Deep learning based Medical Image Classification

Sanghyuck Lee

2022. 09. 14

Overall Progress

AutoML

Problem	Related Work / Survey (Counterparts)	Experiments	Analysis	Proposal
▪ Deep learning based Medical Image Classification	▪ 2021 MICCAI accepted papers survey			

Introduction

- In recent years, with the development of convolutional neural networks (CNN), research on deep learning-based medical image analysis has received significant attention.
- (Existing approaches' limitation)
- (Proposed method)
- (Proposed method in detail)
- (Experiment)

Related Work

AutoML

Recent Trends Survey with MICCAI 2021's accepted papers

- Total 531 papers:
  - Choose 2 Topics among belows: Computer Aided Diagnosis, Outcome disease prediction

Clinical Applications	Machine Learning	Modalities	Others	
✓ Abdomen ✓ Breast ✓ Cardiac ✓ Dermatology ✓ Fetal Imaging ✓ Lung ✓ Musculoskeletal ✓ Neuroimaging - Brain Development ✓ Neuroimaging - DWI and Tractography ✓ Neuroimaging - Functional Brain Networks ✓ Neuroimaging - Others ✓ Oncology ✓ Ophthalmology ✓ Vascular	✓ Active Learning ✓ Attention models ✓ Domain adaptation ✓ Interpretability / Explainability ✓ Reinforcement learning ✓ Self-supervised learning ✓ Semi-supervised learning ✓ Uncertainty ✓ Weakly supervised learning	✓ Biophotonics ✓ CT ✓ EEG/ECG ✓ Histopathology ✓ MRI ✓ Microscopy ✓ PET/SPECT ✓ Text (clinical/radiology reports) ✓ Ultrasound ✓ Video ✓ other	✓ Computational (Integrative) Pathology ✓ Computational Anatomy and Physiology ✓ **Computer Aided Diagnosis** ✓ Image Reconstruction ✓ Image Registration ✓ Image Segmentation ✓ Image-Guided Interventions and Surgery ✓ Integration of Imaging with Non-Imaging Biomarkers ✓ Interventional Imaging Systems	✓ Interventional Simulation Systems ✓ **Outcome/disease prediction** ✓ Population Imaging and Imaging Genetics ✓ Surgical Data Science ✓ Surgical Planning and Simulation ✓ Surgical Skill and Work Flow Analysis ✓ Surgical Visualization and Mixed, Augmented and Virtual Reality ✓ Visualisation in Biomedical Imaging

Recent Trends Survey with MICCAI 2021's accepted papers

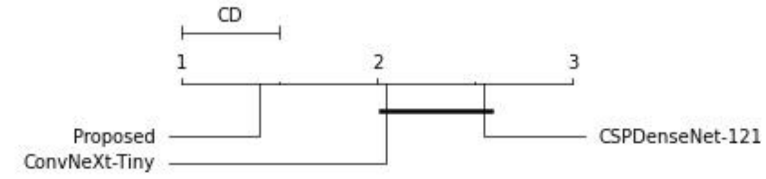
- Investigate papers' abstract
  - Outcome disease prediction (53 papers)
  - Computer Aided Diagnosis (145 papers)

산학-2 실적 현황

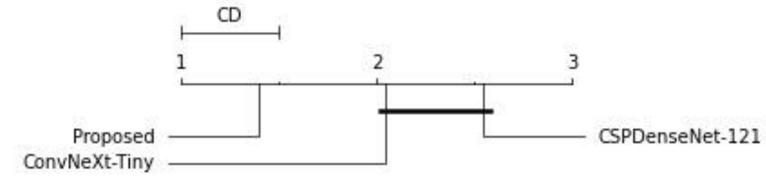
- 국제 논문 2편 :
  - 2022 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia) 투고 완료 (Accepted papers notification: 9.24)
  - IEEE Access 1건 논문 작업 진행중
- 특허 2건
  - 2건 모두 변리사님 다음 연락 대기중

Segmentation & Activity

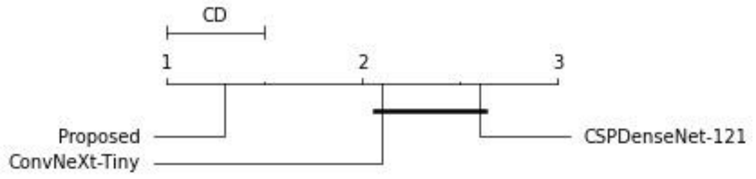
- Segmentation: Manuscript submitted?
- Activity: 이정규 교수님께 영문본 회신, ~9.14(오늘)



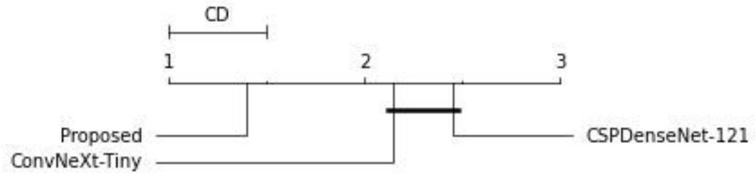
AUROC



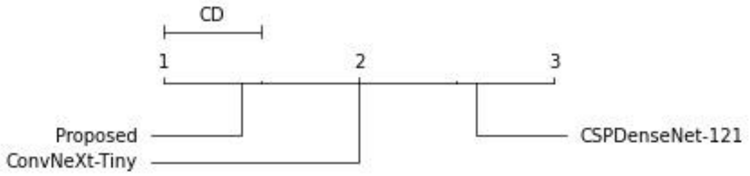
Accuracy



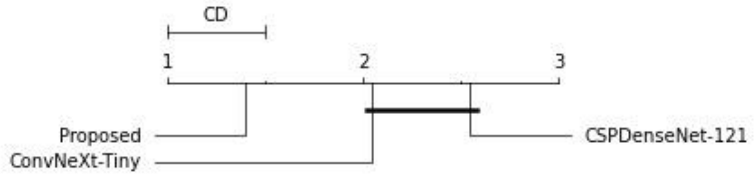
F1 score



Sensitivity



Specificity



Precision

Activity 연구의 Bonferroni-dunn test 결과

TAO studies with CAU Hospital & NVIDIA

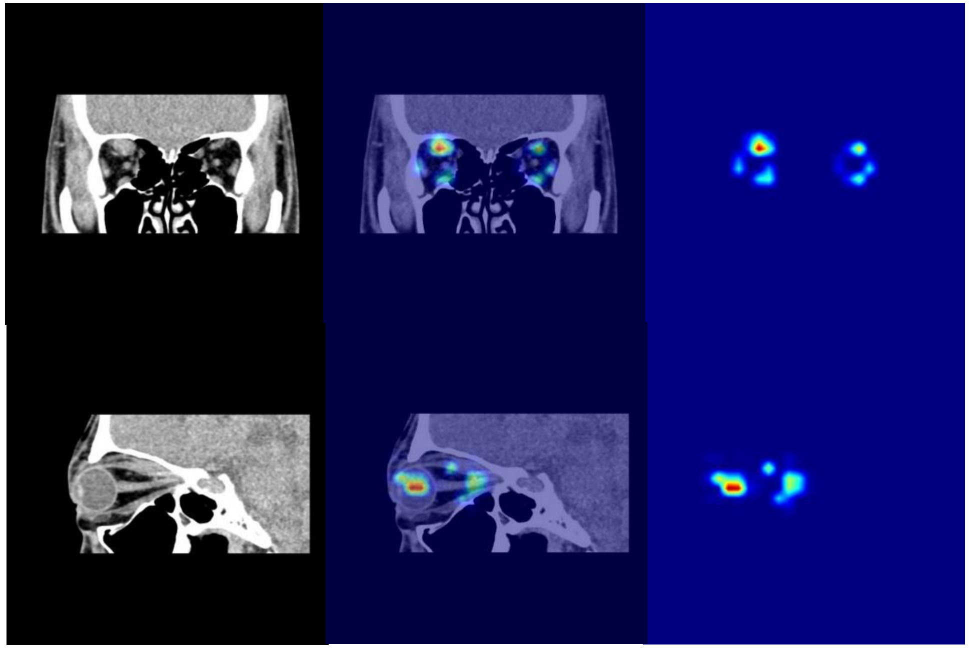
Activity

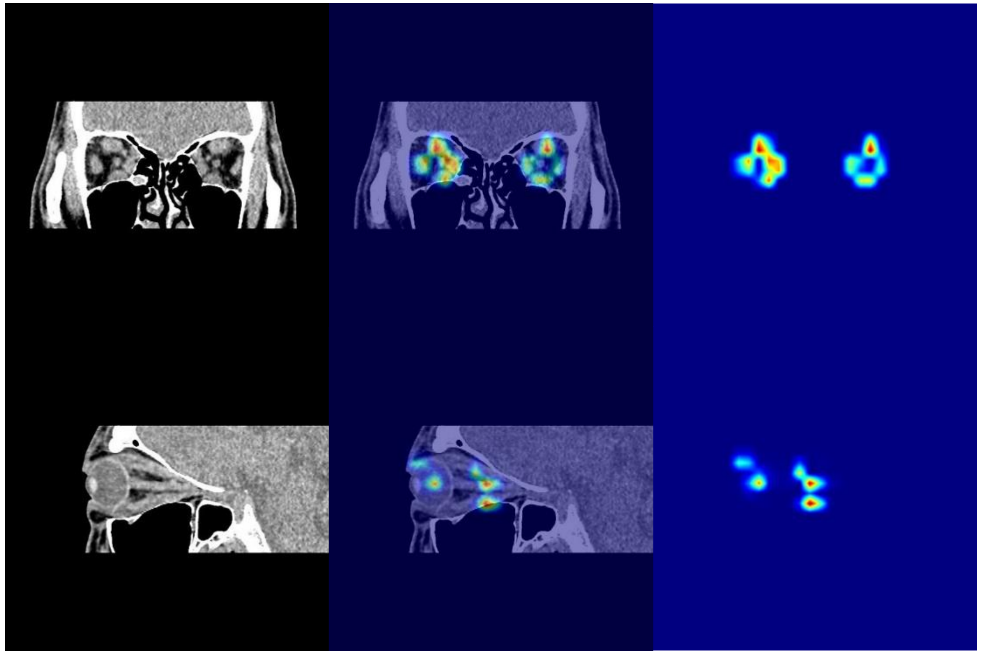
- 기본적으로 입력되는 구조물들을 폭넓게 사용하여 예측을 진행하나, 종종 근육의 두꺼워지는 부분을 중심으로 예측을 진행하기도 함.
- SA_R에서는 CO2에서 볼 수 없는 근육의 단면을 보거나, 혹은 eyeball과 같은 다른 구조물에 대한 정보를 유효하게 사용하기도 하면서, CO2만 입력하였을 때보다 높은 성능을 나타내는 것으로 보임.
- 해당 내용들을 이정규 교수님께서 어떻게 생각하실지 궁금함.

Original Overlay Heatmap

CO2

SA_R





Treatment

- Treatment using 3D dataset (with NVIDIA)
  - 실험 코드 작업 마무리(MONAI기반), ~9.15 ->~9.16
  - 3D 모델 재조사 및 구현, ~9.22
  - 차기 회의는 9월 22일 오전 10시~11시? 혹시 관련 일정 협의하신 내용이 있으실지 문의 드립니다.

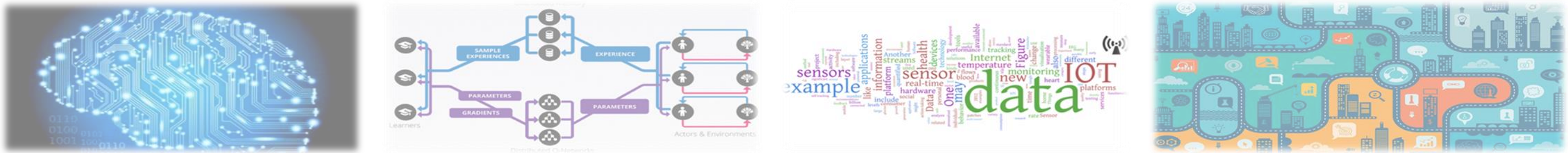
연말 해외 출장

- 목표 컨퍼런스 변경 :
  - 기존에 목표하던 9.16 데드라인 파타야 컨퍼런스가 논문 미완성 등의 이슈로 다른 컨퍼런스를 목표 변경하고자 함(Bangkok, Tokyo 등).
- 예산 축소 :
  - 등록비를 고려하지 못해, 기존 2천만 예산에서 줄여야할듯.

