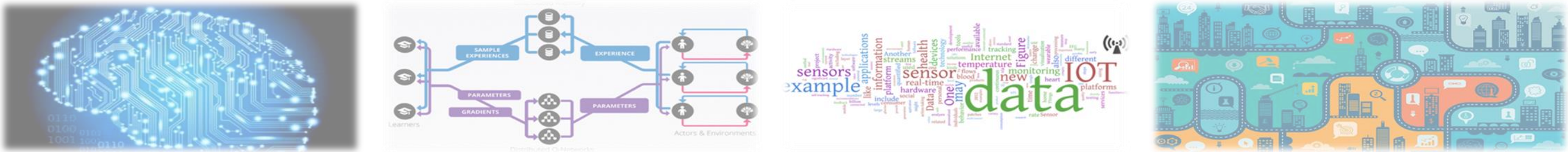
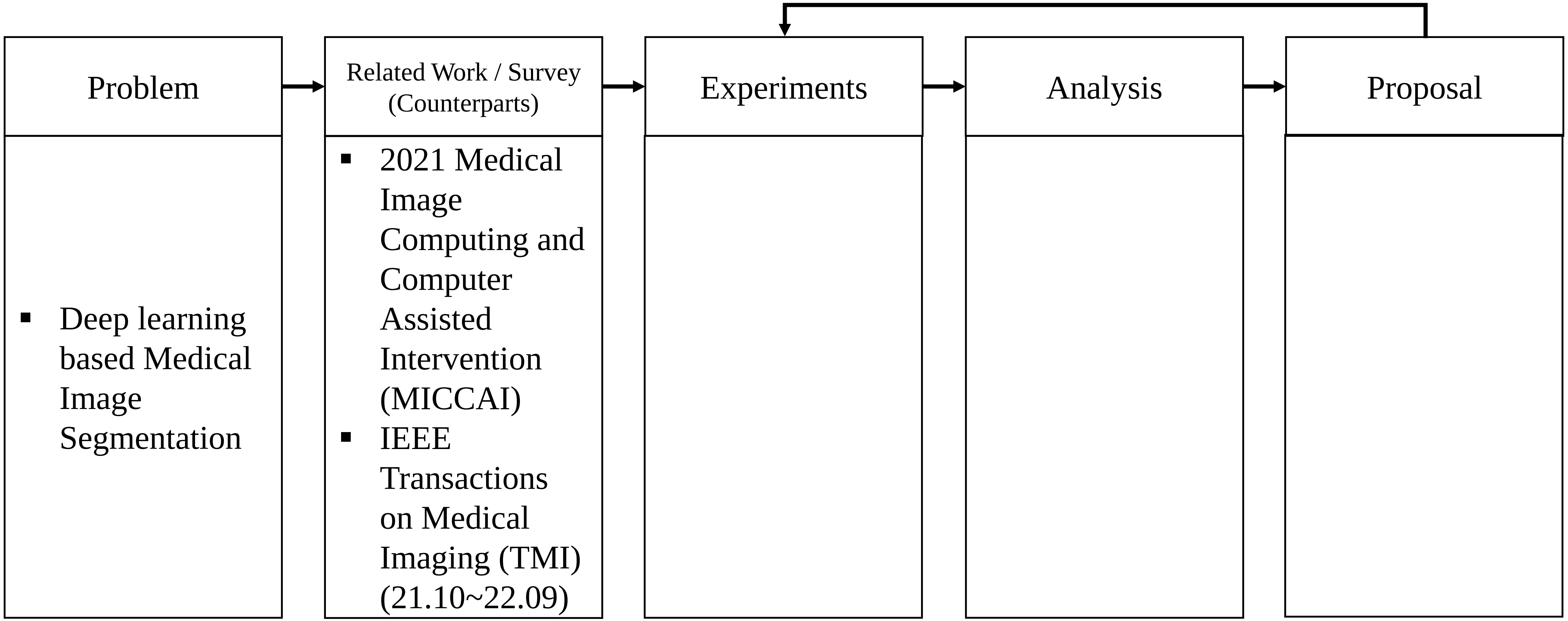


Deep learning based Medical Image Segmentation

Sanghyuck Lee

2022. 09. 28.

Overall Progress



Problem

- Deep learning based Medical Image Segmentation

Related Work / Survey (Counterparts)

- 2021 Medical Image Computing and Computer Assisted Intervention (MICCAI)
- IEEE Transactions on Medical Imaging (TMI) (21.10~22.09)

Experiments

Analysis

Proposal

Related Work

Recent Trends Survey

- With MICCAI 2021’s papers (total 531 papers):
  - Choose 2 Topics (CT, MRI) and filter by title & abstract manual check => 85 papers
- With TMI (21.10~22.09)’s papers (total 339 papers):
  - Search by keyword ‘segmentation’ and filter by title manual check => 63 papers

Future work

Additional RW survey, Dataset Acquisition and Model Implementation

- Additional Related Work survey
  - MICCAI 2020 (or/and 2022)
  - TMI (20.01 or 20.10~21.09)
- Dataset Acquisition & Model Implementation
  - 2D Datasets & Models

산학협력프로젝트(SW중심대학지원사업)

산학-2 실적 현황

- 국제 논문 2편 :
  - 2022 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia) 2건 게재 승인 (* 등록 10월 1일까지)
- 특허 2건
  - 2건 모두 변리사님 다음 연락 대기중

연말 해외 출장

출장 일정

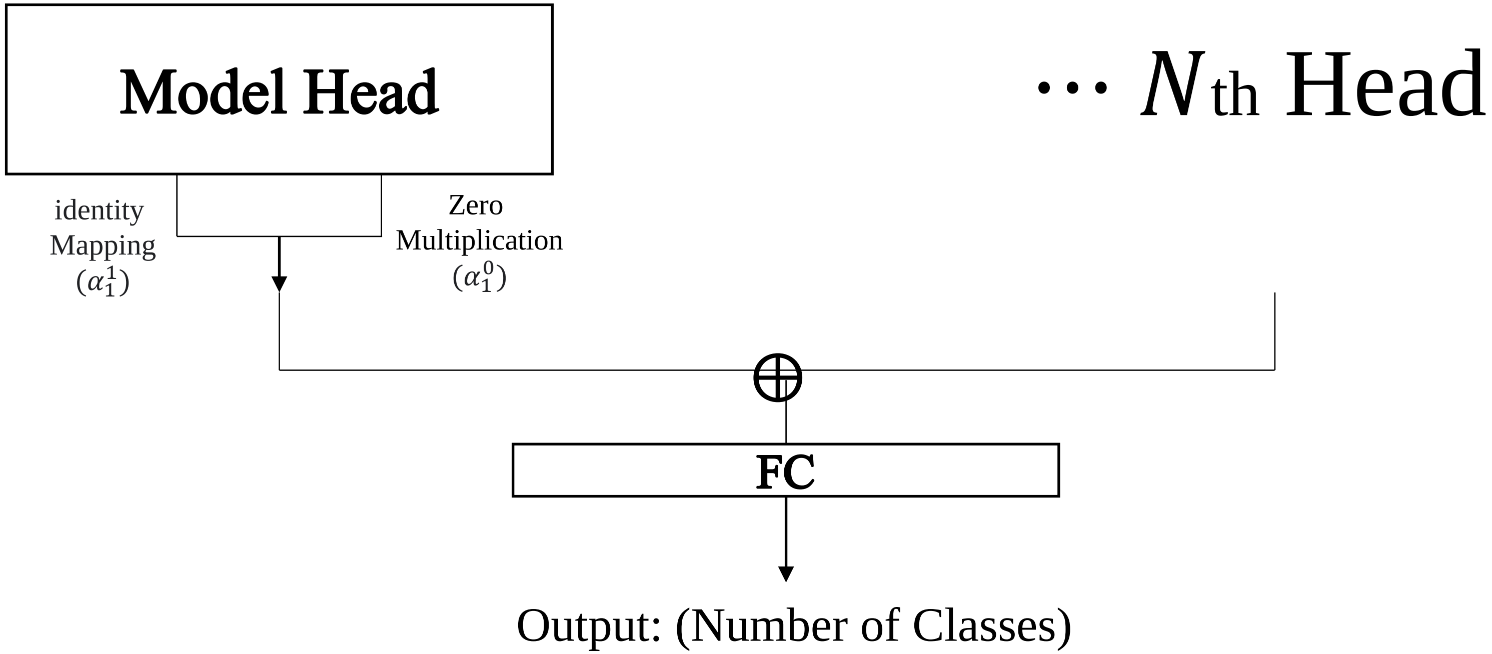
- 참여 행사 변경 : Bangkok or Osaka

이름	위치	기간	논문 데드라인
'2022 6th International Conference on Natural Language Processing and Information Retrieval	Bangkok, Thailand	'December 16-18	Sep 30, 2022
The 4th International Workshop on Big Data Tools, Methods, and Use Cases for Innovative Scientific Discovery (BTSD) 2022	Osaka, Japan	'December 17-20	Oct 1, 2022
Big Data Analytics for Health and Medicine (BDA4HM 2022)			Oct 1, 2022
The 6th International Workshop on Big Data Analytic for Cyber Crime Investigation and Prevention			Oct 5, 2022

연말 해외 출장

논문 작성

- Gradient-based pipeline selection (motivated by Liu et al., 2018)
- Experiment in progress



Model Head

identity Mapping (α_1^1)

Zero Multiplication (α_1^0)

$\cdots$ Nth Head

$\oplus$

FC

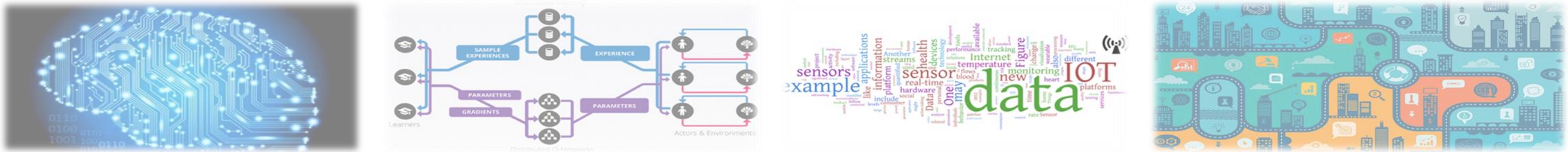
Output: (Number of Classes)

Liu, Hanxiao, Karen Simonyan, and Yiming Yang. "Darts: Differentiable architecture search." *arXiv preprint arXiv:1806.09055* (2018).

TAO studies with CAU Hospital & NVIDIA

Activity & Treatment

- Activity: 재작성 진행중
- Treatment
  - 실험 코드 – Minor 수정 중(가중치, 초매개변수 저장 등)
  - 2D -> 3D 모델 변형 관련 reference 준비 및 구현 중

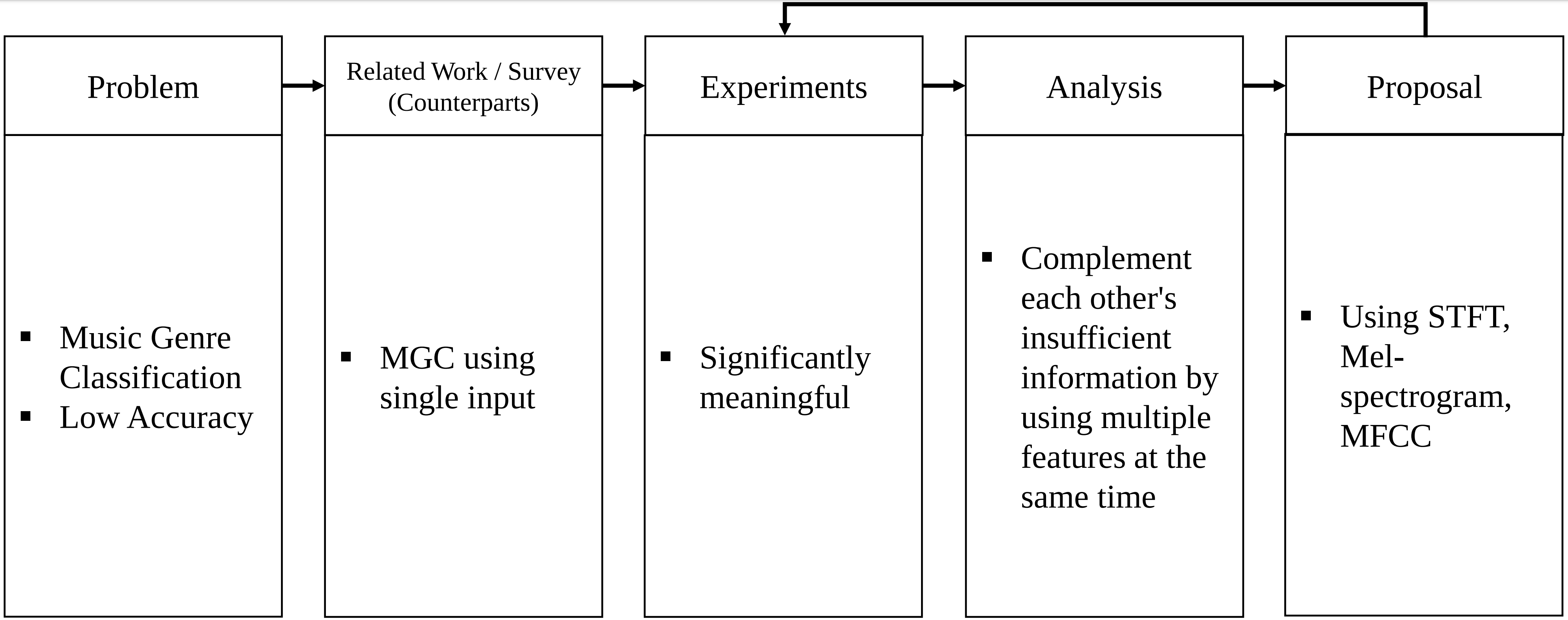


(가제)Effective Music Genre Classification using Late Fusion Convolutional Neural Network with Multiple Spectral Features

Sung-Hyun Cho

2022. 09. 28

Overall Progress (ICCE-Asia)



Problem

- Music Genre Classification
- Low Accuracy

Related Work / Survey (Counterparts)

- MGC using single input

Experiments

- Significantly meaningful

Analysis

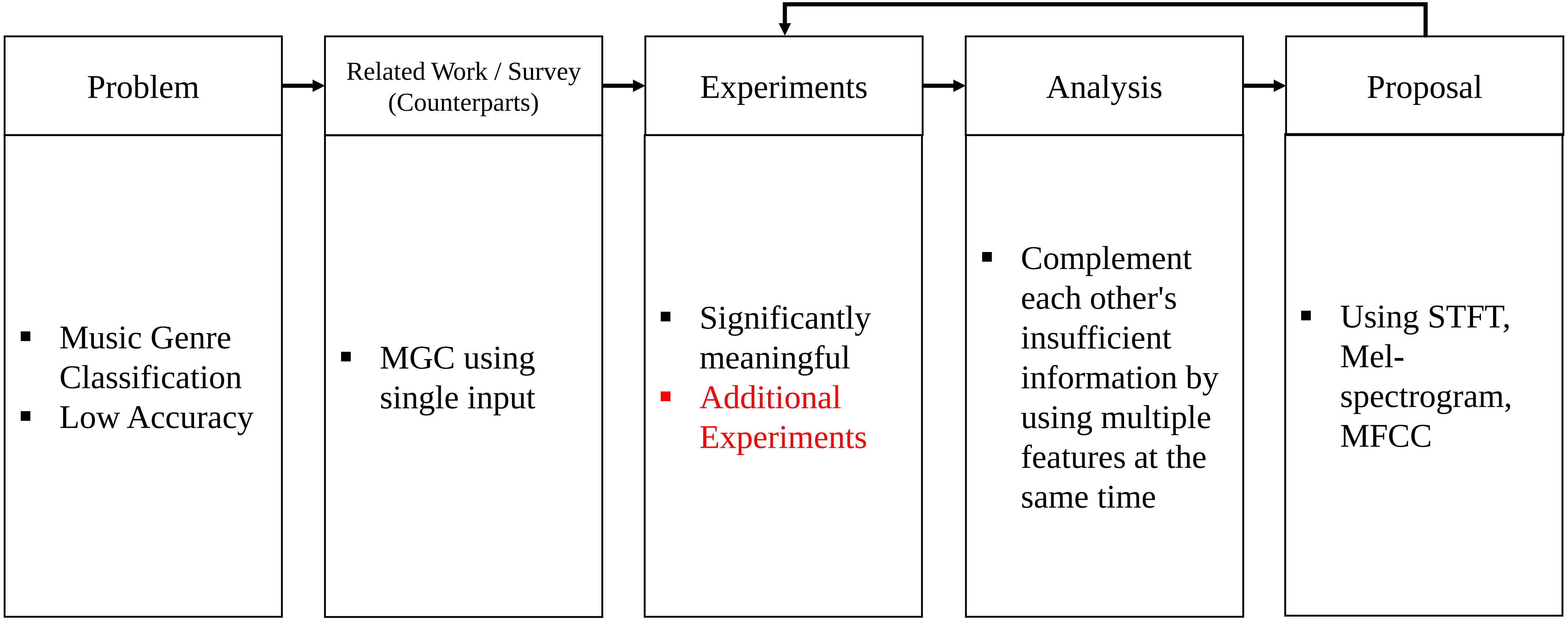
- Complement each other's insufficient information by using multiple features at the same time

Proposal

- Using STFT, Mel-spectrogram, MFCC

- The authors proposed **a new CNN model using three musical features** for music genre classification: Short-Time Fourier Transform, Mel-Spectrogram, and Mel-Frequency Cepstral Coefficient.
- The authors performed MGC using 12 datasets and the proposed model showed **the best performance on all 12 datasets**.
- **Sufficient recent studies related to the work** have been included in the references.
- **Sufficient results** have been shown to support the claims of the proposed work.
- It is recommended **that grammatical errors** are checked throughout the paper before the final submission.
- It will be **very interesting to see the presentation** of the work at the conference.

