키네틱스 휴먼 액션 비디오 데이터 세트

윌 케이
wkay@google.com

Joa˜o Carreira
joaoluis@google.com

카렌 시모니안
simonyan@google.com

브라이언 장
brianzhang@google.com

클로이 힐리어
chillier@google.com

수드헨드라 비자야나라심한
svnaras@google.com

파비오 비올라
fviola@google.com

팀 그린
tfgg@google.com

트레버 백
back@google.com

폴 나세프
natsev@google.com

무스타파 술레이만
mustafasul@google.com

앤드류 지서만
zisserman@google.com

초록

딥마인드 키네틱스의 인간 행동 비디오 데이터 세트에 대해 설명합니다. 이 데이터 세트에는 400개의 인간 행동 클래스와 각 행동에 대해 최소 400개의 비디오 클립이 포함되어 있습니다. 각 클립은 약 10초 분량이며 다른 YouTube 동영상에서 가져온 것입니다. 행동은 사람에 초점을 맞추고 있으며 악기 연주와 같은 인간과 사물 간의 상호작용뿐만 아니라 악수와 같은 인간과 인간 간의 상호작용을 포함한 광범위한 클래스를 포함합니다. 데이터 세트의 통계와 수집 방법을 설명하고, 이 데이터 세트에서 사람의 동작 분류를 위해 훈련 및 테스트된 신경망 아키텍처의 기본 성능 수치를 제공합니다. 또한 데이터 세트의 불균형이 분류기의 편향으로 이어지는지 여부에 대한 예비 분석도 수행합니다.

1. 소개

이 백서에서는 사람의 행동 분류를 위한 새로운 대규모 비디오 데이터 세트를 소개합니다. 이 데이터셋을 주로 개발한 이유는 인간 행동 분류를 위한 데이터셋이 부족하기 때문이며, 이 데이터셋이 딥 네트워크를 처음부터 훈련할 수 있을 만큼 충분히 크고 다른 아키텍처의 장점을 구분할 수 있는 벤치마크 역할을 할 수 있을 만큼 도전적이기 때문에 이 분야의 연구를 촉진할 수 있다고 믿기 때문입니다.

우리의 목표는 시간적 로컬라이제이션이 아닌 사람의 행동 *분류에* 사용할 수 있는 다양한 범위의 인간 행동을 포괄하는 대규모 고품질 데이터 세트를 제공하는 것입니다. 사용 사례가 분류이기 때문에 동작이 포함된 10초 내외의 짧은 클립만 포함되며, 트리밍되지 않은 영상은 포함되지 않습니다. 그러나 클립에는 사운드도 포함되어 있으므로 데이터 세트는 잠재적으로 많은 용도로 사용될 수 있습니다.

다중 모드 분석을 포함한 다양한 목적에 사용할 수 있습니다. 분류를 위한 데이터 세트를 제공하는 데 영감을 준 것은 ImageNet[18]으로, 분류를 위해 이 데이터 세트에서 딥 네트워크를 먼저 학습시킨 다음 학습된 네트워크를 다른 목적(감지, 이미지 분할, 비시각적 모달리티(예: 사운드, 깊이 등) 등)에 사용하는 것의 중요한 이점은 잘 알려져 있습니다.

키네틱스 데이터 세트는 이 분야의 표준 벤치마크로 부상한 두 가지 인간 행동 비디오 데이터 세트의 후속으로 볼 수 있습니다: HMDB-51 [15] 및 UCF-101 [20]. 이러한 데이터 세트는 커뮤니티에 매우 유용하게 사용되어 왔지만 이제 그 유용성이 만료되고 있습니다. 딥러닝을 기반으로 하는 현 세대의 인간 행동 분류 모델을 훈련하고 테스트하기에 충분히 크지 않거나 충분한 변이를 가지고 있지 않기 때문입니다. 공교롭게도 HMDB 데이터셋을 도입하게 된 동기 중 하나는 당시의 행동 데이터셋이 너무 작았기 때문입니다. 당시 10개였던 클래스를 51개로 늘렸고, 이후 400개까지 늘렸습니다.

표 1은 키네틱스의 크기를 여러 가지 인간 행동 데이터 세트와 비교한 것입니다. 변화 측면에서 보면, UCF-101 데이터 세트에는 101개의 동작과 각 동작에 대해 100개 이상의 클립이 포함되어 있지만, 모든 클립은 2.5k 디스크리트 비디오에서 가져온 것입니다. 예를 들어 같은 사람이 머리를 빗는 한 동영상에는 7개의 클립이 있습니다. 즉, 각 클립의 동작이 다른 사람(그리고 다른 시점, 조명 등)에 의해 형성되는 경우보다 훨씬 적은 변이가 있습니다. 키네틱스에서는 각 클립을 다른 동영상에서 가져오기 때문에 이 문제를 피할 수 있습니다.

클립은 YouTube 동영상에서 가져온 것입니다. 따라서 대부분의 경우 TV 및 영화 동영상처럼 전문적으로 촬영 및 편집된 자료가 아닙니다. 카메라 움직임/흔들림, 조도 변화, 그림자, 배경 잡음 등이 상당히 있을 수 있습니다. 더 많은 이미지

데이터 세트	연도	작업	클립	합계	동영상
HMDB-51 [15]	2011	51	최소 102	6,766	3,312
UCF-101 [20]	2012	101	min 101	13,320	2,500
ActivityNet-200 [3]	2015	200	평균 141	28,108	19,994
키네틱스	2017	400	최소 400	306,245	306,245

표 1: 최근 사람 동작 인식 데이터 세트에 대한 통계. '액션'은 액션 클래스 수, '클립'은 클래스당 클립 수, '총'은 총 클립 수, '비디오'는 이러한 클립이 추출된 총 비디오 수를 나타냅니다.

각 클립이 다른 동영상에서 가져온 것이기 때문에 출연자마다 액션 수행 *방식*(예: 속도), 의상, 몸의 포즈와 모양, 나이, 카메라 프레임과 시점에 차이가 있습니다.

이 데이터 세트를 통해 차세대 비디오용 신경망 아키텍처를 개발할 수 있기를 바랍니다. 예를 들어, 여러 가지 입력 스트림(RGB/외관, 광학적 흐름, 사람의 포즈, 사물 카테고리 인식)을 포함하는 아키텍처, 주의력을 사용하는 아키텍처 등이 있습니다. 이를 통해 새로운 아키텍처의 장점(또는 다른 장점)을 입증할 수 있습니다. 정적 예측과 모션 예측 사이의 긴장감, 비디오에서 시간적 집계를 위한 최적의 방법(리커런트 대 컨볼루션)에 대한 미해결 문제 등이 마침내 해결될 수 있습니다.

백서의 나머지 부분은 다음과 같이 구성됩니다: 섹션 2에서는 새로운 데이터 세트에 대한 개요를, 섹션 3에서는 데이터 수집 방법을 설명하고 데이터에서 발생할 수 있는 불균형과 분류기 편향에 대한 결과에 대해 논의합니다. 섹션 4에서는 데이터 세트에 대해 훈련되고 테스트된 여러 ConvNet 아키텍처의 성능을 제공합니다. 동반 논문 [5]에서는 키네틱스에서 행동 분류 네트워크를 사전 학습한 다음, 이 네트워크의 특징을 다른 (더 작은) 데이터 세트의 행동 분류에 사용할 때의 이점을 살펴봅니다.

YouTube 동영상의 URL과 데이터 세트의 시간 간격은 http://deepmind. com/kinetics에서 확인할 수 있습니다.

2. 키네틱스 데이터 세트 개요

콘텐츠: 이 데이터 세트는 활동이나 이벤트가 아닌 사람의 행동에 초점을 맞추고 있습니다. 액션 클래스 목록은 다음과 같습니다: *사람 행동(단수)*(예: 그림 그리기, 술 마시기, 웃기, 주먹 쥐기), *사람-사람 행동*(예: 포옹하기, 키스하기, 악수하기), *사람-물체 행동*(예: 선물 열기, 잔디 깎기, 설거지하기)이 있습니다. 일부 행동은 세분화되어 있으며, 예를 들어 다양한 유형의 수영을 구별하기 위해 시간적 추론이 필요합니다. 다른 행동은 여러 종류의 관악기 연주와 같이 구별할 대상에 더 중점을 두어야 합니다.

깊은 계층 구조는 없지만 대신 음악(드럼, 트롬본, 바이올린 연주, ...), 개인 위생(양치질, 손톱 깎기, 손 씻기, ...), 춤과 같은 여러 (배타적이지 않은) 부모-자식 그룹이 있습니다.

(발레, 마카레나, 탭, ...); 요리 (자르기, 튀기기, 껍질 벗기기, ...). 전체 클래스 목록은 부록에 상위-하위 그룹과 함께 나와 있습니다. 그림 1은 클래스 샘플의 클립을 보여줍니다.

통계: 데이터 세트에는 400개의 인간 행동 클래스가 있으며, 각 행동에 대해 각각 고유한 비디오에서 400~1150개의 클립이 있습니다. 각 클립은 10초 정도 지속됩니다. 현재 버전에는 306,245개의 동영상이 있으며, 클래스당 250~1000개의 동영상이 있는 훈련용, 클래스당 50개의 동영상이 있는 검증용, 클래스당 100개의 동영상이 있는 테스트용의 세 가지 분할로 나뉩니다. 통계는 표 2에 나와 있습니다. 클립은 YouTube 동영상에서 가져온 것으로 해상도와 프레임 속도가 가변적입니다.

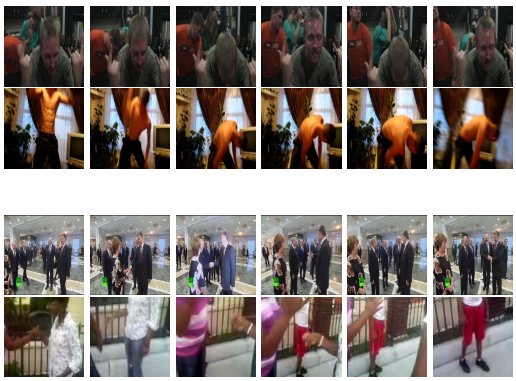
기차	유효성 검사	테스트
250-1000	50	100

표 2: 키네틱스 데이터 세트 통계. 훈련/밸/테스트 파티션의 각 클래스에 대한 클립 수입니다.

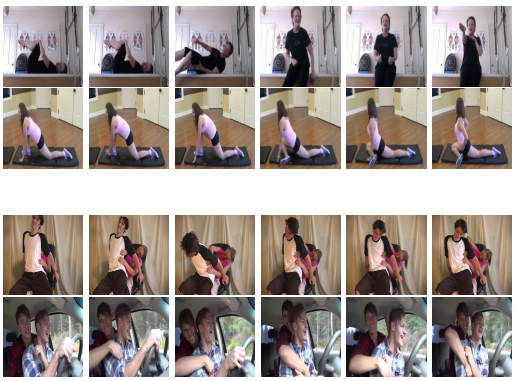
완전하지 않은 주석. 각 클래스에는 해당 동작을 설명하는 클립이 포함되어 있습니다. 그러나 특정 클립에는 여러 동작이 포함될 수 있습니다. 데이터 세트의 흥미로운 예는 다음과 같습니다: '자동차 운전' 중 '문자 메시지 보내기', '우쿨렐레 연주' 중 '훌라후프 돌리기', '춤추기' 중 '양치질하기'(어떤 유형의 춤) 등이 있습니다. 각각의 경우 두 동작 모두 키네틱스 클래스이며, 클립은 이 클래스 중 하나에만 표시되고 둘 다에는 표시되지 않을 수 있습니다. 즉, 클립에는 완전한(완전한) 주석이 없습니다. 이러한 이유로 분류 성능을 평가할 때는 상위 1보다 상위 5 측정값이 더 적합합니다. 이는 상위 5개 측정값을 사용하는 이유 중 하나가 이미지가 여러 클래스를 포함할 수 있지만 단일 클래스에 대해서만 레이블이 지정되기 때문인 ImageNet[18]의 상황과 유사합니다.

