기계학습자동화 연구실
주간 정기 세미나

이상혁, 조성현, 이태원, 윤건일, 한수정, 이한용, 송태경, 진상형

2022. 08. 30.

기계학습자동화 연구실
주간 정기 세미나

이상혁

2022. 08. 30.

Overall Progress (Medial Image Classification)

Problem	Counter Parts	Experiments	Analysis	Proposal
▪ Medical Image Classification ▪ Low Accuracy				

Medical Image

- Modality: X-ray, computed tomography (CT), magnetic resonance image (MRI), fMRI(functional MRI), positron emission tomography (PET), diffusion tensor image (DTI)
- Loss of human lives and others can be reduced through the timely diagnosis of medical anomalies.
- Convolutional Neural Network (CNN):
  - Due to AlexNet (Krizhevsky et al., 2012)
  - GoogleNet 89% but human 70% for detecting cancer
  - placed in the leader board of many challenges: MICCAI biomedical challenge, BRATS challenge

Medical Image Classfication

- Survey based on Sarvamngala et al. (2021)
- Lung diseases (폐 질환)
  - Ensemble CNN (Ciompi et al., 2015)
  - Small-kernel CNN (Anthimopoulos et al., 2016)
  - Whole image CNN (Gao et al., 2018): AlexNet
  - multicrop pooling CNN (Shen et al., 2017): 3 layered CNN
- Covid-19
  - Customized CNN (Maghdid et al., 2021): AlexNet (+ResNet, GoogleNet)
  - Bayesian CNN (Ghoshal et al., 2020): ResNet50-V2
  - PDCOVIDNET (Chowdhury et al., 2020):
  - CVR-Net (Hasan et al., 2020): ResNet, Depthwise separable convolution
  - Twice transfer learning CNN (Bassi et al., 2022): DenseNet201

Medical Image Classfication

- Immune response abnormalities (면역 반응 이상)
  - CUDA ConvNet CNN (Bayramoglu et al., 2015)
  - Six-layer CNN (Gao et al., 2016)
- Breast tumors (유방 종양)
  - Stopping monitoring CNN (Arevalo et al., 2016)
  - Ensemble CNN (Jiao et al., 2016): AlexNet, SVM
  - Semi-supervised CNN (Sun et al., 2017)
- Heart diseases (심장 질환)
  - One-dimensional CNN (Kiranyaz et al., 2015):1-D CNN
  - Fused CNN (Gao et al., 2017)

Medical Image Classfication

- Eye diseases (안과 질환)
  - Gaussian initialized CNN (Pratt et al., 2016)
  - Hyper parameter tuing inception-v4 (Shankar et al., 2020): Inception-v4
- Colon cancer (결장암)
  - Ensemble CNN (Sirinukunwattana et al., 2016): Standard CNN
- Brain disorders (뇌종양)
  - Fused CNN (Gao et al., 2016): 3D CNN
  - Input cascaded CNN (Sajjad et al., 2019): VGG-19

➢ 아키텍쳐 다양성 : 조사된 문헌들에는 방대한 수의 CNN 구조가 포함되어 있음. 이러한 높은 아키텍처 다양성으로 인해, 특정 작업에 가장 적합한 아키텍처를 선택하는 것은 어려움 (Sarvamngala et al., 2021). → 많은 Medical Image에서 Robust한 CNN을 개발하고자 함.

Future

- Related Work
- Public Dataset Gathering
- Popular & SOTA CNN pilot experiments

Overall Progress (Thesis, ~8.31)

Problem	Counter Parts	Experiments	Analysis	Proposal
▪ Medical Image Classification ▪ Low Accuracy	▪ Traditional 2-D approaches ▪ Traditional 3-D approaches	▪ Completed	▪ Too little information in single 2D image ▪ Too much unnecessary information in 3D images.	▪ Segmentation ▪ Proposal of a new model to process multiple 2D images ▪ Independently extract spatial features

Experimental Settings

- Dataset (Thyroid-associated ophthalmopathy, TAO)
  - ✓ Number of Samples: 1068 patients (144 actives and 924 inactives)
  - ✓ Number of Classes: 2 (Active & Inactive)

- Cross-Validation for Test Performance
  - ✓ 30 iterations
  - ✓ Repeated Random Sub-sampling Validation (Stratified)
  - ✓ Data Split Ratio: (Train : Validation : Test) = (0.6 : 0.2 : 0.2)

- Cross-Validation for Test Performance
  - ✓ 30 iteration

- Models for Comparison & Hyper Parameter Setting:

Model	Batch Size	Epochs	Loss Function	Optimizer	LR	Weight Decay
2D Approaches: ResNet152 [1 DenseNet121 [2] EfficientNet-B0 [3] EfficientNetV2-L [4] CovidNet [5] **3D Approaches:** ResNet152-3D [6] DenseNet121-3D [6]	64	30	Cross Entropy Loss	AdamW [7]	1e-3	1e-4

Experimental Results

Table 1. Metrics with 90% confidence interval (▼/△ indicates that the corresponding method is **significantly worse** than the best model written in bold based on paired t-test at 10% significance level).

Input Type	Model	AUROC Average	AUROC 90% CI Lower Upper	AUROC p-value	AUPRC Average	AUPRC 90% CI Lower Upper	AUPRC p-value
78 segmentation approaches	Proposed	0.810	0.823 0.797		0.449	0.502 0.445	
	DenseNet121	0.765	0.784 0.746	0.000 (**▼)	0.451	0.484 0.418	0.096 (*▼)
	EfficientNetV2-L	0.764	0.783 0.745	0.000 (**▼)	0.395	0.437 0.353	0.000 (**▼)
	CovidNet	0.757	0.777 0.737	0.000 (**▼)	0.451	0.486 0.416	0.106
	ResNet152	0.729	0.757 0.702	0.000 (**▼)	0.397	0.431 0.363	0.000 (**▼)
	EfficientNet-B0	0.624	0.649 0.598	0.000 (**▼)	0.252	0.277 0.226	0.000 (**▼)
11 Selection approaches	EfficientNetV2-L	0.711	0.731 0.692	0.000 (**▼)	0.284	0.312 0.255	0.000 (**▼)
	DenseNet121	0.709	0.728 0.690	0.000 (**▼)	0.309	0.333 0.286	0.000 (**▼)
	ResNet152	0.647	0.677 0.616	0.000 (**▼)	0.242	0.267 0.216	0.000 (**▼)
	CovidNet	0.623	0.645 0.600	0.000 (**▼)	0.237	0.267 0.207	0.000 (**▼)
	EfficientNet-B0	0.566	0.601 0.531	0.000 (**▼)	0.213	0.236 0.190	0.000 (**▼)

CI, confidence interval; *: p-value < 0.1; **: p-value < 0.01

Table 2. Metrics with 90% confidence interval (▼/△ indicates that the corresponding method is **significantly worse** than the best model written in bold based on paired t-test at 10% significance level).

Input Type	Model	AUROC Average	AUROC 90% CI Lower Upper	AUROC p-value	AUPRC Average	AUPRC 90% CI Lower Upper	AUPRC p-value
Traditional 3-dimension approaches	DenseNet121-3D	0.700	0.725 0.674	0.000 (**▼)	0.315	0.350 0.281	0.000 (**▼)
	ResNet152-3D	0.665	0.691 0.640	0.000 (**▼)	0.241	0.263 0.219	0.000 (**▼)
Traditional single 2-dimension approaches	EfficientNetV2-L	0.688	0.711 0.665	0.000 (**▼)	0.264	0.290 0.238	0.000 (**▼)
	DenseNet121	0.670	0.704 0.637	0.000 (**▼)	0.253	0.279 0.226	0.000 (**▼)
	CovidNet	0.669	0.691 0.646	0.000 (**▼)	0.251	0.274 0.227	0.000 (**▼)
	ResNet152	0.617	0.645 0.589	0.000 (**▼)	0.215	0.234 0.195	0.000 (**▼)
	EfficientNet-B0	0.557	0.584 0.529	0.000 (**▼)	0.186	0.204 0.167	0.000 (**▼)

CI, confidence interval; *: p-value < 0.1; **: p-value < 0.01

갑상선 안병증 과제 & NVIDIA 협력 연구

- Active 논문 초안 번역 및 압축 작업(~9월 7일)
- Treatment 레이블 수신
- 엔비디아 개발 세팅 진행

산학협력프로젝트(SW중심대학)

- 산학-1 3건 결과보고서 1차 제출 완료, 수정 후 제출 예정
- 산학-2 2건
  - 특허 출원 2건 신청 완료
  - 국제 학회(혹은 저널) 논문 작업 진행중
    - ICCE-Asia 투고 예정(데드라인 : 9월 10일) : 이상혁
    - IEEE Access 투고 예정 : 조성현

References

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems* 25 (2012).

Sarvamangala, D. R., and Raghavendra V. Kulkarni. "Convolutional neural networks in medical image understanding: a survey." *Evolutionary intelligence* (2021): 1-22.

Ciompi, Francesco, et al. "Automatic classification of pulmonary peri-fissural nodules in computed tomography using an ensemble of 2D views and a convolutional neural network out-of-the-box." *Medical image analysis* 26.1 (2015): 195-202.

Anthimopoulos, Marios, et al. "Lung pattern classification for interstitial lung diseases using a deep convolutional neural network." *IEEE transactions on medical imaging* 35.5 (2016): 1207-1216.

Gao, Mingchen, et al. "Holistic classification of CT attenuation patterns for interstitial lung diseases via deep convolutional neural networks." *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization* 6.1 (2018): 1-6.

Shen, Wei, et al. "Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification." *Pattern Recognition* 61 (2017): 663-673.

Maghdid, Halgurd S., et al. "Diagnosing COVID-19 pneumonia from X-ray and CT images using deep learning and transfer learning algorithms." *Multimodal image exploitation and learning 2021*. Vol. 11734. SPIE, 2021.

Ghoshal, Biraja, and Allan Tucker. "Estimating uncertainty and interpretability in deep learning for coronavirus (COVID-19) detection." *arXiv preprint arXiv:2003.10769* (2020).

Chowdhury, Nihad K., Md Rahman, and Muhammad Ashad Kabir. "PDCOVIDNet: a parallel-dilated convolutional neural network architecture for detecting COVID-19 from chest X-ray images." *Health information science and systems* 8.1 (2020): 1-14.

Hasan, Md, et al. "CVR-Net: A deep convolutional neural network for coronavirus recognition from chest radiography images." *arXiv preprint arXiv:2007.11993* (2020).

Bassi, Pedro RAS, and Romis Attux. "A deep convolutional neural network for COVID-19 detection using chest X-rays." *Research on Biomedical Engineering* 38.1 (2022): 139-148.

References

Bayramoglu, Neslihan, Juho Kannala, and Janne Heikkilä. "Human epithelial type 2 cell classification with convolutional neural networks." *2015 IEEE 15th International Conference on Bioinformatics and Bioengineering (BIBE)*. IEEE, 2015.

Gao, Zhimin, et al. "HEp-2 cell image classification with deep convolutional neural networks." *IEEE journal of biomedical and health informatics* 21.2 (2016): 416-428.

Arevalo, John, et al. "Representation learning for mammography mass lesion classification with convolutional neural networks." *Computer methods and programs in biomedicine* 127 (2016): 248-257.

Jiao, Zhicheng, et al. "A deep feature based framework for breast masses classification." *Neurocomputing* 197 (2016): 221-231.

Sun, Wenqing, et al. "Enhancing deep convolutional neural network scheme for breast cancer diagnosis with unlabeled data." *Computerized Medical Imaging and Graphics* 57 (2017): 4-9.

Kiranyaz, Serkan, Turker Ince, and Moncef Gabbouj. "Real-time patient-specific ECG classification by 1-D convolutional neural networks." *IEEE Transactions on Biomedical Engineering* 63.3 (2015): 664-675.

Gao, Xiaohong, et al. "A fused deep learning architecture for viewpoint classification of echocardiography." *Information Fusion* 36 (2017): 103-113.

Pratt, Harry, et al. "Convolutional neural networks for diabetic retinopathy." *Procedia computer science* 90 (2016): 200-205.

Shankar, K., et al. "Hyperparameter tuning deep learning for diabetic retinopathy fundus image classification." *IEEE Access* 8 (2020): 118164-118173.

Sirinukunwattana, Korsuk, et al. "Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images." *IEEE transactions on medical imaging* 35.5 (2016): 1196-1206.

Gao, Xiaohong W., and Rui Hui. "A deep learning based approach to classification of CT brain images." *2016 SAI Computing Conference (SAI)*. IEEE, 2016.

Sajjad, Muhammad, et al. "Multi-grade brain tumor classification using deep CNN with extensive data augmentation." *Journal of computational science* 30 (2019): 174-182.

References (Thesis)

[1] He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[2] Huang, Gao, et al. "Densely connected convolutional networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

[3] Tan, Mingxing, and Quoc Le. "Efficientnet: Rethinking model scaling for convolutional neural networks." International conference on machine learning. PMLR, 2019.

[4] Tan, Mingxing, and Quoc Le. "Efficientnetv2: Smaller models and faster training." International Conference on Machine Learning. PMLR, 2021.

[5] Wang, Linda, Zhong Qiu Lin, and Alexander Wong. "Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images." Scientific Reports 10.1 (2020): 1-12.

[6] Hara, Kensho, Hirokatsu Kataoka, and Yutaka Satoh. "Can spatiotemporal 3d cnns retrace the history of 2d cnns and imagenet?." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. 2018.

[7] Loshchilov, Ilya, and Frank Hutter. "Decoupled weight decay regularization." *arXiv preprint arXiv:1711.05101* (2017).

기계학습자동화 연구실
주간 정기 세미나

조성현

2022. 08. 30.

Individual Studies:

Sung-Hyun Cho

Overall Progress

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

Experiment Settings

- Datasets

Dataset	Patterns	# of Classes	Average Patterns per Class	±	Standard Deviation	Inputs
Ballroom	698	8	87.250	±	17.555	
Ballroom-Extended	4180	13	321.54	±	195.70	
emoMusic-G	744	8	93.00	±	16.1776	
Emotify-G	400	4	100.00	±	0.00	
FMA-SMALL	7996	8	999.500	±	0.500	
GiantStepsKey-G	604	23	24.9130	±	28.0274	(128, 1296), Numpy arrays
GMD-G	1042	18	57.89	±	70.07	
GTZAN	1000	10	100.00	±	0.00	
HOMBURG	1886	9	209.56	±	134.87	
ISMIR04	727	6	121.167	±	95.602	
MICM	1137	7	162.429	±	119.717	
Seyerlehner-Unique	3115	10	311.500	±	248.349	

- Cross-Validation for Test Performance
  - 10 iterations
  - Data Split Ratio: (Train : Test) = (0.8 : 0.2)

Experiment Settings

➢ Hyper Parameter Setting:

Model	Batch Size	Epochs	Loss Function	Optimizer	LR	Weight Decay
Proposed	16	60	Cross Entropy Loss	Adam	1e-3	1e-4
[1] (Li et al., 2021)	64		Proposed Loss Function			
[2] (Pelchat & Gelowitz, 2020)	128		Cross Entropy Loss			
[3] (Cheng et al., 2020)	32		Cross Entropy Loss			
[4] (Shah et al., 2022)	32		Cross Entropy Loss			

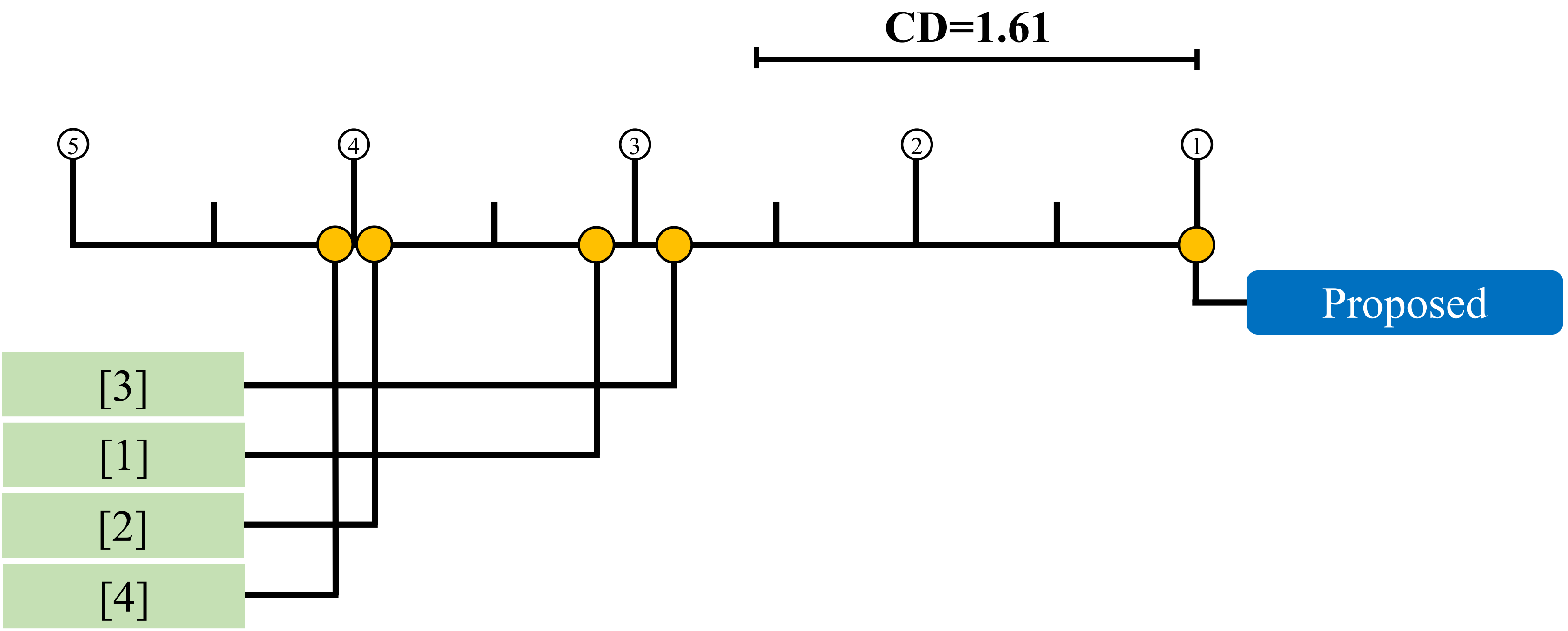
[1] Li, J., Han, L., Li, X., Zhu, J., Yuan, B., & Gou, Z. (2022). An evaluation of deep neural network models for music classification using spectrograms. *Multimedia Tools and Applications*, *81*(4), 4621-4647.
[2] Pelchat, N., & Gelowitz, C. M. (2020). Neural network music genre classification. *Canadian Journal of Electrical and Computer Engineering*, *43*(3), 170-173.
[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.
[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.

Experiment Results

Dataset (Genre)	Proposed	[1]	[2]	[3]	[4]
Ballroom	$\mathbf{55.86 \pm 3.41}$	45.71 ± 2.86	42.43 ± 2.29	48.07 ± 4.79	47.71 ± 4.58
Ballroom-Extended	$\mathbf{70.35 \pm 3.25}$	61.82 ± 3.79	44.68 ± 1.30	62.88 ± 4.70	54.38 ± 5.33
emoMusic-G	$\mathbf{35.77 \pm 2.49}$	27.72 ± 3.50	28.99 ± 1.78	32.95 ± 2.92	31.34 ± 3.66
Emotify-G	$\mathbf{72.00 \pm 4.05}$	62.78 ± 2.70	67.50 ± 4.17	66.00 ± 3.32	65.88 ± 5.62
FMA-SMALL	$\mathbf{55.07 \pm 0.84}$	47.03 ± 1.31	42.19 ± 1.16	46.96 ± 1.62	40.88 ± 1.73
GiantStepsKey-G	$\mathbf{35.29 \pm 2.67}$	28.43 ± 3.15	25.21 ± 2.53	27.19 ± 2.90	24.30 ± 2.11
GMD-G	$\mathbf{65.50 \pm 3.55}$	61.91 ± 3.04	63.44 ± 1.71	39.38 ± 3.36	35.36 ± 6.35
GTZAN	$\mathbf{79.00 \pm 2.94}$	77.50 ± 3.81	63.35 ± 2.08	75.10 ± 3.84	72.50 ± 2.13
HOMBURG	$\mathbf{55.53 \pm 2..39}$	53.49 ± 1.59	51.01 ± 1.51	52.54 ± 2.55	48.15 ± 2.14
ISMIR04	$\mathbf{79.13 \pm 2.39}$	70.96 ± 2.93	67.88 ± 1.70	71.32 ± 1.37	67.67 ± 1.99
MICM	$\mathbf{41.18 \pm 2.94}$	37.81 ± 3.32	39.30 ± 2.21	38.42 ± 3.16	39.34 ± 2.28
Seyerlehner-Unique	$\mathbf{70.85 \pm 2.36}$	63.62 ± 2.56	60.72 ± 2.73	63.50 ± 2.79	61.62 ± 1.70
Avg . Rank	1.00	3.17	3.92	2.83	4.08

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

Bonferroni-Dunn Test



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

Individual Tasks:

Sung-Hyun Cho

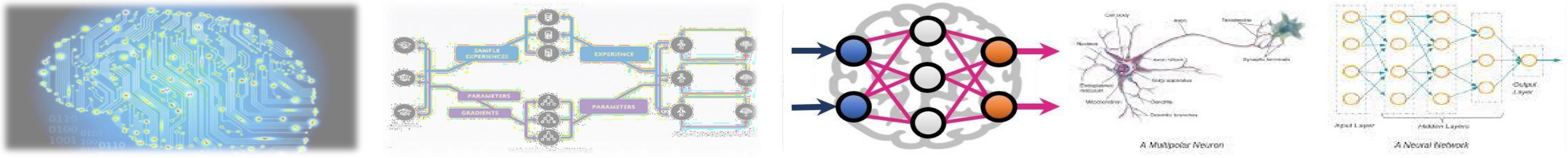
졸업 관련

- 실험 완료

- 논문 작성 중
  - Introduction
  - Related Work
  - Materials and Methods
  - Experimental Results and Result Analysis
  - Conclusion and Future Works

연구실 관련

- 산학프로젝트2 (IEEE Access)
  - Introduction
  - Related Work
  - Materials and Methods
  - Experimental Results and Result Analysis
  - Conclusion and Future Works



Stock Price Prediction

Tae-Won Lee

Overall

Problem

- Stock Price Prediction
- Low Accuracy

Related Works

- CNN-BiLSTM-AM
- AdvLSTM
- DTML

Experiments

- KDD 17 Dataset

Data 개수 : 2518
Feature : 6개
(Open, High, Low, Close, Volume, Adj Close)

- Metric

Accuracy

Analysis

- Need more information (Multimodal)

Proposal

- Multimodal Learning

Post-experimental analysis

- Accuracy improvement (0.506->0.517)
- Ranking mean improvement (2.48->1.72)

Conclusion

- Accuracy improvement
- Statistically significant

Experiment Result

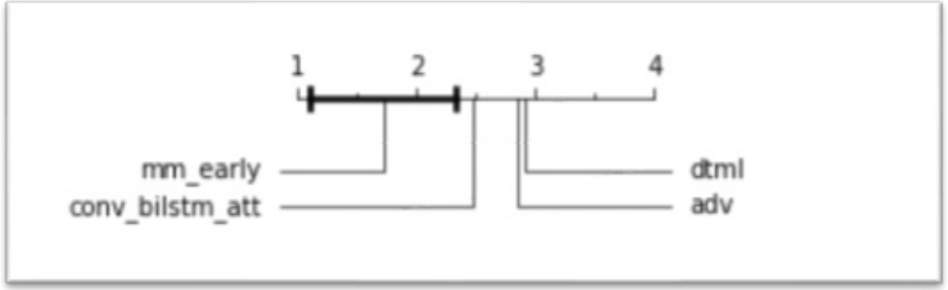
- Test Result

  - Friedman test

    H_0 = 변수(모델)들 간의 순위가 차이가 없다.
    H_1 = 변수(모델)들 간의 순위가 차이가 있다.
    x^2 statistics = 27.97394
    $P - value$ = 3.67805e-06

➢ x^2statistics이 27.97394이고, 유의수준 0.05하에서 P-value 값이 3.67805e-06으로 귀무가설을 기각함.
➢ 변수(모델)들 간의 순위의 차이가 있다.