Overall Progress (IEEE Access & Graduation Paper)



Problem

- Music Genre Classification
- Low Accuracy

Related Work / Survey (Counterparts)

- MGC using single input

Experiments

- Significantly meaningful
- Additional Experiments

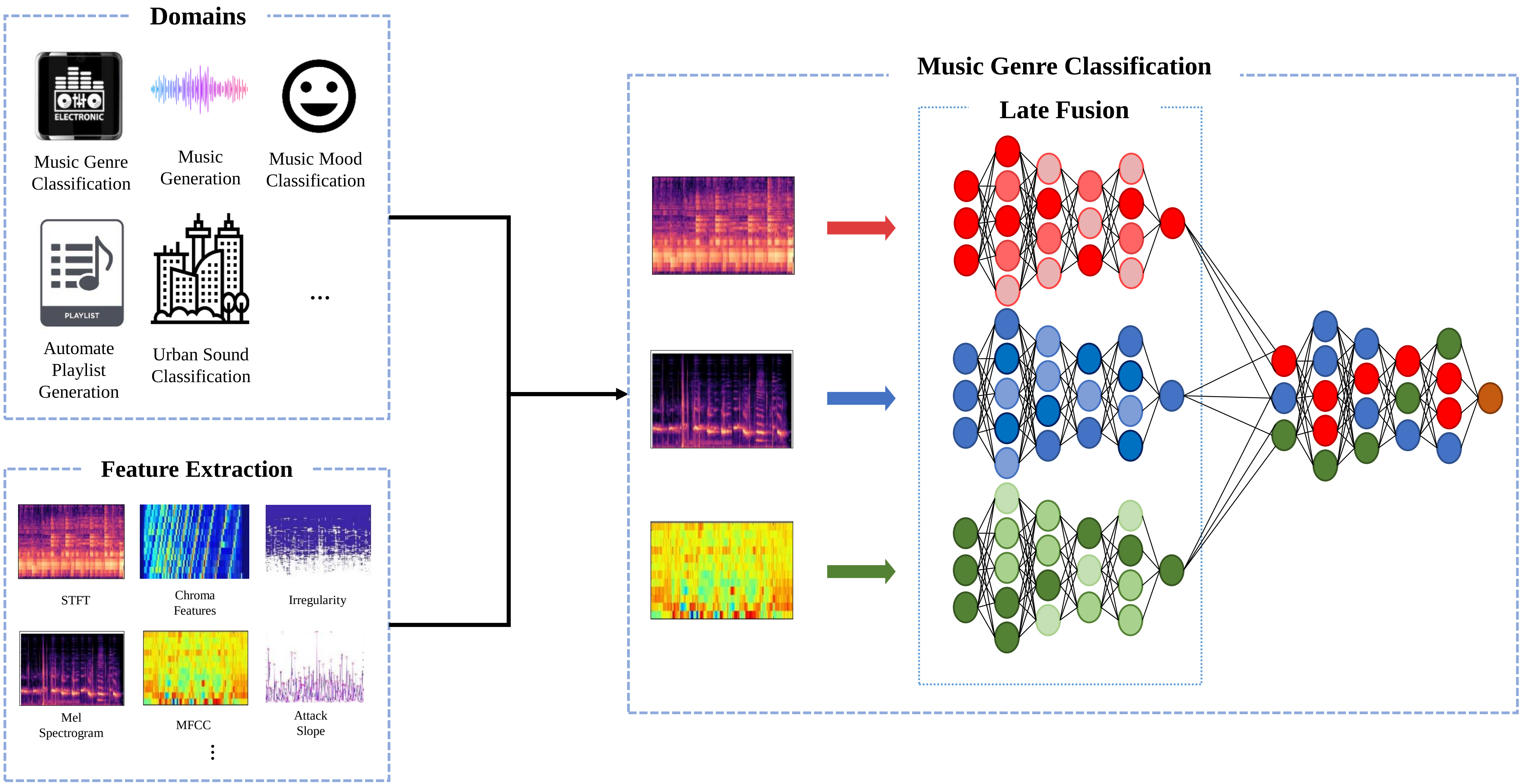
Analysis

- Complement each other's insufficient information by using multiple features at the same time

Proposal

- Using STFT, Mel-spectrogram, MFCC

Introduction



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

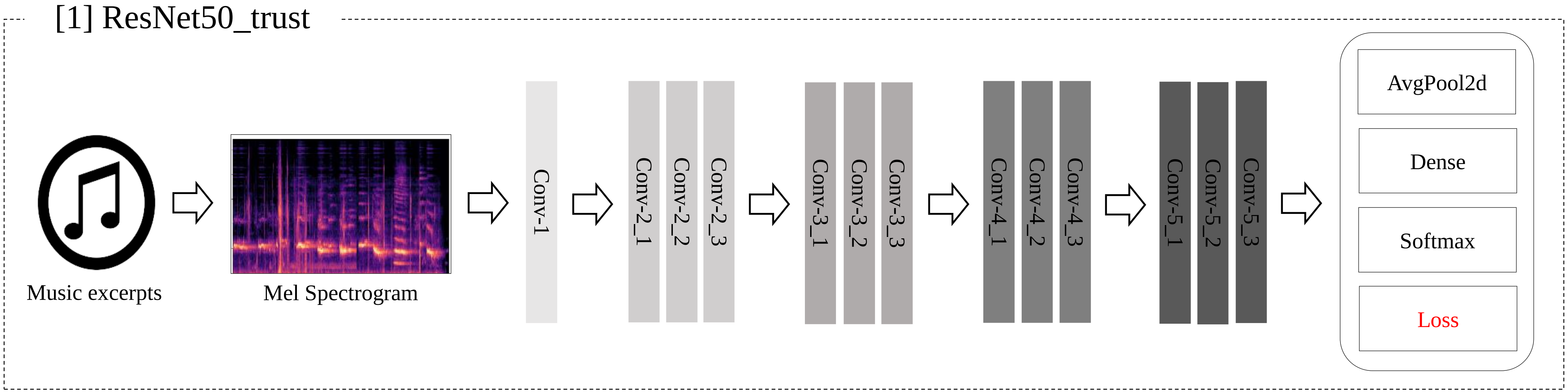
⋮

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] ResNet50_trust

Music excerpts

Mel Spectrogram

Conv-1

Conv-2_1

Conv-2_2

Conv-2_3

Conv-3_1

Conv-3_2

Conv-3_3

Conv-4_1

Conv-4_2

Conv-4_3

Conv-5_1

Conv-5_2

Conv-5_3

AvgPool2d

Dense

Softmax

Loss

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

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

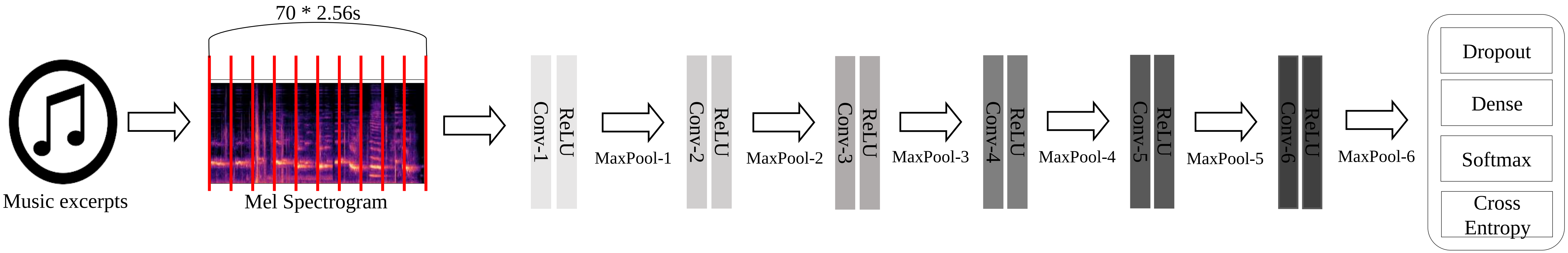
$\text{Loss1} = \text{Cross\_Entropy}(\text{y}_{\text{true}}(\text{Onehot encoding}), \text{y}_{\text{pred}}(\text{Softmax function}))$

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

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.

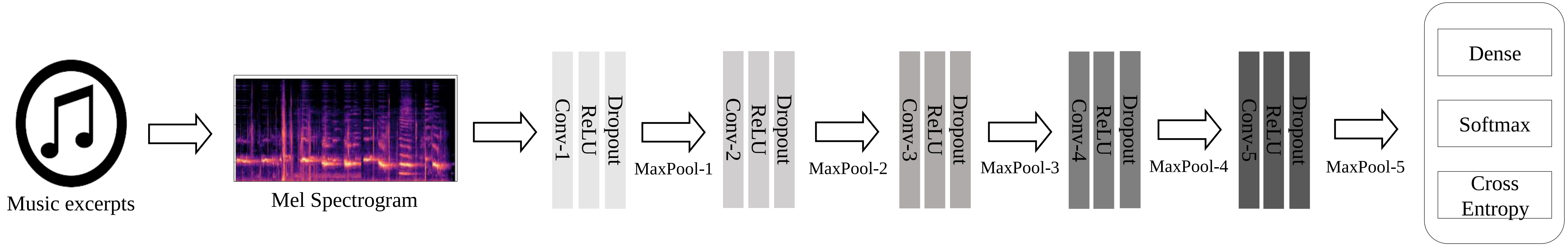
[2] PelchatNet (CNN)



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.

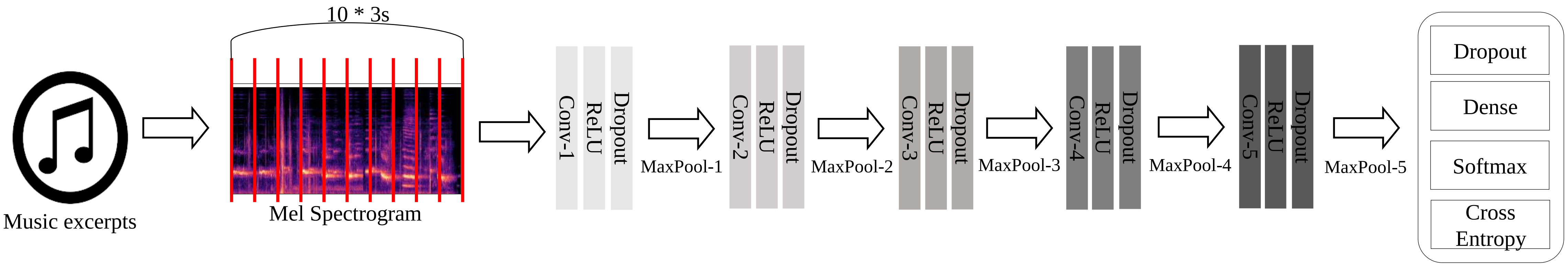
[3] ChengNet (CNN)



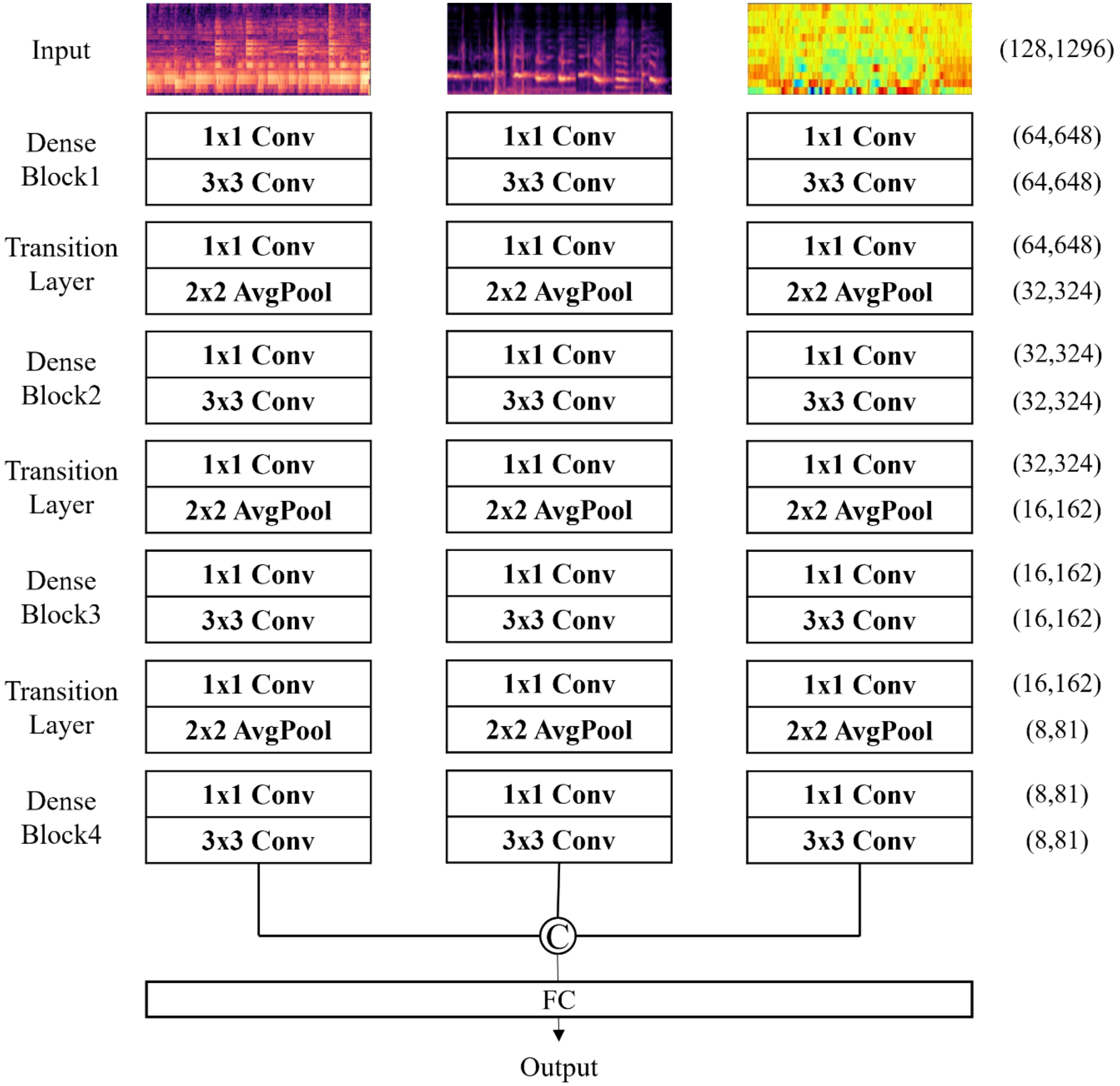
Related Work

[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] ShahNet (CNN)



Proposed Method



Input				(128,1296)
Dense Block1	1x1 Conv	1x1 Conv	1x1 Conv	(64,648)
	3x3 Conv	3x3 Conv	3x3 Conv	(64,648)
Transition Layer	1x1 Conv	1x1 Conv	1x1 Conv	(64,648)
	2x2 AvgPool	2x2 AvgPool	2x2 AvgPool	(32,324)
Dense Block2	1x1 Conv	1x1 Conv	1x1 Conv	(32,324)
	3x3 Conv	3x3 Conv	3x3 Conv	(32,324)
Transition Layer	1x1 Conv	1x1 Conv	1x1 Conv	(32,324)
	2x2 AvgPool	2x2 AvgPool	2x2 AvgPool	(16,162)
Dense Block3	1x1 Conv	1x1 Conv	1x1 Conv	(16,162)
	3x3 Conv	3x3 Conv	3x3 Conv	(16,162)
Transition Layer	1x1 Conv	1x1 Conv	1x1 Conv	(16,162)
	2x2 AvgPool	2x2 AvgPool	2x2 AvgPool	(8,81)
Dense Block4	1x1 Conv	1x1 Conv	1x1 Conv	(8,81)
	3x3 Conv	3x3 Conv	3x3 Conv	(8,81)

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
ResNet50_trust	64		Proposed Loss Function			
PelchatNet (CNN)	128		Cross Entropy Loss			
ChengNet (CNN)	32		Cross Entropy Loss			
ShahNet (CNN)	32		Cross Entropy Loss			

Performance Measure

$$Accuracy(\%) = \frac{TP + TN}{TP + TN + FP + FN} \times 100$$

TP : $True\ Positive$ TN : $True\ Negative$

FP : $False\ Positive$ FN : $False\ Negative$

Evaluates the overall effectiveness of a model

➡ Higher value, Higher Performance

Experimental Results

Model Comparison Based on Accuracy(*▼: $p<0.05$. **▼: $p<0.01$.)

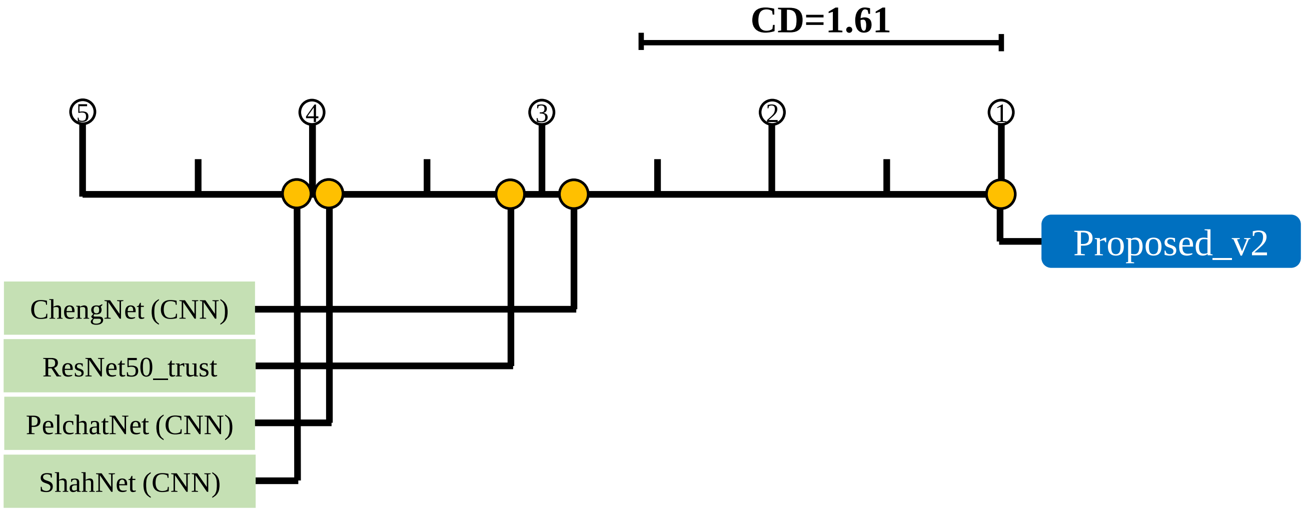
Dataset (Genre)	Proposed	ResNet50_trust	PelchatNet (CNN)	ChengNet (CNN)	ShahNet (CNN)
Ballroom	**56.79 ± 3.77**	45.71 ± 2.86**▼	42.43 ± 2.29**▼	48.07 ± 4.79**▼	47.71 ± 4.58**▼
Ballroom-Extended	**71.87 ± 2.28**	61.82 ± 3.79**▼	44.68 ± 1.30**▼	62.88 ± 4.70**▼	54.38 ± 5.33**▼
emoMusic-G	**34.16 ± 2.20**	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**▼
Total Win / Tie / Lose	**46/2/0**	**16/18/14**	**4/15/29**	**13/22/13**	**4/17/27**
Avg . Rank	**1.00**	3.17	3.92	2.83	4.08

Experimental Results

1. Friedman test

Evaluation measure	Friedman statistics	Critical values ($\alpha = 0.05$)
Accuracy	29.1333	7.3447e-06

2. Bonferroni-Dunn test



Accuracy

Conclusion

Conclusion

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

Contribution

In international Publication (SCI(E))

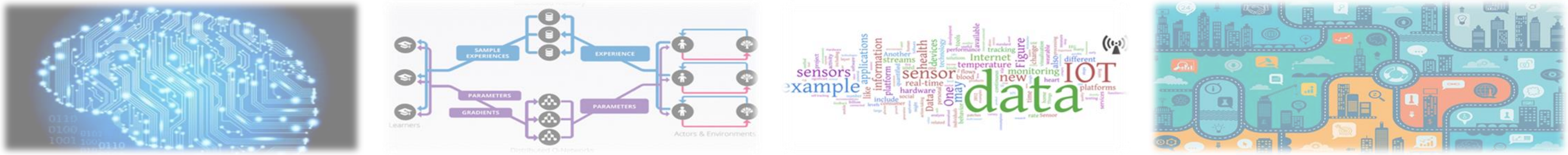
- Toward a Fair Evaluation and Analysis of Feature Selection for Music Tag Classification, ***IEEE Access,*** 2021, **Co-author**

In International Conferences

- An Efficient Neural Network based on Early Compression of Sparse CT Slice Images, ***PlatCon***, 2021, **Co-author**
- Effective Music Genre Classification using Late Fusion Convolutional Neural Network with Multiple Spectral Features ***ICCE-Asia***, 2022, **1^{st} author**

In Domestic Conferences

- Music Genre Classification Using Multiple Musical Visual Features, ***IEIE***, 2022 , **1^{st} author**
- Music Genre Classification using CNN architecture with feature concatenation, ***KISS Autumn Conf***, 2021, **Co-author**

