

말 **vs** 사람 이미지 분류 모델

개발 및 성능 분석

[빅데이터 9기] 신유라, 박만혜, 조아라, 정승연

[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 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, ROC²-AUC³를 기준으로 이루어졌으며, 그 결과 **MobileNetV2**가 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 등 다양한 에듀테크 분야에서 활용될 수 있음을 제시한다.

2. 서론

1) 배경

최근 딥러닝 기술을 통한 이미지 분류 기술은 에듀 테크 분야에서 광범위하게 응용될 수 있는 잠재력을 가지고 있다. 예를 들어, 학습자의 표정이나 행동 패턴을 분석함으로써 학습자의 집중도나 감정 상태를 파악하고, 이를 바탕으로 학습 환경을 조정할 수 있으며 이미지 기반 단어 학습, 해부학 교육 등에도 응용이 가능하다. 이러한 이유로 본 프로젝트는 말과 사람을 구분하는 이진 분류 모델 개발을 통해 에듀 테크 분야에서의 딥러닝 기술 응용 가능성을 탐구하고자 한다.

2) 기획의도

본 프로젝트는 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 도출할 계획이다.

¹ CG : Computer Graphics

² ROC : Receiver Operating Characteristic

³ AUC : Area Under the Curve

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



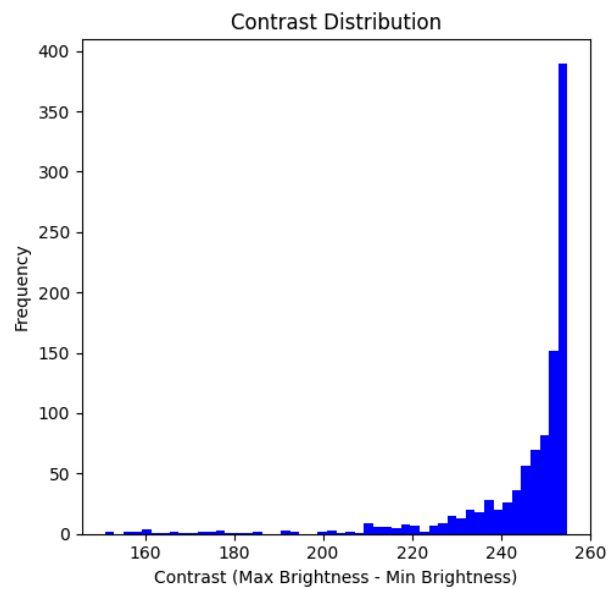
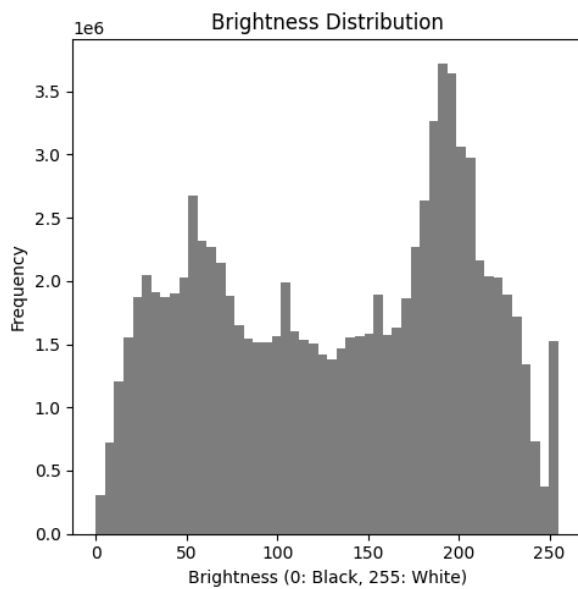
5. EDA 및 전처리

1) 사용 데이터

- 말과 사람의 컬러 CG 이미지 데이터⁴

2) 탐색적 데이터 분석(EDA⁵)

- Train 데이터의 클래스 별 샘플 수
 - 말 : 500개
 - 사람 : 527개
- Train 데이터의 이미지 밝기/명암대비 분포



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>.

⁵ EDA : Exploratory Data Analysis

3) 중복 데이터 확인

- 중복된 이미지가 존재하지 않음

4) 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 확장	Train	기존 Train 데이터 외에 5가지 변형 ⁶ 을 가한 데이터를 추가로 학습시킴
데이터 변경	Test	기존 Test 데이터의 51개 이미지 대신, 이미지 제공 플랫폼 ⁷ 의 API를 통해 horse or human에 해당하는 이미지를 랜덤으로 200개 생성

5) Train 데이터 확장⁸

- 랜덤 시드값 : 12345
- 다음 5가지 중 하나의 변형을 랜덤하게 적용하여 Train 데이터를 확장하였다.

