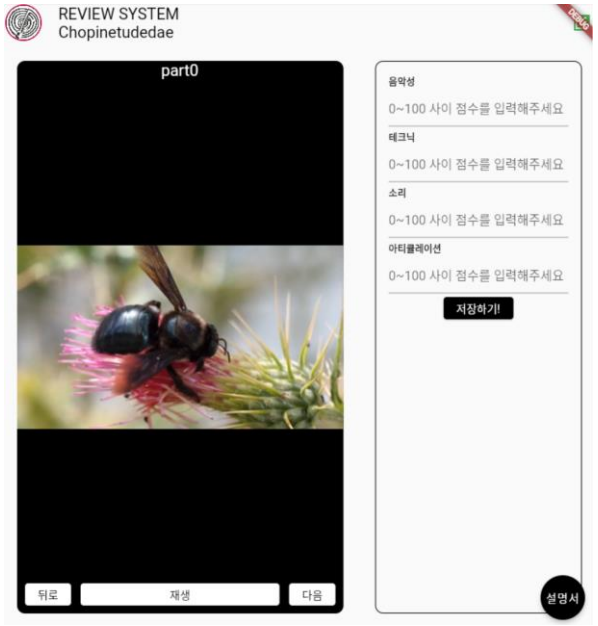
논문의 공헌



1 피아노 연주 데이터베이스 제안 → 데이터 수집 환경 계획 수립 완료 및 환경 셋팅 중에 있음

(1). 피아노 위에서 탑뷰 찍기 (-> 평가에 적합) & db 만들기
(db 만든후 수도 레이블링 통해서 재레이블링 및 디비 구축)

2 hand pose 평가 방법 제안 → Hand pose 평가 모델 훈련 데이터 수집 어플 제작 중 (90% 완료)

(1). 악보랑 소리를 체크하는 metric 제안 (수식적으로 정량화) ex. 음 체크, 박자 체크 등
(2). 딥러닝 기반의 평가 알고리즘 제안 (설문조사로 데이터 수집 (곡 듣고 이게 몇점인지), 자세랑 소리를 같이 학습 및 테스트)

3 hand pose 평가 방법 성능 비교

(1). 두 방법의 correlation 구해보기
(2). 딥러닝이 한게 나은지 hand craft가 나은지 비교

=> 아이디어가 새로워서 복잡한 알고리즘 개발할 필요 X. 기존의 거를 최대한 활용할 예정

Ⅰ. 평가기준

1 참고 자료

(1). 노래방 평가 기준

- 박자
- 음정

(2). 피아노 콩쿨 대회 평가 기준

- 음악의 분위기
- 곡의 기술적인 요소
- Phrasing
- Beat
- Rhythm
- Scale (음계)
- Fingering
- Pedaling
- Trill
- Expression

(3). 기타 피아노 연주 평가 기준

- 곡 선택의 중요성
- 다양한 음색
- 리듬 감각
- 청중과의 소통
- 연주의 전달력 / 표현력

기준 세우기

- 공통인 연주 **평가 기준 선별**
- 평가 기준의 **수식적, 정량화** 여부 판단
- Sound 로 (Audio) 만 평가할 수 있는 기준
- Hand Pose (Video) 로 평가할 수 있는 기준

Ⅱ. 평가 기준 세우기

1 **수식적, 정량화** 평가

(1). Audio (Sound) 로 평가할 수 있는 부분

- 박자 특정 시점에 소리가 있는지 판별
- 음정 음의 높이 (일정 수준 이상과 이하의 높이)
- 템포 음악의 속도나 빠르기
- 리듬 음의 높이 (일정 수준 이상과 이하의 높이)

(2). Video (Hand Pose) 로 평가할 수 있는 부분

- Fingering 어떤 손가락을 사용할 것인지
- Pedaling 피아노의 페달을 사용
- Trill (트릴) 손의 움직임을 보고 두 음을 빠르게 번갈아가면서 연주하는 것

2 **상대적** 평가 기준

(3). 딥러닝 기반으로 학습 시켜 평가할 수 있는 부분

- Technique 곡의 기술적인 요소
- Expression - 음표를 초월하여 음악적 감성과 해석력을 보여주는 능력
  - 기존의 연주자들과 얼마나 유사하게 표현 했는지로 학습

Ⅲ. 평가 알고리즘 1 : Piano Skills Assessment (1/3)

1 Paper Summary

(1). 저널 : Arxiv, 2021
(2). 저자 : Paritosh Parmar et al (University of Nevada, Las Vegas)
(3). 제안하는 방법 :

- Multimodal Pisa Dataset : 피아노 연주 평가를 위한 데이터 셋 제안함
- CNN 모델 기반의 피아노 연주 레벨 예측 모델 제안

