[빅데이터 9기] 팀 프로젝트

2024.09.06

**말 vs 사람 이미지 분류 모델**

**개발 및 성능 분석**

|  |  |
| --- | --- |
| 성명 | 역할 |
| 신유라 | [팀장] VGGNet 모델 학습 및 성능 평가, EDA, 데이터 취합 |
| 조아라 | [팀원] ResNet50 모델 학습 및 성능 평가 |
| 정승연 | [팀원] DenseNet121, Xception, InceptionV3 모델 학습 및 성능 평가 |
| 박만혜 | [팀원] MobileNetV2 모델 학습 및 성능 평가, 네 가지 모델 테스트 |

**[목차]**

1. 요약

2. 서론

3. 개발 환경

4. 개발 프로세스

5. EDA 및 전처리

6. 학습 모델

7. 결과

8. 분석 및 토의

9. 결론

10. 소스코드

# 1. 요약

본 보고서는 말과 사람의 이미지 데이터를 딥러닝 모델에 학습시켜 이진 분류를 효과적으로 수행하기 위한 분석 결과를 다룬다. Laurence Moroney에서 제공하는 **<말과 사람의 컬러 CG[[1]](#footnote-1) 이미지 데이터>**를 활용하였으며, 네 가지 딥러닝 모델(MobileNetV2, ResNet50, VGGNet16, DenseNet121)을 적용하여 학습을 진행하였다. 각 모델의 성능 평가는 정확도(Accuracy), 정밀도(Precision), 재현율(Recall), F1-score, ROC[[2]](#footnote-2)-AUC[[3]](#footnote-3)를 기준으로 이루어졌으며, 그 결과 **MobileNetV2가 가장 우수한 성능을 보였다.** 이때의 성능 지표는 **정확도 96%, 정밀도 93%, 재현율 98%, F1-Score 96%, ROC-AUC 95%**로 나타났다. 본 모델을 통해 이미지 기반 단어 학습, 해부학 교육 등의 에듀테크 분야에서 보다 개인화된 학습 환경을 제공할 수 있을 것으로 기대한다.

# 2. 서론

#### 1) 배경

최근 딥러닝 기술의 급속한 발전은 이미지 분류 문제에 대한 접근 방식을 새롭게 정의하고 있다. 특히 사람과 동물의 이미지를 구분하는 이진 분류 문제는 자율주행, 보안 시스템, 의료 진단 등 다양한 산업 분야에서 중요한 응용을 찾아가고 있으며, 그중 교육 기술(EdTech)은 이를 혁신적으로 활용할 가능성을 보여준다. 교육 현장에서는 학습자 맞춤형 콘텐츠 제공, 학습 패턴 분석, 실시간 피드백 제공 등에서 이미지 분류 기술이 중요한 역할을 할 수 있다.

#### 2) 기획의도

본 연구는 말과 사람의 이미지 데이터를 딥러닝 모델에 학습시켜 이진 분류를 효과적으로 수행하는 방법을 분석하는 것을 목적으로 한다. 이를 위해 MobileNetV2, ResNet50, VGGNet16, DenseNet121 네 가지 딥러닝 모델을 비교하고, 최종적으로 가장 높은 성능을 보이는 모델을 선정하여 최적의 이미지 분류 성능을 도출하고자 한다.

# 3. 개발 환경

|  |  |  |
| --- | --- | --- |
| 언어 및  라이브러리 | 버전 | 라이선스 |
| Python | 3.8.0 | Python Software Foundation License |
| TensorFlow | 2.13.0 | Apache License 2.0 |
| NumPy | 1.24.3 | BSD 3-Clause License |
| Matplotlib | 3.7.5 | Matplotlib License |
| Scikit-learn | 1.3.1 | BSD 3-Clause License |
| OpenCV | 4.10.0.84 | BSD 3-Clause License |