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 '5. EDA 및 전처리 - 6) Train 데이터 확장'에서 자세하게 서술하였다

⁷ 본 테스트에 사용된 모든 이미지의 저작권은 Unsplash에 있으며, Unsplash의 라이선스에 따라 사용되었다

⁸ '크기', '찌그러뜨림', '크롭' 결과 발생한 빈 픽셀은 가장 가까운 픽셀로 채웠다(fill_mode)

6) 최종 데이터 셋 구성

▪ Train 데이터

▫ 총 데이터 수 : 2,054개

- 원본 데이터 : 1,027개 (Horse : 500개, Human : 527개)

- 확장 데이터 : 1,027개 (Horse : 500개, Human : 527개)

- 1 에포크 = 원본 데이터 + 확장 데이터

▪ Validation 데이터

▫ 총 데이터 수 : 256개 (Horse : 128개, Human : 128개)

▪ Test 데이터

▫ 총 데이터 수 : 200개 (Horse : 100개, Human : 100개)

7) 학습 설정

▪ 에포크 수 (Epochs) : 10

▪ 배치 크기 (Batch Size) : 32

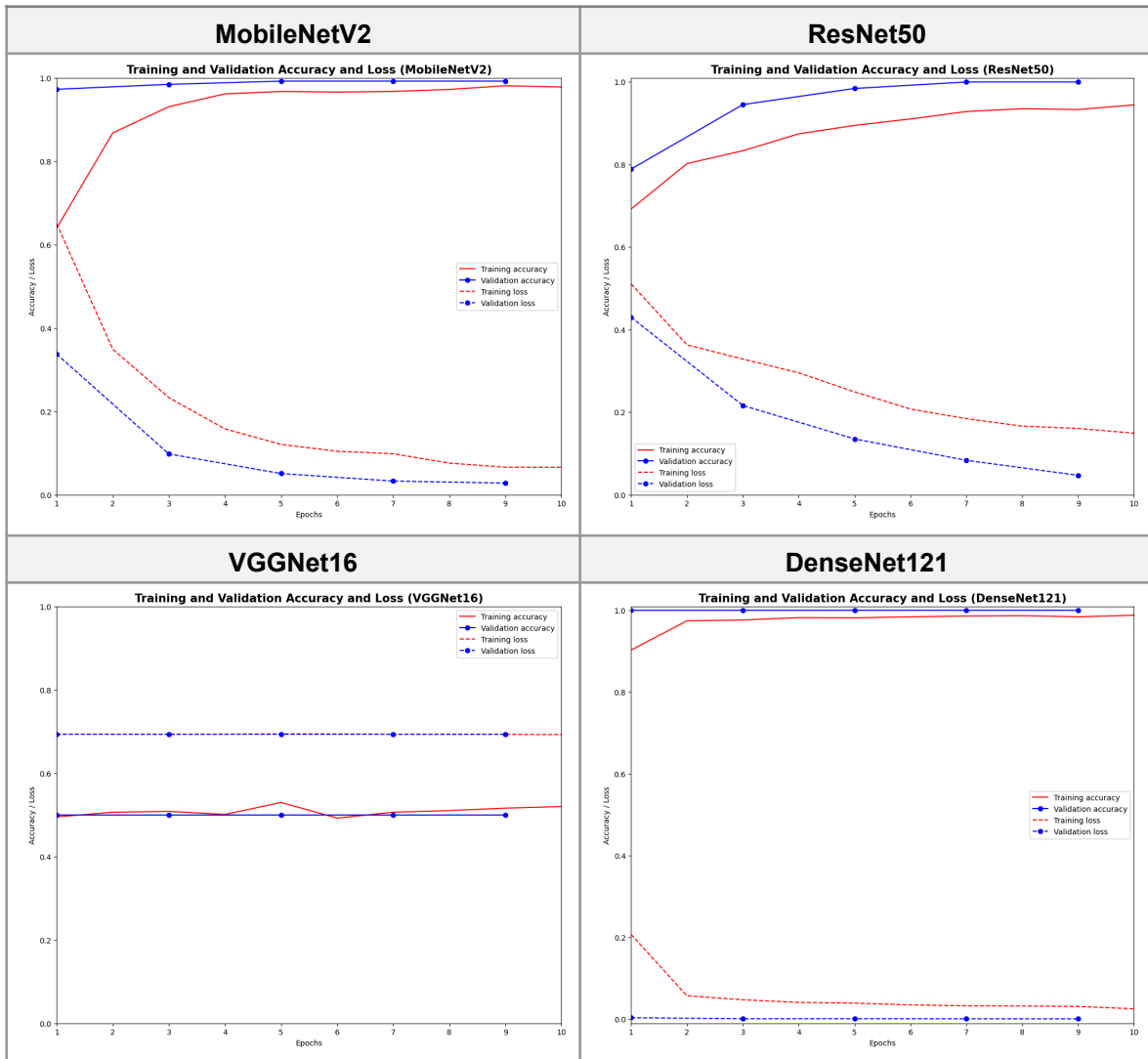
6. 학습 모델

- CNN 기반의 MobileNetV2, ResNet50, VGGNet16, DenseNet121 모델을 활용하여 이진 분류(Binary Classification) 학습을 진행하였다.

CNN 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)	10	32
ResNet50	잔차 학습을 통해 매우 깊은 신경망에서도 기울기 소실 문제 해결	182개 - 기본 레이어 177개 - 추가 레이어 5개 (GlobalAveragePooling2D, BatchNormalization, Dense, Dropout, Dense)	10	32
VGGNet16	단순한 구조로 미세한 이미지 특징까지 잘 학습 가능	20개 - 기본 레이어 16개 - 추가 레이어 4개 (Flatten, Dense, Dropout, Dense)	10	32
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

