특집논문 (Special Paper)
방송공학회논문지 제30권 제2호, 2025년 3월 (JBE Vol.30, No.2, March 2025)
https://doi.org/10.5909/JBE.2025.30.2.122
ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

방송 그래픽 제작을 위한 생성형 AI 기반 맞춤형 시스템 개발 및 적용 가능성 연구

이 윤 재[a], 이 용 건[a], 이 영 화[b], 양 지 희[b], 이 승 주[c], 강 현 석[c], 박 구 만[d]‡

Development of a Generative AI-based Customized System for Broadcast Graphics Production and Study on its Applicability

Yoonjae Lee[a], Yonggun Lee[a], Younghwa Lee[b], Jihee Yang[b], Seungju Lee[c], Hyunseok Kang[c], and Gooman Park[d]‡

요 약

본 연구는 생성형 인공지능을 활용하여 방송 그래픽 제작의 효율성과 품질을 향상시키기 위한 맞춤형 시스템 개발 및 적용 가능성을 탐구한다. Stable Diffusion XL 모델을 기반으로 LoRA 및 Dreambooth 기법을 활용해 특정 그래픽 스타일을 반영하는 AI 모델을 구축하였으며, 트리거 워드를 사용해 간편하게 스타일을 적용할 수 있도록 설계하였다. 또한 폭력적인 단어와 같은 민감한 키워드가 포함된 그래픽 생성 실험을 통해 보도 자료에 필요한 시각적 요소를 윤리적 기준 내에서 효과적으로 생성할 수 있음을 확인하였다. 그라디오(Gradio) 기반의 사용자 중심 웹 인터페이스를 도입하여 텍스트 입력만으로 그래픽을 쉽게 조정하고 실시간으로 결과를 확인할 수 있도록 하였다. 본 연구는 생성형 AI 기술이 방송 그래픽 제작 현장에서 실질적으로 활용될 수 있는 가능성을 제시하며, 다양한 콘텐츠 제작 분야로의 확장 가능성을 보여준다.

Abstract

This study explores the development and applicability of a generative AI-based customized system for broadcast graphics production. By fine-tuning the Stable Diffusion XL model using LoRA and Dreambooth techniques, the system reflects unique broadcast styles. Trigger words enable easy style application, and sensitive keyword experiments demonstrate ethical visual content generation. A Gradio-based interface allows users to adjust graphics and preview results in real-time. This research highlights the potential for generative AI to enhance efficiency and quality in broadcast graphics, with applications extending to broader content creation fields.

Keyword : Generative AI, Text to image, Video generation, Fine tuning, Customized model

Ⅰ. 서 론

인공지능은 최근 몇 년간 괄목할 만한 발전을 이루며, 다양한 사업군에서 핵심 기술로 자리 잡고 있다. 특히, 이미지 생성, 자연어 처리, 음성 인식 등 다양한 AI 기술은 제조, 의료, 금융, 미디어 등 산업 전반에 걸쳐 그 응용 가능성을 입증하고 있다. 이 중에서도 생성형 인공지능(Generative AI)은 기존의 데이터를 학습하여 새로운 데이터를 생성하는 기술로, 특히 창의적인[1] 작업 영역에서 혁신을 이끌고 있다.

방송 산업은 정보 전달과 시각적 표현이 중요한 분야로, 시청자들에게 효과적으로 메시지를 전달하기 위해 고품질의 그래픽 제작이 요구된다. 기존의 그래픽 제작 방식은 높은 기술적 숙련도와 많은 시간 및 비용을 필요로 한다. 이러한 문제를 해결하기 위해 생성형 인공지능 기술을 방송 그래픽 제작에 활용하려는 시도가 점점 더 주목받고 있다.

생성형 AI 기술의 발전은 이미지 및 비디오 생성 모델의 성능 향상으로 이어졌으며, 이는 방송 그래픽 제작 환경에서의 효율성을 크게 높일 수 있는 가능성을 제시한다. 생성형 AI를 활용하면 반복적인 그래픽 작업을 자동화하고, 다양한 스타일과 톤을 유지하면서도 빠른 시간 내에 원하는 결과물을 얻을 수 있다. 이를 통해 제작자는 창의적인 아이디어에 더 많은 시간을 할애할 수 있으며, 비용 절감과 제작 효율성 향상이라는 두 가지 목표를 동시에 달성할 수 있다.

최근 이미지를 생성해 주는 상용 도구들이 많이 출시되고 있는데 Midjourney[2]와 DALL-E 3[3]가 대표적이다. Midjourney는 아트와 디자인 중심의 이미지 생성 플랫폼으로, 사용자들이 텍스트 프롬프트를 통해 예술적인 이미지와 독창적인 비주얼을 만들어낼 수 있도록 설계되었다. 특히, 사용자들이 디스코드(Discord) 서버를 통해 이미지를 생성하는 방식은 커뮤니티 기반의 창작 환경을 조성하여 활발한 피드백과 공유를 가능하게 한다는 특징이 있다. DALL-E 3는 OpenAI에서 개발한 이미지 생성 모델로, 사용자 프롬프트를 정교하게 해석하여 세밀하고 사실적인 이미지를 생성할 수 있다.

또한 텍스트를 기반으로 비디오를 생성하는 상용 도구 또한 활발히 개발되고 있는데, 최근엔 OpenAI의 Sora[4]가 ChatGPT 유료 사용자를 중심으로 공개되고 있으며, 실사에 가까운 영상 생성으로 주목을 받고 있다. 그 외 Runway의 Gen3[5]와 LumaAi의 Dream Machine[6], KlingAi[7] 등 다양한 동영상 생성 상용 도구들이 계속해서 출시되고 있다.

그러나 현재 상용화된 생성형 AI 도구를 방송 그래픽 제작에 직접적으로 적용하는 데는 한계점이 존재한다. 상용 도구는 일반 사용자를 위해 설계되어 있기 때문에 특정 방송국이나 프로그램의 고유한 그래픽 스타일을 반영하기 어렵고 프로그램 제작에 있어 일관성을 유지하는 것이 힘들다. 이러한 한계로 방송 제작 환경에 최적화된 맞춤형 생성 AI 시스템이 필요하다.

본 연구에서는 방송 제작 환경에서 생성형 AI 기반 맞춤형 그래픽 생성 시스템의 적용 가능성을 검증하기 위해 실험적 연구와 성능 평가를 결합한 연구 방법을 적용하였다. 먼저, 생성형 AI 모델 구축을 위해 Stable Diffusion XL(SDXL)[8] 모델을 기반으로 특정 방송사의 그래픽 스타일을 반영할 수 있도록 파인튜닝을 진행하였다. 이를 위해 KBS의 애니메이션 그래픽과 보도 그래픽 데이터 셋을 수집하고, 이를 학습 데이터로 활용하여 LoRA(Low-Rank Adaptation)[9] 및 Dreambooth[10] 기법을 적용하였다. 또한, 이미지 생성뿐 아니라, Image-to-Video 변환을 위해 AnimateAnything 모델을 활용하여 동영상 생성 실험을 수행하였다.

트리거 워드(KBSstyle)를 활용하여 트리거 단어의 입력만으로 스타일을 반영하는 효과를 검증하였으며, ControlNet을 통해 Canny Edge, Depth Map, OpenPose 등의 조건을 활용하여 특정 스타일을 유지하면서도 사용자 요구에 맞춘 세부적인 조정이 가능함을 확인하였다. 또한, AI를 활용한 윤리적 그래픽 생성 가능성을 검증하기 위해 폭력적 사건

a) 한국방송공사(Korean Broadcasting System)
b) 시그마케이 주식회사(Sigmak Co., Ltd.)
c) 서울과학기술대학교 일반대학원 스마트 ICT융합공학과(Department of Smart ICT Convergence Engineering, Seoul National University of Science and Technology)
d) 서울과학기술대학교 스마트 ICT융합공학과(Department of Smart ICT Convergence Engineering, Seoul National University of Science and Technology)
‡ Corresponding Author : 박구만(Gooman Park)
E-mail: gmpark@seoultech.ac.kr
Tel: +82-2-970-6430
ORCID: https://orcid.org/0000-0002-7055-5568
※ 이 논문의 연구 결과 중 일부는 한국방송·미디어공학회 2024년 추계학술대회에서 발표한 바 있음.
· Manuscript January 13, 2025; Revised February 25, 2025; Accepted February 25, 2025.

을 포함한 보도 그래픽 생성 실험을 진행하였으며, 윤리적 기준을 준수하면서도 사실적인 시각적 요소를 효과적으로 생성할 수 있는지 분석하였다.

제안한 AI 기반 맞춤형 시스템은 [그림 1]과 같이 학습 데이터를 이용하여 파인튜닝한 이미지 생성 모듈과 동영상 생성 모듈을 그라디오 Web UI를 기반으로 통합하였으며 Ollama API를 활용하여 한글 프롬프트 입력을 이미지 생성에 알맞은 형식의 영어 프롬프트 입력으로 자동 변환하여 사용자 편의성을 향상시켰다. 또한 동영상 생성 출력 시 여러 편의 후보 영상을 생성하여 사용자로 하여금 원하는 이미지의 영상을 선택할 수 있도록 설계하였다.

