

강아지 품종 분류 모델 개발 및 성능 분석

정보공학과 2020069852 임재서



목차

- 1 서론
- 2 이론적 배경
- 3 데이터셋 소개
- 4 데이터 전처리
- 5 모델 및 학습
- 6 결과 및 분석
- 7 결론
- 8 어려웠던 점 및 느낀 점



서론 – 연구 배경 및 필요성(1)

1. 주제 선정 동기

- 최근 반려동물 양육 인구가 크게 늘어나면서 관련 서비스와 기술에 대한 요구가 높아지고 있다.
- 특히 인공지능(AI) 및 컴퓨터 비전 기술은 동물 인식, 행동 분석 등 다양한 분야에서 혁신적인 가능성을 제시하고 있다.
- 평소 강아지에 대한 관심과 함께, 컴퓨터 비전 분야의 핵심 기술인 이미지 분류를 심도 있게 학습하고자 본 주제를 선정하게 되었다.

2. 강아지 품종 분류의 필요성

- 전 세계적으로 수백 종에 달하는 강아지 품종은 외형이 매우 다양하며, 일반인이 육안으로 정확히 구분하기에는 한계가 있다.
- 품종별 유전적 특성, 질병, 행동 패턴 연구에 기여할 수 있다.
- 따라서 정확하고 자동화된 품종 분류 시스템이 필요하다.



서론 – 연구 배경 및 필요성(2)

3. 과제의 기술적 정의 : Fine-grained image Classification

- 같은 상위 범주(예: '개')에 속하지만, 미묘한 시각적 차이로 구분되는 하위 범주(예: '푸들', '보더콜리')를 식별해야 한다.
- 종종 특정 부위의 미세한 형태, 질감, 패턴 등을 통해 구분해야 하므로 높은 수준의 정확도를 달성하기 어렵다.



-> 같은 품종이어도 형태와 패턴 등이 달라 구분이 어려움. (위 3장의 이미지는 모두 보더콜리)

서론 – 연구 목표 및 범위

1. 연구 목표

- 사전 학습된 EfficientNet-B0 모델을 기반으로 전이 학습을 적용하여, 주어진 강아지 이미지를 120가지의 서로 다른 품종으로 정확하게 분류하는 딥러닝 모델을 개발하는 것을 최종 목표로 한다.

2. 연구 내용 및 범위

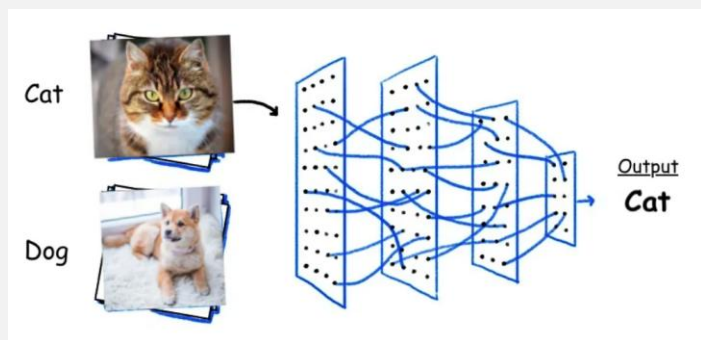
- 데이터는 공개 데이터셋인 Stanford Dogs Dataset을 활용한다. (120개 품종, 약 20,000장의 이미지 포함)
- 데이터 전처리는 XML 어노테이션 정보를 활용하여 강아지 객체 중심의 이미지 크롭(crop)을 수행한다.
- 데이터 증강(Data Augmentation) 기법을 적용하여 모델의 일반화 성능을 높인다.
- EfficientNet-B0 아키텍처를 기반으로, ImageNet으로 사전 학습된 가중치를 활용한 전이 학습(Transfer Learning)을 수행한다.
- 학습된 모델의 성능은 정확도(Accuracy)를 주요 지표로 사용하여 평가하며, 학습 과정에서의 손실(Loss) 변화 등을 함께 분석한다.



이론적 배경(1)

1. 이미지 분류 (Image Classification)

- 컴퓨터 비전의 핵심 분야로, 입력된 이미지에 대해 미리 정의된 여러 범주(레이블) 중 가장 적합한 범주를 할당하는 기술이다.
- (예시) 주어진 사진이 '치와와'인지, '골든 리트리버'인지, 혹은 다른 품종(고양이 등)인지 식별하는 작업이다.
- 본 연구에서는 '강아지 품종'을 분류 대상 범주로 설정한다.