3. 데이터 세트 구축 방법

이 섹션에서는 수집 프로세스, 즉 YouTube에서 후보 동영상을 확보한 방법과 후보를 선택하는 데 사용된 처리 파이프라인에 대해 설명합니다.



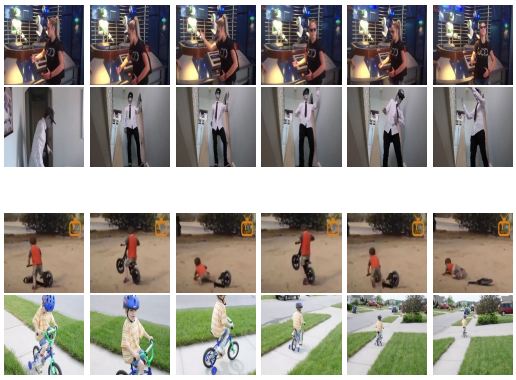
(a) 헤드뱅잉



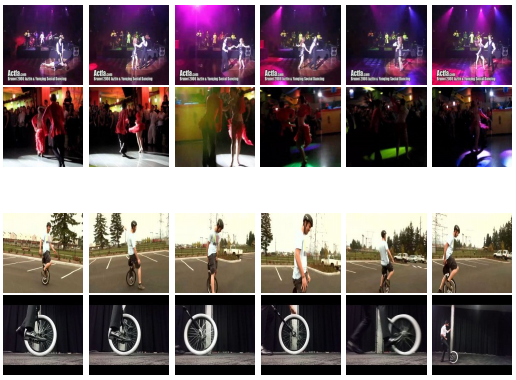
(b) 다리 스트레칭

(c) 악수

(d) 간지럼



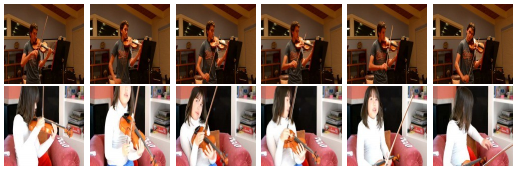
(E) 로봇 댄스



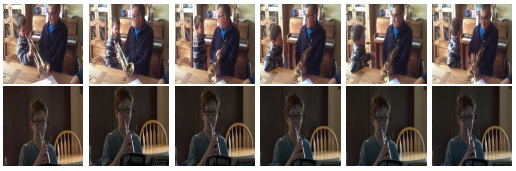
(F) 살사 댄스

(g) 자전거 타기

(h) 외발자전거 타기



(i) 바이올린 연주

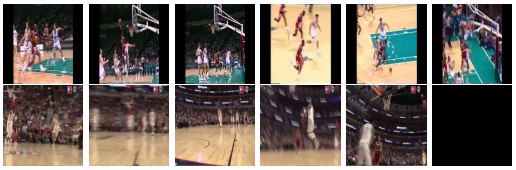


(j) 트럼펫 연주

(케이) 머리 땋기

(엘) 머리 빗기

(남) 드리블 농구



(여) 덩크슛 농구

그림 1: 운동학 데이터 집합의 클래스 예시. 컬러로 확대해서 보는 것이 가장 좋습니다. 어떤 경우에는 단일 이미지만으로는 동작(예: "헤드뱅잉")을 인식하거나 클래스("드리블 농구" 대 "덩크슛 농구")를 구분하는 데 충분하지 않을 수 있다는 점에

유의하세요. 데이터 세트에는 다음이 포함됩니다: 단수 사람 동작(예: "로봇 춤추기", "다리 뻗기"), 사람 대 사람 동작(예: "악수하기", "간지럼"), 사람 대 사물 동작(예: "자전거 타기"), 같은 동사 다른 사물(예: "바이올린 연주", "트럼펫 연주"), 같은 동사 다른 사물(예: "농구 드리블", "덩크슛 농구") 등이 있습니다. 이러한 동영상은 사실적인 (아마추어) 동영상으로, 예를 들어 카메라 흔들림이 심한 경우가 많습니다.

를 클릭하고 데이터 집합을 정리합니다. 그런 다음 수집 프로세스로 인해 데이터 세트에 발생할 수 있는 편향에 대해 논의합니다.

개요: 각 클래스의 클립은 먼저 YouTube에서 후보를 검색한 다음, AMT(Ama- zon Mechanical Turkers)를 사용하여 클립에 액션이 포함되어 있는지 여부를 결정하여 확보했습니다. 5점 만점 중 3점 이상의 확인을 받아야 클립이 허용되었습니다. 각 비디오에서 하나의 클립만 가져왔는지, 클립에 일반적인 비디오 자료가 포함되어 있지 않은지 확인하여 데이터 세트의 중복성을 제거했습니다. 마지막으로 클래스의 중복 및 노이즈 제거를 확인했습니다.

이제 이러한 단계를 더 자세히 설명합니다.

3.1. 1단계: 작업 목록 가져오기

이 정도의 규모에 적합한 시각적 액션 클래스가 있는 단일 목록이 없기 때문에 대규모의 인간 행동 목록을 큐레이팅하는 것은 어려운 작업입니다. 따라서 우리는 우리를 둘러싼 행동에 대한 자체 관찰과 함께 수많은 소스를 결합해야 했습니다. 이러한 소스는 다음과 같습니다: (i) **액션 데이터셋** - Ac- tivityNet [3], HMDB [15], UCF101 [20], MPII Human과 같은 기존 데이터셋
포즈 [2], ACT [25]에는 유용한 클래스가 있으며 이 중 적절한 하위 집합을 사용했습니다. (ii) **모션 캡처** - 수많은 모션 캡처 데이터셋을 살펴보고 파일 제목을 추출했습니다. 이러한 제목은 파일 내의 동작을 설명하며 종종 매우 창의적이었습니다. (iii) **크라우드 소싱** - 클립에 대해 제시했던 라벨이 잘못된 경우 Mechanical Turk 작업자에게 더 적절한 작업을 제안해 달라고 요청했습니다.

3.2. 2단계: 후보 클립 가져오기

선택한 방법과 단계는 여러 가지 내부 노력을 결합하여 아래에 자세히 설명되어 있습니다:

1단계: 비디오 가져오기. 동영상 제목을 키네틱스 작업 목록과 일치시켜 YouTube 말뭉치에서 동영상을 가져옵니다
.

2단계: 비디오 내 시간적 위치 지정. 이미지 분류기는 수많은 사람의 행동에 사용할 수 있습니다. 이러한 분류기는 Google 이미지 검색에서 사용자 행동을 추적하여 얻을 수 있습니다. 예를 들어 '나무 등반'이라는 검색 쿼리의 경우, 해당 검색 쿼리가 여러 번 실행된 횟수를 집계하여 이미지에 대한 사용자 관련성 피드백을 수집합니다. 이 관련성 피드백은 '등산 나무' 이미지 분류기를 훈련하는 데 사용할 수 있는 신뢰도가 높은 이미지 집합을 선택하는 데 사용됩니다. 이러한 분류기는 1단계에서 찾은 동영상에 대해 프레임 수준에서 실행되며, 상위 k*개의* 응답(여기서 $k = 2$)을 중심으로 클립이 추출됩니다.

작업 동사가 다음과 같이 끝나도록 형식이 지정되면 작업 목록이 관련 분류자와 더 잘 일치하는 것으로 나타났습니다.

'ing'. 이미지 검색을 떠올려보면, 일반적으로 어떤 행동을 하는 사람의 예를 검색할 때 '달리는 사람' 또는 '머리를 빗는 사람'과 같은 시제보다 '달리는 사람' 또는 '머리를 빗는 사람'과 같은 쿼리를 사용하는 것이 더 합리적입니다.

이 단계의 결과물은 많은 수의 동영상과 모든 동영상에서 동작 중 하나가 잠재적으로 발생할 수 있는 위치입니다. 해당 위치의 양쪽에서 5초씩 촬영하여 10초 클립을 만듭니다(해당 위치가 동영상의 시작 또는 끝에서 5초 이내인 경우 길이 예외가 적용되어 클립 길이가 더 짧아집니다). 그런 다음 클립은 사람이 라벨링을 통해 다음 정리 단계로 넘어갑니다.

3.3. 3단계: 수동 라벨링 프로세스

이 단계의 핵심 목표는 클립 중에 예상되는 동작이 실제로 발생했는지 여부를 식별하는 것이었습니다. 이 단계에는 사람이 필요했고, 플랫폼을 사용하는 고급 작업자가 많기 때문에 이 작업을 위해 Amazon의 Mechanical Turk(AMT)를 사용하기로 결정했습니다.

라벨링 작업을 위해 단일 페이지 웹앱을 구축했으며, 작업자에게 제공되는 클립 수를 최대화하면서 주석의 품질을 높게 유지하도록 최적화했습니다. 라벨링 인터페이스는 그림 2에 나와 있습니다. 사용자 인터페이스 디자인과 테마는 플랫폼의 다른 많은 작업과 차별화하고 작업을 최대한 자극적이고 흥미롭게 만들기 위해 선택되었습니다. 이 작업은 플랫폼에서 가장 높은 평가를 받았고 새 실행이 시작되자마자 400명 이상의 작업자가 참여하는 경우가 많았기 때문에 확실히 성과를 거두었습니다

.

작업자들은 처음에 명확한 지침을 받았습니다. 두 개의 지시 화면이 있었고, 두 번째 화면은 첫 번째 화면을 다시 반복했습니다. 작업자가 작업을 이해했음을 확인한 후 미디어 플레이어와 몇 가지 응답 아이콘이 표시되었습니다. 인터페이스는 그 순간 작업자가 사용할 수 있는 풀에서 비디오 세트를 가져와 첫 번째 클립을 삽입했습니다. 이 과제는 가능한 한 각기 다른 클래스의 20개 동영상으로 구성되었으며, 작업자의 흥미를 높이고 수익률이 낮은 클래스에 고착되는 것을 방지하기 위해 모든 동영상과 클래스를 무작위로 선정했습니다. 비디오 슬롯 중 두 개는 실사 클립을 삽입하는 데 사용했습니다. 이를 통해 각 작업자의 정확도를 추정할 수 있었습니다. 작업자의 성공률이 50% 미만이면 '낮은 정확도' 경고 화면을 표시했습니다. 이를 통해 많은 낮은 정확도를 해결하는 데 도움이 되었습니다.

라벨링 인터페이스에서 작업자에게 "작업 클래스 이름을 수행하는 사람을 볼 수 있습니까?"라는 질문을 던졌습니다. 인터페이스에는 다음과 같은 응답 옵션이 아이콘으로 표시되었습니다:

- 예, 여기에는 실제 작업의 예가 포함되어 있습니다.
- 아니요, 여기에는 작업의 예가 포함되어 있지 않습니다.

