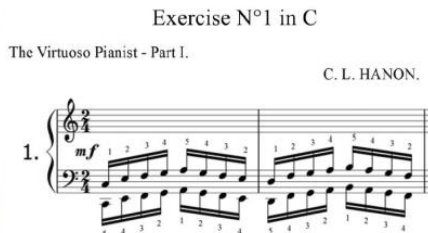


# 기존의 피아노 연주 평가 연구 1 ★

## Piano Tutoring System (ICPR 2018, h5-index 95)



**요약 :** 학생의 손 움직임과 음악적 정확도에 대한 피드백을 제공하는 상호작용 가능한 피아노 교육 시스템 제안

### 제안하는 방법 :

1. 어떤 시점에 어떠한 음이 연주되었는지 체크
2. 몇번째 손가락으로 연주했는지 체크
3. 1, 2를 위하여 two-stream convolutional neural network 제안 (두 데이터 incorporate 위해서 object detection model 개선함)

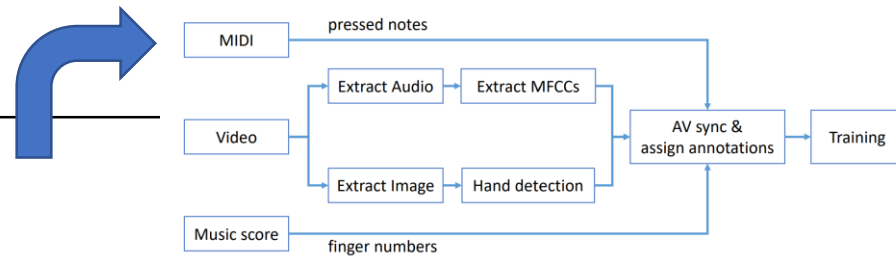
**유사점 :** 피아노 교육 시스템이지만 오디오와 비디오 두개다 체크 한다는 점에서 평가 시스템과 유사

**차이점 :** output이 점수가 아님

### 참고하면 좋을 점 :

어떤 모델 사용했는지, 데이터셋, 환경 어떻게 셋팅했는지

## 데이터셋 :



### [모델의 인풋]

- 비디오(이미지와 오디오 누울)
- 미디 파일
- 악보를 인풋 데이터로 사용

### [One Hand Hanon]

- 하논을 한손으로 연주
- 한명이 1번부터 5번까지 곡 연주  
1 to 3 for training -> 23,555 프레임  
4 to 5 for evaluating -> 11,777 프레임
- 10개의 비디오
- 곡 길이는 50에서 120초
- 각각 손으로 한번씩 연주 -> 총 10번

### [Hand Hanon]

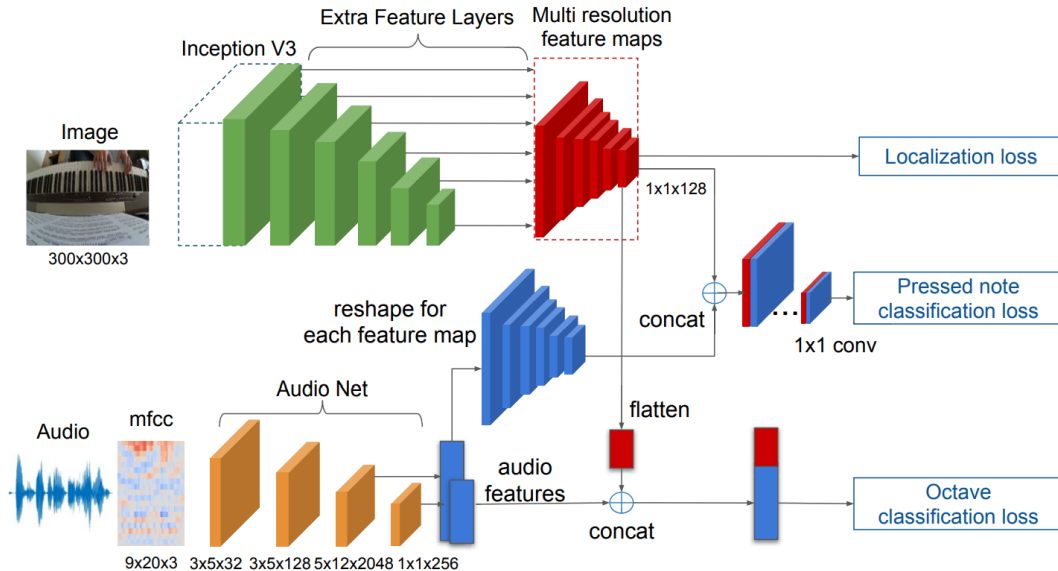
- 한명이 1번부터 5번까지 곡 연주
- 총 51,596 프레임  
2 to 4 for training (36,115 frames)  
1 and 5 for evaluating (15,481 frames)

