

2024년 한국방송·미디어공학회 추계학술대회 후기

서영호 교수 (조직위원장, 광운대학교)



조직위원회

- 대회장 : 박구만 회장 (한국방송·미디어공학회/서울과학기술대학교)
- 조직위원장 : 서영호 교수 (광운대학교)
- 프로그램위원장 : 오민식 교수 (명지대학교)
- 프로그램위원
 - 강제원 교수 (이화여자대학교) 김영태 본부장 (RAPA)
 - 김제우 센터장 (KETI) 류은석 교수 (성균관대학교)
 - 박재형 교수 (서울대학교) 서광덕 교수 (연세대학교)
 - 신춘성 교수 (전남대학교) 심동규 교수 (광운대학교)
 - 오관정 책임 (ETRI) 오병태 교수 (한국항공대학교)
 - 이영호 교수 (목포대학교) 임양미 교수 (덕성여자대학교)
 - 정진우 책임 (KETI) 조병철 교수 (동아방송예술대학교)
 - 추현곤 실장 (ETRI) 홍기훈 실장 (ETRI)



추계학술대회 프로그램 구성 및 내용

최근 방송과 미디어 산업은 급속한 기술 발전과 함께 새로운 시대를 맞이하고 있다. 인공지능(AI), 머신러닝, 빅데이터, 5G 통신, 그리고 초실감형 콘텐츠는 기존의 방송 기술과 융합되어 미디어 환경의 혁신을 이끌고 있다. 특히 메타버스, 디지털 트윈, 확장현실(XR) 등의 첨단 기술은 방송과 미디어가 단순한 정보 전달을 넘어 몰입감 있는 경험을 제공하는 새로운 차원의 산업으로 진화하는 기반이 되고 있는데, 이러한 변화 속에서 한국방송·미디어공학회는 4차 산업혁명 시대에 발맞추어 최신 기술과 표준을 연구·개발하며, 국내 방송과 미디어 산업의 지속적인 발전을 도모하기 위해 2024년도 추계학술대회를 개최하였다. 이번 추계학술대회는 그 어느 때보다 기술적으로 흥미로운 논문들이 발표되었고, 방송 및 미디어 분야에서의 최신 트렌드를 확인할 수 있는 중요한 자리가 되었다. 프로그램은 5개의 특별 세션, 3개의 일반 세션, 그리고 5개의 포스터 세션으로 구성되었다. 특별 세션에서는 <생성형 AI 파인튜닝과 콘텐츠 개발>, <AI 기반의 영상분석 및 통신 기술을 활용한 응용시스템>, <AI 기반 바이오 의료 데이터 분석 기술>, <VCM 및 FCM>, <원격헬스케어기술> 등 최신 기술 연구 성과가 다루어졌다. 또한, 5G와 차세대 방송 전송 기술, 저지연 스트리밍, 그리고 블록체인을 통한 미디어 보안 기술에 대한 논의도 활발히 이루어졌다. 일반 세션과 포스터 세션에서는 딥러닝을 활용한 콘텐츠 생성, 초고해상도 등 다양한 연구 결과가 발표되었다. 이러한 기술들은 기존 미디어의 한계를 넘어서고, 참가자들에게 더욱 몰입감 있는 경험을 제공하는 데 기여할 수 있었다.

이번 학술대회에서는 방송과 미디어 산업의 미래를 이끌어 갈 신진 연구자들이 자신의 연구를 발표하고, 학계 및 산업계 전문가들과 교류할 수 있는 기회도 제공되었다. 더불어 스튜디오메타케이, LG AI 연구원, 백스포트, Shimonoseki City University 등 각계 전문가들의 특별 강연을 통해 미래 미디어 기술의 방향성과 비전을 제시하는 계기를 마련하였다. 스튜디오메타케이 김광집 대표는 ‘AI 버추얼 아티스트 제작 사례와 생성형 AI를 활용한 콘텐츠 제작’을 주제로, LG AI 연구원 이순영 연구위원은 ‘EXAONE: 멀티모달 AI 활용사례’에 대해, (주)백스포트 신윤수 대표는 ‘AI와 ICVFX를 활용한 영상 콘텐츠 제작 프로세스 사례 분석’에 대해, Shimonoseki City University의 Yukihiko Bandoh 교수는 ‘Advances in video coding: from standardization to functional extensions’에 대해 심도있는 내용을 전달하였다.

또한, 미래의 방송·미디어 산업을 이끌어 갈 대학생 및 고등학생들을 위한 <대학생 및 주니어 논문 경진대회>도 개최되었는데, 학생들은 창의적인 아이디어와 연구 결과를 공유하고, 전문가들로부터 피드백을 받을 수 있는 기회를 가짐으로써, 자신의 역량을 키울 수 있는 소중한 경험을 얻을 수 있는 좋은 계기가 되었다.

초청 강연

스튜디오메타케이 김광집 대표는 ‘AI 버추얼 아티스트 제작 사례와 생성형 AI를 활용한 콘텐츠 제작’을 주제로, AI 기술을 활용한 버추얼 아티스트의 제작 방법들과 실증 사례 및 활용방법에 대해서 소개하였다. 또한, 생성형 AI를 활용한 광고, 뮤직비디오, 드라마 트레일러 등의 콘텐츠 제작 과정과 실증 사례를 통해 콘텐츠 제작의 새로운 가능성에 대해서 제시하였다.

LG AI 연구원 멀티모달랩 이순영 연구위원은 ‘EXAONE: 멀티모달 AI 활용사례’에 대해 소개하였다. EXAONE은 전문가를 위한 AI 모델이며, EXAONE Multimodal은 이미지-to-텍스트 모델로 하이레벨 비전 피처와 EXAONE LLM을 이용한 구조로 제품 홍보 문구, SNS를 작성하는 태스크에 활용되고 있다. EXAONE Atelier는 텍스트-to-이미지 모델로 라이선스 이슈 없는 데이터로 학습하였으며 디자인 편집, 도안 생성에 활용되고 있다. EXAONE Path는 병리이미지-Genie 멀티모달 모델로 바이오마커 예측을 통해 임상실험 가속화 태스크에 활용되고 있다.