졸업 논문

AutoML

- 초안 수정 진행중



Effective Music Genre Classification using Late Fusion Convolutional Neural Network with Multiple Spectral Features

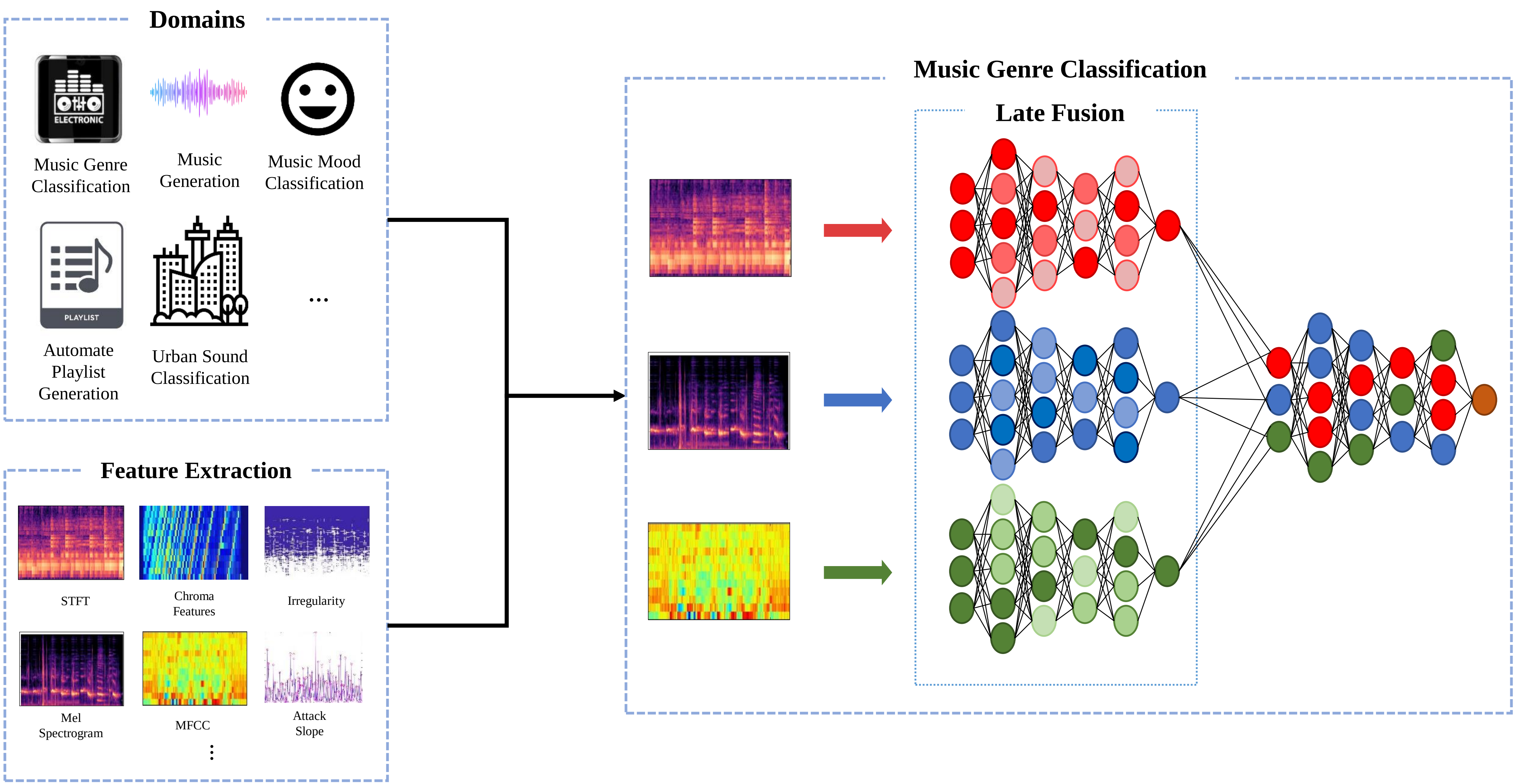
Sung-Hyun Cho

2022. 09. 14

Overall Progress

AutoML

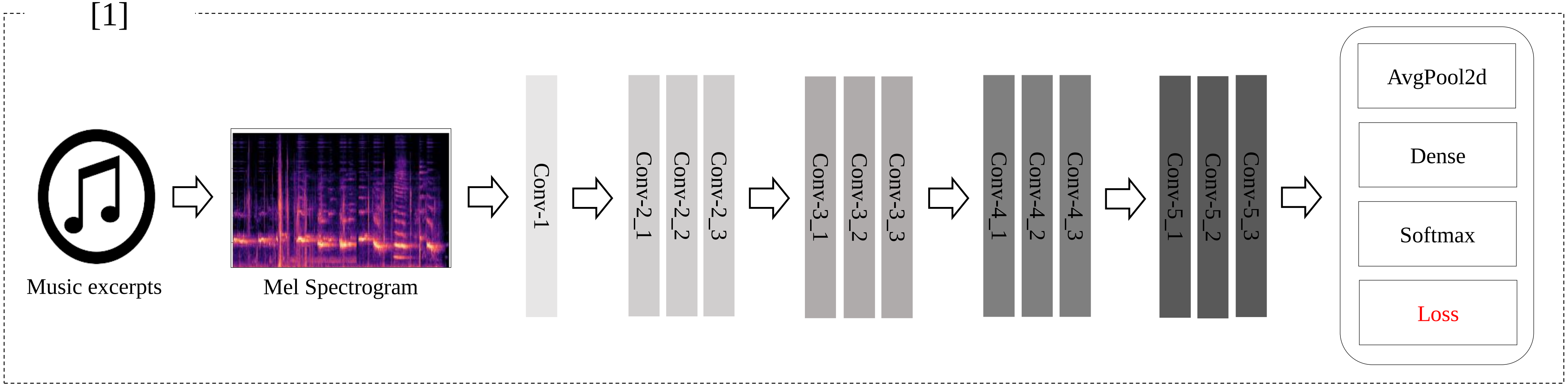
Problem	Related Work / Survey (Counterparts)	Experiments	Analysis	Proposal
▪ Music Genre Classification ▪ Low Accuracy	▪ MGC using single input	▪ Significantly meaningful ▪ Additional Experiments	▪ Complement each other's insufficient information by using multiple features at the same time	▪ Using STFT, Mel-spectrogram, MFCC


Domains
Music Genre Classification
Music Generation
Music Mood Classification
Automate Playlist Generation
Urban Sound Classification
...
Feature Extraction
STFT
Chroma Features
Irregularity
Mel Spectrogram
MFCC
Attack Slope
Music Genre Classification
Late Fusion

Related Work

[1] Li, J., Han, L., Li, X., Zhu, J., Yuan, B., & Gou, Z. (2022). An evaluation of deep neural network models for music classification using spectrograms. Multimedia Tools and Applications, 81(4), 4621-4647.

[1]



$$\boxed{\text{Loss}} = (1 - t) \cdot Loss1 + t \cdot Loss2$$

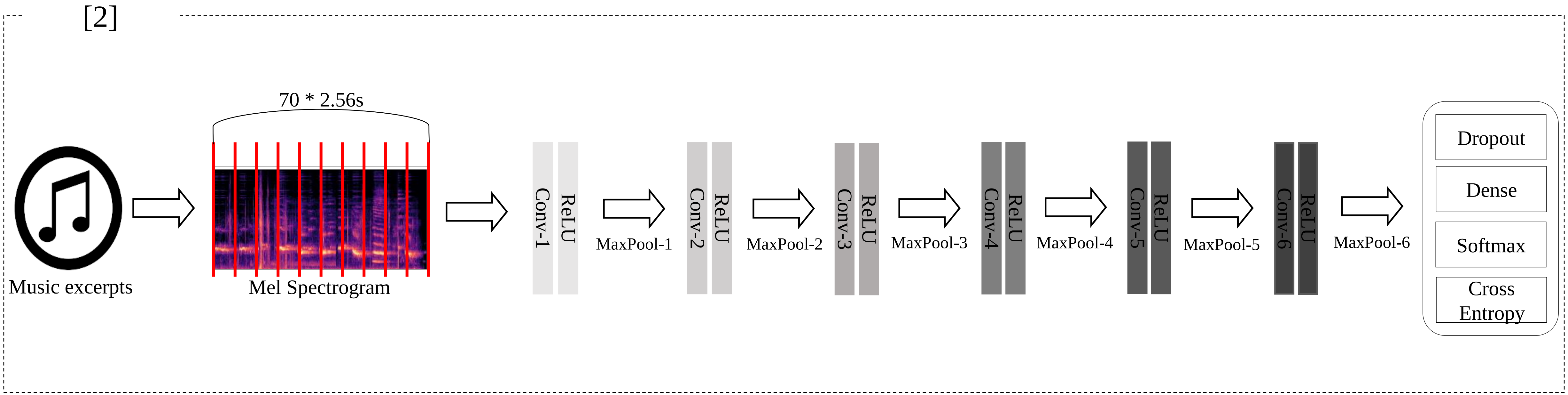
$t = a^{-b}$ $(a : balance\ base, b : balance\ factor)$

$\text{Loss1} = \text{Cross\_Entropy}(\text{y}_{\text{true}}(\text{Onehot encoding}), \text{y}_{\text{pred}}(\text{Softmax function}))$

$\text{Loss2} = \text{Cross\_Entropy}(\text{y}_{\text{true}}(\text{Uniform Distribution function}), \text{y}_{\text{pred}}(\text{Softmax function}))$

Related Work

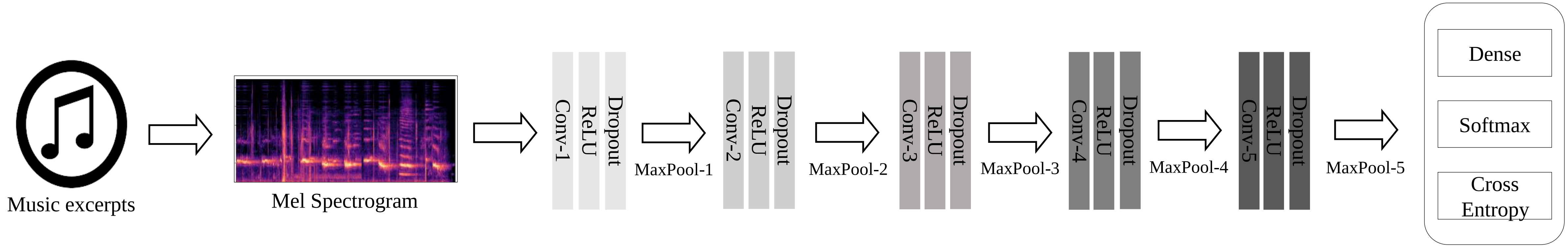
[2] Pelchat, N., & Gelowitz, C. M. (2020). Neural network music genre classification. Canadian Journal of Electrical and Computer Engineering, 43(3), 170-173.



Related Work

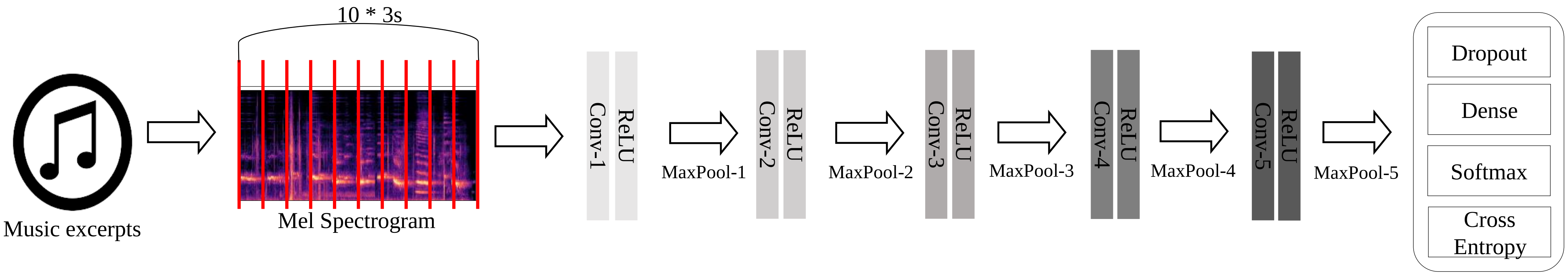
[3] Cheng, Y. H., Chang, P. C., & Kuo, C. N. (2020, November). Convolutional neural networks approach for music genre classification. In 2020 International Symposium on Computer, Consumer and Control (IS3C) (pp. 399-403). IEEE.

