

FBMF Technical Report

기술보고서

FBMF-TR-019

제정일 2024. 11. 29.

AI 기반 메타데이터 추출 및 비디오
화질 개선(기술 보고서)

AI-based metadata extraction and
video quality improvement
(Technical Report)



기술보고서 초안 검토 위원회
기술보고서안 심의 위원회

방송 클라우드 분과 위원회
운영 위원회

	성명	소 속	직위	위원회 및 직위
기술보고서(과제) 제안	박경모	CAST.ERA	CTO	방송 클라우드 분과 위원장
기술보고서 초안 작성자	박성환	CAST.ERA	매니저	방송 클라우드 분과 위원

사무국 담당

함상진

KBS

미래방송표준포럼
사무총장

본 문서에 대한 저작권은 미래방송미디어표준포럼에 있으며, 미래방송미디어표준포럼과 사전 협의 없이 이 문서의 전체 또는 일부를 상업적 목적으로 복제 또는 배포해서는 안 됩니다.

본 표준 발간 이전에 접수된 지식재산권 확약서 정보는 본 표준의 ‘부록(지식재산권 확약서 정보)’에 명시하고 있으며, 이후 접수된 지식재산권 확약서는 미래방송미디어표준포럼 웹사이트에서 확인할 수 있습니다.

본 표준과 관련하여 접수된 확약서 외의 지식재산권이 존재할 수 있습니다.

발행인 : 미래방송미디어표준포럼 의장

발행처 : 미래방송미디어표준포럼

06130, 서울특별시 강남구 테헤란로 7길 22 신관 1108호

Tel : 02-568-3556, Fax : 02-568-3557

발행일 : 2024.11

서 문

1 기술보고서의 목적

이 기술 보고서의 목적은 "AI 기반 메타데이터 추출 및 비디오 화질 개선" 기술에 대한 현황을 분석하고, 대한민국과 글로벌 동향을 비교함으로써 연구 기획 부서가 연구 과제 발굴과 기획을 위한 사전 조사를 하는 데 있다. 이를 통해 AI 기술의 발전과 이를 적용한 메타데이터 추출 및 비디오 화질 개선 기술의 현재 상태를 파악하고, 향후 발전 방향을 제시하고자 한다.

2 주요 내용 요약

이 보고서는 다음과 같은 주요 내용을 포함하고 있다. 첫째, AI 기술의 발전 현황과 이를 적용한 메타데이터 추출 기술에 대해 분석한다. 둘째, AI 기반 비디오 화질 개선 기술의 발전과 적용 사례를 조사한다. 셋째, 대한민국과 글로벌 동향을 비교하여, 각 지역의 기술 개발 및 서비스 출시 현황을 살펴본다. 마지막으로, 이러한 기술들이 미디어 서비스 산업에 미치는 영향을 분석하고, 향후 발전 가능성을 논의한다.

3 인용 표준과의 비교

해당사항 없음

Preface

1 Purpose

The purpose of this technical report is to analyze the current status of "AI-based Metadata Extraction and Video Quality Improvement" technology and compare the trends in South Korea and globally. This report aims to assist the research planning department in preliminary investigation for discovering and planning research projects. By understanding the current state of AI technology and its application in metadata extraction and video quality improvement, the report seeks to present future development directions.

2 Summary

This technical report includes the following main contents. First, it analyzes the current state of AI technology and its application in metadata extraction. Second, it investigates the development and application cases of AI-based video quality improvement technology. Third, it compares the trends in South Korea and globally to examine the technology development and service launch status in each region. Finally, it analyzes the impact of these technologies on the media service industry and discusses potential future developments.

3 Relationship to Reference Standards

N/A

목 차

<제 목 차 례>

1	적용 범위	1
2	인용 표준	1
3	용어 및 약어 정의	1
4.	AI 기반 메타데이터 추출	3
4.1.	개요	3
4.2.	논문 및 연구 사례	4
4.3.	기술 동향	8
4.4.	미래 전망	10
5.	AI 기반 비디오 화질 개선	11
5.1.	개요	11
5.2.	논문 및 연구 사례	11
5.3.	기술 동향	14
5.4.	미래 전망	17
6.	결론	18
부 록 1-1	지식재산권 요약서 정보	19
부 록 1-2	시험인증 관련 사항	20
부 록 1-3	본 기술보고서의 연계(family) 표준	21
부 록 1-4	참고 문헌	22
부 록 1-5	영문기술보고서 해설서	24
부 록 1-6	기술보고서의 이력	25

AI 기반 메타데이터 추출 및 비디오 화질 개선(기술보고서) (AI-based metadata extraction and video quality improvement (Technical Report))

1 적용 범위

본 기술 보고서의 적용 범위는 인공지능(AI) 기술을 활용한 메타데이터 추출 및 비디오 화질 개선 기술에 국한되며 다음의 사항을 포함한다:

- 메타데이터 추출: 비디오 콘텐츠에서 유의미한 정보를 자동으로 분석하고 추출하는 기술. 이는 AI 알고리즘을 통해 텍스트, 이미지, 음성 데이터를 분석하고 메타데이터를 생성하는 과정을 포함한다.
- 비디오 화질 개선: AI 기술을 이용해 저해상도 비디오를 고해상도로 변환하거나, 기존 비디오의 화질을 향상시키는 기술. 이는 딥러닝, 머신러닝 모델을 활용하여 노이즈 제거, 디인터레이싱, 업스케일링 등 다양한 화질 개선 작업을 수행하는 과정을 포함한다.
- 대한민국 및 글로벌 동향 분석: 대한민국과 글로벌 시장에서 AI 기반 메타데이터 추출 및 비디오 화질 개선 기술의 개발 현황, 연구 동향 및 상용 서비스 사례를 분석한다.
- 기술 개발 및 서비스 현황: 현재 개발 중인 기술 및 이미 상용화된 서비스에 대한 현황을 파악하여, 해당 기술의 발전 방향을 제시한다.

이 보고서는 방송 및 미디어 산업에서 AI 기술의 적용 가능성을 탐색하고, 해당 분야의 연구 기획 및 개발 전략 수립에 참고 자료로 활용될 수 있다.