(주)백스포트 신윤수 대표는 ‘AI와 ICVFX를 활용한 영상 콘텐츠 제작 프로세스 사례 분석’이라는 주제로 발표하였다. 2019년 주목받았던 리얼타임 CG 촬영 기술인 ICVFX와 2022년부터 빠르게 발전한 생성형 AI를 영상 업계 제작자에게 교육하고, 이를 실제 제작까지 이어나가 실질적인 활용성을 알아보았다. 이 같은 진행 과정을 복기하여 영상 제작 프로세스를 새롭게 도입하는 과정을 통해 기술에 대한 실증과 영상업계 제작자들의 의견을 듣는 기회를 마련하였다.

일본 시모노세키 대학교의 Yukihiro Bandoh 교수는 ‘Advances in video coding: from standardization to functional extensions’라는 주제로, 지난 30년간 국제 표준화를 통해 발전해 온 비디오 압축 기술 및 표준화 기술을 기반으로 한 자신의 연구를 소개하였다.



특별 세션

특별 세션은 5개로 구성되었으며, 총 24편의 논문이 발표되었다.

1. 특별 세션 1: 생성형 AI 파인튜닝과 콘텐츠 개발

특별 세션 1에서는 ‘생성형 AI 파인튜닝과 콘텐츠 개발’을 주제로 6편의 논문이 발표되었으며, KAI-ST(1편), 서울과학기술대학교(3편), 시그마케이주식회사/KBS/서울과학기술대학교(1편), 서울과학기술대학교/시그마케이주식회사(1편)가 참여하였다. 논문 발표는 생성형 AI 기술의 다양한 응용 방안과 연관된 연구를 중심으로 이루어졌다.

‘Design of an Automatic Adaptive Video Caption Generation Technology for Users in Need of Linguistic Accommodation’은 다양한 언어적 요구를 가진 사용자(청각 장애인, 외국어 학습자, 언어 발달 장애인 등)를 위해 적응형 자막 생성 시스템을 설계하였다. Transformer 기반 대규모 언어 모델(LLMs)을 활용해 사용자 맞춤형 자막을 생성하는 이 시스템은 유튜브 콘텐츠를 사용한 실험에서 그 적용 가능성을 확인하였다. ‘An Experimental Study on the Creation of Film Scenario Using Generative AI’는 생성형 AI 도구를 활용해 영화 시나리오를 창작하는 실험을 진행하였다. 고전 설화와 세 가지 주요 장르(호러, 로맨스, SF)에 기반한 시나리오를 생성하고 그 완성도와 장르 적합성을 분석하였다. 연구는 AI 도구의 가능성과 한계를 평가하며 영화 제작의 새로운 방향을 제시하였다. ‘Development of a Customizing Model for Producing News Broadcast Graphics Using Generative Artificial Intelligence: Focusing on KBS News Graphics’는 KBS 보도 그래픽 스타일을 반영한 생성형 AI 모델을 개발하였다. Stable Diffusion XL 및 LoRA 기법을 활용해 전통적인 그래픽 제작 과정을 단축시키고, Image-to-Video 기술을 통해 정적인 이미지를 동적 비디오로 변환하는 가능성을 탐구하였다. ‘Analysis of Data Preprocessing Experiments for LLM Creative Answers’는 데이터 전처리 기법을 통해 대규모 언어 모델(LLM)의 창의적 답변 생성 성능을 향상시키는 실험을 진행하였다. 실험 결과, 명시적이고 설명적인 데이터 전처리 방식을 적용한 모델이 기존 모델에 비해 모호한 질문에 더 정확히 답변할 수 있음을 확인하였다. ‘A Case Study of Film Production Using Generative AI Commercial Tools’는 Midjourney와 Gen-3 같은 상용화된 생성형 AI 도구를 활용해 영화 제작을 진행하였다. 도구의 신속한 결과 생성 능력과 한계점을 분석하며, 상용 도구를 활용한 영화 제작 가능성을 고찰하였다. ‘User-Customized Real-Time Data Learning-Based Document Retrieval Framework Using RAG’는 RAG(관련성 증강 생성) 기법을 활용한 실시간 문서 검색 프레임워크를 개발하였다. Groq API와 Llama-Index를 활용해 문서 검색과 대화의 일관성을 유지하며, 사용자 맞춤형 응답 생성 가능성을 입증하였다. 본 세션은 생성형 AI 기술의 창의적 응용과 실용적 가능성을 다루며, 다양한 산업에서 활용 가능한 기술적 기여와 응용 사례를 제시하였다.

2. 특별 세션 2: AI 기반의 영상분석 및 통신 기술을 활용한 응용시스템

특별 세션 2에서는 ‘AI 기반의 영상분석 및 통신 기술을 활용한 응용시스템’을 주제로 총 4편의 논문이 발표되었으며 서울시립대학교/KETI(2편), KETI(2편)가 참여하였다. 연구는 AI를 활용한 실시간 데이터 처리와 통신 기술 최적화를 중점적으로 다루었다.