[3]

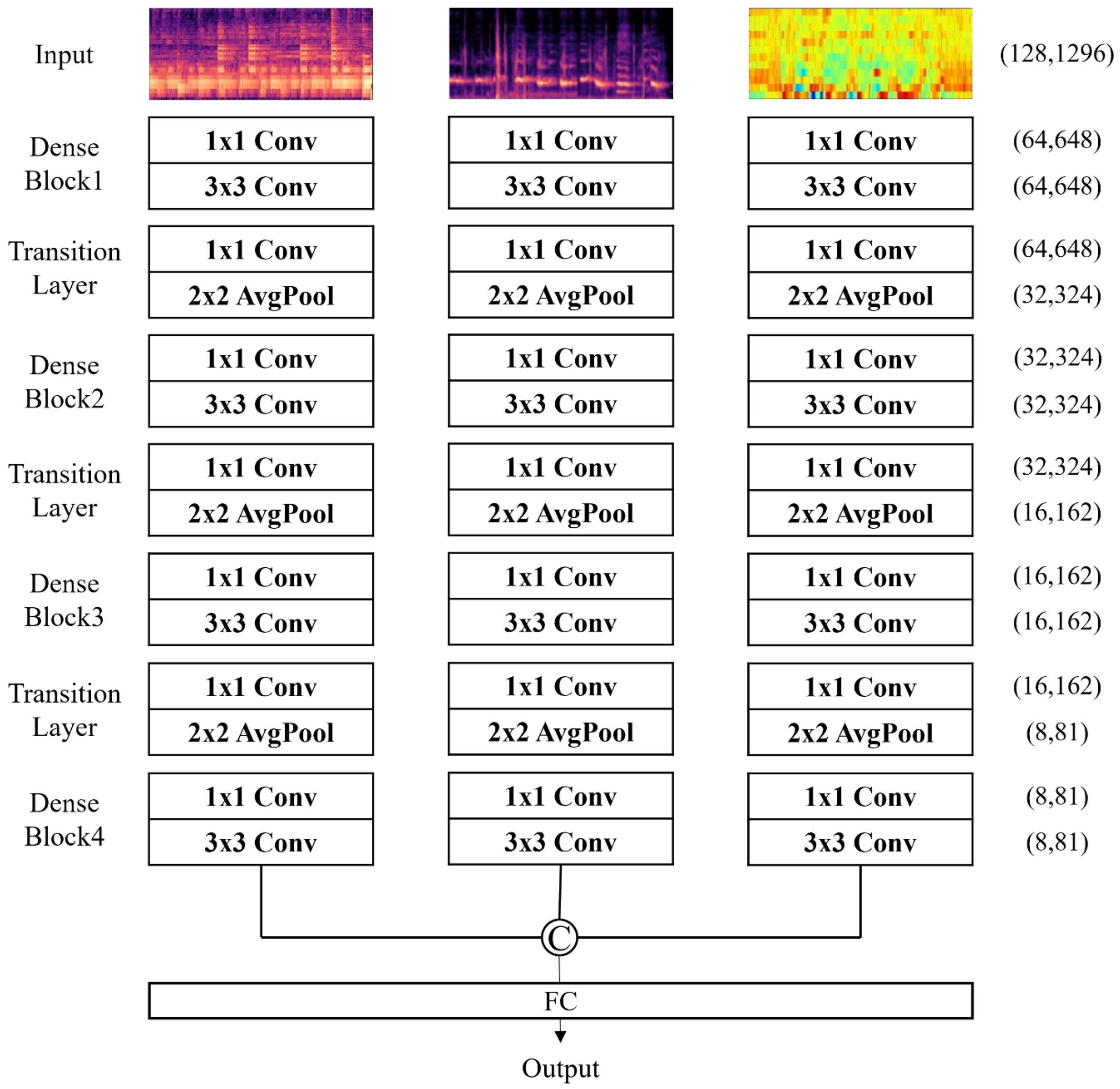


[4] Shah, M., Pujara, N., Mangaroliya, K., Gohil, L., Vyas, T., & Degadwala, S. (2022, March). Music Genre Classification using Deep Learning. In 2022 6th International Conference on Computing Methodologies and Communication (ICCMC) (pp. 974-978). IEEE.

[4]



Proposed Method



Experimental Settings

Datasets

Dataset	Patterns	Avg. Length	# of Classes	Average Patterns per Class	±	Standard Deviation
Ballroom	698	30	8	87.250	±	17.555
Ballroom-Extended	4180	30	13	321.54	±	195.70
emoMusic-G	744	45	8	93.00	±	16.1776
Emotify-G	400	43	4	100.00	±	0.00
FMA-SMALL	7996	30	8	999.500	±	0.500
GiantStepsKey-G	604	80	23	24.9130	±	28.0274
GMD-G	1042	300	18	57.89	±	70.07
GTZAN	1000	30	10	100.00	±	0.00
HOMBURG	1886	10	9	209.56	±	134.87
ISMIR04	727	120	6	121.167	±	95.602
MICM	1137	360	7	162.429	±	119.717
Seyerlehner-Unique	3115	30	10	311.500	±	248.349

Cross=Validation for Test Performance

- 10 iterations
- Data Split Ratio: (Train : Test) = (0.8 : 0.2)

Experimental Settings

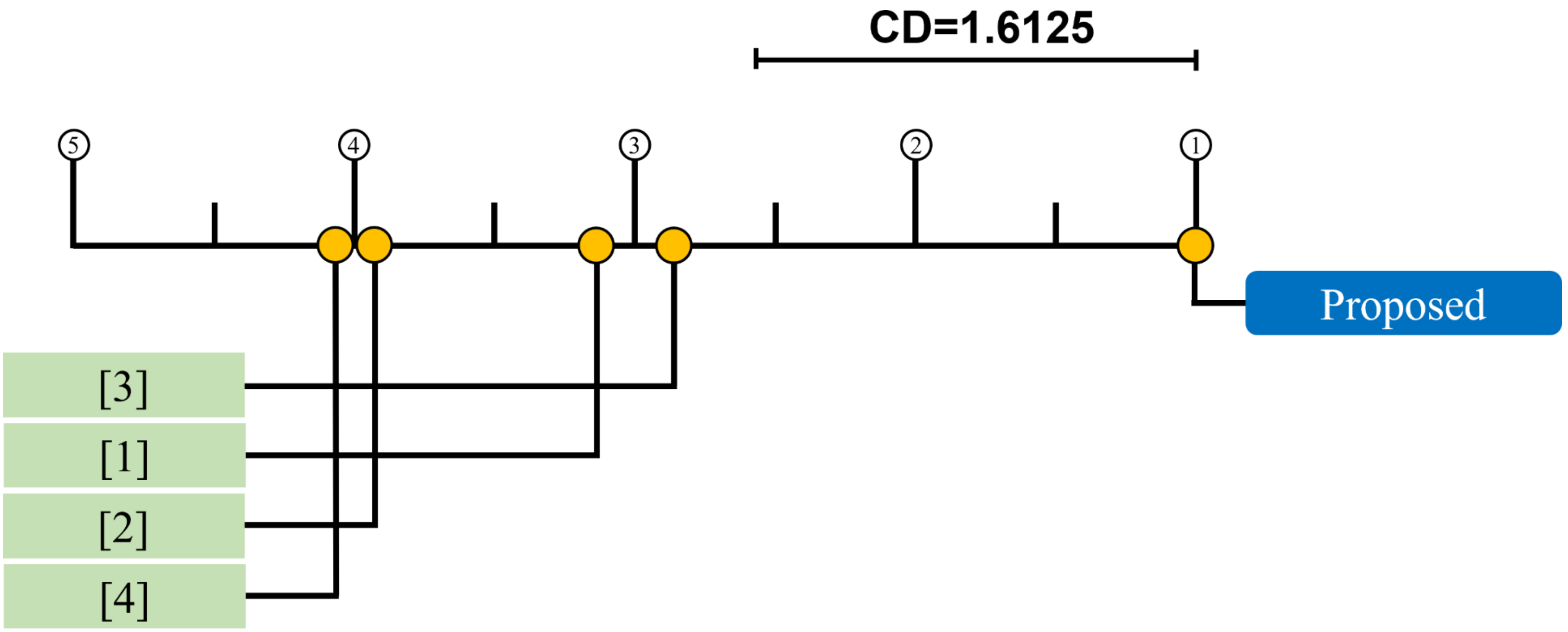
Hyper Parameter Settings

Model	Batch Size	Epochs	Loss Function	Optimizer	LR	Weight Decay
Proposed	16	60	Cross Entropy Loss	Adam	1e-3	1e-4
[1]	64		Proposed Loss Function			
[2]	128		Cross Entropy Loss			
[3]	32		Cross Entropy Loss			
[4]	32		Cross Entropy Loss			

Model Comparison Based on Accuracy

Dataset (Genre)	Proposed	[1]	[2]	[3]	[4]
Ballroom	**56.79 ± 3.77**	45.71 ± 2.86▼	42.43 ± 2.29▼	48.07 ± 4.79▼	47.71 ± 4.58▼
Ballroom-Extended	**71.87 ± 2.28**	61.82 ± 3.79▼	44.68 ± 1.30▼	62.88 ± 4.70▼	54.38 ± 5.33▼
emoMusic-G	**34.16 ± 2.20**	27.72 ± 3.50▼	28.99 ± 1.78▼	32.95 ± 2.92	31.34 ± 3.66▼
Emotify-G	**72.25 ± 2.99**	62.78 ± 2.70▼	67.50 ± 4.17▼	66.00 ± 3.32▼	65.88 ± 5.62▼
FMA-SMALL	**55.98 ± 1.10**	47.03 ± 1.31▼	42.19 ± 1.16▼	46.96 ± 1.62▼	40.88 ± 1.73▼
GiantStepsKey-G	**34.13 ± 2.61**	28.43 ± 3.15▼	25.21 ± 2.53▼	27.19 ± 2.90▼	24.30 ± 2.11▼
GMD-G	**66.36 ± 3.49**	61.91 ± 3.04▼	63.44 ± 1.71▼	39.38 ± 3.36▼	35.36 ± 6.35▼
GTZAN	**80.25 ± 1.98**	77.50 ± 3.81	63.35 ± 2.08▼	75.10 ± 3.84▼	72.50 ± 2.13▼
HOMBURG	**55.34 ± 1.75**	53.49 ± 1.59▼	51.01 ± 1.51▼	52.54 ± 2.55▼	48.15 ± 2.14▼
ISMIR04	**79.29 ± 1.75**	70.96 ± 2.93▼	67.88 ± 1.70▼	71.32 ± 1.37▼	67.67 ± 1.99▼
MICM	**41.36 ± 2.39**	37.81 ± 3.32▼	39.30 ± 2.21▼	38.42 ± 3.16▼	39.34 ± 2.28▼
Seyerlehner-Unique	**70.24 ± 1.09**	63.62 ± 2.56▼	60.72 ± 2.73▼	63.50 ± 2.79▼	61.62 ± 1.70▼
Win / Tie / Lose	**0/ 0/ 0**	**0/ 0/ 0**	**0/ 0/ 0**	**0/ 0/ 0**	**0/ 0/ 0**
Avg . Rank	**1.00**	3.17	3.92	2.83	4.08

Bonferroni-Dunn test



Model Evaluation with Other Features (GTZAN)

Iteration	Proposed	STFT	Mel-spectrogram	MFCC	STFT + Mel-spectrogram	STFT + MFCC	Mel-spectrogram + MFCC
1	81.00	72.50	76.50	64.00			77.00
2	79.50	70.50	79.00	67.00			76.00
3	78.50	74.50	78.50	59.00			73.50
4	81.00	76.00	80.50	61.50			75.50
5	77.00	72.00	76.00	60.00			76.50
6	81.00	73.00	76.50	66.00			77.00
7	81.00	75.50	74.00	68.50			73.00
8	82.00	76.00	80.50	62.00			81.50
9	78.00	79.00	80.00	59.00			77.50
10	83.50	75.50	78.00	62.00			75.00
Avg . Accuracy	80.25	74.4	77.95	62.9			76.25

Conclusion

Contribution of the Study

1. Conducted a novel CNN model with three pipelines.
2. Achieve higher accuracy than conventional methods using a single input.

Future Work

1. Combining new features that can improve accuracy.
2. Ensemble with other models.

Individual Study

Su-jeong Han

2022. 09. 14

Overall Progress

AutoML

Problem

- Hard-to distinguish pixel boundaries
- Wrong position Detection
- Low Accuracy of Segmentation

Related Works

- U-Net
- Deeplab V3+
- HarDNet MSEG
- H-Dense U-Net
- Self-supervised

Experiments

- Dataset
1237 patients
(AX7 내직근, AX7 외직근)

Analysis

- Wrong location
- Use U-Net
(best performance)

Proposal

- Hybrid Loss
- Gaussian Distribution

Post-experimental analysis

Conclusion

Overall Progress

Problem

Related Works

- U-Net
- Deeplab V3+
- 3D U-Net
- H-Dense U-Net
- BAS-Net

Experiments

- Dataset
1237 patients

Analysis

Proposal

Post-experimental analysis

Conclusion

Eyeball Volume Estimation

AutoML

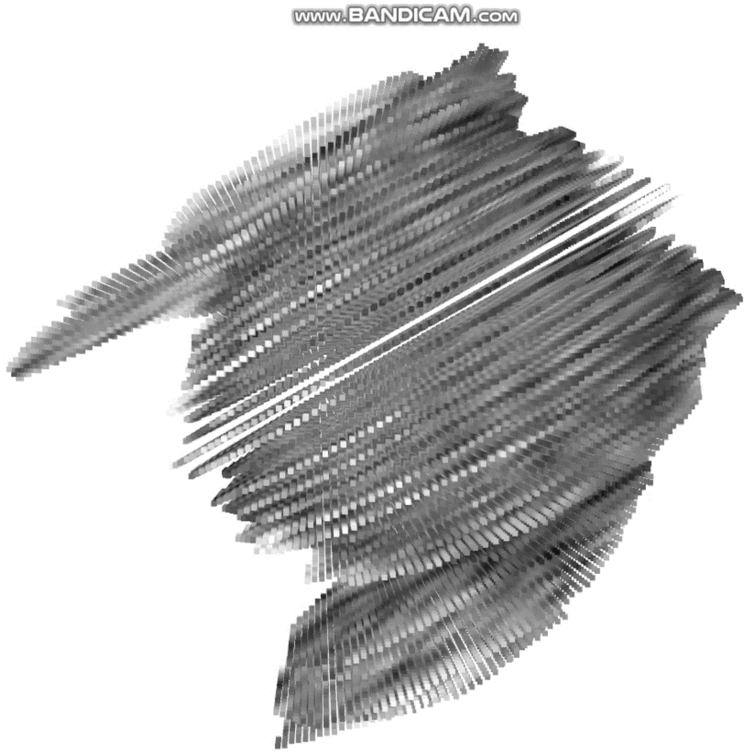
- Datasets
  - 5 images from Ax1 to Ax5 for each patients
  - Number of Samples: 3,505 images (701 patients)
  - Input : 5 eyeball segmentation images from Ax1 to Ax5 for each patients
  - Output : Eyeball Volume