### [Real Pices]

- 총 6곡 (Bach, Reinagle, Glinka, Schumman, Vlkman 등)
- 5명의 피아니스트 녹음 참여. 2 고급, 2 중급, 1 초급
- 65 분 비디오 촬영 (60 fps)

# 기존의 피아노 연주 평가 연구 1

## Detecting Piano Keypress



입력 값 : 이미지와 mfcc

결괏 값 : 옥타브 안에서 연주된 음, 몇번째 옥타브인지

모델 : 비디오 (Inception V3), 오디오 (Audio Net)

## Fingering Identification

입력 값 : 오디오 제외한 비디오만 사용

결괏 값 : finger numbers

모델 :

- 첫번째 방법 : Data Annotation 비용이 많이 들기 때문에, 다른 연구에서 제안한 Bambach 데이터셋으로 훈련. 이 후에 수집한 Piano data로 다시 훈련
- 손가락이 안보여서 annotation이 불가할 경우  
인접 프레임 중 손가락 추출이 잘 된 데이터를 사용 (확인 필요)

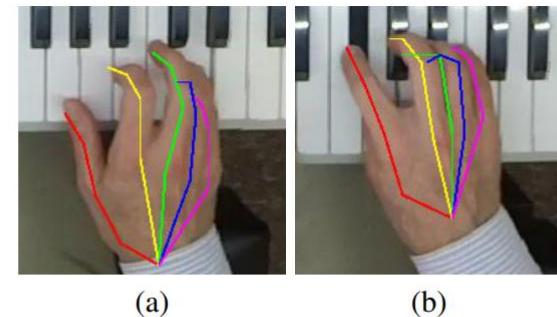


Figure 8. Two failure cases in detecting the fingering. (a) Confusion between middle and ring finger because of vicinity of fingers. (b) Confusion between thumb and index finger due to a crossover and resulting occlusion.

- 두번째 방법 : Pose estimation 방법으로 손가락 스켈레톤 추출 후 인접해 있는 건반(Hough transform of Sobel edges)을 연주한 음으로 판단

## 기존의 피아노 연주 평가 연구 2

Piano Performance Evaluation  
(Wireless Communications and Mobile Computing,  
2023, IF : 2.146)