구축된 시스템에서 생성된 그래픽과 동영상의 성능 평가를 위해 Mean Opinion Score(MOS) 방식을 활용하여 정성 평가를 진행하였다. 콘텐츠 제작 종사자 17명을 대상으로 실험을 수행하였으며, 평가 항목으로는 원본 스타일과의 유사성, 프롬프트 반영 정도, 시각적 품질을 포함하였다.

본 연구는 기존 연구 및 상용 AI 도구와 비교하여 방송 제작 환경에서 생성형 AI의 실무 적용 가능성과 한계를 평가하고 맞춤형 AI 생성 시스템을 설계하였다. 이를 통해 AI 기반 그래픽 생성 기술이 방송 제작 과정에서 어떻게 활용될 수 있는지 검토하고, 향후 발전 방향을 모색하고자 한다.

Ⅱ. 선행 연구 분석

1. 이미지 생성 모델

AI 기반 이미지 생성 기술은 컴퓨터 비전 및 콘텐츠 제작 분야에서 혁신을 이끌고 있다. 초기에는 생성적 적대 신경망(Generative Adversarial Networks, GAN)[11]이 이미지 생성 기술의 주요 방식으로 자리 잡았으나, 최근에는 확산 모델(Diffusion Model)[12]이 더욱 주목받고 있다.

GAN은 2014년 Ian Goodfellow가 제안한 생성 모델로, 생성자와 판별자가 경쟁하며 이미지를 생성하는 방식이다. 생성자는 실제와 유사한 데이터를 만들고, 판별자는 이를 구별하며 학습이 진행된다. DCGAN[13], StyleGAN[14], BigGAN[15] 등 다양한 변형 모델이 등장해 이미지 품질이 향상되었지만, 훈련의 불안정성과 모드 붕괴(Mode Collapse) 등의 한계가 있다.

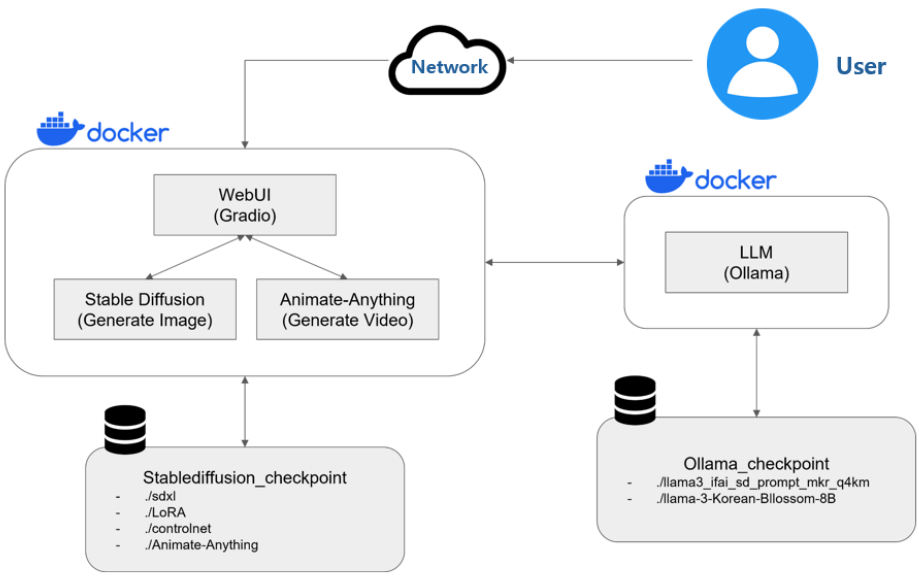


그림 1. 시스템 구성도
Fig. 1. System diagram

이러한 한계를 극복하고자 확산 모델이 등장했다. 확산 모델은 데이터에 점진적으로 노이즈를 추가한 후, 이를 역으로 제거하며 이미지를 생성하는 방식이다. 초기에는 순수한 노이즈에서 시작해 점진적으로 노이즈를 제거하며 이미지를 형성하는데, GAN보다 안정적으로 고품질 이미지를 생성할 수 있다는 점에서 큰 주목을 받았다.

특히 Denoising Diffusion Probabilistic Models(DDPM)[16]과 같은 모델이 확산 모델의 기틀을 마련했으며, 이후 Improved DDPM[17]과 Stable Diffusion[18]이 등장하면서 실험 및 실제 콘텐츠 제작에서 널리 사용되고 있다.

스테이블 디퓨전(Stable Diffusion)은 특히 텍스트에서 이미지로의 전환이 자유롭고 효율적이며, 고해상도 이미지를 빠르게 생성할 수 있다는 점에서 다른 모델들과 차별화된다.

본 연구에서는 SDXL 모델을 사용하여 이미지 생성을 진행하였다. SDXL 모델은 기존 스테이블 디퓨전 모델에 비해 3배 이상의 UNet 백본을 사용하며, 첫 단계인 프롬프트 입력 시 기본(Base) 모델을 사용해 잠재 이미지(latent image)를 생성하고 최종 노이즈 제거 단계에서 특화된 리파이너(Refiner) 모델을 사용하여 해상도가 높은 이미지를 생성한다.

2. 비디오 생성 모델

비디오 생성은 이미지 생성보다 복잡한 기술적 도전과제를 수반한다. 프레임 간의 시간적 일관성을 유지하면서도 시각적 품질을 확보하는 것이 핵심 과제다. 비디오 생성 모델은 이미지를 입력으로 주고 동영상으로 생성하는 Image-to-Video 모델과, 텍스트를 입력으로 넣어주고 동영상을 생성하는 Text-to-Video 모델로 나눌 수 있다.

ConsistI2V(ConsistI2V: Enhancing Visual Consistency for Image-to-Video Generation)[19]는 대표적인 Image-to-Video SOTA(State-of-the-art) 모델로 이미지-비디오 생성 과정에서 시각적 일관성과 움직임의 부드러움을 유지하기 위해 시공간적 주의 메커니즘(Spatiotemporal Attention)을 활용하여 비디오의 프레임 간 일관성을 높였다. 또한 첫 프레임의 저주파(low-frequency) 구성 요소를 초기화된 노이즈에 결합해 비디오 레이아웃의 일관성을 유지하였다.

AnimateAnything(AnimateAnything: Consistent and Controllable Animation for Video Generation)[20]은 다중 조건 제어 신호를 프레임 단위의 광학 흐름(optical flow)으로 변환하여 비디오를 생성한다. 변환된 광학 흐름을 바탕으로 최종 비디오를 생성하는 Video Generation Module을 통해 높은 품질의 일관된 비디오를 제작할 수 있는 특징이 있으며, 대규모 움직임으로 인해 발생하는 깜박임(flickering) 현상을 완화하기 위해 주파수 도메인에서 비디오 특징을 조정한다.

AnimateDiff(AnimateDiff: Animate Your Personalized Text-to-Image Diffusion Models without Specific Tuning)[21]는 Text-to-Video 생성 모델로 한 번의 학습으로 다양한 개인화된 Text-to-Image 모델에 통합할 수 있으며, 프레임 간 움직임을 학습해 부드럽고 자연스러운 애니메이션을 생성하는 특징이 있다. 또한 소량의 참조 비디오와 최소한의 학습으로 특정 모션 패턴(예: 줌인, 팬닝 등)을 효과적으로 적용할 수 있는 경량화된 튜닝 기법인 Motion LoRA를 적용하여 새로운 패턴을 빠르게 학습할 수 있는 장점이 있는 모델이다.

3. 모델 파인튜닝

모델 파인튜닝은 사전 학습된 대규모 생성형 AI 모델을 특정 데이터 셋에 맞게 재학습시켜, 새로운 목적에 적합하도록 성능을 개선하는 방법이다. 파인튜닝을 통해 모델은 특정 스타일, 톤 또는 개별 사용자 요구에 맞는 결과물을 생성할 수 있으므로 방송 그래픽, 이미지 생성, 비디오 편집 등 다양한 콘텐츠 제작 분야에서 활용되고 있다. 대표적인 파인튜닝 기법으로는 LoRA와 Dreambooth 기법이 있다.

LoRA는 대규모 언어 모델 또는 이미지 생성 모델에서 특정 가중치 행렬을 저차원 근사로 분해하여 모델을 부분적으로 업데이트하는 방식이다. 모델의 원래 구조를 보존하며, 역전파 시에도 원래 가중치는 고정되어 있어 메모리 활용에 효율적이다.

Dreambooth는 텍스트-이미지 생성 모델을 특정 대상이나 스타일로 정교하게 개인화하는 기법으로, 디퓨전 모델을 기반으로 한다. Dreambooth는 입력 데이터와 함께 텍스

트 프롬프트를 사용하여 특정 개체의 정체성을 보존하면서도 다양한 맥락에서 해당 개체를 생성할 수 있도록 모델을 학습시킨다. 텍스트 인코더를 활용하여 텍스트와 이미지 간의 연관성을 강화하며, 기존 모델의 크로스 어텐션(Cross-Attention) 레이어를 조정하여 특정 개체에 대한 민감도를 높인다.