- Result
  - 345 patients

Patient Counts	Eye volume(model.ver) mean	Eye volume(model.ver) mean	Error	Correlation
345	8.203 cc	7.962 cc	±0.242	0.311

3D Segmentation

1075 환자 수동 세그멘테이션 시각화



101026 환자 자동 세그멘테이션 시각화

갑상선 연구 데이터셋 구축 현황

- 전체 진행 상황

13,607 (최종 검수 완료)/ **9,633 (자체 검수 완료)**/543,412 (받은 전체 슬라이스 개수)

자체검수 진행사항 : **96,330 (완료)** / 96,330(전체)

검수 완료 : **내직근_left, 외직근_left**

Individual Studies

Taekyung Song

2022. 09. 14

Overall Progress

AutoML

Problem → Related Work / Survey (Counterparts) → Experiments → Analysis → Proposal → Experiments

Problem

- Semantic Segmentation Accuracy

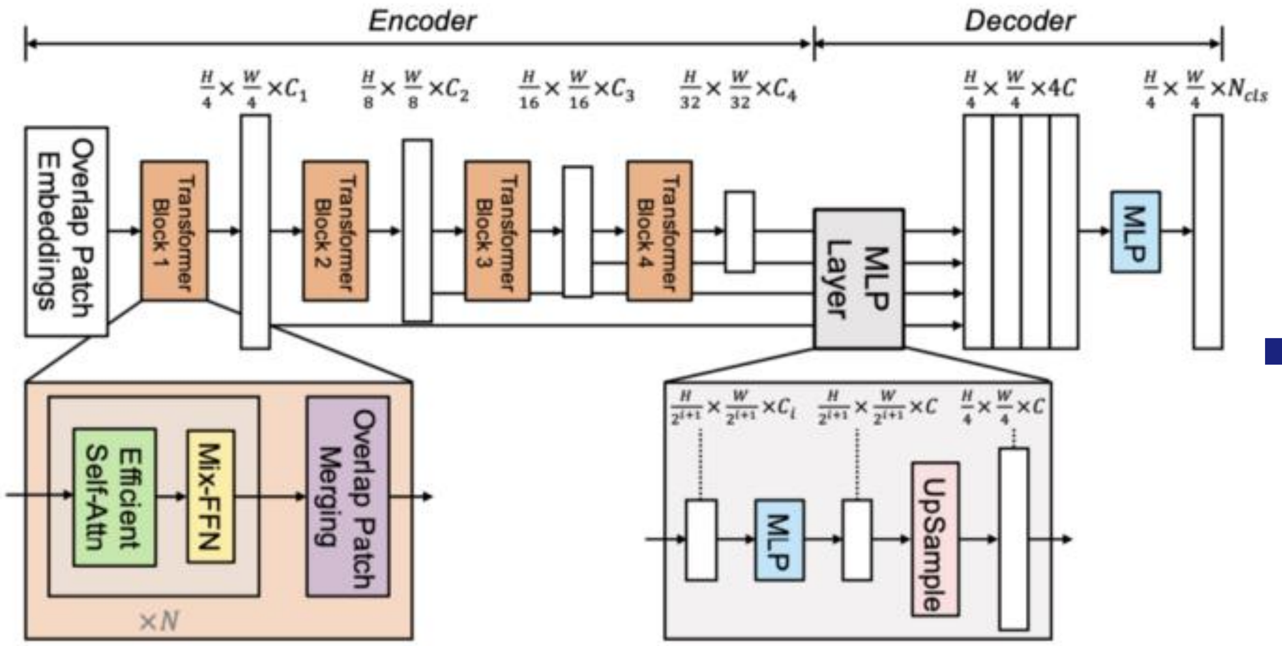
Related Work / Survey (Counterparts)

Experiments

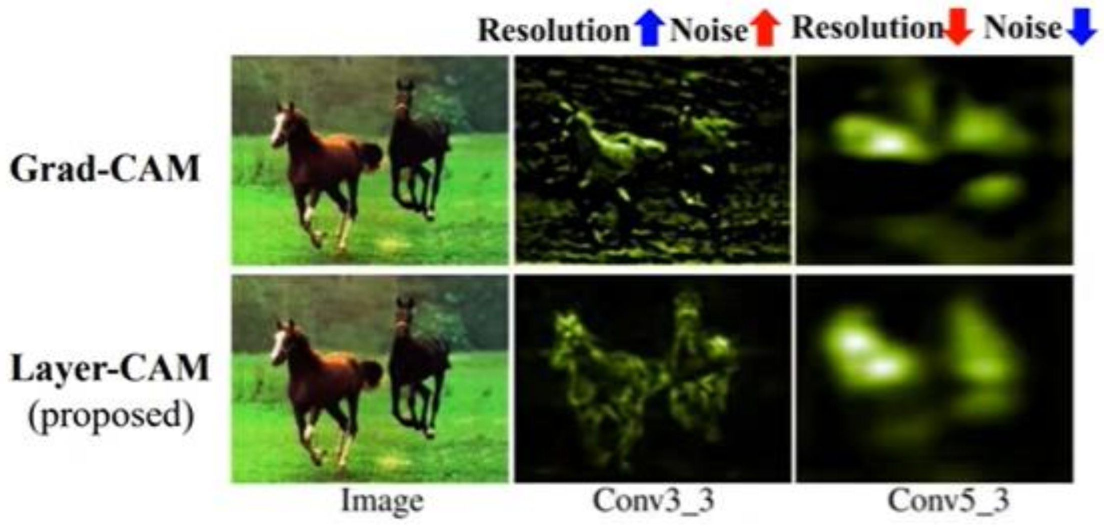
Analysis

Proposal

Segformer & Layer-CAM



< Segformer Architecture >

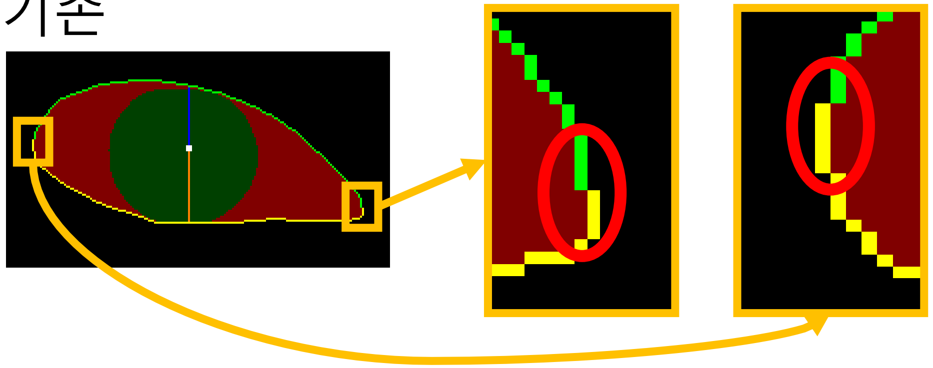


< Layer-CAM >

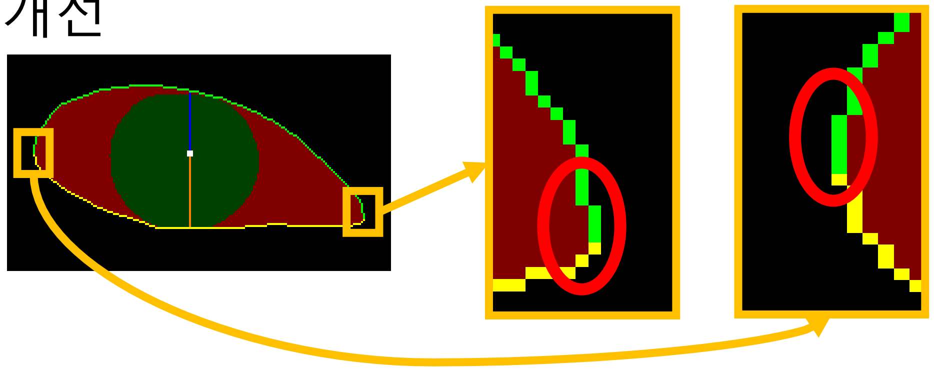
Xie, Enze, et al. "SegFormer: Simple and efficient design for semantic segmentation with transformers." Advances in Neural Information Processing Systems 34 (2021): 12077-12090.
Jiang, Peng-Tao, et al. "Layercam: Exploring hierarchical class activation maps for localization." IEEE Transactions on Image Processing 30 (2021): 5875-5888.

길이 측정 알고리즘 고도화

- 기존



- 개선



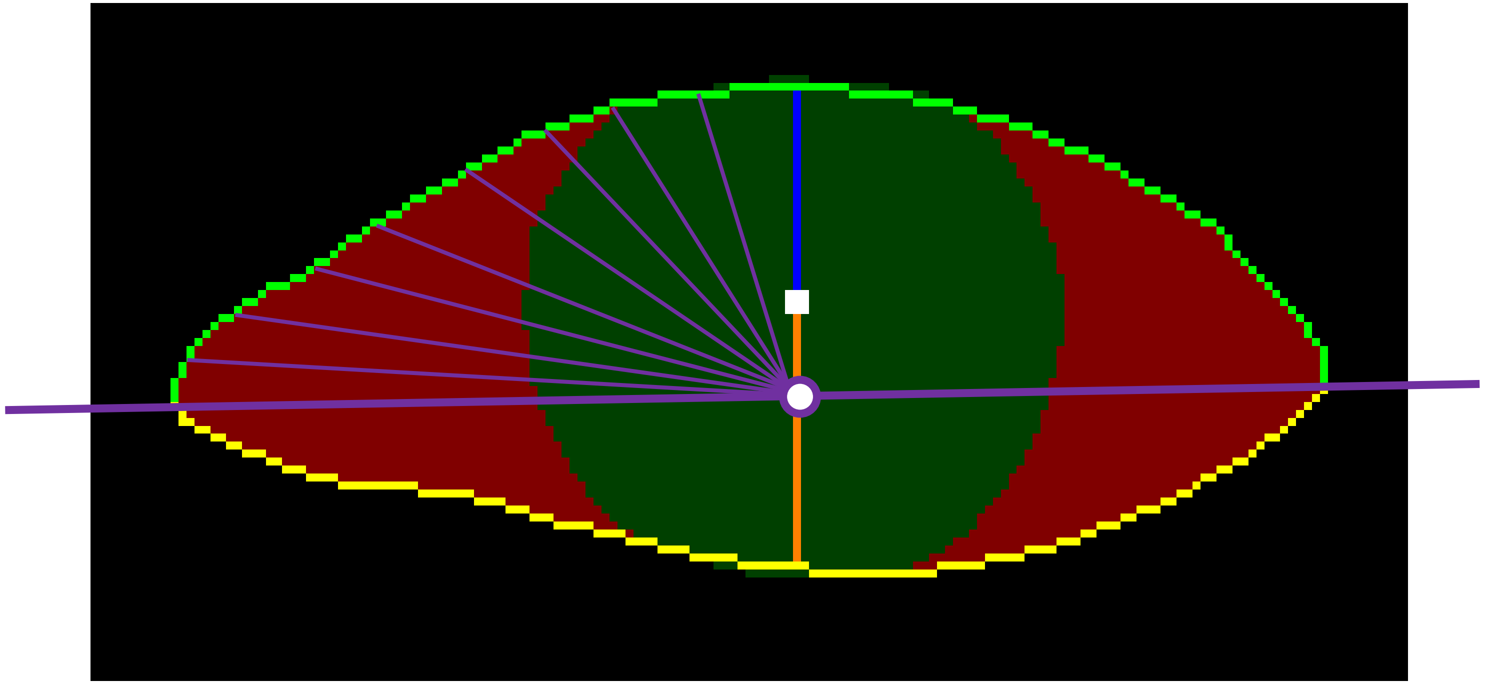
MAE	Original	1차 개선	2차 개선
R_upper_lid	5.39	5.97	1.60
R_lower_lid	3.83	3.94	2.11
L_upper_lid	5.52	6.06	1.76
L_lower_lid	3.64	3.71	1.83

단위: mm

Chun, Yeoun Sook, et al. "Topographic analysis of eyelid position using digital image processing software." *Acta ophthalmologica* 95.7 (2017): e625-e632.