2 인용 표준

해당사항 없음

3 용어 및 약어 정의

AI	Artificial Intelligence
AR	Augmented Reality
CLIP	Contrastive Language-Image Pre-training
DMM	Deep Meta Manager
GEM	Gradient Episodic Memory

Grad-CAM	Gradient-weighted Class Activation Mapping
IoT	Internet of Things
LLM	Large Language Model
NAS	Network Architecture Search
OCR	Optical Character Recognition
VATT	Video Audio Text Transformer
VR	Virtual Reality

4. AI 기반 메타데이터 추출

4.1. 개요

AI 기반 메타데이터 추출 기술은 비디오 콘텐츠에서 자동으로 중요한 정보를 추출하고 구조화하는 기술이다. 이 기술은 컴퓨터 비전, 자연어 처리, 머신러닝 등 다양한 AI 기술을 활용하여 비디오의 내용을 심층적으로 분석하고 이해한다. 주요 구성 요소로는 객체 인식, 얼굴 인식, 음성 인식, 자막 추출, 장면 분할, 감정 분석, 행동 인식, 주제 추출 등이 있다.

이 기술의 작동 원리는 비디오 입력, 전처리, 병렬 처리, 데이터 통합, 메타데이터 생성, 후처리의 단계로 이루어진다. 먼저 비디오 파일이 시스템에 입력되면, 이를 프레임 단위로 분할하고 오디오 트랙을 추출하는 전처리 과정을 거친다. 그 후 각 구성 요소가 동시에 작동하여 데이터를 추출하는 병렬 처리 단계를 거치고, 추출된 정보를 시간 정보와 함께 통합한다. 이렇게 통합된 정보는 구조화된 형태의 메타데이터로 변환되며, 마지막으로 생성된 메타데이터를 검증하고 필요시 수정하는 후처리 과정을 거친다.

AI 기반 메타데이터 추출 기술은 방송 및 미디어, 보안 및 감시, 마케팅, 교육, 의료 등 다양한 분야에서 활용되고 있다. 방송 및 미디어 분야에서는 콘텐츠 관리, 검색 기능 향상, 개인화된 추천 시스템 구축 등에 사용된다. 보안 및 감시 분야에서는 CCTV 영상 분석, 이상 행동 감지 등에 활용된다. 마케팅 분야에서는 소비자 행동 분석, 광고 효과 측정 등에 사용되며, 교육 분야에서는 비디오 기반 학습 자료의 자동 인덱싱 및 요약에 활용된다. 의료 분야에서는 의료 영상 분석 및 진단 보조 도구로 사용된다.

이 기술을 통해 콘텐츠 관리 효율성 증대, 검색 기능 향상, 사용자 경험 개선, 비즈니스 인사이트 도출 등의 효과를 기대할 수 있다. 대량의 비디오 콘텐츠를 빠르고 정확하게 분류하고 관리할 수 있으며, 상세한 메타데이터를 바탕으로 정확하고 세밀한 비디오 검색이 가능해진다. 또한 개인화된 콘텐츠 추천과 더 나은 탐색 기능을 제공할 수 있으며, 비디오 콘텐츠에서 유용한 비즈니스 정보를 자동으로 추출할 수 있다.

AI 기반 메타데이터 추출 기술은 비디오 콘텐츠의 가치를 극대화하고, 사용자들에게 더 나은 경험을 제공하는 핵심 기술로 자리잡고 있다. 앞으로 더욱 발전된 AI 기술과의 결합을 통해 그 활용 범위와 중요성은 계속해서 확대될 것으로 전망된다. 특히, 5G 네트워크의 확산, 8K 해상도의 상용화, VR/AR 기술의 발전 등과 맞물려 AI 기반 메타데이터 추출 기술의 중요성은 더욱 커질 것으로 예상된다.

4.2. 논문 및 연구 사례

4.2.1. CAST.ERA Media Meta Service

CAST.ERA Media Meta Service는 비디오 콘텐츠에서 자동으로 메타데이터를 추출하고 관리하는 시스템을 개발했다. 이 시스템은 객체 인식, 얼굴 인식, 텍스트 인식 등 다양한 AI 기술을 활용하여 비디오의 내용을 분석한다. 주요 기능으로는 미디어 목록 관리, 객체 사전 감지, 레이블 편집, 비디오 클리핑 등이 있다.

시스템의 전체 처리 흐름을 살펴보면, 사용자가 웹 서버에 연결하여 주석 작업을 시작하면 Annotator가 Deep Meta Manager(DMM)에 현재 미디어 목록을 요청한다. 사용자가 주석을 달 미디어를 선택하면, 이미 주석이 있는 경우 DMM이 현재 주석 목록을 제공한다. 주석이 없는 경우, 사용자는 DMM에 자동 주석 생성을 요청한다. DMM은 이 요청을 Deep Meta Extraction Server(DMES)로 전달하고, DMES는 CMS에서 요청된 미디어를 다운로드하여 프레임별로 분석한다. 분석 결과는 DMM을 통해 Annotator로 전달되며, 사용자는 이를 검증하고 수동으로 편집할 수 있다.

