Individual Studies

Taekyung Song

songteagyong@gmail.com

30th Nov, 2022

Contents

1. ICEIC 논문
2. Individual Study
3. 갑상선 연구 – 눈꺼풀 길이 추론
4. To do list

To Do List from Previous Seminar

Contents

- ✓ Individual Study
  - ➢ Sementic Segmentation using localized information

- ✓ 갑상선 연구 – 눈꺼풀 길이 추론
  - ➢ ImageJ 로 만든 mask와의 비교

ICEIC Overall Progress

Problem

- Inaccurate Eyelid Length Measurement

Related Works/Survey (counterparts)

- Analysis of eyelid position using digital image processing software
- Sementic Segmentation

Experiments

- TAO dataset
- Compare model performance
  - U-Net, SegNet, Deeplab V3+
- Metric
  - mIoU

Analysis

- Single-task Multi-task performance comparison
- Length measurement using segmented images

Proposal

- Automated length measurement using deep learning models

ICEIC

Effective Eyelid Length Measurement Using Semantic Segmentation

Effective Eyelid Length Measurement Using Semantic Segmentation

Taekyung Song
Department of Artificial Intelligence
Chung-Ang University
Seoul, Korea
songteagyong@gmail.com

Jaesung Lee*
Department of Artificial Intelligence
Chung-Ang University
Seoul, Korea
curseor@cau.ac.kr

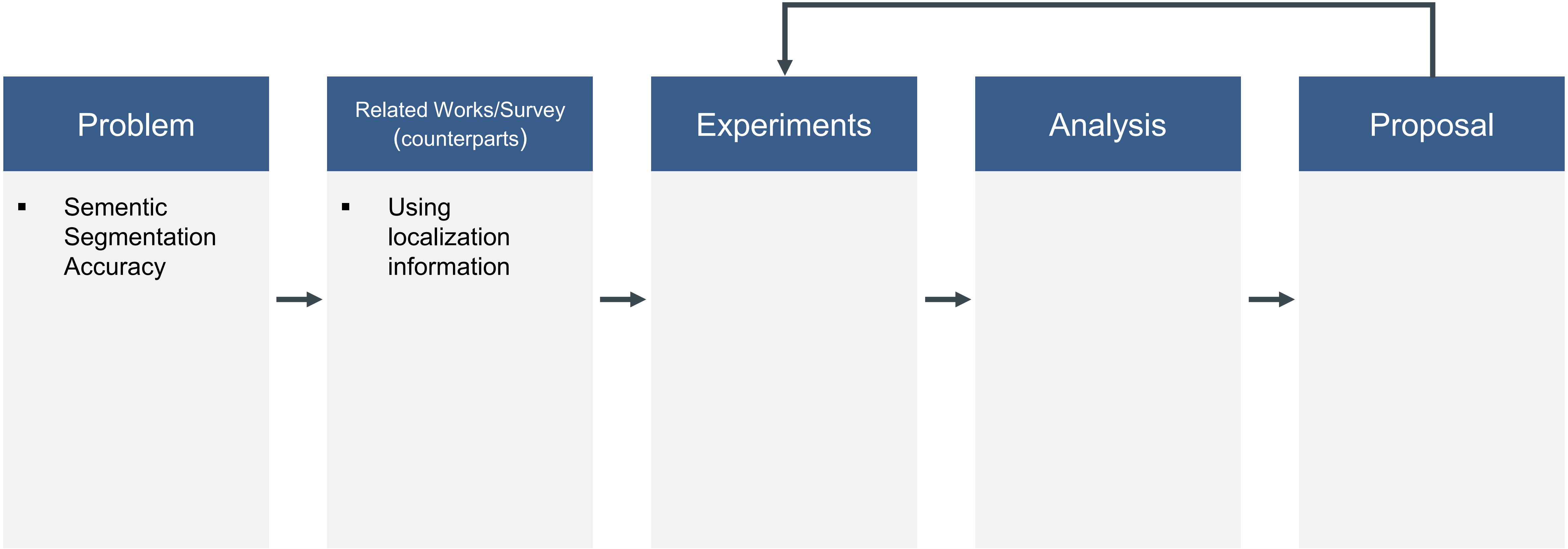
***Abstract*—For patients with Thyroid associated Ophthalmopathy, measuring of eyelid length is essential for disease diagnosis, medical history, and identification of treatment effectiveness. In the previous studies, after segmentation using PC-based commercial software, the length was measured based on the image. However, it is difficult to measure the length accurately because the standards for delineating regions of interest are different for each user. This paper proposes a process of measuring the length of eyelid based on segmented images. The length is measured using the image segmented by the model. Experimental results demonstrated the superiority of the proposed method.**

***Index Terms*—TAO, Semantic Segmentation, Medical Image**

U-shape with left and right symmetry. The left side is called the contraction path and extracts the features of the image. The right side is called the extension path responsible for the decoding process, and restores the convolution. This symmetrical structure is characterized by bringing and combining information on the dilatation path that is symmetrical to the contraction path, enabling more accurate localization.

Segnet [4] was primarily motivated by scene understanding applications. Hence, it is designed to be efficient both in terms of memory and computational time during inference. It is also

Overall Progress



Individual Study

Sementic Segmentation using localized information

Sementic Segmentation using localized information					
no	year	j/c	title	dataset	abstrac
1	2019	*IEEE journal of biomedical a*	Knowledge-aided convolutional neural network for small organ segmentation.		
2	2019	*IEEE/CVF International Conf*	Fine-grained segmentation networks: Self-supervised segmentation for improved long-term visual localization.		
3	2020	*NeuroImage*	An automated framework for localization, segmentation and super-resolution reconstruction of fetal brain MRI.		
4	2020	*IEEE transactions on pattern*	Attention-Based Dropout Layer for Weakly Supervised Single Object Localization and Semantic Segmentation		
5	2021	*Ieee Access*	Real-time polyp detection, localization and segmentation in colonoscopy using deep learning.		
6	2022	*IEEE Transactions on Intellig*	Weakly-Supervised Surface Crack Segmentation by Generating Pseudo-Labels Using Localization With a Classifier and Thresholding		
7	2022	CVPR	Interactive Segmentation and Visualization for Tiny Objects in Multi-Megapixel Images		
8	2022	CVPR	Semantic Segmentation by Early Region Proxy		
9	2022	CVPR	Semi-Supervised Semantic Segmentation With Error Localization Network		
10	2021	CVPR	VS-Net: Voting With Segmentation for Visual Localization		
11	2021	CVPR	Exploit Visual Dependency Relations for Semantic Segmentation		
12	2021	CVPR	Global2Local: Efficient Structure Search for Video Action Segmentation		
13	2021	ICLR	On the Connection between Local Attention and Dynamic Depth-wise Convolution		
14	2021	ICCV	TransPose: Keypoint Localization via Transformer		
15	2022	CVPR	Sparse Non-Local CRF		
16					
17					
18					
19	2021	*IEEE/CVF International Conf*	Swin transformer: Hierarchical vision transformer using shifted windows		

Individual Study

Sementic Segmentation using localized information

Sementic Segmentation using localized information					
no	year	j/c	title	dataset	abstrac
1	2019	*IEEE journal of biomedical ar*	Knowledge-aided convolutional neural network for small organ segmentation.		
2	2019	*IEEE/CVF International Conf*	Fine-grained segmentation networks: Self-supervised segmentation for improved long-term visual localization.		
3	2020	*NeuroImage*	An automated framework for localization, segmentation and super-resolution reconstruction of fetal brain MRI.		
4	2020	*IEEE transactions on pattern*	Attention-Based Dropout Layer for Weakly Supervised Single Object Localization and Semantic Segmentation		
5	2021	Ieee Access	Real-time polyp detection, localization and segmentation in colonoscopy using deep learning		
6			lassifier and Thresholding		
7					
8					
9					
10					
11	2021	CVPR	Exploit Visual Dependency Relations for Semantic Segmentation		
12	2021	CVPR	Global2Local: Efficient Structure Search for Video Action Segmentation		
13	2021	ICLR	On the Connection between Local Attention and Dynamic Depth-wise Convolution		
14	2021	ICCV	TransPose: Keypoint Localization via Transformer		
15	2022	CVPR	Sparse Non-Local CRF		
16					
17					
18					
19	2021	*IEEE/CVF International Conf*	Swin transformer: Hierarchical vision transformer using shifted windows		

cf		
2011	NIPS	Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials
2013	*Journal of mathemati*	Image Segmentation Using Local GMM in a Variational Framework
2018	*IEEE conference on*	Semantic visual localization
2019	*IEEE reviews in biom*	A Review of Automated Methods for the Detection of Sickle Cell Disease

Individual Study

Sementic Segmentation using localized information

- Conditional Random Field (CRF)

Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials

Philipp Krähenbühl
Computer Science Department
Stanford University
philkr@cs.stanford.edu

Vladlen Koltun
Computer Science Department
Stanford University
vladlen@cs.stanford.edu

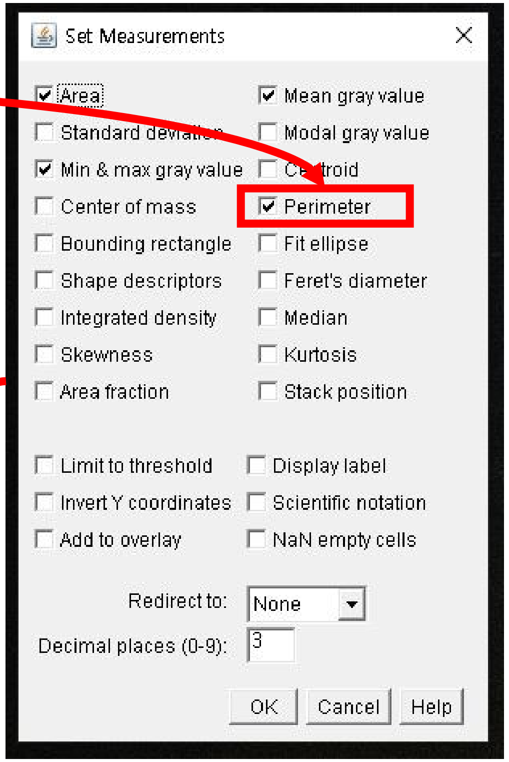
Abstract

Most state-of-the-art techniques for multi-class image segmentation and labeling use conditional random fields defined over pixels or image regions. While region-level models often feature dense pairwise connectivity, pixel-level models are considerably larger and have only permitted sparse graph structures. In this paper, we consider fully connected CRF models defined on the complete set of pixels in an image. The resulting graphs have billions of edges, making traditional inference algorithms impractical. Our main contribution is a highly efficient approximate inference algorithm for fully connected CRF models in which the pairwise edge potentials are defined by a linear combination of Gaussian kernels. Our experiments demonstrate that dense connectivity at the pixel level substantially improves segmentation and labeling accuracy.