LoRA와 Dreambooth는 생성형 AI 모델의 파인튜닝을 위한 도구로, 각각 효율성과 세밀한 조정이 가능하다는 기술적 강점을 지닌다. 따라서 본 연구에서는 LoRA 기법과 Dreambooth 기법을 사용하여 특정 스타일과 사용자의 요구를 반영하는 맞춤형 그래픽 제작 시스템을 설계하였다.

4. 조건부 생성 기능

조건부 생성은 이미지 생성에서 특정 조건이나 입력을 기반으로 원하는 형태의 이미지를 생성하는 기술이다. 일반적인 생성 모델이 단순하게 무작위로 이미지를 생성하는 것과 달리, 사용자가 제공한 정보 또는 조건을 반영하여 더 구체적이고 원하는 결과물을 만들어내는 방식이다. 텍스트, 이미지, 스케치, 키포인트 등 다양한 형식의 조건이 사용될 수 있으며 생성된 결과물을 사용자의 의도에 맞게 조정할 수 있는 장점이 있다.

ControlNet(Adding Conditional Control to Text-to-Image Diffusion Models)[22]은 조건부 이미지 생성 모델로, 디퓨전 모델에 조건을 추가하여 이미지 생성 과정을 제어할 수 있다. 사전 학습된 모델의 파라미터를 고정한 채, 학습 가능한 복사본을 병렬로 추가하여 새로운 조건부 제어 기능을 제공한다. 따라서 기본 디퓨전 모델에 조건부 네트워크를 추가하는 방식으로 기존 모델의 성능을 저하시키지 않으면서 기능을 강화할 수 있으며 하나의 모델에 여러 유형의 조건을 지원하여 다양한 생성 작업에 적용할 수 있다.

5. 기존 방송 그래픽 제작 과정

전통적인 방송 그래픽 제작 과정은 주로 디자인 전문가의 수작업을 기반으로 이루어지며, 높은 수준의 기술적 숙련도와 창의성이 요구된다. 그래픽 제작은 단순히 시각적 요소를 추가하는 것이 아니라, 방송의 목적과 콘텐츠 특성에 맞는 디자인을 구성하고 시청자가 정보를 효과적으로 이해할 수 있도록 시각적 전달력을 극대화하는 과정이다. 이 과정은 기획 단계에서부터 최종 송출에 이르기까지 여러 단계로 진행되며, 각 단계에서 제작자의 의도와 방송사의 브랜드 정체성을 반영해야 한다.

먼저, 방송 그래픽 제작의 첫 번째 단계는 기획 및 디자인이다. 이 단계에서는 프로그램의 성격과 대상 시청층을 고려하여 그래픽 스타일을 결정한다. 예를 들어, 뉴스 프로그램의 경우 정보 전달이 명확해야 하므로 직관적인 레이아웃과 신뢰감을 주는 색상이 사용되며, 예능 프로그램은 밝고 역동적인 그래픽 요소가 강조된다. 또한, 특정 방송사의 브랜드 아이덴티티를 유지하기 위해 기존 방송 그래픽과의 일관성을 고려한 디자인 방향을 설정한다.

기획이 완료되면, 그래픽 제작 단계로 넘어간다. 이 과정에서는 Adobe Photoshop, Illustrator, After Effects, Cinema 4D 등의 그래픽 및 영상 편집 소프트웨어를 활용하여 디자인을 구현한다. 2D 또는 3D 애니메이션을 추가하여 영상의 몰입감을 높이고, 특정 프로그램이나 방송사의 고유 스타일을 반영하기 위해 세부적인 디자인 작업이 이루어진다.

그래픽이 완성되면, 수정 및 피드백 반영 단계를 거친다. 제작된 그래픽은 방송 제작팀과 협의하여 검토되며, 필요에 따라 색상 조정, 텍스트 배치 변경, 영상 효과 조정 등의 수정이 이루어진다. 특히, 뉴스 및 보도 그래픽의 경우 사실 전달이 중요한 만큼 내용의 정확성을 검토하는 과정이 필수적이며, 방송국의 편집 지침을 준수해야 한다. 이 단계에서는 방송 프로듀서(PD)와 그래픽 디자이너 간의 협업이 필수적이며, 최종적으로 방송 콘텐츠에 적합한 형태로 그래픽이 완성된다.

마지막으로, 최종 적용 및 송출 단계에서는 제작된 그래픽이 방송 시스템에 적용되며, 실시간 방송 또는 녹화 콘텐츠에 삽입된다. 이 과정에서는 그래픽의 해상도 및 형식이 방송 송출 환경과 호환되는지 최종 점검이 이루어진다.

이와 같은 기존 방송 그래픽 제작 방식은 높은 제작 비용과 긴 작업 시간이 필요하며, 반복적인 수정 과정이 발생할 가능성이 높다는 단점이 있다. 또한, 대규모 이벤트나 긴급 뉴스 속보와 같이 빠른 그래픽 제작이 요구되는 상황에서는 기존 방식이 한계를 가질 수 있다. 이에 따라, 생성형

AI를 활용한 그래픽 자동화가 방송 제작의 효율성을 높이고, 스타일의 일관성을 유지하면서도 제작 시간을 단축하는 대안으로 주목받고 있다. 본 연구에서는 이러한 한계를 해결하기 위해 생성형 AI 기반의 맞춤형 그래픽 생성 시스템을 개발하고, 기존 방식과 비교하여 방송 제작 환경에서의 활용 가능성을 분석하고자 한다.

Ⅲ. 데이터 셋 구성 및 모델 선정

1. 데이터 셋 구성

본 연구에서는 KBS 그래픽의 스타일과 톤이 반영된 제작물을 생성하기 위해 KBS TV 유치원 애니메이션 캐릭터 “삐아, 빵야”와 보도 방송용 “보도 그래픽” 데이터를 학습 데이터로 사용하였다.

1.1 애니메이션 그래픽

“삐아”는 노란 아기 병아리 캐릭터이며 머리에 새싹을 달고 있는 특징을 가지고 있다. “빵야” 또한 노란 아기 병아리 캐릭터이며 머리에 흰색 꽃을 달고 있다. 두 캐릭터 모두 노란 아기 병아리라는 공통점이 있지만 머리 부분에 있는 장식이 다르며, 몸통의 색상 톤이 조금 다르다. 크기도 상대적으로 빵야가 크다. 데이터의 라벨링은 각 캐릭터의 이름인 “ppia”, “ppangya”로 하였다.

이미지 생성용 학습을 위해 사용한 데이터는 배경 없이 검은 바탕의 캐릭터만 있는 png 파일 189장을 사용했다. 동영상 생성용 학습을 위해서는 1분 가량의 동영상 파일 11편을 사용하였다.

1.2 보도 그래픽

보도 그래픽 데이터는 실제 뉴스와 보도 자료에 사용된 일러스트 이미지와 동영상 클립으로 구성하였다. 이미지 생성 학습을 위해 일러스트 이미지 400장과 동영상 생성 학습을 위해 5초~10초 가량의 동영상 클립 7편을 사용하였다. KBSstyle로 라벨링하여 학습하였으며, 라벨링을 트리거 워드로 사용하여 트리거 워드가 스타일 반영에 미치는 영향을 알아보았다.

KBS 보도 그래픽 일러스트 이미지는 배경보다는 윤곽을 선명하게 표현하여 인물을 강조하는 특징이 있으며 수채화

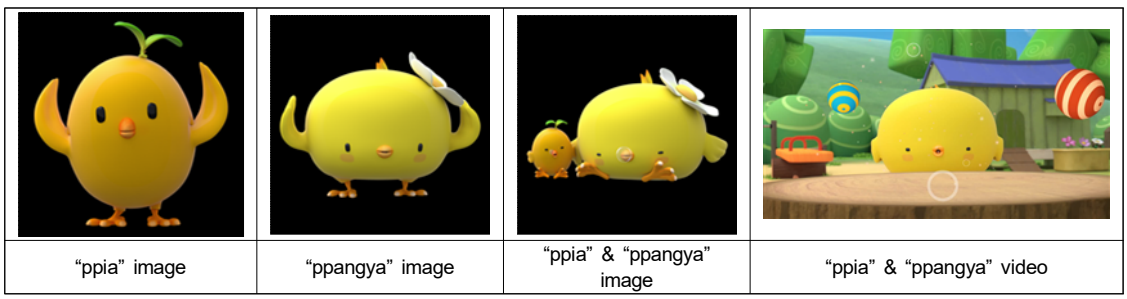


그림 2. 애니메이션 그래픽 학습 데이터
Fig. 2. Animated graphic training data

그림 3. 보도 그래픽 학습 데이터
Fig. 3. News graphic training data

풍의 채색적 특징을 가지고 있다. 동영상 클립은 정확한 내용 전달이 목적인 보도 그래픽의 특성상 이미지 자체에 많은 움직임을 주지 않았다는 특징이 있다.

2. 모델 선정

2.1 이미지 생성 모델

오픈소스 모델을 테스트하며 KBS 콘텐츠에 적합한 이미지 생성 모델을 선정하였다.

SOTA 모델 중 하나인 PixArt-alpha(PixArt-α: Fast Training of Diffusion Transformer for Photorealistic Text-to-Image Synthesis)[23] 모델을 파인튜닝하여 테스트를 진행하였다.