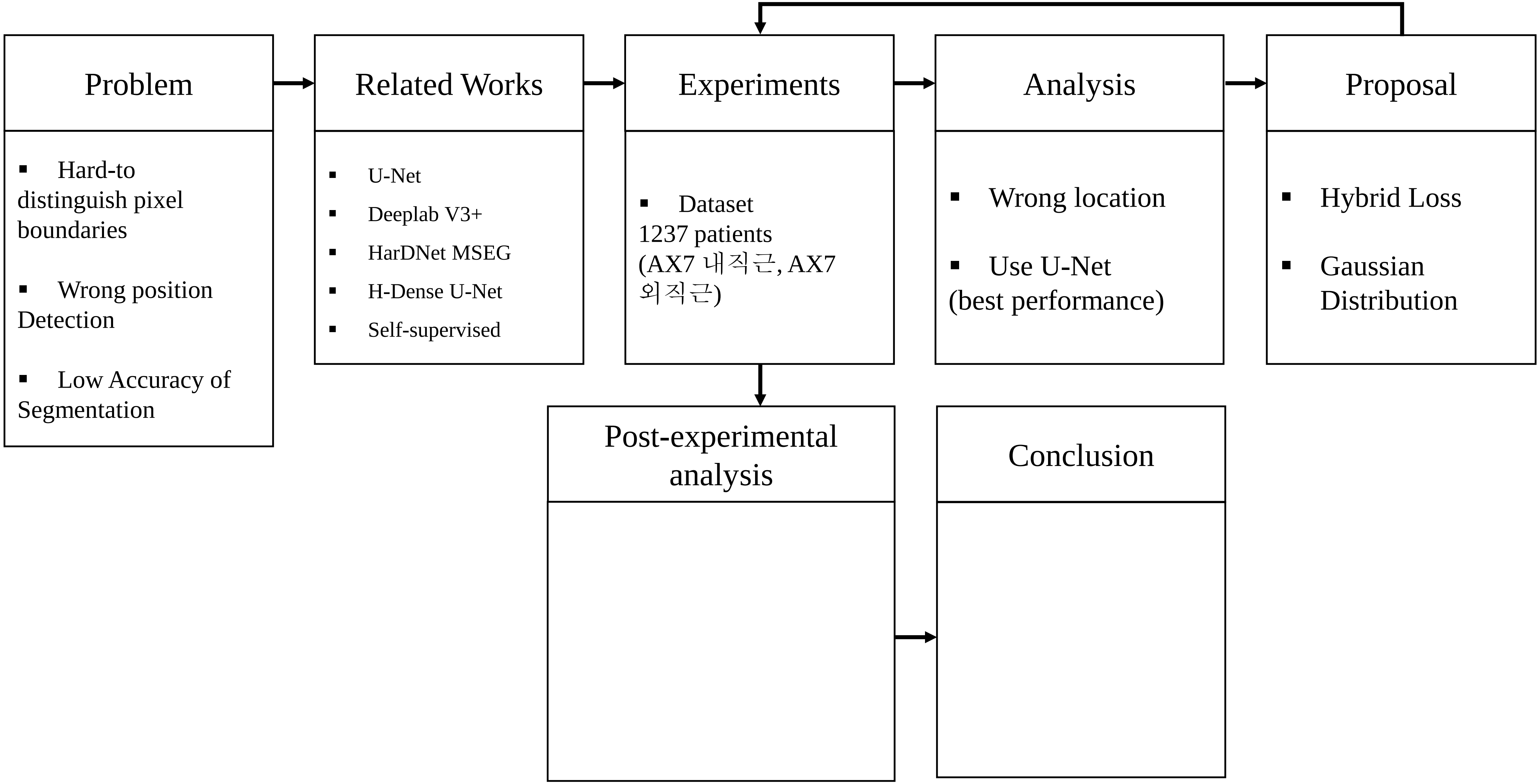


Individual Study

Su-jeong Han

2022. 09. 14

Overall Progress



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)

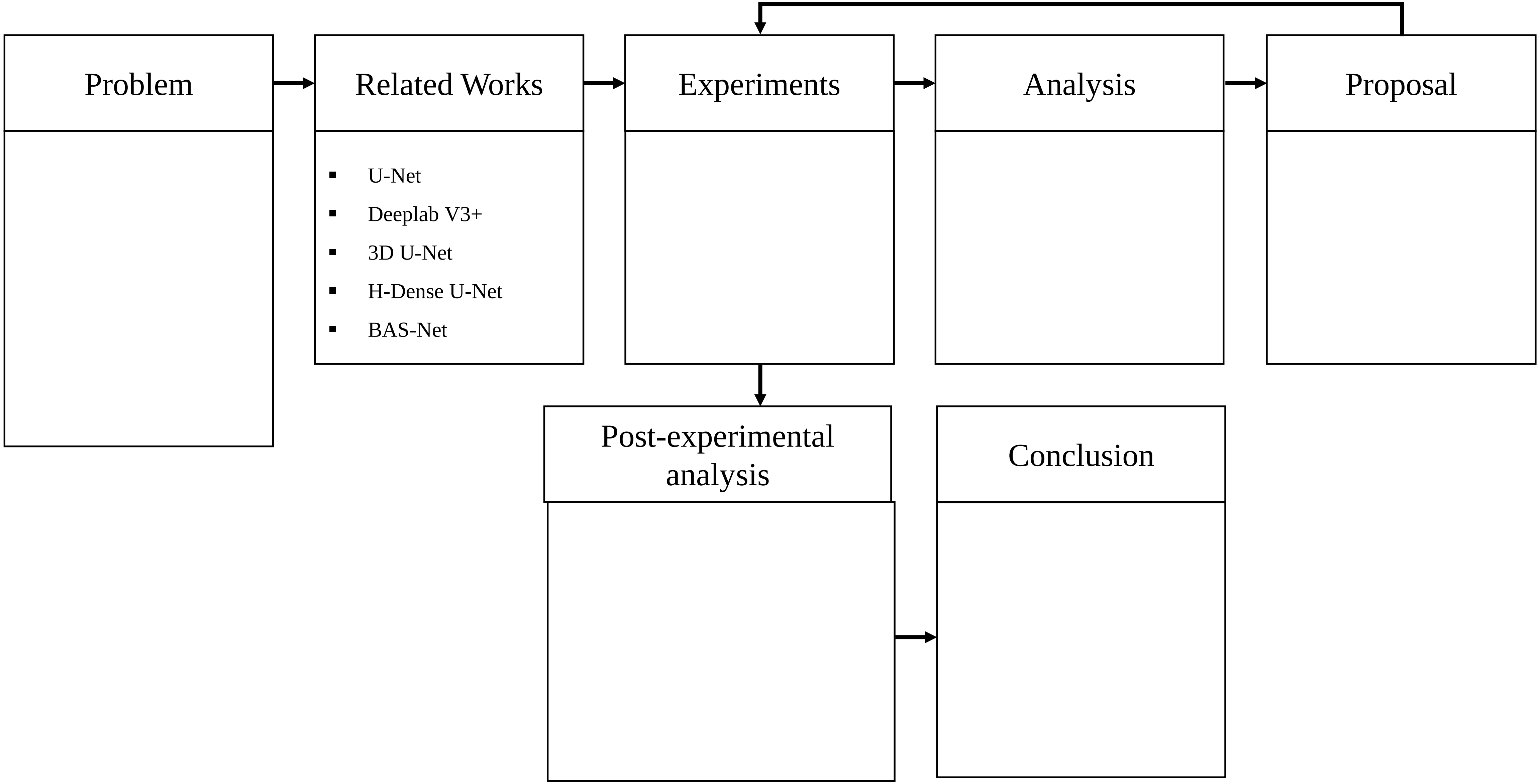
Proposal

- 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

Analysis

Proposal

Post-experimental analysis

Conclusion

Eyeball Volume Estimation

➢ Datasets

- 5 images from Ax1 to Ax5 for each patients
- Number of Samples : 1,237 * 5 images

- 총 환자의 수 : 1,237명
- 사람이 측정한 안구의 부피 : 532명

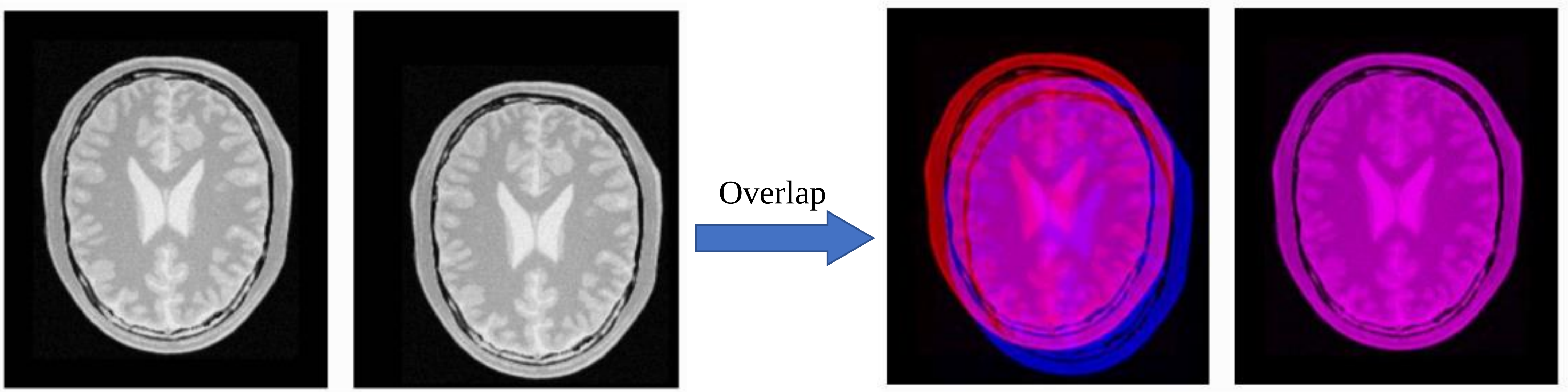
- Input : 5 eyeball segmentation images from Ax1 to Ax5 for each patients
- Output : Eyeball Volume

Individual Study

Experimental Setting

- 5 images from Ax1 to Ax5 for each patients
- Number of Samples : 2,605 images (521 patients)

- Input : 5 eyeball segmentation images from Ax1 to Ax5 for each patients
- Output : Eyeball volume (Before and After Surgery)



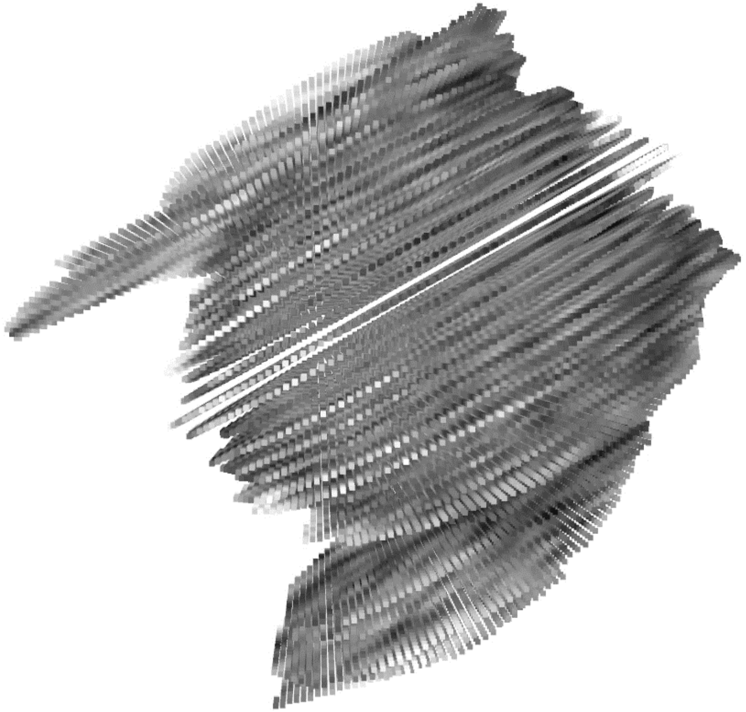
Individual Study

Next Progress

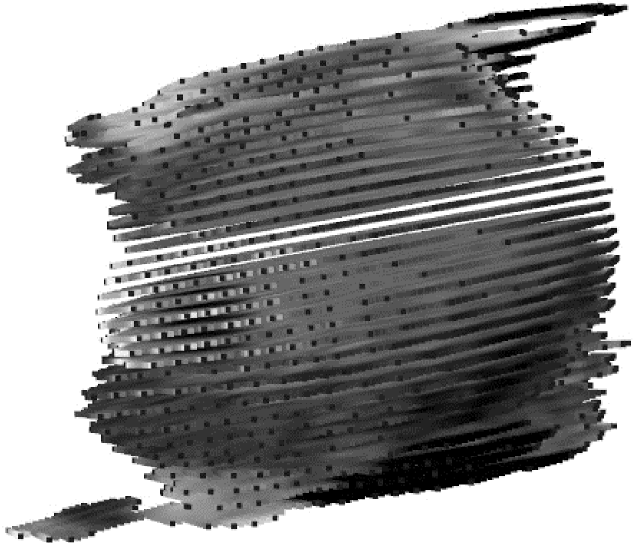
- Segmentation 모델 성능 비교
- Segmentation 모델 구현
- 수술 전후 CT이미지 비교

3D Segmentation

1075 환자 수동 세그멘테이션 시각화



101026 환자 자동 세그멘테이션 시각화



갑상선 연구 데이터셋 구축 현황

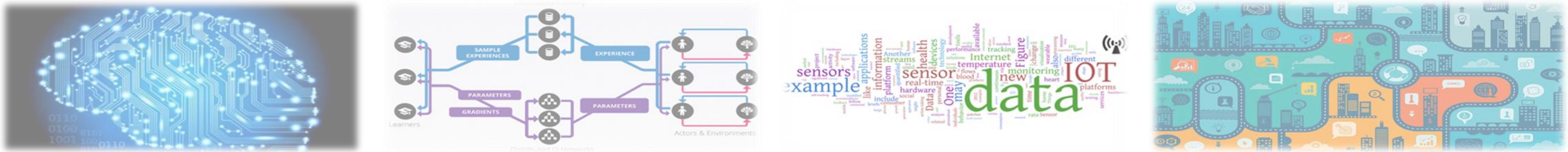
Progress

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 진행중 : **눈_left, 눈_right**

검수 완료 : 내직근 (left, right) , 외직근 (left, right)

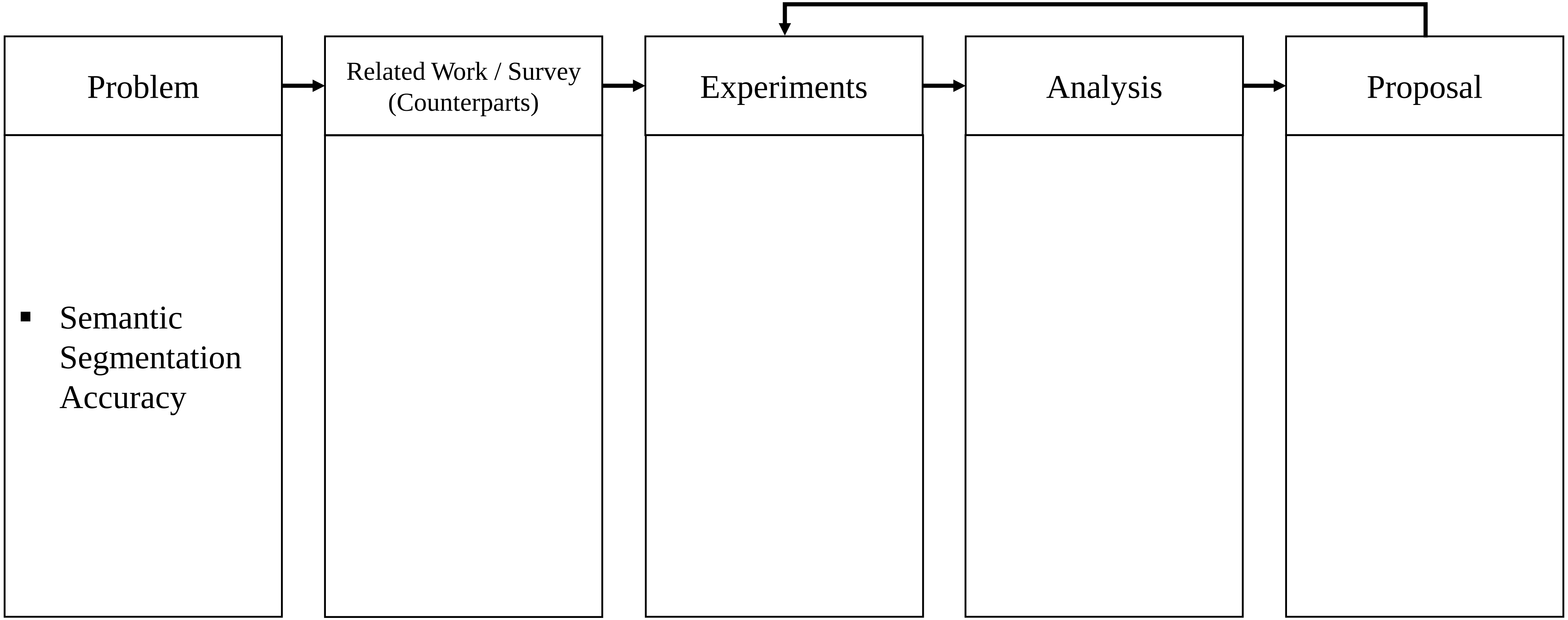


Individual Studies

Taekyung Song

2022. 09. 28

Overall Progress



Problem

Related Work / Survey (Counterparts)