**요약 :** 피아노 교육 산업의 지속 가능한 발전에 일정한 역할을 할 수 있는 과학적인 피아노 연주 평가 체계를 설계하는 것이 목적

### 제안하는 방법 :

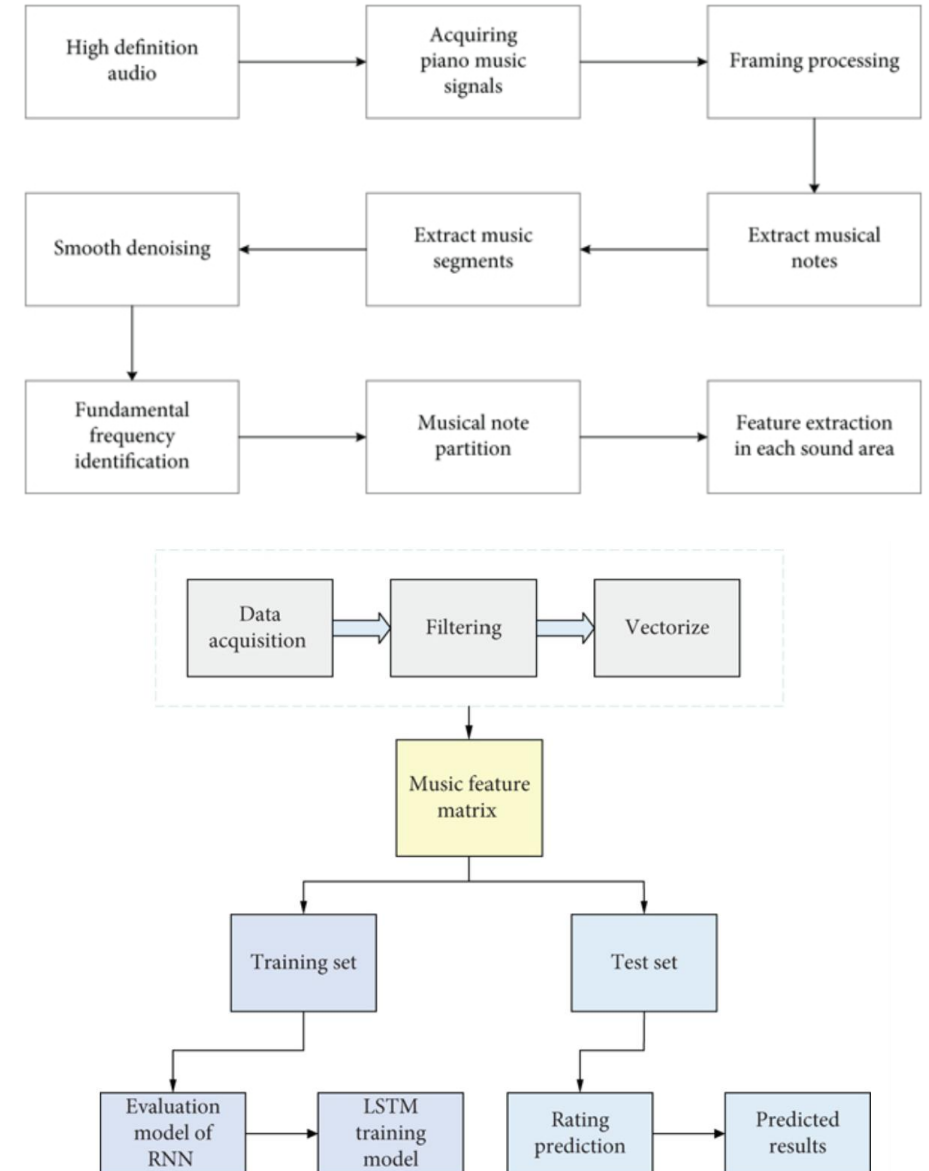
1. 딥러닝의 LSTM 모델을 탐구합니다.
2. 피아노 음악 특징을 추출하고, LSTM 기반 악기 디지털 인터페이스 피아노 연주 평가 모델을 구축
3. 피아노 연주 평가를 위한 장단기 기억 모델에 구현된 숨겨진 레이어 수를 분석

**유사점 :** 소리로 연주 평가

**차이점 :** 자세를 평가하는 부분이 없음, 성능 개선이 없어 보임

### 참고하면 좋을 점 :

점수 예측 시 어떤 걸 평가 요소로 넣었는지, 평가 요소와 점수 예측 방법을 수식으로 쉽고 적절하게 표현함 (리듬, 비트), MIDI로 부터 얻을 수 있는 정보에 어떤 것들이 있는지. 데이터셋 수집이나 환경 구축 관련된 내용이 논문 상에 없는 듯



## 기존의 피아노 연주 평가 연구 3

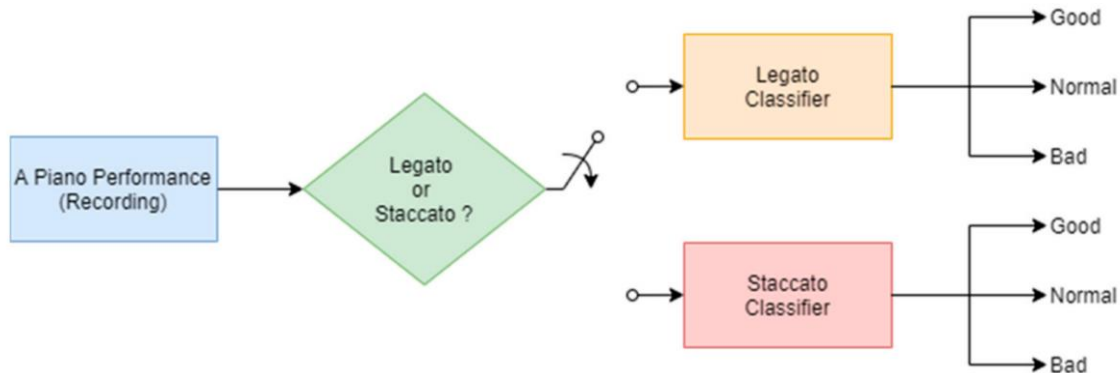
Piano Performance Evaluation  
(Applied Sciences, 2021, IF : 2.7)