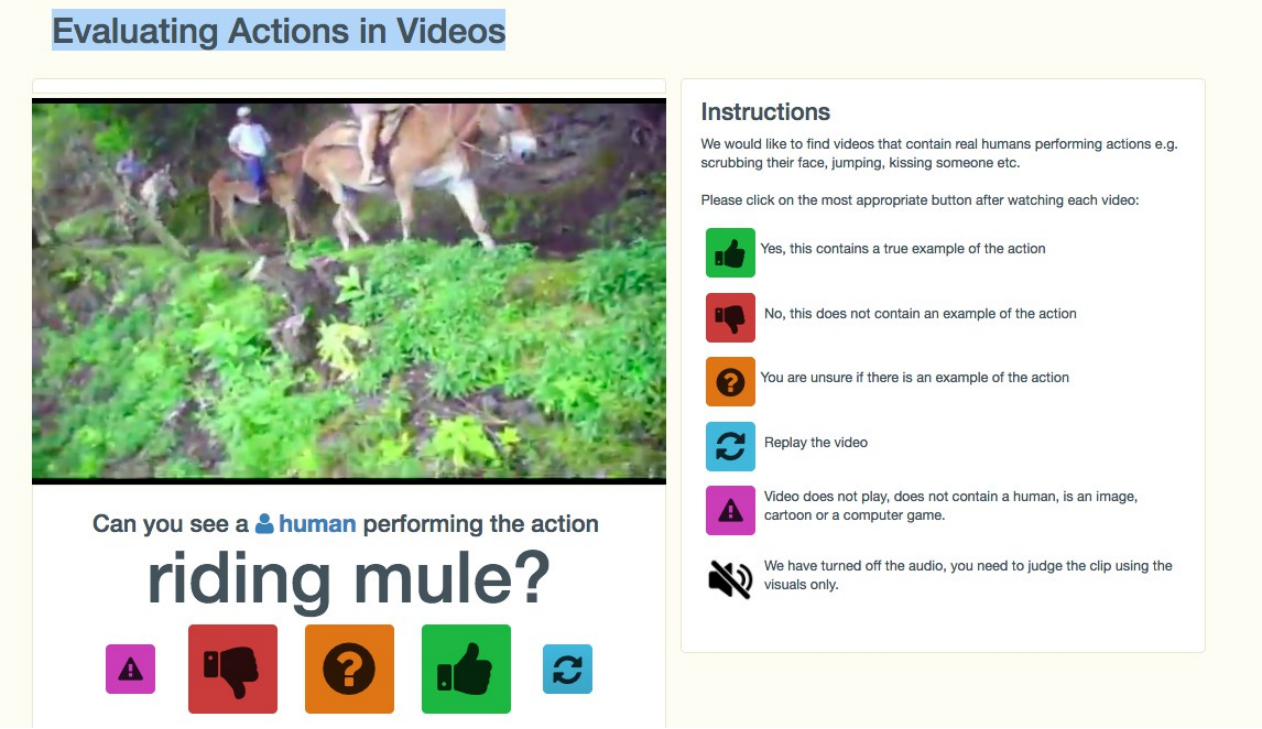


그림 2: Mechanical Turk에서 사용되는 라벨링 인터페이스.

- 작업의 예가 있는지 확실하지 않습니다.
- 동영상 재생
- 동영상이 재생되지 않거나, 사람이 포함되지 않거나, 이미지, 만화 또는 컴퓨터 게임입니다.

작업자가 '예'라고 응답하면 나중에 모델 학습 중에 이 신호를 사용하기 위해 "해당 동작이 전체 클립에서 지속되나요?"라는 질문도 던졌습니다.

비디오가 시각적 콘텐츠만을 기준으로 분류될 수 있도록 하기 위해 AMT 작업자는 오디오에 액세스할 수 없었습니다.

클립이 데이터 세트에 추가되려면 작업자로부터 최소 3개의 긍정적인 응답을 받아야 했습니다. 특정 응답에 2개 이상의 주석이 달린 경우를 제외하고 각 클립에 5번까지 주석을 달 수 있도록 허용했습니다. 예를 들어, 작업자 3명 중 3명이 행동의 예가 포함되어 있지 않다고 답한 경우 즉시 풀에서 삭제하고 5명의 작업자가 주석을 달 때까지 계속 진행하지 않았습니다.

작업의 규모가 크기 때문에 품질이 낮거나 전혀 관련이 없는 후보로 구성된 수업을 신속하게 제거해야 했습니다. 이 작업을 수행하지 못하면 동영상을 부정적 또는 불량으로 표시하는 데 많은 비용을 지불해야 했습니다. 각 클래스의 정확도는 해당 클래스의 클립 20개에 노트를 붙인 후 계산했습니다. 실행할 때마다 정확도 임계값을 조정했지만 일반적으로 50%의 높은 정확도(동영상 2개 중 1개에 해당 동작이 포함될 것으로 예상됨)에서 시작했습니다.

주석을 단 후, 비디오 ID, 클립 시간 및 레이블을 데이터베이스에서 내보내어 모델 학습에 사용할 수 있도록 전달했습니다.

우리가 배운 것: '노새 타기'와 같은 보다 구체적인 클래스가 '승마'와 같은 일반적인 클래스보다 훨씬 적은 소음을 발생시킨다는 사실을 발견했습니다. 그러나 때때로 더 일반적인 클래스를 사용하면 이전에는 존재하지 않았던 몇 가지 다른 클래스로 세분화할 수 있고 지원자가 근로자에게 재전송할 수 있기 때문에 이점이 있었습니다.
예를 들어 '정원 가꾸기'는 '식물에 물주기', '나무 다듬기', '나무 심기'로 나뉩니다.

작업이 생성하는 워커 트래픽의 양이 너무 많아서 적절한 인덱스와 최적화된 쿼리를 사용하더라도 데이터베이스에 직접 가져오기 및 쓰기에 의존할 수 없었습니다. 따라서 각 작업자에 대한 클립 그룹으로 구성된 많은 캐시를 만들었습니다. 작업자가 새 작업을 시작하면 인터페이스는 해당 특정 작업자에 대한 클립 세트를 가져왔습니다. 클립이 충분한 수의 주석을 받으면 백그라운드 프로세스에 의해 캐시가 자주 보충되었습니다. 또한 이전에는 1명이 *넘는* 작업자가 동일한 동영상에 주석을 달기 때문에 하나의 클립에 대해 5개의 응답을 빠르게 초과하는 라벨링 충돌도 방지할 수 있었습니다.

3.4. 4단계: 정리 및 노이즈 제거

데이터 세트 설계 목표 중 하나는 반복적인 동작이 포함된 비디오를 여러 개의 (상호 연관된) 훈련 예시로 분할하는 기존 데이터 세트와는 달리 주어진 비디오 시퀀스에서 단일 클립을 사용하는 것이었습니다. 또한 데이터 세트를 확장하면서 동음이의어 또는 다른 어순으로 인해 반복되는 클래스(예: 오토바이 타기, 오토바이 타기), 너무 일반적이고 다른 많은 클래스(예: 말하기)와 함께 발생하며 일반적인 1-of-K 분류 학습 접근법(다중 라벨 분류 대신)에서 문제가 되는 클래스 등 구조적 문제를 식별하는 메커니즘을 사용했습니다. 이제 이러한 문제를 해결하는 방법을 설명하겠습니다.

동영상 중복 제거. 두 가지 보완적인 접근 방식을 사용하여 동영상 중복을 제거했습니다. 먼저, 각 YouTube 링크에서 하나의 클립만 가져오기 위해 터커가 해당 동영상에 대해 검증한 클립 중에서 무작위로 하나의 클립을 선택했습니다. 이 단계에서는 터커가 승인한 예시 중 약 20%를 걸러냈지만, 여전히 많은 중복이 남아있음을 육안으로 확인할 수 있었습니다. 그 이유는 YouTube 사용자가 동영상 편집이나 홍보 광고의 일부로 다른 동영상의 일부를 재사용하여 동영상을 제작하는 경우가 많기 때문입니다. 때로는 자르고, 크기를 조정하고, 일반적으로 다른 방식으로 사전 처리하기도 합니다(그럼에도 불구하고 이미지 분류기는 동일한 클립을 로컬라이즈할 수 있습니다). 따라서 각 클립이 서로 다른 비디오에서 가져온 것이라도 중복이 발생했습니다.

각 클래스별로 독립적으로 작동하는 YouTube 링크 간 중복 제거 프로세스를 고안했습니다. 먼저 각 동영상에서 25개의 단일 샘플링된 프레임으로 구성된 224×224 센터 크롭에서 Inception-V1[12] 특징 벡터(마지막 평균 풀링 레이어 이후)를 계산한 다음, 이를 에지화했습니다. 그런 다음 이러한 특징 벡터 간의 모든 코사인 유사성을 갖는 클래스별 행렬을 구축하고 임계값을 적용했습니다. 마지막으로 연결된 컴포넌트를 계산하고 각 컴포넌트에서 무작위로 예시를 하나씩 추출했습니다. 0.97이라는 동일한 임계값을 사용하여 대부분의 클래스에서 이 방법이 잘 작동한다는 것을 확인했지만, 눈이나 물속에서 일어나는 일부 클래스와 같이 시각적으로 유사한 몇 가지 경우에 임계값을 조정했습니다. 이 과정을 통해 터키에서 승인한 시험의 수가 15% 더 감소했습니다.

노이즈가 많은 클래스 감지. 클래스는 다른 클래스와 겹치거나 클래스 이름의 모호함으로 인해 (동작 측면에서) 상당히 구분되는 여러 개의 그룹을 포함할 수 있다는 점에서 '노이즈'가 있을 수 있습니다. 예를 들어 '줄넘기'는 '밧줄로 줄넘기'일 수도 있고 '물 건너 돌넘기'일 수도 있습니다. 이러한 노이즈 클래스를 식별하기 위해 데이터 세트 개발 과정에서 2스트림 액션 분류기[19]를 반복적으로 훈련했습니다. 이를 통해 각 클래스에 대한 최고의 융합을 찾을 수 있었는데, 때로는 클래스 이름만 확인해도 명확했지만 (하지만

를 데이터 세트의 규모에 맞게 조정해야 하는 경우도 있었고, 혼동이 정상인지, 아니면 모델의 단점 때문에 클래스를 구분하기 어려운 것인지 파악하기 위해 데이터를 직접 눈으로 확인해야 하는 경우도 있었습니다. 이렇게 감지된 혼동을 기반으로 클래스를 병합, 분할 또는 완전히 제거했습니다.

최종 필터링. 모든 데이터를 수집하고, 중복을 제거하고, 클래스를 선택한 후 최종 수동 클립 필터링 단계를 실행했습니다. 이 단계에서는 두 가지 스트림 모델의 클래스 점수가 다시 유용했는데, 가장 자신 있는 예제부터 가장 자신 없는 예제까지 분류할 수 있어 예제가 얼마나 프로토타입적인지 확인할 수 있었기 때문입니다. 노이즈가 많은 예제들이 가장 낮은 순위에 있는 경우가 많다는 것을 발견하고 이 예제들에 집중했습니다. 또한 순위를 매기다 보니 중복되는 동영상이 인접하게 표시되어 이러한 동영상도 쉽게 필터링할 수 있었습니다.

3.5. 토론: 데이터 세트 편향 I

우리는 일반화 부족으로 이어지는 데이터 세트 편향이라는 개념에 익숙합니다. 한 데이터 세트(예: Caltech 256 [10])에서 훈련된 분류기가 다른 데이터 세트(예: PASCAL VOC [8])에서 테스트했을 때 제대로 작동하지 않는 경우입니다. 실제로 어떤 이미지가 어떤 데이터 세트에 속하는지 식별하도록 분류기를 훈련하는 것도 가능합니다[22].

