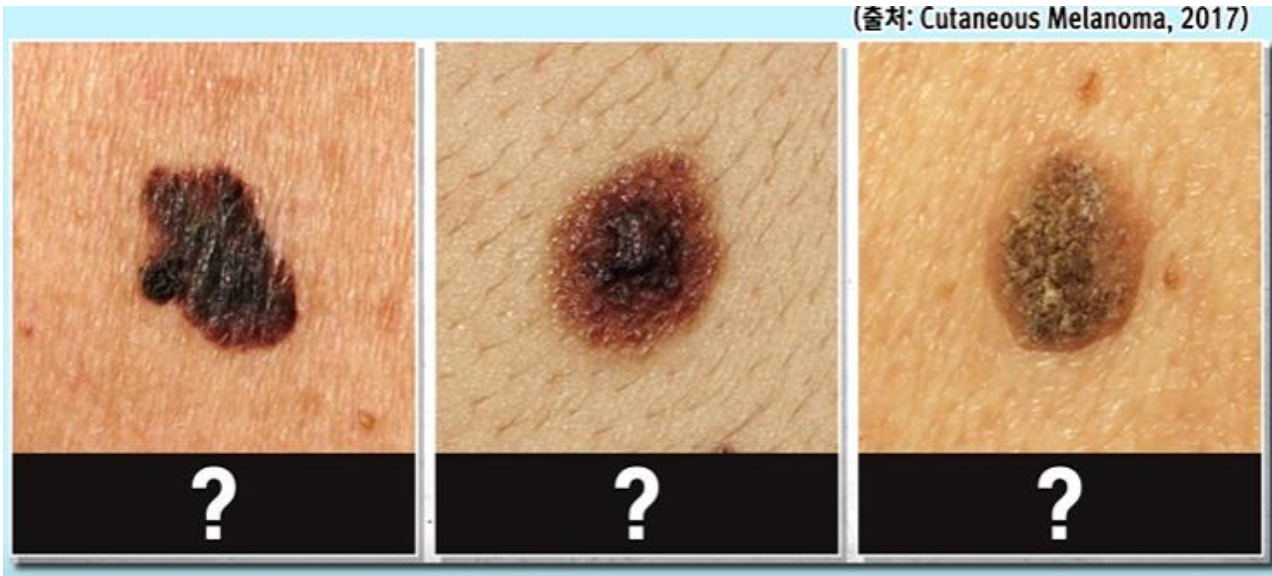


피부암 진단 DNN의 분할 컴퓨팅 구현

[졸업 프로젝트 발표]

저자: 염호진, 김경운 지도교수: 고한열
경희대학교 전자정보대학 전자공학과

INTRODUCTION



• 피부암 진단 DNN의 필요성

- 일반인이 피부암을 정상적인 점과 구분하는 것은 매우 어려움
- 조기진단이 어려워 병을 키우는 일이 빈번함

⇒환자가 **Mobile device**로 사진을 찍어 진단하는
모바일 AI 애플리케이션이 필요함

• 모바일 피부암 진단 DNN이 가지는 문제점

- Mobile Device 단독으로 복잡한 DNN을 수행하기 어려움 → **computing power 문제**
- Cloud Computing 은 환자의 민감한 의료 데이터를 Cloud 서버로 전송해야 함 → **보안 문제**

⇒ **Computing power**가 충분한 **Sever**에게
원본 데이터가 아닌 **intermediate data**를
전송하는 **split computing** 방식을 사용하여
문제 해결