Experiments

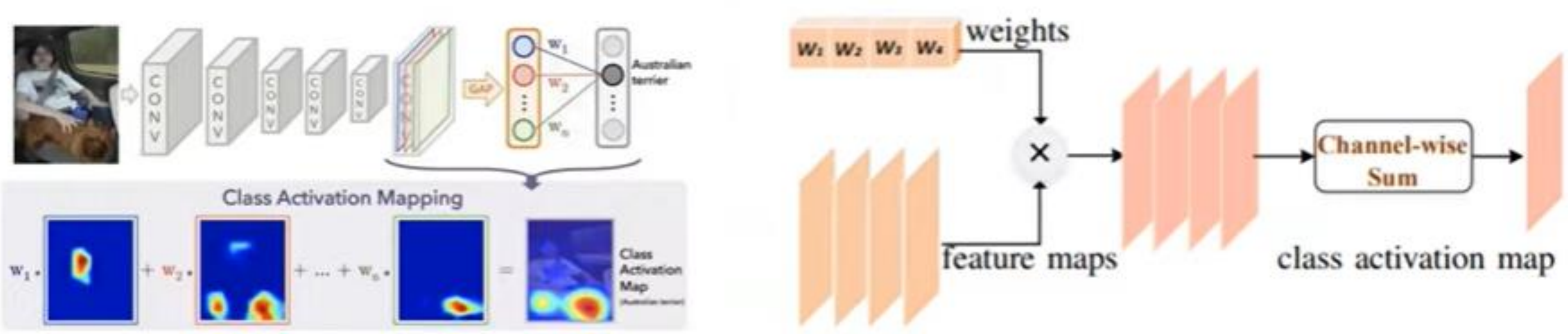
Analysis

Proposal

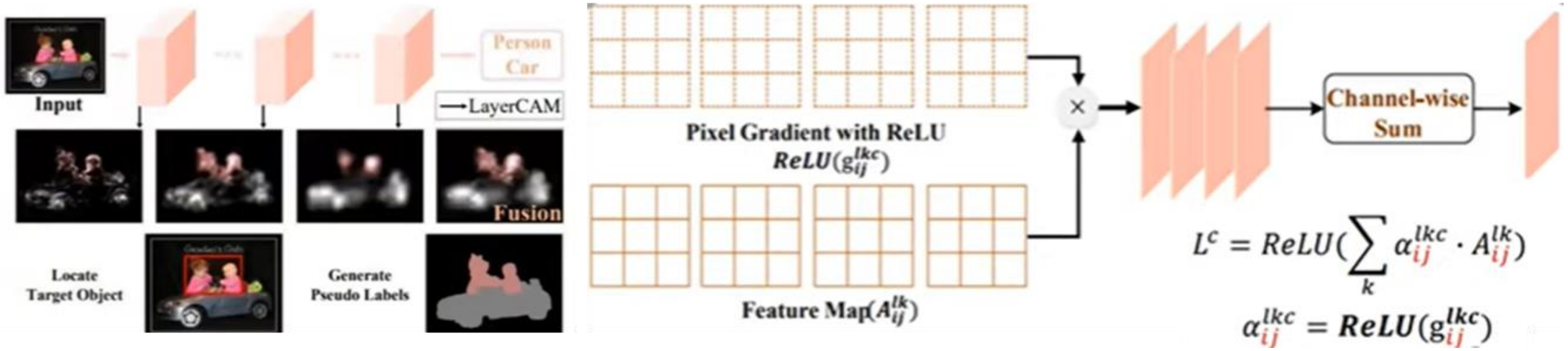
- Semantic Segmentation Accuracy

Individual Study

Layer-CAM



< Concept of Class Activation Maps, CAM >

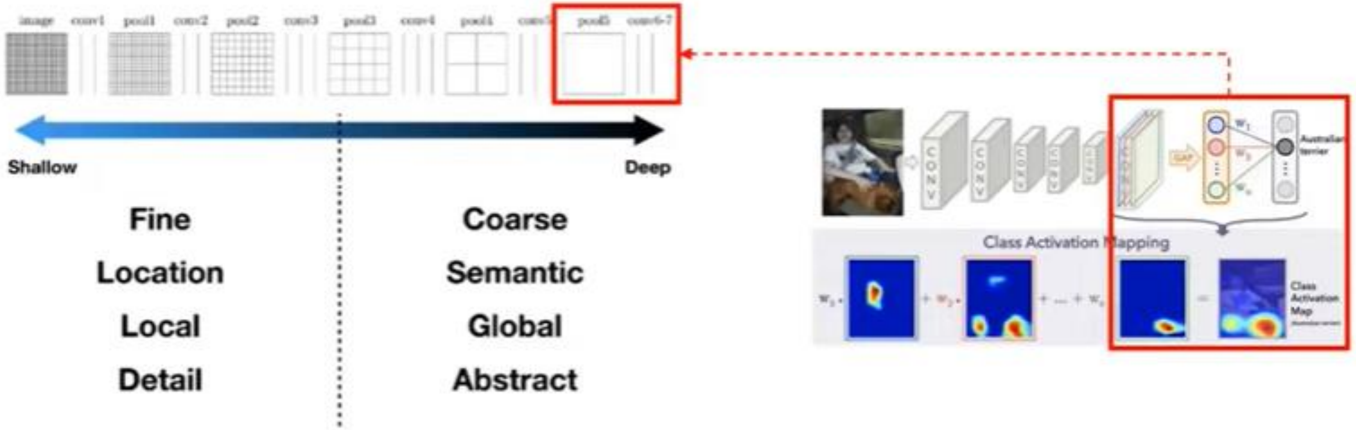


< Concept of LayerCAM >

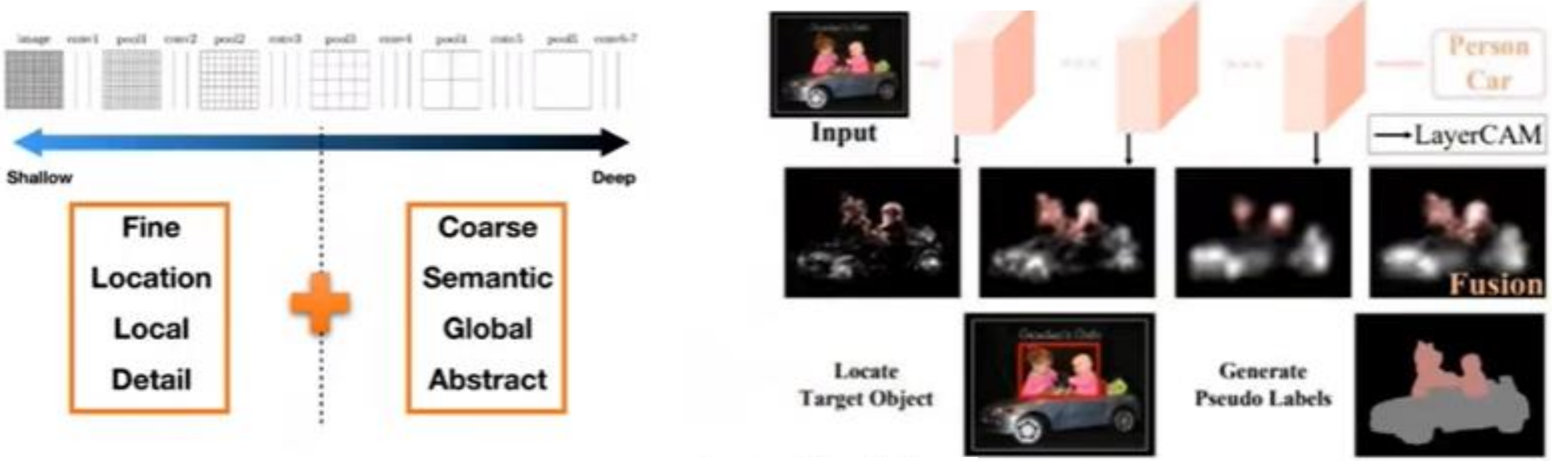
Zhou, Bolei, et al. "Learning deep features for discriminative localization." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
Selvaraju, Ramprasaath R., et al. "Grad-cam: Visual explanations from deep networks via gradient-based localization." *Proceedings of the IEEE international conference on computer vision*. 2017.
Jiang, Peng-Tao, et al. "Layercam: Exploring hierarchical class activation maps for localization." IEEE Transactions on Image Processing 30 (2021): 5875-5888.

Individual Study

Layer-CAM



< Characteristics of each layer in CNN >

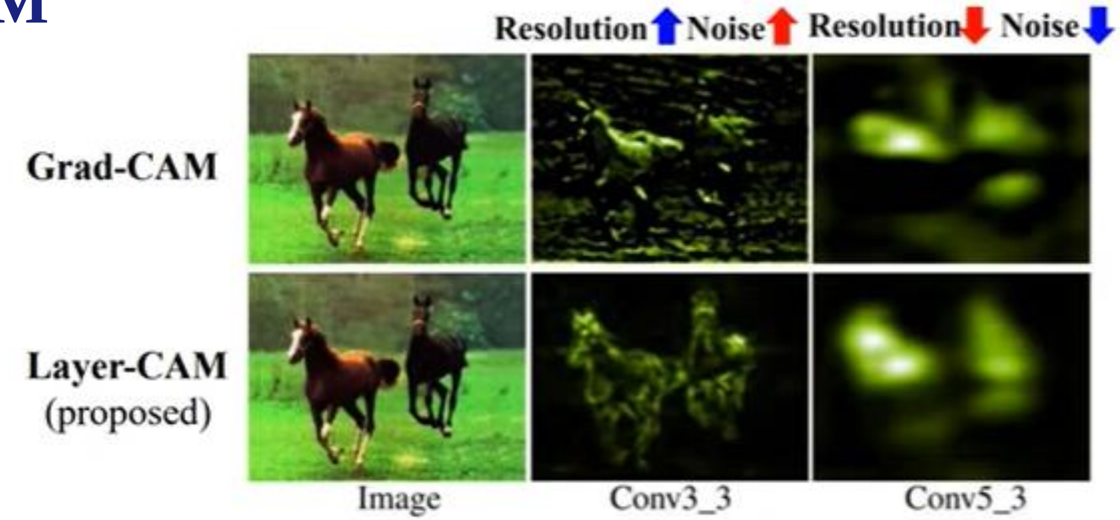


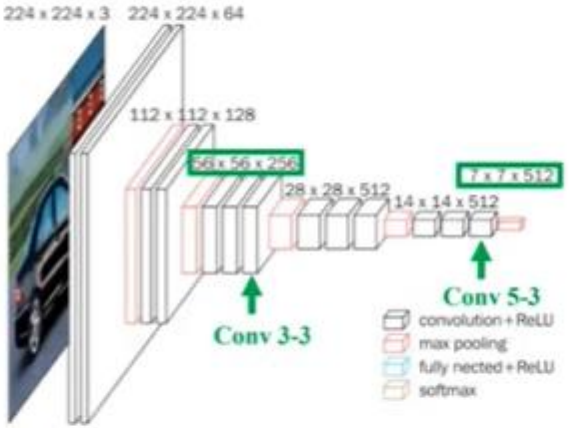
< Concept of LayerCAM >

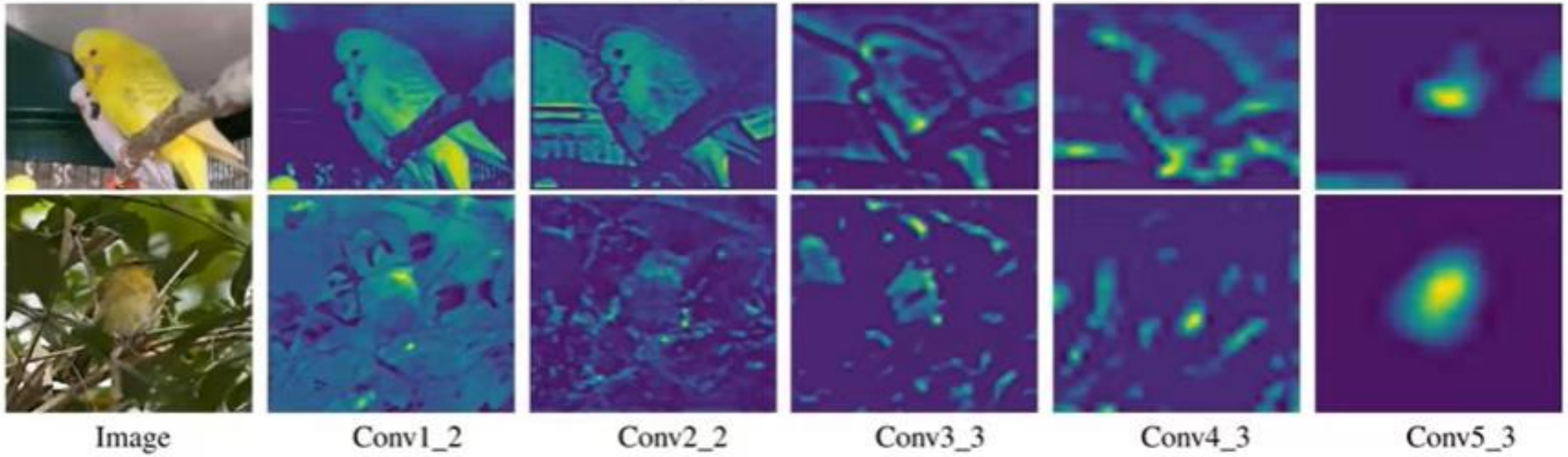
Zhou, Bolei, et al. "Learning deep features for discriminative localization." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
Selvaraju, Ramprasaath R., et al. "Grad-cam: Visual explanations from deep networks via gradient-based localization." *Proceedings of the IEEE international conference on computer vision*. 2017.
Jiang, Peng-Tao, et al. "Layercam: Exploring hierarchical class activation maps for localization." IEEE Transactions on Image Processing 30 (2021): 5875-5888.

Individual Study

Layer-CAM





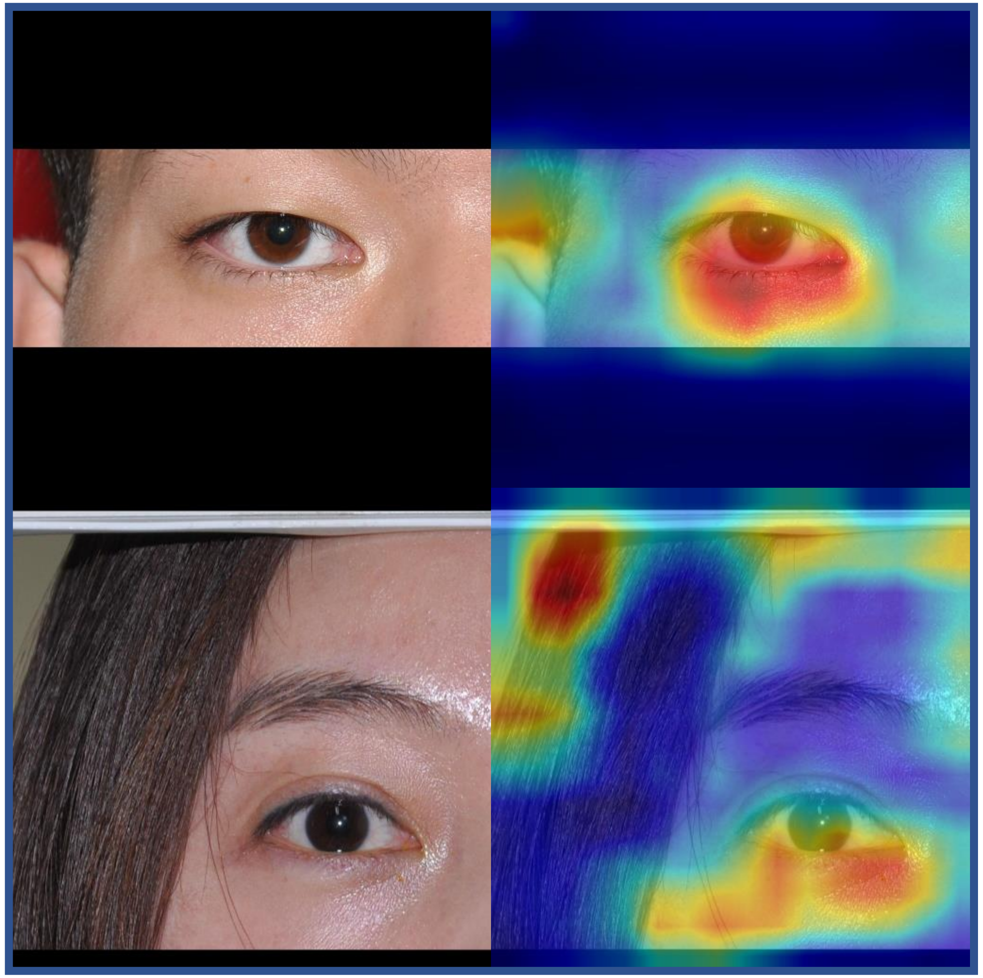


Jiang, Peng-Tao, et al. "Layercam: Exploring hierarchical class activation maps for localization." IEEE Transactions on Image Processing 30 (2021): 5875-5888.

Individual Study

Layer-CAM

< PASCAL VOC 2012 >



< TAO Patient >

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

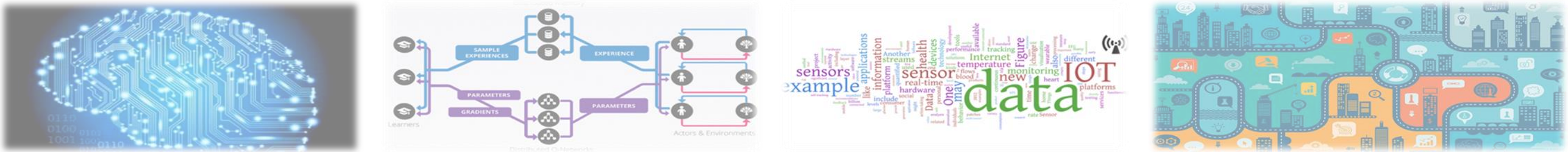
Appendix

References

[1] Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (**2016**). Learning deep features for discriminative localization. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2921-2929).

[2] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (**2017**). Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision* (pp. 618-626).

[3] Jiang, P. T., Zhang, C. B., Hou, Q., Cheng, M. M., & Wei, Y. (**2021**). Layercam: Exploring hierarchical class activation maps for localization. *IEEE Transactions on Image Processing*, *30*, 5875-5888.



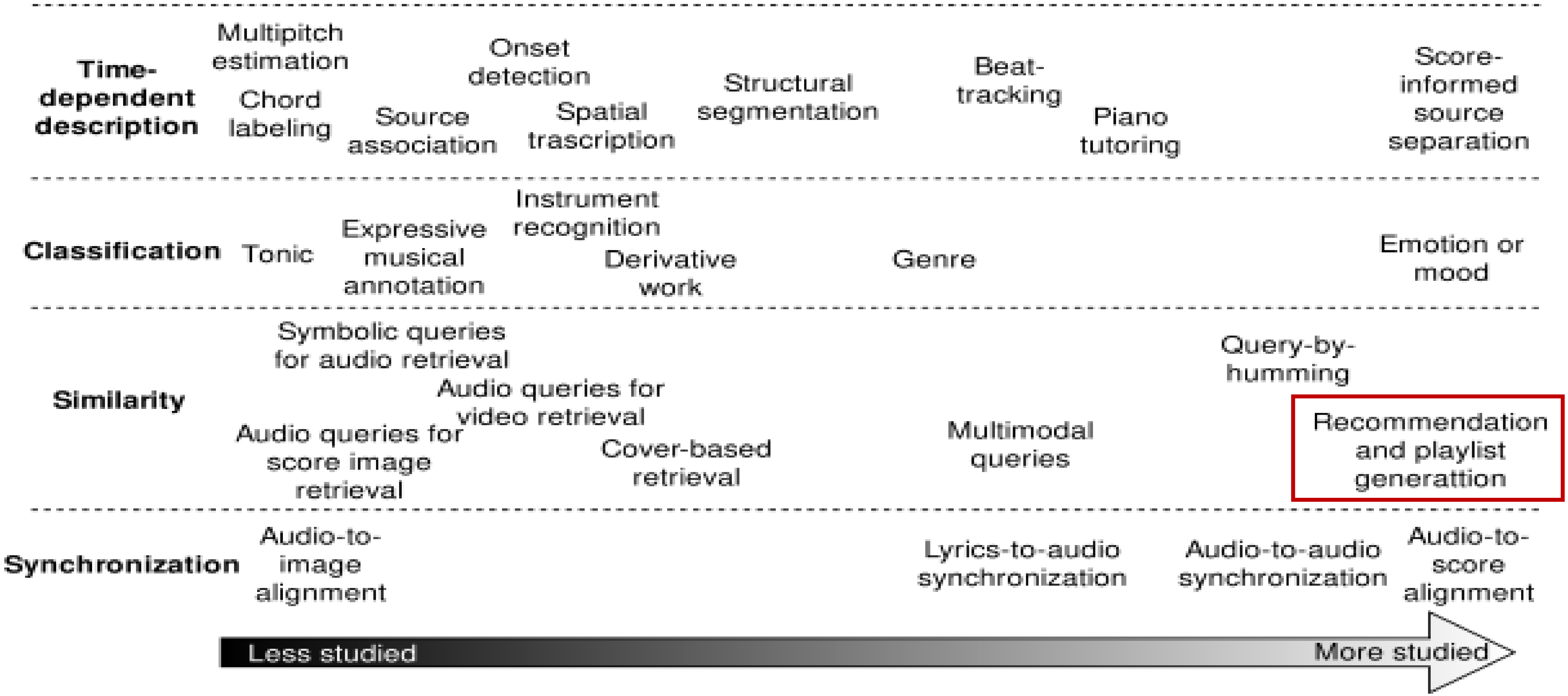
Sequential Skip Prediction
Using Multi-Branch 1D ResNet

Sanghyeong Jin

2022. 09. 28.

