**□ 모듈프로젝트 산출물**

1. 클라우드 AI 융합전문가 양성과정

|  |
| --- |
| 모듈프로젝트 4  8조 : 박준원, 김원일, 김택완, 천성신, 최숙희, 허현준 |

**2021년 11월 09일**

**클라우드 AI 융합 전문가**

이름 : 최숙희

1. **모듈프로젝트 내용**

**- 제시된 모델 외의 모델을 적용해서 성능을 향상시킴**

**- 사용한 각 모델들의 성능을 비교**

**- 새로운 처리방식을 사용하여 성능을 비교**

|  |
| --- |
| **목 차**  **01. 모델 설명 및 성능평가**  **02. 실패 요인 분석**  **03. 프로젝트 소감문** |

**01. 모델 설명 및 성능평가**

|  |
| --- |
| **LeNet 설명**  <LeNet-5 Architecture>    Convolution과 Subsampling을 반복적으로 거치면서, 마지막에 Fully-connceted Multi-layered Network로 Classification 수행. |
| **LeNet 성능평가**  https://cdn.discordapp.com/attachments/902902431816499213/902961751425966171/unknown.png  https://cdn.discordapp.com/attachments/902902431816499213/902961451667435570/unknown.png  간단한 CNN으로 모델을 생성해 특정 성능이 나오지 확인과정을 거침.  정확도 0.72로 저조한 성적을 보임.  결과를 바탕으로 이후 다양한 모델을 이용하여 성능 향상을 도모함. |

