

ICASSP '25 notes and interesting posters

Benedikt Kantz

April 22, 2025

1 Montag Vormittag

1.1 Tutorial: Generative AI and Model Optimization

Problem: (compute) cost, current foundation models not sustainable Solutions:

1.1.1 Sparsity

- scalability, less overfitting, interpretability, adaptive ways to introduce sparsity
- post training: optimal brain damage (OBD)/ optimal brain surgery (OBS)
 - dropout by contribution to error, scale by Hessian \mathcal{H} contribution
- training:
 - L1-loss: Convex optim.; no free lunch: initial model very large!, more eqs.
 - exhaustive: very expensive
 - greedy/evolutionary solutions: StOMP, GOMP based on L0-norm, but very effective
- pre-training
 - SET
 - randomly initial init → evolutionary
- architectural: grow and shrink networks...

Problem: doesn't really work with LMs (empirical study), but well for other networks (esp. low-weight dropout)

1.1.2 Compression

- filter: storage compression
- low rank factorization (\neq LoRA), during train time not fine-tuning
- knowledge distillation

2 Dienstag Nachmittag

2.1 Talk: Underwater Communications

- Problem: very slow comm underwater, ≈ 10 kHz range
- Towards moving target, Doppler correction using active SP correction, very manual work

Comment: interesting manual process, tedious work to sample

3 Mittwoch Nachmittag

3.1 Talk: AI+SP

Comment: just some basics on diffusion/transfomers, a little bit of SP in NNs

4 Donnerstag Vormittag

4.1 Talk: Multiomics

- Genomics: DNA understanding
- Transcriptomics: DNA- $\&$ RNA understanding
- Proteomics: RNA- $\&$ Protein structures
- Knowledge graphs: how do these systems influnce each other
- Flow:
 - identify DNA mutation that triggers illness
 - find possible RNA mechanism
 - find good fitting small ring structure
 - check for side effects in knowledge graph! (certain protein effects unwanted)
 - then test → animal tests, reduce through ML!
- Graph diffusion for drug discovery: noise schedule for diffusion essential, i.e. cosine-square schedule
 - diffuse graphs from atoms & edges as adjacency matrix
 - what is noise: discrete noise: each atom is discrete state \Rightarrow graph structure undergoes state transition change
 - naive: uniform structure, not really chemically sensible - conditional probabilities \Rightarrow not uniform but marginal distribution of molecules in training (just logical!), same for edge (with deletion!)
 - one step further: consider carbon rings, restriction based on maximum bonds of atom (freie Radikale)
 - SMILE-file, QED: Quantitative Estimate of Drug likeness (from RDKit)
 - Existing methods: Time-consuming, progress slow, very few good molecules
 - Their work: jointly perturb rings+nodes
 - other approaches: motives as super-node with rings, difficulty: ring attachments - only $\approx 1\%$ improvement!
 - novelty however high, one molecule of them even patented!
- Knowledge graphs:
 - GNN link prediction
 - none of the existing benchmarks include features!
 - maybe talk to author!

Comment: focused on drug discovery using diffusion, not much on multiomics...

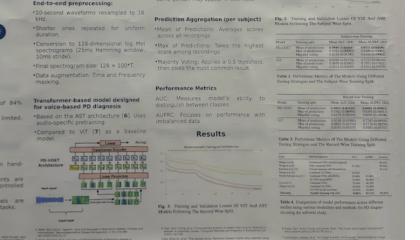
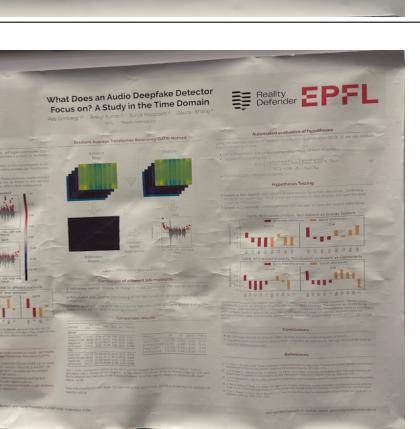
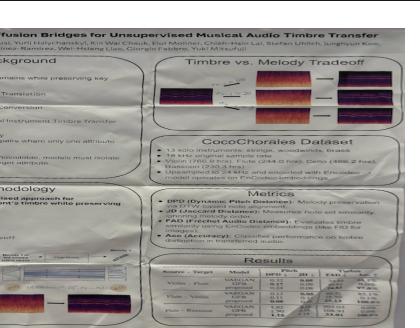
5 Lectures/Orals

Table 1:

Lecture	URL	Notes
Diversity-Seeking Techniques for Red-Teaming LLMs	https://ieeexplore.ieee.org/document/10890844/	Add RL-Loss to train similar to GAN by backpropagating if the model returns very similar output (i.e. discriminator)-; Very fragile learning; Limited further studies
FDR Control for Complex-Valued Data	https://ieeexplore.ieee.org/document/10889705	similar to LASSO; sparsifying system under certain guarantees
SpectralCam: High-Resolution Low-Cost Spectral Imaging Using DSLR Cameras	https://ieeexplore.ieee.org/document/10887725	Interesting concept of applying photo filter to DSLR sensor, Bayesian pattern restoration "learned" using diffusion & attenuation mtx
Fusing Multimodality of Large Language Models and Satellite Imagery	https://ieeexplore.ieee.org/document/10889624	Could be interesting in combination with HEREDITARY geospatial data once we have access
Controllable Forgetting Mechanism for Few-Shot Class-Incremental Learning	https://arxiv.org/pdf/2501.15998	Using embedding space to classify, add new classifier based on distance, seems rather hyperparameter-sensitive