Introduction

Music Information Retrieval



Time-dependent description
Multipitch estimation
Chord labeling
Source association
Onset detection
Spatial trascription
Structural segmentation
Beat-tracking
Piano tutoring
Score-informed source separation

Classification
Tonic
Expressive musical annotation
Instrument recognition
Derivative work
Genre
Emotion or mood

Similarity
Symbolic queries for audio retrieval
Audio queries for score image retrieval
Audio queries for video retrieval
Cover-based retrieval
Multimodal queries
Query-by-humming
Recommendation and playlist generattion

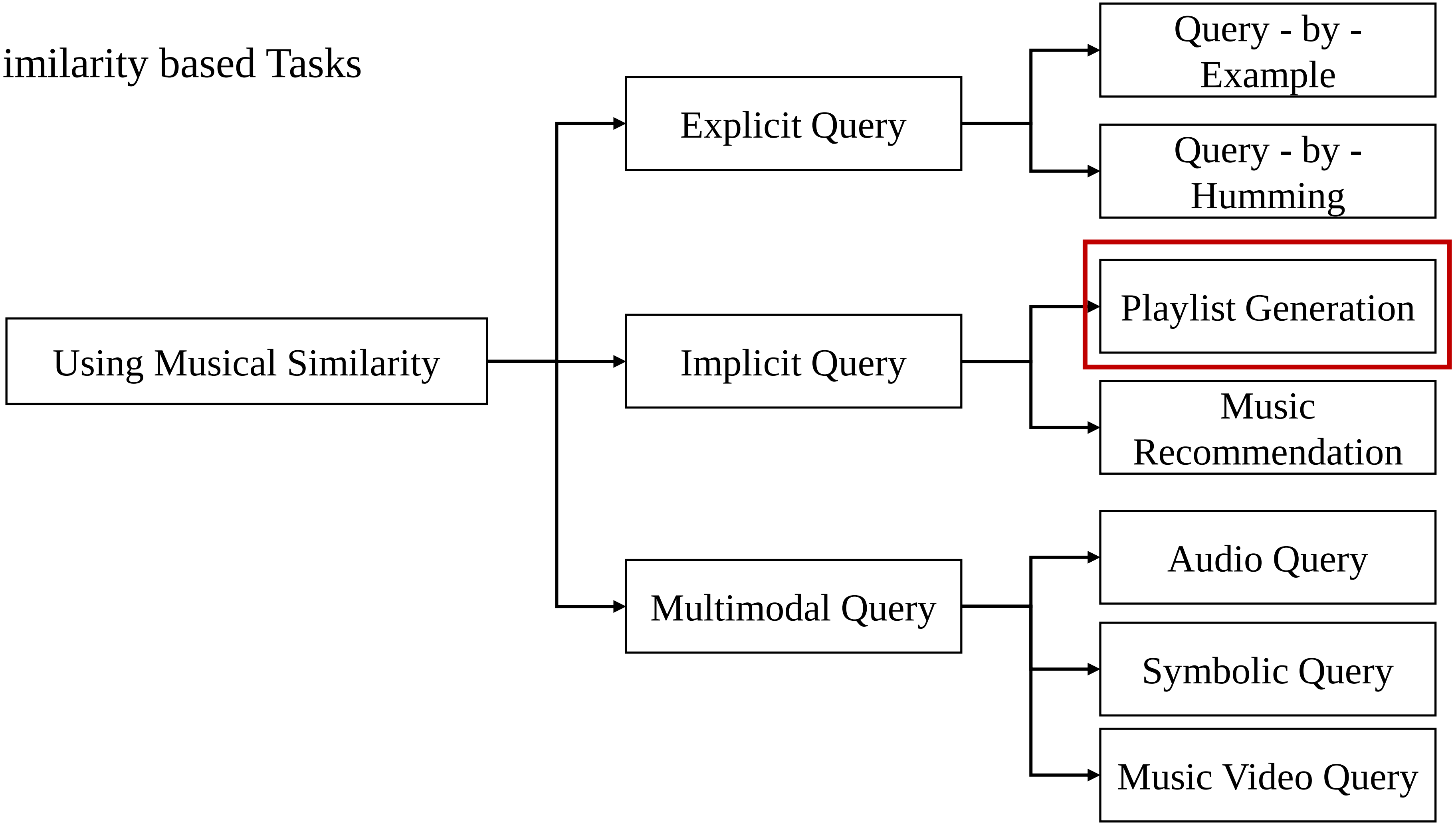
Synchronization
Audio-to-image alignment
Lyrics-to-audio synchronization
Audio-to-audio synchronization
Audio-to-score alignment

Less studied
More studied

Introduction

Music Information Retrieval

- Musical Similarity based Tasks



Using Musical Similarity

Explicit Query

Query - by - Example

Query - by - Humming

Implicit Query

Playlist Generation

Music Recommendation

Multimodal Query

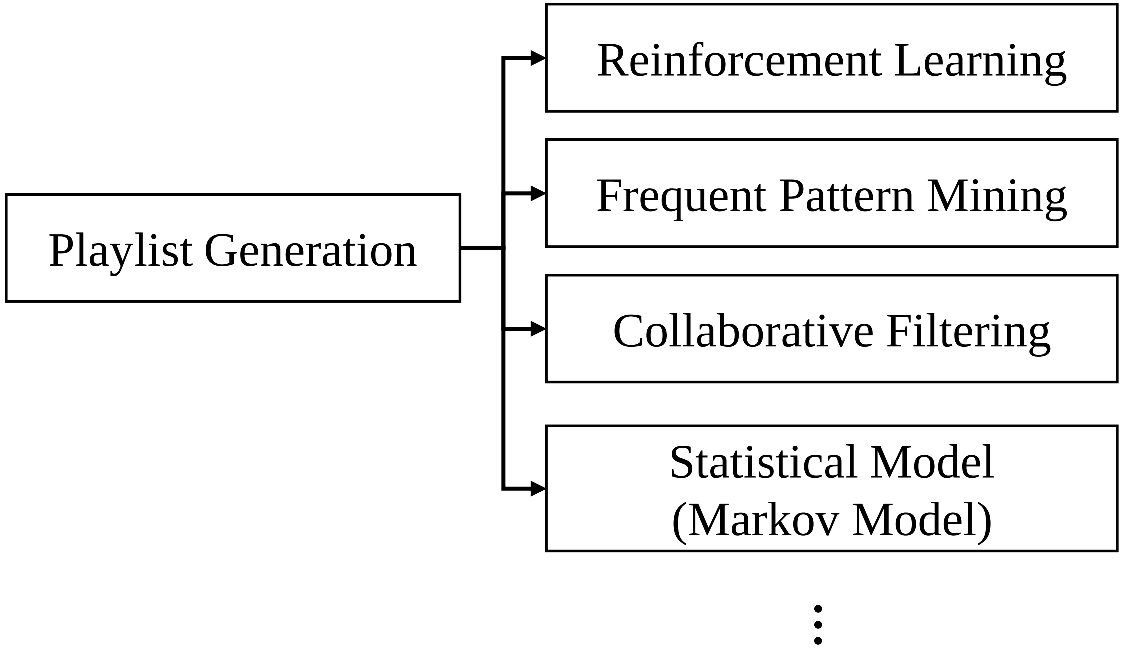
Audio Query

Symbolic Query

Music Video Query

Introduction

Playlist Generation



Playlist Generation

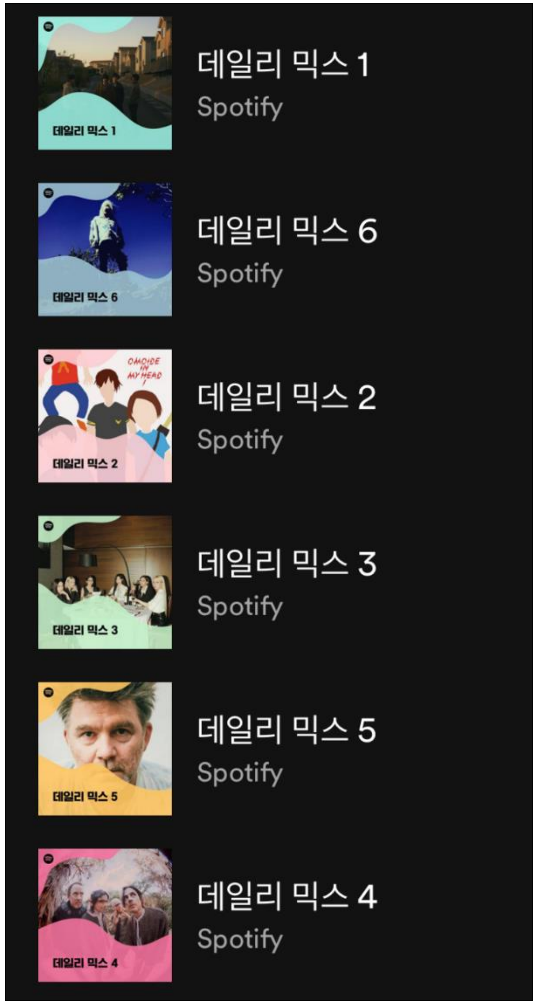
Reinforcement Learning

Frequent Pattern Mining

Collaborative Filtering

Statistical Model
(Markov Model)

⋮



데일리 믹스 1
Spotify

데일리 믹스 6
Spotify

데일리 믹스 2
Spotify

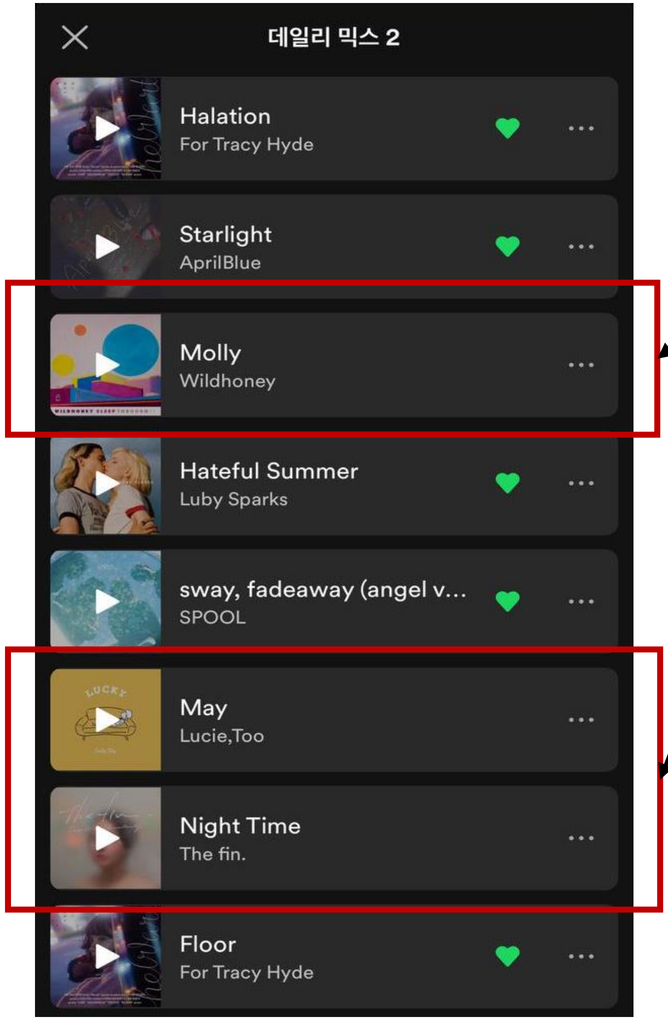
데일리 믹스 3
Spotify

데일리 믹스 5
Spotify

데일리 믹스 4
Spotify

Introduction

Skip Prediction



데일리 믹스 2

Halation
For Tracy Hyde

Starlight
AprilBlue

Molly
Wildhoney

Hateful Summer
Luby Sparks

sway, fadeaway (angel v...
SPOOL

May
Lucie,Too

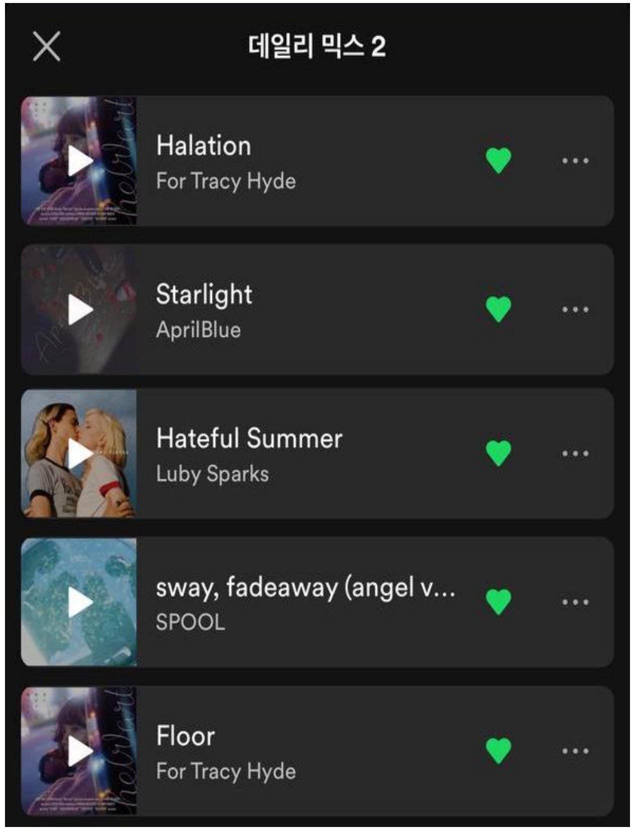
Night Time
The fin.