|  |
| --- |
| **VGGNet 설명**  <VGG Architecture>    기존 딥러닝 알고리즘은 큰 필터와 복잡한 알고리즘을 사용해 성능을 올렸음.  VGG는 필터 자체의 사이즈를 단순한 3 x 3필터를 사용하고 깊이를 늘려 성능을 확보하는 구조를 가짐.  이전의 알고리즘들에 비해 2배 이상 층을 늘리면서 성능을 증가시킴. |
| **VGG-16 성능평가**  https://cdn.discordapp.com/attachments/902902431816499213/902936376541069363/unknown.png  https://cdn.discordapp.com/attachments/902902431816499213/902931425072910336/AyTLVxGWfQDRAAAAAElFTkSuQmCC.png  단독 사용시 epoch 10에서 가장 좋은 성능이 나옴.  Test Data의 임의의 영상프레임을 모델로 예측했을 때 올바른 값이 도출됨. |
| **ResNet설명**  <ImageNet top-1 training error>    VGG 네트워크를 통해 네트워크 모델이 깊어질수록 성능이 좋아짐을 확인함.   * 깊이를 늘리는데 한계가 있음   잔여 학습 블록을 통해 VGG 보다 깊은 네트워크를 만들어도 성능의 향상을 가져오도록함.  잔여 학습 블록 적용 후 더 깊은 네트워크에서 training error가 낮은 것을 확인함. |
| **ResNet성능평가**  https://cdn.discordapp.com/attachments/902902431816499213/902933748847374336/unknown.png  https://cdn.discordapp.com/attachments/902902431816499213/902942914089451570/unknown.png  잔여 블록을 사용하여 VGGNet과 비교하여 성능이 더욱 향상 된 것을 확인함. |
| **ResNext 설명**  <Models Bottle Neck>    128개의 채널을 32개의 그룹으로 분할하여 각 그룹당4개(128/32)의 채널이 됨.  32개의 conv는 4개의 채널 입력 값에 대해서 연산을 수행하여 4채널 feature map을 생성함.  ResNext에서는 Width보다 Cardinality를 증가시키는 것이 더 좋은 성능을 나타냄. |
| **ResNext 성능평가**  **https://cdn.discordapp.com/attachments/902902431816499213/902933755780534272/unknown.png**  **https://cdn.discordapp.com/attachments/902902431816499213/902934124816388167/unknown.png**  **https://cdn.discordapp.com/attachments/902902431816499213/902930995278393404/2021-10-25_225652.png**  ResNet을 개선한 모델로 ResNet에 비하여 성능이 좋은 것을 확인할 수 있었음. |
| **NasNetLarge 설명**  <NasNetLarge Architecture>  **https://cdn.discordapp.com/attachments/902902431816499213/902966943873650718/unknown.png**  사람의 개입 없이 순수하게 Google의 강화 학습 환경(NAS-Neural Architecture Search)에 의해 개발됨.  네트워크는 이미지를 키보드, 마우스, 연필 및 많은 동물과 같은 1000개의 개체 범주로 분류할 수 있으며 결과적으로 네트워크는 광범위한 이미지에 대한 풍부한 기능 표현을 학습했으며 이미지 입력 크키는 331 x 331 로 사용함. |
| **NasNetLarge 성능평가**  **https://cdn.discordapp.com/attachments/902902431816499213/902939792273199114/unknown.png**  **https://cdn.discordapp.com/attachments/902902431816499213/902940199959547964/unknown.png**  정확도가 높게 나오는 이유는 ImageNet 데이터베이스의 백만 개 이상의 이미지에 대해 훈련된 합성곱 신경망이기 때문에 전이 학습을 통해 높은 정확도를 도출함. |
| **GRU 설명**  GRU 구조와 수식    GRU는 기존 LSTM의 구조를 조금 더 간단하게 개선한 모델임.  마지막 출력값에 활성화 함수를 적용하지 않음.  성능 면에서는 LSTM과 비교해서 우월하다고 할 수 없지만 학습할 parameter가 더 적은 것이 장점임. |
| **VGG 16 + GRU 성능평가**  **https://cdn.discordapp.com/attachments/902902431816499213/902933158507462736/unknown.png**  **https://cdn.discordapp.com/attachments/902902431816499213/902932208615030824/unknown.pnghttps://cdn.discordapp.com/attachments/902902431816499213/902971236395675648/unknown.png**  GRU와 VGG-16을 혼용한 모델에 Test data를 예측하여 94%의 성능이 도출됨.  모델 학습시에 교차검증을 이용하고, Model Check point 를 만들어서 최적의 epoch를 활용.  학습과 예측을 할 때 비디오 Generator를 이용하였을 때 예측 성능이 96%가 나옴.  프레임에 뼈대를 입히는 방식으로 전처리한 데이터를 활용했더니 같은 모델로 학습과 테스트 했을 때 93%의 정확도가 도출됨. |
| **VGG-19 + LSTM 성능평가**    VGG19 + LSTM을 돌렸을 때 정확도가 95퍼센트 도출됨. |
| **Xception 설명**    Xception 모델은 기존의 Inception 모델을 바탕으로 완벽히 cross-channel correlations와 spatial correlations를 독립적으로 계산하고 mapping하기 위해 고안된 모델이다. Inception과 합성곱 연산의 순서가 다르고, 비선형 함수를 사용하지 않는 것이 특징이다. 보통의 신경망은 비선형함수의 도움으로 성능을 향상시키지만, Xception의 경우 신경망이 깊지 않기 때문에 오히려 비선형함수를 사용하지 않음으로써 정보손실을 방지할 수 있다. |
| **Xception + GRU 성능평가**    비디오 데이터의 3개의 프레임당 1개의 프레임을 사용하여 총 20개의 프레임을 Xception과 GRU를 혼용한 모델에 적용했더니 약92%의 성능이 도출됐다.  모델 학습시에 교차검증을 이용하고, Model Check point 를 만들어서 최적의 epoch을 활용하려 시도했다. |
| **MobileNet 설명**    고성능의 컴퓨팅 자원을 가진 환경이 아니더라도 딥러닝(CNN)을 사용하기 위한 목적으로 만들어진 모델이 MobileNet이다. |
| **MobileNet + GRU 성능평가**    MobileNet과 GRU를 혼용한 모델에 Test data를 예측하여 91%의 성능이 도출됐다.  모델 학습시에 교차검증을 이용하고, Model Check point 를 만들어서 최적의 epoch을 활용하려 시도했다. |
| **CNN + LSTM 성능평가**    **https://cdn.discordapp.com/attachments/902902431816499213/902960955141525544/cnntest.png**  CNN + LSTM을 20프레임을 5번째 프레임당 1개씩 가져와서 사용하는 프레임의 범위를 늘리고 층을 좀 더 깊게 파서 성능을 60대에서 84로 개선시킴. |
| **CNN + GRU 성능평가**    CNN과 GRU를 20프레임을 3번째 프레임당 1개씩 가져와서 사용하는 프레임의 범위를 늘리고 층이 얕은 상태의 성능은 약78%가 나왔다. 기본 CNN모델의경우, 층의 깊이를 깊게 하는 것이 성능향상에 도움이 될 것이라고 생각된다. |
| **외부 데이터 활용 테스트**    외부 데이터를 활용하여 테스트 했을 때 다음과 같은 결과들이 도출됨. |