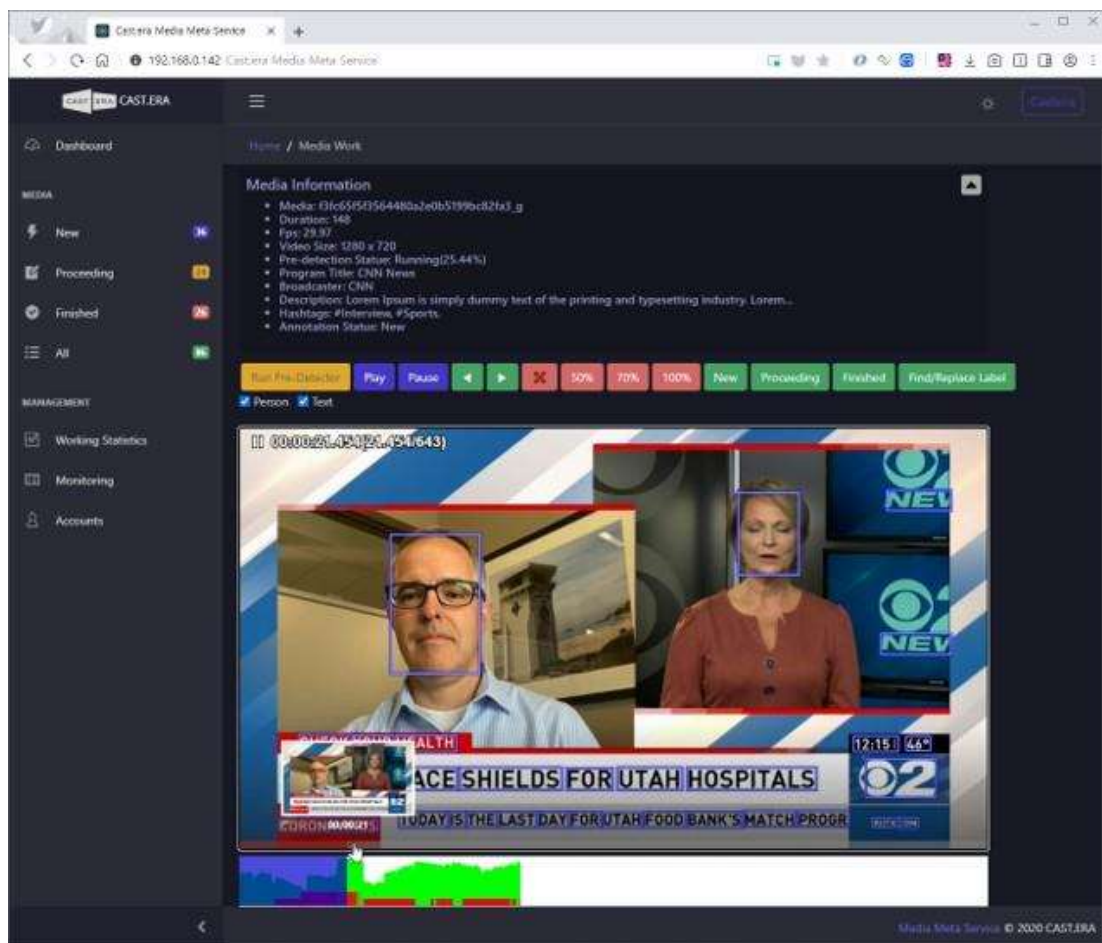


그림 2 CAST.ERA의 Media Meta Service

4.2.2. VideoLLM: Modeling Video Language with Visual and Temporal Context

중국의 여러 대학이 공동으로 연구하여 2023년 제안한 VideoLLM은 비디오의 시각적 및 시간적 맥락을 모두 고려하는 대규모 언어 모델이다. 이 모델은 비디오의 내용을 이해하고 설명하는 데 사용될 수 있으며, 자동 캡션 생성, 비디오 요약, 질의응답 등 다양한 태스크에 적용될 수 있다.

VideoLLM-online: Online Video Large Language Model for Streaming Video (CVPR 2024)

Joya Chen, Zhaoyang Lv, Shiwei Wu, Kevin Qinghong Lin, Chenan Song, Difei Gao, Jia-Wei Liu, Ziteng Gao, Dongxing Mao, Mike Zheng Shou
ShowLab, National University of Singapore Reality Labs Research, Meta

NOTE: 1. Audio is supported by ChatTTS for demonstration purpose. Other aspects is supported by our end-to-end streaming video large language model.
2. Pauses in the video is for waiting the slow audio. The inference speed of the model can approach real-time (5~10 FPS on RTX 3090 GPU, 10~15 FPS on A100 GPU).



그림 3 VideoLLM 데모 영상
VideoLLM의 주요 특징은 다음과 같다:

- 비디오-언어 사전 학습: 대규모 비디오-텍스트 쌍 데이터셋을 사용하여 사전 학습
- 시간적 어텐션 메커니즘: 비디오 프레임 간의 시간적 관계를 모델링
- 멀티모달 퓨전: 비디오의 시각적 특징과 언어적 특징을 효과적으로 결합

VideoLLM은 다양한 비디오 이해 태스크에서 state-of-the-art 성능을 달성했으며, 특히 복잡한 시간적 관계를 포함하는 태스크에서 우수한 성능을 보였다.

4.2.3. CLIP4Clip: An Empirical Study of CLIP for End to End Video Clip Retrieval

중국의 서남 교통대학과 Microsoft가 공동연구하여 2021년 제안한 CLIP4Clip은 OpenAI의 CLIP(Contrastive Language-Image Pre-training) 모델을 비디오 검색 태스크에 적용한 연구이다. 이 모델은 텍스트 쿼리를 사용하여 관련 비디오 클립을 검색할 수 있으며, 이는 비디오 메타데이터 추출 및 검색 시스템에 큰 도움이 될 수 있다.

CLIP4Clip의 주요 특징은 다음과 같다:

- 프레임 레벨 특징 추출: CLIP을 사용하여 각 비디오 프레임의 특징을 추출
- 시간적 집계: 추출된 프레임 특징을 시간적으로 집계하여 비디오 레벨 표현 생성
- 대조 학습: 텍스트 쿼리와 비디오 클립 간의 유사도를 학습

CLIP4Clip은 MSR-VTT, MSVD, LSMDC 등 다양한 비디오-텍스트 검색 벤치마크에서 우수한 성능을 보였으며, 특히 제로샷 설정에서도 높은 성능을 달성했다.

4.2.4. X-CLIP: End-to-End Multi-grained Video-Language Understanding

중국의 대학 및 연구소가 공동연구하여 2023년에 제안한 X-CLIP은 비디오-언 이해를 위한 종단간 다중 입도(multi-grained) 모델이다. 이 모델은 비디오의 다양한 시간적 해상도를 고려하여 메타데이터를 추출하고 비디오-텍스트 매칭을 수행한다.



그림 4 X-CLIP 모델에 대한 Query 및 응답 시연

X-CLIP의 주요 특징은 다음과 같다:

- 다중 입도 비디오 인코딩: 프레임, 클립, 전체 비디오 수준의 특징을 동시에 추출한다.
- 계층적 트랜스포머: 다양한 시간적 해상도의 특징을 효과적으로 통합한다.
- 대규모 사전 학습: 1억 개 이상의 비디오-텍스트 쌍을 사용하여 사전 학습을 수행한다.

X-CLIP은 비디오 검색, 비디오 질의응답, 비디오 캡션 생성 등 다양한 비디오-언어 태스크에서 우수한 성능을 보였다. 특히, 긴 비디오에 대한 이해력이 크게 향상되었다.

4.2.5. Video-ChatGPT: Towards Detailed Video Understanding via Large Vision and Language Models

아랍 에미리트, 오스트레일리아, 스웨덴의 3개 대학이 공동연구하여 2023년 제안한 Video-ChatGPT는 대규모 비전 및 언어 모델을 활용하여 상세한 비디오 이해를 가능하게 하는 연구이다. 이 모델은 비디오의 내용에 대해 자연어로 질의응답을 할 수 있으며, 이를 통해 비디오의 메타데이터를 자동으로 생성하거나 추출할 수 있다.