  - Bonferroni-dunn test (post-hoc)



- Critical Difference = 0.61812

Progress of writing a Thesis

- Introduction

- Related work
  - Stock price prediction
    - Statistical Stock price prediction
    - Machine learning Stock price prediction
    - Deep learning Stock price prediction

---------------------------------- (초안 작성 완료) ------------------------------------

  - Multimodal learning
    - Multimodal deep learning
    - Stock price prediction using Multimodal deep learning

Progress of writing a Thesis

- Multimodal Transformer for stock price prediction

  - Multimodal Transformer
    - 모델 구조 설명 및 그림
  - Data preporcessing
    - 데이터 전처리 및 모델 입력방식 설명

- Experinets
  - Experiments Design
    - 세부 실험 세팅 및 hyperparameter
  - Datasets

- Results
  - 실험 결과 테이블 및 설명
  - 향후 발전 방향

---------------------------------- (초안 작성 완료) ------------------------------------

Appendix

Experiment Result

- Red : High Accuracy
- Asterisk (**) : Low Accuracy

- Accuracy (Index : NASDAQ) – 1/3

Data	Proposed -Early fusion Transformer	DTML [2]	AdvLSTM [7]	CNN BiLSTM AM [8]
AAPL	0.544792	0.52031	**0.47656	0.48281
AMZN	0.520833	0.51250	0.53125	**0.49688
BA	0.5375	**0.50625	0.51563	0.52188
BAC	0.542708	0.51563	0.50781	**0.48438
BHP	0.485417	**0.47813	0.50000	0.48125
BRK-B	0.490625	**0.47188	0.50000	0.49219
CHL	**0.479167	0.52344	0.50781	0.50938
CMCSA	0.516667	**0.49063	0.48750	0.52031
CVX	0.507292	**0.48125	0.49688	0.50313
D	0.517708	0.51094	**0.50313	0.52813
DCM	0.534375	0.52656	**0.48438	0.48906
DIS	0.529167	**0.48750	0.51094	0.51250
DOW	0.51875	0.50000	0.48281	**0.46406
DUK	0.517708	0.48281	0.47344	0.55469
EXC	0.529167	0.49531	**0.46875	0.50781
GE	0.526042	0.51563	0.52188	**0.50469
GOOGL	0.516667	0.52969	0.50938	0.52813
HD	0.514583	**0.50469	0.52813	0.52031

Experiment Result

- Red : High Accuracy
- Asterisk (**) : Low Accuracy

- Accuracy (Index : NASDAQ) – 2/3

Data	Proposed -Early fusion Transformer	DTML [2]	AdvLSTM [7]	CNN BiLSTM AM [8]
INTC	0.525	0.51094	**0.48594	0.53125
JNJ	0.526042	0.52500	**0.48281	0.49219
JPM	0.514583	0.50156	**0.50313	0.50625
KO	0.5125	**0.47344	0.47969	0.52344
MA	0.528125	0.51250	**0.49219	0.50156
MMM	0.519792	0.47969	0.51250	**0.46875
MO	0.544792	**0.48906	0.51250	0.55469
MRK	0.50625	*0.49375	*0.49375	0.49844
MSFT	0.501042	0.53438	0.49531	0.51250
NGG	0.538542	0.53750	0.51563	**0.51406
NTT	0.509375	0.56719	0.51563	**0.49063
NVS	0.51875	**0.49375	0.50781	0.51250
ORCL	0.511458	0.50938	**0.50469	0.52969
PEP	0.514583	**0.48750	0.50313	0.48906
PFE	0.492708	0.48594	**0.47813	0.50625
PG	0.516667	0.48438	0.52500	**0.47344
PTR	0.528125	0.52188	**0.48438	0.52500
RDS-B	0.527083	**0.48750	0.49688	0.51719

Experiment Result

- Red : High Accuracy
- Asterisk (**) : Low Accuracy

- Accuracy (Index : NASDAQ) – 3/3

Data	Proposed -Early fusion Transformer	DTML [2]	AdvLSTM [7]	CNN BiLSTM -AM [8]
RIO	0.501042	0.52344	0.51719	0.49375
SO	0.522917	**0.50000	0.51719	0.51406
SPY	0.55625	**0.47813	0.50313	0.52344
SYT	**0.485417	0.51250	0.50469	0.49688
T	0.514583	0.53750	0.50938	0.50469
TM	0.511458	0.48125	0.50000	**0.47188
TOT	0.515625	**0.48438	0.52031	0.52188
UNH	0.525	**0.46406	0.51563	0.51406
UPS	0.519792	0.51094	0.49844	**0.47813
VALE	0.513542	0.49688	0.49219	0.51875
VZ	0.551042	**0.46406	0.51406	0.52813
WFC	0.523958	0.52188	**0.47656	0.48750
WMT	0.504167	**0.47656	0.51563	0.50000
XOM	0.46875	**0.44844	0.48906	0.51563

Experiment Result

- Red : High Accuracy
- Asterisk (**) : Low Accuracy

- Accuracy (Index : NASDAQ) – early fusion (1/3)

Data	Proposed -Early fusion Transformer	DTML [2]	AdvLSTM [7]	CNN BiLSTM AM [8]	Proposed -Early fusion Transformer	DTML [2]	AdvLSTM [7]	CNN BiLSTM AM [8]
SPY	0.55625	0.478125	0.503125	0.523438	1	4	3	2
VZ	0.551042	0.464063	0.514063	0.528125	1	4	3	2
AAPL	0.544792	0.520313	0.476563	0.482813	1	2	4	3
MO	0.544792	0.489063	0.5125	0.554688	2	4	3	1
BAC	0.542708	0.515625	0.507813	0.484375	1	2	3	4
NGG	0.538542	0.5375	0.515625	0.514063	1	2	3	4
BA	0.5375	0.50625	0.515625	0.521875	1	4	3	2
DCM	0.534375	0.526563	0.484375	0.489063	1	2	4	3
DIS	0.529167	0.4875	0.510938	0.5125	1	4	3	2
EXC	0.529167	0.495313	0.46875	0.507813	1	3	4	2
MA	0.528125	0.5125	0.492188	0.501563	1	2	4	3
PTR	0.528125	0.521875	0.484375	0.525	1	3	4	2
RDS-B	0.527083	0.4875	0.496875	0.517188	1	4	3	2
GE	0.526042	0.515625	0.521875	0.504688	1	3	2	4
JNJ	0.526042	0.525	0.482813	0.492188	1	2	4	3
INTC	0.525	0.510938	0.485938	0.53125	2	3	4	1
UNH	0.525	0.464063	0.515625	0.514063	1	4	2	3
WFC	0.523958	0.521875	0.476563	0.4875	1	2	4	3

Experiment Result

- Red : High Accuracy
- Asterisk (**) : Low Accuracy

- Accuracy (Index : NASDAQ) – early fusion (2/3)

Data	Proposed -Early fusion Transformer	DTML [2]	AdvLSTM [7]	CNN BiLSTM AM [8]	Proposed -Early fusion Transformer	DTML [2]	AdvLSTM [7]	CNN BiLSTM AM [8]
SO	0.522917	0.5	0.517188	0.514063	1	4	2	3
AMZN	0.520833	0.5125	0.53125	0.496875	2	3	1	4
MMM	0.519792	0.479688	0.5125	0.46875	1	3	2	4
UPS	0.519792	0.510938	0.498438	0.478125	1	2	3	4
DOW	0.51875	0.5	0.482813	0.464063	1	2	3	4
NVS	0.51875	0.49375	0.507813	0.5125	1	4	3	2
D	0.517708	0.510938	0.503125	0.528125	2	3	4	1
DUK	0.517708	0.482813	0.473438	0.554688	2	3	4	1
CMCSA	0.516667	0.490625	0.4875	0.520313	2	3	4	1
GOOGL	0.516667	0.529688	0.509375	0.528125	3	1	4	2
PG	0.516667	0.484375	0.525	0.473438	2	3	1	4
TOT	0.515625	0.484375	0.520313	0.521875	3	4	2	1
HD	0.514583	0.504688	0.528125	0.520313	3	4	1	2
JPM	0.514583	0.501563	0.503125	0.50625	1	4	3	2
PEP	0.514583	0.4875	0.503125	0.489063	1	4	2	3
T	0.514583	0.5375	0.509375	0.504688	2	1	3	4
VALE	0.513542	0.496875	0.492188	0.51875	2	3	4	1
KO	0.5125	0.473438	0.479688	0.523438	2	4	3	1

Experiment Result

- Red : High Accuracy
- Asterisk (**) : Low Accuracy

- Accuracy (Index : NASDAQ) – early fusion (3/3)

Data	Proposed -Early fusion Transformer	DTML [2]	AdvLSTM [7]	CNN BiLSTM AM [8]	Proposed -Early fusion Transformer	DTML [2]	AdvLSTM [7]	CNN BiLSTM AM [8]
ORCL	0.511458	0.509375	0.504688	0.529688	2	3	4	1
TM	0.511458	0.48125	0.5	0.471875	1	3	2	4
NTT	0.509375	0.567188	0.515625	0.490625	3	1	2	4
CVX	0.507292	0.48125	0.496875	0.503125	1	4	3	2
MRK	0.50625	0.49375	0.49375	0.498438	1	3	3	2
WMT	0.504167	0.476563	0.515625	0.5	2	4	1	3
MSFT	0.501042	0.534375	0.495313	0.5125	3	1	4	2
RIO	0.501042	0.523438	0.517188	0.49375	3	1	2	4
PFE	0.492708	0.485938	0.478125	0.50625	2	3	4	1
BRK-B	0.490625	0.471875	0.5	0.492188	3	4	1	2
BHP	0.485417	0.478125	0.5	0.48125	2	4	1	3
SYT	0.485417	0.5125	0.504688	0.496875	4	1	2	3
CHL	0.479167	0.523438	0.507813	0.509375	4	1	3	2
XOM	0.46875	0.448438	0.489063	0.515625	3	4	2	1

Data	Proposed -Early fusion Transformer	DTML [2]	AdvLSTM [7]	CNN BiLSTM AM [8]	Proposed -Early fusion Transformer	DTML [2]	AdvLSTM [7]	CNN BiLSTM AM [8]
NASDAQ	0.517563	0.500969	0.501375	0.506344	1.72	2.92	2.86	2.48

Experiment Result

- 특정 모델이 잘 한 종목

early fusion		
SPY	SPDR S&P 500	S&P 500지수를 이용한 ETF
VZ	버라이즌 커뮤니케이션스	통신산업 관련기업(통신사)
AAPL	애플	휴대폰 및 컴퓨터
MO	알트리아그룹	담배(말보로), 와인 판매
BAC	뱅크오브아메리카	은행
NGG	National Grid plc	전기 가스 등 에너지 회사(수송)
BA	보잉	항공기 제조
DCM		
DIS	월트 디즈니	애니메이션 제작사
EXC	Exelon Corporation	에너지 회사
MA	마스터카드	카드회사
PTR	PetroChina Company Limited	중국 에너지 회사
RDS-B		
GE	제너럴 일렉트릭	전력 인프라 기업
JNJ	존슨앤존슨	제약사
INTC	인텔	종합반도체회사
UNH	유나이티드헬스	보험회사
WFC	웰스파고	은행

Related Work

- Modality1 data

data	
Stock price data (7개)	High, Low, Open, Volume, Adj Close, upndown, change
Technical indicator (10개)	위의 Stock data로부터 만들어진 10개의 feature - Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange. Expert Systems with Applications. Yakup Kara, Melek Acar Boyacioglu, Ömer Kaan Baykan. 2011

- Modality2 data

data	
Month Data (12개)	월 정보 데이터를 One-Hot Encoding을 사용하여 추가
Day of the week Data (5개)	요일 정보를 One-Hot Encoding을 사용하여 추가

Related Work

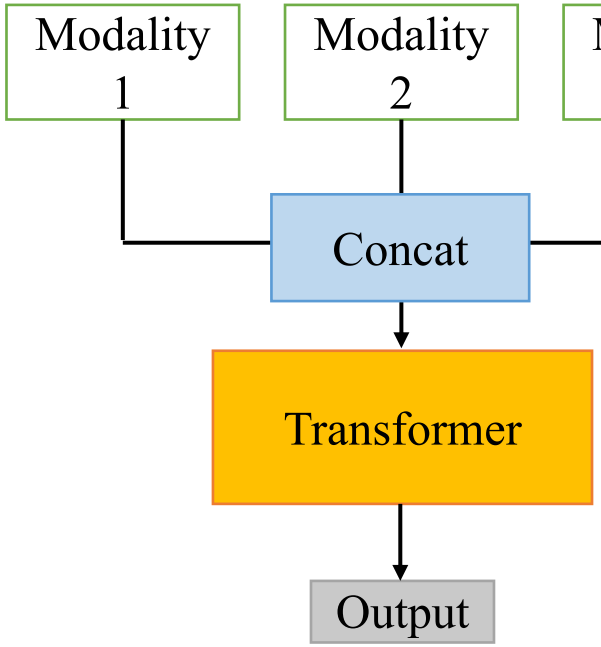
- Modality3 data

data	
nasdaq100	나스닥 상위 100개 종목의 가치의 합을 지수화 한 것
미국 2년물 채권 수익률	‘’
미국 10년물 채권 수익률	‘’
미국 30년물 채권 수익률	‘’
미국 달러 지수	‘’
WTI Oil Price	미국 텍사스산 원유 가격

- 3개의 Modality
- 각 17개,17개, 6개의 feature
- 총 40개를 입력으로 사용

Proposal: *Multimodal Learning*

- **Early Fusion (or data fusion)**



1)

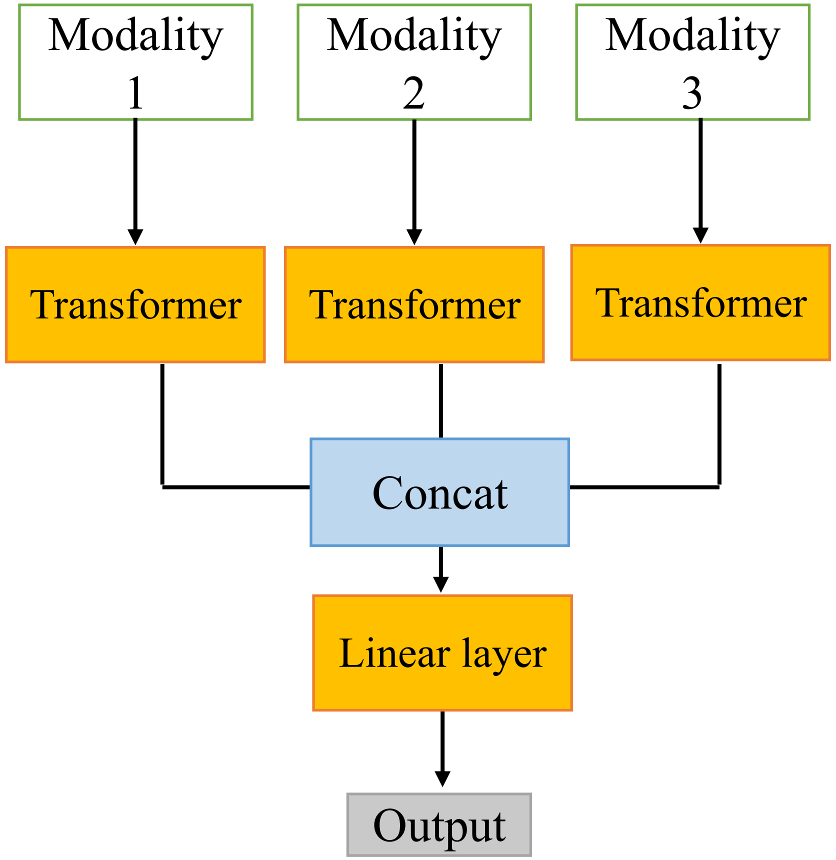
- Combine the data by removing the correlation between two sensors.

2)

- Uses PCA, canonical correlation and independent component analysis

Proposal: *Multimodal Learning*

- **Late Fusion**



- 개별 모델을 통과한 후에 데이터를 합치는 방식으로, 가장 많이 사용되는 방식임

Experiment Result

- Accuracy (Index : NASDAQ)
  - 비교모델 중 가장 좋은 정확성을 보인 모델과 Late fusion의 정확성의 차이가 큰 순서로 정렬

Data	Proposed -Early fusion Transformer	Proposed -Late fusion Transformer	DTML [2]	AdvLSTM [7]	CNN BiLSTM AM [8]	비교모델 중 최대 Accuracy – Late fusion accuracy
PEP	0.514583	0.53854	0.4875	0.50313	0.48906	0.03541
MRK	0.50625	0.53021	0.49375	0.49375	0.49844	0.03177
MA	0.528125	0.54271	0.5125	0.49219	0.50156	0.03021
BAC	0.542708	0.54375	0.51563	0.50781	0.48438	0.02812
RDS-B	0.527083	0.54063	0.4875	0.49688	0.51719	0.02344
SPY	0.55625	0.54271	0.47813	0.50313	0.52344	0.01927
DIS	0.529167	0.52813	0.4875	0.51094	0.5125	0.01563
BHP	0.485417	0.51563	0.47813	0.5	0.48125	0.01563
MMM	0.519792	0.52708	0.47969	0.5125	0.46875	0.01458
NVS	0.51875	0.52604	0.49375	0.50781	0.5125	0.01354
PFE	0.492708	0.51979	0.48594	0.47813	0.50625	0.01354
BRK-B	0.490625	0.50833	0.47188	0.5	0.49219	0.00833
BA	0.5375	0.52813	0.50625	0.51563	0.52188	0.00625
PTR	0.528125	0.53021	0.52188	0.48438	0.525	0.00521
CVX	0.507292	0.50833	0.48125	0.49688	0.50313	0.0052
AAPL	0.544792	0.525	0.52031	0.47656	0.48281	0.00469
MO	0.544792	0.55625	0.48906	0.5125	0.55469	0.00156
TOT	0.515625	0.52292	0.48438	0.52031	0.52188	0.00104

Experiment Result

- Accuracy (Index : NASDAQ)
  - 비교모델 중 가장 좋은 정확성을 보인 모델과 Late fusion의 정확성의 차이가 큰 순서로 정렬

Data	Proposed -Early fusion Transformer	Proposed -Late fusion Transformer	DTML [2]	AdvLSTM [7]	CNN BiLSTM AM [8]	비교모델 중 최대 Accuracy – Late fusion accuracy
GE	0.526042	0.52188	0.51563	0.52188	0.50469	0
JPM	0.514583	0.50521	0.50156	0.50313	0.50625	-0.00104
EXC	0.529167	0.50625	0.49531	0.46875	0.50781	-0.00156
SO	0.522917	0.51563	0.5	0.51719	0.51406	-0.00156
UNH	0.525	0.5125	0.46406	0.51563	0.51406	-0.00313
CMCSA	0.516667	0.51563	0.49063	0.4875	0.52031	-0.00468
TM	0.511458	0.49375	0.48125	0.5	0.47188	-0.00625
WMT	0.504167	0.50833	0.47656	0.51563	0.5	-0.0073
D	0.517708	0.51979	0.51094	0.50313	0.52813	-0.00834
NGG	0.538542	0.52708	0.5375	0.51563	0.51406	-0.01042
KO	0.5125	0.5125	0.47344	0.47969	0.52344	-0.01094
UPS	0.519792	0.49792	0.51094	0.49844	0.47813	-0.01302
JNJ	0.526042	0.51146	0.525	0.48281	0.49219	-0.01354
HD	0.514583	0.51458	0.50469	0.52813	0.52031	-0.01355
WFC	0.523958	0.50833	0.52188	0.47656	0.4875	-0.01355
DCM	0.534375	0.51146	0.52656	0.48438	0.48906	-0.0151
AMZN	0.520833	0.51458	0.5125	0.53125	0.49688	-0.01667
DUK	0.517708	0.53438	0.48281	0.47344	0.55469	-0.02031

Experiment Result

- Accuracy (Index : NASDAQ)
  - 비교모델 중 가장 좋은 정확성을 보인 모델과 Late fusion의 정확성의 차이가 큰 순서로 정렬