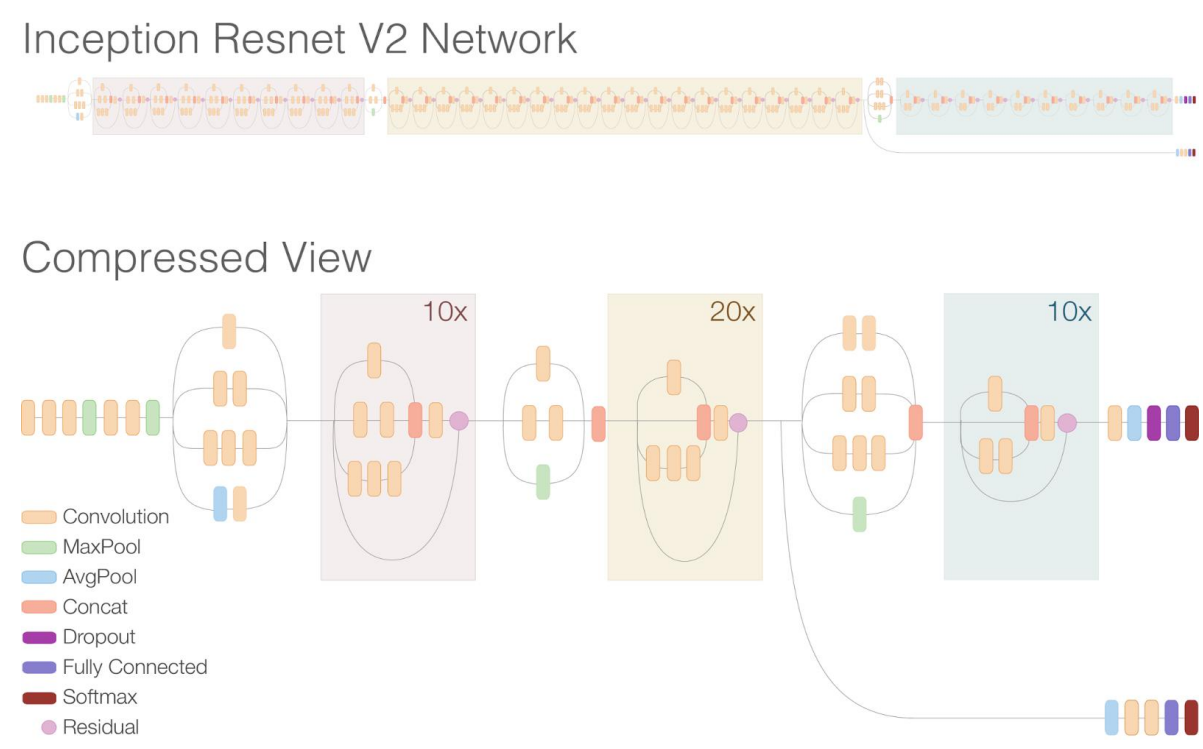
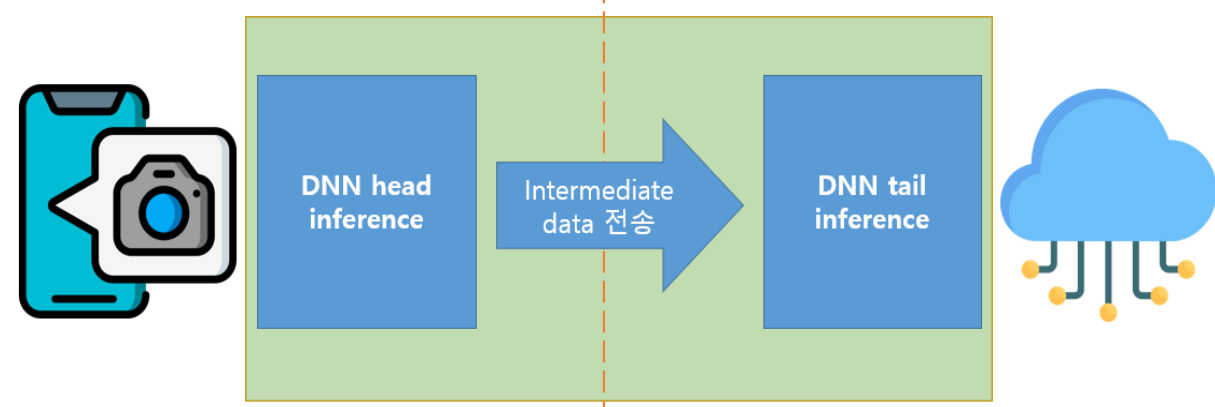
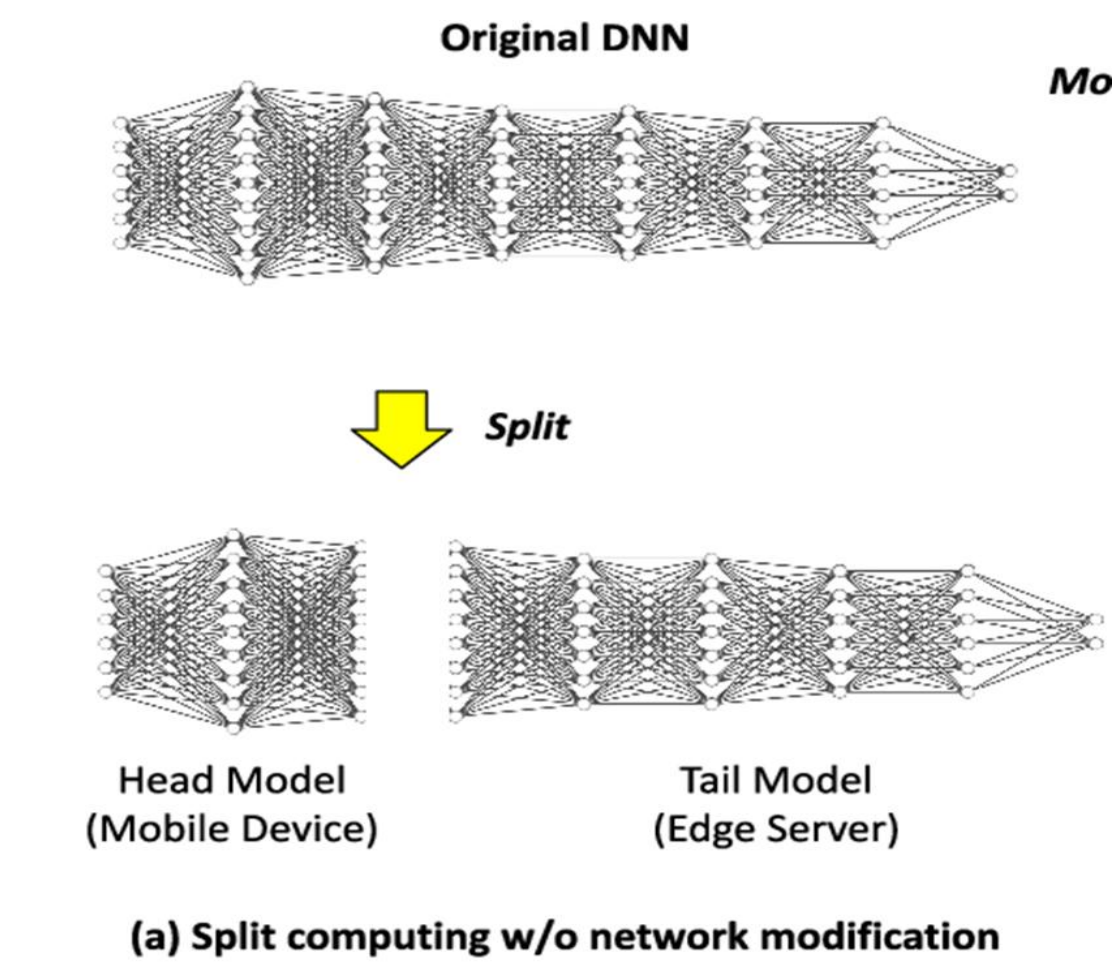
연구 배경 이론