[그림 4]에서 볼 수 있듯이, 애니메이션 그래픽의 경우 “ppangya is walking”을 입력한 결과 파인튜닝 결과에서는 캐릭터의 색상만 학습하고, 캐릭터의 외형이 학습되지 않은 것으로 나타났다. 보도 그래픽의 경우도 “man is walking”을 입력한 결과 원본 스타일과 차이는 있지만, 비교적 유사하게 학습하는 경향을 보였다.

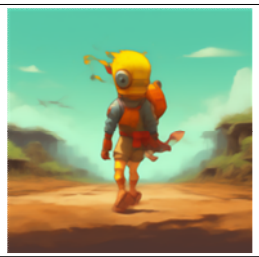
Animated graphic		
Prompt	ppangya is walking	
Generation Results		
News graphic		
Prompt	man is walking	
Generation Results		

그림 4. PixArt-alpha 파인튜닝 모델 결과
Fig. 4. PixArt-alpha fine-tuning model results

Prompt	man and woman are running on the street	
Generation Results		
Model	Stable Diffusion 1.5	Stable Diffusion XL

그림 5. Stable Diffusion 1.5 모델과 Stable Diffusion XL 모델 생성 이미지 비교
Fig. 5. Comparison of generated images of the Stable Diffusion 1.5 model and the Stable Diffusion XL model

스테이블 디퓨전 모델은 디퓨전 기반 이미지 생성 모델 중 하나로, 고해상도 이미지 생성과 세밀한 디테일 표현에 강점을 지닌다. 스테이블 디퓨전 모델 중 안정적인 이미지 생성으로 알려진 Stable Diffusion 1.5 모델과 SDXL 모델을 비교하였으며 [그림 5]와 같이 SDXL 모델이 굵은 선과 수채화 기법의 특징이 있는 KBS 스타일과 유사하게 표현함을 확인하였다. 또한 Dreambooth 및 LoRA 같은 파인튜닝 기법을 통해 특정 개체나 스타일을 쉽게 반영할 수 있어, KBS 그래픽의 톤은 유지하면서도 다양한 그래픽 요소를 생성하는 데 적합하다고 판단하였다. 또한 최대 1024x1024 해상도의 이미지 생성이 가능하므로 업스케일을 통한 방송 콘텐츠 제작도 가능하다.

2.2 동영상 생성 모델

본 연구에서는 이미지 생성 모델을 통해 생성된 이미지를 바탕으로 비디오를 생성하는 것을 목표로 하고 있기 때문에 Image to Video 모델 중 SOTA 모델을 중심으로 테스트를 진행하였다. [그림 6]과 같이 ConsistI2V 모델을 파인튜닝하여 실험한 결과 입력 이미지에 대한 일관성을 잃어버리고 학습한 영상과 비슷한 동영상이 생성되었다.

[그림 7]의 보도 그래픽 예시와 같이 AnimateAnything 모델의 테스트 결과 객체의 움직임은 크지 않지만 캐릭터가 일관성 있게 유지되고 안정적인 영상이 생성됨을 확인할 수 있었다. 따라서 본 연구에서는 AnimateAnything 모델을 파인튜닝하여 동영상 생성 모델로 활용하였다.

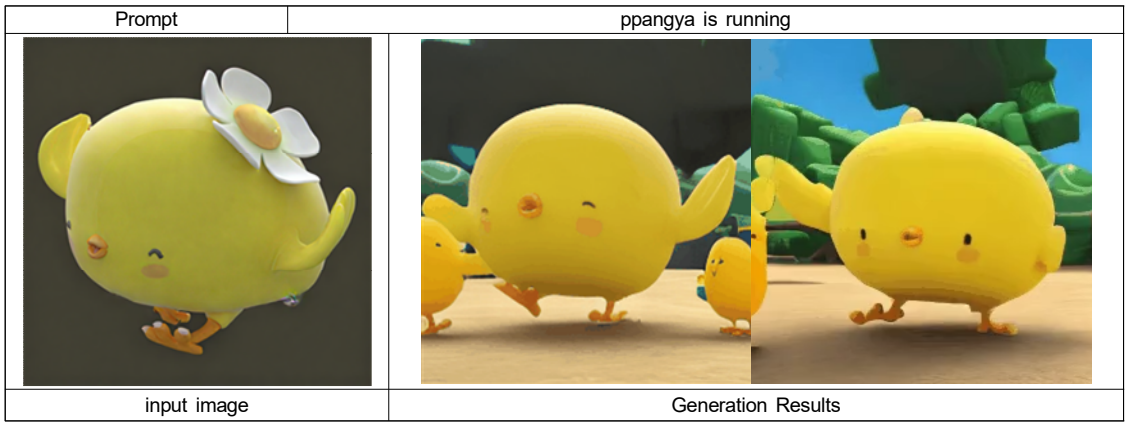


그림 6. ConsistI2V 파인튜닝 모델 테스트 결과
Fig. 6. ConsistI2V fine-tuning model test results

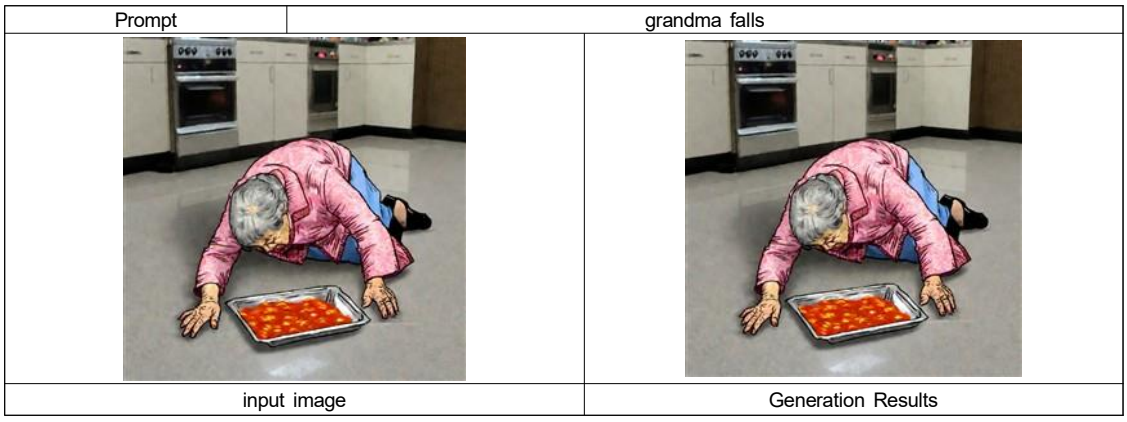


그림 7. AnimateAnything 모델 테스트 결과
Fig. 7. AnimateAnything model test results

Ⅳ. 실험 결과 및 성능 평가 분석

1. 이미지 생성

1.1 애니메이션 그래픽

애니메이션 그래픽은 방송국의 어린이 프로그램, 애니메이션 방송, 캐릭터 중심 콘텐츠에서 활용되는 그래픽 스타일을 유지하는 것이 핵심이다. 따라서, 선명한 색감, 단순화된 형태, 캐릭터 디자인의 일관성을 반영할 수 있도록 SDXL 모델을 기반으로 LoRA와 DreamBooth 기법을 모두 활용하여 파인튜닝을 진행하였고 필수 파라미터를 정제함으로써 이미지 생성 모델의 최적화를 진행하였다.

학습을 위하여 사이즈 1024x1024 pixel 크기의 이미지를 사용하였으며 학습 에포크(epoch)는 20, 배치 사이즈(Btch size)는 1, 학습률(Learning Rate) 5.0e-06, 옵티마이저(Optimizer)는 AdamW, 손실 함수(Loss Function): motion_mask_loss, 랭크(lora_rank)는 16을 사용하였다.

대중적으로 잘 알려진 샘플링 방법인 DPM++ 3M SDE[24], LCM[24], Euler a[24] 방법으로 테스트하였으며 이중 Euler a 샘플링 방법이 삐아, 빵야 캐릭터와 배경을 가장 잘 표현하는 것으로 나타났다. Sampling Steps[24]는 적어도 40 이상의 값을 사용하는 것이 캐릭터와 배경의 세부 묘사를 잘 만들어 내며 이미지가 크게 왜곡되지 않음을 확인할 수 있었다. CFG Scale[24]은 7 이상의 값을 사용해야 프롬프트의 내용을 잘 반영하였으며, 10 이하로 설정하는 것이 왜곡되지 않는 이미지를 생성할 수 있음을 알 수 있었다. [그림 8]은 같은 입력 프롬프트를 기반으로 하이퍼파라미터를 달리하여 생성된 이미지를 보여준다.

1.2 보도 그래픽

보도 그래픽의 경우 SDXL 모델을 기반으로 LoRA 기법만을 활용하여 파인튜닝을 진행하였다.