‘인종에 강인한 딥러닝 기반의 표정 분류 기법’은 다인종 환경에서 표정 인식 모델의 성능을 개선하는 방법을 제안하였다. MTCNN을 이용해 얼굴을 검출하고, DAN을 활용해 표정을 분류하였으며, 다양한 인종 데이터에서 평균 91.25%의 정확도를 기록하였다. ‘삼중항 샘플링 전략과 ANN 알고리즘을 이용한 딥러닝 기반의 유사 이미지 검색 기법’은 삼중항 손실과 ANN 알고리즘을 결합하여 유사 이미지 검색의 효율성과 성능을 향상시켰다. In-class 및 out-of-class negative 샘플을 나누어 학습을 진행한 결과, 기존 방법보다 빠르고 정확한 검색 성능을 보여주었다. ‘인간의 감정 이해를 위한 딥러닝 기반의 미세 행동 인식 기술’은 optical flow를 활용하여 인간의 미세 행동을 인식하는 방법을 제안하였다. iMiGUE 데이터셋을 사용하여 32가지 미세 행동을 학습하고 평가한 결과, 높은 정확도를 달성하였다. ‘ProtoNet을 이용한 패션 이미지의 Few-shot 분류 기법’은 패션 데이터에 적합한 사전 학습 모델 FashionClip과 ProtoNet을 결합하여 few-shot 학습 기법을 제안하였다. 신규 분류 작업에 적응 가능한 모델을 통해 우수한 성능을 확인하였다. 본 세션은 영상 분석과 통신 기술에서 AI를 활용한 최적화와 응용 가능성을 논의하며, 다양한 분야에서 AI 기술의 실질적인 활용 방안을 제시하였다.

3. 특별 세션 3: AI 기반 바이오 메디컬 데이터 분석 기술

특별 세션 3에서는 ‘AI 기반 바이오 메디컬 데이터 분석 기술’을 주제로 총 4편의 논문이 발표되었다. 참여 기관으로는 동국대학교(1편), 수원대학교(1편), 서울대학교/인하대학교(1편), 경북대학교(1편)가 포함되었으며, 발표는 바이오 메디컬 데이터를 활용한 딥러닝 모델 개발과 의료 데이터 분석 기술을 중심으로 진행되었다.

‘공간전사체 데이터의 패스웨이 증강 대조학습 기반 공간 도메인 식별에 관한 연구’는 유전자 간 상호작용과 공간적 맥락을 더 정확히 해석하기 위해 생물학적 경로(biological pathway)를 활용한 새로운 방법을 제안하였다. 유전자 발현 데이터를 그래프 형태로 변환하고 병리학적 이미지와 결합하여 조직의 구조와 기능을 효과적으로 구분할 수 있었다. ‘의료 영상의 병변 영역 분할을 위한 딥러닝 모델의 손실 함수 성능 비교’는 의료 영상의 병변 영역을 분할하는 딥러닝 모델에서 손실 함수의 영향을 비교 분석하였다. 다양한 데이터셋과 모델에 대해 실험한 결과, 복합 손실 함수인 IoU-Focal 손실이 가장 우수한 성능을 보였으며, 손실 함수 선택 가이드라인을 제시하였다. ‘병리 이미지 기반 유방암 아종 예측을 위한 소량 학습 분류 모델 개발’은 병리 이미지와 RNA-seq 데이터를 통합하여 유방암 아종을 분류하는 few-shot 학습 모

델을 제안하였다. The Cancer Genome Atlas 데이터를 활용하여 침윤성 소엽암(ILC)과 침윤성 유관암(IDC)을 분류하였으며, 소량의 데이터로도 높은 예측 성능을 달성하였다. ‘임상정보 기반의 COVID-19 환자들의 중증도 예측 모델’은 COVID-19 환자의 임상 데이터를 기반으로 중증도를 예측하는 Seq2Seq 모델을 개발하였다. 산소치료 수준을 기준으로 92%의 예측 정확도를 달성하였으며, 중증도 예측에 중요한 변수들을 식별하여 모델의 활용 가능성을 입증하였다. 본 세션은 의료 및 생명과학 분야에서 AI 기반 솔루션의 실질적인 기여를 논의하였다.

4. 특별 세션 4: VCM 및 FCM

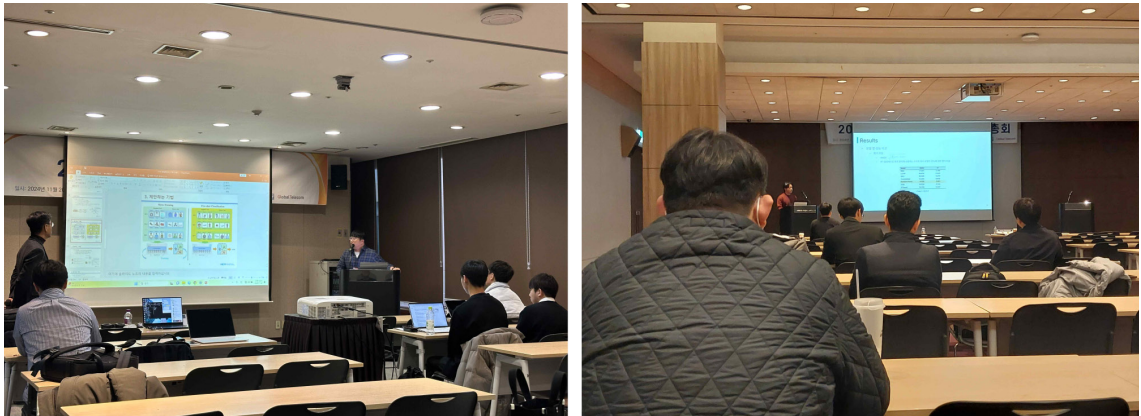
특별 세션 4는 ‘VCM 및 FCM’을 주제로 총 3편의 논문이 발표되었다. 연세대학교/ETRI(1편), 한국항공대학교/ETRI(1편), 건국대학교/ETRI(1편)가 참여하였으며, 발표는 차세대 비디오 코딩 및 데이터 압축 기술의 최적화 방안을 다루었다. VCM(Video Coding for Machines)은 기계 학습 및 비전 시스템을 위한 비디오 코딩 표준으로, 데이터를 인간이 소비하는 대신 기계가 처리할 수 있도록 최적화하는 데 중점을 둔다. FCM(Feature Coding for Machine)은 비디오 데이터에서 추출된 특징을 효율적으로 압축하여 기계 학습 성능을 최적화하는 기술이다.