Next Progress

- 산학연4 S/W 등록
- 실제 길이 측정에 사용된 데이터셋 요청
- New Metric 구현

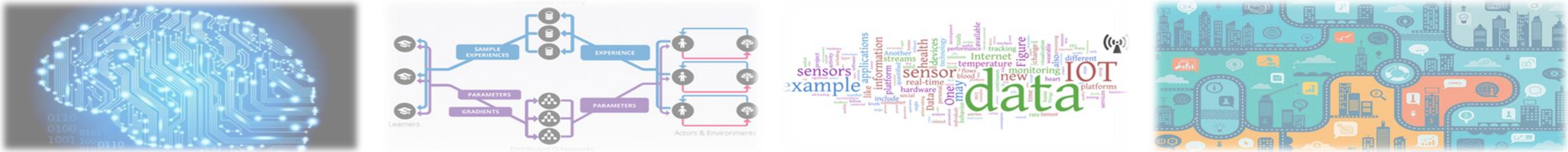


References

[1] Xie, Enze, et al. "SegFormer: Simple and efficient design for semantic segmentation with transformers." *Advances in Neural Information Processing Systems* 34 (2021): 12077-12090.

[2] Jiang, Peng-Tao, et al. "Layercam: Exploring hierarchical class activation maps for localization." *IEEE Transactions on Image Processing* 30 (2021): 5875-5888.

[3] Chun, Yeoun Sook, et al. "Topographic analysis of eyelid position using digital image processing software." *Acta ophthalmologica* 95.7 (2017): e625-e632.