Split Computing이란?

Mobile Device와 Server가 DNN 추론을 분할하여 수행하는 방식이다.

Mobile device 는 복잡한 Deep Neural Network를 단독으로 수행하기에 computing power 가 충분하지 않다. 이에 대한 대안으로 Cloud computing 방식이 제안되었다. Cloud computing 방식에서 Mobile device 는 원본 input 데이터를 서버로 전송하고 서버에서 DNN 추론 작업을 수행한다. Cloud computing 방식은 Mobile device 의 computing power 부족 문제를 효과적으로 극복하지만, 원본 input 데이터를 전송한다는 점에서 traffic overhead 문제와 보안 문제를 가지고 있다. Cloud computing 의 이러한 한계를 극복하기 위해서 Split computing이 제안되었다. Split computing 에서 Mobile device 는 분할된 DNN의 앞부분(head)을 수행하고 중간 결과인 intermediate data를 생성한다. Intermediate data는 서버로 전송되어 서버가 DNN의 뒷부분(tail)을 수행하기 위한 input의 역할을 한다. 이후 DNN의 뒷부분(tail)수행을 마친 서버는 최종 결과를 얻는다. Split computing은 서버로 원본 input 데이터가 아닌 Intermediate data를 보내기 때문에 원본 input 데이터의 유출 우려가 적다는 보안 측면의 장점이 있다.