Data	Proposed -Early fusion Transformer	Proposed -Late fusion Transformer	DTML [2]	AdvLSTM [7]	CNN BiLSTM -AM [8]	비교모델 중 최대 Accuracy – Late fusion accuracy
ORCL	0.511458	0.50729	0.50938	0.50469	0.52969	-0.0224
CHL	0.479167	0.50104	0.52344	0.50781	0.50938	-0.0224
DOW	0.51875	0.47708	0.5	0.48281	0.46406	-0.02292
SYT	0.485417	0.48854	0.5125	0.50469	0.49688	-0.02396
GOOGL	0.516667	0.50417	0.52969	0.50938	0.52813	-0.02552
INTC	0.525	0.50417	0.51094	0.48594	0.53125	-0.02708
VZ	0.551042	0.49896	0.46406	0.51406	0.52813	-0.02917
PG	0.516667	0.49375	0.48438	0.525	0.47344	-0.03125
RIO	0.501042	0.49167	0.52344	0.51719	0.49375	-0.03177
VALE	0.513542	0.48125	0.49688	0.49219	0.51875	-0.0375
MSFT	0.501042	0.49479	0.53438	0.49531	0.5125	-0.03959
XOM	0.46875	0.46979	0.44844	0.48906	0.51563	-0.04584
T	0.514583	0.4875	0.5375	0.50938	0.50469	-0.05
NTT	0.509375	0.51563	0.56719	0.51563	0.49063	-0.05156

Experiment Result

- Accuracy (Index : NASDAQ)
  - 비교모델 중 가장 좋은 정확성을 보인 모델과 Early fusion의 정확성의 차이가 큰 순서로 정렬

Data	Proposed -Early fusion Transformer	Proposed -Late fusion Transformer	DTML [2]	AdvLSTM [7]	CNN BiLSTM AM [8]	비교모델 중 최대 Accuracy – Early fusion accuracy
SPY	0.55625	0.54271	0.47813	0.50313	0.52344	0.03281
BAC	0.542708	0.54375	0.51563	0.50781	0.48438	0.027078
AAPL	0.544792	0.525	0.52031	0.47656	0.48281	0.024482
VZ	0.551042	0.49896	0.46406	0.51406	0.52813	0.022912
EXC	0.529167	0.50625	0.49531	0.46875	0.50781	0.021357
DOW	0.51875	0.47708	0.5	0.48281	0.46406	0.01875
DIS	0.529167	0.52813	0.4875	0.51094	0.5125	0.016667
MA	0.528125	0.54271	0.5125	0.49219	0.50156	0.015625
BA	0.5375	0.52813	0.50625	0.51563	0.52188	0.01562
TM	0.511458	0.49375	0.48125	0.5	0.47188	0.011458
PEP	0.514583	0.53854	0.4875	0.50313	0.48906	0.011453
RDS-B	0.527083	0.54063	0.4875	0.49688	0.51719	0.009893
UNH	0.525	0.5125	0.46406	0.51563	0.51406	0.00937
UPS	0.519792	0.49792	0.51094	0.49844	0.47813	0.008852
JPM	0.514583	0.50521	0.50156	0.50313	0.50625	0.008333
DCM	0.534375	0.51146	0.52656	0.48438	0.48906	0.007815
MRK	0.50625	0.53021	0.49375	0.49375	0.49844	0.00781
MMM	0.519792	0.52708	0.47969	0.5125	0.46875	0.007292

Experiment Result

- Accuracy (Index : NASDAQ)
  - 비교모델 중 가장 좋은 정확성을 보인 모델과 Early fusion의 정확성의 차이가 큰 순서로 정렬

Data	Proposed -Early fusion Transformer	Proposed -Late fusion Transformer	DTML [2]	AdvLSTM [7]	CNN BiLSTM AM [8]	비교모델 중 최대 Accuracy – Early fusion accuracy
NVS	0.51875	0.52604	0.49375	0.50781	0.5125	0.00625
SO	0.522917	0.51563	0.5	0.51719	0.51406	0.005727
CVX	0.507292	0.50833	0.48125	0.49688	0.50313	0.004162
GE	0.526042	0.52188	0.51563	0.52188	0.50469	0.004162
PTR	0.528125	0.53021	0.52188	0.48438	0.525	0.003125
WFC	0.523958	0.50833	0.52188	0.47656	0.4875	0.002078
NGG	0.538542	0.52708	0.5375	0.51563	0.51406	0.001042
JNJ	0.526042	0.51146	0.525	0.48281	0.49219	0.001042
CMCSA	0.516667	0.51563	0.49063	0.4875	0.52031	-0.00364
VALE	0.513542	0.48125	0.49688	0.49219	0.51875	-0.00521
INTC	0.525	0.50417	0.51094	0.48594	0.53125	-0.00625
TOT	0.515625	0.52292	0.48438	0.52031	0.52188	-0.00626
PG	0.516667	0.49375	0.48438	0.525	0.47344	-0.00833
BRK-B	0.490625	0.50833	0.47188	0.5	0.49219	-0.00938
MO	0.544792	0.55625	0.48906	0.5125	0.55469	-0.0099
AMZN	0.520833	0.51458	0.5125	0.53125	0.49688	-0.01042
D	0.517708	0.51979	0.51094	0.50313	0.52813	-0.01042
KO	0.5125	0.5125	0.47344	0.47969	0.52344	-0.01094

Experiment Result

- Accuracy (Index : NASDAQ)
  - 비교모델 중 가장 좋은 정확성을 보인 모델과 Early fusion의 정확성의 차이가 큰 순서로 정렬

Data	Proposed -Early fusion Transformer	Proposed -Late fusion Transformer	DTML [2]	AdvLSTM [7]	CNN BiLSTM -AM [8]	비교모델 중 최대 Accuracy – Early fusion accuracy
WMT	0.504167	0.50833	0.47656	0.51563	0.5	-0.01146
GOOGL	0.516667	0.50417	0.52969	0.50938	0.52813	-0.01302
PFE	0.492708	0.51979	0.48594	0.47813	0.50625	-0.01354
HD	0.514583	0.51458	0.50469	0.52813	0.52031	-0.01355
BHP	0.485417	0.51563	0.47813	0.5	0.48125	-0.01458
ORCL	0.511458	0.50729	0.50938	0.50469	0.52969	-0.01823
RIO	0.501042	0.49167	0.52344	0.51719	0.49375	-0.0224
T	0.514583	0.4875	0.5375	0.50938	0.50469	-0.02292
SYT	0.485417	0.48854	0.5125	0.50469	0.49688	-0.02708
MSFT	0.501042	0.49479	0.53438	0.49531	0.5125	-0.03334
DUK	0.517708	0.53438	0.48281	0.47344	0.55469	-0.03698
CHL	0.479167	0.50104	0.52344	0.50781	0.50938	-0.04427
XOM	0.46875	0.46979	0.44844	0.48906	0.51563	-0.04688
NTT	0.509375	0.51563	0.56719	0.51563	0.49063	-0.05781

윤건일 개인연구 진행상황

220830

중앙대학교 석사과정

윤건일

rjsdlf45@gmail.com

30th Aug, 2022

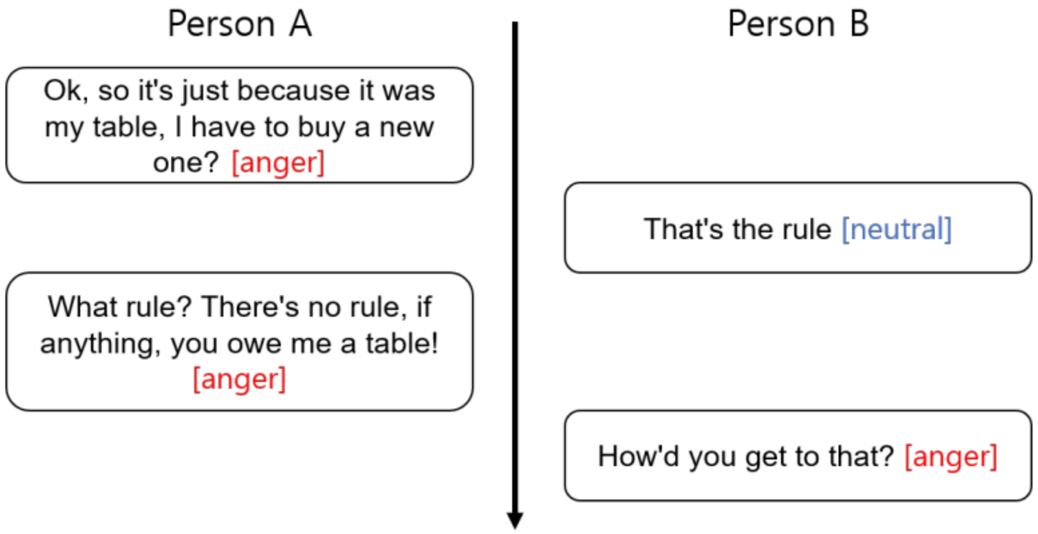
Auto ML

http://ml.cau.ac.kr/

Contents

Emotion Recognition in Conversation

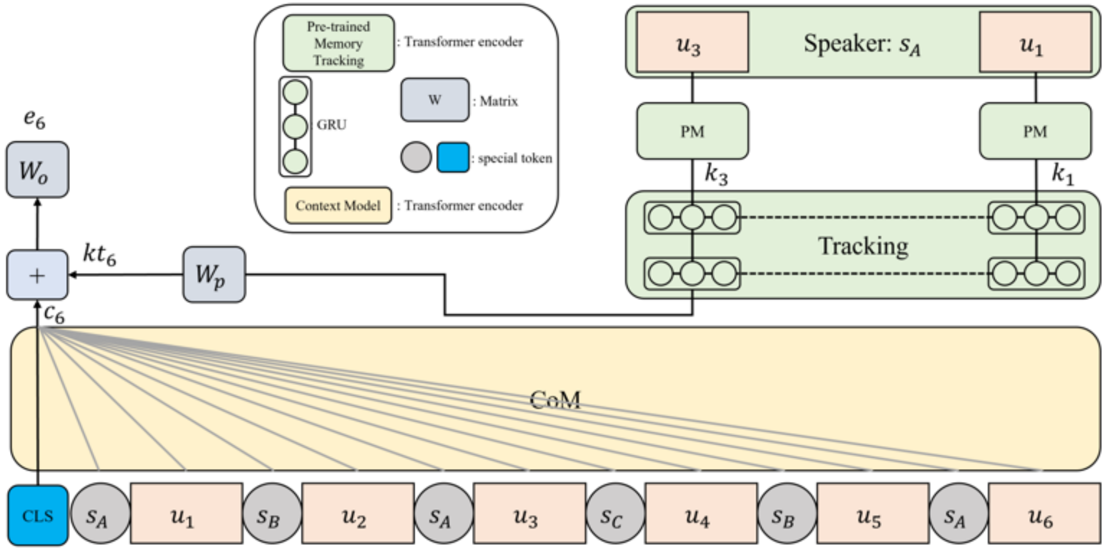
- **Tasks : ERC**



- Recognition of emotions in conversations for **each speaker** and **each utterance.**
- The point is that the current emotion depends on the previous utterance.
- The goal is to recognize **not only positive or negative emotions**, but **also neutral, angry, shy …** emotions.

Emotion Recognition in Conversation

- **CoMPM : Context embedding module + Pre-trained memory Module[1]**



Used PLM

- **BERT**
- **RoBERTa**
- **GPT2**

- **CoMPM** uses two **PLM**s, one for **context modeling** and the other for **speaker pretrained memory**.
- **CoM** is used by concatenating the **previous speakers** and **utterances** as input.
- **PM** uses **only the speaker's utterance** to be predicted as input by using GRU modules.

Emotion Recognition in Conversation

- **Dataset for ERC**

	Dialogues			Utterances			Classes
	Train	Dev	Test	Train	Dev	Test	
IEMOCAP	108	12	31	5,163	647	1,623	6
DailyDialog	11,118	1,000	1,000	87,170	8,069	7,740	7
MELD	1,038	114	280	9,989	1,109	2,610	3, 7
EmoryNLP	713	99	85	9,934	1,344	1,328	3, 7

- **MELD(Multimodal Emotion Lines Dataset)[1] :** Audio, Image, Text for Emotion(anger, disgust, sadness, fear, neutrality), Sentiment(pos, neg, neu)
- **DailyDialog[2]:** 2 speaker dataset (anger, disgust, fear, joy, surprise, sadness, neutral)
- **IEMOCAP[3]:** 10 speaker dataset Emotion(happy, sad, angry, excited, frustrated, neutral), Sentiment(pos, neg, neu)
- **EmoryNLP[4]:** Like MELD, Friends TV show dataset (joyful, peaceful, powerful, scared, mad, sad, neutral)

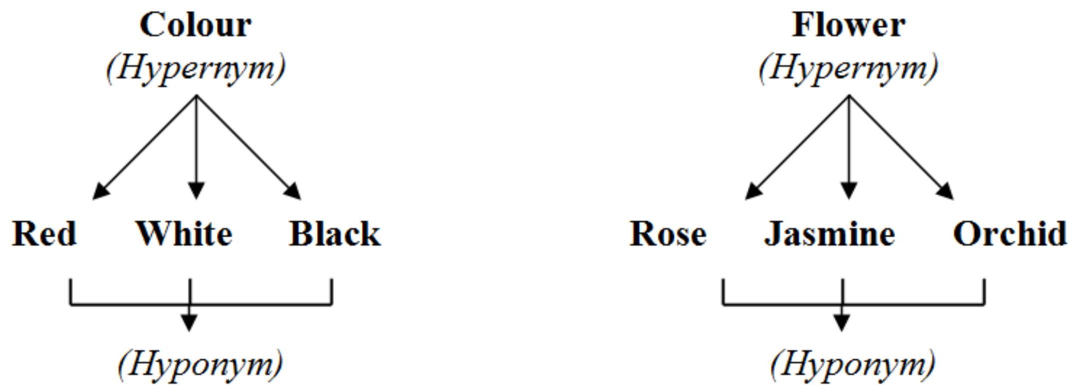
Emotion Recognition in Conversation

- **Experimental results**

	PLM		**Task** (f1score)		**Classes**
	COM	**PM**	**Emotion**	**Sentiment**	
IEMOCAP	RoBERTa	RoBERTa	0.6657		6
DailyDialog					7
MELD					3, 7
EmoryNLP					3, 7

Hypernym Discovery

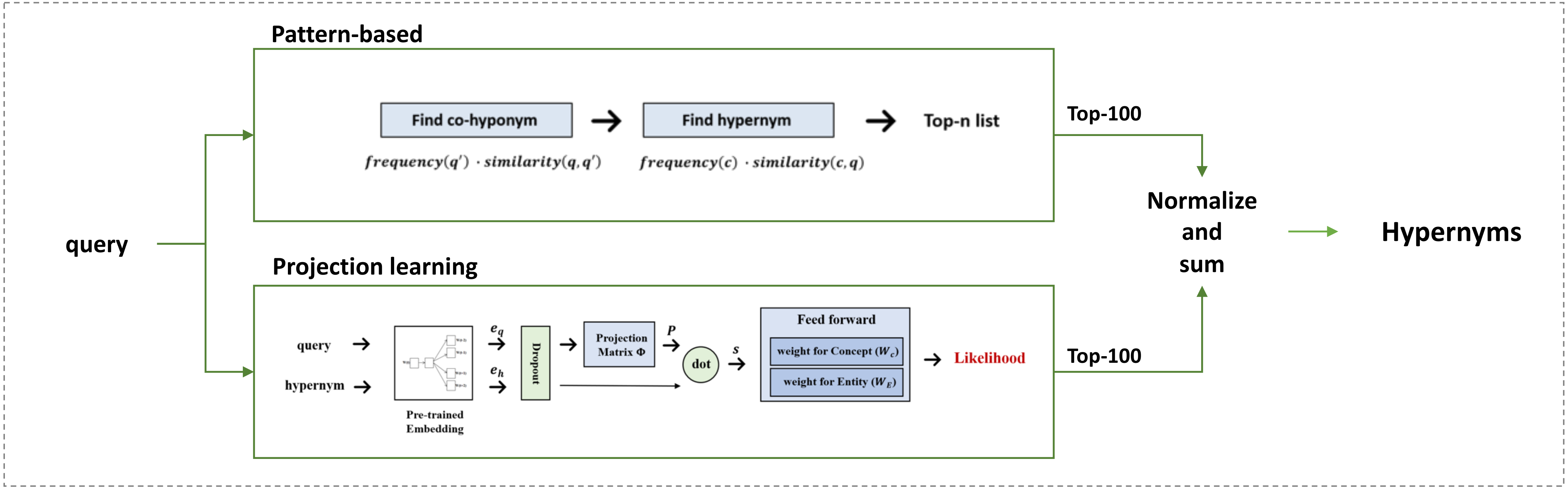
- **Tasks : Hypernym Discovery**



- **Hyponym** is a word or phrase that semantic field is <u>more specific</u> than its **hypernym.**
- **(Color, Red)** is an example of hypernym relation.

Hypernym Discovery

- **CRIM[6] : Hybrid approach – pattern based + projection learning**

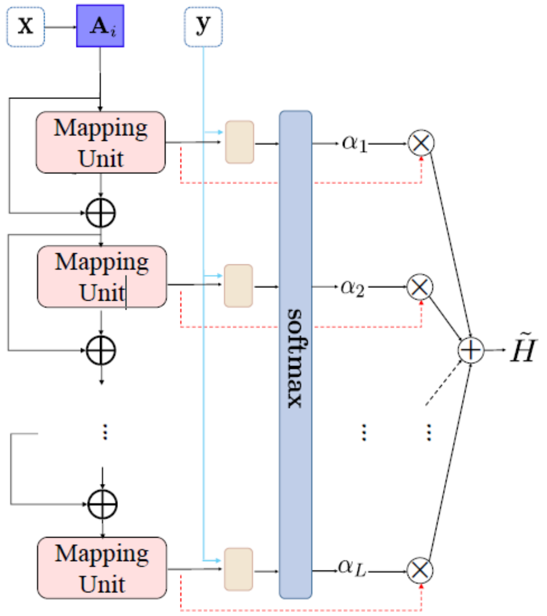


- CRIM model is a hybrid approach that combines **pattern-based** and **projection learning** approach.
- Both approaches use **word2vec embeddings** for <u>word representation in vector space</u>.

Hypernym Discovery

- **RMM[7] : Recurrent Mapping Model**

Model structure



X : embedding of query
y : embedding of hypernym candidate
A_i : transformation matrix (for two types)
Mapping Unit : mapping word to high level by using shared projection matrix

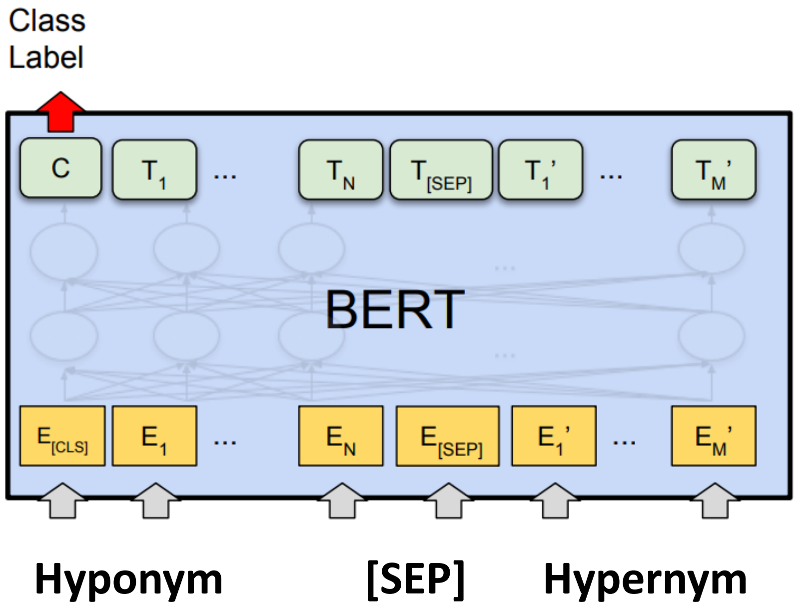
Formulation

- $t = A_i X,\ t \in \mathbb{R}^{d\times 1}, A \in \mathbb{R}^{d\times d}$
- $\hat{h}_{l+1} = W_\phi h_l,\ h_0 = t, W_\phi \in \mathbb{R}^{d\times d}$
- $h_{l+1} = \hat{h}_{l+1} + h_l$
- $\alpha_l = \frac{\exp(h_l \cdot y)}{\sum_{k=1}^{L} \exp(h_k \cdot y)}$
- $\widetilde{H} = \sum_{l=1}^{L} \alpha_l \cdot h_l$
- $s_y = \widetilde{H} \cdot y$

- RMM also use **word2vec** for word embeddings.
- RMM apply **attention mechanism** to aggregate hypernym representations from each hierarchy level.
- In this model, using **list-wise loss function**

Hypernym Discovery

- **Proposed Method : Contextual embedding for word representation**



- Using **Language Model**(Like **BERT, RoBERTa** ..) for contextual embedding.
- Input hyponym and hypernym with [SEP] token.

Thank you

References

[1] Joosung Lee and Wooin Lee. 2022. CoMPM: Context Modeling with Speaker's Pre-trained Memory Tracking for Emotion Recognition in Conversation. In Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 5669–5679, Seattle, United States. Association for Computational Linguistics.

[2] MELD :Soujanya Poria, Devamanyu Hazarika, Navonil Majumder, Gautam Naik, Erik Cambria, and Rada Mihalcea. 2019. MELD: A multimodal multi-party dataset for emotion recognition in conversations. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 527–536, Florence, Italy. Association for Computational Linguistics.

[3] DailyDialog : Yanran Li, Hui Su, Xiaoyu Shen, Wenjie Li, Ziqiang Cao, and Shuzi Niu. 2017. DailyDialog: A manually labelled multi-turn

dialogue dataset. In Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pages 986–995, Taipei, Taiwan. Asian Federation of Natural Language Processing.

[4] IEMOCAP : Carlos Busso, Murtaza Bulut, Chi-Chun Lee, Abe Kazemzadeh, Emily Mower, Samuel Kim, Jeannette N. Chang, Sungbok Lee, and Shrikanth S. Narayanan. 2008. Iemocap: interactive emotional dyadic motion capture database. Lang. Resour. Evaluation, 42(4):335–359.

[5] Sayyed M. Zahiri and Jinho D. Choi. 2018. Emotion detection on TV show transcripts with sequence-based convolutional neural networks.

In The Workshops of the The Thirty-Second AAAI Conference on Artificial Intelligence, New Orleans, Louisiana, USA, February 2-7, 2018, volume WS-18 of AAAI Workshops, pages 44–52. AAAI Press.