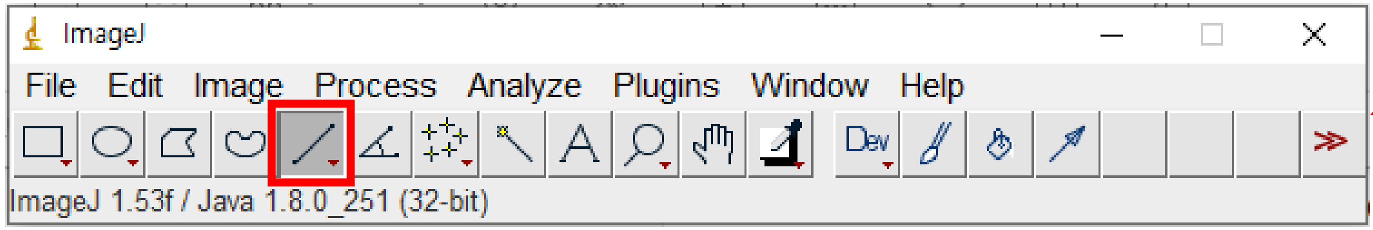
Krähenbühl, P., & Koltun, V. (2011). Efficient inference in fully connected crfs with gaussian edge potentials. *Advances in neural information processing systems*, *24*.

갑상선 연구 – 눈꺼풀 길이 추론

ImageJ



- 자: 50mm
- 스티커: 9mm



To do list

Next progress

- 개인연구 주제 관련 Research

- ImageJ 마스크 확인, 상관관계 분석

Employing Multi-branch Convolution Structure in 1D ResNet for Music Skip Prediction

Sanghyeong Jin

jinsanghyeong@gmail.com

2022. 11. 29.