# 

# 4. 개발 프로세스

**Raw Data (n=1,027)**

말과 사람의 컬러 CG 이미지 데이터

**탐색적 데이터 분석 (EDA)**

데이터의 이미지 밝기/명암대비 분포 확인

**데이터 전처리**

정규화, 데이터 확장

**모델 설계**

MobileNetV2, ResNet50, VGGNet16, DenseNet121

**모델 성능 비교 & 테스트**

Accuracy, Precision, F1-Score, ROC-AUC, Comfusion Matrix

**최종 모델 선정**

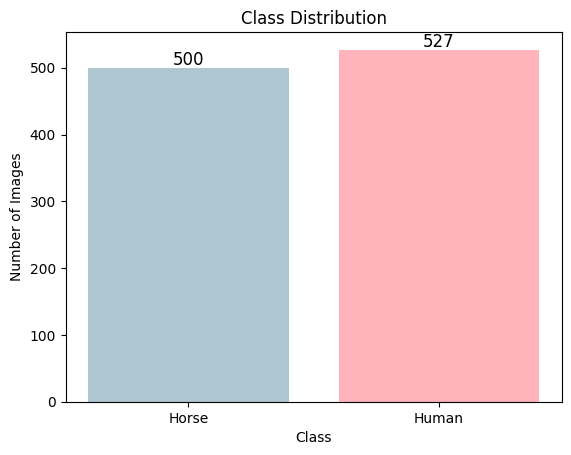
# 5. EDA 및 전처리

#### **1) 사용 데이터**

▪ 말과 사람의 컬러 CG 이미지 데이터[[4]](#footnote-4)

#### **2) 탐색적 데이터 분석(EDA[[5]](#footnote-5))**

▪ Train 데이터의 클래스 별 이미지 개수 시각화



▪ Train 데이터의 이미지 밝기/명암대비 분포

텍스트, 스크린샷, 도표이(가) 표시된 사진

자동 생성된 설명

#### **3) 데이터 셋 구성**

▪ Train 데이터 수 : 1,027개 → 5,135개로 증강

- horse : 500개 → 2,500개

- human : 527개 → 2,635개

▪ Validation 데이터 수 : 256개

- horse : 128개

- human : 128개

▪ Test 데이터 수 : 51개 → 200개로 증강

#### **4) 중복 데이터 확인**

▪ 중복된 이미지가 존재하지 않음

#### **5****) 데이터 변환**

|  |  |  |
| --- | --- | --- |
| 변환 방법 | 적용 데이터 | 설명 |
| 크기 조정 | Train,Validation,Test | 244x244 pixels 로 일괄 조정 |
| 정규화 | Train, Validation, Test | 이미지 데이터의 픽셀값을 0-255 사이의 정수 값에서 0-1 범위의 실수로 변환 |
| 데이터 증강 | Train | 주어진 Train 데이터 외에 5가지 변형[[6]](#footnote-6)을 가한 데이터를 추가로 학습시킴 |
| Test | 주어진 Test 데이터의 51개 이미지 대신 무작위로 선정한 200개의 이미지 데이터를 사용 |

#### **6) Train 데이터 증강[[7]](#footnote-7)**

▪ 다음 7가지 변형을 랜덤하게 적용하여 Train 데이터를 총 5,135개(원본 데이터 1,027개 포함)로 증강하였다.

|  |  |
| --- | --- |
| 변형 종류 | 상세 |
| 회전 | ‘45°, 90°, 135°, 180°, 225°, 270°, 315°’ 범위 내 랜덤 회전 |
| 크기 | 가로∙세로 방향으로 20% 범위 내 랜덤 확대∙축소 |
| 찌그러뜨림 | 가로∙세로 방향으로 20% 범위 내에서 랜덤 찌그러뜨림(shear) 변형 |
| 크롭 | 가로∙세로 방향으로 20% 범위 내 랜덤 이동 |
| 좌우반전 | 랜덤으로 좌우반전 |

# 6. 학습 모델

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 모델 | 사용 이유 | 레이어 수 | 에포크 수 | 배치 크기 |
| MobileNetV2 | 경량화된 구조로 제한된 자원 환경에서 효율적이며 빠른 학습 가능 | 157개   * 기본 레이어 154개 * 추가 레이어 3개   (GlobalAveragePooling2D, Dense, Dense) | 21  (Early Stopping 50) | 32 |
| ResNet50 | 잔차 학습을 통해 매우 깊은 신경망에서도 기울기 소실 문제 해결 | 180개   * 기본 레이어 177개 * 추가 레이어 5개   (Flatten, Dense, BatchNomalization, Dropout, Dense) | 10 | 32 |
| VGGNet16 | 단순한 구조로 미세한 이미지 특징까지 잘 학습 가능 | 20개   * 기본 레이어 16개 * 추가 레이어 4개   (Flatten, Dense, Dropout, Dense) | 11  (Early Stopping 50) | 64 |
| DenseNet121 | 레이어 간 강한 연결로, 피처를 재사용하여 적은 파라미터로도 성능 및 일반화 능력을 높일 수 있음 | 124 개   * 기본 레이어 121개 * 추가 레이어 3개   (Flatten, Dense, Dense) | 10 | 32 |