**02. 실패 요인 분석**

|  |
| --- |
| **실패 1**  가설 1. 이미지내 사람 골격 좌표를 활용하여 행동 인식을 학습.    이미지내 사람 골격 좌표를 활용하여 행동 인식을 학습하면 성능이 좋아질 것이라 예상.  정확도가 61이 나왔고 그 이유로는 사람이 아닌 객체에 골격 좌표 인식이 일어나는 것과 다수의 사람일 때의 문제 해결이 필요하며 주위환경, 조도, 사물 또한 학습에 필요하다 생각함. |
| **실패 2**  가설 2. 이미지내 사람만 학습 시키면 무슨 운동인지 파악 할 수 있을 것이다?    이미지내 사람만 학습 시키면 무슨 운동인지 파악 할 수 있을 것이라 생각하고 객체 탐지 모델을 이용하여 person 기준 프레임으로 자른 후 객체 탐지에 실패할 경우 동일 프레임내 평균 값으로 잘라 예측함.  정확도가 61로 나왔고 그 이유로는 예측에 사용한 단순 사람의 행동 이미지뿐만 아니라 주위 환경, 조도, 사물 등 또한 필요하다고 생각함. |

**03. 소감문**

|  |  |
| --- | --- |
| 성명 | 프로젝트 후 소감 |
| 박준원 | 이번 교육의 마지막 모듈프로젝트가 끝이 났습니다. 약 4개월간 걸쳐 python에 대한 기본 지식과 데이터 전처리, 머신러닝, 딥러닝 배워왔고 이 모든 것을 활용할 수 있었던 좋은 시간이었습니다.  사람의 행동 영상을 프레임 단위로 잘라 이미지로 학습한 기본 모델학습부터, 이를 발전 시킨 시간 흐름에 따른 움직임의 변화를 학습시킨 모델로 정확도를 높여 나아갔습니다. 또한, 프레임단위를 좁히기도 하고 배경정보를 지우기 위해 사람객체 검출 후 학습도 도전해보고 많은 시간이 걸렸지만 정확도 개선을 보면서 만족할 수 있었습니다. 또한, 세상에 다양한 최신 기술과 논문들이 존재했고 새로운 기술을 접목해보며 공부할 수 있었던 좋은 시간이었습니다.  총 6명의 팀원을 이루는 조장을 맡게 되었습니다. 기존에 조원을 배정 받은 것과는 달리 팀을 자체적으로 이루게 되었고, 지금까지 같이 프로젝트를 수행해 보면서 느낀 팀원들의 장점을 생각하며 좋은 팀을 이룰 수 있었습니다. 하지만 기존과 달리 팀원을 뽑아 만든 팀장으로서 해당 임무에서 팀원들에게 만족할 만한 결과를 만들어야겠다는 책임감이 높아졌습니다. 이때, 팀원의 각 장점을 살려 각각의 임무를 나눠주고 시간을 관리하는 능력이 요구되었지만, 아직은 조금 미숙했고, 다행히 팀원들이 있었기에 무사히 프로젝트를 마무리할 수 있었고 이들에게 감사함을 느꼈습니다. |
| 김원일 | 6명으로 이루어진 팀에서 팀프로젝트를 진행하는 건 처음이라 많이 불안 했었습니다. 하지만 막상 팀원들의 도움을 받아 프로젝트를 진행하게 되니 불안함은 사라지고 같이하는데 오히려 즐거웠습니다. 처음 생각했던 나 자신의 능력 부족 때문에 팀원들이 힘들어하거나 싫어할까 많이 걱정했지만 팀원들의 격려와 제 자신이 잘 할 수 있는 일에 대하여 잘 배분해 배정해준 덕분에 프로젝트를 원활하게 진행할 수 있었습니다. 나 자신이 어디서나 주인공이 될 수는 없었지만 사람들과 함께 도움을 주고 받으며 일하는 것이 정말 즐거웠습니다 앞으로도 함께 좋은 관계로 팀원들에게 내 강점을 잘 살려 도움을 줄 수 있는 팀원으로 성장해 나갈 수 있도록 능력향상을 위해 노력해 나갈 것입니다. |