**요약 :** 음악 연주 평가의 표준화 필요성을 언급하며, 레가토/스타카토 등의 피아노 소리를 분석(아티큘레이션)해서 좋음/보통/나쁨의 범주로 분류  
(SVM, NV, CNN, LSTM의 네가지 접근 방식을 조사함)

**데이터셋 :** 13명의 연주자가 제작한 격리된 음계와 동요를 포함한 4670개의 테스트 샘플 사용하여 수행

**유사점 :** 소리로 연주 평가

**차이점 :** 자세를 평가하는 부분이 없음, 소리에서 다양한 특징 추출, 성능 개선이 없어 보임



## 기존의 피아노 연주 평가 연구 4 ★

Piano Performance Evaluation  
(TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, 2021, IF : 5.4)

**요약 :** 초보자를 위한 두 가지 오디오 기반 피아노 연주 평가 시스템을 제안

**제안하는 방법 :**

1. CNN 기반 음향 특징 추출, DTW를 통한 매칭, 성능 점수 회귀의 세 단계를 포함하는 순차적이고 모듈화된 시스템
2. CNN과 어텐션 메커니즘을 갖춘 엔드투엔드 시스템. 두 개의 음향 특징 시퀀스를 입력으로 사용하고 성능 점수를 직접 예측.

**데이터셋 :** YCU-PPE-III(2000개 이상)를 사용하여 제안된 두가지 방법을 평가 (0~100점 범위에서 3.79의 평균 절대 오차 달성)

**유사점 :** 소리로 연주 평가

**차이점 :** 자세를 평가하는 부분이 없음

**참고하면 좋을점 :** 관련연구, 데이터셋 수집 방법 등  
(따로 요청하면 데이터 받을 수 있음)

# 기존의 피아노 연주 평가 연구 4

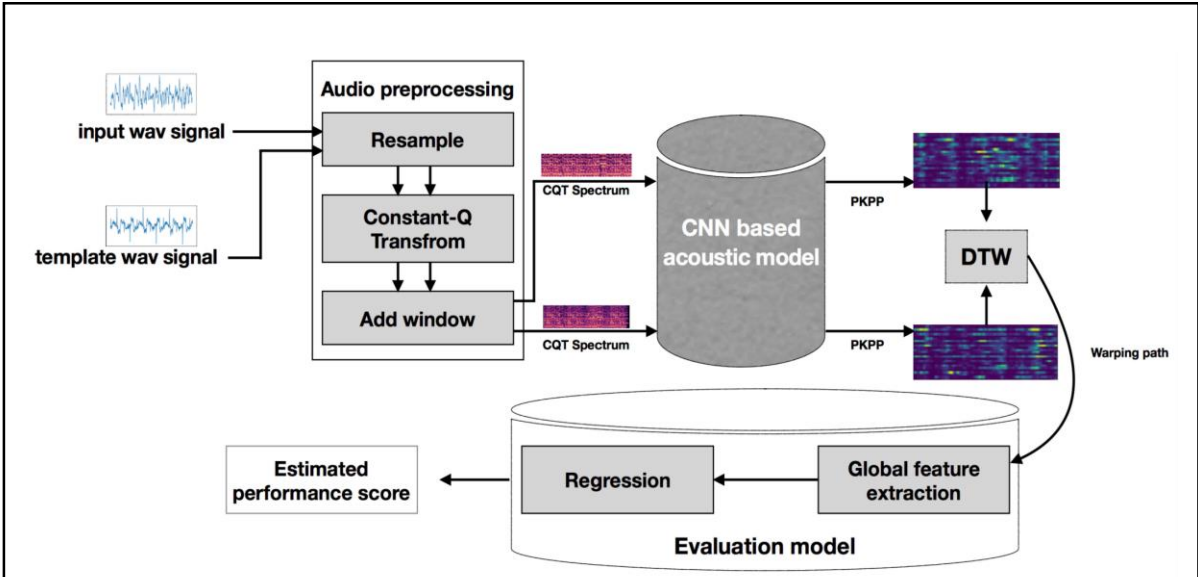


Fig. 1. The architecture of the proposed modularized system.

