization) [23]을 사용해 파라미터를 갱신하였다. 평가 척도로는 영상의학과 전문의가 진단한 뼈 나이 값과 제안한 심층 학습을 통해 추정된 뼈 나이 값과의 오차를 계산하는 MAE 식을 사용하였다.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - y'_i| \quad (2)$$

여기서 n은 테스트 영상의 개수이다. 그리고 y_i는 X선 영상에 대해 전문의가 측정한 뼈 나이 값이고 y'_i는 제안한 네트워크로 추정된 뼈 나이 값이다.

4.2 실험 결과

표 1은 기존의 전이 학습 기반의 뼈 나이 진단 기법[16]과 제안한 기법에 대한 성능 평가 결과를 보여 준다. 표에서 볼 수 있듯이, 기존의 전이 학습 기반의 뼈 나이 진단의 평균 절대 오차는 약 9.7개월인 것을 알 수 있다. 그리고 입력 X선 영상에 히스토그램 평탄화 기법[13]을 적용한 경우, 기존의 전이 학습 기법에 비해 약 1.1개월만큼 정확도가 높아진 것을 볼 수 있다. 이는 X선 영상의 대비를 강화함으로써, 입력 영상에서 추출된 특징의 구별력을 높여 최종 뼈 나이 진단을 개선한 것으로 해석할 수 있다. 즉 대비가 낮은 영상은 특징 구별력이 떨어지므로 히스토그램 평탄화와 같은 대비 강화 기법이 뼈 나이 진단에 상당히 효과적인 것을 알 수 있다. 또한, 표에서 K-평균 군집 알고리즘이나 슈퍼픽셀 알고리즘을 적용할 경우, 뼈 나이 진단의 정확도가 더 향상됨을 알 수 있다. 약 8.6개월에서 8.1개월과 8.0개월까지 떨어지는 것을 볼 수 있다. 이 결과는 배경 영역과 관심 영역에서 추출된 특징 맵을 입력 X선 영상과 퓨전함으로써 뼈 나이 성능을 개선할 수 있음을 나타낸다.

표 1. 성능 평가
Table 1. Performance evaluation

Methods	MAE(months)
Transfer learning [16]	9.7773
Histogram equalization [13]	8.6471
Proposed deep feature fusion based on histogram equalization and K-means clustering (Last fully-connected layer)	8.1127
Proposed deep feature fusion based on histogram equalization and superpixels (Last fully-connected layer)	**8.0236**
Proposed deep feature fusion based on histogram equalization and K-means clustering (5th pooling layer)	9.1649
Proposed deep feature fusion based on histogram equalization and superpixels (5th pooling layer)	8.8497

다시 말하면 대비가 강화된 입력 X선 영상에서 추출된 특징 맵으로 영역에 따른 중요도를 학습함으로써, 특징의 구별성을 높이고 최종 뼈 나이 진단의 정확도를 높인 것으로 해석할 수밖에 없다. 그리고 두 종류의 스트림을 퓨전하는 위치는 풀링 계층보다는 마지막 완전연결계층에서 수행하는 것이 더 좋은 성능을 나타내는 것을 볼 수 있다. 일반적으로 완전연결계층에서는 객체의 얼굴과 같은 의미론적인 고수준의 특징을 뽑아내기 때문에 뼈 나이 진단에서도 동일한 현상이 나타남을 알 수 있다.

실험 결과를 요약하자면 서론에서 언급된 뼈 나이 진단에서 입력 X선 영상의 대비 강화와 배경 및 관심 영역의 분할은 최종 뼈 나이 진단에 긍정적인 영향을 줄 수 있다고 결론을 맺을 수 있다. 다시 말하면 대비 강화를 통해서는 특징의 구별성을 높일 수 있고 군집화 과정을 통해서는 특징의 영역별에 따른 중요도를 학습할 수 있음을 확인할 수 있었다.

Ⅴ. 결론 및 향후 과제

본 연구에서는 뼈 나이 진단을 위한 대비 강화 및 슈퍼픽셀 기반 심층 특징 퓨전 기법을 제안하였다. 핵심 아이디어는 X선 영상의 대비 강화와 배경과 관심 영역 분할이 최종 뼈 나이 진단의 성능을 개선할 수 있다는 것이다. 이를 구현하기 위해, 두 종류의 심층 합성곱 신경망으로 구성된 스트림을 형성하고 최종 퓨전 계층을 통해서 뼈 나이를 추정하는 심층 학습 모델을 제시하였다. 실험 결과에서 보듯이, 제안한 기법은 기존의 전이 학습 기반의 심층 학습 모델에 비해 약 1.7개월의 평균 오차를 줄일 수 있었다.

제안한 대비 강화 및 슈퍼픽셀 기반 심층 특징 퓨전 모델은 공간 영역별 특징 맵의 중요도를 반영한다는 점에서 공간 집중 네트워크(Spatail attention network)로 볼 수 있다. 따라서 향후 계획으로 공간 영역별 가중치를 계산하여 퓨전하는 기법을 새롭게 제안하고자 한다. 그리고 슈퍼픽셀 알고리즘 대신 X선 영상의 라벨링 결과, 예를 들어 영상분할이나 바운딩 박스와 결과를 활용하는 기법도 개발하고자 한다.

감사의 글

뼈 나이 진단을 위한 병원 측 데이터를 지원해 주신 분들께 감사를 드립니다.

References

[1] L. K. Hyong, "Growth assessment and diagnosis of growth disorders in childhood", Journal of the Korean Pediatric Society, Vol. 46, No. 12, pp. 1171-1177, Dec. 2003.

