

딥러닝을 이용한 one-stage 염색체 계수 프로그램 개발

강승영¹, 한중훈¹, 류지승¹, 이세나¹, 이인경², 주해미², 양세정^{1*}

연세대학교 의공학과¹

한국수력원자력(주) 방사선보건원²

One-stage chromosome counting system development based on Deep learning

Seungyoung Kang¹, Junghun Han¹, Jiseung Ryu¹, Sena Lee¹, Inkyoung Lee², Haemi Joo², Sejung Yang^{1*}

Department of Biomedical Engineering, Yonsei University, Korea¹

Radiation Health Institute, Korea Hydro & Nuclear Power²

tigerkang96@gmail.com, *syang@yonsei.ac.kr

Abstract

As the use of radiation increases, interest in the damage caused by radiation is also increasing. When exposed to radiation, changes occur in the chromosome unit, such as abnormal chromosomes or a change in the number of chromosomes. Therefore, karyotype analysis, which analyzes the number of chromosomes or the shape of chromosomes, is important. Since karyotype analysis performed by experts is time-consuming and expensive, in this study, we developed a system for automatically counting chromosomes using deep learning. This system accurately discriminates overlapping chromosomes and showed more than 0.995 in precision and recall, and 0.909 or more in mAP 0.5 and mAP 0.5:0.95.

1. 연구 배경

최근 인류는 방사선을 사용해서 의학적 진단 또는 치료 목적으로 사용하거나 에너지 발전을 위해서 사용하고 있다. 이에 따라 방사선에 의한 피해도 지속해서 관심을 받고 있는데, 방사선이 인체에 미치는 영향은 다음과 같다. 다량의 방사선을 흡수했을 경우 피부 이상, 백혈구 일시 감소, 또는 사망에 이르기까지 한다[1]. 또한 내부적으로 인체를 구성하는 원자가 방사선이 통과되면서 에너지를 흡수하여 세포핵의 DNA 분자가 파괴되어 세포핵 자체에 손상을 받을 수 있다. 방사선에 피폭될 경우 염색체에는 동원체가 두 가지 존재하는 Dicentric, 염색체가 끊어져 발생하는 Fragment 등 여러 형태의 이상 염색체가 발생할 수 있으며 염색체의 수가 감소하거나 증가할 수 있고, 현미경을 통해 염색체를 촬영하였을 때 많은 잡음이 생성될 수 있다. 핵형 분석은 염색체의 형태, 수, 배열상태를 조사하여 방사선 피폭으로 인한 염색체의 변화를 파악할 수 있다. 하지만 전문가가 염색체를 분석할 때는 수동적으로 염색체의 개수를 세기 때문에 시간 소모가 크고 사람에 따라 결과가 달라지기도 한다. 이전에도 딥러닝을 활용하여 염색체를 분석하려는 시도는 지속해서 연구되어 왔다. Song et al.[2]은 Convolutional Neural Network (CNN) 기반의 모델로 영상 분할 기법을 통해 겹쳐진 염색체를 구분하려 하였지만 2개 이상의 염색체가 겹쳤을 경우는 고려되지 않았다. 염색체를 구분하는 데에 있어 겹쳐진 염색체나 주변의 핵과 같은 염색체 이외의 물체는 염색체의 특징을 추출하는 데 어려움이 있었다. 이에 Bai et al.[3]은 CNN을 기반으로 핵과 배경 잡음을 제거함으로써 정확하게 염색체의 특징을 추출하여 정확히 구분하는 연구를 진행했다. 하지만, 위 연구는 여러 개의 CNN 모델을 사용하여 실제로 사용하기에 프로그램의 연산량이 많다는 한계점이 있었다. 따라서 이번 연구에서는 잡음이 존재하는 염색체 영상에서도 잡음제거 없이 자동으로 염색체 계수 핵형 분석을 진행하는 딥러닝 알고리즘을 제시한다.