6 Posters

Table 2:

Poster	Information
 <p>PD-VOST: PARKINSON'S DISEASE VOICE SPECTROGRAM TRANSFORMER Ilias Tougui, Mehdi Zakroum, Ouassim Karakchou, Mounir Ghogho International University of Rabat, Morocco</p> <p>Background •Parkinson's disease (PD) affects over 10 million people worldwide [1]. •More than 50% of Parkinson's patients experience speech changes, such as [2]:</p> <ul style="list-style-type: none"> -Hoarseness -Slow speech rate -Shaky voice -Lack of inflection -Reduced volume -Changes in pitch -Loss of vocal control <p>Motivation •Doctors achieve an accuracy of 84% after 10 years of training [3]. •Access to specialists is limited, particularly in rural areas.</p> <p>Methodology •Data selection from mPawer Study: -16130 patients, 24110 audio samples. -10000 samples were used for 10-fold cross-validation. -Remaining recordings of speech samples for 10-fold were collected via mPawer [5].</p> <p>End-to-end preprocessing: •10-second audio samples resampled to 16 kHz. •Short-time windows for uniform time-frequency representation. •Conversion to 128-dimensional log Mel-Frequency Cepstral Coefficients (MFCCs), 10ms stride. •Final dimension is $128 \times 100T$. •Data augmentation: Time and frequency masking.</p> <p>Transformer-based model design: •Proposed architecture for voice-based PD diagnosis [6]: -Based on the AST architecture [6] used for COVID-19 detection.</p> <p>Performance Metrics •AUROC: Measures model's ability to distinguish between healthy and PD patients. •AUPRC: Focuses on performance with imbalanced datasets. •Mean of Predictions: Averages scores across all subjects. •Majority Voting: Assigns a class based on the prediction of the most frequent class.</p> <p>Results •Fig. 1: Training and Validation Loss of VIT and AST Model Following the Epochs.</p> <p>Table 1: Evaluation Metrics of the Model Using Different Training Strategies.</p> <p>Table 2: Evaluation Metrics of the Model Using Different Window Lengths.</p> <p>Table 3: Comparison of Model Performance.</p> <p>Fig. 2: Training and Validation Loss of VIT and AST Model Following the Epochs.</p>	<p>PD-VOST: Parkinson's Disease Voice Spectrogram Transformer <i>Ilias Tougui, Mehdi Zakroum, Ouassim Karakchou, Mounir Ghogho</i> https://ieeexplore.ieee.org/abstract/document/10889820/</p>
 <p>What Does an Audio Deepfake Detector Focus on? A Study in the Time Domain Surya Raghav, Reality Defender EPFL</p> <p>Motivation •Deepfakes have become a major threat to privacy and security.</p> <p>Methodology •Proposed a deepfake detector that focuses on the time domain.</p> <p>Results •The proposed detector can identify deepfakes with high accuracy.</p>	<p>The EPFL combinational benchmark suite <i>pee/ginbergaepf.ch, facia, surya, gauravia roatty defender al</i></p> <p>https://infoscience.epfl.ch/entities/publication/309aea67{-}b5a1{-}4532{-}8a6f{-}0a141d8f1ab3/full</p>
 <p>Latent Diffusion Bridges for Unsupervised Musical Audio Timbre Transfer Michele Mancusi, Yurii Halychanskyi, Kin Wai Cheuk, Eloi Moliner, Chieh-Hsin Lai, Stefan Uhlich, Junghyun Koo, Michael A. Homa, Michael H. Mandel</p> <p>Background •Timbre transfer: Transferring musical domains while preserving key structure or meaning.</p> <p>Key Challenges •How to handle multiple instruments?</p> <p>Methodology •Proposed an unsupervised approach for timbre transfer.</p> <p>Key Advantages •Unsupervised learning •Fast computation •Timbre vs. Melody Tradeoff</p> <p>Results •Table 1: Results of the CocoChorales Dataset.</p>	<p>Latent Diffusion Bridges for Unsupervised Musical Audio Timbre Transfer <i>Michele Mancusi, Yurii Halychanskyi, Kin Wai Cheuk, Eloi Moliner, Chieh-Hsin Lai, Stefan Uhlich, Junghyun Koo, Michael A. Homa, Michael H. Mandel</i></p> <p>https://ieeexplore.ieee.org/abstract/document/10890708/</p>
 <p>Planetary gear vibration monitoring using synchronous demodulation Sak Yannitsarop, Konstantinos Gyrillas</p> <p>Motivation •We propose a signal processing pipeline for vibration-based monitoring of planetary gearboxes. The pipeline performs demodulation of planetary gears' signals and identifies potential faults.</p> <p>Results •Figure 1: Monitoring indicated areas show a clear trend with respect to the temperature of the planet gears.</p> <p>Conclusions •Synchronous demodulation is chosen to be the main method for planetary gear vibration monitoring due to its robustness against noise and its ability to extract useful information from planetary gear vibrations.</p>	<p>do a crow endus ported ten band mentorine at pear enoses. The pipeline do a crow endus ported ten band mentorine at pear enoses. The pipeline</p>

Continued on next page

Table 2: (Continued)