Contents

1. Introduction
2. Related Work
3. Proposed Method
4. Experimental Settings
5. Experimental Results
6. Conclusion

Appendix

References

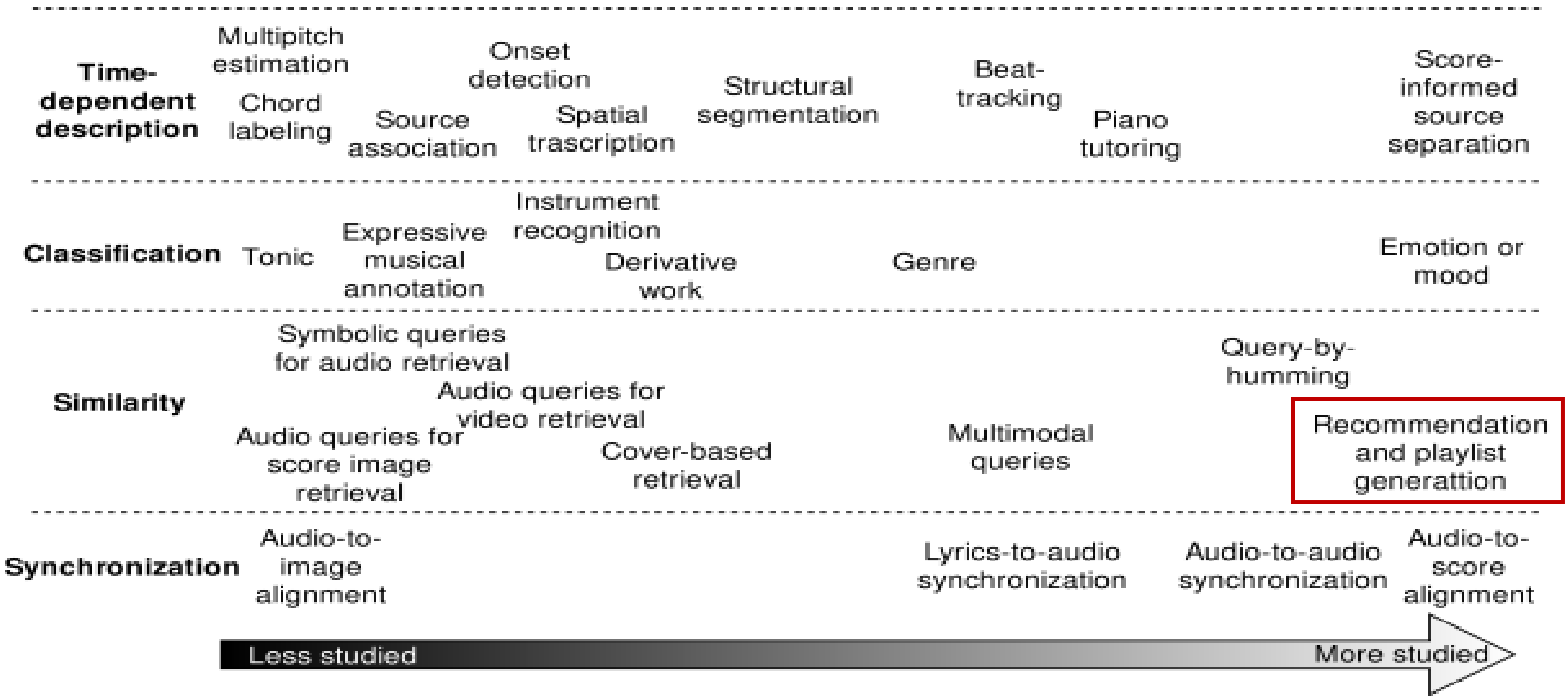
To Do List

Experimental Results Discussion

- Feature map 관련 추가 실험, 조사

Introduction

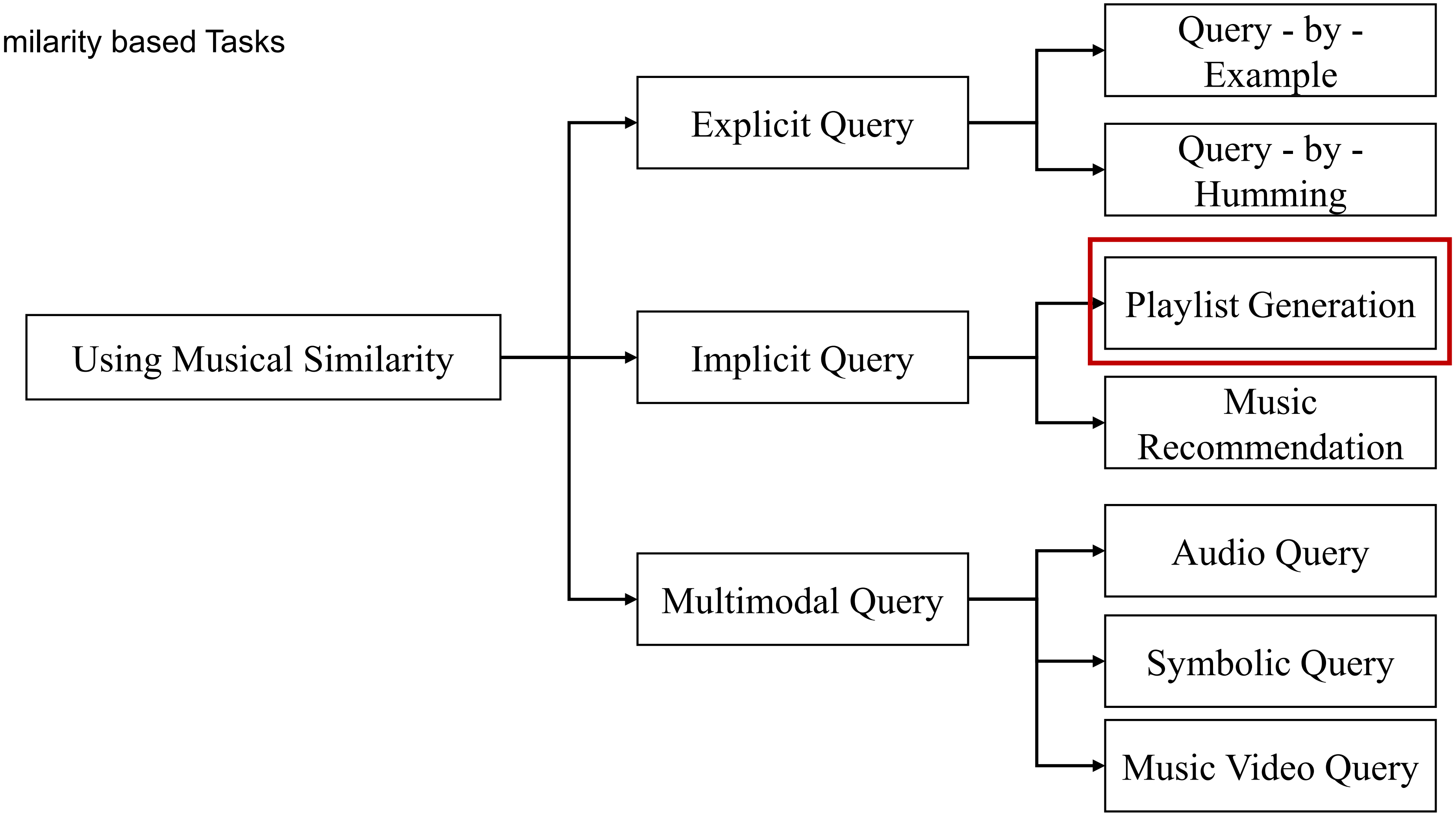
Music Information Retrieval



Introduction

Music Information Retrieval

- Musical Similarity based Tasks

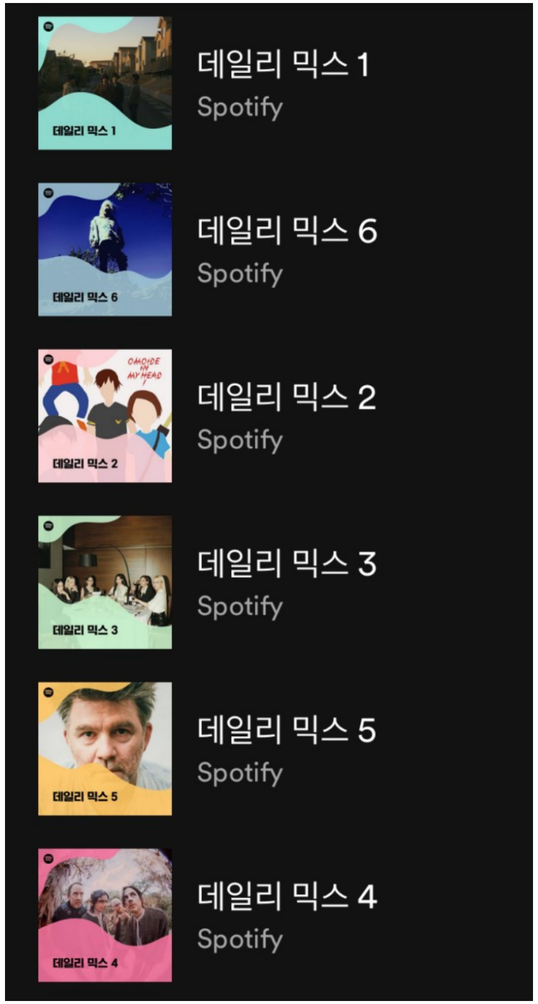


Introduction

Playlist Generation

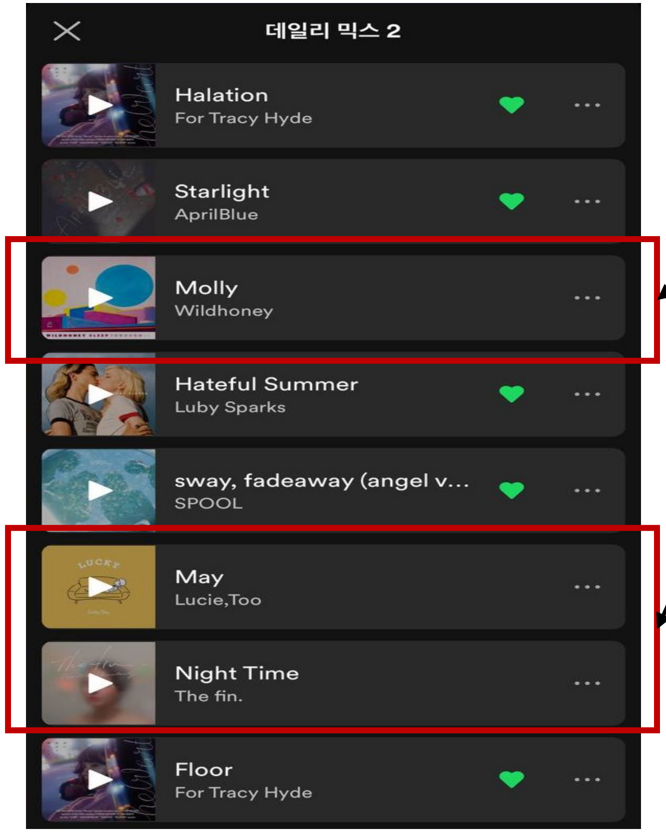
Playlist Generation

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



Introduction

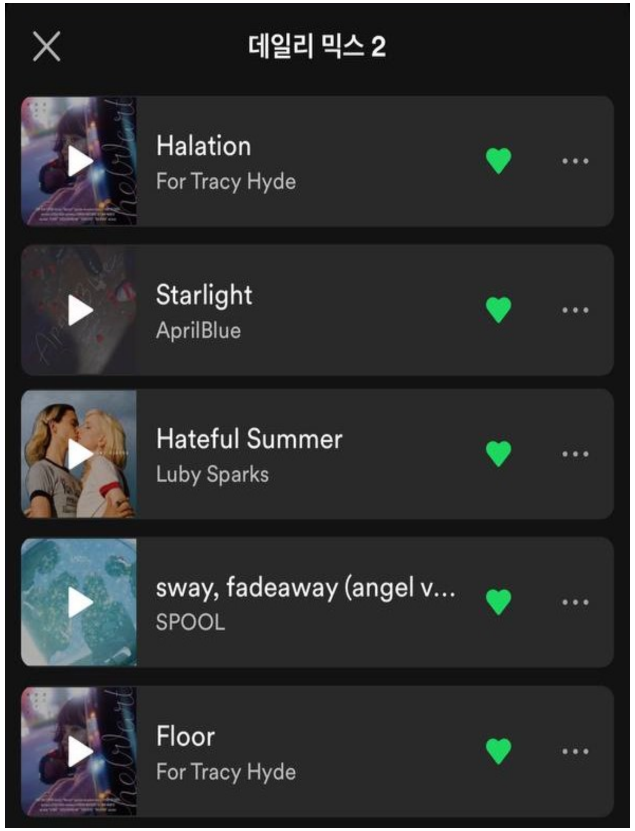
Music Track Skip Prediction



General mix playlist

Annoying to skip these tracks
at the end of each song

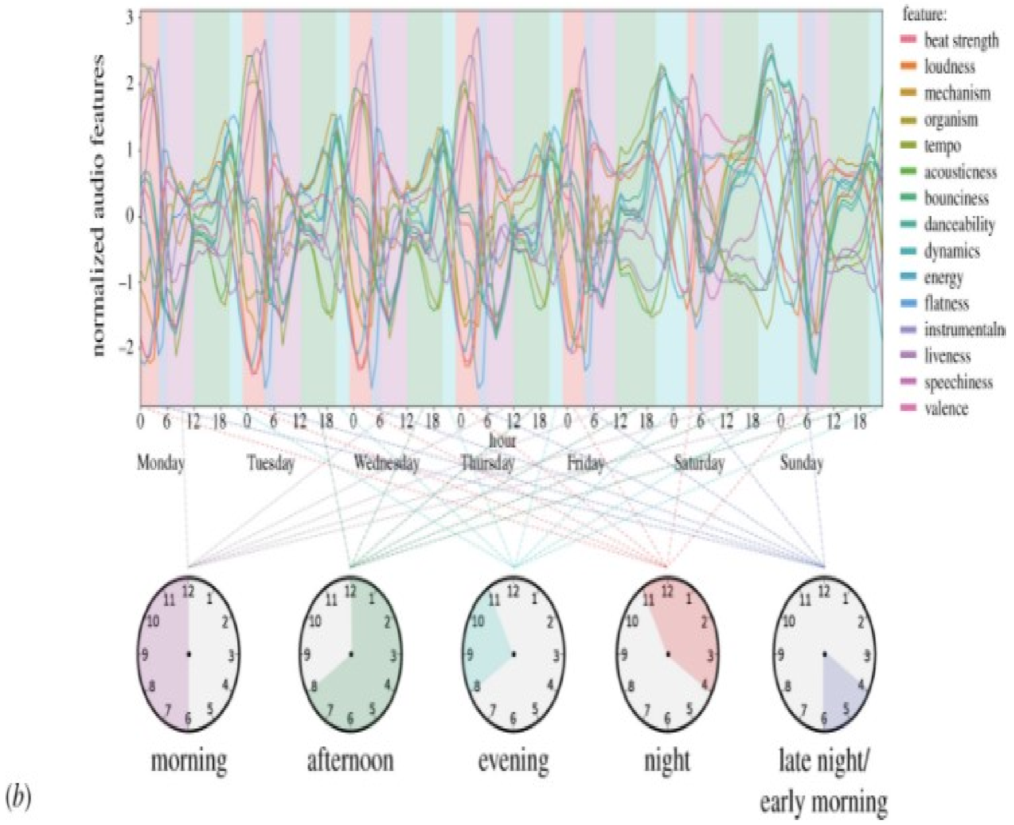
- Minimizing user skip intervention
- Smooth track sequence transition
- Generating user preferred playlist