그림 5 Video ChatGPT의 데모 결과

Video-ChatGPT의 주요 특징은 다음과 같다:

- 멀티모달 인코딩: 비디오의 시각적 정보와 오디오 정보를 동시에 인코딩
- 대화형 인터페이스: 사용자가 자연어로 비디오에 대해 질문하고 답변받을 수 있음
- 시간적 추론: 비디오의 시간적 흐름을 고려한 질의응답 가능
- 제로샷 전이: 사전 학습된 모델을 fine-tuning 없이 다양한 비디오 이해 태스크에 적용 가능

Video-ChatGPT는 NextQA, MSRVTQ-QA 등의 비디오 질의응답 데이터셋에서 우수한 성능을 보였으며, 특히 복잡한 추론을 요구하는 질문에 대해서도 높은 정확도를 달성했다.

이러한 연구들은 AI 기반 메타데이터 추출 기술이 실제 비디오 콘텐츠 관리 시스템에 어떻게 적용될 수 있는지를 보여준다. 이 기술은 대량의 비디오 콘텐츠를 효율적으로 관리하고, 검색 가능성을 높이며, 콘텐츠의 가치를 극대화하는 데 기여할 수 있다.

4.3. 기술 동향

AI 기반 메타데이터 추출 기술의 최근 동향은 다음과 같다.

4.3.1. 자기 지도 학습의 발전

최근 연구들은 레이블이 없는 대량의 비디오 데이터를 활용하여 효과적인 특징 학습을 수행하는 자기 지도 학습 방법에 주목하고 있다. 예를 들어, 2021년 발표된 MoCo-v3 모델은 비디오 프레임 간의 시간적 일관성을 활용한 대조 학습을 통해 강력한 비디오 표현을 학습한다. 이 접근 방식은 레이블링 비용을 줄이면서도 높은 성능의 메타데이터 추출 모델을 학습할 수 있게 한다.

4.3.2. 경량화 모델 개발

모바일 기기나 엣지 디바이스에서의 실시간 메타데이터 추출을 위해 경량화된 모델 개발이 활발히 이루어지고 있다. 2019년 발표된 MobileNetV3는 네트워크 아키텍처 검색(NAS)과 NetAdapt 알고리즘을 결합하여 모바일 환경에 최적화된 경량 모델

을 제시했다. 이러한 경량 모델은 비디오 메타데이터 추출 분야에도 적용되어, 제한된 컴퓨팅 리소스에서도 효율적인 처리를 가능하게 한다.

4.3.3. 멀티태스크 학습 접근법

단일 모델로 여러 메타데이터 추출 태스크를 동시에 수행하는 멀티태스크 학습 접근법이 주목받고 있다. 2019년 발표된 Multitask Self-Supervised Visual Learning 모델은 객체 검출, 동작 인식, 깊이 추정 등 다양한 태스크를 동시에 학습한다. 이 접근법은 모델의 일반화 능력을 향상시키고, 계산 효율성을 높이는 데 기여한다.

4.3.4. 시공간 그래프 신경망의 활용

비디오 내 객체들 간의 복잡한 상호작용을 모델링하기 위해 시공간 그래프 신경망(Spatio-Temporal Graph Neural Networks)이 활용되고 있다. 2018년 발표된 Non-local Neural Networks는 비디오 내 장거리 의존성을 효과적으로 포착하여 더 정확한 메타데이터 추출을 가능하게 한다.

4.3.5. 연속적 학습 및 적응형 모델

실제 환경에서 지속적으로 변화하는 비디오 콘텐츠에 대응하기 위해, 연속적 학습(Continual Learning) 기법을 적용한 적응형 모델 개발이 이루어지고 있다. 2017년 제안된 Gradient Episodic Memory (GEM) 알고리즘은 새로운 태스크를 학습하면서도 이전에 학습한 태스크의 성능을 유지할 수 있게 한다.

4.3.6. 설명 가능한 AI 기술의 통합

메타데이터 추출 결과의 신뢰성과 해석 가능성을 높이기 위해 설명 가능한 AI 기술이 통합되고 있다. 2017년 제안된 Grad-CAM과 같은 기술은 모델의 결정 과정을 시각화하여 사용자가 추출된 메타데이터의 근거를 이해할 수 있게 한다.

이러한 기술 동향들은 AI 기반 메타데이터 추출 기술의 성능, 효율성, 적용 범위를 지속적으로 확장시키고 있다. 향후에는 더욱 정교한 멀티모달 학습, 프라이버시 보호 기술의 통합, 실시간 스트리밍 환경에서의 적용 등이 주요 연구 방향이 될 것으로 전망된다.

4.4. 미래 전망

AI 기반 메타데이터 추출 기술의 미래는 다음과 같은 방향으로 발전할 것으로 예상해 볼 수 있다.

- 메타버스와의 통합: 가상 현실(VR)과 증강 현실(AR) 환경에서 AI 기반 메타데이터 추출 기술이 중요한 역할을 할 것이다. 이는 실시간으로 가상 환경의 객체와 상호 작용을 분석하고 관련 정보를 제공하는 데 활용될 것이다.
- 감정 및 의도 분석: 비디오 콘텐츠에서 등장인물의 감정 상태와 의도를 분석하는 기술이 발전할 것이다. 이는 마케팅, 고객 서비스, 보안 등 다양한 분야에서 활용될 수 있다.
- 크로스 플랫폼 메타데이터 표준화: 다양한 플랫폼과 서비스 간에 메타데이터를 쉽게 공유하고 활용할 수 있도록 표준화가 진행될 것이다. 이는 콘텐츠의 상호운용성을 높이고 새로운 서비스 창출을 촉진할 것이다.
- 생성형 AI와의 결합: 메타데이터 추출 기술이 생성형 AI와 결합하여, 추출된 메타데이터를 바탕으로 새로운 콘텐츠를 자동으로 생성하거나 편집하는 기술이 발전할 것이다.
- 양자 컴퓨팅의 활용: 양자 컴퓨팅 기술이 발전함에 따라, 대규모 비디오 데이터에서 복잡한 패턴을 빠르게 분석하고 메타데이터를 추출하는 데 활용될 수 있다.
- 바이오메트릭 데이터 통합: 시청자의 생체 신호(예: 심박수, 뇌파)를 분석하여 콘텐츠에 대한 반응을 메타데이터로 추출하는 기술이 발전할 것이다. 이는 개인화된 콘텐츠 추천과 사용자 경험 개선에 활용될 수 있다.
- 블록체인 기술과의 융합: 메타데이터의 신뢰성과 무결성을 보장하기 위해 블록체인 기술이 활용될 것이다. 이는 저작권 보호와 콘텐츠 인증에 중요한 역할을 할 것이다.
- 환경 및 지속가능성 메타데이터: 비디오 콘텐츠의 제작 및 소비와 관련된 환경 영향을 분석하고 메타데이터로 추출하는 기술이 발전할 것이다. 이는 지속가능한 콘텐츠 제작과 소비를 촉진하는 데 기여할 것이다.
- 다국어 및 문화적 맥락 이해: AI 모델이 다양한 언어와 문화적 맥락을 더 깊이 이해하고 분석할 수 있게 될 것이다. 이는 글로벌 콘텐츠 유통과 현지화에 큰 도움이 될 것이다.