2. 연구 방법

실험에 사용된 데이터는 방사선보건원에서 제공한

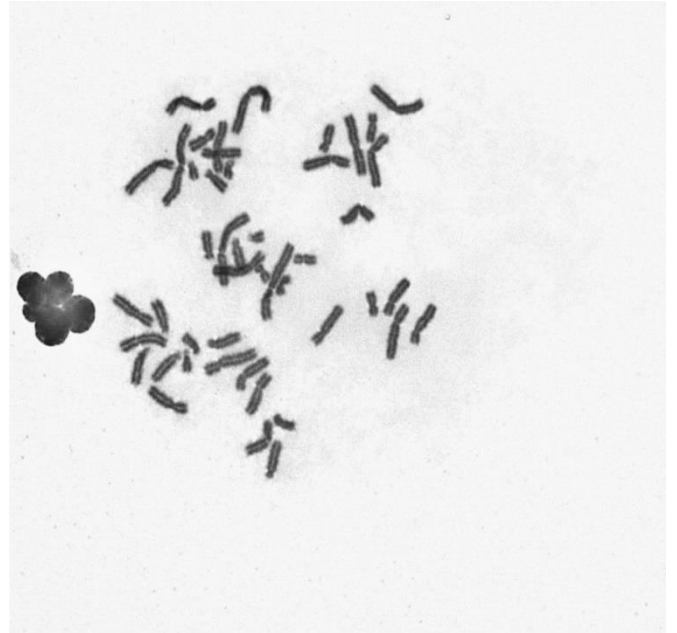


그림 1. 현미경으로 촬영된 염색체 영상 예시

일반적인 사람의 염색체 영상이다. Axio Imager Z2 현미경과 Metafer 4 소프트웨어로 촬영되었으며 그림 1은 염색체 분열 중 중기에 해당하는 영상으로 염색체가 중앙에 배열되어 있는 형태이다. 같은 사람의 세포를 촬영해도 촬영 환경, 세포 상태에 따라 그림 1에서 보이는 바와 같이 염색체가 겹쳐 있거나 잡음이 존재할 수 있다. 염색체 영상의 크기는 1024*1024의 회색조 영상이다. 학습 데이터로 총 2523장을 사용하였으며 검증데이터 100장 테스트 영상 100장으로 실험을 구성했다. 염색체 데이터는 염색체 전문가 2명에게 자문받아 라벨링을 진행하였고, 각 염색체의 위치정보를 바운딩 박스로 표시하였다.

영상 하나당 45~47개의 염색체에 대한 위치정보가 존재하며 겹쳐진 염색체의 경우 일반적인 염색체 보다 그 수가 적기 때문에 데이터 증강 기법을 사용하여 겹쳐진 염색체의 개수를 늘린 후에 학습 데이터로 사용하였다.

염색체 계수를 위해서 객체 검출 분야에서 좋은 성능을 보이는 YOLOv5를 채용했다. YOLOv5는 Cross Stage Patial Network (CSP) 라는 모듈을 사용하여 모델이 추출한 피쳐 맵들을 재사용함으로써 계산 비용을 줄이고 다양한 형태의 특징을 추출할 수 있는 장점이 있다.

3. 연구 결과

본 연구는 사람의 염색체 영상에서 염색체를 정확히 탐지하여 염색체의 개수를 정확히 세는 것을 목표로 실험을 진행했다. 바운딩 박스의 정확도를 추정하기 위한 iou loss와 박스 안에 물체가 있는지 없는지를 판단하기 위한 Binary

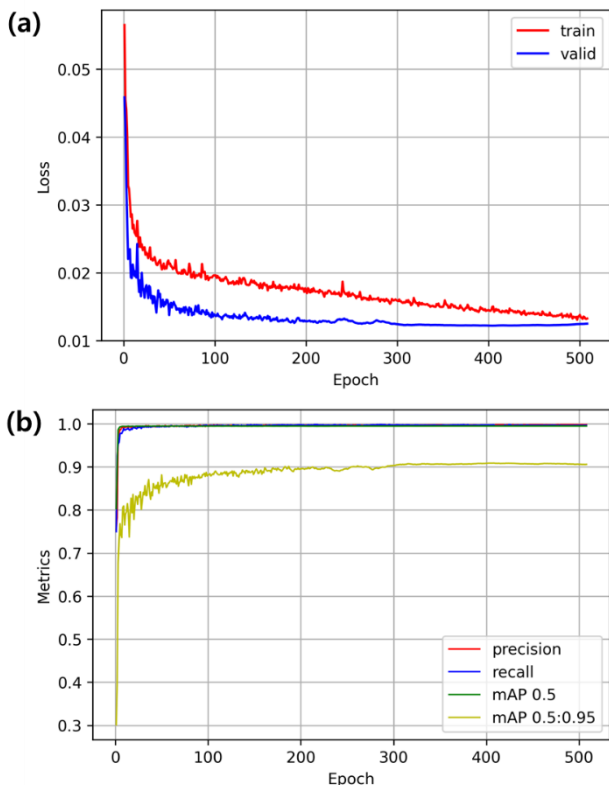


그림 3. 염색체 계수 학습 오차 및 학습 곡선

Cross Entropy loss를 사용하여 모델을 최적화하였다. 그림 2 (a) 에서 보이는 바와 같이 모델의 오차는 학습이 진행됨에 따라 점차 감소하는 양상을 보여주었고, 그림 2 (b) 와 같이 모델의 정확도를 나타내는 precision, recall, mAP 0.5, mAP 0.5:0.95 지표에서도 점차 상승하는 양상을 보여주었다. 모델의 일반화 성능을 보여주기 위한 결과는 표 1과 같다.