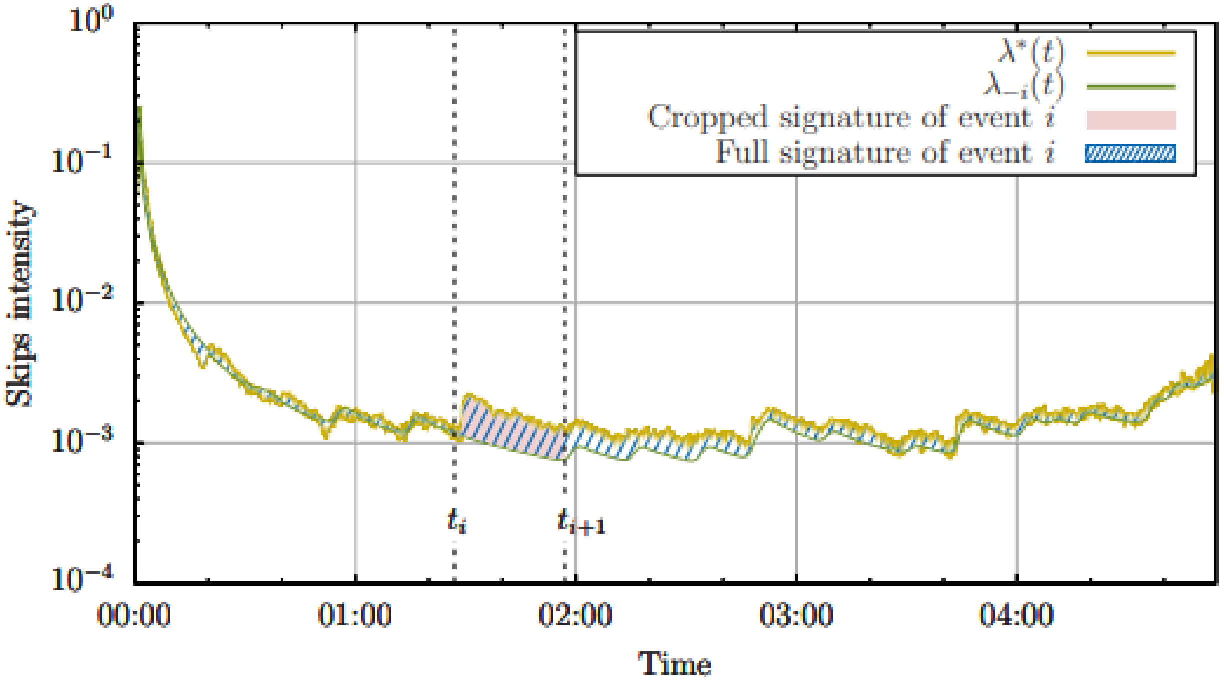
User preferred playlist

Related Work

Track Skip Behavior



Musical diurnal fluctuations

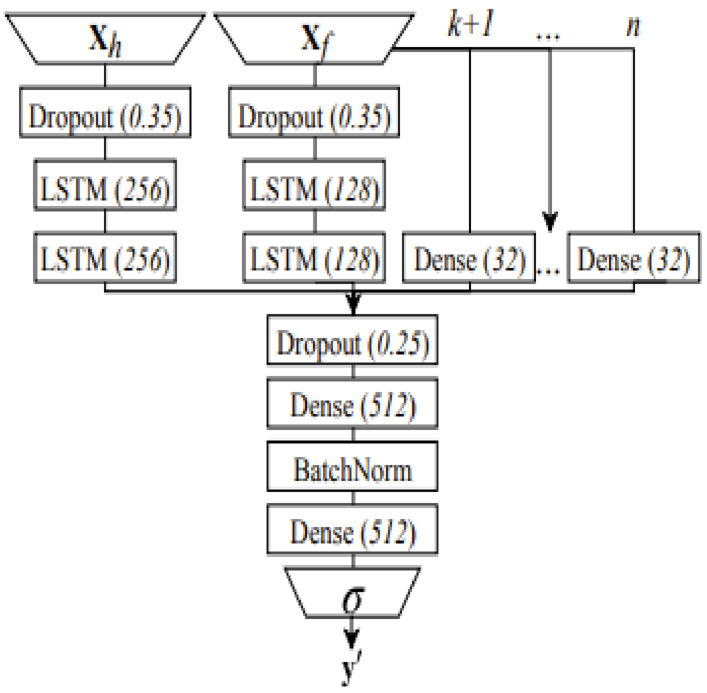


Skip rate based on time

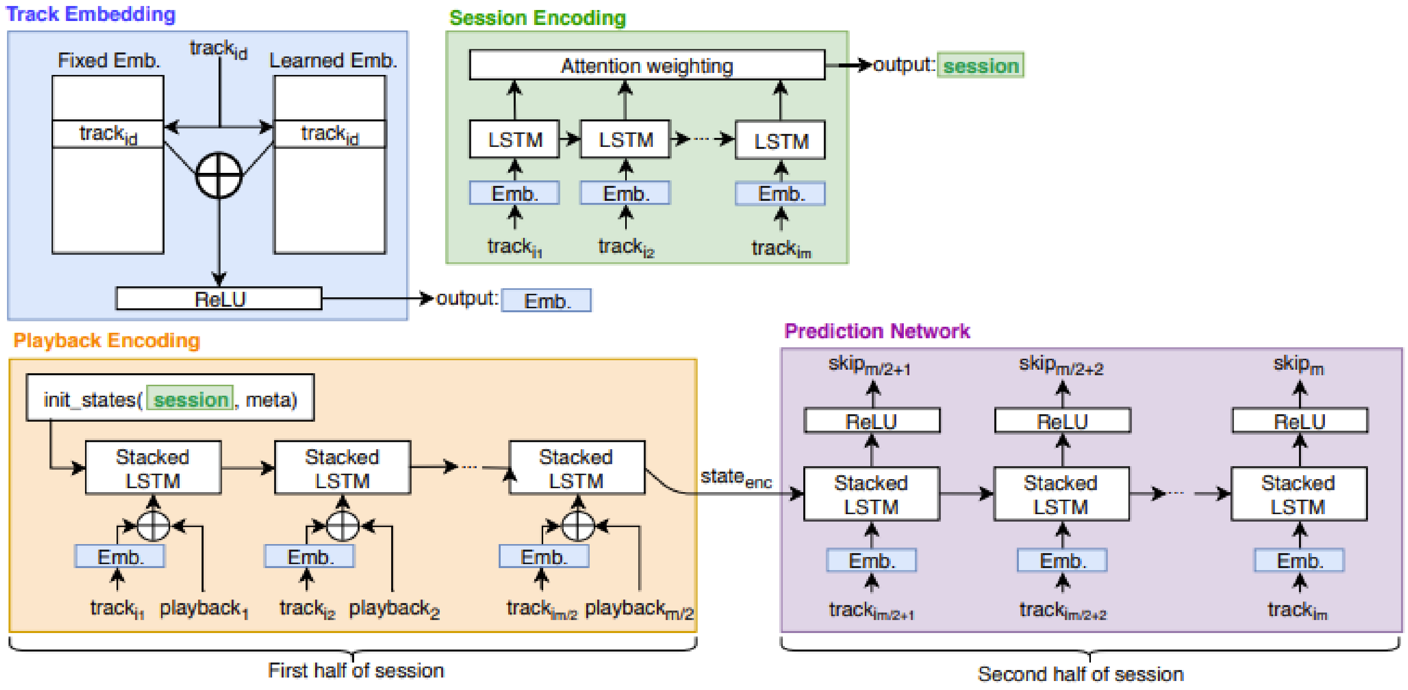
Related Work

Track Skip Prediction

- Mostly based on RNN Models



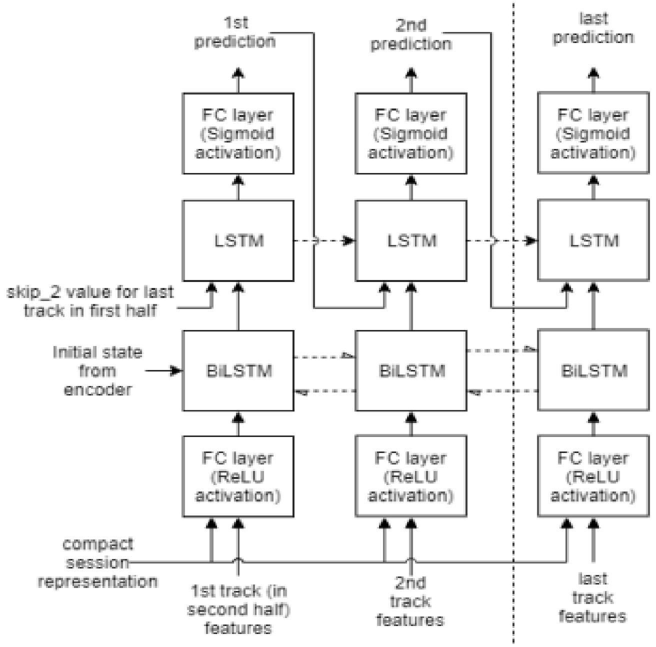
RNN-Dense



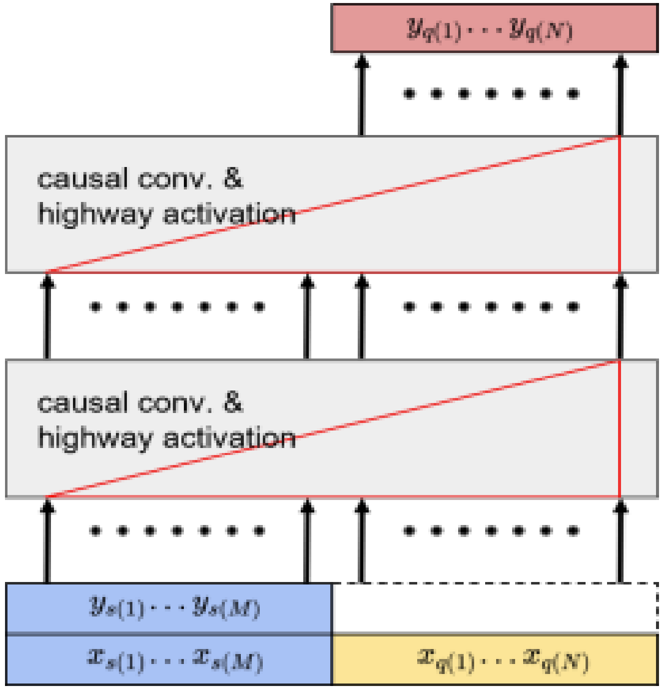
Multi-RNN

Related Work

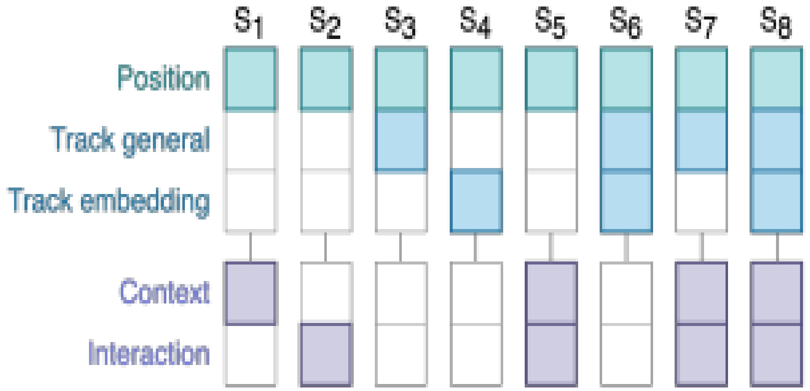
Track Skip Prediction



BiLSTM Encoder-Decoder



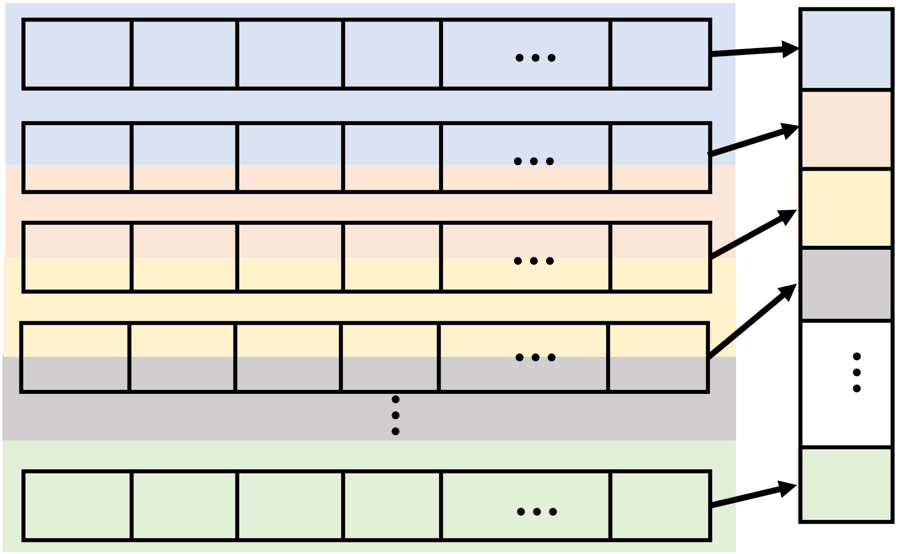
1D convolution with Highway Network



Interpretable Transformer

Related Work

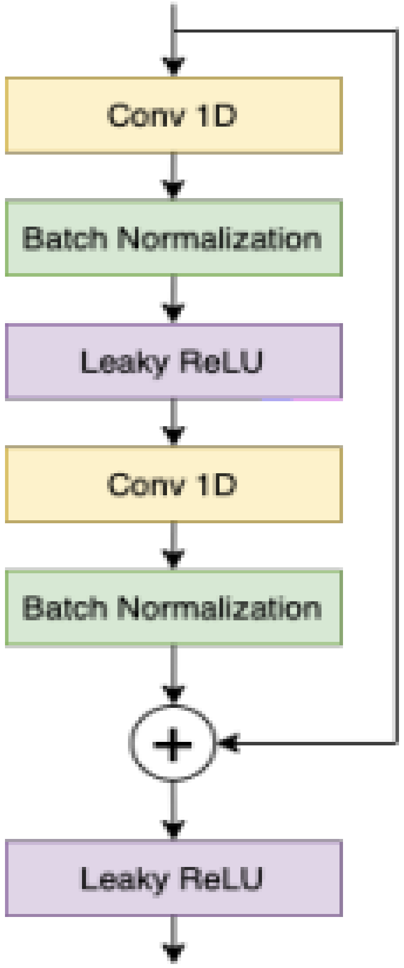
1D ResNet



1D Convolution

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	$H(\mathbf{x}) \cdot T(\mathbf{x}) + \mathbf{x} \cdot C(\mathbf{x})$