2. CNN(합성곱 신경망)

- 이미지 데이터 처리에 특화된 딥러닝 아키텍처로, 이미지 인식, 분류, 객체 탐지 등 다양한 분야에서 뛰어난 성능을 보인다.
- CNN은 이미지 내 객체의 위치 변화나 약간의 왜곡에도 강인한 특징을 학습할 수 있다.



이론적 배경(2)

3. 전이학습

- 대규모 데이터셋으로 이미 학습된 모델의 지식(가중치)을 가져와, 새로운 작업에 적용하는 머신러닝 기법이다.
- 상대적으로 적은 양의 목표 데이터셋으로도 높은 성능을 빠르게 달성할 수 있다.
- 본 연구에서는 ImageNet으로 사전 학습된 EfficientNet-B0 모델을 강아지 품종 분류 문제에 맞게 미세 조정(fine-tuning)하는 방식으로 전이 학습을 활용한다.

4. EfficientNet-B0 모델

- 구글(Google AI)에서 개발한 CNN 아키텍처 시리즈로, 효율성과 성능을 모두 만족하여 이 모델을 선택하였다.
- EfficientNet은 비교적 적은 파라미터 수와 연산량으로도 높은 성능을 내도록 설계되어 제한된 컴퓨팅 성능에서 효과적이다.
- 네트워크의 깊이, 너비, 그리고 입력 이미지의 해상도를 가장 균형 잡힌 비율로 함께 확장하여 최적의 성능을 찾는 방식이다.
- 미세한 차이를 구분해야 하는 Fine-grained Classification 과제에서도 강력한 능력을 발휘할 것으로 기대된다.



데이터셋 소개

1. Stanford Dogs Dataset

- 본 연구는 강아지 품종 분류를 위해 널리 사용되는 공개 데이터셋인 Stanford Dogs Dataset을 활용하였다.
- 출처: 스탠포드 대학교 (Stanford University)
- 모든 이미지에 대해 품종 레이블과 함께 객체의 위치를 나타내는 바운딩 박스(Bounding Box) 정보가 XML 형태로 제공된다.



```
<annotation>
  <folder>02106166</folder>
  <filename>n02106166_739</filename>
  <source>
    <database>ImageNet database</database>
  </source>
  <size>
    <width>375</width>
    <height>500</height>
    <depth>3</depth>
  </size>
  <segment>0</segment>
  <object>
    <name>Border_collie</name>
    <pose>Unspecified</pose>
    <truncated>0</truncated>
    <difficult>0</difficult>
    <bndbox>
      <xmin>0</xmin>
      <ymin>113</ymin>
      <xmax>343</xmax>
      <ymax>488</ymax>
    </bndbox>
  </object>
</annotation>
```



데이터셋 소개

2. 데이터 구조화 : Pandas DataFrame으로 구성하여 활용

학습 데이터

```
학습 데이터 로딩 중...
df_train 생성 완료: 12000개 샘플
--- df_train 데이터 샘플 (상위 5개) ---

      image_path \
0  C:\Users\JT\Desktop\images\Images\n02085620-Ch...
1  C:\Users\JT\Desktop\images\Images\n02085620-Ch...
2  C:\Users\JT\Desktop\images\Images\n02085620-Ch...
3  C:\Users\JT\Desktop\images\Images\n02085620-Ch...
4  C:\Users\JT\Desktop\images\Images\n02085620-Ch...

      annotation_path \
0  C:\Users\JT\Desktop\annotation\Annotation\n020...
1  C:\Users\JT\Desktop\annotation\Annotation\n020...
2  C:\Users\JT\Desktop\annotation\Annotation\n020...
3  C:\Users\JT\Desktop\annotation\Annotation\n020...
4  C:\Users\JT\Desktop\annotation\Annotation\n020...

      label      image_relative \
0      1  n02085620-Chihuahua/n02085620_5927.jpg
1      1  n02085620-Chihuahua/n02085620_4441.jpg
2      1  n02085620-Chihuahua/n02085620_1502.jpg
3      1  n02085620-Chihuahua/n02085620_1916.jpg
4      1  n02085620-Chihuahua/n02085620_13151.jpg

      annotation_relative
0  n02085620-Chihuahua/n02085620_5927
1  n02085620-Chihuahua/n02085620_4441
2  n02085620-Chihuahua/n02085620_1502
3  n02085620-Chihuahua/n02085620_1916
4  n02085620-Chihuahua/n02085620_13151
```