데이터 세트 *내의* 불균형한 카테고리로 인해 발생할 수 있는 또 다른 편향성이 있습니다. 예를 들어, 훈련 세트의 성별 불균형은 이 세트에 대해 훈련된 분류기의 성능 편향으로 이어질 수 있습니다. 이러한 선례는 공개적으로 사용 가능한 얼굴 감지기가 인종에 구애받지 않는 것과 같은 예가 있습니다.[1]그리고 최근에는 서면 텍스트의 의미 편향 학습에서 이러한 선례가 있습니다[4]. 따라서 키네틱스가 이러한 편향을 유발하는지 여부는 중요한 질문입니다.

이를 위해 (i) 키네틱스의 각 액션 클래스에 대한 데이터가 성별에 따라 균형이 잡혀 있는지, (ii) 불균형이 있다면 이로 인해 액션 클래스가 편향된 성능을 보이는지 등에 대한 예비 연구를 수행했습니다.

(i)의 결과는 400개 중 340개의 액션 클래스에서 데이터가 한 성별에 의해 지배되지 않거나 대부분 성별을 확인할 수 없다는 것입니다. 후자는 예를 들어 손만 나타나거나 '배우'가 너무 작거나 옷을 많이 입는 클래스에서 발생합니다. 성별 불균형을 보이는 수업으로는 '수염 깎기'와 '덩크슛 농구'는 대부분 남성이, '눈썹 채우기'와 '치어리딩'은 대부분 여성이 수강하는 수업이 있습니다.

이러한 클래스에 대한 (ii)의 결과, 성별 불균형이 있는 액션 클래스에 대한 분류자 편향의 증거는 거의 발견되지 않았습니다. 예를 들어, 남성 플레이어가 더 많은 경향이 있는 '포커 플레이'의 경우, 여성 플레이어가 등장하는 모든 동영상이 올바르게 분류되었습니다. '해머 던지기'도 마찬가지입니다. 이러한 편향성이 없는 이유는 분류기가 다음과 관련된 객체를 모두 활용할 수 있기 때문이라고 추측할 수 있습니다.

[1] https://www.media.mit.edu/posts/미디어-랩-학생-인정-편향성-대응-학습으로-인정받은-기계학습

단순히 외모가 아닌 동작 패턴과 동작을 분석합니다.

연령과 인종 등 다른 '축'에서도 불균형을 조사할 수 있습니다. 다시 한 번 말씀드리지만, 예비 조사에서 뚜렷한 편향성은 거의 발견되지 않았습니다. 아기에게 명백한 편향성이 있는 한 가지 예외가 있는데, '울음'의 경우 아기가 아닌 동영상이 잘못 분류된 경우가 많았고, '레슬링'은 그 반대의 경우로, 링에서 레슬링하는 성인이 집에서 레슬링하는 어린이보다 더 잘 분류된 것처럼 보이지만 결정적인 요인이 나이인지 아니면 합펜을 하는 장면인지 알기 어렵습니다. 그럼에도 불구하고 이러한 데이터 세트의 불균형과 그로 인한 분류기 편향 문제는 보다 철저한 조사가 필요하며, 이에 대해서는 섹션 5에서 다시 다룹니다.

3.6. 토론: 데이터 세트 편향 II

분류기가 데이터 세트 수집 파이프라인에 관여하기 때문에 또 다른 유형의 편향이 발생할 수 있습니다. 이러한 분류기로 인해 획득한 클립의 시각적 다양성이 감소하여 이러한 클립에 대해 학습된 액션 분류기에 편향이 발생할 수 있습니다. 좀 더 자세히 설명하면, 동영상은 제목(YouTube에 동영상을 업로드하는 사람이 제공)을 기준으로 선택되지만, 동영상 내에서 후보 클립의 위치는 위에서 설명한 대로 이미지(RGB) 분류기에 의해 제공됩니다. 실제로 이 시점에서 분류기를 사용한다고 해서 클립의 다양성이 제한되는 것은 아닙니다. 동영상은 액션에 관한 것이므로 클립의 일부로 선택된 특정 프레임이 중요하지 않을 수 있으며, 어떤 경우에도 클립에는 모양(RGB)과 동작이 상당히 달라질 수 있는 수백 개의 프레임이 더 포함되어 있습니다. 이러한 이유로 이미지 분류기의 중간 사용에 대해서는 그다지 우려하지 않습니다.

4. 벤치마크 성능

이 섹션에서는 먼저 비디오에서 사람의 동작 인식을 위한 세 가지 표준 ConvNet 아키텍처에 대해 간략하게 설명합니다. 그런 다음 이러한 아키텍처를 기준선으로 삼아 Kinetics 데이터 세트에 대한 훈련과 테스트를 통해 성능을 비교합니다. 또한 UCF-101과 HMDB-51에서의 성능도 포함합니다.

비디오 분류를 위한 세 가지 일반적인 접근 방식을 고려합니다: LSTM 이 위에 있는 ConvNet [7, 26], 2스트림 네트워크 [9, 19], 3D ConvNet [13, 21, 23]. 이러한 기본 아키텍처에 대한 많은 개선이 있었지만(예: [9]), 여기서 우리의 의도는 키네틱스에서 가장 좋은 아키텍처가 무엇인지에 대한 철저한 연구를 수행하는 것이 아니라 데이터 세트의 난이도 수준을 나타내는 것입니다. 우리가 비교하는 세 가지 유형의 아키텍처에 대한 대략적인 그래픽 개요는 그림 3에 나와 있으며, 각 아키텍처의 시간적 인터페이스 사양은 표 3에 나와 있습니다.

키네틱스 데이터 세트의 실험을 위해 세 가지 아키텍처 모두 키네틱스를 사용하여 처음부터 학습합니다. 방법

UCF-101 및 HMDB-51에 대한 실험의 경우 3D ConvNet을 제외한 아키텍처는 ImageNet에서 사전 학습되었습니다(이러한 데이터 세트는 처음부터 아키텍처를 학습하기에는 너무 작기 때문입니다).

4.1. ConvNet+LSTM

이미지 분류 네트워크의 높은 성능은 비디오에 대해 가능한 한 최소한의 변경으로 재사용하는 것을 매력적으로 만듭니다. 이는 각 프레임에서 독립적으로 특징을 추출한 다음 전체 비디오에 걸쳐 예측을 풀링하는 데 사용하여 달성할 수 있습니다[14]. 이는 백 오브 워드 이미지 모델링 접근법[16, 17, 24]의 정신과 유사하지만 실제로는 편리하지만 시간적 구조를 완전히 무시하는 문제가 있습니다(예: 모델이 문을 여는 것과 닫는 것을 구분하지 못할 수 있음).

이론적으로 더 만족스러운 접근 방식은 상태를 인코딩하고 시간적 순서와 장거리 종속성을 포착할 수 있는 LSTM과 같은 리커런트 계층을 모델에 추가하는 것입니다[7, 26]. 512개의 숨겨진 유닛이 있는 ResNet-50 모델[11]의 마지막 평균 풀링 레이어 뒤에 배치 노멀라이제이션(Cooijmans 등[6]이 제안한 방식)이 있는 LSTM 레이어를 배치합니다. 그런 다음 다방향 분류를 위해 LSTM의 출력 위에 완전히 연결된 계층을 추가합니다. 테스트 시 마지막 프레임의 모델 출력에서 분류를 수행합니다.

4.2. 투스트림 네트워크

ConvNet의 마지막 레이어에 있는 특징에 대한 LSTM은 높은 수준의 변화를 모델링할 수 있지만, 많은 경우에 중요한 미세한 로우 레벨 동작을 캡처하지 못할 수 있습니다. 또한 역전파 시간을 위해 여러 프레임에 걸쳐 네트워크를 풀어야 하므로 훈련 비용이 많이 듭니다. 시모니안과 지서만[19]이 소개한 다른 매우 실용적인 접근 방식은 단일 RGB 프레임과 외부에서 계산된 10개의 광학 흐름 프레임 스택의 예측을 평균한 후 이를 ImageNet으로 사전 학습된 ConvNet의 두 복제본에 통과시켜 비디오의 짧은 시간 스냅샷을 모델링하는 것입니다. 플로우 스트림에는 플로우 프레임보다 두 배 많은 입력 채널을 가진 적응형 입력 컨볼루션 레이어가 있으며(플로우에는 수평과 수직의 두 채널이 있기 때문), 테스트 시 비디오에서 여러 스냅샷을 샘플링하고 동작 예측을 평균화합니다. 이 방법은 기존 벤치마크에서 매우 높은 성능을 보였으며, 트레이닝 효율이 매우 높았습니다. 테스트.

4.3. 3D 컨버넷

3D 컨볼루션 네트워크[13, 21, 23]는 비디오 모델링에 대한 자연스러운 접근 방식처럼 보입니다. 표준 2D 컨볼루션 네트워크와 비슷하지만 시공간 필터가 있으며, 시공간 데이터의 계층적 표현을 직접 생성한다는 매우 흥미로운 특징이 있습니다. 이러한 모델의 한 가지 문제점은 더 많은 매개변수가 있다는 것입니다.

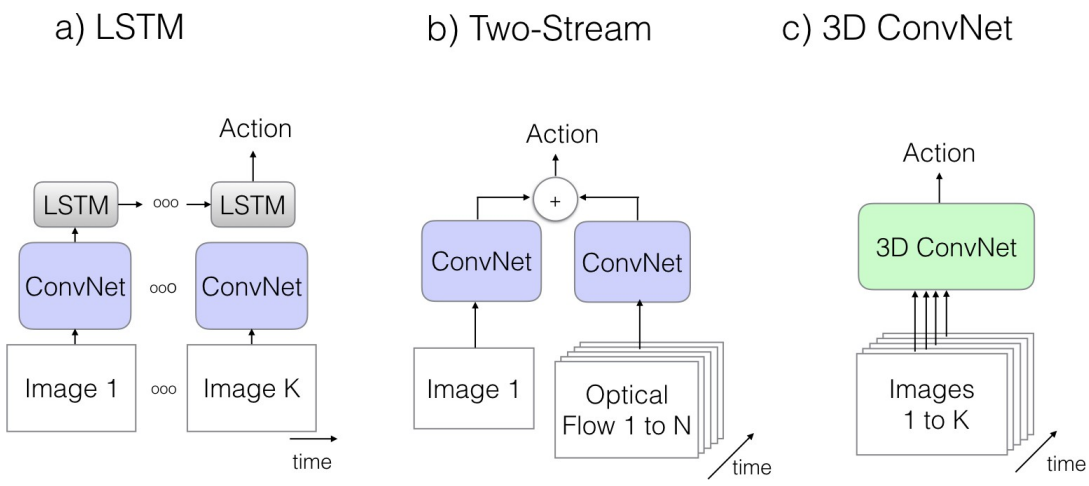


그림 3: 기준 인간 행동 분류자로 사용되는 비디오 아키텍처.

2D 컨브넷보다 커널 크기가 더 커서 훈련하기가 더 어렵습니다. 또한 이미지넷 사전 훈련의 이점을 누리지 못하는 것으로 보이며, 기존 연구에서는 상대적으로 얕은 맞춤형 아키텍처를 정의하고 처음부터 훈련했습니다[13, 14, 21, 23]. 벤치마크 결과는 가능성을 보여주었지만 2D보다 더 많은 훈련 데이터가 필요하기 때문에 아직 최신 기술을 따라잡지는 못했습니다. 따라서 3D ConvNet은 대규모 데이터 세트에 대한 평가를 위한 좋은 후보입니다.