학습을 위하여 사이즈 1024x1024 pixel 크기의 이미지를 사용하였으며 학습 에포크(epoch)는 1, 배치 사이즈(Btch size)는 16, 학습률(Learning Rate) 0.0001, 옵티마이저(Optimizer)는 Adamw8bit, 손실 함수(Loss Function)는

Prompt	A picture of a ppia and a ppangya standing in a subway station		
Hyperparameters	Sampling steps=50, CFG Scale=6		
Generation Results			
Sampling method	Euler a	DPM++ 3M SDE	LCM
Hyperparameters	Sampling Method=Euler a, CFG Scale=6		
Generation Results			
Sampling steps	20	40	60
Hyperparameters	Sampling Method=Euler a, Sampling steps=60		
Generation Results			
CFG Scale	5	7	9

그림 8. 애니메이션 그래픽 최적화
Fig. 8. Optimization of animated graphics

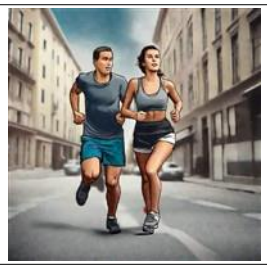
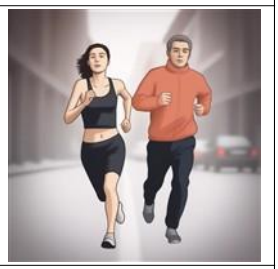
Prompt	man and woman are running on the street.		
Hyperparameters	Network Rank = 8(default)		
Generation Results			
Sampling method	Euler a	DPM++ 3M SDE	LCM
Hyperparameters	Network Rank = 128		
Generation Results			
Sampling method	Euler a	DPM++ 3M SDE	LCM
Hyperparameters	Network Rank = 256		
Generation Results			
CFG Scale	Euler a	DPM++ 3M SDE	LCM

그림 9. 보도 그래픽 최적화
Fig. 9. Optimization of press graphics

l2, 랭크(Network Rank)는 256을 사용하였다. Network Rank가 높을수록, 객체를 또렷하게 강조하는 굵은 선과 수채화풍의 KBS 스타일을 잘 반영하는 것으로 나타났다. 샘플링 기법으로는 DPM++ 3M SDE, LCM, Euler a 중 DPM++ 3M SDE 방법이 좋은 성능을 보였다. 따라서 Network Rank=256, DPM++ 3M SDE을 최종 모델로 선정하였다. [그림 9]는 같은 입력 프롬프트를 기반으로 하이퍼파라미터를 달리하여 생성된 이미지를 보여준다.

보도 그래픽의 경우 데이터 라벨링 과정에서 트리거 워드의 효용성을 실험하기 위해 KBSstyle로 라벨링을 진행하였다. 프롬프트 작성 시 KBSstyle의 트리거 워드를 추가한 것과 추가하지 않은 것을 비교 분석하였다. 그 결과 KBSstyle을 추가하여 생성한 이미지가 좀 더 KBS 콘텐츠의 스타일을 잘 반영하는 것을 확인할 수 있었으며, 트리거 워드만으로도 특정 그래픽의 톤과 스타일을 반영함을 알 수 있었다.

상용 도구를 사용한 이미지 생성에 있어 폭력성이 짙은 프롬프트의 사용은 생성형 알고리즘의 윤리적인 이슈로 생성되지 않는 경우가 많다. 그러나 보도 방송에서는 다소 불편한 내용도 뉴스에서 다뤄져야 하는 부분이 있으므로 폭력적인 단어를 포함하여 이미지를 생성하는 실험을 진행하였다. 실험 결과 [그림 11]과 같이 시각적으로 안전한 선에서 사건의 맥락을 잘 표현하여 생성하는 것을 확인할 수 있었다.

	Without KBSstyle	With KBSstyle
Generation Results		
Prompt	A masked robber hijacked a car at gunpoint in broad daylight, leaving the driver terrified and stranded by the roadside	
Generation Results		
Prompt	a group of people sitting around a table in a dimly lit room with their hands on the table	

그림 10. KBSstyle 트리거 단어 유무에 따른 이미지 생성
Fig. 10. Image generation based on the presence or absence of KBSstyle trigger words

Generation Results		
Prompt	a man assaults a woman, KBSstyle	a man is threatening a woman with a weapon in a dark place, KBSstyle

그림 11. 폭력적인 단어를 포함한 프롬프트 기반 이미지 생성
Fig. 11. Prompt-based image generation including violent words

2. 동영상 생성

오픈 소스인 AnimateAnything 모델을 사용하여 비디오를 생성하였다. 모델 특성상 영상의 가로 세로 비율이 1:1인 512x512 사이즈에 최적화되어 있으나, 16:9 비율인 1024x576 사이즈로 출력 후 업스케일을 통해 방송 송출 가능한 동영상을 생성할 수 있음을 확인하였다. 또한 프레임 수를 높게 설정할수록 조금 더 자연스럽고 모션 강도가 높은 역동적인 영상이 생성되는 것을 확인할 수 있었다.

<table>
<tr><td colspan="3">Animated graphic</td></tr>
<tr><td>Prompt</td><td colspan="2">a ppia and a ppangya are singing in the musium</td></tr>
<tr><td>  </td><td>  </td><td>  </td></tr>
<tr><td>input image</td><td>resolution 1024X576</td><td>resolution 1920X1080</td></tr>
<tr><td colspan="3">News graphic</td></tr>
<tr><td>Prompt</td><td colspan="2">A warm moment of a couple holding their baby indoors, in a cozy living room with natural light streaming in, happy expressions on the parents' faces, the baby wrapped in a soft blanket</td></tr>
<tr><td>  </td><td>  </td><td>  </td></tr>
<tr><td>input image</td><td>14 frame</td><td>30 frame</td></tr>
</table>

그림 12. 애니메이션 그래픽과 보도 그래픽의 동영상 생성
Fig. 12. Video generation of animated graphics and press graphics

3. 조건부 반영 생성

사용자가 원하는 세부적인 스타일이나 특정 요소를 정확히 구현하기 위해 ControlNet과 같은 조건부 생성 기법을 도입하여 프롬프트 기반 그래픽 생성의 정밀도를 높였다. Canny Edge, Depth map, Open Pose의 조건을 지원하는 ControlNet을 도입하였으며, 그 결과는 [그림 13]과 같다.

4. 상용 도구와 비교

같은 이미지와 프롬프트를 입력으로 하여 파인튜닝 모델과 상용 도구의 생성 결과를 비교하였다.

대표적인 이미지 생성 도구인 Midjourney, DALL-E, Wrtn[25]과 파인튜닝 모델을 비교하였다. 애니메이션 그래픽의 경우 프롬프트를 그대로 입력하면 캐릭터가 생성되지 않기 때문에 캐릭터의 세부적인 특징을 추가로 입력하였다. 그 결과 상용 서비스를 사용한 이미지 생성에서 삐아와 빵야의 캐릭터 특징을 추가적으로 프롬프트에 작성하였음에도 불구하고 캐릭터의 특징을 잘 반영하지 못함을 확인할 수 있었다.

보도 그래픽의 경우 상용 서비스에 프롬프트를 그대로 입력하면 일러스트 이미지가 아닌 실사의 사실적인 이미지가 생성되기 때문에 이미지 스타일에 대한 정보를 추가로 입력하였다. 그 결과 실사 이미지가 아닌 그래픽 이미지는 생성하지만 KBS 콘텐츠 스타일을 반영한 이미지는 생성할 수 없음을 확인할 수 있었다.

다음은 동영상 생성을 비교하였다. 대표적인 동영상 생성 도구인 Gen3, Dream Machine, KlingAI와 파인퓨닝 모델을 비교하였다. 이미지 생성 파인튜닝 모델에서 생성된 이미지를 입력으로 같은 텍스트 프롬프트를 입력하였다. 애니메이션 그래픽의 경우 캐릭터의 눈이 갑자기 커지거나 날개가 두, 세개로 표현되는 등 일관성이 무너지는 경향을 확인할 수 있었다. 보도 그래픽의 경우도 마찬가지로 새로운 사람들이 갑자기 만들어지거나 배경의 버스가 실사로 바뀌는 등의 일관성이 깨지는 현상이 발생하였다.

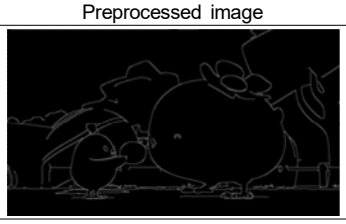
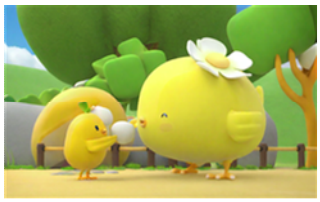
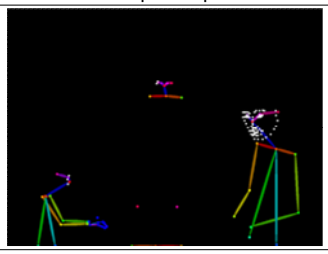
<table>
<tr><td colspan="3">Animated graphic</td></tr>
<tr><td>Prompt</td><td colspan="2">ppia the yellow bird and ppangya the yellow bird are standing in the park</td></tr>
<tr><td>input image</td><td>Preprocessed image</td><td>Generation Results</td></tr>
<tr><td rowspan="3">  </td><td>  </td><td rowspan="2">  </td></tr>
<tr><td>Canny Edge</td></tr>
<tr><td>  </td><td>  </td></tr>
<tr><td></td><td>Depth map</td><td></td></tr>
<tr><td colspan="3">News graphic</td></tr>
<tr><td>Prompt</td><td colspan="2">KBSstyle, A picture of a lawyer talking in front of a judge</td></tr>
<tr><td>input image</td><td>Preprocessed image</td><td>Generation Results</td></tr>
<tr><td rowspan="5">  </td><td>  </td><td rowspan="2">  </td></tr>
<tr><td>Canny Edge</td></tr>
<tr><td>  </td><td rowspan="2">  </td></tr>
<tr><td>Depth map</td></tr>
<tr><td>  </td><td rowspan="2">  </td></tr>
<tr><td>Open Pose</td></tr>
</table>