‘VCM을 위한 공간적 재표본화 기술 개선’은 MPEG의 VCM 표준화 과정에서 효율적인 데이터 압축을 위해 공간적 재표본화 기술을 적용하고 개선 방안을 제시하였다. MI-RPR 기술을 활용한 새로운 재표본화 방식이 데이터 압축률과 기계 성능을 동시에 향상시키는 결과를 보였다. ‘선택적 학습기법을 통한 FCM 시험모델 성능개선’은 FCM 표준의 시험모델인 FCTM의 성능을 개선하기 위해 선택적 학습 기법(SLS)과 QP-적응적 채널 절삭 기법을 도입하였다. 제안된 기법은 기존 대비 최대 35.31% BD-rate 절감 효과를 보여, 머신 비전용 데이터 압축 성능을 크게 향상시켰다. ‘관심 영역 기반 영상 압축에서 배경영역이 머신 비전 성능에 미치는 영향’은 관심 영역 리타겟팅 기술을 활용하여 배경 정보를 줄이는 압축 방식이 머신 비전 성능에 미치는 영향을 분석하였다. Pandaset 데이터셋을 활용한 실험에서 제안된 방식이 특정 작업에서는 기존 방법보다 높은 성능을 보임을 확인하였다. 본 세션은 VCM 및 FCM 표준화와 관련된 최신 기술 개발 동향을 논의하며, 비디오 데이터 압축과 기계 학습 응용의 효율성을 높이기 위한 다양한 접근 방안을 제시하였다.

5. 특별 세션 5: 원격헬스케어기술

특별 세션 5는 ‘원격헬스케어기술’을 주제로 총 7편의 논문이 발표되었다. 광운대학교/서울대학교병원(1편), 광운대학교/LG전자(1편), 광운대학교(5편)가 참여하였으며, 발표는 원격의료 및 헬스케어기술의 발전과 응용 가능성을 중심으로 이루어졌다.

‘두개골 조기 유합증 진단을 위한 세그먼테이션 모델 성능 비교’는 두개골 조기 유합증을 진단하기 위



해 2D X-ray 이미지를 활용한 세그멘테이션 모델의 성능을 비교하였다. SwinUNet, TransUNet, UNet 세 가지 모델을 평가하였으며, 정확도와 처리 시간을 기준으로 최적의 진단 방법을 도출하여 방사선 노출을 줄이고 접근성을 높이는 가능성을 제시하였다. ‘Research Trend on Foreign Language Anxiety’는 외국어 학습에서 불안감(FLA)이 학습 성과에 미치는 영향을 검토하고, 전통적 및 기술 기반 치료 접근법을 논의하였다. 특히, 가상현실(VR)을 활용한 치료법의 가능성을 강조하며 향후 연구 방향을 제안하였다. ‘LLaMA3.2 모델의 비균일 양자화를 위한 인접 파라미터 분포 특성 분석’은 LLaMA3.2 모델에서 가중치 전용 양자화에 적합한 인접 파라미터의 분포 특성을 분석하였다. 분석 결과를 기반으로 양자화 및 모델 압축 가능성을 확인하며, 저자원 환경에서의 모델 최적화 방안을 제안하였다. ‘얼굴 랜드마크 이동 정보를 활용한 안면 근전도 신호 예측 모델 개발’은 얼굴 랜드마크 데이터를 활용하여 비접촉식 안면 근전도(EMG) 신호를 예측하는 모델을 개발하였다. Random Forest 모델이 가장 우수한 성능을 보였으며, 감정 인식 및 HCI 시스템에서의 응용 가능성을 제시하였다. ‘다중 카메라 기반 가상인간의 생성을 위한 wrapping 기반 리토폴로지 기법’은 MediaPipe를 활용한 다중 카메라 기반 가상 인간 생성 기법을 제안하였다. 랜드마크 데이터를 3D 대응점으로 매핑하여 고해상도 메쉬를 생성하며, 모션 캡처 및 애니메이션과 같은 실시간 응용에 적합한 효율적인 파이프라인을 구축하였다. ‘비디오 기반 동적 메쉬 압축을 위한 효과적인 움직임 벡터 예측자 결정 방법’은 동적 메쉬 부호화를 위해 움직임 벡터 예측자를 효율적으로 선택하는 방법을 제안하였다. 기존 방법과 비교하여 BD-rate 성능을 개선하며, MPEG V-DMC 참조 소프트웨어를 기반으로 한 효과적인 압축 가능성을 입증하였다. ‘빅데이터 의미연결망 분석을 활용한 인공지능(AI) 인식 연구’는 2014년부터 2023년까지 인공지능 키워드를 중심으로 빅데이터를 분석하여 주요 기술과 응용 영역의 변화를 파악하였다. 연결 중심성 및 CONCOR 분석을 통해 인공지능의 사회적 인식과 기술적 동향을 시각적으로 표현하였다. 본 세션은 비접촉식 의료 기술, 데이터 분석, 그리고 가상 인간 생성과 같은 다양한 주제를 통해 원격 의료와 헬스케어 기술 및 실용적 응용 가능성을 논의하였다.

일반 세션

일반 세션에서는 ‘신호처리 및 미디어 기술’, ‘인공지능, 지능형 미디어 서비스’, ‘딥러닝, 신경망 기술 및 인지과학’이라는 3가지 주제에 대하여 총 13편의 논문이 발표되었다.

1. 일반 세션 1: 신호처리 및 미디어 기술

‘신호처리 및 미디어 기술’ 세션에서는 총 4편의 논문이 구두 발표되었는데, 한국항공대학교/ETRI(1편), 한국항공대학교(1편), ETRI(1편), 경희대학교(1편)에서 발표하였다. 발표는 데이터 압축, 신호처리 효율성 향상, 그리고 미디어 응용 기술에 대한 연구들로 구성되었다.