이 논문에서는 8개의 컨볼루션 레이어, 5개의 풀링 레이어, 상단에 2개의 완전히 연결된 레이어가 있는 C3D[23]의 작은 변형을 구현했습니다. 모델에 대한 입력은 112×112 픽셀 크롭이 포함된 짧은 16프레임 클립입니다. 원본 논문과 다른 점은 모든 컨볼루션 및 완전히 연결된 레이어 후에 일괄 정규화를 사용한다는 것입니다. 원본 모델과의 또 다른 차이점은 첫 번째 풀링 레이어에서 1 대신 2의 시간적 보폭을 사용하여 메모리 사용량을 줄이고 더 큰 배치를 허용한다는 점인데, 이는 배치 정규화(특히 가중치 연결이 없는 완전 연결 레이어 이후)에 중요했습니다. 이 보폭을 사용하여 표준 K40 GPU를 사용하여 GPU당 배치당 15개의 비디오로 훈련할 수 있었습니다.

테스트 시에는 비디오를 16프레임으로 균일하게 분할하고 각각에 분류기를 개별적으로 적용합니다. 그런 다음 원본 논문에서와 같이 클래스 점수를 평균화합니다.

4.4. 구현 세부 정보

ConvNet+LSTM 및 Two-Stream 아키텍처는 ResNet-50을 기본 아키텍처로 사용합니다. Two-Stream 아키텍처의 경우 각 스트림에 대해 별도의 ResNet-50이 독립적으로 훈련됩니다. 앞서 언급한 바와 같이, 이러한 아키텍처의 경우 ResNet-50 모델은 UCF-101 및 HMDB-51 실험을 위해 Ima-geNet에서 사전 훈련되고, Kinetics 실험을 위해 처음부터 다시 훈련됩니다. 3D-ConvNet은 사전 훈련되지 않습니다.

모든 경우에 모멘텀이 있는 표준 SGD를 사용하여 비디오에 대한 모델을 훈련했으며, 모든 모델에 대해 64개의 GPU에서 동기 병렬화를 수행했습니다. 검증 손실이 포화 상태일 때 학습 속도를 10배 줄이면서 최대 10만 스텝까지 Kinetics로 모델을 훈련하고, Kinetics의 검증 세트에서 가중치 감쇠 및 학습 속도 하이퍼파라미터를 조정했습니다. 모든 모델은 텐서플로우[1]에서 구현되었습니다.

원본 클립의 해상도와 프레임 속도는 가변적입니다. 실험에서는 모두 정규화하여 더 큰 이미지 쪽이 ResNet-50을 사용하는 모델의 경우 340픽셀 너비, 3D ConvNet의 경우 128픽셀 너비가 되도록 했습니다. 또한 초당 25프레임이 되도록 비디오를 리샘플링했습니다.

데이터 증강은 딥 아키텍처의 성능에 매우 중요한 것으로 알려져 있습니다. 공간적으로 299×299 크기의 데이터를 무작위로 자르는 랜덤 크롭을 사용했습니다.

방법	#Params	교육		테스트	
		# 입력 프레임	시간적 발자국	# 입력 프레임	시간적 발자국
(a) ConvNet+LSTM	29M	25 rgb	5s	50 RGB	10s
(b) 투스트림	48M	1 RGB, 10 흐름	0.4s	25 RGB, 250 플로우	10s
(c) 3D-ConvNet	79M	16 RGB	0.64s	240 RGB	9.6s

표 3: 모델의 파라미터 수 및 시간적 입력 크기. ConvNet+LSTM 및 Two-Stream은 ResNet-50 ConvNet 모듈을 사용합니다
.

아키텍처	UCF-101			HMDB-51			키네틱스		
	RGB	흐름	RGB+흐름	RGB	흐름	RGB+흐름	RGB	흐름	RGB+흐름
(a) ConvNet+LSTM	84.3	-	-	43.9	-	-	57.0 / 79.0	-	-
(b) 투스트림	84.2	85.9	92.5	51.0	56.9	63.7	56.0 / 77.3	49.5 / 71.9	61.0 / 81.3
(c) 3D-ConvNet	51.6	-	-	24.3	-	-	56.1 / 79.5	-	-

표 4: 데이터 세트 전반의 기준선 비교: (왼쪽) UCF-101의 분할 1에 대한 훈련 및 테스트, (가운데) HMDB-51의 분할 1에 대한 훈련 및 테스트, (오른쪽) Kinetics에 대한 훈련 및 테스트(상위 1 / 상위 5 성능 표시). ConvNet+LSTM 및 Two-Stream은 UCF-101 및 HMDB-51 예제에서는 ImageNet에서 사전 훈련된 ResNet-50 ConvNet 모듈을 사용하지만 Kinetics 실험에서는 그렇지 않습니다. 개별 RGB 및 플로우 스트림의 투스트림 아키텍처 수치는 균일하게 샘플링된 25개의 프레임에 독립적으로 ConvNet을 적용한 다음 예측값의 평균을 구하는 간단한 기준선으로 해석할 수 있다는 점에 유의하세요.

패치(3D ConvNet의 경우 각각 112 × 112), 그리고 시간적으로는 원하는 프레임 수를 보장할 수 있을 만큼 충분히 빠른 프레임 중에서 시작 프레임을 선택할 때입니다. 더 짧은 비디오의 경우 각 모델의 입력 인터페이스를 만족시키기 위해 필요한 만큼 비디오를 반복했습니다. 또한 훈련 중에 각 동영상에 대해 무작위 좌우 뒤집기를 일관되게 적용했습니다.

테스트 시간에는 최대 10초 분량의 동영상을 샘플링하고 필요한 경우 다시 반복합니다. 테스트 시 좌우로 뒤집힌 비디오도 고려하고 훈련 중에 사진 메트릭과 같은 추가 보강을 추가하면 더 나은 성능을 얻을 수 있습니다. 이 부분은 향후 연구에 맡기겠습니다.

4.5. 기준 평가

이 섹션에서는 학습 및 테스트에 사용되는 데이터 세트를 다양하게 변경하면서 세 가지 기준 아키텍처의 성능을 비교합니다.

표 4는 UCF-101, HMDB-51 또는 Kinetics에서 훈련 및 테스트할 때의 분류 정확도를 보여줍니다. UCF-101과 HMDB-51의 분할 1과 Kinetics의 훈련/값 세트 및 홀드아웃 테스트 세트에 대해 훈련 및 테스트합니다.

몇 가지 주목할 만한 관찰 결과가 있습니다. 첫째, 두 데이터 세트의 난이도가 서로 다르다는 것을 나타내는 지표인 UCF-101보다 Kinetics의 성능이 훨씬 낮습니다. 반면에 HMDB-51의 성능은 Kinetics보다 더 떨어지는데, 이는 테스트 세트가 매우 어렵고, 훈련 데이터가 거의 없는 반면 외관 중심 방식이 어렵도록 설계된 것으로 보입니다. 파라미터가 풍부한 3D-ConvNet 모델은 ImageNet에서 사전 학습되지 않았습니다,

다른 기준선과 달리. 이는 모든 데이터 세트에서 성능 저하로 이어지지만, 특히 UCF-101과 HMDB-51의 경우 훨씬 더 큰 훈련 세트 덕분에 다른 모델의 성능에 훨씬 더 근접합니다.

- **클래스 난이도.** 그림 4에는 두 가지 스트림 모델에서 분류 정확도에 따라 정렬된 운동학 클래스의 전체 목록이 포함되어 있습니다. 먹는 수업은 핫도그, 감자칩, 도넛과 같이 무엇을 먹는지 구분해야 하는 경우가 있고, 영상에서는 이미 부분적으로 합쳐져 작게 보일 수 있기 때문에 가장 어려운 수업 중 하나입니다. 댄스 수업도 어렵고, '발 마사지' 또는 '머리 흔들기'와 같이 특정 신체 부위를 중심으로 하는 수업도 어렵습니다.

- **클래스 혼동**. 상위 10개의 클래스 혼동 사례는 표 5에 나와 있습니다. 대부분 '멀리뛰기'와 '세단뛰기', 햄버거와 도넛을 혼동하는 것과 같이 어렵다고 예상되는 세밀한 구분이 이에 해당합니다. '스윙 댄스'와 '살사 댄스'의 혼동은 '스윙 댄스'가 일반적으로 훨씬 빠른 속도로 진행되며 사람이 살사와 쉽게 구별할 수 있는 독특한 스타일을 가지고 있기 때문에 2스트림 모델에서 모션 모델링이 얼마나 정확한지에 대한 의문을 제기합니다.

- **모션이 가장 중요한 수업** 저희는 어떤 수업에서 모션이 더 중요하고

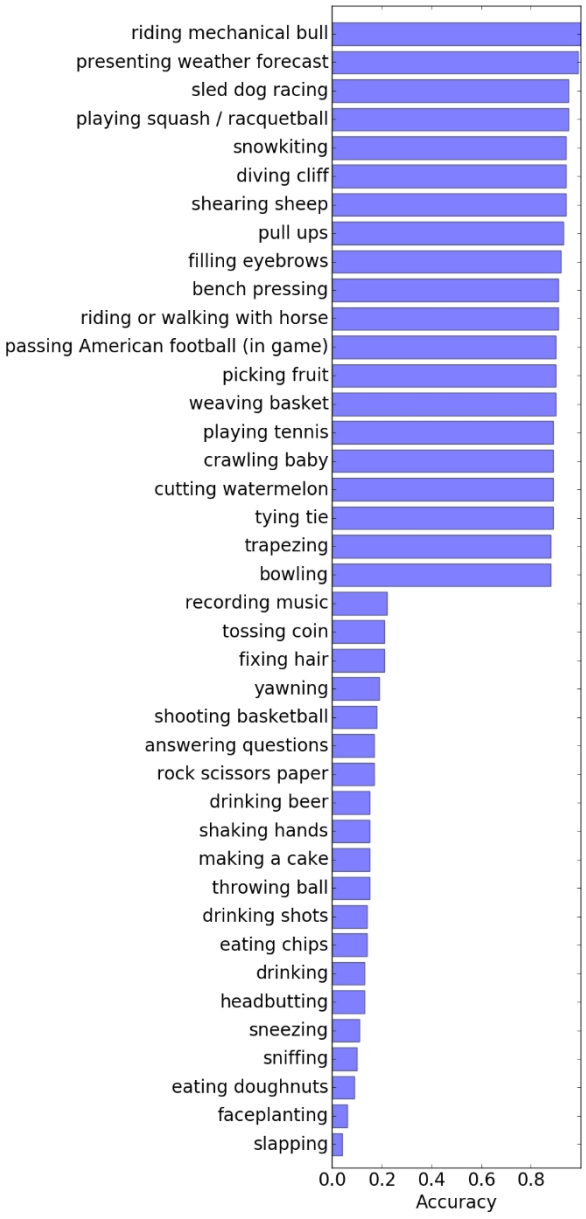


그림 4: 2스트림 모델을 사용하여 얻은 클래스 정확도를 기준으로 정렬된 가장 쉬운 20개 및 가장 어려운 20개의 키네틱스 클래스 목록.

두 스트림 모델의 흐름 스트림과 RGB 스트림을 따로 사용했을 때의 인식 정확도 비율을 비교하여 어떤 스트림이 색상 정보만 사용하여 올바르게 인식되었는지 확인합니다. 이 비율이 가장 크고 가장 작은 5개의 클래스를 표 6에 표시했습니다.