# 7. 결과

#### **1)** 학습 곡선

|  |  |
| --- | --- |
| **MobileNetV2** | **ResNet50** |
| **텍스트, 스크린샷, 도표, 라인이(가) 표시된 사진  자동 생성된 설명** | 텍스트, 스크린샷, 그래프, 라인이(가) 표시된 사진  자동 생성된 설명 |
| **VGGNet16** | **DenseNet121** |
| 텍스트, 스크린샷, 라인, 그래프이(가) 표시된 사진  자동 생성된 설명 |  |

#### **2) 모델 테스트 및 성능 비교**

① 성능 지표 분석[[8]](#footnote-8)[[9]](#footnote-9)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 모델 테스트 결과 | | | |
| **MobileNetV2** | **ResNet50** | **VGGNet16** | **DenseNet121** |
| Accuracy | 0.96 | 0.50 | 0.80 | 0.94 |
| Precision | 0.93 | 0.50 | 0.80 | 0.92 |
| Recall | 0.98 | 0.98 | 0.80 | 0.97 |
| F1-Score | 0.96 | 0.66 | 0.80 | 0.94 |
| ROC-AUC | 0.95 | 0.50 | 0.80 | 0.94 |

② Confusion Matrix

|  |  |
| --- | --- |
| **MobileNetV2** | **ResNet50** |
| 텍스트, 스크린샷, 도표, 직사각형이(가) 표시된 사진  자동 생성된 설명 | 텍스트, 스크린샷, 도표, 직사각형이(가) 표시된 사진  자동 생성된 설명 |
| **VGGNet16** | **DenseNet121** |
| 텍스트, 스크린샷, 도표, 직사각형이(가) 표시된 사진  자동 생성된 설명 | 텍스트, 스크린샷, 도표, 직사각형이(가) 표시된 사진  자동 생성된 설명 |

# 8. 분석 및 토의

**▪ 이미지 변형을 통한 데이터 증강**

기존 train 이미지들의 밝기 분포는 고르게 나타났으며, 명암 대비에서도 큰 차이를 보였다. 이에 따라, 데이터 증강 시 추가적인 밝기 및 명암 대비 조절은 필요하지 않았다. 대신, 회전, 크기 조정, 찌그러뜨림, 크롭, 좌우 반전의 5가지 변형을 적용하여 데이터 증강을 진행하였다.

**▪ MobileNetV2의 우수한 성능**

이번 프로젝트에서는 말과 사람의 이미지를 구분하는 이진 분류 문제를 해결하기 위해 네 가지 딥러닝 모델(MobileNetV2, ResNet50, VGGNet16, DenseNet121)을 사용하여 성능을 비교하였다. 성능 평가 결과, MobileNetV2는 정확도 0.96, 정밀도 0.93, 재현율 0.98, F1-Score 0.96, ROC-AUC 0.95로 다른 모델들보다 높은 성능을 기록하였다. 또한 학습 곡선에서 검증 정확도는 약 0.4에서 시작하여 3 에포크 이후 0.98에 도달하며 안정적인 성능을 유지하였다. 검증 손실은 0.5에서 시작해 3 에포크 이후 0.05 이하로 급격히 감소하여 과적합 없이 꾸준히 성능이 향상되는 것을 확인할 수 있었다.

MobileNetV2가 가장 우수한 성능을 보인 이유는 경량화된 구조(연산 복잡도와 파라미터 수를 줄임), 효율적인 계산 방식, 빠른 학습 속도, 그리고 과적합 방지 특성 때문이라고 분석하였다. 나머지 세 가지 모델들과 에포크 수, 배치 크기 같은 학습 파라미터 설정에 차이가 있었지만 MobileNetV2는 일관되게 최상의 성능을 보였다.

**▪ ResNet50, VGGNet16, DenseNet121의 성능 분석**

**ResNet50**의 경우, confusion matrix 분석 결과에서 대부분의 예측을 human으로만 분류하는 경향을 확인할 수 있었다. 이는 학습 과정에서 human 클래스에 대한 정보만 제대로 학습되었을 가능성이 크며, 모델이 horse 클래스의 이미지를 거의 학습하지 못했음을 시사한다. 결과적으로 비교적 저조한 성능을 보였다(정확도 0.50, 정밀도 0.50, 재현율 0.98, F1-Score 0.66, ROC-AUC 0.50).