InceptionResnetV2는 Inception Network와 ResNet 아키텍처를 결합한 이미지 분류 딥러닝 모델이다. 본 연구에서는 InceptionResNetV2를 이용하여 피부암을 진단하는 DNN 모델 및 split computing을 구현한다.



문제 해결

모델 학습 과정에서 문제 발생

```
Epoch 7/100
825/825 [.....] - 7s 8m/step - loss: 0.8858 - accuracy: 0.9824 - val_loss: 0.8923 - val_accuracy: 0.9822
Epoch 8/100
825/825 [.....] - 7s 8m/step - loss: 0.8858 - accuracy: 0.9824 - val_loss: 0.8884 - val_accuracy: 0.9820
Epoch 9/100
825/825 [.....] - 7s 8m/step - loss: 0.8833 - accuracy: 0.9825 - val_loss: 0.8932 - val_accuracy: 0.9822
Epoch 10/100
825/825 [.....] - 7s 8m/step - loss: 0.8833 - accuracy: 0.9825 - val_loss: 0.8883 - val_accuracy: 0.9822
```

특별한 데이터 전처리 없이도 ISIC의 이미지 데이터셋으로 학습시킨 모델이 98%의 검증 정확도를 보였음.

그러나 이는 양성증양의 데이터는 3만개에 달하는 반면, 양성증양의 데이터는 500개에 불과하기 때문에 발생한 데이터셋의 불균형으로 인한 문제였음.

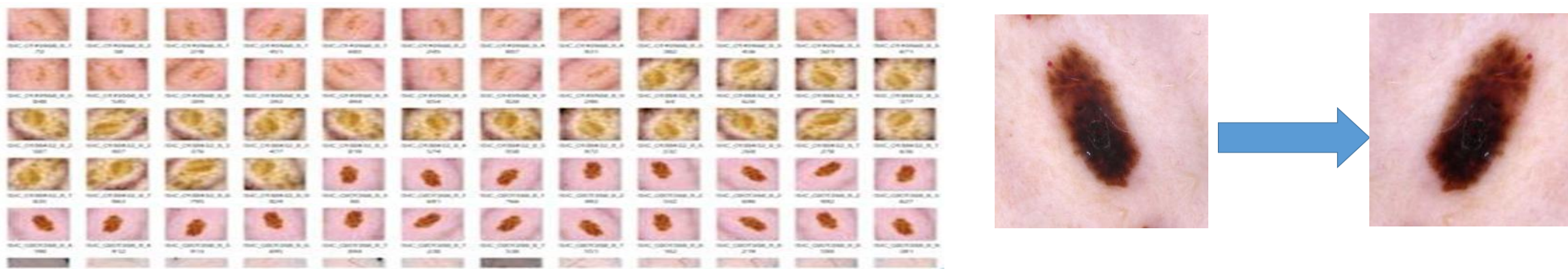
⇒ 양성증양 데이터만을 테스트 했는데 모두 양성이라고 판단하는 잘못된 추론결과를 보인다.

즉, 데이터셋의 불균형 문제를 해결하기 위한 방법이 필요함.

⇒ 학습 데이터 추가 확보 및 데이터 증강

데이터 불균형 해결 방법

- ISIC의 데이터셋만으로는 부족 => (1) 추가적인 데이터 수집
- 양성증양 데이터를 총 2521장 확보했지만, 여전히 부족한 양성증양 이미지 => (2) **증강 기법으로 부족한 이미지 보충**



random 비율로 이미지를 회전, 확대/축소 및 뒤집기로 부족한 양성증양 데이터를 증강함

• 모델의 Overfitting

데이터의 균형을 맞췄으나 InceptionResNetV2 모델의 복잡도가 커서 학습과 검증 데이터를 과도하게 학습했다. 따라서 다른 데이터를 예측하지 못한다.

⇒ **모델을 경량화할 필요가 있다!**

1. Dropout : 모델의 뉴런을 random 하게 drop하여 모델의 복잡도를 줄이는 기법
2. Regularizers : 손실함수는 예측값과 실제값의 차이를 반영한다. 이때, 가중치의 제곱을 추가해 모델이 가중치의 크기를 작게 유지하도록 유도하는 기법

실험 결과

InceptionResNetV2를 이용해 학습한 결과 아래 이미지 의 결과를 가지는 모델을 얻었다. 검증 정확도는 94.75%로 학습이 잘 된 것을 확인할 수 있다.