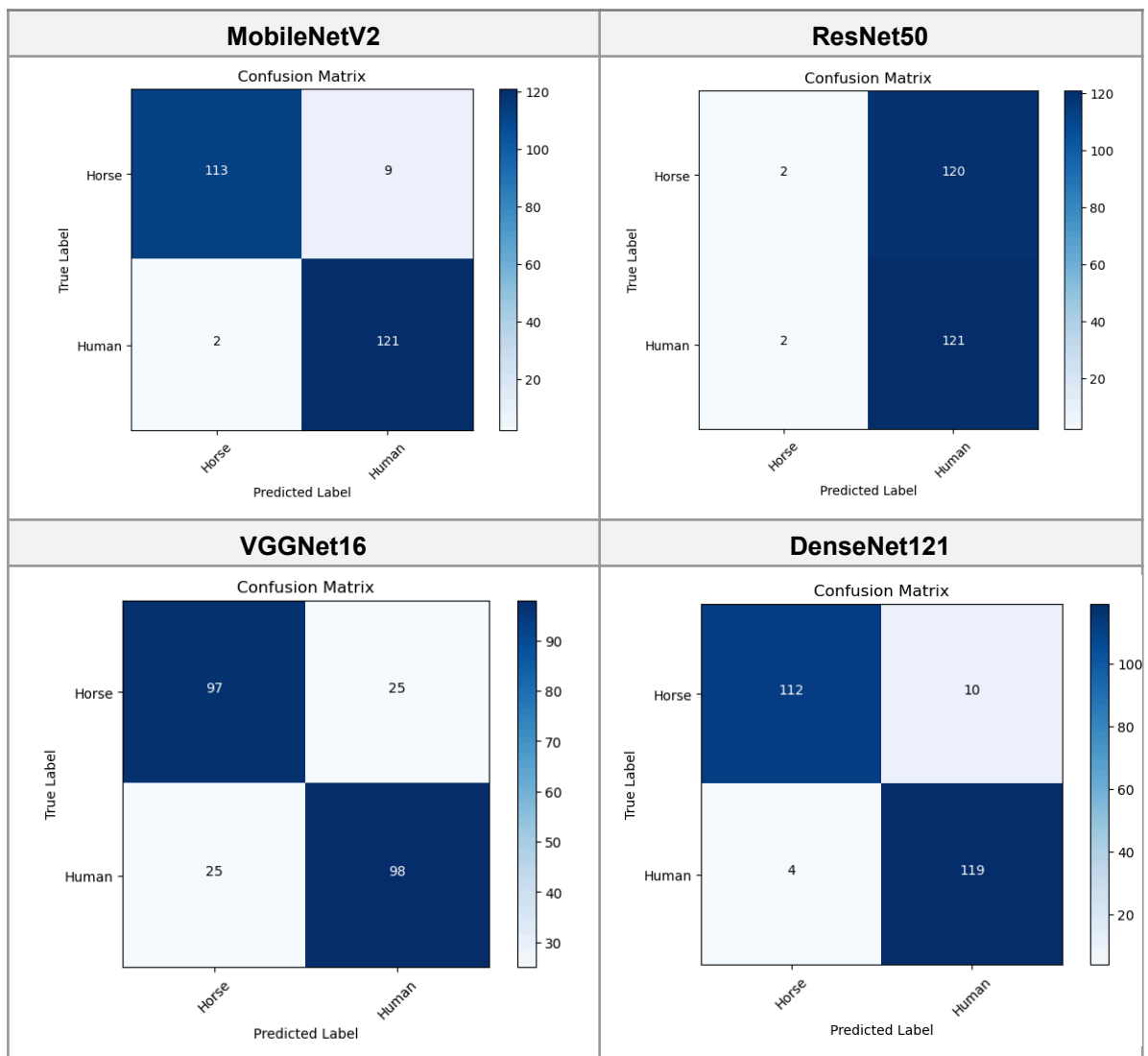


2) 모델 테스트 및 성능 비교

① 성능 지표 분석⁹¹⁰

	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



⁹ 성능 지표 값은 소수점 셋째 자리에서 반올림하였다

¹⁰ 각 성능 지표별 최고 값에 하이라이트를 적용하였다

8. 분석 및 토의

■ 이미지 변형을 통한 데이터 확장

기존 **train** 이미지들의 밝기 분포는 고르게 나타났으며, 명암 대비에서도 큰 차이를 보였다. 이에 따라, 데이터 확장 시 추가적인 밝기 및 명암 대비 조절은 필요하지 않았다. 대신, 회전, 크기 조정, 찌그러뜨림, 크롭, 좌우 반전의 5가지 변형을 적용하여 데이터 확장을 진행하였다. 이를 통해 기존 데이터셋의 제한된 양이나 패턴을 극복하고, 더 많은 경우의 수를 모델이 학습할 수 있게 하여 모델의 일반화 성능을 향상시키고자 하였다.

■ 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 정확도가 약 0.75에서 시작해 9 에포크에 도달했을 때 0.95 이상을 기록하여 안정적인 성능을 유지하였다. 검증 손실 또한 지속적으로 감소하여 0.05 이하로 떨어졌으며, 이는 과적합 없이 꾸준한 성능 향상을 보여준다. 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.

■ ResNet50, VGGNet16, DenseNet121의 성능 분석

ResNet50의 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).

VGGNet16은 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 정확도와 검증 정확도에 거의 변화 없이 일정한 수준에 머물러 있는 것으로 보아, 학습이 제대로 진행되지 않았거나 상대적으로 느리게 학습되고 있음을 추측할 수 있다.