그림 13. 조건부 생성 모델 적용 결과

Fig. 13. Results of applying conditional generation model

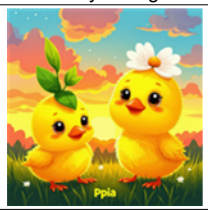
Animated graphic				
Prompt	A picture of a ppia and a ppangya sitting in the field and looking at the sunset, beautiful sky background			
Generation Results				
	Fine tuning model	Midjourney	DALL-E	Wrtn
News graphic				
Prompt	A picture of a young child crossing a crosswalk and a bus coming			
Generation Results				
	Fine tuning model	Midjourney	DALL-E	Wrtn

그림 14. 상용 도구와 파인튜닝 모델의 생성 이미지 비교
Fig. 14. Comparison of generated images of commercial tools and fine-tuning models

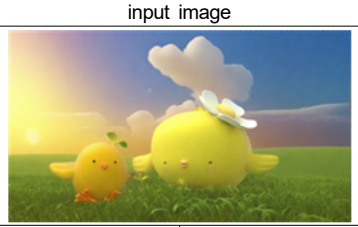
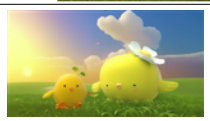
Animated graphic				
Prompt	input image 		A picture of a ppia and a ppangya sitting in the field and looking at the sunset, beautiful sky background	
Generation Results				
	Fine tuning model	Gen3	KlingAI	Dream Machine
News graphic				
Prompt	input image 		A picture of a young child crossing a crosswalk and a bus coming	
Generation Results				
	Fine tuning model	Gen3	KlingAI	Dream Machine

그림 15. 상용 도구와 파인튜닝 모델의 생성 이미지 비교
Fig. 15. Comparison of generated images of commercial tools and fine-tuning models

5. 사용자 중심 시스템 설계

그라디오(Gradio)[26]는 Python으로 개발된 오픈 소스 패키지로 쉽게 커스터마이징할 수 있는 UI 구성 요소를 빠르게 생성할 수 있는 장점이 있다. 따라서 사용자의 편의성을 고려하여 기능의 추가가 용이한 그라디오 기반의 웹 인터페이스를 설계하였다.

해당 인터페이스는 사용자가 텍스트 입력을 통해 그래픽 스타일과 세부 요소를 쉽게 설정할 수 있도록 설계되었으며, 생성된 이미지와 비디오를 실시간으로 미리보기할 수 있는 기능을 제공한다.

또한, 협업 부서의 피드백을 반영하여 인터페이스의 직관성과 활용성을 더욱 향상시켰다. Ollama의 llama3_ifai_sd_prompt_mkr_q4km[27] 모델을 사용하여, 한국어 입력 프롬프트를 스테이블 디퓨전 프롬프트 형식으로 자동으로 변환하는 기능을 추가하여 사용자의 편의성을 향상시켰다.

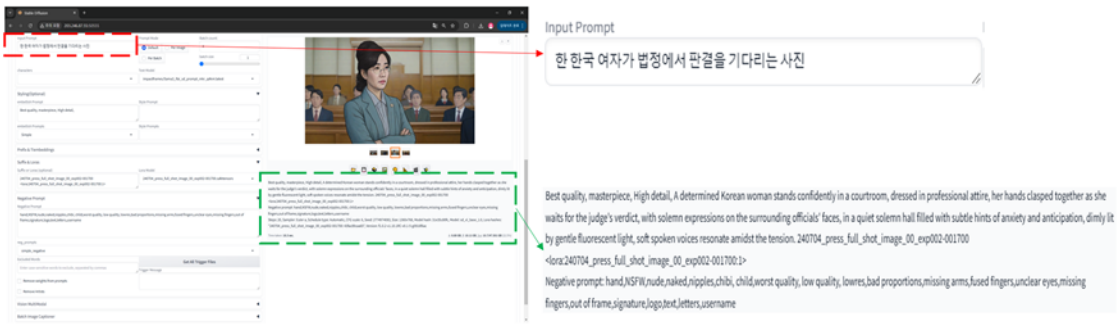


그림 16. 입력 프롬프트의 스테이블 디퓨전 프롬프트 형식으로 자동 변환
Fig. 16. Automatic conversion of input prompt to Stable Diffusion prompt format

그림 17. 한국어 입력 프롬프트의 영문 변환 자동화
Fig. 17. Automating English conversion of Korean input prompts

또한 Ollama의 llama-3-Korean-Bllossom-8B[28] 모델을 사용하여, 한국어 입력 프롬프트를 영문으로 자동 번역하는 기능을 추가하였다. 애니메이션의 경우 한글로 프롬프트를 입력하면, 고유명사인 삐아(PPIA), 빵야(PPANGYA)를 포함하여 영문으로 번역되고 보도 그래픽의 경우 한글 프롬프트를 입력하면, 영문으로 자동 번역되어 입력한다.

동영상 생성 출력 시 motion model의 임계값을 달리하여 총 5개의 영상을 생성하여 사용자로 하여금 선택의 폭을 넓게 하여 선택할 수 있도록 설계하였다. 추가적으로, API를 통해 외부 시스템과의 연동을 지원하여 다양한 제작 환경에서의 활용 가능성을 확보하였다.

6. 성능 평가 분석

본 연구에서는 생성형 AI 기반 방송 그래픽 제작 시스템의 성능을 평가하기 위해 Mean Opinion Score(MOS) 방식을 활용하였다. MOS 평가는 사용자가 직접 AI가 생성한 이미지를 보고, 주어진 기준에 따라 점수를 부여하는 방식으로 진행되었다. 이를 통해 AI가 원본 이미지와 유사한 스타일을 유지하고 있는지, 그리고 프롬프트(지시문)의 내용

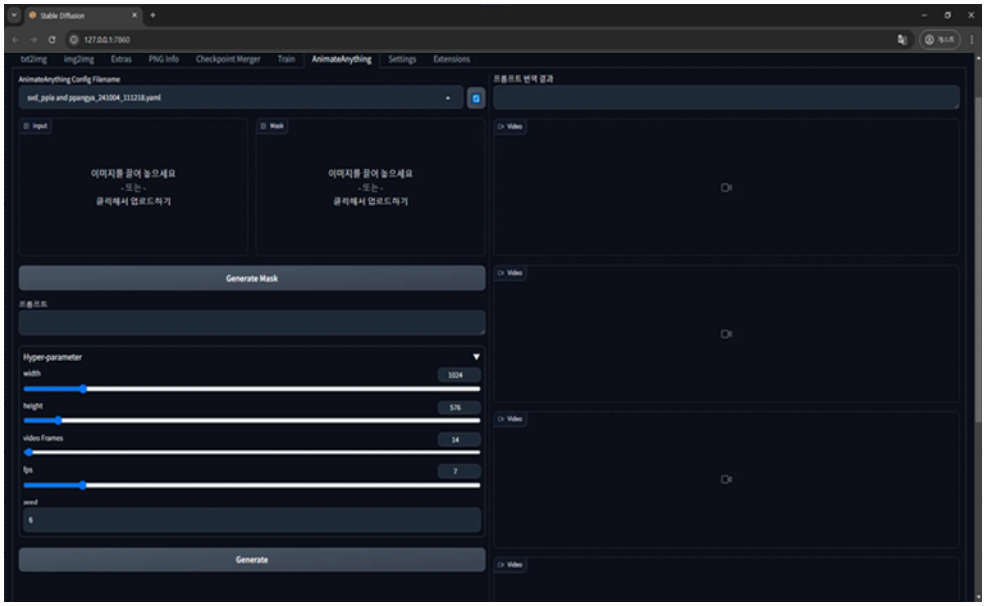


그림 18. 동영상 생성 UI
Fig. 18. Video generation UI

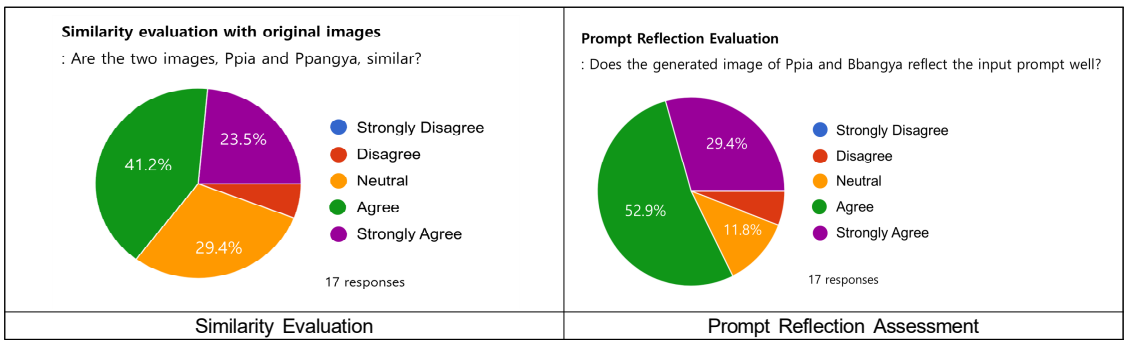


그림 19. 문항별 응답결과
Fig. 19. Response results by question

을 정확히 반영하고 있는지를 평가하였다.