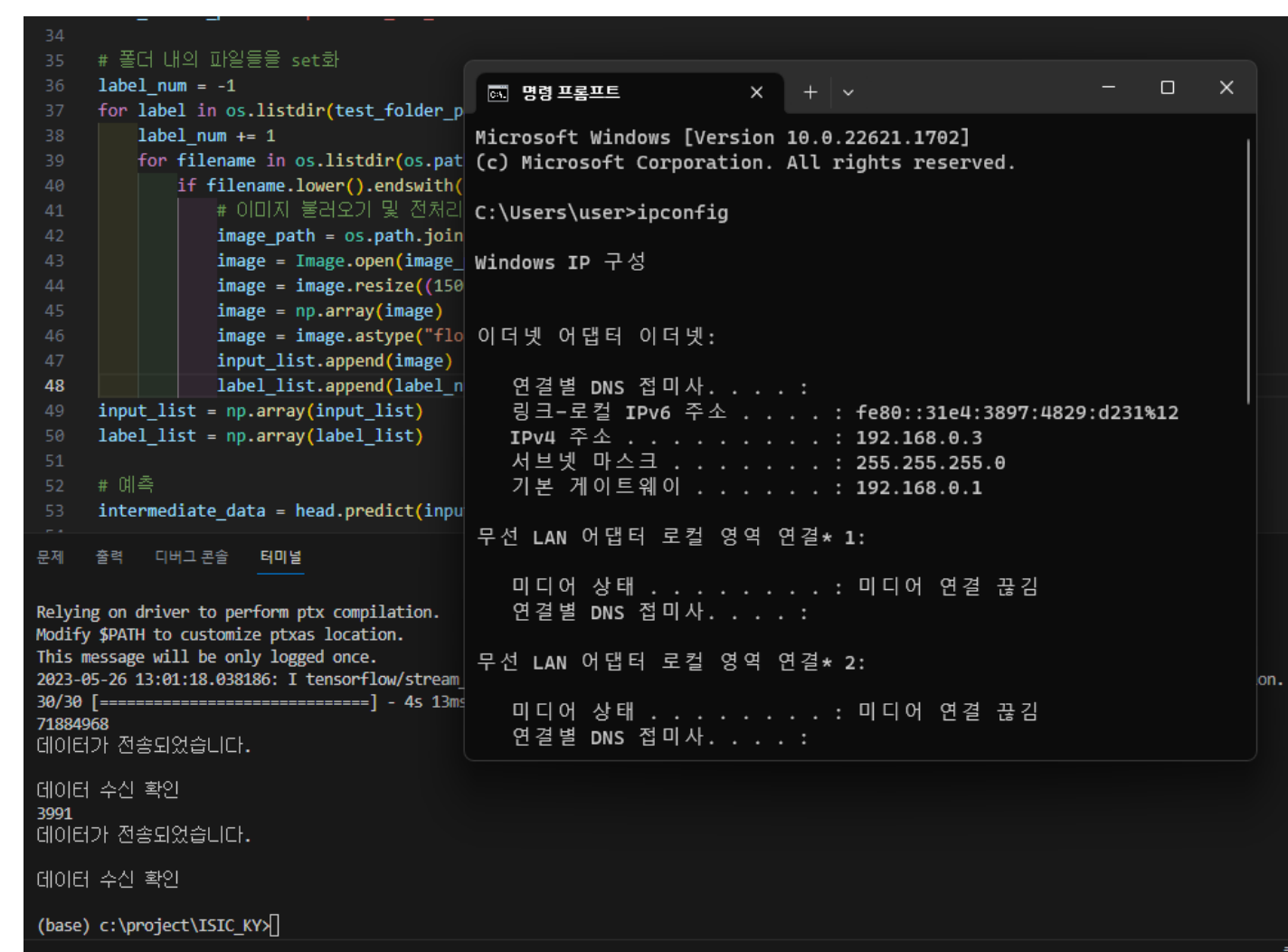
```
814s 1s/step - loss: 0.2504 - accuracy: 0.9197 - val_loss: 0.1722 - val_accuracy: 0.9475
```

```
(base) c:\project\ISIC_V2> cd c:\project\ISIC_V2 && cd /c "C:\Users\user\er\..\debuggy\launcher 13125 -- c:\project\ISIC_V2\inference.py"
2023-05-26 13:21:04.539559: I tensorflow/core/platform/cpu_feature_guard.cc:193]
actions in performance-critical operations.
2023-05-26 13:21:04.861743: I tensorflow/core/common_runtime/gpu/gpu_device.cc:166]
dforce RTX 4090, pci bus id: 0000:01:00.0, compute capability: 8.9
2023-05-26 13:22:04.282321: I tensorflow/stream_executor/cuda/cuda_dnn.cc:384]
2023-05-26 13:22:05.195831: M tensorflow/stream_executor/cuda/cuda_allocator.cc:
Relying on driver to perform gtx compilation.
Modify $PATH to customize ptas location.
This message will be only logged once.
2023-05-26 13:22:05.048821: I tensorflow/stream_executor/cuda/cuda_blas.cc:1614]
30/30 [.....] - 4s 13m/step
30/30 [.....] - 8s 3m/step
<class 'numpy.ndarray'>
711 249
Accuracy = 0.753125
```