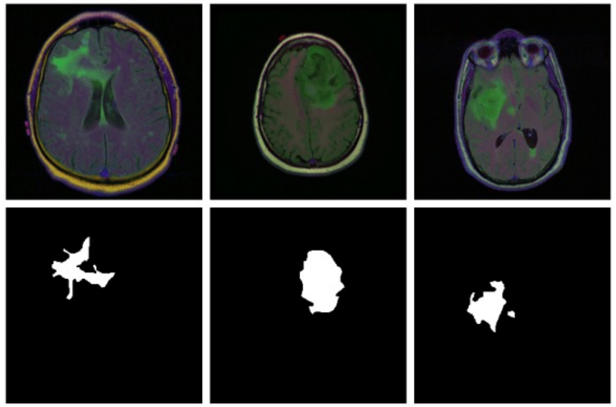
Individual Studies:

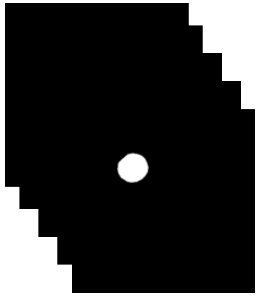
Sujeong Han

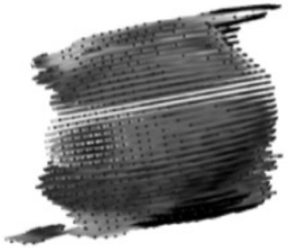
Progress

- 개인연구
- 중앙대병원 갑상선 연구
- ETRI

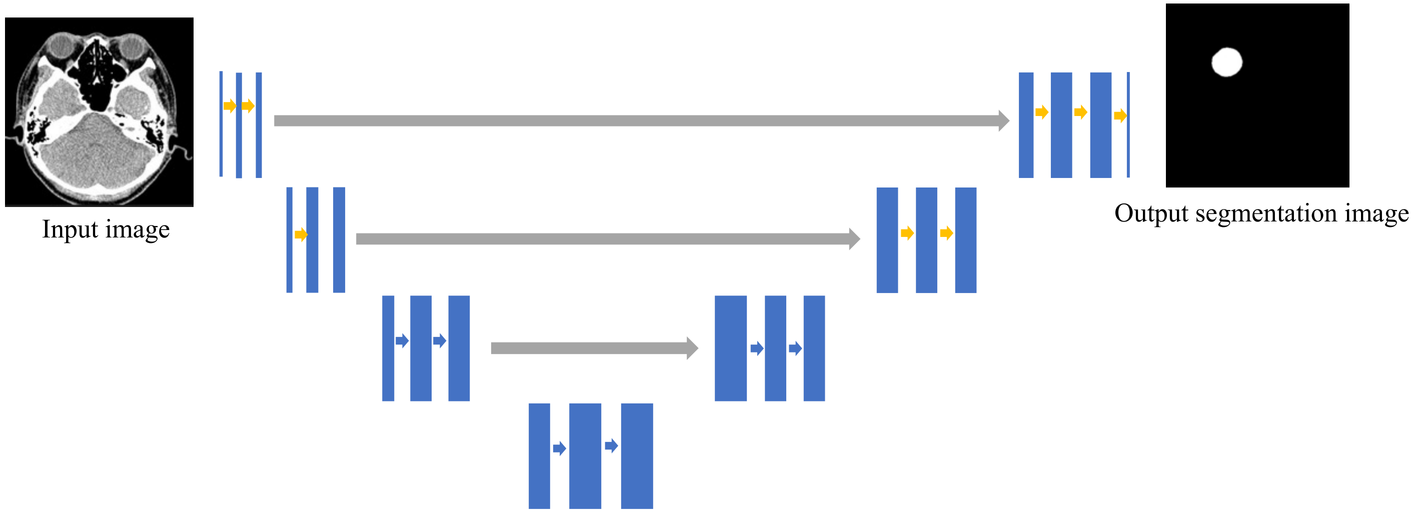
Tasks







Frameworks



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
- MAML
- 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

Experiment Settings

- Datasets (Medial Rectus, Lateral Rectus)
  - Number of Samples: 1237 patients
  - Number of Classes: 2 (Target Anatomical Structure & Background)
  - Input: (512, 512), Gray-scale CT images

- Cross-Validation for Test Performance
  - 10 iterations
  - Repeated Random Sub-sampling Validation (Stratified)

- Hyper Parameter Setting:

Model	Batch Size	Epochs	Loss Function	Optimizer	LR	Weight Decay
Unet (Ronneberger et al., 2015) DeepLab_V3_Plus (Liang-Chieh Chen et al., 2018) HarDNet-MSEG (Huang et al., 2021)	16	30	Tversky Focal Loss (Abraham et al., 2019)	AdamW (Loshchilov et al., 2017)	1e-3	1e-4

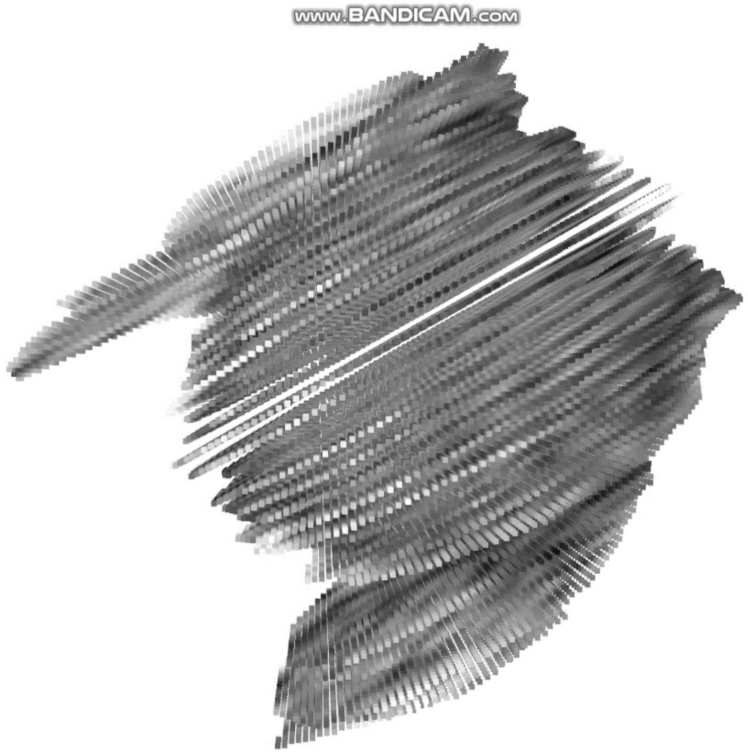
Axial 3D Segmentation

➢ 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)

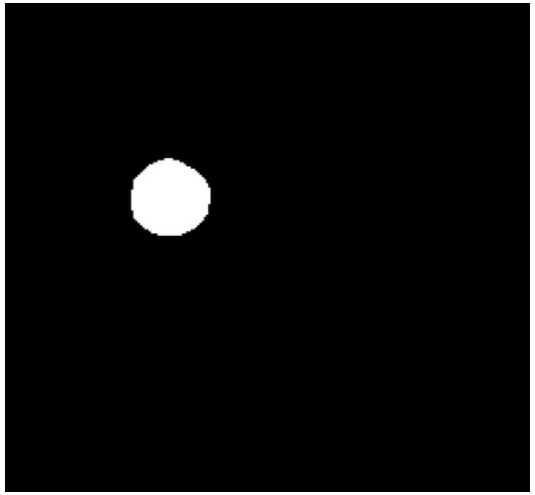
3D Segmentation

1075 환자 수동 세그멘테이션 시각화



101026 환자 자동 세그멘테이션 시각화

Eyeball 부피 구하기



- 10033.dcm ~ 10049.dcm
- Pixel (가로× 세로×높이) : 0.3515625(mm)*0.3515625(mm)*1(mm)
- 101026 eyeball pixel의 총 개수 : 216,441
- 101026 eyeball의 부피 : 8.95 cc

Patient Number	L_Eye volume	L_Eye volume(model. Ver)	Error
101026	8.9	8.95	+0.05
101177	8.76	8.49	-0.27
101414	7.51	8.15	-0.64
101479	9.54	6.34	-3.2
200142	9.65	6.38	-3.24

Restoration of Homoglyph Attack on Spam Using BERT Masked Language Model

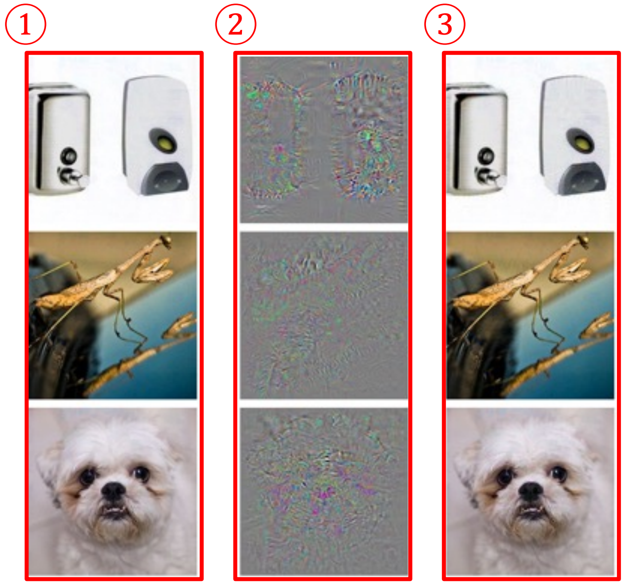
AutoML Lab

Hanyong Lee

Introduction

NLP Attack

Adversarial examples in **Computer Vision**



① correctly predicted samples
② image difference between ① and ③
③ images predicted as “ostrich”

Adversarial examples in **Natural Language Processing**

?

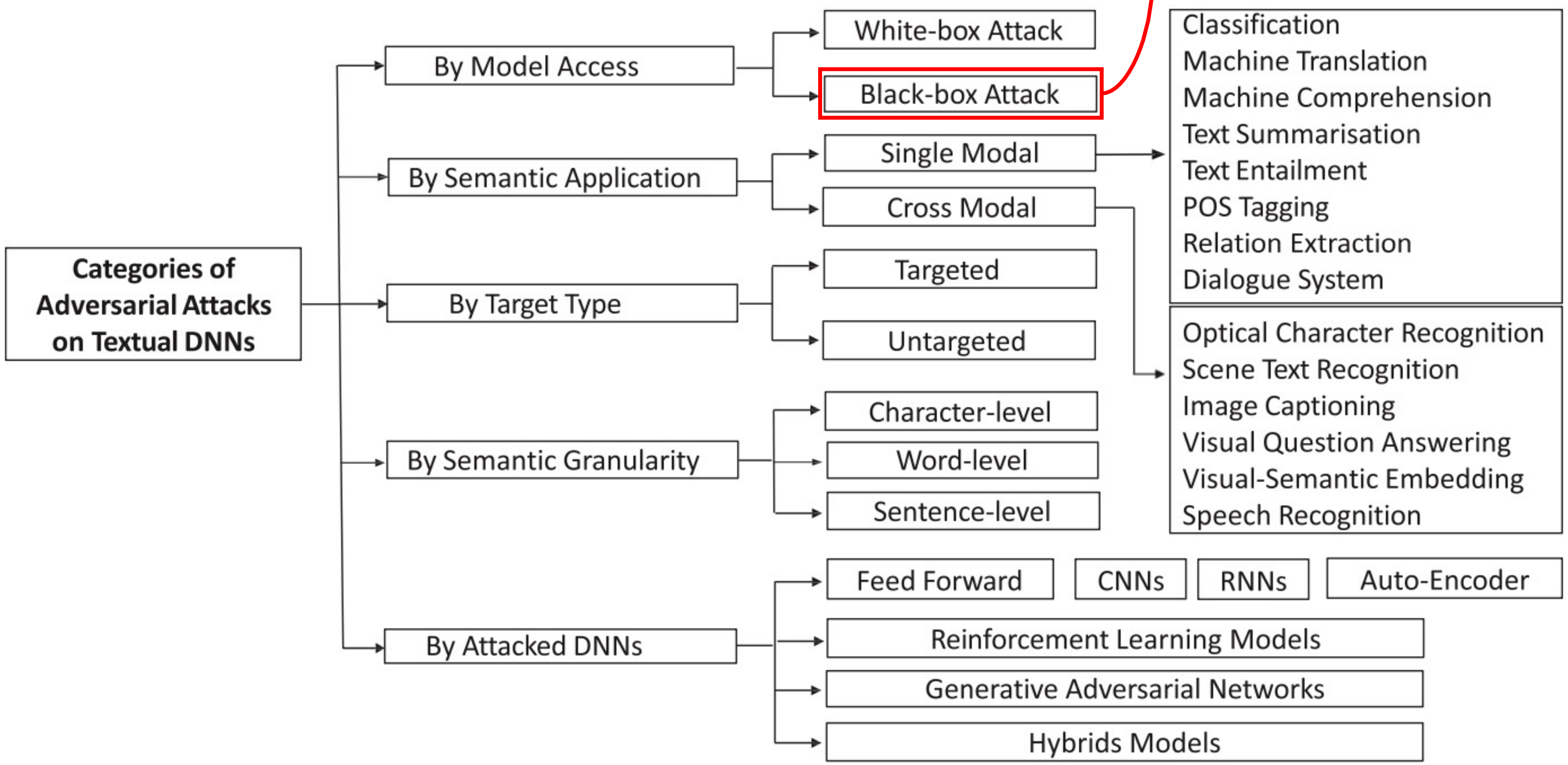
Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., & Fergus, R. (2013). Intriguing properties of neural networks. arXiv preprint arXiv:1312.6199.

Introduction

NLP Attack

Adversarial examples in **Natural Language Processing**

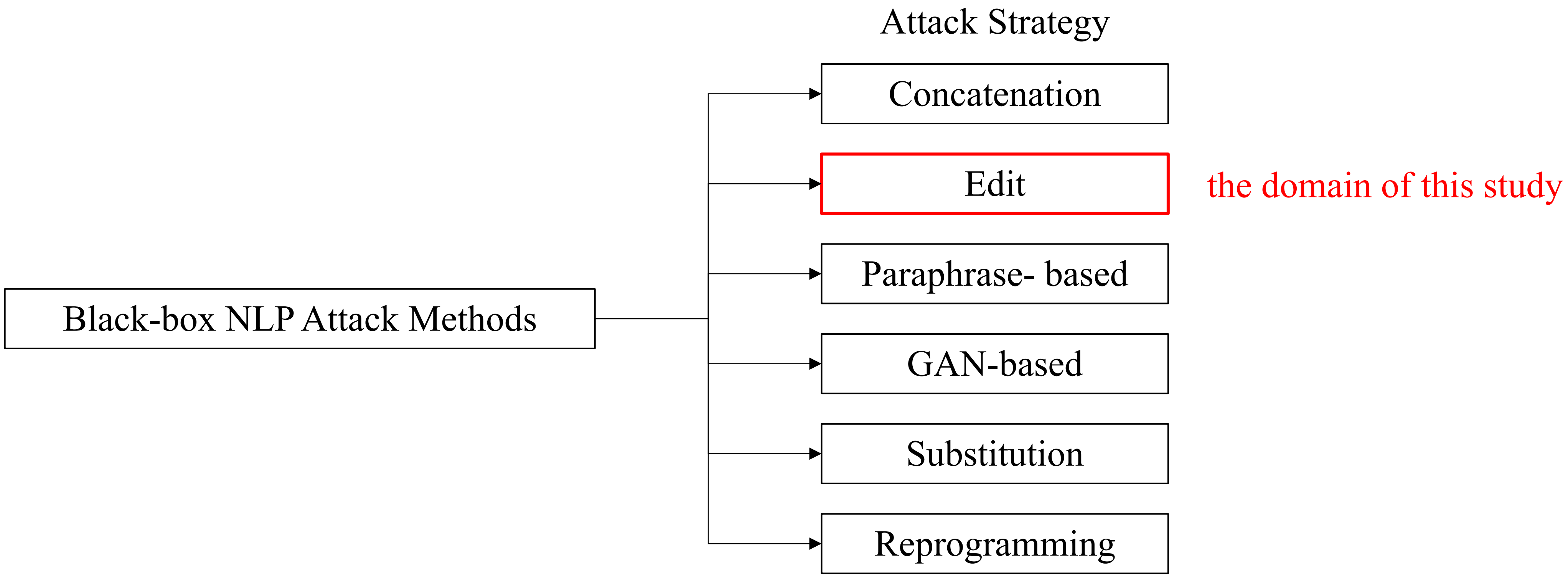
the domain of this study



Akhtar, N., Mian, A., Kardan, N., & Shah, M. (2021). Advances in adversarial attacks and defenses in computer vision: A survey. IEEE Access, 9, 155161-155196.

NLP Attack

Black-box NLP Attack Methods

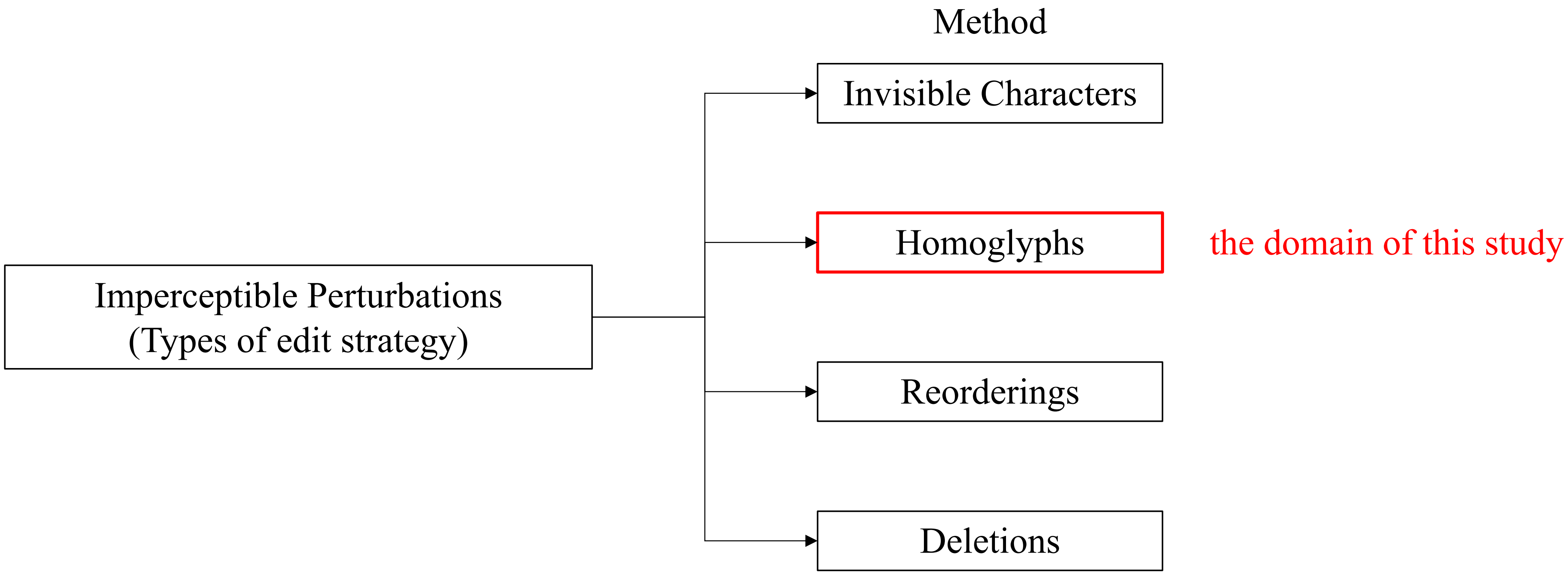


Akhtar, N., Mian, A., Kardan, N., & Shah, M. (2021). Advances in adversarial attacks and defenses in computer vision: A survey. IEEE Access, 9, 155161-155196.

Related Work

Homoglyph Attack

Imperceptible Perturbations

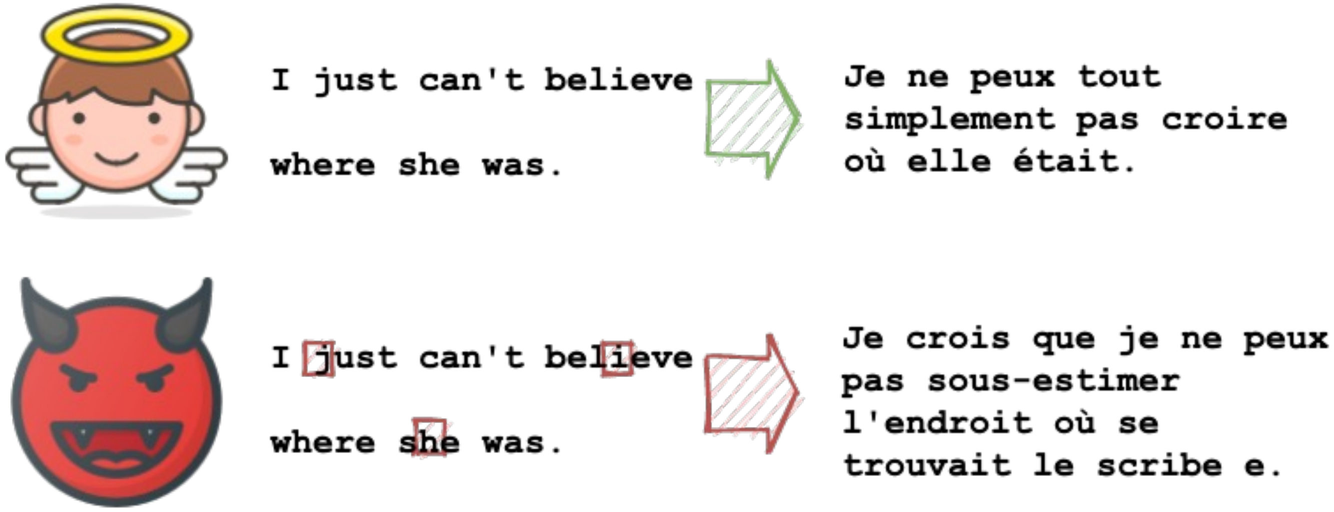


Boucher, N., Shumailov, I., Anderson, R., & Papernot, N. (2022, May). Bad characters: Imperceptible nlp attacks. In 2022 IEEE Symposium on Security and Privacy (SP) (pp. 1987-2004). IEEE.

Related Work

Homoglyph Attack

Example of an attack using homoglyph to confuse English-Franch translation task.



→ wrong result

“j” is replaced with U+3F3(“j”), “i” with U+456(“i”), and “h” with U+4BB(“h”).

Boucher, N., Shumailov, I., Anderson, R., & Papernot, N. (2022, May). Bad characters: Imperceptible nlp attacks. In 2022 IEEE Symposium on Security and Privacy (SP) (pp. 1987-2004). IEEE.