‘블록벡터 기반 화면내 예측의 예측모드 매핑을 통한 ECM의 성능개선’은 차세대 비디오 부호화 표준을 위한 ECM(Enhanced Compression Model)에서 사용되는 화면 내 예측 기술을 개선하였다. 기존 OBIC(Occurrence-Based Intra Coding) 기법에 블록벡터 기반 예측모드를 매핑하여 더 많은 예측 정보를 활용함으로써 BD-rate를 -0.02% 감소시키는 성능 개선을 달성하였다. ‘저지연 고품질 동시지원을 위한 오디오 부호화 기술’은 실시간 통신을 위한 오디오 부호화 기술을 제안하였다. 프레임 간 스위칭 없이 단일 프레임 기반 양자화 기술을 적용하여 50ms 이하의 지연 시간에서 96kbps 스테레오 신호에 대해 MOS(Mean Opinion Score) 4.5의 고품질 압축 성능을 보여주었다. ‘가상 인플루언서의 특성, 모델-제품 적합성, 제품태도, 구매의도, 구전의도와외의 관계에 관한 연구’는 가상 인플루언서의 전문성, 신뢰성, 친밀도와 같은 특성이 소비자의 제품 태도와 구매 의도에 미치는 영향을 분석하였다. 특히, 모델-제품 적합성이 높을수록 긍정적인 소비자 반응을 이끌어내는 것으로 나타나 가상 인플루언서를 활용한 광고 마케팅 전략 수립에 실질적인 정보를 제공하였다. ‘NNCodec을 활용한 LightGaussian 기반의 3DGS 모델 압축 성능분석’은 3D Gaussian Splattering(3DGS)을 위한 LightGaussian 압축 기법과 NNCodec을 조합하여 3D 데이터의 압축률과 렌더링 화질을 분석하였다. 특히, QP(Quantization Parameter) 조정을 통해 압축 효율성과 화질의 균형을 최적화할 가능성을 확인하였다. 위와 같은 발표를 통하여 신호처리와 미디어 동향을 공유하며 실질적인 응용 가능성을 논의하였다.

2. 일반 세션 2: 인공지능, 지능형 미디어 서비스

‘인공지능, 지능형 미디어 서비스’ 세션에서는 총 6편의 논문이 구두 발표되었는데, 경성대학교(2편), 광운대학교(1편), 한양여자대학교(2편), KAIST/ETRI(1편)에서 발표하였다. 발표 논문들은 개인화된 미디어 서비스를 위한 딥러닝 모델 개발, AI 기반 데이터 처리 기술, 그리고 지능형 미디어 추천 시스템에 초점을 맞췄다.

‘강수 예측을 위한 convLSTM 네트워크에서 팽창 커널의 영향 조사’는 다양한 팽창 커널 비율(2,3,4)이 convLSTM 네트워크 성능에 미치는 영향을 분석하였다. 18프레임을 기반으로 3시간 후의 강수량을 예측한 결과, 높은 팽창 커널 비율이 복잡한 시공간 의존성을 잘 포착하며 예측 정확도를 향상시키는 데 기여함을 보여주었다. ‘보안영상에 대한 통계적 분석에 기반한 군중 시뮬레이션’은 Unity 엔진에서 Gaussian 분포를 활용해 군중 에이전트를 생성하고, 영상 데이터를 기반으로 통계적 매개변수를 추출하여 시뮬레이션 환경을 구축하였다. 연구는 현실적인 에이전트 배치와 충돌 회피 알고리즘을 통해 군중의 자연스러운 행동을 재현하였다. ‘3D 페이스 생성 모델의 실시간 동작을 위한 시스템의 구현’은 3D 얼굴 생성 기술을 실시간으로 구현하기 위해 FPGA를 활용한 딥러닝 모델 가속기를 개발하였다. 본 연구는 VR, 게임, 영화 산업에서의 상호작용과 몰입도를 높이는 데 기여할 수 있는 실시간 3D 페이스 생성 디바이스를 제안하였다. ‘트랜스포머 기반 비디오 표정 인식 연구’는 TimeSformer, MViT, Video Swin Transformer 등 트랜스포머 기반 모델을 비디오 얼굴 표정 인식에 적용하였다. 실험 결과, Video Swin Transformer가 가장 우수한 성능을 보이며 비디오 데이터 기반 표정 인식의 가능성을 보여주었다. ‘공간 음향 렌더링용 심층 신경망 훈련을 위한 오디오 데이터 세트 구축’은 공간 음향 렌더링을 위한 다중 오디오 데이터 세트를 구축하였다. 52.5시간 분량의 오디오 데이터와 13가지 음원 유형을 활용하여 음원 분리 및 방향 탐지와 같은 여러 작업을 동시에 수행할 수 있도록 설계되었다. ‘얼굴 영상의 표정이 나이 인식에 미치는 영향 분석’은 표정 변화가 얼굴 나이 인식에 미치는 영향을 분석하였다. 실험 결과, 무표정, 놀람, 두려움, 행복과 같은 긍정적인 표정에서 높은 인식 정확도를 보였으며, 슬픔, 역겨움, 화남과 같은 표정에서는 인식 정확도가 감소하는 경향을 확인하였다. 본 세션은 인공지능 기술을 활용한 미디어 서비스와 데이터 처리 기술의 발전 가능성을 논의하였으며, AI를 통해 개인화된 추천, 공간 음향 처리, 얼굴 표정 인식 등 다양한 응용 사례가 발표되었다.

3. 일반 세션 3: 딥러닝, 신경망 기술 및 인지과학

‘딥러닝, 신경망 기술 및 인지과학’ 세션에서는 총 3편의 논문이 구두 발표되었는데, 광주과학기술원(1편), 광운대학교(1편), 전남대학교(1편)에서 발표하였다. 주요 발표는 딥러닝 모델의 구조 개선, 신경망 기반 이미지 처리, 그리고 인지과학적 접근을 활용한 딥러닝 응용 방안을 다루었다.