테스트 데이터

```
테스트 데이터 로딩 중...
df_test 생성 완료: 8580개 샘플
--- df_test 데이터 샘플 (상위 5개) ---

      image_path \
0  C:\Users\JT\Desktop\images\Images\n02085620-Ch...
1  C:\Users\JT\Desktop\images\Images\n02085620-Ch...
2  C:\Users\JT\Desktop\images\Images\n02085620-Ch...
3  C:\Users\JT\Desktop\images\Images\n02085620-Ch...
4  C:\Users\JT\Desktop\images\Images\n02085620-Ch...

      annotation_path \
0  C:\Users\JT\Desktop\annotation\Annotation\n020...
1  C:\Users\JT\Desktop\annotation\Annotation\n020...
2  C:\Users\JT\Desktop\annotation\Annotation\n020...
3  C:\Users\JT\Desktop\annotation\Annotation\n020...
4  C:\Users\JT\Desktop\annotation\Annotation\n020...

      label      image_relative \
0      1  n02085620-Chihuahua/n02085620_2650.jpg
1      1  n02085620-Chihuahua/n02085620_4919.jpg
2      1  n02085620-Chihuahua/n02085620_1765.jpg
3      1  n02085620-Chihuahua/n02085620_3006.jpg
4      1  n02085620-Chihuahua/n02085620_1492.jpg

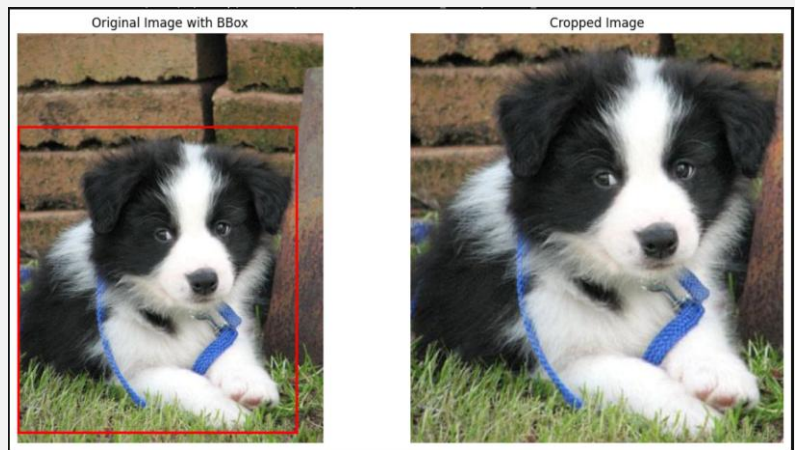
      annotation_relative
0  n02085620-Chihuahua/n02085620_2650
1  n02085620-Chihuahua/n02085620_4919
2  n02085620-Chihuahua/n02085620_1765
3  n02085620-Chihuahua/n02085620_3006
4  n02085620-Chihuahua/n02085620_1492
```



데이터 전처리(1)

1. 바운딩 박스(Bounding Box)를 활용한 객체 중심 이미지 크롭

- 제공된 XML 어노테이션 파일에서 각 이미지 내 강아지의 정확한 위치 정보(바운딩 박스 좌표)를 추출한다.
- 이 좌표를 사용하여 이미지에서 강아지가 포함된 영역만을 잘라낸다.
- 불필요한 배경을 제거하고 강아지 객체 자체에 집중할 수 있도록 한다.