Related Work

Defend Methods of Homoglyph Attack

Translate table for confusable characters

```
⋮
1D4C5 ; 0070 ;  MA     # ( 𝓅 → p ) MATHEMATICAL SCRIPT SMALL P → LATIN SMALL LETTER P  #
1D4F9 ; 0070 ;  MA     # ( 𝓹 → p ) MATHEMATICAL BOLD SCRIPT SMALL P → LATIN SMALL LETTER P    #
1D52D ; 0070 ;  MA     # ( 𝔭 → p ) MATHEMATICAL FRAKTUR SMALL P → LATIN SMALL LETTER P #
1D561 ; 0070 ;  MA     # ( 𝕡 → p ) MATHEMATICAL DOUBLE-STRUCK SMALL P → LATIN SMALL LETTER P   #
⋮
```

It doesn't contain some pairs.

- µ → u
- $ → S
- 5 → S
- …

Unicode Technical Standard #39,"Unicode Security Mechanisms," edited by Mark Davis and Michel Suignard. Version 14.0. 2021-08-25. (https://www.unicode.org/reports/tr39/#Confusable_Detection)

Defend Methods of Homoglyph Attack

SimChar database

H. Suzuki et al. used the following formula to calculate the similarity of two bitmap images of characters.

pixel of image I_1, I_2 at coordinate (i, j)

$$\Delta = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} |I_1(i,j) - I_2(i,j)|$$

character difference

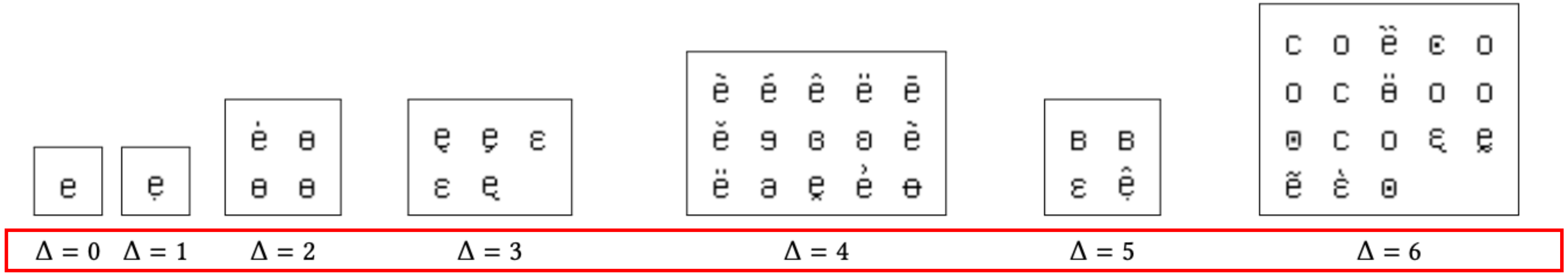


Figure 6: Basic Latin lowercase letter 'e' and characters under different values of the threshold Δ. In this work, characters with Δ ≤ 4 are detected as homoglyphs.

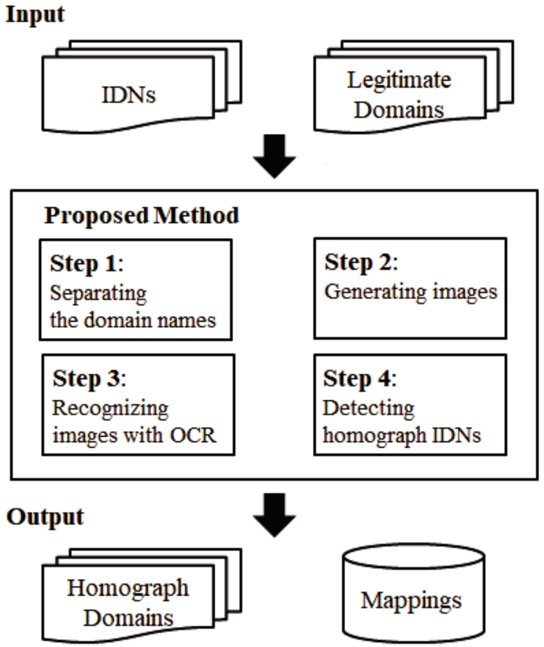
Suzuki, H., Chiba, D., Yoneya, Y., Mori, T., & Goto, S. (2019, October). ShamFinder: An automated framework for detecting IDN homographs. In Proceedings of the Internet Measurement Conference (pp. 449-462).

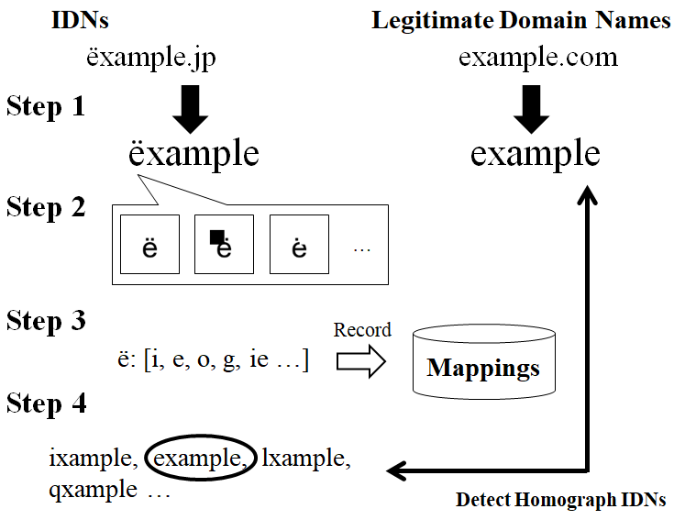
Related Work

Defend Methods of Homoglyph Attack

OCR-based method

Sawabe, Y et al. used the OCR-based Method to detect Homoglyph Attack.





Sawabe, Y., Chiba, D., Akiyama, M., & Goto, S. (2019). Detection method of homograph internationalized domain names with OCR. Journal of Information Processing, 27, 536-544.

Defend Methods of Homoglyph Attack

Spellchecker-based method

Imam, N. H et al. used a spellchecker-based method to restore the homoglyph attacks.

Table 2
Some of the adversarial text examples used in experiments.

Original	Flipping	Swapping	Deletion
Busy	Bu$y	Bsuy	Bsy
Caller	C@1ler	Claler	Cller
Reply	Rep1y	Rpely	Rply
winner	w!nner	wniner	wnner

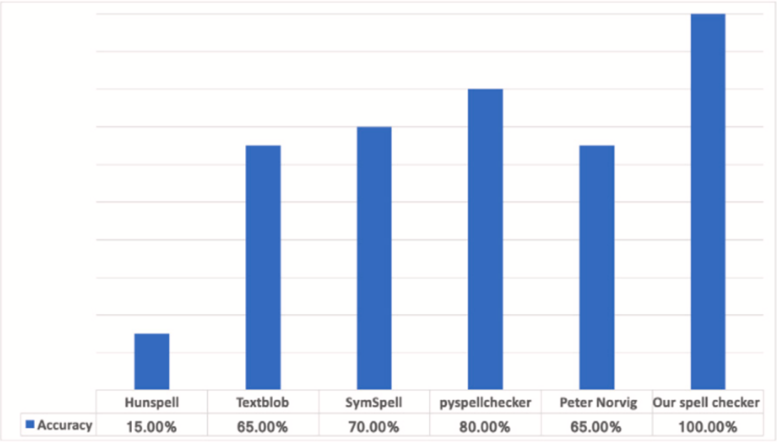
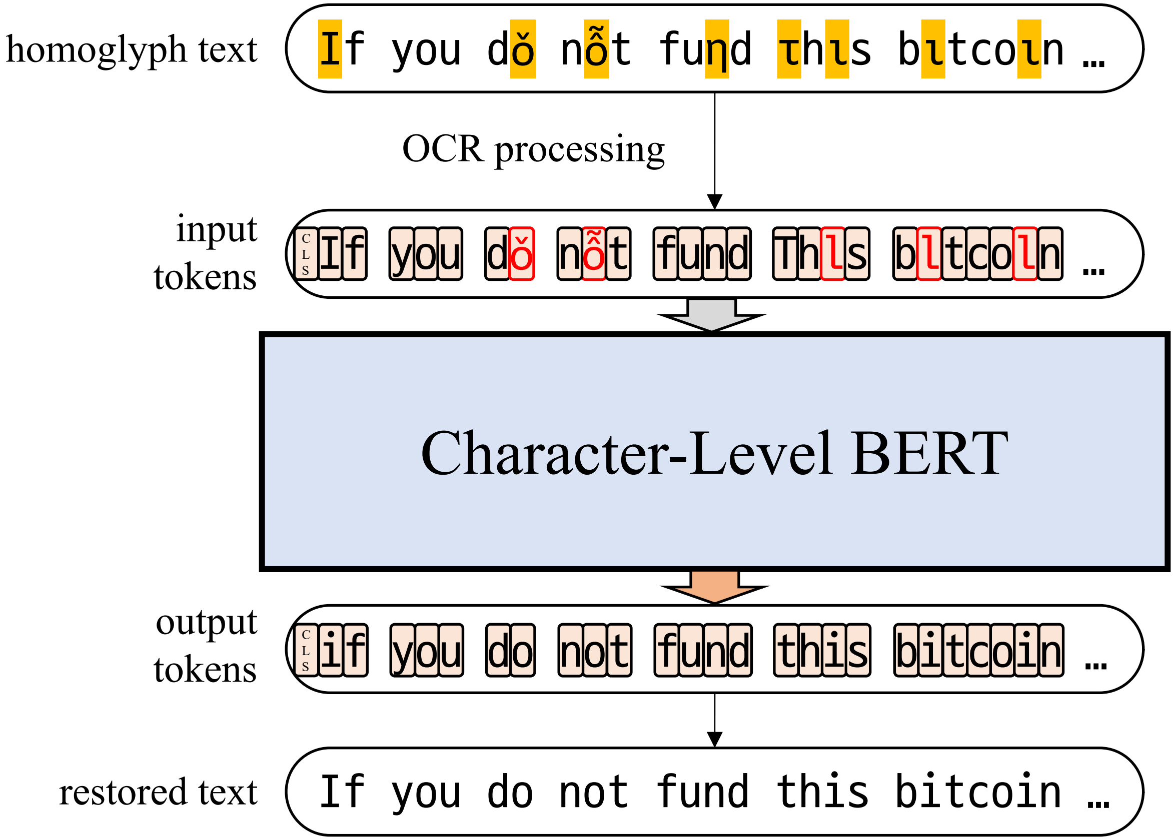


Fig. 9. Flipping.

Imam, N. H., Vassilakis, V. G., & Kolovos, D. (2022). OCR post-correction for detecting adversarial text images. Journal of Information Security and Applications, 66, 103170.

Proposed Method

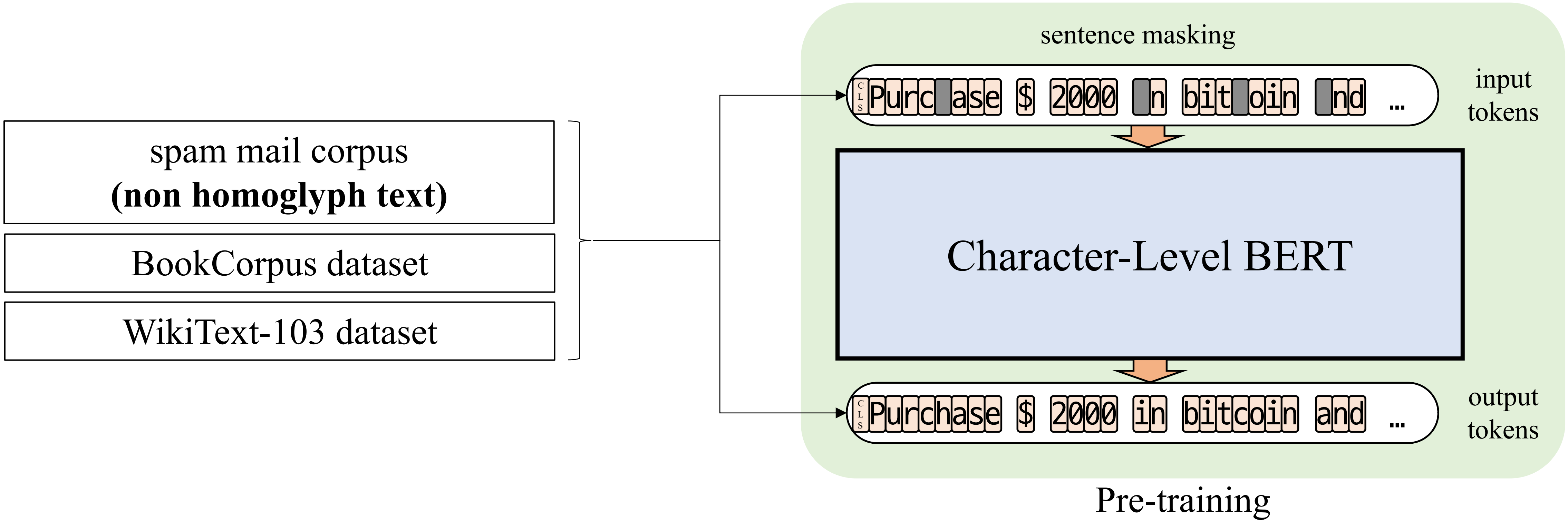
Restoring Process Overview



Experimental Settings

Model Pre-training

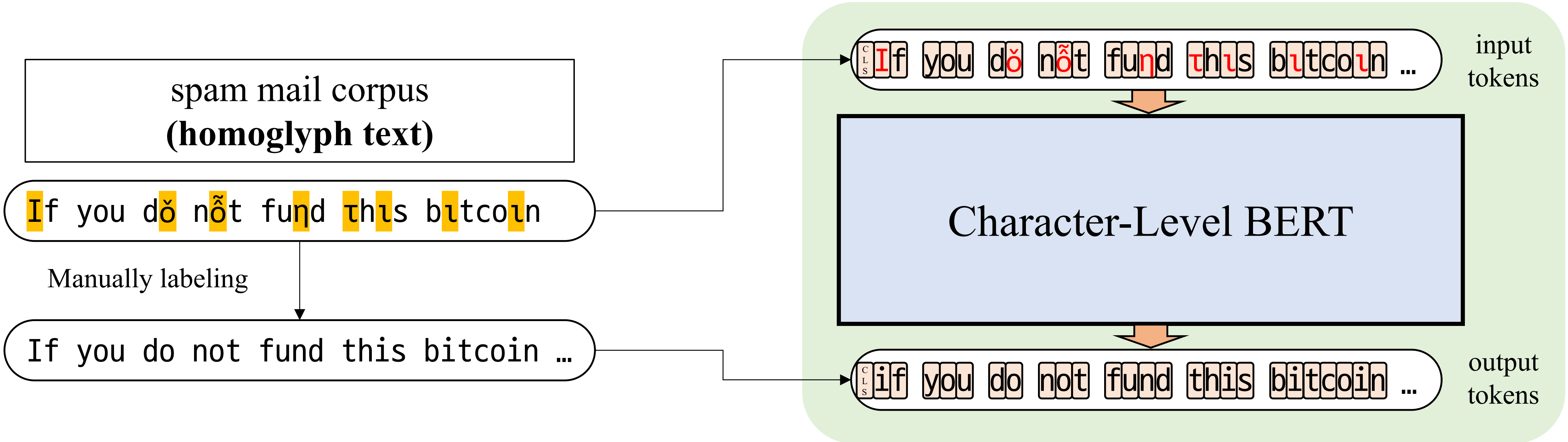
Epochs num: 25 Epochs



Experimental Settings

Model Finetuning

Evaluated with a blind test (Train/Test set split ratio = 8:2 / Iteration = 50 times)



Finetuning

Experimental Results

Finetuned Model Evaluation

Algorithm	Word Accuracy	t-statistic	p-value
OCR	0.6518 ± 0.0028	586.5019	1.5096e-175
SimChar DB	0.5512 ± 0.0046	654.1649	3.4128e-180
Spell Checker	0.9087 ± 0.0011	288.8227	2.0355e-145
Normal BERT	0.6713 ± 0.0029	658.0239	1.9179e-180
Proposed (finetuned)	**0.9959 ± 0.0008**	-	-

Word level accuracy and T-test with the accuracy of the proposed method

Conclusion

Contribution of the Study

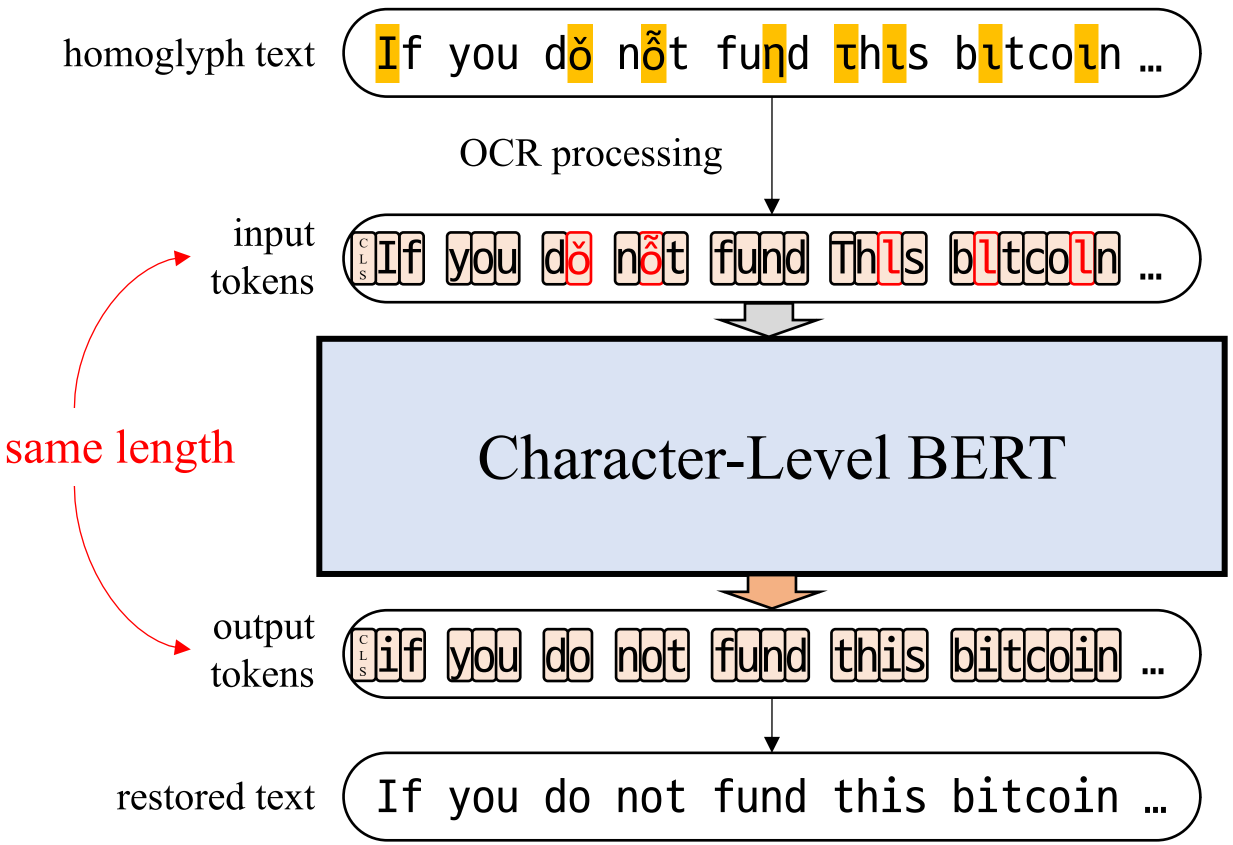
1. Achieve higher performance than traditional methods such as OCR, SimChar DataBase, and spell checker.

2. Propose an automated homoglyph restoration framework by using deep learning.

Conclusion

Limitation of Proposed Model

- The proposed model is a masked language model, limited in that the input/output sentences should be the **same sequence length**.
  - It can be used as a homoglyph of the alphabet ‘m’ by concatenating the letters ‘r’ and ‘n’. But the presented model does not solve this case.
- The proposed model can be applied as a **sequence-to-sequence** model in future studies.



Appendix

A. Proposed Model Information

- Character-level BERT
  - Number of parameters: 86,720,369
  - Max length of input tokens: 512
  - Vocab size: 881

B. Pre-trained Model Evaluation

Evaluated with the **whole test sentences**

Character level BERT is only pretrained with the **WikiText-103, BookCorpus & spam mail datasets**.

Algorithm	Word Accuracy	t-statistic	p-value
OCR	0.6519 ± 0.2976	150.9541	0.0000e+00
SimChar DB	0.5506 ± 0.3627	170.1588	0.0000e+00
Spell Checker	0.9090 ± 0.1620	33.6093	4.5096e-245
Normal BERT	0.6713 ± 0.2679	153.9533	0.0000e+00
Proposed (not finetuned)	**0.9516 ± 0.1289**	-	-

Word level accuracy and T-test with the accuracy of the proposed method

C. Test Output Examples of Pretrained Model

- Example 1
  - input sentence L Neeđ y0ᑌᚱ TOTɑı ατTeNTıOɴ fOʀ The c0MıNg TweNτʏ-fOᑌR hᚱ□,
  - output sentence : i need your total attention for the coming twenty - four hrs,
- Example 2
  - input sentence : ɑftêr thät, i hἀvë ṡtἀrtèd trȧcκi̇ṅġ ɣȯur i̇ṅtérпët ȧctívítìêṡ.
  - output sentence : after that, i have started tracking yo ur internet activities.

C. Test Output Examples of Pretrained Model

- Example 3
  - input sentence : do ṅỏt tгɣ tỏ coṅtáct the ῤoliˈćê àɴd ỏtңêг sëcurity servịces.
  - output sentence : don't try to contact the police and other security services.
- Example 4
  - input sentence : yøuř åccøuňt wås åttåckěd.
  - output sentence : yogr account was attacked.