기계학습자동화 연구실
주간 정기 세미나

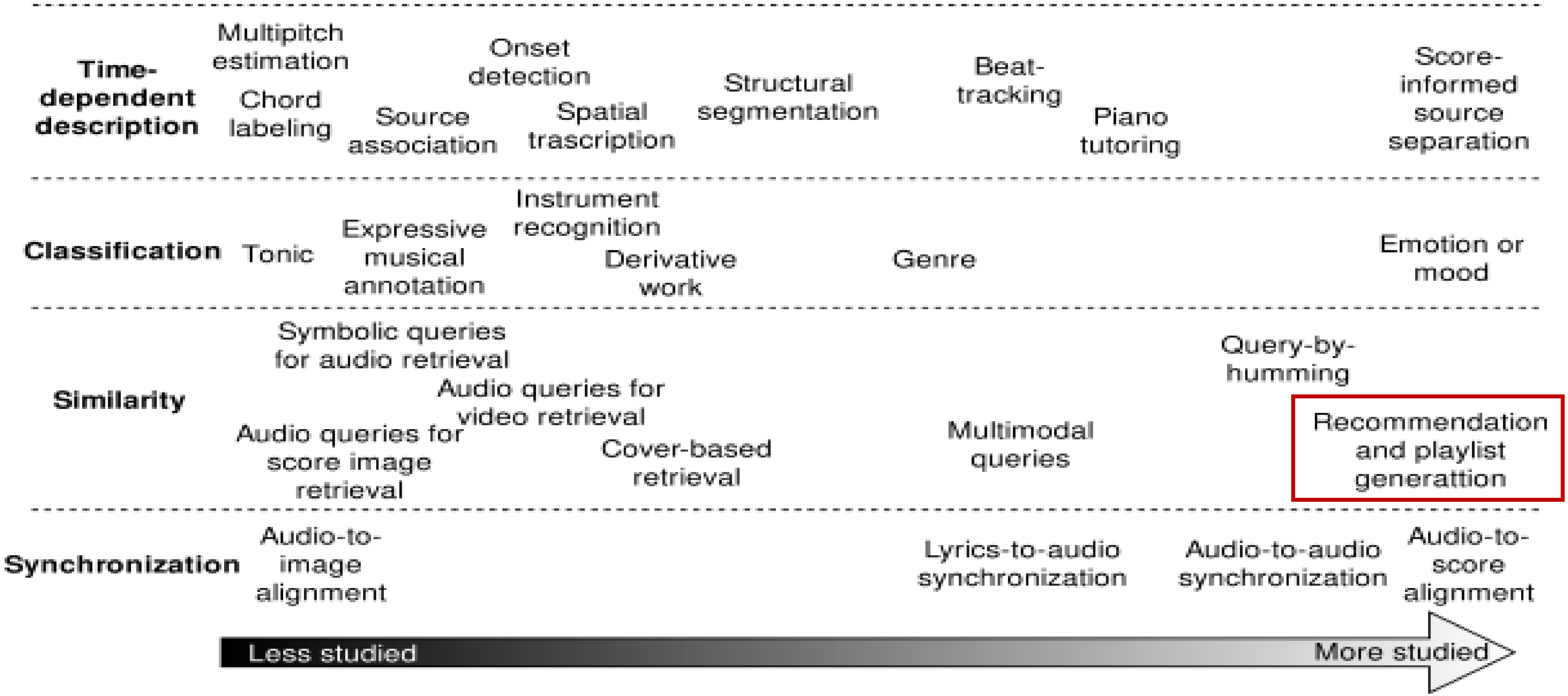
진상형

2022.09.14

Introduction

AutoML

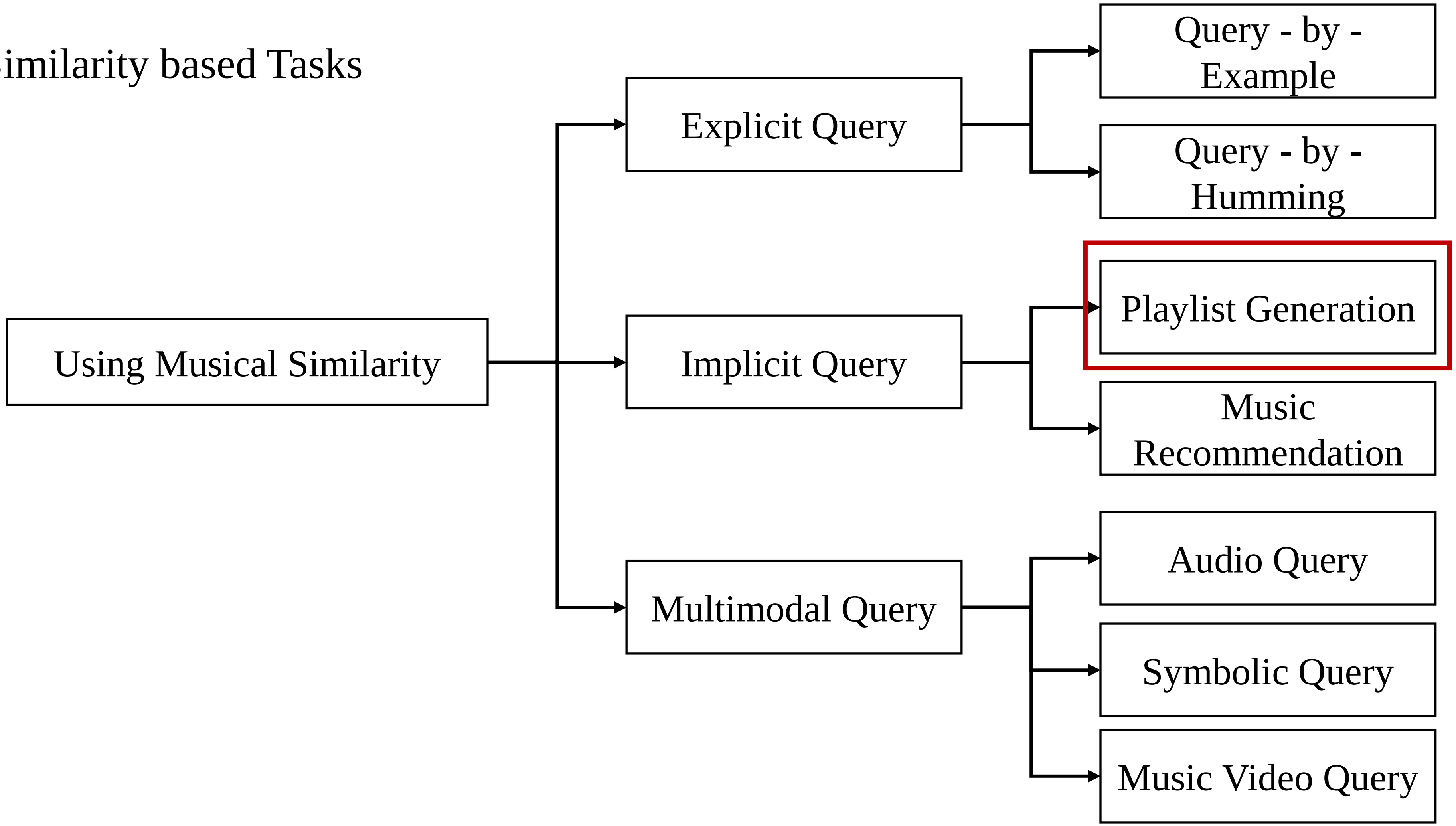
Music Information Retrieval



Introduction

Music Information Retrieval

- Musical Similarity based Tasks

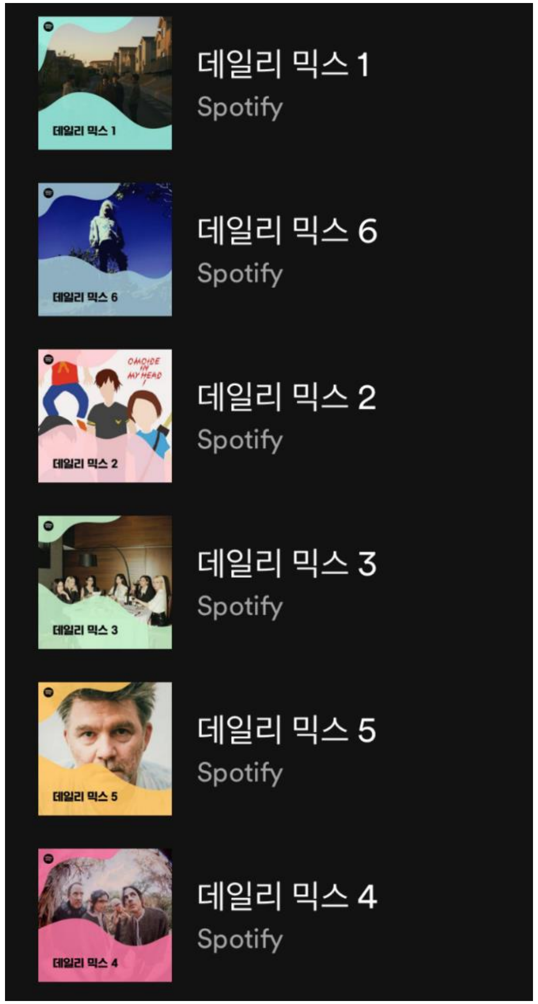


Introduction

Playlist Generation

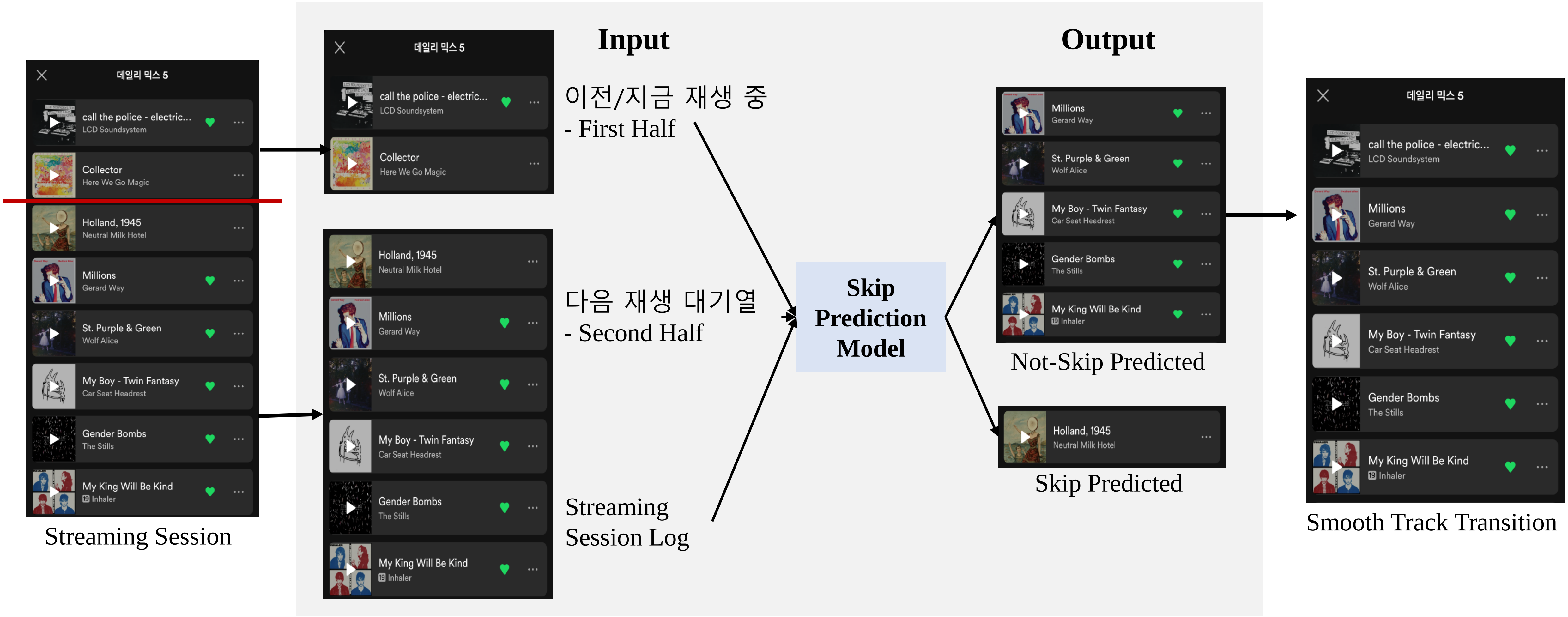
Playlist Generation

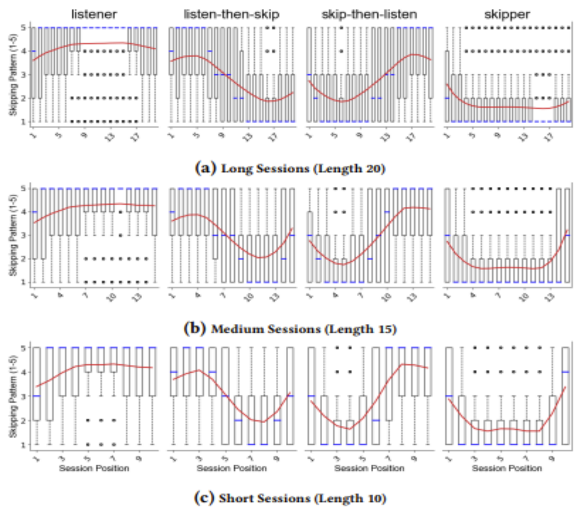
- Reinforcement Learning
- Frequent Pattern Mining
- Collaborative Filtering
- Statistical Model (Markov Model)
- ⋮

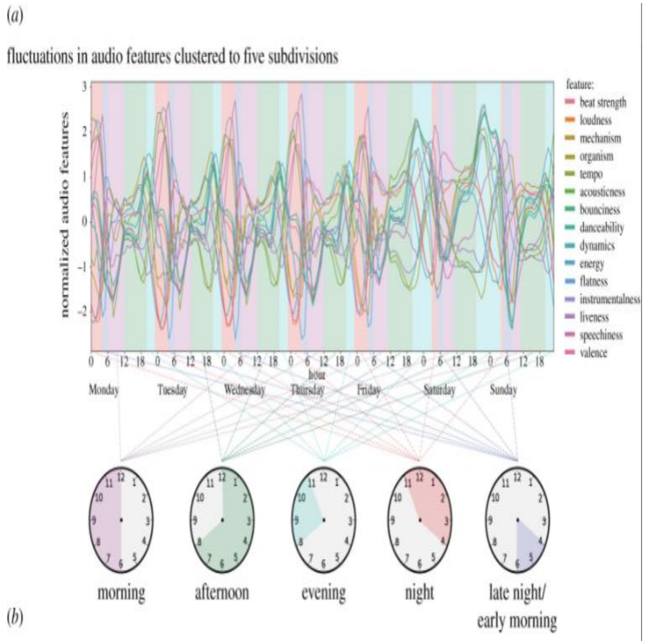


Introduction

Playlist Generation – Process of Skip Prediction







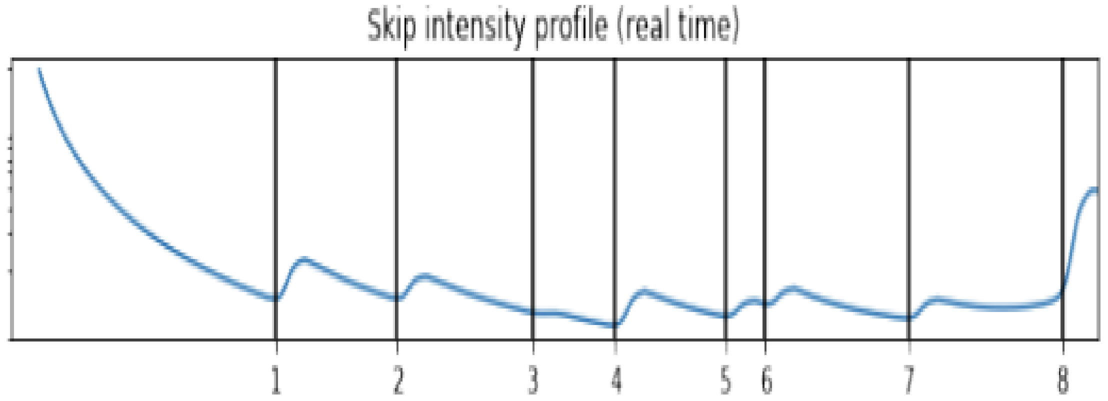
Track Skip Behavior

- Skip 행동 유형은 크게 4가지로 정의할 수 있음

: listener / litsten then skip / skip then listen / skipper

- 오디오 특징의 변화는 일주일 동안 항상 일관된 흐름을 보임

: 하루 동안의 음악적 선호도는 주기적으로, 예측 가능하게 변화함



- 어떤 곡에 대한 스킵 비율은 일반적으로 U자형 패턴을 가짐

: 곡의 초반과 마지막에서 스킵 비율이 높음

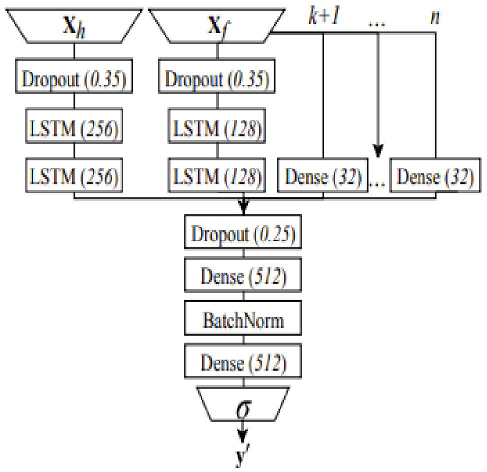
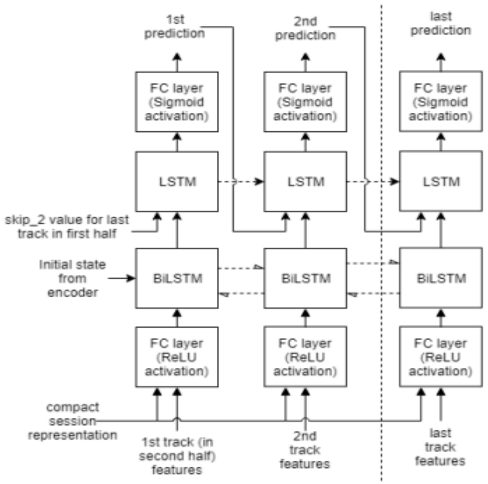
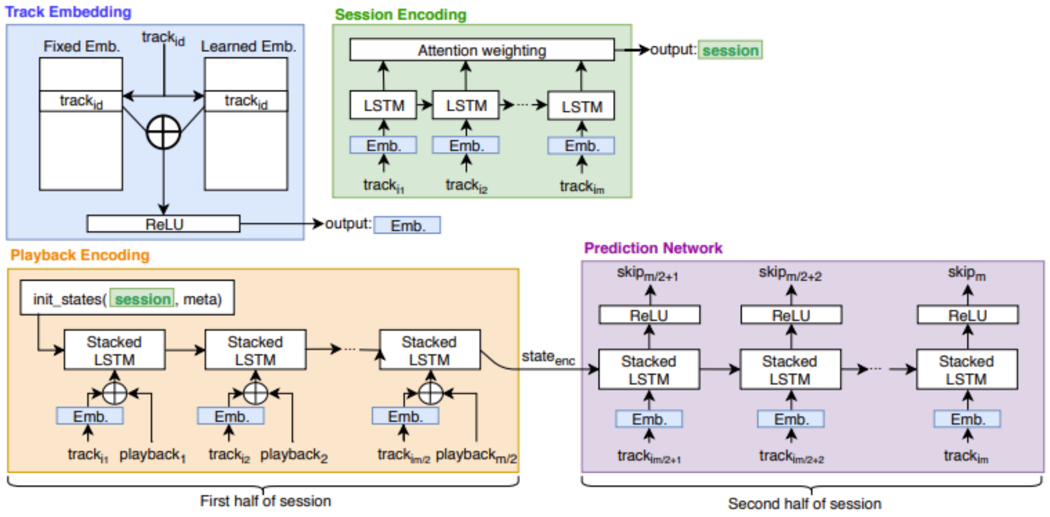


Figure 3: Decoder architecture



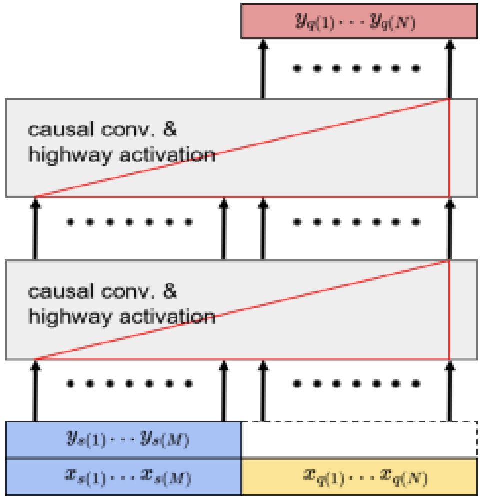


Skip Prediction

: 일반적으로 RNN을 기반으로 사용

- RNN으로 세션을 모델링하고 곡 특징은 Dense를 통과시켜 Skip 예측
- BiLSTM을 기반으로 한 Encoder-Decoder 구조를 사용해 Skip 예측
- Stacked Multi-RNN을 이용하여 Skip 예측

Related Work



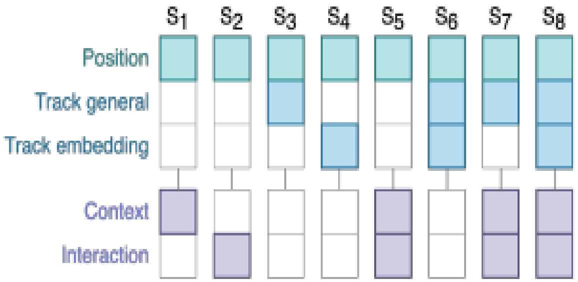


Fig. 7. ***Spotify interpretation subsets*** For each subset S_i, the feature groups we include (*i.e.* X_j) are coloured. The groups of input features drawn in purple are not available in B.

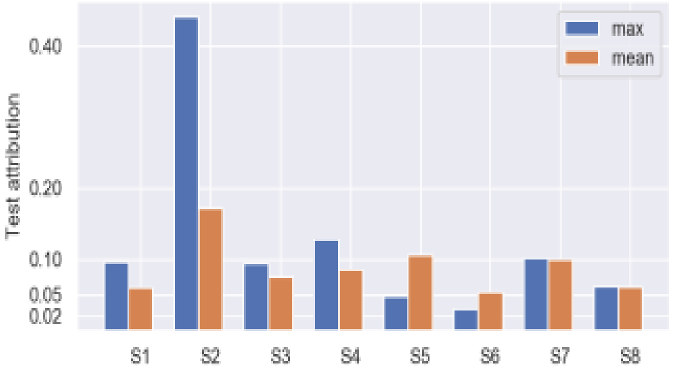


Fig. 8. ***Spotify attribution*** Mean attribution value and mean item-wise max skip attribution over 20′000 test sessions.