	<p>EFFICETH REPARAMNE outperforming models like CV-SOG, Motifs, and VCTree.</p>
	<p>MusicLiME: Explainable multimodal music understanding tasks, achieving 57.34% for genre and 48.53% for emotion Classification https://ieeexplore.ieee.org/abstract/document/10889771/</p>
	<p>OSLO-IC: On-the-Sphere Learned Omnidirectional Image Compression with Attention Modules and Spatial Context Bidgoll, Pascal Frossard?, André Kaup?, Thomas Maugey? https://ieeexplore.ieee.org/abstract/document/10889131/</p>
	<p>Exploiting the Relationship within the Unlabelled Samples by Set Matching for Generalized Category Discovery Qiubo Ma', Hang Yu%, Yuan Shan 3, Pinzhuo Tian 1 https://ieeexplore.ieee.org/abstract/document/10889522/</p>

Continued on next page

Table 2: (Continued)

	<p>Evaluating Contrastive Methodologies for Music Representation Learning Using Playlist Data methods [1, 2] and novel hybrid approaches https://ieeexplore.ieee.org/abstract/document/10888157/</p>
	<p>Fine-tuning and prompt optimization: Two great steps that work better together Dong Sun, Wenya Guo, Xumeng Liu, Ying Zhang*, Zhaoxiang Hou, Zengxiang Li https://arxiv.org/abs/2407.10930</p>
	<p>Digital Twin-Driven Bearing-Fault Detection in Induction Motor and Drives using Graph Sampling and Aggregation Network Haraprasad Badajena, Suryanarayan Majhi, Bivash Chakraborty, Mamata Jenamani, Aurobinda Routray, Ronit Dutta https://ieeexplore.ieee.org/abstract/document/10889484/</p>
	<p>Yi Zhu', Xiangyang Liu!?, Tianqi Pang', Xuncan Xiao!, Xiaofan Zhang33, Chenyou Fan!.* Yi Zhu', Xiangyang Liu!?, Tianqi Pang', Xuncan Xiao!, Xiaofan Zhang33, Chenyou Fan!.*</p>

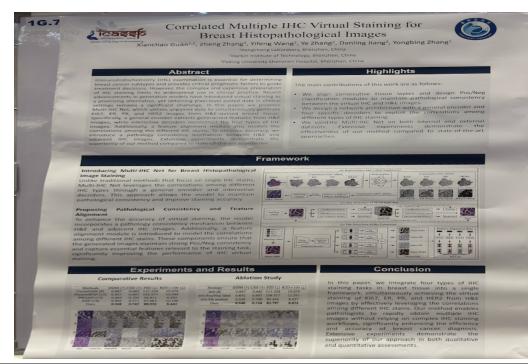
Continued on next page

Table 2: (Continued)

	<p>Text to music audio generation using latent diffusion model: A re-engineering of audiodlm model 1nh tonne Wegner Neteal Ponesin, Deren Hertemans, Rogger Wattenhofer https://www.diva{-}portal.org/smash/record.jsf?pid=diva2:1845150</p>																																																																		
	<p>Exploring the Distribution of Cell Subpopulations in Pancreatic Ductal Adenocarcinoma Slides by Joint Spatial Transcriptomics and Pathology Data Yagi Deng, Wenjie Cai, Bentao Song, Bin Yang, Lingming Kong, Qingfeng Wang*, Jun Huang https://ieeexplore.ieee.org/abstract/document/10890640/</p>																																																																		
	<p>Classification of Eye-Tracking Data Based on Spatiotemporal Attention Encoding Maju Hei, Chen Xia't, Kuan L, Tan Zhangt Beijing University of Chemical Technology / Northwest Polytechnical University *The Affiliated Hospital of Northwest University / **Xian Jiaotong University</p> <p>Task & Contributions Eye tracking has already played an important role in a variety of fields today, such as user interaction, game development, and medical research. Deep learning methods have been proved effective in predicting human eye movement, contributing to disease visual attention.</p> <ol style="list-style-type: none"> We propose an eye movement classification framework based on spatial features and dynamic temporal features to better reconstruct visual attention. We introduce features from a global perspective, accounting for the competitive influence of other features in the classification process. We conducted experiments on three eye tracking datasets and drew three distinct conclusions and improvements about our work model. <p>Main Experiment Results</p> <table border="1"> <thead> <tr> <th colspan="6">AID Identification (w/epoch=0)</th> </tr> <tr> <th></th> <th>AUC</th> <th>F1</th> <th>Sp</th> <th>NPC</th> <th>AUC</th> </tr> </thead> <tbody> <tr> <td>Web</td> <td>0.6412</td> <td>0.5323</td> <td>0.6310</td> <td>0.5647</td> <td>0.6412</td> </tr> <tr> <td>APM</td> <td>0.7086</td> <td>0.5868</td> <td>0.7237</td> <td>0.7870</td> <td>0.7086</td> </tr> <tr> <td>Sphere</td> <td>0.7063</td> <td>0.6556</td> <td>0.6961</td> <td>0.7462</td> <td>0.7063</td> </tr> <tr> <td>SMC</td> <td>0.8350</td> <td>0.7590</td> <td>0.7324</td> <td>0.7479</td> <td>0.8350</td> </tr> </tbody> </table> <p>Methods Overview</p> <p>Fig 1 Diagram of the spatiotemporal attention encoding model</p> <p>Framework This framework extracts spatiotemporal features from spatial (using ViT) and temporal (using enhanced GRU) features and performs sequentially encoded eye movement features, into a classifier to predict class probability.</p> <p>Enhanced GRU We introduce a global temporal modulating factor to GRU for global temporal information, which draws more global hidden states to better represent global sequence patterns, overcoming the limitations of traditional GRU's in capturing long-term dependencies.</p> <p>Fig 2 Global temporal module in the STAE model</p> <p>Ablation Study Ablation analysis of AID identification task</p> <table border="1"> <thead> <tr> <th></th> <th>AUC</th> <th>F1</th> <th>Sp</th> <th>NPC</th> <th>AUC</th> </tr> </thead> <tbody> <tr> <td>Web</td> <td>0.6412</td> <td>0.5323</td> <td>0.6310</td> <td>0.5647</td> <td>0.6412</td> </tr> <tr> <td>APM</td> <td>0.7086</td> <td>0.5868</td> <td>0.7237</td> <td>0.7870</td> <td>0.7086</td> </tr> <tr> <td>Sphere</td> <td>0.7063</td> <td>0.6556</td> <td>0.6961</td> <td>0.7462</td> <td>0.7063</td> </tr> <tr> <td>SMC</td> <td>0.8350</td> <td>0.7590</td> <td>0.7324</td> <td>0.7479</td> <td>0.8350</td> </tr> </tbody> </table> <p>Fig 3 Visualization of feature visual and with temporal modeling of visual task classification</p>	AID Identification (w/epoch=0)							AUC	F1	Sp	NPC	AUC	Web	0.6412	0.5323	0.6310	0.5647	0.6412	APM	0.7086	0.5868	0.7237	0.7870	0.7086	Sphere	0.7063	0.6556	0.6961	0.7462	0.7063	SMC	0.8350	0.7590	0.7324	0.7479	0.8350		AUC	F1	Sp	NPC	AUC	Web	0.6412	0.5323	0.6310	0.5647	0.6412	APM	0.7086	0.5868	0.7237	0.7870	0.7086	Sphere	0.7063	0.6556	0.6961	0.7462	0.7063	SMC	0.8350	0.7590	0.7324	0.7479	0.8350
AID Identification (w/epoch=0)																																																																			
	AUC	F1	Sp	NPC	AUC																																																														
Web	0.6412	0.5323	0.6310	0.5647	0.6412																																																														
APM	0.7086	0.5868	0.7237	0.7870	0.7086																																																														
Sphere	0.7063	0.6556	0.6961	0.7462	0.7063																																																														
SMC	0.8350	0.7590	0.7324	0.7479	0.8350																																																														
	AUC	F1	Sp	NPC	AUC																																																														
Web	0.6412	0.5323	0.6310	0.5647	0.6412																																																														
APM	0.7086	0.5868	0.7237	0.7870	0.7086																																																														
Sphere	0.7063	0.6556	0.6961	0.7462	0.7063																																																														
SMC	0.8350	0.7590	0.7324	0.7479	0.8350																																																														