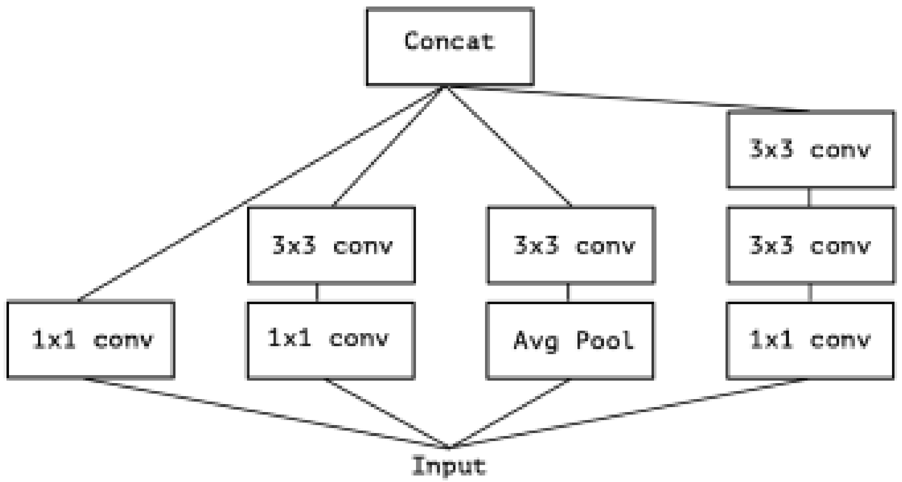
Residual Connection



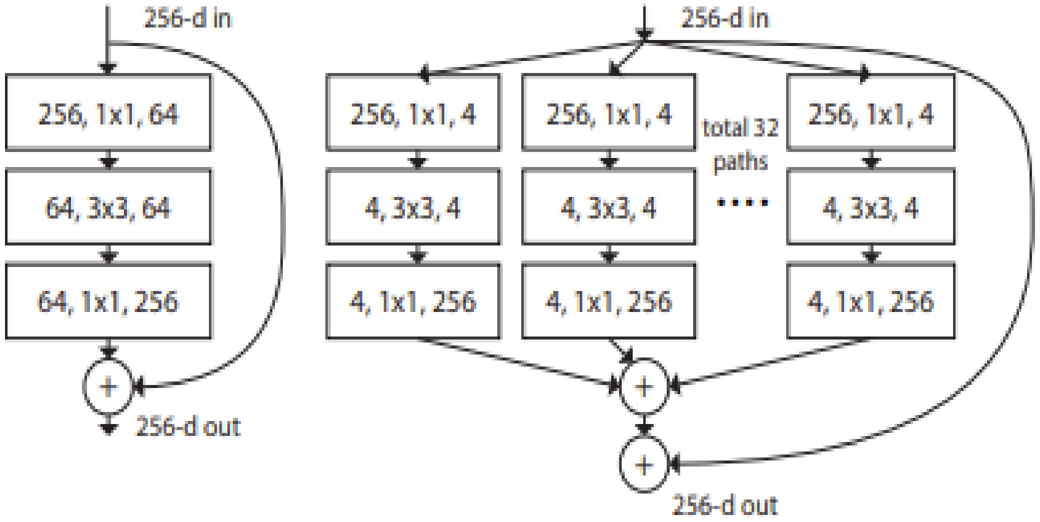
1D Residual Block

Related Work

Multi-branch Convolution



Inception V3



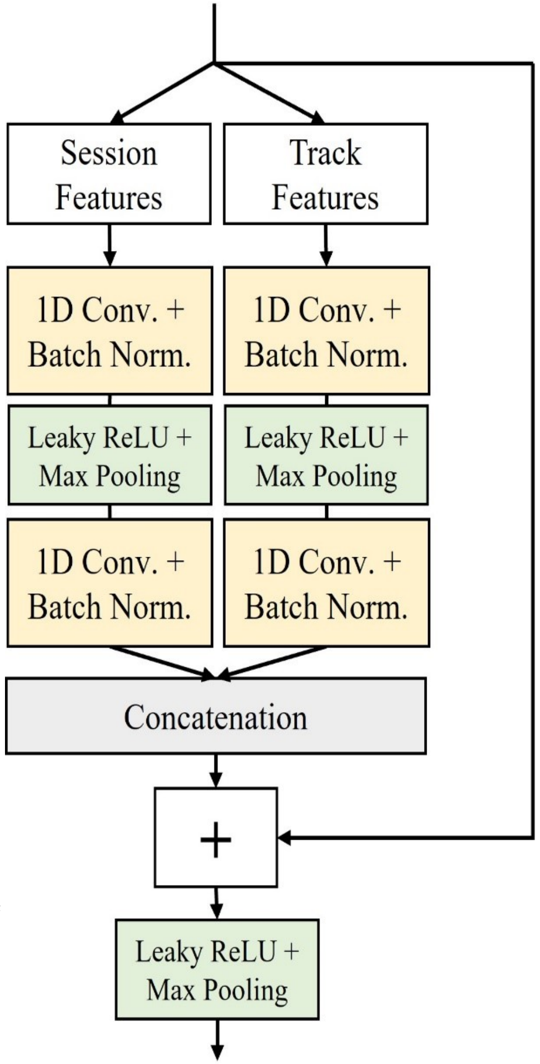
ResNeXt

Proposed Method

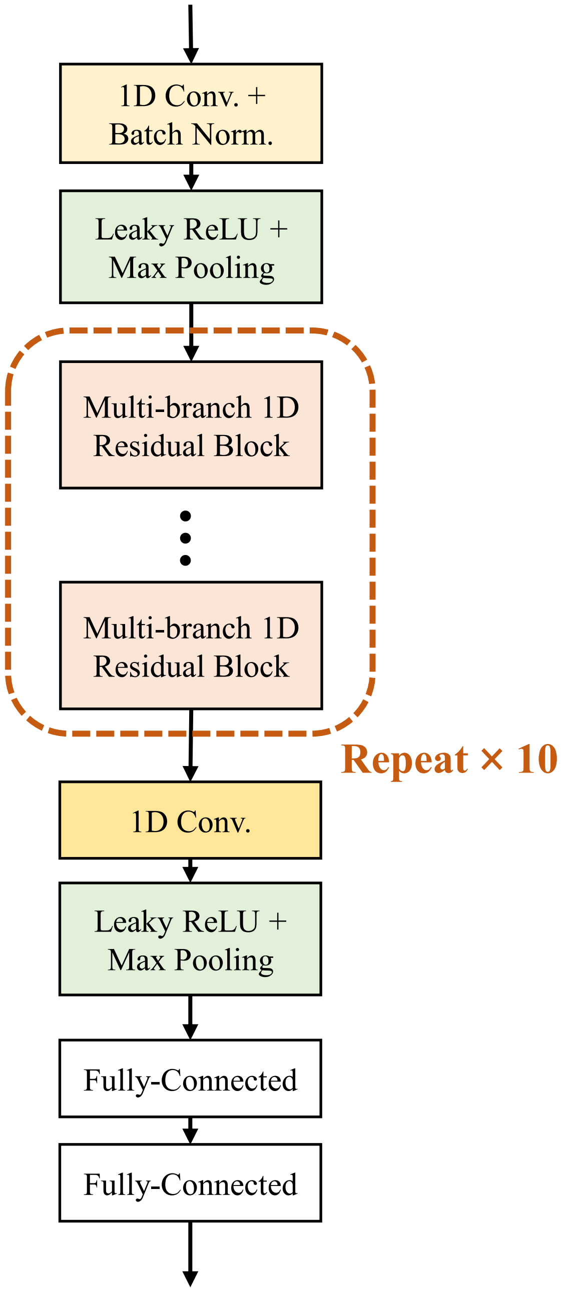
Multi-branch 1D Residual Block

$B \times (S + T) \times L$	
Session Features	Track Features
$B \times S \times L$	$B \times T \times L$
$B \times S \times L$	$B \times T \times L$
Concatenation	
$B \times (S + T) \times L$	

B : Batch size, S : Session size, T : Track size
L : Sequence Length



Multi-branch 1D ResNet



Experimental Settings

MAA (Mean Average Accuracy)

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

T : Number of tracks to be predicted
A(i) : Accuracy to the *i*-th track in sequence
L(i) : Boolean indicator that if *i*-th prediction was correct (0, 1)

Ground Truth	Predicted	AA (Average Accuracy)
1, 1, 1, 1, 1	1, 1, 1, 1, 0	0.800
1, 1, 1, 1, 1	1, 1, 1, 0, 1	0.760
1, 1, 1, 1, 1	1, 1, 0, 1, 1	0.710
1, 1, 1, 1, 1	1, 0, 1, 1, 1	0.643
1, 1, 1, 1, 1	0, 1, 1, 1, 1	0.543

$$\rightarrow \frac{\left(\frac{1}{1} * 1 + \frac{2}{2} * 1 + \frac{3}{3} * 1 + \frac{4}{4} * 1 + \frac{4}{5} * 0\right)}{5} = 0.800$$

$$\rightarrow \frac{\left(\frac{1}{1} * 1 + \frac{2}{2} * 1 + \frac{2}{3} * 0 + \frac{3}{4} * 1 + \frac{4}{5} * 1\right)}{5} = 0.710$$

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

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 : (74, 20) - (Session & Track Features, Sequence Length) / Sequence that less than 20 includes zero padding
  - Output : (10, 1) - Prediction probability for the 10 tracks in second half
- Cross-Validation for Test Performance
  - 10 iterations / Group K-Fold (by Session id)
  - Data Split Ratio: (Train : Test) = (0.9: 0.1)
- Hyper Parameter Setting:

<table>
<tr><th>Model</th><th>Batch Size</th><th>Epochs</th><th>Loss Function</th><th>Optimizer</th><th>LR</th></tr>
<tr><td>Proposed</td><td rowspan="5">16</td><td rowspan="5">10</td><td rowspan="5">Binary Cross Entropy</td><td rowspan="2">Adam</td><td rowspan="3">1e-4</td></tr>
<tr><td>Highway 1D Conv.</td></tr>
<tr><td>Encoder-Decoder</td><td>RMSprop</td></tr>
<tr><td>Multi-RNN</td><td rowspan="2">Adam</td><td>1e-3</td></tr>
<tr><td>RNN-Dense</td><td>1e-4</td></tr>
</table>

Experimental Results

t-test Result

Model	MAA Mean ± Standard Deviation	Average Rank
Proposed	**0.6796 ± 0.0398**	**1.53**
Highway 1D Conv. [12]	0.5976 ± 0.0885 ▼	3.13
Encoder-Decoder [16]	0.5357 ± 0.0912 ▼	3.8
Multi-RNN [14]	0.6027 ± 0.0583 ▼	3.1
RNN-Dense [15]	0.5740 ± 0.0901 ▼	3.43

▼ indicates corresponding model is significantly worse than proposed model at significance level α=0.05