Skip Prediction

: CNN이나 Transformer를 기반으로 사용하기도 함

- Dilated 1D Convolution과 Highway Network를 조합
- Interpretable Transformer를 이용한 Skip 예측 및 설명

Related Work

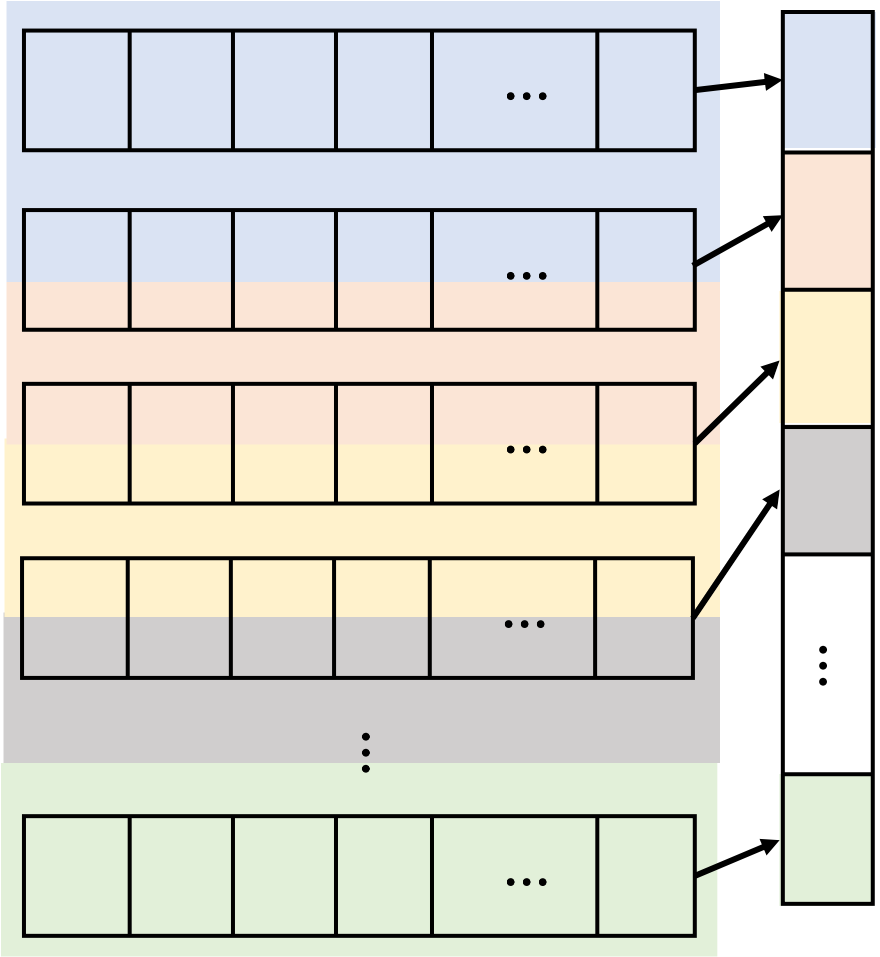
Residual Connection / Highway Network

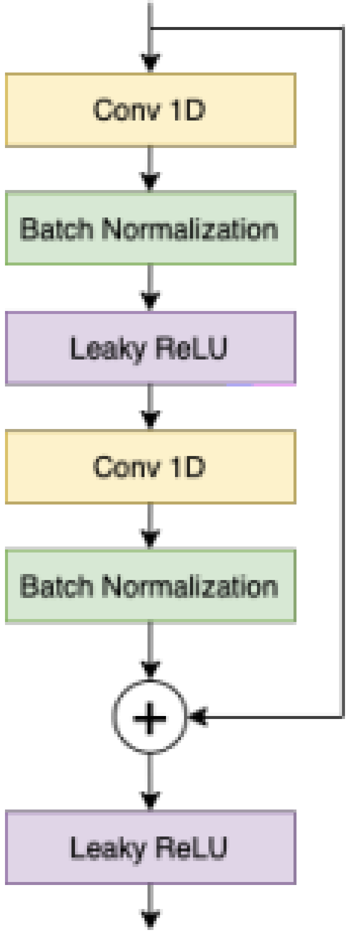
: 재추정 과정에서 이전 값에 가깝게 유지하는 것이 값이 크게 바뀌는 것보다 유리
-> 전반적인 특징을 유지하면서 개별 특징은 능동적으로 학습하기 위함
-> Resnet과 Highway 모두 이전 입력을 유지하는 점은 동일하지만, 이전 입력 변환 과정의 차이가 있음

Variant	Functional Form
Plain	$H(\mathbf{x})$
Residual	$H(\mathbf{x}) + x$
T-Only	$H(\mathbf{x}) \cdot T(\mathbf{x}) + \mathbf{x}$
C-Only	$H(\mathbf{x}) + \mathbf{x} \cdot C(\mathbf{x})$
Coupled	$H(\mathbf{x}) \cdot T(\mathbf{x}) + \mathbf{x} \cdot (1 - T(\mathbf{x}))$
Full	$H(\mathbf{x}) \cdot T(\mathbf{x}) + \mathbf{x} \cdot C(\mathbf{x})$

Residual Connection (ResNet) : identity 연결을 통해 이전 값을 전부 그대로 유지
Highway Network : 어느 한쪽 gate의 비중이 커지면 반대 gate의 비중이 작아져 이전 값이 완전하게 전달되지 않음

Related Work





1D Convolution

- 기존의 CNN은 이미지와 같은 2D 형태의 데이터에서 주로 사용

- 1D CNN을 이용하여 이미지가 아닌 신호, 오디오 등의 시퀀스 데이터에서 RNN보다 빠르면서 비슷한 효과를 얻음

1D ResNet

- 기존의 ResNet 구조에 1D CNN을 적용하여 1D CNN 2개, Batch Normalization, Identity Shortcut으로 1D ResNet 구성

Figure 1. A canonical Inception module (Inception V3).

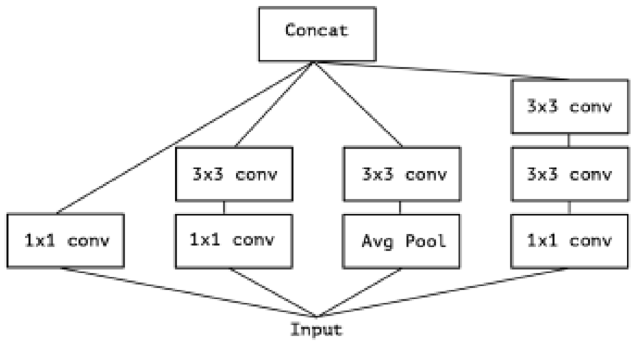
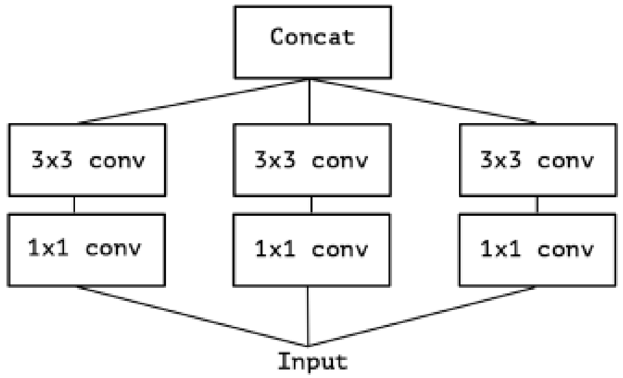
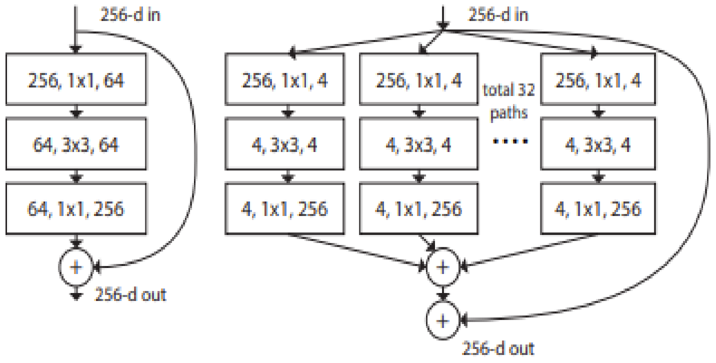


Figure 2. A simplified Inception module.





Multi-Branch Convolution

: Inception, Xception, ResNext 등의 이미지 모델에서 correlation이 높은 그룹끼리 묶어 Convolution하도록 유도해 최적의 학습을 하는 방식

- Inception의 경우 Cross-channel correlation과 Spatial Correlation을 분리
- Xception은 Inception을 확장하여 Cross-channel / Spatial Correlation을 완전히 독립적으로 분리
- ResNeXt의 경우 cardinality라는 개념을 사용하여 동일 크기의 그룹으로 분리하는 Grouped Convolution 사용

Related Work

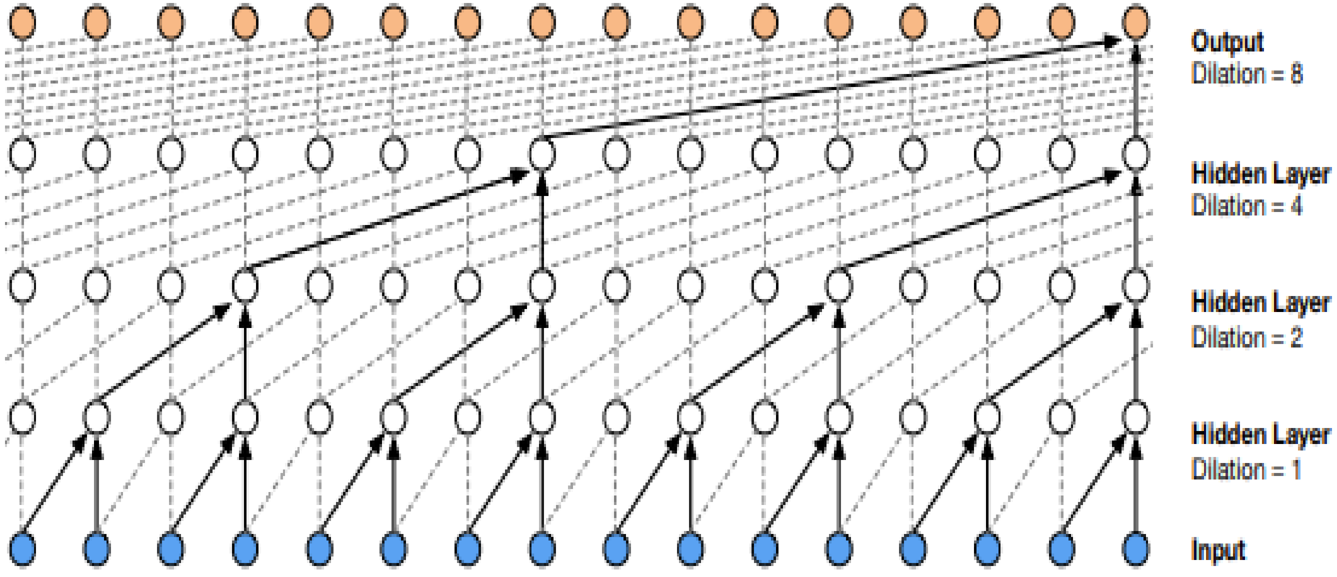


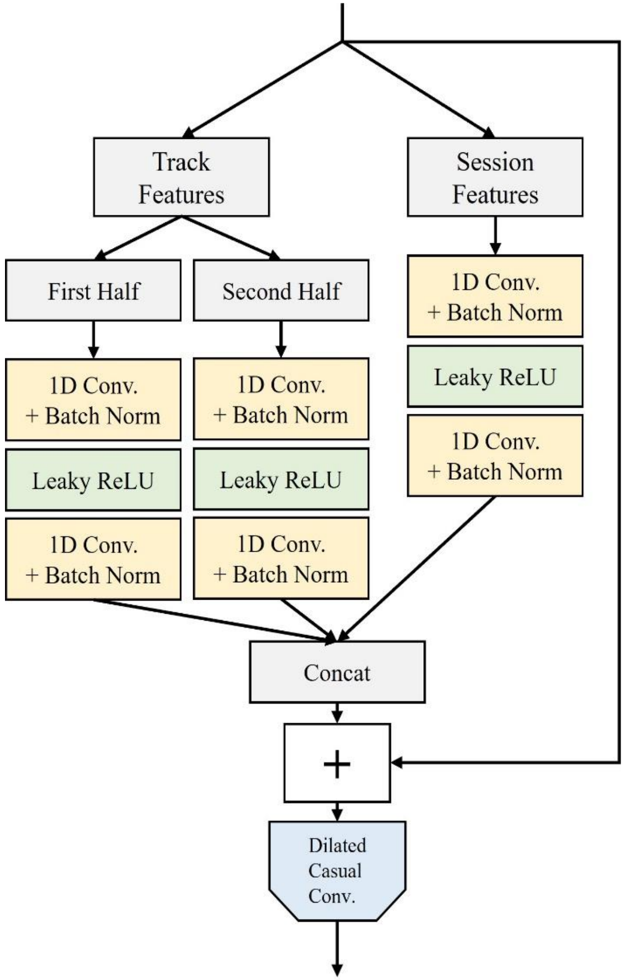
Figure 3: Visualization of a stack of *dilated* causal convolutional layers.