DenseNet121은 정확도 0.94, 정밀도 0.92, 재현율 0.97, F1-Score 0.94, ROC-AUC 0.94로 MobileNetV2에 근접한 성능을 보였다. 하지만, 학습 곡선에서 학습 정확도가 매우 빠르게 1에 도달했으며, 이는 과적합의 징후로 볼 수 있다. 검증 정확도는 0.85을

기록하였으나 그 이후로 더 이상 향상되지 않고, 검증 손실이 지속적으로 유지되는 경향을 보였다. 이로 인해 **DenseNet121**은 학습 데이터에 과도하게 맞춰졌으며 새로운 데이터에 대한 일반화 성능이 저하된다는 것을 알 수 있다. 과적합 현상으로 인해 **DenseNet121**은 **MobileNetV2**에 비해 실시간 응용에는 적합하지 않다고 판단되었다.

9. 결론

네 가지 딥러닝 모델(**MobileNetV2**, **ResNet50**, **VGGNet16**, **DenseNet121**)을 비교 분석한 결과, **MobileNetV2**가 가장 우수한 성능을 보였다. **MobileNetV2**는 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 학습을 수행할 수 있었다. 특히, 과적합 없이 검증 성능을 꾸준히 유지하면서도 최적의 학습 곡선을 보여주었다. 이와 달리, **ResNet50**과 **DenseNet121**은 학습 과정에서 과적합 문제를 겪었고, **VGGNet16**은 상대적으로 느린 학습 속도와 중간 성능을 기록했다.

결론적으로, **MobileNetV2**는 말과 사람의 이미지를 구분하는 이진 분류 문제에서 최적의 성능을 발휘했으며, 빠르고 효율적인 분류 성능은 학습 중 발생하는 복잡한 이미지 처리를 즉각적으로 수행할 수 있기 때문에, 이미지 기반 교육에도 손쉽게 응용될 수 있을 것으로 기대된다.

10. 소스코드

▪ 프로젝트의 소스코드는 별첨 자료('EDA.ipynb', 'MobileNetV2_TRAIN_modify.ipynb', 'MobileNetV2_MODEL_modify.keras')로 제공되며, 다음의 주요 섹션으로 구성되어 있다:

- 탐색적 데이터 분석(EDA)
- 데이터 전처리
- 모델 학습
- 모델 테스트 및 성능 평가