Experimental Results

Friedman-test Result

	Statistics	p-value
Proposed	36.080	2.786e-07

significance level α=0.05

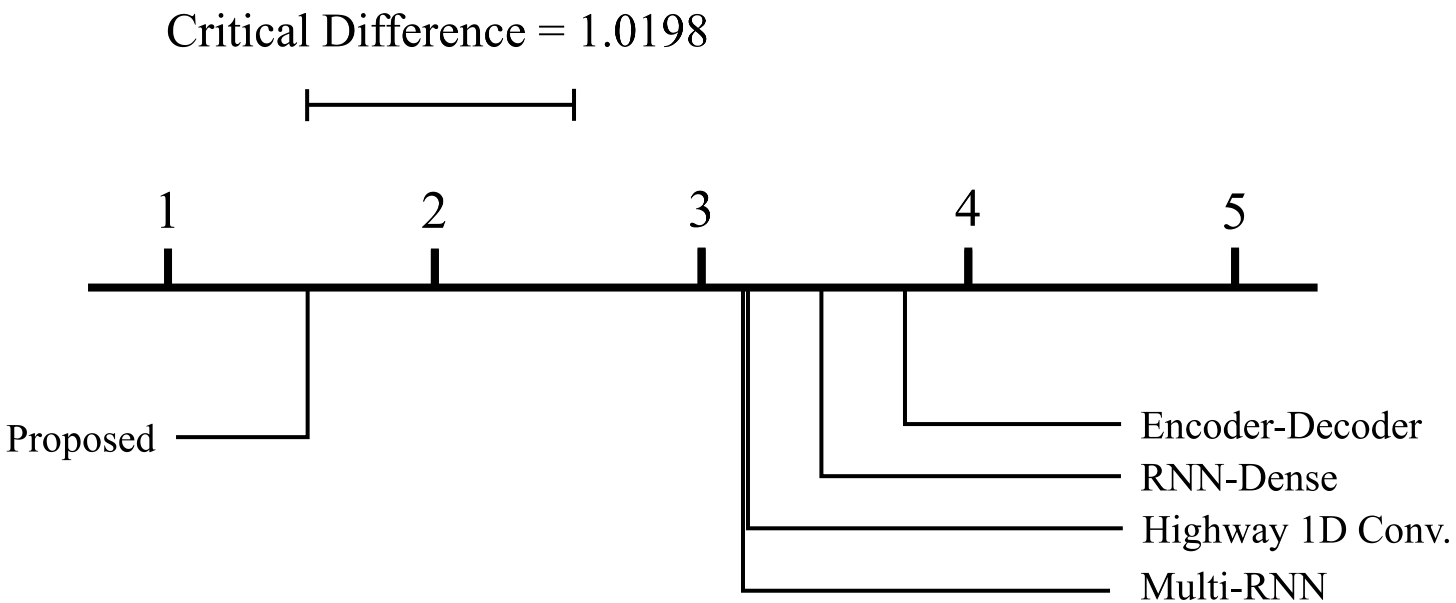
Wilcoxon signed-rank test Result

	Comparison models			
	Highway 1D Conv. [12]	Encoder-Decoder [16]	Multi-RNN [14]	RNN-Dense [15]
Proposed ***versus*** (p-value)	Win (5.31e-05)	Win (4.28e-06)	Win (1.64e-05)	Win (7.51e-05)