2. 이미지 크기 표준화

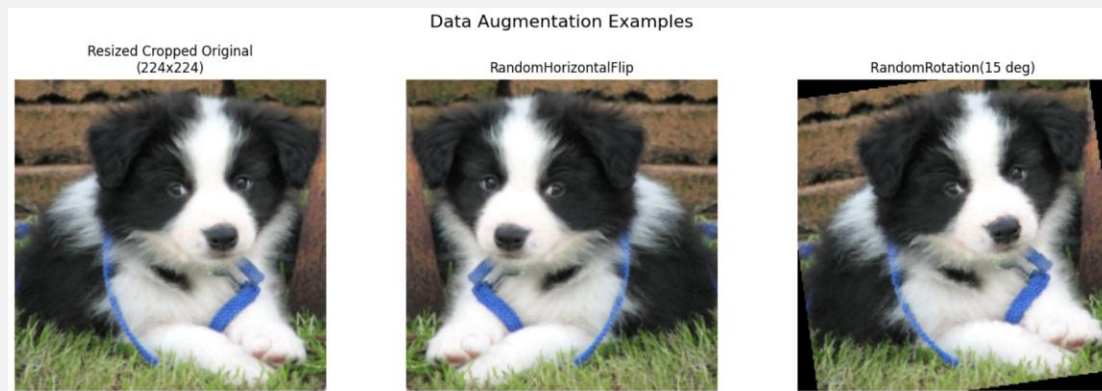
- 모든 이미지를 모델(EfficientNet-B0)의 표준 입력 크기인 224x224 픽셀로 일괄 조정한다.



데이터 전처리(2)

3. 학습 데이터 증강

- 제한된 학습 데이터셋의 다양성을 인위적으로 늘려, 모델의 일반화 성능을 향상시키고 과적합을 방지한다. (학습 데이터에만 적용)
- RandomHorizontalFlip() : 이미지를 수평 방향으로 무작위로 뒤집는다.
- RandomRotation(15) : 이미지를 최대 15도 범위 내에서 무작위로 회전시킨다.



4. 이미지 텐서 변환 및 정규화

- 이미지 객체를 PyTorch 모델이 입력으로 받을 수 있는 다차원 배열 형태인 **텐서(Tensor)**로 변환한다.

텐서로 변환된 이미지의 픽셀 값 범위를 특정 평균과 표준편차를 사용하여 조정한다.



모델 및 학습

1. 모델 아키텍처

- 기반 모델 : EfficientNet-B0 (ImageNet으로 사전 학습된 모델 사용)
- 전이 학습 : 사전 학습된 모델의 가중치를 유지하거나 미세 조정(fine-tuning)하고, 분류기(classifier) 부분을 새로운 작업에 맞게 수정한다.
- EfficientNet-B0의 기존 분류층을 제거하고, 본 연구의 목표인 120개 강아지 품종을 분류하기 위한 새로운 완전 연결 계층(Fully Connected Layer)으로 교체하였다.

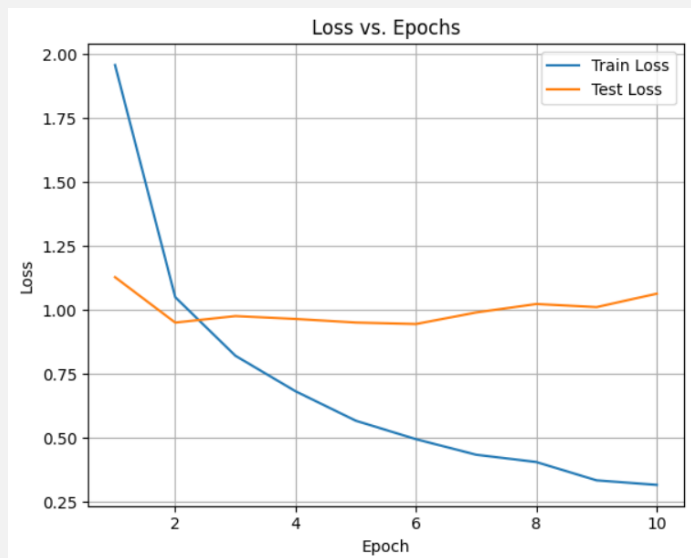
2. 주요 학습 파라미터

- 손실 함수 (Loss Function) : CrossEntropyLoss
- 옵티마이저 (Optimizer) : Adam
- 학습률 (Learning Rate) : 0.001
- 배치 크기 (Batch Size) : 32
- 총 에포크 수 (Number of Epochs) : 10