5. 결론

이전 유형의 데이터 세트보다 훨씬 더 많은 동영상이 포함된 키네틱스 휴먼 액션 비디오 데이터 세트에 대해 설명했습니다. 또한 데이터를 수집하고 품질을 보장하기 위해 사용한 절차에 대해서도 논의했습니다. 이 데이터 세트에 대한 표준 기존 모델의 성능은 UCF-101보다 훨씬 낮고 HMDB-51과 동등한 수준이며, 기존 인간 행동 데이터 세트와 달리 3D ConvNet과 같은 대규모 모델을 처음부터 학습할 수 있다는 것을 보여주었습니다.

또한 데이터 세트의 불균형과 이로 인해 데이터 세트에 대해 학습된 분류기가 편향성을 보이는지 여부에 대한 예비 분석도 수행했습니다. 결과 분류기가 성별과 같이 민감한 축을 따라 편향성을 보인다는 증거는 거의 발견되지 않았습니다. 그러나 이 부분은 더 많은 주의가 필요한 복잡한 영역입니다. 향후 작업을 위해 사회과학자 및 비판적 인문학자 등 다양한 분야의 전문가와 협력하여 철저한 분석을 남겨두고 있습니다.

예를 들어 새로운 액션 클래스의 기능을 생성하는 데 사용할 수 있도록 학습된 기준 모델(텐서플로)을 공개할 예정입니다.

감사:

이 데이터 세트의 수집은 딥마인드의 지원을 받았습니다. Andreas Kirsch, John- Paul Holt, Danielle Breen, Jonathan Fildes, James Besley, Brian Carver의 도움에 깊은 감사를 표합니다. Tom Duerig, Juan Carlos Niebles, Simon Osindero, Chuck Rosenberg, Sean Legassick의 조언과 의견에 감사드리며, 데이터 정리를 도와준 Sandra와 Aditya에게도 감사의 말씀을 전합니다.

참조

[1] M. 아바디, A. 아가왈, P. 바햄, E. 브레브도, Z. 첸, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin 외. Tensorflow: 이종 분산 시스템에 대한 대규모 기계 학습. *arXiv 사전 인쇄물 arXiv:1603.04467*, 2016.

[2] M. 안드릴루카, L. 피쉬출린, P. 겔러, B. 쉴레. 2d

사람 포즈 추정: 새로운 벤치마크와 최 첨

단 분석. *컴퓨터 비전 및 패턴 인식*

(CVPR), 2014 IEEE 컨퍼런스. IEEE,

2014.

[3] F. 카바 하일브론, V. 에스코르시아, B. 가넴, J. C. 니블스.
액티비티넷: 인간의 활동성 이해를 위한

대규모 비디오 벤치마크. *IEEE 컨퍼런스*

논문집 컴퓨터 비전 및 패턴 인식, 2015.

[4] A. 칼리스칸, J. J. 브라이슨, 및 A. 나라야난. 시맨틱 디
언어 말뭉치에서 자동으로 추출한 데이터

에는 인간 편견이 포함되어 있습니다. .

Science, 356(6334):183-186, 2017.

[5] J. 카레이라 및 A. 지서만. 쿼바디스, 액션 인식
새로운 모델과 동역학 데이터 세트. *컴퓨*

터 비전 및 패턴 인식에 관한 IEEE 국제

컨퍼런스 CVPR( ), 2017.

[6] T. Cooijmans, N. Ballas, C. Laurent, 및 A. Courville.

클래스 1	클래스 2	혼란
'라이딩 뮬'	'말 타기 또는 말과 함께 걷기'	40%
'하키 스톱'	'아이스 스케이팅'	36%
'스윙 댄스'	'살사 댄스'	36%
'기타 스트럼밍'	'기타 연주'	35%
'슈팅 농구'	'농구'	32%
'소시지 요리'	'치킨 요리'	29%
'바닥 청소'	'걸레질 바닥'	27%
'트리플 점프'	'멀리뛰기'	26%
'에어로빅을 하다'	'zumba'	26%
'애완 동물(고양이 아님)'	'염소 먹이주기'	25%
'면도 다리'	'왁싱 다리'	25%
'스노우보드'	'스키(활강 또는 크로스컨트리 제외)'	22%

표 5: 두 스트림 모델을 사용한 Kinetics의 상위 12개 클래스 혼동.

클래스	흐름/RGB 정확도 비율
'가위 바위 보'	5.3
'검투'	3.1
'로봇 춤'	3.1
'에어 드럼'	2.8
'운동하는 팔'	2.5
'케이크 만들기'	0.1
'소시지 요리'	0.1
'스니핑'	0.1
'케이크 먹기'	0.0
'샌드위치 만들기'	0.0

표 6: 흐름과 RGB를 사용할 때 인식 정확도의 최대 및 최소 비율을 가진 클래스. 가장 큰 비율은 흐름이 더 나은 경우, 가장 작은 비율은 RGB가 더 나은 경우에 해당합니다. 또한 RGB+흐름과 RGB의 정확도 비율을 평가한 결과, 그 순서는 매우 비슷했습니다.

반복 배치 정규화. *arXiv preprint arXiv:1603.09025*, 2016.

[7] J. 도나휴, L . 앤 헨드릭스, S . 구아다라마, M. Rohrbach, S. Venugopalan, K. Saenko, 및 T. Dar- rell. 시각적 인식 및 설명을 위한 장기 반복 컨볼루션 네트워크. *IEEE 컴퓨터 비전 및 패턴 인식 컨퍼런스 논문집*, , 2625-2634페이지, 2015.

[8] M. 에버링햄, S. A. 에슬라미, L. 반굴, C. K. 윌리엄스, J. 윈, 및 A. 지서만. 파스칼 시각 객체 클래스의 도전: 회고. *국제 컴퓨터 비전 저널- puter Vision*, 111(1):98-136, 2015.

[9] C. Feichtenhofer, A. Pinz, 및 A. Zisserman. 컨볼루션 비디오 동작 인식을 위한 2스트림 네트워크 융합. *IEEE 국제 컴퓨터 비전 및 패치 컨퍼런스(tern Recognition CVPR*, 2016)에서.

[10] G. 그리핀, A. 홀럽, 및 P. 페로나. 칼텍-256 개체 고양이-에고리 데이터 세트. 2007.

[11] K. He, X. Zhang, S. Ren, and J. Sun. 이미지 인식을 위한 심층 잔여 학습. *컴퓨터 비전 및 패턴 인식(CVPR), 2016 IEEE 컨퍼런스*, 2016.

[12] S. 이오페와 C. 세게디. 일괄 정규화: 가속 내부 공변량 이동을 줄여 딥 네트워크 학습을 강화합니다. *arXiv 사전 인쇄물 arXiv:1502.03167*, 2015.

[13] S. Ji, W. Xu, M. Yang, and K. Yu. 인간 행동 인식을 위한 3D 컨볼루션 신경망. *패턴 분석 및 기계 지능에 관한 IEEE 트랜잭션*, 35(1):221- 231, 2013.

[14] A. 카르파시, G. 토데리치, S. 셰티, T. 렁, R. 수크탄카르, 그리고 L. 페이페이. 컨볼 루션 신경망을 사용한 대규모 비디오 분류. *컴퓨터 비전 및 패턴 인식에 관한 IEEE 학술대회 논문집*, 1725-1732, 2014.

[15] H. Kuehne, H. Jhuang, E. Garrote, T. Poggio, 및 T. Serre. HMDB: 사람의 동작 인식을 위한 대규모 비디오 데이터베이스. *국제 컨퍼런스 논문집 컴퓨터 비전(ICCV)*, 2011.

[16] I. Laptev, M. Marszalek, C. Schmid, 및 B. Rozenfeld. 영화에서 사실적인 인간 행동 학습하기. In *컴퓨터 비전 및 패턴 인식, 2008. CVPR 2008. IEEE 컨퍼런스*, 1-8 페이지. IEEE, 2008.

[17] J. C. 니블스, H. 왕, 및 L. 페이페이. 비지도 학습 공간-시간적 단어를 사용한 인간 행동 범주 분류. *국제 컴퓨터 비전 저널*, 79(3):299-318, 2008.

[18] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, S. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. Berg, and F. Li. Imagenet 대규모 시각 인식 도전. *IJCV*, 2015.

[19] K. 시모니안과 A. 지서만. 2스트림 컨볼루션 비디오에서 동작 인식을 위한 네트워크. *신경 정보 처리 시스템의 발전*, 568-576페이지, 2014.

[20] K. Soomro, A. R. Zamir, 및 M. Shah. Ucf101: 데이터 세트 야생의 비디오에서 101개의 인간 행동 클래스를 추출했습니다. *arXiv 사전 인쇄 arXiv:1212.0402*, 2012.

[21] G. W. Taylor, R. Fergus, Y. LeCun, 및 C. Bregler. Convolu-

시공간적 특징을 학습합니다. *유럽에서는*

컴퓨터 비전에 대한 페렌스, 140-153 페이지. 스프링거, 2010.

[22] A. 토랄바 및 A. A. 에프로스. 데이터 세트 편향에 대한 편견 없는 시각. *컴퓨터 비전 및 패턴 인식(CVPR), 2011 IEEE 컨퍼런스*, 1521-1528 페이지. IEEE, 2011.

[23] D. Tran, L. Bourdev, R. Fergus, L. Torresani, 및 M. Paluri. 3D 컨볼루션 네트워크를 사용한 시공간적 특징 학습. *2015 IEEE 컴퓨터 국제 컨퍼런스 비전(ICCV)*, 4489-4497페이지. IEEE, 2015.

[24] H. 왕과 C. 슈미드. 향상된 동작 인식 궤적. *국제 컴퓨터 비전 컨퍼런스* , 2013.

[25] X. Wang, A. Farhadi, 및 A. Gupta. 액션 트랜스포메이션 tions. *CVPR*, 2016.

[26] J. Yue-Hei Ng, M. Hausknecht, S. Vijayanarasimhan, O. 빈얄스, R. 몽가, G. 토데리치. 짧은 저격-펫을 넘어서: 비디오 분류를 위한 딥 네트워크. *IEEE 컴퓨터 비전 및 패턴 인식 컨퍼런스 논문집*, 4694-4702페이지, 2015.

A. 키네틱스 휴먼 액션 클래스 목록

다음은 휴먼 액션 비디오 데이터 세트에 포함된 클래스 목록입니다. 각 액션 클래스의 클립 수는 각 클래스 이름 뒤에 괄호 안에 있는 숫자로 표시됩니다.