Individual Studies:

Taekyung Song

Contents

- 산학연 협력 프로젝트
  - 두산 산업차량
- 갑상선 연구
- 개인연구

산학연 협력 프로젝트- 두산 산업차량

목 적

➢ 전동 지게차의 CAN(Controller Area Network) Data 분석

과제 선정배경

➢ 차량 개발부터 사후 관리까지 활용 가능한 CAN Data를 쉽게 분석할 수 있는 알고리즘을 개발하고자 함

수행 방안

✓ CAN Data 기반의 운행모드 분석 알고리즘 개발
✓ 분석 및 예측 결과 시각화

수행 순서

	Phase 1 : **Train / Validation**	Phase 2 : **Test** Step-1	Phase 2 : **Test** Step-2
Data	• 작업 모드 Data: CAN & TMS Data - DVT, 연비 등 주요 DIVK 평가 모드 • Data Feature : TBA	• TMS Data (대표 기종) – '22.07 ~ 12	• TMS 및 Warranty Data
목표	• CUP 분석 알고리즘 개발 - 작업모드 Classification - Regression (특정 항목)	• Phase 1 알고리즘 이용 TMS Data 분석 - 주요 기종 확대 적용 (알고리즘 고도화) - 업계, 업체, 차량WDA/CUP 분석 결과 Matching을 통한 Field duty 분석, System 등 세부단위 분석	• 최종 목표 도출
Target	• 대표 데이터 기준 • F1-score 0.9 이상 • MAE : Weight 100kg, 속도 3m/s	• 가혹도 별 차량/업체 Clustering	• CUP Platform (Dashboard) - 각 사용목적에 맞는 Guidance 제공 • Digital twin (시각화) - Table/Image/Animation(Optional)

산학연 협력 프로젝트- 두산 산업차량

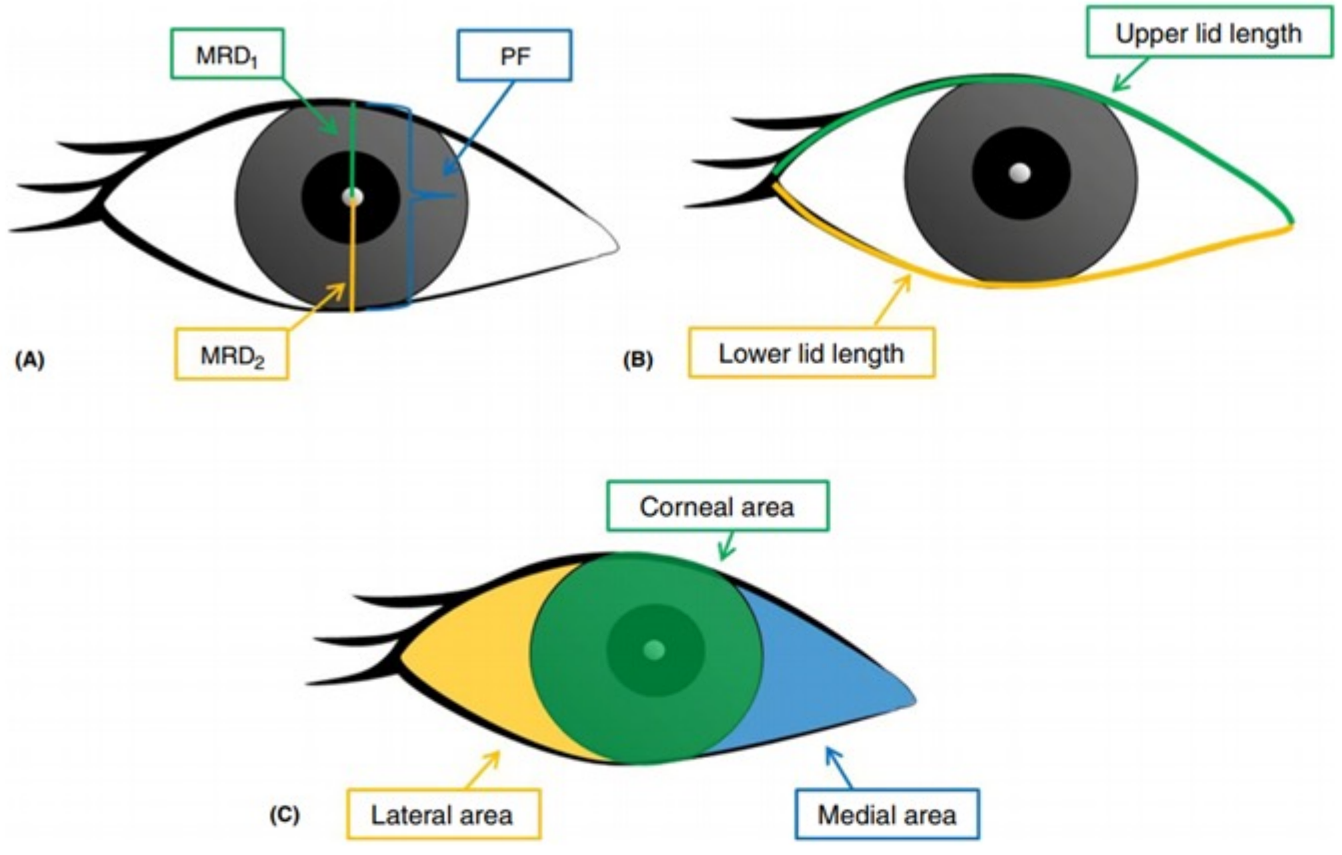
- CAN Dataset

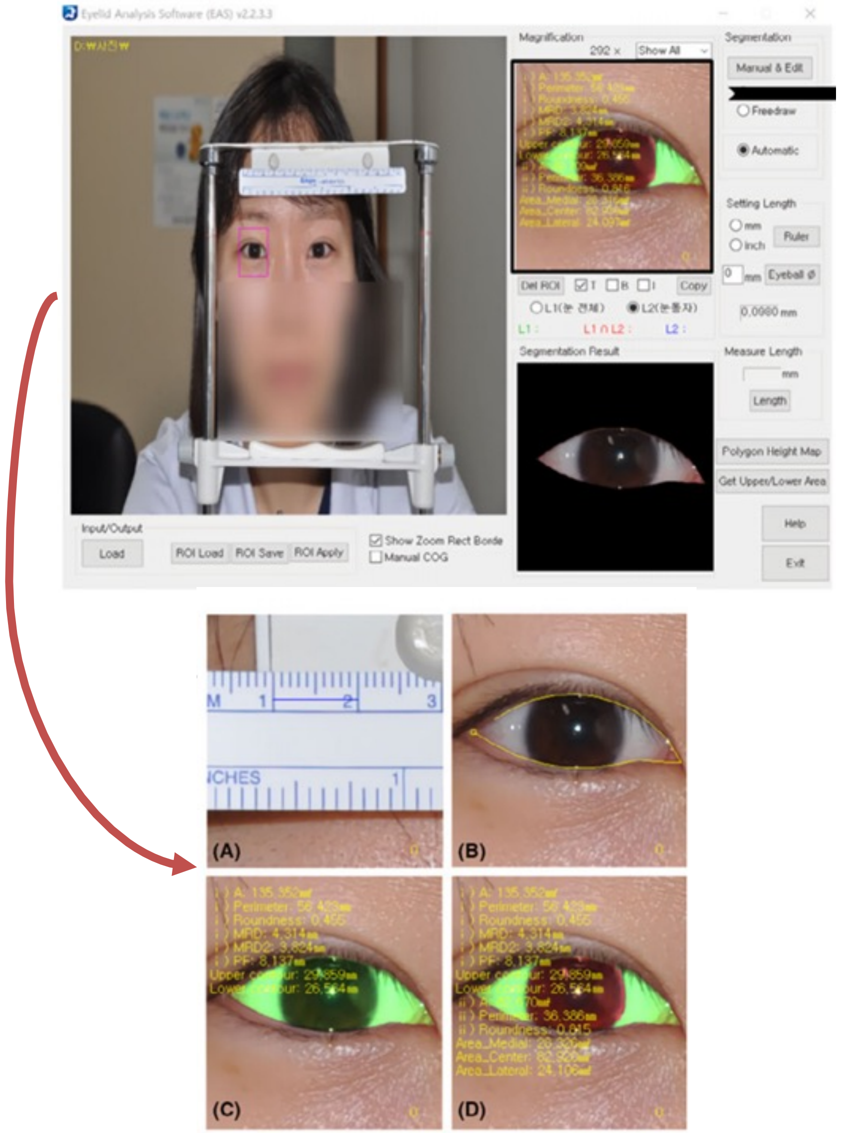
Time	주행모드			운행환경					부하	속도	이동거리	변속(엔진)		단수 (엔진)				모드 (전동)			선회 각도	선회각도				선회방향			인칭 (엔진)	가속도	가속		감속		작업기 모드				작업기 속도		작업기 높이	
	전진	후진	선회	평지	등판	내리막	요철	관성주행				자동	수동	1	2	3	4	H	S	E		30	60	90	90이상	좌회전	우회전	전후진			급가속	가속	급감속	감속	상승	하강	전경	후경	상승속도	하강속도	상승높이	하강높이
	O/X	O/X	O/X	O/X	O/X	O/X	O/X	O/X	kg	km/h	m	O/X	O/X	O/X	O/X	O/X	O/X	O/X	O/X	O/X	°	O/X	O/X	O/X	O/X	O/X	O/X	O/X	O/X	km/h2	O/X	O/X	O/X	O/X	O/X	O/X	O/X	O/X	mm/s	mm/s	mm	mm
1																																										
2																																										
3																																										
4																																										
5																																										
6																																										
7																																										
8																																										
9																																										
10																																										
11																																										
12																																										
13																																										
14																																										
15																																										
16																																										
17																																										
18																																										
19																																										
20																																										
21																																										
22																																										
23																																										
24																																										
25																																										
26																																										
27																																										
28																																										
29																																										
30																																										
31																																										
32																																										
33																																										
34																																										
35																																										
36																																										
37																																										
38																																										
39																																										

[B20T-7 Plus]

갑상선 연구 – 눈꺼풀 길이 추론

- 연구 배경 / 필요성

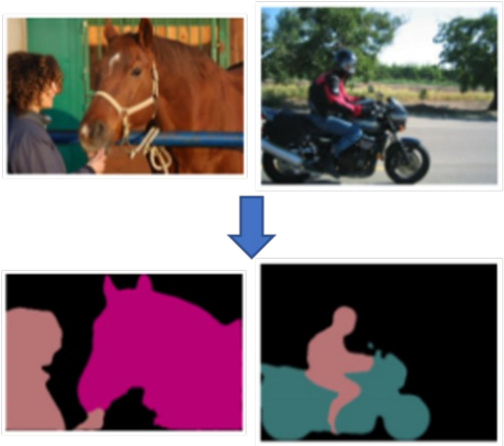




Chun, Yeoun Sook, et al. "Topographic analysis of eyelid position using digital image processing software." Acta ophthalmologica 95.7 (2017)

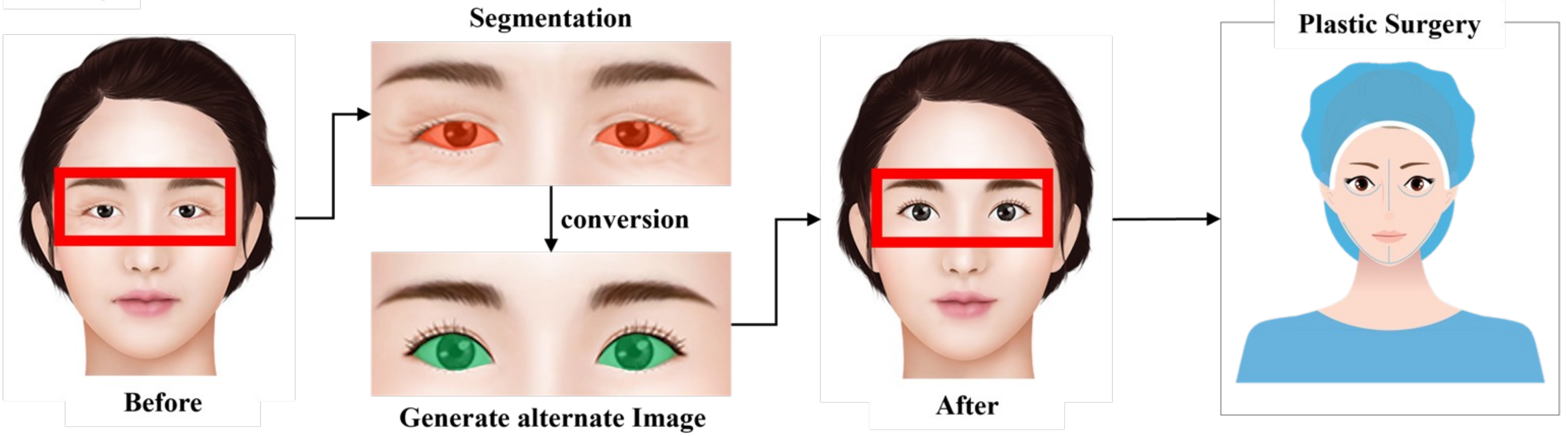
Research

Tasks

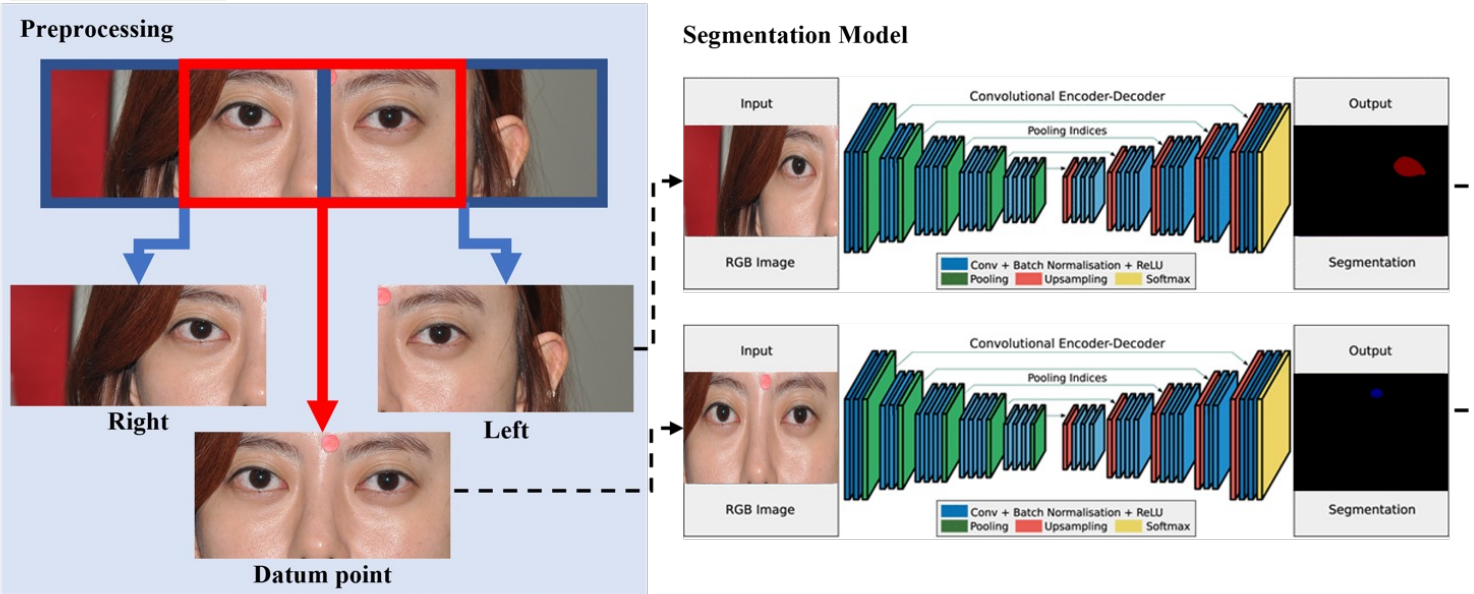


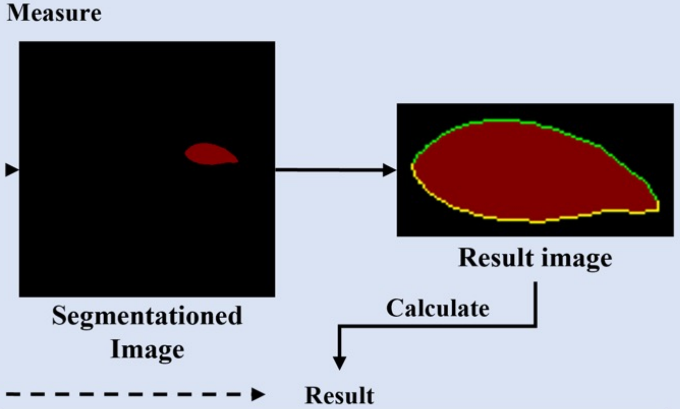
Domains

Beauty



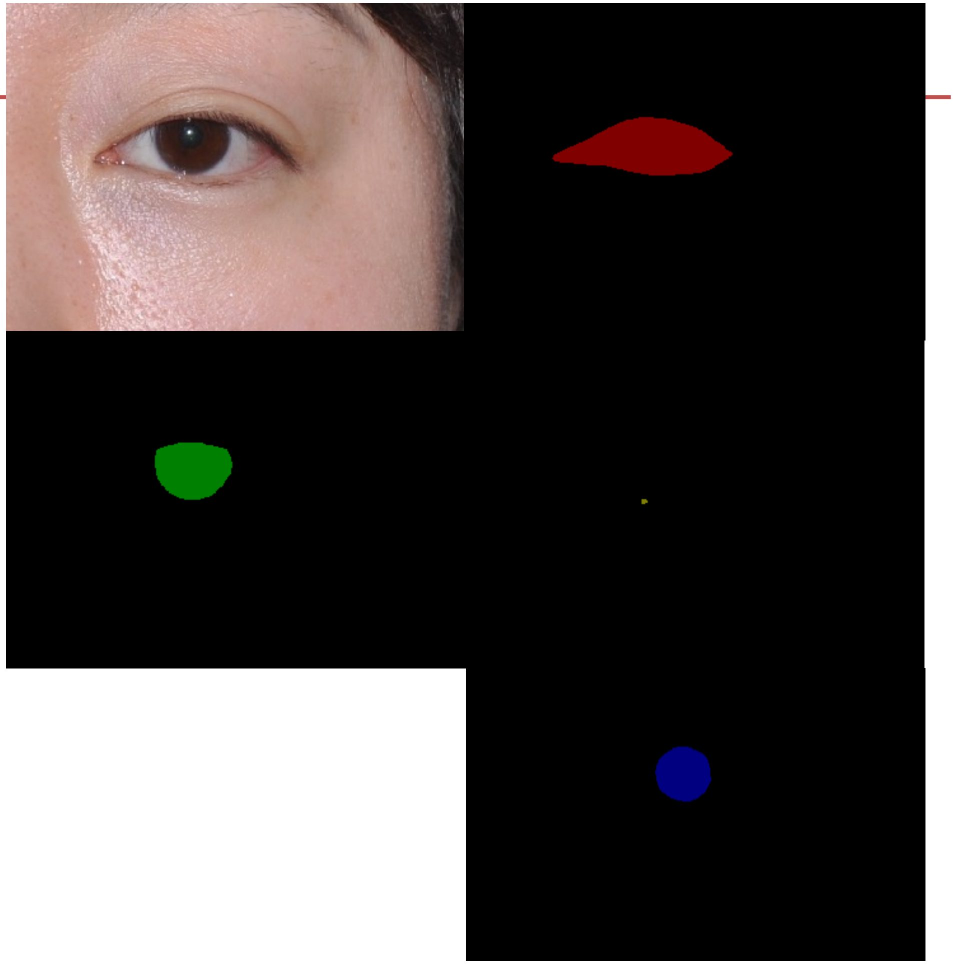
Frameworks





Mesure eyelid length				
Position	Upper_lid	Lower_lid	MRD1	MRD2
mm	39.71	35.21	3.44	6.35

Experiment Settings



- Datasets (patient photograph)
  - Number of Samples: 344 of 708 photograph
  - Number of Classes: 1 (sclera, iris, reflex, sticker)
  - Input: (512, 512), RGB images
  - Output: (512,512,1)
- Augmentation
  - left-right Flip + upside down(only sticker)
- Cross-Validation for Test Performance
  - 10 iterations
  - Data Split Ratio: (Train : Validation : Test) = (0.8 : 0.1 : 0.1)
- Hyper Parameter Setting:

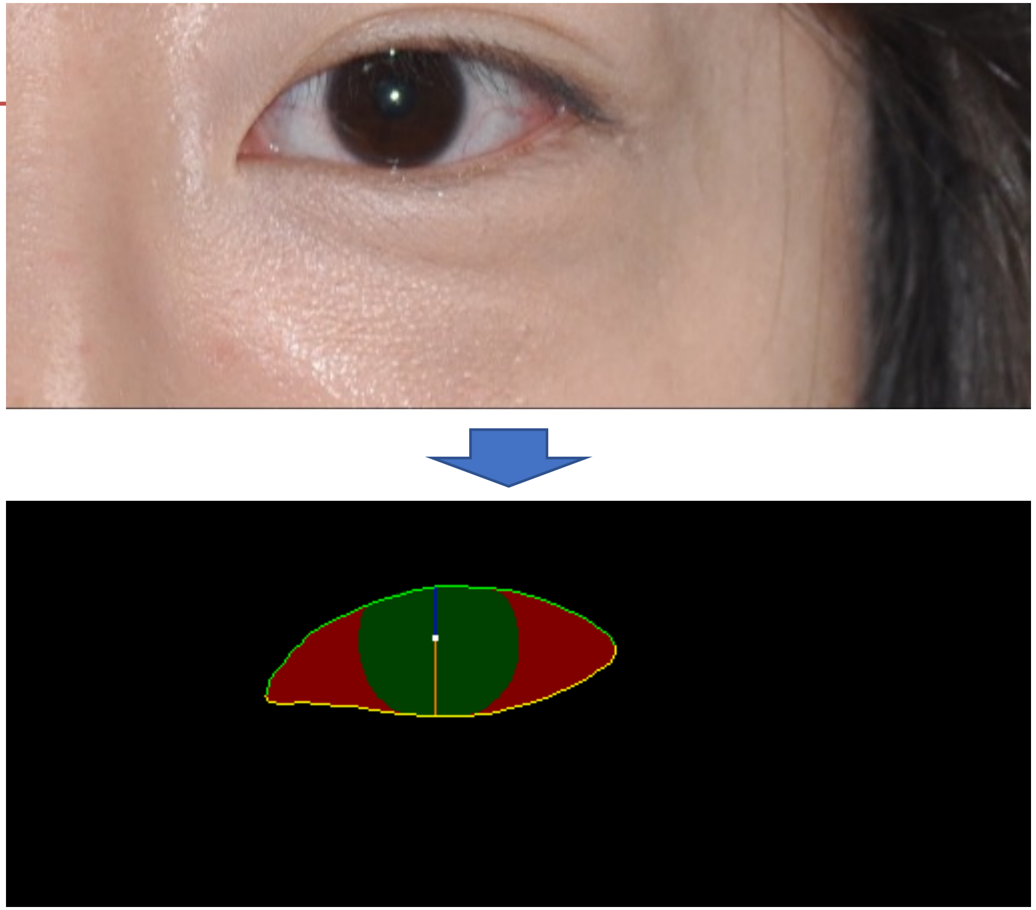
Model	Batch Size	Epochs	Loss Function	Optimizer	LR	Weight Decay
Deeplab V3+ (Chen et al., 2018)	8	100	Cross Entropy Loss	Adam	1e-4	1e-4

Chen, Liang-Chieh, et al. "Encoder-decoder with atrous separable convolution for semantic image segmentation." *Proceedings of the European conference on computer vision (ECCV)*. 2018.