본 평가의 신뢰성을 확보하기 위해, 이미지 또는 동영상 생성 관련 상용 도구를 사용한 경험이 있는 평가자를 대상으로 설문을 진행하였다. 평가 대상자는 서울과학기술대학교 문예창작학과 및 스마트ICT융합공학과의 학부생과 교수진 총 17명으로 구성되었다.

설문 조사는 두 가지 핵심 항목으로 구성되었다. 첫 번째는 원본 이미지와의 유사도 평가이며, 두 번째는 프롬프트 반영 정도 평가이다. 평가에 참여한 응답자들은 각 항목에 대해 5점 척도(1점: 전혀 그렇지 않다 ~ 5점: 매우 그렇다)로 점수를 부여하였다. 설문 문항은 총 4개로 구성되었으며, 각 문항마다 원본 이미지와 AI 생성 이미지를 나란히 배치하여 평가자가 쉽게 비교할 수 있도록 하였다.

총 4개 문항의 유사도 평가 문항에서는 평균 3.72점으로 원본 이미지와 생성 이미지가 유사하다고 답했으며, 총 4개 문항의 프롬프트 반영 정도 평가에서는 평균 4.08점으로 개발된 시스템에서 프롬프트를 통해 이미지를 잘 생성했다고 평가하였다.

이는 생성된 그래픽이 실제 방송 제작 환경에서 활용 가능할 정도의 품질을 갖추었음을 나타내며 방송 제작 과정에서 AI를 도입할 경우, 그래픽 제작의 효율성과 생산성이 높아질 가능성이 크다는 점을 시사한다고 볼 수 있다. 그러나 유사도 평가 점수가 프롬프트 반영 정도 평가 점수 보다 낮은 것에서 알 수 있듯이 기존 상용 AI 도구 대비 방송국의 고유한 스타일을 반영하는 능력이 뛰어나지만, 완전히 유사하게 생성해내지 못함을 의미한다. 따라서 특정 세부 요소를 더욱 정밀하게 구현할 수 있도록 추가적인 데이터 학습과 최적화 과정의 개선이 필요하다는 것을 알 수 있었다.

Ⅴ. 결 론

본 연구는 생성형 인공지능을 활용하여 방송 그래픽 제작에 최적화된 맞춤형 시스템을 개발하고, 이를 통해 제작 효율성과 품질을 향상시킬 수 있는 가능성을 검증하였다. 연구 결과, 기존 상용 AI 그래픽 생성 도구들이 특정 방송국이나 프로그램의 고유한 스타일을 반영하는 데 한계가 있음을 확인하였으며, 이를 극복하기 위해 SDXL 모델을 기반으로 LoRA 및 Dreambooth 기법을 활용한 파인튜닝 모델을 개발하였다.

애니메이션 그래픽과 보도 그래픽을 대상으로 한 실험 결과, 텍스트 프롬프트를 기반으로 방송사의 스타일을 유지하면서도 세부적인 그래픽 요소를 효과적으로 생성할 수 있음을 확인하였다. 특히 Euler a 및 DPM++ 3M SDE 샘플링 기법을 사용하고 하이퍼파라미터 최적화 과정을 통해 그래픽 품질 향상에 기여하였으며, ControlNet을 활용한 조건부 생성 기능은 사용자가 원하는 세부 요소를 보다 정밀하게 구현하였다.

본 연구는 또한 트리거 워드를 활용하여 특정 방송사의 스타일을 쉽게 반영할 수 있는 시스템을 구축하였다. KBSstyle과 같은 트리거 워드를 프롬프트에 추가하는 방식만으로 그래픽의 일관성과 톤을 유지하는 데 효과적이었으며, 특정 스타일을 간편하게 적용할 수 있는 장점을 제공하였다. 이 과정에서 폭력적인 단어나 사회적 사건과 관련된 민감한 키워드를 사용해 보도 그래픽을 생성하는 실험도 진행되었다. 그 결과, 실제 뉴스나 보도 상황을 반영할 수 있는 시각적 콘텐츠를 윤리적 기준 내에서 생성할 수 있음을 확인하였다.

동영상 생성 부분에서는 AnimateAnything 모델을 파인튜닝하여 일관성 있는 결과물을 생성할 수 있었으며, 이를 통해 방송 제작 환경에서 요구되는 고품질 비디오 제작의 가능성을 제시하였다. 상용 도구와의 비교 실험에서도 파인튜닝된 모델이 특정 캐릭터 및 그래픽 스타일을 보다 잘 반영하는 것을 확인할 수 있었다.

또한, 본 연구는 사용자 중심의 시스템 설계에도 중점을 두었다. 사용자가 텍스트 입력을 통해 그래픽 스타일과 세부 요소를 쉽게 설정할 수 있도록 그라디오 기반의 직관적인 웹 인터페이스를 설계하였으며, 생성된 이미지와 동영상을 실시간으로 미리보기할 수 있는 기능을 추가하였다. 한국어로 입력된 프롬프트를 스테이블 디퓨전 형식으로 자동 변환하는 기능을 도입하여 사용자 편의성을 높였으며, 동영상 생성 시 다양한 프레임 옵션을 제공하여 선택의 폭을 넓혔다. API 연동을 통해 외부 시스템과의 협업 및 확장 가능성도 확보하였다.

성능 평가 결과, Mean Opinion Score(MOS)에서 평균

3.9점을 기록하며, 개발된 시스템이 실제 방송 그래픽 제작 현장에서 활용될 수 있는 수준의 품질을 갖추고 있음을 입증하였다.

그러나 연구 과정에서 몇 가지 한계도 발견되었다. 트리거 워드를 활용한 스타일 반영이 일정 수준에서 유효하게 작용했지만, 다양한 스타일의 그래픽에서 완벽한 일관성을 보장하지는 못했다. 또한, AnimateAnything 모델을 활용한 동영상 생성 실험에서는 프레임 간 일관성 부족과 특정 개체의 지속성 유지가 어려운 문제가 확인되었으며, 이는 향후 개선이 필요한 과제로 남았다. 또한, 보도 그래픽 제작은 긴급성을 요하는 업무 특성상 정확한 마감시간을 보장하는 시스템이 필요하며, AI가 생성하는 이미지에 대한 검수 과정이 필수적이다. 보도 그래픽 실무자들은 AI가 제작한 이미지가 뉴스 콘텐츠의 신뢰성과 정확성을 유지할 수 있도록 맞춤법 검수, 얼굴 인식 검수 등의 보완 기능이 필요하다는 점을 강조하였다.

본 연구는 방송 제작 환경에서 생성형 AI의 적용 가능성을 실질적으로 입증하는 동시에, 향후 발전 방향을 제시한다. 첫째, 특정 방송사의 스타일을 더욱 정밀하게 반영할 수 있도록 트리거 워드와 조건부 생성 기법을 결합한 최적화된 스타일 반영 기술을 개발해야 한다. 둘째, Image-to-Video 변환 모델의 개선을 통해 프레임 간 일관성을 유지하고 방송 제작에 적합한 고해상도 동영상 생성 기술을 적용할 필요가 있다. 셋째, 실제 방송 제작 환경에서 AI 그래픽 시스템의 실무 적용성을 검증하기 위해 다양한 형태의 프로그램과 협업한 테스트가 이루어져야 한다. 마지막으로, AI 기반 그래픽 생성이 윤리적 기준을 준수하면서도 보도 자료 제작에 적절히 활용될 수 있도록 윤리적 생성 가이드라인을 마련하는 것이 필요하다.

결과적으로 본 연구는 생성형 AI가 방송 그래픽 제작을 보다 효율적이고 정밀하게 수행할 수 있는 가능성을 보여주었으며, 향후 연구에서는 스타일 반영 기술의 고도화, 동영상 생성 모델 개선, 실무 환경에서의 적용 검증, 그리고 윤리적 문제 해결을 중심으로 발전이 이루어져야 할 것이다. 이를 통해 생성형 AI가 방송 제작 현장에서 실질적으로 활용될 수 있는 핵심 기술로 자리 잡을 수 있을 것으로 기대한다.