Continued on next page

Table 2: (Continued)

 <p>Correlated Multiplex IHC Virtual Staining for Breast Histopathological Images</p> <p>Abstract</p> <p>Highlights</p> <ul style="list-style-type: none"> The main contributions of this work are as follows: We align immunohistochemical images with corresponding histopathological images. We design a network architecture to correlate multiplex IHC images. We validate our method on both simulated and real-world datasets. Our method outperforms the state-of-the-art approaches. <p>Framework</p> <p>Experiments and Results</p> <p>Conclusion</p>	<p>Virtual multiplex immunohistochemistry: application on cell block of effusion and aspiration cytology Xianchao Guan¹², Zheng Zhang^{?, Yifeng Wang[?], Ye Zhang^{?, Danling Jiang[?], Yongbing Zhang[?]}}</p> <p>https://onlinelibrary.wiley.com/doi/abs/10.1002/dc.24344</p>
 <p>Vie sin an Aude Spectresun Taedome 981 on the Speechcommanda</p> <p>Introduction</p> <p>Related Work</p> <p>Methods</p> <p>Results</p> <p>Conclusion</p>	<p>Vie sin an Aude Spectresun Taedome 981 on the Speechcommanda</p>

Continued on next page

Table 2: (Continued)

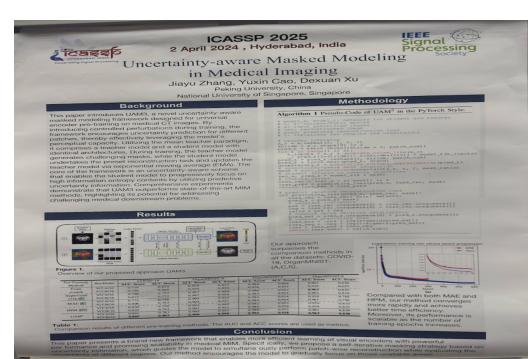
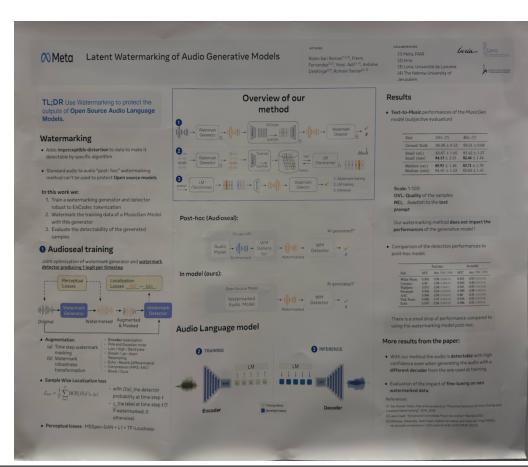
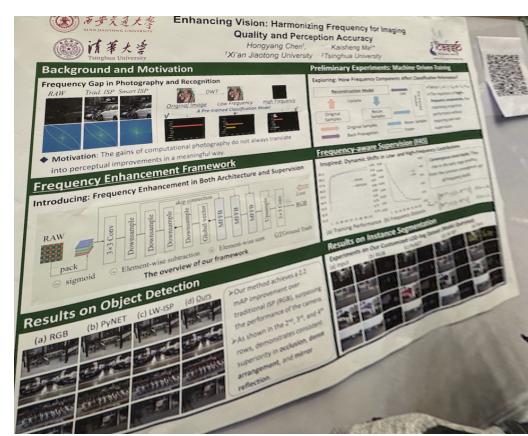
	<p>Exploring the Distribution of Cell Subpopulations in Pancreatic Ductal Adenocarcinoma Slides by Joint Spatial Transcriptomics and Pathology Data Yagi Deng, M cnje Cai, Benao Song, Bin Yang, Limphing Kung , Qedeng Pung*, Jam H https://ieeexplore.ieee.org/abstract/document/10890640/</p>
	<p>NanoGen: A High-affinity Nanobody Generation Model with Guided Diffusion Dezhij Wu*, Xuejiao Liu*, Yiming Qin*, Stephanie M. Linker*, Karin Hrovatin*, Alexander V.Hopp*, Feng Tan** https://ieeexplore.ieee.org/abstract/document/10888039/</p>
	<p>ApinAPDE, a curated repository with physicochemical properties (GRAVY, net charge, isoelectric point, molecular weight). Emas: 2230112006.M.00.u.59. cong na00012e.ntu.edu.sg. as.293th@ntu.edu.5g</p>
	<p>Toward robust early detection of alzheimer's disease via an integrated multimodal learning approach Yifei Chen, Shenghao Zhu. Zhaojie Fang. Chang Liu, Binfeng Zou, Linwei Qiu, Yuhe Wang. Shuo Chang. Fan Jia, Felwel Qin*. Jin Fang. Yong Peng, Changmiao Wang https://ieeexplore.ieee.org/abstract/document/10888363/</p>

Continued on next page

Table 2: (Continued)

Continued on next page

Table 2: (Continued)

 <p>This paper introduces UAMA, a novel uncertainty-aware masked modeling framework for medical image segmentation. It addresses the challenge of learning from incomplete and noisy training data by introducing context-aware uncertainty-aware masked patches. The framework consists of three main components: a teacher network, a student network, and a uncertainty-aware masked patch generation module. The teacher network generates uncertainty maps, while the student model generates corresponding masks. The uncertainty-aware masked patch generation module then generates challenging medical image patches for training. The results show that UAMA achieves state-of-the-art performance on various medical image segmentation tasks.</p>	<p>Masked image modeling advances 3d medical image analysis <i>Der formace and promising scalability in medical MIM. Spect ically, we propose a tel-literative masking stratcoy based on</i> https://openaccess.thecvf.com/content/WACV2023/html/Chen_Masked_Image_Modeling_Advances_3D_Medical_Image_Analysis_WACV_2023_paper.html</p>
 <p>This poster presents a method for protecting the outputs of open-source audio language models. It uses watermarking to detect forged samples and ensure the integrity of generated samples. The method involves adding imperceptible noise to the data for training, generating a watermark, and then extracting it from the generated samples. The results show that the watermarking method can be used to protect open-source audio models.</p>	<p>Latent watermarking of audio generative models <i>Losses*Inlt..</i> https://ieeexplore.ieee.org/abstract/document/10889782/</p>
 <p>This poster introduces HYMAN, a hybrid memory and attention network for unsupervised anomaly detection. It uses a dual-path architecture with a shared feature space to handle complex anomalies. The network consists of an encoder, a hybrid memory module, and an attention module. The results show that HYMAN outperforms existing methods on various datasets.</p>	<p>HYMAN: Hybrid Memory and Attention Network for Unsupervised Anomaly Detection <i>Jiahao Li, Yiqiang Chen, Yunbing Xing, Yang Gu, Xiangyu Lan</i> https://ieeexplore.ieee.org/abstract/document/10890028/</p>
 <p>This poster presents a frequency enhancement framework for improving image quality and perception accuracy. It introduces frequency enhancement in both architecture and supervision. The results show that the proposed framework achieves better performance than traditional methods on various datasets.</p>	<p>A general framework for object detection <i>>As shown in the 29, 3P%, and 4°</i> https://ieeexplore.ieee.org/abstract/document/710772/</p>

Continued on next page

Table 2: (Continued)

Continued on next page

Table 2: (Continued)