significance level α=0.05

Experimental Results

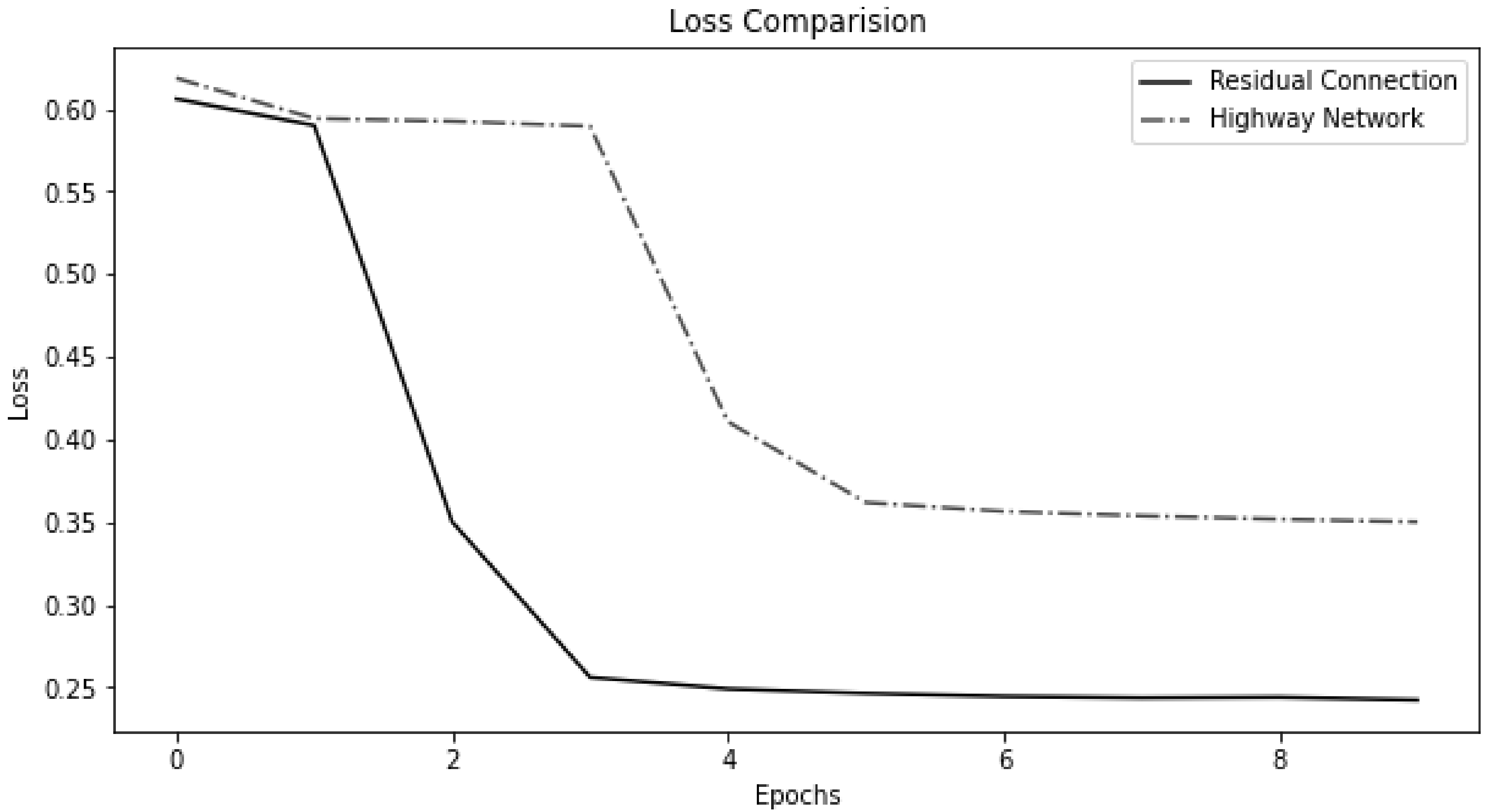
Bonferroni-Dunn test Result



Experimental Results

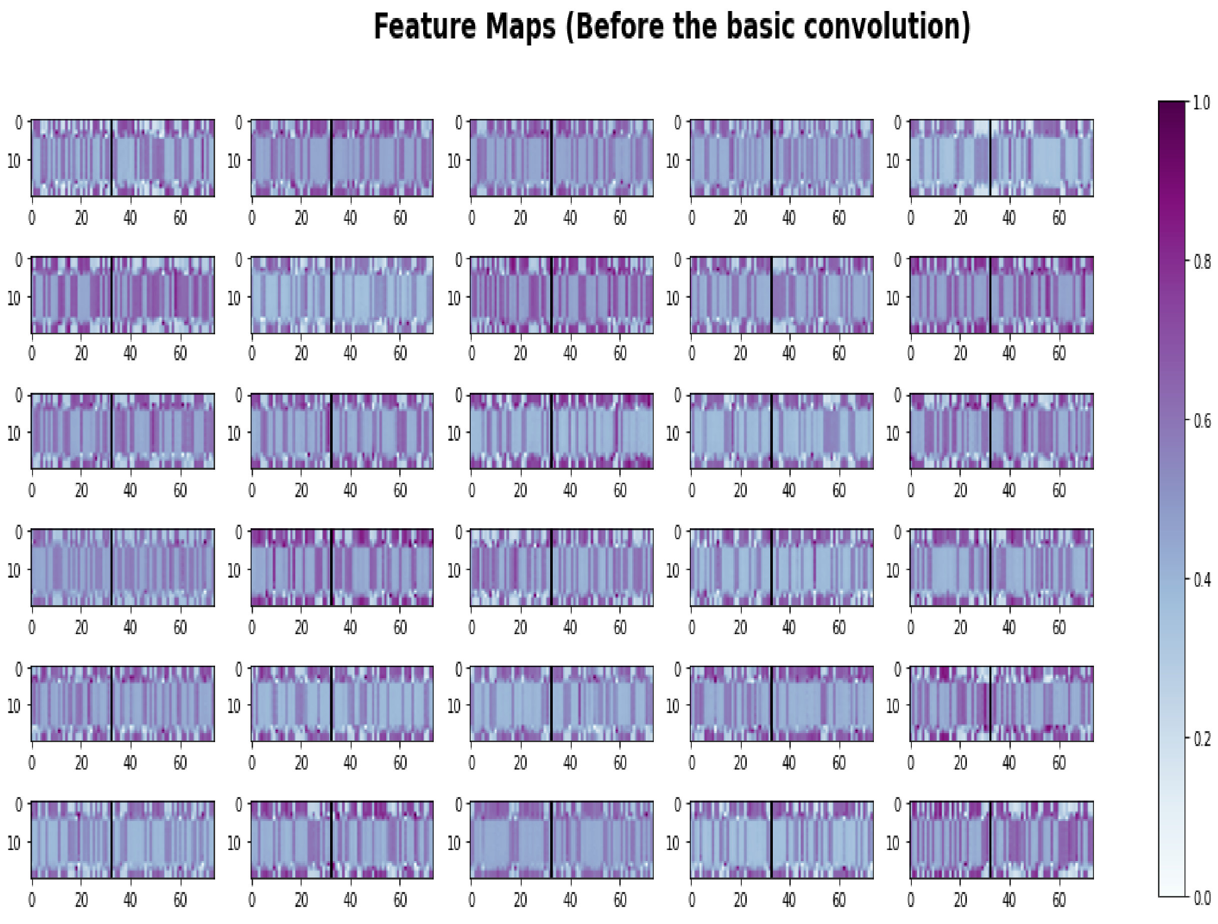
Discussion

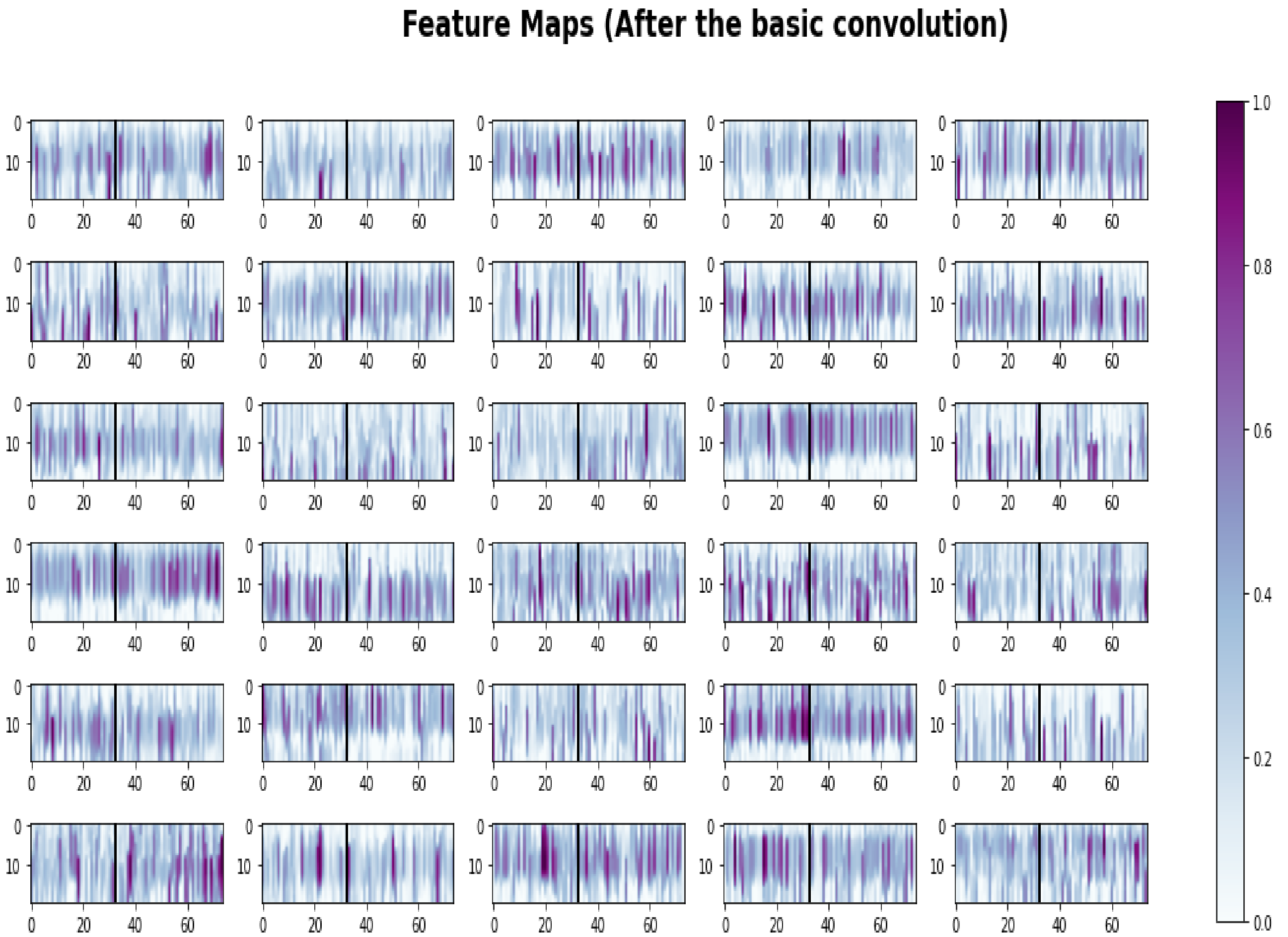
- Highway Network vs. Residual Connection



Experimental Results

Discussion - 1D CNN Feature Maps Visualization

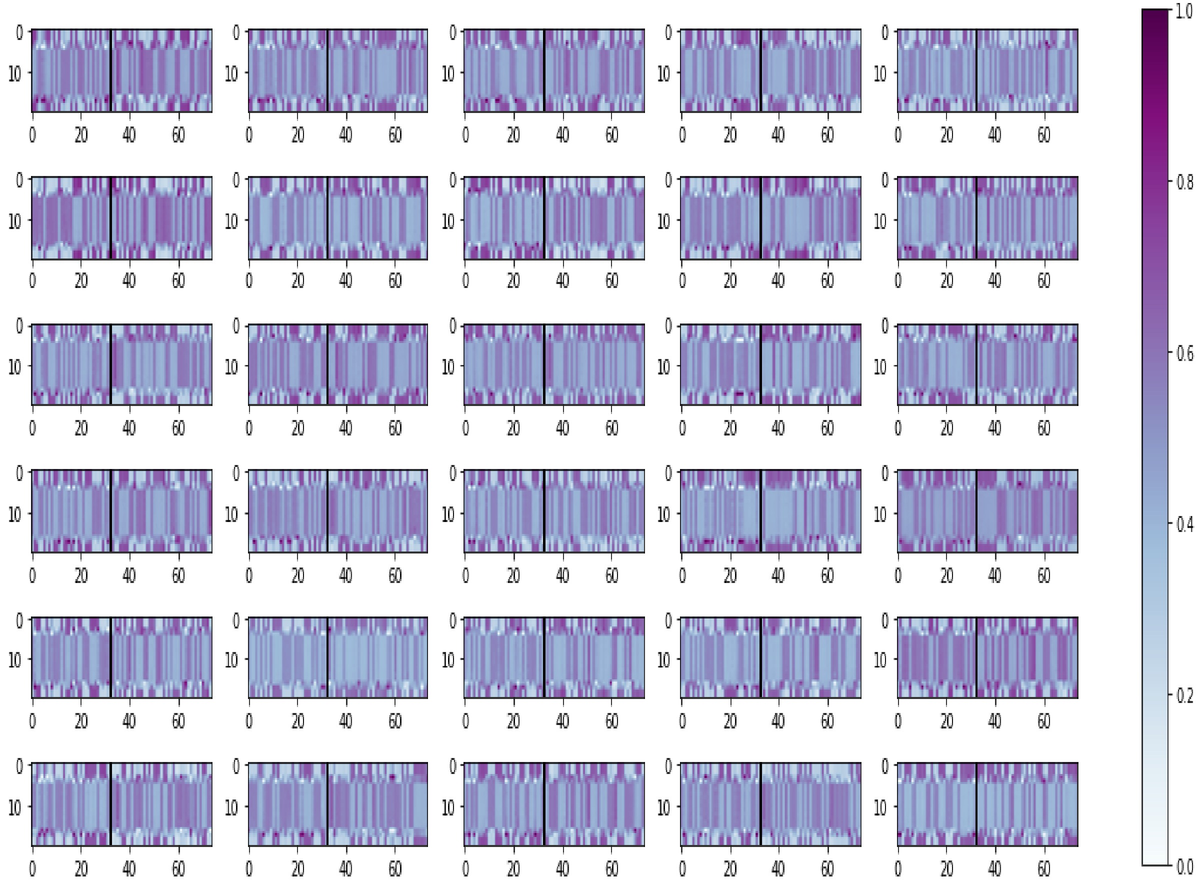




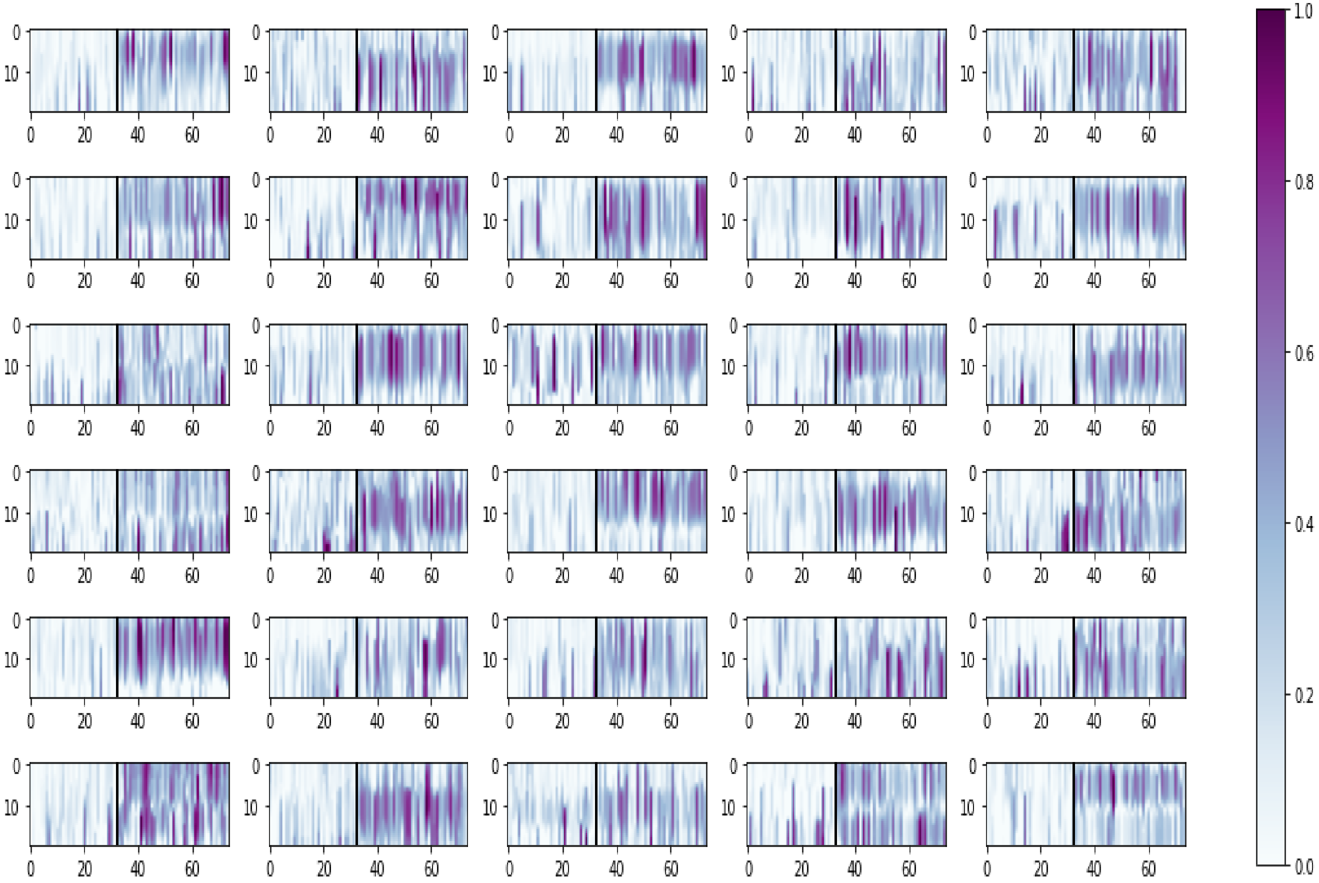
Experimental Results

Discussion - 1D CNN Feature Maps Visualization

Feature Maps (Before the multi-branch convolution)