Dilated Casual Convolution

: Wavenet에서 제안, 다음 시점의 시퀀스를 생성하기 위해 Convolution layer를 이용

1. Casual Convolution
: 현재 input이 과거 데이터의 시간적 순서를 따르도록 강제
- Recurrent connection이 없어 RNN보다 빠름

2. Dilated Convolution
: 필터가 입력을 건너뛰어 원래보다 더 큰 영역에 적용



Muti-Branch 1D ResNet

: 전체적인 구조는 기존 2D ResNet에서 2D Convolution을 1D Convolution으로 변경

- ResNet의 Residual Connection은 재추정을 용이하게 하여 모델의 학습 효과를 높임

- 1D Convolution, Batch Normalization, Leaky ReLU로 내부 블록 구성

- Inception등과 같은 분기 구조를 통해 convolution을 분리하여 각각 학습하고 하나로 Aggregation하여 각각의 분기는 비슷한 Correlation을 가진 특징들을 학습
-> 특히 Skip 문제에서 사용하는 MSSD 데이터는 Tabular 형태로 이미지 데이터와 달리 명시적인 특징 분리가 가능

- 1D ResNet output에 대해 Dilated Casual Convolution을 추가로 적용하여 세션 내의 시간적인 특징 반영

Experimental Settings - Measures

AutoML

Ground truth sequence	Predicted sequence	AA
1, 1, 1, 1, 1	0, 1, 1, 1, 1	0.543
1, 1, 1, 1, 1	1, 0, 1, 1, 1	0.643
1, 1, 1, 1, 1	1, 1, 0, 1, 1	0.710
1, 1, 1, 1, 1	1, 1, 1, 0, 1	0.760
1, 1, 1, 1, 1	1, 1, 1, 1, 0	0.800

$$\frac{\left(\frac{0}{1}*0+\frac{1}{2}*1+\frac{2}{3}*1+\frac{3}{4}*1+\frac{4}{5}*1\right)}{5}=0.543$$

$$\frac{\left(\frac{1}{1}*1+\frac{2}{2}*1+\frac{3}{3}*1+\frac{3}{4}*0+\frac{4}{5}*1\right)}{5}=0.760$$

MAA (Mean Average Accuracy)

- Skip 문제와 관련하여 제안된 평가 방법
- Skip 순서에 있어 첫번째 예측의 정확도가 가장 중요 (First Prediction Accuracy - FPA)
- 후반부 순서의 Skip 결과는 예측 정확도에 대해 낮은 영향력

T : 예측되는 전체 트랙의 개수
A(i) : i번째 트랙까지의 정확도
L(i) : i번째 정답 여부에 대한 Boolean indicator (0, 1)

$$AA=\frac{\sum_{i=1}^{T}A(i)L(i)}{T}$$

Experimental Settings

- Datasets
  - Music Streaming Sessions Dataset (MSSD)
  - Number of Samples: 167,880 session histories (10,000 unique session)
  - Number of Classes: 2 (Skip / Not Skip)
  - Input : (80, 20) - (Session & Track Features, Sequence Length) / 20보다 작은 Sequence는 zero padding 포함
  - Output : (10, 1) - 후반부 10개 트랙에 대한 예측 확률 값
- Cross-Validation for Test Performance
  - 10 iterations / Group K-Fold (by Session id)
  - Data Split Ratio: (Train : Validation : Test) = (0.8 : 0.1 : 0.1)
- Hyper Parameter Setting:

Model	Batch Size	Epochs	Loss Function	Optimizer	LR
Proposed	16	10	Binary Cross Entropy	Adam	1e-4
1D Conv. + Highway					
Encoder-Decoder				RMSprop	
Multi-RNN				Adam	1e-3
RNN + Dense					1e-4

Experimental Results

Model	MAA	MAA P-value
Proposed	**0.6877 ± 0.2537**	**-**
1D Conv. + Highway [12]	0.5774 ± 0.2854	< 0.0001
Encoder-Decoder [16]	0.5303 ± 0.2904	< 0.0001
Multi-RNN [14]	0.6061 ± 0.3015	< 0.0001
RNN + Dense [15]	0.5832 ± 0.2193	< 0.0001

Conclusion

Conclusion

-1D ResNet 구조를 기반으로 데이터 특징에 맞게 Convolution을 분기한 뒤 다시 하나로 Aggregation 하는 방식을 제안

- 각 변형 구조들과의 비교에서 이전 입력을 그대로 유지하고 Dilated Casual Convolution을 적용한 Proposed 방식이 더 효과적인 것으로 나타남

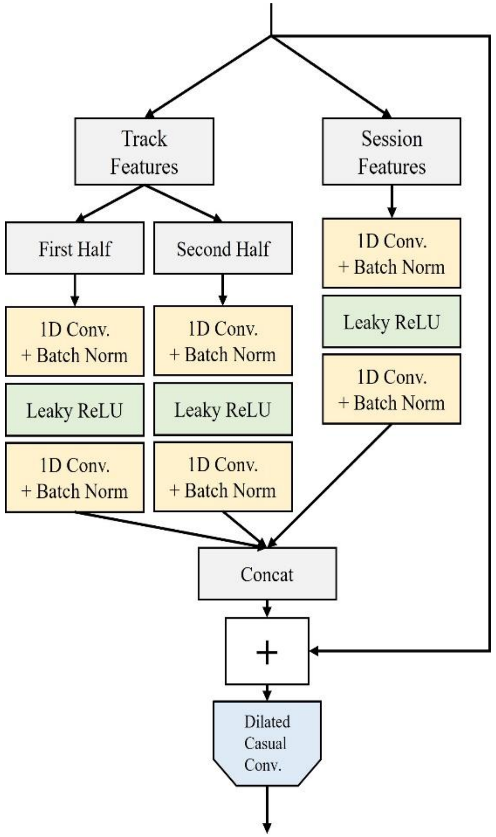
- 기존에 주로 사용되던 RNN 기반의 모델들과의 비교에서 Proposed 방식이 훨씬 더 좋은 성능을 보임

Limitations

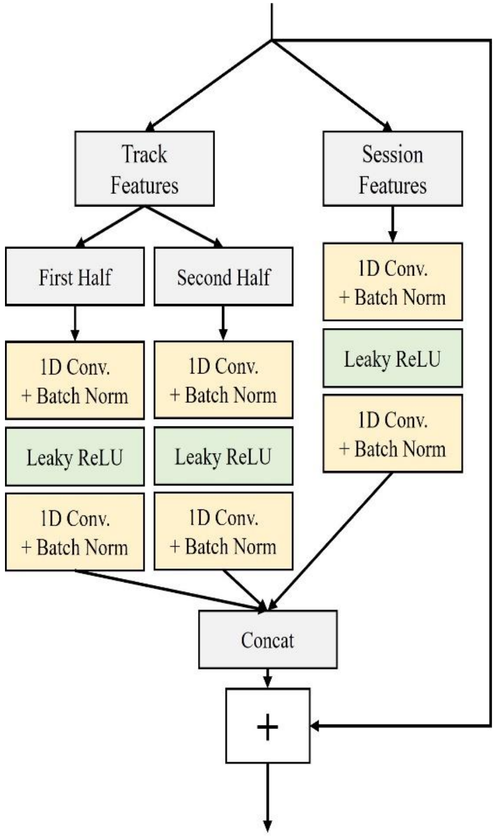
- 현재 사용되는 MSSD 데이터의 구조, 특징에 지나치게 맞춰져 있음
- 다른 데이터에 대한 일반화의 어려움이 있을 것으로 보임

Appendix – Model Comparison

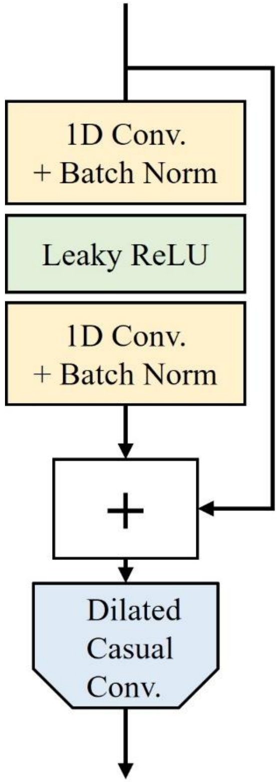
*DCC : Dilated Casual Convolution



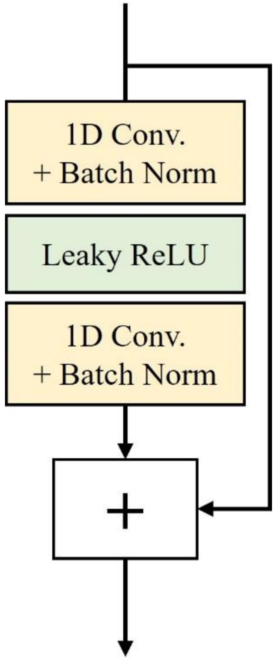
1. Proposed



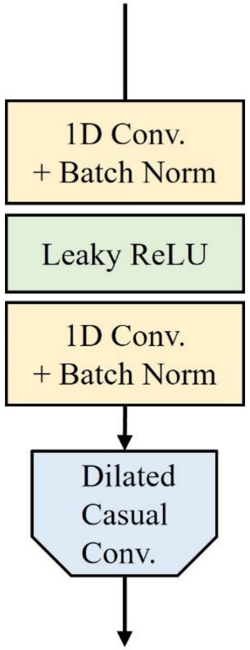
2. Multi-branch 1D ResNet



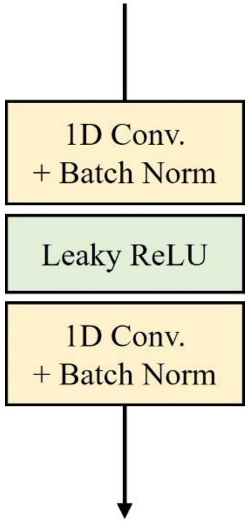
3. Basic 1D ResNet + DCC



4. Basic 1D ResNet



5. Residual Conn. Removed + DCC



6. Residual Conn. Removed

Appendix – Model Comparison Result

	Model	MAA	MAA P-value
1	**Proposed**	**0.6877 ± 0.2537**	**-**
2	Multi-branch 1D ResNet	0.6826 ± 0.2653	0.6600
3	Basic1D ResNet + DCC	0.6458 ± 0.2653	0.0003
4	Basic 1D ResNet	0.6343 ± 0.2583	< 0.0001
5	Residual Connections Removed + DCC	0.4503 ± 0.3101	< 0.0001
6	Residual Connections Removed	0.4149 ± 0.3372	< 0.0001

*DCC : Dilated Casual Convolution

References

[1] Brian Brost, Rishabh Mehrotra, Tristan Jehan. “The Music Streaming Sessions Dataset”, In Proceedings of the ACM Web Conference. 2019.

[2] Aaron Ng et al. "Investigating the Impact of Audio States & Transitions for Track Sequencing in Music Streaming Sessions." 14th ACM Conference on Recommender Systems. 2020.

[3] Ole Adrian Heggli et al. "Diurnal fluctuations in musical preference." Royal Society open science 8.11. 2021.

[4] Jonathan Donier, “The universality of skipping behaviours on music streaming platforms”, arXiv preprint arXiv:2005.06987, 2020.

[5] Serkan Kiranyaz et al. "1D convolutional neural networks and applications: A survey." Mechanical systems and signal processing 151. 2021.

[6] Federico Simonetta et al. "Multimodal music information processing and retrieval: Survey and future challenges." 2019 international workshop on multilayer music representation and processing (MMRP). IEEE, 2019.

[7] Francesco Meggetto, et al. "On skipping behaviour types in music streaming sessions." Proceedings of the 30th ACM International Conference on Information & Knowledge Management. 2021.

[8] Christian Szegedy et al. "Going deeper with convolutions." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.

[9] Francois Chollet. "Xception: Deep learning with depthwise separable convolutions." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

[10] Saining Xie et al. "Aggregated residual transformations for deep neural networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

[11] Casper Hansen et al. "Contextual and sequential user embeddings for large-scale music recommendation.“14th ACM conference on recommender systems. 2020.

[12] Klaus Greff, Rupesh K. Srivastava, Jurgen Schmidhuber. “Highway and Residual Networks learn Unrolled Iterative Estimation”, ICLR 2017 Poster. 2017.

[14]] A. Oord, S. Dieleman, H. Zen, et al. “Wavenet: A generative model for raw audio”. arXiv:1609.03499, 2016.

[15] Sakurai, Keigo, et al. "Music playlist generation based on reinforcement learning using acoustic feature map." 2020 IEEE 9th Global Conference on Consumer Electronics (GCCE). IEEE, 2020.

[16]] Darius Afchar, and Romain Hennequin. "Making neural networks interpretable with attribution: application to implicit signals prediction." 14th ACM Conference on Recommender Systems. 2020.

[17] Christian Hansen, Casper Hansen, et al. “Modelling sequential music track skips using a Multi-RNN approach”, 12th ACM International Conference on Web Search and Data Mining. 2019.

[18] Olivier Jeunen, Bart Goethals. “Predicting Sequential User Behavior with Session-based Recurrent Neural Networks”, 12th ACM International Conference on Web Search and Data Mining. 2019.

[19] Sainath Adapa. “Sequential modeling of Sessions using Recurrent Neural Networks for Skip Prediction”, 12th ACM International Conference on Web Search and Data Mining. 2019.

[20] Geoffray Bonnin, and Dietmar Jannach. "Automated generation of music playlists: Survey and experiments." ACM Computing Surveys (CSUR). 2014

추가 – Dilated Casual Convolution 관련

	Model	MAA	MAA P-value
1	Multi-branch 1D ResNet	0.6826 ± 0.2653	-
2	**Proposed**	**0.6877 ± 0.2537**	**0.6600**
3	1D ResNet + DCC	0.6458 ± 0.2653	0.0019
4	1D ResNet	0.6343 ± 0.2583	< 0.0001
5	Residual Connections Removed + DCC	0.4503 ± 0.3101	< 0.0001
6	Residual Connections Removed	0.4149 ± 0.3372	< 0.0001

*DCC : Dilated Casual Convolution