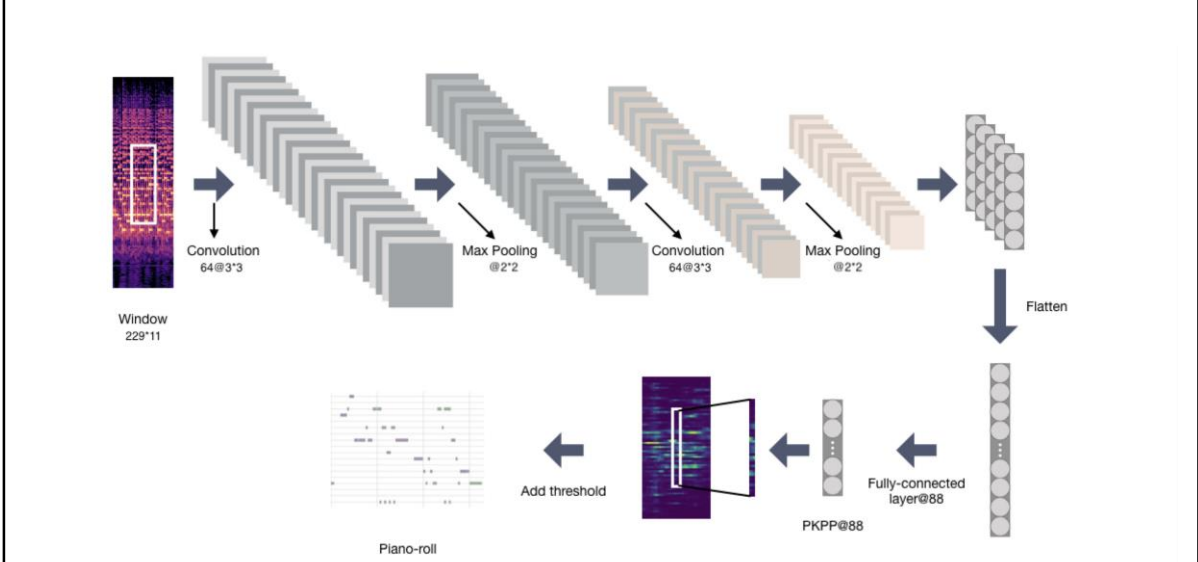


Fig. 2. The architecture of our acoustic model. This figure shows one of our three acoustic models, and the dimension of input and output is 229 and 88. The setups of the other two models are 229-d input to 12-d output and 88-d input to 12-d output, respectively. These three acoustic models have similar architectures.