[2] K. M. Yeon, "Standard bone-age of infant and children in korea", Journal of the Korean Radiological Society, Vol. 34, No. 2, pp. 269-276, Feb. 1996.

[3] W. W. Greulich and S. I. Pyle, "Radiographic atlas of skeletal development of the hand and wrist", Stanford University Press, Jun. 1959.

[4] D. G. Lowe, "Distinct image features from scale-invariant key points", International Journal of Computer Vision, Vol. 60, No. 2, pp. 91-110, Nov. 2004.

[5] Z. Guo, L. Zhang, and D. Zhang, "A completed modeling of local binary pattern operator for texture classification", in IEEE Transactions on Image Processing, Vol. 19, No. 6, pp. 1657-1663, Jun. 2010.

[6] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: spatial pyramid matching for recognizing natural scene categories", in Proc. IEEE International Conference on Computer Vision and Pattern Recognition, New York, NY, pp. 2169-2178, Jun. 2006.

[7] C. Cortes and V. Vapnik, "Support-vector networks", Machine Learning, Vol. 20, No. 3, pp. 273-297, Sept. 1995.

[8] K. Simonyan, and A. Zisserman, "Very deep convolutional networks for large-scale image recognition", arXiv: 1409.1556 [cs. CV], Sept. 2014.

[9] J. W. Kim and P. K. Rhee, "Image Recognition based on Adaptive Deep Learning", The Journal of Institute of Internet, Broadcasting and Communication, Vol. 18, No. 1, pp. 113-117, Feb. 2018.

[10] K. M. Lee and C. H. Lin. "Video Stabilization Algorithm of Shaking image using Deep Learning", The Journal of Institute of Internet, Broadcasting and Communication, Vol. 19, No. 1, pp. 145-152, Feb. 2019.

[11] S. P. Lloyd, "Least squares quantization in PCM", IEEE Transactions on Information Theory, Vol. 28, No. 2, pp. 129-137, Mar. 1982.

[12] R. Achanta et al., "SLIC superpixels compared to state-of-the-art superpixel methods", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 34, No. 11, pp. 2274-2282, Nov. 2012.

[13] K. Zuiderveld, "Contrast limited adaptive histogram equalization", Graphic Gems IV. San Diego: Academic Press Professional, pp. 474–485, Aug. 1994.

[14] Iglovikov, Vladimir I., et al. "Paediatric bone age assessment using deep convolutional neural networks", arXiv: 1712.05053v2 [cs. CV], Jun. 2018.

[15] Lee, Hyunkwang, et al. "Fully automated deep learning system for bone age assessment", Journal of digital imaging, Vol. 30, No. 4, pp. 427-441, Mar. 2017.

[16] C. Spampinato, S. Palazzo, D. Giordano, M. Aldinucci, and R. Leonard, "Deep learning for automated skeletal bone age assessment in X-ray images", Medical Image Analysis, Vol. 36, pp. 41-51, Feb. 2017.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition", in Proc. IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, pp. 770-778, June. 2016.

[18] C. T. Shen and W. L. Hwang, "Color image enhancement using retinex with robust envelope", in Proc. IEEE International Conference on Image Processing, Cairo, Egypt, pp. 3141-3144, Nov. 2009.

[19] J. Seok, J. Kasa-Vubu, M. DiPietro, and A. Girard, "Expert system for automated bone age determination", Expert Systems With Applications, Vol. 50, pp. 75-88, May 2016.

[20] D. Stern, C. Payer, V. Lepetit, and M. Urschler, "Automated age estimation from hand MRI volumes using deep learning", Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016. MICCAI 2016. Lecture Notes in Computer Science, Vol. 9901. Springer, Cham, Oct. 2016.

[21] H. J. Yu and C. H. Son, "Apple leaf disease identification through region-of-interest-aware deep convolutional neural network", arXiv:1903.10356 [cs.CV], Mar. 2019.

[22] M. D. Zeiler, D. Kirshnan, G. W. Taylor, and R. Fergus, "Deconvolutional networks", in Proc. IEEE International Conference on Computer Vision and Pattern Recognition, San Francisco, CA, pp. 2528-2535, Jun. 2010.

[23] S. Ruder, "An overview of gradient descent optimization algorithms", arXiv:1609.04747[cs.LG], Sept. 2016.

저자소개

김 민 정 (Min-Jeong Kim)

2017년 3월 ~ 현재 : 군산대학교 소프트웨어융합공학과 학사과정
관심분야 : 컴퓨터 비전, 영상처리, 딥 러닝, 프로그래밍

손 창 환 (Chang-Hwan Son)

2002년 2월 : 경북대학교 전자전기공학부(공학사)
2004년 2월 : 경북대학교 전자공학과(공학석사)
2008년 8월 : 경북대학교 전자공학과(공학박사)
2017년 4월 ~ 현재 : 군산대학교 소프트웨어융합공학과 조교수
관심분야 : 컴퓨터 비전, 영상처리, 기계학습, 딥 러닝

추 현 승 (Hyunseung Choo)

1996년 12월 : 미국텍사스주립대 컴퓨터공학과(공학박사)
1998년 3월 ~ 현재 : 성균관대학교 소프트웨어학과 교수
2019년 1월 ~ 현재 : 성균관대학교 산학협력단장
2015년 9월 ~ 현재 : 과기정통부 G-ICT센터산업 지능형ICT융합연구센터 센터장
관심분야 : Network Softwarization (SDN/NFV), IoT/WoT/M2M/D2D/D2S、Mobile Ad-hoc and Sensor Networks、Mobile Human Computer Interaction