AI 기반 메타데이터 추출 기술은 이러한 방향으로 발전함에 따라 다양한 산업 분야에서 새로운 가치를 창출하고, 사용자 경험을 혁신적으로 변화시킬 것으로 기대된다.

5. AI 기반 비디오 화질 개선

5.1. 개요

AI 기반 비디오 화질 개선 기술은 인공지능과 머신러닝 알고리즘을 활용하여 저화질 비디오를 고화질로 변환하는 기술이다. 이 기술은 영상 처리, 컴퓨터 비전, 딥러닝 등 다양한 분야의 융합으로 발전해왔으며, 최근 급속한 발전을 이루고 있다. 주요 목적은 저해상도, 노이즈, 압축으로 인한 아티팩트 등 다양한 화질 저하 요인을 개선하여 비디오의 시각적 품질을 향상시키는 것이다.

이 기술은 초해상도, 노이즈 제거, 압축 아티팩트 제거, 프레임 보간, 색상 보정 및 향상 등 다양한 요소 기술로 구성된다. 초기에는 필터링, 에지 검출 등 전통적인 영상 처리 기법을 사용했으나, 머신러닝의 도입으로 성능이 향상되었고, 이후 CNN, GAN 등 딥러닝 기술의 적용으로 획기적인 성능 향상이 이루어졌다. 최근에는 GPU 가속화와 모델 경량화를 통해 실시간 비디오 화질 개선까지 가능해졌다.

AI 기반 비디오 화질 개선 기술은 방송 및 스트리밍 서비스, 보안 및 감시 시스템, 의료 영상, 영화 및 미디어 제작, 모바일 기기 등 다양한 분야에서 응용되고 있다. 이 기술은 저화질 콘텐츠의 화질을 개선하여 시청자 경험을 향상시키고, CCTV 영상의 식별력을 높이며, 의료 영상의 진단 정확도를 개선하는 등 광범위한 영향을 미치고 있다.

그러나 비디오 화질 개선 기술은 여전히 몇 가지 중요한 과제에 직면해 있다. 고화질 비디오의 실시간 처리를 위한 연산 효율성 개선, 다양한 촬영 환경과 비디오 유형에 대응할 수 있는 강건한 모델 개발, 화질 개선 과정에서 발생할 수 있는 새로운 아티팩트 최소화, 그리고 모바일 기기 등 제한된 환경에서의 효율적인 처리 방법 개발 등이 주요 과제로 남아있다.

AI 기반 비디오 화질 개선 기술은 계속해서 발전하고 있으며, 향후 더욱 정교하고 효율적인 알고리즘의 개발이 예상된다. 이를 통해 다양한 분야에서 비디오 콘텐츠의 품질 향상과 새로운 응용 가능성이 열릴 것으로 전망된다. 특히, 5G 네트워크의 확산, 8K 해상도의 상용화, VR/AR 기술의 발전 등과 맞물려 AI 기반 비디오 화질 개선 기술의 중요성은 더욱 커질 것으로 예상된다. 이에 따라 학계와 산업계에서는 지속적인 연구 개발을 통해 기술의 한계를 극복하고 새로운 혁신을 이루어낼 것으로 기대된다.

5.2. 논문 및 연구 사례

AI 기반 비디오 화질 개선 분야에서는 다양한 연구가 활발히 진행되고 있다. 주요 연구 사례와 논문들을 살펴보면 다음과 같다.

5.2.1. 딥러닝 기반 초해상도 기술

초해상도 기술은 저해상도 이미지나 비디오를 고해상도로 변환하는 기술이다. 2014년 Dong 등이 제안한 SRCNN(Super-Resolution Convolutional Neural Network)은 딥러닝 기반 초해상도 기술의 선구자적 역할을 했다. SRCNN은 저해상도 이미지를 고해상도로 변환하는 데 CNN을 사용했으며, 이는 전통적인 보간법보다 우수한 성능을 보였다.

이후 VDSR(Very Deep Super Resolution), EDSR(Enhanced Deep Super Resolution) 등의 모델이 등장하며 성능이 지속적으로 향상되었다. 특히 EDSR은 ResNet 구조를 기반으로 하여 더 깊은 네트워크를 구성함으로써 성능을 크게 개선했다.



그림 6 초기 초해상도 기술 결과 비교

5.2.2. GAN을 활용한 초해상도 기술

2017년 Ledig 등이 제안한 SRGAN(Super-Resolution Generative Adversarial Network)은 GAN 구조를 활용하여 더욱 사실적인 고해상도 이미지를 생성했다. SRGAN은 지각적 손실(perceptual loss)을 도입하여 단순한 픽셀 단위의 비교가 아닌, 인간의 시각 체계를 고려한 손실 함수를 사용했다. 이를 통해 PSNR은 다소 낮더라도 시각적으로 더 만족스러운 결과를 얻을 수 있었다.

5.2.3. 비디오 특화 초해상도 기술

2020년 연세대학교/페이스북 연구팀이 제안한 "STVUN: Deep Space-Time Video Upsampling Networks"는 시공간 정보를 활용하여 비디오의 프레임 레이트와 해상도를 동시에 향상시키는 방법을 제시했다. STVUN은 3D 컨볼루션과 주의 메커니즘을 결합하여 시간적 일관성을 유지하면서 고품질의 비디오를 생성하는 데 초점

을 맞췄다. 이 연구는 비디오의 특성을 고려한 초해상도 기술의 중요한 진전을 보여준다.