	Precision	Recall	mAP 0.5	mAP 0.5:0.95
Valid	0.995	0.997	0.995	0.909
Test	0.999	0.999	0.995	0.943

표 1 염색체 계수 테스트 결과

검증 데이터와 테스트 데이터에서 Precision과 Recall 모두 0.995 이상의 결과가 나타났고, 객체 검출에서 가장 신뢰도가 높은 지표인 mAP 0.5와 mAP 0.5:0.95 에서도 0.909 이상의 결과를 보여줬다.

그림 3은 염색체 계수 및 검출 결과를 보여준다. 그림 3

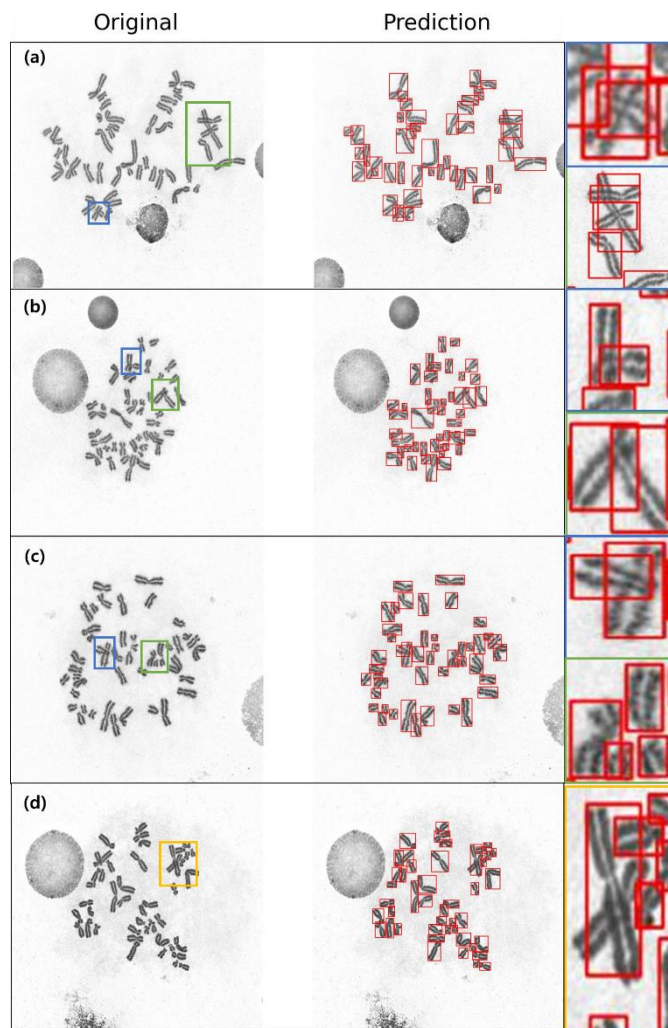


그림 2. 염색체 계수 및 검출 결과

(a), (b) 에서 볼 수 있듯이 2개 이상의 염색체가 겹쳐 있음에도 정확하게 각 염색체를 정확하게 탐지하는 것을 확인할 수 있고, 그림 3 (c) 에서는 큰 염색체 옆에 인접한 작은 염색체들도 염색체로 정확하게 탐지하는 결과를 보여주었다.

실험을 진행하면서 그림 3 (d) 와 같이 비슷한 크기의 염색체가 X 모양으로 겹쳐져 있을 경우에는 바운딩 박스끼리 비교하는 과정인 non-maximum suppression (NMS)에서 두개의 박스 중 하나가 삭제되는 경우가 있었다. 추후에 진행할 연구에서 겹쳐진 염색체에 대해 비슷한 크기의 박스를 무조건 삭제하지 않는 NMS 알고리즘을 적용할 수 있다면 염색체 계수 성능이 향상될 것으로 예상된다.

4. Acknowledgements

이 연구는 한국수력원자력(주) 방사선보건원의 지원을 받아 수행하였음. (A211F03)

5.참고 문헌

[1] Patterson, D. (2009). Molecular genetic analysis of Down syndrome. *Human genetics*, 126(1), 195-214.
 [2] Song, S., Bai, T., Zhao, Y. et al. A New Convolutional Neural Network Architecture for Automatic Segmentation of Overlapping Human Chromosomes. *Neural Process Lett* 54, 285–301 (2022). <https://doi.org/10.1007/s11063-021-10629-0>

[3] H. Bai, T. Zhang, C. Lu, W. Chen, F. Xu and Z. -B. Han, "Chromosome Extraction Based on U-Net and YOLOv3," in *IEEE Access*, vol. 8, pp. 178563-178569, 2020, doi: 10.1109/ACCESS.2020.3026483.

[4] G. Jocher, A. Chaurasia, A. Stoken, J. Borovec, NanoCode012, Y. Kwon, TaoXie, et al. "Ultralytics/yolov5: V6.1 - Tensorrt, Tensorflow Edge TPU and Openvino Export and Inference". Zenodo, February 22, 2022. <https://doi.org/10.5281/zenodo.6222936>.