참 고 문 헌 (References)

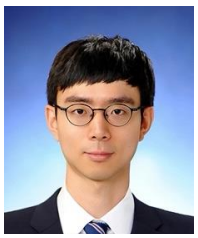
[1] D. Best, “Can creativity be taught?,” British Journal of Educational Studies, vol. 30, no. 3, pp. 280–294, 1982.
[2] Midjourney, “Midjourney Official Website,” Available: https://www.midjourny.com/home (Accessed: Feb. 17, 2025).
[3] OpenAI, “DALL·E 3,” Available: https://openai.com/index/dall-e-3/ (Accessed: Feb. 17, 2025).
[4] OpenAI, “Sora,” Available: https://openai.com/sora/ (Accessed: Feb. 17, 2025).
[5] Runway ML, “Introducing Gen-3 Alpha,” Available: https://runwayml.com/research/introducing-gen-3-alpha (Accessed: Feb. 17, 2025).
[6] Luma Labs, “Dream Machine,” Available: https://lumalabs.ai/dream-machine/creations (Accessed: Feb. 17, 2025).
[7] Kling AI, “Kling AI Official Website,” Available: https://klingai.com/ (Accessed: Feb. 17, 2025).
[8] Podell, Dustin, Zion English, Kyle Lacey, Andreas Blattmann, Tim Dockhorn, Jonas Muller, Joe Penna, and Robin Rombach. “SDXL: Improving Latent Diffusion Models for High-Resolution Image Synthesis.” International Conference on Learning Representations (2024). https://arxiv.org/abs/2307.01952
[9] Hu, Edward J., et al. “LoRA: Low-rank adaptation of large language models.” arXiv preprint arXiv:2106.09685 (2021). doi: https://doi.org/10.48550/arXiv.2106.09685
[10] Nataniel Ruiz, Yuanzhen Li, Varun Jampani, Yael Pritch, Michael Rubinstein, Kfir Aberman. “DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation.” CVPR (2023). doi: https://doi.org/10.48550/arXiv.2208.12242
[11] Goodfellow, Ian, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. “Generative Adversarial Networks.” Advances in Neural Information Processing Systems 27 (2014). https://arxiv.org/abs/1406.2661
[12] Ho, Jonathan, Ajay Jain, and Pieter Abbeel. “Denoising Diffusion Probabilistic Models.” Advances in Neural Information Processing Systems 33 (2020): 6840–6851. https://arxiv.org/abs/2006.11239
[13] Radford, Alec, Luke Metz, and Soumith Chintala. “Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks.” International Conference on Learning Representations (2016). https://arxiv.org/abs/1511.06434
[14] Karras, Tero, Samuli Laine, and Timo Aila. “A Style-Based Generator Architecture for Generative Adversarial Networks.” IEEE Transactions on Pattern Analysis and Machine Intelligence 43, no. 6 (2021): 4217–4231. https://arxiv.org/abs/1812.04948
[15] Brock, Andrew, Jeff Donahue, and Karen Simonyan. “Large Scale GAN Training for High Fidelity Natural Image Synthesis.” International Conference on Learning Representations (2019). https://arxiv.org/abs/1809.11096
[16] Ho, Jonathan, Ajay Jain, and Pieter Abbeel. “Denoising Diffusion Probabilistic Models.” Advances in Neural Information Processing Systems 33 (2020): 6840–6851. https://arxiv.org/abs/2006.11239
[17] Nichol, Alex, and Prafulla Dhariwal. “Improved Denoising Diffusion Probabilistic Models.” International Conference on Machine Learning

(2021): 8162 – 8171. https://arxiv.org/abs/2102.09672
[18] Rombach, Robin, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. “High-Resolution Image Synthesis with Latent Diffusion Models.” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2022): 10684 – 10695. https://arxiv.org/abs/2112.10752
[19] Weiming Ren, Huan Yang, Ge Zhang, Cong Wei, Xinrun Du, Wenhao Huang, Wenhu Chen. “ConsistI2V: Enhancing Visual Consistency for Image-to-Video Generation.” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2024). doi: https://doi.org/10.48550/arXiv.2402.04324
[20] Guojun Lei, Chi Wang, Hong Li, Rong Zhang, Yikai Wang, Weiwei Xu,“AnimateAnything: Consistent and Controllable Animation for Video Generation.” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2024). doi: https://doi.org/10.48550/arXiv.2411.10836
[21] Guo, Yuwei, Tao Yang, Zeyu Chen, Jingjing Chen, and Qingyao Wu. “AnimateDiff: Animate Your Personalized Text-to-Image Diffusion Models without Specific Tuning.” arXiv preprint (2023). https://arxiv.org/abs/2307.04725
[22] Zhang, Lvmin, and Maneesh Agrawala. “Adding Conditional Control to Text-to-Image Diffusion Models.” arXiv preprint (2023). https://arxiv.org/abs/2302.05543
[23] Junsong Chen, Jincheng Yu, Chongjian Ge, Lewei Yao, Enze Xie, Yue Wu, Zhongdao Wang, James Kwok, Ping Luo, Huchuan Lu, Zhenguo Li. “PixArt-α: Fast Training of Diffusion Transformer for Photorealistic Text-to-Image Synthesis.” International Conference on Learning Representations (2024). doi: https://doi.org/10.48550/arXiv.2310.00426
[24] All About Stable Diffusion, https://generative-ai-everything.tistory.com/entry/스테이블-디퓨전Stable-diffusion-매개변수parameters-알아보기
[25] Wrtn, “AI Writing Assistant,” Available: https://wrtn.ai/ (Accessed: Feb. 17, 2025).
[26] Gradio, “Gradio Open Source,” Available: https://github.com/gradio-app/gradio (Accessed: Feb. 17, 2025).
[27] Ollama, “Llama3 Impact Frames,” Available: https://ollama.com/impactframes/llama3_ifai_sd_prompt_mkr_q4km (Accessed: Feb. 17, 2025).
[28] Hugging Face, “Llama-3 Korean Blossom 8B,” Available: https://huggingface.co/MLP-KTLim/llama-3-Korean-Bllossom-8B (Accessed: Feb. 17, 2025).

저 자 소 개

이 윤 재

- 2006년 2월 : 연세대학교 전기전자공학부 공학사
- 2012년 2월 : 연세대학교 정보대학원 석사
- 2019년 3월 ~ 현재 : 연세대학교 기술경영협동과정 박사과정
- 2016년 1월 ~ 현재 : 한국방송공사 미디어연구소 AI전환팀 팀장
- ORCID : https://orcid.org/0000-0001-6342-6947
- 주관심분야 : Computer Vision, AI Transformation, Innovation Management



이 용 건

- 2007년 2월 : 연세대학교 전기전자공학부 공학사
- 2013년 2월 : 연세대학교 전기전자공학부 석사
- 2016년 1월 ~ 현재 : 한국방송공사 미디어연구소 AI전환팀 책임연구원
- ORCID : https://orcid.org/0009-0007-5955-9203
- 주관심분야 : Computer Vision, GenaI, Visual-Language-Model

저 자 소 개

이 영 화

- 2001년 8월 ~ 2021년 12월 : 현대홈쇼핑 방송그래픽팀 NLE 감독
- 2024년 2월 : 서울과학기술대학교 전자IT미디어공학과 박사수료
- 2022년 1월 ~ 현재 : 서울과학기술대학교 산학협력단 IoT융복합연구소 책임연구원
- ORCID : https://orcid.org/0000-0002-2427-7667
- 주관심분야 : Computer Vision & Graphics, Generative AI, Video-Language pre-training

양 지 희

- 2012년 2월 : 강릉원주대학교 멀티미디어공학과 공학사
- 2015년 2월 : 서울과학기술대학교 미디어IT공학과 석사
- 2015년 3월 ~ 2020년 8월 : 서울과학기술대학교 나노IT디자인융합대학원 정보통신미디어공학과 박사
- 2020년 3월 ~ 현재 : 서울과학기술대학교 IoT융복합기술연구소 선임연구원
- ORCID : https://orcid.org/0000-0003-2776-5487
- 주관심분야 : 컴퓨터 비전, 실감미디어

이 승 주

- 2016년 2월 : 상명대학교 컴퓨터소프트웨어공학과 공학사
- 2020년 8월 : 서울과학기술대학교 스마트ICT융합공학과 공학석사
- 2023년 9월 ~ 현재 : 서울과학기술대학교 스마트ICT융합공학과 박사과정
- ORCID : https://orcid.org/0000-0001-8969-4296
- 주관심분야 : 컴퓨터비전, Video Analytics, Object Detection

강 현 석

- 2015년 3월 ~ 2021년 8월 : 강원대학교 삼척캠퍼스 전자공학과 학사
- 2021년 9월 ~ 2023년 8월 : 서울과학기술대학교 일반대학원 스마트ICT융합공학과 석사
- 2023년 9월 ~ 현재 : 서울과학기술대학교 일반대학원 스마트ICT융합공학과 박사과정
- ORCID : https://orcid.org/0000-0003-0783-3841
- 주관심분야 : 3D 컴퓨터 비전, 딥러닝

박 구 만

- 1984년 2월 : 한국항공대학교 전자공학과
- 1986년 2월 : 연세대학교 전자공학과 공학석사
- 1991년 2월 : 연세대학교 전자공학과 공학박사
- 1991년 3월 ~ 1996년 9월 : 삼성전자 신호처리연구소 선임연구원
- 2016년 1월 ~ 2017년 12월 : 서울과학기술대학교 나노IT디자인융합대학원
- 1999년 8월 ~ 현재 : 서울과학기술대학교 스마트ICT융합공학과 교수
- 2006년 1월 ~ 2007년 8월 : Georgia Institute of Technology Dept.of Electrical and Computer Engineering, Visiting Scholar
- ORCID : https://orcid.org/0000-0002-7055-5568
- 주관심분야 : 컴퓨터비전, 지능형실감미디어