<p>Retrieval Augmented Correction of Named Entity Speech Recognition Errors</p> <p>Ernest Pusateri, Anmol Walia, Aniruch Kashi, Bortik Bandyopadhyay, Nadia Hyder, Sayantan Mahinder, Raviteja Anantha, Daben Liu*, and Sashank Gondala*</p> <p>Introduction</p> <ul style="list-style-type: none"> • Motivation: Most end-to-end ASR systems are remarkably accurate, but still struggle to recognize named entities. • Our goal is to recognize named entities using LLMs as versatile tools for service NLP tasks. • When a database of relevant knowledge is available, retrieval-augmented correction (RAC) is effective. • • Retrieval-Augmented Speech Recognition (RASR), performs poorly if no one named entity is present in the knowledge base (KDB). • • Retrieval-Augmented Correction (RAC) is proposed to mitigate this limitation. <p>Query Generation</p> <ul style="list-style-type: none"> • Denote phrases from ASR output that will be used as entities. • Motivation: Query can generated from all ordered word sequence to length N. • Motivation: A large set of entities is generated using hand curated set of regular expressions. • Motivation: A large set of entities is generated using regular expression that starts by named entities. <p>Entity Retrieval</p> <ul style="list-style-type: none"> • Use queries to retrieve entities from an entity knowledge base. • Motivation: Entity retrieval is done using a large set of regular expressions. • Motivation: A large set of entities is generated using regular expression that starts by named entities. <p>Context Construction</p> <ul style="list-style-type: none"> • Create a list for each of the remaining entities. • Create a list for each of the remaining entities. • Create a list for each of the remaining entities. • Create a list for each of the remaining entities. <p>Data</p> <ul style="list-style-type: none"> • STOI: Open-source dataset with eight VM domains. • Synthetic Generated test references from STOI. <p>Results</p> <p>Baseline Methods</p> <table border="1"> <thead> <tr> <th></th> <th>Recall@100</th> <th>Recall@500</th> <th>Recall@1000</th> <th>Recall@5000</th> </tr> </thead> <tbody> <tr> <td>Open-MRIS</td> <td>21.2</td> <td>20.8</td> <td>20.7</td> <td>20.6</td> </tr> <tr> <td>DS</td> <td>21.2</td> <td>20.8</td> <td>20.7</td> <td>20.6</td> </tr> <tr> <td>AM</td> <td>21.2</td> <td>20.8</td> <td>20.7</td> <td>20.6</td> </tr> <tr> <td>AM+Probase</td> <td>32.1</td> <td>31.7</td> <td>31.6</td> <td>31.5</td> </tr> <tr> <td>RAC</td> <td>32.1</td> <td>31.7</td> <td>31.6</td> <td>31.5</td> </tr> </tbody> </table> <p>• phonetic ANEs slightly outperform orthographic ANEs, but require training phonetic orthographic ANEs for experiments</p> <p>Query Generation</p> <ul style="list-style-type: none"> • LLM without hints results in small improvements on RAC. • Overall, 33% (9%) relative recall reduction in synthetic test sets with promptings on STOI. <p>Conclusion</p> <ul style="list-style-type: none"> • Presented a RAC-inspired technique for connecting named entities to ASR output using a pre-trained LLM. • A simple yet effective technique in outperformed methods, retrieving entity references from knowledge bases. • Best system achieved 37% (39%) relative recall reduction in synthetic test sets with promptings on STOI without regressing on multi-domain test set. 		Recall@100	Recall@500	Recall@1000	Recall@5000	Open-MRIS	21.2	20.8	20.7	20.6	DS	21.2	20.8	20.7	20.6	AM	21.2	20.8	20.7	20.6	AM+Probase	32.1	31.7	31.6	31.5	RAC	32.1	31.7	31.6	31.5	<p>Retrieval augmented correction of named entity speech recognition errors Ernest Pusateri, Anmol Walia, Aniruch Kashi, Bortik Bandyopadhyay, Nadia Hyder, Sayantan Mahinder, Raviteja Anantha, Daben Liu*, and Sashank Gondala**</p> <p>https://ieeexplore.ieee.org/abstract/document/10888936/</p>
	Recall@100	Recall@500	Recall@1000	Recall@5000																											
Open-MRIS	21.2	20.8	20.7	20.6																											
DS	21.2	20.8	20.7	20.6																											
AM	21.2	20.8	20.7	20.6																											
AM+Probase	32.1	31.7	31.6	31.5																											
RAC	32.1	31.7	31.6	31.5																											
<p>Self-Prompting Driven SAM2 for 3D Medical Image Segmentation</p> <p>Sheng Wei; Song Qiu*; Mei Zhou; He Zhang; Yan Wang; Qingli Li</p> <p>East China Normal University</p> <p>Abstract</p> <p>The main advantage in large foundation models, SAM2, has demonstrated superior performance in various downstream applications due to their capability to efficiently segment inputs. However, training extensive training on medical images highly requires manual annotations, which is time-consuming and laborious. To alleviate this problem, we propose SAM2-SP, which adopts Low-dose Prompt, self-prompting strategy that provides most cost-effective solution. The low-dose prompt is a small set of images that are injected into the model during training. This strategy achieves state-of-the-art performance on the public Synapse dataset and achieves better performance than other state-of-the-art 3D medical segmentation approaches, the vanilla SAM and SAM2. Method</p> <p>A. Low-dose selected image encoder</p> <p>On the design of Low-dose Selected Image Encoder (SAM2-SP). On the design of Low-dose Selected Image Encoder (SAM2-SP), we propose a novel Selection Strategy to select the selected images. On the design of Low-dose Selected Image Encoder (SAM2-SP), we propose a novel Selection Strategy to select the selected images. While keeping its weight frozen to preserve the original feature map, it introduces a mask to randomly sample a subset of images through Laff-Ackley. These branches are fused with the original feature map. This update applies to the parameters in the image encoder. This update achieves state-of-the-art performance on the public Synapse dataset and achieves better performance than other state-of-the-art 3D medical segmentation approaches, the vanilla SAM and SAM2.</p> <p>B. Efficient Prompting</p> <p>On the design of Low-dose Selected Image Encoder (SAM2-SP). On the design of Low-dose Selected Image Encoder (SAM2-SP), we propose a novel Selection Strategy to select the selected images. While keeping its weight frozen to preserve the original feature map, it introduces a mask to randomly sample a subset of images through Laff-Ackley. These branches are fused with the original feature map. This update applies to the parameters in the image encoder. This update achieves state-of-the-art performance on the public Synapse dataset and achieves better performance than other state-of-the-art 3D medical segmentation approaches, the vanilla SAM and SAM2.