결과 및 분석

1. 학습 진행에 따른 손실(Loss) 변화

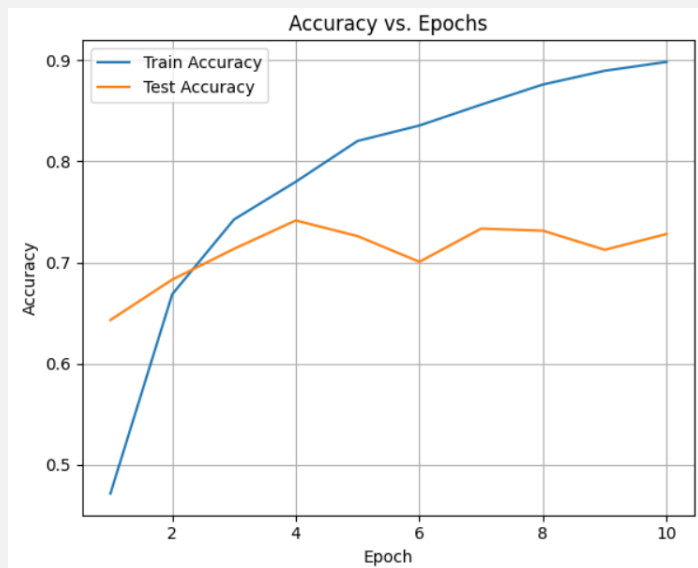


- 학습 손실 (Train Loss)은 에포크가 진행됨에 따라 꾸준히 감소하여 약 0.3 부근까지 도달했다. 이는 모델이 학습 데이터에는 효과적으로 적응하고 있음을 의미한다.
- 테스트 손실(Test Loss)은 초기에만 감소하는 경향을 보이며, 이후로는 뚜렷하게 감소하지 않는다.
- 6 에포크 이후 학습 손실은 계속 감소하는 반면 테스트 손실은 정체되거나 증가하는 양상은 과적합(Overfitting)이 발생하고 있을 가능성을 시사한다.



결과 및 분석

2. 학습 진행에 따른 정확도(Accuracy) 변화



- 학습 정확도 (Train Accuracy)는 에포크가 진행됨에 따라 지속적으로 상승하여 10 에포크 시점에는 약 0.9 (90%)에 도달했다.
- 테스트 정확도 (Test Accuracy)는 학습 초기에 빠르게 상승하여 4 에포크에서 최고치에 도달한 후, 이후로는 그 수준을 유지하거나 다소 하락하였다.
- 학습 정확도와 테스트 정확도 간의 간격이 점차 벌어지는 것 역시 과적합이 진행 중임을 나타내는 지표이다.
- 모델이 학습 데이터에는 능숙해지지만, 새로운 데이터에 대한 일반화 성능은 정체된 것으로 볼 수 있다.



결과 및 분석

3. 주요 성능 지표

- 최고 테스트 정확도 (Best Test Accuracy): 0.7416 (74.16%)
- 평균 테스트 정확도 (Average Test Accuracy over 10 Epochs): 0.7114 (71.14%)
- 평균 테스트 손실 (Average Test Loss over 10 Epochs): 1.0117

4. 모델 성능 평가

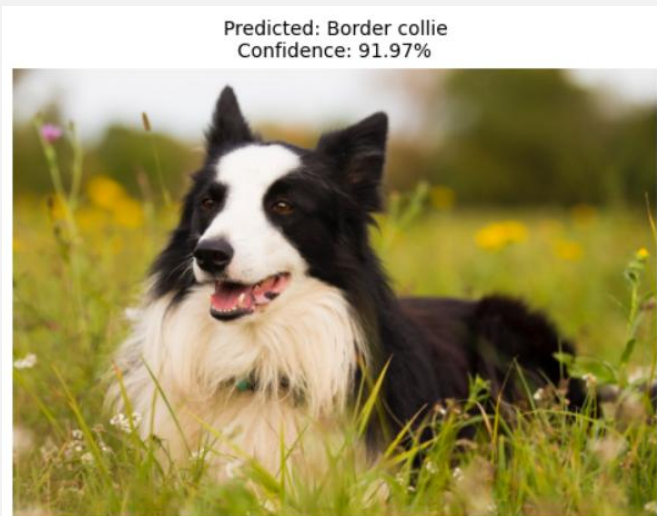
- 120개 품종을 분류하는 Fine-grained image Classification에서, EfficientNet-B0 모델을 10 에포크 동안 학습시킨 결과, 테스트 데이터에 대해 최고 74.16%의 분류 정확도를 얻을 수 있었다.
- 이는 초기 10 에포크 학습만으로는 비교적 어려운 다중 클래스 분류 문제에서 일정 수준의 분류 가능성을 확인한 결과이다.
- 그러나 학습 후반부의 과적합 경향과 테스트 정확도의 정체는 현재 모델 및 학습 전략의 한계를 보여주며, 추가적인 개선이 필요함을 시사한다.