Experiment Results

- Model result

Dataset	Model	mIoU		
Sclera	Deeplab V3+	**0.943**	**±**	**0.009**
Iris		**0.951**	**±**	**0.009**
Reflex		**0.812**	**±**	**0.004**
Sticker		**0.984**	**±**	**0.004**

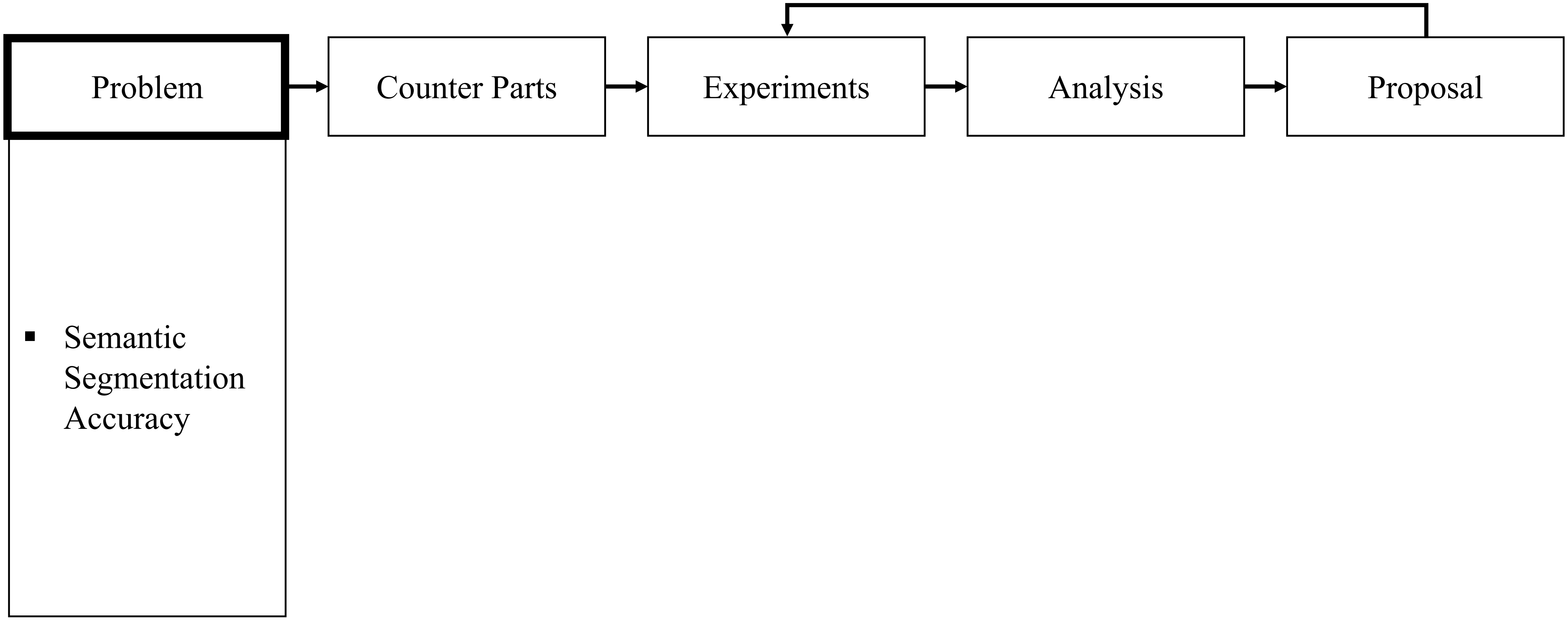


- 실측 결과

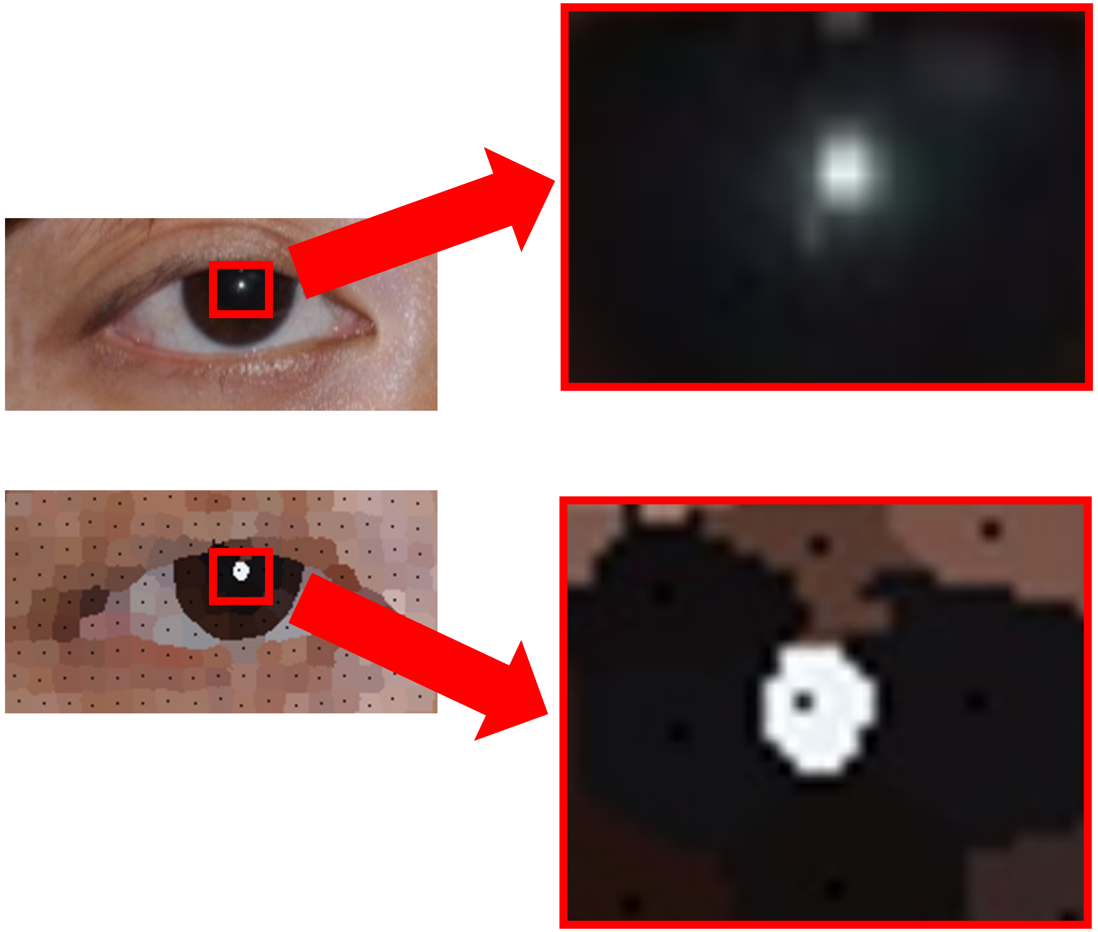
	R_MRD1	R_MRD2	R_upper_lid	R_lower_lid	L_MRD1	L_MRD2	L_upper_lid	L_lower_lid
MAE	0.38	0.46	5.39	3.83	0.43	0.49	5.52	3.64

단위(mm)

Overall Progress



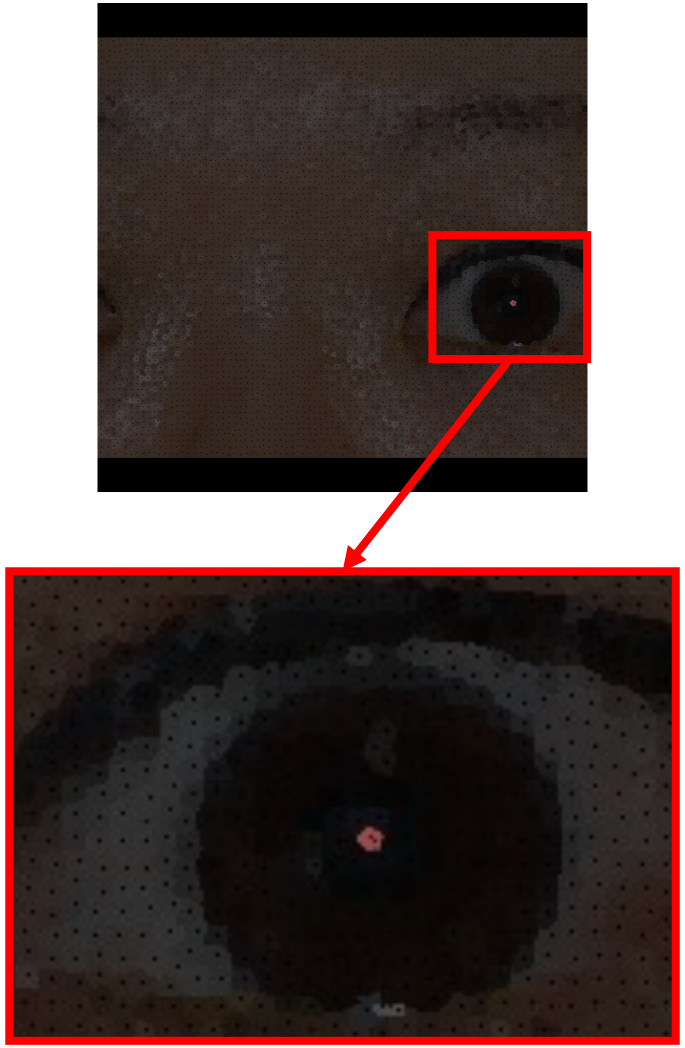
Superpixel



Achanta, Radhakrishna, et al. "SLIC superpixels compared to state-of-the-art superpixel methods." *IEEE transactions on pattern analysis and machine intelligence* 34.11 (2012): 2274-2282.

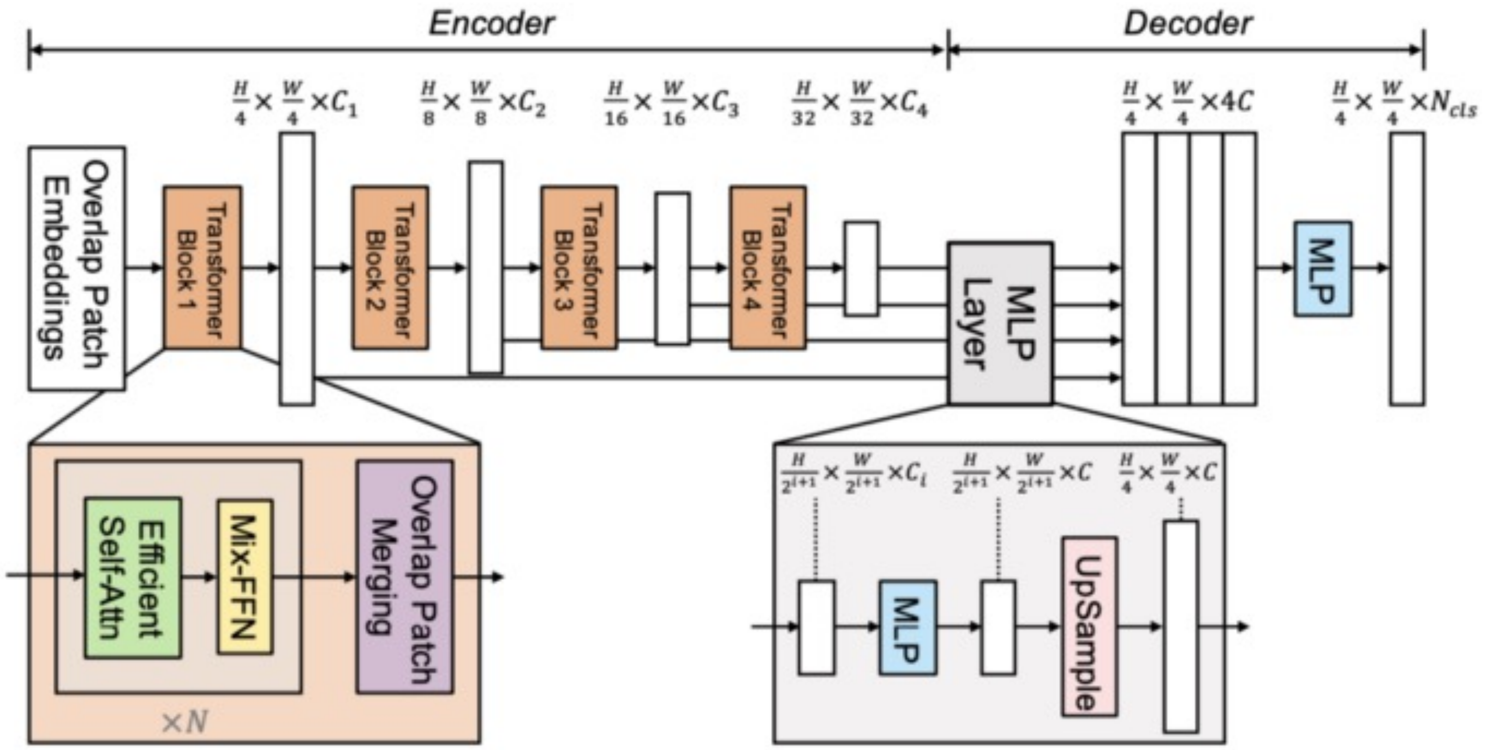
Experiment Results

Dataset	Model	mIoU		
Reflex (superpixel 미적용)	Deeplab V3+	**0.812**	**±**	**0.009**
Reflex (superpixel 적용)		**0.830**	**±**	**0.003**
Reflex (crop, resizing)		**0.858**	**±**	**0.004**



Next study

- Segformer



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

References

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

[2] Chen, Liang-Chieh, et al. "Encoder-decoder with atrous separable convolution for semantic image segmentation." *Proceedings of the European conference on computer vision (ECCV)*. 2018.

[3] Achanta, Radhakrishna, et al. "SLIC superpixels compared to state-of-the-art superpixel methods." *IEEE transactions on pattern analysis and machine intelligence* 34.11 (2012).

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

기계학습자동화 연구실
주간 정기 세미나

진상형

2022. 08. 30.

Individual Studies:

Sanghyeong Jin

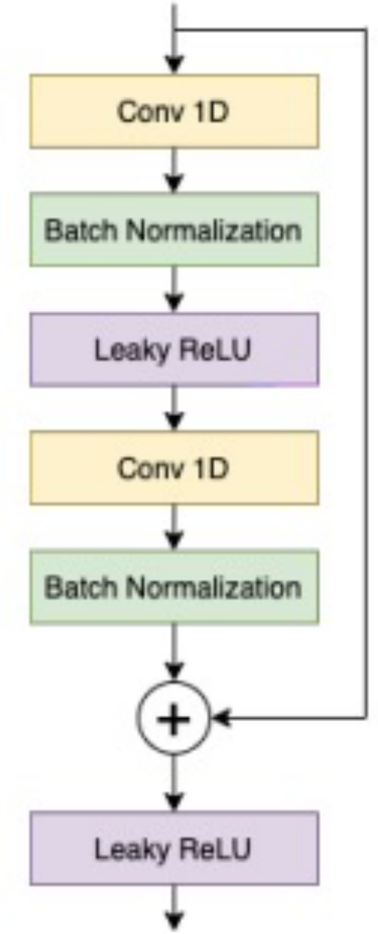
Introduction

Session ID	Track ID	Session Position	…	User Reason Start	User Reason End
1	A	1	…	trackdone	fwdbtn
1	B	2	…	fwdbtn	trackdone
1	C	3	…	trackdone	backbtn

Session Data

Track ID	Acoustincess	…	Release Year	Energy	Tempo
A	0.121	…	1945	0.8123	166.12
B	0.849	…	2002	0.1565	95.45
C	0.461	…	1987	0.3589	100.79

Track Data



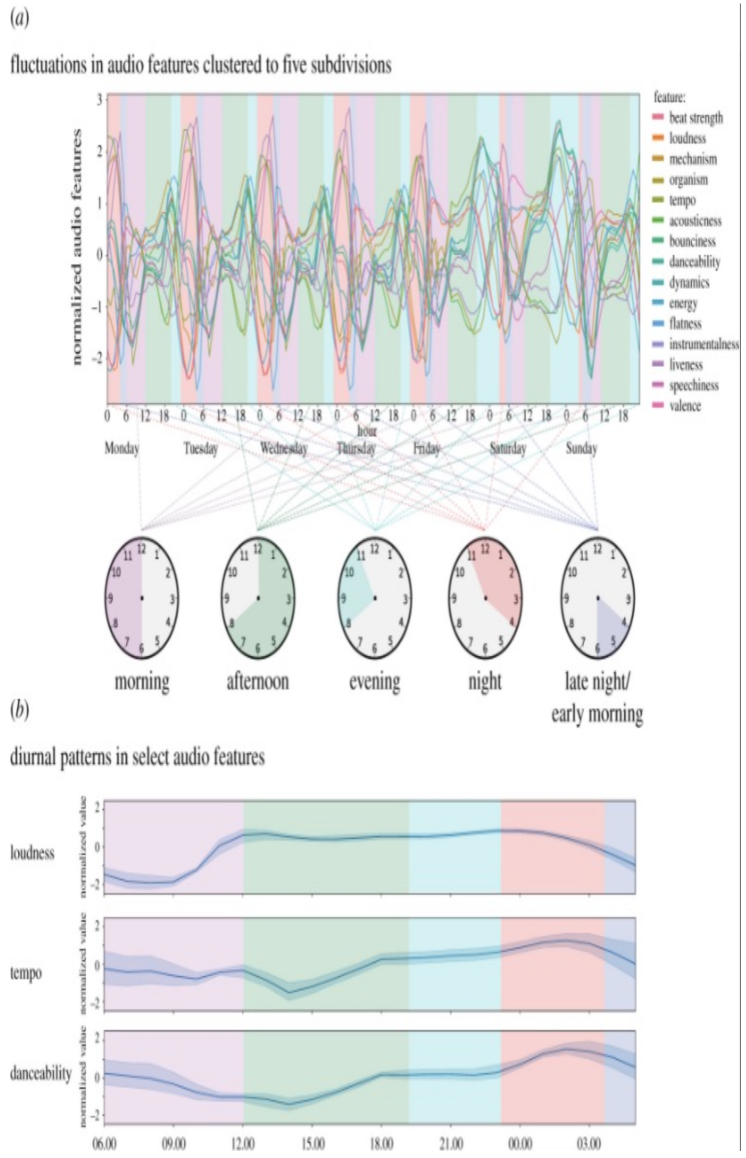
Skip Prediction Model

Predict

Music Streaming Session

Skip / Not-Skip

Introduction



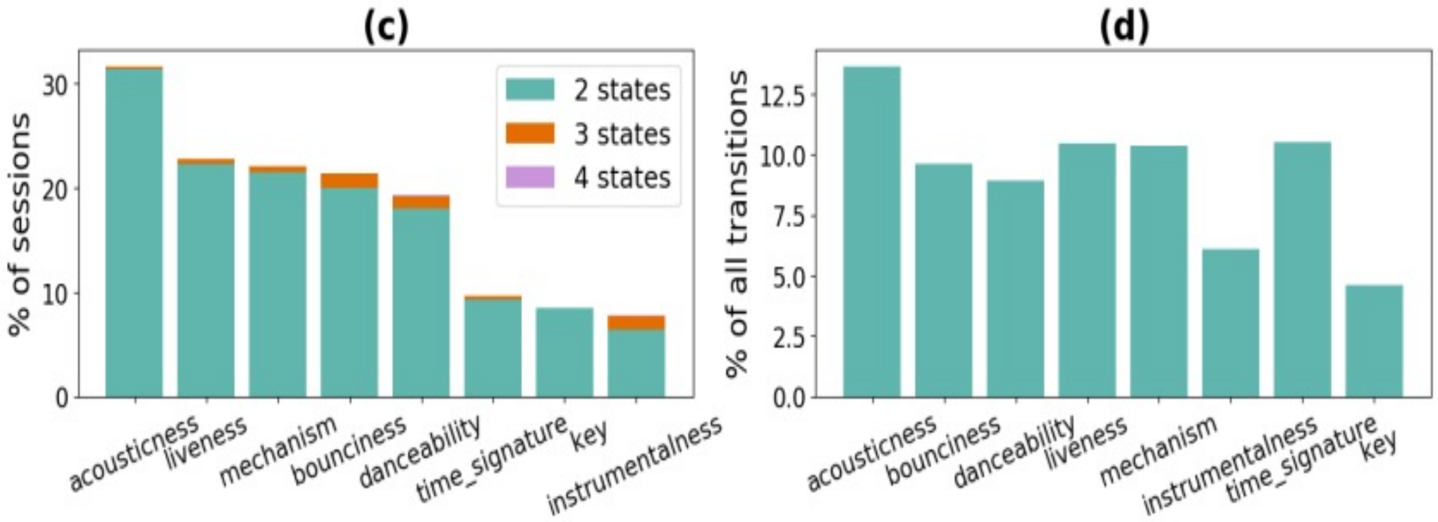
Skip과 관련한 주요 특징

- 오디오 특징의 변화는 일주일 동안 항상 일관된 흐름을 보임
- 하루 동안의 음악적 선호도는 주기적으로, 예측 가능하게 변화함 [3]

- 어떤 곡에 대한 스킵 비율은 일반적으로 U자형 패턴을 가짐 [4]
(곡의 초반과 마지막에서 스킵 비율이 높음)

1. 높은 초반부 스킵 비율, 이후 급격한 지수적 감소
(곡의 스킵 여부를 대부분 초반 몇 초 안에 결정한다는 것)
2. 코러스 시작, 악기 변화, 강도 변화 같은 곡의 특징적인 지점에서 튀는 경향
3. 다음 곡에 대한 기대로 인한 곡이 거의 끝나는 지점에서 스킵 비율 증가