**VGGNet16**은 정확도 0.80, 정밀도 0.80, 재현율 0.80, F1-Score 0.80, ROC-AUC 0.80 으로 중간 성능을 기록했다. VGGNet은 비교적 단순한 구조로 미세한 이미지 특징을 학습하는 데 강점을 보였지만, 성능은 MobileNetV2에 미치지 못했다. 또한 학습 속도가 상대적으로 느렸다.

**DenseNet121**은 정확도 0.94, 정밀도 0.92, 재현율 0.97, F1-Score 0.94, ROC-AUC 0.94 로 MobileNetV2에 근접한 성능을 보였으나, 학습 곡선에서 과적합이 발생한 것이 확인되었다. 에포크 2 이후로 Validation Accuracy가 감소하고 Validation Loss가 상승하는 경향이 나타나면서, 모델이 훈련 데이터에 과도하게 맞춰져 새로운 데이터에 대한 일반화 성능이 저하됨을 알 수 있다. 이로 인해, DenseNet121은 MobileNetV2에 비해 실시간 응용에 적합하지 않다고 판단하였다.

# 9. 결론

네 가지 딥러닝 모델(MobileNetV2, ResNet50, VGGNet16, DenseNet121)을 비교 분석한 결과, **MobileNetV2가 가장 우수한 성능을 보였다. MobileNetV2는 정확도 0.96, 정밀도 0.93, 재현율 0.98, F1-Score 0.96, ROC-AUC 0.95**로 다른 모델들보다 뛰어난 성능을 기록했으며, 경량화된 구조 덕분에 **빠르고 안정적인 학습**을 수행할 수 있었다. 특히, **과적합 없이 검증 성능을 꾸준히 유지**하면서도 최적의 학습 곡선을 보여주었다. 이와 달리, ResNet50과 DenseNet121은 학습 과정에서 과적합 문제를 겪었고, VGGNet16은 상대적으로 느린 학습 속도와 중간 성능을 기록했다.

결론적으로, MobileNetV2는 말과 사람의 이미지를 구분하는 이진 분류 문제에서 최적의 성능을 발휘했으며, 특히 에듀테크 분야에서 실시간 이미지 분류를 기반으로 한 다양한 학습 환경에 적용할 수 있을 것으로 예상된다. 본 모델을 통해 이미지 기반 단어 학습, 해부학 교육 등 다양한 교육 분야에서 보다 개인화된 학습 환경을 제공할 수 있을 것으로 기대된다.

# 10. 소스코드

▪ 프로젝트의 소스코드는 별첨 자료(‘EDA.ipynb’, ‘MobileNetV2\_TRAIN.ipynb', ‘MobileNetV2\_MODEL.keras', ‘MobileNetV2\_TEST.ipynb')로 제공되며, 다음의 주요 섹션으로 구성되어 있다:

▫ 탐색적 데이터 분석(EDA)

▫ 데이터 전처리

▫ 모델 학습

▫ 모델 테스트 및 성능 평가

1. Computer Graphics [↑](#footnote-ref-1)
2. Receiver Operating Characteristic [↑](#footnote-ref-2)
3. Area Under the Curve [↑](#footnote-ref-3)
4. Moroney, L. (발행 연도 미상). *Datasets for machine learning*. Laurence Moroney - The AI Guy. 조회일: 2024년 9월 5일, [https://laurencemoroney.com/datasets.html#horses-or-humans-dataset.](https://laurencemoroney.com/datasets.html#horses-or-humans-dataset) [↑](#footnote-ref-4)
5. Exploratory Data Analysis [↑](#footnote-ref-5)
6. ‘5. EDA 및 전처리 - 6) Train 데이터 증강’에서 자세하게 서술하였다 [↑](#footnote-ref-6)
7. ‘크기’, ‘찌그러뜨림’, ‘크롭’ 결과 발생한 빈 픽셀은 가장 가까운 픽셀로 채웠다(fill\_mode) [↑](#footnote-ref-7)
8. 성능 지표 값은 소수점 셋째 자리에서 반올림하였다 [↑](#footnote-ref-8)
9. 각 성능 지표별 최고 값에 하이라이트를 적용하였다 [↑](#footnote-ref-9)