</p> <p>C. Contribution</p> <p>After generating the saliency map, we need to further evaluate the representations of class slice. To this end, we calculate a confidence score for each slice based on the saliency map:</p> $R_i = \left\ \hat{S}(x, y) \right\ _2 \ \hat{S}(x, y) \ _2$ <p>After obtaining the confidence scores, considering the high similarity between adjacent slices, directly selecting the slice with the highest confidence might fail in evenly covering the entire volume. Therefore, we propose a novel low-to-high scale ratio function to achieve better coverage and more uniform distribution of slices.</p> <p>Experiment Results</p> <p>*Corresponding author: This work was supported by STCSM Research Project and the National Natural Science Foundation of China.</p>	<p>Self-Prompting Driven SAM2 for 3D Medical Image Segmentation a Sorted index from C: s, -(6r,62..., 62...,4)</p> <p>https://ieeexplore.ieee.org/abstract/document/10889344/</p>																														
<p>Integrating Pause Information with Word Embeddings in Language Models for Alzheimer's Disease Detection from Spontaneous Speech</p> <p>Yi Pi, Wei-Cheng Zhang, Tsinghua University, China</p> <p>ABSTRACT</p> <p>Alzheimer's disease (AD) is a progressive neurodegenerative disorder that causes cognitive decline and memory loss. Early detection of AD is crucial for timely intervention and management. In this paper, we propose a novel framework for Alzheimer's disease detection from spontaneous speech. The framework integrates pause information and word embeddings to improve the detection accuracy. We first extract pause features from the speech signals. Then, we use a pre-trained language model (LM) to generate word embeddings. Finally, we combine the pause features and word embeddings to build a classification model. The experimental results show that our proposed framework achieves a higher detection accuracy than the baseline methods.</p> <p>Methods</p> <p>We propose a novel framework for Alzheimer's disease detection from spontaneous speech. The framework integrates pause information and word embeddings to improve the detection accuracy. We first extract pause features from the speech signals. Then, we use a pre-trained language model (LM) to generate word embeddings. Finally, we combine the pause features and word embeddings to build a classification model. The experimental results show that our proposed framework achieves a higher detection accuracy than the baseline methods.</p> <p>Results</p> <p>The experimental results show that our proposed framework achieves a higher detection accuracy than the baseline methods.</p>	<p>Integrating Pause Information with Word Embeddings in Language Models for Alzheimer's Disease Detection from Spontaneous Speech</p> <p>Yi Pi, Wei-Cheng Zhang, Tsinghua University, China</p> <p>https://arxiv.org/abs/2501.06727</p>																														
<p>Multi-scale Context Intertwining for Panoramic Renal Pathology Segmentation</p> <p>Ye Zhang², Xianchao Guan¹³, Hengrui LI¹, Xiangming Yan¹, Ziyue Wang¹, Yongbing Zhang¹</p> <p>¹ Faculty of Computing, Shenzhen Institute of Technology (Shenzhen) ² Department of Computer Science and Technology, Tsinghua University, China</p> <p>Abstract</p> <p>Medical image segmentation is a key task in medical image analysis. In this paper, we propose a novel multi-scale context intertwining framework for panoramic renal pathology segmentation. The framework consists of three main components: a multi-scale context intertwining module, a multi-scale auxiliary network, and a multi-scale feature fusion module. The multi-scale context intertwining module is designed to capture multi-scale context information and intertwine them. The multi-scale auxiliary network is used to provide auxiliary information for the segmentation process. The multi-scale feature fusion module is used to fuse the features from different scales. The experimental results show that our proposed framework achieves state-of-the-art performance in panoramic renal pathology segmentation.</p> <p>Introduction</p> <p>Medical image segmentation is a key task in medical image analysis. In this paper, we propose a novel multi-scale context intertwining framework for panoramic renal pathology segmentation. The framework consists of three main components: a multi-scale context intertwining module, a multi-scale auxiliary network, and a multi-scale feature fusion module. The multi-scale context intertwining module is designed to capture multi-scale context information and intertwine them. The multi-scale auxiliary network is used to provide auxiliary information for the segmentation process. The multi-scale feature fusion module is used to fuse the features from different scales. The experimental results show that our proposed framework achieves state-of-the-art performance in panoramic renal pathology segmentation.</p> <p>Conclusions</p> <p>Our proposed framework achieves state-of-the-art performance in panoramic renal pathology segmentation. By combining multi-scale context intertwining, multi-scale auxiliary network, and multi-scale feature fusion, our model effectively integrates multi-scale context information and auxiliary information, improving segmentation performance. Future work will focus on further optimization of the framework and explore more complex and deeper optimization of the framework.</p> <p>Acknowledgements</p> <p>This work was supported by grants from the National Natural Science Foundation of China (No. 61976111) and the Fundamental Research Funds for the Central Universities (No. HIT.IFC.2022.0356).</p> <p>References</p> <p>[1] D. Cremers, M. Rother, and K. Schindler, "A multi-scale dynamic programming approach for multi-class segmentation," in <i>CVPR</i>, 2009, pp. 209–216.</p> <p>[2] X. Guan, Y. Zhang, H. Li, et al., "Multi-scale organ segmentation via multi-scale feature abstraction," <i>IEEE TMI</i>, vol. 39, pp. 3415–3426, 2020.</p>	<p>Multi-scale Context Intertwining for Panoramic Renal Pathology Segmentation</p> <p>Ye Zhang², Xianchao Guan¹³, Hengrui LI¹, Xiangming Yan¹, Ziyue Wang¹, Yongbing Zhang¹</p> <p>https://ieeexplore.ieee.org/abstract/document/10889659/</p>																														