Floor
For Tracy Hyde

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



데일리 믹스 2

Halation
For Tracy Hyde

Starlight
AprilBlue

Hateful Summer
Luby Sparks

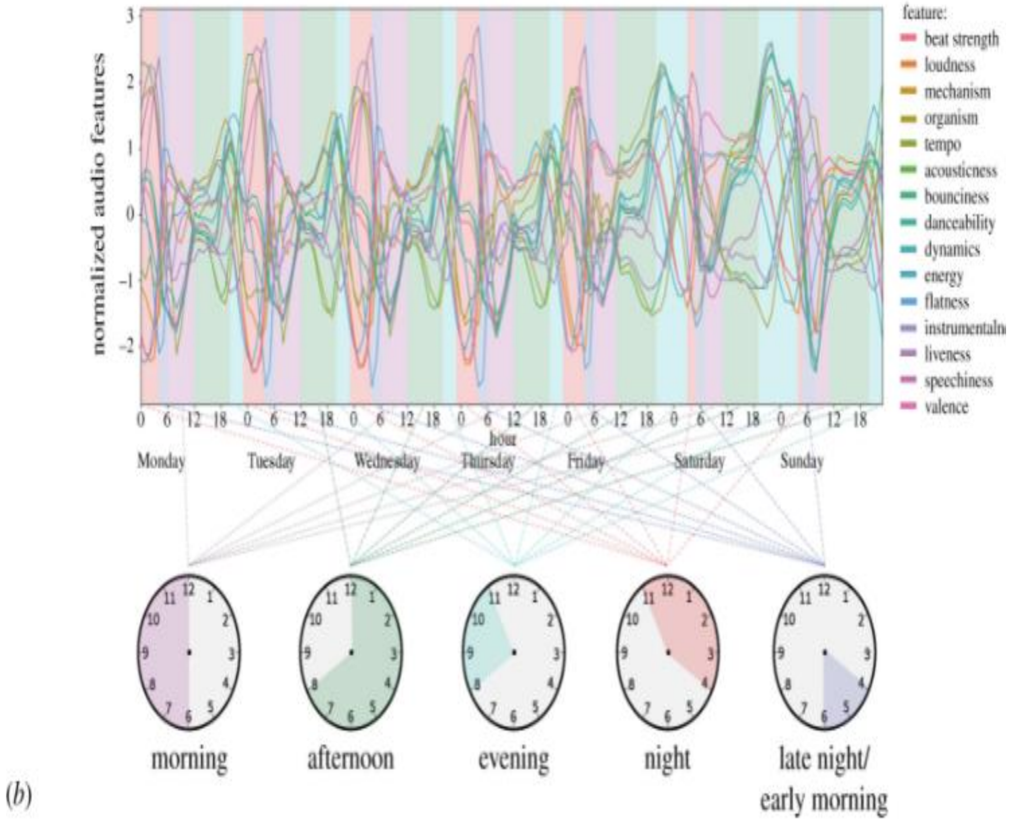
sway, fadeaway (angel v...
SPOOL

Floor
For Tracy Hyde

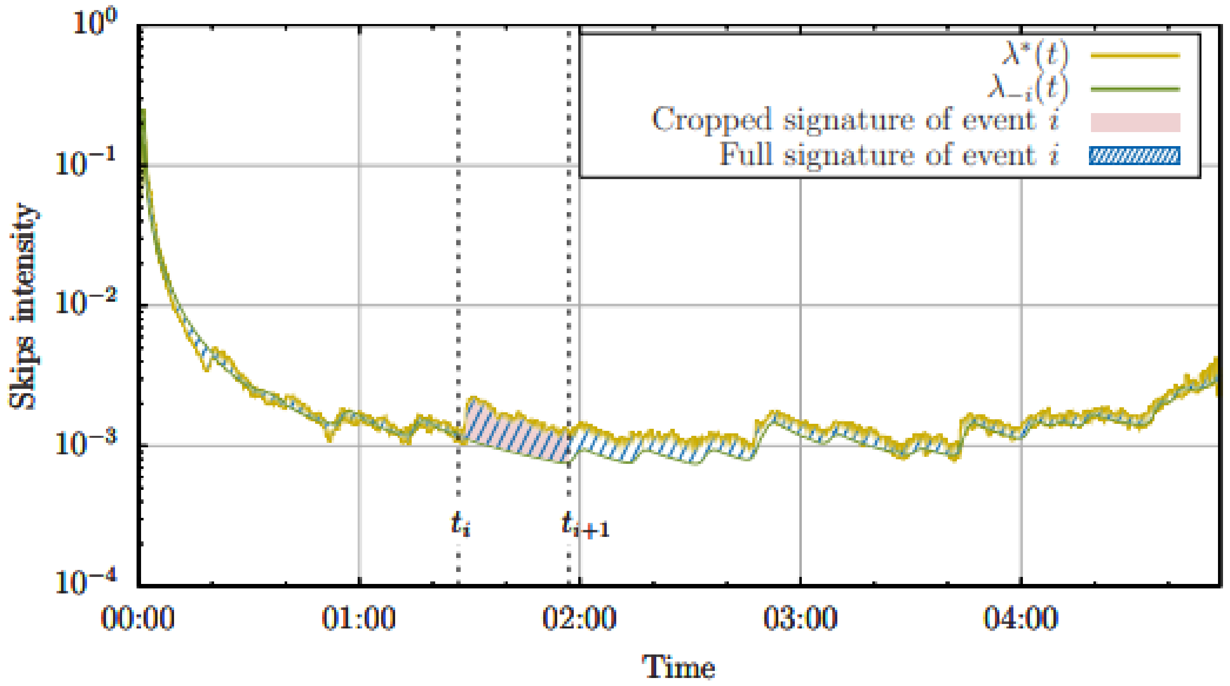
User preferred playlist

Related Work

Track Skip Behavior



Musical diurnal fluctuations

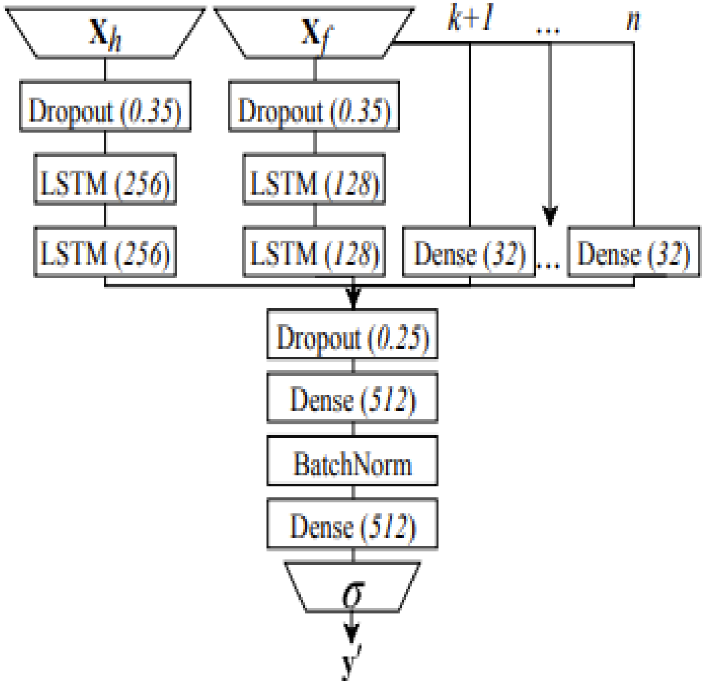


Skip rate based on time

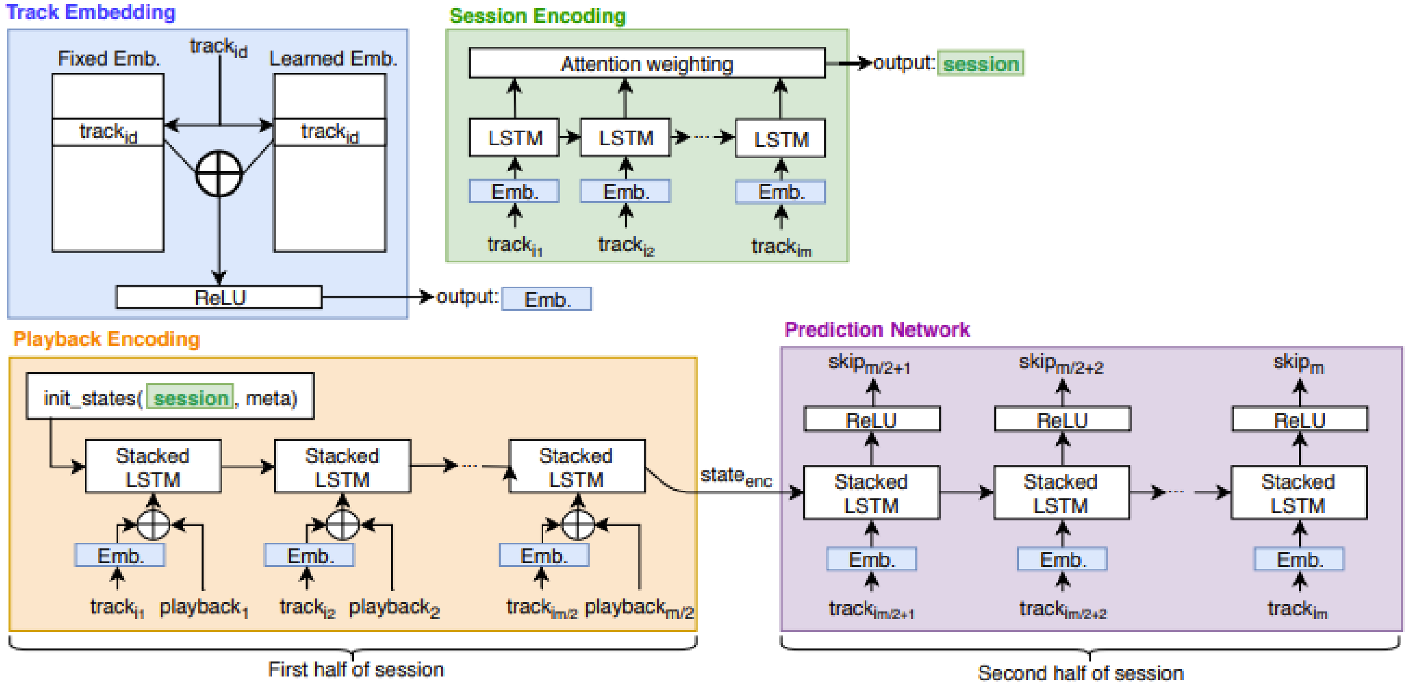
Related Work

Skip Prediction

- Mostly use RNN based Model



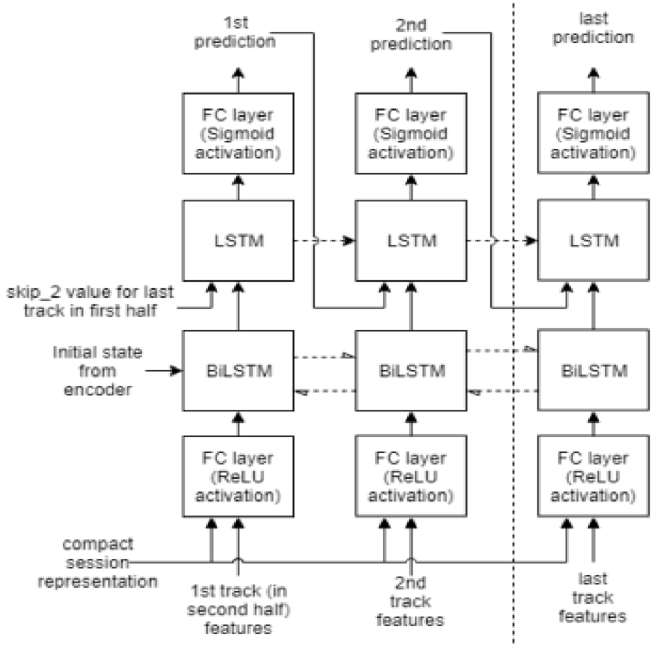
RNN-Dense



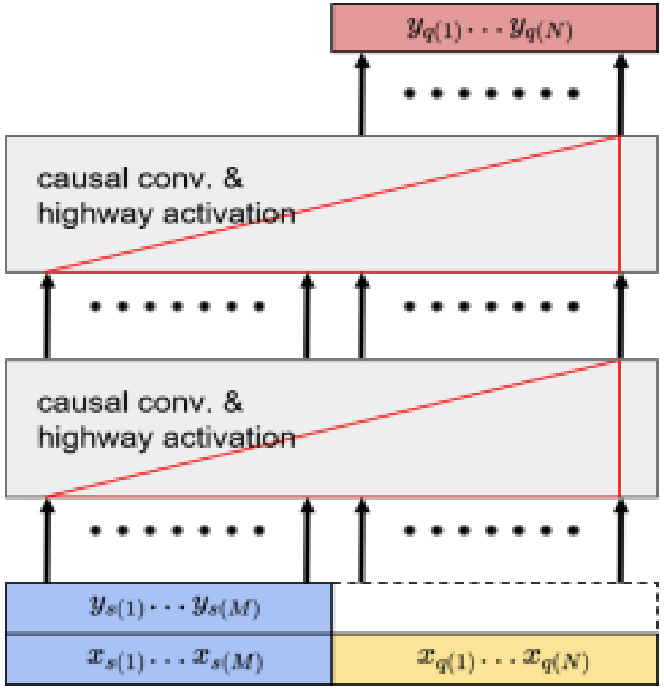
Multi-RNN

Related Work

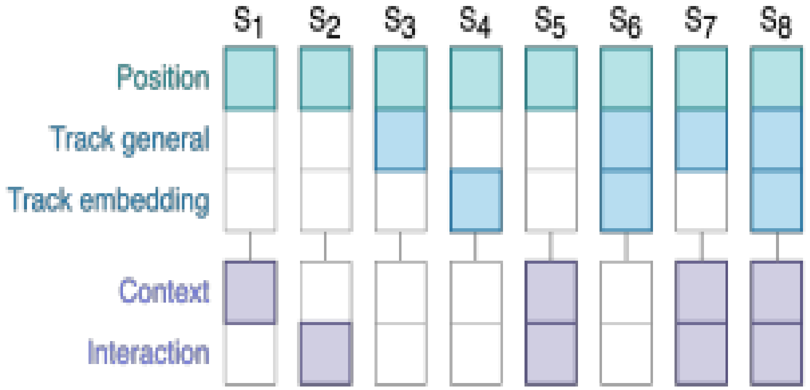
Skip Prediction



BiLSTM Encoder-Decoder



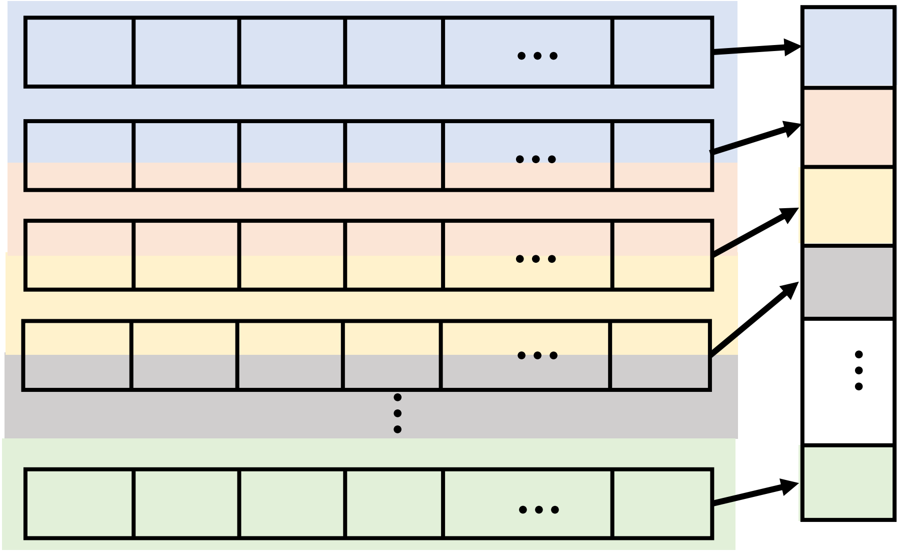
1D convolution with Highway Network



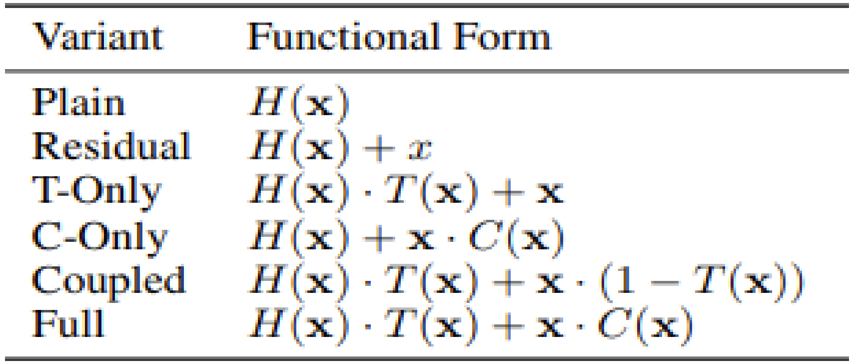
Interpretable Transformer

Related Work

1D ResNet

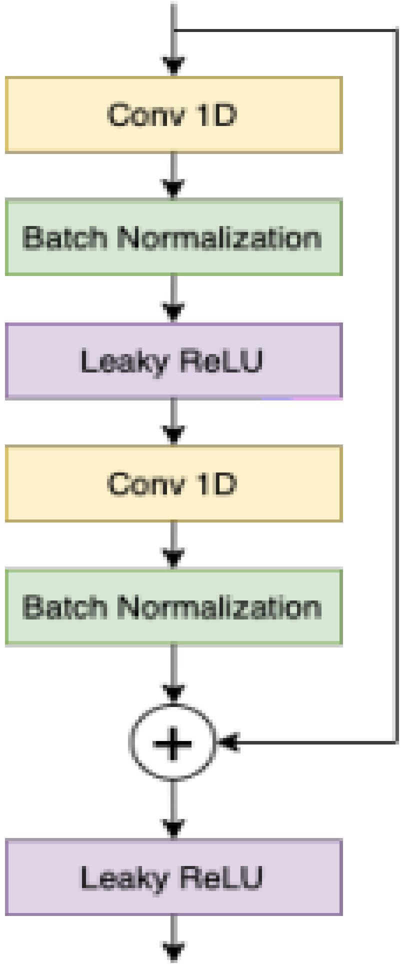


1D Convolution



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

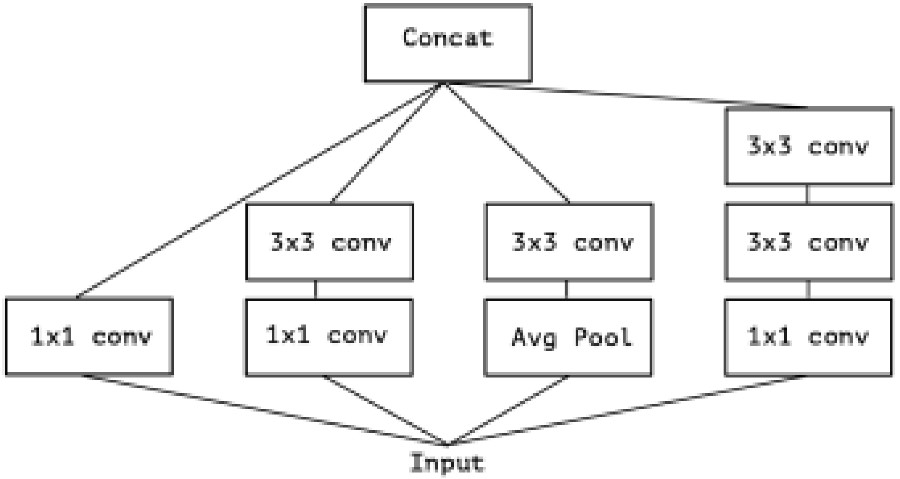
Residual Connection



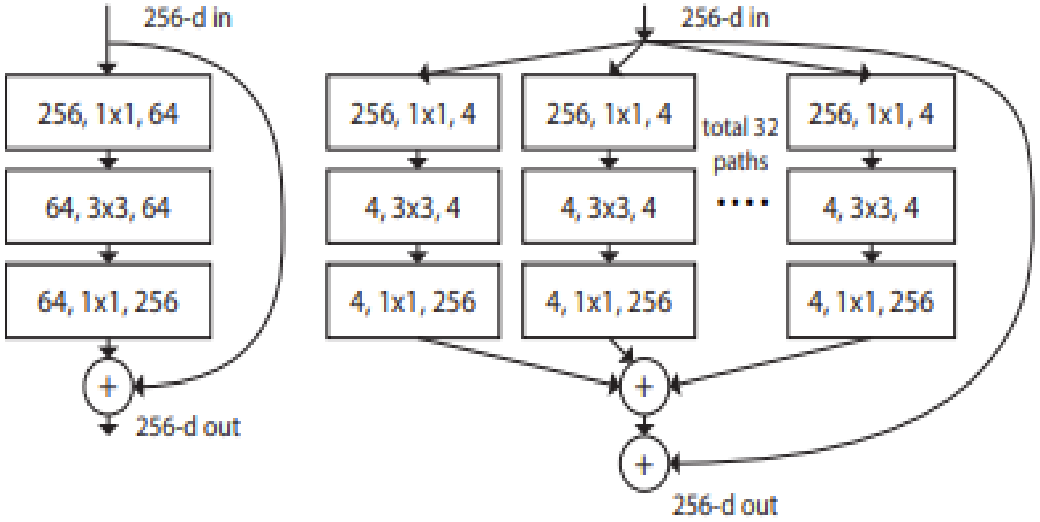
1D Residual Block

Related Work

Multi-Branch Convolution



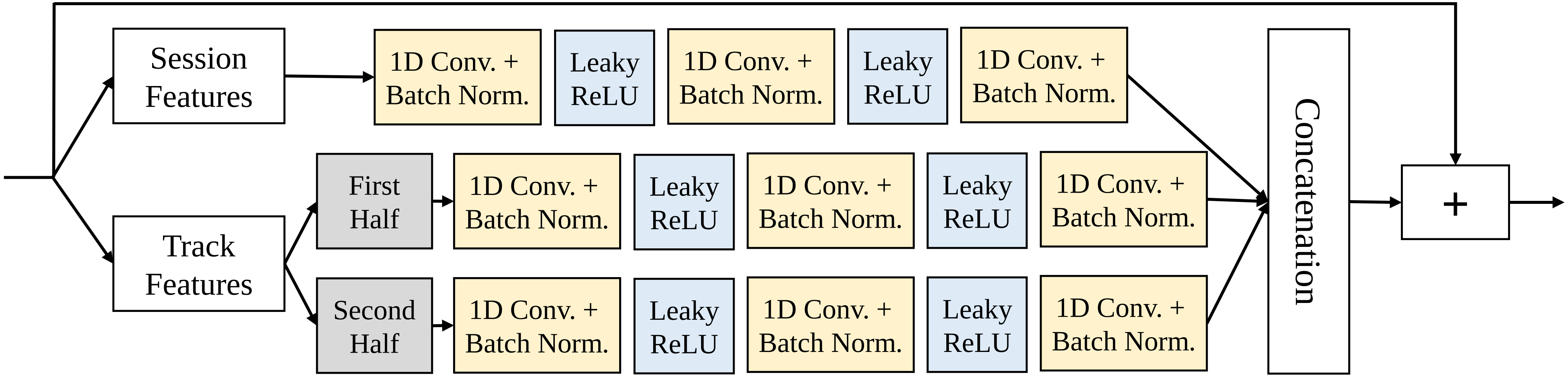
Inception V3



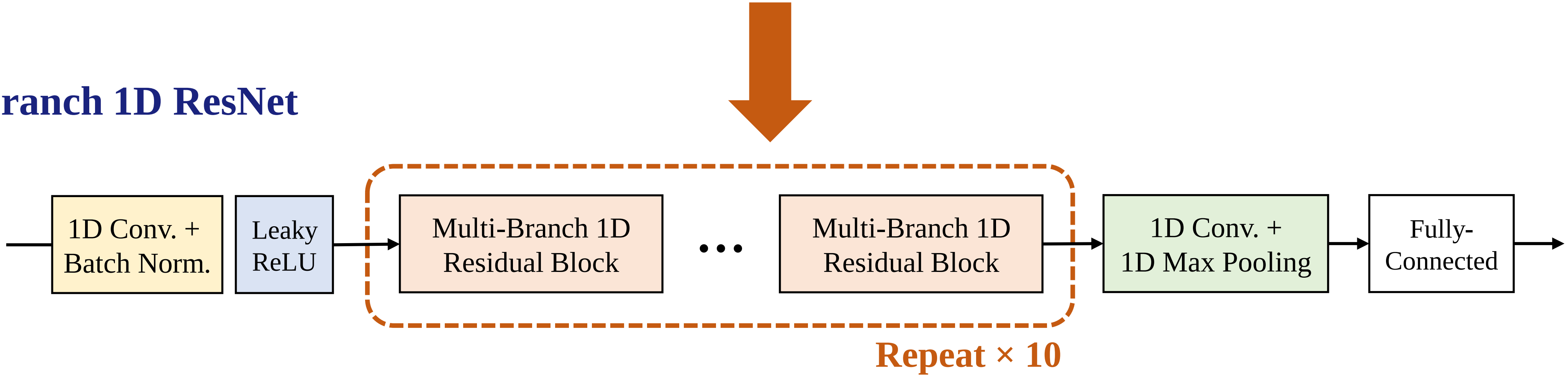
ResNeXt

Proposed Method

Muti-Branch 1D Residual Block



Muti-Branch 1D ResNet



Experimental Settings - Measures

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

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

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

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

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:

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

Experimental Results

T-test Result

Model	MAA Mean ± Standard Deviation	Average Rank
Proposed	**0.6877 ± 0.0922**	**1.4**
Highway 1D Conv. [12]	0.5459 ± 0.1199 ▼	3.47
Encoder-Decoder [16]	0.5399 ± 0.0676 ▼	3.6
Multi-RNN [14]	0.5811 ± 0.0966 ▼	3.2
RNN-Dense [15]	0.5769 ± 0.0815 ▼	3.33

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

Experimental Results

Friedman-test Result

	Statistics	p-value
Proposed	18.560	0.0010

significance level α=0.05

Wilcoxon signed rank test Result

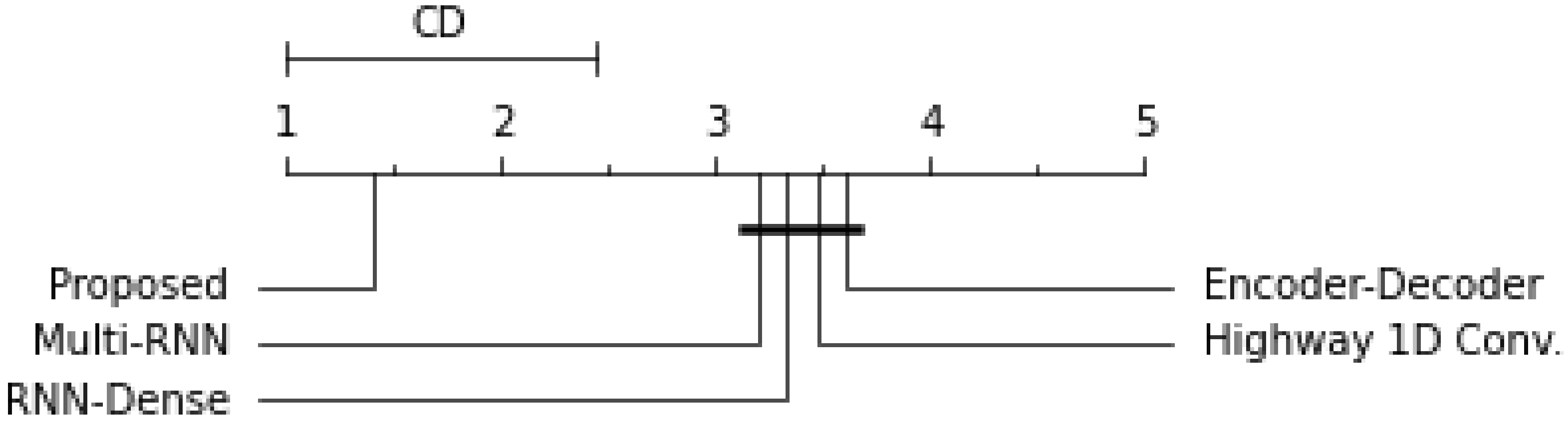
	Proposed ***versus*** Comparison models			
	Highway 1D Conv. [12]	Encoder-Decoder [16]	Multi-RNN [14]	RNN-Dense [15]
Proposed model performance (p-value)	Win (0.0009)	Win (0.0006)	Win (0.0054)	Win (0.0006)

significance level α=0.05

Experimental Results

Bonferroni-Dunn test Result

Critical Difference = 1.4422



Conclusion

Conclusion

- Proposed multi-branch 1D residual block based on 1D convolution and 1D ResNet
- Explicitly separated multi-branch convolution can drive high-correlated feature learning
- Proposed method perform significantly better than conventional RNN-based models