결과 및 분석

5. 예측 수행 성공 사례

- 안전한(정적인) 보더콜리 이미지
- "Border collie"로 예측, 신뢰도 91.97%



6. 예측 수행 실패 사례

- 달리는(동적인) 보더콜리 이미지
- "Boston bull"로 예측, 신뢰도 35.74%



-> 모델이 완벽하지는 않다. 특정 조건에서는 성능이 저하될 수 있다.



결론

1. 본 연구는 120종의 다양한 강아지 품종 이미지를 포함하는 Stanford Dogs Dataset을 대상으로, 사전 학습된 EfficientNet-B0 모델에 전이 학습을 적용하여 품종 분류 모델을 개발하였다. 모델의 학습 효율성과 성능 향상을 위해 XML 어노테이션 기반의 객체 중심 이미지 크롭 및 데이터 증강 기법을 포함한 전처리 과정을 수행했으며, 총 10 에포크 동안 학습을 진행했다.
2. 10 에포크 학습 결과, 개발된 모델은 테스트 데이터셋에서 최고 74.16%의 분류 정확도를 달성했다. 이는 120개라는 다수의 세부 클래스를 구분해야 하는 세밀한 이미지 분류 문제에 대한 초기 실험임에도 불구하고, 딥러닝 모델이 강아지 품종을 식별할 수 있는 유의미한 가능성을 보여준 결과이다. 예측 수행 성공 사례에서, 특정 이미지에서는 모델이 품종의 주요 특징을 효과적으로 학습했음을 확인할 수 있었다.
3. 그러나 학습 과정 분석 결과, 학습 손실은 꾸준히 감소한 반면 테스트 손실은 4-5 에포크 이후 정체 및 소폭 상승하는 경향을 보였으며, 학습 정확도와 테스트 정확도 간의 격차 또한 점차 벌어지는 과적합(Overfitting) 현상이 관찰되었다. 이로 인해 최고 정확도 달성 이후 일반화 성능의 추가적인 향상은 제한적이었다. 예측 수행 실패 사례에서, 특정 조건의 이미지나 유사 품종 간의 구분에는 어려움이 있는 것으로 확인하였다. 따라서 현재 모델의 일반화 성능을 높이고 다양한 상황에 대처하기 위한 추가적인 개선 노력이 필요한 상황이다.



어려웠던 점 및 느낀 점

1. Stanford Dogs Dataset은 120종이라는 많은 품종을 포함하고 있으며, 각 품종 내에서도 개체별 외형, 자세, 촬영 환경 등이 매우 다양했다. 제공된 바운딩 박스를 활용하여 객체 중심으로 이미지를 크롭했지만, 일부 어노테이션의 정확성 문제나 이미지가 가려지는 문제 등으로 인해 일관된 품질의 학습 데이터를 확보하는데 어려움이 있었다. 또한, 최적의 데이터 증강 전략을 수립하기 위한 다양한 시도와 검증이 필요했다.
2. 학습 초기 단계부터 과적합의 징후가 빠르게 나타나, 이를 제어하고 모델의 일반화 성능을 끌어올리는 것이 가장 큰 도전 과제 중 하나였다. 단순히 학습을 오래 시키는 것만으로는 해결되지 않아, 정규화 기법, 데이터 증강의 강도 조절 등 다각적인 접근이 필요함을 느꼈다.
3. 제한된 시간과 컴퓨팅 자원 내에서 학습률, 배치 크기, 옵티마이저 설정 등 수많은 하이퍼파라미터의 최적 조합을 찾는 과정은 많은 인내를 필요로 했다. 각 파라미터 변화가 모델 성능에 미치는 영향을 체계적으로 분석하고 개선 방향을 설정하는 데 어려움을 겪기도 했다.
4. 단순히 모델 코드를 실행하는 것을 넘어, 데이터 수집 및 분석, 전처리, 모델 선정 및 구현, 학습, 평가, 그리고 결과 해석 및 개선에 이르는 프로젝트의 전체 파이프라인을 직접 경험하며 각 단계의 중요성과 연관성을 깊이 이해하게 되었다.



감사합니다.