기술명	주요 기술	장점	발표
SRCNN	CNN 기반	딥러닝 초해상도의 시작	2014
EDSR	깊은 ResNet 구조	높은 PSNR, 효율적 학습	2017
SRGAN	GAN 구조, 지각적 손실	사실적인 텍스처 생성	2017
STVUN	3D 컨볼루션, 주의 메커니즘	시간적 일관성 유지	2020
Real-ESRGAN	합성 데이터 학습, GAN 구조	실제 열화에 강인함	2021

표 4 주요 비디오 화질 개선 연구 비교

5.2.4. 효율적인 비디오 복원 네트워크

2022년 Chan등 난양 기술 대학교 연구팀은 "BasicVSR++: Improving Video Super-Resolution with Enhanced Propagation and Alignment"에서 양방향 전파와 이차 변형 정렬을 활용한 효율적인 비디오 초해상도 네트워크를 제안했다. BasicVSR++는 기존 BasicVSR 모델을 개선하여 더 적은 파라미터로 높은 성능을 달성했다. 이 연구는 비디오 프레임 간의 정보를 효과적으로 활용하여 계산 효율성과 복원 품질을 동시에 향상시켰다.

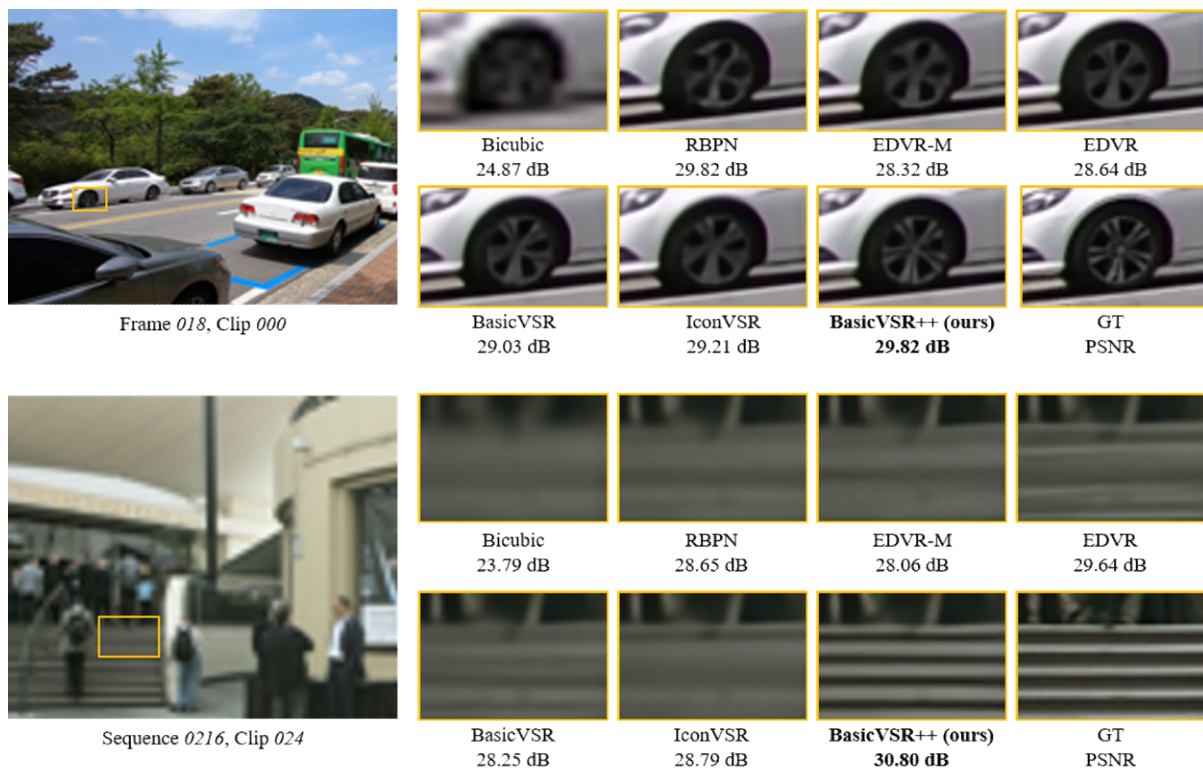


그림 7 최신 초해상도 기술 BasicVSR++ 성능 비교

5.2.5. 실시간 비디오 화질 개선

김문철 교수 연구팀(2023)은 딥러닝 기술을 이용해 풀 HD 비디오 영상을 4K UHD 초고화질 영상으로 실시간 변환할 수 있는 기술을 개발했다. 이 기술은 기존 AI 기반 초해상도 기술의 느린 처리 속도 문제를 해결하여 실시간 방송 환경에서도 적용 가능한 수준의 성능을 달성했다. 연구팀은 경량화된 딥러닝 모델과 GPU 가속을 통해 초당 60프레임의 처리 속도를 실현했다.



그림 8 KAIST의 하드웨어 기반 실시간 4K UHD 변환 기술

이러한 연구들은 AI 기반 비디오 화질 개선 기술의 다양한 접근 방식과 가능성을 보여주고 있다. 최근의 연구 동향은 단순한 해상도 향상을 넘어 시간적 일관성, 실시간 처리, 그리고 다양한 열화 유형에 대한 대응 능력을 향상시키는 데 초점을 맞추고 있다. 향후 연구에서는 더욱 효율적인 네트워크 구조, 개선된 학습 방법, 그리고 실제 응용 환경에서의 성능 향상이 주요 과제가 될 것으로 예상된다.

5.3. 기술 동향

AI 기반 비디오 화질 개선 기술은 빠르게 발전하고 있으며, 최근의 동향은 더욱 정교하고 효율적인 방법론을 추구하고 있다. 이러한 기술 동향을 세부적으로 살펴보면 다음과 같다.

5.3.1. 실시간 처리 기술의 발전

실시간 비디오 화질 개선 기술의 발전이 두드러지고 있다. NVIDIA의 DLSS(Deep Learning Super Sampling) 기술은 이러한 동향을 잘 보여주는 사례이다. DLSS 2.0은 딥러닝을 활용하여 낮은 해상도의 이미지를 실시간으로 고해상도로 변환한다. 이 기술은 게임에서 평균 2-3배의 성능 향상을 제공하며, 4K 해상도에서도 60FPS 이상의 프레임 레이트를 달성할 수 있게 한다.