‘Streaming SEANet을 이용한 실시간 음성 코덱 후처리 모델 연구’는 Streaming SEANet을 이용해 압축된 음성 신호의 품질을 개선하는 실시간 음성향상 모델을 제안하였다. PESQ 점수를 사용한 평가에서 기존 Opus 광대역 음성 압축 시스템 대비 유의미한 품질 향상을 확인하였으며, 가중치 정규화, 편별자, 손실 함수의 조합을 통해 최적 모델 구성을 탐구하였다. ‘RGB 입력을 통한 CNN 기반 DVS 생성 및 행동 분류 성능 향상 연구’는 RGB 데이터를 CNN 기반 DVS(Event) 데이터로 변환하여 행동 인식의 성능을 향상시키는 방법론을 제안하였다. 실험 데이터는 DAVIS346 장비를 사용해 수집되었으며, CNN 학습과 노이즈 필터링을 통해 생성된 이벤트 데이터를 행동 인식 분류 모델에 적용한 결과, 행동 분류 정확도의 유의미한 향상을 확인하였다. ‘생명에너지를 이용한 음악치유가 펼쳐지는 물리적 현상’은 스피커에서 발생

하는 생명에너지와 음파를 활용한 새로운 음악치유의 원리를 제시하였다. 실험을 통해 생명에너지가 특정 신체 장부에 미치는 영향을 측정하였으며, 음악치유의 새로운 접근 방식과 그 연구 방향을 논의하였다. 본 세션은 딥러닝 및 신경망 기술의 최적화와 이를 활용한 응용 사례를 공유하며, 실시간 처리, 이미지 기반 행동 인식을 논의하였다.

포스터 세션

포스터 세션은 총 5개로 구성되었으며, 이 중 3개는 일반 포스터 세션, 1개는 대학생 논문 경진대회, 1개는 대학생/주니어 논문 경진대회로 이루어졌다. 각 세션에서는 다양한 주제와 연구 결과를 발표하며 학술 교류의 장을 마련하였다.

1. 포스터 세션 1

포스터 세션 1에서는 총 17편의 논문이 발표되었다. 참여 기관으로는 ETRI(2편), KBS/현대모비스(1편), 광운대학교(3편), 광운대학교/(주)아르고(1편), 서울대학교(1편), 전남대학교(1편), 충북대학교(1편), 한양대학교(4편), 한화시스템(1편), 홍익대학교(1편), 홍익대학교/(주)시어스랩(1편)이 포함되었다.

차세대 기술과 사용자 경험(UX), 고해상도 영상 처리, 크로스 플랫폼 지원, 데이터캐스팅 송수신 시스템 등 다양한 연구가 논의되었다. ‘XR 환경에서 모달 간의 상관관계를 반영한 UX 설계 프레임 연구’는 AR 안경을 사용할 때의 감각 모달 간 상관관계를 분석하여 사용자 경험에 영향을 미치는 요인을 도출하였다. 이를 통해 감각적 갈등을 최소화하고 모달 간 시너지를 강화하는 설계 전략을 제안하였다. 연구는 사용자 중심의 직관적이고 몰입감 있는 AR 경험 설계에 기여할 수 있는 멀티모달 UX 프레임워크를 제공하였다. ‘비디오 고해상화 기법을 이용한 기하학적 위상 자가간섭 디지털 홀로그래피의 해상도 증대 방법’은 딥러닝 기반 비디오 고해상화 기법을 통해 기존 2K 수준의 디지털 홀로그램 재현영상을 8K로 확장하는 기술을 제안하였다. 잡음 제거를 위한 추가적인 딥러닝 기법을 사용하여 화질을 개선하고, 고해상도 재현영상을 생성하였다. 연구는 데이터셋 생성 과정과 모델 학습 결과를 통해, 디지털 홀로그램의 재현 성능을 크게 향상시키는 방법을 소개하였다. ‘크로스 플랫폼 지원을 위한 음성 기반 딥러닝 모델의 재구성 방법’은 음성 신호 기반 3D 얼굴 애니메이션 모델인 Emotalk을 경량화하고 FPGA와 같은 자원 제한 환경에서도 동작하도록 재구성하였다. PyTorch 기반의 원본 모델을 NumPy로 전환하고, 소켓 프로그래밍을 통해 FPGA에서 후속 처리를 수행하는 구조를 설계하였다. 이러한 경량화는 다양한 플랫폼에서 높은 품질의 애니메이션을 구현할 수 있는 가능성을 제시했다. ‘ATSC 3.0 기반 HP-GNSS 데이터캐스팅 송수신 시스템의 구현’은 방송망을 활용하여 정밀한 RTK 위치 정보를 제공하는 시스템을 구현하였다. ATSC

3.0 지상파 데이터캐스팅 기술을 통해 기존 상용 GNSS 대비 데이터 전송 효율과 정밀도를 개선했다. 연구는 GNSS 수신기와의 성능 비교를 통해 방송망 RTK 시스템의 활용 가능성을 입증하였다. 제시된 논문들 외에도 데이터 기반 건축 평면도 변환, 딥러닝을 활용한 객체 인식 및 탐지, RGB-D 기반 측정 시스템, 비전 트랜스퍼머 모델의 활용 등 다양한 첨단 기술 연구가 다뤄졌다. 사용자 경험(UX) 최적화, 비디오 고해상화, 음성 신호 기반 애니메이션 생성, 정밀 위치 시스템 등의 응용 가능성이 높은 연구들이 포함되어, 차세대 기술의 발전 방향을 제시했다.

2. 포스터 세션 2

포스터 세션 2는 총 8편의 논문이 발표되었다. 참여 기관으로는 광주과학기술원(1편), 서울과학기술대학교(1편), 서울대학교(2편), 시그마케이주식회사/서울과학기술대학교(1편), 이화여자대학교(1편), 특허청(1편), 한국생산기술연구원(1편)이 포함되었다.