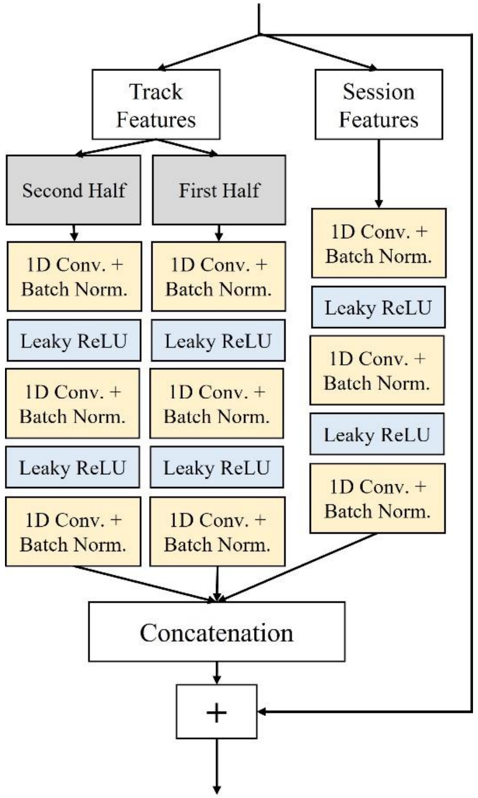
Future Works

- Using track audio features
- Applying grouped convolution (separate inputs into same sizes)

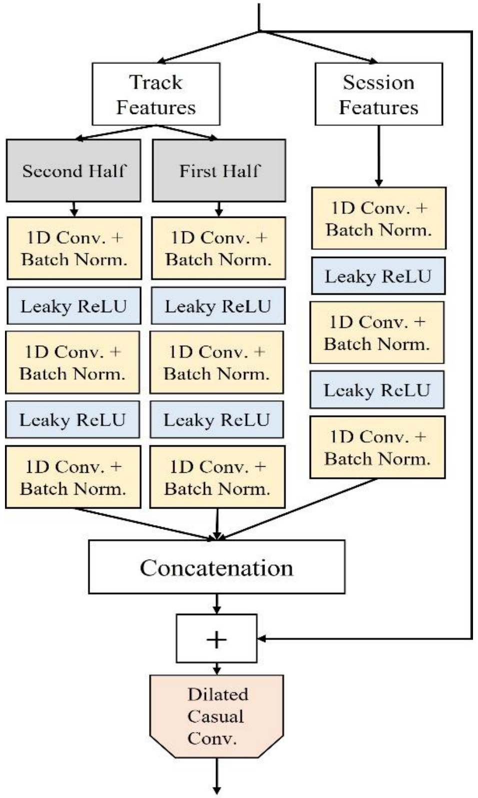
Appendix – 1D ResNet Variation Comparison

1D ResNet Variation Comparison

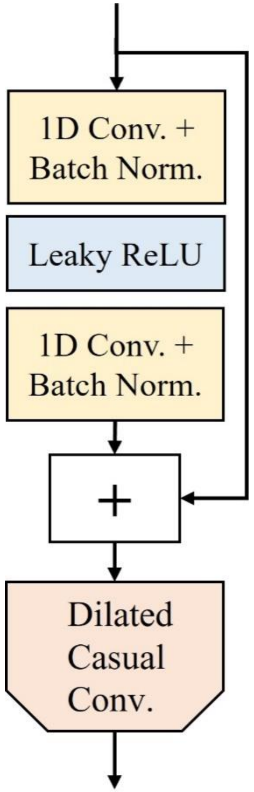
* DCC : Dilated Casual Convolution



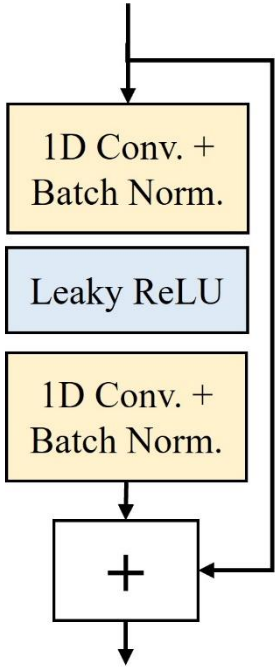
1. Proposed



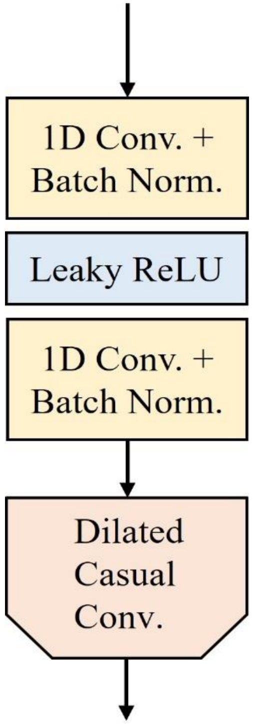
2. Multi-branch
1D ResNet + DCC



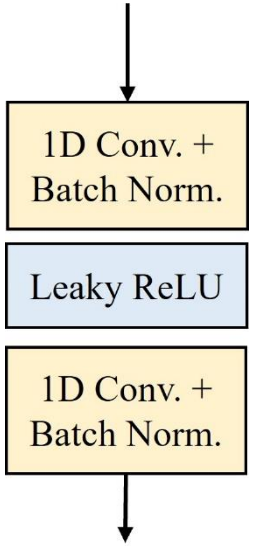
3. Basic 1D
ResNet + DCC



4. Basic 1D
ResNet



5. Residual Conn.
Removed + DCC



6. Residual Conn.
Removed

Appendix – 1D ResNet Variation Comparison Result

AutoML

T-test Result

Model	MAA Mean ± Standard Deviation	Average Rank
Proposed	**0.6877 ± 0.0922**	**1.87**
Multi-Branch 1D ResNet + DCC	0.6912 ± 0.0963	1.93
Basic 1D ResNet + DCC	0.5390 ± 0.1347 ▼	3.8
Basic 1D ResNet	0.5032 ± 0.1214 ▼	4
Residual Conn. Removed + DCC	0.4837 ± 0.1164 ▼	4.4
Residual Conn. Removed	0.4487 ± 0.1195 ▼	5

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

Appendix – 1D ResNet Variation Comparison Result

Friedman-test Result

	Statistics	p-value
Proposed	36.663	6.99e-07

significance level α=0.05

Wilcoxon signed rank test Result

* DCC : Dilated Casual Convolution

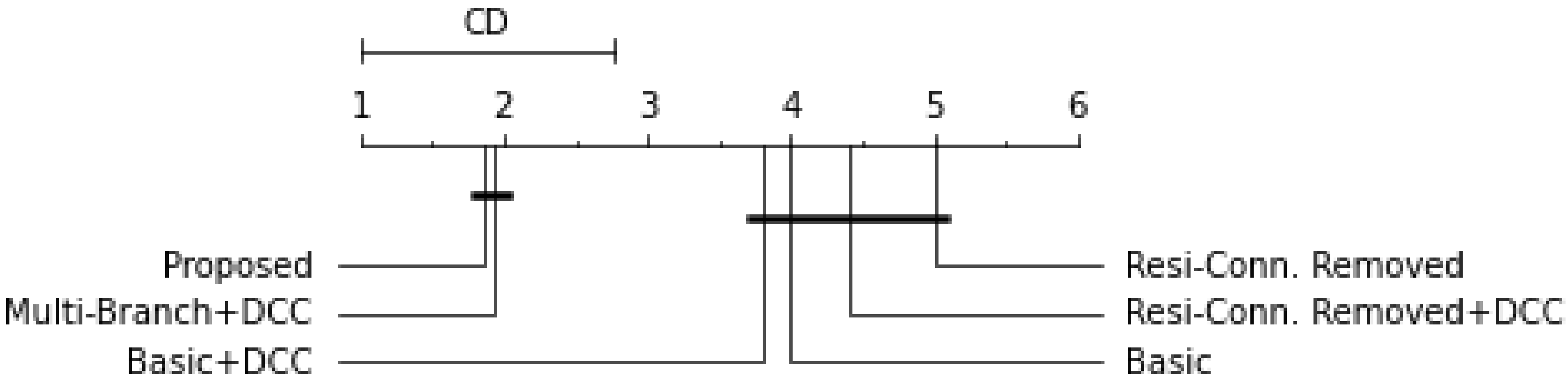
	Proposed ***versus*** Comparison models				
	Multi-Branch 1D ResNet + DCC	Basic 1D ResNet + DCC	Basic 1D ResNet	Residual Conn. Removed + DCC	Residual Conn. Removed
Proposed model performance (p-value)	Tie (0.4212)	Win (0.0003)	Win (0.0015)	Win (6.10e-05)	Win (6.10e-05)

significance level α=0.05

Appendix – 1D ResNet Variation Comparison Result

Bonferroni-Dunn test Result

Critical Difference = 1.7597



* 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.

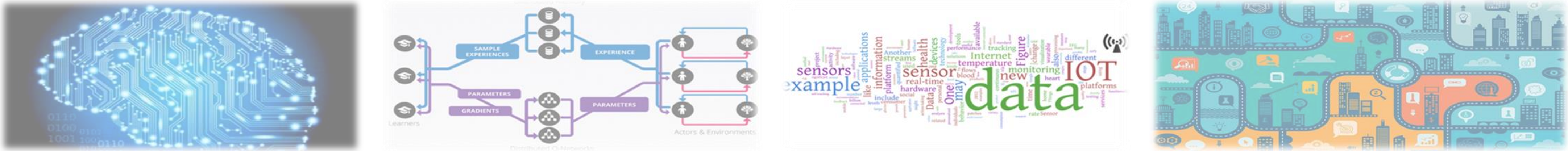
[16]] Darius Afchar, and Romain Hennequin. "Making neural networks interpretable with attribution: application to implicit signals prediction." 14th ACM Conference on Recommender Systems. 2020.

[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



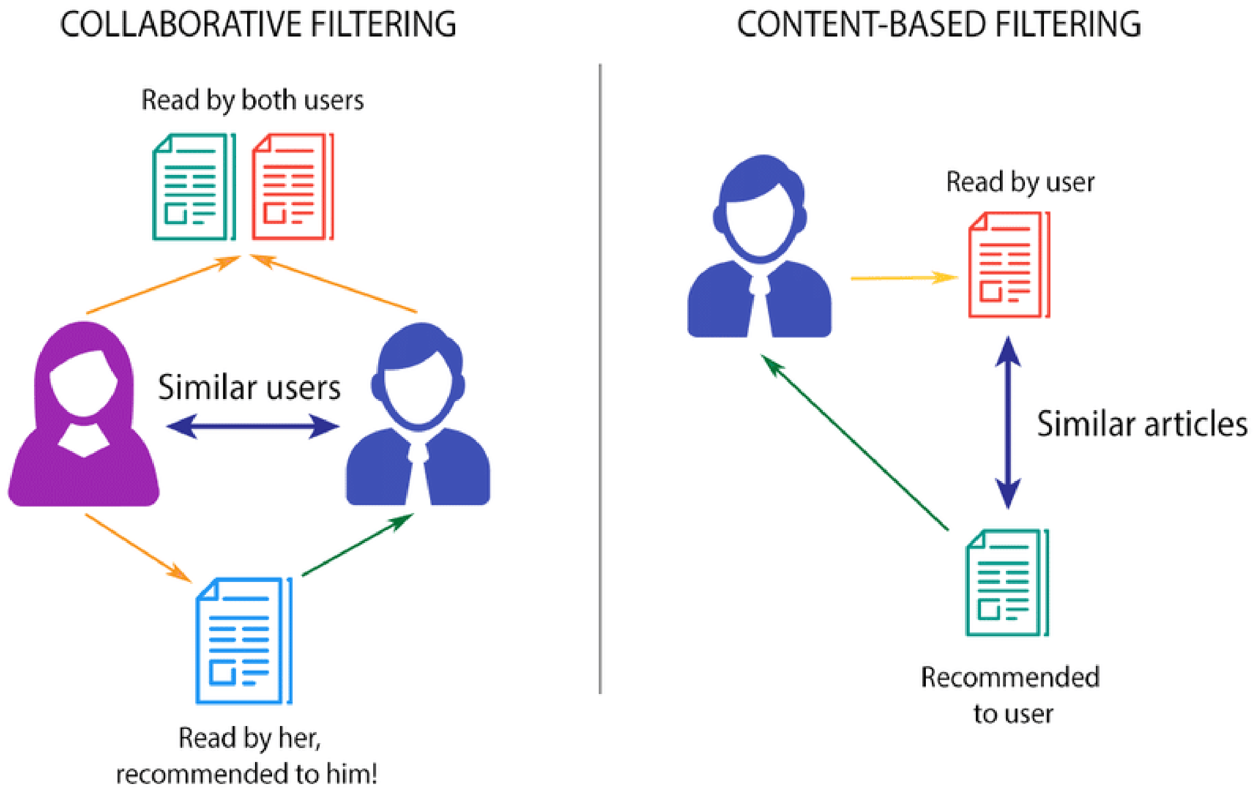
Neural Collaborative Filtering (NCF) Recommendation System

Chae-lyn Lee

2022. 09. 28

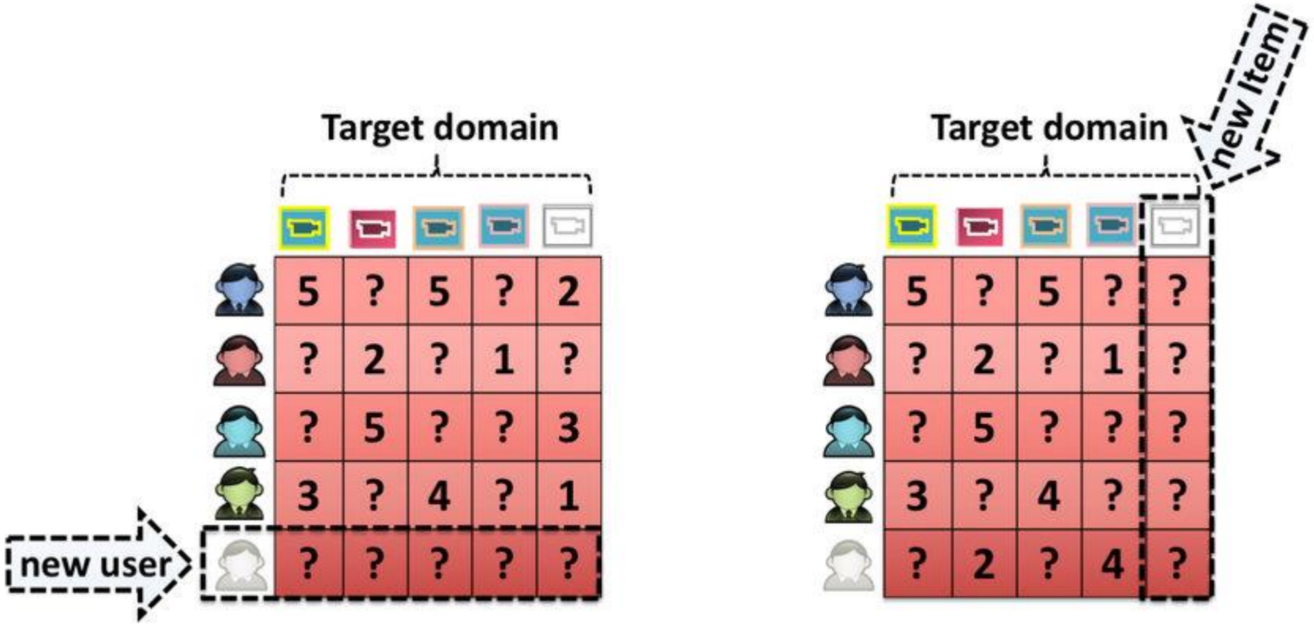
Previous study

The type of recommendation system



Recommendation's Problem

Cold start problem



5	?	5	?	2
?	2	?	1	?
?	5	?	?	3
3	?	4	?	1
?	?	?	?	?

5	?	5	?	?
?	2	?	1	?
?	5	?	?	?
3	?	4	?	?
?	2	?	4	?

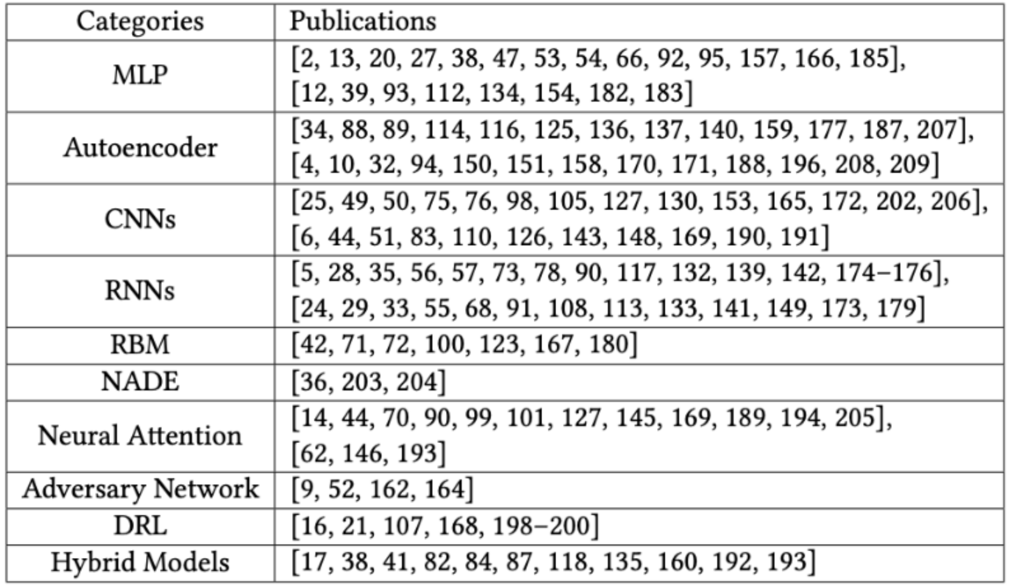
Recommendation Matrix Factorization = Express user-item interaction in inner product

The inherent combination of multiplications between late factors linearly makes it difficult to determine the complex structure of user-item interaction data

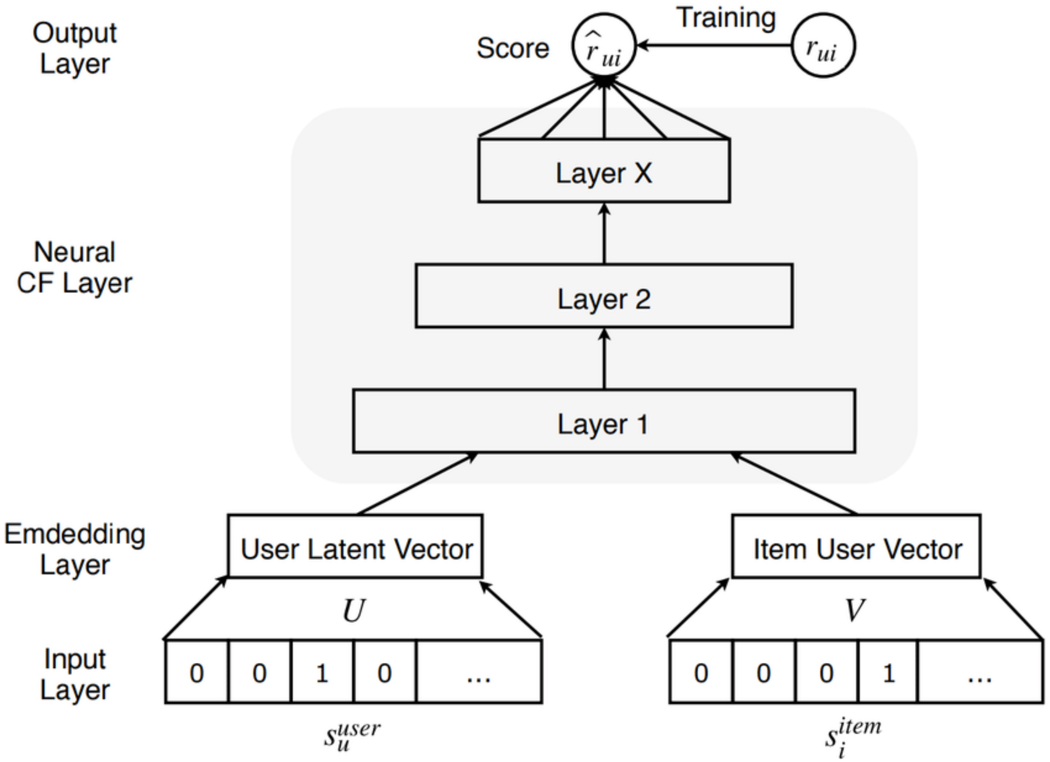
A recent study

Deep-learning recommendation system

Table 1. A lookup table for reviewed publications.