- 게임 "Control"에서 DLSS 2.0을 적용했을 때, 4K 해상도에서 RTX 2080 Ti GPU를 사용하여 평균 프레임 레이트가 53.3 FPS에서 75.5 FPS로 약 41.6% 향상되었다



그림 9 실시간 비디오 화질 개선 기술 DLSS 적용 전후 비교

5.3.2. 멀티모달 학습 접근법

멀티모달 학습 접근법이 주목받고 있다. 예를 들어, MIT의 연구팀이 개발한 "VisualVoice" 시스템은 비디오의 시각 정보와 오디오 정보를 동시에 활용하여 화질을 개선하고 음성을 분리한다. 이 기술은 복잡한 장면에서도 화자의 얼굴과 음성을 정확히 매칭시켜 화질 개선의 정확도를 높인다.

- VisualVoice 시스템은 복잡한 대화 장면에서 각 화자의 얼굴과 음성을 정확히 매칭시켜, 기존 오디오 기반 방식보다 약 20% 향상된 음성 분리 성능을 보였다.

5.3.3. 적대적 생성 신경망(GAN)의 진화

TecoGAN(Temporally Coherent GAN)과 같은 모델은 비디오의 시간적 일관성을 유지하면서 고품질의 비디오를 생성할 수 있다. 이 모델은 4배의 해상도 향상을 제공하면서도 프레임 간의 자연스러운 전환을 유지한다. 또한 TecoGAN은 720p 해상도의 비디오를 4K 해상도로 업스케일링하면서, PSNR 측면에서 기존 SOTA 모델보다 약 0.5dB 높은 성능을 달성했다.

5.3.4. 3D 컨볼루션 네트워크의 발전

3D 컨볼루션을 활용한 "3DSRnet" 모델은 비디오의 시공간적 정보를 효과적으로 활용하여 화질을 개선한다. 이 모델은 2D 컨볼루션 기반 모델에 비해 PSNR이 평균 0.5dB 이상 향상되었다.

5.3.5. 메타학습과 few-shot 학습의 도입

서울대학교에서 개발한 "Meta-Transfer Learning for Zero-Shot Super-Resolution" 모델은 메타학습을 통해 다양한 열화 유형에 빠르게 적응할 수 있는 능력을 보여준다. 이 모델은 새로운 열화 유형에 대해 몇 개의 샘플만으로도 효과적인 화질 개선이 가능하다.

5.3.6. attention 메커니즘의 확장

"Video Super-Resolution Transformer" (VSRT) 모델은 Transformer 아키텍처를 비디오 초해상도 문제에 적용한 사례이다. 이 모델은 REDS4 데이터셋에서 PSNR 기준으로 기존 SOTA 모델 대비 0.3dB의 성능 향상을 달성했으며, 장거리 의존성을 효과적으로 모델링하여 복잡한 움직임이 있는 장면에서도 우수한 성능을 보인다.

5.3.7. 에지 컴퓨팅을 위한 최적화

Qualcomm Snapdragon 888 칩셋에 탑재된 AI Engine은 초당 26 TOPS(Trillion Operations Per Second)의 AI 연산 성능을 제공하여, 모바일 기기에서 실시간 비디오 화질 개선을 가능하게 한다. 이 기술은 모델 양자화와 하드웨어 가속화를 통해 저전력 환경에서도 효율적인 비디오 처리를 지원한다.

이러한 기술 동향들은 AI 기반 비디오 화질 개선 기술의 성능과 적용 범위를 지속적으로 확장시키고 있다. 실제로, Grand View Research의 보고서에 따르면, 글로벌 비디오 인공지능 시장은 2021년부터 2028년까지 연평균 성장률(CAGR) 23.5%로 성장할 것으로 예상된다.

5.4. 미래 전망

AI 기반 비디오 화질 개선 기술은 지속적으로 발전하고 있으며, 향후 더욱 혁신적인 변화가 예상된다. 주요 미래 전망은 다음과 같다

- 실시간 8K 업스케일링: 현재 4K 수준의 실시간 업스케일링이 가능해지고 있는 추세를 고려할 때, 향후에는 8K 해상도까지 실시간으로 업스케일링하는 기술이 상용화될 것으로 예상된다. 이는 방송 및 스트리밍 서비스에서 큰 변화를 가져올 것이다.
- 에지 컴퓨팅의 발전: 모바일 기기와 IoT 장치에서의 AI 기반 화질 개선이 더욱 보편화될 것이다. Qualcomm의 AI Engine과 같은 기술의 발전으로, 저전력 환경에서도 고품질의 비디오 처리가 가능해질 전망이다.
- 딥페이크 탐지와 결합: AI 화질 개선 기술은 딥페이크 탐지 기술과 결합되어, 영상의 진위 여부를 판별하는 데 활용될 것이다. 이는 미디어의 신뢰성 확보에 중요한 역할을 할 것으로 예상된다.
- 메타버스와의 통합: 가상 현실(VR)과 증강 현실(AR) 환경에서 AI 기반 화질 개선 기술이 중요한 역할을 할 것이다. 실시간으로 고품질의 가상 환경을 제공하는 데 필수적인 기술이 될 것이다.
- 개인화된 화질 최적화: 사용자의 시청 환경, 기기 특성, 개인 선호도 등을 고려한 맞춤형 화질 개선 기술이 발전할 것이다. 이는 더욱 향상된 사용자 경험을 제공할 것이다.
- 의료 영상 분야에서의 혁신: AI 기반 화질 개선 기술은 의료 영상의 정확도를 크게 향상시킬 것으로 예상된다. 이는 진단의 정확성을 높이고 조기 질병 발견에 기여할 것이다.

이러한 미래 전망은 AI 기반 비디오 화질 개선 기술이 단순히 영상의 품질을 높이는 것을 넘어, 다양한 산업 분야에서 혁신을 이끌어낼 것임을 시사한다. 앞으로 이 기술은 우리의 일상생활과 산업 전반에 더욱 깊이 통합되어, 새로운 가치를 창출하고 삶의 질을 향상시키는 데 기여할 것으로 전망된다.

6. 결론

AI 기반 메타데이터 추출 기술의 발전은 미디어 서비스 산업에서 중요한 변화를 가져오고 있다. 메타데이터 추출 기술은 비디오 콘텐츠에서 자동으로 인물, 텍스트, 객체 등을 인식하고 태그를 생성함으로써 콘텐츠 관리와 검색의 효율성을 크게 향상시킨다. 이러한 기술은 수동 작업에 비해 시간과 비용을 절감하며, 더 높은 정확성과 일관성을 제공한다.