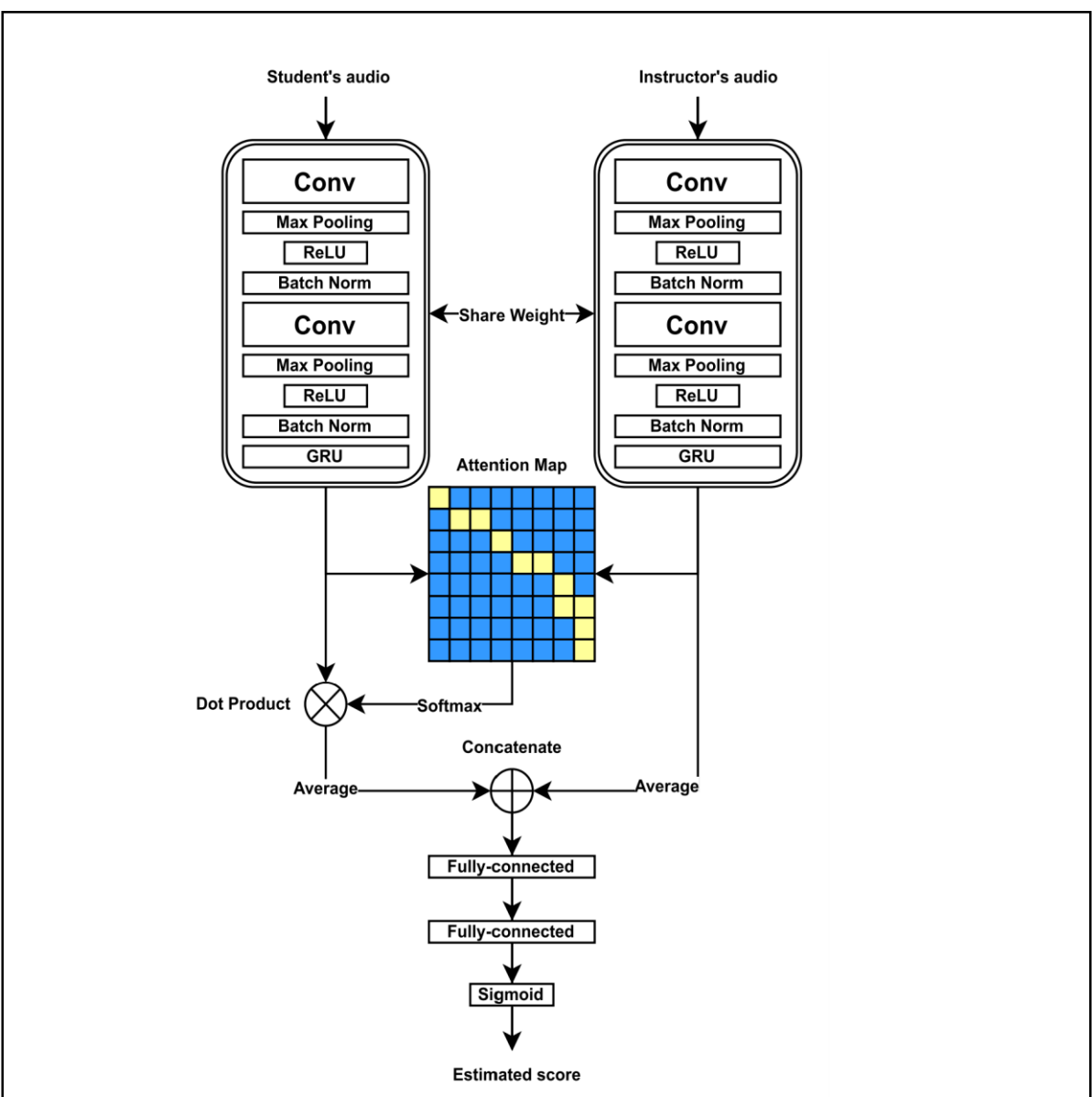


Fig. 5. Architecture of the proposed end-to-end system.

## 기존의 피아노 연주 평가 연구 5

Applied Sciences, 2023, IF : 2.7)

**요약 :** 플레이어의 손가락 움직임 트랙의 시각적 특징과 피치 및 리듬을 포함한 청각적 특징을 모두 추출할 수 있는 ResNet 기반 시청각 융합 모델을 제안함

### 제안하는 방법 :

1. ResNet 기반 시청각 융합 모델 제안
2. 비디오와 오디오 간의 상관 관계 및 보완 정보를 캡처해서 기능 융합 기술을 통해 관절 특징을 획득하여 플레이어의 기술 수준을 포괄적이고 정확하게 평가 가능
3. 피아노 연주에서 복잡한 시간적 (visual brach) 및 주파수 특징을 추출 할 수 있음