1. 암벽 등반 (1146)
2. 에어 드럼 (1132)
3. 질문에 답변하기 (478)
4. 박수 (411)
5. 크림 바르기 (478)
6. 양궁 (1147)
7. 팔씨름 (1123)
8. 꽃꽂이 (583)
9. 컴퓨터 조립 (542)
10. 경매 (478)
11. 아기 깨우기 (611)
12. 베이킹 쿠키 (927)
13. 풍선 불기 (826)
14. 붕대 (569)
15. 바베큐 (1070)
16. 바텐딩 (601)
17. 비트박스 (943)
18. 양봉 (430)
19. 벨리댄스 (1115)
20. 벤치 프레싱 (1106)
21. 뒤로 굽히기 (635)
22. 굽힘 금속 (410)
23. 눈 사이로 자전거 타기 (1052)
24. 발파 모래 (713)
25. 블로잉 글라스 (1145)

26. 나뭇잎 불기 (405)
27. 코 풀기 (597)
28. 촛불 끄기 (1150)
29. 봅슬레이 (605)
30. 제본 (914)
31. 트램펄린에서 튕기기 (690)
32. 볼링 (1079)
33. 머리 땋기 (780)
34. 빵가루 또는 빵가루 (454)
35. 브레이크댄스 (948)
36. 브러시 페인팅 (532)

37. 머리 빗기 (934)
38. 양치질 (1149)
39. 건물 캐비닛 (431)
40. 건물 창고 (427)
41. 번지점프 (1056)
42. 버스킹 (851)
43. 카누 또는 카약 (1146)
44. 카포에이라 (1092)
45. 아기 안고 (558)
46. 카트 휠링 (616)
47. 조각 호박 (711)
48. 물고기 잡기 (671)
49. 야구공 잡기 또는 던지기 (756)
50. 원반 잡기 또는 던지기 (1060)
51. 소프트볼 잡기 또는 던지기 (842)
52. 축하 (751)
53. 오일 교환 (714)
54. 휠 교체 (459)
55. 타이어 점검 (555)
56. 치어리딩 (1145)
57. 나무 자르기 (916)
58. 박수 (491)
59. 점토 도자기 만들기 (513)
60. 클린 앤 저크 (902)
61. 바닥 청소 (874)
62. 홈통 청소 (598)
63. 수영장 청소 (447)
64. 청소용 신발 (706)
65. 화장실 청소 (576)
66. 창문 청소 (695)
67. 로프 등반 (413)
68. 등산 사다리 (662)
69. 등반 나무 (1120)
70. 접촉 저글링 (1135)
71. 치킨 요리 (1000)
72. 계란 요리 (618)
73. 캠프파이어 요리 (403)
74. 소시지 요리 (467)
75. 돈 세기 (674)
76. 컨트리 라인 댄스 (1015)

77. 갈라지는 목 (449)
78. 크롤링 베이비 (1150)
79. 강 건너기 (951)
80. 울음 (1037)
81. 컬링 헤어 (855)
82. 손톱 자르기 (560)
83. 파인애플 자르기 (712)
84. 수박 자르기 (767)
85. 댄싱 발레 (1144)
86. 댄싱 찰스턴 (721)
87. 댄싱 강남 스타일 (836)
88. 춤추는 마카레나 (958)
89. 데드리프트 (805)
90. 크리스마스 트리 장식하기 (612)
91. 파기 (404)
92. 식사 (671)
93. 디스크 골프 (565)
94. 다이빙 절벽 (1075)
95. 피구 (595)
96. 에어로빅하기 (461)
97. 빨래하기 (461)
98. 네일 (949)
99. 그리기 (445)
100. 드리블 농구 (923)
101. 음주 (599)
102. 맥주 마시기 (575)
103. 드링킹 샷 (403)
104. 자동차 운전 (1118)
105. 운전 트랙터 (922)
106. 낙하 발차기 (716)
107. 드럼 연주 손가락 (409)
108. 덩크슛 농구 (1105)
109. 염색 머리 (1072)
110. 햄버거 먹기 (864)
111. 케이크 먹기 (494)
112. 당근 먹기 (516)
113. 칩 먹기 (749)
114. 도넛 먹기 (528)
115. 핫도그 먹기 (570)
116. 아이스크림 먹기 (927)

117. 스파게티 먹기 (1145)
118. 수박 먹기 (550)
119. 달걀 사냥 (500)
120. 팔 운동 (416)
121. 운동 공으로 운동하기 (438)
122. 화재 진압 (602)
123. 안면 성형 (441)
124. 새 먹이 주기 (1150)
125. 물고기 먹이주기 (973)
126. 염소 먹이주기 (1027)
127. 눈썹 채우기 (1085)
128. 손가락 스냅 (825)
129. 머리 고정 (676)
130. 뒤집기 팬케이크 (720)
131. 연 날리기 (1063)
132. 옷 접기 (695)
133. 접이식 냅킨 (874)
134. 접이식 종이 (940)
135. 프론트 레이즈 (962)
136. 야채 튀김 (608)
137. 쓰레기 수거 (441)
138. 양치질 (430)
139. 이발하기 (658)
140. 문신 받기 (737)
141. 상을 주거나 받기 (953)
142. 골프 치핑 (699)
143. 골프 운전 (836)
144. 골프 퍼팅 (1081)
145. 고기 갈기 (415)
146. 미용견 (613)
147. 그루밍 말 (645)
157. 돌 차기 놀이 (726)
158. 호버보드 (564)
159. 포옹 (517)
160. 훌라후프 (1129)
161. 허들 (622)
162. 투포환 (스포츠) (836)
163. 빙벽 등반 (845)
164. 얼음 낚시 (555)
165. 아이스 스케이팅 (1140)
166. 다림질 (535)
167. 창 던지기 (912)
168. 제트스키 (1140)
169. 조깅 (417)
170. 저글링 볼 (923)
171. 저글링 불 (668)
172. 저글링 축구공 (484)
173. 수영장에 뛰어들기 (1133)
174. 점프 스타일 댄스 (662)
175. 필드 골 차기 (833)
176. 축구 공 차기 (544)
177. 키스 (733)
178. 카이트서핑 (794)
179. 뜨개질 (691)
180. 크럼핑 (657)
181. 웃음 (926)
182. 벽돌 쌓기 (432)
183. 멀리뛰기 (831)
184. 런지 (759)
185. 케이크 만들기 (463)
186. 샌드위치 만들기 (440)
187. 침대 만들기 (679)

148. 체조 텀블링 (1143)
149. 해머 던지기 (1148)
150. 헤드뱅잉 (1090)
151. 헤드버팅 (640)
152. 높이뛰기 (954)
153. 하이킥 (825)
154. 타격 야구 (1071)
155. 하키 정류장 (468)
156. 뱀 잡기 (430)
197. 마사지하는 사람 (672)
198. 착유 젖소 (980)
199. 걸레질 바닥 (606)
200. 모터사이클 (1142)
201. 가구 이사 (426)
202. 잔디 깎기 (1147)
203. 뉴스 앵커 (420)
204. 병따개 (732)
205. 개봉 선물 (866)
206. 패러글라이딩 (800)
207. 패러세일링 (762)
208. 파쿠르 (504)
209. 미식축구 패스 (게임 내) (863)
210. 미식축구 패스(경기 중이 아님) (1045)
211. 사과 껍질 벗기기 (592)
212. 감자 껍질 벗기기 (457)
213. 애완용 동물(고양이 제외) (757)
214. 애완용 고양이 (756)
215. 과일 따기 (793)
216. 나무 심기 (557)
217. 미장 (428)
218. 아코디언 연주 (925)
188. 보석 만들기 (658)
189. 피자 만들기 (1147)
190. 눈사람 만들기 (756)
191. 초밥 만들기 (434)
192. 차 만들기 (426)
193. 행진 (1146)
194. 등 마사지 (1113)
195. 발 마사지 (478)
196. 다리 마사지 (592)
237. 키보드 연주 (715)
238. 킥볼 놀이 (468)
239. 독점 놀이 (731)
240. 오르간 연주 (672)
241. 페인트볼 놀이 (1140)
242. 피아노 연주 (691)
243. 포커 플레이 (1134)
244. 레코더 재생 (1148)
245. 색소폰 연주 (916)
246. 스쿼시 또는 라켓볼 경기 (980)
247. 테니스 (1144)
248. 트롬본 연주 (1149)
249. 트럼펫 연주 (989)
250. 우쿨렐레 연주 (1146)
251. 바이올린 연주 (1142)
252. 배구 경기 (804)
253. 실로폰 연주 (746)
254. 장대높이뛰기 (984)
255. 일기 예보 발표(1050)
256. 턱걸이 (1121)
257. 펌핑 주먹 (1009)
258. 가스 펌핑 (544)

219. 배드민턴 경기 (944)
220. 백파이프 연주 (838)
221. 농구하기 (1144)
222. 베이스 기타 연주 (1135)
223. 카드 놀이 (737)
224. 첼로 연주 (1081)
225. 체스 플레이(850)
226. 클라리넷 연주 (1022)
227. 컨트롤러 재생 (524)
228. 크리켓 경기 (949)
229. 심벌즈 연주 (636)
230. 디저리두 연주 (787)
231. 드럼 연주 (908)
232. 플루트 연주 (475)
233. 기타 연주 (1135)
234. 하모니카 연주 (1006)
235. 하프 연주 (1149)
236. 아이스하키 경기 (917)
277. 종이 찢기 (605)
278. 로봇 댄스 (893)
279. 암벽 등반 (1144)
280. 가위 바위 보 (424)
281. 롤러 스케이트 (960)
러닝머신에서 달리기 (428)
283. 항해 (867)
284. 살사 댄스 (1148)
285. 샌딩 바닥 (574)
286. 스크램블 에그 (816)
287. 스쿠버 다이빙 (968)
288. 상차림 (478)
289. 악수 (640)
259. 샌드백 (1150)
260. 펀치하는 사람 (복싱) (483)
261. 밀어 올리기 (614)
262. 자동차 밀기 (1069)
263. 푸시 카트 (1150)
264. 휠체어 밀기 (465)
265. 책 읽기 (1148)
266. 신문 읽기 (424)
267. 음악 녹음 (415)
268. 자전거 타기 (476)
269. 낙타 타기 (716)
270. 코끼리 타기 (1104)
271. 기계식 황소 타기 (698)
272. 산악 자전거 타기 (495)
273. 노새 타기 (476)
274. 승마 또는 말과 함께 걷기 (1131)
275. 라이딩 스쿠터 (674)
276. 외발자전거 타기 (864)
317. 흡연 (1105)
318. 흡연 물 담뱃대 (857)
319. 스내치 웨이트 리프팅 (943)
320. 재채기 (505)
321. 스니핑 (399)
322. 스노클링 (1012)
323. 스노보드 (937)
324. 스노우카이팅 (1145)
325. 스노모빌 (601)
326. 공중제비 (993)
327. 스피닝 포이 (1134)
328. 스프레이 페인팅 (908)
329. 스프레이 (470)

290. 머리 흔들기 (885)
291. 칼 갈기 (424)
292. 연필 깎기 (752)
293. 면도 머리 (971)
294. 다리 면도 (509)
295. 양털 깎기 (988)
296. 구두 닦기 (615)
297. 슈팅 농구 (595)
298. 슈팅 골 (축구) (444)
299. 포환던지기 (987)
300. 눈 삽질 (879)
301. 파쇄 용지 (403)
302. 카드 셔플 (828)
303. 옆차기 (991)
304. 수화 통역 (446)
305. 노래 (1147)
306. 윗몸 일으키기 (817)
307. 스케이트보드 (1139)
308. 스키 점프 (1051)
309. 스키(슬라롬 또는 크로스컨트리 제외) (1140)
310. 스키 크로스컨트리 (477)
311. 스키 슬라럼 (539)
312. 줄넘기 (488)
313. 스카이다이빙 (505)
314. 슬랙라이닝 (790)
315. 때리기 (465)
316. 썰매 개 경주 (775)
330. 스프링보드 다이빙 (406)
331. 스쿼트 (1148)
332. 혀 내밀기 (770)
333. 쿵쾅거리는 포도 (444)
334. 스트레칭 팔 (718)
335. 다리 스트레칭 (829)
336. 기타 스트럼밍 (472)
337. 서핑 군중 (876)
338. 서핑 워터 (751)
339. 스위핑 플로어 (604)
340. 수영 배영 (1077)
341. 수영 평영 (833)
342. 수영 접영 (678)
343. 스윙 댄스 (512)
344. 스윙 다리 (409)
345. 무언가에 스윙 (482)
346. 검술 (473)
347. 태극권 (1070)
348. 샤워하기 (378)
349. 탱고 댄스 (1114)
350. 탭댄스 (947)
351. 태핑 기타 (815)
352. 탭핑 펜 (703)
353. 맥주 시음 (588)
354. 음식 시식 (613)
355. 증언 (497)
356. 문자 메시지 (704)

357. 던지는 도끼 (816)
358. 공 던지기 (634)
359. 원반 던지기 (1104)
360. 간지럼 (610)
361. 터보건 (1147)
362. 동전 던지기 (461)
363. 토싱 샐러드 (463)
364. 훈련견 (481)
365. 사다리타기 (786)
366. 수염 다듬기 또는 면도 (981)
367. 나무 다듬기 (665)
368. 트리플 점프 (784)
369. 나비 넥타이 묶기 (387)
370. 매듭 묶기 (넥타이가 아닌) (844)
371. 넥타이 매기 (673)
372. 개봉 (858)
373. 하역 트럭 (406)
374. 컴퓨터 사용 (937)
375. 리모컨 사용(게임이 아닌 경우) (549)
376. 세그웨이 사용 (387)
377. 금고 (562)
378. 줄서기 대기 (430)
379. 개 산책시키기 (1145)
380. 설거지 (1048)
381. 발 세척 (862)
382. 머리 감기 (423)
383. 손 씻기 (916)
384. 수상 스키 (763)
385. 워터 슬라이딩 (420)
386. 식물에 물주기 (680)
387. 왁싱 백 (537)
388. 왁싱 가슴 (760)
389. 눈썹 왁싱 (720)
390. 다리 왁싱 (948)
391. 직조 바구니 (743)
397. 글쓰기 (735)
398. 하품 (398)
399. 요가 (1140)
400. 줌바 (1093)

B. 부모-자식 그룹 목록

이 목록은 배타적인 것이 아니며 포괄적인 것이 아닙니다. 오히려 관련 휴먼 액션 클래스를 위한 가이드입니다.