비디오 화질 개선 분야에서도 AI의 적용이 두드러진다. 딥러닝 기반의 초해상도 기술은 저해상도 영상을 고해상도로 변환하는 데 큰 성과를 거두고 있으며, 이는 오래된 영상 콘텐츠를 현대의 고해상도 디스플레이에 맞게 업스케일링하는 데 활용되고 있다. AI 기반의 노이즈 제거 및 컬러 보정 기술 또한 열악한 촬영 환경에서 얻은 영상의 품질을 크게 향상시키고 있다.

실시간 처리 기술의 발전도 주목할 만하다. 5G 네트워크와 엣지 컴퓨팅의 발전으로 실시간 메타데이터 추출 및 비디오 화질 개선이 가능해지고 있으며, 이는 라이브 스트리밍 서비스나 실시간 방송에서 큰 가치를 창출할 수 있다.

그러나 이러한 기술의 발전은 개인정보 보호, 윤리적 사용, 데이터 편향성 등의 새로운 과제를 제시하고 있다. 따라서 이러한 문제들을 해결하기 위한 노력도 병행되어야 할 것이다.

향후 AI 기반 메타데이터 추출 및 비디오 화질 개선 기술은 더욱 정교해질 것으로 예상된다. 자연어 처리 기술의 통합과 감정 및 의도 분석 기술의 발전은 더욱 섬세한 콘텐츠 분석을 가능하게 할 것이다.

결론적으로, AI를 적용한 미디어 서비스의 발전은 거스를 수 없는 대세이며, 이는 미디어 산업의 미래를 형성하는 핵심 동력이 될 것이다. 메타데이터 추출과 비디오 화질 개선 기술은 이러한 변화의 최전선에 있으며, 이를 효과적으로 활용하고 관련 과제들을 적절히 해결해 나가는 기업과 서비스가 미래 미디어 시장을 주도할 것으로 전망된다.

부 록 1-1

(본 부록은 기술보고서를 보충하기 위한 내용으로 기술보고서의 일부는 아님)

지식재산권 요약서 정보

해당 사항 없음

부 록 1-2

(본 부록은 기술보고서를 보충하기 위한 내용으로 기술보고서의 일부는 아님)

시험인증 관련 사항

해당 사항 없음

부 록 1-3

(본 부록은 기술보고서를 보충하기 위한 내용으로 기술보고서의 일부는 아님)

본 기술보고서의 연계(family) 표준

해당 사항 없음

부 록 | -4

(본 부록은 기술보고서를 보충하기 위한 내용으로 기술보고서의 일부는 아님)

참고 문헌

[4. AI 기반 메타데이터 추출]

- [1] Xu, J., Mei, T., Yao, T., & Rui, Y. (2023). VideoLLM: Modeling Video Language with Visual and Temporal Context. arXiv preprint arXiv:2305.13292.
- [2] Luo, H., Ji, L., Shi, B., Huang, H., Duan, N., Li, T., ... & Zhou, M. (2021). CLIP4Clip: An Empirical Study of CLIP for End to End Video Clip Retrieval. arXiv preprint arXiv:2104.08860.
- [3] Maaz, M., Rasheed, H., Khan, S., Khan, F. (2023). Video-ChatGPT: Towards Detailed Video Understanding via Large Vision and Language Models. arXiv preprint arXiv:2306.05424.

[5. AI 기반 비디오 화질 개선]

- [1] C. Dong, et al. "Learning a deep convolutional network for image super-resolution," ECCV, 2014.
- [2] C. Ledig, et al. "Photo-realistic single image super-resolution using a generative adversarial network," CVPR, 2017.
- [3] B. Lim, et al. "Enhanced deep residual networks for single image super-resolution," CVPRW, 2017.
- [4] Y. Jo, et al. "STVUN: Deep space-time video upsampling networks," ECCV, 2022.
- [5] K. Chan, et al. "BasicVSR++: Improving video super-resolution with enhanced propagation and alignment," CVPR, 2022.
- [6] 김문철 교수 연구팀, "인공지능 통해 풀HD영상 4K UHD로 실시간 변환", KAIST 뉴스, 2023.
- [7] NVIDIA. (2020). NVIDIA DLSS 2.0: A Big Leap In AI Rendering. <https://www.nvidia.com/en-us/geforce/news/nvidia-dlss-2-0-a-big-leap-in-ai-rendering/>
- [8] Digital Foundry. (2020). Control DLSS 2.0 Analysis: Huge Improvements To Image Quality And Performance. <https://www.eurogamer.net/digitalfoundry-2020-control-dlss-2-dot-0-analysis>
- [9] Gao, R., & Grauman, K. (2021). VisualVoice: Audio-Visual Speech Separation with Cross-Modal Consistency. CVPR 2021.
- [10] Chu, M., Xie, Y., Leal-Taixé, L., & Thurey, N. (2020). Temporally coherent gans for video super-resolution (tecogan). arXiv preprint

- arXiv:2011.09010.
- [11] Kim, S., Li, G., Fuoli, D., Danelljan, M., Huang, Z., Gu, S., & Timofte, R. (2019). The vid3oc and intvid datasets for video super resolution and quality mapping. In 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW) (pp. 3609–3616). IEEE.
 - [12] Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., & Wang, M. (2021). Video Super-Resolution Transformer. arXiv preprint arXiv:2106.06847.
 - [13] Qualcomm. (2020). Qualcomm Unveils Snapdragon 888 5G Mobile Platform. <https://www.qualcomm.com/news/releases/2020/12/qualcomm-unveils-snapdragon-888-5g-mobile-platform>
 - [14] Grand View Research. (2021). Video Artificial Intelligence Market Size, Share & Trends Analysis Report By Solution, By Deployment, By Application, By End-use, By Region, And Segment Forecasts, 2021 – 2028. <https://www.grandviewresearch.com/industry-analysis/video-artificial-intelligence-market>

※ 상기 기재된 참고 문헌의 발간일이 기재된 경우, 해당 표준(문서)의 해당 버전에 대해서만 유효하며, 연도를 표시하지 않은 경우에는 해당 표준(권고)의 최신 버전을 따름

부 록 1-5

(본 부록은 기술보고서를 보충하기 위한 내용으로 기술보고서의 일부는 아님)

영문기술보고서 해설서

해당 사항 없음

부 록 I-6

(본 부록은 기술보고서를 보충하기 위한 내용으로 기술보고서의 일부는 아님)

기술보고서의 이력

판수	채택일	기술보고서번호	내용	담당 위원회
제1판	2024.11.29.	제정 FBMF-TR-019		
오류정 정				
제2판				