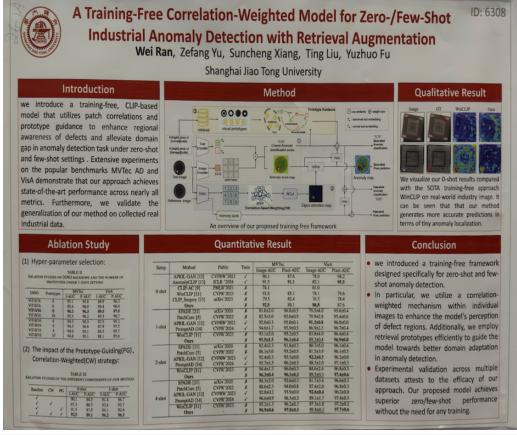
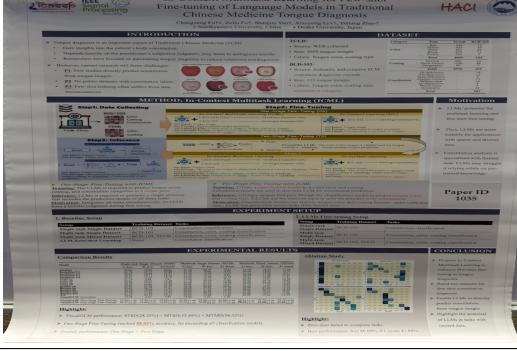
Continued on next page