꽃꽂이
예술 및 공예 (12)
호박 조각 유리
브러시 페인팅
불기
점토 도자기 만들기
크리스마스 트리 그림 꾸미기
문신 받기
뜨개질
보석 스프레이
페인팅 직조 바
구니 만들기

육상 - 점프 (6)
높이뛰기
허들 멀리
뛰기 파쿠
르 장대높
이뛰기 세
단뛰기

육상 - 던지기 + 달리기 (9)
양궁
원반 잡기 또는 던지기 디스크
골프
해머 던지기 창
던지기
포환던
지기

도끼 던지기 공

던지기 원반 던

지기

자동 유지 관리 (4)

오일 교환 휠 교

체

타이어 점검

가스 펌핑

볼링

볼 스포츠 (25)

야구공 잡기 또는 던지기

소프트볼 피구 잡기 또는 던지기
드리블 농구 덩크슛
농구 골프 치핑
골프 운전 골프
퍼팅 타격 야구
투구 (스포츠)
저글링 축구 공 차
기 필드 골 차기 축
구 공 차기
미식축구 패스 (게임 중) 미식축구 패스 (
게임 중이 아님) 농구하기
크리켓 경기 킥볼 경
기
스쿼시 또는 라켓볼 테니스 경기
배구 슈팅 농구 슈팅 골 (
축구) 포환던지기 경기

신체 동작 (16) 에어
드럼 박수치기
등을 구부리고
박수 치며 목이
갈라지는 아기
깨우기
드럼 치는 손가락 손
가락 꺾기 헤드뱅잉
헤드버팅 펌핑 주먹
쥐기 머리 흔들기
팔 뻗기 다리 흔들
기 다리 흔들기

청소 (13) 바닥 청
소 배수구 청소 수
영장 청소 신발 청
소 화장실 청소 창
문 청소 빨래하기
침대 걸레질 바닥

닦기 테이블 세팅 구두 닦기

바닥을 쓸고 설거지하기

옷감 (8)

붕대 감기

세탁하기

옷 접기

냅킨 다림

질하기

침대 묶는 나비

넥타이 만들기

매듭 묶기(넥타이가 아닌)

넥타이 묶기

커뮤니케이션 (11)

질문에 답하기 경매

바텐딩 축하하기

울음

상을 주거나 받는 웃음

뉴스 앵커

일기 예보 수화 통역 증언

발표

요리 (22) 베이

킹 쿠키 바비큐

빵가루 또는 빵가루 요리 닭

고기

모닥불에 계란 요리

하기 소시지 요리하

기 파인애플 자르기

수박 자르기 팬케이

크 뒤집기 야채 튀기

기 고기 갈기 케이크

만들기 샌드위치 만

들기 피자 만들기 초

밥 만들기 차 만들기

사과 껍질 벗기

기 감자 껍질 벗

기기 과일 따기

스크램블 에그

던지기 샐러드

던지기

춤 (18)

밸리 댄스

브레이크 댄스 카포
에라 치어리딩 컨트
리 라인 댄스 댄스
발레 댄스 찰스턴
춤 강남 스타일 춤 마카
레나 점프 스타일 춤 크
럼핑
마칭 로봇 댄
스 살사 댄스
스윙 댄스 탱
고 댄스 탭 댄
스 줌바 댄스

먹기 + **마시기 (17)**
바텐딩 식사
맥주 마시기
샷 마시기 햄
버거 먹기 케
이크 먹기 당
근 먹기 칩 먹
기
도넛 먹기 핫도그
먹기 아이스크림
먹기 스파게티 먹
기 수박 따기 병따
기 맥주 시음하기
음식 시식하기

전자 (5) 조립 컴퓨터 재
생 컨트롤러 문자 메시
지
컴퓨터 사용
리모컨 사용(게임이 아닌 경우)

정원 + **식물 (10)** 나뭇
잎을 불고 호박 조각
나무를 자르는 나무
등반 나무
크리스마스 트리 달걀 사냥 꾸미기
잔디 깎기 나무 심기

식
물
에

물
을

주
는

나
무

다
듬
기

골프 (3)
골
프

치
핑

골
프

드
라
이
빙

골
프

퍼
팅

체
조

(5) 트램펄린 카트 휠링
체조 텀블링 공중제비
뛰기
볼트

머리 (14) 땋기
머리 빗기 머리
컬링 머리 염색
머리 고정 머리
자르기 머리 깎
기 머리 깎기 다
리 깎기
수염 다듬기 또는 면도 머리
감기
왁싱 등 왁싱 가슴
왁싱 눈썹 왁싱 다
리 왁싱

손 (9)
박수 치는 에
어 드럼 박수
치는 손톱 자
르기 손톱 자
르기
드럼 치는 손가락
손가락 스냅 펌핑
주먹으로 손 씻기

머리 + 입 (17) 풍
선 불기 비트박스
불기 코 불기
촛불 끄기 양치질하
기 머리 부딪히기 머
리 흔들기 노래 부
르기

흡연 흡연 물 담
뱃대 재채기 냄
새 맡기
혀를 내밀고 휘파람
불기
하품

높이 (15) 번지
점프 하강 로프
등반 사다리 오
르기 나무 다이
빙 절벽 등반
수영장 패러글라이
딩으로 점프하는 빙
벽 등반
암벽 등반 스카이다
이빙 슬랙라이딩 스
프링보드 다이빙
공중그네를 타는 것

동물과의 상호작용 (19)
꿀벌 물고기 잡
기 물고기 먹이
주기 새 물고기
먹이주기 염소
손질하기 개 손
질하기 뱀을 안
고 있는 말 얼음
낚시 착유하는
젖소
애완용 동물(고양이 아님) 애
완용 고양이
낙타 타기 코끼
리 타기 노새
타기
양털 깎는 말과 함께 승마 또는
걷기
개 산책 훈련

저글링 (6) 접촉 저글링 훌라후프 저글링 볼 저글링 불 저글링
축구공 회전 포이 저글링

메이크업 (5)

크림 바르기

손톱 염색하기

머리 염색하기

눈썹 채우기

문신하기

무술 (10) 팔 레슬링 카포에이라

드롭킥 하이킥

샌드백 펀칭 백

펀칭 사람 옆차기

검술 태극권

레슬링

기타 (9) 파기 소화 쓰레기 수거 벽돌 쌓기 가구 옮기기 스프레이 뿌리기 포도 밟기 펜두드리기 트럭 하역하기

이동성 - 육상 (20)

크롤링 베이비 드라이빙 카

트랙터 운전

페이스 플랜팅

호버보드 조깅

오토바이 파쿠르 자동차 밀기 카트 밀기

휠체어 밀기 자전거 타기

산악 자전거 라이딩

스쿠터

외발자전거 롤러 스케이팅 러닝머신에서 달리기 스케이트보드 서핑 군중 타기

줄
을

서
서

기
다
리
는

세
그
웨
이

이
용

세기 접는 냅킨 접는 종이 열기 선물 독서 책 읽기 신문 찢는 종이

이동성 - 물 (10)

강 건너기 다이빙

절벽 건너기 수영

장으로 뛰어들기

스쿠버 다이빙 스

노클링 스프링보드

다이빙

수영 배영 수영 평영 수영

접영 수영 수중 슬라이딩

음악 (29)

비트박스

버스킹

아코디언 연주 백

파이프 연주 베이

스 기타 연주 첼로

연주 클라리넷 연

주 심벌즈 연주 디

저리두 연주 드럼

연주 플루트 연주

기타 연주 하모니

카 연주 하프 연주

키보드 연주 오르

간 연주 피아노 연

주 리코더 연주 색

소폰 연주 트럼펫

연주 우쿨렐레 연

주 바이올린 연주

실로폰 연주 음악

녹음 노래 부르기

기타 연주 기타

두드리기 기타 휘

파람 불기

종이 (12) 제본 돈

파
쇄
지

개
봉

포
장

선
물

쓰
기

개인 위생 (6)
양
치
질

샤
워
하
기
수
염

다
듬
기

또
는

면
도

발

씻기
머리 씻기 손
씻기

게임하기 (13)
달걀 사냥 연
날리기 돌 차
기 놀이 카드
놀이 체스 놀
이
모노폴리 플레이하기
페인트볼 플레이하기
포커 플레이하기
기계식 황소 가위 바
위 보 종이 셔플 카드
줄넘기 타기
동전 던지기

라켓 + **배트 스포츠 (8)** 야구 잡
기 또는 던지기 소프트볼 타격 야
구 잡기 또는 던지기
던지기 (스포츠)
배드민턴 경기 크
리켓 경기
스쿼시 또는 라켓볼 테니스 경기

눈 + **얼음 (18)** 스노
우 봅슬레이를 통한
자전거 타기
하키 스톱
아이스 클라
이밍 아이스
낚시 아이스
스케이팅
눈사람 만들기 아
이스하키 삽질 스
노우 스키 점프하

기
스
키
(
활
강

또
는

크
로
스
컨
트
리

제
외
)

크
로
스
컨
트
리

스
키
스
키

슬
라
럼

썰
매

개 경주

스노보드 스노

우킥보드 스노

모빌 터보건

수영 (3) 수영 배영 수영

평영 수영
수영 접영 스트로크

아기를 안고 포옹하는

사람(11) 만지기

키스 마사지 등

마사지 발 마사

지 다리 마사지

악수하는 사람의 머리 마사지

간지럼

때리기

도구 사용 (13) 구

부리기 금속 발파

모래 건축 캐비닛

건축 창고 오일 교

환 휠 교체 타이어

점검 미장 펌프 가

스 샌딩 바닥 샌딩

칼 갈기 연필 연필

용접

수상 스포츠 (8) 카누 또

는 카약 제트스키

카이트서핑

패러세일링

세일링 서핑

수상스키 윈

드서핑

왁싱 (4) 등 왁싱

가슴 왁싱 눈썹 왁

싱 다리 왁싱