Categories	Publications
MLP	[2, 13, 20, 27, 38, 47, 53, 54, 66, 92, 95, 157, 166, 185], [12, 39, 93, 112, 134, 154, 182, 183]
Autoencoder	[34, 88, 89, 114, 116, 125, 136, 137, 140, 159, 177, 187, 207], [4, 10, 32, 94, 150, 151, 158, 170, 171, 188, 196, 208, 209]
CNNs	[25, 49, 50, 75, 76, 98, 105, 127, 130, 153, 165, 172, 202, 206], [6, 44, 51, 83, 110, 126, 143, 148, 169, 190, 191]
RNNs	[5, 28, 35, 56, 57, 73, 78, 90, 117, 132, 139, 142, 174–176], [24, 29, 33, 55, 68, 91, 108, 113, 133, 141, 149, 173, 179]
RBM	[42, 71, 72, 100, 123, 167, 180]
NADE	[36, 203, 204]
Neural Attention	[14, 44, 70, 90, 99, 101, 127, 145, 169, 189, 194, 205], [62, 146, 193]
Adversary Network	[9, 52, 162, 164]
DRL	[16, 21, 107, 168, 198–200]
Hybrid Models	[17, 38, 41, 82, 84, 87, 118, 135, 160, 192, 193]



A recent study

Deep-learning recommendation system

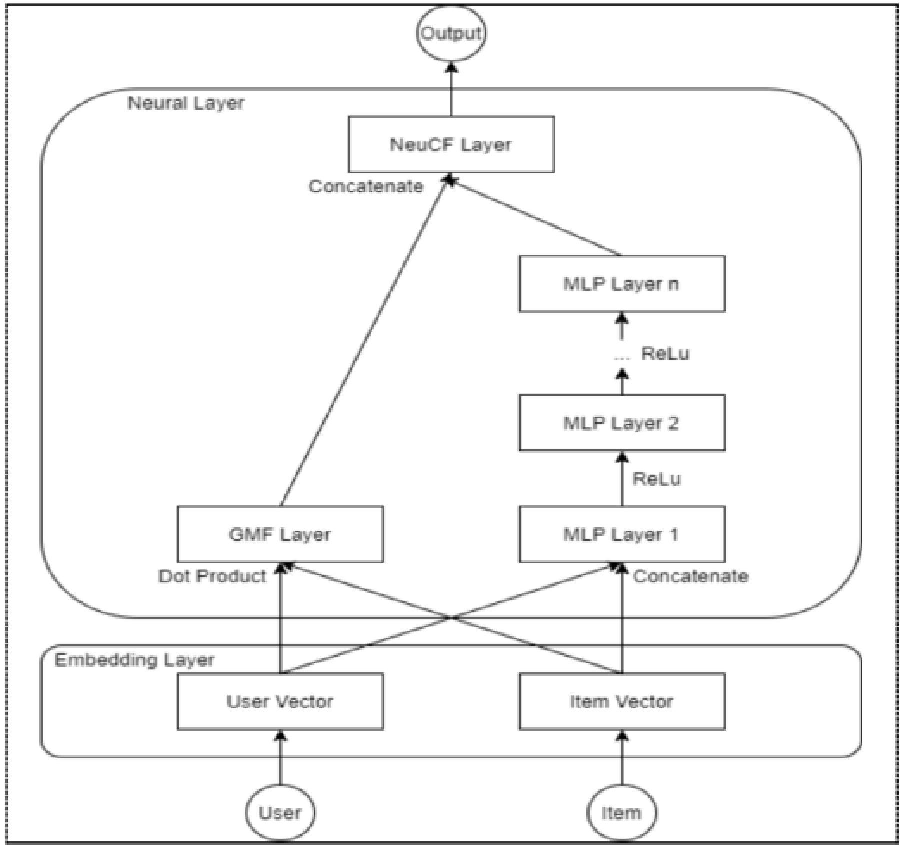


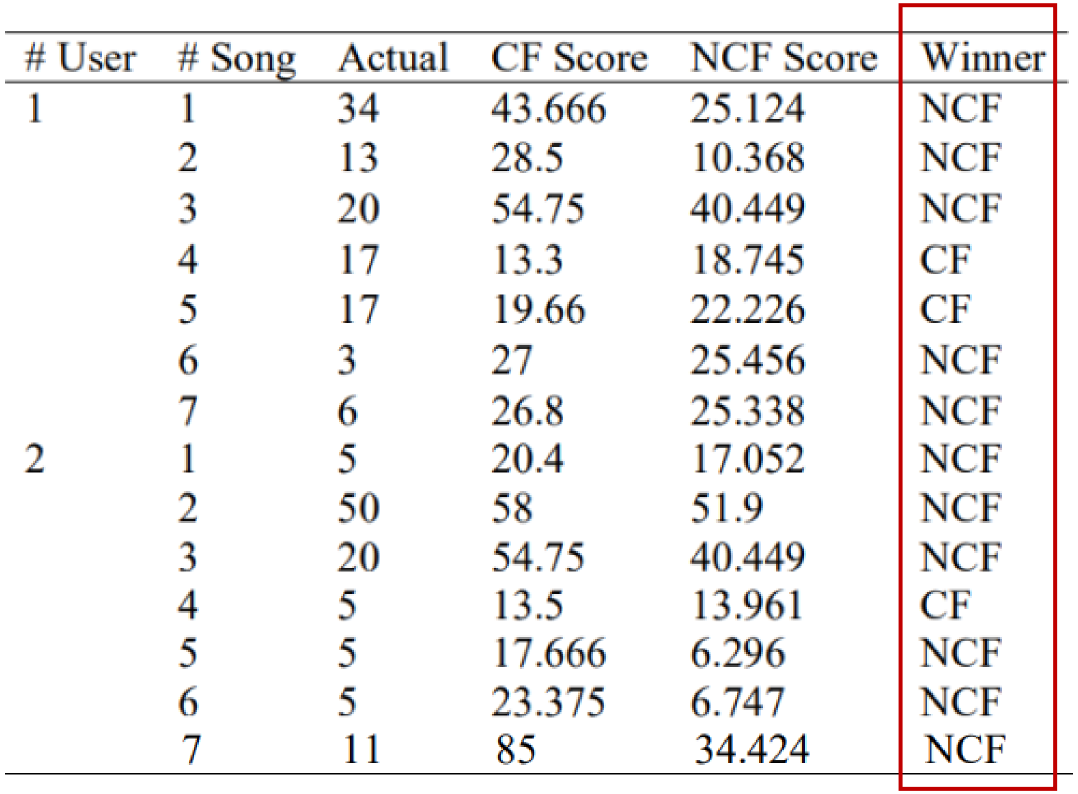
Figure 1. Architecture of Neural Collaborative Filtering.

Neural Collaborative For Music Recommendation System(IOP Conference Series: Materials Science and Engineering,2021)

A recent study

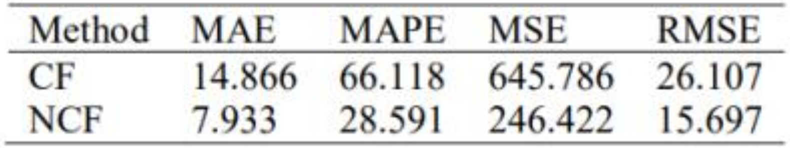
Deep-learning recommendation system

Table 2. Evaluation Results



# User	# Song	Actual	CF Score	NCF Score	Winner
1	1	34	43.666	25.124	NCF
	2	13	28.5	10.368	NCF
	3	20	54.75	40.449	NCF
	4	17	13.3	18.745	CF
	5	17	19.66	22.226	CF
	6	3	27	25.456	NCF
	7	6	26.8	25.338	NCF
2	1	5	20.4	17.052	NCF
	2	50	58	51.9	NCF
	3	20	54.75	40.449	NCF
	4	5	13.5	13.961	CF
	5	5	17.666	6.296	NCF
	6	5	23.375	6.747	NCF
	7	11	85	34.424	NCF

Table 1. Evaluation Error Results



Method	MAE	MAPE	MSE	RMSE
CF	14.866	66.118	645.786	26.107
NCF	7.933	28.591	246.422	15.697

Neural Collaborative For Music Recommendation System(IOP Conference Series: Materials Science and Engineering,2021)

A recent study

Deep-learning recommendation system

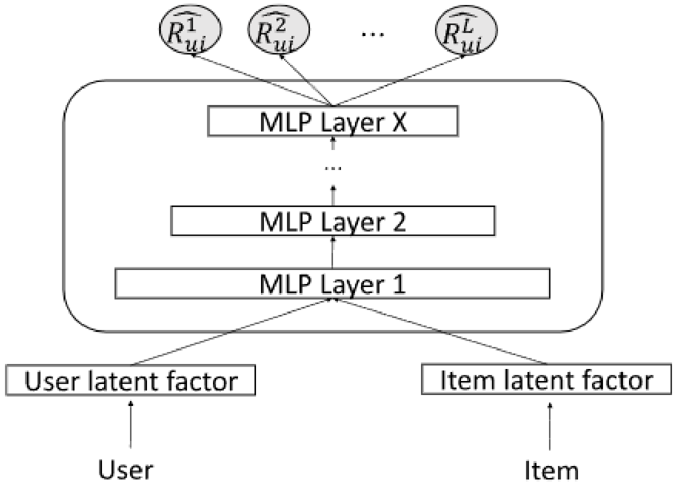


Fig. 1. Neural collaborative filtering for multicriteria model.

NCF-MC Model

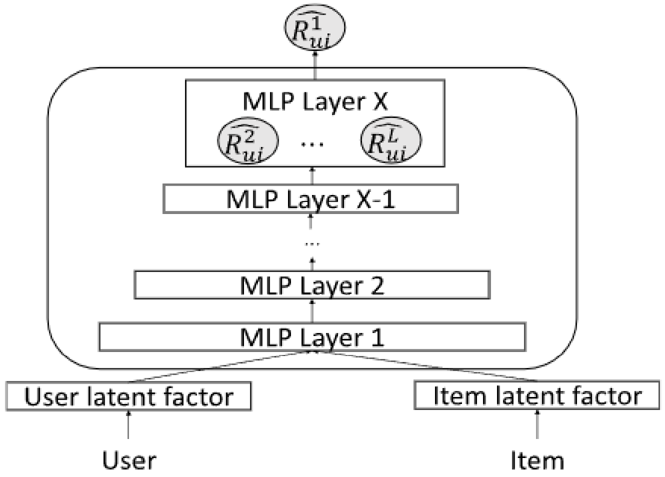


Fig. 2. Neural collaborative filtering for aggregation model.

NCF-AG Model

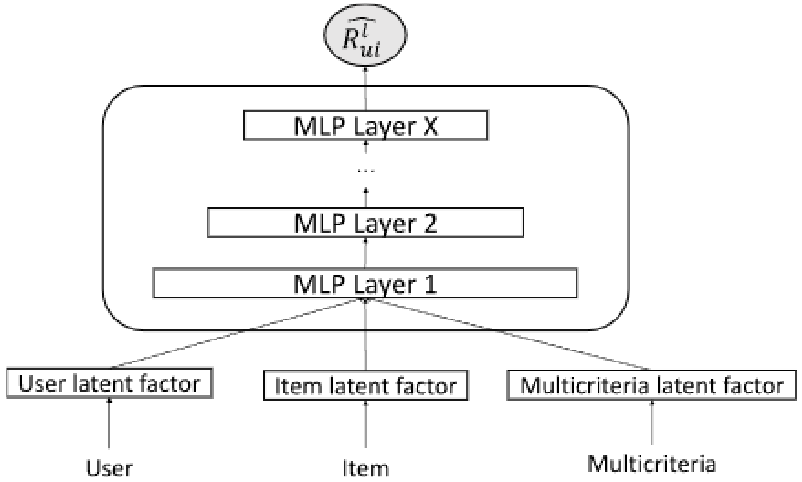


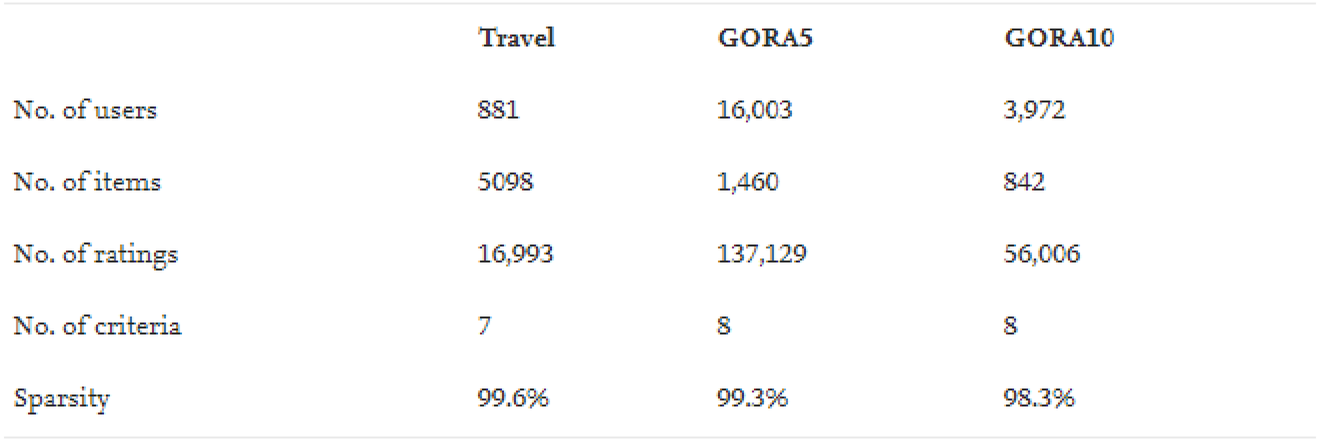
Fig. 3. Neural collaborative filtering for tensor factorization model.

NCF-TF Model

Neural collaborative filtering with multicriteria evaluation data(Applied Soft Computing,2022)

A recent study

Deep-learning recommendation system

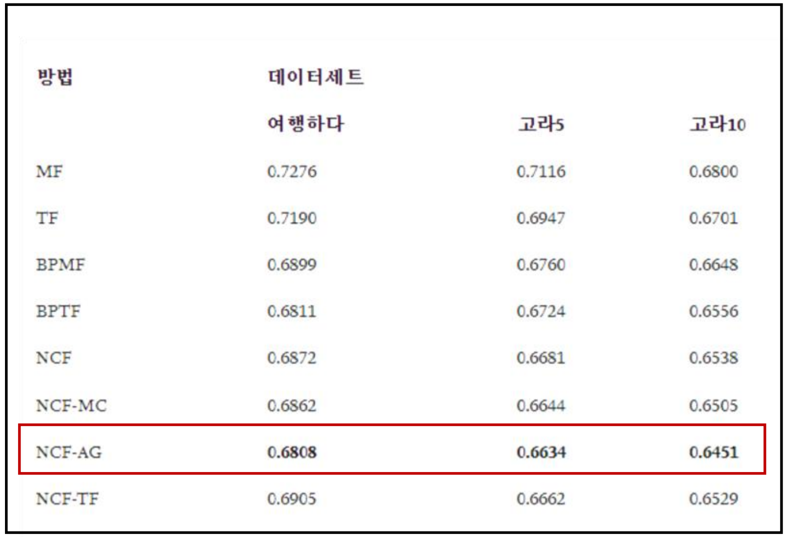


	Travel	GORA5	GORA10
No. of users	881	16,003	3,972
No. of items	5098	1,460	842
No. of ratings	16,993	137,129	56,006
No. of criteria	7	8	8
Sparsity	99.6%	99.3%	98.3%

Neural collaborative filtering with multicriteria evaluation data(Applied Soft Computing,2022)

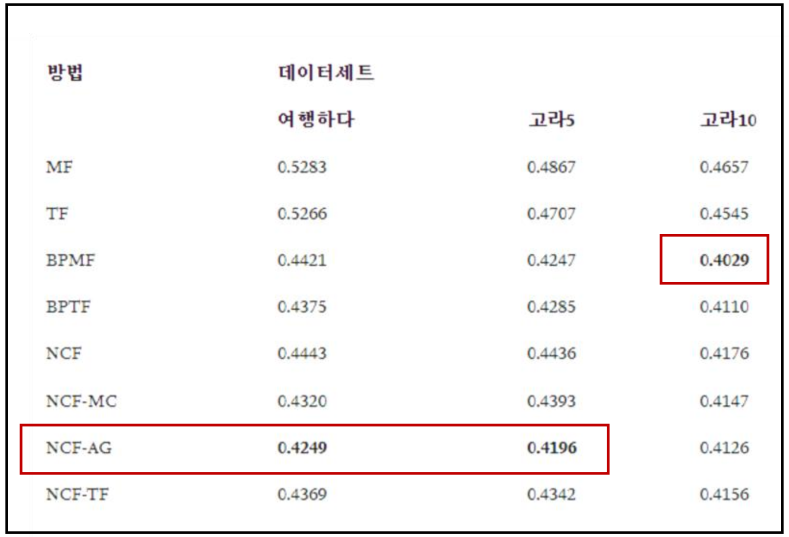
A recent study

Deep-learning recommendation system



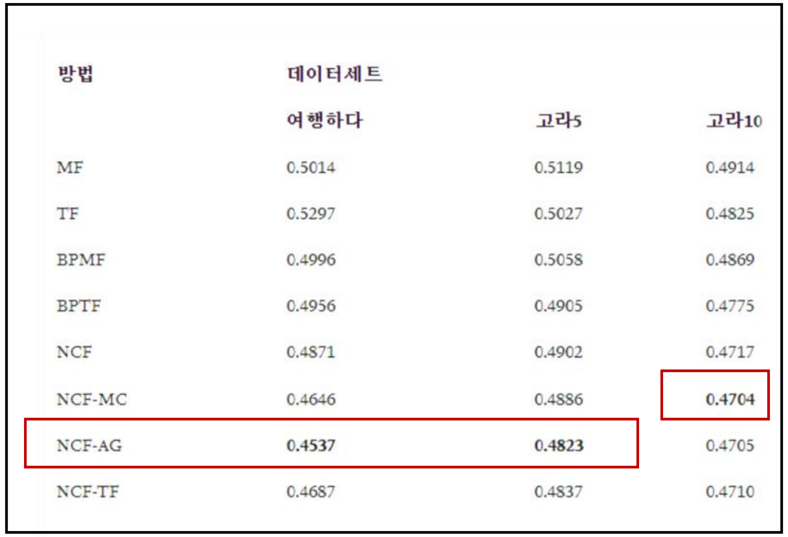
방법	데이터세트		
	여행하다	고라5	고라10
MF	0.7276	0.7116	0.6800
TF	0.7190	0.6947	0.6701
BPMF	0.6899	0.6760	0.6648
BPTF	0.6811	0.6724	0.6556
NCF	0.6872	0.6681	0.6538
NCF-MC	0.6862	0.6644	0.6505
NCF-AG	**0.6808**	**0.6634**	**0.6451**
NCF-TF	0.6905	0.6662	0.6529

RMSE



방법	데이터세트		
	여행하다	고라5	고라10
MF	0.5283	0.4867	0.4657
TF	0.5266	0.4707	0.4545
BPMF	0.4421	0.4247	**0.4029**
BPTF	0.4375	0.4285	0.4110
NCF	0.4443	0.4436	0.4176
NCF-MC	0.4320	0.4393	0.4147
NCF-AG	**0.4249**	**0.4196**	0.4126
NCF-TF	0.4369	0.4342	0.4156

GIM



방법	데이터세트		
	여행하다	고라5	고라10
MF	0.5014	0.5119	0.4914
TF	0.5297	0.5027	0.4825
BPMF	0.4996	0.5058	0.4869
BPTF	0.4956	0.4905	0.4775
NCF	0.4871	0.4902	0.4717
NCF-MC	0.4646	0.4886	**0.4704**
NCF-AG	**0.4537**	**0.4823**	0.4705
NCF-TF	0.4687	0.4837	0.4710

GPIM

Neural collaborative filtering with multicriteria evaluation data(Applied Soft Computing,2022)

A subject of study

Efficient neural networks using NCF (focusing on accuracy)