Table 2: (Continued)

	<p>SELMA: A Speech-Enabled Language Model for Virtual Assistant Interactions <i>Simplified Pigstina, Meduces complarty</i> https://ieeexplore.ieee.org/abstract/document/10890139/</p>
	<p>Glial-neuronal interactions in Alzheimer's disease: the potential role of a 'cytokine cycle' in disease progression <i>hutcmoshini001@Bentuedusg, choc0010@e.ntu.edu.sg, yiha001@e.ntu.edu.sg, congao001@e.ntu.edu.sg, asjagath@ntu.edu.sg</i> https://onlinelibrary.wiley.com/doi/10.1111/j.1750{-}3639.1998.tb00136.x</p>
	<p>Wireless Sensor Networks: 13th China Conference, CWSN 2019, Chongqing, China, October 12–14, 2019, Revised Selected Papers 1. <i>College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing, China</i> https://books.google.com/books?hl=en&lr=&id=o3fADwAAQBAJ&oi=fnd&pg=PR6&dq=1.+College+of+Computer+Science+and+Technology,+Chongqing+University+of+Posts+and+Telecommunications,+Chongqing.+China&ots=xK3vgQuzGb&sig=1R1Jj8oR{-}X4PqqWdWiKUW8crgKE</p>
	<p>AI-Generated Music Detection and its Challenges <i>Suno, Udio, Riffusion, ...</i> https://arxiv.org/abs/2501.10111</p>

Continued on next page

Table 2: (Continued)

 <p>A Training-Free Correlation-Weighted Model for Zero-/Few-Shot Industrial Anomaly Detection with Retrieval Augmentation Wei Ran, Zefang Yu, Suncheng Xiang, Ting Liu, Yuzhuo Fu Shanghai Jiao Tong University</p> <p>Introduction we introduce a training-free, CLP-based model that utilizes prior knowledge and prototype guidance to enhance regional awareness of defects and alleviate domain gap in anomaly detection task under zero-shot or few-shot settings. Extensive experiments on three benchmarks (MoCap, MoAn and Vis4) demonstrate that our approach achieves state-of-the-art performance across nearly all metrics. Furthermore, we validate the generalization of our method on collected real industrial data.</p> <p>Method The proposed framework consists of three main components: 1. Prototype Guiding: A pre-trained CLP model is used to generate prototypes for each category. 2. Region-aware learning: The model uses a correlation-weighted mechanism to emphasize regions of interest. 3. Retrieval Augmentation: The model retrieves prototypes from a database to guide the model towards better domain adaptation in anomaly detection.</p> <p>Qualitative Result We visualize our 0-shot results compared with the SOTA training-free approach. The qualitative results show that our method can be seen that our method generates more accurate predictions in terms of key anomaly locations.</p>	<p>A Training-Free Correlation-Weighted Model for Zero-/Few-Shot Industrial Anomaly Detection with Retrieval Augmentation Wei Ran, Zefang Yu, Suncheng Xiang, Ting Liu, Yuzhuo Fu https://ieeexplore.ieee.org/abstract/document/10890083/</p>
 <p>Parameter-efficient fine-tuning of large-scale pre-trained language models Changzeng Fits, Zelin Fut, Shaojun Yant, Xiaoyong Lyvt, Yuliang Zhaot</p> <p>INTRODUCTION Fine-tuning of Language Models in Traditional Clinical Medicine Disease Diagnosis</p> <p>Dataset HACI</p> <p>EXPERIMENTAL SETUP 1. Data Collection 2. Data Pre-Processing 3. Model Selection 4. Model Training 5. Model Evaluation</p> <p>EXPERIMENTAL RESULTS Comprehensive Results</p> <p>Conclusion Proposed framework can significantly improve the performance of pre-trained language models for clinical medicine disease diagnosis tasks.</p>	<p>Parameter-efficient fine-tuning of large-scale pre-trained language models Changzeng Fits, Zelin Fut, Shaojun Yant, Xiaoyong Lyvt, Yuliang Zhaot</p> <p>https://www.nature.com/articles/s42256{-}023{-}00626{-}4</p>