**데이터셋 :** PISA 데이터 세트

**유사점 :**

**차이점 :**

**참고하면 좋을점 :** 코드있는지, 데이터 PISA 받기, 데이터 속성 정확하게 어떤거 뽑은건지?

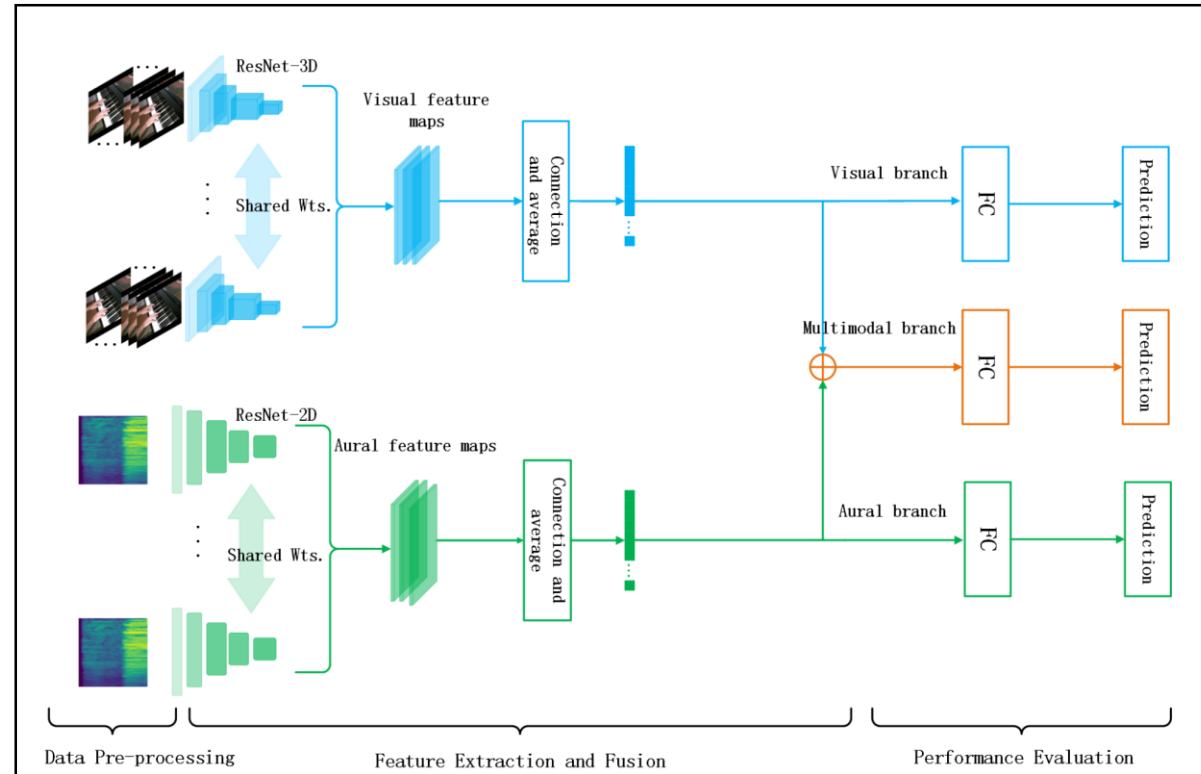



Figure 1. Framework of audio-visual fusion model for piano skills evaluation.



## 기존 컨트리뷰션

1. 피아노 연주 데이터베이스 제안
  - v 피아노 위에서 탐뷰 찍기 (-> 평가에 적합) & db 만들기 (db 만든후 수도 레이블링 통해서 재레이블링 및 디비 구축)
2. hand pose 평가 방법 제안
  - v 악보랑 소리를 체크하는 metric 제안 (수식적으로 정량화) ex. 음 체크, 박자 체크 등
  - v 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
  - v 두 방법의 correlation 구해보기
  - v 딥러닝이 한게 나은지 hand craft가 나은지 비교

## 컨트리뷰션 재설정

1. 피아노 연주 데이터베이스 제안
  - v 기존에는 피아노 연주 평가에 사용되는 특징들이 한쪽에 치우쳐져 있음  근데 이것도 이미 다른 논문에서 한것 같다..
  - > 본 논문에서는 기존 논문들을 종합적으로 정리하여 연주 평가에 사용될 수 있는 모든 요소들을 추출하고 이를 데이터베이스로 제공함
  - v 기존에는 옆에서 찍는거를 사용. 또한, 음성을 사용하기때문에 pose로 인한 정확한 페달정보, 음정 정보의 정리가 어려움
  - > 정확한 평가를 위한 훈련 과정이 중요한데. 부정확한 데이터로 훈련하기 때문에 평가 모델의 신빙성 떨어짐
  - > 전자 피아노로 데이터 수집 및 평가에 적합한 위에서 촬영한 영상 사용. Midi로 페달링, 음정 등 정확한 정보를 입력 받음
2. Hand pose 평가 방법 제안
3. hand pose 평가 방법 성능 비교

# 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

## 향후 연구

1. openpose로 fine tuning 해보까 (이번에 새롭게 수집해놓은 데이터셋으로..)

fine tuning 이 뭘까 vs transfer learning? (심심할때 조금씩 해보기)

- Openpose 기존 모델 훈련 뒤에 fine tuning 할때 피아노 연주 영상에서 소리(다양한 요소 뽑아보기)랑 자세 뽑아서 트레이닝 정확도 테스트
- 뭐가 자세가 디텍트 안될때 소리에 attention 을 더 주도록 하려면 어떻게 모델을 수정해야할지 고민해보기

2. 졸업 논문에서 audio visual 데이터에 대해 tensor fusion 했었는데 이거 하는 방법 있을까?

3. 옆모습 뿐만 아니라 위에서 찍은것도 해서 총 3개의 데이터셋 (pisa, c3pap, 위에서 수집한 새 데이터)에 대해서 성능을 테스트 해보는게 어떨까?