| 김택완 | 이번 프로젝트를 통해 다양한 CNN모델을 사용해보고 이전 까지는CNN이 막연히 이미지를 분류한다 정도 였고 세부적인 모델의 구조나 학습 원리는 몰랐는데 여러가지 모델을 만들어보고 성능을 개선시키기 위해 세부적인 파라미터까지 공부하며 하나씩 바꿔가면서 CNN모델에 대한 이해도가 높아 졌으며 RNN계열 모델을 같이 사용하는 것은 처음 이었는데 아직 더 공부해 보고싶다는 생각이 들었습니다. 같이 고생한 팀원들 감사하며 다음 파이널 프로젝트도 결과를 떠나 좋은 분위기로 마무리 될 수 있도록 열심히 하겠습니다. |
| 천성신 | 벌써 4차 프로젝트까지 마무리하는 시간이 다가왔습니다  4차는 원하는 사람끼리 구성되다 보니 맞는 사람을 직접 찾아 하게 되어 이전 과는 비교되는 엄청난 팀워크와 무엇보다 편했습니다 마음이 편하다 보니 프로젝트에 더 적극적으로 참여하게 되고 퀄리티 있는 결과물이 나오게 되었던 것 같습니다 저는 모델 3개( Lenet, Resnet, Nasnet)를 적용 했었는데 적용하는게 생각 보다 쉽지않아서 고생을 많이 했던 것 같습니다 하지만 질문하면 팀원들이 자기가 하고 있는 일보다 질문한 부분에 먼저 신경 써주고 적극적으로 도와주며 같이 머리 써준 팀원들이 있기에 잘 적용하여 높은 성능을 낼 수 있었습니다 미세조정을 하면서 더 높이고 싶었는데 그게 잘되지 않아 조금 아쉽기도 했던 4차 프로젝트 였습니다 이 경험을 바탕으로 최종프로젝트는 아쉬움없이 최선을 다해서 하도록 하는 것이 목표입니다 |
| 최숙희 | 마지막 모듈 프로젝트인 비디오 분석은 비디오를 프레임화해서 진행했다. 비디오당 프레임의 수를 많이 만들수록 학습시 어려움이 예상되므로 20프레임으로 진행하되, 비디오의 전반적인 내용을 반영하기 위해서 프레임화 가능한 3개의 프레임당 1개의 프레임씩 반영했다. 그러나 프레임간 간격을 좁게 형성했을 때와 더 넓게 형성했을 때의 모델성능의 차이가 미미했는데, 이는 분석에 이용한 비디오 데이터가 10초 가량의 짧은 영상이기 때문인 것으로 보인다.  조원분들과 함께 시간이 오래 걸리는 모델학습 과정에서 효율적으로 모델성능을 향상시킬 방법을 고민하면서 비디오 증강, 교차검증, callback을 사용하게 됐다. 비디오를 증강하여 학습하고, train set을 분리해서 교차검증을 하고, 교차검증시의 손실을 기준으로 최고성능을 checkpoint에 기록해두면서 epochs를 크게 활용해볼 수 있는 callback기능을 사용해봤다.  딥러닝을 이용해보니 컴퓨팅 파워의 중요성을 크게 느꼈다. 모델학습시에 속도를 높이기 위해서 batch size를 키우면 ram이 부족해서 학습을 진행할 수 없기도 했고, 다양한 모델을 분석하기 위해서 전처리된 데이터를 저장해두었다가 불러오는 방식으로 전처리를 거듭하지 않았는데, 이때 전처리된 데이터와 모델(h5)를 저장하는 데에 용량이 많이 필요해서 어려움이 있었다. 이를 통해서 AI를 공부하기 위해서는 데이터처리, 알고리즘 활용능력뿐만 아니라 컴퓨팅 파워를 확보하는 것도 필수적이라고 생각하게 되었다. |
| 허현준 | 마지막 모듈 프로젝트는 앞선 모듈 프로젝트와는 다르게 6명으로 진행되었다.  팀원이 많은 만큼 배울 점도 많았다.  처음 접근 방식은 훈련 시킬 데이터 세트에 사람의 행동만 학습시키면 연속적인 행동 패턴을 분석해  높은 정확도를 기대했다. 따라서 사람 객체를 검출 후 그 부분만 잘라서 신경망에 학습시켰지만 생각보다 높은 정확도가 나오지 않았다.  그 이유는 운동의 종류를 일기 위해서는 사람의 행동 패턴도 중요하지만  사진 내 사람의 주변 환경 또한 필요하기 때문이라고 생각한다.  마지막으로는 다양한 모델 사용, 특히 ,셀프 어텐션 방식인 트랜스포머 방식을 사용하지 못한 것이 아쉬움으로 남는다 |