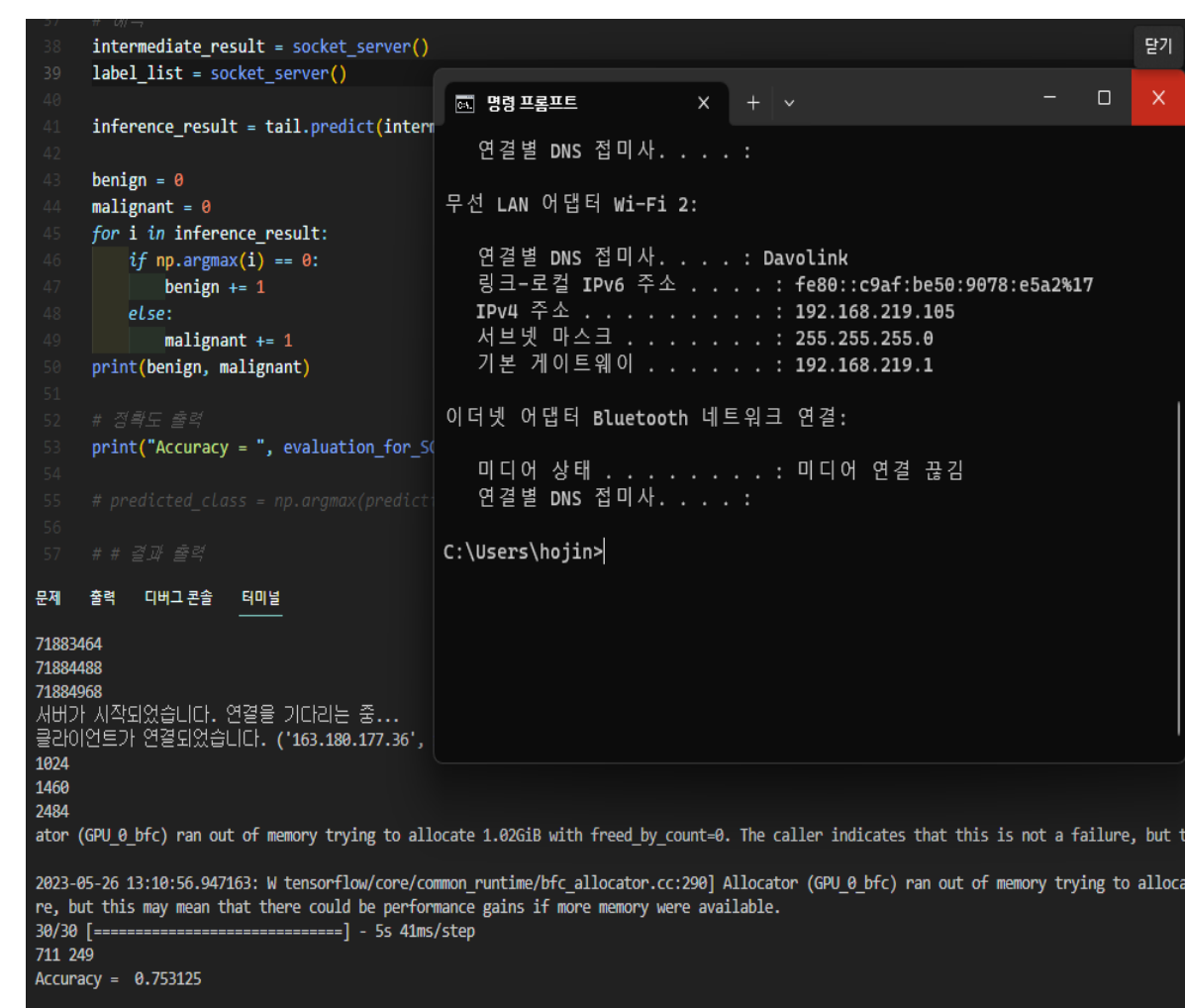
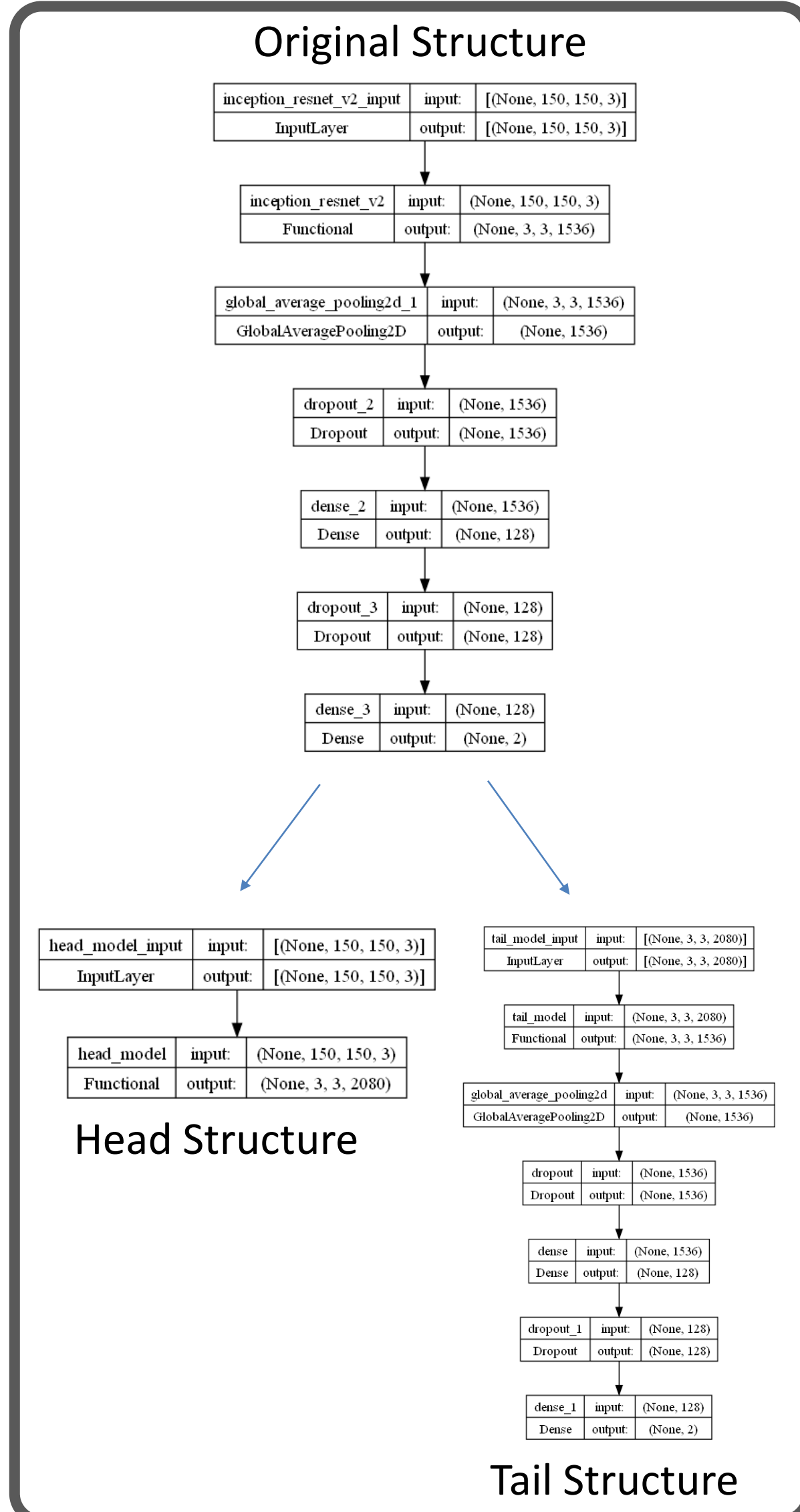
위의 이미지는 학습을 마친 모델로 테스트 데이터를 추론하여 얻은 결과이다. 검증 정확도가 94.75%이었던 것에 비해 낮은 75%가 나온 것을 볼 수 있다. 이는 증강 및 추가된 데이터셋으로 양성증양 데이터의 불균형 문제를 해결하려고 했으나 여전히 절대적인 데이터 수가 부족해 피부암 이미지의 특징을 완벽히 추출하지 못했기 때문으로 예상된다.