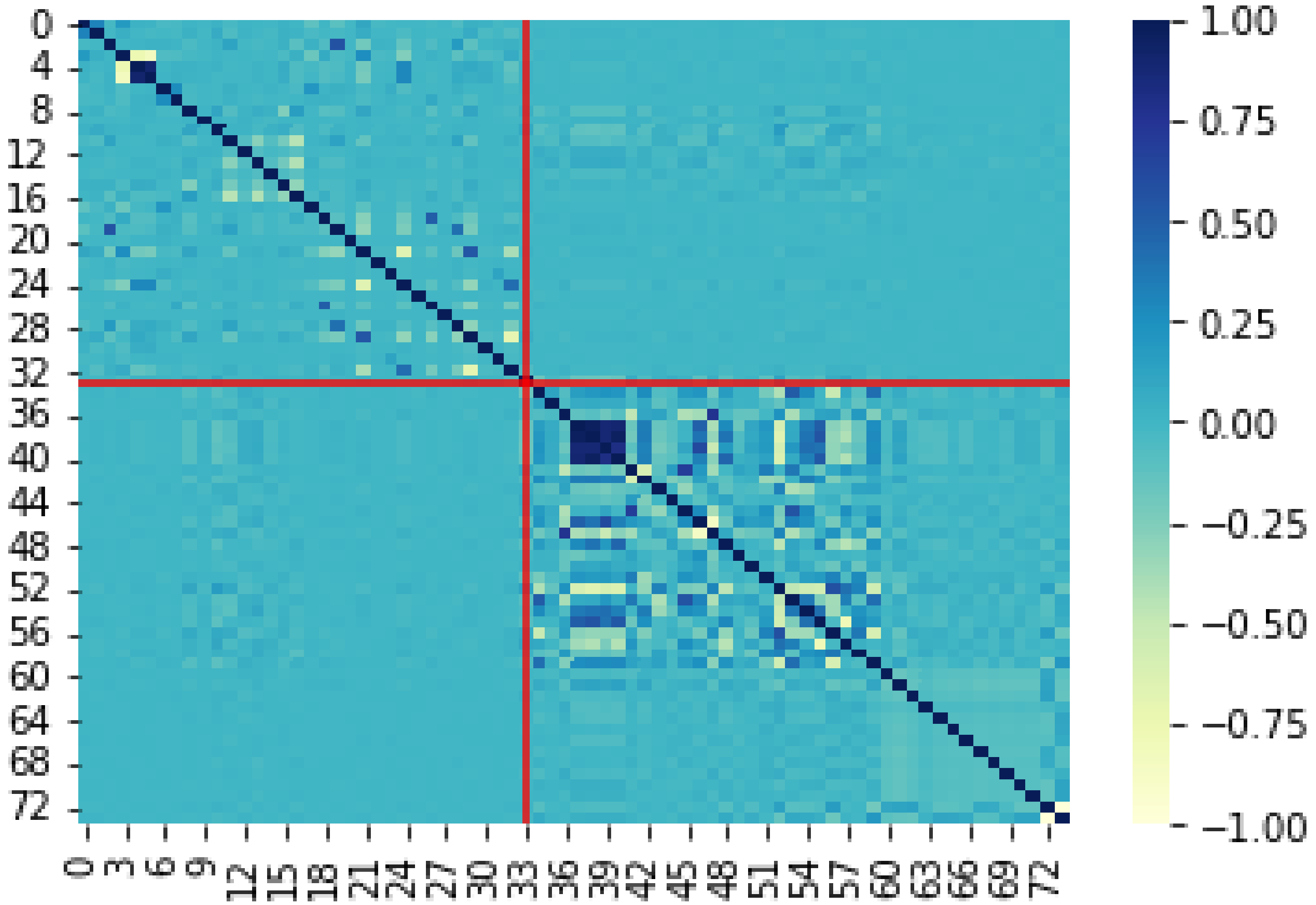


Feature Maps (After the last multi-branch convolution)



Experimental Results

Discussion – Correlations between Session and Track



Conclusion

Summary

- Proposed multi-branch 1D residual block based on 1D convolution and 1D ResNet

- Proposed method perform significantly better than conventional RNN-based models

- Explicitly split multi-branch convolution can drive high-correlated feature learning

Future Works

- Using track more audio features

- Applying grouped convolution (separate inputs into same sizes)

기타 사항

1. ICEIC 논문 관련

- 투고완료

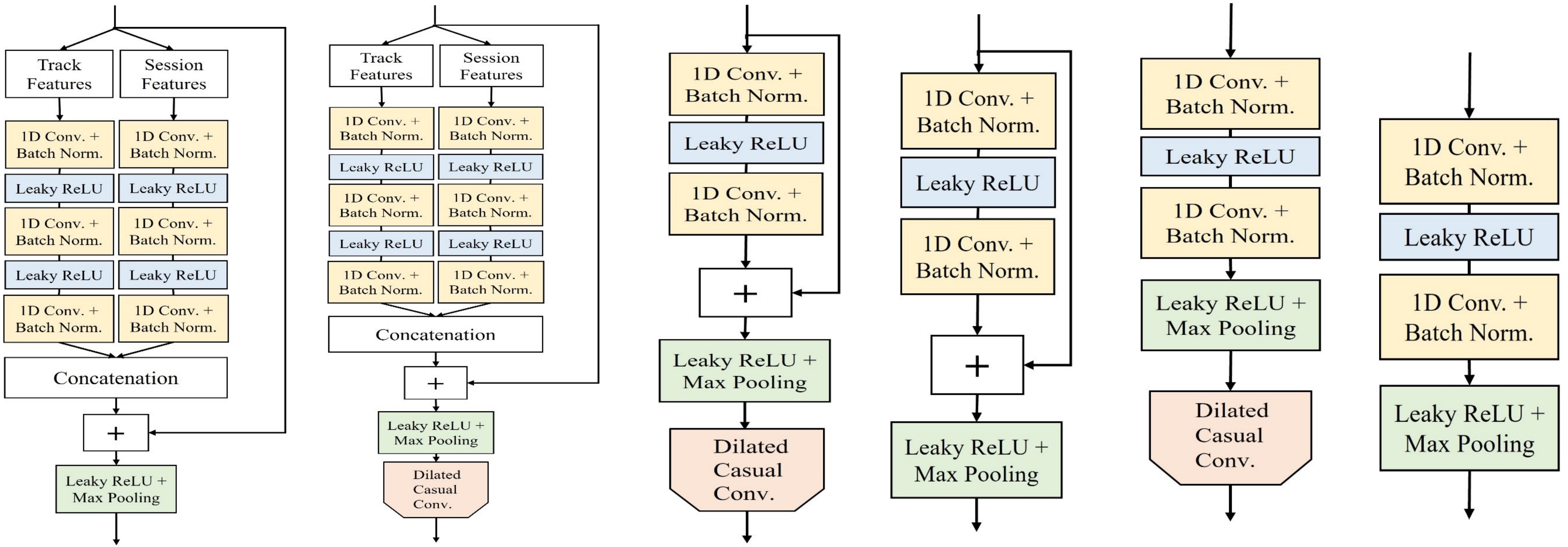
2. 인수인계

- 빠른 시일 내 진행 예정

Appendix

1D ResNet Variation Comparison

* DCC : Dilated Casual Convolution



1. Proposed

2. Multi-branch 1D ResNet + DCC

3. Basic 1D ResNet + DCC

4. Basic 1D ResNet

5. Residual Conn. Removed + DCC

6. Residual Conn. Removed

Appendix

t-test Result

Model	MAA Mean ± Standard Deviation	Average Rank
Proposed	**0.6796 ± 0.0398**	**1.46**
Multi-Branch 1D ResNet + DCC	0.6659 ± 0.0402	1.53
Basic 1D ResNet + DCC	0.5399 ± 0.0242 ▼	3.67
Basic 1D ResNet	0.5418 ± 0.0339 ▼	3.67
w/o Residual + DCC	0.4615 ± 0.0434 ▼	5.27
w/o Residual	0.4616 ± 0.0431 ▼	5.4

▼ indicates corresponding model is significantly worse than proposed model at significance level α=0.05
* DCC : Dilated Casual Convolution

Appendix

Friedman-test Result

	Statistics	p-value
Proposed	126.712	1.18e-25

significance level α=0.05

Wilcoxon signed-rank test Result

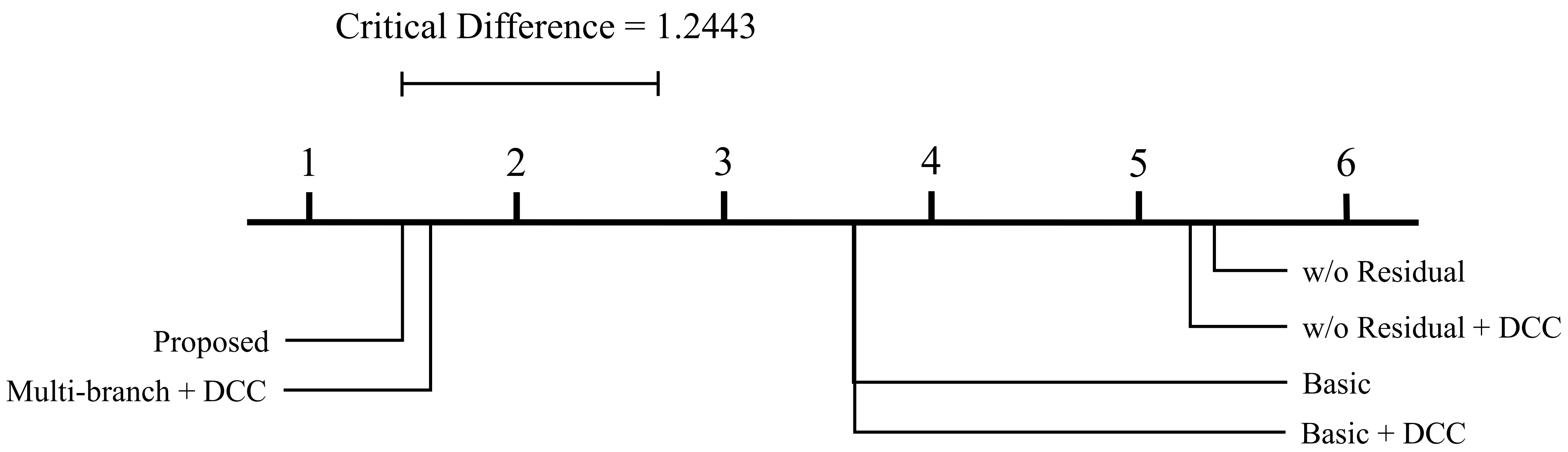
* DCC : Dilated Casual Convolution

	Comparison models				
	Multi-Branch 1D ResNet + DCC	Basic 1D ResNet + DCC	Basic 1D ResNet	w/o Residual + DCC	w/o Residual
Proposed ***versus*** (p-value)	Tie (0.2623)	Win (1.73e-06)	Win (1.73e-06)	Win (1.73e-06)	Win (1.73e-06)

significance level α=0.05

Appendix

Bonferroni-Dunn test Result



* DCC : Dilated Casual Convolution