‘소형 장치에 탑재할 수 있는 딥러닝 기반 실시간 음성 향상 모델 개발’은 CRN(Convolutional Recurrent Network) 모델을 기반으로 제한된 하드웨어에서 실시간 음성 향상을 구현하였다. 잡음 환경에서의 음질 개선을 위해 공기 전도와 골전도 데이터를 활용하였으며, 모델의 크기와 연산량을 줄여 소형 장치에서도 효과적으로 작동하도록 설계되었다. 성능은 PESQ와 STOI 지표로 평가되었으며, 기존 대비 작은 모델 크기와 연산량에도 불구하고 높은 음질 개선 효과를 보였다. ‘반전 라플라시안 정보와 저노출 맵을 활용한 저조도 이미지 화질 개선’은 TriFusionNet이라는 네트워크를 통해 저조도 환경에서 촬영된 이미지의 화질을 개선하는 기술을 제안하였다. 이 네트워크는 이미지의 텍스처 복원을 위해 반전 라플라시안 정보를, 어두운 영역의 복원을 위해 저노출 맵 정보를 활용하였다. 실험 결과, LOL-v1 데이터셋에서 기존 RetinexFormer 모델 대비 평균 0.95dB의 PSNR 향상을 기록하며 뛰어난 성능을 입증했다. 음성 향상, 이미지 화질 개선 외에도, 비편향적 그래프 생성, AI 보호 기술, 제로-샷 학습 기반 분류 기술 등 딥러닝 기술 관련 발표가 이루어졌다.

3. 포스터 세션 3

포스터 세션 3은 총 13편의 논문이 발표되었다. 참여 기관으로는 ETRI(3편), KETI(2편), KETI/전남대학교(1편), 광운대학교(1편), 광운대학교/(주)아르고(1편), 덕성여자대학교(2편), 서울과학기술대학교(1편), 이화여자대학교(2편)가 포함되었다.

‘플렉서블 cspace 기반 다중 앰비소닉 렌더링 방법에 대한 연구’는 MPEG-I Immersive Audio의 CD (Committee Draft)에서 제공하는 재생 기술 중 많은 수의 채널로 구성된 HOA 신호가 여러 개 존재할 때 청취자를 중심으로보다 효과적으로 음장 왜곡 없이 몰입감 높은 음장을 제공하는 방법에 대해서 논한다.

‘얼굴 인식 시스템의 대규모 인원 수용을 위한 임베딩 벡터 보강 및 탐색 알고리즘 연구’는 대규모 인원에 대한 얼굴 인식 시스템에 나이, 성별, 인종 등을 포함하는 것이 미치는 영향을 조사하고, 대용량 데이터를 효율적으로 처리하는 방법론을 제시한다. ‘깊이 정보와 3DGS 기반 가상 시점 영상 생성에서 노말 정보의 적용 방법’은 GPS-Gaussian 등과 같이 입력 영상으로부터 깊이 정보를 추정하고 이를 바탕으로 3D gaussian의 다른 파라미터들을 추정함으로써 3D 장면을 표현하고 이로부터 가상 시점 영상을 생성하는 방법의 가상 시점 영상의 품질을 향상시키는 방법을 제안한다. ‘음악 샘플 데이터셋을 활용한 악기 분류 기술 개발’은 상업적 음악 샘플을 기반으로 멜스펙트로그램을 포함한 악기 분류 기술을 개발한 내용을 담고 있다. ‘사전학습된 이미지-언어 모델을 활용한 효율적인 비디오 시간적 연계성 학습’은 대규모 언어 모델의 한계를 보완하고자 이미지-언어 모델인 BLIP-2에 프레임 간 어텐션 메커니즘을 추가한 변형 모델을 제안하였다. 비디오 클러스터링을 통해 핵심 프레임 추출하고 QLoRA 기법으로 단일 GPU에서도 학습이 가능하도록 최적화하였다. 실험 결과, IoU 평가 지표에서 평균 4~7%의 성능 향상을 보이며 비디오와 텍스트 간의 시간적 맥락과 의미적 연계성을 강화하는 데 성공하였다. 실감 콘텐츠 생성, 효율적인 데이터 처리, 몰입형 음향 기술 등 폭넓은 주제를 다루었다. 이외에도 ‘생성형 AI를 활용한 홀로그램 프린팅 기술’, ‘이벤트 기반 고속 사전 객체 영역 추출 방법’과 같은 다양한 연구들이 논의되었다.

4. 포스터 세션 4: 대학생 논문 경진대회

포스터 세션 4는 ‘대학생 논문 경진대회’로서 총 15편의 논문이 출품되었다. 참여 기관으로는 경북대학교/(주)신라시스템(1편), 국립창원대학교(1편), 국립창원대학교/인도공과대학(1편), 국립한밭대학교(1편), 덕성여자대학교(1편), 동양미래대학교(1편), 서울과학기술대학교(8편), 서울여자대학교(1편)가 포함되었다.