Introduction



Skip 상태 전이 [2]
: Hidden Markov Model 이용하여 MSSD 데이터[1]의 오디오 특징 상태 전이 과정을 분석

- 특히 오디오 특징의 변동은 각 스트리밍 세션에서 상당히 일반적으로 나타나고 이는 사용자의 Skip여부와 큰 관련성을 가짐

- Sequential한 접근을 권장

- Skip 상태 전이 대한 이해는 track sequencing에서 사용자 만족도를 높일 수 있음

Related Work

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

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

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

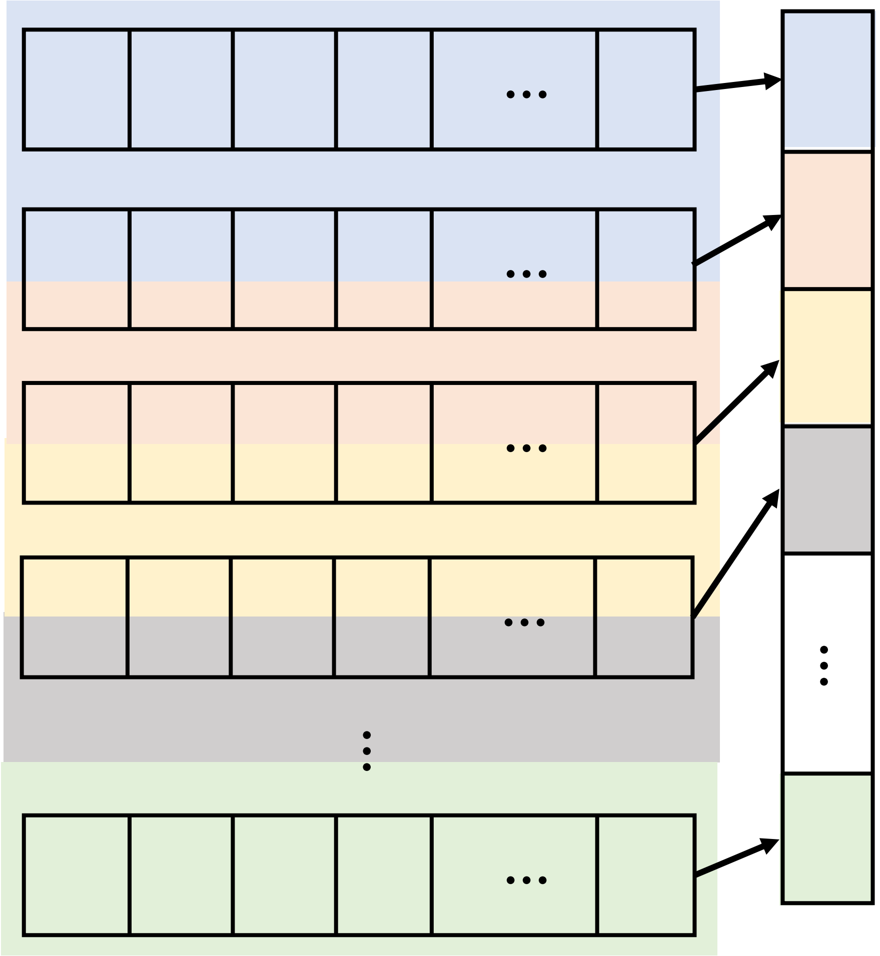
MAA (Mean Average Accuracy)

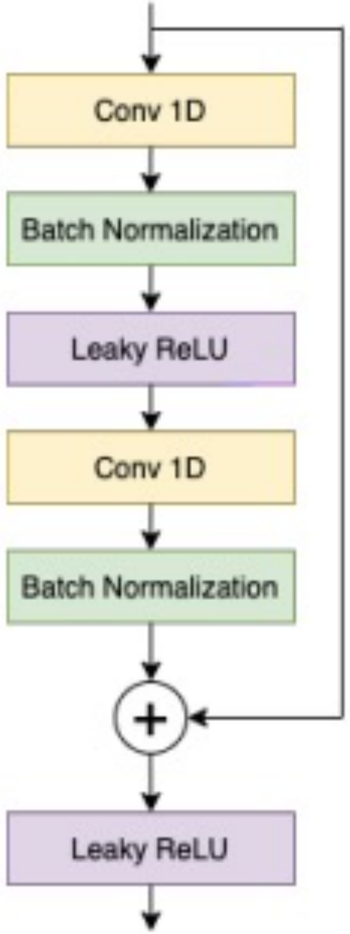
- Skip 문제와 관련하여 제안된 평가 방법
- Skip 순서에 있어 첫번째 예측의 정확도가 가장 중요 (First Prediction Accuracy - FPA)
- 후반부 순서의 Skip 결과는 예측 정확도에 대해 낮은 영향력

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

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

Related Work





1D Convolution [5]

- 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 [13]

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

Related Work

Residual Connection / Highway Network [9]

: 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

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

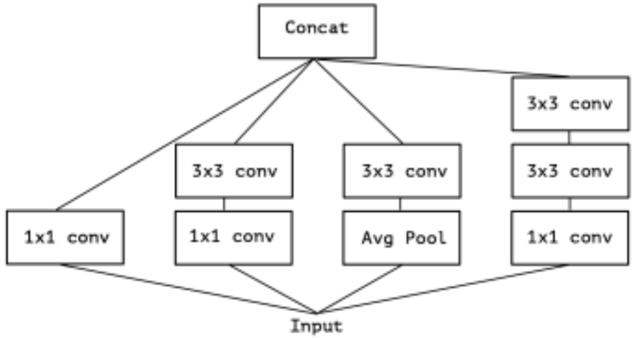
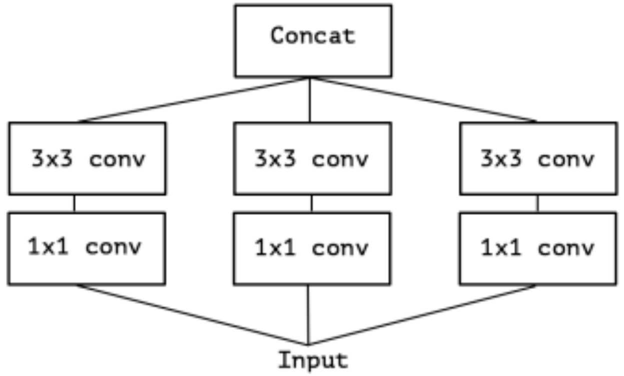
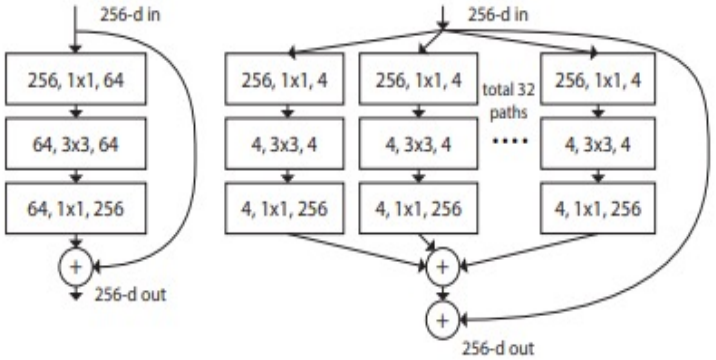


Figure 2. A simplified Inception module.





Multi-Branch Convolution

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

- Inception[6] 의 경우 Cross-channel correlation과 Spatial Correlation을 분리
- Xception[7] 은 Inception을 확장하여 Cross-channel / Spatial Correlation을 완전히 독립적으로 분리
- ResNeXt[8] 의 경우 cardinality라는 개념을 사용하여 동일 크기의 그룹으로 분리하는 Grouped Convolution 사용

Related Work

Dilated Casual Convolution
: Wavenet[10] 에서 제안, 다음 시점의 시퀀스를 생성하기 위해 Convolution layer를 이용

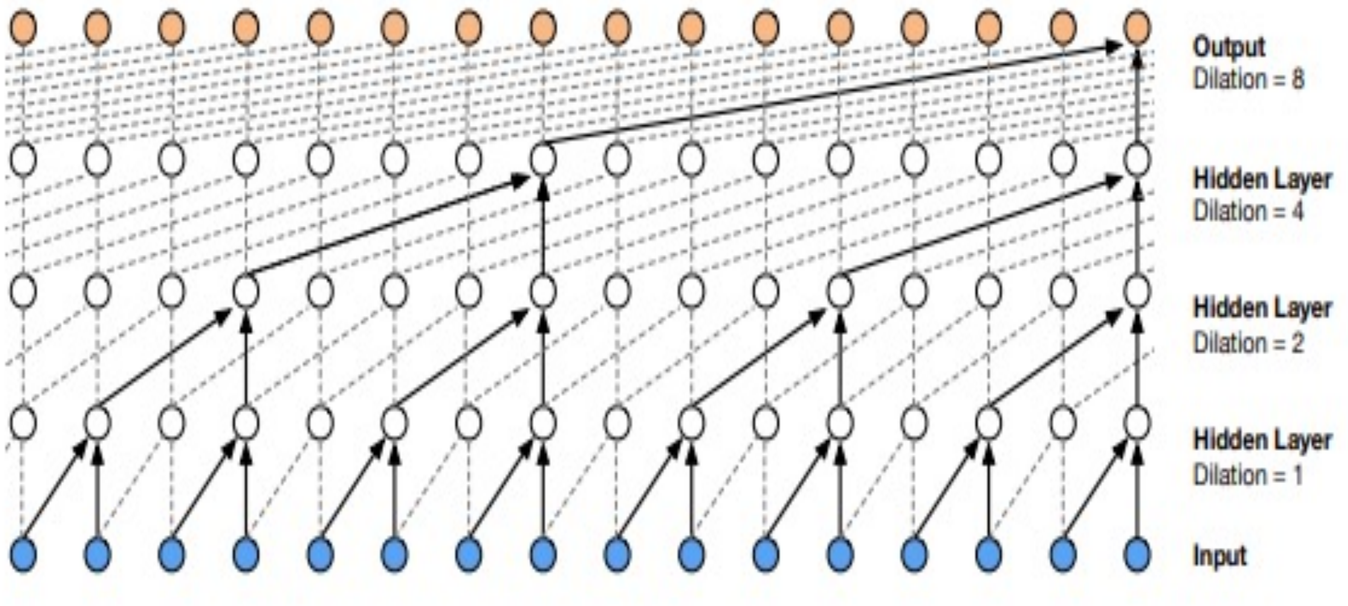
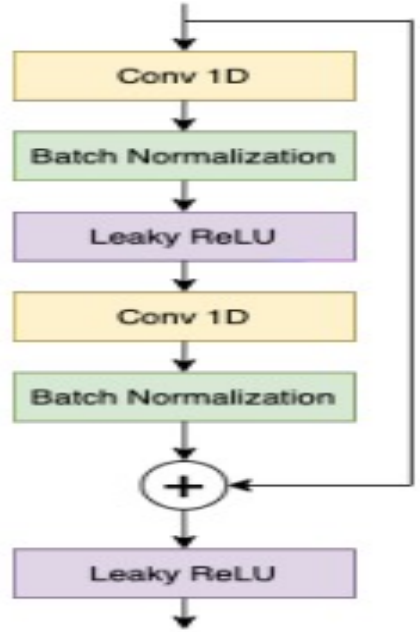


Figure 3: Visualization of a stack of *dilated* causal convolutional layers.

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

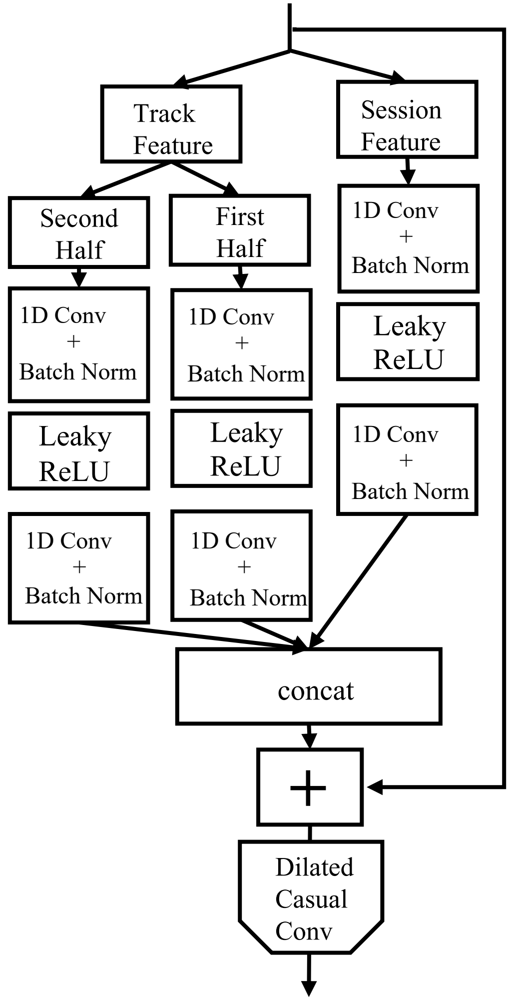
2. Dilated Convolution
: 필터가 입력을 건너뛰어 원래보다 더 큰 영역에 적용
- Zero padding하여 더 큰 필터를 적용하는 것과 동일하나 Dilation을 사용하는 것이 더 효율적임

Proposed Method



Proposed

Multi-branch Convolution
Dilated Casual Convolution
1D ResNet 변형, 적용



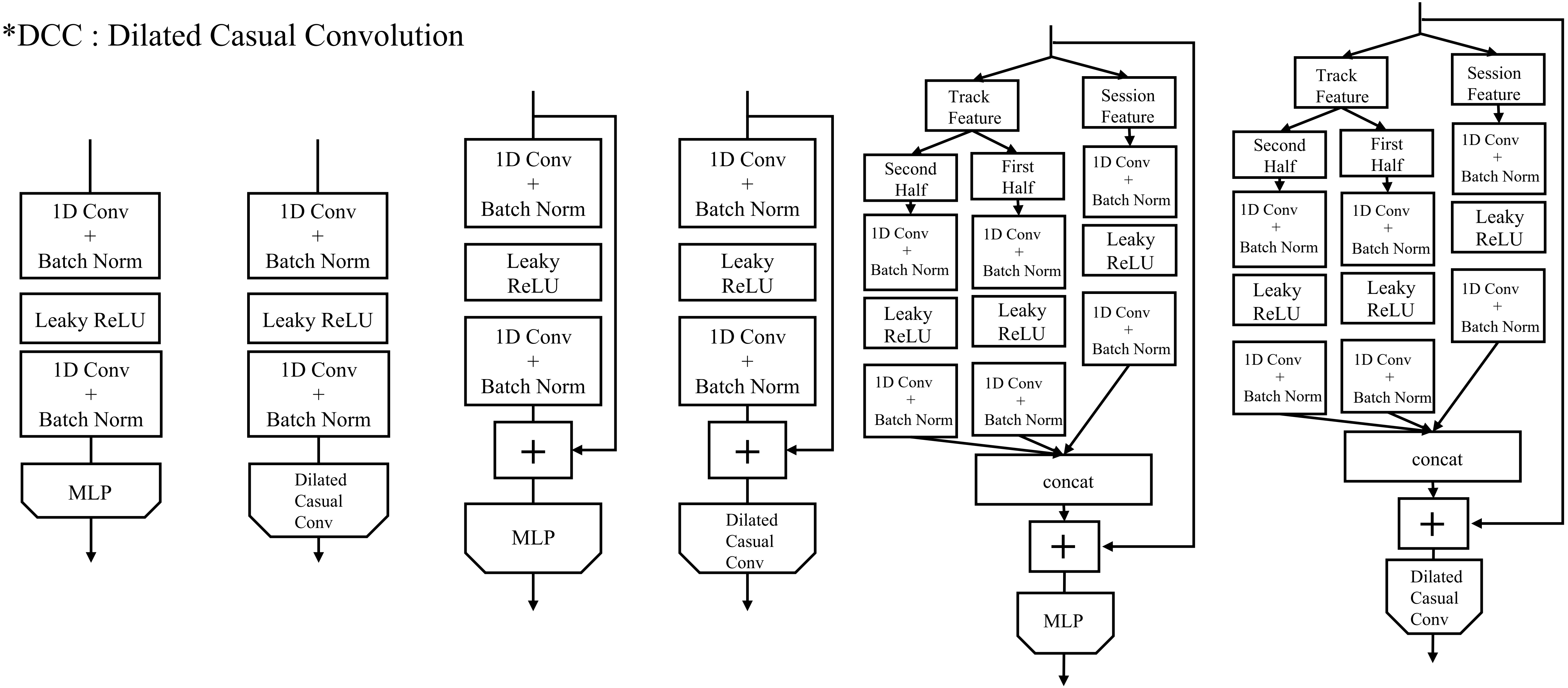
Proposed Method

분기 구조를 통해 convolution을 분리하여 각각 학습하고 하나로 Aggregation하여 각각의 분기에 대해 서로 비슷한 Correlation을 가진 특징들끼리 학습되도록 함

-> Inception, Xception, ResNeXt 등의 앞선 이미지 모델에서도 꾸준히 사용됨
-> Skip 문제에 사용하는 MSSD 데이터는 Tabular 형태로 이미지 데이터와 달리 명시적인 특징 분리가 가능

Proposed Method - Comparison

*DCC : Dilated Casual Convolution



1. Residual 없음
+ MLP

2. Residual 없음
+ DCC

**3. 1D ResNet
+ MLP
(Baseline)**

4. 1D ResNet
+ DCC

5. Multi-branch
1D ResNet + MLP

**6. Multi-branch
1D ResNet + DCC
(Propsed)**

Experiment Settings

- Datasets
  - Music Streaming Sessions Dataset (MSSD) [1]
  - 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
Multi-RNN	16	10	Binary Cross Entropy	Adam	1e-3
RNN + Dense	16	10	Binary Cross Entropy	Adam	1e-4
Encoder-Decoder	16	10	Binary Cross Entropy	RMSprop	1e-4
Interaction Layer + RNN	16	10	Binary Cross Entropy	Adam	1e-4
1D conv. + Highway	16	10	Binary Cross Entropy	Adam	1e-4
1D ResNet	16	10	Binary Cross Entropy	Adam	1e-4
Proposed	16	10	Binary Cross Entropy	Adam	1e-4

Experimental Result - 1

	Model	MAA	MAA P-value	FPA (First Prediction Accuracy)	FPA P-value
1	Residual 없음 + MLP	0.415 ± 0.003	< 0.0001	0.509 ± 0.004	< 0.0001
2	Residual 없음 + DCC	0.450 ± 0.003	< 0.0001	0.577 ± 0.005	< 0.0001
3	**1D ResNet [13] + MLP (Baseline)**	**0.634 ± 0.002**	**< 0.0001**	**0.784 ± 0.004**	**0.4523**
4	1D ResNet + DCC	0.646 ± 0.003	0.0003	0.763 ± 0.004	0.7116
5	Multi-branch 1D ResNet + MLP	0.673 ± 0.002	0.6600	0.802 ± 0.004	0.8111
6	**Multi-branch 1D ResNet + DCC (Proposed)**	**0.688 ± 0.002**	**-**	**0.796 ± 0.004**	**-**

*DCC : Dilated Casual Convolution

Experimental Result - 2

Type	Model	MAA	MAA P-value	FPA (First Prediction Accuracy)	FPA P-value
Non-RNN	Boosting Tree [11]	0.514 ± 0.001	< 0.0001	0.611 ± 0.001	< 0.0001
	1D Conv. + Highway [12]	0.577 ± 0.001	< 0.0001	0.673 ± 0.002	< 0.0001
	1D ResNet [13]	**0.634 ± 0.002**	**< 0.0001**	**0.784 ± 0.004**	**0.4523**
	Proposed	**0.688 ± 0.002**	**-**	**0.796 ± 0.004**	**-**
RNN	Multi-RNN [14]	0.606 ± 0.003	< 0.0001	0.710 ± 0.004	< 0.0001
	RNN + Dense [15]	0.568 ± 0.002	< 0.0001	0.662 ± 0.001	< 0.0001
	RNN+Dense (weighted) [15]	0.583 ± 0.002	< 0.0001	0.681 ± 0.002	< 0.0001
	Encoder-Decoder [16]	0.530 ± 0.004	< 0.0001	0.642 ± 0.002	< 0.0001
	Interaction Layer + RNN [17]	0.595 ± 0.002	< 0.0001	0.697 ± 0.003	< 0.0001

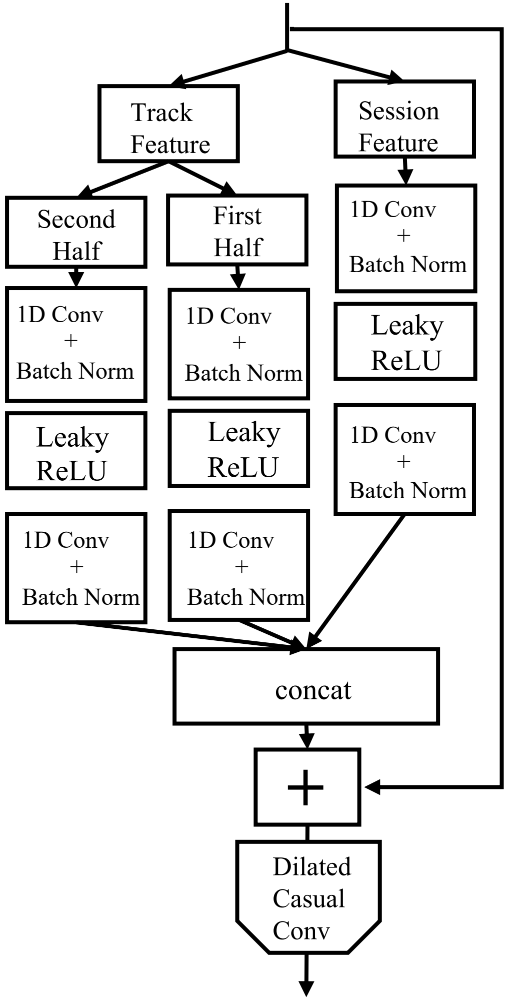
Conclusion

- 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 비교에서는 Proposed 방식이 훨씬 더 좋은 성능을 보임

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



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

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

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

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

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

[11] A. Ferraro, D. Bogdanov, X. Serra. “Skip prediction using boosting trees based on acoustic features of tracks in sessions”, 12th ACM International Conference on Web Search and Data Mining. 2019.

[12] Sungkyun Chang et al. “Sequential Skip Prediction with Few-shot in Streamed Music Contents”, 12th ACM International Conference on Web Search and Data Mining. 2019.

[13] Safaa Allamy, Alessandro Lameiras Koerich. “1D CNN Architectures for Music Genre Classification”, 2021 IEEE Symposium Series on Computational Intelligence, 2021.

[14] 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.

[15] 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.

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

[17] Lin Zhu, Yihong Chen. “Session-based Sequential Skip Prediction via Recurrent Neural Networks”, 12th ACM International Conference on Web Search and Data Mining. 2019.