학습을 마친 모델은 head와 tail로 나누어 각각 클라이언트와 서버에 할당했다. 오른쪽의 이미지는 원본 모델의 구조와 head와 tail의 구조를 plot한 이미지이다.

클라이언트는 head 부분 추론의 결과인 intermediate data를 TCP/IP 방식으로 client에게 전송한다. 아래는 intermediate data 전송을 완료한 클라이언트 측의 결과 화면이다.



서버는 TCP/IP를 통해 수신한 intermediate data를 input으로 하여 tail 모델 추론을 수행하고 최종 결과를 출력한다. 그 결과는 오른쪽의 이미지와 같다. 이때 서버 측 tail 추론 결과의 정확도는 원본 모델의 정확도인 75퍼센트와 동일하다. 이러한 결과를 토대로 TCP/IP를 통한 split computing 추론이 손실없이 잘 이루어졌다는 것을 알 수 있다.



결론

기존의 cloud computing 방식의 DNN 학습은 민감할 수 있는 환자의 개인정보를 외부로 유출해야 한다는 단점이 있었다. 본 연구에서는 이러한 한계를 해결하고자 원본 데이터를 외부로 직접적으로 유출하지 않는 split computing 방식으로 피부암 진단 DNN을 구현하였다. 만약 intermediate data가 유출되더라도 식별이 어려운 다차원 배열 형식이기 때문에 보안적 측면에서 우수하다는 장점을 가진다. 이러한 점에서 split computing이 가지는 의의를 확인할 수 있었다.

하지만 본 연구는 split computing 자체는 성공적으로 구현했지만 피부암 진단의 DNN의 정확도를 만족할 만큼 이끌어내지 못했다는 한계를 가진다. 훈련과 검증 정확도는 높은 반면, 테스트에서는 이에 미치지 못하는 정확도를 얻었다. 가장 큰 원인은 overfitting 문제를 완벽히 해결하지 못했기 때문이다. Overfitting의 근본적인 문제는 우선 데이터의 부족이다. 증강을 통해 부족한 데이터의 양을 늘려 문제를 해결하고자 했지만 어디까지나 인위적으로 증강한 이미지가기 때문에 DNN이 피부암의 특징을 제대로 학습하기에는 부족했다. 즉, 학습 데이터가 부족한 상황에서 이를 과도하게 학습하는 overfitting 때문에 특징을 제대로 추출하지 못한 것이다. 향후 연구에서 이 문제를 해결하기 위해서는 overfitting이 일어나는 것을 방지할 수 있는 증강 기법이나 환자의 나이, 성별, 기타 질병 등의 메타 데이터를 추가하여 학습하는 방법을 제안할 수 있을 것이다.

참고 문헌

임상현, 이명숙(2018), 딥 러닝 기반의 양성흑색종 분류를 위한 컴퓨터 보조진단 알고리즘, 디지털산업정보학회 논문지 제14권 제4호

권 강 외 5명(2015), 국내 한의학계에 보고된 흑색종 관련 실험적 연구들의 비교 고찰, 한방안이비인후피부과학회지 제28권 제3호

Kaiming He Xiangyu Zhang Shaoqing Ren Jian Sun(2015), Deep Residual Learning for Image Recognition, Microsoft Research

염호진, 김현수, 고한열, "무선 모바일 네트워크에서의 트래픽 효율적인 분할 컴퓨팅 메커니즘," JCCI 2023, 2023년 4월.



경희대학교
KYUNG HEE UNIVERSITY