학부생들의 연구 활동을 고취하고, 향후 석사 및 박사 과정으로의 심화 연구로 자연스럽게 유도하기 위해, 학부생 연구원으로서 소속 대학의 연구실에서 수행하여 도출한 연구 성과들을 발표할 수 있는 세션이었다. 이 세션에서는 대학생 연구자들이 창의적인 아이디어를 바탕으로 다양한 연구를 발표하였다. ‘장애인(농인)을 위한 의료 수어 인식 인공지능 에이전트 설계 및 구현’ 논문은 의료기관에서 농인들이 겪는 의사소통 문제를 해결하기 위해 개발된 수어 통역 에이전트를 소개한다. MediaPipe를 활용하여 손과 포즈 랜드마크 데이터를 추출한 후, LSTM 모델을 이용해 수어 단어를 인식한다. 이후 ChatGPT와 TTS 기술을 사용해 인식된 단어를 자연어 문장으로 변환하고 이를 음성으로 전달한다. 최종적으로 TensorFlow Lite를 통해 Android 장치에서도 동작할 수 있도록 구현하였으며, 90% 이상의 인식 정확도를 기록하며 유의미한 성과를 보였다. ‘컨텐츠 손실을 이용한 초해상도 기법 기반의 객체 인식’ 논문은 자율주행 시스템에서 다양한 주행 시나리오에 대응하기 위한 데이터를 확보하는 데 어려움을 해결하기 위해 SRResNet 기반의 초해상도 기법과 YOLOv5 객체 탐지 모델을 결합한 파이프라인을 제안한다. Content Loss를 최적화에 활용하여 객체 인식 정확도를 높였으며, 특히 노이즈와 블러가 포함된 환경에서도 안정적인 성능

을 보여 자율주행 기술의 실용성을 제고하였다. 이외에도 스마트 디바이스와 AI를 활용한 헬스케어 솔루션, 원격탐사 및 데이터 스트리밍 기술, 효율적인 데이터 압축 알고리즘, 그리고 AI 기반 문제 해결 모델 제작 등의 연구가 포함되어 있어, 참가자들이 향후 학술적 성과를 통해 관련 분야에 기여할 수 있는 잠재력을 보여주었다.

5. 포스터 세션 5: 대학생/주니어 논문 경진대회

포스터 세션 5는 ‘대학생/주니어 논문 경진대회’로서 총 12편의 논문이 출품되었다. 참여 기관으로는 국립한밭대학교(1편), 국립한밭대학교/경희대학교(1편), 동아방송예술대학교(5편), 부산일과학교등학교/경성대학교(1편), 서울과학기술대학교(4편)가 포함되었다.

‘멀티모달 모델과 이미지 분할 모델을 활용한 지시 기반 이미지 매칭을 위한 Trimap 생성 방법’에서는 이미지 내 사용자 지시 객체를 효과적으로 매핑하기 위해 멀티모달 모델과 이미지 분할 모델을 활용한 Trimap 생성 방법을 제안하였다. 이 방법은 텍스트를 제외한 최소한의 사용자 개입으로 다수 객체가 포함된 이미지에서도 높은 정확도의 Trimap 생성을 가능하게 하며, 기존 이미지 매칭 모델의 한계를 극복하였다. ‘생성형 AI를 활용한 3D 스캐닝을 기반의 가상 환경 제작에 관한 연구’는 역사 문화 자원의 장소성을 중심으로 생성형 AI와 3D 스캐닝 기술의 결합을 통해 시각적 품질이 높은 가상 환경을 제작하는 방안을 제시하였다. 기존 포토그래메트리 방식의 한계를 극복하며, 일관성 있고 사실적인 배경 공간 생성을 통해 문화 자원의 디지털 재현 가능성을 높였다. 본 세션은 대학생과 주니어 연구자들이 멀티모달 기술, 생성형 AI 기반 가상 환경 제작, UAV 위치 선정 최적화, 실시간 스트리밍, 포터블 중계 시스템, 3D 데이터 압축, 강화학습, 그리고 신경망 비디오 압축 등의 다양한 분야의 실질적인 문제를 해결하기 위한 창의적인 접근과 다양한 기술적 도전을 통해 독창적이고 실용적인 연구 성과를 보여주었다.





맺음말

이번 한국방송·미디어공학회 2024년 추계학술대회는 매우 성공적으로 개최되었고, 중요한 몇 가지의 미래를 남겼다. 이번 학술대회는 인공지능(AI), 5G, 초실감형 콘텐츠, 메타버스 등 최신 기술들이 방송과 미디어 산업에 융합되는 과정에서 발생하는 혁신적인 변화들을 깊이 있게 다룬 자리가 되었다. 방송과 미디어는 단순한 정보 전달을 넘어, 사용자에게 몰입감 있는 경험을 제공하는 새로운 산업으로 발전하고 있으며, 이 학술대회는 이러한 변화의 흐름을 반영하고, 관련 연구 성과를 공유하는 중요한 플랫폼이 되었다. 이를 통해 향후 방송과 미디어 기술의 발전 방향과 새로운 비즈니스 모델을 제시할 수 있었으며, 산업계와 학계의 상호 협력의 장을 마련하였다. 또한 이번 학술대회에서는 대학생 및 고등학생을 위한 논문 경진대회가 열려, 젊은 인재들이 창의적 아이디어를 발표하고 전문가들로부터 피드백을 받으며 성장할 수 있는 기회를 제공하였다. 이는 미래 방송·미디어 산업을 이끌어 나갈 인재들이 실제 연구 현장에서 실력을 쌓고, 학계와 산업계의 전문가들과 교류할 수 있는 중요한 발판이 되었다. 더불어, 대학생 및 주니어 연구자들이 산업 현장의 최신 기술과 동향을 직접 경험하면서, 더 나아가 진로 탐색과 연구 개발에 대한 방향성을 설정할 수 있는 기회를 제공한 점에서 의의가 크다. 마지막으로 이번 대회에서 발표된 기술들은 방송과 미디어 산업의 미래를 여는 중요한 키워드들이다. 생성형 AI, 5G 기반 방송 전송, 블록체인 보안 등은 기존의 방송 환경을 넘어서는 혁신적 변화들을 가능하게 한다. 이를 통해 학계와 산업계의 전문가들이 앞으로 방송과 미디어 기술이 나아갈 방향을 구체적으로 논의하며, 새로운 비전과 전략을 제시하였다. 또한, AI와 XR 기술 등 새로운 기술들이 방송 콘텐츠 제작과 소비의 패러다임을 바꾸어, 미래 미디어의 발전을 이끌어갈 중요한 기초가 될 것이다.