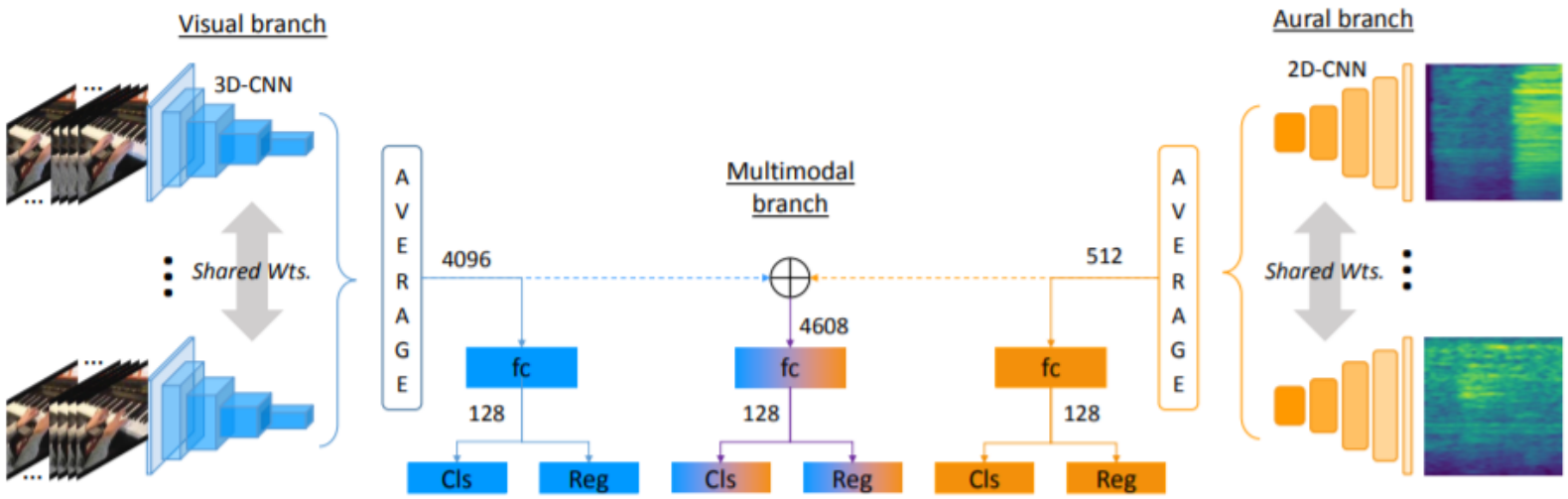


Fig. 4. Our multimodal learning architecture. ⊕ represents concatenation operation.

Ⅲ. 평가 알고리즘 1 : Piano Skills Assessment(2/3)

1 Paper Summary

(1). Visual branch :

- 음악을 오디오로 평가하는게 아닌 비쥬얼로 평가하는게 불가능해보이지만, 비쥬얼로 평가하는게 가능한 몇가지 skill 이 있음. 첫 번째는 테크닉적인 기술이고 두 번째는 virtuoistic skill(professional skill)임
- 테크닉적인 기술 : scale, arpeggio, octave 를 빠른속도내에 칠 수 있는지
- Virtuoistic : 숙련자들은 8음을 도약할때 첫번째 손가락과 세번째 손가락을 이용하지만 초보자들은 그렇게 하지 못함

(2). Aural branch : 오디오로 평가할수 있는 요소들

- Velocity : 음량 (소리의 크기)
- Cadence에 포함되는 음의 개수

But, 곡의 스타일 마다 다름 → CNN 처리를 위해 오디오를 melspectrogram으로 변환

(2. Cadence의 경우 비쥬얼로도 평가할 수 있는 요소)

*Cadence : 연주자의 기교와 음색을 발휘할 수 있는 즉흥 연주

(3). Multimodal branch

- Visual branch와 Aural branch의 feature을 concatenate

(4). Objective function (Player level prediction problem)

- 예측 레벨과 정답 레벨 사이의 거리를 측정하여 제안하는 모델의 성능을 평가함 (L1, L2 distances)

Ⅲ. 평가 알고리즘 1 : Piano Skills Assessment(3/3)

1 Paper Summary

(1). Preprocessing :

- 모든 영상이 아닌 연주자의 레벨을 측정하는데 도움이 될만한 부분들을 crop (forearms, hands 있는 영상들만)
- 오디오는 librosa를 이용하여 melspectrogram으로 변경함

(2). Implementations details

- PyTorch, Adam optimizer, learning rate (0.0001), 100epochs, batch size 4

(3). Visual branch

- 데이터 셋 크기가 작아서 3DCNN 모델 사용
- 데이터과적합 피하기 위해서 UCF101 데이터셋으로 pretrain

(4). Aural branch

- ResNet-18
- ImageNet으로 weight 초기화 (성능 향상에 도움)
- Random cropping 성능 저하되서 삭제 (유의미한 정보 삭제)

Fig. 1. Examples of samples from our dataset. First row

Modality	Sampling Scheme	
	Contiguous	Uniformly Dist.
Video	65.55	73.95
Audio	53.36	64.50
MMDL	61.55	**74.60**

Table 1. Performance (accuracy in %) of single modalities vs a multimodal (MMDL) assessment for contiguous and uniformly distributed sampling schemes.

Ⅲ. 궁금한 점

1 카메라 추천

2 유사한 논문 많이 발견 -> 비교분석해서 차별성 있게 논문 공험 설정 필요해 보임

[piano]

(1) https://www.hindawi.com/journals/wcmc/2022/6727429/
(2) https://www.frontiersin.org/articles/10.3389/fpsyg.2022.954261/full
(3) https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0250299
(4) Automatic Evaluation of Piano Performances for STEAM Education
(5) https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8133499/

[violin]

(1) Quantitative evaluation of violin solo performance
(2) Design of the Violin Performance Evaluation System Based on Mobile Terminal Technology
(3) In-Tool Motion Sensing for Evaluation of Violin Performance

논문 읽으면서 체크해야되는 부분들!
- 오디오/비디오 데이터 사용 여부
- 초급자 대상 or 고급자 대상 평가 방법인지? 따로 언급이 없는지?
- 비디오 데이터 사용한 논문의 경우 손 pose estimation 정확도 고도화 작업을 했는지
- 데이터셋의 크기가 얼마나 되는지도 각 논문별로 추가로 정리